Tight Bounds for Repeated Balls-Into-Bins

Dimitrios Los
Department of Computer Science & Technology, University of Cambridge, UK

Thomas Sauerwald
Department of Computer Science & Technology, University of Cambridge, UK

Abstract

We study the *repeated balls-into-bins* process introduced by Becchetti, Clementi, Natale, Pasquale and Posta [3]. This process starts with m balls arbitrarily distributed across n bins. At each round $t = 1, 2, \ldots$, one ball is selected from each non-empty bin, and then placed it into a bin chosen independently and uniformly at random. We prove the following results:

- For any $n \leqslant m \leqslant \mathrm{poly}(n)$, we prove a lower bound of $\Omega(m/n \cdot \log n)$ on the maximum load. For the special case $m = n$, this matches the upper bound of $\mathcal{O}(\log n)$, as shown in [3]. It also provides a positive answer to the conjecture in [3] that for $m = n$ the maximum load is $\omega(\log n/ \log\log n)$ at least once in a polynomially large time interval. For $m \in [\omega(n), n \log n]$, our new lower bound disproves the conjecture in [3] that the maximum load remains $\mathcal{O}(\log n)$.
- For any $n \leqslant m \leqslant \mathrm{poly}(n)$, we prove an upper bound of $\mathcal{O}(m/n \cdot \log n)$ on the maximum load for all steps of a polynomially large time interval. This matches our lower bound up to multiplicative constants.
- For any $m \geqslant n$, our analysis also implies an $\mathcal{O}(m^2/n)$ waiting time to reach a configuration with a $\mathcal{O}(m/n \cdot \log m)$ maximum load, even for worst-case initial distributions.
- For $m \geqslant n$, we show that every ball visits every bin in $\mathcal{O}(m \log m)$ rounds. For $m = n$, this improves the previous upper bound of $\mathcal{O}(n \log^2 n)$ in [3]. We also prove that the upper bound is tight up to multiplicative constants for any $n \leqslant m \leqslant \mathrm{poly}(n)$.

2012 ACM Subject Classification Mathematics of computing → Probability and statistics; Mathematics of computing → Discrete mathematics; Theory of computation → Randomness, geometry and discrete structures; Theory of computation → Design and analysis of algorithms

Keywords and phrases Repeated balls-into-bins, self-stabilizing systems, balanced allocations, potential functions, random walks

Digital Object Identifier 10.4230/LIPIcs.STACS.2023.45

Related Version *Full Version*: `https://arxiv.org/abs/2203.12400` [25]
Brief Announcement (SPAA'22): `https://doi.org/10.1145/3490148.3538561` [24]

Supplementary Material *Software*: `https://github.com/Dim131/RBB`
archived at `swh:1:dir:106d475ab210c4248cebfdee217a0a7f40fbfcab`

1 Introduction

We consider the allocation processes involving m balls (jobs or data items) to n bins (servers or memory cells), by allowing each ball to choose from a set of randomly chosen bins. The goal is to allocate (or re-allocate) balls efficiently, while also keeping the load distribution balanced. The balls-into-bins framework has found numerous applications in hashing, load balancing, routing (we refer to the surveys [27] and [32] for more details).

A classical sequential allocation algorithm is the d-Choice process introduced by Azar, Broder, Karlin and Upfal [1] and Karp, Richard, Luby, and Meyer auf der Heide [20], where for each ball to be allocated, we sample $d \geqslant 1$ bins uniformly and then place the ball in the least loaded of the d sampled bins. It is well-known that for the One-Choice process ($d = 1$), the maximum load is w.h.p.[1] $\Theta(\log n/ \log\log n)$ for $m = n$ and $m/n + \Theta\big(\sqrt{m/n \cdot \log n}\big)$

[1] In general, with high probability refers to probability of at least $1 - n^{-c}$ for some constant $c > 0$.

40th International Symposium on Theoretical Aspects of Computer Science (STACS 2023).
Editors: Petra Berenbrink, Patricia Bouyer, Anuj Dawar, and Mamadou Moustapha Kanté; Article No. 45; pp. 45:1–45:22
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

for $m = \Omega(n \log n)$. In particular, this gap between maximum and average load grows significantly as $m/n \to \infty$, which is called the *heavily loaded case.* For $d = 2$, [1] proved that the maximum load is only $m/n + \log_2 \log n + \mathcal{O}(1)$ for $m = n$. This result was generalized by Berenbrink, Czumaj, Steger and Vöcking [6] who proved that the same guarantee also holds for $m \geqslant n$, in other words, even as $m/n \to \infty$, the difference between the maximum and average load remains a slowly growing function in n that is independent of m. This improvement of Two-Choice over One-Choice has been widely known as the "power of two choices".

In this work, we investigate the *repeated balls-into-bins (RBB)* process, introduced by Becchetti, Clementi, Natale, Pasquale and Posta [3]. In this process, there are m balls initially allocated arbitrarily across n bins. In each round, one ball is removed from each non-empty bin and then each of these balls is allocated to one bin sampled uniformly at random (see Figure 1). This setting differs from the classical balls into bins setting in that the number of balls is fixed and the amount of balls we re-allocate in each round varies from 1 to n. Unlike Two-Choice (or d-Choice), this re-allocation is performed without inspecting the load of any bin or taking additional samples.

Becchetti et al. [3] proved that for $m = n$, starting from an arbitrary configuration, w.h.p. after $\mathcal{O}(n)$ rounds, the process reaches a maximum load of $\mathcal{O}(\log n)$ and remains in such a configuration for poly(n) rounds. Thus, the RBB process is a natural instance of a self-stabilizing system, and falls into a long line of research on random-walk based algorithms for stabilization and consensus [4, 14, 17, 18, 29]. More recently, Cancrini and Posta [11] proved that the mixing time is $\mathcal{O}(L)$ where L is the maximum load at the initial configuration.

Our Results. In this work, we settle two conjectures stated in [3] and prove tight bounds for the more general case with $n \leqslant m \leqslant \text{poly}(n)$.

Becchetti et al. [3] conjectured that the $\mathcal{O}(\log n)$ upper bound holds for all $m = \mathcal{O}(n \log n)$. They also conjectured that for $m = n$, the maximum load is $\omega(\log n / \log \log n)$. We resolve both conjectures, proving an $\Omega(\frac{m}{n} \cdot \log n)$ lower bound on the maximum load w.h.p. in any interval of length $\Omega(m^2/n^2 \cdot \log^4 n)$ and for any $n \leqslant m \leqslant \text{poly}(n)$ (Lemma 3.3). This disproves the first conjecture, but confirms the second one, showing that for $m = n$, the maximum load is w.h.p. $\Theta(\log n)$.

For the case $m \geqslant n$, we also prove that starting from an arbitrary configuration after $\mathcal{O}(m^2/n)$ rounds, w.h.p. we reach a configuration with a maximum load of $\mathcal{O}(\frac{m}{n} \cdot \log m)$ (Section 4.2). For $n \leqslant m \leqslant \text{poly}(n)$, we show that the process stabilizes in such a configuration there for at least m^2 rounds (Theorem 4.11).

Becchetti et al. [3] also studied the *cover time* (or *traversal time*) of a ball, which is the time required to visit all n bins. For $m = n$, they proved an $\mathcal{O}(n \log^2 n)$ bound on the traversal time. For any $m \geqslant n$, we improve this to $\mathcal{O}(n \log m)$, and also show that it is tight up to constant factors for any $m = \text{poly}(n)$ (Section 5).

Intuition and Techniques. For the upper bound we use an exponential potential Φ with smoothing parameter $\Theta(n/m)$. Provided that Φ is poly(m), we immediately obtain the $\mathcal{O}(m/n \cdot \log m)$ bound on the maximum load. Our analysis exploits that after only $\mathcal{O}((m/n)^2)$ rounds, sufficiently many bins will become empty, which in turn will reduce the number of balls being re-allocated. This then helps to reduce the load of any non-empty bin, since these are guaranteed to lose one ball per round, but only receive in expectation less than one ball in total from the other non-empty bins. As we will prove, the actual equilibrium will have

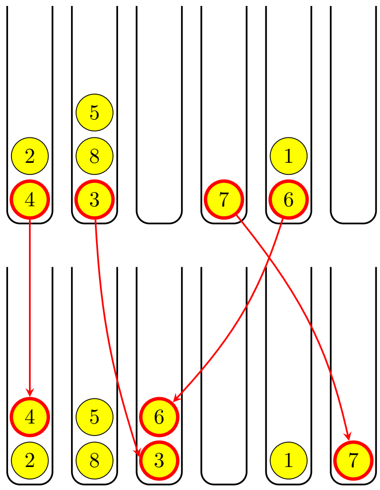

Figure 1 Illustration of one round of RBB with $m = 8$ balls and $n = 6$ bins. The balls highlighted in red are re-allocated to bins chosen randomly among $\{1, 2, \dots, 6\}$.

most bins being empty roughly every $\mathcal{O}(m/n)$ rounds. To establish this, we employ some martingale and drift-arguments to first prove that any bin which starts at load $\mathcal{O}(m/n)$, becomes empty after $\mathcal{O}((m/n)^2)$ rounds with constant probability > 0. Secondly, we prove that if this happens to a fixed bin, the empty load state will be revisited $\Omega(m/n)$ times during the next $\mathcal{O}((m/n)^2)$ rounds. In some sense, this is a generalization of the approach in [3], where they also bounded the fraction of empty bins for the case $m = n$.

A kind of reversed argument is used for the lower bound. Here, the goal is to prove that each bin is only empty every $\mathcal{O}(m/n)$ rounds on average. This shows that the RBB process can be approximated by a One-Choice process where at least an $1 - \mathcal{O}(n/m)$ fraction of the balls are allocated. For $t = \Omega(m^2/n^2 \cdot \log n)$, this yields a maximum load of $\Omega(m/n \cdot \log n)$. To prove that bins are not empty "too often", we establish a link between a quadratic potential and the number of empty bins, similar to that in [26, Lemma 6.2]. This connection essentially implies that whenever the fraction of empty bins is $\omega(n/m)$, then the quadratic potential decreases. By aggregating sufficiently over many rounds, we can conclude that, on average, the number of empty bins cannot be too large.

Further Related Work. Cancrini and Posta investigated the behavior of the RBB process for a large number of rounds, and established "propagation of chaos" [10], meaning that under some conditions on the initial load distribution, the load of the bins become eventually independent. In [10], the authors prove results for the RBB process considered here, while [12] considered more general re-allocation rules. Another variant of the RBB setting was studied in [8], where in each round one ball is deleted from each bin and an expected λn new balls arrive and are distributed in parallel to the bins. In contrast to the RBB model, this means that the number of balls in the system is not fixed.

The RBB is an instance of a discrete time closed Jackson network [19, 21]. However, in RBB, updates are happening synchronously and in parallel, while in most queuing models updates occur asynchronously based on independent point processes. As also pointed out in [10, 12], this leads to a non-reversible Markov Chain, which seems to make the computation of the stationary distribution intractable. Furthermore, formal methods have been used to prove guarantees for RBB with $m = n$ [2]. The RBB setting has also been applied to analyze protocols in short packet communications [33].

