

# A Faster Exponential Time Algorithm for Bin Packing With a Constant Number of Bins via Additive Combinatorics \*

Jesper Nederlof<sup>†</sup>      Jakub Pawlewicz<sup>‡</sup>      Céline M. F. Swennenhuis<sup>§</sup>  
 Karol Węgrzycki<sup>¶</sup>

## Abstract

In the Bin Packing problem one is given  $n$  items with weights  $w(1), \dots, w(n)$  and  $m$  bins with capacities  $c_1, \dots, c_m$ . The goal is to find a partition of the items into sets  $S_1, \dots, S_m$  such that  $w(S_j) \leq c_j$  for every bin  $j$ , where  $w(X)$  denotes  $\sum_{i \in X} w(i)$ .

Björklund, Husfeldt and Koivisto (SICOMP 2009) presented an  $\mathcal{O}^*(2^n)$  time algorithm for Bin Packing (the  $\mathcal{O}^*(\cdot)$  notation omits factors polynomial in the input size). In this paper, we show that for every  $m \in \mathbb{N}$  there exists a constant  $\sigma_m > 0$  such that an instance of Bin Packing with  $m$  bins can be solved in  $\mathcal{O}(2^{(1-\sigma_m)n})$  randomized time. Before our work, such improved algorithms were not known even for  $m$  equals 4.

A key step in our approach is the following new result in Littlewood-Offord theory on the additive combinatorics of subset sums: For every  $\delta > 0$  there exists an  $\varepsilon > 0$  such that if  $|\{X \subseteq \{1, \dots, n\} : w(X) = v\}| \geq 2^{(1-\varepsilon)n}$  for some  $v$  then  $|\{w(X) : X \subseteq \{1, \dots, n\}\}| \leq 2^{\delta n}$ .

\*An extended abstract of this manuscript was presented and included in the proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms.

<sup>†</sup>Utrecht University, The Netherlands, [j.nederlof@uu.nl](mailto:j.nederlof@uu.nl). Supported by the project CRACKNP that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 853234).

<sup>‡</sup>Institute of Informatics, University of Warsaw, Poland, [pan@mimuw.edu.pl](mailto:pan@mimuw.edu.pl).

<sup>§</sup>Eindhoven University of Technology, The Netherlands, [c.m.f.swennenhuis@tue.nl](mailto:c.m.f.swennenhuis@tue.nl). Supported by the Netherlands Organization for Scientific Research under project no. 613.009.031b.

<sup>¶</sup>Saarland University and Max Planck Institute for Informatics, Saarbrücken, Germany, [wegrzycki@cs.uni-saarland.de](mailto:wegrzycki@cs.uni-saarland.de). This work is part of the project TIPEA that has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No. 850979). Author was also supported Foundation for Polish Science (FNP), by the grants 2016/21/N/ST6/01468 and 2018/28/T/ST6/00084 of the Polish National Science Center and project TOTAL that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 677651).



European Research Council  
Established by the European Commission



# 1 Introduction

A central aim in contemporary algorithm design is to minimize the worst-case complexity of an algorithm for a given (supposedly) hard computational problem in a fine-grained sense. The underlying goal is to reveal the optimal running time witnessed by (1) an algorithm with worst-case complexity  $T(n)$  on instances with parameter  $n$ , and (2) a lower bound that excludes improvements to  $T(n)^{1-\varepsilon}$  time for some constant  $\varepsilon > 0$ . For some problems, it is an especially intriguing question whether natural running times of the basic algorithms solving them are optimal. One of the most important instances of such a question for an NP-complete problem is about improvements over a relatively direct dynamic programming algorithm for Set Cover:

**Question 1:** Can Set Cover with  $n$  elements be solved in  $\mathcal{O}^*((2-\varepsilon)^n)$  time, for some  $\varepsilon > 0$ ?

Here and throughout the paper, we use the  $\mathcal{O}^*$  notation to hide factors polynomial in the input size.<sup>1</sup> Unfortunately, Question 1 seems to have a fate similar to the Strong Exponential Time Hypothesis (which is about a similar improvement for the CNF-SAT problem): While there is an increasing interest and dependence on its validity (see e.g. [14, 40]), we seem to be far from resolving it.

Therefore, it is natural to study Question 1 for special cases of Set Cover. And indeed, improved algorithms of the type asked in Question 1 were already presented for instances with small sets [39], (more generally) large solutions [46], and for several other cases (see e.g. [27]).

However, some of the most fundamental NP-complete problems that are special cases of Set Cover such as Graph Coloring and Directed Hamiltonicity<sup>2</sup> still defy considerable research efforts to obtain the type of improved algorithms asked for in Question 1 (see e.g. [10, 22]).

**Bin Packing** We study one such fundamental NP-complete problem, the *Bin Packing problem*: Given item weights  $w(1), \dots, w(n) \in \mathbb{N}$  and capacities  $c_1, \dots, c_m \in \mathbb{N}$ , can we partition items  $\{1, \dots, n\}$  into  $m$  sets  $S_1, \dots, S_m \subseteq \{1, \dots, n\}$  such that  $w(S_j) \leq c_j$  for each  $j \in \{1, \dots, m\}$ ? Here  $w(X)$  denotes  $\sum_{i \in X} w(i)$ . Due to its elegant formulation and clear practical applicability, Bin Packing is a central problem in computer science. For example, it models the most basic non-trivial scheduling problem with multiple machines. While Bin Packing has been extensively studied from an approximation and online algorithms perspective [12], much less research has been devoted to *exact algorithms* for Bin Packing.

The currently fastest algorithm for Bin Packing is a consequence<sup>3</sup> of the aforementioned algorithm for Set Cover from [9], and it runs in  $\mathcal{O}^*(2^n)$  time. With Question 1 on the horizon, we ask whether this can be improved:

**Question 2:** Can Bin Packing with  $n$  items be solved in  $\mathcal{O}((2-\varepsilon)^n)$  time, for some  $\varepsilon > 0$ ?

The only improvement over the  $\mathcal{O}^*(2^n)$  time algorithm for Bin Packing is due to Lente et al. [41], who gave an  $\mathcal{O}^*(m^{n/2})$  time algorithm. Note that this is only an improvement for  $m = 2, 3$  bins and Question 2 remained illusive for  $m = 4$  already. In stark contrast, our main result is an improvement over the  $\mathcal{O}^*(2^n)$  time algorithm for *every* constant number of bins:

**Theorem 1.1** (Main Theorem). *For every  $m \in \mathbb{N}$  there is a constant  $\sigma_m > 0$  such that every Bin Packing instance with  $m$  bins can be solved in  $\mathcal{O}(2^{(1-\sigma_m)n})$  time with high probability.*

<sup>1</sup>In principle, it is natural to assume the Set Cover instance has  $n$  elements and  $\text{poly}(n)$  sets, but an algorithm by Björklund et al. [9] solves Set Cover instances in  $2^n n^{\mathcal{O}(1)}$  time irrespective of the number of sets.

<sup>2</sup>Krauthgamer and Trabelsi [40] rewrite a Directed Hamiltonicity instance efficiently as a Set Cover instance.

<sup>3</sup>Assuming the capacity of each bin equals  $c$ , create a Set Cover instance with all item sets of weight at most  $c$ .

While our algorithm does not resolve Question 2, we believe it makes substantial progress on it because (1) Set Cover with a constant-sized solution is as least as hard as a general Set Cover, and (2) the other extreme, Set Cover with a linear number sets in the solution (and hence Bin Packing with a linear number of bins with equal capacity<sup>3</sup>), can be solved in  $\mathcal{O}((2 - \varepsilon)^n)$  time (see [46]).

## 1.1 Our Approach for Proving Theorem 1.1

As our starting point, we extend the methods from [8, 46] to show that instances of Bin Packing with the following restrictions admit an  $\mathcal{O}^*(2^{(1-\sigma_m)n})$  time randomized algorithm for some  $\sigma_m > 0$ :

**Restriction (1)** the instance has a *low concentration* in the sense that  $\beta(w) \leq 2^{(1-\varepsilon)n}$  for some  $\varepsilon > 0$ , where  $\beta(w)$  is the *maximum frequency*  $\max_v |\{X \subseteq \{1, \dots, n\} : w(X) = v\}|$ , and

**Restriction (2)** the instance is *tight* in the sense that  $\sum_{j=1}^m c_j = w(\{1, \dots, n\})$ .

Fix a set of bins  $L \subseteq \{1, \dots, m\}$  and recall  $(S_1, \dots, S_m)$  denotes a solution. The crux of Restriction (1) and Restriction (2) is that together they imply that the number of distinct sets  $S^L := \bigcup_{j \in L} S_j$  is at most  $2^{(1-\varepsilon)n}$  since  $w(S^L) = \sum_{j \in L} c_j$ . We explain in § 1.1.3 how this allows a faster algorithm via the methods of [8, 46]. In the nutshell, these sets  $S^L$  correspond to the *candidates* for the solutions that we need to check and bounding this number automatically corresponds to the running time of the algorithm.

However, extending this algorithm to an improved algorithm that solves *all* instances with a constant number of bins requires both new combinatorial (for relaxing Restriction (1)) and new algorithmic (for relaxing Restriction (2)) insights that are our main contributions. Therefore we first discuss these insights.

### 1.1.1 Combinatorial Ideas: Lifting Restriction (1) via Littlewood–Offord Theory.

Our main combinatorial contribution is a new structural insight on instances that do not satisfy Restriction (1), i.e. vectors  $w$  with  $|\{X \subseteq \{1, \dots, n\} : w(X) = v\}| \geq 2^{(1-\varepsilon)n}$  for some  $v$  and  $\varepsilon > 0$ .

The challenge of determining the structure of such vectors  $w$  is well-known in additive combinatorics as the *Littlewood–Offord Problem*. Its rich theory has found applications ranging from pure mathematics (such as estimating the singularity of random Bernoulli matrices [52] or zeroes of random polynomials [42]), to database security [28], and to computational complexity theory [18, 35, 44]. See also the designated chapter in the standard textbook on additive combinatorics [51]. However, whereas most works (with notable exceptions being e.g. [29, 49]) assumed inversely *polynomially* small concentration, e.g.  $\beta(w) \geq 2^n / \text{poly}(n)$ , Restriction (1) is about inversely *exponentially* small concentration.

Recent work studied such exponentially small concentration with applications to improved exponential time algorithms for the Subset Sum problem [2, 5]. Specifically, they studied trade-off between the parameters  $\beta(w)$  and  $|w(2^{\{1, \dots, n\}})| := |\{w(X) : X \subseteq \{1, \dots, n\}\}|$ . Two extremal cases are:

$$\begin{array}{ll} \text{If } w_a := (0, 0, \dots, 0) & \text{then } |w_a(2^{\{1, \dots, n\}})| = 1 \text{ and } \beta(w_a) = 2^n \\ \text{If } w_b := (1, 2, \dots, 2^{n-1}) & \text{then } |w_b(2^{\{1, \dots, n\}})| = 2^n \text{ and } \beta(w_b) = 1 \end{array}$$

One may suspect that all vectors  $w \in \mathbb{Z}^n$  are a combination of these two extremes and therefore that a smooth trade-off between  $\beta(w)$  and  $w(2^{\{1, \dots, n\}})$  can be proved. In the case  $w \in \mathbb{F}_2^{n \times n}$  (where

$|w(2^{\{1,\dots,n\}})|\beta(w) = 2^n$ ), this suspicion can be confirmed.<sup>4</sup> Observe that a similar trade-off for  $w \in \mathbb{Z}^n$  would allow us to lift Restriction (1) by a simple  $\mathcal{O}^*(|w(2^{\{1,\dots,n\}})|^m)$  time algorithm for Bin Packing (Lemma 3.5).

Unfortunately, this intuition is not true and the case  $w \in \mathbb{Z}^n$  is far more subtle. For instance, Wiman [55] showed in his remarkable bachelor thesis that, surprisingly, vectors satisfying simultaneously both  $|w(2^{\{1,\dots,n\}})| \geq 2^{(1-\varepsilon)n}$  and  $\beta(w) \geq 2^{0.2563n}$  exist for any  $\varepsilon > 0$ . Our main combinatorial contribution is a proof that instances with the same parameters but the roles of  $\beta(w)$  and  $|w(2^{\{1,\dots,n\}})|$  swapped do *not* exist:<sup>5</sup>

**Theorem 1.2.** *Let  $\varepsilon > 0$ . If  $\beta(w) \geq 2^{(1-\varepsilon)n}$ , then  $|w(2^{\{1,\dots,n\}})| \leq 2^{\delta n}$ , where  $\delta = \delta(\varepsilon) = \mathcal{O}_{\varepsilon \rightarrow 0} \left( \frac{\log \log(1/\varepsilon)}{\sqrt{\log(1/\varepsilon)}} \right)$ .*

The dependency of  $\delta$  on  $\varepsilon$  was recently improved to  $\delta(\varepsilon) = \mathcal{O}(\sqrt{\varepsilon})$  by Jain et al. [32]. The previous best bounds were given by Austrin et al. [2] who found a connection with *Uniquely Decodable Code Pairs (UDCPs)* from information theory (see Subsection 1.2 for details). This implies for example that if  $\beta(w) \geq 2^{(1-\varepsilon)n}$ , then  $|w(2^{\{1,\dots,n\}})| \leq 2^{0.4228n + \sqrt{\varepsilon}}$  by a result on UDCPs from [4]. However, the reduction from [2] is symmetric with respect to swapping the roles of  $\beta(w)$  and  $|w(2^{\{1,\dots,n\}})|$ , and thus by the result from [55] UDCP techniques alone are not enough to decrease the constant 0.4228 beyond 0.2563.

Therefore, we need new ideas to reduce the constant 0.4228 to an arbitrarily small one. To do so, we first investigate the combinatorial structure of the hyperplane  $H := \{x \in \mathbb{Z}^n : \langle w, x \rangle = v\}$ , assuming  $|H \cap \{0, 1\}^n| \geq 2^{(1-\varepsilon)n}$ . Afterwards we apply an argument similar to the UDCP connection from [2]. We formally describe our approach for proving Theorem 1.2 in Section 4.

Note that Theorem 1.2 enables us to lift Restriction (1): We may assume  $\beta(w) \leq 2^{(1-\varepsilon_m)n}$  where  $\varepsilon_m > 0$  depends on  $m$  since otherwise a simple dynamic programming algorithm that runs in  $\mathcal{O}(n \cdot m \cdot |w(2^{\{1,\dots,n\}})|^m)$  time will be fast enough for constant  $m$  (see Lemma 3.5).

### 1.1.2 New Algorithmic Ideas: Lifting Restriction (2)

As mentioned before, Restriction (2) is algorithmically useful because of the following reason: We aim to detect a solution  $S_1, \dots, S_m$  to the Bin Packing instance by listing all candidates for  $S^L := \bigcup_{j \in L} S_j$  for some  $L \subseteq \{1, \dots, m\}$ , and Restriction (2) implies that  $w(S^L) = \sum_{j \in L} c_j$ . This allows us to narrow down the number of candidates to  $2^{(1-\varepsilon)n}$  by Restriction (1) (we explain in §1.1.3 why this is useful). Note this even narrows down the number of candidates for  $S^L$  if all bins have polynomially bounded *slack*, i.e.,  $c_j - w(S_j) \leq \text{poly}(n)$  since the number of possibilities of  $w(S^L)$  is only  $\text{poly}(n)$  as  $m = \mathcal{O}(1)$ .

But generally this strategy does not work whenever a bin has a large *slack*, that is when  $c_j - w(S_j)$  is large. While reductions in several similar situations were able to turn inequalities into equalities via general rounding techniques (such as [47, 54]), we need a more sophisticated method in this paper to deal with this issue: The idea of [47] is to divide the weights by roughly  $c_j - w(S_j)$  and (conservatively) round to an integer. In this case, the bin  $j$  has small slack with respect to the rounded weight function. The major complication however is that for different bins we would then need to work with differently rounded weight functions, which still does not allow us to narrow down the number of options for  $w(S^L)$  and hence (via Restriction (1)) the number of candidates for  $S^L$ .

<sup>4</sup>Here  $w(i)$  should be interpreted as the  $i$ 'th row of  $w$ , so it is a  $n$ -dimensional binary vector for every  $i$ . Then by the rank-nullity theorem  $|w(2^{\{1,\dots,n\}})| = 2^{\text{rk}_2(w)}$  and  $\beta(w) = 2^{n - \text{rk}_2(w)}$ , where  $\text{rk}_2$  is the rank over  $\mathbb{F}_2^n$ .

<sup>5</sup>See Section 2 for the formal definition of the  $\mathcal{O}(\cdot), \mathcal{O}^*(\cdot)$  and  $\mathcal{O}_{\varepsilon \rightarrow 0}(\cdot)$  notation.

Instead, for an integer  $k$  we work with a rounded version  $w_k$  of weights  $w$  where  $w_k(i)$  is obtained from  $w(i)$  by only keeping the  $k$  most significant bits. We will show we can choose integer `crit` (which we call *critical pruner*) such that  $|w_{\text{crit}}(2^{\{1, \dots, n\}})| \approx 2^{\delta n}$ , for some parameter  $\delta$  that depends on  $m$ . We will deal with the bins in two different ways depending on whether it has large slack (i.e. its slack is at least approximately  $n \cdot 2^{\ell - \text{crit}}$ , assuming all weights are  $\ell$ -bit integers) or not:

- **Large Slack Bins:** Our approach for such bins is loosely inspired by rounding approximation algorithms, e.g. the FPTAS for Knapsack (see e.g. [38, Section 11.8]). Observe that if some bin has large slack, we can split it into two parts, and we only need to keep track of the rounded weight of these parts in order to verify whether they indeed jointly fit into the bin. Because we assumed the *upper bound*  $|w_{\text{crit}}(2^{\{1, \dots, n\}})| \leq 2^{\delta n} \cdot \text{poly}(n)$  we can afford to keep track of all combinations of rounded weights as long as  $\delta < 1/m$ .
- **Small Slack Bins:** We deal with all small slack bins jointly by considering a split of the bins  $(L, R)$  such that and all bins in  $L$  have small slack. Now we use the *lower bound*  $|w_{\text{crit}}(2^{\{1, \dots, n\}})| \geq 2^{\delta n} / \text{poly}(n)$  and our additive combinatorics result guarantees  $\beta(w_{\text{crit}}) \leq 2^{(1-\varepsilon(\delta))n}$  for some  $\varepsilon(\delta) > 0$ . Now, we use the fact that all bins have small slack. Note, that there are only  $\text{poly}(n)$  candidates for  $w_{\text{crit}}(S^L)$  and therefore there are at most  $\mathcal{O}^*(2^{(1-\varepsilon(\delta))n})$  candidates for  $S^L$ , which can be algorithmically exploited.

In this informal discussion, we omitted several nontrivial technical issues. In particular, to deal with instances with *both* a substantial number of small slack bins and large slack bins, we need to distinguish several additional cases. Due to the subtle technical issues, we need to deal with each one of them in slightly different ways. Details are postponed to Section 3.

### 1.1.3 Solving Instances that Satisfy Restriction (1) and Restriction (2).

We now discuss how the methods from [8, 46] can be used to solve all instances that satisfy Restriction (1) and Restriction (2) in  $\mathcal{O}^*(2^{(1-\sigma_m)n})$  time for some  $\sigma_m > 0$ . An important subroutine from [8] is an algorithm that, given a set family  $\mathcal{W} \subseteq 2^{\{1, \dots, n\}}$  and set of bins  $L$ , computes for all  $W \in \mathcal{W}$  whether the items in  $W$  can be divided among the bins in  $L$ . That is, it computes whether  $W$  can be a candidate for  $S^L = \bigcup_{j \in L} S_j$ . The running time of this algorithm is  $\mathcal{O}(|\downarrow \mathcal{W}|n)$ , where  $\downarrow \mathcal{W} := \{X \subseteq W : W \in \mathcal{W}\}$  is defined as the *down-closure* of  $\mathcal{W}$ . The analogous *up-closure* of all supersets of elements from  $\mathcal{W}$  is denoted with  $\uparrow \mathcal{W}$ . Let us fix a solution  $(S_1, \dots, S_m)$ . We consider two cases based on how ‘balanced’<sup>6</sup> a solution is, with respect to a small parameter  $0 \leq \alpha \leq 1/2$ :

**Case 1:** There exists an integer  $b \in \{1, \dots, m\}$  such that  $|n/2 - \sum_{j=1}^b |S_j|| \leq \alpha n$ . In this case, observe  $\bigcup_{j=1}^b S_j$  is an element of

$$\mathcal{W} := \left\{ Y \subseteq \{1, \dots, n\} \text{ such that } w(Y) = \sum_{j=1}^b c_j \text{ and } \left| \frac{n}{2} - |Y| \right| \leq \alpha n \right\}.$$

Moreover, by Restriction (1) we have  $|\mathcal{W}| \leq \beta(w) \leq 2^{(1-\varepsilon)n}$ . This means that we can enumerate  $\mathcal{W}$  in essentially  $2^{(1-\varepsilon)n}$  time, because we will present an  $\mathcal{O}((|\downarrow \mathcal{W}| + |\uparrow \mathcal{W}|)n)$  time algorithm that for each  $W \in \mathcal{W}$  computes whether  $W$  can be divided among bins  $1, \dots, b$  and  $\{1, \dots, n\} \setminus W$  among bins  $b+1, \dots, m$  (this algorithm is based on techniques from [8]). This will detect a solution if it exists. We bound the running time using the property  $n/2 - \alpha n \leq |W| \leq n/2 + \alpha n$ . In this case

<sup>6</sup>The actual definition of  $\alpha$ -balancedness (Definition 3.4) will be independent of the ordering of the bins.

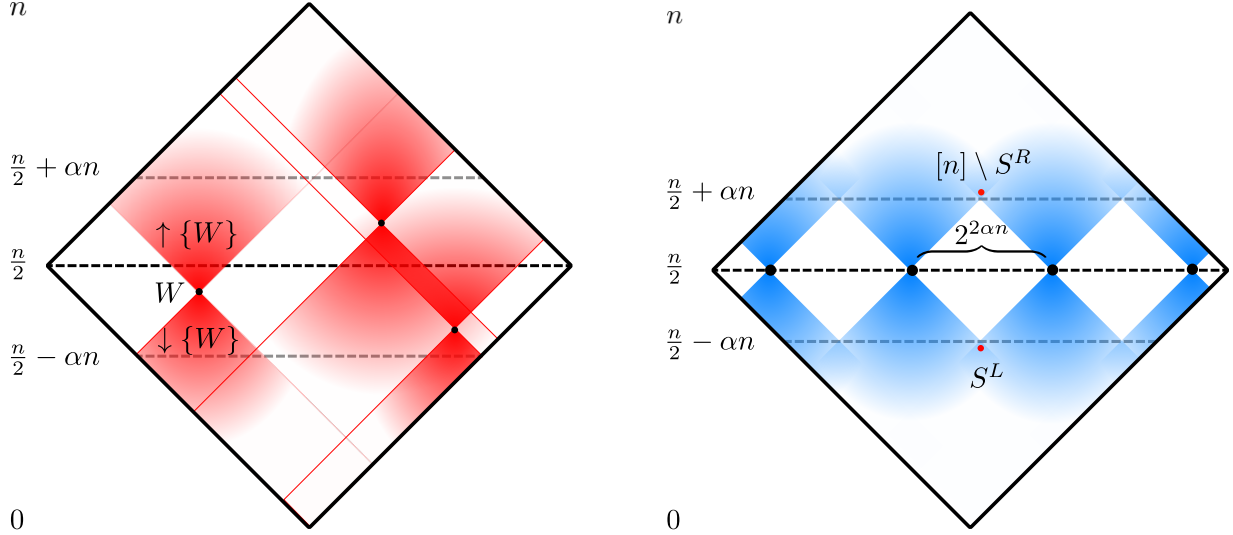


Figure 1: Schematic view of the algorithm from §1.1.3 . A point in the square represents a set in  $2^{\{1, \dots, n\}}$ . The vertical axis corresponds to the cardinality of this set (e.g., longest horizontal line represents all sets in  $\binom{\{1, \dots, n\}}{n/2}$ ). The left figure illustrates the analysis for the case when there exists an  $\alpha$ -balanced solution  $W \subseteq \{1, \dots, n\}$  with  $|n/2 - |W|| \leq \alpha n$ . We iterate through all  $W$  in time proportional to area the of the colored region. The right figure illustrates the case of an  $\alpha$ -unbalanced solution. A division of the solution ( $L, R$ ) is witnessed by the roughly  $2^{2\alpha n}$  sets  $W$  in  $\binom{\{1, \dots, n\}}{n/2}$  satisfying  $S^L \subseteq W \subseteq \{1, \dots, n\} \setminus S^R$ .

we will show  $|\downarrow \mathcal{W}| + |\uparrow \mathcal{W}| \leq 2^{(1-\varepsilon')n}$  and hence the algorithm is fast enough (see left Figure 1 for an illustration).

