# Turbo与LDPC编解码及其应用

肖扬 著

- 紧密联系热点通信标准与数字视频标准
- 直接给出LDPC码的设计与实现方法
- 全面深入，解决实际问题
- Turbo与LDPC编解码及应用研发人员的必备宝典

人民邮电出版社
POSTS & TELECOM PRESS

# Turbo与LDPC编解码及其应用

■ 肖扬 著

人民邮电出版社
北京

# 前　言

与传统的分组码和卷积码相比，Turbo 码与 LDPC 码具有接近仙农限的误码率性能。Turbo 码与 LDPC 码均被多项国际工业技术标准和国内工业技术标准列为信道编码的主要纠错码。Turbo 码的设计、编码与译码较为规范和单一，而 LDPC 码的设计、编码与译码具有诸多方案。出于知识产权保护的需要，国际工业技术标准和国内工业技术标准中 LDPC 码的设计和编码采用了不同的方案，使得国际工业技术标准和国内工业技术标准中的 LDPC 码具有不同的纠错性能，编码复杂度。

国内已出版的有关 Turbo 码与 LDPC 码的书籍局限于翻译国外著名学者的论文，而未深入考虑 Turbo 码与 LDPC 码的系统实现问题与应用问题，未对国际工业技术标准和国内工业技术标准中 LDPC 码的设计、编码和译码方案进行深入研究，未给出具体实现方法，使 LTE 无线通信系统、4G 无线通信系统的研发人员难以实现国际工业技术标准和国内工业技术标准中 LDPC 码编码和译码方案，使得 Turbo 码与 LDPC 码在应用阶段出现断层。

本专著试图解决这一问题，推进 Turbo 码与 LDPC 码在各领域的应用。本书结合国际工业技术标准和国内工业技术标准中 Turbo 码与 LDPC 码的设计与我们在该方面的研究结果(发表论文和发明专利申请)，系统地给出 Turbo 码与 LDPC 码、编码与译码算法和系统设计，指出 Turbo 码与 LDPC 码存在的问题，给出解决方法，同时给出 Turbo 码与 LDPC 码在信息隐藏、保密通信、MIMO 多天线通信系统中的具体应用技术实例。

本书内容可满足目前信息技术领域国家科技重大专项与国家高科技发展课题研究关于 LTE 无线通信系统、4G 无线通信系统中 Turbo 码与 LDPC 码子系统研制和产业化的需要。

全书共有 13 章，内容覆盖 Turbo 码和 LDPC 码的设计、编码算法（包括快速编码算法)、解码算法、短环检验算法、低码重码字和低距离码字检验算法等关键技术，包括设计实例和部分演示程序。本书第 1 章为绪论；第 2 章为 Turbo 编解码；第 3 章为 Turbo 码与 LDPC 码的自适应解码；第 4 章为 LDPC 码的短环、最小码重与最小距离；第 5 章为 LDPC 编码算法与译码算法；第 6 章为准循环 LDPC 码设计；第 7 章为 IEEE 802.16e 标准 LDPC 码；第 8 章为 DVB-S2 标准中的 LDPC 码；第 9 章为原模图 LDPC 码；第 10 章为 CCSDS 标准的 LDPC 码；第 11 章为 GB20600 标准的 LDPC 码；第 12 章为基于 LDPC 码的 MIMO 空间复用系统；第 13 章为基于 Turbo 码与 LDPC 码的保密通信技术。

本书以我们完成的国家自然科学基金课题和教育部博士点基金课题、在研的国家自然科学基金国际合作课题和北京市自然科学基金课题发表的论文和发明专利申请为主要内容，包括参与这些课题研究的博士生与硕士生发表的论文、发明专利申请和他们的实验结果，同时参考和引用了国内外相关工业标准和研究论文。

本书可作为高等院校通信与电子系统方面的研究生教材或参考书，也可供信道编解码系统方面研发人员参考。

肖　扬

2010 年 4 月于北京交通大学

# 目　录

# 第1章 绪 论

## 1.1 Turbo码的起源

随着科学的进步和生活水平的提高，人们对于通信的需求量以及通信质量也日益增长。由于对通信质量的高要求，人们希望找到一些提高通信质量的方法，而纠错码作为信道编码是提高通信质量特别是无线通信质量的好方法之一。提高信息传输的可靠性和有效性，始终是通信工作所追求的目标。纠错码是提高信息传输可靠性的一种重要手段。迄今，纠错码已有50多年的历史，其发展过程大致分以下几个阶段[1,2]。

20世纪50年代至60年代初，这是纠错码从无到有并最初发展的阶段。在此期间，科学工作者研究了各种有效的编、译码方法，奠定了线性分组码的理论基础；提出了著名的BCH码编、译码方法以及卷积码的序列译码等。

20世纪60年代至70年代初，这是纠错码发展过程中最为活跃的时期。在此期间，科学工作者提出了门限译码、迭代译码、软判决译码和卷积码的维特比（Viterbi）译码等有效的译码方法。并且科学工作者们还注意到诸如码的重量分布、译码错误概率和不可检错误概率的计算、信道的模型化等纠错码的实用性问题。在此期间，以代数方法特别是以有限域理论为基础的线性分组码理论已趋于成熟。

