关于非重叠码的 Blackburn 猜想的证明

Yeow Meng Chee

arXiv:2611.28776v2 [cs.IT] 24 Nov 2026

摘要

如果一个码字的任何非空真前缀都不是某个码字的后缀，则称该码为无重叠码。令 $S_n(q)$ 表示在大小为 q 的字母表上，长度为 n 的无重叠码的最大大小。Levenshtein 证明了 $S_n(q) \le c_n q^n$，其中 $c_n = (1/n)\,((n-1)/n)^{n-1}$。当 n 整除 q 时，码 $I^{n-1}J$（其中 I 和 J 划分了字母表）达到此界；Blackburn 猜想，对于每个 n 以及所有充分大的 q，这类码中最优者都是极大无重叠码。我们证明了这一猜想。证明使用生成函数来处理描述极大无重叠码的嵌套划分。我们还证明，只有满足 $|I| = (n-1)q/n$ 的码 $I^{n-1}J$ 及其反转码达到 Levenshtein 界。

索引词

无重叠码、交叉双前后缀自由码、互不相关码、帧同步、Levenshtein 界、生成函数。

I. 引言

令 Σ_q 为一个大小为 q的字母表，并令 Σ_q^n 为由 Σ_q中的字母构成的长度为 n 的字符串集合。对于字符串集合 A 和 B，我们用 AB 表示满足 $a \in A$ 和 $b \in B$的串接 ab 所构成的集合，并用 A^k 表示由 k 个 A字符串串接而成的集合。如果 u 的某个非空真前缀是 v的后缀，则称两个字符串 u，$v \in \Sigma_q^n$ 重叠。如果任意两个码字（包括同一码字与自身）都不重叠，则称码 $C \subseteq \Sigma_q^n$ 为无重叠码。我们用 $S_n(q)$表示 Σ_q^n中无重叠码的最大大小。若 $|C| = S_n(q)$，则称无重叠码 $C \subseteq \Sigma_q^n$为最大码。如果它不被 Σ_q^n中的另一个无重叠码真包含，即对于每个 $u \in \Sigma_q^n \backslash C$，$C \cup \{u\}$都不是无重叠码，则称其为极大码。每个最大码都是极大码。

Levenshtein [15], 提出了无重叠码，并证明这类码恰好是那些存在一种无论初始状态如何都能实现同步的解码自动机的码。Gilbert [12] 研究了用于二进制消息同步的同一类码。词不与自身重叠这一性质称为双前后缀自由性，Nielsen [19], 对此进行了研究，而它也是帧同步 [18]理论的基础。在帧同步中，接收器会在数据流中搜索标记。如果使用多个标记，且这些标记构成无重叠码，那么标记就不会在横跨两个标记的位置被检测到，因此接收器可以同时搜索所有标记。这促成了分布式序列 [9] 的研究，以及对无交叉双前缀集合 [1,2,4,8], 的系统研究；在另一种名称下，如今，基于DNA的数据存储中也出现了相同的性质。在 Tabatabaei Yazdi 等人提出的随机访问架构中， [24],每个数据块都带有一个地址引物，并通过使互补引物退火来选定该数据块。引物不得重叠，以确保引物不会在偏移位置退火。这促成了互不相关码 [17,23], 的提出， [17,23], 这类码是不重叠码，并且还具有平衡性和纠错等额外约束。无重叠码也以遗传密码模型中的强无逗号码形式出现 [10]。

在所有这些应用中，码字充当可区分标记或地址，因此其数量是需要最大化的资源，而 $S_n(q)$则是基本量。

A. 既有工作

Levenshtein [16] 证明了上界

$$S_n(q) \le c_n q^n, \quad \text{where} \quad c_n = \frac{1}{n}\left(\frac{n-1}{n}\right)^{n-1}. \tag{1}$$

Chee等人 [8] 证明了较弱的界 $S_n(q) \le q^n/(2n-1)$，并针对每个固定的 q给出了渐近最优、误差仅相差一个常数因子的码构造。Bilotta、Pergola 和 Pinzani [4] 以及 Barcucci等人 [1]给出了基于格路径的构造。Blackburn [5] 给出了对所有 n 和 q而言，与 (1) 相差一个常数因子的构造。其最简单的情形如下。将$\Sigma_q = I \cup J$ 分划为若干非空集合，并满足 $|I| = s$。 $I^{n-1}J$ 的码字是长度为 n 的词，其前 $n-1$ 个符号属于 I，末符号属于 J。码字的每个真前缀中的所有符号都属于 I，而每个非空后缀都以 J中的符号结尾，因此该码是无重叠的。其大小为 $s^{n-1}(q-s)$。

Y. M. Chee 任职于新加坡科技与设计大学，新加坡（电子邮件：ymchee@sutd.edu.sg）。研究得到 SUTD Grant SKI 2021_07_04资助。

Let

$$B_n(q) = \max_{s\in\{1,2,\dots,q-1\}} s^{n-1}(q-s) \tag{2}$$

表示这种码的最大大小。我们称这些码为 $I^{n-1}J$ *Blackburn* 码。Blackburn证明，当 n 整除 q时， $S_n(q) = B_n(q) = c_n q^n$；对所有 q，有 $S_2(q) = B_2(q)$以及 $S_3(q) = B_3(q)$；并且当 $q\to\infty$时， $S_n(q)/q^n \to c_n$。他提出了以下猜想。

