arXiv:2303.09898v1 [math.CO] 17 Mar 2023

基于交换和失配距离的超立方体与等距词 *

M. Anselmo[1] G. Castiglione[2] M. Flores[1,2]D. Giammarresi[3] M. Madonia[4] S. Mantaci[2]

[1] 萨莱诺大学计算机科学系，意大利萨莱诺。{manselmo, mflores }@unisa.it[2] 巴勒莫大学数学与计算机科学系，意大利巴勒莫{giuseppa.castiglione, sabrina.mantaci }@unipa.it[3] 罗马第二大学数学系，意大利罗马。giammarr@mat.uniroma2.it[4]卡塔尼亚大学数学与计算机科学系，意大利卡塔尼亚。madonia@dmi.unict.it

摘要

维数为 n 的超立方体是一个图，其顶点为长度 n的 2^n 二进制词；当两个顶点之间的汉明距离为 1 时，它们之间有一条边。我们考虑一种基于交换和失配的编辑距离，称为波浪距离，并定义波浪号超立方体，其边连接波浪距离为 1 的词。随后，我们引入并研究波浪号超立方体的一些等距子图，这些子图通过使用称为波浪号等距词的特殊词构造而成。子图仅保留不含给定波浪号等距词作为因子的顶点。对于词 11，这种子图称为波浪号斐波那契立方体，是经典斐波那契立方体的推广。波浪号超立方体和波浪号斐波那契立方体都可以递归定义；它们的边数也具有同样的性质。这使得我们可以渐近估计波浪号斐波那契立方体中的边数，并将其与波浪号超立方体中的总边数进行比较。

关键词： 交换与失配距离、等距词、超立方体。

1 引言

n维超立方体 Q_n 包含所有长度为 n 的二进制字符串，因此它是图论中值得占据重要地位的模型。它被定义为这样一种图：其顶点与长度为 n 的 2^n 字符串一一对应，且当两个顶点所对应的字符串恰有一个位置不同时，这两个顶点之间就有一条边；换言之，它们的汉明距离 为 1。因此，图中两个顶点之间的距离等于对应字符串的汉明距离。多年来，超立方体的概念得到了广泛研究（综述见 [10] ）。超立方体

*本研究部分得到了 INdAM-GNCS 2022 和 2023 项目、萨莱诺大学 FARB 项目 ORSA229894、卡塔尼亚大学 TEAMS 项目以及授予罗马第二大学数学系的 MIUR 卓越系项目 MatMod@TOV 的资助。

被用于设计互连网络，也应用于理论化学（综述见 [14] ）。然而，超立方体有一个显著的局限：其顶点数量呈指数级增长。为此，人们提出了各种改进方案，考虑等距子图；也就是说，这类子图中任意一对顶点之间的距离，都与完整超立方体中的距离相同。为此，Hsu 于 1993年 引入了斐波那契立方体 [11]。它们是从 Q_n 中选取对应字符串不含因子 11 的顶点而得到的等距子图。它们还具有许多与斐波那契数相关的显著性质。

广义斐波那契立方体 $Q_n(f)$ 于 2012年被引入，它是Q_n 的子图，仅保留与不含 f 作为因子的二进制词相关联的顶点，也就是说，保留 f-自由二进制词 [12]。请注意，若要得到 Q_n的等距子图，被避开的词必须满足某些特殊条件；若满足这些条件，则称该词为等距词。具体而言，二进制词 f 是等距（或汉明等距）的，当且仅当对于任意 $n \geq 1$， $Q_n(f)$都能等距嵌入 Q_n；否则非等距 [15]。文献 [13, 15, 18, 21, 22] 已刻画二进制汉明等距词的结构，而这一主题的研究至今仍十分活跃 [7, 19, 20]。

近年来，研究者开始在二维情形下考察二进制汉明等距词，并研究汉明非等距图（也称为坏图） [6]。此外，人们将等距词的概念扩展到大小为 k的字母表情形，其中 $k > 2$，方法是考察 k元 n立方体 Q_n^k，以及避开词 f的 k元 n 立方体 $Q_n^k(f)$。在这种情形下，两个顶点之间的距离不再是它们的汉明距离，而是李距离。考虑这种距离后，人们引入、研究并刻画了李等距 k元词[3, 4, 5]。文献 [4, 8],利用汉明等距词和李等距词的刻画，给出了一些线性时间算法，用于判断一个词是否等距，并提供有关非等距词的一些有趣信息。值得一提的是，汉明等距词和李等距词可以不借助超立方体来定义和研究，而采用一种更接近词组合学的视角。事实上，若对于任意一对长度相同的 f-自由词 u 和 v， f 均可通过一列 f-自由词变换为 v ，且该序列以 u 开头、以 v结尾，其长度等于 u 与 v 之间的汉明距离（李距离），并且序列中每两个相邻词的汉明距离（李距离）均为 1，则称词 f 是汉明等距（分别为李等距）的。

在计算生物学中的某些应用中，考虑词中交换两个相邻且不同符号的交换操作似乎很自然。因此，基于交换和不匹配错误的编辑距离似乎值得探讨 [1, 9]。在[2] 这种距离被称为波浪距离，因为 $\sim$ 符号在某种程度上让人联想到交换操作。人们使用波浪距离而非汉明距离或李距离定义了波浪等距词，并从组合的角度对其进行了研究。

