On Contention Resolution Protocols and Associated Probabilistic Phenomena[*]

P. D. MacKenzie *C. G. Plaxton* *R. Rajaraman*

Department of Computer Science
University of Texas at Austin

Abstract

Consider an on-line scheduling problem in which a set of abstract processes are competing for the use of a number of resources. Further assume that it is either prohibitively expensive or impossible for any two of the processes to directly communicate with one another. If several processes simultaneously attempt to allocate a particular resource (as may be expected to occur, since the processes cannot easily coordinate their allocations), then none succeed. In such a framework, it is a challenge to design efficient contention resolution protocols.

Two recently-proposed approaches to the problem of PRAM emulation give rise to scheduling problems of the above kind. In one approach, the resources (in this case, the shared memory cells) are duplicated and distributed randomly. We analyze a simple and efficient deterministic algorithm for accessing some subset of the duplicated resources. In the other approach, we analyze how quickly we can access the given (nonduplicated) resource using a simple randomized strategy. We obtain precise bounds on the performance of both strategies. We anticipate that our results will find other applications.

1 Introduction

Let k balls be thrown independently and uniformly at random into n bins, and let random variable X denote the maximum number of balls landing in any single bin. If $k = \Theta(n)$, it is well known that $X = \Theta(\lg n / \lg\lg n)$ wvhp. (Throughout this paper, we use "wvhp" to mean "with probability at least $1 - n^{-c}$ for *any* positive constant c", and we use "whp" to mean "with probability at least $1 - n^{-c}$ for *some* positive constant c".) If $k = \Theta(n \lg n)$, it is similarly well known that $X = \Theta(\lg n)$ wvhp. (These claims are straightforward to prove using standard bounds on the tail of the binomial distribution [7].) These sharp threshold phenomena have important consequences in a wide variety of hashing-related applications. In this paper, we explore two similar (but more complex) threshold phenomena, one arising in each of two fundamental families of contention resolution protocols.

To make our discussion of contention resolution protocols more concrete, we will focus our attention primarily on a single application, namely that of efficiently emulating an EREW PRAM on a c-collision crossbar network. We begin by reviewing the definitions of these two computational models. An *EREW PRAM* is a collection of n processors along with a global shared memory. Input and output are provided in the shared memory. In a single computational step, each processor can read or write one memory location. The sole restriction is that no two processors are allowed to access the same memory location in a single step. (If two processors attempt to access the same memory location in a single step, then the machine halts.)

[*] This research was supported by NSF Research Initiation Award CCR–9111591, and the Texas Advanced Research Program under Grant Nos. 003658–461 and 003658–480.

A *c-collision crossbar network* (or simply, a c-collision crossbar) is a more realistic model of parallel computation in which the global shared memory is distributed over n disjoint memory modules. Input and output are provided in the distributed memory. Each computational step consists of a read/write phase followed by an acknowledgement phase. During the read/write phase each processor can issue one read or write request for a specific memory location. If the total number of read/write requests involving memory locations stored in any particular memory module M is less than or equal to c, then all requests involving M succeed and are acknowledged during the acknowledgment phase. On the other hand, if more than c processors attempt to access memory locations stored in the same memory module M, then all requests involving M fail and no corresponding acknowledgements are sent. A *c-arbitrary crossbar* is defined in the same manner as a c-collision crossbar, except that if more than c processors attempt to access some memory module M, then an arbitrary subset of c of the requests succeed and are acknowledged.

The c-collision and c-arbitrary crossbar models have been studied previously under different names. (Our terminology is new.) The local memory PRAM model of Anderson and Miller [5], later studied under the name OCPC (optical communication parallel computer) in [10, 11, 12], corresponds to the 1-collision crossbar model. Valiant's S*PRAM model [24] corresponds to the 1-arbitrary crossbar model. Assuming a complete interconnection between processors and memory modules, the c-collision (resp., c-arbitrary) DMM models of [9] corresponds to the c-collision (resp., c-arbitrary) crossbar model.

We now return to the question of efficiently emulating an EREW PRAM on a c-collision crossbar. More specifically, assume that we wish to emulate a k-processor EREW PRAM on an n-processor c-collision crossbar, where k is some multiple of n. If we map k/n EREW PRAM processors to each processor of the crossbar, and employ a random hash function to map each location of the EREW PRAM shared memory to some memory module of the crossbar, we can easily see the connection between the random "balls and bins" experiment stated at the outset and the desired emulation: Each of the at most k read or write requests generated in a single step of the EREW PRAM computation corresponds to a ball, and each memory module corresponds to a bin.

If $k = n$ and $c = O(1)$, we can conclude that any scheme based on a single hash function requires $\Omega(\lg n/\lg\lg n)$ time to emulate one step of the EREW PRAM. On the other hand, Dietzfelbinger and Meyer auf der Heide [9] have recently shown that a bound of $O(\lg\lg n)$ time per EREW PRAM step is attainable for the same settings of k and c. They present a contention resolution protocol that minimizes the effect of the inevitable "hot-spot" memory modules (e.g., those memory modules receiving $\Theta(\lg n/\lg\lg n)$ requests under a given hash function) by employing three different hash functions. Thus, at the expense of increasing the total storage requirement by a factor of 3, the running time of the emulation is exponentially decreased. The fast performance of the Dietzfelbinger and Meyer auf der Heide emulation relies on a sharp threshold phenomenon that is the focus of Section 2. An overview of our results in this area is given in Section 1.1.

If $k = n\lg n$ and $c = O(1)$, the second "balls in bins" claim made at the outset of the paper shows that the emulation of a single EREW PRAM step will correspond to a $\Theta(\lg n)$-relation routing problem wvhp. (An *h-relation* routing problem is one in which each processor is the source of at most h packets and each memory module is the destination of at most h packets.) Thus, a variety of authors have considered the complexity of h-relation routing on the 1-collision crossbar. We are particularly interested in a natural randomized h-relation routing algorithm proposed by Geréb-Graus and Tsantilas [10]. The idea of their algorithm is to use randomization to break the symmetry between sets of processors with packets to be sent to the same same memory module: In a given "round" a processor with p packets left to send attempts to send a randomly chosen packet with probability roughly p/h, and does nothing with probability roughly $1 - p/h$. The resulting running time is $\Theta(h + \lg n \lg h)$. (Remark: the running time is stated as $\Theta(h + \lg n \lg\lg n)$ in [10], since they only considered the case $h \geq \log n$.) Thus, the algorithm leads to a work-optimal EREW PRAM emulation only for $h = \Omega(\lg n \lg\lg n)$.

The symmetry-breaking idea employed by Geréb-Graus and Tsantilas is central to many other randomized algorithms and communication protocols (e.g., the standard Ethernet protocol [20] and the classic ALOHA packet radio network protocol [1]). As we have seen, such randomized symmetry-breaking does not always lead to performance that is obviously optimal (i.e., that matches a trivial lower bound). (For example, we might have hoped that the algorithm of Geréb-Graus and Tsantilas would run in $\Theta(\lg n)$ time for the case $h = \Theta(\lg n)$.) The main result of the Section 3 is a tight lower bound on the running time of certain randomized symmetry-breaking procedures. In particular, with respect to the problem of randomized h-relation routing on the 1-collision crossbar, we have completely characterized the power of the natural symmetry-breaking paradigm. An overview of our results in this area is given in Section 1.2.

1.1 Redundancy-Based Protocols

In this section we give a brief overview of our emulation results related to protocols employing multiple hash functions. As mentioned earlier, Dietzfelbinger and Meyer auf der Heide [9] have presented a protocol using three hash functions that emulates an n-processor EREW PRAM in $O(\lg\lg n)$ time on an n-processor c-collision crossbar. Under their protocol, a read or write operation of memory location x by EREW PRAM processor i is emulated by having processor i of the c-collision crossbar access 2 out of the 3 copies corresponding to memory location x. (A similar protocol was presented in [17]; the idea that accessing 2 out of 3 copies is sufficient for the purposes of such an emulation was first used by Upfal and Wigderson [23].) The analysis presented in [9] requires some slack in the constants; in particular, they require $c \geq 3$, and are only able to analyze the protocol when it is used to emulate εn processors at a time, where ε is a sufficiently small positive constant. (Thus, the overall running time of the protocol is increased by a factor of $1/\varepsilon$.)

The protocol of [9] is easily generalized to the case where b hash functions are used, and each processor of the c-collision crossbar is required to access a out of b copies of a particular memory location, $a < b$. In Section 2.3, we focus on the case $a = 1$, and pinpoint the asymptotic complexity of the resulting protocol for all possible choices of the parameters b and c. (Furthermore, our analysis goes through with $\varepsilon = 1$, that is, we consider the most basic form of the protocol in which the action of all n EREW PRAM processors is emulated at once.) For $c = 1$, we prove that the protocol runs in $\Theta(\lg\lg n)$ time whp if $b \geq 3$. For $c = 1$ and $b = 2$, we prove that the protocol runs in $\Omega(\lg n)$ time wvhp. For any $b \geq 2$ and any constant $c \geq 2$, we prove that the protocol runs in $\Theta(\lg\lg n)$ time whp. (The protocol will run faster for non-constant c. It would not be difficult to extend our analysis to obtain tight bounds for non-constant c.) In the case of a c-arbitrary crossbar, the protocol runs in $\Theta(\lg\lg n)$ time whp for $b \geq 2$. In Section 2.4 we show that the above results hold even if the hash functions are only $O(\log^{\alpha} n)$-wise independent, where α is a real constant.

In Section 2.5, we observe that any "a out of b" problem with $a > 1$ can be efficiently reduced to a number of "1 out of ℓ" problems, where $\ell = 2$ or $\ell = 3$. Thus, we are able to easily upper bound the complexity of a (new) protocol for essentially any "a out of b" problem. One might suspect that a reduction of this sort, while making the analysis easier, is only doing so at the expense of a significant constant factor in performance. Interestingly, this is not the case; rather, as discussed in Section 2.5, our reduction yields a faster "a out of b" protocol than is obtained via the natural generalization of [9] for virtually all possible values of a and b.

The main idea underlying the aforementioned results may be explained as follows. While the process of throwing k balls independently and uniformly at random into n bins is well understood (e.g., we can easily compute a sharp bound on the number of bins receiving exactly one ball), we find that the sequence of distributions arising in the analysis of any "a out of b" protocol quickly deviates from such simple behavior. In [9, 21], this problem is attacked by studying certain structures related to the sequence of distributions. We are able to provide a more accurate analysis by precisely characterizing this sequence of distributions in terms of "truncated k balls in n bins" distributions (e.g., throw k balls into n bins and then remove all balls contained in bins with less than or equal to c balls).

1.2 Symmetry-Breaking Protocols

In this section we give a brief overview of our results on symmetry-breaking protocols in multiple access channels and for h-relation routing on a 1-collision crossbar.

There has been considerable effort in proving bounds on symmetry-breaking protocols to resolve contention in Ethernet-like multiple access channels [14, 15, 18, 19]. Specifically, it is assumed that some h of n stations wish to transmit to a single shared channel, but a station succeeds in its transmission if and only if it is the only station transmitting at that time. A symmetry-breaking protocol generates a schedule of transmission attempts for each station so that all h stations eventually transmit successfully. Previously studied protocols assume that all stations receive a feedback of 0, 1, or ≥ 2 at each step, depending on how many stations attempt to transmit. We call this the *Ethernet model*. For the Ethernet model, a lower bound of $\Omega((h/\log h)\log n)$ was shown for the time of any deterministic protocol [15], and it was shown that an $O(h\log n)$ time (non-adaptive) deterministic protocol exists [18]. We study protocols in which only the transmitting stations receive feedback (1 or ≥ 2) at a given step. We call the general contention resolution problem on this model the *Control Tower problem*. Protocols to solve the Control Tower problem correspond to non-adaptive protocols for contention resolution on the Ethernet model. Thus the lower bound of $\Omega((h/\log h)\log n)$ above applies to any deterministic solution to the Control Tower problem also. We show a slightly stronger result, that to send even one message (in the Control Tower model, or non-adaptive Ethernet model) requires $\Omega((h/\min\{\log h, \log\log n\})\log n)$ steps. From a technical standpoint, it is most natural to view the Control Tower problem as a problem on hypergraphs. Our lower bound relies on a combinatorial argument for extracting "thick" hypergraphs and another combinatorial argument showing the existence of contention-generating "near transversals".

Much faster protocols can be obtained for the Control Tower problem using randomization. For instance, the randomized protocol in Geréb-Graus and Tsantilas [10] solves the Control Tower problem in $O(h+\log h\log n)$ steps, wvhp. We show a tight lower bound of $\Omega(h+\log h\log n)$ for randomized protocols for the Control Tower problem that succeed with probability at least $1-n^{-3/4}$. (Naturally, this provides the same lower bound for non-adaptive randomized protocols in the Ethernet model.) Again, we find it technically useful to view the Control Tower problem as a problem on hypergraphs. The randomized lower bound then relies on a combinatorial argument for extracting "thick" hypergraphs (where the "thickness" quality is fundamentally different than that in the deterministic lower bound), and a probabilistic argument showing a non-trivial probability of the existence of contention-generating "near transversals" in random sets of vertices.

We now turn to the problem of direct h-relation routing on a 1-collision crossbar. In a *direct* algorithm for a given routing problem, the messages to be routed can only be sent directly from the source to the destination without any intermediate hops, and no additional information can be sent between the processors. Direct algorithms have the advantage of simplicity and low overhead. While non-direct algorithms may have better asymptotic behavior, it is likely that this improved asymptotic behavior is only achieved at the expense of large constant factors. The previously mentioned h-relation routing algorithm of Geréb-Graus and Tsantilas [10] is a direct algorithm, and direct h-relation routing algorithms have also been studied in [11, 12]. We refer the reader to these previous papers for further details. (For previous work on non-direct h-relation routing on the OCPC, see [5, 11, 12, 24]; for recent work that also incorporates redundancy-based techniques, see [13].)

There is a close correspondence between results for the Control Tower problem and results for direct h-relation routing on a 1-collision crossbar. In fact, the lower bound for deterministic protocols for the Control Tower problem directly gives the same lower bound for deterministic direct h-relation routing. The correspondence is not exact in terms of deterministic upper bounds, however, as the upper bound of

$O(h \log n)$ on the Control Tower problem, only indicates that there is an $O(h^2 \log n)$ deterministic direct h-relation routing algorithm. We show that this bound can be improved to $O(h \log h \log n)$, by slightly modifying the deterministic Control Tower protocol. Finally, we show that the tight lower bound for randomized protocols for the Control Tower problem can also be used to prove a tight lower bound of $\Omega(h + \log h \log n)$ for direct randomized h-relation routing on a 1-collision crossbar.

2 Multiple Hash Functions

In this section, we address the a out of b problem discussed in Section 1 on c-arbitrary and c-collision crossbars. In an *a out of b* problem on an n-processor crossbar, each shared memory cell is uniformly and independently hashed b times into the memory modules of the crossbar, and each processor has to access a out of the b copies of a particular shared memory cell.

Consider the 1 out of ℓ problem, $\ell \geq 2$. Let the ℓ hash functions be labeled h_i, $0 \leq i < \ell$, and the shared memory request of processor j be for cell x_j. Processor j needs to access one of the memory locations $h_i(x_j)$, $0 \leq i < \ell$. To solve this problem, the following simple sequence of ℓ rounds can be repeated until each processor has had one successful access: In the jth round, if processor i has not successfully accessed any copy of x_i, then processor i accesses $h_j(x_i)$. (This is analogous to Access Schedule 2 of [9], defined for the 2 out of 3 problem.) On a c-collision crossbar, processor j succeeds on its access if and only if there are at most $c-1$ other processors accessing the same memory module. Each round is executed in a synchronous fashion. We refer to this protocol as the 1 out of ℓ protocol.

We analyze the above process in an equivalent balls-and-bins setup. Let n balls labeled 0 through $n-1$ represent the messages, and n bins labeled 0 through $n-1$ represent the destinations. Each hash function, a random function from $[n]$ to $[n]$, is equivalent to a random throw of n balls uniformly and independently into n bins. Let h^A denote the function h with domain restricted to the set $A \subseteq [n]$. Let R_i denote the set of balls remaining after round i. For convenience, define R_{-1} to be the set of balls left before round 0, i.e., $R_{-1} = [n]$. Note that for $i \geq 0$, R_i is the subset of R_{i-1} given by the following recurrence:

$$R_i = \{j \in R_{i-1} : |f^{-1}(f(j))| > c\},$$

where $f = h^{R_{i-1}}_{i \bmod \ell}$.

Recall that a bag (or multiset) is an unordered set in which repetition is allowed. For any set A we define a bag B to be an *A-bag* if every element of B is also an element of A.

Let $\mathcal{F}_{m,n}$ denote the set of functions from $[m]$ to $[n]$. For each $f \in \mathcal{F}_{m,n}$, note that the bag $\{f(j) : j \in [m]\}$ is an m-size $[n]$-bag. Hence, the uniform distribution over $\mathcal{F}_{m,n}$ induces a probability distribution, which we denote $\mathcal{D}_{m,n}$, over the set of all m-size $[n]$-bags. For any bag B and $A \subseteq [n]$, let $\mathcal{B}_{A,B} = \{f^A : f \in \mathcal{F}_{n,n} \text{ and } f(A) = B\}$. Let S_i and T_i denote the bags $h_{i \bmod \ell}(R_{i-1})$ and $h_{i \bmod \ell}(R_i)$, respectively. Let $t_i = |T_i| = |R_i|$ (thus $t_{-1} = n$) and $s_i = |S_i|$. Note that $\langle t_i \rangle$ is a nonincreasing sequence. The protocol terminates after the first round i for which $t_i = 0$. The protocol fails to terminate if and only if $t_i = t_{i+\ell} > 0$ for some $i \geq -1$. (In such a case, the protocol enters an infinite loop with $t_j = t_i$ for all $j \geq i$.) The goal of our analysis is twofold: (i) to bound the probability that the protocol fails to terminate, and (ii) to analyze the number of rounds required by the protocol when it does terminate. We begin our analysis by establishing some properties of $\mathcal{D}_{m,n}$ and $\mathcal{B}_{A,B}$.

Let random variable X be drawn from $\mathcal{D}_{m,n}$, B be an arbitrary $[n]$-bag of size m, and m_i denote the number of copies of element i in B, $0 \leq i < n$. Then

$$\Pr[X = B] = \frac{m!}{m_0! \cdots m_{n-1}!} \cdot \frac{1}{n^m}. \tag{1}$$

Lemma 2.1 Let $0 \leq m < n$ and assume that X is a random variable drawn from $\mathcal{D}_{m+1,n}$. Let $Y \mid X = X \setminus \{x\}$, where x is an element of X chosen uniformly at random. Then Y is a random variable with probability distribution $\mathcal{D}_{m,n}$.

Proof: Let B be any m-size $[n]$-bag and $B_i = B \cup \{i\}$, $0 \leq i < n$. Let the number of copies of element i in B be m_i. (Hence $\sum_{i=0}^{n-1} m_i = m$.) Using Equation 1 we have

$$\begin{aligned}
\Pr[Y = B] &= \sum_{i=0}^{n-1} \Pr[X = B_i] \cdot \frac{m_i + 1}{m+1} \\
&= \sum_{i=0}^{n-1} \left(\frac{1}{n^{m+1}} \cdot \frac{(m+1)!}{(m_i+1)!} \prod_{j \neq i} \frac{1}{m_j!} \right) \frac{m_i + 1}{m+1} \\
&= \sum_{i=0}^{n-1} \frac{1}{n^{m+1}} \cdot \frac{m!}{m_0! \cdots m_{n-1}!} \\
&= \frac{m!}{m_0! \cdots m_{n-1}!} \cdot \frac{1}{n^m}.
\end{aligned}$$

■

Corollary 2.1.1 Let a, m, and n be integers such that $0 \leq t \leq m \leq n$. Let X be a random variable drawn from $\mathcal{D}_{m,n}$. Let $Y \mid X$ be a random a-size subbag of X. Then Y is a random variable with probability distribution $\mathcal{D}_{a,n}$.

■

Lemma 2.2 Let $h \mid R, B$ be a function drawn uniformly at random from $\mathcal{B}_{R,B}$. For arbitrary $A \subseteq R$, $h(A) \mid R, B$ is a random $|A|$-size subbag of B.

Proof: Consider an arbitrary element $x \in R$. Clearly $h(x)$ is a random element of B. Applying this for each element in A, $h(A)$ is a random $|A|$-size subbag of B. ■

Using Corollary 2.1.1 and 2.2, we prove the following claims related to the 1 out of ℓ protocol.

Lemma 2.3 For all $i \geq 0$, the random variable $h^{R_i}_{i \bmod \ell} \mid R_i, T_i$ is drawn uniformly at random from $\mathcal{B}_{R_i,T_i}$.

Proof: Since $h_{i \bmod \ell}$ is drawn uniformly at random from $\mathcal{F}_{n,n}$, $h^{R_i}_{i \bmod \ell} \mid R_i, T_i$ is drawn uniformly at random from the set of functions that map R_i to $[n]$-bag T_i, namely $\mathcal{B}_{R_i,T_i}$. ■

Lemma 2.4 Let S'_i denote the random variable $S_i \mid \{(R_j, T_j) : 0 \leq j < i\}$. For $0 \leq i < \ell$, S'_i is drawn from $\mathcal{D}_{t_{i-1},n}$ and for $i > \ell$, S'_i is a t_{i-1}-size random subbag of $T_{i-\ell}$.

Proof: By definition, $S_i = h_{i \bmod \ell}(R_{i-1})$. If $0 \leq i < \ell$, then $S'_i = h_{i \bmod \ell}(R_{i-1}) | R_{i-1}$. Let $S = h_{i \bmod \ell}([n])$. Thus S is drawn from $\mathcal{D}_{n,n}$ and $h_{i \bmod \ell} | S$ is drawn uniformly at random from $\mathcal{B}_{[n],S}$. By Lemma 2.2, for arbitrary R_{i-1}, $h_{i \bmod \ell}(M_{i-1}) | S$ is a random t_{i-1}-size subbag of S. Therefore, applying Corollary 2.1.1, with (t_{i-1}, n, n, S, S'_i) for (a, n, n, X, Y), we find that S'_i is drawn from $\mathcal{D}_{t_{i-1},n}$.

