



# Balanced Allocations with the Choice of Noise

DIMITRIOS LOS and THOMAS SAUERWALD, Department of Computer Science and Technology, University of Cambridge, UK

We consider the allocation of  $m$  balls (jobs) into  $n$  bins (servers). In the standard Two-CHOICE process, at each step  $t = 1, 2, \dots, m$  we first sample two randomly chosen bins, compare their two loads and then place a ball in the least loaded bin. It is well-known that for any  $m \geq n$ , this results in a gap (difference between the maximum and average load) of  $\log_2 \log n + \Theta(1)$  (with high probability).

In this work, we consider Two-CHOICE in different settings with noisy load comparisons. One key setting involves an adaptive adversary whose power is limited by some threshold  $g \in \mathbb{N}$ . In each step, such adversary can determine the result of any load comparison between two bins whose loads differ by at most  $g$ , while if the load difference is greater than  $g$ , the comparison is correct.

For this adversarial setting, we first prove that for any  $m \geq n$  the gap is  $O(g + \log n)$  with high probability. Then through a refined analysis we prove that if  $g \leq \log n$ , then for any  $m \geq n$  the gap is  $O(\frac{g}{\log g} \cdot \log \log n)$ . For constant values of  $g$ , this generalizes the heavily loaded analysis of [19, 61] for the Two-CHOICE process, and establishes that asymptotically the same gap bound holds even if load comparisons among “similarly loaded” bins are wrong. Finally, we complement these upper bounds with tight lower bounds, which establish an interesting phase transition on how the parameter  $g$  impacts the gap.

The analysis also applies to settings with outdated and delayed information. For example, for the setting of [18] where balls are allocated in consecutive batches of size  $b = n$ , we present an improved and tight gap bound of  $\Theta(\frac{\log n}{\log \log n})$ . This bound also extends for a range of values of  $b$  and applies to a relaxed setting where the reported load of a bin can be any load value from the last  $b$  steps.

CCS Concepts: • **Theory of computation** → **Randomness, geometry and discrete structures; Distributed algorithms; Parallel algorithms**; • **Mathematics of computing** → **Probabilistic algorithms**;

Additional Key Words and Phrases: Balls-into-bins, balanced allocations, maximum load, heavily-loaded case, power-of-two-choices, adversarial noise, outdated information, parallel allocations, potential functions

## ACM Reference format:

Dimitrios Los and Thomas Sauerwald. 2023. Balanced Allocations with the Choice of Noise. *J. ACM* 70, 6, Article 37 (November 2023), 84 pages.  
<https://doi.org/10.1145/3625386>

## 1 INTRODUCTION

*Motivation.* In this work, we examine balls-and-bins processes where the goal is to allocate  $m$  balls (jobs or tasks) sequentially into  $n$  bins (processors or servers). The balls-and-bins framework a.k.a. balanced allocations [9] is a popular abstraction for various resource allocation and storage problems such as load balancing, scheduling, or hashing (see surveys [52, 63]). In order to allocate

Authors’ address: D. Los and T. Sauerwald, Department of Computer Science and Technology, University of Cambridge, 15 JJ Thomson Ave, Cambridge, Cambridgeshire, UK, CB3 0FD; e-mails: {dimitrios.los, thomas.sauerwald}@cl.cam.ac.uk.



This work is licensed under a Creative Commons Attribution International 4.0 License.

© 2023 Copyright held by the owner/author(s).

0004-5411/2023/11-ART37

<https://doi.org/10.1145/3625386>

the balls in an efficient and decentralized way, randomized strategies are usually employed which are based on sampling a number of bins for each ball, and then allocating the ball into one of those bins.

It is well-known that if each ball is placed in a bin chosen independently and uniformly at random (called ONE-CHOICE), then the maximum load is  $\Theta(\log n / \log \log n)$  w.h.p.<sup>1</sup> for  $m = n$ , and  $m/n + \Theta(\sqrt{(m/n) \log n})$  w.h.p. for  $m \geq n \log n$ . Azar, Broder, Karlin and Upfal [9] (and implicitly Karp, Luby and Meyer auf der Heide [38]) proved the remarkable result that if each ball is placed in the lesser loaded of  $d \geq 2$  randomly chosen bins, then the maximum load drops to  $\log_d \log n + O(1)$  w.h.p., if  $m = n$ . This dramatic improvement from  $d = 1$  (ONE-CHOICE) to  $d = 2$  (TWO-CHOICE) is known as “power-of-two-choices”, and similar effects have been observed in other problems including routing, hashing and randomized rounding [52]. Its practical significance was recognized by the 2020 ACM Paris Kanellakis award [8]:

*[...] it is not surprising that the power of two choices that requires only a local decision rather than global coordination has led to a wide range of practical applications. These include i-Google’s web index, Akamai’s overlay routing network, and highly reliable distributed data storage systems used by Microsoft and Dropbox, which are all based on variants of the power of two choices paradigm. There are many other software systems that use balanced allocations as an important ingredient.*

Later, Berenbrink, Czumaj, Steger, and Vöcking [19] extended the analysis of  $d$ -CHOICE to the so-called “heavily loaded case”, where  $m \geq n$  can be arbitrarily large. In particular, for TWO-CHOICE an upper bound on the gap (the difference between the maximum and average load) of  $\log_2 \log n + O(1)$  w.h.p. was shown. A simpler but slightly weaker analysis was later given by Talwar and Wieder [61].

A crucial ability of TWO-CHOICE (or  $d$ -CHOICE) is to quickly recover from an unbalanced load vector, thanks to the information and flexibility provided by the additional bin sample(s). Hence these processes typically reach an equilibrium at load vectors involving only a small number of different load values (around the mean). This can be seen as some self-stabilizing property, and this even holds in settings where balls can be removed, e.g., [11, 20, 23, 55]. Other applications of the “power-of-two-choices” effect can be found in distributed algorithms for voting and consensus [12, 14, 24, 30], population protocols [3, 5], adaptive filters [15], distributed data structures [4, 56], and online carpooling [36].

It should be noted that the theoretical results for TWO-CHOICE assume that balls arrive sequentially one by one, so that each bin has perfect knowledge about its load at any point in time. This assumption is not always met in practice, e.g., in a concurrent setting, bins may not be able to update their load immediately (even if this were possible, in some applications bins might still prefer not to reveal their true load). These issues had been observed as early as 2001 in a survey on the power-of-two-choices by Mitzenmacher, Richa and Sitaraman [52], who remarked:

*...the general problem of how to cope with incomplete or inaccurate load information and still achieve good load balancing performance appears to be an area with a great deal of research potential.*

Several studies on real-world load balancing systems [27–29, 47, 57] have also identified a discrepancy between the theoretical performance of the (sequential) TWO-CHOICE process and its observed performance in practice.

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

In this work, we are trying to improve our understanding of this problem and consider the following questions:

- (1) What if the load information of a bin at time  $t$  is outdated, e.g., it might be as small as the load at an earlier time  $t - \tau$  for some parameter  $\tau$ ?
- (2) What if the load information of a bin at time  $t$  is subject to some adversarial noise, e.g., the reported load of a bin might be an adversarial perturbation from the exact load within some range  $g$ ?
- (3) What if instead of having adversarial noise, there is a Gaussian perturbation with standard deviation  $\sigma$  on the reported loads?

The analysis in this article bridges the gap in our theoretical understanding of the Two-CHOICE process and its empirical performance, quantifying the influence of the noise parameters on the gap in more realistic settings. In outdated information settings, our results demonstrate that for small values of the delay parameter  $\tau$ , Two-CHOICE is asymptotically optimal. For large values of  $\tau$ , building on the ideas of this work, we subsequently identified and developed processes that asymptotically outperform Two-CHOICE [45].

A closely related setting to question (1) called *batching* was studied by Berenbrink, Czumaj, Englert, Friedetzky, and Nagel [18] (and later in [20]). Here the allocation of the balls proceeds in consecutive batches of size  $b$ , and the load is only updated at the end of each batch. The authors proved that for  $b = n$  the gap is  $O(\log n)$  w.h.p. Recently, the authors of this work proved some further results in the batched setting for a broader class of processes [42]. However, applying these to Two-CHOICE only gives a tight gap bound for  $b \geq n \log n$ , which is  $\Theta(b/n)$ . Mitzenmacher [50] studied the Two-CHOICE process in queuing systems, where the batched setting is referred to as *periodic update model*. His results include some fixed-point analysis and simulations, but no quantitative bounds on the gap are derived. Along similar lines, Dahlin [26] investigated several  $d$ -CHOICE-based processes, and demonstrated through experiments that allocation strategies which interpret load information based on its age, can outperform simpler strategies.

In this work, we will focus on noise settings related to the three questions above, but below we will also discuss other “robustness” aspects of Two-CHOICE that were studied in previous works.

*Our Contribution.* Our first contribution is to present a general framework of four different classes of noise settings (a formal description of these settings can be found in Section 2). This framework contains some of the previously studied processes and settings (like batching [13, 18, 42] and  $g$ -BOUNDED [56]), but it also leads to new settings, which were not studied before.

Perhaps the most important setting involves Two-CHOICE in the **adaptive adversarial setting** with parameter  $g \in \mathbb{N}$ . In each step, if the two bin samples  $i_1, i_2$  have a load difference of at most  $g$ , the adversary can manipulate the outcome of the load comparison arbitrarily and thereby decide whether the ball will be placed in  $i_1$  or  $i_2$ . However, if the load difference is more than  $g$ , the comparison will be correct and Two-CHOICE will place the ball in the less loaded bin. A slightly weaker setting, which we call **myopic setting**, works similarly, but now, in case the load difference is at most  $g$ , the ball is placed in a *random* bin (among  $\{i_1, i_2\}$ ). Through a combination of different lower and upper bounds, we establish the following phase transition on how the parameter  $g$  impacts the gap:

- (1) If  $\log n \leq g$ , then for all  $m \geq n$ ,  $\text{Gap}(m) = O(g)$  in the adversarial setting. Further, there is a matching lower bound in the myopic setting.

- (2) If  $1 < g \leq \log n$ , then for all  $m \geq n$ ,  $\text{Gap}(m) = O(\frac{g}{\log g} \cdot \log \log n)$  in the adversarial setting.<sup>2</sup>

Further, there is a matching lower bound in the myopic setting.

Equivalently, we could say that  $\text{Gap}(m) = \Theta(\frac{g}{\log g} \cdot \log \log n + g)$  for any  $g > 1$ . Both upper bounds improve and generalize a bound of  $O(g \cdot \log \log n)$  which was shown by Nadiradze [56] for the so-called  $g$ -BOUNDED process, where an adversary “greedily” reverts all comparisons if the load difference is at most  $g$ . We believe that our new bounds could be helpful in obtaining a tighter analysis of the *multi-counter data structure* studied in [4, 56]. Note that for constant  $g$ , our upper bound shows that Two-CHOICE maintains a  $O(\log \log n)$  gap even if bins with constant load difference cannot be compared correctly. While it may seem intuitive that such a result should hold, we are not aware of any simple argument. Even in the most basic case  $g = 1$ , there seems to be no ad-hoc method, e.g., using couplings or majorization, which would extend the upper bound of  $O(\log \log n)$  for Two-CHOICE from [19, 61] to this noise setting – not to mention a more general albeit loose upper bound of  $O(g \cdot \log \log n)$  for any  $g > 1$ .

We then proceed to settings with outdated load information. In the general **delay setting**, each load information of a bin can be outdated by at most  $\tau \geq 1$  rounds. This is a generalization of the **batched setting** [18], where the load information of all bins is updated after rounds  $b, 2b, \dots$ . For a range of values for  $\tau$  and  $b$  around  $n$ , we prove tight bounds on the gap in the delay and batched setting. This upper bound is matched by a trivial lower bound for the batched setting, which is based on the gap created by ONE-CHOICE when allocating the first  $b$  balls randomly into the  $n$  bins. Our upper bound demonstrates that the special property of batching to reset all load values to their correct value at the beginning of a batch is not crucial, and it suffices if bin loads are updated asynchronously. Finally, for the batched setting with  $b = n$ , our results improve the gap bound of  $O(\log n)$  from [18] to a tight bound of  $\Theta(\frac{\log n}{\log \log n})$ .

Finally, complementing the adversarial and myopic settings, we also investigate a **probabilistic noise setting**. In this setting, whenever two bins with load difference  $\delta$  are compared, the comparison will be correct with probability  $\rho(\delta) \in [0, 1]$ . One natural instance of this setting is when the reported bin loads are randomly perturbed by some Gaussian noise with variance  $\sigma^2$ , which essentially leads to  $\rho(\delta) = 1 - \frac{1}{2} \exp(-(\delta/\sigma)^2)$ . We do not present a tight analysis of this setting, but our upper and lower bounds demonstrate that the gap is polynomial in  $\sigma$  and poly-logarithmic in  $n$ .

In Section 2, we formally define the various processes and settings, and in Table 1 we give a summary of upper and lower bounds proved in this article.

*Our Techniques.* Since most of our upper bounds use similar techniques as the upper bound for the adversarial setting, we only outline the proof of the latter. As with most of the previous works, e.g., [10, 43, 58], we make extensive use of exponential potential functions in order to prove that the process stabilizes. This technique essentially suffices to prove a weaker gap bound of  $O(g \cdot \log \log n)$  for the adversarial setting. However, in order to prove stronger gap bounds, we require that a technical precondition on the load distribution is satisfied in most of the rounds. Roughly, this precondition says that a linear potential function is small (even though the exponential potential function may be very large). Extending [46], we establish this precondition by studying the interplay between three potential functions: the linear, quadratic and exponential potential functions.

For sub-logarithmic values of  $g$ , we prove sub-logarithmic gap bounds by employing a type of layered induction argument over a series of super-exponential potential functions. These potential functions are extensions of those used in [43] to circumvent rounding issues and prove more fine-

<sup>2</sup>For  $g = 1$ , we prove a bound of  $\Theta(\log \log n)$  and for  $g = 0$ , the process is equivalent to the Two-CHOICE process without noise.

Table 1. Overview of the Lower and Upper Bounds for Different Noise Settings Derived in Previous Works (Rows in Gray) and in This Work (Rows in Green)

Setting/Process	Range	Lower Bound	Reference	Upper Bound	Reference
$g$ -BOUNDED	$1 \leq g$	–	–	$O(g \cdot \log(ng))$	[56, Thm 2.5.12] Thm 4.3
$g$ -ADV-COMP	$1 \leq g$	–	–	$O(g \cdot \log(ng))$	Thm 4.3
$g$ -ADV-COMP	$1 \leq g$	–	–	$O(g + \log n)$	Thm 5.12
$g$ -ADV-COMP	$1 < g \leq \log n$	–	–	$O\left(\frac{g}{\log g} \cdot \log \log n\right)$	Thm 9.2
$g$ -MYOPIC-COMP	$\frac{\log n}{\log \log n} \leq g$	$\Omega(g)$	Pro 11.2	–	–
$g$ -MYOPIC-COMP	$1 < g \leq \frac{\log n}{\log \log n}$	$\Omega\left(\frac{g}{\log g} \cdot \log \log n\right)$	Obs 11.1 Thm 11.3	–	–
$b$ -BATCH	$b = \Omega(n \log n)$	$\Omega\left(\frac{b}{n}\right)$	[42, Prop 7.4]	$O\left(\frac{b}{n}\right)$	[42, Thm 5.1]
$b$ -BATCH	$b = n$	$\Omega\left(\frac{\log n}{\log \log n}\right)$	Obs 11.6	$O(\log n)$	[18, Thm 1]
$\tau$ -DELAY	$\tau = n$	–	–	$O\left(\frac{\log n}{\log \log n}\right)$	Thm 10.2
$\tau$ -DELAY	$\tau \in [n \cdot e^{-\log^c n}, n \log n]$	–	–	$O\left(\frac{\log n}{\log((4n/\tau) \log n)}\right)$	Cor 10.4
$\tau$ -DELAY	$\tau = n^{1-\varepsilon}$	–	–	$O(\log \log n)$	Rem 10.6
$b$ -BATCH	$b \in [n \cdot e^{-\log^c n}, n \log n]$	$\Omega\left(\frac{\log n}{\log((4n/b) \log n)}\right)$	Obs 11.6	–	–
$b$ -BATCH	$b = n^{1-\varepsilon}$	$\Omega(\log \log n)$	Obs 11.1	–	–
$\sigma$ -NOISY-LOAD	$1 \leq \sigma$	–	–	$O(\sigma \sqrt{\log n} \cdot \log(n\sigma))$	Pro 10.1
$\sigma$ -NOISY-LOAD	$2 \cdot (\log n)^{-1/3} \leq \sigma$	$\Omega(\min\{1, \sigma\} \cdot (\log n)^{1/3})$	Pro 11.5	–	–
$\sigma$ -NOISY-LOAD	$32 \leq \sigma$	$\Omega(\min\{\sigma^{4/5}, \sigma^{2/5} \cdot \sqrt{\log n}\})$	Pro 11.5	–	–

All upper bounds hold for all values of  $m \geq n$  w.h.p., while lower bounds may only hold for a suitable value of  $m$  w.h.p. Recall that the  $g$ -BOUNDED and  $g$ -MYOPIC-COMP processes are instances of the  $g$ -ADV-COMP setting, and  $b$ -BATCH an instance of  $\tau$ -DELAY, for  $\tau = b$ . The parameters  $c, \varepsilon$  are any constants in  $(0, 1)$ .

grained gap bounds. The layered induction argument (without use of potential functions) is an established technique for analyzing the lightly case where  $m = n$  [9].

*Organization.* In Section 2 we present in more detail the different noise settings considered in this work. Then, in Section 3 we introduce some basic notation and definitions used in the later analysis. In Section 4, we present our first bound of  $O(g \cdot \log(ng))$  on the gap for the adversarial setting (Theorem 4.3). In Section 5, this gap bound is then refined to  $O(g + \log n)$  (Theorem 5.12). In Section 6, we outline the use of a layered-induction inspired technique to prove a gap bound of  $O\left(\frac{g}{\log g} \cdot \log \log n\right)$  for  $g = o(\log n)$  (Theorem 9.2) and in the following three sections we give the proof of this bound. More specifically, in Section 7, we strengthen the results of Section 5 to obtain the base case of the layered induction, in Section 8 we provide a general analysis of super-exponential potentials and in Section 9, we use this to complete the layered induction step. Next, in Section 10 we derive upper bounds for both the probabilistic noise and the delay settings, essentially through basic reductions to the adversarial setting. Section 11 contains our lower bounds for the different settings, including a tight lower bound of  $\Omega\left(\frac{g}{\log g} \cdot \log \log n + g\right)$  for  $g$ -MYOPIC-COMP (Corollary 11.4). Section 12 presents some experimental results and in Section 13 we conclude with a brief summary of the main results and some open questions.

*Further Related Work.* Mitzenmacher [49] introduced the  $(1 + \beta)$ -process, in which two choices are available with probability  $\beta \in (0, 1]$ , and otherwise only one. This process mixes ONE-CHOICE with TWO-CHOICE steps and is useful in applications where it is costly to always take two bin samples. Peres, Talwar and Wieder [58] proved an upper bound of  $O(\log n / \beta + \log(1/\beta)/\beta)$  on the gap for any  $m \geq n$  showing that we can achieve a gap that is independent of  $m$  for  $\beta := \beta(n)$ . A natural generalization of TWO-CHOICE studied in the same work [58] (and earlier in [39] for

$m = n$ ) is the so-called *graphical allocation setting*. In this setting, bins correspond to vertices of a graph, and for each ball we sample an edge uniformly at random and place the ball in the lesser loaded bin of the two endpoints. The results of [58] show that for any connected and regular graph, the gap remains independent of  $n$ . In [10], Bansal and Feldheim analyzed a sophisticated algorithm that achieves a poly-logarithmic gap on sparse regular graphs in the graphical allocation setting. For a related process where averaging is allowed, a tighter bound for the cycle was shown by Alistarh, Nadiradze and Sabour [6]. Also recently, Greenhill, Mans and Pourmiri studied the graphical allocation setting on dynamic hypergraphs [35].

In the  $(1 + \beta)$ -process the decision whether to take a second bin sample does not depend on the first bin sample. In contrast to that, the Two-THINNING process allocates in a two-stage procedure: Firstly, sample a random bin  $i$ . Secondly, based on the load of bin  $i$  (and additional information based on the history of the process), we can either place the ball into  $i$ , or place the ball into another randomly chosen bin  $j$  (without comparing its load with  $i$ ). This process has received a lot of attention lately, and several variations were studied in [33] for  $m = n$  and [32, 43, 46] for  $m \geq n$ . Finally, Czumaj and Stemmann [25] investigated so-called *adaptive allocation schemes* for  $m = n$ . In contrast to THINNING, after having taken a certain number of bin samples, the ball is allocated into the least loaded bin among all samples. In another related model recently studied by the authors of this work, the load of a sampled bin can only be approximated through binary queries of the form “Is your load at least  $g$ ?” [43]. It was shown that by using  $1 \leq k = O(\log \log n)$  queries for each of the two samples, for any  $m \geq n$ , the gap is  $O(k \cdot (\log n)^{1/k})$  w.h.p.

A setting relaxing the *uniform sampling* assumption was investigated by Wieder [62], which shows that for any  $d > 1$ , the  $d$ -CHOICE gap bounds continue to hold as long as the probability by which the  $d$  bins are sampled is close enough to uniform. A setting with heterogeneous bin capacities was studied in Berenbrink, Brinkmann, Friedetzky and Nagel [17], who showed the gap bound of  $\log_d \log n + O(1)$  for  $d$ -CHOICE continues to hold.

There is also a rich line of work investigating randomized allocation schemes which use fewer random bits. For example, Alon, Gurel–Gurevich and Lubetzky [7] established a tradeoff between the number of bits used for the representation of the load and the number of  $d$  bin choices. For  $d = 2$  choices, Benjamini and Makarychev [16] presented some tight results relating the gap of Two-CHOICE to the available memory.

We remark that parallel versions of Two-CHOICE exist that involve only a very small number of rounds of interactions between all balls and bins. Most of the studies, e.g., [1, 41] focus on  $m = n$ , and only recently the heavily loaded case was addressed in [40].

From a higher perspective, investigating the complexity of problems in the presence of noisy data is a popular area in algorithms, machine learning and distributed computing. For example, of similar flavor as our noisy allocation setting are studies on sorting and ranking with unreliable information [2, 21]. In the paradigm of algorithms with predictions, different scheduling algorithms under noisy job size estimates have been analyzed [51, 54, 60].

## 2 OUR SETTINGS

All of our settings are based on running the Two-CHOICE process with  $m$  balls into  $n$  bins, labeled  $[n] := \{1, 2, \dots, n\}$ . In the normal setting, referred to as “without noise”, the load comparisons between the two bin samples are correct in all steps. Here, we investigate several noise settings where the outcome of the load comparisons may not always be correct. We distinguish between three classes of noise settings: (1) Adversarial Load and Comparison, (2) Adversarial Delay and (3) Probabilistic Load and Comparison. See Figure 1 for the connections between the different settings.



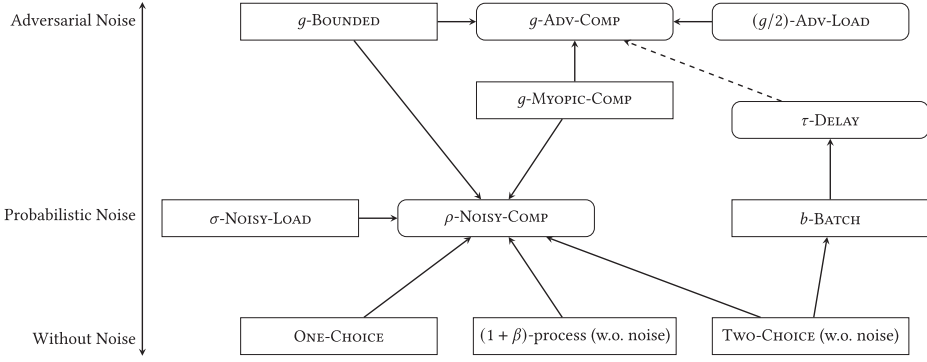


Fig. 1. Overview of settings (rounded rectangles) and processes (rectangles). A directed arrow from setting (process)  $A$  to setting (process)  $B$  means that  $B$  is stronger than  $A$  (that is,  $B$  can simulate  $A$ ). For  $\tau$ -DELAY, a dashed arrow is used for the connection to  $g$ -ADV-COMP, as the (strong) connection is slightly more involved.

#### Two-CHOICE Process (without noise):

Iteration: For each step  $t \geq 1$ ,

- (1) Sample two bins  $i_1 = i_1^t$  and  $i_2 = i_2^t$  with replacement, independently and uniformly at random.
- (2) Let  $i^t \in \{i_1, i_2\}$  be a bin such that  $x_{i^t}^{t-1} = \min\{x_{i_1}^{t-1}, x_{i_2}^{t-1}\}$ , breaking ties arbitrarily.
- (3) Allocate one ball to bin  $i^t$ .

We now present our framework where Two-CHOICE runs in a noisy setting, meaning that step 2 is subject to some noise. One possibility is to have an adversary that is able to replace the load values  $x_{i_1}^{t-1}$  and  $x_{i_2}^{t-1}$  by estimates  $\tilde{x}_{i_1}^{t-1}$  and  $\tilde{x}_{i_2}^{t-1}$ , while the bin  $i^t$  will always be the one with the better load estimate. A more general framework is to have a two-sample process with a constrained decision function that decides in which of the two samples  $i_1$  and  $i_2$  to allocate to. This function may depend on the filtration  $\mathcal{F}^{t-1}$  corresponding to the first  $t - 1$  allocations of the process.<sup>3</sup>

#### Two-CHOICE Process with Noise:

Noise: The adversary  $A^t : \mathcal{F}^{t-1} \times [n] \times [n] \rightarrow [n]$

Iteration: For each step  $t \geq 1$ ,

- (1) Sample two bins  $i_1$  and  $i_2$  with replacement, independently and uniformly at random.
- (2) Let  $i^t = A^t(\mathcal{F}^{t-1}, i_1, i_2) \in \{i_1, i_2\}$  be the bin determined by the adversary.
- (3) Allocate one ball to bin  $i^t$ .

This is a very general setting, and in order to prove any meaningful bounds, we will have to restrict the power of the adversary  $A^t$ , giving rise to different settings defined below. Also note that the adversary  $A^t$  is allowed to use coin-flips, meaning that the above framework encompasses settings in which, for instance, load estimates or comparisons are determined probabilistically.

Clearly we can recover Two-CHOICE without noise as a special case, if the adversary uses

$$A^t(\mathcal{F}^{t-1}, i_1, i_2) = \arg \min_{k \in \{i_1, i_2\}} x_k^{t-1},$$

<sup>3</sup>In particular, filtration  $\mathcal{F}^t$  reveals the load vector  $x^t$ .

breaking ties arbitrarily. On the other hand, for the  $g$ -BOUNDED process,

$$A^t(\mathcal{G}^{t-1}, i_1, i_2) = \begin{cases} \arg \max_{k \in \{i_1, i_2\}} x_k^{t-1} & \text{if } |x_{i_1}^{t-1} - x_{i_2}^{t-1}| \leq g, \\ \arg \min_{k \in \{i_1, i_2\}} x_k^{t-1} & \text{otherwise,} \end{cases}$$

breaking ties arbitrarily.

### Adversarial Load and Comparison

**Settings.** We first present a setting with an *adaptive adversary*, who has direct control on the outcome of load comparisons (provided that the load difference is small). Specifically, in  $g$ -ADV-LOAD for  $g \in \mathbb{N}$ , at each step  $t = 1, 2, \dots$  the adversary first determines for each bin  $k$  a load estimate  $\tilde{x}_k^{t-1} \in [x_k^{t-1} - g, x_k^{t-1} + g]$ . Then our process samples two bins  $i_1$  and  $i_2$ , and allocates into the bin with the smaller load estimate (ties can be broken arbitrarily). A slightly stronger adversary appears in the setting  $g$ -ADV-COMP. Here, at each step  $t = 1, 2, \dots$  the process samples two bins  $i_1$  and  $i_2$ , and then the adversary is able to determine the outcome of the comparison if  $|x_{i_1}^{t-1} - x_{i_2}^{t-1}| \leq g$ , thereby deciding where TWO-CHOICE places the ball. Otherwise, if  $|x_{i_1}^{t-1} - x_{i_2}^{t-1}| > g$ , TWO-CHOICE will place the ball in the less loaded bin. For  $g = 0$ , we recover the TWO-CHOICE process without noise.

We remark that  $g$ -ADV-LOAD can be simulated by  $(2g)$ -ADV-COMP, which is why we will only study the slightly more powerful ADV-COMP setting here.

**Processes.** A special instance of  $g$ -ADV-COMP is one that forces TWO-CHOICE to allocate the ball to the heavier of the two bins when  $|x_{i_1}^{t-1} - x_{i_2}^{t-1}| \leq g$ . This process was studied in [56] under the name  $g$ -BOUNDED and served as an analysis tool for the multi-counter distributed data structure.

Another special instance of the *adversarial* setting is a *myopic* process, where the outcomes of “tight” load comparisons is decided uniformly at random. More precisely, in  $g$ -MYOPIC-COMP, at each step  $t = 1, 2, \dots$ , if  $|x_{i_1}^{t-1} - x_{i_2}^{t-1}| \leq g$ , the ball is allocated to a random bin among  $\{i_1, i_2\}$ , otherwise, the ball is allocated to the less loaded among  $\{i_1, i_2\}$ . By using random coin flips, the adversary of  $g$ -ADV-COMP can trivially simulate  $g$ -MYOPIC-COMP.

### Adversarial Delay

**Setting.** Next we turn to adversarial settings where the load information a bin reports may be *outdated*. The first setting, called  $\tau$ -DELAY for  $\tau \in \mathbb{N}_{\geq 1}$ , is similar to  $g$ -ADV-COMP, but here the range of load values the adversary can choose from is based on a sliding time interval. That is, after bins  $i_1$  and  $i_2$  are sampled in step  $t$ , an adaptive adversary provides us with load estimates that must satisfy  $\tilde{x}_{i_1}^{t-1} \in [x_{i_1}^{t-\tau}, x_{i_1}^{t-1}]$  and  $\tilde{x}_{i_2}^{t-1} \in [x_{i_2}^{t-\tau}, x_{i_2}^{t-1}]$ , and the ball is then allocated to the bin with a smaller load estimate, breaking ties arbitrarily. As an example, this includes an adversary that has the power of delaying load updates to each bin arbitrarily, but any load update (i.e., allocation) that is  $\tau$  or more steps in the past must be processed.

**Processes.** A special case of  $\tau$ -DELAY is  $b$ -BATCH for  $b \in \mathbb{N}_{\geq 1}$ , introduced in [18], where balls are allocated in consecutive batches of size  $b$  each. When the load of a bin  $i$  is queried, the bin  $i$  will report the load the bin had at the beginning of the batch (i.e., all allocations within the current batch will not be considered). More formally, when two bins  $i_1, i_2$  are sampled in step  $t$ , the decision where to allocate is based on comparing  $x_{i_1}^{\lfloor (t-1)/b \rfloor \cdot b}$  and  $x_{i_2}^{\lfloor (t-1)/b \rfloor \cdot b}$ , breaking ties randomly. We recall that in [18], only the special case  $b = n$  was studied, and a gap bound of  $O(\log n)$  was shown. Clearly, for  $\tau = b$ ,  $\tau$ -DELAY can simulate  $b$ -BATCH and for  $\tau = b = 1$ , both are equivalent to the TWO-CHOICE process without noise. Note that  $b$ -BATCH assumes that load updates are perfectly synchronized, while the  $\tau$ -DELAY setting relaxes this assumption, therefore covering a wider class of asynchronous update schemes. The reduction of the  $\tau$ -DELAY (and  $b$ -BATCH) settings to a relaxed  $g$ -ADV-COMP setting will be formalized in Section 10.2.



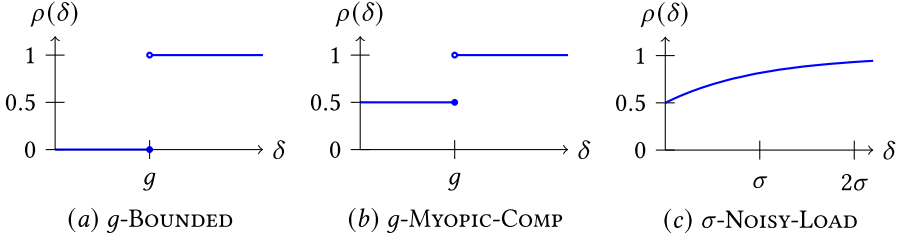


Fig. 2. In the graphs above,  $\delta = |x_{i_1}^{t-1} - x_{i_2}^{t-1}|$  is the load difference of the two sampled bins and  $\rho(\delta)$  is the probability that the load comparison is correct for  $g$ -BOUNDED,  $g$ -MYOPIC-COMP and  $\sigma$ -NOISY-LOAD.

### Probabilistic Noise

**Setting.** We now turn to non-adversarial settings where comparisons are subject to *probabilistic noise*, with a larger chance of getting the correct comparison the larger the load difference. Specifically, we consider the NOISY-COMP setting, where we have an arbitrary non-decreasing function  $\rho : \mathbb{N} \rightarrow [0, 1]$  such that for any  $\delta \in \mathbb{N}$ ,  $\rho(\delta)$  is the probability that a comparison between two bins with absolute load difference  $\delta$  will be correct. As before we assume independence, i.e., for all  $t \geq 1$ , the event of a correct comparison in step  $t$  only depends on the load difference of the two sampled bins. Thus, for a specific function  $\rho$ ,  $\rho$ -NOISY-COMP defines a process.

**Processes.** This is a very expressive setting, encompassing, for example the  $g$ -BOUNDED and  $g$ -MYOPIC-COMP processes by choosing  $\rho$  as step functions with values in  $\{0, 1/2, 1\}$  (see Figure 2). Similarly, Two-CHOICE (without noise), ONE-CHOICE and  $(1 + \beta)$ -process correspond to  $\rho$  being the constant 1,  $1/2$  and  $1/2 + \beta/2$ , respectively.

As a concrete example consider the  $\sigma$ -NOISY-LOAD process. Here, when a bin  $i \in \{i_1, i_2\}$  is sampled at step  $t$ , it reports an unbiased load estimate  $\tilde{x}_i^{t-1} = x_i^{t-1} + Z_i^{t-1}$ , where  $Z_i^{t-1}$  has a normal distribution  $\mathcal{N}(0, \sigma^2)$  (and all  $\{Z_i^t\}_{i \in [n], t \geq 0}$  are mutually independent). Then Two-CHOICE allocates a ball to the bin that reports the smallest load estimate. Thus if  $x_{i_2}^{t-1} - x_{i_1}^{t-1} = \delta > 0$ , the probability to allocate to the lighter bin can be computed as follows:

$$\begin{aligned}
 \Pr \left[ \tilde{x}_{i_1}^{t-1} \leq \tilde{x}_{i_2}^{t-1} \right] &= \Pr \left[ Z_{i_1}^{t-1} - Z_{i_2}^{t-1} \leq \delta \right] \\
 &= 1 - \Pr \left[ \mathcal{N}(0, 2\sigma^2) > \delta \right] \\
 &= 1 - \Pr \left[ \mathcal{N}(0, 1) > \delta / (\sqrt{2}\sigma) \right] \\
 &= 1 - \Phi(\delta / (\sqrt{2}\sigma)).
 \end{aligned}$$

Note that  $\Phi(z) = 1/2$  for  $z = 0$  and  $\Phi(z)$  is increasing in  $z$ . As shown in [37, p. 17], for  $z \geq 0$ ,

$$\frac{1}{\sqrt{2\pi}} \cdot \frac{2}{\sqrt{z^2 + 4} + z} \cdot e^{-z^2/2} \leq 1 - \Phi(z) \leq \frac{1}{\sqrt{2\pi}} \cdot \frac{2}{\sqrt{z^2 + 2} + z} \cdot e^{-z^2/2}.$$

By setting  $z := \delta / (\sqrt{2}\sigma)$ , ignoring the linear term in  $1/z$ , and re-scaling  $\sigma$ , we can define  $\sigma$ -NOISY-LOAD as the process satisfying for all steps  $t$  and samples  $i_1, i_2$  with  $x_{i_2}^{t-1} - x_{i_1}^{t-1} = \delta > 0$ ,

$$\Pr \left[ \tilde{x}_{i_1}^{t-1} \leq \tilde{x}_{i_2}^{t-1} \right] = \rho(\delta) := 1 - \frac{1}{2} \cdot \exp \left( - \left( \frac{\delta}{\sigma} \right)^2 \right), \quad (2.1)$$

meaning that the correct comparison probability exhibits a Gaussian tail behavior.

### 3 NOTATION AND PRELIMINARIES

We sequentially allocate  $m$  balls (jobs) to  $n$  bins (servers) with labels in  $[n] := \{1, 2, \dots, n\}$ . The **load vector** after  $t \geq 0$  steps, i.e.,  $t$  allocations, is  $x^t = (x_1^t, x_2^t, \dots, x_n^t)$  and at the beginning,  $x_i^0 = 0$  for any  $i \in [n]$ . For any  $t \geq 0$ , we also define the normalized load by sorting the entries of  $x^t$  non-increasingly and setting  $y_i^t := x_i^t - \frac{t}{n}$ ; so  $y_1^t \geq y_2^t \geq \dots \geq y_n^t$ . We will analyze the performance of allocation processes via the **gap**, defined for any  $t \geq 0$  as

$$\text{Gap}(t) := \max_{1 \leq i \leq n} x_i^t - \frac{t}{n} = y_1^t,$$

i.e., the difference between maximum and average load after step  $t \geq 0$ . It is well-known that even for Two-CHOICE without noise, the difference between the maximum and minimum load is  $\Omega(\log n)$  for  $m = \Omega(n \log n)$ . Further, for any step  $t \geq 0$ , we define  $B_+^t := \{i \in [n] : y_i^t \geq 0\}$  as the set of **overloaded** bins and  $B_-^t := \{i \in [n] : y_i^t < 0\}$  as the set of **underloaded** bins.

Following [58], many allocation processes can be described by a **probability allocation vector**  $r^t = (r_1^t, r_2^t, \dots, r_n^t)$  for step  $t$ , where  $r_i^t$  is the probability for incrementing the load of the  $i$ th most loaded bin.

Recall the definition of the Two-CHOICE process from Section 2. The Two-CHOICE process without noise has a time-independent probability allocation vector  $p$ , where the probability to allocate to the  $i$ th most loaded bin is given by  $p_i = \frac{2i-1}{n^2}$ , for  $i \in [n]$ . Recall that in ONE-CHOICE, each ball is allocated into a bin sampled independently and uniformly at random. This corresponds to  $r$  being the uniform distribution. We say that a probability vector  $q$  majorizes another probability vector  $r$  if for all  $k \in [n]$ , it satisfies  $\sum_{i=1}^k q_i \geq \sum_{i=1}^k r_i$ . The same definition applies unchanged to sorted load vectors.

For a sequence of random variables  $(X^t)_{t \geq 0}$ , we define  $\Delta X^{t+1} := X^{t+1} - X^t$ . We also use the shorthands  $u^+ := \max\{u, 0\}$  and  $u_1 \wedge u_2 := \min\{u_1, u_2\}$ .

Many statements in this work hold only for sufficiently large  $n$ , and several constants are chosen generously with the intention of making it easier to verify some technical inequalities.

### 4 WARM-UP: UPPER BOUND OF $O(g \log(ng))$ FOR $g$ -ADV-COMP

In this section, we will prove the  $O(g \log(ng))$  gap bound for the  $g$ -ADV-COMP setting (with  $g \geq 1$  being arbitrary), also recovering the  $O(g \log(ng))$  gap bound for the  $g$ -BOUNDED process proven in [56]. The proof relies on a version of the exponential potential function [58], called *hyperbolic cosine potential*  $\Gamma := \Gamma(\gamma)$ , which is defined as

$$\Gamma^t := \Gamma^t(\gamma) := \sum_{i=1}^n \Gamma_i^t = \sum_{i=1}^n \left[ e^{\gamma \cdot y_i^t} + e^{-\gamma \cdot y_i^t} \right], \quad (4.1)$$

for a smoothing parameter  $\gamma \in (0, 1)$ , to be specified in Theorem 4.3. Recall that  $\Delta \Gamma^{t+1} := \Gamma^{t+1} - \Gamma^t$ . We make use of the following general lemma from [58] regarding the expected change of  $\Gamma$ .

**LEMMA 4.1** ([58, LEMMAS 2.1 & 2.3]). *Consider any allocation process with probability allocation vector  $r^t$  and the potential  $\Gamma := \Gamma(\gamma)$  with any  $\gamma \in (0, 1)$ . Then, for any step  $t \geq 0$ ,*

$$\mathbb{E} \left[ \Delta \Gamma^{t+1} \mid y^t \right] \leq \sum_{i=1}^n \left[ \left( r_i^t \cdot (\gamma + \gamma^2) - \left( \frac{\gamma}{n} - \frac{\gamma^2}{n^2} \right) \right) \cdot e^{\gamma y_i^t} + \left( r_i^t \cdot (-\gamma + \gamma^2) + \left( \frac{\gamma}{n} + \frac{\gamma^2}{n^2} \right) \right) \cdot e^{-\gamma y_i^t} \right].$$

For convenience, we rewrite Lemma 4.1 by decomposing the upper bound into the components that are independent of the probability allocation vector  $r^t$  and those that are not, i.e.,

$$\mathbb{E} \left[ \Delta \Gamma^{t+1} \mid y^t \right] \leq h(y^t) + \sum_{i=1}^n r_i^t \cdot f(y_i^t), \quad (4.2)$$

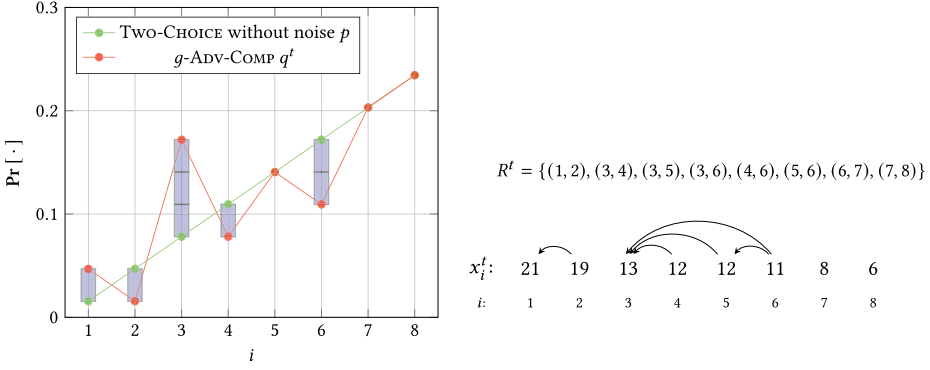


Fig. 3. Illustration of the set  $R^t$  and the change in the probability allocation vector from  $p$  to  $q^t$ , where  $n = 8$  and  $g = 3$ . In the example, each directed arrow moves a probability of  $\frac{2}{n^2}$  (indicated by the blue rectangles) from a bin  $j$  to a heavier bin  $i < j$ . Note that in this example, the adversary decides not to reverse some of the comparisons, e.g., between bins 7 and 8.

where  $h(y^t) := \sum_{i=1}^n -(\frac{\gamma}{n} - \frac{\gamma^2}{n^2}) \cdot e^{\gamma y_i^t} + (\frac{\gamma}{n} + \frac{\gamma^2}{n^2}) \cdot e^{-\gamma y_i^t}$  and  $f(y_i^t) := (\gamma + \gamma^2) \cdot e^{\gamma y_i^t} + (-\gamma + \gamma^2) \cdot e^{-\gamma y_i^t}$ .

We will also make use of the following drop inequality for Two-CHOICE without noise (see also [42, Theorem 3.1]).

LEMMA 4.2 (IMPLIED BY [58, THEOREM 2.9]). *Consider the Two-CHOICE process without noise with probability allocation vector  $p$  and the potential  $\Gamma := \Gamma(\gamma)$  with any  $\gamma \in (0, \frac{1}{6 \cdot 12})$ . Then, there exists a constant  $c > 0$ , such that for any step  $t \geq 0$ ,*

$$\mathbb{E} \left[ \Delta \Gamma^{t+1} \mid y^t \right] \leq h(y^t) + \sum_{i=1}^n p_i \cdot f(y_i^t) \leq -\frac{\gamma}{48n} \cdot \Gamma^t + c.$$

We will analyze  $\Delta \Gamma^{t+1}$  for the  $g$ -ADV-COMP setting by relating it to the change  $\Delta \Gamma^{t+1}$  for the Two-CHOICE process without noise. To this end, it will be helpful to define all pairs of bins (of unequal load), whose comparison is under the control of the adversary:

$$R^t := \{(i, j) \in [n] \times [n] : y_j^t < y_i^t \leq y_j^t + g\}. \quad (4.3)$$

So for each pair  $(i, j) \in R^t$ , the adversary determines the outcome of the load comparison assuming  $\{i, j\}$  are the two bin samples in step  $t + 1$ , which happens with probability  $2 \cdot \frac{1}{n} \cdot \frac{1}{n} = \frac{2}{n^2}$ . This can be seen as moving a probability of up to  $\frac{2}{n^2}$  from bin  $j$  to bin  $i$ , if we relate the probability allocation vector  $p$  (of Two-CHOICE without noise) to the probability allocation vector  $q^t = q^t(\mathcal{F}^t)$  of Two-CHOICE with noise (see Figure 3).

THEOREM 4.3. *Consider the  $g$ -ADV-COMP setting for any  $g \geq 1$ , and the potential  $\Gamma := \Gamma(\gamma)$  with  $\gamma := -\log(1 - \frac{1}{8 \cdot 48})/g < \frac{1}{6 \cdot 12}$ . Then, there exist constants  $c_1 \geq 1, c_2 > 0, c_3 \geq 2$ , such that the following three statements hold for all steps  $t \geq 0$ :*

- (i)  $\mathbb{E} \left[ \Delta \Gamma^{t+1} \mid y^t \right] \leq -\frac{\gamma}{96n} \cdot \Gamma^t + c_1,$
- (ii)  $\mathbb{E} \left[ \Gamma^t \right] \leq c_2 n g,$
- (iii)  $\Pr \left[ \max_{i \in [n]} |y_i^t| \leq c_3 g \log(n g) \right] \geq 1 - (n g)^{-14}.$

PROOF. *First statement.* Consider the probability allocation vector  $q^t$  at step  $t$  in the  $g$ -ADV-COMP setting. By Equation (4.2) we have

$$\mathbb{E} \left[ \Delta \Gamma^{t+1} \mid y^t \right] \leq h(y^t) + \sum_{i=1}^n q_i^t \cdot f(y_i^t).$$

Recall that  $p$  is the probability allocation vector of Two-CHOICE without noise. Then,

$$q^t := p + \sum_{(i,j) \in R^t} (\mathbf{e}_i - \mathbf{e}_j) \cdot \zeta_{i,j}^t + \sum_{(i,j) \in [n] \times [n]: y_i^t = y_j^t} (\mathbf{e}_i - \mathbf{e}_j) \cdot \zeta_{i,j}^t,$$

where  $\mathbf{e}_i \in \mathbb{R}^n$  is the  $i$ th unit vector, and  $\zeta_{i,j}^t$  is a number in  $[0, \frac{2}{n^2}]$ . Hence,

$$\begin{aligned} \mathbb{E} \left[ \Delta \Gamma^{t+1} \mid y^t \right] &\leq h(y^t) + \sum_{i=1}^n p_i \cdot f(y_i^t) + \sum_{(i,j) \in R^t} \zeta_{i,j}^t \cdot (f(y_i^t) - f(y_j^t)) \\ &\quad + \sum_{(i,j) \in [n] \times [n]: y_i^t = y_j^t} \zeta_{i,j}^t \cdot (f(y_i^t) - f(y_j^t)) \\ &= h(y^t) + \sum_{i=1}^n p_i \cdot f(y_i^t) + \sum_{(i,j) \in R^t} \zeta_{i,j}^t \cdot (f(y_i^t) - f(y_j^t)) \\ &\leq -\frac{\gamma}{48n} \cdot \Gamma^t + c + \sum_{(i,j) \in R^t} \zeta_{i,j}^t \cdot (f(y_i^t) - f(y_j^t)), \end{aligned} \quad (4.4)$$

using in the last inequality that by Lemma 4.2 there exists such a constant  $c > 0$  for the Two-CHOICE process without noise and for the same  $\gamma$  (since  $\gamma < \frac{1}{6 \cdot 12}$ ).

For any pair of indices  $(i, j) \in R^t$ , we define

$$\begin{aligned} \xi_{i,j}^t &:= \zeta_{i,j}^t \cdot (f(y_i^t) - f(y_j^t)) \\ &\leq \frac{2}{n^2} \cdot \left( (\gamma + \gamma^2) \cdot e^{\gamma y_i^t} + (-\gamma + \gamma^2) \cdot e^{-\gamma y_i^t} - (\gamma + \gamma^2) \cdot e^{\gamma y_j^t} - (-\gamma + \gamma^2) \cdot e^{-\gamma y_j^t} \right), \end{aligned}$$

and proceed to upper bound  $\xi_{i,j}^t$ , using the following lemma, which is based on a case distinction and Taylor estimates for  $\exp(\cdot)$ .

LEMMA 4.4. *For any pair of indices  $(i, j) \in R^t$ , we have  $\xi_{i,j}^t \leq \frac{\gamma}{96n^2} \cdot (\Gamma_i^t + \Gamma_j^t) + \frac{8}{n^2}$ .*

PROOF OF LEMMA 4.4. Recall that for any  $(i, j) \in R^t$  we have that  $y_j^t < y_i^t \leq y_j^t + g$ . So, now we consider the following three disjoint cases:

**Case 1**  $[y_i^t > g]$ : In this case, we also have that  $y_j^t > 0$ , so

$$\begin{aligned} \xi_{i,j}^t &\stackrel{(a)}{\leq} \frac{2}{n^2} \cdot \left( (\gamma + \gamma^2) \cdot e^{\gamma y_i^t} - (\gamma + \gamma^2) \cdot e^{\gamma y_j^t} + 1 \right) \\ &\stackrel{(b)}{\leq} \frac{2}{n^2} \cdot \left( (\gamma + \gamma^2) \cdot e^{\gamma y_i^t} - (\gamma + \gamma^2) \cdot e^{\gamma(y_i^t - g)} + 1 \right) \\ &= \frac{2}{n^2} \cdot \left( (\gamma + \gamma^2) \cdot e^{\gamma y_i^t} \cdot (1 - e^{-\gamma g}) + 1 \right) \\ &\leq \frac{2}{n^2} \cdot \left( (\gamma + \gamma^2) \cdot \Gamma_i^t \cdot (1 - e^{-\gamma g}) + 1 \right), \end{aligned}$$

using in (a) that  $0 \leq \gamma - \gamma^2 \leq 1$  and  $e^{-\gamma y_k^t} \leq 1$  for any bin  $k \in [n]$  with  $y_k^t > 0$ , and in (b) that  $y_i^t \leq y_j^t + g$ .