Blackburn 猜想([5, 猜想 1])。对于每个 $n \geq 2$ 以及所有充分大的 q，某个码 $I^{n-1}J$ 是最大无重叠码。等价地，对于所有充分大的 q， $S_n(q) = B_n(q)$。

Stanovnik、Moškon 和 Mraz [20] 基于 Fimmel 等人 [10] 提出的算法，通过对字符串进行嵌套划分，描述了所有极大无重叠码。这使得确定 $S_n(q)$ 化简为一个关于各部分大小的整数规划。他们证明了对所有 q， $S_4(q) = B_4(q)$；计算了 $q \leq 6$ 对应的 $S_n(q)$；并针对最大码的结构提出了一个猜想 [20, 猜想 1]，该猜想将推出关于 $q \geq n$ 的 Blackburn 猜想。变长码 [3,21,22]以及只禁止特定长度重叠的码 [7] 相关问题也已得到研究。对于二进制码，Blackburn [6] 最近将 Levenshtein 界 (1) 改进了 5/6 倍，据 $n\to\infty$，将其提升至 $(5/(6e) + O(1/n))2^n/n$。关于具有指定重叠的字符串的生成函数，最早可追溯至 Guibas 和 Odlyzko [14], ; 另见 Goulden 和 Jackson [13] 以及 Flajolet 和 Sedgewick [11]。

Blackburn 猜想之所以困难，原因如下。当 n 不能整除 q时，Blackburn构造的整数最优值 $B_n(q)$ 比 Levenshtein 界 (1) 低一个 q^{n-2} 量级（引理 II.1）。因此，要证明该猜想，上界必须对接近最佳 Blackburn 码的码在二阶项上精确，而 (1) 并非如此。

B. 结果

我们的主要结果如下。

定理 I.1. *Blackburn* 猜想成立：对于每个 $n \geq 2$，都存在一个 $q_0(n)$，使得 $S_n(q) = B_n(q)$ 对所有 $q \geq q_0(n)$均成立。

我们在第三节证明定理 I.1。在此过程中，我们证明，只有当一个码（将每个码字反向后）是满足 $|I| = (n-1)q/n$ 的 Blackburn 码时，它才达到 Levenshtein 界 (1) （定理 II.6 ）。Blackburn 码的大小 $s^{n-1}(q-s)$ 关于 $s \in \mathbb{R}$ 的函数，在 $s = (n-1)q/n$ 处达到最大值，此时其值为 $c_n q^n$。因此，Blackburn 猜想要问的是：最佳 Blackburn 码（其中 s 必须是整数）是否仍然至少与所有其他码一样大。不能去掉 q 必须足够大的条件，因为 Stanovnik 、 Moškon 和 Mraz [20] 对 $q \leq 6$ 的计算显示，在某些情况下有 $S_n(q) > B_n(q)$，例如 $S_5(2) = 2 > B_5(2) = 1$。

C. 证明策略

我们的出发点是通过嵌套划分对极大无重叠码的描述 [10,20]。字母表被分成左部和右部，左词与右词的串接再次进行划分，如此反复，而码由以这种方式产生的长度为 n 的词组成。如果用两个生成函数按长度统计左词和右词，整个构造就归结为对 $1 - qz$ 的一次因式分解，直到次数 n，而码的大小则成为有理函数的一个系数。因此，确定 $S_n(q)$ 等价于求一个系数的最大值，而 Blackburn 码给出了最简单的候选者。

证明随后考察第一次分割，即字母表的分划。如果其中一部分很大，就像 Blackburn 码的情形一样，我们围绕其线性部分展开有理函数。展开式中的每一项后续项都只会减小码的大小，因此该码不可能大于具有相同第一次分割的 Blackburn 码。如果两部分都不大，我们则将该码与 Levenshtein 界进行比较。生成函数给出了这一界的简短证明，其中舍弃了一个非负余项，并且只有对于最佳 Blackburn 码，该余项才会消失。因此，第二类码比 Levenshtein 界小一个固定比例，而最佳 Blackburn 码仅比它小一个低阶项；所以当 q 很大时，最佳 Blackburn 码更大。

II. 分划码及其生成函数

本节将为本文其余部分建立所需的工具。

A. 符号与术语

全文中， $n \geq 2$ 表示长度， $q \geq 2$ 表示字母表大小， c_n 如式 (1) 所定义， $B_n(q)$ 如式 (2) 所定义。对于固定的 n 和 $q\to\infty$，我们使用标准的渐近符号 O 和 Θ。对于长度为 n、定义在 Σ_q 上的码 C，我们称 $|C|/q^n$ 为 C 的归一化大小。对于 Blackburn 码 $I^{n-1}J$，我们称 $\lambda = |I|/q$ 为其比率。比率为 λ 的 Blackburn 码有 $(\lambda q)^{n-1}(q - \lambda q)$ 个码字，因此其归一化大小为 $\beta(\lambda)$，其中

$$\beta(\lambda) = \lambda^{n-1}(1-\lambda), \quad \text{for } \lambda \in \mathbb{R}.$$

因此我们有

$$B_n(q) = q^n \max_{s\in\{1,2,\ldots,q-1\}} \beta\Big(\frac{s}{q}\Big). \tag{3}$$

