

Balanced Allocations: A Simple Proof for the Heavily Loaded Case

Kunal Talwar*

Udi Wieder†

October 22, 2013

Abstract

We provide a relatively simple proof that the expected gap between the maximum load and the average load in the two choice process is bounded by $(1 + o(1)) \log \log n$, irrespective of the number of balls thrown. The theorem was first proven by Berenbrink et al. in [2]. Their proof uses heavy machinery from Markov-Chain theory and some of the calculations are done using computers. In this manuscript we provide a significantly simpler proof that is not aided by computers and is self contained. The simplification comes at a cost of weaker bounds on the low order terms and a weaker tail bound for the probability of deviating from the expectation.

1 A Bit of History

In the Greedy[d] process (sometimes called the d -choice process), balls are placed sequentially into $[n]$ bins with the following rule: Each ball is placed by uniformly and independently sampling d bins and assigning the ball to the least loaded of the d bins. In other words, the probability a ball is placed in one of the i heaviest bins (at the time when it is placed) is exactly¹ $(i/n)^d$. We remark that using this characterization there is no need to assume that d is a natural number (though the process is algorithmically much simpler when d is an integer). The main point is that whenever $d > 1$ the process is *biased*: the lighter bins have a higher chance of getting a ball. In this paper we are interested in the *gap* of the allocation, which is the difference between the number of balls in the heaviest bin, and the average. The case $d = 1$, when balls are placed uniformly at random in the bins, is well understood. In particular when n balls are thrown the bin with the largest number of balls will have $\Theta(\log n / \log \log n)$ balls w.h.p. Since the average is 1 this is also the gap. If $m \gg n$ balls are thrown the heaviest bin will have $m/n + \Theta(\sqrt{m \log n / n})$ balls w.h.p. [8].

In an influential paper Azar et al. [1] showed that when n balls are thrown and $d > 1$ the gap is $\log \log n / \log d + O(1)$ w.h.p. The case $d = 2$ is implicitly shown in Karp et al. [4]. The proof by Azar et al. uses a simple but clever induction; in our proof here we take the same approach. Bounding the number of balls by n (or by $O(n)$) turns out to be a crucial assumption: the proof in [1] breaks down once the number of balls is super-linear in the number of bins. Two other approaches to prove this result, namely, using differential equations or witness trees, also fail when the number of balls is large. See for example the survey [5]. A breakthrough was achieved by Berenbrink et al. in [2]. They proved that the same bound on the gap holds for *any*, arbitrarily large number of balls. Contrast this with the one choice case in which the gap diverges with the number of balls. At a (very) high level their approach was the following: first they show that the gap after m balls are thrown is distributed similarly to the gap after only $\text{poly}(n)$ balls are thrown. This is done by bounding the mixing time of the underlying Markov Chain. The second step is to extend the induction technique of [1] to the case of $\text{poly}(n)$ balls. This turns out to be a major technical challenge which involves four inductive invariants and computer aided calculations. As such, finding a simpler proof remained an interesting open problem. In this paper we provide such a proof. The simplification comes at a minor cost: we get weaker tail bounds and higher lower order terms. While [2] show that for any c , the gap is at most $\log \log n + \gamma(c)$ with probability $(1 - \frac{1}{n^c})$ for a constant $\gamma(c)$ depending on c alone, our proof shows that the gap is $\log \log n + \gamma'(c) \cdot \log \log \log n$ with probability $(1 - \frac{1}{(\log \log n)^c})$ for a constant $\gamma'(c)$ depending on c alone..

*Microsoft Research, kunal@microsoft.com

†Microsoft Research, uwieder@microsoft.com

¹Assume for simplicity and w.l.o.g that ties are broken according to some fixed ordering of the bins.

2 The Proof

We define the *load vector* X^t to be an n dimensional vector where X_i^t is the difference between the load of the i 'th bin after tn balls are thrown and the average t , (so that a load of a bin could be negative and $\sum X_i = 0$). We also assume without loss of generality that the vector is sorted so that $X_1^t \geq X_2^t \geq \dots \geq X_n^t$. We will consider the Markov chain defined by X^t , so one step of the chain consists of throwing n balls according to the d -choice scheme and then sorting and normalizing the load vector.

The main tool we use is the following Theorem proven in [7] using a potential function argument. For the reader's convenience we include a proof in Section A.

Theorem 2.1. *There exists universal constants a and b which may depend on d but not on n or t , such that, $\mathbb{E}[\sum_i \exp(a|X_i^t|)] \leq bn$.*

Let $G_t \stackrel{\text{def}}{=} X_1^t$ denote the gap between maximum and average when sampling from X^t . Theorem 2.1 immediately implies the following:

Lemma 2.2. *For any t , any $c \geq 0$, $\Pr[G^t \geq (c \log n)/a] \leq bn/n^c$. Thus for every c there is a $\gamma = \gamma(c)$ such that $\Pr[G^t \geq \gamma \log n] \leq n^{-c}$.*

Armed with this result, the crucial lemma, that we present next, says that if the gap at time t is L , then after throwing another nL balls, the gap becomes $\log \log n + O(\log L)$ with probability close to 1. A bit more formally, if $L > \log_d \log n + O(\log \log \log n)$, the tail probabilities $\Pr[G^t \geq L]$ and $\Pr[G^{t+L} \geq \log_d \log n + O(\log L)]$ differ by at most an additive $\frac{1}{\text{poly}(L)} + \frac{1}{\text{poly}(n)}$. Then using Lemma 2.2, we will infer a tail bound for $\Pr[G \geq \log_d \log n + O(\log \log \log n)]$.

