# Hitting Sets when the Shallow Cell Complexity is Small * 

Sander Aarts ${ }^{1[0000-0003-1852-9116]}$<br>David B. Shmoys ${ }^{1[0000-0003-3882-901 X]}$

September 26, 2023


#### Abstract

The hitting set problem is a well-known NP-hard optimization problem in which, given a set of elements and a collection of subsets, the goal is to find the smallest selection of elements, such that each subset contains at least one element in the selection. Many geometric set systems enjoy improved approximation ratios, which have recently been shown to be tight with respect to the shallow cell complexity of the set system. The algorithms that exploit the cell complexity, however, tend to be involved and computationally intensive. This paper shows that a slightly improved asymptotic approximation ratio for the hitting set problem can be attained using a much simpler algorithm: solve the linear programming relaxation, take one initial random sample from the set of elements with probabilities proportional to the LP-solution, and, while there is an unhit set, take an additional sample from it proportional to the LP-solution. Our algorithm is a simple generalization of the elegant net-finder algorithm by Nabil Mustafa. To analyze this algorithm for the hitting set problem, we generalize the classic Packing Lemma, and the more recent Shallow Packing Lemma, to the setting of weighted epsilon-nets.


Keywords: Hitting set • Set cover • Approximation algorithms • Computational geometry . Shallow cell complexity • Wireless coverage

## 1 Introduction

The input to the hitting set problem is a finite set system - a ground set $X$ of $m$ elements, or points, and a collection $\mathcal{R}$ of $n$ subsets, or ranges, of $X$. This can also be understood as a hypergraph, with vertices $X$ and hyper-edges $\mathcal{R}$. A hitting set is a subset of elements $H \subseteq X$ such that every set $R \in \mathcal{R}$ is hit by $H$,
*This material is based on work supported by the NSF under Grant CNS-1952063.
i.e. $R \cap H \neq \emptyset$, for all $R \in \mathcal{R}$. This is a vertex cover under the hypergraph view. The set system can be encoded as a set-element incidence matrix $A \in\{0,1\}^{n \times m}$, in which the $(i, j)$ th entry $a_{i j}$ is 1 if range $R_{i}$ contains point $x_{j}$, and 0 otherwise. The IP of the minimum hitting set problem is

$$
\begin{array}{ll}
\min _{y} & \sum_{j: x_{j} \in X} y_{j} \\
\text { s.t. } & \sum_{j: x_{j} \in X} a_{i j} y_{j} \geq 1,  \tag{1}\\
\quad y_{j} \in\{0,1\}, & \forall i: R_{i} \in \mathcal{R} ; \\
& \forall j: x_{j} \in X,
\end{array}
$$

where variable $y_{j} \in\{0,1\}$ indicates whether element $x_{j}$ is in the solution $H$.
Hitting sets and set covers are intimately connected; a hitting set for $A$ is a set cover of $A^{T}$. Both problems' decision versions are NP-complete [10]. There exists an $\mathcal{O}(\log m)$-approximation algorithm, and this bound is tight unless P $=$ NP [9,13]. However, there are algorithms that exploit additional structure in $A$ to attain improved approximation ratios ${ }^{1}$. Indeed, our work is motivated by the problem of exploiting structure when covering large numbers of wireless LoRaWAN transmitters with wireless receivers. Transmitters can be viewed as points, which are considered to be covered if they are in the line of sight of a wireless receiver, which in turn drives transmission quality in LoRaWAN [22]. The area in the line of sight of a receiver roughly resembles a simple shape.

Many geometric set systems enjoy better approximation ratios via epsilonnets, or $\epsilon$-nets. A set system is said to be geometric whenever its elements can be encoded as points in Euclidean space, and sets are derived from containment of the points in geometric shapes, such as half-spaces, balls or rectangles ${ }^{2}$. The seminal work of Brönnimann and Goodrich [3], and Even et al. [8], connects the approximability of a hitting set instance to the size of weighted $\epsilon$-nets. Given non-negative weights on the points, $\mu: X \rightarrow \mathbb{R}_{\geq 0}$, a weighted $\epsilon$-net with respect to weights $\mu$ is a subset $H \subseteq X$ that hits all $\epsilon$-heavy sets:

$$
\begin{equation*}
\forall R \in \mathcal{R} \text { with } \mu(R) \geq \epsilon \cdot \mu(X): \quad R \cap H \neq \emptyset \tag{2}
\end{equation*}
$$

where the weight of any subset $S \subseteq X$ is defined as $\mu(S)=\sum_{x \in S} \mu(x)$. Even et al. [8] reduce the problem of finding a small hitting set to finding a small $\epsilon$-net via a reformulation of the linear programming relaxation of the hitting set problem (1). The reformulated LP (3) is a program for finding the largest $\epsilon$, and corresponding weights $\mu$, subject to the constraint that an $\epsilon$-net with respect to

[^0]weights $\mu$ is a hitting set.
\[

$$
\begin{align*}
\max _{\epsilon, \mu} \epsilon & \\
\text { s.t. } \sum_{j: x_{j} \in X} a_{i j} \mu_{j} \geq \epsilon, & \forall i: R_{i} \in \mathcal{R} \\
\sum_{j: x_{j} \in X} \mu_{j} & =1 ;  \tag{3}\\
\mu_{j} \geq 0, & \forall j: x_{j} \in X
\end{align*}
$$
\]

The first constraint requires that each set $R$ is $\epsilon$-heavy; the second constraint normalizes the weights. Let $\left(\epsilon^{*}, \mu^{*}\right)$ denote an optimal solution to LP (3), with $\mu^{*}=\left(\mu_{1}^{*}, \ldots, \mu_{n}^{*}\right)$. Let $z^{*}$ be the optimal value to the LP relaxation of the original program (1). The first constraint ensures that an $\epsilon^{*}$-net with respect to weights $\mu^{*}$ is a hitting set. Moreover, the reciprocal optimal value $1 / \epsilon^{*}$ is equal to the optimal LP value $z^{*}[8]$. In particular, an $\epsilon^{*}$-net of size $g\left(1 / \epsilon^{*}\right)$ for some function $g(\cdot)$ is a hitting set of size of $g\left(z^{*}\right)$. Hence, to find a small hitting set it suffices to solve LP (3) and find a small $\epsilon^{*}$-net with respect to weights $\mu^{*}$.

Haussler and Welzl [12] show that set systems with bounded VC-dimension admit small $\epsilon$-nets, and develop a simple algorithm to find them. The VCdimension is a measure of the set system's complexity. Given a subset $S \subseteq X$, the projection of $\mathcal{R}$ to $S$ is the set system formed by elements $S$ and sets $\left.\mathcal{R}\right|_{S}=\{R \cap S: R \in \mathcal{R}\}$. The VC-dimension of $\mathcal{R}$ is the size of the largest subset $S \subseteq X$ such that $\left.\mathcal{R}\right|_{S}$ shatters $S$, i.e. the largest set $S$ such that $\left.\mathcal{R}\right|_{S}$ contains all subsets of $S$. In particular, Clarkson [7], and Haussler and Welzl [12], show that any set system with VC-dimension $d$ has a weighted $\epsilon$-net of size $\mathcal{O}\left(\frac{d}{\epsilon} \log \frac{1}{\epsilon}\right)$. This is remarkable, as the size is independent of both the size of $X$ and $\mathcal{R}$. Moreover, the algorithm for finding such an $\epsilon$-net is simple: Select a subset $H \subseteq X$ by sampling each element $x$ in $X$ independently.

Theorem 1 ( $\epsilon$-net Theorem [12,14]). Let $(X, \mathcal{R})$ be a set system with VCdimension $d$, and let $\mu: X \rightarrow \mathbb{R}_{\geq 0}$ be element weights with $\mu(X)=1$. Then for any $\epsilon, \gamma \in(0,1)$ :

$$
H \leftarrow \text { pick each } x \in X \text { with probability } \min \left\{1, \frac{2 \mu(x)}{\epsilon} \cdot \max \left\{\log \frac{1}{\gamma}, d \log \frac{1}{\epsilon}\right\}\right\}
$$

is a weighted $\epsilon$-net with respect to weights $\mu$ with probability at least $1-\gamma$.
Throughout, we define $\mu(S)=\sum_{x \in S} \mu(x)$ for all subsets $S \subseteq X$. For general set systems of VC-dimension $d$, this bound is tight in expectation [14]. However, there are alternative ways to parameterize the complexity of set systems.

