

**Lemma 1** (Constant optimizations and duplication effects for Lemma 4.1 of [hmp01]). Suppose  $(f, g)$  are each functions from  $[n]$  to  $\{0, 1\}^r$ . Let  $F_x = \{i \in [n] : f(i) = x\}$  and  $G_x = \{i \in [n] : g(i) = x\}$ . Let  $\mu : [n] \rightarrow [n]$  be a “deduplication” map, so that for all  $x, y \in \{0, 1\}^r$ ,  $\mu$  maps all elements of  $U_{xy} := \{i \in [n] : f(i) = x \wedge g(i) = y\}$  to a single arbitrary element of  $U_{xy}$ . Then in  $O(n \log n)$  deterministic time and  $O(n \log n)$  bits of space, one can construct  $d : \{0, 1\}^r \rightarrow \{0, 1\}^r$  for which, with function  $h(x) = g(x) \oplus d(f(x))$ , and  $H_x = \{i \in [n] : h(i) = x\}$ , we have:

1.  $\sum_{x \in \{0, 1\}^r} \binom{|H_x|}{2} - \sum_{i \in [n]} \frac{1}{|\mu^{-1}(i)|} \binom{|\mu^{-1}(i)|}{2} \leq \frac{1}{2^r} \binom{n}{2}$
2.  $\sum_{x \in \{0, 1\}^r} \binom{|H_x|}{2} - \sum_{i \in [n]} \binom{|\mu^{-1}(i)|}{2} \leq n \left\lfloor \frac{1}{2^r} \max \left( n - 1, \sum_{x: |F_x| \geq 2} |F_x|^2 \right) \right\rfloor$ .

*Proof.* This is derived from the proofs in Section 4 of [hmp01]. To construct  $d$ , select a permutation  $\pi : \{0, 1\}^r \rightarrow \{0, 1\}^r$  for which  $|F_{\pi(1)}| \geq |F_{\pi(2)}| \geq \dots \geq |F_{\pi(2^r)}|$ . (The last sets in the sequence will all be empty if  $n < 2^r$ .) Then in order, for each  $i \in [2^r]$ , choose  $d(\pi(i))$  to have value  $a \in \{0, 1\}^r$  so that the multiset  $S_{a,i} := (a \oplus g(j) : h \in F_{\pi(i)})$  has no more than the average number of collisions with preceding multisets  $\{S_{d(\pi(j)),j}\}_{j < i}$ . The number of collisions  $c(A, B)$  between two multisets  $A = (a_1, \dots, a_{|A|})$  and  $B = (b_1, \dots, b_{|B|})$  is defined as  $|\{x \in [|A|], y \in [|B|] : a_x = b_y\}|$ . Specifically, we want:

$$\begin{aligned} \sum_{j < i} c(S_{a,i}, S_{d(\pi(j)),j}) &\leq \left\lfloor \frac{1}{2^r} \sum_{b \in \{0, 1\}^r} \sum_{j < i} c(S_{b,i}, S_{d(\pi(j)),j}) \right\rfloor \\ &= \left\lfloor \frac{1}{2^r} \sum_{j < i} \sum_{b \in \{0, 1\}^r} c(S_{b,i}, S_{d(\pi(j)),j}) \right\rfloor \\ &= \left\lfloor \frac{1}{2^r} \sum_{j < i} |F_{\pi(i)}| |F_{\pi(j)}| \right\rfloor \end{aligned}$$

where the last step follows because for each  $x \in F_{\pi(i)}$ ,  $y \in F_{\pi(j)}$ , there is exactly one value of  $b \in \{0, 1\}^r$  for which  $b \oplus g(x) = d(\pi(j)) \oplus g(y)$ . This can be done (even with multi-sets!) using a dynamic search table structure as described in Section 4.3 of [hmp01].

The quantity  $\sum_{j < i} c(S_{a,i}, S_{d(\pi(j)),j})$  counts the total number of colliding pairs  $a, b \in [n]$  where  $f(a) \neq f(b)$  and  $h(a) = h(b)$ . Since  $g(i) = h(i) \oplus d(f(i))$ , the number of colliding pairs where  $a, b \in [n]$  satisfy  $f(a) = f(b)$  and  $h(a) = h(b)$  is equal to  $\sum_{i \in [n]} \binom{|\mu^{-1}(i)|}{2}$  (the number of collisions that  $(f, g)$  have.)