Lemma 2.3. *For any $c > 0$ there is a $\gamma = \gamma(c)$, independent of n , so that for any t, ℓ, L such that $\ell \leq L \leq n^{\frac{1}{4}}$, $\Pr[G^{t+L} \geq \log_d \log n + \ell + \gamma] \leq \Pr[G^t \geq L] + 8bL^3 / \exp(a\ell) + n^{-c}$, where a, b are the constants from Theorem 2.1.*

The lemma is relatively straightforward to prove using the layered induction technique, except that we need a non-trivial “base case” to start the layered induction. Theorem 2.1 provides us with such a base case, for bins with ℓ more balls than average in X^{t+L} . For a specific ball to increase the number of balls in a bin from i to $i+1$, it must pick two bins that already contain at least i balls. If the fraction of bins with at least i balls when this ball is placed is at most β_i , then this probability would be β_i^2 . While this β_i value is a function of time, it is monotonically increasing and using the final β_i value would give us an upper bound on the probability of increase. We get such a bound for the base case using our potential function bound, and use induction and Chernoff bounds to conclude that the gap is likely to be small. We next give the details of such an argument.

Proof. We sample an allocation X^t and let G^t be its gap. Now take an additional L steps of the Markov chain to obtain X^{t+L} : in other words, we throw an additional nL balls using the d -choice process. For brevity, we will use X, G, X', G' to denote $X^t, G^t, X^{t+L}, G^{t+L}$ respectively. We condition on $G < L$ and we prove the bound for G' . Let $L' = \log_d \log n + \ell + \gamma$. Observe that:

$$\Pr[G' \geq L'] \leq \Pr[G' \geq L' \mid G < L] + \Pr[G \geq L] \quad (1)$$

It thus suffices to prove that $\Pr[G' \geq L' \mid G < L] \leq 8bL^3 / \exp(a\ell) + n^{-c}$. We do this using a layered induction similar to the one in [1].

Let ν_i be the fraction of bins with load at least i in X' , we will define a series of numbers β_i such that $\nu_i \leq \beta_i$ with high probability. For convenience, let the balls existing in X be black, and let the new nL balls thrown be white. We define the *height* of a ball to be the load of the bin in which it was placed relative to X' . Let μ_i be the number of balls (out of the nL white balls thrown) that fall at height greater than i in X' . Note that since a total of nL white balls are thrown, the average increases by L , so in order for a black ball to be in height i in X' it had to have been placed in a bin of load $L + i$ in X . The main observation is that conditioned on $G < L$, no black ball is in a bin with load more than L in X and therefore all black balls are below the average of X' . So, for any $i \geq 0$, it must be that $\nu_i n \leq \mu_i$.

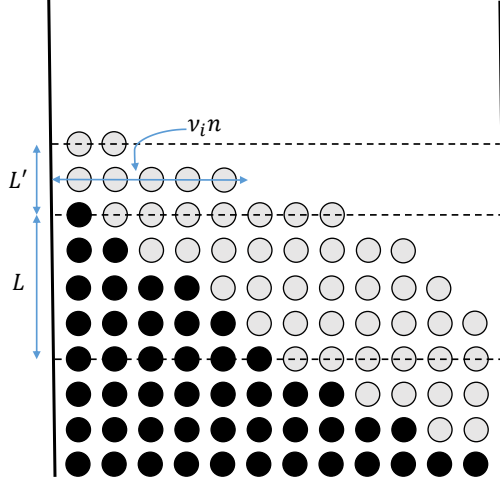


Figure 1: Black balls are in X , nL white balls are thrown to obtain X'

By Theorem 2.1 and Markov's inequality, $\Pr[\nu_\ell \geq \frac{1}{8L^{3/(d-1)}}] \leq \frac{8bL^{3/(d-1)}}{\exp(al)}$, so we can set $\beta_\ell = \frac{1}{8L^{3/(d-1)}}$ as the base of the layered induction. By the standard layered induction argument we have that w.h.p $\nu_{i+1} \leq \mu_{i+1}/n \leq 2L\beta_i^d$ and so we set $\beta_{i+1} = 2L\beta_i^d$. Since $\beta_i < \frac{1}{8L^{3/(d-1)}}$ for $i \geq \ell$, the multiplicative term of L has little impact, and we can derive the claimed bound. For completeness, we give details below. For ease of notation we assume $d = 2$, the generalization for any $d > 1$ is trivial.

Let $i_L = \ell$, $i_H = i_L + \log \log n$ and $c' = 3(c + 1)$. Let $\beta_{i_L} = \frac{1}{8L^3}$ and $\beta_{i+1} = \max(2L\beta_i^2, 2c' \log n/n)$ for $i = i_L, \dots, i_H - 1$. It is easy to check that $\beta_{i_H} = 2c' \log n/n$. Indeed the recurrence

$$\begin{aligned} \log \beta_{i_L} &= -3 \log(2L), \\ \log \beta_{i+1} &= 2 \log \beta_i + \log(2L) \end{aligned}$$

solves to $\log \beta_{i_L+k} = \log(2L)(-3 \cdot 2^k + (2^k - 1))$, which implies the claim. The inductive step in the layered induction is the following:

Lemma 2.4. *For any $i \in [i_L, i_H - 1]$, we have $\Pr[\nu_{i+1} > \beta_{i+1} \mid \nu_i \leq \beta_i] \leq \frac{1}{n^{c+1}}$.*

Proof. For a ball to fall at height at least $i + 1$, it should pick two bins that have load at least i when the ball is placed, and hence at least as much in X' . Thus the probability that a ball falls at height at least $i + 1$ is at most $\nu_i^2 \leq \beta_i^2$ under our conditioning. Since we place nL balls, the expected number of balls that fall at height at least $i + 1$ is bounded by $nL\beta_i^2 \leq n\beta_{i+1}/2$. Finally, since this number is at least $c' \log n$, Chernoff bounds imply that the probability that we get twice the expectation is at most $\exp(-c' \log n/3) \leq 1/n^{c+1}$. The claim follows. \square