### 1.1 Shallow Cell Complexity

The shallow cell complexity (SCC) is a finer parameterization of the complexity of set systems. [1,4,21]. Readers are referred to Mustafa and Varadarajan [20]
for more background. A cell in a binary matrix $A$ is a collection of identical rows. A cell has depth $k$ if the number of 1's in any of its rows is exactly $k$, i.e., if each set in the cell contains $k$ elements. For a non-decreasing function $\varphi(\cdot, \cdot)$ we say binary matrix $A$ has shallow cell complexity (SCC) $\varphi(\cdot, \cdot)$ if, for all $1 \leq k \leq l \leq m$, the number of cells of depth at most $k$ in any submatrix $A^{*}$ of $A$ of at most $l$ columns, is at most $\varphi(l, k)$. A set $\operatorname{system}(X, \mathcal{R})$ is said to have SCC $\varphi(l, k)$ if its set-element incidence matrix $A$ does. Often $\varphi(l, k)=\mathcal{O}\left(\varphi(l) k^{c}\right)$ for some constant $c>0$ and single-variable function $\varphi(\cdot)$, in which case the dependence on $k$ is can be dropped and the SCC denoted by $\varphi(l)$. Examples of geometric set systems with small shallow cell complexity are discs in the plane with $\varphi(l, k)=\mathcal{O}(k)$, and axis-parallel rectangles with $\varphi(l, k)=\mathcal{O}\left(l k^{2}\right)$ [18].

As is true for VC-dimension, there are algorithms that find hitting sets or $\epsilon$-nets with sizes bounded in terms of the shallow cell complexity. A prominent example is the quasi-uniform sampling algorithm of Chan et al. [4]. Given nonnegative weights $\mu: X \rightarrow \mathbb{R}_{\geq 0}$, and a value $\epsilon>0$, the algorithm finds a hitting set while maintaining an upper bound on the probability of selecting any given element.

Theorem 2 (Quasi-uniform sampling [4]). Suppose a set system defined by A has $S C C \varphi(l, k)=\varphi(l) k^{c}$ for some $c>0$. Then there is a randomized polytime algorithm that returns a hitting set of expected size $\mathcal{O}(\max \{1, \log (\varphi(m))\})$ times the LP optimum.

The algorithm attains the optimal approximation ratio with respect to the $\mathrm{SCC}^{3}$. However, the sampling procedure is involved, and may require enumeration over all sets $\mathcal{R}$, of which there can be $n=\Omega\left(m^{c}\right)$ for some constant $c>0$ [17].

Taking a different approach, Mustafa and colleagues [16,17,19] develop a net-finder for asymptotically optimal-sized unweighted $\epsilon$-nets with respect to the SCC. The algorithm is remarkably simple: Take an initial sample from $X$, and while there are unhit sets, choose an unhit set arbitrarily, and add $\mathcal{O}(1)$ randomly chosen elements from this set to the original sample. The algorithm assumes access to an oracle that returns an unhit set. This oracle is called at most $\mathcal{O}(1 / \epsilon)$ times in expectation. While the size of the returned $\epsilon$-net is asymptotically on par with the quasi-uniform sampling algorithm, there are large constants in the upper bound [17].

This algorithm is not directly applicable to the hitting set problem via the LP-reduction above, although it can be used via a standard reduction. The analysis of the algorithm applies to only uniform weights, and the optimal weights $\mu^{*}$ of the LP-formulation (3) are not generally uniform. Nevertheless, it is possible to reduce the problem of finding a weighted $\epsilon$-net to that of finding a uniform $\epsilon^{\prime}$ net following a standard reduction, in which an expanded instance is generated by copying each element $x_{j} \in X$ a number of times roughly proportional to its weight $\mu^{*}\left(x_{j}\right)[3,4]$. This can generate $\Omega(m)$ copies of each element, which can

[^1]have notable consequences. First, to achieve a weighted $\epsilon^{*}$-net in the original instance, one must use a smaller value $\epsilon^{\prime}$ for the expanded instance, on the order of $\mathcal{O}\left(\epsilon^{*} / m\right)$. This results in an approximation ratio of $\mathcal{O}(\log \varphi(\mathcal{O}(m)))$. Secondly, generating copies can increase the number of elements from $m$ to $\Omega\left(m^{2}\right)$. This can increase the runtime considerably. In particular, repeatedly sampling from sets of size $\Theta\left(m^{2}\right)$ can become prohibitive on large instances such as the wireless coverage problem motivating our work.

### 1.2 Our Contributions

This paper generalizes the elegant net-finder algorithm of Mustafa [17] to the setting of weighted $\epsilon$-nets, in order to produce a fast and simple algorithm for the hitting set problem, which attains asymptotically optimal approximation ratios with respect to the shallow cell complexity. The algorithm enjoys a faster runtime that makes solving larger instances, such as LoRaWAN receiver placement at scale, feasible. This is achieved by combining the weighted $\epsilon$-net finder with the reduction of Even et al. [8]. In doing so, we also improve on the asymptotic approximation ratio from $\max \{1, \log \varphi(m)\}$ to $\max \left\{1, \mathcal{O}\left(\log \varphi\left(\mathcal{O}\left(z^{*}\right)\right)\right)\right\}$ where $z *$ is the optimal value to the linear relaxation of the hitting set program (1). While in the worst case $z^{*}=m$, it is often the case that $z^{*} \ll m$. However, the multiplicative constants in our analysis are relatively large, matching those of Mustafa [17]. In addition to the algorithm, our analysis generalizes the classic Packing Lemma of Haussler [11], as well as the Shallow Packing Lemma of Mustafa et al.. [19], to the weighted setting, which may be of independent interest.

Key to our approach are adaptations of Mustafa's [19] Shallow Packing Lemma and Haussler's [11] classic Packing Lemma that accommodate nonuniform weights. Our main technical contribution is to allow a notion of weighted packings. Consider any non-negative weights $\mu: X \rightarrow \mathbb{R}_{\geq 0}$ with $\sum_{x \in X} \mu(X)=1$, and extend it to element subsets via $\mu(S)=\sum_{x \in S} \mu(S) .{ }^{4} \mathrm{~A}(k, \delta)$-packing with respect to weights $\mu$ is a collection of sets $\mathcal{P} \subseteq \mathcal{R}$ in which (i) all sets $R$ in $\mathcal{P}$ are at most $k$-heavy, i.e., have bounded weight $\mu(R) \leq k$; and (ii) all pairs of sets have symmetric differences of weight at least $\delta$. (See Definition 1). Our weighted shallow packing lemma upper bounds the number of sets in $\mathcal{P}$ as a function of the SCC. Our approach accommodates weights $\mu$ by sampling elements from a distribution with probability mass proportional to the weights, rather than from a uniform distribution as in the original proofs. Moreover, our proof uses sampling with replacement rather than without replacement to simplify the analysis. While more generally applicable, our result yields the same bound on the size of $\mathcal{P}$ as in the unweighted setting. An analogous sampling approach is used in proving Theorem 1 [14]. Equipped with our generalized lemma, it is straightforward to adapt Mustafa's [17] analysis to a weighted net-finder. A proof of our Weighted Packing Lemma is included in the extended online version.

[^2]
## 2 Algorithm and Main Result

Our algorithm combines the LP-relaxation of Even et al. [8] with the generalized sampling approach of Mustafa [17]. Our procedure is summarized in Algorithm 1. The algorithm makes use of two global constants, $\beta$ and $\gamma$. These are assumed to be positive, and to satisfy $\gamma \leq 1 / 4$ and $\beta+\gamma \leq 1$.

```
Algorithm 1: A simple hitting set algorithm with details
    Data: A matrix \(A\) with VC-dim \((A) \leq d\) and SCC \(\varphi(\cdot, \cdot)\), constants \(\gamma, \beta>0\)
    \(\epsilon^{*},\left(\mu_{1}^{*}, \ldots, \mu_{n}^{*}\right) \leftarrow\) solve LP \(\left\{\max \epsilon: A \mu \geq \epsilon, \mu^{T} \mathbf{1}=1, \mu \geq \mathbf{0}\right\}\);
    \(H \leftarrow \emptyset ;\)
    for \(x_{j} \in X\) do
        \(H \leftarrow H \cup\left\{x_{j}\right\}\) with probability
        \(\min \left\{1, \frac{2 \mu_{j}^{*}}{\left(\frac{3}{4}-\frac{\beta}{2}\right) \epsilon^{*}} \cdot \max \left\{\log \left(d^{2} \varphi\left(\frac{8 d}{\beta \epsilon^{*}}, \frac{48 d}{\beta}\right)^{2}\right), d \log \left(\frac{1}{\left(\frac{3}{4}-\frac{\beta}{2}\right) \epsilon^{*}}\right)\right\}\right\}\)
    end
    while there is a set \(R \in \mathcal{R}\) not hit by \(H\) do
        Independently add each \(x_{j} \in R\) to \(H\) with probability
        \(\min \left\{1, \frac{2 \mu_{j}^{*}}{\gamma \mu^{*}(R)} \max \left\{\log 2, d \log \frac{1}{\gamma}\right\}\right\}\)
    end
    return \(H\)
```

In the while loop, the weights $\mu^{*}(R)=\sum_{j: x_{j} \in R} \mu_{j}^{*}$ denote the weight of set $R$ under the LP optimal weights $\mu^{*}=\left(\mu_{1}^{*}, \ldots, \mu_{n}^{*}\right)$.

Conceptually, the algorithm is simple; it randomly selects an initial set of elements $H$ from $X$, and proceeds to add additional random subsets of elements to $H$ until this is a hitting set. The algorithm relies on an oracle that returns an arbitrary unhit set. This oracle is treated as a black box. Our main result is twofold: we bound the expected size of the solution hitting set $H$ as a function of the cell complexity, and bound the expected number of oracle calls.