**Case 2:** For every  $b \in \{1, \dots, m\}$  we have  $|n/2 - \sum_{j=1}^b |S_j| > \alpha n|$ . Here we can use a method from [46]: We let  $\mathcal{W}$  consist of  $2^{(1-2\alpha)n}$  independently sampled subsets of  $\{1, \dots, n\}$  of cardinality  $n/2$ . We answer *yes* if there exist  $W \in \mathcal{W}$ , disjoint sets  $S^L = S'_1, \dots, S'_{b-1} \subseteq W$  and  $S^R = S'_{b+1}, \dots, S'_m \subseteq \{1, \dots, n\} \setminus W$  such that  $w(S'_j) = c_j$  for all  $j \in \{1, \dots, m\} \setminus \{b\}$ . This condition can also be computed in  $\mathcal{O}((|\downarrow \mathcal{W}| + |\uparrow \mathcal{W}|)n)$  time by the methods of [8]. The crux is that both conditions together imply our instance is a *yes*-instance, since the remaining elements have total weight  $c_b$  by Restriction (2). Moreover, by the balancedness assumption at least  $2^{2\alpha n}$  sets  $W \subseteq \{1, \dots, n\}$  with the above conditions exist. Therefore the random sampling will include such a  $W$  with good probability (see right Figure 1 for an illustration).

## 1.2 Related Work

**Littlewood–Offord, UDCP’s, and Exponential Time Algorithms.** Two sets  $A, B \subseteq \{0, 1\}^n$  form a Uniquely Decodable Code Pair (UDCP) if  $|A + B| = |A| \cdot |B|$ , where  $A + B := \{a + b : a \in A, b \in B\}$  (and addition is in  $\mathbb{Z}^n$ ). The maximal sizes of UDCP’s have been very well studied in information theory. See e.g. [50, Section 3.5.1] for a (not so recent) overview. Two record upper bounds are  $|A| \cdot |B| \leq 2^{1.5n}$  (from [53]) and  $|A| \leq 2^{(0.4228 + \sqrt{\varepsilon})n}$  whenever  $|B| \leq 2^{(1-\varepsilon)n}$  (from [4]). The study of UDCP’s is relevant for this paper by the following connection shown in [3]: For any vector  $w \in \mathbb{Z}^n$ , there is a UDCP  $A, B \subseteq \{0, 1\}^n$  such that  $|A| = |w(2^{\{1, \dots, n\}})|$  and  $|B| = \beta(w)$ .

A study of the trade-off between the parameters  $|w(2^{\{1, \dots, n\}})|$  and  $\beta(w)$  was already fruitful for

obtaining improved exponential time algorithms in two earlier papers in the context of the Subset Sum problem. In this problem one is given  $w \in \mathbb{Z}^n$  and a target integer  $t$  and one needs to find a subset  $X \subseteq \{1, \dots, n\}$  such that  $w(X) = t$ . First, the aforementioned paper [3] combined their connection to UDCP's with the bound from [53] to show that instances of Subset Sum satisfying  $|w(2^{\{1, \dots, n\}})| \geq 2^{0.997n}$  can be solved in  $\mathcal{O}(2^{0.49991n})$  time, thereby improving the best  $\mathcal{O}^*(2^{n/2})$  worst case running time from [31] for these instances. Second, a slight variant of the trade-off was used in [5] to give a  $\mathcal{O}(2^{0.86n})$  time algorithm that uses polynomial space (assuming read-only access to the exponential number of random bits).

In a recent work by Jain et al. [32], the dependency of  $\delta$  on  $\varepsilon$  in Theorem 1.2 was improved to  $\delta(\varepsilon) = \mathcal{O}(\sqrt{\varepsilon})$ . As a corollary,  $\sigma_m$  in Theorem 1.1 can be bounded with  $\sigma_m = \Omega(m^{-12})$ .

**Exact Algorithms for Set Cover.** Question 1 was for the first time explicitly posed in [15], who showed that a negative answer to (a variant of) the question implies hardness in a fine-grained sense for the Subset Sum, Steiner Tree, and Connected Vertex Cover problems. A main motivation in [15] for posing the question was a curious reduction showing that there is no improved algorithm for counting the number of Set Cover solutions modulo 2 unless improved algorithms for CNF-Sat exist (i.e. the Strong Exponential Time Hypothesis fails). Later the assumption that no improved algorithm exists was dubbed as ‘Set Cover Conjecture’ (see e.g. [16, Conjecture 14.36]). Since then, the conjecture has been used in several works, e.g. in [1, 40].

On the positive side, (especially for this work) important algorithmic tools were developed in [9]: Fast zeta and Möbius transformations were introduced in the area of exponential time algorithms to show that Set Cover can be solved in  $2^n \cdot \text{poly}(n)$  time even when the number of sets in the input is exponential in  $n$ . One major consequence was a  $2^n \cdot \text{poly}(n)$  time algorithm for computing whether an input graph on  $n$  vertices has a proper coloring with  $k$  colors. While for  $k \leq 6$  faster algorithms exist [56] this is still the fastest algorithm for  $k > 6$ .

Improved algorithms for solving Set Cover instances of sets with bounded cardinality were given in [39]. Later, this was generalized to improved algorithms for Set Cover instances where the optimum is linear in the universe size [46]. Other instances that allow improved algorithms were also presented in e.g. [26].

**Exact Algorithms for Bin Packing.** In a textbook on exact exponential time algorithms, it was shown that Bin Packing can be solved in time  $\mathcal{O}(n2^n \cdot \max_i \{w(i)\})$  time [23, Section 4.2.3]. A faster algorithm  $\mathcal{O}^*(2^n)$  time algorithm was given in [9]. Even faster algorithms were given for  $m = 2, 3$  in [41].

In [25] it was shown that Bin Packing can be solved in polynomial time if there are only a constant number of distinct weights of items. Jansen et al. [33] study Bin Packing with a constant number of bins and bounded items weights was studied. They presented a dynamic programming algorithm (similar to the one proposed by us in Lemma 3.5) and show that it runs in time  $n^{\mathcal{O}(m)}$  if the item weights are polynomial in  $n$ . This running time cannot be improved to  $n^{\mathcal{O}(m/\log m)}$ , unless the Exponential Time Hypothesis fails [33].

**Heuristics for Bin Packing.** The applications and combinatorial properties of Bin Packing have been studied since the 1930's [36]. To the best of our knowledge, the first attempt to exactly solve Bin Packing with the assistance of the modern computer was developed in the fifties by Eisemann [20], with the motivation to trim losses in cutting rolls of paper. Starting from the seventies, the research on exact algorithms for Bin Packing focused on the branch-and-bound technique proposed by Eilon

and Christofides [19]. These heuristics work great in practice. Nevertheless, there are no theoretical guarantees on their worst-case performance.

For a modern survey and experimental evaluations of the available software see [43, 17].

**Approximation Algorithms for Bin Packing.** Bin Packing is one of the problems that initiated the study of approximation algorithms. The earliest one is the *First Fit* algorithm analysed by Johnson [34] that requires at most  $1.7 \cdot \text{OPT} + 1$  bins. The major breakthrough was done by Karmarkar-Karp [37] who provided a polynomial time algorithm that requires at most  $\text{OPT} + \mathcal{O}(\log^2(\text{OPT}))$  bins. Recently, a big leap forward was done by Rothvoß [48] who gave a polynomial time algorithm that requires only  $\text{OPT} + \mathcal{O}(\log(\text{OPT}) \log \log(\text{OPT}))$  bins and Hoberg and Rothvoß [30] who improved this even further to  $\text{OPT} + \mathcal{O}(\log(\text{OPT}))$  bins.

### 1.3 Organization

This paper is organized as follows: In Section 2 we present some preliminaries and introduce some notations. In Section 3 we present the algorithm and proof of our main theorem, assuming Theorem 1.2. The latter theorem is proved in the next two Sections 4 and 5. In Appendix A we include the proofs of technical Lemmas from Section 3. In Appendix B we include the proofs of useful inequalities regarding binary entropy.

## 2 Preliminaries

Throughout the paper, we use the  $\mathcal{O}^*$  notation to hide polynomial factors in the input size. The notation  $\tilde{\mathcal{O}}(T)$  means  $\mathcal{O}(T \cdot \text{polylog}(T))$ . The number of bins is assumed to be constant, i.e.  $m = \mathcal{O}(1)$ . We say a function  $f(\varepsilon) = \mathcal{O}_{\varepsilon \rightarrow 0}(g(\varepsilon))$  if there exists a positive number  $C$  and sufficiently small  $\varepsilon_0 > 0$ , such that  $|f(\varepsilon)| \leq C \cdot g(\varepsilon)$  for all  $\varepsilon < \varepsilon_0$ . We use  $\Omega_{\varepsilon \rightarrow 0}$  similarly to express lower bounds. Finally  $\text{poly}(n)$  is a shorthand notation for  $n^{\mathcal{O}(1)}$ . All the logarithms are base 2 unless stated otherwise.

In this paper, we assume that basic arithmetic operations take constant time. We use a result of Frank and Tardos [24], in a similar way to [21], to assume that  $\max_i \{\log w(i)\} \leq \text{poly}(n)$ .

If  $a, b \in \mathbb{R}$  and  $b \geq 0$  we let  $[a \pm b]$  denote the interval  $[a - b, a + b]$ . If  $A$  and  $B$  are sets, we denote by  $B^A$  the set of vectors indexed by  $A$  with values from  $B$ , and we will interchangeably address these vectors as functions from  $A$  to  $B$ . If  $f \in B^A$  and  $b \in B$  we denote  $f^{-1}(b) := \{a \in A : f(a) = b\}$  for its inverse evaluated at  $b$ . For example, when  $x \in \{0, 1\}^{\{1, \dots, n\}}$  then  $x^{-1}(0) = \{i \in \{1, \dots, n\} : x(i) = 0\}$ . If  $x, y \in \mathbb{R}^A$  we denote  $\langle x, y \rangle := \sum_{a \in A} x_a \cdot y_a$  for their inner product.

To quickly refer to the properties of a solution of a Bin Packing instance we use the following notations: The function  $w$  indicates the weights of the input. It is extended to sets  $X \subseteq \{1, \dots, n\}$  by defining  $w(X) := \sum_{i \in X} w(i)$  and to set families  $\mathcal{F} \subseteq 2^{\{1, \dots, n\}}$  by defining  $w(\mathcal{F}) := \{w(X) : X \in \mathcal{F}\}$ . We say a set  $X \subseteq \{1, \dots, n\}$  of items can be *divided* over bins  $L \subseteq \{1, \dots, m\}$  if there is a partition  $X_1, \dots, X_{|L|}$  of  $X$ , such that for all  $j \in \{1, \dots, |L|\}$ , the set  $X_j$  can be placed in bin  $j$ , i.e.,  $w(X_j) \leq c_j$ .

We abstract a simple probabilistic argument that we use several times in our algorithms. It follows easily by observing that a random element from  $X$  is in  $Y$  with probability  $|X|/|Y|$ .

**Observation 2.1.** *Let  $U$  be any universe set and let  $Y \subseteq U$  be an arbitrary nonempty subset of  $U$ . Let  $Z$  be a set obtained by sampling (with replacement)  $\lceil |U|/|Y| \rceil$  times uniformly at random from  $U$  (if at the end of this process an element repeats, we take a single occurrence of this element). Then  $\Pr[Z \cap Y \neq \emptyset] \geq 1 - \frac{1}{e}$ .*



## 2.1 Preliminary Tools: Fast Transformations

Our algorithm will crucially rely on the following algorithmic tools and definitions from [8].

**Definition 2.2** (Zeta and Möbius Transform). *Let  $f : 2^U \rightarrow \mathbb{N}$ . Then the zeta transform  $\zeta f$  and Möbius transform  $\mu f$  are functions from  $2^U$  to  $\mathbb{N}$  such that for every  $X \subseteq U$ :*

$$(\zeta f)(X) := \sum_{Y \subseteq X} f(Y) \quad \text{and} \quad (\mu f)(X) := \sum_{Y \subseteq X} (-1)^{|U \setminus Y|} f(Y).$$

**Definition 2.3.** *Given  $\mathcal{S} \subseteq 2^U$ , the down-closure  $\downarrow \mathcal{S}$  and up-closure  $\uparrow \mathcal{S}$  are defined as follows:*

$$\downarrow \mathcal{S} := \{X : \exists S \in \mathcal{S} \text{ such that } X \subseteq S\} \quad \text{and} \quad \uparrow \mathcal{S} := \{X : \exists S \in \mathcal{S} \text{ such that } X \supseteq S\}.$$

**Theorem 2.4** (Fast zeta and Möbius transform [8]). *Suppose that  $f : 2^U \rightarrow \mathbb{N}$  is such that  $f(X)$  can be evaluated in  $T$  time for any given  $X \subseteq U$ , and let  $\mathcal{S} \subseteq 2^U$  be a set family. There is an algorithm that can compute for every  $X \in \downarrow \mathcal{S}$  the values  $(\zeta f)(X)$  and  $(\mu f)(X)$ . The algorithm runs in  $\mathcal{O}(|\downarrow \mathcal{S}| \cdot |U| \cdot T)$  time.*

**Definition 2.5** (Cover and Entry-Wise Product). *Given  $f, g : 2^U \rightarrow \mathbb{N}$ , the cover product  $f \otimes g = h$  and the entry-wise product  $f \cdot g = h'$  are the functions  $h, h' : 2^U \rightarrow \mathbb{N}$  such that*

$$h(Z) := \sum_{X \cup Y = Z} f(X)g(Y) \quad \text{and} \quad h'(Z) := f(Z) \cdot g(Z).$$

**Theorem 2.6** ([7]).  $\mu((\zeta f) \cdot (\zeta g)) = f \otimes g$ .

**Theorem 2.7.** *Suppose that we have a Bin Packing instance with bin capacities  $c_1, \dots, c_m$  and item weight function  $w$ . Then for any  $B \subseteq \{1, \dots, m\}$  and set  $\mathcal{W} \subseteq 2^{\{1, \dots, n\}}$ , computing for all  $X \in \downarrow \mathcal{W}$  whether  $X$  can be divided over the bins in  $B$  can be done in time  $\mathcal{O}(|\downarrow \mathcal{W}|n)$ . Similarly, for any  $B \subseteq \{1, \dots, m\}$  and set  $\mathcal{W} \subseteq 2^{\{1, \dots, n\}}$ , computing for all  $X \in \uparrow \mathcal{W}$  whether  $\{1, \dots, n\} \setminus X$  can be divided over the bins in  $B$  can be done in time  $\mathcal{O}(|\uparrow \mathcal{W}|n)$ .*

*Proof.* For all  $j = 1, \dots, m$  define a function  $f_j : 2^{\{1, \dots, n\}} \rightarrow \{0, 1\}$  as

$$f_j(X) = \begin{cases} 1, & \text{if } w(X) \leq c_j \\ 0, & \text{otherwise.} \end{cases}$$

Assume without loss of generality that  $B = \{1, \dots, d\}$ . Notice that  $X$  can be divided over the bins in  $B$  if and only if  $(f_1 \otimes f_2 \otimes \dots \otimes f_d)(X) > 0$ . By Theorem 2.6 we have that

$$f_1 \otimes f_2 \otimes \dots \otimes f_d = \mu((\zeta f_1) \cdot (\zeta f_2) \cdot \dots \cdot (\zeta f_d)).$$

Then, the right hand side can be computed in  $\mathcal{O}(|\downarrow \mathcal{W}|)$  time using subsequently fast  $d$  zeta transformation (Theorem 2.4), naïve entry-wise product computation, and one fast Möbius transformation (Theorem 2.4). The proof for the second part of the theorem one takes  $\mathcal{W}' := \{\{1, \dots, n\} \setminus W : W \in \mathcal{W}\}$  and applies the technique above to  $\mathcal{W}'$ . Notice that indeed  $\downarrow \mathcal{W}' = \uparrow \mathcal{W}$ .  $\square$

Note this can be used to obtain the algorithm already mentioned in Section 1. We include the proof to introduce the reader to the state-of-the-art algorithm which will be expanded in the later sections.

**Theorem 2.8** ([8]). *Bin Packing with capacities  $c_1, \dots, c_m$  can be solved in  $\mathcal{O}^*(2^n)$  time.*

*Proof.* For  $i = 1, \dots, m$  define the function  $f_i : 2^{\{1, \dots, n\}} \rightarrow \{0, 1\}$  as

$$f_i(X) = \begin{cases} 1, & \text{if } w(X) \leq c_i \\ 0, & \text{otherwise.} \end{cases}$$

Note that  $(f_1 \otimes f_2 \otimes \dots \otimes f_m)(\{1, \dots, n\}) > 0$  if and only if the answer to Bin Packing is positive. By Theorem 2.6 we have that

$$f_1 \otimes f_2 \otimes \dots \otimes f_m = \mu((\zeta f_1) \cdot (\zeta f_2) \dots (\zeta \cdot f_m)),$$

and the right hand side can be computed in  $\mathcal{O}^*(2^n)$  time using subsequently fast  $m$  zeta transformations (Theorem 2.4), naïve entry-wise product computation, and one fast Möbius transformation (Theorem 2.4).  $\square$

## 2.2 The Entropy Function and Binomial Coefficients

We heavily use properties of the entropy function, which we will now define. For a discrete probability space  $\mathcal{D} = (\Omega, p)$  where  $p : \Omega \rightarrow (0, 1)$ , the *entropy* of  $\mathcal{D}$  is defined as follows:

$$h(\mathcal{D}) := - \sum_{x \in \Omega} p(x) \log p(x). \quad (1)$$

We say  $p = (p_1, \dots, p_k)$  is a probability vector if the  $p_i$ 's are non-negative and satisfy  $\sum_{i=1}^k p_i = 1$ . If no underlying probability space is given, we may interpret  $p$  as a probability measure over  $\{1, \dots, k\}$  and thus (1) gives  $h(p) = - \sum_{i=1}^k p_i \log p_i$ . The *support* of the vector  $p = (p_1, \dots, p_k)$  the set of its non-zero coordinates and the *size* of the support is the number of non-zero coordinates of  $p$ . If  $p \in (0, 1)$ , we use the shorthand notation  $h(p) := h(p, 1 - p)$ . The multinomial coefficient  $\binom{n}{p_1 n, p_2 n, \dots, p_k n}$  can be approximated with  $h(p)$  as follows:

**Lemma 2.9** ([13], Lemma 2.2). *If  $p$  is a probability vector with support of size at most  $s$ , then*

$$\binom{n + s - 1}{s - 1}^{-1} 2^{h(p)n} \leq \binom{n}{p_1 n, \dots, p_k n} \leq 2^{h(p)n}.$$

We will frequently use the special case  $\binom{n}{pn} \leq 2^{h(p)n}$  when  $p \in (0, 1)$ .

The following lemma states the intuitive fact that close probability vectors have close entropy.

**Lemma 2.10.** *Let  $p, q \in \mathbb{R}^k$  be probability vectors such that  $|p_i - q_i| \leq \varepsilon$  for each  $i = 1, \dots, k$ . Then  $|h(p) - h(q)| \leq \frac{1}{\ln(2)} \cdot \varepsilon k \log \frac{1}{\varepsilon}$ .*

*Proof.* Recall that  $h(p) = \sum_{i=1}^k p_i \log \frac{1}{p_i}$ . Thus the lemma follows by applying the following inequalities to all summands of the entropy of  $p$  and  $q$ : When  $x, \varepsilon, x + \varepsilon \in [0, 1]$ , then

$$x \log \frac{1}{x} - \frac{\varepsilon}{\ln(2)} \leq (x + \varepsilon) \log \frac{1}{x + \varepsilon} \leq x \log \frac{1}{x} + \varepsilon \log \frac{1}{\varepsilon}.$$

The second inequality is direct, and the first inequality can be derived as

$$(x + \varepsilon) \log \frac{1}{x + \varepsilon} = x \log \frac{1}{x} + x \log \frac{x}{x + \varepsilon} + \varepsilon \log \frac{1}{x + \varepsilon} \geq x \log \frac{1}{x} - x \log(1 + \frac{\varepsilon}{x}) \geq x \log \frac{1}{x} - \frac{1}{\ln(2)} \cdot \varepsilon,$$

where in the last inequality we use the standard fact that for every  $z \in \mathbb{R}_+$  it holds that  $1 + z \leq \exp(z)$ .  $\square$

### 3 Proof of Theorem 1.1

In this section we prove our main theorem which we first restate for convenience:

**Theorem 3.1.** *For every  $m \in \mathbb{N}$  there is a constant  $\sigma_m > 0$  such that every Bin Packing instance with  $m$  bins can be solved in  $\mathcal{O}(2^{(1-\sigma_m)n})$  time with high probability.*

For the proof of Theorem 1.1, we combine four lemmas, each solving particular types of instances (see Figure 2 for an overview of the algorithm). We will refer to these different types of instances by Case A, Case B, Case C and Case D.

This section is organized as follows: In Subsection 3.1 we introduce definitions that will be used throughout this section, such as the key definition of  $\alpha$ -balanced solutions. We then prove in Subsections 3.2 and 3.3 that ‘easy’ instances of Bin Packing, namely those where  $w$  generates relatively few distinct sums (Case A) and those with  $\alpha$ -unbalanced solutions for some  $\alpha > 0$  (Case B), can be solved fast. We can therefore assume that there are only  $\alpha$ -balanced solutions and that  $|w(2^{\{1, \dots, n\}})| > 2^{\delta n}$  (for some  $\delta < \frac{1}{m}$ ) in the rest of the section.

Subsection 3.4 introduces a few more definitions, such as the ‘slack of a bin’, which is the vacant capacity of a bin in a solution. This is also where we define the ‘crit-pruned item weights’ as the bit representation of the weights, pruned to the  $\text{crit} \in \mathbb{N}$  most significant bits. The parameter  $\text{crit}$  is then chosen such that  $|w_{\text{crit}}(2^{\{1, \dots, n\}})| = \Theta(n2^{\delta n})$ , as discussed in § 1.1.2. These definitions will be central in solving the remaining two types of instances.

In Subsection 3.5, we consider instances where at least roughly half of the items are in a bin with small slack (Case C). This is where we use the approach discussed in § 1.1.1 and apply Theorem 1.2 on the  $\text{crit}$ -pruned item weights, to conclude that  $\beta(w_{\text{crit}}) \leq 2^{(1-\epsilon)n}$  for some  $\epsilon > 0$ .

Subsection 3.6 then solves instances where at least roughly half of the items are in a bin with large slack (Case D). In the proof, we can split the large slack bins into two parts, where we use the  $\text{crit}$ -pruned item weights in each of these parts to determine whether they fit. Because  $|w_{\text{crit}}(2^{\{1, \dots, n\}})|$  is of order  $2^{\delta n}$ , there is of order  $2^{\delta mn}$  different tuples of weights. This we can keep track of since we assumed  $\delta < \frac{1}{m}$ . Furthermore, we split the large slack bins into two parts. This operation guarantees (that with a constant probability) we can correctly guess the partition of the items in small slack bins into two parts.