Czumaj, Riley and Scheideler [15] studied a similar re-allocation process where in each round one random ball is allocated to a random of d bin choices. These are also related to randomized rerouting protocols studied in [7, 9]. In another parallel allocation processes, Berenbrink, Czumaj, Englert, Friedetzky and Nagel [5] proved an $\mathcal{O}(\log n)$ gap for the Two-Choice process where balls are allocated in batches of n balls and was recently improved to $\mathcal{O}(\frac{\log n}{\log \log n})$ in [23].

Organization. In Section 2 we introduce some standard balls-into-bins notations and define the processes. In Section 3, we prove our lower bound on the maximum load. In Section 4, we prove an upper bound on maximum load and also analyze the time until such configuration is reached and preserved (convergence time). In Section 5, we analyze the traversal time. In Section 6, we present some empirical results on the RBB process. We conclude the paper with a summary and a few open problems in Section 7.

2 Notation and Definitions

We consider a set of n bins labeled $[n] := \{1, 2, \ldots, n\}$. By x^t we denote the n-dimensional *load vector* after t rounds, and x^0 is the initial load vector. In our processes, no balls are added or removed, and the existing m balls are only re-allocated; hence, $\sum_{i=1}^n x_i^t = m$ for all $t \geqslant 0$.

By $F^t := |\{i \in [n] \colon x_i^t = 0\}|$ we denote the number of *empty (free) bins* and by $f^t := \frac{1}{n} \cdot F^t$ the fraction of empty bins. Similarly $\kappa^t := n - F^t$ is the number of *non-empty bins*. Since it will be important to track the number of empty bins over a time interval, we also define $F_{t_0}^{t_1}$ as the total number of pairs of empty bins and rounds in the entire interval $[t_0, t_1]$, i.e.,

$$F_{t_0}^{t_1} := \sum_{t=t_0}^{t_1} F^t.$$

RBB (Repeated Balls-into-Bins Process):
Iteration: At each round $t = 1, 2, \ldots$

- For each of the $\kappa^t = n - F^t$ non-empty bins, take one ball and re-allocate it to a bin chosen independently and uniformly at random among $[n]$.

More specifically, in each round we choose κ^t bins $z_1^t, \ldots, z_{\kappa^t}^t \in [n]$ uniformly at random and the load vector at step $t+1$ is given by

$$x_i^{t+1} := x_i^t - \mathbf{1}_{x_i^t > 0} + \sum_{j=1}^{\kappa^t} \mathbf{1}_{z_j^t = i}, \quad \text{for each } i \in [n].$$

Hence, we can express the marginal load distribution of an arbitrary bin $i \in [n]$ at round $t \geqslant 0$ (i.e., having completed t iterations before), as

$$x_i^{t+1} = x_i^t - \mathbf{1}_{x_i^t > 0} + \mathsf{Bin}(\kappa^t, 1/n), \tag{2.1}$$

where with slight abuse of notation, we write $\mathsf{Bin}(\kappa^t, 1/n)$ as a placeholder for a random variable (independent of $\mathfrak{F}^t$, the entire history of the process up to round t) which has distribution $\mathsf{Bin}(\kappa^t, 1/n)$.

Similarly, assuming each bin acts as a FIFO queue on the incoming and departing balls, we can follow the trajectory of an arbitrary single ball. Only if the ball is at the front of its queue, it will be re-allocated to a bin chosen randomly from $[n]$ in the next round. A natural question is the so-called *cover time* (or *traversal time*), the expected time until every ball has been allocated to each bin [3]. This is related to the well-studied *cover time* of parallel random walks on graphs, but with the constraint that only one walk can leave each vertex (=bin) at a time.

3 Lower Bound on the Maximum Load for $n \leqslant m \leqslant \mathrm{poly}(n)$

In this section, we prove that w.h.p. the maximum load becomes $\Omega\left(\frac{m}{n} \cdot \log n\right)$ at least once in every $\mathcal{O}(\frac{m^2}{n^2} \cdot \log^4 n)$ rounds, for any $n \leqslant m \leqslant \mathrm{poly}(n)$. This matches the upper bound in Section 4 up to multiplicative constants and also settles two conjectures in [3].

On a high level, the lower bound follows by showing that in a long enough interval, w.h.p. a constant fraction of the rounds have an $\mathcal{O}(n/m)$ fraction of empty bins. Then, we couple the process with the One-Choice process, to show that the maximum load must be w.h.p. at least $\Omega(\frac{m}{n} \cdot \log n)$.

In order to bound the number of empty bins in an interval we make use of the *quadratic potential function*, defined as

$$\Upsilon^t := \sum_{i=1}^{n} (x_i^t)^2,$$

where x_i^t is the load of bin $i \in [n]$ at round t. We then prove the following relation between the expected change of Υ^t and the number of empty bins F^t in round t:

▶ **Lemma 3.1.** *Consider the RBB setting with any $m \geqslant 1$. Then, for any round $t \geqslant 0$,*

$$\mathbf{E}\left[\, \Upsilon^{t+1} \mid \mathfrak{F}^t \,\right] \leqslant \Upsilon^t - 2 \cdot \frac{m}{n} \cdot F^t + 2n.$$

Proof. Let us define the binomial random variable $Z \sim \mathrm{Bin}(\kappa^t, \frac{1}{n})$. For any bin $i \in [n]$ with load $x_i^t \geqslant 1$,

$$\begin{aligned}
\mathbf{E}\left[\, \Upsilon_i^{t+1} \mid \mathfrak{F}^t \,\right] &= \sum_{z=0}^{\kappa^t} (x_i^t + z - 1)^2 \cdot \binom{\kappa^t}{z} \cdot \frac{1}{n^z} \cdot \left(1 - \frac{1}{n}\right)^{\kappa^t - z} \\
&= (x_i^t)^2 \cdot \sum_{z=0}^{\kappa^t} \binom{\kappa^t}{z} \cdot \frac{1}{n^z} \cdot \left(1 - \frac{1}{n}\right)^{\kappa^t - z} \\
&\quad + 2 \cdot x_i^t \cdot \sum_{z=0}^{\kappa^t} (z-1) \cdot \binom{\kappa^t}{z} \cdot \frac{1}{n^z} \cdot \left(1 - \frac{1}{n}\right)^{\kappa^t - z} \\
&\quad + \sum_{z=0}^{\kappa^t} (z-1)^2 \cdot \binom{\kappa^t}{z} \cdot \frac{1}{n^z} \cdot \left(1 - \frac{1}{n}\right)^{\kappa^t - z} \\
&= (x_i^t)^2 \cdot \mathbf{E}\left[\, Z \,\right] + 2 \cdot x_i^t \cdot \mathbf{E}\left[\, Z - 1 \,\right] + \mathbf{E}\left[\, (Z-1)^2 \,\right] \\
&\overset{(a)}{=} (x_i^t)^2 + 2 \cdot x_i^t \cdot \left(\frac{\kappa^t}{n} - 1\right) + \kappa^t \cdot (\kappa^t - 1) \cdot \frac{1}{n^2} - \frac{\kappa^t}{n} + 1 \\
&\leqslant (x_i^t)^2 + 2 \cdot x_i^t \cdot \left(\frac{\kappa^t}{n} - 1\right) + 2,
\end{aligned}$$

having used in (a) that $\mathbf{E}[\,Z\,] = \frac{\kappa^t}{n}$ and $\mathbf{E}\left[\,Z^2\,\right] = \kappa^t \cdot \frac{1}{n} \cdot (1 - \frac{1}{n}) + (\kappa^t)^2 \cdot (\frac{1}{n})^2$, and thus

$$\mathbf{E}\left[\,(Z-1)^2\,\right] = \kappa^t \cdot \frac{1}{n} \cdot \left(1 - \frac{1}{n}\right) + (\kappa^t)^2 \cdot \left(\frac{1}{n}\right)^2 - 2 \cdot \frac{\kappa^t}{n} + 1 = \kappa^t \cdot (\kappa^t - 1) \cdot \frac{1}{n^2} - \frac{\kappa^t}{n} + 1.$$

Similarly for an empty bin $i \in [n]$ with $x_i^t = 0$, the contribution is

$$\mathbf{E}\left[\,\Upsilon_i^{t+1} \mid \mathfrak{F}^t\,\right] = \sum_{z=0}^{\kappa^t} z^2 \cdot \binom{\kappa^t}{z} \cdot \frac{1}{n^z} \cdot \left(1 - \frac{1}{n}\right)^{\kappa^t - z} = \frac{\kappa^t}{n} + \frac{\kappa^t \cdot (\kappa^t - 1)}{n^2}.$$

Hence, by aggregating the contributions of the κ^t bins non-empty bins and the $n - \kappa^t$ empty bins we obtain

$$\begin{aligned}
\mathbf{E}\left[\,\Upsilon^{t+1} \mid \mathfrak{F}^t\,\right] &\leqslant \Upsilon^t + \sum_{i \in [n] \colon x_i^t \geqslant 1} \left(2 \cdot x_i^t \cdot \left(\frac{\kappa^t}{n} - 1\right) + 2\right) + \sum_{i \in [n] \colon x_i^t = 0} \left(\frac{\kappa^t}{n} + \frac{\kappa^t \cdot (\kappa^t - 1)}{n^2}\right) \\
&\leqslant \Upsilon^t + \left(\frac{\kappa^t}{n} - 1\right) \cdot 2 \cdot m + 2\kappa^t + (n - \kappa^t) \cdot 2 \\
&= \Upsilon^t - 2 \cdot \frac{m}{n} \cdot F^t + 2n,
\end{aligned}$$

where in the last inequality we used that $\kappa^t \leqslant n$. This concludes the proof. ◀

The key insight is that the quadratic potential drops in expectation as soon as the fraction of empty bins is of order $\Omega(n/m)$. This is crucial to upper bound the number of empty bins in an interval. This relation is similar to the ones used in [23, 26], where an interplay between the quadratic potential and the absolute value potential was used to show that the absolute value potential is small in a constant fraction of the rounds.

The next lemma shows that for any sufficiently long interval, either there is a maximum load that is $\Omega(m/n \cdot \log n)$ or the fraction of empty bins in the interval is $\mathcal{O}(n/m)$. Note that we indeed need the interval to be long enough as starting with the perfectly balanced load vector may require several rounds to reach a gap of $\Omega(\frac{m}{n} \log n)$ even for the One-Choice process.

▶ **Lemma 3.2.** *Consider the RBB process with any $n \leqslant m \leqslant n^k$ for some constant $k \geqslant 1$ and any $1 \leqslant \hat{c} \leqslant n$. Then, for any $t_0 \geqslant 0$ and $t_1 := t_0 + \hat{c} \cdot \left(\frac{m}{n} \cdot \log n\right)^2$,*

$$\mathbf{Pr}\left[\,\left\{F_{t_0}^{t_1} < \frac{n^2}{4m} \cdot (t_1 - t_0 + 1)\right\} \cup \bigcup_{t \in [t_0, t_1]} \left\{\max_{i \in [n]} x_i^t > \frac{m}{n} \cdot \log n\right\} \,\middle|\, \mathfrak{F}^{t_0}\,\right] \geqslant 1 - e^{-\frac{\hat{c}}{18}}.$$

Proof. Consider an arbitrary step t_0 and filtration $\mathfrak{F}^{t_0}$. We can further assume that $\{\max_{i \in [n]} x_i^{t_0} \leqslant \frac{m}{n} \cdot \log n\}$ holds, otherwise the conclusion trivially follows.

