# The Parameterized Complexity of Unique Coverage and its Variants *

Neeldhara Misra[1], Hannes Moser[3], Venkatesh Raman[1], Saket Saurabh[1], and Somnath Sikdar[2]

[1] The Institute of Mathematical Sciences, Chennai, India.
{neeldhara|saket|vraman}@imsc.res.in
[2] RWTH Aachen University, Germany.
sikdar@cs.rwth-aachen.de
[3] Friedrich-Schiller-Universität, Jena, Germany.
hannes.moser@uni-jena.de

**Abstract.** In this paper we study the parameterized complexity of the Unique Coverage problem, a variant of the classic Set Cover problem. This problem admits several parameterizations and we show that all, except the standard parameterization and a generalization of it, are unlikely to be fixed-parameter tractable. We use results from extremal combinatorics to obtain the best-known kernel for Unique Coverage and the well-known color-coding technique of Alon et al. [3] to show that a weighted version of this problem is fixed-parameter tractable.
Our application of color-coding uses an interesting variation of $s$-perfect hash families called $(k, s)$-hash families which were studied by Alon et al. [1] in the context of a class of codes called parent identifying codes [5]. To the best of our knowledge, this is the first application of $(k, s)$-hash families outside the domain of coding theory. We prove the existence of such families of size smaller than the best-known $s$-perfect hash families using the probabilistic method [2]. Explicit constructions of such families of size promised by the probabilistic method is open.

## 1 Introduction

The Unique Coverage problem was introduced by Demaine et al. [8] as a natural maximization version of Set Cover and has applications in several areas including wireless networks and radio broadcasting. This problem is defined as follows. Given a ground set $\mathcal{U} = \{1, \ldots, n\}$, a family of subsets $\mathcal{F} = \{S_1, \ldots, S_t\}$ of $\mathcal{U}$ and a nonnegative integer $k$, does there exist a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ such that at least $k$ elements are covered uniquely by $\mathcal{F}'$? An element is covered uniquely by $\mathcal{F}'$ if it appears in exactly one set of $\mathcal{F}'$. The optimization version requires us to maximize the number of uniquely covered elements.

We also consider a weighted version of Unique Coverage, called Budgeted Unique Coverage, defined as follows: Given a set $\mathcal{U} = \{1, \ldots, n\}$, a profit $p_i$ for each element $i \in \mathcal{U}$, a family $\mathcal{F}$ of subsets of $\mathcal{U}$, a cost $c_i$ for each set $S_i \in \mathcal{F}$, a budget $B$ and a nonnegative integer $k$, does there exist a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ with total cost at most $B$ such that the total profit of uniquely covered elements is at least $k$? The optimization version asks for a subset $\mathcal{F}'$ of total cost at most $B$ such that the total profit of uniquely covered elements is maximized.

The original motivation for this problem is a real-world application arising in wireless networks [8]. Assume that we are given a map of the densities of mobile clients along with a set of possible base stations, each with a specified building cost and a specified coverage region. The goal is to choose a set of base stations, subject to a budget on the total building cost, in order to maximize

* Preliminary versions of this paper appeared at ISAAC 2007 [18] and at CSR 2009 [17].

the density of served clients. The difficult aspect of this problem is the interference between base stations. A mobile client's reception is better when it is within the range of only a few base stations. An ideal situation is when every mobile client is within the range of exactly one base station. This is the situation modelled by the Budgeted Unique Coverage problem. The Unique Coverage problem is closely related to a single "round" of the Radio Broadcast problem [4]. For more about this relation, see [8].

One can also view the Unique Coverage problem as a generalization of the well-studied Max Cut problem [8]. The input to the Max Cut problem consists of a graph $G = (V, E)$ and the goal is to find a cut $(T, T')$, where $\emptyset \neq T \subset V$ and $T' = V \setminus T$, that maximizes the number of edges with one endpoint in $T$ and the other endpoint in $T'$. Let $\mathcal{U}$ denote the set of edges of $G$ and for each vertex $v \in V$ define $S_v = \{e \in E : e \text{ is incident to } v\}$. Finally let $\mathcal{F} = \bigcup_{v \in V}\{S_v\}$. Then $G$ has a cut $(T, T')$ with at least $k$ edges across it if and only if $\mathcal{F}' = \bigcup_{v \in T}\{S_v\}$ uniquely covers at least $k$ elements of the ground set.

Demaine et al. [8] considered the approximability of Unique Coverage. On the positive side, they give an $O(\log n)$-approximation for Budgeted Unique Coverage. Moreover, if the ratio between the maximum cost of a set and the minimum profit of an element is bounded by $b$, then there exists an $O(\log b)$-approximation algorithm for the weighted version. They show that Unique Coverage is hard to approximate to within a factor of $O(\log^c n)$ for some constant $c$ depending on $\epsilon > 0$, assuming NP $\nsubseteq$ BPTIME$(2^{n^\epsilon})$ for some $\epsilon$. They strengthen this inapproximability to $\epsilon \log n$ for some $\epsilon > 0$ based on a hardness hypothesis for Balanced Bipartite Independent Set.

Erlebach and van Leeuwen [11] study the approximability of geometric versions of the Unique Coverage problem. Among the many versions that they consider is Unique Coverage on Unit Disks, a variant in which each set is a unit disk in $\mathbb{R}^2$, for which they give a factor-18 approximation algorithm. They also consider a variant called Unique Coverage on Disks of Bounded Ply and design an asymptotic fully polynomial-time approximation scheme (FPTAS$^\omega$) for it.

## 2 Results and Organization of this Paper

We begin with the basic notions in parameterized complexity and describe our notation in Section 3. As with many problems in parameterized complexity, the Unique Coverage problem can be parameterized in a number of ways. We first consider an extensive list of plausible parameterizations of the problem in Section 4 and discuss their parameterized complexity. Our results show that barring the standard parameterized version (where the parameter is the number of uniquely covered elements) and a generalization of it, the remaining parameterized problems are unlikely to be fixed-parameter tractable.

In Section 5 we consider the standard parameterized version of Unique Coverage. We show that a special case of this version where any two sets in the input family intersect in at most $c$ elements is fixed-parameter tractable by demonstrating a polynomial kernel of size $k^{c+1}$. Note that the size of an instance $(\mathcal{U}, \mathcal{F}, k)$ of Unique Coverage is $|\mathcal{U}| + |\mathcal{F}|$. Therefore when we say that there exists a kernel for Unique Coverage of size $g(k)$, what we mean is that there exists a kernelization algorithm that produces an equivalent instance $(\mathcal{U}', \mathcal{F}', k')$ such that $|\mathcal{U}'| + |\mathcal{F}'| \leq g(k)$. This leads to a problem kernel of size $k^k$ for the general case, proving that Unique Coverage is fixed-parameter tractable. Then using results from extremal combinatorics on strong systems of distinct representatives we obtain a $4^k$ kernel. In this context, Dom et al. have shown that Unique Coverage does not admit a polynomial kernel unless the Polynomial Hierarchy collapses to the third level [9]. Therefore it is interesting to explore the question of subexponential kernels for the problem.

In Section 6 we consider the Budgeted Unique Coverage problem. For this problem too, there are several variants. If the profits and costs are allowed to be arbitrary positive rational numbers, then Budgeted Unique Coverage, with parameters $B$ and $k$, is not fixed-parameter tractable unless P = NP. If we restrict the costs and profits to be positive integers and parameterize by $B$, then the problem is W[1]-hard. However if we parameterize with respect to both $B$ and $k$ then we show, using an application of the color-coding technique, that the problem is fixed-parameter tractable. In fact, we show that this remains true even when the costs are positive integers and profits are rationals $\geq 1$ or vice versa. While derandomizing our algorithms, we use a variation of $s$-perfect hash families called $(k, s)$-hash families and show the existence of such families of size smaller than the best-known $s$-perfect hash families. We also modify the algorithms for Budgeted Unique Coverage to obtain improved algorithms for two special cases: Unique Coverage and Budgeted Max Cut.

## 3 Preliminaries

We let $\mathbb{N}$ denote the set of natural numbers and $\mathbb{N}_0$ the set $\mathbb{N} \cup \{0\}$. The symbol $\mathbb{Q}$ denotes the set of rationals and $\mathbb{Q}^{\geq a}$, the set of rationals at least $a$. For a positive integer $n$, we let $[n]$ denote the set $\{1, \ldots, n\}$. Logarithms to base 2 are denoted by log and to base $e$ by ln. In what follows, we let $n$ denote the size of the universe and $m$ denote the number of sets in the family of an instance of the Unique Coverage problem, unless otherwise specified.

A parameterized problem $\Pi$ is a subset of $\Sigma^* \times \mathbb{N}_0$, where $\Sigma$ is a finite alphabet and $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. An instance of a parameterized problem is a tuple $(x, k)$, where $k$ is called the parameter. A central notion in parameterized complexity is that of fixed-parameter tractability (fpt). A parameterized problem is *fixed-parameter tractable (fpt)* if there exists an algorithm that decides each instance $(x, k)$ of the problem in time $f(k) \cdot p(|x|)$, where $f$ is some computable function of $k$ and $p$ is a polynomial in the input size. The class of all parameterized problems that are fixed-parameter tractable is denoted by FPT.

Closely related to the concept of fixed-parameter tractability is the notion of a *kernel*.

**Definition 1.** *A* ***kernelization algorithm*** *for a parameterized problem $\Pi$ is an algorithm that takes as input an instance $(x, k)$ of $\Pi$ and outputs an instance $(x', k')$ of $\Pi$ in time polynomial in $|x| + k$ such that*

1. *$(x, k) \in \Pi$ if and only if $(x', k') \in \Pi$; and,*
2. *$|x'|, k' \leq g(k)$, where $g$ is some computable function.*

*The function $g$ is referred to as the* ***size of the kernel****. If $g(k) = k^{O(1)}$ then we say that $\Pi$ admits a* ***polynomial kernel****; if $g(k) = O(k)$ then $\Pi$ admits a* ***linear kernel****.*

A basic result in this area states that a parameterized problem is in the class FPT if and only if it has a kernelization algorithm [12]. For (Budgeted) Unique Coverage whenever we say we have a kernel of size $g(k)$, we mean that the *number of sets* in the family is bounded above by $g(k)$.

Another important notion is that of a *fixed-parameter reduction*.