本文以波浪距离为基础，定义了波浪号超立方体,Q_n；其顶点仍是所有 n-二进制字符串，但边对应于波浪-˜距离为 1 的情形。这意味着 Q_n 的边数多于 Q_n；特别是，由于一次交换对应两处失配，在 Q_n中距离为 2 的某些顶点，在 Q_n中变为相邻顶点。我们给出波浪号超立方体的递归构造，并枚举˜其边数。接着，对于给定的词 f，我们通过选取对应于 f-自由词的顶点，考察波浪号超立方体 Q_n（f）的子图。容易证明， f 是波浪号等距的，当且仅当 Q_n（f）是 Q_n的等距子图。我们给出一个由波浪等距词组成的无限族，其中的词并非汉明等距。论文最后一部分致力于选取特殊词 f。至于词

~$f = 11$，也就是说，同时满足汉明等距和波浪号等距的词，子图 Q_n(11)称为波浪号斐波那契立方体。我们给出其递归构造，并将其与经典斐波那契立方体进行比较。我们证明，波浪号斐波那契立方体的边数约比整个波浪号超立方体的边数少七分之一。我们还考察 Q_n(1010)，其中 1010 是一个波浪号非等距但汉明等距的~词；以及 Q_n(11100)，其中 11100 是一个汉明非等距且波浪号等距的词。论文最后以一张小表格作结，比较阶为 n = 4, . . . , 16 时 Q_n(1010)和 Q_n(11100)的顶点数、边数、基数及比率~。

2 预备知识

本文仅关注二元字母表 $\Sigma = \{0, 1\}$。长度为 $|w| = n$ 的词（或字符串） w 是 $w = a_1a_2 \cdots a_n$，其中 a_1，a_2，.....，a_n 均为 Σ中的符号。由 Σ 构成的所有词的集合记作 Σ^*。最后，ϵ 表示空词，而 $\Sigma^+ = \Sigma^* - \{\epsilon\}$。对于任意词 $w = a_1a_2 \cdots a_n$，w 的反转是词 $w^{rev} = a_na_{n-1} \cdots a_1$。若$x \in \Sigma$，$x$ 表示 x的相反符号，即 $x = 1$ 当 $x = 0$ 时，反之亦然。于是，我们定义 w 的取补词 $w = a_1a_2 \cdots a_n$。

令 $w[i]$ 表示 w 在位置 i处的符号，即 $w[i] = a_i$。那么，对于 $1 \le i \le j \le n$，$w[i..j] = a_i \dots a_j$表示 w的一个因子。长度为 l 的 w 的前缀（分别为后缀），其中 $1 \le l \le n - 1$，记作 $\text{pre}_l(w) = w[1..l]$（分别记作 $\text{suf}_l(w) = w[n - l + 1..n]$）。当 $\text{pre}_l(w) = \text{suf}_l(w) = u$ 时，u 在此称为 w 的一个长度为 l 的重叠；在其他框架中，它也称为边界或双前后缀。若词 w 不含以 w 为因子的 f，则称其为-自由。

一个编辑操作是一个函数 O：$\Sigma^* \to \Sigma^*$，它将一个词转换为另一个词。

设 OP 为编辑操作集。词 u，$v \in \Sigma^*$ 的编辑距离是指在 OP中将 u转换为 v所需的最少编辑操作数。

本文考虑仅使用交换和替换操作来修正交换和不匹配错误的编辑距离。请注意，这些操作会保持词的长度不变。

定义 1设 $w = a_1a_2 \dots a_n$ 是由 Σ构成的一个词。在位置 i对 w 执行的替换操作（简称替换）定义为

$$R_i(a_1a_2 \dots a_{i-1}\boldsymbol{a_i}a_{i+1} \dots a_n) = a_1a_2 \dots a_{i-1}\overline{\boldsymbol{a}}_{\boldsymbol{i}}a_{i+1} \dots a_n.$$

在位置 i对 w 执行的交换操作（简称交换

其中 $a_i \ne a_{i+1}$，定义为

$$S_i(a_1a_2 \dots a_{i-1}\boldsymbol{a_ia_{i+1}}a_{i+2} \dots a_n) = a_1a_2 \dots a_{i-1}\boldsymbol{a_{i+1}a_i}a_{i+2} \dots a_n.$$

请注意，一次交换相当于替换两个连续符号。

汉明距离 $\text{dist}_H(u, v)$ 定义为将 u，$v \in \Sigma^*$ 转换为 v 所需的最少替换次数。若对任意一对f-free 字符串 u和 v，都存在一个长度为 $\text{dist}_H(u, v)$ 的替换序列，可将 u 转换为 v，且所有中间字符串也都是 f-free，则称词 f 是汉明等距 的。

若存在 l，使得 $\text{pre}_l(w)$ 与 $\text{suf}_l(w)$ 的汉明距离为 2（参见 [18]），则词 w 具有 2-错误重叠。接着证明了汉明等距词的以下刻画。

命题2 ([18]) 词 f 是汉明等距的，当且仅当 f 具有一个 *2*-误差重叠。

设 G 为一个图， $V(G)$ 为其节点集合， $E(G)$ 为其边集合。 $u, v \in V(G)$ 之间的距离 $\text{dist}_G(u, v)$，是连接 u 和 v 的最短路径在 G 中的长度。 G 的直径 记作 $d(G)$，是 G 中两个顶点之间距离的最大值。若（连通）图 G 的子图 S 满足对任意 $u, v \in V(S)$，均有 $\text{dist}_S(u, v) = \text{dist}_G(u, v)$，则称其为等距子图 。