It follows that $\Pr[\nu_{i_H} > \beta_{i_H} \mid G < L] \leq 8bL^3 / \exp(al) + i_H/n^{c+1}$. Now we condition on $\nu_{i_H} \leq \beta_{i_H}$, and let H be the set of bins of height at least i_H in X' . Once a bin reaches this height, an additional ball falls in it with probability at most $(2\beta_{i_H}n + 1)/n^2$. Thus the expected number of balls falling in such a bin is $O(L \log n/n)$. The probability

that any bin in H gets $2c$ balls after reaching height i_H is then at most $O(\log n \exp(-\Omega(c^2 n / 3L \log n))) \leq 1/n^{c+1}$ for large enough n . The claim follows. \square

This lemma allows us to bound $\Pr[G^{t+L} \geq \log \log n + O(\log L)]$ by $\Pr[G^t \geq L] + \frac{1}{\text{poly}(L)}$. Since $\Pr[G^t \geq O(\log n)]$ is small, we can conclude that $\Pr[G^{t+O(\log n)} \geq O(\log \log n)]$ is small. Another application of the lemma then gives that $\Pr[G^{t+O(\log n)+O(\log \log n)} \geq \log \log n + O(\log \log \log n)]$ is small. We formalize these corollaries next.

Corollary 2.5. *There is a universal constant γ such that for any $k \geq 0$, $t \geq (12 \log n)/a$, $\Pr[G^t \geq (5 + \frac{10}{a}) \cdot \log \log n + k + \gamma] \leq \frac{1}{n^{10}} + \frac{\exp(-ak)}{\log^4 n}$.*

Proof. Set $L = 12 \log n/a$, and use Lemma 2.2 to bound $\Pr[G^{t-L} \geq L]$. Set $\ell = k + \log(8bL^3 \log^4 n)/a$ to derive the result. \square

Corollary 2.6. *There are universal constants γ, α such that for any $k \geq 0$, $t \geq \omega(\log n)$, $\Pr[G^t \geq \log \log n + \alpha \log \log \log n + k + \gamma] \leq \frac{1}{n^{10}} + \frac{1}{\log^4 n} + \frac{\exp(-ak)}{(\log \log n)^4}$.*

Proof. Set $L = \log(8b(\frac{12 \log n}{a})^3 \log^4 n)/a = \frac{7 \log \log n}{a} + O_{a,b}(1)$ and use Corollary 2.5 with $k=0$ to bound $\Pr[G^{t-L} \geq L]$. Set $\ell = k + \log(8bL^3 (\log \log n)^4)/a$ to derive the result. \square

Setting $k = 0$ in Corollary 2.6, we conclude that

Corollary 2.7. *There are universal constants γ, α such that for $t \geq \omega(\log n)$, $\Pr[G^t \geq \log \log n + \alpha \log \log \log n + \gamma] \leq \frac{2}{(\log \log n)^4}$.*

Using the above results, we can also conclude

Corollary 2.8. *There are universal constants γ, α such that for $t \geq \omega(\log n)$ $\mathbb{E}[G^t] \leq \log \log n + \alpha \log \log \log n + \gamma$.*

Proof. Let $\ell_1 = \log \log n + \alpha \log \log \log n + \gamma_1$ for α, γ_1 from Corollary 2.6, and let $\ell_2 = (5 + \frac{10}{a}) \cdot \log \log n + \gamma_2$ for γ_2 from Corollary 2.5. Finally, let $\ell_3 = 12 \log n/a$. We bound

$$\mathbb{E}[G^t] \leq \ell_1 + \int_{\ell_1}^{\ell_2} \Pr[G^t \geq x] dx + \int_{\ell_2}^{\ell_3} \Pr[G^t \geq x] dx + \int_{\ell_3}^{\infty} \Pr[G^t \geq x] dx$$

Each of the three integrals are bounded by constants, using Corollaries 2.6 and 2.5 and Lemma 2.2 respectively. The claim follows. \square

The following lemma states that the lower bound condition on t is unnecessary.

Lemma 2.9. *For $t \geq t'$, $G^{t'}$ is stochastically dominated by G^t . Thus $\mathbb{E}[G^{t'}] \leq \mathbb{E}[G^t]$ and for every k , $\Pr[G^{t'} \geq k] \leq \Pr[G^t \geq k]$.*

Proof sketch. We use the notion of majorization, which is a variant of stochastic dominance. See for example [1] for definitions. Observe that trivially X^0 is majorized by $X^{t-t'}$. Now throw nt' balls using the standard coupling and get that $X^{t'}$ is majorized by X^t . The definition of majorization implies the stochastic dominance of the maximum and the bounds on the expectation and the tail follow. \square

3 Extensions

The technique we use naturally extends to other settings.

3.1 The Weighted Case

Previously we assumed all balls are of unit weight. For the case of varying weights we use the model proposed in [9] and also used in [7]. Every ball comes with a weight W independently sampled from a weight distribution \mathcal{W} . Without loss of generality we assume $\mathbb{E}[\mathcal{W}] = 1$. The weight of a bin is the sum of weights of balls assigned to it. The *gap* is naturally defined as the difference between the weight of the heaviest bin and the average bin. In [9] it is shown that if \mathcal{W} has a bounded second moment and satisfies some additional mild smoothness condition, then the expected gap does not depend on the number of balls. The paper does not provide any explicit bounds on the gap though. In [7] it is shown that if \mathcal{W} has a finite exponential generating function the gap is bounded by $O(\log n)$. For some distributions, such as the exponential distribution, this bound is tight. Here we can show that if \mathcal{W} is very concentrated (for instance it is bounded) then better bounds can be proved.

Consider for example the case where the size of each ball is drawn uniformly from $\{1, 2\}$. Previous techniques such as [2] fail to prove an $O(\log \log n)$ bound in this case, and the best bound prior to this work is the $O(\log n)$ via the potential function argument of [7]. The fact that Theorem 2.1 holds means that the technique of this paper can be applied. Moreover, the layered induction still works if we go up in steps of size two instead of one. This shows a bound of $2 \log_d \log n + O(1)$ for this distribution.