Finally, the proof of Theorem 1.1 can be found in Subsection 3.7, where we combine all these results by selecting the appropriate values for  $\delta$  and  $\alpha$  based on the number of bins (see Figure 2 for overview of the algorithm).

#### 3.1 Balanced Solutions and Witnesses

Fix an instance of Bin Packing. We begin by formally defining solutions.

**Definition 3.2** (Solution). *A partition  $S_1, \dots, S_m$  of  $\{1, \dots, n\}$  is a solution of an instance of Bin Packing with  $n$  items and  $m$  bins, if for all  $j \in \{1, \dots, m\}$  the set  $S_j$  fits into the bin  $j$  (i.e.,  $w(S_j) \leq c_j$ ).*

The following notion of a *witness* will be crucial in our approach.

**Definition 3.3** ( $(L, R)$ -witnesses). *Let  $L, R \subseteq \{1, \dots, m\}$ ,  $L \cap R = \emptyset$ . A set  $W \subseteq \{1, \dots, n\}$  is an  $(L, R)$ -witness if there is a solution  $S_1, \dots, S_m$  such that*

$$\bigcup_{j \in L} S_j \subseteq W \quad \text{and} \quad \bigcup_{j \in R} S_j \subseteq \{1, \dots, n\} \setminus W.$$

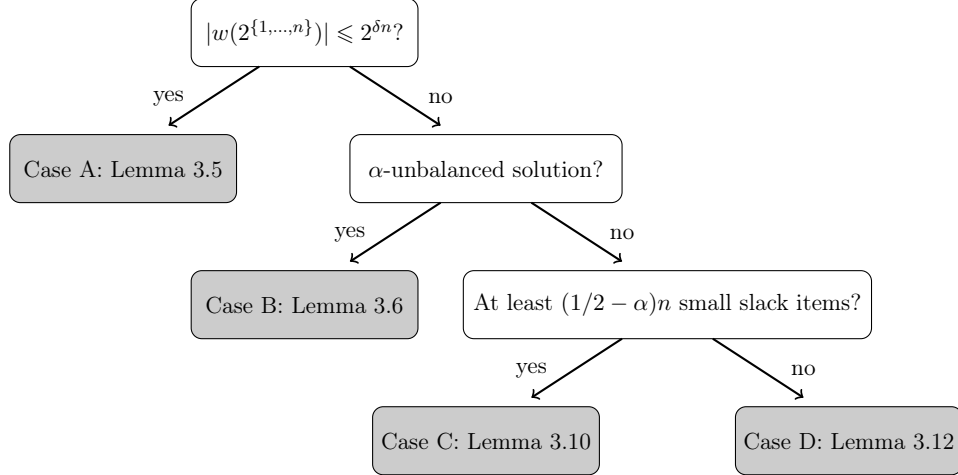


Figure 2: Overview of use of Lemmas proving Theorem 1.1.

We commonly denote  $S^L := \bigcup_{j \in L} S_j$  and  $S^R := \bigcup_{j \in R} S_j$ . Observe that to verify that a set  $W \subseteq \{1, \dots, n\}$  is an  $(L, R)$ -witness, it is sufficient to find  $S^L \subseteq W$  and  $S^R \subseteq \{1, \dots, n\} \setminus W$  with the following properties: (i) the items in  $S^L$  can be distributed to the bins in  $L$ , (ii) items  $S^R$  can be distributed to the bins in  $R$ , and (iii) items  $\{1, \dots, n\} \setminus (S^L \cup S^R)$  can be distributed to the bins in  $\{1, \dots, m\} \setminus (L \cup R)$ . Hence, finding a witness gives us a ‘certificate’ for the existence of a solution. This will be used several times throughout this section.

Our algorithmic approach will heavily depend on whether or not the set of items can be evenly divided, which we formalize as follows:

**Definition 3.4** ( $\alpha$ -balanced solution). *Let  $S_1, \dots, S_m$  be a solution of Bin Packing. The solution is  $\alpha$ -balanced if for all permutations  $\pi : \{1, \dots, m\} \rightarrow \{1, \dots, m\}$  there exists an index  $b \in \{1, \dots, m\}$  such that  $\sum_{j=1}^b |S_{\pi(j)}| \in [n/2 \pm \alpha n]$ . If a solution is not  $\alpha$ -balanced, it is called  $\alpha$ -unbalanced.*

Hence, a solution is  $\alpha$ -unbalanced if and only if there exists a permutation  $\pi : \{1, \dots, m\} \rightarrow \{1, \dots, m\}$  and a  $b \in \{1, \dots, m\}$  such that  $\sum_{j=1}^{b-1} |S_{\pi(j)}| < (1/2 - \alpha)n$  and  $\sum_{j=1}^b |S_{\pi(j)}| > (1/2 + \alpha)n$ .

### 3.2 Solving Case A: Few Distinct Sums

If the instance generates relatively few distinct sums in the sense that  $|w(2^{\{1, \dots, n\}})| \leq 2^{\delta n}$  for some small  $\delta < 1/m$ , we can solve Bin Packing sufficiently fast. This is the algorithm we use in Case A.

**In Case A:** (✓)  $|w(2^{\{1, \dots, n\}})| \leq 2^{\delta n}$

**Lemma 3.5.** *A solution of Bin Packing can be found in  $\mathcal{O}(n \cdot m \cdot |w(2^{\{1, \dots, n\}})|^m)$  time.*

*Proof.* As a first step, we compute the set  $w(2^{\{1, \dots, n\}})$  in  $\mathcal{O}(n \cdot |w(2^{\{1, \dots, n\}})|)$  time with Lemma A.1. Next, we use the following dynamic programming algorithm:

For every  $i \in \{0, \dots, n\}$  and  $a_1, \dots, a_m \in w(2^{\{1, \dots, n\}})$ , define

$$\text{DP}_i[a_1, \dots, a_m] := \begin{cases} \text{true} & \text{if items } \{1, \dots, i\} \text{ can be distributed to} \\ & \text{bins with capacities } a_1, \dots, a_m \text{ exactly,} \\ \text{false} & \text{otherwise.} \end{cases}$$

We initiate a dynamic programming table with  $\text{DP}_0[0, \dots, 0] = \text{true}$  and the remaining entries of  $\text{DP}_0$  are set to **false**, i.e.,  $\text{DP}_0[a_1, \dots, a_m] = \text{false}$  if there exists  $j \in \{1, \dots, m\}$  with  $a_j \neq 0$ . Then the following recurrence relation holds for every  $i \in \{1, \dots, n\}$  and  $a_1, \dots, a_m \in w(2^{\{1, \dots, n\}})$ :

$$\text{DP}_i[a_1, \dots, a_m] = \bigvee_{j \in \{1, \dots, m\}} \text{DP}_{i-1}[a_1, \dots, a_j - w(i), \dots, a_m].$$

This concludes the description of the dynamic programming procedure. Observe that in the above recursion we only need to keep track of **true** entries of table  $\text{DP}_i$  for every  $i \in \{1, \dots, n\}$ . Hence, we only need to check entries with  $a_1, \dots, a_m \in w(2^{\{1, \dots, n\}})$ , since those are all the possible sums that  $w$  generates. Therefore, we have  $\text{DP}_i(a_1, \dots, a_m) = \text{false}$  whenever  $a_j \notin w(2^{\{1, \dots, n\}})$  for some  $j \in \{1, \dots, n\}$ . Since each entry of  $\text{DP}_i$  table can be computed in  $\mathcal{O}(m)$  time, the running time follows.

Finally, observe that when  $c_1, \dots, c_m$  are the capacities of the bins of the bin packing instance, then there exists  $a_1, \dots, a_m$  such that  $\text{DP}_n(a_1, \dots, a_m) = \text{true}$  and  $a_j \leq c_j$  for all  $j \in \{1, \dots, m\}$  if and only if there is a solution to bin packing instance. This condition can be verified by scanning  $\mathcal{O}(w(2^{\{1, \dots, n\}}))$  many **true**-entries of table  $\text{DP}_n$ .  $\square$

### 3.3 Solving Case B: Unbalanced Solutions

Next, we show that  $\alpha$ -unbalanced solutions (for some  $\alpha > 0$ ) can be detected quickly, i.e. we solve the instances in Case B.

**In Case B:** (✓) The instance has an  $\alpha$ -unbalanced solution (for some  $\alpha > 0$ ).

Note that we might also add the assumption that  $|w(2^{\{1, \dots, n\}})| > 2^{\delta n}$  as Case A solves all other instances. However, we do not need this assumption for Case B, as we have the following result.

**Lemma 3.6.** *If a Bin Packing instance has an  $\alpha$ -unbalanced solution, then such a solution can be found in  $\mathcal{O}^*(2^{(1-f_B(\alpha))n})$  time with probability  $\geq 1/2$  where  $f_B(\alpha) = \Omega_{\alpha \rightarrow 0} \left( \frac{\alpha^2}{\log^2(\alpha)} \right)$ .*

*Proof.* The algorithm iterates over all subsets  $L, R \subseteq \{1, \dots, m\}$  such that  $L \cap R = \emptyset$ ,  $|L \cup R| = m - 1$ . Let  $b \in \{1, \dots, m\}$  be the only element not in  $L \cup R$ . For each such  $L$  and  $R$ , the algorithm will search for an  $(L, R)$ -witness of size  $\frac{n}{2}$ . Concretely, it samples a set  $\mathcal{W}$  of  $2^{(1-2\alpha)n}$  random subsets (with replacement, and removing copies afterward) of  $\{1, \dots, n\}$  of size  $\frac{n}{2}$ , and it computes for every  $W \in \mathcal{W}$  whether it is an  $(L, R)$ -witness as follows: First, it computes which sets from  $\downarrow \mathcal{W} \cup \uparrow \mathcal{W}$  are potential candidates for  $S^L$  and  $S^R$ . This is done by computing the booleans  $l_X$  for every  $X \in \downarrow \mathcal{W}$  and  $r_X$  for every  $X \in \uparrow \mathcal{W}$ , where

$$l_X := \begin{cases} \text{true} & \text{if } X \text{ can be distributed to the bins in } L, \\ \text{false} & \text{otherwise,} \end{cases}$$

$$r_X := \begin{cases} \text{true} & \text{if } \{1, \dots, n\} \setminus X \text{ can be distributed to the bins in } R, \\ \text{false} & \text{otherwise.} \end{cases}$$

This can be done in time  $\mathcal{O}((|\downarrow \mathcal{W}| + |\uparrow \mathcal{W}|)n)$  using Theorem 2.7.

Second, for each  $W \in \mathcal{W}$ , we search for sets  $X^L \subseteq W$  and  $X^R \subseteq \{1, \dots, n\} \setminus W$  of maximum weight such that they can be distributed to the bins in  $L$  and  $R$  respectively. To do this, we compute  $l_X^*$  for every  $X \in \downarrow \mathcal{W}$  and  $r_X^*$  for every  $X \in \uparrow \mathcal{W}$ , where

$$l_X^* := \max_{Y \subseteq X: l_Y = \text{true}} w(Y), \quad r_X^* := \max_{Y \supseteq X: r_Y = \text{true}} w(\{1, \dots, n\} \setminus Y).$$

This can be done using dynamic programming with the recurrence relations

$$l_X^* = \begin{cases} w(X) & \text{if } l_X = \mathbf{true}, \\ \max_{i \in X} l_{X \setminus \{i\}}^* & \text{if } l_X = \mathbf{false}, \end{cases} \quad \text{and} \quad r_X^* = \begin{cases} w(\{1, \dots, n\} \setminus X) & \text{if } r_X = \mathbf{true}, \\ \max_{i \notin X} r_{X \cup \{i\}}^* & \text{if } r_X = \mathbf{false}. \end{cases}$$

The running time is only  $\mathcal{O}((|\downarrow\mathcal{W}| + |\uparrow\mathcal{W}|)n)$  since the values  $l_X^*$  for  $X \in \downarrow\mathcal{W}$  do not depend on entries  $l_Y^*$  for  $Y \notin \downarrow\mathcal{W}$ , and the values  $r_X^*$  for  $X \in \uparrow\mathcal{W}$  do not depend on entries  $r_Y^*$  for  $Y \notin \uparrow\mathcal{W}$ . Thus the algorithm only needs to evaluate  $|\downarrow\mathcal{W}| + |\uparrow\mathcal{W}|$  table entries which can be done in time  $\mathcal{O}(n)$  per entry.

Third, the algorithm checks if there exists a  $W \in \mathcal{W}$  such that  $w(\{1, \dots, n\}) - l_W^* - r_W^* \leq c_b$  and returns *yes* if this is the case. If for all different choices of  $L$  and  $R$ , no  $(L, R)$ -witness has been found, the algorithm returns *no*.

### Correctness of Algorithm

Assume that there is an  $\alpha$ -unbalanced solution  $S_1, \dots, S_m$ . Let  $\pi : \{1, \dots, m\} \rightarrow \{1, \dots, m\}$  be a permutation of the bins such that  $\sum_{j=1}^{b-1} |S_{\pi(j)}| < (1/2 - \alpha)n$  and  $\sum_{j=1}^b |S_{\pi(j)}| > (1/2 + \alpha)n$  for some  $b \in \{1, \dots, m\}$ . Thus  $|S_{\pi(b)}| \geq 2\alpha n$ . Take  $L = \{\pi(1), \dots, \pi(b-1)\}$  and  $R = \{\pi(b+1), \dots, \pi(m)\}$ . Recall the notation  $S^L = \bigcup_{j=1}^{b-1} S_{\pi(j)}$ . Since for every  $Y \in \binom{S_{\pi(b)}}{n/2 - |S^L|}$  the set  $Y \cup S^L$  is an  $(L, R)$ -witness of size  $\frac{n}{2}$ , there are at least  $\binom{|S_{\pi(b)}|}{n/2 - |S^L|}$   $(L, R)$ -witnesses of cardinality  $\frac{n}{2}$ , which is at least  $\binom{2\alpha n}{\alpha n}$  since  $|S_{\pi(b)}| \geq 2\alpha n$  and  $n/2 - |S^L| \geq \alpha n$ . Hence, there are  $\binom{n}{n/2}$  subsets of  $\{1, \dots, n\}$  of cardinality  $n/2$  and at least  $\binom{2\alpha n}{\alpha n}$  of those are  $(L, R)$ -witnesses. Observation 2.1 then gives that  $\mathcal{W}$  contains an  $(L, R)$ -witness with probability  $\geq 1/2$  as  $\mathcal{W}$  is a family of  $2^{(1-2\alpha)n} = 2^n / 2^{2\alpha n}$  random subsets of  $\{1, \dots, n\}$  of size  $\frac{n}{2}$ . Notice that for any witness  $W$  it will hold that  $\sum_{i=1}^n w(i) - l_W^* - r_W^* \leq c_{\pi(b)}$  and so the algorithm will return *yes* if  $W \in \mathcal{W}$ .

Moreover, when the algorithm finds a  $W \in \mathcal{W}$  such that  $\sum_{i=1}^n w(i) - l_W^* - r_W^* \leq c_{\pi(b)}$ , it means there exist sets  $X^L \subseteq W$  and  $X^R \subseteq \{1, \dots, n\} \setminus W$  that can be distributed to the bins of  $L$  and  $R$  respectively, such that  $\{1, \dots, n\} \setminus (X^L \cup X^R)$  fits into bin  $\pi(b)$ . Therefore,  $W$  is an  $(L, R)$ -witness and we proved the existence of a solution to the Bin Packing instance.

### Running time Analysis

We are left to prove the running time of the algorithm. Recall that the algorithm will repeat the procedure above for all  $m \cdot 2^{m-1}$  combinations of  $L$  and  $R$ . The running time per one guess of  $L$  and  $R$  is dominated by  $\mathcal{O}((|\downarrow\mathcal{W}| + |\uparrow\mathcal{W}|)n)$ , hence we are left to prove that  $|\downarrow\mathcal{W}| + |\uparrow\mathcal{W}| = \mathcal{O}(2^{(1-f_B(\alpha))n})$ . For this we use Lemma B.2 with  $z = 2\alpha$  and  $c = 0$ . This implies that

$$|\downarrow\mathcal{W}| + |\uparrow\mathcal{W}| \leq \mathcal{O}\left(2^{(1-\rho(2\alpha,0))n}\right),$$

where

$$\rho(2\alpha, 0) = \frac{2}{\ln(2)} \cdot \left(\frac{2\alpha}{4 \log(12/2\alpha)}\right)^2 = \Omega_{\alpha \rightarrow 0} \left(\frac{\alpha^2}{\log^2(1/\alpha)}\right) = \Omega_{\alpha \rightarrow 0} \left(\frac{\alpha^2}{\log^2(\alpha)}\right)$$

□

### 3.4 Pruned Item Weights and Slack

The results from the previous subsection enable us to assume that both  $|w(2^{\{1, \dots, n\}})| \geq 2^{\delta n}$  for some small constant  $\delta > 0$  (that we will fix later) and that there is an  $\alpha$ -balanced solution for some  $\alpha > 0$ .

To solve these instances of Bin Packing, we first need to define different parameters of an instance that determine our proof strategy.

**Definition 3.7** (*s*-pruned item weights). *Let  $l = 1 + \lceil \log(\max_i \{w(i)\}) \rceil$ . For  $s \in \{0, \dots, l\}$ , define the  $s$ -pruned weight of an item  $i$  as*

$$w_s(i) := \lfloor w(i)/2^{l-s} \rfloor.$$

The  $s$ -pruned weight of item  $i$  comes down to pruning the  $l$ -bit representation of  $w(i)$  to the  $s$  most significant bits. Indeed,  $w_s(i) \leq 2^s$ , and  $w_0(i) = 0$  for all items  $i$  and  $w_l = w$ . We will need the fact that the sequence

$$1 = |w_0(2^{\{1, \dots, n\}})|, \quad |w_1(2^{\{1, \dots, n\}})|, \quad \dots, \quad |w_l(2^{\{1, \dots, n\}})| = |w(2^{\{1, \dots, n\}})|,$$

is almost non-decreasing and relatively smooth. Observe that the sequence may not be non-decreasing. For example when  $w = (3, 7, 10)$  the number of bits is  $l = 5$ , and

$$\begin{array}{llll} w_0 = (0, 0, 0), & |w_0(2^{\{1, \dots, n\}})| = 1 & w_1 = (0, 0, 0), & |w_1(2^{\{1, \dots, n\}})| = 1 \\ w_2 = (0, 0, 1), & |w_2(2^{\{1, \dots, n\}})| = 2 & w_3 = (0, 1, 2), & |w_3(2^{\{1, \dots, n\}})| = 1 \\ w_4 = (1, 3, 5), & |w_4(2^{\{1, \dots, n\}})| = 8 & w_5 = (3, 7, 10), & |w_5(2^{\{1, \dots, n\}})| = 7. \end{array}$$

Nevertheless, this is only an artifact of smaller-order rounding errors and that the sequence in fact is *smooth* in the following precise sense (see Lemma A.3): For all  $s \in \{1, \dots, l\}$ :

$$\left(\frac{1}{3n}\right) |w_s(2^{\{1, \dots, n\}})| \leq |w_{s-1}(2^{\{1, \dots, n\}})| \leq \left(\frac{3n}{2}\right) |w_s(2^{\{1, \dots, n\}})|.$$

A part of our strategy is to use techniques from Lemma 3.5 to deal with mostly empty bins. The analogous dynamic programming table needs to be indexed by  $w_s$  for some  $s \in \{1, \dots, l\}$ . To achieve this, we will need a notion of *precision*. The precision parameter we will use is the following:

**Definition 3.8** (Critical pruner). *Let  $\delta \in (0, 1)$  be a fixed parameter<sup>7</sup> such that  $|w(2^{\{1, \dots, n\}})| \geq 2^{\delta n}$ . We define the critical pruner **crit** as*

$$\mathbf{crit} := \mathbf{crit}(\delta) := \min \left\{ s \in \mathbb{N} : |w_s(2^{\{1, \dots, n\}})| \geq 2^{\delta n} \right\}.$$

Observe, that  $|w_{\mathbf{crit}}(2^{\{1, \dots, n\}})| = \Theta(n2^{\delta n})$  by Lemma A.3 and the fact that  $w_0(2^{\{1, \dots, n\}}) = \{0\}$ . Furthermore, by Corollary A.2 the critical pruner **crit** can be computed in  $\mathcal{O}^*(2^{\delta n})$  time.

**Definition 3.9** (Slack). *The slack of a bin  $j$  is  $c_j - \sum_{i \in S_j} w(i)$ . A bin has  $\delta$ -large slack if it has slack at least  $n \cdot 2^{l-\mathbf{crit}}$  and  $\delta$ -small slack otherwise. An item is a large slack item if it is in a bin of large slack and a small slack item otherwise.*

We often omit  $\delta$  in the above notation, because  $\delta$  will be fixed later in Subsection 3.7.

---

<sup>7</sup>Which we will be set later in Subsection 3.7

### 3.5 Solving Case C: Balanced Solution with Many Small Slack Items

In the next lemma we will solve Bin Packing instances with at least  $(1/2 - \alpha)n$  small slack items. We will use this algorithm in Case C.

**In Case C:** (✓)  $|w(2^{\{1, \dots, n\}})| > 2^{\delta n}$ .

(✓) The instance has only  $\alpha$ -balanced solutions.

(✓) There are at least  $(1/2 - \alpha)n$  small slack items with respect to **crit**.

**Lemma 3.10.** *Suppose  $|w(2^{\{1, \dots, n\}})| \geq 2^{\delta n}$  and  $0 < \alpha \leq 2^{-2/\delta^3}$ . If a Bin Packing instance has a solution that is  $\alpha$ -balanced and has at least  $(1/2 - \alpha)n$  items with  $\delta$ -small slack, then such a solution can be found in time  $\mathcal{O}^*(2^{(1-f_C(\delta))n})$  for  $f_C(\delta) = \Omega_{\delta \rightarrow 0}(2^{-3/\delta^3})$ .*

*Proof.* Use Corollary A.2 to compute the critical pruner **crit** in time  $\mathcal{O}^*(2^{\delta n})$ . Then iterate over all combinations of sets  $L, R \subseteq \{1, \dots, m\}$  that form a partition of  $\{1, \dots, m\}$ . For each such a partition, the algorithm searches for  $(L, R)$ -witnesses of size  $[\frac{n}{2} \pm \alpha n]$  as follows: First, enumerate  $\mathcal{W}$ , which is defined as

$$\mathcal{W} := \left\{ W \subseteq \{1, \dots, n\} : \left| |W| - \frac{n}{2} \right| \leq \alpha n, \left( \sum_{j \in L} c_j / 2^{l-\text{crit}} - w_{\text{crit}}(W) \right) \in [0, n \cdot (|L| + 1)] \right\}.$$

We can enumerate  $\mathcal{W}$  in time  $\mathcal{O}(2^{n/2} + |\mathcal{W}|)$  with a standard Meet-in-the-Middle approach (see e.g. [6, Section 3.2] or Lemma [45, Lemma 3.8]). Next, for every  $W \in \mathcal{W}$  we determine whether  $W$  is an  $(L, R)$ -witness. This is done by computing the boolean  $l_X$  for every  $X \in \downarrow \mathcal{W}$  and  $r_X$  for every  $X \in \uparrow \mathcal{W}$ , where

$$l_X := \begin{cases} \text{true} & \text{if } X \text{ can be divided over the bins in } L, \\ \text{false} & \text{otherwise,} \end{cases}$$

$$r_X := \begin{cases} \text{true} & \text{if } \{1, \dots, n\} \setminus X \text{ can be divided over the bins in } R, \\ \text{false} & \text{otherwise.} \end{cases}$$

Using Theorem 2.7 we can do this in time  $\mathcal{O}((|\downarrow \mathcal{W}| + |\uparrow \mathcal{W}|)n)$ . Next, the algorithm checks for all  $W \in \mathcal{W}$ , whether  $l_W = r_W = \text{true}$ , and if so the algorithm returns *yes*. If for no partition  $L, R$  of  $\{1, \dots, m\}$  the algorithm finds a witness, the algorithm returns *no*.

