Reconstructing Binary Words of Unknown Length with #-Subword Queries

Abstract

A guesser must find a hidden binary word of unknown length. For a query u, the answer is the number of times u occurs in the hidden word as a scattered subword. Two deterministic strategies reconstruct every word of length n. The first strategy asks any words and needs $O(\sqrt{\log n \cdot \log\log n})$ queries. The second asks only Lyndon words for $0 < 1$ and needs $O(\log n)$ queries. Both bounds count the two queries that give the length and allow unrestricted computation between queries. A reduction through pair projections gives the same asymptotic bounds for each fixed finite ordered alphabet, including the Lyndon restriction.

1 The game

A guesser must find a hidden binary word. It knows neither the letters of the word nor its length. It asks questions about the word and sees the answers. The cost is the number of questions, so one answer may carry an arbitrarily large integer.

1.1 Queries and answers

Let w be the hidden word, with letters 0 and 1, and let n be its length. A *query* is a word u. Its *answer* is the number $\mathrm{count}(w, u)$ of times that u occurs in w as a scattered subword. The letters of u sit at positions of w that need not be consecutive.

For example, the zeros of `010110` have three, two and no ones to their right, so the query `01` has the answer $3 + 2 + 0 = 5$. If the positions had to be consecutive, `01` would occur only twice.

Definition 1.1 (Occurrences, rules, runs and cost). An *occurrence* of a word u in a word w is a choice of positions $i_1 < \ldots < i_{|u|}$ in w whose letters, read from left to right, spell u. The number of occurrences is $\mathrm{count}(w, u)$. The empty word has exactly one occurrence.

A *history* is a finite list of pairs (q, a) of a query q and an answer a. A *rule* R maps every history to a move. A move is either $\mathrm{ask}(q)$ for a word q, or $\mathrm{stop}(v)$ for a word v, the guess of the rule.

The *run* of R on w starts with the empty history H_0. While $R(H_i) = \mathrm{ask}(q)$, the next history H_{i+1} is H_i followed by $(q, \mathrm{count}(w, q))$. The run ends at the first H_t with $R(H_t) = \mathrm{stop}(v)$. Then R *reconstructs* w if $v = w$, and the *cost* of the run is t, the number of queries.

A word is *Lyndon* if it is nonempty and smaller than each of its proper rotations. Thus $v < yx$ for every split $v = xy$ into nonempty words x and y. The order is the lexicographic order with $0 < 1$.

Let A be a set of words and ρ a function from $\mathbb{N}$ to $\mathbb{R}$. A rule *solves the game for A at rate ρ* if there are a constant $C > 0$ and a threshold θ such that for every word w the run of the

rule on w ends with the guess w, every query of the run is in A, and the cost is at most $C\rho(|w|)$ whenever $|w| \geq \theta$.

The rule sees only the history. It does not know n in advance. Nothing is counted between queries, so the rule may do any computation.

A word v is *consistent* with a history H if $\mathrm{count}(v, q) = a$ for every pair (q, a) in H. The set of these words is written $\mathrm{cons}(H)$. The hidden word is in $\mathrm{cons}(H)$ for every history H that occurs in its run. A rule that stops with a word of $\mathrm{cons}(H)$ picks it with a fixed choice function, and which word it picks does not matter, because $\mathrm{cons}(H) = \{w\}$ at the end of every run (Proposition 5.11 and Proposition 6.8). A rule is described only on the histories of its own runs. On every other history it stops with the empty word.

Theorem 1.2 (Unrestricted queries). There is a rule that reconstructs every binary word with arbitrary queries. For every word of length $n \geq 4$ its cost is at most

$$250\sqrt{\lfloor \log_2 n \rfloor \cdot \lfloor \log_2 \lfloor \log_2 n \rfloor \rfloor}.$$

In particular the cost is $\mathrm{O}(\sqrt{\log n \cdot \log\log n})$.

Theorem 1.3 (Lyndon queries). There is a rule that reconstructs every binary word and asks only Lyndon words. For every word of length n its cost is at most $28(1 + \lfloor \log_2 n \rfloor)$, where $\lfloor \log_2 0 \rfloor = 0$. In particular the cost is $\mathrm{O}(\log n)$.

Both strategies count the initial queries `0` and `1`. The constants bound these particular strategies. No lower bound is claimed.

1.2 Notation

Table 1 lists the symbols that several sections use.

Table 1: Notation used in more than one section.

w, n, z, m	the hidden word, its length, its number of zeros, its number of ones
$\lvert u\rvert$, $\lvert u\rvert_c$	the length of u; the number of letters c in u
uv, u^r	concatenation; r copies of u, with u^0 the empty word
$\overline{c}$	the letter other than c
$\mathrm{count}(w, u)$	the number of occurrences of u in w as a scattered subword
$\binom{x}{y}$	the binomial coefficient, equal to 0 if $y > x$
L	$\lfloor \log_2 n \rfloor + 1$, and 1 if $n = 0$; it satisfies $n < 2^L$
$g_s^c(u)$, $g_s(u)$	the c-groups of u, and $g_s = g_s^0$ (Section 3)
$P_d(s)$	the polynomial $\prod_{i=1}^{d}(1 - s/i)$ (Section 4)
$\mathrm{cons}(H)$	the words consistent with the history H

1.3 The two strategies

Both strategies first ask `0` and `1`. The answers are $z = \mathrm{count}(w, 0)$ and $m = \mathrm{count}(w, 1)$, the numbers of zeros and ones of w, and $n = z + m$. If $z = 0$ or $m = 0$ the word is known. The bounds count these two queries.

A word with J ones is determined by the lengths of the $J + 1$ runs of zeros separated by those ones. These lengths are its *zero groups* (Section 3). The queries 01^r return sums of the zero groups with binomial weights. A few consecutive queries of this form can be combined so that

the first group has weight 1, the next d groups have weight 0, and all later groups have small weights of one sign. A bound on the later groups makes rounding isolate the first group. This step is called *peeling*.

The Lyndon strategy (Section 5) locates the first two zeros by binary search and peels the groups from the left. It then peels the last groups from the right with Lyndon queries that start with two zeros, which the known prefix supplies. The unrestricted strategy (Section 6) reads the word from the left with the queries 01^r and 10^r together, until one letter is rare in the unread part. It finishes with queries 1^r0 or 0^r1.

2 Prior work and scope

2.1 Earlier query bounds

Richomme and Rosenfeld define #-subword queries in Section 2 of Reconstructing Words Using Queries on Subwords or Factors (STACS 2023). They prove a worst-case bound of $O(\sqrt{n \log n})$ queries for binary words and an average-case bound of $O(\log n)$ for a uniformly random binary word of a given length. Their conclusion asks whether $O(\log n)$ is possible in the worst case (STACS 2023, Section 5). Theorem 1.2 gives a worst-case cost of $o(\log n)$.

2.2 The size of the answers

A #-subword query can return a large integer. If $w = 0^n$ and $r = \lfloor n/2 \rfloor$, then $\mathrm{count}(w, 0^r) = \binom{n}{r}$. The bounds count queries. They leave the running time of the guesser unrestricted. These bounds do not decide whether a constant number of unrestricted queries suffices.

2.3 Provenance

The proofs were found with ChatGPT (OpenAI), using the model Astra, in late October 2026. The text and the Lean formalization were rewritten with Claude Code (Anthropic).

3 Zero groups and rows

A word with J ones is a list of $J + 1$ runs of zeros that the ones separate. The query 01^r returns a sum over these runs with binomial weights. The runs determine the word, so recovering their sizes reconstructs it.

3.1 Zero groups

Let u be a word and c a letter. Put $J = |u|_{\overline{c}}$. For $s \geq 0$, let $g_s^c(u)$ be the number of letters c in u that have exactly s letters $\overline{c}$ to their left. These numbers are the *c-groups* of u. Then $g_s^c(u) = 0$ for $s > J$, and

$$u = c^{g_0}\overline{c}c^{g_1}\overline{c}\ldots\overline{c}c^{g_J} \quad \text{with } g_s = g_s^c(u).$$

So u is determined by J and $g_0, \ldots, g_J$, and every list of $J + 1$ natural numbers is the list of groups of some word. Write $g_s(u)$ for $g_s^0(u)$. These numbers are the *zero groups* of u. The one-groups $g_s^1(u)$ are the zero groups of the word with the two letters exchanged.

> **Example 3.1 (Zero groups of `010110`).** The word `010110` has three ones. One zero lies before the first one, one between the first and the second, none between the second and the third, and one after the third. So $(g_0, g_1, g_2, g_3) = (1, 1, 0, 1)$.

3.2 Rows

The *rows* of u are the answers $\mathrm{count}(u, 01^r)$ for $r = 0, 1, 2, \ldots$. Each zero contributes according to the number of ones to its right.