Theorem 3. Let A be a binary matrix encoding a hitting set instance with shallow cell complexity $\varphi(\cdot, \cdot)$ and $\mathrm{VC}-\operatorname{dim}(\mathcal{R}) \leq d$. Let $z^{*}$ be LP optimal value. Then the algorithm returns a hitting set of expected size

$$
\mathcal{O}\left(z^{*} \cdot \max \left\{1, \log \varphi\left(\mathcal{O}\left(z^{*}\right), \mathcal{O}(d)\right)\right\}\right) .
$$

Furthermore it makes $\mathcal{O}\left(z^{*}\right)$ oracle calls in expectation.
Note that the algorithm always returns a hitting set; the randomness is in the size of the solution and the runtime. This is in contrast with the net-finder in Theorem 1. Both algorithms require knowing the VC dimension $d$; ours must additionally know the shallow cell complexity $\varphi(\cdot, \cdot)$. If unknown, these can be searched for using a standard doubling trick [17].

## 3 The Weighted Shallow Packing Lemma

The Weighted Shallow Packing Lemma is key to proving Theorem 3. This section formally defines weighted shallow packings, states the lemma, and proves it. To this end, fix non-negative weights $\mu$ over $X$, and define the weight of a subset of elements $S \subseteq X$ as $\mu(S)=\sum_{j \in S} \mu_{j}$. Assume that $\mu(X)=1$. To contrast, let $\operatorname{card}(S)$ denote the cardinality of any set $S$. Note that the weights $\mu$ induce a probability distribution over the elements $X$. Throughout, whenever an element $u$ of $X$ is randomly sampled, it is assumed to follow a distribution proportional to $\mu(\cdot)$, in which case we say $u$ is sampled from $\mu(\cdot)$, and denote this by $u \sim \mu(\cdot)$. Note that an element $u \sim \mu(\cdot)$ sampled this way lies in subset $S \subseteq X$ with probability $\mu(S)$.

The main purpose of the weighted shallow packing lemma is to bound the number of sets in a set system in terms of its shallow cell complexity. Clearly, an arbitrary set systems can contain large numbers of sets. Instead, we focus on a particular kind of set system called a weighted packing. A set system is a packing if all its sets are "light", and each pair of sets are sufficiently different from each other. Critically, we define "light" and "different" in reference to the weights.
Definition 1. Let $(X, \mathcal{P})$ be a set system with weights $\mu$, and let $k, \delta \in(0,1)$ be constants. If all sets $S$ in $\mathcal{P}$ satisfy $\mu(S) \leq k$, and all pairs of distinct sets $S, R$ in $\mathcal{P}$ have symmetric difference of weight at least $\delta$, i.e.

$$
\begin{equation*}
\mu(\Delta(S, R))=\mu((S \backslash R) \cup(R \backslash S)) \geq \delta \tag{4}
\end{equation*}
$$

then we say $(X, \mathcal{P})$ is a weighted $(k, \delta)$-packing with respect to $\mu$.
We omit the "with respect to $\mu$ "-statement whenever this is clear from context.
The shallow packing lemma bounds the number of sets in a packing as a function of the constants $(k, \delta)$, the VC-dimension, and the shallow cell complexity.

Lemma 1 (Weighted shallow packing lemma). Let $(X, \mathcal{P})$ be a set system on $m$ elements, equipped with weights $\mu$, and let $(X, \mathcal{P})$ be a $(k, \delta)$-packing with respect to $\mu$ for constants $k, \delta>0$. Assume the set system has $\mathrm{VC}-\operatorname{dim}(\mathcal{P}) \leq d$, and shallow cell complexity $\varphi(\cdot, \cdot)$. Then

$$
\operatorname{card}(\mathcal{P}) \leq \frac{24 d}{\delta} \cdot \varphi\left(\frac{8 d}{\delta}, \frac{48 d k}{\delta}\right)
$$

The proof to this lemma makes use of our weighted Packing Lemma. The unweighted Packing Lemma is a classic result by Haussler [11] that bounds the number of sets in a packing. We generalize this to nonuniform weights.

Lemma 2 (Weighted packing lemma). Let $(X, \mathcal{P})$ be a set system with $n$ sets and $m$ elements, equipped with weights $\mu$. Let $\operatorname{VC-dim}(\mathcal{P}) \leq d$ for some integer $d \geq 1$ and assume there is a constant $\delta \in(0,1)$ such that $\mu\left(\Delta\left(S_{i}, S_{k}\right)\right) \geq \delta$ for all $1 \leq i<k \leq n$. Then

$$
\operatorname{card}(\mathcal{P}) \leq 2 \mathbb{E}\left[\operatorname{card}\left(\left.\mathcal{P}\right|_{Y}\right)\right]
$$

where $Y$ is the set of unique elements in a random sample $U=\left(u_{1}, u_{2}, \ldots, u_{s}\right)$ of size $s=\left\lceil\frac{8 d}{\delta}\right\rceil-1$, in which each element $u_{k}$ is sampled iid $u_{k} \sim \mu(\cdot)$ with replacement.
The proof of the latter lemma is in the appendix to the extended online version of the paper; Lemma 1 is proved next.

### 3.1 Proof of the Weighted Shallow Packing Lemma

Proof. Fix a $(k, \delta)$-packing $\mathcal{P}$ and let $U=\left(u_{1}, u_{2}, \ldots, u_{s}\right)$ be a random sample of length $s$, in which each element is sampled $u_{k} \sim \mu(\cdot), k=1, \ldots, s$, independently and with replacement. The number of elements sampled is $s=\left\lceil\frac{8 d}{\delta}\right\rceil-1$. Let $Y \subseteq X$ be the set of unique elements in $U$. For every set $R \in \mathcal{P}$, let $M(R, U):=$ $\sum_{i=k}^{s} \mathbf{1}\left[u_{k} \in R\right]$ denote number of (copies of) elements in $U$ that are in $R$. Define $\mathcal{P}_{L} \subseteq \mathcal{P}$ as the sub-collection of "large" sets in packing $\mathcal{P}$ that contain at least $6\left(\frac{8 \overline{d k}}{\delta}\right)$ (copies of) elements in the random sample $U$ :

$$
\mathcal{P}_{L}=\left\{R \in \mathcal{P}: M(R, U) \geq 6 \cdot \frac{8 d k}{\delta}\right\}
$$

It follows that the probability of a given range $R$ in $\mathcal{P}$ being a member of $\mathcal{P}_{L}$ is

$$
\mathbb{P}\left[R \in \mathcal{P}_{L}\right]=\mathbb{P}\left[M(R, U) \geq 6 \cdot \frac{8 d k}{\delta}\right]
$$

Our goal is to show that the collection of large sets $\mathcal{P}_{L}$ has few members in expectation. To do so, it suffices to bound the probability that a fixed set $R$ is a member of $\mathcal{P}_{L}$. This is achieved using Markov's inequality. Recalling that all sets $R \in \mathcal{P}$ have bounded weight $\mu(R) \leq k$ gives

$$
\mathbb{E}[M(R, U)]=\sum_{k=1}^{s} \mathbb{P}\left[u_{k} \in R\right]=\sum_{k=1}^{s} \mu(R) \leq s \cdot k \leq \frac{8 d k}{\delta}
$$

where we used the fact that we sample from $\mu(\cdot)$, which implies that $\mathbb{P}\left[u_{k} \in\right.$ $R]=\mu(R)$. Now, Markov's inequality bounds the probability of $R$ being in $\mathcal{P}_{L}$ :

$$
\begin{aligned}
\mathbb{P}\left[R \in \mathcal{P}_{L}\right] & =\mathbb{P}\left[M(R, U) \geq 6 \cdot \frac{8 d k}{\delta}\right] \\
& \leq \mathbb{P}[M(R, U) \geq 6 \cdot \mathbb{E}[M(R, U)]] \leq 1 / 6
\end{aligned}
$$

Finally, because $\mathcal{P}_{L} \subseteq \mathcal{P}$, we conclude that

$$
\begin{aligned}
\mathbb{E}\left[\operatorname{card}\left(\left.\mathcal{P}\right|_{Y}\right)\right] & \leq \mathbb{E}\left[\operatorname{card}\left(\mathcal{P}_{L}\right)\right]+\mathbb{E}\left[\operatorname{card}\left(\left.\left(\mathcal{P} \backslash \mathcal{P}_{L}\right)\right|_{Y}\right)\right] \\
& \leq \sum_{R \in \mathcal{P}} \mathbb{P}\left[R \in \mathcal{P}_{L}\right]+\operatorname{card}(Y) \cdot \varphi\left(\operatorname{card}(Y), 6 \cdot \frac{8 d k}{\delta}\right) \\
& \leq \frac{1}{6} \operatorname{card}(\mathcal{P})+\frac{8 d}{\delta} \cdot \varphi\left(\frac{8 d}{\delta}, \frac{48 d k}{\delta}\right)
\end{aligned}
$$

where the second-to-last inequality uses the shallow cell complexity of $\mathcal{P}$; the $\operatorname{system}\left(Y,\left.\left(\mathcal{P} \backslash \mathcal{P}_{L}\right)\right|_{Y}\right)$ has at most $\operatorname{card}(Y) \leq s$ elements, and sets have depth at most ( $6 \cdot \frac{8 d k}{\delta}$ ), as the system consists only of cells that are not "large". The final inequality holds because $\mathbb{P}\left[R \in \mathcal{P}_{L}\right] \leq 1 / 6$. Finally, applying Lemma 2 completes the proof.