If $i \geq \ell$, then $S'_i = h^{R_{i-\ell}}_{i \bmod \ell}(R_{i-1}) | R_{i-\ell}, T_{i-\ell}, R_{i-1}$. By Lemma 2.3, $h^{R_{i-\ell}}_{i \bmod \ell} | R_{i-\ell}, T_{i-\ell}$ is drawn uniformly at random from $\mathcal{B}_{R_{i-\ell},T_{i-\ell}}$. By Lemma 2.2, for any R_{i-1}, $h^{R_{i-\ell}}_{i \bmod \ell}(R_{i-1}) | (R_{i-\ell}, T_{i-\ell})$ is a random t_{i-1}-size subbag of $T_{i-\ell}$. Hence, S'_i is a random t_{i-1}-size subbag of $T_{i-\ell}$. ■

Now we are ready to describe the protocol in terms of the S_i's and T_i's alone. Let *RandomBag*(m, n) return a bag drawn from $\mathcal{D}_{m,n}$. Let *RandomSubbag*(B, m) return a new bag that is a random m-size subbag of B. Let *PrunedBag*(B, c) return a bag that contains exactly those elements of S that have more than c copies. By Lemma 2.4, **Alg1**(n, ℓ, c) describes the random process occuring in the 1-out-of-ℓ protocol on a c-collision n-processor crossbar.

```
Alg1($n, \ell, c$)
(1.1)   $i := 0$;
(1.2)   repeat
(1.3)     if $i < \ell$ then
(1.4)       $S_i := RandomBag(|T_{i-1}|, n)$
(1.5)     else
(1.6)       $S_i := RandomSubbag(T_{i-\ell}, |T_{i-1}|)$;
(1.7)     $T_i := PrunedBag(S_i, c)$;
(1.8)     $i := i + 1$
(1.9)   until $|T_{i-1}| = 0$
```

In order to analyze **Alg1** we will estimate the size of T_i after round i. We propose a modified version of the above algorithm that simplifies the estimation of $|T_i|$. Observe that for $0 \leq i < \ell$, S_i is the bag obtained by throwing $|S_i|$ balls at random into n bins, and T_i is *PrunedBag*(S_i, c). Below we present the modified algorithm **Alg2**(n, ℓ, c) that approximately maintains this invariant after *every* round, under a suitable redefinition of S_i. The analysis in Section 2.3 will make this precise. **Alg2** is the same as **Alg1** except that Lines (1.5) and (1.6) are replaced by Lines (2.1) through (2.7), stated below.

```
(2.1)     else {
(2.2)       $S_i, T_i := S_{i-l}, T_{i-\ell}$;
(2.3)       while $|T_i| > |T_{i-1}|$ {
(2.4)         "Select $x$ at random from $S_i$";
(2.5)         $S_i, T_i := S_i \setminus \{x\}, T_i \setminus \{x\}$
(2.6)       }
(2.7)     };
```

Since each element x in line (2.4) is selected at random from S_i, any element selected from T_i is also random in T_i. Moreover exactly $|T_{i-1}|$ of the elements from $T_{i-\ell}$ are retained after the execution of the **while** loop.

Lemma 2.5 Let S_i^1, T_i^1 (resp., S_i^2, T_i^2) denote bags S_i, T_i in **Alg1** (resp., **Alg2**) after round i, $i \geq 0$. Then T_i^1 and T_i^2 have the same probability distribution.

Proof: We use induction on the number of rounds. For the basis, we observe that $T_0, \ldots, T_{\ell-1}$ in **Alg1** and **Alg2** are obtained in exactly the same way. (Lines (1.5) and (1.6) of **Alg1** and the corresponding lines (2.1) through (2.7) of **Alg2** are not executed.)

Consider round $i \geq \ell$. By the induction hypothesis T_j^1 and T_j^2 have the same probability distribution, $0 \leq j < i$. In Line (1.6), **Alg1** computes S_i^1 by selecting a random subbag of size $|T_{i-1}^1|$ from the subbag $T_{i-\ell}^1$. In Lines (2.3) through (2.6), **Alg2** computes S_i^2 by removing at random elements from $S_{i-\ell}^2$ until $|T_{i-1}^2|$ elements are retained from subbag $T_{i-\ell}^2$. Thus T_i^2 is a $|T_{i-1}^2|$-size subbag chosen randomly from $T_{i-\ell}^2$. By the induction hypothesis, the probability distribution of $T_{i-\ell}^1$ (resp., T_{i-1}^1) is the same as that of $T_{i-\ell}^2$ (resp., T_{i-1}^2). Therefore, S_i^1 after Line (1.6) of **Alg1** and T_i^2 after Line (2.6) of **Alg2** have the same probability distribution. Let S' (resp., T') denote S_i^2 (resp., T_i^2) after Line (2.6) of **Alg2** . Because $T_{i-\ell}^2$ contains all elements of $S_{i-\ell}^2$ with more than c copies, T' contains all elements of S' with more than c copies.

After Line (1.7), T_i^1 is the subbag of S_i^1 containing all elements with more than c copies, and T_i^2 is the subbag of S' containing all elements with more than c copies. Since T' contains all elements of S' with more than c copies, T_i^2 is the subbag of T' containing all elements with more than c copies. Therefore, the probability distribution of T_i^1 after round i is the same as that of T_i^2 after round i. ■

Corollary 2.5.1 The probability that **Alg1**(n, l, c) terminates after round i, $i \geq 0$, is equal to the probability that **Alg2**(n, l, c) terminates after round i. ■

In the remainder of this section, we analyze **Alg2** under different assignments to the parameters ℓ and c. In Section 2.1 we present some results on large deviations which we use for our analysis. In Section 2.2 we analyze certain "balls and bins" experiments. Section 2.3 uses these results to analyze **Alg2**$(n, \ell, 1)$. Among other results, we show that **Alg2**$(n, 3, 1)$ and **Alg2**$(n, 2, 2)$ each terminates in $\Theta(\log\log n)$ rounds whp. The analysis can be easily generalized to apply to **Alg2**(n, ℓ, c) for other values of ℓ and c. Section 2.5 presents simple ideas for extending the above results to general a out of b problems.

2.1 Large Deviations

For our analysis, we make frequent use of the Chernoff bounds for the tail of the binomial distributions [4, 7].

Theorem 1 Let X be a random variable drawn from $B(n, p)$, i.e., X is the number of successes in n independent Bernoulli trials, where each trial succeeds with probability p. Then,

$$\begin{aligned} \Pr[X \leq (1-\varepsilon)np] &\leq e^{-\varepsilon^2 np/2},\ 0 \leq \varepsilon \leq 1 && (2)\\ \Pr[X \geq (1+\varepsilon)np] &\leq e^{-\varepsilon^2 np/3},\ 0 \leq \varepsilon \leq 1 && (3)\\ \Pr[X \geq (1+\varepsilon)np] &\leq [e^{\varepsilon}(1+\varepsilon)^{-(1+\varepsilon)}]^{np} && (4) \end{aligned}$$

■

Lemma 2.6 Let S be a set of s balls, T be a subset of S, $t = |T|$, and $p = t/s$. Let s' balls be chosen uniformly at random from S, and t' be the random variable representing the number of balls that are chosen from T. Then, for any real $\varepsilon \geq 0$,

$$\begin{aligned} \Pr[t' \geq (p+\varepsilon)s'] &\leq e^{-2\varepsilon^2 s'}, \text{ and}\\ \Pr[t' \leq (p-\varepsilon)s'] &\leq e^{-2\varepsilon^2 s'}. \end{aligned}$$

Proof: By [8, 16],

$$\Pr[t' \geq (p+\varepsilon)s'] \leq e^{-2\varepsilon^2 s'}.$$

The lower bound on t' can be proved by using the upper bound on $s' - t'$. Thus,

$$\Pr[t' \leq (p-\varepsilon)s'] = \Pr[s' - t' \geq (1-p+\varepsilon)s'] \leq e^{-2\varepsilon^2 s'}.$$

■

Corollary 2.6.1 Let S be a set of s balls, and T be a subset of S, $t = |T|$. Let s' balls be chosen uniformly at random from S, and t' be the random variable representing the number of balls that are chosen from T. Then,

$$\begin{aligned} \Pr[t' \geq (1 + 1/(2\log^3 n))s't/s] &\leq e^{-s't^2/(2s^2\log^6 n)}, \text{ and}\\ \Pr[t' \leq (1 - 1/(2\log^3 n))s't/s] &\leq e^{-s't^2/(2s^2\log^6 n)}. \end{aligned}$$

Proof: Apply Lemma 2.6 with $\varepsilon = t/(2s\log^3 n)$. ■

Lemma 2.7 Let S be a set of s balls and T be a subset of S, $t = |T|$. Let s' balls be chosen at random from S, and let t' be the random variable representing the number of balls that are chosen from T. If $s't/s \geq \log^2 n$, then $t' \geq s't/(3s)$ wvhp.

Proof: Let $p = t/s$. Consider the s' balls being chosen in s' rounds - one ball in each round. If the number of balls chosen from bag T in rounds $1, \ldots, i-1$ is less than $ps'/3$, the probability that a ball from T is chosen in round i is at least $2p/3$. Let X be a random variable drawn from $B(s', 2p/3)$. The probability that $t' \geq ps'/3$ is at least the probability that $X \geq ps'/3$. By Equation (2), $\Pr[X \geq ps'/3] \geq 1 - e^{-ps'/12}$. Since $ps' \geq \log^2 n$, the lemma is proven. ■

In **Alg2**$(n, \ell, 1)$, T_i is that subbag of S_i, each element of which has at least 2 copies. We call such elements (as well as the associated balls) *non-singletons*. Similarly in **Alg2**$(n, \ell, 2)$, each element of T_i has at least 3 copies. We call these elements (and the associated balls) *non-pairs*. In Section 2.3, we show that the probability distribution of S_i is approximately $\mathcal{D}_{s_i,n}$. Thus, in **Alg2**$(n, \ell, 1)$ (resp., **Alg2**$(n, \ell, 2)$) t_i is approximately the number of non-singletons (resp., non-pairs) in a random bag drawn from $\mathcal{D}_{s_i,n}$. In order to get sharp estimates on the number of non-singletons and non-pairs in a random bag drawn from $\mathcal{D}_{m,n}$, we use a martingale analysis. The following two theorems are used to bound large deviations for martingales [4].

Theorem 2 Let $\Omega = A^B$ denote the set of functions $g : B \to A$. Fix a gradation $\emptyset = B_0 \subset B_1 \subset \cdots \subset B_m = B$. Let $L : \Omega \to \mathbf{R}$ be a functional. Define a martingale $X_0, \ldots, X_m$ by setting

$$X_i(h) = E[L(g) \,|\, g(b) = h(b) \text{ for all } b \in B_i].$$

Assume that for all i, whenever h and h' differ only on $B_{i+1} - B_i$, we have $|L(h') - L(h)| \leq 1$. Then $|X_{i+1}(h) - X_i(h)| \leq 1$, for all $0 \leq i < m$, $h \in \Omega$. ■

Theorem 3 (Azuma's Inequality) Let $X_0, \ldots, X_k$ be a martingale with $|X_{i+1} - X_i| \leq 1$, for all $0 \leq i < k$. Then for real $\lambda > 0$,

$$\Pr\left[|X_k - X_0| > \lambda\sqrt{k}\right] < 2e^{-\lambda^2/2}.$$

■

2.2 Lemmas on Balls and Bins

In this section, we estimate the number of non-singletons and non-pairs in a random bag with distribution $\mathcal{D}_{m,n}$ using some of the large deviations results mentioned in Section 2.1. By linearity of expectation, the expected number of non-singletons (resp., non-pairs) of a random bag X drawn from $\mathcal{D}_{m,n}$ is given by $f(m,n)$ (resp., $g(m,n)$), where

$$\begin{aligned} f(m,n) &= m\left(1 - \left(1 - \frac{1}{n}\right)^{m-1}\right), and \\ g(m,n) &= m\left(1 - \left(1 - \frac{1}{n}\right)^{m-1} - \frac{m-1}{n}\left(1 - \frac{1}{n}\right)^{m-2}\right). \end{aligned}$$

Throughout this section n will be fixed, so we use $f(m)$ (resp., $g(m)$) to denote $f(m,n)$ (resp., $g(m,n)$). Lemmas A.1 and A.2 show that $f(m) = \Theta(m^2/n)$, and $g(m) = \Theta(m^3/n^2)$. Let

$$\begin{aligned} \delta &= 1 - 1/\log^3 n, \text{ and} \\ \Delta &= 1 + 1/\log^3 n. \end{aligned}$$

Now we bound the probability that the number of non-singletons in a random bag drawn from $D_{m,n}$ deviates from the mean $f(m)$. Lemma 2.8 is used to bound deviations to within a $o(1)$ factor for m suitably large, and Lemma 2.9 bounds deviations to within a constant factor for all m.

Lemma 2.8 Let m, n be integers such that $3 \leq m \leq n$, and $h : [m] \to [n]$ be a random function drawn from $\mathcal{F}_{m,n}$, and $t(h)$ be the number of non-singletons in bag $h([m])$. If $m \geq n^{2/3} \log^3 n$, then $\delta f(m) \leq t(h) \leq \Delta f(m)$ wvhp.

Proof: Consider the martingale $X_0, \ldots, X_m$ defined as:

$$X_i(h) = E[t(p) \mid p \text{ and } h \text{ agree on balls in } [i]].$$

If two functions p and p' differ only on ball i, $t(p)$ and $t(p')$ differ by at most 2. We apply Theorem 2 by scaling the random variable t by 2 and thus obtain, $|X_{i+1} - X_i| \leq 2$. Similarly after scaling X_i's by 2, we apply Theroem 3 to get

$$\Pr[|X_m - X_0| > 2\lambda\sqrt{m}] < 2e^{-\lambda^2/2}. \tag{5}$$

The expected value X_0 of the functional t, is $f(m)$. For a function h, $t(h)$ is $X_m(h)$. By Equation 5 with $\lambda = f(m)/(2\sqrt{m}\log^3 n)$, we find that

$$\Pr\left[|t(p) - f(m)| > \frac{f(m)}{\log^3 n}\right] < 2e^{-f(m)^2/(8m\log^6 n)}.$$

Since for all $m > 2$, $f(m) \geq m^2/3n$,

$$\Pr\left[|t(p) - f(m)| > \frac{f(m)}{\log^3 n}\right] < 2e^{-m^3/(72n^2\log^6 n)}.$$

For $m \geq n^{2/3}\log^3 n$, $m^3/(72^2\log^6 n) \geq (\log^3 n)/72$. Therefore, $\delta f(m) \leq t(p) \leq \Delta f(m)$ wvhp. ■

Corollary 2.8.1 Let m and n be integers such that $3 \leq m \leq n$ and S be a random bag drawn from $\mathcal{D}_{m,n}$, and t be the number of non-singletons in S. If $m \geq n^{2/3}\log^3 n$, then $\delta f(m) \leq t \leq \Delta f(m)$ wvhp. ■

Lemma 2.9 Let m and n be integers such that $3 \leq m \leq n$ and S be a random bag drawn from $\mathcal{D}_{m,n}$. Let t represent the number of non-singletons in S. Then,

1. The probability that a particular ball is a non-singleton is at most m/n.
2. For $\sqrt{n}\log^5 n \leq m \leq n$, we have $t \leq 4m^2/n$ wvhp.
3. For $m \leq \sqrt{n}\log^5 n$, we have $t \leq 4\log^{10} n$ wvhp.

Proof: Let the m balls be thrown one-by-one. Since the balls occupy at most m bins, when a ball is thrown the probability that it falls into a non-empty bin is at most m/n. Thus the probability that a particular ball is a non-singleton is at most m/n. This establishes Part 1 of the lemma.

Let X be the random variable representing the number of balls that fall into non-empty bins. The number of non-singletons is at most $2S$. Hence, X is stochastically dominated by the random variable Y drawn from $B(m, m/n)$. The expected value of Y is m^2/n.

For $m \geq \sqrt{n}\log^5 n$, we apply Equation 3 with $\varepsilon = 1$, and obtain $\Pr[Y \geq 2s^2/n] \leq e^{-m^2/3n} \leq e^{-(\log^{10} n)/3}$. Therefore the number of non-singletons is at most $4m^2/n$ wvhp, proving Part 2 of the lemma.

For $m \leq \sqrt{n}\log^5 n$, we upper bound t by the number of non-singletons in a bag drawn from $\mathcal{D}_{\sqrt{n}\log^5 n, n}$. By Part 1, $t \leq 4\log^{10} n$ wvhp, proving Part 3 of the lemma. ■

The following two lemmas establish bounds on the number of non-pairs, analogous to Lemmas 2.8 and 2.9.

Lemma 2.10 Let m, n be integers such that $6 \leq m \leq n$. Let $h : [m] \to [n]$ be a random function drawn from $\mathcal{F}_{m,n}$, and $t(h)$ be the number of non-pairs in bag $h([m])$. If $m \geq n^{4/5} \log^3 n$, then $\delta g(m) \leq t(p) \leq \Delta g(m)$ wvhp.

Proof: Consider the martingale $X_0, \ldots, X_m$ defined as:

$$X_i(h) = E[t(p) \mid p \text{ and } S \text{ agree on balls in } [i]].$$

If two functions p and p' differ only on ball i, $t(p)$ and $t(p')$ differ by at most 3. We apply Theorem 2 by scaling the random variable t by 3 and thus obtain, $|X_{i+1} - X_i| \leq 3$. Similarly, after scaling X_i's by 3, we apply Theorem 3 to get

$$\Pr[|X_m - X_0| > 3\lambda\sqrt{m}] < 2e^{-\lambda^2/2}. \tag{6}$$

The expected value X_0 of the functional t, is $g(m)$. For a function h, $t(h)$ is $X_m(h)$. By Equation 6 with $\lambda = g(m)/(3\sqrt{m}\log^3 n)$, we find that

$$\Pr\left[|t(p) - g(m)| > \frac{g(m)}{\log^3 n}\right] < 2e^{-g(m)^2/(18m\log^6 n)}. \tag{7}$$

Since for $6 \leq m \leq n$, $g(m) \geq m^3/12n^2$,

$$\Pr\left[|t(p) - g(m)| > \frac{g(m)}{\log^3 n}\right] < 2e^{-m^5/(18\cdot 12^2 n^4 \log^6 n)}.$$

If $m \geq n^{4/5}\log^3 n$, $m^5/(18 \cdot 12^2 n^4 \log^6 n) \geq (\log^9 n)/18 \cdot 12^2$. Therefore, for $m \geq n^{4/5}\log^3 n$, $\delta g(m) \leq t(p) \leq \Delta g(m)$ wvhp. ■

Corollary 2.10.1 Let m, n be integers such that $6 \leq m \leq n$, and S be a random bag drawn from $D_{m,n}$, and t be the number of non-pairs in S. If $m \geq n^{4/5}\log^3 n$, then $\delta g(m) \leq t \leq \Delta g(m)$ wvhp. ■

Lemma 2.11 Let m, n be integers such that $6 \leq m \leq n$, and S be a random bag drawn from $D_{m,n}$. Let t be the random variable denoting the number of non-pairs in S.

1. The probability that a particular ball is a non-pair is at most $\max\{3m^2/n^2, 3(\log^{10} n)/n\}$.
2. For $n^{2/3}\log^3 n \leq m \leq n$, t is at most $12m^3/n^2$ wvhp.
3. For $m \leq n^{2/3}\log^3 n$, t is at most $12\log^9 n$ wvhp.

Proof: Let $m \geq \sqrt{n}\log^5 n$. Consider the experiment of throwing balls one-by-one into n bins until either there are $4m^2/n$ non-singletons or all the m balls have been thrown. Let t' be the number of non-pairs in this experiment. Since by Part 1 of Lemma 2.9, the number of non-singletons in a random bag from $\mathcal{D}_{m,n}$ is at most $4m^2/n$ wvhp, when the experiment terminates all the m balls have been thrown wvhp. Therefore any upper bound on t' wvhp (resp., whp) is an upper bound on t wvhp (resp., whp).

During the above experiment, the non-singletons occupy at most $2m^2/n$ bins. Therefore when a ball is thrown the probability that it falls into a bin with non-singletons (referred to as "non-singleton bins") is at most $2m^2/n^2$. Thus the probability that a particular ball is a non-pair is at most $2m^2/n^2 + 1/n^c$ for any real constant $c \geq 0$. Since $m \geq \sqrt{n}\log^5 n$, this probability is at most $3m^2/n^2$. This establishes Part 1 of the lemma. (Note that for $m \leq \sqrt{n}\log^5 n$ we can bound the probability by $3(\sqrt{n}\log^5 n)^2/n^2 = 3(\log^{10} n)/n$.)

Let X be the random variable representing the number of balls that fall into non-singleton bins. The number of non-pairs t' is at most $3X$. The random variable X is stochastically dominated by the random variable Z drawn from $B(m, \min\{1, 2m^2/n^2\})$. The expected value of Z is at most $2m^3/n^2$.

For $m \geq n^{2/3}\log^3 n$, we apply Equation 3 with $\varepsilon = 1$, and obtain $\Pr[Z \geq 4m^2/n] \leq e^{-2m^3/3n^2} \leq e^{-(2\log^9 n)/3}$. Therefore t' (and hence t) is at most $12m^3/n^2$ wvhp, establishing Part 2 of the lemma.

For $m \leq n^{2/3}\log^3 n$, we upper bound t by the number of non-pairs in a bag drawn from $\mathcal{D}_{n^{2/3}\log^3 n, n}$. By Part 1, $t \leq 12\log^9 n$ wvhp, establishing Part 3 of the lemma. ■

2.3 Analysis of Alg2

In this section, we analyze the number of rounds **Alg2** takes before termination. For $0 \leq i < \ell$, we have $s_i = t_{i-1}$. Corollaries 2.12.1 and 2.12.2, and Lemma 2.13 establish bounds on s_i in terms of the s'_j's, $0 \leq j < i$.