More generally, for a weight distribution W with a bounded exponential moment generating function, let M_s be the smallest value such that $\Pr[W \geq M_s] \leq \frac{1}{s(\log \log n)^s}$. Then a proof analogous to Lemma 2.3 shows that the gap is $O(\log \log n) + \sum_{i=i_L}^{i_H} M_{\beta_{i,n}}$. If M_n is $\omega(\log \log n)$, then this is $O(M_n)$, which is tight up to constants. We note however that this proof leaves a “hole”: since majorization does not necessarily hold in the weighted case, our approach proves the bound on the gap when $\Omega(n \log n)$ balls are thrown.

3.2 The Left[d] Scheme

Next we sketch how this approach also proves a tight bound for Vöcking’s Left[d] process [10]. The result had been shown in [2], though there they had to redo large sections of the proof (and the most technical at that), while here we only require minor changes. Recall that in Left[d] the bins are partitioned into d sets of n/d bins each (we assume n is divisible by d). When placing a ball, one bin is sampled uniformly from each set and the ball is placed in the least loaded of the d bins. The surprising feature of this process is that ties are broken according to a fixed ordering of the sets (we think of the sets as ordered from left to right and ties are broken ‘to the left’, hence the name of the scheme). The surprising result is that the gap now drops from $\frac{\log \log n}{\log d}$ to $\frac{\log \log n}{d \ln \phi_d}$ where $\phi_d = \lim_{k \rightarrow \infty} (F_d(k))^{\frac{1}{k}} \in [1.61, 2)$ is the base of the order d Fibonacci number.

The key ingredient in the proof is Theorem 2.1 from [7]. The exponential potential function is Schur-Convex and therefore the theorem holds for any process which is majorized by the Greedy[d] process. It is indeed the case that Vöcking’s Left[d] process [10] is majorized by Greedy[d] (see the proof in [2]). All that remains is to prove the analog of Lemma 2.3. For this we follow the analysis of Mitzenmacher and Vöcking in [6]. Let X_{jd+k} be the number of bins of load at least j from the k ’th set, and set $x_i = X_i/n$. It is easy to verify the recursive equation

$$\mathbb{E}[x_i | x_{<i}] \leq d^d \prod_{j=i-d}^{i-1} x_j$$

From here the proof is similar to that of Lemma 2.3.

4 Discussion

The theorem in [2] states that for every c there is a $\gamma = \gamma(c)$ so that $\Pr[G > \log \log n + \gamma] \leq n^{-c}$. The reason our techniques do not show such a sharp bound is that we do not obtain a small enough tail for the base case of the layered induction, i.e. on $\Pr[\nu_\ell \leq \beta_\ell]$. The reason is that the exponential potential function in Theorem 2.1 is not concentrated enough to yield such a bound. This presents a substantial obstacle, it seems that a different technique is needed in order to recover the results in [2] at full strength.

An interesting corollary from Theorem 2.1 is that the Markov chain X^t has a stationary distribution and that the bounds we prove hold also for the stationary distribution itself. In that sense, while in [2] the mixing of the chain was used to move the interesting events to be closer to the "present", in our technique we allow ourselves to look directly at the distant "future". When balls are unweighted a simple majorization based argument shows that moving closer in time can only improve the bounds on the gap (this is Lemma 2.9). Unfortunately, a similar Lemma does not hold when balls are weighted (see [3]), so we need to specify the time periods we look at. Indeed, while our results hold when considering a large number of balls, we have a 'hole' for a number of balls that is smaller than $n \log n$.

References

- [1] Yossi Azar, Andrei Broder, Anna Karlin, and Eli Upfal. Balanced allocations. *SIAM J. Computing*, 29(1):180–200, 1999.
- [2] Petra Berenbrink, Artur Czumaj, Angelika Steger, and Berthold Vöcking. Balanced allocations: The heavily loaded case. *SIAM J. Computing*, 35(6):1350–1385, 2006.
- [3] Petra Berenbrink, Tom Friedetzky, Zengjian Hu, and Russell Martin. On weighted balls-into-bins games. *Theor. Comput. Sci.*, 409(3):511–520, December 2008.
- [4] Richard M. Karp, Michael Luby, and Friedhelm Meyer auf der Heide. Efficient pram simulation on a distributed memory machine. In *STOC '92: Proceedings of the twenty-fourth annual ACM symposium on Theory of computing*, pages 318–326, New York, NY, USA, 1992. ACM.
- [5] 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*, pages 255–312. Kluwer, 2000.
- [6] Michael Mitzenmacher and Berthold Vöcking. The asymptotics of selecting the shortest of two, improved. In *Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing*, pages 326–327. Karpelevich, 1998.
- [7] Yuval Peres, Kunal Talwar, and Udi Wieder. The $(1 + \beta)$ -choice process and weighted balls-into-bins. In *SODA'10*, pages 1613–1619, 2010.
- [8] Martin Raab and Angelika Steger. "balls into bins" - a simple and tight analysis. In *RANDOM '98: Proceedings of the Second International Workshop on Randomization and Approximation Techniques in Computer Science*, pages 159–170, London, UK, 1998. Springer-Verlag.
- [9] Kunal Talwar and Udi Wieder. Balanced allocations: the weighted case. In *STOC '07: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 256–265, New York, NY, USA, 2007. ACM.
- [10] Berthold Vöcking. How asymmetry helps load balancing. *J. ACM*, 50(4):568–589, 2003.

A Potential Function

In order to make the writeup self contained we next provide a proof of Theorem 2.1.