Consequently,

$$\begin{aligned} \sum_{x \in \{0,1\}^r} \binom{|H_x|}{2} - \sum_{i \in [n]} \frac{1}{|\mu^{-1}(i)|} \binom{|\mu^{-1}(i)|}{2} &= \sum_{i \in \{0,1\}^r} \sum_{j < i} c(S_{a,i}, S_{d(f(j)),j}) \\ &\leq \sum_{i \in \{0,1\}^r} \left\lfloor \frac{1}{2^r} \sum_{j < i} |F_{\pi(i)}| |F_{\pi(j)}| \right\rfloor \end{aligned}$$

There are two ways to bound this. First,

$$\begin{aligned} \sum_{i \in \{0,1\}^r} \left\lfloor \frac{1}{2^r} \sum_{j < i} |F_{\pi(i)}| |F_{\pi(j)}| \right\rfloor &\leq \frac{1}{2^r} \sum_{\{i,j\} \in \binom{\{0,1\}^r}{2}} |F_{\pi(i)}| |F_{\pi(j)}| \\ &\leq \frac{1}{2^r} \cdot \frac{1}{2} \left( \sum_{i,j \in \{0,1\}^r} |F_{\pi(i)}| |F_{\pi(j)}| - \sum_{i \in [n]} |F_{\pi(i)}|^2 \right) \\ &\leq \frac{1}{2^r} \cdot \frac{n^2 - n}{2} = \frac{1}{2^r} \binom{n}{2} \end{aligned}$$

This bound does *not* use the permutation sort order; the following one does (and needs it, when  $(F_{\pi(i)})_{i \in [n]}$  looks like  $\sqrt{n}, \sqrt{n}, 1, 1, 1, \dots, 1$ ). Specifically:

$$\begin{aligned} \sum_{i \in \{0,1\}^r} \left\lfloor \frac{1}{2^r} \sum_{j < i} |F_{\pi(i)}| |F_{\pi(j)}| \right\rfloor &\leq n \max_{i \in \{0,1\}^r} \left\lfloor \frac{1}{2^r} \sum_{j < i} |F_{\pi(i)}| |F_{\pi(j)}| \right\rfloor \\ &\leq n \max_{i \in \{0,1\}^r} \begin{cases} \left\lfloor \frac{1}{2^r} \sum_{j < i} |F_{\pi(j)}|^2 \right\rfloor & \text{if } |F_{\pi(i)}| \geq 2 \\ \left\lfloor \frac{1}{2^r} \sum_{j < i} |F_{\pi(j)}| \right\rfloor & \text{if } |F_{\pi(i)}| \leq 1 \end{cases} \\ &\leq n \max_{i \in \{0,1\}^r} \begin{cases} \left\lfloor \frac{1}{2^r} \sum_{x: |F_x| \geq 2} |F_x|^2 \right\rfloor & \text{if } |F_{\pi(i)}| \geq 2 \\ \left\lfloor \frac{1}{2^r} (n-1) \right\rfloor & \text{if } |F_{\pi(i)}| \leq 1 \end{cases} \\ &\leq n \left\lfloor \frac{1}{2^r} \max \left( n-1, \sum_{x: |F_x| \geq 2} |F_x|^2 \right) \right\rfloor \end{aligned}$$

□

**Lemma 2** (Deterministic double displacement.). *Applying Lemma 1 twice, with  $r = \lceil \log_2(\alpha n) \rceil$ , gives a perfect hash function mapping  $n$  unique pairs  $(f_i, g_i)_{i=1}^n$  to values  $\lambda_i \in \{0,1\}^r$ , when  $\alpha \geq \sqrt{2}$ .*

*Proof.* First, apply Lemma 1 to  $(f_i, g_i)_{i=1}^n$ , producing  $(h_i)_{i=1}^n$  with each  $h_i \in \{0,1\}^r$  satisfying  $h_i = g_i \oplus d_{1,i}(f_i)$  for some displacement function  $d_1$  from  $\{0,1\}^r \rightarrow \{0,1\}^r$ . Then with  $H_x := \{i \in [n] : h_i = x\}$  as defined in Lemma 1, we have  $\sum_{x \in \{0,1\}^r} \binom{|H_x|}{2} \leq \frac{1}{2^r} \binom{n}{2}$ . Next, apply to Lemma 1 to  $(h_i, f_i)_{i=1}^n$ ,

producing  $(\lambda_i)_{i=1}^n$  with each  $\lambda_i \in \{0, 1\}^r$  satisfying  $\lambda_i = f_i \oplus d_{2,i}(h_i)$  for some displacement function  $d_2 : \{0, 1\}^r \rightarrow \{0, 1\}^r$ . Define  $\Lambda_x := \{i \in [n] : \lambda_i = x\}$ . Then:

