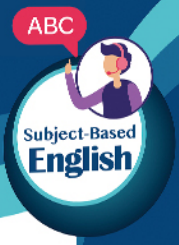
高等学校专业英语教材

电子信息工程专业英语教程（第6版）

◎ 任治刚　编著



中国工信出版集团
電子工業出版社

高等学校专业英语教材
电子信息工程专业英语教程（第6版）
◎ 任治刚　编著
電子工業出版社

高等学校专业英语教材

电子信息工程专业英语教程

(第6版)

任治刚　编著

電子工業出版社
Publishing House of Electronics Industry
北京·BEIJING

内 容 简 介

本书的主要目的是使读者掌握电子信息工程专业英语术语及用法，培养、提高读者阅读和翻译专业英语文献资料的能力。本书由10个单元组成，涵盖了电子信息领域的主要技术分支，主要内容包括电子器件、电子电路、电子系统组件、电子系统、现代数字设计、数字信号处理技术、语音和音频、图像和视频、嵌入式应用、仪器与测量等。每个单元由3篇课文、3篇阅读材料、课文词汇、课文注释和单元练习等组成。每篇课文均有参考译文，书后附有练习参考答案、科技英语语音手册和科技英语词汇手册；全书配有电子教案、视频材料等教辅资源，通过华信教育资源网(www.hxedu.com.cn)免费提供给授课教师。

本书既可作为电子信息工程专业的英语教材，也可供从事相关专业的工程技术人员使用。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目（CIP）数据

电子信息工程专业英语教程 / 任治刚编著. -- 6版.
北京 : 电子工业出版社, 2025. 2. -- ISBN 978-7-121
-49754-4

Ⅰ. G203

中国国家版本馆CIP数据核字第2025Z4X864号

责任编辑：秦淑灵　　特约编辑：张燕虹
印　　刷：
装　　订：
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：21.75　字数：724千字
版　　次：2004年7月第1版
　　2025年2月第6版
印　　次：2025年2月第1次印刷
定　　价：69.00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)88254888，88258888。

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

本书咨询联系方式：qinshl@phei.com.cn。

前　言

电子信息工程是国内外飞速发展的工程领域之一。为了应对国际化竞争，大学生必须在学习阶段打下坚实的专业基础，发展全面的职业技能。“熟练运用专业英语、有效进行科技交流”是电子信息工程专业学生的重要专业素养和必备职业技能。

为了提高学生的专业英语阅读、写作能力，拓展学生对电子信息工程关键技术的认识，培养学生成为具备国际竞争力的技术人才，本书充分吸收最新科技成果和英语教学成果，按照先进实用的选材原则和简明系统的组织原则，为电子信息工程专业学生构建了一个提高英语水平和专业素养的平台。

本书共10个单元，涵盖了电子信息工程领域的主要技术分支。每个单元由3篇课文、3篇阅读材料和单元练习等组成。课文侧重展示基础理论和核心技术；阅读材料着力介绍历史沿革、实用技术和未来前景；单元练习采用完形填空、英汉互译、摘要写作的方式训练语言技能。

本书10个单元包含的主要技术内容如表0.1所示。

表0.1　本书10个单元包含的主要技术内容

单元号	单元名	主要技术内容
1	电子器件	VLSI(超大规模集成)技术、存储器、微处理器、制造集成电路、闪存、微控器等
2	电子电路	运算放大器、低通滤波器、模数转换器(ADC)、开关电容滤波器、数模转换器(DAC)等
3	电子系统组件	开关电源、时钟信号源、互连部件、线缆、总线、USB(通用串行总线)等
4	电子系统	移动电话系统(Ⅰ)、移动电话系统(Ⅱ)、个人计算机(PC)系统、GPS(全球定位系统)、802.11无线局域网(LAN)、机电一体化系统等
5	现代数字设计	FPGAs、VHDL、PLD发展史、Verilog HDL、SoC等
6	数字信号处理技术	数字信号处理(DSP)、数字信号处理器、典型DSP应用、ASP或DSP、软件无线电、医学图像处理等
7	语音和音频	高保真音频、MP3音频压缩、移动电话系统(Ⅲ)、声码器等
8	图像和视频	数字图像基础、数码相机、电视视频信号、JPEG、MPEG、互联网视频等
9	嵌入式应用	选择合适的微处理器内核、嵌入式系统的设计语言、选择实时操作系统、ARM、嵌入式OS等
10	仪器与测量	信号源、示波器、逻辑分析仪、信号完整性、虚拟仪器等

第1～4单元按照“器件”(devices)→“电路”(circuits)→“系统”(systems)的思路，全景展示了一个典型电子信息系统的构建过程。关于“器件”，我们选择了当今广泛应用的两种VLSI芯片——微处理器和存储器，它们是数字时代电子系统的核心。关于“电路”，我们选择了模拟电路的代表——运算放大器和低通滤波器，以及混合电路的代表——ADC和DAC，它们是数字时代电子系统的外围设备。众多功能电路可以构成“系统组件”(system components)，我们选择了任何电子系统都离不开的“电源设备”(power supplies)、“时钟信号源”(clock sources)和“互连部件”(interconnects)，它们分别相当于系统的“血液”“心脏”“筋骨”。关于“系统”，我们

选择了移动电话、PC、GPS 等。

第 5～6 单元聚焦两种数字电子系统核心的设计手段。一方面，可以利用 VHDL、Verilog 等硬件设计语言，在可编程的 FPGA 和 CPLD 芯片上，设计定制的数字电子系统核心；另一方面，可以利用汇编语言、C 语言等软件设计语言，在可编程的 DSP 芯片上，设计功能强大的信号处理核心。在工程实际中，这两种方式往往配合使用。

第 7～8 单元通过介绍 CD、MP3 等具体应用，展示了电子信息工程在音频、视频、图像等多媒体领域的发展成果，同时反映出多姿多彩的设计思路和理论成就。学习这两个单元，有助于激发学生的创新性思维。

第 9～10 单元面向工程实际，展示了嵌入式产品形成过程的主要方面——选择芯片、选择设计语言、选择操作系统，利用仪器进行测试、验证、故障排除等。关于“芯片”，我们选择了嵌入式领域广泛使用的 ARM。关于“语言”，我们介绍了汇编语言、C 语言、C++语言、Java 语言。关于“操作系统”，我们展示了“实时操作系统”(RTOS)和“嵌入式操作系统”(embedded OS)。关于“仪器测量”，我们选择了 3 个必备仪器——信号源、示波器、逻辑分析仪，并讨论了“信号完整性”(SI)问题和“虚拟仪器”(VI)应用。

在英语语言技能训练方面，本书采用了 3 种方式——完形填空、英汉互译、摘要写作。进行“完形填空”练习，有助于雕琢语言细节——语法、词汇、固定搭配、专业术语、专业表达方法等。进行“英汉互译”练习，有助于明辨英汉语言差异、文化差异和思维方式差异，从而提高专业阅读、学术写作和科技交流的水平。进行“摘要写作”练习，有助于熟悉科技文体格式、用词特点和写作技巧，为将来写作科技论文打下良好基础。

任何语言能力的培养，都必须在真实的场景中通过训练才能收到实效。在本书中，各单元练习全部取材于国外大学教材、科技专著、国际会议论文、工程技术文档(芯片使用指南、用户手册、技术白皮书)、业内专家文章，以期达到“利用英语学专业、通过专业练英语”的教学目标。

为了便于学生自学，本书提供了课文参考译文、练习参考答案。此外，本书在附录中还提供了“科技英语语音手册”和“科技英语词汇手册”。

为了便于教师授课及丰富课堂内容，本书配有 PowerPoint 电子讲稿、阅读材料参考译文、模拟试卷及答案、视频材料及脚本等教辅资料。凡使用本书作为教材的教师，均可登录电子工业出版社华信教育资源网(www.hxedu.com.cn)注册下载，或者通过电话(010-88254531)联系免费获得。

为方便读者学习，我们开发了配套音频微课(含词汇、结构、语法、写作)并以二维码的形式给出，扫描下面二维码即可打开。

本书自 2004 年 7 月出版第 1 版以来，受到了高校师生和工程技术人员的普遍认可与欢迎。这次修订是在前 5 版的基础上，充分采纳了读者的合理化建议，全面更新了技术内容和语言内容。希望最新版教程更加贴近教学、科研和工程实际。

由于水平所限，书中难免有纰漏和欠妥之处。本书的读者反馈邮箱为 rzgwriting@163.com，欢迎各位读者不吝赐教！

编著者

使 用 说 明

学生自学建议

在正式学习前，建议通读附录。对于没有系统学习过科技英语课程的学生，最好在学习本书之前，将附录中的“科技英语语音手册”和“科技英语词汇手册”通读一遍。

全面理解核心学术词汇，准确记忆关键科技术语。面对海量英语词汇，必须进行选择。那些出现频率高、词义繁多、用法灵活的词汇是突破的重点。在专业英语学习中，有两类重点词汇——“核心学术词汇”(Core Academic Words)和“关键科技术语”(Key Technical Terms)。请分别参照课文词汇表的 New Words & Expressions 部分和 Technical Terms 部分。

泛读课文内容，聚焦英文表达技巧。在对照参考译文阅读课文的时候，不要仅限于明白大概意思，要把主要精力集中于“这个中文意思，用英文是怎么表达的”。在很多情况下，英文中有若干种表达方法。通过做笔记、摘抄和总结的办法，就可以不断地积累各种常用的英文表达技巧。

翻译阅读材料，训练技术信息提取能力。作为将来的工程师，应具备从英文技术资料中提取关键信息的能力。本书中的阅读材料包含了丰富的技术题材。如果对于某方面内容特别感兴趣，则可以尝试着翻译一下。这也是深入学习词汇、语法和文化的最好方式。

利用单元练习，进行自我测试。本书中的单元练习全部来自真实的技术资料，有利于学生测试自己真实的专业英语水平。

教师授课建议

根据教学实际情况，恰当选择精讲课文及段落。本书的课文难度分为高、中、低，适用于各类型大学、各种水平学生的专业英语教学。课文的选材，既有高难度的科技专著、学术会议论文、技术手册，也有中等难度的大学教材、用户手册、技术白皮书，还有低难度的综述文章、科普读物。根据实际情况，教师要选择适合课堂教学的内容，作为精讲课文和精讲段落。其他课文和段落，可以用于学生课前预习或课后自学，也可以不使用。

利用英语学专业，通过专业练英语。专业英语课程的实践性很强。在课堂上，不但需要精讲，更需要精练。一段精讲课文，既可以训练学生从中快速、准确提取技术信息的阅读能力，又可以训练学生撰写摘要、模仿造句、独立写作的能力。这种“阅读”→“分析”→“写作”的教学模式，被证明是非常成功的。

适时适量引入音频、视频，丰富课堂内容，激发学习兴趣。为了提高专业英语听说能力，可以在课堂教学中引入一定量的音、视频材料，并设法引导学生学会利用网络进行专业英语的听说读写交流。

强调科技英语阅读、写作，提高科技交流职业技能。作为将来的工程师，阅读能力和写作能

力最为重要。在科技交流中，无论是查询资料、进行设计，还是联系业务，都离不开阅读和写作。有效的科技交流要求“准确”“快速”，而这种能力不经过专业训练是无法达到的。

扩展阅读建议

科技交流

[1] 法依弗 W S，阿德金斯 K E. 大学科技英语国际交流教程[M]. 任治刚，译. 北京：电子工业出版社，2015.

[2] KNISELY C W，KNISELY K I. Engineering communication[M]. Boston：Cengage Learning，2014.

[3] ALLE M. The craft of scientific presentations：critical steps to succeed and critical errors to avoid[M]. 2nd ed. New York：Springer，2013.

论文写作

[1] SWALES J M，FEAK C. Academic writing for graduate students：essential tasks and skills[M]. 3rd ed. Ann Arbor：University of Michigan Press ELT，2012.

[2] 盖斯特尔 B，戴 R A. 科技论文写作与发表教程：第 8 版[M]. 任治刚，译. 北京：电子工业出版社，2018.

词汇语法

[1] DAY R A，SAKADUSKI N. Scientific english：a guide for scientists and other professionals[M]. 3rd ed. Santa Barbara：Greenwood Publishing Group，2011.

[2] CAPLAN N A. Grammar choices for graduate and professional writers[M]. 2nd ed. Ann Arbor：University of Michigan Press ELT，2019.

教材选用建议

对于同时开设“电子信息工程专业英语”“通信工程专业英语”的学校，还可以选用《电子通信专业英语简明教程》(任治刚主编，电子工业出版社出版)。该书篇幅适中、内容简明，并专门设立“科技英语、学术写作和职场交流”部分。该书也适合有志留学或深造的学生自修及国际化 IT 企业人士阅读。

对“通信工程专业英语”感兴趣的读者，可以参阅《通信工程专业英语教程》(任治刚主编，电子工业出版社出版)。该书题材广泛、内容实用、练习丰富，并提供“学术英语提要”“科技论文写作”等内容，可以满足读者在就业、深造或留学时的切实需求。

Contents

Unit 1 Electronic Devices …… (1)
Lesson 1 VLSI Technology …… (2)
Lesson 2 Memory Devices …… (5)
Lesson 3 Microprocessors …… (9)
Exercises …… (17)
Reading Materials …… (20)
Passage 1 Manufacturing Integrated Circuits …… (20)
Passage 2 Flash Memory …… (22)
Passage 3 Microcontrollers …… (23)
Unit 2 Electronic Circuits …… (25)
Lesson 4 Operational Amplifiers …… (26)
Lesson 5 Low-pass Filters …… (31)
Lesson 6 Analog-to-Digital Converters …… (38)
Exercises …… (43)
Reading Materials …… (46)
Passage 1 Anti-aliasing Filtering …… (46)
Passage 2 Switched-capacitor Filters …… (47)
Passage 3 Digital-to-Analog Converters …… (48)
Unit 3 Electronic System Components …… (50)
Lesson 7 Switching Power Supply …… (51)
Lesson 8 Clock Sources …… (56)
Lesson 9 Interconnects …… (62)
Exercises …… (66)
Reading Materials …… (69)
Passage 1 Cables …… (69)
Passage 2 Bus Basics …… (70)
Passage 3 USB (Universal Serial Bus) …… (72)
Unit 4 Electronic Systems …… (75)
Lesson 10 The Mobile Telephone System (Ⅰ) …… (76)
Lesson 11 The Mobile Telephone System (Ⅱ) …… (80)
Lesson 12 The Personal Computer System …… (85)
Exercises …… (91)
Reading Materials …… (93)
Passage 1 GPS …… (93)
Passage 2 802.11 Wireless LAN …… (94)
Passage 3 Mechatronic Systems …… (96)
Unit 5 Modern Digital Design …… (99)
Lesson 13 Electronic Design Automation …… (100)
Lesson 14 FPGAs …… (105)

Lesson 15 VHDL …… (109)
Exercises …… (114)
Reading Materials …… (117)
Passage 1 The PLD Evolution …… (117)
Passage 2 Verilog HDL …… (119)
Passage 3 SoC …… (121)
Unit 6 Digital Signal Processing Technology …… (124)
Lesson 16 Digital Signal Processing …… (125)
Lesson 17 Digital Signal Processors …… (130)
Lesson 18 ASP or DSP …… (134)
Exercises …… (138)
Reading Materials …… (142)
Passage 1 Typical DSP Applications …… (142)
Passage 2 Software Radio …… (144)
Passage 3 Medical Image Processing …… (146)
Unit 7 Voice & Audio …… (150)
Lesson 19 High Fidelity Audio …… (151)
Lesson 20 MP3 Audio Compression …… (155)
Lesson 21 The Mobile Telephone System(Ⅲ) …… (159)
Exercises …… (163)
Reading Materials …… (166)
Passage 1 Audio Encoding …… (166)
Passage 2 Sound Quality versus Data Rate …… (168)
Passage 3 Vocoders …… (170)
Unit 8 Image & Video …… (173)
Lesson 22 Digital Image Fundamentals …… (174)
Lesson 23 Digital Cameras …… (177)
Lesson 24 Television Video Signals …… (182)
Exercises …… (184)
Reading Materials …… (187)
Passage 1 JPEG …… (187)
Passage 2 MPEG …… (189)
Passage 3 Internet Video …… (190)
Unit 9 Embedded Applications …… (193)
Lesson 25 Choosing the Right Core …… (194)
Lesson 26 Design Languages for Embedded Systems …… (197)
Lesson 27 Choosing a Real-Time Operating System …… (201)
Exercises …… (205)
Reading Materials …… (207)
Passage 1 ARM …… (207)
Passage 2 Embedded OS …… (209)
Passage 3 Android Tablets …… (211)
Unit 10 Instrument & Measurement …… (213)
Lesson 28 Signal Sources …… (214)

Lesson 29 Oscilloscopes …… (220)
Lesson 30 Logic Analyzers …… (224)
Exercises …… (229)
Reading Materials …… (232)
Passage 1 Understanding Waveforms …… (232)
Passage 2 Signal Integrity …… (234)
Passage 3 Virtual Instrumentation …… (236)
课文参考译文 …… (239)
第 1 课 VLSI(超大规模集成)技术 …… (239)
第 2 课 存储器 …… (240)
第 3 课 微处理器 …… (241)
第 4 课 运算放大器 …… (245)
第 5 课 低通滤波器 …… (247)
第 6 课 模数转换器 …… (249)
第 7 课 开关电源 …… (250)
第 8 课 时钟信号源 …… (251)
第 9 课 互连部件 …… (253)
第 10 课 移动电话系统(Ⅰ) …… (255)
第 11 课 移动电话系统(Ⅱ) …… (256)
第 12 课 个人计算机(PC)系统 …… (259)
第 13 课 电子设计自动化 …… (261)
第 14 课 现场可编程门阵列 …… (264)
第 15 课 VHDL …… (265)
第 16 课 数字信号处理 …… (268)
第 17 课 数字信号处理器 …… (270)
第 18 课 ASP 或 DSP …… (271)
第 19 课 高保真音频 …… (273)
第 20 课 MP3 音频压缩 …… (275)
第 21 课 移动电话系统(Ⅲ) …… (276)
第 22 课 数字图像基础 …… (278)
第 23 课 数码相机 …… (279)
第 24 课 电视视频信号 …… (281)
第 25 课 选择合适的微处理器内核 …… (282)
第 26 课 嵌入式系统的设计语言 …… (284)
第 27 课 选择实时操作系统 …… (286)
第 28 课 信号源 …… (288)
第 29 课 示波器 …… (291)
第 30 课 逻辑分析仪 …… (293)
练习参考答案 …… (296)
附录 A 科技英语语音手册 …… (308)
附录 B 科技英语词汇手册 …… (316)
参考文献 …… (337)

Unit 1

Electronic Devices

Lesson 1 VLSI Technology

Lesson 2 Memory Devices

Lesson 3 Microprocessors

Passage 1 Manufacturing Integrated Circuits

Passage 2 Flash Memory

Passage 3 Microcontrollers

Lesson 1 VLSI Technology

One of the key inventions in the history of electronics[1], and in fact one of the most important inventions over time[2], was the transistor. It was invented by Bell Laboratories[3] in 1947. In short, a transistor is a device that conducts a variable amount of electricity through it, depending on how much electricity is input to it. In other words, it is a digital switch. However, unlike the vacuum tube, it is solid state. This means that it doesn't change its physical form as it switches. There are no moving parts in a transistor.

The advantages of the transistor over the vacuum tube[4] were enormous. Compared to the old technology, transistors were much smaller, faster, and cheaper to manufacture. They were also far more reliable and used much less power. The transistor is what started the evolution of the modern computer industry in motion.[5]

The transistor was originally a single, discrete device, which you could place individually into a circuit much like any other. Today, some special-purpose transistors are still used that way. What allowed the creation of modern processors was the invention of the integrated circuit[6], which is a group of transistors manufactured from a single piece of material and connected together internally, without extra wiring. Integrated circuits are also called ICs or chips.

A special material is used to make these integrated circuits. While most materials either insulate from electrical flow (air, glass, wood) or conduct electricity readily (metals, water), there are some that only conduct electricity by a small amount or under certain conditions. These are called semiconductors. The most commonly used semiconductor is of course silicon.

By careful chemical composition and arrangement, it is possible to create a very small transistor directly on a layer of silicon, using various technologies to manipulate the material into the correct form. These transistors are small, fast and reliable, and use relatively little power. The first integrated circuit was invented in 1958 by Texas Instruments.[7] It contained just six transistors on a single semiconductor surface.

After the invention of the integrated circuit, it took very little time for people to realize the tremendous benefits of miniaturizing and integrating larger numbers of transistors into the same integrated circuit. More transistors (switches) were required in order to implement more complicated functions. Miniaturization was the key to integrating together large numbers of transistors while increasing hardware speed and keeping power consumption and space requirements manageable.

Large-scale integration (LSI) came to refer to the creation of integrated circuits that had previously been made from multiple discrete components. These devices typically contained

hundreds of transistors. Early computers were made from many of these smaller ICs connected together on circuit boards.

As time progressed after the invention of LSI integrated circuits, the technology improved and chips became smaller, faster and cheaper. Building on the success of earlier integration efforts, engineers learned to pack more and more logic into a single circuit. This effort became known as very large-scale integration (VLSI). VLSI circuits can contain millions of transistors.

Originally, the functions performed by a processor were implemented using several different logic chips. Intel[8] was the first company to incorporate all of these logic components into a single chip. This was the first microprocessor, the 4004 (see Figure 1.1), introduced by Intel in 1971. All of today's processors are (highly advanced!) descendants of this original[9] 4-bit CPU.

Figure 1.1 Intel 4004

New Words & Expressions

invention [ɪn'venʃən] *n*. 发明
device [dɪ'vaɪs] *n*. 器件
conduct [kən'dʌkt] *vt*. 传导
variable ['veriəbl] *adj*. 可变的
amount [ə'maʊnt] *n*. 数量
vacuum ['vækjʊəm] *n*. 真空
solid ['sɒlɪd] *adj*. 固体的
advantage [əd'væntɪdʒ] *n*. 优势
enormous [ɪ'nɔːməs] *adj*. 巨大的
manufacture [ˌmænjʊ'fæktʃə] *vt*. 批量生产
reliable [rɪ'laɪəbl] *adj*. 可靠的
evolution [ˌevə'luʃən] *n*. 发展;演变
originally [ə'rɪdʒənəli] *adv*. 起初
individually [ˌɪndɪ'vɪdʒuəli] *adv*. 分别地
internally [ɪn'tɜːnəli] *adv*. 内部地
composition [ˌkɒmpə'zɪʃn] *n*. 合成
arrangement [ə'reɪndʒmənt] *n*. 排列
directly [daɪ'rektli] *adv*. 直接地
layer ['leɪə] *n*. 层
various ['veriəs] *adj*. 多种不同的
manipulate [mə'nɪpjuleɪt] *vt*. 处理
contain [kən'teɪn] *vt*. 包含
surface ['sɜːfɪs] *n*. 表面
tremendous [trə'mendəs] *adj*. 巨大的
benefit ['benɪfɪt] *n*. 益处
miniaturize ['mɪnɪətʃəˌraɪz] *vt*. 使微型化
integrate ['ɪntɪgreɪt] *vt*. 集成
consumption [kən'sʌmpʃən] *n*. 消耗
requirement [rɪ'kwaɪəmənt] *n*. 需求
manageable ['mænɪdʒəbəl] *adj*. 可处理的
refer to ... 是指
previously ['priːviəsli] *adv*. 此前
multiple ['mʌltɪpl] *adj*. 多个的
progress [prə'gres] *vi*. 时光流逝
implement ['ɪmpləment] *vt*. 实现
incorporate [ɪn'kɔːpəreɪt] *vt*. 合并
introduce [ˌɪntrə'djuːs] *vt*. 推出
advanced [əd'vænst] *adj*. 先进的
descendant [dɪ'sendənt] *n*. 后代
original [ə'rɪdʒənəl] *adj*. 最初的

Technical Terms

electronics [ˌɪlekˈtrɒnɪks] *n*. 电子学
transistor [trænˈsɪstə] *n*. 晶体管
electricity [ˌɪlekˈtrɪsəti] *n*. 电
digital [ˈdɪdʒɪtl] *adj*. 数字的
switch [swɪtʃ] *n*. 开关
power [ˈpaʊə] *n*. 功率
discrete [dɪˈskriːt] *adj*. 分立的
circuit [ˈsəːkɪt] *n*. 电路
chip [tʃɪp] *n*. 芯片
insulate [ˈɪnsjuleɪt] *vt*. 绝缘
semiconductor [ˌsemɪkənˈdʌktə] *n*. 半导体
silicon [ˈsɪlɪkən] *n*. 硅
miniaturization [ˈmɪnətʃəˌraɪˈzeɪʃn] *n*. 缩微化
component [kəmˈpəʊnənt] *n*. 元器件;组件;部件
logic [ˈlɒdʒɪk] *n*. 逻辑
microprocessor [maɪkrəʊˈprəʊsesə] *n*. 微处理器
solid state　固态
power consumption　功耗
circuit board　电路板
IC *abbr*. integrated circuit　集成电路
LSI *abbr*. large-scale integration　大规模集成
VLSI *abbr*. very large-scale integration　超大规模集成
CPU *abbr*. central processing unit　中央处理单元

Notes

1. electronics 可译作“电子学”“电子技术”或“电子线路”“电子设备”。例如,“镉是电子工业有毒废物”(Cadmium is a toxic waste product of *electronics* industry)。后缀-cs 表示“学术”,如“美学”(aesthetics)、“伦理学”(ethics)、“经济学”(economics)。
2. 介词 over 可表示时间的跨越,如“此图展示了过去 20 年来 *Y* 随 *X* 变化的情况”(This figure shows how *Y* varied in relationship with *X* *over* the last 20 years),又如“在 4 天之内”(*over* a period of four days)、“此项研究为期 6 个月”(The study will be carried out *over* a six-month period)。
3. Bell Laboratories(又称 Bell Labs)是指贝尔实验室,其总部位于美国新泽西州的默里山(Murray Hill, New Jersey)。1947 年 12 月 23 日,该实验室的 Bill Shockley、Walter Brattain 和 John Bardee 发明了晶体管。

4. The advantages of *Y* over *X* 结构表示“*Y* 相对于 *X* 的优势”。advantage 常见搭配有 a big/great/major advantage、the main advantage、have an advantage 和 the advantage of *sth*.。
5. 此句为强调结构，使用 what 表语从句强调主语(the transistor)。
6. 此句为强调结构，使用 what 主语从句强调表语(the invention of the integrated circuit)。
7. Texas Instruments 是指德州仪器公司，其总部位于美国得克萨斯州的达拉斯(Dallas，Texas)。1958 年，该公司的 Jack Kilby 发明了第一片集成电路。
8. Intel 是指英特尔公司，其总部位于美国加利福尼亚州的圣克拉拉(Santa Clara，California)。1968 年，Robert Noyce 和 Gordon Moore 创建了该公司。
9. original 源自 origin(起点)，意为“最初的”，如“原计划”(*original* plan)、“原始特征”(*original* features)、“将讨论返回到开始的话题”(steer the discussion to the *original* topic)。

Lesson 2　Memory Devices

Memories can be made in mechanical, magnetic, optical, biological and electronic technologies. Examples of magnetic memories are tapes, floppy disks, hard drives and ferroelectric[1] RAMs. Examples of optical memories are CD-ROMs, rewrittable CDs. Electronic memory is used extensively in computer equipment since it is the fastest available. All electronic memory today can be in separate IC format, module format, or can be part of an IC as a macro-function or cell. Table 1.1 is an overview of some electronic memories.

Table 1.1　Overview of common electronic memory devices

Type	Properties	Writable	Non-volatile	Speed	Cost
Flip-flop	One-bit register. Used as basic building blocks in digital circuits	Yes	No	Ultra fast	Very high
Register	Set of flip-flops holding a byte, word or long word. Used in complex chips such as CPUs	Yes	No	Ultra fast	Very high
SRAM	Array of flip-flops that is addressable. Used for temporary storage of data or cache[2]	Yes	No	Very fast	High
DRAM	Array of storage cells which is addressable. Used for main computing data storage	Yes	No	Fast	Moderate
ROM	Array of hard-wired cells that is addressable. Programming done at time of chip manufacture	No	Yes	Very fast	Low
EEPROM	Electrically erasable programmable ROM. Number of write cycles is limited	Yes	Yes	Low	High

Flip-flop

A flip-flop is basically a bi-state circuit in which either a 0 or 1 state can resides. Because of its simplicity, the flip-flop is extremely fast. As a basic element, the flip-flop is used in

digital circuits and ICs. A flip-flop will lose its state when the supply voltage is removed. Therefore, it is volatile.

Register

A register is a set of flip-flops in parallel. Typically a register is 8,16,32 or 64 bits wide. Often a register is used to hold data, address pointers, etc. A register is volatile and very fast just like the flip-flop.

SRAM (static random-access memory)

An SRAM is an array of addressable flip-flops. The array can be configured as such that the data comes out in 1-bit, 4-bit, 8-bit, and etc. SRAM is simple, fast and volatile just like the flip-flop, its basic memory cell. SRAM can be found on microcontroller boards (either on or off the CPU chip), where the amount of memory required is small and it will not pay off to build the extra interface circuitry for DRAMs. In addition, SRAM is often used as cache because of its high speed.

SRAM comes in many speed classes, ranging from several nanoseconds for cache applications to 200ns for low-power applications. SRAM exists in both bipolar and MOS technology. CMOS[3] technology boasts the highest density and the lowest power consumption. Fast cache memory can be constructed in BiCMOS technology, a hybrid technology that uses bipolar transistors for extra drive. The fastest SRAM memories are available in emitter-coupled logic (ECL) bipolar technology. Because of the high power consumption, the memory size is limited in this technology.

A special case of SRAM memory is content-addressable memory (CAM)[4]. In this technology, the memory consists of an array of flip-flops, in which each row is connected to a data comparator. The memory is addressed by presenting[5] data to it (not an address!). All comparators will then check simultaneously[6] if their corresponding RAM register holds the same data. The CAM will respond with the address of the row (register) corresponding to the original data. The main application for this technology is fast look-up tables. These are often used in network routers.

DRAM (dynamic random-access memory)

The word "dynamic" indicates that the data is not held in a flip-flop but rather in a storage cell. The data in a storage cell must be refreshed (read out and re-written) regularly because of leakage. The refresh time interval is usually 4 to 64 ms. The storage cell only requires one capacitor and one transistor, whereas[7] a flip-flop connected in an array requires 6 transistors. In trench capacitor memory technology, which is used in all modern DRAMs, the transistor is constructed above the capacitor so that the space on chip is ultimately minimized. For this reason, DRAM technology has a lower cost per bit than SRAM technology. The disadvantage of the extra circuitry required for refreshing is easily offset by the lower price per bit when using large memory sizes.

DRAM memory is, just like SRAM memory constructed as an array of memory cells. A major difference between SRAM and DRAM, however, lies in the addressing technique. With an SRAM, an address needs to be presented and the chip will respond with presenting the data of the memory cell at the output, or accepting the data at the input and write it into the addressed cell. With DRAM technology, this simple approach is impossible since addressing a row of data without rewriting it will destruct all data in the row because of the dynamic nature.

ROM (read-only memory)

ROMs are also called mask[8]-ROMs or mask-programmed ROMs. This is because a ROM needs to be programmed by setting its cells to either 0 or 1 at the time of manufacture. Usually the 0 or 1 is formed by the presence or absence of an aluminum line. This aluminum pattern is defined by a lithographic mask used in one of the last steps of manufacture. Therefore these devices are often called mask-ROMs.

The advantage of ROM is that it can be manufactured at the lowest price in high volumes. Another advantage in some applications is that it is impossible to alter the data once the chips are made, and that no further programming and testing are required. On the other hand, if the data or code must be changed this can be a small disaster. The rest of the chips will end in the dustbin and new chips will have to be made.

EEPROM (electrically erasable PROM)

This means that the chip can be programmed like an EPROM, but can be erased electrically. As a result, no UV source is required. EEPROMs can be erased on a byte-by-byte basis.

New Words & Expressions

mechanical [mɪˈkænɪkl] *adj*. 机械的
extensively [ɪkˈstensɪvli] *adv*. 大量地;广泛地
available [əˈveɪləbl] *adj*. 可获得的;可用的
format [ˈfɔːmæt] *n*. 格式
module [ˈmɒdʒul] *n*. 模块;功能块
property [ˈprɒpəti] *n*. 特性
volatile [ˈvɒlətaɪl] *adj*. 易失的
array [əˈreɪ] *n*. 阵列
temporary [ˈtempərəri] *adj*. 临时的
moderate [ˈmɒdərət] *adj*. 中等的
simplicity [sɪmˈplɪsəti] *n*. 简易性
parallel [ˈpærəlel] *adj*. 并行的
static [ˈstætɪk] *adj*. 静态的
simultaneously [saɪməlˈteɪniəsli] *adv*. 同时地
corresponding [ˌkɒrɪˈspɒndɪŋ] *adj*. 相应的
respond [rɪˈspɒnd] *vi*. 响应
original [əˈrɪdʒənəl] *adj*. 最初的
dynamic [daɪˈnæmɪk] *adj*. 动态的
indicate [ˈɪndɪkeɪt] *vt*. 标明
regularly [ˈregjələli] *adv*. 定期地
whereas [weərˈæz] *conj*. 然而
ultimately [ˈʌltɪmətli] *adv*. 最终
minimize [ˈmɪnɪmaɪz] *vt*. 减至最低
disadvantage [ˌdɪsədˈvæntɪdʒ] *n*. 缺点;劣势
offset [ˈɒfˌset] *vt*. 抵消;弥补
major [ˈmeɪdʒə] *adj*. 主要的;重要的

random [ˈrændəm] *adj*. 随机的
access [ˈækses] *n*. 访问
configure [kənˈfɪgə(r)] *vt*. 配置
range from ... to ... 从……变化到……
hybrid [ˈhaɪbrɪd] *adj*. 混合的
consist of ... 由……组成
technique [tekˈniːk] *n*. 技术
approach [əˈprəʊtʃ] *n*. 方法
presence [ˈprezəns] *n*. 存在；出现
absence [ˈæbsəns] *n*. 不存在；缺乏
aluminium [ˌæljəˈmɪniəm] *n*. 铝
on a ... basis 按照……方式

Technical Terms

memory [ˈmeməri] *n*. 存储器
magnetic [mægˈnetɪk] *adj*. 有磁性的
optical [ˈɒptɪkl] *adj*. 光学的
ferroelectric [ˌferəʊɪˈlektrɪk] *adj*. 铁电的
register [ˈredʒɪstə] *n*. 寄存器
cache [kæʃ] *n*. 高速缓存
erasable [ɪˈreɪzəbl] *adj*. 可擦除的
programmable [ˈprəʊgræməbl] *adj*. 可编程的
microcontroller [maɪkrəkənˈtrəʊlə] *n*. 微控器
interface [ˈɪntəˌfeɪs] *n*. 接口
circuitry [ˈsɜːkɪtri] *n*. 电路
bipolar [baɪˈpəʊlə] *adj*. 双极性的
density [ˈdensəti] *n*. 密度
drive [draɪv] *n*. 驱动器
comparator [kəmˈpærətə] *n*. 比较器
router [ˈruːtə] *n*. 路由器
refresh [rɪˈfreʃ] *vt*. 刷新
leakage [ˈliːkɪdʒ] *n*. 泄漏
interval [ˈɪntəvl] *n*. 间歇
capacitor [kəˈpæsɪtə] *n*. 电容器
lithographic [ˌlɪθəˈgræfɪk] *adj*. 平版印刷的
source [sɔːs] *n*. 信号源
flip-flop [ˈflɪpˌflɒp] *n*. 触发器
digital circuit 数字电路
address pointer 地址指针
look-up table 查找表
network router 网络路由器
trench capacitor 沟道式电容器
SRAM [ˈesɪræm] *abbr*. static random-access memory 静态随机存取存储器
DRAM [ˈdiːræm] *abbr*. dynamic random-access memory 动态随机存取存储器
CMOS [ˈsiːˌməʊz] *abbr*. complementary metal oxide semiconductor 互补金属氧化物半导体

ECL *abbr*. emitter-coupled logic　射极耦合逻辑
CAM [kæm] *abbr*. content-addressable memory　内容寻址存储器
EEPROM [ˈiːprɒm] *abbr*. electrically erasable PROM　电可擦除可编程只读存储器
UV *abbr*. ultraviolet　紫外线

Notes

1. “铁电”(ferroelectric)现象是指当在某些材料上施加超过特定强度的电场时,该材料就在一个方向上极化或改变原来的极化方向。只要不受干扰,这种极化状态就一直保持下去。
2. “高速缓存”(cache)用于存储频繁使用的或近期使用的数据,从而提高数据存取速度。高速缓存分为两类:处理器中的内部缓存(内存缓存)和主板上的外部缓存(磁盘缓存)。
3. CMOS 集成电路的输出由一个 N 型 MOSFET 和一个 P 型 MOSFET 串联而成。因为 N 型 MOSFET 和 P 型 MOSFET 相互补偿,所以使用“互补”(compementary)来说明它。
4. “内容寻址存储器”(CAM)用于实现高速查找表,其容量要明显小于 DRAM 或 SRAM。
5. 作动词时,present 意为“展示”“呈现”。作形容词时,present 意为“出现的”“存在的”;其名词形式为 presence。absent 是 present 的反义词,其名词形式为 absence。
6. simultaneously 源自形容词 simultaneous。以-aneous 结尾的常用形容词还有 spontaneous(“自发的”)和 instantaneous(“立即的”)。例如,“这种液体自发燃烧”(The liquid *spontaneously* ignited)和“立即反应”(an *instantaneous* response/reply/reaction)。
7. whereas 用来连接对比句,可译作“而……”。例如,“旧系统很复杂,而新系统很简单”(The old system was complicated, *whereas* the new system is very simple)。
8. mask 可译作“掩模”“掩码”或“掩蔽”。此处是指芯片制造过程中的掩模。

Lesson 3　Microprocessors

A microprocessor is a complete computation engine that is fabricated on a single chip.[1] The first microprocessor was the Intel 4004, introduced in 1971. The 4004 was not very powerful — all it could do was to add and subtract, and it could only do that 4 bits at a time. But it was amazing that everything was on one chip. Prior[2] to the 4004, engineers built computers either from collections of chips or from discrete components. The 4004 powered one of the first portable electronic calculators.

The first microprocessor to make it into a home computer was the Intel 8080, a complete 8-bit computer on one chip, introduced in 1974. The first microprocessor to make a real splash in the market was the Intel 8088, introduced in 1979 and incorporated into the IBM PC. The

PC market moved from the 8088 to the 80286 to the 80386 to the 80486 to the Pentium series to the Core series to the Xeon series. All of these microprocessors are made by Intel and all of them are improvements on the basic design of the 8088. An Intel Core i9 processor can have up to eight cores, each of can execute any piece of code that ran on the original 8088, only about 6,700 times faster! Table 1.2 shows the differences between the different processors that Intel has introduced over the years(1974 ~ 2004).

Table 1.2 Intel x86 microprocessors(1974～2004)

Name	Year	Transistors	Microns[3]	Clock speed	Data width[4]	MIPS[5]
8080	1974	6,000	6	2 MHz	8 bits	0.64
8088	1979	29,000	3	5 MHz	16 bits/8-bit bus	0.33
80286	1982	134,000	1.5	6 MHz	16 bits	1
80386	1985	275,000	1.5	16 MHz	32 bits	5
80486	1989	1,200,000	1	25 MHz	32 bits	20
Pentium	1993	3,100,000	0.8	60 MHz	32 bits/64-bit bus	100
Pentium Ⅱ	1997	7,500,000	0.35	233 MHz	32 bits/64-bit bus	～300
Pentium Ⅲ	1999	9,500,000	0.25	450 MHz	32 bits/64-bit bus	～510
Pentium 4	2000	42,000,000	0.18	1.5 GHz	32 bits/64-bit bus	～1,700
Pentium 4 "Prescott"	2004	125,000,000	0.09	3.6 GHz	32 bits/64-bit bus	～7,000

From this table you can see that, in general, there is a relationship between clock speed and MIPS. The maximum clock speed is a function[6] of the manufacturing process and delays within the chip. There is also a relationship between the number of transistors and MIPS. For example, the 8088 clocked at 5 MHz but only executed at 0.33 MIPS (about one instruction per 15 clock cycles). Modern processors can often execute at a rate of two instructions per clock cycle. That improvement is directly related to the number of transistors on the chip.

Inside a Microprocessor

A microprocessor executes a collection of machine instructions that tell the processor what to do. Based on the instructions, a microprocessor does three basic things:

- Using its ALU (arithmetic/logic unit), a microprocessor can perform mathematical operations like addition, subtraction, multiplication and division. Modern microprocessors contain complete floating-point processors that can perform extremely sophisticated operations on large floating-point numbers.
- A microprocessor can move data from one memory location to another.
- A microprocessor can make decisions and jump to a new set of instructions based on those decisions.

There may be very sophisticated things that a microprocessor does, but those are its three basic activities. Figure 1.2 shows an extremely simple microprocessor capable of doing[7] those three things. This microprocessor has an address bus that sends an address to memory, a data

bus that can send data to memory or receive data from memory, an RD (read) and WR (write) line to tell the memory whether it wants to set or get the addressed location, a clock line that lets a clock pulse sequence the processor and a reset[8] line that resets the program counter to zero (or whatever) and restarts execution. And let's assume that both the address and data buses are 8 bits wide here.

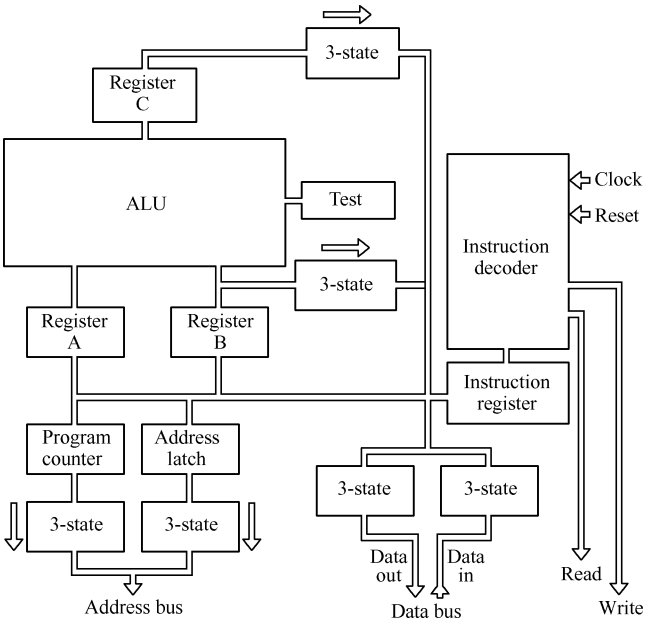


Figure 1.2 A simple microprocessor

Here are the components of this simple microprocessor:

- Registers A, B and C are simply latches made out of flip-flops.
- The address latch is just like registers A, B and C.
- The program counter is a latch with the extra capability to increment by 1 when told to do so, and also to reset to zero when told to do so.
- The ALU could be as simple as an 8-bit adder, or it might be capable of adding, subtracting, multiplying and dividing 8-bit values. Let's assume the latter here.
- The test register is a special latch that can hold values from comparisons performed in the ALU. An ALU can normally compare two numbers and determine if they are equal, if one is greater than the other, etc. The test register can also normally hold a carry bit from the last stage of the adder. It stores these values in flip-flops and then the instruction decoder can use the values to make decisions.
- There are six boxes marked "3-state" in the diagram. These are tri-state buffers[9]. A tri-state buffer can pass a 1, a 0 or it can essentially disconnect its output. A tri-state buffer allows multiple outputs to connect to a wire, but only one of them to actually drive a 1 or a 0 onto the line.
- The instruction register and instruction decoder are responsible for controlling all of the other components.

Although they are not shown in this diagram, there would be control lines from the instruction decoder that would:

1. Tell the register A to latch the value currently on the data bus.
2. Tell the register B to latch the value currently on the data bus.
3. Tell the register C to latch the value currently on the data bus.
4. Tell the program counter to latch the value currently on the data bus.
5. Tell the address register to latch the value currently on the data bus.
6. Tell the instruction register to latch the value currently on the data bus.
7. Tell the program counter to increment.
8. Tell the program counter to reset to zero.
9. Activate any of the six tri-state buffers (six separate lines).
10. Tell the ALU what operation to perform.
11. Tell the test register to latch the ALU's test bits.
12. Activate the RD line.
13. Activate the WR line.

Coming into the instruction decoder are the bits from the test register and the clock line, as well as the bits from the instruction register.

RAM and ROM

The address and data buses, as well as the RD and WR lines connect either to RAM or ROM — generally both. In our sample microprocessor, we have an address bus 8 bits wide and a data bus 8 bits wide. That means that the microprocessor can address (2^8) 256 bytes of memory, and it can read or write 8 bits of the memory at a time. Let's assume that this simple microprocessor has 128 bytes of ROM starting at address 0 and 128 bytes of RAM starting at address 128.

ROM stands for read-only memory. A ROM chip is programmed with a permanent collection of pre-set bytes. The address bus tells the ROM chip which byte to get and place on the data bus. When the RD line changes state, the ROM chip presents the selected byte onto the data bus.

RAM stands for random-access memory. RAM contains bytes of information, and the microprocessor can read or write to those bytes depending on whether the RD or WR line is signaled. One problem with today's RAM chips is that they forget everything once the power goes off. That is why the computer needs ROM.

By the way, nearly all computers contain some amount of ROM (it is possible to create a simple computer that contains no RAM — many microcontrollers do this by placing a handful of RAM bytes on the processor chip itself — but generally impossible to create one that contains no ROM). On a PC, the ROM is called the BIOS (basic input/output system). When the microprocessor starts, it begins executing instructions it finds in the BIOS. The BIOS instructions do things like test the hardware in the machine, and then it goes to the hard disk

to fetch the boot sector. This boot sector is another small program, and the BIOS stores it in RAM after reading it off the disk. The microprocessor then begins executing the boot sector's instructions from RAM. The boot sector program will tell the microprocessor to fetch something else from the hard disk into RAM, which the microprocessor then executes, and so on. This is how the microprocessor loads and executes the entire operating system.

Microprocessor Instructions

Even the incredibly simple microprocessor shown here will have a fairly large set of instructions that it can perform. The collection of instructions is implemented as bit patterns, each one of which has a different meaning when loaded into the instruction register. Humans are not particularly good at remembering bit patterns, so a set of short words are defined to represent the different bit patterns. This collection of words is called the assembly language of the processor. An assembler can translate the words into their bit patterns very easily, and then the output of the assembler is placed in memory for the microprocessor to execute. If you use C language programming, a C compiler will translates the C code into assembly language.

So now the question is, "How do all of these instructions look in ROM?" Each of these assembly language instructions must be represented by a binary number. These numbers are known as opcodes. The instruction decoder needs to turn each of the opcodes into a set of signals that drive the different components inside the microprocessor. Let's take the ADD instruction as an example and look at what it needs to do:

1. During the first clock cycle, we need to actually load the instruction. Therefore the instruction decoder needs to:
   1) activate the tri-state buffer for the program counter.
   2) activate the RD line.
   3) activate the data-in tri-state buffer.
   4) latch the instruction into the instruction register.
2. During the second clock cycle, the ADD instruction is decoded. It needs to do very little:
   1) set the operation of the ALU to addition.
   2) latch the output of the ALU into the register C.
3. During the third clock cycle, the program counter is incremented (in theory this could be overlapped into the second clock cycle).

Every instruction can be broken down as a set of sequenced operations like these that manipulate the components of the microprocessor in the proper order. Some instructions, like this ADD instruction, might take two or three clock cycles. Others might take five or six clock cycles.

Microprocessor Performance

The number of transistors available has a huge effect on the performance of a processor.

As seen earlier, a typical instruction in a processor like an 8088 took 15 clock cycles to execute. Because of the design of the multiplier, it took approximately 80 cycles just to do one 16-bit multiplication on the 8088. With more transistors, much more powerful multipliers capable of single-cycle speeds become possible.

More transistors also allow for a technology called pipelining. In a pipelined architecture, instruction execution overlaps. So even though it might take five clock cycles to execute each instruction, there can be five instructions in various stages of execution simultaneously. That way it looks like one instruction completes every clock cycle.

Many modern processors have multiple instruction decoders, each with its own pipeline. This allows for multiple instruction streams, which means that more than one instruction can complete during each clock cycle. This technique can be quite complex to implement, so it takes lots of transistors.

The trend in processor design has been toward full 32-bit ALUs with fast floating-point processors built in and pipelined execution with multiple instruction streams. There has also been a tendency toward special instructions that make certain operations particularly efficient. There has also been the addition of hardware virtual memory support and L1 caching on the processor chip. All of these trends push up the transistor count, leading to the multi-million transistor powerhouses available today. These processors can execute about one billion instructions per second!

The newest thing in processor design is 64-bit ALUs. Sixty-four-bit processors have been with us since 1992, and in the 21st century they have started to become mainstream. Both Intel and AMD have introduced 64-bit chips. Sixty-four-bit processors have 64-bit ALUs, 64-bit registers, 64-bit buses and so on.

New Words & Expressions

computation [ˌkɒmpjuˈteɪʃn] *n*. 计算
fabricate [ˈfæbrɪkeɪt] *vt*. 构造
at a time　一次
improvement [ɪmˈpruːvmənt] *n*. 改进
in general　大体上
relationship [rɪˈleɪʃənʃɪp] *n*. 关系;联系
maximum [ˈmæksɪməm] *adj*. 最大的
at a rate of ...　以……的速率
related to ...　与……相关
a collection of ...　一批……
based on ...　依据……
sophisticated [səˈfɪstɪˌkeɪtɪd] *adj*. 复杂巧妙的
assume [əˈsjuːm] *vt*. 假定
stand for ...　代表……
permanent [ˈpəːmənənt] *adj*. 永久的
implement [ˈɪmplɪmənt] *vt*. 实现;实施
define [dɪˈfaɪn] *vt*. 定义
represent [ˌreprɪˈzent] *vt*. 代表;表示
overlap [ˌəʊvəˈlæp] *vt*. 重叠
manipulate [məˈnɪpjuˌleɪt] *vt*. 操纵;使用
in the proper order　按照正确顺序
has a huge effect on ...　对……产生巨大影响
performance [pəˈfɔːməns] *n*. 性能
as seen earlier, ...　如前所见,……
approximately [əˈprɒksɪmətli] *adv*. 大约
allow for ...　考虑到……

increment [ˈɪŋkrəmənt] *n*. 增量；增加
comparison [kəmˈpærɪsn] *n*. 比较；对比
normally [ˈnɔːməli] *adv*. 通常
diagram [ˈdaɪəgræm] *n*. 示意图
essentially [ɪˈsenʃəli] *adv*. 基本上
responsible for … 负责……
currently[ˈkʌrəntli] *adv*. 当前；时下
architecture [ˈɑːkɪˌtɛktʃə] *n*. 体系结构
simultaneously [ˌsaɪməlˈteɪniəsli] *adv*. 同时地
trend [trend] *n*. 趋势；动态
tendency[ˈtendənsi] *n*. 趋向；偏好
efficient [ɪˈfɪʃənt] *adj*. 高效的
virtual [ˈvɜːtʃʊəl] *adj*. 虚拟的
mainstream [ˈmeɪnˌstriːm] *n*. 主流

Technical Terms

microprocessor [maɪkrəʊˈprəʊsesə] *n*. 微处理器
execute [ˈeksɪˌkjuːt] *vt*. 执行
micron [ˈmaɪkrɒn] *n*. 微米；10^{-6}米
instruction [ɪnˈstrʌkʃən] *n*. 指令
addition [əˈdɪʃən] *n*. 加法
subtraction [səbˈtrækʃən] *n*. 减法
multiplication [ˌmʌltɪplɪˈkeɪʃn] *n*. 乘法
division [dɪˈvɪʒən] *n*. 除法
pulse [pʌls] *n*. 脉冲
sequence [ˈsiːkwəns] *n*. 序列
counter [ˈkaʊntə] *n*. 计数器
decoder [diːˈkəʊdə] *n*. 译码器
buffer [ˈbʌfə] *n*. 缓存器
activate [ˈæktɪveɪt] *vt*. 激活
assembler [əˈsemblə] *n*. 汇编器
compiler [kəmˈpaɪlə] *n*. 编译器
multiplier [ˈmʌltɪˌplaɪə] *n*. 乘法器
pipelining [ˈpaɪpˌlaɪnɪŋ] *n*. 流水线技术
discrete component 分立元件
machine instruction 机器指令
floating-point processor 浮点处理器
memory location 存储器位置
program counter 程序计数器
address latch 地址锁存器
test register 测试寄存器
carry bit 进位位
instruction decoder 地址译码器
boot sector 引导扇区
operating system 操作系统

bit pattern　比特模式;位模式
assembly language　汇编语言
binary number　二进制数
clock cycle　时钟周期
virtual memory　虚拟内存
MIPS [mɪps] *abbr*. million instructions per second　百万条指令/秒
ALU *abbr*. arithmetic/logic unit　算术逻辑单元
ROM [rɒm] *abbr*. read-only memory　只读存储器
RAM [ræm] *abbr*. random-access memory　随机存取存储器
BIOS [ˈbaɪɒs] *abbr*. basic input/output system　基本输入/输出系统
opcode [ˈɒpkəʊd] *abbr*. operation code　操作码

Notes

1. 此句涉及如何进行术语定义。正式定义采用完整句子形式给出该术语的名称、类属和特征。正式定义可分为四种类型:(1)"组成型"(A resistor is a device that *is composed of a highly resistive material designed to have a definite amount of resistance*.);(2)"类比型"(A resistor is a device that *absorbs electric energy very much like a damping device used in door opener to absorb mechanical energy or in cars for shock absorption* .);(3)"对比型"(A resistor is a device *used for dissipating energy, unlike a capacitor that is used for storing electric energy*.);(4)"功能型"(A resistor is a device *used in circuits to limit and create opposition to current flow or provide voltage drop*.)。除"正式定义"外,还有"非正式定义"(在括号中以单词或短语的形式给出该术语的近义词等最简信息)和"扩展定义"(在正式定义的基础上扩充几段有关历史背景、应用领域、组成部件、图形描述、基本原理、实例展示或术语比照等方面的内容)。简单术语仅需"非正式定义",复杂术语则要进行"正式定义"或"扩展定义"。
2. prior 可表示时间上"更早的"或作用上"更重要的",如"预先通知"(*prior* notice)、"事先安排"(*prior* arrangement)、"事先书面许可"(*prior* written consent)等。短语 prior to sth.表示在某个时间之前或者在某个事件之前。此句可译为"在 4004 出现之前,工程师要用一组芯片或一组分立元件构建计算机"。
3. "微米"(micron)用于度量芯片内最小线宽。其值越小,芯片内集成的晶体管就越多。
4. "数据宽度"(data width),这里是指 ALU 字长。一般而言,外部数据总线宽度和 ALU 数据宽度相同。但是,也存在例外情况。例如,8088 ALU 为 16 bits,其外部数据总线为 8 bits。又如,Pentium ALU 为 32 bits,其外部数据总线为 64 bits。
5. "百万条指令/秒"(MIPS)是表征微处理器性能的指标之一。鉴于现代微处理器的种类繁多、结构各异,MIPS 指标能反映出来的信息是粗略的。
6. 在数学中,function 表示"函数",如"三角函数"(trigonometric *function*)等。Y is a function of X 可译作"Y 是 X 的函数",在这里可译作"Y 取决于 X"。
7. be capable of (doing)*sth*.用于表示人、机构或器物的能力,如"十级风能够掀翻屋顶"(A

force ten wind *is capable of* blowing the roofs off houses.)。而 be able to do *sth*.只用于表示人的能力,如“她能应对此项工作吗?”(Will she *be able to* cope with the work?)。此外,be capable of (doing) *sth*. 还可以表示“可能性”,(the) capability to do *sth*.则表示“做……的能力”。

8. “复位”(reset)是指“使系统状态处于指定的初始值”。
9. “三态缓存器”(tri-state buffer)用来控制向总线输出信号。当控制位有效时,输入信号通过三态缓存器向总线输出;当控制位无效时,三态缓存器输出端呈现高阻状态(记作Z),即不向总线输出任何信号。

Exercises

1. Fill in the blanks with proper words, phrases or clauses.

(1)The microprocessor is the central _____(部件)of the PC._____(你做的全部工作)on your computer is performed directly or _____ by the processor. Obviously, it is one of the most important components of the PC,_____(如果不是)the most important. It is also, scientifically, not only one of the most _____(令人惊奇的)parts of the PC, but one of the most amazing _____ (器件)in the world of technology.

The processor _____(扮演重要的角色)in the following important _____(方面)of your computer system.

Performance: The processor is probably the most important single _____(决定性的因素)of system performance in the PC._____ other components also play a key role in determining performance, the processor s capabilities _____ (控制) the maximum performance of a system. The other devices only allow the processor to _____(达到它的全部潜能).

Software Support: Newer, faster processors _____ the use of the latest software. _____ addition, new processors enable the use of specialized software not usable on earlier machines.

Reliability and Stability: The quality of the processor is one factor that determines _____(系统运行的可靠性). While most processors are very _____(可靠的), some are not. This also depends _____(在某种程度上)on the age of the processor and how much energy it _____(消耗).

Energy Consumption and Cooling: Originally processors consumed relatively little power compared _____ other system devices. Newer processors can consume _____(大量的) power. Power consumption has an impact _____ everything from cooling method selection to overall system reliability.

Motherboard Support: The processor you decide to use in your system will be a _____(主要的) determining factor in what sort of chipset you must use, and hence what motherboard you buy. The motherboard _____(反过来)dictates many facets of your system s capabilities and performance.

(2)Flash memory, also known as flash storage, is a type of _____(非易失的)memory that erases data in units called blocks and rewrites data at the byte level.

Flash memory is widely used _____ storage and data transfer in consumer devices, enterprise systems and industrial applications. Flash memory retains data for an extended period regardless of whether a flash－equipped device is powered on or off.

Flash memory is used in enterprise data center server, storage and networking technology as well as in _____(大量各种各样的)consumer devices, including USB flash drives － also known as memory sticks－SD cards, mobile phones, digital cameras, tablet computers, and PC cards in notebook computers and _____(嵌入式控制器).

There are two types of flash memory: NAND and NOR. NAND flash－based solid－state drives (SSDs) are often used to accelerate the performance of I/O－intensive applications. NOR flash memory is often used to hold control code, _____ the BIOS in a PC. Devices, such as camera phones, can use both NOR and NAND flash－in addition to other memory technologies － to facilitate code execution and data storage.

Flash memory evolved _____ EPROM _____ EEPROM. Flash is technically a variant of EEPROM, but the industry reserves the term EEPROM _____ byte－level erasable memory and applies the term flash memory _____ larger block－level erasable memory.

Flash is the least expensive form of _____(半导体存储器). Unlike DRAM and SRAM, flash memory is non－volatile, offers lower _____(功耗)and can be erased in large blocks. Flash drives consume less power and produce less heat than HDDs. These are some _____(优点)of flash memory. Its main _____(缺点)are the wear－out mechanism and cell－to－cell interference as the dies get smaller.

Dr. Fujio Masuoka is credited _____ inventing flash memory when he worked for Toshiba in the 1980s. Masuoka's colleague, Shoji Ariizumi, reportedly coined the term flash because the process of erasing all the data from a semiconductor chip reminded him _____ the flash of a camera.

2. Translate the following passages into Chinese or English.

(1) Many of the products and services in modern society are based upon the work of electrical engineers and computer scientists. The tremendous reduction over the last decades in the cost of digital electronic devices has led to an explosive growth in the use of computers and computation. At the same time, our increased understanding of computer science has made possible the development of new software systems of increased power, sophistication, and flexibility.

(2) In November 1971, Intel introduced the world's first commercial microprocessor, the 4004, invented by three Intel engineers. Primitive by today's standards, it contained a mere 2,300 transistors and performed about 60, 000 calculations in a second. Today, the microprocessor is the most complex mass-produced product ever, with millions of transistors performing hundreds of millions of calculations each second.

(3) Microprocessors are the brains of your personal computer. Here on this tiny silicon chip are millions of switches and pathways that help your computer make important decisions and perform helpful tasks. And microprocessors don't just think for computers — you might find a processor in many other everyday items like your telephone or car.

(4) Memory is the component of a computer (and an embedded system) that is used to store and retain information. There is always a need to store information in an embedded application, so there will always be some type of memory in the design. Many types of memory are available, thus there are many choices for embedded customers.

(5) Memory can be split into two main categories: volatile and non-volatile. Volatile memory loses its content when the system is powered off. Most types of RAM fall into this category. Non-volatile memory does not lose its data when the system or device is turned off.

(6) 现代电子系统数字化程度越来越高，复杂程度也极高。如果没有超大规模芯片技术，则是无法想象的。现代电子系统太复杂了，其设计原理很像大型软件系统设计。在许多方面，计算机科学和电子系统设计需要相同的背景知识。而作为一个不断扩展的工程领域，现代电子系统设计也离不开计算机辅助设计。

(7) 芯片是一小片嵌入了集成电路的半导体材料。芯片的典型尺寸不超过 1 平方英寸，却能包含几百万只晶体管。芯片类型丰富，例如，CPU 芯片包含一个完整的处理单元；而存储芯片包含空白的存储单元。

(8) 微处理器是一片包含 CPU 的硅片。区分微处理器的基本特征有三个：指令集、带宽和时钟速度。

(9) "使用闪存、不用硬盘"的原因有几个。闪存的速度更快、没有噪声、重量更轻、尺寸更小，也不存在可动部件。

(10) 在计算机系统中，存储器是用来运行程序、存储数据的部件。随机存取存储器芯片构成了主存储器。存储器的可用容量决定了运行程序的大小及能否一次运行多个程序。主存储器暂时存储数据。一旦关闭计算机，数据就丢失了。它不同于内部只读存储器、外部存储器，后者的数据存储时间长久。只读存储器包含计算机的核心程序。总之，任何以机器可读方式存储数据的设备都可称为"存储器"。

3. Writing tasks.

Summary: Write an abstract for *Lesson* 1 *VLSI Technology* within 150～200 words.

Composition: Write an essay about some kind of electronic devices which you are familiar with.

Reading Materials

Passage 1 Manufacturing Integrated Circuits

The manufacturing process for integrated circuits (ICs) is critical to the cost of the chips. Figure 1.3 shows that process. The process starts with a silicon crystal ingot, which looks like a giant sausage. Today, ingots are 8 ~ 12 inches in diameter and about 12 ~ 24 inches long. An ingot is finely sliced into wafers no more than 0.1 inches thick. These wafers then go through a series of processing steps, during which patterns of chemicals are placed on each wafer, creating the transistors, conductors, and insulators. Today's integrated circuits contain only one layer of transistors but may have from two to eight levels of metal conductor, separated by layers of insulators.

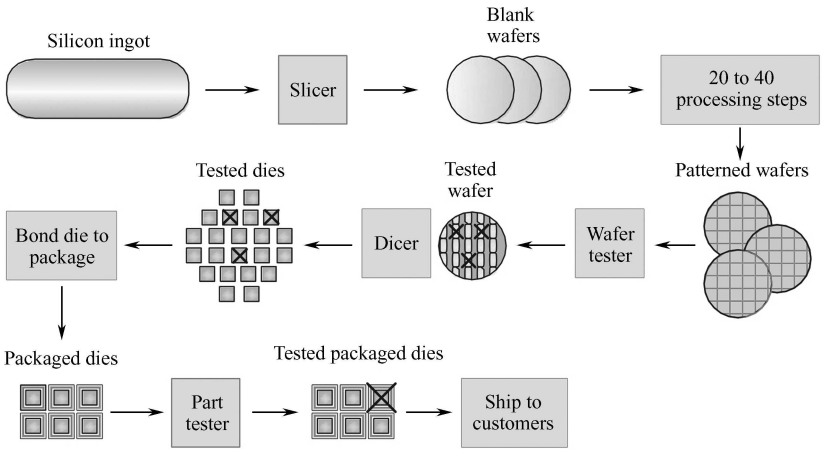


Figure 1.3 The chip manufacturing process

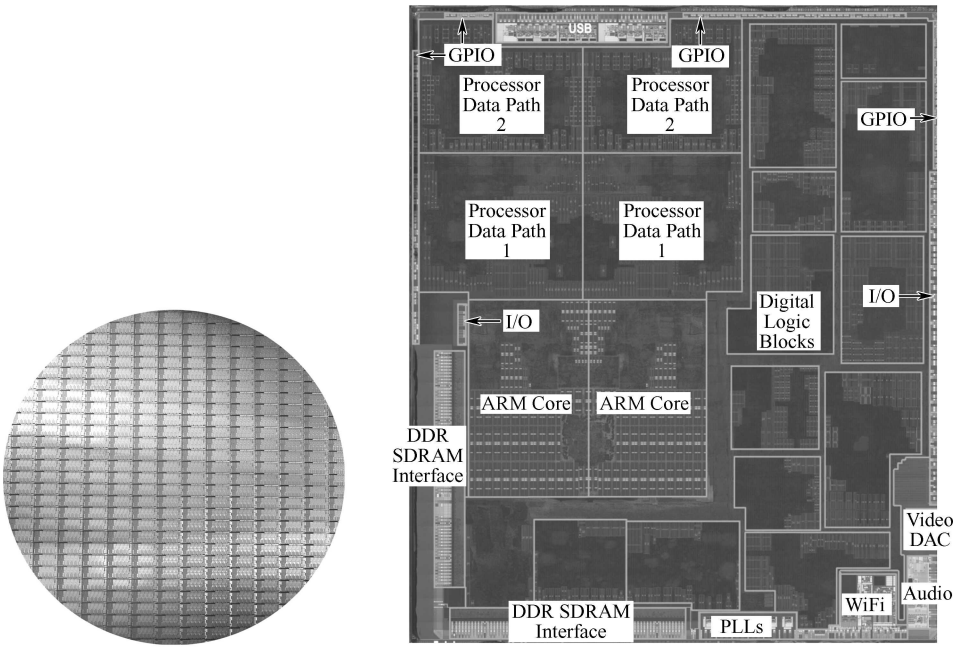
After being sliced from the silicon ingot, blank wafers are put through 20 to 40 steps to create patterned wafers [see Figure 1.4(a)]. These patterned wafers are then tested with a wafer tester, and a map of the good parts is made. Then, the wafers are diced into dies [see Figure 1.4(b)]. In Figure 1.3, one wafer produced 20 dies, of which 17 passed testing. (× means the die is bad.) The yield of good dies in this case was 17/20, or 85%. These good dies are then bonded into packages and tested one more time before shipping the packaged parts to customers. One bad packaged part was found in this final test.

A single microscopic flaw in the wafer itself or in one of the dozens of patterning steps can result in that area of the wafer failing. These defects, as they are called, make it virtually impossible to manufacture a perfect wafer. The simplest way to cope with imperfection is to place many independent components on a single wafer. The patterned wafer is then chopped up, or diced, into these components, called *dies* and more informally known as *chips*.

Figure 1.4(a) shows a photograph of a wafer containing microprocessors before they have been diced. The number of dies on this 300 mm (12 inchs) wafer at 100% yield is 280, each 20.7 by 10.5 mm. The several dozen partially rounded chips at the boundaries of the wafer are

useless; they are included because it's easier to create the masks used to pattern the silicon. This die uses a 32-nanometer technology, which means that the smallest features are approximately 32 nm in size, although they are typically somewhat smaller than the actual feature size, which refers to the size of the transistors as "drawn" versus the final manufactured size.

Figure 1.4(b) shows an individual microprocessor die. The size of chip is 12.1 by 10.1 mm, and it was manufactured originally in a 45-nm process. It has two identical ARM processors or cores in the middle left of the chip and a PowerVR *graphical processor unit* (GPU) with four datapaths in the upper left quadrant. To the left and bottom side of the ARM cores are interfaces to main memory (DRAM).



(a) A 12-inch wafer of Intel Core i7

(b) The microprocessor inside the A5 package

Figure 1.4 The wafers are diced into dies

Dicing enables you to discard only those dies that were unlucky enough to contain the flaws, rather than the whole wafer. This concept is quantified by the yield of a process, which is defined as the percentage of good dies from the total number of dies on the wafer.

The cost of an integrated circuit rises quickly as the die size increases, due both to the lower yield and the smaller number of dies that fit on a wafer. To reduce the cost, using the next generation process shrinks a large die as it uses smaller sizes for both transistors and wires. This improves the yield and the die count per wafer.

Once you've found good dies, they are connected to the input/output pins of a package, using a process called *bonding*. These packaged parts are tested a final time, since mistakes can occur in packaging, and then they are shipped to customers.

The cost of an integrated circuit can be expressed in three simple equations:

$$\text{Cost per die} = \text{Cost per wafer} / (\text{Dies per wafer} \times \text{yield})$$

$$\text{Dies per wafer} \approx \text{Wafer area} / \text{Die area}$$

$$\text{Yield} = 1/ (1 + \text{Defects per area} \times \text{Die area}/2)^2$$

The first equation is straightforward to derive. The second is an approximation, since it does not subtract the area near the border of the round wafer that cannot accommodate the rectangular dies [see Figure 1.4(a)]. The final equation is based on empirical observations of yields at integrated circuit factories, with the exponent related to the number of critical processing steps. Hence, depending on the defect rate and the size of the die and wafer, costs are generally not linear in the die area.

Questions:

1. How is the technical term *yield* defined in this article?
2. What function does the slicer and dicer in Figure 1.3 perform respectively?
3. Could you describe the process *bonding* after good dies have been obtained?
4. For what reasons does the cost of an IC rises quickly as the die size increases?
5. Is the next generation process capable of reducing the cost of chips or not?

Passage 2 Flash Memory

Electronic memory comes in a variety of forms to serve a variety of purposes. Flash memory is used for easy and fast information storage in such devices as digital cameras and home video game consoles. It is used more as a hard drive than as RAM. In fact, flash memory is considered a solid state storage device. Solid state means that there are no moving parts — everything is electronic instead of mechanical. Flash memory is a type of EEPROM chip. It has a grid of columns and rows with a cell that has two transistors at each intersection. The two transistors are separated from each other by a thin oxide layer. One of the transistors is known as a floating gate, and the other one is the control gate. The floating gate's only link to the row, or wordline, is through the control gate. As long as this link is in place, the cell has a value of 1. To change the value to a 0 requires a curious process called Fowler-Nordheim tunneling.

Tunneling is used to alter the placement of electrons in the floating gate. An electrical charge, usually 10 to 13 volts, is applied to the floating gate. The charge comes from the column, or bitline, enters the floating gate and drains to a ground.

This charge causes the floating-gate transistor to act like an electron gun. The excited electrons are pushed through and trapped on other side of the thin oxide layer, giving it a negative charge. These negatively charged electrons act as a barrier between the control gate and the floating gate. A special device called a cell sensor monitors the level of the charge passing through the floating gate. If the flow through the gate is greater than 50 percent of the charge, it has a value of 1. When the charge passing through drops below the 50-percent threshold, the value changes to 0. A blank EEPROM has all of the gates fully open, giving

each cell a value of 1.

The electrons in the cells of a flash-memory chip can be returned to normal ("1") by the application of an electric field, a higher-voltage charge. Flash memory uses in-circuit wiring to apply the electric field either to the entire chip or to predetermined sections known as blocks. This erases the targeted area of the chip, which can then be rewritten. Flash memory works much faster than traditional EEPROMs because instead of erasing one byte at a time, it erases a block or the entire chip, and then rewrites it.

Although standards are flourishing, there are many flash-memory products that are completely proprietary in nature. But it is good to know that as electronic components become increasingly interchangeable and learn to communicate with each other, standardized removable memory will allow you to keep your world close at hand.

Questions:

1. What does the author mean by saying *It is used more as a hard drive than as RAM*?
2. What does the term *solid state device* refer to?
3. Which is the normal state in a flash-memory chip, "1" or "0"?
4. How does flash memory manage to work faster than traditional EEPROMs?
5. What's the difference between a flash-memory and a flash RAM?

Passage 3 Microcontrollers

Microcontrollers are special purpose computers, which do one thing well. If a computer matches a majority of these characteristics, then you can call it a microcontroller:

- Microcontrollers are "embedded" inside some other device (often a consumer product) so that they can control the features or actions of the product.
- Microcontrollers are dedicated to one task and run one specific program. The program is stored in ROM and generally does not change.
- Microcontrollers are often low-power devices. A desktop computer is almost always plugged into a wall socket and might consume 50 watts of electricity. A battery-operated microcontroller might consume 50 milliwatts.
- A microcontroller has a dedicated input device and often (but not always) has a small LED or LCD display for output. For example, the microcontroller inside a TV takes input from the remote control and displays output on the TV screen. The controller controls the channel selector, the speaker system and certain adjustments on the picture tube electronics such as tint and brightness.
- A microcontroller is often small and low cost. The components are chosen to minimize size and to be as inexpensive as possible.
- A microcontroller is often, but not always, ruggedized in some way.

The microcontroller controlling a car's engine, for example, has to work in temperature extremes that a normal computer generally cannot handle. A car's microcontroller in Alaska has to work fine in −30 degrees F (−34℃) weather, while the same microcontroller in Nevada might be operating at 120 degrees F (49℃).

The actual processor used to implement a microcontroller can vary widely. In many products, such as microwave ovens, the demand on the CPU is fairly low and price is an important consideration. In these cases, manufacturers turn to dedicated microcontroller chips — chips that were originally designed to be low-cost, small, low-power, embedded CPUs. The Motorola 6811 and Intel 8051 are both good examples of such chips. There is also a line of popular controllers called "PIC microcontrollers" created by a company called Microchip. By today's standards, these CPUs are incredibly minimalistic; but they are extremely inexpensive when purchased in large quantities and can often meet the needs of a device's designer with just one chip.

Questions:

1. How is a microcontroller defined?
2. What functions does the microcontroller inside a TV perform?
3. Why does a microcontroller often require being ruggedized in some way?
4. What kinds of products are dedicated microcontroller chips often applied to?
5. What does the acronym LED mean in this article?

Unit 2

Electronic Circuits

Lesson 4 Operational Amplifiers

Lesson 5 Low-pass Filters

Lesson 6 Analog-to-Digital Converters

Passage 1 Anti-aliasing Filtering

Passage 2 Switched-capacitor Filters

Passage 3 Digital-to-Analog Converters

Lesson 4 Operational Amplifiers

In 1927 Harold Black[1] commuted from his home in New York City to work at Bell Labs by way of a railroad/ferry. The ferry ride relaxed Harold enabling him to do some conceptual thinking. Harold had a tough problem to solve; when phone lines were extended long distances, they needed amplifiers, and undependable amplifiers limited phone service. First, initial tolerances on the gain were poor, but that problem was quickly solved with an adjustment. Second, even when an amplifier was adjusted correctly at the factory, the gain drifted so much during field operation that the volume was too low or the incoming speech was distorted.

Many attempts had been made to make a stable amplifier, but temperature changes and power supply voltage extremes experienced on phone lines caused uncontrollable gain drift. Passive components had much better drift characteristics than active components had; thus[2] if an amplifier's gain could be made dependent on passive components, the problem would be solved. During one of his ferry trips, Harold's fertile brain conceived a novel solution for the amplifier problem, and he documented the solution while riding on the ferry.

The solution was to first build an amplifier that had more gain than the application required.Then some of the amplifier output signal was fed back to the input in a manner that[3] makes the circuit gain (circuit is the amplifier and feedback components) dependent on the feedback circuit rather than the amplifier gain. Now the circuit gain is dependent on the passive feedback components rather than the active amplifier. This is called negative feedback (see Figure 2.1), and it is the underlying operating principle[4] for all modern-day op amps. Harold had documented the first intentional feedback circuit during a ferry ride. I am sure unintentional feedback circuits had been built prior to that time, but the designers ignored the effect!

The circuit symbols of an open-loop amplifier (without feedback)

The circuit symbols of a closed-loop amplifier (with negative feedback)

Figure 2.1 The negative feedback amplifier

Managers and amplifier designers might say something like this, "it is hard enough to

achieve 30-kHz gain-bandwidth (GBW), and now this fool wants me to design an amplifier with 3-MHz GBW. But, he is still going to get a circuit gain GBW of 30 kHz." Well, time has proven Harold right, but there is a minor problem that Harold didn't discuss in detail, and that is the oscillation[5] problem. It seems that circuits designed with large open-loop gains sometimes oscillate when the loop is closed. A lot of people investigated the instability effect, and it was pretty well understood in the 1940s, but solving stability problems involved long, tedious, and intricate calculations. Years passed without anybody making the problem solution simpler or more understandable.

In 1945, H. W. Bode[6] presented a system[7] for analyzing the stability of feedback systems by using graphical methods (see Figure 2.2). Until this time, feedback analysis was done by multiplication and division, so calculation of transfer functions was a time-consuming and laborious task. Remember, engineers did not have calculators or computers until the 1970s. Bode presented a log technique that transformed the intensely mathematical process of calculating a feedback system's stability into graphical analysis that was simple and perceptive. Feedback system design was still complicated, but it no longer was an art dominated by a few electrical engineers kept in a small dark room. Any electrical engineer could use Bode's methods to find the stability of a feedback circuit, so the application of feedback to machines began to grow. There really wasn't much call for electronic feedback design until computers and transducers become of age.

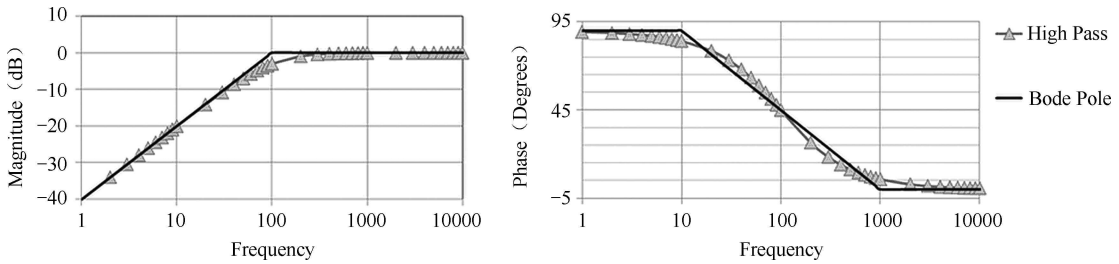


Figure 2.2 The Bode plot of a high-pass filter: magnitude response (left) and phase response (right)

The first real-time computer was the analog computer! This computer used preprogrammed equations and input data to calculate control actions. The programming was hardwired with a series of circuits that performed math operations on the data, and the hard-wiring limitation eventually caused the declining popularity of the analog computer. The heart of the analog computer was a device called operational amplifier because it could be configured to perform many mathematical operations such as multiplication, addition, subtraction, division, integration, and differentiation on the input signals. The name was shortened to the familiar op amp, as we have come to know and love them. The op amp used an amplifier with a large open-loop gain, and when the loop was closed, the amplifier performed the mathematical operations dictated by the external passive components. This amplifier was very large because it was built with vacuum tubes and it required a high-voltage power supply, but it was the heart of the analog computer, thus its large size and huge power requirements were

accepted. Many early op amps were designed for analog computers, and it was soon found out that op amps had other uses and were very handy to have around the physics lab.

At this time general-purpose analog computers were found in universities and large company laboratories because they were critical to the research work done there. There was a parallel requirement for transducer signal conditioning in lab experiments, and op amps found their way into signal conditioning applications. As the signal conditioning applications expanded, the demand for op amps grew beyond the analog computer requirements, and even when the analog computers lost favor to digital computers, the op amp survived because of its importance in universal analog applications. Eventually digital computers replaced the analog computers, but the demand for op amps increased as measurement applications increased.

The first signal conditioning op amps were constructed with vacuum tubes prior to the introduction of transistors, so they were large and bulky. During the1950s, miniature vacuum tubes that worked from lower voltage power supplies enabled the manufacture of op amps that shrunk to the size of a brick used in house construction, so the op amp modules were nicknamed bricks. Vacuum tube size and component size decreased until an op amp was shrunk to the size of a single octal vacuum tube. Transistors were commercially developed in the 1960s, and they further reduced op amp size to several cubic inches. Most of these early op amps were made for specific applications, so they were not necessarily general purpose. The early op amps served a specific purpose, but each manufacturer had different specifications and packages; hence, there was little second sourcing among the early op amps.

ICs were developed during the late 1950s and early 1960s, but it wasn't till the middle 1960s that Fairchild[8] released the μA709. This was the first commercially successful IC op amp. The μA709 had its share of problems, but any competent analog engineer could use it, and it served in many different analog applications. The major drawback of the μA709 was stability; it required external compensation and a competent analog engineer to apply it. Also, the μA709 was quite sensitive because it had a habit of self-destruction under any adverse condition. The μA741 followed the μA709, and it is an internally compensated op amp that does not require external compensation if operated under data sheet conditions. There has been a never-ending series of new op amps released each year since then, and their performance and reliability has improved to the point where present-day op amps can be used for analog applications by anybody.

The IC op amp is here to stay; the latest generation op amps cover the frequency spectrum from 5-kHz GBW to beyond 1-GHz GBW. The supply voltage ranges from guaranteed operation at 0.9 V to absolute maximum voltage ratings of 1000 V. The input current and input offset voltage has fallen so low that customers have problems verifying the specifications during incoming inspection. The op amp has truly become the universal analog IC because it performs all analog tasks.It can function as a line driver[9], comparator (one bit A/D), amplifier, level shifter, oscillator, filter, signal conditioner, actuator driver, current source, voltage source, and etc.. The designer's problem is how to rapidly select the correct

circuit (op amp combination) and then, how to calculate the passive component values that yield the desired transfer function in the circuit.

The op amp will continue to be a vital component of analog design because it is such a fundamental component. Each generation of electronic equipment integrates more functions on silicon and takes more of the analog circuitry inside the IC. As digital applications increase, analog applications also increase because the predominant supply of data and interface applications are in the real world, and the real world is an analog world. Thus, each new generation of electronic equipment creates requirements for new analog circuits; hence, new generations of op amps are required to fulfill these requirements. Analog design, and op amp design, is a fundamental skill that will be required far into the future.

New Words & Expressions

commute [kəˈmjuːt] *vi*. 通勤
by way of ... 经由……
initial [ɪˈnɪʃəl] *adj*. 初始的
tolerance [ˈtɒlərəns] *n*. 公差
adjustment [əˈdʒʌstmənt] *n*. 调整
fertile [ˈfɜːtaɪl] *adj*. 富饶的
conceive [kənˈsiːv] *vt*. 设想
underlying [ˌʌndəˈlaɪɪŋ] *adj*. 根本的
intentional [ɪnˈtenʃənəl] *adj*. 有意的
prior to ... 在……之前
investigate [ɪnˈvestɪˌgeɪt] *vt*. 研究
tedious [ˈtiːdiəs] *adj*. 单调乏味的
intricate [ˈɪntrɪkət] *adj*. 错综复杂的
perceptive [pəˈseptɪv] *adj*. 能理解的
dominate [ˈdɒmɪˌneɪt] *vt*. 主宰
a series of ... 一系列……
dictate [ˈdɪkˌteɪt] *vt*. 决定
bulky [ˈbʌlki] *adj*. 体积大的
miniature [ˈmɪnətʃə] *adj*. 微型的
octal [ˈɒktəl] *adj*. 八引脚的
cubic [ˈkjuːbɪk] *adj*. 立方的
hence [hens] *adv*. 因此
sourcing [ˈsɔːsɪŋ] *n*. 供货
release [rɪˈliːs] *vt*. 发布
competent [ˈkɒmpɪtənt] *adj*. 胜任的
drawback [drɔːˌbæk] *n*. 缺点
compensation [ˌkɒmpenˈseɪʃn] *n*. 补偿
adverse [ˈædvɜːs] *adj*. 不利的
to the point where ... 发展到……程度
range from ... to ... 从……变化到……
guarantee [ˌgærənˈtiː] *vt*. 保证
rating [ˈreɪtɪŋ] *n*. 级别
verify [ˈverɪˌfaɪ] *vt*. 核实
function as ... 实现……功能
etc. [etˈsetərə] *abbr*. 诸如此类
yield [jiːld] *vt*. 得出(结果、答案、信息)
fundamental [ˌfʌndəˈmentl] *adj*. 基础的
integrate [ˈɪntəˌgreɪt] *vt*. 集成
predominant [prɪˈdɒmɪnənt] *adj*. 显著的
fulfill [fʊlˈfɪl] *vt*. 满足

Technical Terms

amplifier [ˈæmplɪˌfaɪə] *n*. 放大器
volume [ˈvɒljuːm] *n*. 音量
passive [ˈpæsɪv] *adj*. 无源的
active [ˈæktɪv] *adj*. 有源的

feedback [ˈfiːdbæk] *n*. 反馈
oscillation [ˌɒsɪˈleɪʃn] *n*. 振荡
stability [stəˈbɪlɪti] *n*. 稳定性
transducer [trænzˈdjuːsə] *n*. 传感器
integration [ˌɪntəˈgreɪʃn] *n*. 积分
differentiation [ˌdɪfərenʃiˈeɪʃn] *n*. 微分
specification [ˌspesɪfɪˈkeɪʃn] *n*. 性能指标
package [ˈpækɪdʒ] *n*. 封装形式
actuator [ˈæktjʊeɪtə] *n*. 激励器
circuitry [ˈsɜːkɪtri] *n*. 电路系统
interface [ˈɪntəˌfeɪs] *n*. 接口电路
phone service　电话业务
field operation　现场运行
power supply voltage　电源电压
gain drift　增益漂移
feedback component　反馈元件
open-loop gain　开环增益
transfer function　传递函数
vacuum tube　真空管
signal conditioning　信号调理
external compensation　外部补偿
data sheet　数据手册
input offset voltage　输入偏置电压
current source　电流源
op amp *abbr*. operational amplifier　运算放大器
GBW *abbr*. gain-bandwidth　增益带宽积
log [lɒg] *abbr*. logarithm [ˈlɒgərɪðəm]　对数
IC *abbr*. integrated circuit　集成电路

Notes

1. 布莱克(Harold Stephen Black,1898—1983)于 1927 年发明负反馈放大器,于 1934 年发表论文 *Stabilized Feedback Amplifiers*。负反馈放大器的发明是 20 世纪电子技术的关键突破。
2. thus 和 therefore 为副词,其功能相当于连词 so,但须注意其用法差别。示例如下:I was hungry, *so* I bought a sandwich. / I was hungry; *therefore*, I bought a sandwich. / I was hungry; *thus* I bought a sandwich. / I was hungry. *Therefore*, I bought a sandwich. / I was hungry. *Thus* I bought a sandwich. 此外,hence 的用法与 therefore 相同。hence 和 thus 仅用于正式写作中。
3. in a manner that ... / in a ... manner 意为“以……方式”。例如,“以无计划的方式去解

决问题”(tackle the problem *in a* haphazard *manner*)。

4. 注意区分 principle 和 principal。前者只作为名词,意为“原理”“准则”。后者可作为名词,意为“大学校长”“负责人”或“本金”;也可作为形容词,意为“首要的”。
5. 与 oscillation(振荡)易混的词有 vibration(振动)、fluctuation(波动)、resonance(谐振)和 vacillation(摆动)。
6. 伯德(Hendrik Wade Bode,1905—1982)是现代控制理论和电子通信技术先驱者之一,他设计的“伯德图”(Bode plot)采用“对数频率”描绘“线性时不变系统”的频率响应。
7. system 此处应译作“方法”,其常见译法有“专家系统”(expert system)、“十进制”(decimal system)、“太阳系”(the solar system)和“填料方法”(filing system)。
8. Fairchild 是指飞兆半导体公司,其总部位于美国加利福尼亚州的圣何塞(San Jose, California)。Fairchild 曾被译作“仙童半导体公司”。
9. “线路驱动器”(line driver)可指用于驱动多个逻辑负载、低阻抗总线 I/O 或远距离传送信号的功率放大器,也可指用于串并转换、控制字符插入或数据缓存的硬件单元。

Lesson 5 Low-pass Filters

First-order Low-pass Filters

An integrator [see Figure 2.3(a)] is the simplest filter mathematically, and it forms the building block for most modern integrated filters. Consider what we know intuitively about an integrator. If you apply a DC signal at the input (i.e., zero frequency), the output will describe a linear ramp that grows in amplitude until limited by the power supplies. Ignoring that limitation, the response of an integrator at zero frequency is infinite, which means that it has a pole at zero frequency. (A pole exists at any frequency for which the transfer function's value becomes infinite.)

We also know that the integrator's gain diminishes with increasing frequency and that at high frequencies the output voltage becomes virtually zero. Gain is inversely proportional[1] to frequency, so it has a slope of −1 when plotted on log/log coordinates [i.e., −20 dB/decade on a Bode plot, see Figure 2.3(b)].

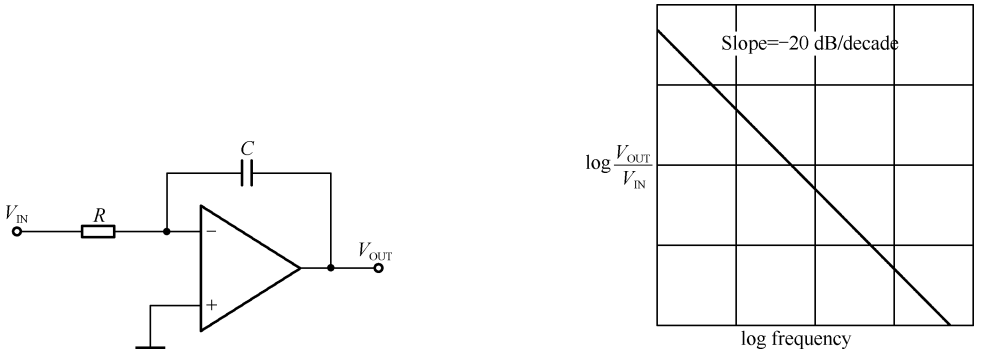
You can easily derive the transfer function as

$$\frac{V_{\mathrm{OUT}}}{V_{\mathrm{IN}}}=\frac{X_{\mathrm{C}}}{R}=\frac{1/sC}{R}=\frac{\omega_0}{s}$$

where s is the complex-frequency variable $\sigma+\mathrm{j}\omega$ and ω_0 is $1/RC$. If we think of s as frequency, this formula confirms the intuitive feeling that gain is inversely proportional to frequency.

A slightly more complex filter is the simple RC low-pass type [see Figure 2.4(a)]. Its transfer function is

$$\frac{V_{\mathrm{OUT}}}{V_{\mathrm{IN}}}=\frac{1/sC}{R+1/sC}=\frac{1}{1+sCR}=\frac{\omega_0}{s+\omega_0}$$



(a) A simple RC integrator (b) A Bode plot of a simple integrator

Figure 2.3 The RC integrator and its Bode plot

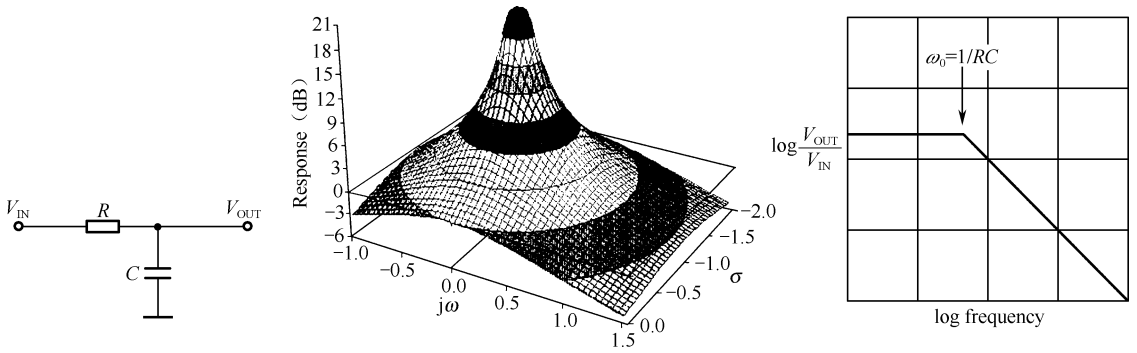
When $s=0$, the function reduces to ω_0/ω_0, i.e., 1. When s increases to infinity, the function approaches zero, so this is a low-pass filter. When $s=-\omega_0$, the denominator is zero and the function's value is infinite, indicating a pole in the complex plane. The magnitude of the transfer function is plotted against s in Figure 2.4(b), where the real component of $s(\sigma)$ is toward us and the positive imaginary part ($j\omega$) is toward the right. The pole at $-\omega_0$ is evident. Amplitude is shown logarithmically to emphasize the function's form. For both the integrator and the RC low-pass filter, frequency response approaches zero at infinite frequency; that is, there is a zero at $s=\infty$. This single zero surrounds the complex plane.

But how does the complex function in s relate to the circuit's response to actual frequencies? When analyzing the response of a circuit to AC signals, we use the expression $j\omega L$ for impedance of an inductor and $1/j\omega C$ for that[2] of a capacitor. When analyzing transient response using Laplace[3] transforms, we use sL and $1/sC$ for the impedance of these elements. The similarity is apparent immediately. The $j\omega$ in AC analysis is in fact the imaginary part of s, which, as mentioned earlier, is composed of[4] a real part σ and an imaginary part $j\omega$.

If we replace s by $j\omega$ in any equation so far, we have the circuit's response to an angular frequency. In the complex plot in Figure 2.4(b), $\sigma=0$ and hence $s=j\omega$ along the positive $j\omega$ axis. Thus, the function's value along this axis is the frequency response of the filter. We have sliced the function along the $j\omega$ axis and emphasized the RC low-pass filter's frequency-response curve[5] by adding a heavy line for function values along the positive $j\omega$ axis. The more familiar Bode plot [see Figure 2.4(c)] looks different in form only because the frequency is expressed logarithmically.

While the complex frequency's imaginary part ($j\omega$) helps describe a response to AC signals, the real part (σ) helps describe a circuit's transient response. Looking at Figure 2.4(b), we can therefore say something about the RC low-pass filter's response as compared to that of the integrator. The low-pass filter's transient response is more stable, because its pole is in the negative real half of the complex plane. That is, the low-pass filter makes a decaying-

exponential response to a step-function input; the integrator makes an infinite response. For the low-pass filter, pole positions further down the $-\sigma$ axis mean a higher ω_0, a shorter time constant, and therefore a quicker transient response. Conversely, a pole closer to the $j\omega$ axis causes a longer transient response.



(a) A simple RC low-pass filter

(b) The complex function of an RC low-pass filter

(c) A Bode plot of a low-pass filter

Figure 2.4 The complex function and Bode plot of RC low-pass filter

So far, we have related[6] the transfer functions of some simple circuits to their associated poles and zeroes in the complex-frequency plane. From these functions, we have derived the circuit's frequency response (and hence its Bode plot) and also its transient response. Because both the integrator and the RC filter have only one s in the denominator of their transfer functions, they each have only one pole. That is, they are first-order filters.

However[7], as we can see from Figure 2.3(b), the first-order filter does not provide a very selective frequency response. To tailor a filter more closely to our needs, we must move on to higher orders. From now on, we will describe the transfer function using $f(s)$ rather than the cumbersome V_{OUT}/V_{IN}.

Second-order Low-pass Filters

A second-order filter has s^2 in the denominator and two poles in the complex plane. You can obtain such a response by using inductance and capacitance in a passive circuit or by creating an active circuit of resistors, capacitors, and amplifiers. Consider the passive RLC filter in Figure 2.5(a), for instance[8]. We can show that its transfer function has the form

$$f(s)=\frac{X_C}{R+X_L+X_C}=\frac{1/sC}{R+sL+1/sC}=\frac{1}{LCs^2+RCs+1}$$

and if we define $\omega_0^2=1/LC$ and $Q=\omega_0 L/R$, then

$$f(s)=\frac{\omega_0^2}{s^2+s\omega_0/Q+\omega_0^2}$$

where ω_0 is the filter's characteristic frequency and Q is the quality factor (lower R means higher Q).

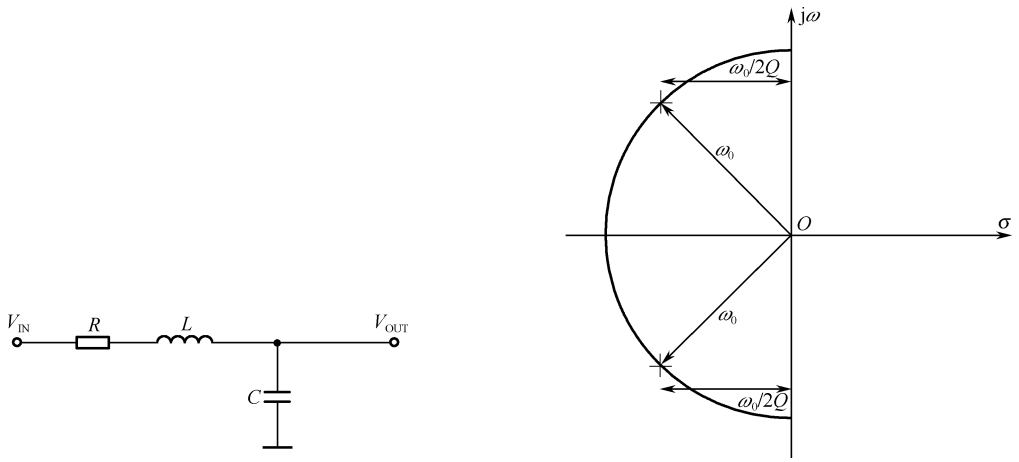
The poles occur at s values for which the denominator becomes zero; that is, when s^2+

$s\omega_0/Q + \omega_0^2 = 0$. We can solve this equation by remembering that the roots of $ax^2 + bx + c = 0$ are given by

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

In this case, $a = 1$, $b = \omega_0/Q$, and $c = \omega_0^2$. The term $(b^2 - 4ac)$ equals $\omega_0^2(1/Q^2 - 4)$, so if Q is less than 0.5 then both roots are real and lie on the negative real axis. The circuit's behavior is much like that of two first-order RC filters in cascade. This case isn't very interesting, so we'll consider only the case where $Q > 0.5$, which means $(b^2 - 4ac)$ is negative and the roots are complex.

The real part is therefore $-b/2a$, which is $-\omega_0/2Q$, and common to both roots. The roots' imaginary parts will be equal and opposite in signs. Calculating the position of the roots in the complex plane, we find that they lie at a distance of ω_0 from the origin, as shown in Figure 2.5(b).



(a) An RLC low-pass filter (b) A pole-zero diagram of an RLC low-pass filter

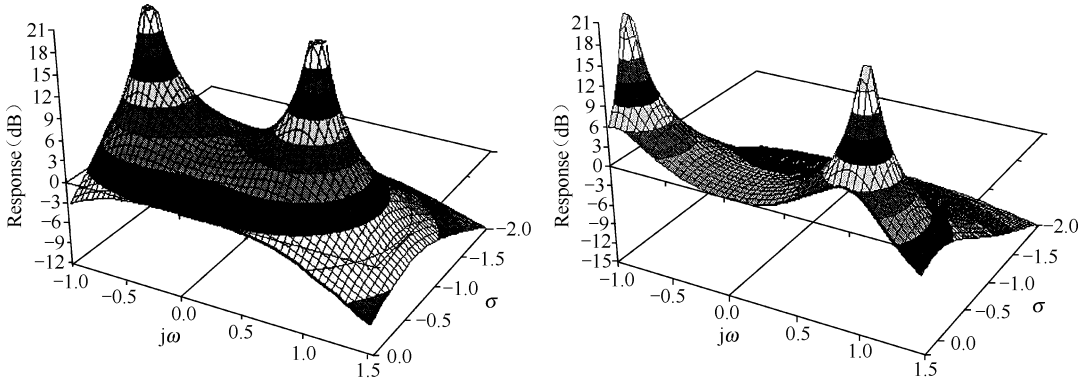
Figure 2.5 The RLC low-pass filter and its pole-zero diagram

Varying ω_0 changes the poles' distance from the origin. Decreasing the Q moves the poles toward each other, whereas increasing the Q moves the poles in a semicircle[9] away from each other and toward the jω axis. When $Q = 0.5$, the poles meet at $-\omega_0$ on the negative real axis. In this case, the corresponding circuit is equivalent to two cascaded first-order filters.

Now let's examine the second-order function's frequency response and see how it varies with Q. As before, Figure 2.6(a) shows the function as a curved surface, depicted in the three-dimensional space formed by the complex plane and a vertical magnitude vector. Further, $Q = 0.707$, and you can see immediately that the response is a low-pass filter.

The effect of increasing the Q is to move the poles in a circular path toward the jω axis. Figure 2.6(b) shows the case where $Q = 2$. Because the poles are closer to the jω axis, they have a greater effect on the frequency response, causing a peak at the high end of the passband.

There is also an effect on the filter's transient response. Because the poles' negative real part is smaller, an input step function will cause ringing at the filter output. Lower values of Q result in less ringing, because the damping is greater. On the other hand, if Q becomes infinite, the poles reach the $j\omega$ axis, causing an infinite frequency response (instability and continuous oscillation) at $\omega=\omega_0$. In the LCR circuit in Figure 2.5(a), this condition would be impossible unless $R=0$. For filters that contain amplifiers, however, the condition is possible and must be considered in the design process.



(a) The complex function of a second-order low-pass filter ($Q=0.707$)

(b) The complex function of a second-order low-pass filter($Q=2$)

Figure 2.6 The complex functions of second-order low-pass filters with different Q values

A second-order filter provides the variables ω_0 and Q, which allow us to place poles wherever we want in the complex plane. These poles must, however, occur as complex-conjugate pairs, in which the real parts are equal and the imaginary parts have opposite signs. This flexibility in pole placement is a powerful tool and one that makes the second-order stage a useful component in many switched-capacitor filters. As in the first-order case, the second-order low-pass transfer function approaches zero as frequency increases to infinity. The second-order function decreases twice as fast, however, because of the s^2 factor in the denominator. The result is a double zero at infinity.

New Words & Expressions

order [ˈɔːdə] *n*. 阶
intuitively [ɪnˈtjʊɪtɪvli] *adv*. 直觉地
describe [dɪˈskraɪb] *vt*. 描述
infinite [ˈɪnfɪnət] *adj*. 无穷的
diminish [dɪˈmɪnɪʃ] *vi*. 减小
proportional [prəˈpɔːʃənl] *adj*. 成比例的
decade [ˈdekeɪd] *n*. 10 倍
derive [dɪˈraɪv] *vt*. 推导
curve [kəːv] *n*. 曲线
conversely [kənˈvəːsli] *adv*. 相反地
associated [əˈsəʊʃiətɪd] *adj*. 相关的
cumbersome [ˈkʌmbəsəm] *adj*. 烦琐的
characteristic [ˌkærəktəˈrɪstɪk] *n*. 特征
factor [ˈfæktə] *n*. 因数
term [təːm] *n*. 项
opposite [ˈɒpəzɪt] *adj*. 相反的

variable [ˈveərɪəbl] *n*. 变量
formula [ˈfɔːmjələ] *n*. 公式
reduce to ... 简化为……
indicate [ˈɪndɪkeɪt] *vt*. 显示
positive [ˈpɒzətɪv] *adj*. 正的
analyze [ˈænəˌlaɪz] *vt*. 分析
expression [ɪkˈspreʃn] *n*. 表达式
transient [ˈtrænziənt] *adj*. 短暂的
similarity [ˌsɪmɪˈlerəti] *n*. 相似性
apparent [əˈpærənt] *adj*. 显然的
distance [ˈdɪstəns] *n*. 距离
origin [ˈɒrɪdʒɪn] *n*. 原点
decrease [diːˈkriːs] *vi*. 减小
increasing [ɪnˈkriːsɪŋ] *adj*. 增大的
whereas [weərˈæz] *conj*. 尽管
corresponding [ˌkɒrəˈspɒndɪŋ] *adj*. 相应的
equivalent [ɪˈkwɪvələnt] *adj*. 等效的
vary with ... 随……变化
vertical [ˈvɜːtɪkəl] *adj*. 垂直的
result in ... 导致……结果
continuous [kənˈtɪnjuːəs] *adj*. 连续的
flexibility [ˌfleksəˈbɪləti] *n*. 灵活性

Technical Terms

integrator [ˈɪntɪgreɪtə] *n*. 积分器
pole [pəʊl] *n*. 极点
slope [sləʊp] *n*. 斜率
coordinate [kəʊˈɔːdɪnət] *n*. 坐标
infinity [ɪnˈfɪnəti] *n*. 无穷大
denominator [dɪˈnɒmɪˌneɪtə] *n*. 分母
impedance [ɪmˈpiːdəns] *n*. 阻抗
inductor [ɪnˈdʌktə] *n*. 电感器
capacitor [kəˈpæsɪtə] *n*. 电容器
angular [ˈæŋgjələ] *adj*. 角度的
axis [ˈæksɪs] *n*. 坐标轴
exponential [ˌekspəˈnenʃl] *adj*. 指数的
selective [sɪˈlektɪv] *adj*. 选择性的
inductance [ɪnˈdʌktəns] *n*. 电感(值)
capacitance [kəˈpæsɪtəns] *n*. 电容(值)
cascade [kæsˈkeɪd] *n*. 级联
vector [ˈvektə] *n*. 向量
circular [ˈsɜːkjʊlə] *adj*. 圆形的
passband [ˈpæsbænd] *n*. 通带
ringing [ˈrɪŋɪŋ] *n*. 振铃振荡
damping [ˈdæmpɪŋ] *n*. 幅度衰减
conjugate [ˈkɒndʒəgeɪt] *adj*. 共轭的
linear ramp 线性斜坡
transfer function 传递函数
Bode plot 伯德图

complex plane　复平面
real part　实部
imaginary part　虚部
transient response　暂态响应
AC analysis　交流分析
angular frequency　角频率
step function　阶跃函数
time constant　时间常数
characteristic frequency　特征频率
quality factor　品质因数
switched-capacitor filter　开关电容滤波器
DC *abbr*. direct current　直流电
AC *abbr*. alternating current　交流电
dB *abbr*. decibel [ˈdesɪˌbel]　分贝
RC *abbr*. resistor capacitor　电阻电容

Notes

1. *Y is directly/inversely proportional to X* 意为“*Y* 与 *X* 成正/反比”。若使用名词 proportion，则表达为 *Y is in direct/inverse proportion to X*。
2. 此处 that 用以指代上文提到的名词(impedance)。若被指代名词为复数，则要使用代词 those。这是一种常见的省略用法，本课中出现了三处：(1) When analyzing the response of a circuit to AC signals, we use the expression jωL for impedance of an inductor and 1/jωC for *that* of a capacitor. (2) Looking at Figure 2.4 (b), we can therefore say something about the RC low-pass filter's response as compared to *that* of the integrator. (3) The circuit's behavior is much like *that* of two first-order RC filters in cascade.
3. 拉普拉斯(Pierre-Simon Laplace，1749—1827)是法国天文学家、数学家和物理学家，因研究太阳系的稳定性而著称。除“拉普拉斯变换”(Laplace transform)外，科技文献中还常见到“拉普拉斯算子”(Laplace operator)“拉普拉斯方程”(Laplace equation)等。
4. “由　　组成”可采用 be composed of .../be made up of .../be comprised of ... / consist of ...这四种表达结构之一，其区别仅在于正式程度和描述对象。
5. curve 意为“曲线”。curve、graph、chart 或 plot 均可表示“曲线图”，但须注意固定搭配。
6. relate *X* and/to/with/*Y* 表示“将 *X* 与 *Y* 联系起来”。relate to sb./sth.则表示“涉及某人/某物”，如“第 9 章是关于 *X* 对 *Y* 的影响”(Chapter nine *relates to* the effects of *X* on *Y*)。
7. 副词 however 用于表达转折，其同义词为 Nevertheless/Nonetheless，但后者仅用于书面语。
8. for instance 相当于 for example，可用在句首、句中或句尾。用在句首时，若前句以分号

结尾，则首字母需小写；若前句以句号结尾，则首字母需大写。

9. semi-意为“一半”或“部分”，如“半圆”(semicircle)、“半球”(semisphere)、“半导体”(semiconductor)、“半自动的”(semiautomatic)等。

Lesson 6 Analog-to-Digital Converters

Analog-to-digital converters (ADCs) come in[1] almost as many flavors as ice creams, and at least as much care is needed in choosing the former as is required with the latter.