## 4 Proof of the Main Theorem

Equipped with the Weighted Shallow Packing Lemma, we follow a similar strategy as Mustafa [17]. We state and prove three key lemmas, and finally prove Theorem 3.

### 4.1 Key lemmas

The proof of our main theorem relies on all sets having similar weight. Let $\epsilon$ and $\mu=\left(\mu_{1}, \ldots, \mu_{n}\right)$ be a feasible solution to the LP relaxation (3). By the constraints of the LP, each set $R \in \mathcal{R}$ has weight $\mu(R)=\sum_{j: x_{j} \in R} \mu_{j} \geq \epsilon$. Partition the collection of sets $\mathcal{R}$ into groups $\ell=0,1, \ldots,\lceil\log \epsilon\rceil$ of sets of similar weight; set $R$ belongs to group $\ell$ if and only if $2^{-\ell-1} \epsilon \leq \mu(R)<2^{-\ell} \epsilon$. Because the algorithm exclusively takes independent samples, we can view one run of the algorithm as multiple parallel, independent runs on each group of sets. All our bounds scale on the order $\mathcal{O}\left(1 /\left(2^{-\ell} \epsilon\right)\right)$, so summing over the groups gives a final bound on the order of $\mathcal{O}(1 / \epsilon)$. Hence, we assume henceforth that all sets $R \in \mathcal{R}$ have weight $\epsilon \leq \mu(R) \leq 2 \epsilon$.

The key idea of the proof is to amortize the elements added from each processed unhit set throughout the run of the algorithm. We say a set is processed each time it is flagged as unhit by the oracle, and a sample is taken from it. We bound the total number of elements sampled using weighted $(k, \delta)$-packings on two levels. The first-level packing is an arbitrary maximal packing $\mathcal{P}$ of sets in $\mathcal{R}$. There are a bounded number of sets in $\mathcal{P}$. Next, each processed set $R_{i}$ is assigned to a set in the first-level packing $\mathcal{P}$. For a fixed set $P^{j}$ in the first-level packing, given that it has been assigned processed sets, we show that the collection of sets $R_{i}$ assigned to $P^{j}$ forms a second-level packing. Each second-level packing also has a bounded number of sets. Finally, by bounding the probability that a set in the first-level packing has any sets assigned to it, the total expected number of times the algorithm processes a set is bounded. Note that the assignments are only a tool for analysis; they need not be computed by the algorithm.

We begin by defining the first-level packing. Fix a maximal $(2 \epsilon, \beta \epsilon)$-packing $\mathcal{P}=\left\{P^{1}, \ldots, P^{p}\right\}$, where $p$ denotes the number of sets in the packing. The Shallow Packing Lemma 1 upper bounds the number of sets in the packing by

$$
\begin{equation*}
p \leq \frac{24 d}{\beta \epsilon} \cdot \varphi\left(\frac{8 d}{\beta \epsilon}, \frac{96 d}{\beta}\right) \tag{5}
\end{equation*}
$$

Now, suppose the algorithm runs for $T$ steps, processing sets $\left(R_{1}, \ldots, R_{T}\right)$ in sequence. One given set may be processed multiple times. Denote the sets of
sampled elements $H_{R_{1}}, \ldots, H_{R_{T}}$. The processed sets $R_{i}$ are assigned to sets $P^{j}$ in the first-level packing $\mathcal{P}$ as follows. Arbitrarily assign each set $R_{i}$ to any index $j \in\{1,2, \ldots, p\}$ satisfying $\mu\left(\Delta\left(R_{i}, P^{j}\right)\right)<\beta \epsilon$. Such an index $j$ exists because $\mathcal{P}$ is a maximal $(2 \epsilon, \beta \epsilon)$-packing. It may be the case that $R_{i}=P^{j}$. The next task is to bound the number of sets $R_{i}$ assigned to any set $P^{j}$ in the first-level packing.

Let $n_{j}$ denote the number of processed sets in $\left(R_{1}, \ldots, R_{T}\right)$ assigned to $P^{j} \in$ $\mathcal{P}$. For now, condition on first-level packing set $P^{j}$ having at least one set assigned to it, i.e. $n_{j} \geq 1$. We study the probability of this event later. Relabel the sets and consider them in the order in which they were processed by the algorithm,

$$
\mathcal{S}^{j}=\left(R_{1}^{j}, \ldots, R_{n_{j}}^{j}\right) .
$$

Claim. For all $j \in\{1,2, \ldots, p\}, i \in\left\{1,2, \ldots, n_{j}\right\}$ we have

$$
\begin{equation*}
\mu\left(P^{j} \cap R_{i}^{j}\right)>\frac{\mu\left(P^{j}\right)+\mu\left(R_{i}^{j}\right)-\beta \epsilon}{2} \tag{6}
\end{equation*}
$$

Proof. Fix $j \in\{1,2, \ldots, p\}$. For all $i \in\left\{1,2, \ldots, n_{j}\right\}$ we have

$$
\begin{aligned}
\mu\left(P^{j}\right)+\mu\left(R_{i}^{j}\right) & =\mu\left(P^{j} \backslash R_{i}^{j}\right)+\mu\left(R_{i}^{j} \backslash P^{j}\right)+2 \mu\left(P^{j} \cap R_{i}^{j}\right) \\
& =\mu\left(\Delta\left(P^{j}, R_{i}^{j}\right)\right)+2 \mu\left(P^{j} \cap R_{i}^{j}\right)<\beta \epsilon+2 \mu\left(P^{j} \cap R_{i}^{j}\right)
\end{aligned}
$$

The first equality follows from straightforward accounting, and the second from the definition of symmetric difference. The inequality follows from the manner in which set $R_{i}^{j}$ is matched to the packing set $P^{j}$. Finally, a simple rearrangement of terms yields the result.
This proves that the intersection of each set $R_{i}^{j}$ with its corresponding first-level packing set $P^{j}$ is heavy. This lets us define a second-level packing using the intersections $R_{i}^{j} \cap P^{j}$.

Rather than directly bounding the the number of processed sets assigned to a first-level packing set, it is easier to first bound the length of a random subsequence of the assigned sets $\mathcal{S}^{j}$. For any $j \in\{1, \ldots, p\}$, define the subsequence $\mathcal{S}^{\prime j}$ as the subsequence of processed sets $R$ in $\mathcal{S}^{j}$ whose corresponding samples $H_{R}$ form $\gamma$-nets for the system $\left(R,\left.\mathcal{R}\right|_{R}\right)$ :

$$
\mathcal{S}^{\prime j}=\left(R \in \mathcal{S}^{j}: H_{R} \text { is a } \gamma \text {-net for }\left(R,\left.\mathcal{R}\right|_{R}\right)\right) .
$$

We proceed by bounding the length of the subsequence $\mathcal{S}^{\prime j}$, and by choosing $\gamma$ so as to make it likely for a set $R$ in $\mathcal{S}^{j}$ to be in $\mathcal{S}^{\prime j}$, using the $\epsilon$-net Theorem 1. We use this to upper bound the expected number of sets in $\mathcal{S}^{j}$. Let len $(\mathcal{S})$ denote the length of a sequence $\mathcal{S}$.

The following claim bounds the length of the subsequence above.
Claim. For any $j \in\{1,2, \ldots, p\}$ :

$$
\operatorname{len}\left(\mathcal{S}^{\prime j}\right) \leq \begin{cases}\frac{24 d}{3 / 2-\beta-\gamma} \cdot \varphi\left(\frac{8 d}{3 / 2-\beta-\gamma}, \frac{48 d}{3 / 2-\beta-\gamma}\right), & \text { if } \beta+\gamma \geq 1 / 2  \tag{7}\\ \mathcal{O}(1), & \text { otherwise }\end{cases}
$$

Proof. Let $n_{j}^{\prime}=\operatorname{len}\left(\mathcal{S}^{\prime j}\right)$ and relabel the sets so that $\mathcal{S}^{\prime j}=\left(R_{1}^{j}, \ldots, R_{n_{j}^{\prime}}^{j}\right)$. Now consider an auxiliary sequence of sets based on intersecting the entries $R_{i}^{j}$ in $\mathcal{S}^{\prime j}$ with $P^{j}$ :

$$
\mathcal{T}^{\prime j}=\left(S_{1}^{j}, \ldots, S_{n_{j}^{\prime}}^{j}\right) \quad \text { with } \quad S_{i}^{j}=R_{i}^{j} \cap P^{j} \text { for each } i \in\left\{1, \ldots, n_{i}^{\prime}\right\}
$$