我们回顾一下超立方体和斐波那契立方体的概念，它们与汉明距离有关。 n-超立方体，或二进制 n-立方体， Q_n，是一个图，含有 2^n 个顶点，每个顶点对应一个长度为 n的二进制字。顶点通常用其对应的字来表示。若 u 和 v 在 Q_n 中的两个顶点所对应的字恰好有 1 个位置不同，则它们相邻，即 $\text{dist}_H(u, v) = 1$。因此，$\text{dist}_{Q_n}(u, v) = \text{dist}_H(u, v)$。

用 f_n 表示第 n个斐波那契数，其定义为 $f_1 = 1, f_2 = 1$ 且 $f_i = f_{i-1} + f_{i-2}$，其中 $i \geq 3$。阶为 n 的斐波那契立方体 F_n 是 Q_n 的子图，其顶点为不含因子 11 的长度为 n 的二进制词。众所周知， F_n是 Q_n 的等距子图（参见 [14]）。超立方体的等距子图也称为偏立方体。

Q_n 和 F_n 的主要性质之一是其递归结构，对此已有大量研究（参见 [11], [16] 和 [14]）。

以下结　　果广为人知，但在此列出以供日后　　参考。

命题 3 设 Q_n 为 n 阶超立方体， F_n 为斐波那契立方体。那么

- $|V(Q_n)| = 2^n$ 和 $|E(Q_n)| = n2^{n-1}$
- $|V(F_n)| = f_{n+2}$
- $|E(F_1)| = 1$， $|E(F_2)| = 2$ 和 $|E(F_n)| = |E(F_{n-1})| + |E(F_{n-2})| + f_n$， $\forall n > 2$

$$|E(F_n)| = \frac{2(n+1)f_n + nf_{n+1}}{5}$$

数列 $|E(F_n)|$ 是 [17]中的数列 A001629。因此，具有 N 个顶点的斐波那契立方体的边数为 $O(N \log\ N)$，渐近地等于具有相同顶点数的超立方体的边数。

3 波浪号等距词

本节讨论基于交换和替换操作的编辑距离，这些操作用于修正两个词之间的交换错误和不匹配错误。该距离称为波浪距离，记作 $dist_{\sim}$。我们回顾 [2] 中给出的波浪号等距词定义，然后介绍一类波浪号等距词。

定义 4 设 u、 $v \in \Sigma^*$ 是长度相同的字符串。 u 与 v 之间的波浪距离 $\text{dist}_{\sim}(u, v)$是将 u 转换为 v 所需的最少替换和交换次数。

定义 5 设 u、 $v \in \Sigma^*$ 是长度相同的字符串。
从 u 到 v、长度为 h 的波浪变换 τ 是一个字符串序列 $(w_0, w_1, \ldots, w_h)$，满足 $w_0 = u, w_h = v$，并且对于任意 $k = 0, 1, \ldots, h-1$，距离$_\sim(w_k, w_{k+1}) = 1$。此外，如果对于任意 $i = 0, 1, \ldots, h$，词 w_i 都是 f-自由的，则称 τ 是 f-自由的。如果其长度等于距离$_\sim(u, v)$，且每个位置上的字符至多修改一次，则称其为最小的。

从 u 到 v 的波浪变换 $(w_0, w_1, \ldots, w_h)$ 与由 h 个操作 $(O_{i_1}, O_{i_2}, \ldots O_{i_h})$ 组成的序列相关联，该序列满足对于任意 $k = 1, \ldots, h$, $O_{i_k} \in \{R_{i_k}, S_{i_k}\}$ 和 $w_k = O_{i_k}(w_{k-1})$；它可以表示如下：

$$u = w_0 \xrightarrow{O_{i_1}} w_1 \xrightarrow{O_{i_2}} \cdots \xrightarrow{O_{i_h}} w_h = v.$$

稍微滥用一下记号，接下来我们将把波浪变换既称为字符串序列，也称为操作序列。下面举几个例子。

例 6 设 $u = 1011, v = 0110$。下面展示了从 u 到 v 的两种不同波浪变换。注意，τ_1 的长度对应于 $\mathrm{dist}_\sim(u, v) = 2$。

$$\tau_1 : 1011 \xrightarrow{S_1} 0111 \xrightarrow{R_4} 0110 \qquad \tau_2 : 1011 \xrightarrow{R_1} 0011 \xrightarrow{R_2} 0111 \xrightarrow{R_4} 0110$$

此外，考虑将 $u' = 100$ 变换为 $v' = 001$ 的以下波浪变换：

$$\tau_1' : 100 \xrightarrow{S_1} 010 \xrightarrow{S_2} 001 \qquad \tau_2' : 100 \xrightarrow{R_1} 000 \xrightarrow{R_3} 001$$

注意，τ_1' 和 τ_2' 的长度相同，均等于 $\mathrm{dist}_\sim(u', v') = 2$，并且在 τ_1' 中，第 *2* 个位置上的符号被更改了两次。

在 [2] 中已证明，在二元情形下，最小波浪变换总是存在的。现在，我们基于波浪距离来定义等距词。

定义7 设 $f \in \Sigma^*$ 是一个长度为 n 且具有 $n \geq 1$ 的词。若对于任意一对等长的 f-自由词 u 和 v，都存在从 u 到 v 的一个最小波浪变换，且该变换是 f-自由的，则称词 f 为波浪号等距。若其不是波浪号等距的，则称其为波浪号非等距。

