线性分组码中盲码率恢复的特性分析

Atreya Vedantam印度理工学院马德拉斯分校电气工程系印度钦奈 ee22b004@smail.iitm.ac.in

RadhaKrishna Ganti印度理工学院马德拉斯分校电气工程系印度钦奈 rganti@ee.iitm.ac.in

arXiv:2603.02031v1 [cs.IT] 2 Mar 2026

摘要——**前向纠错（FEC）广泛应用于通信链路中。我们研究非合作解码，目标是恢复线性分组码的码率。我们提出一种用于表征码率恢复质量的度量，该度量可采用任意基于秩的估计技术。我们推导出该度量关于算法参数和环境参数的闭式表达式，并指出，恢复效果良好时该度量应较低。我们利用该度量推导出高噪声条件下更优码率估计的表达式，并将其与现有估计方法进行比较。最后，我们通过模拟低密度奇偶校验（LDPC）码的恢复，验证所推导的度量表达式以及码率估计性能的提升。我们还据此推导出恢复所需的最优算法参数。**

索引词—**盲检测、码率估计、纠错码、基于秩的技术。**

I. 引言

前向纠错（FEC）码参数的盲恢复长期以来备受关注，并应用于自适应调制与编码（AMC） [1],认知无线电系统 [2] 和频谱监测 [3]。在这种场景下，接收机并不知道码参数，因此无法对消息进行解码。接收机必须依赖通过大量编码帧获得的统计信息。

针对这类非合作场景下的盲恢复问题，人们提出了多种算法。在 [4], Valembois证明，求解低秩矩阵的一般问题是 NP 完全的，并提出了一种用于寻找最小重量码字的启发式算法（最初由 Canteaut 和 Chabaud [5]提出）。在此基础上，Cluzeau [6] 和 Barbier [7]采用秩约简技术寻找对偶码字，并通过权衡较高的检测概率与较高的虚警概率来调整算法参数。然而，上述方法都仅使用硬判决比特，没有利用符号级信息（软信息）。Sicot [8] 引入了利用符号级信息来辅助恢复对偶码字的方法，但未提供性能分析。Xia [9], 、Moosavi [10] 和 Luwei [11] 使用基于对数似然比（LLR）的技术估计码。然而，他们的方法都假设候选码集合已知。Ramabadran [12] 采用了基于秩的方法，但固定了恢复矩阵的纵横比，且未进行分析。Wang [13]采用了基于符号级信息的技术，但没有量化算法参数的最优性。

本文重点研究线性分组码的盲恢复。我们使用符号级信息进行筛选并构建矩阵，方法与Wang等人类似。随后，我们提出一个用于表征码率恢复质量的度量指标。我们阐述该指标的重要性，并推导出以算法参数和环境参数表示的闭式表达式。我们利用该指标推导出高噪声条件下更优码率估计的表达式，并将其与现有估计方法进行比较。最后，我们通过在噪声条件下模拟LDPC码的恢复，验证所推导的度量指标表达式及码率估计性能的提升，并推导出码恢复的最优参数。

II. 数学模型

我们假设信源生成独立同分布的比特。这些比特被划分为大小为 k 的块，每个块称为一条消息，记为$\mathbf{m}$。每条消息乘以一个未知（但相同）的大小为 $k \times n$ 的矩阵，该矩阵称为生成矩阵（记为G），从而生成一个码字，记为$\mathbf{c}$。数学上，

$$\mathbf{c} = \mathbf{mG}, \quad \mathbf{c} \in \{0,1\}^n, \mathbf{m} \in \{0,1\}^k, \mathbf{G} \in \{0,1\}^{k \times n}.$$

本文中对比特进行的所有运算均在 GF(2)中完成。我们不对生成矩阵的形式作任何假设（例如，不假设其为系统形式）。然而$\mathbf{G}$表示一个单射映射，因此其秩为 k。因此，在$\mathbf{c}$ 的 n 个比特中，恰好有 k 个比特是独立 的，其余 $n-k$ 个比特线性依赖于这 k 个比特。码字$\mathbf{c}$ 中的这些独立比特称为信息比特 ，而依赖比特称为校验比特。这是我们在码参数恢复中加以利用的一个关键观察。比值 $\rho = k/n$ 称为码率。我们假设 c中每个比特为0或1的可能性相等。这在实际场景中是一个合理假设。随后，每个码字$\mathbf{c}$ 采用二进制相移键控（BPSK）进行调制，以生成符号向量，记为$\mathbf{b} = 1-2\mathbf{c}$。然后将该向量（逐符号）送入加性高斯白噪声（AWGN）

信道，其信噪比（SNR）未知，并被接收为一个向量，记为 $\mathbf{r} \in \mathbb{R}^n$。因此，

$$\mathbf{r} = \mathbf{b} + \mathbf{v}, \tag{1}$$

其中$\mathbf{v} \in \mathbb{R}^n = [v_1, v_2, \ldots, v_n]$是从方差为 σ^2的正态分布中抽取的独立同分布噪声样本向量，其概率密度函数（PDF）定义为

$$f(v_k) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-v_k^2}{2\sigma^2}\right), \quad k = 1, \ldots, n.$$