This sequence of sets is used to generate a second-level packing. To do this, consider two distinct set-indices $1 \leq k<l \leq n_{j}^{\prime}$. The points $H_{R_{k}^{j}}$ are added before set $R_{l}^{j}$ is considered, so $H_{R_{k}^{j}}$ is a $\gamma$-net for $\left(R_{k}^{j},\left.\mathcal{R}\right|_{R_{k}^{j}}\right)$, whereas the set $R_{l}^{j}$ - because it is subsequently considered by the algorithm - is not hit by this net. This implies that the intersection of $R_{k}^{j}$ and $R_{l}^{j}$ is of bounded weight, as it would be hit by the $\gamma$-net otherwise:

$$
\mu\left(R_{k}^{j} \cap R_{l}^{j}\right)<\gamma \cdot \mu\left(R_{k}^{j}\right)
$$

This implies that the weight of the intersection of $S_{k}^{j}$ and $S_{l}^{j}$ is bounded above:

$$
\begin{equation*}
\mu\left(S_{k}^{j} \cap S_{l}^{j}\right)=\mu\left(R_{k}^{j} \cap R_{l}^{j} \cap P^{j}\right) \leq \mu\left(R_{k}^{j} \cap R_{l}^{j}\right)<\gamma \cdot \mu\left(R_{k}^{j}\right) \tag{8}
\end{equation*}
$$

The fact that sets in $\mathcal{T}^{\prime j}$ have pairwise intersections of small weight implies that their symmetric differences are heavy:

$$
\begin{aligned}
\mu\left(\Delta\left(S_{k}^{j}, S_{l}^{j}\right)\right) & =\mu\left(S_{k}^{j}\right)+\mu\left(S_{l}^{j}\right)-2 \mu\left(S_{k}^{j} \cap S_{l}^{j}\right) \\
& =\mu\left(R_{k}^{j} \cap P^{j}\right)+\mu\left(R_{l}^{j} \cap P^{j}\right)-2 \cdot \mu\left(S_{k}^{j} \cap S_{l}^{j}\right) \\
& >\frac{\mu\left(P^{j}\right)+\mu\left(R_{k}^{j}\right)-\beta \epsilon}{2}+\frac{\mu\left(P^{j}\right)+\mu\left(R_{l}^{j}\right)-\beta \epsilon}{2}-2 \cdot \mu\left(S_{k}^{j} \cap S_{l}^{j}\right) \\
& >\frac{\mu\left(P^{j}\right)+\mu\left(R_{k}^{j}\right)-\beta \epsilon}{2}+\frac{\mu\left(P^{j}\right)+\mu\left(R_{l}^{j}\right)-\beta \epsilon}{2}-2 \gamma \cdot \mu\left(R_{k}^{j}\right) \\
& =\mu\left(P^{j}\right)-\beta \epsilon+\frac{1}{2} \mu\left(R_{l}^{j}\right)+(1 / 2-2 \gamma) \mu\left(R_{k}^{j}\right) \\
& \geq(3 / 2-\beta-\gamma) \cdot \mu\left(P^{j}\right),
\end{aligned}
$$

where the first inequality uses Eq. (6), the second Eq. (8), and the last exploits the fact that sets $R_{k}^{j}, R_{l}^{j}$, and $P^{j}$ are each of measure at least $\epsilon$ and at most $2 \epsilon$, and that $\gamma \leq 1 / 4$. Thus, depending on the constants, the sequence $\mathcal{T}^{\prime j}$ may form a weighted packing.

Finally, reviewing two cases for the constants $\beta$ and $\gamma$ makes the above more precise. First, if $\beta+\gamma<1 / 2$, the inequality above implies that the symmetric difference of $S_{k}^{j}$ and $S_{l}^{j}$ is strictly larger than $\mu\left(P^{j}\right)$. This cannot be the case as both sets are subsets of $P^{j}$. Thus, the only sequence $\mathcal{S}^{\prime j}$ for which $\beta+\gamma$ can be less than a half is if there are no two unique indices, implying that len $\left(\mathcal{S}^{\prime j}\right) \leq 1$.

Secondly, if $\beta+\gamma \geq 1 / 2$, the sets in $\mathcal{T}^{\prime j}$ form a $\left(\mu\left(P^{j}\right),(3 / 2-\beta-\gamma) \mu\left(P^{j}\right)\right)$ packing over $P^{j}$; all sets have measure at most $\mu\left(P^{j}\right)$, and every symmetric difference is at least $(3 / 2-\beta-\gamma) \mu\left(P^{j}\right)$. This is our second-level packing. Now, the Shallow Packing Lemma 1 implies:

$$
\operatorname{len}\left(\mathcal{S}^{\prime j}\right)=\operatorname{len}\left(\mathcal{T}^{\prime j}\right) \leq \frac{24 d}{3 / 2-\beta-\gamma} \cdot \varphi\left(\frac{8 d}{3 / 2-\beta-\gamma}, \frac{48 d}{3 / 2-\beta-\gamma}\right)
$$

where we have used the fact that $\varphi(\cdot, \cdot)$ is non-decreasing and that $\mu\left(P^{j}\right) \leq 1$.
We can now bound the length of the full sequence of sets assigned to the packing set $P^{j}$. Taking expectations sidesteps any dependencies in the sequences. For instance, a set $R$ can only be in $\mathcal{S}^{j}$ if previous samples failed to hit it. However, for each fixed set $R \in \mathcal{R}$, the probability of the sampled points $H_{S}$ forming a $\gamma$-net for $\left(R,\left.R\right|_{\mathcal{R}}\right)$ is independent of previous sampling. Indeed, by Theorem 1, the probability that $H_{R}$ is a $\gamma$-net is at least $1-\gamma \geq 1 / 2$.

Lemma 3 (Mustafa, Lemma 5 [17]).

$$
\mathbb{E}\left[\operatorname{len}\left(\mathcal{S}^{j}\right) \mid n_{j} \geq 1\right] \leq \frac{48}{3 / 2-\beta-\gamma} \cdot \varphi\left(\frac{8 d}{3 / 2-\beta-\gamma}, \frac{48 d}{3 / 2-\beta-\gamma}\right)
$$

Proof. We use a simple application of linearity of expectation, and Theorem 1:

$$
\mathbb{E}\left[\operatorname{len}\left(\mathcal{S}^{\prime j}\right) \mid n_{j} \geq 1\right]=\sum_{R \in \mathcal{S}^{j}} \mathbb{P}\left[H_{R} \text { is a } \gamma \text {-net for }\left(R,\left.\mathcal{R}\right|_{R}\right)\right] \geq \frac{1}{2} \cdot \operatorname{len}\left(\mathcal{S}^{j}\right)
$$

where we drop the conditioning on $n_{j} \geq 1$ because the event that a particular sample $H_{R}$ is a $\gamma$-net is independent of the number of previous samples. On the other hand, Eq. (7) upper bounds the size of $\operatorname{len}\left(\mathcal{S}^{\prime j}\right)$. Piecing these together yields the inequality:
$\frac{1}{2} \operatorname{len}\left(\mathcal{S}^{j}\right) \leq \mathbb{E}\left[\operatorname{len}\left(\mathcal{S}^{\prime j}\right) \mid n_{j} \geq 1\right] \leq \operatorname{len}\left(S^{\prime j}\right) \leq \frac{24 d}{3 / 2-\beta-\gamma} \varphi\left(\frac{8 d}{3 / 2-\beta-\gamma}, \frac{48 d}{3 / 2-\beta-\gamma}\right)$.

Thus far we have conditioned on a set in the first-level packing being assigned at least one processed set. We now bound the probability of this being the case. Later, this probability is used to compute the expected number of processed sets assigned to a first-level packing set.

Lemma 4. Let $H_{0}$ be the initial sample taken by the algorithm. Then for any $j \in\{1, \ldots, p\}$ :

$$
\mathbb{P}\left[n_{j} \geq 1\right]=\mathcal{O}\left(\frac{1}{d^{2} \varphi\left(\frac{8 d}{\beta \epsilon}, \frac{48 d}{\beta}\right)^{2}}\right)
$$

Proof. Fix an index $j \in\{1, \ldots, p\}$. Suppose that $n_{j} \geq 1$. By Eq. (6), for any $i \in\left\{1, \ldots, n_{j}\right\}$ :

$$
\begin{aligned}
\mu\left(P^{j} \cap R_{i}^{j}\right) & >\frac{\mu\left(P^{j}\right)+\mu\left(R_{i}^{j}\right)-\beta \epsilon}{2} \\
& \geq \frac{\mu\left(P^{j}\right)+\mu\left(P^{j}\right) / 2-\beta \mu\left(P^{j}\right)}{2}=\left(\frac{3}{4}-\frac{\beta}{2}\right) \cdot \mu\left(P^{j}\right)
\end{aligned}
$$

The second inequality above follows from the assumption that all sets have weights within a factor 2 of each other. The above implies that, if $H_{0}$ is a $\left(\frac{3}{4}-\frac{\beta}{2}\right)$-net for $\left(P^{j},\left.\mathcal{R}\right|_{P j}\right)$, then any $R \in \mathcal{S}^{j}$ would be hit by $H_{0}$. In other words, $n_{j} \geq 1$ only if $H_{0}$ is not a $\left(\frac{3}{4}-\frac{\beta}{2}\right)$-net for $\left(P^{j},\left.\mathcal{R}\right|_{P^{j}}\right)$ :

$$
\mathbb{P}\left[n_{j} \geq 1\right] \leq \mathbb{P}\left[H_{0} \text { is not a }\left(\frac{3}{4}-\frac{\beta}{2}\right) \text {-net for }\left(P^{j},\left.\mathcal{R}\right|_{P^{j}}\right)\right] .
$$

Beacuse $\frac{\mu_{j}}{\epsilon} \geq \frac{\mu_{j}}{\mu(R)}$, the initial sample includes each element with sufficient probability to apply Theorem 1 to the RHS above, completing the proof.