令 $\lambda_n = (n-1)/n$。由于 $\beta'(\lambda) = \lambda^{n-2}\big((n-1) - n\lambda\big)$，函数 β 在 $[0,\lambda_n]$ 上递增，在 $[\lambda_n, 1]$ 上递减，其最大值为 $\beta(\lambda_n) = c_n$。因此，Levenshtein 界 (1) 是 β 在 $\lambda \in [0,1]$ 上的最大值。由于 Blackburn 码的比率必须是 $1/q$ 的整数倍，除非 $\lambda_n q$ 是整数，否则最佳 Blackburn 码会小于该最大值。下一个引理表明，当 n 不整除 q 时，差值 $c_n q^n - B_n(q)$ 的阶为 q^{n-2}。

引理 II.1。设 $n \ge 2$。如果 n 整除 q，则 $B_n(q) = c_n q^n$。对于 q 不能被 n 整除的情况，我们有

$$c_n q^n - B_n(q) = \Theta(q^{n-2}). \tag{4}$$

证明。如果 n 整除 q，则 $q \ge n$，因此整数 $s = \lambda_n q = q - q/n$ 属于 {1、2、……、$q-1$}。具有 $|I| = s$ 的 Blackburn 码的比率为 λ_n，归一化大小为 $\beta(\lambda_n) = c_n$，所以 $B_n(q) \ge c_n q^n$，根据（1）等号成立。

现在设 n 不整除 q，并设 $s = \lfloor \lambda_n q \rfloor$。由于 $1 \le q/2 \le \lambda_n q < q$，我们有 $s \in$ {1、2、……、$q-1$} 和 $|s/q - \lambda_n| < 1/q$。由于在 $[0,1]$ 上 $\beta'(\lambda_n) = 0$ 且 $|\beta''| \le n(n-1)$，根据泰勒定理可得 $\beta(s/q) \ge c_n - \binom{n}{2}/q^2$，因此由（3）得到 $c_n q^n - B_n(q) \le \binom{n}{2} q^{n-2}$。

对于下界，注意到 $\beta''(\lambda_n) = -(n-1)\lambda_n^{n-3} < 0$ 且 $\beta(\lambda) < c_n$ 对 $\lambda \in [0,1]\setminus\{\lambda_n\}$ 成立。因此，$(c_n - \beta(\lambda))/(\lambda - \lambda_n)^2$ 可延拓为定义在 $[0,1]$ 上的连续正函数，且其正的最小值为 ν_n。对于每个 $s \in \{1, 2, \ldots, q-1\}$，整数 $ns - (n-1)q = n(s-q) + q$ 非零，因为 n 不能整除 q，故 $|s/q - \lambda_n| \ge 1/nq$ 且 $\beta(s/q) \le c_n - \nu_n/(nq)^2$。由 (3)，$c_n q^n - B_n(q) \ge \nu_n q^{n-2}/n^2$，于是得到 (4)。□

令 z 为不定元，并令 $\mathbb{R}[[z]]$ 为 z 上的实系数形式幂级数环。对于 $F \in \mathbb{R}[[z]]$，我们用 $[z^m]F$ 表示 F 中 z^m 的系数。若级数 F 满足 $[z^0]F \ne 0$，则它在 $\mathbb{R}[[z]]$ 中可逆，$1/F$ 表示其逆元。对于 F，$G \in \mathbb{R}[[z]]$，若 $[z^m]F = [z^m]G$ 对 $0 \le m \le n$ 成立，即 F 与 G 在所有不超过 n 的次数上都相同，我们就记作 $F \equiv G \pmod{z^{n+1}}$。这种关系在求和与乘积运算下保持不变；而且，$[z^0]F \ne 0$ 蕴含 $1/F \equiv 1/G$，因为 $1/F$ 中次数不超过 n 的系数只取决于 F 的相应系数。因此，只要读取的系数次数不超过 n，就可以像处理恒等式一样处理模 z^{n+1} 的同余关系。所有幂级数运算都在 $\mathbb{R}[[z]]$ 中进行，因此不存在收敛性问题，并且只有多项式会在实数处取值。

B. 分划码

以下构造可追溯至 Fimmel 等人 [10]，并由 Stanovnik、Moškon 和 Mraz [20] 进行了分析。令 $W_1 = \Sigma_q$。对每个 $i \ge 1$，将 W_i 分划为左部分 L_i 和右部分 R_i，并令 W_{i+1} 为由某个 L_j 中的一个词与 R_{i+1-j} 中的一个词串接而成的集合。因此

$$W_i = L_i \cup R_i, \qquad \text{for } i \ge 1, \text{ and} \tag{5}$$

$$W_i = \bigcup_{j=1}^{i-1} L_j R_{i-j}, \qquad \text{for } i \ge 2, \tag{6}$$

其中 $L_j R_{i-j} = \{uv\colon u \in L_j,\ v \in R_{i-j}\}$。(5) 中的并按定义互不相交，而 (6) 中的集合 $L_j R_{i-j}$ 两两不相交，因为 Stanovnik、Moškon 和 Mraz [20, 在定理 6] 中证明，W_i 中的每个词都有唯一分割点 j。我们称 $(L_i,\ R_i)$ 为第 i 层，对于每个 $n \ge 2$，称由这些层定义的长度为 n 的 $C = W_n$ 分划码。