Lemma 3.2 (Rows). For every word u with J ones and every $r \geq 0$,

$$\mathrm{count}(u, 01^r) = \sum_{s=0}^{J} g_s(u) \binom{J-s}{r}.$$

Proof. An occurrence of 01^r consists of a zero and r ones to its right. A zero in group s has exactly $J - s$ ones to its right, so it starts $\binom{J-s}{r}$ occurrences. There are $g_s(u)$ zeros in group s. Adding over all groups gives the formula. □

Exchanging the two letters gives the rows for the query 10^r.

$$\mathrm{count}(u, 10^r) = \sum_{s=0}^{|u|_0} g_s^1(u) \binom{|u|_0 - s}{r}.$$

For `010110` the formula gives the rows $3, 5, 4, 1, 0$ for $r = 0, \ldots, 4$. For example, $\mathrm{count}(010110, 01^2) = 1 \cdot \binom{3}{2} + 1 \cdot \binom{2}{2} + 0 \cdot \binom{1}{2} + 1 \cdot \binom{0}{2} = 4$. The row for $r = 1$ is the answer 5 computed in Section 1.

3.3 A known prefix

A known prefix contributes a computable term to each row. Subtracting that term gives the row of the unread suffix.

Lemma 3.3 (Known prefix). Let K and u be words, $B = |K|_1$ and $J = |u|_1$. For every $r \geq 0$ the following identities hold.
(1) $\mathrm{count}(Ku, 01^r) = \mathrm{count}(u, 01^r) + F$, where F depends only on K, J and r.
(2) $\mathrm{count}(Ku, 1^r0) = \mathrm{count}(K, 1^r0) + \sum_{j=0}^{J} g_j(u) \binom{B+j}{r}$.

Proof. An occurrence has its zero in K or in u.

(1) If the zero lies in u, the whole occurrence lies in u. If the zero lies in K and has o ones to its right in K, it has $o + J$ ones to its right in Ku, and it starts $\binom{o+J}{r}$ occurrences. The sum of these numbers over the zeros of K is F.

(2) If the zero lies in K, the whole occurrence lies in K. A zero of u in group j has j ones to its left in u, so it has $B + j$ ones to its left in Ku, and it ends $\binom{B+j}{r}$ occurrences. There are $g_j(u)$ such zeros. □

In particular, let Ku and Ku' be two words with the same known prefix K and $|u|_1 = |u'|_1$. Then $\mathrm{count}(Ku, 01^r) = \mathrm{count}(Ku', 01^r)$ holds if and only if $\mathrm{count}(u, 01^r) = \mathrm{count}(u', 01^r)$. The prefix cancels, and the strategies compare only the unread parts.

4 The peeling lemma

Consecutive rows determine a weighted sum in which the first group has weight 1 and the next groups have weight 0. A geometric bound makes the contribution of the later groups smaller than 1. Since group sizes are integers, the weighted sum then determines the first group.

4.1 The polynomial P_d

For $d \geq 0$ let $P_d(s) = \prod_{i=1}^{d}(1 - s/i)$. This is a polynomial of degree d in s, and $P_0 = 1$.

Lemma 4.1 (Values of P_d). $P_d(0) = 1$, and $P_d(s) = (-1)^d\binom{s-1}{d}$ for every integer $s \geq 1$. In particular $P_d(s) = 0$ for $1 \leq s \leq d$.

Proof. At $s = 0$ every factor is 1. For $s \geq 1$,

$$P_d(s) = \prod_{i=1}^{d} \frac{i-s}{i} = (-1)^d \frac{(s-1)(s-2)...(s-d)}{d!} = (-1)^d \binom{s-1}{d}.$$

If $s \leq d$, the factor with $i = s$ is 0, and $\binom{s-1}{d} = 0$. □

For example, $P_2(s) = (1-s)(1-s/2)$ takes the values $1, 0, 0, 1, 3, 6$ at $s = 0, ..., 5$.

4.2 The tail bound

Lemma 4.2 (Tail bound). Let $0 \leq y < 1$, and let d and N be natural numbers. Then

$$\sum_{s=1}^{N} y^s \binom{s-1}{d} \leq \left(\frac{y}{1-y}\right)^{d+1}.$$

Proof. Write S_d for the sum on the left. For $d = 0$ it is $y + ... + y^N \leq y/(1-y)$.

Pascal's rule $\binom{s}{d+1} = \binom{s-1}{d+1} + \binom{s-1}{d}$ gives

$$\sum_{s=1}^{N} y^s \binom{s}{d+1} = S_{d+1} + S_d.$$

Multiply by y and shift the index s by one. The term for $s = 1$ of the shifted sum is 0, because $\binom{0}{d+1} = 0$. So

$$y(S_{d+1} + S_d) = \sum_{s=2}^{N+1} y^s \binom{s-1}{d+1} = S_{d+1} + y^{N+1}\binom{N}{d+1}.$$

Hence $yS_d = (1-y)S_{d+1} + y^{N+1}\binom{N}{d+1} \geq (1-y)S_{d+1}$, so $S_{d+1} \leq \frac{y}{1-y}S_d$. By induction on d, $S_d \leq \left(\frac{y}{1-y}\right)^{d+1}$. □

4.3 The peeling lemma

Lemma 4.3 (Peeling). Let d and N be natural numbers. Let $g_0, ..., g_N$ and $\omega_0, ..., \omega_N$ be nonnegative real numbers with $\omega_0 = 1$, and put

$$S = \sum_{s=0}^{N} g_s \omega_s P_d(s).$$

(1) $(-1)^d(S - g_0) \geq 0$. So $S \geq g_0$ if d is even, and $S \leq g_0$ if d is odd.

Let moreover $0 \leq y < 1$ and $M \geq 0$ satisfy $M\left(\frac{y}{1-y}\right)^{d+1} < 1$, and suppose that $g_s\omega_s \leq My^s$ for $1 \leq s \leq N$.

(2) Then $(-1)^d(S - g_0) < 1$.

(3) Let $g'_0, ..., g'_N$ be natural numbers with $g'_s\omega_s \le My^s$ for $1 \le s \le N$, and let $g_0, ..., g_N$ be natural numbers. If $\sum_s g'_s\omega_s P_d(s) = S$, then $g'_0 = g_0$.

Proof. By Lemma 4.1, $S - g_0 = (-1)^d E$ with $E = \sum_{s=1}^N g_s\omega_s\binom{s-1}{d}$. Every term of E is nonnegative, so $E \ge 0$.

Under the additional hypotheses and by Lemma 4.2,

$$E \le M\sum_{s=1}^N y^s\binom{s-1}{d} \le M\left(\frac{y}{1-y}\right)^{d+1} < 1.$$

In (3), let E' be the error of the second sum, so that $S = g'_0 + (-1)^d E'$. Then $g_0 - g'_0 = (-1)^d(E' - E)$ with $E, E' \in [0, 1)$. So $|g_0 - g'_0| < 1$, and two integers at distance less than 1 are equal. □

4.4 From rows to the peeled sum

Lemma 4.4 (Moments). Let $\pi_0, ..., \pi_d$ be real polynomials with $\deg \pi_i = i$, and let $x_0, ..., x_N$ be real numbers with $\sum_{s=0}^N x_s\pi_i(s) = 0$ for $0 \le i \le d$. Then $\sum_{s=0}^N x_s f(s) = 0$ for every real polynomial f of degree at most d.

Proof. The $d+1$ polynomials π_i have different degrees, so they are linearly independent. The polynomials of degree at most d form a space of dimension $d+1$, so the π_i are a basis of it. Write $f = \sum_i \gamma_i\pi_i$. Then $\sum_s x_s f(s) = \sum_i \gamma_i \sum_s x_s\pi_i(s) = 0$. □

Lemma 4.5 (Forward rows). Let $0 \le k \le N$ and $d \ge 0$ be integers, and put $\omega_s = \binom{N-s}{k}/\binom{N}{k}$ for $0 \le s \le N$.
(1) $\omega_0 = 1$. If $0 \le y \le 1$ and $k \ge (1-y)N$, then $0 \le \omega_s \le y^s$ for all s.
(2) Let $x_0, ..., x_N$ be real numbers with $\sum_{s=0}^N x_s\binom{N-s}{r} = 0$ for $k \le r \le k+d$. Then $\sum_{s=0}^N x_s\omega_s P_d(s) = 0$.

Proof. (1) $\omega_0 = 1$ by definition. If $N - s < k$, then $\omega_s = 0$. Otherwise $\binom{t-1}{k} = \binom{t}{k}(t-k)/t$ for $t \ge k+1$ gives

$$\omega_s = \prod_{l=0}^{s-1}\frac{N-l-k}{N-l} = \prod_{l=0}^{s-1}\left(1 - \frac{k}{N-l}\right).$$

Every factor lies between 0 and $1 - k/N \le y$. So $0 \le \omega_s \le y^s$.

