
This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Self-stabilizing balls and bins in batches: The power of leaky bins

PLEASE CITE THE PUBLISHED VERSION

<https://doi.org/10.1007/s00453-018-0411-z>

PUBLISHER

© Springer

VERSION

AM (Accepted Manuscript)

PUBLISHER STATEMENT

This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at:
<https://creativecommons.org/licenses/by-nc-nd/4.0/>

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Berenbrink, Petra, Tom Friedetzky, Peter Kling, Frederik Mallmann-Trenn, Lars Nagel, and Chris Wastell.
2019. "Self-stabilizing Balls and Bins in Batches: The Power of Leaky Bins". figshare.
<https://hdl.handle.net/2134/28244>.

Self-stabilizing Balls & Bins in Batches

The Power of Leaky Bins

Petra Berenbrink · Tom Friedetzky ·
Peter Kling · Frederik Mallmann-Trenn ·
Lars Nagel · Chris Wastell

Received: date / Accepted: date

Abstract A fundamental problem in distributed computing is the distribution of requests to a set of uniform servers without a centralized controller. Classically, such problems are modelled as static balls into bins processes, where m balls (tasks) are to be distributed among n bins (servers). In a seminal work, Azar et al. [4] proposed the sequential strategy GREEDY[d] for $n = m$. Each ball queries the load of d random bins and is allocated to a least loaded of them. Azar et al. showed that $d = 2$ yields an exponential improvement compared to $d = 1$. Berenbrink et al. [7] extended this to $m \gg n$, showing

A preliminary version of this paper was published in the proceedings of PODC'16

Petra Berenbrink

Universität Hamburg, 22527 Hamburg, Germany

The author did some of the research while affiliated with Simon Fraser University.

E-mail: berenbrink@informatik.uni-hamburg.de

Tom Friedetzky

Durham University, Durham DH1 3LE, U.K.

E-mail: tom.friedetzky@dur.ac.uk

Peter Kling

Universität Hamburg, 22527 Hamburg, Germany

The author did some of the research while affiliated with Simon Fraser University.

E-mail: peter.kling@uni-hamburg.de

Frederik Mallmann-Trenn

École normale supérieure, 75005 Paris, France & Simon Fraser University, Burnaby, B.C., V5A 1S6, Canada

E-mail: frederik.mallmann-trenn@ens.fr

Lars Nagel

Johannes Gutenberg-Universität Mainz

E-mail: nagell@uni-mainz.de

Chris Wastell

Durham University, Durham DH1 3LE, U.K.

E-mail: christopher.wastell@dur.ac.uk

that for $d = 2$ the maximal load difference is independent of m (in contrast to the $d = 1$ case).

We propose a new variant of an *infinite* balls-into-bins process. In each round an expected number of λn new balls arrive and are distributed (in parallel) to the bins, and each non-empty bin deletes one of its balls. This setting models a set of servers processing incoming requests, where clients can query a server’s current load but receive no information about parallel requests. We study the GREEDY[d] distribution scheme in this setting and show a strong self-stabilizing property: for *any* arrival rate $\lambda = \lambda(n) < 1$, the system load is time-invariant. Moreover, for *any* (even super-exponential) round t , the maximum system load is (w.h.p.) $O\left(\frac{1}{1-\lambda} \cdot \log \frac{n}{1-\lambda}\right)$ for $d = 1$ and $O\left(\log \frac{n}{1-\lambda}\right)$ for $d = 2$. In particular, GREEDY[2] has an exponentially smaller system load for high arrival rates.

Keywords balls-into-bins, self-stabilizing, 2-choice, positive recurrent, maximum load

1 Introduction

One of the fundamental problems in distributed computing is the distribution of requests, tasks, or data items to a set of uniform servers. In order to simplify this process and to avoid a single point of failure, it is often advisable to use a simple, randomized strategy instead of a complex, centralized controller to allocate the requests to the servers. In the most naïve strategy (*1-choice*), each client sends its request to a server chosen uniformly at random. A more elaborate scheme (*2-choice*) chooses two servers, queries their current loads, and sends the request to a least loaded of them. Both approaches are typically modelled as balls-into-bins processes [2, 4, 5, 7, 13, 20, 22], where requests are represented as balls and servers as bins. While the latter approach leads to considerably better load distributions [4, 7], it loses some of its power in parallel settings, where requests arrive in parallel and cannot take each other into account [2, 22].

We propose and study a novel infinite batch-based balls-into-bins process to model the client-server scenario. In a round, each server (bin) consumes one of its current tasks (balls). Afterward, expectedly λn tasks arrive and are allocated using a given distribution scheme. The *arrival rate* λ is allowed to be a function of n (e.g., $\lambda = 1 - 1/\text{poly}(n)$). Standard balls-into-bins results imply that, for high arrival rates, with high probability¹ (w.h.p.) in each round there is a bin that receives $\Theta(\log n / \log \log n)$ balls. Most other infinite balls-into-bins-type processes limit the total number of concurrent balls in the system by n [4, 5] and show a fast recovery. Since we do not limit the number of balls, our process can, in principle, result in an arbitrary high system load. In particular, if starting in a high-load situation (e.g., exponentially many balls), we cannot recover in a polynomial number of steps. Instead, we regard the system load as

¹ An event \mathcal{E} occurs *with high probability* (w.h.p.) if $\Pr(\mathcal{E}) = 1 - n^{-\Omega(1)}$.

a Markov chain and adapt the following notion of *self-stabilization*: The system is positive recurrent (expected return time to a typical low-load situation is finite), and taking a snapshot of the load situation at an *arbitrary* (even super-exponential large) time step yields (w.h.p.) a time-independent maximum load. Positive recurrence is a standard notion for stability and basically states that the system load is time-invariant. For irreducible, aperiodic Markov chains it implies the existence of a unique stationary distribution (cf. Section 1.2). While this alone does not guarantee a good load in the stationary distribution, together with the snapshot property we can look at an arbitrary time window of polynomial size (even if it is exponentially far away from the start) and give strong load guarantees. In particular, we give the following bounds on the load in addition to showing positive recurrence:

- 1-choice Process: The maximum load at an arbitrary time is (w.h.p.) bounded by $O\left(\frac{1}{1-\lambda} \cdot \log \frac{n}{1-\lambda}\right)$. We also provide a lower bound which is asymptotically tight for $\lambda \leq 1 - 1/\text{poly}(n)$. While this implies that already the simple 1-choice process is self-stabilizing, the load properties in a “typical” state are poor: even an arrival rate of only $\lambda = 1 - 1/n$ yields a superlinear maximum load.
- 2-choice Process: The maximum load at an arbitrary time is (w.h.p.) bounded by $O\left(\log \frac{n}{1-\lambda}\right)$. This allows to maintain an exponentially better system load compared to the 1-choice process; for any $\lambda \leq 1 - 1/\text{poly}(n)$ the maximum load remains logarithmic.

Note that the resulting processes can be seen as queuing processes.

1.1 Related Work

We will continue with an overview of related work. We start with classical results for sequential and finite balls-into-bins processes, go over to parallel settings, and give an overview of infinite and batch-based processes similar to ours. We also briefly mention some results from queuing theory (which is related but studies slightly different quality of service measures and system models).

Sequential Setting. There are many strong, well-known results for the classical, sequential balls-into-bins process. In the sequential setting, m balls are thrown one after another and allocated to n bins. For $m = n$, the maximum load of any bin is known to be (w.h.p.) $(1 + o(1)) \cdot \ln(n)/\ln \ln n$ for the 1-choice process [13, 20] and $\ln \ln(n)/\ln d + \Theta(1)$ for the d -choice process with $d \geq 2$ [4]. If $m \geq n \cdot \ln n$, the maximum load increases to $m/n + \Theta(\sqrt{m \cdot \ln(n)/n})$ [20] and $m/n + \ln \ln(n)/\ln d + \Theta(1)$ [7], respectively. In particular, note that the number of balls above the average grows with m for $d = 1$ but is independent of m for $d \geq 2$. This fundamental difference is known as the *power of two choices*. A similar (if slightly weaker) result was shown by Talwar and Wieder [24] using a quite elegant proof technique (which we also employ and generalize for our

analysis in Section 3). Czumaj and Stemmann [10] study adaptive allocation processes where the number of a ball's choices depends on the load of queried bins. The authors subsequently analyze a scenario that allows reallocations.

Berenbrink et al. [9] adapt the threshold protocol from [2] (see below) to a sequential setting and $m \geq n$ bins. Here, ball i randomly chooses bins until it sees a load smaller than $1 + i/n$. While this is a relatively strong assumption on the balls, this protocol needs only $O(m)$ choices in total (allocation time) and achieves an almost optimal maximum load of $\lceil m/n \rceil + 1$.

Parallel Setting. Several papers (e.g., [2, 22]) investigated parallel settings of multiple-choice games for the case $m = n$. Here, all m balls have to be allocated in parallel, but balls and bins might employ some (limited) communication. Adler et al. [2] consider a trade-off between the maximum load and the number of communication rounds r the balls need to decide for a target bin. Basically, bounds that are close to the classical (sequential) processes can only be achieved if r is close to the maximum load [2]. The authors also give a lower bound on the maximum load if r communication rounds are allowed, and Stemmann [22] provides a matching upper bound via a collision-based protocol.

Infinite Processes. In infinite processes, the number of balls to be thrown is not fixed. Instead, in each of infinitely many rounds, balls are thrown or reallocated and bins (possibly) delete old balls. Azar et al. [4] consider an infinite, sequential process starting with n balls arbitrarily assigned to n bins. In each round one random ball is reallocated using the d -choice process. For any $t > cn^2 \log \log n$, the maximum load at time t is (w.h.p.) $\ln \ln(n) / \ln d + O(1)$.

Adler et al. [1] consider a system where in each round $m \leq n/9$ balls are allocated. Bins have a FIFO queue, and each arriving ball is stored in the queue of two random bins. After each round, every non-empty bin deletes its frontmost ball (which automatically removes its copy from the second random bin). It is shown that the expected waiting time is constant and the maximum waiting time is (w.h.p.) $\ln \ln(n) / \ln d + O(1)$. The restriction $m \leq n/9$ is the major drawback of this process. A further study of this process, based on differential methods and experiments, was conducted in [6]. The balls' arrival times are binomially distributed with parameters n and $\lambda = m/n$. Their results indicate a stable behaviour for $\lambda \leq 0.86$. A similar model was considered by Mitzenmacher [18], who considers ball arrivals as a Poisson stream of rate λn for $\lambda < 1$. It is shown that the 2-choice process reduces the waiting time exponentially compared to the 1-choice process.

Czumaj [11] presents a framework to study the recovery time of discrete-time dynamic allocation processes. In each round one of n balls is reallocated using the d -choice process. Two models are considered: in the first, the ball to be reallocated is chosen by taking a ball from a random bin. In the second, the ball to be reallocated is chosen by selecting a random ball. From an arbitrary initial assignment, the system is shown to recover to the maximum load from [4] within $O(n^2 \ln n)$ rounds in the former and $O(n \ln n)$ rounds in the latter case. Becchetti et al. [5] consider a similar (but parallel) process. In each round one

ball is chosen from every non-empty bin and reallocated to a randomly chosen bin (one choice per ball). The authors show that (w.h.p.) starting from an arbitrary configuration, it takes $O(n)$ rounds to reach a configuration with maximum load $O(\log n)$. Moreover, if the process starts in a configuration with maximum load $O(\log n)$, then the maximum load stays in $O(\log n)$ for $\text{poly}(n)$ rounds. An interesting connection to our work is that the analysis of [5] is based on an auxiliary TETRIS-process. This process can be seen a special version of our 1-choice process and is defined as follows: starting from a state with at least $n/4$ empty bins, in each round every non-empty bin deletes one ball. Subsequently, exactly $(3/4)n$ new balls are allocated to the bins (one choice per ball).

Batch-Processes. Batch-based processes allocate m balls to n bins in batches of (usually) n balls each, where each batch is allocated in parallel. They lie between (pure) parallel and sequential processes. For $m = \tau \cdot n$, Stemann [22] investigates a scenario with n players each having m/n balls. To allocate a ball, every player independently chooses two bins and allocates copies of the ball to both of them. Every bin has two queues (one for first copies, one for second copies) and processes one ball from each queue per round. When a ball is processed, its copy is removed from the system and the player is allowed to initiate the allocation of the next ball. If $\tau = \ln n$, all balls are processed in $O(\ln n)$ rounds and the waiting time is (w.h.p.) $O(\ln \ln n)$. Berenbrink et al. [8] study the d -choice process in a scenario where m balls are allocated to n bins in batches of size n each. The authors show that the load of every bin is (w.h.p.) $m/n \pm O(\log n)$. As noted in Lemma 3, our analysis can be used to derive the same result by easier means.

Queuing Processes. Batch arrival processes have also been considered in the context of queuing systems. A key motivation for such models stems from the asynchronous transfer mode (ATM) in telecommunication systems. Tasks arrive in batches, are stored in a FIFO queue and served by a fixed number of servers which remove the tasks from the queue and process them. Several papers [3, 15, 16, 21] consider scenarios where the number of arriving tasks is determined by a finite state Markov chain. Results study steady state properties of the system to determine properties of interest (e.g., waiting times or queue lengths). Sohraby and Zhang [21] use spectral techniques to study a multi-server scenario with an infinite queue. Alfa [3] considers a discrete-time process for n identical servers and tasks with constant service time $s \geq 1$. To ensure a stable system, the arrival rate λ is assumed to be at most n/s and tasks are assigned cyclical, allowing to study an arbitrary server (instead of the complete system). Kamal [15] and Kim et al. [16] study a system with a finite capacity. The tasks which arrive when the buffer is full are lost. The authors study the steady state probability and give empirical results to show the decay of waiting times as n increases.