**Case 2** [ $y_j^t < -g$ ]: In this case, we also have that  $y_i^t < 0$ ,

$$\begin{aligned}\xi_{i,j}^t &\stackrel{(a)}{\leq} \frac{2}{n^2} \cdot \left( (-\gamma + \gamma^2) \cdot e^{-\gamma y_i^t} - (-\gamma + \gamma^2) \cdot e^{-\gamma y_j^t} + 1 \right) \\ &\stackrel{(b)}{\leq} \frac{2}{n^2} \cdot \left( (-\gamma + \gamma^2) \cdot e^{-\gamma(y_j^t+g)} - (-\gamma + \gamma^2) \cdot e^{-\gamma y_j^t} + 1 \right) \\ &= \frac{2}{n^2} \cdot \left( (-\gamma + \gamma^2) \cdot e^{-\gamma y_j^t} \cdot (e^{-\gamma g} - 1) + 1 \right) \\ &\leq \frac{2}{n^2} \cdot \left( (\gamma - \gamma^2) \cdot \Gamma_j^t \cdot (1 - e^{-\gamma g}) + 1 \right),\end{aligned}$$

using in (a) that  $\gamma + \gamma^2 \leq 1$  and that  $e^{\gamma y_k^t} \leq 1$  for any bin  $k \in [n]$  with  $y_k^t < 0$  and in (b) that  $y_i^t \leq y_j^t + g$  and  $\gamma^2 \leq \gamma$ .

**Case 3** [ $\max\{|y_i^t|, |y_j^t|\} \leq g$ ]: In this case, we have that

$$\xi_{i,j}^t \leq \frac{2}{n^2} \cdot (\gamma + \gamma^2) \cdot (\Gamma_i^t + \Gamma_j^t) \leq \frac{2}{n^2} \cdot \frac{1}{2} \cdot (2 \cdot 2 \cdot e^{\gamma g}) \leq \frac{2}{n^2} \cdot 4,$$

using that  $\gamma + \gamma^2 \leq \frac{1}{2}$  (since  $\gamma < \frac{1}{6 \cdot 12}$ ) and by the definition of  $\gamma$ , we have  $e^{\gamma g} = e^{-\log(1 - \frac{1}{8 \cdot 48})} < 2$ .

Combining the upper bounds for the three cases, we have that for any  $(i, j) \in R^t$ ,

$$\begin{aligned}\xi_{i,j}^t &\leq \frac{2}{n^2} \cdot \left( (\gamma + \gamma^2) \cdot \Gamma_i^t \cdot (1 - e^{-\gamma g}) + (\gamma - \gamma^2) \cdot \Gamma_j^t \cdot (1 - e^{-\gamma g}) + 4 \right) \\ &\stackrel{(a)}{\leq} \frac{2}{n^2} \cdot \left( 2\gamma \cdot \Gamma_i^t \cdot (1 - e^{-\gamma g}) + 2\gamma \cdot \Gamma_j^t \cdot (1 - e^{-\gamma g}) + 4 \right) \\ &= \frac{4\gamma}{n^2} \cdot (\Gamma_i^t + \Gamma_j^t) \cdot (1 - e^{-\gamma g}) + \frac{8}{n^2} \\ &\stackrel{(b)}{=} \frac{4\gamma}{n^2} \cdot (\Gamma_i^t + \Gamma_j^t) \cdot \frac{1}{8 \cdot 48} + \frac{8}{n^2} \\ &= \frac{\gamma}{96n^2} \cdot (\Gamma_i^t + \Gamma_j^t) + \frac{8}{n^2},\end{aligned}$$

using in (a) that  $\gamma^2 \leq \gamma$  since  $\gamma \leq 1$  and in (b) that  $\gamma = -\log(1 - \frac{1}{8 \cdot 48})/g$ .  $\square$

We continue with the proof of Theorem 4.3. By Lemma 4.4 and Equation (4.4), we have

$$\begin{aligned}\mathbb{E} \left[ \Delta \Gamma^{t+1} \mid y^t \right] &\leq -\frac{\gamma}{48n} \cdot \Gamma^t + c + \sum_{(i,j) \in R^t} \left( \frac{\gamma}{96n^2} \cdot (\Gamma_i^t + \Gamma_j^t) + \frac{8}{n^2} \right) \\ &\stackrel{(a)}{\leq} -\frac{\gamma}{48n} \cdot \Gamma^t + c + \frac{\gamma}{96n} \cdot \Gamma^t + 4 \\ &= -\frac{\gamma}{96n} \cdot \Gamma^t + c_1,\end{aligned} \tag{4.5}$$

for  $c_1 := c + 4 \geq 1$ , where (a) holds since if  $(i, j) \in R^t$  then  $(j, i) \notin R^t$ , so every bin  $k \in [n]$  appears at most  $n$  times in  $R^t$ . This concludes the proof of the first statement.

*Second statement.* By Lemma A.11 (ii) (for  $a = 1 - \frac{\gamma}{96n}$  and  $b = c_1$ ), since  $\Gamma^0 = 2n \leq \frac{96c_1}{\gamma} \cdot n$  (as  $c_1 \geq 1$  and  $\gamma \leq 1$ ) and Equation (4.5) hold, it follows that

$$\mathbb{E} \left[ \Gamma^t \right] \leq \frac{96c_1}{\gamma} \cdot n =: c_2 n g.$$

*Third statement.* Using Markov's inequality, for any step  $t \geq 0$ ,

$$\Pr \left[ \Gamma^t \leq c_2 \cdot (ng)^{15} \right] \geq 1 - (ng)^{-14}.$$

When the event  $\{\Gamma^t \leq c_2 \cdot (ng)^{15}\}$  holds, we have that

$$\max_{i \in [n]} |y_i^t| \leq \frac{1}{\gamma} \cdot (\log c_2 + 15 \log(ng)) \leq \frac{16 \log(ng)}{\gamma} =: c_3 g \log(ng),$$

for sufficiently large  $n$  and for the constant

$$c_3 := \frac{16}{\gamma g} = \frac{16}{-\log(1 - \frac{1}{8 \cdot 48})} \geq 2. \quad (4.6)$$

Therefore we conclude that,

$$\Pr \left[ \max_{i \in [n]} |y_i^t| \leq c_3 g \log(ng) \right] \geq 1 - (ng)^{-14}. \quad \square$$

Next, we will also state a simple corollary that starting with a “small” gap, in any future step, w.h.p. the gap will be small. This corollary will be used in obtaining the tighter  $O(g + \log n)$  gap bound in Section 5. We defer its proof to Appendix B.1.

**COROLLARY 4.5.** *Consider the  $g$ -ADV-COMP setting for any  $g \geq 1$ . Then, for any steps  $t_0 \geq 0$  and  $t_1 \geq t_0$ , we have that*

$$\Pr \left[ \max_{i \in [n]} |y_i^{t_1}| \leq 2g(\log(ng))^2 \mid \mathfrak{F}^{t_0}, \max_{i \in [n]} |y_i^{t_0}| \leq g(\log(ng))^2 \right] \geq 1 - (ng)^{14}.$$

## 5 UPPER BOUND OF $O(g + \log n)$ FOR $g$ -ADV-COMP

In this section, we give the proof of the  $O(g + \log n)$  gap bound, as stated in the theorem below. For  $g = \Omega(\log n)$ , this matches the lower bound for the  $g$ -MYOPIC-COMP process in Proposition 11.2 up to multiplicative constants.

**THEOREM 5.12 (SIMPLIFIED VERSION, PAGE 27).** *Consider the  $g$ -ADV-COMP setting for any  $g \geq 1$ . Then, there exists a constant  $\kappa > 0$ , such that for any step  $m \geq 0$ ,*

$$\Pr \left[ \max_{i \in [n]} |y_i^m| \leq \kappa \cdot (g + \log n) \right] \geq 1 - 2 \cdot (ng)^{-9}.$$

### 5.1 Proof Outline of Theorem 5.12

The proof of this theorem is considerably more involved than that of Theorem 4.3, making use of an interplay between a variant of the hyperbolic cosine, the absolute value, and the quadratic potential (to be defined below).

First, we define the *hyperbolic cosine potential with an offset*, denoted by  $\Lambda$ , as

$$\Lambda^t := \Lambda^t(\alpha, c_4 g) := \sum_{i=1}^n \Lambda_i^t := \sum_{i=1}^n \left[ e^{\alpha \cdot (y_i^t - c_4 g)^+} + e^{\alpha \cdot (-y_i^t - c_4 g)^+} \right], \quad (5.1)$$

for smoothing parameter  $\alpha := \frac{1}{18}$  and offset  $c_4 g$  with  $c_4 := 730$ , and recalling that  $u^+ := \max\{u, 0\}$ . Compared to  $\Gamma$  used in Section 4, the potential  $\Lambda$  has a larger, i.e., constant smoothing parameter  $\alpha$  at the cost of an offset of  $\Theta(g)$ . This means that if at some step  $t \geq 0$ ,  $\Lambda^t = O(n)$  holds, then we can deduce a stronger upper bound of  $\text{Gap}(t) = O(g + \log n)$ .

However, we are not able to show that  $\Lambda$  drops in expectation in every step. More specifically, we only show that  $\Lambda^t$  drops in expectation at step  $t$  when the *absolute value potential*  $\Delta^t := \sum_{i=1}^n |y_i^t|$  satisfies the condition  $\Delta^t \leq Dng$  for  $D := 365$ . This condition implies that there is at most a constant fraction of bins  $i \in [n]$  with  $|y_i^t| \geq \frac{3}{2}Dg$  and hence there is a bias to place away from bins  $j \in [n]$  with  $|y_j^t| \geq 2Dg = c_4 g$ , i.e., the ones with load above the offset of  $\Lambda$ . More specifically, in



such *good* steps  $t$  we show that  $\Delta^t$  satisfies a drop inequality, meaning that when large it drops by a multiplicative factor over the next step in expectation.

LEMMA 5.7 (SIMPLIFIED VERSION, PAGE 20). *Consider the  $g$ -ADV-COMP setting for any  $g \geq 1$  and let  $\varepsilon := \frac{1}{12}$ . Then, for any step  $t \geq 0$ ,*

$$\mathbb{E} \left[ \Delta^{t+1} \mid \mathfrak{F}^t, \Delta^t \leq Dng \right] \leq \Delta^t \cdot \left( 1 - \frac{2\alpha\varepsilon}{n} \right) + 18\alpha.$$

In order to show that there are many steps with  $\Delta^t \leq Dng$ , we relate the expected change of the *quadratic potential*  $\Upsilon^t := \sum_{i=1}^n (y_i^t)^2$  to  $\Delta^t$ . More specifically, by upper bounding the contribution of each pair  $(i, j) \in R^t$  (defined in Equation (4.3)), similarly to the proof of Theorem 4.3, we show that:

LEMMA 5.3 (RESTATED, PAGE 17). *Consider the  $g$ -ADV-COMP setting for any  $g \geq 1$ . Then, for any step  $t \geq 0$ ,*

$$\mathbb{E} \left[ \Delta \Upsilon^{t+1} \mid y^t \right] \leq -\frac{\Delta^t}{n} + 2g + 1.$$

This allows us to prove that in any sufficiently long interval, there is a large constant fraction of good steps, i.e., with  $\Delta^t \leq Dng$  (Section 5.3). Thus using an *adjusted exponential function* based on  $\Delta$  (see Section 5.5 for the definition), we show that w.h.p.  $\Lambda^s = O(n)$  for some  $s \in [m - \Theta(n \cdot (\log(n \cdot g))^2), m]$  (*recovery phase*) and then w.h.p.  $\Delta^t = O(n)$  once every  $O(n \cdot (g + \log n))$  steps (*stabilization phase*). This implies that w.h.p.  $\text{Gap}(m) = O(g + \log n)$ . In more detail:

- (1) **(Base case)** For any  $t \geq 0$ , w.h.p.  $\text{Gap}(t) = O(g \log(n \cdot g))$ . (This follows from Theorem 4.3)
- (2) **(Recovery)** Starting with  $\text{Gap}(t_0) = O(g \log(n \cdot g))$  for  $t_0 = m - \Theta(n \cdot (\log(n \cdot g))^2)$ , w.h.p. there exists a step  $t_1 \in [t_0, m]$  such that  $\Delta^{t_1} = O(n)$  (Lemma 5.9).
- (3) **(Stabilization)** Starting with  $\Delta^{t_1} = O(n)$ , w.h.p. there exists a step  $s \in (t_1, t_1 + \Theta(n \cdot (g + \log n)))$  such that  $\Delta^s = O(n)$  (Lemma 5.10).
- (4) **(Gap deduction)** Hence w.h.p. there is a step  $t \in [m, m + \Theta(n \cdot (g + \log n))]$  with  $\text{Gap}(t) = O(g + \log n)$ , which by smoothness of the gap, implies  $\text{Gap}(m) = O(g + \log n)$  (Section 5.7).

## 5.2 Absolute Value and Quadratic Potentials

Recall that the *absolute value potential* is defined as

$$\Delta^t := \sum_{i=1}^n |y_i^t|, \quad (5.2)$$

and the *quadratic potential* is defined as

$$\Upsilon^t := \sum_{i=1}^n (y_i^t)^2. \quad (5.3)$$

We will upper bound the expected change  $\mathbb{E}[\Delta \Upsilon^{t+1} \mid y^t]$  in the  $g$ -ADV-COMP setting by relating it to the change of the quadratic potential for Two-CHOICE, starting with the same load vector  $y^t$  at step  $t$ .

We will first analyze the expected change of the quadratic potential for the Two-CHOICE process without noise. We will make use of the following general lemma (proven in Lemma B.2), which provides a formula for the change of the quadratic potential:

LEMMA 5.1. Consider any allocation process with probability allocation vector  $r^t$  at step  $t \geq 0$ . Then, for any step  $t \geq 0$ , (i) it holds that

$$\mathbb{E} \left[ \Delta Y^{t+1} \mid y^t \right] = \sum_{i=1}^n 2 \cdot r_i^t \cdot y_i^t + 1 - \frac{1}{n} \leq \sum_{i=1}^n 2 \cdot r_i^t \cdot y_i^t + 1,$$

and (ii) it holds that

$$\left| \Delta Y^{t+1} \right| \leq 4 \cdot \max_{i \in [n]} |y_i^t| + 2.$$

We now use the general formula in Lemma 5.1 (i) to obtain an expression for the expected change of the quadratic potential for Two-CHOICE without noise.

LEMMA 5.2 (CF. [46, LEMMA 6.2]). Consider the Two-CHOICE process without noise with probability allocation vector  $p$ . Then, it holds that for any step  $t \geq 0$ ,

$$\mathbb{E} \left[ \Delta Y^{t+1} \mid y^t \right] \leq \sum_{i=1}^n 2 \cdot p_i \cdot y_i^t + 1 \leq -\frac{\Delta^t}{n} + 1.$$

PROOF. Applying Lemma 5.1 (i) to the probability allocation vector  $p$  yields

$$\mathbb{E} \left[ \Delta Y^{t+1} \mid y^t \right] \leq \sum_{i=1}^n 2 \cdot p_i \cdot y_i^t + 1.$$

Recall that  $B_+^t := \{i \in [n] : y_i^t \geq 0\}$  is the set of overloaded bins at step  $t$  and  $B_-^t := \{i \in [n] : y_i^t < 0\}$ , the set of underloaded bins. The Two-CHOICE process allocates a ball into the set of overloaded bins with probability  $|B_+^t|^2/n^2$ , and thus the average allocation probability across overloaded bins is  $p_+^t = |B_+^t|/n^2$ . Consequently, Two-CHOICE allocates to the set of underloaded bins with probability  $1 - |B_+^t|^2/n^2$ , and thus the average allocation probability across underloaded bins is

$$p_-^t = \frac{1}{|B_-^t|} \cdot \left( 1 - \frac{|B_+^t|^2}{n^2} \right) = \frac{1}{n - |B_+^t|} \cdot \frac{(n + |B_+^t|) \cdot (n - |B_+^t|)}{n^2} = \frac{1}{n} + \frac{|B_+^t|}{n^2}.$$

By splitting the sum  $\sum_{i=1}^n 2 \cdot p_i \cdot y_i^t$  into underloaded and overloaded bins, we get

$$\sum_{i=1}^n 2 \cdot p_i \cdot y_i^t = \sum_{i \in B_+^t} 2 \cdot p_i \cdot y_i^t + \sum_{i \in B_-^t} 2 \cdot p_i \cdot y_i^t.$$

Since  $p_i$  is non-decreasing, we have  $\sum_{i=1}^j p_i \leq \sum_{i=1}^j p_+^t$  for all  $1 \leq j \leq |B_+^t|$ . Further, since  $y_i^t$  is non-increasing over the overloaded bins, by Lemma A.12 we have

$$\sum_{i \in B_+^t} 2 \cdot p_i \cdot y_i^t \leq \sum_{i \in B_+^t} 2 \cdot p_+^t \cdot y_i^t = 2 \cdot p_+^t \cdot \sum_{i \in B_+^t} y_i^t = \frac{|B_+^t|}{n^2} \cdot \Delta^t,$$

since  $\sum_{i \in B_-^t} y_i^t = -\sum_{i \in B_+^t} y_i^t$  and thus  $\sum_{i \in B_+^t} y_i^t = \frac{1}{2} \Delta^t$ . Analogously, since  $y_i^t$  is non-increasing over the underloaded bins,

$$\sum_{i \in B_-^t} 2 \cdot p_i \cdot y_i^t \leq \sum_{i \in B_-^t} 2 \cdot p_-^t \cdot y_i^t = 2 \cdot p_-^t \cdot \sum_{i \in B_-^t} y_i^t = -\left( \frac{1}{n} + \frac{|B_+^t|}{n^2} \right) \cdot \Delta^t.$$

Combining these we get

$$\mathbb{E} \left[ \Delta Y^{t+1} \mid y^t \right] \leq \sum_{i=1}^n 2 \cdot p_i \cdot y_i^t + 1 \leq \frac{|B_+^t|}{n^2} \cdot \Delta^t - \left( \frac{1}{n} + \frac{|B_+^t|}{n^2} \right) \cdot \Delta^t + 1 = -\frac{\Delta^t}{n} + 1. \quad \square$$

Now we relate the change of the quadratic potential for the  $g$ -ADV-COMP setting to the change of the quadratic potential for Two-CHOICE without noise, using that the adversary can determine (and possibly revert) a load comparison between  $y_i^t$  and  $y_j^t$  only if  $|y_i^t - y_j^t| \leq g$ .

LEMMA 5.3. *Consider the  $g$ -ADV-COMP setting for any  $g \geq 1$ . Then, for any step  $t \geq 0$ ,*

$$\mathbb{E} \left[ \Delta Y^{t+1} \mid y^t \right] \leq -\frac{\Delta^t}{n} + 2g + 1.$$

PROOF. By Lemma 5.1 (i), for the  $g$ -ADV-COMP probability allocation vector  $q^t$  we have,

$$\mathbb{E} \left[ \Delta Y^{t+1} \mid y^t \right] \leq \sum_{i=1}^n 2 \cdot q_i^t \cdot y_i^t + 1.$$

This probability vector  $q^t$  is obtained from the probability allocation vector  $p$  of Two-CHOICE without noise by moving a probability of up to  $\frac{2}{n^2}$  from any bin  $j$  to a bin  $i$  with  $y_j^t < y_i^t \leq y_j^t + g$ . Recalling that  $R^t := \{(i, j) \in [n] \times [n] : y_j^t < y_i^t \leq y_j^t + g\}$ ,

$$\begin{aligned} \mathbb{E} \left[ \Delta Y^{t+1} \mid y^t \right] &\leq \sum_{i=1}^n 2 \cdot p_i \cdot y_i^t + 1 + 2 \cdot \sum_{(i,j) \in R^t} \frac{2}{n^2} \cdot (y_i^t - y_j^t) \\ &\leq \sum_{i=1}^n 2 \cdot p_i \cdot y_i^t + 1 + 2 \cdot \sum_{(i,j) \in R^t} \frac{2}{n^2} \cdot g \\ &\leq \sum_{i=1}^n 2 \cdot p_i \cdot y_i^t + 1 + 2g, \end{aligned}$$

using that  $|R^t| < \frac{1}{2}n^2$ . Hence, using Lemma 5.2, we conclude that

$$\mathbb{E} \left[ \Delta Y^{t+1} \mid y^t \right] \leq -\frac{\Delta^t}{n} + 2g + 1. \quad \square$$

### 5.3 Constant Fraction of Good Steps

We define a step  $s \geq 0$  to be a *good step* if  $\mathcal{G}^s := \{\Delta^s \leq Dng\}$  holds, for  $D := 365$ . Further,  $G_{t_0}^{t_1} := G_{t_0}^{t_1}(D)$  denotes the number of good steps in  $[t_0, t_1]$ . Later, in Section 5.4 we will show that in a good step, the exponential potential  $\Lambda$  with any sufficiently small constant  $\alpha$  drops in expectation.

In the following lemma, we show that at least a constant fraction  $r$  of the steps are good in a sufficiently long interval. We will apply this lemma with two different values for  $T$ : (i) in the recovery phase, to prove that there exists a step  $s \in [m - \Theta(ng \cdot (\log(ng))^2), m]$  with  $\Lambda^s = O(n)$  and (ii) in the stabilization phase, to prove that every  $O(n \cdot (g + \log n))$  steps there exists a step  $s$  with  $\Lambda^s = O(n)$ . In the analysis below, we pick  $r := \frac{6}{6+\varepsilon}$ , where  $\varepsilon := \frac{1}{12}$ .

LEMMA 5.4. *Consider the  $g$ -ADV-COMP setting for any  $g \geq 1$  and let  $r := \frac{6}{6+\varepsilon}$  and  $D := 365$ . Then, for any constant  $\hat{c} \geq 1$  and any  $T \in [ng^2, n^2g^3/\hat{c}]$ , we have for any steps  $t_0 \geq 0$  and  $t_1 := t_0 + \hat{c} \cdot T \cdot g^{-1} - 1$ ,*

$$\Pr \left[ G_{t_0}^{t_1}(D) \geq r \cdot (t_1 - t_0 + 1) \mid \tilde{\gamma}^{t_0}, \Upsilon^{t_0} \leq T, \max_{i \in [n]} |y_i^{t_0}| \leq g(\log(ng))^2 \right] \geq 1 - 2 \cdot (ng)^{-12}.$$

PROOF. We define the sequence  $(Z^t)_{t \geq t_0}$  with  $Z^{t_0} := \Upsilon^{t_0}$  and for any  $t > t_0$ ,

$$Z^t := \Upsilon^t + \sum_{s=t_0}^{t-1} \left( \frac{\Delta^s}{n} - 2g - 1 \right).$$

This sequence forms a super-martingale since by Lemma 5.3,

$$\begin{aligned} \mathbf{E} \left[ Z^{t+1} \mid \mathcal{F}^t \right] &= \mathbf{E} \left[ Y^{t+1} + \sum_{s=t_0}^t \left( \frac{\Delta^s}{n} - 2g - 1 \right) \mid \mathcal{F}^t \right] \\ &\leq Y^t - \frac{\Delta^t}{n} + 2g + 1 + \sum_{s=t_0}^t \left( \frac{\Delta^s}{n} - 2g - 1 \right) \\ &= Y^t + \sum_{s=t_0}^{t-1} \left( \frac{\Delta^s}{n} - 2g - 1 \right) = Z^t. \end{aligned}$$

Further, let  $\tau := \inf\{t \geq t_0 : \max_{i \in [n]} |y_i^t| > 2g(\log(ng))^2\}$  and consider the stopped random variable

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

which is then also a super-martingale. Applying Corollary 4.5 and the union bound over steps  $[t_0, t_1]$ , we get

$$\Pr \left[ \tau \leq t_1 \mid \mathcal{F}^{t_0}, Y^{t_0} \leq T, \max_{i \in [n]} |y_i^{t_0}| \leq g(\log(ng))^2 \right] \leq (\hat{c} \cdot T \cdot g^{-1}) \cdot (ng)^{-14} \leq (ng)^{-12}, \quad (5.4)$$

using that  $T \leq n^2 g^3 / \hat{c}$ . This means that the maximum absolute normalized load does not increase above  $2g(\log(ng))^2$  in any of the steps in  $[t_0, t_1]$  w.h.p.

To prove concentration of  $\tilde{Z}^{t+1}$ , we will now derive an upper bound on the difference  $|\tilde{Z}^{t+1} - \tilde{Z}^t|$ :

**Case 1**  $[t \geq \tau]$ : In this case,  $\tilde{Z}^{t+1} = Z^{(t+1) \wedge \tau} = Z^\tau$ , and similarly,  $\tilde{Z}^t = Z^{t \wedge \tau} = Z^\tau$ , so  $|\tilde{Z}^{t+1} - \tilde{Z}^t| = 0$ .

**Case 2**  $[t < \tau]$ : In this case, we have  $\max_{i \in [n]} |y_i^t| \leq 2g(\log(ng))^2$  and by Lemma 5.1 (ii), we have that  $|\Delta Y^{t+1}| \leq 8c_3 g(\log(ng))^2 + 2$ . This implies that

$$|\tilde{Z}^{t+1} - \tilde{Z}^t| \leq |\Delta Y^{t+1}| + \left| \frac{\Delta^t}{n} - 2g - 1 \right| \leq 8g(\log(ng))^2 + 2 + (2g(\log(ng))^2 - 2g - 1) \leq 10g(\log(ng))^2.$$

Combining the two cases above, we conclude that for all  $t \geq t_0$ ,

$$|\tilde{Z}^{t+1} - \tilde{Z}^t| \leq 10g(\log(ng))^2.$$

Using Azuma's inequality for super-martingales (Lemma A.1) for  $\lambda = T$  and  $a_i = 10g(\log(ng))^2$ ,

$$\begin{aligned} \Pr \left[ \tilde{Z}^{t_1+1} - \tilde{Z}^{t_0} \geq T \mid \mathcal{F}^{t_0}, Y^{t_0} \leq T, \max_{i \in [n]} |y_i^{t_0}| \leq g(\log(ng))^2 \right] &\leq \exp \left( -\frac{T^2}{2 \cdot \sum_{t=t_0}^{t_1} (10g(\log(ng))^2)^2} \right) \\ &= \exp \left( -\frac{T^2}{\hat{c} \cdot T \cdot g^{-1} \cdot 200 \cdot g^2 \cdot (\log(ng))^4} \right) \\ &= \exp \left( -\frac{T}{200 \cdot \hat{c} \cdot g \cdot (\log(ng))^4} \right) \\ &\stackrel{(a)}{\leq} \exp \left( -\frac{ng^2}{200 \cdot \hat{c} \cdot g \cdot (\log(ng))^4} \right) = (ng)^{-\omega(1)}, \end{aligned}$$

where in (a) we used that  $T \geq ng^2$ . Hence, we conclude that

$$\Pr \left[ \tilde{Z}^{t_1+1} < \tilde{Z}^{t_0} + T \mid \mathcal{F}^{t_0}, \Upsilon^{t_0} \leq T, \max_{i \in [n]} |y_i^{t_0}| \leq g(\log(ng))^2 \right] \geq 1 - (ng)^{-\omega(1)}.$$

Thus by taking the union bound with Equation (5.4) we have

$$\Pr \left[ Z^{t_1+1} < Z^{t_0} + T \mid \mathcal{F}^{t_0}, \Upsilon^{t_0} \leq T, \max_{i \in [n]} |y_i^{t_0}| \leq g(\log(ng))^2 \right] \geq 1 - 2 \cdot (ng)^{-12}.$$

For the sake of a contradiction, assume now that more than an  $(1-r)$  fraction of the steps  $t \in [t_0, t_1]$  satisfy  $\Delta^t > Dng$ . This implies that

$$\sum_{t=t_0}^{t_1} \frac{\Delta^t}{n} > Dg \cdot (1-r) \cdot (t_1 - t_0 + 1) = D \cdot (1-r) \cdot \hat{c} \cdot T, \quad (5.5)$$

using that  $t_1 - t_0 + 1 = \hat{c} \cdot T \cdot g^{-1}$ . When  $\{Z^{t_1+1} < Z^{t_0} + T\}$  and  $\{\Upsilon^{t_0} \leq T\}$  hold, then we have

$$\Upsilon^{t_1+1} + \sum_{t=t_0}^{t_1} \frac{\Delta^t}{n} - (2g+1) \cdot (t_1 - t_0 + 1) < \Upsilon^{t_0} + T \leq 2T.$$

By rearranging this leads to a contradiction as

$$\begin{aligned} 0 &\leq \Upsilon^{t_1+1} < 2T - \sum_{t=t_0}^{t_1} \frac{\Delta^t}{n} + (2g+1) \cdot (t_1 - t_0 + 1) \\ &\stackrel{(a)}{=} 2T - \sum_{t=t_0}^{t_1} \frac{\Delta^t}{n} + \hat{c}(2g+1) \cdot T \cdot g^{-1} \\ &\stackrel{(b)}{\leq} - \sum_{t=t_0}^{t_1} \frac{\Delta^t}{n} + 5\hat{c} \cdot T \\ &\stackrel{(5.5)}{<} -D \cdot (1-r) \cdot \hat{c} \cdot T + 5\hat{c} \cdot T \\ &\stackrel{(c)}{=} 0, \end{aligned}$$

using in (a) that  $t_1 - t_0 + 1 = \hat{c} \cdot T \cdot g^{-1}$ , in (b) that  $\hat{c} \geq 1$  and  $g \geq 1$ , and in (c) that  $D = \frac{5}{1-r} = 365$  (as  $r = \frac{6}{6+1/12}$ ).

We conclude that when  $\{Z^{t_1+1} < Z^{t_0} + T\}$  holds, then at least an  $r$  fraction of the steps  $t \in [t_0, t_1]$  satisfy  $\Delta^t \leq Dng$ , and thus,

$$\Pr \left[ G_{t_0}^{t_1} \geq r \cdot (t_1 - t_0 + 1) \mid \mathcal{F}^{t_0}, \Upsilon^{t_0} \leq T, \max_{i \in [n]} |y_i^{t_0}| \leq g(\log(ng))^2 \right] \geq 1 - 2 \cdot (ng)^{-12}. \quad \square$$

The following lemma provides two ways of upper bounding the quadratic potential using the exponential potential  $\Lambda$ . These will be used in the recovery and stabilization lemmas, to obtain the starting point condition ( $\Upsilon^{t_0} \leq T$ ) for Lemma 5.4. We defer its proof to Appendix B.2.

**LEMMA 5.5.** *Consider the potential  $\Lambda := \Lambda(\alpha, c_4 g)$  for any constant  $\alpha \in (0, 1)$ , any  $g \geq 1$  and any constant  $c_4 > 0$ . Then (i) for any constant  $\hat{c} > 0$ , there exists a constant  $c_s := c_s(\alpha, c_4, \hat{c}) \geq 1$ , such that for any step  $t \geq 0$  with  $\Lambda^t \leq \hat{c} \cdot n$ ,*

$$\Upsilon^t \leq c_s \cdot ng^2.$$

Furthermore, (ii) there exists a constant  $c_r := c_r(\alpha, c_4) \geq 1$ , such that for any step  $t \geq 0$ ,

$$\Upsilon^t \leq c_r \cdot n \cdot \left( g^2 + (\log \Lambda^t)^2 \right).$$

## 5.4 Exponential Potential

We now prove bounds on the expected change of the  $\Lambda$  potential function over one step. Note that these hold for any sufficiently small constant  $\alpha > 0$ .

We start with a relatively weak bound which holds at any step.

**LEMMA 5.6.** *Consider any allocation process with probability allocation vector  $r^t$  satisfying that  $\max_{i \in [n]} r_i^t \leq \frac{2}{n}$ . Furthermore, consider the potential  $\Lambda := \Lambda(\alpha, c_4 g)$  for any  $\alpha \in (0, \frac{1}{2}]$ , any  $g \geq 1$  and any  $c_4 > 0$ . Then, for any step  $t \geq 0$ , (i) for every bin  $i \in [n]$  it holds that*

$$\mathbb{E} \left[ \Lambda_i^{t+1} \mid \mathfrak{F}^t \right] \leq \Lambda_i^t \cdot \left( 1 + \frac{3\alpha}{n} \right).$$

Furthermore, by aggregating over all bins, (ii) it holds that

$$\mathbb{E} \left[ \Lambda^{t+1} \mid \mathfrak{F}^t \right] \leq \Lambda^t \cdot \left( 1 + \frac{3\alpha}{n} \right).$$

**PROOF.** Consider an arbitrary bin  $i \in [n]$ . We upper bound the change of the overloaded component, i.e.,  $e^{\alpha \cdot (y_i^t - c_4 g)^+}$ , by placing one ball in bin  $i \in [n]$  with probability  $r_i^t \leq \frac{2}{n}$  and ignoring the change of the average load. Also, we upper bound the change for the underloaded component, i.e.,  $e^{\alpha \cdot (-y_i^t - c_4 g)^+}$ , by considering only the change of the average load by  $1/n$ . Hence,

$$\begin{aligned} \mathbb{E} \left[ \Lambda_i^{t+1} \mid \mathfrak{F}^t \right] &\leq e^{\alpha \cdot (y_i^t - c_4 g)^+} \cdot ((1 - r_i^t) + r_i^t \cdot e^\alpha) + e^{\alpha \cdot (-y_i^t - c_4 g)^+} \cdot ((1 - r_i^t) \cdot e^{\alpha/n} + r_i^t \cdot e^{\alpha/n}) \\ &= e^{\alpha \cdot (y_i^t - c_4 g)^+} \cdot (1 + r_i^t \cdot (e^\alpha - 1)) + e^{\alpha \cdot (-y_i^t - c_4 g)^+} \cdot e^{\alpha/n} \\ &\stackrel{(a)}{\leq} e^{\alpha \cdot (y_i^t - c_4 g)^+} \cdot \left( 1 + \frac{2}{n} \cdot 1.5\alpha \right) + e^{\alpha \cdot (-y_i^t - c_4 g)^+} \cdot \left( 1 + 1.5 \frac{\alpha}{n} \right) \\ &\leq \Lambda_i^t \cdot \left( 1 + \frac{3\alpha}{n} \right), \end{aligned}$$

using in (a) that  $e^u \leq 1 + 1.5u$  (for any  $0 \leq u \leq 0.7$ ),  $\alpha \leq 1/2$  and  $r_i^t \leq \frac{2}{n}$ .  $\square$

Now we improve this bound for any good step  $t$ , i.e., when  $\Delta^t \leq Dng$  holds.

**LEMMA 5.7.** *Consider the  $g$ -ADV-COMP setting for any  $g \geq 1$  and the potential  $\Lambda := \Lambda(\alpha, c_4 g)$  for any  $\alpha \in (0, \frac{1}{18}]$ ,  $c_4 := 2D$  and  $D := 365$ . Then, for any step  $t \geq 0$ , (i) for  $\varepsilon := \frac{1}{12}$ , it holds that*

$$\mathbb{E} \left[ \Lambda^{t+1} \mid \mathfrak{F}^t, \Delta^t \leq Dng \right] \leq \Lambda^t \cdot \left( 1 - \frac{2\alpha\varepsilon}{n} \right) + 18\alpha.$$

Furthermore, this also implies that (ii) for  $c := \frac{18}{\varepsilon}$ , it holds that

$$\mathbb{E} \left[ \Lambda^{t+1} \mid \mathfrak{F}^t, \Delta^t \leq Dng, \Lambda^t > cn \right] \leq \Lambda^t \cdot \left( 1 - \frac{\alpha\varepsilon}{n} \right).$$

**PROOF.** *First statement.* Consider an arbitrary step  $t \geq 0$  with  $\Delta^t \leq Dng$ . We bound the expected change of  $\Lambda$  over one step, by considering the following cases for each bin  $i \in [n]$ :

**Case 1** [ $y_i^t \in (-c_4 g - 2, c_4 g + 2)$ ]: Using Lemma 5.6 (i),

$$\begin{aligned} \mathbb{E} \left[ \Lambda_i^{t+1} \mid \mathfrak{F}^t \right] &\leq \Lambda_i^t \cdot \left( 1 + \frac{3\alpha}{n} \right) \\ &= \Lambda_i^t \cdot \left( 1 - \frac{\alpha}{6n} \right) + \Lambda_i^t \cdot \left( \frac{\alpha}{6n} + \frac{3\alpha}{n} \right) \end{aligned}$$



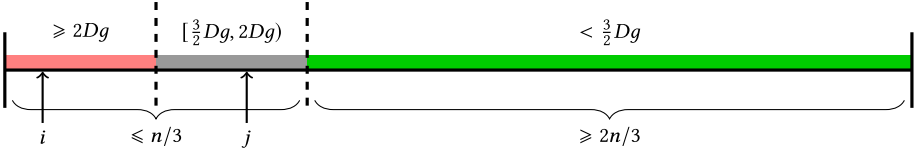


Fig. 4. In any step  $t$  with  $\Delta^t \leq Dng$  there are at most  $\frac{n}{3}$  bins with  $y_i^t \geq \frac{3}{2}Dg$ . In the  $g$ -ADV-COMP, we can distinguish between the red and green bins, so the probability to allocate to a red bin is at most  $\frac{2}{3n}$ .

$$\begin{aligned} &\stackrel{(a)}{\leq} \Lambda_i^t \cdot \left(1 - \frac{\alpha}{6n}\right) + (2 \cdot e^{2\alpha}) \cdot \left(\frac{1}{6} + 3\right) \cdot \frac{\alpha}{n} \\ &\stackrel{(b)}{\leq} \Lambda_i^t \cdot \left(1 - \frac{\alpha}{6n}\right) + \frac{18\alpha}{n}, \end{aligned}$$

using in (a) that  $\Lambda_i^t \leq e^{2\alpha} + 1 \leq 2 \cdot e^{2\alpha}$  by the assumption that  $y_i^t \in (-c_4g - 2, c_4g + 2)$  and in (b) that  $(2 \cdot e^{2\alpha}) \cdot (\frac{1}{6} + 3) \leq 18$ , since  $\alpha \leq 1/4$ .

**Case 2** [ $y_i^t \geq c_4g + 2$ ]: By the condition  $\Delta^t \leq Dng$ , the number of bins  $j$  with  $y_j^t \geq \frac{3}{2}Dg$  is at most  $\frac{1}{2}\Delta^t \cdot \frac{2}{3Dg} \leq \frac{Dng}{3Dg} = \frac{n}{3}$  (see Figure 4). We can allocate to bin  $i \in [n]$  with  $y_i^t \geq c_4g + 2 = 2Dg + 2 \geq \frac{3}{2}Dg + g$  only if we sample bin  $i$  and a bin  $j$  with  $y_j^t \geq \frac{3}{2}Dg$ . Hence,

$$q_i^t \leq 2 \cdot \frac{1}{n} \cdot \frac{1}{3} = \frac{2}{3n}.$$

By assumption, bin  $i$  deterministically satisfies  $y_i^{t+1} \geq c_4g + 2 - 1/n > c_4g$ , so

$$\begin{aligned} &\mathbb{E} \left[ \Lambda_i^{t+1} \mid \mathcal{F}^t, \Delta^t \leq Dng \right] \\ &= e^{\alpha \cdot (y_i^t - c_4g)^+} \cdot \left( (1 - q_i^t) \cdot e^{-\alpha/n} + q_i^t \cdot e^{\alpha(1-1/n)} \right) + 1 \\ &\stackrel{(a)}{\leq} e^{\alpha \cdot (y_i^t - c_4g)^+} \cdot \left( (1 - q_i^t) \cdot \left(1 - \frac{\alpha}{n} + \frac{\alpha^2}{n^2}\right) + q_i^t \cdot \left(1 + \alpha \cdot \left(1 - \frac{1}{n}\right) + \alpha^2\right) \right) + 1 \\ &= e^{\alpha \cdot (y_i^t - c_4g)^+} \cdot \left( 1 + \alpha \cdot \left(q_i^t - \frac{1}{n}\right) + (1 - q_i^t) \cdot \frac{\alpha^2}{n^2} + q_i^t \cdot \alpha^2 \right) + 1 \\ &\stackrel{(b)}{\leq} e^{\alpha \cdot (y_i^t - c_4g)^+} \cdot \left( 1 - \frac{\alpha}{3n} + \frac{4\alpha^2}{3n} \right) + 1 \\ &\stackrel{(c)}{\leq} e^{\alpha \cdot (y_i^t - c_4g)^+} \cdot \left( 1 - \frac{\alpha}{6n} \right) + 1 \\ &= e^{\alpha \cdot (y_i^t - c_4g)^+} \cdot \left( 1 - \frac{\alpha}{6n} \right) + 1 \cdot \left( 1 - \frac{\alpha}{6n} \right) + \frac{\alpha}{6n} \\ &= \Lambda_i^t \cdot \left( 1 - \frac{\alpha}{6n} \right) + \frac{\alpha}{6n}, \end{aligned}$$

using in (a) that  $e^u \leq 1 + u + u^2$  for  $u < 1.75$ ,  $\alpha \leq 1$  and  $(1 - 1/n)^2 \leq 1$ , in (b) that  $q_i^t \leq \frac{2}{3n}$  and  $(1 - q_i^t) \cdot \frac{\alpha^2}{n^2} \leq \frac{2\alpha^2}{3n}$  for  $n \geq 2$ , and in (c) that  $\alpha \leq \frac{1}{8}$ , so  $\frac{4\alpha^2}{3n} \leq \frac{\alpha}{6n}$ .

**Case 3** [ $y_i^t \leq -c_4g - 2$ ]: The number of bins  $j$  with  $y_j^t \leq -\frac{3}{2}Dg$  is at most  $\frac{1}{2}\Delta^t \cdot \frac{2}{3Dg} \leq \frac{Dng}{3Dg} = \frac{n}{3}$  and the number of bins  $j$  with  $y_j^t > -\frac{3}{2}Dg$  is at least  $\frac{2n}{3}$  (see Figure 5). Similarly to Case 2, we can allocate to a bin  $i \in [n]$  with load  $y_i^t \leq -c_4g - 2$  if we sample  $i$  and a bin  $j$  with  $y_j^t > -\frac{3}{2}Dg$ . Hence,

$$q_i^t \geq 2 \cdot \frac{1}{n} \cdot \frac{2}{3} = \frac{4}{3n}.$$

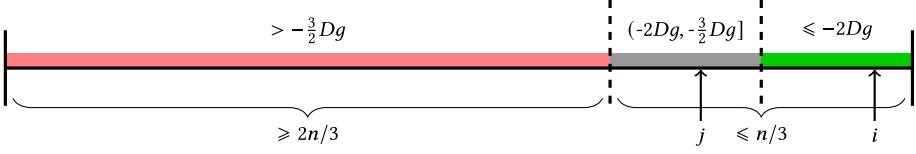


Fig. 5. In any step  $t$  with  $\Delta^t \leq Dng$  there are at most  $n/3$  bins with  $y_i^t \leq -\frac{3}{2}Dg$  and at least  $2n/3$  bins with  $y_i^t > -\frac{3}{2}Dg$ . In the  $g$ -ADV-COMP, we can distinguish between the red and green bins, so the probability to allocate to a green bin is at least  $\frac{4}{3n}$ .

By assumption, bin  $i$  deterministically satisfies  $y_i^{t+1} \leq -c_4g - 2 + 1 < -c_4g$ , so

$$\begin{aligned}
 & \mathbb{E} \left[ \Lambda_i^{t+1} \mid \mathcal{F}^t, \Delta^t \leq Dng \right] \\
 &= 1 + e^{\alpha(-y_i^t - c_4g)^+} \cdot \left( (1 - q_i^t) \cdot e^{\alpha/n} + q_i^t \cdot e^{-\alpha(1-1/n)} \right) \\
 &\stackrel{(a)}{\leq} 1 + e^{\alpha(-y_i^t - c_4g)^+} \cdot \left( (1 - q_i^t) \cdot \left( 1 + \frac{\alpha}{n} + \frac{\alpha^2}{n^2} \right) + q_i^t \cdot \left( 1 - \alpha \cdot \left( 1 - \frac{1}{n} \right) + \alpha^2 \right) \right) \\
 &= 1 + e^{\alpha(-y_i^t - c_4g)^+} \cdot \left( 1 + \alpha \cdot \left( \frac{1}{n} - q_i^t \right) + (1 - q_i^t) \cdot \frac{\alpha^2}{n^2} + q_i^t \cdot \alpha^2 \right) \\
 &\stackrel{(b)}{\leq} 1 + e^{\alpha(-y_i^t - c_4g)^+} \cdot \left( 1 - \frac{\alpha}{3n} + \frac{5\alpha^2}{2n} \right) \\
 &\stackrel{(c)}{\leq} 1 + e^{\alpha(-y_i^t - c_4g)^+} \cdot \left( 1 - \frac{\alpha}{6n} \right) \\
 &= \left( 1 - \frac{\alpha}{6n} \right) + e^{\alpha(-y_i^t - c_4g)^+} \cdot \left( 1 - \frac{\alpha}{6n} \right) + \frac{\alpha}{6n} \\
 &= \Lambda_i^t \cdot \left( 1 - \frac{\alpha}{6n} \right) + \frac{\alpha}{6n},
 \end{aligned}$$

using in (a) that  $e^u \leq 1 + u + u^2$  for  $u < 1.75$ ,  $\alpha \leq 1$  and  $(1 - 1/n)^2 \leq 1$ , in (b) that  $q_i^t \in [\frac{4}{3n}, \frac{2}{n}]$  and  $(1 - q_i^t) \cdot \frac{\alpha^2}{n^2} \leq \frac{\alpha^2}{2n}$  for  $n \geq 2$  and in (c) that  $\alpha \leq \frac{1}{15}$ , so  $\frac{5\alpha^2}{2n} \leq \frac{\alpha}{6n}$ .

Combining these three cases and letting  $\varepsilon := \frac{1}{12}$ , we conclude that

$$\mathbb{E} \left[ \Lambda^{t+1} \mid \mathcal{F}^t, \Delta^t \leq Dng \right] \leq \sum_{i=1}^n \left( \Lambda_i^t \cdot \left( 1 - \frac{\alpha}{6n} \right) + \frac{18\alpha}{n} \right) = \Lambda^t \cdot \left( 1 - \frac{2\alpha\varepsilon}{n} \right) + 18\alpha.$$

*Second statement.* Letting  $c := \frac{18}{\varepsilon} = 18 \cdot 12$ , it follows that

$$\begin{aligned}
 \mathbb{E} \left[ \Lambda^{t+1} \mid \mathcal{F}^t, \Delta^t \leq Dng, \Lambda^t > cn \right] &\leq \Lambda^t \cdot \left( 1 - \frac{2\alpha\varepsilon}{n} \right) + 18\alpha \\
 &= \Lambda^t \cdot \left( 1 - \frac{\alpha\varepsilon}{n} \right) - \Lambda^t \cdot \frac{\alpha\varepsilon}{n} + 18\alpha \\
 &\leq \Lambda^t \cdot \left( 1 - \frac{\alpha\varepsilon}{n} \right).
 \end{aligned}$$

□

## 5.5 Adjusted Exponential Potential

In Lemma 5.7, we proved that in a good step  $t$  with  $\Lambda^t > cn$  (for  $c := 18 \cdot 12$ ), the potential drops in expectation by a multiplicative factor. Our goal will be to show that w.h.p.  $\Lambda^t \leq cn$  at a single step (**recovery**) and then show that it becomes small at least once every  $\mathcal{O}(n \cdot (g + \log n))$  steps (**stabilization**). Since we do not have an expected drop in every step, but only at a constant

fraction  $r$  of the steps, we will define an *adjusted exponential potential function*. First, for any step  $t_0$ , and any step  $s \geq t_0$ , we define the following event:

$$\mathcal{E}_{t_0}^s := \bigcap_{t \in [t_0, s]} \{\Lambda^t > cn\}.$$

Next, we define the sequence  $(\tilde{\Lambda}_{t_0}^s)_{s \geq t_0} := (\tilde{\Lambda}_{t_0}^s)_{s \geq t_0}(\alpha, c_4 g, \varepsilon)$  as  $\tilde{\Lambda}_{t_0}^{t_0} := \Lambda^{t_0}(\alpha, c_4 g)$  and, for any  $s > t_0$ ,

$$\tilde{\Lambda}_{t_0}^s := \Lambda^s(\alpha, c_4 g) \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \exp\left(-\frac{3\alpha}{n} \cdot B_{t_0}^{s-1}\right) \cdot \exp\left(+\frac{\alpha\varepsilon}{n} \cdot G_{t_0}^{s-1}\right), \quad (5.6)$$

recalling that  $G_{t_0}^{t_1}$  is the number of good steps in  $[t_0, t_1]$ , i.e., steps where the event  $\mathcal{G}^s := \{\Delta^s \leq Dng\}$  holds, and  $B_{t_0}^{t_1} := [t_0, t_1] \setminus G_{t_0}^{t_1}$  is the number of bad steps in  $[t_0, t_1]$ .

By its definition and Lemmas 5.6 and 5.7, it follows that the sequence  $(\tilde{\Lambda}_{t_0}^s)_{s \geq t_0}$  is a supermartingale. A proof is given in Appendix B.2.

LEMMA 5.8 (CF. [46, LEMMA 9.1]). *Consider the  $g$ -ADV-COMP setting for any  $g \geq 1$  and the sequence  $(\tilde{\Lambda}_{t_0}^s)_{s \geq t_0} := (\tilde{\Lambda}_{t_0}^s)_{s \geq t_0}(\alpha, c_4 g, \varepsilon)$  for any starting step  $t_0 \geq 0$ , any  $\alpha \in (0, \frac{1}{18}]$  and  $\varepsilon, c_4 > 0$  as defined in Lemma 5.7. Then, for any step  $s \geq t_0$ ,*

$$\mathbb{E} \left[ \tilde{\Lambda}_{t_0}^{s+1} \mid \mathcal{F}^s \right] \leq \tilde{\Lambda}_{t_0}^s.$$

## 5.6 Recovery and Stabilization

We are now ready to prove the recovery lemma, that is the  $\Lambda$  potential becomes small every  $\Theta(n g (\log(n g))^2)$  steps.

