

中短长度 LT 码的展开图构造方法

龚茂康

(扬州大学信息工程学院 扬州 225009)

摘要: 该文提出了中短长度 LT 码的展开图构造方法。该方法以输出节点为根展开 Tanner 图, 当满足停止准则后, 随机选择未被覆盖或最深展开层的输入节点与根相连。展开图法通过展开深度门槛和最大可展深度两个参数的设置, 在控制短环的基础上, 在尽可能大的范围内随机地选择输入节点。仿真结果表明, 该方法与随机图等方法相比, 能够有效地改善中短长度 LT 码的性能。

关键词: 信道编码; LT 码; Tanner 图

中图分类号: TN 911.22

文献标识码: A

文章编号: 1009-5896(2009)04-0885-04

Unfolding Graphs for Constructing of Short and Moderate-Length LT Codes

Gong Mao-kang

(Coll of Inf. Engin, Yangzhou Univ, Yangzhou 225009, China)

Abstract: This paper proposes one construction method for short and moderate-length Luby Transform (LT) codes by unfolding Tanner graphs from every output node as root. For each output node, when the stopping criterion is satisfied, one uncovered input node or one input node on the deepest layer is randomly selected and connected to the root. Two parameters, the unfolding depth threshold and the maximal unfolding depth, are adjusted to control the girth of short cycles and make the range of random selection as large as possible. Simulation shows the performance improvements of the proposed algorithms are obvious compared with the random graph method for short and moderate length LT code.

Key words: Channel coding ; Luby Transform (LT) codes; Tanner Graph (TG)

1 引言

数字喷泉 (Digital Fountain, DF) 码^[1]近几年来已经成为理论界和工程界的关注焦点, 因为它具有的优良的抗擦除特性, 使得它在擦除信道中有着比其他码(如低密度校验码^[2])更好的应用前景^[3,4]。Luby 于 1998 年通过 Luby 变换发明了第 1 种 DF 码即 LT 码^[5]。Shokrollahi 发明的另外一种 DF 码—猛禽码(Raptor Code)^[6]已经成为第 3 代移动通信系统的纠错码标准之一^[7], 它以 LT 码作为其级联码的内码。LT 码的研究已经成为研究猛禽码的重要内容^[8]。因为实时通信的要求, 长度为几千甚至更短的 LT 码的编译码方法研究得到了更加的重视^[3,4,9,10]。

LT 码的输出和输入节点度数分布决定了码的渐进性能^[11], 经典的输出节点度数分布是基于 Soliton 函数的各种分布^[12], 输入节点的度数分布是近似的 Poisson 分布^[10]。LT 码的 Tanner 图是稀疏图, LT 码的编译码方法都是基于图的方法。经典的随机图(RG)编码方法是, 对于某个输出节点, 首先随机地选择其度数, 然后在全部的输入节点中随机地(一般是平均地)选择和该输出节点相连的输入节点。RG 编码方法虽然可以使长码的渐进性能较好, 但会使中短长度的码产

生许多短环, 特别是长度为 4 的环, 削弱了码的性能。针对短环的控制, 文献[13]在构造 LDPC 码时提出了累积边增加(PEG)方法。该方法尽可能深地展开 Tanner 图, 选择具有最小度数的未被覆盖的校验节点, 与某个码字节点亦 Tanner 图的根相连。PEG 方法产生的 Tanner 图具有较大的环长度, 得到校验节点度数分布是平均分布, 而不是 Poisson 分布。本文提出一种中短长度 LT 码展开图构造方法, 有限深度地展开 Tanner 图, 随机选择未被覆盖的输入节点, 既有效控制了环长度, 又使输入节点度数近似于 Poisson 分布。