1.2 Model & Preliminaries

We model our load balancing problem as an infinite, parallel balls-into-bins process. Time is divided into discrete, synchronous rounds. There are n bins and n generators, and the initial system is assumed to be empty. At the start of each round, every non-empty bin deletes one ball. Afterward, every generator generates a ball with a probability of $\lambda = \lambda(n) \in [0, 1]$ (the *arrival rate*). This generation scheme allows us to consider arrival rates that are arbitrarily close to one (like $1 - 1/\text{poly}(n)$). Generated balls are distributed in the system using a distribution process. In this paper we analyze two specific distribution processes:

- The 1-choice process GREEDY[1] assigns every ball to a random bin.
- The 2-choice process GREEDY[2] assigns every ball to a least loaded among two randomly chosen bins.

Notation. The random variable $X_i(t)$ denotes the load (number of balls) of the i -th fullest bin at the end of round t . Thus, the load situation (configuration) after round t can be described by the load vector $\mathbf{X}(t) = (X_i(t))_{i \in [n]} \in \mathbb{N}^n$. We define $\varnothing(t) := \frac{1}{n} \sum_{i=1}^n X_i(t)$ as the average load at the end of round t . The value $\nu(t)$ denotes the fraction of non-empty bins after round t and $\eta(t) := 1 - \nu(t)$ the fraction of empty bins after round t . It will be useful to define $1_i(t) := \min(1, X_i(t))$ and $\eta_i(t) := 1_i(t) - \nu(t)$ (which equals $\eta(t)$ if i is a non-empty bin and $-\nu(t)$ otherwise).

Markov Chain Preliminaries. The Markov chain on the load vectors has the Markov property since $\mathbf{X}(t)$ depends only on $\mathbf{X}(t-1)$ and the random choices during round t . We refer to this Markov chain as \mathbf{X} . Note that \mathbf{X} is time-homogeneous (transition probabilities are time-independent), irreducible (every state is reachable from every other state²), and aperiodic (path lengths have no period; in fact, our chain is lazy). Recall that such a Markov chain is positive recurrent (or ergodic) if the probability to return to the start state is 1 and the expected return time is finite. In particular, this implies the existence of a unique stationary distribution. Positive recurrence is a standard formalization of the intuitive concept of stability. See [17] for an excellent introduction into Markov chains and the involved terminology.

2 The 1-Choice Process

We present two main results for the 1-choice process: Theorem 1 states the stability of the system under the 1-choice process for an arbitrary λ , using the standard notion of positive recurrence (cf. Section 1). In particular, this implies the existence of a stationary distribution for the 1-choice process. Theorem 2 strengthens this by giving a high probability bound on the maximum load

² The state space includes all vectors with non-increasing entries over \mathbb{N}^n .

for an *arbitrary* round $t \in \mathbb{N}$. Together, both results imply that the 1-choice process is self-stabilizing, i.e., the system is positive recurrent, and taking a snapshot of the load situation at an arbitrary time step yields (w.h.p.) a time-independent maximum load.

Theorem 1 (Stability) *Let $\lambda = \lambda(n) < 1$. The Markov chain \mathbf{X} of the 1-choice process is positive recurrent.*

Theorem 2 (Maximum Load) *Let $\lambda = \lambda(n) < 1$. Fix an arbitrary round t of the 1-choice process. The maximum load of all bins is (w.h.p.) bounded by $O\left(\frac{1}{1-\lambda} \cdot \log \frac{n}{1-\lambda}\right)$.*

Note that for high arrival rates of the form $\lambda(n) = 1 - \varepsilon(n)$, the bound given in Theorem 2 is inversely proportional to $\varepsilon(n)$. For example, for $\varepsilon(n) = 1/n$ the maximal load is $O(n \log n)$. Theorem 3 shows that this dependence is unavoidable: the bound given in Theorem 2 is tight for large values of λ . In Section 3, we will see that the 2-choice process features an exponentially better behaviour for large λ .

Theorem 3 *Let $\lambda = \lambda(n) \geq 0.5$ and consider step $t := 9\lambda \log(n)/(64(1-\lambda)^2)$. With probability $1 - o(1)$ there is a bin i in step t with load $\Omega\left(\frac{1}{1-\lambda} \cdot \log n\right)$.*

The proofs of these results can be found in the following subsections. We first prove a bound on the maximum load (Theorem 2), afterward we prove stability of the system (Theorem 1), and finally we prove the lower bound (Theorem 3).

2.1 Maximum Load

Proof (Proof of Theorem 2 (Maximum Load)) In the following, we prove Theorem 2 using a (slightly simplified) drift theorem from Hajek [14] (cf. Theorem 9 in Appendix A). Remember that, as mentioned in Section 1.2, our process is a Markov chain, such that we need to condition only on the previous state (instead of the full filtration from Theorem 9). Our goal is to bound the load of a fixed bin i at time t using Theorem 9 and, subsequently, to use this with a union bound to bound the maximum load over all bins. To apply Theorem 9, we have to prove that the maximum load difference of bin i between two rounds is exponentially bounded (Majorization) and that, given a load high enough, the system tends to lose load (Negative Bias). We start with the majorization. The load difference $|X_i(t+1) - X_i(t)|$ is bounded by $\max(1, B_i(t)) \leq 1 + B_i(t)$, where $B_i(t)$ is the number of tokens resource i receives during round $t+1$. In particular, we have $(|X_i(t+1) - X_i(t)| \mid \mathbf{X}(t)) \prec 1 + B_i(t)$. Note that $B_i(t)$ is binomially distributed with parameters n and λ/n (each of the n balls has probability of $\lambda \cdot 1/n$ to end up in i). Using standard inequalities we bound

$$\Pr(B_i(t) = k) \leq \binom{n}{k} \cdot \left(\frac{\lambda}{n}\right)^k \leq \left(\frac{e \cdot n}{k}\right)^k \cdot \left(\frac{1}{n}\right)^k = \frac{e^k}{k^k} \quad (1)$$

and calculate

$$\begin{aligned} \mathbb{E}\left[e^{B_i(t)+1}\right] &= e \cdot \sum_{k=0}^n e^k \cdot \frac{e^k}{k^k} \leq e \cdot \sum_{k=0}^{\lceil e^3-1 \rceil} \frac{e^{2k}}{k^k} + e \cdot \sum_{k=e^3}^{\infty} \frac{e^{2k}}{k^k} \\ &\leq \Theta(1) + \sum_{k=1}^{\infty} e^{-k} = \Theta(1). \end{aligned} \quad (2)$$

This shows that the Majorization condition from Theorem 9 holds (with $\lambda' = 1$ and $D = \Theta(1)$). To see that the Negative Bias condition is also given, note that if bin i has non-zero load, it is guaranteed to delete one ball and receives in expectation $n \cdot \lambda/n = \lambda$ balls. We get $\mathbb{E}[X_i(t+1) - X_i(t) \mid X_i(t) > 0] \leq \lambda - 1 < 0$, establishing the Negative Bias condition (with $\varepsilon_0 = 1 - \lambda$). We finally can apply Theorem 9 with $\eta := \min(1, (1 - \lambda)/(2D), 1/(2 - 2\lambda)) = (1 - \lambda)/(2D)$ and get for $b \geq 1$

$$\begin{aligned} \Pr(X_i(t) \geq b) &\leq e^{-b \cdot \eta} + \frac{2D}{\eta \cdot (1 - \lambda)} \cdot e^{\eta \cdot (1-b)} \leq \frac{2 \cdot (2D)^2}{(1 - \lambda)^2} \cdot e^{\frac{(1-\lambda) \cdot (1-b)}{2D}} \\ &\leq \frac{2 \cdot (2D)^2}{(1 - \lambda)^2} \cdot e^{\frac{(1-\lambda) \cdot (1-b)}{(4D)^2}} \leq \frac{(4D)^2}{(1 - \lambda)^2} \cdot e^{\frac{-b \cdot (1-\lambda)}{(4D)^2}} \\ &\leq \frac{c}{(1 - \lambda)^2} \cdot e^{-\frac{b \cdot (1-\lambda)}{c}} \end{aligned} \quad (3)$$

where $c \geq (4D)^2$ denotes a constant. Applying a union bound to all n bins and choosing $b := \frac{c}{1-\lambda} \cdot \ln\left(\frac{c \cdot n^{h+1}}{(1-\lambda)^2}\right)$, where $h > 0$ is a constant, yields that $\Pr(\max_{i \in [n]} X_i(t) \geq b) \leq n^{-h}$. The theorem's statement now follows from

$$\begin{aligned} b &= \frac{c}{1 - \lambda} \cdot \ln\left(\frac{c \cdot n^{h+1}}{(1 - \lambda)^2}\right) \leq \frac{c \cdot (h + 1) + 1}{1 - \lambda} \cdot \ln\left(\frac{n}{1 - \lambda}\right) \\ &= O\left(\frac{1}{1 - \lambda} \cdot \ln\left(\frac{n}{1 - \lambda}\right)\right). \end{aligned} \quad (4)$$

□

2.2 Stability

In the following, we provide an auxiliary lemma that will prove useful for deriving the stability of the 1-choice process.

Lemma 1 *Let $\lambda = \lambda(n) < 1$. Fix an arbitrary round t of the 1-choice process and a bin i . There is a constant $c > 0$ such that the expected load of bin i is bounded by $\frac{6c}{1-\lambda} \cdot \ln\left(\frac{e \cdot c}{1-\lambda}\right)$.*

Proof Let $\gamma := \frac{c}{1-\lambda} \cdot \ln\left(\frac{e \cdot c}{(1-\lambda)^2}\right)$ where c is the constant from the proof of Theorem 2. Considering time windows of γ rounds each and using Equation (3), we calculate

$$\begin{aligned}
& \mathbb{E}[X_i(t)] \\
& \leq \sum_{b=1}^{\gamma} b \cdot \Pr(X_i(t) = b) + \sum_{k=1}^{\infty} \sum_{b=k \cdot \gamma}^{(k+1)\gamma} b \cdot \Pr(X_i(t) = b) \\
& \leq \gamma + \sum_{k=1}^{\infty} (k+1) \cdot \gamma \cdot \Pr(X_i(t) \geq k \cdot \gamma) \leq \gamma + \sum_{k=1}^{\infty} (k+1) \cdot \gamma \cdot e^{-k} \\
& \leq 3\gamma \leq \frac{6c}{1-\lambda} \cdot \ln\left(\frac{e \cdot c}{1-\lambda}\right).
\end{aligned} \tag{5}$$

This finishes the proof. \square

Proof (Proof of Theorem 1 (Stability)) We prove Theorem 1 using a result from Fayolle et al. [12] (cf. Theorem 8 in Appendix A). Note that \mathbf{X} is a time-homogeneous irreducible Markov chain with a countable state space. For a configuration \mathbf{x} , we define the auxiliary potential $\Psi(\mathbf{x}) := \sum_{i=1}^n x_i$ as the total system load of configuration \mathbf{x} . Consider the (finite) set $C := \{\mathbf{x} \mid \Psi(\mathbf{x}) \leq n^4/(1-\lambda)^2\}$ of all configurations with not too much load. To prove positive recurrence, it remains to show that Condition 1 (expected potential drop if not in a high-load configuration) and Condition 2 (finite potential) of Theorem 8 hold. In the following, let $\Delta := \frac{n^3}{(1-\lambda)^2}$.

Let us start with Condition 1. So fix a round t and let $\mathbf{x} = \mathbf{X}(t) \notin C$. By definition of C , we have $\Psi(\mathbf{x}) > n^4/(1-\lambda)^3$, such that there is at least one bin i with load $x_i \geq \Psi(\mathbf{x})/n > n^3/(1-\lambda)^2$. In particular, note that $x_i \geq \Delta$, such that during each of the next Δ rounds exactly one ball is deleted. On the other hand, bin i receives in expectation $\Delta \cdot \lambda n \cdot \frac{1}{n} = \lambda \Delta$ balls during the next Δ rounds. We get $\mathbb{E}[X_i(t+\Delta) - x_i \mid \mathbf{X}(t) = \mathbf{x}] = \lambda \Delta - \Delta = -(1-\lambda) \cdot \Delta$. For any bin $j \neq i$, we assume pessimistically that no ball is deleted. Note that the expected load increase of each of these bins can be majorized by the load increase in an empty system running for Δ rounds. Thus, we can use Lemma 1 to bound the expected load increase in each of these bins by $\frac{6c}{1-\lambda} \cdot \ln\left(\frac{e \cdot c}{1-\lambda}\right) \leq \frac{6e \cdot c^2}{(1-\lambda)^2} \leq \Delta/n^2$. We get

$$\begin{aligned}
\mathbb{E}[\Psi(\mathbf{X}(t+\Delta)) \mid \mathbf{X}(t) = \mathbf{x}] & \leq -(1-\lambda) \cdot \Delta + (n-1) \cdot \frac{\Delta}{n^2} \\
& < -\Delta \cdot \left(1 - \lambda - \frac{1}{n}\right) \leq -\Delta \cdot \frac{1-\lambda}{2}.
\end{aligned} \tag{6}$$

This proves Condition 1 of Theorem 8. For Condition 2, assume $\mathbf{x} = \mathbf{X}(t) \in C$. We bound the system load after Δ rounds trivially by

$$\mathbb{E}[\Psi(\mathbf{X}(t+\Delta)) \mid \mathbf{X}(t) = \mathbf{x}] \leq \Psi(\mathbf{x}) + \Delta \cdot n \leq \frac{n^4}{(1-\lambda)^2} + \Delta \cdot n < \infty \tag{7}$$

(note that the finiteness in Theorem 8 is with respect to time, not n). This finishes the proof. \square

2.3 Lower Bound on Maximum Load

Proof (Proof of Theorem 3 (Lower Bound)) To show this result, we will use a bound by Raab and Steger [20, Theorem 1] which lower-bounds the maximum number of balls a bin receives when m balls are allocated into n bins. The idea of the proof is as follows. We assume that we start at an empty system and apply [20, Theorem 1] to $m = \lambda tn$ many balls. The theorem says that one of the bins is likely to get more than λt many balls, which allows us to show that the load of this bin is large, even if the bin was able to delete a ball during each of the t observed time steps.