#### Correctness of Algorithm

Assume that there is an  $\alpha$ -balanced solution  $S_1, \dots, S_m$ . Let  $\pi : \{1, \dots, m\} \rightarrow \{1, \dots, m\}$  be a permutation of the bins such that all bins with small slack have smaller index than the large slack bins, i.e.  $\pi(j) \leq \pi(j')$  for all small slack bins  $j$  and large slack bins  $j'$ . Since we assumed the solution to be  $\alpha$ -balanced, there exists a bin  $b \in \{1, \dots, m\}$  such that  $\sum_{j=1}^b |S_{\pi(j)}| \in [\frac{n}{2} \pm \alpha n]$ . Take  $L = \{\pi(1), \dots, \pi(b)\}$  and  $R = \{\pi(b+1), \dots, \pi(m)\}$ . Notice that  $S^L = \bigcup_{j=1}^b S_{\pi(j)}$  is an  $(L, R)$ -witness. We will prove that in the iteration of the algorithm where the correct partition  $L, R$  is chosen, it holds that  $S^L \in \mathcal{W}$ . Since there are at least  $(1/2 - \alpha)n$  small slack items, all bins in  $L$  have small slack. Hence,

$$\sum_{j=1}^b (c_{\pi(j)} - n \cdot 2^{l-\text{crit}}) \leq w(S^L) \leq \sum_{j=1}^b c_{\pi(j)}.$$



We also have the bound on the pruned weights of the items

$$w_{\text{crit}}(S^L) \leq w(S^L)/2^{l-\text{crit}} \leq \sum_{j=1}^b c_{\pi(j)}/2^{l-\text{crit}}.$$

On the other hand:

$$\begin{aligned} w_{\text{crit}}(S^L) &\geq w(S^L)/2^{l-\text{crit}} - n \\ &\geq \sum_{j=1}^b (c_{\pi(j)}/2^{l-\text{crit}} - n \cdot 2^{l-\text{crit}}) - n \\ &\geq \sum_{j=1}^b c_{\pi(j)}/2^{l-\text{crit}} - (b+1) \cdot n. \end{aligned}$$

Combining this with the fact that  $|S^L| \in [\frac{n}{2} \pm \alpha n]$ , we conclude that the set  $S^L$  is present in

$$\left\{ W \subseteq \{1, \dots, n\} : |W| \in [(\frac{1}{2} \pm \alpha)n], w_{\text{crit}}(W) \in \left[ \sum_{j=1}^b c_j/2^{l-\text{crit}} - (b+1) \cdot n, \sum_{j=1}^b c_j/2^{l-\text{crit}} \right] \right\}$$

which matches the definition of  $\mathcal{W}$ .

Notice that  $l_W = r_W = \text{true}$  if and only if  $W$  is an  $(L, R)$ -witness, since we chose  $L$  and  $R$  to partition  $\{1, \dots, m\}$ . Because  $S^L \in \mathcal{W}$ , the algorithm always returns *yes* in a *yes*-instance. Furthermore, when we find a  $W$  s.t.,  $l_W = r_W = \text{true}$ , we can conclude that there is a solution to the Bin Packing instance since all items are divided over all bins.

## Running time Analysis

It remains to analyze the running time of the algorithm. Recall that the algorithm iterates over all  $m \cdot 2^{m-1}$  combinations of  $L$  and  $R$ . Each iteration takes  $\mathcal{O}((|\downarrow\mathcal{W}| + |\uparrow\mathcal{W}|)n + 2^{n/2})$  time.

Before we can prove that  $|\downarrow\mathcal{W}| + |\uparrow\mathcal{W}| \leq 2^{(1-f_C(\delta))n}$ , we need to bound the size of  $\mathcal{W}$ . Recall that  $\text{crit}$  is the critical pruner, and therefore by definition  $|w_{\text{crit}}(2^{\{1, \dots, n\}})| \geq 2^{\delta n}$ . Theorem 1.2 states that if  $\beta(w_{\text{crit}}) \geq 2^{(1-\varepsilon')n}$ , then  $|w_{\text{crit}}(2^{\{1, \dots, n\}})| \leq 2^{\delta' n}$  where  $\delta' = \mathcal{O}_{\varepsilon' \rightarrow 0} \left( \frac{\log \log(1/\varepsilon')}{\sqrt{\log(1/\varepsilon')}} \right)$ .

Because we assume that  $\varepsilon' \rightarrow 0$  we use a crude bound  $\frac{\log \log(1/\varepsilon')}{\sqrt{\log(1/\varepsilon')}} \leq \frac{1}{\sqrt[3]{\log(1/\varepsilon')}}$  to guarantee that

$$\delta' \leq \mathcal{O}_{\varepsilon' \rightarrow 0} \left( \frac{1}{\sqrt[3]{\log(1/\varepsilon')}} \right).$$

Next, we manipulate this inequality to get:

$$\varepsilon' \leq \mathcal{O}_{\delta' \rightarrow 0} \left( 2^{-\left(\frac{1}{\delta'}\right)^3} \right). \quad (2)$$

Now, we denote  $\varepsilon(\delta) := 2^{-\left(\frac{1}{\delta}\right)^3}$ . By (2) there exists a constant  $\delta_0 > 0$  such that for all  $\delta < \delta_0$  it holds that if  $|w_{\text{crit}}(2^{\{1, \dots, n\}})| \geq 2^{\delta n}$ , then  $\beta(w_{\text{crit}}) \leq 2^{(1-\varepsilon(\delta))n}$ . Note that because we only claim an asymptotic time bound in the lemma, we may assume that  $\delta \leq \delta_0$ . As a consequence, for fixed weight value  $v$ , there are at most  $2^{(1-\varepsilon(\delta))n}$  sets  $W \subseteq \{1, \dots, n\}$  that have a weight  $w_{\text{crit}}(W) = v$ .

Because for all  $W \in \mathcal{W}$  it holds that  $w_{\text{crit}}(W) \in \left[ \sum_{j=1}^b c_j / 2^{l-\text{crit}} - (b+1) \cdot n, \sum_{j=1}^b c_j / 2^{l-\text{crit}} \right]$  there are at most  $mn$  such weights  $v$  and so  $|\mathcal{W}| \leq mn \cdot 2^{(1-\varepsilon(\delta))n}$ .

Therefore, we have a set  $\mathcal{W}$  of size  $\leq mn \cdot 2^{(1-\varepsilon(\delta))n}$  and each set  $W \in \mathcal{W}$  has a size  $|W| \in [\frac{n}{2} \pm \alpha n]$ . Lemma B.2 then bounds the sizes of  $\downarrow\mathcal{W}$  and  $\uparrow\mathcal{W}$ :

$$|\downarrow\mathcal{W}| + |\uparrow\mathcal{W}| \leq \mathcal{O}^* \left( 2^{(1-\rho(\varepsilon(\delta), \alpha))n} \right),$$

where  $\rho(\varepsilon(\delta), \alpha) = \frac{2}{\ln(2)} \left( \frac{\varepsilon(\delta)}{4 \log(12/\varepsilon(\delta))} - \alpha \right)^2$ . Recall that we ensured that  $0 < \alpha \leq 2^{-2/\delta^3}$ . Hence, there is a small enough constant  $\delta_0$  such that for any  $\delta \in (0, \delta_0)$  we have

$$\alpha \leq 2^{-2\delta^3} \leq \frac{2^{-1/\delta^3}}{8 \log(12/2^{-\delta^3})} = \frac{\varepsilon(\delta)}{8 \log(12/\varepsilon(\delta))}.$$

Hence,  $|\downarrow\mathcal{W}| + |\uparrow\mathcal{W}| = \mathcal{O}(2^{(1-f_C(\delta))n})$  where

$$\begin{aligned} f_C(\delta) &= \frac{2}{\ln(2)} \left( \frac{\varepsilon(\delta)}{4 \log(12/\varepsilon(\delta))} - \frac{\varepsilon(\delta)}{8 \log(12/\varepsilon(\delta))} \right)^2 \\ &= \frac{\varepsilon(\delta)^2}{32 \ln(2) (\log^2(12/\varepsilon(\delta)))} \\ &= \Omega_{\delta \rightarrow 0} \left( \frac{\varepsilon(\delta)^2}{\log^2(1/\varepsilon(\delta))} \right) \\ &= \Omega_{\delta \rightarrow 0} \left( \frac{2^{-2/\delta^3}}{\log^2(2^{(1/\delta)^3})} \right) \\ &= \Omega_{\delta \rightarrow 0} \left( \delta^6 \cdot 2^{-2/\delta^3} \right). \end{aligned}$$

Since  $f_C(\delta) \ll 1/2$  for  $\delta \in [0, 1]$ , the  $\mathcal{O}(2^{n/2})$  time bound is subsumed by the  $\mathcal{O}(2^{(1-f_C(\delta))n})$  term. Multiplying this term by  $m2^{m-1}$  different choices for  $L$  and  $R$ , gives us the requested running time.  $\square$

### 3.6 Solving Case D: Detecting a Balanced Solution with Few Small Slack Items

We are left to prove the remaining case, namely Case D, for which we assume that the following list of conditions holds.

- In Case D:** (✓)  $|w(2^{\{1, \dots, n\}})| > 2^{\delta n}$ .
- (✓) Instance has only  $\alpha$ -balanced solutions.
- (✓) There are at most  $(1/2 - \alpha)n$  small slack items with respect to **crit**.

First, we observe the following property of an  $\alpha$ -balanced solution.

**Observation 3.11.** *Let  $S_1, \dots, S_m$  be an  $\alpha$ -balanced solution for some  $0 < \alpha < \frac{1}{4m}$ . Assume  $k, k' \in \{1, \dots, m\}$  to be two different bins with the most items. Then either:*

1.  $|S_k|, |S_{k'}| \in [(1/2 \pm \alpha)n]$ , or

2.  $|S_j| \leq (\frac{1}{2} - \frac{1}{4m})n$  for all bins  $j \in \{1, \dots, m\}$ .

*Proof.* Let  $S_k$  be the bucket with the highest number of items and  $S'_k$  be the second highest number of items. If condition (2) does not hold, then we know that  $|S_k| > (\frac{1}{2} - \frac{1}{4m})n$ . The number of the items in the remaining buckets is  $n - |S_k|$  and the number of remaining buckets is  $m - 1$ . Therefore the second bucket with the highest number of items has a size  $|S_{k'}| \geq \frac{n - |S_k|}{m - 1}$ .

Hence,

$$|S_k| + |S_{k'}| \geq \left( \frac{1}{2} - \frac{1}{4m} + \frac{1 - \frac{1}{2} + \frac{1}{4m}}{m - 1} \right) \cdot n > \left( \frac{1}{2} + \frac{1}{4m} \right) \cdot n.$$

Since the solution is  $\alpha$ -balanced (with  $\alpha \leq \frac{1}{4m}$ ), it means that for all permutations  $\pi : \{1, \dots, m\} \rightarrow \{1, \dots, m\}$ , in particular those with  $\pi^{-1}(k) = 1$  and  $\pi^{-1}(k') = 2$ , there exists  $b \in \{1, \dots, m\}$  such that  $\sum_{j=1}^b |S_{\pi(j)}| \in [\frac{n}{2} \pm \alpha n]$ . Because  $|S_k| + |S_{k'}| > (1/2 + \alpha)n$ , we know that  $b = 1$  and  $|S_k| \in [\frac{n}{2} \pm \alpha n]$ . We can conclude the same for  $k'$  by repeating these last arguments for all permutations  $\pi$  with  $\pi^{-1}(k) = 2$  and  $\pi^{-1}(k') = 1$ , and thus condition (1) must hold, and the observation follows.  $\square$

**Lemma 3.12.** *Assume  $\alpha < \frac{1}{4m}$ . If a solution of a Bin Packing instance with  $m$  bins is  $\alpha$ -balanced and has at most  $(1/2 - \alpha)n$  items that have  $\delta$ -small slack, then with probability at least  $\frac{1}{2}$  a solution can be found in time  $\mathcal{O}^*(2^{(1-f_D(m)+\delta m)n})$  with*

$$f_D(m) = \Omega_{m \rightarrow \infty} \left( \frac{h(\frac{1}{2m})^2}{\log^2(h(\frac{1}{2m}))} \right).$$

*Proof.* For an overview of the algorithm, see Algorithm 1. Compute the critical pruner **crit** and the set  $w_{\text{crit}}(2^{\{1, \dots, n\}})$  in  $\mathcal{O}^*(2^{\delta n})$  time with Corollary A.2. The algorithm will search for  $(L, R)$ -witnesses for all  $L, R \subseteq \{1, \dots, m\}$  such that  $|R| = 1$  and  $L \cap R = \emptyset$ . For notation purposes, assume without loss of generality that  $R = \{1\}$ ,  $L = \{2, \dots, k\}$  and let  $M = \{k + 1, \dots, m\}$ .

Let  $\mathcal{W}$  be obtained by sampling  $2^{(1-g(m))n}$  random subsets (with replacement, and removing repeating items at the end) of  $\{1, \dots, n\}$  of size  $\frac{n}{2}$ , where  $g(m) := \frac{1}{2}h(1/(2m))$ . For a given  $L$  and  $R$ , we guess  $a_{k+1}, \dots, a_m \in w_{\text{crit}}(2^{\{1, \dots, n\}})$ . Then, we compute the boolean  $l_X$  for every  $X \in \downarrow \mathcal{W}$  and  $r_X$  for every  $X \in \uparrow \mathcal{W}$ , where

$$l_X := \begin{cases} \text{true} & \text{if there exists a partition } X_2, \dots, X_k, Y_{k+1}, \dots, Y_m \text{ of } X \text{ such that} \\ & \text{for all } j \in L : w(X_j) \leq c_j \text{ and for all } i \in M : w_{\text{crit}}(Y_i) \leq a_i, \\ \text{false} & \text{otherwise,} \end{cases}$$

$$r_X := \begin{cases} \text{true} & \text{if there exists a partition } X_1, Y_{k+1}, \dots, Y_m \text{ of } \{1, \dots, n\} \setminus X \text{ such that} \\ & w(X_1) \leq c_1 \text{ and for all } i \in M : w_{\text{crit}}(Y_i) \leq c_i / 2^{l-\text{crit}} - n - a_i, \\ \text{false} & \text{otherwise.} \end{cases}$$

To compute  $l_X$  and  $r_X$  we can use the fast zeta transformations. For  $j \in L \cup R$  define the functions  $\text{fit}_j : 2^{\{1, \dots, n\}} \rightarrow \{0, 1\}$  as

$$\text{fit}_j(X) = \begin{cases} 1, & \text{if } w(X) \leq c_j, \\ 0, & \text{otherwise.} \end{cases}$$

For  $i \in M$  define the functions  $\text{lfit}_i, \text{rfit}_i : 2^{\{1, \dots, n\}} \rightarrow \{0, 1\}$  as

$$\text{lfit}_i(X) = \begin{cases} 1, & \text{if } w_{\text{crit}}(X) \leq a_i, \\ 0, & \text{otherwise,} \end{cases}$$

$$\text{rfit}_i(X) = \begin{cases} 1, & \text{if } w_{\text{crit}}(X) \leq c_i/2^{l-\text{crit}} - n - a_i, \\ 0, & \text{otherwise.} \end{cases}$$

Now, we observe the following:

**Claim 3.13.**  $l_X = \text{true}$  if and only if  $(\text{fit}_2 \otimes \dots \otimes \text{fit}_k \otimes \text{lfit}_{k+1} \otimes \dots \otimes \text{lfit}_m)(X) > 0$ .

*Proof.* Let us assume that  $l_X = \text{true}$ . Let  $S'_2, \dots, S'_m$  be such that  $\text{fit}_j(S'_j) = 1$  for every  $j$ . Then  $S'_2, \dots, S'_m$  gives a non-zero contribution to  $(\text{fit}_2 \otimes \dots \otimes \text{fit}_k \otimes \text{lfit}_{k+1} \otimes \dots \otimes \text{lfit}_m)(X)$  and hence it must be positive.

For the other direction, if  $(\text{fit}_2 \otimes \dots \otimes \text{fit}_k \otimes \text{lfit}_{k+1} \otimes \dots \otimes \text{lfit}_m)(X) > 0$  there exist  $S'_2, \dots, S'_m$  such that  $\text{fit}_j(S'_j) = 1$  for every  $j$  and  $S'_2 \cup \dots \cup S'_m = X$ . Observe that we can transform this into a partition  $S''_2, \dots, S''_m$  of  $X$  by choosing  $S''_j \subseteq S'_j$ . Because  $\text{fit}_j$  does not decrease when taking subsets we know that  $1 = \text{fit}_j(S'_j) \leq \text{fit}_j(S''_j)$ , and thus  $l_X = \text{true}$ .  $\square$

Similarly, we can argue that  $r_X = \text{true}$  if and only if  $(\text{fit}_1 \otimes \text{rfit}_{k+1} \otimes \dots \otimes \text{rfit}_m)(\{1, \dots, n\} \setminus X) > 0$ .

We can compute booleans  $l_X$  and  $r_X$  in time  $\mathcal{O}((|\downarrow\mathcal{W}| + |\uparrow\mathcal{W}|)n)$  by combining Theorem 2.4 and Theorem 2.6. Finally, if we find  $W \in \mathcal{W}$ , such that  $l_W = r_W = \text{true}$ , we can return *yes*.

<b>Algorithm:</b> BinPacking( $w_1, \dots, w_n$ )	
<b>Output</b> : Yes (whp.), if an $\alpha$ -balanced solution with $(1/2 - \alpha)n$ small slack items exist	
1	Compute the critical pruner <code>crit</code> with Corollary A.2. <span style="float: right;">// In time <math>\mathcal{O}^*(2^{\delta n})</math></span>
2	Compute the set $w_{\text{crit}}(2^{\{1, \dots, n\}})$ with Lemma A.1. <span style="float: right;">// In time <math>\mathcal{O}^*(2^{\delta n})</math></span>
3	Let $\mathcal{W}$ be a set of $2^{(1-g(m))n}$ random subsets of $n$ of size $\frac{n}{2}$ .
4	<b>for</b> $L, R \subseteq \{1, \dots, m\}$ such that $ R  = 1$ and $L \cap R = \emptyset$ <b>do</b> <span style="float: right;">// <math>m2^{m-1}</math> repetitions</span>
5	Assume without loss of generality that $R = \{1\}$ , $L = \{2, \dots, k\}$ .
6	<b>for</b> $a_{k+1}, \dots, a_m \in w_{\text{crit}}(2^{\{1, \dots, n\}})$ <b>do</b> <span style="float: right;">// <math> w_{\text{crit}}(2^{\{1, \dots, n\}}) ^{m-k}</math> repetitions</span>
7	Compute $l_X$ for all $X \in \downarrow\mathcal{W}$ . <span style="float: right;">// In time <math>\mathcal{O}(( \downarrow\mathcal{W}  +  \uparrow\mathcal{W} )n)</math></span>
8	Compute $r_X$ for all $X \in \uparrow\mathcal{W}$ . <span style="float: right;">// In time <math>\mathcal{O}(( \downarrow\mathcal{W}  +  \uparrow\mathcal{W} )n)</math></span>
9	<b>if</b> $l_W = r_W = \text{true}$ , for some $W \in \mathcal{W}$ <b>then</b>
10	<b>return</b> <i>yes</i>
11	<b>return</b> <i>no</i>

**Algorithm 1:** Overview of the algorithm for Lemma 3.12

### Constant probability of a witness in $\mathcal{W}$

Recall that the set  $\mathcal{W}$  is a random subset of  $\binom{\{1, \dots, n\}}{n/2}$  of size  $2^{(1-g(m))n}$  with  $g(m) = h(1/(2m))/2$ . We first analyze the number of  $(L, R)$ -witnesses that are in  $\binom{\{1, \dots, n\}}{n/2}$ . Assume that there is an  $\alpha$ -balanced solution  $S_1, \dots, S_m$  for some  $0 < \alpha < \frac{1}{4m}$ . We use Observation 3.11 to conclude that either  $|S_j| < (\frac{1}{2} - \frac{1}{4m})$  for all bins, or that  $|S_k|, |S_{k'}| \in [\frac{n}{2} \pm \alpha n]$  for largest bins  $k$  and  $k'$ . Since we assumed that there are at most  $(1/2 - \alpha)n$  small slack items, we know that in the latter case bins

$k$  and  $k'$  are therefore large slack bins. In either case, we conclude that  $|S_j| \leq (\frac{1}{2} - \frac{1}{4m})n$  for all small slack bins. We will assume without loss of generality that bin 1 is the largest small slack bin and that bins  $2, \dots, k$  are the other small slack bins. Then let  $L = \{2, \dots, k\}$ ,  $R = \{1\}$  and thus  $M = \{k+1, \dots, m\}$  are all large slack bins. We will lower bound the number of  $(L, R)$ -witnesses of size  $\frac{n}{2}$ .

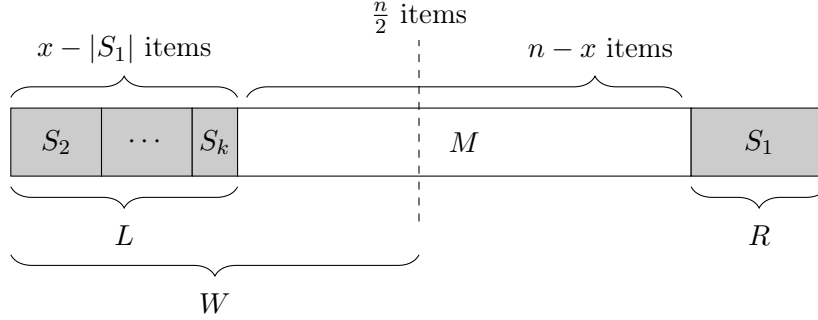


Figure 3: Overview of  $(L, R)$ -witnesses of size  $\frac{n}{2}$  for explanation of equation 3. Let  $x$  be the number of small slack items. Then any such  $(L, R)$ -witness,  $W$ , must include all items of  $S_2, \dots, S_k$  and exclude any items of  $S_1$ . The other items in  $W$  can then be any combination of large slack items, which are exactly the items in the bins of  $M$ .

Let  $x$  be the number of small slack items in the solution. Note that the number of  $(L, R)$ -witnesses of size  $\frac{n}{2}$  is equal to

$$\binom{n-x}{n/2 - (x - |S_1|)}, \quad (3)$$

since the sets  $S_2, \dots, S_k$  together with any subset of  $n/2 - (x - |S_1|)$  large slack items form a witness. See Figure 3 for an illustration of this. Now, we analyse two cases. If  $x \leq n/4$ , then because  $\binom{n-x}{n/2 - (x - |S_1|)} \geq \binom{n-x}{n/2 - x}$  there are at least  $\binom{3n/4}{n/4}$   $(L, R)$ -witnesses of size  $\frac{n}{2}$ .

In the case when  $x \geq \frac{n}{4}$ , then notice that  $\frac{x}{m} \leq |S_1| \leq (\frac{1}{2} - \frac{1}{m})n$  (because  $S_1$  is the largest among the small slack bins). Therefore, the number of witnesses is at least the number of ways to choose  $n/2 - |S_1|$  items from  $M$  to exclude in the witness. Thus we have that the number of witnesses of size  $\frac{n}{2}$  is at least

$$\binom{n-x}{n/2 - |S_1|} \geq \min \left\{ \binom{n/2}{n/2 - x/m}, \binom{n/2}{n/2 - (n/2 - n/m)} \right\} \geq \binom{n/2}{n/4m}.$$

So in both cases for  $x$ , we can conclude that the number of  $(L, R)$ -witnesses of size  $\frac{n}{2}$  is at least  $2^{g(m)n}$ .