2 Tanner 图的展开和环长度的控制

设 LT 码 Tanner 图中 N 个输入节点和输出节点分别为 i_1, i_2, \dots, i_N 和 $v_1, v_2, \dots, v_m, \dots$, 输出节点的数量取决于信道情况。当 v_k 的度数确定后, 在 i_1, i_2, \dots, i_N 中首次随机选取输入节点 $i_{k,0}$ (例如 i_1) 与 v_k 相连。在本文的算法中定义了参数 $UfDepthThreshold$ 为图的展开深度门槛, 当 $UfDepthThreshold=1$ 时, 以 v_k 为根的展开图如图 1(a)所示。在展开图第 0 层和第 1 层所包含的输入节点 i_1, i_2, i_3, i_4 范围以外, 即在展开图未覆盖的输入节点中选择输入节点 i_5 与 v_k 相连, 就可以杜绝因选择 i_2, i_3, i_4 而产生的长度为 4 的短环, 此时以 v_k 为根的扩展图如图 1(b)所示。然后再在未覆盖的输入节点中选择第 3 个输入节点与 v_k 相连, ...。在图扩

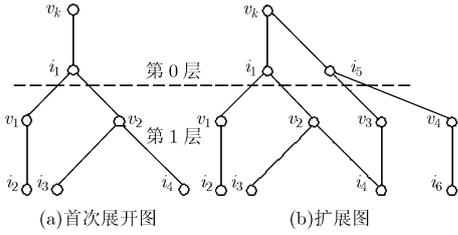


图 1 以 v_k 为根的展开图

展过程中, 可能出现全部输入节点被覆盖的情况, 虽然此时选择一个最深展开层的输入节点会导致环的产生, 但是从全局来说该展开法已经最有效地控制了短环的长度, 减少了短环的数量。当 $UfDepthThreshold \geq 1$ 时, 上述的 Tanner 图迭代展开方法就能够有效地控制环长度为 4 短环的产生。

3 展开图算法

令和 i_1, i_2, \dots, i_N 相连的输出节点集合分别为 $Adj(i_1), Adj(i_2), \dots, Adj(i_N)$, 和 v_k 相连的输入节点集合为 $Adj(v_k)$, $1 \leq k \leq m$ 。令 $L_0 = Adj(v_k)$ 是以 v_k 为根的展开树中第 0 层的输入节点, $L_1 = \bigcup_{i \in L_0} \bigcup_{v \in Adj(i)} Adj(v) - L_0$ 为展开第 1 层后展开树中新增加的输入节点, $L_l = \bigcup_{i \in L_{l-1}} \bigcup_{v \in Adj(i)} Adj(v) - \bigcup_{k < l} L_k$ 为第 l 层的新增输入节点。令 $S_l = \bigcup_{k \leq l} L_k$ 为展开树中前 l 层的输入节点。本算法定义参数 $UfDepthMAX$ 为图的最大可展深度。为保证图在 $UfDepthMAX$ 范围内的有效展开, 提出展开图的停止展开准则如下:

3.1 停止准则

停止准则 1 当 $l < UfDepthMAX$ 时, 如果 $S_l \neq S_{l-1}$ 但是 $S_l = \{i_1, i_2, \dots, i_N\}$, 停止展开, 并且令 S_{l-1} 中的全部输入节点为禁选状态;

停止准则 2 当 $l < UfDepthMAX$ 时, 如果 $S_l = S_{l-1}$ 和 $S_l \neq \{i_1, i_2, \dots, i_N\}$, 停止展开, 并且令 S_{l-1} 中的全部输入节点为禁选状态;

停止准则 3 如果 $l = UfDepthMAX$, 停止展开, 并且令 S_l 中的全部输入节点为禁选状态。

3.2 集合 $Adj(v_k)$ 的选择算法

步骤 1 令 $k = 1$, 集合 $Adj(i_1), Adj(i_2), \dots, Adj(i_N)$ 为空。

步骤 2 令 $Adj(v_k)$ 为空集合。从输出节点度数分布中随机选择 v_k 的度数 d_k ; 从全部输入节点中随机选择一个节点 $i_{k,0}$; 将 $i_{k,0}$ 加到 $Adj(v_k)$, 将 v_k 加到 $Adj(i_{k,0})$; 令 $j = 1$;

