Tight Cell-Probe Lower Bounds for Dynamic Succinct Dictionaries

Tianxiao Li [*]   Jingxun Liang [†]   Huacheng Yu [‡]   Renfei Zhou [§]

Abstract

A dictionary data structure maintains a set of at most n keys from the universe $[U]$ under key insertions and deletions, such that given a query $x \in [U]$, it returns if x is in the set. Some variants also store values associated to the keys such that given a query x, the value associated to x is returned when x is in the set.

This fundamental data structure problem has been studied for six decades since the introduction of hash tables in 1953. A hash table occupies $O(n \log U)$ bits of space with constant time per operation in expectation. There has been a vast literature on improving its time and space usage. The state-of-the-art dictionary by Bender, Farach-Colton, Kuszmaul, Kuszmaul and Liu [BFCK+22] has space consumption close to the *information-theoretic optimum*, using a total of

$$\log \binom{U}{n} + O(n \log^{(k)} n)$$

bits, while supporting all operations in $O(k)$ time, for any parameter $k \leq \log^* n$. The term $O(\log^{(k)} n) = O(\underbrace{\log \cdots \log n}_{k})$ is referred to as the *wasted bits per key*.

In this paper, we prove a matching *cell-probe* lower bound: For $U = n^{1+\Theta(1)}$, any dictionary with $O(\log^{(k)} n)$ wasted bits per key must have expected operational time $\Omega(k)$, in the cell-probe model with word-size $w = \Theta(\log U)$. Furthermore, if a dictionary stores values of $\Theta(\log U)$ bits, we show that *regardless of the query time*, it must have $\Omega(k)$ expected update time. It is worth noting that this is the first cell-probe lower bound on the trade-off between space and update time for general data structures.

---

[*]Institute for Interdisciplinary Information Sciences, Tsinghua University. `litx20@mails.tsinghua.edu.cn`.
[†]Institute for Interdisciplinary Information Sciences, Tsinghua University. `liangjx20@mails.tsinghua.edu.cn`.
[‡]Department of Computer Science, Princeton University. `yuhch123@gmail.com`.
[§]Institute for Interdisciplinary Information Sciences, Tsinghua University. `zhourf20@mails.tsinghua.edu.cn`.

1 Introduction

A dictionary data structure dynamically maintains a set of at most n keys from the universe $[U]$ under key insertions and deletions (the updates), such that given a query $x \in [U]$, it returns if x is in the set. Some variants of dictionaries also maintain a set of key-value pairs with distinct keys and with values from $[V]$, such that given a query x, it further returns the value associated to x when key x is in the set.

Hash tables are classic designs of dictionaries. Using universal hash functions and chaining, hash tables have constant expected update and query times while occupying $O(n \log U)$ bits of space. There has been a vast literature on dictionaries improving the time and space usage [Knu73, ANS09, ANS10, BCFC+21, DdHPP06, DKM+88, DH90, FPSS03, FKS84, Knu63, LYY20, Pag01, PR04, Păt08, RR03, Yu20, BFCK+22]. An ideal dictionary would use nearly information-theoretically optimal space, $\approx \log \binom{U}{n}$ bits, and could process each operation in constant time.

The state-of-the-art dictionary by Bender, Farach-Colton, Kuszmaul, Kuszmaul and Liu [BFCK+22] uses

$$\log \binom{U}{n} + O(n \log^{(k)} n)$$

bits of space, while supporting each operation in $O(k)$ time for any parameter $k \in [\log^* n]$. The term

$$O(\log^{(k)} n) = O(\underbrace{\log \log \cdots \log}_{k} n)$$

is referred to as the *wasted bits per key*, and $O(n \log^{(k)} n)$ is referred to as the *redundancy*. This interesting time-space trade-off may not look natural at the first glance, e.g., when $O(1)$ wasted bits per key are allowed, the data structure has operational time $O(\log^* n)$. Surprisingly, they proved that for any hash table that makes use of "augmented open addressing," which includes all known *dynamic* succinct dictionaries [RR03, ANS10, BFCG+18, BE20, LYY20, BCFC+21, BFCK+22], this trade-off is optimal!

This leads to the question of how general "augmented open addressing" is, and whether one can design a dynamic dictionary that does not fall in this category and has a better time-space trade-off. For example, the best-known *static* dictionary (i.e., a data structure that only needs to support queries) [Yu20], which achieves $O(1)$ expected query time with $O(n^\varepsilon)$ redundancy, does not use augmented open addressing.

In this paper, we show that this is indeed the best possible for *dynamic* dictionaries, by proving a matching *cell-probe* lower bound.[1]

Theorem 1.1. *For $U = n^{1+\Theta(1)}$ and $k \leq \log^* n$, any dynamic dictionary storing at most n keys from $[U]$ with $O(\log^{(k)} n)$ wasted bits per key must have expected insertion, deletion or query time at least $\Omega(k)$, in the cell-probe model with word-size $w = \Theta(\log U)$.*

It is worth noting that when U/n is slightly sub-polynomial, e.g., $U = n^{1+1/\log^{(k)} n}$ for some constant k, Bender et al. [BFCK+22] proposed another data structure with $o(1)$ wasted bits per key supporting constant-time insertions, deletions and queries. Thus, the requirement that U needs to be at least $n^{1+\Omega(1)}$ can barely be relaxed.

Our proof uses a framework similar to the *information transfer tree* argument [PD04b], which is now a widely-used technique for proving cell-probe lower bounds [PD04a, PD06, CJ11, CJS15, CJS16, Yu16, WY16, AWY18, LN18, BHN19, JLN19, LMWY20]. Roughly speaking, it builds a

---

[1] In the cell-probe model [Yao78], it takes unit cost to read or write (probe) one memory cell of w bits, and the computation is free. Thus, a cell-probe lower bound implies the same time lower bound for RAM.

tree on top of a sequence of n operations, and associates each cell-probe to an internal node of the tree. The key step of the proof is to lower bound the number of cell-probes associated to each node. In all prior work using this framework, this is done by reasoning about how the updates and queries in different subtrees must interact (i.e., to answer a query, the data structure must learn sufficient information about previous updates), thus proving a trade-off between update time and query time. In our proof, we are able to reason, via a novel argument, about how *space constraints* can force the operations to spend cell-probes, thus proving a time-space lower bound.

In fact, our new technique already gives an arguably simpler and more intuitive proof of the lower bound against data structures using augmented open addressing that Bender et al. proved. We will present an overview of this simplified proof in the next section as a warm-up.

The technique also allows us to extend the lower bound to $o(n)$ redundancy.

Theorem 1.2. *For $U = n^{1+\Theta(1)}$, any dynamic dictionary storing at most n keys from $[U]$ with $R < n$ bits of redundancy must have expected insertion, deletion or query time at least $\Omega(\log(n/R))$, in the cell-probe model with word-size $w = \Theta(\log U)$.*

Furthermore, if the keys are associated with values of $\Theta(\log U)$ bits, the same technique proves that even if the queries are allowed to take arbitrarily long time, the updates must still follow the same lower bound.

Theorem 1.3. *Consider a dynamic dictionary storing at most n keys from $[U]$, each associated with a value in $[V]$, with R bits of redundancy, where $U \geq 3n$ and $V = U^{2+\Theta(1)}/n^2$. Then, in the cell-probe model with word-size $w = \Theta(\log U)$,*

1. *if $R \geq n$ can be written as $R = O(n \log^{(k)} n)$ for $k \leq \log^* n$, then the expected update time is $\Omega(k)$;*

2. *if $R < n$, then the expected update time is $\Omega(\log(n/R))$.*

This lower bound shows the dictionary by Bender et al. (which can support values) is optimal in a very strong sense: their data structure achieves $O(k)$ *update time* in worst-case except with inverse polynomial probability,[2] $O(1)$ *query time* in worst-case, and it is dynamically resizable;[3] we show that even if we relax all other properties, the $O(k)$ update time is still not improvable just under the space constraint.

We emphasize that this is the first space-update trade-off lower bound in the cell-probe model for any data structure problem. Such lower bounds were proved in [LNN15] for streaming problems against *non-adaptive* update algorithms, which was later shown not to hold for general adaptive update algorithms [AY20]. In fact, by applying *global rebuilding* [Ove83], one cannot hope to prove such a trade-off when the redundancy is linear in the total space without proving a *super-linear RAM lower bound*, a notoriously hard question. Naively, a data structure can always use half of its memory as a buffer to store the unprocessed updates, and only batch-process it when it is full; if there is a batch-processing algorithm with linear time in *RAM*, then this data structure directly has amortized update time $O(1)$ – proving an $\omega(1)$ update time lower bound implies a RAM lower bound for batch processing. Furthermore, to achieve non-amortized update time $O(1)$, global rebuilding suggests building *two* buffers in half of the memory, and using them alternatingly: when one buffer is full, one could gradually execute the batch-processing algorithm during the next $O(n)$ updates, and use the other buffer to store them. Note that such a strategy is not applicable

---

[2] By simply rebuilding the whole data structure when an update takes more than $\Omega(k)$ time, the worst-case-with-high-probability bound implies $O(k)$ expected time.

[3] A dictionary is dynamically resizable if its space usage is in terms of the *current* data size.

in the *succinct* regime, i.e., where the redundancy is asymptotically $o(1)$-fraction of the total space – the size of the buffers cannot exceed the redundancy, thus we would have to flush the buffers too frequently.

With an easy adaption, our technique can also prove tight lower bounds for strongly history-independent dictionaries. A dictionary is said to be *strongly history-independent* (a.k.a. *uniquely representable*) if its memory state is fully determined by the set of elements in it together with some random bits the algorithm uses. We will prove the following result.

Theorem 1.4. *For $U = n^{1+\Theta(1)}$ and $R \geq 1$, any strongly history-independent dynamic dictionary storing at most n keys from $[U]$ with R bits of redundancy must have expected insertion, deletion or query time at least $\Omega\big(\log \frac{n \log U}{R}\big)$, in the cell-probe model with word-size $w = \Theta(\log U)$.*

This lower bound is tight for $R \geq n/\mathrm{poly}\log n$, due to the known data structures [Kus23, LLYZ23]. Note that this time-space trade-off is worse than the optimal trade-off for succinct dynamic dictionaries without the history-independent constraint. We remark that this is the first separation between a data structure problem and its strongly history-independent version under the RAM (or cell-probe) model.

Another related problem is *Stateless Allocation* [Gol08, NT01, GKMT17, BKP+20, Kus23]. It requires the algorithm to put at most $(1 - \varepsilon)n$ elements from universe $[U]$ into n slots, where each slot can only accommodate one element, and εn slots are left empty. We require the mapping between elements and slots to only depend on the current element set and some random bits, i.e., the assignment is strongly history-independent. The performance of the algorithm is measured by its *switching cost*, the number of elements that changes its assigned slot during an insertion/deletion. Kuszmaul [Kus23] showed an algorithm with expected switching cost $O(\log \varepsilon^{-1})$ for $1/n \leq \varepsilon \leq 1$. We will prove that this upper bound is actually tight.

Theorem 1.5. *For $1/n \leq \varepsilon \leq 1$ and $U \geq 3n$, any stateless allocation algorithm that assigns at most $(1 - \varepsilon)n$ elements from universe $[U]$ to n slots must have expected switching cost at least $\Omega(\log \varepsilon^{-1})$.*

2 Technical Overview

Now we present an overview of our proof technique. For simplicity, let us focus on the case with $O(n)$ bits of redundancy, and show a time lower bound of $\Omega(\log^* n)$. Let us also assume $U = V = n^3$.

2.1 Slot Model

As a "proof-of-concept," we first showcase a lower bound in the *slot model*, a generalization of augmented open addressing. Then we move on to discussing how to adapt the proof to the cell-probe model in Section 2.4, before presenting the formal proof in the later sections.

In the slot model, a data structure maintains a set of at most n keys (balls) from $[U]$ under key insertions and deletions, and maps the keys to n slots (bins) indexed by $[n]$, with at most one key in each slot at any point. The data structure must maintain at most $\log \binom{U}{n} + O(n)$ bits of memory, which encodes the set itself and $O(n)$ bits of (arbitrary) auxiliary information, and can also determine the location (i.e., corresponding slot) of each key. We say a key k "is in slot i" if it is mapped to slot i (it is irrelevant to how k is stored in the memory state). Each time we insert or delete a key, the data structure updates its memory and moves (or swaps) the keys between slots. In the slot model, moving a key from one slot to another takes $O(1)$ cost, while accessing the

3

memory is free. Thus, the goal is to minimize the key-moves during the insertions and deletions. We will be proving that each operation must move at least $\Omega(\log^* n)$ keys on average.

We note that the slot model is similar to, yet more general than, the augmented open addressing defined in [BFCK$^+$22]. Augmented open addressing has a list of "hash functions" $h_1, h_2, \ldots, h_n$ mapping $[U]$ to $[n]$. Key x must be put in slot $h_i(x)$ for some i, and the auxiliary information stores the index i using $O(\log i)$ bits.

One can either view a data structure in the slot model as a dictionary with values, where the balls represent the keys' associated values (which are physically stored in their corresponding slots), or view the "quotient" [Knu73, Pag99] as the balls in the key-only setting. It is more limited than general data structures as the values or quotients must be stored in n memory slots atomically, while it is also stronger in the sense that the set of keys is known for free and we only count the number of moves.

Hard distribution. Let us consider the following sequence of operations (similar to the hard instance analyzed in [BFCK$^+$22]): We initialize by inserting a set of n random keys, then repeatedly delete a random key from the initial set and insert a new random key from $[U]$. We call each pair of deletion and insertion after the initialization a *meta-operation*. The analysis will focus on the cost of the n meta-operations.

A simple upper bound. To motivate the quantity $\log^* n$, let us first consider the following algorithm that maintains the keys approximately in sorted order using lazy updates. After the initialization, the data structure spends $O(n)$ swaps to place the keys in sorted order in the n slots. Thus, the key set by itself determines the location of every key.

Next, for each meta-operation, after deleting a key and thus emptying a slot, we can first put the new key in the empty slot, and store its identity and location as auxiliary information. As we delete and insert more keys, the slots gradually become "less sorted." Since we only allow $O(n)$ bits of auxiliary information, this can be done as long as the new keys and their locations take at most $O(n)$ bits to store, i.e., we can store $O(n/\log n)$ new keys in this way.

After $O(n/\log n)$ meta-operations, we rearrange all *new keys* so that they become in sorted order *by themselves*, while the set of new keys still resides in the set of slots that contained deleted keys (see the first three steps in Figure 2.1 for an example). This takes $O(1)$ *amortized* swaps per insertion. It then suffices to only indicate as auxiliary information which keys are the new keys and the *set* of their locations. This takes $\log \binom{n}{O(n/\log n)} = O(n \log \log n / \log n)$ bits. After the rearrangement, every new key effectively only costs $O(\log \log n)$ auxiliary bits.

We do this to every segment of $O(n/\log n)$ consecutive meta-operations: store their identities and locations until it takes $O(n)$ bits; then rearrange so that they become in sorted order by themselves; and further store for every new key and every emptied slot, which segment we inserted the key or emptied the slot. Since every new key and slot needs $O(\log \log n)$ bits to encode, we can repeat this for $O(\log n / \log \log n)$ segments, i.e., $O(n/\log \log n)$ meta-operations in total. Thereafter, we further sort these $O(n/\log \log n)$ keys, as shown in the last step in Figure 2.1. Now we only need to pay $O\left(\log \binom{n}{n/\log \log n}\right) = O(n \log \log \log n / \log \log n)$ total auxiliary bits, or $O(\log \log \log n)$ auxiliary bits per key. Sorting again takes amortized $O(1)$ additional swaps per meta-operation.

This suggests the following strategy based on lazy updates: For each meta-operation, we can place the new key in the slot that just became empty and store its identity and location; after every segment of $O(n/\log n)$ consecutive meta-operations, we sort the new keys inserted in this segment
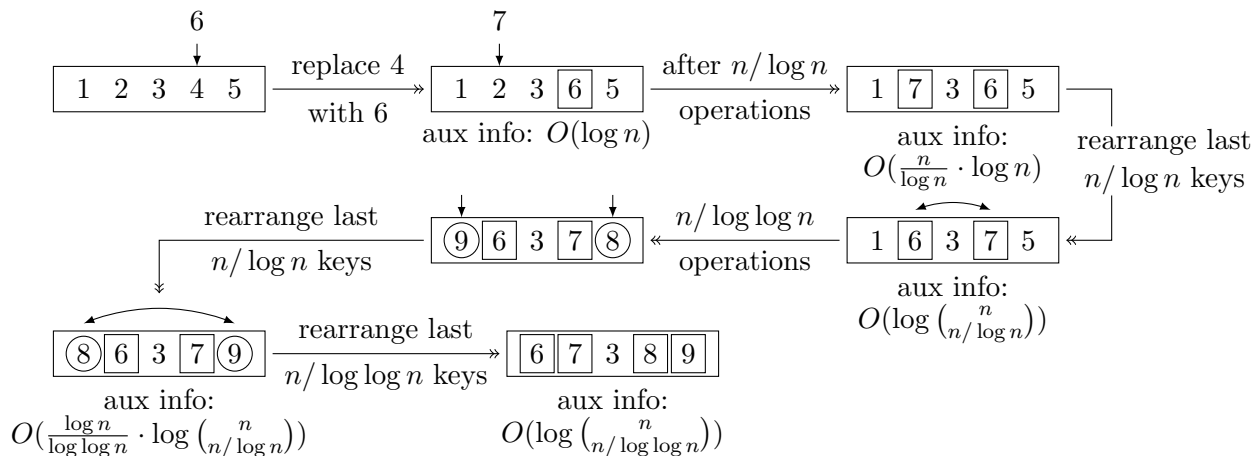
6
↓
| 1 | 2 | 3 | 4 | 5 | $\xrightarrow[\text{with } 6]{\text{replace } 4}$

7
↓
| 1 | 2 | 3 | 6 | 5 | $\xrightarrow[\text{operations}]{\text{after } n/\log n}$
aux info: $O(\log n)$

| 1 | 7 | 3 | 6 | 5 |
aux info:
$O(\frac{n}{\log n} \cdot \log n)$

rearrange last
$n/\log n$ keys

rearrange last
$n/\log n$ keys
| 9 | 6 | 3 | 7 | 8 | $\xleftarrow[\text{operations}]{n/\log\log n}$
| 1 | 6 | 3 | 7 | 5 |
aux info:
$O(\log \binom{n}{n/\log n})$

| 8 | 6 | 3 | 7 | 9 | $\xrightarrow[n/\log\log n \text{ keys}]{\text{rearrange last}}$ | 6 | 7 | 3 | 8 | 9 |
aux info:
$O(\frac{\log n}{\log\log n} \cdot \log \binom{n}{n/\log n})$

aux info:
$O(\log \binom{n}{n/\log\log n})$

In the first $n/\log n$ operations, the new keys are directly placed in the empty slots which are marked by squares (the first two steps), and will be rearranged later (the third step). This process continues for each subsequent group of $n/\log n$ operations, until the $(\log n/\log\log n)$-th group, which is marked by circles (the fourth step). After rearranging this group (the fifth step), we further rearrange all the $n/\log\log n$ new keys (the sixth step).