Because  $\mathcal{W}$  is a random subset of  $\binom{\{1, \dots, n\}}{n/2}$  of size  $2^{(1-g(m))n}$  and at least  $2^{g(m)n}$  of those are  $(L, R)$ -witnesses, Observation 2.1 tells us that with probability at least  $\frac{1}{2}$ ,  $\mathcal{W}$  contains an  $(L, R)$ -witness.

### Correctness of Algorithm

The algorithm returns *yes* if and only if  $l_W = r_W = \mathbf{true}$  for some  $W \in \mathcal{W}$ . So if it returns *yes*, there exists a partition  $X_2, \dots, X_k, Y_{k+1}, \dots, Y_m$  of  $W$  and a partition  $X_1, Y'_{k+1}, \dots, Y'_m$  of  $\{1, \dots, n\} \setminus W$  by definition. Together they partition all items. Notice that by definition we know that  $X_j$  can be

put into bin  $j$  for all  $j \in L \cup R$ . Hence we are left to prove that for all  $j \in M$ :  $X_j = Y_j \cup Y'_j$  can be put into bin  $j$ . Notice that since  $\text{lfit}_j(Y_j) = \text{rfit}_j(Y'_j) = 1$  we have that

$$\begin{aligned} \sum_{i \in X_j} w_{\text{crit}}(i) &\leq a_j + c_j/2^{l-\text{crit}} - n - a_j && \implies \\ \sum_{i \in X_j} \lfloor w(i)/2^{l-\text{crit}} \rfloor &\leq c_j/2^{l-\text{crit}} - n && \implies \\ \sum_{i \in X_j} (w(i) - 2^{l-\text{crit}}) &\leq c_j - n2^{l-\text{crit}} && \implies \\ \sum_{i \in X_j} w(i) &\leq c_j, \end{aligned}$$

and so, indeed the items of  $X_j$  fit into bin  $j$  and we have a yes-instance. For the implication in the other direction, we prove that if there exists a solution, the algorithm finds it with constant probability. We already showed that with constant probability there is an  $(L, R)$ -witness  $W \in \mathcal{W}$ . Next, we will prove that there exist  $a_{k+1}, \dots, a_m \in w_{\text{crit}}(2^{\{1, \dots, n\}})$  such that  $l_W = r_W = 1$  for all witnesses  $W$ . Let  $S_2, \dots, S_k, S_{k+1} \cap W, \dots, S_m \cap W$  be the partition of  $W$  from the definition of  $l_W$ , and let  $S_1, \dots, S_{k+1} \setminus W, \dots, S_m \setminus W$  be the partition of  $\{1, \dots, n\} \setminus W$  from the definition of  $r_W$ .

Note that, for all  $j \in L \cup R$  it holds that  $w(S_j) \leq c_j$  because  $S_1, \dots, S_m$  is a solution. So we are left to prove that for all  $j \in M$  there exists an  $a_j \in w_{\text{crit}}(2^{\{1, \dots, n\}})$  such that

$$w_{\text{crit}}(S_j \cap W) \leq a_j \quad \text{and} \quad w_{\text{crit}}(S_j \setminus W) \leq c_j/2^{l-\text{crit}} - n - a_j.$$

Recall that we assumed that the bins of  $M$  are large slack bins. Hence we know that for  $j \in M$ :

$$\begin{aligned} \sum_{i \in S_j} w(i) &\leq c_j - n2^{l-\text{crit}} && \implies \\ \sum_{i \in S_j} \lfloor w(i)/2^{l-\text{crit}} \rfloor &\leq c_j/2^{l-\text{crit}} - n && \implies \\ \sum_{i \in S_j} w_{\text{crit}}(i) &\leq c_j/2^{l-\text{crit}} - n - a_j + a_j. \end{aligned}$$

So, take  $a_j = w_{\text{crit}}(S_j \cap W) \in w_{\text{crit}}(2^{\{1, \dots, n\}})$  and indeed the correctness of the algorithm follows.

## Running Time Analysis

The algorithm will go through the procedure of computing the booleans  $l_X$  and  $r_X$  for all different sets  $L, R \subseteq \{1, \dots, m\}$  such that  $|R| = 1$  and for all different values of  $a_{k+1}, \dots, a_m \in w_{\text{crit}}(2^{\{1, \dots, n\}})$ . This gives a total of at most  $m \cdot 2^{m-1} \cdot |w_{\text{crit}}(2^{\{1, \dots, n\}})|^m$  repetitions. By Lemma A.3, we have  $|w_{\text{crit}}(2^{\{1, \dots, n\}})| \leq 3n|w_{\text{crit}-1}(2^{\{1, \dots, n\}})|$ . Because  $\text{crit}$  is the critical pruner, and since  $w_0(2^{\{1, \dots, n\}}) = \{0\}$ , we know that  $|w_{\text{crit}-1}(2^{\{1, \dots, n\}})| \leq 2^{\delta n}$ . Hence, the number of repetitions is at most  $\mathcal{O}((2n)^m \cdot 2^{\delta mn})$ .

Now we analyze the time complexity per the choice of  $(L, R)$  and  $a_{k+1}, \dots, a_m$ . Recall that we chose  $\mathcal{W} \subseteq \binom{\{1, \dots, n\}}{n/2}$  as a random set of size  $2^{(1-g(m))n}$ . Computing all the booleans  $l_X$  and  $r_X$  can be done in  $\mathcal{O}((|\downarrow \mathcal{W}| + |\uparrow \mathcal{W}|)n)$  time. We can use Lemma B.2 to find that

$$|\downarrow \mathcal{W}| + |\uparrow \mathcal{W}| = \mathcal{O}(2^{(1-\rho(g(m), 0))n}),$$

where

$$\begin{aligned}\rho(g(m), 0) &= \frac{2}{\ln(2)} \cdot \frac{g(m)^2}{16(\log(12/g(m)))^2} \\ &= \frac{h(\frac{1}{2m})^2}{32 \ln(2) \log(24/h(\frac{1}{2m}))^2} \\ &= \Omega_{m \rightarrow \infty} \left( \frac{h(\frac{1}{2m})^2}{\log^2(h(\frac{1}{2m}))} \right).\end{aligned}$$

Combining this with the number of repetitions we get a running time of  $\mathcal{O}^*(2^{(1-f_D(m)+\delta m)n})$ . This gives us the claimed running time.  $\square$

### 3.7 Proof of Theorem 1.1

We are now ready to prove Theorem 1.1 by combining all work of the previous sections and setting the parameters  $\alpha$  and  $\delta$ :

*Proof.* We will now combine all previous lemmas. An overview of the algorithm can be found in Figure 2. To facilitate the asymptotic analysis, note we can assume the number of bins  $m$  is at least  $m_0$  for some constant  $m_0$ . If this is not the case we can add  $m_0 - m$  artificial bins with unique small capacities and matching items. Since  $m_0$  is constant this does not influence the asymptotic running time of the algorithm. Define  $f_D(m)$  as in Lemma 3.12 as:

$$f_D(m) = \frac{h(\frac{1}{2m})^2}{32 \ln(2) \log(24/h(\frac{1}{2m}))^2} = \Omega_{m \rightarrow \infty} \left( \frac{h(\frac{1}{2m})^2}{\log^2(h(\frac{1}{2m}))} \right).$$

Then, set  $\delta := f_D(m)/(2m)$  and  $\alpha := 2^{-2/\delta^3}$ . Then  $f_D(m) > 0$ ,  $\delta > 0$  and  $\alpha > 0$ .

1. If  $|w(2^{\{1, \dots, n\}})| \leq 2^{\delta n}$  (Case A), the algorithm from Lemma 3.5 solves the instance in time

$$\mathcal{O}^*(|w(2^{\{1, \dots, n\}})|^m) = \mathcal{O}^*(2^{(\delta n)m}) = \mathcal{O}^*(2^{\frac{f_D(m)}{2}n}).$$

2. If the instance has an  $\alpha$ -unbalanced solution (Case B), the algorithm from Lemma 3.6 can detect with probability at least  $\frac{1}{2}$  in time

$$\mathcal{O}^*\left(2^{\left(1 - \Omega\left(\frac{\alpha^2}{\log^2(\alpha)}\right)\right)n}\right) = \mathcal{O}^*\left(2^{\left(1 - \Omega\left(\delta^6 2^{-4/\delta^3}\right)\right)n}\right).$$

3. If the instance has an  $\alpha$ -balanced solution,  $|w(2^{\{1, \dots, n\}})| \leq 2^{\delta n}$ , and a solution with at least  $(1/2 - \alpha)n$  small slack items (Case C), the upper bound  $\alpha \leq 2^{-2/\delta^3}$  ensures that the solution can be detected by the algorithm from Lemma 3.10 in time

$$\mathcal{O}^*\left(2^{\left(1 - \Omega(2^{-3/\delta^3})\right)n}\right).$$

4. Otherwise, if the instance has an  $\alpha$ -balanced solution,  $|w(2^{\{1, \dots, n\}})| \leq 2^{\delta n}$ , and a solution with at most  $(1/2 - \alpha)n$  small slack items (Case D), the algorithm from Lemma 3.12 detect the solution with probability at least  $\frac{1}{2}$  in time

$$\mathcal{O}^*\left(2^{(1-f_D(m)+\delta m)n}\right) = \mathcal{O}^*\left(2^{\left(1 - \frac{f_D(m)}{2}\right)n}\right).$$

Thus, we obtain a probabilistic algorithm for Bin Packing that runs in time  $\mathcal{O}^*(2^{(1-\sigma_m)n})$ , where  $\sigma_m$  is a strictly positive number. Notice that any polynomial factor hidden in the  $\mathcal{O}^*$  notation can be subsumed by  $\mathcal{O}(2^{(1-\sigma_m)n})$ . □

## 4 The Littlewood–Offord Theorem

In this section, we will prove our Additive Combinatorics result which we first restate for convenience:

**Theorem 4.1.** *Let  $\varepsilon > 0$ . If  $\beta(w) \geq 2^{(1-\varepsilon)n}$ , then  $|w(2^{\{1,\dots,n\}})| \leq 2^{\delta n}$ , where*

$$\delta(\varepsilon) = \mathcal{O}_{\varepsilon \rightarrow 0} \left( \frac{\log(\log(1/\varepsilon))}{\sqrt{\log(1/\varepsilon)}} \right).$$

This theorem was recently improved by [32], showing that  $\delta(\varepsilon) = \mathcal{O}(\sqrt{\varepsilon})$ .

For our proof, it will be convenient to use a reformulation of Theorem 1.2 to a version with two set families that attain the parameters and use vector notation (so  $w$  is a vector and  $w(X)$  is the inner product  $\langle w, x \rangle$  of  $w$  with the characteristic vector  $x$  of set  $X$ ):

**Theorem 4.2** (Theorem 1.2 reformulated). *Let  $w = (w_1, \dots, w_n) \in \mathbb{Z}^n$  be a vector with integer weights, and let  $A, B \subseteq \{0, 1\}^n$  be such that  $|a^{-1}(1)| = \alpha n$  for each  $a \in A$  and*

- $\langle w, b \rangle = \tau$  for every  $b \in B$ , and
- if  $a, a' \in A$  and  $\langle w, a \rangle = \langle w, a' \rangle$ , then  $a = a'$ .

*If  $|B| \geq 2^{(1-\varepsilon)n}$ , then  $|A| \leq 2^{\delta(\varepsilon)n}$ , where*

$$\delta(\varepsilon) = \mathcal{O}_{\varepsilon \rightarrow 0} \left( \frac{\log \log(1/\varepsilon)}{\sqrt{\log(1/\varepsilon)}} \right).$$

Recall, that for  $a \in \{0, 1\}^n$  we use  $a^{-1}(1)$  to denote  $\{i \in \{1, \dots, n\} : a(i) = 1\}$  (see Section 2). We first show that this implies Theorem 4.2:

*Proof of Theorem 1.2 assuming Theorem 4.2.* Suppose  $w_1, \dots, w_n$  and  $\tau$  are such that  $|\{X \subseteq \{1, \dots, n\} : |X| \geq 2^{(1-\varepsilon)n}\}| \geq 2^{(1-\varepsilon)n}$ . Then  $B := \{b \in \{0, 1\}^n : \langle w, b \rangle = \tau\}$  satisfies the conditions of Theorem 4.2 and  $|B| \geq 2^{(1-\varepsilon)n}$ . For every  $i \in w(2^{\{1,\dots,n\}})$  arbitrarily choose a vector  $a(i) \in \{0, 1\}^n$  such that  $\langle w, a(i) \rangle = i$ . Define  $A' = \{a(i) : i \in w(2^{\{1,\dots,n\}})\}$ . Since  $|a^{-1}(1)|$  can only take  $n$  different values, there exists an  $\alpha$  such that  $|a^{-1}(1)| = \alpha n$  for at least  $1/n$  fraction of the elements of  $A'$ . This gives a set  $A$  that satisfies the condition of Theorem 4.2, and thus

$$|w(2^{\{1,\dots,n\}})| = |A'| \leq |A| \cdot n \leq 2^{\delta(\varepsilon)n + o(n)}.$$

Note that we use the  $\mathcal{O}(\cdot)$  notation in the term  $\delta(\varepsilon)$  to hide the  $2^{o(n)}$  factors. □

The rest of this section is dedicated to the proof of Theorem 4.2. We use the following standard definitions from Additive Combinatorics: For sets  $X, Y$  we define  $X + Y$  as the sumset  $\{x + y : x \in X, y \in Y\}$ . For an integer  $k$ , we define  $k \cdot X$  as the  $k$ -fold sum

$$k \cdot X := \underbrace{X + X + \dots + X}_{k \text{ times}}.$$

The starting point of the proof of Theorem 4.2 is the following simple lemma that proves that  $|A||k \cdot B| = |A + k \cdot B|$ . It is heavily inspired by the UDCP connection from [3, Proposition 4.2].



**Lemma 4.3.** *If  $a, a' \in A$  and  $b, b' \in k \cdot B$  are such that  $a + b = a' + b'$ , then  $(a, b) = (a', b')$ .*

*Proof.* Note that

$$\langle w, a \rangle + \langle w, b \rangle = \langle w, a + b \rangle = \langle w, a' + b' \rangle = \langle w, a' \rangle + \langle w, b' \rangle.$$

By definition of  $k \cdot B$ , we know that  $\langle w, b \rangle = \langle w, b' \rangle = k \cdot \tau$ , hence  $\langle w, a \rangle = \langle w, a' \rangle$ . Therefore by definition of set  $A$  it has to be that  $a' = a$ . This implies that  $b = b'$ , since  $a + b = a' + b'$ .  $\square$

Thus  $|A|$  is equal to  $|A + k \cdot B|/|k \cdot B|$ , and we may restrict our attention to upper bounding the latter quantity for any integer  $k \in \mathbb{N}$ . Since this is in general not easy, we instead define a set  $P \subseteq A \times k \cdot B$  of pairs such that for each  $(a, b) \in P$  the distribution of the values in the vector  $a + b$  is close to what one would expect for random vectors. This is useful since the control on pairs  $(a, b) \in P$  gives us control on the vectors  $a + b$  which allows us to upper bound  $P$ . Moreover, we also provide a lower bound that shows that  $P$  is not much smaller than  $|A| \cdot |k \cdot B|$ . Combining the two bounds results in the upper bound for  $A$ . We will make this more formal in the next subsections, but first, we give a warm-up result that sets up the notation for the main proof.

#### 4.1 A Warm-up with $B = \{0, 1\}^n$

Let us first investigate what happens in a case when  $B$  is equal to the whole Boolean hypercube  $\{0, 1\}^n$ . While  $|A|$  can be easily upper bounded by direct methods, it is instructive to see what our approach will be in this special case. In this setting, we can think about vectors from  $B$  as sampled uniformly at random. Fix a parameter  $0 < \alpha < 1$ , let  $a \in \{0, 1\}^n$  be a fixed, adversarially chosen vector with  $|a^{-1}(1)| = \alpha n$ , and let  $b_1, \dots, b_k \in \{0, 1\}^n$  be independently sampled random vectors. Let  $b = b_1 + b_2 + \dots + b_k$  and  $c = a + b$ . Observe that for every  $i \in \{0, \dots, k\}$  and  $i' \in \{0, \dots, k+1\}$

$$\mathbb{E} \left[ \frac{|b^{-1}(i)|}{n} \right] = \binom{k}{i} 2^{-k}, \quad \text{and} \quad \mathbb{E} \left[ \frac{|c^{-1}(i')|}{n} \right] = \left( (1 - \alpha) \binom{k}{i'} + \alpha \binom{k}{i' - 1} \right) 2^{-k},$$

For further reference, we now define the found distributions explicitly <sup>8</sup>:

**Definition 4.4** (Altered Binomial Distribution). *For every  $k \in \mathbb{N}$ , we let  $\text{Bin}(k)$  to denote the binomial distribution  $(\{0, \dots, k\}, p)$  where  $p(i) = \binom{k}{i} 2^{-k}$ .*

*For an additional parameter  $\alpha \in (0, 1)$ , we define the altered binomial distribution  $\text{Bin}(k, \alpha)$  as  $(\{0, \dots, k+1\}, p')$  where  $p'(i) = (1 - \alpha) \binom{k}{i} 2^{-k} + \alpha \binom{k}{i-1} 2^{-k}$ .*

Note, that  $\text{Bin}(k+1) = \text{Bin}(k, 1/2)$  by Pascal's Formula. Now, we present the intuition for the random case. We have that:

$$n \cdot h(\text{Bin}(k, \alpha)) = h(c) = h(a, b) = h(a) + h(b) = h(a) + n \cdot h(\text{Bin}(k)),$$

where the second equality follows by Lemma 4.3 and the third equality follows because  $a$  and  $b$  are independent. Thus  $h(a) = n(h(\text{Bin}(k, \alpha)) - h(\text{Bin}(k)))$ , and the proof in the random case can be concluded by using Lemma 4.9 in which we show that for any constant  $\alpha \in (0, 1)$  it holds that  $h(\text{Bin}(k, \alpha)) - h(\text{Bin}(k)) = \mathcal{O}_{k \rightarrow \infty}((\log k)/\sqrt{k})$ , and the standard fact that the support of any uniform random variable of entropy  $h$  is at most  $2^h$ .

This concludes the analysis for the special case of  $B \subset \{0, 1\}^n$ .

---

<sup>8</sup>For convenience we assume that  $\binom{n}{i} = 0$  when  $i < 0$

## 4.2 Balanced Pairs

Now, we consider the general setting where  $B \subset \{0, 1\}^n$ . We need to obtain sufficiently large sample of vectors  $b_1, \dots, b_k$ . More precisely, we will consider the setting with  $\varepsilon \leq 1/2^{\Theta(k)}$ , that will enforce  $k := \Theta(\log(1/\varepsilon))$ . The following chain of (in-)equalities summarizes the strategy of our proof.

$$|A| = \frac{|A + k \cdot B|}{|k \cdot B|} \leq \frac{|P|}{|k \cdot B|} \cdot 2^{f(\varepsilon, k)n} \leq 2^{n(h(\text{Bin}(k+1)) - h(\text{Bin}(k)))} \cdot 2^{f(\varepsilon, k)n} \leq 2^{\delta(\varepsilon)n}$$

Lemma 4.3
Section 4.2
Section 4.3
Lemma 4.9

Now, we will make the above idea more precise.

The following definition quantifies the ‘sufficiently random’ terms from the previous subsection by measuring how far the distribution of the values of a vector is from a given (expected) distribution.

**Definition 4.5** (Balanced vectors). *Let  $\mathcal{D} = (\Omega, p)$  be a discrete probability space. Fix  $\gamma \in (0, 1)$ . Let  $U$  be the finite universe set and let  $X \subseteq U$ . A mapping (or a vector)  $v \in \Omega^U$  is  $\gamma$ - $\mathcal{D}$  balanced for  $X$  if for all  $\omega \in \Omega$  it holds that*

$$\frac{|v^{-1}(\omega) \cap X|}{|X|} \in [p(\omega) \pm \gamma].$$

As a shorthand, we say that a mapping (or a vector)  $v \in \Omega^U$  is  $\gamma$ - $\mathcal{D}$  balanced if it is  $\gamma$ - $\mathcal{D}$  balanced for  $U$ . We denote the set of all  $\gamma$ - $\mathcal{D}$  balanced vectors  $v \in \Omega^U$  with  $(\mathcal{D} \pm \gamma)^U$ .

As an illustration of Definition 4.5, suppose

$$U = \{1, \dots, 6\}, \quad X = \{1, \dots, 4\}, \quad \Omega = \{0, 1\}, \quad p(0) = p(1) = \frac{1}{2}, \quad \mathcal{D} = (\Omega, p).$$

Then  $(0, 1, 1, 1, 1, 1)$  is  $\frac{1}{4}$ - $\mathcal{D}$  balanced for  $X$  but not  $\frac{1}{4}$ - $\mathcal{D}$  balanced. The vector  $(0, 0, 0, 0, 1, 1)$  is not  $\frac{1}{4}$ - $\mathcal{D}$  balanced for  $X$  but it is  $\frac{1}{6}$ - $\mathcal{D}$  balanced.

We will use Definition 4.5 with  $\mathcal{D}$  being the distribution we would get in the random case as outlined in Subsection 4.1 (hence  $\mathcal{D}$  will usually be  $\text{Bin}(k)$  or  $\text{Bin}(k, \alpha)$ ). Now, we prove a general upper bound on the number of  $\gamma$ - $\mathcal{D}$  balanced vectors.

**Lemma 4.6.** *Let  $\mathcal{D} = (\Omega, p)$  be a discrete probability space. The number of  $\gamma$ - $\mathcal{D}$  balanced vectors is at most*

$$2^{(h(\mathcal{D}) + f(\Omega, \gamma))|U|},$$

where  $f(\Omega, \gamma) := \mathcal{O}(|\Omega| \cdot \gamma \log(1/\gamma))$ .

*Proof.* Let  $d$  be the dimension of  $\mathcal{D}$ , i.e.,  $p = (p_1, \dots, p_d) : \Omega \rightarrow (0, 1)^d$ . The number of  $\gamma$ - $\mathcal{D}$  balanced vectors is at most

$$\sum_{q_1, \dots, q_d \in (0, 1)} \binom{|U|}{q_1|U|, \dots, q_d|U|},$$

where the sum is over all the probability distributions  $q_i$  such that  $q_i \cdot |U|$  is integer and  $q_i(\omega) \in [p_i(\omega) \pm \gamma]$  for every  $\omega \in \Omega$  and  $i \in \{1, \dots, d\}$ .

Since the number of possibilities for such a  $q_i$  is at most  $|U|^{|q_i|}$  and  $\binom{|U|}{q_1|U|, \dots, q_d|U|} \leq 2^{h(q_1, \dots, q_d)|U|}$  by Lemma 2.9, we obtain that the number of  $\gamma$ - $\mathcal{D}$  balanced vector is at most

$$|U|^{|q_i|} 2^{h(q_1, \dots, q_d)|U|} \leq |U|^{|q_i|} 2^{(h(\mathcal{D}) + \frac{1}{\ln(2)}|q_i| \gamma \log \frac{1}{\gamma})|U|} \leq 2^{h(\mathcal{D})|U|} \cdot 2^{\mathcal{O}(|U| \gamma \log(\frac{1}{\gamma}))},$$

where the second inequality follows from Lemma 2.10 and the third comes from  $|U|^{|q_i|} \ll 2^{|U| \gamma \log(\frac{1}{\gamma})}$ .  $\square$

For example, Lemma 4.6 bounds the number of  $\gamma$ -Bin( $k$ ) balanced vectors by  $2^{nh(\text{Bin}(k))+nf(\gamma,k)}$  for some positive function  $f(\gamma,k)$  that goes to 0 when  $\gamma \rightarrow 0$ .