Let $m(t')$ be the number of balls allocated during the first t' steps, and let $b_{\max}(t')$ be the maximum number of balls allocated to any bin. Set $t = 9\lambda \log(n)/(64(1-\lambda)^2)$, assume $\lambda > 0.5$ and assume $\lambda nt \leq n \cdot (\log n)^c$ for a constant c . Since the expected number of balls is $\lambda nt \geq n \log n$, we can use Chernoff bounds to show that (w.h.p.) at least $(1-\epsilon) \cdot m(t)$ balls are generated for small ϵ . Then

$$\mathbb{E}[m(t)] = \frac{9\lambda \log n}{64(1-\lambda)^2} \cdot \lambda n. \quad (8)$$

Using Chernoff's inequality we can show that w.h.p. $m(t) \geq (1-\epsilon) \cdot \frac{9\lambda \log n}{64(1-\lambda)^2} \cdot \lambda n$ for an arbitrary small constant ϵ . By Theorem 10 (Case 3) with $\alpha = \sqrt{8/9}$ we get (w.h.p.)

$$b_{\max}(t) \geq (1-\epsilon) \cdot \frac{9\lambda \log n}{64(1-\lambda)^2} \cdot \lambda + \sqrt{(1-\epsilon) \cdot \frac{16 \cdot 9\lambda (\log n)^2}{9 \cdot 64(1-\lambda)^2}} \cdot \lambda. \quad (9)$$

Let $X_{\max}(t)$ denote the load of the bin of maximum load. We derive (using $\epsilon' := 1 - \lambda + \lambda\epsilon > 0$)

$$\begin{aligned} X_{\max}(t) &\geq (1-\epsilon) \cdot \frac{9}{64} \frac{\lambda^2 \log n}{(1-\lambda)^2} + \sqrt{(1-\epsilon) \cdot \frac{16 \cdot 9\lambda^2 \cdot (\log n)^2}{9 \cdot 64(1-\lambda)^2}} - \frac{9\lambda \log n}{64(1-\lambda)^2} \\ &= \lambda \cdot (1-\epsilon) \cdot \frac{9}{64} \frac{\lambda \log n}{(1-\lambda)^2} + \sqrt{\frac{(1-\epsilon)}{4}} \cdot \frac{\lambda \log n}{(1-\lambda)} - \frac{9\lambda \log n}{64(1-\lambda)^2} \\ &= \sqrt{\frac{(1-\epsilon)}{4}} \cdot \frac{\lambda \log n}{(1-\lambda)} + (\lambda \cdot (1-\epsilon) - 1) \cdot \frac{9\lambda \log n}{64(1-\lambda)^2} \\ &= \sqrt{\frac{(1-\epsilon)}{4}} \cdot \frac{\lambda \log n}{(1-\lambda)} - \epsilon' \cdot \frac{9\lambda \log n}{64(1-\lambda)^2} \\ &\geq \sqrt{\frac{(1-\epsilon)}{4}} \cdot \frac{\lambda \log n}{(1-\lambda)} - \epsilon' \cdot \frac{9}{64} \frac{\lambda \log n}{(1-\lambda)} \\ &= \Omega\left(\frac{\lambda \log n}{1-\lambda}\right). \end{aligned}$$

\square

3 The 2-Choice Process

We continue with the study of the 2-choice process. Here, new balls are distributed according to GREEDY[2] (cf. description in Section 1.2). Our main results are the following theorems, which are equivalents to the corresponding theorems for the 1-choice process.

Theorem 4 (Stability) *Let $\lambda = \lambda(n) \in [1/4, 1)$. The Markov chain \mathbf{X} of the 2-choice process is positive recurrent.*

Theorem 5 (Maximum Load) *Let $\lambda = \lambda(n) \in [1/4, 1)$. Fix an arbitrary round t of the 2-choice process. The maximum load of all bins is (w.h.p.) bounded by $O(\log \frac{n}{1-\lambda})$.*

Note that Theorem 5 implies a much better behaved system than we saw in Theorem 2 for the 1-choice process. In particular, it allows for an exponentially higher arrival rate: for $\lambda(n) = 1 - 1/\text{poly}(n)$ the 2-choice process maintains a maximal load of $O(\log n)$. In contrast, for the same arrival rate the 1-choice process results in a system with maximal load $\Omega(\text{poly}(n))$.

Our analysis of the 2-choice process relies to a large part on a good bound on the *smoothness* (the maximum load difference between any two bins). This is stated in the following lemma. This result is of independent interest, showing that even if the arrival rate is $1 - e^{-n}$, where we get a polynomial system load, the maximum load difference is still logarithmic.

Lemma 2 (Smoothness) *Let $\lambda = \lambda(n) \in [1/4, 1]$. Fix an arbitrary round t of the 2-choice process. The load difference of all bins is (w.h.p.) bounded by $O(\ln n)$.*

Analysis Overview. To prove these results, we combine three different potential functions: For a configuration \mathbf{x} with average load \varnothing and for a suitable constant α (to be fixed later), we define

$$\begin{aligned} \Phi(\mathbf{x}) &:= \sum_{i \in [n]} e^{\alpha \cdot (x_i - \varnothing)} + \sum_{i \in [n]} e^{\alpha \cdot (\varnothing - x_i)}, & \Psi(\mathbf{x}) &:= \sum_{i \in [n]} x_i, \quad \text{and} \\ \Gamma(\mathbf{x}) &:= \Phi(\mathbf{x}) + \frac{n}{1-\lambda} \cdot \Psi(\mathbf{x}). \end{aligned} \tag{10}$$

The potential Φ measures the *smoothness* (the maximum load difference to the average) of a configuration and is used to prove Lemma 2 (Section 3.1). The proof is based on the observation that whenever the load of a bin is far from the average load, it decreases in expectation. The potential Ψ measures the *total load* of a configuration and is used, in combination with our results on the smoothness, to prove Theorem 5 (Section 3.2). The potential Γ entangles the smoothness and total load, allowing us to prove Theorem 4 (Section 3.3). The proof is based on the fact that whenever Γ is large (i.e., the configuration is not smooth or it has a huge total load), it decreases in expectation.

Before we continue with our analysis, let us make a simple but useful observation concerning the smoothness: For any configuration \mathbf{x} and value

$b \geq 0$, the inequality $\Phi(\mathbf{x}) \leq e^{\alpha \cdot b}$ implies (by definition of Φ) $\max_i |x_i - \emptyset| \leq b$. That is, the load difference of any bin to the average is at most b and, thus, the load difference between any two bins is at most $2b$. We capture this in the following observation.

Observation 6 *Let $b \geq 0$ and consider a configuration \mathbf{x} with average load \emptyset . If $\Phi(\mathbf{x}) \leq e^{\alpha \cdot b}$, then $|x_i - \emptyset| \leq b$ for all $i \in [n]$. In particular, $\max_i(x_i) - \min_i(x_i) \leq 2b$.*

3.1 Bounding the Smoothness

The goal of this section is to prove Lemma 2. To do so, we show the following bound on the expected smoothness (potential Φ) at an arbitrary time t :

Lemma 3 *Let $\lambda \in [1/4, 1]$. Fix an arbitrary round t of the 2-choice process. There is a constant $\varepsilon > 0$ such that³*

$$\mathbb{E}[\Phi(\mathbf{X}(t))] \leq \frac{n}{\varepsilon}. \quad (11)$$

Note that Lemma 3 together with Observation 6 immediately implies Lemma 2 by a simple application of Markov's inequality to bound the probability that $\Phi(\mathbf{X}(t)) \geq n^2/\varepsilon$.

Our proof of Lemma 3 follows the lines of [19, 24], who used the same potential function to analyze variants of the sequential d -choice process without deletions. While the basic idea of showing a relative drop when the potential is high combined with a bounded absolute increase in the general case is the same, our analysis turns out much more involved. In particular, not only do we have to deal with deletions and throwing balls in batches but the size of each batch is also a random variable. Once Lemma 3 is proven, Lemma 2 emerges by combining Observation 6, Lemma 3, and Markov's inequality as follows:

$$\Pr\left(\max_i X_i(t) - \min_i X_i(t) \geq \frac{4}{\alpha} \cdot \ln\left(\frac{n}{\varepsilon}\right)\right) \leq \Pr\left(\Phi(\mathbf{X}(t)) \geq \frac{n^2}{\varepsilon^2}\right) \leq \frac{\varepsilon}{n}. \quad (12)$$

It remains to prove Lemma 3. Remember the definition of $\Phi(\mathbf{x})$ from Equation (10). We split the potential in two parts

$$\Phi(\mathbf{x}) := \Phi_+(\mathbf{x}) + \Phi_-(\mathbf{x}).$$

Here, $\Phi_+(\mathbf{x}) := \sum_i e^{\alpha \cdot (x_i - \emptyset)}$ denotes the *upper potential* of \mathbf{x} and $\Phi_-(\mathbf{x}) := \sum_i e^{\alpha \cdot (\emptyset - x_i)}$ denotes the *lower potential* of \mathbf{x} . For a fixed bin i , we use $\Phi_{i,+}(\mathbf{x}) := e^{\alpha \cdot (x_i - \emptyset)}$ and $\Phi_{i,-}(\mathbf{x}) := e^{\alpha \cdot (\emptyset - x_i)}$ to denote i 's contribution to the upper and

³ For Φ , the condition $\lambda \geq 1/4$ can be substituted with $\lambda = \Omega(1)$ and only minor changes in the analysis. Moreover, the analysis can be easily adapted for a process that (deterministically) throws $\lambda \cdot n$ balls in each round, even for $\lambda > 1$ as long as it is a constant. Finally, one can easily adapt the analysis to cover the process without deletions by setting $\eta_i(t) = 0$ (see Observation 7). Using Markov's inequality, this yields the same result as [8] using a simpler analysis.

lower potential, respectively. When we consider the effect of a fixed round $t+1$, we will sometimes omit the time parameter and use prime notation to denote the value of a parameter at the end of round $t+1$. For example, we write X_i and X'_i for the load of bin i at the beginning and at the end of round $t+1$, respectively.

We start with two simple but useful identities regarding the potential drop $\Delta_{i,+}(t+1)$ (and $\Delta_{i,-}(t+1)$) due to a fixed bin i during round $t+1$.

Observation 7 *Fix a bin i , let K denote the number of balls that are placed during round $t+1$ and let $k \leq K$ be the number of these balls that fall into bin i . Then*

1. $\Delta_{i,+}(t+1) = \Phi_{i,+}(\mathbf{X}(t)) \cdot (e^{\alpha \cdot (k - \eta_i(t) - K/n)} - 1)$ and
2. $\Delta_{i,-}(t+1) = \Phi_{i,-}(\mathbf{X}(t)) \cdot (e^{-\alpha \cdot (k - \eta_i(t) - K/n)} - 1)$.

Proof Remember that 1_i is an indicator value which equals 1 if and only if the i -th bin is non-empty in configuration \mathbf{X} . Bin i loses exactly 1_i balls and receives exactly k balls, such that $X'_i - X_i = -1_i + k$. Similarly, we have $\varnothing' - \varnothing = -\nu + K/n$ for the change of the average load. With the identity $\eta_i = 1_i - \nu$ (see Section 1.2), this yields

$$\begin{aligned} \Delta_{i,+}(t) &= e^{\alpha \cdot (X'_i - \varnothing')} - e^{\alpha \cdot (X_i - \varnothing)} \\ &= e^{\alpha \cdot (X_i - \varnothing)} \cdot \left(e^{\alpha \cdot (-1_i + k + \nu - K/n)} - 1 \right) = \Phi_{i,+} \cdot \left(e^{\alpha \cdot (k - \eta_i - K/n)} - 1 \right), \end{aligned} \quad (13)$$

proving the first statement. The second statement follows similarly. \square

We now derive the main technical lemma that states general bounds on the expected upper and lower potential change during a single round. This will be used to derive bounds on the potential change in different situations. For this, let $p_i := \binom{i}{n}^2 - \binom{i-1}{n}^2 = \frac{2i-1}{n^2}$ (the probability that a ball thrown with GREEDY[2] falls into the i -th fullest bin). We also define $\hat{\alpha} := e^\alpha - 1$ and $\check{\alpha} := 1 - e^{-\alpha}$. Note that $\hat{\alpha} \in (\alpha, \alpha + \alpha^2)$ and $\check{\alpha} \in (\alpha - \alpha^2, \alpha)$ for $\alpha \in (0, 1.7)$. This follows easily from the Taylor approximation $e^x \leq 1 + x + x^2$, which holds for any $x \in (-\infty, 1.7]$ (we will use this approximation several times in the analysis). Finally, let $\hat{\delta}_i := \lambda n \cdot (1/n \cdot \check{1} - p_i \cdot \hat{\alpha}/\alpha)$ and $\check{\delta}_i := \lambda n \cdot (1/n \cdot \hat{1} - p_i \cdot \check{\alpha}/\alpha)$, where $\check{1} := 1 - \alpha/n < 1 < \hat{1} := 1 + \alpha/n$. These $\hat{\delta}_i$ and $\check{\delta}_i$ values can be thought of as upper/lower bounds on the expected difference in the number of balls that fall into bin i under the 1-choice and 2-choice process, respectively (note that $\hat{1}$, $\check{1}$, $\hat{\alpha}/\alpha$, and $\check{\alpha}/\alpha$ are all close to 1).