### 4.2 Proof of Theorem 3

Proof of Theorem 3. At this stage, the analysis closely follows Mustafa's [17]. Clearly, the algorithm proceeds until $H$ is an $\epsilon^{*}$-net with respect to measure $\mu^{*}$, i.e., a hitting set. It suffices to bound the expected size of the hitting set $H$, as well as the expected number of oracle calls. These quantities are related, since the number of points added depends on the number of times a set is processed.

First, consider the expected size of the hitting set. There are two contributions to the set: the initial sample $H_{0}$, and the samples from the processed sets $H_{R_{1}}, \ldots, H_{R_{T}}$. We bound the expected size of the initial sample first.

Claim. The expected size of the initial sample, $\mathbb{E}\left[\operatorname{card}\left(H_{0}\right)\right]$ is bounded by

$$
\begin{equation*}
\mathcal{O}\left(\frac{1}{\left(\frac{3}{2}-\frac{\beta}{2}\right) \epsilon} \max \left\{\log \left(d \varphi\left(\frac{8 d}{\beta \epsilon}, \frac{48 d}{\beta}\right)\right), \frac{d}{\left(\frac{3}{2}-\frac{\beta}{2}\right)} \log \frac{1}{\left(\frac{3}{2}-\frac{\beta}{2}\right) \epsilon}\right\}\right) . \tag{9}
\end{equation*}
$$

This follows by summing the probability of sampling $x$ for each $x \in X$. An analogous result is used for the number of points added during the processing of a set $R \in \mathcal{R}$, provided it is processed:

Claim. For any fixed set $R \in \mathcal{R}$, conditional on being processed, the expected number of points added each time it is processed is

$$
\begin{equation*}
\mathbb{E}\left[\operatorname{card}\left(H_{R}\right)\right] \leq 2\left(\frac{\log 2}{\gamma}+\frac{d}{\gamma} \log \frac{1}{\gamma}\right)=\mathcal{O}(1) . \tag{10}
\end{equation*}
$$

This bound applies irrespective of whether or not a set was processed previously.
The number of points added during processing, and the number of oracle calls, can be bounded together. Recalling that $R_{1}, \ldots, R_{T}$ are the processed sets, and using the claim above, the number of added elements is at most

$$
\begin{equation*}
\mathbb{E}\left[\sum_{i=1}^{T} \operatorname{card}\left(H_{R_{i}}\right)\right] \leq \mathbb{E}\left[\sum_{i=1}^{T} 2\left(\frac{\log 2}{\gamma}+\frac{d}{\gamma} \log \frac{1}{\gamma}\right)\right]=\mathbb{E}[T] \cdot 2\left(\frac{\log 2}{\gamma}+\frac{d}{\gamma} \log \frac{1}{\gamma}\right) \tag{11}
\end{equation*}
$$

Thus, it suffices to bound the expected number of oracle calls $\mathbb{E}[T]$. This is where we employ both the first-, and second-level packings. In particular

$$
\begin{equation*}
\mathbb{E}[T]=\mathbb{E}\left[\sum_{j=1}^{p} \operatorname{len}\left(\mathcal{S}^{j}\right)\right]=\underbrace{\sum_{j=1}^{p}}_{\text {(i) }} \cdot \underbrace{\mathbb{E}\left[\operatorname{len}\left(\mathcal{S}^{j}\right) \mid n_{j} \geq 1\right]}_{\text {(ii) }} \cdot \underbrace{\mathbb{P}\left[n_{j} \geq 1\right]}_{\text {(iii) }} . \tag{12}
\end{equation*}
$$

The terms (i), (ii) and (iii) are bounded using Eq. (5), Lemma 3, and Lemma 4, respectively. In addition, using $\frac{3}{2}-\beta-\gamma \geq \frac{1}{2} \geq \max \{\beta \epsilon, \beta / 2\}$ :
(i) $p \leq \frac{24 d}{\beta \epsilon} \varphi\left(\frac{8 d}{\beta \epsilon}, \frac{24 d}{\beta}\right)$
(ii) $\mathbb{E}\left[\operatorname{len}\left(\mathcal{S}^{j}\right) \mid n_{j} \geq 1\right] \leq \frac{48 d}{3 / 2-\beta-\gamma} \varphi\left(\frac{8 d}{3 / 2-\beta-\gamma}, \frac{24 d}{3 / 2-\beta-\gamma}\right) \leq \frac{48 d}{3 / 2-\beta-\gamma} \varphi\left(\frac{8 d}{\beta \epsilon}, \frac{24 d}{\beta}\right)$;
(iii) $\mathbb{P}\left[n_{j} \geq 1\right] \leq\left(d^{2} \varphi\left(\frac{8 d}{\beta \epsilon}, \frac{24 d}{\beta}\right)^{2}\right)^{-1}$.

Combining the right-hand-side terms, we obtain the bound

$$
\begin{equation*}
\mathbb{E}[T] \leq \frac{24 \cdot 48}{\beta \cdot(3 / 2-\beta-\gamma)} \frac{1}{\epsilon}=\mathcal{O}\left(\frac{1}{\epsilon}\right) \tag{13}
\end{equation*}
$$

This is minimized by choosing a small $\gamma$, e.g. $\gamma=1 / 100$, and setting $\beta=3 / 4$.
Finally, summing over the $\ell$ groups of sets, and adding the expected number of initial samples to the expected number added points completes the proof. Note that Eq. (13) also bounds the expected number of oracle calls made during the run of the algorithm.

## References

1. Aronov, B., Ezra, E., Sharir, M.: Small-size $\varepsilon$-nets for axis-parallel rectangles and boxes. In: Proceedings of the Forty-First Qnnual ACM Symposium on Theory of Computing. pp. 639-648 (2009)
2. Bar-Yehuda, R., Even, S.: A linear-time approximation algorithm for the weighted vertex cover problem. Journal of Algorithms 2(2), 198-203 (1981)
3. Brönnimann, H., Goodrich, M.T.: Almost Optimal Set Covers in Finite VCDimension. Discrete Comput. Geom. 14, 263-279 (1995)
4. Chan, T.M., Grant, E., Könemann, J., Sharpe, M.: Weighted capacitated, priority, and geometric set cover via improved quasi-uniform sampling. In: Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms. p. 1576-1585. SODA '12, Society for Industrial and Applied Mathematics, USA (2012)
5. Chazelle, B.: A note on Haussler's packing lemma (1992), a note on Haussler's packing lemma
6. Chvatal, V.: A greedy heuristic for the set-covering problem. Mathematics of Operations Research 4(3), 233-235 (1979)
7. Clarkson, K.L.: A randomized algorithm for closest-point queries. SIAM Journal on Computing 17(4), 830-847 (1988)
8. Even, G., Rawitz, D., Shahar, S.M.: Hitting sets when the VC-dimension is small. Information Processing Letters 95(2), 358-362 (2005)
9. Feige, U.: A threshold of $\ln n$ for Approximating Set Cover. Journal of the ACM (JACM) 45(4), 634-652 (1998)
10. Garey, M.R., Johnson, D.S.: Computers and Intractability, vol. 174. Freeman San Francisco (1979)
11. Haussler, D.: Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik-Chervonenkis dimension. Journal of Combinatorial Theory, Series A 69(2), 217-232 (1995)
12. Haussler, D., Welzl, E.: Epsilon-nets and simplex range queries. In: Proceedings of the Second Annual Symposium on Computational Geometry. pp. 61-71 (1986)
13. Johnson, D.S.: Approximation algorithms for combinatorial problems. In: Proceedings of the Fifth Annual ACM Symposium on Theory of Computing. pp. 38-49 (1973)
14. Komlós, J., Pach, J., Woeginger, G.J.: Almost tight bounds for epsilon-nets. Discret. Comput. Geom. 7, 163-173 (1992)
15. Matoušek, J.: Geometric discrepancy: An illustrated guide. Springer (1999)
16. Mustafa, N.H.: A simple proof of the shallow packing lemma. Discret. Comput. Geom. 55(3), 739-743 (2016)
17. Mustafa, N.H.: Computing optimal epsilon-nets is as easy as finding an unhit set. In: Baier, C., Chatzigiannakis, I., Flocchini, P., Leonardi, S. (eds.) 46th International Colloquium on Automata, Languages, and Programming, ICALP 2019, July 9-12, 2019, Patras, Greece. LIPIcs, vol. 132, pp. 87:1-87:12. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2019)
18. Mustafa, N.H.: Sampling in Combinatorial and Geometric Set Systems, vol. 265. American Mathematical Society (2022)
19. Mustafa, N.H., Dutta, K., Ghosh, A.: A simple proof of optimal epsilon nets. Combinatorica 38(5), 1269-1277 (2018)
20. Mustafa, N.H., Varadarajan, K.: Epsilon-approximations \& epsilon-nets. In: Handbook of Discrete and Computational Geometry, pp. 1241-1267. Chapman and Hall/CRC (2017)
21. Varadarajan, K.: Epsilon nets and union complexity. In: Proceedings of the Twenty-Fifth Annual Symposium on Computational Geometry. pp. 11-16 (2009)
22. Yousuf, A.M., Rochester, E.M., Ghaderi, M.: A low-cost LoRaWAN testbed for IoT: Implementation and measurements. In: 2018 IEEE 4th World Forum on Internet of Things (WF-IoT). pp. 361-366 (2018)

## A Proof of the Weighted Packing Lemma

This section proves the Weighted Packing Lemma 2. Our proof closely follows the original in Haussler [11]. The reader is referred to Matoušek [15] (Sec 5.3) for an excellent treatment of the unweighted proof. We begin by restating the weighted lemma.