With Definition 4.5 in hand, we are ready to define the set of pairs mentioned at the beginning of this section:

$$P := \{(a, b) \in A \times k \cdot B : b \in B \text{ is } \varepsilon^{0.01}\text{-Bin}(k) \text{ balanced for } a\}.$$

In Section 5 we prove the following somewhat technical lemma:

**Lemma 4.7.** *Let  $k < 0.01 \cdot \log(1/\varepsilon)$ . Then, for every  $a \in \{0, 1\}^n$  with  $|a^{-1}(1)| > \varepsilon^{0.01}n$ , there exists  $E_a \subseteq k \cdot B$ , such that  $|E_a| \geq 2^{(h(\text{Bin}(k))-\varepsilon^{0.1})n}$  and  $(a, b) \in P$  for every  $b \in E_a$ .*

Note, that we can assume that  $\alpha > \varepsilon^{0.01}$  because otherwise  $|A| \leq \binom{n}{\varepsilon^{0.01}n} \leq 2^{\varepsilon^{0.01} \log(4/\varepsilon)n} \leq 2^{\delta n}$  and Theorem 4.2 follows automatically.

Thus, we may apply Lemma 4.7 for each  $a \in A$  and obtain that

$$|P| \geq |A| \cdot 2^{(h(\text{Bin}(k))-\varepsilon^{0.1})n}. \quad (4)$$

On the other hand, the balancedness property can be used to give an upper bound on  $P$  via Lemma 4.3. To do so, the following will be useful:

**Lemma 4.8.** *If  $(a, b) \in P$ , then  $a + b$  is  $(2\varepsilon^{0.01})$ -Bin( $k, \alpha$ ) balanced.*

*Proof.* From the definition of  $P$ , vector  $b$  is  $\varepsilon^{0.01}$ -Bin( $k$ ) balanced for  $a$ . So, for every  $i \in \{0, \dots, k\}$ :

$$|a^{-1}(1) \cap b^{-1}(i)| \in \left[ \binom{k}{i} \frac{\alpha n}{2^k} \pm \varepsilon^{0.01}n \right]$$

And similarly,

$$|a^{-1}(0) \cap b^{-1}(i)| \in \left[ \binom{k}{i} \frac{(1-\alpha)n}{2^k} \pm \varepsilon^{0.01}n \right]$$

It follows that for every  $i \in \{0, \dots, k+1\}$  it holds that:

$$|(a+b)^{-1}(i)| \in \left[ \binom{k}{i} \frac{(1-\alpha)n}{2^k} + \binom{k}{i-1} \frac{\alpha n}{2^k} \pm 2\varepsilon^{0.01}n \right].$$

□

Next we define function  $\eta(a, b) := a + b$ . Observe that  $\eta$  is an injective function on  $A \times k \cdot B$  by Lemma 4.3, and since  $P \subseteq A \times k \cdot B$  we have  $|\eta(P)| = |P|$ . By Lemma 4.8, every vector in  $\eta(P)$  is  $(2\varepsilon^{0.01})$ -Bin( $k, \alpha$ ) balanced, and thus Lemma 4.6 implies

$$|P| \leq 2^{n \cdot h(\text{Bin}(k, \alpha))} \cdot 2^{\mathcal{O}(nk\varepsilon \log(1/\varepsilon))}. \quad (5)$$

### 4.3 Proof of Theorem 4.2

By combining (4) and (5) we obtain the following bound:

$$|A| \leq 2^{n(h(\text{Bin}(k, \alpha)) - h(\text{Bin}(k)))} \cdot 2^{\mathcal{O}((\varepsilon^{0.01} + k\varepsilon \log \frac{1}{\varepsilon})n)}. \quad (6)$$

By Lemma B.4 we have that  $h(\text{Bin}(k, \alpha)) \leq h(\text{Bin}(k+1))$ , and thus it remains to bound the difference in entropy of two consecutive binomial distributions as follows:

**Lemma 4.9.** For large enough  $k$ , we have that  $h(\text{Bin}(k)) - h(\text{Bin}(k-1)) \leq \frac{\log k}{\sqrt{k}}$ .

Before we present the proof of Lemma 4.9, let us see how to use it. We choose  $k := \Theta(\log(1/\varepsilon))$ . Thus Lemma 4.9 implies that

$$|A| \leq 2^{n(\log k/\sqrt{k} + \varepsilon^{0.01} \log(1/\varepsilon))} = 2^{\mathcal{O}(n \cdot \delta(\varepsilon))},$$

where  $\delta(\varepsilon) = \mathcal{O}_{\varepsilon \rightarrow 0} \left( \frac{\log(\log(1/\varepsilon))}{\sqrt{\log(1/\varepsilon)}} \right)$ , because  $\varepsilon^{0.01} \log(1/\varepsilon) \ll \delta(\varepsilon)$  for small enough  $\varepsilon$ . This finishes the proof of Theorem 4.2.

*Proof of Lemma 4.9.* For every  $i, k \in \mathbb{N}$  such that  $i \leq k$  let us define an auxiliary function:

$$f(k, i) := \frac{\binom{k}{i}}{2^k} \log \left( \frac{2^k}{\binom{k}{i}} \right).$$

Thus we have  $h(\text{Bin}(k)) = \sum_{i=0}^k f(k, i)$ . To relate  $h(\text{Bin}(k))$  with  $h(\text{Bin}(k-1))$ , the following will be useful:

**Claim 4.10.**

$$f(k, i) \leq \begin{cases} f(k-1, i), & \text{if } i < \lfloor k/2 \rfloor, \\ f(k-1, i-1), & \text{if } i \geq k/2. \end{cases} \quad (7)$$

*Proof.* Define  $g(x) = x \cdot \log(1/x)$ . Since its derivative is  $g'(x) = -\frac{\ln(x)+1}{\ln(2)}$  we have that  $g(x) \leq g(x')$  whenever  $x \leq x' \leq 1/e$ .

Note  $f(k, i) = g(\binom{k}{i}/2^k)$ , and since  $\binom{k}{i}/2^k \leq 1/\sqrt{k}$  by the standard bound  $\binom{k}{i} \leq 2^k/\sqrt{k}$ , we have  $\binom{k}{i}/2^k \leq 1/e$  for  $k \geq 9$ . Thus to prove the claim it remains to show that

$$\binom{k}{i} 2^{-k} \leq \begin{cases} \binom{k-1}{i} 2^{-(k-1)}, & \text{if } i < \lfloor k/2 \rfloor, \\ \binom{k-1}{i-1} 2^{-(k-1)}, & \text{if } i \geq k/2. \end{cases}$$

To see this first suppose  $i < \lfloor k/2 \rfloor$ . Then we have that

$$\binom{k}{i} 2^{-k} = \binom{k-1}{i} \frac{k}{k-i-1} 2^{-k} \leq \binom{k-1}{i} 2^{-(k-1)}.$$

Second, if  $i \geq k/2$ , then we have that

$$\binom{k}{i} 2^{-k} = \binom{k-1}{i-1} \frac{k-1}{i-1} 2^{-k} \leq \binom{k-1}{i-1} 2^{-(k-1)}.$$

□

Now we can use Claim 4.10 to give the required upper bound:

$$\begin{aligned}
h(\text{Bin}(k)) &= \sum_{i=0}^k f(k, i) \\
&= \left( \sum_{i=0}^{\lfloor k/2 \rfloor - 1} f(k, i) \right) + \left( \sum_{i=\lfloor k/2 \rfloor + 1}^k f(k, i) \right) + f(k, \lfloor k/2 \rfloor) \\
&\leq \left( \sum_{i=0}^{\lfloor k/2 \rfloor - 1} f(k-1, i) \right) + \left( \sum_{i=\lfloor k/2 \rfloor + 1}^k f(k-1, i-1) \right) + f(k, \lfloor k/2 \rfloor) \\
&= h(\text{Bin}(k-1)) + f(k, \lfloor k/2 \rfloor) \\
&\leq h(\text{Bin}(k-1)) + \log(k)/\sqrt{k},
\end{aligned}$$

where we use Claim 4.10 in the first inequality, and  $\binom{k}{\lfloor k/2 \rfloor} \leq 2^k/\sqrt{k}$  in the second inequality. Hence  $h(\text{Bin}(k)) - h(\text{Bin}(k-1)) \leq \log(k)/\sqrt{k}$ .  $\square$

## 5 Properties of $k \cdot B$ : Proof of Lemma 4.7

In this section, we prove the Lemma 4.7.

**Lemma 5.1.** *Let  $k < 0.01 \cdot \log(1/\varepsilon)$ . Then, for every  $a \in \{0, 1\}^n$  with  $|a^{-1}(1)| > \varepsilon^{0.01}n$ , there exists  $E_a \subseteq k \cdot B$ , such that  $|E_a| \geq 2^{(h(\text{Bin}(k)) - \varepsilon^{0.1})n}$  and  $(a, b) \in P$  for every  $b \in E_a$ .*

Recall that

$$P = \{(a, b) \in A \times k \cdot B : b \in B \text{ is } \varepsilon^{0.01}\text{-Bin}(k) \text{ balanced for } a\}.$$

Intuitively, we prove that for any fixed set  $B \subseteq \{0, 1\}^n$  there exists a large set  $E_a \subseteq k \cdot B$  with the following property: for every  $b \in E_a$  we can perturb  $\varepsilon^{0.01}n$  entries in  $b$ , such that it is indistinguishable from a vector randomly sampled from the binomial distribution, even if we focus on a concrete subset of coordinates  $a^{-1}(1) \subseteq \{1, \dots, n\}$ .

First, observe that we can interpret a tuple  $(b_1, \dots, b_k) \in B^k$  as the  $n \times k$  matrix with the  $i$ 'th column equal to  $b_i$ . Note that we interpret an  $x \times y$  matrix as a tuple of  $y$  vectors of dimension  $x$ . We interchangeably address such a tuple as an  $n \times k$  matrix and as an element of  $\{0, 1\}^{n \times k}$ . To emphasize the type of such variables, we denote such matrices with boldface. For example,  $(b_1, \dots, b_k)$  is denoted with  $\mathbf{b} \in \{0, 1\}^{n \times k}$ .

The notation  $\mathbf{b}^T$  denotes the transpose of a matrix. Next,  $C := \{\mathbf{b}^T : \mathbf{b} \in B^k\} \subseteq \{0, 1\}^{k \times n}$  denotes the set of matrices  $B^k$  interpreted in the transposed way.

In Section 5.1 we show how to select a subset  $D \subseteq C$  of matrices in  $C$ , in such a way that for all  $\mathbf{b} \in D \subseteq \{0, 1\}^{k \times n}$ , any column  $z \in \{0, 1\}^k$  occurs  $\frac{n}{2^k} \pm f(\varepsilon)n$  times in  $\mathbf{b}$ , that is  $|\mathbf{b}^{-1}(z)| \in [\frac{n}{2^k} \pm f(\varepsilon)n]$ .

Next, in Section 5.2 we define the operation  $\zeta((a_1, \dots, a_x)) := \sum_{i=1}^x a_i$ , that sums the columns of a matrix  $\mathbf{a} \in \mathbb{Z}^{y \times x}$  to a single column  $\zeta(\mathbf{a}) \in \mathbb{Z}^y$  (see Figure 4). We consider the set  $E := \{\zeta(\mathbf{b}^T) : \mathbf{b} \in D\} \subseteq \{0, \dots, k\}^n$ , and argue that each vector in  $E$  is  $\gamma$ -Bin( $k$ ) balanced for some small  $\gamma > 0$ .

Finally, in Section 5.3 we take care of  $a \in \{0, 1\}^n$  and select the set  $E_a \subseteq E$  to be all vectors in  $E$  that are  $\varepsilon^{0.01}$ -Bin( $k$ ) balanced for  $a^{-1}(1)$ .

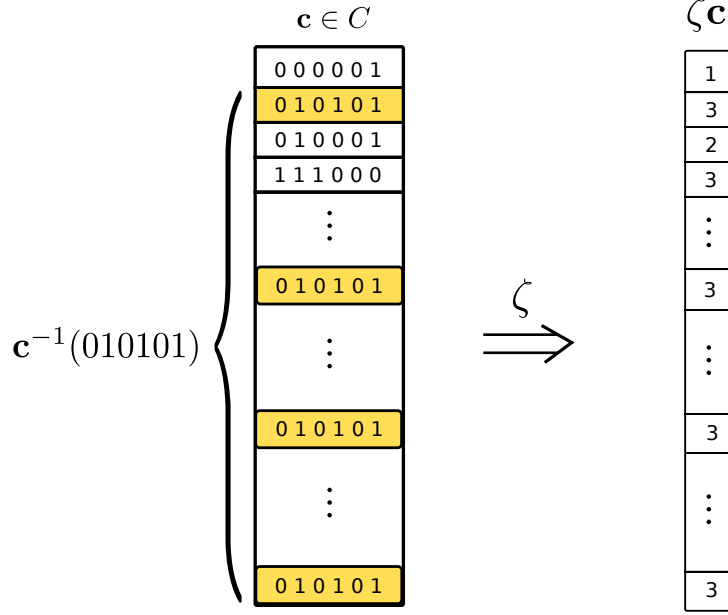


Figure 4: The  $\zeta$  operation takes an  $x \times y$  matrix  $\mathbf{c} \in \mathbb{Z}^{x \times y}$  as input and outputs a vector  $\zeta(\mathbf{c}) \in \mathbb{Z}^x$  by adding all columns.

**Uniform distribution.** We define the *uniform distribution* to be  $\text{Uni}(\Omega) = (\Omega, p)$  if  $p(\omega) = \frac{1}{|\Omega|}$  for each  $\omega \in \Omega$ . We will focus on the special cases when  $\Omega = \{0, 1\}$  and  $\Omega = \{0, 1\}^k$ . Thus,  $v \in (\text{Uni}(\{0, 1\}) \pm \gamma)^{\{1, \dots, n\}}$  means that  $v^{-1}(i) \in [n/2 \pm \gamma n]$  for all  $i \in \{0, 1\}$ . Similarly,  $\mathbf{v} \in (\text{Uni}(\{0, 1\}^k) \pm \gamma)^{\{1, \dots, n\}}$  means that  $\mathbf{v}^{-1}(z) \in [n/2^k \pm \gamma n]$  for all  $z \in \{0, 1\}^k$ . If  $\Omega$  is clear from the context,  $v \in \Omega^U$  and  $X \subseteq U$ , we also say that a vector is  $\gamma$ -uniform for  $X$  to refer to the statement that it is  $\gamma$ - $\text{Uni}(\Omega)$  balanced for  $X$ .

**Inequalities.** Through the section we assume that  $\varepsilon \leq 1/2^{4k}$ ,  $\gamma = 4\sqrt{\varepsilon}$  and  $k > 100$  is an integer. This means that the following inequalities hold:

$$\varepsilon^{1/2} \leq 2^k \gamma \leq \varepsilon^{1/4}, \quad (8)$$

$$100 \cdot k 2^k \gamma \log(1/(2^k \gamma)) \leq \varepsilon^{1/5}. \quad (9)$$

## 5.1 Constructing a set $D$ of uniform $k$ -tuples

We first prove the following result that will be helpful to obtain the aforementioned set  $C$ .

**Lemma 5.2** (Most vectors in  $B$  are uniform). *Let  $U_1 \uplus \dots \uplus U_\ell = \{1, \dots, n\}$  be a partition such that  $|U_i| \geq \mu n$  for all  $i \in [\ell]$ . Let  $\lambda \in (0, 1/2)$ . For every  $B \subseteq \{0, 1\}^n$  with  $|B| \geq 2^{(1-\mu\lambda^2)n-o(n)}$  it holds that:*

$$|\{b \in B : b \text{ is } \lambda\text{-uniform in } U_i \text{ for every } i\}| \geq |B|/2.$$

*Proof.* For a fixed  $i$  we argue that the number of vectors that are not  $\lambda$ -uniform for  $U_i$  is bounded by  $2^{(1-\lambda^2\mu)n+o(n)}$ . This will finish the proof since we can sum this bound over all partitions.

Let  $s = |U_i|/n$  and note that  $s \geq \mu$ . Observe that the number of vectors  $v \in \{0, 1\}^n$  such that  $|v^{-1}(1) \cap U_i| \notin [sn/2 \pm \lambda sn]$  is at most:

$$\sum_{\lambda' \notin [-\lambda, \lambda]} \binom{sn}{sn/2 - \lambda' sn} 2^{n-sn}$$

because a vector  $v$  that is not  $\lambda$ -uniform on  $U_i$  can be arbitrary in  $\{1, \dots, n\} \setminus U_i$ . We upper bound this with the binary entropy function

$$\sum_{0 \leq \lambda' \leq \lambda} 2^{sn \cdot h(1/2 - \lambda') + o(n)} \cdot 2^{n-sn}.$$

The expression is maximized when  $\lambda' = \lambda$  because the  $h(p)$  entropy function is increasing in  $[0, 1/2]$ . Hence we can upper-bound the expression with

$$n 2^{(1+s(h(1/2-\lambda)-1))n+o(n)}.$$

Now, we use a bound  $h(1/2 - x) \leq 1 - x^2$  when  $0 \leq x \leq 1/2$  (recall that  $\lambda \in [0, 1/2]$ ) and obtain that the number of vectors that are not  $\lambda$ -uniform for  $U_i$  is at most

$$2^{(1-s\lambda^2)n+o(n)} \leq 2^{(1-\mu\lambda^2)n+o(n)}.$$

Thus, by summing over all  $U_i$ , the number of vectors that are not  $\lambda$ -uniform for some  $U_i$  is at most  $2^{(1-\mu\lambda^2)n+o(n)}$ , and the number of vectors in  $B$  that are  $\lambda$ -uniform for all  $U_i$  is at least

$$|B| - 2^{(1-\mu\lambda^2)n+o(n)} \geq |B|/2,$$

and the claim follows. □

Set a balance parameter  $\gamma = 4\sqrt{\varepsilon}$ , and define

$$D := C \cap (\text{Uni}(\{0, 1\}^k) \pm \gamma)^{\{1, \dots, n\}}.$$

**Lemma 5.3** (Most  $k$ -tuples are uniform). *Let  $k \in \mathbb{N}$  be such that  $\varepsilon < 1/4^{k+2}$ . Then it holds that*

$$|D| \geq \left(\frac{|B|}{2}\right)^k.$$

*Proof.* We will use set of matrices  $C_j \subseteq \{0, 1\}^{j \times n}$  to select all matrices obtained by keeping the first  $j$  columns of matrices of  $C$ , namely

$$C_j := \{\mathbf{b}^T : \mathbf{b} \in B^j\}.$$

Thus,  $C = C_k$ . For  $j \in \{1, \dots, k\}$ , let

$$D_j := C_j \cap (\text{Uni}(\{0, 1\}^j) \pm \gamma)^{\{1, \dots, n\}}.$$

We prove that  $|D_j| \geq (|B|/2)^j$  by induction on  $k$ . First, we prove the base case  $j = 1$  of the induction, so  $|D_1| \geq |B|/2$ . This follows by applying Lemma 5.2 with  $\lambda = \sqrt{\varepsilon}$  and partition  $U_1 = \{1, \dots, n\}$ , since it implies that

$$|B|/2 \leq \left| B \cap (\text{Uni}(\{0, 1\}) \pm \sqrt{\varepsilon})^{U_1} \right| \leq \left| C_1 \cap (\text{Uni}(\{0, 1\}) \pm \gamma)^{U_1} \right| = |D_1|.$$

The induction step with  $j > 1$  is a direct consequence of the following claim, which therefore is sufficient to finish the proof.

**Claim 5.4.** Let  $\mathbf{b} \in D_{j-1}$ . Then there are at least  $|B|/2$  vectors  $b_j \in B$ , such that  $\mathbf{b}_+ \in (\text{Uni}(\{0, 1\}^j) \pm \gamma)^{\{1, \dots, n\}}$ , where  $\mathbf{b}_+$  is obtained from  $\mathbf{b}$  by appending  $b_j$  as the  $j$ 'th row to it.

*Proof.* Define a partition  $\{U_z\}_{z \in \{0,1\}^{j-1}}$  of  $\{1, \dots, n\}$  by  $U_z = \mathbf{b}^{-1}(z)$ . Because  $\mathbf{b} \in (\text{Uni}(\{0, 1\}^{(j-1)}) \pm \gamma)^{\{1, \dots, n\}}$  we know that:

$$\mu = \min_{z \in \{0,1\}^{j-1}} |U_z|/n \geq \frac{1}{2^{j-1}} - \gamma.$$

Note, that  $\mu > 1/2^j$  because we assumed that  $\varepsilon < 1/4^{k+2}$  (hence  $\gamma < 1/2^j$ ). Now, we use Lemma 5.2 with partition  $\{U_z\}_{z \in \{0,1\}^{j-1}}$  and  $\lambda = 2^{j-3} \cdot \gamma$ . First let us assert that the condition  $|B| \geq 2^{(1-\mu\lambda^2)n+o(n)}$  holds. Recall that we assumed  $|B| \geq 2^{(1-\varepsilon)n+o(n)}$  and  $\mu\lambda^2 \geq \frac{1}{2^j} (2^{j-3} \cdot 4\sqrt{\varepsilon})^2 \geq 2^{j-2}\varepsilon \geq \varepsilon$  (for  $j \geq 2$ ). Hence  $|B| \geq 2^{(1-\varepsilon)n} \geq 2^{(1-\mu\lambda^2)n+f(n)}$  for some function  $f \in o(n)$ .

Lemma 5.2 states that there are at least  $|B|/2$  vectors  $b_j \in B$  such that for each  $z \in \{0, 1\}^{j-1}$

$$|U_z \cap b_j^{-1}(1)| \in \left[ \frac{|U_z|}{2} \pm \lambda|U_z| \right]. \quad (10)$$

We know that  $|U_z| \in [\frac{n}{2^{j-1}} \pm \gamma n]$  (because  $\mathbf{b} \in (\text{Uni}(\{0, 1\}^{j-1}) \pm \gamma)^{\{1, \dots, n\}}$ ). Thus in fact (10) can be rewritten to

$$|U_z \cap b_j^{-1}(1)| \in \left[ \frac{n}{2^j} \pm (\lambda|U_z| + (\gamma/2)n) \right].$$

We bound  $\lambda|U_z|$  by

$$\begin{aligned} \lambda|U_z| &= 2^{j-3}\gamma|U_z| \leq 2^{j-3}\gamma \left( \frac{n}{2^{j-1}} + \gamma n \right) = \gamma n \left( \frac{1}{4} + 2^{j-3}\gamma \right) = \gamma n \left( \frac{1}{4} + 2^{j-3}4\sqrt{\varepsilon} \right) \\ &< \gamma n \left( \frac{1}{4} + 2^{j-3}4\sqrt{1/4^{k+2}} \right) = \gamma n \left( \frac{1}{4} + 2^{j-1}/2^{k+2} \right) < \gamma n \left( \frac{1}{4} + \frac{1}{4} \right) = (\gamma/2)n, \end{aligned}$$

where we use the assumption  $\varepsilon \leq \frac{1}{4^{k+2}}$  in the second line of the inequality. Thus, for every  $z \in \{0, 1\}^{j-1}$  we have

$$|U_z \cap b_j^{-1}(1)| \in \left[ \frac{n}{2^j} \pm 2(\gamma/2)n \right].$$

Now, observe that for all  $z \in \{0, 1\}^{j-1}$  it holds that:

$$U_z \cap b_j^{-1}(1) = \mathbf{b}^{-1}(z) \cap b_j^{-1}(1) = \mathbf{b}_+^{-1}(z'),$$

where  $z' \in \{0, 1\}^j$  is the vector obtained from  $z$  by adding a  $j$ -th entry with value 1. Thus vector  $z'$  fulfills the condition for  $\mathbf{b}_+$  to be in  $(\text{Uni}(\{0, 1\}^j) \pm \gamma)^{\{1, \dots, n\}}$ . Similarly, we can prove this condition by concatenating a 0 to the vector  $z$ . Hence, for every  $\mathbf{b} \in D_{j-1}$  there are at least  $|B|/2$  vectors  $b_j \in B$ , such that  $\mathbf{b}_+ \in (\text{Uni}(\{0, 1\}^j \pm \gamma)^{\{1, \dots, n\}}$ .  $\square$

Thus this claim proves our induction hypothesis and hence the lemma.  $\square$

## 5.2 Summing tuples from $D$ gives many distinct sums

As mentioned at the beginning of this section we define the operation  $\zeta((a_1, \dots, a_x)) := \sum_{i=1}^x a_i$ , that sums the columns of a matrix  $\mathbf{a} \in \mathbb{Z}^{y \times x}$  to a single column  $\zeta(\mathbf{a}) \in \mathbb{Z}^y$  (see Figure 4).