Flash[2] ADC

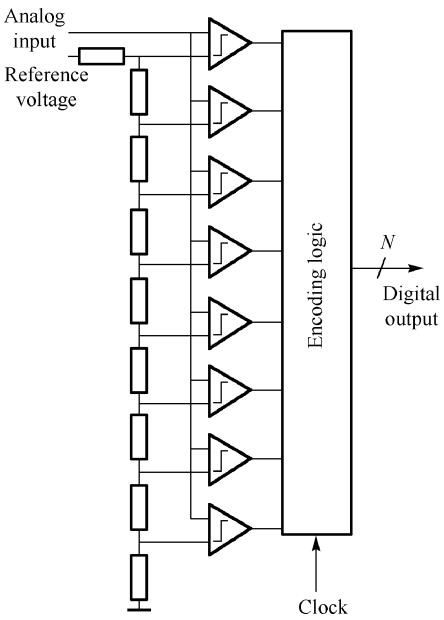


Figure 2.7 Flash ADC block diagram

A popular and readily understood type of ADC is the Flash ADC (see Figure 2.7). This is capable of very high-speed conversion and thus can accommodate high sampling rates, but in its basic form is very power hungry.

The flash converter operates by simultaneously presenting the input signal to a bank[3] of $2^N - 1$ comparators, whose reference voltages are set by a resistor chain to exactly correspond[4] to all of the possible sample levels which can be represented by the converter. The output from each comparator (either a 1 or a 0) is then encoded into an N-bit word representing the input sample level. This approach is the most simple, most intuitive and also the fastest solution for ADC implementation. For large numbers of bits (e.g. $N > 14$ bits), the number of resistors needed ($2^N - 1$) becomes prohibitively large for most practical applications. Also, the power consumption is considerably higher than some of the slightly slower and more exotic solutions.

Figure 2.8 shows the output spectrum of a typical flash converter as determined by taking an FFT (fast Fourier transform) of the converter output samples for a pure sine wave input. One thing is immediately apparent. The spectrum does not simply consist of the pure input sine wave component, but also has a mass of other components spread throughout the measurement band. These largely arise[5] from the inevitable quantization error (noise), because the converter is trying to represent the analog input level from a finite number of available sample values (dictated by the number of bits in the ADC).

For our test full-scale sine wave, the effective SNR (signal-to-noise ratio) as a function of

converter resolution can be readily determined (with a few basic assumptions) to be SNR = (6.02N + 1.76) dB, where N is the number of bits for the converter.

The converter resolution for the ADC generating the plot in Figure 2.8 is 12 bits, giving a theoretical SNR of 74 dB. Looking at the plot, the difference in levels between the sine wave component and the individual noise components is much nearer 104 dB. The reason for the difference between these two values is that the SNR formula[6] refers to the whole noise contribution, comprising the sum of all the individual noise components making up the FFT. The reason for pointing out this feature is it can be used to increase the effective resolution of a converter by trading off sampling rate as shown below.

Over-sampling to Achieve Processing Gain

Assume that we need to achieve a minimum 70 dB SNR in the conversion process for a given audio application. The formula above suggests a minimum of 12-bit converter resolution is required (full-scale[7] sine wave input and ideal converter), where the noise within the whole band from 0 Hz to $f_s/2$ is included in the measurement. Now, if we employ for example 8 × bandwidth sampling (see Figure 2.9), we see that the actual audio signal only occupies 1/4 of the baseband (0 to $f_s/2$) space whereas the noise is spread uniformly across the band. If we were to now digitally filter this sampled signal, we could remove approximately 3/4 of the noise, increasing the SNR by a factor of[8] 4 or 6 dB. This effective increase in SNR is termed the processing gain, achieved by over-sampling the input relative to the 2 × bandwidth rule.

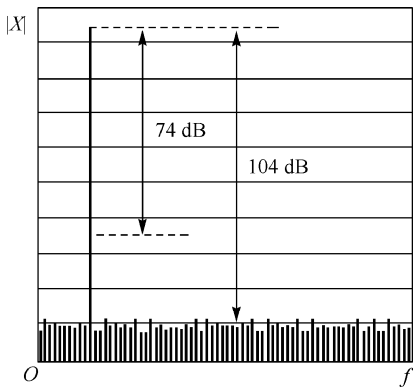


Figure 2.8 FFT of typical ADC samples

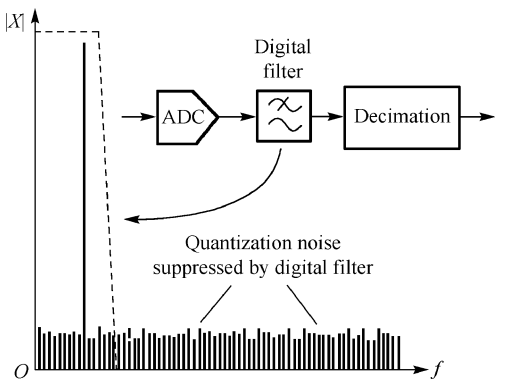


Figure 2.9 Over-sampling and quantization noise

A simple formula for the maximum processing gain can easily be derived: processing gain (dB) = 10 log(sample rate/2 × signal bandwidth), where it is assumed that digital filtering is employed to restrict the sample bandwidth to exactly match the wanted input signal bandwidth and that the noise is uniformly distributed.

Thus, if we use a 128 times over-sampling design (typically found in minidisk recorders and PC sound cards), we can achieve a real 18 dB improvement in SNR, effectively increasing the resolution of the converter from 12 bits to 15 bits. By way of an example, consider using

this method to improve the performance of a data converter within a digital cellular phone. The signal bandwidth for a GSM cellular channel is 200 kHz. It is now possible to obtain high speed ADCs with a sampling rate of 80 MSPS (million samples per second) and 14-bit resolution, giving an impressive measured 75 dB SNR over the full 0 to $f_s/2$ bandwidth. The processing gain possible is thus: processing gain (dB) = 10 log (80 000 000/2 × 200 000) = 26 dB, resulting in a very respectable 75 + 26 = 101 dB SNR for the sampled GSM signal.

Sigma-delta ADC

This concept of processing gain leads us nicely into the topic of sigma-delta converters. These take the notion of processing gain to the extreme to achieve very high performance, low cost, low power devices ideally suited for audio applications, with simple analog interfacing (little or no anti-aliasing filters). A block diagram of a basic sigma-delta converter is shown in Figure 2.10.

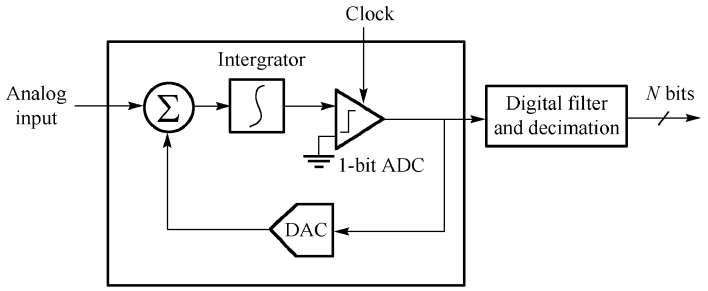


Figure 2.10 Basic sigma-delta converter

The converter is essentially a highly over-sampling 1-bit ADC (the comparator) followed by digital filtering and decimation to realize the processing gain. The effective performance of the converter is greatly enhanced by the addition of circuitry to shape the quantization noise such that, instead of being uniformly spread throughout the 0 to $f_s/2$ band, it is minimized in the band of interest[9] (see Figure 2.11).

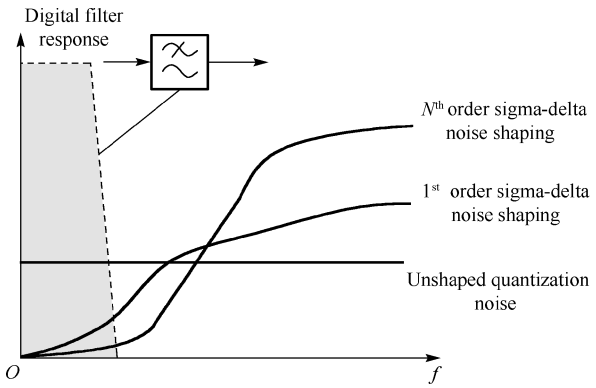


Figure 2.11 Noise shaping in sigma-delta converters

For a typical 128 × bandwidth over-sampling system, the processing gain alone would

give an extra 3 bits of effective resolution (i.e. a 4-bit converter). Noise shaping however can extend this effective resolution much further, with some sigma-delta converters now achieving 24-bit accuracy for audio band applications. The modern converters use a much more sophisticated form of noise shaping processing than that shown in Figure 2.11, which is a simple first-order sigma-delta design, but the basic principle of exploiting processing gain with noise shaping remains the same.

In fact, there are many tens of ADC methods in use, Successive Approximation, Multipass, Interpolating, Subranging, Bit-Per-Stage, to name but a few of the flavors, each potentially having some benefit in performance over its rivals. Luckily, it is not usually necessary to understand how the ADC works in order to make the correct choice of converter for the application. Instead, careful study of the performance specifications on the data sheet will determine the best choice for your application.

New Words & Expressions

conversion [kənˈvɜːʃn] *n*. 转换
accommodate [əˈkɒmədeɪt] *vt*. 容纳
simultaneously [saɪməlˈteɪnɪəsli] *adv*. 同时地
correspond [ˌkɒrɪˈspɒnd] *vi*. 相一致
represent [ˌreprɪˈzent] *vt*. 代表
approach [əˈprəʊtʃ] *n*. 方法
intuitive [ɪnˈtjuːɪtɪv] *adj*. 直觉的；直观的
solution [səˈluːʃn] *n*. 解决方案
implementation [ˌɪmplɪmənˈteɪʃn] *n*. 实现
prohibitively [prəʊˈhɪbɪtɪvli] *adv*. 过高地
application [ˌæplɪˈkeɪʃn] *n*. 应用
exotic [ɪgˈzɒtɪk] *adj*. 特殊的
apparent [əˈpærənt] *adj*. 显然的
component [kəmˈpəʊnənt] *n*. 部件
inevitable [ɪnˈevɪtəbl] *adj*. 不可避免的
finite [ˈfaɪnaɪt] *adj*. 有限的
available [əˈveɪləbl] *adj*. 可获得的
dictate [dɪkˈteɪt] *vt*. 决定；规定
effective [ɪˈfektɪv] *adj*. 有效的
assumption [əˈsʌmpʃən] *n*. 假设；假定
theoretical [ˌθɪəˈretɪkəl] *adj*. 理论上的
trade off 折中；权衡
uniformly [ˈjuːnɪfɔːmli] *adv*. 均匀地
approximately [əˈprɒksɪmətli] *adv*. 大约
relative [ˈrelətɪv] *adj*. 相对的
derive [dɪˈraɪv] *vt*. 推导
employ [ɪmˈplɔɪ] *vt*. 使用
distribute [ˈdɪstrɪbjuːt] *vi*. 分布
improvement [ɪmˈpruːvmənt] *n*. 改进
impressive [ɪmˈpresɪv] *adj*. 可观的
notion [ˈnəʊʃən] *n*. 概念
extreme [ɪkˈstriːm] *n*. 极致；极限
device [dɪˈvaɪs] *n*. 器件；设备
enhance [ɪnˈhæns] *vt*. 增强
minimize [ˈmɪnɪmaɪz] *vt*. 降至最低
principle [ˈprɪnsəpl] *n*. 原理
exploit [ɪkˈsplɔɪt] *n*. 利用
successive [səkˈsesɪv] *adj*. 依次的
approximation [əˌprɒksɪˈmeɪʃən] *n*. 逼近
potentially [pəˈtenʃəli] *adv*. 有潜在可能性地

Technical Terms

comparator [kəmˈpærətə] *n*. 比较器

resistor [rɪˈzɪstə] *n*. 电阻器
spectrum [ˈspektrəm] *n*. 频谱
resolution [rezəˈluːʃn] *n*. 分辨率
bandwidth [ˈbændˌwɪdθ] *n*. 带宽
baseband [ˈbeɪsˌbænd] *n*. 基带
filter [ˈfɪltə] *n*. 滤波器
cellular [ˈseljʊlə] *n*. 使用蜂窝技术的
channel [ˈtʃænəl] *n*. 信道
audio [ˈɔːdiəʊ] *adj*. 音频的
decimation [desɪˈmeɪʃn] *n*. 抽取
circuitry [ˈsɜːkɪtri] *n*. 电路
accuracy [ˈækjʊrəsi] *n*. 精度
interpolation [ɪntɜːpəˈleɪʃn] *n*. 插值
sampling rate　采样率
reference voltage　参考电压
power consumption　功耗;能耗
sine wave　正弦波
quantization error (noise)　量化误差(噪声)
full scale　满幅度;满量程
converter resolution　转换器分辨率
processing gain　处理增益
digital filtering　数字滤波
data converter　数据转换器
digital cellular phone　数字蜂窝电话
analog interfacing　模拟接口(技术)
anti-aliasing filter　抗混叠滤波器
block diagram　方框图
noise shaping　噪声整形
performance specification　性能指标
data sheet　数据手册
ADC *abbr*. analog-to-digital converter　模数转换器
FFT *abbr*. fast Fourier transform　快速傅里叶变换
SNR *abbr*. signal-to-noise ratio　信噪比
GSM *abbr*. Global System for Mobile communications　全球移动通信系统
MSPS *abbr*. million samples per second　每秒百万样本数

Notes

1. 短语... come in ... 意为“……以……形式存在着”。例如,“如今移动电话的形状和大小

多种多样”(Mobile phones *come in* all sorts of shape and size these days),“这种 T 恤有黑色的吗?”(Does this T-shirt *come in* black?)。此结构可用于表达分类,其功能与 ... be classified into ... / ... fall into ...相同,如“无线电话可分为无绳电话、移动电话两大类”(Wireless telephones *come in* two basic varieties: cordless phones and mobile phones. / Wireless telephones *fall into* two basic categories: cordless phones and mobile phones. / Wireless telephones *can be classified into* two types: cordless phones and mobile phones.)。

2. flash 可指“闪光灯”“闪存”等。此处是指“闪式模数转换器”。
3. 此处 bank 是指几个排成行或列的相似器物,a bank of ... 可译为“一行/列/排/组……”,如“一组开关”(*a bank of* switches)、“滤波器组”(filter *bank*)。
4. 注意区分 correspond 和 respond,前者是指“相应”,后者是指“响应”。
5. *Y arise from X* / *Y arise out of X* 意为“*Y* 是由 *X* 引起的”,即“*X* 导致 *Y*”(*X result in Y*)。
6. formula 可以表示“公式”“配方”“处方”“方程式”等意思,其复数形式为 formulae。具有特殊复数形式的常用名词还有“数据”(datum →data)、“介质”(medium →media)、“现象”(phenomenon →phenomena)和“标准”(criterion →criteria)等。
7. 此处 scale 意为“幅度”。“满幅正弦波输入”(full-scale sine wave input)是指输入正弦信号振幅等于 ADC 允许的输入信号最大幅值。scale 还用作“比例”“规模”“坐标”,如“1∶50000的比例”(a *scale* of 1∶50000)、“大规模的”(on a large *scale*)、“对数坐标”(a logarithmic *scale*)。
8. by a factor of ...用以表达数量增减的倍数。例如,“核反应堆附近地区的白血病数量上升了 4 倍”(Cases of leukaemia in the area near the nuclear reactor have risen *by a factor of* four.)。
9. 此处 interest 是指事物“值得关注的品质或特征”,可译作“价值”“意义”“趣味”“吸引力”。结构 of interest (to *sb.*) 可作表语或后置定语,如“具有科学价值”(*of* scientific *interest*)、“具有历史意义”(*of* historical *interest*)、“毫无趣味”(*of* no *interest*)等。

Exercises

1. Fill in the blanks with proper words, phrases or clauses.

(1) What is a filter? A filter is a device that passes electric signals _____ certain frequencies or frequency ranges _____ preventing the passage of others.

Filter circuits are used _____ a wide variety of applications. _____(在电信领域), band-pass filters are used in the audio frequency range (0 kHz to 20 kHz) for _____(调制解调器和语音处理). High-frequency band-pass filters (several hundred MHz) are used for _____(信道选择)in telephone central offices. Data acquisition systems usually require _____(抗混叠低通滤波器)as well as low-pass noise filters in their preceding _____(信号调理)stages. System power supplies often use band-rejection filters to suppress the 60-Hz line frequency and high frequency transients. In addition, there are filters that do not _____

(滤除)any frequencies of a complex input signal, but just add _____(线性相移)to each frequency component, thus contributing _____ a constant _____(时延). These are called _____(全通滤波器).

At high frequencies (> 1 MHz), all of these filters usually consist _____ passive components _____ inductors (L), resistors (R), and capacitors (C). They are then _____ LRC filters. In the lower frequency range (1 Hz to 1 MHz), however, the inductor value becomes very large and the inductor itself gets quite bulky, _____ economical production difficult. In these _____, _____(有源滤波器)become important. Active filters are circuits that use _____(运算放大器)as the active device _____ combination with some _____(电阻和电容)to provide an LRC-like filter performance at low frequencies.

(2) Choosing the right _____(无源器件)for an analog design is important. In most cases, a right passive component will fit on the same pads _____ a wrong passive component, but not always. Start the design process _____ carefully considering the high frequency characteristics of passive components, and putting the correct part outline on the board _____ the start.

Most designers are totally _____(不懂)the frequency limitations of the passive components they use in _____(模拟电路). Passive components have _____(有限的频率范围), and operation of the part outside of that range can have some very unexpected results. One might think that this discussion only applies _____ high-speed analog circuits. But high frequencies that are _____(辐射或者感应)into a low-speed circuit will affect passive components _____. For example, a simple op amp low-pass filter may well turn into _____(高通滤波器) at RF frequencies.

2. Translate the following passages into Chinese or English.

(1) It is extremely hard to define dynamic range (DR) for an op amp, so lets start with a digital-to-analog converter (DAC) where DR is defined as the ratio of the maximum output voltage to the smallest output voltage the DAC can produce.

(2) The same definition of DR can be used for an op amp, and the maximum output voltage swing equals V_{OUTMAX}. This output voltage swing is defined as the maximum output voltage the op amp can achieve (V_{OH}) minus the minimum output voltage the op amp can achieve (V_{OL}). V_{OH} and V_{OL} are easily obtainable from an op amp IC data sheet.

(3) Noise fluctuates randomly over a period of time, so instantaneous signal or noise levels don't describe the situation adequately. Averages over a long period of time (root mean squared or RMS) are used to describe both the signal and the noise. Signal-to-noise ratio (SNR) was initially established as a measure of the quality of the signal that exists in the presence of noise. This SNR was a power ratio, and it was established at the output of a circuit. The SNR that we are interested in is a voltage ratio because the impedance is constant, and it is established at the input to the op amp. This means that all noise voltages, including resistor noise voltage, must be calculated in RMS volts at the op amp input.

(4) The Chebyshev equal ripple filter distributes the roll-off across the whole passband. It introduces more ripples in the passband but provides a sharper roll-off in the transition region. This type of filter has poorer transient and step responses due to its higher Q values in the stages of the filter.

(5) Chip package is the housing that chips come in for plugging into (socket mount) or soldering onto (surface mount) the printed circuit board. Creating a mounting for a chip might seem trivial to the uninitiated, but chip packaging is a huge and complicated industry. The ability to provide more and more I/O interconnections to a die (bare chip) that is increasingly shrinking in size is an ever-present problem. In addition, the smaller size of the package contributes as much to the miniaturization of cell phones and other handheld devices as the shrinking of the semiconductor circuits.

(6) 电压和水有个共同点：仅当供应不足时，其价值才会被认识到。低电压系统(低于5 V的单电源供电系统)让我们认识到电压的价值。

(7) 欧姆定律为$U=I\cdot R$，它是电子学基本定律。欧姆定律可用于单个器件、一组器件或者一个完整电路。当已知流经线路各部分的电流时，其电压降就是电阻和电流的乘积。

(8) 电路就是无源器件和有源器件的组合。器件按照特定方式排列，从而完成期望的功能。这种器件排列就叫电路(或电路结构)。设计模拟电路就是设计电路结构。

(9) 噪声决定了系统处理数据和信号的范围。放大器、接收器等设备的信号辨别能力被噪声降低了。噪声混入输入信号中，运放噪声、电阻噪声和电源噪声最终决定了可重建与可测量信号的大小。

(10) 巴特沃斯滤波器(也叫“最大平坦”滤波器)是最常见的通用滤波器。巴特沃斯滤波器有一个单调变化的通带，通带衰减抵达其 3 dB(即人们熟知的特征频率点)。不论是几阶滤波器，该频率都是一样的。然而，通过增加滤波器阶数，通带滚降会更靠近其特征频率，在特征频率和阻带间的过渡区滚降也变得更陡。

3. Writing tasks.

Summary: Write an abstract for *Lesson 4 Operational Amplifiers* within 150～200 words.

Composition: Write an essay about a typical circuit you have built or used in college labs.

Reading Materials

Passage 1 Anti-aliasing Filtering

Generally, every circuit in which an analog signal converts to the digital domain should have an analog low-pass filter. This situation is true whether the analog-to-digital converter (ADC) is a successive-approximation-register (SAR), sigma-delta, pipeline, dualslope, or any other type of converter you may contrive. The placement of this type of filter in the circuit must always be on the analog side, in front of the converter.

Why do you need the analog low-pass filter? Remember that every analog signal has high- and low-frequency noise, whether or not you acknowledge it. The reason you need a low-pass filter goes back to the Nyquist theorem, which illustrates that to accurately convert a signal without contamination, you must first eliminate out-of-band frequencies.

Any signal that passes through the ADC has a magnitude associated with it. The ADC usually faithfully converts the magnitude of that signal as long as the signal frequency is below the converter's input-stage bandwidth. Although the ADC preserves the magnitude, the same is not true for the signal's frequencies. The frequencies above half of the ADC's sampling frequency contaminate the conversion to the point that you won't be able to tell the difference between in-band and out-of-band signals at the converter's output. This phenomenon is known as signal aliasing (see Figure 2.12). After conversion, the ADC preserves the magnitude of the input signal but aliases the frequencies above half of the sampling frequency, f_s, back below $f_s/2$.

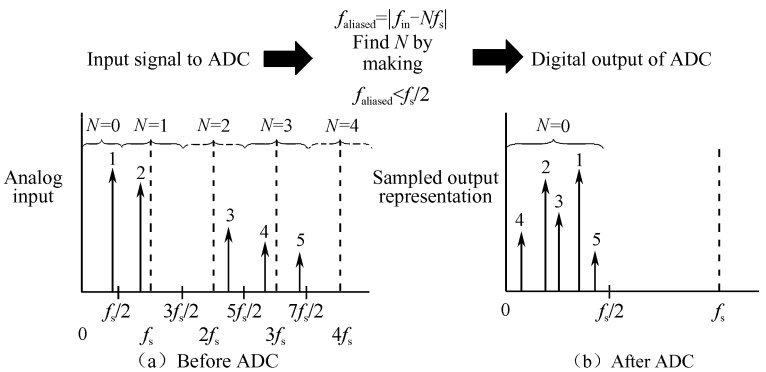


Figure 2.12 The spectrum of signal

Anti-aliasing filters can be built using one of three approaches. The first is to design a passive filter using resistors and capacitors. A second alternative is to design an active filter by using an operational amplifier, a few resistors and capacitors. A third option is to use a switched-capacitor filter.

Questions:

1. What does the acronym SAR stand for?

2. What does Nyquist theorem state?
3. What symbol is used to represent *sampling frequency* in this article?
4. What is *signal aliasing*?
5. How is a typical active filter constructed?

Passage 2 Switched-capacitor Filters

The characteristics of all active filters, regardless of architecture, depend on the accuracy of their RC time constants. Because the typical precision achieved for integrated resistors and capacitors is approximately±30%, a designer is handicapped when attempting to use absolute values for the components in an integrated filter circuit. The ratio of capacitor values on a chip can be accurately controlled, however, to about one part in 2000. Switched-capacitor filters use these capacitor ratios to achieve precision without the need for precise external components.

In the switched-capacitor integrator shown in Figure 2.13, the combination of C_1 and the switch simulates a resistor.

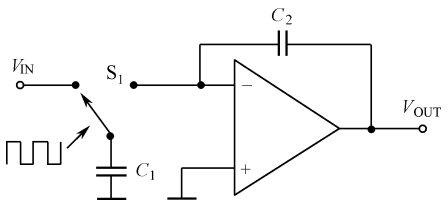


Figure 2.13 A switched-capacitor integrator

The switch S_1 toggles continuously at a clock frequency f_{CLK}. Capacitor C_1 charges to V_{IN} when S_1 is to the left. When it switches to the right, C_1 dumps charge into the integrator's summing node, from which it flows into the capacitor C_2. The charge on C_1 during each clock cycle is $Q = C_1 \quad V_{IN}$, and thus the average current transferred to the summing junction is

$$I = Q \quad f_c = C_1 \quad V_{IN} \quad f_c$$

Notice that the current is proportional to V_{IN}, so we have the same effect as a resistor of value

$$R = \frac{V_{IN}}{I} = \frac{1}{C_1 \quad f_c}$$

The integrator's ω_0 is therefore

$$\omega_0 = \frac{1}{R \quad C_2} = \frac{C_1 \quad f_c}{C_2}$$

Because ω_0 is proportional to the ratio of the two capacitors, its value can be controlled with great accuracy. Moreover, the value is proportional to the clock frequency, so you can vary the filter characteristics by changing f_{CLK}, if desired.

Questions:

1. What's the meaning of *handicapped* in the first paragraph?

2. What does the verb *toggle* mean in the third paragraph?
3. What symbol is used to represent clock frequency in this article?
4. What advanges does the switched-capacitor filter have over any active filter?
5. What does the term *time constant* mean in this article?

Passage 3 Digital-to-Analog Converters

Figure 2.14 shows the components that make up the DAC (digital-to-analog conversion) chain. In addition to the conversion of the digital word to a discrete voltage or current level, a zero-order hold is used to "hold" the signal level until the next update. To smooth out the recovered waveform, reconstruction filtering is needed.

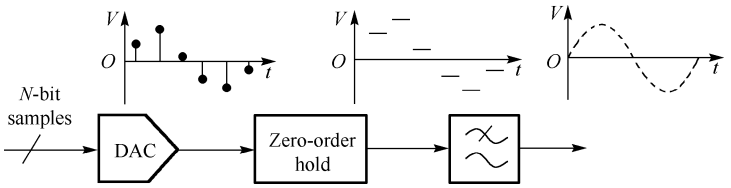


Figure 2.14 The DAC chain

As we know, over-sampling on the output (providing samples more frequently than is strictly necessary to satisfy the Shannon/Nyquist 2×bandwidth rule) can greatly reduce the specification of the reconstruction filter. We shall consider just two of the many varieties of DAC architectures. The zero-order hold process introduces a small error into the frequency response of the output of the DAC, giving it a $\sin x/x$ or sinc weighting. This can be overcome by implementing an "inverse sinc" digital compensation filter prior to the DAC chain. Some DAC devices have an in-built compensation filter, so check the data sheet carefully.

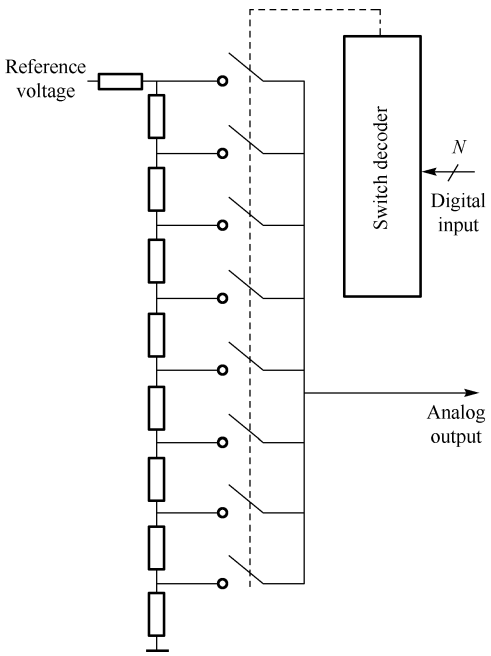


Figure 2.15 Basic resistor divider DAC

Figure 2.15 shows the components of the most straightforward DAC, which essentially operates as the reverse of the flash ADC. A known reference voltage is applied across a chain of $2^N - 1$ resistors such that the voltage measured at any of the tap points exactly corresponds to one of the 2^N available output sample values for an N-bit converter. By switching in the tap point that corresponds to the value of the digital word to be converted, a basic DAC operation is achieved.

The over-sampling approach with a 1-bit DAC is just as effective for output conversion as the 1-bit ADC with over-sampling and noise shaping is for input conversion. A simple sigma-delta DAC is shown in Figure 2.16.

The first element in the sigma-delta DAC is the interpolation process, which inserts zero samples between each valid sample to realize the increase in data rate, together with sigma-delta DAC digital filtering to perform part of the signal reconstruction. The digital sigma-delta modulator performs a shaping function on the quantization noise, such that in the final reconstructed output, most of the quantization noise is pushed out of the band of interest. The 1-bit DAC is basically a switch, selecting typically a 0 or positive voltage reference level (single supply operation), operating at the very high sampling rate. The analog filter smoothes out the transitions on the output to yield a continuous and high-fidelity output waveform.

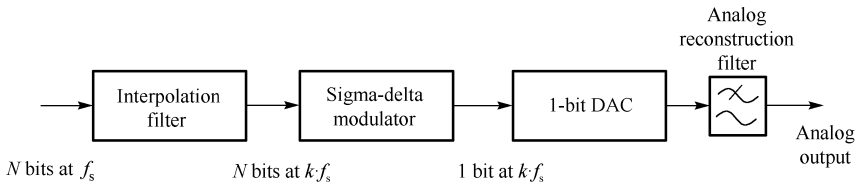


Figure 2.16 Basic sigma-delta DAC

Table 2.1 lists the major specifications of some typical DACs. Audio DACs are used in consumer hi-fi products, while high-speed DACs are used in the fields of video and wireless communications. In Table 2.1, we can see the tradeoffs between sampling rate and resolution, resolution and power.

Table 2.1 Typical DAC specifications

Sampling rate	Resolution	Dynamic range	Total harmonic distortion	Power
Audio DAC				
192 kSPS	16	95 dB	90 dB	40 mW
192 kSPS	24	120 dB	100 dB	70 mW
High-speed DAC				
500 MSPS	10 bit	80 dB		250 mW
300 MSPS	12 bit	85 dB		300 mW
300 MSPS	14 bit	88 dB		350 mW

Questions:

1. What components should a DAC chain have?
2. What does Shannon/Nyquist rule state?
3. What's the relation between resistor divider DACs and flash ADCs?
4. What function does the interpolation filter perform in the sigma-delta DAC?
5. How is a basic sigma-delta DAC constructed?

Unit 3

Electronic System Components

Lesson 7 Switching Power Supply

Lesson 8 Clock Sources

Lesson 9 Interconnects

Passage 1 Cables

Passage 2 Bus Basics

Passage 3 USB (Universal Serial Bus)

Lesson 7 Switching Power Supply

Every new electronic product, except those that are battery-powered, requires converting off-line 115 V_{ac} or 230 V_{ac} power to some DC voltage for powering the electronics. Efficient conversion of electrical power is becoming a primary concern to companies and to society as a whole. Switching power supplies[1] offer not only higher efficiencies but also offer greater flexibility to the designer. Recent advances in semiconductor, magnetic and passive technologies make the switching power supply an ever more popular choice in the power conversion arena today.

Linear versus Switching Power Supplies

Historically, the linear regulator was the primary method of creating a regulated output voltage. It operates by reducing a higher input voltage down to the lower output voltage by linearly controlling the conductivity of a series[2] pass power device in response to changes in its load. This results in a large voltage being placed across the pass unit with the load current flowing through it.

This headroom loss ($V_{drop} \cdot I_{load}$) causes the linear regulator to only be 30 to 50 percent efficient. That means that for each watt delivered to the load, at least a watt has to be dissipated in heat. The cost of the heatsink actually makes the linear regulator uneconomical[3] above 10 watts for small applications. Below that point, however, they are cost-effective in step-down applications.

The switching regulator operates the power devices in the full-on and cutoff states. This then results in either large currents being passed through the power devices with a low "on" voltage or no current flowing with high voltage across the device. This results in a much lower power being dissipated within the supply. The average switching power supply exhibits efficiencies of between 70 to 90 percent, regardless of the input voltage.

Higher levels of integration have driven the cost of switching power supplies downward which makes it an attractive choice for output powers greater than 10 watts or where multiple outputs are desired.

Forward-mode[4] Converters

The most elementary[5] forward-mode converter is the buck or step-down converter which can be seen in Figure 3.1. Its operation can be seen as having two distinct[6] time periods which occur when the series power switch is on and off. When the power switch is on, the input

voltage is connected to the input of the inductor. The output of the inductor is the output voltage, and the rectifier is back-biased. During this period, since there is a constant voltage source connected across the inductor, the inductor current begins to linearly ramp upward which is described by:

$$i_{L(on)} = \frac{(V_{in} - V_{out}) \cdot t_{on}}{L}$$

During the "on" period, energy is being stored within the core material of the inductor in the form of flux. There is sufficient energy stored to carry the requirements of the load during the next off period.

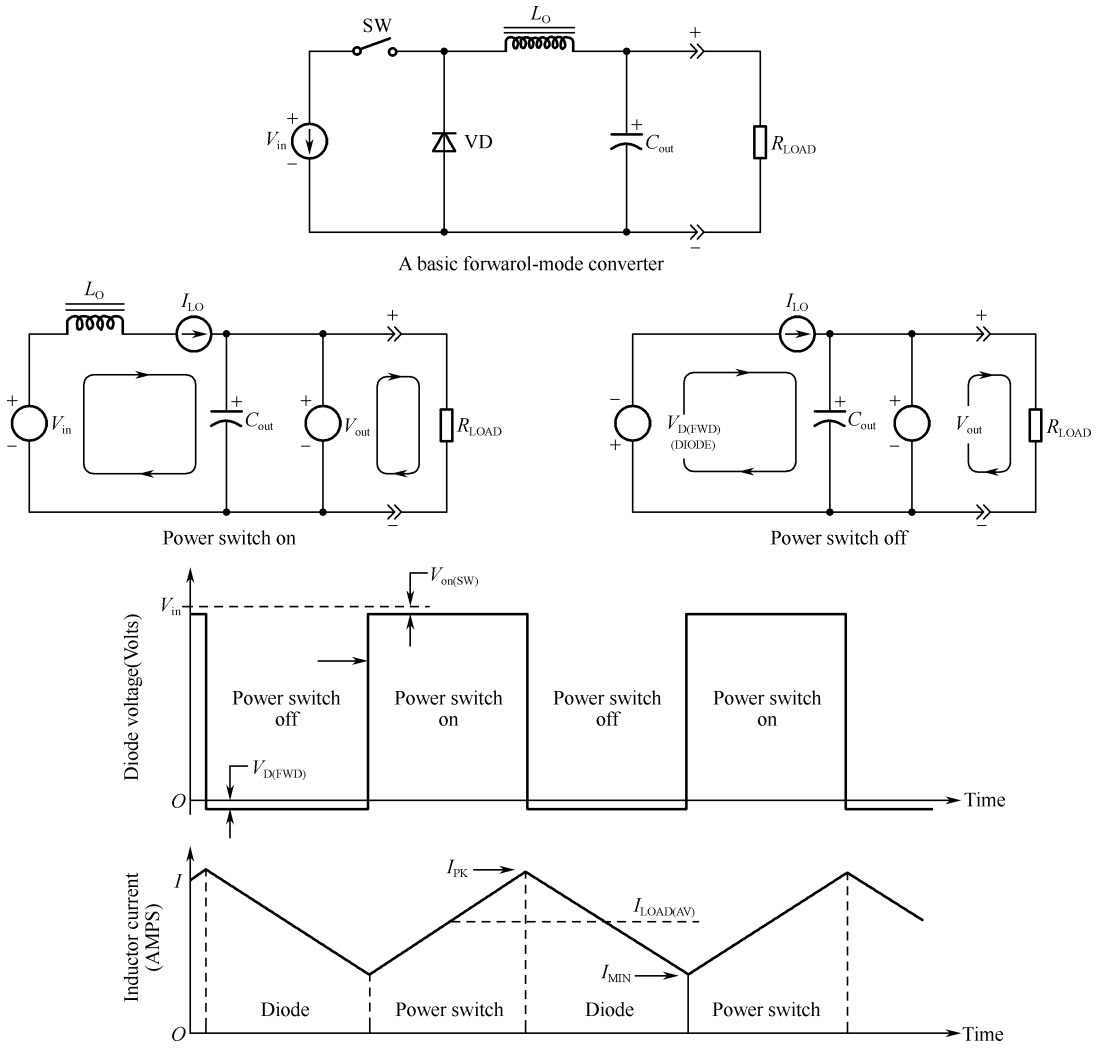


Figure 3.1 Forward-mode converter operation (buck converter shown)

The next period is the "off" period of the power switch. When the power switch turns off, the input voltage of the inductor flies below ground and is clamped at one diode drop below ground by the catch diode. Current now begins to flow through the catch diode thus

maintaining the load current loop. This removes the stored energy from the inductor. The inductor current during this time is:

$$i_{L(off)} = \frac{(V_{out} - V_D) \cdot t_{off}}{L}$$

this period ends when the power switch is once again turned on.

Regulation is accomplished by varying the on-to-off duty cycle of the power switch. The relationship which approximately describes its operation is:

$$V_{out} \approx \partial \cdot V_{in}$$

where ∂ is the duty cycle $[\partial = t_{on}/(t_{on} + t_{off})]$.

The buck converter is capable of kilowatts of output power, but suffers from[7] one serious shortcoming which would occur if the power switch were to fail short-circuited, the input power source is connected directly to the load circuitry with usually produces catastrophic results. To avoid this situation, a crowbar[8] is placed across the output. A crowbar is a latching SCR (silicon controlled rectifier) which is fired when the output is sensed as entering an overvoltage condition. The buck converter should only be used for board-level regulation.

Flyback or Boost-mode Converters

The most elementary flyback-mode converter is the boost or step-up converter. Its schematic can be seen in Figure 3.2. Its operation can also be broken into two distinct periods where the power switch is on and off. When the power switch turns on, the input voltage source is placed directly across the inductor. This causes the current to begin linearly ramping upwards from zero and is described by:

$$i_{L(on)} = \frac{V_{in} \cdot t_{on}}{L}$$

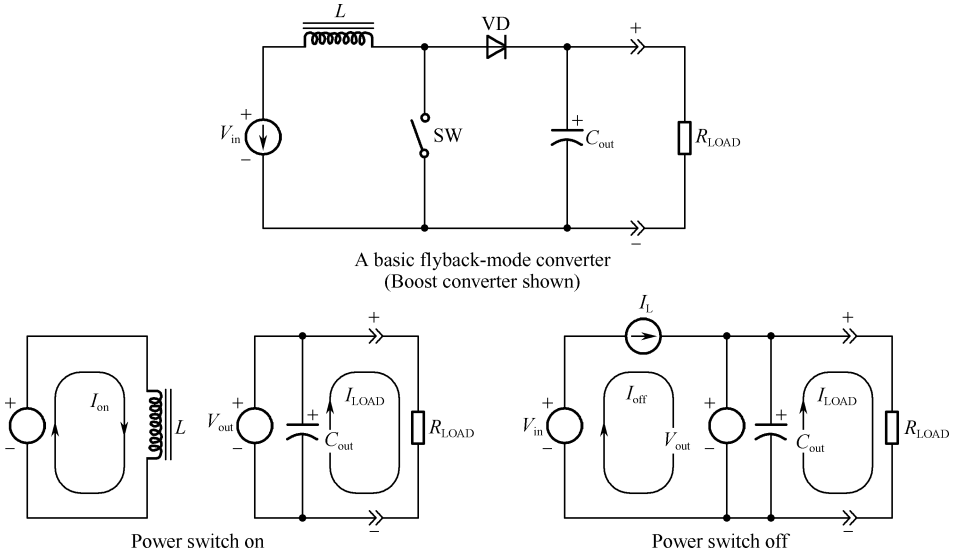


Figure 3.2 Schematic of a boost converter

Once again, energy is being stored within the core material.

The amount of energy stored during each cycle times[9] the frequency of operation must be higher than the power demands of the load or,

$$P_{sto}=0.5L \cdot I_{pk}^2 \cdot f_{op} > P_{out}$$

The power switch then turns off and the inductor voltage flys back above the input voltage and is clamped by the rectifier at the output voltage. The current then begins to linearly ramp downward until the energy within the core is completely depleted. Its waveform which is shown in Figure 3.3 is determined by:

$$i_{L(off)}=\frac{(V_{out}-V_{in}) \cdot t_{off}}{L}$$

The boost converter should also be only used for board-level regulation.

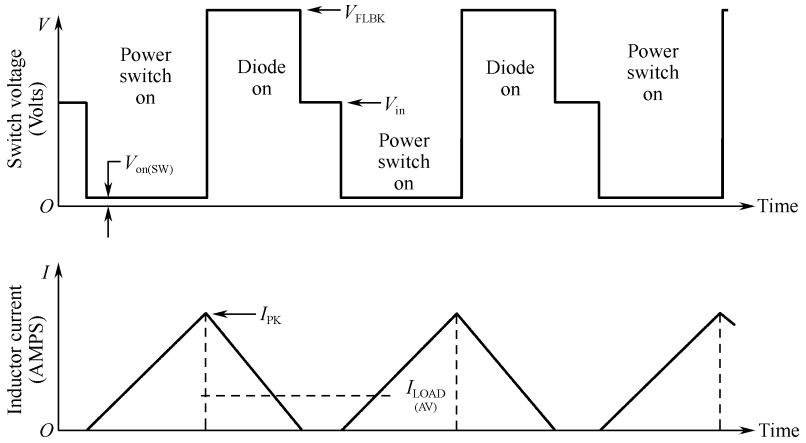


Figure 3.3 Waveforms for a boost converter

New Words & Expressions

product [ˈprɒdʌkt] *n*. 产品
efficient [ɪˈfɪʃənt] *adj*. 高效的
conversion [kənˈvɜːʃn] *n*. 转换
primary [ˈpraɪməri] *adj*. 主要的
concern [kənˈsɜːn] *n*. 关心
as a whole 全体
efficiency [ɪˈfɪʃənsi] *n*. 效率
flexibility [ˌfleksəˈbɪləti] *n*. 灵活性
advance [ədˈvæns] *n*. 进展
arena [əˈriːnə] *n*. 竞技场
linear [ˈlɪniə] *adj*. 线性的
versus [ˈvɜːsəs] *prep*. 对比
... in response to ... 应 而
downward [ˈdaʊnwəd] *adv*. 向下地
multiple [ˈmʌltɪpl] *adj*. 多个的
elementary [ˌeləˈmentəri] *adj*. 基本的
forward [ˈfɔːwəd] *adj*. 向前的
mode [məʊd] *n*. 模式
boost [buːst] *n*. & *vt*. 增大
distinct [dɪˈstɪŋkt] *adj*. 不同的
accomplish [əˈkʌmplɪʃ] *vt*. 完成
vary [ˈveəri] *vi*. 变化
relationship [rɪˈleɪʃənʃɪp] *n*. 关系
approximately [əˈprɒksɪmətli] *adv*. 大约
capable [ˈkeɪpəbl] *adj*. 有能力的
shortcoming [ˈʃɔːtˌkʌmɪŋ] *n*. 缺点

deliver [dɪˈlɪvə] *vt*. 传送
uneconomical [ˌʌnˌekəˈnɒmɪkl] *adj*. 不合算的
effective [ɪˈfektɪv] *adj*. 有效的
average [ˈævərɪdʒ] *adj*. 典型的;普通的
exhibit [ɪgˈzɪbɪt] *vt*. 表现
regardless of ... 不管……
catastrophic [ˌkætəˈstrɒfɪk] *adj*. 灾难性的
condition [kənˈdɪʃən] *n*. 状态
amount [əˈmaʊnt] *n*. 数量
demand [dɪˈmænd] *n*. 需要
deplete [dɪˈpliːt] *vt*. 耗尽
determine [dɪˈtɜːmɪn] *vt*. 确定

Technical Terms

switching [ˈswɪtʃɪŋ] *n*. 交换,切换
battery [ˈbætəri] *n*. 电池
voltage [ˈvəʊltɪdʒ] *n*. 电压;伏特数
electronics [ˌilekˈtrɒnɪks] *n*. 电子电路
semiconductor [ˌsemɪkənˈdʌktə] *n*. 半导体
magnetic [mægˈnetɪk] *adj*. 磁的
passive [ˈpæsɪv] *adj*. 无源的
regulator [ˈregjʊˌleɪtə] *n*. 稳压器
converter [kənˈvɜːtə] *n*. 整流器
conductivity [kɒndʌkˈtɪvɪti] *n*. 导电性
series [ˈsɪəriːz] *n*. 串联
load [ləʊd] *n*. 负载
headroom [ˈhedˌruːm] *n*. 净空,活动空间
loss [lɒs] *n*. 损耗
dissipate [ˈdɪsɪˌpeɪt] *vi*. 耗散
heatsink [ˈhiːtˈsɪŋk] *n*. 散热片
cutoff [ˈkʌtɒf] *n*. 截止
integration [ˌɪntəˈgreɪʃn] *n*. 集成
inductor [ɪnˈdʌktə] *n*. 电感器
rectifier [ˈrektɪfaɪə] *n*. 整流器
biased [ˈbaɪəst] *adj*. 加偏压的
flux [flʌks] *n*. 通量
diode [ˈdaɪəʊd] *n*. 二极管
crowbar [ˈkrəʊˌbɑː] *n*. 短路器
overvoltage [ˈəʊvəˈvəʊltɪdʒ] *n*. 过压
flyback [ˈflaɪbæk] *n*. 回扫
schematic [skiːˈmætɪk] *n*. 原理图
waveform [ˈweɪvˌfɔːm] *n*. 信号波形
electrical power 电能
linear regulator 线性稳压器

load current　负载电流
duty cycle　占空比
SCR *abbr*. silicon-controlled rectifier　可控硅整流器

Notes

1. switching power supply 又名 switch-mode power supply，后者缩写为 SMPS。
2. series 表示“串联的”，如“串联电路”(*series* circuit)、“串联地”(in *series*)，对应于“并联电路”(parallel circuit)、“并联地”(in parallel)。作为名词时，series 单复同形，如“系列讲座”(a lecture *series*)、“一串数字”(a *series* of numbers)、“泰勒级数”(Taylor *series*)。
3. uneconomical 意为“不合算的”(not economical)。下句中的 cost-effective 与 economical 同义，均可表示“成本低、收益高”，即“合算的”。effective 意为“收到预期效果的”“成功的”。
4. mode 属于学术词汇，可译作“模式”。例如，“模式滤波器”(*mode* filter)、“切换机器到音频模式”(Switch the machine to audio *mode*.)、“公司各部门均有其运行模式”(Each department in the company has its own *mode* of operation.)。在统计学中，mode 译作“众数”。
5. elementary 与 basic 同义。其来源名词 element 可指“化学元素”或“电路元件”。
6. distinct 可表示“明显的”或“不同的”，如“明显的下降”(a *distinct* decline)、“这两个概念截然不同”(The two concepts are quite *distinct*.)。课文中的 distinct 取第二种意思。
7. suffer from ... 此处意为“呈现弊端或反映出不良效果”，可译为“表现出……”。例如，“这部电影有点太长”(This film *suffers from* being a bit too long.)。
8. “短路器”(crowbar)是一种保护电子系统免受高电压浪涌破坏的电路。
9. times 在此处作为介词，表示“乘以”(multiplied by)，如“矩形的面积就是其高度乘以宽度”(The area of a rectangle is its height *times* its width.)。

Lesson 8　Clock Sources

There are a variety[1] of clock devices available today. Some of them are described below.

Crystals

A crystal is a basic piezoelectric quartz crystal. On its own, it cannot generate electrical clocks. It has to be connected to a clock oscillator to get a clock waveform. There are two kinds of crystals: Series Resonant, which can be modeled as a high Q series LC circuit, and Parallel Resonant, which can be modeled as a high Q parallel LC circuit. The series resonant crystal has minimum impedance at the resonating frequency, while[2] the parallel resonant crystal has maximum impedance at the resonating frequency.

Crystal Oscillators

A crystal oscillator is an oscillator with the crystal as the feedback element. There are other kinds of oscillators with active or passive feedback components, but the crystal oscillator provides the most accurate and stable output frequency. Crystal oscillators are the preferred clock source in most high-speed digital systems requiring clocks.

Compensated Oscillators

The output frequency of a crystal oscillator varies with temperature and voltage. Applications that require a highly stable clock usually use compensated oscillators. Compensated oscillators try to adjust the variation in frequency due to temperature and voltage. Temperature compensated crystal oscillators (TCXO) contain circuitry that compensates for temperature changes, and hence combat frequency variations. Oven-controlled oscillators encase their crystals in a temperature-controlled oven, and so maintain a precise operating temperature at the crystal. Double oven oscillators contain two ovens, with the crystal encased in the inner oven, and the temperature control circuitry and the inner oven encased in the outer oven. Such oscillators provide even better temperature stability than oven-controlled oscillators. Obviously, as the frequency stability improves, the cost of the oscillator increases.

Voltage-Controlled Oscillators

The output of voltage-controlled oscillators (VCO) is controlled by a voltage control input pin. Variation between control voltage and frequency is usually nonlinear over the entire frequency range but is linear within subset ranges.

Frequency Synthesizers

Frequency synthesizers use one or more phase-locked loops (PLL) to generate one to many different frequencies on their outputs, from one or more reference sources. The reference frequency is usually generated by a crystal attached to the synthesizer. The design goal of frequency synthesizers is to replace multiple oscillators in a system, and hence reduce board space and cost. Figure 3.4 shows a block diagram of a PLL.

A PLL has two inputs, a reference input and a feedback input. A PLL corrects frequency in two ways[3]. The first, frequency correction, corrects large differences in frequency between the reference input and the feedback input. Frequency correction is akin to "rough" tuning and occurs when F_{vco} is less than $0.5F_{ref}$ or greater than $2F_{ref}$. Phase correction is the "fine" tuning and occurs when $0.5F_{vco} < F_{ref} < 2F_{vco}$.

The Phase/Frequency Detector detects differences in phase and frequency between the reference and feedback inputs and generates compensating "U_p" and "D_n" signals depending on whether the feedback frequency is lagging or leading the reference frequency respectively. These control signals are then passed through a charge pump and a loop filter to generate a

control voltage, which controls a VCO. The frequency of this oscillator is dependent on the V_{ctrl} input. At steady state, the VCO frequency is:

$$F_{vco} = \frac{F_{ref} \cdot P}{Q}$$

The output frequency of the PLL can be expressed as

$$F_{out} = \frac{F_{ref} \cdot P}{Q \cdot N}$$

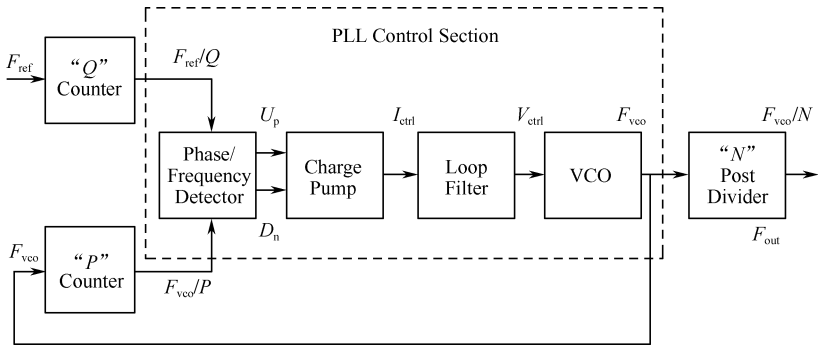


Figure 3.4 Block diagram of a phase-locked loop (PLL)

The sample rate of a frequency synthesizer determines how often the inputs are sampled in order to perform phase and frequency correction. It is expressed as F_{ref}/Q.

The acquisition/lock time of a PLL-based frequency synthsizer is the amount of time taken by the frequency synthesizer to attain the target frequency after power-up, or after a programmed output frequency change.

The resolution of a PLL-based frequency synthesizer is based on the number of bits in the P and Q counter. The resolution will determine in what size increments the frequency can change.

The deadband of a PLL-based frequency synthesizer is the largest phase difference between the reference and the feedback inputs, which will not be corrected by the PLL.

Multiple PLLs are needed within a single frequency synthesizer to generate multiple unrelated frequencies. Frequency synthesizers are gaining in popularity as system complexity increases and systems utilize multiple clocks. The term "clock generator" is interchangeably used with "frequency synthesizer."

Clock Buffers

A clock buffer is a device in which the output waveform directly follows the input waveform. The input waveform propagates through the device and is re-driven by the output buffers. Hence, such devices have a propagation delay associated with them. In addition, due to the differences between the propagation delay through the device on each input-output path, skew will exist on the outputs.

Clock Parameters

Clock Jitter Jitter can be defined as the deviations in a clock's output transitions from their ideal positions. The deviation can either be leading or lagging the ideal position. Hence, jitter is expressed in ±ns. Jitter can be classified into three categories: cycle-cycle jitter, period jitter, and long-term jitter.

Skew Skew is the variation in arrival time of two signals specified to arrive at the same time. Skew is composed of two parts, the output skew of the driving device, and board design skew, caused by layout variation of board traces.

Tolerance/Accuracy/Precision Tolerance/Accuracy/Precision is a measure of how close the part operates to the specified (nominal) frequency, typically referenced at ambient temperature (25 ℃±5 ℃). For example, if a part is specified with a 25.000-MHz output, and the long-term (user-defined) average of its output frequency is 25.001 MHz at ambient temperature, the part has +40 ppm (parts per million) accuracy. Frequency tolerance is affected or controlled by controlling the accuracy of the manufacturing and calibrating process for the crystal.

Stability Stability is a parameter usually associated with crystals and oscillators. Stability is defined as the variation in operating frequency from the ambient temperature frequency (frequency tolerance value) over the operating temperature range and is expressed in ppm. This parameter is specified with a maximum and minimum frequency deviation, expressed in percent or parts per million.

Aging Aging is defined as the systematic change in frequency over time due to internal changes in crystal/oscillator. It is usually expressed in ppm/year, and may be incorporated in the stability spec, if it is not drawn out separately. It is a parameter usually associated with crystal oscillators. New crystals age faster than old crystals.

Error On a PLL-based device, it may not always be possible to get the specified frequency on the outputs. The limitation is due to the size of the internal "*P*" and "*Q*" counters in the PLL. If, for example, the specified frequency is 25.000 MHz, and the PLL can output 24.998 MHz, the error is −80 ppm. Error can be expressed as: Error=(Nominal Freq. − Target Freq.)/Target Freq.

Slew The rate[4] of change of voltage or frequency is called slew. Slew is usually measured on the rising and falling edges of digital signals. However, rise times and fall times are more commonly specified, instead of slew, in vendor's catalogs. Recently, with the advent of[5] low-power devices, slew is being used to define a rate of change of frequency.

Duty Cycle Duty cycle is the ratio of the output high time to the total cycle time. It is expressed as a percentage. 50% is the ideal duty cycle, though most clock manufacturers specify duty cycles from 40% ~ 60%. Duty cycle is important in systems that use both the rising and falling clock edges. Duty cycles can be expressed for both TTL and CMOS devices. For TTL devices, since the voltage swing is from 0 ~ 3 V, the high time is measured at the

1.5 V level. For CMOS devices, since the voltage swing is from 0 ~ V_{dd} volts, the high time is measured at $V_{dd}/2$.

New Words & Expressions

variety [vəˈraɪəti] *n*. 品种;类型
compensate [ˈkɒmpənseɪt] *vi*. 补偿
adjust [əˈdʒʌst] *vt*. 调节
variation [ˌveərɪˈeɪʃn] *n*. 变化
hence [hens] *adv*. 因此
combat [ˈkɒmbæt] *vt*. 防止
oven [ˈʌvən] *n*. 恒温箱
encase [ɪnˈkeɪs] *vt*. 装入
inner [ˈɪnə] *adj*. 内部的
outer [ˈaʊtə] *adj*. 外围的
subset [ˈsʌbˌset] *n*. 子集
reference [ˈrefərəns] *n*. & *vt*. 参考
attach [əˈtætʃ] *vt*. 连接
correction [kəˈrekʃn] *n*. 校正
difference [ˈdɪfərəns] *n*. 差别
akin[əˈkɪn] *adj*. 同类的,相似的
rough [rʌf] *adj*. 粗略的
fine [faɪn] *adj*. 精细的
respectively [rɪˈspektɪvli] *adv*. 分别地
acquisition [ˌækwɪˈzɪʃn] *n*. 采集
attain [əˈteɪn] *vt*. 达到
increment [ˈɪŋkrəmənt] *n*. 增量
gain in 提高
popularity [ˌpɒpjʊˈlærəti] *n*.普及,流行
complexity [kəmˈpleksəti] *n*. 复杂度
utilize [ˈjuːtɪlaɪz] *vt*. 使用
parameter [pəˈræmɪtə] *n*. 参数
category [ˈkætəgəri] *n*. 类别
tolerance [ˈtɒlərəns] *n*. 公差
accuracy [ˈækjʊrəsi] *n*. 准确度
precision [prɪˈsɪʒən] *n*. 精密度
ambient [ˈæmbiənt] *adj*. 环境的
calibrate [ˈkæləˌbreɪt] *vt*. 标定
systematic [ˌsɪstəˈmætɪk] *adj*. 系统的
incorporate [ɪnˈkɒpəˌreɪt] *vt*. 并入
percentage [pəˈsentɪdʒ] *n*. 百分比

Technical Terms

crystal [ˈkrɪstl] *n*. 晶体
piezoelectric [paɪzəʊɪˈlektrɪk] *adj*. 压电的
quartz [kwɒts] *n*. 石英
oscillator [ˈɒsɪˌleɪtə] *n*. 振荡器
series [ˈsɪəriːz] *n*. 串联
parallel [ˈpærəlel] *adj*. 并联的
resonant [ˈrezənənt] *adj*. 谐振的
synthesizer [ˈsɪnθəsaɪzə] *n*. 合成器
tuning [ˈtjuːnɪŋ] *n*. 调谐
detector [dɪˈtektə] *n*. 检波器
lagging [ˈlæɡɪŋ] *adj*. 滞后的
leading [ˈliːdɪŋ] *adj*. 超前的
propagate [ˈprɒpəɡeɪt] *vi*. 传播

skew [skjuː] *n*. 相位偏移
jitters [ˈdʒɪtəz] *n*. 时钟抖动
deviation [ˌdiːviˈeɪʃn] *n*. 偏差
nominal [ˈnɒmɪnəl] *adj*. 标称的
slew [sluː] *n*. 摆率
volt [vəʊlt] *n*. 伏特
series resonant　串联谐振
parallel resonant　并联谐振
resonating frequency　谐振频率
frequency synthesizer　频率合成器
charge pump　电荷泵
loop filter　环路滤波器
steady state　稳态
clock generator　时钟发生器
propagation delay　传输延迟
duty cycle　占空比
voltage swing　电压摆幅
TCXO *abbr*. temperature compensated crystal oscillator　温度补偿晶体振荡器
VCO *abbr*. voltage-controlled oscillator　压控振荡器
PLL *abbr*. phase-locked loop　锁相环
ppm *abbr*. parts per million　百万分之一
spec [spek] *abbr*. specification　性能指标

Notes

1. variety 常用搭配有 a wide/large/infinite variety；a whole variety of *sth*.；offer/provide a variety of *sth*.。例如，“该设备可用于多种教学领域”(The equipment could be used for *a variety of* educational purposes)。
2. 连词 while 既可表示“让步”，也可表示“转折”。课文中用作转折。
3. 当 way 表示“方式”“方法”时，常用结构有：(1)in … way；(2)way to do *sth*.；(3)way of doing *sth*.。例如，“解决此问题的方法有很多种”(There are many *ways of* solving the problem.)。注意：(1)当表示“某人做某事的方法”时，要使用 one's way of doing *sth*.(而不用 one's way to do *sth*.)；(2)要使用 means of *sth*.或 method of *sth*.，而不用 way of *sth*.，如“通信方式”(*means of* communication)、“交通方式”(*means of* transport)。
4. 与 rate 易混的有 ratio 和 ratings。ratio 表示“比率”(二者数量比)，如“性价比”(price-quality *ratio*)、“信噪比”(signal-to-noise *ratio*)、“男女比为 1∶3”(The *ratio* of boys to girls is 1 to 3)等。rate 表示“速率”(发生的快慢)，如“心率”(heart *rate*)、“比特率”(bit *rate*)、“以　　速率”(at a *rate* of …)等；rate 在经济领域表示“百分比率”，如“税率”(tax *rate*)、“利率”(interest *rate*)、“汇率”(exchange *rate*)等。ratings 表示“额定值”“等

级”“规格”，是基于评估得到的质量指标。

5. 短语 the advent of ... 意为“……的发生/出现/到来”，如“移动电话的出现改变了人们的生活”(Life has been transformed by *the advent of* the mobile phone)。

Lesson 9 Interconnects

Interconnect devices are used for interconnecting components, circuit boards, and modules. The needs for these devices must be considered during the design phase[1]. Interconnect devices must meet certain performance requirements in most or all of the following categories:

(1) Signal integrity.

(2) Power loss.

(3) Electrical characteristics: contact resistance, inductance and capacitance; voltage and power ratings; shielding; filtering.

(4) Mechanical characteristics: contact count, contact spacing, contact forces; shock and vibration; size; soldering technique.

(5) Environmental issues: cleaning materials; thermal; corrosion protection.

Signal integrity is defined by loss of quality of the input signal. Ideally, the signal at the output of an interconnect should be equal in all characteristics with the signal at the input of the interconnect. In reality, signal degradation occurs. The user must define what level of degradation is acceptable or, alternatively, must define minimum acceptable signal levels at the output. Tests that indicate signal integrity include:

(1) Voltage standing wave ratio (VSWR)[2].

(2) Frequency response.

(3) Rise time of signal edges.

(4) Current flow.

High-speed Connector Problems

In today s designs, with clock rates over 100 MHz and rise times commonly 1 ns or less, designers cannot ignore the role interconnections play in a logic design. Interconnect effects can play a significant part in the timing and noise characteristics of a circuit. The faster clock rates and rise times increase both capacitive and inductive coupling effects, which makes crosstalk problems greater. They also mean shorter time for reflections to decay before the data is clocked and read, which decreases the maximum line length that can be used for unterminated systems. This all means that one of the major interconnect challenges is to ensure signal integrity as the high-speed pulses move along the total interconnect path, from device to PCB, through the PCB to the backplane, and on out to any network connections

which may be present.

An interconnect that must pass the short rise time of a high-speed signal pulse can be detrimental to maintaining signal integrity due to an unwanted reflection. The shorter the rise time, the greater the risk of degradation of the signal. The ideal situation is that the connector will have the appropriate termination characteristics to create no degradation of the signal.

Crosstalk

In addition to general signal degradation, another potential impact[3] of a connector on signal integrity is increased crosstalk, which is caused by inductive and capacitive coupling between contacts and/or lines feeding contacts. This can cause noise to be induced in one contact/line from a switching signal in another contact/line. Crosstalk can be expressed in a percentage of the voltage swing of the signal, or in decibels (dB), where 0 dB represents 100% crosstalk, −20 dB represents 10% crosstalk, and −40 dB represents 1% crosstalk.

Two types of crosstalk can be generated: backward crosstalk and forward crosstalk. Backward, also called near-end crosstalk, is measured at the driving end of the connector and represents the sum of the capacitive and inductive coupling. Forward, also called far-end crosstalk, is measured at the receiving end of the connector and represents capacitive minus inductive coupling.

Crosstalk can also be single-line or multiline. Single-line crosstalk is the result of one driven line inducing a voltage on one quiet line. Multiline crosstalk is two or more driven lines inducing voltage(s) on one quiet line. In most systems, many lines/contacts will be in use at once, so multiline crosstalk is a more realistic figure[4].

To completely understand the crosstalk specifications of a connector, a user needs the following information:

(1) Is the crosstalk specification measured on the actual connector, or is it the result of a simulation?

(2) What is the value of the signal rise time used for the crosstalk measurement, and the digital technology used to generate it (TTL and CMOS rise times are typically 10% to 90% rise times, while ECL is typically a 20% to 80% rise time)?

(3) What is the signal pattern on the connector, i.e., the location of the ground lines relative to the measured lines?

(4) How many lines are being driven by the switched signal?

(5) What are the source and load impedances?

(6) Is the crosstalk specification for forward or backward, or the total?

(7) Is the crosstalk specification for single-line or multiline?

(8) If multiline, how many lines are being driven?

One common way to minimize[5] crosstalk is to intersperse signal contacts with ground contacts. However, this has the disadvantage[6] of minimizing signal contact density.

Levels of Interconnects

Granitz has defined six levels of interconnects:

(1) Levels 1. Chip pad to package leads (e.g., wire bonds).

(2) Levels 2. Components to circuit board (e.g., PLCC socket).

(3) Levels 3. Circuit board connector (e.g., edge-card for mother board to daughter board[7] connection).

(4) Levels 4. Subassembly interconnects (e.g., ribbon cables).

(5) Levels 5. Subassembly to I/O (e.g., BNC[8] connector).

(6) Levels 6. System interconnects (e.g., Ethernet connectors).

Each level of interconnection may be accomplished in several ways. For example, a level 4 interconnect between circuit boards may use a card edge connector, a ribbon cable, or even discrete wires.

New Words & Expressions

device [dɪˈvaɪs] *n*. 装置
component [kəmˈpəʊnənt] *n*. 部件
module [ˈmɒdjuːl] *n*. 模块
performance [pəˈfɔːməns] *n*. 性能
category [ˈkætəgəri] *n*. 种类
integrity [ɪnˈtegrəti] *n*. 完整
characteristic [ˌkærɪktəˈrɪstɪk] *n*. 特性
contact [ˈkɒntækt] *n*. 接触
vibration [vaɪˈbreɪʃn] *n*. 振动
technique [tekˈniːk] *n*. 技术
thermal [ˈθɜːməl] *adj*. 热的
corrosion [kəˈrəʊʒən] *n*. 腐蚀
degradation [ˌdegrəˈdeɪʃn] *n*. 恶化
acceptable [əkˈseptəbəl] *adj*. 可容许的
alternatively [ɔːlˈtɜːnətɪvli] *adv*. 或者
ignore [ɪgˈnɔː] *vt*. 忽视
significant [sɪgˈnɪfɪkənt] *adj*. 重要的
reflection [rɪˈflekʃn] *n*. 反射
decay [dɪˈkeɪ] *vi*. 衰减
major [ˈmeɪdʒə] *adj*. 主要的
detrimental [ˌdetrɪˈmentl] *adj*. 有害的
appropriate [əˈprəʊpriˌeɪt] *adj*. 合适的
potential [pəˈtenʃl] *adj*. 潜在的
impact [ˈɪmpækt] *n*. 巨大影响
represent [ˌreprɪˈzent] *v*. 相当于
minus [ˈmaɪnəs] *prep*. 减去
realistic [ˌrɪəˈlɪstɪk] *adj*. 实际的
figure [ˈfɪgə] *n*. 数字
specification [ˌspesɪfɪˈkeɪʃn] *n*. 规格
measurement [ˈmeʒəmənt] *n*. 测量
minimize [ˈmɪnɪˌmaɪz] *vt*. 最小化
intersperse [ˌɪntəˈspɜːs] *vt*. 散置
disadvantage [ˌdɪsədˈvæntɪdʒ] *n*. 弊端
density [ˈdensəti] *n*. 密度
accomplish [əˈkʌmplɪʃ] *vt*. 实现
discrete [dɪˈskriːt] *adj*. 分立的

Technical Terms

interconnect [ˌɪntəkəˈnekt] *n*. 互连
shielding [ˈʃiːldɪŋ] *n*.屏蔽

filtering [ˈfɪltərɪŋ] *n*. 滤波
soldering [ˈsɒldərɪŋ] *n*. 焊接
timing [ˈtaɪmɪŋ] *n*. 时序
capacitive [kəˈpæsɪtɪv] *adj*. 容性的
inductive [ɪnˈdʌktɪv] *adj*. 感性的
coupling [ˈkʌplɪŋ] *n*. 耦合
crosstalk [ˈkrɒsˌtɔːk] *n*. 串扰
backplane [ˈbækˌpleɪn] *n*. 背板;底板
termination [ˌtɜːmɪˈneɪʃn] *n*. 端接
simulation [ˌsɪmjʊˈleɪʃn] *n*. 模拟
pad [pæd] *n*. 焊盘
lead [liːd] *n*. 引线
bond [bɒnd] *n*. 接头
socket [ˈsɒkɪt] *n*. 插座
subassembly [ˌsʌbəˈsembli] *n*. 部件
ethernet [ˈiːθənet] *n*. 以太网
signal integrity　信号完整性
power loss　功率损耗
frequency response　频率响应
clock rate　时钟速率
termination characteristics　端接特性
source and load impedances　源阻抗和负载阻抗
ribbon cable　带状电缆;扁平柔性电缆
VSWR *abbr*. voltage standing wave ratio　电压驻波比
PCB *abbr*. printed circuit board　印制电路板
TTL *abbr*. transistor-transistor logic　晶体管-晶体管逻辑
CMOS *abbr*. complementary metal oxide semiconductor　互补金属氧化物半导体
ECL *abbr*. emitter-coupled logic　射极耦合逻辑
PLCC *abbr*. plastic leadless chip carrier　塑料无引线芯片承载封装
BNC *abbr*. bayonet Neill-Concelman　BNC 连接器

Notes

1. phase 属于学术词汇,此处含义为“阶段”。例如,“设计阶段”(design *phase*)、“这种新药尚处于试验阶段。”(The new drug is still in the experimental *phase*.)。在电子通信专业文献中,phase 译作“相位”,in phase 译作“同相”,out of phase 译作“异相”。
2. “电压驻波比”(VSWR)用于度量传输线与负载间的阻抗失配,其值越大,失配越严重。
3. impact of *X* on *Y* 意为“*X* 对 *Y* 的影响”,如课文中的“连接器对信号完整性的潜在影响”(potential *impact of* a connector *on* signal integrity)。
4. figure 可表示“数字”,如“目前职位空缺为 2 位数”(The job vacancies are now in double

figures.)、“这些数字表示为总数的百分比”(These *figures* are expressed as a percentage of total)。figure 可表示“图形”(略作 fig.),如“见图 8 和图 9”(Please see *figures* 8 and 9.)。

5. minimize 属于学术词汇,意为“最小化”“减至最低”。例如,“尽量减小石油泄漏造成的影响”(*minimize* the impact of the oil spill)、“我们一定要将感染的风险降至最低”(We must *minimize* the risk of infection.)。minimize 相应名词为 minimum,相应形容词为 minimal。
6. disadvantage 是指事物的弊端、劣势或缺陷。例如,“我们需要斟酌此方案是否弊大于利”(We need to consider whether the *disadvantages* of the plan outweigh the advantages.)、“不会使用计算机的人处于劣势”(Anyone who can not use a computer is at a *disadvantage*.)、“缺乏公共交通是此地生活的不便之一”(One *disadvantage* of living in this area is the lack of public transport.)。
7. “子板”(daughter board)是指插入“母板”(mother board)中的“印制电路板”(PCB)。
8. BNC 连接器即“同轴电缆卡环形接头”,它在小功率射频测试中用于连接同轴电缆。BNC 连接器的发明者是保罗·尼尔(Paul Neill,1882—1968)和卡尔·康塞曼(Carl Concelman,1912—1975)。

Exercises

1. Fill in the blanks with proper words, phrases or clauses.

(1) Printed circuit board is a flat board that holds chips and other electronic components. The board is made of _____(层)(typically 2 to 10) that interconnects components _____(经由)copper pathways. The main printed circuit board in a system is called a “system board” or “mother board,”_____ smaller ones that plug into the slots in the main board are called “boards” or “cards”.

The printed circuit board of the 1960s connected _____(分立元件)together. The circuit board of the 1990s interconnects chips, each containing _____(几十万个)or millions of elementary components. The “printed” circuit is really an _____(蚀刻电路). A _____(铜箔)is placed over the _____(玻璃纤维)or plastic base of each layer and covered with a photoresist. Light is shined through a _____(负像)of the circuit paths onto the photoresist, hardening the areas that will remain _____ etching. When passed through an acid bath, the unhardened areas are washed away. The finished layers are then glued together. A _____(相似的处理)creates the microminiaturized circuits _____ a chip.

(2) The interface is a connection and interaction between hardware, software and the user. Hardware interfaces are the plugs, sockets, wires and the _____(电脉冲)traveling through them _____ a particular pattern. Also included are electrical timing considerations. Examples _____ RS-232 transmission, the Ethernet and Token Ring network topologies and the IDE, ESDI, SCSI, ISA, EISA and Micro Channel interfaces. Software interfaces are the languages, codes and messages programs use _____ communicate with each other and to the hardware. Examples are the applications that run under the Mac, DOS and Windows

operating systems _____ the SMTP E-mail and LU 6.2 _____（通信协议）. User interfaces are the keyboards, mice, commands and menus used for communication between you and the computer. Examples are the command lines in DOS and UNIX, and the Mac, Windows and Motif _____(图形界面).

Interfacing is a major part of what _____(工程技术人员、程序设计人员和咨询人员) do. Users "talk to" the software. The software "talks to" the hardware and other software. Hardware "talks to" other hardware. All this is interfacing. Every interface _____(暗含着一种结构). Electrical signals are made up of _____(电压电平、频率和持续时间). The data passed from one device or program to another has a precise format (header, body, trailer, etc.). Every interface implies a function. At the _____ level, electronic signals activate functions; data are _____, written, _____, received, analyzed for error, etc. At the software level, instructions _____ the hardware (access methods, data link protocols, etc.). At higher levels, the data transferred or transmitted may itself request functions to be performed (client/server, program to program, etc.).

2. Translate the following passages into Chinese or English.

(1) Use of custom gate array logic, application-specific integrated circuits (ASICs), ball grid arrays (BGAs), multichip modules (MCMs), and digital devices operating in the sub-nanosecond range present new and challenging opportunities for EMC engineers.

(2) A resistor at high frequency acts like a series combination of inductance with the resistor in parallel with a capacitor. A capacitor at high frequency acts like an inductor and resistor in series combination with the capacitor plates.

(3) SPICE (Simulation Program with Integrated Circuit Emphasis) is a program widely used to simulate the performance of analog electronic systems and mixed-mode analog and digital systems. SPICE solves sets of non-linear differential equations in the frequency domain, steady state and time domain and can simulate the behavior of transistor and gate designs.

(4) IBIS (I/O Buffer Information Specification) is a format for defining the analog characteristics of the input and output of integrated circuits. IBIS models are ASCII files that provide the behavioral information required to model the device without divulging the proprietary design of the circuit.

(5) In multiconductor systems, excessive line-to-line coupling, or crosstalk, can cause two detrimental effects. First, crosstalk will change the performance of the transmission lines in a bus by modifying the effective characteristic impedance and propagation velocity. Additionally, crosstalk will induce noise onto other lines. These aspects of crosstalk make system performance heavily dependent on data patterns, line-to-line spacing, and switching rates.

(6) 晶体是含有统一排列分子的透明石英材料。

(7) 随着高速数字系统日益复杂，实现时序和信号完整性也越来越困难。

(8) 串扰是从一条导线到另一条导线的能量耦合。当结构不同的电磁场相互作用时，串扰

就会发生。

(9) 在数字设计中，串扰是普遍存在的——出现在芯片上、印制电路板上、连接器上、芯片封装上和连接电缆上。

(10) 许多系统工作在高频。在这些频率，导体不仅表现为导线，还展现出高频效应，表现为传输线。如果没有正确处理这些传输线，它们可能在无意中破坏了系统时序。

3. Writing tasks.

Summary: Write an abstract for *Lesson 8 Clock Sources* within 150～200 words.

Composition: Write an essay about some electronic component you often use.

Reading Materials

Passage 1 Cables

Coaxial Cable

Coaxial cable (coax) consists of a center conductor within an insulating dielectric material. A copper braided or foil shield surrounds the dielectric to protect the conductor against electromagnetic interference (EMI). The shield is encased in a protective insulating jacket, as shown in Figure 3.5. BNC (Bayonet Neill-Concelman) connectors are typically used for coaxial connections. Coax is used in data transmission applications with data rates up to about 1 GHz. Two common applications for coax are cable TV and Internet connections.

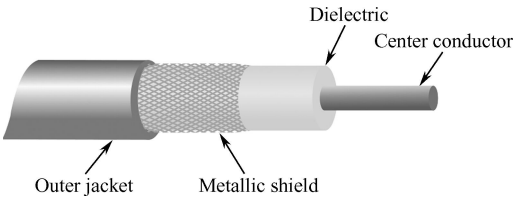


Figure 3.5 Construction view of a coaxial cable

Twisted Pair Cable

Unshielded twisted pair (UTP) cable is used extensively for indoor telephone application as well as some outdoor uses. It is found in many computer networks and video applications, such as security cameras, and also in the USB (universal serial bus) cable. UTP is color-coded according to a standard 25-pair color code. Most cables use a subset of these standard colors.

Crosstalk, a type of distortion, is minimized when twisted pairs are bundled together. The two wires in each pair are twisted so that they cross each other at nearly 90°, ideally cancelling any electromagnetic fields generated by the signals in the wires. UTP cables are limited to use in low-noise environments and to lower signal frequencies than coax, such as audio and other signals up to about 1 MHz. UTP cables use standard RJ-45 connectors. A common four-pair UTP cable is shown in Figure 3.6(a), and an RJ-45 connector is shown in Figure 3.6(b). Shielded twisted pair (STP) cable encased in a metal sleeve or conduit is also available and provides more protection from EMI.

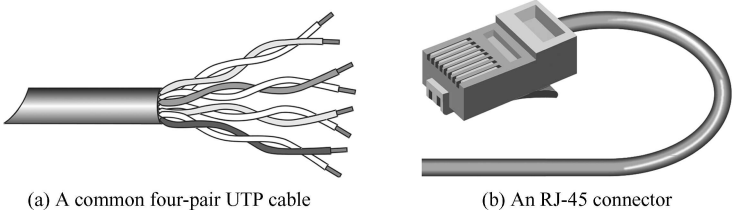


Figure 3.6 Example of an unshielded twisted pair (UTP) cable and connector

Optical Fiber Cable

The structure of a single optical fiber is shownin Figure 3.7(a). An optical fiber can be as small as a human hair, so many single fibers can be bundled into a cable, as shown in Figure 3.7(b).

Instead of using electrical pulses to transmit information through copper lines, fiber optics uses light pulses transmitted through optical fibers. Fiber-optic systems have several advantages over electrical transmission media. Advantages include faster data rates, higher signal capacity (more signals at a time), and better transmission over longer distances; optical fibers are not susceptible to EMI. The main disadvantage of fiber optics is the cost, which is higher than that of coax, UTP, and STP.

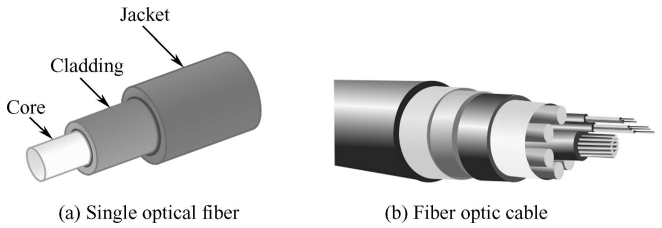


Figure 3.7 Optical fiber cables

Optical fiber is commonly used as a medium for telecommunication and networking. Because light propagates through the fiber with little attenuation compared to electrical cables, optical fiber is useful for long-distance transmission. Data rates from 10 GHz to 40 GHz are common, although rates over 100 GHz are used.

Questions:

1. What does the acronym EMI stand for?
2. For what applications are coaxial cables mainly used?
3. What does the technical term *crosstalk* mean in this article?
4. Are BNC connectors typically used for coaxial connections or UTP connections?
5. What do you think is the main reason for optical fiber cables' popularity in telecommunication?

Passage 2 Bus Basics

A bus allows two or more devices to communicate with each other, generally for the purpose of transmitting data. A bus is a set of physical wires and connectors and a set of electrical specifications. The physical properties of a typical bus include the number of wires or PCB conductors (width), the configuration and length of the wires or conductors, and the types and configurations of the connectors. The electrical properties of a typical bus include but are not limited to some or all of the following: signal format, signal voltage levels, clock frequency, data transfer speed, bandwidth, data frame format, data rate, handshaking protocol, error detection, impedances, and line termination.

Internal and External Buses Internal buses carry information within a system, that is, from one part of the system to another part of the same system. External buses (also known as expansion buses) are used to connect one system to another separate system. For example, a computer connects to peripheral units such as a monitor, keyboard, mouse, and printer through external buses.

Parallel and Serial Buses Buses can be either parallel or serial. A parallel bus carries data bits simultaneously, and a serial bus carries data bits sequentially one at a time. It would seem that a parallel bus would transmit data faster than a serial bus because multiple data bits can be sent simultaneously. However, this is not always the case. As data rates increase, things like crosstalk across parallel bus lines, timing skew between bus lines, and EMI (electromagnetic interference) become problems that limit the speed. Serial buses are not limited by those factors and can actually transmit data at higher rates than parallel buses in many situations.

Synchronous and Asynchronous Buses A synchronized bus includes a clock in the control lines and has a fixed protocol that is relative to the clock. A synchronous bus is fast but has the disadvantage that every device connected to it must operate at the same clock frequency. Also, the physical length of the bus may be limited because of having to carry a high-frequency clock signal. An asynchronous bus is not clocked, so it can serve various devices with different clock rates. The asynchronous bus uses a handshake protocol to establish communication.

Single-Ended vs. Differential Buses The single-ended bus uses one wire for data and one wire for ground, where the signal voltage on the wire is with respect to ground. The differential bus uses two wires for data and one wire for ground. The data signal is sent on one wire in the twisted pair and its complement (inversion) is sent on the other wire. The difference between the two data wires is the differential signal. Figure 3.8 shows both single-ended and differential operation.

A single-ended transmission is simpler and lower in cost compared to a differential transmission. Differential operation is much less sensitive to noise because of common-mode rejection. This means that when a common noise signal appears on each line, the two noise signals cancel due to the differential operation where the difference between them is zero. Since the data signals are the same but opposite in phase, they are effectively added and the data signal is reinforced while the noise signal is cancelled. Due to the twisted pair, crosstalk is reduced at higher frequencies, thus allowing longer cables.

A bus is typically described by the following parameters (Table 3.1 lists some typical buses and their characteristics):

- **Width**: The number of bits that a bus can transmit at one time. The width of typical buses can vary from 1 bit for a serial bus up to 64 bits for a parallel bus.
- **Frequency**: The clock frequency at which a bus can operate.
- **Transfer speed**: The number of bytes per clock cycle.

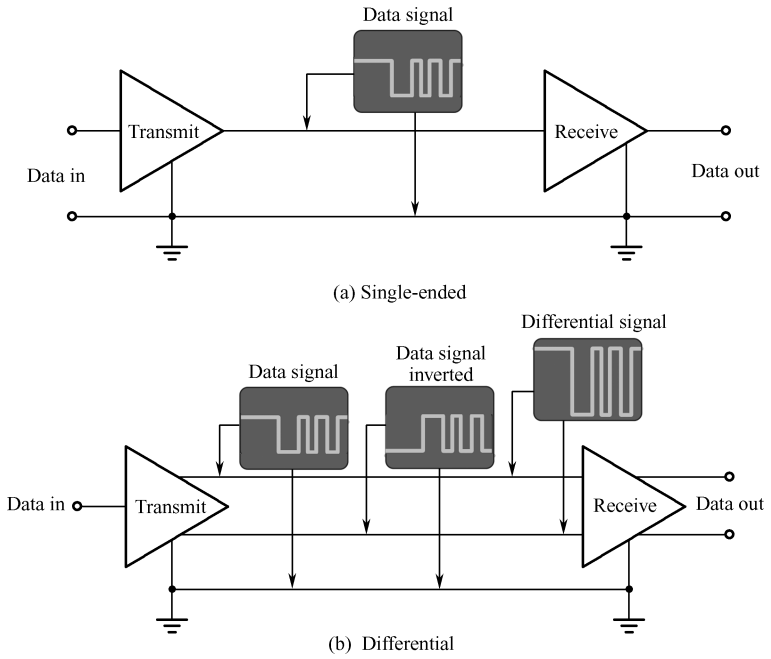


Figure 3.8 A comparison of single-ended and differential buses operation

- **Bandwidth**: The number of bytes per clock cycle times the number of clock cycles per second; that is, transfer speed times frequency. Bus bandwidth is sometimes called *throughput*.

Table 3.1 Some typical buses and their characteristics

Bus	Width (bits)	Frequency (MHz)	Transfer Speed (bytes/cycle)	Bandwidth (Mbps)	
				Decimal	Binary
ISA(16 bit)	16	8.3	2	16.6	15.9
PCI	32	33	4	132	125.9
PCI-X	32	66	4	264	251.8
AGP	32	66	4	264	251.8

Questions:

1. How does a computer connect to its peripheral units?
2. What value is the parameter "width" for a serial bus?
3. What does the technical term *throughput* mean in this article?
4. Is "handshaking protocol" the electrical property or physical property of a bus?
5. What advatages and disadvatages does the differential bus have compared to the single-ended bus?

Passage 3 USB (Universal Serial Bus)

The USB is a widely used standard serial bus for connecting peripherals to a computer.

There are typically two or more USB ports on computers and, with USB hubs, up to 127 devices can be connected. USB allows the devices to be connected or disconnected while the computer is running (hot swapping). Figure 3.9 shows the USB in a typical computer system.

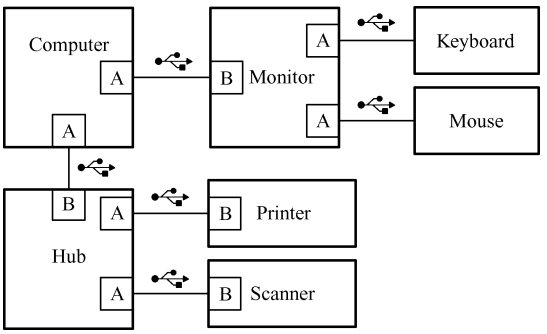


Figure 3.9 Example of USB applications

There have been several major USB standards (see Table 3.2), with USB4 2.0 the newest version available. Most USB devices and cables today work with USB 2.0, and a growing number are updating to USB 3.0.

Table 3.2 Major USB standards

Standard	Date	Maximum transmission rate	Maximum cable length
USB 1.1	1998	12 Mbps (full-speed USB)	3.0 m (9 ft)
USB 2.0	2000	480 Mbps (high-speed USB)	5.0 m (16 ft)
USB 3.0	2008	5Gbps (superspeed USB)	2.0 ~ 3.0 m (6 ~ 9 ft)
USB 3.1	2013	10Gbps (superspeed+ USB)	3.0 m (9 ft)
USB 3.2	2017	20Gbps (superspeed+ USB dual-lane)	3.0 m (9 ft)
USB4	2019	40Gbps	0.8 m (31 in)
USB4 2.0	2022	80Gbps	0.8 m (31 in)

USB versions up to and including 2.0 have a four-wire cable that includes a twisted pair to reduce or eliminate noise for data transmission, a +5V wire, and a ground wire, as shown in Figure 3.10(a). The standard Type-A and Type-B connectors are shown in parts (b) and (c) with pin designations. USB hosts (computer) and devices (peripherals) have sockets, and all USB cables have a Type-A plug at one end and a Type-B plug at the other. The sockets on a host are Type-A, and the sockets on peripheral devices are Type-B. Hubs have both Type-A and Type-B. The USB standard also specifies smaller connectors designated USB-mini and USB-micro.

The USB 3.0 connectors are different than version 2.0 because they now have nine contacts instead of four. The 3.0 Type-A connector looks about the same as that for 2.0 except that the extra five pins are further inside and make contact only with a 3.0 mating connector. The connector is compatible with a 2.0 device where the front four pins are accessible.

The USB Type-C connector supports the USB 3.1 standard, which enables data transfer

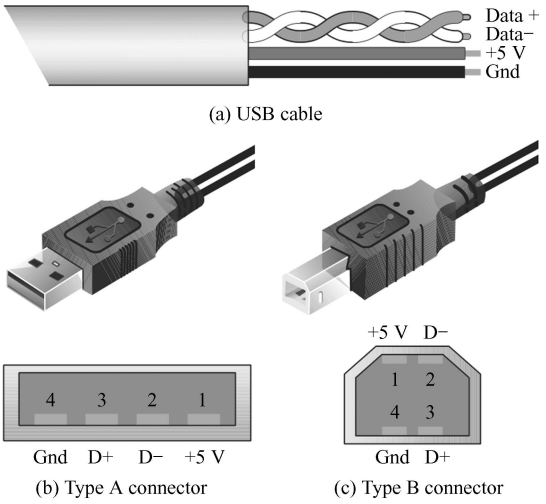


Figure 3.10 USB cable and connectors for USB standards through 2.0

speeds of up to 10Gbps and power delivery of up to 100 watts, facilitating faster charging and efficient data handling. A USB-mini connector connects a variety of smaller devices like digital cameras, MP3 players and older mobile phones to computers for data transfer and charging. A USB-micro connector is a compact version of the USB interfaces. It is mostly for small, portable devices like smartphones, tablets and digital cameras.

Questions:

1. What does the acronym USB stand for?
2. What does the technical term *peripheral* mean in this article?
3. How many devices can be connected to a computer via USB with USB hubs?
4. What is the main difference between USB 3.0 connectors and USB 2.0 connectors?
5. What is the maximum power delivery in USB 3.1 standard?

Unit 4

Electronic Systems

Lesson 10 The Mobile Telephone System (Ⅰ)

Lesson 11 The Mobile Telephone System (Ⅱ)

Lesson 12 The Personal Computer System

Passage 1 GPS

Passage 2 802.11 Wireless LAN

Passage 3 Mechatronic Systems

Lesson 10 The Mobile Telephone System (Ⅰ)

Wireless telephones come in two basic varieties: cordless phones and mobile phones (sometimes called cell phones). Cordless phones are devices consisting of a base station and a handset sold as a set for use within the home. These are never used for networking. Mobile phones have gone through three distinct generations, with different technologies: (1) Analog voice; (2) Digital voice; (3) Digital voice and data.

The first mobile system was devised in the U.S. by AT&T[1] and mandated for the whole country by the FCC. As a result, the entire U.S. had a single (analog) system and a mobile phone purchased in California also worked in New York. In contrast, when mobile came to Europe, every country devised its own system, which resulted in a fiasco.

Europe learned from its mistake and when digital came around, the government-run PTTs[2] got together and standardized on a single system (GSM), so any European mobile phone will work anywhere in Europe. By then, the U.S. had decided that government should not be in the standardization business, so it left digital to the marketplace. This decision resulted in different equipment manufacturers producing different kinds of mobile phones.

Despite an initial lead by the U.S., mobile phone ownership and usage in Europe is now far greater than in the U.S. Having a single system for all of Europe is part of the reason, but there is more. A second area where the U.S. and Europe differed is in the humble matter of phone numbers. In the U.S. mobile phones are mixed in with regular (fixed) telephones. Thus, there is no way for a caller to see if, say, (212) 234-5678 is a fixed telephone or a mobile phone. To keep people from getting nervous about using the telephone, the telephone companies decided to make the mobile phone owner pay for incoming calls. As a consequence, many people hesitated to buy a mobile phone for fear of running up a big bill by just receiving calls. In Europe, mobile phones have a special area code (analogous to 800 and 900 numbers) so they are instantly recognizable. Consequently, the usual rule of "caller pays" also applies to mobile phones in Europe (except for international calls where costs are split).

A third issue that has had a large impact on adoption is the widespread use of prepaid mobile phones in Europe (up to 75% in some areas). These can be purchased in many stores with no more formality than buying a radio. You pay and you go. They are preloaded with, for example, 20 or 50 euro and can be recharged when the balance drops to zero. As a consequence, practically every teenager and many small children in Europe have (usually prepaid) mobile phones so their parents can locate them, without the danger of the child

running up a huge bill. If the mobile phone is used only occasionally, its use is essentially free since there is no monthly charge or charge for incoming calls.

IMTS (Improved Mobile Telephone System)

Mobile radiotelephones were used sporadically for maritime and military communication during the early decades of the 20th century. In 1946, the first system for car-based telephones was set up in St. Louis. This system used a single large transmitter on top of a tall building and had a single channel, used for both sending and receiving. To talk, the user had to push a button that enabled the transmitter and disabled the receiver. Such systems, known as push-to-talk systems, were installed in several cities beginning in the late 1950s. CB-radio, taxis, and police cars on television programs often use this technology.

In the 1960s, IMTS (improved mobile telephone system) was installed. It, too, used a high-powered (200-watt[3]) transmitter, on top of a hill, but now had two frequencies, one for sending and one for receiving, so the push-to-talk button was no longer needed. Since all communication from the mobile telephones went inbound on a different channel than the outbound signals, the mobile users could not hear each other (unlike the push-to-talk system used in taxis). IMTS supported 23 channels spread out from 150 MHz to 450 MHz. Due to the small number of channels, users often had to wait a long time before getting a dial tone. Also, due to the large power of the hilltop transmitter, adjacent systems had to be several hundred kilometers apart to avoid interference. All in all, the limited capacity made the system impractical.

AMPS (Advanced Mobile Phone System)

All that changed with AMPS (advanced mobile phone system), invented by Bell Labs and first installed in the United States in 1982. It was also used in England, where it was called TACS, and in Japan, where it was called MCS-L1. Many of its fundamental properties have been directly inherited by its digital successor, D-AMPS, in order to achieve backward compatibility.

In all mobile phone systems, a geographic region is divided up into cells, which is why the devices are sometimes called cell phones. In AMPS, the cells are typically 10 to 20 km across; in digital systems, the cells are smaller. Each cell uses some set of frequencies not used by any of its neighbors. The key idea that gives cellular systems far more capacity than previous systems is the use of relatively small cells and the reuse of transmission frequencies in nearby (but not adjacent) cells. Whereas an IMTS system 100 km across can have one call on each frequency, an AMPS system might have 100 10-km cells in the same area and be able to have 10 to 15 calls on each frequency, in widely separated cells. Thus, the cellular design increases the system capacity by at least an order of magnitude[4], more as the cells get smaller. Furthermore, smaller cells mean that less power is needed, which leads to smaller

and cheaper transmitters and handsets. Hand-held telephones put out 0.6 watts; transmitters in cars are 3 watts, the maximum allowed by the FCC.

The idea of frequency reuse is illustrated in Figure 4.1(a). The cells are normally roughly circular, but they are easier to model as hexagons. In Figure 4.1(a), the cells are all the same size. They are grouped in units of seven cells. Each letter indicates a group of frequencies. Notice that for each frequency set, there is a buffer about two cells wide where that frequency is not reused, providing for good separation and low interference.

In an area where the number of users has grown to the point that the system is overloaded, the power is reduced, and the overloaded cells are split into smaller microcells to permit more frequency reuse, as shown in Figure 4.1(b). Telephone companies sometimes create temporary microcells[5], using portable towers with satellite links at sporting events, rock concerts, and other places where large numbers of mobile users congregate for a few hours.

(a) Frequencies are not reused in adjacent cells

(b) To add more users, smaller cells can be used

Figure 4.1 The concept of frequency reuse in cellular networks

At the center of each cell is a base station to which all the telephones in the cell transmit. The base station consists of a computer and transmitter/receiver connected to an antenna. In a small system, all the base stations are connected to a single device called an MTSO (mobile telephone switching office) or MSC (mobile switching center). In a larger one, several MTSOs may be needed, all of which are connected to a second-level MTSO, and so on. The MTSOs are essentially end offices as in the telephone system, and are, in fact, connected to at least one telephone system end office. The MTSOs communicate with the base stations, each other, and the PSTN using a packet-switching network.

New Words & Expressions

mandate [mænˈdeɪt] *vt*. 委任，授权

fiasco [fiˈæskəʊ] *n*. 惨败

humble [ˈhʌmbl] *adj*. 不重要的

as a consequence 因此

for fear of 为了避免

install [ɪnˈstɔːl] *vt*. 安装

all in all 总而言之

successor [səkˈsesə] *n*. 后续事物

geographic [ˌdʒiːəˈgræfɪk] *adj*. 地理的

cell [sel] *n*. 蜂窝

analogous [ə'nælə̇gəs] *adj*. 类似的
formality [fɔː'mælətɪ] *n*. 手续
run up　积欠
incoming ['ɪnkʌmɪŋ] *adj*. 输入的
cellular ['seljʊlə] *adj*. 蜂窝状的
put out　消除
circular ['sɜːkjʊlə] *adj*. 圆形的
hexagon ['heksəgən] *n*. 六边形

Technical Terms

handset ['hænset] *n*. 手持设备
high-powered [ˌhaɪ'paʊəd] *adj*. 大功率的
watt [wɒt] *n*. 瓦特
inbound ['ɪnbaʊnd] *adj*. 输入的
outbound ['aʊtbaʊnd] *adj*. 输出的
interference [ɪntə'fɪərəns] *n*. 干扰
capacity [kə'pæsɪti] *n*. 容量
buffer ['bʌfə] *n*. 缓冲区
overload [ˌəʊvə'ləʊd] *n*. 过载
antenna [æn'tenə] *n*.天线
microcell ['maɪkrəʊˌsel] *n*. 微蜂窝
cordless phone　无绳电话
base station　基站
dial tone　拨号音
order of magnitude　数量级
frequency reuse　频率复用
packet switching　分组交换
end office　端局
FCC *abbr*. Federal Communications Commission　联邦通信委员会
PTT *abbr*. post telephone and telegraph administration　邮政电话电报管理局
CB *abbr*. citizen's band　民用波段
MTSO *abbr*. mobile telephone switching office　移动电话交换局
MSC *abbr*. mobile switching center　移动电话交换中心
PSTN *abbr*. public switched telephone network　公共交换电话网

Notes

1. "美国电话电报公司"(AT&T，American Telephone and Telegraph Corporation)曾是美国最大的本地和长途电话公司，其前身是在 1877 年创建的美国贝尔电话公司。
2. "邮政电话电报管理局"(PTT，Post Telephone and Telegraph Administration)是对除美国、加拿大外的国家经营的公用电信部门的通称。
3. 功率单位"瓦特"得名自英国工程师、发明家詹姆斯　瓦特(James Watt)。他对蒸汽机

做出了重大改进，使之发展成现代的高压蒸汽机，并于 1769 年获得专利。

4. “数量级”(order of magnitude)指数量大小的级别，每个级别间存在固定比例。常用比例有 2、10、1000、1024 和 e，最常用的是十进制下的数量级。如“两数相差 3 个数量级”即“一个数比另一个大 1000 倍”。
5. “微蜂窝”(microcell)的发射功率较小(一般在 2 W 左右)，覆盖半径为 100 m～1 km(如一幢建筑物或一个街区)。

Lesson 11 The Mobile Telephone System (Ⅱ)

The first generation of mobile phones was analog, the second generation was digital. Just as there was no worldwide standardization during the first generation, there was also no standardization during the second, either. Four systems were used: D-AMPS, GSM, CDMA, and PDC. PDC was used only in Japan and was basically D-AMPS modified for backward compatibility with the first-generation Japanese analog system.

D-AMPS (The Digital Advanced Mobile Phone System)

The second generation of the AMPS systems is D-AMPS and is fully digital. It is described in International Standard IS-54 and its successor IS-136. D-AMPS was carefully designed to coexist with AMPS so that both first- and second- generation mobile phones could operate simultaneously in the same cell. In particular, D-AMPS uses the same 30 kHz channels as AMPS and at the same frequencies so that one channel can be analog and the adjacent ones can be digital.

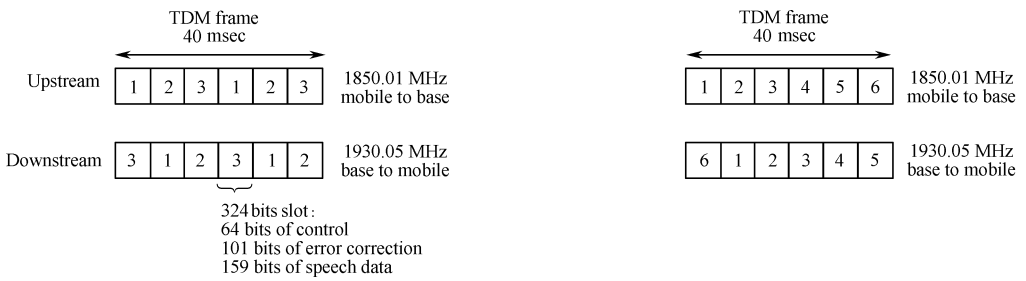
When D-AMPS was introduced as a service, a new frequency band was made available to handle the expected increased load. The upstream channels were in the 1850-1910 MHz range, and the corresponding downstream channels were in the 1930-1990 MHz range, again in pairs, as in AMPS. In this band, the waves are 16 cm long, so a standard 1/4-wave antenna is only 4 cm long, leading to smaller phones. However, many D-AMPS phones can use both the 850-MHz and 1900-MHz bands to get a wider range of available channels.

On a D-AMPS mobile phone, the voice signal picked up by the microphone is digitized and compressed using a model that is more sophisticated than the delta modulation and predictive encoding schemes. Compression takes into account detailed properties of the human vocal system to get the bandwidth from the standard 56-kbps PCM encoding to 8 kbps or less. The compression is done by a circuit called a vocoder [1]. The compression is done in the telephone, rather than in the base station or end office, to reduce the number of bits sent over the air link. With fixed telephony, there is no benefit to having compression done in the telephone, since reducing the traffic over the local loop does not increase system capacity at all.

With mobile telephony there is a huge gain from doing digitization and compression in the handset, so much so that in D-AMPS, three users can share a single frequency pair using time-division multiplexing. Each frequency pair supports 25 frames/sec of 40 msec each. Each frame is divided into six time slots of 6.67 msec each, as illustrated in Figure 4.2(a) for the lowest frequency pair.

Each frame holds three users who take turns using the upstream and downstream links. During slot 1 of Figure 4.2(a), for example, user 1 may transmit to the base station and user 3 is receiving from the base station. Each slot is 324 bits long, of which 64 bits are used for guard times [2], synchronization, and control purposes, leaving 260 bits for the user payload. Of the payload bits, 101 are used for error correction over the noisy air link, so ultimately only 159 bits are left for compressed speech. With 50 slots/sec, the bandwidth available for compressed speech is just under 8 kbps, 1/7 of the standard PCM bandwidth.