It would be convenient to define the load vector $x(t)$ to the sorted vector of gaps after t balls are thrown, where t is not necessarily a multiple of n , as in the previous section. In other words, $x(t)_i$ is the difference between the number of balls in the i th most loaded bin and the average t/n . Note that in the notation of the previous section, X^t is $x(nt)$. The load of a bin now is not necessarily an integer. We define p_i to be the probability the i 'th loaded bin receives a ball, so $p_i = \left(\frac{i}{n}\right)^d - \left(\frac{i-1}{n}\right)^d$. Recall that we also have a weight distribution \mathcal{W} . The Markov chain is thus the following:

- sample $j \in_{\mathcal{P}} [n]$, i.e. pick j with probability p_j .
- sample $W \in \mathcal{W}$

- set $y_i = x(t)_i + W - \frac{W}{n}$ for $i = j$ and $z_i = x(t)_i - \frac{W}{n}$ for $i \neq j$
- obtain $x(t+1)$ by sorting y

We make the following two observations which hold whenever $d > 1$. It turns out to be all we need:

$$p_i \leq p_{i+1} \text{ for } i \in [n-1] \quad (2)$$

For some $\epsilon > 0$ it holds that

$$\sum_{i \geq \frac{3n}{4}} p_i \geq \frac{1}{4} + \epsilon \quad \text{and} \quad \sum_{i \leq \frac{n}{4}} p_i \leq \frac{1}{4} - \epsilon \quad (3)$$

For the distribution \mathcal{W} , we assume that there is a $\lambda > 0$ such that the moment generating function $M[\lambda] = \mathbb{E}[e^{\lambda W}] < \infty$. Further, without loss of generality, $\mathbb{E}[W] = 1$. Note that

$$M''(z) = \mathbb{E}[W^2 e^{zW}] \leq \sqrt{\mathbb{E}[W^4] \mathbb{E}[e^{2zW}]}.$$

The above assumption implies that there is an $S \geq 1$ such that for every $|z| < \lambda/2$ it holds that $M''(z) < 2S$. For simplicity, we assume throughout that n is bounded below by a large enough constant.

Let $\alpha = \min(\frac{\epsilon}{6S}, \lambda/2)$. We can assume that $\epsilon \leq 1/4$ and thus that $\alpha \leq 1/6$. Define the following potential functions

$$\begin{aligned} \Phi(x(t)) &:= \sum_{i=1}^n \exp(\alpha x(t)_i) \\ \Psi(x(t)) &:= \sum_{i=1}^n \exp(-\alpha x(t)_i) \\ \Gamma(x(t)) &:= \Phi(x(t)) + \Psi(x(t)) \end{aligned}$$

We start by calculating the expected change of Φ and Ψ individually. For ease of notation we write Φ or $\Phi(t)$ when the context is clear.

Lemma A.1. *For Φ defined as above,*

$$\mathbb{E}[\Phi(t+1) - \Phi(t) \mid x(t)] \leq \sum_{i=1}^n \left(p_i(\alpha + S\alpha^2) - \left(\frac{\alpha}{n} - S\frac{\alpha^2}{n^2} \right) \right) e^{\alpha x_i}. \quad (4)$$

Proof. Let Δ_i denote the change in $\Phi_i = \exp(\alpha x_i)$, i.e. $\Delta_i = \exp(\alpha y_i) - \exp(\alpha x_i)$, where $y_i = x_i + W - \frac{W}{n}$ with probability p_i , and $y_i = x_i - \frac{W}{n}$ otherwise. In the first case, when the ball is placed in bin i , the expected change (taken over randomness in W) Δ_i is

$$\begin{aligned} \mathbb{E}[e^{\alpha(x_i + W - \frac{W}{n})}] - e^{\alpha x_i} &= e^{\alpha x_i} (M(\alpha(1 - \frac{1}{n})) - 1) \\ &= e^{\alpha x_i} (M(0) + M'(0)\alpha(1 - \frac{1}{n}) + M''(\zeta)(\alpha(1 - \frac{1}{n}))^2/2 - 1) \end{aligned}$$

for some $\zeta \in [0, \alpha(1 - \frac{1}{n})]$. By the assumption on \mathcal{W} and α , $M''(\zeta) \leq 2S$. Moreover, $M(0) = 1$ and $M'(0) = \mathbb{E}[W] = 1$. Thus the above expression can be bounded from above by

$$e^{\alpha x_i} (\alpha(1 - \frac{1}{n}) + S\alpha^2).$$

Similarly, in the case that the ball goes to a bin other than i , the expected value of Δ_i can be bounded by $(-\frac{\alpha}{n} + S\frac{\alpha^2}{n^2})e^{\alpha x_i}$. Thus

$$\mathbb{E}[\Delta_i] \leq p_i(\alpha(1 - \frac{1}{n}) + S\alpha^2)e^{\alpha x_i} - (1 - p_i)(\frac{\alpha}{n} - S\frac{\alpha^2}{n^2})e^{\alpha x_i} \leq \left(p_i(\alpha + S\alpha^2) - \left(\frac{\alpha}{n} - S\frac{\alpha^2}{n^2} \right) \right) e^{\alpha x_i}.$$

The claim follows. \square

Corollary A.2.

$$\mathbb{E}[\Phi(t+1) - \Phi(t) \mid x(t)] \leq \frac{2\alpha}{n} \Phi(t) \quad (5)$$

Proof. Note that $S\alpha \leq \frac{1}{6} < 1$ so that

$$\mathbb{E}[\Phi(t+1) - \Phi(t) \mid x(t)] \leq \sum_{i=1}^n 2\alpha p_i e^{\alpha x_i}. \quad (6)$$

The claim follows by observing that p_i 's are increasing and x_i 's are decreasing, so that the expression is at most what it would be if the p_i 's were all equal. \square

Similar arguments show that

Lemma A.3. *Let Ψ be defined as above. Then*

$$\mathbb{E}[\Psi(t+1) - \Psi(t) \mid x(t)] \leq \sum_{i=1}^n \left(p_i(-\alpha + S\alpha^2) + \left(\frac{\alpha}{n} + S\frac{\alpha^2}{n^2} \right) \right) e^{-\alpha x_i}. \quad (7)$$

Corollary A.4.

$$\mathbb{E}[\Psi(t+1) - \Psi(t) \mid x(t)] \leq \frac{2\alpha}{n} \Psi(t) \quad (8)$$

Proof. This follows immediately as $p_i > 0$ and $S\alpha < \frac{1}{6}$. \square

We start by showing that for reasonably balanced configurations, both Φ and Ψ have the right decrease in expectation. More precisely, if $x_{\frac{3n}{4}} \leq 0$, then Φ decreases in expectation, and if $x_{\frac{n}{4}} \geq 0$, then Ψ decreases in expectation.