要证明一个词是波浪号非等距的，只需举出一对违反定义7的词 (u, v)。证明一个词是波浪号等距的则更具挑战性。

例8 词 $f = 1010$ 是波浪号非等距的。事实上，令 $u = 11000$ 和 $v = 10110$；u 和 v 是 f-自由的；此外，从 u 到 v 唯一可能的最小波浪变换是 $11000 \xrightarrow{S_2} 10100 \xrightarrow{R_4} 10110$ 和 $11000 \xrightarrow{R_4} 11010 \xrightarrow{S_2} 10110$，而在这两种情况下，1010 都会在第一步后作为因子出现。另一方面，注意到 f 由命题*2*可知是汉明等距的。

注9 当一个波浪变换中包含相邻的交换和替换时，可能存在涉及不同操作集合的最小波浪变换。例如，数对 (u, v)，其中包含 $u = 010$ 和 $v = 101$，有两个最小波浪变换：$010 \xrightarrow{S_1} 100 \xrightarrow{R_3} 101$ 和 $010 \xrightarrow{S_2} 001 \xrightarrow{R_1} 101$。

仅允许替换时不会出现这种情况，因此，相较于汉明距离的情形，证明等距性会面临新的困难。

我们强调一下波浪等距二进制词的以下一条显而易见的性质，它有助于简化证明。

注 10 词 f 是波浪号等距词，当且仅当 f 是波浪号等距词；当且仅当 f^{rev} 是波浪号等距词。

鉴于注 10，我们将重点关注以 1 开头的词。下面的命题明确探讨了长度为 2、3 和 4 的所有字符串的波浪号等距性。

命题11 以下陈述成立。

1. 所有长度为2 的字符串都是波浪号等距的。*2.* 除101 和010外，所有长度为3的字符串都是波浪号等距的。*3.* 字符串1111、1110、1000、0000、0001、0111 都是波浪号等距的。其他所有长度为4 的字符串都是波浪号非等距的。

证明： 考虑以下不同情形：

1. 若 $|f| = 2$，在反转和取补意义下会出现两种情形： $f = 10$ 和 $f = 11$。根据下一个命题12，它们都是波浪号等距词。

2. 若 $|f| = 3$，在取补和反转意义下会出现三种情况。

   - $f = 111$和 $f = 100$是波浪号等距的，这是因为命题12。
   - $f = 101$是波浪号非等距的。事实上， $u = 1111$和 $v = 1001$与 f的等距性相矛盾，因为它们不含 f，而将 u 变换为 v 的所有最小变换都需要把 u 变为一个以101为因子的词。

3. 若 $|f| = 4$，在取补和反转意义下会出现以下几种情况。

   - $f = 1111$和 $f = 1110$ 是波浪号等距的，这是因为命题12。
   - $f = 1100$。于是， $u = 110100$ 和 $v = 101010$ 与 f的等距性相矛盾。
   - $f = 1001$。于是， $u = 11011$ 和 $v = 10001$ 与 f的等距性相矛盾。
   - $f = 1010$是波浪号非等距的（见例8）。
   - $f = 1011$。于是， $u = 11111$ 和 $v = 10011$ 与 f的等距性相矛盾。

□

根据命题2，我们来展示一个波浪号等距但非 Ham- 等距的字符串无限族。

命题12 设 $f_{h,k} = 1^h0^k$，其中 $h, k \geq 0$。那么，对于任意$h, k \geq 0$，$f_{h,k}$ 都是波浪号等距的，但 $h = k = 2$ 除外；也就是说，$f_{2,2} = 1100$ 是波浪号非等距的。

证明： 假设 $f = f_{h,k} = 1^h0^k$, $f \neq 1100$ 是波浪号非等距的，并设 (u, v) 是一对字符串，其中 $u, v \in \Sigma^m$，它们与定义7相矛盾；在所有长度为 m 且满足此条件的字符串对中，$d = \text{dist}_{\sim}(u, v)$ 最小。

设 $\{O_{i_1}, O_{i_2}, \ldots, O_{i_d}\}$ 为从 u 到 v、 $1 \leq i_1 < i_2 < \cdots < i_d \leq m$ 的最小波浪变换所涉及的操作集合，其中对于任意 $j = 1, 2, \ldots, d$, $O_{i_j} \in \{R_{i_j}, S_{i_j}\}$。那么，对于 $j = 1, 2, \ldots, d$， $O_{i_j}(u)$ 在一个

区间中有 $I_t = [k_t..k_t + n - 1]$ 的一次出现，记作 I_j，该区间至少包含一个被 O_{i_j} 修改的位置。注意，从 u 到 v 的波浪变换中，这次 f 的出现必定消失，因为 v 是无 f 的。因此， I_j 包含一个被 $\{O_{i_1}, O_{i_2}, \ldots, O_{i_d}\}$ 中另一操作修改的位置。根据鸽巢原理，存在 s、$t \in \{i_1, i_2, \ldots i_d\}$，使得 $O_s(u)$ 在 $I_s = [k_s..k_s + n - 1]$ 中有 f 的一次出现，且该出现至少包含一个被 O_t 修改的位置；并且 $O_t(u)$ 在 $I_t = [k_t..k_t + n - 1]$ 中有 f 的一次出现，且该出现至少包含一个被 O_s 修改的位置。不失一般性，设 $k_s < k_t$。现在，令 $I = [p..q]$ 为 I_s 与 I_t 的交集；该区间截取了 O_s 中 f 的一个后缀，以及 $O_t(u)$ 中 f 的一个前缀，二者长度相同，均为 l，且 $l = q - p + 1$。换言之，$(O_s(u))[p..q] = suf_l(f)$ 且 $(O_t(u))[p..q] = pref_l(f)$。注意，这意味着 $f \neq 1^k$。