(2) For $0 \le i \le d$ and every integer $t \ge 0$,

$$\binom{t}{k+i} = \binom{t}{k}q_i(t), \quad \text{where } q_i(t) = \frac{\prod_{l=0}^{i-1}(t-k-l)}{\prod_{l=1}^{i}(k+l)}.$$

For $t < k$ both sides are 0. For $t \ge k$ cancel the factorials. If $t - k < i$, the numerator of $q_i(t)$ contains the factor 0, and $\binom{t}{k+i} = 0$. The polynomial q_i has degree i. Put $\pi_i(s) = q_i(N-s)$, a polynomial of degree i in s, and $Y_s = x_s\binom{N-s}{k}$. By the hypothesis, for $0 \le i \le d$,

$$\sum_{s=0}^N Y_s\pi_i(s) = \sum_{s=0}^N x_s\binom{N-s}{k+i} = 0.$$

By Lemma 4.4, $\sum_s Y_s P_d(s) = 0$. Divide by $\binom{N}{k}$. □

Corollary 4.6 (First group from consecutive rows). Let $0 \le k \le N$ and $d \ge 0$ be integers, $0 \le y < 1$ and $M \ge 0$ with $k \ge (1-y)N$ and $M\left(\frac{y}{1-y}\right)^{d+1} < 1$. Let $\xi_0, ..., \xi_N$ and $\xi'_0, ..., \xi'_N$ be natural numbers that are at most M, with

$$\sum_{s=0}^{N} \xi_s \binom{N-s}{r} = \sum_{s=0}^{N} \xi'_s \binom{N-s}{r} \quad \text{for } k \le r \le k+d.$$

Then $\xi_0 = \xi'_0$.

Proof. Apply Lemma 4.5 (2) to $x_s = \xi_s - \xi'_s$. With the weights ω_s of that lemma, $\sum_s \xi_s \omega_s P_d(s) = \sum_s \xi'_s \omega_s P_d(s)$. By Lemma 4.5 (1), $\xi_s \omega_s \le M y^s$ and $\xi'_s \omega_s \le M y^s$. Now Lemma 4.3 (3) gives $\xi_0 = \xi'_0$. □

Example 4.7 (Peeling `010110` with two rows). Take $u =$ `010110`, so $J = 3$ and $(g_0, ..., g_3) = (1, 1, 0, 1)$. Let $k = 1$ and $d = 1$. The weights are $\omega_s = \binom{3-s}{1}/3 = 1, 2/3, 1/3, 0$, and $P_1(s) = 1 - s$. For $t = 3 - s$,

$$3\omega_s P_1(s) = t(t-2) = 2\binom{t}{2} - \binom{t}{1}.$$

So $3S = 2\,\mathrm{count}(u, 01^2) - \mathrm{count}(u, 01) = 2 \cdot 4 - 5 = 3$, and $S = 1$. In general $S = g_0 - g_2/3$ for every word with three ones. The combination of the two rows removes group 1 and gives group 2 the weight $-1/3$. Here $g_2 = 0$, so $S = g_0$. With more rows the weights of all later groups shrink further, and Lemma 4.2 bounds their sum.

5 The Lyndon strategy

The Lyndon strategy locates the first two zeros by binary search. Two bands of queries 01^r determine a further common prefix of all consistent words. A band of queries 001^A01^p determines the unread suffix. The strategy stops with any consistent word, and uniqueness makes that guess correct.

5.1 The queries are Lyndon

Lemma 5.1 (Lyndon queries). The words 0, 1, 01^r for $r \ge 0$, and 001^A01^p for $A, p \ge 1$ are Lyndon.

Proof. A word with one letter has no proper rotation, and $01^0 = 0$. For $r \ge 1$, every proper rotation of 01^r starts with 1, and 01^r starts with 0, so 01^r is smaller. For 001^A01^p, a proper rotation that starts with 1 is larger than the word, because the word starts with 0. The other proper rotations start at the second or at the third zero. They are 01^A01^p0 and 01^p001^A. Both have the letter 1 at the second place, where the word has 0. So both are larger than the word. □

For example, `00101` is Lyndon because its proper rotations `01010`, `10100`, `01001` and `10010` are larger. The word `01101` is not Lyndon, because its rotation `01011` is smaller. The initial two zeros ensure that the guarded queries are Lyndon.

5.2 Locating the first two zeros

Let w have $z \geq 2$ zeros and m ones. Let J_1 and J_2 be the numbers of ones to the right of the first zero and of the second zero of w. There are $m - J_1$ ones before the first zero and $J_1 - J_2$ ones between the first two zeros. So

$$w = K_0 u \quad \text{with } K_0 = 1^{m-J_1} 0 1^{J_1 - J_2} 0,$$

where u has J_2 ones and $z - 2$ zeros.

Lemma 5.2 (First two zeros). Let w have $z \geq 2$ zeros and m ones, and let $r \geq 0$.

(1) $\mathrm{count}(w, 01^r) > 0$ if and only if $r \leq J_1$.

(2) $\mathrm{count}(w, 01^r) - \binom{J_1}{r} > 0$ if and only if $r \leq J_2$.

(3) If $m + 1 \leq 2^L$, binary search finds J_1 with at most L queries 01^r with $r \geq 1$. If $J_1 + 1 \leq 2^L$, binary search finds J_2 with at most L queries of this form, when J_1 is known.

Proof. Let o_x be the number of ones to the right of a zero x. Then $\mathrm{count}(w, 01^r) = \sum_x \binom{o_x}{r}$, summed over the zeros x of w. The term of x is positive if and only if $r \leq o_x$. The numbers o_x do not increase from the left to the right. The first zero has $o_x = J_1$, and the second has $o_x = J_2$. This proves (1). Removing the term $\binom{J_1}{r}$ of the first zero leaves the terms of the other zeros, of which the second has the most ones. This proves (2).

(3) Let J be J_1 or J_2. The test "$r \leq J$" costs one query by (1) or (2), it is true for $r = 0$, and it is false for $r > J$. Keep integers $a < b$ with $a \leq J < b$, starting with $a = 0$ and $b = m + 1$. While $b - a \geq 2$, test $r = \lfloor (a + b)/2 \rfloor$, which satisfies $r \geq 1$, and replace a by r if the test is true and b by r otherwise. Each test replaces $b - a$ by at most $\lceil (b - a)/2 \rceil$. Since $b - a \leq 2^L$ at the start, $b - a \leq 1$ after L tests, so $J = a$. For J_2 start with $b = J_1 + 1$. □

5.3 The rule

Let z and m be the answers to the first two queries, $n = z + m$ and $L = \lfloor \log_2 n \rfloor + 1$, with $L = 1$ if $n = 0$. Then $n < 2^L$, so $m + 1 \leq 2^L$ and $J_1 + 1 \leq 2^L$. Put $e = 2L + 6$.

Algorithm 5.3 (Rule Y).

(1) Ask `0` and then `1`.

(2) If $z = 1$, ask `01`.

(3) If $z \geq 2$, find J_1 and then J_2 by the binary searches of Lemma 5.2. If $J_2 \geq 1$, put $\rho_0 = J_2$ and, for $i = 1, 2$,

$$k_i = \max(\lceil 3\rho_{i-1}/5 \rceil, 1), \quad \rho_i = k_i - 1,$$

where $\lceil 3\rho/5 \rceil = \lfloor (3\rho + 4)/5 \rfloor$, and put $A = \rho_0 - \rho_2$. Then ask the following queries.

- The first band asks 01^r for $k_1 \leq r < k_1 + e$.
- The second band asks 01^r for $k_2 \leq r < k_2 + e$.
- The guarded band asks $001^A 01^p$ for $1 \leq p \leq e$.

(4) Stop with a word of $\mathrm{cons}(H)$, where H is the history.

Every number in this rule is computed from the answers in the history, so Y is a rule in the sense of Definition 1.1. The binary searches choose each query from the earlier answers. The queries of the three bands depend only on e and J_2.

Example 5.4 (The numbers ρ_i). For $J_2 = 7$ the rule computes $k_1 = \lceil 21/5 \rceil = 5$, $\rho_1 = 4$, $k_2 = \lceil 12/5 \rceil = 3$, $\rho_2 = 2$ and $A = 5$. Band 1 starts at the row $r = 5$ and band 2 at the row $r = 3$. For $J_2 = 1$ the rule computes $k_1 = 1$, $\rho_1 = 0$, $k_2 = 1$ and $\rho_2 = 0$.

Lemma 5.5 (Sizes). Let $\rho_0 \geq 1$. Then $k_1 \geq 3\rho_0/5$ and $k_2 \geq 3\rho_1/5$, and $k_1, k_2 \geq 1$. Moreover $25\rho_2 < 9\rho_0$, so $A > \rho_2$ and $A \geq 1$.