Lemma 5 (Weighted Packing Lemma). Let $(X, \mathcal{P})$ be a set system with $m$ elements and $n$ sets $\mathcal{P}=\left\{S_{1}, \ldots, S_{n}\right\}$, equipped with weights $\mu: X \rightarrow \mathbb{R}_{\geq 0}$ with $\mu(X)>0$. Let $\mathrm{VC}-\operatorname{dim}(\mathcal{P}) \leq d$ for some integer $d \geq 1$, and let $\delta>0$ be a constant such that $\mu\left(\Delta\left(S_{i}, S_{j}\right)\right) \geq \delta$ for all $1 \leq i<j \leq n$. Then

$$
\operatorname{card}(\mathcal{P}) \leq 2 \mathbb{E}\left[\operatorname{card}\left(\left.\mathcal{P}\right|_{Y}\right)\right]
$$

where $Y$ is the set of unique elements in an random sample $U=\left(u_{1}, u_{2}, \ldots, u_{s}\right)$ of size $s=\left\lceil\frac{8 d}{\delta}\right\rceil-1$, in which each element $u_{k}$ is sampled iid $u_{k} \sim \mu(\cdot)$ with replacement.

The proof strategy is the following. First, we consider a random sample $U=$ $\left(u_{1}, \ldots, u_{s}\right)$ sampled from $\mu(\cdot)$ with replacement. The sample $U$ may contain repeated elements; let $Y \subseteq X$ denote the set of unique elements in $U$. Next, we generate a unit-distance graph; a weighted graph that depends on the random set $Y$, and derive three claims about the total weight of its edges: (i) an upper bound on the total weight, (ii) a partial lower bound, and (iii) a complete lower bound. Combining the bounds completes the proof. The main differences between our approach and that of Chazelle, Haussler, and Mustafa are twofold [5,11,16]. Firstly, we permit non-uniform weights $\mu$ as opposed to weighing each set by its cardinality, and for we sample from a probability distribution proportional to $\mu$. Secondly we use sampling with replacement as opposed to without replacement. This makes the analysis more straightforward under non-uniform sampling using $\mu$.

A weighted unit-distance graph over the sampled set-system takes a central stage in the proof. In this graph, sets are viewed as vertices, and edges are drawn between any two sets at unit-distance of each other.

Definition 2 (Unit distance graph). Let $(X, \mathcal{P})$ be a set system. The unit distance graph of $(X, \mathcal{P})$ is a graph $G=\left(\mathcal{P}, E_{\mathcal{P}}\right)$ with vertex set $\mathcal{P}$ and edges

$$
E_{\mathcal{P}}=\left\{\left\{S_{i}, S_{j}\right\} \in \mathcal{P} \times \mathcal{P}: \operatorname{card}\left(\Delta\left(S_{i}, S_{j}\right)\right)=1 \text { and } i \neq j\right\}
$$

In other words, edges represents pairs of sets that differ on exactly one element. The following lemma connects the number of edges with the VC-dimension.

Lemma 6 (Haussler [11]). Fix a set system $(X, \mathcal{P})$ with $\operatorname{VC}-\operatorname{dim}(\mathcal{P})=d$. Let $G=\left(\mathcal{P}, E_{\mathcal{P}}\right)$ be its unit distance graph. Then $\operatorname{card}\left(E_{\mathcal{P}}\right) \leq d \operatorname{card}(\mathcal{P})$.

We use this lemma to bound the total edge weight in our unit-distance graph. But first, we define our particular unit-distance graph and its edge weights.

We construct a graph that depends on the random set $Y \subseteq X$. Consider the projection of $\mathcal{P}$ to $Y$, denoted $\left.\mathcal{P}\right|_{Y}$. Let $G_{Y}=\left(\left.\mathcal{P}\right|_{Y}, E_{\left.\mathcal{P}\right|_{Y}}\right)$ be a unit distance graph over the projected system. For each vertex $\left.S^{\prime} \in \mathcal{P}\right|_{Y}$, define the vertex weight as the number of sets $S \in \mathcal{P}$ that are projected to $S^{\prime}$ in $\left.\mathcal{P}\right|_{Y}$, that is

$$
w\left(S^{\prime}\right)=\operatorname{card}\left(\left\{S \in \mathcal{P}: S \cap Y=S^{\prime}\right\}\right)
$$

Moreover, define the weight of an edge $\left\{S_{i}^{\prime}, S_{j}^{\prime}\right\} \in E_{\left.\mathcal{P}\right|_{Y}}$ as the minimum over the weights of its two vertices, $w\left(\left\{S_{i}^{\prime}, S_{j}^{\prime}\right\}\right)=\min \left\{w\left(S_{i}^{\prime}\right), w\left(S_{j}^{\prime}\right)\right\}$. Finally, let the total edge weight be $W=\sum_{e \in E_{\left.\mathcal{P}\right|_{Y}}} w(e)$. Note that the weights are random variables because they depend on the random selection $Y$. We proceed by bounding the total edge weight. This will allow us to bound the size of the packing in a way that "looks like a magician's trick" [15].

First, we find an upper bound for the total edge weight. This is naturally also an upper bound on the expected total edge weight.

Claim. The total edge weight is upper-bounded by $W \leq 2 d \operatorname{card}(\mathcal{P})$.
Proof. This is the proof of Haussler and Chazelle [11,5]. Lemma 6 implies that:

$$
\sum_{\left.S^{\prime} \in \mathcal{P}\right|_{Y}} \operatorname{deg}\left(S^{\prime}\right)=2 \operatorname{card}\left(E_{\left.\mathcal{P}\right|_{Y}}\right) \leq 2 d \operatorname{card}\left(\left.\mathcal{P}\right|_{Y}\right)
$$

Hence there exists a vertex $S^{\prime}$ in $\left.\mathcal{P}\right|_{Y}$ with degree at most $2 d$. Each edge incident to $S^{\prime}$ has weight at most $w\left(S^{\prime}\right)$, so the vertex $S^{\prime}$ is responsible for edges of total weight at most $2 d w\left(S^{\prime}\right)$. Applying this inductively proves the bound on the total weight.

$$
W \leq \sum_{\left.S^{\prime} \in \mathcal{P}\right|_{Y}}\left(\text { Weight due to } S^{\prime}\right) \leq 2 d \sum_{\left.S^{\prime} \in \mathcal{P}\right|_{Y}} w\left(S^{\prime}\right)=2 d \operatorname{card}(\mathcal{P})=2 d n
$$

Next we derive a lower bound. It suffices to derive the bound for a reduced problem. Let $U^{\prime}$ be the subsequence of sample $U=\left(u_{1}, \ldots, u_{s}\right)$ containing all but the last element $u_{s}$. Similarly, let $Y^{\prime}$ be the set of unique elements in the subsample $U^{\prime}$, just as $Y$ denotes the unique elements in the full sample $U$. Now, when the final element $u_{s}$ is added to $U^{\prime}$, some vertices in $\left.\mathcal{P}\right|_{Y}$ may form an edge due to $u_{s}$. This occurs exactly when the last sampled element $u_{s}$ falls in the symmetric difference of two sets that were previously equal in the projection $\left.\mathcal{P}\right|_{Y^{\prime}}$. Let $W_{s}$ be the sum of the weights of the edges due to element $u_{s}$. In other words, the weight $W_{s}$ is the weight of the edges generated by adding a random element $u_{s} \sim \mu(\cdot)$ to the random sequence $U^{\prime}=\left(u_{1}, \ldots, u_{s-1}\right)$. Because the samples are iid, given a squence, the ordering of the random elements in the sequence is uniform over all permutations, so we can apply symmetry of expectation:

$$
\mathbb{E}[W]=s \mathbb{E}\left[W_{s}\right]
$$

Hence, it suffices to derive an upper bound on $\mathbb{E}\left[W_{s}\right]$.