Lemma 4 *Consider a bin i after round t and a constant $\alpha \leq 1$.*

1. *For the expected change of i 's upper potential during round $t+1$ we have*

$$\frac{\mathbb{E}[\Delta_{i,+}(t+1) \mid \mathbf{X}(t)]}{\Phi_{i,+}(\mathbf{X}(t))} \leq -\alpha \cdot (\eta_i + \hat{\delta}_i) + \alpha^2 \cdot (\eta_i + \hat{\delta}_i)^2. \quad (14)$$

2. For the expected change of i 's lower potential during round $t + 1$ we have

$$\frac{\mathbb{E}[\Delta_{i,-}(t+1) \mid \mathbf{X}(t)]}{\Phi_{i,-}(\mathbf{X}(t))} \leq \alpha \cdot (\eta_i + \check{\delta}_i) + \alpha^2 \cdot (\eta_i + \check{\delta}_i)^2. \quad (15)$$

Proof For the first statement, we use Observation 7 to calculate $\mathbb{E}[\Delta_{i,+}(t) \mid \mathbf{X}]/\Phi_{i,+}$

$$\begin{aligned} &= \sum_{K=0}^n \sum_{k=0}^K \binom{n}{K} \binom{K}{k} (p_i \lambda)^k \cdot ((1-p_i)\lambda)^{K-k} \cdot (1-\lambda)^{n-K} \cdot \left(e^{\alpha \cdot (k - \eta_i - K/n)} - 1 \right) \\ &= \sum_{K=0}^n \binom{n}{K} (1-\lambda)^{n-K} \lambda^K \sum_{k=0}^K \binom{K}{k} \cdot p_i^k \cdot (1-p_i)^{K-k} \cdot \left(e^{\alpha \cdot (k - \eta_i - K/n)} - 1 \right) \\ &= \sum_{K=0}^n \binom{n}{K} (1-\lambda)^{n-K} \lambda^K \cdot \left(e^{-\alpha(\eta_i + K/n)} \sum_{k=0}^K \binom{K}{k} (e^{\alpha \cdot p_i})^k (1-p_i)^{K-k} - 1 \right) \\ &= \sum_{K=0}^n \binom{n}{K} (1-\lambda)^{n-K} \lambda^K \cdot \left(e^{-\alpha(\eta_i + K/n)} \cdot (1 + \hat{\alpha} \cdot p_i)^K - 1 \right), \end{aligned}$$

where we first apply the law of total expectation together with Observation 7 and, afterward, twice the binomial theorem. Continuing the calculation using the aforementioned Taylor approximation $e^x \leq 1 + x + x^2$ (which holds for any $x \in (-\infty, 1.7]$), and the definition of $\hat{\delta}_i$ yields

$$\begin{aligned} &= e^{-\alpha\eta_i} \cdot \left(1 - \lambda + \lambda e^{-\alpha/n} \cdot (1 + \hat{\alpha} \cdot p_i) \right)^n - 1 \\ &\leq e^{-\alpha\eta_i} \cdot \left(1 - \lambda(1 - e^{-\alpha/n}) + \lambda \cdot \hat{\alpha} \cdot p_i \right)^n - 1 \\ &\leq e^{-\alpha\eta_i} \cdot \left(1 - \frac{\lambda \cdot \alpha}{n} \cdot (1 - \alpha/n) + \lambda \cdot \hat{\alpha} \cdot p_i \right)^n - 1 \\ &\leq e^{-\alpha\eta_i} \cdot \left(1 - \frac{\alpha}{n} \cdot \hat{\delta}_i \right)^n - 1 \\ &\leq e^{-\alpha \cdot (\eta_i + \hat{\delta}_i)} - 1. \end{aligned}$$

Now, the claim follows by another application of the Taylor approximation. The second statement follows similarly. \square

Using Lemma 4, we derive different bounds on the potential drop that will be used in the various situations. First we provide some auxiliary claims (Proposition 1, Proposition 2, and Proposition 3).

Proposition 1 Consider a bin i and the values $\hat{\delta}_i$ and $\check{\delta}_i$ as defined before Lemma 4. If $\alpha \leq \ln(10/9)$, then $\max(|\hat{\delta}_i|, |\check{\delta}_i|) \leq \frac{5}{4}\lambda$.

Proof Remember that $\hat{\delta}_i := \lambda n \cdot (1/n \cdot \check{1} - p_i \cdot \check{\alpha}/\alpha)$ and $\check{\delta}_i := \lambda n \cdot (1/n \cdot \hat{1} - p_i \cdot \hat{\alpha}/\alpha)$, where $\check{1} = 1 - \alpha/n < 1 < 1 + \alpha/n = \hat{1}$ (see proof of Lemma 4). Note that if $\alpha \leq \ln(10/9)$, we have $\hat{1} < 5/4$ and $\check{1} > 8/9$. The claims hold trivially for $i = 1$, since then $p_i = (2i - 1)/n^2 = 1/n^2$ and both $|1/n \cdot \check{1} - p_i \cdot \check{\alpha}/\alpha| \leq 1/n$ and $|1/n \cdot \hat{1} - p_i \cdot \hat{\alpha}/\alpha| \leq \hat{1}/n$. For the other extreme, $i = n$, we have $p_n \leq 2/n$.

The first statement follows from this and the definition of $\hat{\alpha} = e^\alpha - 1$ since $\frac{2}{n} \cdot \frac{\hat{\alpha}}{\alpha} - \frac{1}{n} \cdot \hat{1} \leq \frac{2}{n} \frac{10/9-1}{\ln(10/9)} - \frac{1}{n} \cdot \hat{1} < \frac{5}{4n}$. Similarly, the second statement follows together with $\frac{2}{n} \frac{\hat{\alpha}}{\alpha} - \frac{1}{n} \cdot \hat{1} < \frac{1}{n}$ (which holds for any $\alpha > 0$). \square

Proposition 2 *There is a constant $\varepsilon > 0$ such that*

$$\sum_{i \leq \frac{3}{4}n} p_i \cdot \Phi_{i,+} \leq (1 - 2\varepsilon) \cdot \frac{\Phi_+}{n} \quad (16)$$

and

$$\sum_{i \in [n]} p_i \cdot \Phi_{i,-} \geq (1 + 2\varepsilon) \cdot \frac{\Phi_- - \sum_{i \leq \frac{n}{4}} \Phi_{i,-}}{n}. \quad (17)$$

Proof For Equation (16) recall that $\sum_{i < 3n/4} \Phi_{i,+} \leq \Phi_+$ (by definition). Since $\Phi_{i,+}$ for $i = 1, \dots, n$ is non-increasing where i is the i -th loaded bin, the above equation is maximized where all $\Phi_{i,+} = \frac{4\Phi_+}{3n}$. Combined with the following observation (from [23])

$$\sum_{i \geq 3n/4} p_i \geq \frac{1}{4} + \varepsilon \implies \sum_{i \leq 3n/4} p_i \leq 1 - \frac{1}{4} - \varepsilon = \frac{3}{4} - \varepsilon, \quad (18)$$

the result follows by

$$\sum_{i \leq \frac{3}{4}n} p_i \cdot \Phi_{i,+} \leq \left(\frac{3}{4} - \varepsilon\right) \frac{4\Phi_+}{3n} = \left(1 - \frac{4\varepsilon}{3n}\right) \cdot \Phi_+ \leq (1 - 2\varepsilon) \cdot \frac{\Phi_+}{n}. \quad (19)$$

Equation (17) follows similarly. \square

Proposition 3 *Consider a round t and a constant $\alpha \geq 0$. The following inequalities hold:*

1. $\sum_{i \in [n]} \alpha \eta_i (\alpha \eta_i - 1) \cdot \Phi_{i,+}(\mathbf{X}(t)) \leq \alpha^2 \eta \nu \cdot \min(n, \Phi_+(\mathbf{X}(t)))$.
2. $\sum_{i \in [n]} \alpha \eta_i (\alpha \eta_i + 1) \cdot \Phi_{i,-}(\mathbf{X}(t)) \leq \alpha^2 \eta \nu \cdot \Phi_-(\mathbf{X}(t))$.

Proof For the first statement, we calculate $\sum_{i \in [n]} \alpha \eta_i (\alpha \eta_i - 1) \cdot \Phi_{i,+}(\mathbf{X}(t))$

$$\begin{aligned} &= \sum_{i \leq \nu n} \alpha \eta_i (\alpha \eta_i - 1) \cdot \Phi_{i,+}(\mathbf{X}(t)) + \sum_{i > \nu n} \alpha \eta_i (\alpha \eta_i - 1) \cdot \Phi_{i,+}(\mathbf{X}(t)) \\ &= \alpha \eta (\alpha \eta - 1) \cdot \sum_{i \leq \nu n} \Phi_{i,+}(\mathbf{X}(t)) + \alpha \nu (1 + \alpha \nu) \cdot \sum_{i > \nu n} \Phi_{i,+}(\mathbf{X}(t)) \\ &\leq \alpha \eta (\alpha \eta - 1) \cdot \nu \cdot \Phi_+(\mathbf{X}(t)) + \alpha \nu (1 + \alpha \nu) \cdot \eta \cdot \min(n, \Phi_+(\mathbf{X}(t))) \\ &\leq \alpha^2 \eta \nu \cdot \min(n, \Phi_+(\mathbf{X}(t))), \end{aligned} \quad (20)$$

where the first inequality uses that $\Phi_{i,+}(\mathbf{X}(t))$ is non-increasing in i and that $\Phi_{i,+}(\mathbf{X}(t)) \leq 1$ for all $i > \nu n$. The claim's second statement follows by a similar calculation, using that $\Phi_{i,-}(\mathbf{X}(t))$ is non-decreasing in i (note that we cannot apply the same trick as above to get $\min(n, \Phi_-(\mathbf{X}(t)))$ instead of $\Phi_-(\mathbf{X}(t))$). \square

The remaining proof follows roughly the outline of the corresponding proof from [23] for the simpler case of a non-parallel balls-into-bins game. We start with a relative bound on the upper and lower potential change.

Lemma 5 *Consider a round t and a constant $\alpha \leq \ln(10/9)$ ($< 1/8$). Let $R \in \{+, -\}$ and $\lambda \in [1/4, 1]$. For the expected upper and lower potential drop during round $t + 1$ we have*

$$\mathbb{E}[\Delta_R(t+1) \mid \mathbf{X}(t)] < 2\alpha\lambda \cdot \Phi_R(\mathbf{X}(t)). \quad (21)$$

Proof We prove the statement for $R = +$. The case $R = -$ follows similarly. Using Lemma 4 and summing up over all $i \in [n]$ we get

$$\begin{aligned} & \mathbb{E}[\Delta_+(t+1) \mid \mathbf{X}] \\ & \leq \sum_{i \in [n]} \left(-\alpha \cdot (\eta_i + \hat{\delta}_i) + \alpha^2 \cdot (\eta_i + \hat{\delta}_i)^2 \right) \cdot \Phi_{i,+} \\ & = \sum_{i \in [n]} \left(\eta_i \alpha (\eta_i \alpha - 1) + \alpha^2 \cdot (2\eta_i \hat{\delta}_i + \hat{\delta}_i^2) - \alpha \cdot \hat{\delta}_i \right) \cdot \Phi_{i,+} \quad (22) \\ & \leq \sum_{i \in [n]} \left(\eta_i \alpha (\eta_i \alpha - 1) + 5\alpha^2 \lambda + \frac{5}{4} \alpha \lambda \right) \cdot \Phi_{i,+}. \end{aligned}$$

Here, the last inequality uses $\lambda \leq 1$ and $|\hat{\delta}_i| \leq \frac{5}{4} \lambda$ (Proposition 1). We now apply Proposition 3, $\nu \eta \leq 1/4 \leq \lambda$, and $\alpha < 1/8$ to get

$$\mathbb{E}[\Delta_+(t) \mid \mathbf{X}] \leq \left(\alpha^2 \lambda + 5\alpha^2 \lambda + \frac{5}{4} \alpha \lambda \right) \cdot \Phi_+ < 2\alpha\lambda \cdot \Phi_+. \quad (23)$$

\square

The next two lemmas derive bounds that are used to bound the upper/lower potential change in reasonably balanced configurations.

Lemma 6 *Consider a round t and the constants ε (from Proposition 2) and $\alpha \leq \min(\ln(10/9), \varepsilon/4)$. Let $\lambda \in [1/4, 1]$ and assume $X_{\frac{3}{4}n}(t) \leq \varnothing(t)$. For the expected upper potential drop during round $t + 1$ we have*

$$\mathbb{E}[\Delta_+(t+1) \mid \mathbf{X}(t)] \leq -\varepsilon\alpha\lambda \cdot \Phi_+(\mathbf{X}(t)) + 2\alpha\lambda n. \quad (24)$$

Proof To calculate the expected upper potential change, we use Lemma 4 and sum up over all $i \in [n]$ (using similar inequalities as in the proof of Lemma 5 and the definition of $\hat{\delta}_i$):

$$\begin{aligned} \mathbb{E}[\Delta_+(t+1) \mid \mathbf{X}] &\leq 6\alpha^2\lambda \cdot \Phi_+ - \sum_{i \in [n]} \alpha \cdot \hat{\delta}_i \cdot \Phi_{i,+} \\ &= (6\alpha^2\lambda - \alpha\lambda \cdot \check{1}) \cdot \Phi_+ + \hat{\alpha}\lambda n \sum_{i \in [n]} p_i \cdot \Phi_{i,+}. \end{aligned} \quad (25)$$

We now use that $\Phi_{i,+} = e^{\alpha \cdot (X_i - \varnothing)} \leq 1$ for all $i > \frac{3}{4}n$ (by our assumption on $X_{\frac{3}{4}n}$). This yields

$$\mathbb{E}[\Delta_+(t+1) \mid \mathbf{X}] \leq (6\alpha^2\lambda - \alpha\lambda \cdot \check{1}) \cdot \Phi_+ + \hat{\alpha}\lambda n \sum_{i \leq \frac{3}{4}n} p_i \cdot \Phi_{i,+} + 2\alpha\lambda n. \quad (26)$$

Finally, we apply Proposition 2 and the definition of $\check{1}$ and $\hat{\alpha}$ to get

$$\begin{aligned} \mathbb{E}[\Delta_+(t+1) \mid \mathbf{X}] &\leq (6\alpha^2\lambda - \alpha\lambda \cdot \check{1} + (1 - 2\varepsilon) \cdot \hat{\alpha}\lambda) \cdot \Phi_+ + 2\alpha\lambda n \\ &\leq (4\alpha^2\lambda - 2\varepsilon \cdot \alpha\lambda) \cdot \Phi_+ + 2\alpha\lambda n. \end{aligned} \quad (27)$$