Figure 2.1: Key Arrangement

and store the set of locations as a batch[4]; after every $O(\log n/\log\log n)$ consecutive segments, we sort all new keys inserted in them and store the set of locations, and so on. This process continues until we have inserted $O(n)$ keys, in which case, we can afford to re-sort all keys with $O(1)$ additional swaps per key. There will be a total of $O(\log^* n)$ levels of sorting. Since sorting in each level takes $O(1)$ swaps per key, the total number of swaps for $O(n)$ meta-operations is $O(n\log^* n)$.

2.2 Slot Model Lower Bound

The slot model lower bound is inspired by the above algorithm. Note that in the algorithm, for every $O(n/\log n)$ consecutive meta-operations, we must rearrange the keys, otherwise the auxiliary information will take more than $O(n)$ bits. The rearrangement takes $O(1)$ cost per key. We will show that this $O(1)$-cost per key cannot be avoided.

We say a slot is *accessed* if we move a key from or to this slot. Formally, we will show that for every consecutive $m = \Omega(n/\log n)$ meta-operations, there must be $\Omega(m)$ times in expectation that a slot is accessed during multiple meta-operations. In other words, we go over all m meta-operations, and examine during each meta-operation, which slots are accessed *both in this and in a previous meta-operation*, and add the number of such slots for every meta-operation together. We assert that there must be $\Omega(m)$ times in expectation that a slot is accessed in a meta-operation and also in a previous one (thus, if the same slot is accessed during t different meta-operations, it is counted $t-1$ times).

Similarly, for every consecutive $m = \Theta(n/\log\log n)$ meta-operations, consisting of $\Theta(\log n/\log\log n)$ segments of size $\Theta(n/\log n)$, we assert that there must be $\Omega(m)$ times in expectation that a slot is accessed during multiple *segments*, and so on. We formally state it in the following lemma, which is the slot-model variant of Lemma 3.2.

---

[4]Strictly, we also store which keys belong to the new batch, whose space usage is no more than storing the set of locations.
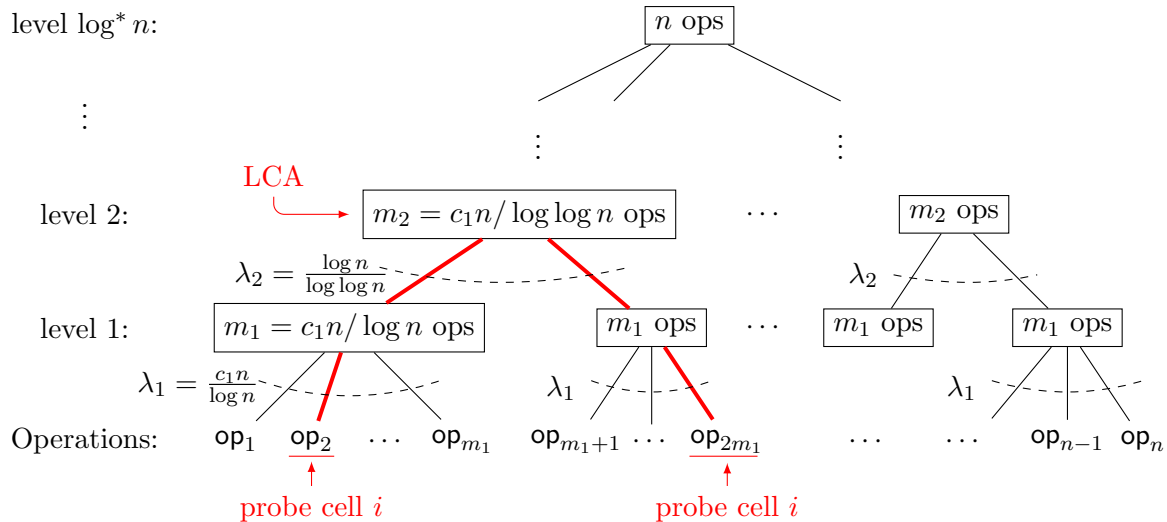
Figure 2.2: Tree Structure

Lemma. *Let $1 \leq \ell \leq \log^* n$. For any consecutive $m = c_1 \cdot n/\log^{(\ell)} n$ meta-operations for a sufficiently large constant c_1, partitioned into $\lambda = \Theta(\log^{(\ell-1)} n/\log^{(\ell)} n)$ segments of size $\Theta(n/\log^{(\ell-1)} n)$, there must be $\Omega(m)$ times that a slot is accessed during multiple segments.*

We used the convention that $\log^{(0)} n := n$. To see why it implies a lower bound of $\Omega(\log^* n)$, we build a tree on top of the n meta-operations with depth $\log^* n$. Each leaf of the tree (level-0 node) corresponds to one meta-operation, and we place the leaves in the same order as the meta-operations. Then we divide the leaves into groups of $\Theta(n/\log n)$ consecutive meta-operations, and assign a common parent (a new node) at level 1 to each group. Next, $\Theta(\log n/\log \log n)$ consecutive level-1 nodes are grouped together, and assigned a common parent at level 2, and so on. In general, we group $\Theta(\log^{(\ell-1)} n/\log^{(\ell)} n)$ nodes at level $\ell - 1$, and assign them a common parent at level ℓ. See Figure 2.2 for details.

Now consider an access to a slot during a meta-operation (a leaf), and consider when it was accessed the *previous time* (another leaf), we associate this access to the *lowest common ancestor* (LCA) of these two leaves. To lower bound the total number of accesses, it suffices to sum over all (internal) nodes the number of accesses associated to it. On the other hand, whenever an access is associated to a node u, then with respect to u, this is a slot that is accessed in two different segments (two different children of u). Hence, the above lemma lower bounds the number of accesses associated to each u by the size of the subtree rooted at u. Thus, the sum of every level is $\Omega(n)$, and the total sum over all internal nodes is $\Omega(n \log^* n)$. This way of counting the number of accesses has been used in many cell-probe lower bounds [PD04a, PD06, CJ11, CJS15, CJS16, Yu16, WY16, AWY18, LN18, BHN19, JLN19, LMWY20], although the nodes in the tree are usually set to the same degree, in which case, one obtains a logarithmic lower bound.

2.3 Number of Swaps per Node

Next, let us prove the above lemma. For simplicity, we first focus on $\ell = 1$, i.e., $m = c_1 \cdot n/\log n$ and "segments" consisting of one meta-operation ($\lambda = m$). To begin with, let us see what would go wrong if there was *zero* slot accessed during multiple meta-operations, then we will extend this proof-by-contradiction to $\Omega(m)$ slots.

6

Intuitively, we are going to show that the data structure stores too much information about the *order* of insertions (given the deletions) as auxiliary information, which corresponds to storing the new keys and locations in the upper bound. Fix a sequence of operations sampled from the distribution, and fix an index k. We will analyze the $(k+1)$-th to the $(k+m)$-th meta-operation.

The outer game. To formalize this intuition, let us consider the following communication game between two players Alice and Bob, which we call the *outer game*.

- Both players get the memory state C_{st} after the k-th meta-operation, of $\log \binom{U}{n} + O(n)$ bits. It determines the set of n initial keys K before the m meta-operations, as well as the slot allocation. The players also both know the keys that are deleted and the order of the deletions $\mathbf{d}$, and the (unordered) *set* of all inserted keys $\mathcal{A}$.

- Alice further knows the order of insertions $\mathbf{a}$, a permutation of $\mathcal{A}$.

- Alice sends one message to Bob, and their goal is for Bob to recover the permutation $\mathbf{a}$.

A communication protocol. Next, let us consider a protocol for the outer game. Alice's (only) message of this protocol will be an encoding of the memory state after the m meta-operations C_{end}. Since Bob knows the initial set K and the set of deleted keys and inserted keys, he also knows the set of keys K_{end} after the m meta-operations. On the other hand, C_{end} also determines K_{end} – the entropy of C_{end} *conditioned on* K_{end} is at most $O(n)$, since

$$H(C_{\mathrm{end}} \mid K_{\mathrm{end}}) = H(C_{\mathrm{end}}, K_{\mathrm{end}}) - H(K_{\mathrm{end}}) = H(C_{\mathrm{end}}) - H(K_{\mathrm{end}})$$

$$\leq \left(\log \binom{U}{n} + O(n) \right) - \log \binom{U}{n} = O(n).$$

Thus, by encoding C_{end} optimally *conditioned on* K_{end}, Alice can send $O(n)$ bits in expectation so that Bob is able to recover C_{end} from the message.

Now we claim that Bob can fully recover the permutation $\mathbf{a}$ if every slot is accessed in at most one meta-operation. Let us consider the first insertion. Bob knows the first deleted key d_1, but only knows that the first inserted key a_1 is among $\mathcal{A}$. Without loss of generality, we may assume that the data structure always first puts a_1 in the slot where d_1 was, then immediately applies an arbitrary permutation over the slots. Recall that any permutation can be decomposed into a collection of disjoint cycles: In particular, suppose d_1 (and a_1 temporarily) was in slot s_1, then there exist slots $s_2, \ldots, s_c$ for some $c \geq 1$ such that the key in s_1 is moved to s_2, the key in s_2 is moved to s_3, and finally, the key in s_c is moved to s_1. Now observe that all slots $s_1, \ldots, s_c$ are accessed during this meta-operation. By our assumption that every slot can be accessed in at most one meta-operation, these keys must remain in the same slots in the final state C_{end}. Next, since Bob knows d_1, he knows s_1, then he can examine which key x is in slot s_1 in C_{end}. If x is *not* in K, then he knows immediately that it must be a_1, as this slot is not accessed afterwards. Otherwise, x must be the key that was initially in s_c and was moved to s_1. Thus, Bob learns s_c by examining the location of x in initial memory state C_{st}. Then Bob can further examine which key is in s_c in C_{end}, and it must be the key that was initially in s_{c-1}. Bob can trace this cycle by alternatingly examining the keys in the slots in C_{st} and C_{end}. Eventually, he is able to identify a key that is not in K, which must be a_1.

After knowing a_1, Bob simply simulates the data structure for the first meta-operation by deleting d_1 and inserting a_1, then proceeds to the second meta-operation. Note that this process does not require any communication. Finally, Bob is able to recover the order of all insertions

7

after Alice sends the single message consisting of an encoding of C_{end}. The message has expected length $O(n)$, but the order of insertions has entropy $\Theta(m \log m)$, which yields a contradiction when $m = c_1 n / \log n$ for sufficiently large constant c_1.

Extending to $\Omega(m)$ cost and level $\ell > 1$. In general, if there are at most $0.01m$ times that a slot is accessed more than once, then except for at most $0.01m$ meta-operations, all slots accessed in other meta-operations are not accessed later. We then simply apply the above argument to iteratively figure out the remaining $0.99m$ insertions. For the other $0.01m$, Alice can simply send Bob the inserted keys within $\mathcal{A}$ using $\log m$ bits each (and $O(m)$ bits indicating which rounds Bob needs to run the above algorithm on his own). Alice's message has total length $0.01m \log m + O(n)$, which again yields a contradiction.

For levels $\ell > 1$, the players' goal becomes for Bob to figure out that, for each insertion in $\mathcal{A}$, which *segment* it belongs to (we give the order *within each segment* to both players). In this case, Bob needs to learn $m \log \lambda \approx c_1 n$ bits of information. It turns out that the same protocol works, and in each segment, by applying the same strategy more carefully, Bob can again find the inserted keys by following the cycles of the permutations. Thus, the lemma holds, and the lower bound in the slot model follows.

2.4 Cell-Probe Lower Bound

The lower bound in the cell-probe model is proved using the same framework. Instead of the slots, we will work with *memory cells*. In general, a data structure does not necessarily have one part to store the key set and auxiliary information, and another part for the slots. Moreover, (the "quotient" of) a key do not necessarily occupy a complete memory cell, and may be encoded arbitrarily. We need a more careful argument in this case.

Fix a data structure in the cell-probe model with word-size $w = \Theta(\log U)$ that uses $\log \binom{U}{n} + O(n)$ bits of space. We prove the following lemma.

Lemma. *Let $2 \le \ell \le \log^* n$. For any consecutive $m = c_1 \cdot n / \log^{(\ell)} n$ meta-operations for a sufficiently large constant c_1, partitioned into $\lambda = \Theta(\log^{(\ell-1)} n / \log^{(\ell)} n)$ segments of size $\Theta(n / \log^{(\ell-1)} n)$, there must be $\Omega(m)$ times that a memory cell is probed during multiple segments.*

By the same argument as in the slot model using trees, this lemma implies an $\Omega(\log^* n)$ cell-probe lower bound. It turns out that in order to prove it, the only part that we need to "upgrade" from the slot model is the step in the protocol of the outer game, where we showed that Bob can figure out which segment contains each insertion based (essentially) only on the starting and ending memory states. For the slot model, this step uses the fact that (the quotient of) the keys are atomic and can only be stored in slots and moved between the slots. For general data structures, we will have to do it differently.

Recovering the segments. Now let us fix $\ell \ge 2$ (for technical reasons, the proof does not work for $\ell = 1$). We again consider the outer communication game, where

- both players know the initial memory state C_{st}, the (unordered) set of m insertions $\mathcal{A}$, the (ordered) sequence of m deletions $\mathbf{d}$;

- Alice further knows the order of insertions $\mathbf{a}$;

- Bob only knows the order of insertions *within each segment* (e.g., Bob knows for each segment, the first insertion is the $\pi(1)$-th lexicographically smallest inserted key in this segment, the

8

second is the $\pi(2)$-th smallest inserted key, etc, but without knowing the actual set of inserted keys);

- their goal is to let Bob learn the set of inserted keys $A^{[i]}$ in each segment $i \in [\lambda]$.

For simplicity, we again will first derive a contradiction when *no* cell was probed in more than one segment. Similar to the slot model, Alice can send Bob the final memory state C_{end} using $O(n)$ bits. Bob will then first try to identify $A^{[1]}$, the insertions in the first segment. To this end, Bob decodes C_{end} from the message, then for all possible (m/λ)-sets of insertions $B \subset \mathcal{A}$, Bob pretends that $A^{[1]} = B$, and simulates the data structure on B together with the deletions in the first segment, which are known to Bob.

Now observe that a necessary condition for $A^{[1]} = B$ is that the set of cells probed in this simulation have the same new cell-contents as in C_{end}. This is because we assumed that no cell is probed in more than one segment, thus their contents must not be updated later. We call such a B *qualified*. Note that Bob can check if a set B is qualified, and clearly, the correct set of inserted keys must be qualified. Moreover, the key property we will prove is that, in expectation, only very few B are qualified. Thus, it suffices for Alice to send few bits to tell Bob which B is the correct $A^{[1]}$ *among the qualified sets*.

To prove this key property, consider the correct set of insertions $A = A^{[1]}$ and another qualified set B. Let S_A [resp. S_B] be the set of cells probed when $A^{[1]} = A$ [resp. $A^{[1]} = B$], and let C_A [resp. C_B] be the memory state after processing the first segment. Since A and B are both qualified, the contents of $S_A \cap S_B$ in C_A and C_B must be the same, because in particular, they are the same as in C_{end}. For such A and B, we say they are *consistent*. The following technical lemma says that two random subsets $A, B \subset \mathcal{A}$ of size m/λ with intersection at most $m/2\lambda$ are consistent with very low probability.

Lemma. *In expectation, two random m/λ-subsets A and B of $\mathcal{A}$ conditioned on $|A \cap B| \leq m/2\lambda$ are consistent with probability $U^{-\Omega(m/\lambda)}$.*

This is a simplified version of Lemma 4.4 to fit in the regime in this overview. Note that our distribution guarantees that $A^{[1]}$ is a random subset of $\mathcal{A}$ of size m/λ.

In particular, the above lemma implies that, on average, the sets that have a small intersection with the correct $A^{[1]}$ are very unlikely to qualify. On the other hand, there are only $\binom{m}{\leq m/2\lambda}\binom{m/\lambda}{\geq m/2\lambda} = O(\lambda)^{m/2\lambda}$ sets that have intersection size at least $m/2\lambda$ with A, while there are $\approx O(\lambda)^{m/\lambda}$ many m/λ-subsets of $\mathcal{A}$. Thus, by encoding the correct $A^{[1]}$ among the qualified sets, Alice can save a factor of two in the message length. This constant-factor saving in the communication cost turns out to already be sufficient for a contradiction.

The remaining step is to prove the above lemma, showing that two random sets conditioned on their intersection size being small are very unlikely to be consistent. For simplicity, let us consider two random sets A and B conditioned on $A \cap B = \varnothing$. If the contents of $S_A \cap S_B$ are the same after inserting either A or B, then intuitively, these cells should not be storing much useful information about A or B – we are wasting space proportional to $|S_A \cap S_B| \cdot w$. On the other hand, the *same* set D is deleted in the segment in both cases, one must probe the cells storing the information about D. This is because intuitively, the redundancy is only $O(n)$ bits, which is much smaller than the entropy of D (or A, B) when $\ell \geq 2$; the data structure must "empty most of the space storing D" to make room for A or B. Thus, the set $S_A \cap S_B$ should be sufficiently large to contain D. Combining the two, we are wasting $\Omega(H(D)) = \Omega(m/\lambda \cdot \log U) \gg \Omega(n)$ bits. This leads to a contradiction (this is also where we need $\ell \geq 2$).