Proof. By definition $k_i \geq \lceil 3\rho_{i-1}/5 \rceil$ and $k_i \geq 1$. Since $\rho_0 \geq 1$, $k_1 = \lceil 3\rho_0/5 \rceil$, so $\rho_1 = k_1 - 1 < 3\rho_0/5$. If $\rho_1 \geq 1$, the same argument gives $\rho_2 < 3\rho_1/5 < 9\rho_0/25$. If $\rho_1 = 0$, then $k_2 = 1$ and $\rho_2 = 0$. In both cases $25\rho_2 < 9\rho_0$. Hence $\rho_2 < \rho_0/2$ and $A = \rho_0 - \rho_2 > \rho_0/2 > \rho_2 \geq 0$. □

5.4 The forward bands

Every consistent word follows the same binary searches, so the searches determine a common prefix.

Lemma 5.6 (Common shape). Let w have $z \geq 2$ zeros, let H be the history of the run of Y on w, and let $w' \in \mathrm{cons}(H)$. Then w and w' have the same z, m, J_1 and J_2. So $w = K_0 u$ and $w' = K_0 u'$ with $K_0 = 1^{m-J_1} 0 1^{J_1 - J_2} 0$, where u and u' have J_2 ones and $z - 2$ zeros.

Proof. The history H contains the answers to 0 and 1, so w and w' have the same z and m. The word w' gives the same answers as w to all queries of the run. Each test of a binary search asks whether $r \leq J$ for the word that gives the answer, by Lemma 5.2 (1) and (2). So the tests have the same outcomes for w and w', and the intervals $[a, b)$ are the same for both words. At the end a is J_1 for w and for w'. The same holds for J_2, because the second search uses the same J_1. The shape of w and w' is the one described above for every word with at least two zeros. □

Lemma 5.7 (Forward bands). Let $J \geq 1$, let $k_1, k_2, \rho_1, \rho_2, A$ be computed from $\rho_0 = J$ as in rule Y, and let $n < 2^L$ and $e = 2L + 6$. Let u and u' be words with J ones each and at most n zeros each. If

$$\mathrm{count}(u, 01^r) = \mathrm{count}(u', 01^r) \quad \text{for } k_i \leq r < k_i + e, i = 1, 2,$$

then $g_s(u) = g_s(u')$ for $0 \leq s < A$.

Proof. Write $g_s = g_s(u)$ and $g'_s = g_s(u')$. Every g_s and g'_s is at most n. Let $s < A$ and suppose that $g_{s'} = g'_{s'}$ for all $s' < s$. It remains to show $g_s = g'_s$. The lemma then follows by induction on s.

Put $N = J - s$. Since $s \leq A - 1 = J - \rho_2 - 1$, $N \geq \rho_2 + 1 = k_2$. If $N \geq k_1$, let $k = k_1$. Then $k \leq N$, and $k \geq 3J/5 \geq 3N/5$ by Lemma 5.5. Otherwise $k_2 \leq N < k_1$, so $N \leq \rho_1$. Let $k = k_2$. Then $k \leq N$, and $k \geq 3\rho_1/5 \geq 3N/5$. In both cases $k \leq N$ and $k \geq (1 - 2/5)N$.

The words u and u' give the same answers to the queries 01^r with $k \leq r < k + e$, which belong to band 1 or band 2. By Lemma 3.2 this is

$$\sum_{s'=0}^{J} g_{s'} \binom{J - s'}{r} = \sum_{s'=0}^{J} g'_{s'} \binom{J - s'}{r}.$$

The terms with $s' < s$ are equal, so they cancel. With $s' = s + i$ and $\xi_i = g_{s+i}$, $\xi'_i = g'_{s+i}$ the equation becomes

$$\sum_{i=0}^{N} \xi_i \binom{N-i}{r} = \sum_{i=0}^{N} \xi'_i \binom{N-i}{r} \quad \text{for } k \le r < k+e.$$

Now apply Corollary 4.6 with $d = e-1$, $y = 2/5$ and $M = n$. Here $\xi_i, \xi'_i \le n$. Also $n(2/3)^e < 1$ because $n < 2^L$ and $(2/3)^e = (4/9)^{L+3}$, so

$$n\left(\frac{2}{3}\right)^e < 2^L \left(\frac{4}{9}\right)^{L+3} = \left(\frac{8}{9}\right)^L \left(\frac{4}{9}\right)^3 \le \frac{64}{729} < 1.$$

So Corollary 4.6 gives $\xi_0 = \xi'_0$, so $g_s = g'_s$. □

5.5 The guarded band

From now on let w have $z \ge 2$ and $J_2 \ge 1$, and let $w' \in \mathrm{cons}(H)$. Write $J = J_2$. The words w and w' give the same answers to the queries of bands 1 and 2, so by Lemma 3.3 (1) the words u and u' do. They have J ones and $z - 2 \le n$ zeros, so Lemma 5.7 gives $g_s(u) = g_s(u')$ for $s < A$. Write $g_s = g_s(u)$. Then $u = fv$ and $u' = fv'$, where $f = 0^{g_0}10^{g_1}1 \ldots 0^{g_{A-1}}1$ has A ones, and v and v' have $\rho_2 = J - A$ ones. Put $K = K_0 f$. Then $w = Kv$ and $w' = Kv'$. The second zero of K has exactly A ones to its right in K, and $\rho_2 < A$ by Lemma 5.5.

The known prefix K supplies the first two zeros of each guarded query. The inequality $\rho_2 < A$ forces both zeros to lie in K.

Lemma 5.8 (Guarded counts). Let K be a word with at least two zeros whose second zero has exactly $A \ge 1$ ones to its right in K. Let v be a word with $\rho < A$ ones. Put $G(t) = \mathrm{count}\big(K1^t, 001^A\big)$ for $t \ge 0$. For every $p \ge 0$,

$$\mathrm{count}\big(Kv, 001^A01^p\big) = \Gamma_p + \sum_{t=0}^{\rho} g_t(v) G(t) \binom{\rho - t}{p},$$

where $\Gamma_p = \mathrm{count}\big(K1^\rho, 001^A01^p\big)$. The function G also satisfies the following bounds.
(1) $G(t) \ge 1$ for all $t \ge 0$.
(2) $(A + \rho)G(t-1) \le \rho G(t)$ for $1 \le t \le \rho$.

Proof. Call the zeros of an occurrence of 001^A01^p its first, second and third zero. Sort the occurrences by the place of the third zero.

If the third zero lies in K, then the first two zeros and the A ones following the second zero lie in K. The p final ones lie to the right of the third zero, in K or among the ρ ones of v. The number of such occurrences does not depend on the zeros of v, only on K and the number ρ of ones of v. It is the number of occurrences in $K1^\rho$, which has no zero outside K. Call it Γ_p.

Let the third zero lie in v, in group t. Thus t ones of v lie to its left. The second zero cannot lie in v, because A ones lie to its right before the third zero, and v has only $\rho < A$ ones. So the first two zeros lie in K. The A ones lie in K or among the t ones of v before the third zero. The number of choices of the first two zeros and of the A ones is $\mathrm{count}\big(K1^t, 001^A\big) = G(t)$. The p final ones are among the $\rho - t$ ones of v to the right of the third zero. There are $g_t(v)$ zeros in group t. Adding over t gives the formula.

For (1), the first two zeros of K and the A ones to the right of the second give an occurrence of 001^A in K, so $G(t) \geq G(0) \geq 1$.