Lemma 2.12 In **Alg2**(n, ℓ, c) let $i \geq \ell$, $s_+ = \Delta s_{i-\ell} t_{i-1}/t_{i-\ell}$ and $s_- = \delta s_{i-\ell} t_{i-1}/t_{i-\ell}$. Then

$$\begin{aligned} \Pr[s_i \geq s_+] &\geq e^{-s+t_{i-\ell}^2/(2s_{i-\ell}^2 \log^6 n)}, \text{ and} \\ \Pr[s_i \leq s_-] &\leq e^{-s-t_{i-\ell}^2/(2s_{i-\ell}^2 \log^6 n)}. \end{aligned}$$

Proof: In round i, **Alg2** removes elements at random from $S_{i-\ell}$ until t_{i-1} elements are left from the subbag $T_{i-\ell}$ of $S_{i-\ell}$. Hence, $\Pr[s_i \geq s_+]$ equals the probability that less than t_{i-1} elements are left from $T_{i-\ell}$ after $s_{i-\ell} - s_+$ elements are removed. This is equal to the probability that less than t_{i-1} elements are chosen from $T_{i-\ell}$ in a random selection of s_+ elements from $S_{i-\ell}$. Applying Corollary 2.6.1 with $(s, t, s') = (s_{i-\ell}, t_{i-\ell}, s_+)$, the desired probability is at most $e^{-s+t_{i-\ell}^2/(2s_{i-\ell}^2 \log^6 n)}$. (Here we use the fact that for n sufficiently large, $(1 - 1/(2\log^3 n))\Delta \geq 1$.)

Similarly $\Pr[s_i \leq s_-]$ equals the probability that more than t_{i-1} elements are left from $T_{i-\ell}$ after $s_{i-\ell} - s_-$ elements are removed from $S_{i-\ell}$. This is equal to the probability that more than t_{i-1} elements are chosen from $T_{i-\ell}$ in a random selection of s_- elements from $S_{i-\ell}$. Applying Corollary 2.6.1 with $(s, t, s') = (s_{i-\ell}, t_{i-\ell}, s_-)$, the desired probability is at most $e^{-s-t_{i-\ell}^2/(2s_{i-\ell}^2 \log^6 n)}$. (Here we use that fact that for n sufficiently large, $(1 + 1/(2\log^3 n))\delta \leq 1$.) ■

Corollary 2.12.1 In **Alg2**(n, ℓ, c), if $i \geq \ell$, $s_{i-\ell} t_{i-1}/t_{i-\ell} \geq 2n^{2/3} \log^3 n$ and $t_{i-\ell} \geq s_{i-\ell}^2/4n$, then wvhp,

$$\delta s_{i-\ell} t_{i-1}/t_{i-\ell} \leq s_i \leq \Delta s_{i-\ell} t_{i-1}/t_{i-\ell}.$$

Proof: Let s_+, s_- be as defined in Lemma 2.12. By Lemma 2.12, we have

$$\Pr[s_i \geq \Delta s_{i-\ell} t_{i-1}/t_{i-\ell}] \leq e^{-s+t_{i-\ell}^2/(2s_{i-\ell}^2 \log^6 n)}.$$

Since $s_+, s_{i-\ell} \geq 2n^{2/3} \log^3 n$ and $t_{i-\ell} \geq s_{i-\ell}^2/4n$, the right hand side of the above inequality is at most $e^{-s+s_{i-\ell}^2/32n^2 \log^6 n} \leq e^{-\log^3 n/4}$. Similarly we can prove the desired lower bound on s_i wvhp using the lower bound in Lemma 2.12. (Note that $s_- \geq 2\delta n^{2/3} \log^3 n \geq n^{2/3} \log^3 n$ for n sufficiently large.) ■

Corollary 2.12.2 In **Alg2**(n, ℓ, c), if $s_{i-\ell} t_{i-1}/t_{i-\ell} \geq 2n^{4/5} \log^3 n$ and $t_{i-\ell} \geq s_{i-\ell}^3/13n^2$, then wvhp,

$$\delta s_{i-\ell} t_{i-1}/t_{i-\ell} \leq s_i \leq \Delta s_{i-\ell} t_{i-1}/t_{i-\ell}.$$

Proof: Let s_+, s_- be as defined in Lemma 2.12. By Lemma 2.12, we have

$$\Pr[s_i \leq \Delta s_{i-\ell} t_{i-1}/t_{i-\ell}] \leq e^{-s+t_{i-\ell}^2/(2s_{i-\ell}^2 \log^6 n)}.$$

Since $s_+, s_{i-\ell} \geq 2n^{4/5} \log^3 n$ and $t_{i-\ell} \geq s_{i-\ell}^3/13n^2$, the right hand side of the above inequality is at most $e^{-s+s_{i-\ell}^4/(2 \cdot 13^2 n^4 \log^6 n)} \leq e^{-(16 \log^9 n)/(\cdot 13^2)}$. Similarly we can prove the desired lower bound on s_i wvhp using the lower bound in Lemma 2.12. (Note that $s_- \geq 2\delta n^{4/5} \log^3 n \geq n^{4/5} \log^3 n$ for n sufficiently large.) ■

Lemma 2.13 Let $i \geq \ell$. In **Alg2**(n, ℓ, c), if $t_{i-1} \geq \log^2 n$, then $s_i \leq 3s_{i-\ell} t_{i-1}/t_{i-\ell}$ wvhp. If $t_{i-1} \leq \log^2 n$, then $s_i \leq 3s_{i-\ell}(\log^2 n)/t_{i-\ell}$ wvhp.

Proof: In **Alg2**, $\Pr[s_i \leq 3s_{i-\ell}t_{i-1}/t_{i-\ell}]$ is equal to the probability that more than t_{i-1} elements are selected from $T_{i-\ell}$ in a random selection of $3s_{i-\ell}t_{i-1}/t_{i-\ell}$ elements from $S_{i-\ell}$. If $t_{i-1} \geq \log^2 n$, then we apply Lemma 2.7 with $(s, t, s') = (s_{i-\ell}, t_{i-\ell}, 3s_{i-\ell}t_{i-1}/t_{i-\ell})$ to establish that $s_i \leq 3s_{i-\ell}t_{i-1}/t_{i-\ell}$ wvhp. Similarly $\Pr[s_i \leq 3s_{i-\ell}(\log^2 n)/t_{i-\ell}]$ is equal to the probability that more than t_{i-1} elements are selected from $T_{i-\ell}$ in a random selection of $3s_{i-\ell}(\log^2 n)/t_{i-\ell}$ elements from $S_{i-\ell}$. If $t_{i-1} \geq \log^2 n$, then we apply Lemma 2.7 with $(s, t, s') = (s_{i-\ell}, t_{i-\ell}, 3s_{i-\ell}(\log^2 n)/t_{i-\ell})$ to establish that $s_i \geq 3s_{i-\ell}(\log^2 n)/t_{i-\ell}$ wvhp. ■

Lemma 2.14 relates t_i to s_i for $i \geq \ell$.

Lemma 2.14 Let m balls be thrown uniformly and independently into n bins and S be the associated random bag. Let balls be removed at random from S until the remaining bag, denoted by S', satisfies condition C. Let X denote the set of balls that are non-singletons, $m = |S'|$, and $t' = |X|$. Let condition C be such that there exist integers d, u satisfying $d \leq m' \leq u$ wvhp.

1. If $d, u \geq n^{2/3} \log^3 n$, then $\delta f(d) \leq t' \leq \Delta f(u)$ wvhp.
2. If $u \geq \sqrt{n} \log^5 n$, then $t' \leq 4u^2/n$ wvhp.
3. If $u \leq \sqrt{n} \log^5 n$, then $t' \leq 4 \log^{10} n$ wvhp.
4. For any ball x, $\Pr[x \in X] \leq u^2/(mn) + 1/n^c$ for any real constant $c \geq 0$.

Proof: Consider the experiment of removing balls one-by-one at random from S. Let S_1 (resp., X_1) be the bag (resp., set of non-singleton balls) obtained when $m - u$ balls have been removed and S_2 be the bag obtained when $m - d$ balls have been removed. Therefore $|S_1| = u$ and $|S_2| = d$. Also, S_2 is a subbag of S_1. Wvhp, the condition C occurs after $m - u$ balls are removed and before $m - d$ balls are removed from S. Thus wvhp, S' is a subbag of S_1 and a superbag of S_2. Let t_1 (resp. t_2) denote the number of non-singletons in S_1 (resp., S_2). Hence $t_2 \leq t' \leq t_1$ wvhp. Note that by Corollary 2.1.1, S_1 and S_2 have probability distributions $\mathcal{D}_{u,n}$ and $\mathcal{D}_{d,n}$, respectively.

1. If $d, u \geq n^{2/3} \log^3 n$, then by Corollary 2.8.1, $t_2 \geq \delta f(d)$ and $t_1 \leq \Delta f(u)$ wvhp, thus establishing Part 1 of the lemma.
2. If $u \geq \sqrt{n} \log^5 n$, then by Part 2 of Lemma 2.9, $t_1 \leq 4u^2/n$ wvhp, thus establishing Part 2 of the lemma.
3. If $u \leq \sqrt{n} \log^5 n$, then by Part 3 of Lemma 2.9, $t_1 \leq 4 \log^{10} n$ wvhp. Hence $t' \leq 4 \log^{10} n$ wvhp, thus establishing Part 3 of the lemma.
4. For any ball x, $\Pr[x \in X] \leq \Pr[x \in X_1] + 1/n^c$ for any $c \geq 0$. By symmetry, the probability that x remains when u balls are left is u/m. Since S_1 is drawn uniformly at random from $\mathcal{D}_{u,n}$, by Part 1 of Lemma 2.9, $\Pr[x \in X_1] \leq (u/m)(u/n) = u^2/(mn)$, thus establishing Part 4 of the lemma.

■

Corollary 2.14.1 In **Alg2**$(n, \ell, 1)$, let $i \geq \ell$ and $d, u \geq 0$ be integers such that $d \leq s_i \leq u$ wvhp. If $t' = t_i$, then Parts 1 through 3 of Lemma 2.14 hold. Also, for any ball $x \in [n]$, the probability that x remains after round i is at most $(u^2/n^2) + 1/n^c$ for any real constant $c \geq 0$.

Proof: Fix integer $i \geq \ell$. Let $k = i \bmod \ell$. Consider the sequence of bags $\{S_{j\ell+k} \mid j \geq 0\}$ in **Alg2**. Bag S_k is obtained by throwing t_{k-1} (n if $k = 0$) balls into n bins. Bag $S_{j\ell+k}$, $j > 0$, is obtained by removing at random balls from $S_{(j-1)\ell+k}$ until $t_{(j-1)\ell+k-1}$ balls are left in a particular subbag $T_{(j-1)\ell+k}$ of $S_{(j-1)\ell+k}$.

Bag S_k can be obtained equivalently the following way: remove $n - t_{k-1}$ (0 if $k = 0$) balls at random from S that is a random bag drawn from $\mathcal{D}_{n,n}$. Thus each bag $S_{j\ell+k}$, $j \geq 0$ (S_i, in particular), can be viewed as having been obtained from bag S by removing balls at random until a certain condition (say C) holds. For bag S_i thus obtained, it is given that $d \leq |S_i| \leq u$ wvhp. We invoke Lemma 2.14 substituting $(S, S_i, s_i, t', n, d, u, C)$ for $(S, S', m', t', m, d, u, C)$ to establish the desired claims. ■

Similar to Corollary 2.14.1 we establish the following result for **Alg2**$(n, \ell, 2)$ using Lemma 2.11

Lemma 2.15 In **Alg2**$(n, \ell, 2)$, let $i \geq \ell$ and $d, u \geq 0$ be integers such that $d \leq s_i \leq u$ wvhp.

1. If $d, u \geq n^{4/5} \log^3 n$, then $\delta g(d) \leq t_i \leq \Delta g(u)$ wvhp.
2. If $u \geq n^{2/3} \log^3 n$, then $t_i \leq 12u^3/n^2$ wvhp.
3. If $u \leq n^{2/3} \log^3 n$, then $t_i \leq 12 \log^9 n$ wvhp.
4. For any $x \in [n]$ the probability that x remains after round i is at most $\max\{3u^3/n^3, (u \log^{10} n)/n^2\} + 1/n^c$ for any real constant $c \geq 0$.

■

2.3.1 Analysis of the 1-collision crossbar

Using results from Section 2.2, we show that the probability that **Alg2**$(n, \ell, 1)$ deviates significantly from the "expected" behavior is polynomially small. Let s'_i be defined as follows:

$$s'_i = \begin{cases} n & \text{if } i = 0, \\ f(s'_{i-1}) & \text{if } 0 < i < \ell, \text{ and} \\ s'_{i-\ell} \cdot \frac{f(s'_{i-1})}{f(s'_{i-\ell})} & \text{otherwise.} \end{cases}$$

Let $t'_i = f(s'_i)$ for all $i \geq 0$. (Note that for all $i \geq 0$, s'_i is the expected value of s_i given that $(s_j, t_j) = (s'_j, t'_j)$ for $0 \leq j < i$ and t'_i is the expected value of t_i given that $(s_j, t_j) = (s'_j, t'_j)$ for $0 \leq j < i$ and $s_i = s'_i$.)

Lemma 2.16 In **Alg2**$(n, \ell, 1)$, for all $0 \leq i \leq \frac{5}{2} \log_3 \log n$, if $s'_i \geq 4n^{2/3} \log^3 n$ and n is sufficiently large, then wvhp,

$$\delta^{3^i} s'_i \leq s_i \leq \Delta^{3^i} s'_i, \text{ and} \tag{8}$$

$$\delta^{2 \cdot 3^i + 1} t'_i \leq t_i \leq \Delta^{2 \cdot 3^i + 1} t'_i. \tag{9}$$

Proof: We use induction on i. For the basis, $i = 0$ and $s_0 = n = s'_0$. By Lemma 2.8, $\delta f(n) \leq t_0 \leq \Delta f(n)$ wvhp and since $t'_0 = f(n)$ the desired claims hold for $i = 0$.

Assume the claim holds for all $j < i$. We first establish Equation 8 from which we then derive Equation 9. We consider two cases. If $i < \ell$, then $s_i = t_{i-1}$. Since $s'_{i-1} \geq s'_i \geq 4n^{2/3} \log^3 n$, by the induction hypothesis, we have $\delta^{2 \cdot 3^{i-1}+1} t'_{i-1} \leq t_{i-1} \leq \Delta^{2 \cdot 3^{i-1}+1} t'_{i-1}$ wvhp. Since $3^i \geq 2 \cdot 3^{i-1} + 1$ for $i < \ell$, $\delta^{c^i} s'_i \leq s_i \leq s'_i \Delta^{c^i}$ wvhp.

If $i \geq \ell$, we use Corollary 2.12.1 to bound s_i. By induction hypothesis, since $s'_{i-1}, s'_{i-\ell} \geq 4n^{2/3} \log^3 n$,

$$\begin{aligned} \delta^{c^{i-\ell}} s'_{i-\ell} &\leq s_{i-\ell} \leq \Delta^{c^{i-\ell}} s'_{i-\ell}, \\ \delta^{2 \cdot 3^{i-\ell}+1} t'_{i-\ell} &\leq t_{i-\ell} \leq \Delta^{2 \cdot 3^{i-\ell}+1} t'_{i-\ell}, \text{ and} \\ \delta^{2 \cdot 3^{i-1}+1} t'_{i-1} &\leq t_{i-1} \leq \Delta^{2 \cdot 3^{i-1}+1} t'_{i-1}. \end{aligned}$$

Substituting appropriate bounds on $s_{i-\ell}, t_{i-\ell}$, and t_{i-1}, we get the following bounds on $s = s_{i-\ell}t_{i-1}/t_{i-\ell}$ wvhp:

$$\frac{\delta^{2\cdot 3^{i-1}+3^{i-\ell}+1} s'_{i-\ell}t'_{i-1}}{\Delta^{2\cdot 3^{i-\ell}+1} t'_{i-\ell}} \leq s \leq \frac{\Delta^{2\cdot 3^{i-1}+3^{i-\ell}+1} s'_{i-\ell}t'_{i-1}}{\delta^{2\cdot 3^{i-\ell}+1} t'_{i-\ell}}$$

By the fact that $\delta \leq \Delta^{-1}$, and $\Delta^2 \geq \delta^{-1}$ for n sufficiently large, we have

$$\frac{\delta^{2\cdot 3^{i-1}+3\cdot 3^{i-\ell}+2} s'_{i-\ell}t'_{i-1}}{t'_{i-\ell}} \leq s \leq \frac{\Delta^{2\cdot 3^{i-1}+5\cdot 3^{i-\ell}+3} s'_{i-\ell}t'_{i-1}}{t'_{i-\ell}} \tag{10}$$

Since $3^\ell \geq 2\cdot 3^{\ell-1}+9$ for $\ell \geq 3$, $3^i \geq 2\cdot 3^{i-1}+3\cdot 3^{i-\ell}+2$. Therefore $s \geq \delta^{3^i} s'_{i-\ell}t'_{i-1}/t'_{i-\ell} = \delta^{3^i} s'_i$ wvhp. Since $i \leq \frac{5}{2}\log_3 \log n$, $3^i \leq \log^{5/2} n$, and $\delta^{3^i} \geq \alpha$ for any real $\alpha < 1$ for n sufficiently large. Hence $s \geq 2n^{2/3}\log^3 n$. Next we show that $t_{i-\ell} \geq s^2_{i-\ell}/4n$ wvhp. By the induction hypothesis, $t_{i-\ell} \geq \delta^{2\cdot 3^{i-\ell}+1} t'_{i-\ell} = \delta^{2\cdot 3^{i-\ell}+1} f(s'_{i-\ell})$ wvhp. Since $f(s'_{i-\ell}) \geq (s'_{i-\ell})^2/3n$ and $s_{i-\ell} \leq \Delta^{3^{i-\ell}} s'_{i-\ell}$ wvhp, we have

$$t_{i-\ell} \geq \frac{\delta^{2\cdot 3^{i-\ell}+1} s^2_{i-\ell}}{3\Delta^{2\cdot 3^{i-\ell}} n} \geq \frac{\delta^{4\cdot 3^{i-\ell}+1} s^2_{i-\ell}}{3n},$$

wvhp. In the last step we use $\delta \leq \Delta^{-1}$. For any real $\alpha < 1$, $\delta^{3^{i-\ell}} \geq \delta^{3^i} \geq \alpha$ for n sufficiently large. Therefore, $\delta^{4\cdot 3^{i-\ell}+1} \geq 3/4$ for n sufficiently large and thus it follows that $t_{i-\ell} \geq s^2_{i-\ell}/4n$ wvhp.

Now we apply Corollary 2.12.1 to obtain $\delta s \leq s_i \leq \Delta s$ wvhp. By Equation 10, wvhp,

$$\frac{\delta^{2\cdot 3^{i-1}+3\cdot 3^{i-\ell}+3} s'_{i-\ell}t'_{i-1}}{t'_{i-\ell}} \leq s_i \leq \frac{\Delta^{2\cdot 3^{i-1}+5\cdot 3^{i-\ell}+4} s'_{i-\ell}t'_{i-1}}{t'_{i-\ell}}.$$

Since for $\ell \geq 3$, $3^\ell \geq 2\cdot 3^{\ell-1}+9$, we have $3^i \geq 2\cdot 3^{i-1} + 5\cdot 3^{i-\ell} + +4$ and $3^i \geq 2\cdot 3^{i-1} + 3\cdot 3^{i-\ell} + 3$. Since $s'_i = s'_{i-\ell}t'_{i-1}/t'_{i-\ell}$, Equation 8 holds wvhp.

Now we invoke Part 1 of Corollary 2.14.1 to obtain bounds on t_i. Note that $\delta^{3^i} s'_i, \Delta^{3^i} s'_i \geq n^{2/3}\log^3 n$ for n sufficiently large. Thus wvhp,

$$\delta f(\delta^{3^i} s'_i) \leq t_i \leq \Delta f(\Delta^{3^i} s'_i),$$

and hence by Corollary A.3.1,

$$\delta^{2\cdot 3c^i+1} f(s'_i) \leq t_i \leq \Delta^{2\cdot 3^i+1} f(s'_i).$$

Since $t'_i = f(s'_i)$, Equation 9 follows wvhp. ■

Lemma 2.16 implies that we can analyze **Alg2**$(n, \ell, 1)$ by studying how s'_i decreases as i increases.

Lemma 2.17 For all $0 \leq i < \ell$, we have

$$\prod_{0\leq j<i+1} s'_j = s'_0 \prod_{0\leq j<i} f(s'_j),$$

and for $i \geq \ell$, we have

$$\prod_{i-\ell+1\leq j<i+1} s'_j = s'_0 \prod_{i-\ell+1\leq j<i} f(s'_j).$$

Proof: For $0 \le i < \ell$, the desired claim follows directly from the definition of s'_j, $0 \le j < i+1$. Observe that for $i = \ell - 1$, we have $\prod\limits_{i-\ell+1 \le j < i+1} s'_j = s'_0 \prod\limits_{i-\ell+1 \le j < i} f(s'_{j-1})$. We use this fact as a basis for the case $i \ge \ell$. Assume that for $\ell - 1 \le k < i$, we have $\prod\limits_{k-\ell+1 \le j < k+1} s'_j = s'_0 \prod\limits_{k-\ell+1 \le j < k} f(s'_j)$. Then

$$\begin{aligned}
\prod_{i-\ell+1 \le j < i+1} s'_j &= \frac{s'_i}{s'_{i-\ell}} \prod_{i-\ell \le j < i} s'_j \\
&= \frac{s'_0 f(s'_{i-1})}{f(s'_{i-\ell})} \prod_{i-\ell \le j < i-1} f(s'_j) \\
&= s'_0 \prod_{i-\ell+1 \le j < i} f(s'_j).
\end{aligned}$$

■

Lemma 2.18 For all $1 \le i < \ell$, if s'_{i-1} and n are sufficiently large, then

$$\frac{1}{3^{i-1}} \prod_{0 \le j < i} \frac{s'_j}{n} \le \frac{s'_i}{n} \le \prod_{i \le j < i} \frac{s'_j}{n}.$$

For $i \ge \ell$, if s'_{i-1} and n are sufficiently large, then

$$\frac{1}{3^{\ell-1}} \prod_{i-\ell+1 \le j < i} \frac{s'_j}{n} \le \frac{s'_i}{n} \le \prod_{i-\ell+1 \le j < i} \frac{s'_j}{n}.$$

Proof: By Lemma 2.17 and Lemma A.1, if s'_{i-1} and n are sufficiently large, then for all $0 \le i < \ell$, we have

$$\frac{s'_0}{3^{i-1}} \prod_{0 \le j < i} \frac{(s'_j)^2}{n} \le \prod_{0 \le j < i+1} s'_j \le s'_0 \prod_{0 \le j < i} \frac{(s'_j)^2}{n},$$

and the claim of the lemma follows after dividing by $s'_0 \prod\limits_{0 \le j < i} s'_j$. By Lemma 2.17 and Lemma A.1, if s'_{i-1} and n are sufficiently large, then for all $i \ge 0$, we have

$$\frac{s'_0}{3^{\ell-1}} \prod_{i-\ell+1 \le j < i} \frac{(s'_j)^2}{n} \le \prod_{i-\ell+1 \le j < i+1} s'_j \le s'_0 \prod_{i-\ell+1 \le j < i} \frac{(s'_j)^2}{n},$$

and the claim of the lemma follows after dividing by $s'_0 \prod\limits_{i-\ell+1 \le j < i} s'_j$. ■

Lemma 2.18 can be used to analyze **Alg2**$(n, \ell, 1)$ for any $\ell \ge 2$. In the remainder of this subsection, we restrict our attention to the case $\ell = 3$. Let $w_i = \log_r(n/s_i)$ and $w'_i = \log_r(n/s'_i)$, where $r = n/f(n)$. (Note that $e/(e-1) \le r \le 2$ for all $n \ge 2$.)