Using better compression algorithms, it is possible to get the speech down to 4 kbps, in which case six users can be stuffed into a frame, as illustrated in Figure 4.2(b). From the operator's perspective, being able to squeeze three to six times as many D-AMPS users into the same spectrum as one AMPS user is a huge win and explains much of the popularity of PCS (personal communication service).



(a) A D-AMPS channel with 3 users (b) A D-AMPS channel with 6 users

Figure 4.2 The TDM frames in D-AMPS channels

GSM (The Global System for Mobile Communications)

D-AMPS is widely used in the U. S. and (in modified form) in Japan. Virtually everywhere else in the world, a system called GSM (global system for mobile communications) is used. To a first approximation, GSM is similar to D-AMPS. Both are cellular systems. In both systems, frequency-division multiplexing is used, with each mobile transmitting on one frequency and receiving on a higher frequency (80 MHz higher for D-AMPS, 55 MHz higher for GSM). Also in both systems, a single frequency pair is split by time-division multiplexing into time slots shared by multiple mobiles. However, the GSM channels are much wider than the AMPS channels (200 kHz versus 30 kHz) and hold relatively few additional users (8 versus 3), giving GSM a much higher data rate per user than D-AMPS.

Each frequency band is 200 kHz wide, as shown in Figure 4.3. A GSM system has 124

pairs of simplex channels. Each simplex channel is 200 kHz wide and supports eight separate connections on it, using time-division multiplexing. Each currently active station is assigned one time slot on one channel pair. Theoretically, 992 channels can be supported in each cell, but many of them are not available, to avoid frequency conflicts with neighbouring cells. In Figure 4.3, the eight shaded time slots all belong to the same connection, four of them in each direction. Transmitting and receiving does not happen in the same time slot because the GSM radios cannot transmit and receive at the same time and it takes time to switch from one to the other. If the mobile station assigned to 890.4/935.4 MHz and time slot 2 wanted to transmit to the base station, it would use the lower four shaded slots (and the ones following them in time), putting some data in each slot until all the data had been sent.

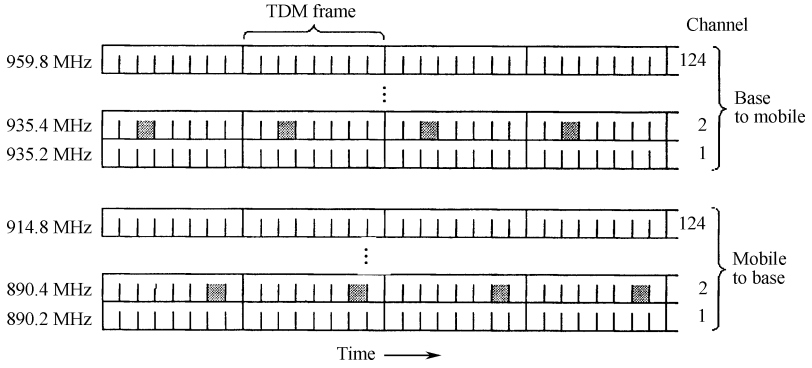


Figure 4.3 GSM uses 124 frequency channels, each of which uses an eight-slot TDM system

The TDM slots shown in Figure 4.3 are part of a complex framing hierarchy. Each TDM slot has a specific structure, and groups of TDM slots form multiframes, also with a specific structure. A simplified version of this hierarchy is shown in Figure 4.4. Here we can see that each TDM slot consists of a 148-bit data frame that occupies the channel for 577 μsec (including a 30-μsec guard time after each slot). Each data frame starts and ends with three 0 bits, for frame delineation purposes. It also contains two 57-bit Information fields, each one having a control bit that indicates whether the following Information field is for voice or data. Between the Information fields is a 26-bit Sync (training) field that is used by the receiver to synchronize to the sender's frame boundaries.

A data frame is transmitted in 547 μsec, but a transmitter is only allowed to send one data frame every 4.615 msec, since it is sharing the channel with seven other stations. The gross rate of each channel is 270,833 bps, divided among eight users. This gives 33.854 kbps gross, more than double D-AMPS' 50 times 324 bits per second for 16.2 kbps.

As can be seen from Figure 4.4, eight data frames make up a TDM frame and 26 TDM frames make up a 120-msec multiframe. Of the 26 TDM frames in a multiframe, slot 12 is used for control and slot 25 is reserved for future use, so only 24 are available for user traffic.

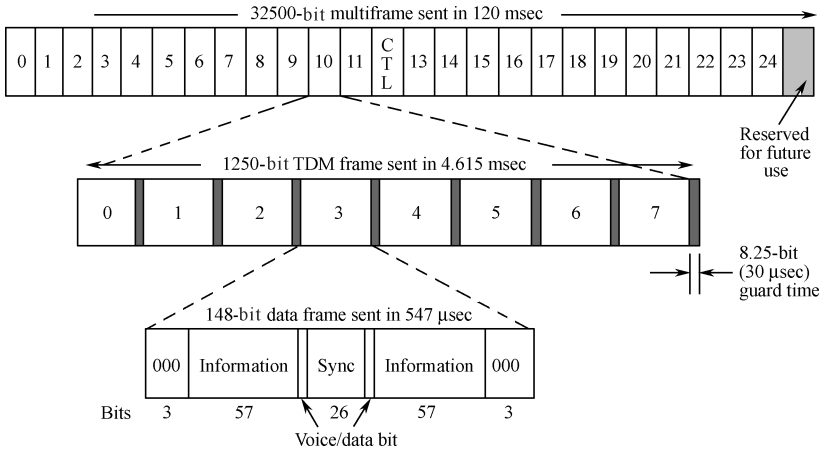


Figure 4.4 A portion of the GSM framing structure

CDMA (Code Division Multiple Access)

CDMA is completely different from AMPS, D-AMPS, and GSM. Instead of dividing the allowed frequency range into a few hundred narrow channels, CDMA allows each station to transmit over the entire frequency spectrum all the time. Multiple simultaneous transmissions are separated using coding theory. CDMA also relaxes the assumption that colliding frames are totally garbled. Instead, it assumes that multiple signals add linearly.

Let us consider an analogy: an airport lounge with many pairs of people conversing. TDM is comparable to all the people being in the middle of the room but taking turns speaking. FDM is comparable to the people being in widely separated clumps, each clump holding its own conversation at the same time as, but still independent of, the others. CDMA is comparable to everybody being in the middle of the room talking at once, but with each pair in a different language. The French-speaking couple just hones in on the French, rejecting everything that is not French as noise. Thus, the key to CDMA is to be able to extract the desired signal while rejecting everything else as random noise.

In CDMA, each bit time is subdivided into m short intervals called chips. Typically, there are 64 or 128 chips per bit, but in the example given below we will use 8 chips/bit for simplicity.

Each station is assigned a unique m-bit code called a chip sequence. To transmit a 1 bit, a station sends its chip sequence. To transmit a 0 bit, it sends the one's complement of its chip sequence. No other patterns are permitted. Thus, for $m=8$, if station A is assigned the chip sequence 00011011, it sends a 1 bit by sending 00011011 and a 0 bit by sending 11100100.

Increasing the amount of information to be sent from b bits/sec to mb chips/sec can only be done if the bandwidth available is increased by a factor of m, making CDMA a form of spread spectrum communication (assuming no changes in the modulation or encoding techniques). If we have a 1-MHz band available for 100 stations, with FDM each one would have 10 kHz and could send at 10 kbps (assuming 1 bit per Hz). With CDMA, each station

uses the full 1 MHz, so the chip rate is 1 megachip per second. With fewer than 100 chips per bit, the effective bandwidth per station is higher for CDMA than FDM, and the channel allocation problem is also solved.

In an ideal, noiseless CDMA system, the capacity (i.e., number of stations) can be made arbitrarily large, just as the capacity of a noiseless Nyquist channel can be made arbitrarily large by using more and more bits per sample. In practice, physical limitations reduce the capacity considerably. First, we have assumed that all the chips are synchronized in time. In reality, such synchronization is impossible. What can be done is that the sender and receiver synchronize by having the sender transmit a predefined chip sequence that is long enough for the receiver to lock onto. All the other (unsynchronised) transmissions are then seen as random noise. As one might expect, the longer the chip sequence is, the higher the probability of detecting it correctly in the presence of noise. For extra reliability, the bit sequence can use an error-correcting code. Chip sequences never use error-correcting codes.

An implicit assumption in our discussion is that the power levels of all stations are the same as perceived by the receiver. CDMA is typically used for wireless systems with a fixed base station and many mobile stations at varying distances from it. The power levels received at the base station depend on how far away the transmitters are. A good heuristic here is for each mobile station to transmit to the base station at the inverse of the power level it receives from the base station. In other words, a mobile station receiving a weak signal from the will use more power than one getting a strong signal. The base station can also give explicit commands to the mobile stations to increase or decrease their transmission power.

We have also assumed that the receiver knows who the sender is. In principle, given enough computing capacity, the receiver can listen to all the senders at once by running the decoding algorithm for each of them in parallel. In real life, suffice it to say that this is easier said than done. CDMA also has many other complicating factors that have been glossed over here. Nevertheless, CDMA is a clever scheme that is being rapidly introduced for wireless mobile communication. It normally operates in a band of 1.25 MHz (versus 30 kHz for D-AMPS and 200 kHz for GSM), but it supports many more users in that band than either of the other systems. In practice, the bandwidth available to each user is at least as good as GSM and often much better.

New Words & Expressions

coexist [ˌkəʊɪgˈzɪst] *vi*. 共存
approximation [əˌprɒksɪˈmeɪʃn] *n*. 类似物
mobile [ˈməʊbaɪl] *adj*. 移动的
delineation [dɪˈlɪnɪeɪʃn] *n*. 描绘，叙述
gross [grəʊs] *adj*. 总计的
garble [ˈgɑːbl] *vt*. 混淆
hone [həʊn] *vt*. 渴望
subdivide [sʌbdɪˈvaɪd] *vt*. 细分
in practice 实际上
heuristic [hjʊˈrɪstɪk] *n*. 直观推断
inverse [ɪnˈvɜːs] *n*. 倒数
computing [kəmˈpjuːtɪŋ] *n*. 计算，处理

collide [kəˈlaɪd] *vi*. 碰撞
analogy [əˈnælədʒi] *n*. 类似
lounge [laʊndʒ] *n*. 休息室
comparable [ˈkɑmpərəbl] *adj*. 可比较的
clump [klʌmp] *n*. 丛
in parallel　并行地
real life　现实生活
suffice it to say that　说……就足够了
complicate [ˈkɒmplɪkeɪt] *vt*. 使……变复杂
scheme [skiːm] *n*. 方案

Technical Terms

upstream [ˈʌpstriːm] *n*. 上行比特流
downstream [ˈdaʊnˈstriːm] *n*. 下行比特流
vocoder [ˈvəʊˌkəʊdə] *n*. 声码器
synchronization [ˌsɪŋkrənaɪˈzeɪʃn] *n*. 同步
payload [ˈpeɪˌləʊd] *n*. 有效载荷
simplex [ˈsɪmpleks] *n*. 单工，单向通信
multiframe *n*. 复帧
delta modulation　增量调制(Δ调制)
predictive encoding　预测编码
frequency conflict　频率冲突
coding theory　编码理论
random noise　随机噪声
one s complement　二进制反码
spread spectrum communication　扩频通信
transmission power　发射功率
TDM *abbr*. time-division multiplexing　时分复用
PCM *abbr*. pulse code modulation　脉冲编码调制
FDM *abbr*. frequency-division multiplexing　频分复用
PCS *abbr*. personal communication service　个人通信业务

Notes

1. “声码器”(vocoder)是由“语音”(voice)和“编码器”(coder)拼接而成的。vocoder 还可译为“语音编码器”“自动语音合成器”。
2. “保护时间”(guard time)是为防止 TDMA 系统中脉冲信号的重叠而提供的时间间隔。

Lesson 12　The Personal Computer System

When you mention the word “technology”, most people think about computers. Virtually every facet of our lives has some computerized component. The appliances in our homes have microprocessors built into them, as do our televisions. Even our cars have a computer. But the

computer that everyone thinks of first is typically the personal computer, or PC.

A PC is a general-purpose tool built around a microprocessor. It has lots of different parts — memory, a hard disk, a modem, etc. —that work together. "General purpose" means that you can do many different things with a PC. You can use it to type documents, send E-mail, browse the Web and play games.

Core PC Components

- **Case**— If you're using a laptop, the computer case includes keyboard and screen. For desktop PCs, the case is typically some type of box with lights, vents, and places for attaching cables. A larger case doesn't always imply a more powerful computer; it's what's inside that counts. PC builders design or select a case based on the type of mother board that should fit inside.
- **Mother board** — The primary circuit board inside your PC is its mother board. All components, inside and out, connect through the mother board in some way. Several important components are attached directly to the mother board. These include the complementary metal oxide semiconductor (CMOS), which stores some information, such as the system clock, when the computer is powered down. Mother boards come in different sizes and standards, the most common being ATX and MicroATX.
- **Power supply**—Other than its CMOS, which is powered by a replaceable CMOS battery on the mother board, every component in your PC relies on its power supply. The power supply connects to some type of power source, whether that's a battery in the case of mobile computers, or a power outlet in the case of desktop PCs.
- **Central processing unit (CPU)** — The CPU, often just called the processor, is the component that contains the microprocessor. That microprocessor is the heart of all the PC's operations, and the performance of both hardware and software rely on the processor's performance. Intel and AMD are the largest CPU manufacturers for PCs. The two common CPU architectures are 32-bit and 64-bit, and you'll find that certain software relies on this architecture distinction.
- **Random-access memory (RAM)** — Even the fastest processor needs a buffer to store information while it's being processed. The RAM is to the CPU as a countertop is to a cook: It serves as the place where the ingredients and tools you're working with wait until you need to pick up and use them. Both a fast CPU and an ample amount of RAM are necessary for a speedy PC.
- **Drives**—A drive is a device intended to store data when it's not in use. A hard drive or solid state drive stores a PC's operating system and software. This category also includes optical drives such as those used for reading and writing CD, DVD and Blu-ray media. A drive connects to the mother board based on the type of drive controller technology it uses, including the older IDE standard and the newer SATA standard.
- **Cooling devices** — The more your computer processes, the more heat it generates. The

CPU and other components can handle a certain amount of heat. However, if a PC isn't cooled properly, it can overheat, causing costly damage to its components and circuitry. Fans are the most common device used to cool a PC. In addition, the CPU is covered by a metallic block called a heat sink, which draws heat away from the CPU. Some serious computer users, such as gamers, sometimes have more expensive heat management solutions, like a water-cooled system, designed to deal with more intense cooling demands.

- **Cables**—All the components we've mentioned so far are connected by some combination of cables. These cables are designed to carry data, power or both. PCs should be constructed so that the cables fold neatly within the case and do not block air flow throughout it.

Ports, Peripherals and Expansion Slots

- **Graphics components**— While some mother boards have on-board graphics, others include what's called an expansion slot, where you can slide in a separate video card. In both cases, the video components in a PC process some of the complex graphics data going to the screen, taking some of the load off your CPU. A mother board accepts video cards based on a specific interface, such as the older AGP standard or one of the newer PCI standards.
- **Ports**— The word port [1] is often used to describe a place on the outside of your PC where you can plug in a cable. Describe a port by its use, such as a USB port or an Ethernet port. Many ports are affixed directly to the mother board. Some of the ports you'll find on a PC include the following: (1) USB ports; (2) network ports, typically Ethernet and FireWire[2]; (3) video ports, typically some combination of VGA, DVI, RCA[3], S-Video[4] and HDMI; (4) audio ports, typically some combination mini analog audio jacks or RCA; (5)legacy ports, or ports that follow old standards which are rarely used in modern computers, such as parallel printer ports and PS2 ports for a keyboard and mouse.
- **Peripherals**— Any piece of hardware that isn't mounted inside a PC's case is called a peripheral. This includes your basic input and output devices: monitors, keyboards and mice. It also includes printers, speakers, headphones, microphones, webcams and USB flash drives. Anything you can plug in to a port on the PC is one of the PC's peripherals.
- **Expansion slots**—On occasion, you'll want to add components to a PC that don't have a designated slot somewhere on the mother board. That's why the mother board will include a series of expansion slots. The removable components designed to fit into expansion slots are called cards, probably because of their flat, card-like structure. Using expansion slots, you can add extra video cards, network cards, printer ports, TV receivers and many other custom additions.

Powering Up a PC

When you first power up a PC, the machine goes through several internal processes before it's ready for you to use. This is called the boot process, or booting the PC. Boot is short for bootstrap, a reference to the old adage, "Pull yourself up by the bootstraps", which means to start something from the very beginning. The boot process is controlled by the PC's basic input/output system (BIOS).

The BIOS is software stored on a flash memory chip. In a PC, the BIOS is embedded on the mother board. Occasionally, a PC manufacturer will release an update for the BIOS, and you can carefully follow instructions to "flash the BIOS" with the updated software.

Besides controlling the boot process, the BIOS provides a basic configuration interface for the PC's hardware components. In that interface, you can configure such things as the order to read drives during boot and how fast the processor should be allowed to run. Check your PC's documentation to find out how to enter its BIOS interface. This information is often displayed when you first boot the computer, too, with a message such as, "Press DEL to enter Setup Menu."

The following is a summary of the boot process in a PC:

1. The power button activates the power supply in the PC, sending power to the mother board and other components.
2. The PC performs a power-on self-test (POST). The POST is a small computer program within the BIOS that checks for hardware failures. A single beep after the POST signals that everything's okay. Other beep sequences signal a hardware failure, and PC repair specialists compare these sequences with a chart to determine which component has failed.
3. The PC displays information on the attached monitor showing details about the boot process. These include the BIOS manufacturer and revision, processor specs, the amount of RAM installed, and the drives detected. Many PCs have replaced displaying this information with a splash screen showing the manufacturer's logo. You can turn off the splash screen in the BIOS settings if you'd rather see the text.
4. The BIOS attempts to access the first sector of the drive designated as the boot disk. The first sector is the first kilobytes of the disk in sequence, if the drive is read sequentially starting with the first available storage address. The boot disk is typically the same hard disk or solid-state drive that contains your operating system. You can change the boot disk by configuring the BIOS or interrupting the boot process with a key sequence (often indicated on the boot screens).
5. The BIOS confirms there's a bootstrap loader, or boot loader, in that first sector of the boot disk, and it loads that boot loader into memory (RAM). The boot loader is a small program designed to find and launch the PC's operating system.
6. Once the boot loader is in memory, the BIOS hands over its work to the boot loader,

which in turn begins loading the operating system into memory.

7. When the boot loader finishes its task, it turns control of the PC over to the operating system. Then, the OS is ready for user interaction.

PC Operating Systems

After a PC boots, you can control it through an operating system, or OS for short.Most non-Apple PCs run a version of Microsoft Windows or a Linux distribution. These operating systems are designed to run on various kinds of PC hardware, while OS X is designed primarily for Apple hardware.

An operating system is responsible for several tasks. These tasks fall into the following broad categories.

- **Processor management**—Breaks down the processor's work into manageable chunks and prioritizes them before sending them to the CPU.
- **Memory management**—Coordinates the flow of data in and out of RAM, and determines when to use virtual memory on the hard disk to supplement an insufficient amount of RAM.
- **Device management** — Provides a software-based interface between the computer's internal components and each device connected to the computer.
- **Storage management**—Directs where data should be stored permanently on hard drives, solid state drives, USB drives and other forms of storage. For example, storage management tasks assist when creating, reading, editing, moving, copying and deleting documents.
- **Application interface**—Provides data exchange between software programs and the PC. An application must be programmed to work with the application interface for the operating system you're using.
- **User interface (UI)**—Provides a way for you to interact with the computer.

New Words & Expressions

facet [ˈfæsɪt] *n*. 方面
appliance [əˈplaɪəns] *n*. 装置
browse [braʊz] *vt*. 浏览
vent [vent] *n*. 通风孔
imply [ɪmˈplaɪ] *vt*. 意味
attach [əˈtætʃ] *vt*. 使依附
countertop [ˈkaʊntəˌtɒp] *n*. 工作台面
ingredient [ɪnˈgriːdɪənt] *n*. 原料
ample [ˈæmpl] *adj*. 足够的
category [ˈkætɪgəri] *n*. 种类
legacy [ˈlegəsi] *adj*. 老式的
designate [ˈdezɪgneɪt] *vt*. 指定
custom [ˈkʌstəm] *adj*. 定制的
addition [əˈdɪʃn] *n*. 增加物
adage [ˈædɪdʒ] *n*. 谚语
activate [ˈæktɪveɪt] *vt*. 激活
sector [ˈsektə] *n*. 扇区
hand over 移交
interaction [ɪntəˈækʃn] *n*. 相互作用
chunk [tʃʌŋk] *n*. 大块
prioritize [praɪˈɒrəˌtaɪz] *vt*. 分出轻重缓急
coordinate[kəʊˈɔːdɪnət] *vt*. 协调

Technical Terms

modem [ˈməʊdəm] *n*. 调制解调器
laptop [ˈlæptɒp] *n*. 笔记本电脑
drive [draɪv] *n*. 驱动器
cable [ˈkeɪbl] *n*. 电缆
battery [ˈbætəri] *n*. 电池
outlet [ˈaʊtlet] *n*. 电源插座
ethernet [ˈiːθənet] *n*. 以太网
jack [dʒæk] *n*. 音频插口
peripheral [pəˈrɪfərəl] *n*. 外设
webcam [ˈwebkæm] *n*. 网络摄像头
spec [spek] *n*. 规格
power supply　电源
heat sink　散热片
CMOS *abbr*. complementary metal oxide semiconductor　互补金属氧化物半导体
CPU *abbr*. central processing unit　中央处理单元
RAM [ræm] *abbr*. random-access memory　随机存取存储器
CD *abbr*. compact disc　激光唱盘
DVD *abbr*. digital video disk　数字激光视盘
IDE [aɪd] *abbr*. integrated drive electronics　集成驱动器电路
SATA *abbr*. serial advanced technology attachment　串行高级技术附件
USB *abbr*. universal serial bus　通用串行总线
VGA *abbr*. video graphics array　视频图形阵列
DVI *abbr*. digital video interactive　交互式数字视频
HDMI *abbr*. high-definition multimedia interface　高清晰度多媒体接口
RCA *abbr*. Radio Corporation of America　美国无线电公司
POST [pəʊst] *abbr*. power-on self-test　开机自检
OS *abbr*. operating system　操作系统

Notes

1. “接口”(port)一词还可以用来指硬件通信中的软件接口。
2. “火线接口”(FireWire)即 IEEE 1394，是由 Apple(苹果)公司领导的设计联盟开发的一种高速传送接口。
3. “RCA 接口”得名自“美国无线电公司”(Radio Corporation of America，RCA)。在 20 世纪 30 年代，RCA 公司的收音机多采用这样的连接方式。RCA 接口分为“黄、白、红”三色——黄色用于复合视频信号，白色用于左声道音频信号，红色用于右声道音频信号。
4. “S-视频接口”(S-Video)往往称作“S-端子”(S 代表 Separate)。“独立视频信号”

(separate video)不同于“复合视频信号”(composite video)，它将模拟视频信号分为亮度信号、色度信号两个独立信号进行传送。

Exercises

1. Fill in the blanks with proper words，phrases or clauses.

(1) In a TDMA scheme several calls _____(公用一个频道). The scheme is useful for _____(数字化语音)or other digital data. Each call is allocated _____(若干时隙) based on its _____(数据率)within a frame for upstream _____ downstream._____(除去) the user data，each time slot also carries other data for _____(同步)，guard times，and control information.

The transmission _____ base station _____ mobile is done _____ time division multiplex (TDM) mode whereas _____ the upstream direction each mobile transmits in its own time slot. The overlap between different slots resulting _____ different propagation delay is prevented _____ using guard times and precise slot synchronization schemes.

The TDMA scheme is used along with the FDMA scheme because _____ (在一个蜂窝内使用几个频道). The traffic in two directions is separated either by using two separate frequency channels _____ by alternating in time. The two schemes _____ (被称作) frequency-division duplex (FDD) and time-division duplex (TDD)，_____ . The FDD scheme uses less bandwidth than TDD schemes use and does not require as precise synchronization of data flowing in two directions _____ in the TDD method. The latter，however，is useful when flexible bandwidth allocation is required for upstream and downstream traffic.

(2) The CDMA scheme is a direct sequence spread spectrum (DSSS) method. It uses _____(线性调制)with wideband pseudonoise (PN) sequences to generate signals. These sequences，_____(也叫作)codes，spread the spectrum of the modulating signal _____ a large bandwidth，simultaneously reducing the _____(信号谱密度). Thus，various CDMA signals occupy the same bandwidth and appear _____ noise to each other.

In the CDMA scheme，each user is assigned an individual code _____(在……时候)call initiation. This code is used both for spreading the signal at the time of transmission _____ dispreading the signal at the time of reception. Cellular systems _____ (使用 CDMA 方式) use FDD，thus employing two frequency channels _____ forward and reverse links.

On forward-link a mobile transmits to all users synchronously and _____ preserves the orthogonality of various codes assigned to different users. The orthogonality，however，is not preserved between different components arriving from different paths _____ (在多径情形中). On reverse links each user transmits independently from other users _____(由于)their individual locations. Thus，the transmission on reverse link is _____ (异步的)and the various signals are _____(不一定)orthogonal.

_____(应当指出的是)these PN sequences are designed to be orthogonal to each other. In _____ words，the cross correlation between different code sequences is zero and thus the

signal modulated with one code appears to be orthogonal to a receiver using a different code if the orthogonality is preserved during the transmission. This is the case on forward-link and _____ the absence of multipath the signal received by a mobile is not affected by _____ (基站发送给其他移动用户的信号).

2. Translate the following passages into Chinese or English.

(1) The available spectrum bandwidth is shared in a number of ways by various wireless radio links. The way in which this is done is referred to as a multiple access scheme. There are basically four principle schemes. These are frequency-division multiple access (FDMA), time-division multiple access (TDMA), code-division multiple access (CDMA), and space-division multiple access (SDMA).

(2) The area served by mobile phone systems is divided into small areas known as cells. Each cell contains a base station that communicates with mobiles in the cell by transmitting and receiving signals on radio links. The transmission from the base station to a mobile is typically referred to as downstream, forwardlink, or downlink. The corresponding terms for the transmission from a mobile to the base are upstream, reverse-link, and uplink. Each base station is associated with a mobile switching center (MSC) that connects calls to and from the base to mobiles in other cells and the public switched telephone network.

(3) A base station communicates with mobiles using two types of radio channels, control channels to carry control information and traffic channels to carry messages.

(4) In an FDMA scheme the available spectrum is divided into a number of frequency channels of certain bandwidth and individual calls use different frequency channels. All first-generation cellular systems use this scheme.

(5) In TDMA and FDMA systems a frequency channel used in a cell is not used in adjacent cells to prevent co-channel interference. In a CDMA system it is possible to use the same frequency channel in adjacent cells and thus increase the system capacity.

(6) “信道”通常是指“频分多址”系统中的一个频率、“时分多址”系统中的一个时隙、“码分多址”系统中的一个代码，或者是指“组合系统”中频率、时隙、代码的某种组合。

(7) 在“频分多址”系统和“时分多址”系统中，信道(频率或时隙)一旦被分配给用户，则该信道在非活动期间也不能被利用。

(8) “蜂窝”概念是由贝尔实验室提出的。1983 年 10 月，第一个商用模拟话音系统出现在美国芝加哥市。

(9) 移动站是蜂窝无线电业务中应用于运动中的非指定位置站点。移动站可能是手持个人设备，也可能安装在车辆中。

(10) 基站使用两类无线信道和移动用户进行通信：传送控制信息的“控制信道”和传送消息的“通信信道”。

3. Writing tasks.

Summary: Write an abstract for *Lesson* 12 *The Personal Computer System* within 150～200 words.

Composition: Write an essay about an electronic gadget you often use.

Reading Materials

Passage 1 GPS

GPS, which stands for Global Positioning System, is the first system able to show you your exact position on the Earth anytime, in any weather, anywhere. GPS satellites, 24 in all, orbit 11,000 nautical miles above the Earth (see Figure 4.5). They are continuously monitored by ground stations located worldwide. The satellites transmit signals that can be detected by anyone with a GPS receiver. Using the receiver, you can determine your location with great precision.

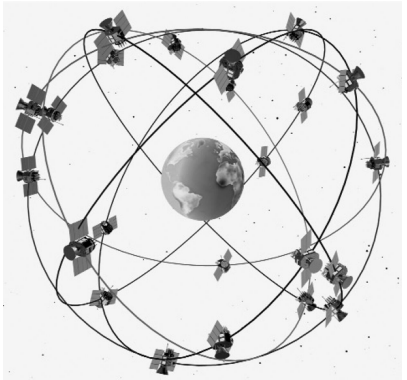


Figure 4.5 GPS satellites

GPS has three parts: the space segment, the user segment, and the control segment. The space segment consists of 24 satellites, each in its own orbit 11,000 nautical miles above the Earth. Satellites are equipped with very precise clocks that keep accurate time to within three nanoseconds. This precision timing is important because the receiver must know exactly how long it takes for its signal to get to each satellite and return. By knowing the exact amount of time the signal has taken to get back from each satellite, it can calculate its position. The first GPS satellite was launched in 1978. The first 10 satellites were developmental satellites, called Block Ⅰ. From 1989 to 1993, 23 production satellites, called Block Ⅱ, were launched. The launch of the 24th satellite in 1994 completed the system.

The user segment consists of receivers, which you can hold in your hand or mount in your car. These receivers detect, decode, and process GPS satellite signals. The typical hand-held receiver is about the size of a cellular telephone, and the newer models are even smaller.

The control segment consists of unmanned monitor stations located around the world (Hawaii and Kwajalein in the Pacific Ocean; Diego Garcia in the Indian Ocean; Ascension Island in the Atlantic Ocean; and Colorado Springs, Colorado); a master ground station at Falcon Air Force Base in Colorado Springs, Colorado; and four large ground antenna stations that broadcast signals to the satellites. The stations also track and monitor the GPS satellites.

One trip around the Earth in space equals one orbit. The GPS satellites each take 12 hours to orbit the Earth. Each satellite is equipped with an accurate clock to let it broadcast signals coupled with a precise time message. The ground unit receives the satellite signal, which travels at the speed of light. Even at this speed, the signal takes a measurable amount of time to reach the receiver.

The difference between the time the signal is sent and the time it is received, multiplied

by the speed of light, enables the receiver to calculate the distance to the satellite. To measure precise latitude, longitude, and altitude, the receiver measures the time it took for the signals from four separate satellites to get to the receiver.

The GPS system can tell you your location anywhere on or above the Earth to within about 300 feet. Even greater accuracy, usually within less than three feet, can be obtained with corrections calculated by a GPS receiver at a known fixed location.

The GPS system was developed to meet military needs of the Department of Defense, but new ways to use its capabilities are continually being found. We'll mention just a few.

During construction of the tunnel under the English Channel, British and French crews started digging from opposite ends: one from Dover, England, one from Calais, France. They relied on GPS receivers outside the tunnel to check their positions along the way and to make sure they met exactly in the middle. Otherwise, the tunnel might have been crooked.

Mapping and surveying companies use GPS extensively. In the field of wildlife management, endangered species such as Montana elk and Mojave Desert tortoises are being fitted with GPS receivers and tiny transmitters to help determine population distribution patterns and possible sources of disease.

GPS-equipped balloons are monitoring holes in the ozone layer over the polar regions, and air quality is being monitored using GPS receivers. Buoys tracking major oil spills transmit data using GPS.

Archaeologists and explorers are using the system. Anyone equipped with a GPS receiver can use it as a reference point to locate or find another location. With a basic knowledge of math and science, plus a hand-held GPS receiver, you could be an instant hero if you and friends got temporarily lost on a camping trip.

The future of GPS is as unlimited as your imagination. New applications will continue to be created as technology evolves. The GPS satellites, like handmade stars in the sky, will be guiding you well into the future.

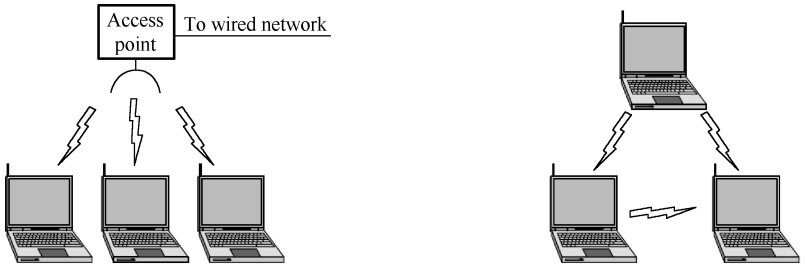
Questions:

1. How is GPS constructed?
2. What does the term *orbit* mean in this article?
3. Why is timing precision important in GPS?
4. At what time was GPS completed?
5. Please give some GPS applications in your daily life.

Passage 2 802.11 Wireless LAN

802.11 networks (a common slang name for it is Wi-Fi) are made up of clients, such as laptops and mobile phones, and infrastructure called APs (access points) that is installed in buildings. Access points are sometimes called base stations. The access points connect to the wired network, and all communication between clients goes through an access point. It is also possible for clients that are in radio range to talk directly, such as two computers in an office

without an access point. This arrangement is called an ad hoc network. It is used much less often than the access point mode. Both modes are shown in Figure 4.6.



(a) Wireless network with an access point (b) Ad hoc network

Figure 4.6 Two operating modes of 802.11 wireless LAN

802.11 transmission is complicated by wireless conditions that vary with even small changes in the environment. At the frequencies used for 802.11, radio signals can be reflected off solid objects so that multiple echoes of a transmission may reach a receiver along different paths. The echoes can cancel or reinforce each other, causing the received signal to fluctuate greatly. This phenomenon is called multipath fading, and it is shown in Figure 4.7.

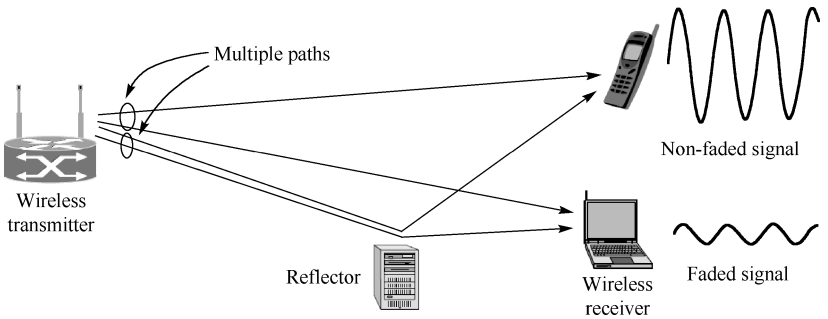


Figure 4.7 Multipath fading

The key idea for overcoming variable wireless conditions is path diversity, or the sending of information along multiple, independent paths. In this way, the information is likely to be received even if one of the paths happens to be poor due to a fade. These independent paths are typically built into the digital modulation scheme at the physical layer. Options include using different frequencies across the allowed band, following different spatial paths between different pairs of antennas, or repeating bits over different periods of time.

Different versions of 802.11 have used all of these techniques. The initial (1997) standard defined a wireless LAN that ran at either 1 Mbps or 2 Mbps by hopping between frequencies or spreading the signal across the allowed spectrum. The spread spectrum design was extended and became the (1999) 802.11b standard running at rates up to 11 Mbps. The 802.11a (1999) and 802.11g (2003) standards switched to a different modulation scheme called OFDM (orthogonal frequency-division multiplexing). It divides a wide band of spectrum into many narrow slices over which different bits are sent in parallel. This improved scheme boosted the

802.11a/g bitrates up to 54 Mbps. That is a significant increase, but people still wanted more throughput to support more demanding uses.

Since wireless is inherently a broadcast medium, 802.11 radios also have to deal with the problem that multiple transmissions that are sent at the same time will collide, which may interfere with reception. To handle this problem, 802.11 uses a CSMA (carrier sense multiple access) scheme that draws on ideas from classic wired Ethernet, which, ironically, drew from an early wireless network developed in Hawaii and called ALOHA. Computers wait for a short random interval before transmitting, and defer their transmissions if they hear that someone else is already transmitting. This scheme makes it less likely that two computers will send at the same time.

Another problem is that of mobility. If a mobile client is moved away from the access point it is using and into the range of a different access point, some way of handing it off is needed. The solution is that an 802.11 network can consist of multiple cells, each with its own access point, and a distribution system that connects the cells. The distribution system is often switched Ethernet, but it can use any technology. As the clients move, they may find another access point with a better signal than the one they are currently using and change their association.

Finally, there is the problem of security. Since wireless transmissions are broadcast, it is easy for nearby computers to receive packets of information that were not intended for them. To prevent this, the 802.11 standard included an encryption scheme known as WEP (wired equivalent privacy). The idea was to make wireless security like that of wired security. It is a good idea, but unfortunately the scheme was flawed and soon broken. It has since been replaced with newer schemes that have different cryptographic details in the 802.11i standard, also called Wi-Fi Protected Access(WPA).

802.11 has caused a revolution in wireless networking that is set to continue. Beyond buildings, it is starting to be installed in trains, planes, boats, and automobiles so that people can surf the Internet wherever they go. Mobile phones and all manner of consumer electronics, from game consoles to digital cameras, can communicate with it.

Questions:

1. How is an 802.11 network configured?
2. Which mode is more popular, ad hoc mode or access point mode?
3. What does term *multipath fading* mean in this article?
4. What problem does 802.11 use CSMA to handle?
5. What does the acronym *WPA* stand for?

Passage 3 Mechatronic Systems

Both digital and analog electronics are used in the control of various mechanical systems. The interdisciplinary field that comprises both mechanical and electronic components is known as *mechatronics*. Figure 4.8 is a basic block diagram of a mechatronic system.

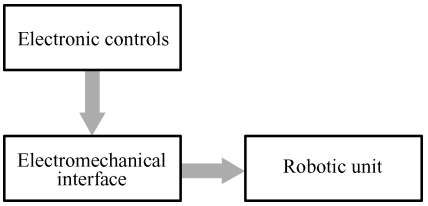


Figure 4.8 The block diagram of a mechatronic system

Mechatronic systems are found in homes, industry, and transportation. Most home appliances consist of both mechanical and electronic components. Electronics controls the operation of a washing machine in terms of water flow, temperature, and type of cycle. Manufacturing industries rely heavily on mechatronics for process control and assembly.

In automotive and other types of manufacturing, robotic arms perform precision welding, painting, and other functions on the assembly line. Automobiles themselves are mechatronic machines; a digital computer controls functions such as braking, engine parameters, fuel flow, safety features, and monitoring.

A simple robotic arm is shown in Figure 4.9, which is used on an automotive assembly line. The movement of the arm in any quadrant and to any specified position is accomplished with some type of digital control such as a microcontroller.

Figure 4.9 A simple robotic arm

The journey of mechatronics as a field began in the late 1960s, emerging as an intersection of mechanics and electronics. The term *mechatronics* itself was first coined by Tetsuro Mori, an engineer at Yaskawa Electric Corporation, to describe the integration of these disciplines in the development of industrial products. Initially, it was a marketing term for the seamless blend of mechanical and electronic engineering, but it quickly evolved into a recognized engineering discipline.

In its early stages, mechatronics was primarily focused on adding electronic control to mechanical systems, enhancing their functionality and precision. The advent of microprocessors and advancements in digital electronics in the 1970s and 1980s propelled mechatronics into a new era. This period saw the introduction of more sophisticated control systems and feedback mechanisms, allowing for greater automation and precision in mechanical systems.

The 1990s and 2000s marked a period of rapid growth for mechatronics, fueled by leaps in computer technology and software. These advancements expanded the scope of mechatronics beyond traditional fields, incorporating elements of computer engineering, software development, and systems design. This era saw the birth of many modern mechatronic systems, such as advanced robotics, automated manufacturing processes, and intelligent vehicle systems.

Today, mechatronics stands at the forefront of technological innovation, integrating cutting-edge developments in artificial intelligence, machine learning, and the Internet of Things (IoT). It plays a critical role in developing smart systems and technologies that are reshaping industries, healthcare, and daily life.

Questions:

1. Who first coined the term *mechatronics* ? And what does it mean?
2. What components does a basic mechatronic system comprise according to Figure 4.8?
3. What is the object shown in Figure 4.9? And what function does it have?
4. Does mechatronics play a critical role in developing today's smart systems or not?
5. What does the acronym IoT stand for?

Unit 5

Modern Digital Design

Lesson 13 Electronic Design Automation

Lesson 14 FPGAs

Lesson 15 VHDL

Passage 1 The PLD Evolution

Passage 2 Verilog HDL

Passage 3 SoC

Lesson 13 Electronic Design Automation

Electronic circuit design has traditionally fallen into two main areas: analogue and digital. These subjects are usually taught separately, and electronics engineers tend to specialize in one area. Within these two groupings there are further specializations, such as radio frequency analogue design; digital integrated circuit design; and, where the two domains meet, mixed-signal design. In addition, of course, software engineering plays an increasingly important role in embedded systems.

Digital electronics is ever more significant in consumer goods. Cars have sophisticated control systems. Many homes now have personal computers. Products that used to be thought of as analog, such as radio, television and telephones, are all becoming digital. Digital compact discs have almost entirely replaced analog LPs for recorded audio. With these changes, the lifetimes of products have lessened. In a period of less than a year, new models will probably have replaced all the digital electronic products in your local store.

Electronic Design Automation

To keep pace with this rapid change, electronics products have to be designed extremely quickly. Analog design is still a specialized (and well-paid) profession. Digital design has become very dependent on computer-aided design (CAD) — also known as electronic design automation (EDA). The EDA tools allow two tasks to be performed: synthesis, in other words the translation of a specification into an actual implementation of the design; and simulation, in which the specification or the detailed implementation can be exercised in order to verify correct operation [1].

Synthesis and simulation EDA tools require that the design be transferred from the designer's imagination into the tools themselves. This can be done by drawing a diagram of the design using a graphical package. This is known as schematic capture. Alternatively, the design can be represented in a textual form, much like a software program. Textual descriptions of digital hardware can be written in a modified programming language, such as C, or in a hardware-description language (HDL). Over the past fifty years or so, a number of HDLs have been designed. Two HDLs are in common usage today: Verilog and VHDL (VHSIC hardware-description language, where VHSIC stands for very-high-speed integrated circuit). Standard HDLs are important because they can be used by different CAD tools from different tool vendors. In the days before Verilog and VHDL, every tool had its own HDL,

requiring laborious translation between HDLs, for example, to verify the output from a synthesis tool with another vendor's simulator.

Logic Gates

The basic building blocks of digital circuits are gates. A gate is an electronic component with a number of inputs and, generally, a single output. The inputs and the outputs are normally in one of two states: logic 0 or logic l. These logic values are represented by voltages (for instance, 0 V for logic 0 and 3 3 V for logic 1) or currents. The gate itself performs a logical operation using all of its inputs to generate the output. Ultimately, of course, digital gates are really analog components, but for simplicity we tend to ignore their analog nature.

It is possible to buy a single integrated circuit containing, say, four identical gates, as shown in Figure 5.1. (Note that two of the connections are for the positive and negative power supplies to the device. These connections are not normally shown in logic diagrams.) A digital system could be built by connecting hundreds of such devices together — indeed many systems have been designed in that way. Although the individual integrated circuits might cost as little as 10 cents each, the cost of designing the printed circuit board for such a system and the cost of assembling the board are very significant and this design style is no longer cost-effective.

Much more complicated functions are available as mass-produced integrated circuits, ranging from flip-flops through to microprocessors. With increasing complexity comes flexibility—a microprocessor can be programmed to perform a near-infinite variety of tasks. Digital system design therefore consists, in part, of taking standard components and connecting them together. Inevitably, however, some aspect of the functionality will not be available as a standard device. The designer is then left with the choice of implementing this functionality from discrete gates or of designing a specialized integrated circuit to perform that task. While this latter task may appear daunting, it should be remembered that the cost of a system will depend to a great extent not on the cost of the individual components but on the cost of connecting those components together.

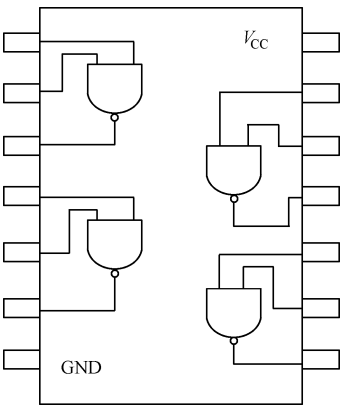


Figure 5.1 A small-scale integrated circuit

ASICs and FPGAs

The design of a high-performance, full-custom integrated circuit (IC) is, of course, a difficult task. In full-custom IC design, everything, down to and including individual transistors may be designed (although libraries of parts are, of course, used). For many

years, however, it has been possible to build semi-custom integrated circuits using gate arrays. A gate array, as its name suggests, is an integrated circuit on which an array of logic gates has been created. The design of an application-specific integrated circuit (ASIC) using a gate array therefore involves the definition of how the gates in the array should be connected. In practical terms, this means that one or two layers of metal interconnect must be designed. Since an integrated circuit requires seven or more processing stages, all the processing steps other than the final metallization can be completed in advance. Because the uncommitted gate arrays can be produced in volume, the cost of each device is relatively small.

The term ASIC is often applied to full-custom and semi-custom integrated circuits. Another class of integrated circuit is that of programmable logic. The earliest programmable logic devices (PLDs) were programmable logic arrays (PLAs). Like gate arrays, these consist of arrays of uncommitted logic, but unlike mask-programmable gate arrays, the configuration of the array is determined by applying a large (usually negative) voltage to individual connections. The general structure of a PLA is shown in Figure 5.2. The PLA has a number of inputs (A, B, C) and outputs (X, Y, Z), an AND-plane and an OR-plane. Connections between the inputs and the product terms (P, Q, R, S) and between the product terms and outputs are shown; the remaining connections have been removed as part of the programming procedure. Some PLAs may be reprogrammed electrically or by restoring the connections by exposing the device to ultraviolet light. PALs (programmable array logic) extend the idea of PLAs to include up to 12 flip-flops. In recent years, programmable devices have become much more complex and include CPLDs (complex PLDs) and FPGAs (field-programmable gate arrays).

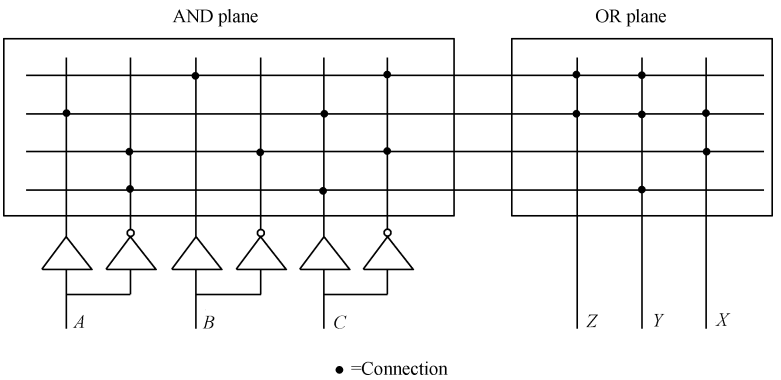


Figure 5.2 PLA structure

Design Flow

Most digital systems are sequential, that is they have states, and the outputs depend on the present state. Some early designs of computer were asynchronous; in other words, the transition to a new state happened as soon as inputs had stabilized. For many years, digital

systems have tended to be synchronous. In a synchronous system, the change of state is triggered by one or more clock signals. In order to design reliable systems, formal design methodologies have been defined. The design of a synchronous sequential digital system using discrete gates would therefore proceed as follows.

1. Write a specification.
2. Partition the design into smaller parts and write a specification for each part (when it is necessary).
3. From the specification draw a state machine chart. This shows each state of the system and the input conditions that cause a change of state, together with the outputs in each state.
4. Minimize the number of states. This is optional and may not be useful in all cases.
5. Assign Boolean[2] variables to represent each state.
6. Derive next state and output logic.
7. Optimize the next state and output logic. It is to minimize the number of gates needed.
8. Choose a suitable placement for the gates (in terms of which gates share integrated circuits and in terms of where each integrated circuit is placed on the printed circuit board).
9. Design the routing between the integrated circuits.

In general, steps 1 and 2 cannot be avoided. This is where the creativity of the designer is needed. Most books on digital design concentrate on steps 3 to 7. Steps 8 and 9 can be performed manually, but placement and routing was one of the first tasks to be successfully automated. It is possible to simulate the design at different stages if it is converted into a computer-readable form. Typically, in order to perform the placement and routing, a schematic capture program would be used at around step 7, such that the gate-level structure of the circuit would be entered. This schematic could be converted to a form suitable for a logic simulator. After step 9 had been completed, the structure of the circuit, including any delays generated by the resistance and capacitance of the interconnect could be extracted and again simulated.

The implementation of digital designs on ASICs or FPGAs therefore involves the configuration of connections between predefined logic blocks. As noted, we cannot avoid steps 1 and 2, above and steps 8 and 9 can be done automatically. The use of an HDL means that the design can be entered into a CAD system and simulated at step 3 or 4, rather than step 7. So-called register transfer level (RTL) synthesis tools automate steps 6 and 7. Step 4 still has to be done by hand. Step 5 can be automated, but now the consequences of a particular state assignment can be assessed very quickly. Behavioral synthesis tools are starting to appear that automate the process from about step 2 onwards. Figure 5.3 shows the overall design flow for RTL synthesis-based design.

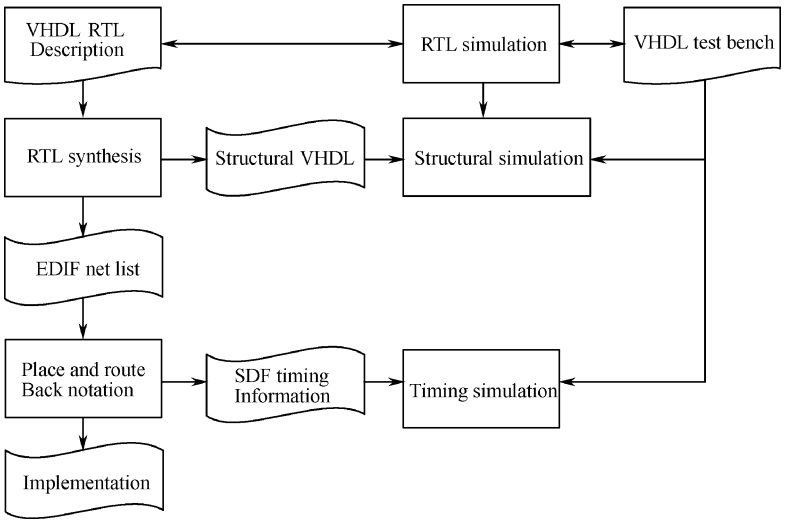


Figure 5.3 The overall design flow for RTL synthesis-based design

New Words & Expressions

lifetime [ˈlaɪftaɪm] *n*. 寿命，使用期
lessen [ˈlesən] *vt*. 减少
specialized [ˈspeʃəˈlaɪzd] *adj*. 专业的
well-paid [ˈwel ˈpeɪd] *adj*. 高收入的
profession [prəˈfeʃn] *n*. 职业；专业
diagram [ˈdaɪəgræm] *n*. 图表
textual [ˈtekstjʊəl] *adj*. 文本的
vendor [ˈvendə] *n*. 卖主
laborious [ləˈbɔːrɪəs] *adj*. 费劲的
cost-effective [ˈkɒst ɪˈfektɪv] *adj*. 合算的
in practical terms 实际上
mass-produced [ˈmæs prəˈdjuːst] *adj*. 量产的
inevitably [ɪˈnevɪtəbli] *adv*. 无法避免地
daunting [ˈdɔːntɪŋ] *adj*. 令人畏缩的
in terms of ... 在……方面
uncommitted [ʌnkəˈmɪtɪd] *adj*. 未确定用途的
methodology [me əˈdɒlədʒi] *n*. 方法学
specification [ˌspesɪfɪˈkeɪʃn] *n*. 规格
partition [pɑːˈtɪʃən] *n*. 分割
placement [ˈpleɪsmənt] *n*. 布置

Technical Terms

synthesis [ˈsɪn ɪsɪs] *n*. 综合
simulation [ˌsɪmjʊˈleɪʃn] *n*. 模拟
schematic [skiːˈmætɪk] *n*. 原理图
capture [ˈkæptʃə] *vt*. 输入，记录
metallization [ˌmetəlaɪˈzeɪʃən] *n*. 金属化
sequential [sɪˈkwenʃəl] *adj*. 时序的
asynchronous [əˈsɪŋkrənəs] *adj*. 异步的
synchronous [ˈsɪŋkrənəs] *adj*. 同步的
trigger [ˈtrɪgə] *vt*. 触发

discrete [dɪˈskriːt] *adj*. 离散的，分立的
embedded system　嵌入式系统
design flow　设计流程
present state　现态
next state　次态
CD *abbr*. compact disc　激光唱盘
LP *abbr*. long playing　密纹唱片
CAD *abbr*. computer-aided design　计算机辅助设计
EDA *abbr*. electronic design automation　电子设计自动化
HDL *abbr*. hardware-description language　硬件描述语言
ASIC *abbr*. application-specific integrated circuit　专用集成电路
PLD *abbr*. programmable logic device　可编程逻辑器件
PLA *abbr*. programmable logic array　可编程逻辑阵列
FPGA *abbr*. field programmable gate array　现场可编程门阵列
RTL *abbr*. register transfer level　寄存器传输级

Notes

1. “综合”(synthesis)是将设计指标转化为实际的设计实现，而“模拟”(simulation)是对设计指标或详细的实现方案进行“演习”以确保其运行正确。
2. 乔治·布尔(George Boole，1815－1864)是英国数学家和哲学家，“布尔代数”(Boolean Algebra)的发明人。

Lesson 14　FPGAs

What Are FPGAs?

Field-programmable gate arrays (FPGAs)[1] are digital integrated circuits (ICs) that contain configurable (programmable) blocks of logic along with configurable interconnects between these blocks. Design engineers can configure (program) such devices to perform a tremendous variety of tasks. See Figure 5.4.

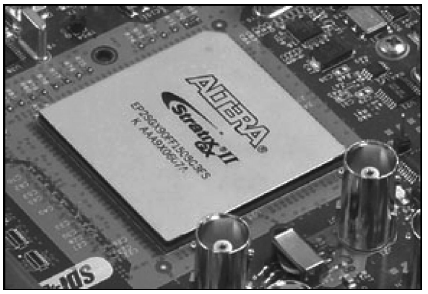


Figure 5.4　Altera Stratix Ⅱ GX FPGA

Depending on the way in which they are implemented, some FPGAs may only be programmed a single time, while others may be reprogrammed over and over again. Not surprisingly, a device that can be

programmed only one time is referred to as one-time programmable (OTP).

The "field-programmable" portion of the FPGA's name refers to the fact that its programming takes place "in the field" (as opposed to devices whose internal functionality is hardwired by the manufacturer). This may mean that FPGAs are configured in the laboratory, or it may refer to modifying the function of a device resident in an electronic system that has already been deployed in the outside world. If a device is capable of being programmed while remaining resident in a higher-level system, it is referred to as being in-system programmable (ISP).

Why Are FPGAs of Interest?

There are many different types of digital ICs, including "jelly-bean logic" (small components containing a few simple, fixed logical functions), memory devices, and microprocessors (μPs). Of particular interest to us here, however, are programmable logic devices (PLDs), application-specific integrated circuits (ASICs), application-specific standard products (ASSPs), and FPGAs.

We should be aware that PLDs are devices whose internal architecture is predetermined by the manufacturer, but which are created in such a way that they can be configured (programmed) by engineers in the field to perform a variety of different functions. In comparison to an FPGA, however, these devices contain a relatively limited number of logic gates, and the functions they can be used to implement are much smaller and simpler.

At the other end of the spectrum are ASICs and ASSPs, which can contain hundreds of millions of logic gates and can be used to create incredibly large and complex functions. ASICs and ASSPs are based on the same design processes and manufacturing technologies. Both are custom-designed to address a specific application, the only difference being that an ASIC is designed and built to order for use by a specific company, while an ASSP is marketed to multiple customers.

Although ASICs offer the ultimate in size (number of transistors), complexity, and performance; designing and building one is an extremely time-consuming and expensive process, with the added disadvantage that the final design is "frozen in silicon" and cannot be modified without creating a new version of the device.

Thus, FPGAs occupy a middle ground between PLDs and ASICs because their functionality can be customized in the field like PLDs, but they can contain millions of logic gates and be used to implement extremely large and complex functions that previously could be realized only using ASICs.

The cost of an FPGA design is much lower than that of an ASIC (although the ensuing ASIC components are much cheaper in large production runs). At the same time, implementing design changes is much easier in FPGAs, and the time to market for such designs is much faster. Thus, FPGAs make a lot of small, innovative design companies viable because — in addition to their use by large system design houses — FPGAs facilitate "Fred-in-

the-shed" — type operations. This means they allow individual engineers or small groups of engineers to realize their hardware and software concepts on an FPGA-based test platform without having to incur the enormous nonrecurring engineering (NRE) costs or purchase the expensive toolsets associated with ASIC designs.

What Can FPGAs Be Used for?

When they first arrived on the scene in the mid-1980s, FPGAs were largely used to implement glue logic[2], medium complexity state machines, and relatively limited data processing tasks. During the early 1990s, as the size and sophistication of FPGAs started to increase, their big markets at that time were in the telecommunications and networking arenas, both of which involved processing large blocks of data and pushing that data around. Later, toward the end of the 1990s, the use of FPGAs in consumer, automotive, and industrial applications underwent a humongous growth spurt.

FPGAs are often used to prototype ASIC designs or to provide a hardware platform on which to verify the physical implementation of new algorithms. However, their low development cost and short time-to-market mean that they are increasingly finding their way into final products (some of the major FPGA vendors actually have devices that they specifically market as competing directly against ASICs).

By the early-2000s, high-performance FPGAs containing millions of gates had become available. Some of these devices feature embedded microprocessor cores, high-speed input/output (I/O) interfaces, and the like. The end result is that today's FPGAs can be used to implement just about anything, including communications devices and software-defined radios (SDR); radar, image, and other digital signal processing (DSP) applications; all the way up to system-on-chip (SoC)[3] components that contain both hardware and software elements.

To be just a tad more specific, FPGAs are currently eating into four major market segments: ASIC and custom silicon, DSP, embedded microcontroller applications, and physical layer communication. Furthermore, FPGAs have created a new market in their own right: reconfigurable computing (RC).

- **ASIC and custom silicon** — Today's FPGAs are increasingly being used to implement a variety of designs that could previously have been realized using only ASICs and custom silicon.
- **Digital signal processing** — High-speed DSP has traditionally been implemented using specially tailored microprocessors called digital signal processors (DSP). However, today's FPGAs can contain embedded multipliers, dedicated arithmetic routing, and large amounts of on-chip RAM, all of which facilitate DSP operations. When these features are coupled with the massive parallelism provided by FPGAs, the result is to outperform the fastest DSP chips by a factor of 500 or more.
- **Embedded microcontrollers** — Small control functions have traditionally been handled by special-purpose embedded processors called microcontrollers. These low-cost devices

contain on-chip program and instruction memories, timers, and I/O peripherals wrapped around a processor core. FPGA prices are falling, however, and even the smallest devices now have more than enough capability to implement a soft processor core combined with a selection of custom I/O functions. The end result is that FPGAs are becoming increasingly attractive for embedded control applications.

- **Physical layer communications** — FPGAs have long been used to implement the glue logic that interfaces between physical layer communication chips and high-level networking protocol layers. The fact that today's high-end FPGAs can contain multiple high-speed transceivers means that communications and networking functions can be consolidated into a single device.
- **Reconfigurable computing** — This refers to exploiting the inherent parallelism and reconfigurability provided by FPGAs to "hardware accelerate" software algorithms. Various companies are currently building huge FPGA-based reconfigurable computing engines for tasks ranging from hardware simulation to cryptography analysis to discovering new drugs.

New Words & Expressions

spectrum [ˈspektrəm] *n*. 范围
address [əˈdres] *vt*. 应对
be referred to as ... 被称作
henceforth [hensˈfɔː] *adv*. 此后
interpretation [ɪntɜːprɪˈteɪʃn] *n*. 解释
inconsistent [ɪnkənˈsɪstənt] *adj*. 不一致的
ultimate [ˈʌltɪmət] *adj*. 最终的
in the field 在现场
disadvantage [dɪsədˈvɑːntɪdʒ] *n*. 缺点
ensue [ɪnˈsjuː] *vi*. 继起
viable [ˈvaɪəbl] *adj*. 可行的
innovative [ˈɪnəvətɪv] *adj*. 创新的
shed [ʃed] *n*. 小屋
as apposed to ... 和 相反
incur [ɪnˈkɜː] *vt*. 招致
guesstimate [ˈgestəmɪt] *n*. 大概估计
nonrecurring [ˈnʌn rɪˈkɜːɪŋ] *adj*. 一次性的
arena [əˈriːnə] *n*. 竞技场
humongous [hjuːˈmʌŋgəs] *adj*. 巨大的
for the nonce 目前
spurt [spɜːt] *vi*. 迸发
tailor [ˈteɪlə] *vt*. 修改
dedicate [ˈdedɪkeɪt] *vt*. 致力于
facilitate [fəˈsɪlɪteɪt] *vt*. 使 容易
in one's own right 依靠自身素质
couple [ˈkʌpl] *vt*. 结合
wrap [ræp] *vt*. 紧围
consolidate [kənˈsɒlɪdeɪt] *vt*. 合并

Technical Terms

parallelism [ˈpærəlelɪzəm] *n*. 并行度
prototype [ˈprəʊtətaɪp] *n*. 样机
cryptography [krɪpˈtɒgrəfi] *n*. 密码学
state machine 状态机

data processing　数据处理
OTP *abbr*. one-time programmable　一次性编程
ISP *abbr*. in-system programmable　在系统可编程
μP *abbr*. microprocessor　微处理器
PLD *abbr*. programmable logic device　可编程逻辑器件
ASIC *abbr*. application-specific integrated circuit　专用集成电路
ASSP *abbr*. application-specific standard product　专用标准产品
NRE *abbr*. nonrecurring engineering　一次性工程
SDR *abbr*. software-defined radio　软件无线电
DSP *abbr*. digital signal processing　数字信号处理
SoC *abbr*. system-on-chip　片上系统
RC *abbr*. reconfigurable computing　可重配计算

Notes

1. 本课出现了大量“首字母缩略语”(acronyms)，其发音大多是将这些字母逐一读出。例如，FPGA 的发音就是“F-P-G-A”。IC、OTP、ISP、PLD、SPLD、CPLD、ASSP、NRE、I/O、SoC、RC、DSP 的发音都是逐一读出字母即可。μP 中的 μ 是希腊字母，发音为[mjuː]，含义为“微小的”(相当于英语前缀 micro-)。例外情况有：(1)完全符合英语发音规则的，要依照规则发音(例如，RAM 读作[ræm])；(2)部分符合英语发音规则的，该部分要依照规则发音，余者发字母音(例如，ASIC 读作[ˈeɪsɪk])。
2. “胶连逻辑”(glue logic)是指那些用来连接大型逻辑块、逻辑功能或逻辑器件的小型简易逻辑电路。
3. “片上系统”(system-on-chip，SoC)是指在单一芯片上实现的完整电子系统。不过，现实情况中还是无法避免使用片外的附加部件。

Lesson 15　VHDL

What Is VHDL?

VHDL is a programming language that has been designed and optimized for describing the behavior of digital systems. VHDL has many features appropriate for describing the behavior of electronic components ranging from simple logic gates to complete microprocessors and custom chips. Features of VHDL allow electrical aspects of circuit behavior (such as rise and fall times of signals, delays through gates, and functional operation) to be precisely described. The resulting VHDL simulation models can then be used as building blocks in larger circuits (using schematics, block diagrams or system-level VHDL

descriptions) for the purpose of simulation.

VHDL is also a general-purpose programming language; just as high-level programming languages allow complex design concepts to be expressed as computer programs, VHDL allows the behavior of complex electronic circuits to be captured into a design system for automatic circuit synthesis or for system simulation. Like Pascal, C and C++, VHDL includes features useful for structured design techniques, and offers a rich set of control and data representation features. Unlike these other programming languages, VHDL provides features allowing concurrent events to be described. This is important because the hardware described using VHDL is inherently concurrent in its operation.

One of the most important applications of VHDL is to capture the performance specification for a circuit, in the form of what is commonly referred to as a test bench. Test benches are VHDL descriptions of circuit stimuli and corresponding expected outputs that verify the behavior of a circuit over time. Test benches should be an integral part of any VHDL project and should be created in tandem with other descriptions of the circuit.

A Standard Language

One of the most compelling reasons for you to become experienced with and knowledgeable in VHDL is its adoption as a standard in the electronic design community. Using a standard language such as VHDL virtually guarantees that you will not have to throw away and recapture design concepts simply because the design entry method you have chosen is not supported in a newer generation of design tools. Using a standard language also means that you are more likely to be able to take advantage of the most up-to-date design tools and that you will have access to a knowledge base of thousands of other engineers, many of whom are solving problems similar to your own.

A Brief History of VHDL

VHDL, which stands for VHSIC (Very-High-Speed Integrated Circuit) Hardware-Description Language, was developed in the early 1980s as a spin-off of a high-speed integrated circuit research project funded by the U.S. Department of Defense. During the VHSIC program, researchers were confronted with the daunting task of describing circuits of enormous scale (for their time) and of managing very large circuit design problems that involved multiple teams of engineers. With only gate-level design tools available, it soon became clear that better, more structured design methods and tools would be needed.

To meet this challenge, a team of engineers from three companies (IBM, Texas Instruments and Intermetrics) were contracted by the Department of Defense to complete the specification and implementation of a new, language-based design description method. The first publicly available version of VHDL, version 7.2, was released in 1985. In 1986, the Institute of Electrical and Electronics Engineers (IEEE)[1] was presented with a proposal to standardize the language, which it did in 1987 after substantial enhancements and modifications were

made by a team of commercial, government and academic representatives. The resulting standard, IEEE 1076-1987, is the basis for virtually every simulation and synthesis product sold today. An enhanced and updated version of the language, IEEE 1076-1993, was released in 1994, and VHDL tool vendors have been responding by adding these new language features to their products.

Although IEEE 1076 defines the complete VHDL, there are aspects of the language that make it difficult to write completely portable design descriptions (descriptions that can be simulated identically using different vendors' tools). The problem stems from the fact that VHDL supports many abstract data types, but it does not address the simple problem of characterizing different signal strengths or commonly used simulation conditions such as unknowns and high-impedance.

Soon after IEEE 1076-1987 was adopted, simulator companies began enhancing VHDL with new, non-standard types to allow their customers to accurately simulate complex electronic circuits. This caused problems because design descriptions entered into one simulator were often incompatible with other simulation environments. VHDL was quickly becoming a non-standard.

To get around the problem of non-standard data types, another standard was developed by an IEEE committee. This standard, numbered 1164, defines a standard package (a VHDL feature that allows commonly used declarations to be collected into an external library) containing definitions for a standard nine-valued data type [2]. This standard data type is called std_logic, and the IEEE 1164 package is often referred to as the standard logic package.

The IEEE 1076-1987 and IEEE 1164 together form the complete VHDL standard in widest use today. (IEEE 1076-1993 is slowly working its way into the VHDL mainstream, but it does not add significant new features for synthesis users.)

IEEE 1076.3 (often called the Numeric Standard or Synthesis Standard) defines standard packages and interpretations for VHDL data types as they relate to actual hardware. This standard, which was released at the end of 1995, is intended to replace the many custom (non-standard) packages that vendors of synthesis tools have created and distributed with their products.

IEEE 1076.3 does for synthesis users what IEEE 1164 did for simulation users: increase the power of IEEE 1076, while at the same time ensuring compatibility between different vendors' tools. IEEE 1076.3 includes, among other things:

- A documented hardware interpretation of values belonging to the bit and boolean types defined by IEEE 1076, as well as interpretations of the std_ulogic type defined by IEEE 1164.
- A function that provides "don't care" or "wild card" testing of values based on the std_ulogic type. This is of particular use for synthesis, since it is often helpful to express logic in terms of "don't care" values.
- Definitions for standard signed and unsigned arithmetic data types, along with

arithmetic, shift, and type conversion operations for those types.

The annotation of timing information to a simulation model is an important aspect of accurate digital simulation. The VHDL 1076 standard describes a variety of language features that can be used for timing annotation. However, it does not describe a standard method for expressing timing data outside of the timing model itself.

The capability to separate the behavioral description of a simulation model from the timing specifications is important for many reasons. One of the major strengths of Verilog HDL (VHDL's closest rival) is the fact that Verilog HDL includes a feature specifically intended for timing annotation. This feature, the Standard Delay Format, or SDF, allows timing data to be expressed in a tabular form and included into the Verilog timing model at the time of simulation.

The IEEE 1076.4, published by the IEEE in late 1995, adds this capability to VHDL as a standard package. A primary impetus behind this standard effort was to make it easier for ASIC vendors and others to generate timing models applicable to both VHDL and Verilog HDL. For this reason, the underlying data formats of IEEE 1076.4 and Verilog's SDF are quite similar.

When Should You Use VHDL?

Why choose to use VHDL for your design efforts? There are many likely reasons. If you ask most VHDL tool vendors this question, the first answer you will get is, "It will improve your productivity." But just what does this mean? Can you really expect to get your projects done faster using VHDL than by using your existing design methods?

The answer is yes, but probably not the first time you use it, and only if you apply VHDL in a structured manner. VHDL (like a structured software design language) is most beneficial when you use a structured, top-down approach to design. Real increases in productivity will come later, when you have climbed higher on the VHDL learning curve and have accumulated a library of reusable VHDL components.

Productivity increases will also occur when you begin to use VHDL to enhance communication between team members and when you take advantage of the more powerful tools for simulation and design verification that are available. In addition, VHDL allows you to design at a more abstract level. Instead of focusing on a gate-level implementation, you can address the behavioral function of the design.

How will VHDL increase your productivity? By making it easy to build and use libraries of commonly-used VHDL modules. VHDL makes design reuse feel natural. As you discover the benefits of reusable code, you will soon find yourself thinking of ways to write your VHDL statements in ways that make them general purpose. Writing portable code will become an automatic reflex.

Another important reason to use VHDL is the rapid pace of development in electronic design automation (EDA) tools and in target technologies. Using a standard language such as

VHDL can greatly improve your chances of moving into more advanced tools (for example, from a basic low-cost simulator to a more advanced one) without having to re-enter your circuit descriptions. Your ability to retarget circuits to new types of device targets (for example, ASICs, FPGAs, and complex PLDs) will also be improved by using a standard design entry method.

New Words & Expressions

appropriate [ə'prəʊprɪət] *adj*. 适当的
integral ['ɪntɪgrəl] *adj*. 整体的
tandem ['tændəm] *adj*. 串联的
in tandem with ... 与……前后相连
compelling [kəm'pelɪŋ] *adj*. 强制的
virtually ['vɜːtjʊəli] *adv*. 实质上
reusable [rɪ'juzəbl] *adj*. 可重复使用的
proposal [prə'pəʊzəl] *n*. 提议，建议
beneficial [benɪ'fɪʃəl] *adj*. 有益的
reflex ['riːfleks] *n*. 反射
have access to ... 可以使用……
enhancement [ɪn'hɑːnsmənt] *n*. 增强
modification [ˌmɒdɪfɪ'keɪʃn] *n*. 修改
unknown [ʌn'nəʊn] *n*. 未知事物
get around ... 避开……
work one s way into ... 努力实现……
mainstream ['meɪnstriːm] *n*. 主流
variety [və'raɪəti] *n*. 种类
productivity [prɒdʌk'tɪvɪti] *n*. 工作效率
tabular ['tæbjələ] *adj*. 表格式的

Technical Terms

concurrent [kən'kʌrənt] *adj*. 并发的
stimuli ['stɪmjʊlaɪ] *n*. 激励源
impedance [ɪm'piːdəns] *n*. 阻抗
package ['pækɪdʒ] *n*. 程序包
declaration [deklə'reɪʃn] *n*. 声明
annotation [ænə'teɪʃn] *n*. 标注
timing ['taɪmɪŋ] *n*. 时序
statement ['steɪtmənt] *n*. 语句
test bench 测试台
high impedance 高阻
don t care 无关项
wild card 通配符
type conversion 类型转换
top-down approach “自顶而下”设计法
learning curve 学习曲线
HDL *abbr*. hardware-description language 硬件描述语言
IEEE *abbr*. Institute of Electrical and Electronic Engineers 电气电子工程师学会
EDA *abbr*. electronic design automation 电子设计自动化

ASIC *abbr*. application-specific integrated circuit　专用集成电路
FPGA *abbr*. field-programmable gate array　现场可编程门阵列
PLD *abbr*. programmable logic device　可编程逻辑器件

Notes

1. “电气电子工程师学会”(Institute of Electrical and Electronics Engineers，IEEE)是一个国际性学术组织。1963 年，美国“电气工程师学会”和“无线电工程师学会”合并组成 IEEE。IEEE 总部设在纽约。
2. feature 和 characteristic 均可以表示“特征”“特性”“特点”“特色”。characteristic 更强调内在本质，feature 更强调外观表象。此句可译为：该标准(标准号为 1164)定义了一个包含一个标准 9 值数据类型定义的标准包(标准包是 VHDL 的特色之一，即允许将公用数据类型声明集中于外部库函数中)。

Exercises

1. Fill in the blanks with proper words，phrases or clauses.

(1)_____(随着 20 世纪 80 年代超大规模集成技术的到来)engineers began to realize the advantages of designing an IC that was customized or tailored to a particular system or application rather than using _____(标准集成电路)alone. Microelectronic system design then becomes a matter _____ defining the functions that you can _____(实现)using standard ICs and then implementing the remaining _____(逻辑功能)(sometimes called glue logic) with one or more _____(定制集成电路). As VLSI became possible，you could build a system from _____(数量更少的部件)by combining many standard ICs _____ a few custom ICs. Building a microelectronic system with fewer ICs allows you to _____(降低成本和改进可靠性).

Of course，there are many situations _____ which it is not appropriate _____ use a custom IC for each and every part of an microelectronic system._____(如果您需要大容量的存储器)，for example，it is still best to use standard memory ICs，either DRAM or SRAM，_____ conjunction with custom ICs.

One of the first conferences to be devoted _____ this rapidly emerging segment of the IC industry was the IEEE Custom Integrated Circuits Conference (CICC)，and the proceedings of this annual conference form a useful reference _____ the development of custom ICs. As different types of custom ICs began to evolve for different types of applications，these new ICs gave rise to a new term：application-specific IC，or ASIC. Now we have the IEEE International ASIC Conference，which tracks advances in ASICs separately from other types of custom ICs.

(2) PLD (programmable logic device) is an umbrella term for _____(一类芯片) that are programmable at the customer's site (_____ this case，the customer is the circuit

developer, not the end user). There are three physical structures.

The first is the permanent fuse type which blows apart lines or fuses them together _____ electrically melting an aluminum trace or insulator. This was the first type of PLD, _____ "programmable array logic" or PAL.

The second is reprogrammable and uses EEPROM or flash memory. It causes a transistor _____ open or close _____ (依据) the contents of its associated memory cell.

The third type is RAM-based, _____ makes it dynamic and volatile. Its contents are loaded _____(每次) it starts up. This control memory is not a self-contained memory-bank-like one that is used to input data. It is a separate masking layer in the chip, and each programmable point is connected _____ its counterpart bit in this layer.

CPLDs (complex PLDs) and FPGAs (field-programmable gate arrays) are _____(最常见的可编程逻辑器件芯片). CPLDs are mostly EEPROM and flash-based and are reprogrammable. FPGAs use all three methods.

_____(和门阵列不同), which require the final masking fabrication process, PLDs are easily programmable _____(在现场). PLDs are always used for logical functions, but _____(可编程存储芯片) such as PROMs and EPROMs might also be considered PLDs if they contain program code _____ just data.

2. Translate the following passages into Chinese or English.

(1) Some digital logic ICs and their analog counterparts (analog-to-digital converters, for example) are standard parts, or standard ICs. You can select standard ICs from catalogs and data books and buy them from distributors. Systems manufacturers and designers can use the same standard part in a variety of different microelectronic systems (systems that use microelectronics or ICs).

(2) A modern submicron CMOS process is now just as complicated as a submicron bipolar or BiCMOS (a combination of bipolar and CMOS) process. However, CMOS ICs have established a dominant position, are manufactured in much greater volume than any other technology, and therefore, because of the economy of scale, the cost of CMOS ICs is less than a bipolar or BiCMOS IC for the same function. Bipolar and BiCMOS ICs are still used for special needs. For example, bipolar technology is generally capable of handling higher voltages than CMOS. This makes bipolar and BiCMOS ICs useful in power electronics, cars, telephone circuits, and so on.

(3) An ASIC (application-specific integrated circuit) is a chip that is custom designed for a specific application rather than a general-purpose chip such as a microprocessor. The use of ASICs improve performance over general-purpose CPUs, because ASICs are "hardwired" to do a specific job and do not incur the overhead of fetching and interpreting stored instructions. An ASIC chip performs an electronic operation as fast as it is possible to do so, providing, of course, that the circuit design is efficiently architected.

(4) A soft core is a block of logic for a particular function that is designed to be implemened in a programmable logic chip (PLD) or on a programmable section of a

microcontroller chip or system-on-chip (SoC). Depending on the vendor, the logic can come as a schematic, netlist or HDL code. Soft cores may be implemented with hard cores on a chip.

(5) A hard core is a block of logic for a particular function on a chip that is designed at the circuit level. Microprocessors are typically hard cores, which ties them to a particular foundry's semiconductor process. Microcontrollers or SoCs may be entirely hard cores or made up of hard (hardwired) cores and soft (programmable) cores.

(6)“电子设计自动化”利用计算机对芯片上的电子线路进行功能设计和仿真。

(7)“硬件描述语言”是一种描述电子线路功能的语言，它用于编制文档、进行仿真或逻辑综合。

(8)“逻辑综合”是将电子线路的高级描述转换为一张逻辑门及其互连列表(所谓“网表”)的过程。每个逻辑综合程序都理解 Verilog 和 VHDL 的某个子集。

(9)“复杂可编程逻辑器件”(CPLD)是一种在逻辑模块之间具有可编程互连的可编程逻辑器件。多数 CPLD 是基于“电可擦写可编程只读存储器”(EEPROM)和闪存的。

(10)“现场可编程门阵列”(FPGA)是一种由高密度门电路组成的可编程逻辑器件。FPGA 包含多达几十万个门电路，而且具有各种不同的结构。有些 FPGA 是非常复杂的，不仅包括可编程逻辑模块，而且包括可编程逻辑模块之间的可编程互连和开关。FPGA 多数是可再编程的(基于“电可擦写可编程只读存储器”和闪存)或者是动态的(基于随机存取存储器)。

3. Writing tasks.

Summary: Write an abstract for *Lesson 14 FPGAs* within 150～200 words.

Composition: Write an essay about your experience of designing and implementing a digital system.

Reading Materials

Passage 1 The PLD Evolution

PROMs

The first programmable integrated circuits were programmable read-only memories (PROMs), which appeared on the scene in 1970. By today's standards, these really were rather simple, but everyone was far too polite to mention it. One way to visualize the manner in which these devices perform their magic is to consider them as consisting of a fixed array of AND functions driving a programmable array of OR functions.

PROMs were originally intended for use as computer memories in which to store program instructions and constant data values. However, design engineers also used them to implement simple logical functions such as look-up tables and state machines. In fact, a PROM can be used to implement any block of combinational logic so long as it doesn't have too many inputs or outputs.

When PROMs are employed to implement combinational logic, they are useful for equations requiring a large number of product terms, but they can only support relatively few inputs because every input combination is always decoded and used. This led engineers to start considering alternative architectures.

PLAs

In order to address the limitations imposed by the PROM architecture, the next step up the PLD evolutionary ladder was that of programmable logic arrays (PLAs), which first became available circa 1975. These were the most user-configurable of the simple PLDs, because both the AND and OR arrays were programmable.

On the downside, signals take a relatively long time to pass through programmable links as opposed to their predefined counterparts. This was especially true in the early technologies. Thus, the fact that both their AND and OR arrays were programmable meant that PLAs were significantly slower than PROMs.

PALs and GALs

In order to address the speed problems posed by PLAs, a new class of device called programmable array logic (PAL) was introduced in the late 1970s. Conceptually, a PAL is almost the exact opposite to a PROM, because it has a programmable AND array driving a predefined OR array.

In 1983, Lattice Semiconductor Corporation introduced a suite of generic array logic (GAL) devices, which provided sophisticated CMOS-based electrically erasable (E^2) variations on the PAL concept. The advantage of PALs and GALs (as compared to PLAs) is that they are faster because only one of their arrays is programmable. On the downside, PALs

and GALs are more limited because they only allow a restricted number of product terms to be OR-ed together.

CPLDs

The one truism in electronics is that everyone is always looking for things to get bigger (in terms of functional capacity), smaller (in terms of physical size), lower (in terms of power consumption), faster, more powerful, and cheaper—surely that's not too much to ask, is it? Thus, the tail end of the 1970s and the early 1980s began to see the emergence of more-sophisticated PLD devices. In order to distinguish these little scamps from their less-sophisticated ancestors (which still find use to this day), these new devices were referred to as complex PLDs (CPLDs). Perhaps not surprisingly, it subsequently became common practice to refer to the original, less-pretentious versions as simple PLDs (SPLDs).

Leading the fray were the inventors of the original PAL devices—the guys and gals at Monolithic Memories Inc. (MMI)—who introduced a component they called a MegaPAL. This was an 84-pin device that essentially comprised four standard PALs with some interconnect linking them together. Unfortunately, the MegaPAL consumed a disproportionate amount of power and it was generally perceived to offer little advantage compared to using four individual devices.

The big leap forward occurred in 1984, when the recently-formed Altera Corporation introduced a CPLD based on a combination of CMOS and EPROM technologies. Using CMOS allowed Altera to achieve tremendous functional density and complexity while consuming relatively little power. And basing the programmability of these devices on EPROM cells made them ideal for use in development and prototyping environments.

FPGAs

Around the beginning of the 1980s, it became apparent that there was a "gap" in the digital integrated circuit continuum. On the one hand, there were programmable devices like SPLDs and CPLDs, which were highly configurable and had fast design and modification times, but which couldn't support large or complex functions. At the other end of the spectrum were ASICs (application-specific integrated circuits).

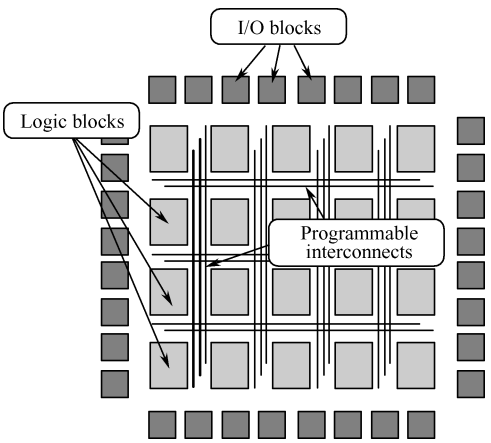


Figure 5.5 The FPGA architecture

ASICs can support extremely large and complex functions, but they are expensive and time-consuming to design. Furthermore, once a design is implemented as an ASIC, it's effectively "frozen in silicon". In order to address this gap, Xilinx introduced a new class of integrated circuit to the market in 1984. This new component (see Figure 5.5) was called a field-

programmable gate array (FPGA). The first devices hit the streets in 1985, and—in many respects— the world (of electronics) has never been the same since.

Questions:

1. Why are PROM inefficient for realizing logic circuits?
2. What is the main difference between PLAs and PALs?
3. What does the acronym SPLD stand for?
4. Which company first introduced CPLD?
5. Which company first introduced FPGA?

Passage 2 Verilog HDL

Verilog HDL is one of the two most common hardware-description languages (HDL) used by integrated circuit (IC) designers. The other one is VHDL. HDLs allows the design to be simulated earlier in the design cycle in order to correct errors or experiment with different architectures. Designs described in HDL are technology-independent, easy to design and debug, and are usually more readable than schematics, particularly for large circuits.

Abstraction Levels

Verilog can be used to describe designs at four levels of abstraction:

- **Algorithmic level** (much like C code with if, case and loop statements).
- **Register transfer level** (RTL uses registers connected by Boolean equations).
- **Gate level** (interconnected AND, NOR etc.).
- **Switch level** (the switches are MOS transistors inside gates).

The language also defines constructs that can be used to control the input and output of simulation.

More recently Verilog is used as an input for synthesis programs which will generate a gate-level description (a netlist) for the circuit. Some Verilog constructs are not synthesizable. Also the way the code is written will greatly effect he size and speed of the synthesized circuit. Most readers will want to synthesize their circuits, so nonsynthesizable constructs should be used only for test benches. These are program modules used to generate I/O needed to simulate the rest of the design.

Code Types

There are two types of code in most HDLs:

- **Structural**—which is a verbal wiring diagram without storage.

```
assign a= b&c|d;
assign d= e&(~ c);
```

  Here the order of the statements does not matter. Changing e will change a.

- **Procedural**—which is used for circuits with storage, or as a convenient way to write conditional logic.

```
always @ (posedge clk);
count < = count+ 1;
```

Procedural code is written like C code and assumes every assignment is stored in memory until over written. For synthesis, with flip-flop storage, this type of thinking generates too much storage. However people prefer procedural code because it is usually much easier to write, for example, if and case statements are only allowed in procedural code. As a result, the synthesizers have been constructed which can recognize certain styles of procedural code as actually combinational. They generate a flip-flop only for left-hand variables which truly need to be stored. However if you stray from this style, beware. Your synthesis will start to fill with superfluous latches.

Verilog vs. VHDL

VHDL is a strongly and richly typed language. Derived from the Ada programming language, its language requirements make it more verbose than Verilog. The additional verbosity is intended to make designs self-documenting. Also, the strong typing requires additional coding to explicitly convert from one data type to another (integer to bit-vector, for example). The creators of VHDL emphasized semantics that were unambiguous and designs that were easily portable from one tool to the next. Hence, race conditions, as an artifact of the language and tool implementation, are not a concern for VHDL users.

Several related standards have been developed to increase the utility of the language. Any VHDL design today depends on at least IEEE-Std 1164 (std_logic type), and many also depend on standard Numeric and Math packages as well. The development of related standards is due to another goal of VHDL's authors: namely, to produce a general language and allow development of reusable packages to cover functionality not built into the language.

VHDL does not define any simulation control or monitoring capabilities within the language. These capabilities are tool dependent. Due to this lack of language-defined simulation control commands and also because of VHDL's user-defined type capabilities, the VHDL community usually relies on interactive GUI environments for debugging design problems.

Verilog is a weakly and limited typed language. Its heritage can be traced to the C programming language and an older HDL called Hilo. All data types in Verilog are predefined in the language. Verilog recognizes that all data types have a bit-level representation. The supported data representations (excluding strings) can be mixed freely in Verilog.

Simulation semantics in Verilog are more ambiguous than in VHDL. This ambiguity gives designers more flexibility in applying optimizations, but it can also (and often does) result in race conditions if careful coding guidelines are not followed. It is possible to have a design that generates different results on different vendors' tools or even on different releases of the same vendor's tool.

Unlike the creators of VHDL, Verilog's authors thought that they provided designers

everything they would need in the language. The more limited scope of the language combined with the lack of packaging capabilities makes it difficult, if not impossible, to develop reusable functionality not already included in the language.

Verilog defines a set of basic simulation control capabilities (system tasks) within the language. As a result of these predefined system tasks and a lack of complex data types, Verilog users often run batch or command-line simulations and debug design problems by viewing waveforms from a simulation results database.

Questions:

1. What's the main goals of VHDL creators?
2. How many types of code exist in most HDLs?
3. Which programming language is Verilog derived from?
4. What does the term *ambiguity* mean in this article? Why is it important?
5. What difference do you think the most considerable between VHDL and Verilog?

Passage 3 SoC

What Is an SoC?

SoC is the acronym for system-on-chip. An SoC is an integrated circuit that contains various electronic components designed to work together to achieve a common goal. The first part of the term—**system**—says that it's all about a complex electronic assembly, while the last part—**chip**—tells you that all the components of that system are squeezed together on a single integrated circuit.

To get a better idea of what an SoC is, imagine it as a full computer that's miniaturized and compressed to fit on a single chip. For instance, an SoC could be compared to a miniature system that has a motherboard, a processor, a graphics card, a network card and so on.

What Are the Common Uses of SoCs?

SoCs are widely used in many industries for all kinds of purposes such as for smartphones, tablets, wearables, digital cameras, wireless routers and so on. However, probably their most common uses today are for powering smartphones. If you're wondering why, keep in mind that both smartphones and tablets are small devices that need a lot of processing power to work and they all need to meet users' requirements, which are increasingly more demanding.

For instance, people want to be able to use their smartphones to browse the internet, listen to music, watch videos, use GPS navigation, shoot photos and film videos, play games, be always connected to social networks, and so on. All these are things that need not only a good processor, but also a good graphic chip, a fast wireless and Bluetooth chip, support for connecting to mobile telephone networks, a GPS chip and the list can go on. And all that must

happen with the least power consumption possible. After all, nobody wants their devices to shut down after very few hours of use. The answer is to miniaturize everything that can be miniaturized and squeeze as many components as possible on a smaller surface. The consequence is a higher processing power and a lower power consumption. That is exactly what an SoC offers.

Which Are the Usual Components of an SoC?

An SoC can have a variety of elements, depending on its purpose. However, as the overwhelming majority of SoCs are used on smartphones, here's a list of the most common components of such devices:

CPU—The central processing unit or the processor is the core of all things inside an SoC. It's the part of the SoC that's responsible for making most calculations and decisions. It receives input from other hardware components and software and delivers appropriate output responses. Without a CPU, there would be no SoC. Most CPUs today have two, four or eight cores inside, to be able to do parallel processing.

GPU—It is short for graphical processing unit and it's also known as the video chip. The GPU is responsible for those 3D games you can play on your smartphone and also for the neat visual transitions that you see in the user interface of any device using an SoC.

RAM—All computing devices need memory to work. For being able to run apps and software data must be used, and to be able to do that, an SoC must have RAM.

ROM—Just as a device needs RAM to be able to run apps and other software, it must also have ROM to be able to store software like the firmware or the operating system that it runs.

Modem—No smartphone would be a phone if it weren't able to connect to radio networks. And by that, we refer to both wireless networks and to mobile networks. Modems take care of network or cellular connectivity.

Besides a processor and memory, other SoCs, like the ones found on wireless routers and other similar networking devices, might also include PCIe interfaces designed to connect the radio transceivers, SATA and/or USB interfaces to connect to storage devices.

The list of components above shows only a few core, must-have elements of an SoC. However, SoCs can have other parts as well. For instance, many have components designed to work with location services, sensors, security, or interfaces capable of working with better cameras or larger screens. The list goes on.

What Are the Advantages and Disadvantages of an SoC?

You might wonder still: why would somebody want to design, manufacture and sell an SoC? There are quite a few reasons for that.

An SoC is usually tiny and doesn't take a lot of space inside an electronic device, which makes it a perfect solution for small devices like smartphones, tablets, or wearables, for

instance. And these three examples only are enough to fill the world with hundreds of billions of dollars.

When you have to design a device that's large enough to fit in your palm at best, integration is a thing you want. That means that you can put a processor, a graphic chip, a modem and many other components in the same square inch.

Because an SoC is small and because it integrates many different parts on a single chip, it means that its manufacturer doesn't have to spend time, money and resources on wiring significant physical parts and creating long circuits, which in its turn means lower manufacturing costs.

SoCs are a lot more power efficient than a system with dedicated, separate components like a desktop PC or a laptop for instance. An SoC can run on batteries for longer times, which makes it a right choice for any mobile device.

On the other hand, SoCs have the big disadvantage of not being adaptable at all. By their nature, SoCs are so tightly integrated and so small that they cannot adapt or, in other words, they cannot be upgraded. An SoC will usually die the same as it was born: nothing changes during its lifetime. And that's also a reason for why if something breaks inside an SoC, you cannot repair or change only that part—you have to replace the whole SoC.

Questions:

1. Which is the most common use of an SoC—the smartphone, the tablet, or the wearable?
2. How many components does an SoC usually have? What are they?
3. Why would people want to design, manufacture and sell an SoC?
4. What does the term PCIe mean in this article?
5. What disadvantages does an SoC have?

Unit 6

Digital Signal Processing Technology

Lesson 16 Digital Signal Processing

Lesson 17 Digital Signal Processors

Lesson 18 ASP or DSP

Passage 1 Typical DSP Applications

Passage 2 Software Radio

Passage 3 Medical Image Processing

Lesson 16 Digital Signal Processing

We don't speak in a digital signal. A digital signal is a language of 1s and 0s that can be processed by mathematics. We speak in real-world, analog signals. Analog signals are real world signals that we experience everyday — sound, light, temperature, and pressure. A digital signal is a numerical representation of the analog signal. It may be easier and more cost-effective to process these signals in the digital world. In the real world, we can convert these signals into digital signals through the analog-to-digital converter, process the signals, and if needed, bring the signals back out to the analog world through the digital-to-analog converter.

Analog-to-Digital Conversion

The first essential step in analog-to-digital (A/D) conversion is to sample an analog signal (see Figure 6.1). This step is performed by a sample-and-hold circuit, which samples at regular intervals called sampling intervals. The length of the sampling interval is the same as the sampling period, and the reciprocal of the sampling period is the sampling frequency f_s. According to the Nyquist theorem, a signal with a maximum frequency of W Hz (called a band-limited signal) must be sampled at least $2W$ samples per second to ensure accurate recording. When this minimum is not respected, distortion called aliasing occurs. Aliasing causes high frequency signals to appear as lower frequency signals. To make sure aliasing will not occur, sampling is always preceded by low-pass filtering. The low-pass filter, called the anti-aliasing filter, removes all frequencies above half the selected sampling rate.

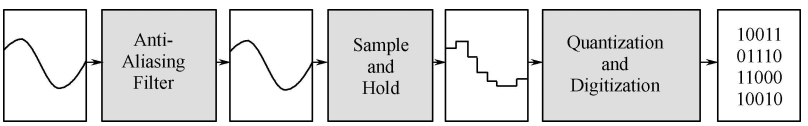


Figure 6.1 Analog-to-digital conversion

After a brief acquisition time, during which a sample is acquired, the sample-and-hold circuit holds the sample steady for the remainder of the sampling interval. This hold time is needed to allow time for an A/D converter to generate a digital code that best corresponds to the analog sample.

The A/D converter chooses a quantization level for each analog sample. An N-bit converter chooses among 2^N possible quantization levels. The larger the number of levels, the smaller the quantization errors, calculated as the difference between the quantized level and the true sample level. Most quantization errors are limited in size to half a quantization step Q.

The quantization step size is calculated as $Q = R/2^N$, where R is the full scale range of the analog signal and N is the number of bits used by the converter. The strength of the signal compared to that of the quantization errors is measured by dynamic range [1] and signal-to-noise ratio.

A digital signal is represented by a set of vertical lines with circles at the top to mark the quantization levels selected for each sample. The bit rate for an A/D converter is the Nf_s, where f_s is the sampling rate.

Finally, each digital sample is assigned a digital code, which completes the A/D process. The result is a digital bit stream. It is this collection of digital codes that is processed in digital signal processing.

To summarize, A/D comprises anti-aliasing, sampling, quantization and digitization.

Digital-to-Analog Conversion

Once digital signal processing is complete, digital-to-analog (D/A) conversion must occur (see Figure 6.2). This process begins by converting each digital code into an analog voltage that is proportional in size to the number represented by the code. This voltage is held steady through zero-order hold until the next code is available, one sampling interval later. This creates a staircase-like signal that contains frequencies above W Hz. These signals are removed with a smoothing low-pass filter, the last step in D/A conversion.

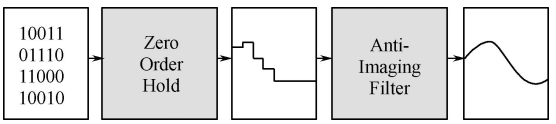


Figure 6.2 Digital-to-analog conversion

The images of each frequency f present in a sampled signal appear, through sampling, at the infinite number of frequencies $kf_s \pm f$ Hz. When the sampling rate is lower than the required Nyquist rate, that is $f_s < 2W$, images of high frequency signals erroneously appear in the baseband [2] (or Nyquist range) due to aliasing. While this undersampling is normally avoided, it can be exploited. For example, signals whose frequencies are restricted to a narrow band of high frequencies can be sampled at a rate similar to twice the width of the band instead of twice the maximum frequency. All of the important signal characteristics can be deduced from the copy of the spectrum that appears in the baseband through sampling. Depending on the relationship between the signal frequencies and the sampling rate, spectral inversion may cause the shape of the spectrum in the baseband to be inverted from the true spectrum of the signal.

Technologies for Digital Signal Processing

If a universal microprocessor solution existed with which every design could be realized, the electronics industry wouldn't be a very competitive place[3]. However, typically in most

electronic designs, more than one processor technology can be used to implement the required functions. The trick is, of course, to choose the one that best delivers the performance, size, power consumption, features, software and tools to get the job done fast—without breaking the budget.

Digital Signal Processors

After several decades of development, digital signal processors (DSP) continue to take the place of competitive processors. DSP are, after all, at the center of signal processing. A DSP is a type of microprocessor—one that is incredibly fast and powerful. It is unique because it processes data in real time. This real-time capability makes a DSP perfect for applications that cannot tolerate any delays. For example, did you ever talk on a cell phone where two people couldn't talk at once? You had to wait until the other person finished talking. If you both spoke simultaneously, the signal was cut—you didn't hear the other person. With today's digital cell phones, which use DSP, you can talk normally. The DSP processors inside cell phones process sounds so rapidly you hear them as quickly as you can speak—in real time.

Here are just some of the advantages of designing with DSP over other microprocessors:

- Single-cycle multiply-accumulate operations.
- Real-time performance simulation and emulation[4].
- Flexibility.
- Reliability.
- Increased system performance.
- Reduced system cost.

However, there are some of the other alternatives available for digital signal processing. How do they compare to DSP?

The FPGA Alternative

Field-Programmable Gate Arrays have the capability of being reconfigurable within a system, which can be a big advantage in applications that need multiple trial versions within development, offering reasonably fast time to market. They also offer greater raw performance per specific operation because of the resulting dedicated logic circuit. However, FPGAs are significantly more expensive and typically have much higher power dissipation than DSP with similar functionality. As such, even when FPGAs are the chosen performance technology in designs such as wireless infrastructure, DSP are typically used in conjunction with FPGAs to provide greater flexibility, better price/performance ratios, and lower system power.

The ASIC Alternative

Application-Specific ICs can be tailored to perform specific functions extremely well, and can be made quite power-efficient. However, since ASICs are not field-programmable, their

functionality cannot be iteratively changed or updated while in product development. As such, every new version of the product requires a redesign and trips through the foundry[5], an expensive proposition, and an impediment to rapid time to market. Programmable DSP, on the other hand, can be updated without changing the silicon, merely change the software program, greatly reducing development costs, and availing aftermarket feature enhancements with mere code downloads. Consequently, more often than not, when you see ASICs in real-time signal processing applications, they are typically employed as bus interfaces, glue logic, and/or functional accelerators for a programmable DSP-based system.

The GPP Alternative

In contrast to ASICs that are optimized for specific functions, general-purpose processors (GPPs) are best suited for performing a broad array of tasks. However, for applications in which the end product must process answers in real time, or must do so while powered by consumer batteries, GPPs' comparatively poor real-time performance and high power consumption all but rules them out [6]. More and more, these processors are being seen as the dinosaurs of the industry, too encumbered with PC compatibility and desktop features to adapt to the changing real-time market place. As the world embraces tiny hand-held wireless-enabled products that require power dissipation measured in milliwatts — not the watts that these processors consume — DSP are the programmable technology of choice. That trend is bound to continue as digital Internet appliances get smaller, faster and more portable.

New Words & Expressions

numerical [nju:ˈmerɪkl] *adj*. 数值的
reciprocal [rɪˈsɪprəkl] *n*. 倒数
precede [prɪˈsi:d] *vt*. 在　　前面
proportion [prəˈpɔ:ʃn] *n*. 比例
infinite [ˈɪnfɪnɪt] *adj*. 无限的
erroneous [ɪˈrəʊnɪəs] *adj*. 错误的
deduce [dɪˈdju:s] *vt*. 推导
incredibly [ɪnˈkredɪbli] *adv*. 难以置信地
budget [ˈbʌdʒɪt] *n*. 预算
simultaneously [ˌsaɪmlˈteɪnɪəsli] *adv*. 同时地
alternative [ɔ:lˈtə:nətɪv] *n*. 供替代的选择
in conjunction with ...　与　　协力
iterative [ˈɪtərətɪv] *adj*. 反复的
proposition [prɒpəˈzɪʃn] *n*. 主张
impediment [ɪmˈpedɪmənt] *n*. 妨碍
more often than not 时常
encumber [ɪnˈkʌmbə] *vt*. 拖累
be encumbered with ...　为　　所累
embrace [ɪmˈbreɪs] *vt*. 包含
appliance [əˈplaɪəns] *n*. 装置

Technical Terms

aliasing [ˈeɪlɪəsɪŋ] *n*. 混叠
digitization [ˌdɪdʒɪtɪˈzeɪʃn] *n*. 数字化
smoothing [ˈsmu：ɪŋ] *n*. 平滑(滤波)
spectrum [ˈspektrəm] *n*. 频谱

simulation [ˌsɪmjʊˈleɪʃn] *n*. 模拟
emulation [ˌemjʊˈleɪʃn] *n*. 仿真
foundry [ˈfaʊndri] *n*. 晶圆代工厂
compatibility [kəmˌpætɪˈbɪlɪti] *n*. 兼容性
sample and hold circuit　采样保持电路
sampling interval　采样间隔
Nyquist theorem　奈奎斯特定理
anti-aliasing filter　抗混叠滤波器
acquisition time　采集时间
quantization level　量化电平
full scale range　满量程范围
dynamic range　动态范围
signal-to-noise ratio　信噪比
zero order hold　零阶保持器
spectral inversion　频谱反转
real time　实时
time to market　上市时间
wireless infrastructure　无线基础设施
price/performance ratio　性价比
bus interface　总线接口
functional accelerator　性能加速器
end product　最终产品
power dissipation　功耗
A/D *abbr*. analog-to-digital　模数转换
D/A *abbr*. digital-to-analog　数模转换
DSP *abbr*. digital signal processor　数字信号处理器
FPGA *abbr*. field-programmable gate array　现场可编程门阵列
ASIC *abbr*. application-specific integrated circuit　专用集成电路
GPP *abbr*. general-purpose processor　通用处理器

Notes

1. “动态范围”(dynamic range)是指信号最小值和最大值之间的范围，通常用信号最大值和信号最小值之比来表示。
2. “基带”(baseband)是指不加任何调制过程的数字信号传输。基带带宽全部用来传送数字脉冲信号，并可以通过“时分复用”方式进行多路数字信号的传输。
3. 本句含虚拟语气，因为并不存在可实现任何设计的所谓“万能微处理器”(universal microprocessor)。
4. “模拟”(simulation)和“仿真”(emulation)是不同的。“模拟”通常是指用软件方法模拟

微处理器功能，而“仿真”则是指用硬件设备(仿真器)监测微处理器的实时运行。

5. foundry 是指为第三方制造芯片的半导体制造商，既可以指那些出售其富余生产能力的大型芯片制造企业，也可以指专为其他公司制造芯片的企业，通常译作“晶圆代工厂”。
6. all but 在此处作为 rule them out 的状语，意为“几乎”“差点儿”。rule ... out 的意思是“将……排除在外”。

Lesson 17 Digital Signal Processors

Digital signal processing tasks can be performed by all processors. Specialized digital signal processors (DSP), however, perform these tasks most efficiently and most quickly.

Architecture

While traditional processors follow the Von Neumann architecture[1] model, which assumes a shared single memory to be used for both program instructions and data (see Figure 6 3), DSP use the Harvard or modified Harvard architecture[2], which includes multiple program and data memories, along with multiple buses to access them (see Figure 6 4). This arrangement means that much less waiting is required when instructions or numbers are fetched from memory. In fact, at least one of each can be fetched simultaneously. Such overlapping of tasks is called pipelining. In addition to multiple memories and buses, all DSP have fast multipliers, accumulators, and shifters, and many have hardware support for circular buffers. Address generators can speed up accesses to memory locations referenced by registers.