We define  $E$  to be all sums of tuples from  $D$ :

$$E := \{\zeta(\mathbf{b}^T) : \mathbf{b} \in D\} \subseteq \{0, \dots, k\}^n$$



In fact, by the assumption on  $D$  we have the following control on the distributions of the values in the vectors in  $E$ :

**Lemma 5.5.** *If  $v \in E$ , then for  $j \in \{0, \dots, k\}$  it holds that  $|v^{-1}(j)| \in \left[ \binom{k}{j} \frac{n}{2^k} \pm \binom{k}{j} \gamma n \right]$ , i.e., every vector in  $E$  is a  $(2^k \gamma)$ -Bin( $k$ ) balanced vector.*

*Proof.* Consider an arbitrarily vector  $v \in E$  and fix  $j \in \{0, \dots, k\}$ . From the definition of  $E$ , there exists a vector  $\mathbf{b} \in D = C \cap (\text{Uni}(\{0, 1\}^k) \pm \gamma)^{\{1, \dots, n\}}$  such that  $\zeta(\mathbf{b}^T) = v$ . Hence for every  $z \in \{0, 1\}^k$ :

$$|\mathbf{b}^{-1}(z)| \in \left[ \frac{n}{2^k} \pm \gamma n \right].$$

Hence, if we sum over all vectors  $z \in \{0, 1\}^k$  such that  $|z^{-1}(1)| = j$  we have:

$$|v^{-1}(j)| \leq \sum_{\substack{z \in \{0, 1\}^k \\ |z^{-1}(1)| = j}} \frac{n}{2^k} + \gamma n = \binom{k}{j} \frac{n}{2^k} + \binom{k}{j} \gamma n,$$

and analogously  $|v^{-1}(j)| \geq \binom{k}{j} \frac{n}{2^k} - \binom{k}{j} \gamma n$ . Thus indeed  $v$  is a  $(2^k \gamma)$ -Bin( $k$ ) balanced vector, as desired.  $\square$

We now show that  $E$  is sufficiently large:

**Lemma 5.6.** *It holds that  $|E| \geq 2^{(h(\text{Bin}(k)) - \varepsilon^{0.2})n}$ .*

*Proof.* For a vector  $v \in E$  we define

$$D_v := \{\mathbf{b} \in D : \zeta(\mathbf{b}^T) = v\}.$$

By grouping all elements of  $D$  on their image with respect to  $\zeta$ :

$$|D| = \sum_{v \in E} |D_v| \leq |E| \max_{v \in E} |D_v|,$$

which can be rewritten into the following lower bound on  $|E|$ :

$$|E| \geq |D| / \max_{v \in E} |D_v| \geq (|B|/2)^k / \max_{v \in E} |D_v| \geq 2^{k(n - \varepsilon n + 1)} / \max_{v \in E} |D_v|. \quad (11)$$

Thus in the remainder of the proof we can focus on showing that for any vector  $v \in E$ ,  $|D_v| \leq 2^{n(k - h(\text{Bin}(k)) + \varepsilon^{0.2})}$ ; the Lemma would then follow by the bound in (11).

Let  $\mathbf{b} \in D_v$ . This means that for every  $j \in \{0, \dots, k\}$ :

$$\bigcup_{\substack{z \in \{0, 1\}^k \\ |z^{-1}(1)| = j}} \mathbf{b}^{-1}(z) = v^{-1}(j).$$

Thus the number of possibilities for  $\mathbf{b}$  is

$$\prod_{j=0}^k \binom{k}{j}^{|v^{-1}(j)|}.$$

We multiply this quantity with  $\binom{\{1, \dots, n\}}{|v^{-1}(0)|, \dots, |v^{-1}(k)|}$  and obtain

$$\binom{\{1, \dots, n\}}{|v^{-1}(0)|, \dots, |v^{-1}(k)|} \cdot \prod_{j=0}^k \binom{k}{j}^{|v^{-1}(j)|} \leq 2^{kn},$$

where the inequality follows since the left-hand side counts partitions of  $\{1, \dots, n\}$  into  $\sum_{i=0}^k \binom{k}{i} = 2^k$  parts. Now, let  $\mathbf{r} = (|v^{-1}(0)|/n, \dots, |v^{-1}(k)|/n)$  and observe that by Lemma 2.9 we have  $|D_v| \leq 2^{n(k-h(\mathbf{r}))}$ . Because  $v$  is  $\gamma$ -Bin( $k$ ) balanced (since it is in  $E$ ), we have

$$h(\mathbf{r}) \geq h(\text{Bin}(k)) - \frac{1}{\ln(2)} \cdot k2^k \gamma \log \frac{1}{2^{k\gamma}} \geq h(\text{Bin}(k)) - \varepsilon^{0.2},$$

where the first inequality is by Lemma 2.10, and the second inequality uses that  $\gamma = 4\sqrt{\varepsilon}$  (see Inequality 9).  $\square$

### 5.3 Selecting a set $E_a \subseteq E$ for every $a \in A$

**Lemma 5.7.** *Let  $\mathcal{D} = (\Omega, p)$  be a discrete probability space, and let  $X \subseteq \{1, \dots, n\}$  with  $|X| = \alpha n$ . The number of vectors  $v \in (\mathcal{D} + 0)^{\{1, \dots, n\}}$  that are not  $\rho$ - $\mathcal{D}$  balanced for  $X$  and  $\{1, \dots, n\} \setminus X$  is at most  $2^{n(h(\mathcal{D}) - \alpha^2 \min(\rho^2, \log \alpha))}$ .*

*Proof.* We define a relation  $R \subseteq (\mathcal{D} + 0)^{\{1, \dots, n\}} \times \binom{\{1, \dots, n\}}{\alpha n}$  as follows:

$$(v, X) \in R \Leftrightarrow v \text{ is not } \rho\text{-}\mathcal{D} \text{ balanced for } X.$$

Additionally, let

$$\begin{aligned} R_v &= R \cap \left( \{v\} \times \binom{\{1, \dots, n\}}{\alpha n} \right), & \text{for } v \in (\mathcal{D} + 0)^{\{1, \dots, n\}}, \\ R_X &= R \cap \left( (\mathcal{D} + 0)^{\{1, \dots, n\}} \times \{X\} \right), & \text{for } X \in \binom{\{1, \dots, n\}}{\alpha n}. \end{aligned}$$

Note that  $|R_X|$  is the value we want to bound. Note that the mapping  $(v, X) \mapsto (v \circ \pi, \pi(X))$  for any permutation  $\pi : \{1, \dots, n\} \leftrightarrow \{1, \dots, n\}$  of the index set  $\{1, \dots, n\}$  is an automorphism of  $R$  (i.e.,  $(v, X) \in R$  if and only if  $(\pi(v), \pi(X)) \in R$ ). Therefore, we have

$$|R| = |(\mathcal{D} + 0)^{\{1, \dots, n\}}| \cdot |R_v| = |R_X| \binom{n}{\alpha n} \quad (12)$$

for a fixed  $v$  and  $X$ . By (12) we can focus on bounding  $|R_v|$  instead of  $|R_X|$ . To do so, note that if  $(v, X) \in R$  for  $X \in \binom{\{1, \dots, n\}}{\alpha n}$ , there must exist  $\omega \in \Omega$  such that  $|X \cap v^{-1}(\omega)| \notin [p(\omega)\alpha n \pm \rho\alpha n]$ . We can construct any such  $X$  by first selecting a subset of  $v^{-1}(\omega)$  (which has cardinality  $p(\omega)n$ ), and then choosing the remaining elements. Hence:

$$|R_v| \leq \sum_{\omega \in \Omega} \sum_{x \notin [-\rho, \rho]} \binom{p(\omega)n}{(p(\omega) + x)\alpha n} \binom{(1 - p(\omega))n}{(1 - p(\omega) - x)\alpha n}.$$

Next, we use Lemma B.3. In our case (with parameters  $\beta = p(\omega)$ ,  $\alpha = \alpha$  and  $\gamma = -\alpha x$ ) it implies:

$$|R_v| \leq \sum_{j \in \{0, \dots, k\}} \sum_{x \notin [-\rho, \rho]} \binom{n}{\alpha n} 2^{-c \cdot n}$$

where

$$c = \begin{cases} (\alpha x)^2, & \text{if } |\alpha x| < \alpha(1 - \alpha)p(\omega), \\ \alpha^2 \log(1/(2\alpha)), & \text{otherwise.} \end{cases}$$

Since  $|x| \geq \rho$  we have  $c \geq \alpha^2 \min\{\rho^2, \log(1/(2\alpha))\}$ , thus

$$|R_v| \leq \binom{n}{\alpha n} 2^{-\alpha^2 \min\{\rho^2, \log(1/(2\alpha))\}n},$$

which plugged into (12) gives the desired inequality.  $\square$

*Proof of Lemma 4.7.* Recall that we assume that  $\alpha > \varepsilon^{0.01}$ . Lemma 5.6 gives us a lower bound on  $E$ , and by Lemma 5.5 each vector in  $E$  is a  $(2^k \gamma)$ -Bin( $k$ ) balanced vector. Hence, by the pigeonhole principle there exists a distribution  $\mathcal{D} = (\{0, 1, \dots, k\}, p)$  where  $p = (p_0, \dots, p_k)$  such that  $|p_0 - \binom{k}{j} 2^{-k}| \leq 2^k \gamma$  for each  $j$  and  $E$  has a subset  $E'$  of at least  $|E|/n^k$  vectors that are in  $\mathcal{D}^{\{1, \dots, n\}}$ . Hence:

$$|E'| \geq |E|/n^k \geq 2^{(h(\mathcal{D}) - \varepsilon^{0.2}n) - o(n)}$$

Now for each  $a \in A$ , define  $E_a$  to be all vectors in  $E'$  that are  $\varepsilon^{0.05}$ - $\mathcal{D}$  balanced for  $a^{-1}(1)$ . Observe that this means that vectors in  $E_a$  are  $\varepsilon^{0.01}$ -Bin( $k$ ) balanced (because  $\varepsilon^{0.05} + 2^k \gamma \ll \varepsilon^{0.01}$ ).

Applying Lemma 5.7 with  $E'$  and  $a^{-1}(1)$ , we get that there are at most  $2^{h(\mathcal{D}) - \alpha^2 \min(\varepsilon^{0.1}, \log(1/(2\alpha)))} \leq 2^{(h(\mathcal{D}) - \varepsilon^{0.12}n)}$  vectors in  $E'$  that are not  $\varepsilon^{0.05}$ - $\mathcal{D}$  balanced. Hence:

$$|E' \setminus E_a| \leq 2^{(h(\mathcal{D}) - \varepsilon^{0.12}n)} \leq |E'|/2$$

Now the lemma follows because

$$|E_a| \geq |E'|/2 \geq 2^{(h(\mathcal{D}) - \varepsilon^{0.2}n) - o(n)} \geq 2^{(h(\text{Bin}(k) - \frac{1}{\ln(2)}) \cdot 2^k \gamma \log(1/(2^k \gamma)) - \varepsilon^{0.2}n)} \geq 2^{(h(\text{Bin}(k) - \varepsilon^{0.1})n)}$$

where the last inequality follows from Lemma 2.10 and Inequality 9 (since  $\varepsilon$  is small enough).  $\square$

## 6 Conclusion and Open Problems

In this paper, we present a randomized  $\mathcal{O}(2^{(1-\sigma_m)n})$  time algorithm for the Bin Packing problem, where  $\sigma_m > 0$  and  $m$  denotes the number of bins. This is an improvement over the state-of-the-art algorithm of Björklund et al. [9] that runs in  $\mathcal{O}^*(2^n)$  time for small  $m$ . Nevertheless, it remains to give an algorithm for Bin Packing that works in  $\mathcal{O}^*((2 - \varepsilon)^n)$  time for an unbounded number of bins for some fixed constant  $\varepsilon > 0$ . We believe our algorithm made significant progress on this question. One open end for further research is how the number of bins influences the complexity of an instance. By the methods of [46], instances of Bin Packing with a linear number of bins (with equal capacity) can also be solved in time  $\mathcal{O}(2^{(1-\varepsilon)n})$  based on a witness sampling technique similar to what we used in some of our cases. It is thus natural to wonder whether (an extension) of the methods presented in this paper are enough to give improved algorithms for all numbers of bins.

We believe our Additive Combinatorics result is natural and may have applications beyond the scope of this paper. As mentioned in the introduction, Littlewood–Offord’s theory has a wide variety

of applications, and it is natural to expect that the setting that we address may be of interest in any of these settings.

In the introduction we mentioned Question 1 as one motivation for studying improved exact exponential time algorithms for the Bin Packing problem. While it is not clear if we made direct progress on this question, we do believe that some of our ideas such as the approach to narrow down the number of witnesses may inspire future work on improved algorithms for Set Cover.

## Acknowledgement

The research leading to the results presented in this paper was partially carried out during the Parameterized Algorithms Retreat of the University of Warsaw, PARUW 2020, held in Krynica-Zdrój in February 2020. This workshop was supported by a project that has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme under grant agreement No 714704 (PI: Marcin Pilipczuk).

## References

- [1] A. Abboud. Fine-grained reductions and quantum speedups for dynamic programming. In C. Baier, I. Chatzigiannakis, P. Flocchini, and S. Leonardi, editors, *46th International Colloquium on Automata, Languages, and Programming, ICALP 2019, July 9-12, 2019, Patras, Greece*, volume 132 of *LIPICs*, pages 8:1–8:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.
- [2] P. Austrin, P. Kaski, M. Koivisto, and J. Nederlof. Subset Sum in the Absence of Concentration. In E. W. Mayr and N. Ollinger, editors, *32nd International Symposium on Theoretical Aspects of Computer Science, STACS 2015, March 4-7, 2015, Garching, Germany*, volume 30 of *LIPICs*, pages 48–61. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2015.
- [3] P. Austrin, P. Kaski, M. Koivisto, and J. Nederlof. Dense subset sum may be the hardest. In N. Ollinger and H. Vollmer, editors, *33rd Symposium on Theoretical Aspects of Computer Science, STACS 2016, February 17-20, 2016, Orléans, France*, volume 47 of *LIPICs*, pages 13:1–13:14. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2016.
- [4] P. Austrin, P. Kaski, M. Koivisto, and J. Nederlof. Sharper upper bounds for unbalanced uniquely decodable code pairs. *IEEE Trans. Inf. Theory*, 64(2):1368–1373, 2018.
- [5] N. Bansal, S. Garg, J. Nederlof, and N. Vyas. Faster Space-Efficient Algorithms for Subset Sum, k-Sum, and Related Problems. *SIAM J. Comput.*, 47(5):1755–1777, 2018.
- [6] A. Becker, J. Coron, and A. Joux. Improved Generic Algorithms for Hard Knapsacks. In K. G. Paterson, editor, *Advances in Cryptology - EUROCRYPT 2011 - 30th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tallinn, Estonia, May 15-19, 2011. Proceedings*, volume 6632 of *Lecture Notes in Computer Science*, pages 364–385. Springer, 2011.
- [7] A. Björklund, T. Husfeldt, P. Kaski, and M. Koivisto. Fourier meets Möbius: fast subset convolution. In D. S. Johnson and U. Feige, editors, *Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, California, USA, June 11-13, 2007*, pages 67–74. ACM, 2007.

- [8] A. Björklund, T. Husfeldt, P. Kaski, and M. Koivisto. Counting Paths and Packings in Halves. In A. Fiat and P. Sanders, editors, *Algorithms - ESA 2009, 17th Annual European Symposium, Copenhagen, Denmark, September 7-9, 2009. Proceedings*, volume 5757 of *Lecture Notes in Computer Science*, pages 578–586. Springer, 2009.
- [9] A. Björklund, T. Husfeldt, and M. Koivisto. Set Partitioning via Inclusion-Exclusion. *SIAM J. Comput.*, 39(2):546–563, 2009.
- [10] A. Björklund, P. Kaski, and I. Koutis. Directed Hamiltonicity and Out-Branchings via Generalized Laplacians. In I. Chatzigiannakis, P. Indyk, F. Kuhn, and A. Muscholl, editors, *44th International Colloquium on Automata, Languages, and Programming, ICALP 2017, July 10-14, 2017, Warsaw, Poland*, volume 80 of *LIPICs*, pages 91:1–91:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017.
- [11] C. Calabro. *The exponential complexity of satisfiability problems*. PhD thesis, UC San Diego, 2009.
- [12] E. G. Coffman Jr., J. Csirik, G. Galambos, S. Martello, and D. Vigo. *Bin Packing Approximation Algorithms: Survey and Classification*, pages 455–531. Springer New York, New York, NY, 2013.
- [13] I. Csiszár and P. C. Shields. *Information theory and statistics: A tutorial*. Now Publishers Inc, 2004.
- [14] M. Cygan, H. Dell, D. Lokshtanov, D. Marx, J. Nederlof, Y. Okamoto, R. Paturi, S. Saurabh, and M. Wahlström. On problems as hard as CNF-SAT. In *Proceedings of the 27th Conference on Computational Complexity, CCC 2012, Porto, Portugal, June 26-29, 2012*, pages 74–84. IEEE Computer Society, 2012.
- [15] M. Cygan, H. Dell, D. Lokshtanov, D. Marx, J. Nederlof, Y. Okamoto, R. Paturi, S. Saurabh, and M. Wahlström. On problems as hard as CNF-SAT. *ACM Trans. Algorithms*, 12(3):41:1–41:24, 2016.
- [16] M. Cygan, F. V. Fomin, L. Kowalik, D. Lokshtanov, D. Marx, M. Pilipczuk, M. Pilipczuk, and S. Saurabh. *Parameterized Algorithms*. Springer, 2015.
- [17] M. Delorme, M. Iori, and S. Martello. Bin packing and cutting stock problems: Mathematical models and exact algorithms. *European Journal of Operational Research*, 255(1):1–20, 2016.
- [18] I. Diakonikolas and R. A. Servedio. Improved approximation of linear threshold functions. *Comput. Complex.*, 22(3):623–677, 2013.
- [19] S. Eilon and N. Christofides. The loading problem. *Management Science*, 17(5):259–268, 1971.
- [20] K. Eisemann. The trim problem. *Management Science*, 3(3):279–284, 1957.
- [21] M. Etscheid, S. Kratsch, M. Mnich, and H. Röglin. Polynomial kernels for weighted problems. *J. Comput. Syst. Sci.*, 84:1–10, 2017.
- [22] F. V. Fomin and P. Kaski. Exact exponential algorithms. *Commun. ACM*, 56(3):80–88, 2013.
- [23] F. V. Fomin and D. Kratsch. *Exact Exponential Algorithms*. Texts in Theoretical Computer Science. An EATCS Series. Springer, 2010.

- [24] A. Frank and É. Tardos. An application of simultaneous Diophantine approximation in combinatorial optimization. *Combinatorica*, 7(1):49–65, 1987.
- [25] M. X. Goemans and T. Rothvoß. Polynomiality for bin packing with a constant number of item types. In C. Chekuri, editor, *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 830–839. SIAM, 2014.
- [26] A. Golovnev, A. S. Kulikov, and I. Mihajlin. Families with infants: A general approach to solve hard partition problems. In J. Esparza, P. Fraigniaud, T. Husfeldt, and E. Koutsoupias, editors, *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part I*, volume 8572 of *Lecture Notes in Computer Science*, pages 551–562. Springer, 2014.
- [27] A. Golovnev, A. S. Kulikov, and I. Mihajlin. Families with Infants: Speeding Up Algorithms for NP-Hard Problems Using FFT. *ACM Trans. Algorithms*, 12(3):35:1–35:17, 2016.
- [28] J. R. Griggs. Database security and the distribution of subset sums in  $\mathbb{R}^m$ . In *Graph Theory and Combinatorial Biology*, 1998.
- [29] G. Halász. Estimates for the concentration function of combinatorial number theory and probability. *Periodica Mathematica Hungarica*, 8(3-4):197–211, 1977.
- [30] R. Hoberg and T. Rothvoss. A logarithmic additive integrality gap for bin packing. In P. N. Klein, editor, *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, Barcelona, Spain, Hotel Porta Fira, January 16-19*, pages 2616–2625. SIAM, 2017.
- [31] E. Horowitz and S. Sahni. Computing Partitions with Applications to the Knapsack Problem. *J. ACM*, 21(2):277–292, 1974.
- [32] V. Jain, A. Sah, and M. Sawhney. Anticoncentration versus the number of subset sums. *Advances in Combinatorics*, page 24872, 2021.
- [33] K. Jansen, S. Kratsch, D. Marx, and I. Schlotter. Bin packing with fixed number of bins revisited. *J. Comput. Syst. Sci.*, 79(1):39–49, 2013.
- [34] D. S. Johnson. *Near-optimal bin packing algorithms*. PhD thesis, Massachusetts Institute of Technology, 1973.
- [35] D. M. Kane and R. Williams. Super-linear gate and super-quadratic wire lower bounds for depth-two and depth-three threshold circuits. In D. Wichs and Y. Mansour, editors, *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 633–643. ACM, 2016.
- [36] L. V. Kantorovich. Mathematical methods of organizing and planning production. *Management science, English Translation of a 1939 paper written in Russian*, 6(4):366–422, 1960.
- [37] N. Karmarkar and R. M. Karp. An efficient approximation scheme for the one-dimensional bin-packing problem. In *23rd Annual Symposium on Foundations of Computer Science (sfcs 1982)*, pages 312–320. IEEE, 1982.
- [38] J. M. Kleinberg and É. Tardos. *Algorithm design*. Addison-Wesley, 2006.

- [39] M. Koivisto. Partitioning into sets of bounded cardinality. In J. Chen and F. V. Fomin, editors, *Parameterized and Exact Computation, 4th International Workshop, IWPEC 2009, Copenhagen, Denmark, September 10-11, 2009, Revised Selected Papers*, volume 5917 of *Lecture Notes in Computer Science*, pages 258–263. Springer, 2009.
- [40] R. Krauthgamer and O. Trabelsi. The Set Cover Conjecture and Subgraph Isomorphism with a Tree Pattern. In R. Niedermeier and C. Paul, editors, *36th International Symposium on Theoretical Aspects of Computer Science (STACS 2019)*, volume 126 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 45:1–45:15, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- [41] C. Lenté, M. Liedloff, A. Soukhal, and V. T’Kindt. On an extension of the Sort & Search method with application to scheduling theory. *Theor. Comput. Sci.*, 511:13–22, 2013.
- [42] J. E. Littlewood and A. C. Offord. On the number of real roots of a random algebraic equation. *Journal of the London Mathematical Society*, s1-13(4):288–295, 1938.
- [43] S. Martello and P. Toth. *Knapsack Problems: Algorithms and Computer Implementations*. Wiley Series in Discrete Mathematics and Optimization. Wiley, 1990.
- [44] R. Meka, O. Nguyen, and V. Vu. Anti-concentration for polynomials of independent random variables. *Theory Comput.*, 12(1):1–17, 2016.
- [45] M. Mucha, J. Nederlof, J. Pawlewicz, and K. Węgrzycki. Equal-Subset-Sum Faster Than the Meet-in-the-Middle. In M. A. Bender, O. Svensson, and G. Herman, editors, *27th Annual European Symposium on Algorithms, ESA 2019, September 9-11, 2019, Munich/Garching, Germany*, volume 144 of *LIPIcs*, pages 73:1–73:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.
- [46] J. Nederlof. Finding Large Set Covers Faster via the Representation Method. In P. Sankowski and C. D. Zaroliagis, editors, *24th Annual European Symposium on Algorithms, ESA 2016, August 22-24, 2016, Aarhus, Denmark*, volume 57 of *LIPIcs*, pages 69:1–69:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016.
- [47] J. Nederlof, E. J. van Leeuwen, and R. van der Zwaan. Reducing a Target Interval to a Few Exact Queries. In B. Rován, V. Sassone, and P. Widmayer, editors, *Mathematical Foundations of Computer Science 2012 - 37th International Symposium, MFCS 2012, Bratislava, Slovakia, August 27-31, 2012. Proceedings*, volume 7464 of *Lecture Notes in Computer Science*, pages 718–727. Springer, 2012.
- [48] T. Rothvoß. Approximating bin packing within  $O(\log OPT \cdot \log \log OPT)$  bins. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 20–29, 2013.
- [49] M. Rudelson and R. Vershynin. The Littlewood–Offord problem and invertibility of random matrices. *Advances in Mathematics*, 218(2):600 – 633, 2008.
- [50] C. Schlegel and A. Grant. *Coordinated multiuser communications*. Springer, 2006.
- [51] T. Tao and V. H. Vu. *Additive combinatorics*, volume 105 of *Cambridge studies in advanced mathematics*. Cambridge University Press, 2007.
- [52] K. Tikhomirov. Singularity of random Bernoulli matrices. *Annals of Mathematics*, 191(2):593–634, 2020.