$\sigma^2 = N_0/2$，其中 N_0 为噪声功率谱密度。向量$\mathbf{r}$ 也对应于发送符号的对数似然比，因此称为对数似然比向量。接收机对该向量进行硬判决译码，生成接收比特向量$\mathbf{y} \in \{0,1\}^n$，其各分量的生成方式如下：

$$y_i = \begin{cases} 1 & r_i < 0, \\ 0 & r_i \geq 0. \end{cases}$$

在任何通信系统中，多个这样的码字都会在较长时间内被发送和接收。我们假设发送 M个码字。记为$\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_M\}$的 M 个码字采用BPSK映射进行调制，并通过加性高斯白噪声信道发送。相应的接收符号向量记为 $\mathcal{R} = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M\}$。在本文余下部分， $\Phi(\cdot)$表示标准正态分布的累积分布函数（CDF），定义如下：

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(\frac{-u^2}{2}\right) du,$$

且 $Q(\cdot)$ 表示标准 Q函数，其定义为

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} \exp\left(\frac{-u^2}{2}\right) \mathrm{d}u.$$

本文假设已使用 [6] 中的技术实现时间同步，并使用 [12] 中的技术确定了码字 n的长度。接收机移除接收帧中的导频比特、前导码和帧头，从而获得码字帧。我们还假设恢复过程有足够数量的码字 M，其具体数量将在后文量化。目标是利用 $\mathcal{R}$ 以稳健的方式恢复码率 ρ。

III. 算法

在高信噪比场景下（或码字长度较短时），可以使用诸如 [14] 的秩估计方法，在盲场景下估计码率。然而，在存在噪声的情况下，尤其是分组长度较大时，现有的基于秩的方法会失效，因为接收帧中的错误数量会导致秩估计不准确。穷举法的计算开销也会变得难以承受。本节介绍一种新的稳健算法，该算法基于软信息，通过观测 $\mathcal{R}$ 来恢复码率 ρ。该算法以参数 $\{\mathcal{R}, t_1, t_2, n\}$ 作为输入，并输出码率估计 $\hat{\rho}$。

1) 使用以下方法估计噪声方差 σ^2

$$\frac{1}{M}\sum_{i=1}^{M} \mathrm{Var}(\mathbf{r}_i) = 1 + \hat{\sigma}^2.$$

这是因为在式 **(1)** 中，噪声向量与符号向量相互独立，因此它们的方差相加。对所有消息的该估计值取平均，可得到噪声方差的良好估计。

2）估计信噪比：

$$\hat{\mathrm{SNR}} = \hat{\sigma}^{-2}.$$

估计 $Q\left(\sqrt{\quad}\right)$ p_e p_e =，即比特错误的概率，为 ˆSNR。

3) 对所有 $i \in [1, 2, \ldots, M]$迭代执行以下过程。假设 $t_1 \in [0,1]$ 和 $t_2 \in \mathbb{N}$ 是算法的两个参数。对于每个$\mathbf{r}_i$，考察其第 k个元素 r_i^k。如果$|r_i^k| < t_1$，则将相应的 r_i^k标记为不可靠。令j_i 为$\mathbf{r}_i$中不可靠比特的总数。如果 $j_i < t_2$，则将接收向量$\mathbf{r}_i$ 标记为合适。

4) 假设找到 M_s 个合适的接收向量。对这些合适的接收向量进行解码（采用 BPSK阈值判决），并将解码后的比特向量（写作行向量）逐行排列，生成一个大小为 $M_s \times n$的矩阵，称为字矩阵，记为$\mathbf{W}$。注意，$\mathbf{W}$ 中的每一列要么包含全部信息比特，要么包含全部校验比特，因为所有码字都属于同一个码 $\mathcal{C}$。这些列分别称为信息列（记为$\mathbf{W}_m$）和校验列（记为$\mathbf{W}_p$）。

5) 对字矩阵 $\mathbf{W}$ 应用高斯消元算法，得到如下形式的行最简阶梯形（RREF）（这需要进行列置换），

$$\mathbf{R} = \begin{pmatrix} \mathbf{I}^{k\times k} & \mathbf{D}^{k\times n} \\ \mathbf{0}^{(M_s-k)\times k} & \mathbf{0}^{(M_s-k)\times(n-k)} \end{pmatrix},$$

其中，$\mathbf{D}$ 是一个矩阵，用于表示字矩阵各列之间的所有线性依赖关系，而$\mathbf{I}$是标准单位矩阵。

6) 求每列的均值 i，记为 $\mu(R_i)$，其中$\mathbf{R}$。

令 k' 表示 $\mathbf{R}$ 中满足$\mu(R_i) \leq 1/M_s$的列数。则码率估计为

$$\hat{\rho} = \frac{k' - \mathbb{E}[C]}{n - \mathbb{E}[C]}, \tag{2}$$

其中，$\mathbb{E}[C]$ 定义见式（5）。$\mathbb{E}[C]$的近似值见式（6）。

IV. 分析

本节将阐释上述算法的直观原理，并引入一种用于分析码率恢复质量的度量。注意，每个校验比特都是若干信息比特的线性组合（模 2）。这样的线性相关集合共有 $n-k$个。因此，在无噪声码字