命题 II.2 ([20, 定理 3 和命题 7])。每个分划码都是无重叠的，并且每个极大无重叠码都是分划码。

Stanovnik、Moškon 和 Mraz 的构造针对 $n \ge 3$。对于 $n = 2$，两个陈述都是显然的，因为 Σ_q^2 中的码是无重叠的，当且仅当没有码字的首符号是某个码字的末符号，因此对于分划 $\Sigma_q = L_1 \cup R_1$，极大码就是 $L_1 R_1$。由于每个最大码都是极大的，因此可知 $S_n(q)$ 是分划码的最大大小。将每个码字反转，会把一个分划码映射为一个左右部分互换的分划码，并且保持无重叠性质和大小不变。因此，不妨设 $|L_1| \ge |R_1|$，下面我们始终作此假设。Blackburn 码 $I^{n-1}J$ 是满足 $L_1 = I$、$R_1 = J$ 和 $L_i = \emptyset$ 的 $i \ge 2$ 的分划码。

C. 生成函数

记 $\ell_i = |L_i|$、$r_i = |R_i|$ 和 $w_i = |W_i|$。码 $C = W_n$ 仅取决于截至 n 的层，因此我们舍弃 n 以上的层，并为 $i > n$ 设定 $\ell_i = r_i = w_i = 0$。于是，(5) 和 (6) 在 n 层以上所施加的方程不再成立，而由此得到的恒等式 (7) 和 (10) 仅模 z^{n+1} 成立。此外，由于 $W_n = L_n \cup R_n$，W_n 本身的分划并不重要；我们约定 $L_n = \emptyset$，从而 $\ell_n = 0$ 和 $r_n = w_n = |C|$。我们用按长度的生成函数（即多项式）来记录左词和右词，它们是多项式

$$L(z) = \sum_{m=1}^{n-1} \ell_m z^m \quad \text{and} \quad R(z) = \sum_{m=1}^{n} r_m z^m.$$

引理 II.3。 对于任意的层序列，我们有

$$(1-L)(1-R) \equiv 1-qz \pmod{z^{n+1}}, \quad \textit{and} \tag{7}$$

$$|C| = [z^n]R. \tag{8}$$

证明。 令 $W(z) = \sum_{m=1}^{n} w_m z^m$ 为 $\bigcup_{m\le n} W_m$ 按长度计数的生成函数。分解式 (5) 和 (6) 构成了 Goulden 和 Jackson [13, 第 2 节2]意义下的递归分解。根据 (5)， $\bigcup_{m\le n} W_m$ 是不相交并集$\bigcup_{m\le n} L_m$ 和 $\bigcup_{m\le n} R_m$，因此和引理 [13, 引理 2.2 12] 给出

$$W = L + R. \tag{9}$$

根据 (6) 及其各项互不相交这一性质，连接 $(u,\ v) \mapsto uv$ 是从 $\left(\bigcup_{m\le n} L_m\right) \times \left(\bigcup_{m\le n} R_m\right)$中满足 $|u| + |v| \le n$的配对到 $\bigcup_{2\le m\le n} W_m$的双射，且会将长度相加。根据乘积引理 [13,引理 2.214],，这些配对由 LR 中次数不超过 n的项计数，因此有

$$W - qz \equiv LR \pmod{z^{n+1}}. \tag{10}$$

从 (9) 和 (10) 中消去 W，得到 $L + R - qz \equiv LR$（模 z^{n+1}），这就是 (7)；又因 $r_n = |C|$，故 (8) 成立。 □

若由非负整数系数组成的多项式 $L = \sum_{m=1}^{n-1} \ell_m z^m$ 和 $R = \sum_{m=1}^{n} r_m z^m$ 满足 $\ell_1 \ge r_1 > 0$ 且同余式 (7) 成立，则称多项式对 (L, R) 为可行对 。根据引理 II.3，每个非空分划码 C 都给出一个可行对 (L, R)，其中 $[z^n]R = |C|$；因为 $R_1 = \emptyset$ 会导致 $W_i = \emptyset$ 对 $i \ge 2$ 成立，从而 $r_1 > 0$。因此，我们称 $[z^n]R$ 为可行对 (L, R) 的大小 ，并将其记作 $|C|$，即使 (L, R) 并非来自某个码也是如此。记 $w_i = \ell_i + r_i$ 为 $1 \le i \le n$。比较 (7) 中 $1 \le i \le n$ 对应的 z^i 的系数，可知该同余式等价于

$$w_1 = q \quad \text{and} \quad w_i = \sum_{j=1}^{i-1} \ell_j r_{i-j}, \quad \text{for } 2 \le i \le n. \tag{11}$$

由于 $\deg L \le n-1$，我们有 $|C| = r_n = w_n$。对于分划码，根据 (5) 有 $w_i = |W_i|$，而 (11) 统计了 (6) 中的字符串数量。