For any $t \geqslant t_0$, we define the sequence

$$Z^t := \Upsilon^t - 2 \cdot (t - t_0) \cdot n + 2 \cdot \frac{m}{n} \cdot F_{t_0}^{t-1},$$

where $F_{t_0}^{t_0 - 1} = 0$. This sequence forms a super-martingale since by Lemma 3.1,

$$\begin{aligned}
\mathbf{E}\left[\,Z^{t+1} \mid \mathfrak{F}^t\,\right] &= \mathbf{E}\left[\,\Upsilon^{t+1} - 2 \cdot (t - t_0 + 1) \cdot n + 2 \cdot \frac{m}{n} \cdot F_{t_0}^t \,\middle|\, \mathfrak{F}^t\,\right] \\
&= \mathbf{E}\left[\,\Upsilon^{t+1} \mid \mathfrak{F}^t\,\right] - 2 \cdot (t - t_0 + 1) \cdot n + 2 \cdot \frac{m}{n} \cdot F_{t_0}^t \\
&\leqslant \Upsilon^t + 2 \cdot n - 2 \cdot \frac{m}{n} \cdot F^t - 2 \cdot (t - t_0 + 1) \cdot n + 2 \cdot \frac{m}{n} \cdot F_{t_0}^t
\end{aligned}$$

$$= \Upsilon^t - 2 \cdot (t - t_0) \cdot n + 2 \cdot \frac{m}{n} \cdot F_{t_0}^{t-1}$$
$$= Z^t.$$

Further, let $\tau := \min\{t \geqslant t_0 \colon \max_{i \in [n]} x_i^t > \frac{m}{n} \cdot \log n\}$ and consider the stopped random variable

$$\widetilde{Z}^t := Z^{t \wedge \tau},$$

which is then also a super-martingale.

To prove concentration of $\widetilde{Z}^t$, we will now derive an upper bound on $\left|\widetilde{Z}^{t+1} - \widetilde{Z}^t\right|$ conditional on $\mathfrak{F}^t$.

Case 1: $t \geqslant \tau$. In this case, $\widetilde{Z}^{t+1} = Z^{(t+1)\wedge\tau} = Z^\tau$, and similarly, $\widetilde{Z}^t = Z^{t\wedge\tau} = Z^\tau$, so $|\widetilde{Z}^{t+1} - \widetilde{Z}^t| = 0$.

Case 2: $t < \tau$. Hence for t we have $\max_{i\in[n]} x_i^t \leqslant \frac{m}{n} \cdot \log n$ and thus Lemma A.2 implies that the biggest change in the quadratic potential is w.h.p. at most $2 \cdot m \cdot \log n + 4n$ and under this condition, using that $m \geqslant n$,

$$|\widetilde{Z}^{t+1} - \widetilde{Z}^t| \leqslant 2 \cdot m \cdot \log n + 4n + 2 \cdot \frac{m}{n} \cdot n \leqslant 3 \cdot m \cdot \log n.$$

Combining the two cases above, we conclude,

$$\mathbf{Pr}\left[\bigcap_{t\in[t_0,t_1-1]} \left\{|\widetilde{Z}^{t+1} - \widetilde{Z}^t| \leqslant 3 \cdot m \cdot \log n\right\}\right] \geqslant 1 - n^{-\omega(1)} \cdot (t_1 - t_0) \geqslant 1 - n^{-\omega(1)},$$

since $t_1 - t_0 \leqslant \mathrm{poly}(n)$.

Using the concentration inequality Theorem A.4 with bad event, $\mathcal{B}^t := \neg \bigcap_{t\in[t_0,t]}\{|\widetilde{Z}^{t+1} - \widetilde{Z}^t| \leqslant 3 \cdot m \cdot \log n\}$ and $\lambda = \hat{c} \cdot \frac{m^2}{n} \cdot \log^2 n$, we get

$$\mathbf{Pr}\left[\widetilde{Z}^{t_1+1} - \widetilde{Z}^{t_0} > \lambda\right] \leqslant \exp\left(-\frac{\lambda^2}{2 \cdot \sum_{t=t_0}^{t_1} (3 \cdot m \cdot \log n)^2}\right) + \mathbf{Pr}[\mathcal{B}]$$
$$= \exp\left(-\frac{\hat{c}^2 \cdot \left(\frac{m^2}{n} \cdot \log^2 n\right)^2}{18 \cdot \hat{c} \cdot \left(\frac{m}{n} \cdot \log n\right)^2 \cdot (m \cdot \log n)^2}\right) + \mathbf{Pr}[\mathcal{B}]$$
$$\leqslant e^{-\frac{\hat{c}}{18}} + n^{-\omega(1)} \leqslant 2 \cdot e^{-\frac{\hat{c}}{18}}.$$

Thus,

$$\mathbf{Pr}\left[\{Z^{t_1+1} \leqslant Z^{t_0} + \lambda\} \cup \bigcup_{t\in[t_0,t_1]} \left\{\max_{i\in[n]} x_i^t \geqslant \frac{m}{n} \cdot \log n\right\}\right] < 1 - 2 \cdot e^{-\frac{\hat{c}}{18}}.$$

Assume that $\{Z^{t_1+1} \leqslant Z^{t_0} + \lambda\}$ holds. Our aim is to show that $\{F_{t_0}^{t_1} < \frac{4n^2}{m} \cdot (t_1 - t_0 + 1)\}$ also holds. For the sake of a contradiction, assume that

$$F_{t_0}^{t_1} \geqslant \frac{4n^2}{m} \cdot (t_1 - t_0 + 1).$$

By $\{Z^{t_1+1} \leqslant Z^{t_0} + \lambda\}$, we have that

$$\Upsilon^{t_1+1} - 2 \cdot (t_1 - t_0 + 1) \cdot n + 2 \cdot \frac{m}{n} \cdot F_{t_0}^{t_1} \leqslant \Upsilon^{t_0} + \lambda.$$

Rearranging the inequality above gives

$$\begin{aligned}\Upsilon^{t_1+1} &\leqslant \Upsilon^{t_0} + \lambda + 2 \cdot (t_1 - t_0 + 1) \cdot n - 2 \cdot \frac{m}{n} \cdot F_{t_0}^{t_1} \\ &\leqslant \Upsilon^{t_0} + \lambda + 2 \cdot (t_1 - t_0 + 1) \cdot n - 8 \cdot n \cdot (t_1 - t_0 + 1) \\ &\leqslant \Upsilon^{t_0} + \lambda - 6 \cdot (t_1 - t_0 + 1) \cdot n. \end{aligned} \quad (3.1)$$

Recall that we start from a round t_0 where $\{\max_{i\in[n]} x_i^{t_0} \leqslant \frac{m}{n} \cdot \log n\}$ holds, and therefore also $\{\Upsilon^{t_0} \leqslant n \cdot \big(\frac{m}{n} \cdot \log n\big)^2\}$ holds. Thus, by (3.1) we have

$$\Upsilon^{t_1+1} \leqslant n \cdot \left(\frac{m}{n} \cdot \log n\right)^2 + \hat{c} \cdot n \cdot \left(\frac{m}{n} \cdot \log n\right)^2 - 6 \cdot \hat{c} \cdot \left(\frac{m}{n} \cdot \log n\right)^2 \cdot n < 0$$

which is a contradiction for large n since $\hat{c} \geqslant 1$. We conclude that if $Z^{t_1+1} \leqslant Z^{t_0} + \lambda$, then $F_{t_0}^{t_1} < \frac{n^2}{4m} \cdot (t_1 - t_0 + 1)$ or the stopping time was reached, i.e.

$$\mathbf{Pr}\left[\left\{F_{t_0}^{t_1} < \frac{n^2}{4m} \cdot (t_1 - t_0 + 1)\right\} \cup \bigcup_{t\in[t_0,t_1]} \left\{\max_{i\in[n]} x_i^t \geqslant \frac{m}{n} \cdot \log n\right\}\right] \geqslant 1 - 2 \cdot e^{-\frac{\hat{c}}{18}}. \quad ◀$$

To complete the derivation of the lower bound we need to show that in an interval of length $T = \Theta((\frac{m}{n} \cdot \log n)^2)$ with an $\mathcal{O}(n/m)$ fraction of empty bins, the maximum load is $\Omega(\frac{m}{n} \cdot \log n)$. This follows by coupling the allocations of the RBB process in the interval with a One-Choice process with $T \cdot (1 - \mathcal{O}(n/m))$ balls. By the following standard expression, for the maximum load, setting $c := \frac{(1-\gamma)^2}{200} \cdot \frac{1}{\gamma^2}$ (for $\gamma = \Theta(\frac{n}{m})$), we get the desired lower bound on the maximum load for the RBB setting.

[cf. [26, Lemma 10.4]] Consider the One-Choice process with $m = cn \log n$ balls, for any $c \geqslant 1/\log n$. Then, we have

$$\mathbf{Pr}\left[\max_{i\in[n]} x_i^m \geqslant \left(c + \frac{\sqrt{c}}{10}\right) \cdot \log n\right] \geqslant 1 - n^{-2}.$$

Putting the lemmas together, we get the desired lower bound.

▶ **Lemma 3.3.** *Consider the RBB process with any $n \leqslant m \leqslant n^k$ for some constant $k \geqslant 1$ and let $\gamma := \frac{n}{4m}$. Then, for any round $t_0 \geqslant 0$ and for $t_1 := t_0 + \frac{1-\gamma}{200} \cdot \frac{1}{\gamma^2} \cdot \log^4 n$,*

$$\mathbf{Pr}\left[\bigcup_{t\in[t_0,t_1]} \left\{\max_{i\in[n]} x_i^t \geqslant 0.008 \cdot \frac{m}{n} \cdot \log n\right\}\right] \geqslant 1 - n^{-1}.$$

Proof. Using Lemma 3.2 (for $k := \frac{1-\gamma}{200} \cdot 16 \cdot \log^2 n \geqslant 3 \cdot 18 \cdot \log n$), we have for $t_1 = t_0 + \frac{1-\gamma}{200} \cdot \frac{1}{\gamma^2} \cdot \log^4 n$,

$$\mathbf{Pr}\left[\left\{F_{t_0}^{t_1} < \frac{n^2}{4m} \cdot (t_1 - t_0 + 1)\right\} \cup \bigcup_{t\in[t_0,t_1]} \left\{\max_{i\in[n]} x_i^t \geqslant \frac{m}{n} \cdot \log n\right\}\right] \geqslant 1 - n^{-2}. \quad (3.2)$$

Consider the $\log^3 n$ sub-intervals $\mathcal{I}_1, \ldots, \mathcal{I}_{\log^3 n}$ of length $\Delta = \frac{1-\gamma}{200} \cdot \frac{1}{\gamma^2} \cdot \log n$ with starting points $s_j := t_0 + \Delta \cdot (j - 1)$. We also define the events for $j \in [\log^3 n]$,

$$\mathcal{C}_j := \left\{F_{s_j}^{s_j+\Delta} < \frac{n^2}{4m} \cdot \Delta\right\}.$$