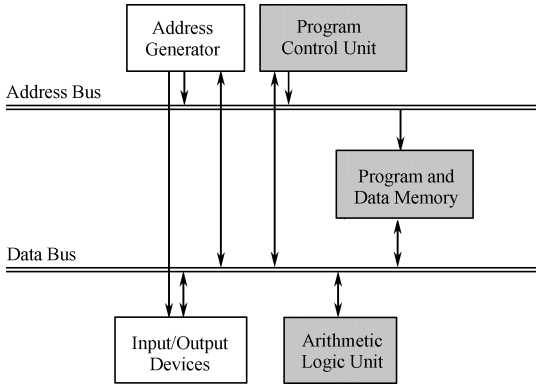


Figure 6.3 Von Neumann architecture

Classification

DSP are available in two major classes: fixed-point and floating-point. The fixed-point class represents real numbers in a fixed number of bits. The position of the binary point (similar to the decimal point) can be controlled by the programmer, and determines the range of numbers that can be represented. As the range increases, though, the available precision goes down, since fewer bits lie to the right of the binary point. In 16 bits, the formats 16.0, 15.1, 14.2, 13.3, 12.4, 11.5, 10.6, 9.7, 8.8, 7.9, 6.10, 5.11, 4.12, 3.13, 2.14, and 1.15 are possible. The dynamic range, calculated as 20 log (Full Scale Range/Smallest Resolvable

Difference), remains the same for all 16-bit formats, 20 log $2^{16}=96.3$ dB.

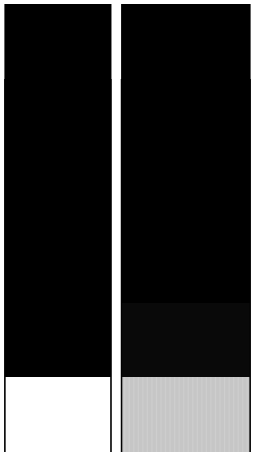


Figure 6.4 Harvard architecture

Floating-point DSP represent real numbers using a mantissa and an exponent, similar to scientific notation: many combine mantissa and exponent into a 32-bit number. The dynamic range for floating point devices is calculated from the largest and smallest multipliers 2^E, where E is the exponent. Thus, for a representation that uses 24 bits for the mantissa and 8 bits for the signed exponent, the dynamic range is 20 log$(2^{127}/2^{-128})=1535.3$ dB. A large dynamic range means the system has great power to represent a wide range of input signals, from very small to very large.

Selection

Choosing a DSP for a particular application is not always easy. The first decision is on whether to choose a fixed-point or a floating-point device. Generally, fixed-point devices are cheaper and quicker, but floating-point devices are more convenient to program and more suited to calculation-intensive algorithms. Second, the data width of the DSP determines how accurately it can represent numbers. Speed is another issue, not only how many cycles occur in each second, but also how many instructions execute in each cycle and how much work each of these instructions accomplishes. One way to assess the minimum requirements for the DSP is to estimate how many instructions must be executed for each received sample. When this number is multiplied by the sampling frequency, the minimum required number of instructions per second is obtained.

The specific hardware and software features offered by a particular DSP can make one choice better than another, as can the amount of on-chip memory available. Sometimes DSP are chosen because well-matched supporting hardware, particularly A/D and D/A converters, is obtainable. Frequently, the quality and convenience of the software tools, for both low-level and high-level programming languages, are also major factors, as is the availability of third-party software. As always, cost is a factor. In fact, quite often, the DSP that is fastest and

offers the most features, but also fits the budget, is the one selected.

Buying DSP

DSP can be purchased in three forms, as a core, as a processor, and as a board level product. In DSP, the term "core" refers to the section of the processor where the key tasks are carried out, including the data registers, multiplier, ALU, address generator, and program sequencer. A complete processor requires combining the core with memory and interfaces to the outside world. While the core and these peripheral sections are designed separately, they will be fabricated on the same piece of silicon, making the processor a single integrated circuit.

Suppose you build cellular telephones and want to include a DSP in the design. You will probably want to purchase the DSP as a processor, that is, an integrated circuit that contains the core, memory and other internal features. To incorporate this IC in your product, you have to design a printed circuit board where it will be soldered in next to your other electronics. This is the most common way that DSP are used.

Now, suppose the company you work for manufactures its own integrated circuits. In this case, you might not want the entire processor, just the design of the core. After completing the appropriate licensing agreement, you can start making chips that are highly customized to your particular application. This gives you the flexibility of selecting how much memory is included, how the chip receives and transmits data, how it is packaged, and so on. Custom devices of this type are an increasingly important segment of the DSP marketplace.

There are several dozen companies that will sell you DSP already mounted on a printed circuit board. These have such features as extra memory, A/D and D/A converters, multiple processors on the same board, and so on. While some of these boards are intended to be used as stand-alone computers, most are configured to be plugged into a host, such as a personal computer. Companies that make these types of boards are called third-party developers. The best way to find them is to ask the manufacturer of the DSP you want to use. Look at the DSP manufacturer's website; if you don't find a list there, send them an E-mail. They will be more than happy to tell you who are using their products and how to contact them.

Trends

Keep in mind that the distinction between DSP and other microprocessors is not always a clear line. For instance, look at how Intel describes the MMX technology addition to its Pentium processor: "Intel engineers have added 57 powerful new instructions specifically designed to manipulate and process video, audio and graphical data efficiently. These instructions are oriented to the highly parallel, repetitive sequences often found in multimedia operations."

In the future, we will undoubtedly see more DSP-like functions merged into traditional microprocessors and microcontrollers. The Internet and other multimedia applications are a strong driving force for these changes. These applications are expanding so rapidly, in a number of years it is very possible that the DSP may be the "traditional" microprocessor.

New Words & Expressions

specialized [ˈspeʃəˈlaɪzd] *adj*. 专门的
shared [ʃeəd] *adj*. 共享的
reference [ˈrefərəns] *vt*. 查阅
match [mætʃ] *vt*. 相配
solder [ˈsəʊldə] *vt*. 焊接
architecture [ˈɑːkɪtektʃə] *n*. 体系结构
customize [ˈkʌstəmaɪz] *vt*. 定制
merge [mɜːdʒ] *vt*. 合并
fixed [fɪkst] *adj*. 固定的
floating [ˈfləʊtɪŋ] *adj*. 浮动的
point [pɒɪnt] *n*. 小数点
algorithm [ˈælgərɪðəm] *n*. 算法

Technical Terms

multiplier [ˈmʌltɪplaɪə] *n*. 乘法器
accumulator [əˈkjuːmjʊleɪtə] *n*. 累加器
shifter [ˈʃɪftə] *n*. 移位器
mantissa [mænˈtɪsə] *n*. 尾数
exponent [ɪkˈspəʊnənt] *n*. 指数
Von Neumann architecture　冯·诺依曼结构
Harvard architecture　哈佛结构
circular buffer　循环缓存器
address generator　地址产生器
smallest resolvable difference　最小可分辨值
scientific notation　科学记数法
parallel architecture　并行结构
looping scheme　循环机制
calculation-intensive algorithm　运算密集型算法
third-party developer　第三方开发者
licensing agreement　专利使用权转让协定
DSP *abbr*. digital signal processor　数字信号处理器
A/D *abbr*. analog-to-digital　模数转换
D/A *abbr*. digital-to-analog　数模转换
MMX *abbr*. multimedia extension　多媒体增强(指令集)

Notes

1. “冯·诺依曼结构”得名自美籍数学家约翰·冯·诺依曼(John Von Neumann，1903—1957)。20 世纪初，许多重大数学发现都源于冯·诺依曼。其主要成就包括：(1)提出“存储程序计算机”概念；(2)对量子力学的公式化。

2. “哈佛结构”得名自 20 世纪 40 年代在 Howard Aiken(1900—1973)领导下的哈佛大学(Harvard University)的研究工作。

Lesson 18 ASP or DSP

Signals may be processed using analog techniques (analog signal processing, or ASP), digital techniques (digital signal processing, or DSP), or a combination of analog and digital techniques (mixed signal processing, or MSP). In some cases, the choice of techniques is clear; in others, there is no clear-cut choice, and second-order considerations may be used to make the final decision[1].

With respect to DSP, the factor that distinguishes it from traditional computer analysis of data is its speed and efficiency in performing sophisticated digital processing functions such as filtering, FFT analysis, and data compression in real time.

The term *mixed signal processing* implies that both analog and digital processing is done as part of the system. The system may be implemented in the form of a printed circuit board or a single integrated circuit chip. In the context of this broad definition, ADCs and DACs are considered to be mixed signal processors, since both analog and digital functions are implemented in each. Recent advances in very large-scale integration (VLSI) processing technology allow complex digital processing as well as analog processing to be performed on the same chip. The very nature of DSP itself implies that these functions can be performed in real time.

ASP vs. DSP

Today's engineer faces a challenge in selecting the proper mix of analog and digital techniques to solve the signal processing task at hand. It is impossible to process real-world analog signals using purely digital techniques, since all sensors (microphones, thermocouples, strain gages, piezoelectric crystals, disk drive heads, etc.) are analog sensors. Therefore, some sort of signal conditioning circuitry is required in order to prepare the sensor output for further signal processing, whether it be analog or digital. Signal conditioning circuits are, in reality, analog signal processors, performing such functions as multiplication (gain), isolation (instrumentation amplifiers and isolation amplifiers), detection in the presence of noise (high common-mode instrumentation amplifiers etc.), dynamic range compression (log amps, LOGDACs, and programmable gain amplifiers), and filtering (both passive and active)[2]. Several methods of accomplishing signal processing are shown in Figure 6.5. The top portion of the figure shows the purely analog approach. The latter parts of the figure show the DSP approach. Note that once the decision has been made to use DSP techniques, the next decision must be where to place the ADC in the signal path.

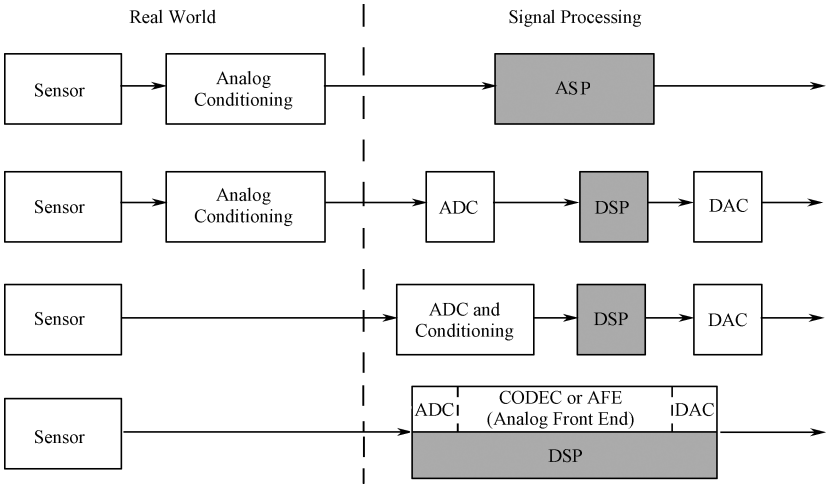


Figure 6.5 Signal processing approaches

In general, as the ADC is moved closer to the actual sensor, more of the analog signal conditioning burden is now placed on the ADC. The added ADC complexity may take the form of increased sampling rate, wider dynamic range, higher resolution, input noise rejection, input filtering and programmable gain amplifiers (PGAs), on-chip voltage references, etc., all of which add functionality and simplify the system[3]. With today's high-resolution/high-sampling-rate data converter technology, significant progress has been made in integrating more and more of the conditioning circuitry within the ADC/DAC itself. In the measurement area, for instance, 24-bit ADCs are available with built-in programmable gain amplifiers (PGAs) which allow full-scale bridge signals of 10 mV to be digitized directly with no further conditioning. At voiceband and audio frequencies, complete coder-decoders (Codecs, or Analog Front Ends) are available which have sufficient on-chip analog circuitry to minimize the requirements for external conditioning components. At video speeds, analog front ends are also available for such applications as CCD image processing and others.

A Practical Example

As a practical example of the power of DSP, consider the comparison between an analog and a digital low-pass filter, each with a cutoff frequency of 1 kHz. The digital filter is implemented in a typical sampled data system shown in Figure 6.6. Note that there are several implicit requirements in the diagram. First, it is assumed that an ADC/DAC combination is available with sufficient sampling frequency, resolution, dynamic range to accurately process the signal. Second, the DSP must be fast enough to complete all its calculations within the sampling intervals, $1/f_s$. Third, analog filters are still required at the ADC input and DAC output for anti-aliasing and anti-imaging, but the performance demands are not as great. Assuming these requirements are met, the following offers a comparison between the digital and analog filters. The required cutoff frequency of both filters is 1 kHz. The analog filter is

realized as a 6-pole Chebyshev Type 1 filter (ripple in passband, no ripple in stopband), and the response is shown in Figure 6.7. In practice, this filter would probably be realized using three 2-pole stages, each of which requires an op amp, and several resistors and capacitors. Modern filter design CAD packages make the 6-pole design relatively straightforward, but maintaining the 0.5 dB ripple specification requires accurate component selection and matching.

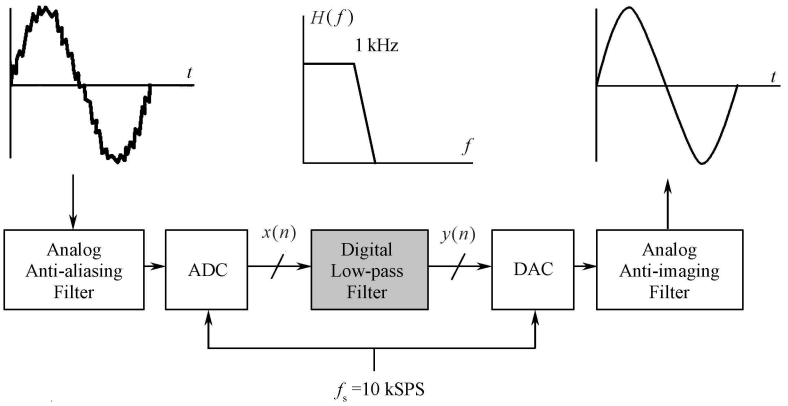
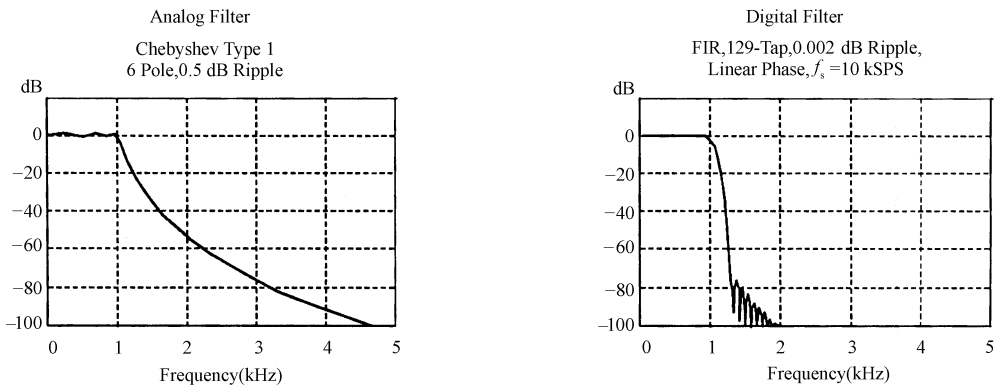


Figure 6.6 A typical sampled data system



(a) The magnitude response in dB of an analog filter (b) The magnitude response in dB of a digital filter

Figure 6.7 The analog filter versus the digital filter in terms of magnitude

On the other hand, the 129-tap digital FIR filter shown has only 0.002 dB passband ripple, linear phase, and a much sharper roll-off. In fact, it could not be realized using analog techniques! Another obvious advantage is that the digital filter requires no component matching, and it is not sensitive to drift since the clock frequencies are crystal-controlled. The 129-tap filter requires 129 MACs in order to compute an output sample. This processing must be completed within the sampling interval, $1/f_s$, in order to maintain real-time operation. In this example, the sampling frequency is 10k SPS, therefore 100 μs is available for processing, assuming no significant additional overhead requirement. Most DSP can complete the entire multiply-accumulate process (and other functions necessary for the filter) in a single instruction cycle.

Therefore, a 129-tap filter requires that the instruction rate be greater than 129/100 μs= 1.3 million instructions per second (MIPS). DSP are available with instruction rates much greater than this, so the DSP certainly is not the limiting factor in this application.

In a practical application, there are certainly many other factors to consider when evaluating analog versus digital filters, or analog versus digital signal processing in general. Most modern signal processing systems use a combination of analog and digital techniques in order to accomplish the desired function and take advantage of the best of both the analog and the digital world.

New Words & Expressions

technique [tekˈniːk] *n*. 技术
order [ˈɔːdə] *n*. 阶，次
with respect to ... 关于……
context [ˈkɒntekst] *n*. 背景
in the context of ... 在……的情况下
at hand 在手边
in reality 事实上
strain [streɪn] *n*. 张力
gage [geɪdʒ] *n*. 计量器
package [ˈpækɪdʒ] *n*. 软件包
desired [dɪˈzaɪəd] *adj*. 期望的
tap [tæp] *n*. 抽头
straightforward [streɪtˈfɔːwəd] *adj*. 简单易懂的
overhead [əʊvəˈhed] *n*. 开销
versus [ˈvɜːsəs] *prep*. 与……相比
compression [kəmˈpreʃən] *n*. 压缩

Technical Terms

filtering [ˈfɪltərɪŋ] *n*. 滤波
microphone [ˈmaɪkrəfəʊn] *n*. 扩音器
thermocouple [ɜːməʊˈkʌpəl] *n*. 热电偶
piezoelectric [paɪˌizəʊɪˈlektrɪk] *adj*. 压电的
crystal [ˈkrɪstl] *n*. 晶体
passive [ˈpæsɪv] *adj*. 无源的
active [ˈæktɪv] *adj*. 有源的
pole [pəʊl] *n*. 极点
ripple [ˈrɪpl] *n*. 波纹
passband [ˈpɑːsbænd] *n*. 通带
stopband [ˈstɒpbænd] *n*. 阻带
data compression 数据压缩
strain gage 应力计
piezoelectric crystal 压电晶体
disk drive head 磁盘驱动器磁头
signal conditioning circuit 信号调理电路
voltage reference 参考电压
cutoff frequency 截止频率

Chebyshev Type 1 filter　切比雪夫 I 型滤波器
roll-off　滚降
op amp *abbr*. operational amplifier　运算放大器
FFT *abbr*. fast Fourier transform　快速傅里叶变换
CAD *abbr*. computer-aided design　计算机辅助设计
PCB *abbr*. printed circuit board　印制电路板
MAC *abbr*. multiplication and accumulation　乘法累加
PGA *abbr*. programmable gain amplifier　可编程增益放大器
ADC *abbr*. analog-to-digital converter　模数转换器
DAC *abbr*. digital-to-analog converter　数模转换器
CCD *abbr*. charge-coupled device　电荷耦合器件
DSP *abbr*. digital signal processing　数字信号处理
MIPS *abbr*. million instructions per second　百万条指令/秒
SPS *abbr*. samples per second　样本/秒

Notes

1. 此句可译为：在有些情况下，非常容易明确选择技术；但在有些情况下，却无法一下子明确选择技术，可能需要考虑次要因素才能做出最终决定。其中，clear-cut 是复合形容词，表示“界线分明的”；second-order 是复合名词做形容词，表示“再一次的”。
2. 此句可译为：实际上，信号调理电路就是模拟信号处理器，它能够完成乘法（增益）、隔离（测量放大器和隔离放大器）、噪声环境下的信号检测（高共模抑制比的测量放大器）、动态范围压缩（对数放大器、对数数模转换器和增益可编程放大器）和滤波（包括无源滤波和有源滤波）等功能。
3. 此句可译为：ADC 复杂性的提高可表现为如下几种形式：提高采样率、加宽动态范围、提高精度、抑制输入噪声、对输入信号进行滤波、增益可编程放大器（PGA）、片内参考电压等。所有这些都增加了 ADC 的功能并简化了系统。

Exercises

1. Fill in the blanks with proper words, phrases or clauses.

(1) What does “real-time” mean? In an _____（模拟系统），every task is performed in “real time” with _____（连续的）signals and processing. In a digital signal processing (DSP) system, signals are represented with sets of samples, i.e., values at _____（离散点）in time. Thus the time for processing a given number of samples in a DSP system can have an arbitrary interpretation in “real time”, depending on the _____（采样率）. In real-time applications, the sampling frequency must be at least _____ the frequency of the highest _____（频率分量）of interest in the analog signal. The time between samples _____（被称为）the sampling interval. To consider a system as operating in “real time”, all processing of _____

(一组给定的数据)(one or more samples, depending on the algorithm) must be completed _____(在新数据到来之前).

This definition of real time implies that, for a processor operating _____(以给定的时钟速率), the speed and quantity of the input data determines how much processing can be applied to the data without falling behind the data stream. The idea of having a limited amount of time with which to handle data may seem _____(奇怪的)to analog designers _____ this concept does not have a parallel in analog systems. In analog systems, signals are processed continuously. The only penalty in a slow system is limited frequency response. By comparison, digital systems process parts of the signal, enough for very accurate approximations, but only within a limited block of time. Real-time DSP can be limited by _____(数据量)or type of processing that can be completed within the algorithm's time budget. For example, a given DSP processor handling data values sampled at, say, 48-kHz (audio signals), has less time to process those data values, including execution of all necessary tasks, than one sampling 8-kHz voice-band data.

(2) One can program a DSP to process data using one of several strategies for handling the "event", the arrival of data. A _____(状态位)or flag pin could be read periodically to determine whether new data is available. But — "polling" wastes processor cycles. The data may arrive just after the last poll, but it can't make its presence known until the next poll. This _____(使实时系统开发变得困难).

The second strategy is for the data to interrupt the processor on arrival. Using interrupts to notify the processor is _____(有效率的), though not as easy to program; _____(时钟周期)can be wasted during the wait for an interrupt. Nevertheless, _____(事件驱动的中断编程)being well-suited to processing _____(现实世界信号)promptly, most DSP are designed to deal efficiently with them. In fact, they are designed to _____(响应)very quickly to interrupts. In many DSP-based systems, the interrupt rates, based on the input data sampling rate, are often totally unrelated to the DSP's clock rate.

Because interrupt processing is such a vital element in DSP systems, processors typically have built-in hardware mechanisms to handle interrupts efficiently. Hard-wired mechanisms are more efficacious than software alone because a DSP's interrupt service routines (ISRs) may have to _____(满足以下所有要求):

- **Fast context switching** — Switch from working on one task and its data (a context) to another context without the time loss and complication associated with writing programs to save register contents and chip status information.
- **Nested-interrupt handling** — Handle _____(多个)interrupts of different priorities "simultaneously". The DSP handles one interrupt at a time, but an interrupt of higher priority can take precedence over the handling of a lower-priority interrupt.
- **Continue to accept data/record status** — While the DSP services an interrupt, events keep on occurring in the real world and data keeps on arriving. To _____(跟上)the "real-world", the DSP must record these events and accept the data — then process

them when it has finished servicing the interrupt.

In DSP systems, interrupts are typically generated by the arrival of data or the requirement to provide new output data. Interrupts may occur with each sample, or they may occur after _____(一帧数据)has been collected. The differences greatly influence how the DSP algorithm deals with data.

For algorithms that operate on a sample-by-sample basis, DSP software may be required to handle each incoming and outgoing data value. Each DSP _____(串行口)incorporates two data I/O registers, a receive register (Rx), and a transmit register (Tx). When a serial word is received, the port will typically generate a Receive interrupt. The processor _____(停下正在做的事情), begins executing code at the _____(中断向量)location, reads the incoming value from the Rx register into a processor data register, and either operates _____ that data value or returns to its background task.

_____(为了发送数据), the serial port can generate a Transmit interrupt, indicating that new data can be written to the Tx register. The DSP can then begin code execution at the Tx interrupt vector and typically _____(转移一个数值)from a data register to the Tx register. If data input and output are controlled by the same sampling clock, only one interrupt is necessary. For example, if _____(一个程序段)is initiated by Receive interrupt timing, new data would be read during the interrupt routine; then either the previously computed result, which is being held in a register, would be transmitted, _____ a new result would be computed and immediately transmitted — as the final step of the interrupt routine.

These mechanisms help a DSP to approach the ability _____ emulate what an analog system does naturally — continuously process data in real time — but with digital _____(精度和灵活性). In addition, in an efficiently programmed digital system, spare processor cycles left between processing data sets can be used to handle other tasks.

2. Translate the following passages into Chinese or English.

(1) The primary reason for processing real-world signals is to extract information from them. This information normally exists in the form of signal amplitude (absolute or relative), frequency or spectral content, phase, or timing relationships with respect to other signals. Once the desired information is extracted from the signal, it may be used in a number of ways.

(2) In some cases, it may be desirable to reformat the information contained in a signal. This would be the case in the transmission of a voice signal over a frequency-division multiple access (FDMA) telephone system. In this case, analog techniques are used to "stack" voice channels in the frequency spectrum for transmission via microwave relay, coaxial cable, or fiber. In the case of a digital transmission link, the analog voice information is first converted into digital using an ADC. The digital information representing the individual voice channels is multiplexed in time (time-division multiple access, or TDMA) and transmitted over a serial digital transmission link (as in the T-Carrier system).

(3) Another requirement for signal processing is to compress the frequency content of the signal (without losing significant information) then format and transmit the information at

lower data rates, thereby achieving a reduction in required channel bandwidth. High speed modems and adaptive pulse code modulation systems (ADPCM) make extensive use of data reduction algorithms, as do digital mobile radio systems, MPEG recording and playback, and high-definition television (HDTV).

(4) The roots of DSP are in the 1960s and 1970s when digital computers first became available. Computers were expensive during this era, and DSP was limited to only a few critical applications. Pioneering efforts were made in four key areas: radar & sonar, where national security was at risk; oil exploration, where large amounts of money could be made; space exploration, where the data are irreplaceable; and medical imaging, where lives could be saved.

(5) The first step in converting an analog signal into a digital signal is sampling. This step is accomplished using a sample-and-hold circuit, which grabs a sample from the signal and holds it steady until the next sampling time. Samples are normally collected at regular time intervals, called sampling periods. If the rate of the sampling is not adequate, distortion called aliasing can occur. The output from the sample and hold circuit is passed to an A/D converter, which chooses a quantization level that is closest to the actual amplitude of the signal. This is the second step in A/D conversion. The number of levels to choose from depends on the number of bits used by the converter: for N bits, 2^N levels are possible. Since analog amplitudes cannot in general be represented perfectly in a digital system, quantization errors occur.

(6) "数字信号处理"(DSP)是无数家用和商用系统中的关键部分，其应用领域与日俱增。因此，DSP 正在成为技术专家和工程师们专业知识的重要组成部分。

(7) "信号"(如声音、光或者电压)就是携带信息的变化。模拟信号来自现实世界中。模拟信号在每个时间点都有定义，其幅度取值没有限制。模拟信号不适合计算机处理。通过采样、量化，可以将模拟信号转换为数字信号。数字信号只在采样点有定义，其幅度只能取有限的离散值。经过处理后，数字信号被转换回模拟信号。

(8) 数字信号处理可分成定点和浮点两大类。定点与浮点是指器件内数据的存储格式和运算格式。定点数字信号处理器通常用最少 16 位来表示数据，浮点数字信号处理器往往用最少 32 位来存储数据。

(9) 和其他科学工程应用一样，数字信号处理器通常使用汇编语言或者 C 语言编程。汇编语言程序执行速度快，C 语言程序更易于开发和维护。在 PC 程序等传统应用中，C 语言几乎总是首选。如果使用了汇编语言，则仅限于那些必须以极高速度运行的短小子程序。然而，和传统软件任务相比，DSP 程序存在两大差异：① 程序往往很短(如 100 行，而非 10000 行)；② 执行速度是关键。

(10) 为使常见任务执行起来更加便捷，数字信号处理器(DSP)往往使用特殊指令进行编程。例如，为了充分利用并行结构，多数 DSP 提供多功能指令。又如，针对运算中的大量重复操作，许多 DSP 指令设计了高效的循环机制。

3. Writing tasks.

Summary: Write an abstract for *Lesson* 17 *Digital Signal Processors* within 150～200 words.

Composition: Write an essay about your understanding of digital filters.

Reading Materials

Passage 1 Typical DSP Applications

Radar

Radar is an acronym for radio detection and ranging. In the simplest radar system, a radio transmitter produces a pulse of radio frequency energy a few microseconds long. This pulse is fed into a highly directional antenna, where the resulting radio wave propagates away at the speed of light. Aircraft in the path of this wave will reflect a small portion of the energy back toward a receiving antenna, situated near the transmission site. The distance to the object is calculated from the elapsed time between the transmitted pulse and the received echo. The direction to the object is found more simply; you know where you pointed the directional antenna when the echo was received.

The operating range of a radar system is determined by two parameters: how much energy is in the initial pulse, and the noise level of the radio receiver. Unfortunately, increasing the energy in the pulse usually requires making the pulse longer. In turn, the longer pulse reduces the accuracy and precision of the elapsed time measurement. This results in a conflict between two important parameters: the ability to detect objects at long range, and the ability to accurately determine an object's distance.

DSP(digital signal processing) has revolutionized radar in three areas, all of which relate to this basic problem. First, DSP can compress the pulse after it is received, providing better distance determination without reducing the operating range. Second, DSP can filter the received signal to decrease the noise. This increases the range, without degrading the distance determination. Third, DSP enables the rapid selection and generation of different pulse shapes and lengths. Among other things, this allows the pulse to be optimized for a particular detection problem. Now the impressive part: much of this is done at a sampling rate comparable to the radio frequency used, at high as several hundred megahertz! When it comes to radar, DSP is as much about high-speed hardware design as it is about algorithms.

Sonar

Sonar is an acronym for sound navigation and ranging. It is divided into two categories, active and passive. In active sonar, sound pulses between 2 kHz and 40 kHz are transmitted into the water, and the resulting echoes detected and analyzed. Uses of active sonar include: detection & localization of undersea bodies, navigation, communication, and mapping the sea floor. A maximum operating range of 10 to 100 kilometers is typical. In comparison, passive sonar simply listens to underwater sounds, which includes: natural turbulence, marine life, and mechanical sounds from submarines and surface vessels.

DSP has revolutionized sonar in many of the same areas as radar: pulse generation, pulse compression, and filtering of detected signals. In one view, sonar is simpler than radar

because of the lower frequencies involved. In another view, sonar is more difficult than radar because the environment is much less uniform and stable. Sonar systems usually employ extensive arrays of transmitting and receiving elements, rather than just a single channel. By properly controlling and mixing the signals in these many elements, the sonar system can steer the emitted pulse to the desired location and determine the direction that echoes are received from. To handle these multiple channels, sonar systems require the same massive DSP computing power as radar.

CT and MRI

In 1895, Wilhelm Conrad Rötgen discovered that x-rays could pass through substantial amounts of matter. Medicine was revolutionized by the ability to look inside the living human body. Medical x-ray systems spread throughout the world in only a few years. In spite of its obvious success, medical x-ray imaging was limited by four problems until DSP and related techniques came along in the 1970s. First, overlapping structures in the body can hide behind each other. Second, it is not always possible to distinguish between similar tissues. Third, x-ray images show anatomy, the body's structure, and not physiology, the body's operation. Fourth, x-ray exposure can cause cancer, requiring it to be used sparingly and only with proper justification.

The problem of overlapping structures was solved in 1971 with the introduction of the first computerized tomography scanner (formerly called computerized axial tomography, or CAT scanner). Computerized tomography (CT) is a classic example of DSP. X-rays from many directions are passed through the section of the patient's body being examined. Instead of simply forming images with the detected x-rays, the signals are converted into digital data and stored in a computer. The information is then used to calculate images that appear to be slices through the body. These images show much greater detail than conventional techniques, allowing significantly better diagnosis and treatment. In 1979, two of CT's principle contributors, Godfrey N. Hounsfield and Allan M. Cormack, shared the Nobel Prize in Medicine. That's good DSP!

The last three x-ray problems have been solved by using penetrating energy other than x-rays, such as radio and sound waves. DSP plays a key role in all these techniques. For example, magnetic resonance imaging (MRI) uses magnetic fields in conjunction with radio waves to probe the interior of the human body. Properly adjusting the strength and frequency of the fields cause the atomic nuclei in a localized region of the body to resonate between quantum energy states. This resonance results in the emission of a secondary radio wave, detected with an antenna placed near the body. The strength and other characteristics of this detected signal provide information about the localized region in resonance. Adjustment of the magnetic field allows the resonance region to be scanned throughout the body, mapping the internal structure. This information is usually presented as images, just as in computed tomography. Besides providing excellent discrimination between different types of soft tissue,

MRI can provide information about physiology, such as blood flow through arteries. MRI relies totally on DSP techniques, and could not be implemented without them.

Questions:

1. How DSP revolutionized radar system?
2. What is the most important application of passive sonar?
3. Who contributed to CT's principle and won Noble Prize?
4. What problems used to limit medical x-ray imaging?
5. Can you tell some DSP applications in your daily life?

Passage 2 Software Radio

What Is a Software Radio?

The term software radio was coined by Joe Mitola in 1991 to refer to the class of reprogrammable or reconfigurable radios. In other words, the same piece of hardware can perform different functions at different times.

The exact definition of a software radio is controversial, and no consensus exists about the level of reconfigurability needed to qualify a radio as a software radio. A radio that includes a microprocessor or digital signal processor (DSP) does not necessarily qualify as a software radio. However, a radio that defines in software its modulation, error correction, and encryption processes, exhibits some control over the RF hardware, and can be reprogrammed is clearly a software radio. A good working definition of a software radio is a radio that is substantially defined in software and whose physical layer behavior can be significantly altered through changes to its software. The degree of reconfigurability is largely determined by a complex interaction between a number of common issues in radio design, including systems engineering, antenna form factors, RF electronics, baseband processing, speed and reconfigurability of the hardware, and power supply management.

The functionality of conventional radio architectures is usually determined primarily by hardware with minimal configurability through software. The hardware consists of the amplifiers, filters, mixers (probably several stages), and oscillators. The software is confined to controlling the interface with the network, stripping the headers and error correction codes from the data packets, and determining where the data packets need to be routed based on the header information. Because the hardware dominates the design, upgrading a conventional radio design essentially means completely abandoning the old design and starting over again. In upgrading a software radio design, the vast majority of the new content is software and the rest is improvements in hardware component design. In short, software radios represent a paradigm shift from fixed, hardware-intensive radios to multiband, multimode, software-intensive radios.

Characteristics and Benefits of a Software Radio

A model of a practical software radio is shown in Figure 6.8. The receiver begins with a smart antenna that provides a gain versus direction characteristic to minimize interference, multipath, and noise. The smart antenna provides similar benefits for the transmitter. Most practical software radios digitize the signal as early as possible in the receiver chain while keeping the signal in the digital domain and converting to the analog domain as late as possible for the transmitter using a digital-to-analog converter (DAC). Often the received signal is digitized in the intermediate frequency (IF) band. Conventional radio architectures employ a super heterodyne receiver, in which the RF signal is picked up by the antenna along with other spurious/unwanted signals, filtered, amplified with a low noise amplifier (LNA), and mixed with a local oscillator (LO) to an IF. Depending on the application, the number of stages of this operation may vary. Finally, the IF is then mixed exactly to baseband.

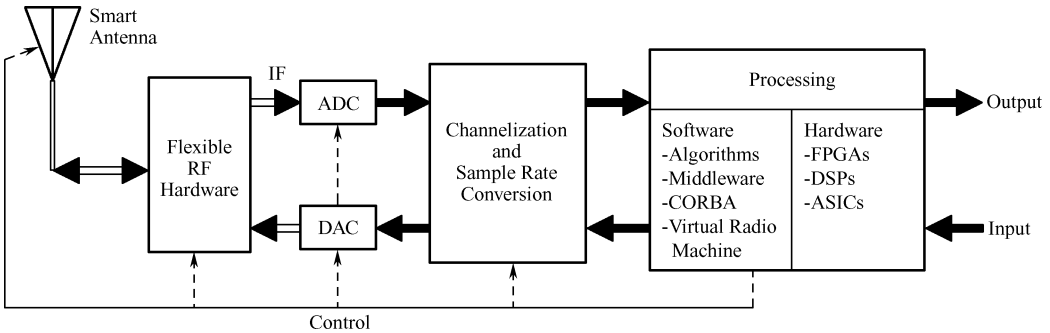


Figure 6.8 Model of a software radio

Digitizing the signal with an analog-to-digital converter (ADC) in the IF range eliminates the last stage in the conventional model in which problems like carrier offset and imaging are encountered. When sampled, digital IF signals give spectral replicas that can be placed accurately near the baseband frequency, allowing frequency translation and digitization to be carried out simultaneously. Digital filtering (channelization) and sample rate conversion are often needed to interface the output of the ADC to the processing hardware to implement the receiver. Likewise, digital filtering and sample rate conversion are often necessary to interface the digital hardware that creates the modulated waveforms to the digital to analog converter. Processing is performed in software using DSPs, field-programmable gate arrays (FPGAs), or application-specific integrated circuits (ASICs). The algorithm used to modulate and demodulate the signal may use a variety of software methodologies, such as middleware, e.g., common object request broker architecture (CORBA), or virtual radio machines, which are similar in function to Java virtual machines. This forms a typical model of a software radio.

The software radio provides a flexible radio architecture that allows changing the radio personality, possibly in real time, and in the process somewhat guarantees a desired QoS. The flexibility in the architecture allows service providers to upgrade the infrastructure and market

new services quickly. This flexibility in hardware architecture combined with flexibility in software architecture, through the implementation of techniques such as object-oriented programming and object brokers, provides software radio with the ability to seamlessly integrate itself into multiple networks with wildly different air and data interfaces. In addition, software radio architecture gives the system new capabilities that are easily implemented with software. For example, typical upgrades may include interference rejection techniques, encryption, voice recognition and compression, software-enabled power minimization and control, different addressing protocols, and advanced error recovery schemes. Such capabilities are well-suited for advanced wireless networking approaches.

Questions:

1. What does the concept of software radio refer to?
2. What does the term *reconfigurability* mean in this article?
3. What benefits will software radio bring according to the author?
4. What is the function of ADC block in Figure 6.8?
5. What is the function of DAC block in Figure 6.8?

Passage 3 Medical Image Processing

There are numerous concepts and approaches for structuring the field of medical image processing that focus on different aspects of its core areas. These areas shape three major processes underlying this field—image formation, image computing, and image management.

Image Formation

The first integral step in the image formation is an acquisition of raw imaging data. It contains the original information about captured physical quantities describing internal aspects of the body. This information becomes the primary subject for all subsequent steps of image processing.

Different types of imaging modalities may utilize different physical principles and thus involve detection of different physical quantities. For example, in digital radiography (DR) or computerized tomography (CT), it is the energy of incident photons; in positron emission tomography (PET), it is the photons energy and their detection time; in magnetic resonance imaging (MRI), it is the parameters of a radio-frequency signal emitted by the excited atoms; and in ultrasonography, it is the parameters of the acoustic echoes.

However, regardless of the type of imaging modality, the data acquisition process can be subdivided into detection of a physical quantity that also includes its conversion into an electrical signal, preconditioning of the acquired signal, and its digitization. A generic block diagram representing all these steps applicable to most of the medical imaging modalities is schematically depicted in Figure 6.9.

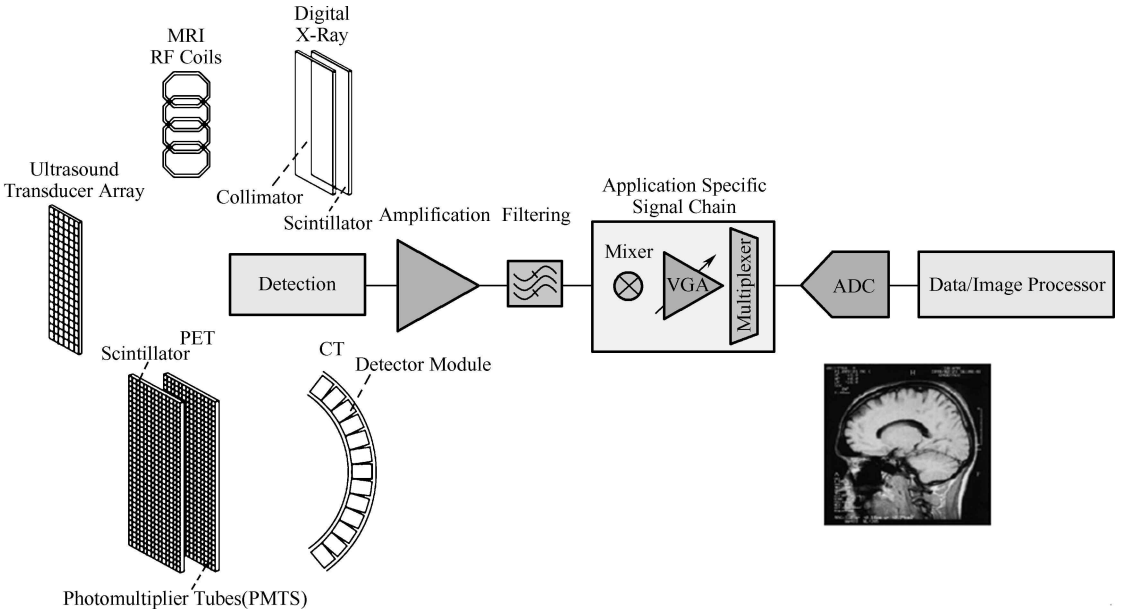


Figure 6.9 Generic block diagram of the data acquisition process

Image reconstruction is a mathematical process of forming an image using the acquired raw data. For multidimensional imaging, this process also includes a combination of multiple data sets captured at different angles or different time steps. This part of medical image processing deals with inverse problems, which is a fundamental subject of the field. There are two primary algorithms used to solve this type of problems—analytical and iterative.

Typical examples of analytical methods include filtered backprojection (FBP), widely used in tomography; Fourier transform (FT), particularly important in MRI; and delay-and-sum (DAS) beamforming, a technique which is integral to ultrasonography. These algorithms are elegant and efficient in terms of required processing power and computational time.

However, they are based on idealized models and thus have some distinctive limitations, including their inability to handle such complex factors as statistical properties of the measurement noise and physics of an imaging system.

Iterative algorithms overcome those limitations to enable significant improvement in insensitivity to noise and the capability to reconstruct an optimal image using incomplete raw data. Iterative methods typically use a system and statistical noise model to calculate projections based on the initial object model with assumed coefficients. The difference between the calculated projections and the original data defines new coefficients used to update the object model. This procedure is repeated using multiple iteration steps until a cost function, which maps the estimated and true values, is minimized—resulting in a convergence of the reconstruction process to the final image.

There is a large variety of iterative methods including maximum likelihood expectation maximization (MLEM), maximum a posteriori (MAP), algebraic reconstruction (ARC) technique, and many others widely used across medical imaging modalities today.

Image Computing

Image computing deals with computational and mathematical methods operated on reconstructed imaging data to extract clinically relevant information. These methods are applied for enhancement, analysis, and visualization of the imaging results.

Image enhancement refines a transform representation of an image to improve interpretability of the contained information. Its methods can be subdivided into spatial and frequency domain techniques.

The spatial domain techniques operate directly on image pixels, which is particularly useful for contrast optimization. These techniques typically rely on logarithmic, histogram, and power law transforms. The frequency domain methods use frequency transform and are best suited for smoothening and sharpening the images by applying different kinds of filters.

Utilization of all these techniques allow for noise and inhomogeneity reduction, contrast optimization, enhancement of edges, elimination of artifacts, and improvement of other relevant properties that are crucial for the subsequent image analysis and its accurate interpretation.

Image analysis is the central process in image computing that uses a broad variety of methods which can be grouped into three main categories: image segmentation, image registration, and image quantification.

The image segmentation process partitions the image into meaningful contours of different anatomical structures. Image registration ensures correct alignment of multiple images, which is particularly important for analysis of temporal changes or a combination of images acquired using different modalities. The process of quantification determines properties of the identified structures such as volume, diameter, composition, and other relevant anatomical or physiological information. All these processes have a direct impact on the inspection quality of the imaging data and the accuracy level of medical findings.

The visualization process renders the image data to visually represent anatomical and physiological imaging information in a specific form over defined dimensions. Through direct interaction with data, the visualization can be performed both at the initial and intermediate phases of imaging analysis—for instance, to assist segmentation and registration processes, and at the final stage to display the refined results.

Image Management

The final part of medical image processing deals with management of the acquired information and encompasses various techniques for storage, retrieval, and communication of image data. There are several standards and technologies developed to address various aspects of image management.

For example, the medical imaging technology picture archiving and communication system (PACS) provides economical storage and access to images from multiple modalities

and the digital imaging and communication medicine (DICOM) standard is used for storing and transmitting medical images. Special techniques for image compression and streaming provide efficient realization of these tasks.

Questions:

1. How many steps does the data acquisition process have? What are they?
2. What goals does the image computing process aim have?
3. What does the term *histogram* mean in this article?
4. What functions does the image management have?
5. What does the acronym MLEM mean in this article?

Unit 7

Voice & Audio

Lesson 19 High Fidelity Audio

Lesson 20 MP3 Audio Compression

Lesson 21 The Mobile Telephone System(Ⅲ)

Passage 1 Audio Encoding

Passage 2 Sound Quality versus Data Rate

Passage 3 Vocoders

Lesson 19 High Fidelity Audio

Audiophiles demand the utmost sound quality, and all other factors are treated as secondary. If you had to describe the mindset in one word, it would be: overkill [1]. Rather than just matching the abilities of the human ear, these systems are designed to exceed the limits of hearing[2]. It's the only way to be sure that the reproduced music is pristine[3]. Digital audio was brought to the world by the compact laser disc, or CD. This was a revolution in music; the sound quality of the CD system far exceeds older systems, such as records and tapes.

The CD Audio System

Figure 7.1 illustrates the surface of a compact laser disc, such as viewed through a high power microscope. The main surface is shiny (reflective of light), with the digital information stored as a series of dark pits burned on the surface with a laser. The information is arranged in a single track that spirals from the outside to the inside, the same as a phonograph record. The rotation of the CD is changed from about 210 to 480 rpm as the information is read from the outside to the inside of the spiral, making the scanning velocity a constant 1.2 meters per second. (In comparison, phonograph records spin at a fixed rate, such as 33, 45 or 78 rpm.) During playback, an optical sensor detects if the surface is reflective or non-reflective, generating the corresponding binary information.

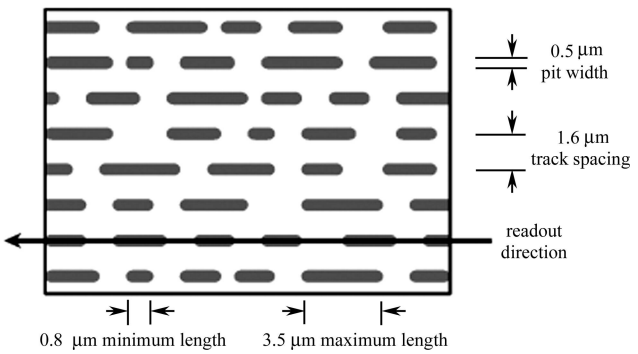


Figure 7.1 CD surface

As shown by the geometry in Figure 7.1, the CD stores about 1 bit per $(\mu m)^2$, corresponding to 1 million bits per $(mm)^2$, and 15 billion bits per disk. This is about the same feature size used in integrated circuit manufacturing, and for a good reason. One of the properties of light is that it cannot be focused to smaller than about one-half wavelength, or

0.3 μm. Since both integrated circuits and laser disks are created by optical means, the fuzziness of light below 0.3 μm limits how small of features can be used.

Figure 7.2 shows a block diagram of a typical compact disc playback system. The raw data rate is 4.3 million bits per second, corresponding to 1 bit each 0.28 μm of track length. However, this is in conflict [4] with the specified geometry of the CD; each pit must be no shorter than 0.8 μm, and no longer than 3.5 μm. In other words, each binary one must be part of a group of 3 to 13 ones. This has the advantage of reducing the error rate due to the optical pickup, but how do you force the binary data to comply with this strange bunching?

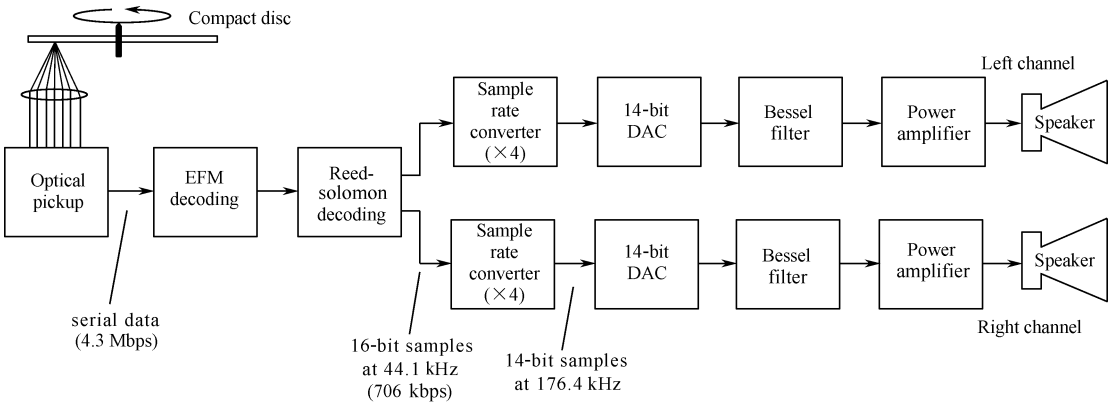


Figure 7.2 A typical CD playback system

The answer is an encoding scheme called eight-to-fourteen modulation (EFM). Instead of directly storing a byte of data on the disc, the 8 bits are passed through a look-up table that pops out 14 bits. These 14 bits have the desired bunching characteristics, and are stored on the laser disc. Upon playback, the binary values read from the disc are passed through the inverse of the EFM look-up table, resulting in each 14-bit group being turned back into the correct 8 bits.In addition to EFM, the data are encoded in a format called two-level Reed-Solomon coding. This involves combining the left and right stereo channels along with data for error detection and correction. Digital errors detected during playback are either: corrected by using the redundant data in the encoding scheme, concealed by interpolating between adjacent samples, or muted by setting the sample value to zero[5]. These encoding schemes result in the data rate being tripled, i.e., 1.4 Mbps for the stereo audio signals versus 4.3 Mbps stored on the disc.

After decoding and error correction, the audio signals are represented as 16-bit samples at a 44.1 kHz sampling rate. In the simplest system, these signals could be run through a 16-bit DAC, followed by a low-pass analog filter[6]. However, this would require high performance analog electronics to pass frequencies below 20 kHz, while rejecting all frequencies above 22.05 kHz, 1/2 of the sampling rate. A more common method is to use a multirate technique, that is, convert the digital data to a higher sampling rate before the DAC. A factor of four is commonly used, converting from 44.1 kHz to 176.4 kHz. This is called interpolation, and can be explained as a two-step process (although it may not actually be carried out this way).

First, three samples with a value of zero are placed between the original samples, producing the higher sampling rate. In the frequency domain, this has the effect of duplicating the 0 to 22.05 kHz spectrum three times, at 22.05 to 44.1 kHz, 44.1 to 66.15 kHz, and 66.15 to 88.2 kHz. In the second step, an efficient digital filter is used to remove the newly added frequencies.

The sample rate increase makes the sampling interval smaller, resulting in a smoother signal being generated by the DAC. The signal still contains frequencies between 20 Hz and 20 kHz; however, the Nyquist frequency has been increased by a factor of four. This means that the analog filter only needs to pass frequencies below 20 kHz, while blocking frequencies above 88.2 kHz. This is usually done with a three-pole Bessel[7] filter.

Since there are four times as many samples, the number of bits per sample can be reduced from 16 bits to 14 bits, without degrading the sound quality. The sinc[8] correction needed to compensate for the zero-order hold of the DAC can be part of either the analog or digital filter.

The Stereo Systems

Audio systems with more than one channel are said to be in stereo (from the Greek word for solid, or three-dimensional). Multiple channels send sound to the listener from different directions, providing a more accurate reproduction of the original music. Music played through a monaural (one channel) system often sounds artificial and bland. In comparison, a good stereo reproduction makes the listener feel as if the musicians are only a few feet away. Since the 1960s, high fidelity music has used two channels (left and right), while motion pictures have used four channels (left, right, center, and surround). In early stereo recordings, individual singers can often be heard in only one channel or the other. This rapidly progressed into a more sophisticated mix-down, where the sound from many microphones in the recording studio is combined into the two channels. Mix-down is an art, aimed at providing the listener with the perception of being there.

The four channel sound used in motion pictures is called Dolby Stereo[9], with the home version called Dolby Surround Pro Logic (as shown in Figure 7.3). The four channels are encoded into the standard left and right channels, allowing regular two-channel stereo systems to reproduce the music. A Dolby decoder is used during playback to recreate the four channels of sound. The left and right channels, from speakers placed on each side of the movie or television screen, is similar to that of a regular two-channel stereo system. The speaker for the center channel is usually placed directly above or below the screen. Its purpose is to reproduce speech and other visually-connected sounds, keeping them firmly centered on the screen, regardless of the seating position of the viewer/listener. The surround speakers are placed to the left and right of the listener, and may involve as many as twenty speakers in a large auditorium. The surround channel only contains midrange frequencies (say, 100 Hz to 7 kHz), and is delayed by 15 to 30 milliseconds. This delay makes the listener perceive that speech is coming from the screen, and not the sides. That is, the listener hears the speech coming from the front, followed by a delayed

version of the speech coming from the sides. The listener's mind interprets the delayed signal as a reflection from the walls, and ignores it.

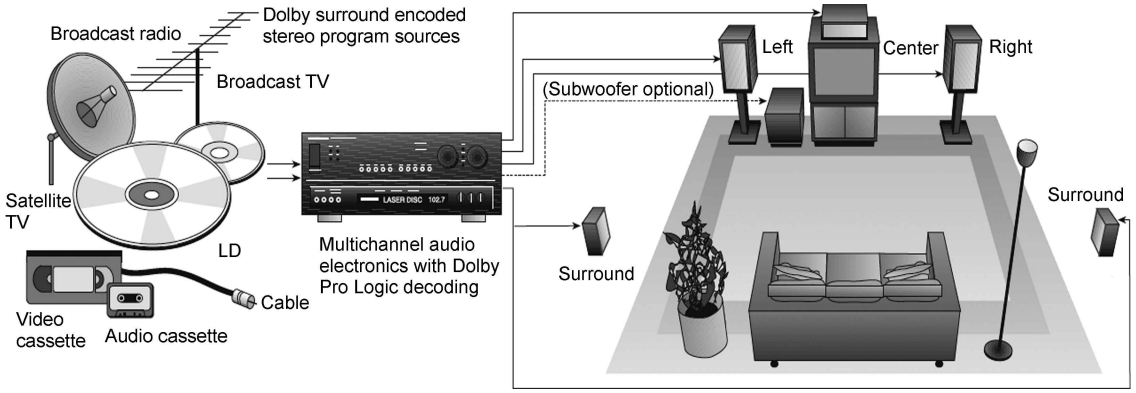


Figure 7.3 The 5.1-channel Dolby Pro Logic II

New Words & Expressions

audiophile [ˈɔːdɪəʊfaɪl] *n*. 高保真音响爱好者
utmost [ˈʌtməʊst] *adj*. 极度的
secondary [ˈsekəndəri] *adj*. 次要的
mindset [ˈmaɪndset] *n*. 心态
overkill [ˈəʊvəkɪl] *n*. 过度杀伤
hearing [ˈhɪərɪŋ] *n*. 听力
reproduce [riːprəˈdjuːs] *vt*. 重现
pristine [ˈprɪstiːn] *adj*. 原始的
microscope [ˈmaɪkrəskəʊp] *n*. 显微镜
shiny [ˈʃaɪni] *adj*. 闪闪发光的
reflective [rɪˈflektɪv] *adj*. 反射的
pit [pɪt] *n*. 坑陷
spiral [ˈspaɪrəl] *n*. 螺旋线
focused [ˈfəʊkəst] *adj*. 聚焦的
fuzzy [ˈfʌzi] *adj*. 模糊的
in conflict with ... 与 冲突
comply with ... 遵从
inverse [ˈɪnvɜːs] *adj*. 相反的
mute [mjuːt] *vt*. 使声音减弱或消失
triple [ˈtrɪpl] *adj*. 3 倍的
artificial [ɑːtɪˈfɪʃl] *adj*. 人造的
bland [blænd] *adj*. 乏味的
auditorium [ɔːdɪˈtɔːrɪəm] *n*. 礼堂
phonograph [ˈfəʊnəʊgræf] *n*. 留声机

Technical Terms

playback [ˈpleɪbæk] *n*. 重放
encoding [ɪnˈkəʊdɪŋ] *n*. 编码
bunching [ˈbʌntʃɪŋ] *n*. 聚束
scanning velocity 扫描速度
feature size 特征尺寸
block diagram 框图
data rate 数据率
look-up table 查找表

Reed-Solomon coding　里德-所罗门编码(RS 编码)
Bessel filter　贝塞尔滤波器
Dolby Stereo　杜比立体声
sinc correction　抽样函数校正
rpm *abbr*. revolutions per minute　每分钟的转数
laser [ˈleɪzə] *abbr*. light amplification by stimulated emission of radiation　激光

Notes

1. 此句含虚拟语气，可译为：假如要用一个词来描述这种心理状态，那就是“过分”。其中，mindset 是指“思维定式”“固有观念”“习惯心理”；overkill 原意是指战场上的“过度杀戮”，这里用其引申含义。
2. 此句可译为：这些系统不是设计得刚好满足人类听觉的需求，而是设计得超越了人类听觉的极限。
3. 此句可译为：这是能够确保再现音乐无任何失真的唯一方法。其中，pristine 是指通过音响系统播放出来的音乐要和原来的现场音乐完全一样，没有任何失真或者畸变。
4. 用作名词时，conflict 分别可以和 with，of，between ... and ...连用，表示“和……的矛盾”“……上的冲突”“……和……之间的争执”。
5. either 用作代词时，用于指代两者之一。如果涉及多于两者的情况，一般应当使用 any。either 用作连词时，可以连接多于两个的从句。本句中使用 either ...，... or ...结构连接了三个具有相同主语的从句。
6. 这两句为隐含的虚拟语气。因为实际当中并没有采用所谓“最简系统”的方案。
7. 弗里德里希　威廉　贝塞尔(Friedrich Wilhelm Bessel，1784—1846)是德国数学家、天文学家。他系统总结了贝塞尔函数(该函数由 Daniel Bernoulli 发现)。贝塞尔滤波器是一类线性相位滤波器，其通带具备“恒定群延迟”特点，常用于高保真音响系统中。
8. 在数字信号处理和通信理论中，归一化抽样函数的定义为 $\mathrm{sinc}(x)=\sin px/px$。在数学中，未归一化抽样函数的定义为 $\mathrm{sinc}(x)=\sin x/x$。
9. “杜比立体声”(Dolby Stereo)，又称“杜比数码”(Dolby Digital)，是“杜比公司”(Dolby Laboratories，Inc.)开发的数字音频编解码系统，于 20 世纪 70 年代中期首先应用于电影院。杜比公司是用于消费类音视产品、娱乐媒体和专业音频的多声道音频技术的全球供应商。

Lesson 20　MP3 Audio Compression

CD-quality audio requires a transmission bandwidth of 1.411 Mbps. Clearly, substantial compression is needed to make transmission over the Internet practical. For this reason, various audio compression algorithms have been developed. Probably the most popular one is MPEG audio, which has three layers (variants), of which MP3 (MPEG audio layer 3) is the

most powerful and best known. Large amounts of music in MP3 format are available on the Internet, not all of it legal, which has resulted in numerous lawsuits from the artists and copyright owners. MP3 belongs to the audio portion of the MPEG video compression standard.

Audio compression can be done in one of two ways. In waveform coding the signal is transformed mathematically by a Fourier transform into its frequency components. The amplitude of each component is then encoded in a minimal way. The goal is to reproduce the waveform accurately at the other end in as few bits as possible.

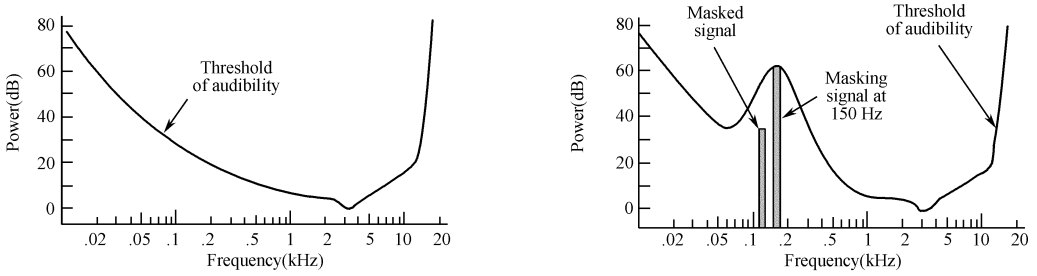
The other way, perceptual coding, exploits certain flaws in the human auditory system to encode a signal in such a way that it sounds the same to a human listener, even if it looks quite different on an oscilloscope. Perceptual coding is based on the science of psychoacoustics — how people perceive sound. MP3 is based on perceptual coding.

The key property of perceptual coding is that some sounds can mask other sounds. Imagine you are broadcasting a live flute concert on a warm summer day. Then all of a sudden, a crew of workmen nearby turn on their jackhammers and start tearing up the street. No one can hear the flute any more. Its sounds have been masked by the jackhammers. For transmission purposes, it is now sufficient to encode just the frequency band used by the jackhammers because the listeners cannot hear the flute anyway. This is called frequency masking — the capability of a loud sound in one frequency band to hide a softer sound in another frequency band that would have been audible in the absence of the loud sound[1]. In fact, even after the jackhammers stop, the flute will be inaudible for a short period of time because the ear turns down its gain when they start and it takes a finite time to turn it up again. This effect is called temporal masking.

To make these effects more quantitative, imagine experiment 1. A person in a quiet room puts on headphones connected to a computer's sound card. The computer generates a pure sine wave at 100 Hz at low, but gradually increasing power. The person is instructed to strike a key when he or she hears the tone. The computer records the current power level and then repeats the experiment at 200 Hz, 300 Hz, and all the other frequencies up to the limit of human hearing. When averaged over many people, a log-log graph of how much power it takes for a tone to be audible looks like that of Figure 7.4(a).

A direct consequence of this curve is that it is never necessary to encode any frequencies whose power falls below the threshold of audibility [2]. For example, if the power at 100 Hz were 20 dB in Figure 7.4(a), it could be omitted from the output with no perceptible loss of quality because 20 dB at 100 Hz falls below the level of audibility.

Now consider Experiment 2. The computer runs experiment 1 again, but this time with a constant-amplitude sine wave at, say, 150 Hz, superimposed on the test frequency [3]. What we discover is that the threshold of audibility for frequencies near 150 Hz is raised, as shown in Figure 7.4(b).



(a) The threshold of audibility as a function of frequency

(b) The masking effect

Figure 7.4 The experiment results explaining perceptual coding

The consequence of this new observation is that by keeping track of which signals are being masked by more powerful signals in nearby frequency bands, we can omit more and more frequencies in the encoded signal, saving bits [4]. In Figure 7.4, the 125-Hz signal can be completely omitted from the output and no one will be able to hear the difference. Even after a powerful signal stops in some frequency band, knowledge of its temporal masking properties allow us to continue to omit the masked frequencies for some time interval as the ear recovers. The essence of MP3 is to Fourier-transform the sound to get the power at each frequency and then transmit only the unmasked frequencies, encoding these in as few bits as possible.

With this information as background, we can now see how the encoding is done. The audio compression is done by sampling the waveform at 32 kHz, 44.1 kHz, or 48 kHz. Sampling can be done on one or two channels, in any of four configurations:

1. Monophonic (a single input stream).
2. Dual monophonic (e.g., an English and a Japanese soundtrack).
3. Disjoint stereo (each channel compressed separately).
4. Joint stereo (inter-channel redundancy fully exploited).

First, the output bit rate is chosen. MP3 can compress a stereo rock'n roll CD down to 96 kbps with little perceptible loss in quality, even for rock'n roll fans with no hearing loss. For a piano concert, at least 128 kbps are needed. These differ because the signal-to-noise ratio for rock'n roll is much higher than for a piano concert (in an engineering sense, anyway). It is also possible to choose lower output rates and accept some loss in quality.

Then the samples are processed in groups of 1152 (about 26 ms worth). Each group is first passed through 32 digital filters to get 32 frequency bands. At the same time, the input is fed into a psychoacoustic model in order to determine the masked frequencies. Next, each of the 32 frequency bands is further transformed to provide a finer spectral resolution.

In the next phase the available bit budget is divided among the bands, with more bits allocated to the bands with the most unmasked spectral power, fewer bits allocated to unmasked bands with less spectral power, and no bits allocated to masked bands.

Finally, the bits are encoded using Huffman encoding, which assigns short codes to numbers that appear frequently and long codes to those that occur infrequently.

There is actually more to the story. Various techniques are also used for noise reduction, anti-aliasing, and exploiting the inter-channel redundancy, if possible, but these are beyond the scope of this article.

New Words & Expressions

lawsuit [ˈlɔːsuːt] *n*. 诉讼
flaw [flɔː] *n*. 缺陷
live [laɪv] *adj*. 实况转播的
flute [fluːt] *n*. 长笛
all of a sudden 突然
jackhammer [ˈdʒækˌhæmə] *n*.手提钻
mask [mɑːsk] *vt*. 掩盖
in the absence of ... 当……不在时
inaudible [ɪnˈlɔːdɪbl] *adj*. 听不见的
quantitative[ˈkwɒntɪˌtətɪv] *adj*. 定量的
superimpose [ˌsuːpərɪmˈpəʊz] *vt*. 叠加
threshold [ˈθreʃəʊld] *n*. 临界值
monophonic [ˌmɒnəʊˈfɒnɪk] *adj*. 单声道的
disjoint [dɪsˈdʒɒɪnt] *vt*. 分开
amplitude [ˈæmplɪtjuːd] *n*. 振幅
resolution [rezəˈluːʃn] *n*. 分辨率
perceptible [pəˈseptɪbl] *adj*. 可察觉的
loss [lɒs] *n*. 损失
budget [ˈbʌdʒɪt] *n*. 预算
allocate [ˈæləkeɪt] *vt*. 分配

Technical Terms

redundancy [rɪˈdʌndənsi] *n*. 冗余
oscilloscope[əˈsɪləˌskəʊp] *n*. 示波器
psychoacoustics [ˌsaɪkəʊəˈkuːstɪks] *n*. 心理声学
transmission bandwidth 传输带宽
waveform coding 波形编码
perceptual coding 知觉编码
frequency masking 频率掩蔽
temporal masking 暂时掩蔽
sound card 声卡
signal-to-noise ratio 信噪比
spectral resolution 频率分辨率
MPEG *abbr*. Motion Picture Experts Group 运动图像专家组

Notes

1. 此句可译为：这种现象称作“频率掩蔽”——某个频率上响度较大的声音能够掩盖另一个频率上响度较小的声音；假如响度大的声音不存在，这个响度小的声音本来是可以听得到的。在句中，修饰 a softer sound in another frequency band 的定语从句 would have been audible in the absence of the loud sound 中含有虚拟语气。in the absence of the loud sound 是假设条件，而 would have been audible 是相应结果。

2. 此句可译为：根据该曲线，可以直接得知完全没必要对功率低于可听门限的频率成分进行编码。
3. 此句可译为：计算机再次运行实验 1，但这次使用一个等幅正弦波(如 150 Hz)叠加到测试频率上。在句中，with a constant-amplitude sine wave at, say, 150 Hz, superimposed on the test frequency 是由 with 引导的独立主格结构。
4. 此句可译为：通过这次新观测可以得出：通过跟踪哪些信号会被邻近频带的更强信号屏蔽，我们就可以在编码信号中忽略更多的频率成分，从而节省数据位数。

Lesson 21 The Mobile Telephone System(Ⅲ)

IMT-2000

In 1992, ITU issued a blueprint for the third generation mobile telephony, which is called IMT-2000. IMT stood for International Mobile Telecommunications. The number 2000 stood for three things: (1) the year it was supposed to go into service, (2) the frequency it was supposed to operate at (in MHz), and (3) the bandwidth the service should have (in kHz).

But nothing was implemented by 2000. ITU recommended that all governments reserve spectrum at 2 GHz so devices could roam seamlessly from country to country. Later, it was recognized that 2 Mbps is not feasible for users who are too mobile (due to the difficulty of performing handoffs [1] quickly enough). More realistic is 2 Mbps for stationary indoor users (which will compete head-on with ADSL), 384 kbps for people walking, and 144 kbps for connections in cars.

The basic services that the IMT-2000 network is supposed to provide to its users are:

1. High-quality voice transmission.
2. Messaging (replacing E-mail, fax, SMS, chat, etc.).
3. Multimedia (playing music, viewing videos, films, television, etc.).
4. Internet access (Web surfing, including pages with audio and video).

Additional services might be video conferencing, telepresence[2], group game playing, and m-commerce. Furthermore, all these services are supposed to be available worldwide (with automatic connection via a satellite when no terrestrial network can be located), instantly (always on), and with quality-of-service guarantees.

ITU envisioned a single worldwide technology for IMT-2000, so that manufacturers could build a single device that could be sold and used anywhere in the world (like CD players and computers and unlike mobile phones and televisions). Having a single technology would also make life much simpler for network operators and would encourage more people to use the services. Format wars, such as the Betamax versus VHS battle when video recorders first came out, are not good for business.

W-CDMA versus CDMA 2000

Several proposals were made, and after some winnowing, it came down to two main ones. The first one, W-CDMA (wideband CDMA), was proposed by Ericsson. This system uses direct sequence spread spectrum (DSSS). It runs in a 5 MHz bandwidth and has been designed to interwork with GSM networks although it is not backward compatible with GSM. It does, however, have the property that a caller can leave a W-CDMA cell and enter a GSM cell without losing the call. This system was pushed hard by the European Union, which called it UMTS (Universal Mobile Telecommunications System).

The other contender was CDMA 2000, proposed by Qualcomm. It, too, is a direct sequence spread spectrum design, basically an extension of IS-95 and backward compatible with it. It also uses a 5-MHz bandwidth, but it has not been designed to interwork with GSM and cannot hand off calls to a GSM cell. Other technical differences with W-CDMA include a different chip rate, different frame time, different spectrum used, and a different way to do time synchronization.

If the Ericsson and Qualcomm engineers were put in a room and told to come to a common design, they probably could. After all, the basic principle behind both systems is CDMA in a 5 MHz channel and nobody is willing to die for his preferred chip rate. The trouble is that the real problem is not engineering. Europe wanted a system that interworked with GSM; the U.S. wanted a system that was compatible with one already widely deployed in the U.S. (IS-95). Each side also supported its local company (Ericsson is based in Sweden; Qualcomm is in California). Finally, Ericsson and Qualcomm were involved in numerous lawsuits over their respective CDMA patents.

In March 1999, the two companies settled the lawsuits when Ericsson agreed to buy Qualcomm's infrastructure. They also agreed to a single 3G standard, but one with multiple incompatible options, which to a large extent just papers over the technical differences.

EDGE and GPRS

While waiting for the fighting over 3G to stop, some operators are gingerly taking a cautious small step in the direction of 3G by going to what is sometimes called 2.5G. One such system is EDGE (enhanced data rates for GSM evolution), which is just GSM with more bits per baud [3]. The trouble is, more bits per baud also means more errors per baud, so EDGE has nine different schemes for modulation and error correction, differing on how much of the bandwidth is devoted to fixing the errors introduced by the higher speed.

Another 2.5G scheme is GPRS (general packet radio service), which is an overlay packet network on top of D-AMPS or GSM. It allows mobile stations to send and receive IP packets in a cell running a voice system. When GPRS is in operation, some time slots on some frequencies are reserved for packet traffic. The number and location of the time slots can be dynamically managed by the base station, depending on the ratio of voice to data traffic in the cell.

The available time slots are divided into several logical channels, used for different purposes. The base station determines which logical channels are mapped onto which time slots. One logical channel is for downloading packets from the base station to some mobile station, with each packet indicating who it is destined for. To send an IP packet, a mobile station requests one or more time slots by sending a request to the base station. If the request arrives without damage, the base station announces the frequency and time slots allocated to the mobile for sending the packet. Once the packet has arrived at the base station, it is transferred to the Internet by a wired connection.

3G, 4G and 5G

In October 2001, the first commercial 3G network was launched in Japan. In Europe, the first 3G commercial network was opened in December 2001.In South Korea, 3G network went commercially in January 2002 … Gradually, 3G was adopted around the world.

In March 2008, the International Telecommunications Union-Radio communications sector (ITU-R) specified a set of requirements for 4G standards, named the International Mobile Telecommunications Advanced (IMT-Advanced) specification, setting peak speed requirements for 4G service at 100 Mbps for high mobility communication (such as from trains and cars) and 1 Gbps for low mobility communication (such as pedestrians and stationary users).

In early 2012, ITU-R embarked on a program to develop "IMT for 2020 and beyond", setting the stage for 5G research activities that are emerging around the world.

In March 2015, the NGMN (next generation mobile networks) White Paper on 5G was published with the vision: "5G is an end-to-end ecosystem to enable a fully mobile and connected society. It foresees 4G and its evolution as a component part of the 5G revolution."

New Words & Expressions

roam [rəʊm] *vi*. 漫游
seamlessly [ˈsiːmlɪsli] *adv*. 无缝地
head-on [ˈhedɒn] *adv*. 正面地
stationary [ˈsteɪʃənəri] *adj*. 静止的
paper over … 掩盖……
gingerly [ˈdʒɪndʒəli] *adv*. 慎重地
overlay [əʊvəˈleɪ] *vt*. 叠加
destined [ˈdestɪnd] *adj*. 去往……的
deploy [dɪˈplɔɪ] *vt*. 部署
sector [ˈsektə] *n*. 部门
specify [ˈspesɪfaɪ] *vt*. 详细说明
preferred [prɪˈfɜːd] *adj*. 首选的
numerous [ˈnjuːmərəs] *adj*. 许多的
patent [ˈpætənt] *n*. 专利
infrastructure [ˈɪnfrəstrʌktʃə] *n*. 基础设施
incompatible [ɪnkəmˈpætɪbl] *adj*. 不兼容的
terrestrial [təˈrestrɪəl] *adj*. 地面的
envision [enˈvɪʒən] *vt*. 想象
winnow [ˈwɪnəʊ] *vt*. 挑选
come down to … 归结为……
interwork [ɪntəˈwɜːk] *vi*. 交织工作
come to … 得出……
requirement [rɪˈkwaɪəmənt] *n*. 要求
pedestrian [pəˈdestrɪən] *n*. 步行者

Technical Terms

handoff [ˈhændɒf] *n*. 越区切换
telepresence [ˈtelɪˌprezəns] *n*. 远程在位
wideband [ˈwaɪdˈbænd] *adj*. 宽频带的
synchronization [ˌsɪŋkrənaɪˈzeɪʃn] *n*. 同步
baud [bɔːd] *n*. 波特
packet [ˈpækɪt] *n*. 信息分组
video conference　视频会议
m-commerce　移动商务
network operator　网络运营商
backward compatible　向下兼容
chip rate　码片速率
IP packet　IP 分组
time slot　时隙
logical channel　逻辑通道
base station　基站
ADSL *abbr*. asymmetrical digital subscriber line　非对称数字用户线路
SMS *abbr*. short message service　短信息服务
QoS *abbr*. quality-of-service　服务质量
VHS *abbr*. Video Home System　家用录像系统
DSSS *abbr*. direct sequence spread spectrum　直序扩频
IP *abbr*. Internet Protocol　互联网协议

Notes

1. “越区切换”(handoff)是指当移动电话用户进入新的蜂窝区域时，信号传输将从一个蜂窝切换到另一个蜂窝。这个切换过程在大约 0.25 s(秒)内完成，这样通话者一般就不会注意到这种切换了。
2. telepresence 是“虚拟现实”(virtual reality)技术的一种。该技术要创建一个多维信息构成的可操纵空间。虚拟现实技术最重要的目标就是实现真实体验和自然的人机交互，能够达到或者部分达到这样目标的系统就可称为虚拟现实系统。虚拟现实技术的其他形式还有“人工现实”(artificial reality)、“虚拟环境”(virtual environment)、“赛博空间或网络空间”(cyberspace)等。
3. “波特”(baud)源自法国工程师 Jean Maurice Emile(1845—1903)。波特是一种数据传输速率单位，等于每秒内出现的离散状态或离散信号数。在二进制信号序列中，1 波特等于 1 比特每秒。

Exercises

1. Fill in the blanks with proper words, phrases or clauses.

(1)The primary advantage of having two ears is the ability to _____(辨别声音的方向). Human listeners can detect the difference _____ two sound sources that are placed as little as three degrees apart, about the width of a person at 10 meters. This directional information is obtained _____ two separate ways. First, _____(约 1 kHz 以上的频率) are strongly shadowed by the head. In other words, the ear nearest the sound receives a stronger signal _____ the ear _____(在头的另一侧). The second clue to directionality is that the ear on the far side of the head hears the sound _____(比……稍迟些)the near ear, _____(由于它距离音源更远). _____(根据)a typical head size (about 22 cm) and the speed of sound (about 340 meters per second), an angular discrimination of three degrees requires a timing precision of about 30 microseconds. Since this timing requires the volley principle, this clue to directionality is predominately used for sounds less than about 1 kHz.

Both these sources of directional information are greatly aided by the ability to turn the head and observe the change in the signals. An interesting sensation occurs when a listener is presented _____ exactly the same sounds to both ears, such as listening to monaural sound through headphones. The brain concludes that the sound is _____(来自听者头部的中央)!

_____(尽管)human hearing can determine the direction a sound is _____, it does poorly in identifying the distance _____ the sound source. _____(这是因为)there are few clues available in a sound wave _____(可以提供这种信息). Human hearing weakly perceives that high frequency sounds are nearby, while low frequency sounds are distant. This is because sound waves dissipate their higher frequencies _____ they propagate long distances. Echo content is another weak clue to distance, providing a perception of the room size. For example, sounds in a large auditorium will contain echoes at about 100 millisecond intervals, while 10 milliseconds is typical for a small office. Some species have solved this ranging problem _____ using active sonar. For example, bats and dolphins produce clicks and squeaks that reflect _____ nearby objects. _____(通过测量发送信号和回波信号之间的时间间隔), these animals can locate objects _____ about 1 cm resolution. Experiments have shown that some humans, particularly the blind, can also use active _____(回声定位)to a small extent.

(2)In a CD (and any other digital recording technology), the goal is _____ create a recording with very high _____(保真度)(very high similarity _____ the original signal and the reproduced signal) and perfect reproduction (the recording sounds the same every single time you play it no matter how many times you play it).

_____ (为了实现这两个目标), digital recording converts the _____(模拟波形)into a stream of numbers and records the numbers instead of the wave. The conversion is done by _____(一种叫作模数转换器的器件). To play back the music, the stream of numbers is

converted back to an analog wave _____ a digital-to-analog converter (DAC). The analog wave produced by the DAC _____ amplified and fed to the speakers to produce the sound.

The analog wave produced by the DAC will be the same every time, _____(只要)the numbers are not corrupted. The analog wave produced by the DAC will also _____(和……非常相似)the original analog wave if the analog-to-digital converter sampled _____ a high rate and produced accurate numbers.

You can understand why CDs have such high fidelity if you understand the analog-to-digital conversion process better. Let s say you have a sound wave, and you wish to sample it with an ADC. When you sample the wave with an analog-to-digital converter, you have control over two variables:

- **The sampling rate**—Controls how many samples are taken per second.
- **The sampling precision**—Controls how many different gradations (quantization levels) are possible when taking the sample.

One thing about the CD s sampling rate and precision is _____ it produces a lot of data. On a CD, the digital numbers produced by the ADC _____(以字节形式存储), and it takes 2 bytes to represent 65,536 gradations. There are two sound streams being recorded (one for each of the speakers on a stereo system). _____(一张光盘可以存储多达 74 分钟的音乐), so the total amount of digital data that must be stored on a CD is: 44,100 samples/(channel × second) × 2 bytes/sample × 2 channels × 74 minutes × 60 seconds/minute = 783,216,000 bytes. That is a lot of bytes! To store that many bytes on a cheap piece of plastic that is tough enough to survive the abuse most people put a CD through is no small task, especially when you consider that the first CDs came out in 1980.

2. Translate the following passages into Chinese or English.

(1) The range of human hearing is generally considered to be 20 Hz to 20 kHz, but it is far more sensitive to sounds between 1 kHz and 4 kHz. For example, listeners can detect sounds as low as 0 dB SPL at 3 kHz, but require 40 dB SPL at 100 Hz (an amplitude increase of 100). Listeners can tell that two tones are different if their frequencies differ by more than about 0.3% at 3 kHz. This increases to 3% at 100 Hz. For comparison, adjacent keys on a piano differ by about 6% in frequency.

(2) The term octave means a factor of two in frequency. On the piano, one octave comprises eight white keys, accounting for the name (octo is Latin for eight). In other words, the piano s frequency doubles after every seven white keys, and the entire keyboard spans a little over seven octaves. The range of human hearing is generally quoted as 20 Hz to 20 kHz, corresponding to about 1/2 octave to the left, and two octaves to the right of the piano keyboard.

(3) Companding can be carried out in three ways: ① run the analog signal through a nonlinear circuit before reaching a linear 8 bit ADC, ② use an 8 bit ADC that internally has unequally spaced steps, or ③ use a linear 12 bit ADC followed by a digital look-up table (12 bits in, 8 bits out). Each of these three options requires the same nonlinearity, just in a

different place: an analog circuit, an ADC, or a digital circuit.

(4) The perception of a continuous sound, such as a note from a musical instrument, is often divided into three parts: loudness, pitch, and timbre (pronounced "timber"). Loudness is a measure of sound wave intensity. Pitch is the frequency of the fundamental component in the sound. While there are subtle effects in both these perceptions, they are a straightforward match with easily characterized physical quantities. Timbre is more complicated, being determined by the harmonic content of the signal.

(5) Nearly all techniques for speech synthesis and recognition are based on the model of human speech production. Most human speech sounds can be classified as either voiced or fricative. Voiced sounds occur when air is forced from the lungs, through the vocal cords, and out of the mouth and/or nose. The vocal cords are two thin flaps of tissue stretched across the air flow, just behind the Adam's apple. In response to varying muscle tension, the vocal cords vibrate at frequencies between 50 and 1000 Hz, resulting in periodic puffs of air being injected into the throat. Vowels are an example of voiced sounds.

(6) 音频信号就是一维声波。当声波进入耳朵后，鼓膜振动并引发内耳小骨振动，将神经脉冲发送给大脑。收听者将这些脉冲感知为声音。

(7) 音乐是一种音频信号，是一种非常重要的音频信号。语音是另一种重要的音频信号。人类语音通常在 600 ～ 6000 Hz 范围内。语音分为元音和辅音，二者的特点不同。

(8) 由于样本为 16 bit、采样率为 44,100 样本/秒，音频 CD 所需的单声道带宽为 705.6 kbps，立体声带宽为 1.411 Mbps。尽管这比视频带宽低，但实时传输未压缩 CD 音质立体声仍要占用几乎整个 T1 信道。

(9) 电话系统中的"脉冲编码调制"(PCM)使用每秒 8000 个 8 bit 样本。在北美和日本，7 bit用于数据，1 bit 用于控制；在欧洲，所有 8 bit 均用于数据。PCM 系统的数据率为 56000 bps 或 64000 bps。由于采样率只有 8000 样本/秒，所以 4kHz 以上的频率就丢掉了。

(10) 计算机通过软件可以很容易地处理数字化声音。有几十个用于个人计算机的程序允许用户录制、显示、编辑、混合和存储来自多种音源的声波。事实上，目前所有的专业化录音和编辑都是数字化的。

3. Writing tasks.

Summary: Write an abstract for *Lesson* 19 *High Fidelity Audio* within 150～200 words.

Composition: Write an essay about the experience you had when using digital audio gadgets.

Reading Materials

Passage 1 Audio Encoding

An audio signal is a one-dimensional acoustic wave. When an acoustic wave enters the ear, the eardrum vibrates, causing the tiny bones of the inner ear to vibrate along with it, sending nerve pulses to the brain. These pulses are perceived as sound by the listener. In a similar way, when an acoustic wave strikes a microphone, the microphone generates an electrical signal, representing the sound amplitude as a function of time. A one-second segment of a voice signal is plotted in Figure 7.5. Such a plot is often called a waveform.

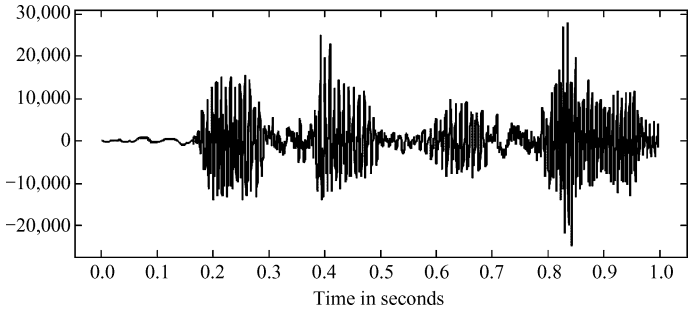


Figure 7.5 Waveform of a speech fragment

The frequency range of the human ear runs from 20 Hz to 20,000 Hz (see Table 7.1). The ear hears logarithmically, so the ratio of two sounds with amplitudes A and B is conventionally expressed in dB (decibels) according to the formula $dB = 20\log_{10}(A/B)$.

Table 7.1 The frequency range of the human ear

Acoustic Power	Sound Power Level	Example Sound
10^{-2} watts/cm^2	140 dB	Pain
10^{-3} watts/cm^2	130 dB	
10^{-4} watts/cm^2	120 dB	Discomfort
10^{-5} watts/cm^2	110 dB	Jack hammers, rock concerts
10^{-6} watts/cm^2	100 dB	
10^{-7} watts/cm^2	90 dB	OSHA limit for industrial noise
10^{-8} watts/cm^2	80 dB	
10^{-9} watts/cm^2	70 dB	
10^{-10} watts/cm^2	60 dB	Normal conversation
10^{-11} watts/cm^2	50 dB	
10^{-12} watts/cm^2	40 dB	Weakest audible at 100 Hz
10^{-13} watts/cm^2	30 dB	
10^{-14} watts/cm^2	20 dB	Weakest audible at 10 kHz

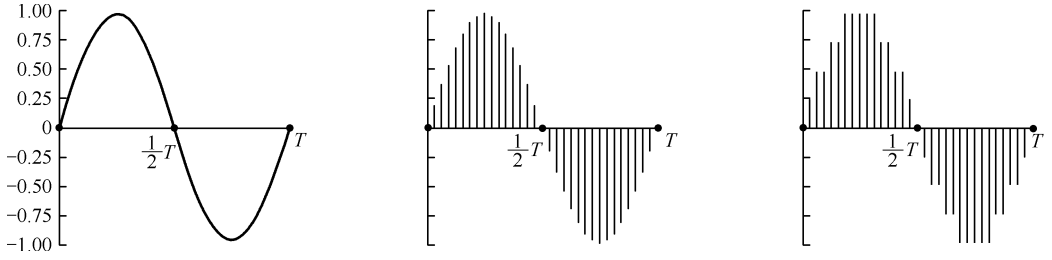
续表

Acoustic Power	Sound Power Level	Example Sound
10^{-15} watts/cm^2	10 dB	
10^{-16} watts/cm^2	0 dB	Weakest audible at 3 kHz

If we define the lower limit of audibility (a pressure of about 0.0003 dyne/cm^2) for a 1-kHz sine wave as 0 dB, an ordinary conversation is about 50 dB and the pain threshold is about 120 dB, a dynamic range of a factor of more than 1 million. To avoid any confusion, A and B above are amplitudes. If we were to use the power level, which is proportional to the square of the amplitude, the coefficient of the logarithm would be 10, not 20.

Digital audio is a digital representation of an audio wave that can be used to recreate it. Audio waves can be converted to digital form by an ADC (analog-to-digital converter). An ADC takes an electrical voltage as input and generates a binary number as output. In Figure 7.6(a) we see an example of a sine wave. To represent this signal digitally, we can sample it every ΔT seconds, as shown by the bar heights in Figure 7.6(b). If a sound wave is not a pure sine wave, but a linear superposition of sine waves where the highest frequency component present is f then it is sufficient to make samples at a frequency $2f$. Sampling more often is of no value since the higher frequencies that such sampling could detect are not present. The reverse process takes digital values and produces an analog electrical voltage. It is done by a DAC (digital-to-analog converter). A loudspeaker can then convert the analog voltage to acoustic waves so that people can hear sounds.

Digital samples are never exact. The samples of Figure 7.6(c) allow only nine values, from −1.00 to +1.00 in steps of 0.25. Consequently, 4 bits are needed to represent all of them. An 8-bit sample would allow 256 distinct values. A 16-bit sample would allow 65,536 distinct values. The error introduced by the finite number of bits per sample is called the quantization noise. If it is too large, the ear detects it.



(a) A sine wave (b) Sampling the sine wave (c) Quantizing the samples to 4 bits

Figure 7.6 The sampling and quantization of a sine wave

Two well-known examples of sampled sound are the telephone and audio CD. Pulse code modulation is used within the telephone system and uses 7-bit (North America and Japan) or 8-bit (Europe) samples 8000 times per second. This system gives a data rate of 56 kbps or 64

kbps. With only 8000 samples/sec, frequencies above 4 kHz are lost.

Audio CDs are digital with a sampling rate of 44,100 samples/sec, enough to capture frequencies up to 22,050 Hz. The samples are 16 bits each, and are linear over the range of amplitudes. Note that 16-bit samples allow only 65,536 distinct values, even though the dynamic range of the ear is more than 1 million. Thus using only 16 bits per sample introduces some quantization noise. With 44,100 samples/sec of 16 bits each, an audio CD needs a bandwidth of 705.6 kbps for monaural and 1.411 Mbps for stereo. Audio compression is possible based on psychoacoustic models of how human hearing works. A compression of 10× is possible using the MPEG layer 3 (MP3) system.

Questions:

1. What is the frequency range of human hearing?
2. What is the dynamic range (in dB) of human hearing?
3. What device do we use when converting audio waves to digital form?
4. What device do we use when converting digital values to an analog electrical voltage?
5. What does the term *quantization noise* mean in this article?

Passage 2 Sound Quality versus Data Rate

When designing a digital audio system there are two questions that need to be asked: (1) how good does it need to sound? and (2) what data rate can be tolerated? The answer to these questions usually results in one of three categories. First, high fidelity music, where sound quality is of the greatest importance, and almost any data rate will be acceptable. Second, telephone communication, requiring natural sounding speech and a low data rate to reduce the system cost. Third, compressed speech, where reducing the data rate is very important and some unnaturalness in the sound quality can be tolerated. This includes military communication, cellular telephones, and digitally stored speech for voice mail and multimedia.

Table 7.2 shows the tradeoff between sound quality and data rate for these three categories. High fidelity music systems sample fast enough (44.1 kHz), and with enough precision (16 bits), that they can capture virtually all of the sounds that humans are capable of hearing. This magnificent sound quality comes at the price of a high data rate, 44.1 kHz × 16 bits = 706 kbps. This is pure brute force.

Table 7.2 Sound quality versus data rate

Sound Quality	Bandwidth	Sampling rate	Number of bits	Data rate
High fidelity music (CD)	5 Hz to 20 kHz	44.1 kHz	16 bit	706 kbps
Telephone quality speech	200 Hz to 3.2 kHz	8 kHz	12 bit	96 kbps
Telephone quality speech(with companding)	200 Hz to 3.2 kHz	8 kHz	8 bit	64 kbps
Speech encoded by LPC	200 Hz to 3.2 kHz	8 kHz	12 bit	4 kbps

Whereas music requires a bandwidth of 20 kHz, natural sounding speech only requires about 3.2 kHz. Even though the frequency range has been reduced to only 16% (3.2 kHz out of 20 kHz), the signal still contains 80% of the original sound information (8 out of 10

octaves). Telecommunication systems typically operate with a sampling rate of about 8 kHz, allowing natural sounding speech, but greatly reduced music quality. You are probably already familiar with this difference in sound quality: FM radio stations broadcast with a bandwidth of almost 20 kHz, while AM radio stations are limited to about 3.2 kHz. Voices sound normal on the AM stations, but the music is weak and unsatisfying.

Voice-only systems also reduce the precision from 16 bits to 12 bits per sample, with little noticeable change in the sound quality. This can be reduced to only 8 bits per sample if the quantization step size is made unequal. This is a widespread procedure called companding. There are two companding schemes in common use. In North America, the μ-Law compander is common in telephony; Europe uses the A-Law compander. Both are similar in their calculation.

The μ-Law is defined as

$$y = \frac{\ln(1+\mu x)}{\ln(1+\mu)} \quad \text{where } \mu = 255 \text{ and } 0 < x < 1$$

The A-Law compander uses

$$y = \begin{cases} \dfrac{1+\ln(Ax)}{1+\ln A} & \text{for } 0 < x < \dfrac{1}{A}, \text{ with } A = 87.6 \\ \dfrac{Ax}{1+\ln A} & \text{for } \dfrac{1}{A} < x < 1, \text{ with } A = 87.6 \end{cases}$$

An 8 kHz sampling rate, with an ADC precision of 8 bits per sample, results in a data rate of 64 kbps. This is the brute force data rate for natural sounding speech. Notice that speech requires less than 10% of the data rate of high fidelity music.

The data rate of 64 kbps represents the straightforward application of sampling and quantization theory to audio signals. Techniques for lowering the data rate further are based on compressing the data stream by removing the inherent redundancies in speech signals. One of the most efficient ways of compressing an audio signal is Linear Predictive Coding (LPC), of which there are several variations and subgroups. Depending on the speech quality required, LPC can reduce the data rate to as little as 2～6 kbps.

LPC uses a model of human speech generation as the basis of compression. Over short intervals, about 2 to 40 milliseconds, human speech can be modeled using three parameters, as shown in Figure 7.7. The first parameter is a choice of sound source, either random noise or a pulse train. The noise source corresponds to the frictive sound of a consonant, such as [s], [f], and [v]. The pulse source corresponds to voiced sounds such as vowels. Because voiced sounds may vary in pitch, the model includes a frequency for the pulse source as the second parameter. The third parameter is a set of coefficients for a recursive linear filter that models the acoustic response of the mouth, throat, and nasal passages during that moment of speech.

This model can be used to synthesize speech, although it sounds mechanical and not human. It is also the core of code-excited linear predictive (CELP) coding. The voice quality that results from CELP coding is somewhat less than that of other compression methods, but it still offers adequate voice fidelity for normal conversations. The advantage of CELP,

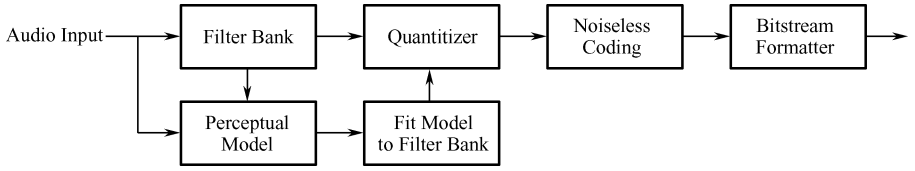


Figure 7.7 The general structure of a perceptual audio codec

however, outweighs the slight loss of sound quality. CELP codecs can compress voice signals to as little as 4.8 kbps.

Questions:

1. Does the sound quality of a digitized audio signal depend on its data rate?
2. What is data rate and how can it be figured out?
3. Which audio system has a letter sound quality, a CD player or a mobile phone?
4. Why is the quality of music poor on AM stations while voices sound good?
5. Do you know any application of compressed speech?

Passage 3 Vocoders

The PCM coding and decoding processes are concerned with reproducing waveforms as accurately as possible. The precise nature of the wave is unimportant as long as it occupies the voice-band frequency range. When digitizing speech signals only, special voice encoders/decoders called vocoders are often used, which are based on human's speech generation model (see Figure 7.8). To achieve acceptable speech communications, the short-term power spectrum of speechis all that must be preserved. The human ear is relatively insensitive to the phase relationship between individual frequency components within a voice waveform. Therefore, vocoders are designed to reproduce only the short-term power spectrum, and the decoded time waveforms often only vaguely resemble the original input signal. Vocoders cannot be used in applications where analog signals other than voice are present, such as output signals from voice-band data modems. Vocoders typically produce unnatural sounding speech and, therefore, are generally used for recorded information such as "wrong number" messages, encrypted voice for transmission over analog telephone circuits, computer output signals, and educational games.

The purpose of a vocoder is to encode the minimum amount of speech information necessary to reproduce a perceptible message with fewer bits than those needed by a conventional encoder/decoder. Vocoders are used primarily in limited bandwidth applications. Essentially, there are four vocoding techniques available: the channel vocoder, the formant vocoder, the cepstrum vocoder and the linear predictive coder.

Channel Vocoders The first channel vocoder was developed by Homer Dudley in 1928. Dudley's vocoder compressed conventional speech waveforms into an analog signal with a total bandwidth of approximately 300 Hz. Present-day digital vocoders operate at less than 2 kbps. Digital channel vocoders use bandpass filters to separate the speech waveform into narrower

subbands. Each subband is full-wave rectified, filtered, then digitally encoded. The encoded signal is transmitted to the destination receiver, where it is decoded. Generally speaking, the quality of the signal at the output of a vocoder is quite poor. However, some of the more advanced channel vocoders operate at 2400 bps and can produce a highly intelligible, although slightly synthetic sounding speech.

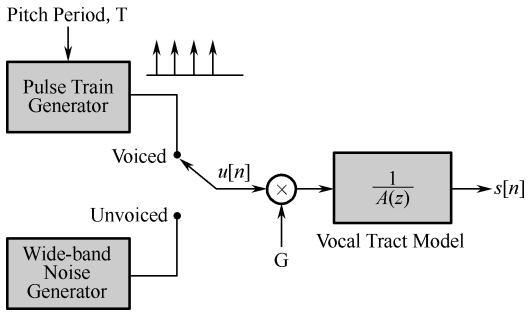


Figure 7.8 The speech generation model

Formant Vocoders A formant vocoder takes advantage of the fact that the short-term spectral density of typical speech signals seldom distributes uniformly across the entire voice-band spectrum (300 Hz to 3000 Hz). Instead, the spectral power of most speech energy concentrates at three or four peak frequencies called formants, A formant vocoder simply determines the location of these peaks and encodes and transmits only the information with the most significant short-term components. Therefore, formant vocoders can operate at lower bit rates and, thus, require narrower bandwidths. Formant vocoders sometimes have trouble tracking changes in the formants. However, once the formants have been identified, a formant vocoder can transfer intelligible speech at less than 1000 bps.

Cepstrum Vocoders The cepstrum vocoder separates the excitation and vocal tract spectrum by inverse Fourier transforming of the log magnitude spectrum to produce the cepstrum of the signal. The low frequency coefficients in the cepstrum correspond to the vocal tract spectral envelope, with the high frequency coefficients forming a periodic pulse train at multiples of the sampling period. Linear filtering is performed to separate the vocal tract cepstral coefficients from the excitation coefficients. In the receiver, the vocal tract cepstral coefficients are Fourier transformed to produce the vocal tract impulse response. By convolving this impulse response with a synthetic excitation signal (random noise or periodic pulse train), the original speech is reconstructed.

Linear Predictive Coders A linear predictive coder extracts the most significant portions of speech information directly from the time waveform rather than from the frequency spectrum as with the channel and formant vocoders. A linear predictive coder produces a time-varying model of the vocal tract excitation and transfer function directly from the speech waveform. At the receive end, a synthesizer reproduces the speech by passing the specified excitation through a mathematical model of the vocal tract. Linear predictive coders provide more natural sounding speech than either the channel or formant vocoder. Linear predictive

coders typically encode and transmit speech at between 1.2 kbps and 2.4 kbps.

Questions:

1. What does the acronym PCM stand for?
2. What is the purpose of a vocoder?
3. What is the mathematical model for vocal tract according to Figure 7.8?
4. What does the term *formant* mean in this article?
5. Can you give the data rate range of linear predictive coders?

Unit 8

Image & Video

Lesson 22 Digital Image Fundamentals

Lesson 23 Digital Cameras

Lesson 24 Television Video Signals

Passage 1 JPEG

Passage 2 MPEG

Passage 3 Internet Video

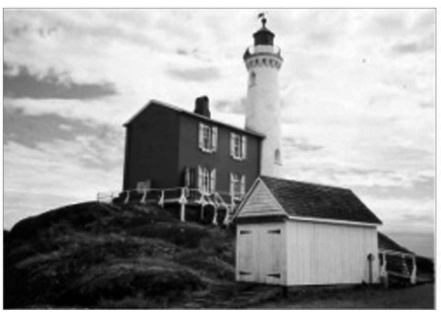
Lesson 22 Digital Image Fundamentals

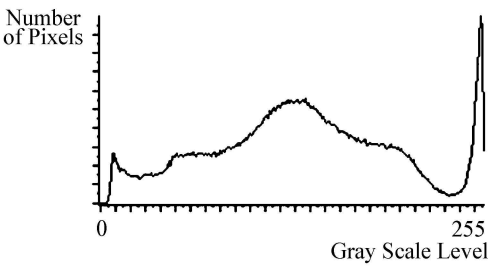
Digital Image Resolution

A digital image is made up of many rows and columns of pixels. For gray scale images, each pixel is assigned a number that represents the gray shade assigned to that pixel. The larger the number of pixels in an image, and the larger the number of available gray scale levels, the better the resolution of the image. Figure 8.1(a) shows an 8-bit digital image, with $2^8=256$ possible gray scale levels. The number of row and column are 808×562.

Histograms

The gray scales present in a digital image can be summarized by its histogram, as shown in Figure 8.1(b). The histogram reports the number of pixels for each gray scale level present in the image, as a bar graph. When an image uses only a small portion of the available gray scale levels, histogram equalization can be used to spread out the usage of gray scale levels over the entire available range. This procedure re-assigns gray scale levels so that image contrast is improved.





(a) An 808 row×562 column×8 bit digital image (b) Histogram of the left image

Figure 8.1 Digital image and histogram

Addition and Subtraction of Images

Digital images can be added and subtracted pixel-by-pixel. Adding two images can combine two sets of objects into a single image. Moreover, several noisy images of the same scene can be averaged together to reduce the effect of noise. Image subtraction, on the other hand, can be used to remove an unwanted background from an image. Subtraction of two time-lapsed photographs will show where motion has occurred between the two.

When two images are added or subtracted, the resultant matrix will frequently contain

illegal gray scale values. For example, when a pixel in one 8-bit image has the gray scale level 129 and the corresponding pixel in a second 8-bit image has the gray scale level 201, the sum pixel is 129+201=330. This is outside the legal range for an 8-bit image, which may only contain gray scale levels between 0 and 255. When the same two images are subtracted, the difference pixel is 129-201= -72, again a value outside the legal range. For these reasons, scaling follows most image arithmetic. Scaling to the range [0, GSLmax] may be accomplished as follows:

$$\text{Scaled Output}[m,n]=\frac{\text{Output}[m,n]-\text{Min(Output)}}{\text{Max(Output)}-\text{Min(Output)}}\text{GSL}_{\max}$$

Warping and Morphing

Warping and morphing are digital image techniques that are finding application not only in entertainment but also in medical imaging. Warping stretches or re-shapes an object in an image, while morphing transforms one image into another. These transformations may be accomplished by marking control points, control lines, or triangles in a source image and choosing their new positions in a destination image. The transition between source and destination images is then accomplished by smoothly transforming not only the control element locations, but also their colors. The locations and colors of pixels not explicitly marked as control elements are determined by the locations and colors of the control elements that are nearest.

Image Filtering

Digital images can be filtered using two-dimensional convolution with a convolution kernel. When an $N \times N$ image is filtered by an $M \times M$ convolution kernel, $(M-1)/2$ rows and columns on each side of the image are lost to boundary effects. Low-pass filters blur images, high-pass filters emphasize sharp changes in gray scale level, and edge filters locate edges in an image.

Dilation and Erosion

Dilation adds a layer of pixels to all objects in an image. Erosion removes one layer of pixels from all objects. When dilation is followed by erosion, gaps in broken boundaries identified through edge detection can be filled in. Conversely, when erosion is followed by dilation, spots of noise in an image are removed.

Successfully detecting the edges in an image is the first step towards confident identification of object boundaries and then object recognition. From boundary information, shape characteristics like perimeter and area can be calculated, which can be used to classify an object.

Image Spectra

Two-dimensional FFTs are used to analyze the spectra of digital images. Just as in the

one-dimensional case, a two-dimensional spectrum comprises a magnitude spectrum and a phase spectrum. The phase spectrum carries the best information about the locations of the objects in the image. When all magnitudes are set to one, the phases alone still show a facsimile of the original image. When all phases are set to zero, the magnitudes alone show no trace of it.

Image spectra form the basis for both CT (computerized tomography) and MRI (magnetic resonance imaging) scan displays. CT scans are X-rays taken in many directions in a single plane of an object. MRI scans depend instead on the magnetic properties of an object placed in a varying magnetic field. Both types of scans permit non-invasive investigations of three-dimensional objects.

Image Compression

In part due to the Internet, digital images are transmitted from place to place more often than ever. To save time and bandwidth (space), both images and other files are often compressed before being transmitted. Lossless compression means that a file is compacted without losing any information, so that the reconstructed file is identical to the original. Lossy compression means that some information from the original file is irretrievably lost, but generally the reconstructed file is extremely close to the original. The compression ratio is the ratio of the original file size to the compressed file size.