For (2), let a zero x of K have q_x zeros of K to its left and o_x ones of K to its right. If x is the second zero of an occurrence of 001^A in $K1^t$, there are q_x choices for the first zero and $\binom{o_x+t}{A}$ choices for the ones. So $G(t) = \sum_x q_x \binom{o_x+t}{A}$. Only zeros with $q_x \geq 1$ contribute. They are the second zero of K or lie to its right, so $o_x \leq A$. For $j = o_x + t \geq 1$, $\binom{j-1}{A} = (1 - A/j)\binom{j}{A}$ if $j \geq A$, and both sides are 0 if $j < A$. Since $j \leq A + t$,

$$1 - \frac{A}{j} \leq 1 - \frac{A}{A+t} = \frac{t}{A+t} \leq \frac{\rho}{A+\rho}$$

for $t \leq \rho$. So $(A+\rho)\binom{o_x+t-1}{A} \leq \rho\binom{o_x+t}{A}$. Multiply by q_x and add over x. □

Example 5.9 (Guarded counts). Let $w = 00\,1^5\,0^2\,1\,0^3\,1\,0$, with $z = 8$ and $m = 7$. Both of the first two zeros have seven ones to their right, so $J_1 = J_2 = 7$ and $u = 1^5\,0^2\,1\,0^3\,1\,0$. The numbers of Example 5.4 give $A = 5$ and $\rho_2 = 2$. The first five groups of u are empty, so $K = 001^5$ and $v = 0^2\,1\,0^3\,1\,0$, whose zero groups are $(2, 3, 1)$. Here $G(0) = \binom{5}{5} = 1$ and $G(1) = \binom{6}{5} = 6$, and $\Gamma_p = 0$. Lemma 5.8 gives

$$\operatorname{count}(w, 001^5 01) = 2 \cdot 1 \cdot \binom{2}{1} + 3 \cdot 6 \cdot \binom{1}{1} = 22, \quad \operatorname{count}(w, 001^5 01^2) = 2 \cdot 1 \cdot \binom{2}{2} = 2.$$

Lemma 5.10 (Guarded band). Let K, A, ρ be as in Lemma 5.8, with $25\rho < 9(A+\rho)$. Let v and v' have ρ ones each and the same number of zeros, and let $n < 2^L$ be at least this number of zeros. Let $e = 2L + 6$. If

$$\operatorname{count}\big(Kv, 001^A 01^p\big) = \operatorname{count}\big(Kv', 001^A 01^p\big) \quad \text{for } 1 \leq p \leq e,$$

then $v = v'$.

Proof. Write $a_t = g_t(v)$ and $a'_t = g_t(v')$ and $\delta_t = a_t - a'_t$. Suppose that $a_t \neq a'_t$ for some $t < \rho$, and let $t_0 < \rho$ be the largest such index.

Subtract the formulas of Lemma 5.8 for v and v'. The terms Γ_p cancel. The terms with $t_0 < t < \rho$ vanish because $\delta_t = 0$, and the term $t = \rho$ vanishes because $\binom{0}{p} = 0$. So

$$\sum_{t=0}^{t_0} \delta_t G(t) \binom{\rho - t}{p} = 0 \quad \text{for } 1 \leq p \leq e.$$

Put $\lambda = \rho - t_0 \geq 1$ and $s = t_0 - t$. Then $\binom{\rho-t}{p} = \binom{\lambda+s}{p}$, and $p\binom{\lambda+s}{p} = (\lambda+s)\binom{\lambda+s-1}{p-1}$. With $i = p - 1$ this gives

$$\sum_{s=0}^{t_0} \delta_{t_0-s} G(t_0 - s)(\lambda + s) \binom{\lambda + s - 1}{i} = 0 \quad \text{for } 0 \leq i \leq e - 1.$$

For every integer $s \geq 0$, $\binom{\lambda+s-1}{i} = (\lambda+s-1)(\lambda+s-2)\ldots(\lambda+s-i)/i!$, a polynomial in s of degree i. By Lemma 4.4 with $d = e - 1$ and the polynomial P_{e-1},

$$\sum_{s=0}^{t_0} \delta_{t_0-s} G(t_0 - s)(\lambda + s) P_{e-1}(s) = 0.$$

Divide by $\lambda G(t_0) > 0$ and put $\omega_s = (\lambda + s)G(t_0 - s)/(\lambda G(t_0))$. Then

$$\sum_{s=0}^{t_0} a_{t_0-s}\omega_s P_{e-1}(s) = \sum_{s=0}^{t_0} a'_{t_0-s}\omega_s P_{e-1}(s).$$

Then $\omega_0 = 1$ and $\omega_s \geq 0$. Applying Lemma 5.8 (2) s times, which bounds each ratio $G(t-1)/G(t)$ by $\rho/(A+\rho)$, and using $(\lambda+s)/\lambda \leq 1+s$ and $\rho/(A+\rho) < 9/25$, gives

$$\omega_s \leq (1+s)\left(\frac{9}{25}\right)^s \leq 10\left(\frac{2}{5}\right)^s.$$

For the last step, Bernoulli's inequality gives $(10/9)^s \geq 1 + s/9$, so $10(10/9)^s \geq 1+s$. Since $a_t, a'_t \leq n$, this gives $a_{t_0-s}\omega_s \leq 10n(2/5)^s$, and the same for a'. Moreover $10n(2/3)^e < 10(4/9)^3 < 1$ by the computation in the proof of Lemma 5.7 and $640 < 729$. So Lemma 4.3 (3) with $y = 2/5$, $M = 10n$ and $d = e-1$ gives $a_{t_0} = a'_{t_0}$. This contradicts the choice of t_0. So $a_t = a'_t$ for all $t < \rho$.

The words v and v' have the same number of zeros, so $a_\rho = a'_\rho$ as well. A word is determined by its zero groups, so $v = v'$. □

5.6 Correctness and cost

Proposition 5.11 (Uniqueness). Let H be the history of the run of Y on w. Then $\mathrm{cons}(H) = \{w\}$.

Proof. Let $w' \in \mathrm{cons}(H)$. The words w and w' have the same z and m.

If $z = 0$, both words are 1^m. If $z = 1$, the history contains the answer a to the query `01`, which is the number of ones to the right of the only zero. So $w' = 1^{m-a}01^a = w$.

Let $z \geq 2$. By Lemma 5.6, $w = K_0 u$ and $w' = K_0 u'$, where u and u' have J_2 ones and $z-2$ zeros. If $J_2 = 0$, then $u = u' = 0^{z-2}$. Let $J_2 \geq 1$. As above, Lemma 3.3 (1) and Lemma 5.7 give $u = fv$ and $u' = fv'$, and $w = Kv$, $w' = Kv'$ with $K = K_0 f$. The word K satisfies the hypotheses of Lemma 5.8, and $25\rho_2 < 9(A+\rho_2)$ because $A + \rho_2 = \rho_0$ and Lemma 5.5. The words v and v' have the same number of zeros, at most n. The answers to band 3 agree. By Lemma 5.10, $v = v'$. So $w' = w$. □

Lemma 5.12 (Cost of rule Y). The run of Y asks at most $8L + 20$ queries, for every z.

Proof. If $z \geq 2$ the run asks two queries, at most L queries in each binary search by Lemma 5.2, and $3e = 6L + 18$ queries in the bands. The total is at most $2 + 2L + 6L + 18 = 8L + 20$. If $z = 1$ the run asks three queries, and if $z = 0$ two, and $L \geq 1$. □

Proof. Every query of Y is Lyndon by Lemma 5.1. The queries 01^r of the binary searches and the bands have $r \geq 1$, because $k_1, k_2 \geq 1$. The queries of band 3 have $A \geq 1$ by Lemma 5.5 and $p \geq 1$. The rule stops with a word of $\mathrm{cons}(H)$, which is w by Proposition 5.11. So Y reconstructs every word. By Lemma 5.12 the cost is at most $8L + 20 \leq 28L = 28(1 + \lfloor \log_2 n \rfloor)$ for $n \geq 1$, and the cost is 2 for $n = 0$. □

6 The unrestricted strategy

The unrestricted strategy asks both 01^r and 10^r in each round. Their answers force the unread count of at least one letter to fall by a factor smaller than $\frac{3}{4}$. Repeating the rounds makes one

letter rare in the unread suffix and common in the known prefix. Queries $c^r\overline{c}$ then determine the suffix.

6.1 Parameters

Let z and m be the answers to the first two queries, $n = z + m$ and $L = \lfloor \log_2 n \rfloor + 1$, with $L = 1$ if $n = 0$. For a letter a let n_a be the number of letters a in the hidden word, so $n_0 = z$ and $n_1 = m$. Put

$$\ell = \lceil \log_2 L \rceil + 2, \quad j = \left\lceil \sqrt{L/\ell} \right\rceil, \quad h = 3(j+1), \quad D = \lceil L/j \rceil, \quad T = 3L.$$

Let ℓ_0 be the even number and ℓ_1 the odd number among ℓ and $\ell + 1$, so $\ell_0 + \ell_1 = 2\ell + 1$.

Lemma 6.1 (Parameters).
(1) $n < 2^L$ and $n \cdot 3^T < 4^T$, so $n(3/4)^T < 1$.
(2) $3L \le 4L \le 2^\ell \le 2^{\ell_0}$ and $2^\ell \le 2^{\ell_1}$.
(3) $3^h(2^j + 1) \le 4^h$, so $(4/3)^h - 1 \ge 2^j$.
(4) $jD \ge L$.

Proof. (1) $n < 2^L$ by the definition of L. Then $n \cdot 3^T = n \cdot 27^L < 2^L \cdot 27^L = 54^L \le 64^L = 4^T$.

(2) $2^\ell = 4 \cdot 2^{\lceil \log_2 L \rceil} \ge 4L$.

(3) $3^h = 27^{j+1}$ and $4^h = 64^{j+1}$, and $2^j + 1 \le 2^{j+1}$. So $3^h(2^j + 1) \le 54^{j+1} \le 64^{j+1} = 4^h$.