进入20世纪70年代，信道编码开始进入了另一个大的发展时期。在这个时期有三大重要进展：一是一类不展宽频带的编码调制技术的提出；二是级连编码概念的提出；三是作为分组码的另一类——卷积码的软判决译码算法。

进入20世纪80年代，人们开始运用几何观点讨论分析码，即利用代数曲线构造了一类代数几何码，并取得了许多成果。

20世纪90年代，Turbo码开始进入人们的视线。1993年，Berrou等人提出的Turbo码的并行级联卷积码，性能非常接近仙农限，同时复杂度较低可以实现，为信道编码领域带来了一场革命。由于Turbo码自提出之日起就成为信息论与编码理论界热切关注的焦点[1-20]，但事实上迄今为止并没有一套完整的理论对Turbo码的编译码原理做出解释，因此，人们对于Turbo码的研究热点也从最初的理论探索而转入了其在通信领域的应用，已被三大无线通信标准采纳[3-5]。在无线通信系统中，它被应用于信道的估计与译码、多用户检测、Turbo均衡、OFDM（正交频分复用）以及空时处理等；在信息安全领域中，它被应用于数字水印和信息加密。

## 1.2 Turbo码的性能及特点分析

### 1.2.1 Turbo码的性能特点及比较

Turbo码的优越性在于：即使在较低的信噪比下也能获得较好的性能。Turbo码的性能提

高是通过增加交织长度和迭代次数来实现的。同时，在 Turbo 码中存在错误平台问题，即在开始阶段，误比特率随着信噪比的增加急剧下降，但是到了一个特定点后，曲线下降会变得非常缓慢。这种错误平台的存在是 Turbo 码的距离特性造成的，即虽然 Turbo 码具有良好的性能，但其自由距离较小。Turbo 码性能优越的原因在于：尽管其自由距离特性差，但低距离的路径数目即距离的重复度比较小。虽然卷积码的自由距离可以设计得很大，但是低距离的重复度也会很大。在低信噪比的情况下，重复度对编码性能的影响比较大；在高信噪比的情况下，自由距离的作用会更大。

Shannon 有噪信道编码定理指出：对任何信道，只要其信息传输速率 $R$ 不超过信道容量 $C$，总存在一种编码方法，当采用最大似然（ML）译码时，其误码率可以任意小。该定理的证明中引用了 3 个基本条件：采用随机性编、译码；编码长度 $L \to \infty$，即分组的码组长度无限；译码采用最佳的似然译码（MLD）方案。

在 $R<C$ 的前提下，只有在码组长度无限的码集合中随机地选择编、译码字，并在接收端采用最大似然译码算法，才能使误码率趋于 0。但最大似然译码的复杂度随编码长度的增加而增大，当编码长度趋于无穷大时，最大似然译码不可能实现。所以，人们认为随机性编译仅是为证明定理存在性而引入的一种数学方法和手段，在实际的编码构造中是不可能实现的。直到 1993 年 Turbo 码的提出，很好地应用了 Shannon 信道编码定理中的随机性编译码，才获得了几乎接近 Shannon 理论极限的优异性能。

下面从 Turbo 码自身因素并结合 Shannon 有噪信道编码定理分析 Turbo 码性能，并从联合界的角度分析 Turbo 码的性能。

### 1.2.2 交织器对 Turbo 码性能的影响

在 Turbo 码生成中，交织器起着重要作用，在很大程度上影响着 Turbo 性能[18]。它在分量码编码器 RSC2 之前将输入信息比特的位置进行随机置换，使得长码的构造具有了随机性。我们知道，随机化是贯穿 Shannon 理论的核心思想之一。可以说，Turbo 码之所以具有如此令人惊异的优异性能，其根本原因就在于 Turbo 码实现方案中由于交织器的引入而实现的伪随机性。在发送端，伪随机性是通过编码器中的交织器及分量码的并行级联方式实现的；在接收端，则是通过带有交织器的具有软输入、软输出特性的反馈递推迭代译码来实现的。由于交织器的引入，整个 Turbo 码编码器可以看成是一个伪随机编码器。（可参见第 2 章图 2-1Turbo 编码器结构框图）

1995 年，Svirid 考虑了分量码是分组码的情形。设码字 $C=(m \mid mP \mid m'P)$，其中，$m$ 是 $k$ 维信息矢量，$\boldsymbol{P}$ 是 $k \times r$ 矩阵，$m'$ 是 $m$ 通过随机置换得来的。Svirid 指出交织器的目的在于使 Turbo 码的最小重量尽可能大，即交织器直接影响 $m'P$ 的重量，起到随机化作用。当 $mP$ 的汉明重量小时，$m'P$ 的汉明重量应该大；反之亦然。在 Turbo 码的编码中，交织器起着“窄谱化”的作用，使得 Turbo 码中重量小的码字数目减少，而这正是影响 Turbo 码性能的主要因素之一。1996 年 S. Benedetto 和 G. Montorsi 引入均匀交织器的概念，指出好的交织器是存在的。这些都说明了 Turbo 码优异性能在发送端主要是由编码交织器的伪随机化带来的。