We formally prove the lemma by studying another communication game, which we call the *inner game*. We show that whenever two sets A and B are consistent, it is possible to encode sets K, A, B, D (recall that K is the initial set of keys) using roughly $H(K, A, B, D) - \Omega(m/\lambda \cdot \log U)$ bits. This implies an upper bound on the possible number of such tuples (K, A, B, D), i.e., only $U^{-\Omega(m/\lambda)}$-fraction of A and B are consistent on average. See Section 5.1 for the protocol.

For the general case where A and B can have a small intersection, and at most $0.01m$ cells can be probed in more than one segment, we generalize the definition of consistency to require that the number of different cells in $S_A \cap S_B$ is bounded by $O(m/\lambda)$. It turns out that the above argument still goes through, and Bob can recover $A^{[1]}$ after Alice sends few extra bits. Then, Bob simulates the operations in the first segment, and the players apply the same strategy for Bob to iteratively recover $A^{[2]}, \ldots, A^{[\lambda]}$. Bob learns the partition of $\mathcal{A}$ into $A^{[1]}, \ldots, A^{[\lambda]}$, which has $\approx m \log \lambda$ bits of information, while Alice's message turns out to have only $\approx (1 - \Omega(1))m \log \lambda + O(n)$ bits, yielding a contradiction. See Section 4.2 and Section 4.3 for the detailed analysis.

3 Hard Distribution and Proof of Main Theorem

In the following sections, we present the formal proof of our lower bound for the dynamic succinct dictionary problem. Let C be a dynamic dictionary data structure that maintains a set T of at most n distinct keys in range $[U]$, i.e., C supports the following operations:

- Initialize T to $\varnothing$.

- Insert a key x, assuming $|T| < n$ and $x \notin T$.

- Delete a key x, assuming $x \in T$.

- Query whether x is in T.

We will focus on the following regime:

- We assume $U = n^{1+\Theta(1)}$. Without loss of generality, we set the word size $w = \log U$ so that a key can be stored in one word.

- The memory of C is fixed to be $\log \binom{U}{n} + R$ bits, where the term $\log \binom{U}{n}$ is the optimal space required to store n keys, and R is the redundancy. Moreover, by the convention of prior work, we say the data structure incurs $r = \frac{R}{n}$ *wasted bits per key*.

The running time of the dictionary is measured by the expected amortized number of cell-probes per operation, while the redundancy is measured by r, the number of wasted bits per key. As mentioned in Section 2, we will consider a sequence of n insertions followed by n *meta-operations*. Each meta-operation will consist of a query (for technical reasons), a deletion, followed by an insertion. A formal description of the hard distribution is presented in Distribution 1.

Distribution 1: Hard Distribution
---
1 Initialize an empty dictionary with capacity n and key-universe $[U]$
2 $K \leftarrow$ uniform random n-element subset of $[U]$
3 Insert keys in K into the dictionary one by one, using n insertions
4 for $i = 1$ *to* n **do**
5 $d_i \leftarrow$ a uniform random key in K that has not been removed
6   Query(d_i)
7   Delete(d_i) from the dictionary
8 $a_i \leftarrow$ a uniform random key in $[U]$ which is neither in K nor in the current dictionary
9   Insert(a_i) to the dictionary

The following theorem, a rephrase of Theorem 1.1 with respect to Distribution 1, is the main result of this section, proving a space-time lower bound for dynamic dictionaries.

Theorem 3.1. *For sufficiently large n and any positive integer $k \leq \log^* n$, if a dictionary incurs $r \leq \log^{(k)} n$ wasted bits per key, then it must perform $\Omega(nk)$ cell-probes to process a random sequence of operations sampled from Distribution 1 in expectation, i.e., the expected amortized time to process one operation is $\Omega(k)$.*

We first fix a data structure using $r \leq \log^{(k)} n$ wasted bits per key. By Yao's Minimax Principle, we may assume without loss of generality that it is deterministic.

Recall that we denote Step 5-9 in the hard distribution by a meta-operation. To prove the lower bound, we will build a tree over the sequence of n meta-operations. Each meta-operation is a leaf of the tree, which is also called a *level-0* node. The tree is then constructed bottom-up. After constructing level $\ell - 1$, we group every λ_ℓ *consecutive* level-$(\ell-1)$ nodes, and assign a level-ℓ node to be the parent of every group. Thus, a level-ℓ node u represents an interval of consecutive meta-operations, which is the union of the intervals of u's children.

Let m_ℓ denote the number of consecutive meta-operations a level-ℓ node represents. Then, we have $m_\ell = m_{\ell-1} \cdot \lambda_\ell$. We will set $m_\ell := cn \log^{(k)} n / \log^{(\ell)} n$ for levels $\ell \geq 1$, where $c = 10^6$ is a large constant, and set $m_0 = 1$ for the leaf nodes. For simplicity, we assume $m_{\ell-1} \mid m_\ell$. Therefore, for $\ell > 1$, we have $\lambda_\ell = \log^{(\ell-1)} n / \log^{(\ell)} n$.

The process of grouping level-$(\ell - 1)$ nodes to form level-ℓ nodes terminates at level h when $cn \log^{(k)} n / \log^{(h)} n \geq n$, i.e., all meta-operations are grouped into the same level-h node. Formally, $h = \min\{\ell \in \mathbb{N} : \log^{(\ell)} n \leq c \log^{(k)} n\}$. Here we let $m_h = n$, and the only level-h node is the root of the tree. Clearly, the height of the tree h is $\Omega(k)$. See Fig. 2.2 in Section 2.

Now let us return to the data structure. In the cell-probe model, its memory contains N cells, where $N = \left(\log \binom{U}{n} + R \right) \Big/ w = O(n)$. The addresses of these cells are from 1 to N. We will lower bound the total number of cell-probes in two ways. To this end, we introduce two quantities for each node u in the tree, cost_u and probe_u, to measure the number of probes.

For cost_u, if cell i is probed in meta-operations t_1 and t_2 $(t_1 < t_2)$ but not between them, we *assign* the probe of i at t_2 to node u, the *lowest common ancestor (LCA)* of t_1 and t_2 in the tree, and let cost_u be the total number of probes assigned to u. Clearly, since each cell-probe is assigned at most once, the total number of cell-probes during the whole operation sequence is at least the sum of cost_u over all the nodes.

For probe_u, we simply define it as the number of probes incurred while processing the operations in u's corresponding interval. For a single level ℓ, the sum of probe_u over all level-ℓ nodes u exactly equals the total time cost in the whole process (unlike cost_u, we cannot use the sum of probe_u over all nodes as a lower bound, because each cell-probe will be counted $h + 1$ times).

11

The lemma below gives a lower bound on the expectations of cost_u and probe_u.

Lemma 3.2 (Outer Lemma). *Let t be a parameter. Suppose $U = n^{1+\alpha}$ for a constant α, and let $\gamma = \frac{\alpha}{6(1+\alpha)}$. For a level-$\ell$ node u, if*

- $2^{64} \leq \lambda_\ell \leq \frac{\alpha}{12} \log n$,

- $m_\ell \log \lambda_\ell \geq 100R$,

- $\frac{1}{2}\gamma m_\ell \log U \geq R + 5m_\ell + 2(\beta + 1)m_\ell t$,

- $m_\ell \geq n^{1-\gamma/2}$,

where β is a global fixed constant to be determined later, then at least one *of the two proposition holds:*

- *Prop 1.* $\mathbb{E}[\mathsf{cost}_u] \geq \frac{\gamma}{100}m_\ell$;

- *Prop 2.* $\mathbb{E}[\mathsf{probe}_u] \geq \frac{1}{16}m_\ell t$.

Note that although we have set the tree parameters $\lambda_{\ell'}$, $m_{\ell'}$ above for all levels ℓ', Lemma 3.2 only focuses on a single level ℓ and has no constraint for the tree parameters for other levels $\ell' \neq \ell$. Hence, even if we use another set of tree parameters, Lemma 3.2 still applies as long as the required conditions are met (which will be the case in Section 6).

We will prove Lemma 3.2 in the next section. Now we use it to prove Theorem 3.1.

Proof of Theorem 3.1. The main idea is to apply Lemma 3.2 on all nodes in levels $2 \leq \ell < h$, with parameter $t = \frac{\gamma}{8(\beta+1)} \log U$. After applying the lemma, consider two cases:

- If there is a level ℓ $(2 \leq \ell < h)$ such that at least half of the nodes in this level satisfy *Prop 2*, the expected time can be bounded by

$$\sum_{u \text{ in level } \ell} \mathbb{E}[\mathsf{probe}_u] \geq \frac{1}{2} \cdot \frac{n}{m_\ell} \cdot \frac{1}{16}m_\ell t = \frac{1}{32}nt = \frac{\gamma}{256(\beta + 1)} \cdot n \log U.$$

  As the coefficient $\frac{\gamma}{256(\beta+1)}$ is a constant and $\log U \geq \log n \geq \log^* n \geq k$, the time per operation is at least $\Omega(k)$.

- Otherwise, it means for every level ℓ $(2 \leq \ell < h)$, at least half of the nodes in it satisfy *Prop 1*. Taking summation of the lower bound of $\mathbb{E}[\mathsf{cost}_u]$ for each node u in these levels, we know that the total time cost is at least $\sum_{\ell=2}^{h-1} \frac{\gamma}{100}m_\ell \cdot \frac{1}{2} \cdot \frac{n}{m_\ell} = \frac{\gamma}{200}n(h - 2)$ in expectation, which means the expected time cost per operation is at least $\Omega(h) = \Omega(k)$, as desired.

It remains to verify the premises of Lemma 3.2. Notice that we require $2 \leq \ell < h$, i.e., when we apply Lemma 3.2 on some node u, u and its children are not root or leaf nodes. Thus we always have $m_\ell = cn \log^{(k)} n / \log^{(\ell)} n$ and $\lambda_\ell = \log^{(\ell-1)} n / \log^{(\ell)} n$.

- $\lambda_\ell \geq 2^{64}$. According to the definition of h, for every $2 \leq \ell < h$, we have $\log^{(\ell)} n > c \log^{(k)} n \geq c = 10^6$, so $\lambda_\ell = \log^{(\ell-1)} n / \log^{(\ell)} n = 2^{\log^{(\ell)} n} / \log^{(\ell)} n \geq 2^c/c \gg 2^{64}$.

- $\lambda_\ell \leq \frac{\alpha}{12} \log n$. As n is sufficiently large, we only need to show $\lambda_\ell = o(\log n)$, that is, $\log^{(\ell-1)} n / \log^{(\ell)} n = o(\log n)$. All levels except for level 1 satisfy this condition.

12

- $m_\ell \log \lambda_\ell \geq 100R$. We can calculate

$$m_\ell \log \lambda_\ell = m_\ell(\log^{(\ell)} n - \log^{(\ell+1)} n) \geq \frac{1}{2} m_\ell \log^{(\ell)} n = \frac{1}{2} cn \log^{(k)} n \gg 100n \log^{(k)} n,$$

$$100R = 100nr \leq 100n \log^{(k)} n,$$

where the first inequality holds because $\log^{(\ell)} n \geq 2 \log^{(\ell+1)} n$ as $\log^{(\ell)} n \geq c$. So we have $m_\ell \log \lambda_\ell \geq 100R$ for all $2 \leq \ell < h$.

- $\frac{1}{2} \gamma m_\ell \log U \geq R + 5m_\ell + 2(\beta + 1)m_\ell t$. Substituting $t = \frac{\gamma}{8(\beta+1)} \log U$ into this inequality, we only need to prove $\frac{1}{4} \gamma m_\ell \log U \geq R + 5m_\ell$. When n is sufficiently large, we have

$$\gamma m_\ell \log U \geq \gamma \frac{cn \log^{(k)} n}{\log \log n} \log U \geq \gamma \frac{\log n}{\log \log n} \cdot cn \log^{(k)} n \geq 24cn \log^{(k)} n \geq 4(R + 5m_\ell).$$

- $m_\ell \geq n^{1-\gamma/2}$. As $m_\ell \geq cn \cdot \frac{\log^{(k)} n}{\log \log n}$, this inequality naturally holds with sufficiently large n.

This proves the theorem. $\qquad\square$

4   Outer Lemma

Suppose u is a level-ℓ node in the tree structure. It also represents an interval of m_ℓ consecutive meta-operations. In this section, we focus on the operations in u and prove Lemma 3.2, which asserts that either $\mathbb{E}[\mathsf{cost}_u] \geq \frac{\gamma}{100} m_\ell$ or $\mathbb{E}[\mathsf{probe}_u] \geq \frac{1}{16} m_\ell t$.

4.1   Proof Overview

We index the meta-operations in u by $\{1, \ldots, m_\ell\}$. Let a_i, d_i represent the keys to be inserted and deleted in the i-th meta-operation in u, respectively. Also define sequences $\mathbf{a} := (a_1, \ldots, a_{m_\ell})$, $\mathbf{d} := (d_1, \ldots, d_{m_\ell})$. These meta-operations are further divided into λ_ℓ subintervals of length $m_{\ell-1}$, each corresponding to a child node of u. We call each subinterval a *segment*. For each $i \in [\lambda_\ell]$, the keys inserted and deleted in the i-th segment are denoted by $\mathbf{a}^{[i]} := \big(a_{(i-1)m_{\ell-1}+1}, \ldots, a_{im_{\ell-1}}\big)$ and $\mathbf{d}^{[i]} := \big(d_{(i-1)m_{\ell-1}+1}, \ldots, d_{im_{\ell-1}}\big)$. We further define the following quantities of $\mathbf{a}$:

- $\mathcal{A} := \{a_1, \ldots, a_{m_\ell}\}$ is the unordered version of $\mathbf{a}$, and $A^{[i]} := \big\{a_{(i-1)m_{\ell-1}+1}, \ldots, a_{im_{\ell-1}}\big\}$ is the unordered version of $\mathbf{a}^{[i]}$.

- Inside the i-th segment, we use a permutation $\pi^{[i]}$ to represent the order of key insertions in $A^{[i]}$. Formally, $\pi^{[i]}$ is a permutation over $[m_{\ell-1}]$ such that the j-th key inserted in the i-th segment is the $\pi_j^{[i]}$-*th largest* element in $A^{[i]}$. Denote by $\boldsymbol{\pi} = \big(\pi^{[1]}, \ldots, \pi^{[\lambda_\ell]}\big)$ the sequence of all these permutations.

Let C be a (memory) state of the deterministic data structure at a given time, i.e., the contents in all N memory cells. It is clear that $H(C) \leq \log \binom{U}{n} + R$. We also define $C(S)$ to be the state of the cells S, where $S \subseteq [N]$. Formally, $C(S) := \{(i, \mathrm{cont}(i)) : i \in S\}$, where $\mathrm{cont}(i)$ is the w-bit content of the i-th cell in the state C. Notice that $C(S)$ also includes the address information S when S is a random variable. Hence, the entropy of $C(S)$ might be larger than $w \cdot |S|$.

Denote by C_{st} the state of the data structure just before executing the first operation in u; denote by C_{end} the state right after executing the last operation in u. We will not consider states before C_{st} or after C_{end} in this section.

We will prove Lemma 3.2 by contradiction via a communication game. In this game,

- Alice is given $A^{[1]}, \ldots, A^{[\lambda_\ell]}$, which determines $\mathcal{A}$, and is also given $C_{\text{st}}, \mathbf{d}, \boldsymbol{\pi}$;

- Bob is given $\mathcal{A}, C_{\text{st}}, \mathbf{d}, \boldsymbol{\pi}$;

- the goal of the game is for Alice to tell Bob $\left(A^{[1]}, \ldots, A^{[\lambda_\ell]}\right)$ by sending one message.

Note that $\left(A^{[1]}, \ldots, A^{[\lambda_\ell]}\right)$ is simply a random partition of $\mathcal{A}$, which has m_ℓ elements, into λ_ℓ groups of size $m_{\ell-1}$ each. Since all other inputs are independent of the partition, the entropy of $\left(A^{[1]}, \ldots, A^{[\lambda_\ell]}\right)$ given Bob's input is

$$H(A^{[1]}, \ldots, A^{[\lambda_\ell]} \mid \mathcal{A}, C_{\text{st}}, \mathbf{d}, \boldsymbol{\pi}) = H(A^{[1]}, \ldots, A^{[\lambda_\ell]} \mid \mathcal{A})$$
$$= \log \binom{m_\ell}{m_{\ell-1}, \ldots, m_{\ell-1}} = \log m_\ell! - \lambda_\ell \log m_{\ell-1}!. \qquad (1)$$

4.2 Outer Game Protocol

Let us now assume for contradiction that the conclusion of Lemma 3.2 does not hold, i.e., $\mathbb{E}[\mathsf{cost}_u] < \frac{\gamma m_\ell}{100}$ and $\mathbb{E}[\mathsf{probe}_u] < \frac{1}{16} m_\ell t$, we will show a protocol where Alice sends less than $(\log m_\ell! - \lambda_\ell \log m_{\ell-1}!)$ bits of information, but Bob can still recover $A^{[1]}, \ldots, A^{[\lambda_\ell]}$, which leads to a contradiction.

Recall the definition of cost_u: While a cell j is probed at different segments in u, we increase cost_u by one. Conversely, since cost_u is small, we know that most cells are only probed in at most one segment. We call the cells that are probed in at least two different segments *bad cells*, and use S_{bad} to denote the set of bad cells. The expected number of bad cells is at most $\mathbb{E}[|S_{\text{bad}}|] \leq \mathbb{E}[\mathsf{cost}_u] < \frac{\gamma m_\ell}{100}$.

The framework of our protocol is shown in Protocol 1. Below, we will explain it in detail.