Running the repeated balls-into-bins process over the interval $[s_j, s_j + \Delta]$ involves reallocating $\Delta \cdot n - F_{s_j}^{s_j+\Delta}$ balls, meaning we sample $\Delta \cdot n - F_{s_j}^{s_j+\Delta}$ many times a bin uniformly at random. So if the event $\mathcal{C}_j$ holds, then we sample in total

$$\Delta \cdot n - \frac{n^2}{4m} \cdot \Delta = (1-\gamma) \cdot \Delta \cdot n =: \overline{m}$$

bins. Hence these re-allocations correspond to a ONE-CHOICE process with $\overline{m}$ balls into n bins; let us denote its load vector by y^t for any round $t \geqslant 0$ starting from the empty load configuration. By Section 3 with $c := \frac{(1-\gamma)^2}{200} \cdot \frac{1}{\gamma^2}$, we obtain

$$\mathbf{Pr}\left[\max_{i\in[n]} y_i^{\overline{m}} \geqslant \left(c + \frac{\sqrt{c}}{10}\right) \cdot \log n\right] \geqslant 1 - n^{-2}.$$

Further, note that

$$\begin{aligned}\max_{i\in[n]} y_i^{\overline{m}} &\geqslant \left(\frac{(1-\gamma)^2}{200\gamma^2} + \frac{1}{10} \cdot \sqrt{\frac{(1-\gamma)^2}{200\gamma^2}}\right) \cdot \log n \\ &\geqslant \frac{1-\gamma}{200\gamma^2} \cdot \log n + 0.002 \cdot \frac{\log n}{\gamma} = \Delta + 0.002 \cdot \frac{\log n}{\gamma}.\end{aligned}$$

In Δ rounds, at most Δ balls can be removed from any single bin $i \in [n]$, so for any bin $i \in [n]$, $x_i^{s_j+\Delta} \geqslant x_i^{s_j} + y_i^{\overline{m}} - \Delta \geqslant y_i^{\overline{m}} - \Delta$. and hence

$$\max_{i\in[n]} x_i^{s_j+\Delta} \geqslant \max_{i\in[n]} y_i^{\overline{m}} - \Delta \geqslant 0.002 \cdot \frac{\log n}{\gamma}.$$

Defining for any round $t \geqslant 0$

$$\mathcal{E}^t := \left\{\max_{i\in[n]} x_i^t \geqslant 0.008 \cdot \frac{m}{n} \cdot \log n\right\},$$

we have shown that for any $1 \leqslant j \leqslant \log^3 n$,

$$\mathbf{Pr}\left[\, \mathcal{E}^{s_j+\Delta} \cup \neg\mathcal{C}_j \,\right] \geqslant 1 - n^{-2}.$$

By taking the union bound over the $\log^3 n$ sub-intervals, we conclude

$$\mathbf{Pr}\left[\bigcap_{j\in[\log^3 n]} \left(\left\{\max_{i\in[n]} x_i^{s_j+\Delta} \geqslant 0.008 \cdot \frac{m}{n} \cdot \log n\right\} \cup \neg\mathcal{C}_j\right)\right] \geqslant 1 - (\log^3 n) \cdot n^{-2}. \tag{3.3}$$

Assuming that $\left\{F_{t_0}^{t_1} < \frac{n^2}{4m} \cdot (t_1 - t_0 + 1)\right\}$ holds, then using the pigeonhole principle, at least one of these intervals j satisfies $\mathcal{C}_j$, i.e., $\{\cup_{j\in[\log^3 n]} \mathcal{C}_j\}$ holds. Hence, by the union bound of Equation (3.2) and Equation (3.3) we conclude that

$$\begin{aligned}\mathbf{Pr}\left[\bigcup_{t\in[t_0,t_1]} \mathcal{E}^t\right] &\geqslant \mathbf{Pr}\left[\bigcup_{j\in[\log^3 n]} \mathcal{E}^{s_j+\Delta}\right] \\ &\geqslant 1 - \mathbf{Pr}\left[\neg \bigcap_{j\in[\log^3 n]} \left(\mathcal{E}^{s_j+\Delta} \cup \neg\mathcal{C}_j\right) \cup \bigcap_{j\in[\log^3 n]} \neg\mathcal{C}_j\right] \\ &\geqslant \mathbf{Pr}\left[\bigcup_{j\in[\log^3 n]} \left(\mathcal{E}^{s_j+\Delta} \cup \neg\mathcal{C}_j\right)\right] - \mathbf{Pr}\left[\bigcap_{j\in[\log^3 n]} \neg\mathcal{C}_j\right] \\ &\geqslant 1 - n^{-2} - (\log^3 n) \cdot n^{-2} \geqslant 1 - n^{-1}.\end{aligned}$$

◀

4 Upper Bounds on the Maximum Load and Convergence Time

In this section, we outline the proofs for the $\mathcal{O}(\frac{m}{n} \cdot \log n)$ matching upper bound on the maximum load and the $\mathcal{O}(m^2/n)$ upper bound on the convergence time of the RBB process for any $n \leqslant m \leqslant \mathrm{poly}(n)$. The omitted proofs and details can be found in the full version [25]. In Section 4.1, we introduce the exponential potential function and demonstrate its use on the simpler setting with $m \leqslant n$ balls, and in Section 4.2 we outline the proof for the more challenging case with $m \geqslant n$ balls.

4.1 The Exponential Potential and an Upper Bound for $m \leqslant n$

For the upper bounds, we make use of the *exponential potential function* defined as

$$\Phi^t := \Phi^t(\alpha) := \sum_{i=1}^{n} \Phi_i^t := \sum_{i=1}^{n} e^{\alpha x_i^t},$$

where x_i^t is the load of bin $i \in [n]$ at round $t \geqslant 0$ and $\alpha > 0$ is a smoothing parameter. For any round t with $\Phi^t = \mathrm{poly}(n)$, we can deduce that

$$\max_{i \in [n]} x_i^t = \mathcal{O}\Big(\frac{\log n}{\alpha}\Big).$$

By choosing a smoothing parameter $\alpha = \Theta(n/m)$, this will give the desired bound on the maximum load.

We start by giving a general formula for the expected change of Φ over one step.

▶ **Lemma 4.1.** *Consider the RBB process with any $m \geqslant n$ and the potential $\Phi := \Phi(\alpha)$ for any $\alpha > 0$. Then, for any round $t \geqslant 0$,*

$$\mathbf{E}\left[\, \Phi^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \Phi^t \cdot e^{-\alpha} \cdot e^{\frac{e^\alpha - 1}{n} \cdot \kappa^t} + (n - \kappa^t) \cdot e^{\frac{e^\alpha - 1}{n} \cdot \kappa^t}.$$

Proof. Consider the expected contribution of a bin $i \in [n]$ with $x_i^t \geqslant 1$,

$$\begin{aligned}
\mathbf{E}\left[\, \Phi_i^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] &= \sum_{z=0}^{\kappa^t} e^{\alpha(x_i^t + z - 1)} \cdot \binom{\kappa^t}{z} \cdot \Big(\frac{1}{n}\Big)^z \cdot \Big(1 - \frac{1}{n}\Big)^{\kappa^t - z} \\
&= \Phi_i^t \cdot e^{-\alpha} \cdot \sum_{z=0}^{\kappa^t} \binom{\kappa^t}{z} \cdot \Big(\frac{e^\alpha}{n}\Big)^z \cdot \Big(1 - \frac{1}{n}\Big)^{\kappa^t - z} \\
&\overset{(a)}{=} \Phi_i^t \cdot e^{-\alpha} \cdot \Big(1 - \frac{1}{n} + \frac{e^\alpha}{n}\Big)^{\kappa^t} \\
&\overset{(b)}{\leqslant} \Phi_i^t \cdot e^{-\alpha} \cdot e^{\frac{e^\alpha - 1}{n} \cdot \kappa^t},
\end{aligned}$$

using in (a) the binomial identity $\sum_{z=0}^{k} \binom{k}{z} p^z q^{k-z} = (p+q)^k$ and in (b) that $1 + z \leqslant e^z$ for any $z \geqslant 0$.

For an empty bin $i \in [n]$, its expected contribution is

$$\mathbf{E}\left[\, \Phi_i^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] = \sum_{z=0}^{\kappa^t} \binom{\kappa^t}{z} \cdot e^{\alpha z} \cdot \Big(\frac{1}{n}\Big)^z \cdot \Big(1 - \frac{1}{n}\Big)^{\kappa^t - z} = \Big(1 - \frac{1}{n} + \frac{e^\alpha}{n}\Big)^{\kappa^t} \leqslant e^{\frac{e^\alpha - 1}{n} \cdot \kappa^t}.$$

Aggregating over all bins, we have

$$\begin{aligned}
\mathbf{E}\left[\, \Phi^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] &= \sum_{i \in [n] : x_i^t \geqslant 1} \mathbf{E}\left[\, \Phi_i^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] + \sum_{i \in [n] : x_i^t = 0} \mathbf{E}\left[\, \Phi_i^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \\
&\leqslant \Phi^t \cdot e^{-\alpha} \cdot e^{\frac{e^\alpha - 1}{n} \cdot \kappa^t} + (n - \kappa^t) \cdot e^{\frac{e^\alpha - 1}{n} \cdot \kappa^t}.
\end{aligned}$$

◀

Now, we will investigate the simpler setting where m is much smaller than n, to demonstrate the use of the exponential potential function. This implies that in each round, deterministically at least $n - m$ bins are empty. As we prove below, this implies for example, that for $m = \frac{n}{\log n}$ we get w.h.p. a maximum load of $\mathcal{O}(\frac{\log n}{\log \log n})$ after $\mathcal{O}(\frac{n}{\log n})$ rounds.

▶ **Lemma 4.2.** *Consider the RBB process with $m \leqslant \frac{1}{e^2} n$. Then for any round $t \geqslant 2m$,*

$$\mathbf{Pr}\left[\max_{i \in [n]} x_i^t \leqslant 4 \cdot \frac{\log n}{\log\left(\frac{n}{em}\right)} \right] \geqslant 1 - n^{-2}.$$

Proof. We will use the potential $\Phi := \Phi(\alpha)$ with $\alpha := \log\left(\frac{n}{em}\right) \geqslant 1$ (since $m \leqslant \frac{1}{e^2}n$). Note that for $m = o(n)$ the potential is super-exponential, as in [22]. Since $\kappa^t \leqslant m$, we have

$$\frac{e^\alpha - 1}{n} \cdot \kappa^t \leqslant e^\alpha \cdot \frac{m}{n} = \frac{1}{e}.$$

Hence, by using Lemma 4.1,

$$\begin{aligned} \mathbf{E}\left[\Phi^{t+1} \mid \mathfrak{F}^t \right] &\leqslant \Phi^t \cdot e^{-\alpha} \cdot e^{\frac{e^\alpha - 1}{n} \cdot \kappa^t} + (n - \kappa^t) \cdot e^{\frac{e^\alpha - 1}{n} \cdot \kappa^t} \\ &\leqslant \Phi^t \cdot e^{-\alpha} \cdot e^{1/e} + e \cdot n \\ &\leqslant \Phi^t \cdot e^{-\frac{\alpha}{2}} + e \cdot n, \end{aligned}$$

using in the last inequality that $\alpha \geqslant 1$.