Turbo 码的译码器结构是一类具有反馈结构的伪随机译码器，两个码可以交替互不影响地译码，并可以通过关于系统码信息位的软判决输出相互传递信息，进行递推式迭代译码。通过若干次迭代，每个码元都可以得到来自序列中几乎所有码元的信息。它具体是通过迭代中反复交织反馈及解交织来实现的，这实际上就实现了译码的伪随机化。

### 1.2.3 长码的构造及对 Turbo 码性能的影响

Shannon 有噪信道编码定理指出，随着编码长度 $L \to \infty$，译码错误概率趋于 0。因此为了使码有效就必须使用长码。但随着码长的增加，译码器的复杂度和计算量也相应增加以致难以实现。为解决这个问题，人们提出了级联码概念，把编制长码的过程分几级完成（通常分 2 级），即以现有的短码为基础构成等效长码[16]。如分别采用确知的短码作为内码和外码，希望通过 2 次纠错串行级联方式，即外码可以继续纠正内码未能纠正的错误，其总的纠错能力取决于内、外码的纠错能力。级联码方框图如图 1-1 所示。

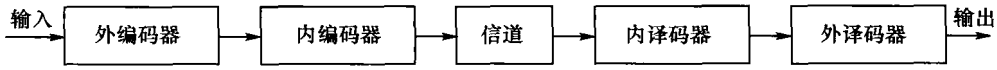


图 1-1 级联码方框图

但传统级联码并没有摆脱短码性能的束缚，当其接近信道容量的渐近状态时，一般传统短码的译码过程不但不能纠正错误，反而有可能使错误增大。然而，Berrou 等人提出的 Turbo 码，其编码器由递归系统卷积码（RSC）并行级联而成，并采用了反馈迭代译码，真正挖掘了级联码潜力，以其类似于随机的编译码方式，使其更加逼近了 Shannon 随机码的性能。

图 1-2 的 Turbo 码译码性能曲线反映了码长对译码性能的影响。方针采用 Log-MAP 算法，交织长度分别为 128、256 和 512，生成矩阵 $G=[15,17]$，码率 $R=1/2$ 且迭代次数为 8 次的译码性能曲线比较。

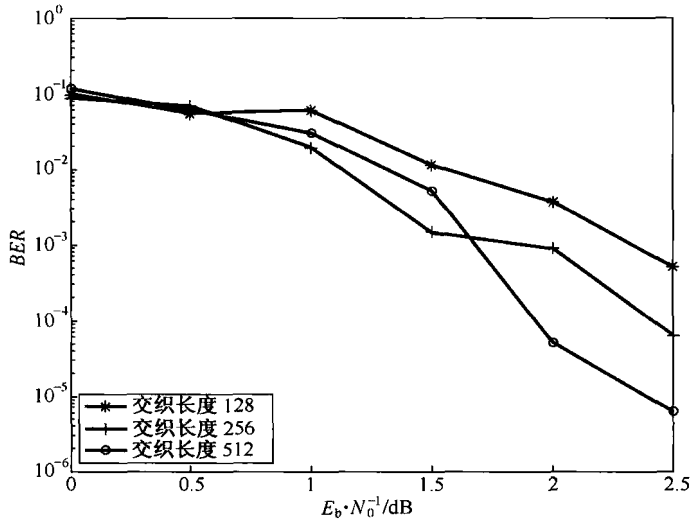


图 1-2 不同交织长度对 Turbo 译码性能的影响

从图 1-2 的仿真结果可以看出，迭代译码的误比特率都随着信息序列长度的增加而降低。当信噪比较小时，增加信息序列长度对误比特率性能的改进不大，但对于信噪比稍大的区域，Turbo 码的误比特率性能曲线与信息序列长度近似呈线性关系，这也从实验的角度证明了交织器增益与交织长度成线性关系；在信噪比较大时，误比特率曲线随着信息序列长度增加而变化趋缓，从而也验证了错误平层的存在。

### 1.2.4 译码算法对 Turbo 码性能的影响

Berrou 在 Turbo 码中采用了修正的 BCJR 算法，这使得分量码译码器 DEC1 与 DEC2 均可采用性能优良的卷积码。虽然此时仍然是由短码构成长码，但由于采用了反馈译码结构，实现了软输入、软输出递推迭代式译码，使编译码过程实现了伪随机化，并简化了最大似然译码的算法，使其性能达到了接近 Shannon 极限的优异性能。

Turbo 码主要的译码算法有 3 种[19]：（1）MAP 算法，该算法是修正的 BCJR 算法，也叫最大后验概率算法，它采用对数似然比（LLR，即后验概率（APP）的比值的对数值）作为其软判决的输出；（2）Max-log-MAP 算法及 Log-MAP 算法，此两种算法是在对数域中计算信息比特的后验概率；（3）SOVA 算法，即软输出 Viterbi 译码算法。