LEMMA 5.9 (RECOVERY). *Consider the  $g$ -ADV-COMP setting for any  $g \geq 1$  and the potential  $\Lambda := \Lambda(\alpha, c_4 g)$  with  $\alpha = \frac{1}{18}$ , and  $c_4 > 0$  as defined in Lemma 5.7. Further, let the constants  $c, \varepsilon > 0$  be as defined in Lemma 5.7,  $c_r := c_r(\alpha, c_4) \geq 1$  as in Lemma 5.5 (ii),  $r \in (0, 1)$  as in Lemma 5.4 and  $c_3 \geq 2$  as in Theorem 4.3. Then, for any step  $t_0 \geq 0$ , (i) for  $\Delta_r := \Delta_r(g) := \frac{60c_3^2 c_r}{\alpha \varepsilon r} \cdot ng \cdot (\log(n g))^2$ , it holds that*

$$\Pr \left[ \bigcup_{t \in [t_0, t_0 + \Delta_r]} \{\Lambda^t \leq cn\} \mid \mathcal{F}^{t_0}, \max_{i \in [n]} |y_i^{t_0}| \leq c_3 g \log(n g) \right] \geq 1 - 3 \cdot (ng)^{-12}.$$

Further, (ii) it holds that,

$$\Pr \left[ \bigcup_{t \in [t_0, t_0 + \Delta_r]} \{\Lambda^t \leq cn\} \right] \geq 1 - (ng)^{-11}.$$

PROOF. *First statement.* Consider an arbitrary step  $t_0$  with  $\max_{i \in [n]} |y_i^{t_0}| \leq c_3 g \log(n g)$ . Our aim is to show that w.h.p.  $\Lambda^t \leq cn$  for some step  $t \in [t_0, t_0 + \Delta_r]$ . We will do this by first showing that w.h.p. there is a significant number of good steps, i.e.,  $G_{t_0}^{t_1} \geq r \cdot \Delta_r$ , and when this happens w.h.p.  $\tilde{\Lambda}_{t_0}^{t_1} = 0$ , which implies the conclusion.

We start by upper bounding  $\Lambda^{t_0}$  as follows,

$$\Lambda^{t_0} \leq 2n \cdot e^{\alpha c_3 g \log(n g)} \leq e^{c_3 g \log(n g)} =: \lambda, \quad (5.7)$$

since  $\alpha \leq \frac{1}{18}$  and  $c_3 \geq 2$ . Hence, by Lemma 5.5 (ii), there exists a constant  $c_r := c_r(\alpha, c_4)$ , such that

$$\Upsilon^{t_0} \leq c_r \cdot n \cdot (g^2 + (\log \Lambda^{t_0})^2) \leq 2c_r \cdot n \cdot (c_3 g \log(n g))^2 =: T.$$

Let  $t_1 := t_0 + \Delta_r$ . Applying Lemma 5.4 with  $T = 2c_r \cdot n \cdot (c_3 g \log(\log ng))^2 = o(n^2 g^3)$  (and  $\geq ng^2$ ) and  $\hat{c} = \frac{\Delta_r \cdot g}{T} \geq \frac{30}{\alpha \varepsilon r} \geq 1$  as  $\alpha, \varepsilon, r \leq 1$ , we get

$$\begin{aligned} & \Pr \left[ G_{t_0}^{t_1-1} \geq r \cdot \Delta_r \mid \mathcal{F}^{t_0}, \Lambda^{t_0} \leq \lambda, \max_{i \in [n]} |y_i^{t_0}| \leq c_3 g \log(\log ng) \right] \\ & \geq \Pr \left[ G_{t_0}^{t_1-1} \geq r \cdot \Delta_r \mid \mathcal{F}^{t_0}, Y^{t_0} \leq T, \max_{i \in [n]} |y_i^{t_0}| \leq g(\log(\log ng))^2 \right] \\ & \geq 1 - 2 \cdot (ng)^{-12}. \end{aligned} \quad (5.8)$$

By Lemma 5.8,  $(\tilde{\Lambda}_{t_0}^t)_{t \geq t_0}$  is a super-martingale, so  $\mathbb{E}[\tilde{\Lambda}_{t_0}^{t_1} \mid \mathcal{F}^{t_0}] \leq \tilde{\Lambda}_{t_0}^{t_0} = \Lambda^{t_0}$ . Hence, using Markov's inequality we get  $\Pr[\tilde{\Lambda}_{t_0}^{t_1} > \Lambda^{t_0} \cdot (ng)^{12} \mid \mathcal{F}^{t_0}, \Lambda^{t_0} \leq \lambda] \leq (ng)^{-12}$ . Thus, by the definition of  $\tilde{\Lambda}_{t_0}^{t_1}$  in Equation (5.6), we have

$$\Pr \left[ \Lambda^{t_1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{t_1-1}} \leq \Lambda^{t_0} \cdot (ng)^{12} \cdot \exp \left( \frac{3\alpha}{n} \cdot B_{t_0}^{t_1-1} - \frac{\alpha \varepsilon}{n} \cdot G_{t_0}^{t_1-1} \right) \mid \mathcal{F}^{t_0}, \Lambda^{t_0} \leq \lambda \right] \geq 1 - (ng)^{-12}. \quad (5.9)$$

Further, if in addition to the two events  $\{\tilde{\Lambda}_{t_0}^{t_1} \leq \Lambda^{t_0} \cdot (ng)^{12}\}$  and  $\{\Lambda^{t_0} \leq \lambda\}$ , also the event  $\{G_{t_0}^{t_1-1} \geq r \cdot \Delta_r\}$  holds, then

$$\begin{aligned} \Lambda^{t_1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{t_1-1}} & \leq \Lambda^{t_0} \cdot (ng)^{12} \cdot \exp \left( \frac{3\alpha}{n} \cdot B_{t_0}^{t_1-1} - \frac{\alpha \varepsilon}{n} \cdot G_{t_0}^{t_1-1} \right) \\ & \leq e^{c_3 g \log(\log ng)} \cdot (ng)^{12} \cdot \exp \left( \frac{3\alpha}{n} \cdot (1-r) \cdot \Delta_r - \frac{\alpha \varepsilon}{n} \cdot r \cdot \Delta_r \right) \\ & \stackrel{(a)}{=} e^{c_3 g \log(\log ng)} \cdot (ng)^{12} \cdot \exp \left( \frac{\alpha \varepsilon}{n} \cdot \frac{r}{2} \cdot \Delta_r - \frac{\alpha \varepsilon}{n} \cdot r \cdot \Delta_r \right) \\ & = e^{c_3 g \log(\log ng)} \cdot (ng)^{12} \cdot \exp \left( -\frac{\alpha \varepsilon}{n} \cdot \frac{r}{2} \cdot \Delta_r \right) \\ & = e^{c_3 g \log(\log ng)} \cdot (ng)^{12} \cdot \exp \left( -\frac{\alpha \varepsilon}{n} \cdot \frac{r}{2} \cdot \frac{60c_3^2 c_r}{\alpha \varepsilon r} \cdot ng \cdot (\log(\log ng))^2 \right) \\ & \stackrel{(b)}{\leq} e^{c_3 g \log(\log ng)} \cdot (ng)^{12} \cdot \exp(-30c_3 g \log(\log ng)) \\ & \leq 1, \end{aligned}$$

where we used in (a) that  $r = \frac{6}{6+\varepsilon}$  implies  $\frac{3\alpha}{n} \cdot (1-r) = \frac{3\alpha}{n} \cdot \frac{\varepsilon}{6+\varepsilon} = \frac{\alpha \varepsilon}{n} \cdot \frac{r}{2}$ , in (b) that  $g \geq 1$ ,  $c_r \geq 1$  and  $c_3 \geq 1$ . By the definition of  $\Lambda$ , we have that  $\Lambda^{t_1} \geq n$  holds deterministically, and so we can deduce from the above inequality that  $\mathbf{1}_{\mathcal{E}_{t_0}^{t_1-1}} = 0$ , that is,

$$\Pr \left[ \neg \mathcal{E}_{t_0}^{t_1-1} \mid \mathcal{F}^{t_0}, \tilde{\Lambda}_{t_0}^{t_1} \leq \Lambda^{t_0} \cdot (ng)^{12}, \Lambda^{t_0} \leq \lambda, G_{t_0}^{t_1-1} \geq r \cdot \Delta_r \right] = 1.$$

Recalling the definition of  $\mathcal{E}_{t_0}^{t_1-1} := \bigcap_{t \in [t_0, t_1-1]} \{\Lambda^t > cn\}$  and taking the union bound over Equations (5.8) and (5.9) yields

$$\Pr \left[ \bigcup_{t \in [t_0, t_0 + \Delta_r]} \{\Lambda^t \leq cn\} \mid \mathcal{F}^{t_0}, \Lambda^{t_0} \leq \lambda \right] \geq 1 - 2 \cdot (ng)^{-12} - (ng)^{-12} = 1 - 3 \cdot (ng)^{-12}. \quad (5.10)$$

*Second statement.* Using Theorem 4.3 (iii), for the constant  $c_3 \geq 2$ , it holds that,

$$\Pr \left[ \max_{i \in [n]} |y_i^{t_0}| \leq c_3 g \log(\log ng) \right] \geq 1 - (ng)^{-14}.$$

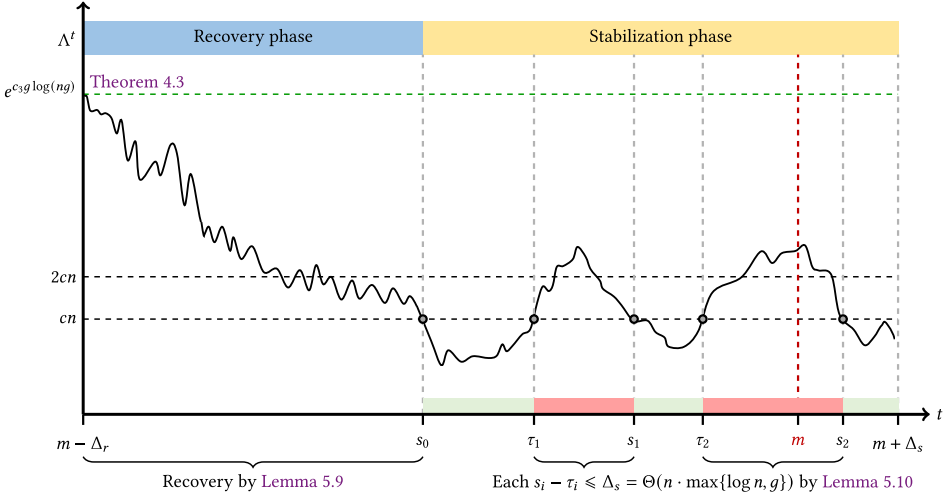


Fig. 6. Visualization of the recovery (Lemma 5.9) and stabilization (Lemma 5.10) phases. Note that the red intervals w.h.p. will have length  $\leq \Delta_s$ , implying  $\text{Gap}(m) = O(g + \log n)$ .

Recall by Equation (5.7), that this implies that

$$\Pr \left[ \Lambda^{t_0} \leq \lambda \right] \geq 1 - (ng)^{-14}.$$

Hence, combining with Equation (5.10), we conclude that

$$\Pr \left[ \bigcup_{t \in [t_0, t_0 + \Delta_r]} \{ \Lambda^t \leq cn \} \right] \geq (1 - 3 \cdot (ng)^{-12}) \cdot (1 - (ng)^{-14}) \geq 1 - (ng)^{-11}. \quad \square$$

The derivation of the lemma below is similar to that of Lemma 5.9, with the main difference being the tighter condition that  $\Lambda^{t_0} \leq 2cn$ , which allows us to choose a slightly shorter time window of  $\Theta(n \cdot \max\{\log n, g\})$  steps (see Figure 6). We defer the proof for Appendix B.2.

**LEMMA 5.10 (STABILIZATION).** *Consider the  $g$ -ADV-COMP setting for any  $g \geq 1$  and the potential  $\Lambda := \Lambda(\alpha, c_4 g)$  with  $\alpha = \frac{1}{18}$ , and  $c_4 > 0$  as defined in Lemma 5.7. Further, let the constants  $c, \varepsilon > 0$  be as defined in Lemma 5.7,  $c_s := c_s(\alpha, c_4, 2c) \geq 1$  as in Lemma 5.5 (i) and  $r \in (0, 1)$  as in Lemma 5.4. Then, for  $\Delta_s := \Delta_s(g) := \frac{60c_s}{\alpha \varepsilon r} \cdot n \cdot \max\{\log n, g\}$ , we have that for any step  $t_0 \geq 0$ ,*

$$\Pr \left[ \bigcup_{t \in [t_0, t_0 + \Delta_s]} \{ \Lambda^t \leq cn \} \mid \mathcal{F}^{t_0}, \Lambda^{t_0} \leq 2cn \right] \geq 1 - (ng)^{-11}.$$

## 5.7 Completing the Proof of Theorem 5.12

We will now prove that starting with  $\Lambda^{t_0} \leq 2cn$  implies that for any step  $t_1 \in [t_0, t_0 + (ng)^2]$ , we have w.h.p.  $\text{Gap}(t_1) = O(g + \log n)$ .

**LEMMA 5.11.** *Consider the  $g$ -ADV-COMP setting for any  $g \geq 1$ , the potential  $\Lambda := \Lambda(\alpha, c_4 g)$  with  $\alpha = \frac{1}{18}$ ,  $c_4 > 0$  as defined in Lemma 5.7 and  $\Delta_s > 0$  as defined in Lemma 5.10. Then, there exists a constant  $\kappa \geq \frac{1}{\alpha}$  such that for any steps  $t_0 \geq 0$  and  $t_1 \in (t_0, t_0 + (ng)^2]$ ,*

$$\Pr \left[ \max_{i \in [n]} |y_i^{t_1}| \leq \kappa \cdot (g + \log n) \mid \mathcal{F}^{t_0}, \Lambda^{t_0} \leq cn \right] \geq 1 - (ng)^{-9}.$$

PROOF. Consider any step  $t_0$  with  $\Lambda^{t_0} \leq cn$ . We define the event for any step  $t_1 \geq t_0$

$$\mathcal{M}_{t_0}^{t_1} := \{\text{for all } t \in [t_0, t_1] \text{ there exists } s \in [t, t + \Delta_s] \text{ such that } \Lambda^s \leq cn\},$$

that is, if  $\mathcal{M}_{t_0}^{t_1}$  holds then we have  $\Lambda^s \leq cn$  at least once every  $\Delta_s := \frac{60c_s}{\alpha \varepsilon r} \cdot n \cdot \max\{\log n, g\}$  steps.

Assume now that  $\mathcal{M}_{t_0}^{t_1}$  holds. We will show that

$$\max_{i \in [n]} |y_i^{t_1}| \leq \kappa \cdot (g + \log n).$$

Choosing  $t = t_1$ , implies that there exists  $s \in [t_1, t_1 + \Delta_s]$  such that  $\Lambda^s \leq cn$ , which in turn implies by definition of  $\Lambda$  that  $\max_{i \in [n]} |y_i^s| \leq \frac{1}{\alpha} \cdot \log(cn) + c_4 g < \frac{2}{\alpha} \cdot \log n + c_4 g$ . Clearly, any  $y_i^t$  can decrease by at most  $1/n$  in each step, and from this it follows that if  $\mathcal{M}_{t_0}^{t_1}$  holds, then

$$\max_{i \in [n]} y_i^{t_1} \leq \max_{i \in [n]} |y_i^s| + \frac{\Delta_s}{n} \leq \kappa \cdot (g + \log n),$$

for the constant

$$\kappa := \frac{2}{\alpha} + c_4 + \frac{\Delta_s}{n \cdot \max\{\log n, g\}} = \frac{2}{\alpha} + c_4 + \frac{60c_s}{\alpha \varepsilon r} > 0. \quad (5.11)$$

If  $t_1 \geq t_0 + \Delta_s$  and  $\mathcal{M}_{t_0}^{t_1}$  holds, then choosing  $t = t_1 - \Delta_s$ , there exists  $s \in [t_1 - \Delta_s, t_1]$  such that  $\Lambda^s \leq cn$ . (In case  $t_1 < t_0 + \Delta_s$ , then we arrive at the same conclusion by choosing  $s = t_0$  and using the precondition  $\Lambda^{t_0} \leq cn$ ). This in turn implies  $\max_{i \in [n]} |y_i^s| \leq \frac{2}{\alpha} \cdot \log n + c_4 g$ . Hence

$$\min_{i \in [n]} y_i^{t_1} \geq -\max_{i \in [n]} |y_i^s| - \frac{\Delta_s}{n} \geq -\kappa \cdot (g + \log n).$$

Hence,  $\mathcal{M}_{t_0}^{t_1}$  together with the precondition on  $\Lambda^{t_0} \leq cn$  implies that  $\max_{i \in [n]} |y_i^{t_1}| \leq \kappa \cdot (g + \log n)$ . It remains to bound  $\Pr[\neg \mathcal{M}_{t_0}^{t_1} \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \leq cn]$ .

Note that if for some step  $j_1$  we have that  $\Lambda^{j_1} \leq cn$  and for some  $j_2 \geq j_1$  that  $\Lambda^{j_2} > 2cn$ , then there must exist  $j \in (j_1, j_2)$  such that  $\Lambda^j \in (cn, 2cn]$ , since for every  $t \geq 0$  it holds that  $\Lambda^{t+1} \leq \Lambda^t \cdot e^\alpha \leq 2\Lambda^t$ , as  $\alpha \leq 1/2$ . Let  $t_0 < \tau_1 < \tau_2 < \dots$  and  $t_0 =: s_0 < s_1 < \dots$  be two interlaced sequences defined recursively for  $i \geq 1$  by

$$\tau_i := \inf \{\tau > s_{i-1} : \Lambda^\tau \in (cn, 2cn]\}, \quad \text{and} \quad s_i := \inf \{s > \tau_i : \Lambda^s \leq cn\}.$$

Thus we have

$$t_0 = s_0 < \tau_1 < s_1 < \tau_2 < s_2 < \dots,$$

and since  $\tau_i > \tau_{i-1}$  we have  $\tau_{t_1-t_0} \geq t_1$ . Therefore, if the event  $\cap_{i=1}^{t_1-t_0} \{s_i - \tau_i \leq \Delta_s\}$  holds, then also  $\mathcal{M}_{t_0}^{t_1}$  holds.

Recall that by Lemma 5.10 we have for any  $i = 1, 2, \dots, t_1 - t_0$  and any  $\tau = t_0 + 1, \dots, t_1$

$$\Pr \left[ \bigcup_{t \in [\tau_i, \tau_i + \Delta_s]} \{\Lambda^t \leq cn\} \mid \mathfrak{F}^\tau, \Lambda^\tau \in (cn, 2cn], \tau_i = \tau \right] \geq 1 - (ng)^{-11},$$

and by negating and the definition of  $s_i$ ,

$$\Pr \left[ s_i - \tau_i > \Delta_s \mid \mathfrak{F}^\tau, \Lambda^\tau \in (cn, 2cn], \tau_i = \tau \right] \leq (ng)^{-11}.$$

Since the above bound holds for any  $i \geq 1$  and  $\mathfrak{F}^\tau$ , with  $\tau_i = \tau$ , it follows by the union bound over all  $i = 1, 2, \dots, t_1 - t_0$ , as  $t_1 - t_0 \leq (ng)^2$ ,

$$\Pr \left[ \neg \mathcal{M}_{t_0}^{t_1} \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \leq cn \right] \leq (t_1 - t_0) \cdot (ng)^{-11} \leq (ng)^{-9}. \quad \square$$



Finally, we deduce that for the  $g$ -ADV-COMP setting, for an arbitrary step  $m$  w.h.p.  $\text{Gap}(m) = O(g + \log n)$ .

**THEOREM 5.12.** *Consider the  $g$ -ADV-COMP setting for any  $g \geq 1$ , the constant  $\kappa \geq \frac{1}{\alpha}$  defined in Equation (5.11) in Lemma 5.11 and  $\alpha = \frac{1}{18}$ . Then, for any step  $m \geq 0$ ,*

$$\Pr \left[ \max_{i \in [n]} |y_i^m| \leq \kappa \cdot (g + \log n) \right] \geq 1 - 2 \cdot (ng)^{-9}.$$

**PROOF.** Consider an arbitrary step  $m \geq 0$  and recall that  $\Delta_r := \frac{60c_3^2 c_r}{\alpha \varepsilon r} \cdot ng \cdot (\log(ng))^2$ . If  $m < \Delta_r$ , then the claim follows by Lemma 5.11 as  $\Lambda^0 = 2n \leq cn$  and  $\Delta_r < (ng)^2$ .

Otherwise, let  $t_0 := m - \Delta_r$ . Firstly, by the recovery lemma (Lemma 5.9 (ii)), we get

$$\Pr \left[ \bigcup_{t \in [t_0, t_0 + \Delta_r]} \{ \Lambda^t \leq cn \} \right] \geq 1 - (ng)^{-11}. \quad (5.12)$$

Hence for  $\tau := \inf\{s \geq t_0 : \Lambda^s \leq cn\}$  we have  $\Pr[\tau \leq m] \geq 1 - (ng)^{-11}$ , as  $t_0 + \Delta_r = m$ .

Secondly, using Lemma 5.11, there exists a constant  $\kappa := \kappa(\alpha, \varepsilon) > 0$  such that for any step  $s \in [t_0, m]$ ,

$$\Pr \left[ \max_{i \in [n]} |y_i^m| \leq \kappa \cdot (g + \log n) \mid \mathcal{F}^s, \Lambda^s \leq cn \right] \geq 1 - (ng)^{-9}. \quad (5.13)$$

Combining the two inequalities from above, we conclude the proof

$$\begin{aligned} \Pr \left[ \max_{i \in [n]} |y_i^m| \leq \kappa \cdot (g + \log n) \right] &\geq \sum_{s=t_0}^m \Pr \left[ \max_{i \in [n]} |y_i^m| \leq \kappa \cdot (g + \log n) \mid \tau = s \right] \cdot \Pr[\tau = s] \\ &\geq \sum_{s=t_0}^m \Pr \left[ \max_{i \in [n]} |y_i^m| \leq \kappa \cdot (g + \log n) \mid \mathcal{F}^s, \Lambda^s \leq cn \right] \cdot \Pr[\tau = s] \\ &\stackrel{(5.13)}{\geq} (1 - (ng)^{-9}) \cdot \Pr[\tau \leq m] \\ &\stackrel{(5.12)}{\geq} (1 - (ng)^{-9}) \cdot (1 - (ng)^{-11}) \geq 1 - 2 \cdot (ng)^{-9}. \quad \square \end{aligned}$$

## 6 UPPER BOUND OF $O(\frac{g}{\log g} \cdot \log \log n)$ FOR $g$ -ADV-COMP WITH $g \leq \log n$ : OUTLINE

In this section, we will outline the proof for the upper bound on the gap of  $O(\frac{g}{\log g} \cdot \log \log n)$  for the  $g$ -ADV-COMP setting with  $1 < g \leq \log n$  (which also implies for  $g \in \{0, 1\}$  the  $O(\log \log n)$  gap bound by monotonicity). This matches the lower bound for the  $g$ -MYOPIC-COMP process proven later in Theorem 11.3 up to multiplicative constants. The upper bound proof will be completed in Sections 7 to 9.

**THEOREM 9.2 (RESTATED, PAGE 52).** *Consider the  $g$ -ADV-COMP setting for any  $g \in (1, \log n]$ . Then, there exists a constant  $\tilde{\kappa} > 0$  such that for any step  $m \geq 0$ ,*

$$\Pr \left[ \text{Gap}(m) \leq \tilde{\kappa} \cdot \frac{g}{\log g} \cdot \log \log n \right] \geq 1 - n^{-3}.$$

### 6.1 Definitions of Super-Exponential Potential Functions

The proof of this theorem employs some kind of layered induction over  $k$  different, super-exponential potential functions. We will be using the following definition of super-exponential potentials.

The *super-exponential potential function* with smoothing parameter  $\phi \geq 1$  and integer offset  $z := z(n) > 0$  is defined at any step  $t \geq 0$  as

$$\Phi^t := \Phi^t(\phi, z) := \sum_{i=1}^n \Phi_i^t := \sum_{i=1}^n e^{\phi \cdot (y_i^t - z)^+}, \quad (6.1)$$

where  $u^+ := \max\{u, 0\}$ .

These are similar to the potential functions used in [43, Section 6], but with two extensions: (i) the potential functions are “smoother”, e.g., allowing us to deduce gap bounds between  $\sqrt{\log n}$  and  $\log n$ , and (ii) the concentration of the base potential in the layered induction follows from strengthening the stabilization theorem in Section 5.

Let

$$\alpha_1 := \frac{1}{6\kappa} \leq \frac{1}{6 \cdot 18}, \quad (6.2)$$

for  $\kappa \geq \frac{1}{\alpha} = 18 > 0$  the constant in Equation (5.11) in Lemma 5.11 and

$$\alpha_2 := \frac{\alpha_1}{84} \leq \frac{1}{84 \cdot 6 \cdot 18}. \quad (6.3)$$

We define the function

$$f(k) := (\alpha_1 \log n)^{1/k} = e^{\frac{1}{k} \cdot \log(\alpha_1 \log n)},$$

which is monotone decreasing in  $k > 0$ , and for  $k = 1$ ,  $f(1) = \alpha_1 \log n$ . This implies that for every  $1 < g < \alpha_1 \log n$ , there exists a unique integer  $k := k(g) \geq 2$  satisfying,

$$(\alpha_1 \log n)^{1/k} \leq g < (\alpha_1 \log n)^{1/(k-1)}.$$

This definition implies that  $k = \Theta(\frac{\log \log n}{\log g})$  and that  $k = O(\log \log n)$ , since  $g > 1$ .

Keeping in mind the previous inequality, we will be making the slightly stronger assumption for  $g = \Omega(1)$  (see Claim B.1) that

$$(\alpha_1 \cdot (\log n))^{1/k} \leq g < \left( \frac{\alpha_2}{4} \cdot (\log n) \right)^{1/(k-1)}. \quad (6.4)$$

For any  $g$  satisfying  $(\frac{\alpha_2}{4} \cdot (\log n))^{1/k} \leq g < (\alpha_1 \cdot (\log n))^{1/k}$ , we will obtain the stated  $O(\frac{g}{\log g} \cdot \log \log n)$  bound by analyzing the  $\tilde{g}$ -ADV-COMP setting for  $\tilde{g} = (\alpha_1 \cdot (\log n))^{1/k} > g$ , since

$$\frac{\tilde{g}}{\log \tilde{g}} \leq \frac{\tilde{g}}{\log g} = \frac{\tilde{g}}{g} \cdot \frac{g}{\log g} \leq \left( \frac{4\alpha_1}{\alpha_2} \right)^{1/k} \cdot \frac{g}{\log g} = O\left( \frac{g}{\log g} \right).$$

We will now define the super-exponential potential functions  $\Phi_0, \dots, \Phi_{k-1}$ . The base potential function  $\Phi_0$  is just an exponential potential (i.e., has a constant smoothing parameter) defined as

$$\Phi_0^s := \Phi_0^s(\alpha_2, z_0) := \sum_{i=1}^n \Phi_{0,i}^s := \sum_{i=1}^n \exp\left(\alpha_2 \cdot (y_i^s - z_0)^+\right), \quad (6.5)$$

where  $\alpha_2 := \frac{\alpha_1}{84}$  and  $z_0 := c_5 \cdot g$  for some sufficiently large constant integer  $c_5 > 0$  (to be defined in Equation (7.14) in Lemma 7.7). Further, we define for any integer  $1 \leq j \leq k-1$ ,

$$\Phi_j^s := \Phi_j^s(\alpha_2 \cdot (\log n) \cdot g^{j-k}, z_j) := \sum_{i=1}^n \Phi_{j,i}^s := \sum_{i=1}^n \exp\left(\alpha_2 \cdot (\log n) \cdot g^{j-k} \cdot (y_i^s - z_j)^+\right), \quad (6.6)$$

where the offsets are given by

$$z_j := c_5 \cdot g + \left\lceil \frac{4}{\alpha_2} \right\rceil \cdot j \cdot g. \quad (6.7)$$

Note that for most choices of  $g$  and  $j$ , the smoothing parameter is  $\omega(1)$ , motivating the term “super-exponential”. For  $g = (\log n)^{1/k}$ , the potential functions match in form those defined in [43].

As concrete example, consider  $g = (\log n)^{5/12}$ , for which we have  $k = 3$  and the potential functions are:

$$\begin{aligned}\Phi_0^s &= \sum_{i=1}^n \exp\left(\alpha_2 \cdot (y_i^s - z_0)^+\right), \\ \Phi_1^s &= \sum_{i=1}^n \exp\left(\alpha_2 \cdot (\log n)^{2/12} \cdot (y_i^s - z_1)^+\right), \\ \Phi_2^s &= \sum_{i=1}^n \exp\left(\alpha_2 \cdot (\log n)^{7/12} \cdot (y_i^s - z_2)^+\right).\end{aligned}$$

When each of these potentials is  $O(n)$ , this implies increasingly stronger bounds on the gap:  $O(g + \log n)$ ,  $O(2g + (\log n)^{10/12})$  and finally  $O(3g + (\log n)^{5/12}) = O(g)$ .

In general, we will employ this series of potential functions  $(\Phi_j)_{j=0}^{k-1}$  to analyze the process over the time-interval  $[m - n \log^5 n, m]$ , with the goal of eventually establishing an  $O(g \cdot k)$  gap at time  $m$ . By definition of  $k$ , this will imply that  $\text{Gap}(m) = O(\frac{g}{\log g} \cdot \log \log n)$ .

## 6.2 The Layered Induction Argument

The next lemma (Lemma 9.3) formalizes the inductive argument outlined in Section 6.1. It shows that if for all steps  $s$  within some suitable time-interval, the number of bins with load at least  $z_j$  is small, then the number of bins with load at least  $z_{j+1}$  is even smaller. This “even smaller” is captured by the (non-constant) base of  $\Phi_j$ , which increases in  $j$ ; however, this comes at the cost of reducing the time-interval slightly by an additive  $2n \log^4 n$  term. Finally, for  $j = k - 1$ , we conclude that at step  $m$ , there are no bins with load at least  $z_k := c_5 \cdot g + \left\lceil \frac{4}{\alpha_2} \right\rceil \cdot k \cdot g$ , implying that  $\text{Gap}(m) = O(\frac{g}{\log g} \cdot \log \log n)$  and establishing Theorem 9.2.

**LEMMA 9.3 (INDUCTION STEP – SIMPLIFIED VERSION, PAGE 53).** *Consider the  $g$ -ADV-COMP setting for any  $g = \Omega(1)$  satisfying  $(\alpha_1 \log n)^{1/k} < g \leq (\frac{\alpha_2}{4} \log n)^{1/(k-1)}$  for some integer  $k \geq 2$ , for  $\alpha_1, \alpha_2 > 0$  defined in Equations (6.2) and (6.3). Then, for any integer  $1 \leq j \leq k - 1$  and any step  $m \geq 0$ , if it holds that*

$$\Pr \left[ \bigcap_{s \in [m - 2n(k-j+1) \cdot \log^4 n, m]} \{\Phi_{j-1}^s = O(n)\} \right] \geq 1 - \frac{(\log n)^{8(j-1)}}{n^4},$$

then it also follows that

$$\Pr \left[ \bigcap_{s \in [m - 2n(k-j) \cdot \log^4 n, m]} \{\Phi_j^s = O(n)\} \right] \geq 1 - \frac{(\log n)^{8j}}{n^4}.$$

*Base case.* The base case follows by strengthening the stabilization lemma for a variant of the  $\Lambda$  potential used in Section 5, so that the potential becomes  $O(n)$  every  $O(ng)$  steps (instead of  $\Lambda$  becoming  $O(n)$  every  $O(n \cdot (g + \log n))$  steps which was proven in Lemma 5.10). By making the constant  $c_5 > 0$  sufficiently large in the offset  $c_5 g$  of the  $\Phi_0$  potential, we obtain concentration for  $\Phi_0$  and establish the base case for the layered induction.

**THEOREM 7.8 (BASE CASE – SIMPLIFIED VERSION, PAGE 40).** *Consider the  $g$ -ADV-COMP setting for any  $g \in [1, O(\log n)]$ . Then, for any step  $m \geq 0$ ,*

$$\Pr \left[ \bigcap_{s \in [m - n \log^5 n, m]} \{\Phi_0^s = O(n)\} \right] \geq 1 - n^{-4}.$$

### 6.3 Analysis of Super-Exponential Potential Functions

In order to prove the induction step (Lemma 9.3), we will need to prove concentration for super-exponential potentials. We will now outline the proof of this and the challenges involved.

For super-exponential potentials, unlike the hyperbolic cosine potential, there exist load configurations, where a super-exponential potential may increase in expectation. We will show that in each step where the probability to allocate to a bin with load at least  $z - 1$  is sufficiently small, the potential function  $\Phi := \Phi(\phi, z)$  drops in expectation over one step. More specifically, we show that the following event is sufficient:

$$\mathcal{K}^s := \mathcal{K}_{\phi, z}^s(q^s) := \left\{ \forall i \in [n]: y_i^s \geq z - 1 \Rightarrow q_i^s \leq \frac{1}{n} \cdot e^{-\phi} \right\},$$

where  $q^s$  is the probability allocation vector used by the process at step  $s$ .

We prove the following drop inequality in Section 8.

**LEMMA 8.1 (RESTATED, PAGE 42).** *Consider any allocation process and any super-exponential potential  $\Phi := \Phi(\phi, z)$  with  $\phi \in [4, n]$ . For any step  $s \geq 0$  where  $\mathcal{K}^s$  holds, we have that*

$$\mathbb{E} \left[ \Phi^{s+1} \mid \mathcal{K}^s, \mathcal{K}^s \right] \leq \Phi^s \cdot \left( 1 - \frac{1}{n} \right) + 2.$$

Using this drop inequality, we are able to prove the following concentration lemma for super-exponential potentials.

**THEOREM 8.7 (RESTATED, PAGE 47).** *Consider any allocation process for which there exist super-exponential potential functions  $\Phi_1 := \Phi_1(\phi_1, z)$  and  $\Phi_2 := \Phi_2(\phi_2, z)$  with integer offset  $z := z(n) > 0$  and smoothing parameters  $\phi_1, \phi_2 \in (0, (\log n)/6]$  with  $\phi_2 \leq \frac{\phi_1}{84}$ , such that they satisfy for any step  $s \geq 0$ ,*

$$\mathbb{E} \left[ \Phi_1^{s+1} \mid \mathcal{K}^s, \mathcal{K}^s \right] \leq \Phi_1^s \cdot \left( 1 - \frac{1}{n} \right) + 2,$$

and

$$\mathbb{E} \left[ \Phi_2^{s+1} \mid \mathcal{K}^s, \mathcal{K}^s \right] \leq \Phi_2^s \cdot \left( 1 - \frac{1}{n} \right) + 2,$$

where  $\mathcal{K}^s := \mathcal{K}_{\phi_1, z}^s$ . Further, let  $P \in [n^{-4}, 1]$ . Then, for any steps  $t \geq 0$  and  $\tilde{t} \in [t, t + n \log^5 n]$ , which satisfy

$$\Pr \left[ \left\{ \text{Gap}(t - 2n \log^4 n) \leq \log^2 n \right\} \cap \bigcap_{s \in [t - 2n \log^4 n, \tilde{t}]} \mathcal{K}^s \right] \geq 1 - P,$$

they must also satisfy

$$\Pr \left[ \bigcap_{s \in [t, \tilde{t}]} \{\Phi_2^s \leq 8n\} \right] \geq 1 - (\log^8 n) \cdot P.$$

The statement of this theorem concerns steps in  $[t - 2n \log^4 n, \tilde{t}]$  with  $\tilde{t} \in [t, t + n \log^5 n]$ . The interval  $[t, \tilde{t}]$  is the *stabilization interval*, i.e., the interval where we want to show that  $\Phi_2^s \leq 8n$  for every  $s \in [t, \tilde{t}]$ . The interval  $[t - 2n \log^4 n, t]$  is the *recovery interval* where we will show that w.h.p.  $\Phi_2$  becomes  $O(n)$  at least *once*, provided we start with a “good”  $O(\log^2 n)$  gap at step  $t - 2n \log^4 n$ . For both the recovery and stabilization intervals we will condition that the event  $\mathcal{K}$  holds at every step.

Using Theorem 8.7 to prove Lemma 9.3. In the concentration lemma (Theorem 8.7), the potentials  $(\Phi_j)_{j=0}^{k-1}$  will take the role of  $\Phi_2$  and the role of  $\Phi_1$  will be taken by the potentials  $(\Psi_j)_{j=0}^{k-1}$ , defined as  $(\Phi_j)_{j=0}^{k-1}$  but with a smoothing parameter  $\alpha_1$  that is a constant factor larger, i.e.,  $\alpha_1 := 84 \cdot \alpha_2$ . More specifically,

$$\Psi_0 := \Psi_0(\alpha_1, z_0) := \sum_{i=1}^n \Psi_{0,i}^s := \sum_{i=1}^n \exp\left(\alpha_1 \cdot (y_i^s - z_0)^+\right), \quad (6.8)$$

and for  $1 \leq j \leq k-1$ ,

$$\Psi_j^s := \Psi_j^s(\alpha_1 \cdot (\log n) \cdot g^{j-k}, z_j) := \sum_{i=1}^n \Psi_{j,i}^s := \sum_{i=1}^n \exp\left(\alpha_1 \cdot (\log n) \cdot g^{j-k} \cdot (y_i^s - z_j)^+\right). \quad (6.9)$$

For the induction step (Lemma 9.3), it trivially follows that when  $\Phi_{j-1}^t = O(n)$ , then we also have that  $\text{Gap}(t) \leq \log^2 n$  and the following lemma establishes that the event  $\mathcal{K}_{\psi_j, z_j}^t$  (associated with the drop of  $\Phi_j$  and  $\Psi_j$ ) also holds.

**LEMMA 9.1 (SIMPLIFIED VERSION, PAGE 51).** *For any integer  $1 \leq j \leq k-1$  and any step  $s \geq 0$  such that  $\Phi_{j-1}^s = O(n)$  holds, then also  $\mathcal{K}_{\psi_j, z_j}^s$  holds.*

Hence, combining this lemma (Lemma 9.1) with the super-exponential potential drop (Lemma 8.1), the preconditions of super-exponential concentration lemma (Theorem 8.7) are satisfied and so layered induction step (Lemma 9.3) follows.

*Road map.* The proof of Theorem 9.2 is split into three parts. In Section 7, we strengthen the analysis of Section 5 to establish the base case (Theorem 7.8). In Section 8, we prove the general concentration for super-exponential potentials (Theorem 8.7). Finally, in Section 9, we use this concentration theorem to prove the layered induction step (Lemma 9.3) and complete the proof for the bound on the gap (Theorem 9.2).

## 7 UPPER BOUND OF $O(\frac{g}{\log g} \cdot \log \log n)$ FOR $g$ -ADV-COMP WITH $g \leq \log n$ : BASE CASE

In this section, we will obtain for the  $g$ -ADV-COMP setting a stronger guarantee for a variant of the  $\Lambda$  potential used in Section 5. The precise upper bound that we need on  $g$  is  $g \leq c_6 \log n$ , where  $c_6 > 0$  is a sufficiently small constant defined as

$$c_6 := \frac{r}{9 \cdot 20 \cdot \tilde{c}_s \cdot \log(2ce^{2\alpha_1})} \leq 1, \quad (7.1)$$

where  $r \in (0, 1)$  is as defined in Lemma 5.4,  $\tilde{c}_s := \tilde{c}_s(\alpha_1, c_4, e^{2\alpha_1}c) \geq 1$  as defined in Lemma 5.5 (for  $c_4 := 730$ ),  $c > 0$  as defined in Lemma 5.7 and  $\alpha_1 := \frac{1}{6\kappa} \leq \frac{\alpha}{6}$ , for  $\kappa > 0$  the constant defined in Equation (5.11) and  $\alpha := \frac{1}{18}$  used in Section 5.

In Section 5, we showed that w.h.p.  $\Lambda^t = O(n)$  at least *once* every  $O(n \cdot (g + \log n))$  steps. Here, we will show that w.h.p. for any step  $t$ , we have for *all* steps  $s \in [t, t + n \log^5 n]$  that  $\Psi_0^s = O(n)$ . This will serve as the base case for the layered induction in Section 9.

We start by defining the potential function  $V := V(\alpha_1, c_4g)$  which is a variant of the  $\Lambda := \Lambda(\alpha, c_4g)$  potential function (defined in Equation (5.1)), with the same offset  $c_4g = 2Dg = 730g$ , but with a smaller smoothing parameter  $\alpha_1 \leq \frac{\alpha}{6}$ ,

$$V^t := V^t(\alpha_1, c_4g) := \sum_{i=1}^n V_i^t := \sum_{i=1}^n \left[ e^{\alpha_1(y_i^t - c_4g)^+} + e^{\alpha_1(-y_i^t - c_4g)^+} \right]. \quad (7.2)$$

In Section 5, we proved that w.h.p. every  $O(n \cdot (g + \log n))$  steps the potential  $\Lambda$  satisfies  $\Lambda^s \leq cn$ . In this section, we will strengthen this to show that every  $O(ng)$  steps (for  $g = O(\log n)$ ) the potential  $V$  satisfies  $V^s \leq e^{O(\alpha_1g)} \cdot n$ . By Lemma 7.3, this implies the base case of the layered induction in Section 9, i.e., that for all steps  $s \in [m - n \log^5 n, m]$ ,  $\Psi_0^s \leq Cn$  for  $C := 2e^{2\alpha_1} \cdot c + 1$  and recalling that  $\Psi_0 := \Psi_0(\alpha_1, c_5g)$  for some sufficiently large constant  $c_5 > 0$  (to be defined in Equation (7.14) in Lemma 7.7) is given by

$$\Psi_0^t = \sum_{i=1}^n \exp\left(\alpha_1 \cdot (y_i^t - z_0)^+\right) = \sum_{i=1}^n \exp\left(\alpha_1 \cdot (y_i^t - c_5g)^+\right).$$

The proof follows along the lines of Theorem 5.12 in Section 5, but it further conditions on the gap being  $O(g + \log n)$  at every step of the analysis. In particular, by conditioning on  $\max_{i \in [n]} |y_i^t| \leq \kappa \cdot (g + \log n)$ , we obtain that  $|\Delta V^{t+1}| = O(n^{1/3})$  (Lemma 7.4), which allows us to apply Azuma's inequality (Lemma A.1) to deduce that w.h.p.  $V$  remains small. This bounded difference condition is similar to the one used in [43, Proof of Theorem 5.3].

### 7.1 A Modified Process

Let  $\mathcal{P}$  be the process in the  $g$ -ADV-COMP setting (with arbitrary  $1 \leq g \leq c_6 \log n$ ) that we want to analyze. We would like to condition on the event that  $\mathcal{P}$  satisfies  $\max_{i \in [n]} |y_i^t| \leq \kappa \cdot (g + \log n)$ , for every step  $t$  in an interval of  $2n \log^5 n$  steps, which holds w.h.p., as implied by Theorem 5.12.

We implement this conditioning by defining a *modified process*  $Q_{g,r_0} := Q_{g,r_0}(\mathcal{P})$  for the same  $g$  and some arbitrary step  $r_0$ . Consider the stopping time  $\sigma := \inf\{s \geq r_0 : \max_{i \in [n]} |y_i^s| > \kappa \cdot (g + \log n)\}$ , then the process  $Q_{g,r_0}$  is defined so that

- in steps  $s \in [0, \sigma)$  makes the same allocations as  $\mathcal{P}$ , and
- in steps  $s \in [\sigma, \infty)$  allocates to the currently least loaded bin, i.e., it uses the probability allocation vector  $r^s = (0, 0, \dots, 0, 1)$ .

Let  $y_Q^s$  be the normalized load vector of  $Q_{g,r_0}$  at step  $s \geq 0$ . By Theorem 5.12, it follows that for any interval  $[r_0, m]$  with  $m - r_0 \leq n^2$ , with high probability the two processes agree, i.e.,

$$\begin{aligned} \Pr \left[ \bigcap_{s \in [r_0, m]} \{y_Q^s = y^s\} \right] &\geq \Pr \left[ \bigcap_{s \in [r_0, m]} \left\{ \max_{i \in [n]} |y_i^s| \leq \kappa \cdot (g + \log n) \right\} \right] \\ &\geq 1 - 2 \cdot (ng)^{-9} \cdot n^2 \geq 1 - 2n^{-7}. \end{aligned} \quad (7.3)$$

The process  $Q_{g,r_0}$  is defined in a way to satisfy the following property:

- **(Property 1):** The  $Q_{g,r_0}$  process satisfies the drop inequalities for the potential functions  $\Lambda_Q$ ,  $V_Q$  and  $Y_Q$  (Lemmas 5.3, 5.6 (ii) and 5.7) for any step  $s \geq 0$ . This holds because for any step  $s < \sigma$ , the process follows  $\mathcal{P}$  and so it is an instance of the  $g$ -ADV-COMP setting. For any step  $s \geq \sigma$ , the process allocates to the currently least loaded bin and therefore minimizes the potential  $\Lambda_Q^{s+1}$  given any  $\mathfrak{F}^s$ , which means that  $\Lambda_Q^{s+1} \leq \mathbb{E}[\Lambda^{s+1} \mid \mathfrak{F}^s]$  and so it trivially satisfies any drop inequality of the original process (and similarly for  $V_Q$  and  $Y_Q$ ).



Further, we define the event that the maximum normalized load in absolute value is small at step  $r_0$  as,

$$\mathcal{Z}^{r_0} := \left\{ \max_{i \in [n]} |y_{Q,i}^{r_0}| \leq \min \{ \kappa \cdot (g + \log n), c_3 g \log(n g) \} \right\}, \quad (7.4)$$

where  $c_3 \geq 2$  is the constant defined in Equation (4.6). We are primarily interested in the  $\kappa \cdot (g + \log n)$  bound on the gap and the second bound is only needed for very small values of  $g = O(1)$ . When the event  $\mathcal{Z}^{r_0}$  holds, then the process  $Q_{g,r_0}$  also satisfies the following property (which “implements” the conditioning that the gap is  $O(g + \log n)$ ):

– **(Property 2)**: For any step  $s \geq r_0$ , it follows that

$$\max_{i \in [n]} |y_i^s| \leq \kappa \cdot (g + \log n) + 1 \leq 2\kappa \log n,$$

using that  $g \leq c_6 \log n$  with  $c_6 \leq \frac{1}{4}$  by Equation (7.1). At any step  $s \in [r_0, \sigma)$ , this holds by the definition of  $\sigma$ . For any step  $s \geq \sigma$ , a ball will never be allocated to a bin with  $y_i^s > 0$  and in every  $n$  steps the at most  $n$  bins with load equal to the minimum load (at step  $s$ ) will receive at least one ball each. Hence, over any  $n$  steps the maximum absolute normalized load does not increase and in the steps in between this can be larger by at most 1.

## 7.2 Preliminaries

We now define the adjusted potential  $\tilde{V}$ , analogously to  $\tilde{\Lambda}$  in Equation (5.6). Note that Lemma 5.7 with constants  $\varepsilon = \frac{1}{12}$ ,  $c = 12 \cdot 18$  also applies to the potential  $V$ , since  $V$  has the same form as  $\Lambda$  but a smaller smoothing parameter  $\alpha_1 \leq \alpha$ . Next, we define the sequence  $(\tilde{V}_{t_0}^s)_{s \geq t_0} := (\tilde{V}_{t_0}^s)_{s \geq t_0}(\alpha_1, c_4 g, \varepsilon)$  as  $\tilde{V}_{t_0}^{t_0} := V^{t_0}(\alpha_1, c_4 g)$  and, for any  $s > t_0$ ,

$$\tilde{V}_{t_0}^s := V^s(\alpha_1, c_4 g) \cdot \mathbf{1}_{\tilde{\mathcal{E}}_{t_0}^{s-1}} \cdot \exp\left(-\frac{3\alpha_1}{n} \cdot B_{t_0}^{s-1}\right) \cdot \exp\left(+\frac{\alpha_1 \varepsilon}{n} \cdot G_{t_0}^{s-1}\right), \quad (7.5)$$

where  $G_t^{s-1}$  (and  $B_t^{s-1}$ ) is the number of good (bad) steps in  $[t_0, s-1]$  (as defined in Section 5.5) and

$$\tilde{\mathcal{E}}_{t_0}^s := \tilde{\mathcal{E}}_{t_0}^s(V, c) := \bigcap_{t \in [t_0, s]} \{V^t > cn\}.$$

Similarly, to  $\tilde{\Lambda}$  in Section 5, we have that  $\tilde{V}$  is a super-martingale.

**LEMMA 7.1** (CF. LEMMA 5.8). *Consider the  $Q_{g,r_0}$  process for any  $g \geq 1$ , any step  $r_0 \geq 0$ , the sequence  $(\tilde{V}_{t_0}^s)_{s \geq t_0} := (\tilde{V}_{t_0}^s)_{s \geq t_0}(\alpha_1, c_4 g, \varepsilon)$  for any  $t_0 \geq r_0$  with  $\alpha_1 > 0$  as defined in Equation (6.2) and  $\varepsilon, c_4 > 0$  as defined in Lemma 5.7. For any step  $s \geq t_0$ , we have that,*

$$\mathbb{E}[\tilde{V}_{t_0}^{s+1} \mid \mathcal{F}^s] \leq \tilde{V}_{t_0}^s.$$

**PROOF.** The proof is analogous to that of Lemma 5.8, by substituting  $\Lambda$  with  $V$  and  $\tilde{\Lambda}$  with  $\tilde{V}$ . The drop inequalities follow from Lemma 5.6 and Lemma 5.7, since  $V$  has the same form as  $\Lambda$  and a smaller smoothing parameter  $\alpha_1 \leq \alpha$ . The process  $Q_{g,r_0}$  also satisfies the drop inequalities by Property 1 (see Section 7.1).  $\square$

We defer the proofs of the next two lemmas to Appendix B.4. The first one is a simple smoothness argument for the potential  $V$  defined in Equation (7.2).

**LEMMA 7.2.** *Consider the potential  $V := V(\alpha_1, c_4 g)$  for any  $\alpha_1 > 0$ , any  $c_4 > 0$  and any  $g \geq 1$ . Then, (i) for any step  $t \geq 0$ , we have that*

$$e^{-\alpha_1} \cdot V^t \leq V^{t+1} \leq e^{\alpha_1} \cdot V^t.$$

Further, (ii) for any  $\hat{c} > 0$ , for any integer  $T > 0$  and any step  $t \geq 0$ , for which there exist steps  $s_0 \in [t - T, t]$  and  $s_1 \in [t, t + T]$ , such that  $V^{s_0} \leq \hat{c}n$  and  $V^{s_1} \leq \hat{c}n$ , we have that

$$V^t \leq e^{\alpha_1 \frac{T}{n}} \cdot 2\hat{c}n.$$

The following lemma shows that by choosing a large enough offset  $c_5 > 0$  in the potential  $\Psi_0 := \Psi_0(\alpha_1, c_5g)$  (defined in Equation (6.8)), when  $V^t = e^{O(\alpha_1g)} \cdot cn$ , then  $\Psi_0^t = O(n)$ .