区间 I 中， O_s 和 O_t修改的位置可能有四个、三个或两个，其中至少一个由 O_s 修改，且至少一个由 O_t修改。可以证明，所有情况下都会导致矛盾。我们只详细讨论涉及交换操作的部分情况；其他情况也可用类似方法处理。

考虑 I 包含 O_s 和 O_t修改的四个位置的情况。因此， O_s 和 O_t 是交换，即 $O_s = S_s$ 和 $O_t = S_t$，其中 s， $t \in [p..q]$。由于$(S_s(u))[p..q] = suf_l(f)$，可得

$u[p..q] \in 1^*010^*$。但是，这时 $u[p..q]$ 中不存在其他交换操作能够产生 f的前缀，而 S_t本应如此。

现在考虑 I 包含 O_s 和 O_t修改的两个位置的情况。根据 O_s 和 O_t 是两个替换操作还是两个交换操作，有三种可能的情况

（其中 $s = p - 1$ 和 $t = q$），或者一个是交换，另一个是替换（其中$s = p - 1$ 和 $t \in [p + 1..q]$ 或 $s \in [p..q - 1]$ 和 $t = q$）。

我们来考虑它们

都是交换操作的情况。如果 $O_s = S_{p-1}$ 和 $O_t = S_q$，那么在 $u[p..q]$ 中由 O_s 和 O_t 修改的两个位置必须是位置 p 和 q。

假设 $q = p + 1$。

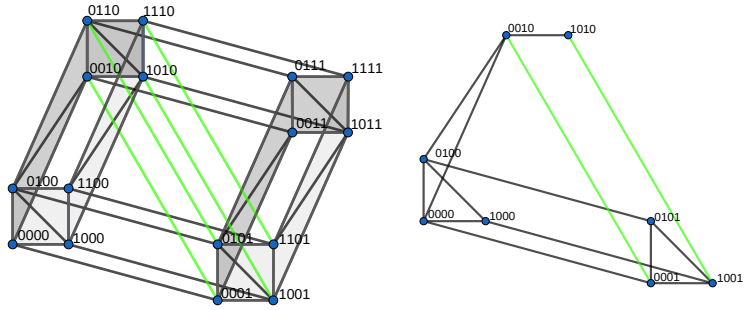
如果 $u[p..q] = 10$，那么 $u[p - 1..q + 1] = 0101$。将 S_{p-1} 应用于 u 意味着 f 以 **100** 结尾，而将 S_q 应用于 f 意味着其以 **110** 开头。因此， $f = 1100$, 这与假设相矛盾。在 $u[p..q] = 00, \mathbf{01}, \mathbf{11}$ 的情况下，分别将 S_{p-1} 应用于 u 会使 f的后缀变为 **010**、**011**、**101**，这就产生了矛盾。现在假设 $q > p + 1$。

若 $u[p] = 1$，则 $u[p..q] \in 10^*$，因为 $(S_{p-1}(u))[p..q] = suf_l(f)$，但在 $u[p..q]$ 中不存在其他交换操作能产生 f的前缀，而这正是 S_q所应具备的性质。类似的推理表明， $u[p] = 0$也不成立。因此，在这种情况下同样会导出矛盾。 □

波浪号等距性与汉明等距性不可比较。此外，以下结果成立。

命题13 词 11100 是最短的、并非汉明等距的波浪号等距词。词1010 是最短的、并非波浪号等距的汉明等距词。

证明： 词 **11100** 是波浪号等距的（命题12），但汉明非等距（命题2）。另一方面，**1010** 是波浪号非等距的（命题11），而且汉明等距（命题2）。这些词的长度之所以最小，是由命题11得出的。 □



(a) (b)

图 1：4 阶波浪号超立方体 (a)；4 阶波浪号斐波那契立方体 (b)

4 波浪号超立方体

经典超立方体按照顶点之间的汉明距离连接顶点，而 k 元 n 维超立方体中顶点之间的距离表示它们的李距离。这启发我们研究基于其他距离的超立方体。本文介绍波浪号超立方体，其顶点为二进制字，边连接波浪距离为1的顶点。随后，我们探讨了它的递归结构。

定义 14 n波浪号超立方体 Q_n是一个图，含有 2^n 个顶点，每个顶点对应一个长度为 n的二进制字。当两个顶点在 Q_n中，其波浪号距离为1时，它们相邻。

图 1(a) 展示了 4 阶波浪号超立方体。

注记 15 Q_n 是 Q_n的真子图。事实上，对于 $u, v \in \Sigma^*$, $dist_H(u, v) = 1$意味着 $dist_\sim(u, v) = 1$。此外，对于任意 $n \geq 2$，存在长度为 n 的词 u_n, v_n，使得 $dist_\sim(u_n, v_n) = 1$ 且 $dist_H(u_n, v_n) \neq 1$，例如 $u_n = 0^{n-2}01$ 和 $v_n = 0^{n-2}10$，因此 (u_n, v_n) 是 $\tilde{Q}_n$ 中的一条边，但不是 Q_n中的边。

下面的引理是展示波浪号超立方体递归定义的主要工具，这与经典超立方体的情形类似。

引理 16 对于任意 $u, v \in \Sigma^{n-1}$, $dist_\sim(u0, v0) = dist_\sim(u, v) = dist_\sim(u1, v1)$ 以及 $dist_\sim(u0, u1) = 1$。此外，对于任意 $u' \in \Sigma^{n-2}$, $dist_\sim(u'01, u'10) = 1$。