**Definition 2.** *Let $\Pi_1$ and $\Pi_2$ be parameterized problems. We say that $\Pi_1$* ***fixed-parameter reduces*** *to $\Pi_2$, denoted by $\Pi_1 \leq_{FPT} \Pi_2$, if there exist functions $f, g\colon \mathbb{N}_0 \to \mathbb{N}_0$, $\Phi\colon \Sigma^* \times \mathbb{N}_0 \to \Sigma^*$ and a polynomial $p(\cdot)$ such that for any instance $(I, k)$ of $\Pi_1$,*

1. *$(\Phi(I, k), g(k))$ is an instance of $\Pi_2$ computable in time $f(k) \cdot p(|I|)$;*
2. *$(I, k)$ is a* YES*-instance of $\Pi_1$ if and only if $(\Phi(I, k), g(k))$ is a* YES*-instance of $\Pi_2$.*

*We say that $\Pi_1$ and $\Pi_2$ are* ***fixed-parameter equivalent*** *if and only if $\Pi_1 \leq_{FPT} \Pi_2$ and $\Pi_2 \leq_{FPT} \Pi_1$.*

To capture the notion of hardness, Downey and Fellows [10] defined a hierarchy of classes of parameterized problems

$$\text{FPT} \subseteq \text{W}[1] \subseteq \text{W}[2] \subseteq \cdots \subseteq \text{W}[P]$$

and showed that the parameterized version of CLIQUE is W[1]-complete under fixed parameter reductions, and that of DOMINATING SET is complete for W[2]. We will not formally define the W-hierarchy. It has been conjectured that FPT $\neq$ W[1] and W[1]-hardness is considered as evidence of the fact that a parameterized problem is unlikely to be fixed parameter tractable. For a comprehensive introduction to parameterized complexity see [10, 12, 20].

Given a graph $G(V, E)$, the *open neighborhood* of a vertex $v \in V$, denoted by $N(v)$, is the set of neighbors of $v$ in $G$. The *closed neighborhood* of $v$ is defined as $N[v] := N(v) \cup \{v\}$.

## 4 Unique Coverage: Which Parameterization?

First consider the following parameterized problem: given $(\mathcal{U}, \mathcal{F})$, find a subfamily $\mathcal{F}'$ of $\mathcal{F}$ that covers all of $\mathcal{U}$, each element being covered at most $k$ times and at least once, assuming $k$ as parameter. This is a practical parameterization for mobile-computing applications. Unfortunately, this problem is not fixed-parameter tractable unless P = NP as the case $k = 1$ reduces to the NP-complete EXACT COVER problem [13].

An alternative parameterization with a similar motivation of covering each element a small number of times is as follows: given $(\mathcal{U}, \mathcal{F})$, find a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ of size at most $|\mathcal{F}| - k$ that covers all of $\mathcal{U}$. Call this problem ALL BUT $k$ COVERAGE. We show that this problem is W[1]-hard by a fixed-parameter reduction from the W[1]-complete RED/BLUE NONBLOCKER problem [10]:

| | |
|---|---|
| *Input:* | A bipartite graph $G = (R \uplus B = V, E)$ with its vertex set partitioned into a red set $R$ and a blue set $B$, and a nonnegative integer $k$. |
| *Parameter:* | The integer $k$. |
| *Question:* | Does there exist a set $T$ of at least $k$ red vertices such that every vertex in $B$ is adjacent to at least one vertex in $R - T$? |

A word regarding terminology. If a vertex $u$ is adjacent to $v$, we say that $u$ *dominates* $v$ and that $v$ *is dominated by* $u$. Thus in the RED/BLUE NONBLOCKER problem, one has to decide whether there exists a set $T$ of at least $k$ red vertices such that $R - T$ dominates $B$.

**Theorem 1.** ALL BUT $k$ COVERAGE *is* W[1]*-hard with respect to $k$ as parameter.*

*Proof.* Given an instance $(G = (R \uplus B, E), k)$ of Red/Blue Nonblocker, let $(\mathcal{U}, \mathcal{F}, k')$ be an instance of All But $k$ Coverage defined as follows: $\mathcal{U} := B$, $\mathcal{F} := R$, where each red vertex is interpreted as the set of blue vertices it dominates in $G$, and $k' = k$. Now it is easy to see that there exist at least $k$ red vertices such that the remaining red vertices dominate all blue vertices if and only if there exists a subfamily of $\mathcal{F}$ size at most $|\mathcal{F}| - k$ that covers all of $\mathcal{U}$. □

We also note that the following parameterized versions of Unique Coverage are unlikely to be fixed-parameter tractable. Given $(\mathcal{U}, \mathcal{F})$ and nonnegative integers $k$ and $t$ as parameters,

1. Does there exist a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ of size at most $k$ that covers all of $\mathcal{U}$ with each element being covered at most $t$ times? This version is not fixed-parameter tractable as the case $t = 1$ is W[1]-hard by a reduction from Perfect Code [6] as shown below.
2. Does there exist a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ of size at most $k$ that covers all of $\mathcal{U}$ with each element being covered at most $|\mathcal{F}| - t$ times? When $t = 0$, this is precisely the Set Cover problem, which is W[2]-complete.

Call the version in Item 1 above with $t = 1$ the Disjoint Set Cover problem which we now show to be W[1]-hard by a reduction from Perfect Code [6].

Perfect Code
*Input:* A graph $G = (V, E)$ and an integer $k$.
*Parameter:* The integer $k$.
*Question:* Does $G$ have a $k$-element perfect code?

A perfect code is a vertex subset $V' \subseteq V$ such that for all $u \in V$ we have $|N[u] \cap V'| = 1$, where $N[u]$ is the closed neighborhood of vertex $u$, that is, every vertex is dominated by exactly one vertex in $V'$. Given an instance $(G, k)$ of Perfect Code construct an instance $(\mathcal{U}, \mathcal{F}, k)$ of Disjoint Set Cover by setting $\mathcal{U} := V(G)$ and $\mathcal{F} := \{N[v] : v \in V\}$.

**Lemma 1.** *The graph $G$ has a $k$-element perfect code if and only if there exists a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ of pairwise disjoint sets of size at most $k$ that covers all of $\mathcal{U}$.*

*Proof.* Let $\{v_1, \ldots, v_k\} \subseteq V$ be a $k$-element perfect code of $G$. Clearly $\bigcup_{i=1}^{k} N[v_i]$ covers all of $\mathcal{U}$ and the sets $N[v_1], \ldots, N[v_k]$ are pairwise disjoint. For if $x \in N[v_i] \cap N[v_j]$, $i \neq j$, then $|N[x] \cap \{v_1, \ldots, v_k\}| \geq 2$, which contradicts the definition of a perfect code. Conversely if $N[v_1], \ldots, N[v_k]$ is a collection of pairwise disjoint sets that covers all of $\mathcal{U}$ then clearly $v_1, \ldots, v_k$ is a $k$-element perfect code for $G$. □

**Theorem 2.** Disjoint Set Cover *is* W[1]*-hard with respect to the number of sets in the solution as parameter.*

Finally we consider a generalization of the standard parameterized version: Gen Unique Coverage: Given $(\mathcal{U}, \mathcal{F})$ and nonnegative integers $k$ and $t$, does there exist a subfamily of $\mathcal{F}$ that covers $k$ elements at least once and at most $t$ times, where $k$ is the parameter? Setting $t = 1$ gives us the standard parameterized version of Unique Coverage. We note that Gen Unique Coverage is fixed-parameter tractable as the kernelization algorithm for the standard parameterized version works for this problem also. We elaborate this further in Section 5.2.

## 5 Unique Coverage: The Standard Version

In this section we study the standard parameterized version of UNIQUE COVERAGE. Let $(\mathcal{U} = \{1, \ldots, n\}, \mathcal{F} = \{S_1, \ldots, S_m\}, k)$ be an instance of this problem, where $k$ is the parameter. Apply the following rules on this instance until no longer applicable.

**R1** If there exists $S_i \in \mathcal{F}$ such that $|S_i| \geq k$, then the given instance is a yes-instance.
**R2** If there exists $S_1, S_2 \in \mathcal{F}$ such that $S_1 = S_2$, then delete $S_1$.

It is easy to see that these are indeed reduction rules for UNIQUE COVERAGE. For in the first case $S_i$ is a solution and in the second case, it is clear that no solution need have both $S_1$ and $S_2$. In the following we always assume that the given instance of UNIQUE COVERAGE is reduced with respect to the above rules.

As a warm-up, we first begin with the simple case where each element of $\mathcal{U}$ is contained in at most $b$ sets of $\mathcal{F}$. A special case of this situation is MAX CUT where $b = 2$.

**Lemma 2.** *If each element $e \in \mathcal{U}$ occurs in at most $b$ sets of $\mathcal{F}$ then the* UNIQUE COVERAGE *problem admits a kernel of size $b(k-1)$, that is, the number of sets in the resulting family is at most $b(k-1)$.*

*Proof.* Find a maximal collection $\mathcal{F}'$ of pairwise disjoint sets in $\mathcal{F}$. If $|\bigcup_{S \in \mathcal{F}'} S| \geq k$, we are done. Therefore assume that $|\bigcup_{S \in \mathcal{F}'} S| \leq k-1$. Since every set in $\mathcal{F} - \mathcal{F}'$ intersects some set in $\mathcal{F}'$, and every element of $\mathcal{U}$ occurs in at most $b$ sets in $\mathcal{F}$, we have $|\mathcal{F} - \mathcal{F}'| \leq (k-1)(b-1)$. But $|\mathcal{F}'| \leq k-1$ and so $|\mathcal{F}| \leq b(k-1)$. □

### 5.1 Bounded Intersection Size

Now consider the situation where for all $S_i, S_j \in \mathcal{F}$, $i \neq j$, we have $|S_i \cap S_j| \leq c$, for some constant $c$. In this case we say that the problem instance has *intersection size bounded by* $c$ and show that the problem admits a polynomial kernel of size $O(k^{c+1})$. First consider the case when $|S_i \cap S_j| \leq 1$.

**Lemma 3.** *Suppose that for all $S_i, S_j \in \mathcal{F}$, $i \neq j$, we have $|S_i \cap S_j| \leq 1$. If an element $e \in \mathcal{U}$ is covered by at least $k+1$ sets, then one can obtain a solution covering $k$ elements uniquely in polynomial time.*

*Proof.* Suppose an element $e \in \mathcal{U}$ is covered by the sets $S_1, \ldots, S_{k+1}$. Then by reduction rule R2, at most one of these sets can have size 1. The remaining $k$ sets uniquely cover at least one distinct element each. □

One can now easily obtain a kernel of size $k^2$ for the case when the intersection size is at most 1.