矩阵中，有 $n-k$ 个校验列依赖于信息列。假设信息列的一个子集$\mathbf{w}_i \subseteq \mathbf{W}m$ 以及一个校验列 $p_i \in \mathbf{W}p$ （其中 $i \in \{1,2,\cdots,n-k\}$）构成一个线性相关集合。也就是说，存在一个二进制向量 $\lambda_i \in \{0,1\}^k$和一个标量 $\alpha_i \in \{0,1\}$，使得 $\lambda_i \cdot \mathbf{w}_i + \alpha_i p_i = 0$ （其中“$\cdot$”运算是 GF（2）上的通常内积）。我们用以下方式表示这个集合：

$$\mathcal{S}_i = \{\mathbf{w}_i, p_i\}, i \in \{1,2,\cdots,n-k\}.$$

现在假设整个字矩阵中恰好有一个错误，且该错误出现在校验列 p_i 中。只有 $\mathcal{S}_1$ 会受到此错误的影响，秩估计也会增加 1。另一方面，假设该错误出现在信息列 w_i 中。这会影响所有满足 $w_i \in \mathcal{S}_j$ 的子集 $\mathcal{S}_j$ ，可能导致多个秩估计值增大。因此，要实现良好的恢复，多个列都不应出现比特错误。

我们认为，字矩阵中至少出现一个比特错误的列数可以表征恢复质量。为此，我们将说明，某列中至少出现一个比特错误时，秩估计很可能会增大。

我们考虑一个玩具模型，其中有 d 个信息列，以及一个与所有信息列线性相关的校验列。根据上述讨论，单个错误会使秩估计从 d 增至 $d+1$。在字矩阵中至少存在一个比特错误的条件下，只有当这些比特错误保留线性相关性时，我们才能得到正确的秩估计。换言之，出错的校验列必须仍位于出错的信息列的张成空间中。我们将说明，对于较大的字矩阵，这种情况发生的概率极小。

定理 1. 假设 p、m_1、m_2、.....、$m_d \in \{0,1\}^{M_s}$ 是字矩阵的列，且满足 $p = \bigoplus_{i=1}^{d} m_i$。令 p'_e 为码字矩阵中发生比特错误的概率。假设 $M_s >> d$ 和 p'_e 较小，使得 $M_s p'_e$ 较大。那么，在至少发生一个错误（记为 p_1）的情况下，秩增加1 的概率下界如下。

$$p_1 \geq \Phi\left(\frac{(M_s-d)^2 - 2M_s^2(d+1)p'_e}{2M_s(d+1)\sqrt{M_s p'_e(1-p'_e)}}\right). \tag{3}$$

证明。详细证明见附录。证明概要：我们（几乎必然）给出由 $\{m_1, m_2, \ldots, m_d\}$ 的张成空间构成的码的最小距离下界。该距离随 M_s 的增大而增大。现在，假设矩阵中至少发生一个比特错误，那么若要保持线性相关性，错误数量就必须足以“跨越”此最小距离。对于较小的比特错误概率和较高的 M_s，这种情况发生的概率趋近于零，证毕。□

将 (3) 简化为 $M_s >> d$ 并假设 $p'_e << 1/(d+1)$，我们得到

$$p_1 \approx \Phi\left(\frac{\sqrt{M_s}}{2(d+1)\sqrt{p'_e(1-p'_e)}}\right).$$

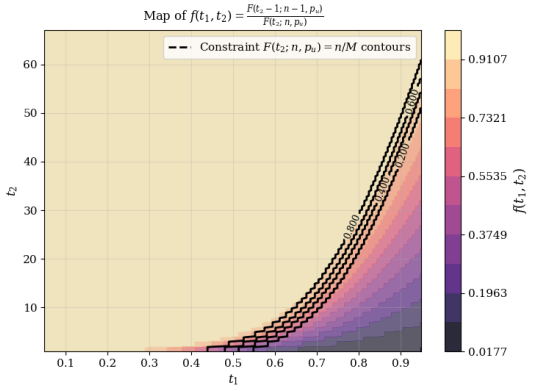


图 1. 在信噪比为 10 dB 时，长度为 $n = 136$ 的码的算法误差随 t_1 和 t_2 变化的等高线图。$t_1^* = 0.99$ 和 $t_2^* = 1$ 以及最小 f（t_1^*, t_2^*）$= 0.0153$。黑线表示恒定的 M 曲线（见公式 (9)）。

对于 $M_s = 10^3$、$d = 10$、$p'_e = 0.02$ 这样的合理选择，p_1 几乎为 1。因此，当噪声足够大，使得 $M_s p'_e$ 很大时，码恢复将不准确。这也说明了为什么至少包含一个比特错误的列数能够体现良好的码恢复质量。