实数松弛是由多项式 $L = \sum_{m=1}^{n-1} \ell_m z^m$ 和 $R = \sum_{m=1}^{n} r_m z^m$ 构成的有序对 (L, R) 的集合，其中系数均为非负实数，满足 $\ell_1 \ge r_1 > 0$ ，且同余式 (7) 成立。它包含每个可行对。与可行对一样，实数松弛中元素 (L, R) 的大小为 $|C| = [z^n]R$，我们称 $|C|/q^n$ 为其归一化大小。根据 (11)，对于 $2 \le i \le n$，我们有 $w_1 = q$ 和 $w_i = \sum_{j=1}^{i-1} \ell_j (w_{i-j} - \ell_{i-j})$，因此通过对 i 进行归纳，每个 w_i 都是 q、ℓ_1、.....、ℓ_{i-1} 的一个具有整数系数的多项式。因此， L 确定了 w_1、.....、w_n，从而也确定了 R，因为 $r_i = w_i - \ell_i$。所以，实数松弛中的元素 (L, R) 由 L 确定，我们将其与 L 视为同一对象。于是，对于固定的 q，我们可以将实数松弛取为满足多项式不等式 $0 \le \ell_i \le w_i$（其中 $1 \le i \le n-1$ 和 $q/2 \le \ell_1 < q$）的、具有实系数的多项式 $L = \sum_{m=1}^{n-1} \ell_m z^m$ 的集合；该集合为 $\ell_1 \ge r_1 > 0$，而 $|C| = w_n$ 是其上的一个多项式函数。将 ℓ_i 和 r_i 按 q^{-i} 进行缩放，会把 q 的实数松弛映射到 $q = 1$ 的实数松弛，并将 $|C|$ 除以 q^n，因为当 ℓ_i 和 r_i 被赋予次数 i 时，(11) 和 $|C|$ 都是齐次的。我们称实数松弛中元素 L 的 $\lambda = \ell_1/q$ 为其比率。对于 Blackburn 码，这与其比率一致，因为 $\ell_1 = |I|$。由于 $\ell_1 \ge r_1 > 0$ 和 $\ell_1 + r_1 = q$，我们有 $1/2 \le \lambda < 1$，且缩放保持比率不变。我们称具有比率 λ 和 $\ell_i = 0$（对于 $2 \le i \le n-1$）的元素为具有比率 λ 的 Blackburn 码的实数类比。其归一化大小为 $\beta(\lambda)$。

同余式 (7) 仅通过 L 表示 $|C|$ 。

引理 II.4。 对于实数松弛中的每个元素 L， 我们有

$$|C| = -[z^n]\,\frac{1-qz}{1-L(z)}. \tag{12}$$

证明。 由于 $1-L$ 的常数项为 1，因此它是可逆的，由 (7) 得到 $1-R \equiv (1-qz)/(1-L)$（模 z^{n+1}）。取 z^n 的系数得到 $-r_n = [z^n](1-qz)/(1-L)$，以及 $r_n = |C|$。 □

根据命题 II.2 和引理 II.4， $S_n(q)$ 是来自划分码的多项式 L 中系数 (12) 的最大值，因此确定 $S_n(q)$ 等价于求一个有理函数的某个系数的最大值。

D. Levenshtein 界及其等号情形

现在我们证明 Levenshtein 界 (1)，并确定其何时取等号。为证明该界，我们在一个实数—— $1 - L$ 的一个根——处进行求值，这需要用一个精确恒等式代替同余式 (7)。由于 L 和 R 是多项式，因此 $(1 - L)(1 - R)$ 也是多项式，而下一个引理计算其次数高于 n 的各项。

引理 II.5。 对于实数松弛中的每个元素 L， 我们有

$$(1-L(z))(1-R(z)) = 1 - qz + |C|z^n L(z) + E(z), \tag{13}$$

其中

$$E(z) = \sum_{\substack{1\le i,j\le n-1\\ i+j>n}} \ell_i r_j z^{i+j}$$

具有非负系数。

证明。由 (11)，$[z^1](L+R) = q$且 $[z^i](L+R) = w_i = [z^i]LR$，对于 $2 \le i \le n$。由于 $\deg L \le n-1$，LR中满足 $i+j > n$的项 $\ell_i r_j z^{i+j}$具有 $i \le n-1$。由于 $r_n = |C|$，满足 $j = n$的项之和为 $|C|z^n L$，而满足 $j \le n-1$的项构成 E。因此 $LR = L + R - qz + |C|z^n L + E$，两边加上 $1 - L - R$后即得 (13)。□

定理 II.6。实数松弛中的每个元素 满足 $|C| \le c_n q^n$。当且仅当 $L(z) = \lambda_n qz$时，等号成立 .

证明。由于 $L(0) = 0$且 $L(x) \ge \ell_1 x$，对于 $x \in [0,\infty)$，多项式 $1 - L$有一个根 $\rho \in (0, 1/\ell_1]$。在 $z = \rho$处，(13) 左侧为零，且 $L(\rho) = 1$；因此，由于 E 具有非负系数，且 $\rho > 0$，我们有

$$|C|\rho^n = q\rho - 1 - E(\rho) \le q\rho - 1. \tag{14}$$

两边除以 ρ^n，得到 $|C| \le (q\rho - 1)\rho^{-n} = q^n \beta(1/q\rho) \le c_n q^n$，因为 $\beta \le c_n$在（0, 1]上且 $\beta \le 0$在 $[1,\infty$上）。

设等号成立。则（14）中等号成立，因此 $E(\rho) = 0$。由于 $\rho > 0$ 且 E 的每个系数都非负，E的每个系数都为零，所以当 $i + j > n$时，$\ell_i r_j = 0$。现在，$w_1 = q > 0$、$w_2 = \ell_1 r_1 > 0$，并且对于 $3 \le i \le n-1$，（11）中 w_i 的求和包含互不相同的项 $\ell_1 r_{i-1}$和 $\ell_{i-1} r_1$，因此由于 $\ell_1 \ge r_1$，我们有