$$\frac{1}{2^r} \binom{n}{2} \geq \sum_{x \in \{0,1\}^r} \binom{|H_x|}{2} \geq \frac{1}{4} \sum_{x \in \{0,1\}^r : |H_x| \geq 2} |H_x|^2$$

so by the second bound in Lemma 1:

$$\begin{aligned} \sum_{x \in \{0,1\}^r} \binom{|\Lambda_x|}{2} &\leq n \left\lfloor \frac{1}{2^r} \max \left( n-1, 4 \frac{1}{2^r} \binom{n}{2} \right) \right\rfloor \\ &= n \left\lfloor \frac{2}{2^{2r}} n(n-1) \right\rfloor && \text{if } 2^r \geq n \\ &= 0 && \text{if } 2^r \geq n\sqrt{2} \end{aligned}$$

□

*Note 3.* The first and second bounds of Lemma 1 do not fit together when  $i \mapsto (f(i), g(i))$  is not one-to-one. It is *possible* that, when  $\sum_{x \in \{0,1\}^r} \binom{|F_x|}{2} - \sum_{i \in [n]} \frac{1}{|\mu^{-1}(i)|} \binom{|\mu^{-1}(i)|}{2} \leq n$ , the bound  $\sum_{x \in \{0,1\}^r} \binom{|H_x|}{2} - \sum_{i \in [n]} \frac{1}{|\mu^{-1}(i)|} \binom{|\mu^{-1}(i)|}{2} = 0$  holds, but proving or disproving this may require going into the details of the search table procedure. (If there is a hard instance, it might have each nonempty multiset  $F_x$  contain two distinct  $g$  values (possibly with duplicates) structured to trick the search procedure into using a small branch of the table.)

Say that for some  $x$ ,  $F_x$  has  $k = |\mu(F_x)|$  equivalence classes by  $\mu$ , of sizes  $a_1, \dots, a_k$ , with all  $a_j \geq 1$ . Because  $\sum_{j \in [k]} (a_j - 1)^2 \leq \left( \sum_{j \in [k]} (a_j - 1) \right)^2$ :

$$\begin{aligned} &\binom{|F_x|}{2} - \sum_{i \in F_x} \frac{1}{|\mu^{-1}(i)|} \binom{|\mu^{-1}(i)|}{2} \\ &= \binom{|F_x|}{2} - \sum_{j \in [k]} \binom{a_j}{2} \\ &= \binom{|F_x|}{2} - \frac{1}{2} \sum_{j \in [k]} [(a_j - 1)^2 + (a_j - 1)] \\ &= \binom{|F_x|}{2} - \frac{1}{2} (|F_x| - k) - \frac{1}{2} \sum_{j \in [k]} (a_j - 1)^2 \\ &\geq \binom{|F_x|}{2} - \frac{1}{2} (|F_x| - k) - \frac{1}{2} (|F_x| - k)^2 \\ &= \binom{|F_x|}{2} - \binom{|F_x| - k}{2} \\ &= \frac{1}{2} (|F_x|^2 - |F_x| - (|F_x| - k)^2 + (|F_x| - k)) \\ &= \frac{k}{2} (2|F_x| - k - 1) \end{aligned}$$

Therefore, the nontrivial collision count  $\kappa$  satisfies:

$$\begin{aligned}\kappa &:= \sum_{x \in \{0,1\}^r} \binom{|F_x|}{2} - \sum_{i \in [n]} \frac{1}{|\mu^{-1}(i)|} \binom{|\mu^{-1}(i)|}{2} = \sum_{x \in \{0,1\}^r} \left( \binom{|F_x|}{2} - \sum_{i \in F_x} \frac{1}{|\mu^{-1}(i)|} \binom{|\mu^{-1}(i)|}{2} \right) \\ &\geq \sum_{x \in \{0,1\}^r} \frac{|\mu(F_x)|}{2} (2|F_x| - |\mu(F_x)| - 1)\end{aligned}$$