### 1.2.5 Turbo 码性能的联合界分析

基于概率论中 $P_r(\cup E_i) \leqslant \sum_i P_r(E_i)$ 这一结论，可得到采用最大似然（ML）译码时的平均译码错误概率上界。为得到 Turbo 码的联合界，先对一个（$N$, $K$）线性分组码 $C$ 定义如下的重量枚举函数。定义码输入冗余重量枚举函数（IRWEF）为：

$$A^C(W,Z)=\sum_{\omega,j} A_{\omega,j} W^{\omega} Z^j \tag{1}$$

其中，$A_{\omega,j}$ 代表码字集中信息位重量为 $\omega$、校验位重量为 $j$ 的码字的个数。

定义给定信息位重量为 $\omega$ 时校验位的条件重量枚举函数（CWEF）为：

$$A_{\omega}^C(Z)=\sum_j A_{\omega,j} Z^j \tag{2}$$

为求得 2 个分量码并行级联后的 IRWEF，可引入均匀交织器的概念。所谓均匀交织器，是指一种概率器件，它对所有可能的交织器进行统计平均。这样重量为 $\omega$ 和长为 $K$ 的信息码共有 $\begin{bmatrix} K \\ \omega \end{bmatrix}$ 种交织结果，每种概率为 $\begin{bmatrix} K \\ \omega \end{bmatrix}^{-1}$。令 $A_{\omega}^{C_1}$ 和 $A_{\omega}^{C_2}$ 分别是分量码编码器 RSC1 和 RSC2 的条件重量枚举函数，则 Turbo 码整体条件重量枚举函数为[20]：

$$A_{\omega}^{C_p}(Z)=\frac{A_{\omega}^{C_1}(Z)A_{\omega}^{C_2}(Z)}{\begin{bmatrix} K \\ \omega \end{bmatrix}} \tag{3}$$

上式中，上标 $C_p$，$C_1$ 及 $C_2$ 分别表示并行级联码和 2 个分量码。这样就可以得到给码下的 Turbo 码平均性能，也就是说必然存在一种交织器使 Turbo 码优于这个平均性能，于是对 AWGN 信道输出误码率的联合界可写成：

$$P_b(e) \leqslant \frac{1}{2}\sum_{\omega=1}^{K} \frac{\omega}{K} W^{\omega} A_{\omega}^{C_p}(Z) \Big|_{W=Z=\exp(-R_C E_b/N_0)} \tag{4}$$

但是，给定一 Turbo 码，其 AWGN 信道输出误码率的联合界见式（4），并不能解释其误码平台出现的原因。笔者在第 2 章提出的最小码重搜索算法解决了这一问题。

除了以上方面对于 Turbo 码的译码性能有影响外，还有迭代次数、所采用的分量码以及编码速率等方面，第 2 章将分别作介绍。

第2章从实验过程中的仿真以及现实Turbo码的概况发现了Turbo码存在着以下缺点。

（1）计算量大。要得到高码率往往需要很大的交织器，这就增加了译码复杂性，而较短的交织器不可能达到高码率，因此往往要根据实际需要来确定码率和计算复杂性之间的平衡来设计相应的Turbo码。

（2）交织和交错译码造成的时延，使Turbo码在某些对时延要求高的通信系统中的应用受到限制。

（3）所谓的错误平层效应，即误码率下降到一定程度后，再下降就很慢了。

（4）理论分析困难。至今尚未对Turbo码译码的复杂性以及比特误差概率的完整理论分析和估计。一般是通过数值模拟与单个卷积码、乘积码、级联码进行比较或在不同译法之间作性能比较。

通过对不同设计参数下Turbo码的仿真，可得到如下结论。

（1）在有限范围内增加分量码的约束长度可以提高Turbo码的性能。

（2）Turbo码的性能随迭代次数的增加而提高。通常在一定的迭代次数后Turbo码的误比特率曲线收敛到一定的范围，采用迭代停止准则动态设置迭代次数可以有效降低译码复杂性。此外，交织长度小的Turbo码的收敛速度要比交织长度大的Turbo码的收敛速度快。交织长度越大，误比特率曲线误平层出现的越早，相应的信噪比越小。

（3）Turbo码的性能随交织长度的增加而提高。

（4）降低码率可以有效地提高Turbo码的性能。

## 1.3 LDPC码概述

由于文献[21]提出的LDPC码可采用复杂度较低的迭代消息传递（BP）译码算法，且译码复杂度不会随着码长的增加而加大，所以性能优异的LDPC码设计受到极大关注[21-64]。目前国际上对LDPC码的理论研究已取得重要进展，而且已经进入工程应用和超大规模集成电路实现阶段[38-40]。