$$w_i \ge \ell_1 r_{i-1} + \ell_{i-1} r_1 \ge r_1 w_{i-1} > 0.$$

由于 $w_{n-1} = \ell_{n-1} + r_{n-1} > 0$，要么 $\ell_{n-1} > 0$，此时由 $\ell_{n-1} r_j = 0$可得 $r_j = 0$对 $2 \le j \le n-1$成立；要么 $r_{n-1} > 0$，此时由 $\ell_i r_{n-1} = 0$可得 $\ell_i = 0$对 $2 \le i \le n-1$成立。因此，$L = \ell_1 z$，或者对 $2 \le i \le n-1$有 $r_i = 0$。在第一种情况下，$|C| = \ell_1^{n-1}(q - \ell_1) = q^n \beta(\ell_1/q)$，这仅当 $\ell_1 = \lambda_n q$时才等于 $c_n q^n$。在第二种情况下，如果 $L \ne \ell_1 z$，则 $n \ge 3$，同理，$|C| = r_1^{n-1}(q - r_1) = q^n \beta(r_1/q)$，这仅当 $r_1 = \lambda_n q > q/2$时才等于 $c_n q^n$。这与 $\ell_1 \ge r_1$矛盾。反之，$L(z) = \lambda_n qz$ 是比率为 λ_n的 Blackburn 码的实数类比，因此 $|C| = q^n \beta(\lambda_n) = c_n q^n$。□

该证明通过两个不等式——式 (14) 中的不等式以及 $\beta(1/q\rho) \le c_n$——得到 Levenshtein 界。第一个不等式当且仅当 $E(\rho) = 0$ 时取等号，第二个不等式当且仅当 $1/q\rho = \lambda_n$，即 $\rho = n/(n-1)q$ 时取等号。当 n整除 q 时，具有 $|I| = \lambda_n q$的 Blackburn 码 $I^{n-1}J$的归一化大小为 $\beta(\lambda_n) = c_n$，因此它满足 (1)，并且对它而言两个不等式都取等号。事实上，此时 $L(z) = \lambda_n qz$，所以 $\rho = 1/\lambda_n q = n/(n-1)q$。定理 II.6 表明，除该码外，没有其他码能使两个不等式同时取等号，实数松弛中也没有其他元素能使二者同时取等号。用码的术语来说，$L(z) = \ell_1 z$ 意味着对于 $2 \le i \le n-1$有 $L_i = \emptyset$。于是 (6) 给出 $W_i = L_1 R_{i-1} = L_1 W_{i-1}$对于 $3 \le i \le n$，因此 $C = W_n = L_1^{n-1} R_1$ 就是具有 $I = L_1$ 和 $J = R_1$ 的码 $I^{n-1}J$。达到 Levenshtein 界的码是最大的，因而根据命题 II.2，它是分划码。因此，定理 II.6 表明，一个码达到 Levenshtein 界的必要条件是：除将每个码字反转之外，它必须是具有 $|I| = (n-1)q/n$的 Blackburn 码。特别地，这要求 $n \mid q$。

Levenshtein 自己的证明 [16], 另见 Blackburn [5,定理 2], 对不包含任何码字作为因子的字符串进行计数，其生成函数是 $1 - qz + |C|z^n$ 的倒数，并由此推出该多项式有一个正实根。我们的证明则在 $1 - L$ 的根 ρ 处求值，而它舍弃的项 $E(\rho)$ 正是刻画等号成立的关键。

推论 II.7. $|C|$ 在实数松弛上的最大值为

$$c_n q^n = q^n \max_{0\le\lambda\le 1} \beta(\lambda),$$

且仅当取 *Blackburn* 码的实数类比时才能达到，其比率为 λ_n。

证明。由定理 II.6，$|C| \le c_n q^n$，且等号仅在所述元素处成立，而该元素的 $|C| = q^n \beta(\lambda_n)$ 为 $= c_n q^n$。□

III. Blackburn 猜想的证明

本节证明定理 I.1。我们假设 $n \ge 3$，因此在下文中等式的中等码定义里，$(n-2)/(n-1) \ge 1/2$。

定义 III.1。实数松弛中的一个元素，或给出该元素的码，如果其比率 λ满足 $\lambda \ge (n-2)/(n-1)$，则称为偏斜；否则称为中等，也就是说，如果 $1/2 \le \lambda < (n-2)/(n-1)$。

A. 偏斜码

我们首先证明，偏斜码不大于比率相同的 Blackburn 码 λ，后者有 $q^n \beta(\lambda)$ 个码字。

定理 III.2。实数松弛中比率为 λ 的每个偏斜元素 L 都满足 $|C| \le q^n \beta$（λ）。特别地，每个偏斜码至多有 B_n（q） 个码字。

证明。 写作 $L(z) = \ell_1 z + H(z)$，其中 H 具有非负系数，且不含次数低于 2 的项。由于 $H(z)/(1-\ell_1 z)$ 不含次数低于 2 的项，其 j 次幂不含次数低于 $2j$ 的项。因此，几何级数 $\sum_{j\ge 0}(H(z)/(1-\ell_1 z))^j$ 的每个系数都是有限和，并且该级数是 $1 - H(z)/(1-\ell_1 z)$ 在 $\mathbb{R}[[z]]$ 中的逆元。因此我们有

$$\begin{aligned}\frac{1-qz}{1-L(z)} &= \frac{1-qz}{(1-\ell_1 z)\Big(1-\dfrac{H(z)}{1-\ell_1 z}\Big)}\\ &= \sum_{j\ge 0} H(z)^j \,\frac{1-qz}{(1-\ell_1 z)^{j+1}}.\end{aligned}$$