This can be used to bound the cost of *delayed* deduplication for the second displacement round. For example, for each  $F_x$ , one can by a variant of insertion sort construct a sorted list of unique elements in  $O(|F_x| |\mu(F_x)|)$  time, which summed over all  $x \in \{0,1\}^r$  is  $O(n)$ . Or the search table design can be modified so that, when it is time to update table frequencies after choosing  $d(\pi(i))$ , the leaves are updated to indicate just *whether* an element has been used, not *how many times* it has been used, and the remaining values derived from the leaves. Then the key quantity to bound would be:

$$\begin{aligned}\max_{i \in \{0,1\}^r} |F_{\pi(i)}| \sum_{j < i} |\mu(F_{\pi(j)})| &\leq \max \left( n-1, \sum_{x \in \{0,1\}^r: |F_x| \geq 2} |F_x| |\mu(F_x)| \right) \\ &\leq \max(n-1, 4\kappa)\end{aligned}$$

The last inequality follows because if  $|F_x| > |\mu(F_x)|$ , then  $|\mu(F_x)| (2|F_x| - |\mu(F_x)| - 1) \geq |F_x| |\mu(F_x)|$ , while if  $|F_x| = |\mu(F_x)|$  and  $|F_x| \geq 2$  then  $|\mu(F_x)| (2|F_x| - |\mu(F_x)| - 1) \geq \frac{1}{2} |\mu(F_x)| |F_x|$ .

**Definition 4.** Odd-multiply-shift (OMS) hashing. Used by [r95], proved 2-approximately universal by [dhkp97], with the hash function type known at latest since TAOCP Vol 3 in 1973. The hash family  $\mathcal{H}_{b,s}$  maps  $\{0, \dots, 2^u - 1\} \equiv \{0,1\}^u$  to  $\{0, \dots, 2^s - 1\} \equiv \{0,1\}^s$  and is parameterized by odd integers  $a \in \{0, \dots, 2^u - 1\}$ ; individual functions are given by

$$h_\alpha(x) = (ax \bmod 2^u) \operatorname{div} 2^{u-s}.$$

**Lemma 5.** For any distinct  $x, y \in \{0,1\}^u$ , with  $\alpha$  drawn uniformly at random from the odd integers in  $\{0, \dots, 2^u - 1\}$ , we have the following two bounds:

$$\begin{aligned}\Pr_a[h_a(x) = h_a(y)] &\leq \begin{cases} 0 & \text{if } (x - y \bmod 2^{u-s}) = 0 \\ 1 & \text{otherwise} \end{cases} \\ \Pr_a[h_a(x) = h_a(y)] &\leq \Pr_a[a(x - y) \bmod 2^u \operatorname{div} 2^{u-s} = 0]\end{aligned}$$

*Proof.* We have

$$\Pr_a[h_a(x) = h_a(y)] = \Pr_a[(ax \bmod 2^u) \operatorname{div} 2^{u-s} = (bx \bmod 2^u) \operatorname{div} 2^{u-s}]$$

Let  $i$  be the largest integer for which  $a \bmod 2^i = b \bmod 2^i$ . Decompose  $x = x_h 2^i + c$  and  $y = y_h 2^i + c$  where  $c = a \bmod 2^i = b \bmod 2^i$ . If  $i \geq u - s$ , then

$$\begin{aligned}
(ax \bmod 2^u) \operatorname{div} 2^{u-s} &= (ax_h 2^i + ac) \bmod 2^u \operatorname{div} 2^{u-s} \\
&= (((ax_h \bmod 2^{u-i}) 2^i + (ac \bmod 2^u)) \bmod 2^u) \operatorname{div} 2^{u-s} \\
&= \left( (ax_h \bmod 2^{u-i}) 2^{i-(u-s)} + (ac \bmod 2^u \operatorname{div} 2^{u-s}) \right) \bmod 2^{u-s} \\
&= \left( (ax_h \bmod 2^{u-i}) 2^{i-(u-s)} + d \right) \bmod 2^{u-s}
\end{aligned}$$

where  $d = (ac \bmod 2^u \operatorname{div} 2^{u-s})$ . Since  $a$  is odd, for  $z \in \{0, \dots, 2^{u-i}\}$ , the map  $z \rightarrow (az \bmod 2^{u-i})$  is a bijection; and since  $x_h \neq y_h$ , it follows  $h_a(x) \neq h_a(y)$ .