步骤 3 在展开深度小于 $UfDepthThreshold$ 且不满足停止准则 1, 在展开深度不小于 $UfDepthThreshold$ 且不满足停止准则 2 或停止准则 3 时, 以 v_k 为根展开 Tanner 图, 并且令展开树上所有输入节点为禁选状态; 满足停止准则后, 在可供选择的输入节点中随机选择一个输入节点 $i_{k,j}$; 将 $i_{k,j}$ 加到 $Adj(v_k)$, 将 v_k 加到 $Adj(i_{k,j})$;

步骤 4 $j = j + 1$, 直至 $j = d_k$, 重复步骤 3。

步骤 5 $k = k + 1$, 直至达到某个与信道条件相适应的输出节点数 m , 重复步骤 2。

本算法中参数 $UfDepthMAX$ 的选取影响输入节点的最小选择范围。较大的 $UfDepthMAX$, 缩小了输入节点随机选择的范围, 伤害了码的渐进性能, 但是能够间接起到增加 Tanner 图环长的作用; 减小 $UfDepthMAX$, 虽然扩大了输入节点随机选择的范围, 但 $UfDepthThreshold$ 会随之减小, 容易产生长度小的环, 提升错误平层。一般取 $UfDepthMAX \geq UfDepthThreshold \geq 1$, 使得展开图法即能够控制短环, 又使输入节点的选择范围尽可能地大。当 $UfDepthMAX = 0$ 和 $UfDepthThreshold = 0$ 时, 本法就是 RG 方法。

4 输入节点分布的理论分析和参数测试结果

设 N 个输入节点的平均度数为 b , m 个输出节点的平均度数为 a , 某个 $i_{k,0}$ 被选为 v_k 的首个邻居的概率 $p_0 = 1/N$; 在展开到 l 层后, 某个未被覆盖的输入节点被选为 v_k 的第 s 个邻居的概率为 $p_s = \frac{1}{N - s(ba)^l}$ 。因此某个输入节点被选为 v_k 的第 s 个邻居的概率 $q_s = p_s \prod_{i=0}^{s-1} (1 - p_i)$, 被选为 v_k 邻居的概率为 $\sum_{i=0}^{A-1} q_i$, A 为 v_k 的度数; 故某个输入节点是 r 个输出节点的概率为 $\binom{m}{r} \left(\sum_{i=0}^{A-1} q_i \right)^r \left(1 - \sum_{i=0}^{A-1} q_i \right)^{m-r}$, 输入节点的概率分布为 $o(x) = \sum_r \binom{m}{r} \left(\sum_{i=0}^{A-1} q_i \right)^r \left(1 - \sum_{i=0}^{A-1} q_i \right)^{m-r} x^r$ 。当 $N \rightarrow \infty$ 时,

$$\lim_{N \rightarrow \infty} N \sum_{i=0}^{A-1} q_i = \sum_{i=0}^{A-1} \lim_{N \rightarrow \infty} N p_i \prod_{j=0}^{i-1} (1 - p_j) = \sum_{i=0}^{A-1} \lim_{N \rightarrow \infty} N p_i = \sum_{i=0}^{A-1} 1 = A \quad (1)$$

$$E \left[\sum_{i=0}^{A-1} q_i \right] = \frac{E[A]}{N} = \frac{a}{N} \quad (2)$$

$$o(x) \approx \sum_r \binom{m}{r} \left(\frac{a}{N} \right)^r \left(1 - \frac{a}{N} \right)^{m-r} x^r \quad (3)$$

$$\lim_{N \rightarrow \infty} o(x) = \left[1 - \frac{a(1-x)}{N} \right]^m \quad (4)$$

由文献[6]可知 $o(x)$ 满足近似的泊松分布。本文所用的 LT 码的输出节点度数分布均为

$$\Omega(x) = 0.008x + 0.493x^2 + 0.166x^3 + 0.073x^4 + 0.083x^5 + 0.056x^8 + 0.037x^9 + 0.056x^{19} + 0.025x^{65} + 0.003x^{66}$$