Protocol 1: Outer Game Protocol Framework

1 Alice sends C_{end} to Bob
2 for $i = 1$ *to* λ_ℓ **do**
3 $\quad$ Bob computes C_{bef}, the state of the data structure before the i-th segment
4 $\quad$ $\mathcal{A}_{\text{rest}} := \mathcal{A} \setminus (A^{[1]} \cup \cdots \cup A^{[i-1]})$
5 $\quad$ $\mathcal{F}_{\text{all}}^{[i]} := \{B \subseteq \mathcal{A}_{\text{rest}} : |B| = m_{\ell-1}\}$
6 $\quad$ Alice decides $X_i \in \{0, 1\}$ and sends it to Bob $\qquad \qquad \triangleright$ Explained below
7 $\quad$ **if** $X_i = 1$ **then**
8 $\quad \quad$ Bob runs *tests* on all $B \in \mathcal{F}_{\text{all}}^{[i]}$ $\qquad \qquad \triangleright$ Explained below
9 $\quad \quad$ $\mathcal{F}_{\text{qual}}^{[i]} := \{B \in \mathcal{F}_{\text{all}}^{[i]} : B \text{ passes the test}\}$
10 $\quad \quad$ Alice sends the index of $A^{[i]}$ in $\mathcal{F}_{\text{qual}}^{[i]}$ $\quad \triangleright$ It is guaranteed that $A^{[i]}$ will pass the test
11 $\quad$ **else**
12 $\quad \quad$ Alice sends the index of $A^{[i]}$ in $\mathcal{F}_{\text{all}}^{[i]}$

Sending C_{end}. The first step is to send C_{end}, the final state of the data structure after processing all operations in u, to Bob. We show that the entropy of this message given Bob's knowledge is in fact much smaller than the number of bits in the memory, because C_{end} has a large mutual information with C_{st}. To upper bound this entropy, we consider the key set stored in C_{end}, which we denote by K_{end}. Now observe that K_{end} can be inferred from both C_{end} and from $(C_{\text{st}}, \mathcal{A}, \mathbf{d})$ (by running all possible queries on C_{end}, or on C_{st} and replacing keys $\mathbf{d}$ with $\mathcal{A}$). Thus, the mutual

information $I(C_{\mathrm{end}}\,;C_{\mathrm{st}},\mathcal{A},\mathbf{d},\boldsymbol{\pi}) \geq H(K_{\mathrm{end}}) = \log\binom{U}{n}$. Furthermore, we have

$$H(C_{\mathrm{end}} \mid C_{\mathrm{st}},\mathcal{A},\mathbf{d},\boldsymbol{\pi}) = H(C_{\mathrm{end}}) - I(C_{\mathrm{end}}\,;C_{\mathrm{st}},\mathcal{A},\mathbf{d},\boldsymbol{\pi}) \leq \log\binom{U}{n} + R - \log\binom{U}{n} = R. \quad (2)$$

Hence, Alice can send C_{end} *conditioned on* $(C_{\mathrm{st}},\mathcal{A},\mathbf{d},\boldsymbol{\pi})$, using at most R bits in expectation.

The remainder of the protocol consists of λ_ℓ rounds. In each round, Bob aims to learn a single $A^{[i]}$. Notice that since Bob knows the order of insertions $\pi^{[i]}$ and the sequence of deleted keys $\mathbf{d}^{[i]}$, Bob would also be able to infer the entire operation sequence in segment i.

Preparation steps. At the beginning of the i-th round, Bob already knows $A^{[1]},\dots,A^{[i-1]}$; combined with the common knowledge $\boldsymbol{\pi}$, $\mathbf{d}$, Bob can recover all operations in the first $i-1$ segments. So he can simulate the process of the data structure from C_{st} throughout the first $i-1$ segments, obtaining the state before the i-th segment, C_{bef}. At this point, the set of all possible keys that can be inserted in segment i *in Bob's view* is $\mathcal{A}_{\mathrm{rest}} := \mathcal{A} \setminus (A^{[1]} \cup \dots \cup A^{[i-1]})$. Line 5 defines $\mathcal{F}_{\mathrm{all}}^{[i]}$ as the collection of all $m_{\ell-1}$-element subsets of $\mathcal{A}_{\mathrm{rest}}$. Every set in $\mathcal{F}_{\mathrm{all}}^{[i]}$ is currently a candidate of $A^{[i]}$. It is clear that the correct $A^{[i]}$ is in $\mathcal{F}_{\mathrm{all}}^{[i]}$.

The most trivial way for Alice to reveal $A^{[i]}$ to Bob is to send the index of $A^{[i]}$ in $\mathcal{F}_{\mathrm{all}}^{[i]}$. This takes $\log\left|\mathcal{F}_{\mathrm{all}}^{[i]}\right| = \log\binom{|\mathcal{A}_{\mathrm{rest}}|}{m_{\ell-1}}$ bits. However, this approach would not lead to a contradiction, as $\sum_{i=1}^{\lambda_\ell} \log\left|\mathcal{F}_{\mathrm{all}}^{[i]}\right| = \log\binom{m_\ell}{m_{\ell-1},\dots,m_{\ell-1}}$ exactly equals $\log m_\ell! - \lambda_\ell \log m_{\ell-1}!$, the entropy that Bob needs to learn.

Compared to this simple method, our protocol lets Bob eliminate most of the candidates by running *tests* based on C_{bef} and C_{end}. Ideally, he can narrow down the candidates to a small subset $\mathcal{F}_{\mathrm{qual}}^{[i]} \subseteq \mathcal{F}_{\mathrm{all}}^{[i]}$ while the correct $A^{[i]} \in \mathcal{F}_{\mathrm{qual}}^{[i]}$ is not eliminated. In this case, Alice only needs to send $\log\left|\mathcal{F}_{\mathrm{qual}}^{[i]}\right| \ll \log\left|\mathcal{F}_{\mathrm{all}}^{[i]}\right|$ bits. On the other hand, we observe that occasionally, $A^{[i]}$ is eliminated by Bob's tests. In this *bad case*, Alice will still send $A^{[i]}$ using the index in $\mathcal{F}_{\mathrm{all}}^{[i]}$. An additional bit X_i is sent to tell Bob whether his tests will be successful.

Alice's Decision. We define $S_A^{[i]}$ to be the set of cells probed in the i-th segment. Recall that we use S_{bad} to represent the cells probed in multiple segments (Alice knows all operations, and thus, knows S_{bad}). X_i in Line 6 is determined according to the relationship between $S_A^{[i]}$ and S_{bad}: If $\left|S_A^{[i]} \cap S_{\mathrm{bad}}\right| > \gamma m_{\ell-1}/2$ (recall $\gamma := \frac{\alpha}{6(1+\alpha)}$), we call it the *bad case*, and Alice sends $X_i = 0$ and the index of $A^{[i]}$ in $\mathcal{F}_{\mathrm{all}}^{[i]}$ in Line 12. Otherwise, Alice sends $X_i = 1$.

Bob's Tests. For each candidate $B \in \mathcal{F}_{\mathrm{all}}^{[i]}$, Bob runs the following test. He pretends that B is the correct set $A^{[i]}$, and permutes B according to the permutation $\pi^{[i]}$, obtaining an ordered sequence of $m_{\ell-1}$ keys $\mathbf{b}^{[i]}$. Then, he simulates the operations in the i-th segment according to $\mathbf{b}^{[i]}$ and $\mathbf{d}^{[i]}$. The simulation starts with the state C_{bef}, and we denote by C_B the end state after the simulation. Let S_B be the set of probed cells during the simulation.

B passes the test if most of these probed cells have the same content as their final value in C_{end}, i.e., if $C_B(S_B)$ and $C_{\mathrm{end}}(S_B)$ differ by at most $\gamma m_{\ell-1}/2$ cells. In this case, we say that B is *qualified*.

Next we show that $A^{[i]}$ is always qualified when $X_i = 1$. Let $C_A^{[i]}$ be the true state after the i-th segment. Note that $C_A^{[i]}(S_A^{[i]})$ and $C_{\mathrm{end}}(S_A^{[i]})$ can only differ in $S_A^{[i]} \cap S_{\mathrm{bad}}$: If some cell $j \in S_A^{[i]}$

has different values in $C_A^{[i]}$ and C_{end}, it must be probed in a later segment; but it is also probed in segment i, which tells us that $j \in S_{\text{bad}}$ by the definition of S_{bad}. Combined with the fact $\left|S_A^{[i]} \cap S_{\text{bad}}\right| \leq \gamma m_{\ell-1}/2$ (since $X_i = 1$), we see that $A^{[i]}$ must be qualified.

Thus, in either case ($X_i = 0$ or $X_i = 1$), Bob is able to recover the set $A^{[i]}$. By iterating over all $i = 1, \ldots, \lambda_\ell$, Bob successfully recovers the partition $(A^{[1]}, \ldots, A^{[\lambda_\ell]})$ as required.

Analysis. Denote by M_i the message sent by Alice in the i-th round. M_i contains an indicator X_i and an index of $\log \left|\mathcal{F}_{\text{all}}^{[i]}\right|$ bits (when $X_i = 0$) or $\log \left|\mathcal{F}_{\text{qual}}^{[i]}\right|$ bits (when $X_i = 1$). Formally,

$$|M_i| = X_i \log \left|\mathcal{F}_{\text{qual}}^{[i]}\right| + (1 - X_i) \log \left|\mathcal{F}_{\text{all}}^{[i]}\right| + 1.$$

Thus, by (2), the message sent by Alice has at most

$$H(C_{\text{end}} \mid \mathcal{A}, C_{\text{st}}, \mathbf{d}, \boldsymbol{\pi}) + \sum_{i=1}^{\lambda_\ell} \mathbb{E}[|M_i|] \ \leq \ R + \sum_{i=1}^{\lambda_\ell} \mathbb{E}[|M_i|]$$

bits of information in total. On the other hand, since Bob can always recover the partition $(A^{[1]}, \ldots, A^{[\lambda_\ell]})$ given his input $(\mathcal{A}, C_{\text{st}}, \mathbf{d}, \boldsymbol{\pi})$ and Alice's message, we must have

$$R + \sum_{i=1}^{\lambda_\ell} \mathbb{E}[|M_i|] \ \geq \ H(A^{[1]}, \ldots, A^{[\lambda_\ell]} \mid \mathcal{A}, C_{\text{st}}, \mathbf{d}, \boldsymbol{\pi}),$$

$$\Leftrightarrow \quad \sum_{i=1}^{\lambda_\ell} \mathbb{E}[|M_i|] \ \geq \ \log m_\ell! - \lambda_\ell \log m_{\ell-1}! - R. \tag{3}$$

Next, we will derive an upper bound on $\mathbb{E}[|M_i|]$ that will contradict with this inequality when $\mathbb{E}[\text{cost}_u] < \frac{\gamma m_\ell}{100}$ and $\mathbb{E}[\text{probe}_u] < \frac{1}{16} m_\ell t$.

4.3 Estimate the Message Length

We introduce an intermediate set $\mathcal{F}_{\text{nbr}}^{[i]} := \{B \in \mathcal{F}_{\text{all}}^{[i]} : |A^{[i]} \cap B| \geq m_{\ell-1}/2\}$ to help us estimate $\mathcal{F}_{\text{qual}}^{[i]}$. This is the set of all possible B that are not too different from the correct $A^{[i]}$. Define indicator variables $Y_i := \mathbb{1}\left[\mathcal{F}_{\text{qual}}^{[i]} \subseteq \mathcal{F}_{\text{nbr}}^{[i]}\right]$ and $Z_i := \mathbb{1}\left[\mathbb{E}[|S_A^{[i]}|] \leq m_{\ell-1}t\right]$. The variable Y_i indicates that the only sets that can pass the test are the ones similar to the correct $A^{[i]}$. Note that Y_i is a random variable, while Z_i is not (since the only random variable in the definition of Z_i is placed within the expection), which can be fully determined by the location of the current segment (i.e., determined by the node u's location in the tree and the segment index i). Later, we will demonstrate that segments with $Z_i = 1$ are *likely* to have $Y_i = 1$.

Then, when $Y_i = Z_i = 1$, we can apply $\left|\mathcal{F}_{\text{qual}}^{[i]}\right| \leq \left|\mathcal{F}_{\text{nbr}}^{[i]}\right|$ on the inequality; otherwise, we just use the trivial bound $\left|\mathcal{F}_{\text{qual}}^{[i]}\right| \leq \left|\mathcal{F}_{\text{all}}^{[i]}\right|$. Recall $X_i = \mathbb{1}\left[\left|S_A^{[i]} \cap S_{\text{bad}}\right| \leq \gamma m_{\ell-1}/2\right]$ indicates if this is the bad case, we then have

$$\begin{aligned} |M_i| &\leq (1 - X_i Y_i Z_i) \log \left|\mathcal{F}_{\text{all}}^{[i]}\right| + X_i Y_i Z_i \log \left|\mathcal{F}_{\text{nbr}}^{[i]}\right| + 1 \\ &= \log \left|\mathcal{F}_{\text{all}}^{[i]}\right| - X_i Y_i Z_i \left(\log \left|\mathcal{F}_{\text{all}}^{[i]}\right| - \log \left|\mathcal{F}_{\text{nbr}}^{[i]}\right|\right) + 1. \end{aligned} \tag{4}$$

The following claims give lower bounds on X_i, Y_i, Z_i and $\log \left|\mathcal{F}_{\text{all}}^{[i]}\right| - \log \left|\mathcal{F}_{\text{nbr}}^{[i]}\right|$ respectively.

Claim 4.1. *We have*

$$\mathbb{E}\left[\sum_{i=1}^{\lambda_\ell/2} X_i\right] \geq \frac{1}{4}\lambda_\ell.$$

Proof. First, we reconsider how $S_A^{[i]} \cap S_{\mathrm{bad}}$ in the i-th segment is related to cost_u. If some cell is probed in $s > 1$ different segments of u, it incurs $s - 1 \geq s/2$ costs on u. Hence, if we think every probe of a bad cell incurs $1/2$ cost, we will get a lower bound on cost_u. Noting that $S_A^{[i]} \cap S_{\mathrm{bad}}$ is exactly the set of bad cells probed in segment i, where each probe contributes at least $1/2$ to cost_u, we have

$$\frac{1}{2}\mathbb{E}\left[\sum_{i=1}^{\lambda_\ell} \left|S_A^{[i]} \cap S_{\mathrm{bad}}\right|\right] \leq \mathbb{E}[\mathsf{cost}_u] \leq \frac{\gamma}{100}m_\ell. \tag{5}$$

Then we put X_i into (5). Since $X_i = 0$ implies $\left|S_A^{[i]} \cap S_{\mathrm{bad}}\right| > \gamma m_{\ell-1}/2$,

$$\frac{1}{2}\mathbb{E}\left[\sum_{i=1}^{\lambda_\ell} \frac{\gamma m_{\ell-1}}{2}(1 - X_i)\right] \leq \frac{1}{2}\mathbb{E}\left[\sum_{i=1}^{\lambda_\ell} \left|S_A^{[i]} \cap S_{\mathrm{bad}}\right|\right] \leq \frac{\gamma}{100}m_\ell = \frac{\gamma m_{\ell-1} \cdot \lambda_\ell}{100}$$

$$\Rightarrow \quad \mathbb{E}\left[\sum_{i=1}^{\lambda_\ell}(1 - X_i)\right] \leq \frac{\lambda_\ell}{25},$$

$$\mathbb{E}\left[\sum_{i=1}^{\lambda_\ell/2} X_i\right] \geq \frac{\lambda_\ell}{2} - \mathbb{E}\left[\sum_{i=1}^{\lambda_\ell}(1 - X_i)\right] \geq \frac{\lambda_\ell}{2} - \frac{\lambda_\ell}{25} \geq \frac{\lambda_\ell}{4}. \qquad \square$$

The next claim gives a bound on Y_i.

Claim 4.2. *For any $1 \leq i \leq \lambda_\ell/2$, if $Z_i = 1$, we have*

$$\Pr[Y_i = 0 \wedge X_i = 1] \leq \frac{1}{8}.$$

Proof. Suppose $X_i = 1$. For any $B \in \mathcal{F}_{\mathrm{qual}}^{[i]}$, we know $A^{[i]}$ and B are both qualified, which means $C_A^{[i]}(S_A^{[i]})$ and $C_{\mathrm{end}}(S_A^{[i]})$ differ in at most $\gamma m_{\ell-1}/2$ cells; so do $C_B(S_B)$ and $C_{\mathrm{end}}(S_B)$. Thus $C_A^{[i]}(S_A^{[i]} \cap S_B)$ and $C_B(S_A^{[i]} \cap S_B)$ have at most $\gamma m_{\ell-1}$ different cells, in which case we say $A^{[i]}$ and B are *consistent*. Formally:

Definition 4.3. *For two insertion sets $A, B \in \mathcal{F}_{\mathrm{all}}^{[i]}$, we say A and B are consistent, if $C_A(S_A \cap S_B)$ and $C_B(S_A \cap S_B)$ have at most $\gamma m_{\ell-1}$ different cells.*

Under this definition, if $X_i = 1$ and $B \in \mathcal{F}_{\mathrm{qual}}^{[i]}$, then $A^{[i]}$ and B must be consistent. Hence,

$$\Pr[Y_i = 0 \wedge X_i = 1] \leq \Pr\left[\left(\exists B \in \mathcal{F}_{\mathrm{all}}^{[i]} \setminus \mathcal{F}_{\mathrm{nbr}}^{[i]}, \ B \in \mathcal{F}_{\mathrm{qual}}^{[i]}\right) \wedge (X_i = 1)\right]$$

$$\leq \Pr\left[\exists B \in \mathcal{F}_{\mathrm{all}}^{[i]} \setminus \mathcal{F}_{\mathrm{nbr}}^{[i]}, \ B \text{ and } A^{[i]} \text{ are consistent}\right].$$

Thus, to prove Claim 4.2, it suffices to show

$$\Pr\left[\exists B \in \mathcal{F}_{\mathrm{all}}^{[i]} \setminus \mathcal{F}_{\mathrm{nbr}}^{[i]}, \ B \text{ and } A^{[i]} \text{ are consistent}\right] \leq \frac{1}{8}.$$

Next, we show that a random $B \notin \mathcal{F}_{\mathrm{nbr}}^{[i]}$ is unlikely to be consistent with $A^{[i]}$.