On the other hand, if  $i < u - s$ , then

$$\begin{aligned}
&\Pr_a [(ax \bmod 2^u) \operatorname{div} 2^{u-s} = (ay \bmod 2^u) \operatorname{div} 2^{u-s}] \\
&\leq \Pr_a [((ax \bmod 2^u) - (ay \bmod 2^u)) \operatorname{div} 2^{u-s} = 0] \\
&\quad + \Pr_a [(bx \bmod 2^u) - (ay \bmod 2^u)) \operatorname{div} 2^{u-s} = 0] \\
&= \Pr_a [a(x - y) \bmod 2^u \operatorname{div} 2^{u-s} = 0] \\
&\quad + \Pr_a [b(x - y) \bmod 2^u \operatorname{div} 2^{u-s} = 0]
\end{aligned}$$

The first inequality follows since  $w \operatorname{div} 2^{u-s} = z \operatorname{div} 2^{u-s}$  implies both  $w$  and  $z$  lie in the same set  $\{2^{u-s}k, \dots, 2^{u-s}(k+1) - 1\}$ ; if  $w \leq z$ , then  $(z - w) \bmod 2^u < 2^{u-s}$  so  $(z - w) \bmod 2^u \operatorname{div} 2^{u-s} = 0$ , while if  $w \geq z$ , then  $(w - z) \bmod 2^u \operatorname{div} 2^{u-s} = 0$ . Note that  $(x - y) \bmod 2^u = ((x_h - y_h) 2^i) \bmod 2^u$  and  $x_h - y_h \bmod 2 = 1$ . Multiplying by  $x_h - y_h$  under  $\bmod 2^{u-i}$  permutes odd values in  $\{0, \dots, 2^{u-i} - 1\}$ , so  $(a(x_h - y_h) 2^i) \bmod 2^u$  is uniformly distributed over *odd* multiples of  $2^i$  in  $2^u$ . Since  $i < u - s$ , and  $\{0, \dots, 2^{u-s}\}$  contains  $2^{u-s-i-1}$  odd multiples of  $2^i$  (out of  $2^{u-i-1}$  possible values).

$$\Pr [(a(x - y) \bmod 2^u) \in \{0, \dots, 2^{u-s}\}] = \frac{1}{2^s}.$$

The  $\Pr_a [h_a(x) = h_a(y)] \leq \frac{1}{2^{s-1}}$  upper bound is tight. For example, with  $u = 4, s = 3$ ,  $\Pr [h_a(1) = h_a(y)] = \frac{1}{4}$  holds  $y = 3$  and  $y = 11$ .  $\square$

*Note 6.* For comparison, the upper bound from [r95] on  $\Pr_a [h_a(x) = h_a(y)]$  is

$$\Pr_a [a(x - y) \bmod 2^u \in \{0, \dots, 2^{u-s} - 1\} \cup \{2^u - 2^{u-s} + 1, \dots, 2^{u-1}\}]$$

for which a  $\leq \frac{1}{2^{s-1}}$  upper bound also holds. The bound of Lemma 5 has the advantage of being slightly easier to compute in the incremental  $a$ -finding hash construction approach of [r95]. Also, either bound can be used to find that the expected number of collisions when uniformly randomly hashing an  $n$ -element set  $n$  from  $\{0, 1\}^s$  to  $\{0, 1\}^b$  is  $\leq \frac{1}{2^{s-1}} \binom{n}{2}$ . Consequently, for a perfect hash

function from  $\mathcal{H}_{b,s}$  to exist for all sets of size  $n$ , one needs  $\frac{1}{2^{s-1}} \binom{n}{2} < 1$  which holds iff  $s \geq \lceil \log(n(n-1) + 1) \rceil$ .

However, this is *not* a necessary condition, and one can do slightly better in some cases. When  $n = 1$ ,  $2^{\lceil \log(n(n-1)+1) \rceil} = 1$ , but when  $n = 2$ ,  $2^{\lceil \log(n(n-1)+1) \rceil} = 4$ . However, given any  $x, y \in \{0, 1\}^u$ , let  $j \in \{0, \dots, u-1\}$  be the highest bit for which  $x_j \neq y_j$ . If  $j = u-1$ , then  $h_1(x) = x_j \neq y_j = h_1(y)$  and choosing  $a = 1$  separates the two. Otherwise, let  $a = 2^{u-1-j}$  (which is admittedly *even*); then  $ax = 2^{u-1-j}x + x$ ; the multiplication does not overflow  $2^u$  so  $ax \div 2^{u-1} = x_j$ , and similarly  $ay \div 2^{u-1} = y_j$ , so  $h_a(x) \neq h_a(y)$ . Thus a perfect hash function for  $n = 2$  with  $s = 1$  always exists. It *may* be possible that similar minimal perfect hashing can be done with odd multiply-shift hashes for  $n = 2, 3, 4, \dots$  up to some small set size threshold, although space lower bounds imply the threshold is at most  $O(\log u)$ .