LEMMA 7.3. Consider any  $c, \hat{c} > 0$  and the potential  $V := V(\alpha_1, c_4g)$  for any  $\alpha_1 > 0$ , any  $c_4 > 0$  and any  $g \geq 1$ . Further, consider the potential  $\Psi_0 := \Psi_0(\alpha_1, c_5g)$  with offset  $c_5 := 2 \cdot \max\{c_4, \hat{c}\}$  and  $C := 2e^{2\alpha_1} \cdot c + 1$ . Then, for any step  $t \geq 0$  with  $V^t \leq e^{\alpha_1 \cdot \hat{c} \cdot g} \cdot 2e^{2\alpha_1}cn$ , it holds that  $\Psi_0^t \leq Cn$ .

In comparison to Section 5, where we proved stabilization over an interval of length  $\Omega(n \cdot \max\{\log n, g\})$ , here we will be using a shorter interval of length

$$\tilde{\Delta}_s := \frac{20 \cdot \tilde{c}_s \cdot \log(2ce^{2\alpha_1})}{\alpha_1 \varepsilon} \cdot ng, \quad (7.6)$$

where constants  $\varepsilon := \frac{1}{12}$ ,  $r := \frac{6}{6+\varepsilon} > 0$  are as defined in Section 5 and  $\tilde{c}_s := \tilde{c}_s(\alpha_1, c_4, e^{2\alpha_1}c) \geq 1$  is defined in Lemma 5.5.

We now prove the bounded difference condition for the  $\tilde{V}$  potential. This follows from the second property of  $\mathcal{Q}_{g,r_0}$  that the maximum normalized load in absolute value is  $O(g + \log n)$  for any step  $s \geq r_0$ .

LEMMA 7.4. Consider the  $\mathcal{Q}_{g,r_0}$  process for any  $g \in [1, c_6 \log n]$  with  $c_6 > 0$  as defined in Equation (7.1), any step  $r_0 \geq 0$ , and  $\mathcal{Z}^{r_0}$  as defined in Equation (7.4). Further, consider the sequence  $(\tilde{V}_{t_0}^s)_{s \geq t_0} := (\tilde{V}_{t_0}^s)_{s \geq t_0}(\alpha_1, c_4g, \varepsilon)$  for any step  $t_0 \geq r_0$  with  $\alpha_1 > 0$  as defined in Equation (6.2) and  $\varepsilon, c_4 > 0$  as defined in Lemma 5.7. Then, for any step  $s \geq t_0 \geq r_0$  we have that  $\tilde{V}_{t_0}^{s+1} = 0$  or

$$(\left| \Delta \tilde{V}_{t_0}^{s+1} \right| \mid \mathcal{Z}^{r_0}, \tilde{\mathcal{G}}^s) \leq 16 \cdot e^{\alpha_1 \varepsilon \cdot \frac{s-t_0}{n}} \cdot n^{1/3}.$$

PROOF. Consider an arbitrary step  $s \geq r_0$  and assume that the event  $\mathcal{Z}^{r_0}$  holds. By Property 2 of the  $\mathcal{Q}_{g,r_0}$  process (see Section 7.1), we have that

$$\max_{i \in [n]} |y_i^s| \leq \kappa \cdot (g + \log n) + 1,$$

which also implies for  $c_4 := 730 > 0$ , since  $g \leq c_6 \log n \leq \log n$  and  $\kappa > 1$  that

$$\max_{i \in [n]} \{(y_i^s - c_4g)^+, (-y_i^s - c_4g)^+\} \leq \kappa \cdot (g + \log n) + 1 - c_4g \leq 2\kappa \log n. \quad (7.7)$$

We will now show that  $|\Delta V^{s+1}| \leq 5n^{1/3}$  and then use this to bound  $|\Delta \tilde{V}^{s+1}|$ . By Equation (7.7) for any bin  $i \in [n]$ ,

$$V_i^s \leq 2 \cdot e^{2\alpha_1 \kappa \log n} = 2 \cdot e^{\frac{1}{3} \cdot \log n} = 2n^{1/3},$$

using that  $\alpha_1 := \frac{1}{6\kappa}$ . Hence, by aggregating over all bins,  $V^s \leq 2n^{4/3}$ . Furthermore, if the ball at step  $s+1$  is allocated to bin  $j \in [n]$ , then

$$\Delta V^{s+1} \leq e^{\alpha_1/n} \cdot V^s + e^{\alpha_1} \cdot V_j^s - V^s \leq \frac{2\alpha_1}{n} V^s + 2 \cdot 2n^{1/3} \leq \frac{2\alpha_1}{n} \cdot (2n^{4/3}) + 4n^{1/3} \leq 5n^{1/3},$$

using that  $e^{\alpha_1/n} \leq 1 + 2 \cdot \frac{\alpha_1}{n}$  and  $e^{\alpha_1} \leq 2$ , which both hold as  $\alpha_1 \leq 1/4$ . Similarly,

$$\Delta V^{s+1} \geq e^{-\alpha_1/n} \cdot V^s - e^{\alpha_1} \cdot V_j^s - V^s \geq -\frac{\alpha_1}{n} V^s - 2 \cdot 2n^{1/3} \geq -\frac{\alpha_1}{n} \cdot (2n^{4/3}) - 4n^{1/3} \geq -5n^{1/3},$$

using that  $e^{-\alpha_1/n} \geq 1 - \frac{\alpha_1}{n}$  and  $e^{\alpha_1} \leq 2$ , as  $\alpha_1 \leq 1/4$ .

Now, we turn to upper bounding  $|\Delta \widetilde{V}_{t_0}^{s+1}|$  by proving lower and upper bounds on  $\widetilde{V}_{t_0}^{s+1}$ . If  $\widetilde{V}_{t_0}^{s+1} = 0$ , then the conclusion follows. Otherwise, since  $\widetilde{V}_{t_0}^{s+1} > 0$ , we have that  $1_{\widetilde{\mathcal{E}}_{t_0}^s} = 1_{\widetilde{\mathcal{E}}_{t_0}^{s-1}} = 1$ , so by definition of  $\widetilde{V}$  in Equation (7.5),

$$\widetilde{V}_{t_0}^s = V^s \cdot \exp\left(-\frac{3\alpha_1}{n} \cdot B_{t_0}^{s-1}\right) \cdot \exp\left(+\frac{\alpha_1 \varepsilon}{n} \cdot G_{t_0}^{s-1}\right) \leq 2 \cdot e^{\alpha_1 \varepsilon \cdot \frac{s-t_0}{n}} \cdot n^{4/3}, \quad (7.8)$$

using that  $G_{t_0}^{s-1} \leq s - t_0$  and  $V^s \leq 2n^{4/3}$ .

Now, we upper bound  $\widetilde{V}_{t_0}^{s+1}$ , recalling that  $1_{\widetilde{\mathcal{E}}_{t_0}^s} = 1$ ,

$$\begin{aligned} \widetilde{V}_{t_0}^{s+1} &= V^{s+1} \cdot \exp\left(-\frac{3\alpha_1}{n} \cdot B_{t_0}^s\right) \cdot \exp\left(+\frac{\alpha_1 \varepsilon}{n} \cdot G_{t_0}^s\right) \\ &\leq (V^s + 5n^{1/3}) \cdot \exp\left(-\frac{3\alpha_1}{n} \cdot B_{t_0}^s\right) \cdot \exp\left(+\frac{\alpha_1 \varepsilon}{n} \cdot G_{t_0}^s\right) \\ &\leq (V^s + 5n^{1/3}) \cdot \exp\left(-\frac{3\alpha_1}{n} \cdot B_{t_0}^{s-1}\right) \cdot \exp\left(+\frac{\alpha_1 \varepsilon}{n} \cdot G_{t_0}^{s-1}\right) \cdot \exp\left(\frac{\alpha_1 \varepsilon}{n}\right) \\ &= \widetilde{V}_{t_0}^s \cdot \exp\left(\frac{\alpha_1 \varepsilon}{n}\right) + 5 \cdot n^{1/3} \cdot \exp\left(-\frac{3\alpha_1}{n} \cdot B_{t_0}^{s-1}\right) \cdot \exp\left(+\frac{\alpha_1 \varepsilon}{n} \cdot G_{t_0}^{s-1}\right) \cdot \exp\left(\frac{\alpha_1 \varepsilon}{n}\right) \\ &\stackrel{(a)}{\leq} \widetilde{V}_{t_0}^s \cdot \left(1 + \frac{3\alpha_1}{n}\right) + 5 \cdot n^{1/3} \cdot e^{\alpha_1 \varepsilon \cdot \frac{s-t_0}{n}} \cdot 2 \\ &\stackrel{(b)}{\leq} \widetilde{V}_{t_0}^s + (2 \cdot e^{\alpha_1 \varepsilon \cdot \frac{s-t_0}{n}} \cdot n^{4/3}) \cdot \frac{3\alpha_1}{n} + 10 \cdot n^{1/3} \cdot e^{\alpha_1 \varepsilon \cdot \frac{s-t_0}{n}} \\ &\leq \widetilde{V}_{t_0}^s + 16 \cdot e^{\alpha_1 \varepsilon \cdot \frac{s-t_0}{n}} \cdot n^{1/3}, \end{aligned} \quad (7.9)$$

using in (a) that  $e^{\alpha_1 \varepsilon / n} \leq 1 + \frac{3\alpha_1}{n}$  as  $\alpha_1 \leq 1$  and  $\varepsilon = \frac{1}{12}$ ,  $G_{t_0}^{s-1} \leq s - t_0$ , and  $e^{\alpha_1 \varepsilon / n} \leq 2$  and in (b) using Equation (7.8).

Similarly, we lower bound  $\widetilde{V}_{t_0}^{s+1}$ ,

$$\begin{aligned} \widetilde{V}_{t_0}^{s+1} &= V^{s+1} \cdot \exp\left(-\frac{3\alpha_1}{n} \cdot B_{t_0}^s\right) \cdot \exp\left(+\frac{\alpha_1 \varepsilon}{n} \cdot G_{t_0}^s\right) \\ &\geq (V^s - 5n^{1/3}) \cdot \exp\left(-\frac{3\alpha_1}{n} \cdot B_{t_0}^s\right) \cdot \exp\left(+\frac{\alpha_1 \varepsilon}{n} \cdot G_{t_0}^s\right) \\ &\stackrel{(a)}{\geq} (V^s - 5n^{1/3}) \cdot \exp\left(-\frac{3\alpha_1}{n} \cdot B_{t_0}^{s-1}\right) \cdot \exp\left(+\frac{\alpha_1 \varepsilon}{n} \cdot G_{t_0}^{s-1}\right) \cdot \exp\left(-\frac{3\alpha_1}{n}\right) \\ &= \widetilde{V}_{t_0}^s \cdot \exp\left(-\frac{3\alpha_1}{n}\right) - 5 \cdot n^{1/3} \cdot \exp\left(-\frac{3\alpha_1}{n} \cdot B_{t_0}^{s-1}\right) \cdot \exp\left(+\frac{\alpha_1 \varepsilon}{n} \cdot G_{t_0}^{s-1}\right) \cdot \exp\left(-\frac{3\alpha_1}{n}\right) \\ &\stackrel{(b)}{\geq} \widetilde{V}_{t_0}^s \cdot \left(1 - \frac{3\alpha_1}{n}\right) - 5 \cdot n^{1/3} \cdot e^{\alpha_1 \varepsilon \cdot \frac{s-t_0}{n}} \\ &\stackrel{(c)}{\geq} \widetilde{V}_{t_0}^s - (2 \cdot e^{\alpha_1 \varepsilon \cdot \frac{s-t_0}{n}} \cdot n^{4/3}) \cdot \frac{3\alpha_1}{n} - 5 \cdot n^{1/3} \cdot e^{\alpha_1 \varepsilon \cdot \frac{s-t_0}{n}} \\ &\geq \widetilde{V}_{t_0}^s - 16 \cdot e^{\alpha_1 \varepsilon \cdot \frac{s-t_0}{n}} \cdot n^{1/3}, \end{aligned} \quad (7.10)$$

using in (a) that  $V^s \geq 2n \geq 5n^{1/3}$  holds deterministically, in (b) that  $e^{-3\alpha_1/n} \geq 1 - \frac{3\alpha_1}{n}$  and  $G_{t_0}^{s-1} \leq s - t_0$  and in (c) using Equation (7.8).

Hence, combining the two upper bounds in Equations (7.9) and (7.10), we conclude that  $|\Delta \widetilde{V}_{t_0}^{s+1}| \leq 16 \cdot e^{\alpha_1 \varepsilon \cdot \frac{s-t_0}{n}} \cdot n^{1/3}$ .  $\square$

### 7.3 Strong Stabilization

We will now prove the following slightly stronger version of Lemma 5.10, meaning that stabilization is over intervals of length  $\Theta(n\mathfrak{g})$  instead of  $\Theta(n \cdot (g + \log n))$ .

**LEMMA 7.5 (STRONG STABILIZATION).** *Consider the  $\mathcal{Q}_{g,r_0}$  process for any  $g \in [1, c_6 \log n]$  for  $c_6 > 0$  as defined in Equation (7.1), any step  $r_0 \geq 0$  and  $\mathcal{Z}^{r_0}$  as defined in Equation (7.4). Then, for the potential  $V := V(\alpha_1, c_4 g)$  with  $\alpha_1 > 0$  as defined in Equation (6.2),  $c_4, c > 0$  as defined in Lemma 5.7 and  $\tilde{\Delta}_s > 0$  as defined in Equation (7.6), it holds that for any step  $t_0 \geq r_0$ ,*

$$\Pr \left[ \bigcup_{s \in [t_0, t_0 + \tilde{\Delta}_s]} \{V^s \leq e^{\alpha_1 cn}\} \mid \mathcal{Z}^{r_0}, \mathfrak{F}^{t_0}, e^{\alpha_1 cn} < V^{t_0} \leq e^{2\alpha_1 cn} \right] \geq 1 - n^{-11}.$$

**PROOF.** The proof of this lemma proceeds similarly to that of Lemma 5.10, but we will apply Azuma's inequality for  $\tilde{V}_{t_0}^s$  instead of Markov's inequality. However, we cannot directly apply concentration to  $\tilde{V}_{t_0}^s$  because the bounded difference condition (Lemma 7.4) holds only when  $\tilde{V}_{t_0}^s$  is positive. So instead, we apply it to a stopped random variable  $X_{t_0}^s$  to be defined in a way that ensures it is always positive.

Let  $t_1 := t_0 + \tilde{\Delta}_s$ . We define the stopping time  $\tau := \inf\{s \geq t_0 : V^s \leq e^{\alpha_1 cn}\}$  and for any  $s \in [t_0, t_1]$ ,

$$X_{t_0}^s := \tilde{V}_{t_0}^{s \wedge \tau}.$$

We will now verify that  $X_{t_0}^s > 0$  for all  $s \in [t_0, t_1]$ . Firstly, consider any  $s < \tau$ . Since  $V^s > e^{\alpha_1 cn}$ , by Lemma 7.2 (i), we have that  $V^{s+1} \geq V^s \cdot e^{-\alpha_1} > cn$  and hence  $X_{t_0}^{s+1} = \tilde{V}_{t_0}^{s+1} > 0$ . Secondly, for any  $s \geq \tau$ , it trivially holds that  $X_{t_0}^{s+1} = X_{t_0}^s > 0$ .

We proceed to verify the preconditions of Azuma's inequality for super-martingales (Lemma A.1) for the sequence  $(X_{t_0}^s)_{s \in [t_0, t_1]}$ . Firstly, using Lemma 7.1 it forms a super-martingale, i.e., that  $\mathbb{E}[X_{t_0}^s \mid \mathfrak{F}^{s-1}] \leq X_{t_0}^{s-1}$ . Secondly, by Lemma 7.4, since  $X_{t_0}^s > 0$ , for any filtration  $\mathfrak{F}^{s-1}$  where  $\mathcal{Z}^{r_0}$  holds, we have that

$$\begin{aligned} \left( |X_{t_0}^s - X_{t_0}^{s-1}| \mid \mathcal{Z}^{r_0}, \mathfrak{F}^{s-1} \right) &\leq \left( |\tilde{V}_{t_0}^s - \tilde{V}_{t_0}^{s-1}| \mid \mathcal{Z}^{r_0}, \mathfrak{F}^{s-1} \right) \\ &\leq 16 \cdot e^{\alpha_1 \varepsilon \cdot \frac{\tilde{\Delta}_s}{n}} \cdot n^{1/3} \\ &\stackrel{(a)}{=} (16 \cdot e^{20 \cdot \tilde{c}_s \cdot r^{-1} \cdot \log(2ce^{2\alpha_1}g)}) \cdot n^{1/3} \\ &\stackrel{(b)}{\leq} 16 \cdot e^{20 \cdot \tilde{c}_s \cdot r^{-1} \cdot \log(2ce^{2\alpha_1}) \cdot c_6 \log n} \cdot n^{1/3} \\ &= 16 \cdot n^{4/9}, \end{aligned}$$

using in (a) that  $\tilde{\Delta}_s := \frac{20 \cdot \tilde{c}_s \cdot \log(2ce^{2\alpha_1})}{\alpha_1 \varepsilon r} \cdot ng$  and in (b) that  $g \leq c_6 \log n$  and  $c_6 := \frac{r}{9 \cdot 20 \cdot \tilde{c}_s \cdot \log(2ce^{2\alpha_1})}$ .

Hence, applying Lemma A.1 for  $\lambda := n$ ,  $N := \tilde{\Delta}_s$  and  $a_i := 16 \cdot n^{4/9}$ ,

$$\begin{aligned} \Pr \left[ X_{t_0}^{t_1} \geq X_{t_0}^{t_0} + n \mid \mathcal{Z}^{r_0}, \mathfrak{F}^{t_0}, e^{\alpha_1 cn} < V^{t_0} \leq e^{2\alpha_1 cn} \right] &\leq \exp \left( -\frac{n^2}{2 \cdot \tilde{\Delta}_s \cdot (16 \cdot n^{4/9})^2} \right) \\ &\leq n^{-\omega(1)}, \end{aligned} \tag{7.11}$$

using that  $\tilde{\Delta}_s = O(n\mathfrak{g}) = O(n \log n)$ .

As we condition on  $\{V^{t_0} \leq e^{2\alpha_1 cn}\}$ , we have that

$$\max_{i \in [n]} |y_i^{t_0}| \leq c_4 g + \frac{\log(e^{2\alpha_1 cn})}{\alpha_1} \leq g(\log(n\mathfrak{g}))^2,$$

for sufficiently large  $n$ , using that  $\alpha_1, c_5, c > 0$  are constants. Further, by Lemma 5.5 (i) (since  $V$  has the same form as  $\Lambda$ ), we have that  $Y^{t_0} \leq \tilde{c}_s n g^2$ , for some constant  $\tilde{c}_s := \tilde{c}_s(\alpha_1, c_4, e^{2\alpha_1} c) \geq 1$ .

Applying Lemma 5.4 for  $T := \tilde{c}_s n g^2$  (since  $T \in [n g^2, o(n^2 g^3)]$ ) and  $\hat{c} := \frac{\tilde{\Delta}_s \cdot g}{T} \geq \frac{20}{\alpha_1 \varepsilon r} \geq 1$  (since  $\alpha_1, \varepsilon, r \leq 1$ ),

$$\begin{aligned} & \Pr \left[ G_{t_0}^{t_1-1} \geq r \cdot \tilde{\Delta}_s \mid \mathcal{Z}^{r_0}, \mathfrak{F}^{t_0}, e^{\alpha_1} c n < V^{t_0} \leq e^{2\alpha_1} c n \right] \\ & \geq \Pr \left[ G_{t_0}^{t_1-1} \geq r \cdot \tilde{\Delta}_s \mid \mathfrak{F}^{t_0}, Y^{t_0} \leq T, \max_{i \in [n]} |y_i^{t_0}| \leq g(\log(n g))^2 \right] \\ & \geq 1 - 2 \cdot n^{-12}. \end{aligned} \quad (7.12)$$

Taking the union bound over Equations (7.11) and (7.12), we have

$$\Pr \left[ \{X_{t_0}^{t_1} < X_{t_0}^{t_0} + n\} \cap \{G_{t_0}^{t_1-1} \geq r \cdot \tilde{\Delta}_s\} \mid \mathcal{Z}^{r_0}, \mathfrak{F}^{t_0}, e^{\alpha_1} c n < V^{t_0} \leq e^{2\alpha_1} c n \right] \geq 1 - n^{-11}.$$

Assume that  $\{X_{t_0}^{t_1} < X_{t_0}^{t_0} + n\}$  and  $\{G_{t_0}^{t_1-1} \geq r \cdot \tilde{\Delta}_s\}$  hold. Then, we consider two cases based on whether the stopping time occurred before  $t_1$ .

**Case 1** [ $\tau \leq t_1$ ]: Here, clearly there is an  $s \in [t_0, t_1]$ , namely  $s = \tau$ , such that  $V^s \leq e^{\alpha_1} c n$  and the conclusion follows.

**Case 2** [ $\tau > t_1$ ]: Here, using that  $\{\tau > t_1\}$  and  $\{X_{t_0}^{t_1} < X_{t_0}^{t_0} + n\}$  both hold, it follows that

$$\tilde{V}_{t_0}^{t_1} < \tilde{V}_{t_0}^{t_0} + n,$$

and so, by definition of  $\tilde{V}_{t_0}^{t_1}$  (Equation (7.5)),

$$V^{t_1} \cdot \mathbf{1}_{\tilde{\mathcal{E}}_{t_0}^{t_1-1}} \cdot \exp \left( -\frac{3\alpha_1}{n} \cdot B_{t_0}^{t_1-1} \right) \cdot \exp \left( +\frac{\alpha_1 \varepsilon}{n} \cdot G_{t_0}^{t_1-1} \right) < V^{t_0} + n.$$

By re-arranging and using that  $\{G_{t_0}^{t_1-1} \geq r \cdot \tilde{\Delta}_s\}$  holds, we have that

$$\begin{aligned} V^{t_1} \cdot \mathbf{1}_{\tilde{\mathcal{E}}_{t_0}^{t_1-1}} & < (V^{t_0} + n) \cdot \exp \left( \frac{3\alpha_1}{n} \cdot B_{t_0}^{t_1-1} - \frac{\alpha_1 \varepsilon}{n} \cdot G_{t_0}^{t_1-1} \right) \\ & \leq (e^{2\alpha_1} c n + n) \cdot \exp \left( \frac{3\alpha_1}{n} \cdot (1-r) \cdot \tilde{\Delta}_s - \frac{\alpha_1 \varepsilon}{n} \cdot r \cdot \tilde{\Delta}_s \right) \\ & \stackrel{(a)}{=} 2e^{2\alpha_1} \cdot c n \cdot \exp \left( \frac{\alpha_1 \varepsilon}{n} \cdot \frac{r}{2} \cdot \tilde{\Delta}_s - \frac{\alpha_1 \varepsilon}{n} \cdot r \cdot \tilde{\Delta}_s \right) \\ & = 2e^{2\alpha_1} \cdot c n \cdot \exp \left( -\frac{\alpha_1 \varepsilon}{n} \cdot \frac{r}{2} \cdot \tilde{\Delta}_s \right) \\ & = 2e^{2\alpha_1} \cdot c n \cdot \exp \left( -\frac{\alpha_1 \varepsilon}{n} \cdot \frac{r}{2} \cdot \frac{20 \cdot \tilde{c}_s \cdot \log(2ce^{2\alpha_1})}{\alpha_1 \varepsilon r} \cdot n g \right) \\ & \stackrel{(b)}{\leq} 2e^{2\alpha_1} \cdot c n \cdot \exp \left( -10 \log(2ce^{2\alpha_1}) \right) \\ & \stackrel{(c)}{\leq} n/2, \end{aligned}$$

where we used in (a) that  $r = \frac{6}{6+\varepsilon}$  implies  $\frac{3\alpha_1}{n} \cdot (1-r) = \frac{3\alpha_1}{n} \cdot \frac{\varepsilon}{6+\varepsilon} = \frac{\alpha_1 \varepsilon}{n} \cdot \frac{r}{2}$  and  $c \geq 1$ , in (b) that  $g \geq 1$ ,  $c \geq 1$  and  $\tilde{c}_s \geq 1$  and in (c) that  $c = 18 \cdot 12$ . Since deterministically we have that  $V^{t_1} \geq n$ , it must be that  $\mathbf{1}_{\tilde{\mathcal{E}}_{t_0}^{t_1-1}} = 0$ , implying that there exists  $s \in [t_0, t_1]$  such that  $V^s \leq c n \leq e^{\alpha_1} c n$ .  $\square$

We will now show that the potential  $V$  becomes small every  $\Theta(n g)$  steps. The proof, given in Appendix B.4, proceeds similarly to the proof of Lemma 5.11.

LEMMA 7.6. Consider the  $\mathcal{Q}_{g,r_0}$  process for any  $g \in [1, c_6 \log n]$  for  $c_6 > 0$  as defined in Equation (7.1) and any step  $r_0 \geq 0$ . Then, for the potential  $V := V(\alpha_1, c_4 g)$  with  $\alpha_1$  as defined in Equation (6.2),  $c_4, c > 0$  as defined in Lemma 5.7 and  $\tilde{\Delta}_s > 0$  as defined in Equation (7.6), it holds that for any step  $t_0 \geq r_0$  and  $t_1$  such that  $t_0 < t_1 \leq t_0 + 2n \log^5 n$ ,

$$\Pr \left[ \bigcap_{t \in [t_0, t_1]} \bigcup_{s \in [t, t + \tilde{\Delta}_s]} \{V^s \leq e^{2\alpha_1 cn}\} \mid \mathcal{Z}^{r_0}, \mathcal{F}^{t_0}, V^{t_0} \leq cn \right] \geq 1 - n^{-9},$$

#### 7.4 Completing the Proof of Theorem 7.8

In this section, we will complete the base case proof using the stronger stabilization for the  $V$  potential (see Lemma 7.5) and upper bounding  $\Psi_0$  using  $V$  (see Lemma 7.3). We will first prove the result for the modified process and then relate the results to the  $g$ -ADV-COMP setting.

LEMMA 7.7. Consider the  $\mathcal{Q}_{g,r_0}$  process for any  $g \in [1, c_6 \log n]$  and any step  $r_0 \geq -\Delta_r - n \log^5 n$ , where  $c_6 > 0$  is as defined in Equation (7.1) and  $\Delta_r := \Delta_r(g) > 0$  is as in Lemma 5.9. Further, let  $c > 0$  be as defined in Lemma 5.7 and  $\alpha_1 > 0$  as in Equation (6.2), then for the constant  $C := 2e^{2\alpha_1 \cdot c} + 1 \geq 8$  and the potential  $\Psi_0 := \Psi_0(\alpha_1, c_5 g)$  with constant integer  $c_5 > 0$  (to be defined in Equation (7.14)), we have that,

$$\Pr \left[ \bigcap_{s \in [r_0 + \Delta_r, r_0 + \Delta_r + n \log^5 n]} \{\Psi_0^s \leq Cn\} \mid \mathcal{F}^{r_0}, \mathcal{Z}^{r_0} \right] \geq 1 - n^{-8}.$$

PROOF. This proof proceeds similarly to that of Theorem 5.12 having a recovery and a stabilization phase to show that the potential  $V := V(\alpha_1, c_4 g)$  stabilizes at  $\leq e^{O(\alpha_1 g)} \cdot cn$ . By Lemma 7.3, this implies that  $\Psi_0 := \Psi_0(\alpha_1, c_5 g)$  stabilizes at  $\leq Cn$  for sufficiently large constants  $c_5, C > 0$ .

Consider first the case where  $r_0 \geq 0$ . For the recovery phase, we will use the potential function  $\Lambda := \Lambda(\alpha, c_4 g)$  defined in Equation (5.1). More specifically, we have

$$\begin{aligned} & \Pr \left[ \bigcup_{t \in [r_0, r_0 + \Delta_r]} \{V^t \leq cn\} \mid \mathcal{F}^{r_0}, \mathcal{Z}^{r_0} \right] \\ & \stackrel{(a)}{\geq} \Pr \left[ \bigcup_{t \in [r_0, r_0 + \Delta_r]} \{\Lambda^t \leq cn\} \mid \mathcal{F}^{r_0}, \mathcal{Z}^{r_0} \right] \\ & \stackrel{(b)}{\geq} \Pr \left[ \bigcup_{t \in [r_0, r_0 + \Delta_r]} \{\Lambda^t \leq cn\} \mid \mathcal{F}^{r_0}, \max_{i \in [n]} |y_i^{r_0}| \leq c_3 g \log(ng) \right] \\ & \stackrel{(c)}{\geq} 1 - 3 \cdot (ng)^{-12} \geq 1 - n^{-11}, \end{aligned}$$

using in (a) that  $V^t \leq \Lambda^t$  for any step  $t \geq 0$ , as  $V$  and  $\Lambda$  have the same form, but  $V$  has a smoothing parameter  $\alpha_1 < \alpha$ , in (b) recalling that  $\mathcal{Z}^{r_0} := \{\max_{i \in [n]} |y_i^{r_0}| \leq \min\{\kappa \cdot (g + \log n), c_3 g \log(ng)\}\}$  for  $c_3 \geq 2$  the constant defined in Equation (4.6) and in (c) applying Lemma 5.9 (i).

Therefore, for the stopping time  $\tau := \inf\{t \geq r_0 : V^t \leq cn\}$ , it holds that

$$\Pr [\tau \leq r_0 + \Delta_r \mid \mathcal{F}^{r_0}, \mathcal{Z}^{r_0}] \geq 1 - n^{-11}. \quad (7.13)$$

Consider any  $t_0 \in [r_0, r_0 + \Delta_r]$  ( $t_0$  will play the role of a concrete value of  $\tau$ ). By Lemma 7.6 (for  $t_0 := t_0$  and  $t_1 := r_0 + \Delta_r + n \log^5 n$ , since  $t_1 - t_0 \leq n \log^5 n + \Delta_r \leq 2n \log^5 n$ ), we have that

$$\Pr \left[ \bigcap_{t \in [t_0, r_0 + \Delta_r + n \log^5 n]} \bigcup_{s \in [t, t + \tilde{\Delta}_s]} \{V^s \leq e^{2\alpha_1} cn\} \mid \mathcal{Z}^{r_0}, \mathcal{F}^{t_0}, V^{t_0} \leq cn \right] \geq 1 - n^{-9}.$$

When the above event holds, then for every  $t \in [t_0, r_0 + \Delta_r + n \log^5 n]$ , there exists  $s_0 \in [t - \tilde{\Delta}_s, t]$  and  $s_1 \in [t, t + \tilde{\Delta}_s]$  such that  $V^{s_0} \leq e^{2\alpha_1} cn$  and  $V^{s_1} \leq e^{2\alpha_1} cn$ , using that for any  $t \in [t_0, t_0 + \tilde{\Delta}_s]$  we can set  $s_0 = t_0$  since by the conditioning  $V^{t_0} \leq cn \leq e^{2\alpha_1} cn$ . So, by Lemma 7.2 (ii) (for  $\hat{c} := e^{2\alpha_1} c$  and  $T := \tilde{\Delta}_s$ ), it follows that

$$\Pr \left[ \bigcap_{t \in [t_0, r_0 + \Delta_r + n \log^5 n]} \left\{ V^t \leq e^{\alpha_1 \frac{\tilde{\Delta}_s}{n}} \cdot 2e^{2\alpha_1} cn \right\} \mid \mathcal{Z}^{r_0}, \mathcal{F}^{t_0}, V^{t_0} \leq cn \right] \geq 1 - n^{-9}.$$

Next, adjusting the range of the big intersection using that  $t_0 \leq r_0 + \Delta_r$ , it follows that

$$\Pr \left[ \bigcap_{t \in [r_0 + \Delta_r, r_0 + \Delta_r + n \log^5 n]} \left\{ V^t \leq e^{\alpha_1 \frac{\tilde{\Delta}_s}{n}} \cdot 2e^{2\alpha_1} cn \right\} \mid \mathcal{Z}^{r_0}, \mathcal{F}^{t_0}, V^{t_0} \leq cn \right] \geq 1 - n^{-9}.$$

By Lemma 7.3 (for  $\hat{c} := \lceil \frac{\tilde{\Delta}_s}{ng} \rceil$ ), we conclude that for  $\Psi_0 := \Psi_0(\alpha_1, c_5 g)$  for the constant integer

$$c_5 := 2 \cdot \max \left\{ c_4, \left\lceil \frac{\tilde{\Delta}_s}{ng} \right\rceil \right\} = 2 \cdot \max \left\{ c_4, \left\lceil \frac{20 \cdot \tilde{c}_s \cdot \log(2ce^{2\alpha_1})}{\alpha_1 \varepsilon r} \right\rceil \right\}, \quad (7.14)$$

and for the constant  $C := 2e^{2\alpha_1} \cdot c + 1$ , it holds that

$$\Pr \left[ \bigcap_{t \in [r_0 + \Delta_r, r_0 + \Delta_r + n \log^5 n]} \left\{ \Psi_0^t \leq Cn \right\} \mid \mathcal{Z}^{r_0}, \mathcal{F}^{t_0}, V^{t_0} \leq cn \right] \geq 1 - n^{-9}. \quad (7.15)$$

Finally,

$$\begin{aligned} & \Pr \left[ \bigcap_{t \in [r_0 + \Delta_r, r_0 + \Delta_r + n \log^5 n]} \left\{ \Psi_0^t \leq Cn \right\} \mid \mathcal{Z}^{r_0}, \mathcal{F}^{r_0} \right] \\ & \geq \sum_{t_0=r_0}^{r_0+\Delta_r} \Pr \left[ \bigcap_{t \in [r_0 + \Delta_r, r_0 + \Delta_r + n \log^5 n]} \left\{ \Psi_0^t \leq Cn \right\} \mid \mathcal{F}^{r_0}, \mathcal{Z}^{r_0}, \tau = t_0 \right] \cdot \Pr[\tau = t_0 \mid \mathcal{Z}^{r_0}, \mathcal{F}^{r_0}] \\ & \geq \sum_{t_0=r_0}^{r_0+\Delta_r} \Pr \left[ \bigcap_{t \in [r_0 + \Delta_r, r_0 + \Delta_r + n \log^5 n]} \left\{ \Psi_0^t \leq Cn \right\} \mid \mathcal{Z}^{r_0}, \mathcal{F}^{t_0}, V^{t_0} \leq cn \right] \cdot \Pr[\tau = t_0 \mid \mathcal{Z}^{r_0}, \mathcal{F}^{r_0}] \\ & \stackrel{(7.15)}{\geq} (1 - n^{-9}) \cdot \Pr[\tau \leq r_0 + \Delta_r \mid \mathcal{Z}^{r_0}, \mathcal{F}^{r_0}] \\ & \stackrel{(7.13)}{\geq} (1 - n^{-9}) \cdot (1 - n^{-11}) \geq 1 - n^{-8}. \end{aligned}$$

This concludes the claim for  $r_0 \geq 0$ .

If  $r_0 < 0$ , then deterministically  $\tau = 0$ , since  $V^\tau = n \leq cn$ . Hence, the rest of the proof follows for an interval of length at most  $2n \log^5 n$ .  $\square$

**THEOREM 7.8 (BASE CASE).** *Consider the  $g$ -ADV-COMP setting for any  $g \in [1, c_6 \log n]$ , where  $c_6 > 0$  is as defined in Equation (7.1). For constant  $C := 2e^{2\alpha_1} \cdot c + 1 \geq 8$  with  $\alpha_1 > 0$  as defined in Equation (6.2) and  $c > 0$  as in Lemma 5.7, and the potential  $\Phi_0 := \Phi_0(\alpha_2, c_5 g)$  with  $\alpha_2 > 0$  as in Equation (6.3), and constant integer  $c_5 > 0$  as in Equation (7.14), we have that for any step  $m \geq 0$ ,*

$$\Pr \left[ \bigcap_{s \in [m - n \log^5 n, m]} \{\Phi_0^s \leq Cn\} \right] \geq 1 - n^{-4}.$$

**PROOF.** Let  $\mathcal{P}$  be the original process in the  $g$ -ADV-COMP setting. Consider the modified process  $Q_{g, r_0}$  for  $r_0 := m - n \log^5 n - \Delta_r$  where  $\Delta_r := \Theta(ng \cdot (\log(ng))^2)$  is as defined in Lemma 5.9. Let  $y_Q$  be its normalized load vector and  $\Psi_{Q,0}$  be its  $\Psi_0$  potential function. Recall from Equation (7.3) that the two processes  $Q_{g, r_0}$  and  $\mathcal{P}$  agree at every step w.h.p., since  $m - r_0 \leq 2n \log^5 n$ ,

$$\Pr \left[ \bigcap_{s \in [r_0, m]} \{y_Q^s = y^s\} \right] \geq 1 - 2n^{-7}. \quad (7.16)$$

Taking the union bound of the conclusions in Theorem 5.12 and Theorem 4.3 (iii), we have that

$$\Pr [\mathcal{Z}^{r_0}] = \Pr \left[ \max_{i \in [n]} |y_i^{r_0}| \leq \min\{\kappa \cdot (g + \log n), c_3 g \log(ng)\} \right] \geq 1 - 3 \cdot (ng)^{-9}. \quad (7.17)$$

By Lemma 7.7 we have that,

$$\Pr \left[ \bigcap_{s \in [m - n \log^5 n, m]} \{\Psi_{Q,0}^s \leq Cn\} \mid \mathcal{Z}^{r_0}, \mathcal{F}^{r_0} \right] \geq 1 - n^{-8}.$$

By combining with Equation (7.17),

$$\Pr \left[ \bigcap_{s \in [m - n \log^5 n, m]} \{\Psi_{Q,0}^s \leq Cn\} \right] \geq (1 - n^{-8}) \cdot (1 - 3 \cdot (ng)^{-9}) \geq 1 - n^{-7}.$$

Taking the union bound with Equation (7.16), we get that

$$\Pr \left[ \bigcap_{s \in [r_0, m]} \{y_Q^s = y^s\} \cap \bigcap_{s \in [m - n \log^5 n, m]} \{\Psi_{Q,0}^s \leq Cn\} \right] \geq 1 - n^{-7} - 2n^{-7} \geq 1 - n^{-6}.$$

When this event holds we have that  $\Psi_{Q,0}^s = \Psi_0^s$  for every step  $s \in [r_0, m]$  and hence we can deduce for the original process  $\mathcal{P}$  that

$$\Pr \left[ \bigcap_{s \in [m - n \log^5 n, m]} \{\Phi_0^s \leq Cn\} \right] \geq 1 - n^{-6}.$$

Finally, since  $\alpha_2 \leq \alpha_1$ , we have that  $\Phi_0^s \leq \Psi_0^s$  for any step  $s \geq 0$  and hence, the conclusion follows.  $\square$

## 8 SUPER-EXPONENTIAL POTENTIAL FUNCTIONS

In this section, we will analyze super-exponential potential functions, i.e., exponential potentials with  $\Omega(1)$  smoothing parameters. In Section 8.1, we recall the definition of the super-exponential potentials and give an overview of the theorems we prove. In Section 8.2, we give a sufficient condition for a super-exponential potential to satisfy a drop inequality over one step, and in Section 8.3, we prove the concentration theorem for super-exponential potentials.



### 8.1 Outline

Recall from Equation (6.1) that a super-exponential potential function is defined as

$$\Phi^t := \Phi^t(\phi, z) := \sum_{i=1}^n \Phi_i^t := \sum_{i=1}^n e^{\phi \cdot (y_i^t - z)^+},$$

for smoothing parameter  $\phi > 0$  and integer offset  $z := z(n) > 0$ .

Note that if  $\Phi^t = O(\text{poly}(n))$  at some step  $t \geq 0$ , then

$$\text{Gap}(t) = O\left(z + \frac{\log n}{\phi}\right).$$

There are two differences in the *form* compared to the hyperbolic cosine potential  $\Gamma$  that we used in Section 4: (i) there is no underloaded component, as w.h.p., its contribution would be  $\omega(n)$  for any process making a constant number of samples in each step<sup>4</sup> and (ii) there is this  $(\dots)^+$  operation which is not essential, but simplifies some of the derivations.

However, the *main* difference is that there exist load vectors where the potential *could increase* in expectation, even when sufficiently large.<sup>5</sup> Here, we will prove a sufficient condition for the potential  $\Phi$  to drop in expectation over one step; requiring that the probability to allocate to a bin with normalized load at least  $z - 1$  is sufficiently small. More specifically, the event that we require is the following:

$$\mathcal{K}^s := \mathcal{K}_{\phi, z}^s(q^s) := \left\{ \forall i \in [n]: y_i^s \geq z - 1 \Rightarrow q_i^s \leq \frac{1}{n} \cdot e^{-\phi} \right\},$$

where  $q^s$  is the probability allocation vector used by the process at step  $s$ .

LEMMA 8.1 (RESTATED, PAGE 42). *Consider any allocation process and any super-exponential potential  $\Phi := \Phi(\phi, z)$  with  $\phi \in [4, n]$ . For any step  $s \geq 0$  where  $\mathcal{K}^s$  holds, we have that*

$$\mathbb{E} \left[ \Phi^{s+1} \mid \mathcal{F}^s, \mathcal{K}^s \right] \leq \Phi^s \cdot \left( 1 - \frac{1}{n} \right) + 2.$$

We will try to establish that this event  $\mathcal{K}^s$  holds for a sufficiently long interval and then show that in this interval the potential  $\Phi$  becomes small. In the analysis of  $g$ -ADV-COMP, this event will arise from the concentration of the hyperbolic cosine potential or from that of a super-exponential potential with smaller smoothing parameter. Now we are ready to state the main theorem.

THEOREM 8.7 (RESTATED, PAGE 47). *Consider any allocation process for which there exist super-exponential potential functions  $\Phi_1 := \Phi_1(\phi_1, z)$  and  $\Phi_2 := \Phi_2(\phi_2, z)$  with integer offset  $z := z(n) > 0$  and smoothing parameters  $\phi_1, \phi_2 \in (0, (\log n)/6]$  with  $\phi_2 \leq \frac{\phi_1}{84}$ , such that they satisfy for any step  $s \geq 0$ ,*

$$\mathbb{E} \left[ \Phi_1^{s+1} \mid \mathcal{F}^s, \mathcal{K}^s \right] \leq \Phi_1^s \cdot \left( 1 - \frac{1}{n} \right) + 2, \quad (8.1)$$

and

$$\mathbb{E} \left[ \Phi_2^{s+1} \mid \mathcal{F}^s, \mathcal{K}^s \right] \leq \Phi_2^s \cdot \left( 1 - \frac{1}{n} \right) + 2, \quad (8.2)$$

<sup>4</sup>It follows by a coupon collector's argument that the minimum load is w.h.p.  $\frac{t}{n} - \Omega(\log n)$ , for sufficiently large  $t$ .

<sup>5</sup>An example of such a configuration for the Two-CHOICE process without noise is one where  $n - 1$  bins have normalized loads  $z + 10/\phi$  and one bin is underloaded.

where  $\mathcal{K}^s := \mathcal{K}_{\phi_1, z}^s$ . Further, let  $P \in [n^{-4}, 1]$ . Then, for any steps  $t \geq 0$  and  $\tilde{t} \in [t, t + n \log^5 n]$ , which satisfy

$$\Pr \left[ \left\{ \text{Gap}(t - 2n \log^4 n) \leq \log^2 n \right\} \cap \bigcap_{s \in [t - 2n \log^4 n, \tilde{t}]} \mathcal{K}^s \right] \geq 1 - P, \quad (8.3)$$

they must also satisfy

$$\Pr \left[ \bigcap_{s \in [t, \tilde{t}]} \{\Phi_2^s \leq 8n\} \right] \geq 1 - (\log^8 n) \cdot P.$$

The statement of this theorem concerns steps in  $[t - 2n \log^4 n, \tilde{t}]$  with  $\tilde{t} \in [t, t + n \log^5 n]$ . The interval  $[t, \tilde{t}]$  is the *stabilization interval*, i.e., the interval where we want to show that  $\Phi_2^s \leq 8n$  for every  $s \in [t, \tilde{t}]$ . The interval  $[t - 2n \log^4 n, t]$  is the *recovery interval* where we will show that w.h.p.  $\Phi_2$  becomes  $O(n)$  at least *once*, provided we start with a “weak”  $\leq \log^2 n$  gap at step  $t - 2n \log^4 n$ . For both the recovery and stabilization intervals we will condition on the event  $\mathcal{K}$  holding at every step.

We will write  $\mathcal{H}^{t - 2n \log^4 n}$  as a shorthand for the event  $\{\text{Gap}(t - 2n \log^4 n) \leq \log^2 n\}$ .

*Dealing with negative steps.* In the analysis, we deal with negative (integer) steps, by extending the definition of the process to allocate zero weight balls in steps  $t < 0$ , i.e., we have that  $x^t = 0$ . Hence, for any  $t < 0$  it also follows deterministically that  $\Phi_1^t = \Phi_2^t = n$  and so the potentials trivially satisfy the drop inequalities Equations (8.1) and (8.2) at these steps.

## 8.2 General Drop Inequality

We now show that the drop inequality is satisfied in every step  $s$  where  $\mathcal{K}^s$  holds.

**LEMMA 8.1 (GENERAL DROP INEQUALITY).** *Consider any allocation process and any super-exponential potential  $\Phi := \Phi(\phi, z)$  with  $\phi \in [4, n]$ . For any step  $s \geq 0$  where  $\mathcal{K}^s$  holds, we have that*

$$\mathbb{E} \left[ \Phi^{s+1} \mid \mathcal{F}^s, \mathcal{K}^s \right] \leq \Phi^s \cdot \left( 1 - \frac{1}{n} \right) + 2.$$

**PROOF.** We consider the following three cases for the contribution of a bin  $i \in [n]$ :

**Case 1**  $[y_i^s < z - 1]$ : The contribution of  $i$  will remain  $\Phi_i^{s+1} = \Phi_i^s = 1$ , even if a ball is allocated to bin  $i$ . Hence,

$$\mathbb{E} \left[ \Phi_i^{s+1} \mid \mathcal{F}^s, \mathcal{K}^s \right] = \Phi_i^s = \Phi_i^s \cdot \left( 1 - \frac{1}{n} \right) + \frac{1}{n}.$$

**Case 2**  $[y_i^s \in [z - 1, z]]$ : By the condition  $\mathcal{K}^s$ , the probability of allocating a ball to bin  $i$  with  $y_i^s \geq z - 1$  is  $q_i^s \leq \frac{1}{n} \cdot e^{-\phi}$ . Hence, the expected contribution of this bin is at most

$$\begin{aligned} \mathbb{E} \left[ \Phi_i^{s+1} \mid \mathcal{F}^s, \mathcal{K}^s \right] &\leq e^\phi \cdot q_i^s + \Phi_i^s \cdot (1 - q_i^s) \\ &\leq e^\phi \cdot \frac{1}{n} \cdot e^{-\phi} + \Phi_i^s \\ &= \frac{1}{n} + \Phi_i^s \\ &= \Phi_i^s \cdot \left( 1 - \frac{1}{n} \right) + \frac{2}{n}, \end{aligned}$$

using in the last equation that  $\Phi_i^s = 1$ .

**Case 3**  $[y_i^s > z]$ : Again, by the condition  $\mathcal{K}^s$ , the probability of allocating a ball to bin  $i$  with  $y_i^s > z$  is  $q_i^s \leq \frac{1}{n} \cdot e^{-\phi}$ . Hence,

$$\begin{aligned}
 \mathbb{E} \left[ \Phi_i^{s+1} \mid \mathcal{F}^s, \mathcal{K}^s \right] &\stackrel{(a)}{=} \Phi_i^s \cdot e^{\phi \cdot (1-1/n)} \cdot q_i^s + \Phi_i^s \cdot e^{-\phi/n} \cdot (1 - q_i^s) \\
 &\leq \Phi_i^s \cdot e^{\phi \cdot (1-1/n)} \cdot q_i^s + \Phi_i^s \cdot e^{-\phi/n} \\
 &= \Phi_i^s \cdot e^{-\phi/n} \cdot \left( 1 + e^{\phi} \cdot q_i^s \right) \\
 &\stackrel{(b)}{\leq} \Phi_i^s \cdot \left( 1 - \frac{\phi}{2n} \right) \cdot \left( 1 + e^{\phi} \cdot \frac{1}{n} \cdot e^{-\phi} \right) \\
 &\stackrel{(c)}{\leq} \Phi_i^s \cdot \left( 1 - \frac{2}{n} \right) \cdot \left( 1 + \frac{1}{n} \right) \\
 &\leq \Phi_i^s \cdot \left( 1 - \frac{1}{n} \right),
 \end{aligned}$$

using in (a) that  $y_i^s > z$  implies that  $y_i^s \geq z + \frac{1}{n}$  since  $z$  is an integer, in (b) that  $e^u \leq 1 + \frac{1}{2}u$  (for any  $-1.5 \leq u < 0$ ) and that  $\phi \leq n$  and in (c) that  $\phi \geq 4$ .

Aggregating over the three cases, we get the claim:

$$\mathbb{E} \left[ \Phi^{s+1} \mid \mathcal{F}^s, \mathcal{K}^s \right] = \sum_{i=1}^n \mathbb{E} \left[ \Phi_i^{s+1} \mid \mathcal{F}^s, \mathcal{K}^s \right] \leq \sum_{i=1}^n \left( \Phi_i^s \cdot \left( 1 - \frac{1}{n} \right) + \frac{2}{n} \right) = \Phi^s \cdot \left( 1 - \frac{1}{n} \right) + 2. \quad \square$$

### 8.3 Concentration

**8.3.1 Proof Outline of Theorem 8.7.** We will now give a summary of the main technical steps in the proof of Theorem 8.7 (an illustration of the key steps is shown in Figure 7). The proof uses an interplay between two instances  $\Phi_1 := \Phi_1(\phi_1, z)$  and  $\Phi_2 := \Phi_2(\phi_2, z)$  of the super-exponential potential function with  $\phi_2 \leq \phi_1/84$ , such that in steps  $s \geq 0$  when  $\Phi_1^s$  is small, then the change of  $\Phi_2^s$  is very small.

**Recovery.** By the third precondition Equation (8.3) of Theorem 8.7, we start with  $\text{Gap}(t - 2n \log^4 n) \leq \log^2 n$ , which implies that  $\Phi_1^{t-2n \log^4 n} \leq e^{\frac{1}{2} \cdot \log^4 n}$  (Claim 8.2). Using the drop inequality for the potential  $\Phi_1$  (first precondition Equation (8.1)), it follows that  $\mathbb{E}[\Phi_1^s] \leq 6n$ , for any step  $s \in [t - n \log^4 n, \tilde{t}]$  (Lemma 8.5). By using Markov's inequality and a union bound, we can deduce that w.h.p.  $\Phi_1^s \leq 6n^{12}$  for all steps  $s \in [t - n \log^4 n, \tilde{t}]$ . By a simple interplay between two potentials, this implies  $\Phi_2^{t-n \log^4 n} \leq n^{7/6}$  (Lemma 8.4 (i)). Now using a drop inequality for the potential  $\Phi_2$  (second precondition Equation (8.2)), guarantees that w.h.p.  $\Phi_2^{r_0} \leq 6n$  for *some single* step  $r_0 \in [t - n \log^4 n, \tilde{t}]$  (Lemma 8.6).