Lemma 4.4 (Inner Lemma). *Assume C_{bef} (the state before the i-th segment) and $\mathcal{A}_{\text{rest}}$ are random variables whose distribution is induced from Distribution 1. Let A, B be two uniformly random $m_{\ell-1}$-sized subsets of $\mathcal{A}_{\text{rest}}$ conditioned on $|A \cap B| < m_{\ell-1}/2$. If for some parameter t, inequalities $\frac{1}{2}\gamma m_\ell \log U \geq R + 5m_\ell + 2(\beta+1)m_\ell t$ (β is a global fixed constant), $m_\ell \geq n^{1-\gamma/2}$, and $\mathbb{E}[|S_A|] \leq m_{\ell-1}t$ hold, then we have*

$$\Pr_{C_{\text{bef}}, \mathcal{A}_{\text{rest}}, A, B}\left[A \text{ and } B \text{ are consistent} \;\middle|\; |A \cap B| < m_{\ell-1}/2\right] \leq U^{-\gamma m_{\ell-1}}.$$

We will prove this lemma in Section 5. Now, we continue the proof of Lemma 3.2 using this lemma.

To apply this lemma, observe that the first two inequalities ($\frac{1}{2}\gamma m_\ell \log U \geq R + 5m_\ell + 2(\beta+1)m_\ell t$ and $m_\ell \geq n^{1-\gamma/2}$) hold because they are also premises of Lemma 3.2; the third constraint ($\mathbb{E}[|S_A|] \leq m_{\ell-1}t$) holds because $Z_i = 1$. Moreover, note that $A^{[i]}$ is a random subset of $\mathcal{A}_{\text{rest}}$; if we randomly select a $B \in \mathcal{F}_{\text{all}}^{[i]} \setminus \mathcal{F}_{\text{nbr}}^{[i]}$, the distribution of $(A^{[i]}, B)$ matches the required distribution of (A, B) in the lemma.

Combining with the discussion above, as long as $Z_i = 1$, we have

$$\Pr_{C_{\text{bef}}, \mathcal{A}_{\text{rest}}, A^{[i]}}[Y_i = 0 \wedge X_i = 1] \leq \Pr\left[\exists B \in \mathcal{F}_{\text{all}}^{[i]} \setminus \mathcal{F}_{\text{nbr}}^{[i]}, \; A^{[i]} \text{ and } B \text{ are consistent}\right]$$

$$\leq \mathbb{E}\left[\left|\left\{B \in \mathcal{F}_{\text{all}}^{[i]} \setminus \mathcal{F}_{\text{nbr}}^{[i]} : A^{[i]} \text{ and } B \text{ are consistent}\right\}\right|\right]$$

$$= \left|\mathcal{F}_{\text{all}}^{[i]} \setminus \mathcal{F}_{\text{nbr}}^{[i]}\right| \cdot \Pr_{\substack{C_{\text{bef}}, \mathcal{A}_{\text{rest}}, A^{[i]} \\ B \in \mathcal{F}_{\text{all}}^{[i]} \setminus \mathcal{F}_{\text{nbr}}^{[i]}}}\left[A^{[i]} \text{ and } B \text{ are consistent}\right]$$

$$\leq \left|\mathcal{F}_{\text{all}}^{[i]}\right| \cdot U^{-\gamma m_{\ell-1}} \tag{$*$}$$

$$\leq 2^{m_\ell} \cdot U^{-\gamma m_{\ell-1}} = 2^{\lambda_\ell m_{\ell-1}} \cdot n^{-\alpha m_{\ell-1}/6}, \tag{6}$$

where Step ($*$) is derived by applying Lemma 4.4 with $(A, B) = (A^{[i]}, B)$. (Recall $\gamma = \frac{\alpha}{6(1+\alpha)}$.)

Recall that in the statement of Lemma 3.2, we require $\lambda_\ell \leq \log n^{\alpha/12}$. Plugging this into (6), we obtain $\Pr[Y_i = 0 \wedge X_i = 1] \leq n^{-\alpha m_{\ell-1}/12}$. Since α is a constant in our assumption $U = \text{poly}(n)$, with sufficiently large n, we have $\Pr[Y_i = 0 \wedge X_i = 1] \leq 1/8$. (This probability can be smaller than any fixed constant, but $1/8$ is enough for us.) $\qquad\square$

Using the condition $\mathbb{E}[\mathsf{probe}_u] \leq \frac{1}{16}m_\ell t$, we can derive a bound of $\sum Z_i$ as below.

Claim 4.5. *We have*

$$\sum_{i=1}^{\lambda_\ell/2}(1 - Z_i) \leq \frac{1}{16}\lambda_\ell.$$

Proof. Note that $\sum_{i=1}^{\lambda_\ell/2}\left|S_A^{[i]}\right| \leq \sum_{i=1}^{\lambda_\ell}\left|S_A^{[i]}\right| \leq \mathsf{probe}_u$. Combined with $Z_i := \mathbb{1}\left[\mathbb{E}[|S_A^{[i]}|] \leq m_{\ell-1} \cdot t\right]$, we get

$$\sum_{i=1}^{\lambda_\ell/2}(1 - Z_i) \cdot m_{\ell-1} \cdot t \leq \sum_{i=1}^{\lambda_\ell/2}\mathbb{E}[|S_A^{[i]}|] \leq \mathbb{E}[\mathsf{probe}_u] \leq \frac{1}{16}m_\ell \cdot t \quad \Rightarrow \quad \sum_{i=1}^{\lambda_\ell/2}(1 - Z_i) \leq \frac{1}{16}\lambda_\ell. \quad\square$$

Next, a simple counting argument implies a lower bound on $\log\left|\mathcal{F}_{\text{all}}^{[i]}\right| - \log\left|\mathcal{F}_{\text{nbr}}^{[i]}\right|$.

18

Claim 4.6. *For $1 \leq i \leq \lambda_\ell/2$, we have*

$$\log \left| \mathcal{F}_{\text{all}}^{[i]} \right| - \log \left| \mathcal{F}_{\text{nbr}}^{[i]} \right| \geq \frac{1}{4} m_{\ell-1} \log \lambda_\ell.$$

Proof. By definition, $\left| \mathcal{F}_{\text{all}}^{[i]} \right| = \binom{|\mathcal{A}_{\text{rest}}|}{m_{\ell-1}}$, and since $i \leq \lambda_\ell/2$, $\mathcal{A}_{\text{rest}}$ satisfies $|\mathcal{A}_{\text{rest}}| \geq m_\ell/2$, so

$$\log \left| \mathcal{F}_{\text{all}}^{[i]} \right| \geq \log \binom{m_\ell/2}{m_{\ell-1}} \geq \log \left(\frac{m_\ell/2}{m_{\ell-1}} \right)^{m_{\ell-1}} = m_{\ell-1} \log \lambda_\ell - m_{\ell-1}. \tag{7}$$

On the other hand, we have

$$\left| \mathcal{F}_{\text{nbr}}^{[i]} \right| = \sum_{j=m_{\ell-1}/2}^{m_{\ell-1}} \binom{m_{\ell-1}}{j} \binom{|\mathcal{A}_{\text{rest}}| - m_{\ell-1}}{m_{\ell-1} - j}$$

$$\leq m_{\ell-1} \binom{m_{\ell-1}}{m_{\ell-1}/2} \binom{|\mathcal{A}_{\text{rest}}|}{m_{\ell-1}/2},$$

$$\log \left| \mathcal{F}_{\text{nbr}}^{[i]} \right| \leq \log m_{\ell-1} + m_{\ell-1} + \log \binom{m_\ell}{m_{\ell-1}/2}$$

$$\leq 2 m_{\ell-1} + \frac{m_{\ell-1}}{2} \log 2e\lambda_\ell$$

$$\leq 4 m_{\ell-1} + \frac{m_{\ell-1}}{2} \log \lambda_\ell.$$

On the first line above, we count the number of elements $B \in \mathcal{F}_{\text{nbr}}^{[i]}$ by enumerating $j = |A^{[i]} \cap B|$; the two binomial coefficient factors represent the number of ways to select $A^{[i]} \cap B$ and $B \setminus A^{[i]}$, respectively. The inequality on the second line holds because both binomial coefficients reach their maximum values when $j = m_{\ell-1}/2$. Combining the last line with (7), we have

$$\log \left| \mathcal{F}_{\text{all}}^{[i]} \right| - \log \left| \mathcal{F}_{\text{nbr}}^{[i]} \right| \geq \frac{m_{\ell-1}}{2} \log \lambda_\ell - 5 m_{\ell-1}$$

$$\geq \frac{1}{4} m_{\ell-1} \log \lambda_\ell$$

as we have assumed $\log \lambda_\ell \geq 64$. This proves the claim. $\qquad \square$

Now we are ready to finish the proof of Lemma 3.2:

Proof of Lemma 3.2. Suppose the lemma does not hold. Recall (4):

$$|M_i| \leq \log \left| \mathcal{F}_{\text{all}}^{[i]} \right| - X_i Y_i Z_i \left(\log \left| \mathcal{F}_{\text{all}}^{[i]} \right| - \log \left| \mathcal{F}_{\text{nbr}}^{[i]} \right| \right) + 1.$$

For the segments with $i \leq \lambda_\ell/2$, we use the previous claims to bound each term in this inequality. From Claims 4.1, 4.2 and 4.5, we can bound the expectation of $X_i Y_i Z_i$ by writing it as

$$\mathbb{E}[X_i Y_i Z_i] \geq \mathbb{E}[X_i Y_i Z_i - (1 - X_i)(1 - Z_i)]$$

$$= \mathbb{E}[X_i - (1 - Z_i) - X_i Z_i (1 - Y_i)]$$

$$= \mathbb{E}[X_i] - (1 - Z_i) - Z_i \Pr[Y_i = 0 \wedge X_i = 1].$$

Claim 4.6 is a direct upper bound on the term $\log \left| \mathcal{F}_{\text{all}}^{[i]} \right| - \log \left| \mathcal{F}_{\text{nbr}}^{[i]} \right|$.

For other segments with $i > \lambda_\ell/2$, we simply apply the trivial bound $|M_i| \leq \log \left|\mathcal{F}_{\mathrm{all}}^{[i]}\right| + 1$. Now taking summation over all the segments, i.e., $i = 1, \ldots, \lambda_\ell$, we get

$$\sum_{i=1}^{\lambda_\ell} \mathbb{E}[|M_i|] \;\leq\; \lambda_\ell + \sum_{i=1}^{\lambda_\ell} \log \left|\mathcal{F}_{\mathrm{all}}^{[i]}\right| - \sum_{i=1}^{\lambda_\ell/2} \Big(\mathbb{E}[X_i] - (1 - Z_i) - Z_i \Pr[Y_i = 0 \wedge X_i = 1]\Big) \cdot \frac{1}{4} m_{\ell-1} \log \lambda_\ell$$

$$= \lambda_\ell + \sum_{i=1}^{\lambda_\ell} \log \left|\mathcal{F}_{\mathrm{all}}^{[i]}\right| - \left(\mathbb{E}\left[\sum_{i=1}^{\lambda_\ell/2} X_i\right] - \sum_{i=1}^{\lambda_\ell/2} (1 - Z_i) - \sum_{\substack{1 \leq i \leq \lambda_\ell/2 \\ Z_i = 1}} \Pr[Y_i = 0 \wedge X_i = 1] \right) \cdot \frac{1}{4} m_{\ell-1} \log \lambda_\ell$$

$$\leq \lambda_\ell + \sum_{i=1}^{\lambda_\ell} \log \left|\mathcal{F}_{\mathrm{all}}^{[i]}\right| - \left(\frac{\lambda_\ell}{4} - \frac{\lambda_\ell}{16} - \frac{1}{8} \cdot \frac{\lambda_\ell}{2}\right) \cdot \frac{1}{4} m_{\ell-1} \log \lambda_\ell$$

$$= \lambda_\ell + \log m_\ell! - \lambda_\ell \log m_{\ell-1}! - \frac{m_\ell}{32} \log \lambda_\ell. \tag{8}$$

On the other side, recall (3):

$$\sum_{i=1}^{\lambda_\ell} \mathbb{E}[|M_i|] \geq \log m_\ell! - \lambda_\ell \log m_{\ell-1}! - R.$$

Combining it with (8), we get

$$R \geq \frac{m_\ell}{32} \log \lambda_\ell - \lambda_\ell > \frac{m_\ell}{64} \log \lambda_\ell, \tag{9}$$

where the second inequality is from $\lambda_\ell \leq m_\ell$ and $\log \lambda_\ell \geq 64$. Finally, observe that (9) contradicts with the premise of the lemma, $R \leq \frac{m_\ell}{100} \log \lambda_\ell$. This completes the proof of Lemma 3.2. $\qquad \square$

5 Inner Lemma

In this section, we prove Lemma 4.4, the last piece of our lower bound.

Lemma 4.4 (Restated). *Assume C_{bef} (the state before the i-th segment) and $\mathcal{A}_{\mathrm{rest}}$ are random variables whose distribution is induced from Distribution 1. Let A, B be two uniformly random $m_{\ell-1}$-sized subsets of $\mathcal{A}_{\mathrm{rest}}$ conditioned on $|A \cap B| < m_{\ell-1}/2$. If for some parameter t, inequalities $\frac{1}{2}\gamma m_\ell \log U \geq R + 5m_\ell + 2(\beta+1)m_\ell t$ (β is a global fixed constant), $m_\ell \geq n^{1-\gamma/2}$, and $\mathbb{E}[|S_A|] \leq m_{\ell-1} t$ hold, then we have*

$$\Pr_{C_{\mathrm{bef}}, \mathcal{A}_{\mathrm{rest}}, A, B}\left[A \text{ and } B \text{ are consistent} \;\Big|\; |A \cap B| < m_{\ell-1}/2\right] \leq U^{-\gamma m_{\ell-1}}.$$

The view here is a bit different from previous sections as we focus on a *single* segment of length $m_{\ell-1}$ instead of all λ_ℓ segments within a node u. We start by restating the procedure within the single segment using more specific notations.

1. Parameters. We still use n to represent the number of keys stored in the data structure, and $U = n^{1+\alpha}$ to represent the size of the key universe. Define $m < n/2$ as the number of meta-operations in the segment. When applying this lemma in Section 4, we set $m = m_{\ell-1}$.

   In the statement of Lemma 4.4 above, the probability is conditioned on $|A \cap B| < m/2$. Below, we first prove the lemma conditioned on $|A \cap B| = g$ for every fixed $g < m/2$. Then Lemma 4.4 can be directly implied via the law of total probability.

The parameters n, U, m, g are fixed, while all other involved variables are random according to the hard distribution. Their distributions will be explicitly stated below.

2. Initial keys. Let K be the key set stored in C_{bef}, the state of the data structure just before the current segment. It is chosen uniformly at random among all n-sized subsets of the universe $[U]$. C_{bef} is also a random state, about which we only know that its key set K follows the above uniform distribution.

3. Deletion sequence. In Section 4, we used $\mathbf{d}^{[i]}$ to denote the keys to be deleted in the i-th segment. Now we omit the superscript i since we are focusing on a single segment, using $\mathbf{d}$ to denote the deletion sequence in the segment. We further divide the information of $\mathbf{d}$ into two parts: the set of keys $D := \{d_1, \ldots, d_m\}$ to be deleted, and the order in which the keys in D are deleted. The latter one is formulated by a permutation σ over $[m]$.

   Conditioned on K, it is easy to see that D follows the uniform distribution over all m-element subsets of K and σ is a random permutation.

4. Insertion sets. The two random insertion sets A and B are uniformly chosen over m-element subsets of $[U] \setminus K$, conditioned on $|A \cap B| = g$. Note that in the original description, A, B are sampled within $\mathcal{A}_{\text{rest}}$, but $\mathcal{A}_{\text{rest}}$ is a random set by itself. Thus, we can avoid using the intermediate variable $\mathcal{A}_{\text{rest}}$ and directly describe the distribution of (A, B) as above. $\mathcal{A}_{\text{rest}}$ will not appear throughout the proof of Lemma 4.4. Also note that if we hide B and only observe the distribution of A (conditioned on K), it is a uniformly random m-sized subset of $[U] \setminus K$, which exactly matches the distribution of the insertion set in the current segment (induced from the hard instance).

   We use a permutation π to represent the order in which the keys are inserted. It was denoted by $\pi^{[i]}$ in the outer game. We denote by $\mathbf{a}, \mathbf{b}$ the ordered sequences of inserted keys, which are obtained by permuting A, B with the same permutation π. ($\mathbf{a}$ was denoted by $\mathbf{a}^{[i]}$ in the outer game.)

5. Ending states. For any $1 \le i \le m$, recall that meta-operation i consists of three operations: $\texttt{Query}(d_i)$, $\texttt{Delete}(d_i)$, $\texttt{Insert}(a_i)$, assuming the insertion sequence $\mathbf{a}$ is processed. The state of the data structure after executing all m meta-operations on C_{bef} is denoted by C_A. Similarly, if we replace $\mathbf{a}$ with $\mathbf{b}$ and execute these operations on C_{bef}, we obtain another ending state C_B. Let S_A, S_B be the sets of probed cells during these two procedures, respectively. Note that from the condition of Lemma 4.4, $\mathbb{E}[|S_A|] = \mathbb{E}[|S_B|] \le mt$.

6. Consistency. Recall Definition 4.3: We say A and B are consistent if $C_A(S_A \cap S_B)$ and $C_B(S_A \cap S_B)$ have at most γm different cells, where $\gamma := \frac{\alpha}{6(1+\alpha)}$ is a fixed constant.