Following Haussler [15], an intermediate step is to lower bound the expectation of the weight $W_{s}$ due to the $s$ th element $u_{s}$ conditional on the previous $s-1$ elements $U^{\prime}$. The lower bound follows from considering pairs of vertices that lack edges in $E_{\left.\mathcal{P}\right|_{Y}}$, but that may share an edge after adding $u_{s}$. This happens exactly when the new element $u_{s}$ falls in their symmetric difference. This event occurs with probability at least $\delta$, because we assume $\mu(\Delta(S),) \geq \delta$ for all pairs $S \neq R$ in our packing $\mathcal{P}$. This argument uses the fact that we sample elements proportional to the weight $\mu$.

Claim. For any set $Y^{\prime} \subseteq X$ generated by a fixed sequence $u^{\prime}=\left(u_{1}, \ldots, u_{s-1}\right)$, with $u_{k} \in X$ for $k=1, \ldots, s-1$, it holds that

$$
\mathbb{E}_{\mu}\left[W_{s} \mid U^{\prime}=u^{\prime}\right] \geq \frac{\delta}{2}\left(n-\operatorname{card}\left(\left.\mathcal{P}\right|_{Y^{\prime}}\right)\right)
$$

where $\mathbb{E}_{\mu}$ denotes expectation over is over a single random element $u_{s} \sim \mu(\cdot)$.
Proof. Fix a set $Y^{\prime} \subseteq X$ corresponding to the unique elements in the partial selection $u^{\prime}=\left(u_{1}, \ldots, u_{s-1}\right)$. Consider an arbitrary set $Q$ in the projection $\left.\mathcal{P}\right|_{Y^{\prime}}$. There may be many sets in $\mathcal{P}$ that map to $Q$ in $\left.\mathcal{P}\right|_{Y^{\prime}}$. Let $\mathcal{P}_{Q}$ be the collection of these sets, and let $b$ denote the number of such sets. Note that for any pair of sets $S_{i}, S_{j} \in \mathcal{P}_{Q}, Y^{\prime}$ cannot contain any element in the symmetric difference $\Delta\left(S_{i}, S_{j}\right)$, or else the two sets would not map to the same $Q$. However, when an additional element $u_{s}$ is sampled and added to $Y^{\prime}$ (with the possibility that $u_{s}$ is already in $Y^{\prime}$ ) the collection $\mathcal{P}_{Q}$ is partitioned into two groups: (i) sets that contain $u_{s}$, and (ii) sets that do not contain $u_{s}$. Let $b_{1}$ and $b_{2}$ denote the number of sets in these groups, respectively, with $b=b_{1}+b_{2}$. These two groups are at a unit-distance in $\left.\mathcal{P}\right|_{Y}$. The weight of the resulting edges in the unit-distance graph is exactly $\min \left\{b_{1}, b_{2}\right\}$.

By adding up the expected weights due to each pair we get $\mathbb{E}\left[W_{s}\right]$. For every pair $S_{1}^{\prime}, S_{2}^{\prime} \in \mathcal{P}_{Q}$, the probability that $u_{s}$ hits $\Delta\left(S_{1}^{\prime}, S_{2}^{\prime}\right)$ is $\mu\left(\Delta\left(S_{1}^{\prime}, S_{2}^{\prime}\right)\right) \geq \delta$. Thus, the expected contribution of each pair to the product $b_{1} b_{2}$ is at least $\delta$. Note that the sum $b_{1}+b_{2}=b$ depends on $Y^{\prime}$, however is independent of $u_{s}$. Hence

$$
\begin{aligned}
\mathbb{E}\left[\min \left\{b_{1}, b_{2}\right\}\right] & \geq \mathbb{E}\left[\frac{b_{1} b_{2}}{b}\right] \\
& =\sum_{S_{1}^{\prime}, S_{2}^{\prime} \in \mathcal{P}_{Q}} \frac{\mathbb{P}\left[U_{s} \in \Delta\left(S_{1}^{\prime}, S_{2}^{\prime},\right)\right]}{b} \\
& =\frac{\operatorname{card}\left(\mathcal{P}_{Q}\right)\left(\operatorname{card}\left(\mathcal{P}_{Q}\right)-1\right) \delta}{2 \operatorname{card}\left(\mathcal{P}_{Q}\right)}=\frac{\delta}{2}\left(\operatorname{card}\left(\mathcal{P}_{Q}\right)-1\right) .
\end{aligned}
$$

The first inequality follows from the fact that $\min \left\{b_{1}, b_{2}\right\} \geq b_{1} b_{2} / b$. The first equality uses the fact that there $\mathcal{R}_{Q}$ is partitioned into two groups, each at unit distance, so $b_{1} b_{2}$ is the total number of edges between the two groups. Taking
the above inequality and summing over all vertices $\left.Q \in \mathcal{P}\right|_{Y^{\prime}}$ gives the result

$$
\begin{aligned}
\mathbb{E}\left[W_{s} \mid U^{\prime}=u^{\prime}\right] & \geq \sum_{\left.Q \in \mathcal{P}\right|_{Y^{\prime}}} \frac{\delta}{2}\left(\operatorname{card}\left(\mathcal{P}_{Q}\right)-1\right) \\
& =\frac{\delta}{2}\left(\sum_{\left.Q \in \mathcal{P}\right|_{Y^{\prime}}} \operatorname{card}\left(\mathcal{P}_{Q}\right)-\sum_{\left.Q \in \mathcal{P}\right|_{Y^{\prime}}} 1\right)=\frac{\delta}{2}\left(n-\operatorname{card}\left(\left.\mathcal{P}\right|_{Y^{\prime}}\right)\right) .
\end{aligned}
$$

By using the claim above it is now straightforward to produce a lower bound on the expected total edge weight.

Claim. $\mathbb{E}[W] \geq 4 d n-4 d \mathbb{E}\left[\operatorname{card}\left(\left.\mathcal{P}\right|_{Y^{\prime}}\right)\right]$.
Proof. We employ the reduction from above as well as the partial lower bound. Let $X^{k}$ denote the Cartesian product of element set $X$. For any length sequence $u^{\prime}=\left(u_{1}, \ldots, u_{s-1}\right)$ of elements in $X^{s-1}$, let $Y^{\prime}\left(u^{\prime}\right) \subseteq X$ be the set of unique elements in $u^{\prime}$. Analogously let let $Y^{\prime}\left(U^{\prime}\right)$ denote the random set of unique elements in a random sequence $U^{\prime}$ in $X^{s-1}$. It then follows that:

$$
\begin{aligned}
\mathbb{E}[W] & =s \mathbb{E}\left[W_{s}\right] \\
& =s \sum_{u^{\prime} \in X^{s-1}} \mathbb{E}\left[W_{s} \mid Y^{\prime}\left(U^{\prime}\right)=Y\left(u^{\prime}\right)\right] \cdot \mathbb{P}\left[U^{\prime}=u^{\prime}\right] \\
& \geq s \sum_{u^{\prime} \in X^{s-1}} \frac{\delta}{2}\left(n-\operatorname{card}\left(\left.\mathcal{P}\right|_{Y\left(u^{\prime}\right)}\right)\right) \cdot \mathbb{P}\left[U^{\prime}=u^{\prime}\right] \\
& =\frac{s \delta}{2}\left(n \sum_{u^{\prime} \in X^{s-1}} \mathbb{P}\left[U^{\prime}=u^{\prime}\right]-\sum_{u^{\prime} \in X^{s-1}} \operatorname{card}\left(\left.\mathcal{P}\right|_{Y\left(u^{\prime}\right)}\right) \cdot \mathbb{P}\left[U^{\prime}=u^{\prime}\right]\right) \\
& =\frac{s \delta}{2}\left(n-\mathbb{E}\left[\operatorname{card}\left(\left.\mathcal{P}\right|_{Y^{\prime}\left(U^{\prime}\right)}\right)\right]\right),
\end{aligned}
$$

where the first inequality uses the partial lower bound, and the last equality follows from the definition the random set $Y\left(U^{\prime}\right)$.

Finally, all the pieces are in place to prove the Packing Lemma. Using the lower and upper bounds it follows that cardinality of $\mathcal{P}$ is bounded above by twice the expected size of $\left.\mathcal{P}\right|_{Y^{\prime}}$ :

$$
4 d n-4 d \mathbb{E}\left[\operatorname{card}\left(\left.\mathcal{P}\right|_{Y^{\prime}}\right)\right] \leq \mathbb{E}[W] \leq 2 d n .
$$

This yields the statement of the weighted packing lemma, completing the proof.


[^0]:    ${ }^{1}$ For example when $A$ has bounded row or column sums [2,6].
    ${ }^{2}$ Some definition allow for uncountably many geometric shapes in $\mathcal{R}$, e.g. all squares. However, because the number of points $X$ is finite, there are nevertheless a finite number of unique sets induced by these shapes.

[^1]:    ${ }^{3}$ In addition, it is worth noting that this algorithm can solve the more general weighted hitting set problem, in which each element has a given weight, and the goal is to find the minimum weight hitting set.

[^2]:    ${ }^{4}$ Any non-negative weights $w: X \rightarrow \mathbb{R}_{\geq 0}$ with $w(X)>0$ can be normalized as $\mu(x)=w(x) / w(X)$.