**Stabilization.** To obtain the stronger statement which holds for *all* steps  $s \in [t, \tilde{t}]$ , we will use a concentration inequality. The key point is that for any step  $r$  with  $\Phi_1^r \leq 6n^{12}$  the absolute difference  $|\Phi_2^{r+1} - \Phi_2^r|$  is at most  $n^{1/3}$ , because  $\phi_2 \leq \frac{\phi_1}{84}$  (by preconditions Equations (8.1) and (8.2)). This is crucial for applying the Azuma's inequality for super-martingales (Lemma A.1) to  $\Phi_2$  which yields that  $\Phi_2^s \leq 8n$  for all steps  $s \in [t, \tilde{t}]$  using a smoothing argument (Claim 8.3).

**8.3.2 Deterministic Relations between the Potential Functions.** We collect several basic facts about the super-exponential potential functions  $\Phi_1 := \Phi_1(\phi_1, z)$  and  $\Phi_2 := \Phi_2(\phi_2, z)$  satisfying the preconditions of Theorem 8.7.

We start with a simple upper bound on  $\Phi_1^s$  using a weak upper bound on the gap at step  $s$ .

**CLAIM 8.2.** *For any step  $s \geq 0$  where  $\text{Gap}(s) \leq \log^2 n$ , we have that  $\Phi_1^s \leq e^{\frac{1}{2} \cdot \log^4 n}$ .*

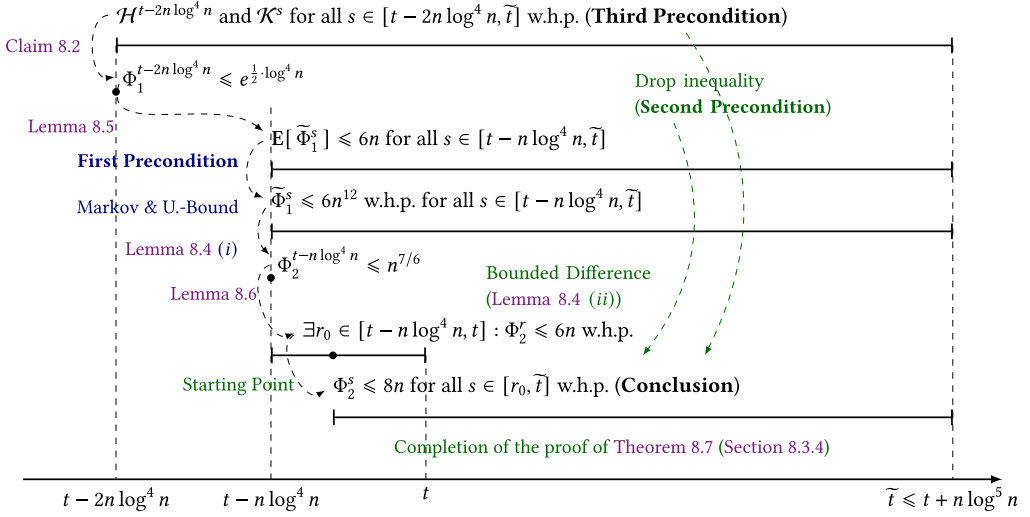


Fig. 7. Outline for the proof of Theorem 8.7. Results in blue are given in Section 8.3.2 and 8.3.3, while results in green are used in the completion of the proof in Section 8.3.4.

PROOF. Since  $\phi_1 \leq (\log n)/6$ , we have that  $\Phi_1^s = \sum_{i=1}^n e^{\phi_1 \cdot (y_i^s - z)^+} \leq n \cdot e^{\phi_1 \cdot \log^2 n} \leq e^{\frac{1}{2} \cdot \log^4 n}$ .  $\square$

The next claim is a simple “smoothness” argument showing that the potential cannot decrease quickly within  $\lceil n/\log^2 n \rceil$  steps. The derivation is elementary and relies on the fact that the average load changes by at most  $2/\log^2 n$ .

CLAIM 8.3. *For any step  $s \geq 0$  and any step  $r \in [s, s + \lceil n/\log^2 n \rceil]$ , we have that  $\Phi_2^r \geq 0.99 \cdot \Phi_2^s$ .*

PROOF. The normalized load after  $r - s$  steps can decrease by at most  $\frac{r-s}{n} \leq \frac{2}{\log^2 n}$ . Hence, for any bin  $i \in [n]$ ,

$$\begin{aligned} \Phi_{2,i}^r &= e^{\phi_2 \cdot (y_i^r - z)^+} \geq e^{\phi_2 \cdot (y_i^s - \frac{r-s}{n} - z)^+} \geq e^{\phi_2 \cdot (y_i^s - z)^+ - \phi_2 \cdot \frac{2}{\log^2 n}} = \Phi_{2,i}^s \cdot e^{-\frac{2\phi_2}{\log^2 n}} \geq \Phi_{2,i}^s \cdot e^{-o(1)} \\ &\geq 0.99 \cdot \Phi_{2,i}^s, \end{aligned}$$

for sufficiently large  $n$ , using that  $\phi_2 \leq (\log n)/6$ . By aggregating over all bins, we get the claim.  $\square$

The next claim is crucial for applying the concentration inequality, since the second statement bounds the maximum additive change of  $\Phi_2^s$  in any step  $s$  where  $\Phi_1^s$  is poly( $n$ ):

LEMMA 8.4. *For any step  $s \geq 0$  where  $\Phi_1^s \leq 6n^{12}$ , we have that*

- (i)  $\Phi_2^s \leq n^{7/6}$ ,
- (ii)  $|\Phi_2^{s+1} - \Phi_2^s| \leq n^{1/3}$ .

PROOF. Consider an arbitrary step  $s \geq 0$  with  $\Phi_1^s \leq 6n^{12}$ . We start by upper bounding the normalized load  $y_i^s$  of any bin  $i \in [n]$ ,

$$y_i^s \leq z + \frac{\log \Phi_{1,i}^s}{\phi_1} \leq z + \frac{\log \Phi_1^s}{\phi_1} \leq z + \frac{\log(6n^{12})}{\phi_1} \leq z + \frac{14 \log n}{\phi_1}.$$

*First statement.* Now, we upper bound the contribution of any bin  $i \in [n]$  to  $\Phi_2^s$ ,

$$\Phi_{2,i}^s = \exp\left(\phi_2 \cdot (y_i^s - z)^+\right) \leq \exp\left(\frac{14 \cdot \phi_2}{\phi_1} \cdot \log n\right) \leq n^{1/6}, \quad (8.4)$$

using that  $\phi_2 \leq \frac{\phi_1}{84}$ . Hence, by aggregating over all bins,

$$\Phi_2^s \leq n \cdot n^{1/6} = n^{7/6}. \quad (8.5)$$

*Second statement.* We will derive lower and upper bounds for  $\Phi_2^{s+1}$ . For the upper bound, let  $i = i^{s+1} \in [n]$  be the bin where the  $(s+1)$ -th ball is allocated, then

$$\Phi_2^{s+1} \leq \Phi_2^s + \Phi_{2,i}^s \cdot e^{\phi_2} \leq \Phi_2^s + n^{1/6} \cdot n^{1/6} = \Phi_2^s + n^{1/3},$$

using that  $\phi_2 \leq (\log n)/6$  and Equation (8.4). For the lower bound, we pessimistically assume that all bin loads decrease by  $1/n$  in step  $s+1$ , so

$$\Phi_2^{s+1} \geq \Phi_2^s \cdot e^{-\phi_2/n} \stackrel{(a)}{\geq} \Phi_2^s \cdot \left(1 - \frac{\phi_2}{n}\right) \stackrel{(b)}{\geq} \Phi_2^s - \frac{n \cdot n^{1/6}}{n} \cdot \log n \geq \Phi_2^s - n^{1/3},$$

using in (a) that  $e^u \geq 1+u$  (for any  $u$ ) and in (b) that  $\phi_2 \leq \log n$  and  $\Phi_2^s \leq n \cdot n^{1/6}$  by Equation (8.5). Combining the two bounds we get the second statement.  $\square$

**8.3.3 Recovery Phase.** In this section, we will show for an auxiliary process  $\tilde{\mathcal{P}}$  (to be defined below) that the potential function  $\Phi_2$  satisfies  $\Phi_2^s \leq 6n$  in *at least one step*  $s \in [t - n \log^4 n, t]$  w.h.p.

First, we show that for the original process  $\mathcal{P}$  in the statement of Theorem 8.7, the potential

$$\tilde{\Phi}_1^s := \tilde{\Phi}_1^s(t) := \Phi_1^s \cdot \mathbf{1}_{\cap_{r \in [t-2n \log^4 n, s]} \mathcal{K}^r \cap \mathcal{H}^{t-2n \log^4 n}}, \quad (8.6)$$

is small in expectation for *all* steps  $s \geq t - n \log^4 n$ . Note that there is a “recovery time” until the expectation becomes small, of at most  $n \log^4 n$  steps after the “weak” bound  $\text{Gap}(t - 2n \log^4 n) \leq \log^2 n$  which follows from the third precondition Equation (8.3) in Theorem 8.7.

LEMMA 8.5. *Consider the potential  $\tilde{\Phi}_1 := \tilde{\Phi}_1(t)$  for any step  $t \geq 0$ . Then, for any step  $s \geq t - n \log^4 n$ ,*

$$\mathbb{E} \left[ \tilde{\Phi}_1^s \right] \leq 6n.$$

PROOF. By the precondition Equation (8.1) of Theorem 8.7, for any step  $s \geq t - n \log^4 n$ ,

$$\mathbb{E} \left[ \Phi_1^{s+1} \mid \mathcal{F}^s, \mathcal{K}^s \right] \leq \Phi_1^s \cdot \left(1 - \frac{1}{n}\right) + 2.$$

Next note that whenever  $\neg \mathcal{K}^s$  holds, it follows deterministically that  $\{\tilde{\Phi}_1^s = \tilde{\Phi}_1^{s+1} = 0\}$ , and hence

$$\mathbb{E} \left[ \tilde{\Phi}_1^{s+1} \mid \mathcal{F}^s \right] \leq \tilde{\Phi}_1^s \cdot \left(1 - \frac{1}{n}\right) + 2. \quad (8.7)$$

We will now upper bound  $\mathbb{E}[\tilde{\Phi}_1^s \mid \mathcal{F}^{t-2n \log^4 n}, \mathcal{H}^{t-2n \log^4 n}]$  for any step  $s \geq t - n \log^4 n$ . When  $\mathcal{H}^{t-2n \log^4 n}$  holds, by Claim 8.2, it also follows that  $\tilde{\Phi}_1^{t-2n \log^4 n} \leq \Phi_1^{t-2n \log^4 n} \leq e^{\frac{1}{2} \cdot \log^4 n}$ . Hence applying Lemma A.11 (i) (for  $a = 1 - \frac{1}{n}$  and  $b = 2$ ) using Equation (8.7),

$$\begin{aligned} \mathbb{E} \left[ \tilde{\Phi}_1^s \mid \mathcal{F}^{t-2n \log^4 n}, \mathcal{H}^{t-2n \log^4 n} \right] &\leq \mathbb{E} \left[ \tilde{\Phi}_1^s \mid \mathcal{F}^{t-2n \log^4 n}, \tilde{\Phi}_1^{t-2n \log^4 n} \leq e^{\frac{1}{2} \cdot \log^4 n} \right] \\ &\leq e^{\frac{1}{2} \cdot \log^4 n} \cdot \left(1 - \frac{1}{n}\right)^{s-(t-2n \log^4 n)} + 2n \\ &\stackrel{(a)}{\leq} e^{\frac{1}{2} \cdot \log^4 n} \cdot \left(1 - \frac{1}{n}\right)^{n \log^4 n} + 2n \end{aligned}$$

$$\stackrel{(b)}{\leq} e^{\frac{1}{2} \cdot \log^4 n} \cdot e^{-\log^4 n} + 2n \leq 1 + 2n \leq 6n,$$

using in (a) that  $s \geq t - n \log^4 n$  and in (b) that  $e^u \geq 1 + u$  (for any  $u$ ). Hence, the claim follows,

$$\mathbb{E} [\tilde{\Phi}_1^s] = \mathbb{E} [\tilde{\Phi}_1^s \mid \mathcal{H}^{t-2n \log^4 n}] \cdot \Pr [\mathcal{H}^{t-2n \log^4 n}] + 0 \cdot \Pr [-\mathcal{H}^{t-2n \log^4 n}] \leq 6n. \quad \square$$

*The auxiliary process  $\tilde{\mathcal{P}}$ .* Let  $\mathcal{P}$  be the process in the statement of Theorem 8.7, define  $t_1 := t - n \log^4 n$  and the stopping time  $\sigma := \inf\{s \geq t_1 : \{\Phi_1^s > 6n^{11}\} \cup \neg \mathcal{K}^s\}$ . Now we define the auxiliary process  $\tilde{\mathcal{P}}_{t_1}$  such that

- in steps  $s \in [0, \sigma)$  it follows the allocations of  $\mathcal{P}$ , and
- in steps  $s \in [\sigma, \infty)$  it allocates to the (currently) least loaded bin, which is a bin with normalized load  $\leq 1 \leq z$ .

This way the process trivially satisfies  $\mathcal{K}^s$  and therefore also the drop inequalities (Equations (8.1) and (8.2)) for any step  $s \geq t_1$ . Furthermore, starting with  $\Phi_1^{t_1} \leq 6n^{11}$ , for any  $s \geq t_1$  it deterministically satisfies  $\Phi_1^s \leq 6n^{12}$ , since  $\Phi^\sigma \leq 6n^{11} \cdot e^{\phi_1} \leq 6n^{12}$  (as  $\phi_1 \leq (\log n)/6$ ) and for  $s \geq \sigma$  the potential does not increase.

For this starting condition, we will also show that the other potential function  $\Phi_2$  w.h.p. becomes linear in *at least one* step in  $[t - n \log^4 n, t]$ , by using Lemma 8.5.

LEMMA 8.6 (RECOVERY). *For any step  $t \geq 0$  and the auxiliary process  $\tilde{\mathcal{P}}_{t-n \log^4 n}$ , it holds that*

$$\Pr \left[ \bigcup_{s \in [t-n \log^4 n, t]} \{\Phi_2^s \leq 6n\} \mid \mathcal{F}^{t-n \log^4 n}, \Phi_1^{t-n \log^4 n} \leq 6n^{11} \right] \geq 1 - n^{-6}.$$

PROOF. Fix any step  $t \geq 0$  and let  $t_1 := t - n \log^4 n$  be the starting point of the analysis. Recall that the auxiliary process  $\tilde{\mathcal{P}}_{t_1}$  satisfies  $\mathcal{K}^s$  for any step  $s \geq t_1$ . Further, for any  $s \geq t_1$ , we define the “killed” potential function

$$\hat{\Phi}_2^s := \Phi_2^s \cdot \mathbf{1}_{\cap_{r \in [t_1, s]} \{\Phi_2^r > 6n\}}.$$

We will show that  $\hat{\Phi}_2$  drops in expectation by a multiplicative factor in every step. We start by showing for  $\Phi_2$  that when  $\Phi_2^s > 6n$ , then it drops in expectation by a multiplicative factor. By the second precondition Equation (8.2) of Theorem 8.7,

$$\begin{aligned} \mathbb{E} [\Phi_2^{s+1} \mid \Phi_1^{t_1} \leq 6n^{11}, \mathcal{F}^s, \Phi_2^s > 6n] &= \mathbb{E} [\Phi_2^{s+1} \mid \Phi_1^{t_1} \leq 6n^{11}, \mathcal{F}^s, \mathcal{K}^s, \Phi_2^s > 6n] \\ &\leq \Phi_2^s \cdot \left(1 - \frac{1}{n}\right) + 2 \\ &\leq \Phi_2^s \cdot \left(1 - \frac{1}{2n}\right) - 6n \cdot \frac{1}{2n} + 2 \leq \Phi_2^s \cdot \left(1 - \frac{1}{2n}\right). \end{aligned}$$

Whenever the event  $\{\Phi_2^s \leq 6n\}$  holds, it follows deterministically that  $\hat{\Phi}_2^s = \hat{\Phi}_2^{s+1} = 0$ . So, for the potential  $\hat{\Phi}_2$  we obtain the drop inequality (with one fewer condition),

$$\mathbb{E} [\hat{\Phi}_2^{s+1} \mid \Phi_1^{t_1} \leq 6n^{11}, \mathcal{F}^s] \leq \hat{\Phi}_2^s \cdot \left(1 - \frac{1}{2n}\right). \quad (8.8)$$

Using Lemma 8.4 (i),  $\Phi_1^{t_1} \leq 6n^{11}$  implies that  $\hat{\Phi}_2^{t_1} \leq \Phi_2^{t_1} \leq n^{7/6}$  and so inductively applying Equation (8.8) for  $n \log^4 n$  steps starting at  $t_1$ , we have that

$$\mathbb{E} [\hat{\Phi}_2^t \mid \mathcal{F}^{t_1}, \Phi_1^{t_1} \leq 6n^{11}] \leq \mathbb{E} [\hat{\Phi}_2^t \mid \mathcal{F}^{t_1}, \hat{\Phi}_2^{t_1} \leq n^{7/6}] \leq \hat{\Phi}_2^{t_1} \cdot \left(1 - \frac{1}{2n}\right)^{n \log^4 n} \leq n^{7/6} \cdot e^{-\frac{1}{2} \log^4 n} \leq n^{-6},$$

for sufficiently large  $n$ , using that  $e^u \geq 1 + u$  (for any  $u$ ). Hence, by Markov's inequality,

$$\Pr \left[ \widehat{\Phi}_2^t \leq 1 \mid \mathcal{F}^{t_1}, \Phi_1^{t_1} \leq 6n^{11} \right] \geq 1 - n^{-6}.$$

Since it deterministically holds that  $\{\Phi_2^t \geq n\}$  for any step  $t \geq 0$ , it follows that if  $\{\widehat{\Phi}_2^t \leq 1\}$  holds, then also  $\{\widehat{\Phi}_2^t = 0\}$ . So, we conclude that  $\mathbf{1}_{\cap_{r \in [t_1, t]} \{\Phi_2^r > 6n\}} = 0$  holds, i.e.,

$$\Pr \left[ \bigcup_{r \in [t_1, t]} \{\Phi_2^r \leq 6n\} \mid \mathcal{F}^{t_1}, \Phi_1^{t_1} \leq 6n^{11} \right] \geq 1 - n^{-6}. \quad \square$$

**8.3.4 Completing the Proof of Theorem 8.7.** For some values of  $g$ , the number of steps  $k$  in the layered induction may be  $\omega(1)$  and so we may need to apply Theorem 8.7 up to  $\omega(1)$  times, which means that we cannot afford to lose a polynomial factor in the error probability. To overcome this, we will partition the time-interval into consecutive intervals of length  $\lceil n/\log^2 n \rceil$ . Then, we will prove that at the end of each such interval the potential is small w.h.p., and finally use a simple smoothness argument (Claim 8.3) to show that the potential is small w.h.p. in *all* steps.

**THEOREM 8.7.** *Consider any allocation process for which there exist super-exponential potential functions  $\Phi_1 := \Phi_1(\phi_1, z)$  and  $\Phi_2 := \Phi_2(\phi_2, z)$  with integer offset  $z := z(n) > 0$  and smoothing parameters  $\phi_1, \phi_2 \in (0, (\log n)/6]$  with  $\phi_2 \leq \frac{\phi_1}{84}$ , such that they satisfy for any steps  $s \geq 0$ ,*

$$\mathbb{E} \left[ \Phi_1^{s+1} \mid \mathcal{F}^s, \mathcal{K}^s \right] \leq \Phi_1^s \cdot \left( 1 - \frac{1}{n} \right) + 2,$$

and

$$\mathbb{E} \left[ \Phi_2^{s+1} \mid \mathcal{F}^s, \mathcal{K}^s \right] \leq \Phi_2^s \cdot \left( 1 - \frac{1}{n} \right) + 2,$$

where  $\mathcal{K}^s := \mathcal{K}_{\phi_1, z}^s$ . Further, let  $P \in [n^{-4}, 1]$ . Then, for any steps  $t \geq 0$  and  $\tilde{t} \in [t, t + n \log^5 n]$ , which satisfy

$$\Pr \left[ \left\{ \text{Gap}(t - 2n \log^4 n) \leq \log^2 n \right\} \cap \bigcap_{s \in [t - 2n \log^4 n, \tilde{t}]} \mathcal{K}^s \right] \geq 1 - P,$$

they must also satisfy

$$\Pr \left[ \bigcap_{s \in [t, \tilde{t}]} \{\Phi_2^s \leq 8n\} \right] \geq 1 - (\log^8 n) \cdot P.$$

We will start by proving the following lemma for the auxiliary process  $\widetilde{\mathcal{P}}_{t-n \log^4 n}$ .

**LEMMA 8.8.** *For any step  $t \geq 0$  and the auxiliary process  $\widetilde{\mathcal{P}}_{t-n \log^4 n}$ , it holds that*

$$\Pr \left[ \bigcap_{s \in [t, \tilde{t}]} \{\Phi_2^s \leq 8n\} \mid \Phi_1^{t-n \log^4 n} \leq 6n^{11} \right] \geq 1 - \frac{1}{4} \cdot (\log^8 n) \cdot P.$$

**PROOF.** Our goal is to apply Azuma's inequality to  $\Phi_2$ . However, there are two challenges: (i) we cannot afford to take the union bound over all  $\text{poly}(n)$  steps and (ii)  $\Phi_2$  is a super-martingale only when it is sufficiently large. To deal with (i) we will apply Azuma's inequality to sub-intervals of length at most  $\lceil n/\log^2 n \rceil$  and then use a smoothness argument (Claim 8.3) to deduce that  $\Phi_2$  is small in the steps in between. For (ii) we will define  $X^s := X^s(\Phi_2^s)$  in a way to ensure that it is super-martingale at every step and also satisfies the bounded difference inequality.

More specifically, consider an arbitrary step  $r \in [t - n \log^4 n, t]$  and partition the interval  $(r, \tilde{t}]$  into

$$\mathcal{I}_1 := (r, r + \Delta], \mathcal{I}_2 = (r + \Delta, r + 2\Delta], \dots, \mathcal{I}_q := (r + (q-1)\Delta, \tilde{t}],$$

where  $\Delta := \lceil n / \log^2 n \rceil$  and  $q := \lceil \frac{\tilde{t}-r}{\Delta} \rceil \leq \lceil \frac{t+n \log^5 n - r}{\Delta} \rceil \leq 2 \log^7 n$ . In order to prove that  $\Phi_2$  is at most  $8n$  in every step in  $(r, \tilde{t}]$ , we will use our auxiliary lemmas (Section 8.3.2 and 8.3.3) and Azuma's super-martingale concentration inequality (Lemma A.1) to establish that  $\Phi_2$  is at most  $7n$  at each of the steps  $r + \Delta, r + 2\Delta, \dots, r + (q-1)\Delta, \tilde{t}$ . Finally, by using a smoothness argument (Claim 8.3), we will establish that  $\Phi_2$  is at most  $8n$  at all steps in  $(r, \tilde{t}]$ , which is the conclusion of the theorem.

For each interval  $i \in [q]$ , we define  $X_i^{r+(i-1)\Delta} := \max\{\Phi_2^{r+(i-1)\Delta}, 5n + n^{1/3}\}$  and for any  $s \in (r + (i-1)\Delta, r + i\Delta]$ ,

$$X_i^s := \begin{cases} \Phi_2^s & \text{if there exists } u \in [r + (i-1)\Delta, s] \text{ such that } \Phi_2^u \geq 5n, \\ 5n + n^{1/3} & \text{otherwise.} \end{cases}$$

Note that whenever the first condition in the definition of  $X_i^s$  is satisfied, it remains satisfied until the end of the interval, i.e., until step  $r + i \cdot \Delta$ . Our next aim is to establish the preconditions of Azuma's inequality (Lemma A.1) for  $X_i^s$ .

For convenience, we define  $t_1 := t - n \log^4 n$  and the event

$$\mathcal{Z}^{t_1} := \{\Phi_1^{t_1} \leq 6n^{11}\}.$$

LEMMA 8.9. Fix any interval  $\mathcal{I}_i$  for  $i \in [q]$ . Then for the random variables  $X_i^s$ , for any step  $s \in (r + (i-1)\Delta, r + i\Delta]$ , it holds that

$$\mathbb{E} [X_i^s \mid \mathcal{Z}^{t_1}, \mathcal{F}^{s-1}] \leq X_i^{s-1},$$

and

$$(|X_i^s - X_i^{s-1}| \mid \mathcal{Z}^{t_1}, \mathcal{F}^{s-1}) \leq 2n^{1/3}.$$

PROOF OF LEMMA 8.9. Recall that by conditioning on  $\mathcal{Z}^{t_1}$ , the auxiliary process  $\tilde{\mathcal{P}}_{t_1}$  satisfies  $\{\Phi_1^{s-1} \leq 6n^{12}\}$  and  $\mathcal{K}^{s-1}$ , for any  $s \in (r + (i-1)\Delta, r + i\Delta]$ , since  $r + (i-1)\Delta \geq t_1$ .

By precondition Equation (8.2), when  $\{\Phi_2^{s-1} \geq 4n\}$  also holds, we have that,

$$\begin{aligned} \mathbb{E} [\Phi_2^s \mid \mathcal{Z}^{t_1}, \mathcal{F}^{s-1}, \Phi_2^{s-1} \geq 4n] &= \mathbb{E} [\Phi_2^s \mid \mathcal{Z}^{t_1}, \mathcal{F}^{s-1}, \mathcal{K}^{s-1}, \Phi_2^{s-1} \geq 4n] \\ &\leq \Phi_2^{s-1} \cdot \left(1 - \frac{1}{n}\right) + 2 \leq \Phi_2^{s-1} - 4n \cdot \frac{1}{n} + 2 \leq \Phi_2^{s-1}. \end{aligned} \quad (8.9)$$

Further, by Lemma 8.4 (ii), we also have that

$$(|\Phi_2^s - \Phi_2^{s-1}| \mid \mathcal{Z}^{t_1}, \mathcal{F}^{s-1}) \leq n^{1/3}. \quad (8.10)$$

**Case 1**  $[\Phi_2^{r+(i-1)\Delta} \geq 5n + n^{1/3}]$ : In this case  $X_i^{s-1} = \Phi_2^{s-1}$  for all  $s \in (r + (i-1)\Delta, r + i\Delta]$ . By Claim 8.3, we also have  $\Phi_2^{s-1} \geq 0.99 \cdot (5n + n^{1/3}) \geq 4n$  (as  $\Delta \leq \lceil n / \log^2 n \rceil$ ) and the two statements follow by Equations (8.9) and (8.10).

**Case 2**  $[\Phi_2^{r+(i-1)\Delta} < 5n + n^{1/3}]$ : Let  $\sigma := \inf\{u \geq r + (i-1)\Delta : \Phi_2^u \geq 5n\}$ . We consider the following three cases (see Figure 8):

- **Case 2(a)**  $[s - 1 < \sigma]$ : Here  $X_i^s = X_i^{s-1} = 5n + n^{1/3}$ , so the two statements hold trivially.
- **Case 2(b)**  $[s - 1 = \sigma]$ : We will first establish that

$$5n \leq \Phi_2^{s-1} \leq 5n + n^{1/3}. \quad (8.11)$$



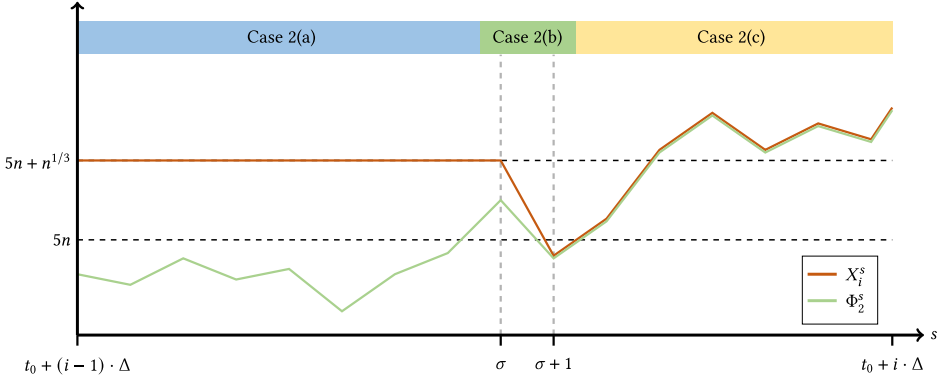


Fig. 8. Visualization of three cases in the proof of Lemma 8.9 for interval  $i \in [q]$ .

The lower bound  $\Phi_2^{s-1} \geq 5n$  follows by definition of  $\sigma$ . For the upper bound, we consider the following two cases. If  $s - 1 = r + (i - 1)\Delta$ , then this follows by the assumption for Case 2. Otherwise, we have that  $\Phi_2^{s-1} \geq 5n$  and  $\Phi_2^{s-2} < 5n$ . By Equation (8.10), we obtain that  $\Phi_2^{s-1} \leq \Phi_2^{s-2} + n^{1/3} < 5n + n^{1/3}$ .

Next, by definition,  $X_i^{s-1} = 5n + n^{1/3}$  and  $X_i^s = \Phi_2^s$ , so by Equation (8.9),

$$\mathbb{E} [X_i^s \mid \mathcal{Z}^{t_1}, \mathcal{F}^{s-1}] = \mathbb{E} [\Phi_2^s \mid \mathcal{Z}^{t_1}, \mathcal{F}^{s-1}] \leq \Phi_2^{s-1} < X_i^{s-1},$$

which establishes the first statement. For the second statement, by Equations (8.10) and (8.11) we have

$$|X_i^s - X_i^{s-1}| = |\Phi_2^s - 5n - n^{1/3}| \leq |\Phi_2^{s-1} - 5n - n^{1/3}| + |\Phi_2^s - \Phi_2^{s-1}| \leq 2n^{1/3},$$

using in the last inequality that Equations (8.10) and (8.11).

– **Case 2(c)** [ $s - 1 > \sigma$ ]: Here,  $X_i^{s-1} = \Phi_2^{s-1}$  and  $X_i^s = \Phi_2^s$ . Since  $\Phi_2^\sigma \geq 5n$ , by Claim 8.3 (as  $s - \sigma \leq \lceil n/\log^2 n \rceil$ ), we also have that

$$\Phi_2^{s-1} \geq 0.99 \cdot \Phi_2^\sigma \geq 0.99 \cdot 5n \geq 4n,$$

and thus by Equation (8.9), the first statement follows. The second statement follows by Equation (8.10).  $\square$

Now we return to the proof of Lemma 8.8. By Lemma 8.9, we have verified that  $X_i^s$  satisfies the preconditions of Azuma's inequality for any filtration  $\mathcal{F}^{s-1}$  where  $\mathcal{Z}^{t_1}$  holds. So, applying Lemma A.1 for  $\lambda = \frac{n}{2 \log^7 n}$ ,  $N \leq \Delta$  and  $D = 2n^{1/3}$ , we get for any  $i \in [q]$ ,

$$\Pr [X_i^{r+i\Delta} \geq X_i^{r+(i-1)\Delta} + \lambda \mid \mathcal{Z}^{t_1}, \mathcal{F}^r] \leq \exp \left( -\frac{n^2/(4 \log^{14} n)}{2 \cdot \Delta \cdot (4n^{2/3})} \right) + 2P \leq 3P,$$

since  $\Delta \leq \lceil n/\log^2 n \rceil$  and  $P \geq n^{-4}$ . Taking the union bound over the at most  $2 \log^7 n$  intervals  $i \in [q]$ , it follows that

$$\Pr \left[ \bigcup_{i \in [q]} \left\{ X_i^{r+i\Delta} \geq X_i^r + i \cdot \frac{n}{2 \log^7 n} \right\} \mid \mathcal{Z}^{t_1}, \mathcal{F}^r \right] \leq (2 \log^7 n) \cdot 3P \leq \frac{1}{8} \cdot (\log^8 n) \cdot P.$$

Next, conditional on  $(\mathcal{Z}^{t_1}, \mathfrak{F}^r, \Phi_2^r \leq 6n)$ , we have the following chain of inclusions:

$$\begin{aligned} \bigcap_{i \in [q]} \left\{ X_i^{r+i\Delta} \leq X_1^r + i \cdot \frac{n}{2 \log^7 n} \right\} &\stackrel{(a)}{\leq} \bigcap_{i \in [q]} \{X_i^{r+i\Delta} \leq 6n + n\} \\ &\stackrel{(b)}{=} \bigcap_{i \in [q]} \{\Phi_2^{r+i\Delta} \leq 7n\} \\ &\stackrel{(c)}{\subseteq} \bigcap_{s \in [r, \tilde{t}]} \left\{ \Phi_2^s \leq \frac{7}{0.99} \cdot n \right\} \\ &\stackrel{(d)}{\subseteq} \bigcap_{s \in [t, \tilde{t}]} \{\Phi_2^s \leq 8n\}, \end{aligned}$$

where (a) holds since  $i \leq q \leq 2 \log^7 n$  and  $X_1^r \leq \max\{\Phi_2^r, 5n + n^{1/3}\}$ , (b) holds since  $\Phi_2^{r+i\Delta} \leq X_i^{r+i\Delta}$ , (c) holds by applying the smoothness argument of Claim 8.3 to each interval  $i \in [q]$ , as the length of each interval is at most  $\lceil n/\log^2 n \rceil$  and (d) holds since  $8 \geq \frac{7}{0.99}$  and  $r \leq t$ . This implies that

$$\Pr \left[ \bigcap_{s \in [t, \tilde{t}]} \{\Phi_2^s \leq 8n\} \mid \mathcal{Z}^{t_1}, \mathfrak{F}^r, \Phi_2^r \leq 6n \right] \geq 1 - \frac{1}{8} \cdot (\log^8 n) \cdot P. \quad (8.12)$$

Next define  $\tau := \inf\{r \geq t - n \log^4 n : \Phi_2^r \leq 6n\}$ . By Lemma 8.6,

$$\Pr [\tau \leq t \mid \mathcal{Z}^{t_1}] \geq \Pr [\tau \leq t \mid \mathfrak{F}^{t_1}, \mathcal{Z}^{t_1}] \geq 1 - n^{-6}. \quad (8.13)$$

We get the conclusion for the  $\tilde{\mathcal{P}}_{t_1}$  process, by combining this with Equations (8.12) and (8.13),

$$\begin{aligned} &\Pr \left[ \bigcap_{s \in [t, \tilde{t}]} \{\Phi_2^s \leq 8n\} \mid \mathcal{Z}^{t_1} \right] \\ &\geq \sum_{r=t-n \log^4 n}^t \Pr \left[ \bigcap_{s \in [t, \tilde{t}]} \{\Phi_2^s \leq 8n\} \mid \mathcal{Z}^{t_1}, \mathfrak{F}^r, \Phi_2^r \leq 6n \right] \cdot \Pr [\tau = r \mid \mathcal{Z}^{t_1}] \\ &\stackrel{(8.12)}{\geq} \left(1 - \frac{1}{8} \cdot (\log^8 n) \cdot P\right) \cdot \Pr [\tau \leq t \mid \mathcal{Z}^{t_1}] \\ &\stackrel{(8.13)}{\geq} \left(1 - \frac{1}{8} \cdot (\log^8 n) \cdot P\right) \cdot (1 - n^{-6}) \geq 1 - \frac{1}{4} \cdot (\log^8 n) \cdot P. \quad \square \end{aligned}$$

We now return to the proof of Theorem 8.7 for the original process  $\mathcal{P}$ .

**PROOF OF THEOREM 8.7.** Let  $t_1 := t - n \log^4 n$ . We start by showing that the processes  $\mathcal{P}$  and  $\tilde{\mathcal{P}}_{t_1}$  agree with high probability in the interval  $[t_1, \tilde{t}]$  (and so at every step  $s \in [t_1, \tilde{t}]$  we have  $\Phi_2^s = \Phi_{\tilde{\mathcal{P}}_{t_1}, 2}^s$ ),

$$\begin{aligned} \Pr \left[ \bigcap_{s \in [t_1, \tilde{t}]} \{y^s = y_{\tilde{\mathcal{P}}_{t_1}}^s\} \right] &\geq \Pr \left[ \bigcap_{s \in [t-2n \log^4 n, \tilde{t}]} \{\Phi_1^s \leq 6n^{11}\} \cap \bigcap_{s \in [t-2n \log^4 n, \tilde{t}]} \mathcal{K}^s \right] \\ &\stackrel{(a)}{\geq} \Pr \left[ \bigcap_{s \in [t-2n \log^4 n, \tilde{t}]} \{\tilde{\Phi}_1^s \leq 6n^{11}\} \cap \mathcal{H}^{t-n \log^4 n} \cap \bigcap_{s \in [t-2n \log^4 n, \tilde{t}]} \mathcal{K}^s \right] \end{aligned}$$

$$\stackrel{(b)}{\geq} 1 - n^2 \cdot n^{-11} - P \stackrel{(c)}{\geq} 1 - 2P, \quad (8.14)$$

using in (a) the definition of  $\tilde{\Phi}_1$  in Equation (8.6), in (b) Lemma 8.5, Markov's inequality and union bound over  $\tilde{t} - (t - 2n \log^4 n) \leq n^2$  steps and precondition Equation (8.3) and in (c) that  $P \geq n^{-4}$ . Note that this also implies that

$$\Pr \left[ \Phi_1^{t-n \log^4 n} \leq 6n^{11} \right] \geq 1 - 2P. \quad (8.15)$$

By Lemma 8.8, for the auxiliary process  $\tilde{\mathcal{P}}_{t_1}$ , we have that

$$\Pr \left[ \bigcap_{s \in [t, \tilde{t}]} \left\{ \Phi_{\tilde{\mathcal{P}}_{t_1}, 2}^s \leq 8n \right\} \mid \Phi_1^{t-n \log^4 n} \leq 6n^{11} \right] \geq 1 - \frac{1}{4} \cdot (\log^8 n) \cdot P.$$

By combining with Equation (8.15) (since by definition  $\mathcal{P}$  and  $\tilde{\mathcal{P}}_{t_1}$  agree in steps  $s \leq t_1$ ) we get,

$$\Pr \left[ \bigcap_{s \in [t, \tilde{t}]} \left\{ \Phi_{\tilde{\mathcal{P}}_{t_1}, 2}^s \leq 8n \right\} \right] \geq \left( 1 - \frac{1}{4} \cdot (\log^8 n) \cdot P \right) \cdot (1 - 2P) \geq 1 - \frac{1}{2} \cdot (\log^8 n) \cdot P.$$

Finally, using  $\Pr[A \cap B] \geq \Pr[A] - \Pr[\neg B]$  with Equation (8.14), we conclude that  $\mathcal{P}$  and  $\tilde{\mathcal{P}}_{t_1}$  agree in all steps of the interval  $[t_1, \tilde{t}]$  and so

$$\Pr \left[ \bigcap_{s \in [t, \tilde{t}]} \left\{ \Phi_2^s \leq 8n \right\} \right] \geq 1 - \frac{1}{2} \cdot (\log^8 n) \cdot P - 2P \geq 1 - (\log^8 n) \cdot P. \quad \square$$

## 9 UPPER BOUND OF $O(\frac{g}{\log g} \cdot \log \log n)$ FOR $g$ -ADV-COMP WITH $g \leq \log n$ : LAYERED INDUCTION

In this section, we will complete the proof of the  $O(\frac{g}{\log g} \cdot \log \log n)$  upper bound for any  $g \leq \log n$ . We do this by first proving the key lemma for the drop of the super-exponential potentials (Lemma 9.1) and then complete the layered induction in Theorem 9.2.

We begin by recalling from Section 6.1 the definition for the number of steps in the layered induction  $k$ . Recall that  $\alpha_1 := \frac{1}{6\kappa}$ , for  $\kappa > 0$  the constant in Lemma 5.11 and  $k := k(g) \geq 2$  is the unique integer satisfying,

$$(\alpha_1 \log n)^{1/k} \leq g < (\alpha_1 \log n)^{1/(k-1)}.$$

Note that the integer  $k$  always satisfies  $k = O(\log \log n)$ , since  $g > 1$ .

Further, recall that for any integer  $1 \leq j \leq k-1$  we are using the super-exponential potentials  $\Phi_j := \Phi_j(\phi_j, z_j)$  with smoothing parameters  $\phi_j := \alpha_2 \cdot (\log n) \cdot g^{j-k}$  with  $\alpha_2 := \frac{1}{84} \cdot \alpha_1$  and offsets  $z_j := c_5 \cdot g + \left\lceil \frac{4}{\alpha_2} \right\rceil \cdot j \cdot g$  (for the constant  $c_5 > 0$  defined in Equation (7.14)), as well as the potentials  $\Psi_j := \Psi_j(\psi_j, z_j)$  with smoothing parameters  $\psi_j := \alpha_1 \cdot (\log n) \cdot g^{j-k}$ .

We are now ready to prove the key lemma for the layered induction.

**LEMMA 9.1.** *Consider the  $g$ -ADV-COMP setting for any  $g \in [\log(2C), \alpha_1 \log n]$ , for the constant  $C > 0$  defined in Theorem 7.8 and  $\alpha_1 > 0$  defined in Equation (6.2). Further, let  $k := k(g) \geq 2$  be the unique integer such that  $(\alpha_1 \log n)^{1/k} \leq g < (\alpha_1 \log n)^{1/(k-1)}$ . Then, for any integer  $1 \leq j \leq k-1$  and any step  $s \geq 0$ ,  $\Phi_{j-1}^s \leq Cn$  implies  $\mathcal{K}_{\psi_j, z_j}^s$ .*

**PROOF.** Consider an arbitrary step  $s$  with  $\Phi_{j-1}^s \leq Cn$ . Recall the definition of  $\mathcal{K}_{\psi_j, z_j}^s$ ,

$$\mathcal{K}_{\psi_j, z_j}^s(q^s) := \left\{ \forall i \in [n]: y_i^s \geq z_j - 1 \Rightarrow q_i^s \leq \frac{1}{n} \cdot e^{-\psi_j} \right\}.$$

Thus, we want to bound the probability to allocate to a bin  $i \in [n]$  with load  $y_i^s \geq z_j - 1$ . In order to do this, we will bound the number of bins  $\ell \in [n]$  for which the adversary can reverse the comparison of  $i$  and  $\ell$ , by bounding the ones with load  $y_\ell^s \geq z_j - 1 - g$ . Recall that  $z_j := c_5 \cdot g + \lceil \frac{4}{\alpha_2} \rceil \cdot j \cdot g$ . In the analysis below we make use of the following simple bound for  $1 \leq j \leq k - 1$ ,

$$z_j - 1 - g - z_{j-1} = \left\lceil \frac{4}{\alpha_2} \right\rceil \cdot g - 1 - g \geq \frac{3}{\alpha_2} \cdot g, \quad (9.1)$$

using that  $\alpha_2 \leq 1/2$ . We consider the cases  $j = 1$  and  $j > 1$  separately as  $\Phi_0$  has a slightly different form than  $\Phi_{j-1}$  for  $j > 1$ .

**Case 1 [ $j = 1$ ]:** The contribution of any bin  $\ell \in [n]$  with load  $y_\ell^s \geq z_1 - 1 - g$  to  $\Phi_0^s$  is,

$$\Phi_{0,\ell}^s = e^{\alpha_2 \cdot (y_\ell^s - z_0)^+} \geq e^{\alpha_2 \cdot (z_1 - 1 - g - z_0)^+} \stackrel{(9.1)}{\geq} e^{3g}.$$

Hence, when  $\{\Phi_0^s \leq Cn\}$  holds, the number of such bins is at most

$$Cn \cdot e^{-3g} = Cn \cdot e^{-g} \cdot e^{-2g} \leq \frac{n}{2} \cdot e^{-2g},$$

using that  $g \geq \log(2C)$ . Hence, the probability of allocating a ball to a bin  $i \in [n]$  with  $y_i^s \geq z_1 - 1$  is at most that of sampling  $i$  and a bin  $\ell \in [n]$  with  $y_\ell^s \geq z_1 - 1 - g$ , i.e., at most

$$q_i^s \leq 2 \cdot \frac{1}{n} \cdot \frac{1}{2} \cdot e^{-2g} \leq \frac{1}{n} \cdot e^{-\psi_1},$$

using that  $\psi_1 := \alpha_1 \cdot (\log n) \cdot g^{1-k} \leq g$ , as  $g \geq (\alpha_1 \log n)^{1/k}$ .

**Case 2 [ $j > 1$ ]:** The contribution of any bin  $\ell \in [n]$  with load  $y_\ell^s \geq z_j - 1 - g$  to  $\Phi_{j-1}^s$  is,

$$\Phi_{j-1,\ell}^s = e^{\alpha_2 \cdot (\log n) \cdot g^{j-1-k} \cdot (y_\ell^s - z_{j-1})} \stackrel{(9.1)}{\geq} e^{\alpha_2 \cdot (\log n) \cdot g^{j-1-k} \cdot (\frac{3}{\alpha_2} \cdot g)} = e^{3 \cdot (\log n) \cdot g^{j-k}}.$$

Hence, when  $\{\Phi_{j-1}^s \leq Cn\}$  holds, the number of such bins is at most

$$Cn \cdot e^{-3 \cdot (\log n) \cdot g^{j-k}} \leq \frac{n}{2} \cdot e^{-2 \cdot (\log n) \cdot g^{j-k}},$$

using that  $e^{-(\log n) \cdot g^{j-k}} \leq e^{-(\log n) \cdot g^{2-k}} \leq e^{-(\log n) \cdot g \cdot \frac{1}{\alpha_1 \log n}} \leq e^{-\log(2C)} = \frac{1}{2C}$ , since  $j > 1$ ,  $g \leq (\alpha_1 \log n)^{1/(k-1)}$ ,  $g \geq \log(2C)$  and  $\alpha_1 \leq 1$ .

Hence, the probability of allocating a ball to a bin  $i \in [n]$  with  $y_i^s \geq z_j - 1$  is at most that of sampling  $i$  and a bin  $\ell \in [n]$  with  $y_\ell^s \geq z_j - 1 - g$ , i.e., at most

$$q_i^s \leq 2 \cdot \frac{1}{n} \cdot \frac{1}{2} \cdot e^{-2 \cdot (\log n) \cdot g^{j-k}} = \frac{1}{n} \cdot e^{-2 \cdot (\log n) \cdot g^{j-k}} \leq \frac{1}{n} \cdot e^{-\psi_j},$$

recalling that  $\psi_j := \alpha_1 \cdot (\log n) \cdot g^{j-k}$  for  $\alpha_1 \leq 1$ .

Combining the two cases, we conclude that the event  $\mathcal{K}_{\psi_j, z_j}^s$  holds at step  $s$ .  $\square$

**THEOREM 9.2.** *Consider the  $g$ -ADV-COMP setting for any  $g \in (1, \log n]$ . Then, there exists a constant  $\tilde{\kappa} > 0$  such that for any step  $m \geq 0$ ,*

$$\Pr \left[ \text{Gap}(m) \leq \tilde{\kappa} \cdot \frac{g}{\log g} \cdot \log \log n \right] \geq 1 - n^{-3}.$$

**PROOF.** Let  $g_{\min} := \max\{\log(2C), \frac{\alpha_2}{4\sqrt{\alpha_1}}\}$ . We consider three cases depending on the value of  $1 \leq g \leq \log n$ :

**Case 1 [ $\min\{\frac{\alpha_2}{4}, c_6\} \cdot \log n \leq g \leq \log n$ ]:** (for  $c_6 > 0$  as defined in Equation (7.1)) In this case, the  $O(\log n)$  upper bound follows by the  $O(g + \log n)$  upper bound of Theorem 5.12.

**Case 2** [ $1 < g < g_{\min}$ ]: For  $1 < g \leq g_{\min}$ ,  $g$  is constant and the  $O(\log \log n)$  upper bound on the gap will follow by considering the  $\tilde{g}$ -ADV-COMP setting with  $\tilde{g} = \lceil g_{\min} \rceil$ . This setting encompasses  $g$ -ADV-COMP as  $\tilde{g} \geq g$  and is analyzed in Case 3.

**Case 3** [ $g_{\min} \leq g < \min\{\frac{\alpha_2}{4}, c_6\} \cdot \log n$ ]: Recall that for any  $g < \alpha_1 \log n$ , we defined the unique integer  $k := k(g) \geq 2$  satisfying,

$$(\alpha_1 \log n)^{1/k} \leq g < (\alpha_1 \log n)^{1/(k-1)},$$

and as explained in Equation (6.4), since  $g < \frac{\alpha_2}{4} \log n$ , we may assume that the following stronger condition holds

$$(\alpha_1 \log n)^{1/k} \leq g < \left(\frac{\alpha_2}{4} \log n\right)^{1/(k-1)},$$

where the inequalities are valid using Claim B.1 and that  $g \geq \frac{\alpha_2}{4\sqrt{\alpha_1}}$ .

Let  $t_j := m - 2n(k - j) \log^4 n$  for any integer  $0 \leq j \leq k - 1$ . We will proceed by induction on the potential functions  $\Phi_j$ . The base case follows by applying Theorem 7.8 (using  $g \leq c_6 \log n$  and  $t_0 \geq m - n \log^5 n$ ),

$$\Pr \left[ \bigcap_{t \in [t_0, m]} \{\Phi_0^t \leq Cn\} \right] \geq 1 - n^{-4}. \quad (9.2)$$

We will now prove the induction step.