LDPC码是一种分组码，其校验矩阵只含有很少量非零元素。正是校验矩阵 $\boldsymbol{H}$ 的这种稀疏性，保证了译码复杂度和最小码距都只随码长呈现线性增加[21-30]。校验矩阵 $\boldsymbol{H}$ 的每一行对应一个校验方程，每一列对应码字中的一比特。因此，对于一个二进制码，如果它有 $m$ 个奇偶校验约束关系，码字的长度为 $n$，则校验矩阵是一个尺寸为 $m \times n$ 的二进制矩阵。对于 $m \times n$ 维校验矩阵为 $\boldsymbol{H}$，当且仅当向量 $\boldsymbol{c}=[c(1) \quad c(2) \quad \cdots \quad c(m)]$ 满足：

$$\boldsymbol{H} \cdot \boldsymbol{c}^{\mathrm{T}} = 0 \tag{5}$$

时，它才是该码的一个有效码字。

除了校验矩阵 $\boldsymbol{H}$ 是稀疏矩阵外，LDPC码本身与任何其他的分组码并无二致。如果现有的分组码可以被稀疏矩阵所表达，那么用于LDPC码的迭代译码算法也可以成功地移植到它身上。然而，一般来说，为现有的分组码找到一个稀疏矩阵并不实际。不同的是，LDPC码的设计是以构造一个校验矩阵开始的，然后才通过它确定一个生成矩阵进行后续编码。

译码方法是LDPC码与经典的分组码之间的最大区别。经典的分组码一般是用ML类的译码算法进行译码的，所以它们一般码长较小，并通过代数设计以减低译码工作的复杂

度。但是 LDPC 码较长，并通过其校验矩阵 $\boldsymbol{H}$ 的两部图而进行迭代译码，所以它的设计以校验矩阵 $\boldsymbol{H}$ 的特性为核心考虑之一。目前的研究均表明 LDPC 码是信道编码中纠错能力最强的一种码，而且其译码器结构简单，可以用较少的资源消耗获得极高的吞吐量，因此应用前景相当广泛。

LDPC 码在结构上可以分为规则 LDPC 码和不规则 LDPC 码[21-40]，规则 LDPC 码的校验矩阵每行的非零元素的数目相同，记为 $w_r$，每列的非零元素的数目也相同，记为 $w_c$；而不规则 LDPC 码则不受此规则限制。构造二进制 LDPC 码实际上就是要找到一个稀疏矩阵 $\boldsymbol{H}$ 作为 LDPC 码的校验矩阵，基本方法是将一个全零矩阵的一小部分元素替换成 1，使得替换后的矩阵各行和各列具有所要求的数目的非零元素。文献[41-54]指出，如果要使构造出的 LDPC 码具有良好的纠错性能，则必须满足 3 个条件，分别是无短环、无低码重码字、码间最小距离要尽可能大。现在分别阐述 3 个条件。

短环的存在使译码器不能快速收敛，甚至不能收敛[25, 25, 41-54]。码的环指码的 Tanner 图从一个比特节点出发，交替地经过校验节点和比特节点，跳了若干步后回到原来的比特节点这一过程所形成的封闭回路。显然，码环的长度只能是大于或等于 4 的偶数。虽然没有明确的标准指出码无短环指的到底是不能有长度是几的环，但是有一点是确定的，即码至少不能有四环。从 LDPC 校验矩阵上看，如果校验矩阵 $\boldsymbol{H}$ 有任何 4 个 1 分别在矩形的 4 个顶点上，则这 4 个 1 就构成了一个四环。如果 LDPC 码的 Tanner 图是无环的，则和积 SP(Sum-Product)译码算法可以实现最佳译码[25,26]，如果存在环，那么由和积算法计算所得的概率并非真正的后验概率（因为迭代过程中的独立性假设不成立）。因此，译码并不是最优的逐符号最大后验概率译码。可见，环的存在使得译码的最优性能得不到满足。但在代码长固定的情况下，实现码字无环是不可能的。围长是码中最小的环长度，增大围长可以提高码字的性能，围长达到一定的值就可以接近无环时的译码性能。所以如何检测 LDPC 码是否具有短环是 LDPC 码构造中的重要问题。

第 4 章提出一种基于逐点搜索法的短环检验算法，根据校验矩阵中环的几何形状依次搜索组成环的“1”的组合。该算法计算复杂度低，但只能检验几何形状比较简单的四环和六环。文献[52]提出一种基于校验矩阵自相关性来检验校验矩阵中是否有四环的方法。利用校验矩阵与它的转置矩阵相乘，然后检查积矩阵中除主对角线上的元素以外，其他元素是否存在大于 1 的整数来判断原校验矩阵是否存在四环。该方法设计简单易行，但只对校验矩阵中有无四环进行判断，对六环及更大的环不能判断。文献[53]提出一种基于 Tanner 图的方法来计算 LDPC 码短环分布的方法。该方法是针对每一个变量节点，利用 Tanner 图逐步增长的方法求得其包含的最小环。对于给定节点 U，其最小环的计算方法为，以 U 为根，经过 $f$ 步（$f$ 大于或等于 $k$），所有与根节点 U 距离为 $f$ 的节点都包含在树中。假定在第 $k$ 步，树中第一次出现了相同节点叶子，则 $2k$ 便是 U 的最小环。这样依次可以求得每一个变量节点所包含的最小环，然后求出所有变量节点的平均最小环。利用这种方法从一组 LDPC 码中搜寻出平均最小环最大的 LDPC 码，即可得到该组 LDPC 码中性能最优的码。