Lemma A.5. *Let Φ be defined as above. If $x_{\frac{3n}{4}}(t) \leq 0$, then $\mathbb{E}[\Phi(t+1) \mid x(t)] \leq (1 - \frac{\alpha\epsilon}{n})\Phi(t) + 1$.*

Proof. We upper bound $\sum_{i=1}^n p_i(\alpha + S\alpha^2)e^{\alpha x_i}$ for a fixed $\Phi(x)$, for x which is non increasing with $\sum_i x_i = 0$. We first write

$$\begin{aligned} \sum_{i=1}^n p_i(\alpha + S\alpha^2)e^{\alpha x_i} &\leq \sum_{i < \frac{3n}{4}} p_i(\alpha + S\alpha^2)e^{\alpha x_i} + \sum_{i \geq \frac{3n}{4}} p_i(\alpha + S\alpha^2)e^0 \\ &\leq \sum_{i < \frac{3n}{4}} p_i(\alpha + S\alpha^2)e^{\alpha x_i} + 1 \end{aligned} \quad (9)$$

since $\alpha + S\alpha^2 \leq \frac{6\epsilon + \epsilon^2}{36S} \leq 1$ by our assumptions that $\epsilon \leq 1$ and $S \geq 1$.

Now set $y_i := e^{\alpha x_i}$. The first term above is no larger than the maximum value of

$$(\alpha + S\alpha^2) \sum_{i < \frac{3n}{4}} p_i y_i$$

subject to

$$\sum_{i < \frac{3n}{4}} y_i \leq \Phi$$

$$y_{i-1} \geq y_i \quad \forall 1 < i < \frac{3n}{4}.$$

Since \mathbf{p} is non-decreasing and \mathbf{y} is non-increasing, the maximum is achieved when $y_i = \frac{4\Phi}{3n}$ for each i , and is at most $(\alpha + S\alpha^2)(\frac{3}{4} - \epsilon)\frac{4\Phi}{3n}$.

We can now plug this bound in (9), and substituting in (4) we upper-bound the expected change in Φ .

$$\begin{aligned}\mathbb{E}[\Phi(t+1) - \Phi(t) \mid x(t)] &\leq (\alpha + S\alpha^2)\left(\frac{3}{4} - \epsilon\right)\frac{4\Phi}{3n} - \left(\frac{\alpha}{n} - S\frac{\alpha^2}{n^2}\right)\Phi + 1 \\ &\leq \frac{\alpha\Phi}{n}\left((1+S\alpha)\left(1 - \frac{4\epsilon}{3}\right) - 1 + S\frac{\alpha}{n}\right) + 1\end{aligned}$$

Assuming $S\alpha \leq \epsilon/6$ we have

$$\begin{aligned}&\leq \frac{\alpha}{n}\Phi\left(\frac{\epsilon}{6} - \frac{4\epsilon}{3} + \frac{\epsilon}{6n}\right) + 1 \\ &\leq -\frac{\alpha\epsilon}{n}\Phi + 1\end{aligned}$$

The claim follows. \square

Lemma A.6. *Let Ψ be defined as above. If $x_{\frac{n}{4}}(t) \geq 0$, then $\mathbb{E}[\Psi(t+1) \mid x(t)] \leq (1 - \frac{\alpha\epsilon}{n})\Psi(t) + 1$.*

Proof. We first upper bound $\sum_{i=1}^n p_i(-\alpha + S\alpha^2)e^{-\alpha x_i}$ for a fixed $\Psi(x)$, for x which is non increasing with $\sum_i x_i = 0$. Since $(-\alpha + S\alpha^2)$ is negative, we have

$$\sum_{i=1}^n p_i(-\alpha + S\alpha^2)e^{-\alpha x_i} \leq (-\alpha + S\alpha^2) \sum_{i \geq \frac{n}{4}} p_i e^{-\alpha x_i}$$

Now set $z_i := e^{-\alpha x_i}$. Under the assumption on $x_{\frac{n}{4}}$, the sum $\sum_{i \geq \frac{n}{4}} z_i$ is at least $\Psi - \frac{n}{4}$. Since $(-\alpha + S\alpha^2)$ is negative, to upper bound the second term, we need to find the minimum value of

$$\begin{aligned}&\sum_{i \geq \frac{n}{4}} p_i z_i \\ &\text{subject to} \\ &\sum_{i \geq \frac{n}{4}} z_i \geq \Psi - \frac{n}{4} \\ &z_{i-1} \geq z_i \quad \forall i > \frac{n}{4}.\end{aligned}$$

Since both \mathbf{p} and \mathbf{z} are (weakly) increasing, the minimum is achieved when $z_i = \frac{4(\Psi - \frac{n}{4})}{3n}$ for each i . Using the assumption that $\sum_{i \geq n/4} p_i \geq \frac{3}{4} + \epsilon$ we can bound the expression above by $(-\alpha + S\alpha^2)\left(\frac{3}{4} + \epsilon\right)\frac{4(\Psi - \frac{n}{4})}{3n}$. We can now upper-bound the expected change in Ψ by plugging this bound in (7).