**LEMMA 9.3 (INDUCTION STEP).** *Consider the  $g$ -ADV-COMP setting for any  $g \geq \max\{\log(2C), \frac{\alpha_2}{4\sqrt{\alpha_1}}\}$  satisfying  $(\alpha_1 \log n)^{1/k} \leq g < (\frac{\alpha_2}{4} \log n)^{1/(k-1)}$  for some integer  $k \geq 2$ , where  $C > 0$  is the constant defined in Theorem 7.8 and  $\alpha_1, \alpha_2 > 0$  are defined in Equations (6.2) and (6.3). Then, for any integer  $1 \leq j \leq k - 1$  and any step  $m \geq 0$ , if it holds that*

$$\Pr \left[ \bigcap_{t \in [t_{j-1}, m]} \{\Phi_{j-1}^t \leq Cn\} \right] \geq 1 - \frac{(\log n)^{8(j-1)}}{n^4},$$

*then it also follows that*

$$\Pr \left[ \bigcap_{t \in [t_j, m]} \{\Phi_j^t \leq Cn\} \right] \geq 1 - \frac{(\log n)^{8j}}{n^4}.$$

**PROOF OF LEMMA 9.3.** Consider an arbitrary integer  $j$  with  $1 \leq j \leq k - 1$  and assume that

$$\Pr \left[ \bigcap_{t \in [t_{j-1}, m]} \{\Phi_{j-1}^t \leq Cn\} \right] \geq 1 - \frac{(\log n)^{8(j-1)}}{n^4}.$$

By Lemma 9.1, we have that  $\{\Phi_{j-1}^t \leq Cn\}$  implies  $\mathcal{K}_{\psi_j, z_j}^t$ . Furthermore,  $\{\Phi_{j-1}^{t_{j-1}} \leq Cn\}$  also implies  $\{\text{Gap}(t_{j-1}) \leq \log^2 n\}$ . Hence, it also holds that

$$\Pr \left[ \left\{ \text{Gap}(t_{j-1}) \leq \log^2 n \right\} \cap \bigcap_{t \in [t_{j-1}, m]} \mathcal{K}_{\psi_j, z_j}^t \right] \geq 1 - \frac{(\log n)^{8(j-1)}}{n^4}. \quad (9.3)$$

Applying Lemma 8.1 for the potentials  $\Phi_j$  and  $\Psi_j$ , since  $\psi_j \geq \phi_j = \alpha_2 \cdot (\log n) \cdot g^{j-k} \geq \alpha_2 \cdot (\log n) \cdot g^{1-k} \geq 4$  (as  $g < (\frac{\alpha_2}{4} \log n)^{1/(k-1)}$ ), for any step  $t \geq 0$  it holds that

$$\mathbb{E} \left[ \Phi_j^{t+1} \mid \mathcal{F}^t, \mathcal{K}_{\psi_j, z_j}^t \right] \leq \Phi_j^t \cdot \left( 1 - \frac{1}{n} \right) + 2, \quad (9.4)$$

and

$$\mathbb{E} \left[ \Psi_j^{t+1} \mid \mathcal{F}^t, \mathcal{K}_{\psi_j, z_j}^t \right] \leq \Psi_j^t \cdot \left( 1 - \frac{1}{n} \right) + 2. \quad (9.5)$$

Hence, by Equation (9.4), Equations (9.5) and (9.3), the preconditions of Theorem 8.7 are satisfied for starting step  $t_j := m - 2n(k - j) \cdot \log^4 n$ ,  $P := (\log n)^{8(j-1)}/n^4$  and terminating step at  $\tilde{t} := m$ , and so we conclude (since  $C \geq 8$ ) that

$$\Pr \left[ \bigcap_{t \in [t_j, m]} \{ \Phi_j^t \leq Cn \} \right] \geq 1 - \frac{(\log n)^{8j}}{n^4}. \quad \square$$

Returning to the proof of Theorem 9.2, inductively applying Lemma 9.3 for  $k - 1$  times and using Equation (9.2) as a base case, we get that

$$\Pr \left[ \bigcap_{t \in [t_{k-1}, m]} \{ \Phi_{k-1}^t \leq Cn \} \right] \geq 1 - \frac{(\log n)^{8(k-1)}}{n^4} \geq 1 - n^{-3},$$

using in the last step that  $k = O(\log \log n)$ . When  $\{ \Phi_{k-1}^m \leq Cn \}$  occurs, the gap at step  $m$  cannot be more than  $z_k := c_5 g + \lceil \frac{4}{\alpha_2} \rceil \cdot k \cdot g$ , since otherwise we would get a contradiction

$$\begin{aligned} Cn &\geq \Phi_{k-1}^m \geq \exp \left( \alpha_2 \cdot (\log n) \cdot g^{(k-1)-k} \cdot (z_k - z_{k-1}) \right) \\ &= \exp \left( \alpha_2 \cdot (\log n) \cdot g^{-1} \cdot \left( \lceil \frac{4}{\alpha_2} \rceil \cdot g \right) \right) \geq \exp(4 \cdot \log n) = n^4. \end{aligned}$$

Hence,  $\text{Gap}(m) \leq z_k = c_5 g + \lceil \frac{4}{\alpha_2} \rceil \cdot k \cdot g$ .

By the assumption on  $g$ , we have

$$g < \left( \frac{\alpha_2}{4} \log n \right)^{1/(k-1)} \Rightarrow \log g < \frac{1}{k-1} \cdot \log \left( \frac{\alpha_2}{4} \log n \right) \Rightarrow k < 1 + \frac{\log \left( \frac{\alpha_2}{4} \log n \right)}{\log g},$$

using that  $g > 1$ . Since  $\alpha_2 > 0$  and  $c_5 > 0$  are constants, we conclude that there exists a constant  $\tilde{\kappa} > 0$  such that

$$\Pr \left[ \text{Gap}(m) \leq \tilde{\kappa} \cdot \frac{g}{\log g} \cdot \log \log n \right] \geq 1 - n^{-3}. \quad \square$$

In the above, we actually proved the following slightly stronger corollary, which we will use in Section 10.2. This is based on the insight that, in order to prove the above gap bound at step  $m$ , the only assumption on the steps  $[0, t_0)$ , with  $t_0 := m - n \log^5 n - \Delta_r$ , is that the coarse bound of  $O(g \log(ng))$  on the difference between maximum and minimum load must hold at step  $t_0$  (see, e.g., Lemma 5.9 and Theorem 7.8). While during the interval  $[t_0, m]$  the process is required to be an instance of  $g$ -ADV-COMP, in the interval  $[0, t_0)$  the process can be arbitrary as long as the coarse gap bound holds at step  $t_0$ .

**COROLLARY 9.4.** *Consider any  $g \in (1, \log n]$ ,  $m \geq 0$ ,  $t_0 := m - n \log^5 n - \Delta_r$  for  $\Delta_r := \Delta_r(g) > 0$  as defined in Lemma 5.9 and  $c_3 > 0$  the constant in Theorem 4.3. Further, consider a process which, in steps  $[t_0, m]$ , is an instance of  $g$ -ADV-COMP setting. Then, there exists a constant  $\tilde{\kappa} > 0$ , such that*

$$\Pr \left[ \text{Gap}(m) \leq \tilde{\kappa} \cdot \frac{g}{\log g} \cdot \log \log n \mid \mathcal{F}^{t_0}, \max_{i \in [n]} |y_i^{t_0}| \leq c_3 g \log(ng) \right] \geq 1 - n^{-3}.$$

## 10 UPPER BOUNDS FOR PROBABILISTIC NOISE AND DELAY SETTINGS

In this section, we present upper bounds for the probabilistic noise settings and the delay settings. These upper bounds follow from the results (and analysis) for  $g$ -ADV-COMP in Sections 4–9.

### 10.1 An Upper Bound for the Probabilistic Noise Setting

PROPOSITION 10.1. *Consider the  $\rho$ -NOISY-COMP setting with  $\rho(\delta)$  being any non-decreasing function in  $\delta$  with  $\lim_{\delta \rightarrow \infty} \rho(\delta) = 1$ . For any  $n \in \mathbb{N}$ , define  $\delta^* := \delta^*(n) = \min\{\delta \geq 1 : \rho(\delta) \geq 1 - n^{-4}\}$ . Then, there exists a constant  $\kappa > 0$ , such that for any step  $m \geq 0$ ,*

$$\Pr \left[ \max_{i \in [n]} |y_i^m| \leq \kappa \cdot \delta^* \log(n\delta^*) \right] \geq 1 - n^{-3}.$$

Note that for the  $\sigma$ -NOISY-LOAD process where  $\rho(\delta)$  has Gaussian tails (see Equation (2.1)), we have  $\delta^* = O(\sigma \cdot \sqrt{\log n})$ . The choice of  $\delta^*$  in Proposition 10.1 ensures that in most steps, all possible comparisons among bins with load difference greater than  $\delta^*$  will be correct, implying that the process satisfies the condition of  $g$ -ADV-COMP with  $g = \delta^*$ .

PROOF. We will analyze the hyperbolic cosine potential  $\Gamma := \Gamma(\gamma)$  as defined in Equation (4.1), with  $\gamma := -\log(1 - \frac{1}{8.48})/\delta^*$ . We first state a trivial upper bound on  $\mathbb{E}[\Gamma^{t+1} | \mathcal{F}^t]$  in terms of  $\Gamma^t$ , which holds deterministically for all steps  $t \geq 0$  (cf. Lemma 7.2 (i)),

$$\Gamma^{t+1} = \sum_{i=1}^n \Gamma_i^{t+1} \leq \sum_{i=1}^n e^\gamma \cdot \Gamma_i^t = e^\gamma \cdot \Gamma^t \leq (1 + 2\gamma) \cdot \Gamma^t,$$

using that  $e^\gamma \leq 1 + 2\gamma$  for  $0 < \gamma \leq 1$ .

We will now provide a better upper bound, exploiting that with high probability all possible comparisons between bins that differ by at least  $\delta^*$  will be correct. Again, consider any step  $t \geq 0$ . Let us assume that in step  $t$ , we first determine the outcome of the load comparisons among all  $n^2$  possible bin pairs. Only then we sample two bins, and allocate the ball following the pre-determined outcome of the load comparison. For any two bins  $i_1, i_2 \in [n]$  with  $|x_{i_1}^t - x_{i_2}^t| \geq \delta^*$ , we have

$$\rho(|x_{i_1}^t - x_{i_2}^t|) \geq \rho(\delta^*) \geq 1 - n^{-4}.$$

Hence by the union bound over all  $n^2$  pairs, we can conclude that with probability at least  $1 - n^{-2}$ , all comparisons among bin pairs with load difference at least  $\delta^*$  are correct. Let us denote this event by  $\mathcal{G}^t$ , so  $\Pr[\mathcal{G}^t] \geq 1 - n^{-2}$  for every step  $t \geq 0$ . Conditional on  $\mathcal{G}^t$ , the process in step  $t$  is an instance of  $g$ -ADV-COMP with  $g = \delta^*$ . Therefore, by Theorem 4.3 (i), there exists a constant  $c_1 \geq 1$ , so that the hyperbolic cosine potential satisfies

$$\mathbb{E}[\Gamma^{t+1} | \mathcal{F}^t, \mathcal{G}^t] \leq \Gamma^t \cdot \left(1 - \frac{\gamma}{96n}\right) + c_1.$$

Now combining our two upper bounds on  $\mathbb{E}[\Gamma^{t+1} | \mathcal{F}^t]$ , we conclude

$$\begin{aligned} \mathbb{E}[\Gamma^{t+1} | \mathcal{F}^t] &\leq \mathbb{E}[\Gamma^{t+1} | \mathcal{F}^t, \mathcal{G}^t] \cdot \Pr[\mathcal{G}^t] + \mathbb{E}[\Gamma^{t+1} | \mathcal{F}^t] \cdot (1 - \Pr[\mathcal{G}^t]) \\ &\stackrel{(a)}{\leq} \mathbb{E}[\Gamma^{t+1} | \mathcal{F}^t, \mathcal{G}^t] \cdot (1 - n^{-2}) + \mathbb{E}[\Gamma^{t+1} | \mathcal{F}^t] \cdot n^{-2} \\ &\leq \left(\Gamma^t \cdot \left(1 - \frac{\gamma}{96n}\right) + c_1\right) \cdot (1 - n^{-2}) + (1 + 2\gamma) \cdot \Gamma^t \cdot n^{-2} \\ &\leq \Gamma^t \cdot \left(1 - \frac{\gamma}{96n}\right) + c_1 - \Gamma^t \cdot n^{-2} + \Gamma^t \cdot \frac{\gamma}{96n} \cdot n^{-2} + \Gamma^t \cdot n^{-2} + \Gamma^t \cdot 2\gamma \cdot n^{-2} \\ &\leq \Gamma^t \cdot \left(1 - \frac{\gamma}{100n}\right) + c_1, \end{aligned}$$

using in (a) that  $\Pr [\mathcal{G}^t] \geq 1 - n^{-2}$ . So, using Lemma A.11 (ii) (with  $a = 1 - \frac{\gamma}{100n}$  and  $b = c_1$ ) since  $\Gamma^0 = 2n \leq \frac{100c_1}{\gamma} \cdot n$ , we get for any  $t \geq 0$ ,

$$\mathbb{E} [\Gamma^t] \leq \frac{100c_1}{\gamma} \cdot n.$$

Using Markov's inequality yields,  $\Pr[\Gamma^t > \frac{100c_1}{\gamma} \cdot n^4] \leq n^{-3}$ . Now the claim follows, since the event  $\{\Gamma^t \leq \frac{100c_1}{\gamma} \cdot n^4\}$  for  $\gamma = \Theta(\frac{1}{\delta^*})$ , implies that

$$\max_{i \in [n]} |y_i^t| \leq \frac{1}{\gamma} \log \left( \frac{100c_1}{\gamma} \cdot n^4 \right) = O(\delta^* \cdot \log(n\delta^*)). \quad \square$$

## 10.2 Upper Bounds for Delay Settings

In this section, we will prove tight upper bounds for the Two-CHOICE process in the  $\tau$ -DELAY setting and the  $b$ -BATCH process for a range of values for the delay parameter  $\tau$  and batch size  $b$ . In particular for  $\tau = n$ , we show:

**THEOREM 10.2.** *Consider the  $\tau$ -DELAY setting with  $\tau = n$ . Then, there exists a constant  $\kappa > 0$  such that for any step  $m \geq 0$ ,*

$$\Pr \left[ \text{Gap}(m) \leq \kappa \cdot \frac{\log n}{\log \log n} \right] \geq 1 - n^{-2}.$$

This implies the same upper bound for the  $b$ -BATCH process for  $b = n$ , since it is an instance of the  $\tau$ -DELAY setting with  $\tau = n$ . This bound improves the  $O(\log n)$  bound in [18, Theorem 1] and can be easily seen to be asymptotically tight due to the  $\Omega(\frac{\log n}{\log \log n})$  lower bound for ONE-CHOICE with  $n$  balls (Observation 11.6).

We will analyze  $\tau$ -DELAY using a more general approach which also works for other choices of the delay parameter  $\tau \leq n \log n$ . First note that  $\tau$ -DELAY is an instance of  $g_1$ -ADV-COMP with  $g_1 := \tau - 1$ . This follows since for all steps, a bin could be sampled (and be allocated to) at most  $\tau - 1$  times during the last  $\tau - 1$  steps. However, for a typical execution we expect each bin to be incremented much less frequently during the last  $\tau - 1$  steps, and thus the process is with high probability an instance of  $g_2$ -ADV-COMP for some  $g_2 \ll g_1$ .

To make this more specific, consider an arbitrary instance of  $g_1$ -ADV-COMP and define  $D^t$  as the set of pairs of bins (of unequal load) whose comparison is reversed with non-zero probability by the adversary in step  $t + 1$ ,

$$D^t(\mathcal{F}^t) := \{(i, j) \in [n] \times [n] : y_i^t > y_j^t \wedge \Pr[A^{t+1}(\mathcal{F}^t, i, j) = i] > 0\},$$

and then define the largest load difference that could be reversed by the process at step  $t + 1$ ,

$$g^t(\mathcal{F}^t) := \max_{(i, j) \in D^t(\mathcal{F}^t)} |y_i^t - y_j^t|.$$

This can be seen as the “effective  $g$ -bound” of the process in step  $t + 1$ . Note that  $g^t \leq g_1$  holds deterministically.

**LEMMA 10.3.** *Consider the  $g_1$ -ADV-COMP setting for any  $g_1 \in [1, n \log n]$ , and consider any  $g_2 \in [1, \log^2 n]$ . If for every step  $t \geq 0$ , we have*

$$\Pr[g^t \leq g_2] \geq 1 - n^{-6},$$

*then there exists a constant  $\tilde{\kappa} > 0$  such that for any step  $m \geq 0$ ,*

$$\Pr \left[ \text{Gap}(m) \leq \tilde{\kappa} \cdot \frac{g_2}{\log g_2} \cdot \log \log n \right] \geq 1 - n^{-2}.$$



We will now use Lemma 10.3 to prove Theorem 10.2.

**PROOF OF THEOREM 10.2.** Any bin can be allocated to at most  $n - 1$  times during an interval of  $\tau - 1 = n - 1$  steps, so an adversary with  $g_1 = n - 1$  who is also aware of the entire history of the process, can simulate  $\tau$ -DELAY at any step  $t$ , by using the allocation information in steps  $[t - n, t]$ .

To obtain the bound for  $g_2$ , note that in TWO-CHOICE, at each step two bins are sampled for each ball. So in  $n - 1$  steps of TWO-CHOICE, there are  $2(n - 1)$  bins sampled using ONE-CHOICE. By the properties of ONE-CHOICE (Lemma A.6), we have that for any consecutive  $n - 1$  allocations, with probability at least  $1 - n^{-6}$  we sample (and allocate to) no bin more than  $11 \log n / \log \log n$  times. In such a sequence of bin samples, we can simulate  $\tau$ -DELAY using  $g_2$ -ADV-COMP with  $g_2 = 11 \log n / \log \log n$ . Hence, for any step  $t \geq 0$ ,

$$\Pr \left[ g^t \leq g_2 \right] \geq 1 - n^{-6}.$$

Since the precondition of Lemma 10.3 holds for  $g_1 = n - 1$  and  $g_2 = 11 \log n / \log \log n$ , we get that there exists a constant  $\kappa > 0$ , such that

$$\Pr \left[ \text{Gap}(m) \leq \kappa \cdot \frac{\log n}{\log \log n} \right] \geq 1 - n^{-2}. \quad \square$$

The same argument also applies for any  $\tau \in [n \cdot e^{-\log^c n}, n \log n]$ . For ONE-CHOICE with  $2(\tau - 1)$  balls the gap is w.h.p. polylog( $n$ ) (e.g., see Lemma A.5) and so we can apply Lemma 10.3 with  $g_2 = \text{polylog}(n)$ , to obtain the gap bound of

$$\frac{g_2}{\log g_2} \cdot \log \log n = \Theta(g_2).$$

**COROLLARY 10.4.** *There exists a constant  $\kappa > 0$ , such that the  $\tau$ -DELAY setting with any  $\tau \in [n \cdot e^{-\log^c n}, n \log n]$ , where  $c > 0$  is any constant, for any step  $m \geq 0$ , it holds that*

$$\Pr \left[ \text{Gap}(m) \leq \kappa \cdot \frac{\log n}{\log \left( \frac{4n}{\tau} \cdot \log n \right)} \right] \geq 1 - n^{-2}.$$

*Remark 10.5.* A matching lower bound holds for the  $b$ -BATCH process for any batch size  $b \in [n \cdot e^{-\log^c n}, n \log n]$ . This follows by the lower bound for ONE-CHOICE with  $b$  balls (e.g., see Lemma A.10) which matches the gap of  $b$ -BATCH in the first batch (Observation 11.6).

Therefore, Corollary 10.4 and Remark 10.5 establish that w.h.p.  $\text{Gap}(m) = \Theta\left(\frac{\log n}{\log((4n/b) \cdot \log n)}\right)$  for the  $b$ -BATCH process for any  $b \in [n \cdot e^{-\log^c n}, n \log n]$ . However, the following remark (which also applies to the  $\tau$ -DELAY setting), establishes that there are regions where the  $b$ -BATCH process has an asymptotically worse gap than ONE-CHOICE with  $b$  balls (see also some empirical results in Figure 11 in Section 12):

*Remark 10.6.* For any  $\tau$  (or  $b$ ) being  $n^{1-\varepsilon}$  for any constant  $\varepsilon \in (0, 1)$ , the ONE-CHOICE process has  $\text{Gap}(b) = O(1)$  w.h.p. (see Lemma A.7). Hence, by Lemma 10.3 with  $g_2 = O(1)$ ,  $\tau$ -DELAY has for any step  $m \geq 0$ ,  $\text{Gap}(m) = O(\log \log n)$  w.h.p., which is asymptotically tight by Observation 11.1 for  $m = n$ .

We now return to proving Lemma 10.3.

**PROOF OF LEMMA 10.3.** Let  $\mathcal{P}$  be a process satisfying the preconditions in the statement and we will define the auxiliary process  $\tilde{\mathcal{P}}_{t_0}$  for some step  $t_0 \geq 0$  (to be specified below). Consider the stopping time  $\sigma := \inf\{s \geq t_0 : g^s > g_2\}$ . Then, the auxiliary process  $\tilde{\mathcal{P}}_{t_0}$  is defined so that

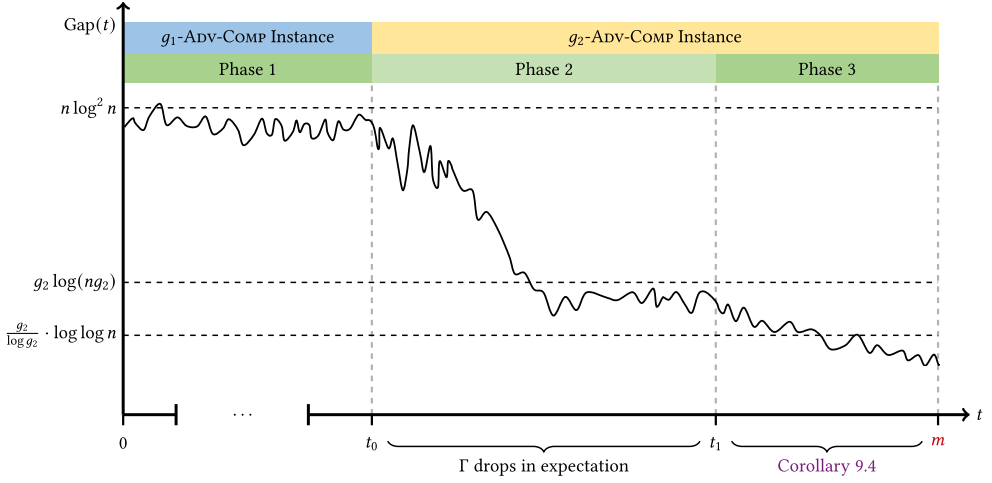


Fig. 9. The three phases in the proof of Lemma 10.3.

- in steps  $s \in [0, \sigma)$ , it makes the same allocations as  $\mathcal{P}$ , and
- in steps  $s \in [\sigma, \infty)$ , it makes the same allocations as the  $g_2$ -BOUNDED process.

This way  $\tilde{\mathcal{P}}_{t_0}$  is a  $g_2$ -ADV-COMP process for all steps  $s \geq t_0$ . Let  $\tilde{y}$  be the normalized load vector for  $\tilde{\mathcal{P}}_{t_0}$ , then it follows by the precondition that w.h.p. the two processes agree for any interval  $[t_0, m]$  with  $m - t_0 \leq n^3$ , i.e.,

$$\Pr \left[ \bigcap_{s \in [t_0, m]} \{y^s = \tilde{y}^s\} \right] \geq \Pr \left[ \bigcap_{s \in [t_0, m]} \{g^s \leq g_2\} \right] \geq 1 - n^{-6} \cdot n^3 = 1 - n^{-3}. \quad (10.1)$$

For  $m \leq n^3$ , the upper bound follows directly by Theorem 9.2 for  $\tilde{\mathcal{P}}_0$  and taking the union bound with Equation (10.1), i.e., that  $\tilde{\mathcal{P}}_0$  agrees with  $\mathcal{P}$ .

For  $m > n^3$ , the analysis is slightly more challenging. We need to show that the process recovers from the weak upper bound obtained from the  $g_1$ -ADV-COMP setting. Let  $\Gamma := \Gamma(\gamma)$  be as defined in Equation (4.1) with  $\gamma := -\log(1 - \frac{1}{8 \cdot 48})/g_2$  for the  $\tilde{\mathcal{P}}_{t_0}$  process (i.e., the  $\tilde{y}$  load vector). Also, let  $t_0 := m - n^3$  and  $t_1 := m - n \log^5 n - \Delta_r$  where  $\Delta_r := \Delta_r(g_2) = \Theta(n g_2 (\log(n g_2))^2)$  is the recovery time defined in Lemma 5.9. In this analysis, we consider the following three phases (see Figure 9):

- $[0, t_0]$ : The process  $\mathcal{P}$  is an instance of the  $g_1$ -ADV-COMP setting with  $g_1 = n \log n$ . Hence, by Theorem 4.3, it follows that at step  $t_0$  w.h.p.  $\text{Gap}(t_0) = O(n \log^2 n)$ .
- $(t_0, t_1]$ : The process  $\mathcal{P}$  w.h.p. agrees with  $\tilde{\mathcal{P}}_{t_0}$  which is a  $g_2$ -ADV-COMP process. We will use an analysis similar to that in Section 4 to prove that w.h.p.  $\text{Gap}(t_1) = O(g_2 \log(n g_2))$ .
- $(t_1, m]$ : The process  $\mathcal{P}$  w.h.p. continues to agree with  $\tilde{\mathcal{P}}_{t_0}$  which is a  $g_2$ -ADV-COMP process and so by Corollary 9.4 this implies that w.h.p.  $\text{Gap}(m) = O\left(\frac{g_2}{\log g_2} \cdot \log \log n\right)$ .

**Phase 1**  $[0, t_0]$ : Using Theorem 4.3 (iii) (for  $g := n \log n$ ) and for  $c_3 > 0$  being the constant defined in Equation (4.6), we have that,

$$\Pr \left[ \max_{i \in [n]} |\tilde{y}_i^{t_0}| \leq 3c_3 n \log^2 n \right] \geq 1 - n^{-14}, \quad (10.2)$$

using  $c_3 g \log(n g) \leq c_3 n \cdot \log(n^2 \cdot \log n) \leq 3c_3 n \log^2 n$ .

**Phase 2** ( $t_0, t_1$ ): Let us now turn our attention to the interval  $(t_0, t_1]$ , where  $\tilde{\mathcal{P}}_{t_0}$  is a  $g_2$ -ADV-COMP process. So, by Theorem 4.3 (i) (for  $g := g_2$ ), there exists a constant  $c_1 \geq 1$ , such that for any step  $t \geq t_0$

$$\mathbb{E} \left[ \Gamma^{t+1} \mid \mathcal{F}^t \right] \leq \Gamma^t \cdot \left( 1 - \frac{\gamma}{96n} \right) + c_1.$$

At step  $t_0$ , when  $\{\max_{i \in [n]} |\tilde{y}_i^{t_0}| \leq 3c_3 n \log^2 n\}$  holds, we also have that  $\Gamma^{t_0} \leq 2n \cdot e^{3\gamma c_3 n \log^2 n}$ . Hence, applying Lemma A.11 (i) (with  $a = 1 - \frac{\gamma}{96n}$  and  $b = c_1$ ), for step  $t_1$  we have

$$\begin{aligned} \mathbb{E} \left[ \Gamma^{t_1} \mid \left| \mathcal{F}^{t_0}, \max_{i \in [n]} |\tilde{y}_i^{t_0}| \leq 3c_3 n \log^2 n \right. \right] &\leq \mathbb{E} \left[ \Gamma^{t_1} \mid \mathcal{F}^{t_0}, \Gamma^{t_0} \leq 2n \cdot e^{3\gamma c_3 n \log^2 n} \right] \\ &\leq \Gamma^{t_0} \cdot \left( 1 - \frac{\gamma}{96n} \right)^{t_1 - t_0} + \frac{96c_1}{\gamma} \cdot n \\ &\stackrel{(a)}{\leq} 2n \cdot e^{3\gamma c_3 n \log^2 n} \cdot e^{-\frac{\gamma}{96n} \cdot \frac{1}{2} n^3} + \frac{96c_1}{\gamma} \cdot n \\ &\leq \frac{100c_1}{\gamma} \cdot n, \end{aligned}$$

using in (a) that  $e^u \geq 1 + u$  and  $t_1 - t_0 \geq \frac{1}{2} n^3$ . By Markov's inequality, we have that,

$$\Pr \left[ \Gamma^{t_1} \leq \frac{100c_1}{\gamma} \cdot n^4 \mid \left| \mathcal{F}^{t_0}, \max_{i \in [n]} |\tilde{y}_i^{t_0}| \leq 3c_3 n \log^2 n \right. \right] \geq 1 - n^{-3}.$$

When the event  $\{\Gamma^{t_1} \leq \frac{100c_1}{\gamma} \cdot n^4\}$  holds, it implies that

$$\text{Gap}(t_1) \leq \frac{1}{\gamma} \cdot \left( \log \left( \frac{100c_1}{\gamma} \right) + 4 \log n \right) \stackrel{(a)}{=} \frac{c_3 g_2}{16} \cdot \left( O(1) + \log \left( \frac{c_3 g_2}{16} \right) + 4 \log n \right) \leq c_3 g_2 \log(n g_2),$$

using in (a) that  $c_3 := \frac{16}{\gamma g_2}$  (defined in Equation (4.6)). Therefore,

$$\Pr \left[ \max_{i \in [n]} |\tilde{y}_i^{t_1}| \leq c_3 g_2 \log(n g_2) \mid \left| \mathcal{F}^{t_0}, \max_{i \in [n]} |\tilde{y}_i^{t_0}| \leq 3c_3 n \log^2 n \right. \right] \geq 1 - n^{-3}. \quad (10.3)$$

**Phase 3** ( $t_1, m$ ): Now, we turn our attention to the steps in  $(t_1, m]$ , where  $\tilde{\mathcal{P}}_{t_0}$  is again a  $g_2$ -ADV-COMP process. Therefore, applying Corollary 9.4 (for  $t_0 := t_1 = m - n \log^5 n - \Delta_r$  and  $g := g_2$ ), there exists a constant  $\tilde{\kappa} > 0$  such that

$$\Pr \left[ \text{Gap}_{\tilde{\mathcal{P}}_{t_0}}(m) \leq \tilde{\kappa} \cdot \frac{g_2}{\log g_2} \cdot \log \log n \mid \left| \mathcal{F}^{t_1}, \max_{i \in [n]} |\tilde{y}_i^{t_1}| \leq c_3 g_2 \log(n g_2) \right. \right] \geq 1 - n^{-3}. \quad (10.4)$$

By combining Equations (10.2), (10.3) and (10.4), we have that

$$\Pr \left[ \text{Gap}_{\tilde{\mathcal{P}}_{t_0}}(m) \leq \tilde{\kappa} \cdot \frac{g_2}{\log g_2} \cdot \log \log n \right] \geq (1 - n^{-14}) \cdot (1 - n^{-3}) \cdot (1 - n^{-3}) \geq 1 - 3n^{-3}.$$

Finally, by Equation (10.1) we have that w.h.p.  $\mathcal{P}$  and  $\tilde{\mathcal{P}}_{t_0}$  agree in every step in  $[t_0, m]$ , so by taking the union bound we conclude

$$\Pr \left[ \text{Gap}_{\mathcal{P}}(m) \leq \tilde{\kappa} \cdot \frac{g_2}{\log g_2} \cdot \log \log n \right] \geq 1 - 3n^{-3} - n^{-3} \geq 1 - n^{-2}. \quad \square$$

Table 2. Overview of the Lower Bounds for Different Noise Settings

Process	Range	Lower Bound	Reference
Any $g$ -ADV-COMP	$0 \leq g$	$\log_2 \log n + \Omega(1)$	Obs 11.1
$g$ -MYOPIC-COMP	$2 \leq g$	$\Omega(g)$	Prop 11.2
$g$ -MYOPIC-COMP	$10 \leq g \leq \frac{1}{8} \cdot \frac{\log n}{\log \log n}$	$\Omega\left(\frac{g}{\log g} \cdot \log \log n\right)$	Thm 11.3
$\sigma$ -NOISY-LOAD	$2 \cdot (\log n)^{-1/3} \leq \sigma$	$\Omega(\min\{1, \sigma\} \cdot (\log n)^{1/3})$	Prop 11.5
$\sigma$ -NOISY-LOAD	$32 \leq \sigma$	$\Omega(\min\{\sigma^{4/5}, \sigma^{2/5} \cdot \sqrt{\log n}\})$	Prop 11.5
$b$ -BATCH	$b \in [n \cdot e^{\log^{\Omega(1)} n}, n \log n]$	$\Omega\left(\frac{\log n}{\log(\frac{4n}{b} \cdot \log n)}\right)$	Obs 11.6

All of these hold for a particular value of  $m$  with high probability.

## 11 LOWER BOUNDS

In this section, we will state and prove lower bounds on the gap of various processes and settings. Recall that the  $g$ -BOUNDED and  $g$ -MYOPIC-COMP processes are specific instances of the  $g$ -ADV-COMP setting. Our main result is for  $g$ -MYOPIC-COMP, where we will prove a lower bound of  $\Omega(g + \frac{g}{\log g} \cdot \log \log n)$ , which matches the upper bounds of Sections 5–9 for all  $g \geq 0$  (Corollary 11.4). Table 2 gives a summary of the results.

We first start with a simple lower bound which follows immediately by majorization with the TWO-CHOICE process without noise. This lower bound holds regardless of which strategy the adversary uses.

**OBSERVATION 11.1.** *There is a constant  $\kappa > 0$  such that for any  $g \geq 0$  and any instance of the  $g$ -ADV-COMP setting,*

$$\Pr [\text{Gap}(n) \geq \log_2 \log n - \kappa] \geq 1 - n^{-1}.$$

**PROOF.** For the TWO-CHOICE process without noise, it was shown in [9, Theorem 3.3] that there is a constant  $\kappa > 0$  such that  $\Pr [\text{Gap}(n) \geq \log_2 \log n - \kappa] \geq 1 - n^{-1}$ .

At any step  $t \geq 0$ , probability allocation vector  $p$  of TWO-CHOICE without noise is majorized the probability allocation vector  $q^t$  of any instance of TWO-CHOICE in the  $g$ -ADV-COMP setting, as  $q^t$  is formed by  $p$  and possibly reallocating some probability mass from light to heavy bins. Hence, the lower bound follows by majorization (see Lemma A.13).  $\square$

We proceed by analyzing the  $g$ -MYOPIC-COMP process by coupling its allocations with that of a ONE-CHOICE process.

**PROPOSITION 11.2.** *Consider the  $g$ -MYOPIC-COMP process. Then, (i) for any  $g \in [2, 6 \log n]$  and for  $m = \frac{1}{2} \cdot ng$ , it holds that*

$$\Pr \left[ \text{Gap}(m) \geq \frac{1}{35} \cdot g \right] \geq 1 - n^{-2}.$$

*Further, (ii) for any  $g \geq 6 \log n$  and for  $m = ng^2/(32 \log n)$ ,*

$$\Pr \left[ \text{Gap}(m) \geq \frac{1}{60} \cdot g \right] \geq 1 - 2n^{-2}.$$

Recall that for  $g = \Omega(\text{polylog}(n))$ , our upper bound on the gap is

$$O\left(g + \frac{g}{\log g} \cdot \log \log n\right) = O(g).$$

This means that the lower bound in Proposition 11.2 is matching for those  $g$ . For smaller values of  $g$ , a stronger lower bound will be presented in Theorem 11.3.

**PROOF.** *First statement.* Consider  $g$ -MYOPIC-COMP with  $m = \frac{1}{2} \cdot ng$  balls and define the stopping time  $\tau := \inf\{t \geq 0 : \max_{i \in [n]} x_i^t \geq g\}$ . Note that  $\tau \leq m$  implies there is a bin  $j$  with  $x_j^\tau \geq g$ , and hence

$$\text{Gap}(m) \geq x_j^m - \frac{m}{n} \geq x_j^\tau - \frac{1}{2}g \geq g - \frac{1}{2}g > \frac{1}{35}g.$$

Let us now assume  $\tau > m$ . In that case, all bins have an absolute load in  $[0, g]$  and are therefore indistinguishable. Hence during steps  $1, 2, \dots, m$ , the  $g$ -MYOPIC-COMP process behaves exactly like ONE-CHOICE.

By Lemma A.9 (for  $c := \frac{1}{2} \cdot \frac{g}{\log n} \geq \frac{1}{\log n}$  since  $g \geq 2$ ) it holds for ONE-CHOICE that,

$$\Pr \left[ \text{Gap}(m) \geq \frac{1}{10\sqrt{2}} \cdot \sqrt{g \cdot \log n} \right] \geq 1 - n^{-2},$$

which, as  $g \cdot \log n \geq \frac{1}{6}g^2$ , implies that

$$\Pr \left[ \text{Gap}(m) \geq \frac{1}{35} \cdot g \right] \geq 1 - n^{-2}.$$

*Second statement.* We will first prove for the  $g$ -MYOPIC-COMP process that w.h.p. for any step  $1 \leq t \leq m$  and any bin  $i \in [n]$ , it holds that  $|y_i^t| \leq \frac{g}{2}$ . To this end, fix any bin  $i \in [n]$  and define for any  $1 \leq t \leq m$ ,  $Z^t = Z^t(i) := \sum_{j=1}^t (Y_j - \frac{1}{n})$ , where the  $Y_j$ 's are independent Bernoulli variables with parameter  $1/n$  each. Clearly,  $Z^t$  forms a martingale, and  $E[Z^t] = 0$  for all  $1 \leq t \leq m$ . Further, let  $\tau := \inf\{t \geq 1 : |Z^t| > \frac{g}{2}\}$ . Then  $Z^{t \wedge \tau}$  is also martingale. Further, we have

$$\text{Var}[Z^{t+1} \mid Z^t] = \frac{1}{n} \cdot \left(1 - \frac{1}{n}\right) \leq \frac{1}{n} =: \sigma^2.$$

Also  $|Z^{t+1} - Z^t| \leq 1 =: M$ . Hence by a martingale inequality Lemma A.2,

$$\Pr \left[ |Z^{m \wedge \tau}| \geq \lambda \right] \leq 2 \cdot \exp \left( -\frac{\lambda^2}{2(\sum_{i=1}^m \sigma^2 + M\lambda/3)} \right),$$

with  $\lambda = \frac{g}{2}$ ,

$$\Pr \left[ |Z^{m \wedge \tau}| \geq \frac{g}{2} \right] \leq 2 \cdot \exp \left( -\frac{g^2}{8(m \cdot \frac{1}{n} + \frac{g}{6})} \right) = 2 \cdot \exp \left( -\frac{g^2}{8 \cdot (\frac{g^2}{32 \log n} + \frac{g}{6})} \right) \leq 2n^{-3},$$

where the last inequality holds since  $8 \cdot (\frac{g^2}{32 \log n} + \frac{g}{6}) \leq 8 \cdot (\frac{g^2}{32 \log n} + \frac{g^2}{32 \log n}) = \frac{g^2}{2 \log n}$ , using that  $g \geq 6 \log n$ .

If the event  $|Z^{m \wedge \tau}| \leq \frac{g}{2}$  (or equivalently  $\tau > m$ ) occurs, then this implies that bin  $i$  deviates from the average load by at most  $\frac{g}{2}$  in all steps  $1, 2, \dots, m$ . By the union bound, this holds with probability  $1 - 2n^{-2}$  for all  $n$  bins. Consequently, the  $g$ -MYOPIC-COMP process behaves exactly like ONE-CHOICE until time  $m$  with probability  $1 - 2n^{-2}$ .

By Lemma A.9 (for  $c := \frac{g^2}{32 \log^2 n}$ ) it holds for ONE-CHOICE that,

$$\Pr \left[ \text{Gap}(m) \geq \frac{g}{10 \cdot \sqrt{32}} \right] \geq 1 - n^{-2}.$$

The claim follows by taking the union bound and using that  $\frac{g}{10 \cdot \sqrt{32}} \geq \frac{g}{60}$ .  $\square$

**THEOREM 11.3.** *Consider the  $g$ -MYOPIC-COMP process for any  $g \in [10, \frac{1}{8} \cdot \frac{\log n}{\log \log n}]$ . Then, there exists  $\ell := \ell(g, n)$  (defined in Equation (11.1)), such that for  $m = n \cdot \ell$ , it holds that*

$$\Pr \left[ \text{Gap}(m) \geq \frac{1}{8} \cdot \frac{g}{\log g} \cdot \log \log n \right] \geq 1 - n^{-\omega(1)}.$$

The proof of this lower bound is similar to the method used by Azar, Broder, Karlin and Upfal [9] to prove the  $\Omega(\log \log n)$  lower bound for Two-CHOICE without noise, in the sense that it follows a layered induction approach, but some additional care is needed. For example, the induction step size in the load is  $g$  and not 1, and the outcome of a load comparison depends on the load difference.

**PROOF.** In the proof we will divide the allocation of  $m = n \cdot \ell$  balls into  $\ell$  consecutive phases, each of which lasts for  $n$  steps, where

$$\ell := \left\lfloor \frac{\log(\frac{1}{8} \log n / \log g)}{\log g} \right\rfloor. \quad (11.1)$$

First, we verify that  $\ell \geq 1$ ,

$$\ell = \left\lfloor \frac{\log(\frac{1}{8} \log n / \log g)}{\log g} \right\rfloor \stackrel{(a)}{\geq} \left\lfloor \frac{\log(\frac{1}{8} \log n / \log g)}{\log(\frac{1}{8} \log n / \log n)} \right\rfloor \stackrel{(b)}{\geq} \left\lfloor \frac{\log(\frac{1}{8} \log n / \log n)}{\log(\frac{1}{8} \log n / \log n)} \right\rfloor = 1,$$

where (a) used that  $g \leq \frac{1}{8} \log n / \log \log n$  and (b) that  $g \leq \log n$ .

Next we define for any  $k = 1, 2, \dots, \ell$  the following event:

$$\mathcal{E}_k := \left\{ \left| \left\{ i \in [n] : x_i^{k \cdot n} \geq k \cdot g \right\} \right| \geq n \cdot g^{-\sum_{j=1}^k g^j} \right\}.$$

The main goal of this proof is to show that  $\mathcal{E}_\ell$  occurs with high probability. Assuming for the moment that  $\mathcal{E}_\ell$  indeed occurs, let us verify that the lower bound on the gap follows. First, note that  $\mathcal{E}_\ell$  implies the existence of a bin  $i \in [n]$  with  $x_i^m \geq \ell \cdot g$  and thus  $\text{Gap}(m) \geq x_i^m - \frac{m}{n} \geq \ell \cdot g - \ell = \ell \cdot (g - 1)$ , since by the choice of  $\ell = \lfloor \frac{\log(\frac{1}{8} \log n / \log g)}{\log g} \rfloor$ ,

$$g^{-\sum_{j=1}^k g^j} \stackrel{(a)}{\geq} g^{-4g^\ell} \geq g^{-\frac{1}{2} \log n / \log g} = n^{-1/2}, \quad (11.2)$$

where in (a) we used that for any  $g \geq 2$ ,

$$\sum_{j=1}^k g^j \leq 2 \sum_{j=0}^{\ell} g^j = 2 \cdot \frac{g^{\ell+1} - 1}{g - 1} \leq 2 \cdot \frac{g^{\ell+1}}{\frac{1}{2}g} = 4g^\ell. \quad (11.3)$$

Thus  $n \cdot g^{-\sum_{j=1}^k g^j} \geq 1$ . Secondly, we verify that  $\ell \cdot (g - 1) = \Omega(\frac{g}{\log g} \cdot \log \log n)$ ,

$$\begin{aligned} \ell \cdot (g - 1) &\stackrel{(a)}{\geq} \frac{1}{4} \cdot \frac{\log(\frac{1}{8} \log n / \log g)}{\log g} \cdot g \\ &= \frac{1}{4} \cdot \frac{\log(1/8) + \log \log n - \log \log g}{\log g} \cdot g \\ &\stackrel{(b)}{\geq} \frac{1}{8} \cdot \frac{\log \log n}{\log g} \cdot g, \end{aligned} \quad (11.4)$$

where (a) holds since  $\lfloor u \rfloor \geq u/2$  for  $u \geq 1$  (as  $\ell \geq 1$ ) and  $g - 1 \geq g/2$ , since  $g \geq 2$ , and (b) holds since for sufficiently large  $n$ ,  $(1/4) \cdot \log \log n \geq -\log(1/8)$  and  $(1/4) \cdot \log \log n \geq \log \log g$  as  $g \leq \log n$ .

In order to establish that  $\mathcal{E}_\ell$  occurs w.h.p., we will proceed by induction and prove that for any  $k \geq 1$ , with  $\varepsilon_k := n^{-\omega(1)}$  and  $\mathcal{E}_0 := \Omega$ ,

$$\Pr[\mathcal{E}_k \mid \mathcal{E}_{k-1}] \geq 1 - \varepsilon_k.$$

**Induction Base** ( $k = 1$ ). Here we consider the allocation of the first  $n$  balls into  $n$  bins. We are interested in the number of bins which reach load level  $g$  during that phase. Note that as long as the loads of both sampled bins are smaller than  $g$ , the allocation follows that of ONE-CHOICE; if one of the bins has a load which is already larger than  $g$ , then the load difference may force the process to place a ball in the lighter of the two bins. In the following, we will (pessimistically) assume that the allocation of *all*  $n$  balls follows ONE-CHOICE.

Instead of the original ONE-CHOICE process which produces a load vector  $(x_i^n)_{i \in [n]}$ , we consider the Poisson Approximation (Lemma A.3) and analyze the load vector  $(\tilde{x}_i^n)_{i \in [n]}$ , where the  $\tilde{x}_i^n$ ,  $i \in [n]$  are independent Poisson random variables with mean  $n/n = 1$ . Clearly, for any  $i \in [n]$ ,

$$\Pr[\tilde{x}_i^n \geq g] \geq \Pr[\tilde{x}_i^n = g] = e^{-1} \cdot \frac{1^g}{g!} \geq 2 \cdot g^{-g},$$

using in the last inequality Stirling's approximation (e.g., [53, Lemma 5.8]) and  $g \geq 5$ ,

$$g! \leq e \cdot g \cdot \left(\frac{g}{e}\right)^g = \frac{e^{-1}}{2} \cdot g^g \cdot \left(\frac{2e^2 g}{e^g}\right) \leq \frac{e^{-1}}{2} \cdot g^g.$$

Let  $Y := |\{i \in [n] : \tilde{x}_i^n \geq g\}|$ . Then  $\mathbb{E}[Y] \geq 2n \cdot g^{-g}$ . By a standard Chernoff Bound,

$$\Pr[Y \leq n \cdot g^{-g}] \leq \Pr\left[Y \leq \frac{1}{2} \cdot \mathbb{E}[Y]\right] \leq \exp\left(-\frac{1}{8} \cdot \mathbb{E}[Y]\right) \leq n^{-\omega(1)},$$

where we have used  $g^{-g} \geq g^{-4g^\ell} \geq n^{-1/2}$ , due to Equation (11.2). Hence by the Poisson Approximation (Lemma A.3),  $\Pr[\mathcal{E}_1] \geq 1 - 2 \cdot n^{-\omega(1)} = 1 - n^{-\omega(1)}$ .

**Induction Step** ( $k - 1 \rightarrow k, k \geq 2$ ). For the induction step, we analyze phase  $k = 2, \dots, \ell$  and we will lower bound  $\Pr[\mathcal{E}_k \mid \mathcal{E}_{k-1}]$ . Assuming  $\mathcal{E}_{k-1}$  occurs, there are at least  $n \cdot g^{-\sum_{j=1}^{k-1} g^j}$  bins whose load is at least  $(k-1) \cdot g$  at the beginning of phase  $k$ . Let us call such a set of bins  $\mathcal{B}_{k-1}$ , which can be assumed to satisfy with equality:

$$|\mathcal{B}_{k-1}| = n \cdot g^{-\sum_{j=1}^{k-1} g^j}. \quad (11.5)$$

Next note that whenever we sample two bins  $i_1, i_2$  from  $\mathcal{B}_{k-1}$ , we allocate the ball to a random bin among  $\{i_1, i_2\}$  if both bins have not reached load level  $k \cdot g$ . Therefore, in order to lower bound the number of bins in  $\mathcal{B}_{k-1}$  which reach load level  $k \cdot g$  by the end of phase  $k$ , we may pessimistically assume that if two bins in  $\mathcal{B}_{k-1}$  are sampled, the ball will be *always* placed in a randomly sampled bin. Note that the probability that a ball will be allocated into the set  $\mathcal{B}_{k-1}$  is lower bounded by  $(\frac{|\mathcal{B}_{k-1}|}{n})^2$ . Let  $Z$  denote the number of balls allocated to  $\mathcal{B}_{k-1}$  in phase  $k$ . Then  $\mathbb{E}[Z] \geq n \cdot (\frac{|\mathcal{B}_{k-1}|}{n})^2$ . Using a Chernoff Bound for  $\tilde{m} := \frac{2}{3} \cdot n \cdot (\frac{|\mathcal{B}_{k-1}|}{n})^2$ ,

$$\Pr[Z \leq \tilde{m}] \leq \Pr\left[Z \leq \frac{2}{3} \cdot \mathbb{E}[Z]\right] \leq \exp\left(-\frac{1}{18} \cdot \mathbb{E}[Z]\right) \leq n^{-\omega(1)},$$

where we used that

$$\mathbb{E}[Z] \geq n \cdot g^{-2 \sum_{j=1}^{k-1} g^j} \stackrel{(a)}{\geq} n \cdot g^{-2g^k} \stackrel{(b)}{\geq} n \cdot n^{-1/2},$$

using in (a) that for any integer  $g \geq 2$ ,  $\sum_{j=1}^{k-1} g^j \leq \frac{g^k - 1}{g - 1} \leq g^k$ , and in (b) that  $k \leq \ell$  and the property of  $\ell$  in Equation (11.2).