Using $\alpha \leq \varepsilon/4$ yields the desired result. \square

Lemma 7 *Consider a round t and the constants ε (from Proposition 2) and $\alpha \leq \min(\ln(10/9), \varepsilon/8)$. Let $\lambda \in [1/4, 1]$ and assume $X_{\frac{n}{4}}(t) \geq \varnothing(t)$. For the expected lower potential drop during round t we have*

$$\mathbb{E}[\Delta_-(t+1) \mid \mathbf{X}(t)] \leq -\varepsilon\alpha\lambda \cdot \Phi_-(\mathbf{X}(t)) + \frac{\alpha\lambda n}{2}. \quad (28)$$

Proof To calculate the expected lower potential change, we use Lemma 4 and sum up over all $i \in [n]$ (as in the proof of Lemma 6):

$$\begin{aligned} \mathbb{E}[\Delta_-(t+1) \mid \mathbf{X}] &\leq 6\alpha^2\lambda \cdot \Phi_- + \sum_{i \in [n]} \alpha \cdot \check{\delta}_i \cdot \Phi_{i,-} \\ &= (6\alpha^2\lambda + \alpha\lambda \cdot \hat{1}) \cdot \Phi_- - \check{\alpha}\lambda n \sum_{i \in [n]} p_i \cdot \Phi_{i,-}. \end{aligned} \quad (29)$$

We now use that $\Phi_{i,-} = e^{\alpha \cdot (\varnothing - X_i)} \leq 1$ for all $i \leq \frac{n}{4}$ (by our assumption on $X_{\frac{n}{4}}$) and apply Proposition 2 to get

$$\begin{aligned} \mathbb{E}[\Delta_-(t) \mid \mathbf{X}] &\leq (6\alpha^2\lambda + \alpha\lambda \cdot \hat{1}) \cdot \Phi_- - (1 + 2\varepsilon) \cdot \check{\alpha}\lambda n \cdot \frac{\Phi_- - \frac{n}{4}}{n} \\ &= (6\alpha^2\lambda + \alpha\lambda \cdot \hat{1} - (1 + 2\varepsilon) \cdot \check{\alpha}\lambda) \cdot \Phi_- + (1 + 2\varepsilon) \cdot \frac{\check{\alpha}\lambda n}{4} \\ &\leq (8\alpha^2\lambda - 2\varepsilon \cdot \alpha\lambda) \cdot \Phi_- + \frac{\alpha\lambda n}{2}, \end{aligned} \quad (30)$$

where the last inequality used the definitions of $\hat{1}$, $\check{\alpha}$, as well as $\check{\alpha} > \alpha - \alpha^2$. Using $\alpha \leq \varepsilon/8$ yields the desired result. \square

The following two lemmas bound the potential drop in configurations with many balls far below the average to the right (Lemma 8) and with many balls far above the average to the left (Lemma 9).

Lemma 8 *Consider a round t and constants $\alpha \leq 1/46$ ($< \ln(10/9)$) and $\varepsilon \leq 1/3$. Let $\lambda \in [1/4, 1]$ and assume $X_{\frac{3}{4}n}(t) \geq \varnothing(t)$ and $\mathbb{E}[\Delta_+(t+1) \mid \mathbf{X}(t)] \geq -\frac{\varepsilon\alpha\lambda}{4} \cdot \Phi_+(\mathbf{X}(t))$. Then $\Phi_+(\mathbf{X}(t)) \leq \frac{\varepsilon}{4} \cdot \Phi_-(\mathbf{X}(t))$ or $\Phi(\mathbf{X}(t)) = \varepsilon^{-8} \cdot O(n)$.*

Proof Let $L := \sum_{i \in [n]} \max(X_i - \varnothing, 0) = \sum_{i \in [n]} \max(\varnothing - X_i, 0)$ be the “excess load” above and below the average. First note that the assumption $X_{\frac{3}{4}n} \geq \varnothing$ implies $\Phi_- \geq \frac{n}{4} \cdot \exp(\frac{\alpha L}{n/4})$ (using Jensen’s inequality). On the other hand, we can use the assumption $\mathbb{E}[\Delta_+(t+1) \mid \mathbf{X}] \geq -\frac{\varepsilon\alpha\lambda}{4} \cdot \Phi_+$ to show an upper bound on Φ_+ . To this end, we use Lemma 4 and sum up over all $i \in [n]$ (as in the proof of Lemma 6):

$$\begin{aligned} \mathbb{E}[\Delta_+(t+1) \mid \mathbf{X}] &\leq 6\alpha^2\lambda \cdot \Phi_+ - \sum_{i \in [n]} \alpha \cdot \hat{\delta}_i \cdot \Phi_{i,+} \\ &= 6\alpha^2\lambda \cdot \Phi_+ - \sum_{i \leq \frac{n}{3}} \alpha \cdot \hat{\delta}_i \cdot \Phi_{i,+} - \sum_{i > \frac{n}{3}} \alpha \cdot \hat{\delta}_i \cdot \Phi_{i,+}. \end{aligned} \quad (31)$$

For $i \leq n/3$ we have $p_i = \frac{2i-1}{n^2} \leq \frac{2}{3n}$ and, using the definition of $\check{1}$ and $\hat{\alpha}$, $\hat{\delta}_i = \lambda n \cdot (1/n \cdot \check{1} - p_i \cdot \hat{\alpha}/\alpha) \geq (1-5\alpha)\lambda/3$. Setting $\Phi_{\leq n/3,+} := \sum_{i \leq n/3} \Phi_{i,+}$ and $\Phi_{> n/3,+} := \sum_{i > n/3} \Phi_{i,+}$, together with Proposition 1 this yields

$$\begin{aligned} &\mathbb{E}[\Delta_+(t+1) \mid \mathbf{X}] \\ &\leq 6\alpha^2\lambda \cdot \Phi_+ - \frac{\alpha(1-5\alpha)\lambda}{3} \cdot \Phi_{\leq n/3,+} + \frac{5}{4}\alpha\lambda \cdot \Phi_{> n/3,+} \\ &= \left(6\alpha^2\lambda - \frac{\alpha(1-5\alpha)\lambda}{3}\right) \cdot \Phi_+ + \left(\frac{5}{4}\alpha\lambda + \frac{\alpha(1-5\alpha)\lambda}{3}\right) \cdot \Phi_{> n/3,+} \\ &\leq -\frac{\varepsilon\alpha\lambda}{2} \cdot \Phi_+ + 2\alpha\lambda \cdot \Phi_{> n/3,+}, \end{aligned} \quad (32)$$

where the last inequality uses $\alpha \leq 1/46 \leq \frac{1}{23} - \frac{3}{46}\varepsilon$. With this, the assumption $\mathbb{E}[\Delta_+(t+1) \mid \mathbf{X}] \geq -\frac{\varepsilon\alpha\lambda}{4} \cdot \Phi_+$ implies $\Phi_+ \leq \frac{8}{\varepsilon} \cdot \Phi_{> n/3,+} \leq \frac{8}{\varepsilon} \cdot \frac{2n}{3} e^{\frac{\alpha L}{n/3}} = \frac{16n}{3\varepsilon} e^{\frac{3\alpha L}{n}}$ (the last inequality uses that none of the $2n/3$ remaining bins can have a load higher than $L/(n/3)$). To finish the proof, assume $\Phi_+ > \frac{\varepsilon}{4} \cdot \Phi_-$ (otherwise the lemma holds). Combining this with the upper bound on Φ_+ and with the lower bound on Φ_- , we get

$$\frac{16n}{3\varepsilon} e^{\frac{3\alpha L}{n}} \geq \Phi_+ > \frac{\varepsilon}{4} \cdot \Phi_- \geq \frac{\varepsilon n}{16} \cdot e^{\frac{4\alpha L}{n}}. \quad (33)$$

Thus, the excess load can be bounded by $L < \frac{n}{\alpha} \cdot \ln\left(\frac{256}{3\varepsilon^2}\right)$. Now, the lemma’s statement follows from $\Phi = \Phi_+ + \Phi_- < \frac{5}{\varepsilon} \cdot \Phi_+ \leq \frac{80n}{3\varepsilon^2} e^{\frac{3\alpha L}{n}} = \varepsilon^{-8} \cdot O(n)$. \square

Lemma 9 Consider a round t and constants $\alpha \leq 1/32$ ($< \ln(10/9)$) and $\varepsilon \leq 1$. Let $\lambda \in [1/4, 1]$ and assume $X_{\frac{n}{4}}(t) \leq \varnothing(t)$ and $\mathbb{E}[\Delta_-(t+1) \mid \mathbf{X}(t)] \geq -\frac{\varepsilon\alpha\lambda}{4} \cdot \Phi_-(\mathbf{X}(t))$. Then $\Phi_-(\mathbf{X}(t)) \leq \frac{\varepsilon}{4} \cdot \Phi_+(\mathbf{X}(t))$ or $\Phi(\mathbf{X}(t)) = \varepsilon^{-8} \cdot O(n)$.

Proof Let $L := \sum_{i \in [n]} \max(X_i - \varnothing, 0) = \sum_{i \in [n]} \max(\varnothing - X_i, 0)$ be the “excess load” above and below the average. First note that the assumption $X_{\frac{n}{4}} \leq \varnothing$ implies $\Phi_+ \geq \frac{n}{4} \cdot e^{\frac{\alpha L}{n^4}}$ (using Jensen’s inequality). On the other hand, we can use the assumption $\mathbb{E}[\Delta_-(t+1) \mid \mathbf{X}] \geq -\frac{\varepsilon\alpha\lambda}{4} \cdot \Phi_-$ to show an upper bound on Φ_- . To this end, we use Lemma 4 and sum up over all $i \in [n]$ (as in the proof of Lemma 7):

$$\begin{aligned} \mathbb{E}[\Delta_-(t+1) \mid \mathbf{X}] &\leq 6\alpha^2\lambda \cdot \Phi_- + \sum_{i \in [n]} \alpha \cdot \check{\delta}_i \cdot \Phi_{i,-} \\ &= 6\alpha^2\lambda \cdot \Phi_- + \sum_{i \leq \frac{2n}{3}} \alpha \cdot \check{\delta}_i \cdot \Phi_{i,-} + \sum_{i > \frac{2n}{3}} \alpha \cdot \check{\delta}_i \cdot \Phi_{i,-}. \end{aligned} \quad (34)$$

For $i \geq 2n/3$ we have $p_i = \frac{2i-1}{n^2} \geq \frac{4}{3n} - \frac{1}{n^2}$. Using this with $p_i \leq p_n \leq 2/n$ and $\check{\alpha} \geq \alpha - \alpha^2$, we can bound $\check{\delta}_i = \lambda n \cdot (1/n \cdot \check{1} - p_i \cdot \check{\alpha}/\alpha) \leq \lambda \cdot (-1/3 + \frac{1+\alpha}{n}) + 2\alpha\lambda \leq -\lambda/6 + 2\alpha\lambda$. Setting $\Phi_{\leq 2n/3,-} := \sum_{i \leq 2n/3} \Phi_{i,-}$ and $\Phi_{> 2n/3,-} := \sum_{i > 2n/3} \Phi_{i,-}$, together with Proposition 1 this yields

$$\begin{aligned} &\mathbb{E}[\Delta_-(t+1) \mid \mathbf{X}] \\ &\leq 6\alpha^2\lambda \cdot \Phi_- + \frac{5}{4}\alpha\lambda \cdot \Phi_{\leq 2n/3,-} - \frac{\alpha\lambda}{6} \cdot \Phi_{> 2n/3,-} + 2\alpha^2\lambda \cdot \Phi_{> 2n/3,-} \\ &\leq (8\alpha^2\lambda - \alpha\lambda/6) \cdot \Phi_- + \left(\frac{5}{4}\alpha\lambda + \alpha\lambda/6\right) \cdot \Phi_{\leq 2n/3,-} \\ &\leq -\frac{\varepsilon\alpha\lambda}{2} \cdot \Phi_- + 2\alpha\lambda \cdot \Phi_{\leq 2n/3,-}, \end{aligned} \quad (35)$$

where the last inequality uses $\alpha \leq 1/32 \leq \frac{1}{16} - \frac{1}{48}\varepsilon$. With this, the assumption $\mathbb{E}[\Delta_-(t+1) \mid \mathbf{X}] \geq -\frac{\varepsilon\alpha\lambda}{4} \cdot \Phi_-$ implies that $\Phi_- \leq \frac{8}{\varepsilon} \cdot \Phi_{\leq 2n/3,-} \leq \frac{8}{\varepsilon} \cdot \frac{2n}{3} e^{\frac{\alpha L}{n/3}} = \frac{16n}{3\varepsilon} e^{\frac{3\alpha L}{n}}$ (the last inequality uses that none of the $2n/3$ remaining bins can have a load higher than $L/(n/3)$). To finish the proof, assume $\Phi_- > \frac{\varepsilon}{4} \cdot \Phi_+$ (otherwise the lemma holds). Combining this with the upper bound on Φ_- and with the lower bound on Φ_+ , we get

$$\frac{16n}{3\varepsilon} e^{\frac{3\alpha L}{n}} \geq \Phi_- > \frac{\varepsilon}{4} \cdot \Phi_+ \geq \frac{\varepsilon n}{16} \cdot e^{\frac{4\alpha L}{n}}. \quad (36)$$

Thus, the excess load can be bounded by $L < \frac{n}{\alpha} \cdot \ln\left(\frac{256}{3\varepsilon^2}\right)$. Now, the lemma’s statement follows from $\Phi = \Phi_+ + \Phi_- < \frac{5}{\varepsilon} \cdot \Phi_- \leq \frac{80n}{3\varepsilon^2} e^{\frac{3\alpha L}{n}} = \varepsilon^{-8} \cdot O(n)$. \square

Putting all these lemmas together, we can derive the following bound on the potential change during a single round.

