Lemma 1 (Constant optimizations and duplication effects for Lemma 4.1 of [1])**.** *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 $\mu : [n] \to [n]$ be a "deduplication" map, so that for all $x, y \in \{0,1\}^r$, μ 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 $d : \{0,1\}^r \to \{0,1\}^r$ for which, with function $h(x) = g(x) \oplus d(f(x))$, and $H_x = \{i \in [n] : h(i) = x\}$, we have:*

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

Proof. This is derived from the proofs in Section 4 of [1]. To construct d, select a permutation $\pi : \{0,1\}^r \to \{0,1\}^r$ for which $|F_{\pi(1)}| \ge |F_{\pi(2)}| \ge \ldots \ge |F_{\pi(2^r)}|$. (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 $S_{a,i} := \left(a \oplus g(j) : h \in F_{\pi(i)}\right)$ has no more than the average number of collisions with preceding multisets $\left\{S_{d(\pi(j)),j}\right\}_{j<i}$. The number of collisions $c(A, B)$ between two multisets $A = \left(a_1, \ldots, a_{|A|}\right)$ and $B = \left(b_1, \ldots, b_{|B|}\right)$ is defined as $|\{x \in [|A|], y \in [|B|] : a_x = b_y\}|$. Specifically, we want:

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

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 [1].

The quantity $\sum_{j<i} c\left(S_{a,i}, S_{d(\pi(j)),j}\right)$ counts the total number of colliding pairs $a, b \in [n]$ where $f(a) \ne f(b)$ 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,

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

There are two ways to bound this. First,

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

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

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

□

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

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 $\{0,1\}^r \to \{0,1\}^r$. 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 $d_2 : \{0,1\}^r \to \{0,1\}^r$. Define $\Lambda_x := \{i \in [n] : \lambda_i = x\}$. Because $n \leq 2^r$,

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

so by the second bound in Lemma 1:

$$\sum_{x\in\{0,1\}^r} \binom{|\Lambda_x|}{2} \leq n \left\lfloor \frac{1}{2^r} \max(n-1, n-1) \right\rfloor = 0$$

because $2^r \geq n$. (Note: the proof also works if one sets $r = \lceil \log_2 n \rceil$ for the first round and $r = \lceil \log_2 n \rceil + 1$ for the second, ensuring $\sum_{x \in \{0,1\}^r : |H_x| \geq 2} |H_x|^2 \leq 2(n-1)$ and $\sum_{x \in \{0,1\}^r} \binom{|\Lambda_x|}{2} = 0$; although doing this requires that all g_i are in $\{0,1\}^{\lceil \log_2 n \rceil}$.) □

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 $\sum_{x \in \{0,1\}^r} \binom{|F_x|}{2} - \sum_{i \in [n]} \binom{|\mu^{-1}(i)|}{2} \leq n$, the bound $\sum_{x \in \{0,1\}^r} \binom{|H_x|}{2} - \sum_{i \in [n]} \binom{|\mu^{-1}(i)|}{2} = 0$ 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.)

One approach to work around this issue is to use different search table design in the second phase, which at the leaves of the dynamic frequency table only counts *whether* an element has been used, not *how many times* it has been used; and derives inner node values from the leaves. Then we'd only need $|F_{\pi(i)}| \sum_{j<i} |\mu(F_{\pi(j)})| \leq 2^r$, which might be more easily provable when $\sum_{x \in \{0,1\}^r} \binom{|F_x|}{2} - \sum_{i \in [n]} \binom{|\mu^{-1}(i)|}{2} \leq n$.

References

[1] Hagerup, Miltersen, Pagh, "Deterministic Dictionaries", 2001, https://doi.org/10.1006/jagm.2001.1171.