**Lemma 4.** *Suppose that for all $S_i, S_j \in \mathcal{F}$, $i \neq j$, we have $|S_i \cap S_j| \leq 1$. If $|\mathcal{F}| \geq k^2$, then there exists $\mathcal{T} \subseteq \mathcal{F}$ that covers at least $k$ elements uniquely.*

*Proof.* If an element appears in at least $k+1$ sets then we are done by Lemma 3. Otherwise every element appears in at most $k$ sets and by Lemma 2 we have a kernel of size $k(k-1) < k^2$. □

Next, we generalize these observations to the case when $|S_i \cap S_j| \leq c$, for some constant $c$.

**Theorem 3.** *Suppose that for all $S_i, S_j \in \mathcal{F}$, $i \neq j$, we have $|S_i \cap S_j| \leq c$, for some positive constant $c$. If $|\mathcal{F}| \geq k^{c+1}$ then in polynomial time one can find a subset $\mathcal{T} \subseteq \mathcal{F}$ that covers $k$ elements uniquely.*

*Proof.* By induction on $c$. For $c = 1$, this follows from Lemma 4. Assume the theorem to hold for $c \geq 1$. Let $c \geq 2$. Greedily obtain a maximal collection $\mathcal{F}' = \{S_1, \ldots, S_p\}$ of pairwise disjoint sets. If $|\bigcup_{S_i \in \mathcal{F}'} S_i| \geq k$ then we are done. Therefore assume $|\bigcup_{S \in \mathcal{F}'} S| \leq k-1$ (this also implies $p \leq k-1$). Since $|\mathcal{F}| \geq k^{c+1}$, and since every set in $\mathcal{F}$ intersects with at least one set in $\mathcal{F}'$, there exists $e \in \bigcup_{S \in \mathcal{F}'} S$ such that at least $k^c + 1$ sets in $\mathcal{F} - \{S_1, \ldots, S_p\}$ contain $e$. For otherwise, every element occurs in at most $k^c$ sets and by Lemma 2, $|\mathcal{F}| \leq (k-1)k^c + p < k^{c+1}$, a contradiction. Let $T_1, \ldots, T_{k^c+1}$ be some $k^c + 1$ such sets. Delete $e$ from each of these sets. We obtain at least $k^c$ nonempty distinct sets $T'_1, \ldots, T'_{k^c}$ (there is at most one set consisting only of the element $e$ which is deleted in this process). Note that any two of these sets intersect in at most $c-1$ elements. By induction hypothesis, there exists a collection $\mathcal{T}' \subseteq \{T'_1, \ldots, T'_{k^c}\}$ that uniquely covers at least $k$ elements, and thus there exists a collection $\mathcal{T} \subseteq \{T_1, \ldots, T_{k^c}\}$ that uniquely covers at least $k$ elements (just take the solution for $T'$ and add $e$ to every set in it). This proves the theorem. $\square$

**Corollary 1.** Unique Coverage *admits a kernel of size $k^{c+1}$ in the case where any two sets in the family intersect in at most $c$ elements.*

By reduction rule R1 we have $c \leq k-1$ and therefore for the general case we have a kernel of size $k^k$.

**Corollary 2.** Unique Coverage *is fixed-parameter tractable and admits a problem kernel of size $k^k$.*

An algorithm that checks all possible subsets of a family of size $k^k$ to see whether any of them uniquely covers at least $k$ elements is an FPT-algorithm with time complexity $O(2^{k^k} \cdot n^{O(1)})$. But note that we can assume without loss of generality that every set in the solution covers at least one element uniquely. Thus it suffices to check whether subfamilies of size at most $k$ uniquely cover at least $k$ elements. This can be done in time $O(k^{k^2} \cdot n^{O(1)}) = O(2^{k^2 \log k} \cdot n^{O(1)})$. However, it turns out that a much better kernel and a better running time can be obtained for the general case.

### 5.2 A Better Kernel for the General Case

We now show that Unique Coverage has a kernel of size $4^k$ using a result on strong systems of distinct representatives. Given a family of sets $\mathcal{F} = \{S_1, \ldots, S_m\}$, *a system of distinct representatives* for $\mathcal{F}$ is an $m$-tuple $(x_1, \ldots, x_m)$ where the elements $x_i$ are distinct and $x_i \in S_i$ for all $i = 1, 2, \ldots, m$. Such a system is *strong* if we additionally have $x_i \notin S_j$ for all $i \neq j$.

**Theorem 4 ([14]).** *In any family of more than $\binom{r+s}{s}$ sets of cardinality at most $r$, at least $s+2$ of its members have a strong system of distinct representatives.*

Given an instance $(\mathcal{U}, \mathcal{F}, k)$ of Unique Coverage, put $r = k-1$ and $s = k-2$ in the statement of the above theorem and we have a kernel of size $\binom{2k-3}{k-2} \leq \binom{2k}{k} \leq 2^{2k}$.

**Corollary 3.** Unique Coverage *admits a problem kernel of size $4^k$.*

*Proof.* Given an instance $(\mathcal{U}, \mathcal{F}, k)$ of Unique Coverage, if $|\mathcal{F}| \geq 4^k$ then the instance is trivially a yes-instance. Hence the result follows.

As noted before, Dom et al. have shown that Unique Coverage does not admit a kernel of size polynomial in $k$ unless the Polynomial Hierarchy collapses to the third level [9]. It is open as to whether there exists a subexponential (or even better than $4^k$) kernel for Unique Coverage.

**Corollary 4.** *There is an $O(4^{k^2} \cdot n^{O(1)})$ time algorithm for the* Unique Coverage *problem.*

*Proof.* A subfamily that covers $k$ elements uniquely has size at most $k$. Therefore it is sufficient to consider all possible size-$k$ subfamilies of the $4^k$-kernel. This takes time $O(4^{k^2} \cdot n^{O(1)})$. ◻

In Section 7, we provide a better algorithm with running time $O(2^{O(k \log \log k)} \cdot mk + m^2)$ for Unique Coverage. Now we note that Gen Unique Coverage (see Section 4) is fixed-parameter tractable. Recall the problem definition: Given $(\mathcal{U}, \mathcal{F})$ and nonnegative integers $k$ and $t$, does there exist a subfamily of $\mathcal{F}$ that covers $k$ elements at least once and at most $t$ times? Here $k$ is the parameter. An instance $(\mathcal{U}, \mathcal{F}, k, t)$ is trivially a yes-instance if $|\mathcal{F}| > 4^k$. Otherwise $|\mathcal{F}| \leq 4^k$ and we have a kernel.

**Corollary 5.** *The problem* Gen Unique Coverage*, a generalization of* Unique Coverage*, where one has to cover at least $k$ elements once and at most $t$ times is fixed-parameter tractable w.r.t. $k$ as parameter.*

For the case where each set of the input family has size at most $b$, for some constant $b$, there is a better kernel. By Theorem 4, if there exists at least $\binom{b+k}{k}$ sets in the input family, then there exists at least $k$ sets with a strong system of distinct representatives.

**Corollary 6.** *If each set $S \in \mathcal{F}$ has size at most $b$ then the* Unique Coverage *problem has a kernel of size $O(2^{b+k})$.*

## 6 Budgeted Unique Coverage

In this section we consider the Budgeted Unique Coverage problem where each set in the input family has a cost and each element in the universe has a profit; the goal is to decide whether there exists a subfamily of total cost at most $B$ that uniquely covers elements of total profit at least $k$. By parameterizing on $k$ or $B$ or both we obtain different parameterized versions of this decision question.

### 6.1 Intractable Parameterized Versions

We first consider the Budgeted Max Cut problem which is a specialization of the Budgeted Unique Coverage problem. An instance of this problem is an undirected graph $G = (V, E)$ with a cost function $c : V \to \mathbb{Q}^+$ on the vertex set and a profit function $p : E \to \mathbb{Q}^+$ on the edge set; positive rational numbers $B$ and $k$. The question is whether there exists a cut $(T, T')$ such that the total cost of the vertices in $T$ is at most $B$ and the total profit of the edges crossing the cut is at least $k$.

We first show that the Budgeted Max Cut problem with arbitrary positive rational costs and profits is probably not in FPT.

**Lemma 5.** *The* BUDGETED MAX CUT *problem with arbitrary positive rational costs and profits with parameters $B$ and $k$ is not in* FPT*, unless* P = NP*.*

*Proof.* Suppose there exists an algorithm for the BUDGETED MAX CUT problem (with arbitrary positive costs and profits) with run-time $O(f(k, B) \cdot p(n))$, where $p$ is a polynomial in $n$. We will use this to solve the decision version of MAX CUT in polynomial time. Let $(G = (V, E), k)$ be an instance of the MAX CUT problem, where $|V| = n$. Assign each vertex of the input graph cost $1/n$ and each edge profit $1/k$. Let the budget $B = 1/2$ and the profit $k' = 1$. Clearly, $G$ has a maximum cut of size at least $k$ if and only if there exists $S \subseteq V$ of total cost at most $B$ such that the total profit of the edges crossing the cut $(S, V - S)$ is at least $k'$. And this can be answered in time $O(f(1, 1/2) \cdot p(|V|))$, implying P = NP. □

**Theorem 5.** *The* BUDGETED UNIQUE COVERAGE *problem with arbitrary positive rational costs and profits is not in* FPT*, unless* P = NP*.*

We next show that the BUDGETED MAX CUT problem parameterized by the budget $B$ alone is W[1]-hard even when the costs and profits are positive integers.

**Lemma 6.** *The* BUDGETED MAX CUT *problem with positive integer costs and profits, parameterized by the budget $B$ is* W[1]*-hard.*

*Proof.* To show W[1]-hardness, we exhibit a fixed-parameter reduction from the INDEPENDENT SET problem to the BUDGETED MAX CUT problem with unit costs and profits. Let $(G = (V, E), B)$ be an instance of INDEPENDENT SET with $|V| = n$. For every vertex $u \in V$ add $|V| - 1 - \deg(u)$ new vertices and connect them to $u$. Call the resulting graph $G'$. Define $c(v) = 1$ for all $v \in V(G')$ and $p(e) = 1$ for all $e \in E(G')$. Note that every vertex $u \in G$ has degree $|V| - 1$ in $G'$. We let $(G' = (V', E'), B, k = B(n - 1))$ be the instance of BUDGETED MAX CUT.