Lemma 10 *Consider an arbitrary round $t+1$ of the 2-choice process and the constants ε (from Proposition 2) and $\alpha \leq \min(\ln(10/9), \varepsilon/8)$. For $\lambda \in [1/4, 1]$ we have*

$$\mathbb{E}[\Phi(\mathbf{X}(t+1)) \mid \mathbf{X}(t)] \leq \left(1 - \frac{\varepsilon\alpha\lambda}{4}\right) \cdot \Phi(\mathbf{X}(t)) + \varepsilon^{-8} \cdot O(n). \quad (37)$$

Proof The proof is via case analysis

Case 1: $x_{\frac{n}{4}} \geq \emptyset$ and $x_{\frac{3n}{4}} \leq \emptyset$. In this case the desired bound follows from Lemma 6 and Lemma 7.

Case 2: $x_{\frac{n}{4}} \geq x_{\frac{3n}{4}} \geq \emptyset$. For $\mathbb{E}[\Delta_+(t+1) \mid \mathbf{X}(t)] \leq \frac{-\varepsilon\alpha}{4}\Phi_+$ the results follows from Lemma 7.

Recall $\mathbb{E}[\Delta(t+1) \mid \mathbf{X}(t)] = \mathbb{E}[\Delta_+(t+1) \mid \mathbf{X}(t)] + \mathbb{E}[\Delta_-(t+1) \mid \mathbf{X}(t)]$. By Lemma 7 we can derive

$$\mathbb{E}[\Delta(t+1) \mid \mathbf{X}(t)] \leq \mathbb{E}[\Delta_+(t+1) \mid \mathbf{X}(t)] - \varepsilon\alpha\lambda \cdot \Phi_-(\mathbf{X}(t)) + \frac{\alpha\lambda n}{2}. \quad (38)$$

We now show that the RHS can be bounded by the RHS of Lemma 10. The result holds when

$$\begin{aligned} & \mathbb{E}[\Delta_+(t+1) \mid \mathbf{X}(t)] \\ & \leq -\frac{\varepsilon\alpha\lambda}{4} \cdot \Phi(\mathbf{X}(t)) + \varepsilon^{-8} \cdot O(n) + \varepsilon\alpha\lambda \cdot \Phi_-(\mathbf{X}(t)) - \frac{\alpha\lambda n}{2} \\ & \leq -\frac{\varepsilon\alpha\lambda}{4} \cdot \Phi(\mathbf{X}(t)) + \varepsilon\alpha\lambda \cdot \Phi_-(\mathbf{X}(t)) \\ & \leq -\frac{\varepsilon\alpha\lambda}{4} \cdot (\Phi_+(\mathbf{X}(t)) + \Phi_-(\mathbf{X}(t))) + \varepsilon\alpha\lambda \cdot \Phi_-(\mathbf{X}(t)) \\ & \leq -\frac{\varepsilon\alpha\lambda}{4} \cdot \Phi_+(\mathbf{X}(t)). \end{aligned} \quad (39)$$

For $\mathbb{E}[\Delta_+(t+1) \mid \mathbf{X}(t)] \geq \frac{-\varepsilon\alpha}{4}\Phi_+$ Lemma 8 gives two subcases.

Case 2.1: $\Phi_+(\mathbf{X}(t)) \leq \frac{\varepsilon}{4} \cdot \Phi_-(\mathbf{X}(t))$. Using Lemma 5 and Lemma 7 we obtain

$$\begin{aligned} \mathbb{E}[\Delta(t+1) \mid \mathbf{X}(t)] & \leq 2\alpha\lambda \cdot \Phi_+(\mathbf{X}(t)) - \varepsilon\alpha\lambda \cdot \Phi_-(\mathbf{X}(t)) + \frac{\alpha\lambda n}{2} \\ & \leq \frac{2\alpha\lambda\varepsilon}{4} \cdot \Phi_-(\mathbf{X}(t)) - \varepsilon\alpha\lambda \cdot \Phi_-(\mathbf{X}(t)) + \frac{\alpha\lambda n}{2} \\ & = -\frac{\varepsilon\alpha\lambda}{2} \cdot \Phi_-(\mathbf{X}(t)) + \frac{\alpha\lambda n}{2} \\ & \leq -\frac{\varepsilon\alpha\lambda}{4} \cdot \Phi(\mathbf{X}(t)) + \varepsilon^{-8} \cdot O(n). \end{aligned} \quad (40)$$

Case 2.2: $\Phi(\mathbf{X}(t)) = \varepsilon^{-8} \cdot O(n)$. Using Lemma 5 we get that $\mathbb{E}[\Delta(t+1) \mid \mathbf{X}(t)] \leq 2\alpha\lambda\varepsilon^{-8} \cdot O(n)$. It remains to show that

$$2\alpha\lambda\varepsilon^{-8} \cdot O(n) \leq -\frac{\varepsilon\alpha\lambda}{4} \cdot \Phi(\mathbf{X}(t)) + \varepsilon^{-8} \cdot O(n). \quad (41)$$

Since $\Phi(\mathbf{X}(t)) = \varepsilon^{-8} \cdot O(n)$,

$$2\alpha\lambda\varepsilon^{-8} \cdot O(n) \leq \left(1 - \frac{\varepsilon\alpha\lambda}{4}\right) \cdot \varepsilon^{-8} \cdot O(n). \quad (42)$$

This holds where $2\alpha\lambda \leq \left(1 - \frac{\varepsilon\alpha\lambda}{4}\right)$. By definition $\alpha \leq 1/8$ and $\lambda < 1$. The result follows.

Case 3: $x_{\frac{3n}{4}} \leq x_{\frac{n}{4}} \leq \emptyset$. This case is similar to case 2. For $\mathbb{E}[\Delta_-(t+1) \mid \mathbf{X}(t)] \leq \frac{-\varepsilon\alpha n}{4}\Phi_-$ the results follows from Lemma 6. For $\mathbb{E}[\Delta_-(t+1) \mid \mathbf{X}(t)] \geq \frac{-\varepsilon\alpha n}{4}\Phi_-$ two subcases are given by Lemma 9.

Case 3.1: $\Phi_-(\mathbf{X}(t)) \leq \frac{\varepsilon}{4} \cdot \Phi_+(\mathbf{X}(t))$. The result follows from applying Lemma 9 and Lemma 5.

Case 3.2: $\Phi(\mathbf{X}(t)) = \varepsilon^{-8} \cdot O(n)$. This result follows from Lemma 5. \square

We can use this result in a simple induction to prove Lemma 3.

Proof (Proof of Lemma 3) Lemma 10 gives us a $\gamma < 1$ and $c > 0$ such that $\mathbb{E}[\Phi(\mathbf{X}(t+1)) \mid \mathbf{X}(t)] \leq \gamma \cdot \Phi(\mathbf{X}(t)) + c$ holds for all rounds $t \geq 0$. Taking the expected value on both sides yields $\mathbb{E}[\Phi(\mathbf{X}(t+1))] \leq \gamma \cdot \mathbb{E}[\Phi(\mathbf{X}(t))] + c$. Using induction and the linearity of expectation, it is easy to check that $\mathbb{E}[\Phi(\mathbf{X}(t))] \leq \frac{c}{1-\gamma}$ solves this recursion. Using the values from Lemma 10 for γ and c (substituting ε' for ε) we get $\mathbb{E}[\Phi(\mathbf{X}(t))] \leq \frac{4\varepsilon'^{-8}}{\varepsilon'\alpha\lambda} \cdot O(n)$. The lemma's statement follows for the constant $\varepsilon = O(\varepsilon'^{-9}/(\alpha\lambda))$. \square

3.2 Maximum Load

The goal of this section is to prove Theorem 5. Remember the definitions of $\Phi(\mathbf{x})$ and $\Psi(\mathbf{x})$ from Equation (10). For any fixed round t , we will prove that (w.h.p.) $\Psi(\mathbf{X}(t)) = O(n \cdot \ln n)$, so that the average load is $\emptyset = O(\ln n)$. Using a union bound and Lemma 2, we see that (w.h.p.) the maximum load at the end of round t is bounded by $\emptyset + O(\ln n) = O(\ln n)$.

It remains to prove a high probability bound on $\Psi(\mathbf{X}(t))$ for arbitrary t . To get an intuition for our analysis, consider the toy case $t = \text{poly}(n)$ and assume that exactly $\lambda \cdot n \leq n$ balls are thrown each round. Here, we can combine Observation 6 and Lemma 3 to bound (w.h.p.) the load difference between any pair of bins and for all $t' < t$ by $O(\ln n)$ (via a union bound over $\text{poly}(n)$ rounds). Using the combinatorial observation that, while the load distance to the average is bounded by some $b \geq 0$, the bound $\Psi \leq 2b \cdot n$ is

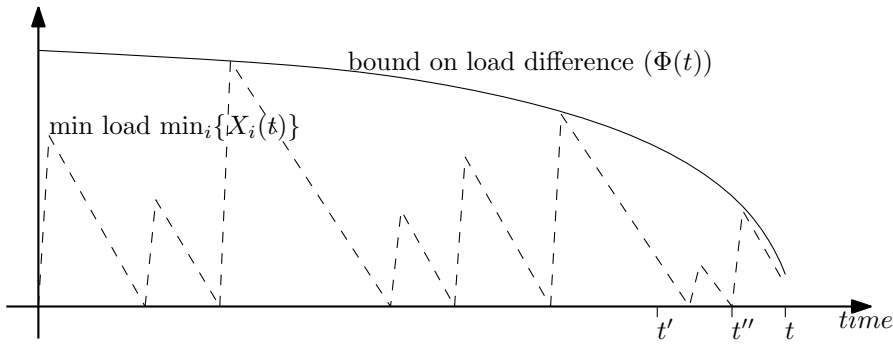


Fig. 1: To bound the system load at time t , consider the minimum load and our bound on the load difference over time. There was a last time t'' when there was an empty bin. The system load can only increase if there is an empty bin, and this increase is bounded by our bound on the load difference. Using that the system load decreases linearly in time while every increase is bounded by our logarithmic bound on the load difference, we find a small interval $[t', t]$ containing t'' .

invariant under the 2-choice process (Lemma 11), we get for $b = O(\ln n)$ that $\Psi(\mathbf{X}(t)) \leq 2b \cdot n = O(n \cdot \ln n)$, as required. The case for $t = \omega(\text{poly}(n))$ is considerably more involved. In particular, the fact that the number of balls in the system is only guaranteed to decrease when the total load is high *and* the load distance to the average is low makes it challenging to design a suitable potential function that drops fast enough when it is high. Thus, we deviate from this standard technique and elaborate on the idea of the toy case: Instead of bounding (w.h.p.) the load difference between any pair of bins by $O(\ln n)$ for all $t' < t$ (which is not possible for $t \gg \text{poly}(n)$), we prove (w.h.p.) an *adaptive bound* of $O(\ln(t - t') \cdot f(\lambda))$ for all $t' < t$, where f is a suitable function (Lemma 12). Then we consider the last round $t'' < t$ with an empty bin. Observation 6 yields a bound of $\Psi(\mathbf{X}(t'')) = 2 \cdot O(\ln(t - t'')) \cdot f(\lambda) \cdot n$ on the total load at time t'' . Using the same combinatorial observation as in the toy case, we get that (w.h.p.) $\Psi(\mathbf{X}(t)) \leq \Psi(\mathbf{X}(t'')) = 2 \cdot O(\ln(t - t'')) \cdot f(\lambda) \cdot n$. The final step is to show that the load at time t'' (which is logarithmic in $t - t''$) decreases *linearly* in $t - t''$, showing that the time interval $t - t''$ cannot be too large (or we would get a negative load at time t). See Figure 1 for an illustration.

Lemma 11 *Let $b \geq 0$ and consider a configuration \mathbf{x} with $\Psi(\mathbf{x}) \leq 2b \cdot n$ and $\Phi(\mathbf{x}) \leq e^{\alpha \cdot b}$. Let \mathbf{x}' denote the configuration after one step of the 2-choice process. Then $\Psi(\mathbf{x}') \leq 2b \cdot n$.*

Proof We distinguish two cases: if there is no empty bin, then all n bins delete one ball. Since the maximum number of new balls is n , the number of balls cannot increase. That is, we have $\Psi(\mathbf{x}') \leq \Psi(\mathbf{x}) \leq 2b \cdot n$. Now consider the case that there is at least one empty bin. Let $\eta \in (0, 1]$ denote the fraction of empty bins (i.e., there are exactly $\eta \cdot n > 0$ empty bins). Since the minimal load is zero, Observation 6 implies $\max_i x_i \leq 2b$. Thus, the total number of balls in configuration \mathbf{x} is at most $(1 - \eta)n \cdot 2b$. Exactly $(1 - \eta)n$ balls are deleted (one

from each non-empty bin) and at most n new balls enter the system. We get $\Psi(\mathbf{x}') \leq (1 - \eta)n \cdot 2b - (1 - \eta)n + n = (1 - \eta)n \cdot (2b - 1) + n \leq 2b \cdot n$. \square

Lemma 12 *Let $\lambda \in [1/4, 1)$. Fix a round t . For $i \in \mathbb{N}$ with $t - i \cdot \frac{8 \log n}{1 - \lambda} \geq 0$ define $\mathcal{I}_i := [t - i \cdot \frac{8 \log n}{1 - \lambda}, t]$. Let Y_i be the number of balls which spawn in \mathcal{I}_i .*

1. *Define the (good) smooth event $\mathcal{S}_t := \bigcap_{t' < t} (\Phi(\mathbf{X}(t')) \leq |t - t'|^2 \cdot n^2)$. Then $\Pr(\mathcal{S}_t) = 1 - O(n^{-1})$.*
2. *Define the (good) bounded balls event $\mathcal{B}_t := \bigcap_i (Y_i \leq \frac{1 + \lambda}{2} \cdot |\mathcal{I}_i| \cdot n)$. Then $\Pr(\mathcal{B}_t) = 1 - O(n^{-1})$.*

Proof Consider an arbitrary time $t' < t$. By Lemma 3 we have $\mathbb{E}[\Phi(t')] \leq n/\varepsilon$. Using Markov's inequality, this implies

$$\Pr(\Phi(t') \geq |t - t'|^2 \cdot n^2) \leq 1/(\varepsilon \cdot |t - t'|^2 \cdot n).$$

Using the union bound over all $t' < t$ we calculate

$$\Pr(\bar{\mathcal{S}}_t) \leq \sum_{t' < t} \Pr(\Phi(t') \geq |t - t'|^2 \cdot n^2) \leq \frac{1}{\varepsilon n} \cdot \sum_{t' < t} \frac{1}{|t - t'|^2} \leq \frac{\pi^2}{6\varepsilon \cdot n} = O(n^{-1}), \quad (43)$$

where the last inequality uses the solution to the Basel problem. This proves the first statement.