Based on these definitions, we restate Lemma 4.4 as follows:

Lemma 5.1 (Inner Lemma Restated). *For integers U, n, m, g satisfying $g < m/2$, $m < n/2$, $U = \mathrm{poly}(n)$, assume the random variables $C_{\text{bef}}, \mathbf{d}, \pi, A, B$ are randomly sampled according to the procedure above. Let $t > 0$ be a parameter such that the constraints $\frac{1}{2}\gamma m \log U \ge R + 5m + 2(\beta+1)mt$ (β is a global fixed constant), $m \ge n^{1-\gamma/2}$, and $\mathbb{E}[|S_A|] \le mt$ are satisfied, we must have*

$$\Pr\Big[A \text{ and } B \text{ are consistent} \,\Big|\, |A \cap B| = g\Big] \le U^{-\gamma m}.$$

Next, we introduce a new communication game to prove this lemma.

5.1 Inner Game

Sender Alice and receiver Bob are playing a communication game based on the random process above. The public information known by them are the fixed parameters g, m, n, U. In addition, Alice knows the initial state C_{bef} and sequences $\mathbf{a}, \mathbf{b}, \mathbf{d}$; Bob only knows permutations π and σ. Now Alice wants to send Bob A, B, D and the initial key set K when A and B are consistent. Since Bob knows the two permutations, after learning A, B, D, he knows the complete sequences $\mathbf{a}, \mathbf{b}, \mathbf{d}$. To achieve this goal, they follow the protocol introduced below, which makes use of the following data structure from [CKRT04].

Theorem 5.2 (Bloomier Filter [CKRT04]). *Let A, B be two given disjoint subsets of $[U]$. There is a set S separating A, B, i.e., $A \subseteq S$ and $B \cap S = \varnothing$, that we can encode within $O(|A| + |B|)$ bits in expectation. We denote the leading factor within the big-O notation by β, which is a global fixed constant that appeared in Lemmas 3.2 and 4.4. (Formally, the expected encoding length is at most $\beta(|A| + |B|)$ for all $|A| + |B| > 0$.)*

The protocol is as follows:

1. Check consistency. First, Alice determines an indicator variable $W := \mathbb{1}[A \text{ and } B \text{ are consistent}]$ and sends it to Bob. If A and B are not consistent, the game is terminated immediately. Otherwise, they continue with the subsequent steps.

2. Send B. Alice directly sends B using $\log \binom{U}{m}$ bits. Since Bob knows π, he learns $\mathbf{b}$ after receiving the message.

3. Send cell contents. Alice sends an artificial memory state C_{mix}, which is a combination of C_{bef} and C_A. For the cells in S_B, $C_{\mathrm{mix}}(S_B) = C_{\mathrm{bef}}(S_B)$; for the cells outside S_B, $C_{\mathrm{mix}}(\overline{S_B}) = C_A(\overline{S_B})$. Sending this state takes $Nw = \log \binom{U}{n} + R$ bits.

   To help Bob use this mixed state correctly, Alice also needs to tell Bob which cells in C_{mix} come from C_{bef} and which cells come from C_A. The most trivial way of sending this "partition information" is to send one bit for each cell to indicate where it comes from, which takes a total of $O(n)$ bits. Unfortunately, this trivial way does not suffice for all the cases we concern,[5] and we use a better approach as follows.

   Notice that $C_{\mathrm{bef}}(\overline{S_A \cup S_B}) = C_A(\overline{S_A \cup S_B})$. For cells in $\overline{S_A \cup S_B}$, it is correct to treat them as coming from either C_{bef} or C_A. Hence Alice only needs to send *some* partition $(S_B^*; \overline{S_B^*})$ that agrees with the true partition $(S_B; \overline{S_B})$ in $S_A \cup S_B$ (i.e., $S_B \subseteq S_B^*$ and $S_A \setminus S_B \subseteq \overline{S_B^*}$).

   Fortunately, the Bloomier Filter stated in Theorem 5.2 solves this task. It can encode such a partition $(S_B^*; \overline{S_B^*})$ using $\beta|S_A \cup S_B|$ bits in expectation for a global constant $\beta > 0$. Alice directly sends the Bloomier Filter to Bob, which takes $\beta|S_A \cup S_B| \leq 2\beta mt$ bits in expectation, since we have assumed $\mathbb{E}[|S_A|] = \mathbb{E}[|S_B|] \leq mt$. Then Bob learns the partition $(S_B^*; \overline{S_B^*})$; in particular, Bob now knows $C_{\mathrm{bef}}(S_B^*)$.

4. Do simulation and recover D. Bob runs the following test for all possible choices of D on $C_{\mathrm{bef}}(S_B^*)$. He enumerates all m-sized subsets $D^* \subset [U]$, and permutes it according to σ to get the deletion sequence $\mathbf{d}^*$. He then performs m meta-operations from C_{bef} of the form "$\mathtt{Query}(d_i^*) - \mathtt{Delete}(d_i^*) - \mathtt{Insert}(b_i)$", pretending that D^* is the true deletion set. During

---

[5]Actually, it is enough to use this $O(n)$-bit approach to prove Theorem 3.1 where we only use Lemma 5.1 for $m \geq \frac{n}{\log \log n}$. The following improved approach is needed in the later discussion about an extension of Theorem 3.1 (i.e., Theorem 6.1), where we need to use Lemma 5.1 for much smaller m where an $O(n)$ cost is unaffordable.

the test, once a cell outside S_B^* is probed, or any `Query` returns "not exist", the test terminates and D^* *fails the test.* If no such exception occurs during the whole simulation, we say D^* *passes the test.*

It is clear that the correct D will pass the test. In fact, to pass the test, D^* must satisfy $D^* \subseteq K \cup B$, as the data structure will not return "exist" for any key it does not store. Let $\mathcal{F} = \{D^* : D^* \text{ passes the test}\}$, then $|\mathcal{F}| \leq \binom{|K \cup B|}{m} = \binom{n+m}{m}$. So Alice can send Bob an index in $\mathcal{F}$ within $\log \binom{n+m}{m}$ bits, to tell him the index of D in $\mathcal{F}$. After receiving the message, Bob recovers D and thus knows **d**.

5. Recover S_B and $S_A \cap S_B$. After knowing **d** and **b**, Bob can simulate the correct operation sequence again, recording which cells are probed during the process, namely S_B. Then, Alice sends the subset $S_A \cap S_B \subseteq S_B$ to Bob *given* S_B. It takes at most $\mathbb{E}[|S_B|] \leq mt$ bits in expectation.

6. Recover A and K. Alice sends Bob the cells where $C_A(S_A \cap S_B)$ and $C_B(S_A \cap S_B)$ are different, including both the address information (i.e., the set of indices of these cells) and their cell contents. The former costs $\mathbb{E}[|S_B|] \leq mt$ bits in expectation, as it is a subset of S_B. For the latter, since A and B are consistent, the number of such cells will not exceed γm by definition. Thus, we need totally $mt + \gamma m w = mt + \gamma m \log U$ bits to send the message (we have assumed $w = \log U$).

   After that, Bob can use this message to obtain C_A:

   - In the cells $S_A \cap S_B$, C_A and C_B have limited difference, which are already sent to Bob, so Bob can edit $C_B(S_A \cap S_B)$ to obtain $C_A(S_A \cap S_B)$.
   - The cells $\overline{S_A}$ are not probed before reaching C_A, so in these cells we have $C_A(\overline{S_A}) = C_{\text{bef}}(\overline{S_A}) = C_{\text{mix}}(\overline{S_A})$.
   - Moreover, $C_A(S_A \setminus S_B) = C_{\text{mix}}(S_A \setminus S_B)$.

   Specifically, the last two parts imply that $C_A(\overline{S_A \cap S_B}) = C_{\text{mix}}(\overline{S_A \cap S_B})$. Bob learns C_A by merging it with the first part $C_A(S_A \cap S_B)$. Note that Bob does not know S_A (yet); he only knows S_B and $S_A \cap S_B$. But S_A is not needed for this merging process.

   At this point, Bob learns the whole state C_A, and hence knows its key set $(K \setminus D) \cup A$. The final step is to send a $\log \binom{n}{m}$-bit message specifying A given $(K \setminus D) \cup A$. Then Bob can infer A and K respectively.

Thus, Bob recovers the four sets A, B, D, K in this whole process when A and B are consistent ($W = 1$). Denote the message Alice sends (including W) by M_{send}, and denote the information Bob learns about A, B, D, K by M_{learn}. Here M_{send} and M_{learn} can be seen as random variables about the random process. Since M_{learn} can be inferred from M_{send} under any condition, we have

$$H(M_{\text{send}} \mid W = 1) \geq H(M_{\text{learn}} \mid W = 1). \tag{10}$$

In the next subsection, we will analyze the entropy of M_{send} and M_{learn} according to each step of communication.

5.2 Entropy Calculation

For M_{send}, we add the message sent in each step together:

$$H(M_{\text{send}} \mid W = 1) \leq \log \binom{U}{m} + \left(\log \binom{U}{n} + R\right) + 2\beta mt + \log \binom{n+m}{m}$$
$$+ mt + (mt + \gamma m \log U) + \log \binom{n}{m}, \tag{11}$$

where the terms on the RHS correspond to the cost of sending B, C_{mix}, the Bloomier Filter, the index of D, $S_A \cap S_B$ conditioned on S_B, the difference between $C_A(S_A \cap S_B)$ and $C_B(S_A \cap S_B)$, and A conditioned on $(K \setminus D) \cup A$, respectively.

All information Bob learns is $M_{\text{learn}} = (A, B, D, K)$, so we have

$$H(M_{\text{learn}} \mid W = 1) \geq \log\left(1 \middle/ \max_{A^*, B^*, D^*, K^*} \Pr_{A,B,D,K}\left[A = A^*, B = B^*, D = D^*, K = K^* \mid W = 1\right]\right). \tag{12}$$

We can see that for any A^*, B^*, D^*, K^*, the probability term can be rewritten as

$$\Pr[A = A^*, B = B^*, D = D^*, K = K^* \mid W = 1]$$
$$\leq \frac{\Pr[A = A^*, B = B^*, D = D^*, K = K^*]}{\Pr[W = 1]}$$
$$= \left(\Pr[W = 1] \cdot \binom{U}{n} \cdot \binom{n}{m} \cdot \binom{U-n}{m} \cdot \binom{m}{g} \cdot \binom{U-n-m}{m-g}\right)^{-1}, \tag{13}$$

where the binomial coefficients in the last line represent the number of ways to choose K, $(D \mid K)$, $(A \mid K)$, $(A \cap B \mid A)$, and $(B \setminus A \mid K, A)$, respectively.

Substituting (13) into (12), we get

$$H(M_{\text{learn}} \mid W = 1)$$
$$\geq \log \Pr[W = 1] + \log \binom{U}{n} + \log \binom{n}{m} + \log \binom{U-n}{m} + \log \binom{m}{g} + \log \binom{U-n-m}{m-g}.$$

Again, substituting this bound and (11) into (10) gives

$$\log \frac{1}{\Pr[W = 1]}$$
$$\geq \log \binom{U}{n} + \log \binom{n}{m} + \log \binom{U-n}{m} + \log \binom{m}{g} + \log \binom{U-n-m}{m-g}$$
$$- \log \binom{U}{m} - \log \binom{U}{n} - R - 2m(\beta+1)t - \log \binom{n+m}{m} - \gamma m \log U - \log \binom{n}{m}$$
$$= \log \binom{m}{g} + \log \binom{U-n-m}{m-g} - \left(\log \binom{U}{m} - \log \binom{U-n}{m}\right)$$
$$- \log \binom{n+m}{m} - R - 2m(\beta+1)t - \gamma m \log U. \tag{14}$$

We simplify the RHS with the following facts:

- $\log \binom{U}{m} - \log \binom{U-n}{m} \leq m \log \frac{U-m}{U-n-m} \leq m$ as long as n is sufficiently large.

- Similarly, $\log \binom{U-n-m}{m-g} \geq \log \binom{U}{m-g} - m$.

- $\log \binom{n+m}{m} \leq m \log \frac{e(n+m)}{m} \leq m \log \frac{2en}{m} \leq m\left(3 + \log \frac{n}{m}\right)$; since we required $m \geq n^{1-\gamma/2}$, this term is at most $\frac{\gamma}{2} m \log n + 3m \leq \frac{\gamma}{2} m \log U + 3m$.

- Recall that we have $g < \frac{m}{2}$, $m < \frac{U}{2}$, and $\gamma = \frac{\alpha}{6(1+\alpha)}$, so

$$\log \binom{U}{m-g} \geq \log \binom{U}{m/2} \geq \frac{m}{2} \log \frac{U}{m} \geq \frac{m}{2} \log \frac{U}{U^{1/(1+\alpha)}} = \frac{m}{2}\left(1 - \frac{1}{1+\alpha}\right) \log U = 3\gamma m \log U,$$

  where the third inequality holds as $m \leq n = U^{1/(1+\alpha)}$. Therefore, $\log \binom{U-n-m}{m-g} \geq \log \binom{U}{m-g} - m \geq 3\gamma m \log U - m$.

- $\log \binom{m}{g} \geq 0$, we just omit it.

Therefore, Eq. (14) can be rewritten as

$$\log \frac{1}{\Pr[W=1]} \geq (3\gamma m \log U - m) - m - \left(\frac{\gamma}{2} m \log U + 3m\right) - R - 2m(\beta+1)t - \gamma m \log U$$

$$\geq \frac{3}{2}\gamma m \log U - (5m + R + 2m(\beta+1)t).$$

Recall that we have $R + 5m + 2m(\beta+1)t \leq \frac{1}{2}\gamma m \log U$ as a premise of Lemma 5.1, so we have $\log(1/\Pr[W=1]) \geq \gamma m \log U$. Thus,

$$\Pr\left[A, B \text{ are consistent} \,\middle|\, |A \cap B| = g\right] = \Pr[W=1] \leq 2^{-\gamma m \log U} = U^{-\gamma m},$$

which concludes the proof of Lemma 5.1. Lemma 4.4 is a direct corollary of Lemma 5.1, as we only need to apply the law of total probability over $g < m/2$.

6   Extended Lower Bound

In the previous sections, we have proved a space-time lower bound for dynamic dictionaries: If the data structure has redundancy at most $n \log^{(k)} n$ (i.e., incurs at most $\log^{(k)} n$ wasted bits per key), its expected amortized time complexity is at least $\Omega(k)$. In this section, we extend this result to dictionaries with sublinear redundancy $R < n$. The result is stated as the following theorem, rephrasing Theorem 1.2 with respect to Distribution 1.

Theorem 6.1. *For any dynamic dictionary with redundancy $R \leq n$, running it on Distribution 1 takes at least $\Omega(\log(n/R))$ expected time per operation.*

Proof. The proof is similar to Section 3, based on a tree structure, but we will use different parameters.

Same as before, there are n (level-0) leaf nodes on the tree, each corresponds to a single meta-operation. Then, we let each level-1 node represent an interval of $m_1 := \max(R, n^{1-\gamma/2})$ consecutive meta-operations, i.e., it is a parent of m_1 leaf nodes.[6] Beyond level 1, we fix $\lambda = 2^{64}$ as a large constant and build a λ-ary tree, letting every node (except the root) have exactly λ children. Finally, there is a single root node at level $h = \log_\lambda(n/m_1) = \Omega(\log(n/R))$.[7]

---

[6]For simplicity, we assume that all parameters are integers, as is the case throughout this paragraph.

[7]If $m_1 = R$, it is clear that $h = \log_\lambda(n/R) = \Theta(\log(n/R))$; if $m_1 = n^{1-\gamma/2}$, we also have $h = \log_\lambda n^{\gamma/2} = \Theta(\log n) = \Omega(\log(n/R))$.

Recall that m_ℓ represents the number of operations within each level-ℓ node. For $1 \le \ell < h$, we have $m_\ell = m_1 \lambda^{\ell-1} \ge \max(R, n^{1-\gamma/2})$.

We apply Lemma 3.2 on every node in level $\ell \in [2, h]$. First, we verify the remaining premises in Lemma 3.2:

- $2^{64} \le \lambda \le \alpha \log n/12$. This condition is satisfied as long as n is sufficiently large.

- $m_\ell \log \lambda \ge 100R$. For every $\ell \ge 2$, the width m_ℓ satisfies $m_\ell \log \lambda > \lambda m_1 > 100R$.

- $\frac{1}{2}\gamma m_\ell \log U \ge R + 5m_\ell + 2(\beta+1)m_\ell t$. We still set $t = \frac{\gamma}{8(\beta+1)} \log U$, so we only need to show $\frac{1}{4}\gamma m_\ell \log U \ge R + 5m_\ell$. When n is sufficiently large, there is

$$\gamma m_\ell \log U \ge 24 m_\ell \ge 4(R + 5m_\ell).$$

- $m_\ell \ge n^{1-\gamma/2}$ holds due to the value of m_1.

Therefore, for the levels $\ell \in [2, h]$, we can apply Lemma 3.2 on every node. Finally, similar to the final step of Theorem 3.1, we can consider the following two cases to finish the proof:

- If there exists a level ℓ in which at least half of the nodes satisfy $\mathbb{E}[\mathsf{probe}_u] \ge \frac{1}{16}m_\ell t$, then we bound the expected time by $\frac{1}{2} \cdot \frac{n}{m_\ell} \cdot \frac{1}{16}m_\ell t = \frac{\gamma}{256(\beta+1)} \cdot n \log U$. As the coefficient is a constant and $\log U \ge \log(n/R)$, the expected time per operation is at least $\Omega(\log(n/R))$.

- Otherwise, for every level ℓ, at least half of the nodes in level ℓ satisfy $\mathbb{E}[\mathsf{cost}_u] \ge \frac{\gamma}{100}m_\ell$. Taking summation of all $\mathbb{E}[\mathsf{cost}_u]$, the total time cost is at least $\sum_{\ell=2}^{h} \frac{\gamma}{100}m_\ell \cdot \frac{1}{2} \cdot \frac{n}{m_\ell} = \frac{\gamma}{200}n(h-1)$ in expectation, so the expected time cost per operation is at least $\Omega(h) = \Omega(\log(n/R))$. □

7 Key-Value Reduction

So far we have only considered the lower bound for key-only dictionaries that store n different keys in the universe $[U]$. In this section, we will extend this result to dictionaries with values associated with keys:

Definition 7.1. *A* key-value dictionary, *denoted by D_{kv}, is a dictionary that stores n key-value pairs $(k, v) \in [U] \times [V]$, where all keys are distinct. D_{kv} supports insertions and deletions of key-value pairs. Moreover, when querying some key k, it returns whether k is present, together with the corresponding value v of k if k is in D_{kv}.*

In this definition, we call $[U]$ the key universe and $[V]$ the value universe. Without loss of generality, we assume that the word-size w satisfies $w = \log U + \log V$, so we can store both the key and the value in a single word.