*Claim.* The graph $G$ has an independent set of size $B$ if and only if $G'$ has a cut $(S, V' - S)$ such that $|S| = B$ and at least $k = B(n - 1)$ edges lie across it.

If $G$ has an independent set $S$ of size $B$, then clearly $S$ is independent in $G'$. The cut $(S, V' - S)$ does indeed have $B(n - 1)$ edges crossing it, as every vertex of $S$ has degree $n - 1$. Next suppose that $G'$ has a cut $(S, V' - S)$ with $B(n - 1)$ edges crossing it such that $|S| = B$. Note that every vertex in $S$ must be a vertex from $G$. Otherwise the cut cannot have $B(n - 1)$ edges crossing it. Suppose two vertices $u$ and $v$ in $S$ are adjacent. Then both $u$ and $v$ contribute less than $n - 1$ edges to the cut. Since each vertex in $S$ contributes at most $n - 1$ edges to the cut, the number of edges crossing the cut must be less than $B(n - 1)$, a contradiction. Hence $S$ is independent in $G'$ and and $G$ has an independent set of size $B$. □

Since the BUDGETED UNIQUE COVERAGE problem is a generalization of BUDGETED MAX CUT we have the following theorem.

**Theorem 6.** *The* BUDGETED UNIQUE COVERAGE *problem with positive integer costs and profits, parameterized by the budget $B$ is* W[1]*-hard.*

### 6.2 Parameterizing by both $B$ and $k$

We now assume that, unless otherwise mentioned, both costs and profits are positive integers and that both $B$ and $k$ are parameters. Let $(\mathcal{U}, \mathcal{F}, c, p, B, k)$ be an instance of Budgeted Unique Coverage. We may assume that for all $S_i, S_j \in \mathcal{F}$, $i \neq j$, we have

1. $S_i \neq S_j$;
2. $c(S_i) \leq B$;
3. $p(S_i) \leq k - 1$;
4. $B \geq 2$.

For if $c(S_i) > B$ then $S_i$ cannot be part of any solution and may be discarded; if $p(S_i) \geq k$ then the given instance is trivially a yes-instance. Note that the condition $p(S_i) \leq k - 1$ implies that $|S_i| \leq k - 1$. Also observe that if the third condition above holds then we must have $B \geq 2$, for otherwise the given instance is a no-instance.

For some of the results in this section, we need the following lemma.

**Lemma 7.** *For all nonnegative integers $t$ and $k$ such that $t \geq 2k$, we have $\left(\frac{t-k}{t}\right)^t \geq (2e)^{-k}$.*

*Proof.* We show that $\left(\frac{t}{t-k}\right)^t \leq (2e)^k$ for all $t \geq 2k$.

$$\begin{aligned}
\left(\frac{t}{t-k}\right)^t &= \left(1 + \frac{k}{t-k}\right)^{t-k} \cdot \left(1 + \frac{k}{t-k}\right)^k \\
&\leq e^k \cdot \left(1 + \frac{k}{t-k}\right)^k \\
&\leq e^k \cdot 2^k = (2e)^k
\end{aligned}$$

The last inequality follows since $t \geq 2k$. □

Demaine et al. [8] show that there exists an $\Omega(1/\log n)$-approximation algorithm for Budgeted Unique Coverage (Theorem 4.1). We use the same proof technique to show the following.

**Lemma 8.** *Let $(\mathcal{U}, \mathcal{F}, c, p, B, k)$ be an instance of* Budgeted Unique Coverage *and let $c : \mathcal{F} \to \mathbb{Q}^{\geq 1}$ and $p : \mathcal{U} \to \mathbb{Q}^{\geq 1}$. Then either*

1. *in polynomial time, one can find a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ with total cost at most $B$ such that the total profit of elements uniquely covered by $\mathcal{F}'$ is at least $k$; or*
2. *for every subfamily $\mathcal{H}$ with total cost at most $B$, we have $|\bigcup_{S \in \mathcal{H}} S| \leq 18k \log B$.*

*Proof.* Let $\mathcal{H} = \{S_1, \ldots, S_r\}$ be a subfamily of $\mathcal{F}$ with budget at most $B$ which maximizes $|\mathcal{U}'|$, where $\mathcal{U}' = \bigcup_{S \in \mathcal{H}} S$. For $u \in \mathcal{U}$, let $f_u$ denote the number of sets of $\mathcal{H}$ containing $u$. Partition $\mathcal{U}'$ into sets $C_0, C_1 \ldots C_{t-1}$, such that $u \in C_i$ if $2^i \leq f_u \leq 2^{i+1} - 1$. Note that $i$ ranges from 0 to $\log(r+1) - 1$, since the frequency of any element in $C_{t-1}$ is at most $2^t - 1$, which in turn is at most $r$. So $t \leq \log(r+1)$.

Clearly there exists $j$ such that $|C_j| \geq |\mathcal{U}'|/\log(r+1)$. Fix $j$ to be an index for which $|C_j| \geq |\mathcal{U}'|/\log(r+1)$. Fix $u \in C_j$ and note that $2^j \leq f_u \leq 2^{j+1} - 1$. Construct a subfamily $\mathcal{F}'$ from $\mathcal{H}$

by going through each set in $\mathcal{H}$ and including it in $\mathcal{F}'$ with probability $1/2^{j+1}$. Denote by $l_u$ the probability that $u$ is covered uniquely by $\mathcal{F}'$. Then

$$\begin{aligned}
l_u &= \left(\tfrac{f_u}{2^{j+1}}\right)\left(1-\tfrac{1}{2^{j+1}}\right)^{(f_u-1)} \\
&\geq \tfrac{1}{2}\left(1-\tfrac{1}{2^{j+1}}\right)^{(f_u-1)} && \text{(since } f_u \geq 2^j\text{).} \\
&\geq \tfrac{1}{2}\left(1-\tfrac{1}{2^{j+1}}\right)^{(2^{j+1}-2)} && \text{(since } f_u \leq 2^{j+1}-1\text{).} \\
&\geq \tfrac{1}{2}\left(1-\tfrac{1}{2^{j+1}}\right)^{(2^{j+1})} \\
&\geq \tfrac{1}{4e}. && \text{(by Lemma 7).}
\end{aligned}$$

Let $X_u$ be an indicator random variable which takes value 1 if $u$ is covered uniquely by the subfamily $\mathcal{F}'$, and 0 otherwise. Also, let $X = \sum_u X_u$. Then

$$E(X) = \sum_u E(X_u) \geq \sum_{u \in C_j} E(X_u) = \sum_{u \in C_j} l_u \geq \frac{|\mathcal{U}'|}{4e\log(r+1)}.$$

Let the set of elements uniquely covered by $\mathcal{F}'$ be $Q$. Then the total profit of these uniquely covered elements is at least $|Q|$ and

$$E(|Q|) \geq \frac{1}{4e} \cdot \frac{|\mathcal{U}'|}{\log(r+1)},$$

as every uniquely covered element contributes at least 1 to the total profit. Since $r$ is bounded above by $B$, the total expected profit of these elements is at least

$$\frac{1}{4e} \cdot \frac{|\mathcal{U}'|}{\log(B+1)}. \tag{1}$$

This implies that if $k$ is at most the quantity in expression 1 then the total profit of uniquely covered elements of subfamily $\mathcal{F}'$ is at least $k$; else, we have

$$|\mathcal{U}'| \leq 4ek\log(B+1) \leq 12k\log B.$$

We can design an algorithm that finds a subfamily of total budget at most $B$ that uniquely covers elements with total profit at least $k$ as follows. Define $p'$ to be the unit profit function, that is, $p'(u) = 1$ for all $u \in \mathcal{U}$. We find a subfamily $\mathcal{H}'$ in place of $\mathcal{H}$ using an approximation algorithm for the Maximum Coverage problem (where the objective is to maximize the elements covered by the subfamily). To do this we run the polynomial-time $(1-1/e)$-ratio approximate algorithm described in [15] for the Budgeted Maximum Coverage problem with universe $\mathcal{U}$, family $\mathcal{F}$, cost function $c$, profit function $p'$ and budget $B$. This returns a subfamily $\mathcal{H}'$ of total cost at most $B$ that covers at least $(1-1/e)\cdot \textsc{opt} = d \cdot \textsc{opt}$ elements, where $\textsc{opt}$ denotes the maximum number of elements in any family of total cost at most $B$. From $\mathcal{H}'$ one can deterministically obtain a subfamily $\mathcal{F}'$ that uniquely covers elements with total profit at least

$$\frac{1}{4e} \cdot \frac{d \cdot \textsc{opt}}{\log(B+1)}$$

by the method of conditional probabilities (see [19]).

Hence depending on the value of $k$ we can, in polynomial time, either obtain a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ with budget at most $B$ such that the total profit of uniquely covered elements is at least $k$; or every subfamily $\mathcal{H}$ with budget at most $B$ contains at most $\frac{4e}{1-1/e} \cdot k\log(B+1) \leq 18k\log B$ elements. $\square$

The first step of our algorithm is to apply Step 1 of Lemma 8. Therefore from now on we assume that every subfamily of total cost at most $B$ covers at most $18k \log B$ elements of the universe.

We now proceed to show that Budgeted Unique Coverage is in FPT by an application of the color-coding technique. We first show this for the case when the costs and profits are all one and then handle the more general case of integral costs and profits. Therefore let $(\mathcal{U}, \mathcal{F}, B, k)$ be an instance of Budgeted Unique Coverage with unit costs and profits. For this version of the problem, we have to decide whether there exists a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ of size at most $B$ that uniquely covers at least $k$ elements.

To develop our color-coding algorithm, we use two sets of colors $\mathcal{C}_g$ and $\mathcal{C}_b$ with the understanding that the (good) colors from $\mathcal{C}_g$ are used for the elements that are uniquely covered and the (bad) colors from $\mathcal{C}_b$ are used for the remaining elements. In the present setting, $\mathcal{C}_g = \{1, \ldots, k\}$ and $\mathcal{C}_b = \{k+1\}$.