Conditioning on this event, in phase  $k$  we have a ONE-CHOICE process with  $\tilde{m}$  balls into  $\tilde{n} := |\mathcal{B}_{k-1}|$  bins, which w.l.o.g. will be labeled  $1, 2, \dots, \tilde{n}$ . Again, we apply the Poisson approximation and define  $(\tilde{x}^m)_{i \in [\tilde{n}]}$  as  $\tilde{n}$  independent Poisson random variables with mean  $\lambda$  given by

$$\lambda := \frac{\tilde{m}}{\tilde{n}} = \frac{\frac{2}{3} \cdot n \cdot \left(\frac{|\mathcal{B}_{k-1}|}{n}\right)^2}{|\mathcal{B}_{k-1}|} = \frac{2}{3} \cdot \frac{|\mathcal{B}_{k-1}|}{n} = \frac{2}{3} \cdot g^{-\sum_{j=1}^{k-1} g^j},$$

using Equation (11.5). With that, it follows for any bin  $i \in \mathcal{B}_{k-1}$ ,

$$\Pr[\tilde{x}_i \geq g] \geq \Pr[\tilde{x}_i = g] = e^{-\lambda} \cdot \frac{\lambda^g}{g!} \stackrel{(a)}{\geq} e^{-1} \cdot \frac{\left(\frac{2}{3} \cdot g^{-\sum_{j=1}^{k-1} g^j}\right)^g}{g!} \stackrel{(b)}{\geq} 2 \cdot g^{-\sum_{j=1}^k g^j},$$

using in (a) that  $\lambda \leq 1$  and in (b) Stirling's approximation (e.g., [53, Lemma 5.8]) and that  $g \geq 10$ ,

$$g! \leq e \cdot g \cdot \left(\frac{g}{e}\right)^g = \frac{e^{-1}}{2} \cdot \left(\frac{2}{3}\right)^g \cdot g^g \cdot \left(2e^2 g \cdot \left(\frac{3}{2e}\right)^g\right) \leq \frac{e^{-1}}{2} \cdot \left(\frac{2}{3}\right)^g \cdot g^g.$$

Let  $Y := |\{i \in [\tilde{n}] : \tilde{x}_i \geq g\}|$ . Then  $\mathbb{E}[Y] \geq 2\tilde{n} \cdot g^{-\sum_{j=1}^k g^j}$ . Thus by a Chernoff Bound,

$$\Pr\left[Y \leq \tilde{n} \cdot g^{-\sum_{j=1}^k g^j}\right] \leq \Pr\left[Y \leq \frac{1}{2} \cdot \mathbb{E}[Y]\right] \leq \exp\left(-\frac{1}{8} \cdot \mathbb{E}[Y]\right) \leq n^{-\omega(1)},$$

where we used that

$$\mathbb{E}[Y] \geq 2\tilde{n} \cdot g^{-\sum_{j=1}^k g^j} = 2n \cdot g^{-\sum_{j=1}^{k-1} g^j} \cdot g^{-\sum_{j=1}^k g^j} \geq 2n \cdot g^{-2\sum_{j=1}^k g^j} \stackrel{(11.3)}{\geq} 2n \cdot g^{-4g^k} \stackrel{(a)}{\geq} 2n^{1/2},$$

where in (a) we used  $k \leq \ell$  and Equation (11.2). Thus by the union bound and Poisson Approximation,  $\Pr[\mathcal{E}_k \mid \mathcal{E}_{k-1}] \geq 1 - n^{-\omega(1)} - 2 \cdot n^{-\omega(1)} = 1 - n^{-\omega(1)}$ , which completes the induction.

Finally, by a simple union bound,

$$\Pr[\mathcal{E}_\ell] \geq 1 - \sum_{k=1}^{\ell} \Pr[\neg \mathcal{E}_k \mid \mathcal{E}_{k-1}] \geq 1 - \ell \cdot n^{-\omega(1)} \geq 1 - n^{-\omega(1)}.$$

As verified in Equation (11.4),  $\mathcal{E}_\ell$  implies  $\text{Gap}(m) \geq \frac{1}{8} \cdot \frac{g}{\log g} \cdot \log \log n$ , and therefore the proof is complete.  $\square$

Combining Observation 11.1, Proposition 11.2 and Theorem 11.3, we get:

**COROLLARY 11.4.** *Consider the  $g$ -MYOPIC-COMP process for any  $g \geq 1$ . Then, there exists an  $m := m(g) \geq 0$ , such that*

$$\Pr\left[\text{Gap}(m) = \Omega\left(g + \frac{g}{\log g} \cdot \log \log n\right)\right] \geq 1 - n^{-1}.$$

We proceed with two lower bounds for the  $\sigma$ -NOISY-LOAD process.

**PROPOSITION 11.5.** *Consider the  $\sigma$ -NOISY-LOAD process with  $\rho(\delta) := 1 - \frac{1}{2} \cdot \exp(-(\delta/\sigma)^2)$  for some (not necessarily constant)  $\sigma > 0$ . Then, (i) for any  $\sigma \geq 2 \cdot (\log n)^{-1/3}$ ,*

$$\Pr\left[\text{Gap}(n) \geq \min\left\{\frac{1}{8} \cdot (\log n)^{1/3}, \frac{1}{2} \sigma \cdot (\log n)^{1/3}\right\}\right] \geq 1 - 2n^{-1}.$$

*Further, (ii) for any  $\sigma \geq 32$ , it holds for  $m = \frac{1}{2} \sigma^{4/5} \cdot n$ ,*

$$\Pr\left[\text{Gap}(m) \geq \min\left\{\frac{1}{2} \sigma^{4/5}, \frac{1}{30} \sigma^{2/5} \cdot \sqrt{\log n}\right\}\right] \geq 1 - 2n^{-2}.$$



PROOF. *First statement.* Let  $\tau := \inf\{t \geq 1 : \max_{i \in [n]} x_i^t \geq \sigma \cdot (\log n)^{1/3}\}$ . If  $\tau \leq n$ , then

$$\text{Gap}(n) \geq x_i^n - \frac{n}{n} \geq x_i^\tau - 1 \geq \sigma \cdot (\log n)^{1/3} - 1 \geq \frac{1}{2} \sigma \cdot (\log n)^{1/3}.$$

Consider now the case  $\tau > n$ . For any step  $t \leq \tau$ , we will perform a “sub-sampling” of the correct comparison in (possibly) two stages as follows. Let  $i_1 = i_1^t$  and  $i_2 = i_2^t$  be the two sampled bins, and  $\delta$  be their load difference. Let  $Z_1 \sim \text{Ber}(1 - \exp(-(\delta/\sigma)^2))$  and  $Z_2 \sim \text{Ber}(1/2)$  be two independent random variables, which can be thought as the outcome of two biased coin flips that will be used to determine whether the load comparison is correct. If  $Z_1 = 1$ , then the comparison is correct (regardless of what  $Z_2$  is). However, if  $Z_1 = 0$ , then the comparison is correct if and only if  $Z_2 = 1$ . Overall, the probability of a correct comparison (if  $\delta > 0$ ) is equal to

$$\begin{aligned} \Pr[Z_1 = 1] + \Pr[Z_1 = 0] \cdot \Pr[Z_2 = 1 \mid Z_1 = 0] &= 1 - \exp(-(\delta/\sigma)^2) + \exp(-(\delta/\sigma)^2) \cdot \frac{1}{2} \\ &= 1 - \frac{1}{2} \exp(-(\delta/\sigma)^2). \end{aligned}$$

Further, conditional on  $Z_1 = 0$ , the ball will be placed in a random bin among  $\{i_1, i_2\}$ . Hence as long as  $t \leq \tau$ , we can couple the allocation of each ball by  $g$ -MYOPIC-COMP to an allocation by ONE-CHOICE with probability at least  $\exp(-(\delta/\sigma)^2) \geq \exp(-(\log n)^{2/3})$ . Using a Chernoff bound, with probability  $1 - n^{-\omega(1)}$ , we can couple the allocation of at least  $n/2 \cdot \exp(-(\log n)^{2/3})$  balls out of the first  $n$  balls with that of ONE-CHOICE. Consequently, using Lemma A.8 the maximum load (and gap) is at least  $\frac{1}{8}(\log n)^{1/3}$  with probability at least  $1 - n^{-1}$ . Combining the two cases we get the claim.

*Second Statement.* Consider any  $\sigma \geq 32$  and let  $m := \frac{1}{2}\sigma^{4/5} \cdot n$ . Next, define the stopping time  $\tau := \inf\{t \geq 1 : \max_{i \in [n]} x_i^t \geq \sigma^{4/5}\}$ . If  $\tau \leq m$ , then there is a bin  $i \in [n]$  with  $x_i^\tau \geq \sigma^{4/5}$ , and

$$\text{Gap}(m) \geq x_i^m - \frac{m}{n} \geq x_i^\tau - \frac{1}{2}\sigma^{4/5} \geq \frac{1}{2}\sigma^{4/5}.$$

Otherwise, in each step until  $m$ , the load difference between any two sampled bins is at most  $\sigma^{4/5}$ , and therefore each ball is coupled with the allocation of a ONE-CHOICE process with probability at least  $\exp(-(\sigma^{4/5}/\sigma)^2) \geq 1 - \frac{1}{\sigma^{2/5}}$ .

Let  $X$  be the number of ONE-CHOICE allocations in the first  $m$  steps. Using the following standard Chernoff bound,

$$\Pr[X \geq (1 - \delta) \cdot \mathbb{E}[X]] \geq 1 - e^{-\frac{1}{2}\delta^2 \cdot \mathbb{E}[X]},$$

with  $\delta = \frac{1}{\sigma^{2/5}} \leq 1$  (since  $\sigma \geq 1$ ) and  $\mathbb{E}[X] \geq (1 - \frac{1}{\sigma^{2/5}}) \cdot m$ , we get that

$$\Pr\left[X \geq \frac{1}{2} \left(1 - 2 \cdot \frac{1}{\sigma^{2/5}}\right) \cdot \sigma^{4/5} \cdot n\right] \geq 1 - e^{-\frac{1}{2}\delta^2 \cdot \mathbb{E}[X]} \geq 1 - n^{-\omega(1)}.$$

Therefore, using Lemma A.9 (for  $c := \frac{1}{2} \cdot (1 - 2 \cdot \frac{1}{\sigma^{2/5}}) \cdot \sigma^{4/5} \cdot \frac{1}{\log n} \geq \frac{1}{\log n}$ ), we get that

$$\begin{aligned} \text{Gap}(m) &\geq \frac{1}{10} \cdot \sqrt{\frac{1}{2} \left(1 - \frac{2}{\sigma^{2/5}}\right) \cdot \sigma^{4/5} \cdot \log n} + \frac{1}{2} \left(1 - \frac{1}{\sigma^{2/5}}\right) \cdot \sigma^{4/5} - \frac{m}{n} \\ &= \frac{1}{10} \cdot \sqrt{\frac{1}{2} \left(1 - \frac{2}{\sigma^{2/5}}\right) \cdot \sigma^{4/5} \cdot \log n} + \frac{1}{2} \left(1 - \frac{1}{\sigma^{2/5}}\right) \cdot \sigma^{4/5} - \frac{1}{2}\sigma^{4/5} \end{aligned}$$

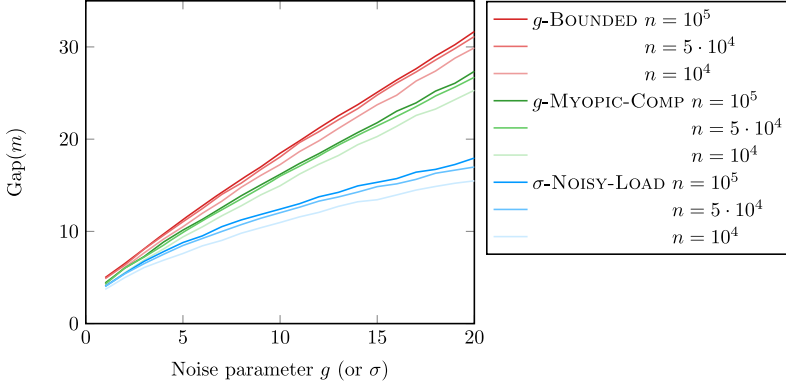


Fig. 10. Average gap for the  $g$ -BOUNDED and  $g$ -MYOPIC-COMP with  $g \in \{1, \dots, 20\}$  and  $\sigma$ -NOISY-LOAD with  $\sigma \in \{1, \dots, 20\}$ , for  $n \in \{10^4, 5 \cdot 10^4, 10^5\}$  and  $m = 1000 \cdot n$ , over 100 runs.

$$\begin{aligned}
 &= \frac{1}{10} \cdot \sqrt{\frac{1}{2} \left(1 - \frac{2}{\sigma^{2/5}}\right) \cdot \sigma^{4/5} \cdot \log n} - \frac{1}{2} \sigma^{2/5} \\
 &\stackrel{(a)}{\geq} \frac{1}{20} \sigma^{2/5} \cdot \sqrt{\log n} - \frac{1}{2} \cdot \sigma^{2/5} \\
 &\geq \frac{1}{30} \sigma^{2/5} \cdot \sqrt{\log n},
 \end{aligned}$$

using in (a) that  $\sigma \geq 32$ . So, combining the two cases we get the claim.  $\square$

Finally, we will prove a simple lower bound for  $b$ -BATCH, which already holds in the first batch of  $b$  balls, by a ONE-CHOICE argument. This upper bound is tight by Corollary 10.4.

**OBSERVATION 11.6.** *There exists a constant  $\kappa > 0$ , such that for the  $b$ -BATCH process for any  $b \in [n \cdot e^{\log^c n}, n \log n]$  with  $c > 0$  any constant, it holds that*

$$\Pr \left[ \text{Gap}(b) \geq \kappa \cdot \frac{\log n}{\log \left( \frac{4n}{b} \cdot \log n \right)} \right] \geq 1 - n^{-1}.$$

**PROOF.** During the allocation of the first batch consisting of  $b$  balls, the load information of all bins is never updated, i.e., their load “estimate” equals zero. Hence in these first  $n$  steps, as ties are broken randomly,  $b$ -BATCH behaves exactly like ONE-CHOICE. Hence, by Lemma A.10 there exists a constant  $\kappa > 0$ ,

$$\Pr \left[ \text{Gap}(b) \geq \kappa \cdot \frac{\log n}{\log \left( \frac{4n}{b} \cdot \log n \right)} \right] \geq 1 - n^{-1}. \quad \square$$

## 12 EXPERIMENTAL RESULTS

In this section, we empirically analyze the gap for the  $g$ -BOUNDED,  $g$ -MYOPIC-COMP and  $\sigma$ -NOISY-LOAD processes for various values of  $g$ ,  $\sigma$ ,  $n$  and  $m$  (Figure 10 and Table 3). For sufficiently large  $g$ , the gap plots for  $g$ -BOUNDED and  $g$ -MYOPIC-COMP are almost linear, as suggested by our upper bound of  $\Theta(g)$  for any  $g = \log^{\Omega(1)} n$ . We also present the empirical gap for  $b$ -BATCH for various values of batch size  $b$  around  $n$  and compare with the ONE-CHOICE gap for the case of  $b$  balls (Figure 11 and Table 4). As our analysis in Section 10.2 suggests, these gaps are close when  $b$  is close to  $n$  and for smaller  $b$ , e.g., when  $b = n^{1-\Omega(1)}$ , the gaps of the two processes diverge (Remark 10.6).

Table 3. Empirical Gap Distribution for  $g$ -BOUNDED,  $g$ -MYOPIC-COMP and  $\sigma$ -NOISY-LOAD with  $g, \sigma \in \{0, 1, 2, 4, 8, 16\}$ , for  $n \in \{10^4, 5 \cdot 10^4, 10^5\}$  and  $m = 1000 \cdot n$ , Over 100 Runs

$g$ -BOUNDED						
$n$	$g = 0$	$g = 1$	$g = 2$	$g = 4$	$g = 8$	$g = 16$
$10^4$	2 : 46% 3 : 54%	4 : 74% 5 : 26%	5 : 1% 6 : 89% 7 : 10%	8 : 1% 9 : 82% 10 : 17%	13 : 1% 14 : 35% 15 : 51% 16 : 11% 17 : 2%	23 : 4% 24 : 37% 25 : 43% 26 : 11% 27 : 5%
$5 \cdot 10^4$	2 : 4% 3 : 96%	4 : 13% 5 : 86% 6 : 1%	6 : 67% 7 : 33%	9 : 46% 10 : 51% 11 : 3%	14 : 3% 15 : 72% 16 : 24% 17 : 1%	25 : 25% 26 : 47% 27 : 23% 28 : 4% 29 : 1%
$10^5$	3 : 100%	4 : 1% 5 : 99%	6 : 50% 7 : 50%	9 : 32% 10 : 67% 11 : 1%	15 : 39% 16 : 57% 17 : 4%	25 : 9% 26 : 50% 27 : 33% 28 : 7% 29 : 1%

$g$ -MYOPIC-COMP						
$n$	$g = 0$	$g = 1$	$g = 2$	$g = 4$	$g = 8$	$g = 16$
$10^4$	2 : 46% 3 : 54%	4 : 97% 5 : 3%	5 : 49% 6 : 51%	7 : 2% 8 : 87% 9 : 11%	12 : 37% 13 : 50% 14 : 12% 15 : 1%	20 : 14% 21 : 47% 22 : 29% 23 : 8% 25 : 2%
$5 \cdot 10^4$	2 : 4% 3 : 96%	4 : 73% 5 : 27%	5 : 1% 6 : 97% 7 : 2%	8 : 50% 9 : 50%	12 : 1% 13 : 50% 14 : 44% 15 : 5%	21 : 10% 22 : 44% 23 : 39% 24 : 6% 26 : 1%
$10^5$	3 : 100%	4 : 59% 5 : 41%	6 : 99% 7 : 1%	8 : 19% 9 : 78% 10 : 3%	13 : 21% 14 : 72% 15 : 7%	22 : 24% 23 : 51% 24 : 24% 26 : 1%

$\sigma$ -NOISY-LOAD						
$n$	$\sigma = 0$	$\sigma = 1$	$\sigma = 2$	$\sigma = 4$	$\sigma = 8$	$\sigma = 16$
$10^4$	2 : 46% 3 : 54%	3 : 29% 4 : 71%	4 : 9% 5 : 84% 6 : 7%	6 : 20% 7 : 73% 8 : 7%	9 : 36% 10 : 50% 11 : 12% 12 : 2%	12 : 2% 13 : 33% 14 : 42% 15 : 16% 16 : 6% 18 : 1%
$5 \cdot 10^4$	2 : 4% 3 : 96%	4 : 98% 5 : 2%	5 : 61% 6 : 39%	7 : 51% 8 : 48% 10 : 1%	9 : 1% 10 : 37% 11 : 52% 12 : 8% 13 : 2%	14 : 24% 15 : 45% 16 : 24% 17 : 6% 18 : 1%
$10^5$	3 : 100%	4 : 95% 5 : 5%	5 : 58% 6 : 41% 7 : 1%	7 : 26% 8 : 69% 9 : 4% 10 : 1%	10 : 13% 11 : 56% 12 : 26% 13 : 4% 14 : 1%	14 : 1% 15 : 49% 16 : 35% 17 : 8% 18 : 6% 19 : 1%

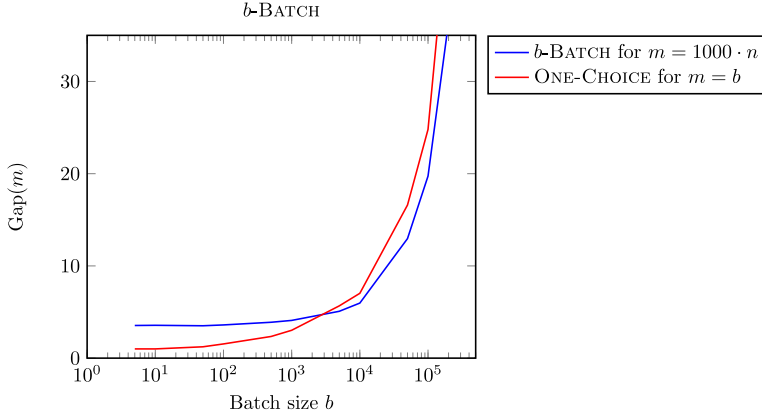


Fig. 11. Average gap for the  $b$ -BATCH process with  $b \in \{5, 10, 5 \cdot 10, 10^2, \dots, 10^5, 5 \cdot 10^5\}$ , for  $n = 10^4$  and  $m = 1000 \cdot n$  and for ONE-CHOICE for  $b$  balls, over 100 runs.

Table 4. Empirical Gap Distribution for  $b$ -BATCH with  $b \in \{5, 10, 5 \cdot 10, 10^2, \dots, 10^5, 5 \cdot 10^5\}$ , for  $n = 10^4$  and  $m = 1000 \cdot n$  and for ONE-CHOICE with  $m = b$ , Over 100 Runs

<b><math>b</math>-BATCH/ONE-CHOICE</b>					
Process	$b = 10$	$b = 10^2$	$b = 10^3$	$b = 10^4$	$b = 10^5$
$b$ -BATCH $m = 1000 \cdot n$	<b>3 : 44%</b> <b>4 : 56%</b>	<b>3 : 40%</b> <b>4 : 60%</b>	<b>4 : 91%</b> <b>5 : 9%</b>	<b>5 : 29%</b> <b>6 : 49%</b> <b>7 : 18%</b> <b>8 : 4%</b>	<b>16 : 1%</b>
					<b>17 : 8%</b>
					<b>18 : 15%</b>
					<b>19 : 28%</b>
					<b>20 : 18%</b>
					<b>21 : 12%</b>
					<b>22 : 14%</b>
					<b>24 : 1%</b>
					<b>25 : 2%</b>
					<b>26 : 1%</b>
ONE-CHOICE $m = b$ balls	<b>1 : 100%</b>	<b>1 : 47%</b> <b>2 : 52%</b> <b>3 : 1%</b>	<b>2 : 5%</b> <b>3 : 88%</b> <b>4 : 7%</b>	<b>6 : 22%</b> <b>7 : 56%</b> <b>8 : 19%</b> <b>9 : 3%</b>	<b>21 : 2%</b>
					<b>22 : 12%</b>
					<b>23 : 13%</b>
					<b>24 : 21%</b>
					<b>25 : 18%</b>
					<b>26 : 17%</b>
					<b>27 : 4%</b>
					<b>28 : 8%</b>
					<b>29 : 4%</b>
					<b>31 : 1%</b>

### 13 CONCLUSIONS

In this work, we proposed a number of different noise settings, and analyzed the gap of the Two-CHOICE process. As our main result, we established for all  $g \geq 1$  the asymptotic value of  $\text{Gap}(\cdot)$  in the adversarial setting  $g$ -ADV-COMP. This shows that  $\text{Gap}(\cdot)$  undergoes a delicate phase transition

at the point  $g = \log^{\Theta(1)} n$ , around which the impact of  $g$  on the gap switches from superlinear to linear. Further, we presented tight bounds for the delay and batched settings for a range of batch sizes  $b$  and delay parameters  $\tau$ , where load information of the bins may be stale. Finally, we also proposed a smoothed probabilistic setting where each load comparison between two bins  $i_1$  and  $i_2$  is correct with some probability  $\rho(\delta)$ , where  $\delta = |x_{i_1}^t - x_{i_2}^t|$  is their load difference. Assuming that the load estimates of bins are independent Gaussian perturbations leads to  $\rho(\delta) = 1 - \frac{1}{2} \exp(-(\delta/\sigma)^2)$ . For this choice of  $\rho$ , we proved that the gap is some function that is polynomial in  $\sigma$  and logarithmic in  $n$ .

There are several directions for future work. One interesting avenue is establishing tight(er) bounds for  $\sigma$ -NOISY-LOAD or other choices of  $\rho$ , as well as determining tight(er) bounds for the remaining values of  $b$  and  $\tau$  in the delay settings. Also, it would be interesting to investigate the noisy setting for other balanced allocations processes, such as the MEAN-THINNING or the  $(1 + \beta)$ -process. Lastly, one could consider extending the noisy settings to graphical allocation [58] or processes where balls get removed. We think our techniques might apply if removals and insertions are both random and uncorrelated, e.g., like in [9]. However, for adversarial removals as in [15, 23], this appears to be more challenging.

## APPENDICES

### A ANALYSIS TOOLS

#### A.1 Concentration Inequalities

We state the following well-known concentration inequalities.

LEMMA A.1 (AZUMA'S INEQUALITY FOR SUPER-MARTINGALES [31, PROBLEM 6.5]). *Consider a super-martingale  $X^0, \dots, X^N$  satisfying  $|X^i - X^{i-1}| \leq a_i$  for any  $i \in [N]$ , then for any  $\lambda > 0$ ,*

$$\Pr [X^N \geq X^0 + \lambda] \leq \exp\left(-\frac{\lambda^2}{2 \cdot \sum_{i=1}^N a_i^2}\right).$$

LEMMA A.2 ([22, THEOREMS 6.1 & 6.5]). *Consider a martingale  $X^0, \dots, X^N$  with filtration  $\mathcal{F}^0, \dots, \mathcal{F}^N$  satisfying  $|X^i - X^{i-1}| \leq M$  and  $\text{Var}[X^i | \mathcal{F}^{i-1}] \leq \sigma_i^2$  for any  $i \in [N]$ , then for any  $\lambda > 0$ ,*

$$\Pr [ |X^N - \mathbb{E}[X^N]| \geq \lambda ] \leq 2 \cdot \exp\left(-\frac{\lambda^2}{2 \cdot (\sum_{i=1}^N \sigma_i^2 + M\lambda/3)}\right).$$

#### A.2 Facts about ONE-CHOICE

In this part, we collect several (mostly folklore) results for the maximum load and the gap of ONE-CHOICE. In contrast to e.g., [59], we are aiming at obtaining bounds on the gap with probability at least  $1 - n^{-6}$  and not just with probability at least  $1 - o(1)$ , but which are not necessarily tight up to lower order terms. Furthermore, the bounds in [59] hold for any  $m = \Omega(\frac{n}{\text{polylog}(n)})$ , but we wish to also cover  $m = n \cdot \exp(-\log^{\Theta(1)} n)$ . In Section 10, we use these bounds to analyze the  $b$ -BATCH and  $\tau$ -DELAY settings for various values of  $b$  and  $\tau$ .

We first restate the so-called Poisson Approximation method.

LEMMA A.3 ([53, COROLLARY 5.11]). *Let  $(x^T)_{i \in [n]}$  be the load vector after  $T$  steps of ONE-CHOICE. Further, let  $(\tilde{x}^T)_{i \in [n]}$  be  $n$  independent Poisson random variables with parameter  $\lambda = T/n$  each. Further, let  $\mathcal{E}$  be any event which is determined by  $x^T$ , and further assume that  $\Pr[\mathcal{E}]$  is either monotonically increasing in  $T$  or monotonically decreasing in  $T$ . Further, let  $\tilde{\mathcal{E}}$  be the corresponding event determined by  $\tilde{x}^T$ . Then,*

$$\Pr[\mathcal{E}] \leq 2 \cdot \Pr[\tilde{\mathcal{E}}].$$

LEMMA A.4. Consider the ONE-CHOICE process, for any  $\alpha > 0$ , any bin  $i \in [n]$ , and any step  $m \geq 0$ ,

$$\mathbb{E}[e^{\alpha x_i^m}] \leq e^{\frac{m}{n} \cdot (e^\alpha - 1)}.$$

PROOF. We will proceed inductively to show that  $\mathbb{E}[e^{\alpha x_i^m}] \leq e^{\frac{m}{n} \cdot (e^\alpha - 1)}$ . The base case follows since  $\mathbb{E}[e^{\alpha x_i^0}] = 1 \leq 1$ . For  $m \geq 1$ , let  $Z_i^m$  indicate whether the  $m$ -th ball was allocated to bin  $i \in [n]$  and assume that  $\mathbb{E}[e^{\alpha x_i^{m-1}}] \leq e^{\frac{m-1}{n} \cdot (e^\alpha - 1)}$  holds, then

$$\begin{aligned} \mathbb{E}[e^{\alpha x_i^m}] &= \mathbb{E}[e^{\alpha(Z_i^m + x_i^{m-1})}] = \mathbb{E}[e^{\alpha Z_i^m}] \cdot \mathbb{E}[e^{\alpha x_i^{m-1}}] \\ &= \left(\frac{1}{n} \cdot e^\alpha + \left(1 - \frac{1}{n}\right) \cdot e^0\right) \cdot \mathbb{E}[e^{\alpha x_i^{m-1}}] = \left(1 + \frac{1}{n} \cdot (e^\alpha - 1)\right) \cdot \mathbb{E}[e^{\alpha x_i^{m-1}}] \\ &\stackrel{(a)}{\leq} e^{\frac{1}{n} \cdot (e^\alpha - 1)} \cdot \mathbb{E}[e^{\alpha x_i^{m-1}}] \\ &\leq e^{\frac{1}{n} \cdot (e^\alpha - 1)} \cdot e^{\frac{m-1}{n} \cdot (e^\alpha - 1)} = e^{\frac{m}{n} \cdot (e^\alpha - 1)}, \end{aligned}$$

using in (a) that  $1 + u \leq e^u$  (for any  $u$ ).  $\square$

We now proceed to obtain an upper bound for the maximum load of ONE-CHOICE for any  $m \leq 2n \log n$ . As we will show in Lemma A.8, this bound is asymptotically tight.

LEMMA A.5 (CF. [1, LEMMA 14]). Consider the ONE-CHOICE process for any  $m \leq 2n \log n$ . Then,

$$\Pr \left[ \max_{i \in [n]} x_i^m \leq 11 \cdot \frac{\log n}{\log(\frac{4n}{m} \cdot \log n)} \right] \geq 1 - n^{-6}.$$

PROOF. Using Claim A.4, for the given  $m$  and  $\alpha = \log(\frac{4n}{m} \cdot \log n) > 0$ , we have that for any bin  $i \in [n]$ ,

$$\mathbb{E}[e^{\alpha x_i^m}] \leq e^{\frac{m}{n} \cdot e^{\log(\frac{4n}{m} \cdot \log n)}} = n^4.$$

Hence, by Markov's inequality,

$$\Pr[e^{\alpha x_i^m} \leq n^{11}] \geq 1 - n^{-7}.$$

When this event holds, we have

$$x_i^m \leq \frac{1}{\alpha} \cdot \log(n^{11}) \leq 11 \cdot \frac{\log n}{\log(\frac{4n}{m} \cdot \log n)}.$$

By taking the union bound over all bins  $i \in [n]$ , we get the claim.  $\square$

For  $m = O(n)$ , this recovers the well-known  $O(\frac{\log n}{\log \log n})$  bound.

COROLLARY A.6 (CF. [53, CHAPTER 5]). Consider the ONE-CHOICE process for  $m = 2n$ . Then,

$$\Pr \left[ \max_{i \in [n]} x_i^m \leq 11 \cdot \frac{\log n}{\log \log n} \right] \geq 1 - n^{-6}.$$

For  $m = O(n^{1-\epsilon})$  for some constant  $\epsilon \in (0, 1)$ , this shows that  $\text{Gap}(m) = O(1)$ .

COROLLARY A.7. Consider the ONE-CHOICE process with  $m = 2n^{1-\epsilon}$  for any constant  $\epsilon \in (0, 1)$ . Then,

$$\Pr \left[ \max_{i \in [n]} x_i^m \leq \frac{11}{\epsilon} \right] \geq 1 - n^{-6}.$$

In the following lemma, we prove that the ONE-CHOICE bound obtained in Lemma A.5 is asymptotically tight.

LEMMA A.8 (CF. LEMMA 14 IN [1]). *Consider the ONE-CHOICE process with  $m \leq n \log n$ . Then, there exists a constant  $\kappa > 0$ , such that*

$$\Pr \left[ \max_{i \in [n]} x_i^m \geq \frac{1}{4} \cdot \frac{\log n}{\log(\frac{4n}{m} \cdot \log n)} \right] \geq 1 - n^{-1}.$$

PROOF. We will bound the probability of event  $\mathcal{E}$ , that the maximum load is less than  $M = \frac{1}{4} \cdot \frac{\log n}{\log(\frac{4n}{m} \cdot \log n)}$ . Clearly, the probability of the event  $\mathcal{E}$  is monotonically increasing in the number of balls (while keeping  $M$  fixed).

Following Lemma A.3, it suffices to bound the probability of the event  $\tilde{\mathcal{E}}$  which is that the maximum value of  $n$  independent Poisson random variables with parameter  $\lambda = \frac{m}{n}$  is less than  $M$ . We want to show that

$$\Pr [\tilde{\mathcal{E}}] \leq \left( 1 - \frac{e^{-\frac{m}{n}} \left(\frac{m}{n}\right)^M}{M!} \right)^n \leq \exp \left( -n \cdot \frac{e^{-m/n} \left(\frac{m}{n}\right)^M}{M!} \right) < n^{-1}.$$

This is equivalent to showing that

$$\begin{aligned} -n \cdot \frac{e^{-m/n} \left(\frac{m}{n}\right)^M}{M!} < -\log n &\iff \log n - \frac{m}{n} + M \cdot \log \left(\frac{m}{n}\right) - \log M! > \log \log n \\ &\iff \log n + M \cdot \log \left(\frac{m}{n}\right) > \frac{m}{n} + \log M! + \log \log n. \end{aligned}$$

Using  $\log M! \leq M \cdot (\log M - 1) + \log M$  (e.g., in [34, Equation 9.1]), we deduce that

$$\begin{aligned} &\frac{m}{n} + M \cdot \left( \log M - 1 - \log \left(\frac{m}{n}\right) \right) + \log M + \log \log n \\ &< \frac{3}{4} \cdot \log n + \frac{\log n}{4 \cdot \log(\frac{4n \log n}{m})} \cdot \left( \log \log n - \log \left( 4 \cdot \log \left(\frac{4n \log n}{m}\right) \right) - \log \left(\frac{m}{n}\right) \right) \\ &= \frac{3}{4} \cdot \log n + \frac{\log n}{4 \cdot \log(\frac{4n \log n}{m})} \cdot \left( \log \left(\frac{4n \log n}{m}\right) - \log \left( 16 \cdot \log \left(\frac{4n \log n}{m}\right) \right) \right) \\ &\leq \frac{3}{4} \cdot \log n + \frac{\log n}{4 \cdot \log(\frac{4n \log n}{m})} \cdot \log \left(\frac{4n \log n}{m}\right) \\ &= \frac{3}{4} \cdot \log n + \frac{1}{4} \cdot \log n = \log n, \end{aligned}$$

for sufficiently large  $n$  and using that  $\log(\frac{4n \log n}{m}) > 0$  as  $m \leq n \log n$ . Hence, we get the desired lower bound.  $\square$

The next standard result was also used in [58, Section 4] and is based on [59].

LEMMA A.9 (CF. [44, LEMMA A.2]). *Consider the ONE-CHOICE process for  $m = cn \log n$  where  $c \geq 1/\log n$ . Then,*

$$\Pr \left[ \text{Gap}(m) \geq \frac{\sqrt{c}}{10} \cdot \log n \right] \geq 1 - n^{-2}.$$

Combining the previous two lemmas, we also get the asymptotically tight bound on the gap

LEMMA A.10 (CF. [1, LEMMA 14]). *Consider the ONE-CHOICE process with  $m \leq n \log n$ . Then, there exists a constant  $\kappa > 0$ , such that*

$$\Pr \left[ \text{Gap}(m) \geq \kappa \cdot \frac{\log n}{\log(\frac{4n}{m} \cdot \log n)} \right] \geq 1 - n^{-1}.$$

PROOF. For sufficiently small constant  $C \in (0, 1)$ , for any  $m \leq Cn \log n$  we have that

$$\frac{1}{4} \cdot \frac{\log n}{\log(\frac{4n}{m} \cdot \log n)} \geq 2 \cdot \frac{m}{n},$$

and hence the conclusion follows by Lemma A.8 for  $\kappa = \frac{1}{8}$ . For  $m > Cn \log n$ , the stated bound follows from Lemma A.9.  $\square$

### A.3 Auxiliary Probabilistic Claims

For convenience, we state and prove the following well-known result.

LEMMA A.11. *Consider any sequence of random variables  $(X^i)_{i \in \mathbb{N}}$  for which there exist  $0 < a < 1$  and  $b > 0$ , such that every  $i \geq 1$ ,*

$$\mathbb{E} [X^i \mid X^{i-1}] \leq X^{i-1} \cdot a + b.$$

*Then, (i) for every  $i \geq 0$ ,*

$$\mathbb{E} [X^i \mid X^0] \leq X^0 \cdot a^i + \frac{b}{1-a}.$$

*Further, (ii) if  $X^0 \leq \frac{b}{1-a}$  holds, then for every  $i \geq 0$ ,*

$$\mathbb{E} [X^i] \leq \frac{b}{1-a}.$$

PROOF. *First statement.* We will prove by induction that for every  $i \in \mathbb{N}$ ,

$$\mathbb{E} [X^i \mid X^0] \leq X^0 \cdot a^i + b \cdot \sum_{j=0}^{i-1} a^j.$$

For  $i = 0$ ,  $\mathbb{E} [X^0 \mid X^0] \leq X^0$ . Assuming the induction hypothesis holds for some  $i \geq 0$ , then since  $a > 0$ ,

$$\begin{aligned} \mathbb{E} [X^{i+1} \mid X^0] &= \mathbb{E} [\mathbb{E} [X^{i+1} \mid X^i] \mid X^0] \leq \mathbb{E} [X^i \mid X^0] \cdot a + b \\ &\leq \left( X^0 \cdot a^i + b \cdot \sum_{j=0}^{i-1} a^j \right) \cdot a + b \\ &= X^0 \cdot a^{i+1} + b \cdot \sum_{j=0}^i a^j. \end{aligned}$$

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

*Second statement.* We will prove this claim by induction. Then, assuming that  $\mathbb{E} [X^i] \leq \frac{b}{1-a}$  holds for  $i \geq 0$ , we have for  $i + 1$

$$\mathbb{E} [X^{i+1}] = \mathbb{E} [\mathbb{E} [X^{i+1} \mid X^i]] \leq \mathbb{E} [X^i] \cdot a + b \leq \frac{b}{1-a} \cdot a + b = \frac{b}{1-a}. \quad \square$$



## A.4 Majorization

We begin with the following well-known majorization inequality which appears in Theorem 1 [48, Chapter XII]. For a proof see e.g., [46, Lemma A.2].

LEMMA A.12 (CF. [46, LEMMA A.2]). *Let  $(a_k)_{k=1}^n, (b_k)_{k=1}^n$  be non-negative and  $(c_k)_{k=1}^n$  be non-negative and non-increasing. If  $\sum_{k=1}^i a_k \leq \sum_{k=1}^i b_k$  holds for all  $1 \leq i \leq n$  then,*

$$\sum_{k=1}^n a_k \cdot c_k \leq \sum_{k=1}^n b_k \cdot c_k.$$

Next we state an auxiliary result, which is implicit in [58]. Theorem 3.1 in [58] proves the required majorization for time-independent probability allocation vectors, but as mentioned in the proof of Theorem 3.2 in [58], the same result generalizes to time-dependent probability allocation vectors.

LEMMA A.13 (CF. [58, SECTION 3]). *Consider two allocation processes  $\mathcal{Q}$  and  $\mathcal{P}$ . The allocation process  $\mathcal{Q}$  uses at each step a fixed probability allocation vector  $q$ . The allocation process  $\mathcal{P}$  uses a time-dependent probability allocation vector  $p^t$ , which may depend on  $\mathfrak{F}^t$  but majorizes  $q$  at each step  $t \geq 0$ . Let  $y^t(\mathcal{Q})$  and  $y^t(\mathcal{P})$  be the two normalized load vectors, sorted non-increasingly. Then there is a coupling such that for all steps  $t \geq 0$ ,  $y^t(\mathcal{P})$  majorizes  $y^t(\mathcal{Q})$ , so in particular,  $y_1^t(\mathcal{P}) \geq y_1^t(\mathcal{Q})$ .*

## B OMITTED PROOFS AND AUXILIARY CLAIMS

### B.1 Omitted Proofs from Section 4

COROLLARY 4.5 (RESTATED, PAGE 14). *Consider the  $g$ -ADV-COMP setting for any  $g \geq 1$ . Then, for any steps  $t_0 \geq 0$  and  $t_1 \geq t_0$ , we have that*

$$\Pr \left[ \max_{i \in [n]} |y_i^{t_1}| \leq 2g(\log(ng))^2 \mid \mathfrak{F}^{t_0}, \max_{i \in [n]} |y_i^{t_0}| \leq g(\log(ng))^2 \right] \geq 1 - (ng)^{14}.$$

PROOF. We will be using the hyperbolic cosine potential  $\Gamma := \Gamma(\gamma)$  with smoothing parameter  $\gamma := -\log(1 - \frac{1}{8 \cdot 48})/g$  as we did in the proof of Theorem 4.3.

Consider an arbitrary step  $t_0$  with  $\max_{i \in [n]} |y_i^{t_0}| \leq g(\log(ng))^2$ . Then, it follows that

$$\Gamma^{t_0} \leq 2n \cdot e^{\gamma g(\log(ng))^2}.$$

By Theorem 4.3 (i), there exists a constant  $c_1 \geq 1$  such that

$$\mathbb{E} \left[ \Gamma^{t+1} \mid \mathfrak{F}^t \right] \leq \left( 1 - \frac{\gamma}{96n} \right) + c_1,$$

and using Lemma A.11 (i) (for  $a := 1 - \frac{\gamma}{96n}$  and  $b := c_1$ ) at step  $t_1 \geq t_0$ , we have that

$$\mathbb{E} \left[ \Gamma^{t_1} \mid \mathfrak{F}^{t_0}, \Gamma^{t_0} \leq 2n \cdot e^{\gamma g(\log(ng))^2} \right] \leq 2n \cdot e^{\gamma g(\log(ng))^2} \cdot \left( 1 - \frac{\gamma}{96n} \right)^{t_1 - t_0} + \frac{96c_1 n}{\gamma} \leq 4n \cdot e^{\gamma g(\log(ng))^2},$$

recalling that  $\gamma = \Theta(\frac{1}{g})$ . Hence, by Markov's inequality, we have that

$$\begin{aligned} \Pr \left[ \Gamma^{t_1} \leq 4n \cdot e^{\gamma g(\log(ng))^2} \cdot (ng)^{14} \mid \mathfrak{F}^{t_0}, \max_{i \in [n]} |y_i^{t_0}| \leq g(\log(ng))^2 \right] \\ \geq \Pr \left[ \Gamma^{t_1} \leq 4n \cdot e^{\gamma g(\log(ng))^2} \cdot (ng)^{14} \mid \mathfrak{F}^{t_0}, \Gamma^{t_0} \leq 2n \cdot e^{\gamma g(\log(ng))^2} \right] \geq 1 - (ng)^{-14}. \end{aligned}$$

When the event  $\{\Gamma^{t_1} \leq 4n \cdot e^{\gamma g(\log(ng))^2} \cdot (ng)^{14}\}$  holds, then it also follows that

$$\max_{i \in [n]} |y_i^{t_1}| \leq \frac{\log \Gamma^{t_1}}{\gamma} \leq g(\log(ng))^2 + \frac{1}{\gamma} \cdot (\log(4n) + 14 \log(ng)) \leq 2g(\log(ng))^2,$$

for sufficiently large  $n$  and using that  $\gamma = \Theta(\frac{1}{g})$ . Hence, we get the conclusion.  $\square$

## B.2 Omitted Proofs from Section 5

LEMMA 5.1 (RESTATED, PAGE 15). *Consider any allocation process with probability allocation vector  $r^t$  at step  $t \geq 0$ . Then, for any step  $t \geq 0$ , (i) it holds that*

$$\mathbb{E} \left[ \Delta Y^{t+1} \mid y^t \right] = \sum_{i=1}^n 2 \cdot r_i^t \cdot y_i^t + 1 - \frac{1}{n} \leq \sum_{i=1}^n 2 \cdot r_i^t \cdot y_i^t + 1,$$

and (ii) it holds that

$$|\Delta Y^{t+1}| \leq 4 \cdot \max_{i \in [n]} |y_i^t| + 2.$$

PROOF. *First statement.* For any bin  $i \in [n]$ , its expected contribution to  $Y^{t+1}$  is given by,

$$\begin{aligned} \mathbb{E} \left[ Y_i^{t+1} \mid y^t \right] &= \left( y_i^t + 1 - \frac{1}{n} \right)^2 \cdot r_i^t + \left( y_i^t - \frac{1}{n} \right)^2 \cdot (1 - r_i^t) \\ &= (y_i^t)^2 + 2 \cdot \left( 1 - \frac{1}{n} \right) \cdot y_i^t \cdot r_i^t - 2 \cdot \frac{1}{n} \cdot y_i^t \cdot (1 - r_i^t) + \left( 1 - \frac{1}{n} \right)^2 \cdot r_i^t + \frac{1}{n^2} \cdot (1 - r_i^t) \\ &= (y_i^t)^2 + 2 \cdot \left( r_i^t - \frac{1}{n} \right) \cdot y_i^t + \left( 1 - \frac{1}{n} \right)^2 \cdot r_i^t + \frac{1}{n^2} \cdot (1 - r_i^t). \end{aligned}$$

Hence, by aggregating over all bins we get,

$$\begin{aligned} \mathbb{E} \left[ Y^{t+1} \mid y^t \right] &= \sum_{i=1}^n \left[ (y_i^t)^2 + 2 \cdot \left( r_i^t - \frac{1}{n} \right) \cdot y_i^t + \left( 1 - \frac{1}{n} \right)^2 \cdot r_i^t + \frac{1}{n^2} \cdot (1 - r_i^t) \right] \\ &= Y^t + \sum_{i=1}^n 2 \cdot \left( r_i^t - \frac{1}{n} \right) \cdot y_i^t + \left( 1 - \frac{1}{n} \right)^2 + \frac{1}{n} - \frac{1}{n^2} \\ &= Y^t + \sum_{i=1}^n 2 \cdot \left( r_i^t - \frac{1}{n} \right) \cdot y_i^t + 1 - \frac{1}{n} \\ &= Y^t + \sum_{i=1}^n 2 \cdot r_i^t \cdot y_i^t - \sum_{i=1}^n \frac{2}{n} \cdot y_i^t + 1 - \frac{1}{n} \\ &\stackrel{(a)}{=} Y^t + \sum_{i=1}^n 2 \cdot r_i^t \cdot y_i^t + 1 - \frac{1}{n} \\ &\leq Y^t + \sum_{i=1}^n 2 \cdot r_i^t \cdot y_i^t + 1, \end{aligned}$$

using in (a) that  $\sum_{i=1}^n y_i^t = 0$ . Therefore, by subtracting  $Y^t$ , statement (i) follows.

*Second statement.* Let  $M := \max_{i \in [n]} |y_i^t|$ . We will upper bound the change  $\Delta Y_i^{t+1}$  for an arbitrary bin  $i \in [n]$ , by considering the following two cases:

**Case 1:** Ball at step  $t + 1$  is allocated to bin  $i$ . So,

$$|\Delta Y_i^{t+1}| = \left| \left( y_i^t + 1 - \frac{1}{n} \right)^2 - (y_i^t)^2 \right| = \left| 2 \cdot \left( 1 - \frac{1}{n} \right) \cdot y_i^t + \left( 1 - \frac{1}{n} \right)^2 \right| \leq 2M + 1.$$

**Case 2:** Ball at step  $t + 1$  is not allocated to bin  $i$ . So,

$$|\Delta Y_i^{t+1}| = \left| \left( y_i^t - \frac{1}{n} \right)^2 - (y_i^t)^2 \right| = \left| -\frac{2}{n} \cdot y_i^t + \frac{1}{n^2} \right| \leq \frac{2M}{n} + \frac{1}{n}.$$

Aggregating over all bins  $i \in [n]$  yields

$$|\Delta Y^{t+1}| \leq \sum_{i=1}^n |\Delta Y_i^{t+1}| \leq 2M + 1 + (n-1) \cdot \left( \frac{2M}{n} + \frac{1}{n} \right) \leq 4M + 2. \quad \square$$

LEMMA 5.8 (RESTATED, PAGE 23). *Consider the  $g$ -ADV-COMP setting for any  $g \geq 1$  and the sequence  $(\tilde{\Lambda}_{t_0}^s)_{s \geq t_0} := (\tilde{\Lambda}_{t_0}^s)_{s \geq t_0}(\alpha, c_4 g, \varepsilon)$  for any starting step  $t_0 \geq 0$ , any  $\alpha \in (0, \frac{1}{18}]$  and  $\varepsilon, c_4 > 0$  as defined in Lemma 5.7. Then, for any step  $s \geq t_0$ ,*

$$\mathbb{E} \left[ \tilde{\Lambda}_{t_0}^{s+1} \mid \mathcal{F}^s \right] \leq \tilde{\Lambda}_{t_0}^s.$$

PROOF. Recalling the definition of  $\tilde{\Lambda}$  in Equation (5.6),

$$\begin{aligned} \mathbb{E}[\tilde{\Lambda}_{t_0}^{s+1} \mid \mathcal{F}^s] &= \mathbb{E}[\Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \mid \mathcal{F}^s] \cdot \exp\left(-\frac{3\alpha}{n} \cdot B_{t_0}^s\right) \cdot \exp\left(\frac{\alpha\varepsilon}{n} \cdot G_{t_0}^s\right) \\ &= \mathbb{E}[\Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \mid \mathcal{F}^s] \cdot \exp\left(\frac{\alpha\varepsilon}{n} \cdot \mathbf{1}_{\mathcal{G}^s} - \frac{3\alpha}{n} \cdot \mathbf{1}_{\neg\mathcal{G}^s}\right) \cdot \exp\left(-\frac{3\alpha}{n} \cdot B_{t_0}^{s-1}\right) \cdot \exp\left(\frac{\alpha\varepsilon}{n} \cdot G_{t_0}^{s-1}\right). \end{aligned}$$

Thus, to prove the statement, it suffices to show that

$$\mathbb{E}[\Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \mid \mathcal{F}^s] \cdot \exp\left(\frac{\alpha\varepsilon}{n} \cdot \mathbf{1}_{\mathcal{G}^s} - \frac{3\alpha}{n} \cdot \mathbf{1}_{\neg\mathcal{G}^s}\right) \leq \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}}. \quad (\text{B.1})$$

To show Equation (B.1), we consider two cases based on whether  $\mathcal{G}^s$  holds.