For the second statement, let $Z_i := |\mathcal{I}_i| \cdot n - Y_i$ be the number of balls that did not spawn during \mathcal{I}_i . Note that Z_i is a sum of $|\mathcal{I}_i| \cdot n$ independent indicator variables with $\mathbb{E}[Z_i] = (1 - \lambda) \cdot |\mathcal{I}_i| \cdot n = 8i \cdot \ln n$. Chernoff yields $\Pr(Z_i \leq (1 - \lambda) \cdot |\mathcal{I}_i| \cdot n/2) \leq e^{-8i \cdot \ln n/8} = n^{-i}$. The desired statement follows from applying the identity $Z_i = |\mathcal{I}_i| \cdot n - Y_i$ and taking the union bound. \square

Lemma 13 *Fix a round t and assume that both \mathcal{S}_t and \mathcal{B}_t hold. Then $\Psi(\mathbf{X}(t)) \leq \frac{9n}{\alpha} \cdot \ln\left(\frac{n}{1 - \lambda}\right)$.*

Proof Let $t' < t$ be the last time when there was an empty bin and set $\Delta := t - t'$. Note that t' is well defined, as we have $X_i(0) = 0$ for all $i \in [n]$. Since \mathcal{S}_t holds, we have $\Phi(\mathbf{X}(t')) \leq \Delta^2 \cdot n^2 = \exp(\ln(\Delta^2 \cdot n^2))$. By choice of t' we have $\min_i X_i(t') = 0$. Together with Observation 6 we get $\max_i X_i(t') \leq 2 \ln(\Delta^2 \cdot n^2)/\alpha$. Summing up over all bins (and pulling out the square), this implies $\Psi(\mathbf{X}(t')) \leq 4n \cdot \ln(\Delta \cdot n)/\alpha$. Applying Lemma 11 yields $\Psi(\mathbf{X}(t' + 1)) \leq 4n \cdot \ln(\Delta \cdot n)/\alpha$. By choice of t' , there is no empty bin in $\mathbf{X}(t'')$ for all $t'' \in \{t' + 1, t' + 2, \dots, t - 1\}$. Thus, during each of these rounds exactly n balls are deleted. To bound the number of deleted balls, let i be maximal with $\mathcal{I}_i \subseteq [t', t]$ (as defined in Lemma 12). Since \mathcal{B}_t holds and using the maximality of i , the number of balls Y that spawn during $[t', t]$ is at most $(1 + \lambda)|\mathcal{I}_i| \cdot n/2 + \frac{8 \log n}{1 - \lambda} \cdot n \leq$

$(1 + \lambda)\Delta \cdot n/2 + \frac{8 \ln n}{1-\lambda} \cdot n$. We calculate

$$\begin{aligned}
& \Psi(\mathbf{X}(t)) \\
& \leq \Psi(\mathbf{X}(t' + 1)) - \Delta \cdot n + Y \leq \frac{4n}{\alpha} \ln(\Delta \cdot n) - \frac{1-\lambda}{2} \Delta \cdot n + \frac{8 \ln n}{1-\lambda} \cdot n \\
& = \frac{1-\lambda}{2} \cdot n \cdot \left(\frac{8}{\alpha(1-\lambda)} \cdot \ln(\Delta \cdot n) - \Delta + \frac{16 \ln n}{(1-\lambda)^2} \right) \\
& \leq \frac{1-\lambda}{2} \cdot \Delta \cdot n \cdot \left(\frac{24}{\alpha(1-\lambda)^2} \cdot \frac{\ln(\Delta \cdot n)}{\Delta} - 1 \right). \tag{44}
\end{aligned}$$

With $f = f(\lambda) := 24/(\alpha(1-\lambda)^2)$ the last factor becomes $f \cdot \ln(\Delta \cdot n)/\Delta - 1$. It is negative if and only if $\Delta > f \cdot \ln(\Delta \cdot n)$. This inequality holds for any $\Delta > -f \cdot W_{-1}(-\frac{1}{f \cdot n})$, where W_{-1} denotes the lower branch of the Lambert W function⁴. This implies that $\Delta \leq -f \cdot W_{-1}(-1/fn)$, since otherwise we would have $\Psi(\mathbf{X}(t)) < 0$, which is clearly a contradiction. Using the Taylor approximation $W_{-1}(x) = \ln(-x) - \ln(\ln(-1/x)) - o(1)$ as $x \rightarrow -0$, we get

$$\Delta \leq -f \cdot W_{-1}\left(-\frac{1}{f \cdot n}\right) \leq f \cdot \ln(f \cdot n) + f \cdot \ln(\ln(f \cdot n)) + f \leq 2f \cdot \ln(f \cdot n). \tag{45}$$

Finally, we use this bound on Δ to get

$$\begin{aligned}
\Psi(\mathbf{X}(t)) & \leq \Psi(\mathbf{X}(t' + 1)) \leq \frac{4n}{\alpha} \cdot \ln(\Delta \cdot n) \leq \frac{4n}{\alpha} \cdot \ln(2fn \cdot \ln(fn)) \\
& \leq \frac{4n}{\alpha} \cdot \ln\left(\frac{48n}{\alpha(1-\lambda)^2} \cdot \ln\left(\frac{24n}{\alpha(1-\lambda)^2}\right)\right) \leq \frac{9n}{\alpha} \cdot \ln\left(\frac{n}{1-\lambda}\right).
\end{aligned}$$

□

Now, by combining Lemma 13 with the fact that the events \mathcal{S}_t and \mathcal{B}_t hold with high probability (Lemma 12), we immediately get that (w.h.p.) $\Psi(\mathbf{X}(t)) = O(n \cdot \ln n)$. As described at the beginning of this section, combining this with Lemma 2 proves Theorem 5.

3.3 Stability

This section proves Theorem 4. The first auxiliary lemma states that for a sufficiently high value of Γ , this potential decreases.⁵

Lemma 14 (Negative Bias Γ) *Let $\lambda \in [1/4, 1)$. If $\Gamma(\mathbf{X}(t)) \geq 2\frac{n^4}{(1-\lambda)^2\lambda}$, then*

$$\mathbb{E}[\Gamma(\mathbf{X}(t+1)) - \Gamma(\mathbf{X}(t)) \mid \mathbf{X}(t)] \leq -1.$$

⁴ Note that $-\frac{1}{f \cdot n} \geq -1/e$, so that $W_{-1}(-\frac{1}{f \cdot n})$ is well defined.

⁵ It might look tempting to use Γ together with Hajek's theorem to bound the maximum system load. However, this would require (exponentially) sharper bounds on Φ .

Proof Assume $\mathbf{X}(t) = x$ is fixed. By definition of $\Gamma(\cdot)$, we have $\Phi(x) \geq \Gamma(x)/2$ or $\Psi(x) \geq \Gamma(x)/2$. We now show that in both cases

$$\mathbb{E}[\Gamma(\mathbf{X}(t+1)) - \Gamma(x) \mid \mathbf{X}(t) = x] \leq -1.$$

1. If $\Phi(x) \geq \Gamma(x)/2$, then we have, by Lemma 10, a potential drop of

$$\begin{aligned} \mathbb{E}[\Phi(\mathbf{X}(t+1)) - \Phi(x) \mid \mathbf{X}(t) = x] &\leq -(\varepsilon\alpha\lambda/4) \cdot \Phi(x) + n \log n \\ &\leq -(\varepsilon\alpha\lambda/8) \cdot \Gamma(x) + n \log n. \end{aligned}$$

Note that, by definition of Ψ , $\Psi(\mathbf{X}(t+1)) - \Psi(x) \leq n$. Together with $\Gamma(x) \geq \frac{8(n \log n + n^2/(1-\lambda)+1)}{\varepsilon\alpha\lambda}$,

$$\begin{aligned} \mathbb{E}[\Gamma(\mathbf{X}(t+1)) - \Gamma(x) \mid \mathbf{X}(t) = x] &\leq -\frac{\varepsilon\alpha\lambda}{8}\Gamma(x) + n \log n + (n/(1-\lambda)) \cdot n \\ &\leq -1. \end{aligned}$$

2. Otherwise, i.e., if $\Phi(x) < \Gamma(x)/2$, we have that

(i) the load difference is, by Observation 6, bounded by $2 \ln(\Gamma(x)/2)/\alpha$, and

(ii) $\Psi(x) \geq \Gamma(x)/2$ must hold. This implies that $\varnothing \geq \frac{1}{n} \left(\frac{\Gamma(x)/2}{1-\lambda} \right) = \frac{(1-\lambda) \cdot \Gamma(x)}{2n^2}$.

From (i) and (ii) we have that the minimum load is at least $\frac{(1-\lambda) \cdot \Gamma(x)}{2n^2} - \ln(\Gamma(x)/2)/\alpha$. From Lemma 15 and $\Gamma(x) \geq 2 \frac{n^4}{(1-\lambda)^2\lambda}$, it follows that every bin has load at least load 1. Thus each bin will delete one ball and the number of balls arriving is λn in expectation. Hence,

$$\mathbb{E}[\Psi(\mathbf{X}(t+1)) - \Psi(x) \mid \mathbf{X}(t) = x] = -\frac{n}{1-\lambda}(1-\lambda)n.$$

Now,

$$\begin{aligned} \mathbb{E}[\Gamma(\mathbf{X}(t+1)) - \Gamma(x) \mid \mathbf{X}(t) = x] &= \\ &= \mathbb{E}[\Phi(\mathbf{X}(t+1)) - \Phi(x) \mid \mathbf{X}(t) = x] - \frac{n}{1-\lambda}(1-\lambda)n \quad (46) \\ &\leq n \log n - \frac{n}{1-\lambda}(1-\lambda)n \leq -1. \end{aligned}$$

Thus, $\mathbb{E}[\Gamma(\mathbf{X}(t+1)) - \Gamma(x) \mid \mathbf{X}(t) = x] \leq -1$, which yields the claim. \square

Lemma 15 For all $x \geq 2 \frac{n^4}{(1-\lambda)^2\lambda}$ it holds that $\frac{(1-\lambda) \cdot x}{2n^2} - 2 \ln(x/2)/\alpha \geq 1$.

Proof Define $f(x) = \frac{(1-\lambda) \cdot x}{2n^2} - 2 \ln(x/2)/\alpha$. We have $f\left(2 \frac{n^4}{(1-\lambda)^2\lambda}\right) \geq \frac{n^2}{(1-\lambda)\lambda} - \frac{2}{\alpha} \ln\left(\frac{n^4}{(1-\lambda)^2\lambda}\right) \geq 1$, where the last inequality holds for large enough of n since α is a constant. Moreover, for all $x \geq 2 \frac{n^4}{(1-\lambda)^2\lambda}$ we have $f'(x) = \frac{1-\lambda}{n^2} - \frac{2}{\alpha x} \geq 0$. Thus, the claim follows. \square

We are ready to prove Theorem 4.

Proof (Proof of Theorem 4) The proof proceeds by applying Theorem 8. We now define the parameters of Theorem 8. Let $\zeta(t) = \mathbf{X}(t)$ and hence Ω is the state space of X . First we observe that Ω is countable since there are a constant number of bins (n is consider a constant in this matter) each having a load which is a natural number. We define $\phi(\mathbf{X}(t))$ to be $\Gamma(\mathbf{X}(t))$. We define $C = \{x \mid \Gamma(x) \leq 2\frac{n^4}{(1-\lambda)^2\lambda}\}$. Define $\beta(x) = 1$ and $\eta = 1$. We now show that the preconditions (a) and (b) of Theorem 8 are fulfilled.

– Let $x \notin C$. By definition of C and $\phi(\mathbf{X}(t))$, and from Lemma 14 we have

$$E[\phi(X(t+1)) - \phi(x) \mid \mathbf{X}(t) = x] \leq E[\Gamma(\mathbf{X}(t+1)) - \Gamma(x) \mid \mathbf{X}(t) = x] \leq -1. \quad (47)$$

– Let $x \in C$. Recall that $\Gamma(\mathbf{X}(t)) = \Phi(\mathbf{X}(t)) + \Psi(\mathbf{X}(t))$. By Lemma 9 and the fact that the number of balls arriving in one round is bounded by n , we derive,

$$\begin{aligned} E[\phi(X(t+1)) \mid \mathbf{X}(t) = x] &= \\ &= E[\Phi(\mathbf{X}(t+1)) \mid \mathbf{X}(t) = x] + E[\Psi(\mathbf{X}(t+1)) \mid \mathbf{X}(t) = x] \\ &\leq \left(\left(1 - \frac{\varepsilon\alpha\lambda}{4} \right) 2\frac{n^4}{(1-\lambda)^2\lambda} \right) + \frac{n}{1-\lambda}n < \infty. \end{aligned} \quad (48)$$

The claim follows by applying Theorem 8 with Equations (47) and (48). \square

Compliance with Ethical Standards

Funding: Petra Berenbrink is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). Peter Kling is partly supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Pacific Institute for the Mathematical Sciences (PIMS). Lars Nagel is supported by the German Ministry of Education and Research under Grant 01IH13004. Christopher Wastell is supported by EPSRC.