We now describe the notion of a *good configuration*. Given a function $h\colon \mathcal{U} \to \mathcal{C}_g \uplus \mathcal{C}_b$ and $\mathcal{F}' \subseteq \mathcal{F}$, define $\mathcal{U}(\mathcal{F}')$ to be the set of elements covered (not necessarily uniquely) by $\mathcal{F}'$ and $h(\mathcal{F}')$ to be the set of colors assigned to the elements in $\mathcal{U}(\mathcal{F}')$. That is,

$$\mathcal{U}(\mathcal{F}') := \bigcup_{S \in \mathcal{F}'} S; \qquad h(\mathcal{F}') := \bigcup_{i \in \mathcal{U}(\mathcal{F}')} \{h(i)\}.$$

**Definition 3.** *Given $h\colon \mathcal{U} \to \mathcal{C}_g \uplus \mathcal{C}_b$ and $\mathcal{C}'_g \subseteq \mathcal{C}_g$, we say that $\mathcal{F}' \subseteq \mathcal{F}$* ***has a good configuration with respect to*** *$h$* ***and*** *$\mathcal{C}'_g$ if*

1. *$h(\mathcal{F}') \cap \mathcal{C}_g = \mathcal{C}'_g$, and*
2. *there are exactly $|\mathcal{C}'_g|$ elements in $\mathcal{F}'$ that are assigned colors from $\mathcal{C}'_g$ and these elements are uniquely covered by $\mathcal{F}'$.*

*We also say that $\mathcal{F}$* has a good configuration with respect to $h$ and $\mathcal{C}'_g$ *if there exists a subfamily $\mathcal{F}'$ with a good configuration with respect to $h$ and $\mathcal{C}'_g$. Call $\mathcal{F}'$ a* ***witness subfamily****.*

A *solution subfamily* (for the unit costs and profits version) is a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ with at most $B$ sets and which uniquely covers at least $k$ elements.

The next lemma shows that if $h$ is chosen uniformly at random from the space of all functions from $\mathcal{U}$ to $[k+1]$ and $(\mathcal{U}, \mathcal{F}, B, k)$ is a yes-instance of Budgeted Unique Coverage with unit costs and profits, then with probability roughly $2^{-k \log k}$ a solution subfamily $\mathcal{F}'$ has a good configuration with respect to $h$ and $\mathcal{C}_g$. Note that such a uniformly chosen $h$ maps every element from $\mathcal{U}$ uniformly at random to an element in $[k+1]$.

**Lemma 9.** *Let $(\mathcal{U}, \mathcal{F}, B, k)$ be a yes-instance of* Budgeted Unique Coverage *with unit costs and profits and let $h : \mathcal{U} \to [k+1]$ be a function chosen uniformly at random. Then a solution subfamily $\mathcal{F}'$ has a good configuration with respect to $h$ and $\mathcal{C}_g$ with probability at least*

$$2^{-k(18 \log B \log(k+1) - \log k + \log e)}.$$

*Proof.* Let $\mathcal{F}'$ be a solution subfamily with at most $B$ sets that covers the elements $Q = \{i_1, \ldots, i_k\}$ uniquely. Then by Lemma 8, we have $p := |\mathcal{U}(\mathcal{F}')| \leq 18k \log B$. To complete the proof, we show that $\mathcal{F}'$ has a good configuration with respect to $h$ and $\mathcal{C}_g$ with the stated probability. For $\mathcal{F}'$ to have a good configuration, two conditions must be satisfied: $h(i) = k+1$ for all $i \in \mathcal{U}(\mathcal{F}') \setminus Q$ and $h(i_1), \ldots, h(i_k)$ must be a permutation of $1, \ldots, k$. The probability **Pr** that this happens is:

$$\begin{aligned}
\mathbf{Pr} &= \frac{1}{(k+1)^{|\mathcal{U}(\mathcal{F}')\setminus Q|}} \times \frac{k!}{(k+1)^k} \\
&\geq \left(\frac{k}{e}\right)^k \frac{1}{(k+1)^p} = e^{k\ln(k/e) - p\ln(k+1)} \\
&\geq e^{k\ln(k/e) - 18k\log B\ln(k+1)} \\
&= 2^{-k(18\log B\log(k+1) - \log k + \log e)}
\end{aligned}$$

This proves the lemma. □

Given a coloring $h$, how do we find out whether $\mathcal{F}$ has a good configuration with respect to $h$ and $\mathcal{C}_g$? We answer this next.

*Finding a good configuration.* Observe that if $\mathcal{F}$ has a good configuration with respect to $h$ and $\mathcal{C}_g$, then any witness subfamily $\mathcal{F}'$ covers at least $k$ elements uniquely. To locate such a family of size at most $B$ we use dynamic programming over subsets of $\mathcal{C}_g$. To this end, let $W$ be a $2^k \times B$ array where we identify the rows of $W$ with subsets of $\mathcal{C}_g$ and the columns with the size of a subfamily. For a fixed coloring function $h$, a subset $\mathcal{C}'_g \subseteq \mathcal{C}_g$ and $1 \leq i \leq B$, define $W[\mathcal{C}'_g][i]$ as follows:

$$W[\mathcal{C}'_g][i] = \begin{cases} 1, & \text{if there exists } \mathcal{F}' \subseteq \mathcal{F}, \text{ with } |\mathcal{F}'| \leq i, \text{ with a} \\ & \text{good configuration with respect to } \mathcal{C}'_g \text{ and } h. \\ 0, & \text{otherwise.} \end{cases}$$

The entry corresponding to $W[\emptyset][i]$ is set to 1 for all $1 \leq i \leq B$, as a convention. We fill this array in increasing order of the sizes of subsets of $\mathcal{C}_g$. Let $\mathcal{T}$ be the family of all sets $S \in \mathcal{F}$ such that $g(S) := h(S) \cap \mathcal{C}_g \subseteq \mathcal{C}'_g$ and for each $c \in g(S)$ there exists exactly one element $e \in S$ with $h(e) = c$. Then

$$W[\mathcal{C}'_g][i] = \bigvee_{S \in \mathcal{T}} W[\mathcal{C}'_g \setminus g(S)][i-1].$$

The correctness of the algorithm is immediate. Clearly if $W[\mathcal{C}_g][B] = 1$, then a subfamily with at most $B$ sets that uniquely covers at least $k$ elements exists, and can be found out by simply storing the witness families $\mathcal{F}'$ for every entry in the table and backtracking. The time taken by the algorithm is $O(2^k Bmk)$, since the size of the array is $2^k B$ and each entry of the array can be filled in time $O(mk)$, where $m = |\mathcal{F}|$.

**Lemma 10.** *Let $(\mathcal{U}, \mathcal{F}, B, k)$ be an instance of* Budgeted Unique Coverage *with unit costs and profits and $h: \mathcal{U} \to \mathcal{C}$ a coloring function. Then we can find a subfamily $\mathcal{F}'$ of size at most $B$ which has a good configuration with respect to $h$ and $\mathcal{C}_g$, if one exists, in time $O(2^k Bmk)$.*

A randomized algorithm for Budgeted Unique Coverage with unit costs and profits is as follows.

1. Randomly choose a coloring function $h: \mathcal{U} \to \{1, \ldots, k+1\}$.
2. Apply Lemma 10 and check whether there exists a family $\mathcal{F}'$ of size at most $B$ that is witness to a good configuration with respect to $h$ and $\mathcal{C}_g$. If such a family exists, return yes, else go to Step 1.

By Lemma 9, if the given instance is a yes-instance, the probability that a solution subfamily $\mathcal{F}'$ has a good configuration with respect to a randomly chosen function $h : \mathcal{U} \to \mathcal{C}$ and $\mathcal{C}_g$ is at least $2^{-k(18 \log B \log(k+1) - \log k + \log e)}$. By Lemma 10, we can find such a subfamily in time $O(2^k Bmk)$.

**Theorem 7.** *Let $(\mathcal{U}, \mathcal{F}, B, k)$ be an instance of* Budgeted Unique Coverage *with unit costs and profits. There exists a randomized algorithm that finds a subfamily $\mathcal{F}'$ of size at most $B$ covering at least $k$ elements uniquely, if one exists, in expected $O(2^{18k \log B \log(k+1)} \cdot Bmk)$ time.*

Note that the expected running time of the randomized algorithm is better than that in Corollary 2

**Improving the Run-time.** It is clear that if a solution subfamily $\mathcal{F}'$ is to have a good configuration with respect to a randomly chosen coloring function $h$ and $\mathcal{C}_g$, then $h$ must assign all the non-uniquely covered elements of $\mathcal{F}'$ the color in $\mathcal{C}_b$. Intuitively, if we increase the number of colors in $\mathcal{C}_b$, we increase the probability that a specific target subfamily has a good configuration with respect to a randomly chosen coloring function. We formalize this intuition below.

**Lemma 11.** *Let $(\mathcal{U}, \mathcal{F}, B, k)$ be a yes-instance of* Budgeted Unique Coverage *with unit costs and profits; let $\mathcal{C}_g = [k]$, $\mathcal{C}_b = \{k+1, \ldots, q\}$ and $\mathcal{C} = [q]$ so that $q \geq 2k$. If $h : \mathcal{U} \to \mathcal{C}$ is chosen uniformly at random then every solution subfamily $\mathcal{F}'$ with $p$ elements of the universe has a good configuration with respect to $h$ and $\mathcal{C}_g$ with probability at least $e^{-k}\left(\frac{k}{q-k}\right)^k (2e)^{-\frac{kp}{q}}$.*

*Proof.* Let the set of elements uniquely covered by $\mathcal{F}'$ be $Q = \{i_1, \ldots, i_k\}$. For $\mathcal{F}'$ to have a good configuration, the function $h$ must map every element of $\mathcal{U}(\mathcal{F}') \setminus Q$ to $\mathcal{C}_b$ and map $Q$ to $\mathcal{C}_g$ injectively. Therefore the probability $\mathbf{Pr}$ that $\mathcal{F}'$ has a good configuration with respect to $\mathcal{C}_g$ and a randomly chosen $h$ is:

$$\begin{aligned}
\mathbf{Pr} &= \frac{(q-k)^{p-k}}{q^{p-k}} \times \frac{k!}{q^k} \\
&\geq \left(\frac{q-k}{q}\right)^p \cdot \left(\frac{1}{q-k}\right)^k \cdot k^k e^{-k} \\
&\geq e^{-k} \cdot \left(\frac{k}{q-k}\right)^k \cdot \left(1 - \frac{k}{q}\right)^p \\
&\geq e^{-k} \cdot \left(\frac{k}{q-k}\right)^k \cdot (2e)^{-\frac{kp}{q}} \qquad \text{(by Lemma 7).}
\end{aligned}$$