One simple compression scheme is run-length encoding, which codes more than three repetitions of a number as three copies of the number followed by a count of the additional copies needed. Another compression scheme is Huffman encoding[1], which uses shorter codes to represent the most common signal elements, and longer codes to represent the least common signal elements.

JPEG, an extremely common image compression scheme, uses the discrete cosine transform (DCT)[2] to concentrate most of the information about an 8×8 sub-block of an image into a few significant coefficients. It then uses both run-length encoding and Huffman encoding to provide further compression.

New Words & Expressions

row [rəʊ] *n*. 行
column [ˈkɒləm] *n*. 列
equalization [ˌɪkwəlɪˈzeɪʃn] *n*. 均衡
matrix [ˈmeɪtrɪks] *n*. 矩阵
arithmetic [əˈrɪθmətɪk] *n*. 算术
warping [ˈwɔːpɪ] *n*. 扭曲
morphing [ˈmɔːfɪ] *n*. 变态,变形
dilation [daɪˈleɪʃn] *n*. 膨胀
erosion [ɪˈrəʊʒən] *n*. 腐蚀
blur [blɜː] *vt*. 使……模糊
kernel [ˈkɜːnəl] *n*. 核心
identification [aɪˌdentɪfɪˈkeɪʃn] *n*. 识别
perimeter [pəˈrɪmɪtə] *n*. 周长
facsimile [fækˈsɪmɪli] *n*. 复制本,传真
tomography [təˈmɒgrəfi] *n*. 层析成像
lossy [ˈlɔsi] *adj*. 有损的

Technical Terms

pixel [ˈpɪksəl] *n*. 像素
shade [ʃeɪd] *n*. 明暗度
histogram [ˈhɪstəgræm] *n*. 直方图
gray scale level　灰度级
bar graph　柱状图
histogram equalization　直方图均衡
image contrast　图像对比度
edge detection　边缘检测
object recognition　目标识别
magnitude spectrum　幅度谱
phase spectrum　相位谱
run-length encoding　行程编码
Huffman encoding　哈夫曼编码
CT *abbr*. computerized tomography　计算机层析成像
MRI *abbr*. magnetic resonance imaging　磁共振成像
DCT *abbr*. discrete cosine transform　离散余弦变换

Notes

1. "哈夫曼编码"(Huffman encoding)是一种用于无损数据压缩的熵编码算法，是于 1952 年由哈夫曼(David A. Huffman)在其论文 *A Method for the Construction of Minimum Redundancy Codes* 中提出的。
2. "离散余弦变换"(discrete cosine transform，DCT)是 Ahmed N. T. Natarajan 和K. R. Rao于 1974 年提出的。1984 年，W. H. Chen 和 W. K. Pratt 首先将其应用到图像压缩中。

Lesson 23　Digital Cameras

Image Sensors[1]

Instead of film，a digital camera[2] has a image sensor that converts light into electrical charges. The image sensor employed by most digital cameras is a charge-coupled device (CCD). Some cameras use complementary metal oxide semiconductor (CMOS) technology instead. Both CCD and CMOS image sensors convert light into electrons. A simplified way to think about these sensors is to think of a 2-D array[3] of thousands or millions of tiny solar cells.

Once the sensor converts the light into electrons, it reads the value (accumulated charge) of each cell in the image. This is where the differences between the two main sensor types kick in:

- A CCD transports the charge across the chip and reads it at one corner of the array. An analog-to-digital converter (ADC) then turns each pixel's value into a digital value by measuring the amount of charge at each photosite[4] and converting that measurement to binary form.
- CMOS devices use several transistors at each pixel to amplify and move the charge using more traditional wires.

Differences between the two types of sensors lead to a number of pros and cons[5]:

- CCD sensors create high-quality, low-noise images. CMOS sensors are generally more susceptible to noise.
- CMOS sensors traditionally consume little power. CCDs, on the other hand, use a process that consumes lots of power. CCDs consume as much as 100 times more power than an equivalent CMOS sensor.
- CCD sensors have been mass-produced for a longer period of time, so they are more mature. They tend to have higher quality pixels, and more of them.

Digital Camera Resolution

The amount of detail that the camera can capture is called the resolution, and it is measured in pixels. The more pixels a camera has, the more detail it can capture and the larger pictures can be without becoming blurry or "grainy."

Some typical resolutions include:

- **640 ×480 pixels** — This is the low end on most *real* cameras. This resolution is ideal for E-mailing pictures or posting pictures on a website.
- **1216 ×912 pixels** — This is a "megapixel" image size, which is good for printing pictures.
- **1600 ×1200 pixels** — With almost 2 million total pixels, this is "high resolution." You can print a 4 ×5 inch print taken at this resolution with the same quality that you would get from a photo lab.
- **2240 ×1680 pixels** — Found on 4 megapixel cameras, this resolution allows even larger printed photos, with good quality for prints up to 16 ×20 inches.
- **4064 ×2704 pixels** — A top-of-the-line[6] digital camera with 11.1 megapixels takes pictures at this resolution. At this setting, you can create 13.5 ×9 inch prints with no loss of picture quality.

High-end consumer cameras can capture over 12 million pixels. Some professional cameras support over 16 million pixels, or 20 million pixels for large-format cameras. For comparison, Hewlett Packard estimates that the quality of 35mm film is about 20 million pixels.

Storing Digital Photos

Digital cameras use a number of storage systems. These are like reusable, digital film, and they use a card reader to transfer the data to a computer. Many involve fixed or removable flash memory. Digital camera manufacturers often develop their own proprietary flash memory devices, including SmartMedia cards, CompactFlash cards and Memory Sticks.[7]

No matter what type of storage they use, all digital cameras need lots of room for pictures. They usually store images in one of two formats — TIFF, which is uncompressed, and JPEG, which is compressed, but some use RAW[8] format. Most cameras use the JPEG file format for storing pictures, and they sometimes offer quality settings (such as medium or high). The following chart (Table 8.1) will give you an idea of the file sizes you might expect with different image sizes.

Table 8.1 Image file formats and sizes

Image Size	TIFF(uncompressed)	JPEG(high quality)	JPEG(medium quality)
640×480	1.0 MB	300 KB	90 KB
800×600	1.5 MB	500 KB	130 KB
1024×768	2.5 MB	800 KB	200 KB
1600×1200	6.0 MB	1.7 MB	420 KB

To make the most of their storage space, almost all digital cameras use some sort of data compression to make the files smaller. Two features of digital images make compression possible. One is repetition. The other is irrelevancy.

Taking a Picture

It takes several steps for a digital camera to take a picture. Here's a review of what happens in a CCD camera, from beginning to end:

1. You aim the camera at the subject and adjust the optical zoom to get closer or farther away.
2. You press lightly on the shutter release.
3. The camera automatically focuses on the subject and takes a reading of the available light.
4. The camera sets the aperture and shutter speed for optimal exposure.
5. You press the shutter release all the way.
6. The camera resets the CCD and exposes it to the light, building up an electrical charge, until the shutter closes.
7. The ADC measures the charge and creates a digital signal that represents the values of the charge at each pixel.
8. A processor interpolates the data from the different pixels to create natural color.

9. A processor may perform a preset level of compression on the data.

10. The information is stored in some form of memory device (probably a flash memory card).

New Words & Expressions

convert [kənˈvɜːt] *vt*. 转换
employ [ɪmˈplɔɪ] *vt*. 使用
simplify [ˈsɪmplɪˌfaɪ] *vt*. 简化
two-dimensional [ˌtuˈdaɪmenʃnəl] *adj*. 二维的
solar [ˈsəʊlə] *adj*. 太阳的;太阳能的
accumulate [əˈkjuːmjʊˌleɪt] *vt*. 积累;积聚
transport [trænsˈpɔːt] *vt*. 运输;输送
amplify [ˈæmplɪˌfaɪ] *vt*. 放大;增强
pros and cons 利弊;优缺点;正反面
susceptible [səˈseptəbl] *adj*. 易受影响的
consume [kənˈsuːm] *vt*. 消耗;消费
process [ˈprəʊses] *n*. 工艺流程
equivalent [ɪˈkwɪvələnt] *adj*. 等效的
mass-produce [ˌmæsprəˈdjuːs] *vt*. 批量生产
mature [məˈtjʊə] *adj*. 成熟的
detail [dɪˈteɪl] *n*. 细节
capture [ˈkæptʃə] *vt*. 采集;拍摄
format [ˈfɔːmæt] *n*. 格式
estimate [ˈestɪmeɪt] *vt*. 估计;估算
transfer [ˈtrænsˌfɜː] *vt*. 转移
proprietary [prəˈpraɪəˌteri] *adj*. 专有的
repetition [ˌrepəˈtɪʃn] *n*. 重复
irrelevancy [ɪˈreləvəns] *n*. 不相关性
subject [ˈsʌbˌdʒekt] *n*. 主体
adjust [əˈdʒʌst] *vt*. 调整
available [əˈveɪləbl] *adj*. 可获得的
optimal [ˈɒptɪməl] *adj*. 最优的
represent [ˌreprɪˈzent] *vt*. 代表

Technical Terms

electron [iˈlektrɒn] *n*. 电子
pixel [ˈpɪksəl] *n*. 像素
photosite [ˈfəʊtəʊˌsaɪt] *n*. 感光点
transistor [trænˈsɪstə] *n*. 晶体管
resolution [rezəˈluːʃn] *n*. 分辨率
shutter [ˈʃʌtə] *n*. 快门
focus [ˈfəʊkəs] *n*. 焦点;*v*. 聚焦
aperture [ˈæpətʃə] *n*. 光圈
exposure [ɪksˈpəʊʒə] *n*. 曝光
interpolate [ɪnˈtɜːpəleɪt] *vt*. 对……进行插值处理
image sensor 图像传感器
electrical charge 电荷
solar cell 太阳能电池
binary form 二进制形式
card reader 读卡器
flash memory 闪存

file format　文件格式
storage space　存储空间
data compression　数据压缩
optical zoom　光学变焦
memory device　存储设备
CCD *abbr*. charge-coupled device　电荷耦合器件
CMOS *abbr*. complementary metal oxide semiconductor　互补金属氧化物半导体
ADC *abbr*. analog-to-digital converter　模数转换器
TIFF *abbr*. tag image file format　标记图像文件格式
JPEG *abbr*. Joint Photographic Experts Group　联合图像专家组

Notes

1. image sensor 译作“图像传感器”。1969 年，贝尔实验室的史密斯(George Smith)和博伊尔(Willard Boyle)成功研制了 CCD 图像传感器。1995 年，喷气推进实验室的富瑟姆(Eric Fossum)成功研制了 CMOS 图像传感器。2009 年，史密斯和博伊尔因“发明成像半导体电路——CCD 传感器”而获得诺贝尔物理学奖。
2. Camera 涵盖：(1)主要拍照片的 digital still camera(照相机)；(2)主要拍视频的 video camera(摄像机)、cinema camera(电影摄像机)、video surveillance camera(监控摄像头)等。
3. 2-D array 可译作“二维阵列”。2-D 即 two-dimensional 的缩写形式。同理，three-dimensional 也可缩写为 3-D。注意区分“向量”(vector)、“矩阵”(matrix)、“阵列”(array)。
4. photosite 可译作“感光点”、“感光单元”或“像点”。前缀 photo-既可表示“与光有关的”，如“光合作用”(photosynthesis)；也可表示“与摄影有关的”，如“摄影新闻报道”(photojournalism)。此外，还需注意区分 site，sight，cite 这三个同音异义词。
5. pros and cons 即 advantages and disadvantages。例如，“任何技术都存在优缺点”(Each technology has its pros and cons.)。又如，“我们正在权衡新计划的利与弊”(We're just weighing up the pros and cons of the new plan.)。其中，pro 和 con 均为可数名词，分别读作[prəʊ]和[kɒn]。
6. 复合形容词 top-of-the-line 或 top-of-the-range 均表示“一流的”、“顶级的”或“最昂贵的”。
7. Smart Media 简称 SM 卡，在 1995 年由东芝公司(Toshiba)推出。Compact Flash 简称为 CF 卡，在 1994 年由闪迪公司(Sandisk)推出。Memory Stick 译作“记忆棒”，由索尼公司(Sony)推出。
8. RAW 文件包含的是图像传感器捕获到的“原始”(raw)图像数据。RAW 格式也称“拜耳格式”，取名自“数字图像之父”、曾任职于柯达公司的布莱斯　拜耳(Bryce Bayer，1929—2012 年)。

Lesson 24 Television Video Signals

Figure 8.2 shows how the television signal appears on an oscilloscope. This is called composite video, meaning that there are vertical and horizontal synchronization (sync) pulses mixed with the actual picture information. These pulses are used in the television receiver to synchronize the vertical and horizontal deflection circuits to match the video being displayed. Each second of standard video contains 30 complete images, commonly called frames. A video engineer would say that each frame contains 525 lines, the television jargon for what programmers call rows. This number is a little deceptive because only 480 to 486 of these lines contain video information; the remaining 39 to 45 lines are reserved for sync pulses to keep the television's circuits synchronized with the video signal.

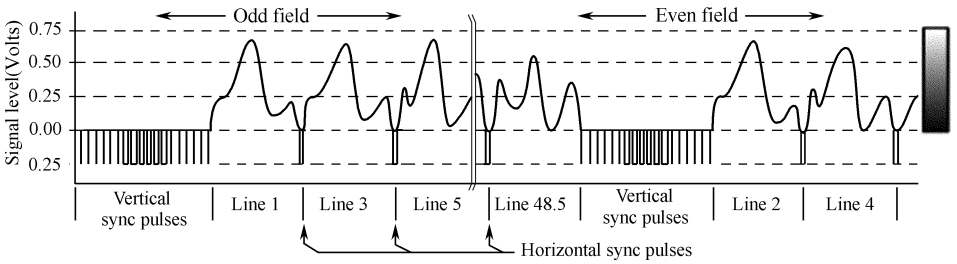


Figure 8.2 Composite video

Standard television uses an interlaced format to reduce flicker in the displayed image. This means that all the odd lines of each frame are transmitted first, followed by the even lines. The group of odd lines is called the odd field, and the group of even lines is called the even field. Since each frame consists of two fields, the video signal transmits 60 fields per second. Each field starts with a complex series of vertical sync pulses lasting 1.3 milliseconds. This is followed by either the even or odd lines of video. Each line lasts for 63 5 microseconds, including a 10.2 microsecond horizontal sync pulse, separating one line from the next. Within each line, the analog voltage corresponds to the gray scale of the image, with brighter values being in the direction away from the sync pulses. This places the sync pulses beyond the black range. In video jargon, the sync pulses are said to be blacker than black.

The hardware used for analog-to-digital conversion of video signals is called a frame grabber[1]. This is usually in the form of an electronics card that plugs into a computer, and connects to a camera through a coaxial cable. Upon command from software, the frame grabber waits for the beginning of the next frame, as indicated by the vertical sync pulses. During the following two fields, each line of video is sampled many times, typically 512, 640 or 720 samples per line, at 8 bits per sample. These samples are stored in memory as one row of the digital image.

This way of acquiring a digital image results in an important difference between the vertical and horizontal directions. Each row in the digital image corresponds to one line in the video signal, and therefore to one row of wells in the CCD. Unfortunately, the columns are not so straightforward. In the CCD, each row contains between about 400 and 800 wells (columns), depending on the particular device used. When a row of wells is read from the CCD, the resulting line of video is filtered into a smooth analog signal, such as in Figure 8.2. In other words, the video signal does not depend on how many columns are present in the CCD. The resolution in the horizontal direction is limited by how rapidly the analog signal is allowed to change. This is usually set at 3.2 MHz for color television, resulting in a rise time of about 100 nanoseconds, i.e., about 1/500th of the 53.2 microsecond video line.

When the video signal is digitized in the frame grabber, it is converted back into columns. However, these columns in the digitized image have no relation to the columns in the CCD. The number of columns in the digital image depends solely on how many times the frame grabber samples each line of video. For example, a CCD might have 800 wells per row, while the digitized image might only have 512 pixels (i.e., columns) per row.

The number of columns in the digitized image is also important for another reason. The standard television image has an aspect ratio of 4 to 3, i.e., it is slightly wider than it is high. Motion pictures have the wider aspect ratio of 25 to 9. CCDs used for scientific applications often have an aspect ratio of 1 to 1, i.e., a perfect square. In any event, the aspect ratio of a CCD is fixed by the placement of the electrodes, and cannot be altered. However, the aspect ratio of the digitized image depends on the number of samples per line. This becomes a problem when the image is displayed, either on a video monitor or in a hardcopy. If the aspect ratio isn't properly reproduced, the image looks squashed horizontally or vertically.

The 525 line video signal described here is called NTSC[2] (National Television Systems Committee), a standard defined way back in 1952. This is the system used in the United States and Japan. In Europe there are two similar standards called PAL[3] (Phase-Alternation Line) and SECAM[4] (Sequential Color and Memory). The basic concepts are the same, just the numbers are different. Both PAL and SECAM operate with 25 interlaced frames per second, with 625 lines per frame. Just as with NTSC, some of these lines occur during the vertical sync, resulting in about 576 lines that carry picture information. Other more subtle differences relate to how color and sound are added to the signal.

The most straightforward way of transmitting color television would be to have three separate analog signals, one for each of the three colors the human eye can detect: red, green and blue. Unfortunately, the historical development of television did not allow such a simple scheme. The color television signal was developed to allow existing black and white television sets to remain in use without modification. This was done by retaining the same signal for brightness information, but adding a separate signal for color information. In video jargon, the brightness is called the luminance signal, while the color is the chrominance signal. The chrominance signal is contained on a 3.58 MHz carrier wave added to the black and white video signal. Sound is added in this same way, on a

4.5 MHz carrier wave. The television receiver separates these three signals, processes them individually, and recombines them in the final display.

New Words & Expressions

deflection [dɪˈflekʃn] *n*. 偏转
frame [freɪm] *n*. 画面
jargon [ˈdʒɑːgən] *n*. 行话
deceptive [dɪˈseptɪv] *adj*. 欺骗性的
interlace [ɪntəˈleɪs] *vt*. 使交错
flicker [ˈflɪkə] *vi*. 闪烁
squash [skwɒʃ] *vt*. 挤压
in any event 无论如何

Technical Terms

luminance [ˈluːmɪnəns] *n*. 亮度
chrominance [ˈkrəʊmɪnəns] *n*.色度
odd field　奇数场
even field　偶数场
frame grabber　帧采集器
carrier wave　载波
NTSC *abbr*. National Television Systems Committee　国家电视系统委员会
PAL [pæl] *abbr*. Phase-Alternation Line　逐行倒相
SECAM [ˈsiːkæm] *abbr*. Sequential Color and Memory　顺序彩色与存储

Notes

1. “帧采集器”(frame grabber)是计算机系统中用于模拟视频信号数字化的设备，其功能和显卡正好相反。
2. NTSC 制(又称 N 制)是在 1952 年 12 月由美国“国家电视系统委员会”(NTSC)制定的彩色电视广播标准。
3. PAL 制(帕尔制)是在 1962 年由前联邦德国在 NTSC 技术上研制出来的改进方案。
4. SECAM 制式(塞康制)是 1966 年由法国研制成功的。

Exercises

1. Fill in the blanks with proper words, phrases or clauses.

(1)_____(为了理解视频), it is best to start with simple, old-fashioned black-and-white television. To represent _____(二维图像)in front of it as a one-dimensional voltage as _____(时间函数), the camera scans an _____(电子束)rapidly across the image and slowly down it, recording _____(光的强度)as it goes. At the end of the scan,_____(称作一帧), the beam retraces. This intensity as a function of time is broadcast, and receivers repeat the scanning process to _____(重建图像).

The exact scanning parameters _____ from country to country. The system used in North and South America and Japan has 525 _____(扫描线), a horizontal-to-vertical aspect ratio of 4∶3, and 30 frames/sec. The European system has 625 scan lines, the same aspect ratio of 4∶3, and 25 frames/sec._____(在这两个系统中), the top few and bottom few lines are not displayed (to approximate a rectangular image on the original round CRTs). Only 483 of the 525 NTSC scan lines (and 576 of the 625 PAL/SECAM scan lines) are displayed. The beam _____(被关闭)during the vertical retrace, so many stations (especially in Europe) use this time to broadcast TeleText (text pages containing news, weather, sports, stock prices, etc.).

While 25 frames/sec is enough _____ capture smooth motion, at that frame rate many people, especially older ones, will perceive the image to flicker (because the old image _____ faded off the retina before the new one appears)._____ increase the frame rate, which would require using more scarce bandwidth, a different approach is taken._____ the scan lines being displayed _____(依序), first all the odd scan lines are displayed, then the even ones are displayed. Each of these half frames is called a field. Experiments have shown _____ although people notice flicker at 25 frames/sec, they do not notice it at 50 fields/sec. This technique is called interlacing. Non-interlaced television or video is called progressive. Note that movies run _____ 24 fps, but each frame is fully visible for 1/24 sec.

(2) Color video uses the same scanning pattern _____ monochrome (black and white), except _____ instead of displaying the image with one moving beam, it uses three beams moving _____ unison. One beam is used _____ each of _____(加性三基色): red, green, and blue (RGB). This technique works because any color can be constructed from _____(红、绿、蓝的线性叠加)with the appropriate intensities. However, for transmission on a single channel, the three color signals _____(必须合成为单一复合信号).

When color television was invented, _____(各种色彩显示方法)were technically possible, and different countries _____ made different choices, leading to systems that are still _____(不兼容). (Note that these choices have nothing to do with VHS versus Betamax versus P2000, which are recording methods.) In all countries, a political requirement was that programs transmitted in color had to be receivable on _____(现存的)black-and-white television sets. Consequently, the simplest scheme, just encoding the RGB signals separately, was not acceptable._____(RGB 并非效率最高的方案).

2. Translate the following passages into Chinese or English.

(1) The three-dimensional world is imaged by the lens of the human eye onto the retina, which is populated with photoreceptor cells that respond to light having wavelengths in the range of about 400 nm to 700 nm. In an imaging system, we build a camera having a lens and a photosensitive device, to mimic how the world is perceived by vision.

(2) In computing, a display is described by the count of pixels across the width and height of the image. Conventional television would be denoted 644×483, which indicates 483 picture lines. But any display system involves some scanning overhead, so the total number of

lines in the raster of conventional video is necessarily greater than 483.

(3) Video scanning systems have traditionally been denoted by their total number of lines including sync and blanking overhead, the frame rate in hertz, and an indication of interlace (2∶1) or progressive (1∶1) scan. 525/59.94/2∶1 scanning is used in North America and Japan, with an analog bandwidth for studio video of about 5.5 MHz. 625/50/2∶1 scanning is used in Europe and Asia, with an analog bandwidth for studio video of about 6.5 MHz.1125/60/2∶1 scanning is in use for high-definition television (HDTV), with an analog bandwidth of about 30 MHz.

(4) A video system conveys image data in the form of a component that represents brightness, and two other components that represent color. It is important to convey the brightness component in such a way that noise introduced in transmission, processing, and storage has a perceptually similar effect across the entire tone scale from black to white.

(5) Although the shape of the retina is roughly a section of a sphere, it is topologically two-dimensional. In a camera, for practical reasons, we employ a flat image plane, instead of a spherical image surface. Image system theory concerns analyzing the continuous distribution of power that is incident on the image plane. A photographic camera has, in the image plane, film that is subject to chemical change when irradiated by light. The active ingredient of photographic film is contained in a thin layer of particles having carefully controlled size and shape, in a pattern with no coherent structure. If the particles are sufficiently dense, an image can be reproduced that has sufficient information for a human observer to get a strong sense of the original scene. The finer the particles and the more densely they are arranged in the film medium, the higher will be the capability of the film to record spatial detail.

(6) 人眼有这样的特点：出现在视网膜上的图像保留几毫秒后才会消失。如果图像序列以50幅每秒的速度逐行扫描显示的话，眼睛不会感到那是一幅幅离散图像。视频系统都是利用该原理产生运动画面的。

(7) 压缩系统需要两种算法：一种算法用于信源压缩数据，另一种算法用于信宿解压数据。文献中将这两种算法分别称作“编码算法”和“解码算法”。

(8) JPEG(联合图像专家组)用于压缩连续色调的静止图像(如照片)。它是由ITU、ISO和IEC共同支持的图像专家开发出来的。

(9)“宽高比”是图像宽度和高度之比。常规电视的宽高比是4∶3。高清晰度电视使用更宽的16∶9宽高比。摄像机常用的宽高比是1.85∶1或2.35∶1。

(10) MPEG代表“运动图像专家组”。它是对数字压缩格式视听信息(电影、录像、音乐)编码的一组标准的统称。与其他的视频、音频的编码格式相比较，MPEG的主要优势是同质量下的文件小得多。这是因为MPEG使用了非常复杂的压缩技术。

3. Writing tasks.

Summary: Write an abstract for *Lesson* 23 *Digital Cameras* within 150～200 words.

Composition: Write an essay about your experience of using digital cameras.

Reading Materials

Passage 1 JPEG

The Joint Photographic Experts Group (JPEG) standard provides lossy compression that is used in most implementations. The JPEG standard can be used for both color and gray images. However, for simplicity, we discuss only the gray scale pictures; the method can be applied to each of the three channels in a color image. In JPEG, a gray scale picture is divided into blocks of 8 × 8 pixels. The compression and decompression each go through three steps, as shown in Figure 8.3.

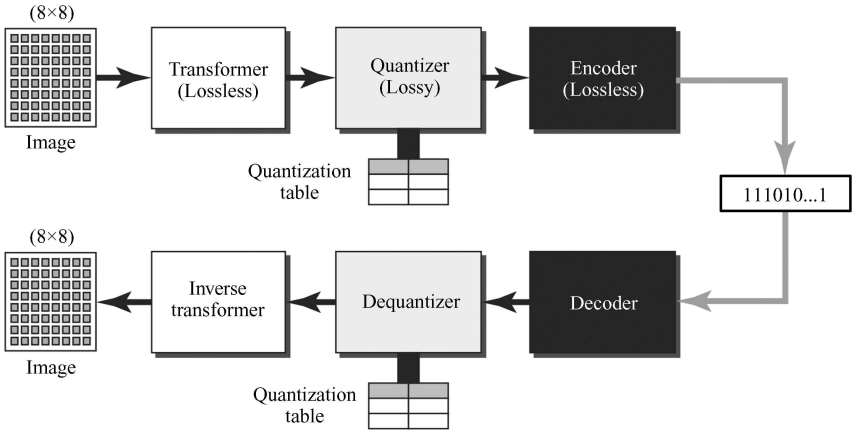


Figure 8.3 Compression in each channel of JPEG

Transformation JPEG normally uses disctrete cosine transform (DCT) in the first step in compression and inverse DCT in the last step in decompression. Transformation and inverse transformation are applied on 8 × 8 blocks.

Quantization The output of DCT transformation is a matrix of real numbers. The precise encoding of these real numbers requires a lot of bits. JPEG uses a quantization step that not only rounds real values in the matrix, but also changes some values to zeros. The zeros can be eliminated in the encoding step to achieve a high compression rate. The result of DCT transformation defines the weights of different frequencies in the source matrix. Since high frequencies mean sudden changes in the value of pixels, they can be eliminated because human vision cannot recognize them.

The quantization step creates a new matrix in which each element, $\boldsymbol{C}(m, n)$, is defined as shown below.

$$\boldsymbol{C}(m, n) = \text{round}\,[\boldsymbol{M}(m, n) / \boldsymbol{Q}(m, n)]$$

in which $\boldsymbol{M}(m, n)$ is an entry in the transformed matrix and $\boldsymbol{Q}(m, n)$ is an entry in the quantization matrix. The round function first adds 0.5 to a real value and then truncates the value to an integer. This means that 3.7 is rounded to integer 4, but 3.2 is rounded to integer 3.

JPEG has defined 100 quantization matrices, $\boldsymbol{Q}_1$ to $\boldsymbol{Q}_{100}$, in which $\boldsymbol{Q}_1$ gives the poorest image quality but the highest level of compression and $\boldsymbol{Q}_{100}$ gives the best image quality but the lowest level of compression. It is up to the implementation to choose one of these matrices. Figure 8.4 shows some of these matrices.

$$\underset{\boldsymbol{Q}_{10}}{\begin{bmatrix} 80 & 60 & 50 & 80 & 120 & 200 & 255 & 255 \\ 55 & 60 & 70 & 95 & 130 & 255 & 255 & 255 \\ 70 & 65 & 80 & 120 & 200 & 255 & 255 & 255 \\ 70 & 85 & 110 & 145 & 255 & 255 & 255 & 255 \\ 90 & 110 & 185 & 255 & 255 & 255 & 255 & 255 \\ 120 & 175 & 255 & 255 & 255 & 255 & 255 & 255 \\ 245 & 255 & 255 & 255 & 255 & 255 & 255 & 255 \\ 255 & 255 & 255 & 255 & 255 & 255 & 255 & 255 \end{bmatrix}} \quad \underset{\boldsymbol{Q}_{50}}{\begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 110 & 103 & 99 \end{bmatrix}} \quad \underset{\boldsymbol{Q}_{90}}{\begin{bmatrix} 3 & 2 & 2 & 3 & 5 & 8 & 10 & 12 \\ 2 & 2 & 3 & 4 & 5 & 12 & 12 & 11 \\ 3 & 3 & 3 & 5 & 8 & 11 & 14 & 11 \\ 3 & 3 & 4 & 6 & 10 & 17 & 16 & 12 \\ 4 & 4 & 7 & 11 & 14 & 22 & 21 & 15 \\ 5 & 7 & 11 & 13 & 16 & 12 & 23 & 18 \\ 10 & 13 & 16 & 17 & 21 & 24 & 24 & 21 \\ 14 & 18 & 19 & 20 & 22 & 20 & 20 & 20 \end{bmatrix}}$$

Figure 8.4 Three different quantization matrices

Note that the only phase in the process that is not completely reversible is the quantizing phase. We lose some information here that is not recoverable. As a matter of fact, the only reason that JPEG is called *lossy compression* is because of this quantization phase.

Encoding After quantization, the values are reordered in a zigzag sequence before being input into the encoder. The zigzag reordering of the quantized values is done to let the values related to the lower frequency feed into the encoder before the values related to the higher frequency. Since most of the higher-frequency values are zeros, this means non-zero values are given to the encoder before the zero values. The encoding in this case is a lossless compression using either run-length coding or arithmetic coding. Figure 8.5 shows the encoding process of a block of gray image in which the bit depth for each pixel is 20.

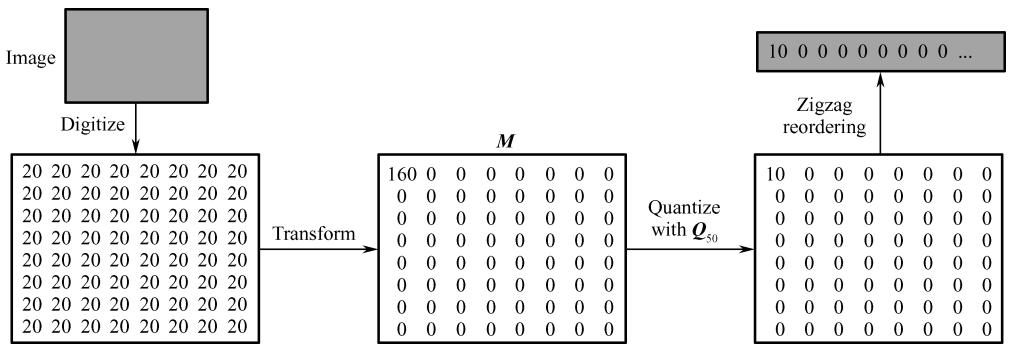


Figure 8.5 The JPEG encoding process of a block of gray image with uniform gray scale

Questions:

1. What does the acronym DCT stand for?
2. For what reason is JPEG called *lossy compression*?
3. Could you describe the zigzag reordering in JPEG encoding?
4. Is the run-length coding in JPEG compression lossy or lossless?
5. What does the technical term *round function* mean in this article?

Passage 2 MPEG

Motion Picture Experts Group (MPEG) is a method to compress video. As shown in Figure 8.6, a motion picture is a rapid flow of a set of frames, where each frame is an image. In other words, a frame is a spatial combination of pixels, and a video is a temporal combination of frames that are sent one after another. Compressing video, then, means spatially compressing each frame and temporally compressing a set of frames.

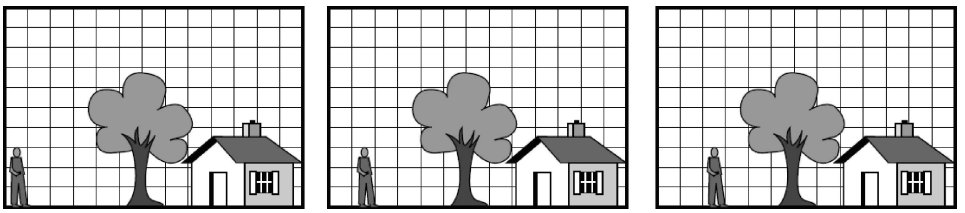


Figure 8.6 Three consecutive video frames

Spatial Compression The spatial compression of each frame is done with JPEG (or a modification of it). Each frame is a picture that can be independently compressed.

Temporal Compression In temporal compression, redundant frames are removed. When we watch television, we receive 50 frames per second. However, most of the consecutive frames are almost the same. For example, when someone is talking, most of the frame is the same as the previous one except for the segment of the frame around the lips, which changes from one frame to another.

To temporally compress data, the MPEG method first divides a set of frames into three categories: I-frames, P-frames, and B-frames. Figure 8.7 shows how a set of frames (7 in the figure) are compressed to create another set of frames.

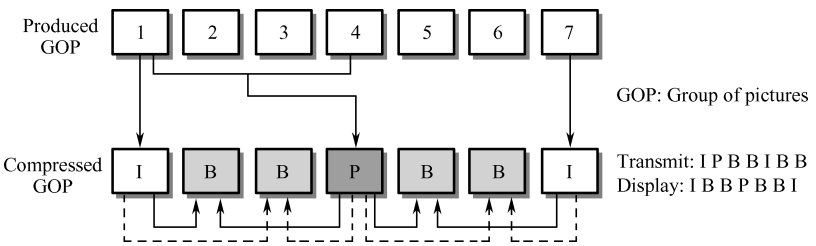


Figure 8.7 MPEG frames

- **I-frames.** An *intracoded frame* (I-frame) is an independent frame that is not related to any other frame (not to the frame sent before or after). They are present at regular intervals. An I-frame must appear periodically to handle some sudden change in the frame that the previous and following frames cannot show. Also, when a video is broadcast, a viewer may tune in at any time. If there is only one I-frame at the beginning of the broadcast, the viewer who tunes in late will not receive a complete

picture. I-frames are independent of other frames and cannot be constructed from other frames.

- **P-frames.** A *predicted frame* (P-frame) is related to the preceding I-frame or P-frame. In other words, each P-frame contains only the changes from the preceding frame. P-frames can be constructed only from previous I- or P-frames. P-frames carry much less information than other frame types and carry even fewer bits after compression.
- **B-frames.** A *bidirectional frame* (B-frame) is related to the preceding and following I-frame or P-frame. In other words, each B-frame is relative to the past and the future. Note that a B-frame is never related to another B-frame.

The MPEG-1 standard(International Standard 11172) was first published in 1993 and is still widely used. Its goal was to produce video-recorder-quality output that was compressed 40:1 to rates of around 1 Mbps. This video is suitable for broad Internet use on Web sites.

The MPEG-2 standard (International Standard 13818), released in 1996, was designed for compressing broadcast-quality video. It is very common now, as it is used as the basis for video encoded on DVDs and for digital broadcast television. DVD quality video is typically encoded at rates of 4 ~ 8 Mbps.

The MPEG-4 standard has two video formats. The first format, released in 1999, encodes video with an object-based representation. This allows for the mixing of natural and synthetic images and other kinds of media, for example, a weatherperson standing in front of a weather map. With this structure, it is easy to let programs interact with movie data.

The second format, released in 2003, is known as H.264 or AVC (Advanced Video Coding). Its goal is to encode video at half the rate of earlier encoders for the same quality level, all the better to support the transmission of video over networks. This encoder is used for HDTV on most Blu-ray discs.

Questions:

1. What does the technical term *frame* mean in this article?
2. Could you describe the spatial compression and temporal compression in MPEG?
3. For what reason must an I-frame appear periodically?
4. Is MPEG-2 commonly used in DVD quality video or not?
5. What does the acronym AVC stand for?

Passage 3 Internet Video

Many Internet companies today provide streaming video, including, Netflix, YouTube (Google), Amazon, and TikTok. As shown in Figure 8.8, Netflix video distribution has two major components: the Amazon cloud and its own private *content delivery network* (CDN) infrastructure. Netflix has a website that handles numerous functions, including user registration and login, billing, movie catalogue for browsing and searching, and a movie recommendation system. This website (and its associated backend databases) run entirely on

Amazon servers in the Amazon cloud.

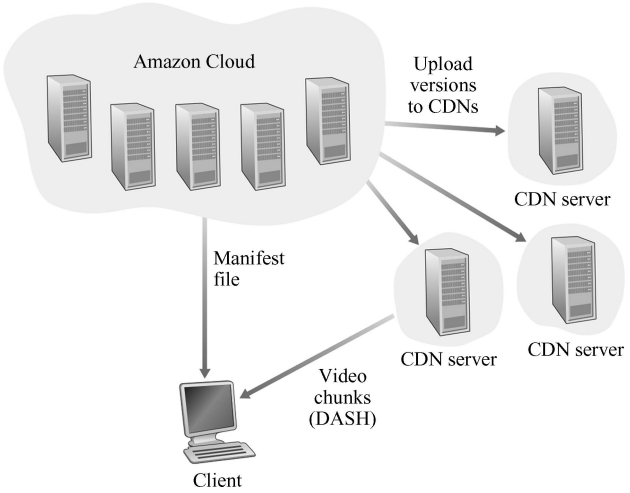


Figure 8.8 Netflix video streaming platform

Dynamic Adaptive Streaming over HTTP (DASH)

In HTTP streaming, the video is simply stored at an HTTP server as an ordinary file with a specific URL. When a user wants to see the video, the client establishes a TCP connection with the server and issues an HTTP GET request for that URL. The server then sends the video file, within an HTTP response message, as quickly as the underlying network protocols and traffic conditions will allow.

On the client side, the bytes are collected in a client application buffer. Once the number of bytes in this buffer exceeds a predetermined threshold, the client application begins playback—specifically, the streaming video application periodically grabs video frames from the client application buffer, decompresses the frames, and displays them on the user's screen. Thus, the video streaming application is displaying video as it is receiving and buffering frames corresponding to latter parts of the video.

Although HTTP streaming has been extensively deployed in practice, it has a major shortcoming: All clients receive the same encoding of the video, despite the large variations in the amount of bandwidth available to a client, both across different clients and also over time for the same client. This has led to the development of a new type of HTTP-based streaming, often referred to as *dynamic adaptive streaming over HTTP* (DASH).

In DASH, the video is encoded into several different versions, with each version having a different bit rate and, correspondingly, a different quality level. The client dynamically requests chunks of video segments of a few seconds in length. When the amount of available bandwidth is high, the client naturally selects chunks from a high-rate version; and when the available bandwidth is low, it naturally selects from a low-rate version. The client selects different chunks one at a time with HTTP GET request messages.

DASH allows clients with different Internet access rates to stream in video at different

encoding rates. Clients with low-speed connections can receive a low bit-rate (and low-quality) version, and clients with fiber connections can receive a high-quality version. DASH also allows a client to adapt to the available bandwidth if the available end-to-end bandwidth changes during the session. This feature is particularly important for mobile users, who typically see their bandwidth availability fluctuate as they move with respect to the base stations.

Content Distribution Networks (CDNs)

For an Internet video company, perhaps the most straightforward approach to providing streaming video service is to build a single massive data center, store all of its videos in the data center, and stream the videos directly from the data center to clients worldwide. But there are three major problems with this approach.

First, if the client is far from the data center, server-to-client packets will cross many communication links and likely pass through many *Internet service providers* (ISPs), with some of the ISPs possibly located on different continents. If one of these links provides a throughput that is less than the video consumption rate, the end-to-end throughput will also be below the consumption rate, resulting in annoying freezing delays for the user. The likelihood of this happening increases as the number of links in the end-to-end path increases.

A second drawback is that a popular video will likely be sent many times over the same communication links. Not only does this waste network bandwidth, but the Internet video company itself will be paying its provider ISP (connected to the data center) for sending the same bytes into the Internet over and over again. A third problem with this solution is that a single data center represents a single point of failure—if the data center or its links to the Internet goes down, it would not be able to distribute any video streams.

In order to meet the challenge of distributing massive amounts of video data to users distributed around the world, almost all major video-streaming companies make use of *content distribution networks* (CDNs). A CDN manages servers in multiple geographically distributed locations, stores copies of the videos (and other types of Web content, including documents, images, and audio) in its servers, and attempts to direct each user request to a CDN location that will provide the best user experience. The CDN may be a private CDN, that is, owned by the content provider itself. The CDN may alternatively be a third-party CDN that distributes content on behalf of multiple content providers.

Questions:

1. What does the technical term *streaming* mean in this article?
2. What components does Netflix video distribution have?
3. Could you list the drawbacks of building a single massive data center?
4. Are CDN servers in Netflix video streaming platform private or third-party?
5. What does the acronym DASH stand for?

Unit 9

Embedded Applications

Lesson 25 Choosing the Right Core

Lesson 26 Design Languages for Embedded Systems

Lesson 27 Choosing a Real-Time Operating System

Passage 1 ARM

Passage 2 Embedded OS

Passage 3 Android Tablets

Lesson 25 Choosing the Right Core

Generally speaking, there are five very-widely-used 32-bit cores on the market at the moment: Intel x86, PowerPC, MIPS, SuperH, and ARM.[1] Numerous less popular or proprietary architectures also exist, of course; many of these are associated with specific applications such as laser printers or DVD players.

Architectures based around the high-end x86 family have some immediate advantages:

- You can use almost any PC-compatible operating system, and free software development tools.
- Installing operating systems is simple; in most cases there are automated installers that will probe your hardware combination and automatically install appropriate kernels, drivers etc. Compare this to the norm with embedded systems, where you will need to look at the board, work out the hardware configuration yourself, and sysgen[2] the kernel and driver set on external hardware, probably using a cross-compiler.
- It is simple to interface literally thousands of peripheral components for almost any imaginable function. Because these components are produced for the consumer market, with its enormous volumes and bloodthirsty price competition, peripheral components are cheap and fairly easy to acquire.
- Driver support exists (within the framework of most off-the-shelf operating systems) for almost any piece of hardware you could want to attach to your system.
- Highly-integrated mainboards are available with many possible combinations of peripherals, in a wide variety of form factors.
- Migrating to a slightly different hardware platform due to shortages of support parts or evolving customer needs is relatively simple, in many cases, it simply involves recompiling and reinstalling the operating system and preparing a new master disk image for duplication.

The obvious virtues of these parts having been extolled, some of the downsides must be pointed out:

- x86 parts are very expensive, in production quantities, compared to RISC alternatives of comparable performance. This may affect your ability to commercialize your device.
- There are relatively few x86 variants that are true "system-on-chip" devices, so you are likely to need quite a bit of external hardware in addition to the microprocessor itself. Often, in order to obtain one specific function, you will need to add a complex multifunction part because the single function you want isn't available as a discrete

component. Again, this brings up your system complexity and total bill-of-materials cost.

- x86 has significant power consumption, heat and size disadvantages.
- Modern x86 parts and their support chips are very high-speed devices in dense packages. It is virtually impossible to hand-prototype your own design based around these parts; unless you want to spend many thousands of dollars on equipment, at the very least you will have to contract out some assembly work.
- PC peripheral ICs often have very short production lifespans; twelve to eighteen months is not uncommon, so ongoing sourcing may be an issue.
- Code to cold-boot a "bare" PC platform is usually very complicated, because you have to replace numerous layers — motherboard BIOS, expansion card BIOS, and various OS layers. The CPU architecture is also complex.
- It bears pointing out that JTAG[3]-based or other hardware debugging systems aren't usually available on commercial single-board x86 computers.

I recommend x86 as the platform of choice if you are either building just a few of your appliance, or if you are prototyping something and want to pull together a lot of miscellaneous hardware features without spending a lot of time debugging the hardware design. It's also a good choice for an initial production run that you can ship to early adopters while you are developing a cheaper second-round customized hardware design. There are other special situations where you might find x86 to be a good choice, but these are the major ones.

Moving onto the RISC platforms, MIPS, SuperH and PowerPC are good candidates for many applications, and in particular the SuperH family is large and contains a wide variety of useful devices, though MIPS seems to be a more widely licensed core in third-party ASICs and ASSPs. PowerPC seems to be found mainly in applications requiring very high performance. In evaluating all of these parts for various projects, I have found them to be fairly difficult to develop with on a shoestring budget; evaluation hardware is usually costly, and most variants of these parts are not readily available to buyers who are unable to demonstrate a need for large quantities. However, all of these cores are likely to remain available and well-supported for the foreseeable future, so they are all viable choices as long as you can obtain development systems and parts.

At least in the case of SuperH and MIPS, your cheapest path to a prototype based on these parts is generally to repurpose some existing piece of hardware; for PowerPC, I would suggest buying a commercial single-board industrial control computer based around the chip of interest. Be warned that this is likely to be expensive; PowerPC boards don't have the same kind of mass-market pricing as x86-compatible boards and you can expect to pay between two and three times as much for a PowerPC SBC as for a comparable x86-based board.

Bearing the above discussion in mind, unless some of the Intel arguments apply to your case, my primary recommendation for a 32-bit embedded platform is ARM. This architecture has many important advantages (some of these are also applicable to the other RISC platforms

mentioned above, of course):

- It is a mature, well-understood architecture with a solid engineering history and many refinements. The large number of current licensees and now-shipping parts makes ARM a very safe bet for future availability.
- The cores are small and have excellent power consumption vs. performance characteristics.
- Many features — coprocessors, external bus widths, memory management unit, cache size, etc. — are tunable by the chip designer, meaning that a core variant can be found to meet almost any performance/size/power requirement.
- There are a huge number of attractively priced standard, custom and semi-custom parts on the market with a wide variety of integrated peripherals.
- Since ARM provides reference designs for many different peripherals as well as the core itself, there are often similarities in peripheral control on different ARM implementations, even from different vendors. To take a trivial example, code to send data out of a serial port can usually be ported from one ARM variant to another with little effort.
- Partly because of the above factors, there is a huge amount of freely available intellectual property already extant for this core.

The cliché often used is that "ARM is the 32-bit 8051", meaning that it is the universal 32-bit microcontroller core known to everybody and used everywhere. This is barely an exaggeration: ARM is to the embedded world what x86 is to the desktop PC world.

New Words & Expressions

proprietary [prə'praɪətəri] *adj*. 专属的
immediate [ɪ'miːdiət] *adj*. 立即的
probe [prəʊb] *vt*. 查看
norm [nɔːm] *n*. 常态
attach to ... 连接到……
a wide variety of ... 大量各种的……
shortage ['ʃɔːtɪdʒ] *n*. 缺少
framework ['freɪmwɜːk] *n*. 框架
duplication [ˌduplɪ'keɪʃn] *n*. 复制
virtue ['vɜːtʃuː] *n*. 优点;长处
extol [ɪk'stəʊl] *vt*. 赞美
downside ['daʊnsaɪd] *n*. 不利方面
variant ['veəriənt] *n*. 变体
ongoing ['ɒngəʊɪŋ] *adj*. 持续存在的
complicated ['kɒmplɪˌeɪtɪd] *adj*. 复杂的
appliance [ə'plaɪəns] *n*. 器具
shoestring ['ʃuːstrɪŋ] *n*. 小额资本
viable ['vaɪəbl] *adj*. 可行的
refinement [rɪ'faɪnmənt] *n*. 改进
licensee [laɪsən'siː] *n*. 获许可人
extant [ek'stænt] *adj*. 现存的
cliché ['kliːʃeɪ] *n*. 套话
universal [ˌjuːnɪ'vɜːsl] *adj*. 通用的
exaggeration [ɪgˌzædʒə'reɪʃn] *n*. 夸张

Technical Terms

cross-compiler 交叉编译器
bill of materials 材料单

memory management unit　内存管理单元
RISC *abbr*. reduced-instruction-set computer　精简指令集计算机
JTAG *abbr*. Joint Test Action Group　联合测试行动组
ASIC *abbr*. application-specific integrated circuit　专用集成电路
ASSP *abbr*. application-specific standard product　专用标准产品
IP *abbr*. intellectual property　知识产权

Notes

1. PowerPC 是在 1991 年由 Apple、IBM、Motorola 联合推出的 RISC 微处理器架构。MIPS(Microprocessor without Interlocked Pipeline Stages)是由 MIPS 公司推出的 RISC 微处理器架构。SuperH(或 SH)是在 20 世纪 90 年代初由日立(Hitachi)公司推出的 32 位 RISC 微处理器架构。ARM(Advanced RISC Machine)是由 ARM 公司推出的 RISC 微处理器架构。
2. sysgen 这个词是由“系统”(system)和“产生”(generation)合成的，可译为“系统生成”。
3. JTAG 即 IEEE 1149.1 标准，此标准使用“边界扫描”技术测试印制电路板。

Lesson 26　Design Languages for Embedded Systems

An embedded system is a computer masquerading as a non-computer that must perform a small set of tasks cheaply and efficiently. A typical system might need to do communication, signal processing, and user-interface tasks. Because the tasks must solve diverse problems, a general-purpose language to solve all embedded applications would be difficult to write, analyze, and compile. Instead, developers have introduced a variety of languages, each best suited to a particular problem domain.

Hardware Languages

Verilog and VHDL are the most popular languages for hardware description and modeling. Each hardware-description language (HDL) models systems with discrete-event semantics that ignore idle portions of the design for efficient simulation. Both describe systems with structural hierarchy: a system consists of blocks that contain instances of primitives, other blocks, or concurrent processes. In addition, each HDL explicitly lists connections.

Verilog provides more primitives geared specifically toward hardware simulation. On the other hand, VHDL's primitives are assignments such as a = b + c or procedural code. Verilog adds transistor and logic-gate primitives, and lets you define new primitives with truth tables.

Both languages allow concurrent processes to be described procedurally. Such processes sleep

until awakened by an event that causes them to run, read and write variables, and suspend. Processes may wait for a period of time (for example, # 10 in Verilog, wait for 10ns in VHDL), a value change (@(a or b), wait on a,b), or an event (@(posedge clk), wait on clk until clk='1').

VHDL communication is more disciplined and flexible. Verilog communicates through wires or regs: shared memory locations that can cause race conditions [1]. VHDL's signals behave like wires, but the resolution function may be user-defined. VHDL's variables are local to a single process unless declared as shared variables.

Verilog's type system models hardware with four-valued bit vectors and arrays for modeling memory. VHDL does not include four-valued vectors, but its type system lets a user add them. Furthermore, you can define composite types such as C structs.

Overall, Verilog is the leaner language, more directly geared toward simulating digital integrated circuits. VHDL is a much larger, more verbose language capable of handing a wider class of simulation and modeling tasks.

Software Languages

Software languages describe sequences of instructions for a processor to execute. As such, most languages list imperative instructions, executed in order that communicate through memory — an array of storage locations that hold their values until changed.

Each machine instruction typically does little more than, for example, add two numbers, so high-level languages aim to specify many instructions concisely and intuitively. Arithmetic expressions are typical — coding an expression such as $ax^2 + bx + c$ in machine code is straightforward, tedious, and best done by a compiler. The C language provides such expressions, control-flow constructs such as loops and conditionals, and recursive functions. In addition, the C++ language provides classes as a way to build new data types, templates for polymorphic code, exceptions for error handling, and a standard library of common data structures. Java is a still higher-level language that provides automatic garbage collection, threads, and monitors to synchronize the threads.

Assembly Languages

An assembly language program is a list of processor instructions written in a symbolic, human-readable form. Each instruction consists of an operation, such as addition, along with some operands. For example, add r5, r2, r4 might add the contents of registers r2 and r4 and then write the result to r5. An assembly language executes these types of arithmetic instructions in order, but branch instructions can perform conditionals and loops by changing the processor's program counter — the address of the instruction being executed.

A processor's assembly language is defined by its opcodes, addressing modes, registers, and memories. The opcode distinguishes, for example, addition from a conditional branch, and an addressing mode defines how and where data is gathered and stored (such as, from a

register or from a particular memory location). Registers can be thought of as small, fast, and easy-to-access pieces of memory.

The C Language

A C program contains functions built from arithmetic expressions structured with loops and conditionals. Instructions in a C program run sequentially, but control flow constructs such as loops of conditionals can affect the order in which instructions execute. When control reaches a function call in an expression, the program passes control to the called function. This function runs until it produces a result, and control returns to continue evaluating the expression that called the function.

C derives its types from those a processor manipulates directly: signed and unsigned integers ranging from bytes to words, floating-point numbers, and pointers. These can be further aggregated into arrays and structures — groups of named fields.

C programs use three types of memory. The program allocates space for global data upon program compilation. The stack stores automatic variables the program allocates and releases the variables when the program calls the function and returns a value. The heap supplies arbitrarily sized regions of memory that the program can deallocate in any order. Although the C language is an ISO standard, many people consult the book by Kernighan and Ritchie, since Ritchie designed the language.

C++

C++extends C with structuring mechanisms for big programs. These mechanisms include user-defined data types, a way to reuse code with different types, names to group objects and avoid accidental name collisions when program pieces are assembled, and exceptions to handle errors. The C++ standard library includes a collection of efficient polymorphic data types such as arrays, trees, and strings for which the compiler generates custom implementations.

A class defines a new data type by specifying its representation and the operations that may access and modify it. Classes may be defined by inheritance, which extends and modifies existing classes. For example, a rectangle class might add length and width fields and an area method to a shape class.

A template is a function or class that can work with multiple types. The compiler generates custom code for each different use of the template. For example, a program could use the same min template for both integers and floating-point numbers.

Java[2]

Sun's Java language resembles, but is incompatible with, C++. Like C++, Java is object-oriented, providing classes and inheritance. It is a higher-level language than C++, since it uses object references, arrays, and strings instead of pointers. Java's automatic

garbage collection frees the programmer from memory management.

Java provides concurrent threads. Creating a thread involves extending the Thread class, creating instances of these objects, and calling their start methods to start a new thread of control that executes the objects' run methods.

Synchronizing a method or block uses a per-object lock to resolve contention when two or more threads attempt to access the same object simultaneously. A thread that attempts to gain a lock owned by another thread will block until the lock is released — you can use this feature grant a thread exclusive access to a particular object.

RTOS

Many embedded systems use a real-time operating system (RTOS) to simulate concurrency on a single processor. An RTOS manages multiple running processes, each written in a sequential language such as C. The processes perform the system's computations and the RTOS schedules them. The scheduling attempts to meet deadlines by deciding which processes run at what times and in what order.

Most real-time operating systems use fixed-priority preemptive scheduling in which each process is given a particular priority (a small integer) when the system is designed. At any time, the RTOS runs the highest-priority running process, which is expected to run for a short period of time before suspending itself to wait for more data. Priorities are usually assigned using rate-monotonic analysis, which assigns higher priorities to processes that must meet deadlines that are more frequent.

New Words & Expressions

masquerade [ˌmɑːskəˈreɪd] *vi*. 伪装
semantics [sɪˈmæntɪks] *n*. 语义
hierarchy [ˈhaɪərɑːki] *n*. 层级体系
instance [ˈɪnstəns] *n*. 实例
primitive [ˈprɪmɪtɪv] *n*. 元件
explicitly [ɪkˈsplɪsɪtli] *adv*. 明确地
gear [gɪə] *vt*. 使　　适合
assignment [əˈsaɪnmənt] *n*. 赋值
disciplined [ˈdɪsɪplɪnd] *adj*. 规矩的
verbose [vɜːˈbəʊs] *adj*. 冗长的
imperative [ɪmˈperətɪv] *adj*. 命令的
straightforward [streɪtˈfɔːwəd] *adj*. 简明的
tedious [ˈtiːdɪəs] *adj*. 乏味的
loop [luːp] *vi*. 循环
conditional [kənˈdɪʃnəl] *adj*. 有条件的
polymorphic [ˈpɒlɪˌmɔːfɪk] *adj*. 多态的
symbolic [sɪmˈbɒlɪk] *adj*. 使用符号的
integer [ˈɪntɪdʒə] *n*. 整数
aggregate [ˈægrɪgeɪt] *vt*. 聚集
allocate [ˈæləkeɪt] *vt*. 分配
release [rɪˈliːs] *vt*. 释放
mechanism [ˈmekənɪzəm] *n*. 机制
accidental [æksɪˈdentəl] *adj*. 意外的
collision [kəˈlɪʒən] *n*. 冲突
contention [kənˈtenʃn] *n*. 竞争
schedule [ˈskedʒuːl] *vt*. 调度
preemptive [prɪˈemptɪv] *adj*. 先占的
monotonic [mɒnəˈtɒnɪk] *adj*. 单调的

Technical Terms

expression [ɪkˈspreʃn] *n*. 表达式
compiler [kəmˈpaɪlə] *n*. 编译器
template [ˈtemplɪt] *n*. 模板
operand [ˈɒpərænd] *n*. 操作数
pointer [ˈpɒɪntə] *n*. 指针
array [əˈreɪ] *n*. 数组
structure [ˈstrʌktʃə] *n*. 结构体
field [fiːld] *n*. 字段
stack [stæk] *n*. 堆栈
class [klɑːs] *n*. 类
thread [θred] *n*. 线程
string [strɪ] *n*. 字符串
concurrent process　并发进程
truth table　真值表
resolution function　判决函数
bit vector　位向量
branch instruction　分支指令
program counter　程序计数器
addressing mode　寻址模式
global data　全局数据
automatic variable　自动变量
program call　程序调用
ISO *abbr*. International Organization for Standardization　国际标准化组织
RTOS *abbr*. real-time operating system　实时操作系统

Notes

1. “竞争状态”(race condition)是指当两个实体(如两个进程)对同一资源进行竞争时，因系统没有判定执行顺序的机制而导致的结果不可预测状态。
2. Java 语言是由 Sun 公司(Sun Microsystems)于 1995 年推出的一种面向对象的、跨平台解释性程序设计语言，其语法规则和 C＋＋类似。Java 已成为互联网上广泛使用的编程语言之一。

Lesson 27　Choosing a Real-Time Operating System

Engineers often use the term “real time” to describe computing problems for which a late answer is as bad as a wrong one. These problems are said to have deadlines, and embedded

systems frequently operate under such constraints. For example, if the embedded software that controls your anti-lock brakes misses one of its deadlines you might find yourself in an accident. So it's extremely important that the designers of real-time embedded systems know everything they can about the behavior and performance of their hardware and software.

The designers of real-time systems spend a large portion of their time worrying about worst-case performance. They must constantly ask themselves questions like: what is the worst-case time between the human operator pressing the brake pedal and an interrupt signal arriving at the processor? What is the worst-case interrupt latency? And what is the worst-case time for the software to respond by triggering the braking mechanism? Average or expected-case analysis simply will not suffice. Most of the commercial embedded operating systems available today are designed for possible inclusion in real-time systems. In the ideal case, their worst-case performance is well understood and documented.

RTOS' Characteristics

To earn the distinctive title "Real-Time Operating System" (RTOS), an operating system should be deterministic and have guaranteed worst-case interrupt latency and context-switch times. Given these characteristics and the relative priorities of the tasks and interrupts in your system, it is possible to analyze the worst-case performance of the software using a technique such as rate-monotonic analysis.

An operating system is said to be deterministic if the worst-case execution time of each of its system calls is calculable. An operating system vendor that takes the real-time behavior of its product seriously will usually publish a datasheet providing the minimum, average, and maximum number of clock cycles required by each system call. These numbers may be different for different processors, but it is reasonable to expect that if the algorithm[1] is deterministic on one processor it will be so on any other. (The actual times may differ, however.)

Interrupt latency[2] is the total length of time from an interrupt signal arriving at the processor to the start of the associated interrupt service routine (ISR). When an interrupt occurs, the processor must take several steps before executing the ISR. First, the processor must finish executing the current instruction. Next, the interrupt type must be recognized. This is done by the hardware and does not slow or suspend the running task. Finally, and only if interrupts are enabled, the CPU's context is saved and the ISR associated with the interrupt is started.

Of course, if interrupts are ever-disabled (say within a system call), the worst-case interrupt latency increases by the maximum amount of time that they are turned off. Each operating system will internally disable interrupts in several places and for different lengths of time, so it is important that you know what your system's requirements are. One real-time project might require a guaranteed interrupt response time as short as 1 μs, while another may require only 100 μs.

The third real-time characteristic of an operating system is the amount of time required to

perform a context switch. This is important because it represents overhead across your entire system. For example, imagine if the average execution time of any task before it blocks is 100ms but that the context-switch time is also 100 ms. In that case, fully one half of the processor's time is spent within the context-switch routine!

Again, there is no magic number and the actual times are usually processor-specific, since they are dependent on the number of registers that must be saved and where. Be sure to get these numbers from any operating system vendor you are thinking of using. That way, there won't be any last minute surprises.

Selection Process

Considering the cost of engineering time these days, a few thousand dollars is a bargain for a commercial RTOS. A wide variety of operating systems are available to suit most projects and pocketbooks. Commercial operating systems form a continuum of functionality, performance, and price. Those at the lower end of the spectrum offer just a basic preemptive scheduler and a few other key system calls. These operating systems are usually inexpensive, come with source code that you can modify, and do not require payment of any royalties. Operating systems at the other end of the spectrum typically include a lot of functionality beyond the basic scheduler. These operating systems can be quite expensive, though, with startup costs ranging from $10,000 to $50,000 and royalties due on every copy shipped in ROM. However, this price often includes free technical support and training and a set of integrated development tools. Between these two extremes are the operating systems that have modest up-front costs and/or royalties, but do not include source code; technical support may cost extra. Most commercial operating systems fall in this category.

With such a variety of operating systems and features to choose from, it can be difficult to decide which is the best for your project. Try putting your processor, real-time performance, and budgetary requirements first. These are criteria that you cannot change, so you can use them to narrow the possible choices to a smaller set of products. Then contact all of the vendors of the remaining operating systems for more detailed technical information.

At this point, many people make their decision based on compatibility with their choice of cross-compiler, debugger, and other development tools. But it's really up to you to decide what additional features are most important for your project. No matter what you decide to buy, the basic kernel will be about the same. The differences will most likely be measured in processor support, minimum and maximum memory requirements, availability of add-on software modules (for example, networking protocol stacks and device drivers), and compatibility with third-party development tools.

The best reason to choose a commercial operating system is the advantage of using something that is better tested and, therefore, more reliable than a kernel you have developed inhouse. So one of the most important things you should be looking for from your OS vendor is experience.

New Words & Expressions

constraint [kənˈstreɪnt] *n*. 约束
brake [breɪk] *n*. 刹车
suffice [səˈfaɪs] *vi*. 足够
inclusion [ɪnˈkluːʒən] *n*. 内含物
documented [ˈdɒkjʊməntɪd] *adj*. 有文字说明的
deterministic [dɪˈtɜːmɪˌnɪstɪk] *adj*. 确定性的
context [ˈkɒntekst] *n*. 环境
calculable [ˈkælkjʊləbl] *adj*. 可计算的
disable [dɪsˈeɪbl] *vt*. 禁用
bargain [ˈbɑːgɪn] *n*. 便宜货
a portion of ... 一部分……
pocketbook [ˈpɒkɪtˌbʊk] *n*. 财力
continuum [kənˈtɪnjʊəm] *n*. 连续区间
spectrum [ˈspektrəm] *n*. 范围
preemptive [prɪˈemptɪv] *adj*. 抢先的
royalty [ˈrɒɪəlti] *n*. 特许权使用费
a variety of ... 各种各样的……
extreme [ɪkˈstriːm] *n*. 极端
budgetary [ˈbʌdʒɪtəri] *adj*. 预算的
criteria [kraɪˈtɪərɪə] *n*. 标准
protocol [ˈprəʊtəkɒl] *n*. 协议
stack [stæk] *n*. 堆栈

Technical Terms

latency [ˈleɪtənsi] *n*. 反应时间
algorithm [ˈælgərɪðəm] *n*. 算法
scheduler [ˈʃedjuːələ] *n*. 调度程序
debugger [diːˈbʌgə] *n*. 调试程序
kernel [ˈkɜːnəl] *n*. 内核程序
interrupt latency 中断等待时间
system call 系统调用
execution time 执行时间
device driver 设备驱动程序
source code 源代码
startup cost 启动成本
integrated development tool 集成开发工具
cross compiler 交叉编译程序
RTOS *abbr*. real-time operating system 实时操作系统
ISR *abbr*. interrupt service routine 中断服务程序

Notes

1. 所谓“算法”(algorithm)，就是一种在有限步骤内解决问题的计算程序。
2. 任务的“反应时间”(latency)或“迟滞时间”(tardiness)是指其实际启动(或结束)的时刻和应该启动(或结束)的时刻之差。决定“反应时间”的主要因素有：(1)处理器、总线、存储器和外设的时序特性；(2)操作系统的调度特性；(3)操作系统内核的抢占性；(4)系统载荷；(5)任务切换时间。

Exercises

1. Fill in the blanks with proper words, phrases or clauses.

(1)An embedded system is any electronic system that uses a CPU chip, but that is not a _____(通用的)workstation, desktop _____ laptop computer._____(这样的系统一般使用微处理器), or they may use _____(定制芯片)or both. They are used in automobiles, planes, trains, space vehicles, machine tools, cameras, consumer and office appliances, cell-phones and other hand-helds _____ robots and toys. The uses are endless, and billions of microprocessors are shipped every year for a myriad of applications. Although there are embedded versions of popular _____(操作系统), _____(低成本消费类产品)can use chips that cost less than a dollar and have _____(非常有限的指令存储器)._____(在此类情况下), the OS and application may be combined into one program.

In embedded systems, the software is permanently set into a read-only memory such as a ROM or flash memory chip, in contrast _____ a general-purpose computer that _____(将其程序载入随机存取存储器中)each time. Sometimes, single board and rack mounted general-purpose computers _____(被称作嵌入式计算机)if used to control a single printer, drill press or other such device.

(2)Multithreading means multitasking within _____(一个程序). It allows multiple streams of execution to _____(并行发生) within the same program, each stream processing a different _____(事务或者消息). In order _____(使一个多线程程序真能获得性能提高), it must be run _____ a multitasking or multiprocessing environment, which allows multiple operations _____ take place.

Certain types of applications lend themselves to multithreading. For example, in an order processing system, each order can be entered independently of the other orders. In an image editing program, a _____ (运算密集的)filter can be performed on one image, while the user works on another. In a _____(对称的)multiprocessing (SMP) operating system, its multithreading allows multiple CPUs to be controlled _____ the same time. It is also used to create synchronized _____(音视频) applications.

Multithreading generally uses _____(可重用代码), which cannot be modified when executing, _____ the same code can be shared by multiple programs.

2. Translate the following passages into Chinese or English.

(1) Hardware means machinery and equipment (CPU, disks, tapes, modem, cables, etc.). In operation, a computer is both hardware and software. One is useless without the other. The hardware design specifies the commands it can follow, and the instructions tell it what to do.

(2) Firmware is a category of memory chips that hold their content without electrical power and include ROM, PROM, EPROM and EEPROM technologies. Firmware becomes "hard software" when holding program code.

(3) Software is "logic and language". Software deals with the details of an ever-changing business and must process transactions in a logical fashion. Languages are used to program the software. The "logic and language" involved in analysis and programming is generally far more complicated than specifying a storage and transmission requirement.

(4) Software package is an application program developed for sale to the general public. Packaged software is generally designed to appeal to a large audience of users, and although the programs may be tailored to a user's taste by setting various preferences, it is not as individualized as custom-designed and custom-programmed software.

(5) Process is the unique execution of a particular piece of code by a specific user on the same computer. Thus, if Alice runs the Microsoft Word at 8:30 and then again at 8:36, she has generated two processes. If Alice and Bob both run the Word at 8:30 on the same computer, they have also generated two processes.

(6) "实时系统"能够对输入信号做出快速响应，从而维持所需的处理速度。

(7) "实时操作系统"是一种专门为"实时计算机系统"设计的操作系统。

(8) 系统软件是用于控制计算机和开发、运行应用程序的程序。它包括操作系统、TP 监控程序、网络操作系统和数据库管理器。

(9) 硬件表现为"存储和传输"。一部计算机拥有的内存和磁盘存储器越多，它能做的工作也越多。内存和磁盘向 CPU 传送数据和指令的速度越快，CPU 完成得也越快。硬件需求基于将创建的数据库大小和用户数量或需同时服务的应用数量。

(10) 软件是计算机指令。"程序"就是用以完成某项工作的一组指令。软件分为"系统软件"和"应用软件"两大类。"系统软件"由操作系统、数据库管理系统等控制程序组成；"应用软件"是用户处理数据的程序(如电子制表软件、文字处理软件等)。人们常把"软件"误认为是"数据"，其实不然。软件告诉硬件应当如何处理数据。软件是"被运行的"，而数据是"被处理的"。

3. Writing tasks.

Summary: Write an abstract for *Lesson 26 Design Languages for Embedded Systems* within 150～200 words.

Composition: Write an essay about the design language you often use.

Reading Materials

Passage 1 ARM

What Is ARM?

ARM, Ltd. is a company in England that develops and designs a processor architecture. The ARM abbreviation for the processor design stands for Acorn RISC Machine, and the ARM abbreviation for the company that designs and sells the license to use that architecture stands for Advanced RISC Machines.

RISC stands for reduced-instruction-set computer. The Intel or AMD processor in your laptop or desktop computer is likely a CISC (complex-instruction-set computer) processor. The two different types are designed for different needs. A RISC processor is designed to run a smaller amount of instructions than a CISC processor. Because they can do fewer things, they can have a higher frequency and perform more MIPS (millions of instructions per second) than a CISC processor.

When you reduce the number of instructions the processor can compute, you can create a simpler circuit inside the chip. A RISC processor uses fewer transistors which in turn use less power. Because the circuits are simple, a smaller die size can be used to build the processor. Die size is the measurement of one chip on the silicon wafer that a processor is built on. When the die size is smaller, more components with less wiring can be placed on the processor surface. This makes ARM processors small and much less power-hungry.

Small, fast and simple processors are perfect for things like phones. A phone doesn't try to run hundreds of threads on its limited number of cores. Mobile software, both the operating system and applications that run on it, are coded and optimized for the reduced instruction set the ARM processor uses. But that doesn't mean ARM CPUs aren't powerful in their own right.

The current ARM specification allows for 32-bit and 64-bit design, hardware virtualization, advanced power management that can interface with user software, and a load/store architecture that is mostly single-cycle execution and orthogonal.

The ARM Business Model

ARM Ltd. does not manufacture and sell CPU devices based on their own designs, but rather, licenses the processor architecture to interested parties (see Figure 9.1). ARM offers a variety of licensing terms, varying in cost and deliverables. To all licensees, ARM provides an integratable hardware description of the ARM core, as well as complete software development toolset (compiler, debugger, SDK), and the right to sell manufactured silicon containing the ARM CPU.

Fabless licensees, who wish to integrate an ARM core into their own chip design, are

usually only interested in acquiring a ready-to-manufacture verified IP core. For these customers, ARM delivers a gate netlist description of the chosen ARM core, along with an abstracted simulation model and test programs to aid design integration and verification.

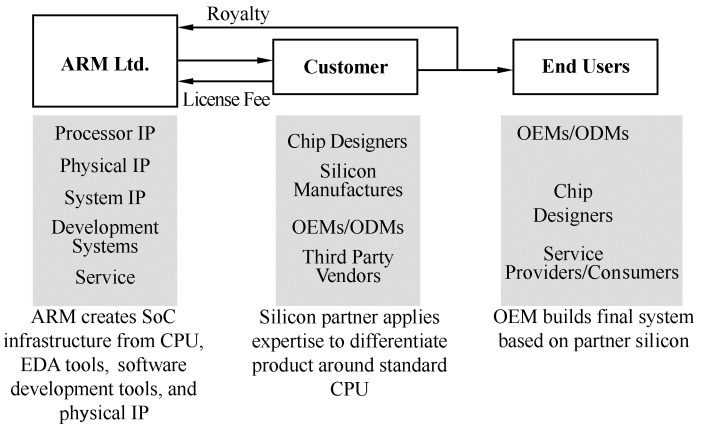


Figure 9.1 The ARM business model

More ambitious customers, including integrated device manufacturers (IDM) and foundry operators, choose to acquire the processor IP in synthesizable RTL (Verilog) form. With the synthesizable RTL, the customer has the ability to perform architectural level optimizations and extensions. This allows the designer to achieve exotic design goals not otherwise possible with an unmodified netlist (high clock speed, very low power consumption, instruction set extensions, etc.).

While ARM does not grant the licensee the right to resell the ARM architecture itself, licensees may freely sell manufactured product (chip devices, evaluation boards, complete systems, etc.). Merchant foundries can be a special case: not only are they allowed to sell finished silicon containing ARM cores, they generally hold the right to remanufacture ARM cores for other customers.

ARM creates SoC infrastructure from CPU EDA tools, software development tools, and physical IP silicon partner applies expertise to differentiate product around standard CPU OEM (original equipment manufacturer) builds final system based on partner silicon. Many semiconductor firms hold ARM licenses.

Although ARM's license terms are covered by NDA(non-disclosure agreement), within the IP industry, ARM is widely known to be among the most expensive CPU cores. A single customer product containing a basic ARM core can incur a one-time license fee in excess of $ 200,000. Where significant quantity and architectural modification are involved, the license fee can exceed $10 M.

With the diversity of ARM IP and the broad ecosystem of supporting silicon and software for ARM-based solutions, the world's leading OEMs use ARM technology in a wide variety of applications ranging from mobile handsets and digital set top boxes to car braking systems and network routers. Today, ARM technology is used in more than 95% of the world's

mobile handsets and over one-quarter of all electronic devices.

Questions:

1. What does the acronym ARM stand for?
2. For what application are ARM processors mainly used?
3. Could you describe the ARM business model?
4. What do you think is the main reason for ARM's popularity?
5. What does the term *netlist* mean in this article?

Passage 2 Embedded OS

Embedded operating systems have some features that distinguish them from real-time and general-purpose operating systems. But the definition of an "embedded operating system "is probably even more ambiguous than that of an RTOS, and they come in a zillion different forms. But you'll recognize one when you see one, although the boundary between general-purpose operating systems and embedded operating systems is not sharp.

Embedded systems are being installed in tremendous quantities: they control lots of functions in modern cars; they show up in household appliances and toys; they control vital medical instrumentation; they make remote controls and GPS work; they make your portable phones work; etc.

Classification

The simplest classification between different kinds of embedded operating systems is as follows:

High-end embedded OS. These systems are often downsized derivatives of an existing general purpose OS, but with much of the "ballast" removed. Linux has given rise to a large set of such derivatives, because of its highly modular structure and the availability of source code.

Deeply embedded OS. These OSes must be really very small, and need only a handful of basic functions. Therefore, they are mostly designed from the ground up for a particular application. Two typical functions deeply embedded systems (used to) lack are high-performance graphical user interfacing or network communication.

Features

The most important features that make an OS into an embedded OS are:

It should be as cheap as possible. Embedded systems are often produced in quantities of several thousands or even millions. Decreasing the unit price even a little bit boils down to enormous savings.

It should have small footprint. Designers are continuously trying to put more computing power in smaller housings, using cheaper CPUs, with on-board digital and/or analog IO; and they want to integrate these CPUs in all kinds of small objects. A small embedded OS also

often uses only a couple of kilobytes of RAM and ROM memory.

It should never fail. The embedded system should run for years without manual intervention. This means that the hardware and the software should never fail. Hence, the system should preferably have no mechanical parts. Not only because mechanical parts are more sensitive to failures, but they also take up more space, need more energy, take longer to communicate with, and have more complex drivers (e.g., due to motion control of the mechanical parts).

It should reset the system when necessary. Many embedded systems have to control devices that can be dangerous if they don't work exactly as designed. Therefore, the status of these devices has to be checked regularly. The embedded computer system itself, however, is one of these critical devices, and has to be checked too! Hence, one often sees hardware watchdogs included in embedded systems. These watchdogs are usually retriggerable monostable timers attached to the processor's reset input. The operating system checks within specified intervals whether everything is working as desired, for example by examining the contents of status registers. It then resets the watchdog. So, if the OS doesn't succeed in resetting the timer, that means that the system is not functioning properly and the timer goes off, forcing the processor to reset.

If something went wrong but the OS is still working (e.g., a memory protection error in one of the tasks), the OS can activate a software watchdog, which is nothing else but an interrupt that schedules a service routine to handle the error. One important job of the software watchdog could be to generate a core dump, to be used for analysis of what situations led to the crash.

If the system does fail despite its designed robustness (e.g., caused by a memory protection fault), there is usually no user around to take the appropriate actions. Hence, the system itself should reboot autonomously, in a "safe" state, and "instantly" if it is supposed to control other critical devices. Compare this to the booting of your desktop computer, which needs a minute or more before it can be used, and always comes up in the same default state.

It should support dynamic linking. Some embedded systems are not physically reachable anymore after they have been started (e.g., launched satellites) in order to add software updates. However, more and more of them can still be accessed remotely. Therefore, they should support dynamic linking: object code that did not exist at the time of start is uploaded to the system, and linked in the running OS without stopping it.

Some applications require all features of embedded and real-time operating systems. The best-known example are mobile phones and handheld computers: they must be small, consume little power, and yet be able to execute advanced signal processing algorithms, while taking up as little space as possible.

The above-mentioned arguments led embedded OS developers to design systems with the absolute minimum of software and hardware. Roughly speaking, developers of general purpose and real-time operating systems approach their clients with a "Hey, look how much

we can do! " marketing strategy; while EOS developers say "Hey, look how little we need to do what you want! " Hence, embedded systems often come without a memory management unit (MMU), multi-tasking, a networking "stack" or file systems.

Taking out more and more features of a general-purpose operating system makes its footprint smaller and its predictability higher. On the other hand, adding more features to an EOS makes it look like a general purpose OS. Most current RTOS and EOS operating systems are expanding their ranges of application, and cover more of the full "feature spectrum."

Questions:

1. How are embedded operating systems classified?
2. Can you tell something about Linux?
3. Why hardware watchdogs are often seen included in embedded systems?
4. What applications require all features of embedded and real-time operating systems?
5. Can you give the reasons why general-purpose OSes and embedded OSes take different marketing strategy?

Passage 3 Android Tablets

More than other tablet operating systems (OS), Android goes from low to high. The broad flexibility of Google's OS lets manufacturers build useful products ranging from $50 to more than $500, fitting a broader range of niches and tastes than the Apple iPad, which has dominated the tablet market since it first came out.

While Android dukes it out with the iPad and Windows tablets at the top of its range, and doesn't always win, its strengths at smaller sizes and lower price points can be hard to beat. Android tablets also sometimes show up with crazy experimental features, like a pico projector or a separate Wacom touchpad for artists. But all of those innovative ideas make Android a little harder to program for than iOS or Windows, so you may find that hot games and top productivity apps come to one of the other platforms first.

Start With Software

Unfortunately, Android tablet manufacturers don't tend to keep their software up to date. While the most recent version of Android is issued, most tablets are still on an older iteration.

Companies like Amazon and Samsung like to put their own spin on Google's OS, adding a bevy of features, new app stores, and completely revamped user interfaces. It's not for everyone, but those who are familiar with Amazon's Fire OS or Samsung's user experience should give the latest devices from each company a try.

Even if you don't have the most recent OS iteration, Android continues to be the most configurable tablet operating system you'll find. It's a master multitasker, with an excellent notification system and top-notch integration with Google services like Gmail, Hangouts, and Google Maps.

Android is home to plenty of great apps, but there still aren't as many tablet-specific options as you'll find for the iPad.

Cellular or Wi-Fi?

All of the major carriers are currently selling cellular-enabled tablets. You can add them to a monthly family plan, or pay for data by the day. T-Mobile tablets are unlocked and can be used on any service that the tablets' radios are compatible with, while AT&T and Sprint tablets tend to be locked to that service.

If you have a hotspot option on your smartphone plan, you can save money by just using that with a Wi-Fi-enabled tablet. But that will burn your smartphone's battery. And unless you're on an unlimited data plan, be aware that streaming video on a tablet will quickly eat up an LTE(long term evolution) service plan.

Don't Forget About Design

Performance on even the least expensive devices tends to be passable these days, so pay extra attention to design and display quality. The software experience might improve over time, but that chintzy plastic body isn't getting any upgrades.

Low-cost tablets also tend to have dull, 1,024-by-600 or 1,280-by-800 screens, that can look grainy to eyes used to even midrange smartphones. Higher-quality tablets are often in the 2,048-by-1,536 range, which is notably sharper. We recommend IPS(in-plane switching) and AMOLED(active matrix organic light emitting display) displays for their wide viewing angles, vibrant colors, and excellent contrast.

There are plenty of bargain bin options out there promising the same Android experience as big names. While many of these off-label tabs are perfectly serviceable, we recommend choosing a brand you can count on for software support and hardware quality control.

Questions:

1. What does the phrase *niches and tastes* mean in the first paragraph?
2. Compared with iOS or Windows, what advantages and disadvantages does Android have?
3. What does the term *streaming video* mean in this article?
4. Why design and display quality are also important when choosing a tablet?
5. What does the acronym IPS mean in this article?

Unit 10

Instrument & Measurement

Lesson 28 Signal Sources

Lesson 29 Oscilloscopes

Lesson 30 Logic Analyzers

Passage 1 Understanding Waveforms

Passage 2 Signal Integrity

Passage 3 Virtual Instrumentation

Lesson 28 Signal Sources

The Complete Measurement System

An acquisition instrument — usually an oscilloscope or logic analyser — is probably the first thing that comes to mind when you think about making electronic measurements. But these tools can only make a measurement when they are able to acquire a signal of some kind. And there are many instances in which no such signal is available unless it is externally provided.

A strain gauge amplifier, for example, does not produce signals; it merely increases the power of the signals it receives from a sensor. Similarly, a multiplexer on a digital address bus does not originate signals; it directs signal traffic from counters, registers, and other elements. But inevitably it becomes necessary to test the amplifier or multiplexer before it is connected to the circuit that feeds it. In order to use an acquisition instrument to measure the behavior of such devices, you must provide a stimulus signal at the input.

To cite another example, engineers must characterize their emerging designs to ensure that the new hardware meets design specifications across the full range of operation and beyond. This is known as margin or limit testing. It is a measurement task that requires a complete solution; one that can generate signals as well as make measurements.

The toolset for digital design characterization differs from its counterpart in analog/mixed signal design, but both must include stimulus instruments and acquisition instruments. The signal source, or signal generator, is the stimulus source that pairs with an acquisition instrument to create the two elements of a complete measurement solution.

What Is a Signal Source?

A signal source is nothing less than the cornerstone of almost any instrumentation setup used in hardware design, debug, or evaluation projects [1]. It is a key engineering tool. It is an essential troubleshooting aid for the technician. It is a surrogate for an automotive ignition pulse, a heart pacemaker, or a guided missile's gyro output. Second only to the ubiquitous DMM, signal sources are perhaps the most universal class of electronic test instruments [2].

Unless you're working with a purely DC circuit, your circuit is likely to require some kind of AC[3] stimulus signal as you evaluate components, functional blocks, and subsystems. The waveform from the signal source emulates a signal coming in from the outside world, such as a sensor output. Similarly, it can be used as a stand-in for waveforms that will appear in as-yet-unavailable parts of the circuit design.

Interestingly, the signal source's job is not simply to provide an "ideal" waveform. Often the instrument must add known, repeatable amounts and types of distortion (or errors) to the signal it delivers. This characteristic is one of the signal source's strongest virtues, since it is often impossible to create predictable distortion exactly when and where it's needed using only the circuit itself. The response of the unit under test (UUT) in the presence of these distorted signals reveals its capability to handle stresses that fall outside the normal performance envelope.

A stimulus signal may take the form of a low-distortion sine wave, a stream of logic pulses, a high-frequency radio carrier wave, a mobile telephone transmission, and many other formats. Traditionally, the task of producing these diverse waveforms has been filled by separate, dedicated signal sources, from ultra-pure audio sine-wave generators to multi-GHz RF signal generators. While there are many commercial solutions, the user must often customize or modify a signal source for the project at hand. It can be very difficult to design an instrumentation-quality signal generator, and of course, the time spent designing ancillary test equipment is a costly distraction from the project itself.

Fortunately, digital sampling technology and signal processing techniques have brought us a solution that answers almost any kind of signal generation need with just one type of instrument—the arbitrary generator.

Types of Digital Signal Sources

Broadly divided into arbitrary waveform generators (AWG), arbitrary function generators (AFG), and data or pattern generators (DG), digital signal sources span the whole range of signal-producing needs. Each of these types has its unique strengths:

- **AWG** — Whether you want a data stream shaped by a precise Lorentzian pulse for disk-drive characterization, or a complex modulated RF signal to test a GSM — or CDMA — based telephone handset, the AWG can produce any waveform you can imagine. You can use a variety of methods — from mathematical formulae to "drawing" the waveform—to create the needed output.
- **AFG** — Typically this instrument offers fewer waveform variations, but with excellent stability and fast response to frequency changes. If the UUT requires the classic "sine and square" waveforms (to name a few) and the capability to switch almost instantly between two frequencies, the AFG is the right tool. An additional virtue is the AFG's low cost, which makes it very attractive for applications that do not require an AWG's versatility.
- **DG** — This third type of signal source meets the special stimulus needs of digital devices that require long, continuous streams of binary data, with specific information content and timing characteristics.

Signal Generation Techniques

There are several ways to create waveforms with a signal source. The choice of methods

depends upon the information available about the device under test (DUT) and its input requirements; whether there is a need to add distortion or error signals, and other variables. Modern high-performance signal sources offer at least three ways to develop waveforms:

- **Simulation** — "Building" an event or sequence of events, based on a specific waveform definition (often from a simulator or a library of waveforms).
- **Replication** — Capturing an existing signal on an oscilloscope and sending it to the signal source for reproduction (see Figure 10.1).
- **Substitution** — Creating and/or modifying a defined signal to substitute for a signal from unavailable circuitry (see Figure 10.1).

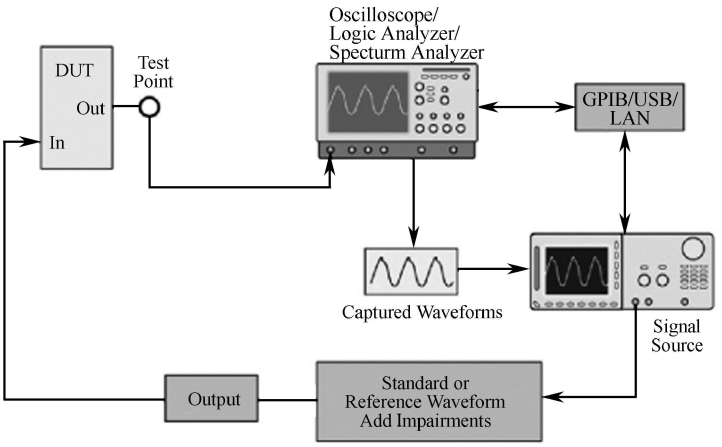


Figure 10.1 Signal sources can use standard, user-created, or captured waveforms, adding impairments where necessary for special test applications

Basic Signal Source Applications

Signal sources have hundreds of different applications. But in the electronic measurement context, they fall into three basic categories: verification, characterization, and stress/margin testing. Some representative applications include:

- **Verification — Analyzing Digital Modulation**: Wireless equipment designers developing new transmitter and receiver hardware must simulate baseband I&Q signals — with and without impairments — to verify conformance with emerging and proprietary wireless standards. Some high-performance arbitrary waveform generators can provide the needed low-distortion, high-resolution signals at rates up to 12.5 gigabit per second (12.5 Gbps), with two independent channels, one for the "I" phase and one for the "Q" phase.
- **Characterization — Testing D/A and A/D Converters**: Newly-developed digital-to-analog converters (DAC) and analog-to-digital converters (ADC) must be exhaustively tested to determine their limits of linearity, monotonicity, and distortion. A state-of-the-art[4] AWG can generate simultaneous, in-phase analog and digital signals to drive such devices at speeds up to 12.5 Gbps.

- **Stress/Margin Testing — Stressing Communication Receivers**: Engineers working with serial data stream architectures (commonly used in digital communications buses and disk drive amplifiers) need to stress their devices with impairments, particularly jitter and timing violations. Advanced signal sources save the engineer untold hours of calculation by providing efficient built-in jitter editing and generation tools. These instruments can shift critical signal edges as little as 20 ps.

AWG Hardware Architecture

Fundamentally, an AWG is a sophisticated "playback" system that delivers waveforms based on stored digital data that describes the constantly changing voltage levels of an AC signal. It is a tool whose block diagram is deceptively simple.

To understand the AWG, first it's necessary to grasp the broad concepts of digital sampling. Digital sampling is exactly what its name implies: defining a signal using samples, or data points, that represent a series of voltage measurements along the slope of the waveform. These samples may be determined by actually measuring a waveform with an instrument such as an oscilloscope, or by using graphical or mathematical techniques.

Figure 10.2(a) depicts a series of sampled points. All of the points are sampled at uniform time intervals, even though the curve makes their spacing appear to vary. In an AWG, the sampled values are stored in binary form in a fast random-access memory (RAM).

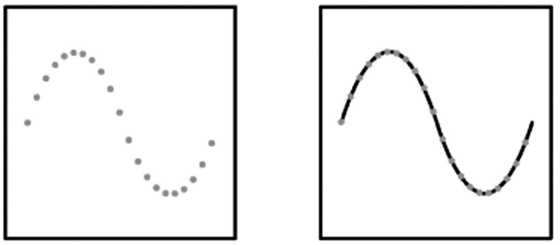
With the stored information, the signal can be reconstructed at any time by reading back the memory locations and feeding the data points through a DAC. Figure 10.2(b) depicts the result. Note that the AWG's output circuitry filters between the points to connect the dots and create a clean, uninterrupted waveform shape. The UUT does not "see" these dots as discrete points, but as a continuous analog waveform. Figure 10.3 is a simplified block diagram of an AWG that implements these operations.

As always, there are some terms and conditions that affect the fidelity of the signal reproduction. Chief among these is the Nyquist Sampling Theorem, which states that the sampling frequency must be at least twice that of the highest frequency component of the sampled signal. To sample a 1 MHz signal, for instance, it is necessary to acquire points at a frequency of at least 2 MS/s. Although the theorem is usually cited as a guideline for acquisition (as with an oscilloscope), its pertinence to AWGs is clear: stored waveforms must have enough points to faithfully retrace the details of the desired signal.

An AWG can take these points and read them out of memory at any frequency within its specified limits. If a set of stored points conforms to the Nyquist Theorem and describes a sine wave, then the AWG's output will be a sine wave. However, there is a finite maximum frequency, or sample rate, at which the instrument can operate. This is usually specified in terms of megasamples or gigasamples per second.

Other AWG hardware characteristics, particularly vertical resolution and memory depth, are just as important as sample rate. Vertical resolution (amplitude) expresses the voltage-

measuring precision of the sample points described above. Resolution pertains to the binary word width, in bits, of the instrument's DAC, with more bits equating to higher resolution. While "more is better", higher-frequency AWGs usually have lower resolution — 8 or 10 bits — than general-purpose AWGs offering 12 or 14 bits.



(a) A series of sampled points (b) The result

Figure 10.2 AWGs use stored digital samples to construct a waveform

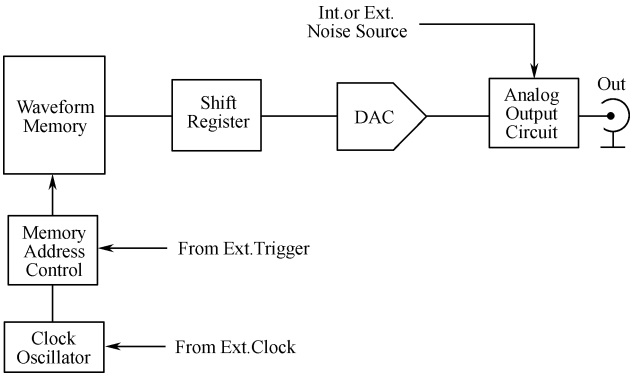


Figure 10.3 AWG block diagram (simplified)

An AWG with 10-bit resolution provides 1024 sample levels spread across the full voltage range of the instrument. If, for example, this 10-bit AWG has a total voltage range of 2 $V_{p\text{-}p}$, each sample represents a step of approximately 2 mV — the smallest increment the instrument could deliver, assuming it is not constrained by other factors in its architecture.

Memory depth in an AWG plays a key role in the instrument's flexibility. More (deeper) memory provides either of two benefits:

- **More cycles of the desired waveform can be stored.** This is useful because it reduces the number of "endpoints". An endpoint is the last memory location occupied by the waveform, after which the AWG must wrap around and return to the beginning in order to continue to produce the output signal. There are unavoidable errors that occur at this transition, so it is desirable to minimize the number of endpoints.
- **More waveform detail can be stored.** Complex waveforms have high-frequency information in their pulse edges and transients. It is difficult to interpolate these fast transitions as we did with the simple, predictable sine wave. To faithfully reproduce a complex signal, the available waveform memory capacity must be used to store more transitions and fluctuations rather than more cycles of the signal.

When combined with the high sample rate that some top models deliver, these instruments can store and reproduce complex RF waveforms, even including pseudo-random bit streams for use in physical-layer[5] testing of network equipment. Similarly, these fast AWGs with deep memory can generate very brief digital pulses and transients.

New Words & Expressions

acquisition [ˌækwɪˈzɪʃn] *n*. 采集
originate [əˈrɪdʒɪneɪt] *vt*. 引起
stimulus [ˈstɪmjʊləs] *n*. 激励
margin [ˈmɑːdʒɪn] *n*. 边缘
cornerstone [ˈkɔːnəˌstəʊn] *n*. 基石
evaluation [ɪˌvæljʊˈeɪʃn] *n*. 评估
troubleshooting [ˈtrʌblˌʃuːtɪŋ] *n*. 检修
surrogate [ˈsʌrəgət] *n*. 代用品
automotive [ˌɔːtəˈməʊtɪv] *adj*. 汽车的
ignition [ɪgˈnɪʃn] *n*. 点火
pacemaker [ˈpeɪsˌmeɪkə] *n*. 心脏起搏器
ubiquitous [juːˈbɪkwɪtəs] *adj*. 无所不在的
stand-in *n*. 替身
ancillary [ænˈsɪləri] *adj*. 补助的，副的
formulae [ˈfɔːmjʊliː] *n*. 公式(复数)
characterization [ˌkærəktəraɪˈzeɪʃn] *n*. 特性描述
impairment [ɪmˈpeəmənt] *n*. 损害
conformance [kənˈfɔːməns] *n*. 一致
proprietary [prəˈpraɪətəri] *adj*. 专有的
untold [ʌnˈtəʊld] *adj*. 数不清的
built-in [ˈbɪltˈɪn] *adj*. 内置的
spacing [ˈspeɪsɪŋ] *n*. 间隔
pertinence [ˈpɜːtɪnəns] *n*. 相关性
constrain [kənˈstreɪn] *vt*. 约束
endpoint [ˈendˌpɒɪnt] *n*. 端点
pseudo [ˈsjuːdəʊ] *adj*. 假的

Technical Terms

oscilloscope[əˈsɪləˌskəʊp] *n*. 示波器
multiplexer [ˈmʌltɪˌpleksə] *n*. 多路复用器
instrumentation [ˌɪnstrʊmenˈteɪʃn] *n*. 仪器
debug [diːˈbʌg] *vt*. 调试
gyro [ˈdʒaɪrəʊ] *n*. 陀螺仪
subsystem [ˈsʌbsɪstəm] *n*. 子系统
distortion [dɪˈstɔːʃn] *n*. 畸变
linearity [ˌlɪnɪˈærɪti] *n*. 线性
monotonicity [mɒnətəˈnɪsɪti] *n*. 单调性
jitter [ˈdʒɪtə] *n*. 抖动
transient [ˈtrænzɪənt] *n*. 暂态过程
in-phase 同相
logic analyzer 逻辑分析仪
stimulus signal 激励信号
guided missile 导弹
full range 满量程

design specification　设计规格
desired signal　期望信号
vertical resolution　垂直分辨率
DMM *abbr*. digital multimeter　数字万用表
UUT *abbr*. unit under test　被测单元
AWG *abbr*. arbitrary waveform generator　任意波形发生器
AFG *abbr*. arbitrary function generator　任意函数发生器
DG *abbr*. data generator　数据发生器

Notes

1. 此句可译为：信号源是用于硬件设计、调试和评价工程中几乎任何一种测试配置的基础。nothing less than 是“不亚于”“完全”的意思。
2. 此句可译为：在电子测试仪器中，信号源也许是仅次于无处不在的数字万用表的最通用的仪器。second only to 是“仅次于”的意思。
3. 通常情况下，AC 是指在 0 V 上下交替改变方向的信号(即交流信号)。在有些情况下，AC 可以指任何变化的信号(如幅度在+1～+3 V 之间变化的信号)。
4. state-of-the-art 是指一种设备、一门技术或一个科学领域在某个特定时代取得的最高发展水平。
5. 物理层(physical-layer)是数据通信中的开放系统互连(Open System Interconnection，OSI)参考模型的底层。OSI 参考模型由 7 层组成，其余 6 层依次为数据链路层(data link layer)、网络层(network layer)、传输层(transport layer)、会话层(session layer)、表示层(presentation layer)和应用层 (application layer)。OSI 模型中的每一层执行一项特定的任务。其中，物理层负责位流的传输。

Lesson 29　Oscilloscopes

Nature moves in the form of a sine wave, be it an ocean wave, earthquake, sonic boom [1], explosion, sound through air, or the natural frequency of a body in motion [2]. Energy, vibrating particles and other invisible forces pervade our physical universe. Even light — part particle, part wave — has a fundamental frequency, which can be observed as color. Sensors can convert these forces into electrical signals that you can observe and study with an oscilloscope. Oscilloscopes enable scientists, engineers, technicians, educators and others to “see” events that change over time.

Oscilloscopes are indispensable tools for anyone designing, manufacturing or repairing electronic equipment. In today's fast-paced world, engineers need the best tools available to solve their measurement challenges quickly and accurately. As the eyes of the engineer,

oscilloscopes are the key to meeting today's demanding measurement challenges.

The usefulness of an oscilloscope is not limited to the world of electronics. With the proper transducer, an oscilloscope can measure all kinds of phenomena. A transducer is a device that creates an electrical signal in response to physical stimuli, such as sound, mechanical stress, pressure, light, or heat. A microphone is a transducer that converts sound into an electrical signal.

Oscilloscopes are used by everyone from physicists to television repair technicians. An automotive engineer uses an oscilloscope to measure engine vibrations. A medical researcher uses an oscilloscope to measure brain waves. The possibilities are endless.

The Types of Oscilloscopes

Electronic equipment can be classified into two categories: analog and digital. Analog equipment works with continuously variable voltages, while digital equipment works with discrete binary numbers that represent voltage samples.

Oscilloscopes can be classified similarly — as analog and digital types. For many applications, either an analog or digital oscilloscope will do. However, each type has unique characteristics that may make it more or less suitable for specific applications. Digital oscilloscopes can be further classified into digital storage oscilloscopes (DSOs), digital phosphor oscilloscopes (DPOs) and digital sampling oscilloscopes.

Analog Oscilloscopes

Fundamentally, an analog oscilloscope works by applying the measured signal voltage directly to the vertical axis of an electron beam that moves from left to right across the oscilloscope screen — usually a cathode-ray tube (CRT). The backside of the screen is treated with luminous phosphor that glows wherever the electron beam hits it. The signal voltage deflects the beam up and down proportionally as it moves horizontally across the display, tracing the waveform on the screen. The more frequently the beam hits a particular screen location, the more brightly it glows.

The CRT limits the range of frequencies that can be displayed by an analog oscilloscope. At very low frequencies, the signal appears as a bright, slow-moving dot that is difficult to distinguish as a waveform. At high frequencies, the CRT's writing speed defines the limit. When the signal frequency exceeds the CRT's writing speed, the display becomes too dim to see. The fastest analog oscilloscopes can display frequencies up to about 1 GHz.

When you connect an oscilloscope probe to a circuit, the voltage signal travels through the probe to the vertical system of the oscilloscope. Figure 10.4 illustrates how an analog oscilloscope displays a measured signal. Depending on how you set the vertical scale (volts/div control), an attenuator reduces the signal voltage and an amplifier increases the signal voltage.

Next, the signal travels directly to the vertical deflection plates of the CRT. Voltage applied to these deflection plates causes a glowing dot to move across the screen. The glowing dot is created by an electron beam that hits the luminous phosphor inside the CRT. A positive voltage causes the dot to

move up while a negative voltage causes the dot to move down. The signal also travels to the trigger system to start, or trigger, a horizontal sweep. Horizontal sweep refers to the action of the horizontal system that causes the glowing dot to move across the screen.

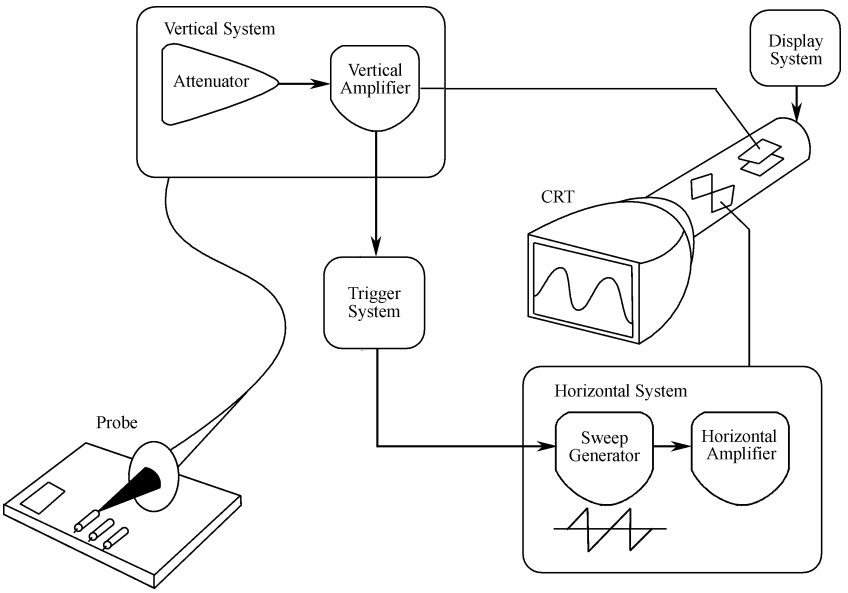


Figure 10.4 An analog oscilloscope

Triggering the horizontal system causes the horizontal time base to move the glowing dot across the screen from left to right within a specific time interval. Many sweeps in rapid sequence cause the movement of the glowing dot to blend into a solid line. At higher speeds, the dot may sweep across the screen up to 500,000 times per second.

Together, the horizontal sweeping action and the vertical deflection action trace a graph of the signal on the screen. The trigger is necessary to stabilize a repeating signal — it ensures that the sweep begins at the same point of a repeating signal.

In addition, analog oscilloscopes have focus and intensity controls that can be adjusted to create a sharp, legible display. People often prefer analog oscilloscopes when it is important to display rapidly varying signals in "real time" — or, as they occur. The analog oscilloscope's chemical phosphor-based display has a characteristic known as intensity grading that makes the trace brighter wherever the signal features occur most often. This intensity grading makes it easy to distinguish signal details just by looking at the trace's intensity levels.

Digital Oscilloscopes

In contrast to an analog oscilloscope, a digital oscilloscope uses an ADC to convert the measured voltage into digital information. It acquires the waveform as a series of samples, and stores these samples until it accumulates enough samples to describe a waveform. The digital oscilloscope then re-assembles the waveform for display on the screen. Digital oscilloscopes can be classified into digital storage oscilloscopes (DSOs), digital phosphor oscilloscopes (DPOs), and digital sampling oscilloscopes.

The digital approach means that the oscilloscope can display any frequency within its range with stability, brightness, and clarity. For repetitive signals, the bandwidth of the digital oscilloscope is a function of the analog bandwidth of the front-end components of the oscilloscope, commonly referred to as the −3 dB point. For single-shot and transient events, such as pulses and steps, the bandwidth can be limited by the oscilloscope's sample rate.

- **Digital Storage Oscilloscopes.** A conventional digital oscilloscope is known as a digital storage oscilloscope (DSO). Its display typically relies on a raster-type screen rather than luminous phosphor. DSOs allow you to capture and view events that may happen only once — known as transients. Because the waveform information exists in digital form as a series of stored binary values, it can be analyzed, archived, printed, and otherwise processed, within the oscilloscope itself or by an external computer. The waveform need not be continuous; it can be displayed even when the signal disappears. Unlike analog oscilloscopes, digital storage oscilloscopes provide permanent signal storage and extensive waveform processing. However, DSOs typically have no real-time intensity grading; therefore, they cannot express varying levels of intensity in the live signal.
- **Digital Phosphor Oscilloscopes.** The digital phosphor oscilloscope (DPO) offers a new approach to oscilloscope architecture. This architecture enables a DPO to deliver unique acquisition and display capabilities to accurately reconstruct a signal. While a DSO uses a serial-processing architecture (see Figure 10.5) to capture, display and analyze signals, a DPO employs a parallel-processing architecture to perform these functions, as shown in Figure 10.6. The DPO architecture dedicates unique ASIC hardware to acquire waveform images, delivering high waveform capture rates that result in a higher level of signal visualization. This performance increases the probability of witnessing transient events that occur in digital systems, such as runt pulses, glitches and transition errors.