命题 17 Q_n 可以递归定义。

证明： 如果 $n = 1, Q_1$ 只有两个顶点 0 和 1，且它们由一条边连接。

假设维数小于 n 的波浪号超立方体已经定义。取两个 Q_{n-1}的副本。第一个副本中的所有顶点 u 均替换为 $u0$，第二个副本中则替换为 $u1$。根据引理 16，如果 u 和 v 在 Q_{n-1}中相连，那么 $u0$ 和 $v0$（分别为 $u1$ 和 $v1$）在 $\tilde{Q}_n$ 中相连。此外，对于第一个副本中的任意 $u \in \Sigma^{n-1}$、$u0$，以及第二个副本中的 $u1$，它们彼此相连。最后，对于 Q_{n-1} 中每个以 1 结尾的顶点，例如 $u = u'1$，第一个 Q_n 副本中的 $u'10$ 与第二个 Q_n 副本中的 $u'01$ 之间有一条边（见图 1(a) 中的绿色边）。对于其他任意一对词 u，$v \in \{0, 1\}^n$，有 $dist_\sim$（u，v）> 1。 □

˜**推论 18** 设 Q_n 为阶 n 的波浪号超立方体。那么

$$|E(\tilde{Q}_n)| = 2|E(\tilde{Q}_{n-1})| + 2^{n-1} + 2^{n-2},\ \textit{with}\ |E(\tilde{Q}_1)| = 1$$

˜**证明：** 根据命题 17 中的递归构造， Q_n 的边数是 $\tilde{Q}_{n-1}$ 的两倍（因为它有两个副本），再加上 2^{n-1} 条边，每个$\tilde{Q}_{n-1}$的顶点对应一条；此外还有 2^{n-2} 条边， $\tilde{Q}_{n-1}$ 中每个以 1 结尾的顶点对应一条。 □

通过求解递推关系，我们得到精确解 $|E(\tilde{Q}_n)| = (3n-1)\cdot 2^{n-2}$ (序列 A053220, 见 [17])。设 $EQ(N)$表示具有 N 个顶点的波浪号超立方体的边数。那么，

$$E\tilde{Q}(N) = N(3\log N - 1)/4 \tag{1}$$

5 避开一个词的波浪号超立方体

所谓的广义斐波那契立方体在 [12] 中被定义为超立方体的一个子图：其中，作为因子包含给定词的顶点均被移除。类比之下，我们引入波浪号超立方体和

波浪号斐波那契立方体。

˜**定义 19** 避开词 f 的 n-波浪号超立方体，记作 $Q_n(f)$，是从 Q_n 中移除那些以 f 为因子的顶点后得到的子-˜图。

下面的命题阐述了波浪等距词与避开某个词的波浪号超立方体子图之间的关系。由定义很容易得出证明。

˜**命题 20** 词 $f \in \Sigma^*$ 是波浪号等距词，当且仅当对所有 $n \geq |f|$， $Q_n(f)$ ˜都是 Q_n 的等距子图。

例 21 所有长度为 2 的二进制词既是波浪号等距词（命题*11*）˜，也是汉明等距词；对每个 $n \geq 1$， $Q_n(10)$ 与 $Q_n(10)$ 都重合。事实上，$V(\tilde{Q}_n(10))$ $= \{0^h1^k |\ h,k \geq 0,\ h{+}k = n\}$，且 $E(\tilde{Q}_n(10)) = \{(0^i1^j,\ 0^{i-1}1^{j+1})|\ 1 \leq i,\ j \leq n\}$。词 11 的情形值得单独一节讨论。长度为2 的其他词根据取补也都是波浪号等距词（见注 *10*）。

5.1 波浪号斐波那契立方体

不含词 11 的波浪号超立方体称为波浪号斐波那契立方体，这一命名是类比 Hsu [11] 引入的斐波那契立方体。这里，我们展示波浪号斐波那契立方体的一种递归构造；该构造使我们能够计算其边数，并将其与顶点数相同的波浪号超立方体的边数进行比较。

˜ ˜ ˜**定义 22** 称 n-波浪号斐波那契立方体，记作 F_n，为 $F_n = Q_n(11),\ n \geq 1$。

˜根据命题 3， $|V(F_n)| = |V(F_n)| = f_{n+2}$。在这些顶点中， f_{n+1} 以 0 结尾， f_n 以 1 结尾。图 1(b) 展示了阶为 4 的波浪号斐波那契立方体。

˜**注 23** 设 $u \in V(F_{n-1})$， $x \in \Sigma$。若 u 以1结尾，则 $ux \in V(F_n)$ 当且仅当 $x = 0$。˜若 u 以0 结尾，则对任意 $x \in \{0,1\}$， $ux \in V(F_n)$。

命题 24 波浪号斐波那契立方体 F_n 可以递归定义。

证明： 如果 $n = 1$, F_1 有两个顶点 0 和 1，并由一条边连接。如果 $n = 2$, F_2 有三个顶点 00、01 和 10，并且 $E(F_2) = \{(00, 10), (00, 01), (01, 10)\}$。