第 3 个条件是保证码字间的最小码距要尽可能大[56-66]。保持码字间的最小码距尽可能大的原因是保证码的纠错性能。同样，现在尚没有完善的方法可以求出码的最小码距，只能通过计算机搜索来近似确定码字间的最小码距。低码重码的存在使 LDPC 译码器纠错能力低下，不能纠正多个比特的错误。虽然没有明确的界限指出 LDPC 码的码重要是多少才

不算低，但可以认为低码重码的码重小于 10。如果存在低码重码字，在发送端没有发送信息，或者说发送的是全 0 信号时，如果全 0 信号受到信道的干扰成为含有少量 1 的非全 0 信号，则接收端会错误地认为发送端发送了非 0 码字。现在尚没有办法准确求出 LDPC 码的最小码重，只能通过计算机搜索来寻找低码重码。在搜寻范围很大的情况下，仍没有出现码重较低的码字，则认为该 LDPC 码不存在低码重码。虽然码的最低码重无法准确求出，也就意味着无法避免设计出的码具有低码重码字，但是仍然可以在设计中注意一些问题来减少设计出的码具有低码重码字的可能。由文献[56，57，59]可知，生成矩阵 $\boldsymbol{G}$ 的各行都是码字，它们相互独立，整个码字集合就是由这 $k$ 个码字的线型组合构成的。所以，设计 LDPC 码时必须保证生成矩阵 $\boldsymbol{G}$ 的各行不是低码重码字，即首先保证矩阵 $\boldsymbol{G}$ 无低码重行；另外，还可尽量分别考虑矩阵 $\boldsymbol{G}$ 的每两行、每三行等的线性组合不出现低码重码字[56, 57, 59]，考虑得越周全则设计出的码出现低码重码字的可能性越小。在 LDPC 码设计时，做到无四环并不难，难的是无低码重码[60]。第 4 章给出了我们提出的低码重码搜寻算法，可以检测所设计的码是否存在低码重码。

## 1.4 LDPC 码的基本构造方法

### 1.4.1 Gallager 构造的规则随机 LDPC 码

最早的 LDPC 码是由 Gallager 构造的[21]，这是一种规则随机 LDPC 码，而且这种 LDPC 码的校验矩阵 $\boldsymbol{H}$ 的结构是有规定的。设 Gallager 所构造的校验矩阵 $\boldsymbol{H}$ 的尺寸为 $m \times n$，每行，每列具有 1 的数目分别是 $w_r$ 和 $w_c$。矩阵 $\boldsymbol{H}$ 的各行被平分为 $w_c$ 块，每一块含有 $m/w_r$ 行，每一块的列数规定为 $(m/w_c)\cdot w_r$。第一块矩阵从第一行到最后一行，从左到右依次在不同的连续 $w_r$ 个位置上是 1，即当 $i \leqslant m/w_c$ 时，第 $i$ 行的第 $(i-1)\cdot w_r+1$ 列到第 $i\cdot w_r$ 列的元素是 1。构造了校验矩阵 $\boldsymbol{H}$ 的第一块后，其余各块根据第一块按列进行随机置换而得到。结果是，校验矩阵 $\boldsymbol{H}$ 的每一行有 $w_r$ 个 1，每一列有 $w_c$ 个 1。以下是一个按照 Gallager 方法构造出的规则 LDPC 码的校验矩阵，其中 $w_r$=4，$w_c$=3，$\boldsymbol{H}$ 的尺寸为 $9 \times 12$。

$$\boldsymbol{H}=\left[\begin{array}{cccccccccccc}
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
\hdashline
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
\hdashline
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1
\end{array}\right] \tag{6}$$

用 Gallager 方法构造校验矩阵简单，而且校验矩阵的行重和列重容易控制，只需先根据码长和行重构造一个子矩阵，校验矩阵就可以利用此子矩阵得到。但是，Gallager 码存在的问题有 3 个：（1）由于随机构造的原因，在码长较短时二分图中可能会出现较短的环，影响码的性能，但随着码长的增加，短环出现的概率也会变得很小；（2）用 Gallager 方法构造的

校验矩阵为奇异，无法得到同样尺寸的生成矩阵；（3）Gallager 码的是一种随机 LDPC 码，其校验矩阵与生成矩阵并不具有准循环特性，使得其编码和解码的复杂度甚高，难以应用。因此在几个工业标准中[38-41]并未采用 Gallager 码。

采用 Gallager 的方法构造 1/2 码率的随机 LDPC 码的校验矩阵的程序见附录。在该程序中，补充了四环检验算法，以解决 Gallager 码的四环问题。读者可利用该程序获取 Gallager 的方法构造校验矩阵，验证上述结论。