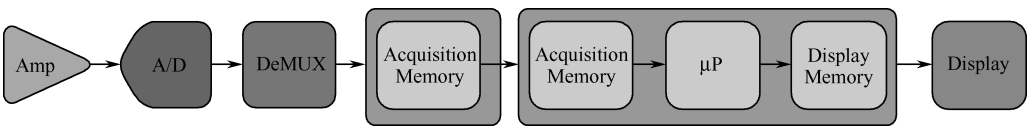


Figure 10.5 The serial-processing architecture of a digital storage oscilloscope (DSO)

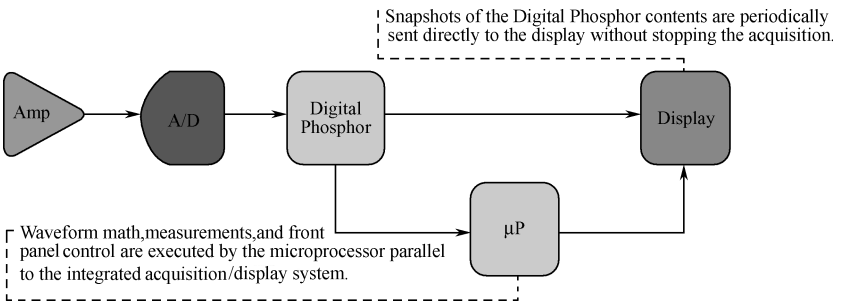


Figure 10.6 The parallel-processing architecture of a digital phospor oscilloscope (DPO)

- **Digital Sampling Oscilloscopes.** When measuring high-frequency signals, the oscilloscope may not be able to collect enough samples in one sweep. A digital sampling oscilloscope is an ideal tool for accurately capturing signals whose frequency components are much higher than the oscilloscope's sample rate. This oscilloscope is capable of measuring signals of up to an order of magnitude faster than any other oscilloscope. It can achieve bandwidth and high-speed timing ten times higher than other oscilloscopes for repetitive signals. Sequential equivalent-time sampling oscilloscopes are available with bandwidths to 80 GHz.

New Words & Expressions

vibrating [ˈvaɪbreɪtɪŋ] *adj*. 振动的
particle [ˈpɑːtɪkl] *n*. 粒子
pervade [pəˈveɪd] *vt*. 遍及
indispensable [ɪndɪˈspensəbl] *adj*. 不可缺少的
demanding [dɪˈmɑːndɪŋ] *adj*. 要求高的
phosphor[ˈfɒsfə] *n*. 磷
backside [bækˈsaɪd] *n*. 背面
luminous [ˈluːmɪnəs] *adj*. 发光的
legible [ˈledʒɪbl] *adj*. 清晰的
grading [ˈgreɪdɪŋ] *n*. 分级
intensity [ɪnˈtensɪti] *n*. 强度
archive [ˈɑːkaɪv] *vt*. 把　　存档

Technical Terms

transducer [trænzˈdjuːsə] *n*. 传感器
cathode [ˈkæ əʊd] *n*. 阴极
attenuator [əˈtenjʊeɪtə] *n*. 衰减器
raster [ˈræstə] *n*. 光栅
single-shot　单脉冲
electron beam　电子束
fundamental frequency　基频
CRT *abbr*. cathode-ray tube　阴极射线管

Notes

1. "声爆"(sonic boom)是一种在以音速或超音速飞行的飞行器前方的冲击波引起的爆炸声。
2. 此句可译为：自然界以正弦波的形式运动。不论是海浪、地震、声爆、爆炸、空气中传播的声音，还是运动物体的固有频率，都是如此。

Lesson 30　Logic Analyzers

The logic analyzer has different capabilities than the oscilloscope. The most obvious

difference between the two instruments is the number of channels (inputs). Typical digital oscilloscopes have up to four signal inputs. Each channel inputs one digital signal. Some complex system designs require thousands of input channels. Appropriately-scaled logic analyzers are available for those tasks as well.

A logic analyzer measures and analyzes signals differently than an oscilloscope. The logic analyzer doesn't measure analog details. Instead, it detects logic threshold levels. When you connect a logic analyzer to a digital circuit, you're only concerned with the logic state of the signal. A logic analyzer looks for just two logic levels. When the input is above the threshold voltage is said to be "high" or "1"; conversely, the level below the threshold voltage is a "low" or "0". When a logic analyzer samples input, it stores a "1" or a "0" depending on the level of the signal relative to the voltage threshold.

A logic analyzer's waveform timing display is similar to that of a timing diagram found in a data sheet or produced by a simulator. All of the signals are time-correlated, so that setup-and-hold time, pulse width, extraneous or missing data can be viewed. In addition to their high channel count, logic analyzers offer important features that support digital design verification and debugging. Among these are:

- **Sophisticated triggering** — Lets you specify the conditions under which the logic analyzer acquires data.
- **High-density probes and adapters** — Simplify connection to the system under test (SUT).
- **Analysis capabilities** — Translate captured data into processor instructions and correlate it to source code.

Architecture and Operation

The logic analyzer connects to, acquires, and analyzes digital signals. These are the four steps of using a logic analyzer:

① **Probe.**

② **Setup** (clock mode and triggering).

③ **Acquisition.**

④ **Analysis and display.**

Figure 10.7 is a simple logic analyzer block diagram. Each block symbolizes several hardware and/or software elements. The block numbers correspond to the four steps listed above.

Probe

The large number of signals that can be captured at one time by the logic analyzer is what sets it apart from the oscilloscope. The acquisition probes connect to the SUT. The probe's internal comparator is where the input voltage is compared against the threshold voltage (V_{th}), and where the decision about the signal's logic state (1 or 0) is made. The threshold value is set by the user, ranging from TTL levels to, CMOS, ECL, and user-definable.

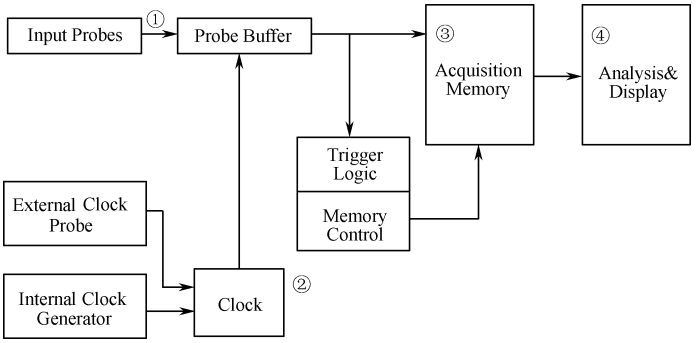


Figure 10.7 Simplified logic analyzer block diagram

The probes are capable of acquiring high-quality signals, and have a minimal impact on the SUT. The logic analyzer probe provides a high-quality signal path to the logic analyzer, minimizes electrical loading on the SUT and adapts to the various types of connections on circuit boards and devices [1].

The impedance of the logic analyzer's probes (capacitance, resistance, and inductance) becomes part of the overall load on the circuit being tested. All probes exhibit loading characteristics. The logic analyzer probe should introduce minimal loading on the SUT, and provide an accurate signal to the logic analyzer.

Probe capacitance tends to "roll off" the edges of signal transitions. Why is this important? Because a slower edge crosses the logic threshold of the circuit later, introducing timing errors in the SUT. This is a problem that becomes more severe as clock rates increase. In high-speed systems, excessive probe capacitance can potentially prevent the SUT from working! It is always critical to choose a probe with the lowest possible total capacitance.

Setup (Clock Mode and Triggering)

Logic analyzers are designed to capture data from multi-pin devices and buses. The term *capture rate* refers to how often the inputs are sampled [2]. It is the same function as the time base in an oscilloscope. There are two types of data acquisition:

- **Asynchronous acquisition**—Captures signal timing information. In this mode, a clock internal to the logic analyzer is used to sample data. The faster that data is sampled, the higher will be the resolution of the measurement. There is no fixed timing relationship between the target device and the data acquired by the logic analyzer. This acquisition mode is primarily used when the timing relationship between SUT signals is of primary importance.
- **Synchronous acquisition**—Is used to acquire the "state" of the SUT. A signal from the SUT defines the sample point (when and how often data will be acquired). The signal used to clock the acquisition may be the system clock, a control signal on the bus, or a signal that causes the SUT to change states. Data is sampled on the active edge and it represents the condition of the SUT when the logic signals are stable. The logic

analyzer samples when, and only when, the chosen signals are valid.

Acquisition: Real-time Acquisition Memory

The logic analyzer's probing, triggering, and clocking systems exist to deliver data to the real-time acquisition memory. This memory is the heart of the instrument — the destination for all of the sampled data from the SUT, and the source for all of the instrument's analysis and display.

Logic analyzers have memory capable of storing data at the instrument's sample rate. This memory can be envisioned as a matrix having width and depth.

The instrument accumulates a record of all signal activity until a trigger event or the user tells it to stop. The result is an acquisition — essentially a multi-channel waveform display that lets you view the interaction of all the signals you've acquired, with a very high degree of timing precision.

Both width and depth are key factors in choosing a logic analyzer. Following [3] are some tips to help you determine your channel count and memory depth:

- **How many signals do you need to capture and analyze**? Your logic analyzer's channel count maps directly to the number of signals you want to capture. Digital system buses come in various widths, and there is often a need to probe other signals (clocks, enables, etc.) at the same time the full bus is being monitored. Be sure to consider all the buses and signals you will need to acquire simultaneously.
- **How much "time" do you need to acquire**? This determines the logic analyzer's memory depth requirement, and is especially important for asynchronous acquisition. For a given memory capacity, the total acquisition time decreases as the sample rate increases. For example, the data stored in a 1M memory spans 1 second of time when the sample rate is 1 ms. The same 1M memory spans only 10 ms of time for an acquisition clock period of 10 ns. Acquiring more samples (time) increases your chance of capturing both an error, and the fault that caused the error. When it comes to memory capacity, you can't have too much!

Analysis and Display

The data stored in the real-time acquisition memory can be used in a variety of display and analysis modes. Once the information is stored within the system, it can be viewed in formats ranging from timing waveforms to instruction mnemonics correlated to source code.

The waveform display is a multi-channel detailed view that lets you see the time relationship of all the captured signals, much like the display of an oscilloscope. The waveform display is commonly used in timing analysis, and it is ideal for:

- **Diagnosing timing problems** (in SUT hardware).
- **Verifying correct hardware operation** (by comparing the recorded results with simulator output or data sheet timing diagrams).

- **Measuring hardware timing-related characteristics** (race conditions, propagation delays, absence or presence of pulses).
- **Analyzing glitches.**

The listing display provides state information in user-selectable alphanumeric form. The data values in the listing are developed from samples captured from an entire bus and can be represented in hexadecimal or other formats. The intent of the listing display is to show the state of the SUT. The listing display lets you see the information flow exactly as the SUT sees it — a stream of data words.

State data is displayed in several formats. The real-time instruction trace disassembles every bus transaction and determines exactly which instructions were read across the bus. It places the appropriate instruction mnemonic, along with its associated address, on the logic analyzer display.

An additional display, the source code debug display, makes your debug work more efficient by correlating the source code to the instruction trace history. It provides instant visibility of what's actually going on when an instruction executes.

New Words & Expressions

threshold [ˈθreʃəʊld] *n*. 门限
concerned [kənˈsɜːnd] *adj*. 有关的
diagram [ˈdaɪəgræm] *n*. 图表
setup [ˈsetʌp] *n*. 设置
extraneous [ɪkˈstreɪnɪəs] *adj*. 外来的
alphanumeric [ˌælfənjʊˈmerɪk] *adj*. 含字母与数字的
asynchronous [əˈsɪ krənəs] *adj*. 异步的
synchronous [ˈsɪ krənəs] *adj*. 同步的
envision [enˈvɪʒən] *vt*. 想象
enable [ɪnˈeɪbl] *vt*. 使能

Technical Terms

probe [prəʊb] *n*. 探头
loading [ˈləʊdɪ] *n*. 负载
impedance [ɪmˈpiːdəns] *n*. 阻抗
mnemonics [nɪˈmɒnɪks] *n*. 助记符
time base 时基
data acquisition 数据采集
timing diagram 时序图
pulse width 脉冲宽度
propagation delay 传输延迟
SUT *abbr*. system under test 被测系统

Notes

1. 此句可译为：逻辑分析仪探头为逻辑分析仪提供了高质量的信号通道，并将对 SUT 造成的电气负载降至最小，而且能够适配于各种电路板和器件。
2. 结构 *A refers to B* 用来表达“A 指代 B”的意思。*how often* ... 用来表达频度。此句可译为：“采集速率”是指对输入信号进行采样的频率。
3. Following 可用作名词、形容词或介词。在这里，Following 用作名词，意思为“下列各项”。

Exercises

1. Fill in the blanks with proper words, phrases or clauses.

(1)Proper grounding is an important step _____ setting up to take measurements or work _____ a circuit. Proper grounding of the oscilloscope protects you from a hazardous shock and grounding yourself protects your circuits _____ damage.

To ground the oscilloscope means to connect it to an electrically neutral _____(参考点), such as earth ground. Ground your oscilloscope by plugging its three-pronged power cord into an outlet grounded to earth ground.

Grounding the oscilloscope is necessary for safety. If _____(一个高电压)contacts the case of an ungrounded oscilloscope — any part of the case, _____ knobs that appear insulated — it can give you a shock. However, with a properly grounded oscilloscope, the current travels through the grounding path to earth ground _____ through you to earth ground.

Grounding is also necessary for _____(进行精确测量)with your oscilloscope. The oscilloscope needs to _____(公用)the same ground as any circuits you are testing.

Some oscilloscopes do not require separate connection to earth ground. These oscilloscopes have insulated cases and controls, which keeps any possible shock hazard away from the user.

If you are working with _____(集成电路), you also need to ground yourself. Integrated circuits have tiny _____(导电通路)that can be damaged by _____(静电)that builds up on your body. You can ruin an expensive IC simply by walking across a carpet or taking off a sweater and then touching the _____(引线)of the IC. To solve this problem, wear a grounding strap. This strap safely sends static _____(电荷)on your body to earth ground.

(2)Many engineers look at tasks like troubleshooting and _____(设计验证) as purely “measurement” challenges, and tend to think reflexively of their _____(示波器或者逻辑分析仪) as the whole solution _____ the problem. But these acquisition instruments have an important _____(伙伴) in their work: the _____(激励) instrument — the signal source.

Stimulus and acquisition instruments together _____（构成）a complete solution that can drive a _____（被测设备）with complex _____（现实世界信号）and acquire the resulting outputs. The oscilloscope is the _____（工业标准工具）for acquisition. But only with a signal source _____ engineers really control what goes into the device. And that is often a necessity _____ making sense of what comes out of the device.

_____（类似地），the signal source makes margin testing and characterization possible. Working with a signal source and an oscilloscope or logic analyzer，engineers can explore the limits of _____（其设计的性能）— introducing deliberate stresses with the source and measuring the results _____ the oscilloscope，or capturing data with the logic analyzer when digital errors occur.

In applications ranging _____ disk drive design to telecommunications conformance testing，the signal source and the acquisition instruments work together as _____（一个完整的测量方案）.

2. Translate the following passages into Chinese or English.

(1) The term "wave" can be defined as a pattern of varying quantitative values that repeats over some interval of time. Waves are common in nature：sound waves，brain waves，ocean waves，light waves，voltage waves，and many more. All are periodically repeating phenomena. Signal sources are usually concerned with producing electrical (typically voltage) waves that repeat in a controllable manner.

(2) Waveforms have many characteristics but their key properties pertain to amplitude，frequency，and phase. The amplitude，frequency，and phase characteristics of a waveform are the building blocks a signal source uses to optimize waveforms for almost any application.

(3) A multimeter is an instrument for measuring electricity (volts，amps，ohms) that is widely used and available in numerous shapes and sizes. An analog multimeter displays results by moving a pointer across a printed scale. The movement of the pointer helps to show gradual changes，but is less precise than its digital counterpart. A digital multimeter (DMM) displays the exact measurement on a numeric LCD or LED readout.

(4) An oscilloscope is a test instrument that displays electronic signals (waves and pulses) on a screen. It creates its own time base against which signals can be measured.

(5) Selecting the right probe for your application is key to attaining the best signal fidelity in your measurements. Active probes provide truer signal reproduction and fidelity for high-frequency measurements.

(6) 信号幅度变化不都是以 0V 参考地电位为中心的。电路地电位和信号幅度中心之间的电压之差称作"偏置电压"。

(7) "差分信号"是指那些使用两路互补通道的信号。两路互补通道分别以正、反极性传输同一个信号。

(8) "上升时间"和"下降时间"是指信号边沿进行状态转换所用的时间。在现代数字电路中，其数值通常是几纳秒。

(9) "带宽"是指通信线路、计算机总线等电气信道的传输能力。在数字线路中，"带宽"是

用“位速率”或“字节速率”度量的；在模拟线路中，“带宽”是指最高频率和最低频率之差，并以赫兹为单位。

（10）“噪声”是侵入电气传输中的无关信号。它可能来自附近电线的强电磁信号，也可能来自匹配欠佳的电气连接，还可能来自电力线尖峰信号。

3. Writing tasks.

Summary：Write an abstract for *Lesson* 29 *Oscilloscopes* within 150～200 words.

Composition：Write an essay about some electronic measurement instrument you often use.

Reading Materials

Passage 1 Understanding Waveforms

Waveforms come in many shapes and forms. Most electronic measurements use one or more of the following wave shapes, often with noise or distortion added.

Sine Waves Sine waves are perhaps the most recognizable wave shape. Most AC power sources produce sine waves. Household wall outlets deliver power in the form of sine waves. The sine wave is the result of a basic mathematical function-graphing a sine curve through 360 degrees will produce a definitive sine wave image. The damped sine wave is a special case in which a circuit oscillates from an impulse, and then winds down over time. Many applications such as frequency response tests call for a "swept" sine wave — one that increases in frequency over some period of time.

Square and Rectangular Waves Square and rectangular waves are basic forms that are at the heart of all digital electronics. A square wave is a voltage that switches between two fixed voltage levels at equal intervals. It is routinely used to test amplifiers, which should be able to reproduce the fast transitions between the two voltage levels. The square wave makes an ideal timekeeping clock for digital systems. A rectangular wave has switching characteristics similar to those of a square wave, except that its high and low time intervals are not of equal length.

Sawtooth and Triangle Waves Sawtooth and triangle waves look very much like the geometric shapes they are named for. The sawtooth ramps up slowly and evenly to a peak in each cycle, then falls off quickly. The triangle has more symmetrical rise and fall times. These waveforms are often used to control other voltages in systems such as analog oscilloscopes and televisions.

Step and Pulse Shapes A "step" is simply a waveform that shows a sudden change in voltage. The "pulse" is related to the rectangular wave. Like the rectangle, it is produced by switching up and then down, or down and then up, between two fixed voltage levels. Pulses are inherently binary and therefore are the basic tool for carrying information in digital systems. A pulse might represent one bit of information traveling through a computer. A collection of pulses traveling together creates a pulse train. A synchronized group of pulse trains makes up a digital pattern. Sometimes, circuit anomalies produce pulses spontaneously. Usually these transient signals occur non-periodically, and have come to be described with the term "glitch." One of the challenges of digital troubleshooting is distinguishing glitch pulses from valid but narrow data pulses.

In operational electronic systems, waveforms rarely look like the textbook examples explained above. Certain clock and carrier signals are pure, but most other waveforms will exhibit some unintended distortion or deliberate modulation. Some waveforms may even include elements of sines, squares, steps, and pulses.

Signal Modulation In modulated signals, amplitude, phase and/or frequency variations

embed lower-frequency information into a carrier signal of higher frequency. The resulting signals may convey anything from speech to data to video. The waveforms can be a challenge to reproduce unless the signal source is specifically equipped to do so.

Analog Modulation — Amplitude modulation (AM) and frequency modulation (FM) are most commonly used in broadcast communications. The modulating signal varies the carrier's amplitude and/or frequency. At the receiving end, demodulating circuits interpret the amplitude and/or frequency variations, and extract the content from the carrier. Phase modulation (PM) modulates the phase rather than the frequency of the carrier waveform to embed the content.

Digital Modulation — Digital modulation is based on two states which allow the signal to express binary data. In amplitude-shift keying (ASK), the digital modulating signal causes the output frequency to switch between two amplitudes; in frequency-shift keying (FSK), the carrier switches between two frequencies (its center frequency and an offset frequency); and in phase-shift keying (PSK), the carrier switches between two-phase settings. Pulse-width modulation (PWM) is yet another common digital format; it is often used in digital audio systems.

Quadrature Modulation — Today's digital wireless communications networks are built on a foundation of quadrature (IQ) modulation technology. Two carriers—an in-phase (I) waveform and a quadrature-phase (Q) waveform that is delayed by exactly 90 degrees relative to the "I" waveform—are modulated to produce four states of information. The two carriers are combined and transmitted over one channel, then separated and demodulated at the receiving end. The IQ format delivers far more information than other forms of analog and digital modulation: it increases the effective bandwidth of the system.

Digital Patterns and Formats A digital pattern consists of multiple synchronized pulse streams that make up "words" of data that may be 8, 12, 16, or more bits wide. One class of signal source, the digital pattern generator, specializes in delivering words of data to digital buses and processors via parallel outputs. The words in these patterns are transmitted in a steady march of cycles, with the activity for each bit in each cycle determined by the chosen signal format. The formats affect the width of the pulses within the cycles that compose the data streams.

Bit Streams Pseudo-random bit streams (PRBS) and pseudo-random word streams (PRWS) exist to make up for an innate limitation in digital computers: their inability to produce truly random numbers. Yet random events can have beneficial uses in digital systems. To create random noise, digital systems produce a stream of numbers that has the effect of randomness even though the numbers follow a predictable mathematical pattern. These "pseudo-random" numbers are actually a set of sequences repeated at a random rate. The result is a PRBS. A pseudo-random word stream defines how multiple PRBS streams are presented across the signal source's parallel outputs. PRWS is often used when testing serializers or multiplexers. These elements reassemble the PRWS signal into a serial stream of

pseudo-random bits.

Questions:

1. What wave shape do you think is the most recognizable one? Why?
2. What waveform is widely used in digital systems?
3. What does term *glitch* mean in this article?
4. Why is quadrature modulation widely used in telecommunication networks?
5. What does the acronym PRBS stand for?

Passage 2 Signal Integrity

According to Webster's, "integrity" can be definded as "the state of being unimpaired". In electronics, signal integrity means the signal is unimpaired with regard to functionality. For example, a system with good signal integrity has data that arrives early enough to guarantee setup timing requirements are met and clocks that make only one logic transition per physical transition.

Digital systems are tolerant of many signal-integrity effects including delay, mismatches and signal ringing. However, sufficiently large signal-integrity problems can cause systems to fail or, worse yet, work only intermittently.

One common signal-integrity problem is false clocking, where a clock crosses a logic threshold more than once on a transition, due to ringing on the line. If the ringing occurs on a data line, the system may exhibit data faults due to vialations of setup and hold times. In either case, the system behaves incorrectly producing intermittent "bad" data at its output. While it is impossible (and unnecessary) to eliminate all noise on a signal, signal integrity is required to preserve a system's functionality.

Signal integrity problems arise from the physical nature of the interconnecting wires. A wire is not the same as a connection line on a schematic. A real wire has resistance, capacities the ground and to otherwise, and inductance. At higher frequencies, the capacity and inductance cause the wire to act like a transmission line, and antena effects can cause crosstalk and electromagnetic interference (EMI).

Why aresignal integrity problems so much more prevalent today than in years past? The answer is speed. Transactions that once took microseconds are now measured in nanoseconds. To achieve this improvement, edge speeds too have accelerated. It's important to remember that the edge speed (rise time) of a digital signal can carry much higher frequency components than its repetition rate might imply. Circuit board traces just six inches long become transmission lines when driven with signals exhibiting edge rates below four to six nanoseconds, irrespective of the cycle rate. In addition, fast edge speeds require generally higher currents to produce them. Higher currents tend to cause ground bounce, especially on wide buses in which many signals switch at once. Moreover, higher current increases the amount of radiated magnetic energy and with it, crosstalk.

The most common signal integrity testing equipments include oscilloscope, time-domain

reflectometer (TDR), vector network analyzer, near-field probe. The basic things to investigate in signal integrity testing are as follows:

- **Transmission line and circuit impedance**, which must often be done on a test board with attached fixtures that have known S-parameters.
- **Channel losses, jitter, and distortion**, which can be done with direct signal measurements using test fixtures in a prototype or a test board.
- **EMI from high-speed channels**, which may be indicative of another problem in a circuit or with impedance matching in a high-speed channel.
- **Crosstalk between channels**, which can only reliably be measured on test boards with multiple lines running in parallel.

Before gathering measurements from test boards or your prototype, simulations should be used to extract important electrical parameters and performance data. This gives you a chance to generate benchmarks for a design before gathering measurements as well as identify simple design problems before producing a prototype. Some of the important points to simulate in your design and the corresponding measurements are summarizedin Table 10.1.

Table 10.1 Simulation and measurement for signal integrity problems

Electrical behavior	Simulation	Measurement
Reflections and attenuation	Not simulated directly, can be quantified by extracting impedance and propagation constant	TDR measurement
Losses	2-port S-parameters	Impulse
Bit errors, ISI, and jitter	Eye diagram	Eye diagram
Crosstalk	Multi-port S-parameters, direct time-domain simulation, or method of moments	Multi-port S-parameters or oscilloscope measurements
EMI (frequency domain)	Full-wave near-field or far-field simulations	Anechoic chamber measurements

Signal integrity is just one of the many alligators in the swamp called "design". For today's high-speed boards and ICs, signal-integrity performance is as important to specify digital functionality or clock speed. Signal integrity problems including delays, ringing, crosstalk and EMI can be identified through careful use of simulation. Once potential problems have been identified, layout can be constrained to improve performance and termination strategies can be applied to minimize the impact of signal integrity on the overall performance of a high-speed design.

By addressing signal integrity problem early, a high performance product can be produced with higher reliability and lower cost. It will get to market faster, and it will win market share that would otherwise be lost. And best of all, because you will not need to debug as many

prototypes, you will be able to move on to the next exciting design sooner than you thought possible.

Questions:

1. What does the techincal term *signal integrity* mean in electronics?
2. Is it possible to eliminate all noise on a signal? Why or why not?
3. What do you think is the main reason for a wire's behavior like a transmission line?
4. Why does the fast edge speed of a digital signal lead to signal integrity problems?
5. What does the acronym TDR stand for?

Passage 3 Virtual Instrumentation

What Is Virtual Instrumentation?

As shown in Figure 10.8, a virtual instrument (VI) consists of an industry-standard computer or workstation equipped with powerful application software, cost-effective hardware such as plug-in boards, and driver software, which together perform the functions of traditional instruments.

Figure 10.8 The virtual instrumentation system

Virtual instruments represent a fundamental shift from traditional hardware-centered instrumentation systems to software-centered systems that exploit the computing power, productivity, display, and connectivity capabilities of popular desktop computers and workstations.

Virtual Instruments versus Traditional Instruments

Stand-alone traditional instruments such as oscilloscopes and waveform generators are very powerful, expensive, and designed to perform one or more specific tasks defined by the vendor. However, the user generally cannot extend or customize them. The knobs and buttons on the instrument, the built-in circuitry, and the functions available to the user, are all specific to the nature of the instrument. In addition, special technology and costly components must be developed to build these instruments, making them very expensive and slow to adapt.

Virtual instruments, by virtue of being PC-based, inherently take advantage of the benefits from the latest technology incorporated into off-the-shelf PCs. These advances in technology and performance, which are quickly closing the gap between stand-alone instruments and PCs, include powerful processors and operating systems and technologies.

In addition to incorporating powerful features, these platforms also offer easy access to powerful tools such as the Internet. Traditional instruments also frequently lack portability, whereas virtual instruments running on notebooks automatically incorporate their portable nature.

Use of LabVIEW in Data Acquisition and Postprocessing of Signals

LabVIEW is a software package that provides control and a user interface for the data acquisition process. Figure 10.9 depicts a schematic of data flow in the data acquisition process. Note that the physical system maybe a mechanical system such as a beam subjected to stress, a chemical process such as a distillation column, a DC motor with both mechanical and electrical components, and so forth. The key issue here is that certain measurements are taken from the given physical system and are acquired and processed by the PC-based data acquisition system.

LabVIEW plays a pivotal role in the data acquisition process. Through the use of VIs, LabVIEW directs the real-time sampling of sensor data through the DAQ card and is capable of storing, processing, and displaying the collected data. In most cases, one or more sensors transmit analog readings to the DAQ card in the computer. This analog data is then converted to individual digital values by the DAQ card and is made available to LabVIEW, at which point it can be displayed to the user. Although LabVIEW is capable of some data analysis functions, it is often preferable to export the data to a format that is readable by MATLAB or spreadsheet programs.

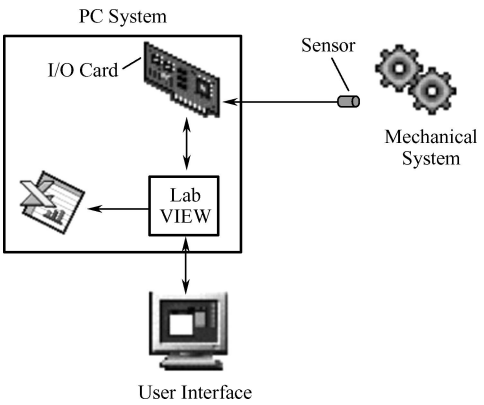


Figure 10.9 Schematic of the data acquisition process

A VI is a program created in theLabVIEW programming environment that simulates physical or hard instruments such as oscilloscopes and function generators. A simple VI used to produce a waveform is depicted in Figure 10.10. The front panel of the VI acts as the user

interface while the data acquisition process is performed by the combination of the PC and the DAQ card.

Much like the front panel of a real instrument, the front panel of the VI contains controls (i.e., knobs, switches, etc.) that allow the user to modify certain parameters during the experiment. These include a selector to choose the type of waveform, numerical controls to choose the frequency and amplitude of the generated waveform as well as its phase and amplitude offset. The front panel of a VI also typically contains indicators that display the acquired or generated data or other important information related to the experiment.

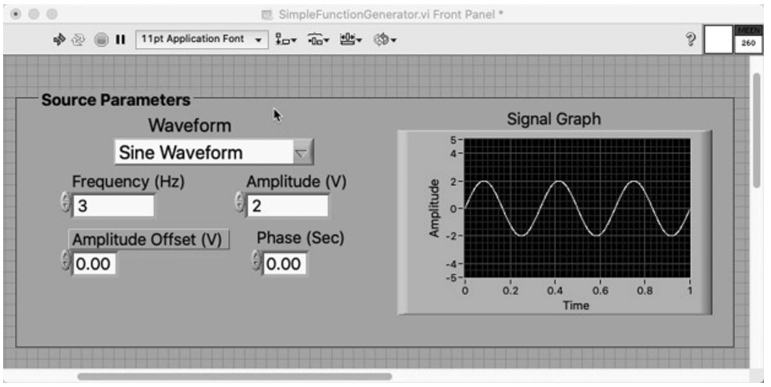


Figure 10.10 A simple function generator virtual instrument (VI)

Questions:

1. What components does a typical virtual instrument have?
2. How can virtual instruments be user-defined?
3. What role does LabVIEW play in the data acquisition process?
4. Is the front panel of a LabVIEW VI different from that of a real instrument?
5. What does the acronym DAQ mean in this article?

课文参考译文

第 1 课　VLSI(超大规模集成)技术

晶体管(transisitor)是电子技术发展史上的关键发明之一。它也是人类历史上最重要的发明之一。1947 年,贝尔实验室(Bell Labs)发明了晶体管。简言之,晶体管是一种依据输入电压传导可变电流的器件。换言之,晶体管是一种数字开关。但是,晶体管不同于真空管(vacuum tube),它是固态(solid state)的。“固态”是指晶体管在切换状态时物理形式保持不变。在晶体管中,不存在可移动部件。

和真空管相比,晶体管具备巨大优势——尺寸小、速度快、成本低,性能更可靠、耗能也更少。晶体管的发明带来了计算机工业的蓬勃发展。

最早的晶体管是分立元件(discrete device)。和其他分立元件一样,晶体管可以单独置于电路之中。如今,某些专用晶体管还是这样使用的。集成电路(integrated circuit)的发明引发了现代处理器的诞生。集成电路就是一组用单片材料制造的、内部互连的(无外部连线的)晶体管。集成电路也称作 IC 或芯片。

制造集成电路,要使用一种特殊材料。多数材料不是对电流绝缘(如空气、玻璃、木材),就是很容易传导电流(如金属、水溶液)。但某些材料只传导少量电流,或只在特定条件下传导电流,这些材料称作半导体(semiconductor)。显然,硅(silicon)是最常用的半导体材料。

通过精密的化学合成和设计、运用多种材料处理技术,可以在硅层上直接构造极小的晶体管。这种晶体管的体积小、速度快、功能可靠,耗能也少。1958 年,Texas Instruments 公司发明了第一片集成电路,它由 6 只晶体管在单一半导体表面构成。

发明集成电路后,人们很快就意识到了“缩微化”(miniaturization)和将大量晶体管集成到单片集成电路中的巨大好处。要实现更复杂的功能,就需要更多的晶体管(数字开关)。在提高硬件速度、控制功耗的同时,缩微化是集成大量晶体管的关键。

“大规模集成”(LSI)是指将先前多个分立元件构成的电路集成化。LSI 器件通常包含几百只晶体管。早期的计算机就是在电路板上将许多这种较小的集成电路互连而成的。

发明 LSI 电路后,集成技术日新月异。芯片越来越小、运算速度越来越快,价格也更加便宜。在既有成果的基础上,工程师们掌握了将更多逻辑集成到单片集成电路中的技术,这就是“超大规模集成”(VLSI)技术。VLSI 电路可以包含数百万只晶体管。

早期处理器的功能是由多个逻辑芯片实现的。Intel 公司率先将它们集成到单个芯片中。这就是最早的微处理器——Intel 公司在 1971 年推出的 4004(见图 1.1)①。今天的高级处理器都是这个 4 位中央处理器(CPU)的后代。

---

① 第 1～30 课中的图皆略。

第2课　存　储　器

制造存储器，可以采用机械技术、磁技术、光技术、生物技术和电子技术。磁存储器有磁带、软盘、硬盘和铁电随机存储器(ferroelectric RAM)等。光存储器件有只读光盘(CD-ROM)、可写光盘等。电子存储器广泛用于计算机设备中——因为它是当前速度最快的存储器。今天的电子存储器既可能采用独立的集成电路(IC)形式、模块形式，又可能以宏功能(存储单元)的形式作为集成电路的一部分。表1.1概括了几类电子存储器。

表1.1　电子存储器概览

类型	特点	可写	非易失	速度	成本
触发器	一位寄存器。 用作数字电路基本构件	是	否	超快	很高
寄存器	一组触发器，保存一字节、字或长字。 用于CPU等复杂芯片中	是	否	超快	很高
静态随机存取存储器	可寻址触发器阵列。 用来临时存储数据或用作高速缓存	是	否	很快	高
动态随机存取存储器	可寻址存储单元阵列。 用作计算机的主要数据存储器	是	否	快	中
只读存储器	可寻址硬连线存储单元阵列。 进行芯片制造时完成编程	否	是	很快	低
电可擦除可编程只读存储器	电可擦除可编程只读存储器。 写周期数是有限的	是	是	慢	高

触发器

触发器(flip-flop)其实就是一种双态电路，存储着一个“0”状态或者一个“1”状态。因为结构简单，所以触发器速度极快。在数字电路和集成电路中，触发器用作最基本的元件。当去掉电源电压后，触发器的状态就失去了。因此，触发器是“易失的”(volatile)。

寄存器

寄存器(register)是一组并行的触发器。寄存器的典型宽度为8位、16位、32位或64位。寄存器常用于保存数据、地址指针等。和触发器一样，寄存器也是“易失的”，而且速度很快。

静态随机存取存储器(SRAM)

SRAM是可寻址的触发器阵列。该阵列可配置成1位、4位、8位等数据格式。和触发器一样，SRAM结构简单，存取速度快，具有“易失的”特点。在微控器电路板上(CPU芯片内部或外部)可以发现SRAM，因为在这些应用中所需的存储量不大，不值得为了使用DRAM而去构建额外的接口电

路。此外，因其存取速度快，SRAM 也常用作“高速缓存”(cache)。

SRAM 的速度级别很多——从几纳秒(ns，高速缓存应用)到 200 纳秒(低功率应用)。SRAM 既有双极性技术，也有金属氧化物半导体(MOS)技术。互补金属氧化物半导体(CMOS)技术密度最高、功耗最低。高速缓存可以使用 BiCMOS 技术构建：BiCMOS 是一种混合技术，它使用双极性晶体管作为附加驱动。速度最快的 SRAM 采用“射极耦合逻辑”(ECL)双极性技术，因为该技术功耗大，所以存储器容量有限。

“内容寻址存储器”(CAM)是一类特殊的 SRAM。CAM 由触发器阵列构成，阵列中每一行都和一个数据比较器相连。存储器寻址不是向其提供地址，而是向其提供数据。所有的数据比较器将同时检查其对应的 RAM 寄存器是否保存着相同的数据。CAM 输出该数据对应的行地址。这项技术的主要应用是实现“快速查找表”(fast lookup tables)。在网络路由器中，经常使用快速查找表。

动态随机存取存储器(DRAM)

“动态”(dynamic)一词是指“数据不是保存在触发器中，而是保存在存储单元中”。因为存在泄漏，所以存储单元中的数据必须定期更新(读出重写)。更新时间间隔通常为 4～64ms。存储单元只需要 1 个电容器和 1 只晶体管，而在阵列中连接的触发器则需要 6 只晶体管。现代 DRAM 都采用“沟道电容器存储技术”—— 为了最小化芯片空间，将晶体管构建于电容器之上。因此，DRAM 的比特(位)成本要低于 SRAM。当所需存储量很大时，“数据更新需要额外电路”的缺点很容易就被较低的比特成本弥补了。

和 SRAM 存储器一样，DRAM 存储器也是由存储单元构成的。不过，SRAM 和 DRAM 还存在一个重大差别，那就是寻址方法不同。对于 SRAM，需要为存储器提供地址，存储器芯片输出存储单元中的数据；或者在输入端接收数据，并将其写入存储单元中。对于 DRAM，这种简单的存取方法是不可行的。由于动态特性，读出一行数据而不再次将其写入将破坏该行内的所有数据。

只读存储器(ROM)

ROM 也叫“掩模 ROM”或者“掩模编程 ROM”。这是因为在制造 ROM 时，要通过对存储单元“置 0”或“置 1”来进行编程。在通常情况下，“0”“1”就是铝线的“有”“无”。在芯片制造的最后一道工序中，铝层图案是由光刻掩模决定的。因此，这类器件常被称作“掩模 ROM”。

ROM 的优点是量产价格最低。在某些应用中，ROM 还有另一个优点：芯片一旦制成，数据就无法改动了，也就无须进一步编程、测试了。但是，若数据或代码必须改动，那么这就是个不小的麻烦。剩余的芯片只好丢进垃圾桶，还得重新制造芯片。

电可擦除可编程只读存储器(EEPROM)

EEPROM 能够像“可擦除可编程只读存储器”(EPROM)一样可编程，但采用的是电擦除方法。因此，不需要紫外线信号源。EEPROM 支持按字节擦除。

第 3 课　微 处 理 器

微处理器是一种在单芯片上构造的完整计算引擎。第一款微处理器是于 1971 年推出的 Intel

4004。4004 功能不强——它只能实现加减法，一次只能处理 4 位数据。但“全部功能集于单个芯片”是 4004 引人注目之处。在 4004 出现之前，工程师要用一组芯片，或一组分立元件构建计算机。4004 是第一款便携式电子计算器的计算引擎。

第一款用于家用计算机的微处理器是 Intel 8080，它是于 1974 年推出的单片 8 位计算机。第一款真正获得商业成功的微处理器是于 1979 年推出的 Intel 8088，用于国际商用机器公司(IBM)的个人计算机(PC)中。PC 市场从 8088、80286、80386、80486 到奔腾系列、酷睿系列、至强系列。这些微处理器均由 Intel 公司制造，都是在 8088 基础上改进的。一个酷睿 i9 处理器可含 8 个内核，每个内核都能够执行任何一段曾在 8008 上运行的代码，运行速度却要快约 6700 倍。表 1.2 展示了多年来(1974—2004 年)由 Intel 公司推出的各种处理器之间的差别。

表 1.2　Intel x86 微处理器(1974—2004 年)

名称	年份	晶体管	微米	时钟速度	数据宽度	MIPS
8080	1974	6,000	6	2 MHz	8 位	0.64
8088	1979	29,000	3	5 MHz	16 位/8 位总线	0.33
80286	1982	134,000	1.5	6 MHz	16 位	1
80386	1985	275,000	1.5	16 MHz	32 位	5
80486	1989	1,200,000	1	25 MHz	32 位	20
Pentium	1993	3,100,000	0.8	60 MHz	32 位/64 位总线	100
Pentium Ⅱ	1997	7,500,000	0.35	233 MHz	32 位/64 位总线	～300
Pentium Ⅲ	1999	9,500,000	0.25	450 MHz	32 位/64 位总线	～510
Pentium 4	2000	42,000,000	0.18	1.5 GHz	32 位/64 位总线	～1,700
Pentium 4 “Prescott”	2004	125,000,000	0.09	3.6 GHz	32 位/64 位总线	～7,000

从这张表中，我们可以看出：时钟速度与 MIPS(百万条指令/秒)之间存在着某种联系。最高时钟速度是由生产工艺和片内延迟决定的。片内晶体管数量与 MIPS 之间也存在着某种联系。例如，时钟频率为5 MHz 的 8088 的运行速率只有 0.033 MIPS(大约每 15 个时钟周期执行一条指令)，而现代处理器的运行速度通常为每个时钟周期执行两条指令。运行速度的提高和片内的晶体管数量有直接关系。

微处理器的内部结构

微处理器执行一组机器指令。这些指令告诉微处理器去做什么。利用这些指令，微处理器可以完成以下三项基本任务：

- 微处理器使用“算术逻辑单元”(ALU)来完成加、减、乘、除等数学运算。现代微处理器包含完整的浮点处理器，可以对大量浮点数据进行极其复杂的运算。
- 微处理器可将数据从存储器的一个位置搬移到另一个位置。
- 微处理器可以做出判断，并根据这些判断跳转到新的一组指令。

微处理器可以完成非常复杂的任务，但上述三项任务是最基本的。图 1.2 展示了一个能够完成上述三项任务的简易微处理器。

该微处理器有一套地址总线(向存储器发送地址)、一套数据总线(向存储器发送数据或者接收存储器数据)、一条读信号线 RD、一条写信号线 WR(用于通知存储器是从寻址地址读取数据还是写入数据)、一条时钟信号线(为处理器安排时序的时钟脉冲)和一条复位信号线(将程序计数器置零并重新开始执行)。这里假定数据总线和地址总线的宽度都是 8 位。

这个简易微处理器由如下组件构成。

- **寄存器 A、寄存器 B 和寄存器 C:** 是由触发器构成的简易锁存器。
- **地址锁存器:** 和寄存器 A、B、C 一样。
- **程序计数器:** 一种具备“加 1”功能和“置零”功能的锁存器。
- **算术逻辑单元:** 既可以简单到只是一个 8 位加法器,也可以是能够完成 8 位加、减、乘、除运算的单元(此处,我们假定为后者)。
- **测试寄存器:** 一种保存 ALU 比较结果的专用锁存器。通常,ALU 能够将两个数进行比较,并判断出二者是否相等或者哪个更大。测试寄存器也可以保存加法运算最后一步的进位位。这些数值保存在触发器当中,指令译码器利用这些数值做出判断。
- **三态缓存器:**就是图中标有“3-state”的 6 个方框。它可以传送逻辑 1,逻辑 0,或者和输出断开。三态缓存器允许在一条信号线上连接多个输出信号,但只有一个信号输出。
- **指令寄存器、指令译码器:**负责控制所有其他组件。

尽管图中没有显示出来,但需要从指令译码器引出完成如下功能的控制信号线:

1. 通知寄存器 A 锁定出现在数据总线上的数值。
2. 通知寄存器 B 锁定出现在数据总线上的数值。
3. 通知寄存器 C 锁定出现在数据总线上的数值。
4. 通知程序计数器锁定出现在数据总线上的数值。
5. 通知地址寄存器锁定出现在数据总线上的数值。
6. 通知指令寄存器锁定出现在数据总线上的数值。
7. 通知程序计数器加 1。
8. 通知程序计数器复位置零。
9. 激活 6 个三态缓存器之一(6 根独立的信号线)。
10. 通知 ALU 需要完成的操作。
11. 通知测试寄存器锁定 ALU 的测试位。
12. 激活 RD 信号线。
13. 激活 WR 信号线。

指令译码器的数据位不仅来自指令寄存器,而且来自测试寄存器和时钟信号线。

随机存取存储器(RAM)和只读存储器(ROM)

数据总线、地址总线、读/写信号线都连接到 ROM 或者 RAM 上(通常二者都有)。在这个微处理器例子中,有一套 8 位地址总线和一套 8 位数据总线。这意味着微处理器可寻址 256 字节的存储器,一次可以读/写 8 位数据。假定该微处理器有 128 字节(地址从 0 开始)的 ROM 和 128 字节(地址从 128 开始)的 RAM。

ROM 是“只读存储器”。ROM 芯片是用一组预设字节编程得到的。地址总线告知 ROM 芯片要将哪个字节取出并置于数据总线上。当 RD 信号线改变状态时,ROM 芯片将选中的字节输出到数据总线上。

RAM 是“随机存取存储器”。RAM 中包含着以字节为单位的信息,微处理器能够依据

RD/WR 信号哪个有效来决定字节的读/写。当前 RAM 芯片的一个问题是：掉电后，所有保存在 RAM 上的内容全部丢失。这就是计算机需要 ROM 的原因。

顺便提一下，几乎所有计算机都有一定数量的 ROM(可以建造一种简单的不含 RAM 的计算机——许多微控器在片内集成了一定数量的 RAM——但是一般不可能建造出一种不含 ROM 的计算机)。在 PC 中，ROM 被称作“基本输入/输出系统”(BIOS)。当启动后，微处理器就执行在 BIOS 中找到的指令。这些 BIOS 指令完成对机内硬件的测试，然后从硬盘中读取引导扇区。引导扇区也是一个小程序，BIOS 将其从硬盘中读出之后，这个小程序就存储在 RAM 中。之后，微处理器开始从 RAM 执行引导扇区的指令。这个程序将告知微处理器从硬盘其他位置读取信息到 RAM 中，然后微处理器执行相应的指令等。这就是微处理器装载和执行整个操作系统的过程。

微处理器指令

这个简易微处理器的指令集也不小。指令集是以“比特组合”方式实现的；当装载到指令寄存器的时候，每条指令都有各自的含义。由于人们不善于记忆“比特组合”，因此定义了一组短字来代表不同的比特组合。短字的集合就叫作“处理器汇编语言”。汇编器可以很容易地将这些短字翻译成相应的比特组合，汇编器的输出被放置到存储器中以便微处理器执行。假如使用 C 语言进行编程，那么编译器会将 C 代码翻译为汇编语言。

这些指令在 ROM 中是什么样的呢？每一条汇编语言指令都用一个二进制数代表，这个二进制数称作“操作码”。指令译码器需要将操作码转换为驱动微处理器内部各组件的一组信号。以 ADD 指令为例，我们看一下具体要做什么。

1. 在第一个时钟周期内，需要装载指令。指令译码器需要进行以下动作：
   1) 激活程序计数器的三态缓存器。
   2) 激活 RD 信号线。
   3) 激活三态缓存器中的数据。
   4) 将 ADD 指令锁存至指令寄存器。
2. 在第二个时钟周期内，对 ADD 指令进行译码。微处理器仅需要完成：
   1) 设置操作使 ALU 完成加法。
   2) 将 ALU 运算结果锁存到 C 寄存器。
3. 在第三个时钟周期内，程序计数器计数加 1(在理论上，该操作可以并入第二个时钟周期)。

每条指令都可以这样分解为一组前后相继的操作，使微处理器的各个组件按照正确顺序工作。有些指令(如 ADD)可能需要 2～3 个时钟周期，有些指令可能需要 5～6 个时钟周期。

微处理器性能

晶体管数量对于微处理器的性能有很大影响。如前所述，像 8088 这样的处理器执行一条典型指令需要 15 个时钟周期。由于设计乘法器的缘故，在 8088 上完成一次 16 位乘法需要大约 80 个时钟周期。晶体管越多，具备单周期乘法能力的乘法器就会越多。

有了更多的晶体管，就可以使用流水线技术。在流水线结构中，指令的执行是重叠的。尽管执行每条指令可能需要 5 个周期，但可以在不同阶段同时执行 5 条指令。这样看上去好像每个周期都能完成一条指令。

许多现代处理器有多个指令译码器，而每个指令译码器都有各自的流水线。这样，就可以实现“多指令流”——在一个周期内完成多条指令。“多指令流”技术实现起来相当复杂，需要使用大量晶体管。

处理器的设计趋势已经是全 32 位 ALU、内置快速浮点处理器和多指令流流水线。还有一个趋势是采用特殊指令，以便高效执行某些特定操作。此外，在处理器芯片上附加“硬件虚拟存储器”和“L1 高速缓存”也是微处理器的发展趋势之一。所有上述趋势都需要增加晶体管才能够实现。因此，今天已经出现了集成度高达几百万个晶体管的微处理器，可以在一秒内执行约 10 亿条指令。

64 位 ALU 是最新的处理器设计技术。64 位处理器是从 1992 年开始出现的，进入 21 世纪后已经成为主流。Intel 公司和 AMD 公司都推出了 64 位芯片。64 位处理器拥有 64 位 ALU、64 位寄存器和 64 位总线等。

第 4 课　运算放大器

1927 年，哈罗德·布莱克从位于纽约市的家出发，乘坐火车或渡船，去贝尔实验室上班。在上班途中，哈罗德可以放松下来，思考一些理论问题。哈罗德需要解决一个棘手问题——用于长途传输时，电话线需要放大器，而性能不可靠的放大器限制了电话业务的扩展。首先，放大器增益的容差性能不好。不过，通过使用调节器，这个问题很快就得到解决。其次，即使放大器在工厂里调节正确了，现场工作时的增益还是漂移得很厉害，以至音量太低或输入语音发生畸变。

为了制造出稳定的放大器，哈罗德已经进行了多次尝试。但是，温度变化和电话线上出现的供电电压极限使得增益漂移无法被控制。无源器件的漂移特性比有源器件好得多，假如能使放大器增益仅由无源器件决定，那么这个问题就会被迎刃而解。在乘渡船上下班的途中，哈罗德构思出了一个新颖的、解决放大器问题的办法，并在途中将它记录下来。

哈罗德的解决方案是：首先，设计一个增益比实际需求高的放大器；然后，将放大器的一部分输出信号反馈到输入端，使电路增益(这里的电路是由放大器和反馈器件组成的)由反馈电路决定，而不是由放大器增益决定。这样，电路增益就取决于无源反馈器件，而不是有源放大器。这个方案被称为“负反馈”(见图 2.1)，它是现代运算放大器的基本工作原理。哈罗德在乘坐渡船的途中记录了第一个有意加入反馈的电路。此前，一定有人无意中使用过反馈电路，但设计者却忽视了这种效应！

管理者和放大器设计者可能会发出痛苦的抱怨。他们也许会说：“获得 30 kHz 的‘增益带宽积’(GBW)就够难的了，现在这个傻瓜却要我设计 GBW 为 3 MHz 的放大器，而他还是想得到一个 GBW 为 30 kHz 的电路。”然而，时间已经证明哈罗德是正确的。不过，有一个小问题，哈罗德没有详细讨论，那就是振荡问题。在环路闭合的时候，开环增益很大的电路有时会发生振荡。很多人都研究过这种不稳定现象，不过，直到 20 世纪 40 年代，人们对它才有了清晰的认识。可是，解决稳定性问题需要长时间单调、复杂的计算。好多年过去了，没有人能将解法简化或者使之易于理解。

1945 年，H.W. 伯德提出了一种用于分析反馈系统稳定性的图形化方法(见图 2.2)。在这之前，反馈分析是用乘法和除法完成的。因此，计算传递函数是一项耗时、费力的工作。要知道，20 世纪 70 年代，工程师们才有了计算器和计算机。伯德采用了一种对数方法——将分析反馈系统稳定性的数学过程转换为既简单又好理解的图形化分析。反馈系统设计依然很复杂，却再也不是一项只为少数电气工程师所掌握的技术了。任何电气工程师都能用“伯德法”去判定反馈电路的稳定性，反馈技术也越来越多地应用于机器设计中。而对电子反馈设计真正迫切的需求是在计算机和传感器成为成熟技术后才产生的。

第一台实时计算机是模拟计算机。这台计算机使用预先编程的数学公式和输入数据来计算控制动作。编程是对一系列电路进行“硬连线”，这些电路对输入数据进行数学操作。“硬连线”的局限性导致模拟计算机无法普及。运算放大器是模拟计算机的核心部件，配置好的运算放大器可以对输入数据进行各种数学运算(如乘法、加法、减法、除法、积分和微分)。随着人们逐步了解、喜欢运算放大器，其名称就简化为大家熟知的“运放”。运放采用了具有很大开环增益的放大器；当电路形成闭合环路时，放大器就会执行由外部无源器件控制的数学运算。这种放大器是由真空管制作的，而且需要大电压电源供电，所以体积很大。不过，人们还是容忍了运放的大体积和大电压——因为它是模拟计算机的核心部件。早期的运放多是为模拟计算机设计的，但人们很快就发现运放还有其他用途。此后，在物理实验室里运放就很常见了。

那时，在大学和大公司的实验室里，能够看到通用模拟计算机——因为计算机对研究工作是至关重要的。同时，在实验中也需要对传感器信号进行调理。在信号调理方面，运放也找到了用武之地。随着信号调理应用范围的拓展，其对运放需求的增长超过了模拟计算机。在模拟计算机落伍于数字计算机之后，运放并未受到冷落，因为它在通用模拟应用中具有重要性。最终，数字计算机替代了模拟计算机。而随着测量应用的增加，运放的需求量也增长了。

在晶体管出现前，第一个用于信号调理的运放是用真空管构建的，因此它的体积很大。20 世纪 50 年代，低电压工作的小型真空管使运放的体积缩小到一块砖的大小，因此运放模块有了个绰号“砖块”。真空管和元器件体积不断缩小，直至运放缩至一只八引脚真空管的大小。在 20 世纪 60 年代，晶体管实现了商业开发，这进一步将运放的体积减至几立方英寸。早期的运放大多是为特定应用制造的，所以它们不一定通用；早期的运放是为某种特定应用服务的，而每个生产商制定的技术指标和封装都各不相同。所以，早期的运放几乎都不存在另外的货源。

在 20 世纪 50 年代末、60 年代初，集成电路开发出来了。但直到 20 世纪 60 年代中期，飞兆半导体公司才发布了 μA709。μA709 是首片取得商业成功的集成运放。虽然 μA709 存在自身的问题，但任何一位称职的模拟工程师都会使用它，在多种不同模拟应用中都可以用到它。μA709 的主要缺陷在于稳定性。它需要外部补偿，需要称职的模拟工程师去使用它。μA709 也很敏感，一旦不利条件出现，它就容易自毁。在 μA709 之后，出现了内部补偿的 μA741。当在数据手册要求的条件下工作时，μA741 是不需要外部补偿的。从那以后，新型运放就逐年不断出现。如今，运放的功能和可靠性已经提高到了“人人都能在模拟应用中使用”的程度。

今天，集成运放依然在用。最新一代运放能够覆盖 5 kHz～1 GHz GBW 的频谱范围。供电电压范围从可靠运行的 0.9 V 到最大绝对指标的 1000 V。输入电流和输入偏置电压低到用户在验证指标时遇到难题。运放已经成为名副其实的“通用”模拟集成电路，因为它可以完成所有模拟任务——线路驱动器、比较器(1 位模数转换器)、放大器、电平转换器、振荡器、滤波器、信号调理器、执行部件驱动器、电流源和电压源等。而设计人员所面临的则是如何迅速选定恰当的电路(运放的组合)，以及如何计算出无源元件值来产生期望的传递函数。

运放太重要了，它将继续作为模拟设计的关键部件。每一代新型电子设备都在硅片上集成了更多的功能，并将更多的模拟电路置于集成电路之中。随着数字应用的增加，模拟应用也会增加——因为大量的数据应用和接口应用是在现实世界中，而现实世界是一个模拟的世界。因此，每一代新型电子设备都需要新型模拟电路，也就需要新一代运放去满足相应的需求。在未来，模拟设计和运放设计依然是一项必备技能。

第 5 课　低通滤波器

一阶低通滤波器

从数学角度看，一个积分器[见图 2.3(a)]就是最简单的滤波器。它是构成大多数现代滤波器的基本模块。我们怎么从直观上理解积分器呢？假设在输入端加一个直流信号(即频率为 0)，那么在输出端将会出现一个线性斜坡信号——其幅度不断增大，直到受到电源电压的限制为止。如果不考虑电源电压对输出信号的限制，积分器在零频率上的响应将是无穷大的，这意味着它在零频率上存在一个极点(在使传递函数为无穷大的频率上存在一个极点)。

我们还知道，积分器的增益随着频率的增加而减小，在高频上，它的输出电压实际上就是零。因为增益和频率成反比，所以在对数坐标系中就是一条斜率为-1 的直线[见图 2.3(b)]。

积分器的传递函数很容易推导出来：

$$V_{OUT}/V_{IN}=X_C/R=(1/sC)/R=\omega_0/s)。$$

其中，s 是复频率变量 $\sigma+j\omega$，ω_0 等于 $1/RC$。假如把 s 当作频率，这个公式就印证了我们有关“增益和频率成反比”的直观判断。

简单的 RC 低通滤波器[见图 2.4(a)]是一种比积分器稍复杂的滤波器。其传递函数为：

$$V_{OUT}/V_{IN}=(1/sC)/(R+1/sC)=1/(1+sCR)=\omega_0/(s+\omega_0)。$$

当 $s=0$ 时，函数简化为 1；当 s 趋近于无穷大时，函数简化为 0。因此，这是低通滤波器。当 $s=-\omega_0$ 时，分母为 0 且函数值为无穷大，这意味着在复平面上出现一个极点。图 2.4(b)画出了传递函数的“幅度-频率”关系，s 的实部(σ)指向读者，s 的正虚部($j\omega$)朝右。$-\omega_0$ 处的极点非常明显。为了突出函数的形状，幅度是用对数值显示的。在无穷大频率处，积分器和 RC 低通滤波器的频率响应都趋近于 0，也就是说，在 s 等于无穷大处存在一个零点，这个零点环绕着复平面。

s 变量复函数和电路的实际频率响应是如何联系起来的呢？在分析电路对交变信号的响应时，我们用 $j\omega L$ 来表示电感的阻抗，用 $1/j\omega C$ 来表示电容的阻抗。在使用拉普拉斯变换分析瞬态响应时，我们用 sL 和 $1/sC$ 分别表示电感的阻抗和电容的阻抗。这种相似性是显而易见的。实际上，交流分析中的 $j\omega$ 就是 s 的虚部；正如上面提到的那样，s 是由实部 σ 和虚部 $j\omega$ 组成的。

假如将上面任一公式中的 s 用 $j\omega$ 替代，就可以得到电路的角频率响应。在图 2.4(b)中，$\sigma=0$，$s=j\omega$ 且沿正 $j\omega$ 坐标轴 。因此，沿着该坐标轴的函数值就是滤波器的频率响应。我们沿着 $j\omega$ 坐标轴将函数切割，并且在沿正坐标轴的函数值处加粗线，从而突出了 RC 低通滤波器的频率响应曲线。因为频率是用对数表示的，所以曲线和人们熟悉的伯德图看上去不一样[见图 2.4(c)]。

复频率的虚部($j\omega$)有助于描述电路对交流信号的响应，而实部(σ)有助于描述电路的瞬态响应。从图 2.4(b)中，我们可以看出 RC 低通滤波器和积分器之间的区别。RC 低通滤波器的瞬态响应更加稳定，因为其极点位于复平面的左半部，即对于阶跃输入，滤波器的响应是衰减指数形式的；而积分器的响应则是无限大。对于 RC 低通滤波器而言，极点沿-σ 坐标轴离原点越远意味着 ω_0 越大，时间常数越短，瞬态响应越快。与此相反，极点离 $j\omega$ 坐标轴越近，瞬态响应越慢。

至此，我们已经阐述了几个简单电路的传递函数与其复平面上的极点、零点关系。从传递函数中，我们可以推导出电路的频率响应(继而可以得到伯德图)及其瞬态响应。在积分器和 RC 滤波器各自的传递函数中，分母中都只有一个 s，所以它们只有一个极点。也就是说，它们都属于一阶滤

波器。

从图 2.3(b)中,我们也可以看出:一阶滤波器的频率选择性不是很好。为了得到更接近需求的滤波器,我们必须增加滤波器的阶数。从现在开始,我们将使用 $f(s)$表示传递函数,不再使用 V_{OUT}/V_{IN}。

二阶低通滤波器

二阶滤波器的分母中出现了 s^2,所以在复平面就出现了两个极点。使用电感和电容的无源电路或者包含电阻、电容和放大器的有源电路都具有这样的响应。例如,考虑图 2.5(a)中的无源 RLC 滤波器。其传递函数具有如下形式:

$$f(s)=X_C/(R+X_L+X_C)=(1/sC)/(R+sL+1/sC)=1/(LCs^2+RCs+1)。$$

当定义 $\omega_0^2=1/LC$ 和 $Q=\omega_0 L/R$ 之后,该式变为:

$$f(s)=\omega_0{}^2/(s^2+s\omega_0/Q+\omega_0^2),$$

此处的 ω_0是滤波器的特征频率,而 Q 是品质因数(R 越小,Q 越大)。

极点出现在使分母为零的 s 值处,即当 $s^2+s\omega_0/Q+\omega_0^2=0$ 的时候。只要记得 $ax^2+bx+c=0$ 的根为

$$x=[-b\pm(b^2-4ac)^{0.5}]/2a,$$

我们就可以解出这个方程。

在这个方程中,$a=1$,$b=\omega_0/Q$,$c=\omega_0^2$,$b^2-4ac=\omega_0^2(1/Q^2-4)$。如果 $Q<0.5$,那么两个根都是实数且位于负实轴上。其电路行为如同两个一阶 RC 滤波器级联。这种情况的意义不太大,所以我们只考虑 $Q>0.5$ 的情况。这意味着$(b^2-4ac)<0$,而方程的根均为复数。

因此,两个根实部相同,均为$-b/2a=-\omega_0/2Q$;两个根的虚部互为相反数。计算根在复平面的位置,我们就会发现它们位于距原点 ω_0处[见图 2.5(b)]。

改变 ω_0会改变极点距离原点的位置。减小 Q 值会将两个极点靠近,而增大 Q 值会将极点沿着半圆弧拉开并趋近 $j\omega$ 轴。当 $Q=0.5$ 时,两极点在负实轴上的$-\omega_0$处汇合。在这种情况下,对应电路就等效于两个一阶滤波器级联。

现在,让我们考察一下二阶函数频率响应及其随 Q 变化的情况。图 2.6(a)展示了在由复平面和垂直幅度向量构成的三维空间中绘制的函数曲线面。图中 $Q=0.707$,可以立即看出这是一个低通滤波器的响应。

增大 Q 值的效果是将极点沿着圆弧路径移向 $j\omega$ 轴。图 2.6(b)展示的是 $Q=2$ 的情形。由于极点靠近 $j\omega$ 轴,其对频率响应的影响就更大,这导致在通带高频端出现了一个峰值。

此外,对滤波器的瞬态响应也有影响。因为两个极点的负实部更小了,所以输入阶跃函数就会引起滤波器输出振铃振荡。Q 值越小,振铃振荡就越小,因为阻尼更大。另一方面,假如 Q 为无穷大,极点抵达 $j\omega$ 轴就会引起在$\omega=\omega_0$处的频率响应为无穷大(不稳定及持续振荡)。在图 2.5(a)的 RLC 电路中,这种情形是不可能出现的,除非 $R=0$。然而,对于包含放大器的滤波器而言,这种情形是有可能出现的。在设计过程中,必须考虑到这一点。

二阶滤波器提供了 ω_0和 Q 这两个变量,这就允许我们将极点放置在复平面上任何需要的地方。然而,这两个极点必须成共轭对。两者实部相同,虚部符号相反。极点放置的灵活性是一种强大的手段,它使二阶滤波器成为众多开关电容滤波器中的有用部件之一。随着频率趋近于无穷大,二阶低通滤波器传递函数和一阶低通滤波器传递函数一样趋近于零。然而,二阶低通滤波器传递函数的下降速度比一阶低通滤波器快了 2 倍,原因是分母中出现了 s^2因子。因此,在无穷大处就存在一个二重零点。

第 6 课　模数转换器

“模数转换器”(ADC)的类型像冰激凌口味一样多,挑选 ADC 要像挑选冰激凌一样细心。

闪式 ADC

闪式 ADC 是一种常见、易懂的 ADC(见图 2.7),它的转换速度很快,因此能够提供很高的采样率。但是,基本形式的闪式 ADC 的功耗很大。

闪式 ADC 将输入信号同时置于一列(2^N-1 个)比较器之前,比较器的参考电压由一组电阻设定,参考电压精确对应于转换器能代表的全部采样电压。比较器的输出(0 或 1)被编码成 N 位码字,以代表输入采样信号电平。这是最简单、直观和最快速的 ADC 实现方案。对于多数实际应用而言,当位数较多(如大于 14 位)时,所需的电阻数量(2^N-1)就太大了。与一些速度稍慢、特殊的模数转换方案相比,闪式 ADC 的功耗明显要高。

图 2.8 展示了典型闪式 ADC 的输出频谱。这张图是通过对 ADC 转换后的纯正弦样本进行“快速傅里叶变换”(FFT)得到的。可以看出:频谱不是简单地由输入正弦分量组成,还有一些分布在测量带宽内的其他成分。这些成分主要来自无法避免的“量化误差”(噪声)。这是因为转换器要使用有限个可用样值(由 ADC 位数决定)来代表模拟输入电平。

对于满量程正弦波,有效“信噪比”(SNR)是转换器精度的函数。当满足某些条件时,它可由如下公式决定:SNR=$(6.02N+1.76)$dB。其中,N 是转换器的位数。

产生图 2.8 中曲线的 ADC 具备 12 位转换精度,其理论 SNR 值为 74 dB。从该曲线中,我们可以看出:正弦分量电平和单个噪声分量电平之差接近 104 dB。这两个数值之间存在差异的原因是 SNR 公式使用了全部噪声,其中包括构成 FFT 的全部噪声分量之和。指出这一点的原因是:通过提高采样率,可以提高转换器的有效精度(见下文)。

过采样实现“处理增益”

在某个给定的音频应用中,假定我们需要转换过程中的信噪比至少为 70 dB。上述公式表明:需要最少 12 位的转换精度(前提是对于满幅正弦波输入和理想转换器而言),测量中包含($0\sim f_s/2$)Hz 带宽中的全部噪声。如果采用 8 倍带宽的采样率(见图 2.9),我们可以看到实际的音频信号仅占用了基带(即 $0\sim f_s/2$)的 1/4,而噪声在带宽内均匀分布。假如现在准备对采样信号进行数字滤波,我们可以去掉约 3/4 的噪声,从而将信噪比提高 4 dB 或 6 dB。这种有效 SNR 的提高称作“处理增益”,是通过对输入信号进行“过采样”(相对于 2 倍带宽原则而言)来实现的。

“处理增益”的简易计算公式可以很容易推导出来:处理增益(dB)=10log(采样率/2×信号带宽)。在该公式中,我们假定:(1)为了准确符合期望的输入信号带宽,我们使用数字滤波器对样本带宽进行限制;(2)噪声是均匀分布的。

这样的话,假如我们使用 128 倍过采样设计(常见于迷你盘录音设备和 PC 声卡),那么就可以获得 18 dB 的“信号/量化噪声比”的提高,从而将转换器的有效精度从 12 位提高至 15 位。顺便说一个例子,我们可以考虑使用该方法提高数字移动电话中数据转换器的性能。一个 GSM 蜂窝信道的带宽为 200 kHz。目前,我们可以得到采样率为 80 MSPS、精度为 14 位的高速模数转换器,这种

ADC 在整个 0～$f_s/2$ 范围内提供 75 dB 信噪比。其处理增益如下：处理增益（dB）＝10log(80000000/2×200000)＝26 dB。因此，GSM 采样信号具备很高的信噪比：75＋26＝101 dB。

Σ-Δ 转换器

“处理增益”的概念恰好可以引领我们进入 Σ-Δ 转换器这个话题。由于充分利用了“处理增益”的概念，Σ-Δ 转换器获得了高性能、低成本和低功耗的优势；Σ-Δ 转换器只需简单的模拟接口（需要很少的或者不需要抗混叠滤波器），所以非常适合于音频应用。图 2.10 展示的是基本 Σ-Δ 转换器框图。

这个转换器实质上是一个后跟数字滤波和抽取的高过采样 1 位 ADC（即比较器），数字滤波和抽取用来实现处理增益。由于增加了噪声整形电路，该转换器的有效性能获得了极大提高。噪声整形电路将原本在 0～$f_s/2$ 频带内均匀分布的噪声最大限度地从有用频带中去除（见图 2.11）。

对于典型的 128 倍过采样系统，仅“处理增益”一项就可将有效精度提高 3 位（相当于 4 位转换器），而“噪声整形”技术能将有效精度进一步大幅提高。对于某些用于音频的 Σ-Δ 转换器而言，现在已经可以达到 24 位精度了。现代转换器使用了比图 2.11 复杂得多的噪声整形处理方法。尽管图 2.11 只是简单的一阶 Σ-Δ 设计，但采用“处理增益”和“噪声整形”的基本原则却是一致的。

目前，在实际中使用的模数转换方法有几十种。常见的方法有逐次逼近法、多通道法、插值法、子区法和逐位处理法。每种方法都具备一些其他方法不具备的潜在优势。幸运的是，在选择 ADC 器件时，用户一般不需要理解其工作原理。只要用心研究一下数据手册上提供的性能指标，就可以做出最佳选择。

第 7 课　开 关 电 源

除电池供电的电子产品外，新型电子产品都需要将 115 V 或 230 V 交流电源转换成直流电源为电路供电，而电能转换效率正在成为业界及整个社会关注的重点。开关电源不仅提供了较高的转换效率，而且为设计者提供了更大的灵活性。半导体技术、磁器件和无源器件技术的发展使开关电源在当今功率转换的舞台上成为日益流行的选择。

线性电源与开关电源

在历史上，线性稳压器曾经是产生稳压输出电压的主要方法。通过对级联功率通过器件的导电性进行线性控制以响应负载变化，线性稳压器将输入高电压降为输出低电压。这种方法导致在负载电流流经的通过单元两端出现一个大电压。

这种功率损耗（$V_{drop}\cdot I_{load}$）使线性稳压电源的效率只有 30%～50%。这意味着每向负载输送 1 W 的功率，就会有至少 1 W 的功率以热能形式消耗了。对于那些功率超过 10 W 的小型应用而言，散热装置的成本就导致使用线性稳压电源不合算。

开关电源中的功率器件工作在“全开”和“截断”状态。这样，要么在大电流流经功率器件时，导通电压很低；要么在大电压时，没有电流通过器件。因此，电源内部消耗的功率就很少。开关电源的平均效率为 70%～90%，而且和输入电压无关。

集成度的提高带动了开关电源成本的下降，这使开关电源在输出功率超过 10 W 及多输出应用

中成为具有吸引力的选择。

前向转换器

图 3.1 中的降压转换器是最基本的前向转换器。其工作过程可视为两个不同的时间周期，分别出现在串联功率开关处于“接通”和“断开”的状态。当功率开关接通时，输入电压连接到电感的输入端。电感的输出电压就是转换器的输出电压，整流器处于反向偏置。在这个周期，由于在电感两端存在恒定电压源，所以电感电流开始按照如下公式线性增加：

$$i_{L(on)}=(V_{in}-V_{out})\cdot t_{on}/L$$

在“接通期”，能量以磁通的形式存储在电感的铁芯材料中。存储的能量足以满足负载在下一个“断开期”的需求。

下一个周期就是功率开关的“断开期”。当功率开关断开时，电感输入电压被捕获二极管钳位在地电位以下一个二极管电压降。电流开始流过捕获二极管以维持负载环路电流。存储在电感中的能量被移走。这段时间内的电感电流可表示为：

$$i_{L(off)}=(V_{out}-V_{D})\cdot t_{off}/L$$

当功率开关再次接通时，这个周期就结束了。

电压调节是通过改变功率开关的“接通-断开”的占空比来实现的。下面的关系式近似描述这个工作过程：$V_{out}\approx\partial\cdot V_{in}$，式中的 ∂ 是占空比，$\partial=t_{on}/(t_{on}+t_{off})$。

该转换器可以输出 kW 级功率，但却有一个严重缺点——假如功率开关非正常短路，输入电源就会直接连接到负载电路，这通常会造成灾难性后果。为了避免出现这种情况，在输出端跨置了一个“短路器”。短路器是一个闭塞的“可控硅整流器”。当输出进入“过压”状态，就会激发这个整流器。降压转换器仅用于板级电源管理。

增压转换器

最基本的增压转换器是升压稳压器，其原理图见图 3.2。它的工作过程也可以分为两个不同的周期，分别对应着功率开关的“导通”和“断开”状态。当功率开关导通时，输入电压源接到电感两端。电流从零开始按照如下公式增加：

$$i_{L(on)}=V_{in}\cdot t_{on}/L$$

能量被存储在铁芯材料中。

每个周期存储的能量乘以工作频率的积必须大于负载的功率需求，即满足下式：

$$P_{sto}=0.5L\cdot I_{pk}^{2}\cdot f_{op}>P_{out}$$

然后，功率开关断开，电感电压超过输入电压并被整流器钳位在输出电压上。电流开始线性下降，直至铁芯中的能量全部耗尽。如图 3.3 所示，其电流波形由下式决定：

$$i_{L(off)}=(V_{out}-V_{in})\cdot t_{off}/L$$

升压变换器也仅用于板级电源管理。

第 8 课　时钟信号源

当今的时钟器件多种多样，下面阐述其中的几种。

晶体

晶体是一种基本的压电石英晶体。它本身是不能产生时钟信号的，必须和时钟振荡器连接在一起才能得到时钟波形。晶体有两种：串联谐振晶体(可视作高品质因数的串联 LC 电路)和并联谐振晶体(可视作高品质因数的并联 LC 电路)。串联谐振晶体在谐振频率点的阻抗最小，而并联谐振晶体在谐振频率点的阻抗最大。

晶体振荡器

晶体振荡器是一种用晶体反馈元件组成的振荡器，其他振荡器采用有源器件或无源器件作为反馈元件。晶体振荡器的输出频率最为精确和稳定。晶体振荡器是多数高速数字系统时钟源的首选。

补偿振荡器

随着温度和电压的变化，晶体振荡器的输出频率也会发生变化。在需要高稳定度时钟的应用中，人们通常使用补偿振荡器。补偿振荡器能够调整由电压和温度引起的频率变化。“温度补偿振荡器”(TCXO)包含了用于补偿温度变化的电路，从而防止频率的变化。恒温控制振荡器将晶体放置在一个温控恒温箱中，这样保持晶体工作在一个精确温度之下。双恒温箱振荡器有两个恒温箱，晶体在内层恒温箱中，而控制电路和内层恒温箱又包含在外层恒温箱中。双恒温箱振荡器比恒温控制振荡器的温度稳定性更好。显然，随着温度稳定性的提高，振荡器的成本也提高了。

压控振荡器

“压控振荡器”(VCO)的输出受输入电压引脚的控制。在整个频率范围内，控制电压和输出频率的关系是非线性的，但是在部分频率范围内是线性的。

频率合成器

通过使用一个或多个“锁相环”(PLL)，频率合成器从一个或多个参考时钟源产生一个或多个不同的输出频率。参考频率通常是由连接到合成器上的晶体产生的。设计频率合成器的目的是用以替代系统中的多个振荡器，从而减少电路板空间，降低系统成本。图 3.4 展示了锁相环的框图。

PLL 有两个输入，一个参考输入和一个反馈输入。PLL 采用两种方法校正频率。“频率校正”先对参考输入和反馈输入间的大频差进行校正。“频率校正”类似于“粗调”，当 VCO 的频率低于参考频率的一半或者高于参考频率的 2 倍时，要进行“频率校正”。当 VCO 的频率在参考频率的一半和 2 倍之间时，要进行“相位校正”，“相位校正”是“微调”。

“相位/频率检测器”检测参考输入和反馈输入之间的频率差与相位差，并依据反馈频率是超前还是滞后于参考频率分别产生用于补偿的 U_p 信号和 D_n 信号。然后，这些控制信号通过“电荷泵”和“环路滤波器”产生控制 VCO 的控制电压。振荡器的频率取决于控制电压 V_{ctrl} 的输入。VCO 的稳态频率为 $F_{vco}=(F_{ref}\cdot P)/Q$。PLL 的输出频率可以表示为 $F_{out}=(F_{ref}\cdot P)/(Q\cdot N)$。

频率合成器的采样率决定了为进行相位和频率校正而对输入信号采样的频率，其表达式为 F_{ref}/Q。

基于 PLL 的频率合成器的：(1)“采集/锁定时间”是频率合成器在加电后(或在可编程输出频

率发生改变之后)达到目标频率所用的时间;(2)“精度”是 P、Q 计数器的位数,精度决定了频率改变的增量;(3)“死区”是指无法被 PLL 校正的参考输入和反馈输入之间的最大相位差。

产生多个不相关频率的频率合成器需要使用多个 PLL。随着系统复杂性的提高和系统中多个时钟的使用,频率合成器应用得越来越普遍。“时钟信号产生器”和“频率合成器”这两个词可以互换使用。

时钟缓存器

时钟缓存器是一种输出波形直接跟随输入波形的器件。输入波形通过该器件并被输出缓存器重新驱动。因此,该类器件存在传输延迟。此外,由于在每个输入、输出通道间存在传输延迟的差别,因此输出端将出现“相位抖动”。

时钟参数

时钟抖动 “抖动”可以定义为“时钟输出状态转换位置偏离了其理想位置”。这种偏离既可能超前于理想位置,也可能滞后于理想位置。因此,抖动表示为 ±ns。抖动分为三类:周期间抖动、周期内抖动和长期抖动。

相位偏移 “相位偏移”是指应该同时到达的两路信号在抵达时间上的差异。相位偏移由两部分组成:驱动器件的输出相位偏移和布线导致的电路板设计相位偏移。

容差/精度 该参数度量的是器件工作频率和标称频率(一般是指在 25 ℃±5 ℃的常温下)的接近程度。例如,某个器件的标称输出频率为 25.000 MHz,其在常温下输出频率的“长期”平均值(由用户定义的)为 25.001 MHz,那么该器件的精度为 +40 ppm。“频率容差”会受到晶体生产和校准工艺中的“精度控制能力”的影响或控制。

稳定性 “稳定性”参数通常和晶体、振荡器相关。稳定性是指在常温范围内工作频率和常温标称频率之间的差别,表示为 ppm。该参数由频率偏差的最大值和最小值给出,表示为%或 ppm。

老化 “老化”定义为“内部晶体或振荡器随时间变化而引起的频率系统误差”。“老化”通常用 ppm/年来表示。若不单列,则老化归于“稳定性”指标中。该参数通常与晶体振荡器有关。新晶体的老化速度要高于旧晶体。

误差 在基于锁相环的设备中,无法在输出端总是获得指定频率,原因在于锁相环内部的 P、Q 计数器大小有限。例如,若指定频率为 25.000 MHz,但锁相环输出频率可能是 24.998 MHz,则误差为-80 ppm。误差计算公式为:误差 =(标称频率-目标频率)/目标频率。

摆率 电压或频率的变化速率称作“摆率”(slew)。摆率通常在数字信号的上升沿和下降沿进行测量。不过,在生产商的产品目录中,往往给出的是上升时间和下降时间,而不是摆率。随着低功率器件的出现,摆率也用来描述频率的变化速率。

占空比 “占空比”是输出信号高电平时间与整个周期时间之比。该参数用百分比表示。理想占空比为 50%,但多数时钟生产商指定的占空比为 40%~60%。在既使用时钟上升沿也使用时钟下降沿的系统中,占空比很重要。TTL 器件和 CMOS 器件均有占空比。对于 TTL 器件,电压在 0~3 V之间摆动,高电平时间在 1.5 V 测量。对于 CMOS 器件,电压在 0~V_{dd}之间摆动,高电平时间在 $V_{dd}/2$ 处测量。

第 9 课 互连部件

互连部件用于连接器件、电路板和模块。在设计阶段,就必须考虑到对互连部件的需求。互连

部件必须满足下列大多数或者全部性能要求：(1)信号完整性；(2)功率损耗；(3)电气特性(接触电阻、电感和电容，电压和功率额定值，屏蔽，滤波)；(4)机械特性(接触点数量、间隔和力、冲击和振动、尺寸、焊接技术)；(5)环境问题(清洗材料、热性能、防腐保护)。

“信号完整性”是指输入信号的质量损失。在理想情况下，互连部件输出端信号的所有特性和输入端信号完全相同。在实际中，输出端信号的质量会下降。用户必须明确可接受的质量下降程度，或者必须定义输出端信号的最小可接受电平。反映信号完整性的测试有：(1)“电压驻波比”(VSWR)；(2)频率响应；(3)信号边沿上升时间；(4)电流。

高速连接器问题

今天，时钟速率超过 100 MHz，上升时间通常为 1 ns 以下，设计者不能无视互连部件对逻辑设计的影响。对电路的时序特性和噪声特性，互连部件都有显著的影响。时钟速率与上升时间的加快增加了容性耦合效应和感性耦合效应，这使得串扰问题更加严重。这也意味着在数据定时、读取前，反射信号衰退的时间变短了，这就要求减小“未端接系统”的最大导线长度。这些将带来一个重大挑战：确保高速脉冲经过互连路径(从器件到印制电路板、通过印制电路板到底板并进而延伸到任何正在使用的网络连接)时的信号完整性。

由于反射的出现，在通过互连路径传送高速信号脉冲时，短促的上升时间就会破坏信号的完整性。上升时间越短，信号质量下降的风险越大。理想情况是：连接器拥有合适的端接特性，没有造成信号质量的下降。

串扰

除信号质量下降外，“串扰”的出现是又一个影响信号完整性的潜在因素。串扰是由接触点间的感性耦合和容性耦合引起的。这会将一条导线(或接触点)中的开关信号噪声感应到另一条导线(或接触点)。串扰可以表示为信号电压摆幅的百分比或者分贝数(0 dB 代表 100％的串扰，-20 dB 代表 10％的串扰，-40 dB 代表 1％的串扰)。

串扰分为“后向串扰”和“前向串扰”两类。“后向串扰”也叫“近端串扰”，它在连接器的驱动端进行测量，代表了容性耦合和感性耦合之和。“前向串扰”也叫“远端串扰”，它在连接器的接收端进行测量，代表了容性耦合和感性耦合之差。

串扰还可分为“单线串扰”和“多线串扰”。“单线串扰”是一条驱动线感应电压至一条静止线的结果。“多线串扰”是两条或两条以上驱动线感应电压至一条静止线。由于多数系统同时使用多条导线(多个接触点)，“多线串扰”就更为实用。

为了全面理解连接器的串扰指标，用户需要获取以下信息：(1)串扰指标是实际测量结果，还是仿真结果？(2)串扰测量时使用的信号“上升时间”是多少？是用何种数字技术产生的(TTL 和 CMOS“上升时间”的典型定义为幅度的 10％～90％，而 ECL“上升时间”的典型定义为幅度的 20％～80％)？(3)连接器上的信号模式是什么？例如，地线相对于可测量信号线的位置。(4)信号状态切换影响了多少条线？(5)源阻抗和负载阻抗是多少？(6)串扰指标是指“前向串扰”或“后向串扰”，还是全部串扰？(7)串扰指标是指“单线串扰”，还是“多线串扰”？(8)如果是“多线串扰”，有几条驱动线？

减小串扰的常用方法是在信号接触点间分布接地点。当然，这种方法的缺点是降低了信号接触点密度。

互连的层次

戈兰尼兹定义了互连的 6 个层次：(1)芯片焊盘到封装引线的连接(如导线接头)；(2)元器件

到电路板的连接(如 PLCC 插座);(3)电路板间的连接(如用于主板和子板连接的边缘卡);(4)部件互连(如扁平电缆);(5)部件和输入/输出的互连(如 BNC 连接器);(6)系统互连(如以太网连接器)。

每个层次的互连都可以有多种实现方式。例如,第 4 层电路板间互连可以采用边缘卡连接器、扁平电缆或导线。

第 10 课　移动电话系统(Ⅰ)

无线移动电话分为两大类:无绳电话和移动电话。无绳电话由基站和室内手持设备组成。无绳电话不支持网络传输。移动电话的发展经历了三个时期,各有不同的技术:(1)模拟语音;(2)数字语音;(3)数字语音和数据。

第一套移动系统是由美国 AT&T 公司设计的,美国"联邦通信委员会"(FCC)委托该公司将其应用于全美国。因此,美国各地使用了统一(模拟)系统——在加州购买的移动电话到了纽约州也可以使用。在欧洲,移动电话系统的情形与美国正好相反——欧洲各国都设计了自己的系统。这种做法导致了移动电话在欧洲的失败。

从这次失败中,欧洲汲取了教训。当数字移动通信技术到来时,政府运营的邮电公司联合推出了统一的系统标准(GSM);在欧洲境内任何地点,任何一部欧洲移动电话都能够使用。而当时,美国决定政府不应参与标准化事务,而将数字移动通信的发展交由市场决定。该决定导致不同的设备制造商生产出不同类型的移动电话。

尽管美国最早引领了移动通信技术的应用,但是现在欧洲移动电话的拥有率和使用率远远超过了美国。"拥有欧洲全境统一的系统"是原因之一,但还有其他原因。美国和欧洲另一个不同之处在于"电话号码"这件小事上。在美国,移动电话号码和固定电话号码是不分的。这样一来,用户看不出某个号码,如(212)-234-5678 是固定电话号码还是移动电话号码。为了防止人们对使用电话有顾虑,电话公司决定移动电话用户只为来电付费。这样做的结果是:由于担心仅接电话就会积累大额账单,许多人在购买移动电话时很犹豫。在欧洲,移动电话都有一个特殊的区位号(类似 800 号码和 900 号码),很容易识别移动电话号码。这样,"主叫付费"的惯例也可以应用于欧洲的移动电话(费用分开的国际电话除外)。

还有一个对采用移动电话产生巨大影响的因素是预付费移动电话在欧洲的广泛使用(在某些地区,高达 75%的移动电话是预付费的)。在许多商店里,人们都可以像买收音机一样地买到这种电话——付完钱就可直接拿走。你既可以预存话费(如 20 或 50 欧元),也可以在余额降为零时对其充值。因此,欧洲的儿童和许多小孩儿都拥有移动电话(通常是预付费电话);这样,孩子的父母就可以知道孩子在哪里,而孩子们也不会花掉大量话费。如果仅偶尔使用移动电话,那么使用移动电话实质上是免费的,因为不存在月租费和来电费。

改进型移动电话系统(IMTS)

在 20 世纪前叶,移动电话偶尔用于航海和军事通信。1946 年,第一套车载电话系统在圣路易斯架设。该系统位于一座高楼的顶部,使用一台发射机和一个同时用于接收、发送的信道。为了进行通话,用户必须按下一个按钮开启发射机、关闭接收机。20 世纪 50 年代末,有几个城市开始架设这种系统(称为"随按即说"系统)。民用无线电设备、出租车和警车经常使用这种技术。

20 世纪 60 年代，“改进型移动电话系统”(IMTS)架设。IMTS 也在山顶上使用大功率(200 W)发射机，不过现在拥有两个频率(一个用于发射，另一个用于接收)。由于移动电话的输入信号和输出信号占用不同的信道，故移动用户无法听到对方的声音(不同于出租车上的“按键—通话”系统)。IMTS 提供 150～450 MHz 之间的 23 个信道。由于信道数量很少，用户往往得等待很长时间才能听到拨号音。此外，位于山顶的发射机功率很大，临近系统必须相距几千米以避免互相干扰。总之，由于容量有限，IMTS 并不实用。

高级移动电话系统(AMPS)

随着“高级移动电话系统”(AMPS)于 1982 年在贝尔实验室发明及其在美国架设，一切都发生了改变。AMPS 也用于英国，并被称作 TACS；AMPS 在日本被称作 MCS-L1。为了实现向下兼容，AMPS 的许多基本特性都由其后来者 D-AMPS 继承了。

在移动电话系统中，地理区域分成一个个蜂窝小区。这就是“移动电话”也称作“蜂窝电话”的原因。在 AMPS 中，蜂窝小区的典型大小为 10～20 km；在数字系统中，蜂窝小区要小一些。每个小区使用一组特定的频率，这组频率在其临近的任何小区中都不能使用。使蜂窝系统具备很大容量的关键在于使用了较小的蜂窝小区和在附近(但不是临近的)小区中重复使用频率。在方圆 100 km内，IMTS 的每个频率上只能拥有一路通话；在同样区域内，AMPS 拥有 100 个 10 km 大小的蜂窝小区，而且每个频率拥有 10～15 路通话。这样，蜂窝设计将系统容量至少提高了一个数量级，蜂窝越小、提高量越大。此外，蜂窝越小意味着需要的功率越小，发射机和手持设备会更小、更便宜。手持电话输出功率为 0.6 W，车载发射机功率为 3 W，这是 FCC 允许的最大功率值。

图 4.1(a)展示了频率重用概念。通常，蜂窝小区都接近圆形，但用六边形进行建模更加容易。在图 4.1(a)中，蜂窝小区的大小都相同，每 7 个小区分成一组，每个字母代表一组频率。需要注意：对于每组频率，存在约两个小区宽的缓冲地带；为了实现更好的隔离以降低干扰，在这些缓冲地带，频率是不能重用的。

在用户数量达到导致系统过载程度的地区中，为了实现更多的频率重用，需要降低功率并将过载蜂窝小区分为更小的微蜂窝单元，见图 4.1(b)。在体育赛事、摇滚音乐会及其他在几个小时内有大量移动用户云集的场所，电话公司有时会使用卫星链接的便携发射塔创建临时微蜂窝。

基站位于蜂窝小区的中心，小区内的所有呼叫都传送到基站。基站由计算机和发射机/接收机(与天线相连)组成。在较小的系统中，所有基站都和一个称作“移动电话交换局”(MTSO)或“移动电话交换中心”(MSC)的设备相连。在更大的系统中，需要几个 MTSO。所有 MTSO 都和二级 MTSO 相连，……，以此类推。MTSO 实质上就是电话系统中的端局；实际上，它至少和一个电话系统端局相连。MTSO 使用分组交换网络和基站、其他 MTSO 及“公共交换电话网”(PSTN)进行通信。

第 11 课　移动电话系统(Ⅱ)

第一代移动电话是模拟的，第二代移动电话是数字的。第一代移动电话没有世界通用标准，第二代移动电话也没有世界通用标准。曾经使用的系统包括 D-AMPS、GSM、CDMA 和 PDC 这 4 个。PDC 只用于日本，它基本上就是为了与日本第一代模拟系统兼容而对 D-AMPS 做的修改。

D-AMPS(数字高级移动电话系统)

第二代 AMPS 是全数字的 D-AMPS。国际标准 IS-54 及后来的 IS-136 描述了D-AMPS。为了使同一蜂窝小区中的第一代、第二代移动电话都能够同时使用，对 D-AMPS 进行了精心设计，以便和 AMPS 共存。特别需要指出的是：在相同频率上，D-AMPS 采用了和 AMPS 相同的 30 kHz 信道，这样一来，某个模拟信道的附近可能是数字信道。

当 D-AMPS 开始服务时，为了解决载荷增加的问题，就需要使用一个新频带。上行信道在 1850～1910 MHz 范围内，相应的下行信道在 1930～1990 MHz 范围内；上行信道和下行信道是成对出现的，这一点和 AMPS 完全相同。该频段的波长是 16 cm，标准 1/4 波长天线仅为 4 cm 长，这样一来，电话尺寸就减小了。然而，为了获得更大范围的可用信道，许多 D-AMPS 电话机既使用 850 MHz 频段，也使用 1900 MHz 频段。

在 D-AMPS 移动电话上，语音信号由麦克风拾取之后，先要对其进行数字化，然后要使用比增量调制和预测编码更复杂的模型进行压缩。由于考虑到人类发声系统的详尽特性，这种压缩算法能把标准 PCM 编码的 56 kbps 带宽压缩到 8 kbps 或更低。压缩过程是由声码器电路完成的。为了减小空中链路传送的数据量，压缩过程是在话机内完成的，而不是在基站或端局完成的。对固定电话来说，在话机内完成压缩不会带来任何好处——因为降低本地环路的通信量根本不会增加系统容量。

对于移动电话而言，在话机内完成数字化和压缩会带来巨大收益——利用时分复用技术，一个频率对可被 3 位 D-AMPS 用户共享。每个频率对支持的帧速率为 25 帧/秒(即每帧占用 40 ms)，而每帧又分为 6 个 6.67 ms 的时隙。最低频率对的情形如图 4.2(a)所示。

每帧拥有 3 位用户，这 3 位用户轮流使用上行链路和下行链路。例如，在图 4.2(a)中的时隙 1 期间，用户 1 可能在向基站发送数据，而用户 3 在接收基站传来的数据。每个时隙的长度是 324 位。其中，64 位用于保护时间、同步和控制，其余 260 位供用户使用。在这 260 位中，101 位用于噪声环境下空中链路所需的误码校正，因此最终只有 159 位用于压缩语音数据。由于每秒包含 50 个时隙，所以压缩语音数据的可用带宽恰好低于 8 kbps——标准 PCM 带宽的 1/7。

如果使用更好的压缩算法，还可以把语音数据率降至 4 kbps，这样的话，1 帧中就可以容纳 6 位用户了，见图 4.2(b)。在运营商看来，“3～6 位 D-AMPS 用户只需占用相当于 1 位 AMPS 用户所需的带宽”就是巨大成功，而且在很大程度上解释了个人通信业务普及的原因。

GSM(全球移动通信系统)

D-AMPS 广泛应用于美国，其改进形式应用于日本。世界其他地区使用的是一种称作 GSM 的系统。乍一看，GSM 和 D-AMPS 很相似，两者都是蜂窝系统、都使用“频分复用”(一个频率用于发送，另一个更高的频率用于接收；D-AMPS 的接收频率比发送频率高 80 MHz，而 GSM 的接收频率比发送频率高55 MHz)、都采用“时分复用”方法将一个频率对分成多用户共享的几个时隙。不过，GSM 的信道比 D-AMPS 要宽很多(GSM 信道宽度为 200 kHz，D-AMPS 信道宽度为 30 kHz)，而增加的用户数相对要小(一个 GSM 信道的用户数为 8，一个 D-AMPS 信道的用户数为 3)。因此，GSM 为用户提供的数据率比 D-AMPS 高得多。

如图 4.3 所示，每个频带的宽度为 200 kHz。GSM 拥有 124 对单工信道，每个单工信道带宽为 200 kHz，并以时分复用的方式提供了 8 路独立连接。每个正在使用的基站都会分配到频率对上的一个时隙。从理论上讲，每个蜂窝小区可以支持 992 个信道，但为了避免和邻近蜂窝小区发生频率冲突，许多信道并不能使用。在图 4.3 中，8 个阴影时隙全部属于同一个连接，每个方向4 个。在同

一时隙中，不会同时进行发送和接收，这是因为 GSM 不能同时发送和接收射频信号，发送和接收之间的切换是需要时间的。假定分配到 890.4/935.4 MHz 频点上时隙 2 的移动站向基站发送数据，它会使用靠下的 4 个阴影时隙(以及时间上相继的时隙)，并且向每个时隙添加数据直至所有数据发送完毕。

如图 4.3 所示的 TDM 时隙只是一个复杂组帧结构的局部。每个 TDM 时隙都有特定的结构，而若干组 TDM 时隙又构成了具有特定结构的复帧。该结构的简化示意如图 4.4 所示。从这张图中，我们可以看到：每个 TDM 时隙包含一个 148 位的数据帧，它占用了 577 μs 的时间(包括每个时隙之后的 30 μs 保护时间)。数据帧的开头和结尾都是 3 个 0，这几位用来对帧进行描述。数据帧还包含 2 个 57 位的“信息字段”，每个字段包含 1 位控制位。控制位用来标明下一个信息字段是用于语音还是用于数据。信息字段之间有一个 26 位的同步字段(又称训练字段)，接收机用它来同步发送方的帧边界。

一个数据帧占用的发送时间是 547 μs，而发射机只允许每隔 4.615 ms 发送一个数据帧，原因是每个发射机都要和其他 7 个发射机共享信道。每个信道的总数据率为 270.833 kbps，这个数据率要被 8 位用户等分。这样，每位用户得到的总数据率为 33.854 kbps，这是 D-AMPS 中单用户总数据率(50×324 bps＝16.2 kbps)的 2 倍还多。

从图 4.4 中可以看出，8 个数据帧构成一个 TDM 帧，而 26 个 TDM 帧构成一个 120 ms 的复帧。在这 26 个 TDM 帧中，时隙 12 用于控制，时隙 25 保留使用，所以只有 24 个时隙用于传输用户数据。

CDMA(码分多址)

CDMA 和 AMPS、D-AMPS 及 GSM 完全不同。CDMA 没有将可用频带分割为几百个窄带信道，而是允许信号在整个频谱范围内传送。利用编码原理，CDMA 可以将多个同时传送的信号分离开来。在 CDMA 中，“相遇数据帧会造成数据混淆”的想法不存在了——CDMA 认为多路信号之间是线性相加的。

我们不妨打个比方：在机场候机厅内，三三两两的人正在攀谈。TDM 就像一个人站在屋子中间轮流和别人交谈；FDM 就像分散在各处的人群，每群人都同时进行交谈，但互不干扰；CDMA 就像屋子中一直在说话的人，但每对谈话者都使用不同的语言。讲法语的那一对想听到的语言是法语，而把其他语言当作噪声。因此，CDMA 的关键在于能够提取出希望得到的信号，而把其他信号当作随机噪声加以抑制。

CDMA 的位时间被分成 m 个称作“码片”的、更短的时间间隔。每位包含的典型码片数为 64 或 128。为简单起见，下例采用的是 8 码片/位。

每站分配到唯一的称为“码片序列”的 m 位码。要传送比特“1”，就需要发送该站的码片序列；要传送比特“0”，就需要发送该站码片序列的反码。发送其他序列是不允许的。因此，对于 $m=8$ 的情形，假如 A 站分配的码片序列为 00011011，它通过发送 00011011 发送“1”，通过发送 11100100 发送“0”。

如果可用带宽提高至原来的 m 倍，发送信息量就可从 b 位/秒提高到 mb 码片/秒。这使得 CDMA 成为扩频通信方式之一(假定调制和编码技术没有发生变化)。假设 100 个站的可用带宽为 1 MHz，如果采用 FDM，则每个站将拥有 10 kHz 带宽，数据率可达 10 kbps(假定信道利用率为 1 bit/Hz)；如果采用 CDMA，则每个站可以利用 1 MHz 带宽，码片速率为 1 Mc/s。如果每一比特所含码片数小于 100，那么每个 CDMA 站的有效带宽要高于 FDM，信道分配的问题也解决了。

随着样本位数越来越多，零噪声“奈奎斯特”信道具有任意大容量；同理，在理想的零噪声

CDMA 系统中，系统容量(即站点数)也为任意大。在实际应用中，因受到各种物理限制，系统容量明显减小了。首先，我们假定所有码片在时间上是同步的——而这在实际中是无法实现的。可以做到的是：发送方和接收方通过传送预先确定的码片序列来实现同步，该码片序列的长度足以使接收机能够锁定。所有其他的(未同步的)传送均被视为随机噪声。正如人们可以预期的那样，码片序列越长，在噪声环境中正确检测出信号的概率就越高。为了更加可靠，比特序列可以使用纠错编码，而码片序列从不使用纠错编码。

我们的讨论隐含了这样的假设：在接收机看来，所有发射站的功率电平是一样的。无线通信是 CDMA 的典型应用领域。在无线通信系统中，有一个固定的基站和许多与基站距离随时变化的移动站。基站接收到的信号功率电平取决于它和发射机之间的距离。可以直观地推断出如下结论：移动站向基站发送信号的功率是接收到的基站信号功率电平的倒数。换言之，移动站从基站接收到一个弱信号要比接收一个强信号耗费更多的功率。基站也可以向移动站发出明确指令以使之提高或降低发射功率。

我们还假定接收机知道发送方是谁。理论上讲，如果提供了足够的计算能力，通过并行运行解码算法，接收机可以同时监听全部发送方。但实际的情况是“说起来容易，做起来难”。在这里，我们也没有谈到许多其他使问题更加复杂化的因素。但无论如何，CDMA 确实是一种非常巧妙的设计方案，并正在迅速地应用于无线移动通信之中。CDMA 的工作带宽为 1.25 MHz(D-AMPS 的工作带宽为 30 kHz，GSM 的工作带宽为 200 kHz)，在这个带宽范围内，CDMA 支持的用户数要远远超过其他系统。在实际中，CDMA 用户的可用带宽至少和 GSM 一样出色，而且往往比 GSM 要好很多。

第 12 课　个人计算机(PC)系统

提到“技术”，很多人都会联想到计算机。其实，我们生活的方方面面都存在着和计算机有关的事物。电视机等家用电器的内部就有微处理器，甚至汽车里也有一个计算机。不过，人们首先想到的还是“个人计算机”或者叫作 PC。

PC 是一种以微处理器为中心的通用工具，它由存储器、硬盘、调制解调器等很多不同的部件共同构成。所谓“通用”是指使用 PC 可以做许多不同的事情，既可以用它录入文件，也可以用它发送电子邮件、浏览网页和玩游戏。

PC 核心部件

- **机箱**：如果你用的是笔记本电脑，则机箱还包括键盘和显示屏；如果是台式计算机，则机箱通常就是带有指示灯和通风口的箱体，其中留有连接电缆的空间。机箱大并不一定意味着功能强，功能是由机箱里面的东西决定的。PC 制造商是根据主板类型设计、选择机箱的。
- **主板**：PC 最主要的电路板叫主板。内外部件都是通过主板互连的。有些重要部件直接置于主板之上，比如，计算机掉电后存储系统时钟等信息的 CMOS。主板大小和标准很多，当前最常见的是 ATX 和 MicroATX。
- **电源**：除 CMOS 由主板上的 CMOS 电池供电外，PC 的所有部件都要靠电源供电。电源连接着电池或者插座。
- **中央处理单元(CPU)**：CPU 也叫中央处理器，是包含微处理器的部件。微处理器是 PC 的运算

中心，软硬件的性能好坏都要依赖它。Intel 公司和 AMD 公司是最大的 PC CPU 制造商。32 位 CPU 和 64 位 CPU 是两种常见的 CPU 结构，不同软件的划分都基于这两种 CPU 结构。

- **内存(RAM)**：最快的处理器工作时也需要缓存来存储信息。对于 CPU 来说，RAM 的作用就像厨师的工作台——那是放置备用食材和刀具的地方。高速 PC 不仅需要高速 CPU，而且还需要足量的 RAM。
- **驱动器**：驱动器旨在存储备用数据。硬盘驱动器(或称固态驱动器)存储着 PC 操作系统和应用软件。驱动器还包括用来读/写 CD、DVD、蓝光碟等的光盘驱动器。驱动器连接主板的方式取决于其使用的驱动控制器技术类型——是早期的 IDE 标准，还是近来的 SATA 标准。
- **冷却设备**：计算机处理量越大，它产生的热量越多。CPU 和其他部件可以承受一定的热量。但如果无法正常冷却，PC 就会过热，从而导致部件和电路的严重损坏。风扇是最常见的冷却设备。此外，CPU 外套着金属散热片，从 CPU 上把热量吸走。有的计算机用户(如游戏玩家)还会采用更昂贵的冷却方案(如水冷系统)来满足更高的需求。
- **电缆**：至此谈及的所有部件都要通过某种电缆进行连接。电缆用于传输数据和电能。PC 电缆应当整齐地布置在机箱内，保证空气流畅通过。