Lemma 2.19 In **Alg2**$(n, 3, 1)$, for all $i > 0$, if $s'_{i-1} \ge 3$, then

$$w'_{i-2} + w'_{i-1} \le w'_i \le w'_{i-2} + w'_{i-1} + 2\log_r 3.$$

Proof: Follows directly from the definition of w'_i and Lemma 2.18. ■

Lemma 2.20 In **Alg2**$(n, 3, 1)$ for all $i > 0$, if $s'_{i-1} \ge 3$, then $p_1^{i-1} \le w'_i \le p_2^{i-1}$, where $p_1, p_2 > 1$ satisfy

$$p_1^2 - p_1 - 1 \le 0 \tag{11}$$

$$p_2^2 - p_2 - 2\log_r 3 - 1 \ge 0 \tag{12}$$

Proof: The proof is by induction on i. We have $s'_1 = n/r$, hence $p_1^0 = 1 = w'_1 = p_2^0$.

Let the claimed lower and upper bounds on w'_i hold for all $0 < j < i$, $i > 1$. By Lemma 2.19 and the induction hypothesis,

$$p_1^{i-3} + p_1^{i-2} \leq w'_i \leq p_2^{i-3} + p_2^{i-2} + 2\log_r 3$$

Equations 11 and 12, together with the fact that $p_2 > 1$, establish that $p_1^{k-1} \leq w'_k \leq p_2^{k-1}$. ■

We now place tight bounds on the number of rounds **Alg2**$(n, 3, 1)$ takes before termination.

Lemma 2.21 There exists integer $j = O(\log\log n)$ such that $s_j \leq n^{2/5}$ wvhp in **Alg2**$(n, 3, 1)$.

Proof: Let $\phi = (1+\sqrt{5})/2$. Since $\phi^2 - \phi - 1 = 0$, Lemma 2.20 implies that $w'_i \geq \phi^{i-1}$ for all $i > 0$. Let $k = \min\{i : w'_i \geq \log_r(\frac{n^{1/3}}{4\log^3 n})\}$. For $i = \lceil \log_\phi \log_r \frac{n^{1/3}}{4\log^3 n} \rceil + 1$, $w'_i \geq \log_r(\frac{n^{1/3}}{4\log^3 n})$. Therefore, $k \leq \log_\phi \log_r \frac{n^{1/3}}{4\log^3 n} + 2$. (Also note that since $w'_2 \leq 1 + 2\log_r 3$, $k \geq 3$ for n sufficiently large.) Since $\phi^{5/2} > 3$, $k \leq 5/2 \log_3 \log n$ for n sufficiently large. Thus, Equations 8 and 9 of Lemma 2.16 hold for all $i < k$. (Also note that $s'_k = n/r^{w'_k} \leq 4n^{2/3}\log^3 n$.)

By Lemma 2.16, $t_{k-1} \geq \delta^{2\cdot 3^{k-1}+1} t'_{k-1}$ wvhp. Since $t'_{k-1} = f(s'_{k-1}) \geq (s'_{k-1})^2/3n$, $t_{k-1} \geq 16\delta^{2\cdot 3^{k-1}+1}(n^{1/3}\log^6 n)/3 \geq \log^2 n$ wvhp for n sufficiently large. By Lemma 2.13, $s_k \leq 3s_{k-3}t_{k-1}/t_{k-3}$ wvhp. Substituting appropriate bounds on s_{k-3}, t_{k-3}, and t_{k-1} from Lemma 2.16, we have wvhp

$$s_k \quad \leq \quad 3\Delta^{2\cdot 3^{k-1}+5\cdot 3^{k-3}+4} s'_k \leq 3\Delta^{3^k} s'_k \leq 4s'_k. \tag{13}$$

The last step follows from the fact that $\Delta^{3^k} < \alpha$ for any real $\alpha < 1$ and for n sufficiently large. We consider two cases depending on the value of s'_k.

Case 1: $s'_k \leq \sqrt{n}\log^5 n$. By Equation 13, $s_k \leq 4\sqrt{n}\log^5 n$ wvhp. Therefore, by Part 3 of Lemma 2.14.1, $t_k \leq 64\log^{10} n$ wvhp. We consider two cases. If $t_k \geq \log^2 n$, by Lemma 2.13, $s_{k+1} \leq 3s_{k-2}t_k/t_{k-2}$ wvhp. If $t_k \leq \log^2 n$, then $s_{k+1} \leq 3s_{k-2}\log^2 n/t_{k-2}$. In any case, $s_{k+1} \leq (192 s_{k-2}\log^{10} n)/t_{k-2}$ wvhp. Now we substitute appropriate bounds on s_{k-2} and t_{k-2} from Lemma 2.16 and obtain wvhp,

$$\begin{aligned} s_{k+1} \quad &\leq \quad \frac{192\Delta^{3^{k-2}} s'_{k-2}\log^{10} n}{\delta^{2\cdot 3^{k-2}+1} t'_{k-2}} \\ &\leq \quad \frac{576 n \Delta^{5\cdot 3^{k-2}+2}\log^{10} n}{s'_{k-2}} \\ &\leq \quad 144\Delta^{3^k} n^{1/3}\log^7 n \\ &\leq \quad n^{2/5} \end{aligned}$$

for n sufficiently large. (Note: The penultimate step follows from the fact that $3^k \geq 5\cdot 3^{k-2} + 2$ and $s'_{k-2} \geq 4n^{2/3}\log^3 n$.)

Case 2: $s'_k \geq \sqrt{n}\log^5 n$. By Equation 13, $s_k \leq 4s'_k$ wvhp. We again consider two cases, depending on whether $t_k \leq \log^2 n$ or $t_k \geq \log^2 n$.

If $t_k \leq \log^2 n$ then Lemma 2.13 implies that $s_{k+1} \leq 3s_{k-2}\log^2 n/t_{k-2}$ wvhp. Arguing as in Case 2, s_{k+1} is at most $n^{2/5}$ wvhp.

If $t_k \geq \log^2 n$ then Lemma 2.13 implies that $s_{k+1} \leq 3s_{k-2}t_k/t_{k-2}$ wvhp. Since $s_k \leq 4s'_k$, by Part 2 of Lemma 2.14.1, $t_k \leq 64(s'_k)^2/n \leq 192t'_k$ wvhp. Substituting this upper bound on t_k and appropriate bounds on s_{k-2} and t_{k-2} obtained from Lemma 2.16, we have $s_{k+1} \leq 1000s'_{k+1}$ wvhp for n sufficiently large. We now derive an upper bound on s'_{k+1}.

By Lemma 2.19, $w'_k \leq w'_{k-1} + w'_{k-2} + 2\log_r 3$. Since $w'_{k-1} \geq w'_{k-2}$, we have

$$w'_{k-1} \geq \frac{1}{2}(w'_k - 2\log_r 3) \geq \frac{\log_r n}{6} - \frac{3\log_r \log n}{2} - \log_r 6.$$

Thus, by Lemma 2.19,

$$\begin{aligned} w'_{k+1} &\ge w'_k + w'_{k-1} \\ &\ge \frac{\log_r n}{2} - \frac{9 \log_r \log n}{2} - \log_r 24, \text{ and} \\ s'_{k+1} &\le \sqrt{n} \log^5 n, \end{aligned}$$

for n sufficiently large. Now we apply an analysis similar to Case 2 with k replaced by $k+1$ to establish that s_{k+2} is at most $n^{2/5}$ wvhp.

Cases 1 and 2 establish that wvhp after $j = k+2 = O(\log\log n)$ rounds s_j is at most $n^{2/5}$ wvhp. ■

Lemma 2.22 For any ball $x \in [n]$, the probability that x remains after $O(\log\log n)$ rounds of **Alg2**$(n, 3, 1)$ is at most $2/n^{6/5}$ for n sufficiently large.

Proof: By Lemma 2.21, after $j = O(\log\log n)$ rounds, $s_j \le n^{2/5}$ wvhp. By Corollary 2.14.1, the probability that x remains after round j is at most $2n^{4/5}/n^2$ for n sufficiently large. Since $2n^{4/5}/n^2 = 2/n^{6/5}$, the desired claim follows. ■

The following theorem is an easy consequence of the above lemma.

Theorem 4 **Alg2**$(n, 3, 1)$ terminates in $O(\log\log n)$ rounds whp. ■

Theorem 5 **Alg2**$(n, 3, 1)$ terminates in $\Omega(\log\log n)$ rounds wvhp.

Proof: A possible solution to Equation 12 is $p_2 = 3$. Thus by Lemma 2.20, $w'_i \le 3^{i-1}$ for all $i > 0$. After $k = \lfloor \log_3((\log_r n)/4) \rfloor$ rounds, $w'_k \le (\log_r n)/4$ and $s'_k \ge n^{3/4}$. For n sufficiently large $n^{3/4} \ge 4n^{2/3} \log^3 n$. Therefore, by Lemma 2.16, $t_k \ge \delta^{2 \cdot 3^k + 1} t'_k \ge \delta^{2 \cdot 3^k + 1} (s'_k)^2 / 3n > 0$ for n sufficiently large. This shows that **Alg2**$(n, 3, 1)$ executes at least $\log_3((\log_r n)/4) \ge \log_3((\log n)/4) = \Omega(\log\log n)$ rounds before termination. ■

The recurrence in Lemma 2.18 for $\ell = 2$ yields $s'_{i+1}/n \ge s'_i/3n$ for all $i \ge 0$. Thus $w'_i = O(i)$. Using an analysis similar to the above theorem we establish an $\Omega(\log n)$ lower bound for **Alg2**$(n, 2, 1)$.

Theorem 6 **Alg2**$(n, 2, 1)$ terminates in $\Omega(\log n)$ rounds wvhp. ■

2.3.2 Analysis of the 2-collision crossbar

The analysis of **Alg2** with the collision factor set to 2 is similar to the 1-collision case. Analogous to Section 2.3.1 we define s'_i as follows:

$$s'_i = \begin{cases} n & \text{if } i = 0, \\ g(s'_{i-1}) & \text{if } 0 < i < \ell, \\ s'_{i-\ell} \cdot \frac{g(s'_{i-1})}{g(s'_{i-\ell})} & \text{otherwise.} \end{cases}$$

For all $i \ge 0$ let $t'_i = g(s'_i)$.

Lemma 2.23 Let c be the positive root of $c^2 = 4c + 13$. In **Alg2**$(n, \ell, 2)$, for all $0 \le i \le (11/4) \log_c \log n$, if $s'_i \ge 4n^{4/5} \log^3 n$, then wvhp,

$$\delta^{c^i} s'_i \le s_i \le \Delta^{c^i} s'_i \tag{14}$$

$$\delta^{4c^i+1} t'_i \le t_i \le \Delta^{4c^i+1} t'_i \tag{15}$$

$$\tag{16}$$

Proof: We use induction on i. For the basis, $i = 0$ and $s_0 = n = s'_0$. By Corollary 2.10, $\delta g(n) \leq t_0 \leq \Delta g(n)$ wvhp and since $t'_0 = g(n)$, the desired claims hold for $i = 0$.

Assume the claim holds for all $j < i$, $i \geq 1$. We first establish Equation 14, from which we then derive Equation 15. We consider two cases: $i < \ell$ and $i \geq \ell$.

If $i < \ell$, then $s_i = t_{i-1}$ and $s'_i = t'_{i-1}$. Since $s'_{i-1} \geq s'_i \geq 4n^{4/5}\log^3 n$, by the induction hypothesis, $\delta^{4c^{i-1}+1}t'_{i-1} \leq t_{i-1} \leq \Delta^{4c^{i-1}+1}t'_{i-1}$ wvhp. Since $c \geq 5$, we have $c^i \geq 4c^{i-1}+1$ for all $i \geq 1$. Hence $\delta^{c^i}s'_i \leq s_i \leq \Delta^{c^i}s'_i$ wvhp. If $i \geq \ell$, we use Corollary 2.12.2 to bound s_i. By the induction hypothesis, since $s'_{i-1}, s'_{i-\ell} \geq 4n^{4/5}\log^3 n$, we have wvhp,

$$\begin{aligned}
\delta^{c^{i-\ell}}s'_{i-\ell} &\leq s_{i-\ell} \leq \Delta^{c^{i-\ell}}s'_{i-\ell}, \\
\delta^{4c^{i-\ell}+1}t'_{i-\ell} &\leq t_{i-\ell} \leq \Delta^{4c^{i-\ell}+1}t'_{i-\ell}, \text{ and} \\
\delta^{4c^{i-1}+1}t'_{i-1} &\leq t_{i-1} \leq \Delta^{4c^{i-1}+1}t'_{i-1}.
\end{aligned}$$

Substituting appropriate bounds on $s_{i-\ell}, t_{i-\ell}$, and t_{i-1}, we obtain the following bounds on $s = s_{i-\ell}t_{i-1}/t_{i-\ell}$ wvhp:

$$\frac{\delta^{4c^{i-1}+c^{i-\ell}+1}s'_{i-\ell}t'_{i-1}}{\Delta^{4c^{i-\ell}+1}t'_{i-\ell}} \leq s \leq \frac{\Delta^{4c^{i-1}+c^{i-\ell}+1}s'_{i-\ell}t'_{i-1}}{\delta^{4c^{i-\ell}+1}t'_{i-\ell}}.$$

Since $\delta \leq \Delta^{-1}$, and $\Delta^2 \geq \delta^{-1}$ we have

$$\frac{\delta^{4c^{i-1}+5c^{i-\ell}+2}s'_{i-\ell}t'_{i-1}}{t'_{i-\ell}} \leq s \leq \frac{\Delta^{4c^{i-1}+9c^{i-\ell}+3}s'_{i-\ell}t'_{i-1}}{t'_{i-\ell}}. \qquad (17)$$

Since $\ell \geq 2$, we have $c^\ell \geq 4c^{\ell-1}+13$. Hence, $c^i \geq 4c^{i-1}+5c^{i-\ell}+2$. Therefore $s \geq \delta^{c^i}s'_{i-\ell}t'_{i-1}/t'_{i-\ell} = \delta^{c^i}s'_i$ wvhp. Since $i \leq (11/4)\log_c \log n$, we have $c^i \leq \log^{11/4} n$ and $\delta^{c^i} \geq \alpha$ for any real $\alpha < 1$ for n sufficiently large. Hence $s \geq 2n^{4/5}\log^3 n$ wvhp. Next we show that $t_{i-\ell} \geq s^3_{i-\ell}/(13n^2)$ wvhp. By the induction hypothesis, $t_{i-\ell} \geq \delta^{4c^{i-\ell}+1}t'_{i-\ell} = \delta^{4c^{i-\ell}+1}g(s'_{i-\ell})$ wvhp. Since $g(s'_{i-\ell}) \geq (s'_{i-\ell})^3/12n^2$ and $s_{i-\ell} \leq \Delta^{c^{i-\ell}}s'_{i-\ell}$ wvhp, we have

$$t_{i-\ell} \geq \frac{\delta^{4c^{i-\ell}+1}s^3_{i-\ell}}{12\Delta^{3c^{i-\ell}}n^2} \geq \frac{\delta^{7c^{i-\ell}+1}s^3_{i-\ell}}{12n^2},$$

wvhp. For any real $\alpha < 1$, $\delta^{c^{i-\ell}} \geq \delta^{c^i} \geq \alpha$ for n sufficiently large. Therefore, $\delta^{7c^{i-\ell}+1} \geq 12/13$ for n sufficiently large and thus $t_{i-\ell} \geq s^3_{i-\ell}/(13n^2)$ wvhp.

Now we apply Corollary 2.12.2 to obtain $\delta s \leq s_i \leq \Delta s$ wvhp. By Equation 17,

$$\frac{\delta^{4c^{i-1}+5c^{i-\ell}+3}s'_{i-\ell}t'_{i-1}}{t'_{i-\ell}} \leq s_i \leq \frac{\Delta^{4c^{i-1}+9c^{i-\ell}+4}s'_{i-\ell}t'_{i-1}}{t'_{i-\ell}}.$$

Since $c^\ell \geq 4c^{\ell-1}+13$, we have $c^i \geq 4c^{i-1}+5c^{i-\ell}+3$ and $c^i \geq 4c^{i-1}+9c^{i-\ell}+4$. Hence Equation 8 holds wvhp.

Now we invoke Part 1 of Lemma 2.15 to obtain bounds on t_i. Note that $\delta^{c^i}s'_i, \Delta^{c^i}s'_i \geq n^{4/5}\log^3 n$ for n sufficiently large. Thus wvhp,

$$\delta g(\delta^{c^i}s'_i) \leq t_i \leq \Delta g(\Delta^{c^i}s'_i),$$

and hence by Corollary A.4.1,

$$\delta^{4c^i+1}g(s'_i) \leq t_i \leq \Delta^{4c^i+1}g(s'_i).$$

Since $t'_i = g(s'_i)$, Equation 9 follows wvhp. ■

Lemmas 2.24 and 2.25 determine the rate at which s'_i decreases with increasing i. Using Lemma 2.23, we can then determine the rate of change of s_i as i increases.

Lemma 2.24 For all $0 \le i < \ell$, we have

$$\prod_{0 \le j < i+1} s'_j = s'_0 \prod_{0 \le j < i} g(s'_j),$$

and for $i \ge \ell$, we have

$$\prod_{i-\ell+1 \le j < i+1} s'_j = s'_0 \prod_{i-\ell+1 \le j < i} g(s'_j).$$

Proof: Similar to the proof of Lemma 2.17. ■

Lemma 2.25 For all $1 \le i < \ell$, if s'_{i-1} and n are sufficiently large, then

$$\frac{1}{12^{i-1}} \prod_{0 \le j < i} \left(\frac{s'_j}{n}\right)^2 \le \frac{s'_i}{n} \le \prod_{i \le j < i} \left(\frac{s'_j}{n}\right).$$

For $i \ge \ell$, if s'_{i-1} and n are sufficiently large, then

$$\frac{1}{12^{\ell-1}} \prod_{i-\ell+1 \le j < i} \left(\frac{s'_j}{n}\right)^2 \le \frac{s'_i}{n} \le \prod_{i-\ell+1 \le j < i} \left(\frac{s'_j}{n}\right)^2.$$

Proof: By Lemma 2.24 and Lemma A.2, if s'_{i-1} and n are sufficiently large, then for all $0 \le i < \ell$, we have

$$\frac{s'_0}{12^{i-1}} \prod_{0 \le j < i} \frac{(s'_j)^3}{n^2} \le \prod_{0 \le j < i+1} s'_j \le s'_0 \prod_{0 \le j < i} \frac{(s'_j)^3}{n^2},$$

and the claim of the lemma follows after dividing by $s'_0 \prod\limits_{0 \le j < i} s'_j$. By Lemma 2.17 and Lemma A.1, if s'_{i-1} and n are sufficiently large, then for all $i \ge 0$, we have

$$\frac{s'_0}{12^{\ell-1}} \prod_{i-\ell+1 \le j < i} \frac{(s'_j)^3}{n^2} \le \prod_{i-\ell+1 \le j < i+1} s'_j \le s'_0 \prod_{i-\ell+1 \le j < i} \frac{(s'_j)^3}{n^2},$$

and the claim of the lemma follows after dividing by $s'_0 \prod\limits_{i-\ell+1 \le j < i} s'_j$. ■

In the remainder of this section, we restrict our attention to the case where $\ell = 2$. We set $r = n/g(n)$. (Note that $e/(e-2) \le r \le 9$ for $n \ge 3$.)

Lemma 2.26 In **Alg2**$(n, 2, 2)$, for all $i > 0$, if $s'_{i-1} \ge 6$, then

$$2w'_{i-1} \le w'_i \le 2w'_{i-1} + \log_r 12$$

Proof: Follows directly from the definition of w'_i and Lemma 2.25. ■

Lemma 2.27 In **Alg2**$(n, 2, 2)$, for all $i > 0$, if $s'_{i-1} \ge 6$, then $p_1^{i-1} \le w'_i \le p_2^{i-1}$, where $p_1 \le 2$ and $p_2 \ge 2 + \log_r 12$.

Proof: The proof is by induction on i. We have $s'_1 = n/r$, hence $p_1^0 = 1 = w'_1 = p_2^0$.

Let the claimed lower and upper bounds on w'_i hold for all $0 < j < i$, $i > 1$. Then by Lemma 2.26, and the induction hypothesis,

$$p_1^{i-3} + p_1^{i-2} \le w'_i \le p_2^{i-3} + p_2^{i-2} + 2\log_r 3$$

Since $p_1 \le 2$ and $p_2 \ge 2 + \log_r 12$, it follows that $p_1^{k-1} \le w'_k \le p_2^{k-1}$. ■

Now we place tight bounds on the number of rounds **Alg2**$(n, 2, 2)$ takes before termination.

Lemma 2.28 There exists integer $j = O(\log\log n)$ such that $s_j \le n^{5/8}$ wvhp in $\mathbf{Alg2}(n, 2, 2)$.