At round $t = 0$ we have $\Phi^0 \leqslant e^{\alpha m}$. Hence applying Lemma A.5, we have for any $t \geqslant 2m$,

$$\mathbf{E}\left[\Phi^t \right] \leqslant e^{\alpha m} \cdot e^{-\frac{1}{2} \cdot \alpha t} + \frac{e \cdot n}{1 - e^{-\frac{\alpha}{2}}} \leqslant 1 + \frac{e \cdot n}{1 - e^{-1/2}} \leqslant 3e \cdot n.$$

By applying Markov's inequality for $t \geqslant 2m$,

$$\mathbf{Pr}\left[\Phi^t \leqslant 3e \cdot n^3 \right] \geqslant 1 - n^{-2}.$$

When $\{\Phi^t \leqslant 3e \cdot n^3\}$ holds, we have for any bin $i \in [n]$,

$$x_i^t \leqslant \frac{1}{\alpha} \cdot (\log(3e) + 3 \log n) \leqslant 4 \cdot \frac{\log n}{\log\left(\frac{n}{em}\right)},$$

completing the proof. ◀

4.2 Upper Bound for $m \geqslant n$

We now turn to outlining the proof for the more challenging case of $m \geqslant n$. The omitted proofs an be found in the full version [25].

On a high level, we show that in a large enough interval, an $\Omega(m/n)$ fraction of the bins are empty, the opposite of what we had in Section 3. This follows through a coupling with an idealized version of the process, which is simpler to analyze. Then, in an interval with $\Omega(m/n)$ fraction of empty bins, the exponential potential with a sufficiently small smoothing parameter $\alpha = \Theta(n/m)$ drops in expectation, at some point becoming $\text{poly}(n)$ and implying the $\mathcal{O}(\frac{m}{n} \cdot \log n)$ maximum load (**convergence**). Then, with a similar analysis over a slightly smaller interval we show that it remains in such a configuration for $\mathcal{O}(\frac{m}{n} \cdot \log n)$ rounds (**stabilization**).

In the analysis, we make use of the following bound on the expected change of Φ, which is a restatement of Lemma 4.1 based on the fraction of empty bins f^t.

▶ **Lemma 4.3.** *Consider the RBB process with any $m \geqslant n$ and the potential $\Phi := \Phi(\alpha)$ with any $0 < \alpha < 1.5$. Then, for any round $t \geqslant 0$,*

$$\mathbf{E}\left[\ \Phi^{t+1} \mid \mathfrak{F}^t\ \right] \leqslant \Phi^t \cdot e^{\alpha^2 - \alpha f^t} + 6n,$$

In particular, when the fraction of empty bins satisfies $f^t = \Omega(\alpha)$, the potential drops in expectation over one round, as was trivially the case for $m \ll n$. So it will be central to our analysis to prove a lower bound on the fraction of empty bins in a sufficiently long interval. This idea is inspired by [3, Lemma 19], who proved that in case of $m = n$, for each round, a constant fraction of the bins are empty with very high probability. This is useful, as it implies a constant additive drift for the load of each non-empty bin, which will drop by a constant term > 0 in expectation.

However, for general $m \gg n$, there will be starting configurations in which all bins remain non-empty for several rounds. Only if the process runs for a sufficiently long time, a small fraction of bins will become (and, to some extent, remain) empty. The following lemma quantifies this behavior and proves that, after a waiting time of $\mathcal{O}((m/n)^2)$ (the square of the average load), a fraction of $\mathcal{O}(n/m)$ of the bins will be empty per round on average. Hence for a time interval of length $(m/n)^2$, the aggregated "empty bin/round pairs" will be $\approx (m/n)^2 \cdot n \cdot (n/m) = m$.

[Key Lemma for the Upper Bound] Consider the RBB process with $m \geqslant n$ and any round $t_0 \geqslant 0$. Then, for round $t_3 := t_0 + 744(m/n)^2$ it holds that

$$\mathbf{Pr}\left[\ F_{t_0}^{t_3} \geqslant \frac{1}{384} \cdot m \ \middle|\ \mathfrak{F}^{t_0}\ \right] \geqslant 1 - e^{-\Omega(n)}.$$

First, let us remark that for the simpler case $m = n$, a stronger result was shown in [3, Lemma 1], proving that for *any* round $t \geqslant 1$, $F^t = \Omega(n)$ holds with probability $1 - \exp(-\Omega(n))$. In fact adjusting the proof in [3] slightly, the same result holds for any $m = \mathcal{O}(n)$. Therefore, we may assume in the following proof for convenience, that $m \geqslant C \cdot n$ for a sufficiently large constant $C > 0$ (we will choose $C := 6$). Alternatively, we can also reduce the case with m balls for some $m \in [n, C \cdot m]$ balls to the case with $C \cdot m$ balls, by using the fact that $F_{t_0}^{t_3}$ becomes stochastically smaller if we add more balls.

In order to establish Section 4.2, we will relate the RBB process to a simpler process, which we call the *idealized process.* In the idealized process, we also remove one ball from each non-empty bin at each round, but we allocate exactly n balls, regardless of how many bins are empty.

Formally, fix any load configuration of m balls with load vector x^{t_0}. The load vector of the idealized process is denoted by $y^t, t \geqslant t_0$ and defined as follows. For any bin $i \in [n]$, $y_i^{t_0} := x_i^{t_0}$. Further, for any $t \geqslant t_0$, let $Z_1^t, Z_2^t, \ldots, Z_n^t \in \{1, \ldots, n\}$ be n independent, uniform random samples. Then define,

$$y_i^{t+1} := y_i^t - \mathbf{1}_{y_i^t > 0} + \sum_{j=1}^{n} \mathbf{1}_{Z_j^t = i}. \tag{4.1}$$

Note that the marginal distribution of y_i^{t+1} can be expressed as

$$y_i^{t+1} = y_i^t - \mathbf{1}_{y_i^t > 0} + \mathsf{Bin}(n, 1/n).$$

Comparing this to the RBB process (see Equation (2.1)) we have the same distribution apart from that $\mathsf{Bin}(n, 1/n)$ is replaced by $\mathsf{Bin}(\kappa^t, 1/n)$. Thus we see that the *idealized process* is a bit simpler and also has the advantage that the number of balls that are added to the bins does not depend on the load configuration.

▶ **Lemma 4.4.** *For any round $t_0 \geqslant 0$ and load vector x^{t_0}, there is a coupling between the load vectors $(x^t)_{t \geqslant t_0}$ and $(y^t)_{t \geqslant t_0}$ such that for all rounds $t \geqslant t_0$ and for all bins $i \in [n]$, $x_i^t \leqslant y_i^t$.*

Based on this coupling, we also define for two rounds $t_0 \leqslant t_3$,

$$G_{t_0}^{t_3} := \sum_{t=t_0}^{t_3} \sum_{i \in [n]} \mathbf{1}_{y_i^t = 0}.$$

Note that Lemma 4.4 implies that $F_{t_0}^{t_3}$ is stochastically larger than $G_{t_0}^{t_3}$, therefore it suffices to analyze $G_{t_0}^{t_3}$ in the following.

Our first lemma proves that starting from any load configuration with m balls at time t_0, any bin $i \in [n]$ whose load is close to the average load, has a constant probability > 0 of reaching zero load after $\mathcal{O}((m/n)^2)$ rounds.

▶ **Lemma 4.5.** *Consider the idealized process with an arbitrary initial load configuration at time t_0 with $m \geqslant 6n$ balls. Let $i \in [n]$ be any bin with $y_i^{t_0} \leqslant 2 \cdot m/n$. Then,*

$$\mathbf{Pr}\left[\bigcup_{t_1 \in [t_0, t_0 + 720 \cdot \frac{m^2}{n^2}]} \{y_i^{t_1} = 0\} \;\middle|\; \mathfrak{F}^{t_0}, y_i^{t_0} \leqslant 2 \cdot \frac{m}{n} \right] \geqslant \frac{1}{4}.$$

The next lemma shows that once $y_i^t = 0$ occurs, then with constant probability bin i will have zero load in $\Omega(m/n)$ further rounds until time $\mathcal{O}((m/n)^2)$.

▶ **Lemma 4.6.** *Consider the idealized process with an arbitrary load configuration at round t_1 with $m \geqslant 6n$ balls, such that there is a bin $i \in [n]$ with $y_i^{t_1} = 0$. Then, for round $t_2 := t_1 + 24 \cdot (m/n)^2$,*

$$\mathbf{Pr}\left[\sum_{t=t_1}^{t_2} \mathbf{1}_{y_i^t = 0} \geqslant \frac{1}{6} \cdot \frac{m}{n} \;\middle|\; \mathfrak{F}^{t_1}, y_i^{t_1} = 0 \right] \geqslant \frac{1}{4}.$$

By combining Lemma 4.5 and Lemma 4.6, we derive the following lower bound on $\mathbf{E}\left[G_{t_0}^{t_3} \right]$, which by the coupling also holds for $\mathbf{E}\left[F_{t_0}^{t_3} \right]$.

▶ **Lemma 4.7.** *Consider the idealized process with m balls, where $m \geqslant 6n$. Then, for any round $t_0 \geqslant 0$ and for $t_3 := t_0 + 744 \cdot (m/n)^2$,*

$$\mathbf{E}\left[G_{t_0}^{t_3} \mid \mathfrak{F}^{t_0} \right] \geqslant \frac{1}{192} \cdot m.$$

Using the Method of Bounded Differences (Theorem A.3), we get this w.h.p.

▶ **Lemma 4.8.** *Consider the idealized process with m balls, where $m \geqslant 6n$. Then, for any round $t_0 \geqslant 0$ and for $t_3 := t_0 + 744 \cdot (m/n)^2$,*

$$\mathbf{Pr}\left[G_{t_0}^{t_3} \geqslant \frac{1}{384} \cdot m \;\middle|\; \mathfrak{F}^{t_0} \right] \geqslant 1 - e^{-\Omega(n)}.$$

Upper Bound on Convergence Time

To bound the convergence time for $m \geqslant n$, we use the Φ with $\alpha = \Theta(n/m)$ and show that in $\mathcal{O}(m^2/n)$ rounds the process reaches a configuration with $\Phi^t < \frac{48}{\alpha^2} \cdot n$. In such a step, the maximum load is $\mathcal{O}(m/n \cdot \log m)$, which becomes $\mathcal{O}(m/n \cdot \log n)$ for $n \leqslant m \leqslant \text{poly}(n)$.

We start by proving that potential drops in expectation when it is sufficiently large and there is a large fraction of empty bins.

▶ **Lemma 4.9.** *Consider the RBB process for any $m \geqslant n$, and the potential $\Phi := \Phi(\alpha)$ with $\alpha := \frac{1}{2\cdot 384\cdot 744} \cdot \frac{n}{m}$. Then for any round $t \geqslant 0$,*

$$\mathbf{E}\left[\, \Phi^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \Phi^t \cdot e^{\alpha^2 - \alpha f^t} + 6n.$$

In particular,

$$\mathbf{E}\left[\Phi^{t+1} \,\middle|\, \mathfrak{F}^t, \Phi^t > \frac{48}{\alpha^2} \cdot n \right] \leqslant \Phi^t \cdot e^{1.5\alpha^2 - \alpha f^t}.$$

We now define the event

$$\mathcal{E}^t := \left\{ \Phi^t \leqslant \frac{48}{\alpha^2} \cdot n \right\}.$$

When $\mathcal{E}^t$ holds, the potential is small enough to imply a maximum load of $\mathcal{O}(m/n \cdot \log m)$. When it is large, it drops in expectation by a multiplicative factor in any round with $f^t = \Omega(m/n)$. We now define for any $t_0 \geqslant 0$, the *adjusted exponential potential function* $\widetilde{\Phi}_{t_0}^s := \widetilde{\Phi}_{t_0}^s(\alpha)$, with $\widetilde{\Phi}_{t_0}^{t_0} := \Phi^{t_0}(\alpha)$ and for any $s > t_0$

$$\widetilde{\Phi}_{t_0}^s := \mathbf{1}_{\cap_{t \in [t_0, s)} \neg \mathcal{E}^t} \cdot \Phi^s(\alpha) \cdot \exp\left(\sum_{t=t_0}^{s-1} (\alpha f^t - 1.5\alpha^2) \right).$$

This forms a super-martingale. Using Section 4.2, we will show that in a $\Theta(m^2/n)$ interval w.h.p. the potential becomes small at least once, implying the $\mathcal{O}(m/n \cdot \log m)$ bound.