现在，我们通过解析推导至少包含一个比特错误的列的期望数量。

定理 2. 令 p_e 表示比特错误的概率。令p_u 表示比特不可靠的概率。

(1) 那么， p_u 取决于参数 t_1 ，其关系为

$$p_u = Q\left(\frac{1-t_1}{\sigma}\right) - Q\left(\frac{1+t_1}{\sigma}\right). \tag{4}$$

(2) 若 $F(k;n,p) = \sum_{i=0}^{k}\binom{n}{i}p^i(1-p)^{n-i}$ 表示二项分布累积分布函数，则至少包含一个错误的列的期望数量为

$$\mathbb{E}[C] \approx n - n\left(1 - \frac{p_e F(t_2-1;n-1,p_u)}{F(t_2,n,p_u)}\right)^{M_s}. \tag{5}$$

当 p_e 足够小，使得 $M_s p_e << 1$时，我们有

$$\mathbb{E}[C] = n\mathbb{E}[M]p_e F(t_2-1;n-1,p_u), \tag{6}$$

其中， $\mathbb{E}[M]$ 是获得 M_s 个合适码字所需的噪声码字总期望数量。其表达式为

$$\mathbb{E}[M] = \frac{M_s}{F(t_2;n,p_u)}. \tag{7}$$

证明。详细证明请参见附录。证明概要：第一部分可通过对“非可靠”区域的正态概率密度函数积分得出。对于第二部分， $\mathbb{E}[C]$ 与某一列至少出现一个比特错误的概率成正比。该概率取决于每个码字中预期的比特错误数。可以利用每个码字中非可靠比特和可靠比特数量的分布，以及其对比特错误数的影响，计算这一数量。

□

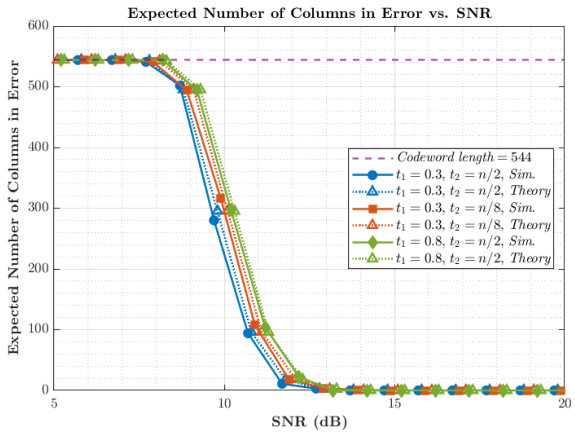


图 2. 三种情况下的理论值与仿真值 $\mathbb{E}[C]$： $t_1 = 0.3$， $t_2 = n/2$、$t_1 = 0.3$ $t_2 = n/8$ $t_1 = 0.8$ $t_2 = n/2$ 1,000 544，、，针对长度为 的 LDPC 码字消息，在信噪比为 5 至 20 dB 时的结果。两者高度吻合，证明了理论结果在广泛的 t_1 和 t_2 范围内均成立。

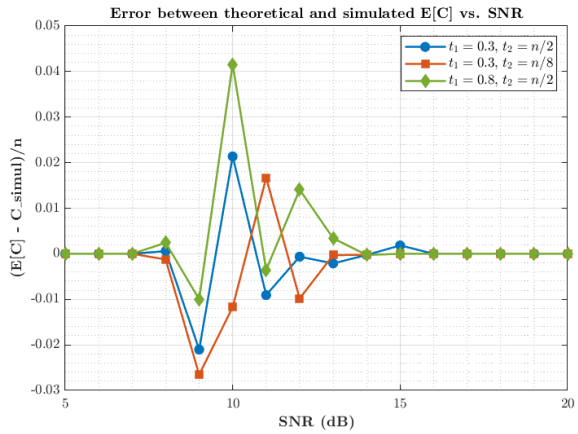


图3. 理论 $\mathbb{E}[C]$ 与观测出错列数之间的误差，已按码长（544）归一化。该误差在高信噪比下趋于0。

公式 (6) 具有直观的含义。我们注意到， nM_s是字矩阵中的比特总数。将其乘以 p_e ，即可得到这些比特中因环境因素而出错的比例。因此，我们将项 nM_sp_e 称为环境错误。另一项则表示算法优化降低错误的倍数。我们将这一项称为算法错误。最优的 t_1 和 t_2 值可使 $\mathbb{E}[C]$最小。因此，最优值会使算法误差最小。

$$t_1^*, t_2^* = \arg\min_{t_1,t_2} \frac{F(t_2 - 1; n - 1, p_u)}{F(t_2; n, p_u)}.$$

需要注意的是，这也对恢复消息的期望数量提出了（相当严格的）要求（见 (7)）。这一点可从图 (1) 中看出。

图 (1) 展示了码长为 $n = 136$ 、信噪比为 10 dB 时，算法误差随 t_1 和 t_2 变化的等高线图。四条黑线表示函数 $F(t_2; n, p_u)$ 分别等于 0。2、0。4、0。6 和 0。8 时的等高线。较低的值对应较高的（期望）消息需求，如式 (7) 所示。图中可以看出，随机选择 t_1 和 t_2 会产生不利影响，因为大部分区域的算法误差都很高。我们还观察到，目标函数的等高线几乎与常数 $F(t_2; n, p_u)$ 的等高线平行。因此，最小化算法误差也会大幅降低 $F(t_2; n, p_u)$。这意味着我们需要观察更多消息，才能找到合适的 M_s 码字。在某些情况下，这可能无法实现。当 M 成为固定值时，我们将重新表述该问题。