The memory usage of D_{kv} can be represented by $\log \binom{U}{n} + n \log V + R$, where the first two terms are for succinct storage of n keys and n values, and R is the redundancy of D_{kv}. Below we will derive a reduction from D_{kv} to some key-only dictionary denoted by $D_{\mathrm{k\text{-}only}}$, and prove that we can use D_{kv} to simulate $D_{\mathrm{k\text{-}only}}$ with little additional time and space consumption. This theorem implies a time-space lower bound similar to the key-only version.

Theorem 7.2. *Assume $U = n^{1+\alpha}$ where α is a constant. For sufficiently large n, if we have a key-value dictionary D_{kv} with redundancy $R = \Omega(n^{1-\alpha/2})$ and running time T, then setting $U' = U \times V$, we can construct a key-only dictionary $D_{\mathrm{k\text{-}only}}$ on universe $[U']$, which can solve Distribution 1 with redundancy $R + o(R)$ and running time $T + O(1)$ in expectation.*

Proof. The intuition is to split a key $x \in [U']$ in $D_{\text{k-only}}$ into two parts (k, v) with length $\log U$ and $\log V$ bits, respectively. We regard the two parts as a key-value pair, and transform the operation on $D_{\text{k-only}}$ to an operation on D_{kv}. Roughly speaking, an insertion/deletion of x in $D_{\text{k-only}}$ is turned into an insertion/deletion of (k, v) in D_{kv}, and a query of x in $D_{\text{k-only}}$ will return "true" if and only if k is in D_{kv} while the corresponding value equals v.

If no key collisions occur, i.e., all pairs (k, v) from the transformation have different k, we can directly simulate $D_{\text{k-only}}$ with D_{kv} to complete all operations. In fact, among all $2n$ keys inserted in Distribution 1, the expected number of collisions is only $O(n^2/U) = O(n^{1-\alpha})$; we maintain a separate compact hash table to store those collided keys, which takes additional $O(n^{1-\alpha}w) = o(n^{1-\alpha/2}) = o(R)$ space and $O(1)$ update/query time. $\square$

When $R = \Theta(n^{1-\alpha/2})$, the above theorem combined with Theorem 6.1 already gives the best possible time lower bound $T \geq \Omega(\log(n/R)) = \Omega(\log n)$, which also applies for $R = o(n^{1-\alpha/2})$. Therefore, we deduce that the time-space trade-off of key-value dictionaries is not weaker than that of key-only dictionaries. Thus both Theorem 3.1 and Theorem 6.1 can be applied to show

- $T \geq \Omega(k)$ when $R = O(n \log^{(k)} n)$;

- $T \geq \Omega(\log(n/R))$ for all $R \leq O(n)$.

8 Update-Only Lower Bound

In this section, we extend the lower bound from the previous sections by relaxing the restriction on query time and focusing only on the trade-off between space and update time. Specifically, we prove that if the keys are associated with long values, the updates must follow the same time-space lower bound even if the queries are allowed to take arbitrarily long time.

First, note that although Distribution 1 combines a query, a deletion, and an insertion into every meta-operation, the existence of the query is only used once in the inner game. In Step 4 of the inner game protocol, Bob needs to learn the deleted set D by performing queries on $C_{\text{bef}}(S_B^*)$. The inclusion of queries in the meta-operations ensures that the cell set S_B^* is sufficient to answer queries for the true set D, allowing D to pass the test.

This observation suggests that, if we can modify the protocol of the inner game to avoid using queries in Step 4, we can eliminate queries from the hard distribution and prove a trade-off between space and update time (without query-time requirements). We demonstrate below that for the *key-value dictionary* with a relatively long value length, we can make such a modification and prove an update-only lower bound.[8]

Theorem 1.3 (Restated). *Consider a dynamic dictionary storing at most n keys from $[U]$, each associated with a value in $[V]$, with R bits of redundancy, where $U \geq 3n$ and $V = U^{2+\Theta(1)}/n^2$. Then, in the cell-probe model with word-size $w = \Theta(\log U)$,*

1. *if $R \geq n$ can be written as $R = O(n \log^{(k)} n)$ for $k \leq \log^* n$, then the expected update time is $\Omega(k)$;*

2. *if $R < n$, then the expected update time is $\Omega(\log(n/R))$.*

---

[8]However, the problem of proving an update-only lower bound for a key-only dictionary remains open and will be discussed in Section 10.

The proof of Theorem 1.3 is based on directly modifying the proof of Theorem 1.1, instead of reductions as in Section 7. It is worth noting that the combination of Theorem 7.2 and Theorem 1.3 covers all cases for key-value dictionaries: Theorem 7.2 works for the cases with long keys, i.e., $U = n^{1+\Theta(1)}$, while Theorem 1.3 covers the cases with short keys, i.e., $U = n^{1+o(1)}$ and $V = n^{\Theta(1)}$.

Proof. The proof of this theorem follows the same framework as the proof of Theorem 1.1. Similar to Distribution 1, the hard distribution used here is a sequence of n meta-operations, each consisting of a deletion and an insertion only (no query). When we insert a key, its associated value is sampled uniformly at random from $[V]$. We build a tree on top of these n meta-operations and assign each cell-probe to an internal node of the tree as before. We bound the total cost assigned to an internal node using the *outer lemma*.

The proof of the outer lemma remains the same as in Lemma 3.2, which makes use of a communication game (the *outer game*). Before the game starts, we give both Alice and Bob the associated values to all keys. During the game, Alice sends a message to tell Bob each key is inserted in which segment, where the length of the message is analyzed to complete the proof-by-contradiction. The proof of the outer lemma relies solely on the randomness of the *order* of inserted keys, so it works for short keys ($U = n^{1+o(1)}$) as well.

It remains to modify the inner lemma to avoid using query operations. The modified inner lemma is shown below. (Some of the constants are different from those in Lemma 4.4, but they are not essential to the application in the outer lemma.) Throughout the remainder of this section, we adopt the notations from Section 5 unless otherwise specified.

Lemma 8.1. *For integers U, V, n, m, g and real number $\gamma > 0$ satisfying $g < m/2$, $m < n/2$, $U \geq 3n$, $V \geq U^{2+5\gamma}/n^2$, assume the random variables $C_{\text{bef}}, K, \mathbf{d}, \pi, A, B$ are randomly sampled according to the procedure in Section 5, and assume the associated values of all keys are sampled independently and uniformly at random from $[V]$. Let $t > 0$ be a parameter such that the constraints $\frac{1}{2}\gamma m \log U \geq 2mt + 2R + 4m$ and $\mathbb{E}[|S_A|] \leq mt$ are satisfied, we must have*

$$\Pr\Big[A \text{ and } B \text{ are consistent} \,\Big|\, |A \cap B| = g\Big] \leq U^{-\gamma m}.$$

The proof of Lemma 8.1 leverages the randomness of the values in addition to the keys. Let $\mathbf{v}_A$ denote the values associated with keys in the inserted key set A. Similarly, we can define $\mathbf{v}_B$, $\mathbf{v}_D$, $\mathbf{v}_K$, $\mathbf{v}_{K \setminus D}$, and $\mathbf{v}_{A \cup B}$ for B, D, K, $K \setminus D$, and $A \cup B$, respectively. The inner game is modified as follows by making the key sequences pre-given.

- In addition to the pre-given fixed parameters g, m, n, U, and V, the sequences of keys $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{d}$ are also given to both Alice and Bob before the game. (Bob does not know K in advance.)

- Alice further knows C_{bef}, $\mathbf{v}_A$, $\mathbf{v}_B$, and $\mathbf{v}_D$ before the game.

- The goal of the game is to let Bob learn $\mathbf{v}_{A \cup B}$, $\mathbf{v}_K$, and $K \setminus D$ when A and B are consistent.

The protocol is as follows.

1. Check consistency. Alice sends $W := \mathbb{1}[A \text{ and } B \text{ are consistent}]$ to Bob, and the game terminates if $W = 0$.

2. Send $\mathbf{v}_B$. Alice directly sends $\mathbf{v}_B$ using $m \log V$ bits.

3. Send cell contents. Alice sends the artificial memory state C_{mix} using $Nw = n \log V + \log \binom{U}{n} + R$ bits. Note that Alice does not need to send the partition information by Bloomier Filter as before, since S_B can be learned by Bob in the next step.

4. Recover S_B and $C_B(S_B)$. As Bob already knows $\mathbf{b}$, $\mathbf{v}_B$ and $\mathbf{d}$, he can simulate the meta-operation sequence of $\mathbf{b}$ on C_{mix}. (Note that deleting a key from the data structure only requires knowledge of the key but not its value, thus Bob does not need $\mathbf{v}_D$ to complete the simulation.) We still let S_B represent the set of cells probed during this simulation, which Bob can learn via the simulation. Since $C_{\text{mix}}(S_B) = C_{\text{bef}}(S_B)$, the simulation process is the same as doing these operations on C_{bef}, so Bob can also learn $C_B(S_B)$, the memory state after the simulation.

5. Recover $\mathbf{v}_A, \mathbf{v}_{K \setminus D}$ and $K \setminus D$. Alice sends $S_A \cap S_B$ (conditioned on S_B) together with the difference between $C_A(S_A \cap S_B)$ and $C_B(S_A \cap S_B)$ using $2mt + \gamma m \log U$ bits in expectation. Then, Bob can learn C_A by combining $C_A(S_A \cap S_B)$ and $C_{\text{mix}}(\overline{S_A \cap S_B})$. He can further extract $\mathbf{v}_A, \mathbf{v}_{K \setminus D}$, and $K \setminus D$ from C_A.[9]

6. Recover $\mathbf{v}_D$. This step is different from Section 5 in the sense that we do not rely on query operations here.

   - First, Alice sends C_B: Since all the keys stored in C_B (i.e., $K \setminus D$ and $\mathbf{b}$) and their corresponding values ($\mathbf{v}_{K \setminus D}$ and $\mathbf{v}_B$) are known to Bob, Alice only needs to send C_B *conditioned on* these keys and values. This can be done using at most R bits.
   - Next, Bob computes C_{bef}: It is obtained by combining $C_{\text{bef}}(S_B) = C_{\text{mix}}(S_B)$ and $C_{\text{bef}}(\overline{S_B}) = C_B(\overline{S_B})$. The first equality holds due to the definition of C_{mix}, while the second holds because cells in $\overline{S_B}$ are not probed in the process of transforming C_{bef} to C_B.
   - Finally, Bob extracts $\mathbf{v}_D$ from C_{bef}.

Similar to Lemma 4.4, we can derive the following inequality from the modified inner game:

$$H(M_{\text{send}} \mid W = 1) \ \geq \ H(M_{\text{learn}} \mid W = 1). \tag{15}$$

To compute the left-hand side, we add up the messages sent in each step:

$$H(M_{\text{send}} \mid W = 1) \ \leq \ m \log V + \left(n \log V + \log \binom{U}{n} + R \right) + (2mt + \gamma m \log U) + R.$$

For the right-hand side, we can apply the same technique as in Lemma 4.4 to handle the condition

---

[9]Extracting information from the data structure can be done by querying all possible elements in the key universe. This can take arbitrarily long time.

$W = 1$. Here, $M_{\text{learn}} = (\mathbf{v}_{A \cup B}, \mathbf{v}_K, K \setminus D \mid \mathbf{a}, \mathbf{b}, \mathbf{d})$, so we can obtain

$$H(M_{\text{learn}} \mid W = 1)$$

$$\geq \log \left(1 \Big/ \max_{\mathbf{v}^*_{A \cup B}, \mathbf{v}^*_K, K^*} \Pr_{\mathbf{v}_{A \cup B}, \mathbf{v}_K, K} \left[(\mathbf{v}_{A \cup B}, \mathbf{v}_K, K \setminus D) = (\mathbf{v}^*_{A \cup B}, \mathbf{v}^*_K, K^* \setminus D) \,\Big|\, \mathbf{a}, \mathbf{b}, \mathbf{d}, W = 1 \right] \right)$$

$$\geq \log \left(\Pr[W = 1] \Big/ \Pr_{\mathbf{v}_{A \cup B}, \mathbf{v}_K, K} \left[(\mathbf{v}_{A \cup B}, \mathbf{v}_K, K \setminus D) = (\mathbf{v}^*_{A \cup B}, \mathbf{v}^*_K, K^* \setminus D) \,\Big|\, \mathbf{a}, \mathbf{b}, \mathbf{d} \right] \right)$$

$$= \log \Pr[W = 1] + (2m - g + n) \log V + \log \binom{U - 3m + g}{n - m}$$

$$\geq \log \Pr[W = 1] + (2m - g + n) \log V + \log \binom{U - 3m}{n - m}.$$

Plugging these bounds back into (15), we get

$$\log \frac{1}{\Pr[W = 1]} \geq (m - g) \log V + \log \binom{U - 3m}{n - m} - \log \binom{U}{n} - (\gamma m \log U + 2mt + 2R)$$

$$\geq \frac{1}{2} m \log V - \left(m \log \frac{U}{n} + 4m \right) - (\gamma m \log U + 2mt + 2R)$$

$$\geq \frac{5}{2} \gamma m \log U - (\gamma m \log U + 2mt + 2R + 4m)$$

$$\geq \gamma m \log U.$$

Here the second inequality is due to $g < m/2$ and

- $\log \binom{U}{n-m} - \log \binom{U-3m}{n-m} \leq (n - m) \log \frac{U - (n-m)}{U - 3m - (n-m)} \leq (n - m) \log \frac{2n+m}{2n-2m} \leq \frac{3m}{2 \ln 2} < 3m$ (here we used $\log(1 + x) \leq x/\ln 2$);

- $\log \binom{U}{n} - \log \binom{U}{n-m} \leq m \log \frac{U-n+m}{n-m+1} \leq m \log \frac{U}{n/2} = m \log \frac{U}{n} + m$.

The third inequality is due to the condition $V \geq U^{2+5\gamma}/n^2$, and the fourth inequality is due to the condition $\frac{1}{2} \gamma m \log U \geq 2mt + 2R + 4m$. This proves Lemma 8.1. $\qquad \square$

9 Lower Bounds for Related Problems

9.1 Strongly History-Independent Dictionaries

In this subsection, we show a brief overview of the lower bounds for strongly history-independent dictionaries. Recall that a strongly history-independent dictionary's memory state only depends on the current set of keys stored in it, and possibly some random bits; moreover, by Yao's Minimax Principle, we may assume without loss of generality that the algorithm is deterministic, which means we can fully recover its memory state by knowing only the current key set.

This fact benefits our Protocol 1 for the outer game: Bob already knows the starting state C_{st}, the keys to insert $\mathcal{A}$, and the keys to delete $\mathbf{d}$, which together can infer the key set K_{end} at the end, and further, the memory state C_{end} at the end. Hence, Alice no longer needs to send C_{end} to Bob, which costed R bits of information and was the only step involving the redundancy R in Section 4. After eliminating this cost, the proof in Section 4 works regardless of the redundancy R, except that it still relies on the Inner Lemma 4.4.

We do not change the statement (or the proof) of the inner lemma; however, it has a larger tolerance of redundancy R than the initial outer lemma: $R \leq \frac{1}{100}\gamma m_\ell \log U = \Theta(m_\ell \log U)$ suffices for its premise, where m_ℓ is the number of meta-operations the current node represents.

Recall that the entire proof is based on a tree on top of n meta-operations. Now, we set the tree parameters similarly to Section 6: the branching factor is a fixed large constant $\lambda = 2^{64}$, while every level-1 node (parent of leaves) represents $m_1 := \max\left(\frac{100R}{\gamma \log U}, n^{1-\gamma/2}\right)$ consecutive meta-operations. Under these parameters, the height of the tree is $\Theta\left(\log \frac{n \log U}{R}\right)$, and for every internal node of the tree,

$$\frac{1}{100}\gamma m_\ell \log U \geq \frac{1}{100}\gamma m_1 \log U \geq R,$$

which means the inner lemma's premise is satisfied, thus the outer lemma holds on levels $\ell \geq 2$. By a similar argument as the proof of Theorem 3.1, we conclude the following result.

Theorem 1.4 (Restated). *For $U = n^{1+\Theta(1)}$ and $R \geq 1$, any strongly history-independent dynamic dictionary storing at most n keys from $[U]$ with R bits of redundancy must have expected insertion, deletion or query time at least $\Omega\left(\log \frac{n \log U}{R}\right)$, in the cell-probe model with word-size $w = \Theta(\log U)$.*

9.2 Stateless Allocation

We recall the *stateless allocation* problem mentioned in Section 1: The algorithm is given a set $S \subseteq [U]$ of at most $(1-\varepsilon)n$ elements, and it should allocate these elements to n slots $\{1, 2, \ldots, n\}$, where each slot can accommodate at most one element. The allocation should be an injection from S to $[n]$ that only depends on the current set S as well as random bits r that are fixed over time, written $\sigma_{S,r} : S \to [n]$. When an insertion/deletion changes S to S', we define the expected switching cost to be

$$\mathbb{E}_r\left[\left|\{x \in S \cap S' \mid \sigma_{S,r}(x) \neq \sigma_{S',r}(x)\}\right|\right],$$

which equals the number of elements that change their assigned slots during the update. This problem is very similar to the *slot model* for dictionaries introduced in Section 2, where we have to assign n keys to n slots, with the following minor differences:

- Stateless allocation allows ε-fraction of the slots to be empty, while the slot model utilizes all n slots.

- Slot model additionally allows the mapping from keys to slots to be determined by not only the current key set but also $O(n)$ bits of redundancy, making it not strongly history-independent.