### 1.4.2 MacKay 和 Neal 构造的规则随机 LDPC 码

另一种常用的规则随机 LDPC 码的构造方法由 MacKay 和 Neal 提出[22,23]。在这种方法中，校验矩阵 $\boldsymbol{H}$ 的列从左至右逐列增加，最终形成整个校验矩阵。每一列的重量，即 1 的数目，根据需要而定，而 1 的位置随机选择，但是必须使得行重不超出规定范围。如果直到设置最后一列的时候，仍然有一些行的 1 的数目不能满足要求，则要重新设置 $\boldsymbol{H}$ 或者从右往左取消某些列进行重新设置，直到最后的 $\boldsymbol{H}$ 的行列重量都满足要求。以下是一个按照 MacKay 和 Neal 所提出的方法构造出的规则随机 LDPC 码的校验矩阵，其中 $w_r$=4，$w_c$=3，$\boldsymbol{H}$ 的尺寸为 9 × 12：

$$\boldsymbol{H}=\begin{bmatrix}1&0&0&0&0&1&0&1&0&1&0&0\\1&0&0&1&1&0&0&0&0&0&1&0\\0&1&0&0&1&0&1&0&1&0&0&0\\0&0&1&0&0&1&0&0&0&0&1&1\\0&0&1&0&0&0&1&1&0&0&0&1\\0&1&0&0&1&0&0&0&1&0&1&0\\1&0&0&1&0&0&1&0&0&1&0&0\\0&1&0&0&0&1&0&1&0&1&0&0\\0&0&1&1&0&0&0&0&1&0&0&1\end{bmatrix} \tag{7}$$

当设置第 11 列时，第 2、4、5、6、9 行的 1 的数目都未达到 4，所以选择该列的 1 的位置时在它们之中选择，最终选择第 2、4、6 列。

MacKay 和 Neal 所提出规则随机 LDPC 码可以在无四环的约束条件下设计，但是其校验矩阵与生成矩阵并不具有准循环特性，使得其编码和解码复杂度甚高，难以应用。本书第 13 章中基于 LDPC 码的保密通信，使用了该码以获得较高的信息加密性能。

### 1.4.3 重复累加设计法构造的不规则 LDPC 码

另一种 LDPC 码构造法针对不规则 LDPC 码，叫做重复累加设计法。这种方法中，校验矩阵 $\boldsymbol{H}$ 的前 $k$ 列的列重根据具体情况而定，但是后 $m$ 列的列重为 2，并且具有双对角下三角结构。这样的结构非常特殊，可以使 LDPC 码具有系统化特征，更重要的是，这样的结构使编码容易，DVB-s2 标准采用了这种码结构[38]。在第 11 章中将对这个结构的不规则 LDPC 码进行单独而详细的讨论。以下的校验矩阵 $\boldsymbol{H}$ 是根据重复累加法而设计的，它的尺寸为 9 × 12，码率为 1/4：

$$\boldsymbol{H}=\begin{bmatrix}1&0&0&1&0&0&0&0&0&0&0&0\\1&0&0&1&1&0&0&0&0&0&0&0\\0&1&0&0&1&1&0&0&0&0&0&0\\0&0&1&0&0&1&1&0&0&0&0&0\\0&0&1&0&0&0&1&1&0&0&0&0\\0&1&0&0&0&0&0&1&1&0&0&0\\1&0&0&0&0&0&0&0&1&1&0&0\\0&1&0&0&0&0&0&0&0&1&1&0\\0&0&1&0&0&0&0&0&0&0&1&1\end{bmatrix} \tag{8}$$

显然，$\boldsymbol{H}$ 的前 3 列的列重是 3，而后 9 列的列重都是 2，而且具有双对角下三角结构。现在简要介绍一下这种结构的校验矩阵如何使得编码更为简便。矩阵的前 3 列对应 3 位信息比特，后 9 列对应的是 9 位校验比特。设编码后的码字向量为：

$$\boldsymbol{c}=\begin{bmatrix}c(1) & c(2) & c(3) & c(4) & c(5) & c(6) & c(7) & c(8) & c(9) & c(10) & c(11) & c(12)\end{bmatrix}$$

观察校验矩阵 $\boldsymbol{H}$ 的结构，很容易得到 $c(4)=c(1)$，然后有 $c(5)=c(4)\oplus c(1)$，下一个是 $c(6)=c(5)\oplus c(2)$。用这种方法编码，每次只编出 1 位校验比特，当前的校验比特只需要通过前一位的校验比特和一位信息比特即可得到，使得编码存储空间大大减少。

式（8）给出的检验矩阵可以分解为两个子矩阵，其中一个子矩阵为双对角线矩阵。利用该双对角线矩阵可以直接用检验矩阵编码，无须生成矩阵，第 5 章和第 8 章详细给出了这种码的设计与编码算法。这种码的缺点是可能存在低码重码[60]。

除此以外还有其他 LDPC 码基本构造方法，下面将逐一做简单介绍。

### 1.4.4 π–旋转矩阵构造法

旋转矩阵构造法的核心思想是利用单位置换矩阵来构造校验矩阵，如文献[29]中就应利用单位置换矩阵和一个双对角矩阵构成 LDPC 码校验矩阵。该方法和 Gallager 构造法有些相似，都是由子矩阵构成校验矩阵，但此种方法中子矩阵为单位置换矩阵，其行重和列重均为 1。多个子矩阵在纵向和横向排列构造得到校验矩阵，有更大的灵活性。子矩阵的大小要根据码长、码率以及校验矩阵的行重、列重进行设计，而且子矩阵的大小对码的性能有较大的影响。