F_n 可以由 F_i 对所有 $i < n$都有定义。对于任意 $n \geq 3$，假设一份 F_{n-1} 的副本（记作 $\mathbf{F}_{n-1}$）和一份 F_{n-2} 的副本（记作 $\mathbf{F}_{n-2}$）构造，其中 F_{n-1} 中的每个顶点 u 都替换为 $\mathbf{F}_{n-1}$中的 $u0$，而 F_{n-2} 中的每个顶点 v 都替换为 $\mathbf{F}_{n-2}$中的 $v01$。此外，对于任意长度为 $n-2$的 v，$\mathbf{F}_{n-1}$ 中的 $v00$ 与 $\mathbf{F}_{n-2}$中的 $v01$ 之间有一条边，$\mathbf{F}_{n-1}$ 中的 $v10$ 与 $\mathbf{F}_{n-2}$ 中的 $v01$ 之间也有一条边（见图 1 中的绿色边）。根据注 23 和引理 16，F_n中不存在其他边。 □

推论 25 设 F_n 为波浪号斐波那契立方体。则 $|E(F_1)| = 1, |E(F_2)| = 3$以及

$$|E(\tilde{F}_n)| = |E(\tilde{F}_{n-1})| + |E(\tilde{F}_{n-2})| + f_{n+1}, \forall n \geq 2.$$

证明： 根据命题 24 的证明，F_1 有 1 条边，而 F_2 有 3 条边。此外，$|E(F_n)|$ 等于 $|E(F_{n-1})|$ 与 $|E(F_{n-2})|$之和，再加上$\mathbf{F}_{n-1}$中每个顶点各对应的一条边，即 f_{n+1}，由命题 3 得出。 □

通过求解推论 25 中的递推关系，我们得到以下精确解

$$|E(\tilde{F}_n)| = \frac{(n+1)f_{n+3} + (n-2)f_{n+1}}{5}$$

（[17] 中的数列 A023610，对应 $|E(F_{n+1})|$）。

由于 F_n 的顶点数为 f_{n+2}，由前面的公式可知，波浪号斐波那契立方体有 $O(N \log N)$条边，其中 N为顶点数，与波浪号超立方体的情况相同（见公式 (1)）。

为了比较斐波那契立方体和超立方体的边数，在 [11]中，作者证明了具有 N 个顶点的斐波那契立方体与超立方体的边数 $EF(N)$和 $EQ(N)$之比渐近地有界于 $0.79 < EF(N)/EQ(N) < 0.80$。类比这一结果，采用 [11], 中的相同方法，我们得到以下推论。

推论 26 设 $EF(N)$和 $EQ(N)$ 分别表示具有 N 个顶点的波浪号斐波那契立方体和波浪号超立方体的边数。那么，它们的比率渐近地有界于

$$0.85 < \frac{\tilde{EF}(N)}{\tilde{EQ}(N)} < 0.86$$

证明： 由公式 (1)和命题 3，$EQ(f_{n+2}) = f_{n+2}(3 \log f_{n+2} - 1)/4$；由推论 25，$EF(f_{n+2}) = ((n+1)f_{n+3} + (n-2)f_{n+1})/5$。渐近地考察$EF(f_{n+2})/EQ(f_{n+2})$，即可得证。 □

这证明，在顶点数固定时，波浪号斐波那契立方体的边数约比 $1/7$ 波浪号超立方体的边数少。这个比率仅略高于汉明情形。这并不令人意外，因为交换操作会增加新的边，但另一方面，由于一次交换相当于两次替换操作，因此会缩短平均距离。更正式地说，我们有如下备注。

n	4	5	6	7	8	9	10	11	12	13	14	15	16
$\|V(Q_n)\|=\|V(\tilde{Q}_n)\|$	16	32	64	128	256	512	1024	2048	4096	8192	16384	32768	65536
$\|E(Q_n)\|$	32	80	192	448	1024	2304	5120	11264	24576	53248	114688	245760	524288
$\|E(\tilde{Q}_n)\|$	44	112	272	640	1472	3328	7424	16384	35840	77824	167936	360448	770048
$\|V(\tilde{Q}_n(11100))\|$	16	31	60	116	224	432	833	1606	3096	5968	11504	22175	42744
$\|E(\tilde{Q}_n(11100))\|$	44	106	245	550	1208	2609	5569	11773	24691	51440	106566	219696	451005
$R_{\sim}$ (11100)	1	0,99	0,98	0,97	0,96	0,96	0,95	0,95	0,944	0,941	0,939	0,937	0,935
$\|V(Q_n(1010))\|$	15	28	53	100	188	354	667	1256	2365	4454	8388	15796	29747
$\|E(Q_n(1010))\|$	28	62	138	299	632	1323	2746	5645	11520	23377	47192	94830	189808
R_H (1010)	0,96	0,92	0,91	0,90	0,89	0,88	0,88	0,87	0,869	0,866	0,863	0,861	0,859

表1：顶点 a　　和边的基数以及某些阶数立方体的比率　　er $n = 4, \ldots, 16$

备注27 在 *[11]* 中已证明，直径 $d(F_n)$ 为 n ，且当 n为偶数时，最大距离对应的词为$(10)^{n/2}$ 和$(01)^{n/2}$ ；当 n为奇数时，则对应的词为$(01)^{\lfloor n/2 \rfloor}0$ 和$(10)^{\lfloor n/2 \rfloor}1$ 。若考虑波浪距离，则 $d(\tilde{F}_n) = \lceil n/2 \rceil$。事实上，相同的字符串具有最大的波浪距离，而将一个字符串变换为另一个字符串所需的最少波浪变换次数为：当 n 为偶数时，进行 $n/2$ 次交换；当 n为奇数时，进行 $\lfloor n/2 \rfloor$ 次交换和一次替换。