平均度数 $\alpha = 5.87$ 。表 1 给出了用展开图法得到的几种码输入节点分布与泊松分布的差平方和, 泊松分布的数学期望 $\lambda = (m/N)\alpha = 2 \times 5.87 = 11.74$ 。码代号的含义是编码时 $UfDepthThreshold$ 和 $UfDepthMAX$ 的取值, 即当 $UfDepthThreshold$ 分别取 0, 1 和 2 时, $UfDepthMAX$ 分别取 $UfDepthThreshold$, $UfDepthThreshold + 1$ 和 N , 码代号就分别记为 00, 01, 0N; 11, 12, 1N; 22, 23, 2N。

表 1 本文所用几种码的参数测试结果

N/m 码代号	1000/2000		2000/4000				3000/6000	
	00	11	00	11	1N	22	00	11
输入节点分布与泊松分布的差平方和	0.0015	0.0040	0.0016	0.0036	0.0052	0.0034	0.0006	0.0039
输出节点环 4 密度	0.551	0.3545	0.4017	0.2147	0.227	0.22	0.331	0.104

由表 1 可见，随着码长的增加，输入节点分布越接近泊松分布，随着展开深度的增加接近效果越差，用 RG 方法得到的 00 码输入节点分布最接近泊松分布。虽然从表 1 还可以看出用本文方法得到码(例如 11 码)的输入节点分布接近效果稍差，但与 RG 方法相比环 4 密度大大减少，这与本法的设计目标和理论分析完全吻合。从表 1 中 $N=2000$ 的例子还可见，对中短长度码编码时，再增加 $UfDepthThreshold$ 和 $UfDepthMAX$ ，对于进一步减少环 4 数量效果不大。

5 仿真结果

本节介绍在误比特率为 0.02 的 BSC 信道条件下对输入节点长度 N 分别为 1000, 2000 和 3000 的若干种 LT 码进行译码仿真的结果。仿真时用 BP 算法对最多 50000 个全零码字进行 20 次迭代译码，直到有 10 个码字错误产生或 50000 码字全部被仿真时停止，计算 BER。图 2 给出了码长 2000 OVERHEAD(超量比)从 1.2 到 2 时，用 PEG, RG 和展开图法编码得到的 LT 码的译码性能。从图中可以看出，因为用 PEG 方法得到 LT 码的输入节点是平均分布，故其译码性能明显劣于用展开图法得到的 00 码和 11 码；因为从表 1 可见 00 码输入节点分布最接近泊松分布，故其渐进性能略优于 11 码；因为从表 1 可见 00 码输出节点的环 4 密度明显大于 11 码，故其误码平层明显高于 11 码。图 3 表示 $N=2000$ OVERHEAD 1.2 到 2 时，用本文方法构造出的 3 种码译码性能相差不多，这与从表 1 看到的这 3 种码的环 4 密度以及与泊松分布的差平方和都相差不多相互印证。图 4 表示使用相同编码参数的展开图法得到不同码长的 3 个 11 码，随着码长的增加，码的渐进性能变好，错误平层降低；从表 1 也可以看到随着码长的增加，这 3 个码的输入节点的分布越接近泊松分布，输出节点环 4 密度随之减少。