### 1.4.5 PEG 构造法

Hu XiaoYu 等提出的 PEG（Progressive Edge-Growth）方法是一种使得变量节点的局部环长最大的构造方法[30]。假设在二分图上前 $i-1$ 个变量节点的边已经构造出来，在构造下一个变量节点的边时，每次向二分图上添加一条边，新添加的边使得经过第 $i$ 个变量节点的最小环长最大化。由于短环的存在对译码性能存在很大的影响，利用 PEG 方法构造的校验矩阵使得最小环长最大，因而一般来说，利用 PEG 方法构造的校验矩阵比利用 Gallager、旋转矩阵等方法构造的校验矩阵性能好。但此种方法要对边的分布进行优化，其复杂度随之增加。

### 1.4.6 准循环构造法

准循环（Quasi-Cyclic）LDPC 码的校验矩阵由一些零矩阵和循环置换单位子矩阵构成[25,28, 38-51]。定义 $Z_i$ 为 $z\times z$ 阶单位阵循环移动 $i$ 次得到的循环置换子矩阵，其中，$Z_\infty$ 意味着尺寸为 $z\times z$ 的零矩阵。校验矩阵 $\boldsymbol{H}$ 为 $mz\times nz$ 阶，可如下构造：

$$\boldsymbol{H}=\begin{bmatrix} \boldsymbol{Z}_{\alpha_{11}} & \boldsymbol{Z}_{\alpha_{12}} & \cdots & \boldsymbol{Z}_{\alpha_{1(n-1)}} & \boldsymbol{Z}_{\alpha_{1n}} \\ \boldsymbol{Z}_{\alpha_{21}} & \boldsymbol{Z}_{\alpha_{22}} & \cdots & \boldsymbol{Z}_{\alpha_{2(n-1)}} & \boldsymbol{Z}_{\alpha_{2n}} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \boldsymbol{Z}_{\alpha_{m1}} & \boldsymbol{Z}_{\alpha_{m2}} & \cdots & \boldsymbol{Z}_{\alpha_{m(n-1)}} & \boldsymbol{Z}_{\alpha_{mn}} \end{bmatrix} \tag{9}$$

其中 $\alpha_{ij}$ 的取值范围是 $\{0,1,\cdots,z-1,\infty\}$。

对于准循环 LDPC 码只需存储循环置换单位子矩阵第一行中“1”元素的位置以及循环置换单位子矩阵在校验矩阵中的位置，所需要的存储量大为减少，变为原来的 $1/z$。准循环 LDPC 码具有循环码的一些特性，也可以采用移位寄存器完成编码，硬件实现复杂度低。在标准 CCSDS131.1-O-2（Ex-perimental Specification）中针对近地业务和深空业务提出的 LDPC 码方案都是准循环 LDPC 码[40]，由此也可见其良好的应用前景。这种准循环结构的 LDPC 码是一类重要的 LDPC 码，已被 IEEE 802.16e 标准和 GB20600 标准所采纳[39, 41]。

### 1.4.7 代数构造法

代数构造方法中的一类是利用有限几何（Finite Geometries）来分析构造好的 LDPC 码。Yu Kou 和 Shu Lin 等[25, 26]提出一种在有限域上基于欧几里德空间和投影几何的点和线的几何方法来设计 LDPC 码。Y. Kou 给出了 4 种类型的几何构造的 LDPC 码，这些码具有很好的最小距离特性，对应的 Tanner 图不含周长为 4 的环，最小环周长为 6。有限几何 LDPC 码可以采用很多种译码方法，采用置信传播迭代译码可以获得很好的性能，更重要的是它可以表示成循环或半循环的形式，利用生成多项式决定的简单反馈移位寄存器就可以实现编码，编码时间与码长呈线性关系，在实际应用中非常重要。同时几何 LDPC 码可以用不同的方法扩展或截短得到其他的码。在代数构造方法中，是利用循环置换矩阵来构造 LDPC 码[8]。利用这种方法构造的 LDPC 码校验矩阵大部分是准循环矩阵，这类构造方法具有线性的编码时间。如 Lin Shu 等人提出一种对利用有限几何构造的循环矩阵进行分解的准循环 LDPC 码的构造方法，这种方法具有线性的编码时间，而且可以用线性的移位寄存器来实现。

### 1.4.8 Tanner 图

LDPC 码可通过伪随机的方法构造，需要在给定一些设计规范后得到所有 LDPC 码的集合，而不仅限于如何选择一些特殊的校验矩阵使之满足设计规范。

LDPC 码常常通过 Tanner 图来表示[28]，而 Tanner 图所表示的其实是 LDPC 码的校验矩阵。Tanner 图包含两类顶点：$n$ 个码字比特顶点（称为比特节点），分别与校验矩阵的各列相对应；$m$ 个校验方程顶点（称为校验节点），分别与校验矩阵的各行对应。校验矩阵的每行代