(4) $D = \lceil L/j \rceil \ge L/j$. □

6.2 The rule

Let H be a history that contains the answers to 0 and 1. All words of $\mathrm{cons}(H)$ have z zeros and m ones. Let $K = \mathrm{lcp}(\mathrm{cons}(H))$ be their longest common prefix. For two words, $\mathrm{lcp}(w, w')$ is likewise their longest common prefix. Let $\mu_a = n_a - |K|_a$ be the number of letters a that are not in K.

For example, if $\mathrm{cons}(H)$ has the two elements `0101` and `0110`, then K is `01` and $\mu_0 = \mu_1 = 1$.

If $\mathrm{cons}(H)$ has at least two elements, then $\mu_0 \ge 1$ and $\mu_1 \ge 1$. Indeed, if $\mu_1 = 0$, every word of $\mathrm{cons}(H)$ is $K0^{\mu_0}$, so $\mathrm{cons}(H)$ has one element, and likewise for $\mu_0 = 0$.

Algorithm 6.2 (Rule U).
(1) Ask `0` and then `1`.
(2) For each round $i = 1, \ldots, 2h - 1$, if $\mathrm{cons}(H)$ has one element, stop with it. Otherwise let K, μ_0 and μ_1 be as above, and ask the queries
- 01^r for $k_0 \le r < k_0 + \ell_0$, where $k_0 = \lceil 3\mu_1/4 \rceil$,
- 10^r for $k_1 \le r < k_1 + \ell_1$, where $k_1 = \lceil 3\mu_0/4 \rceil$.

(3) If $\mathrm{cons}(H)$ has one element, stop with it. Otherwise let K, μ_0 and μ_1 be as above, let c be a letter with $4^h\mu_c < 3^h n_c$ (the smaller one if both qualify, and $c = 1$ if none does), and put $B = |K|_c$. Ask $c^r\overline{c}$ for $B \le r < B + D$.
(4) Stop with a word of $\mathrm{cons}(H)$.

Every round asks $\ell_0 + \ell_1 = 2\ell + 1$ queries, so the rule knows from the length of H which round it is in. The queries of a round depend only on the history before the round. So U is a rule in

the sense of Definition 1.1. Lemma 6.6 below shows that on a run a letter c as in step 3 exists, so the default $c = 1$ never matters. It only makes the rule defined for every history.

6.3 Complementary groups

Lemma 6.3 (Complementary groups). Let v be a word and $t < T$. If $g_t^1(v) \geq T$, then $g_s(v) \leq t$ for all $s < T$.

Proof. The ones of v with exactly t zeros to their left form a block of $g_t^1(v) \geq T$ consecutive ones, and exactly t zeros lie to the left of this block. A zero to the right of the block has at least T ones to its left. So every zero with fewer than T ones to its left lies to the left of the block. There are at most $t < T$ such zeros, and hence $g_s(v) \leq t$ for every $s < T$. □

For example, let $v =$ `0011111000` and $T = 4$. The five ones have two zeros to their left, so $g_2^1(v) = 5 \geq 4$. Only the first two zeros have fewer than four ones to their left, and indeed $g_0(v) = 2 \leq t$ and $g_1(v) = g_2(v) = g_3(v) = 0$.

6.4 One round

Lemma 6.4 (Round). Let n_0, n_1 be numbers, $n = n_0 + n_1$, let K be a word and $\mu_a = n_a - |K|_a$. Let $w = x0u$ and $w' = x1u'$ be words with n_a letters a for $a = 0, 1$, such that K is a prefix of x. Suppose that w and w' give the same answers to the queries of one round of U for μ_0 and μ_1. Then $4|0u|_a < 3\mu_a$ for some letter a.

Proof. Put $v = 0u$ and $v' = 1u'$. Both w and w' have n_a letters a, so $J_a = |v|_a = |v'|_a$. Since K is a prefix of x, $J_a \leq \mu_a$. Also $J_0 \geq 1$ because v starts with 0, and $J_1 \geq 1$ because v' starts with 1.

Suppose that $4J_a \geq 3\mu_a$ for both letters. Since J_a is an integer, $J_1 \geq k_0$ and $J_0 \geq k_1$.

Rows of zeros. Let $J = J_1$, $k = k_0$ and $d = \ell_0 - 1$, which is odd. Then $k \leq J$ and $k \geq 3\mu_1/4 \geq 3J/4$. The words w and w' give the same answers to the queries 01^r with $k \leq r < k + \ell_0$. By Lemma 3.3 (1) with the prefix x, and by Lemma 3.2,

$$\sum_{s=0}^{J} g_s(v) \binom{J-s}{r} = \sum_{s=0}^{J} g_s(v') \binom{J-s}{r} \quad \text{for } k \leq r \leq k + d.$$

Put $\omega_s = \binom{J-s}{k} / \binom{J}{k}$. By Lemma 4.5 (2), $S(v) = S(v')$, where $S(v) = \sum_s g_s(v)\omega_s P_d(s)$. By Lemma 4.5 (1) with $y = 1/4$, $0 \leq \omega_s \leq 4^{-s}$.

Rows of ones. Exchange the roles of the letters. Let $J' = J_0$, $k' = k_1$ and $d' = \ell_1 - 1$, which is even. In the same way, $S'(v) = S'(v')$, where $S'(v) = \sum_s g_s^1(v)\omega'_s P_{d'}(s)$ and the weights $\omega'_s = \binom{J'-s}{k'} / \binom{J'}{k'}$ satisfy $0 \leq \omega'_s \leq 4^{-s}$.

Since v starts with 0 and v' starts with 1, $g_0(v') = 0$ and $g_0^1(v') \geq 1$, and $g_0(v) \geq 1$ and $g_0^1(v) = 0$. By Lemma 4.3 (1), $S(v') \leq g_0(v') = 0$ because d is odd, and $S'(v') \geq g_0^1(v') \geq 1$ because d' is even. Hence

$$S(v) \leq 0 \quad \text{and} \quad S'(v) \geq 1.$$

The hypotheses of peeling. Put $M = 3(T-1)/4$ and $y = 1/3$. Then $M \geq 1$. Suppose that $g_s(v) < T$ for all $s < T$. Here $g_s(v) \leq n$ for every s. For $1 \leq s < T$,

$$g_s(v)\omega_s \le (T-1)4^{-s} = (T-1)\left(\frac{3}{4}\right)^s 3^{-s} \le M3^{-s}.$$

For $s \ge T$, $n3^T < 4^T$ by Lemma 6.1 (1), so

$$g_s(v)\omega_s \le n\left(\frac{3}{4}\right)^s 3^{-s} \le n\left(\frac{3}{4}\right)^T 3^{-s} < M3^{-s}.$$

Also $M(y/(1-y))^{d+1} = M2^{-\ell_0} < T2^{-\ell} \le 1$ by Lemma 6.1 (2). So Lemma 4.3 (2) applies, and since d is odd it gives $g_0(v) - S(v) < 1$. Then $S(v) > g_0(v) - 1 \ge 0$, which contradicts $S(v) \le 0$. So some $s < T$ has $g_s(v) \ge T$.

In the same way, if $g^1_s(v) < T$ for all $s < T$, then Lemma 4.3 (2) with d' even gives $S'(v) - g^1_0(v) < 1$, so $S'(v) < 1$. This contradicts $S'(v) \ge 1$. So some $t < T$ has $g^1_t(v) \ge T$. But then Lemma 6.3 gives $g_s(v) \le t < T$ for all $s < T$, which was just excluded. This contradiction proves the lemma. □

6.5 The rounds shrink the unread part

For the run of U on a word w, let H_0 be the history of the first two queries, and let H_i be the history after the rounds $1, ..., i$, as long as the rule has not stopped. Write $K(H) = \mathrm{lcp}(\mathrm{cons}(H))$ and $\mu_a(H) = n_a - |K(H)|_a$.

Lemma 6.5 (Contraction). Let H contain the answers to 0 and 1, let $w \in \mathrm{cons}(H)$, and let H' be H followed by the queries of one round of U with the answers of w. If $\mathrm{cons}(H')$ has at least two elements, then $K(H)$ is a prefix of $K(H')$, so $\mu_a(H') \le \mu_a(H)$ for both letters, and $4\mu_a(H') < 3\mu_a(H)$ for at least one letter a.