如果 $M_s < n$，字矩阵的秩将受到M_s的限制。由于我们不希望出现这种情况，因此必须有 $M_s \geq n$。对于较大的 M_s，$\mathbb{E}[C]$ 也会变大，因为每一列更长，更容易受到比特错误的影响。因此，为了在秩估计的限制与更多错误之间取得平衡，最优做法是设定 $M_s = n$。于是(7)对 t_1 和 t_2 可以取的值施加了约束。在这种情况下，最优值由下式给出

$$t_1^*, t_2^* = \arg\min_{t_1^*,t_2^*} F(t_2 - 1; n - 1, p_u) \tag{8}$$

$$\text{s.t. } F(t_2; n, p_u) = \frac{n}{M}. \tag{9}$$

由于等高线几乎平行，沿该约束无法进行太多优化。由此可直观得出结论：环境误差是恢复中更为显著的限制因素。

假设 k' 是字矩阵的观测秩。显然$k' > k$。差值 $k'-k$ 表示字矩阵的线性相关子集中因比特错误而不再呈现该关系的列数。平均而言，至少含有一个比特错误的 $\mathbb{E}[C]$ 列中有 $1-k/n$ 比例为校验列。这些列中的每一列必然贡献一次秩增加。因此我们可以说

$$k' - k \geq \left(1 - \frac{k}{n}\right) \mathbb{E}[C].$$

整理后可得一个码率估计，该估计用于算法中的方程（2）。

$$\hat{\rho} = \frac{k}{n} \approx \frac{k' - \mathbb{E}[C]}{n - \mathbb{E}[C]}. \tag{10}$$

方程（10）也可以作直观解释。如果我们设想从估计的信息列中移除 $\mathbb{E}[C]$ 列，则这个新码的码率是更好的估计。

V. 仿真结果

本节中，我们展示了$\mathbb{E}[C]$ 的理论估计与实际噪声场景中的观测值高度吻合。随后，我们展示了该算法中针对码率估计提出的修正方法在较宽的信噪比范围内均表现良好。

我们使用 MATLAB 对实验进行仿真。我们生成 1000 条随机二进制消息，并使用码长为 544 的 LDPC 码对其进行编码。我们采用 5G 新空口（NR）的标准 LDPC 编码。该码未打孔，其真实

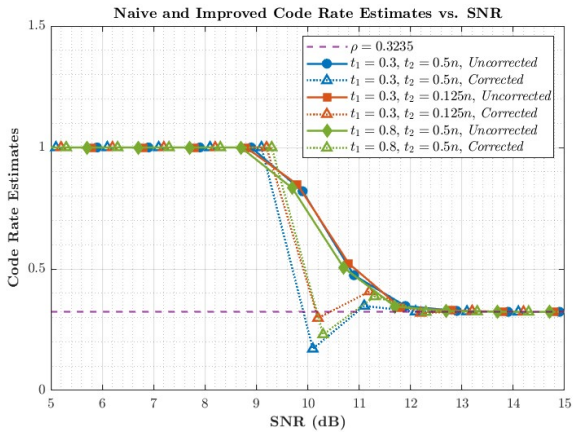


Fig. 4朴素码率估计 k'/n （实线）和改进估计（$k' - \mathbb{E}[C]$）/（$n - \mathbb{E}[C]$）（虚线）随信噪比变化的情况，针对 t_1 和 t_2的三种不同情形。仿真采用长度为544的LDPC码的1000条消息进行，其真实码率为0.3235。

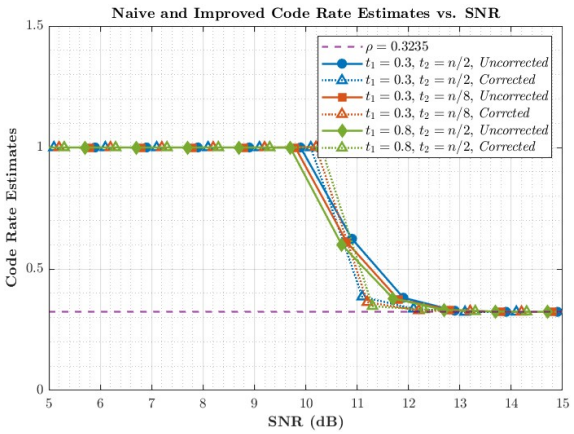


Fig. 5三种不同 t_1 和 t_2情况下，朴素码率估计 k'/n（实线）和改进后的估计（$k' - \mathbb{E}[C]$）/（$n - \mathbb{E}[C]$）（虚线）随信噪比的变化。仿真针对一个长度为 1088、真实码率为 0.3235 的 LDPC码，使用 2000 条消息进行。