该项 $j=0$ 贡献了 $[z^n](1-qz)/(1-\ell_1 z) = \ell_1^n - q\ell_1^{n-1} = -\ell_1^{n-1}(q-\ell_1)$。对于 $j\ge 1$ 和 $m\ge 1$ ，

$$\begin{aligned}[z^m]\frac{1-qz}{(1-\ell_1 z)^{j+1}} &= \binom{m+j}{j}\ell_1^m - q\binom{m+j-1}{j}\ell_1^{m-1}\\ &= \binom{m+j-1}{j}\ell_1^{m-1}\Big(\frac{m+j}{m}\,\ell_1 - q\Big),\end{aligned}$$

且 $m=0$ 的系数为 1。由于 H^j 没有次数低于 $2j$的项， j的第 $[z^n]$ 项的系数是 H^j的系数与 $(1-qz)(1-\ell_1 z)^{-j-1}$ 在次数 $m \le n-2j \le n-2$处的系数的乘积之和。对于这些 m和$j\ge 1$，我们有 $m/(m+j) \le (n-2)/(n-1)$，因此每个这样的系数都是非负的，因为 $\ell_1/q = \lambda \ge (n-2)/(n-1)$。由于 H^j也具有非负系数，每一项 $j\ge 1$都会向 $[z^n](1-qz)/(1-L)$增加一个非负量。因此我们有$[z^n](1-qz)/(1-L) \ge -\ell_1^{n-1}(q-\ell_1)$。由引理 II.4， $|C| = -[z^n](1-qz)/(1-L)$，从而得到 $|C| \le \ell_1^{n-1}(q-\ell_1)$ $= q^n\beta(\lambda)$。对于一个可行对， ℓ_1和 $r_1 = q-\ell_1$是正整数，因此 $\ell_1 \in \{1,2,\dots,q-1\}$，且 $\ell_1^{n-1}(q-\ell_1) \le B_n(q)$。□

B. 中等码

通过紧致性论证，我们给出中等码的界。首先利用 II-C 节的缩放消去对 q 的依赖；该缩放将 q 的实数松弛映射为 $q=1$的实数松弛，将 $|C|$ 除以 q^n ，并保持比率不变。对于 $q=1$，比率为 ℓ_1。令 M_n 为实数松弛中满足 $q=1$ 且比率至多为$(n-2)/(n-1)$的元素集合。它包含 $q=1$对应的适度元素。该集合非空，因为其中包含 Blackburn 码的实数类比，其比率为$(n-2)/(n-1)$，且具有 $\ell_1 = (n-2)/(n-1)$和 $r_1 = 1/(n-1)$。

引理 III.3。 存在一个仅依赖于 n的常数 $\mu_n < c_n$，使得每个适度可行对都满足 $|C| \le \mu_n q^n$。

证明。 我们首先证明 M_n 是紧致的。由于 $\ell_1 \le (n-2)/(n-1) < 1$，条件 $r_1 = 1-\ell_1 > 0$ 在 M_n上自动成立，因此 M_n 由关于 L的系数的多项式不等式 $0 \le \ell_i \le w_i$ 以及$1/2 \le \ell_1 \le (n-2)/(n-1)$定义，且它是闭集。它是有界的，因为 $0 \le \ell_i,\ r_i \le w_i$，并且由于 $w_1 = \ell_1 + r_1 = 1$，我们有 $w_i \le \sum_{j=1}^{i-1} w_j w_{i-j}$，因此通过归纳法， w_i至多为卡特兰数 $\binom{2i-2}{i-1}/i$。因此 M_n是紧致的。

紧致性意味着多项式 w_n 在 M_n上取得其最大值 μ_n ，即，对于 M_n中的某个元素，有 $\mu_n = w_n$。根据定理 II.6，对于 M_n中的每个元素都有 $w_n < c_n$ ，因为 c_n 要求比率 $\lambda_n = (n-1)/n$，而该比率超过 $(n-2)/(n-1)$。因此， $\mu_n < c_n$。将一个适度元素进行缩放以得到 q ，即可得到 M_n中的一个元素，因此每个适度可行对都有 $|C| \le \mu_n q^n$。□

C. 定理 *I.1* 的证明

现在我们将定理 III.2 与引理 III.3 结合起来。剩下的工作是将 $\mu_n q^n$ 与 $B_n(q)$ 进行比较，后者是最佳 Blackburn 码的大小 (2)。

定理 *I.1* 的证明。 令 $n\ge 3$，并令 C为 Σ_q^n中的一个最大无重叠码。那么 C是极大的，因此根据命题 II.2，它是一个分划码；必要时将每个码字反转后，我们有 $|L_1| \ge |R_1|$。由于某个 Blackburn 码具有$B_n(q)$个码字，我们有 $|C| = S_n(q) \ge B_n(q)$。

由引理 II.1， $B_n(q)/q^n \to c_n$趋于 $q\to\infty$。由于根据引理 III.3， $\mu_n < c_n$，对于所有充分大的 q，我们有 $B_n(q) > \mu_n q^n$。对于这样的 q，若 C是适中的，则引理 III.3 会给出 $|C| \le \mu_n q^n < B_n(q)$，矛盾。因此， C是偏斜的，而定理 III.2 给出 $|C| \le B_n(q)$。所以， $S_n(q) = B_n(q)$，并且满足 $|I| = s$的 Blackburn 码 $I^{n-1}J$（其中 $s\in\{1,\ 2,\ \dots\dots,\ q-1\}$ 使$s^{n-1}(q-s)$最大，是一个最大码。□