- [53] H. C. A. van Tilborg. An upper bound for codes in a two-access binary erasure channel (corresp.). *IEEE Trans. Inf. Theory*, 24(1):112–116, 1978.
- [54] V. V. Williams and R. R. Williams. Subcubic Equivalences Between Path, Matrix, and Triangle Problems. *J. ACM*, 65(5):27:1–27:38, 2018.
- [55] M. Wiman. Improved Constructions of Unbalanced Uniquely Decodable Code Pairs, 2017. Bachelor Thesis KTH.
- [56] O. Zamir. Breaking the  $2^n$  Barrier for 5-Coloring and 6-Coloring. In N. Bansal, E. Merelli, and J. Worrell, editors, *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, volume 198 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 113:1–113:20, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

## A Computing the Number of Distinct Sums and Critical Pruner

**Lemma A.1.** *Let  $w : \{1, \dots, n\} \rightarrow \mathbb{N}$  be an item weight function. Then the set  $w(2^{\{1, \dots, n\}})$  can be computed in time  $\mathcal{O}(n \cdot |w(2^{\{1, \dots, n\}})|)$ .*

*Proof.* Algorithm is a simple dynamic programming procedure. For all  $i \in \{0, \dots, n\}$  define the set  $\text{DP}_i$  as:

$$\text{DP}_i = \{w(X) : X \subseteq \{0, \dots, i\}\}.$$

Notice that  $\text{DP}_n = w(2^{\{1, \dots, n\}})$ . We iterate over  $i \in \{0, \dots, n\}$  to compute these sets. In the base case we set  $\text{DP}_0 = \{0\}$ . Then, for  $i \in \{1, \dots, n\}$  given  $\text{DP}_{i-1}$  we compute  $\text{DP}_i$  as follows:

$$\text{DP}_i = \text{DP}_{i-1} \cup \{x + w(i) \mid x \in \text{DP}_{i-1}\}.$$

Note that in a single iteration each item in  $\text{DP}_{i-1}$  is touched at most twice. Hence, the total number of arithmetic operations can be upper bounded by  $\sum_{i=0}^n 2 \cdot |\text{DP}_i| = \mathcal{O}(n \cdot |\text{DP}_n|)$ .  $\square$

**Corollary A.2.** *Let  $\delta \in (0, 1)$  be a fixed parameter. If  $|w(2^{\{1, \dots, n\}})| \geq 2^{\delta n}$ , then the critical pruner  $\text{crit}(\delta)$  can be computed in time  $\mathcal{O}^*(2^{\delta n})$ .*

*Proof.* Recall the following definitions. Let  $l = 1 + \lceil \log(\max_i \{w(i)\}) \rceil$ . For  $s \in \{0, \dots, l\}$ , the  $s$ -pruned weight of item  $i$  is  $w_s(i) := \lfloor w(i)/2^{l-s} \rfloor$ . The critical pruner,  $\text{crit}$ , is  $\text{crit}(\delta) = \min\{s \in \mathbb{N} \mid |w_s(2^{\{1, \dots, n\}})| \geq 2^{\delta n}\}$ . Notice that we can assume  $l = \text{poly}(n)$  by [24].

The algorithm finds  $\text{crit}$  by computing  $|w_s(2^{\{1, \dots, n\}})|$  using Lemma A.1 for consecutive  $s = 1, 2, \dots$  until  $|w_s(2^{\{1, \dots, n\}})| \geq 2^{\delta n}$ . At the end it returns this last  $s$  as  $\text{crit}$ . Because  $w_0(2^{\{1, \dots, n\}}) = \{0\}$  and Lemma A.3 tells us that  $(\frac{1}{3n}) |w_s(2^{\{1, \dots, n\}})| \leq |w_{s-1}(2^{\{1, \dots, n\}})|$  for any  $s$ , we know that  $|w_{\text{crit}}(2^{\{1, \dots, n\}})| = \mathcal{O}(2^{\delta n})$ . Analogously, the algorithm takes  $\mathcal{O}(n^2 \cdot 2^{\delta n})$  time per iteration and the number of iterations is at most  $l$ , which gives the claimed running time.  $\square$

**Lemma A.3.** *Let  $w : \{1, \dots, n\} \rightarrow \mathbb{N}$  be an item weight function and let  $l = 1 + \lceil \log(\max_i \{w(i)\}) \rceil$ . Then for all  $s \in \{1, \dots, l\}$ :*

$$\left(\frac{1}{3n}\right) |w_s(2^{\{1, \dots, n\}})| \leq |w_{s-1}(2^{\{1, \dots, n\}})| \leq \left(\frac{3n}{2}\right) |w_s(2^{\{1, \dots, n\}})|.$$



*Proof.* Let  $l = 1 + \lceil \log(\max_i \{w(i)\}) \rceil$ . We are given  $w_s(A)$  for all  $A \in 2^{\{1, \dots, n\}}$ . Observe, that we can bound the value of  $w_{s-1}(A)$  by the following:

$$\begin{aligned}
& \sum_{i \in A} (w(i)/2^{l-s+1} - 1) && \leq w_{s-1}(A) \leq && \sum_{i \in A} w(i)/2^{l-s+1} \\
\Rightarrow & \frac{1}{2} \sum_{i \in A} (w(i)/2^{l-s} - 2) && \leq w_{s-1}(A) \leq && \frac{1}{2} \sum_{i \in A} w(i)/2^{l-s} \\
\Rightarrow & \frac{1}{2} \sum_{i \in A} (\lfloor w(i)/2^{l-s} \rfloor - 2) && \leq w_{s-1}(A) \leq && \frac{1}{2} \sum_{i \in A} (\lfloor w(i)/2^{l-s} \rfloor + 1) \\
\Rightarrow & \frac{1}{2} (w_s(A) - 2n) && \leq w_{s-1}(A) \leq && \frac{1}{2} (w_s(A) + n)
\end{aligned}$$

Hence, for each value in  $w_s(2^{\{1, \dots, n\}})$ , there are at most  $\frac{3n}{2}$  values in  $w_{s-1}(2^{\{1, \dots, n\}})$ , i.e.

$$|w_{s-1}(2^{\{1, \dots, n\}})| \leq \left(\frac{3n}{2}\right) |w_s(2^{\{1, \dots, n\}})|.$$

Analogously, for a given  $w_{s-1}(A)$  and any subset  $A \in 2^{\{1, \dots, n\}}$  we can bound the value of  $w_s(A)$  by the following:

$$\begin{aligned}
& \sum_{i \in A} (w(i)/2^{l-s} - 1) && \leq w_s(A) \leq && \sum_{i \in A} w(i)/2^{l-s} \\
\Rightarrow & 2 \sum_{i \in A} \left( w(i)/2^{l-s+1} - \frac{1}{2} \right) && \leq w_s(A) \leq && 2 \sum_{i \in A} w(i)/2^{l-s+1} \\
\Rightarrow & 2 \sum_{i \in A} \left( \lfloor w(i)/2^{l-s+1} \rfloor - \frac{1}{2} \right) && \leq w_s(A) \leq && 2 \sum_{i \in A} (\lfloor w(i)/2^{l-s+1} \rfloor + 1) \\
\Rightarrow & 2 \left( w_{s-1}(A) - \frac{n}{2} \right) && \leq w_s(A) \leq && 2(w_{s-1}(A) + n)
\end{aligned}$$

Hence, for each value in  $w_{s-1}(2^{\{1, \dots, n\}})$ , there are at most  $3n$  values in  $w_s(2^{\{1, \dots, n\}})$ , i.e.

$$|w_{s-1}(2^{\{1, \dots, n\}})| \geq \left(\frac{1}{3n}\right) |w_s(2^{\{1, \dots, n\}})|.$$

□

## B Inequalities with Binomials and Entropy

Let us start with the useful facts about binary entropy function.

$$h(x) := -x \log(x) - (1-x) \log(1-x).$$

The first derivative of binary entropy is:

$$h'(x) := \log(1-x) - \log(x)$$

The second derivative:

$$h''(x) := -\frac{1}{(\ln 2)x(1-x)}$$

and we will also need third derivative

$$h'''(x) := \frac{1 - 2x}{(\ln 2)x^2(1-x)^2}$$

Observe, that for  $x \in [0, \frac{1}{2}]$  we have that  $h'(x) \geq 0$ ,  $h''(x) \leq 0$  and  $h'''(x) \geq 0$ . From 4th derivative we will only need that  $h^{(4)}(x) \leq 0$  when  $x \in [0, \frac{1}{2}]$ . Hence from Taylor expansion for  $x \in [0, \frac{1}{2}]$  it holds that:

$$h(x + \varepsilon) := h(x) + h'(x)\varepsilon + \frac{h''(x)}{2}\varepsilon^2 + \frac{h'''(x)}{6}\varepsilon^3 + \mathcal{O}(\varepsilon^4).$$

If we assume, that  $x \in [0, \frac{1}{2}]$  and  $\varepsilon \leq \frac{3h''(x)}{2h'''(x)}$  then:

$$h(x + \varepsilon) \leq h(x) + h'(x)\varepsilon + \frac{h''(x)}{4}\varepsilon^2. \quad (13)$$

because  $h^{(4)}(x) \leq 0$  when  $x \in [0, \frac{1}{2}]$ .

**Lemma B.1** (Theorem 2.2 from [11]).

$$\begin{aligned} \forall x \in [0, 1] : \quad & 1 - 4 \left(x - \frac{1}{2}\right)^2 \leq h(x) \leq 1 - \frac{2}{\ln(2)} \left(x - \frac{1}{2}\right)^2, \\ \forall x \in [0, 1] : \quad & \frac{x}{2 \log(\frac{6}{x})} \leq h^{-1}(x) \leq \frac{x}{\log \frac{1}{x}}, \end{aligned}$$

where the inverse entropy function  $h^{-1} : [0, 1] \rightarrow [0, 1]$  is the inverse of  $h$  restricted to the interval  $[0, \frac{1}{2}]$ .

**Lemma B.2.** Given  $\mathcal{W} \subseteq \{W \subseteq \{1, \dots, n\} : |W| \in [\frac{n}{2} \pm c \cdot n]\}$  for some  $c \in [0, \frac{1}{2}]$  with  $|\mathcal{W}| \leq 2^{(1-z)n} \text{poly}(n)$  and  $z \in (0, 1)$  such that  $\frac{z}{4 \log(12/z)} > c$ . Then

$$|\downarrow \mathcal{W}| + |\uparrow \mathcal{W}| \leq \mathcal{O}\left(2^{(1-\rho(z,c))n}\right)$$

where  $\rho(z, c) = \frac{2}{\ln(2)} \left(\frac{z}{4 \log(12/z)} - c\right)^2$ .

*Proof.* We will bound  $|\downarrow \mathcal{W}|$ . Take  $\lambda = \frac{z}{4 \log(12/z)} - c$  and note that by assumption  $\lambda > 0$ . We can describe any  $X \in \downarrow \mathcal{W}$  either as a set in  $\binom{\{1, \dots, n\}}{|X|}$  (if  $|X| \leq (\frac{1}{2} - \lambda)n$ ) or as an element  $W \in \mathcal{W}$  together with the items on which  $W$  and  $X$  differ (if  $|X| \geq (\frac{1}{2} - \lambda)n$ ). In the latter case,  $|W \setminus X| \leq (\lambda + c)n$ . This, together with the fact that  $|X|$  can only take  $n$  distinct values implies:

$$\begin{aligned} |\downarrow \mathcal{W}| &\leq n \cdot \binom{n}{(\frac{1}{2} - \lambda)n} + |\mathcal{W}| \cdot \binom{n}{(\lambda + c)n} \\ &\leq n \cdot 2^{h(\frac{1}{2} - \lambda)n} + \text{poly}(n) \cdot 2^{(1-z+h(\lambda+c))n} \end{aligned}$$

Which means that:

$$\frac{\log |\downarrow \mathcal{W}|}{n} \leq \max \left\{ h \left( \frac{1}{2} - \lambda \right), 1 - z + h(\lambda + c) \right\} + o(1)$$

First, we will show that  $h(\frac{1}{2} - \lambda) \geq 1 - z + h(\lambda + c)$ . Note that

$$\lambda + c = \frac{z}{4 \log(12/z)} = \frac{z/2}{2 \log(6/(z/2))} \leq h^{-1}(z/2)$$

by Lemma B.1. Hence,  $h(\lambda + c) \leq z/2$  as  $h$  is monotonic on  $[0, \frac{1}{2}]$  and  $\lambda + c \leq \frac{1}{2}$ . Therefore:

$$\begin{aligned}
h\left(\frac{1}{2} - \lambda\right) &\geq 1 - 4\lambda^2 \\
&\geq 1 - 4\left(\frac{z}{4\log(12/z)} - c\right)^2 \\
&\geq 1 - 4\left(\frac{z^2}{16\log^2(12/z)}\right) \\
&\geq 1 - 4\left(\frac{z}{16\log^2(12)}\right) \\
&\geq 1 - z/2 \\
&\geq 1 - z + h(\lambda + c).
\end{aligned}$$

Where the first inequality follows from Lemma B.1, the second is because  $z/\log^2(12/z)$  is an increasing function for  $z < 1$  and the last because  $h(\lambda + c) \leq z/2$ . Finally, we know, using Lemma B.1 that  $h\left(\frac{1}{2} - \lambda\right) < 1 - \frac{2}{\ln(2)}\lambda^2$ . Here the inequality is strict since  $\lambda > 0$ . Therefore,

$$|\downarrow\mathcal{W}| \leq \mathcal{O}\left(2^{(1-\rho(z,c))n}\right),$$

where  $\rho(z, c) = \frac{2}{\ln(2)}\left(\frac{z}{4\log(12/z)} - c\right)^2$  and we use that the inequality on  $h\left(\frac{1}{2} - \lambda\right)$  is strict to omit  $\text{poly}(n)$  factors. The same argument establishes  $|\uparrow\mathcal{W}| \leq \mathcal{O}\left(2^{(1-\rho(z,c))n}\right)$   $\square$

**Lemma B.3.** *For every  $\beta, \alpha, \gamma \in [0, \frac{1}{2}]$  it holds that :*

$$\binom{\beta n}{\alpha\beta n - \gamma n} \binom{(1-\beta)n}{\alpha(1-\beta)n + \gamma n} \leq \binom{n}{\alpha n} \cdot 2^{-f(\gamma, \alpha, \beta)n}.$$

where

$$f(\gamma, \alpha, \beta) := \begin{cases} \gamma^2 & \text{if } |\gamma| < \alpha(1-\alpha) \min\{\beta, (1-\beta)\} \\ -\alpha^2 \log 2\alpha & \text{otherwise} \end{cases}$$

*Proof.* First, observe that when  $|\gamma| > \alpha(1-\alpha) \min\{\beta, (1-\beta)\}$  our expression is upper bounded by:

$$\binom{\beta n}{\alpha\beta n(1-\alpha)} \binom{(1-\beta)n}{\alpha(1-\alpha)(1-\beta)n} \leq \binom{n}{\alpha(1-\alpha)n}.$$

This however is bounded by  $2^{h(\alpha(1-\alpha))n}$ . Observe that  $h(\alpha - \alpha^2) \leq h(\alpha) - h'(\alpha)\alpha^2 = h(\alpha) - \alpha^2(\log(\alpha) - \log(1-\alpha)) \leq h(\alpha) - \alpha^2 \log(2\alpha)$ . Hence when  $|\gamma|$  is large we upper bound our expression with:

$$\binom{n}{\alpha n} 2^{\alpha^2 \log(2\alpha)n}.$$

Now, we consider the case of small  $|\gamma|$ . We upper bound the expression with binary entropy.

$$\binom{\beta n}{\alpha\beta n - \gamma n} \binom{(1-\beta)n}{\alpha(1-\beta)n + \gamma n} = 2^{n(\beta h(\alpha - \frac{\gamma}{\beta}) + (1-\beta)h(\alpha + \frac{\gamma}{1-\beta}))}$$

Let us consider an exponent:

$$\beta h\left(\alpha - \frac{\gamma}{\beta}\right) + (1 - \beta)h\left(\alpha + \frac{\gamma}{1 - \beta}\right)$$

We use Inequality 13 with  $x = \alpha$  and  $\varepsilon := -\frac{\gamma}{\beta}$  for  $h(\alpha - \gamma/\beta)$  and with  $\varepsilon := \frac{\gamma}{1 - \beta}$  for  $h(\alpha + \gamma/(1 - \beta))$ .

Observe that at the beginning we assumed that  $|\gamma| \leq \alpha(1 - \alpha) \min\{\beta, (1 - \beta)\}$  hence  $|\frac{\gamma}{\beta}|$  and  $|\frac{\gamma}{1 - \beta}|$  are upper bounded by  $|\frac{3h''(\alpha)}{2h'''(\alpha)}|$ . So, by Inequality 13:

$$\beta h\left(\alpha - \frac{\gamma}{\beta}\right) + (1 - \beta)h\left(\alpha + \frac{\gamma}{1 - \beta}\right) \leq h(\alpha) + \frac{h''(\alpha)}{4} \frac{\gamma^2}{\beta(1 - \beta)}.$$

Observe that the first order factors cancel out. Hence

$$\binom{\beta n}{\alpha\beta n - \gamma n} \binom{(1 - \beta)n}{\alpha(1 - \beta)n + \gamma n} \leq \binom{n}{\alpha n} 2^{\frac{h''(\alpha)\gamma^2}{4\beta(1 - \beta)}}$$

Finally, observe that  $h''(\alpha) < -1$  for all  $\alpha \in [0, \frac{1}{2}]$  and  $\frac{1}{\beta(1 - \beta)} \geq 4$  for all  $\beta \in [0, \frac{1}{2}]$  hence:

$$\binom{\beta n}{\alpha\beta n - \gamma n} \binom{(1 - \beta)n}{\alpha(1 - \beta)n + \gamma n} \leq \binom{n}{\alpha n} 2^{-\gamma^2 n}$$

□

**Lemma B.4.** For all  $k \in \mathbb{N}$  and  $\alpha \in [0, 1]$  we have:

$$h(\text{Bin}(k, \alpha)) \leq h(\text{Bin}(k + 1)).$$

*Proof.* Let us fix  $k \in \mathbb{N}$ . Recall that  $\text{Bin}(k + 1) = (\{0, \dots, k + 1\}, p(i))$  and  $\text{Bin}(k, \alpha) = (\{0, \dots, k + 1\}, p_\alpha(i))$  where  $p(i) = \binom{k+1}{i} \frac{1}{2^{k+1}}$  and  $p_\alpha(i) = \binom{k}{i} \frac{(1 - \alpha)}{2^k} + \binom{k}{i-1} \frac{\alpha}{2^k}$ . Hence, we need to prove that for all  $\alpha \in [0, 1]$ :

$$h(p_\alpha(0), \dots, p_\alpha(k + 1)) \leq h(p(0), \dots, p(k + 1)).$$

Let us denote  $\phi(\alpha) = h(p_\alpha(0), \dots, p_\alpha(k + 1))$ . First, observe that  $\phi(\frac{1}{2}) = h(p(0), \dots, p(k + 1))$  because  $\binom{k+1}{i} = \binom{k}{i} + \binom{k}{i-1}$ . Therefore we need to prove that for all  $\alpha \in [0, 1]$  it holds that:

$$\phi(\alpha) \leq \phi\left(\frac{1}{2}\right).$$

Recall that the binary entropy of a multinomial is  $h(a_0, \dots, a_{k+1}) = -a_0 \log(a_0) - \dots - a_{k+1} \log(a_{k+1})$  and  $(x \ln(x))' = \ln(x) + 1$ . Observe that function  $\phi(\alpha)$  is well defined for  $\alpha = 0$  and  $\alpha = 1$  as limits. Moreover  $\phi(\alpha) \geq 0$  for all  $\alpha \in [0, 1]$ .

Now, we compute the first derivative.

$$\phi'(\alpha) = -\frac{1}{2^k \ln 2} \sum_i \left( \binom{k}{i-1} - \binom{k}{i} \right) (1 + \ln(p_\alpha(i))).$$

Because  $\sum_i \binom{k}{i-1} = \sum_i \binom{k}{i}$  the first derivative simplifies to:

$$\phi'(\alpha) = -\frac{1}{2^k \ln 2} \sum_i \left( \binom{k}{i-1} - \binom{k}{i} \right) \ln(p_\alpha(i))$$

Now the second derivative is

$$\phi''(\alpha) = -\frac{1}{4^k \ln 2} \sum_i \left( \binom{k}{i-1} - \binom{k}{i} \right)^2 \cdot \frac{1}{p_\alpha(i)} \leq 0,$$

thus  $\phi(\alpha)$  is concave for all  $\alpha \in [0, 1]$ . So in order to show that the  $\phi(\alpha)$  function has exactly one maximum in  $\alpha = \frac{1}{2}$  it is sufficient to show that  $\phi'(\frac{1}{2}) = 0$ .

Let us rearrange the sum:

$$\begin{aligned} \phi' \left( \frac{1}{2} \right) &= \sum_i \left[ \binom{k}{i-1} - \binom{k}{i} \right] \ln(p_{\frac{1}{2}}(i)) = \sum_i \binom{k}{i} \ln(p_{\frac{1}{2}}(i+1)) - \sum_i \binom{k}{i} \ln(p_{\frac{1}{2}}(i)) \\ &= \sum_i \binom{k}{i} \ln \frac{p_{\frac{1}{2}}(i+1)}{p_{\frac{1}{2}}(i)}. \end{aligned}$$

Because

$$p_{\frac{1}{2}}(i) = \frac{1}{2^{k+1}} \binom{k+1}{i}$$

we can simplify the fraction:

$$\frac{p_{\frac{1}{2}}(i+1)}{p_{\frac{1}{2}}(i)} = \frac{\binom{k+1}{i+1}}{\binom{k+1}{i}} = \frac{k+1-i}{i+1},$$

thus

$$\begin{aligned} \phi' \left( \frac{1}{2} \right) &= \sum_i \binom{k}{i} \ln \left( \frac{k+1-i}{i+1} \right) = \sum_i \binom{k}{i} \ln(k+1-i) - \sum_i \binom{k}{i} \ln(i+1) \\ &= \sum_i \binom{k}{i} \ln(k+1-i) - \sum_i \binom{k}{k-i} \ln(k-i+1) = 0, \end{aligned}$$

which finishes the proof. □