码率为 $\rho = 0.3235$（接近 $1/3$）。每个码字均采用 BPSK 调制，并通过信噪比从 5 dB 到 20 dB 不等的信道发送。我们使用第 2 节中描述的恢复算法。

$\mathbb{E}[C]$的公式首先在广泛的t_1 、 t_2 取值及信噪比范围内进行了验证。如图2所示。由图3可见，错误小于5%。

图4展示了不同信噪比下的朴素码率估计以及所提出的校正方法。为避免重叠并使图示更清晰，图中的点在水平方向上略微错开。在信噪比为5至9 dB时，由于错误，字矩阵变为满秩，因此不难看出所提出的校正方法没有帮助（$k' = n \implies k' - \mathbb{E}[C] = n - \mathbb{E}[C]$）。在信噪比为10 dB时，字矩阵不会变为满秩，而所提出的校正方法使估计码率迅速接近真实码率0.3235。在这种情况下，朴素码率估计直到信噪比达到12 dB时才收敛到真实码率。该经验结果表明，校正方法能够使我们在比朴素估计低2 dB的信噪比下准确估计码率。我们还看到，码率估计的准确度恰好遵循图2中的 $\mathbb{E}[C]$ 曲线。这验证了 $\mathbb{E}[C]$能够表征码率估计的质量。

下面我们将这一修正后的估计与现有方法进行比较：1) Ramabadran [12] 对采用 BPSK 调制的 [648, 162] LDPC码进行仿真，并在信噪比为 $\geq$ 10 dB 时恢复码率，其性能与我们的方法相当。使用的码字数量为 1296，与我们的仿真相近。2) Wang [13] 对 [648, 540] LDPC码进行仿真，并在误码率为 1.5×10^{-3} 时恢复该码；对于采用 BPSK 调制的信号流，这相当于 8.5 dB 的信噪比。然而，恢复过程需要 3,840,000 个码字，而我们提出的算法仅用 1000 个码字就能恢复大小相近的码。

我们还观察到，与其他选择相比，使用 $t_1 = 0.3$ 和 $t_2 = n/2$ 时，10 dB 信噪比下的码率估计较差。这突显了优化 t_1 和t_2的重要性。在 10 dB 信噪比下，发现 $t_1 = 0.07$ 和 $t_2 = 0.007$ 接近通过求解 (9) 得到的最优参数。这是通过对参数 t_1 和 t_2 在$[0, 1]$范围内进行网格搜索得出的。优化景观较为平滑，因此在合理的分辨率下，网格搜索能够得到接近最优的值。

对于较长的分组长度，由于字矩阵规模增大，恢复会更加困难。我们生成了 2000 条长度为 1088、码率相同为 $\rho = 0.3235$ 的 LDPC 码的随机二进制消息。该仿真的结果如图 5 所示。可以看到，当字矩阵秩亏时（信噪比为 ≥ 11 dB），所提出的校正方法能显著地将朴素估计值拉回真实值。

对于数量级为 10^4，的超长分组长度，实际中很难获得超过（数量级为） 10^4 个码字。因此，环境误差为 $nM_s p_e = 10^8 p_e$ ，与 10^4 相比应当足够小。由此，实践中要实现良好的恢复，我们需要 $p_e \approx 10^{-6}$，这要求信噪比至少达到 10 dB。

VI. 结论

我们提出了一种度量指标（$\mathbb{E}[C]$），用于表征线性分组码中任何基于秩的估计技术对码参数进行盲恢复的质量。我们阐述了采用该度量指标的理由：当至少存在一个错误时，字矩阵的秩以较高概率增加 1。随后，我们推导出了该度量指标的闭式表达式，该表达式取决于两个因素：环境误差和算法误差。

随后，我们利用该指标推导出更优的码率估计。最后，我们最小化算法误差，从而针对两种实际约束推导出恢复的最优参数值：一种是接收机能够控制所接收含噪码字数量的情况，另一种是接收机无法控制的情况。我们通过在不同信噪比下仿真长度为544的低密度奇偶校验码，并将其与现有类似方法进行比较，验证了上述各项理论结果。