$$\begin{aligned}\mathbb{E}[\Psi(t+1) - \Psi(t) \mid x(t)] &\leq (-\alpha + S\alpha^2)\left(\frac{3}{4} + \epsilon\right)\frac{4(\Psi - \frac{n}{4})}{3n} + \frac{\alpha}{n}(1 + S\frac{\alpha}{n})\Psi \\ &= \frac{\alpha}{n}\left((1 + S\frac{\alpha}{n})\Psi + (-1 + S\alpha)\left(\frac{3}{4} + \epsilon\right)\frac{4\Psi - n}{3}\right) \\ &= \frac{\alpha}{n}\left((1 + S\frac{\alpha}{n})\Psi + S\alpha\left(\frac{3}{4} + \epsilon\right)\frac{4\Psi - n}{3} - \left(\frac{3}{4} + \epsilon\right)\frac{4\Psi - n}{3}\right) \\ &\leq \frac{\alpha\Psi}{n}\left(1 + S\frac{\alpha}{n} + S\alpha\left(\frac{3}{4} + \epsilon\right)\frac{4}{3} - \left(\frac{3}{4} + \epsilon\right)\frac{4}{3}\right) + \frac{\alpha}{3}\left(\frac{3}{4} + \epsilon\right) \\ &\leq -\frac{\alpha\epsilon}{n}\Psi + 1\end{aligned}$$

where the last inequality follows since $\epsilon \leq \frac{1}{4}$ and $S\alpha \leq \frac{\epsilon}{6}$. \square

The next lemma will be useful in the case that $x_{\frac{3n}{4}} > 0$.

Lemma A.7. *Suppose that $x_{\frac{3n}{4}} > 0$ and $\mathbb{E}[\Delta\Phi \mid x(t)] \geq -\frac{\alpha\epsilon}{4n}\Phi$. Then either $\Phi < \frac{\epsilon}{4}$ or $\Gamma < cn$ for some $c = \text{poly}(\frac{1}{\epsilon})$.*

Proof. First note that the expected increase in Φ is at most

$$\begin{aligned}
\sum_i (p_i(\alpha + S\alpha^2) - \frac{\alpha}{n} + S\frac{\alpha^2}{n^2})e^{\alpha x_i} &\leq \sum_{i \leq n/3} (p_i(\alpha + S\alpha^2) - \frac{\alpha}{n} + S\frac{\alpha^2}{n^2})e^{\alpha x_i} + (\alpha + S\alpha^2) \sum_{i > n/3} p_i e^{\alpha x_i} \\
&\leq -\frac{\alpha\epsilon}{2n}\Phi_{\leq n/3} + \frac{2\alpha}{n}\Phi_{> n/3} \\
&\leq -\frac{\alpha\epsilon}{2n}\Phi + \frac{3\alpha}{n}\Phi_{> n/3}
\end{aligned} \tag{10}$$

where in the next to last inequality we used that for $i \leq n/3$, $p_i \leq \frac{1-4\epsilon}{n}$ and that for given Φ , $\sum p_i e^{\alpha x_i}$ is maximized when \mathbf{p} is uniform.

Thus $\mathbb{E}[\Delta\Phi|x(t)] \geq -\frac{\alpha\epsilon}{4n}\Phi$ implies that

$$\frac{3\alpha}{n}\Phi_{> \frac{n}{3}} \geq \frac{\alpha\epsilon}{4n}\Phi.$$

Let $B = \sum_i \max(0, x_i) = \frac{1}{2}\|x\|_1$. Note that $\Phi_{\geq \frac{n}{3}}$ is upper bounded by $\frac{2n}{3}e^{\frac{3\alpha B}{n}}$. Thus

$$\Phi \leq \frac{12}{\epsilon}\Phi_{> \frac{n}{3}} \leq \frac{8n}{\epsilon}e^{\frac{3\alpha B}{n}}. \tag{11}$$

On the other hand, $x_{\frac{3n}{4}} > 0$ implies that $\Psi \geq \frac{n}{4}e^{\frac{4\alpha B}{n}}$.

If $\Phi < \frac{\epsilon}{4}\Psi$, we are already done. Otherwise,

$$\frac{8n}{\epsilon}e^{\frac{3\alpha B}{n}} \geq \Phi \geq \frac{\epsilon}{4}\Psi \geq \frac{\epsilon n}{16}e^{\frac{4\alpha B}{n}}$$

so that $e^{\frac{\alpha B}{n}} \leq \frac{128}{\epsilon^2}$. It follows that

$$\Gamma \leq \frac{5}{\epsilon}\Phi \leq \frac{40n}{\epsilon^2}\left(\frac{128}{\epsilon}\right)^3 \leq cn.$$

□

Similarly,

Lemma A.8. Suppose that $x_{\frac{n}{4}} < 0$ and $\mathbb{E}[\Delta\Psi|x(t)] \geq -\frac{\alpha\epsilon}{4n}\Psi$. Then either $\Psi < \frac{\epsilon}{4}\Phi$ or $\Gamma < cn$ for some $c = \text{poly}(\frac{1}{\epsilon})$.

Proof. First observe that for any $i > \frac{2n}{3}$, $p_i > \frac{1+\epsilon}{n}$ so that $p_i(-\alpha + S\alpha^2) + (\frac{\alpha}{n} + S\frac{\alpha^2}{n^2}) \leq -\frac{\alpha\epsilon}{2n}$. Since $p_i \geq 0$ it holds that $p_i(-\alpha + S\alpha^2) + (\frac{\alpha}{n} + S\frac{\alpha^2}{n^2}) \leq \frac{2\alpha}{n}$ for every i . Using the upper bound from (7) we get

$$\begin{aligned}
\mathbb{E}[\Delta\Psi | x(t)] &\leq -\frac{\alpha\epsilon}{2n}\Psi_{> \frac{2n}{3}} + \frac{2\alpha}{n}\Psi_{\leq \frac{2n}{3}} \\
&= -\frac{\alpha\epsilon}{2n}\Psi + \frac{4\alpha + \alpha\epsilon}{2n}\Psi_{\leq \frac{2n}{3}} \\
&\leq -\frac{\alpha\epsilon}{2n}\Psi + \frac{3\alpha}{n}\Psi_{\leq \frac{2n}{3}}.
\end{aligned}$$