接口、外设和扩展槽

- **显卡**：有些主板提供集成显卡，有些主板提供显卡插槽——你可以将独立显卡插入。PC 显卡对复杂的图形数据进行处理，减轻了 CPU 的工作量。主板允许使用诸如符合 AGP 接口标准或符合 PCI 标准的显卡。
- **接口**：接口常用来指 PC 外部插入电缆的地方。人们根据用途给接口起名字，如“USB 口”或者“以太网口”。许多接口直接固定在主板上。在 PC 上，你会发现以下一些接口：(1)USB 接口；(2)网络接口——以太网接口和 FireWire 接口；(3)视频接口——VGA 接口、DVI 接口、RCA 视频接口、S-视频接口和 HDMI 接口；(4)音频接口——迷你模拟音频插口或 RCA 音频接口；(5)老式接口(即当前计算机罕用的接口)——连接打印机的并行接口和连接键盘、鼠标的 PS2 接口。
- **外设**：不在机箱内的设备称作外设。外设包括显示器、键盘、鼠标等基本输入/输出设备，还包括打印机、扬声器、耳机、话筒、网络摄像头和 USB 闪存驱动器。凡是可插入 PC 接口的都叫外设。
- **扩展槽**：偶尔你想增加一个主板上没有指定插槽的 PC 部件，这就是为什么需要扩展槽的原因。能插入扩展槽的部件称作“卡”(盖因其结构扁平、类似卡片吧)。有了扩展槽，你就可以增添额外的视频卡、网卡、打印机接口、电视接收器等许多定制附件。

启动 PC

启动 PC 时，机器内部要经历一个准备过程——启动过程。“启动”(boot)一词是“bootstrap”的简写形式——来自那句老话“Pull yourself up by the bootstraps”(意思是“一切都要从零开始”)。这个启动过程由 PC 的“基本输入/输出系统”(BIOS)控制完成。

BIOS 是储存在闪存芯片中的软件，闪存芯片嵌在 PC 的主板上。有时候，PC 制造商会发布 BIOS 更新，你可以按照说明“刷新 BIOS”。

除了控制启动过程，BIOS 还提供了 PC 硬件的基础配置界面。通过此界面，用户可以配置启动时驱动器读取顺序、处理器运行速率等。查阅 PC 说明书，就能找到进入 BIOS 界面的方法。启动计算机时也会显示相关信息——例如，你会看到“Press DEL to enter Setup Menu.”(按下 DEL 键

进入配置菜单)。

PC 启动过程概述如下。

1. 电源按钮激活 PC 电源，将电能提供给主板和其他部件。
2. PC 进行“开机自检”(POST)。POST 是 BIOS 中的硬件故障检测程序。如果一切正常，蜂鸣器就会鸣叫一声。如果蜂鸣器发出其他声响，就说明出现了硬件故障。根据鸣叫情况，PC 维修人员对照图表判定哪个部件出现了故障。
3. PC 显示器呈现启动过程的详细信息，包括 BIOS 制造商和版本号、处理器技术指标、内存(RAM)容量、检测到的驱动器。很多 PC 不显示这些信息，而显示制造商徽标。如果你不想看徽标而想看信息，则可以在 BIOS 配置界面中关闭徽标显示。
4. BIOS 尝试访问启动盘第一扇区。第一扇区就是该盘的前几千字节(如果驱动器从可用外存地址的开始处顺序读取)。启动盘通常就是包含操作系统的“硬盘”(或固态驱动器)。配置 BIOS 可以修改启动盘，按下某几个特定的组合键(往往提示在启动屏幕上)就可以中断启动过程。
5. BIOS 确认启动盘第一扇区存在启动程序，并将启动程序载入内存(RAM)。启动程序用于寻找和启动 PC 操作系统。
6. 启动程序进入内存后，BIOS 就将 PC 控制权交给它。接下来，启动程序开始将操作系统载入内存。
7. 启动程序完成任务后，将 PC 控制权交给操作系统。接下来，操作系统就可以为用户服务了。

PC 操作系统

PC 启动完毕后，用户就可以通过操作系统(OS)来控制计算机了。多数非苹果 PC 使用 Microsoft Windows 操作系统或 Linux 操作系统，这些操作系统可以运行在不同的 PC 硬件上。然而，OS X 操作系统主要运行在苹果计算机中。

操作系统负责多项任务，任务分类如下。

- **处理器管理：**将处理器工作分解为易管理的小块，并为其分配优先级；然后，将它们发送到 CPU。
- **内存管理：**协调进出内存的数据流，判定何时使用硬盘上的虚拟内存来弥补内存的不足。
- **设备管理：**提供计算机内部部件和外部设备之间的软件接口。
- **外存管理：**将数据导入永久存储设备——硬盘驱动器、固态驱动器、USB 驱动器或者其他形式的外存中。例如，外存管理工作有助于文件的创建、阅读、编辑、移动、复制和删除。
- **应用程序接口：**提供软件程序和 PC 之间的数据交换。应用程序必须通过应用程序接口和操作系统联系。
- **用户接口(UI)：**提供用户和计算机之间的交互。

第 13 课　电子设计自动化

传统上，电子电路设计可分为模拟设计、数字设计两大类。这两类课程通常是分别教授的，电子工程师一般只精通一个领域。这两类设计还可以进行专业细分。例如，射频模拟电路设计、数字集成电路设计和混合信号电路设计(当数字和模拟两大领域相遇时)。此外，在嵌入式系统设计中，

软件工程也扮演着越来越重要的角色。

在消费类电子产品中，数字电子技术更加重要。汽车拥有复杂的控制系统；许多家庭如今也拥有了个人计算机；曾被视为模拟产品的收音机、电视机和电话也都在数字化；CD 几乎完全替代了唱片。这些变化的出现使得产品的生命周期缩短了。在不到一年的时间里，附近商店里的原有产品很可能就完全被新型数字电子产品替代了。

电子设计自动化

为了紧跟如此迅速的变化步伐，电子产品必须以极快的速度设计出来。模拟设计依然是专业化程度很高、收入颇丰的职业之一。而数字设计已经非常依赖“计算机辅助设计”(CAD)，它又称“电子设计自动化”(EDA)。EDA 工具可以完成“综合”和“模拟”两类任务。“综合”是将设计指标转变为实际的设计实现，而“模拟”是对设计指标或详细的实现方案进行“演习”，以确保其正确运行。

用于综合和模拟的 EDA 工具要求设计者将想法转化成工具设计。这可以通过使用“图形化软件包”绘制设计原理图来实现，这种方法称为“原理图输入法”。

另一种方法是用文本形式来表示设计，这种方法和软件编程很相似。对数字电路硬件的文本描述既可以采用编程语言(如 C 语言)，也可以采用“硬件描述语言”(HDL)。在过去的约 50 年中，设计出了几种 HDL。Verilog 和 VHDL(VHSIC 硬件描述语言，VHSIC 代表超高速集成电路)是两种广泛应用的硬件描述语言。标准化 HDL 是很重要的，因为标准化语言可以被不同生产商提供的不同 CAD 工具使用。在 Verilog 和 VHDL 出现之前，每个 CAD 工具都有自己的 HDL。在不同 HDL 之间进行转换(例如，使用其他模拟器来验证某个综合工具的输出)既费时又费力。

逻辑门

门电路是构成数字电路的基本模块。门电路是由几个输入和(通常情况下)一个输出组成的电子器件。在通常情况下，输入和输出是“逻辑 0”与“逻辑 1”两种状态之一。逻辑值可以由电压(例如，0 V代表“逻辑 0”，3.3 V 代表“逻辑 1”)或电流来表示。门电路利用全部输入实现某种逻辑运算以产生输出。当然，数字门电路说到底也是模拟器件。不过，为了简单起见，我们倾向于忽略其模拟本质。

我们可以买到一片如图 5.1 所示的集成电路，它包含 4 个完全一样的门电路(注意：有两个连接用于器件的正、负电源。一般情况下，这些连接不显示在逻辑原理图中)。将成百上千个这样的器件连接起来，可以构成一个数字系统——实际上，许多系统就是这样设计的。虽然每片集成电路的成本也许只有区区 10 美分，可用于印制电路板设计和系统组装的花销却是非常可观的——采用这种设计方法并不合算。

大规模生产的集成电路(从“触发器”到“微处理器”)集成了越来越多的复杂功能。日益增加的复杂度带来了灵活性的提高——微处理器通过编程可以完成几乎无限种任务。因此，某些数字系统的设计过程就是“选择标准元器件，将其连接起来”。不过，标准器件可能无法提供某种功能，这是避免不了的。设计者只好选用“分立门电路”来实现这种功能，或者设计专用集成电路来实现这种功能。“设计专用集成电路”或许令人畏缩，但不要忘记——系统成本在很大程度上取决于器件互连成本，而非单个器件成本。

ASIC(专用集成电路)和 FPGA(现场可编程门阵列)

高性能“全定制集成电路”设计是一项难度很大的工作。在全定制集成电路设计中，所有事情(甚至每个晶体管)都需要设计(尽管可以使用元件库)。不过，利用“门阵列”构造“半定制集成电路”已有很多年。顾名思义，“门阵列”就是集成了逻辑门阵列的集成电路。所以，使用门阵列设计

“专用集成电路”(ASIC)就是要定义阵列中的门电路之间的连接。从实现角度看，就是要设计一层或两层的金属互连线。集成电路的制造流程分为 7 个步骤(或更多)，除最后一步“金属化”外，其他步骤都可提前完成。由于门阵列可以大量生产，所以单个器件的成本比较低。

ASIC 通常是指全定制和半定制集成电路。“可编程逻辑”是另外一类集成电路。最早的“可编程逻辑器件”(PLD)是“可编程逻辑阵列”(PLA)。和门阵列一样，PLA 也是由“功能待定逻辑”组成的阵列。PLA 和“掩模编程”门阵列的区别是“用大电压(极性通常为负)可以对 PLA 进行配置”。图 5.2 是 PLA 的通用结构。PLA 有若干输入(A，B，C)和若干输出(X，Y，Z)，一个“与平面”和一个“或平面”。图中显示了输入和乘积项(P，Q，R，S)之间的连接，以及乘积项和输出之间的连接。剩余的连接在编程时就被去除了。有些 PLA 是电可编程的，也可以使用紫外线照射进行编程。“可编程阵列逻辑”(PAL)扩展了 PLA，PAL 包含 12 个触发器。近年来，可编程器件变得更复杂了——包括“复杂可编程逻辑器件”(CPLD)和“现场可编程门阵列”(FPGA)。

设计流程

数字系统大多是“时序的”，即系统存在各种“状态”，输出取决于“现态”。早期的某些计算机设计是“异步的”，即输入稳定下来后就向新状态转移。数字系统的“同步”设计趋势已经有多年了。在同步系统中，状态的改变是由一个或多个时钟信号触发的。为了设计可靠的系统，人们定义了规范的设计方法。采用“分立门电路”设计(同步时序)数字系统，可按照如下步骤进行。

1. 写出设计规格。
2. 如果需要，则将设计分解为若干个部分，并为每个部分写出设计规格。
3. 根据设计规格，画出状态机转换图。该图包含系统中的每个状态、引起状态变化的输入条件和每个状态的输出。
4. 将状态数最小化。这是一个可选步骤，并不是在所有情况下都有用。
5. 为每个状态分配布尔变量。
6. 推导出次态逻辑和输出逻辑。
7. 对次态逻辑和输出逻辑进行优化，从而将所需的门数量减至最少。
8. 为逻辑门选择合理的布局，即集成电路中包含哪些逻辑门及各个集成电路在印制电路板上的位置。
9. 设计集成电路之间的走线。

一般而言，步骤 1 和步骤 2 是不能被忽略的。因为这里需要设计者的创造性工作。大多数数字设计书籍集中讨论了步骤 3 和步骤 7。步骤 8 和步骤 9 可以手工完成，而“元件布局和布线”是最早成功实现自动化的设计任务之一。如果将设计转换成计算机可读的形式，那么就可以在设计的不同阶段进行仿真。为了进行布局和布线，在步骤 7 中往往使用原理图输入程序，这样就可以输入电路的门级结构。原理图可以转换成适合逻辑仿真器的形式。在步骤 9 完成之后，电路结构(包括互连线的电阻和电容引起的延迟)可以提取出来，然后进行再次仿真。

因此，在 ASIC 和 FPGA 上实现数字设计需要配置逻辑模块之间的连接。正像刚才提到的，我们不能忽略步骤 1 和步骤 2，而步骤 8 和步骤 9 可以自动完成。使用 HDL 意味着可以将设计输入 CAD 系统中，并可以在步骤 3 或步骤 4 中进行仿真，而不是在步骤 7 中。所谓的“寄存器传输级”(RTL)综合工具可以自动完成步骤 6 和步骤 7。步骤 4 仍需手工完成。步骤 5 可以自动完成，并可以对某种状态分配结果进行快速评估。“行为综合”工具开始出现在自动化设计流程中的步骤 2 附近。图 5.3 显示了基于 RTL 综合设计的整个设计流程。

第 14 课　现场可编程门阵列

FPGA 是什么？

“现场可编程门阵列”(FPGA)是数字集成电路之一。FPGA 由“可编程逻辑块”和块间的“可编程互连”组成。利用 FPGA，设计工程师可以编程实现各种各样的功能。见图 5.4。

依据使用方式的不同，FPGA 可分为两类。一类 FPGA 只能编程一次，而另一类 FPGA 可多次编程。显然，只能编程一次的器件被称为“一次编程型(OTP)”FPGA。

FPGA 名称中的“现场可编程”是指研发人员“在现场”对 FPGA 编程，而不是由芯片制造商以“硬连线”方式实现器件的内部功能。也就是说，既可以在实验室中对 FPGA 进行配置，也可以对“已投入使用的”电子系统中的 FPGA 进行配置。如果系统中的某种器件具备可编程能力，那么就可称作“在系统可编程”器件。

FPGA 引人关注的原因

数字集成电路的种类有许多，如“豆粒糖逻辑”(仅含少量简单、固定逻辑功能的小规模器件)、存储器件和“微处理器”(μP)。这里，我们特别关注的是“可编程逻辑器件”(PLD)、“专用集成电路”(ASIC)、“专用标准产品”(ASSP)和 FPGA。

我们需要知道：芯片制造商预先限定了 PLD 的内部结构。这种内部结构允许工程师“现场”编程实现各种功能。不过，和 FPGA 相比，PLD 包含的逻辑门比较少，而且 PLD 实现的功能也少得多、简单得多。

在该领域的另一端是 ASIC 和 ASSP，其逻辑门数量多达数亿个，可实现的功能也极其庞大和复杂。ASIC 和 ASSP 的设计流程与制造工艺是相同的，都是为特定应用而定制的芯片。二者的唯一区别是，ASIC 是为某一家公司设计制造的，而 ASSP 面向的是众多的客户(此后当提及 ASIC 时，除非特别说明或与上下文不符，该术语同时也指 ASSP)。

尽管 ASIC 在大小(即晶体管数量)、复杂度和性能上都是最好的，但设计与制造一片 ASIC 却是极其耗时和昂贵的。ASIC 还有一个缺点，就是最终设计“固化”在芯片中，无法修改。

而 FPGA 的位置处在 PLD 和 ASIC 的中间——FPGA 的功能可以“现场”定制，这一点像 PLD；FPGA 中的几百万个逻辑门可以实现“以前只有 ASIC 才能实现的”庞大、复杂功能。

开发基于 FPGA 的设计成本要比 ASIC 低得多(尽管 ASIC 在量产后非常便宜)。此外，基于 FPGA 的设计更易于修改，而且上市时间更快。FPGA 为众多具备创新能力的小型设计公司提供了生存机会。因为 FPGA 不仅可供“环境优越”的大公司进行系统开发，也便于“条件简陋”的小公司进行产品设计——即无须投入巨额的“一次性工程”(NRE)成本和购买昂贵的 ASIC 设计工具，几个工程师就能在 FPGA 测试平台上实现他们的软、硬件设想。

FPGA 的用途是什么？

20 世纪 80 年代中期，当 FPGA 刚刚出现时，它主要用于实现“胶连逻辑”“中等复杂度”状态机和比较有限的数据处理任务。20 世纪 90 年代早期，随着 FPGA 容量和复杂度的提升，当时的市场

主要是通信和网络领域，FPGA 用来处理、传输大量数据。到 20 世纪 90 年代末，在消费产品、汽车和工业等领域，FPGA 的使用量出现了高速增长。

FPGA 常用于 ASIC 原型设计，或者作为新算法硬件实现的验证性平台。然而，由于开发成本低和上市时间短，FPGA 越来越多地出现在最终产品中。实际上，几家大生产商已经拥有可以和 ASIC 竞争的 FPGA 器件。

在 21 世纪的头几年，几百万门的高性能 FPGA 就已经出现。它们用在嵌入式微处理器内核、高速输入/输出接口等应用中。现实情况说明：今天的 FPGA 几乎无所不能——它可以用来实现通信设备、软件无线电；它可以用来实现雷达、图像及其他数字信号处理应用；它可以用来实现构成“片上系统”(SoC)的全部软、硬组件。

具体地说，FPGA 正在占领以下 4 个主要市场：ASIC 和定制芯片、“数字信号处理”(DSP)、嵌入式微控制器和物理层通信芯片。此外，FPGA 还为自己开辟了一个新市场——“可重配计算”(RC)。

- **ASIC 和定制芯片：**如前所述，当今的 FPGA 越来越多地用于实现过去用 ASIC 和定制芯片才能实现的各种设计。
- **数字信号处理：**高速数字信号处理早已由特制微处理器——“数字信号处理器”来实现了。然而，今天的 FPGA 包含了嵌入式乘法器、专用算术布线和大量片上 RAM——这些硬件为数字信号处理运算提供了方便。当上述特点结合 FPGA 强大的并行处理功能后，FPGA 的性能比最快的数字信号处理芯片高 500 倍以上。
- **嵌入式微控制器：**小型的控制功能通常由“微控制器”这种专用嵌入式处理器来实现。微控制器是一种低成本器件，在其处理器核周围集成了片上程序和指令存储器、计时器和输入/输出接口。而 FPGA 的价格不断下降，最小的 FPGA 器件也足以实现一个微处理器内核及一组定制的输入/输出功能。结果是，对嵌入式控制应用而言，使用 FPGA 的吸引力与日俱增。
- **物理层通信：**长久以来，FPGA 用于实现通信芯片和高级网络协议之间的“胶连逻辑”。实际上，今天的高端 FPGA 可以容纳多个高速无线收发器，即可将通信和网络功能固化于单个器件中。
- **可重配计算：**利用 FPGA 的固有并行性和“可重配”特性对软件算法进行“硬件加速”。目前，许多公司正在建造基于 FPGA 的可重配计算引擎。这些引擎用于完成硬件仿真、密码分析、新药品研发等领域的计算任务。

第 15 课　VHDL

VHDL 是什么？

VHDL 是一种为描述数字系统而设计、优化的编程语言。VHDL 具备多项适合描述电子元器件(从简单逻辑门到微处理器、定制芯片)行为的功能。这些功能可以精确描述诸如信号的上升时间和下降时间、门电路传输延迟、函数操作等电路行为的电气特性。将既有的 VHDL 仿真模型作为构成模块，可以实现对更大型电路(采用原理图、方框图或系统级 VHDL 等描述方法)的仿真。

VHDL 是一种通用编程语言。和其他高级编程语言一样，VHDL 将复杂的设计概念表达为计算机程序；为了实现电路自动综合或系统仿真，VHDL 可将复杂电子电路的行为输入一个设计系统中。和 PASCAL、C、C++一样，VHDL 也具有结构化设计的特点，并提供了丰富的控制指令集

和数据表示法。和上述编程语言不同的是：VHDL 能够描述并发事件。这一点是非常重要的，因为硬件在运行时就是并发的。

VHDL 最重要的应用之一是以所谓“测试平台”的形式记录电路的性能指标。测试平台是验证电路时域行为的激励源及相应期望输出的 VHDL 描述。在 VHDL 工程项目中，测试平台是必不可少的，应该和其他电路描述一同创建。

VHDL 是标准语言

熟悉并掌握 VHDL 最具说服力的原因之一是，VHDL 在电子设计领域是作为一种标准被采纳的。使用 VHDL 这样的标准语言，实际上保证了设计者将不必因新一代设计工具不支持已选择的输入方法，而放弃原有的设计思路，重新进行设计。使用标准语言意味着设计者可以更加充分地利用最新设计工具，可以了解数以千计的其他工程师所积累的经验、知识——他们中的许多人也正在解决类似的问题。

VHDL 简史

VHDL 代表“超高速集成电路硬件描述语言”。它是在 20 世纪 80 年代早期作为美国国防部资助的一个高速集成电路研究计划的副产品而开发出来的。在这个计划中，研究人员反复遇到描述“庞大规模”(相对于当时的技术水平而言)电路的难题和如何管理多个工程师小组参与进行的大型电路设计的难题。由于当时只有门级设计工具，人们很快就认识到需要性能更好、更加结构化的设计方法和工具。

为了应对这个挑战，由来自“国际商用机器公司”(IBM)、“德州仪器公司”(TI)和 Intermetrics 三家公司的工程师组成的一个小组与美国国防部签署了一份协议——要设计和实现一种崭新的、基于语言设计的描述方法。

VHDL 的最早版本(7.2 版本)是在 1985 年公布的。1986 年，“电气电子工程师学会”(IEEE)提出了语言标准化建议。由商业、政府和学术代表组成的小组对该建议进行了实质增强与改进之后，IEEE 于 1987 年完成了标准化工作。

该标准(IEEE 1076-1987)就是今天市场上所有仿真和综合产品的基础。1994 年，IEEE 发布了 VHDL 的增强和更新版本(IEEE 1076-1993)，VHDL 工具提供商马上做出反应——在其产品中加入了 VHDL 的最新特点。

虽然 IEEE 1076 完整定义了 VHDL，但是 VHDL 的某些方面使之很难实现完全可移植的设计描述(所谓“完全可移植的设计描述”是指该设计描述在使用不同工具生产商提供的仿真工具进行仿真时，会得到完全相同的仿真结果)。这个问题源于 VHDL 支持多种抽象数据类型，却没有很好地描述各种信号强度和一些常用仿真条件(如“未知状态”和“高阻状态”)。

在 IEEE 1076-1987 被采纳后不久，为了使用户能够对复杂电路进行精确仿真，仿真器提供商开始提供非标准数据类型以对其进行改进。这就带来了问题，因为在某个仿真器中输入的设计描述在其他仿真环境下是不兼容的。VHDL 很快就变成非标准语言了。

为了避开非标准数据类型这个问题，IEEE 又研究出了另外一个标准。该标准(标准号为 1164)定义了一个包含标准 9 值数据类型定义在内的标准包(标准包是 VHDL 的特色之一，即允许将公用数据类型声明集中在一个外部库函数中)。标准数据类型称为 std_logic，而 IEEE 1164 标准包常被称作“标准逻辑包”。

IEEE 1076-1987 和 IEEE 1164 共同构成了当前广泛使用的 VHDL 标准。(IEEE 1076-1993 正逐步进入 VHDL 主流，但它并没有为综合用户增添任何明显的新功能。)

IEEE 1076.3(常被称作“数值标准”或“综合标准”)定义了标准包和对与实际硬件相关的 VHDL 数据类型的解释。该标准于 1995 年年底发布，目的在于替代由许多综合工具商创建的、在其产品中提供的定制(非标准)包。

IEEE 1076.3 为综合用户做的和 IEEE 1169 为仿真用户做的完全相同：在确保兼容不同生产商工具的同时，提高 IEEE 1076 的功能。IEEE 1076.3 包含以下几项：

- IEEE 1076 定义了比特型和布尔型数值的硬件解释文档，IEEE 1164 定义了std_ulogic 类型的硬件解释文档。
- 提供对 std_ulogic 类型数值进行“无关项”或者“通配项”测试的功能。该功能对综合尤为有用，因为用“无关项”来表达逻辑往往很有用。
- 对标准有符号和无符号算术数据类型的定义及其算术、移位、类型转换运算的定义。

“对仿真模型进行时序信息标注”是精确数字仿真的重要方面之一。VHDL 1076 标准描述了一组可用于时序标注的语言特点。然而，它并未描述出一种在时序模型外表达时序数据的标准方法。

出于多种原因，将仿真模型的行为描述和时序技术规格分开是很重要的。Verilog HDL (VHDL 的竞争对手)的一项主要优势是：Verilog HDL 包含了一种旨在进行时序标注的功能，该功能(即“标准延迟格式”——SDF)允许以表格形式表示时序数据，并可以在仿真时将其包含进 Verilog 时序模型中。

在 1995 年的晚些时候，IEEE 公布的 IEEE 1076.4 将这种功能以标准包的形式加入 VHDL 中。推动这种标准化的主要原因是为了使 ASIC 等生产商更容易生成可同时应用于 VHDL 和 Verilog HDL 的时序模型。因此，IEEE 1076.4 和 Verilog SDF 底层数据格式是非常相似的。

何时采用 VHDL?

为什么要用 VHDL 进行设计？原因可能有许多。当向 VHDL 工具提供商请教这个问题时，你首先得到的答案是“VHDL 将提高工作效率”。不过，这是什么意思呢？你真指望使用 VHDL 比使用现有设计方法能更快完成手中的项目吗？回答是肯定的。不过，刚刚使用的时候很可能不会这样，只有以结构化方法应用 VHDL 才会这样。

VHDL 和结构化软件设计方法是相同的。当你使用结构化的、自顶向下的设计方法的时候，VHDL 才是有益的。当你超越了 VHDL 学习曲线，并积累了一定数量的可重用 VHDL 组件之后，工作效率很快就会出现真正的提高。

当你为增进项目小组成员间的交流而开始使用 VHDL 时，当你利用更强大的工具进行仿真和设计验证时，工作效率就会提高。此外，VHDL 允许在更抽象的层次上设计。你不再集中精力于门级实现，你可以去思考设计的行为功能。

VHDL 是如何提高工作效率的？是通过简化公用 VHDL 模块库的创建和使用。VHDL 使设计重用变得顺理成章。当发现了可重用代码的好处之后，你很快会发觉自己正在考虑采用 VHDL，从而使自己编写的语句更加通用。编写可移植代码将会成为自觉的行为。

使用 VHDL 的另外一个重要原因是 EDA 工具和目标器件技术的快速发展。使用 VHDL 这样的标准语言可以极大提高转用更高级工具(例如，从基本的低价位仿真器到更高级的仿真器)的可能性，而不必重新输入电路描述。使用标准输入方法也将提高将电路重定位于新型目标器件(如 ASICs、FPGAs 和 CPLDs)的能力。

第 16 课　数字信号处理

我们说的话不是数字信号。数字信号是一种由“1”和“0”组成的、能用数学方法处理的语言。我们讲出的话是现实世界中的模拟信号。我们天天遇到的现实世界信号都是模拟信号——如声音、光、温度和压力。数字信号是模拟信号的数值表示。在数字世界里，对这些信号进行处理可能会更容易、成本更低。在现实世界中，我们可以通过“模数转换器”将信号转换为数字信号，然后对信号进行处理；如果需要的话，再用“数模转换器”将信号转换回模拟世界中。

模数转换

模数(A/N)转换的第一步是对模拟信号进行采样(见图 6.1)。这一步是由“采样保持电路”完成的，采样保持电路按照固定的采样间隔进行采样。采样间隔的长度和采样周期的长度相同，而采样周期的倒数就是采样频率 f_s。根据奈奎斯特采样定理，为了确保准确记录信号，最高频率为 W Hz 的信号(“带限信号”)在每秒内必须采集至少 $2W$ 个样本。如果不能满足这个最低要求，就会出现混叠失真。混叠会使高频信号出现在较低频段。为了保证不会出现混叠，在采样前要进行低通滤波。这个低通滤波器(“抗混叠滤波器”)将频率超过所选采样频率一半的信号全部滤除。

经过一段短暂的采集时间(其间采集到一个样本)之后，采样保持电路将这个样本一直保持到下一个采样间隔的开始。A/D 转换器需要利用这个保持时间来生成模拟样本对应的码字。

A/D 转换器为每个模拟样值选择一个量化电平。一个 N 位转换器可以在 2^N 个可选量化电平中进行选择。量化电平数越多，量化误差(量化电平和实际电平之差)越小。最大量化误差不会超过“量阶”的一半。量阶计算公式为 $Q=R/2^N$，其中 R 是模拟信号的满量程范围，N 是转换器所用的位数。相对于量化误差的信号强度使用“动态范围”和“信噪比”来度量。

数字信号是用一组竖线表示的，竖线顶部的圆圈标出样本选用的量化电平。A/D 转换器的比特率为 Nf_s，这里的 f_s是采样率。

最后，要为每个数字化样本分配一个码字，这样就完成了 A/D 转换过程。A/D 转换的结果就是得到一串数字比特流。数字信号处理就是对这些码字集合进行处理。

总结一下，A/D 转换过程由抗混叠滤波、采样、量化和数字化四步组成。

数模转换

数字信号处理完成之后，就必须进行数模(D/A)转换(见图 6.2)。首先，将码字转换为和码字所代表数字的大小成正比的模拟电压。这个电压值要在零阶保持器中保持到下一个码字出现，即需要保持一个采样间隔。这样，信号就是阶梯状的，信号中包含着频率超过 W Hz 的成分。D/A 转换的最后一步就是：使用平滑滤波器滤除那些频率超过 W Hz 的成分。

对于采样信号中的任意频率 f，其因采样而产生的镜像频率将出现在无数个频率处($kf_s \pm f$ Hz)。当采样率低于要求的奈奎斯特率(即 $f_s<2W$)时，高频信号的镜像就会因为混叠而错误地出现在基带内(或奈奎斯特范围内)。通常要避免“欠采样”发生，但“欠采样”也有用处。例如，高频窄带信号的采样率可以是信号带宽的 2 倍，而不必是信号最高频率的 2 倍。信号的所有重要特征都可以从因采样而出现在基带的频谱副本中得到。由于信号频率和采样率之间的关系不同，可能会

出现“频谱反转”现象——基带频谱的形状与信号真实频谱的形状正好相反。

数字信号处理技术

假如存在一种可以实现所有设计的“万能微处理器”，电子行业的竞争就不会如此激烈了。对于多数电子设计而言，可用于实现所需功能的处理器技术不止一种。需要面临的问题是：如何在预算范围内选择一种处理器技术，它不仅具备最佳的性能、大小、功耗特点，而且拥有易于快速开发的软件工具。

数字信号处理器

经过了几十年的发展，“数字信号处理器”(DSP)依然是极具竞争力的微处理器。数字信号处理器是信号处理的核心。数字信号处理器是一种速度极快、功能强大的微处理器。DSP 的与众不同之处在于它能够实时处理数据。实时处理能力使 DSP 非常适合于那些不容许任何延迟的应用。例如，你用过两人无法同时说话的手机吗？你必须等对方说完才能说话。如果你们两人同时说话，信号就会掉线——你听不到对方的声音。今天的数字手机使用了 DSP，这样人们就可以正常通话了。手机中 DSP 处理声音的速度非常快——你在说话的同时就可以听到对方的声音。

和其他处理器相比，使用 DSP 进行设计有如下几点优势：

- 可实现单周期“乘法-累加”运算。
- 能够对实时性能进行模拟和仿真。
- 灵活性好。
- 可靠性高。
- 有利于提高系统性能。
- 有利于降低系统成本。

不过，实现数字信号处理还可以选择其他技术。和 DSP 相比，这些技术如何呢？

现场可编程门阵列

现场可编程门阵列(FPGA)具有“在系统内可重新配置”的能力——对于开发需要多次试用的应用，这种能力是巨大的优势，能提供快速的上市时间。由于生成了专用逻辑电路，所以对于指定任务 FPGA 能提供更好的性能。可是，FPGA 相当昂贵，功耗也往往大大高于功能相近的 DSP。在无线通信基础设备等设计中选用了 FPGA 技术，但通常是将 FPGA 和 DSP 联合使用——这样做灵活性更大，价性比更好，系统功耗更低。

专用集成电路

通过设计，专用集成电路(ASIC)可以极好地完成某些特定功能，而且功率效率也可以设计得很好。不过，ASIC 无法进行现场编程，即使是在产品开发阶段，也无法对其功能进行修改和升级。因此，每次推出一个新产品都需要重新进行设计，并由晶圆代工厂制造出来——这样做不仅造价高昂，也不利于迅速上市。可编程 DSP 与 ASIC 不同，可以不需要修改硬件而进行升级，升级只需修改软件程序——这样大大降低了系统成本，并可通过下载代码而实现(售后)功能增强。因此，在实时信号处理应用中，可编程 DSP 是系统的基础，而 ASIC 常用作系统的总线接口、胶连逻辑及功能加速器。

通用微处理器

ASIC 是为特定用途优化的。与 ASIC 相反，通用微处理器(GPP)最适合完成多种不同的任务。但在最终产品必须实时响应的应用中，或者必须在电池驱动下实现实时响应的应用中，由于实时性能较差、功耗大，所以 GPP 就被排除在外了。GPP 正在越来越多地被视为业界的庞然大物——为了适应不断变化的实时应用市场，不断增强的 PC 兼容性和台式计算机性能使 GPP 受到了拖累。因为人们越来越喜欢小巧的手持无线产品(其处理器功耗为毫瓦级，而不是瓦级)，所以 DSP 就成为可编程技术的最佳选择。随着数字互联网设备变得更小、更快、更便于携带，这种趋势一定会保持下去。

第 17 课　数字信号处理器

只要是处理器，就能进行数字信号处理。不过，数字信号处理器(DSP)的速度最快、效率最高。

结构

传统处理器遵循“冯·诺依曼结构”，该结构采用一个共享存储器，同时存储程序和数据(见图 6.3)；而 DSP 采用的是“哈佛结构”或“改进型哈佛结构”，该结构包含多个程序存储器和数据存储器，以及访问这些存储器的多套总线(见图 6.4)。如此设计意味着从存储器取指令或取数据所花费的等待时间要少很多。实际上，至少可以同时取得一条指令和一个数据。这种任务的重叠称为“流水线”。除多存储器和多总线外，DSP 都有高速乘法器、累加器、移位器，许多 DSP 都有硬件“循环缓存器”。“地址产生器”可以加速对寄存器寻址的存储器访问。

分类

DSP 分为定点和浮点两大类。定点 DSP 使用固定的位(比特)数来代表实数。二进制小数点的位置可以由编程人员决定，这个位置决定了可以表示的实数范围。可用精度会随着表示范围的增加而下降，因为二进制小数点右边的位数减少了。在 16 位数据中，可能出现的格式有 16.0、15.1、14.2、13.3、12.4、11.5、10.6、9.7、8.8、7.9、6.10、5.11、4.12、3.13、2.14 和 1.15。16 位定点数据格式的动态范围都是一样的，即 $20\log 2^{16}=96.3$ dB。动态范围的计算方法为 20 log(满量程范围/最小可分辨差别)。

浮点 DSP 使用尾数和指数来表示实数，这种方法很像科学记数法——将尾数和指数组合成一个 32 位数。浮点器件的动态范围是用 2^E 的最大值和最小值进行计算的，此处 E 是指数。对于24 位尾数和 8 位有符号指数的表示方法而言，动态范围是 $20\log(2^{127}/2^{-128})=1535.3$ dB。动态范围大，意味着系统具备更大的表示很宽范围输入信号的能力，从很小的信号到很大的信号。

选用

为特定应用选择合适的 DSP 不太容易。第一，需要明确是选择定点器件还是浮点器件。一般而言，定点器件比较便宜，速度也快，但浮点器件更便于编程，更适于运算密集型算法。第二，DSP 的数据宽度决定了它所代表数据的精度。速度也是要考虑的问题之一，速度不仅指一秒钟包含多少个

机器周期，还包括每个周期能够执行多少条指令，以及这些指令中的每一条能够完成多少工作。一种评估 DSP 最低要求的办法是：估计对每个到来的样本必须执行多少条指令。这个数乘以采样频率就得到了所需每秒指令数的最小值。

一种特定 DSP 提供的某些特定的软、硬件功能会使一种选择优于另一种选择，可用片内存储器的容量也是这样。有时，选择 DSP 的理由是支撑硬件的匹配性很好，尤其是片上集成的 A/D 转换器和 D/A 转换器。对于低级编程语言和高级编程语言，软件开发工具是否高效便捷也是经常要考虑的主要因素；是否能够得到第三方软件也是主要因素之一。而“成本”当然永远都是因素之一。事实上，被选中的 DSP 往往速度快、功能多，也符合资金预算要求。

购买 DSP

购买 DSP 有三种形式：购买内核、购买处理器和购买板级产品。在 DSP 中，内核是指处理器中运行关键任务的部分，它包括数据寄存器、乘法器、算术逻辑单元、地址产生器和程序定序器。一个完整的处理器需要将内核、存储器和外部接口组合起来。尽管内核设计和片内外设设计是分别进行的，但是却被制作在同一片硅片上，这样，处理器就能成为单片集成电路。

假定你要制造蜂窝电话并希望在设计中使用 DSP，你很可能会以处理器的形式购买 DSP，也就是说，你会购买一片包含内核、存储器和其他内部功能的集成电路。为了在产品中使用这片集成电路，你必须设计一个印制电路板，而 DSP 将被焊接到这块电路板上，和其他电子元器件相连。这是使用 DSP 最常见的方式。

假设你服务的公司要制造自己的集成电路。在这种情况下，你可能不会购买整个处理器，而仅购买处理器的内核设计。在完成了相应的授权协议后，你就可以开始制造为特殊用途而定制的芯片了。这样做给了你选择片内存储器数量、数据收发方式和封装形式等方面的自由。在 DSP 市场中，这种定制器件越来越成为重要的一部分。

有几十家公司提供安装了 DSP 的印制电路板。在这些电路板上，有附加存储器、A/D 转换器、D/A 转换器和多个处理器等。虽然有些电路板可用作独立工作的计算机，但大部分电路板配置成了主机(如 PC)插板的形式。制造这种电路板的公司被称为“第三方开发商”。寻找第三方开发商的最好方法是询问你要使用的 DSP 的生产商，查看 DSP 生产商的网站；如果在那里找不到，就给生产商发个电子邮件。生产商非常乐于告诉你：谁在使用他们的产品及如何与他们联系。

趋势

别忘了：DSP 和其他微处理器的界限并不是很清晰。例如，让我们看一看 Intel 公司是如何描述其奔腾处理器新增的“多媒体增强指令集”(MMX)技术的：“为了高效操作和处理视频、音频和图形数据，Intel 工程师新增了 57 条功能强大的指令。这些指令面向的是多媒体操作中经常出现的、高度并行和重复的程序。”

在将来，我们一定会看到更多的数字信号处理功能融合到传统微处理器和微控制器当中。推动这种变革的强大动力之一就是网络和其他多媒体应用。此类应用的发展速度非常快，以至于若干年后的 DSP 很可能也会变成传统微处理器了。

第 18 课　ASP 或 DSP

信号处理既可以使用模拟技术——“模拟信号处理”(ASP)，也可以使用数字技术——“数字信

号处理”(DSP)，还可以使用模拟数字混合技术——“混合信号处理”(MSP)。在有些情况下，非常容易明确选择技术；但在有些情况下，却无法一下子明确选择技术，可能需要考虑次要因素才能做出最终决定。

说到 DSP，它和传统计算机数据分析的区别在于 DSP 在实时进行滤波、FFT 分析与数据压缩等复杂信号处理时的速度和效率。

“混合信号处理”这个词包含“模拟处理和数字处理都是系统的一部分”这样的意思。系统的实现形式既可能是一块印制电路板，也可能是一片集成电路。在这个宽泛的定义之下，ADC 和 DAC 也可视作“混合信号处理器”，因为它们既实现了模拟功能，也实现了数字功能。“超大规模集成电路”(VLSI)技术的新近进展允许在一片芯片内既可实现模拟处理，也可实现复杂的数字处理。DSP 本身的特性就表明：它可以实时完成这些功能。

ASP 对比 DSP

今天，工程师们面临着这样的挑战：为了解决手头的信号处理任务，应怎样选择模拟技术和数字技术的恰当组合。单靠纯数字技术是不能完成处理现实世界模拟信号的任务的，因为所有传感器(话筒、热偶、压力感应器、压电晶体、磁盘驱动器头等)都是模拟的。因此，为了使传感器的输出信号能够进一步地实现信号处理(不论是模拟信号处理，还是数字信号处理)，就需要某种形式的“信号调理电路”。实际上，信号调理电路就是模拟信号处理器，它能够完成乘法(增益)、隔离(测量放大器和隔离放大器)、噪声环境下的信号检测(高共模抑制比的测量放大器)、动态范围压缩(对数放大器、对数数模转换器和增益可编程放大器)和滤波(包括无源滤波器和有源滤波器)等功能。图 6.5 展示了实现信号处理的几种方法，图的最上方是纯模拟方法，接下来是 DSP 的方法。注意：一旦决定使用 DSP 技术，接下来就要决定将 ADC 置于信号通道中的哪个位置。

一般而言，ADC 越靠近实际的传感器，ADC 将要担负的信号调理任务就越重。ADC 复杂性的提高可表现为如下几种形式：提高采样率、加宽动态范围、提高精度、抑制输入噪声、对输入信号进行滤波、增益可编程放大器(PGA)、片内参考电压等。所有这些都增加了 ADC 的功能并简化了系统。有了当前高精度、高采样率的数据转换技术，ADC/DAC 在集成更多调理电路方面就取得了重大进展。例如，在测量领域，人们能够得到具有内置可编程增益放大器的 24 位 ADC，它可以对满量程的 10 mV 桥式整流信号直接进行数字化，而不需要进一步调理。在音频和声频领域，出现了完整的编解码器(CODEC，又称模拟前端)，它拥有数量充足的片内模拟电路，从而最大限度地减少了对外部调理器件的需求。在视频领域，CCD 图像处理等应用也有了模拟前端。

实例分析

为了理解 DSP 的功能，这里给出一个实例。对一个模拟低通滤波器和一个数字低通滤波器进行比较，二者的截止频率都是 1 kHz。这个数字滤波器是通过如图 6.6 所示的典型采样数据系统来实现的。注意，图中有几个隐含的要求：第一，假定 ADC 和 DAC 有足够高的采样率、精度和足够大的动态范围，能够精确地进行信号处理；第二，DSP 必须足够快，能够在采样间隔 $1/f_s$ 内完成全部运算；第三，在 ADC 输入端与 DAC 输出端依然需要模拟滤波器来完成抗混叠和抗镜像功能，但是对其性能的要求不像对数字滤波器那么严格。假定这些条件都能够得到满足，那么下面就对这两种滤波器进行比较。两个滤波器要求的截止频率都是 1 kHz。模拟滤波器用六阶切比雪夫 I 型滤波器(通带有波纹，阻带没有波纹)实现，其频率响应见图 6.7。在实际应用中，这个滤波器很可能是用三个二阶滤波器实现的，每个二阶滤波器需要一个运算放大器、一些电阻器和一些电容器。现代滤波器“计算机辅助设计”(CAD)软件能够比较容易地进行六阶滤波器设计，但“0.5 dB 波纹”的

指标还是需要对元器件进行精心选择和匹配才能实现。

另一方面，图中的 129 抽头 FIR 滤波器仅有 0.002 dB 的通带波纹，而且具有线性相位特性，其滚降特性也陡峭得多。实际上，这个滤波器是无法用模拟技术实现的。数字滤波器的另一个优势是不需要元器件匹配，而且对于漂移不敏感(因为时钟频率是受晶体控制的)。这个 129 抽头滤波器需要进行 129 次乘法累加运算，才能计算出一个输出样本。为了保证实时运行，处理过程必须在采样间隔 $1/f_s$ 内完成。在本例中，采样频率为 10 kSPS，因此可利用的处理时间有 100 μs(假定没有重大额外开销)。多数 DSP 能够在一个指令周期内完成全部的“乘法累加”处理(和其他滤波器所需的功能)。

因此，一个 129 抽头滤波器需要的指令速率大于 129/100 μs=1.3 MIPS。我们可以得到指令速率大大超过 1.3 MIPS 的数字信号处理器，因此，在这个应用中，数字信号处理器肯定不是限制性因素。

在实际应用中，在评价模拟滤波器和数字滤波器(或模拟信号处理和数字信号处理)的时候，肯定还要考虑许多其他因素。为了实现期望的功能和充分利用两种技术各自的优势，多数现代信号处理系统都采用了模拟技术和数字技术相结合的方式。

第 19 课　高保真音频

高保真音响爱好者对声音质量的要求极高，所有其他因素都被视为次要的。假如要用一个词来描述这种心理状态，那就是“过分”。这些系统不是设计得刚好满足人类听觉的需求，而是设计得超越了人类听觉的极限。这是能够确保再现音乐无任何失真的唯一方法。CD 带给世人数字音频的享受。这是音乐领域里的巨大变化，CD 系统的音质远远超过了传统的唱片、磁带。

CD 音频系统

图 7.1 显示了高倍显微镜下的 CD 盘面。盘面的主要部分是闪闪发光的(反射光线)，数字信息的存储形式是激光烧制在盘面上的一组暗坑。信息排列在由外到内的一条螺旋线形轨道上(这一点和唱片是一样的)。当从 CD 外侧向 CD 内侧读取信息时，CD 的转速从 210 r/min 变到 480 r/min，从而保持扫描速率为恒定的 1.2 m/s(唱片转速是固定的，如 33 r/min、45 r/min 或 78 r/min)。播放时，光传感器检测表面是否反射光，并产生相应的数字信息。

从图 7.1 显示的几何尺寸可以看出，CD 每平方微米存储1 位信息(相当于每平方毫米存储100 万位信息)，共存储150 亿位信息。这个尺寸和集成电路制造中的特征尺寸大致相同。光是无法聚焦到小于大约 1/2 波长(或 0.3 μm)的，这是光的特性之一。由于集成电路和光盘都是使用光学方法制造的，因此在 0.3 μm 以下光的模糊性限制了可用特征尺寸的大小。

图 7.2 是一个典型 CD 播放系统的组成框图。原始数据率为 430 万位每秒，相当于每 0.28 μm 轨道长度存储 1 位。然而，这和指定的 CD 几何尺寸发生了矛盾。每个凹坑的长度不准小于 0.8 μm，也不准大于 3.5 μm。换言之，每个二进制数“1”必须是 3～13 个“1”中的一部分。这样做的好处是降低了光学读取造成的错误率。但是，如何让二进制数据遵循这种奇怪的安排？

答案是使用一种称作“8-14 调制”(EFM)的编码方法。这种方法不是直接将 1 字节的数据存储到光盘上，而是使用查找表将这个 8 位数据转换为 14 位数据。这种 14 位数据具备所需的特性而被存储在光盘上。播放时，从光盘读取的 14 位二进制数据通过反向搜索 EFM 查找表找到对应的 8

位数据。除 EFM 外，数据以二级“里德-所罗门编码”方式进行编码。这种编码方法对左、右声道和检错、纠错数据进行组合。对播放时检测到的错误，可以使用编码机制中的冗余数据进行纠正，也可以用邻近样本插值的方法去除，还可以用“样本置零”的方法使之“静音”。这些编码机制使得数据率增至原来的 3 倍——即 1.4 Mbps 立体声音频信号将以 4.3 Mbps 的数据率存储在光盘上。

在完成解码和纠错后，音频信号将用 44.1 kHz 采样率的 16 位样本表示。在最简系统中，信号通过 16 位“数模转换器”(DAC)，之后由低通模拟滤波器进行滤波。然而，这样做就需要高性能的模拟器件去实现如下滤波功能：让 20 kHz 以下的频率成分通过，而阻止 22.05 kHz(采样率的一半)以上的频率成分通过。“多采样率”技术是更常用的方法——在 DAC 前将数据转换为更高的采样率。常用的倍增因子是 4，即将 44.1 kHz 转换至 176.4 kHz。这就是所谓的“插值”，这个过程可以理解为两个处理步骤(尽管实际上不可能用这种方式实现)。第一步，在原有数据的每两个样本之间，插入 3 个零值样本，从而产生更高的采样率。在频域，这样做会产生将 0～22.05 kHz 频谱复制三次的作用；这 3 个频谱副本分别位于 22.05～44.1 kHz、44.1～66.15 kHz 和 66.15～88.2 kHz。第二步，用一个高效数字滤波器去掉这些新增的频率分量。

提高采样率就减小了采样间隔，DAC 生成的信号就更平滑了。这个信号中依然包含着 20 Hz～20 kHz的频率成分，而奈奎斯特频率已经提高至原来的 4 倍。这意味着模拟滤波器仅需使 20 kHz 以下的频率成分通过，而阻止 88.2 kHz 以上的频率成分。通常，一个三阶贝塞尔滤波器就可完成这个功能。

因为样本数是原来的 4 倍，所以代表样本的位数可以由 16 位降至 14 位，但却不会降低声音质量。用于补偿 DAC 零阶保持器的“抽样函数校正”既可以是模拟滤波器的一部分，也可以是数字滤波器的一部分。

立体声系统

多于一个声道的音频系统称为“立体声系统”(stereo 源于表示“立体”“三维”的希腊单词)。多个声道从不同方向向听众发送声音，这样能更准确地重现原来的音乐。单声道系统播放出来的音乐往往听上去生硬而又平淡。相比之下，一个好的立体声音响系统会使听者感到演奏家们距离自己似乎只有几步之遥。从 20 世纪 60 年代以来，高保真音乐就已开始使用双声道(左声道和右声道)，而电影也已开始使用四声道(左声道、右声道、中央声道、环绕声道)。在早期的立体声录音中，往往只能在一个声道听到歌唱家的声音。很快，这就发展成为更复杂的“混音”技术，录音棚中来自多个麦克风的声音被合成到两个声道中。“混音”是一种旨在为听者提供“现场感”的技术。

电影中的四声道立体声称作“杜比立体声”，其用于家庭的版本称作“杜比环绕 Pro Logic”(见图 7.3)。四声道声音被编码成标准的左、右声道，这样常规的双声道立体声系统也可播放音乐。杜比解码器用于在播放时重现四声道声音。在电影或电视荧屏两侧的扬声器中传出的两个声道的声音与常规双声道声音是很相似的。产生中央声道的扬声器常被安置在荧屏的正上方(或者正下方)。目的是不论观众或听众的座位如何，重现语音和其他视觉相关声音都固定在荧屏中央。环绕扬声器放置在听众的左、右，而在大型礼堂中可能有多达 20 个环绕声道；环绕声道仅包含中间频率(如 100 Hz～7 kHz)，而且有 15～30 ms 的延迟。这个延迟使听众认为声音来自荧屏而不是来自两侧。也就是说，听众首先听到来自荧屏的声音，然后听到来自两侧经过延迟的声音。听众的大脑会把延迟信号当作墙的反射，而将其忽略。

第 20 课　MP3 音频压缩

CD 音质的音频信号需要 1.411 Mbps 传输带宽。如果要使网络传输成为现实，显然需要进行大幅度压缩。各种各样的音频压缩算法已经开发出来，MPEG 音频算法或许是最流行的一种。MPEG 算法分为三层，其中 MP3(MPEG 音频第三层)最有效、最著名。在互联网上，人们可以获得大量 MP3 格式的音乐。不过，并非所有这些音乐都是合法取得的，因此，就出现了大量由艺术家和版权所有人提起的诉讼。

MP3 是 MPEG 视频压缩标准中的音频部分。音频压缩可以通过两种方式实现。在波形编码中，采用傅里叶变换将波形信号变换为频率分量。各分量幅度采用最小方式进行编码，目的是用尽可能少的位数在另一端准确重现波形。

另一种方法是知觉编码。这种方法利用了人类听觉系统的缺陷，采用让人耳听不出来差别的方式对信号进行编码，尽管在示波器上观看重放波形差别很大。知觉编码技术是建立在心理声学基础之上的。心理声学研究的是人类感知声音的方式。MP3 建立在知觉编码基础之上。

知觉编码的一个关键特性是：一些声音可以掩蔽另一些声音。想象一下：在温暖的夏日，你正在直播一场长笛演奏会。突然，附近有一组工人开动了手提钻，并开始切割街道路面。此时谁也听不见长笛的声音了，它的声音被手提钻的声音掩盖了。从传输的角度看，现在只要对手提钻所在频率进行编码就足够了，因为听众再也无法听到长笛声。这种现象称作“频率掩蔽”——某个频率上响度较大的声音能够掩盖另一个频率上响度较小的声音；假如响度大的声音不存在，这个响度小的声音本来是可以听得到的。事实上，即使在手提钻停止工作后的一小段时间内，还是听不到长笛的声音。因为手提钻开始工作的时候，人耳调低了其增益，而将人耳增益再次调高需要一段时间。此效应称作“暂时掩蔽”。

为了使这些效应更加量化，想象一下实验 1。在安静的房间里，一个人将耳机连至计算机的声卡上，计算机产生一个小功率 100 Hz 纯净正弦波，正弦波的功率在缓慢增加。这个人被告知：当他听到这个音的时候，就敲击一下键。计算机记录着当前的功率电平，然后在 200 Hz、300 Hz 和其他频率上重复这个实验，直到人耳听觉的极限。在对多个实验结果进行平均之后，就得到了一张和图 7.4(a)相像的有关“具备多大功率的单音才能被听到”的对数-对数图。

根据该曲线，可以直接得知完全没必要对功率低于可听门限的频率成分进行编码。例如，在图 7.4(a)中，如果 100 Hz 频率信号的功率为 20 dB，那么就可以将这个信号从输出信号中略去，却不会出现可察觉的音质降低，因为 100 Hz 上 20 dB 的功率在可听电平之下。

现在考察一下实验 2。计算机再次运行实验 1，但这次使用一个等幅正弦波(如 150 Hz)叠加到测试频率上。我们发现频率位于 150 Hz 附近的信号的可听门限提高了，见图 7.4(b)。

通过这次新观测可以得出：通过跟踪哪些信号会被邻近频带的更强信号屏蔽，我们就可以在编码信号中忽略更多的频率成分，从而节省数据位数。在图 7.4 中，输出完全可以忽略 125 Hz 信号，而没有人能听出差别；甚至某个频带上的一个强信号消失了，但由于“暂时掩蔽”特性，在接下来的一段(人耳)恢复期内也可以忽略被屏蔽的频率。MP3 算法的核心就是利用傅里叶变换获得声音在每个频率上的功率，然后只输出那些不被屏蔽的频率，并用尽可能少的位数对其进行编码。

有了这些背景知识，我们现在可以看一下 MP3 编码是如何进行的。音频压缩使用 32 kHz、44.1 kHz或 48 kHz 对波形进行采样。采样既可以是单声道的，也可以是双声道的，并且可以选用

如下配置之一：

1. 单声道(一路输入信号流)。

2. 双单声道(如一路英语音轨和一路日语音轨)。

3. 非联合立体声(各声道分别压缩)。

4. 联合立体声(充分利用声道间冗余)。

首先，要选择输出数据率。MP3 能将一张立体声摇滚乐 CD 压缩至 96 kbps，而几乎没有可察觉的音质下降，即便是摇滚乐爱好者也听不出音质下降。对于钢琴音乐会而言，至少需要 128 kbps。两个数据率的不同，源于摇滚乐的“信噪比”远远高于钢琴音乐会(当然，这是在工程意义上讲的)。也可以选择更低的输出数据率，但在音质上会出现一些下降。

其次，对样本以 1152(约 26 ms)为一组进行处理。每组样本首先通过 32 个数字滤波器，从而得到 32 个频带。同时，输入信号进入心理声学模型以决定被屏蔽的频率。下一步，对 32 频带中的每个频带进一步变换以得到更好的频谱分辨率。

再次，将可用的位数分配给每个频带，谱功率大的“未屏蔽”频带分配到较多的位数，谱功率小的“未屏蔽”频带分配到较少的位数，而完全被屏蔽的频带不分配位数。

最后，使用哈夫曼编码方法对这些数据进行编码。哈夫曼编码将短码字分配给出现频繁的数据，而将长码字分配给出现不频繁的数据。

其实，相关内容不止这些。还有各种不同的技术用来消除噪声、抗混叠和挖掘声道间冗余，但是这些内容已超出了本文范围。

第 21 课　移动电话系统(Ⅲ)

IMT-2000

1992 年，“国际电联”(ITU)公布了一份称作“IMT-2000”的第三代移动电话蓝图。其中，IMT 代表“国际移动通信”，而数字 2000 有以下三个含义：(1)是指开通这项业务的年份；(2)是指该业务的工作频段(以 MHz 为单位)；(3)是指该业务将拥有的带宽(以 kHz 为单位)。

可是，到了 2000 年，什么也没有实现。ITU 建议各国政府预留 2 GHz 附近的频谱，以便设备能够在国家间实现无缝漫游。后来，人们认识到：对于移动性太强的用户，2 Mbps 是不现实的(因为很难实现快速的越区切换)。比较现实的做法是将 2 Mbps 用于室内用户(这将会和 ADSL 直接交锋)，384 kbps 用于步行用户，144 kbps 用于车载用户。

IMT-2000 网络提供给用户的基本业务有：

1. 高质量的话音传输。

2. 消息(取代电子邮件、传真、短信息服务、聊天等)。

3. 多媒体(播放音乐，观看视频、电影和电视等)。

4. 互联网访问(网上冲浪，包括含音、视频的网页)。

IMT-2000 的附加业务有视频会议、遥现、多人游戏和移动商务。不仅如此，还可以在世界范围内瞬时地、高质量地提供所有这些业务。在没有地面网的地方，可以通过卫星实现自动连接。

ITU 提出的 IMT-2000 是一种全球适用的技术。生产商可以生产世界任何地方均可销售、使用的设备(就像 CD 播放器和计算机一样，而不像现在的移动电话和电视)。统一的技术也将简化网络

运营商的工作，并可以鼓励更多的人使用。像 Betamax 和 VHS 在录像机问世之初时的格式之争，对商业并没有什么好处。

W-CDMA 与 CDMA 2000

业界提出的 3G 方案不止一个，经过筛选，最终得到两个方案。一个是爱立信公司(Ericsson)提出的W-CDMA。W-CDMA 采用直序扩频技术，工作带宽为 5 MHz。虽然 W-CDMA 并不向下兼容 GSM，但却可以和 GSM 网络一同工作。W-CDMA 具备以下特性：当通话者从 W-CDMA 小区进入 GSM 小区时，通话不会中断。W-CDMA 得到欧盟的大力推介，并称之为“通用移动电信系统”(UMTS)。

另一个 3G 方案是高通公司(Qualcomm)提出的 CDMA 2000。CDMA 2000 也采用直序扩频技术。它基本上就是 IS-95 的扩展，并向下兼容 IS-95。CDMA 2000 也使用 5 MHz 带宽，但它并没有设计成和 GSM 一同工作，因此无法将通话切换到 GSM 小区。此外，CDMA 2000 和 W-CDMA 在码片速率、帧长、所用频谱、时间同步方式这些方面也存在差异。

假如爱立信公司和高通公司的工程师们被请到一起来，并告知完成某个共同设计，他们很可能会这样做。毕竟，这两个系统的基本原理都是 5 MHz 信道上的 CDMA，而且没有人愿意追求码片速率的不同。但问题的真正难点并不是工程技术。欧洲想要能和 GSM 一同工作的系统，而美国想要能和美国广泛使用的系统(IS-95)相兼容的系统。双方也支持各自的本地公司(爱立信公司总部在欧洲的瑞典，而高通公司在美国的加利福尼亚州)。最终结果是：爱立信公司和高通公司在各自的 CDMA 专利权问题上卷入了一起又一起的官司。

1999 年 3 月，两个公司的官司解决了——爱立信公司同意购买高通公司的基础设施。双方也达成了“同意建立单一 3G 标准”的协议。不过，该标准包含了多个不兼容可选项。在很大程度上，这么做的目的只是掩饰双方在技术上的分歧。

EDGE 与 GPRS

在等待 3G 之争停止的同时，有的运营商向 3G 的方向迈出了谨慎的一小步(有时被称作2.5G)。“增强数据率 GSM”(EDGE)就是这样的系统，EDGE 在 GSM 的基础上提高了每波特包含的比特数。由此带来的问题是：每波特比特数越多，意味着每波特误码率越高。因此，在调制和纠错方面，EDGE 采取了与 GSM 不同的 9 项机制，这种不同体现为：将带宽的多少用于纠正高速传输引起的误码。

“通用分组无线业务”(GPRS)是另一个 2.5G 方案。GPRS 是一种建立在 D-AMPS 或 GSM 上的分组网络。它允许移动站在运行话音系统的小区中发送、接收 IP 分组。当 GPRS 运行时，要将一些频率上的若干时隙预留用于分组传输。时隙的数量和位置可由基站进行动态管理，这取决于小区内话音业务量和数据业务量之比。

可用时隙被分成几个不同用途的逻辑信道，基站决定逻辑信道和时隙之间的映射关系。一个逻辑信道用于将分组从基站下载到某个移动站，而每个分组都指明了自己要去谁那里。为了发送一个 IP 分组，移动站向基站请求一个或几个时隙。如果这个请求正确抵达基站，基站将声明分配给用于发送该分组的频率和时隙。一旦分组抵达基站，它就通过有线连接被转交给互联网。

3G、4G 和 5G

2001 年 10 月，第一个 3G 商用网络在日本投入运营。欧洲的 3G 商用网络开始于 2001 年 12 月。韩国的 3G 商用网络于 2002 年 1 月启动……在世界范围内，3G 技术逐步被采纳了。

2008 年 3 月,“国际电联无线通信分部”(ITU-R)为 4G 制定了标准,称为“高级国际移动通信”(IMT-Advanced)。其中,对乘坐火车、汽车的快移动用户而言,4G 通信业务最高速率可达 100 Mbps;对静立者、步行者等慢移动用户而言,4G 通信业务最高速率可达 1 Gbps。

2012 年年初,ITU－R 着手规划“面向 2020 年及以后的国际移动通信”,为世界范围出现的 5G 研究活动搭建平台。

2015 年 3 月,“下一代移动网络”(NGMN)关于 5G 的白皮书发布了,其中是这么描述 5G 的:“5G 是一个端对端的生态系统,能够实现一个全移动、全互联的社会。5G 预见到 4G 演进,并将其融入到 5G 革命之中。”

第 22 课　数字图像基础

数字图像分辨率

一幅数字图像是由多行多列像素组成的。对灰度图像而言，每个像素被赋予一个代表其灰度级的数字。一幅图像拥有的像素越多，可用灰度级越多，这幅图像的分辨率就越高。图 8.1(a)是一幅 8 bit 数字图像，其可用灰度级为 $2^8=256$,行列数为 808×562。

直方图

如图 8.1(b)所示，一幅数字图像的灰度级可以用直方图来概括。直方图以柱状图的形式反映了在图像中出现的每个灰度级对应的像素数量。当一幅图像只用了可用灰度级的一小部分时，可以使用“直方图均衡”的方法将灰度级的使用扩展到整个可用范围。直方图均衡是指对灰度分布重新进行安排，提高了图像对比度。

图像的加与减

数字图像可以逐像素地相加与相减。两幅图像相加可以把两组物体合到一幅图像之中。此外，对同一场景的几幅含噪图像进行平均可以降低噪声的影响。而图像相减可用于去掉图像中不需要的背景。将拍摄时间点前后相继的两幅照片相减，会显示出两幅照片中出现变化的地方。

当对两幅图像进行相加、减时，结果矩阵往往包含不正确的灰度值。例如，在一幅 8 bit 图像中，某个像素灰度级为 129；在另一幅 8 bit 图像中，相应像素的灰度级为 201；那么，和像素的灰度级是 330。这个数值超出了 8 bit 图像灰度级的有效范围——8 bit 图像能够包含的灰度值是 0～255。这两幅图像相减的时候，差像素为-72，这个数值也超出了有效范围。鉴于上述情况，大多数算术运算之后都要进行数据缩放。可以按照下面的方法将数据缩放到范围[0，GSL_{max}]。(公式略)

扭曲和变态

扭曲和变态是应用于娱乐、医学成像领域中的数字图像技术。扭曲是对图像中的物体进行拉伸或重塑，而变态是将一幅图像变换成另外一幅图像。这种变换的实现过程如下：在源图像中标记控制点、控制线或控制三角形，并为其选择在目标图像中的新位置。然后，通过对控制元素的位置、颜色进行平滑变换，实现两幅图像间的变换。那些没有明确标记为控制元素的像素的位置、颜色由离其最近的控制元素的位置、颜色来确定。

图像滤波

通过使用卷积算子进行二维卷积运算，可实现对数字图像的滤波。一幅 $N \times N$ 图像被 $M \times M$ 卷积算子滤波时，边缘效应会使图像的每一边都失去 $(M-1)/2$ 行和列的像素。低通滤波器会使图像变得模糊，而高通滤波器突出了图像的灰度锐变，边缘滤波器对图像边缘进行定位。

膨胀和腐蚀

膨胀是向图像中的所有物体添加一层像素，腐蚀是从所有物体上去掉一层像素。在膨胀之后进行腐蚀，就可以填补(由边缘检测识别产生的)破碎边界中的缝隙。相反，在腐蚀之后进行膨胀，就可以去除图像中的噪声点。要实现可靠的物体边界识别，首先要检测图像边缘。利用边界信息，可以计算出周长、面积等形状特征，这些特征可用于物体分类。

图像频谱

“快速傅里叶变换”(FFT)可以用来分析数字图像的频谱。二维频谱包括幅度谱和相位谱，这和一维频谱是相同的。相位谱携带着图像中目标位置的信息。当幅度全部设置为 1 时，仅靠相位谱也能显示出原始图像的大概。而当相位全部设置为零时，单靠幅度谱却什么都显示不出来。

图像频谱是“计算机层析成像”(CT)与“磁共振成像”(MRI)这两种扫描成像技术的基础。CT 技术使用 X 射线从不同方向对目标的某个平面进行扫描。MRI 扫描取决于物体在变化磁场中表现出来的磁学特性。这两种扫描方式都允许对三维物体进行非侵入性探测。

图像压缩

部分原因出于互联网，数字图像比以往更频繁地从一地传送到另一地。为了节省时间和带宽(空间)，(和其他文件一样)经常需要在传输前对图像文件进行压缩。用无损压缩方法压缩了文件，却没有损失原始文件的任何信息——重建文件和原始文件是完全相同的。有损压缩是指原始文件中的有些信息永远丢失了；不过，重建文件和原始文件非常相近。“压缩比”是指原始文件大小和压缩文件大小之比。

“行程编码”(RLE)是一种简单的压缩方法。行程编码将一个数字超过三次的重复次数附加在这个数字后面。哈夫曼编码是另外一种编码方法。这种方法用较短的代码来表示经常出现的信号，而使用较长的代码来代表不经常出现的信号。

JPEG 是一种极为常见的图像压缩方法。JPEG 使用“离散余弦变换”(DCT)将一幅图像 8×8 小块中的大部分信息集中到少数几个重要系数上。然后，JPEG 使用行程编码和哈夫曼编码对图像完成进一步压缩。

第 23 课　数 码 相 机

图像传感器

数码相机不使用胶卷，而使用“图像传感器”——将光转化为电荷。多数数码相机采用 CCD(电

荷耦合器件)传感器。有些数码相机采用 CMOS(互补金属氧化物半导体)传感器。它们都将光转换为电子。简言之,图像传感器就像一个“二维阵列”,该阵列由几千个或者几百万个“微型太阳能电池”组成。

当光转换为电子后,图像传感器就对阵列中每个“微型太阳能电池”的值(积聚的电荷)进行“读取”——CCD 传感器与 CMOS 传感器的差异就发生在这里。

- CCD 传感器:它把电荷传输到芯片的一角,并在那里进行读取。之后,ADC(模数转换器)对每个感光点的电荷量进行测量,并将测量值转换为二进制形式,从而完成了每个像素值的数字化。
- CMOS 传感器:它在每个像素处使用若干个晶体管对电荷进行放大,并使用导线对电荷进行移动。

这种差异造成了两类图像传感器的诸多优缺点:

- CCD 传感器可生成“高质量、低噪声”图像。CMOS 传感器往往易受噪声影响。
- CMOS 传感器的功耗通常很小。而 CCD 传感器所采用的工艺则功耗很大。CCD 传感器的功耗是等效 CMOS 传感器的 100 倍。
- CCD 传感器量产时间更长,因而更加成熟——像素质量更高、像素数量更多。

数码相机分辨率

数码相机能够采集到的“细节”的数量称为“分辨率”,其度量单位是“像素”。数码相机拥有的“像素”越多,它能够采集到的“细节”就越多;照片就能够放得更大,而不会变得模糊(“颗粒状”)。

数码相机的典型分辨率有如下几种:

- **640 × 480 像素**:该分辨率见于大多数低端“真实”相机,非常适合通过电子邮件发送照片或者在网站上发布照片。
- **1216 × 912 像素**:该尺寸图像已达“百万像素”,适合打印照片。
- **1600 × 1200 像素**:像素总量接近 200 万,属于“高分辨率”。该分辨率拍摄的照片,以 4 × 5 英寸打印出来,其画质与从照相馆获得的画质相同。
- **2240 × 1680 像素**:该分辨率出现在 400 万像素的相机上,照片可以打印得更大。最大可达 16 × 20 英寸,画质不错。
- **4064 × 2704 像素**:顶级数码相机以此分辨率拍摄照片,可获得 1110 万像素。在此设置下,打印 13.5 × 9 英寸的照片,画质不存在任何损失。

高端消费类相机的像素总量可超过 1200 万。有些专业相机的像素总量可超过 1600 万,大画幅相机的像素总量可达 2000 万。根据惠普公司(Hewlett Packard)的估计,35 mm 胶片的画质大约相当于 2000 万像素。

存储数码照片

数码相机使用几种存储系统。这些存储系统就像“可重复使用的数字胶片”,它们使用读卡器将数据传输到计算机。许多存储系统都含有固定闪存或可移动闪存。数码相机制造商往往会开发其专有的闪存设备——如 SM 卡、CF 卡、记忆棒。

无论使用哪种存储设备,数码相机都需要大量空间来存储照片。数码相机通常采用 TIFF(未压缩格式)、JPEG(压缩格式)这两种格式存储照片。但是,有些数码相机采用的是 RAW 格式。数码相机大多采用 JPEG 文件格式来存储照片,有时还会提供画质设置(中等质量或高质量)。表 8.1 是不同图像尺寸所对应的文件大小。

表 8.1　图像文件格式与文件大小

图像尺寸	TIFF(未压缩)	JPEG(高质量)	JPEG(中等质量)
640×480	1.0 MB	300 KB	90 KB
800×600	1.5 MB	500 KB	130 KB
1024×768	2.5 MB	800 KB	200 KB
1600×1200	6.0 MB	1.7 MB	420 KB

为了充分利用存储空间，几乎所有的数码相机都对文件进行某种“数据压缩”。“压缩”之所以成为可能，源于数字图像的两个特征——重复性、不相关性。

拍照过程

为了拍摄照片，数码相机要采取若干步骤。CCD 数码相机的完整拍照过程如下：

1. 摄影者将相机瞄准拍摄对象，并根据对象远近调整光学变焦镜头。
2. 摄影者轻轻按下快门。
3. 相机自动对焦于拍摄对象，并获取现场光线强度。
4. 相机设置光圈和快门速度，以获得最佳曝光效果。
5. 摄影者完全按下快门。
6. 相机对 CCD 进行复位，CCD 曝光——积聚电荷直至快门关闭。
7. ADC 测量电荷，并生成“代表像素电荷值”的数字信号。
8. 为了获得自然色彩，处理器对“源自不同像素的数据”进行插值处理。
9. 按照预先设定的压缩级别，处理器对数据进行压缩。
10. 图像信息保存于某种存储设备(很可能是闪存卡)。

第 24 课　电视视频信号

图 8.2 是在示波器上看到的电视信号。这个信号被称为“复合视频信号”，即在图像信号中加入了“垂直同步脉冲”和“水平同步脉冲”。这些脉冲用在电视接收机的“同步垂直电路”和“水平偏转电路”中，用来配合正在显示的视频信号。标准视频信号在一秒(1 s)内包含 30 幅完整图像，每幅图像通常称作一帧。视频工程师会说每帧包含 525 线。“线”也是电视术语，而程序员称之为“行”。525 这个数字有点名不副实，因为只有 480～486 线包含视频信息，其余 39～45 线保留用于视频信号同步。

标准电视信号采用交错形式来减少画面的闪烁。也就是说，首先发送一帧画面全部的奇数线，然后发送偶数线。奇数线组被称作“奇数场”，而偶数线组被称作“偶数场”。每帧由两场组成，因此视频信号每秒发送 60 场。每场都是从一列持续 1.3 ms 的复杂垂直同步脉冲开始的，其后就是视频信号(偶数线或者奇数线)。每条线持续 63.5 ms，包括 10.2 ms 的水平同步脉冲，将一条线和另一条线分开。在每条线内部，模拟电压对应着图像的灰度，越亮的值表示沿黑色方向离开同步脉冲越远。因此，同步脉冲就在黑色范围之外。用视频术语来讲，就是同步脉冲的电压值比黑色电压值还低。

在视频信号中，用于 A/D 转换的硬件称作帧接收器。帧接收器通常是一张电子卡，它插入计算机并通过同轴电缆连接到摄像机上。根据软件指令，帧接收器等候由垂直同步脉冲指示的下一帧的开始。在接下来的两场中，每条视频信号线要被采样多次，采样率的典型值为 512 个样本/线、640 个样本/线、720 个样本/线和 8 位/样本。这些样本就被存入存储器而成为数字图像中的某一行数据。

这种数字图像的采集方式导致在垂直方向和水平方向上存在显著区别。数字图像中的一行对应着视频信号中的一条线，也对应着 CCD 中的一行势阱。而列就没有这么简单的对应关系。在 CCD 中，每行包含 400～800 个势阱(列)，具体数目是由所用的特定器件决定的。从 CCD 读取一行势阱时，视频信号线经过滤波后形成如图 8.2 所示的平滑模拟信号。换言之，视频信号并不取决于 CCD 有多少列。水平方向分辨率是受模拟信号可变速度限制的。在通常情况下，彩色电视信号设置为 3.2 MHz，上升时间约为 100 ns，即 53.2 ms 视频线的 1/500。

视频信号在帧接收器中被数字化之后，就被转换回列。然而，数字化图像中的列和 CCD 中的列是不相关的。数字化图像中列的数目仅由帧接收器对每条视频线的采样次数决定。例如，某个 CCD 的每行可能包含 800 个势阱，而数字化图像的每行可能仅含 512 像素(512 列)。

此外，还有一个原因使得数字化图像中的列数很重要。标准电视图像的宽高比为 4∶3，即宽度稍大于高度，而电影图像的宽高比更宽(25∶9)。科学应用中 CCD 的宽高比通常为 1∶1，即一个完美的正方形。不论是哪种情形，CCD 的宽高比已经由电极的放置情况决定了，不能改变。然而，数字化图像的宽高比取决于每条线的采样数。在视频监视器上显示数字化图像或者对图像进行硬拷贝时，这就成了一个问题。假如不能正确重现宽高比，图像看上去就像在水平方向或者垂直方向受到了挤压。

这里描述的 525 线视频信号称为“国家电视系统委员会”(NTSC)制。NTSC 制是在 1952 年定义的标准。NTSC 系统用于美国和日本。在欧洲，存在两个相似的标准——“逐行倒相(PAL)制”和“顺序彩色与存储(SECAM)制”。PAL 制和 SECAM 制的基本概念是相同的，只是具体数字不同。PAL 制和 SECAM 制都采用每秒 25 个交错帧，每帧 625 线。和 NTSC 制一样，有些线在垂直同步时出现，使得约 576 线携带图像信息。其他不易理解的区别和色彩、声音添加到信号的方式有关。

传送彩色电视的最直接方法可能就是使用三路独立的模拟信号，分别代表人眼可以检测到的红、绿、蓝三种颜色。但令人遗憾的是，电视在其历史发展过程中并没有采用这种简单的机制。人们要求彩色电视信号做到：在不做任何修改的前提下，黑白电视机能够继续使用。这是通过保留亮度信号，并增加一路独立的用于传输色彩信息的信号来实现的。在视频术语中，“亮度”被称作“辉度信号”，而“色彩”被称作“色度信号”。3.58 MHz 载波上的色度信号被叠加到黑白视频信号上。声音信号以相同的方式加到黑白视频信号上，声音信号的载波是 4.5 MHz。电视接收机将这三个信号分离，分别进行处理，然后在终端显示器中再将其合成。

第 25 课　选择合适的微处理器内核

目前，在市场上广泛采用的 32 位微处理器内核(简称内核)大致有 5 种：Intel x86、PowerPC、MIPS、SuperH 和 ARM。当然，还存在着很多应用不太普遍的或专有性不太强的微处理器结构。其中，许多结构是和激光打印机、DVD 播放机等特定应用联系在一起的。

基于高端 x86 系列微处理器的结构有如下明显优势：

- 可以使用几乎任何一种与 PC 兼容的操作系统和免费的软件开发工具。
- 操作系统的安装非常简单。在多数情况下，“自动安装器”将检测系统的硬件构成，并自动安装合适的内核、驱动程序等。这一点和嵌入式系统的规矩是不同的——(嵌入式系统)用户需要查看电路板，自己完成硬件配置的设置，(很可能要使用交叉编译器)生成系统内核和外部硬件的驱动程序。
- 对于几乎任何一种能想到的功能，x86 微处理器可以很容易地和数以千计的周边组件进行接口。这些组件是为消费市场生产的，巨大的生产量和激烈的价格竞争致使其价格便宜、容易获得。
- 所有你想添加到系统中的硬件的驱动程序几乎都有(大多数现成的操作系统中都包含这些程序)。
- 可以得到多种不同形式的、包含多种可能的外设组合的高集成度主板。
- 因缺乏支撑部件或因用户需求变化而进行的向大致相同硬件平台的移植相对要容易些。在许多情况下，这种移植仅需要对操作系统进行重新编译和安装，并为复制准备好一个新的主磁盘映像。

关于 x86 系列器件的优点已经叙述过了，而其不足之处也必须指出：

- 和同样性能的 RISC 处理器比较，x86 器件非常昂贵。这就会影响到设备的商业化能力。
- 因为能够称得上是真正“片上系统”的 x86 器件比较少，所以在处理器之外很可能还需要一些外部硬件。为了获得某种特殊功能，往往需要添加一个多功能元件——因为得不到所需的单功能分立元件。这就提高了系统的复杂性和材料成本。
- x86 器件在功耗、散热、尺寸方面都存在明显的劣势。
- 现代 x86 器件及其支撑芯片都是高密度封装形式的高速器件。实际上，不可能以这种器件为基础手工设计样机，除非你想在设备上投入数万美元，但也不得不将其中某些装配工作承包出去。
- PC 外设集成电路的生产周期往往很短，12～18 个月是很平常的。因此，供货可能会成为一个问题。
- 用于裸机冷启动的代码往往很复杂，因为它必须对若干层(主板 BIOS、扩展卡 BIOS 及操作系统的各层)进行替换。CPU 结构也很复杂。
- 需要指出，在商用 x86 单板计算机上，基于 JTAG 的硬件调试系统或者其他硬件调试系统是不存在的。

当设备数量很少或者当样机需要很多硬件功能而不想花费时间进行硬件调试时，建议选用 x86 作为硬件平台。当你既向早期用户供货又在开发价格便宜的第二轮定制硬件设计时，x86 也是不错的选择。虽然还有其他适合选用 x86 的情况，但上述情况是主要的两种。

我们接着谈 RISC 平台。对于许多应用而言，MIPS、SuperH 和 PowerPC 都是很好的选择。其中，SuperH 是一个庞大的、包含各种各样有用器件的微处理器系列，而 MIPS 则是在第三方专用集成电路和专用标准产品中获得广泛授权的微处理器内核。PowerPC 主要在高性能应用中使用。对所有这些器件在不同工程项目中的表现进行评价后，我发现用很少的资金去开发这些微处理器是很困难的——评估硬件通常很贵，而且那些表现不出将会大量订货的购买者也不会立即得到这些器件。然而，在可预见的未来，所有这些处理器内核很可能都会存在下去，并且也会得到很好的技术支持。因此，只要你能够得到开发系统和器件，这些器件都是可行的选择。

至少对于 SuperH 和 MIPS，设计基于该类处理器样机最省钱的途径是对现有某种硬件进行重新定位。对于 PowerPC，我建议购买一块含有感兴趣芯片的单板工控计算机，但需要注意这样做

也许很贵。PowerPC 不具有同 x86 兼容板一样的大规模市场价格，一块 PowerPC 单板机的价格是 x86 板价格的 2～3 倍。

我主要推荐的 32 位嵌入式平台是 ARM，除非某些 Intel 芯片的特点适合你的应用。该结构具有许多重要的优点(当然，这些优点中的一些也适用于上述其他 RISC 平台)。

- ARM 是一个成熟的、为人熟知的结构。它有翔实可靠的工程应用历史，并且做过许多改进。目前大量的 ARM 授权和大批 ARM 器件的出货可以确保其未来的供货能力。
- ARM 内核比较小，而且具有卓越的“功耗－性能”特性。
- 芯片设计者可以改变许多特性(协处理器、外部总线宽度、存储器管理单元和高速缓存大小等)。这意味着总能够找到一款满足几乎任何一种“性能/尺寸/功率”要求的 ARM 内核。
- 市场上存在大量价格诱人、集成外设丰富的标准器件、定制器件和半定制器件。
- ARM 不仅提供了处理器内核的参考设计，而且提供了许多不同外设的参考设计。因此，在外设控制方面，不同的 ARM 实现方案(甚至来自不同厂商的 ARM 实现方案)往往具有相似性。举个小例子，通常可以不怎么费力就把一个用于串口数据发送的代码从一个 ARM 器件移植到另一个 ARM 器件。
- 部分由于上述因素，业界存在着大量用于 ARM 内核的免费 IP。

人们常说“ARM 就是 32 位的 8051”，就是在说 ARM 是一种“人人皆知、处处在用”的通用 32 位微处理器。这绝非夸张——在嵌入式应用领域中，ARM 的地位就如同台式 PC 领域中的 x86。

第 26 课　嵌入式系统的设计语言

嵌入式系统是一种“看似不是计算机”的计算机，它能低成本、高效率地完成一组任务。典型的嵌入式系统可能需要完成通信、信号处理和用户接口的任务。因为这些任务需要解决的问题不同，所以能够解决所有嵌入式应用问题的通用语言都是很难编写、分析和编译的。开发人员采用了多种不同的语言，每种语言都有最适合的应用领域。

硬件语言

Verilog 和 VHDL 是最常用的硬件描述和建模语言。每种“硬件描述语言”(HDL)都采用了(为高效仿真而忽略设计空闲部分的)离散事件语法对系统进行建模。这两种语言都使用了具有结构化层次的描述系统——将一个系统分成若干模块，系统模块包含了基本元件实例、其他模块和并发进程。此外，每种 HDL 将这些连接都明确罗列出来。

Verilog 提供了更多的适于硬件仿真的基本元件，而 VHDL 的基本元件是像 a＝b＋c 这样的赋值表达式或程序代码。Verilog 增加了晶体管和逻辑门作为基本元件，而且允许用户使用真值表定义新的基本元件。

这两种语言都允许使用编写程序的方法来描述并发进程。在驱动事件被唤醒之前，这些进程处于休眠状态；进程被唤醒后，进行变量读/写，然后终止。进程可能等待一段时间(如 Verilog 中的 #10，VHDL 中的 wait for 10 ns)，也可能等待某个值发生变化(@(a or b)，wait on a，b)，或者等待某个事件的发生(@(posedge clk)，wait on clk ＝'1')。

VHDL 的通信更加规范和灵活。Verilog 通过导线或寄存器进行通信——共享存储器位

置会导致出现竞争状态。VHDL 信号量的行为类似导线，但其判决函数可由用户定义。除被声明为共享变量外，VHDL 变量只在进程内部有效。

Verilog 的类型系统使用四值位向量对硬件建模，用数组对存储器建模。VHDL 不含四值向量，但其类型系统允许用户添加数据类型。此外，用户还可以定义 C 结构体这样的复合数据类型。

总的来说，Verilog 是一种更简洁、更直接面向数字集成电路仿真的语言，而 VHDL 是一种更庞大、更详尽、能处理更多类型仿真和建模任务的语言。

软件语言

软件语言描述的是微处理器要执行的指令序列。大多数语言列出了要执行的指令，这些指令按照和存储器(存储器是存储数据的阵列)通信的顺序执行。

一般而言，每条机器指令只能做“两数相加”这么简单的操作。因此，高级语言的设计目标是能够简明、直观地指定指令。算术表达式是一种典型指令——用机器代码对 ax^2+bx+c 这样的表达式进行编码虽然直接，但冗长，编译器会完成得很好。C 语言提供了这种表达式，而且还提供了循环、条件等控制流结构和递归函数。此外，C++语言提供了类的方法来创建新的数据类型、用于多态代码的模板、用于处理错误的异常机制和公用数据结构的标准库。Java 是更高级的语言，它提供了自动回收功能、线程和用来进行线程同步的监控程序。

汇编语言

汇编语言程序就是一张符号化的、人类可读形式的处理器指令列表，每条指令都包含一种操作(如加法)和一些操作数。例如，add r5，r2，r4 可以表示将寄存器 r2 和 r4 的内容相加的结果写入 r5。汇编语言依序执行这样的算术指令，而跳转指令可以通过改变程序计数器的值来执行条件语句和循环语句。程序计数器中存放着正在执行的指令地址。

某种微处理器的汇编语言是由操作码、寻址模式、寄存器和存储器来定义的。操作码用于区分不同的操作，例如，是加法还是条件转移。寻址模式定义了数据聚集、存储(例如，来自寄存器还是来自特定存储器位置)的方式。寄存器可被视作速度快且易于访问的小容量存储器。

C 语言

一个 C 程序中包含着由算术表达式构成的循环结构或条件结构的函数。C 程序指令按照顺序执行。但是，循环、条件等控制流结构会影响指令的执行顺序。当表达式中出现了函数调用，程序就将“控制权”交给被调用函数。该函数一直运行到输出结果，控制权返回并继续计算调用函数表达式的值。

根据微处理器能够直接操作的数据，C 语言可以推出自己的数据类型。这些数据类型包括有符号整数和无符号整数(从字节到字)、浮点数和指针。还可进一步将这些数据类型组合为数组和结构体(结构体由一些带名称的字段组成)。

C 程序使用三种类型的存储器，在程序编译阶段，程序为全局数据分配空间。堆栈存储自动变量，程序在函数调用和返回数值时分配与释放这些变量。堆为程序提供了可以按照任意顺序重新分配的、大小任意的存储区域。尽管 C 语言是一种国际标准，但许多人还是参考 Kernighan 和 Ritchie 写的书，因为 Ritchie 设计了这种语言。

C＋＋语言

为了适于开发大型程序，C++使用结构化机制对 C 语言进行了扩展。这些机制包括用户定义的数据类型、不同类型代码重用的方法、群对象命名规则、避免(程序片段进行汇编时出现的)名字冲突的命名规则和错误异常处理机制。C++标准库包含了数组、树、字符串等高效多态数据类型，编译器会为这些多态数据类型生成定制的实现。

“类”通过指定其表现形式及可以访问和修改的操作，定义了一种新的数据类型。类可以用继承性对已有类进行扩展和修改。例如，一个“矩形类”可以将长度字段、宽度字段和一种计算面积的方法加到一个“形状类”中。

“模板”是一种能和多种数据类型工作的函数或类。编译器为每次不同的模板应用生成定制的代码。例如，某个程序会对整数、浮点数应用同一个“求最小值”(min)模板。

Java 语言

Sun 公司的 Java 语言集合了 C++，但不兼容 C++。和 C++一样，Java 语言是面向对象的，它提供类和继承性。Java 是比 C++更高级的语言，因为 Java 使用了对象引用、数组、字符串，而不使用指针。Java 的自动回收功能将程序员从存储器管理的工作中解脱出来。

Java 提供并发线程。创建一个线程需要对“线程类”进行扩展、为这些对象创建实例、调用它们的启动方法(以启动一个新的执行对象运行方法的控制线程)。

方法(或模块)的同步使用为每个对象分配的锁来解决(两个或更多进程试图同时访问同一个对象出现的)竞争问题。试图获得另一个线程拥有的锁的线程将停止运行，直至该锁被释放为止。利用这个特点，可以将特定对象的独享访问权授予某个线程。

实时操作系统(RTOS)

许多嵌入式系统使用“实时操作系统”(RTOS)来模拟单片处理器的并发性。RTOS 管理多个正在运行的进程，每个进程都是用 C 语言这样的顺序执行语言编写的。进程完成系统的运算，而 RTOS 对它们进行调度。为了满足时限要求，调度机制将决定哪个进程在何时、以何种顺序执行。

多数 RTOS 使用固定优先级的先占调度机制。在系统设计之时，每个进程就被赋予某个特定优先级(即一个小整数)。在任意时刻，RTOS 都运行具有最高优先级的进程；在运行一段时间之后，该进程就挂起，并等待更多数据的到来。通常，RTOS 使用“率单调分析”来分配优先级，最高优先级被分配给那些需要频繁满足“时限”的进程。

第 27 课　选择实时操作系统

工程师们常用“实时”来描述这种计算问题：“应答迟滞”和“应答错误”同样糟糕。此类问题对时限是有限制的，嵌入式系统常常在这种约束条件下运行。例如，假如控制防抱死刹车装置的嵌入式软件的某个时限没有得到满足，那么车祸就可能发生。所以，对于实时嵌入式系统的设计人员而言，尽可能了解其软、硬件的行为和性能是至关重要的。

实时系统设计人员要花费大量时间来考虑“最坏情况”下的系统性能。他们必须不断地问自己这样的问题：在最坏情况下，“驾驶员踩刹车踏板”和“中断信号抵达处理器”这两个事件的时间差是多少？在最坏情况下，“中断反应时间”是多少？在最坏情况下，软件需要多长时间才能触发刹车系统？仅进行“平均情况分析”或“期望情况分析”是不够的。今天，大多数商用嵌入式操作系统在设计时就都将实时系统考虑在内了。在理想情况下，这些嵌入式操作系统的最坏性能是明确的、可信的。

RTOS 特征

要想获得“实时操作系统”(RTOS)之名，这个操作系统应该是确定的，并要确保最坏情况下的“中断反应时间”和“任务切换时间”。当给定这些特征、系统中的任务和中断的“相对优先级”时，就有可能使用“率单调分析”对软件的最坏情况性能进行分析了。

如果每个系统调用在最坏情况下的执行时间都是可计算的，那么这个操作系统就是“确定的”。如果一个操作系统提供商能够认真看待其产品的实时性能，那么一般都会通过数据手册公布每个系统调用所需时钟周期数的最小值、平均值和最大值。对于不同处理器而言，这些数值可能不尽相同；但是，假如算法在某个处理器上运行时是“确定的”，那么可以合理地推断这个算法在任何其他处理器上运行都是确定的。(当然，实际情况可能会和这个推断不同。)

“中断反应时间”是指从中断信号到达处理器到“中断服务程序”(ISR)开始运行之间的全部时间。当中断出现时，在执行 ISR 之前，处理器必须完成以下几步。首先，处理器必须停止正在执行的指令；然后，必须辨认中断类型，这是由硬件完成的，不会减慢或挂起正在运行的任务；最后，只有当中断使能、CPU 环境设置被保存之后，相关的 ISR 才会启动执行。

当然，假如中断会被禁用(例如，在系统调用中)，那么最坏情况下的中断反应时间将增加，增加量是中断的最长关闭时间。每个操作系统都会在一些地方禁用中断一段时间，因此，就需要知道系统的需求是什么。一个实时工程项目可能需要中断响应时间在 1 μs 之内，而另一个实时工程项目可能仅需要在 100 μs 之内。

操作系统的第三个实时特性是任务切换所需的时间。这个特性之所以重要，是由于它反映了整个系统的开销。例如，假定任何一个任务的平均执行时间是 100 μs，而任务切换时间也是 100 μs，这样的话，整整一半的处理器时间都用在任务切换例程上了！

这里也没有什么特定的数据。实际时间通常是由处理器决定的，因为这取决于应予保护的寄存器数量和在何处进行保护。一定要从操作系统提供商那里得到这些数据，这样，在购买前就不会出现任何意外。

选择过程

考虑到当今工程的时间成本，花几千美元买一个商用实时操作系统是合算的。对于大多数工程项目和资金预算，是可以买到合适的操作系统的。商用操作系统在功能、性能和价格上是一致的。低端商用操作系统仅仅提供基本的“先占式调度器”和几个关键的系统调用。这种操作系统往往价格低廉，源代码无法修改，无须支付版权费。高端操作系统往往包含除基本调度器外的许多功能。这种操作系统都相当昂贵，尽管启动成本在 10000～50000 美元之间，每个 ROM 复制都要付版权费。不过，这个价格往往包含了免费技术支持、技术培训和一套集成开发工具。在这两种极端情况之间，还有一些操作系统价格适中，但不包含源代码，技术支持也会额外收费。大多数商用操作系统属于这一类。

操作系统种类如此繁多，特点各不相同，很难决定哪个操作系统最适合你的工程项目。首先，

要考虑你对处理器、实时性能和经费预算的要求。你无法改变这些指标，因此使用这些指标可以将可选范围缩小一些。然后，和这些操作系统的提供商联系，获取更加详尽的技术信息。

此时，许多人就会根据操作系统和已选交叉编译器、调试器及其他开发工具之间的兼容性做出决定了。不过，你也应该搞清楚：对于你的工程项目，哪些附加功能是很重要的？不论你购买什么附加功能，基本内核大体都是一样的。区别很可能在于：处理器的支持情况、最大和最小存储器要求、有无附加软件模块(如网络协议栈和驱动程序)，以及是否和第三方开发工具兼容。

选择商业操作系统最具说服力的理由是可以利用经过充分验证的内核，这要比使用自己开发内核的可靠性高得多。因此，在选择操作系统提供商时，考察其开发经验如何是很重要的一点。

第28课　信　号　源

完整的测量系统

在进行电子测量的时候，人们首先想到的很可能是一个“采集仪器”——通常是示波器或逻辑分析仪。只有在能够获得某种类型的信号时，这些工具才能进行测量。但是，在许多情形下，这样的信号是不存在的，除非外部提供。

举例来说，一个应力测量放大器不会产生信号，它仅能够将传感器接收到的信号放大；同样，数据总线上的“多路复用器”也不会产生信号，它为来自计数器、寄存器和其他元件的信号指引方向。但是，在将放大器或多路复用器连接到提供信号的线路之前，不可避免地要进行必要的测试。为了使用采集工具测量这些器件的行为，必须在其输入端提供激励信号。

举另外一个例子。为了确保新硬件在整个工作范围内，以及该范围外满足性能设计要求，工程师必须对即将完成的设计进行特性评价。这就是所谓的“边际测试或极限测试”。这项测量工作就需要一个“完整的”解决方案——一个既能进行测量又能产生信号的方案。

用于数字设计特性评价的测量仪器和用于模拟/混合信号设计的测量仪器是不一样的，但两者一定都包含激励信号产生仪器和信号采集仪器。信号源(或信号发生器)和采集仪器是构成一个完整测量解决方案的两大基本组成部分。

什么是信号源？

信号源是用于硬件设计、调试和评价工程中几乎任何一种测量配置的基础。信号源是一种关键的工程工具。对技术人员而言，它是必备的故障诊治助手。它可以产生用来代替汽车点火脉冲、心脏起搏器或导弹陀螺仪的输出信号。在电子测试仪器中，信号源也许是仅次于无处不在的数字万用表的最通用的仪器。

当对元器件、功能模块和子系统进行验证时，电路中很可能需要某种类型的交流激励源，除非你只使用纯直流电路。信号源输出的波形模拟了来自外部世界的信号(如传感器输出信号)。同样，信号源还可用于替代电路设计中尚未完成部分所产生的波形。

有趣的是，信号源不仅提供一个理想波形，它还经常在输出波形中添加已知的、数量和类型均可重复的“畸变”(误差)。由于往往不可能使用电路准确地在所需时间、位置产生可预测畸变，所以信号源的这个功能是极为有用的。“被测单元”(UUT)对畸变信号的响应反映了该单元在正常环境下处理问题的能力。

激励信号可以采取低畸变正弦波、逻辑脉冲、高频无线载波、移动电话发射信号等多种形式。传统上，这些波形是由独立的专用信号源产生的——从超纯净音频正弦波发生器到 GHz 射频信号发生器。尽管市场上存在着多种商业解决方案，但是为了应用到手头的项目中，用户常常要对信号源进行定制设计或修改。设计一个仪器级的信号发生器是非常困难的，而设计辅助设备耗费的时间也将人们的宝贵精力从项目中分散开。

幸运的是，数字采样技术和信号处理技术已经带给人们一种使用一种仪器就可以满足几乎所有类型信号产生需求的方案——任意信号发生器。

数字信号源的类型

数字信号发生器囊括了全部信号产生需求，它可分为“任意波形发生器”(AWG)、“任意函数发生器”(AFG)和“数据/模式发生器”(DG)三大类。每类都有其独特的优点。

- **AWG**：可以产生你能想象到的任何波形——不论是用于测试磁盘驱动器的洛伦兹脉冲整形数据流，还是测试 GSM 手机或 CDMA 手机的复杂调制射频信号。你可以使用多种方法——从数学公式到波形绘制——创建所需的输出信号。
- **AFG**：提供的波形种类一般比较少，但它具备卓越的稳定性，而且能快速响应频率的改变。假如 UUT 需要典型“正弦波和方波”及在两个频率间的瞬时切换能力，那么 AFG 就是正确的选择。AFG 的另外一个优点是价格低，这使其在不需要 AWG 灵活性的应用中非常具有吸引力。
- **DG**：能够满足需要具有特定内容和时序特征的长时间连续二进制数据流的数字器件对特殊激励源的需求。

信号产生技术

使用信号源产生波形的方法有许多种。选择何种方法取决于“被测设备”(DUT)的已知信息和该设备对输入信号的要求——是否需要添加畸变、误差信号及其他变量。现代高性能信号源提供至少三种波形产生的方法。

- **模拟**：以波形的定义(通常来自模拟器或者波形库)为基础，创建一个事件或一个事件序列。
- **复制**：记录示波器上存在的波形并将其发送到信号源进行重建(见图 10.1)。
- **替代**：创建或者修改一个已定义信号，用它替代一个本应由电路产生但却无法得到的信号(见图 10.1)。

信号源的主要用途

信号源有几百种用途。在电子测量领域中，信号源的应用分为以下三类：功能验证、特征量提取和极限性能测试。一些具有代表性的应用如下。

- **功能验证——数字调制分析**：开发新型发射、接收硬件的无线设备设计商必须对基带 IQ、信号(包括有损和无损)进行仿真，以确保他们研发的设备能够符合刚出现的和已有的无线通信标准。某些高性能 AWG 能够提供数据率高达 12.5 Gbps 的两路独立的低畸变、高精度信号——一路产生 I 信号，另一路产生 Q 信号。
- **特征量提取——测试“数模转换器”(DAC)和“模数转换器”(ADC)**：为了确定线性度、单调性和畸变的范围，对新开发出的 DAC 和 ADC 必须进行全面测试。AWG 的现有技术水平能够同时产生同相模拟和数字信号，并驱动数据率高达 12.5 Gbps 的器件。
- **极限性能测试——通信接收机的极限性能**：使用串行数据流结构(普遍用于数据通信总线和

磁盘驱动放大器中)的工程师需要用“损害”(抖动和时序违例)来测试器件的极限性能。高级的信号源提供了高效的内置式“抖动编辑和生成工具”，这省去了工程师大量的、花在计算上的时间。这些仪器能够将关键的信号边沿移动 20 ps。

信号源硬件结构：任意波形发生器

从根本上说，AWG 是一种复杂的重放系统——波形的产生是以(描述交流信号不断变化电压的)存储数据为基础的。AWG 的结构框图看似很简单。

为了理解 AWG，先要掌握“数字采样”这个概念。“数字采样”是指使用(沿波形变化测量到的)一组电压样本(数据点)来定义一个信号。这些样本既可能是通过使用示波器等仪器对波形进行实际测量得到的，也可能是利用图形化技术或数学技术得到的。

图 10.2(a)画出了一组采样点。所有采样点的时间间隔都是相等的，尽管曲线使得这些间隔看上去不一样。在 AWG 中，采样值以二进制形式存储在快速 RAM 中。

通过读取存储器地址和输入数据到 DAC，可以利用存储信息随时重建信号。图 10.2(b)描述了结果。注意：AWG 的输出电路在样点间进行滤波，将这些点连接起来，产生纯净的、不间断的波形。UUT 看到的不是离散的点，而是连续的模拟波形。图 10.3 是 AWG 实现上述过程的简化框图。

在这里，还需要定义一些术语和说明影响信号重建保真度的条件。在这些术语和条件中，首要的是奈奎斯特定理。该定理指出：采样频率至少为被采样信号最高频率分量的 2 倍。例如，为了对 1 MHz 的信号进行采样，至少需要 2 MS/s 的频率进行样本采集。尽管该定理通常作为信号采集(如示波器)的指南，但它和 AWG 的关系是显而易见的——为了能够真实再现期望信号的细节，存储波形必须拥有足够多的采样点。

AWG 按照指定范围内的任意频率获取并从存储器读出这些样点。如果一组存储样点满足奈奎斯特定理，而且描述了一个正弦波，那么 AWG 的输出将是一个正弦波形。当然，该仪器存在一个最高工作频率(采样率)。通常，这个频率用 MS/s 或 GS/s 表示。

AWG 的其他硬件特点也同样重要，特别是垂直分辨率、存储深度和采样率。垂直分辨率(幅度分辨率)表示的是上述样点幅度测量的精度。这个精度和仪器 DAC 的二进制字长(以 bit 为单位)有关——位数越多，精度越高。尽管越多越好，但高频 AWG 的精度(8 位或者 10 位)往往比通用 AWG(12 位或 14 位)要低。

在仪器满幅电压范围内，10 位精度的 AWG 提供了 1024 个采样电平。例如，若一个 10 位的 AWG 电压范围为 $2V_{p\text{-}p}$(峰-峰值)，则每个样本代表着约 2 mV 的量阶——假定不受结构中其他因素的限制，这就是该仪器提供的最小增量。

在仪器的灵活性方面，“存储深度”具有关键作用。大存储深度提供两个好处：

- **可以存储更多周期的期望波形。**这很有用，因为减少了“端点”数。“端点”是指波形的最后存储器位置，在这个位置之后，AWG 必须折回存储器的开始位置，才能连续产生输出信号。由于转换时发生的误差是不可避免的，所以希望减少端点数量。
- **可以存储更多的波形细节。**在复杂波形的脉冲边沿处和波形过渡处存在着高频信息。在这些快变区域，很难采用和简单的、可预测的正弦波形一样的插值方法。为了真实再现一个复杂信号，现有的波形存储容量必须用于存储更多的过渡波形和振荡波形，而不是更多的信号周期。

某些顶级型号的 AWG 同时提供大存储深度和高采样率。这些仪器可以存储和重现复杂的射频波形，甚至包括用于网络设备物理层测试的伪随机数据流。同样，这些 AWG 能够产生非常短暂

的数字脉冲和暂态波形。

第 29 课　示　波　器

自然界以正弦波的形式运动，不论是海浪、地震、声爆、爆炸、空气中传播的声音，还是运动物体的固有频率，都是如此。整个宇宙遍布着能量、振动粒子和其他人眼看不见的作用。甚至是光线——半波、半粒子——也有基波频率，这个频率在人看来就是颜色。传感器可将各种作用转化成使用示波器能测量和研究的电信号。有了示波器，科学家、工程师和教师等人员就能够看到事件随时间变化的过程。

对于电子设备的设计、生产和维修人员而言，示波器是不可或缺的工具。在如今这个迅速变化的世界里，工程师需要最好的工具迅速、准确地解决测量问题。示波器就像工程师的眼睛一样，它是完成当今苛刻测量任务的关键。

示波器的用途不仅限于电子技术领域。通过使用适当的传感器，示波器可以测量各种现象。传感器(换能器)是一种能够受其他物理量(如声音、机械张力、压力、光和热)激励而产生电信号的器件。麦克风就是一种传感器，它将声音转换为电信号。

每个人都可以使用示波器——从物理学家到电视维修技工。汽车工程师使用示波器测量发动机的振动，而医学研究人员使用示波器测量脑电波。示波器具有无数种可能的用途。

示波器类型

电子设备可分为模拟设备和数字设备两大类。模拟设备处理的是连续变化的电压，而数字设备处理的是代表电压样值的离散二进制数字。

示波器也有类似的分类——模拟示波器和数字示波器。对很多应用而言，二者都可胜任。然而，它们也有各自的独特之处，对特定应用而言，这些独特之处可能使之更适用，也可能使之不是很适用。数字示波器还可进一步分为数字存储示波器(DSO)、数字荧光示波器(DPO)和数字采样示波器三类。

模拟示波器

模拟示波器的基本工作原理是将被测信号电压直接加到电子束的垂直轴上，控制电子束在示波器荧光屏(通常是“阴极射线管”，CRT)上从左向右扫描。荧光屏背面涂有一层荧光物质，电子束撞击到的地方就会发光。当信号在显示屏水平移动时，信号电压就成比例地将电子束上、下偏转，从而在荧光屏上显示出信号波形。电子束撞击荧光屏上哪个位置越频繁，哪个位置就越亮。

CRT 限制了模拟示波器可显示的信号频率范围。当频率很低时，信号看上去就是一个明亮的缓慢移动的亮点，很难让人辨别出波形。当频率很高时，CRT 的显示速度又形成了限制。当信号频率超过 CRT 的显示速度时，信号就会变得太暗而无法看清。最快的模拟示波器能够显示大约 1 GHz 的信号。

当把示波器探头连接到电路中的时候，电压信号将通过探头抵达示波器的垂直系统。图 10.4 展示了模拟示波器显示被测信号的方法。根据垂直刻度(伏/格)的设置，电压信号可能经由衰减器而被衰减，也可能经由放大器而被放大。

然后，信号就直接抵达 CRT 的垂直偏转板。偏转板两端的电压会导致示波器荧光屏上出现一个移动的发光点。这个发光点是由电子束轰击 CRT 内壁的荧光物质而产生的。正电压会使发光点上移，而负电压会使发光点下移。信号也会抵达示波器的触发系统，从而启动或者触发一次水平扫描。水平扫描是指水平系统使发光点在荧光屏上移动。

触发水平系统能使水平时基驱动发光点在指定时间间隔内在荧光屏上从左到右移动。多次快速扫描会使发光点汇合为一条实线。在更高的速度上，发光点在荧屏上往返的速度可达 500000 次/秒。

水平扫描和垂直偏转一起在荧光屏上勾勒出信号的形状。为了稳定一个重复信号，需要进行触发——它保证每次扫描重复信号时的起始点都一样。

此外，模拟示波器都有聚焦、亮度调节控制功能，能够产生清晰的显示。当需要实时显示(即刻显示)快变波形时，人们往往想起模拟示波器。模拟示波器的荧光显示器具备一种亮度分级特性——频繁出现的信号特征具有更亮的扫描线。有了这种亮度分级特性，仅通过查看扫描线的亮度级别就可以很容易地区分信号细节。

数字示波器

和模拟示波器相反，数字示波器采用“模数转换器”(ADC)将被测电压转换为数字信息。数字示波器采集到的波形是一组样本，然后将这些样本保存，直至样本数达到足以描述波形的数量。之后，数字示波器将这些样本重新组合、在示波器显示屏上显示出来。数字示波器可分为数字存储示波器(DSO)、数字荧光示波器(DPO)和数字采样示波器。

数字技术意味着示波器可以稳定、清晰地显示示波器范围内的任何频率信号。对于重复性信号，数字示波器的带宽是示波器前端器件模拟带宽(一般是指 3dB 带宽)的函数。对于只出现一次的信号和暂态信号(如脉冲信号和阶跃信号)，数字示波器的带宽可由示波器的采样率决定。

数字存储示波器

数字存储示波器(DSO)是一种常见的数字示波器。DSO 的显示取决于光栅显示屏，而不是荧光显示屏。DSO 能够捕捉、观测那些可能只出现一次的事件——暂态事件。由于波形信息是以一组二进制值的数字形式存储的，所以，可以对这些信息进行分析、存档、打印和处理。这个过程既可以在示波器内部进行，也可以在计算机上进行。波形不必是连续的，在信号消失之后，DSO 仍然可以显示波形。和模拟示波器不同的是，DSO 能够永久保存信号，并可以对波形进行各种处理。然而，DSO 一般没有实时亮度分级能力，因此，它无法反映信号亮度级别的变化。

数字荧光示波器

数字荧光示波器(DPO)采用了一种新型示波器结构，该结构使得 DPO 具备精确重建信号的独特的采集、显示能力。DSO 使用串行处理结构(见图 10.5)对信号进行采集、显示和分析，而 DPO 使用并行结构来实现这些功能，如图 10.6 所示。DPO 结构使用专用、独特的 ASIC 硬件采集波形的映像，波形采集速率高，信号可视化程度就高。这就提高了观测到数字系统暂态事件(短脉冲、毛刺和转换错误)的概率。

数字采样示波器

在测量高频信号时，示波器在一次扫描中可能无法搜集足够多的样本。当需要准确记录那些频

率成分远远超过示波器采样频率的信号时，数字采样示波器就是一种理想的工具。这种示波器能够测量的信号速度要比其他示波器能够测量的信号速度快一个数量级。对于重复信号而言，数字采样示波器的带宽和高速时序是其他示波器的 10 倍。连续等效时间采样示波器的带宽可达 80 GHz。

第 30 课　逻辑分析仪

逻辑分析仪的功能和示波器是不同的，最明显的区别是两者输入信号通道数不一样。一般而言，数字示波器的通道数可达四路，每个通道输入一路数字信号。有些复杂的系统设计需要几千路输入通道，也存在相应规模的逻辑分析仪。

逻辑分析仪测量、分析信号的方式也和示波器不一样。逻辑分析仪并不测量信号的模拟细节，而是检测信号的逻辑门限电平。当把逻辑分析仪连接到数字电路的时候，人们只关心信号的逻辑状态。逻辑分析仪只看两个逻辑电平：当输入信号电压超过门限电压时，输入信号就被认为是“高电平”或者“逻辑 1”；当输入信号电压低于门限电压时，输入信号就被认为是“低电平”或者“逻辑 0”。逻辑分析仪对输入信号进行采样的时候，它将根据信号电平和门限电平的相对大小存储一个“1”或者一个“0”。

逻辑分析仪的“波形时序显示”和数据手册上提供(或者模拟器产生)的“时序图”很相似。所有信号都和时间相关，这样就可以观察到“建立－保持时间”、脉冲宽度、外来数据或丢失数据。除通道数量不同外，逻辑分析仪还提供了许多支持数字设计和验证的功能，其中包括：

- **复杂的触发机制**——允许用户指定逻辑分析仪进行数据采集的条件。
- **高密度探头和适配器**——简化了逻辑分析仪和“被测系统”(SUT)之间的连接。
- **分析能力**——将采集到的数据转化为处理器指令，并将这些指令和源代码相关联。

结构与操作

逻辑分析仪和数字信号相连，采集、分析数字信号。使用逻辑分析仪有如下四步：

① **探头(将逻辑分析仪连接到 SUT 上)。**

② **设置(时钟模式和触发模式)。**

③ **采集。**

④ **分析和显示。**

图 10.7 是逻辑分析仪的简图，每个方框代表了若干软、硬件模块，方框上的编号和上述四个步骤相对应。

探头

逻辑分析仪可以一次采集大量信号，这一点就和示波器不同。探头连接到 SUT 上，探头的内部比较器将输入信号电压和门限电压 V_{th} 进行比较，并判决信号的逻辑状态(1 或者 0)。门限电压值是由用户设置的，设置范围从 TTL 电平到 CMOS 电平、ECL 电平和用户定义电平。

探头能够采集高质量信号，而且对 SUT 的影响最小。逻辑分析仪探头为逻辑分析仪提供了高质量的信号通道，并将对 SUT 造成的电气负载降至最小，而且能够适配于各种电路板和器件。

逻辑分析仪探头的阻抗(电容、电阻和电感)会成为整个被测电路负载的一部分，所有探头都

会表现出负载特性。逻辑分析仪探头引入 SUT 中的负载应该尽量小，并为逻辑分析仪提供准确的信号。

探头的电容会将信号边沿的过渡过程变得缓慢。这一点为什么很重要呢？因为过渡缓慢的边沿穿越电路门限电平的时间要晚一些，这样就会引入时序误差。随着时钟频率的提高，这个问题会变得很严重。在高速系统中，过量的探头电容可能会(潜在地)阻止被测系统工作！选择一个总电容尽可能小的探头往往是很关键的。

设置(时钟模式和触发模式)

逻辑分析仪可以从多引脚器件或者总线上采集数据。“采集速率”是指对输入信号进行采样的频率。其功能和示波器中的时基电路是一样的。数据采集有两种类型：

- **异步采集**——对信号的时序信息进行采集。在这种模式下，逻辑分析仪的内部时钟用于数据采集。数据采样越快，测量精度就越高。在目标器件和逻辑分析仪的数据采集之间不存在固定的时序关系。这种采集模式主要用于观测 SUT 信号间的时序关系。
- **同步采集**——用于获取 SUT 的状态。SUT 的信号决定了样本点(何时、以什么频率采集数据)。作为同步采集时钟的信号可能是系统时钟、总线控制信号或者一个引发 SUT 状态发生改变的信号。在有效边沿对数据进行采样，这个数据反映了逻辑信号稳定时的 SUT 状态。当且仅当被选信号有效时，逻辑分析仪才进行采样。

数据采集：实时采集存储器

逻辑分析仪的探测、触发和定时系统将数据传送给实时采集存储器。这个存储器是逻辑分析仪的心脏——它既是所有 SUT 的采样数据的目的地，也是逻辑分析仪所有分析显示功能的数据来源。

逻辑分析仪的存储器能够按照仪器的采样率存储数据。这种存储器可想象为有一定宽度和一定深度的矩阵。

在一个触发事件出现或者用户告知停止前，逻辑分析仪一直在积累数据。这就是采集——具有很高时序准确性的多通道波形显示(可用于观测所有被采集信号间的相互作用)。

选择逻辑分析仪时，宽度和深度都是关键因素。下面几条提示会有助于确定通道数量和存储深度：

- **你需要记录、分析多少路信号？** 逻辑分析仪的通道数和希望记录的信号数是直接对应的。数字系统总线的宽度有很多种，在总线检测的同时还需探测一些其他信号(时钟信号、使能信号等)。一定要考虑到你要同时记录的所有总线和信号。
- **你需要记录多长时间？** 这将决定逻辑分析仪的存储深度，而这一点对于异步采集尤为重要。对于给定的存储容量，随着采样率的提高，总的采集时间会下降。例如，当采样率为 1 ms 时，1 M 存储器存储的数据对应的时间跨度为 1 s。当采样时钟周期为 10 ns 时，同样的 1 M 存储器只能存储 10 ns 的数据。采集的样本多(采集的信号时间长)，捕获错误、(导致错误的)问题的机会就大。说到存储容量，当然是越多越好！

分析和显示

实时采集存储器中存储的数据可用于几种不同的显示和分析模式。信息一旦存到系统当中，就可以按照多种格式(从时序波形到和源代码指令相关的助记符)观察这些信息。

通过波形显示可以看到多个通道的详细情况，你可以看到所有记录信号之间的时序关系，这很

像示波器。波形显示常用于时序分析，它极适合完成如下工作：

- **诊断时序问题**(针对 SUT 硬件)。
- **验证硬件工作是否正常**(通过对记录结果和仿真器输出或数据手册时序图进行比较)。
- **测量硬件时序特性**(竞争状态、传输延迟、脉冲有无)。
- **分析毛刺。**

“列表显示”可以提供用户选择的、字母数字形式的状态信息。列表中的数值是从整个总线上记录的样值得来的，这些数值可以表示为十六进制格式，也可以表示为其他格式。“列表显示”的目的在于展示 SUT 的状态。“列表显示”展示给用户的信息流就是 SUT 看到的信息流——一组数据字。

状态数据的显示格式有好几种。“实时指令跟踪”将每个总线操作进行反汇编，准确判定总线读取的是哪条指令。它将对应的助记符指令及其地址显示在逻辑分析仪显示屏上。

还有一个附加显示，就是“源码调试显示”。通过将源码和过去的指令跟踪情况关联起来，它能使调试工作更有效率。有了“源码调试显示”，人们即刻就可以看到在一条指令执行时，到底发生了什么。

练习参考答案

第 1 单 元

1. 完形填空

(1) Component, All work that you do, indirectly, if not, amazing, devices, plays a significant role, aspects, determinant, While, dictate, reach its full potential, enable, In, how reliably your system will run, dependable, to some extent, consumes, to, a great deal of, on, major, in turn

(2) non－volatile, for, a wide range of, embedded controllers, such as, from, to, for, to, semiconductor memory, power consumption, advantages, disadvantages, with, of

2. 英汉互译

(1) 许多现代社会产品和服务都基于电气工程师和计算机科学家的工作。在过去几十年中，数字电子器件成本的大幅下降带来了计算机应用的迅猛增长。与此同时，随着人们对计算机科学认识的提高，开发更强大、更复杂、更灵活的新型软件系统已经成为可能。

(2) 1971 年 11 月，Intel 公司的三位工程师发明了世界上首片商用微处理器 4004。按照今天的标准，4004 相当原始。它仅含 2300 个晶体管，在 1 s 内可进行约 60000 次运算。今天，微处理器是最复杂的量产产品。它包含几百万个晶体管，1 s 可完成几亿次运算。

(3) 微处理器是个人计算机的神经中枢。在这个小小的硅片上集成了几百万个开关和通路，它们帮助计算机做出重要决定、完成有意义的任务。但微处理器不是只为计算机而设计的。在电话、汽车等许多日常物品中，人们都可以发现微处理器。

(4) 存储器是计算机和嵌入式系统中用来保存信息的部件。在嵌入式应用中需要存储信息，因此，设计中总含有某种类型的存储器。现在有许多种存储器件，因此对于嵌入式用户存在多种选择。

(5) 存储器可以分为两大类：易失的和非易失的。系统掉电后，易失存储器会丢失内容。多数随机存取存储器是易失的。非易失存储器在系统或器件掉电后不会丢失数据。

(6) Modern electronic systems are increasingly digital in nature, exceedingly complex, and would be inconceivable without today's VLSI chip technology. Indeed, such systems are so complex that the principles of their design bear great similarities to the design principles of large software systems. Thus, computer science and electronic system design require similar backgrounds in many respects, and computer aids to modern electronic systems design are essential in this ever-expanding domain of engineering.