我们讨论了在码字长度较大时，实现合理码恢复所需信噪比的一个实用下界。

参考文献

[1] C. Mehlfuehrer、S. Caban 和 M. Rupp，“有限反馈条件下MIMO WiMAX中自适应调制与编码的实验评估”，*EURASIP*信号处理进展期刊，第2008卷，第837102页，2008年。[2] S. Haykin，“认知无线电：大脑赋能的无线通信”，*IEEE*通信领域精选期刊，第23卷，第2期，第201–220页，2005年。[3] T. Yücek 和 H. Arslan，“认知无线电应用中的频谱感知算法综述”，*IEEE*通信综述与教程，第11卷，第1期，第116–130页，2009年。[4] A. Valembois，“二元线性码的检测与识别”，离散应用数学，第111卷，第199–218页，2001年。[5] A. Canteaut 和 F. Chabaud，“在线性码中寻找最小重量码字的新算法：其在McEliece密码系统和长度为511的窄义BCH码中的应用”，*IEEE*信息论汇刊，第44卷，第1期，第367–378页，1998年。[6] M. Cluzeau 和 M. Finiasz，“从噪声截获比特流中恢复码长和同步信息”，载于*IEEE*国际信息论研讨会（*ISIT*）会议论文集，韩国首尔，2009年，第2737–2741页。[7] J. Barbier、G. Sicot 和 S. Houcke，“线性纠错码与卷积纠错码重构的代数方法”，载于第三届国际计算机科学与工程会议（*CISE 2006*）论文集，C. Ardil 编，第16卷。意大利威尼斯：World Enformatika Society，2006年，第66–71页。[8] G. Sicot 和 S. Houcke，“交织器参数的盲检测”，载于*IEEE*国际声学、语音与信号处理会议（*ICASSP*）论文集，美国费城，2005年。[9] T. Xia 和 H.-C. Wu，“利用综合征的平均LLR进行LDPC码的新型盲识别：后验概率”，*IEEE*信号处理汇刊，第62卷，第3期，第632–640页，2014年。[10] R. Moosavi 和 E. G. Larsson，“信道码的快速盲识别”，*IEEE*通信汇刊，第62卷，2014年，第1405、1393页，第5期。[11] L. Luwei 和 L. Yingke，“LDPC码码长和码率的编码识别算法”，载于*2018*年*IEEE*第四届信息技术与机电工程会议（*ITOEC*）论文集，中国重庆，2018年，第1460–1463页。[12] S. Ramabadran、A. M. Kumar、W. Guohua 和 T. S. Kee，“有误差的信道条件下LDPC码参数的盲识别”，*IET*信号处理，第13卷，第86–95页，2019年。[13] P. Wang、Y. L. Guan、L. Wang 和 P. Cheng，“噪声信道中线性分组码的快速盲恢复”，载于*2023*年*IEEE*国际信息论研讨会（*ISIT*）论文集，中国台湾台北，2023年，第264–269页。[14] V. Kuvaja，“信号情报中的纠错码识别”，硕士论文，电气工程学院，阿尔托大学，埃斯波，芬兰，2015年。[15] A. Barg 和 G. D. Forney，“随机码：最小距离与错误指数”，*IEEE*信息论汇刊，第48卷，第9期，第2568–2573页，2002年。

附录 A 定理 1 的证明

证明。我们首先观察到，列向量$m_1, m_2, \ldots, m_d$是随机向量。这些向量的张成空间构成一个随机线性码 S。该码 d_m 的最小距离几乎必然有界于

$$H_2\left(\frac{d_m}{M_s}\right) \geq 1 - \frac{k}{M_s},$$

其中， $H_2(\cdot)$ 是二元熵函数 [15]。因此，我们得到

$$2\sqrt{\frac{d_m}{M_s}\left(1-\frac{d_m}{M_s}\right)} \geq 1 - \frac{k}{M_s},$$

并忽略相对于 1 的 d_m/M_s ，可得

$$d_m \geq \frac{(M_s-k)^2}{2M_s}. \tag{11}$$

现在假设每个 m_i 和 p 分别被噪声向量e_i和 e 破坏。因此 $m_i' = m_i \oplus e_i$和 $p' = p \oplus e$。无噪声的 p 位于 S之内。为保持秩， p' 必须位于 $m_1', m_2', m_3', \ldots, m_d'$的张成空间中。在最坏情况下，$p' = p \oplus e = \bigoplus_{i=1}^{d} m_i' = \bigoplus_{i=1}^{d}(m_i \oplus e_i)$意味着$p \oplus (e \bigoplus_{i=1}^{d} e_i)$位于 $m_1, m_2, m_3, \ldots, m_d$的张成空间中。若上述 $d+1$ 个错误向量之和的总重量不超过 d_m，则出错向量 p' 不可能位于新的张成空间中，因此秩必须增加1。因此我们有

$$p_1 \geq \mathbb{P}\left(\left|e \bigoplus_{i=1}^{d} e_i\right| \leq d_m\right).$$

当所有错误向量的支撑集互不相交时，上式取得下界，因此

$$p_1 \geq \mathbb{P}\left(|e| \leq \frac{d_m}{d+1}\right).$$

由于 e 是由服从伯努利分布的1构成的向量，因此$|e|$服从二项分布，均值为 $M_s p_e'$ ，方差为$M_s p_e'(1-p_e')$。对于 $M_s >> d$ ，在利用(11)代入 d_m 之后，这可用正态累积分布函数来近似。至此证明完毕。 □

附录 B 定理 2 的证明

证明。一个比特不可靠的概率与它是0还是1无关。因此 p_u 简单地由

$$p_u = \int_{1-t_1}^{1+t_1} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-\nu^2}{2\sigma^2}\right) d\nu,$$