Proof: By Lemma 2.27, $w'_i \ge 2^{i-1}$ for all $i > 0$. Let $k = \min\{i \mid w'_i \ge \log_r \frac{n^{1/3}}{4\log^3 n}\}$. Therefore $k \le \lceil \log\log_r \frac{n^{1/5}}{4\log^3 n}\rceil + 1 \le \log\log_r \frac{n^{1/5}}{4\log^3 n} + 2$. (Also note that since $w'_1 = 1$, we have $k \ge 2$ for n sufficiently large.) Now we apply Lemma 2.23 with $\ell = 2$. Let c be the root of the equation $c^2 = 4c + 13$. Since $2^{11/4} > c$, $k \le (11/4)\log_c \log_r n$ for n sufficiently large. Therefore by Lemma 2.23, $t_{k-1} \ge \delta^{4c^{k-1}+1} t'_{k-1}$ wvhp. Since $t'_{k-1} \ge (s'_{k-1})^3/(12n^2)$, we have $t_{k-1} \ge \delta^{4c^{k-1}+1}(16n^{2/5}\log^9 n)/3 \ge \log^2 n$ for n sufficiently large. By Lemma 2.13, $s_k \le 3s_{k-2}t_{k-1}/t_{k-2}$ wvhp. Substituting the appropriate bounds on s_{k-2}, t_{k-1} and t_{k-2} given by Lemma 2.23, we have wvhp,

$$\begin{aligned} s_k &\le \frac{3\Delta^{4c^{k-1}+c^{k-2}+1} s'_{k-2} t'_{k-1}}{\delta^{4c^{k-2}+1} t'_{k-2}} \\ &\le \frac{3\Delta 4c^{k-1} + 9c^{k-2} + 3s'_{k-2} t' k - 1}{t'_{k-2}}. \end{aligned}$$

Since $c^2 = 4c + 13$, we have $c^k \ge 4c^{k-1} + 9c^{k-2} + 3$. Therefore,

$$s_k \le 3\Delta^{c^k} s'_k \le 4s'_k, \tag{18}$$

wvhp for n sufficiently large.

We consider two cases, depending on whether $t_k \le \log^2 n$ or $t_k > \log^2 n$.

If $t_k \le \log^2 n$, then by Lemma 2.13, $s_{k+1} \le 3s_{k-1}\log^2 n / t_{k-1}$ wvhp. Substituting appropriate bounds on s_{k-1} and t_{k-1} given by Lemma 2.23, we have wvhp,

$$\begin{aligned} s_{k+1} &\le \frac{3\Delta^{c^{k-1}} s'_{k-1} \log^2 n}{\delta^{4c^{k-1}+1} t'_{k-1}} \\ &\le \frac{36\Delta^{9c^{k-1}+2} n^2 \log^2 n}{(s'_{k-1})^2} \\ &\le \frac{9\Delta^{c^{k+1}} n^{2/5}}{4\log^4 n} \\ &\le \frac{92^c n^{2/5}}{4\log^4 n} \\ &\le n^{5/8} \end{aligned}$$

for n sufficiently large. (The second step follows from the lower bound on t'_{k-1} given by Lemma A.2. In the third step we use $s'_{k-1} \ge n^{4/5}\log^3 n$. And in the penultimate step we use $\Delta^{c^k} \le 2$ for n sufficiently large.)

If $t_k > \log^2 n$ then by Lemma 2.13, $s_{k+1} \le 3s_{k-1}t_k/t_{k-1}$ wvhp. If $s'_k \le (n^{2/3}\log^3 n)/4$, then since $s_k \le 4s'_k$ wvhp, by Part 3 of Lemma 2.15, $t_k \le 12\log^9 n$ wvhp. Hence, as in the case $t_k \le \log^2 n$ above, we can establish that t_{k+1} is zero whp. If $s'_k \ge (n^{2/3}\log^3 n)/4$, then by Lemma 2.15, $t_k \le 768(s'_{k-1})^3/n^2$ wvhp. Therefore, by Lemma A.2, $t_k \le 12 \cdot 768 t'_k$. Substituting this bound on t_k and appropriate bounds on s_{k-1} and t_{k-1} given by Lemma 2.23, we have wvhp,

$$\begin{aligned} s_{k+1} &\le \frac{36 \cdot 768\Delta^{c^{k-1}} s'_{k-1} t'_k}{\delta^{4c^{k-1}+1} t'_{k-1}} \\ &\le 36 \cdot 768\Delta^{9c^{k-1}+2} s'_{k+1} \\ &\le 36 \cdot 768 \cdot 2^c s'_{k+1}. \end{aligned}$$

By Lemma 2.26 with $\ell = 2$, $w'_{k+1} \geq 2w'_k$. Thus $w'_{k+1} \geq 2\log_r(\frac{n^{1/5}}{4\log^3 n})$, and $s'_{k+1} \leq 16n^{3/5}\log^6 n$. Hence, $s_{k+1} \leq n^{5/8}$ for n sufficiently large. ■

In Lemma 2.29 we place a bound on the probability that a particular ball remains after $O(\log\log n)$ rounds.

Lemma 2.29 For any ball $x \in [n]$, the probability that x remains after $O(\log\log n)$ rounds of **Alg2**$(n, 3, 1)$ is at most $4/n^{9/8}$ for n sufficiently large.

Proof: By Lemma 2.28, after $j = O(\log\log n)$ rounds, $s_j \leq n^{5/8}$ wvhp. By Part 4 of Lemma 2.15, the probability that x remains after round j is at most $4n^{15/8}/n^3$ for n sufficiently large. Since $4n^{15/8}/n^3 = 4/n^{9/8}$, the desired claim follows. ■

The following theorem follows easily from Lemma 2.29.

Theorem 7 **Alg2**$(n, 2, 2)$ terminates in $O(\log\log n)$ rounds whp. ■

Theorem 8 **Alg2**$(n, 2, 2)$ terminates in $\Omega(\log\log n)$ rounds wvhp.

Proof: For $n \geq 3$, since $r \geq e/(e-2)$ we have $2 + \log_r 12 \leq 10$. Therefore, by Lemma 2.27, $w'_i \leq 10^{i-1}$ for all $i > 0$. After $k = \lfloor \log_{10}((\log_r n)/6) \rfloor$, rounds $w'_k \leq (\log_r n)/6$, and $s'_k \geq n^{5/6}$. For n sufficiently large $n^{5/6} \geq 4n^{4/5}\log^3 n$. Therefore, by Lemma 2.23, $t_k \geq \delta^{4c^k+1} t'_k \geq \delta^{4c^k+1}(s'_k)^3/12n^2 > 1$ wvhp for n sufficiently large. (Here c is the positive root of $c^2 = 4c + 13$. Note that $\delta^{4c^k+1} > 12\sqrt{n}$ for n sufficiently large.) This shows that **Alg2**$(n, 2, 2)$ executes at least $\lfloor \log_{10}((\log_r n)/6) \rfloor \geq \lfloor \log_{10}((\log_9 n)/6) \rfloor = \Omega(\log\log n)$ rounds wvhp before termination. ■

2.4 Limited Independence

In this section we analyze the 1 out of ℓ protocol when the ℓ hash functions are chosen from a k-wise independent family of hash functions. As before, we view each hash function as a throw of n balls into n bins. We show that on any c-collision crossbar, the probability that a particular ball remains after r rounds of the k-wise independent 1 out of ℓ protocol is close to that of the fully independent protocol for $r = O(\log\log n)$, even when $k \ll n$. Importing the results in Lemma 2.22 and 2.29 in Section 2.3, we obtain the following main theorems.

Theorem 9 For integers $\ell \geq 3$ and $c \geq 1$, the 1 out of ℓ problem is solved on a c-collision crossbar in $O(\log\log n)$ rounds whp, when the ℓ hash functions are chosen independently and uniformly at random from a k-wise independent family of hash functions for $k = \Omega(\log^\alpha n)$, where α is a real constant. ■

Theorem 10 For integers $\ell \geq 2$ and $c \geq 2$, the 1 out of ℓ problem is solved on a c-collision crossbar in $O(\log\log n)$ rounds whp, when the ℓ hash functions are chosen independently and uniformly at random from a k-wise independent family of hash functions for $k = \Omega(\log^\alpha n)$ where α is a real constant. ■

Let $\mathcal{F}^k_{m,n}$ denote a k-wise independent family of functions from $[m]$ to $[n]$, that is, for $\{x_i : 0 \leq i < j\} \subseteq [m]$, $y_0, \ldots, y_{\ell-1} \in [n]^j$, $0 \leq j \leq k$, it holds that if h is drawn uniformly at random from $\mathcal{F}^k_{m,n}$, then

$$\Pr[h(x_i) = y_i \text{ for } 0 \leq i < j] = 1/n^j.$$

If $k \leq \sqrt{n}$, $\mathcal{F}^k_{m,n}$ can be constructed as in [17] using the families $\overline{H}_{n^d,n}$ and H^1_{m,n^d} defined in [6] and [22] respectively. (Here d is an appropriate constant.) A hash function h chosen uniformly at random from

$\mathcal{F}^k_{m,n}$ is defined as $r \circ s$, where r and s are chosen uniformly at random from $\overline{H}_{n^d,n}$ and H^1_{m,n^d} respectively. Both r and s can be evaluated in constant time [22, 6], and hence so can be h.

In order to analyze the 1 out of ℓ protocol, we restrict our attention to the at most n memory requests of the processors. The hash functions with the domain restricted to this set of requests can be viewed as mapping $m \leq n$ memory locations into n memory modules k-wise independently. First, we establish a few simple properties of k-wise independent hash functions.

Lemma 2.30 Let k, m, and n be integers such that $0 < k \leq m \leq n$. Let h be drawn uniformly at random from $\mathcal{F}^k_{m,n}$. For any $A \subseteq [n]$, $|A| \leq (k-1)/e^2$, we have

$$\Pr[h^{-1}(A) = \emptyset] \leq (1 - |A|/n)^m (1 + e^{-(k-1)/3}).$$

Proof: If k is even, let $k' = k$; otherwise let $k' = k - 1$. By inclusion-exclusion we have

$$\begin{aligned}
\Pr[h^{-1}(A) = \emptyset] &= 1 + \sum_{i=1}^{m} \sum_{0 \leq x_0 < \ldots < x_{i-1} < m} (-1)^i \Pr[h(x_0), \ldots, h(x_{i-1}) \in A] \\
&\leq 1 + \sum_{i=1}^{k'} \sum_{0 \leq x_0 < \ldots < x_{i-1} < m} (-1)^i \Pr[h(x_0), \ldots, h(x_{i-1}) \in A] \\
&= 1 + \sum_{i=1}^{k'} \sum_{0 \leq x_0 < \ldots < x_{i-1} < m} (-1)^i (|A|/n)^i \\
&= 1 + \left(\sum_{i=1}^{k'-1} \sum_{0 \leq x_0 < \ldots < x_{i-1} < m} (-1)^i (|A|/n)^i \right) + \binom{m}{k'} (|A|/n)^{k'} \\
&\leq 1 + \left(\sum_{i=1}^{m} \sum_{0 \leq x_0 < \ldots < x_{i-1} < m} (-1)^i (|A|/n)^i \right) + \binom{m}{k'} (|A|/n)^{k'} \\
&\leq (1 - |A|/n)^m + \binom{m}{k'} (|A|/n)^{k'} \\
&\leq (1 - |A|/n)^m (1 + (em/k')^{k'} (|A|/n)^{k'} e^{2m|A|/n}) \\
&\leq (1 - |A|/n)^m (1 + (e|A|/k')^{k'} e^{2m|A|/n}) \\
&\leq (1 - |A|/n)^m (1 + e^{-k'} e^{2|A|}) \\
&\leq (1 - |A|/n)^m (1 + e^{-k'/3}).
\end{aligned}$$

(In the seventh step we use the inequalities $1 - x \geq e^{-2x}$ for $0 \leq x \leq 1/2$ and $|A| \leq k'/e^2 \leq n/2$. The last step follows since $|A| \leq k'/e^2$.) ■

Lemma 2.31 Let k, m, and n be integers such that $0 < k \leq m \leq n$. Let h be drawn uniformly at random from $\mathcal{F}^k_{m,n}$. Let $B \subseteq [n]$ satisfy $|B| \leq k/\beta$, where real $\beta > 0$. If $S = h^{-1}(B)$, then $\Pr[|S| \geq \beta|B|] \leq (e/c)^{c|B|}$.

Proof: By the definition of S, $\Pr[|S| \geq \beta|B|]$ is the probability that there exists a set $T \subseteq [m]$, $|T| = \beta|B|$, such that $h(T) \subseteq B$. Since $\beta|B| \leq k$ and h is chosen uniformly from a k-wise independent family of hash functions, this is at most $\binom{m}{\beta|B|}(|B|/n)^{\beta|B|} \leq (e/\beta)^{\beta|B|}$. ■

Corollary 2.31.1 Let k, m, n, and p be integers such that $0 \leq p < k \leq m \leq n$. Let h be drawn uniformly at random from $\mathcal{F}^k_{m,n}$. For $0 \leq i < p$, let $X = \{x_i : 0 \leq i < p\} \subseteq [m]$ and $y = (y_0, \ldots, y_{p-1}) \in [n]^p$. Let E be the event that for $0 \leq i < p$, $h(x_i) = y_i$. Let $A \subseteq [n]$, $|A| \leq \min\{p, (k-p-1)/e^2\}$, and E' be the event that for all $x \notin X$, $h(x) \notin A$. Let $B \subseteq [n]$ satisfy $|B| \leq (k-p-1)/\beta$, where real $\beta \geq 0$. If $S = h^{-1}(B)$, then $\Pr[|S| \geq \beta|B| + p \mid E \cap E'] \leq 2e^{2|A|}(e/\beta)^{\beta|B|}$.

Proof: Let Y denote $[m] \setminus X$. Thus, $g = h^Y \mid E$ is drawn uniformly from a $(k-p)$-wise independent family of functions from Y to $[n]$. The event $E' \mid E$ is equivalent to the event that $g^{-1}(A) = \emptyset$. If $k - p$ is odd, let $k' = k - p$; otherwise let $k' = k - p - 1$. By inclusion-exclusion we have

$$\begin{aligned}
\Pr[E' \mid E] &\geq (1 - |A|/n)^{m-p} - \binom{m-p}{k'} (|A|/n)^{k'} \\
&\geq e^{-2|A|(m-p)/n} - (e|A|/(k'))^{k'} \\
&\geq e^{-2|A|} - (2e)^{-2|A|} \\
&\geq e^{-2|A|}/2.
\end{aligned}$$

(In the second step we use the inequality $(1 - |A|/n) \geq e^{-2|A|/n}$ since $|A| \leq n/2$. The third step follows from the inequality $k' \geq 2e^2|A| \geq 2|A|$.)

$$\begin{aligned}
\Pr[(|S| \geq \beta|B| + p) \mid E \cap E'] &= \frac{\Pr[((|S| \geq \beta|B| + p) \cap E') \mid E]}{\Pr[E' \mid E]} \\
&\leq \frac{\Pr[(|S| \geq \beta|B| + p) \mid E]}{\Pr[E' \mid E]} \\
&\leq \frac{\Pr[g^{-1}(B) \geq \beta|B|]}{\Pr[E']} \\
&\leq 2e^{2|A|}(e/\beta)^{\beta|B|}.
\end{aligned}$$

(In the last step we invoke Lemma 2.31 substituting $(m-p, n, k-p, \beta, B, S)$ for (m, n, k, α, B, S).) ■

For the rest of the section, we fix integers $\ell, c \geq 1$, and analyze the 1 out of ℓ protocol on the c-collision crossbar. Let $\vec{h} = (h_0, \ldots, h_{\ell-1})$ represent a tuple of ℓ hash functions, where $h_i : [m] \to [n]$, $0 \leq i < \ell$. For $x \in [m]$, let $AFFECT_i(\vec{h}, x)$ denote the set of balls that could affect the success of ball x in round j for all $0 \leq j \leq i$. Formally, we define

$$AFFECT_i(\vec{h}, x) = \begin{cases} \{x\} & \text{if } i = -1, \\ \{z \in [m] : h_{i \bmod \ell}(z) = h_{i \bmod \ell}(y) \text{ for some } y \in AFFECT_{i-1}(\vec{h}, x)\} & \text{otherwise.} \end{cases}$$

Lemma 2.32 Let k, m, and n be integers such that $0 \leq k \leq m \leq n$. Let $\vec{h} = (h_0, \ldots, h_{\ell-1})$, denote ℓ hash functions chosen independently and uniformly at random from $\mathcal{F}^k_{m,n}$, For any $x \in [m]$ and $i \geq 0$, if $k \geq \max\{4\log^2 n, 10|AFFECT_{i-1}(\vec{h}, x)|\}$, then $|AFFECT_i(\vec{h}, x)| \leq \max\{4\log^2 n, 10|AFFECT_{i-1}(\vec{h}, x)|\}$ wvhp.

Proof: In the following we use A_i as a shorthand for $AFFECT_i(\vec{h}, x)$. Let $j = i \bmod \ell$. Let $A_{i-1} = \{x_0, \ldots, x_{p-1}\}$, where $0 \leq p \leq m$. If $i \geq \ell - 1$, let $A = A_{i-1} \setminus A_{i-\ell}$; otherwise let $A = A_{i-1}$. Let $B = h_j(A_{i-\ell})$, $C = h_j(A)$, and $S = A_j \setminus A_{j-1}$. Thus $S \subseteq h_j^{-1}(C)$. Fix $y = (y_0, \ldots, y_{p-1}) \in [n]^p$ and let E be the event that for all $0 \leq q < a$, $(h_j(x_0), \ldots, h_j(x_{p-1})) = y$. Let E' be the event that for all $x \notin A_{i-1}$ $h_j(x) \notin C$. Set $\beta = \max\{(e^2 p)/|C|, (\log^2 n)/|C|\}$. We now apply Corollary 2.31.1 substituting $(k, m, n, h_j, p, X, y, B, C, S, E, E', \beta)$ for $(k, m, n, h, p, X, y, A, B, S, E, E', \beta)$ to obtain $|S| \leq \beta|C| + a$ with probability at least $1 - 2e^{2|B|}(e/\beta)^{\beta|C|}$. Since $\beta \geq e^2 p/|C| \geq e^2$, we have

$$\begin{aligned}
2e^{2|B|}(e/\beta)^{\beta|C|} &\leq 2e^{2|B| - \beta|C|} \\
&\leq 2e^{-\beta|C|/2} \\
&\leq 2e^{-\log^2 n}
\end{aligned}$$

(In the second step we use the inequality $2|B| \leq 2p \leq 2\beta|C|/e^2 \leq \beta|C|/2$.) Thus,

$$|A_j| \leq |A_{j-1}| + |S| \leq \max\{e^2 p, 2\log^2 n\} + 2p \leq \max\{4\log^2 n, 10|A_{j-1}|\}$$

wvhp. ■

For $r \geq 0$, $\vec{h} = (h_0, \ldots, h_{\ell-1})$, $h_i : [m] \rightarrow [n]$ for $0 \leq i < \ell$, and $x \in [m]$, define $\mathit{ASSIGN}_r(\vec{h}, x)$ as $\{(y, h_0(y), \ldots, h_{\ell-1}(y)) : y \in \mathit{AFFECT}_r(\vec{h}, x)\}$. We note that given $\vec{h}$, $\mathit{ASSIGN}_r(\vec{h}, x)$ completely determines whether x succeeds in r rounds under $\vec{h}$.

Let $x_i \in [m]$, $y_{i,j} \in [n]$, $0 \leq i < p$, $0 \leq j < \ell$. Let $A = \{(x_i, y_{i,0}, \ldots, y_{i,\ell-1}) : 0 \leq i < p\}$. We call A an *assignment*. Let X denote $\{x_i : 0 \leq i < p\}$. For any $r \geq 0$, we call A a *valid r-assignment* for $x \in [m]$ if $\mathit{ASSIGN}_r(\vec{g}, x) = A$, where $\vec{g} = (g_0, \ldots, g_{\ell-1})$, $g_j : X \rightarrow [n]$ defined by $g_j(x_i) = y_{i,j}$, $0 \leq i < p$, for all $0 \leq j < \ell$.

In the following let $\Pr_k[\mathit{EVENT}(\vec{h})]$ denote the probability of $\mathit{EVENT}(\vec{h})$ when each hash function in $\vec{h}$ is chosen independently and uniformly from $\mathcal{F}^k_{m,n}$.

Lemma 2.33 Let k, m, n and p be integers such that $0 \leq k \leq m \leq n$ and $0 \leq p \leq (k-1)/(e^2+1)$. Let $x_i \in [m]$, $0 \leq i < p$ be distinct integers and $y_{i,j} \in [n]$, $0 \leq i < p$, $0 \leq j < \ell$. Let $A = \{(x_i, y_{i,0}, \ldots, y_{i,\ell-1}) : 0 \leq i < p\}$. For arbitrary $x \in [m]$ and integer $r \geq 0$, we have

$$\Pr_k[\mathit{ASSIGN}_r(\vec{h}, x) = A] \leq \Pr_m[\mathit{ASSIGN}_r(\vec{h}, x) = A](1 + e^{-(k-p)/3})^\ell.$$

Proof: Let E be the event that $h_j(x_i) = y_{i,j}$ for $0 \leq i < p$, $0 \leq j < \ell$. Let $X = \{x_i : 0 \leq i < p\}$ and $Y_j = \{y_{i,j} : 0 \leq i < p\}$, $0 \leq j < \ell$. (Note that $|Y_j| \leq p$ for $0 \leq j < \ell$.) Let Z denote $[m] \setminus X$ and E' be the event that $\mathit{AFFECT}_r(\vec{h}, x) = X$. Thus $\mathit{ASSIGN}_r(\vec{h}, x) = A$ if and only if E and E' occur. Since $p \leq k$, we have $\Pr_k[E] = \Pr_m[E]$. The event E' occurs if A is an affecting r-assignment and for some $B_j \subseteq Y_j$, $0 \leq j < \ell$, determined by A, $h_j^{-1}(B_j) \cap Z = \emptyset$. (The B_j's are determined from A as follows. Let $\vec{g} = (g_0, \ldots, g_{\ell-1})$, where $g_j : X \rightarrow [n]$ is defined by $g_j(x_i) = y_{i,j}$, $0 \leq i < p$ for all $0 \leq j < \ell$. For $0 \leq j < \ell$, if $r < j$ then $B_j = \emptyset$. Otherwise, we consider two cases. If $r \bmod \ell \leq j$, then $B_j = g_j(\mathit{AFFECT}_{(\lfloor r/\ell \rfloor - 1)\ell + j}(\vec{g}, x))$; otherwise $B_j = g_j(\mathit{AFFECT}_{\lfloor r/\ell \rfloor \ell + j}(\vec{g}, x))$.)