(7) A chip is a small piece of semiconducting material on which an integrated circuit is embedded. A typical chip is less than 1 square inches and can contain millions of transistors. There are different types of chips. For example, CPU chips contain an entire processing unit, whereas memory chips contain blank memory.

(8) A microprocessor is a silicon chip that contains a CPU. There are three basic characteristics differentiate microprocessors: instruction set, bandwidth and clock speed.

(9) There are several reasons to use flash memory instead of a hard disk. Flash memory allows faster access. It is noiseless, lighter, smaller in size and has no moving parts.

(10) Memory is the part of a computer system that is used to run programs and store data. The main memory is built from random-access memory (RAM) chips. The amount of memory available determines the size of programs that can be run, and whether more than one program can run at a time. Main memory is temporary, and the data is lost when the computer is turned off. It is distinguished from more permanent internal read-only memory (ROM), and external storage. ROM contains the computer's essential programs. In the general sense, Memory can be any device that can hold data in machine-readable format.

3. 写作练习(答案略)

第 2 单 元

1. 完形填空

(1) at, while, in, In the field of telecommunication, modems and speech processing, channel selection, anti-aliasing low-pass filters, signal conditioning, filter, a linear phase shift, to, time delay, all-pass filters, of, such as, called, making, cases, active filters, an operational amplifier, in, resistors and capacitors

(2) passive components, as, by, from, ignorant of, analog circuitry, limited frequency ranges, to, radiated or conducted, as well, a high-pass filter

2. 英汉互译

(1) 由于很难定义运算放大器的“动态范围”(DR)，因此先给“数模转换器”(DAC)的动态范围下定义。DAC 的动态范围就是其最大输出电压和最小输出电压之比。

(2) 这个“动态范围”的定义也可用于运算放大器，其最大输出电压摆幅为 V_{OUTMAX}。输出电压摆幅定义为运算放大器能达到的最高输出电压(V_{OH})减去最低输出电压(V_{OL})。从运放集成电路数据手册中，可以很容易获得 V_{OH} 和 V_{OL}。

(3) 噪声是在一段时间内随机变化的，信号或噪声电平的瞬时值不足以描述这种情形。信号和噪声要用其在较长时段内的平均值(均方根值，RMS)来描述。“信噪比”(SNR)用于度量噪声环境中的信号质量。SNR 是功率比，是在电路输出端得到的。我们感兴趣的 SNR 是电压比(因为阻抗是常数)，是在运算放大器的输入端得到的。就是说，全部噪声(包括电阻噪声)电压必须在运算放大器的输入端以均方根值进行计算。

(4) 切比雪夫等波纹滤波器的频域滚降分布在整个通带范围内。其通带内波纹较多，而过渡区的滚降比较陡峭。因其在滤波过程中的 Q 值较高，切比雪夫等波纹滤波器的暂态响应和阶跃响应较差。

(5) 芯片封装就是芯片的外罩。有了它，才能把芯片插入(插座安装)或焊至(表面安装)印制电路板上。在外行看来，芯片安装设计似乎不太重要。事实上，芯片封装技术是一个庞大、复杂的行业。为尺寸不断缩减的电路小片(裸片)提供越来越多的“输入/输出互连”是经常出现的问题。除此之外，封装和半导体电路的“缩微化”均有助于减小蜂窝电话和其他手

持设备的尺寸。

(6) In one respect, voltage is like water: you don't appreciate its value until your supply runs low. Low-voltage systems, defined here as a single power supply less than 5 V, teach us to appreciate voltage.

(7) Ohm's law is stated as $U=I\quad R$, and it is fundamental to all electronics. Ohm's law can be applied to a single component, to any group of components, or to a complete circuit. When the current flowing through any portion of a circuit is known, the voltage dropped across that portion of the circuit is obtained by multiplying the current times the resistance.

(8) Circuits are a mix of passive and active components. The components are arranged in a manner that enables them to perform some desired function. The resulting arrangement of components is called a circuit or sometimes a circuit configuration. The art portion of analog design is developing the circuit configuration.

(9) Noise sets a limit on the information and signals that can be handled by a system. The ability of an amplifier, receiver, or other device to discern a signal is degraded by noise. Noise mixed with the incoming signal, noise generated by the op amp, resistor noise, and power supply noise ultimately determine the size of the signal that can be recovered and measured.

(10) A Butterworth (maximally flat) filter is the most commonly used general-purpose filter. It has a monotonic passband with the attenuation increasing up to its 3-dB point, which is known as the natural frequency. This frequency is the same regardless of the order of the filter. However, by increasing the order of the filter, the roll-off in the passband moves closer to its natural frequency and the roll-off in the transition region between the natural frequency and the stopband becomes sharper.

3. 写作练习(答案略)

第 3 单 元

1. 完形填空

(1) layers, via, while, discrete components, hundreds of thousands, etched circuit, copper foil, glassfiber, negative image, after, similar process, on

(2) electrical pulses, in, are to, as well as, communication protocols, graphical interfaces, engineers, programmers and consultants, implies a structure, voltage levels, frequencies and duration, hardware, read, transmitted, activate

2. 英汉互译

(1) 定制门阵列逻辑、专用集成电路、球栅阵列、多芯片模块和亚纳秒数字器件的使用为电磁兼容工程师提供了新的挑战性机遇。

(2) 工作在高频的电阻表现为电感和并联的电阻、电容之间的级联。工作在高频的电容表现为电感、电阻和电容的级联。

(3) "侧重集成电路的仿真程序"(SPICE)是一个广泛应用于模拟电路和混合电路仿真的程序。SPICE 解决了多组建立在频域、稳态和时域的非线性方程，从而能够对晶体管和门电路的

设计行为进行模拟。

(4) “输入/输出缓存规范”(IBIS)是一种用来对集成电路输入/输出模拟特性进行定义的格式。IBIS 模型是 ASCII 文件。在不泄露电路设计专利的前提下，IBIS 模型提供了对器件建模所需的行为信息。

(5) 在多导线系统中，导线间的过量耦合(即串扰)会引发两种有害效应。首先，串扰改变了总线中传输线的有效特性阻抗和传输速度，从而使总线传输线的性能发生改变；此外，串扰将噪声感应到其他导线上。串扰现象的这些特点使得系统性能严重地依赖于数据模式、导线间距和切换速率。

(6) Crystal is a transparent quartz material that contains a uniform arrangement of molecules.

(7) As high-speed digital systems become more complex, it is becoming increasingly difficult to manage timings and signal integrity.

(8) Crosstalk is the coupling of energy from one line to another. It will occur whenever the electromagnetic fields from different structures interact.

(9) In digital designs the occurrence of crosstalk is very widespread. Crosstalk will occur on the chip, on the PCB board, on the connectors, on the chip package, and on the connector cables.

(10) Many systems operate at high frequencies at which conductors no longer behave as simple wires, but instead exhibit high-frequency effects and behave as transmission lines. If these transmission lines are not handled properly, they can unintentionally ruin system timing.

3. 写作练习(答案略)

第 4 单 元

1. 完形填空

(1) share afrequency channel, digitized speech, a number of time slots, data rate, as well as, Apart from, synchronization, from, to, in, in, from, by, there are several frequency channels used in a cell, or, are referred to as, respectively, as that

(2) linear modulation, also known as, over, spectral density of the signal, as, at the time of, and, using CDMA scheme, for, this, in multipath situations, because of, asynchronous, not necessarily, It should be noted that, other, in, signals transmitted by the base station to other mobiles

2. 英汉互译

(1) 不同的无线射频链路以不同的方式公用频谱带宽，这些公用方式称为“多址模式”。“频分多址”(FDMA)、“时分多址”(TDMA)、“码分多址”(CDMA)和“空分多址”(SDMA)是四种主要的多址模式。

(2) 移动电话系统的服务区划分为一个个蜂窝小区。每个蜂窝小区都有一座基站，通过在无线链路上发送、接收信号，基站和蜂窝内的移动用户通信。基站向移动用户发送信号称作“前向链路”或“下行链路”，移动用户向基站发送信号称作“后向链路”或“上行链路”。每个基站都要和“移动交换中心”(MSC)联络，而 MSC 的作用是把来往于基站的呼叫和其他蜂

窝的移动用户、公共交换电话网连接起来。

(3) 基站使用两类无线信道和移动用户进行通信：传送控制信息的“控制信道”和传送消息的“通信信道”。

(4) 在 FDMA 方式中，可用频谱划分为若干具有一定带宽的频道，每个呼叫都使用不同的频道。第一代蜂窝系统采用的都是 FDMA 方式。

(5) 在 TDMA 和 FDMA 系统中，为了防止信道间干扰，某个蜂窝小区使用的频道不会被邻近蜂窝小区使用。在 CDMA 系统中，可以在邻近蜂窝小区中使用相同频率，从而增加了系统容量。

(6) The generic term channel is normally used to denote a frequency in FDMA system, a time slot in TDMA system, and a code in CDMA system or a combination of these in a mixed system.

(7) In FDMA and TDMA sytems once a channel (frequency and/or time slot) is allocated to a user, that channel cannot be used during nonactivity periods.

(8) The cellular concept was invented by Bell Laboratories and the first commercial analog voice system was introduced in Chicago in October 1983.

(9) The mobile station is a station in the cellular radio service intended for use while in motion at unspecified locations. Mobile stations may be hand-held personal units or installed in vehicles.

(10) A base station communicates with mobiles using two types of radio channels, control channels to carry control information and traffic channels to carry messages.

3. 写作练习(答案略)

第 5 单 元

1. 完形填空

(1) With the advent of VLSI in the 1980s, standard ICs, of, implement, logic functions, custom ICs, a smaller number of components, into, reduce cost and improve reliability, in, to, If you need a large amount of memory, in, to, to

(2) a variety of chips, in, by, known as, to, depending on, which, each time, to, the most common PLD chips, Unlike gate arrays, in the field, programmable storage chips, rather than

2. 英汉互译

(1) 有些数字逻辑集成电路和模拟集成电路(如模数转换器)是标准器件或者标准集成电路。可以从产品目录和数据手册当中选择标准集成电路，从经销商处购买。系统制造商和设计人员可以在各种不同的微电子系统(“微电子系统”是指应用微电子技术和集成电路的系统)中使用相同的标准器件。

(2) 现代亚微米 CMOS 工艺的复杂度和亚微米双极性工艺、BiMOS(即双极性和 CMOS 的组合)工艺是一样的。不过，CMOS 集成电路现已居于主导地位——其产量比采用其他技术制造的集成电路都大。由于规模带来的成本节约，CMOS 集成电路的成本要低于同等功

能的双极性或 BiMOS 集成电路。在某些特殊应用中，仍要使用双极性集成电路和 BiMOS 集成电路。例如，和 CMOS 集成电路相比，双极性集成电路往往能够处理更高的电压。因此，在电力电子、汽车、电话电路等领域，双极性集成电路和 BiMOS 集成电路是非常有用的。

(3) “专用集成电路”(ASIC)不是微处理器这样的通用芯片，而是为特殊用途设计的定制芯片。和通用 CPU 相比，使用 ASIC 可以提高性能。因为 ASIC 是使用硬连线完成具体工作的，并不需要为取指令和解释指令付出开销。ASIC 能够以尽可能快的速度完成一项电子操作，当然，其电路设计应该具有高效的结构。

(4) “软核”是在可编程逻辑芯片上或在微控器、片上系统的可编程部分实现的。根据生产商的不同，逻辑功能可能以原理图、网表或者硬件描述语言代码的形式提供。“软核”可以用芯片上的“硬核”来实现。

(5) “硬核”是一种在芯片电路级设计的、具有特定功能的逻辑模块。微处理器就是典型的“硬核”，微处理器和特定芯片制造商的半导体生产工艺是联系在一起的。微控器或片上系统可能全部都是“硬核”，也可能是由“硬核”(硬连线)和“软核”(可编程)组合而成的。

(6) EDA (electronic design automation) uses the computer to design and simulate the performance of electronic circuits on a chip.

(7) Hardware-description language is a language used to describe the functions of an electronic circuit for documentation, simulation or logic synthesis.

(8) Logic synthesis is the conversion of a high-level electronic circuit description into a list of logic gates and their interconnections, called the “netlist”. Every logic synthesis program understands some subset of Verilog and VHDL.

(9) CPLD (Complex PLD) is a programmable logic device that includes a reprogrammable interconnect between the logic blocks. CPLDs are mostly EEPROM and flash based.

(10) FPGA is a programmable logic chip with a high density of gates. Containing up to hundreds of thousands of gates, there are a variety of architectures. Some are very sophisticated, including not only programmable logic blocks, but programmable interconnects and switches between the blocks. FPGAs are mostly reprogrammable (EEPROM or flash-based) or dynamic (RAM-based).

3. 写作练习(答案略)

第 6 单 元

1. 完形填空

(1) analog system, continuous, discrete points, sampling rate, twice, frequency component, is referred to as, a given set of data, before new data arrives, at a given clock rate, odd, because, the amount of data

(2) status bit, makes it difficult to develop real-time systems, efficient, clock cycles, event-driven interrupt programming, real-world signals, respond, meet all of the following demands, multiple, keep up with, a frame of data, serial port, stops what it is doing, interrupt vector, on, To transmit data, transfer a value, a program segment, or, to,

precision and flexibility

2. 英汉互译

(1) 处理现实世界信号的主要原因是“提取信息”。信息往往存在于信号的幅度(绝对幅度或相对幅度)、频率(频谱成分)、相位(与其他信号间的时序关系)。当将有用信息从信号中提取出来后，就能以多种方式使用它。

(2) 有时，人们希望对信号中的信息赋以新的格式。在频分多址电话系统中传输的语音信号就是这种情况。模拟技术用于在频谱范围内对语音信道进行“堆叠”，并通过微波中继、同轴电缆或光纤进行传输。在数字传输链路中，首先使用模数转换器将模拟语音信息转化为数字信号。代表单路语音信号的数字信息在时域进行多路复用(时分多址)，并通过一条串行数字链路进行传输(如 T 载波系统)。

(3) 信号处理还可以满足以下需求：在不损失重要信息的前提下，对信号的频率成分进行压缩、赋以新的格式并以低数据率进行传输，从而减小可用带宽。和数字移动射频系统、运动图像专家组(MPEG)录放设备和高清晰度电视一样，在高速调制解调器和“自适应差分脉冲编码调制系统”(ADPCM)中也大量使用数据压缩算法。

(4) 数字信号处理技术源于 20 世纪六七十年代，当时刚刚出现数字计算机。那时的计算机是相当昂贵的，数字信号处理仅用于某些关键领域，并在雷达和声呐(关乎国家安全)、石油勘探(可以创造巨大利润)、空间探索(数据具有不可替代性)和医学图像(可用于拯救生命)这四个方面获得了开拓性进展。

(5) “采样”是模数转换的第一步。通过使用“采样保持电路”可以实现采样——“采样保持电路”从信号中获取一个样本，并保持其稳定到下一个采样时刻。样本通常是按照固定时间间隔(称为采样周期)采集的。假如采样率不够高，就会出现一种称为“混叠”的信号失真。“采样保持电路”的输出传送到“模数转换器”，转换器会选择一个最接近信号实际幅度的“量化电平”。这就是模数转换的第二步。可供选择的电平数取决于转换器的位数——对于 N 位转换器，可用电平数为 2^N。在数字系统中，模拟幅度是无法得到完美表示的，所以就会出现“量化误差”。

(6) Digital signal processing is an essential element of countless home and business systems. Its domain can only increase as time proceeds. Thus, DSP is becoming an essential area of expertise for technologists and engineers.

(7) Signals, like sound, light, or voltage, are information-bearing variations. Analog signals are real-world signals. They are defined at every point in time and may take an infinite number of possible amplitudes. Analog signals are not well-suited to processing by computer. They may be converted to digital signals through sampling and quantization. Digital signals, on the other hand, are defined only at sampling points, and may take only a finite number of discrete amplitudes. After processing, a digital signal is converted back into an analog signal.

(8) DSP can be divided into two categories: fixed-point and floating-point. These refer to the format used to store and manipulate numbers within the devices. Fixed-point DSP usually represent each number with a minimum of 16 bits, floating-point DSP typically use a minimum of 32 bits to store each value.

(9) DSP are programmed in the same languages as other scientific and engineering applications, usually assembly or C. Programs written in assembly can execute faster, while programs written in C are easier to develop and maintain. In traditional applications, such as programs

run on personal computers, C is almost always the first choice. If assembly is used at all, it is restricted to short subroutines that must run with the utmost speed. However, DSP programs are different from traditional software tasks in two important respects. First, the programs are usually much shorter, say, one hundred lines versus ten-thousand lines. Second, the execution speed is often a critical part of the application.

(10) DSP often have specialized instructions that make programming for common DSP tasks more convenient and more efficient. For example, most DSP offer multi-function instructions that exploit their parallel architecture. Other constructs that are frequently offered are efficient looping schemes, since so many DSP operations involve a great deal of repetition.

3. 写作练习(答案略)

第 7 单 元

1. 完形填空

(1) identify the direction of the sound, between, in, frequencies above about 1 kHz, than, on the opposite side of the head, slightly later than, due to its greater distance from the source, Based on, with, coming from the center of the listener's head, While, from, to, This is because, that can provide this information, as, by, from, By measuring the interval between transmission and echo, with, echo localization

(2) to, fidelity, between, To accomplish these two goals, analog wave, a device called an analog-to-digital converter, by, is, as long as, be very similar to, at, that, are stored as bytes, A CD can store up to 74 minutes of music

2. 英汉互译

(1) 通常认为，人类听力范围在 20 Hz～20 kHz 之间，但人耳对 1～4 kHz 之间的声音要敏感得多。举例来说，收听者能察觉到声压级为 0 dB 的 3 kHz 声音，而 100 Hz 声音却需要具有 40 dB 的声压级(幅度增至 100 倍)。收听者可以辨别 3 kHz 频率处大于 0.3%左右的频差。在 100 Hz 频率处，可分辨频差要达到 3%。可以比较一下，钢琴相邻两键的频差约为 6%。

(2) “八度音阶”或“倍频程”(octave)是指频率上的两倍关系。在钢琴上，一个“八度音阶”包括八个白色琴键，这就是 octave 得名的原因(在拉丁文中，octo 代表“八”)。换言之，每隔七个白色琴键之后，钢琴频率就加倍，而整个键盘跨越了七个多一点的“八度音阶”。通常认为人类听力范围在 20 Hz～20 kHz 之间，这两个频率分别对应于钢琴键盘左边约 1/2“八度”处的频率和钢琴键盘右边两个“八度”处的频率。

(3) 实现“压扩”的方法有三种：① 先用非线性电路对模拟信号进行处理，然后输出到 8 位线性 ADC；② 使用内部采用非均匀量化间隔的 8 位 ADC；③ 使用 12 位线性 ADC，之后使用一张“12 位进－8 位出”的数字查找表。这三种方法所需的非线性处理是一样的，只不过这种非线性处理出现在不同的地方：模拟电路中、ADC 中或者数字电路中。

(4) 人类对连续声音(如某件乐器发出的音符)的感知常被划分为三个部分：响度、音调和音品。响度度量的是声波强度，音调是声音中的基波频率。虽然这两个感知参量起到的某些

作用不好理解，但是却和易于描述的物理量之间存在着明确的关系。而音品就比较复杂，它是由信号中的谐波分量决定的。

(5) 语音合成和语音识别技术几乎都是基于人类语音产生模型的。大多数人类语声可分为“浊音”和“摩擦音”。当空气由肺压出、并经由声带、从口/鼻呼出时，就产生了“浊音”。声带是沿气流截面拉伸的两片薄片组织，声带正位于喉结之后。声带以 50～1000 Hz 频率的振动来响应不断变化的肌肉拉力，这样形成了周期性短促气流注入人的喉咙。元音就属于一种“浊音”。

(6) An audio signal is a one-dimensional acoustic wave. When an acoustic wave enters the ear, the eardrum vibrates, causing the tiny bones of the inner ear to vibrate along with it, sending nerve pulses to the brain. These pulses are perceived as sound by the listener.

(7) Music is an case of audio signal, but an very important one. Another important special case is speech. Human speech tends to be in the 600-Hz to 6000-Hz range. Speech is made up of vowels and consonants, which have different properties.

(8) With 44100 samples/sec of 16 bits each, an audio CD needs a bandwidth of 705.6 kbps for monaural and 1.411 Mbps for stereo. While this is lower than what video needs, it still takes almost a full T1 channel to transmit uncompressed CD quality stereo sound in real time.

(9) Pulse code modulation, as used within the telephone system, uses 8-bit samples made 8000 times per second. In North America and Japan, 7 bits are for data and 1 is for control; in Europe all 8 bits are for data. This system gives a data rate of 56000 bps or 64000 bps. With only 8000 samples/sec, frequencies above 4 kHz are lost.

(10) Digitized sound can be easily processed by computers in software. Dozens of programs exist for personal computers to allow users to record, display, edit, mix, and store sound waves from multiple sources. Virtually all professional sound recording and editing are digital nowadays.

3. 写作练习(答案略)

第 8 单 元

1. 完形填空

(1) To understand video, the two-dimensional image, a function of time, electron beam, the light intensity, called a frame, reconstruct the image, vary, scan lines, In both systems, is turned off, to, has, Rather than, Instead of, in order, that, at

(2) as, that, in, for, the three additive primary colors, a linear superposition of red, green and blue, must be combined into a single composite signal, various methods for displaying color, have, incompatible, existing, RGB is also not the most efficient scheme

2. 英汉互译

(1) 通过人眼成像系统，三维世界成像在视网膜上。视网膜上遍布着光接收细胞，这些细胞对 400～700 nm 的光有反应。在成像系统当中，我们用镜头和光敏器件组成照相机来模拟人类视觉观察世界的方式。

(2) 在计算机领域，显示器用图像宽、高方向上的像素数进行描述。常规电视机表示为 644×483，这意味着 483 条图像线。不过，显示系统都包含一些用于扫描的开销，因此，

常规视频扫描的总线数一定大于 483。

(3) 传统上，视频扫描系统是由其总线数(包括同步、消隐开销)、帧速率(以 Hz 为单位)和扫描方式(隔行扫描的 2∶1 或逐行扫描的 1∶1)来表征的。在北美和日本，使用 525/59.94/2∶1 扫描系统，其视频模拟带宽约为 5.5 MHz。在欧洲和亚洲，使用 625/50/2∶1 扫描系统，其模拟带宽约为 6.5 MHz。1125/60/2∶1 扫描系统用于高清晰度电视中，其模拟带宽约为 30MHz。

(4) 视频系统采用一个亮度分量和两个色彩分量的形式传送图像数据。亮度分量的传输是很重要的，因为在传输、处理和存储过程中引入的噪声也会引起和亮度相似的感知效应，而且覆盖整个“从黑到白”的范围。

(5) 尽管视网膜的形状大致是球体的一部分，但从局部解剖角度讲却是二维的。出于实际原因，我们在照相机中使用了平面像平面，而没有使用球面像平面。成像系统理论关注的是对像平面上能量连续分布情况的分析。在光学照相机的像平面处有光照下发生化学变化的胶片。光学胶片的有效成分含在一层由大小和形状仔细控制的微粒组成的薄层当中。如果微粒的密度足够大，就可以重现出具备很强原始场景感的图像。在胶片介质中的微粒越细小、排列密度越大，胶片记录空间细节的能力就越强。

(6) The human eye has the property that when an image appears on the retina, the image is retained for some number of milliseconds before decaying. If a sequence of images is drawn line by line at 50 images/sec, the eye does not notice that it is looking at discrete images. All video systems exploit this principle to produce moving pictures.

(7) All compression systems require two algorithms: one for compressing the data at the source, and another for decompressing it at the destination. In the literature, these algorithms are referred to as the encoding and decoding algorithms, respectively.

(8) The JPEG (Joint Photographic Experts Group) is used for compressing continuous-tone still pictures (e.g., photographs). JPEG was developed by photographic experts working under the joint auspices of ITU, ISO, and IEC.

(9) Aspect ratio is the ratio of image width to height. Conventional television has an aspect ratio of 4∶3. High-definition television uses a wider ratio of 16∶9. Cinema commonly uses 1.85∶1 or 2.35∶1.

(10) MPEG stands for Motion Picture Experts Group, is the name of family of standards used for coding audio-visual information (e.g., movies, video, music) in a digital compressed format. The major advantage of MPEG compared to other video and audio coding formats is that MPEG files are much smaller for the same quality. This is because MPEG uses very sophisticated compression techniques.

3. 写作练习(答案略)

第 9 单 元

1. 完形填空

(1) general-purpose, or, Such systems generally use microprocessors, custom-designed chips, as well as, operating systems, low-cost consumer products, very limited storage for

instructions, In such cases, to, loads its programs into RAM, are called "embedded computers"

(2) a single program, take place concurrently, transaction or message, for a multithreaded program to achieve true performance gains, in, to, calculation-intensive, symmetric, at, audio and video, reentrant code, so that

2. 英汉互译

(1) 硬件就是机器和设备(如处理器、磁盘、磁带、调制解调器和电缆等)。计算机在运行时既需要硬件，也需要软件，缺少哪一个都不行。硬件设计指定了计算机能够遵循的指令，而指令告诉计算机该做什么。

(2) 固件是不需要电源供电而能保存内容的存储芯片，包括 ROM、PROM、EPROM 和 EEPROM。在存储了程序代码之后，固件就成为一种"硬的软件"。

(3) 软件涉及逻辑和语言。软件要对不断变化的事件进行处理，它必须遵循某种逻辑方式。语言用来编写软件。分析和编程中的逻辑和语言往往比明确存储与传输需求更为复杂。

(4) 软件包是一种为向公众出售而开发的应用程序。软件包通常是为了满足大量用户而设计的。尽管通过对各种参数进行选择设定，程序也可以做到适合某个用户的喜好，但它不如定制软件更具个性。

(5) 进程是指计算机上某个用户一段特定代码的执行过程。假如爱丽丝在 8∶30 运行微软 Word 程序，而在 8∶36 再次运行，那么她产生了两个进程。如果爱丽丝和鲍伯在 8∶30 在同一台计算机上运行 Word 程序，他们也产生了两个进程。

(6) A real-time system is a computer system that responds to input signals fast enough to keep an operation moving at its required speed.

(7) Real-time operating system is an operating system designed for use in a real-time computer system.

(8) System software are programs used to control the computer and develop and run application programs. It includes operating systems, TP monitors, network operating systems and database managers.

(9) Hardware is "storage and transmission". The more memory and disk storage a computer has, the more work it can do. The faster the memory and disks transmit data and instructions to the CPU, the faster it gets done. A hardware requirement is based on the size of the databases that will be created and the number of users or applications that will be served at the same time.

(10) The software is instructions for the computer. A series of instructions that performs a particular task is called a "program". The two major categories of software are "system software" and "application software". System software is made up of control programs such as the operating system and database management system (DBMS). Application software is any program that processes data for the user (spreadsheet, word processor, etc.). A common misconception is that software is data. It is not. Software tells the hardware how to process the data. Software is "run". Data are "processed".

3. 写作练习(答案略)

第 10 单 元

1. 完形填空

(1) when, on, from, reference point, a high voltage, including, rather than, taking accurate measurements, share, integrated circuits, conduction paths, static electricity, leads, charges

(2) design verification, oscilloscope or logic analyzer, to, partner, stimulus, make up, device-under-test, real-world signals, industry-standard tool, can, for, Similarly, their designs' performance, with, from, a complete measurement solution

2. 英汉互译

(1) "波"可被定义为每隔一段时间重复出现、由变化数值构成的曲线。在自然界中，波是很常见的——如声波、脑电波、海浪、光波和电压波形等——这都是周期性出现的现象。信号源通常产生重复出现、可控的电气波形(通常为电压)。

(2) 波形具备许多特性，而其关键属性是和幅度、频率、相位相关的。波形的幅度特性、频率特性和相位特性是信号源用来优化波形的。

(3) 数字万用表是一种用来测量电学量(电压、电流、电阻)的仪表。数字万用表的用途广泛、形状多样、尺寸各异。模拟万用表在印制刻度上移动指针来显示结果，虽然指针移动有利于显示渐进的变化，但精确度却不如数字显示。数字万用表在数字液晶读出器或者发光二极管读出器上显示精确的测量结果。

(4) 示波器是一种在屏幕上显示电信号(波形和脉冲)的测量仪器。示波器能够产生时基用来进行信号测量。

(5) 在测量应用中，选择合适的探头是获得最佳信号保真度的关键。在高频测量中，采用有源探头可以获得更加真实的信号再现效果和保真度。

(6) Not all signals have their amplitude variations centered on a ground (0 V) reference. The "offset" voltage is the voltage between circuit ground and the center of the signal's amplitude.

(7) Differential signals are those that use two complementary paths carrying copies of the same signal in equal and opposite polarity (relative to ground).

(8) Rise and fall time are measures of the time it takes the signal edge to make a transition from one state to another. In modern digital circuitry, these values are usually several nanoseconds.

(9) Bandwidth means the transmission capacity of an electronic pathway such as a communications line or computer bus. In a digital line, it is measured in bits per second or bytes per second. In an analog line, it is the difference between the highest and lowest frequencies and is measured in Hertz.

(10) The noise is an extraneous signal that invades an electrical transmission. It can come from strong electrical or magnetic signals in nearby lines, from poorly fitting electrical contacts, and from power line spikes.

3. 写作练习(答案略)

附录 A　科技英语语音手册

A.1　数字读法

数值

-5	negative five / minus five
64	sixty-four
256	two hundred (and) fifty-six
8,192	eight thousand one hundred (and) ninety-two
4,294,967,296	four billion two hundred (and) ninety-four million nine hundred (and) sixty-seven thousand two hundred (and) ninety-six
1/2	a half / one half
1/3	a third / one third
1/4	a quarter / one fourth
1/5	a fifth / one fifth
2/3	two thirds / two over three / two by three
$5\frac{7}{13}$	five and seven thirteenth
4%	four per cent
4‰	four per mil
0.1	zero point one / naught point one
0.01	point naught one / point oh one
3.1415926535	three point one four one five nine two six five three five
0.333…	point three repeating / point three recurring
0.3%	point three percent / decimal three percent
1.23×10^{-4}	one point two three times ten to the minus four
2^3	two cubed
6^2	six squared
7^5	seven to the fifth power / seven to the fifth
$\sqrt{25}$	the square root of twenty-five / twenty-five to the one half power
$\sqrt[3]{27}$	the cube root of twenty-seven
$\sqrt[8]{32}$	the eighth root of thirty-two
30°	thirty degrees
21st	twenty-first
255th	two hundred (and) fifty-fifth

日期

221 B.C.	two twenty-one B C

79 A.D.	seventy-nine A D
1034	ten thirty-four
1999	nineteen ninety-nine
2000	the year two thousand
2012	two thousand (and) twelve
1950's	nineteen fifties
6 January,1974	the sixth of January nineteen seventy-four
January 6,1974	January (the) sixth of nineteen seventy-four

时间

10:56	ten fifty-six / four to eleven
9:25	nine twenty-five / twenty-five past nine
8:30 a.m.	eight thirty a m
5:45 p.m.	five forty-five p m
42h.15′38″	forty-two hours fifteen minutes thirty-eight seconds

号码

1223-9066	one two two three, nine zero six six /one double two three, nine zero double six
room 316	room three one six

其他

5:3	five to three
$9.25	nine dollars twenty-five cents
5 ft. 5 in	five foot five (inches)
10 gal. 2 qt.	ten gallons two quarts
8×6 feet	eight by six feet
9″×7×3	nine inches by seven by three
Vol. IV	Volume four

A.2 符号读法

数学符号

加号/正号	+	plus
减号/负号	-	minus
乘号	×	times / multiplied by
		times / multiplied by
除号	÷	divided by
	/	quotient of
正负号	±	plus or minus
等号	=	is equal to / equals
约等于	≈	is approximately equal to
全等号	≡	is equivalent to
不等号	≠	is not equal to

	＜＞	is not equal to
	！＝	is not equal to
大于号	＞	greater than
	≫	much greater than
小于号	＜	less than
	≪	much less than
小于或等于	≤	is equal to or less than / is less than or equal to
大于或等于	≥	is equal to or greater than / is greater than or equal to
不小于	≮	is not less than
不大于	≯	is not more than
无限大	∞	infinity
和……成比例	∝	varies as
因为	∵	since; because
所以	∴	hence
垂直	⊥	is perpendicular to
圆	⊙	circle
半圆	⌒	semicircle
圆周	○	circumference
三角形	△	triangle
角	∠	angle
度	°	degree
分	′	minute
秒	″	second
正弦	sin	sine
余弦	cos	cosine
正切	tan	tangent
余切	cot	cotangent
正割	sec	secant
余割	csc	cosecant

标点符号

句号	.	period
逗号	,	comma
冒号	:	colon
分号	;	semicolon
叹号	!	exclamation mark
撇号	’	apostrophe
问号	?	question mark
破折号	—	dash
省略号	…	ellipsis
圆括号	( )	parentheses

单引号	‘ ’	single quotation marks
双引号	“ ”	double quotation marks
方括号	[ ]	(square) brackets
大括号	{ }	braces

其他符号

分节符	§	section
编号符	#	number
位置符	@	at
联合符	&	and
点	.	dot
星号	*	asterisk / star
箭号	→	arrow
波浪号	~	tilde
双线号	‖	parallel
斜线	/	slash
反斜线	\	backslash

A.3 数学表达式读法

逻辑

$\exists$	there exists / there is a
$\forall$	for all / for every
$p \Rightarrow q$	p implies q / if p, then q
$p \Leftrightarrow q$	p if and only if q / p is equivalent to q

集合

$\varnothing$	(the) empty set / null set
A^c	the complement of A
$x \in A$	x belongs to A / x is an element (or a member) of set A
$x \notin A$	x does not belong to A / x is not an element of set A
$A \subset B$	A is contained in B / A is a subset of B
$A \supset B$	A contains B / B is a subset of A
$A \cap B$	A meet B / intersection of sets A and B
$A \cup B$	A join B / union of sets A and B
$A \backslash B$	A minus B / the difference between A and B
$A \times B$	A cross B / the Cartesian product of A and B

实数

$\lvert x \rvert$	absolute value of x; mod x / modulus x
$x+1$	x plus one
$x-1$	x minus one

$x \pm 1$	x plus or minus one
xy	x y / x multiplied by y
$(x-y)(x+y)$	x minus y, x plus y
$\frac{x}{y}$	x over y / x divided by y
$\frac{x+y}{z}$	x plus y, all divided z / x plus y, the quantity, divided by z
$x=5$	x equals five / x is equal to five
$x \neq 5$	x (is) not equal to five
$x \equiv y$	x is equivalent to y / x is identical to y
$x > y$	x is greater than y
$x \geqslant y$	x is greater than or equal to y
$x < y$	x is less than y
$x \leqslant y$	x is less than or equal to y
$0 < x < 1$	zero is less than x is less than one / x is between zero and one
$0 \leqslant x \leqslant 1$	zero is less than or equal to x
$x : y = u : v$	x is to y as u is to v
x^2	x squared / x (raised) to the power two
$4x^2$	four times x squared
x^3	x cubed
x^4	x to the fourth / x to the power four
x^n	x to the n th / x to the power n
x^{-1}	x inverse
x^{-n}	x to the (power) minus n
$\sqrt{x}$	(square) root x / the square root of x
$\sqrt[3]{x}$	cube root (of) x
$\sqrt[4]{x}$	fourth root (of) x
$\sqrt[n]{x}$	n th root (of) x
$(x+y)^2$	x plus y all squared
$\left(\frac{x}{y}\right)^2$	x over y all squared
$n!$	n factorial
$\hat{x}$	x hat
$\bar{x}$	x bar
$\tilde{x}$	x tilde
x_i	x i/x subscript i/x suffix i
$x_1, \cdots, x_n$	x one up to x n
$\sum_{i=1}^{n} a_i$	the sum from i equals one to n a i / the sum as i runs from one to n of the a

$\prod_{k=0}^{m} b_k$	the product from k equals zero to m bk / the product as k runs from zero to m of the b

复数

$\lvert z \rvert$	modulus of z
$\mathrm{Re}(z)$	real part of z
$\mathrm{Im}(z)$	imaginary part of z

线性代数

$\lVert x \rVert$	the norm of x / the modulus of x
$\boldsymbol{OA}$	$\boldsymbol{O}\ \boldsymbol{A}$ / vector $\boldsymbol{O}\ \boldsymbol{A}$
$\overline{\boldsymbol{OA}}$	$\boldsymbol{O}\ \boldsymbol{A}$ / the length of the segment $\boldsymbol{O}\ \boldsymbol{A}$
$\boldsymbol{A}^{\mathrm{T}}$	$\boldsymbol{A}$ transpose / the transpose of $\boldsymbol{A}$
$\boldsymbol{A}^{-1}$	$\boldsymbol{A}$ inverse / the inverse of $\boldsymbol{A}$
$<v, w>$	scalar product of v and w

微积分

$f(x)$	$f\ x$ / f of x / the function f of x
$f^{-1}(x)$	f inverse of x
$f: S \rightarrow T$	a function f from S to T
$x \mapsto y$	x maps to y / x is sent to y / x is mapped to y
$f'(x)$	f prime x / f dash x / the (first) derivative of f with respect to x
$f''(x)$	f double-prime x / f double-dash x / the second derivative of f with respect to x
$f'''(x)$	f triple-prime x / f triple-dash x / the third derivative of f with respect to x
$f^{(4)}(x)$	f four x / the fourth derivative of f with respect to x
$\frac{\partial f}{\partial x}$	the partial (derivative) of f with respect to x
$\frac{\partial^2 f}{\partial x^2}$	the second partial (derivative) of f with respect to x
$\int_0^{\infty}$	the integral from zero to infinity
$\int_a^b f(x)\mathrm{d}x$	the integral from a to b (of) $f\ x$ d x
$\lim_{x \rightarrow 0}$	the limit as x approaches zero
$\lim_{x \rightarrow +0}$	the limit as x approaches zero from above
$\lim_{x \rightarrow -0}$	the limit as x approaches zero from below
$\lim_{x \rightarrow \infty} \frac{x}{\sin x}$	the limit as x goes to infinity of x over sine x / the limit as x approaches infinity of x divided by sine x
$\sin^2(x)$	sine squared of x
$\sin(x^2)$	sine of x squared

$(\sin x)^2$	sine of x, quantity squared / sine of x all squared
$\tan^{-1}(x)$	inverse tangent of x
$\exp(x)$, e^x	exponential of x / e to the x
$\log(x)$	log x
$\log_b(x)$	log x to the base b
$\log_e(x)$, $\ln(x)$	log x to the base e / log to the base e of x / natural log (of) x

A.4 国际单位前缀发音表

数量级	前缀	缩写	发音	示例		
太(10^{12})	tera	T	[ˈterə]	terabyte	[ˈterəˌbaɪt]	太字节(TB)
吉(10^9)	giga	G	[ˈdʒɪgə]	gigahertz	[ˈdʒɪgəˌhɜːts]	吉赫(GHz)
兆(10^6)	mega	M	[ˈmegə]	megahertz	[ˈmegəˌhɜːts]	兆赫(MHz)
千(10^3)	kilo	k	[ˈkɪlə]	kilovolt	[ˈkɪləʊˌvəʊlt]	千伏(kV)
毫(10^{-3})	mili	m	[ˈmɪlɪ]	milliwatt	[ˈmɪlɪˌwɒt]	毫瓦(mW)
微(10^{-6})	micro	μ	[ˈmaɪkrəʊ]	microsecond	[ˈmaɪkrəʊˌsekənd]	微秒(μs)
纳(10^{-9})	nano	n	[ˈnænəʊ]	nanometer	[ˈnænəʊˌmiːtə]	纳米(nm)
皮(10^{-12})	pico	p	[ˈpɪkə]	picofarad	[ˌpɪkəˈfærəd]	皮法(pF)

A.5 首字母缩略语发音表

拼写特点	发音方法	示例
完全符合英语发音规则	完全按照英语单词发音	RAM [ræm] *abbr*. random-access memory 随机存取存储器 ROM [rɒm] *abbr*. read-only memory 只读存储器 GUI [ˈguːiː] *abbr*. graphical user interface 图形用户界面 LAN [læn] *abbr*. local-area network 局域网 WAN [wæn] *abbr*. wide-area network 广域网 BIOS [ˈbaɪɒs] *abbr*. basic input output system 基本输入/输出系统 POST [pəʊst] *abbr*. power-on self-test 开机自检 RISC [rɪsk] *abbr*. reduced-instruction-set computer 精简指令集计算机 SATA [ˈseɪtə] *abbr*. serial advanced technology attachment 串行高级技术附件 MIDI [ˈmɪdi] *abbr*. musical instrument digital interface 乐器数字接口
部分符合英语发音规则	部分按照英语单词发音，余者读字母本音	JPEG [ˈdʒeɪˌpeg] *abbr*. Joint Photographic Experts Group 联合图像专家组 MPEG [ˈemˌpeg] *abbr*. Motion Picture Experts Group 运动图像专家组 ASIC [ˈeɪsɪk] *abbr*. application-specific integrated circuit 专用集成电路 CMOS [ˈsiːˌməʊz] *abbr*. complementary metal oxide semiconductor 互补金属氧化物半导体 JTAG [ˈdʒeɪˌtæg] *abbr*. Joint Test Action Group 联合测试行动组 WLAN [ˈdʌbljuːˌlæn] *abbr*. wireless local-area network 无线局域网

续表

拼写特点	发音方法	示例
不符合英语发音规则	逐一读字母本音，重音落在最后一个字母上	UI [ˌjuːˈaɪ] *abbr*. user interface 用户界面 UPS [ˌjuːpiːˈes] *abbr*. uninterrupted power supply 不间断电源 CPU [ˌsiːpiːˈjuː] *abbr*. central processing unit 中央处理单元 DSP [ˌdiːesˈpiː] *abbr*. digital signal processing 数字信号处理 LCD [ˌelsiːˈdiː] *abbr*. liquid crystal display 液晶显示屏 FPGA [ˌefpiːdʒiːˈeɪ] *abbr*. field-programmable gate array 现场可编程门阵列

A.6 希腊字母发音表

大写	小写	英语拼写	英语发音	科技用途
Α	α	alpha	[ˈælfə]	角度；系数
Β	β	beta	[ˈbiːtə]	角度；系数；磁通系数
Γ	γ	gamma	[ˈgæmə]	电导系数
Δ	δ	delta	[ˈdeltə]	增量；密度；屈光度
Ε	ε	epsilon	[ˈepsɪlɒn]	
Ζ	ζ	zeta	[ˈziːtə]	系数；方位角；阻抗；相对黏度；原子序数
Η	η	eta	[ˈiːtə]	磁滞系数；效率
Θ	θ	theta	[ˈθiːtə]	温度；相位角
Ι	ι	iot	[aɪˈəʊtə]	
Κ	κ	kappa	[ˈkæpə]	介质常数
Λ	λ	lambda	[ˈlæmdə]	波长
Μ	μ	mu	[mjuː]	磁导系数；微(10^{-3})；放大倍数
Ν	ν	nu	[njuː]	磁阻系数
Ξ	ξ	xi	[ksaɪ]	粒子
Ο	o	omicron	[əʊˈmaɪkrən]	
Π	π	pi	[paɪ]	圆周率
Ρ	ρ	rho	[rəʊ]	电阻系数
Σ	σ	sigma	[ˈsɪgmə]	总和；表面密度；跨导
Τ	τ	tau	[tɔː]	时间常数
Υ	υ	upsilon	[juːˈpsaɪlən]	位移
Φ	φ	phi	[faɪ]	磁通；角度
Χ	χ	chi	[kaɪ]	
Ψ	ψ	psi	[psaɪ]	角度；角速度；介质电通量
Ω	ω	omega	[ˈəʊmɪgə]	欧姆；角速度；角度

附录 B　科技英语词汇手册

B.1　构词法

派生法

电阻器	resistor	resist→resistor
电感器	inducer	induce→inducer
导体	conductor	conduct→conductor
半导体	semiconductor	conduct→semiconductor
导电性	conductivity	conduct→conductivity
绝缘体	insulator	insulate→insulator

组合法

软件	software	soft＋ware→software
波形	waveform	wave＋form→waveform
带宽	bandwidth	band＋width→bandwidth
主板	motherboard	mother＋board→motherboard
配置	setup	set＋up→setup
网表	netlist	net＋list→netlist

截短法

照片	photo	**photo**graph→photo
冰箱	fridge	re**frige**rator→fridge
传真	fax	**facs**imile→fax
信息	info	**info**rmation→info
技术	tech	**tech**nology→tech
应用程序	app	**app**lication→app

拼缀法

比特	bit	**b**inary **i**nformation uni**t**→bit
像素	pixel	**pic**ture **el**ement→pixel
声码器	vocoder	**vo**ice **coder**→vocoder
系统生成	sysgen	**sys**tem **gen**eration→sysgen
编解码器	codec	**co**der-**dec**oder→codec
调制解调器	modem	**mo**dulator-**dem**odulator→modem

首字母缩略法

雷达	radar	**ra**dio **d**etection **a**nd **r**anging→radar
声呐	sonar	**so**und **n**avigation **a**nd **r**anging→sonar
激光	laser	**l**ight **a**mplification by **s**timulated **e**mission of **r**adiation→laser

液晶显示器	LCD	**l**iquid **c**rystal **d**isplay→LCD
发光二极管	LED	**l**ight-**e**mitting **d**iode→LED
电荷耦合器件	CCD	**c**harge-**c**oupled **d**evice→CCD

B.2 构词成分(词根、前缀、后缀)

构词成分	含义、用途
agogue，agog	领导的；激励的
al	关于
ali	其他的
ambi	并；双方都……；两者皆……；既……又……
animus，anima	精神；理智；灵魂；心灵
anni，annu，enni	年
anthrop，anthropo	男人的；人类的
aqua，aqui	水
ar，er，or	某人；某物
arch	统治的；主要的
archae，archaeo	古代的；古老的
astro	星；星球；天体
audi	听到；听；审理
aut，auto	自己
bello，belli	战争；冲突；对抗
bene	好处；善行；慷慨的；虔诚的
bi	两个
biblio	书
bio	生命；生活
brevi	短的；摘要；概要
cap，cep	带走；接收
capit	头；主要的
cata	向下；相反；离开
cede，ceed	放弃；割让
cent，centi	成百的；百分之一的
cess	走
chron，chrono	时间
cide	谋杀；杀人
civ，civis	公民；市民；老百姓
contra	反对
corp，corpor	身体；主体
cred	相信
crypto，crypt	秘密；机密；隐藏
cura	照管；看护；监护；关心；关怀

cycl，cyclo	圆圈；周期；循环；轮子
de	离开；除去；降低；完全
dec，deca	十
dem，demo	人们；公民；民族
derm，dermo	皮肤
dia	通过；横跨；在……之间；穿越
dic，dict	说；讲
dis	离去；分开；不
drom	睡；睡眠
duc	领导；引导；致使
ego	我；自己；自我；本质
em，en	里面；深入；进入
epi	在……上；在……周围；在……后面
equi	平等的；相等的；匹配；比得上
err	徘徊；迷路；离题
ex-	前任的或以前的
ex，e	向外；从；无
fac，fect，fic	使得；制造；进行
feder，fid，fide	信任；信赖；信念
fer	带来；拿来；承受；忍受；屈服；放弃
fin	结束；范围
frater，fratr	兄弟；同事；战友
gamy	结婚；婚姻；合并
gen，geno	种族的；种类的；血统的
geo	陆地；泥土
gnosi，gnosis	知识；学问；认知
gram	写；画图；记录
graph	写出来的东西；机器
gyn，gyno	妇女；女性
hom，homo	相同的；男人的；人类的
hyper	越过；结束；在……之上的；过多的
hypn，hypno	睡；睡眠
hypo	下面；次等
in，im，il，ir	无；不
ist	……的人；……者
leg，legis，lex	法律
less	无；没有；不
loc，loco	放置；地方
log，logo	语言；词汇
luc，lum	点亮；光亮；明亮；清楚；清晰
magna	大的；伟大的；好的

mal	坏；恶；不良；不当；不完全
man，manu	手工
meter	测量；计量
micro	非常小
milli	成千的；千分之一的
mis，miso	厌恶；错误
mitt，miss	送；发送；派出
mon，mono	单个
mors，mort	死亡
nasc，nat	出生；天生的
naut	水手；海员
nov	新的；新鲜的
ology	一门学问；一种学科
omni	所有
onym，nomin	名字；名称；命名
pac，pax	和平；平静；和睦；秩序
pathy	感觉；触觉；苦楚
ped，pedo	小孩
ped，pod	脚
peri	四处；周围；大约；附近
phob，phobo	害怕；恐怕
phon，phono	声音
polis	城市；都市；政治
poly	许多的
pop	居住于；人类；民族
port	搬运；携带
pos，pon	放置；使位于；布置；规定的
post	在……之后；之后的；过去的
poten	强大的；强有力的
pro	向前；在前
pseudo	错误的；伪造的
re	再；回
ri，ridi，risi	笑；笑声
sci，scio	知道；认知
scope	看到；观察
scrib，scrip	写
semi	半；不完全；一部分；一段时间内两次的
sequi	接着；次序
sist，sta	站立；立场；位于；停止；抵抗
spect	看；观察
soph	使知道；教导；聪明的；博学的

sub	低于；在底下；在下面；等级低
syl，sym，syn	相同；一起；沿着
tele	来自远方
temp，tempo，tempor	时间；定期的
ten，tent，tain	拿住；抓住
tend，tens，tent	伸展；延伸；拉紧
tox，toxo	毒药；污染；中毒
uni	一个
ven，veni，vent	来；开始；变成
vis，vid	看见
voc，vox	表达；吐露；声音；呼叫；召集

B.3 学术派生词表

拼写	读音	词义
ab**brevi**ation	[əbriːvɪˈeɪʃn]	*n*.缩写；缩写词
ac**cred**itation	[əˌkredɪˈteɪʃn]	*n*.委派；信赖
af**fect**	[əˈfekt]	*vt*.影响
af**fin**ity	[əˈfɪnɪti]	*n*.密切关系；吸引力；类同
a**gnos**tic	[ægˈnɒstɪk]	*n*.不可知论者　*adj*.不可知论的
alias	[ˈeɪlɪəs]	*n*.别名；化名
alien	[ˈeɪlɪən]	*n*.外侨　*adj*.外国的
alienate	[ˈeɪlɪəneɪt]	*vt*.使疏远
al**loc**ate	[ˈæləkeɪt]	*vt*.分配；配给；拔出
ambiguous	[æmˈbɪgjʊəs]	*n*.有两种或多种解释　*adj*.模棱两可的
a**mort**ize	[əˈmɔːtaɪz]	*vt*.摊销；分期偿还
ana**chron**ism	[əˈnækrəˌnɪzəm]	*n*.时代错误；不合潮流的人或物
animate	[ˈænɪmeɪt]	*vt*.使活泼；鼓舞；推动　*adj*.有生命的
anniversary	[ænɪˈvɜːsəri]	*n*.周年纪念
annual	[ˈænjʊəl]	*adj*.每年的；年度的
an**onym**ous	[əˈnɒnɪməs]	*adj*.无名的；匿名的
anthropoid	[ˈænθrəpɒɪd]	*adj*.类人猿的；类人的
anthropology	[ˌænθrəˈpɒlədʒi]	*n*.人类学
anti**pathy**	[ænˈtɪpəθi]	*n*.反感；厌恶；不相容
anti**tox**in	[æntɪˈtɒksɪn]	*n*.抗毒素；抗毒素血清
ant**onym**	[ˈæntənɪm]	*n*.反义词
a**pathy**	[ˈæpəθɪ]	*n*.无感情；无感觉；冷淡
a**pod**al	[ˈæpədəl]	*adj*.无足的；无腹鳍的
aqua**naut**	[ˈækwənɔːt]	*n*.海底潜水员
aquarium	[əˈkweərɪəm]	*n*.养鱼池；水族馆
aquatic	[əˈkwætɪk]	*adj*.长在或栖于水中的；在水中或水上举行的

archae**ology**	[ˌɑːkɪˈɒlədʒi]	*n*.考古学
archetype	[ˈɑːkɪtaɪp]	*n*.原型；典范；样板
astro**naut**	[ˈæstrənɔːt]	*n*.宇航员
astronomy	[əˈstrɒnəmi]	*n*.天文学
at**tent**ion	[əˈtenʃn]	*n*.注意力；关心；立正姿势
audible	[ˈɔːdəbl]	*adj*.可听见的
audience	[ˈɔːdɪəns]	*n*.听众；观众
audi**ology**	[ˌɔːdɪˈɒlədʒi]	*n*.听觉学
audio**meter**	[ˌɔːdɪˈɒmɪtə]	*n*.听度计；测听计
audiovisual	[ˌɔːdɪəʊˈvɪzjʊəl]	*adj*.听觉视觉的
audition	[ɔːˈdɪʃn]	*n*.试听；听　*v*.试听
auditorium	[ɔːdɪˈtɔːrɪəm]	*n*.礼堂；音乐厅
auto**bio**graphy	[ɔːtəbaɪˈɒgrəfi]	*n*.自传
auto**graph**	[ˈɔːtəgrɑːf]	*n*.亲笔签名
a**voc**ation	[ˌævəʊˈkeɪʃn]	*n*.副业；业余爱好；嗜好
belligerent	[bəˈlɪdʒərənt]	*adj*.交战的；好战的；交战国的　*n*.交战国
benediction	[ˌbenɪˈdɪkʃn]	*n*.祝福；赐福
benefactor	[ˈbenɪfæktə]	*n*.捐助人
beneficiary	[benɪˈfɪʃəri]	*n*.受益者；封臣　*adj*.拥有封地的
benefit	[ˈbenɪfɪt]	*n*.好处；利益　*v*.有益于
bi**annu**al	[baɪˈænjʊəl]	*adj*.一年两次的
bibliography	[ˌbɪblɪˈɒgrəfi]	*n*.书目提要；文献目录
bi**cent**ennial	[baɪsenˈtenɪəl]	*adj*.200 年纪念的；200 年一次的
biennial	[baɪˈenɪəl]	*adj*.两年一次的；持续两年的
bimonthly	[baɪˈmʌnθli]	*adj*.双月的；一月两次的
bio**graph**er	[baɪˈɒgrəfə]	*n*.传记作者
biography	[baɪˈɒgrəfi]	*n*.人物传记；生平
biology	[baɪˈɒlədʒi]	*n*.生物学
bi**ped**	[ˈbaɪped]	*n*.两足动物
bi**sexual**	[baɪˈseksʃʊəl]	*adj*.两性的；雌雄同体的
bi**week**ly	[baɪˈwiːkli]	*adj*.两周一次的；一周两次的
capable	[ˈkeɪpəbl]	*adj*.有能力的，有才能的
capital	[ˈkæpɪtl]	*n*.首都；大写字母　*adj*.极好的
cata**log**	[ˈkætəlɒg]	*n*.目录；一览表　*vt*.把……编目录
centennial	[senˈtenɪəl]	*adj*.百年纪念的；百年的
centi**meter**	[ˈsentɪmiːtə]	*n*.厘米
centi**pede**	[ˈsentɪpiːd]	*n*.百足虫(蜈蚣)
century	[ˈsentʃəri]	*n*.一世纪；百年
chronic	[ˈkrɒnɪk]	*adj*.慢性的；长期的；习惯性的
chronological	[krɒnəˈlɒdʒɪkl]	*adj*.按年代顺序排列的
chrono**meter**	[krəˈnɒmɪtə]	*n*.高度精确的钟表；航海经线仪

circumference	[sə'kʌmfərəns]	*n*.圆周的长度；圆周
civil	['sɪvəl]	*adj*.公民的；文职的；有礼貌的；根据民法的
civilian	[sɪ'vɪliən]	*adj*.民用的；百姓的　*n*.平民
civilization	[ˌsɪvɪlaɪ'zeɪʃn]	*n*.文明；文化
con**cede**	[kən'siːd]	*vt*.承认；退让；给予　*vi*.让步
con**cess**ion	[kən'seʃn]	*n*.让步；特许；承认；退位
con**duc**tor	[kɒn'dʌktə]	*n*.导体；售票员；领导者；管理人
con**fer**ence	['kɒnfərəns]	*n*.商谈；会议
con**fide**	[kən'faɪd]	*vt*.吐露；委托　*vi*.信赖；吐露秘密
con**sequ**ence	['kɒnsɪkwəns]	*n*.结果；重要性；推论
con**sist**	[kən'sɪst]	*vi*.组成；在于；符合
con**tempor**ary	[kən'tempərəri]	*n*.同时代的人或物　*adj*.当代的；同时代的
con**tent**	[kən'tent]	*n*.内容；容量　*adj*.满意的　*vt*.使满足
contradiction	[kɒntrə'dɪkʃn]	*n*.矛盾
contrary	['kɒntrərɪ]	*adj*.相反的
contrast	['kɒntrɑːst]	*n*.差异；对比　*vt*.对照
con**vene**	[kən'viːn]	*vt*.集合
con**veni**ent	[kən'viːnɪənt]	*adj*.合适的；便利的
con**ven**tion	[kən'venʃn]	*n*.会议；习俗；常规
con**voc**ation	[ˌkɒnvə'keɪʃn]	*n*.召集；集会；评议会
corporation	[kɔːpə'reɪʃn]	*n*.公司；社团；法人
credential	[krɪ'denʃl]	*n*.证书；凭据
credible	['kredəbl]	*adj*.可信赖的
credit	['kredɪt]	*n*.相信；名誉　*vt*.相信；信任
creditor	['kredɪtə]	*n*.债权人；贷方
cryptic	['krɪptɪk]	*adj*.神秘的；含义模糊的；隐藏的
curator	[kjʊə'reɪtə]	*n*.监护人；管理者
decade	['dekeɪd]	*n*.10 年
deca**meter**	['dekəˌmiːtə]	*n*.10 米；10 公尺
decimal	['desɪml]	*n*.小数　*adj*.十进位的
decimate	['desɪmeɪt]	*vt*.取十分之一；大批杀害
deci**meter**	['desɪˌmiːtə]	*n*.十分之一米；公寸
decode	[diː'kəʊd]	*vt*.译码
decolor	[diː'kʌlə]	*vi*.使脱色
de**duc**tion	[dɪ'dʌkʃn]	*n*.扣除；减除；推论；减除额
de**fer**	[dɪ'fɜː]	*vi*.推迟；服从　*vt*.使推迟
de**fer**ence	['defərəns]	*n*.顺从；尊重
de**fin**itive	[dɪ'fɪnətɪv]	*n*.限定词　*adj*.决定性的；最后的
demo**graphy**	[dɪ'mɒgrəfi]	*n*.人口统计学
de**port**ment	[dɪ'pɔːtmənt]	*n*.举止；行为；态度
dermatologist	[ˌdɜːmə'tɒlədʒɪst]	*n*.皮肤专家

de**scrip**tion	[dɪˈskrɪpʃn]	*n*.描述
dia**gram**	[ˈdaɪəgræm]	*n*.图解；图表
dia**lect**	[ˈdaɪəlekt]	*n*.方言；行话　*adj*.方言的
dia**log**ue	[ˈdaɪəlɒg]	*n*.对话
dia**meter**	[daɪˈæmɪtə]	*n*.直径
dictation	[dɪkˈteɪʃn]	*n*.听写；口授笔录
diction	[ˈdɪkʃn]	*n*.措词
dictionary	[ˈdɪkʃənrɪ]	*n*.词典；字典
dis**approve**	[dɪsəˈpruːv]	*vt*.不赞成；不同意
dis**miss**	[dɪsˈmɪs]	*vt*.解散；解雇；不予理会　*vi*.解散
dormant	[ˈdɔːmənt]	*adj*.休眠的；静止的；隐匿的
ef**fect**	[ɪˈfekt]	*n*.效果；结果；后果
ef**fec**tive	[ɪˈfektɪv]	*adj*.有效的
em**pathy**	[ˈempəθi]	*n*.执着；身临其境的感觉
en**vision**	[enˈvɪʒn]	*vt*.想象；预想
epi**derm**is	[ˌepɪˈdɜːmɪs]	*n*.上皮；表皮
epi**log**ue	[ˈepɪlɒg]	*n*.结尾部分；后记
equivalent	[ɪˈkwɪvələnt]	*adj*.等价的；相等的；同意义的
erratic	[ɪˈrætɪk]	*adj*.不稳定的；古怪的
error	[ˈerə]	*n*.误差；错误；过失
e**vid**ence	[ˈevɪdəns]	*n*.证据
e**vid**ent	[ˈevɪdənt]	*adj*.明显的；明白的
ex**cep**tion	[ɪkˈsepʃn]	*n*.除外；异议
ex**clude**	[ɪkˈsklɪuːd]	*vt*.把……排除在外
ex**port**	[ɪkˈspɔːt]	*vt*.运走；输出
ex**tempor**aneous	[ɪkˌstempəˈreɪnɪəs]	*adj*.临时的；无准备的
fac**simil**e	[fækˈsɪmɪli]	*n*.传真　*adj*. 复制的　*vt*.传真；临摹
federal	[ˈfedərəl]	*adj*.联邦的；同盟的
fertile	[ˈfɜːtaɪl]	*adj*.肥沃的；多产的
final	[ˈfaɪnl]	*adj*.决定性的；最终的
finale	[fɪˈnɑːli]	*n*.结局；尾声
finite	[ˈfaɪnaɪt]	*adj*.有限的；能测量的
general	[ˈdʒenərəl]	*adj*.总的；全面的
generate	[ˈdʒenəreɪt]	*vt*.使形成；发生
generic	[dʒɪˈnerɪk]	*adj*.类的；一般的
genus	[ˈdʒiːnəs]	*n*.类；种；属
geography	[dʒɪˈɒgrəfi]	*n*.地理学
geologist	[dʒɪˈɒlədʒɪst]	*n*.地质学家；地质学者
geo**metry**	[dʒɪˈɒmɪtri]	*n*.几何学
gynecologist	[ˌgaɪnɪˈkɒlədʒɪst]	*n*.妇科医生
harmless	[ˈhɑːmləs]	*adj*.无害的；无恶意的

homogeneous	[ˌhɒmə'dʒiːnɪəs]	*adj*.相同的；同性的
homo**graph**	['hɒməgrɑːf]	*n*.同形异义字
hom**onym**	['hɒmənɪm]	*n*.同音异义词
hyper**tension**	[haɪpə'tenʃn]	*n*.高血压；过度紧张
hypnosis	[hɪp'nəʊsɪs]	*n*.催眠；催眠状态
hypo**derm**ic	[haɪpə'dɜːmɪk]	*n*.皮下注射；皮下注射器　*adj*.皮下的
hypo**thesis**	[haɪ'pɒθɪsɪs]	*n*.假设
il**leg**al	[ɪ'liːgəl]	*adj*.不合法的
im**perfect**	[ɪm'pɜːfɪkt]	*adj*.不完全的；不完善的
im**port**	[ɪm'pɔːt]	*vt*.进口；输入　*n*.输入品
im**potent**	['ɪmpətənt]	*adj*.无行动能力的；无力的
in**corpor**ate	[ɪn'kɔːpəreɪt]	*vt*.合并；使结合
in**cred**ible	[ɪn'kredəbl]	*adj*.难信的
in**fer**ence	['ɪnfərəns]	*n*.推理；推断
in**finite**	['ɪnfɪnət]	*adj*.无限的；不能测量的
in**finite**simal	[ˌɪnfɪnɪ'tesɪml]	*adj*.无穷小的　*n*.无限小；极微量
in**nat**e	[ɪ'neɪt]	*adj*.先天的；固有的
in**nov**ation	[ɪnə'veɪʃn]	*n*.创新；革新
inter**cede**	[ˌɪntə'siːd]	*vi*.调解；求情
inter**cept**	[ˌɪntə'sept]	*vt*.拦截；窃听　*n*.拦截；截获的情报
inter**miss**ion	[ɪntə'mɪʃn]	*n*.幕间休息；暂停；中断
inter**mitt**ent	[ɪntə'mɪtənt]	*adj*.间歇的；断断续续的
inter**vene**	[ɪntə'viːn]	*vi*.干涉；调停
in**vis**ible	[ɪn'vɪzəbl]	*adj*.看不见的
kilo**meter**	['kɪləˌmiːtə]	*n*.千米
legal	['liːgl]	*adj*.法律上的；合法的
legislature	['ledʒɪsletʃə]	*n*.立法机关
local	['ləʊkl]	*adj*.本地的；局部的
location	[ləʊ'keɪʃn]	*n*.位置；地方；定位
logical	['lɒdʒɪkl]	*adj*.符合逻辑的
lucid	['luːsɪd]	*adj*.明晰的；透明的；易懂的
magnanimous	[mæg'nænɪməs]	*adj*.宽宏大量的；宽大的
magnify	['mægnɪfaɪ]	*vt*.放大；赞美；夸大
main**tain**	[meɪn'teɪn]	*vt*.维持；继续；主张；供养
mal**function**	[mæl'fʌ kʃn]	*n*.故障；疾病
manipulation	[məˌnɪpjʊ'leɪʃn]	*n*.操纵；操作；处理；篡改
manual	['mænjʊəl]	*adj*.用手操作的　*n*. 手册；指南
manufacture	[mænjʊ'fæktʃə]	*n*.制造；加工
manuscript	['mænjʊskrɪpt]	*adj*.手写的　*n*.抄本；原稿；手稿
megalo**polis**	[ˌmegə'lɒpəlɪs]	*n*.特大都市；人口稠密地带
metro**poli**tan	[metrə'pɒlɪtən]	*n*.大城市人；大主教

micro**scope** [ˈmaɪkrəskəʊp] *n*.显微镜
milli**meter** [ˈmɪlɪˌmiːtə] *n*.毫米
million [ˈmɪljən] *n*.百万；无数 *adj*.百万的；很多的
mission [ˈmɪʃn] *n*.使命；任务；代表团； *vt*.派遣
mono**rail** [ˈmɒnəʊreɪl] *n*.单轨
mono**syllable** [ˈmɒnəsɪləbl] *n*.单音节词
monotone [ˈmɒnətəʊn] *adj*.单调的
monotonous [məˈnɒtənəs] *adj*.单调的；无变化的
mortal [ˈmɔːtl] *adj*.必死的；致命的
mortality [mɔːˈtælɪti] *n*.必死的状态；死亡率；死亡比例
mortgage [ˈmɔːgɪdʒ] *n*.抵押 *v*.抵押
mortify [ˈmɔːtɪfaɪ] *vt*.抑制；使……感到屈辱 *vi*. 约束
nature [ˈneɪtʃə] *n*.自然；性质；本性；种类
nautical [ˈnɔːtɪkl] *adj*.航海的；海上的
novel [ˈnɒvl] *adj*.新奇的；异常的
novice [ˈnɒvɪs] *n*.初学者；新手
omnibus [ˈɒmnɪbəs] *adj*.综合的；全集；总括的
omnipresent [ɒmnɪˈprezənt] *adj*.无所不在的
pedestrian [pəˈdestrɪən] *n*.步行者
pediatrician [ˌpiːdɪəˈtrɪʃn] *n*.儿科医生
per**cep**tion [pəˈsepʃn] *n*.知觉
per**cept**ive [pəˈseptɪv] *adj*.感知的；知觉的
peri**meter** [pəˈrɪmɪtə] *n*.周长；周界
period [ˈpɪərɪəd] *n*.周期；期间；时期 *adj*.某一时代的
periodic [ˌpɪərɪˈɒdɪk] *adj*.周期的；定期的
periodical [pɪərɪˈɒdɪkl] *adj*.周期的；定期的
peri**scope** [ˈperɪskəʊp] *n*.潜望镜
per**sist** [pəˈsɪst] *vi*.存留；坚持；持续 *vt*.坚持说；反复说
personi**fic**ation [pəˌsɒnɪfɪˈkeɪʃn] *n*.人格化；拟人法；象征
philo**soph**y [fɪˈlɒsəfi] *n*.哲学；哲理；人生观
phobia [ˈfəʊbɪə] *n*.恐惧；憎恶
podiatrist [pəˈdaɪətrɪst] *n*.足病医生
polygon [ˈpɒlɪgən] *n*.多边形；多角形物体
popular [ˈpɒpjʊlə] *adj*.流行的；通俗的；普及的
population [pɒpjʊˈleɪʃn] *n*.人口
port [pɔːt] *n*.港口；装卸货口岸
portable [ˈpɔːtəbl] *adj*.便于携带的；可搬运的
position [pəˈzɪʃn] *n*.位置；方位；职位；工作
post**nat**al [pəʊstˈneɪtl] *adj*.出生后的
post**script** [ˈpəʊstskrɪpt] *n*.附言；后记
posture [ˈpɒstʃə] *n*.姿势；态度；情形

potent	[ˈpəʊtənt]	*adj*.有力的；有影响的
potential	[pəˈtenʃl]	*n*.可能；潜力　*adj*.可能的；潜在的
pre**cede**	[prɪˈsiːd]	*vt*.领先　*vi*.领先；在前面
pre**fer**ence	[ˈprefrəns]	*n*.优先选择　*adj*.偏爱的；特别喜爱的
pre**nat**al	[priːˈneɪtl]	*adj*.产前的；胎儿期的
pro**ceed**	[prəˈsiːd]	*vi*.开始；继续进行
pro**cess**ion	[prəˈseʃn]	*n*.行列；列队行进　*vi*.列队行进　*vt*.沿着……行进
pro**fic**ient	[prəˈfɪʃnt]	*adj*.熟练的；精通的　*n*.精通；专家；能手
pro**gnosis**	[prɒgˈnəʊsɪs]	*n*.预知
pro**log**ue	[ˈprəʊlɒg]	*n*.序言；开场白；开端
pro**pos**al	[prəˈpəʊzl]	*n*.提议；建议；求婚
provision	[prəˈvɪʒn]	*n*.准备；预备；措施
pseud**onym**	[ˈsjuːdənɪm]	*n*.笔名；假名
pseudo**pod**ium	[ˌsjuːdəʊˈpəʊdɪəm]	*n*.虚足；假肢
re**cep**tion	[rɪˈsepʃn]	*n*.接待
re**cess**ion	[rɪˈseʃn]	*n*.衰退；后退；凹处
re**fer**ence	[ˈrefrəns]	*n*.委托或受委托；提交；参考人
re**write**	[riːˈraɪt]	*vt*.重写；改写
ridicule	[ˈrɪdɪkjuːl]	*n*.嘲笑；愚弄
ridiculous	[rɪˈdɪkjələs]	*adj*.可笑的；反常的
script	[skrɪpt]	*n*.草写体；手稿；脚本
se**cede**	[sɪˈsiːd]	*vi*.脱离；退出　*vt*. 从……中脱离；割让
semi**annu**al	[ˌsemɪˈænjʊəl]	*adj*.一年两次的；每半年一次的
sequence	[ˈsiːkwəns]	*n*.顺序；续发事件　*vt*. 按顺序排好
sophisticated	[səˈfɪstɪkeɪtɪd]	*adj*.复杂的；精致的；久经世故的；富有经验的
spectacle	[ˈspektəkl]	*n*.公开展示；场面；怪相；洋相
spectacular	[spekˈtækjʊlə]	*adj*.引人注目的；非凡的；惊人的
spectator	[spekˈteɪtə]	*n*.观众；旁观者
sub**marine**	[ˈsʌbməriːn]	*n*.潜水艇
sub**sequ**ent	[ˈsʌbsɪkwənt]	*adj*.后来的；随后的
suc**ceed**	[səkˈsiːd]	*vi*. 成功；继承　*vt*. 继承；接替
suf**fer**	[ˈsʌfə]	*vt*.受苦或忧伤
sus**cept**ible	[səˈseptəbl]	*adj*.易受影响的　*n*.易得病的人
syllable	[ˈsɪləbl]	*n*.音节　*vt*.划分音节　*vi*.按音节发音；讲话
symbol	[ˈsɪmbl]	*n*. 象征；符号；标志
symmetry	[ˈsɪmətri]	*n*.整齐；匀称
sympathy	[ˈsɪmpəθi]	*n*.同情；同感
symphony	[ˈsɪmfəni]	*n*.交响乐；谐声；和声
symptom	[ˈsɪmptəm]	*n*.症状；征兆
syn**chron**ize	[ˈsɪ krənaɪz]	*vt*.使……合拍　*vi*.同步；同时发生
syn**onym**	[ˈsɪnənɪm]	*n*.同义词

synthesis	[ˈsɪnθɪsɪs]	*n*.综合；综合体
tele**gram**	[ˈtelɪgræm]	*n*.电报
tele**graph**	[ˈtelɪgrɑːf]	*n*.电报；打电报
tele**phone**	[ˈtelɪfəʊn]	*n*.电话； *vi*.打电话
tele**scope**	[ˈtelɪskəʊp]	*n*.望远镜
tele**vis**ion	[ˈtelɪvɪʒn]	*n*.电视；电视机
temporary	[ˈtempərərɪ]	*adj*.暂时的 *n*.临时工
tension	[ˈtenʃn]	*n*.张力；拉力；紧张 *vt*.使紧张；使拉紧
toxic	[ˈtɒksɪk]	*adj*.有毒的；中毒的
toxicologist	[ˌtɒksɪˈkɒlədʒɪst]	*n*.毒理学家
trans**fer**	[trænsˈfɜː]	*vt*.搬；传递 *n*.转送；调任
trans**luc**ent	[trænsˈluːsnt]	*adj*.透明的；半透明的
uni**form**	[ˈjuːnɪfɔːm]	*adj*.相同的；一致的
union	[ˈjuːniən]	*n*.联合；团结；联盟；联合会
unique	[juːˈniːk]	*adj*.唯一的；独一无二的
unison	[ˈjuːnɪsn]	*n*.一致；调和；齐声
universal	[juːnɪˈvɜːsl]	*adj*.普遍的；通用的
universe	[ˈjuːnɪvɜːs]	*n*.宇宙；万象；万物；全人类
un**ten**able	[ʌnˈtenəbl]	*adj*.站不住脚的；难以防守的
venture	[ˈventʃə]	*vt*.敢于 *n*.风险 *vi*.投机
visage	[ˈvɪzɪdʒ]	*n*.面貌；容颜
visible	[ˈvɪzəbl]	*adj*.可见的；明显的；明白的
vision	[ˈvɪʒn]	*n*.视觉
visionary	[ˈvɪʒənri]	*adj*.梦想的；幻影的
visor	[ˈvaɪzə]	*n*.遮阳板 *vt*.遮护
vocabulary	[vəˈkæbjələri]	*n*.词汇；词表
vocation	[vəʊˈkeɪʃn]	*n*.职业；天职

B.4 常用信息技术缩略语

A

ACF	auto-correlation function 自相关函数
ADC	analog-to-digital converter 模数转换器
ADPCM	adaptive differential pulse code modulation 自适应差分脉冲编码调制
ADSL	asymmetrical digital subscriber line 非对称数字用户线路
AES	Advanced Encryption Standard 高级加密标准
AFC	automatic frequency control 自动频率控制
AFE	analog front end 模拟前端
AGC	automatic gain control 自动增益控制
AHDL	analog hardware-description language 模拟硬件描述语言
AI	artificial intelligence 人工智能

ALU arithmetic/logic unit 算术逻辑单元
AM amplitude modulation 调幅
ANSI American National Standards Institute 美国国家标准学会
APD avalanche photo diode 雪崩光电二极管
API application program interface 应用程序接口
ASIC application-specific integrated circuit 专用集成电路
ASSP application-specific standard product 专用标准产品
ATM asynchronous transfer mode 异步传输模式
AWG American wire gauge 美国线规

B

B-CDMA broadband code-division multiple access 宽带码分多址
BER bit error rate 误码率
BFSK binary frequency-shift keying 二进制频移键控
BGA ball grid array 球栅阵列封装
BJT bipolar junction transistor 双极性结型晶体管
BPF band-pass filter 带通滤波器
BPSK binary phase-shift keying 二进制相移键控

C

CAD computer-aided design 计算机辅助设计
CAM content-addressable memory 内容寻址存储器
CAM computer-aided manufacturing 计算机辅助制造
CAN controller area network 控制器局域网
CCD charge-coupled device 电荷耦合器件
CCF cross-correlation function 互相关函数
CCS common channel signaling 公共信道信令
CDMA code-division multiple access 码分多址
CELP code-excited linear prediction 码激励线性预测编码
CISC complex-instruction-set computer 复杂指令集计算机
CMOS complementary metal oxide semiconductor 互补金属氧化物半导体
CMRR common-mode rejection ratio 共模抑制比
codec coder-decoder 编解码器
CPLD complex programmable logic device 复杂可编程逻辑器件
CRC cyclic redundancy check 循环冗余校验码
CRT cathode-ray tube 阴极射线管
CSMA/CD carrier-sense multiple access/collision detection 载波侦听多路访问/冲突检测
CW continuous wave 连续波

D

DAB digital audio broadcasting 数字音频广播
DAC digital-to-analog converter 数模转换器
DAS data acquisition system 数据采集系统

DBPSK differential binary phase shift keying 差分二进制相移键控
DCS digital communications system 数字通信系统
DCT discrete cosine transform 离散余弦变换
DDR double data rate 双倍数据率
DDS direct digital synthesis 直接数字合成
DES Data Encryption Standard 数据加密标准
DFT discrete Fourier transform 离散傅里叶变换
DIP dual in line package 双列直插封装
DMA direct memory access 直接存储器存取
DMT discrete multi tone 离散多频(调制解调器)
DNL differential nonlinearity 差分非线性(数模转换器性能指标)
DNS domain name system 域名系统
dpdt double pole, double throw 双刀双掷开关
dpi dots per inch 点数/英寸
dpst double pole, single throw 双刀单掷开关
DQPSK differential quadrature phase-shift keying 差分正交相移键控
DRAM dynamic random-access memory 动态随机存取存储器
DSL digital subscriber line 数字用户线路
DSO digital storage oscilloscope 数字存储示波器
DSO digital sampling oscilloscope 数字采样示波器
DSP digital signal processing 数字信号处理
DSSS direct sequence spread spectrum 直序扩频技术
DTE data termination equipment 数据终端设备
DTV digital television 数字电视
DUT device under test 被测器件
DVI digital video interface 数字视频接口
DVM digital voltmeter 数字电压表
DWDM dense-wavelength-division multiplexing 密集波分复用技术
DWT discrete wavelet transform 离散小波变换

E

ECC error-correction code 纠错码
ECL emitter-coupled logic 射极耦合逻辑
EDA electronic design automation 电子设计自动化
EEPROM electrically erasable PROM 电可擦除可编程只读存储器
EMC electromagnetic compatibility 电磁兼容
EMF electromagnetic field 电磁场
EMI electromagnetic interference 电磁干扰
ENOB effective number of bits 有效位数(比特数)
ESD electrostatic discharge 静电放电
EUT equipment under test 被测设备

F

FAT	file allocation table 文件分配表
FC	Fibre Channel 光纤信道
FDDI	fibre distributed data interface 光纤分布数据接口
FDMA	frequency-division multiple access 频分多址
FDX	full duplex 全双工
FEC	forward error correction 前向纠错
FET	field-effect transistor 场效应管
FFT	fast Fourier transform 快速傅里叶变换
FHSS	frequency hopping spread spectrum 跳频扩谱技术
FIFO	first in, first out 先进先出(存储器)
FIR	finite impulse response 有限冲激响应
FLOPS	floating-point operations per second 每秒浮点运算数/秒
FM	frequency modulation 调频
FMCW	frequency-modulated continuous wave 调频连续波(雷达)
FPGA	field-programmable gate array 现场可编程门阵列
FR	full rate 全速率
FRAM	ferroelectric random-access memory 铁电随机存取存储器
FRF	frequency response function 频率响应函数
FS	full scale 满量程
FSK	frequency-shift keying 频移键控
FSM	finite state machine 有限状态机
FSS	fixed satellite service 固定卫星业务
FTP	File Transfer Protocol 文件传输协议

G

GaAs	gallium arsenide 砷化镓
GCA	gain-controlled amplifier 增益控制放大器
GIF	graphics interchange format 图形交换格式
GMSK	gaussian-filtered minimum-shift keying 高斯滤波最小频移键控
GPIB	general-purpose interface bus 通用接口总线
GPRS	General Packet Radio Service 通用无线分组业务
GPS	Global Positioning System 全球定位系统
GSM	Global System for Mobile communications 全球移动通信系统
GUI	graphical user interface 图形用户界面

H

HDD	hard-disk drive 硬盘驱动器
HDL	hardware-description language 硬件描述语言
HDTV	high-definition television 高清晰度电视
HDX	half duplex 半双工
HLL	high-level language 高级语言

HR half rate 半速率
HSL hue，saturation，lightness 色度、饱和度、亮度
HTML Hypertext Markup Language 超文本标记语言
HTTL high-power transistor-to-transistor logic 大功率晶体管-晶体管逻辑
HTTP Hypertext Transfer Protocol 超文本传输协议

I

I/O input/output 输入/输出
IC integrated circuit 集成电路
IDFT inverse discrete Fourier transform 离散傅里叶逆变换
IEEE Institute of Electrical and Electronic Engineers 电气电子工程师学会
IF intermediate frequency 中频
IFFT inverse fast Fourier transform 快速傅里叶逆变换
INL integral non-linearity 积分非线性(模数转换器性能指标)
IP intellectual property 知识产权
IP Internet Protocol 互联网协议
IR infrared 红外线
IrDA Infrared Data Association 红外数据协会
IRQ interrupt request 中断请求(软件)
ISA industry-standard architecture 工业标准结构
ISDN integrated services digital network 综合业务数字网
ISO International Organization for Standardization 国际标准化组织
ISP in-system programmable 在系统可编程
ISP Internet services provider 互联网服务提供商
ITU International Telecommunication Union 国际电信联盟

J

JEDEC Joint Electronic Device Engineering Council 联合电子设备工程委员会
JFET junction field-effect transistor 结型场效应管
JPEG Joint Photographic Experts Group 联合图像专家组
JTAG Joint Test Action Group 联合测试行动组

K

L

LAB logic array block 逻辑阵列块
LAN local-area network 局域网
laser light amplification by stimulated emission of radiation 激光
LCD liquid crystal display 液晶显示器
LCR inductance-capacitance-resistance 电感电容电阻
LED light-emitting diode 发光二极管
LF low frequency 低频(30～300 kHz)
LIFO last-in，first-out 后进先出(存储器)

lm lumen 流明
LNA low noise amplifier 低噪声放大器
LO local oscillator 本地振荡器
LPF low-pass filter 低通滤波器
LSB least significant bit 最低有效位
LSI large-scale integration 大规模集成
LUT look-up table 查找表
LVDS low-voltage differential signaling 低电压差分信号

M

MAC multiplication and accumulation 乘法累加
MAC media-access control 介质访问控制(通信协议)
MCU microcontroller unit 微控器单元
MD minidisk 迷你光盘
MDCT modified discrete cosine transform 改进离散余弦变换
MEMS micro-electro-mechanical system 微电子机械系统
MIDI musical instrument digital interface 乐器数字接口
MIMD multiple-instruction-stream, multiple-data-stream 多指令流,多数据流
MIPS million instructions per second 百万条指令/秒
MIS management information system 管理信息系统
MLP meridian lossless packing 最优无损压缩编码
MMDS multichannel multipoint distribution service 多路多点分布式业务
MMIC monolithic microwave integrated circuit 单片微波集成电路
MMX multimedia extension 多媒体增强(指令集)
modem modulator-demodulator 调制解调器
MOSFET metal oxide semiconductor field-effect transistor 金属氧化物半导体场效应晶体管
MPEG Motion Picture Experts Group 运动图像专家组
MPU microprocessor unit 微处理器单元
MRI magnetic resonance imaging 磁共振成像
MSB most significant bit 最高有效位
MSDSL multirate symmetric digital subscriber line 多速率对称数字用户线
MSI medium-scale integration 中规模集成
MSK minimum shift keying 最小频移键控
MTD moving-target detection 动目标检测(雷达)
mux multiplexer 复用器

N

NCO numerically-controlled oscillator 数控振荡器
NMOS n-type metal oxide semiconductor N 型金属氧化物半导体
NOP no-operation instruction 空操作(指令)
NOS network operating system 网络操作系统
NRE nonrecurring engineering 一次性工程

NRZ non-return-to-zero 非归零码
NTSC National Television Systems Committee 国家电视系统委员会
NVRAM nonvolatile random-access memory 非易失随机存取存储器

O

OC optical carrier 光载波
OFDM orthogonal frequency-division multiplexing 正交频分复用
OOP object-oriented programming 面向对象编程技术
OQPSK offset-quadrature phase-shift keying 偏置正交相移键控
OS operating system 操作系统
OSR oversampling ratio 过采样倍数
OTDR optical time-domain reflectometer 光时域反射测量仪

P

p-p peak-to-peak 峰-峰值
PAL Phase-Alternation Line 逐行倒相
PAL programmable array logic 可编程阵列逻辑
PAM pulse amplitude modulation 脉幅调制
PASC perceptual audio sub-band coding 感知音频子带编码
PBX private branch exchange 专用分组交换机
PC personal computer 个人计算机
PCB printed circuit board 印制电路板
PCI Peripheral Component Interconnect 外设部件接口(总线标准)
PCM pulse code modulation 脉冲编码调制
PCMCIA Personal Computer Memory Card International Association 个人计算机存储卡国际协会
PD pulsed doppler 脉冲多普勒(雷达)
PDP plasma display panel 等离子显示屏
PFC power factor correction 功率因数校正
PGA pin grid array 针栅阵列封装
PHY physical layer 物理层
pixel picture element 像素
PLD programmable logic device 可编程逻辑器件
PLL phase-locked loop 锁相环
PM phase modulation 相位调制
PMOS p-type metal oxide semiconductor P 型金属氧化物半导体
POST power-on self-test 开机自检
PRN pseudo-random noise 伪随机噪声
PROM programmable ROM 可编程只读存储器
PRS pattern recognition system 模式识别系统
PSD power spectral density 功率谱密度
PSK phase-shift keying 相移键控
PSPDN packet switched public data network 分组交换公共数据网

PSTN public switched telephone network 公共交换电话网
PWM pulse-width modulation 脉宽调制

Q

QAM quadrature amplitude modulation 正交调幅
QoS quality-of-service 服务质量
QPSK quadrature phase-shift keying 正交相移键控

R

RAM random-access memory 随机存取存储器
RAMDAC random-access memory，digital-to-analog converter 随机存取存储器和数模转换器
RF radio frequency 射频
RGB red，green，blue 红绿蓝
RISC reduced-instruction-set computer 精简指令集计算机
RLC resistance-inductance-capacitance 电阻电感电容
RLE run-length encoding 行程编码
rms root-mean-square 均方根值
ROM read-only memory 只读存储器
RTOS real-time operating system 实时操作系统
RW read/write 读/写

S

S/H sample-and-hold 采样保持电路
SAR synthetic aperture radar 合成孔径雷达
SAW surface acoustic wave 表面波(器件)
SDRAM synchronous DRAM 同步动态随机存取存储器
SDTV standard-definition digital television 标准清晰度电视
SECAM Sequential Color and Memory 顺序彩色与存储
SFC Shannon-Farro coding 香农-范诺编码
SFDR spurious-free dynamic range 无失真动态范围
SHA sample-and-hold amplifier 采样保持放大器
SHF super high frequency 超高频
SiGe silicon germanium 硅锗
SIMD single-instruction-stream，multiple-data-stream 单指令流，多数据流
SINAD signal-to-noise-and-distortion 信号-噪声畸变比
SIP single in-line package 单列直插封装
SMD surface-mount device 表面安装器件
SNR signal-to-noise ratio 信噪比
SoC system-on-chip 片上系统
SONET synchronous optical network 同步光纤网络
spdt single-pole，double-throw 单刀双掷开关
SPI serial peripheral interface 串行外设接口
SPICE Simulation Program with Integrated Circuit Emphasis 侧重集成电路仿真程序

SPS samples per second 样本数/秒
spst single-pole，single-throw 单刀单掷开关
SRAM static random-access memory 静态随机存取存储器
SS spread spectrum 扩频技术
SSB single sideband 单边带
SSI small-scale integration 小规模集成
SST solid state technology 固态技术
STB set-top box 机顶盒

T

T/H track-and-hold 跟踪保持电路
T/R transmit-and-receive 发射/接收
TCM trellis-coded modulation 格形编码调制
TCP/IP Transmission Control Protocol/Internet Protocol 传输控制协议/互联网协议
TDM time-division multiplexing 时分复用
TDMA time-division multiple access 时分多址
TDR time-domain reflectometer 时域反射测量仪
TEM transverse electromagnetic 横向电磁场
TF transfer function 传递函数
THD total harmonic distortion 总谐波失真
TIFF tag image format file 标记图像文件格式
TTL transistor-transistor logic 晶体管-晶体管逻辑

U

UART universal asynchronous receiver-transmitter 通用异步收发器
UHF ultrahigh frequency 特高频
ULSI ultra large-scale integration 极大规模集成
UPS uninterruptible power source 不间断电源
USB universal serial bus 通用串行总线
UTP unshielded twisted pair 未屏蔽双绞线
UV ultraviolet 紫外线

V

V/F voltage-to-frequency 电压-频率转换
VC virtual channel 虚拟信道
VCO voltage-controlled oscillator 压控振荡器
VF voice frequency 音频(300～3400 Hz)
VFC voltage-to-frequency converter 压频转换器
VGA video graphics array 视频图形阵列
VHF very high frequency 甚高频(30 kHz～300 MHz)
VLF very low frequency 甚低频(30 kHz 以下)
VNA vector network analyzer 矢量网络分析仪
VoIP voice over Internet Protocol 互联网协议电话

VPN virtual private network 虚拟私人网络
VQ vector quantization 矢量量化
VR virtual reality 虚拟现实
VSB vestigial sideband 残留边带

W

WAN wide-area network 广域网
WAP Wireless Application Protocol 无线应用协议
WDM wavelength-division multiplexing 波分复用
WLAN wireless local-area network 无线局域网
WWW World Wide Web 万维网

X

XML eXtensible Markup Language 可扩展标记语言

Y

YUV luminance/chrominance 亮度/色度(彩色电视图像)

Z

ZOH zero-order hold 零阶保持器

参 考 文 献

[1] ORSAK G C. Engineering our digital future[M]. London: Pearson, 2004.

[2] CARTER B. Op amps for everyone[M].5th ed. Boston: Newns, 2017.

[3] FLOYD T L. Digital fundamentals[M].11th ed. London: Pearson, 2014.

[4] HALL S H, HALL G W, MCCALL J A. High-speed digital system design: a handbook of interconnect theory and design practices[M]. Hoboken: John Wiley & Sons, 2000.

[5] ZWOLINSKI M. Digital system design with VHDL[M]. 2nd ed. London: Pearson, 2004.

[6] MAXFIELD C. The design warrior's guide to FPGAs: devices, tools and flows[M]. Boston: Newnes, 2004.

[7] VEGTE J V. Fundamentals of digital signal processing[M]. London: Pearson, 2002.

[8] KESTER W. Mixed signal processing and DSP design techniques[M]. Boston: Newnes, 2003.

[9] SMITH S W. Digital signal processing: a practical guide to engineers and scientists[M]. Boston: Newnes, 2002.

[10] BATEMAN A, STEPHENS I P. The DSP handbook: algorithms, applications and design techniques[M]. London: Pearson, 2002.

[11] TANENBAUM A S, FEAMSTER N, WETHERALL D J. Computer networks[M]. 6th ed. London: Pearson, 2021.

[12] KUROSE J, ROSS K. Computer networking: a top-down approach[M]. 8th ed. London: Pearson, 2022.

[13] FOROUZAN B. Data communications and networking with TCP/IP protocol suite[M]. 6th ed. New York: McGraw-Hill, 2021.

[14] FOROUZAN B. Foundations of computer science[M]. 5th ed. Boston: Cengage Learning, 2023.

[15] REED J H. Software radio: a modern approach to radio engineering[M]. London: Pearson, 2002.

[16] EDWARDS L. Embedded system design on a shoestring: achieving high performance with a limited budget[M]. Boston: Newnes, 2003.

[17] EDWARDS S A. Languages for digital embedded systems[M]. Berlin: Springer, 2000.

[18] MORRIS A S, LANGARI R. Measurement and instrumentation: theory and application [M]. 3rd ed. Amsterdam: Elsevier, 2021.

[19] MCINTOSH C. Cambridge advanced learner's dictionary[M]. 4th ed. Cambridge: Cambridge University Press, 2013.

[20] McGraw-Hill. McGraw-Hill dictionary of scientific and technical terms[M]. 7th ed. New York: McGraw-Hill, 2016.

[21] ROBIN D. Gaining word power[M]. 8th ed. London: Pearson, 2008.