[Convergence] Consider the RBB process for any $m \geqslant n$ and the potential $\Phi := \Phi(\alpha)$ for $\alpha > 0$ as defined in Lemma 4.9. Let $c_r := 16 \cdot 384^2 \cdot 744^2$. For any round $t_0 \geqslant 0$, for $t_1 := t_0 + c_r \cdot \frac{m^2}{n}$, we have

$$\mathbf{Pr}\left[\bigcup_{t \in [t_0, t_1]} \left\{ \Phi^t \leqslant \frac{48}{\alpha^2} \cdot n \right\} \right] \geqslant 1 - e^{-\Omega(n)}.$$

In particular, this implies that for $m = \mathrm{poly}(n)$, there exists a constant $C > 0$ such that

$$\mathbf{Pr}\left[\bigcup_{t \in [t_0, t_1]} \left\{ \max_{i \in [n]} x_i^t \leqslant C \cdot \frac{m}{n} \cdot \log m \right\} \right] \geqslant 1 - e^{-\Omega(n)}.$$

Upper Bound on the Maximum Load

We will now show that, for any $n \leqslant m \leqslant \mathrm{poly}(n)$, once a configuration with $\Phi^t \leqslant \frac{48}{\alpha^2} \cdot n$ is reached, then w.h.p. the process will re-visit such a configuration in the next $\mathcal{O}(m^2/n \cdot \log n)$ rounds. The proof is quite similar to Section 4.2, but with intervals of shorter lengths. By a One-Choice argument we will deduce that the maximum load in every of the in-between rounds is $\mathcal{O}(m/n \cdot \log n)$ and so the maximum load remains small for $\mathrm{poly}(n)$ rounds.

▶ **Lemma 4.10.** *Consider the RBB process with $n \leqslant m \leqslant n^k$ for some constant $k \geqslant 1$ and the potential $\Phi := \Phi(\alpha)$ for $\alpha > 0$ as defined in Lemma 4.9. Further, let $c_s := 8k \cdot 16 \cdot 384^2 \cdot 744^2$. Then, for any round $t_0 \geqslant 0$ and for $t_1 := t_0 + c_s \cdot \frac{m^2}{n^2} \cdot \log n$, we have*

$$\mathbf{Pr}\left[\bigcup_{t \in [t_0, t_1]} \left\{ \Phi^t \leqslant \frac{48}{\alpha^2} \cdot n \right\} \,\middle|\, \mathfrak{F}^{t_0}, \Phi^{t_0} \leqslant e^{\alpha \log n} \cdot \frac{48}{\alpha^2} \cdot n \right] \geqslant 1 - n^{-7k}.$$

Finally, combining Section 4.2 and Lemma 4.10 we can derive the following upper bound on the maximum load, which holds for poly(n) rounds.

▶ **Theorem 4.11** (Stabilization). *Consider the RBB process with any $n \leqslant m \leqslant n^k$ for some constant $k \geqslant 1$. There exists a constant $C > 0$ such that, for any $t \geqslant c_r \cdot \frac{m^2}{n}$, where $c_r > 0$ is the constant defined in Section 4.2,*

$$\mathbf{Pr}\left[\bigcap_{s \in [t, t+m^2]} \left\{ \max_{i \in [n]} x_i^s \leqslant C \cdot \frac{m}{n} \cdot \log n \right\} \right] \geqslant 1 - n^{-2k}.$$

5 The Multi-Token Traversal Time

As mentioned in [3], it is natural to regard the RBB process as a multi-token traversal problem, in which each ball should visit all bins as frequently as possible. This can be seen as a "cover time" of parallel and dependent random walks, which is the first time until each ball has been allocated at least once to every bin. In [3, Corollary 1], a w.h.p. bound of $\mathcal{O}(n \log^2 n)$ on this quantity was established (it was also shown that this bound holds even in an adversarial setting, where an adversary is able to re-allocate all tokens arbitrarily every $\mathcal{O}(n)$ rounds). For the original setting without the adversary, we show:

Consider the RBB with any $m \geqslant n$. Then, with probability $1 - m^{-2}$, each of the m balls traverses all n bins within $28m \cdot \log m$ rounds. Furthermore, any fixed ball needs with probability at least $1 - o(1)$ at least $1/16 \cdot m \cdot \log n$ rounds until all n bins are traversed.

6 Experiments

We complement our analysis with some experimental results in Figure 2 and Figure 3.

In Figure 2, we the plot the maximum load vs the average number of balls for $n \in \{10^2, 10^3, 10^4\}$ and $m \in \{n, 2n, \ldots 50n\}$ after 10^6 rounds starting with the uniform distribution. The trend seems to be linear in m/n as m grows, which is in accordance with the $\Theta(m/n \cdot \log n)$ bound on the maximum load shown by our theoretical analysis in Lemma 3.3 and Theorem 4.11. In Figure 3, we plot of the fraction of empty bins vs the average number of balls for $n \in \{10^2, 10^3, 10^4\}$ and $m \in \{n, 2n, \ldots 50n\}$ averaged over 10^6 rounds, starting from the uniform load vector. The trend supports that the fraction is $\Theta(n/m)$ in the steady state, as proven in Lemma 3.2 and Section 4.2.

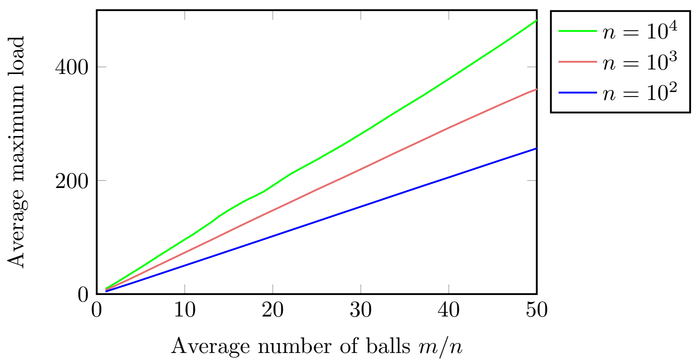

Figure 2 Maximum load vs average number of balls for $n \in \{10^2, 10^3, 10^4\}$ and $m \in \{n, 2n, \ldots 50n\}$ after 10^6 rounds, starting from the uniform load vector (averaged over 25 runs).

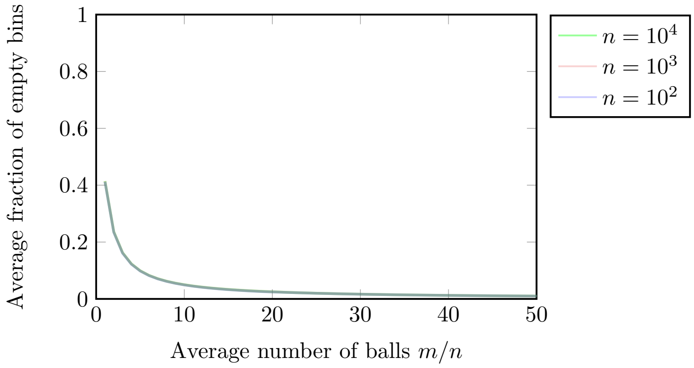

Figure 3 Fraction of empty bins vs the average load for $n \in \{10^2, 10^3, 10^4\}$ and $m \in \{n, 2n, \ldots 50n\}$ averaged over 10^6 rounds, starting from the uniform load vector (averaged over 25 runs). Note that for all values of n, the curves are very close to one another.

7 Conclusions

We revisited the RBB process and proved that for any $m \geqslant n$ that w.h.p. after $\mathcal{O}(m^2/n)$ rounds it achieves an $\mathcal{O}(m/n \cdot \log m)$ maximum load. For $n \leqslant m \leqslant \text{poly}(n)$ we show that it stabilizes in a configuration with an $\mathcal{O}(m/n \cdot \log n)$ maximum load, for at least m^2 rounds and also prove a lower bound matching up to multiplicative constants. This resolved two conjectures in [3]. We also obtained an upper bound of $\mathcal{O}(m \cdot \log m)$ on the traversal time for the balls, which was shown to be tight for any $m = \text{poly}(n)$.

There are several possible extensions, such as generalizing the stabilization result for $m = n^{\omega(1)}$, determining whether the $\mathcal{O}(m^2/n)$ convergence time is tight for $m = \omega(n)$ and determining tight bounds for the maximum load when $m < n$.

Finally, as mentioned in [3], an interesting but also challenging generalization is the RBB process on graphs. We hope that at least some of our arguments could be leveraged, for example, the insight in Section 4.2 that many bins become empty within $\mathcal{O}((m/n)^2)$ rounds might extend to graphs.

References

1 Yossi Azar, Andrei Z. Broder, Anna R. Karlin, and Eli Upfal. Balanced allocations. *SIAM Journal on Computing*, 29(1):180–200, 1999. `doi:10.1137/S0097539795288490`.

2 Jialu Bao, Marco Gaboardi, Justin Hsu, and Joseph Tassarotti. A separation logic for negative dependence. In *Proceedings of the ACM on Programming Languages (POPL)*, volume 6, New York, NY, USA, January 2022. Association for Computing Machinery. `doi:10.1145/3498719`.

3 Luca Becchetti, Andrea E. F. Clementi, Emanuele Natale, Francesco Pasquale, and Gustavo Posta. Self-stabilizing repeated balls-into-bins. *Distributed Computing*, 32(1):59–68, 2019. (earlier version in SPAA'15). `doi:10.1007/s00446-017-0320-4`.

4 Luca Becchetti, Andrea E. F. Clementi, Emanuele Natale, Francesco Pasquale, and Luca Trevisan. Stabilizing consensus with many opinions. In *Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 620–635, Philadelphia, PA, 2016. SIAM. `doi:10.1137/1.9781611974331.ch46`.

5 Petra Berenbrink, Artur Czumaj, Matthias Englert, Tom Friedetzky, and Lars Nagel. Multiple-choice balanced allocation in (almost) parallel. In *Proceedings of the 16th International Workshop on Randomization and Computation (RANDOM)*, pages 411–422, Berlin Heidelberg, 2012. Springer-Verlag. `doi:10.1007/978-3-642-32512-0_35`.

6 Petra Berenbrink, Artur Czumaj, Angelika Steger, and Berthold Vöcking. Balanced allocations: the heavily loaded case. *SIAM Journal on Computing*, 35(6):1350–1385, 2006. doi:10.1137/S009753970444435X.

7 Petra Berenbrink, Tom Friedetzky, Leslie Ann Goldberg, Paul W. Goldberg, Zengjian Hu, and Russell Martin. Distributed selfish load balancing. *SIAM Journal on Computing*, 37(4):1163–1181, 2007. doi:10.1137/060660345.

8 Petra Berenbrink, Tom Friedetzky, Peter Kling, Frederik Mallmann-Trenn, Lars Nagel, and Chris Wastell. Self-stabilizing balls and bins in batches: the power of leaky bins. *Algorithmica. An International Journal in Computer Science*, 80(12):3673–3703, 2018. doi:10.1007/s00453-018-0411-z.