If A is not a valid r-assignment the claim holds trivially. So we assume that A is a valid r-assignment. For any $p \leq q \leq m$, if h_j is drawn from a q-wise independent family of hash functions from $[m]$ to $[n]$, then $g_j = h_j^Z \mid E$ can be viewed as having been drawn from a $(q-p)$-wise independent family of hash functions from Z to $[n]$. We invoke Lemma 2.30 substituting $(k-p, m-p, n, g_j, B_j)$ for (k, m, n, h, A) to obtain

$$\Pr_k[g_j^{-1}(B_j) = \emptyset] \leq \Pr_m[g_j^{-1}(B_j) = \emptyset](1 + e^{-(k-p)/3})$$

for $0 \leq j < \ell$. Therefore,

$$\begin{aligned}
\Pr_k[E \cap E'] &= \Pr_k[E]\Pr_k[E' \mid E] \\
&\leq \Pr_m[E]\Pr_m[E' \mid E](1 + e^{-(k-p)/3})^\ell \\
&= \Pr_m[E \cap E'](1 + e^{-(k-p)/3})^\ell.
\end{aligned}$$

■

Lemma 2.34 Let k, m and n be integers such that $0 \leq k < m \leq n$. For any real $\gamma \geq 0$, $x \in [m]$ and integer $r \leq \gamma \log\log n$, if $k \geq 8\log^{4\gamma+2} n$, then

$$\Pr_k[x \text{ remains after } r \text{ rounds under } \vec{h}] \leq \Pr_m[x \text{ remains after } r \text{ rounds under } \vec{h}] + 1/n^2$$

for n sufficiently large.

Proof: Let $\mathcal{A}$ be the set of valid r-assignments for x under which x fails. By Lemma 2.32, $|ASSIGN_r(\vec{h}, x)| \leq 4(\log^2 n)10^r\} \leq 4\log^{4\gamma+2} n$ wvhp. By Lemma 2.33, for any assignment A such that $|A| \leq 4\log^{4\gamma+2} n$,

$$\begin{aligned}
\Pr_k[ASSIGN_r(\vec{h}, x) = A] &\leq \Pr_m[ASSIGN_r(\vec{h}, x) = A](1 + e^{-(k-|A|)/3})^{\ell} \\
&\leq \Pr_m[ASSIGN_r(\vec{h}, x) = A](1 + 1/n^3),
\end{aligned}$$

for n sufficiently large. (Here we use the fact that $k \geq 8\log^{4\gamma+2} n \geq 2|A|$.) Thus,

$$\begin{aligned}
\Pr_k[x \text{ fails in } r \text{ rounds under } \vec{h}] &\leq \Pr_k[ASSIGN_r(\vec{h}, x) \in \mathcal{A}] \\
&\leq \Pr_k[(ASSIGN_r(\vec{h}, x) \in \mathcal{A}) \text{ and } |ASSIGN_r(\vec{h}, x)| \leq 4\log^{4\gamma+2} n] + 1/n^3 \\
&\leq \sum_{\substack{A \in \mathcal{A} \\ |A| \leq 4\log^{4\gamma+2} n}} \Pr_k[ASSIGN_r(\vec{h}, x) = A] + 1/n^3 \\
&\leq \sum_{\substack{A \in \mathcal{A} \\ |A| \leq 4\log^{4\gamma+2} n}} \Pr_m[ASSIGN_r(\vec{h}, x) = A](1 + 1/n^3) + 1/n^3 \\
&\leq \Pr_m[ASSIGN_r(\vec{h}, x) \in \mathcal{A}](1 + 1/n^3) + 1/n^3 \\
&\leq \Pr_m[x \text{ fails in } r \text{ rounds under } \vec{h}] + 1/n^2
\end{aligned}$$

for n sufficiently large. ■

By Lemma 2.22, for any $x \in [n]$, $\Pr_n[x$ remains after $O(\log\log n)$ rounds$]$ is at most $2/n^{6/5}$ in the 1 out of 3 protocol on the 1-collision crossbar. Similarly, for the 1 out of 2 protocol on the 2-collision crossbar, Lemma 2.29 implies that $\Pr_n[x$ remains after $O(\log\log n)$ rounds$]$ is at most $4/n^{9/8}$ for any $x \in [n]$. We now apply Lemma 2.34 to establish Theorems 9 and 10.

2.5 Generalizations

Alg1 can be generalized to apply to any a out of b problem by changing the *RandomSubbag* and *PrunedBag* routines appropriately; after each step, we need to keep track of how many successes each processor has had, and only those processors with fewer than a successes participate. In the following discussion, we refer to this protocol as the *generic* protocol. For given a and b, the analysis of the generic protocol can be done using the approach of Subsection 2.3, but involves more complicated calculations and recurrences. In [9], the authors use a different analysis of this protocol for the 2 out of 3 case and show an $O(\log\log n)$ upper bound when the collision factor is greater than 3. In this section, we present a simple variant of the generic protocol that solves any a out of b problem on a 2-collision crossbar in $O(\log\log n)$ time whp.

In particular we can solve any a out of $a+1$ problem by running $\mathbf{Alg1}(n, 2, 1)$ with $\binom{a+1}{2}$ different hash-function pairs. Since each run fails with a polynomially small probability, and there are only a constant number of runs, the entire algorithm succeeds whp. For instance, in the case of 2 out of 3, we simply perform 3 runs of **Alg1**. At first glance, it may appear that this revised protocol is only of interest because it is simpler to analyze. Actually, the new protocol is competitive with the generic one for small a and is much faster for large a. Comparing it for the 2 out of 3 problem, we first note that since each of the 3 runs use 2 hash functions only, while the generic protocol uses 3, the revised protocol will be at most twice as slow as the generic one. Moreover, the 1 out of 2 problem is clearly a simpler problem than the 2 out of 3 problem. So each run will involve a fewer rounds than in the generic algorithm. For large a, this phenomenon is exaggerated. For a generic a out of $a+1$ protocol to make "progress", a number of processors must have a large number of successes. But at the outset, the fraction of processors that have succeeded on $d \leq a$ hash functions decreases exponentially with d. Therefore, while the revised protocol experiences only a quadratic slowdown in running time, the generic procotol will suffer an exponential increase in running time with increasing a.

The basic idea outlined above can be used to solve any a out of b problem by choosing any $a+1$ hash functions and solving the corresponding a out of $a+1$ problem.

Theorem 11 For integer constants a and b with $1 \leq a < b$, the corresponding a out of b problem can be solved on a 2-collision crossbar in $O(\log \log n)$ time whp. ■

The above generalizations can be made for the 1-collision crossbar as well. Since **Alg1** solves the 1 out of 3 problem on a 1-collision crossbar in $O(\log \log n)$ time whp, any a out of $a+2$ problem can be solved in the same asymptotic time bound by running **Alg1**$(n, 3, 1)$ on $\binom{a+2}{3}$ different triples of hash functions.

Theorem 12 For integer constants a and b with $1 \leq a < b-1$, the corresponding a out of b problem can be solved on a 1-collision crossbar in $O(\log \log n)$ time whp. ■

It is worth noting that by the above result, a 1-collision crossbar can solve the 3 out of 5 problem and hence can simulate an EREW PRAM with n processors in $O(\log \log n)$ time whp using 5 hash functions. Thus a 1-collision crossbar is asymptotically as powerful as a 2-collision one.

3 Symmetry Breaking

In this section we analyze algorithms for the Control Tower problem. In the Control Tower problem, there is one central control tower and n remote stations, h of which contain a message destined for the control tower. Each station can only transmit a message to and receive a message from the control tower, and only in discrete time slots. When two or more stations attempt to transmit a message in the same time slot, all of the transmitted messages are lost. Note that the control tower can transmit a message to only one remote station in one time slot. We will consider each time slot to be a *step*; the goal is to transmit all h messages to the control tower in as few steps as possible. (Throughout our analysis, we will assume that the remote stations are numbered from 1 to n, and that the control tower sends an acknowledgement upon receipt of a message.)

The Control Tower problem is related to the problem of direct routing of h-relations on a 1-collision crossbar, as discussed in Section 1. Specifically, since each processor is only permitted to transmit a message directly to its destination, we are able to analyze each destination independently. Transmitting a set of at most h messages to a destination is then equivalent to the Control Tower problem. Thus the lower bound we obtain for the Control Tower problem can be used to obtain a lower bound for direct h-relation routing.

To give some intuition into the Control Tower problem, let us first examine the case when $h = 2$. In this case, the problem is that two stations are trying to transmit messages, but if they transmit at the same time, then both of the transmissions are blocked. Note that some sort of symmetry breaking is required, or the messages may never be successfully transmitted. A simple randomized strategy to break the symmetry would be for each station to flip a coin and transmit if and only if it comes up "heads". Then the expected number of steps before both messages are successfully transmitted is constant. To achieve success wvhp requires $\Theta(\log n)$ steps, however. Still, this turns out to be the best strategy possible, as shown by Goldberg, Jerrum, Leighton, and Rao [12]. Generalizing to $h > 2$, we can easily verify that if there are k messages left to transmit ($k \leq h$), then by having each station transmit with probability $1/k$, we maximize the probability of obtaining a successful transmission. The difficulty is that the stations are not able to communicate with each other, and consequently they do not know exactly how many messages are left to transmit at any intermediate step of the algorithm. If the number of messages left to transmit were known to all stations at every step, then the Control Tower problem could be solved in $\Theta(h + \log n)$ steps wvhp. If not, Geréb-Graus and Tsantilas [10] showed that the Control Tower problem could be solved in $O(h + \log h \log n)$ time wvhp, but it was not known whether the extra $\log h$ factor could be eliminated. We show that the extra factor of $\log h$ is indeed necessary.

We also examine deterministic solutions for the Control Tower problem and direct h-relation routing. As mentioned in Section 1.2, a deterministic lower bound of $\Omega((h/\log h)\log n)$ for the Control Tower problem follows from the same lower bound for routing all h messages in the Ethernet model [15]. We show a lower bound of $\Omega((h/\min\{\log h, \log\log n\})\log n)$ (which improves the previous lower bound for h larger than *polylog*(n)), and we show this lower bound holds for successfully transmitting *any* of the h messages in the Control Tower problem. (This result does not hold in the Ethernet model, where it is trivial to successfully transmit one of the messages in $\Theta(\log n)$ time.) Finally, we prove the existence of a $\Theta(h\log h\log n)$ time deterministic algorithm for direct h-relation routing.

Lower bounds for the Control Tower problem can most easily be studied in the context of hypergraphs, so here we review the definition of a hypergraph and related concepts.

Let V be a set of elements, called *vertices*. Let E be a set of nonempty subsets of V, called *edges*. Then a *hypergraph* is given by a pair (V,E). Given a hypergraph $H=(V,E)$, the hypergraph *induced* by a set of vertices $V'\subseteq V$ is $H'=(V',E')$, where $E'=\{e\in E : e\cap V'\neq\emptyset\}$. Also the hypergraph *induced* by a set of edges $E'\subseteq E$ is $H'=(V,E')$. A subset of vertices $T\subseteq V$ *covers* an edge $e\in E$ if $e\cap T\neq\emptyset$. A *transversal* of H is a set of vertices $T\subseteq V$ that covers every edge in E. For convenience, we define an *a-transversal* to be a set of vertices $T\subseteq V$ that covers at least a edges of E.

We define a hypergraph $H=(V,E)$ to be *a-thick* if $\min_{e\in E}|e|\geq a$, Note that if E is empty, then H is a-thick for every a. A hypergraph $H=(V,E)$ is *(a,b)-thick* if H is a-thick and $\sum_{k\geq 0} f_k 2^{-k}\leq b$, where f_k denotes the number of edges $e\in E$ such that $a2^k\leq|e|\leq a2^{k+1}$, $k\geq 0$.

We will make use of the inequalities: (i) $1+x\leq e^x$ for all real x, and (ii) $1-x\geq e^{-2x}$ for $0\leq x\leq 1/2$.

3.1 Deterministic Algorithm for Direct Routing

Here we give an improved upper bound for deterministic direct routing on the 1-collision crossbar. This improved bound is obtained by slightly modifying a technique in Goldberg and Jerrum [11]. They show that one can solve the Control Tower problem deterministically in $\Theta(h\log n)$ steps. (Their proof is much simpler than the proof in Komlós and Greenberg [18].) This directly implies the existence of a deterministic direct h-relation routing algorithm for the 1-collision crossbar which takes $\Theta(h^2\log n)$ steps. By modifying their technique, however, we are able to obtain an upper bound of $\Theta(h\log h\log n)$ on deterministic direct h-relation routing on the 1-collision crossbar. Thus, in combination with the lower bounds shown later, we obtain an exponential decrease in the gap between the upper and lower bounds for this problem.

As in Goldberg and Jerrum [11], define an (h,k)-relation as a routing problem in which each processor is the source of at most h packets and each memory module is the destination of at most k packets.

Lemma 3.1 There exists a deterministic algorithm that, given an arbitrary (h,k)-relation ($k\leq h\leq n$), terminates after $O(h\log n)$ communication steps such that the remaining communication problem is a $(h,\lceil k/2\rceil)$-relation.

Proof: We employ the probabilistic method to prove this. Assume each processor has h slots numbered $1,\ldots,h$ and each slot contains at most one message. Consider the following randomized communication step: Each processor selects one of its slots at random, and if the slot has a message, attempts to transmit it.

At any step, if there are ℓ messages, $k/2\leq\ell\leq k$, yet to be received by a memory module M, the probability of a successful transmission to M is at least:

$$\ell\frac{1}{h}\left(1-\frac{1}{h}\right)^{\ell-1}\geq\ell\frac{1}{h}\left(1-\frac{1}{h}\right)^{h-1}\geq\frac{k}{2eh}.$$

The probability that in $ch\log n$ communication steps there exist at least $ch\log n - \lfloor k/2 \rfloor$ failed transmissions to M is at most

$$\binom{ch\log n}{k/2}\left(1-\frac{k}{2eh}\right)^{ch\log n - k/2} \leq (ch\log n)^{k/2}e^{-kc\log n/4e} \leq \frac{1}{n^{3k}}$$

for n sufficiently large. (The last inequality is obtained by choosing an appropriate c.) Since the number of choices for the sources and the slots of the messages destined to M are at most

$$\binom{nh}{k} \leq (nh)^k \leq n^{2k},$$

the expected number of message assignments such that this algorithm fails to send $\lfloor k/2 \rfloor$ messages is bounded above by $1/n^k$. Thus there exists a deterministic algorithm that transmits at least $\lfloor k/2 \rfloor$ messages destined to M. Now we note that any assignment of messages destined to another module P is also a possible assignment for messages to M. Therefore the same deterministic algorithm that routes at least $\lfloor k/2 \rfloor$ messages to M will also route at least $\lfloor k/2 \rfloor$ messages to P, for any other P. ■

Theorem 13 There exists a deterministic algorithm that realizes any h-relation , $h \leq n$, in $O((h\log h)\log n)$ steps.

Proof: There are $\lceil \log h \rceil$ phases in the algorithm. The ith phase reduces an $(h, \lceil h/2^{i-1} \rceil)$-relation problem to an $(h, \lceil h/2^i \rceil)$-relation using the algorithm in Lemma 3.1. Thus after $\lceil \log h \rceil$ phases, the remaining communication problem is an $(h,1)$-relation that can be realized in $O(h)$ steps. ■

3.2 Deterministic Lower Bound

To obtain a lower bound on the number of steps required for a deterministic solution to the Control Tower problem, we will first find a subset of stations such that a good fraction of those stations transmit at each step, and then show that there are two small disjoint groups of stations which always transmit during exactly the same steps. By placing messages at all stations in both groups, no station in either group would succeed in transmitting its message, due to contention.

Some preliminary results which will aid in our proof are presented here.

Given that the number of edges in a hypergraph is small (compared to the number of vertices), the following lemma shows that we can find a large subset of vertices which all cover exactly the same set of edges. By using this relatively simple lemma, we would be able to prove a logarithmic lower bound on the Control Tower problem. It will be much more difficult, however, to prove a superlogarithmic lower bound.

Lemma 3.2 Any hypergraph $H = (V, E)$ with n vertices and m edges has a subset V' of V with $|V'| \geq n2^{-m}$ in which every vertex in V' covers exactly the same set of edges.

Proof: By induction on m. If $m = 0$ then $E = \emptyset$, so let $V' = V$. Then $|V'| = n = n2^0$, and every vertex in V' covers no edges. If $m > 0$, choose an edge $e \in E$, and consider the hypergraph $H_0 = (V, E - \{e\})$ induced by $E - \{e\}$. By induction we can find a subset V'' of V with $|V''| \geq n2^{-(m-1)}$ in which every vertex in V'' covers exactly the same set of edges. Then let V' be the larger of $V'' \cap e$ and $V'' - e$, one of which is guaranteed to be at least of size $|V''|/2 \geq n2^{-m}$. Also, all vertices in V' cover exactly the same set of edges in $E - \{e\}$, and either all vertices in V' cover e or all do not cover e. ■

The following lemma shows that we can find a subset of vertices such that every edge induced by that subset contains a large fraction of those vertices.

Lemma 3.3 Let $x > 0$. Then any hypergraph $H = (V, E)$ with n vertices and m edges has a subset V' of V with $|V'| \geq n(1 - 1/x)^m$ that induces a $|V'|/x$-thick hypergraph $H' = (V', E')$.

Proof: Construct V' from V iteratively by removing vertices for m steps. Let V_i be the set of vertices remaining after step i, with $V_0 = V$, and $V' = V_m$. At step i, if V_{i-1} induces a hypergraph which is $|V_{i-1}|/x$-thick, let $V_i = V_{i-1}$. Otherwise, let $e \in E$ be an edge with $|e \cap V_{i-1}| < |V_{i-1}|/x$, and let $V_i = V_{i-1} - e$. By induction, we have $|V_i| > n(1 - 1/x)^i$, and thus $|V'| = |V_m| > n(1 - 1/x)^m$. If at some step i, the hypergraph induced by V_i is $|V_i|/x$-thick, then by construction, $V' = V_i$, and V' induces a hypergraph which is $|V'|/x$-thick. Otherwise, the number of edges in the hypergraph induced by V_i is at least one less than the number of edges in the hypergraph induced by V_{i-1}. Thus the hypergraph induced by $V' = V_m$ contains no edges, and consequently, it is $|V'|/x$-thick. ■

Corollary 3.3.1 Let $x \geq 2$. Then any hypergraph $H = (V, E)$ with n vertices and m edges has a subset V' of V with $|V'| \geq ne^{-2m/x}$ that induces a $|V'|/x$-thick hypergraph $H' = (V', E')$. ■

Next we formulate some results about transversals and "near-transversals" (small subsets of vertices which cover almost all of the edges).

The following lemma is given by (Alon [2], Proposition 2.1, with α set to $\frac{\ln(m/x)}{\ln k}$)

Lemma 3.4 (Alon [2]) Let $x \geq 1$. Then any n/x-thick hypergraph $H = (V, E)$ with n vertices and m edges has a transversal of size at most $x + x\ln(m/x)$. ■

We will need to use the following simple corollary.

Corollary 3.4.1 Let $x \geq 1$. Then any $2n/x$-thick hypergraph $H = (V, E)$ with n vertices and m edges where $2n/x > 2(x + x\ln(m/x))$ has two disjoint transversals of size at most $x + x\ln(m/x)$.

Proof: Using the fact that H is n/x thick, we can construct one transversal of size at most $x + x\ln(m/x)$ using Lemma 3.4. Note that after removing that transversal from the set of vertices, the remaining hypergraph is also n/x thick. Thus we can construct another (disjoint) transversal of size $x + x\ln(m/x)$. ■

Using these results, we would only be able to prove the desired lower bound on the Control Tower problem for $h = \Omega(\log n)$. In order to prove our bound for small h, we must use a much more involved argument using near-transversals. We first present a lemma which is similar to the lemma by Alon above, except that it allow us to trade off the size of a near-transversal with the number of edges covered by the near-transversal.

Lemma 3.5 Let $x, z \geq 1$. Then any n/x-thick hypergraph $H = (V, E)$ with n vertices and m edges has an $(m - m/z)$-transversal of size at most $\lceil x \ln z \rceil$.

Proof: Iteratively choose a set of vertices to be in the $(m - m/z)$-transversal. Let T_i be the set of vertices chosen by step i, and let E_i be the set of edges which are covered by T_i. Then $T_0 = \emptyset$ and $E_0 = \emptyset$. We will proceed for at most $t = \lceil x \ln z \rceil$ steps. At step i, if $E = E_{i-1}$, then let $T_i = T_{i-1}$ and $E_i = E_{i-1}$. Otherwise, T_{i-1} does not intersect any edge in $E - E_{i-1}$ and thus each edge in $E - E_{i-1}$ contains at least n/x vertices of $V - T_{i-1}$. Then by an averaging argument, some vertex $v \in V - T_{i-1}$ must be contained in at least $|E - E_{i-1}|/x$ edges. Let $T_i = T_{i-1} \cup \{v\}$. Then $E_i = E_{i-1} \cup \{e : e \cap \{v\} \neq \emptyset\}$. By induction, we have $|E - E_i| \leq m(1 - 1/x)^i$. Then $|E - E_i| \leq me^{-i/x}$. Thus $|E - E_t| \leq me^{-\ln z} = m/z$, implying $|E_t| \geq m - m/z$. Then T_t is an $(m - m/z)$-transversal, and $|T_t| \leq t = \lceil x \ln z \rceil$. ■

For explanation purposes, call an $(m - m/z)$-transversal constructed in Lemma 3.5 a *near-transversal*. To obtain our desired result, we will need two disjoint near-transversals. It would be trivial to simply find two of these by applying the previous lemma twice, but it is necessary that these two near-transversals have the property that they cover exactly the same edges. (Notice that when using full transversals, this property is trivially satisfied.) Here we show that for certain hypergraphs there exist two small disjoint near-transversals covering almost all of the edges, and furthermore, they cover exactly the same edges.