6 结束语

本文提出的 LT 码展开图构造方法，首先选择合适的 $UfDepthThreshold$ 以有效地控制短环的长度，根据码长选

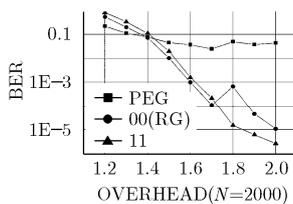


图 2 3 种编码方法的译码性能比较

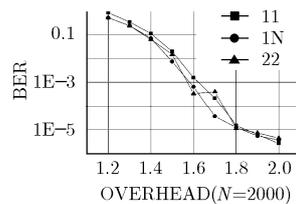


图 3 不同编码参数的译码性能比较

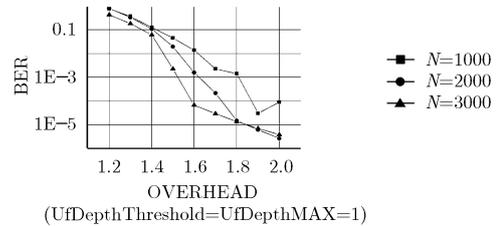


图 4 3 种码长的 11 码译码性能比较

择合适的 $UfDepthMAX$ 以保证输入节点一定的选择范围，然后展开以输出节点为根的 Tanner 图，在满足停止展开准则后随机选择输入节点与根相联。本法既满足了输入节点泊松分布的要求，又取得了短环控制和输入节点选择范围之间的均衡，进而取得了改善码的渐进性能和降低错误平层的效果。对于输入节点选择范围有限的中短长度 LT 码，本法编码效果优于传统的随机图方法和 PEG 方法。对于长 LT 码，虽然输入节点选择范围变大，短环产生的概率降低，但由于展开图法能够有效控制短环和确定输入节点的最小选择范围，因此它的设计思想对长 LT 码的编码仍有一定的借鉴作用。

参考文献

- [1] Mackay D J C. Fountain codes[J]. *IEEE Proc. Commun.*, 2005, 152(6): 1062-1068.
- [2] Mackay D J C. Good error correcting codes based on very sparse matrix[J]. *IEEE Trans. on Info. Theory*, 1999, 45(2): 399-431.
- [3] Wagner J P, Chakareski J, and Frossard P. Streaming of scalable video from multiple servers using rateless codes[C]. XVII International Conference on Electrical Machines, Chania, Crete Island, Greece, September 2-5, 2006: 1501-1504.
- [4] Luby M, Watson M, Gasiba T, Stockhammer T, and Xu Wen. Raptor codes for reliable download delivery in wireless broadcast systems[C]. IEEE Consumer Communications and Networking Conference Las Vegas, Nevada, USA, January 8-10, 2006: 192-197.
- [5] Luby M. LT codes[C]. Proc. 43rd Ann. IEEE Symp. on Foundations of Computer Science, Vancouver, BC, Canada, November 16-16, 2002: 271-282.
- [6] Shokrollahi A. Raptor codes[J]. *IEEE Trans. on Info. Theory*, 2006, 52(6): 2551-2567.

- [7] Luby M, Gasiba T, Stockhammer T, and Watson M. Reliable multimedia download delivery in cellular broadcast network[J]. *IEEE Trans. on Broadcasting*, 2007, 53(1): 235-246.
- [8] Pakzad P and Shokrollahi A. Design principles for raptor codes[C]. 2006IEEE Info. Theory Workshop, Uruguay, March 13-17, 2006: 165-169.
- [9] Zhu Hongjie, Zhang Chao, and Lu Jianhua. Designing of fountain codes with short code-length[C]. The Third International Workshop on Signal Design and Its Applications in Communications, Chengdu, China, September 23-27, 2007: 65-68.
- [10] Puducheri S, Klierer J, and Fuja T E. The design and performance of distributed LT codes[J]. *IEEE Trans. on Info. Theory*, 2007, 53(10): 3740-3754.
- [11] Etesami O and Shokrollahi A. Raptor codes on binary memoryless symmetric channels[J]. *IEEE Trans. on Info. Theory*, 2006, 52(5): 2033-2051.
- [12] Hyytia E, Tirronen T, and Virtamo J. Optimal degree distribution for LT codes with small message length[C]. IEEE INFOCOM 2007, Anchorage, Alaska, USA, May 6-12, 2007: 2576-2580.
- [13] Hu X Y, Eleftheriou E, and Arnold D M. Regular and irregular progressive edge-growth tanner graphs[J]. *IEEE Trans. on Info. Theory*, 2005, 51(1): 386-398.
- 龚茂康: 男, 1957年生, 副教授, 研究方向为信道编码及其实现、数字信号处理.