This proves the lemma. $\square$

If $(\mathcal{U}, \mathcal{F}, B, k)$ is a yes-instance of Budgeted Unique Coverage with unit costs and profits then $p \leq 18k \log B$. As we observed earlier, we have $B \geq 2$. By Lemma 11, a solution subfamily $\mathcal{F}'$ has a good configuration with respect to a randomly chosen coloring function $h$ and $\mathcal{C}_g$ with probability at least $e^{-k} \cdot (\frac{q-k}{k})^{-k} \cdot (2e)^{-kp/q}$. Setting $q = k + p$, this expression works out to be:

$$\begin{aligned}\mathbf{Pr} &\geq e^{-k} \cdot \left(\frac{k}{p}\right)^{k} \cdot (2e)^{-kp/(k+p)} \\ &\geq 2^{-k\log e - k\log\frac{p}{k} - \frac{kp}{k+p}\cdot\log 2e} \\ &= 2^{-k\log e - k\log\frac{p}{k} - \frac{k}{1+k/p}\cdot\log 2e} \\ &\geq 2^{-k\log e - k\log\frac{p}{k} - k\log 2e}.\end{aligned}$$

Now setting $p = 18k\log B$ this expression works out to $2^{-8.1k - k\log\log B}$. Combining this with Lemma 10, we obtain:

**Theorem 8.** *Let $(\mathcal{U}, \mathcal{F}, B, k)$ be an instance of* Budgeted Unique Coverage *with unit costs and profits. Then we can find a subfamily $\mathcal{F}'$ of size at most $B$ covering at least $k$ elements uniquely, if one exists, in $O(2^{9.1k + k\log\log B} \cdot Bmk)$ expected time.*

**Derandomization.** We now discuss how to derandomize the algorithms described in the last subsection. In general, randomized algorithms based on the color-coding method are derandomized using a suitable family of hash functions or "universal sets". We need a family of functions from $\mathcal{U}$ to $[t]$, where $t \geq k+1$, such that for all $S \subseteq \mathcal{U}$ of size $s = \lceil 18k\log B \rceil$ and all $X \subseteq S$ of size $k$, there exists a function $h$ in the family which maps $X$ injectively and the colors it assigns to the elements in $S \setminus X$ are different from the ones it assigns to those in $X$.

Such hash families are called $(k, s)$-hash families (with domain $[n]$ and range $[t]$) and they were introduced by Barg et al. [5] in the context of a particular class of codes called parent identifying codes. At this point, we recall the definition of an $(n, t, s)$*-perfect hash family.*

**Definition 4.** *A family $\mathcal{H}$ of functions from $[n]$ to $[t]$ is called an $(n, t, s)$**-perfect hash family** if for every subset $X \subseteq [n]$ of size $s$, there is a function $h \in \mathcal{H}$ that maps $X$ injectively.*

Note that an $(n, t, s)$-perfect hash family is a $(k, s)$-hash family with domain $[n]$ and range $[t]$ for any $k \leq s$, and a $(k, s)$-hash family with domain $[n]$ and range $[t]$ is an $(n, t, k)$-perfect hash family. Therefore $(k, s)$-hash families may be thought of as standing in between $(n, t, k)$-perfect and $(n, t, s)$-perfect hash families.

Our deterministic algorithm simply uses functions from these families $\mathcal{H}$ for coloring and is described in Figure 1. Given an instance $(\mathcal{U}, \mathcal{F}, B, k)$ of Budgeted Unique Coverage with unit costs and profits, we let $n = |\mathcal{U}|$, $\mathcal{C} = [t]$, and $s$ to be the closest integer to our estimate in Lemma 8, which is $O(k\log B)$. The correctness of the algorithm follows from the description—if a witness subfamily for the given $\mathcal{F}$ exists, at least one $h \in \mathcal{H}$ will color all the uniquely covered elements of the witness subfamily distinctly, thereby resulting in a good configuration. The running time of the algorithm is $O\left(|\mathcal{H}| \cdot \binom{t}{k} \cdot 2^k Bmk\right)$.

**Theorem 9 (Alon et al. [1]).** *Let $2 \leq k < s$. There is an explicit construction of a $(k, s)$-hash family $\mathcal{H}$ with domain $[n]$ and*

- *range $[k+1]$ of size at most $2^{ck\log s} \cdot \log_{k+1} n$, for some absolute constant $c > 0$;*
- *range $[ks]$ of size $O(k^2 s^2 \log n)$.*

**for** each $h \in \mathcal{H}$ **do**
  **for** each subset $X \subseteq \mathcal{C}$ of size $k$ **do**

1. Define $\mathcal{C}_g = X$ and $\mathcal{C}_b = \mathcal{C} \setminus X$;
2. Apply Lemma 10 and check whether there exists a subfamily $\mathcal{F}'$ of size at most $B$ which has a good configuration with respect to $\mathcal{C}_g$ and $h$;
3. **if** yes, then **return** the corresponding $\mathcal{F}'$;

**return** NO;

**Fig. 1.** Deterministic algorithm for BUDGETED UNIQUE COVERAGE.

If $t = k + 1$, then by the above theorem, the running time of our deterministic algorithm is

$$O(2^{O(k \log k + k \log \log B)} \cdot Bmk \cdot \log n);$$

when $t = ks$, the running time works out to be

$$O(2^{O(k \log k + k \log \log B)} \cdot mk^5 \cdot B \log^2 B \cdot \log n).$$

Note that for the unit costs and profits case $B \leq k$ and hence we have:

**Theorem 10.** *Let $(\mathcal{U}, \mathcal{F}, B, k)$ be an instance of the* BUDGETED UNIQUE COVERAGE *problem with unit costs and profits. Then we can find a subfamily $\mathcal{F}'$ of size at most $B$ covering at least $k$ elements uniquely, if one exists, in time $O(2^{O(k \log k)} \cdot Bmk \cdot \log n)$.*

We next give alternative running time bounds using standard $(n, t, s)$-perfect hash families for derandomizing our algorithm.

**Theorem 11** ([3, 21, 7]). *There exist explicit constructions of $(n, t, s)$-perfect hash families of size*

- $2^{O(s)} \log n$ *when $t = s$; and*
- $s^{O(1)} \log n$ *when $t = s^2$.*

*In fact, when $t = s$, one can construct an $s$-perfect hash family of size $6.4^s \log^2 n$ in time $6.4^s n \log^2 n$.*

For $t = s$, using the construction of $s$-perfect hash families by Chen et al. [7], we obtain a running time of $O(6.4^s \log^2 n \cdot \binom{s}{k} \cdot 2^k \cdot Bkm)$. Since $s = O(k \log B)$, this expression simplifies to $O(2^{O(k \log B)} \cdot \log^2 n \cdot Bmk)$. For $t = s^2$, we can use a hash family of size $s^{O(1)} \log n$ [3], and the expression for the running time then works out to be $O(2^{O(k \log k + k \log \log B)} \cdot \log n \cdot Bmk)$. Since the costs and profits are all 1, we have $B \leq k$.

**Theorem 12.** *Let $(\mathcal{U}, \mathcal{F}, B, k)$ be an instance of the* BUDGETED UNIQUE COVERAGE *problem with unit costs and profits. Then we can find a subfamily $\mathcal{F}'$ of size at most $B$ covering at least $k$ elements uniquely, if one exists, in time $O(2^{O(k \log B)} \cdot \log^2 n \cdot Bmk)$.*

In general, there is no relation between the parameters $k$ and $B$, but if $B$ is much smaller than $k$ then the run-time in Theorem 12 is better than that in Theorem 10.

We now consider existential results concerning hash families. The following is known about $(n, t, s)$-hash families.

**Theorem 13 ([16]).** *For all positive integers $n \geq t \geq s \geq 2$, there exists an $(n,t,s)$-perfect hash family $\Delta(n,t,s)$ of size $e^{s^2/t} s \ln n$.*

Alon et al. [1] provide existential bounds for $(k,s)$-hash families when the range $t = k+1$ and one such bound (see Theorem 3 in [1]) is the following. If $\mathcal{H}$ is a $(k,s)$-hash family with a domain $[n]$ and range $[k+1]$ then

$$\frac{\log_{k+1} n}{|\mathcal{H}|} = \frac{k!(s-k)^{s-k}}{s^s(s-1)\ln(k+1)} - o(1),$$

which implies that

$$|\mathcal{H}| \leq \frac{s^s(s-1)\ln(k+1)\log_{k+1} n}{k!(s-k)^{s-k}}.$$

If we assume that $s \geq 2k$, then using Lemma 7, this expression can be bounded above by $(2e)^k \cdot s^{k+1} \ln(k+1) \log_{k+1} n$. As stated in Theorem 9, Alon et al. [1] describe explicit constructions of such $(k,s)$-hash family of size $s^{ck} \cdot \log_{k+1} n$ which is still exponentially larger than the existential bound.

In the lemmas that follow, we provide existential bounds for an arbitrary range.

**Lemma 12.** *Let $k \leq s \leq n$ be positive integers and let $t \geq 2k$ be an integer. There exists a $(k,s)$-hash family $\mathcal{H}$ with domain $[n]$ and range $[t]$ of size $e^k(2e)^{sk/t} \cdot s \log n$.*

*Proof.* Let $\mathcal{A} = \{h : [n] \to [t]\}$ be the set of all functions from $[n]$ to $[t]$. For $h \in \mathcal{A}$, $S \subseteq [n]$ of size $s$ and $X \subseteq S$ of size $k$, define $h$ to be an $(X,S)$*-hash function* if $h$ maps $X$ injectively such that $h(X) \cap h(S \setminus X) = \emptyset$ and *not an* $(X,S)$*-hash function* otherwise.