Lemma 3.6 Let $y \geq 2$, $k \geq 1$, and $t = \lceil 2y \ln y^k \rceil$. Then any n/y-thick hypergraph $H = (V, E)$ with n vertices and m edges where

$$\left\lfloor \frac{n}{2yt} \right\rfloor > \sum_{i=0}^{\lfloor m/y^k \rfloor} \binom{m}{i},$$

has two disjoint $(m - m/y^k)$-transversals each of size at most t which cover exactly the same edges.

Proof: If $m = 0$, then the lemma holds trivially, so assume that $m \geq 1$. Iteratively construct $\lfloor n/(2yt) \rfloor$ disjoint $(m - m/y^k)$-transversals, each having size at most t, using the method from Lemma 3.5 with $x = 2y$ and $z = y^k$. After constructing each near-transversal, remove the vertices in that near-transversal from the hypergraph. This ensures that the near-transversals constructed will be disjoint. Note that after constructing i near-transversals of size t, we will have removed at most it vertices. Therefore, we will remove a total of at most $\lfloor n/(2yt) \rfloor t \leq n/2y$ vertices, which implies that the remaining hypergraph will be $n/2y$-thick. This guarantees that we can use the value $x = 2y$ in Lemma 3.5 at each step.

Now the number of ways to choose at most m/y^k edges from m edges is

$$\sum_{i=0}^{\lfloor m/y^k \rfloor} \binom{m}{i}$$

By the condition stated in the lemma, this is less than the number of near-transversals we created, and thus two of these near-transversals must cover exactly the same edges. ■

Corollary 3.6.1 Let $y \geq 2$. Then any n/y-thick hypergraph $H = (V, E)$ with n vertices and m edges, where $y \geq 4m/\ln n$ and

$$\left\lfloor \frac{n}{2y\lceil 6y \ln y \rceil} \right\rfloor > mn^{1/4},$$

has two disjoint $(m - m/y^3)$-transversals each of size at most $\lceil 6y \ln y \rceil$ which cover exactly the same edges.

Proof: If $m = 0$, the corollary holds trivially, so assume $m \geq 1$. Use Lemma A.5 to show that the conditions of Lemma 3.6 apply (with k set to 3), as follows

$$\begin{aligned} \sum_{i=0}^{\lfloor m/y^k \rfloor} \binom{m}{i} &\leq me^{m/y} \\ &\leq me^{(\ln n)/4} \\ &\leq mn^{1/4} \\ &< \left\lfloor \frac{n}{2y\lceil 6y \ln y \rceil} \right\rfloor. \end{aligned}$$

Then Lemma 3.6 states that there are two disjoint $(m - m/y^3)$-transversals, each of size at most $\lceil 6y \ln y \rceil$ which cover exactly the same edges. ■

Now we can place a bound on the number of transmission steps required to solve the Control Tower problem.

Theorem 14 A deterministic algorithm for the Control Tower problem with h messages ($2 \leq h \leq n^{1/10}$) to transmit requires at least $\Omega((h/\min\{\log h, \log\log n\})\log n)$ steps to successfully transmit any message.

Proof: Let V be the set of all n stations, and E be a set of edges, where edge i contains all stations that would attempt to transmit a signal in step i, if they had a message to transmit. Let $m = |E|$. We will show that if $m < (h/128\min\{\ln h, \ln\ln n\})\ln n$, then we can select a group of h stations such that no station succeeds in transmitting its message. To select a group such that at least two stations do not succeed, we would need to select two disjoint subgroups of stations that try to transmit at exactly the same time steps. In terms of hypergraphs, this is equivalent to selecting two disjoint sets of vertices which cover exactly the same set of edges. (Note that this would be enough to prove a lower bound for the Control Tower problem.) To select a group such that no stations succeed, we need to make sure that in all steps where a station is attempting to transmit, another station is also attempting to transmit. In terms of hypergraphs this is equivalent to selecting two disjoint sets of vertices which cover exactly the same set of edges, where the size of the union of the two disjoint sets is exactly h.

If $h < 64$, then $m \leq \frac{1}{2}\log n$. Then by Lemma 3.2, we can find a group of $\sqrt{n}$ vertices which cover exactly the same set of edges. Then we simply select h vertices from this group. (This case is also shown in Goldberg, Jerrum, Leighton, Rao [12].)

If $h \geq 64$ and $h \geq \ln n$, then $m < (h/128\ln\ln n)\ln n$, and we choose x such that $h = 4x\ln\ln n$. Note that $x \geq h/4\ln\ln n \geq 32m/\ln n$, and $x \geq 2$. By Corollary 3.3.1, there is a subset V' of V such that $|V'| \geq ne^{-4m/x} \geq n^{7/8}$ and V' induces a $2|V'|/x$-thick hypergraph $H' = (V', E')$. Let $m' = |E'|$. By Corollary 3.4.1, there are two disjoint transversals of H' whose sizes sum to at most $2(x + x\ln\ln n) \leq h$ Then we can select the vertices in these two transversals, and select the rest of the h vertices arbitrarily from V'. Then none of the h stations corresponding to the selected vertices successfully transmits its message.

If $64 \leq h \leq \ln n$, then choose x such that $h = 4 + 12x\ln x$. Note that $x \geq h/16\ln h > 8m/\ln n$ and $x \geq 2$. By Corollary 3.3.1, there is a subset V' of V such that $|V'| \geq ne^{-2m/x} \geq n^{3/4}$ and V' induces a $|V'|/x$-thick hypergraph $H' = (V', E')$. Let $m' = |E'|$. Then by Corollary 3.6.1, there are two disjoint $(m' - m'/x^3)$-covers T_1 and T_2 of H' which cover exactly the same set of edges and whose sizes sum to at most $2(6x\ln x + 1) \leq h - 2$. Then to obtain a lower bound for the Control Tower problem, we could select the vertices in T_1 and T_2 and select the other $h - |T_1 \cup T_2|$ vertices arbitrarily.

To prove the theorem, however, requires that we obtain two disjoint subsets of V which cover the same edges and whose sizes sum to h. We proceed as follows. First we note that we will be selecting the vertices from V', and thus they will not cover any edge in $E - E'$. We will start with the two disjoint subsets T_1 and T_2 found above. Let E'' be the set of edges in E' that are not covered by T_1 or T_2. Note that $|E''| \leq m'/x^3 \leq m'/x \leq \frac{\ln n}{4}$. By Lemma 3.2, there is a subset V'' of V' with $|V''| \geq |V'|2^{-(\ln n)/4} \geq n^{1/2}$ in which all vertices cover exactly the same set of edges in E''. Then add one vertex from V'' to T_1, and another $h - |T_1 \cup T_2| - 1$ from V'' to T_2. ■

Corollary 14.1 Any deterministic algorithm for direct routing of an h-relation ($2 \leq h \leq n^{1/10}$) on a 1-collision crossbar with n processors requires $\Omega((h/\min\{\log h, \log\log n\})\log n)$ steps.

Proof: Consider one memory module to be the control tower, and the other n processors to be the stations. The collision protocol is the same in the 1-collision crossbar and the Control Tower problem. By the previous theorem then, there is a way to place h messages on the n processors such that they do not all succeed in being sent to the destination memory module in less than $(h/128\min\{\ln h, \ln\ln n\})\ln n$ steps. ■

3.3 Randomized Lower Bound

A randomized algorithm can solve the Control Tower problem in expected $O(h)$ steps. However, for many applications, including the analysis of direct h-relation routing on a c-collision crossbar, it is necessary to determine the number of steps required to achieve a polynomially small probability of failure. That is what we analyze in this section.

Our lower bound will make use of a result of Yao [25], which states that any lower bound for deterministic algorithms on random inputs implies the same lower bound for randomized algorithms on worst case inputs. In particular, we will proceed by developing a lower bound for deterministic algorithms solving the Control Tower problem, where we assume that the h stations that have messages to transmit are chosen randomly from the n stations.

To begin, we convert the problem into the hypergraph domain. Then we show how to reduce the hypergraph corresponding to a given deterministic algorithm to a thick hypergraph, and show that with a significant probability, the processors corresponding to vertices in this thick hypergraph contain messages that are not successfully transmitted.

The following lemma is a modification of a result of Alon, Bar-Noy, Linial and Peleg [3, Lemma 3.1], with a similar proof. For completeness, we will present the entire proof.

Lemma 3.7 Given $r_1, r_2 \geq 1$, where $r_1 < r_2$, let $r = r_2 - r_1$ and $H = (V, E)$ be a hypergraph with n vertices and m edges. Then for some k, $r_1 \leq k < r_2$, there is a subset V' of V with $|V'| \geq ne^{-8m2^{-k}/r}$ that induces a $(|V'|2^{-k}, 4m/r)$-thick hypergraph $H' = (V', E')$.

Proof: Define a permutation $e_1, \ldots, e_m$ of the edges inductively as follows. Let e_1 be a minimum size edge in E. Then assuming edges $e_1, \ldots, e_i$ have already been chosen $(1 \leq i < m)$, let e_{i+1} be the edge in $E \setminus \{e_1, \ldots, e_i\}$ such that $\left| e_{i+1} \setminus \bigcup_{1 \leq j \leq i} e_j \right|$ is minimum. For $1 \leq i \leq m$, define $p_i = 0$ if $\left| \bigcup_{1 \leq j < i} e_j \right| = n$, else

$$p_i = \frac{\left| e_i \setminus \bigcup_{1 \leq j < i} e_j \right|}{n - \left| \bigcup_{1 \leq j < i} e_j \right|}.$$

For each $k \geq 0$, $r_1 < k \leq r_2$, let $j(k)$ be the smallest i such that $p_i \geq 2^{-k}$. (If there is no such i, let $j(k) = m + 1$). Notice that by the definition of the permutation $e_1, \ldots, e_m$, for every $k \geq 0$ and every $j' \geq j(k)$,

$$\frac{\left| e_{j'} \setminus \bigcup_{1 \leq i < j(k)} e_i \right|}{n - \left| \bigcup_{1 \leq i < j(k)} e_i \right|} \geq 2^{-k}.$$

Now for each $k \geq 0$, let

$$\begin{aligned} d_k &= |\{i : 1 \leq i \leq m \text{ and } 2^{-k} \leq p_i < 2^{-k+1}\}|, \text{ and} \\ d'_k &= \sum_{i \geq 1} \frac{d_{k+i}}{2^i}. \end{aligned}$$

Then

$$\sum_{k \geq 0} d'_k \leq \sum_{k \geq 0} d_k \leq m.$$

Call an index k *good* if $r_1 \leq k < r_2$ and $d'_k \leq 2m/r$. The average value of d'_k over $r_1 \leq k < r_2$ is at most m/r, and thus at least half of the indices k, $r_1 < k \leq r_2$, are good. Note that if k is good, then we have

$$n - \left| \bigcup_{1 \leq i < j(k)} e_i \right| = n \prod_{1 \leq i < j(k)} (1 - p_i)$$

$$\begin{aligned} &\geq n \prod_{k'>k} (1 - 2^{1-k'})^{d_{k'}} \\ &\geq ne^{-(2\sum_{k'>k} d_{k'} 2^{1-k'})} \\ &= ne^{-2^{2-k} d'_k} \\ &\geq ne^{-8m2^{-k}/r} \end{aligned}$$

Thus, for any good k there exists some subset V' of the required size that induces a $|V'|2^{-k}$-thick hypergraph.

Next we show that for at least for some good k, there exists a subset V' that induces a $(|V'|2^{-k}, 4m/r)$-thick hypergraph. For each good k and each edge $e_{j'}$ with $j(k) \leq j' \leq m$, define $s(k, j')$ to be the unique integer ℓ if

$$2^{\ell-k} \leq \frac{\left|e_{j'} \setminus \bigcup_{1 \leq i < j(k)} e_i\right|}{n - \left|\bigcup_{1 \leq i < j(k)} e_i\right|} < 2^{\ell-k+1}.$$

Note that if k',k are both good and $k' < k$, then $j(k') \geq j(k)$. Thus, if $j' \geq j(k')$ then

$$\frac{\left|e_{j'} \setminus \bigcup_{1 \leq i < j(k')} e_i\right|}{n - \left|\bigcup_{1 \leq i < j(k')} e_i\right|} \leq \frac{\left|e_{j'} \setminus \bigcup_{1 \leq i < j(k)} e_i\right|}{n - \left|\bigcup_{1 \leq i < j(k)} e_i\right|}.$$

Consequently, $s(k', j') \leq s(k, j') - 1$. Therefore, for every fixed $j' \leq m$,

$$\sum_{\substack{k \text{ good} \\ j(k) \leq j'}} 2^{-s(k,j')} \leq 2.$$

For each good k, define $y_k = \sum_{j' \geq j(k)} 2^{-s(k,j')}$. Then

$$\sum_{k \text{ good}} y_k \leq \sum_{1 \leq j' < m+1} \sum_{\substack{k \text{ good} \\ j(k) \leq j'}} 2^{-s(k,j')} \leq 2m.$$

Since at least $r/2$ indices are good, there is some index k with $y_k \leq 4m/r$. Hence $V' = V \setminus \bigcup_{1 \leq i < j(k)} e_j$ satisfies the conditions of the lemma. ■

Next we show that large random sets of vertices in a thick hypergraph cover many edges with non-trivial probability.

Lemma 3.8 Let $x \geq 1$ and $t = \lceil 13x \rceil$. Given an $(n/x, y)$-thick hypergraph $H = (V, E)$ with n vertices and m edges, and a randomly chosen subset T of V with $|T| = t$, then with probability at least $(2x)^{-t}$, for every $j \geq 0$, T will cover all but $y2^{-(j+1)}$ of the edges with thickness in the range $[n2^j/x, n2^{j+1}/x)$.

Proof: Note that $\lceil 13x \rceil \geq \lceil 1 + \lfloor \log x \rfloor + 4x \sum_{i \geq 0} (\frac{3}{2})^{-i} \rceil$. For each j, $0 \leq j \leq \log x$, let

$$E_j = \{e \in E : n2^j/x \leq |e| < n2^{j+1}/x\}.$$

Then $E = \bigcup_{j=0}^{\lfloor \log x \rfloor} E_j$. By the fact that H is $(n/x, y)$-thick, $|E_j| \leq y2^j$. View T as a collection of disjoint randomly chosen subsets $T_0, \ldots, T_{\lfloor \log x \rfloor}, T'$, where $t_j = |T_j| = \lceil 4x(1.5)^{-j} \rceil$. Note that $t \geq \sum_{j=0}^{\lfloor \log x \rfloor} t_j$. We will show that the vertices from the set T_0 cover all but $y2^{-1}$ edges in E_0 with probability at least $(2x)^{-t_0}$, the vertices from the set T_1 cover all but $y2^{-2}$ edges in E_1 that were not covered by T_1 with probability at least $(2x)^{-t_1}$, and in general the vertices from the set T_j cover all but $y2^{-(j+1)}$ of the edges in E_j that were not covered by $\bigcup_{0 \leq i < j} T_i$ with probability at least $(2x)^{-t_j}$. The lemma then follows easily from this.

Assume we have chosen the vertices in sets $T_0, \ldots, T_{j-1}$, $j \geq 0$. Let

$$E'_j = E_j \setminus \left\{ e : \left| e \cap \bigcup\nolimits_{0 \leq i < j} T_i \right| \neq 0 \right\}.$$

Consider choosing the vertices in T_j one at a time. Let $T_j^{(i)}$ be the set containing the first i vertices chosen. Let $E_j^{(i)}$ be the set of edges in E'_j not covered by $T_j^{(i)}$, with $E_j^{(0)} = E'_j$. Let

$$V_j^{(i)} = V \setminus \left(T_j^{(i)} \cup \bigcup\nolimits_{0 \leq i < j} T_i \right).$$

The hypergraph $H_i = (V_j^{(i)}, E_j^{(i)})$ is $|V_j^{(i)}|2^j/x$-thick, so the average number of edges covered by a vertex in $V_j^{(i)}$ is at least $|E_j^{(i)}|2^j/x$, and the maximum number of edges covered by a vertex in $V_j^{(i)}$ is $|E_j^{(i)}|$. Hence, with probability at least $2^j/(2x) > (2x)^{-1}$, the number of edges of $E_j^{(i)}$covered by the next vertex added is at least $|E_j^{(i)}|2^j/(2x)$. Therefore, with probability at least $(2x)^{-i}$, $|E_j^{(i)}| \leq |E'_j|(1-2^j/(2x))^i \leq |E'_j|e^{-i2^j/(2x)}$. Thus with probability at least $(2x)^{-t_j}$, $|E_j^{(t_j)}| \leq |E'_j|e^{-2(4/3)^j} \leq y2^j e^{-2(4/3)^j} \leq y2^{-(j+1)}$. (Here we use the fact that for all integers $j \geq 0$, $2^j e^{-2(4/3)^j} \leq 2^{-(j+1)}$). ■

Corollary 3.8.1 Let $x \geq 1$ and $t = \lceil 13x \rceil$. Given an $(n/x, y)$-thick hypergraph $H = (V, E)$ with n vertices and m edges, there are at least $\binom{n}{t}(2x)^{-t}$ subsets of vertices of size t that cover, for every $j \geq 0$, all but $y2^{-(j+1)}$ of the edges with thickness in the range $[n2^j/x, n2^{j+1}/x)$. ■

Using Corollary 3.8.1 we establish a lower bound on the probability that a random set of vertices contains two near transversals covering exactly the same set of edges.

Lemma 3.9 Let $x \geq 1$ and $t = \lceil 13x \rceil$. Given an $(n/x, y)$-thick hypergraph $H = (V, E)$ with n vertices and m edges, where $n \geq 4t$, the probability that two randomly chosen disjoint t-size subsets of V cover exactly the same set of edges is at least

$$\frac{1}{2(2x)^t} \left(\frac{1}{2y^{\log y}(2x)^t e^{18y}} - \frac{4t^2}{n} \right).$$

Proof: If $m = 0$, then the lemma holds trivially, so assume $m \geq 1$. Call a t-size subset of V an *attempt*, and an attempt that satisfies the condition of Corollary 3.8.1 a *good attempt.* From Lemma A.7, the number of different subsets of edges that could possibly fail to be covered by a good attempt is at most $y^{\log y} e^{18y}$. Let A be the set containing exactly these subsets. For a given attempt, define the *uncovered set* to be the subset of edges not covered by that attempt. Using Corollary 3.8.1, we find that on average each subset in A is the uncovered set for at least

$$\alpha = \binom{n}{t} (2x)^{-t} \frac{1}{y^{\log y} e^{18y}}$$

good attempts. Let $B \subseteq A$ be the set containing the subsets of edges that are the uncovered set for at most $\alpha/2$ good attempts. Then at most

$$\frac{\alpha y^{\log y} e^{18y}}{2} = \frac{1}{2}\binom{n}{t}(2x)^{-t}$$

good attempts correspond to uncovered sets in B. Hence, the probability that a single random attempt T_1 is one of the good attempts with associated uncovered set in $A \setminus B$ is at least $\frac{1}{2}(2x)^{-t}$. If this is so, then the number of attempts that are disjoint from T_1 and that cover the same edges as T_1 is at least

$$\frac{1}{2}\binom{n}{t}\frac{1}{(2x)^t y^{\log y} e^{18y}} - \left[\binom{n}{t} - \binom{n-t}{t} \right].$$

Note that we are subtracting the total number of subsets of size t that intersect T_1. Hence, the probability that a random attempt T_2 covers the same edges as T_1 without intersecting T_1 is at least

$$\left\{\frac{1}{2}\binom{n}{t}\frac{1}{(2x)^t y^{\log y} e^{18y}} - \left[\binom{n}{t} - \binom{n-t}{t}\right]\right\} \Big/ \binom{n}{t}.$$

By Lemma A.8, the probability that two randomly chosen t-size disjpoint subsets of V cover exactly the same set of vertices is at least

$$\frac{1}{2(2x)^t}\left(\frac{1}{2y^{\log y}(2x)^t e^{18y}} - \frac{4t^2}{n}\right).$$

■

Now we can place a bound on the number of transmission steps required to solve the Control Tower problem with probability of failure polynomially small in n.

Theorem 15 Let $\mathcal{A}$ be a deterministic algorithm for the Control Tower problem where $h \geq 2$ messages are placed randomly at h of the n stations. If $\mathcal{A}$ succeeds in $T(n,h)$ steps with probability at least $1 - n^{-3/4}$, then for some constant $\varepsilon > 0$, $T(n,h) \geq \varepsilon \log h \log n$.

Proof: Let $\varepsilon = 10^{-5}$. For $h \geq \log n \log\log n$, the result is trivial, since $T(n,h) \geq h \geq 10^{-5} \log h \log n$. So assume $2 \leq h \leq \log n \log\log n$. Let V be the set of n stations, and E be a set of edges, where edge i contains each station that would attempt to transmit a message in step i, if it had a message to transmit. Let $m = |E|$, and assume that $m < 10^{-5} \log h \log n < 10^{-4} \log h \ln n$. In what follows, we show that with probability at least $n^{-3/4}$ there are two disjoint sets of stations that attempt to transmit at exactly the same time steps. In terms of hypergraphs, this is equivalent to showing that with probability at least $n^{-3/4}$, a random set of h vertices contains two disjoint sets of vertices that cover exactly the same set of edges.

If $h < 2^{1000}$, then $m \leq \frac{1}{10}\log n$. If $n < 2^{10}$, then $m = 0$, and the theorem holds trivially. Otherwise, by Lemma 3.2, we can find a group of $n^{9/10}$ vertices that cover exactly the same set of edges. The probability that two of the h randomly placed messages lie in this group is at least

$$n^{-1/10}((n^{9/10} - 1)/(n-1)) \geq n^{-3/4}.$$

(Goldberg, Jerrum, Leighton, and Rao [12] prove a similar result.)