Thus $\mathbb{E}[\Delta\Psi | x(t)] \geq -\frac{\alpha\epsilon}{4n}\Psi$ implies that

$$\frac{3\alpha}{n}\Psi_{\leq \frac{2n}{3}} \geq \frac{\alpha\epsilon}{4n}\Psi.$$

Let $B = \sum_i \max(0, x_i) = \frac{1}{2}\|x\|_1$. Note that $\Psi_{\leq \frac{2n}{3}}$ is upper bounded by $\frac{2n}{3}e^{\frac{3\alpha B}{n}}$. Thus

$$\Psi \leq \frac{12}{\epsilon}\Psi_{\leq \frac{2n}{3}} \leq \frac{8n}{\epsilon}e^{\frac{3\alpha B}{n}}. \tag{12}$$

On the other hand, $x_{\frac{n}{4}} < 0$ implies that $\Phi \geq \frac{n}{4}e^{\frac{4\alpha B}{n}}$.

If $\Psi < \frac{\epsilon}{4}\Phi$, we are already done. Otherwise,

$$\frac{8n}{\epsilon} e^{\frac{3\alpha B}{n}} \geq \Psi \geq \frac{\epsilon}{4}\Phi \geq \frac{n\epsilon}{16} e^{\frac{4\alpha B}{n}}$$

so that $e^{\frac{\alpha B}{n}} \leq \frac{128}{\epsilon^2}$. It follows that

$$\Gamma \leq \frac{5}{\epsilon}\Psi \leq \frac{40n}{\epsilon^2} \left(\frac{128}{\epsilon}\right)^3 \leq cn.$$

□

We are now ready to prove the supermartingale-type property of Γ .

Theorem A.9. *Let Γ be as above. Then $\mathbb{E}[\Gamma(t+1) \mid x(t)] \leq (1 - \frac{\alpha\epsilon}{4n})\Gamma(t) + c$, for a constant $c = c(\epsilon) = \text{poly}(\frac{1}{\epsilon})$.*

Proof. The proof proceeds via a case analysis. In case the conditions, $x_{\frac{n}{4}} \geq 0$ and $x_{\frac{3n}{4}} \leq 0$ hold, we show both Φ and Ψ decrease in expectation. If one of these is violated Lemmas A.7 and A.8 come to the rescue.

Case 1: $x_{\frac{n}{4}} \geq 0$ and $x_{\frac{3n}{4}} \leq 0$. In this case the theorem follows from Lemmas A.5 and A.6.

Case 2: $x_{\frac{n}{4}} \geq x_{\frac{3n}{4}} > 0$. Intuitively, this means that the allocation is very non symmetric with big holes in the less loaded bins. While Φ may sometimes grow in expectation, we will show that if that happens, then the asymmetry implies that Γ is dominated by Ψ which decreases. Thus the decrease in Ψ offsets the increase in Φ and the expected change in Γ is negative.

Formally, if $\mathbb{E}[\Delta\Phi|x] \leq -\frac{\alpha\epsilon}{4n}\Phi$, Lemma A.6 implies the result. Otherwise, by Lemma A.7 there are two subcases:

Case 2.1: $\Phi < \frac{\epsilon}{4}\Psi$. In this case, using Lemma A.6 and Corollary A.2

$$\mathbb{E}[\Delta\Gamma|x] = \mathbb{E}[\Delta\Phi|x] + \mathbb{E}[\Delta\Psi|x] \leq \frac{2\alpha}{n}\Phi - \frac{\alpha\epsilon}{n}\Psi + 1 \leq -\frac{\alpha\epsilon}{2n}\Psi + 1 \leq -\frac{\alpha\epsilon}{4n}\Gamma + 1$$

Case 2.2: $\Gamma < cn$. In this case, Corollaries A.2 and A.4 imply that

$$\mathbb{E}[\Delta\Gamma|x] \leq \frac{2\alpha}{n}\Gamma \leq 2c\alpha.$$

On the other hand, $c - \frac{\alpha\epsilon}{4n}\Gamma \geq c(1 - \frac{\alpha\epsilon}{4}) > 2c\alpha$.

Case 3: $x_{\frac{3n}{4}} \leq x_{\frac{n}{4}} < 0$. This case is similar to case 2. If $\mathbb{E}[\Delta\Psi|x] \leq -\frac{\alpha\epsilon}{4n}\Psi$, Lemma A.5 implies the result. Otherwise, by Lemma A.8 there are two subcases:

Case 3.1: $\Psi < \frac{\epsilon}{4}\Phi$. In this case, using Lemma A.5 and Corollary A.4, the claim follows.

Case 3.2: $\Gamma < cn$. This case is the same as case

□

Once we have shown that Γ decreases in expectation when large, we can use that to bound the expected value of Γ .

We are now ready to prove Theorem 2.1.

Theorem A.10. *For any $t \geq 0$, $\mathbb{E}[\Gamma(t)] \leq \frac{4c}{\alpha\epsilon}n$.*

Proof. We show the claim by induction. For $t = 0$, it is trivially true. By Theorem A.9, we have

$$\begin{aligned} \mathbb{E}[\Gamma(t+1)] &= \mathbb{E}[\mathbb{E}[\Gamma(t+1) \mid \Gamma(t)]] \\ &\leq \mathbb{E}[(1 - \frac{\alpha\epsilon}{4n})\Gamma(t) + c] \\ &\leq \frac{4c}{\alpha\epsilon}n(1 - \frac{\alpha\epsilon}{4n}) + c \\ &\leq \frac{4c}{\alpha\epsilon}n - c + c \end{aligned}$$

The claim follows. □