Conflict of Interest: The authors declare that they have no conflict of interest.

References

1. Micah Adler, Petra Berenbrink, and Klaus Schröder. Analyzing an infinite parallel job allocation process. In *Proceedings of the 6th Annual European Symposium on Algorithms*, ESA '98, pages 417–428, London, UK, UK, 1998. Springer-Verlag. ISBN 3-540-64848-8. URL <http://dl.acm.org/citation.cfm?id=647908.740138>.

2. Micah Adler, Soumen Chakrabarti, Michael Mitzenmacher, and Lars Rasmussen. Parallel randomized load balancing. *Random Structures & Algorithms*, 13(2):159–188, 1998. ISSN 1098-2418. doi: 10.1002/(SICI)1098-2418(199809)13:2<159::AID-RSA3>3.0.CO;2-Q. URL [http://dx.doi.org/10.1002/\(SICI\)1098-2418\(199809\)13:2<159::AID-RSA3>3.0.CO;2-Q](http://dx.doi.org/10.1002/(SICI)1098-2418(199809)13:2<159::AID-RSA3>3.0.CO;2-Q).
3. Attahiru Sule Alfa. Algorithmic analysis of the bmap/d/k system in discrete time. *Adv. in Appl. Probab.*, 35(4):1131–1152, 12 2003. doi: 10.1239/aap/1067436338. URL <http://dx.doi.org/10.1239/aap/1067436338>.
4. 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.
5. Luca Becchetti, Andrea E. F. Clementi, Emanuele Natale, Francesco Pasquale, and Gustavo Posta. Self-stabilizing repeated balls-into-bins. In Guy E. Blelloch and Kunal Agrawal, editors, *Proceedings of the 27th ACM on Symposium on Parallelism in Algorithms and Architectures, SPAA 2015, Portland, OR, USA, June 13-15, 2015*, pages 332–339. ACM, 2015. ISBN 978-1-4503-3588-1. doi: 10.1145/2755573.2755584. URL <http://doi.acm.org/10.1145/2755573.2755584>.
6. Petra Berenbrink, Artur Czumaj, Tom Friedetzky, and Nikita D. Vvedenskaya. Infinite parallel job allocation (extended abstract). In *Proceedings of the Twelfth Annual ACM Symposium on Parallel Algorithms and Architectures, SPAA '00*, pages 99–108, New York, NY, USA, 2000. ACM. ISBN 1-58113-185-2. doi: 10.1145/341800.341813. URL <http://doi.acm.org/10.1145/341800.341813>.
7. 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.
8. Petra Berenbrink, Artur Czumaj, Matthias Englert, Tom Friedetzky, and Lars Nagel. Multiple-choice balanced allocation in (almost) parallel. In Anupam Gupta, Klaus Jansen, José Rolim, and Rocco Servedio, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 7408 of *Lecture Notes in Computer Science*, pages 411–422. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-32511-3. doi: 10.1007/978-3-642-32512-0_35. URL http://dx.doi.org/10.1007/978-3-642-32512-0_35.
9. Petra Berenbrink, Kamyar Khodamoradi, Thomas Sauerwald, and Alexandre Stauffer. Balls-into-bins with nearly optimal load distribution. In *Proceedings of the Twenty-fifth Annual ACM Symposium on Parallelism in Algorithms and Architectures, SPAA '13*, pages 326–335, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1572-2. doi: 10.1145/2486159.2486191. URL <http://doi.acm.org/10.1145/2486159.2486191>.
10. A. Czumaj and V. Stemann. Randomized allocation processes. In *Foundations of Computer Science, 1997. Proceedings., 38th Annual Symposium on*, pages 194–203, Oct 1997. doi: 10.1109/SFCS.1997.646108.

11. Artur Czumaj. Recovery time of dynamic allocation processes. In *Proceedings of the Tenth Annual ACM Symposium on Parallel Algorithms and Architectures*, SPAA '98, pages 202–211, New York, NY, USA, 1998. ACM. ISBN 0-89791-989-0. doi: 10.1145/277651.277686. URL <http://doi.acm.org/10.1145/277651.277686>.
12. G. Fayolle, V.A. Malyshev, and M.V. Menshikov. *Topics in the Constructive Theory of Countable Markov Chains*. Cambridge University Press, 1995. ISBN 9780521461979. URL <https://books.google.ca/books?id=1TJ1tFEnnHcC>.
13. Gaston H. Gonnet. Expected length of the longest probe sequence in hash code searching. *J. ACM*, 28(2):289–304, April 1981. ISSN 0004-5411. doi: 10.1145/322248.322254. URL <http://doi.acm.org/10.1145/322248.322254>.
14. B. Hajek. Hitting-time and occupation-time bounds implied by drift analysis with applications. *Advances in Applied Probability*, 14(3):502–525.
15. A.E. Kamal. Efficient solution of multiple server queues with application to the modeling of atm concentrators. In *INFOCOM '96. Fifteenth Annual Joint Conference of the IEEE Computer Societies. Networking the Next Generation. Proceedings IEEE*, volume 1, pages 248–254 vol.1, Mar 1996. doi: 10.1109/INFCOM.1996.497900.
16. Nam K Kim, Mohan L Chaudhry, Bong K Yoon, and Killwan Kim. A complete and simple solution to a discrete-time finite-capacity bmap/d/c queue. 2012.
17. David A. Levin and Yuval Peres. *Markov Chains and Mixing Times*. American Mathematical Society, December 2008. ISBN 978-0-8218-4739-8.
18. Michael Mitzenmacher. The power of two choices in randomized load balancing. *IEEE Trans. Parallel Distrib. Syst.*, 12(10):1094–1104, 2001. doi: 10.1109/71.963420. URL <http://doi.ieeecomputersociety.org/10.1109/71.963420>.
19. Yuval Peres, Kunal Talwar, and Udi Wieder. The $(1 + \beta)$ -choice process and weighted balls-into-bins. In *Proceedings of the 21st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, SODA '10, pages 1613–1619, Philadelphia, PA, USA, 2010. Society for Industrial and Applied Mathematics. ISBN 978-0-898716-98-6.
20. Martin Raab and Angelika Steger. "balls into bins" - A simple and tight analysis. In Michael Luby, José D. P. Rolim, and Maria J. Serna, editors, *Randomization and Approximation Techniques in Computer Science, Second International Workshop, RANDOM'98, Barcelona, Spain, October 8-10, 1998, Proceedings*, volume 1518 of *Lecture Notes in Computer Science*, pages 159–170. Springer, 1998. ISBN 3-540-65142-X. doi: 10.1007/3-540-49543-6_13. URL http://dx.doi.org/10.1007/3-540-49543-6_13.
21. Khosrow Sohraby and Ji Zhang. Spectral decomposition approach for transient analysis of multi-server discrete-time queues. In *INFOCOM'92. Eleventh Annual Joint Conference of the IEEE Computer and Communi-*

- cations Societies, IEEE*, pages 395–404. IEEE, 1992.
22. Volker Stemann. Parallel balanced allocations. In *Proceedings of the Eighth Annual ACM Symposium on Parallel Algorithms and Architectures, SPAA '96*, pages 261–269, New York, NY, USA, 1996. ACM. ISBN 0-89791-809-6. doi: 10.1145/237502.237565. URL <http://doi.acm.org/10.1145/237502.237565>.
 23. Kunal Talwar and Udi Wieder. Balanced allocations: A simple proof for the heavily loaded case. *CoRR*, abs/1310.5367, 2013. URL <http://arxiv.org/abs/1310.5367>.
 24. Kunal Talwar and Udi Wieder. Balanced allocations: A simple proof for the heavily loaded case. In Javier Esparza, Pierre Fraigniaud, Thore Husfeldt, and Elias Koutsoupias, editors, *Automata, Languages, and Programming*, volume 8572 of *Lecture Notes in Computer Science*, pages 979–990. Springer Berlin Heidelberg, 2014. ISBN 978-3-662-43947-0. doi: 10.1007/978-3-662-43948-7_81. URL http://dx.doi.org/10.1007/978-3-662-43948-7_81.

A Auxiliary Results

Theorem 8 (Fayolle et al. [12, Theorem 2.2.4]) *A time-homogeneous irreducible aperiodic Markov chain ζ with a countable state space Ω is positive recurrent if and only if there exists a positive function $\phi(x), x \in \Omega$, a number $\eta > 0$, a positive integer-valued function $\beta(x), x \in \Omega$, and a finite set $C \subseteq \Omega$ such that the following inequalities hold:*

1. $E[\phi(\zeta(t + \beta(x))) - \phi(x) | \zeta(t) = x] \leq -\eta\beta(x), x \notin C$
2. $E[\phi(\zeta(t + \beta(x))) | \zeta(t) = x] < \infty, x \in C$

Theorem 9 (Simplified version of Hajek [14, Theorem 2.3]) *Let $(Y(t))_{t \geq 0}$ be a sequence of random variables on a probability space (Ω, \mathcal{F}, P) with respect to the filtration $(\mathcal{F}(t))_{t \geq 0}$. Assume the following two conditions hold:*

- (i) *(Majorization) There exists a random variable Z and a constant $\lambda' > 0$, such that $\mathbb{E}[e^{\lambda' Z}] \leq D$ for some finite D , and $(|Y(t+1) - Y(t)| | \mathcal{F}(t)) \prec Z$ for all $t \geq 0$; and*
- (ii) *(Negative Bias) There exist $a, \varepsilon_0 > 0$, such for all t we have*

$$\mathbb{E}[Y(t+1) - Y(t) | \mathcal{F}(t), Y(t) > a] \leq -\varepsilon_0.$$

Let $\eta = \min \{ \lambda', \varepsilon_0 \cdot \lambda'^2 / (2D), 1 / (2\varepsilon_0) \}$. Then, for all b and t we have

$$\Pr(Y(t) \geq b | \mathcal{F}(0)) \leq e^{\eta(Y(0)-b)} + \frac{2D}{\varepsilon_0 \cdot \eta} \cdot e^{\eta(a-b)}.$$

Proof The statement of the theorem provided in [14] requires besides (i) and (ii) to choose constants η , and ρ such that $0 < \rho \leq \lambda'$, $\eta < \varepsilon_0/c$ and $\rho = 1 - \varepsilon_0 \cdot \eta + c\eta^2$ where $c = \frac{\mathbb{E}[e^{\lambda' Z}] - (1 + \lambda' \mathbb{E}[Z])}{\lambda'^2} = \sum_{k=2}^{\infty} \frac{\lambda'^{k-2}}{k!} \mathbb{E}[Z^k]$. With these requirements it then holds that for all b and t

$$\Pr(Y(t) \geq b | \mathcal{F}(0)) \leq \rho^t e^{\eta(Y(0)-b)} + \frac{1 - \rho^t}{1 - \rho} \cdot D \cdot e^{\eta(a-b)}. \quad (49)$$

In the following we bound (49) by setting $\eta = \min \{ \lambda', \varepsilon_0 \cdot \lambda'^2 / (2D), 1 / (2\varepsilon_0) \}$. The following upper and lower bound on ρ follow.

- $\rho = 1 - \varepsilon_0 \cdot \eta + c\eta^2 \leq 1 - \varepsilon_0 \cdot \eta + \varepsilon_0 \cdot \eta \cdot c \cdot \lambda'^2 / (2D) \leq 1 - \varepsilon_0 \cdot \eta + \varepsilon_0 \cdot \eta / 2 = 1 - \varepsilon_0 \cdot \eta / 2$,
where we used $c \leq D / \lambda'^2$.
- $\rho = 1 - \varepsilon_0 \cdot \eta + c\eta^2 \geq 1 - \varepsilon_0 / (2\varepsilon_0) \geq 0$.

We derive, from (49) using that for any $t \geq 0$ we have $0 \leq \rho^t \leq 1$

$$\begin{aligned} \Pr(Y(t) \geq b \mid \mathcal{F}(0)) &\leq \rho^t e^{\eta(Y(0)-b)} + \frac{1-\rho^t}{1-\rho} \cdot D \cdot e^{\eta(a-b)} \leq e^{\eta(Y(0)-b)} + \frac{1}{1-\rho} \cdot D \cdot e^{\eta(a-b)} \\ &\leq e^{\eta(Y(0)-b)} + \frac{2D}{\varepsilon_0 \cdot \eta} \cdot e^{\eta(a-b)}, \end{aligned} \tag{50}$$

since $\frac{1}{(1-\rho)} \leq \frac{2}{\varepsilon_0 \cdot \eta}$. This yields the claim. \square

Theorem 10 (Raab and Steger [20, Theorem 1]) *Let M be the random variable that counts the maximum number of balls in any bin, if we throw m balls independently and uniformly at random into n bins. Then $\Pr(M > k_\alpha) = o(1)$ if $\alpha > 1$ and $\Pr(M > k_\alpha) = 1 - o(1)$ if $0 < \alpha < 1$, where*

$$k_\alpha = \begin{cases} \frac{\log n}{\log \frac{n \log n}{m}} \left(1 + \alpha \frac{\log \log \frac{n \log n}{m}}{\log \frac{n \log n}{m}} \right) & \text{if } \frac{n}{\text{polylog}(n)} \leq m \ll n \log n \\ (d_c - 1 + \alpha) \log n & \text{if } m = c \cdot n \log n \text{ for some constant } c \\ \frac{m}{n} + \alpha \sqrt{2 \frac{m}{n} \log n} & \text{if } n \log n \ll m \leq n \text{ polylog}(n) \\ \frac{m}{n} + \sqrt{2 \frac{m}{n} \log n \left(1 - \frac{1}{\alpha} \frac{\log \log n}{2 \log n} \right)} & \text{if } m \gg n (\log n)^3, \end{cases}$$

where d_c is largest solution of $1 + x(\log c - \log x + 1) - c = 0$. We have $d_1 = e$ and $d_{1.00001} = 2.7183$.