9 Petra Berenbrink, Martin Hoefer, and Thomas Sauerwald. Distributed selfish load balancing on networks. In *Proceedings of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1487–1497. SIAM, Philadelphia, PA, 2011.

10 Nicoletta Cancrini and Gustavo Posta. Propagation of chaos for a balls into bins model. *Electronic Communications in Probability*, 24:Paper No. 1, 9, 2019. doi:10.1214/18-ECP204.

11 Nicoletta Cancrini and Gustavo Posta. Mixing time for the repeated balls into bins dynamics. *Electronic Communications in Probability*, 25:Paper No. 60, 14, 2020. doi:10.1214/20-ecp338.

12 Nicoletta Cancrini and Gustavo Posta. Propagation of chaos for a general balls into bins dynamics. *Electronic Journal of Probability*, 26:Paper No. 23, 20, 2021. doi:10.1214/21-EJP590.

13 Fan Chung and Linyuan Lu. Concentration inequalities and martingale inequalities: a survey. *Internet Mathematics*, 3(1):79–127, 2006. URL: http://projecteuclid.org/euclid.im/1175266369.

14 Colin Cooper. Random walks, interacting particles, dynamic networks: Randomness can be helpful. In *Proceedings of the 18th International Colloquium on Structural Information and Communication Complexity (SIROCCO)*, volume 6796 of *Lecture Notes in Computer Science*, pages 1–14, Berlin, Heidelberg, 2011. Springer. doi:doi.org/10.1007/978-3-642-22212-2_1.

15 Artur Czumaj, Chris Riley, and Christian Scheideler. Perfectly balanced allocation. In *Proceedings of the 7th International Workshop on Randomization and Computation (RANDOM)*, volume 2764 of *Lecture Notes in Computer Science*, pages 240–251. Springer, Berlin, 2003. doi:10.1007/978-3-540-45198-3_21.

16 Devdatt P. Dubhashi and Alessandro Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, Cambridge, 2009. URL: http://www.cambridge.org/gb/knowledge/isbn/item2327542/.

17 Yehuda Hassin and David Peleg. Distributed probabilistic polling and applications to proportionate agreement. *Information and Computation*, 171(2):248–268, 2001. doi:10.1006/inco.2001.3088.

18 Amos Israeli and Marc Jalfon. Token management schemes and random walks yield self-stabilizing mutual exclusion. In *Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 119–131, New York, NY, USA, 1990. ACM. doi:10.1145/93385.93409.

19 James R. Jackson. Jobshop-like queueing systems. *Management Science*, 50(12 Supplement):1796–1802, December 2004. doi:10.1287/mnsc.1040.0268.

20 Richard M. Karp, Michael Luby, and Friedhelm Meyer auf der Heide. Efficient PRAM simulation on a distributed memory machine. *Algorithmica. An International Journal in Computer Science*, 16(4-5):517–542, 1996. doi:10.1007/BF01940878.

21 Frank P. Kelly. Networks of queues. *Advances in Applied Probability*, 8(2):416–432, 1976. doi:10.2307/1425912.

22 Dimitrios Los and Thomas Sauerwald. Balanced Allocations with Incomplete Information: The Power of Two Queries. In *Proceedings of the 13th Innovations in Theoretical Computer Science Conference (ITCS)*, volume 215 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 103:1–103:23, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ITCS.2022.103.

23 Dimitrios Los and Thomas Sauerwald. Balanced allocations with the choice of noise. In *Proceedings of the 2022 ACM Symposium on Principles of Distributed Computing (PODC)*, pages 164–175, New York, NY, USA, 2022. Association for Computing Machinery. doi:10.1145/3519270.3538428.

24 Dimitrios Los and Thomas Sauerwald. Brief announcement: Tight bounds for repeated balls-into-bins. In *Proceedings of the 34th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA)*, pages 419–421, New York, NY, USA, 2022. ACM. doi:10.1145/3490148.3538561.

25 Dimitrios Los and Thomas Sauerwald. Tight bounds for repeated balls-into-bins, 2022. doi:10.48550/arXiv.2203.12400.

26 Dimitrios Los, Thomas Sauerwald, and John Sylvester. Balanced Allocations: Caching and Packing, Twinning and Thinning. In *Proceedings of the 33rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1847–1874, Alexandria, Virginia, 2022. SIAM. doi:10.1137/1.9781611977073.74.

27 Michael Mitzenmacher, Andréa W. Richa, and Ramesh Sitaraman. The power of two random choices: a survey of techniques and results. In *Handbook of randomized computing, Vol. I, II*, volume 9 of *Combinatorial Optimization*, pages 255–312. Kluwer Acad. Publ., Dordrecht, Netherlands, 2001. doi:10.1007/978-1-4615-0013-1_9.

28 Michael Mitzenmacher and Eli Upfal. *Probability and computing*. Cambridge University Press, Cambridge, second edition, 2017. Randomization and probabilistic techniques in algorithms and data analysis.

29 David Peleg and Eli Upfal. The token distribution problem. *SIAM Journal on Computing*, 18(2):229–243, 1989. doi:10.1137/0218015.

30 Yuval Peres, Kunal Talwar, and Udi Wieder. Graphical balanced allocations and the $(1+\beta)$-choice process. *Random Structures & Algorithms*, 47(4):760–775, 2015. doi:10.1002/rsa.20558.

31 Martin Raab and Angelika Steger. "Balls into bins"—a simple and tight analysis. In *Proceedings of the 2nd International Workshop on Randomization and Computation (RANDOM)*, volume 1518, pages 159–170. Springer, Barcelona, Spain, 1998. doi:10.1007/3-540-49543-6_13.

32 Udi Wieder. Hashing, load balancing and multiple choice. *Foundations and Trends in Theoretical Computer Science*, 12(3-4):275–379, 2017. doi:10.1561/0400000070.

33 Mingxi Yin, Yuli Yang, Jen-Ming Wu, and Bingli Jiao. Opportunistic bits in short-packet communications: A finite blocklength perspective. *IEEE Transactions on Communications*, 69(12):8085–8099, 2021. doi:10.1109/TCOMM.2021.3117606.

A Tools

A.1 Facts about the One-Choice Process

In this section, we prove two well-known properties of the One-Choice process. In the first lemma, we show that the quadratic potential is w.h.p. $\mathcal{O}(n)$ for n balls into n bins.

► **Lemma A.1.** *Consider the* One-Choice *process for n balls into n bins. Then, for any $t \geqslant 0$*

$$\mathbf{Pr}\left[\,\Upsilon^t \leqslant 3n\,\right] \geqslant 1 - n^{-\omega(1)}.$$

Proof. Recall that

$$\Upsilon^t := \sum_{i=1}^{n} \Upsilon_i^t = \sum_{i=1}^{n} (x_i^t)^2.$$

Since x_i^t has distribution $\mathsf{Bin}(t, 1/n)$ for any $t \geqslant 0$, $\mathbf{E}\left[\, (x_i^n)^2 \,\right] = (1-1/n)+1 = 2 - \frac{1}{n}$. Hence by linearity of expectations,

$$\mathbf{E}\left[\, \Upsilon^n \,\right] \leqslant 2n.$$

Define

$$\widetilde{\Upsilon}^n := \sum_{i=1}^n \min\{\Upsilon_i^n, \log^2 n\},$$

and note that $\Upsilon^n = \widetilde{\Upsilon}^n$ if and only if the maximum load is at most $\log n$. Then,

$$\mathbf{E}\left[\, \widetilde{\Upsilon}^n \,\right] \leqslant \mathbf{E}\left[\, \Upsilon^n \,\right] \leqslant 2n.$$

Further, $\widetilde{\Upsilon}^n$ is a function of n independent random variables (the random bin choices of the n balls), and changing one of these choices can change $\widetilde{\Upsilon}^n$ by at most $(\log n)^2 - (\log n - 1)^2 \leqslant 2\log n$. Hence by the Method of Bounded Differences (Theorem A.3),

$$\mathbf{Pr}\left[\, \widetilde{\Upsilon}^n - \mathbf{E}\left[\, \widetilde{\Upsilon}^n \,\right] \geqslant \lambda \,\right] \leqslant \exp\left(- \frac{\lambda^2}{2\sum_{i=1}^n 4(\log n)^2} \right),$$

and choosing $\lambda = n$ yields,

$$\mathbf{Pr}\left[\, \widetilde{\Upsilon}^n \geqslant 3n \,\right] \leqslant \mathbf{Pr}\left[\, \widetilde{\Upsilon}^n \geqslant \mathbf{E}\left[\, \widetilde{\Upsilon}^n \,\right] + n \,\right] \leqslant n^{-\omega(1)}.$$

Further, since the maximum load is larger than $\log n$ with probability $1 - n^{-\omega(1)}$, we have by the union bound

$$\mathbf{Pr}\left[\, \Upsilon^n \geqslant 3n \,\right] \leqslant \mathbf{Pr}\left[\left\{ \widetilde{\Upsilon}^n \geqslant 3n \right\} \cup \left\{ \max_{i\in[n]} x_i^n > \log n \right\} \right] \leqslant n^{-\omega(1)} + n^{-\omega(1)} = 2n^{-\omega(1)}. \blacktriangleleft$$

The next standard result was also used in [30, Section 4] and is based on [31]. For convenience of the reader, we give a self-contained proof, obtaining high probability bounds. [cf. [26, Lemma 10.4]] Consider the One-Choice process with $m = cn\log n$ balls, for any $c \geqslant 1/\log n$. Then, we have

$$\mathbf{Pr}\left[\max_{i\in[n]} x_i^m \geqslant \left(c + \frac{\sqrt{c}}{10} \right) \cdot \log n \right] \geqslant 1 - n^{-2}.$$

Proof. In order to use the Poisson Approximation [28, Chapter 5], let $Y_1, Y_2, \ldots, Y_n$ be n independent Poisson random variables with parameter $\lambda = \frac{m}{n} = c\log n$. Then,

$$\mathbf{Pr}\left[Y_i \geqslant \lambda + \frac{\sqrt{c}}{10} \cdot \log n \right] \geqslant \mathbf{Pr}\left[Y_i = \lambda + \frac{\sqrt{c}}{10} \cdot \log n \right] = e^{-\lambda} \cdot \frac{\lambda^{\lambda + \frac{\sqrt{c}}{10} \cdot \log n}}{(\lambda + \frac{\sqrt{c}}{10} \cdot \log n)!}.$$

Using that $z! \leqslant \sqrt{2\pi z} \left(\frac{z}{e}\right)^z e^{\frac{1}{12z}}$ for any integer $z \geqslant 1$,

$$\begin{aligned}\mathbf{Pr}\left[Y_i = \lambda + \frac{\sqrt{c}}{10} \cdot \log n \right] &\geqslant \frac{1}{4 \cdot \sqrt{2\pi\lambda}} \cdot e^{-\lambda} \cdot \left(\frac{e\lambda}{\lambda + \frac{\sqrt{c}}{10} \cdot \log n} \right)^{\lambda + \frac{\sqrt{c}}{10} \cdot \log n} \\ &\geqslant \frac{1}{4 \cdot \sqrt{2\pi\lambda}} \cdot e^{\frac{\sqrt{c}}{10} \log n} \cdot \left(1 + \frac{1}{10\sqrt{c}} \right)^{-\lambda - \frac{\sqrt{c}}{10} \cdot \log n}\end{aligned}$$

$$\geqslant \frac{1}{4\cdot\sqrt{2\pi\lambda}} \cdot e^{\frac{\sqrt{c}}{10}\log n} \cdot e^{-\frac{1}{10\sqrt{c}}\cdot(\lambda+\frac{\sqrt{c}}{10}\cdot\log n)}$$
$$\geqslant \frac{1}{4\cdot\sqrt{2\pi\lambda}} \cdot e^{\frac{\sqrt{c}}{10}\log n - \frac{1}{10\sqrt{c}}\lambda - \frac{1}{100}\log n}$$
$$\geqslant \frac{1}{4\cdot\sqrt{2\pi\lambda}} \cdot e^{-\frac{1}{100}\log n}$$