6 结论与未来工作

本文介绍了波浪号超立方体和波浪号斐波那契立方体，将其作为相应经典概念的推广，并以波浪距离取代汉明距离。

我们证明了，与经典情形一样，波浪号超立方体和波浪号斐波那契立方体可以递归定义。

这使得我们能够给出关于其边数与阶之间关系的递归公式和闭式公式。我们利用这些结果，量化了具有相同顶点数的波浪号斐波那契立方体与波浪号超立方体相比所拥有的边数，结果发现这一数值非常接近经典情形。不过，这项研究确实值得进一步深入，因为超立方体和波浪号超立方体是基于不同的距离定义的，而这两种距离预计适用于不同的应用。

此外，我们还考虑了避开某些特殊词（即 1010 和 11100）的超立方体。超立方体 Q_n(1010) 是 Q_n 的等距子图，而 ~ ~Q_n(1010) 则不是 Q_n 的等距子图。相反，波浪号超立方体 ~ ~Q_n(11100) 是 Q_n 的等距子图，而 Q_n(11100) 则不是 Q_n 的等距子图（参见命题13和20）。表1汇总了 Q_n、 Q_n、 Q_n(11100) 和 Q_n(1010) 的顶点数和边数的前几项数值，其中顶点数为 ~ ~。此外，对于每个 n，第 n 列中 R_H(1010) 的值，是 $|E(Q_n(1010))|$ 与顶点数等于 $N = |V(Q_n(1010))|$ 的超立方体的边数之比，即 $(N \log N)/2$（参见命题 3）。此外，对于每个 n，第 n 列中 $R_{\sim}$(11100) ~ 的值，是 $|E(Q_n(11100))|$ 与顶点数等于 $N = |V(Q_n(11100))|$ 的波浪号超立方体的边数之比，即 $N(3 \log N - 1)/4$（参见方程1）。

我们猜想，就减少边数而言，斐波那契立方体和波浪号斐波那契立方体都是避免某个词的最佳等距立方体，但目前相关研究还处于萌芽阶段。我们计划继续沿这一方向开展研究，尤其深入研究波浪号斐波那契立方体的结构和拓扑性质。

参考文献

[1] Amir, A., Eisenberg, E., Porat, E.：交换与失配编辑距离。Algorithmica **45**(1), 109–120 (2006)[2]Anselmo, M., Castiglione, G., Flores, M., Giammarresi, D., Madonia, M., Mantaci, S.：基于交换与失配距离的等距词。CORR abs/2303.03086 (2023)[3] Anselmo, M., Flores, M., Madonia, M.：四元 n-立方体与等距词。载于：词组合学。计算机科学讲义，卷 12842，第 27–39 页 (2021)[4] Anselmo, M., Flores, M., Madonia, M.：趣味老虎机与避免因子的词变换。载于：第11届算法趣味国际会议。LIPIcs，卷 226，第 4:1–4:15 页 (2022)[5] Anselmo, M., Flores, M., Madonia, M.：关于 k 元 n 立方体与等距词。理论计算机科学**938**, 50–64 (2022)[6] Anselmo, M., Giammarresi, D., Madonia, M., Selmi, C.：坏图：与重叠相关的一些结构性质。载于：DCFS 2020。计算机科学讲义，卷 12442，第 13–25 页 (2020)[7] Azarija, J., Klavˇzar, S., Lee, J., Pantone, J., Rho, Y.：关于广义斐波那契立方体的同构类。欧洲组合学杂志**51**, 372–379 (2016)[8] B´eal, M., Crochemore, M.：线性时间内检验一个词是否为汉明等距。理论计算机科学**933**, 55–59 (2022)[9] Faro, S., Pavone, A.：一种高效的交换匹配跳跃搜索方法。Com- put. J.**61**(9), 1351–1360 (2018)[10] Harary, F., Hayes, J., Wu, H.：超立方体图理论综述。Com- put. Math. Appl.**15**(4), 277–289 (1988)[11] Hsu, W.J.：斐波那契立方体——一种新的互连拓扑。IEEE Transactions on Parallel and Distributed Systems**4**(1), 3–12 (1993)[12] Ili´c, A., Klavˇzar, S., Rho, Y.：广义斐波那契立方体。离散数学**312**(1), 2–11 (2012)[13] Ili´c, A., Klavˇzar, S., Rho, Y.：二进制字的指标。理论计算机科学**452**, 100–106 (2012)[14] Klavˇzar, S.：斐波那契立方体的结构：综述。J. Comb. Optim. **25**(4), 505–522 (2013)[15] Klavˇzar, S., Shpectorov, S.V.：等距广义斐波那契立方体的渐近数量。欧洲组合学杂志 **33**(2), 220–226 (2012)[16] Munarini, E., Salvi, N.Z.：斐波那契立方体的结构与枚举性质。离散数学**255**(1-3), 317–324 (2002)[17] Sloane, N.：在线整数序列百科全书。http://oeis.org/

[18] Wei, J.：坏词的结构。欧洲组合学杂志**59**，204–214（2017）[19] Wei, J.：关于2-等距词猜想的证明。理论计算机科学**855**，68–73（2021）[20] Wei, J.、Yang, Y.、Wang, G.：等距词的循环嵌入性。离散数学**343**(10)，112024（2020）[21] Wei, J.、Yang, Y.、Zhu, X.：非等距二进制词的刻画。欧洲组合学杂志**78**，121–133（2019）[22] Wei, J.、Zhang, H.：广义斐波那契立方体的两个猜想的证明。欧洲组合学杂志**51**，419–432（2016）