IV. 结论

我们证明了 Blackburn 猜想对每个长度都成立：对于每个 n，当 q 足够大时，某个 Blackburn 码 $I^{n-1}J$ 就是最大无重叠码。利用生成函数对划分码进行计数，可将确定 $S_n(q)$的问题转化为求有理函数某个系数最大值的问题。当第一层偏斜时，Blackburn构造是最优的；Levenshtein 界的等号情形也表明，Levenshtein 界就是实数松弛的值，并由此将中等码与该界区分开来。该证明没有给出明确的阈值。还有两个相关的公开问题：对于每个 n，确定定理 I.1 中最小的 $q_0(n)$；以及确定所有满足 $S_n(q) > B_n(q)$的数对 (q,n)。

参考文献

[1] E. Barcucci、S. Bilotta、E. Pergola、R. Pinzani 和 J. Succi，“通过 Motzkin路径生成无交叉双前缀集合”，RAIRO Theor. Inform. Appl.，第 50 卷，第 1 期，第 81–91 页，2016 年。[2] D. Bajić 和 T. Lončar-Turukalo，“一种简单的次优交叉双前后缀自由码构造”，Cryptogr. Commun.，第 6 卷，第 1 期，第 27–37 页，2014 年。[3] S. Bilotta，“变长无重叠码”，IEEE Trans. Inf. Theory，第 63 卷，第 10 期，第 6530–6537 页，2017 年。[4] S. Bilotta、E. Pergola 和 R. Pinzani，“无交叉双前缀集合的新方法”，IEEE Trans. Inf. Theory，第 58 卷，第 6 期，第 4058–4063 页，2012 年。[5] S. R. Blackburn，“非重叠码”，IEEE Trans. Inf. Theory，第 61 卷，第 9 期，第 4890–4894 页，2015 年。[6] S. R. Blackburn，“非重叠码的紧上界”，arXiv:2611.21913，2026 年。[7] S. R. Blackburn、N. N. Esfahani、D. L. Kreher 和 D. R. Stinson，“受限重叠码的构造与界”，IEEETrans. Inf. Theory，第 70 卷，第 4 期，第 2479–2490 页，2024 年。[8] Y. M. Chee、H. M. Kiah、P. Purkayastha 和 C. Wang，“ 渐近达到最优的常数因子内的交叉双前后缀自由码 ”，IEEE Trans. Inf. Theory，第 59 卷，第 7 期，第 4668–4674 页，2013 年。[9] A. J. de Lind van Wijngaarden 和 T. J. Willink，“利用分布式序列进行帧同步”，IEEE Trans. Commun.，第 48 卷，第 12 期，第 2127–2138 页，2000 年。[10] E. Fimmel、C. J. Michel、F. Pirot、J.-S. Sereni 和 L. Strüngmann，“有限字母表上的双字母和三字母无逗号码”，Australas. J. Combin.，第 86 卷，第 2 期，第 233–270 页，2023 年。[11] P. Flajolet 和 R. Sedgewick，《解析组合学》。剑桥：剑桥大学出版社，2009 年。[12] E. N. Gilbert，“二进制消息同步”，IRE Trans. Inf. Theory，第 6 卷，第 4 期，第 470–477 页，1960 年。[13] I. P. Goulden 和 D. M. Jackson，《组合计数》。纽约：Wiley，1983 年。Dover 于 2004 年再版。[14] L. J. Guibas 和 A. M. Odlyzko，“字符串重叠、模式匹配与非传递博弈”，J. Combin. Theory Ser. A，第 30 卷，第 2 期，第 183–208 页，1981 年。[15] V. I. Levenshtein，“对初始状态不变的解码自动机”，Problemy Kibernet.，第 12 卷，第 125–136 页，1964 年（俄文）。[16] V. I. Levenshtein，“关于无重叠码中字符串的最大数量”，Problemy Peredachi Informatsii，第 6 卷，第 4 期，第 88–90 页，1970 年（俄文）。[17] M. Levy 和 E. Yaakobi，“用于DNA存储的互不相关码”，IEEE Trans. Inf. Theory，第 65 卷，第 6 期，第 3671–3691 页，2019 年。[18] J. L. Massey，“最优帧同步”，IEEE Trans. Commun.，第 20 卷，第 2 期，第 115–119 页，1972 年。[19] P. T. Nielsen，“关于无双前缀序列的一则注记”，IEEE Trans. Inf. Theory，第 19 卷，第 5 期，第 704–706 页，1973 年。[20] L. Stanovnik、M. Moškon 和 M. Mraz，“寻找最大无重叠码”，Des. Codes Cryptogr.，第 92 卷，第 5 期，第 1299–1326 页，2024 年。[21] G. Wang 和 Q. Wang，“*Q*元无重叠码：一种生成函数方法”，IEEE Trans. Inf. Theory，第 68 卷，第 8 期，第 5154–5164 页，2022 年。[22] G. Wang 和 Q. Wang，“关于变长无重叠码的最大规模”，Des. Codes Cryptogr.，第 93 卷，第 4 期，第 871–878 页，2025 年。[23] S. M. H. Tabatabaei Yazdi、H. M. Kiah、R. Gabrys 和 O. Milenkovic，“用于基于DNA的数据存储的互不相关引物”，IEEE Trans. Inf. Theory，第 64 卷，第 9 期，第 6283–6296 页，2018 年。[24] S. M. H. Tabatabaei Yazdi、Y. Yuan、J. Ma、H. Zhao 和 O. Milenkovic，“一种可重写、随机访问的基于DNA的存储系统”，Sci. Rep.，第 5 卷，文章编号 14138，2015 年。