Fix $S \subseteq [n]$ of size $s$ and $X \subseteq S$ of size $k$. The probability $\mathbf{Pr}$ that a function $h$ picked uniformly at random from $\mathcal{A}$ is an $(X,S)$-hash function, is given by:

$$\begin{aligned}
\mathbf{Pr} &= \frac{\binom{t}{k} k!(t-k)^{s-k}}{t^s} \\
&> \left(\frac{t}{k}\right)^k \cdot \left(\frac{k}{e}\right)^k \cdot \frac{1}{t^k} \cdot \left(\frac{t-k}{t}\right)^{s-k} \\
&= \frac{1}{e^k} \left(\frac{t-k}{t}\right)^{s-k} \\
&\geq \frac{1}{e^k} \cdot \left(\frac{1}{2e}\right)^{k(s-k)/t} \qquad \text{(By Lemma 7.)} \\
&\geq \frac{1}{e^k} \cdot \left(\frac{1}{2e}\right)^{ks/t}.
\end{aligned}$$

The probability that the function $h$ is not an $(X,S)$-hash function is less than $1 - e^{-k}(2e)^{-ks/t}$. If we pick $N$ functions uniformly at random from $\mathcal{A}$ then the probability that none of these functions is an $(X,S)$-hash function is less than $(1 - e^{-k}(2e)^{-ks/t})^N$. The probability that none of these $N$ functions is an $(X,S)$-hash function for some $(S,X)$ pair is less than

$$\binom{n}{s}\binom{s}{k}(1 - e^{-k}(2e)^{-ks/t})^N,$$

which in turn is less than $n^s(1 - e^{-k}(2e)^{-ks/t})^N$. For this family of $N$ functions to contain an $(X, S)$-hash function for every $(S, X)$ pair, we would want

$$n^s(1 - e^{-k}(2e)^{-ks/t})^N \leq 1. \tag{2}$$

Inequality 2 implies that

$$s \ln n + N \ln(1 - e^{-k}(2e)^{-ks/t}) \leq 0.$$

One can show that given any $\epsilon \in (0, 1)$ there exists $c_\epsilon > 0$ such that for all $x \in (0, \epsilon)$, we have $-c_\epsilon x \leq \ln(1 - x)$. In fact, one may choose $c_\epsilon = (|\ln(1 - \epsilon)| + 1)/\epsilon$. For $k$ and $s$ sufficiently large we have $e^{-k}(2e)^{-ks/t} < 1/4$ and we may choose $\epsilon = 1/4$. Then $c_\epsilon \approx 5.2$ and we have:

$$s \ln n - 5.2N \cdot e^{-k}(2e)^{-sk/t} \leq 0.$$

which shows that $N \geq e^k(2e)^{ks/t} \cdot s \log n$. □

**Lemma 13.** *Let $k \leq s \leq n$ be positive integers and let $t \geq k + 1$. Then there exists a $(k, s)$-hash family with domain $[n]$ and range $[t]$ of size $2^{O(k \log(s/k))} \cdot s \log n$ .*

*Proof.* Let $F = \Delta(n, m, s)$, the $(n, m, s)$-perfect hash family obtained from Theorem 13, where we set $m = \lceil s^2/(k \log(s/k)) \rceil$. Let $G$ be a family of functions $g_X$ from $[m]$ to $[t]$, indexed by $k$-element subsets $X$ of $[m]$ as follows. The function $g_X$ maps $X$ in an one-one, onto fashion to $\{1, \ldots, k\}$ and maps an element of $[m] - X$ to an arbitrary element in $\{k + 1, \ldots, t\}$. Our required family $T$ of functions from $[n]$ to $[t]$ is obtained by composing the families $F$ and $G$. It is easy to see that $T$ is an $(k, s)$-hash family and has the claimed bound for its size. □

Note that Lemma 12 requires that $t \geq 2k$ and that for Lemma 13 we have no restriction on $t$. Also note that the bound in Lemma 13 is existential as it uses the existential bound of Theorem 13. If we had an explicit construction of a $(k, s)$-hash family satisfying the bound in Lemma 12, then by setting $s = t = O(k \log B)$, we would have obtained a running time of $O(2^{O(k \log \log B)} \cdot k \log B \log n)$ which is significantly better than that given in Theorem 12. We believe that this is motivation for studying explicit constructions of $(k, s)$-hash families for an arbitrary range.

**Generalized Costs and Profits.** Recall that an instance of the Budgeted Unique Coverage problem with unit costs and profits consisted of a finite nonempty set $\mathcal{U}$, a family $\mathcal{F}$ of subsets of $\mathcal{U}$ and integers $B$ and $k$. The question is to decide whether there exists a subfamily with $B$ sets that covers at least $k$ elements uniquely. Observe that the algorithms for Budgeted Unique Coverage with unit costs and profits had two components—one component ensured that the probability that a solution subfamily has a good configuration with respect to a random coloring $h$ and $\mathcal{C}_g$ is sufficiently high (see Lemma 9), and the other had to do with finding a witness subfamily given a coloring (see Lemma 10). Note that Lemma 8 and the discussions proceeding Theorem 8 continue to hold when either the cost or the profit is integral. The derandomization procedures given in the last subsection also go through for these general cases. We now show how to modify the dynamic programming algorithm when the costs and profits are arbitrary integers.

*Dynamic programming with integral costs and profits.* Consider an instance $(\mathcal{U}, \mathcal{F}, B, k)$ of the problem with cost function $c$ and profit function $p$. Recall that the colors used are from $\mathcal{C} = \mathcal{C}_g \cup \mathcal{C}_b$ and that we are given a coloring function $h$. As before, we define an array $W$ of size $2^k \times B$ and associate the rows with subsets of $\mathcal{C}_g$ and columns with the cost of a subfamily. For a subfamily $\mathcal{H}$, let $p(\mathcal{H})$ denote the total profit of elements uniquely covered by $\mathcal{H}$. For $\mathcal{C}'_g \subseteq \mathcal{C}_g$ and $1 \leq i \leq B$, let $\mathcal{W}_i[\mathcal{C}'_g]$ denote the set of all subfamilies of budget at most $i$ which has a good configuration with respect to $h$ and $\mathcal{C}'_g$.

For $\mathcal{C}'_g \subseteq \mathcal{C}_g$ and $1 \leq i \leq B$, define $W[\mathcal{C}'_g][i]$ as follows:

$$W[\mathcal{C}'_g][i] = \begin{cases} \max_{\mathcal{H} \in \mathcal{W}_i[\mathcal{C}'_g]}\{p(\mathcal{H})\}, & \text{if } \mathcal{W}_i[\mathcal{C}'_g] \neq \emptyset; \\ 0, & \text{otherwise.} \end{cases}$$

The entry corresponding to $W[\emptyset][i]$ is set to 0 for all $1 \leq i \leq B$ as a convention. We fill this array in increasing order of sizes of the subsets of $\mathcal{C}_g$. For $S \in \mathcal{F}$, define $g(S) = h(S) \cap \mathcal{C}_g$ to be the set of good colors assigned to elements in $S$ and $p(S)$ to be the total profit of elements in $S$ assigned colors from $g(S)$. Let $\mathcal{T}$ be the subfamily containing all sets $S \in \mathcal{F}$ such that $g(S) \subseteq \mathcal{C}'_g$ and for each $c \in g(S)$ there exists exactly one element $e \in S$ with $h(e) = c$. Then

$$W[\mathcal{C}'_g][i] = \max_{S \in \mathcal{T}} \left\{p(S) + W[\mathcal{C}'_g \setminus g(S)][i - c(S)]\right\}$$

The correctness of the algorithm is immediate. Clearly, if $W[\mathcal{C}_g][B] \geq k$, then there exists a subfamily with total cost at most $B$ which uniquely covers elements with total profit at least $k$. Such a family can be found by simply storing the witness set for every entry in the table and backtracking. The time taken by the algorithm is $O(2^k Bmk + |\mathcal{C}|) = O(2^k Bmk)$, where $m = |\mathcal{F}|$.

**Lemma 14.** *Let $(\mathcal{U}, \mathcal{F}, B, k, c, p)$ be an instance of* Budgeted Unique Coverage *with integral costs and profits, and $h : \mathcal{U} \to \mathcal{C}$ be a coloring function. Then we can find a subfamily $\mathcal{H}$ of total cost at most $B$ which has a good configuration with respect to coloring $h$ and $\mathcal{C}_g$, if one exists, in time $O(2^k Bmk)$.*

Note that Lemma 14 holds even when profits are from $\mathbb{Q}^{\geq 1}$. When the costs are from $\mathbb{Q}^{\geq 1}$ and profits are positive integers then we can modify the dynamic programming algorithm as follows. Define an array $W$ of size $2^k \times k$ and identify its rows with subsets of $\mathcal{C}_g$ and columns with the profit of a subfamily. For $\mathcal{C}'_g \subseteq \mathcal{C}_g$ and $1 \leq i \leq k$, define $\mathcal{W}_i[\mathcal{C}'_g]$ and $W[\mathcal{C}'_g][i]$ as follows: $\mathcal{W}_i[\mathcal{C}'_g]$ denotes the set of all subfamilies with profit at least $i$ and which have a good configuration with respect to the given coloring function $h$ and $\mathcal{C}'_g$; and

$$W[\mathcal{C}'_g][i] = \begin{cases} \min_{\mathcal{H} \in \mathcal{W}_i[\mathcal{C}'_g]}\{c(\mathcal{H})\}, & \text{if } \mathcal{W}_i[\mathcal{C}'_g] \neq \emptyset; \\ \infty, & \text{otherwise.} \end{cases}$$

The entry $W[\emptyset][i]$ is set to 0 for all $1 \leq i \leq k$. Any entry of the form $W[\mathcal{C}'_g][i]$ where $\mathcal{C}'_g \subseteq \mathcal{C}_g$ and $i \leq 1$ is identified with the entry $W[\mathcal{C}'_g][1]$. Given $\mathcal{C}'_g \subseteq \mathcal{C}_g$, let $\mathcal{T}$ be the subfamily containing all sets $S \in \mathcal{F}$ such that $g(S) := h(S) \cap \mathcal{C}_g \subseteq \mathcal{C}'_g$ and for each $c \in g(S)$ there exists exactly one $e \in S$ with $h(e) = c$. Then

$$W[\mathcal{C}'_g][i] = \min_{S \in \mathcal{T}}\{c(S) + W[\mathcal{C}'_g \setminus g(S)][i - p(S)]\}.$$

If $W[\mathcal{C}_g][k] \leq B$ then there exists a subfamily with total cost at most $B$ that uniquely covers elements with total profit at least $k$. The time taken by the algorithm is $O(2^k \cdot mk^2)$. We thus obtain:

**Theorem 14.** *Let $(\mathcal{U}, \mathcal{F}, B, k)$ be an instance of the* Budgeted Unique Coverage *problem with either integral costs and rational profits $\geq 1$ or with rational costs $\geq 1$ and integral profits. Then one can find a subfamily $\mathcal{F}'$ of total cost at most $B$ that uniquely covers elements with total profit at least $k$, if one exists, in time*

$$O(\min\{2^{O(k \log B)}, 2^{O(k \log k + k \log \log B)}\} \cdot Bmk^2 \log^2 n).$$

## 7 Algorithms for Special Cases

We now consider deterministic algorithms for two special cases of Budgeted Unique Coverage: Unique Coverage and Budgeted Max Cut.