Since for any $k \geqslant 0$,

$$\frac{\mathbf{Pr}\left[\, Y_i = k+1 \,\right]}{\mathbf{Pr}\left[\, Y_i = k \,\right]} = \frac{\lambda}{k+1},$$

we conclude that

$$\begin{aligned}
\mathbf{Pr}\left[Y_i \geqslant \lambda + \frac{\sqrt{c}}{10}\cdot\log n \right] &\geqslant \sum_{k=0}^{\sqrt{\lambda}-1} \mathbf{Pr}\left[Y_i = \lambda + \frac{\sqrt{c}}{10}\cdot\log n + k \right] \\
&\geqslant \sqrt{\lambda}\cdot \mathbf{Pr}\left[Y_i = \lambda + \frac{\sqrt{c}}{10}\cdot\log n + \sqrt{\lambda} \right] \\
&\geqslant \sqrt{\lambda}\cdot \mathbf{Pr}\left[Y_i = \lambda + \frac{\sqrt{c}}{10}\cdot\log n \right] \cdot \prod_{k=1}^{\sqrt{\lambda}} \left(\frac{\lambda}{\lambda + \frac{\sqrt{c}}{10}\cdot\log n + k} \right) \\
&\geqslant \sqrt{\lambda}\cdot \frac{1}{4\cdot\sqrt{2\pi\lambda}} \cdot e^{-\frac{1}{100}\log n} \cdot \left(\frac{\lambda}{\lambda + \frac{\sqrt{c}}{10}\cdot\log n + \sqrt{\lambda}} \right)^{\sqrt{\lambda}} \\
&\geqslant \sqrt{\lambda}\cdot \frac{1}{4\cdot\sqrt{2\pi\lambda}} \cdot e^{-\frac{1}{100}\log n} \cdot \left(1 + \frac{1}{5\sqrt{c}} \right)^{-\sqrt{\lambda}} \\
&\geqslant \sqrt{\lambda}\cdot \frac{1}{4\cdot\sqrt{2\pi\lambda}} \cdot e^{-\frac{1}{100}\log n} \cdot e^{-\frac{1}{5}\sqrt{\log n}} \\
&\geqslant e^{-\frac{1}{99}\log n} = n^{-1/99},
\end{aligned}$$

where the last inequality holds for sufficiently large n. Hence,

$$\mathbf{Pr}\left[\bigcup_{i=1}^{n} \left\{ Y_i \geqslant \lambda + \frac{\sqrt{c}}{10}\cdot\log n \right\} \right] \geqslant 1 - \left(1 - n^{-1/99}\right)^n \geqslant 1 - n^{-3}.$$

Hence for $\widetilde{\mathcal{E}} := \left\{\max_{i\in[n]} Y_i \geqslant \lambda + \frac{\sqrt{c}}{10}\cdot\log n\right\}$, we have $\mathbf{Pr}\left[\neg\widetilde{\mathcal{E}}\right] \leqslant n^{-3}$. Note that $\widetilde{\mathcal{E}}$ is a monotone event under adding balls, and thus with $\mathcal{E} := \left\{\max_{i\in[n]} x_i^m \geqslant \lambda + \frac{\sqrt{c}}{10}\cdot\log n\right\}$, we have by [28, Corollary 5.11])

$$\mathbf{Pr}\left[\,\neg\mathcal{E}\,\right] \leqslant 2\cdot\mathbf{Pr}\left[\neg\widetilde{\mathcal{E}}\right] \leqslant 2\cdot n^{-3} \leqslant n^{-2}.$$ ◀

A.2 A Simple Bound for the RBB process

In this section, we show that given that the maximum load is small in the current step, then the change of the quadratic potential is w.h.p. small over the next step.

▶ **Lemma A.2.** *Consider the RBB process with $m \geqslant n$ balls and n bins. For any round $t \geqslant 0$, we have,*

$$\mathbf{Pr}\left[|\Upsilon^{t+1} - \Upsilon^t| \leqslant 2\cdot m\cdot\log n + 4n \;\middle|\; \max_{i\in[n]} x_i^t \leqslant \frac{m}{n}\cdot\log n \right] \geqslant 1 - n^{-\omega(1)}.$$

Proof. Let $k_i \in [0, n]$ be the number of balls that each bin receives at round t. For any bin $i \in [n]$ with $x_i^t > 0$,

$$|\Upsilon_i^{t+1} - \Upsilon_i^t| = |(x_i^t + k_i - 1)^2 - (x_i^t)^2| = |2 \cdot (k_i - 1) \cdot x_i^t + (k_i - 1)^2| \\ \leqslant 2 \cdot x_i^t \cdot k_i + (k_i)^2 + 1.$$

For any bin $i \in [n]$ with $x_i^t = 0$,

$$|\Upsilon_i^{t+1} - \Upsilon_i^t| = k_i^2.$$

Aggregating over all bins, we have

$$|\Upsilon^{t+1} - \Upsilon^t| \leqslant \sum_{i=1}^{n} 2 \cdot x_i^t \cdot k_i + \sum_{i=1}^{n} k_i^2 + n. \tag{A.1}$$

Using Lemma A.1 we have

$$\mathbf{Pr}\left[\sum_{i=1}^{n} k_i^2 \leqslant 3n\right] \geqslant 1 - n^{-\omega(1)}.$$

When the event $\{\sum_{i=1}^{n} k_i^2 \leqslant 3n\}$ holds, and by the condition $\{\max_{i\in[n]} x_i^t \leqslant \frac{m}{n} \cdot \log n\}$ and $m \geqslant n$, we finally conclude from Equation (A.1)

$$|\Upsilon^{t+1} - \Upsilon^t| \leqslant 2 \cdot n \cdot \frac{m}{n} \cdot \log n + 4n = 2 \cdot m \cdot \log n + 4n.$$ ◀

A.3 Concentration Inequalities

In this section, we state the Method of Bounded Differences and a concentration inequality with a bad event.

▶ **Theorem A.3** ([16, Corollary 5.2]). *Consider a function $f : \prod_{i\in[N]} \Omega_i \to \mathbb{R}$ such that it satisfies the Lipschitz condition with bounds $(c_i)_{i\in[N]}$. For independent random variables $X^1, \ldots, X^N$ with X^i taking values in Ω_i, we have that for any $\lambda > 0$*

$$\mathbf{Pr}\left[\, f(X^1, \ldots, X^N) \geqslant \mathbf{E}\left[\, f(X^1, \ldots, X^N)\,\right] + \lambda\,\right] \leqslant \exp\Big(-\frac{2 \cdot \lambda^2}{\sum_{i=1}^{N} c_i^2}\Big).$$

In order to state the concentration inequality for supermartingales conditional on a bad event not occurring, we introduce the following definitions from [13]. Consider any random variable X (in our case it will be the Z^t, the adjusted quadratic potential in Lemma 3.2) that can be evaluated by a sequence of decisions $Y^1, Y^2, \ldots, Y^N$ of finitely many outputs (the allocated balls). We can describe the process by a *decision tree* T, a complete rooted tree with depth n with vertex set $V(T)$. Each edge uv of T is associated with a probability p_{uv} depending on the decision made from u to v.

We say $f : V(T) \to \mathbb{R}$ satisfies an *admissible condition* P if $P = \{P_v\}$ holds for every vertex v. For an admissible condition P, the associated bad set $\mathcal{B}^i$ over the X_i is defined to be

$$\mathcal{B}^i = \{v \mid \text{the depth of } v \text{ is } i, \text{ and } P_u \text{ does not hold for some ancestor } u \text{ of } v\}.$$

▶ **Theorem A.4** (Theorem 8.3 in [13]). *For a filtration* $\mathbf{F}$,

$$\{\emptyset, \Omega\} = \mathfrak{F}^0 \subseteq \mathfrak{F}^1 \subseteq \ldots \subseteq \mathfrak{F}^N,$$

suppose that the random variable X^i *is* $\mathfrak{F}^i$*-measurable for* $0 \leqslant i \leqslant N$*. Let* $\mathcal{B} = \mathcal{B}^N$ *denote the bad set with the following admissible condition:*

$$\mathbf{E}\left[\, X^i \mid \mathfrak{F}^{i-1} \,\right] \leqslant X^{i-1},$$
$$|X^i - X^{i-1}| \leqslant c_i,$$

for $1 \leqslant i \leqslant N$ *and for* $c_1, \ldots, c_N \geqslant 0$*. Then, we have*

$$\mathbf{Pr}\left[\, X^N \geqslant X^0 + \lambda \,\right] \leqslant \exp\left(-\frac{\lambda^2}{2 \cdot \sum_{i=1}^N c_i^2}\right) + \mathbf{Pr}\left[\, \mathcal{B} \,\right].$$

A.4 Auxiliary Probabilistic Claim

▶ **Lemma A.5.** *Consider a sequence of random variables* $(Z_i)_{i\in\mathbb{N}}$ *such that there are* $0 < a < 1$ *and* $b > 0$ *such that every* $i \geqslant 1$,

$$\mathbf{E}\left[\, Z_i \mid Z_{i-1} \,\right] \leqslant Z_{i-1} \cdot a + b.$$

Then for every $i \geqslant 1$,

$$\mathbf{E}\left[\, Z_i \mid Z_0 \,\right] \leqslant Z_0 \cdot a^i + \frac{b}{1-a}.$$

Proof. We will prove by induction that for every $i \in \mathbb{N}$,

$$\mathbf{E}\left[\, Z_i \mid Z_0 \,\right] \leqslant Z_0 \cdot a^i + b \cdot \sum_{j=0}^{i-1} a^j.$$

For $i = 0$, $\mathbf{E}\left[\, Z_0 \mid Z_0 \,\right] \leqslant Z_0$. Assuming the induction hypothesis holds for some $i \geqslant 0$, then since $a > 0$,

$$\begin{aligned}
\mathbf{E}\left[\, Z_{i+1} \mid Z_0 \,\right] &= \mathbf{E}\left[\, \mathbf{E}\left[\, Z_{i+1} \mid Z_i \,\right] \mid Z_0 \,\right] \leqslant \mathbf{E}\left[\, Z_i \mid Z_0 \,\right] \cdot a + b \\
&\leqslant \Big(Z_0 \cdot a^i + b \cdot \sum_{j=0}^{i-1} a^j\Big) \cdot a + b \\
&= Z_0 \cdot a^{i+1} + b \cdot \sum_{j=0}^{i} a^j.
\end{aligned}$$

The claims follows using that for $a \in (0,1)$, $\sum_{j=0}^{\infty} a^j = \frac{1}{1-a}$. ◀