If $2^{1000} \leq h \leq \log n \log\log n \leq \log^2 n$, then $n \geq 2^{2^{500}}$ and $h \leq n^{1/10}$. If $m \leq \log h$ then $m \leq \frac{\log n}{10}$, and the argument in the previous paragraph holds; otherwise, from Lemma 3.7 with $r_1 = \frac{\log h}{10}$ and $r_2 = \frac{3\log h}{10}$, there is a subset V' of V with

$$|V'| \geq ne^{-(40m)/(x\log h)},$$

that induces a $(|V'|/x, 20m/\log h)$-thick hypergraph, for some x, $h^{1/10} \leq x \leq h^{3/10}$. By our assumption that $m \leq 10^{-4}\log h \ln n$, we have

$$|V'| \geq ne^{-\log n/(250x)} = n^{1-(250x)^{-1}},$$

and the induced hypergraph is $(|V'|/x, (\ln n)/500)$-thick. Note that for $h \geq 2^{1000}$, we have $|V'| \geq 52x + 4$ and $h^{1/2}/5 \geq 2\lceil 13x \rceil$. Furthermore, with y set to $(\ln n)/500$ and t set to $\lceil 13x \rceil$, we have

$$\frac{1}{2(2x)^t}\left(\frac{1}{2y^{\log y}(2x)^t e^{18y}} - \frac{4t^2}{n}\right) \geq n^{-1/5}.$$

Hence, the probability that at least $2\lceil 13x\rceil$ of the stations with messages belong to V' is at least

$$\begin{aligned}\left(\frac{n^{1-(250x)^{-1}}}{2n}\right)^{2\lceil 13x\rceil} &\geq n^{-1/5}2^{-2\lceil 13x\rceil}\\ &\geq n^{-2/5}.\end{aligned}$$

By Lemma 3.9, the probability that two disjoint $\lceil 13x\rceil$-size random subsets of these vertices each cover exactly the same set of edges is at least $n^{-1/5}$. Hence, the probability that two disjoint subsets of the randomly chosen set of h vertices covers exactly the same set of edges is at least $n^{-2/5}n^{-1/5} \geq n^{-3/4}$. ■

Corollary 15.1 Let $\mathcal{A}$ be a randomized algorithm for the Control Tower problem with $h \geq 2$ messages and n processors. If $\mathcal{A}$ succeeds in $T(n,h)$ steps with probability at least $1 - n^{-3/4}$, then for some constant $\varepsilon > 0$, $T(n,h) \geq \varepsilon \log h \log n$.

Proof: With $\varepsilon = 10^{-5}$, this Corollary follows from Theorem 15 and Yao [25, Theorem 1] ■

Corollary 15.2 Let $h \geq 2$, and let $\mathcal{A}$ be a randomized algorithm for direct routing of an h-relation on a 1-collision crossbar with n processors. If the expected time of $\mathcal{A}$ is $T(n,h)$ then for some constant $\varepsilon > 0$, $T(n,h) \geq \max\{h, \varepsilon \log h \log n\}$.

Proof: Let $\varepsilon = 10^{-6}$. The claim that $T(n,h) \geq h$ is trivial. Also, for all $h \geq 2$, if $n < 2^{100}$ or $h \geq n^{1/3}$, $h \geq 10^{-6} \log h \log n$, so the claimed bound on $T(n,h)$ is trivial.

In the case of $n \geq 2^{100}$ and $h < n^{1/3}$, by Yao [25, Theorem 1], we only need to show that the lower bound holds for the expected time of any deterministic algorithm over some input distribution. We choose an input distribution as follows. Let $S = \{S_1, \ldots, S_{\lceil n^{1/3}\rceil}\}$ be a set of $\lceil n^{1/3}\rceil$ disjoint groups of $\lfloor n^{1/3}\rfloor \geq n^{1/5}$ processors, and let $T = \{p_1, \ldots, p_{\lceil n^{1/3}\rceil}\}$ be a set of processors that is disjoint from each S_i, $1 \leq i \leq \lceil n^{1/3}\rceil$. For all i, $1 \leq i \leq \lceil n^{1/3}\rceil$, let h messages, each with destination p_i, be placed randomly at h processors in S_i. By Theorem 15, for any deterministic algorithm using at most $(10^{-5}/5) \log h \log n$ steps, the probability of a group S_i successfully transmitting all h messages to p_i is at most $1 - n^{-1/4}$. Hence the probability of success in all groups is at most $(1 - n^{-1/4})^{n^{1/3}} \leq e^{-n^{1/12}}$. (Notice that the analysis for each group is independent, since we are dealing with direct algorithms.) Therefore the expected number of steps required is at least $(10^{-6}) \log h \log n$. ■

4 Concluding Remarks

For appropriate choices of a and b, the a out of b protocol can be used to emulate a single step of an n-processor EREW PRAM on an n-processor crossbar. As shown in Section 2, the delay associated with this simulation is $O(\log\log n)$ whp. One way to optimize delay is to introduce parallel slackness. By simulating an m-processor PRAM on an n-processor crossbar ($m > n$) with delay $O(m/n)$, *time-processor-optimal simulations* can be obtained [17]. It would be interesting to see if the techniques of Section 2 used to analyze the a out of b protocol can be applied to obtain tight analysis of time-processor-optimal simulations.

The randomized lower bound for the Control Tower and the h-relation routing problems in Section 3 matches the upper bound of [10] up to a constant factor. In the deterministic case, however, there is factor of $\min\{\log h, \log\log n\}$ (resp., $\min\{\log h, \log\log n\} \log h$) separating the best known upper and lower bounds for the Control Tower problem (resp., h-relation routing problem). Closing this gap is an important open problem.

Acknowledgements

The authors would like to thank Noga Alon, Leslie Goldberg, Yossi Matias, Friedhelm Meyer auf der Heide, Aravind Srinivasan, Torsten Suel, and David Zuckerman for many helpful discussions.

References

[1] N. Abramson. The ALOHA system. In N. Abramson and F. Kuo, editors, *Computer-Communication Networks*. Prentice-Hall, Englewood Cliffs, NJ, 1973.

[2] N. Alon. Transversal numbers of uniform hypergraphs. *Graphs and Combinatorics*, 6:1–4, 1990.

[3] N. Alon, A. Bar-Noy, N. Linial, and D. Peleg. A lower bound for radio broadcast. *JCSS*, 43:290–298, 1991.

[4] N. Alon and J. H. Spencer. *The Probabilistic Method.* Wiley-Interscience, New York, NY, 1992.

[5] R. J. Anderson and G. L. Miller. Optical communication for pointer based algorithms. Technical Report CRI–88–14, Computer Science Department, University of Southern California, 1988.

[6] J. L. Carter and M. N. Wegman. Universal classes of hash functions. *JCSS*, 18:143–154, 1979.

[7] H. Chernoff. A measure of the asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23:493–509, 1952.

[8] V. Chvátal. The tail of the hypergeometric distribution. *Discrete Mathematics*, 25:285–287, 1979.

[9] M. Dietzfelbinger and F. Meyer auf der Heide. Simple, efficient shared memory simulations. In *Proceedings of the 5th Annual ACM Symposium on Parallel Algorithms and Architectures*, pages 110–119, June 1993.

[10] M. Geréb-Graus and T. Tsantilas. Efficient optical communication in parallel computers. In *Proceedings of the 4th Annual ACM Symposium on Parallel Algorithms and Architectures*, pages 41–48, June 1992.

[11] L. A. Goldberg and M. Jerrum. A sub-logarithmic communication algorithm for the completely connected optical communication parallel computer. Technical Report ECS–LFCS–92–234, Laboratory for Foundations of Computer Science, Department of Computer Science, University of Edinburgh, September 1992.

[12] L. A. Goldberg, M. Jerrum, F. T. Leighton, and S. B. Rao. A doubly logarithmic communication algorithm for the completely connected optical communication parallel computer. In *Proceedings of the 5th Annual ACM Symposium on Parallel Algorithms and Architectures*, pages 300–309, June 1993.

[13] L. A. Goldberg, Y. Matias, and S. B. Rao. An optical simulation of shared memory. In *Proceedings of the 6th Annual ACM Symposium on Parallel Algorithms and Architectures*, June 1994. To appear.

[14] A. G. Greenberg, P. Flajolet, and R. E. Ladner. Estimating the multiplicities of conflicts to speed their resolution in multiple access channels. *JACM*, 34:289–325, 1987.

[15] A. G. Greenberg and S. Winograd. A lower bound on the time needed in the worst case to resolve conflicts deterministically in multiple access channels. *JACM*, 32:589–596, 1985.

[16] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[17] R. Karp, M. Luby, and F. Meyer auf der Heide. Efficient PRAM simulation on a distributed memory machine. In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing*, pages 318–326, May 1992.

[18] J. Komlós and A. G. Greenberg. An asymptotically optimal nonadaptive algorithm for conflict resolution in multiple-access channels. *IEEE Transactions on Information Theory*, IT–31:302–306, 1985.

[19] C. U. Martel and T. P. Vayda. The complexity of selection resolution, conflict resolution, and maximum finding on multiple access channels. In *Proceedings of the 3rd International Workshop on Parallel Computation and VLSI Theory*, pages 401–410, 1988.

[20] R. Metcalfe and D. Boggs. Ethernet: Distributed packet switching for local computer networks. *Communications of the ACM*, 19:395–404, 1976.

[21] F. Meyer auf der Heide, C. Scheideler, and V. Stemann. Fast, simple dictionaries and shared memory simulations on distributed memory machines; upper and lower bounds. Manuscript, January 1994.

[22] A. Siegel. On universal classes of fast high performance hash functions their time-space tradeoff, and their applications. In *Proceedings of the 30th IEEE Foundations of Computer Science*, pages 20–25, November 1989. Revised Version.

[23] E. Upfal and A. Wigderson. How to share memory in a distributed system. *JACM*, 34:116–127, 1987.

[24] L. G. Valiant. General purpose parallel architectures. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science, Volume A: Algorithms and Complexity*, pages 943–971. Elsevier/MIT Press, 1990.

[25] A. C. Yao. Lower bounds by probabilistic arguments. In *Proceedings of the 24th Annual IEEE Symposium on Foundations of Computer Science*, pages 420–428, October 1983.

A Technical Lemmas

Lemma A.1 For all integers m and n such that $3 \le m \le n$, we have

$$m^2/3n \le f(m) \le m^2/n.$$

Proof: By definition,

$$f(m) = m(1 - (1 - 1/n)^{m-1}).$$

Since $(1 - 1/n)^{m-1} \ge 1 - (m-1)/n$,

$$\begin{aligned} f(m) &\le m(1 - 1 + (m-1)/n) \\ &\le m^2/n. \end{aligned}$$

For the lower bound, since $(1 - 1/n)^{m-1} \le 1 - \binom{m-1}{1}/n + \binom{m-1}{2}/n^2$,

$$\begin{aligned} f(m) &\ge m(1 - 1 + \binom{m-1}{1}/n - \binom{m-1}{2}/n^2) \\ &\ge (m(m-1)/n)(1 - (m-2)/2n) \\ &\ge m(m-1)/(2n) \\ &\ge m^2/3n. \end{aligned}$$

In the penultimate step we use $(m-2)/2n \le 1/2$, and in the last step we use $(m-1)/2 \ge m/3$ for $m \ge 3$. ■

Lemma A.2 For all integers m and n such that $6 \leq m \leq n$, we have

$$m^3/12n^2 \leq g(m) \leq m^3/n^2.$$

Proof: By definition,

$$g(m) = m(1 - (1 - 1/n)^{m-1} - ((m-1)/n)(1 - 1/n)^{m-2}.$$

Since $(1 - 1/n)^{m-1} \geq 1 - (m-1)/n$ and $(1 - 1/n)^{m-2} \geq 1 - (m-2)/n$,

$$\begin{aligned} g(m) &\leq m(1 - 1 + \binom{m-1}{1}/n - (m-1)/n + (m-1)\binom{m-2}{1}/n^2) \\ &\leq m^3/n^2. \end{aligned}$$

For the lower bound,

$$\begin{aligned} g(m) &\geq m(1 - 1 + \binom{m-1}{1}/n - \binom{m-1}{2}/n^2 + \binom{m-1}{3}/n^3 - \binom{m-1}{4}/n^4 \\ &\quad -(m-1)/n + (m-1)\binom{m-2}{1}/n^2 - (m-1)\binom{m-2}{2}/n^3 \\ &\quad +(m-1)\binom{m-2}{3}/n^4 - (m-1)\binom{m-2}{4}/n^5) \\ &\geq m((m-1)(m-2)/2n^2 - (m-1)(m-2)(m-3)/3n^3 \\ &\quad +(m-1)(m-2)(m-3)(m-4)/24n^4 - (m-1)\binom{m-2}{4}/n^5) \\ &\geq (m(m-1)(m-2)/n^2)(1/2 - (m-3)/3n) \\ &\geq m(m-1)(m-2)/6n^2 \\ &\geq m^3/12n^2. \end{aligned}$$

In the last step we use $(m-1)(m-2) \geq m^2/2$ for $m \geq 6$. ■

Lemma A.3 If n and m are integers such that $2\sqrt{n} \leq m \leq n$, then for all real $x \geq 0$,

$$f(m(1+x)) \leq (1+x)^2 f(m).$$

Proof: By the definition of f,

$$f(m(1+x)) = m(1+x)(1 - (1 - 1/n)^{m(1+x)-1}).$$

We establish the desired inequality by proving that $(1-(1-1/n)^{m-1+mx}) \leq (1+x)(1-(1-1/n)^{m-1})$, which is equivalent to showing that $(1-1/n)^{m-1+mx} \geq (1-1/n)^{m-1}(1+x) - x$. Since $(1-1/n)^{mx} \geq (1 - mx/n)$,

$$\begin{aligned} (1 - 1/n)^{m-1+mx} &\geq (1 - 1/n)^{m-1}(1 - mx/n) \\ &= (1 - 1/n)^{m-1}(1+x) - x(1 + m/n)(1 - 1/n)^{m-1} \\ &\geq (1 - 1/n)^{m-1}(1+x) - x. \end{aligned}$$

The last step follows from the fact that for $m \geq 2\sqrt{n}$, $(1 + m/n) \leq (1 - 1/n)^{1-m}$. ■

Corollary A.3.1 If n and m are integers and x is real such that $0 \leq x < 1$ and $2\sqrt{n} \leq m(1-x) \leq m \leq n$, then:

$$\begin{aligned} f(m(1+x)) &\leq (1+x)^2 f(m), \text{ and} \\ f(m(1-x)) &\geq (1-x)^2 f(m). \end{aligned}$$

Proof: The first inequality follows directly from Lemma A.3. The second inequality is proved by applying Lemma A.3 substituting $(m(1-x), 1/(1-x)-1)$ for (m, x). ■

Lemma A.4 If n and m are integers such that $n \geq 9$ and $10\sqrt{n} \leq m \leq n$, then for real $x \geq 0$,

$$g(m(1+x)) \leq (1+x)^4 g(m)$$

Proof: By the definition of g,

$$g(m(1+x)) = m(1+x)(1-(1-1/n)^{m(1+x)-1} - ((m(1+x)-1)/n)(1-1/n)^{m(1+x)-2}).$$

We establish the desired inequality by showing that $(1-(1-1/n)^{m(1+x)-1} - ((m(1+x)-1)/n)(1-1/n)^{m(1+x)-2}) \leq (1+x)^3(1-(1-1/n)^{m-1} - ((m-1)/n)(1-1/n)^{m-2})$. This is equivalent to showing that $(1-1/n)^{m-1+mx} + ((m(1+x)-1)/n)(1-1/n)^{m-2+mx} \geq (1+x)^3(1-1/n)^{m-1} + (1+x)^3((m-1)/n)(1-1/n)^{m-2} - x^3 - 3x^2 - 3x$.

$$\begin{aligned}
& (1-1/n)^{m-1+mx} + ((m(1+x)-1)/n)(1-1/n)^{m-2+mx} \\
\geq\ & (1-1/n)^{m-1}(1-mx/n) + ((m(1+x)-1)/n)(1-1/n)^{m-2} \\
& -((mx(m(1+x)-1))/n^2)(1-1/n)^{m-2} \\
\geq\ & (1-1/n)^{m-1}(1-mx/n) + ((m-1)(1+x)/n)(1-1/n)^{m-2} \\
& +(x/n)(1-1/n)^{m-2} - ((mx(m(1+x)-1))/n^2)(1-1/n)^{m-2} \\
=\ & (1+x)^3(1-1/n)^{m-1} - (x^3+3x^2+3x+mx/n)(1-1/n)^{m-1} \\
& +(1+x)^3((m-1)/n)(1-1/n)^{m-2} - (x^3+3x^2+2x)((m-1)/n)(1-1/n)^{m-2} \\
& +(x/n - m^2x/n^2 - m^2x^2/n^2 + mx/n^2)(1-1/n)^{m-2} \\
\geq\ & (1+x)^3(1-1/n)^{m-1} - (x^3+3x^2+3x+mx/n)(1-1/n)^{m-2} \\
& +(1+x)^3((m-1)/n)(1-1/n)^{m-2} - (x^3+3x^2+2x)((m-1)/n)(1-1/n)^{m-2} \\
& +(x/n - m^2x/n^2 - m^2x^2/n^2 + mx/n^2)(1-1/n)^{m-2} \\
\geq\ & (1+x)^3(1-1/n)^{m-1} + (1+x)^3((m-1)/n)(1-1/n)^{m-2} \\
& -(1-1/n)^{m-2}(x^3+3x^2+3x+mx/n+mx^3/n-x^3/n \\
& +3mx^2/n - 3x^2/n + 2mx/n - 2x/n - x/n \\
& +m^2x/n^2 + m^2x^2/n^2 - mx/n^2) \\
\geq\ & (1+x)^3(1-1/n)^{m-1} + (1+x)^3((m-1)/n)(1-1/n)^{m-2} \\
& -(1-1/n)^{m-2}(x^3(1+m/n-1/n) \\
& +3x^2(1+m/n-1/n+m^2/3n^2) + 3x(1+m/n-1/n+m^2/3n^2-m/3n^2)) \\
\geq\ & (1+x)^3(1-1/n)^{m-1} + (1+x)^3((m-1)/n)(1-1/n)^{m-2} - (x^3+3x^2+3x).
\end{aligned}$$

The last step follows from the fact that $(1+(m-1)/n+m^2/3n^2) \leq (1-1/n)^{-(m-2)}$ for $n \geq 9$ and $10\sqrt{n} \leq m \leq n$. ■

Corollary A.4.1 If n and m are integers and x is real such that $0 \leq x < 1$, $n \geq 9$, and $10\sqrt{n} \leq m(1-x) \leq m \leq n$, then:

$$\begin{aligned} g(m(1+x)) &\leq (1+x)^4 g(m), \text{ and} \\ g(m(1-x)) &\geq (1-x)^4 g(m). \end{aligned}$$

Proof: The first inequality follows directly from Lemma A.4. The second inequality is proved by applying Lemma A.4 substituting $(m(1-x), 1/(1-x)-1)$ for (m, x). ■

Lemma A.5 Let $m \geq 1$ be an integer, and let $y \geq 2$ be a real. The number of ways to choose at most m/y^3 items from a set of m items is at most $me^{m/y}$.

Proof: If $m/y^3 < 1$, then we are considering the number of ways to choose 0 items, which is $1 \leq me^{m/y}$. If $m/y^3 \geq 1$, then the number of ways to choose at most m/y^3 items can be bounded as follows. Note that we use the fact that for $y \geq 2$, $1 + \ln 2 + 3 \ln y \leq y^2$.

$$\begin{aligned}
\sum_{i=0}^{\lfloor m/y^3 \rfloor} \binom{m}{i} &\leq \frac{m}{y^3} \binom{m}{\lfloor m/y^3 \rfloor} \\
&\leq \frac{m}{y^3} \left(\frac{2em}{m/y^3} \right)^{\lfloor m/y^3 \rfloor} \\
&\leq me^{(1+\ln 2+3\ln y)m/y^k} \\
&\leq me^{m/y}
\end{aligned}$$

■

Lemma A.6 Let $y \geq 1$ be a real number, and $j \geq 0$ be an integer. The number of ways to choose at most $y2^{-(j+1)}$ items from a set of at most $y2^j$ items is at most $ye^{6y(3/2)^{-j}}$.

Proof: If $y2^{-(j+1)} < 1$ then we are considering the number of ways to choose 0 items, which is $1 \leq ye^{6y(3/2)^{-j}}$. Otherwise, we have

$$\begin{aligned}
\sum_{i=0}^{\lfloor y2^{-(j+1)} \rfloor} \binom{y2^j}{i} &\leq y2^{-(j+1)} \binom{y2^j}{\lfloor y2^{-(j+1)} \rfloor} \\
&\leq y \left(\frac{2ey2^j}{y2^{-(j+1)}} \right)^{y2^{-(j+1)}} \\
&= ye^{(1+(2j+2)\ln 2)y2^{-(j+1)}} \\
&\leq ye^{6y(3/2)^{-j}},
\end{aligned}$$

where the last inequality holds since $(1 + (2j+2)\ln 2)2^{-(j+1)} \leq 6(3/2)^{-j}$ for all integers $j \geq 0$. ■

Lemma A.7 Let $y \geq 1$ be a real number. Let $\langle S_j \rangle$ be a sequence of disjoint sets with $|S_j| \leq y2^j$ for all $j \geq 0$. The number of ways to choose a sequence of sets $\langle T_j \rangle$ with $T_j \subseteq S_j$ and $|T_j| \leq y2^{-(j+1)}$ for all $j \geq 0$, is at most $y^{\log y} e^{18y}$.

Proof: We will consider the product over all $j \geq 0$, of the number of ways to choose T_j from S_j. Note that we only need to be concerned with $j \leq \log y$, because for $j > \log y$, $y2^{-(j+1)} < 1$, so we would be choosing 0 items. By Lemma A.6, we can bound the desired product by

$$\begin{aligned}
\prod_{j=0}^{\log y} ye^{6y(3/2)^{-j}} &\leq y^{\log y} e^{6y \sum_{j \geq 0} (1.5)^{-j}} \\
&\leq y^{\log y} e^{18y}.
\end{aligned}$$

■

Lemma A.8 For all positive integers n and t such that $n \geq 4t$, we have

$$1 - \binom{n-t}{t} \Big/ \binom{n}{t} \leq 4t^2/n.$$

Proof:

$$\begin{aligned}
1 - \binom{n-t}{t} \Big/ \binom{n}{t} &\leq 1 - \left(\frac{n-2t+1}{n-t+1}\right)^t \\
&= 1 - \left(1 - \frac{t}{n-t+1}\right)^t \\
&\leq 1 - \left(1 - \frac{2t}{n}\right)^t \\
&\leq 1 - e^{-4t^2/n} \\
&\leq \frac{4t^2}{n}.
\end{aligned}$$

■