Below, we first show an $\Omega(\log n)$ lower bound on the expected switching cost of any stateless allocation algorithm with $\varepsilon = 0$, which one can think of as the slot model with 0 redundancy.

Proof Sketch for $\varepsilon = 0$. The proof structure is again similar to that of Theorem 3.1: We build a tree structure over n meta-operations each consisting of one deletion followed by one insertion, where the branching factor $\lambda = 2^{64}$ is a fixed constant for every node, which implies that the height of the tree equals $\Theta(\log n)$. When some key (element) is moved in two meta-operations $t_1 < t_2$ but not in between, we add one cost to the LCA of these two meta-operations on the tree. Then, using almost the same argument as Section 2, we are able to prove that the expected cost on every node u that represents m_ℓ meta-operations is at least $\Omega(m_\ell)$. Note that the branching factors used here are smaller than those we used in Section 2 to prove the $\Omega(\log^* n)$ lower bound for the slot model, and the proof still works because the redundancy $R = 0$. Finally, we sum up the cost on all internal nodes and conclude an $\Omega(\log n)$ lower bound. $\square$

31

When $\varepsilon > 0$, there are εn slots left empty. We add εn placeholder elements $\perp_1, \ldots, \perp_{\varepsilon n}$ and put $\perp_i$ into the i-th empty slot, which leads to an algorithm allocating $S \cup \{\perp_1, \ldots, \perp_{\varepsilon n}\}$ to all n slots, without leaving any slot empty, i.e., the new problem with placeholders is a stateless allocation problem with $\varepsilon = 0$.

We adapt the proof for $\varepsilon = 0$ to the new instance with placeholders. Again, we build a λ-ary tree over n meta-operations, and when some element is moved in two meta-operations $t_1 < t_2$ but not in between, we add one cost to the LCA of these two leaves: This cost is said to be a *real cost* if the moved element is a real element in S, or a *virtual cost* if the moved element is a placeholder. The proof of $\varepsilon = 0$ tells us that, for an internal node u representing m_ℓ meta-operations, the real and virtual cost add up to $\Omega(m_\ell)$. However, only real cost will cause switching cost of the initial stateless allocation problem.

Fortunately, the virtual cost on any node u, which is maximized when every placeholder is moved in all λ segments (children), cannot exceed $\lambda \varepsilon n$. When $m_\ell \geq c\varepsilon n$ for some large constant c, the sum of virtual and real cost is at least $\Omega(m_\ell) \geq 2\lambda \varepsilon n$, which implies that the real cost on this node is at least $\Omega(m_\ell) - \lambda \varepsilon n \geq \frac{1}{2} \cdot \Omega(m_\ell) = \Omega(m_\ell)$. This inequality can apply to the top $\Theta(\log \varepsilon^{-1})$ levels of the tree where $m_\ell \geq c\varepsilon n$, thus we conclude an $\Omega(\log \varepsilon^{-1})$ lower bound on the switching cost, as stated in the following theorem.

Theorem 1.5 (Restated). *For $1/n \leq \varepsilon \leq 1$ and $U \geq 3n$, any stateless allocation algorithm that assigns at most $(1-\varepsilon)n$ elements from universe $[U]$ to n slots must have expected switching cost at least $\Omega(\log \varepsilon^{-1})$.*

(Similar to Section 8, $U \geq 3n$ suffices for the proof, because the outer communication game only relies on the randomness of the *order* of inserted keys rather than the keys themselves.)

10   Discussions

We have proved tight bounds for dynamic succinct dictionaries in the previous sections. Now let us discuss the limitations of our methods.

The case when $U = n^{1+o(1)}$. Recall that all our conclusions are based on the assumption $U = n^{1+\alpha}$ for $\alpha = \Theta(1)$. If we allow slightly subconstant α here, the proof of Theorem 3.1 can only prove a bound of $\Omega(\alpha k)$ time when the wasted bits per key is $\log^{(k)} n$. This multiplicative factor α comes from Lemma 3.2 which gives the proposition $\mathsf{cost}_u \geq \gamma m/100$. When α was a constant, we could regard γ as a constant and sum up all cost_u to prove a time lower bound $\Omega(k)$; however, when $\alpha = o(1)$, $\gamma = \Theta(\alpha)$ is no longer a constant, so we can only obtain the bound $\Omega(\alpha k)$. This still implies an $\omega(1)$ time lower bound when $\alpha = \omega(1/\log^* n)$ under $O(1)$ wasted bits per key. In contrast, [BFCK$^+$22] showed that when $\alpha = 1/\log^{(t)} n$ for any constant t, one can achieve $o(1)$ wasted bits per key with constant running time. So the $\Omega(\alpha k)$ lower bound cannot be improved much. Similarly, when $R < n$, Theorem 6.1 is weakened to $\Omega\big(\min\{\alpha \log(n/R), \alpha^2 \log n\}\big)$.[10]

Key-only dictionaries with no query time constraints. In previous sections, we have obtained the time-space trade-off for key-only dictionaries with query time requirements, as well as key-value dictionaries without query time requirements. However, the problem of proving a lower bound for key-only dictionaries without query time constraints still remains open. Below is a simple example that shows why our method does not work for this goal.

---

[10]The second α factor comes from the tree depth $h \approx \min\{\log(n/R), \log n^{\gamma/2}\} = \Theta(\min\{\log(n/R), \alpha \log n\})$.

Recall that in Lemma 5.1, we claim that for two m-sized sets A, B, the probability that $C_A(S_A \cap S_B)$ and $C_B(S_A \cap S_B)$ has at most γm different cells is at most $U^{-\gamma m}$. The proof of this lemma makes use of the queries. If there are no queries in the meta-operations, we can construct a data structure that contradicts Lemma 5.1:

Suppose we use three cells C_1, C_2, C_3 to maintain keys $x_1, x_2, x_3 \in [0, 5U)$, allowing $O(1)$-bit redundancy. If $x_1 \in [0, 3U), x_2 \in [3U, 4U), x_3 \in [4U, 5U)$ (which happens with constant probability, and we call it the *good case*), we organize the keys according to the three cases listed in Table 1. Otherwise, we give up, succinctly store all keys, and check all cells for every operation. We store in an $O(1)$-bit extra memory whether the good case applies, and if so, which case in Table 1 applies. It is easy to see that the whole data structure incurs $O(1)$ bits of redundancy.

Table 1: Data structure organization under the good case.

	C_1	C_2	C_3
Case 1: $x_1 \in [0, U)$	$x_2 \oplus x_1$	x_3	$x_2 \oplus x_3$
Case 2: $x_1 \in [U, 2U)$	x_2	$x_3 \oplus x_1$	$x_2 \oplus x_3$
Case 3: $x_1 \in [2U, 3U)$	x_2	x_3	$x_2 \oplus x_3 \oplus x_1$

We let $m = 1$ in Lemma 5.1, i.e., there is only one meta-operation. We randomly delete a key and consider two possible keys to be inserted, denoted by x_A and x_B (they correspond to A, B in Lemma 5.1). With constant probability, we will be deleting x_1 and inserting back two new x_1's $(x_A, x_B \in [0, 3U))$, which means the good case still applies after both possible insertions. In this case, we denote the initial x_1 by x_{1D}, and denote the two inserted keys by $x_{1A} := x_A$ and $x_{1B} := x_B$.

While x_{1D}, x_{1A}, x_{1B} are selected uniformly at random, there is a constant probability that x_{1D}, x_{1A}, x_{1B} lead to Case 2, 1, and 3, respectively. Suppose we are deleting x_{1D} and inserting x_{1A}, we can complete the meta-operation by probing only C_1 and C_2:

- C_2 initially stores $x_3 \oplus x_{1D}$. We read x_{1D} from the deletion operation itself and read x_3 by probing C_2.

- We probe C_1 to read x_2.

- After knowing all three keys x_{1A}, x_2, x_3 that should be stored after the insertion, we directly write $x_2 \oplus x_{1A}$ and x_3 into C_1 and C_2, respectively. (Recall that Case 1 is the target case.)

Similarly, when we are deleting x_{1D} and inserting x_{1B}, we can probe only C_2 and C_3. The only commonly probed cell is $S_A \cap S_B = \{C_2\}$, whose content will be the same after both processes, which means that $C_A(S_A \cap S_B)$ and $C_B(S_A \cap S_B)$ have a constant probability to be equal. With a sufficiently large U, this example contradicts Lemma 5.1.[11] Thus, to prove a key-only lower bound without query time constraints, we need to develop a new method that does not rely on Lemma 5.1.

References

[ANS09]   Yuriy Arbitman, Moni Naor, and Gil Segev. De-amortized cuckoo hashing: Provable worst-case performance and experimental results. In *Proceedings of the 36th Inter-*

---

[11]Strictly speaking, it contradicts the following variant of Lemma 5.1: An $O(1)$-bit extra memory can be accessed which is not taken into account in the definition of consistency, and we only require U (but not n) to be sufficiently large. If the meta-operations include queries, then this variant can also be proved similarly to the proof of Lemma 5.1.

national Colloquium on Automata, Languages and Programming: Part I, ICALP '09, pages 107–118, Berlin, Heidelberg, 2009. Springer-Verlag.

[ANS10]   Yuriy Arbitman, Moni Naor, and Gil Segev. Backyard cuckoo hashing: Constant worst-case operations with a succinct representation. In *Proceedings of the 51st IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 787–796, 2010.

[AWY18]   Josh Alman, Joshua R. Wang, and Huacheng Yu. Cell-probe lower bounds from on-line communication complexity. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 1003–1012. ACM, 2018.

[AY20]   Josh Alman and Huacheng Yu. Faster update time for turnstile streaming algorithms. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1803–1813. SIAM, 2020.

[BCFC+21]   Michael A Bender, Alex Conway, Martin Farach-Colton, William Kuszmaul, and Guido Tagliavini. All-purpose hashing. *arXiv preprint arXiv:2109.04548*, 2021.

[BE20]   Ioana O. Bercea and Guy Even. A dynamic space-efficient filter with constant time operations. In *Scandinavian Workshop on Algorithm Theory*, 2020.

[BFCG+18]   Michael A. Bender, Martin Farach-Colton, Mayank Goswami, Rob Johnson, Samuel McCauley, and Shikha Singh. Bloom filters, adaptivity, and the dictionary problem. In *Proceedings of the 59th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 182–193, 2018.

[BFCK+22]   Michael A. Bender, Martín Farach-Colton, John Kuszmaul, William Kuszmaul, and Mingmou Liu. On the optimal time/space tradeoff for hash tables. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, pages 1284–1297, New York, NY, USA, 2022. Association for Computing Machinery.

[BHN19]   Sayan Bhattacharya, Monika Henzinger, and Stefan Neumann. New amortized cell-probe lower bounds for dynamic problems. *Theor. Comput. Sci.*, 779:72–87, 2019.

[BKP+20]   Aaron Berger, William Kuszmaul, Adam Polak, Jonathan Tidor, and Nicole Wein. Memoryless worker-task assignment with polylogarithmic switching cost. In *Proceedings of International Colloquium on Automata, Languages and Programming (ICALP)*, 2020.

[CJ11]   Raphaël Clifford and Markus Jalsenius. Lower bounds for online integer multiplication and convolution in the cell-probe model. In *Proceedings of the 38th International Colloquium on Automata, Languages and Programming (ICALP), Part I*, volume 6755 of *Lecture Notes in Computer Science*, pages 593–604. Springer, 2011.

[CJS15]   Raphaël Clifford, Markus Jalsenius, and Benjamin Sach. Cell-probe bounds for online edit distance and other pattern matching problems. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 552–561. SIAM, 2015.

[CJS16]   Raphaël Clifford, Markus Jalsenius, and Benjamin Sach. Cell-probe lower bounds for bit stream computation. In *Proceedings of the 24th Annual European Symposium on Algorithms (ESA)*, volume 57 of *LIPIcs*, pages 31:1–31:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016.

[CKRT04] Bernard Chazelle, Joe Kilian, Ronitt Rubinfeld, and Ayellet Tal. The Bloomier filter: An efficient data structure for static support lookup tables. In *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '04, pages 30–39, USA, 2004. Society for Industrial and Applied Mathematics.

[DdHPP06] Erik D. Demaine, Friedhelm Meyer auf der Heide, Rasmus Pagh, and Mihai Pătraşcu. De dictionariis dynamicis pauco spatio utentibus. In *LATIN 2006: Theoretical Informatics*, pages 349–361, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[DH90] Martin Dietzfelbinger and Friedhelm Meyer auf der Heide. A new universal class of hash functions and dynamic hashing in real time. In *Proceedings of the 17th International Colloquium on Automata, Languages and Programming (ICALP)*, pages 6–19, Berlin, Heidelberg, 1990. Springer-Verlag.

[DKM+88] M. Dietzfelbinger, A. Karlin, K. Mehlhorn, F.M. auf der Heide, H. Rohnert, and R.E. Tarjan. Dynamic perfect hashing: upper and lower bounds. In *Proceedings of the 29th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 524–531, 1988.

[FKS84] Michael L. Fredman, János Komlós, and Endre Szemerédi. Storing a sparse table with $O(1)$ worst case access time. *J. ACM*, 31(3):538–544, Jun 1984.

[FPSS03] Dimitris Fotakis, Rasmus Pagh, Peter Sanders, and Paul G. Spirakis. Space efficient hash tables with worst case constant access time. In *Proceedings of the 20th Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 271–282, Berlin, Heidelberg, 2003. Springer-Verlag.

[GKMT17] Michael T. Goodrich, Evgenios M. Kornaropoulos, Michael Mitzenmacher, and Roberto Tamassia. Auditable data structures. In *Proceedings of IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 285–300, 2017.

[Gol08] Daniel Golovin. *Uniquely Represented Data Structures with Applications to Privacy*. PhD thesis, Carnegie Mellon University, 2008.

[JLN19] Riko Jacob, Kasper Green Larsen, and Jesper Buus Nielsen. Lower bounds for oblivious data structures. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2439–2447. SIAM, 2019.

[Knu63] Don Knuth. Notes on "open" addressing, 1963. Available at `https://jeffe.cs.illinois.edu/teaching/datastructures/2011/notes/knuth-OALP.pdf`.

[Knu73] Donald E. Knuth. *The Art of Computer Programming, Volume III: Sorting and Searching*. Addison-Wesley, 1973.

[Kus23] William Kuszmaul. Strongly history-independent storage allocation: New upper and lower bounds. In *Proceedings of the 64th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2023.

[LLYZ23] Tianxiao Li, Jingxun Liang, Huacheng Yu, and Renfei Zhou. Dynamic dictionary with subconstant wasted bits per key. Manuscript, 2023.

[LMWY20] Kasper Green Larsen, Tal Malkin, Omri Weinstein, and Kevin Yeo. Lower bounds for oblivious near-neighbor search. In *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1116–1134. SIAM, 2020.

[LN18] Kasper Green Larsen and Jesper Buus Nielsen. Yes, there is an oblivious RAM lower bound! In *Proceedings of the 38th Annual International Cryptology Conference (CRYPTO), Part II*, volume 10992 of *Lecture Notes in Computer Science*, pages 523–542. Springer, 2018.

[LNN15] Kasper Green Larsen, Jelani Nelson, and Huy L. Nguyên. Time lower bounds for nonadaptive turnstile streaming algorithms. In *Proceedings of the 47th Annual ACM on Symposium on Theory of Computing (STOC)*, pages 803–812. ACM, 2015.

[LYY20] Mingmou Liu, Yitong Yin, and Huacheng Yu. Succinct filters for sets of unknown sizes. In *47th International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 168 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 79:1–79:19, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[NT01] Moni Naor and Vanessa Teague. Anti-persistence: History independent data structures. In *Proceedings of the 33th Annual ACM Symposium on Theory of Computing (STOC)*, pages 492–501, New York, NY, USA, 2001. ACM.

[Ove83] Mark H. Overmars. *The Design of Dynamic Data Structures*, volume 156 of *Lecture Notes in Computer Science*. Springer, 1983.

[Pag99] Rasmus Pagh. Low redundancy in static dictionaries with $O(1)$ worst case lookup time. In *Proceedings of the 26th International Colloquium on Automata, Languages and Programming (ICALP)*, volume 1644 of *Lecture Notes in Computer Science*, pages 595–604. Springer, 1999.

[Pag01] Rasmus Pagh. Low redundancy in static dictionaries with constant query time. *SIAM Journal on Computing*, 31(2):353–363, 2001.

[Păt08] Mihai Pătraşcu. Succincter. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 305–313. IEEE, 2008.

[PD04a] Mihai Pătraşcu and Erik D. Demaine. Lower bounds for dynamic connectivity. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC)*, pages 546–553. ACM, 2004.

[PD04b] Mihai Pătraşcu and Erik D. Demaine. Tight bounds for the partial-sums problem. In *Proceedings of the 15th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 20–29. SIAM, 2004.

[PD06] Mihai Pătraşcu and Erik D. Demaine. Logarithmic lower bounds in the cell-probe model. *SIAM J. Comput.*, 35(4):932–963, 2006.

[PR04] Rasmus Pagh and Flemming Friche Rodler. Cuckoo hashing. *Journal of Algorithms*, 51(2):122–144, 2004.

[RR03]   Rajeev Raman and Satti Srinivasa Rao. Succinct dynamic dictionaries and trees. In *Proceedings of the 30th International Conference on Automata, Languages and Programming (ICALP)*, pages 357–368, Berlin, Heidelberg, 2003. Springer-Verlag.

[WY16]   Omri Weinstein and Huacheng Yu. Amortized dynamic cell-probe lower bounds from four-party communication. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 305–314. IEEE Computer Society, 2016.

[Yao78]   Andrew Chi-Chih Yao. Should tables be sorted? (extended abstract). In *Proceedings of the 19th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 22–27. IEEE Computer Society, 1978.

[Yu16]   Huacheng Yu. Cell-probe lower bounds for dynamic problems via a new communication model. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 362–374. ACM, 2016.

[Yu20]   Huacheng Yu. Nearly optimal static las vegas succinct dictionary. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 1389–1401. ACM, 2020.