这可平凡地化为方程(4)。我们定义若干随机变量以辅助证明方程(5)。下面定义的所有随机变量均以码字选自字矩阵为条件，因而具有 t_2个或更少的不可靠比特。对每个码字，令 U_k 和 R_k 为随机变量，分别表示给定的 k 个不可靠比特和 $n-k$ 个可靠比特中的比特错误数。令E_k 表示在一个码字具有 k 个不可靠比特的条件下该码字中的比特错误数。于是我们有

$$E_k = U_k + R_k.$$

令 C_i 为一个随机变量：若字矩阵的第 i列至少有一个比特错误，则取值为 1，否则取值为 0。令 C 为一个随机变量，表示字矩阵中至少有一个比特错误的列数。于是有 $C = \sum_{i=1}^{n} C_i$。令 E 为表示一个码字中比特错误总数的随机变量，令 K 为表示一个码字中不可靠比特数的随机变量。令 M 表示所需的

码字数，直到我们获得足够数量的合适码字用于恢复。因此，根据 C_i的相同分布，

$$\mathbb{E}[C] = \sum_{i=1}^{n} \mathbb{E}[C_i] = n\mathbb{E}[C_1] = n\mathbb{P}[C_1 = 1]. \tag{12}$$

根据全期望定理，

$$\mathbb{P}[C_1 = 0] = \left(\sum_{e=0}^{n} \left(1 - \frac{e}{n}\right) \mathbb{P}[E = e] \right)^{M_s}$$

$$\implies \mathbb{P}[C_1 = 0] = \left(1 - \frac{\mathbb{E}[E]}{n}\right)^{M_s}. \tag{13}$$

现在， $\mathbb{E}[E]$ 可以写为

$$\mathbb{E}[E] = \sum_{k=0}^{n} \mathbb{E}[E_k]\mathbb{P}[K = k] = \sum_{k=0}^{n} (\mathbb{E}[U_k] + \mathbb{E}[R_k])\mathbb{P}[K = k].$$

令$p_{e|u}$ 表示不可靠比特出错的概率， $p_{e|r}$ 表示可靠比特出错的概率。则 U_k 和 R_k 分别服从二项分布：$U_k \sim$ Binomial(k, $p_{e|u}$) 和 $R_k \sim$Binomial($n-k$, $p_{e|r}$)。代入它们的期望值后，

$$\mathbb{E}[E] = \sum_{k=0}^{n} (kp_{e|u} + (n-k)p_{e|r})\mathbb{P}[K = k].$$

K 服从截断二项分布（因为我们只考虑字矩阵）。因此，K 的概率质量函数（PMF）为

$$\mathbb{P}[K = k] = \begin{cases} \frac{\binom{n}{k} p_u^k (1-p_u)^{n-k}}{F(t_2; n, p_u)}, k \le t_2 \\ 0, k > t_2. \end{cases}$$

因此，经过一些化简后，该求和可简化为

$$= \sum_{k=0}^{t_2} (kp_{e|u} + (n-k)p_{e|r}) \left\{ \frac{\binom{n}{k} p_u^k (1-p_u)^{n-k}}{F(t_2; n, p_u)} \right\},$$

$$= \frac{np_{e|u}p_u}{F(t_2; n, p_u)} \sum_{k=1}^{t_2} \binom{n-1}{k-1} p_u^{k-1} (1-p_u)^{n-k}$$

$$+ \frac{np_{e|r}(1-p_u)}{F(t_2; n, p_u)} \sum_{k=0}^{t_2} \binom{n-1}{k} p_u^k (1-p_u)^{n-k-1}.$$

上面的两项都是二项分布累积分布函数。根据我们的记号，上式可简化为

$$= \frac{np_{e|u}p_u F(t_2 - 1; n-1, p_u) + np_{e|r}p_r F(t_2; n-1, p_u)}{F(t_2; n, p_u)}.$$

我们观察到 $p_e = p_{e|u}p_u + p_{e|r}p_r$，这恰好可以作为一个公因子提取出来。剩余项可以忽略，因为当t_2足够大时，$p_{e|r}p_r$相对于 $p_{e|u}p_u$ 很小，而且 $\binom{n-1}{t_2} p_u^{t_2} (1-p_u)^{1-t_2}$相对于 $F(t_2 - 1;\ n-1, p_u)$ 也很小。因此，我们有

$$\mathbb{E}[E] \approx \frac{np_e F(t_2 - 1; n-1, p_u)}{F(t_2; n, p_u)}.$$

将其代回式（13），得到

$$\mathbb{P}[C_1 = 0] \approx \left(1 - \frac{p_e F(t_2 - 1; n-1, p_u)}{F(t_2; n, p_u)}\right)^{M_s}.$$

将其代入方程（12），即可得到方程（5）。如果 p_e较低（即 $M_s p_e$ 较小），则可将其化简为

$$\mathbb{E}[C] = \frac{nM_s p_e F(t_2 - 1; n-1, p_u)}{F(t_2; n, p_u)}. \tag{14}$$

M 服从成功概率为 F（t_2； n， p_u）的负二项分布。因此，所需消息数的期望值为

$$\mathbb{E}[M] = \frac{M_s}{F(t_2; n, p_u)}, \tag{15}$$

将其代入方程（14），即可完成证明。 □