Proof. Since $\mathrm{cons}(H') \subseteq \mathrm{cons}(H)$, every word of $\mathrm{cons}(H')$ starts with $K(H)$. So $K(H)$ is a prefix of $K' = K(H')$. All words of $\mathrm{cons}(H')$ have the same length. If they all had the same letter after K', then K' would not be their longest common prefix, unless all of them were equal to K', and there are at least two of them. So there are $w_1, w_2 \in \mathrm{cons}(H')$ with $w_1 = K'0u$ and $w_2 = K'1u'$. Both have n_a letters a, and they give the same answers as w to the queries of the round. Apply Lemma 6.4 with $K = K(H)$ and $x = K'$. It gives a letter a with $4|0u|_a < 3\mu_a(H)$, and $|0u|_a = n_a - |K'|_a = \mu_a(H')$. □

Lemma 6.6 (One letter contracts often). Suppose that $\mathrm{cons}(H_{2h-1})$ has at least two elements. Then there is a letter c with $4^h\mu_c < 3^h n_c$, where $\mu_c = \mu_c(H_{2h-1})$. For every such letter put $B = n_c - \mu_c$ and $b = \mu_c$. Then $b \ge 1$ and $2^j b < B$, so $b < B$.

Proof. Write $\mu_a(i) = \mu_a(H_i)$. For $i \le 2h-1$, $\mathrm{cons}(H_i)$ has at least two elements, because $\mathrm{cons}(H_{2h-1}) \subseteq \mathrm{cons}(H_i)$. By Lemma 6.5, each round $i = 1, ..., 2h-1$ contracts some letter, so $4\mu_a(i) < 3\mu_a(i-1)$ for some a. Choose one such letter for each round, and let $\gamma_a(i)$ be the number of rounds up to i that chose a. So $\gamma_0(i) + \gamma_1(i) = i$.

Invariant. For both letters a, $4^{\gamma_a(i)}\mu_a(i) \le 3^{\gamma_a(i)}n_a$, with strict inequality if $\gamma_a(i) > 0$. For $i = 0$ this holds, because $\mu_a(0) \le n_a$. In round $i+1$, a letter a that was not chosen has γ_a unchanged and $\mu_a(i+1) \le \mu_a(i)$ by Lemma 6.5, so both claims persist. For the chosen letter, with $\gamma = \gamma_a(i)$,

$$4^{\gamma+1}\mu_a(i+1) < 4^\gamma \cdot 3\mu_a(i) \le 3 \cdot 3^\gamma n_a.$$

Since $\gamma_0(2h-1)+\gamma_1(2h-1)=2h-1$, some letter c has $\gamma=\gamma_c(2h-1)\geq h\geq 1$. Then $4^\gamma\mu_c<3^\gamma n_c$. Multiply the left side by $4^{\gamma-h}\geq 3^{\gamma-h}$ to get $4^h\mu_c 3^{\gamma-h}\leq 4^\gamma\mu_c<3^\gamma n_c$, and divide by $3^{\gamma-h}$. So $4^h\mu_c<3^h n_c$.

Now let c be any letter with $4^h\mu_c<3^h n_c$. If $b=0$, then $K(H_{2h-1})$ contains all letters c of the words of $\mathrm{cons}(H_{2h-1})$. Two words with the same numbers of zeros and ones, a common prefix, and no letter c after it are equal, so $\mathrm{cons}(H_{2h-1})$ has one element. So $b\geq 1$. Since $n_c=B+b$, $4^h b<3^h(B+b)$. By Lemma 6.1 (3), $3^h 2^j b+3^h b\leq 4^h b<3^h B+3^h b$, so $2^j b<B$. □

6.6 Completion

Lemma 6.7 (Completion). Let w and w' be words with the same numbers of zeros and ones and a common prefix K. Let c be a letter, $B=|K|_c$, $b=|w|_c-B$ and $q=|w|_{\overline{c}}-|K|_{\overline{c}}$. Let $D\geq 1$, $1\leq b<B$ and $qb^D<B^D$. If

$$\mathrm{count}(w,c^r\overline{c})=\mathrm{count}(w',c^r\overline{c})\quad\text{for } B\leq r<B+D,$$

then $w=w'$.

Proof. Exchanging the letters if necessary, let $c=1$. Write $w=Ku$ and $w'=Ku'$. Then u and u' have b ones and q zeros. By Lemma 3.3 (2), the answers $\mathrm{count}(Ku,1^r0)$ and $\mathrm{count}(Ku',1^r0)$ differ by

$$\sum_{t=0}^{b}(g_t(u)-g_t(u'))\binom{B+t}{r}.$$

So for $B\leq r<B+D$ this sum is 0. Put $N=B+b$ and $\xi_s=g_{b-s}(u)$ and $\xi'_s=g_{b-s}(u')$ for $0\leq s\leq b$, and $\xi_s=\xi'_s=0$ for $b<s\leq N$. Since $B+t=N-s$ for $s=b-t$,

$$\sum_{s=0}^{N}\xi_s\binom{N-s}{r}=\sum_{s=0}^{N}\xi'_s\binom{N-s}{r}\quad\text{for } B\leq r<B+D.$$

Suppose that $u\neq u'$. Then $\xi_s\neq\xi'_s$ for some s, because the groups determine the word. Let $s_0\leq b$ be the least such s. Cancel the terms with $s<s_0$ and put $N'=N-s_0\geq B$. With $s=s_0+s'$ the equation becomes

$$\sum_{s'=0}^{N'}\xi_{s_0+s'}\binom{N'-s'}{r}=\sum_{s'=0}^{N'}\xi'_{s_0+s'}\binom{N'-s'}{r}\quad\text{for } B\leq r<B+D.$$

Apply Corollary 4.6 with $k=B$, $d=D-1$, $y=b/(B+b)$ and $M=q$. Here $k\geq(1-y)N'$, because $(1-y)N'=BN'/(B+b)\leq B$. The numbers $\xi_{s_0+s'}$ and $\xi'_{s_0+s'}$ are at most q. And $M(y/(1-y))^D=q(b/B)^D<1$ because $qb^D<B^D$. So $\xi_{s_0}=\xi'_{s_0}$, which contradicts the choice of s_0. □

6.7 Proof of the theorem

Proposition 6.8 (Uniqueness). Let H be the history of the run of U on w. Then $\mathrm{cons}(H)=\{w\}$.

Proof. If the rule stops in a round or in step 3 because $\mathrm{cons}(H)$ has one element, the claim holds. Otherwise the rule asks the $2h-1$ rounds and the completion queries, and $\mathrm{cons}(H_{2h-1})$

has at least two elements. Let c, B and b be as in step 3 of the rule. By Lemma 6.6, $1 \le b < B$ and $2^j b < B$. The unread part has $q \le n < 2^L$ letters $\overline{c}$. By Lemma 6.1 (4), $2^L \le 2^{jD}$, so

$$q b^D < 2^{jD} b^D = (2^j b)^D < B^D.$$

Let $w' \in \mathrm{cons}(H)$. Then w and w' have the same numbers of zeros and ones and the common prefix K, and they give the same answers to the queries $c^r \overline{c}$ with $B \le r < B + D$. By Lemma 6.7, $w' = w$. □

Proof. By Proposition 6.8 the rule U reconstructs every word. The cost is at most

$$2 + (2h - 1)(2\ell + 1) + D = 2 + (6j + 5)(2\ell + 1) + \lceil L/j \rceil.$$

Let $n \ge 4$, $\alpha = \lfloor \log_2 n \rfloor$ and $\beta = \lfloor \log_2 \alpha \rfloor$. Then $\alpha \ge 2$ and $\beta \ge 1$, $L = \alpha + 1$, and $\lceil \log_2 L \rceil = \beta + 1$ because $2^\beta < \alpha + 1 \le 2^{\beta+1}$. So $\ell = \beta + 3$. Put $S = \sqrt{L\ell}$. Then the following bounds hold.
(1) $L \ge 3$ and $\ell \ge 4$, so $S \ge 3$.
(2) $\ell \le L + 1 \le 4L$, because $\lceil \log_2 L \rceil \le L - 1$ as $2^{L-1} \ge L$. So $\ell^2 \le 4L\ell$, so $\ell \le 2S$.
(3) $\sqrt{L/\ell} \le j < \sqrt{L/\ell} + 1$ gives $j\ell < S + \ell$ and $L/j \le S$.

With $6j + 5 \le 11j$ and $2\ell + 1 \le 3\ell$, the cost is at most

$$2 + 33j\ell + (S + 1) < 3 + 33(S + \ell) + S \le 3 + 34S + 66S \le 101S.$$

Finally $L\ell = (\alpha + 1)(\beta + 3) \le (3\alpha/2)(4\beta) = 6\alpha\beta$. So the cost is at most

$$101\sqrt{6\alpha\beta} \le 250\sqrt{\alpha\beta},$$

because $6 \cdot 101^2 = 61206 \le 62500 = 250^2$. □

7 Reduction to binary projections

Let Σ be a finite alphabet of size k. For $B \subseteq \Sigma$, the *projection* $\pi_B(w)$ erases every letter of w outside B. For distinct letters a and b, choose a relabelling ι_{ab} that sends 0 to a and 1 to b. Write v_{ab} for the binary word obtained from $\pi_{\{a,b\}}(w)$ by the inverse relabelling.