**Definition 7.** A hash function  $h$  from  $A$  to  $B$  is  $c$ -colliding on a set  $S \subseteq A$  of size  $n$  if  $\sum_{\{x,y\} \in \binom{S}{2}} 1_{h(x)=h(y)} \leq c$ .

**Lemma 8.** Say that  $h : A \rightarrow B$  is  $c$ -colliding on the set  $S$ . Let  $b_1, \dots, b_k$  be the sizes of the equivalence classes of  $S$  under  $h$ , so that  $\sum_i b_i = n = |S|$  and  $\sum_i \binom{b_i}{2} \leq c$ . Let

$$s(b) = \begin{cases} b & \text{if } b \leq 2 \\ 2^{\lceil \log(b_i(b_i-1)+1) \rceil} & \text{if } b \geq 3 \end{cases}$$

Then

$$\sum_i s(b_i) < n + 4c$$

$$\max \left( n, \sum_{b_i \geq 2} s(b_i) \right) \leq \max(n, 4c)$$

(These give bounds on the total output sizes for the FKS two-level hashing scheme ([fks84]) if placing leaf hash tables disjointly using offset values, and if packing size-1 tables into the gaps of the other tables. Size 2-tables are also contiguous, so one may be able to pack those into the  $\max(\frac{1}{2}s(b) - b, 0)$  gaps of length 2 left when hashing  $b$ -sized equivalence classes; to what extent this works depends on the exact equivalence class size distribution.)

*Proof.* Let  $r(b) = \frac{s(b)}{b(b-1)/2}$  (i.e, the ratio of table size to collision count for an equivalence class). We have  $r(1) = \infty$ ,  $r(2) = 2$ ,  $r(3) = r(4) = \frac{8}{3}$ ,  $r(5) = \frac{16}{5}$ , and  $\sup_{b \geq 6} r(b) = 4$  (although *exactly* 4 is never reached; in the limit,  $r(b)$  oscillates between 2 and 4 depending on how much the  $\lceil \cdot \rceil$  operation in  $s(b)$  adds.)

$$\sum_{b_i \geq 2} s(b_i) < 4 \sum_{b_i \geq 2} \frac{b_i(b_i-1)}{2} \leq 4c$$

Since  $s(1) = 1$ ,  $\sum_{b_i=1} s(b_i) = |\{i : b_i = 1\}| \leq n$ , so

$$\sum_i s(b_i) = \sum_{b_i=1} s(b_i) + \sum_{b_i=2} s(b_i) \leq n + 4c$$

(This is roughly tight; if all collisions are concentrated in a single equivalence class of size  $d = O(\sqrt{c})$ , then the large class will contribute  $s(d) = \Theta(c)$  while the small ones add  $n - d$  for the singleton equivalence classes.  $\square$ )

*Remark 9.* If the FKS hashing scheme ([fks84]) is implemented using OMS hashing (Definition 4), with hash functions picked to have a sub-average number of collisions, using size-2 leaf tables for size-2 leaves, tightly packing size-1 sub-tables in the free space of leaf tables, using  $\alpha$  and  $\beta$  bits per primary and secondary table entries, and with main hash output bit count  $s$ , the total space usage is

$$\alpha 2^s + \beta \max \left( n, \frac{4}{2^{s-1}} \binom{n}{2} \right)$$

which is minimized when

$$s = \min \left( \lceil \log(n-1) \rceil, \left\lceil \frac{1}{2} \log \left( \frac{4\beta \binom{n}{2}}{\alpha} \right) \right\rceil \right)$$

(Choosing  $s = \lceil \frac{1}{2} \log(\frac{x}{2}) \rceil$  minimizes  $2^s + \frac{x}{2^s}$  with value  $\leq \frac{3}{\sqrt{2}}\sqrt{x}$ , so the total space usage for the table is  $\leq \max \left( \frac{6\sqrt{\beta\alpha}}{\sqrt{2}}n, \beta n + 2\alpha n \right)$ .)

## References

- [fks84] Fredman, Komlós, Szemerédi, “Storing a Sparse Table with  $O(1)$  Worst-case Access Time”, 1984. <https://doi.org/10.1145/828.1884>
- [hmp01] Hagerup, Miltersen, Pagh, “Deterministic Dictionaries”, 2001, <https://doi.org/10.1006/jagm.2001.1171>.
- [r95] Raman, “Improved data structures for predecessor queries in integer sets”, 1995, technical report.
- [dhkp97] Dietzfelbinger, Hagerup, Katajainen, Penttonen, 1997, <https://doi.org/10.1006/jagm.1997.0873>