**Case 1 [ $\mathcal{G}^s$  holds]:** Recall that when  $\mathcal{G}^s$  holds, then  $\Lambda^s \leq Dng$ . Further, when  $\Lambda^s \leq cn$  holds (for  $c > 0$  the constant in Lemma 5.7), then  $\mathbf{1}_{\mathcal{E}_{t_0}^s} = 0$ . Thus, using Lemma 5.7 (ii),

$$\mathbb{E}[\Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \mid \mathcal{F}^s, \mathcal{G}^s] \leq \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \left(1 - \frac{\alpha\varepsilon}{n}\right) \leq \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \exp\left(-\frac{\alpha\varepsilon}{n}\right).$$

Hence, since in this case  $\mathbf{1}_{\mathcal{G}^s} = 1$ , the left hand side of Equation (B.1) is equal to

$$\mathbb{E}[\Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \mid \mathcal{F}^s, \mathcal{G}^s] \cdot \exp\left(\frac{\alpha\varepsilon}{n}\right) \leq \left(\Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \exp\left(-\frac{\alpha\varepsilon}{n}\right)\right) \cdot \exp\left(\frac{\alpha\varepsilon}{n}\right) = \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}}.$$

**Case 2 [ $\mathcal{G}^s$  does not hold]:** By Lemma 5.6 (ii), we get

$$\mathbb{E}[\Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \mid \mathcal{F}^s, \neg\mathcal{G}^s] \leq \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \left(1 + \frac{3\alpha}{n}\right) \leq \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \exp\left(\frac{3\alpha}{n}\right).$$

Hence, since in this case  $\mathbf{1}_{\mathcal{G}^s} = 0$ , the left hand side of Equation (B.1) is equal to

$$\mathbb{E}[\Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \mid \mathcal{F}^s, \neg\mathcal{G}^s] \cdot \exp\left(-\frac{3\alpha}{n}\right) \leq \left(\Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \exp\left(\frac{3\alpha}{n}\right)\right) \cdot \exp\left(-\frac{3\alpha}{n}\right) = \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}}.$$

Since Equation (B.1) holds in either case, we deduce that  $(\tilde{\Lambda}_{t_0}^s)_{s \geq t_0}$  forms a super-martingale.  $\square$

LEMMA 5.5 (RESTATED, PAGE 19). *Consider the potential  $\Lambda := \Lambda(\alpha, c_4 g)$  for any constant  $\alpha \in (0, 1)$ , any  $g \geq 1$  and any constant  $c_4 > 0$ . Then (i) for any constant  $\hat{c} > 0$ , there exists a constant  $c_s := c_s(\alpha, c_4, \hat{c}) \geq 1$ , such that for any step  $t \geq 0$  with  $\Lambda^t \leq \hat{c} \cdot n$ ,*

$$\Upsilon^t \leq c_s \cdot ng^2.$$

Furthermore, (ii) there exists a constant  $c_r := c_r(\alpha, c_4) \geq 1$ , such that for any step  $t \geq 0$ ,

$$\Upsilon^t \leq c_r \cdot n \cdot \left(g^2 + (\log \Lambda^t)^2\right).$$

PROOF. *First statement.* We begin by proving some basic inequalities between exponential, quadratic and linear terms. Let  $\hat{u} := (4/\alpha) \cdot \log(4/\alpha)$ . Note that  $e^u \geq u$  (for any  $u \geq 0$ ) and hence for any  $u \geq \hat{u}$ ,

$$e^{\alpha u/2} = e^{\alpha u/4} \cdot e^{\alpha u/4} \geq \frac{\alpha u}{4} \cdot e^{\alpha \hat{u}/4} = \frac{\alpha u}{4} \cdot \frac{4}{\alpha} = u,$$

and  $e^{\alpha u} = e^{\alpha u/2} \cdot e^{\alpha u/2} \geq u \cdot u = u^2$ . Therefore, for every  $u \geq 0$ ,

$$u^2 \leq \max \{ \hat{u}^2, e^{\alpha u} \}. \quad (\text{B.2})$$

Recall that for any bin  $i \in [n]$ ,  $\Lambda_i^t := e^{\alpha \cdot (y_i^t - c_4 g)^+} + e^{\alpha \cdot (-y_i^t - c_4 g)^+}$ . Hence,

$$\left( (y_i^t - c_4 g)^+ \right)^2 + \left( (-y_i^t - c_4 g)^+ \right)^2 \stackrel{(a)}{\leq} \max \{ 2\hat{u}^2, \Lambda_i^t \} \stackrel{(b)}{\leq} \max \{ 2\hat{u}^2 \cdot \Lambda_i^t, \Lambda_i^t \} \stackrel{(c)}{=} 2\hat{u}^2 \cdot \Lambda_i^t, \quad (\text{B.3})$$

where in (a) we used Equation (B.2) first with  $u = (y_i^t - c_4 g)^+$  and then with  $u = (-y_i^t - c_4 g)^+$ , in (b) that  $\Lambda_i^t \geq 1$  for any  $i \in [n]$  and in (c) that  $\hat{u} \geq 1$ , since  $\alpha \in (0, 1)$ .

We now proceed to upper bound the quadratic potential,

$$\begin{aligned} \Upsilon^t &\leq \sum_{i=1}^n \left[ \left( (y_i^t - c_4 g)^+ + c_4 g \right)^2 + \left( (-y_i^t - c_4 g)^+ + c_4 g \right)^2 \right] \\ &\stackrel{(a)}{\leq} 2 \cdot \sum_{i=1}^n \left[ \left( (y_i^t - c_4 g)^+ \right)^2 + \left( (-y_i^t - c_4 g)^+ \right)^2 + 2 \cdot (c_4 g)^2 \right] \\ &\stackrel{(\text{B.3})}{\leq} 4\hat{u}^2 \cdot \Lambda^t + 4c_4^2 \cdot ng^2 \\ &\stackrel{(b)}{\leq} 4\hat{c}\hat{u}^2 \cdot n + 4c_4^2 \cdot ng^2 \\ &\leq (4\hat{c}\hat{u}^2 + 4c_4^2) \cdot ng^2, \end{aligned}$$

using in (a) that  $(a+b)^2 \leq 2 \cdot (a^2 + b^2)$  (for any  $a, b$ ) and in (b) that  $\Lambda^t \leq \hat{c} \cdot n$ . Therefore, for the constant  $c_s := c_s(\alpha, c_4, \hat{c}) := 4\hat{c}\hat{u}^2 + 4c_4^2$ , we get the first statement.

*Second statement.* For any bin  $i \in [n]$  we have,

$$|y_i^t| \leq c_4 g + \frac{1}{\alpha} \log \Lambda^t.$$

Hence, using that  $(a+b)^2 \leq 2 \cdot (a^2 + b^2)$ ,

$$(y_i^t)^2 \leq \left( c_4 g + \frac{1}{\alpha} \log \Lambda^t \right)^2 \leq 2 \cdot \left( c_4^2 g^2 + \frac{1}{\alpha^2} \cdot (\log \Lambda^t)^2 \right) \leq c_r \cdot \left( g^2 + (\log \Lambda^t)^2 \right),$$

for some constant  $c_r := c_r(\alpha, c_4) = \max\{2c_4^2, \frac{2}{\alpha^2}\} \geq 1$ . By aggregating the contributions over all bins, we get the second statement.  $\square$

LEMMA 5.10 (RESTATEd, PAGE 25). *Consider the  $g$ -ADV-COMP setting for any  $g \geq 1$  and the potential  $\Lambda := \Lambda(\alpha, c_4 g)$  with  $\alpha = \frac{1}{18}$ , and  $c_4 > 0$  as defined in Lemma 5.7. Further, let the constants  $c, \varepsilon > 0$  be as defined in Lemma 5.7,  $c_s := c_s(\alpha, c_4, 2c) \geq 1$  as in Lemma 5.5 (i) and  $r \in (0, 1)$  as in Lemma 5.4. Then, for  $\Delta_s := \Delta_s(g) := \frac{60c_s}{\alpha \varepsilon r} \cdot n \cdot \max\{\log n, g\}$ , we have that for any step  $t_0 \geq 0$ ,*

$$\Pr \left[ \bigcup_{t \in [t_0, t_0 + \Delta_s]} \left\{ \Lambda^t \leq cn \right\} \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \leq 2cn \right] \geq 1 - (ng)^{-11}.$$

PROOF. By Lemma 5.5 (i),  $\Lambda^{t_0} \leq 2cn$  implies that deterministically  $\Upsilon^{t_0} \leq c_s ng^2$  for constant  $c_s := c_s(\alpha, c_4, 2c) \geq 1$  and

$$\max_{i \in [n]} |y_i^{t_0}| \leq c_4 g + \frac{1}{\alpha} \cdot \log(2cn) \leq g(\log(ng))^2,$$

for sufficiently large  $n$  using that  $c_4, \alpha, c > 0$  are constants. Let  $t_1 := t_0 + \Delta_s$ . By Lemma 5.4 with  $T := c_s ng \cdot \max\{\log n, g\} \geq c_s ng^2$  (and  $T = o(n^2 g^3)$ ) and  $\hat{c} := \frac{\Delta_s \cdot g}{T} = \frac{60}{\alpha \varepsilon r} \geq 1$  as  $\alpha, \varepsilon, r \leq 1$ , we have that

$$\begin{aligned} & \Pr \left[ G_{t_0}^{t_1-1} \geq r \cdot \Delta_s \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \leq 2cn \right] \\ & \geq \Pr \left[ G_{t_0}^{t_1-1} \geq r \cdot \Delta_s \mid \mathfrak{F}^{t_0}, \Upsilon^{t_0} \leq T, \max_{i \in [n]} |y_i^{t_0}| \leq g(\log(ng))^2 \right] \\ & \geq 1 - 2 \cdot (ng)^{-12}. \end{aligned} \quad (\text{B.4})$$

By Lemma 5.8,  $(\tilde{\Lambda}_{t_0}^t)_{t \geq t_0}$  is a super-martingale, so  $\mathbb{E}[\tilde{\Lambda}_{t_0}^{t_1} \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \leq 2cn] \leq \tilde{\Lambda}_{t_0}^{t_0} = \Lambda^{t_0}$ . Hence, using Markov's inequality we get  $\Pr[\tilde{\Lambda}_{t_0}^{t_1} > \Lambda^{t_0} \cdot (ng)^{12} \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \leq 2cn] \leq (ng)^{-12}$ . Thus, by the definition of  $\tilde{\Lambda}_{t_0}^{t_1}$  in Equation (5.6), we have

$$\Pr \left[ \Lambda^{t_1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{t_1-1}} \leq \Lambda^{t_0} \cdot (ng)^{12} \cdot \exp \left( \frac{3\alpha}{n} \cdot B_{t_0}^{t_1-1} - \frac{\alpha \varepsilon}{n} \cdot G_{t_0}^{t_1-1} \right) \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \leq 2cn \right] \geq 1 - (ng)^{-12}. \quad (\text{B.5})$$

Further, if in addition to the two events  $\{\tilde{\Lambda}_{t_0}^{t_1} \leq \Lambda^{t_0} \cdot (ng)^{12}\}$  and  $\{\Lambda^{t_0} \leq 2cn\}$ , also the event  $\{G_{t_0}^{t_1-1} \geq r \cdot \Delta_s\}$  holds, then

$$\begin{aligned} \Lambda^{t_1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{t_1-1}} & \leq \Lambda^{t_0} \cdot (ng)^{12} \cdot \exp \left( \frac{3\alpha}{n} \cdot B_{t_0}^{t_1-1} - \frac{\alpha \varepsilon}{n} \cdot G_{t_0}^{t_1-1} \right) \\ & \leq 2cn \cdot (ng)^{12} \cdot \exp \left( \frac{3\alpha}{n} \cdot (1-r) \cdot \Delta_s - \frac{\alpha \varepsilon}{n} \cdot r \cdot \Delta_s \right) \\ & \stackrel{(a)}{=} 2cn \cdot (ng)^{12} \cdot \exp \left( -\frac{\alpha \varepsilon}{n} \cdot \frac{r}{2} \cdot \Delta_s \right) \\ & = 2cn \cdot (ng)^{12} \cdot \exp \left( -\frac{\alpha \varepsilon}{n} \cdot \frac{r}{2} \cdot \frac{60c_s}{\alpha \varepsilon r} \cdot n \cdot \max\{\log n, g\} \right) \\ & \stackrel{(b)}{\leq} 2cn \cdot (ng)^{12} \cdot \exp(-30 \cdot \max\{\log n, g\}) \\ & \leq 2cn \cdot (ng)^{12} \cdot \exp(-15 \cdot \log(ng)) \\ & \leq 1, \end{aligned}$$

where we used in (a) that  $r = \frac{6}{6+\varepsilon}$  implies  $\frac{3\alpha}{n} \cdot (1-r) = \frac{3\alpha}{n} \cdot \frac{\varepsilon}{6+\varepsilon} = \frac{\alpha \varepsilon}{n} \cdot \frac{r}{2}$  and in (b) that  $c_s \geq 1$  and  $\alpha \leq 1/2$ . Also  $\Lambda^{t_1} \geq 2n$  holds deterministically, so we can deduce from the above inequality that  $\mathbf{1}_{\mathcal{E}_{t_0}^{t_1-1}} = 0$ , that is,

$$\Pr \left[ \neg \mathcal{E}_{t_0}^{t_1-1} \mid \mathfrak{F}^{t_0}, \tilde{\Lambda}_{t_0}^{t_1} \leq \Lambda^{t_0} \cdot (ng)^{12}, \Lambda^{t_0} \leq 2cn, G_{t_0}^{t_1-1} \geq r \cdot \Delta_s \right] = 1.$$

Recalling the definition of  $\mathcal{E}_{t_0}^{t_1-1} := \bigcap_{t \in [t_0, t_1-1]} \{\Lambda^t > cn\}$ , and taking the union bound over Equations (B.4) and (B.5) yields

$$\Pr \left[ \bigcup_{t \in [t_0, t_0 + \Delta_s]} \{\Lambda^t \leq cn\} \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \leq 2cn \right] \geq 1 - 2 \cdot (ng)^{-12} - (ng)^{-12} \geq 1 - (ng)^{-11}. \quad \square$$

### B.3 Omitted Proofs from Section 6

CLAIM B.1. Consider  $\alpha_1, \alpha_2 > 0$  as defined in Equations (6.2) and (6.3) respectively. Then, for any  $g \geq \frac{\alpha_2}{4\sqrt{\alpha_1}}$  and  $k := k(g) \geq 2$  being the unique integer such that  $(\alpha_1 \log n)^{1/k} < g \leq (\alpha_1 \log n)^{1/(k-1)}$ , it holds that

$$(\alpha_1 \cdot (\log n))^{1/k} \leq \left( \frac{\alpha_2}{4} \cdot (\log n) \right)^{1/(k-1)}.$$

PROOF. Let  $R := \frac{\alpha_2}{4\sqrt{\alpha_1}} = \frac{\sqrt{\alpha_1}}{4 \cdot 84} \leq 1$  (using that  $\alpha_2 = \frac{\alpha_1}{84}$  and  $\alpha_1 \leq 1$ ). By rearranging the target inequality,

$$\begin{aligned} \left( \frac{\alpha_2}{4} \cdot (\log n) \right)^{1/(k-1)} \cdot (\alpha_1 \cdot (\log n))^{-1/k} &= \exp \left( \frac{1}{k-1} \cdot \log \left( \frac{\alpha_2}{4} \log n \right) - \frac{1}{k} \cdot \log(\alpha_1 \log n) \right) \\ &\stackrel{(a)}{\geq} \exp \left( \frac{1}{k-1} \cdot \left( \log R + \frac{1}{k} \cdot \log \log n \right) \right) \\ &\stackrel{(b)}{\geq} \exp \left( \frac{1}{k-1} \cdot (\log R - \log R) \right) = 1, \end{aligned}$$

using in (a) that  $-\frac{1}{k} \log \alpha_1 \geq -\frac{1}{2(k-1)} \log \alpha_1$  and in (b) that  $k \leq \frac{1}{-\log R} \cdot \log \log n$  since  $g \geq R^{-1}$ .  $\square$

### B.4 Omitted Proofs from Section 7

Next, we proceed with a simple smoothness argument for the potential  $V$  defined in Equation (7.2).

LEMMA 7.2 (RESTATE, PAGE 33). Consider the potential  $V := V(\alpha_1, c_4 g)$  for any  $\alpha_1 > 0$ , any  $c_4 > 0$  and any  $g \geq 1$ . Then, (i) for any step  $t \geq 0$ , we have that

$$e^{-\alpha_1} \cdot V^t \leq V^{t+1} \leq e^{\alpha_1} \cdot V^t.$$

Further, (ii) for any  $\hat{c} > 0$ , for any integer  $T > 0$  and any step  $t \geq 0$ , for which there exist steps  $s_0 \in [t - T, t]$  and  $s_1 \in [t, t + T]$ , such that  $V^{s_0} \leq \hat{c}n$  and  $V^{s_1} \leq \hat{c}n$ , we have that

$$V^t \leq e^{\alpha_1 \frac{T}{n}} \cdot 2\hat{c}n.$$

PROOF. First statement. In each step the normalized load of any bin can change by at most 1, i.e.,  $|y_i^{t+1} - y_i^t| \leq 1$  and so  $e^{-\alpha_1} \cdot V_i^t \leq V_i^{t+1} \leq e^{\alpha_1} \cdot V_i^t$ . By aggregating over all bins, we get the claim.

Second statement. For any bin  $i \in [n]$ , in  $T$  steps the normalized load can decrease by at most  $T/n$ , i.e.,  $y_i^{s_1} \geq y_i^t - \frac{T}{n}$ . So, the overload term is bounded by

$$e^{\alpha_1(y_i^t - c_4 g)^+} \leq e^{\alpha_1 \frac{T}{n}} \cdot e^{\alpha_1(y_i^{s_1} - c_4 g)^+} \leq e^{\alpha_1 \frac{T}{n}} \cdot V_i^{s_1}.$$

Similarly,  $y_i^t \geq y_i^{s_0} - \frac{T}{n}$ , and so the underload term is bounded by

$$e^{\alpha_1(-y_i^t - c_4 g)^+} \leq e^{\alpha_1 \frac{T}{n}} \cdot e^{\alpha_1(-y_i^{s_0} - c_4 g)^+} \leq e^{\alpha_1 \frac{T}{n}} \cdot V_i^{s_0}.$$

Hence, by aggregating over all bins and using the preconditions  $V^{s_0} \leq \hat{c}n$  and  $V^{s_1} \leq \hat{c}n$ ,

$$V^t = \sum_{i=1}^n \left[ e^{\alpha_1(y_i^t - c_4 g)^+} + e^{\alpha_1(-y_i^t - c_4 g)^+} \right] \leq e^{\alpha_1 \frac{T}{n}} \cdot \sum_{i=1}^n (V_i^{s_1} + V_i^{s_0}) = e^{\alpha_1 \frac{T}{n}} \cdot (V^{s_1} + V^{s_0}) \leq e^{\alpha_1 \frac{T}{n}} \cdot 2\hat{c}n. \quad \square$$

The following lemma shows that by choosing a large enough offset  $c_5 > 0$  in the potential  $\Psi_0 := \Psi_0(\alpha_1, c_5 g)$  (defined in Equation (6.8)), when  $V^t = e^{O(g)} \cdot cn$ , then  $\Psi_0^t = O(n)$ .

LEMMA 7.3 (RESTATEd, PAGE 34). *Consider any  $c, \hat{c} > 0$  and the potential  $V := V(\alpha_1, c_4 g)$  for any  $\alpha_1 > 0$ , any  $c_4 > 0$  and any  $g \geq 1$ . Further, consider the potential  $\Psi_0 := \Psi_0(\alpha_1, c_5 g)$  with offset  $c_5 := 2 \cdot \max\{c_4, \hat{c}\}$  and  $C := 2e^{2\alpha_1} \cdot c + 1$ . Then, for any step  $t \geq 0$  with  $V^t \leq e^{\alpha_1 \cdot \hat{c} \cdot g} \cdot 2e^{2\alpha_1} cn$ , it holds that  $\Psi_0^t \leq Cn$ .*

PROOF. We start by upper bounding  $\Psi_0^t$ ,

$$\begin{aligned} \Psi_0^t &= \sum_{i=1}^n e^{\alpha_1(y_i^t - c_5 g)^+} \\ &= \sum_{i \in [n]: y_i^t \geq c_5 g} e^{\alpha_1(y_i^t - c_5 g)} + \sum_{i \in [n]: y_i^t < c_5 g} e^0 \\ &\leq e^{-\alpha_1(c_5/2)g} \sum_{i \in [n]: y_i^t \geq c_5 g} e^{\alpha_1(y_i^t - (c_5/2)g)} + n \\ &\stackrel{(a)}{\leq} e^{-\alpha_1(c_5/2)g} \sum_{i \in [n]: y_i^t \geq c_4 g} e^{\alpha_1(y_i^t - c_4 g)} + n \\ &\leq e^{-\alpha_1(c_5/2)g} \sum_{i=1}^n e^{\alpha_1(y_i^t - c_4 g)^+} + n \\ &= e^{-\alpha_1(c_5/2)g} \cdot V^t + n, \end{aligned}$$

where in (a) we used that  $c_5/2 \geq c_4$ . Now, using that  $V^t \leq e^{\alpha_1 \cdot \hat{c} \cdot g} \cdot 2e^{2\alpha_1} cn$  and  $c_5 \geq 2 \cdot \hat{c}$ , we conclude

$$\Psi_0^t \leq e^{-\alpha_1(c_5/2)g} \cdot V^t + n \leq 2e^{2\alpha_1} \cdot cn + n = Cn. \quad \square$$

LEMMA 7.6 (RESTATEd, PAGE 38). *Consider the  $\mathcal{Q}_{g, r_0}$  process for any  $g \in [1, c_6 \log n]$  for  $c_6 > 0$  as defined in Equation (7.1) and any step  $r_0 \geq 0$ . Then, for the potential  $V := V(\alpha_1, c_4 g)$  with  $\alpha_1$  as defined in Equation (6.2),  $c_4, c > 0$  as defined in Lemma 5.7 and  $\tilde{\Delta}_s > 0$  as defined in Equation (7.6), it holds that for any step  $t_0 \geq r_0$  and  $t_1$  such that  $t_0 < t_1 \leq t_0 + 2n \log^5 n$ ,*

$$\Pr \left[ \bigcap_{t \in [t_0, t_1]} \bigcup_{s \in [t, t + \tilde{\Delta}_s]} \{V^s \leq e^{2\alpha_1} cn\} \mid \mathcal{Z}^{r_0}, \mathcal{F}^{t_0}, V^{t_0} \leq cn \right] \geq 1 - n^{-9},$$

PROOF. Analogously to the proof of Lemma 5.11, we begin by defining the event

$$\widetilde{\mathcal{M}}_{t_0}^{t_1} = \left\{ \text{for all } t \in [t_0, t_1] \text{ there exists } s \in [t, t + \tilde{\Delta}_s] \text{ such that } V^s \leq e^{2\alpha_1} cn \right\},$$

that is, if  $\widetilde{\mathcal{M}}_{t_0}^{t_1}$  holds then we have that  $V^s \leq e^{2\alpha_1} cn$  at least once every  $\tilde{\Delta}_s$  steps and so the claim follows.

Note that if for some step  $j_1$  we have that  $V^{j_1} \leq e^{\alpha_1} cn$  and for some  $j_2 \geq j_1$  that  $V^{j_2} > e^{2\alpha_1} cn$ , then there must exist  $j \in (j_1, j_2)$  such that  $V^j \in (e^{\alpha_1} cn, e^{2\alpha_1} cn]$ , since for every  $t \geq 0$  it holds that  $V^{t+1} \leq e^{\alpha_1} \cdot V^t$  (Lemma 7.2 (i)). Let  $t_0 < \tau_1 < \tau_2 < \dots$  and  $t_0 =: s_0 < s_1 < \dots$  be two interlaced sequences defined recursively for  $i \geq 1$  by

$$\tau_i := \inf\{\tau > s_{i-1} : V^\tau \in (e^{\alpha_1} cn, e^{2\alpha_1} cn]\}, \quad \text{and} \quad s_i := \inf\{s > \tau_i : V^s \leq e^{\alpha_1} cn\}.$$

Thus we have

$$t_0 = s_0 < \tau_1 < s_1 < \tau_2 < s_2 < \dots,$$

and since  $\tau_i > \tau_{i-1}$  we have  $\tau_{t_1-t_0} \geq t_1 - t_0$ . Therefore, if the event  $\cap_{i=1}^{t_1-t_0} \{s_i - \tau_i \leq \tilde{\Delta}_s\}$  holds, then also  $\tilde{\mathcal{M}}_{t_0}^{t_1}$  holds.

Recall that by the strong stabilization (Lemma 7.5) we have for any  $i = 1, 2, \dots, t_1 - t_0$  and any  $\tau = t_0 + 1, \dots, t_1$ ,

$$\Pr \left[ \bigcup_{t \in [\tau_i, \tau_i + \tilde{\Delta}_s]} \{V^t \leq e^{\alpha_1} cn\} \mid \mathcal{Z}^{r_0}, \mathfrak{F}^\tau, e^{\alpha_1} cn < V^\tau \leq e^{2\alpha_1} cn, \tau_i = \tau \right] \geq 1 - n^{-11},$$

and by negating and the definition of  $s_i$ ,

$$\Pr \left[ s_i - \tau_i > \tilde{\Delta}_s \mid \mathcal{Z}^{r_0}, \mathfrak{F}^\tau, e^{\alpha_1} cn < V^\tau \leq e^{2\alpha_1} cn, \tau_i = \tau \right] \leq n^{-11}.$$

Since the above bound holds for any  $i \geq 1$  and  $\mathfrak{F}^\tau$ , with  $\tau_i = \tau$ , it follows by the union bound over all  $i = 1, 2, \dots, t_1 - t_0$ , as  $t_1 - t_0 \leq 2n \log^5 n$ ,

$$\Pr \left[ \neg \tilde{\mathcal{M}}_{t_0}^{t_1} \mid \mathcal{Z}^{r_0}, \mathfrak{F}^{t_0}, V^{t_0} \leq cn \right] \leq (t_1 - t_0) \cdot n^{-11} \leq n^{-9}. \quad \square$$

## C INDEX OF POTENTIAL FUNCTIONS AND CONSTANTS FOR UPPER BOUNDS

Table 5. Symbols, Equation and Page References for the Various Potential Functions used (Excluding Some Killed Potentials Used Only in Single Lemmas)

Sym	Potential Function	Reference	Page
$\Gamma$	Hyperbolic cosine potential with $\gamma = \Theta(1/g)$	Equation (4.1)	10
$\Lambda$	Hyperbolic cosine potential with $\alpha = \frac{1}{18}$ and offset $c_4 g$	Equation (5.1)	14
$\Delta$	Absolute value potential	Equation (5.2)	15
$\Upsilon$	Quadratic potential	Equation (5.3)	15
$\tilde{\Lambda}$	Adjusted potential for $\Lambda$	Equation (5.6)	23
$V$	Same as $\Lambda$ with constant smoothing parameter $\alpha_1 \leq \alpha$	Equation (7.2)	32
$\tilde{V}$	Adjusted potential for $V$	Equation (7.5)	33
$\Phi_j$	Super-exponential with $\alpha_2(\log n) \cdot \max\{g^{j-k}, 1\}$ and offset $z_j$	Equations (6.5) and (6.6)	28
$\Psi_j$	Super-exponential with $\alpha_1(\log n) \cdot \max\{g^{j-k}, 1\}$ and offset $z_j$	Equations (6.8) and (6.9)	31



Table 6. Definition of Constants and Other Variables

Sym	Definition	Usage	Reference
$\alpha$	$\gamma := -\log(1 - \frac{1}{8.48})/g$	Smoothing parameter for $\Gamma$	Theorem 4.3
$c_1$	$c_1 := c' + 4$ , where $c'$ is from [58]	$\mathbb{E}[\Delta\Gamma^{t+1} \mid \tilde{\gamma}^t] \leq -\frac{\gamma}{96n} \cdot \Gamma^t + c_1$	Theorem 4.3
$c_2$	$c_2 := \frac{96c_1}{-\log(1-1/(8.48))}$	$\mathbb{E}[\Gamma^t] \leq c_2 ng$	Theorem 4.3
$c_3$	$c_3 := \frac{16}{-\log(1-1/(8.48))}$	$\max_{i \in [n]}  y_i^t  \leq c_3 g \log(ng)$	Equation (4.6)
$D$	$D := 365$	Defines when $\Delta$ is small ( $\Delta^t \leq Dng$ )	Lemma 5.4
$\alpha$	$\alpha := \frac{1}{18}$	Smoothing parameter for $\Lambda$	Equation (5.1)
$c_4$	$c_4 := 2D = 730$	Offset $c_4 g$ for $\Lambda$	Equation (5.1)
$r$	$r := \frac{6}{6+\varepsilon}$	Fraction of good steps	Lemma 5.4
$\varepsilon$	$\varepsilon := \frac{1}{12}$	Appears in drop inequalities for $\Lambda$ and $V$	Lemma 5.7
$c$	$c := 12 \cdot 18$	Defines when $\Lambda$ and $V$ are large	Lemma 5.7
$c_s$	$c_s := c_s(\alpha, c_4, 2c) \geq 1$ in Lemma 5.5	$\Upsilon^t \leq c_s \cdot ng^2$ when $\Lambda^t \leq 2cn$	Lemma 5.10
$\Delta_s$	$\Delta_s := \frac{60c_s}{\alpha \varepsilon r} \cdot n \cdot \max\{\log n, g\}$	$\Lambda^t \leq cn$ every $\Delta_s$ steps	Lemma 5.10
$\kappa$	$\kappa := \frac{2}{\alpha} + c_4 + \frac{\Delta_s}{n \cdot \max\{\log n, g\}}$	$\text{Gap}(m) \leq \kappa \cdot (g + \log n)$	Lemma 5.11
$c_r$	$c_r := c_r(\alpha, c_4) \geq 1$ in Lemma 5.5	$\Upsilon^t \leq 2c_r \cdot n \cdot (c_3 g \log(ng))^2$ when $\Lambda^t \leq e^{c_3 g \log(ng)}$	Lemma 5.9
$\Delta_r$	$\Delta_r := \frac{60c_r^2}{\alpha \varepsilon r} \cdot ng \cdot (\log(ng))^2$	For any $t_0$ , $\Lambda^t \leq cn$ for some $t \in [t_0, t_0 + \Delta_r]$	Lemma 5.9
$\alpha_1$	$\alpha_1 := \frac{1}{6\kappa}$	Part of smoothing parameters in $V$ and $\Psi_j$	Equation (6.2)
$\tilde{c}_s$	$\tilde{c}_s := \tilde{c}_s(\alpha_1, c_4, e^{2\alpha_1} c) \geq 1$ in Lemma 5.5	$\Upsilon^t \leq \tilde{c}_s \cdot ng^2$ when $V^t \leq e^{2\alpha_1} cn$	Lemma 7.5
$\tilde{\Delta}_s$	$\tilde{\Delta}_s := \frac{20 \cdot \tilde{c}_s \cdot \log(2e^{2\alpha_1})}{\alpha_1 \varepsilon r} \cdot ng$	$V^t \leq e^{\alpha_1} n$ every $\tilde{\Delta}_s$ steps	Equation (7.6)
$c_6$	$c_6 := \frac{r}{9 \cdot 20 \cdot \tilde{c}_s \cdot \log(2e^{2\alpha_1})}$	Strong stabilization works for $g \leq c_6 \log n$	Equation (7.1)
$c_5$	$c_5 := 2 \cdot \max\{c_4, \frac{\Delta_s}{ng}\}$	Offset $c_5 g$ for $\Psi_0$ and $\Phi_0$	Equation (7.14)
$C$	$C := 2e^{2\alpha_1} \cdot c + 1$	Defines when $\Phi_j$ and $\Psi_j$ are large	Lemma 7.3
$\alpha_2$	$\alpha_2 := \frac{1}{84 \cdot 6\kappa}$	Part of smoothing parameter in $\Phi_j$	Equation (6.3)
$z_j$	$z_j := c_5 \cdot g + \left\lceil \frac{4}{\alpha_2} \right\rceil \cdot j \cdot g$	Offsets for $\Phi_j$ and $\Psi_j$	Equation (6.7)
$k$	$k \in \mathbb{N}$ s.t. $(\alpha_1 \log n)^{1/k} \leq g \leq (\frac{\alpha_2}{4} \log n)^{1/(k-1)}$	Number of layered induction steps	Section 9

## REFERENCES

- [1] Micah Adler, Soumen Chakrabarti, Michael Mitzenmacher, and Lars Rasmussen. 1998. Parallel randomized load balancing. *Random Structures and Algorithms* 13, 2 (1998), 159–188. DOI : [https://doi.org/10.1002/\(SICI\)1098-2418\(199809\)13:2<159::AID-RSA3>3.3.CO;2-Z](https://doi.org/10.1002/(SICI)1098-2418(199809)13:2<159::AID-RSA3>3.3.CO;2-Z)
- [2] Nir Ailon, Moses Charikar, and Alanthan Newman. 2005. Aggregating inconsistent information: Ranking and clustering. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC'05)*. ACM, 684–693. DOI : <https://doi.org/10.1145/1060590.1060692>
- [3] Dan Alistarh, James Aspnes, and Rati Gelashvili. 2018. Space-optimal majority in population protocols. In *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'18)*. SIAM, 2221–2239. DOI : <https://doi.org/10.1137/1.9781611975031.144>
- [4] Dan Alistarh, Trevor Brown, Justin Kopinsky, Jerry Zheng Li, and Giorgi Nadiradze. 2018. Distributionally linearizable data structures. In *Proceedings of the 30th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'18)*. ACM, New York, NY, 133–142. DOI : <https://doi.org/10.1145/3210377.3210411>
- [5] Dan Alistarh, Rati Gelashvili, and Joel Rybicki. 2022. Fast graphical population protocols. In *Proceedings of the 25th International Conference on Principles of Distributed Systems (OPODIS'21)*. Quentin Bramas, Vincent Gramoli, and Alessia Milani (Eds.), Vol. 217, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 14:1–14:18. DOI : <https://doi.org/10.4230/LIPIcs.OPODIS.2021.14>
- [6] Dan Alistarh, Giorgi Nadiradze, and Amirmojtaba Sabour. 2020. Dynamic averaging load balancing on cycles. In *Proceedings of the 47th International Colloquium on Automata, Languages, and Programming (ICALP'20)*. Artur Czumaj, Anuj Dawar, and Emanuela Merelli (Eds.), Vol. 168, Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 7:1–7:16. DOI : <https://doi.org/10.4230/LIPIcs.ICALP.2020.7>

- [7] Noga Alon, Ori Gurel-Gurevich, and Eyal Lubetzky. 2010. Choice-memory tradeoff in allocations. *Annals of Applied Probability* 20, 4 (2010), 1470–1511. DOI : <https://doi.org/10.1214/09-AAP656>
- [8] Yossi Azar, Andrei Z. Broder, Anna R. Karlin, Michael Mitzenmacher, and Eli Upfal. 2020. The ACM Paris Kanellakis Theory and Practice Award. Retrieved from <https://www.acm.org/media-center/2021/may/technical-awards-2020>. Accessed 06-01-2023.
- [9] Yossi Azar, Andrei Z. Broder, Anna R. Karlin, and Eli Upfal. 1999. Balanced allocations. *SIAM Journal on Computing* 29, 1 (1999), 180–200. DOI : <https://doi.org/10.1137/S0097539795288490>
- [10] Nikhil Bansal and Ohad N. Feldheim. 2022. The power of two choices in graphical allocation. In *Proceedings of the 54th Annual ACM Symposium on Theory of Computing (STOC'22)*. Stefano Leonardi and Anupam Gupta (Eds.), ACM, New York, NY, 52–63. DOI : <https://doi.org/10.1145/3519935.3519995>
- [11] Nikhil Bansal and William Kuszmaul. 2022. Balanced allocations: The heavily loaded case with deletions. In *Proceedings of the 63rd IEEE Annual Symposium on Foundations of Computer Science (FOCS'22)*. IEEE, 801–812. DOI : <https://doi.org/10.1109/FOCS54457.2022.00081>
- [12] Luca Becchetti, Andrea Clementi, Emanuele Natale, Francesco Pasquale, Riccardo Silvestri, and Luca Trevisan. 2014. Simple dynamics for plurality consensus. In *Proceedings of the 26th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'14)*. ACM, New York, NY, 247–256. DOI : <https://doi.org/10.1145/2612669.2612677>
- [13] Luca Becchetti, Andrea E. F. Clementi, Emanuele Natale, Francesco Pasquale, and Gustavo Posta. 2019. Self-stabilizing repeated balls-into-bins. *Distributed Comput.* 32, 1 (2019), 59–68. DOI : <https://doi.org/10.1007/s00446-017-0320-4>
- [14] Luca Becchetti, Andrea E. F. Clementi, Emanuele Natale, Francesco Pasquale, and Luca Trevisan. 2016. Stabilizing consensus with many opinions. In *Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'16)*. Robert Krauthgamer (Ed.), SIAM, 620–635. DOI : <https://doi.org/10.1137/1.9781611974331.ch46>
- [15] Michael A. Bender and William Kuszmaul. 2021. Randomized cup game algorithms against strong adversaries. In *Proceedings of the 32nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'21)*. Dániel Marx (Ed.), SIAM, 2059–2077. DOI : <https://doi.org/10.1137/1.9781611976465.123>
- [16] Itai Benjamini and Yury Makarychev. 2012. Balanced allocation: Memory performance tradeoffs. *Annals of Applied Probability* 22, 4 (2012), 1642–1649. DOI : <https://doi.org/10.1214/11-AAP804>
- [17] Petra Berenbrink, André Brinkmann, Tom Friedetzky, and Lars Nagel. 2014. Balls into non-uniform bins. *IEEE International Symposium on Parallel & Distributed Processing* 74, 2 (2014), 2065–2076. DOI : <https://doi.org/10.1016/j.jpdc.2013.10.008>
- [18] Petra Berenbrink, Artur Czumaj, Matthias Englert, Tom Friedetzky, and Lars Nagel. 2012. Multiple-choice balanced allocation in (almost) parallel. In *Proceedings of the 16th International Workshop on Randomization and Computation (RANDOM'12)*. Springer-Verlag, Berlin, 411–422. DOI : [https://doi.org/10.1007/978-3-642-32512-0\\_35](https://doi.org/10.1007/978-3-642-32512-0_35)
- [19] Petra Berenbrink, Artur Czumaj, Angelika Steger, and Berthold Vöcking. 2006. Balanced allocations: The heavily loaded case. *SIAM Journal on Computing* 35, 6 (2006), 1350–1385. DOI : <https://doi.org/10.1137/S00975397044435X>
- [20] Petra Berenbrink, Tom Friedetzky, Peter Kling, Frederik Mallmann-Trenn, Lars Nagel, and Chris Wastell. 2018. Self-stabilizing balls and bins in batches: The power of leaky bins. *Algorithmica* 80, 12 (2018), 3673–3703. DOI : <https://doi.org/10.1007/s00453-018-0411-z>
- [21] Mark Braverman and Elchanan Mossel. 2008. Noisy sorting without resampling. In *Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'08)*. Shang-Hua Teng (Ed.), SIAM, 268–276. DOI : <https://doi.org/10.1080/15427951.2006.10129115>
- [22] Fan Chung and Linyuan Lu. 2006. Concentration inequalities and martingale inequalities: A survey. *Internet Mathematics Journal* 3, 1 (2006), 79–127. DOI : <https://doi.org/10.1080/15427951.2006.10129115>
- [23] Richard Cole, Alan M. Frieze, Bruce M. Maggs, Michael Mitzenmacher, Andréa W. Richa, Ramesh K. Sitaraman, and Eli Upfal. 1998. On balls and bins with deletions. In *Proceedings of the 2nd International Workshop on Randomization and Computation (RANDOM'98)*. Michael Luby, José D. P. Rolim, and Maria J. Serna (Eds.), Vol. 1518, Springer, 145–158. DOI : [https://doi.org/10.1007/3-540-49543-6\\_12](https://doi.org/10.1007/3-540-49543-6_12)
- [24] Colin Cooper, Robert Elsässer, and Tomasz Radzik. 2014. The power of two choices in distributed voting. In *Proceedings of the 41st International Colloquium on Automata, Languages, and Programming (ICALP'14)*. Javier Esparza, Pierre Fraigniaud, Thore Husfeldt, and Elias Koutsoupias (Eds.), Vol. 8573, Springer, Berlin, 435–446. DOI : [https://doi.org/10.1007/978-3-662-43951-7\\_37](https://doi.org/10.1007/978-3-662-43951-7_37)
- [25] Artur Czumaj and Volker Stemann. 2001. Randomized allocation processes. *Random Structures and Algorithms Journal* 18, 4 (2001), 297–331. DOI : <https://doi.org/10.1002/rsa.1011>
- [26] Michael Dahlin. 2000. Interpreting stale load information. *IEEE Parallel Distributed Systems* 11, 10 (2000), 1033–1047. DOI : <https://doi.org/10.1109/71.888643>
- [27] Pamela Delgado, Diego Didona, Florin Dinu, and Willy Zwaenepoel. 2016. Job-aware scheduling in eagle: Divide and stick to your probes. In *Proceedings of the 7th ACM Symposium on Cloud Computing (SoCC'16)*. Marcos K. Aguilera, Brian Cooper, and Yanlei Diao (Eds.), ACM, 497–509. DOI : <https://doi.org/10.1145/2987550.2987563>

- [28] Pamela Delgado, Florin Dinu, Anne-Marie Kermarrec, and Willy Zwaenepoel. 2015. Hawk: Hybrid datacenter scheduling. In *Proceedings of the 2015 USENIX Annual Technical Conference (USENIX'15)*. Shan Lu and Erik Riedel (Eds.), USENIX, 499–510.
- [29] Christina Delimitrou, Daniel Sánchez, and Christos Kozyrakis. 2015. Tarcil: Reconciling scheduling speed and quality in large shared clusters. In *Proceedings of the 6th ACM Symposium on Cloud Computing (SoCC'15)*. Shahram Ghaharizadeh, Sumita Barahmand, Magdalena Balazinska, and Michael J. Freedman (Eds.), ACM, 97–110. DOI: <https://doi.org/10.1145/2806777.2806779>
- [30] Benjamin Doerr, Leslie Ann Goldberg, Lorenz Minder, Thomas Sauerwald, and Christian Scheideler. 2011. Stabilizing consensus with the power of two choices. In *Proceedings of the 23rd Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'11)*. Rajmohan Rajaraman and Friedhelm Meyer auf der Heide (Eds.), ACM, Dagstuhl, Germany, 149–158. DOI: <https://doi.org/10.1145/1989493.1989516>
- [31] Devdatt P. Dubhashi and Alessandro Panconesi. 2009. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, Cambridge. DOI: <https://doi.org/10.1017/CBO9780511581274>
- [32] Ohad Noy Feldheim, Ori Gurel-Gurevich, and Jiange Li. 2021. Long-term Balanced Allocation via Thinning. Retrieved from <https://doi.org/10.48550/ARXIV.2110.05009>. Accessed 06-01-2023.
- [33] Ohad Noy Feldheim and Jiange Li. 2020. Load balancing under  $d$ -thinning. *Electronic Communications in Probability* 25 (2020), 13. DOI: <https://doi.org/10.1214/19-ecp282>
- [34] William Feller. 1971. *An Introduction to Probability Theory and Its Applications. Vol. II*. John Wiley & Sons, Inc., New York-London-Sydney.
- [35] Catherine S. Greenhill, Bernard Mans, and Ali Pourmiri. 2020. Balanced allocation on dynamic hypergraphs. In *Proceedings of the 24th International Workshop on Randomization and Computation (RANDOM'20)*. Jaroslav Byrka and Raghu Meka (Eds.), Vol. 176, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 11:1–11:22. DOI: <https://doi.org/10.4230/LIPIcs.APPROX/RANDOM.2020.11>
- [36] Anupam Gupta, Ravishankar Krishnaswamy, Amit Kumar, and Sahil Singla. 2020. Online carpooling using expander decompositions. In *Proceedings of the 40th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS'20)*. Nitin Saxena and Sunil Simon (Eds.), Vol. 182, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 23:1–23:14. DOI: <https://doi.org/10.4230/LIPIcs.FSTTCS.2020.23>
- [37] Kiyoshi Itô and Henry P. McKean, Jr. 1965. *Diffusion Processes and Their Sample Paths*. Academic Press, Inc., Publishers, New York; Springer-Verlag, Berlin-New York, Berlin. DOI: <https://doi.org/10.1007/978-3-642-62025-6>
- [38] Richard M. Karp, Michael Luby, and Friedhelm Meyer auf der Heide. 1996. Efficient PRAM simulation on a distributed memory machine. *Algorithmica* 16, 4–5 (1996), 517–542. DOI: <https://doi.org/10.1007/BF01940878>
- [39] Krishnamurthy K. Kenthapadi and Rina Panigrahy. 2006. Balanced allocation on graphs. In *Proceedings of 17th ACM-SIAM Symposium on Discrete Algorithms (SODA'06)*. Society for Industrial and Applied Mathematics, 434–443. DOI: <https://doi.org/10.1145/1109557.1109606>
- [40] Christoph Lenzen, Merav Parter, and Eylon Yogev. 2019. Parallel balanced allocations: The heavily loaded case. In *Proceedings of the 31st Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'19)*. Christian Scheideler and Petra Berenbrink (Eds.), ACM, New York, NY, 313–322. DOI: <https://doi.org/10.1145/3323165.3323203>
- [41] Christoph Lenzen and Roger Wattenhofer. 2011. Tight bounds for parallel randomized load balancing: Extended abstract. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing (STOC'11)*. ACM, 11–20. DOI: <https://doi.org/10.1145/1993636.1993639>
- [42] Dimitrios Los and Thomas Sauerwald. 2022. Balanced allocations in batches: Simplified and generalized. In *Proceedings of the 34th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'22)*. ACM, New York, NY, 389–399. DOI: <https://doi.org/10.1145/3490148.3538593>
- [43] Dimitrios Los and Thomas Sauerwald. 2022. Balanced allocations with incomplete information: The power of two queries. In *Proceedings of the 13th Innovations in Theoretical Computer Science Conference (ITCS'22)*. Mark Braverman (Ed.), Vol. 215, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 103:1–103:23. DOI: <https://doi.org/10.4230/LIPIcs.ITCS.2022.103>
- [44] Dimitrios Los and Thomas Sauerwald. 2022. Brief announcement: Tight bounds for repeated balls-into-bins. In *Proceedings of the 34th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'22)*. ACM, New York, NY, 419–421. DOI: <https://doi.org/10.1145/3490148.3538561>
- [45] Dimitrios Los and Thomas Sauerwald. 2023. Balanced allocations in batches: The tower of two choices. In *Proceedings of the 35th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'23)*. Kunal Agrawal and Julian Shun (Eds.), ACM, 51–61. DOI: <https://doi.org/10.1145/3558481.3591088>
- [46] Dimitrios Los, Thomas Sauerwald, and John Sylvester. 2022. Balanced allocations: Caching and packing, twinning and thinning. In *Proceedings of the 33rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'22)*. SIAM, Alexandria, Virginia, 1847–1874. DOI: <https://doi.org/10.1137/1.9781611977073.74>

- [47] Yi Lu, Qiaomin Xie, Gabriel Kliot, Alan Geller, James R. Larus, and Albert G. Greenberg. 2011. Join-idle-queue: A novel load balancing algorithm for dynamically scalable web services. *Perform. Evaluation* 68, 11 (2011), 1056–1071. DOI: <https://doi.org/10.1016/j.peva.2011.07.015>
- [48] Dragoslav S. Mitrinović, Josip E. Pečarić, and Arlington M. Fink. 1993. *Classical and New Inequalities in Analysis*. Mathematics and its Applications (East European Series), Vol. 61. Kluwer Academic Publishers Group, Dordrecht. DOI: <https://doi.org/10.1007/978-94-017-1043-5>
- [49] Michael Mitzenmacher. 1999. On the analysis of randomized load balancing schemes. *Theory Comput. Syst.* 32, 3 (1999), 361–386. DOI: <https://doi.org/10.1007/s002240000122>
- [50] Michael Mitzenmacher. 2000. How useful is old information? *IEEE Trans. Parallel Distributed Syst.* 11, 1 (2000), 6–20. DOI: <https://doi.org/10.1109/71.824633>
- [51] Michael Mitzenmacher. 2020. Scheduling with predictions and the price of misprediction. In *Proceedings of the 11th Innovations in Theoretical Computer Science Conference (ITCS’20)*. Thomas Vidick (Ed.), Vol. 151, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 14:1–14:18. DOI: <https://doi.org/10.4230/LIPIcs.ITCS.2020.14>
- [52] Michael Mitzenmacher, Andréa W. Richa, and Ramesh Sitaraman. 2001. The power of two random choices: A survey of techniques and results. In *Proceedings of the Handbook of Randomized Computing, Vol. I, II*. Kluwer Acad. Publ., Dordrecht, Netherlands, 255–312. DOI: [https://doi.org/10.1007/978-1-4615-0013-1\\_9](https://doi.org/10.1007/978-1-4615-0013-1_9)
- [53] Michael Mitzenmacher and Eli Upfal. 2017. *Probability and Computing* (2nd. ed.). Cambridge University Press, Cambridge. Randomization and probabilistic techniques in algorithms and data analysis.
- [54] Michael Mitzenmacher and Sergei Vassilvitskii. 2021. *Algorithms with Predictions*. Cambridge University Press, Cambridge, 646–662. DOI: <https://doi.org/10.1017/9781108637435.037>
- [55] Michael Mitzenmacher and Berthold Vöcking. 1999. *The Asymptotics of Selecting the Shortest of Two, Improved*. Harvard Computer Science Group Technical Report TR-08-99. Harvard University.
- [56] Giorgi Nadiradze. 2021. *On Achieving Scalability through Relaxation*. Ph.D. Dissertation. IST Austria. DOI: <https://doi.org/10.15479/at:ista:10429>. A revised version of reference [4].
- [57] Kay Ousterhout, Patrick Wendell, Matei Zaharia, and Ion Stoica. 2013. Sparrow: Distributed, low latency scheduling. In *Proceedings of the ACM SIGOPS 24th Symposium on Operating Systems Principles (SOSP’13)*. Michael Kaminsky and Mike Dahlin (Eds.), ACM, 69–84. DOI: <https://doi.org/10.1145/2517349.2522716>
- [58] Yuval Peres, Kunal Talwar, and Udi Wieder. 2015. Graphical balanced allocations and the  $(1 + \beta)$ -choice process. *Random Structures and Algorithms Journal* 47, 4 (2015), 760–775. DOI: <https://doi.org/10.1002/rsa.20558>
- [59] Martin Raab and Angelika Steger. 1998. “Balls into bins”—A simple and tight analysis. In *Proceedings of the 2nd International Workshop on Randomization and Computation (RANDOM’98)*. Vol. 1518, Springer, Barcelona, Spain, 159–170. DOI: [https://doi.org/10.1007/3-540-49543-6\\_13](https://doi.org/10.1007/3-540-49543-6_13)
- [60] Ziv Scully, Isaac Grosf, and Michael Mitzenmacher. 2022. Uniform bounds for scheduling with job size estimates. In *Proceedings of the 13th Innovations in Theoretical Computer Science Conference (ITCS’22)*. Mark Braverman (Ed.), Vol. 215, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 114:1–114:30. DOI: <https://doi.org/10.4230/LIPIcs.ITCS.2022.114>
- [61] Kunal Talwar and Udi Wieder. 2014. Balanced allocations: A simple proof for the heavily loaded case. In *Proceedings of the 41st International Colloquium on Automata, Languages, and Programming (ICALP’14)*. Javier Esparza, Pierre Fraigniaud, Thore Husfeldt, and Elias Koutsoupias (Eds.), Vol. 8572, Springer, 979–990. DOI: [https://doi.org/10.1007/978-3-662-43948-7\\_81](https://doi.org/10.1007/978-3-662-43948-7_81)
- [62] Udi Wieder. 2007. Balanced allocations with heterogenous bins. In *Proceedings of the 19th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA’07)*. ACM, New York, NY, 188–193. DOI: <https://doi.org/10.1145/1248377.1248407>
- [63] Udi Wieder. 2017. Hashing, load balancing and multiple choice. *Foundations and Trends® in Theoretical Computer Science* 12, 3–4 (2017), 275–379. DOI: <https://doi.org/10.1561/04000000070>

Received 13 January 2023; revised 10 September 2023; accepted 16 September 2023