Lemma 7.1 (Projection answers). For every binary word u and every pair of distinct letters a, b,

$$\mathrm{count}(w, \iota_{ab}(u)) = \mathrm{count}(v_{ab}, u).$$

Proof. Every occurrence of $\iota_{ab}(u)$ uses positions labelled a or b. Erasing the other positions keeps the order of the chosen positions and gives an occurrence in $\pi_{\{a,b\}}(w)$. Each occurrence in the projection lifts to the original positions in w. These operations are inverse, and relabelling gives the stated equality. □

Lemma 7.2 (Pair projections determine the word). Two words over Σ are equal if they have the same number of each letter and the same projection onto every unordered pair of distinct letters.

Proof. Induct on the common length, which is the sum of the letter counts. At length 0 both words are empty. Suppose that the words are nonempty and start with a and b. If $a \ne b$, their projections onto $\{a, b\}$ start with different letters, contradicting equality of these projections.

Thus both words start with a. Delete that first letter from both words. The letter counts still agree. A projection onto a pair containing a loses the same first letter on both sides, so cancellation leaves equal projections of the suffixes. Every other pair projection is unchanged. The induction hypothesis makes the suffixes equal, hence the original words are equal. □

Let R be a deterministic binary rule that reconstructs every binary word. Write $q_R(v)$ for the cost of its run on v, and let A be a set containing every query on these runs. Each pair uses a fixed orientation and a fixed place in a list of all unordered pairs. These choices make the reduction deterministic.

Theorem 7.3 (Finite alphabet reduction). There is a rule that reconstructs every word w over Σ. It first asks each of the k singleton letters and then simulates R on v_{ab} for each unordered pair $\{a, b\}$. Its cost is

$$k + \sum_{\{a,b\}\subseteq\Sigma, a\neq b} q_R(v_{ab}),$$

where the sum takes each unordered pair once. Every query is either a singleton letter or $\iota_{ab}(u)$ for some $u \in A$.

Proof. The singleton queries give all letter counts. For each pair, simulate R from its empty history. Whenever R asks u, ask $\iota_{ab}(u)$ about w and return the answer to the simulation. By Lemma 7.1 the simulated history is the run of R on v_{ab}. The simulation therefore ends and returns v_{ab} after $q_R(v_{ab})$ queries. The rule performs every query of each simulated run, including any singleton queries that R repeats.

All pair projections are now known. The hidden word satisfies these projections and the known letter counts, so at least one word satisfies the collected data. By Lemma 7.2 at most one word satisfies them. The rule chooses that word and stops. There are finitely many words of the known length, so unrestricted computation can make this choice. The initial block has length k, and concatenating the simulated runs gives the stated cost and query restriction. If k is 0 or 1, the letter counts alone determine the word and the pair sum is empty. □

Corollary 7.4 (Uniform query bound). Fix n and a nonnegative number $Q(n)$. Suppose that $q_R(v) \le Q(n)$ for every binary word v of length at most n. Every word over Σ of length n is then reconstructed with at most

$$k + \binom{k}{2} Q(n)$$

queries, subject to the query restriction of Theorem 7.3.

Proof. Each projection has length at most n, and there are $\binom{k}{2}$ unordered pairs. Bound each summand in Theorem 7.3 by $Q(n)$. □

Corollary 7.5 (Fixed ordered alphabets). For every fixed finite ordered alphabet, arbitrary queries reconstruct words of unknown length n in $\mathrm{O}(\sqrt{\log n \cdot \log\log n})$ queries. Lyndon queries reconstruct them in $\mathrm{O}(\log n)$ queries.

Proof. Apply Theorem 7.3 to the binary rules of Theorem 1.2 and Theorem 1.3. Their explicit bounds give the required uniform bounds for projection lengths at most n. For the unrestricted

rule, projection lengths below 4 contribute a fixed constant. For larger lengths, both factors $\lfloor \log_2 n \rfloor$ and $\lfloor \log_2 \lfloor \log_2 n \rfloor \rfloor$ are nondecreasing. The Lyndon bound is nondecreasing as well. Since k is fixed, Corollary 7.4 gives the two asymptotic bounds.

For the Lyndon rule, orient every pair so that $a < b$. Relabelling then keeps lexicographic comparisons between binary words. It also commutes with rotations, so a relabelled Lyndon word is smaller than every proper rotation. Every initial singleton letter is Lyndon. Thus every query of the reduced rule is Lyndon over Σ. □

8 The Lean formalization

The binary reconstruction theorems are proved in Lean. Table 2 identifies the formal statements supporting the paper proofs. All files lie in the directory `proofs/Lax100001Proofs` of the entry. Names in the Lyndon and unrestricted groups lie in the namespaces `Lax100001Proofs.Lyndon` and `Lax100001Proofs.Unrestricted`, and the first group lies in `Lax100001Proofs`.

Table 2: Paper items and the Lean theorems that state them.

Paper item	Lean name	File
Rows (Lemma 3.2)	`count_zero_ones`	`Results`
Known prefix (Lemma 3.3)	`count_append_zero_ones`, `count_append_ones_zero`	`Prefix`
Tail bound (Lemma 4.2)	`tail_le`	`Results`
Peeling (Lemma 4.3)	`peel_sign_nonneg`, `peel_lt_one`, `peel_head_eq`	`Results`
First group (Corollary 4.6)	`head_eq_of_rows_eq`, `eq_of_rows_eq_of_eq_below`	`Peel`
Lyndon queries (Lemma 5.1)	`isLyndon_singleton`, `isLyndon_zeroOnes`, `isLyndon_guarded`	`Lyndon/Words`
First two zeros (Lemma 5.2)	`zero_lt_count_firstTwoZeros_iff`, `choose_lt_count_firstTwoZeros_iff`, `bisect_found_eq`, `found_firstSearch`, `found_secondSearch`	`Lyndon/Words`, `Lyndon/Search`
Rule Y (Algorithm 5.3)	`rule`, `search`, `bandQueries`	`Lyndon/Strategy`, `Lyndon/Search`
Sizes (Lemma 5.5)	`sizes`	`Lyndon/Bands`
Common shape (Lemma 5.6)	first part of the proof of `eq_of_agree`	`Lyndon/Strategy`
Forward bands (Lemma 5.7)	`groups_eq_of_forward_bands`, `exists_common_prefix`	`Lyndon/Bands`
Guarded counts (Lemma 5.8)	`count_append_guarded`, `guardCoeff_pos`, `guardCoeff_decay`	`Lyndon/Guard`
Guarded band (Lemma 5.10)	`eq_of_count_append_guarded_eq`	`Lyndon/Guard`
Uniqueness (Proposition 5.11)	`eq_of_agree`	`Lyndon/Strategy`
Cost (Lemma 5.12)	`transcript_rule`	`Lyndon/Strategy`
Theorem 1.3	`reconstructs`, `lyndon_reconstruction`	`Lyndon/Strategy`, `Results`
Parameters (Lemma 6.1)	`mul_pow_threshold_lt`, `three_pow_mul_le`, `lt_two_pow_bitLength`, `four_mul_bitLength_le`	`Unrestricted/Parameters`
Rule U (Algorithm 6.2)	`roundQueries`, `completionQueries`, `program`	`Unrestricted/Rule`, `Unrestricted/Strategy`
Complementary groups (Lemma 6.3)	`getElem!_groups_le_of_complement`	`Unrestricted/Complementary`
Round (Lemma 6.4)	`round`	`Unrestricted/Round`

Paper item	Lean name	File
Contraction (Lemma 6.5)	`contraction`	`Unrestricted/Contraction`
One letter contracts often (Lemma 6.6)	`one_letter_contracts_often`	`Unrestricted/Contraction`
Completion (Lemma 6.7)	`eq_of_count_eq`	`Unrestricted/Completion`
Uniqueness (Proposition 6.8)	`cons_transcript_program`	`Unrestricted/Strategy`
Cost (proof of Theorem 1.2)	`cost_le`, `length_transcript_program`	`Unrestricted/Cost`, `Unrestricted/Strategy`
Theorem 1.2	`reconstructs`, `unrestricted_reconstruction`	`Unrestricted/Cost`, `Results`

The alphabet reduction uses the namespace `Lax100001Proofs.Alphabet`. The theorem `count_pairProjection` in `Alphabet` proves Lemma 7.1, and `pairwise_unique` proves Lemma 7.2. The rule composition in `AlphabetRule` simulates one binary run after another. In `AlphabetLift`, the theorems `liftRule_run`, `liftHistory_length` and `liftRule_allowed` prove the reconstruction, exact cost and query restriction of Theorem 7.3. The theorem `liftRule_envelope` gives Corollary 7.4 under its fixed-length uniform bound.