### 7.1 Unique Coverage

An instance $(\mathcal{U}, \mathcal{F}, k)$ of Unique Coverage can be viewed as an instance of Budgeted Unique Coverage where the costs and profits are all one and the budget $B = k$ as we do not need more than $k$ sets to cover $k$ elements uniquely. Using Theorem 12, we immediately obtain an algorithm with run-time

$$O(2^{O(k \log k)} \cdot |\mathcal{F}| \cdot k^2 \log^2 n).$$

In this subsection we present an algorithm for Unique Coverage that runs in deterministic time

$$O(2^{O(k \log \log k)} \cdot |\mathcal{F}| \cdot k + |\mathcal{F}|^2).$$

We first need some lower bounds on the number of elements that can be uniquely covered in any instance of Unique Coverage.

Define the *frequency* $f_u$ of an element $u \in \mathcal{U}$ to be the number of sets in the family $\mathcal{F}$ that contain $u$. Let $\gamma$ denote the maximum frequency, that is, $\gamma = \max_{u \in \mathcal{U}}\{f_u\}$.

**Lemma 15.** *Given an instance $(\mathcal{U}, \mathcal{F}, k)$ of* Unique Coverage*, one can in polynomial time obtain a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ such that $\mathcal{F}'$ covers at least $|\mathcal{U}|/(4e \log \gamma)$ elements uniquely, where $\gamma = \max_{u \in \mathcal{U}}\{f_u\}$.*

*Proof.* The proof of the lemma is similar to Lemma 8 but we provide a proof here for the sake of completeness.

Partition the elements of $\mathcal{U}$ into sets $C_0, C_1 \ldots C_{r-1}$, such that $u \in C_i$ if $2^i \leq f_u \leq 2^{i+1} - 1$. Note that $i$ ranges from 0 to $\log(\gamma + 1) - 1$, since the frequency of any element in $C_{r-1}$ is at most $2^r - 1$, which in turn is at most $\gamma$. So the total number of sets that the universe gets partitioned into is $\log(\gamma + 1)$.

Clearly, there exists $j$ such that $|C_j| \geq n/\log(\gamma + 1)$, where $n := |\mathcal{U}|$. Fix $j$ to be the index for which $|C_j| \geq n/\log(\gamma + 1)$. Let $u \in C_j$ and note that $2^j \leq f_u \leq 2^{j+1} - 1$. Construct a subfamily $\mathcal{F}'$ from $\mathcal{F}$ by going through each set in $\mathcal{F}$ and including it in $\mathcal{F}'$ with probability $1/2^{j+1}$. Denote by $l_u$ the probability that $u$ is covered uniquely by $\mathcal{F}'$. Then

$$\begin{aligned} l_u &= \left(\tfrac{f_u}{2^{j+1}}\right)\left(1 - \tfrac{1}{2^{j+1}}\right)^{(f_u - 1)} \\ &\geq \tfrac{1}{2}\left(1 - \tfrac{1}{2^{j+1}}\right)^{(f_u - 1)} \qquad \text{(since } f_u \geq 2^j\text{).} \\ &\geq \tfrac{1}{2}\left(1 - \tfrac{1}{2^{j+1}}\right)^{(2^{j+1} - 2)} \qquad \text{(since } f_u \leq 2^{j+1} - 1\text{).} \\ &\geq \tfrac{1}{2}\left(1 - \tfrac{1}{2^{j+1}}\right)^{(2^{j+1})} \geq \tfrac{1}{4e}. \text{ (by Lemma 7).} \end{aligned}$$

Now, let $X_u$ be an indicator random variable which takes value 1 if $u$ is covered uniquely by the subfamily $\mathcal{F}'$, and 0 otherwise. Also, define $X = \sum_u X_u$.

$$E(X) = \sum_{u \in \mathcal{U}} E(X_u) \geq \sum_{u \in C_j} E(X_u) = \sum_{u \in C_j} l_u \geq \frac{n}{4e \log \gamma}.$$

This immediately implies the existence of a subfamily that covers $n/4e \log \gamma$ elements (or more) uniquely. We can derandomize the above algorithm using the method of conditional probabilities to obtain the desired subfamily. ◻

**Lemma 16.** *Given an instance $(\mathcal{U}, \mathcal{F}, k)$ of* Unique Coverage*, one can in polynomial time obtain a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ such that $\mathcal{F}'$ covers at least $|\mathcal{U}|/(8e \log M)$ elements uniquely, where $M = \max_{S \in \mathcal{F}}\{|S|\}$.*

*Proof.* We begin by constructing a subfamily $\mathcal{F}'$ from $\mathcal{F}$ that is *minimal* in the sense that every set in $\mathcal{F}'$ covers at least one element in $\mathcal{U}$ uniquely. Such a subfamily is easily obtained, by going over every set in the family and checking if it has at least one element which is not contained in any other set. Let $m'$ denote the size of the subfamily $\mathcal{F}'$. Note that $|\bigcup_{S \in \mathcal{F}'} S| = n := |\mathcal{U}|$. For the proof of the lemma we distinguish two cases based on $m'$:

*Case 1:* ($m' \geq n/2$) As the subfamily is minimal, by construction, we are immediately able to cover at least $n/2$ elements uniquely. Thus $\mathcal{F}'$ itself satisfies the claim of the lemma.

*Case 2:* ($m' < n/2$) In this case, we first claim that $|\{u \in \mathcal{F}' : f_u < M\}| \geq n/2$. If not, then there would be more than $n/2$ elements whose frequency is at least $M$, which implies that $\sum_{S \in \mathcal{F}'} |S| > Mn/2$. On the other hand, $\sum_{S \in \mathcal{F}'} |S|$ is clearly at most $M(n/2 - 1)$ (because there are strictly less than $n/2$ sets in the family and the size of any set in the family is bounded by $M$). The claim implies that there exists a set of at least $n/2$ elements whose frequency is less than $M$. Denote this set of elements by $\mathcal{V}$. Consider the family $\mathcal{F}''$ obtained from $\mathcal{F}'$ as follows: $\mathcal{F}'' = \{S \cap \mathcal{V} \mid S \in \mathcal{F}'\}$. Applying Lemma 15 to the instance $(\mathcal{V}, \mathcal{F}'')$, we obtain a subfamily $\mathcal{T}$ of $\mathcal{F}''$ that covers at least $n/(8e \log M)$ elements uniquely. The corresponding subfamily of $\mathcal{F}'$ will clearly cover the same set of elements uniquely in $\mathcal{U}$. This completes the proof of the lemma. ◻

Using these lower bounds on the number of elements that are uniquely covered, we can upper bound the size of a yes-instance of the Unique Coverage problem as a function of the parameter $k$. Let $(\mathcal{U}, \mathcal{F}, k)$ be an instance of Unique Coverage reduced with respect to Rules R1 and R2 described in Section 5. Then $|S| \leq k - 1$ for all $S \in \mathcal{F}$ and $M$ is bounded above by $k - 1$. If $k \leq n/8e \log(k-1)$, then there exists a subfamily that covers $k$ elements uniquely. If not, we have $k > n/8e \log k$, which implies that $n < 8ek \log k$.

**Lemma 17.** *Let $(\mathcal{U}, \mathcal{F}, k)$ be an instance of* Unique Coverage*. Then, in polynomial time, we can either find a subfamily covering at least $k$ elements uniquely, or an equivalent instance where the size of the universe is $O(k \log k)$.*

Note that the above lemma shows that Unique Coverage admits a kernel with $k^{O(k)}$ sets which is what was shown in Corollary 2.

An improved algorithm for Unique Coverage first applies Lemma 17 and obtains an instance of Unique Coverage, $(\mathcal{U}, \mathcal{F}, k)$, where $n = |\mathcal{U}| \leq O(k \log k)$. Now we examine all $k$-sized subsets $X$

of the universe $\mathcal{U}$ and check whether there exists a subfamily that covers it uniquely. Let $X = \{u_{i_1}, u_{i_2}, \ldots, u_{i_k}\}$, and let $h$ be a function that maps $X$ injectively to $[k]$ and each element in $\mathcal{U} \setminus X$ to the color $k+1$. Applying Lemma 10 to the instance $(\mathcal{U}, \mathcal{F}, B = k, k)$, with the coloring function $h$ described above gives us an algorithm to find the desired $\mathcal{F}'$ in time $O(2^k k^2 m)$. Note that a factor of $k$ can be avoided by directly applying dynamic programming over subsets of $X$. The size of $\mathcal{U}$ is upper bounded by $8ek \log k$ and hence the total number of subsets that need to be examined is at most $\binom{8ek \log k}{k}$, which is bounded above by $(8e \log k)^k \leq 2^{4.5k + k \log \log k}$. Combining this with the above discussion results in:

**Theorem 15.** *Let $(\mathcal{U}, \mathcal{F}, k)$ be an instance of* Unique Coverage. *Then one can decide whether there exists a subfamily covering at least k elements of $\mathcal{U}$ uniquely in time $O(2^{5.5k+k \log \log k} \cdot mk + m^2)$.*

### 7.2 Budgeted Max Cut

An instance of Budgeted Max Cut consists of an undirected graph $G = (V, E)$ on $n$ vertices and $m$ edges; a cost function $c : V \to \mathbb{Z}^+$; a profit function $p : E \to \mathbb{Z}^+$; and positive integers $k$ and $B$. The question is whether there exists a cut $(T, V - T)$, $\emptyset \neq T \neq V$, such that the total cost of the vertices in $T$ is at most $B$ and the total profit of the edges crossing the cut is at least $k$. This problem can be modelled as an instance of Budgeted Unique Coverage by taking $\mathcal{U} = E$ and $\mathcal{F} = \{S_v : v \in V\}$, where $S_v = \{e \in E : e$ is incident on $v\}$.

The algorithm we describe has running time $O(2^{O(k)} \cdot Bmk \cdot \log^2 n)$. Given $S \subseteq V$, we let $c(S)$ denote the total cost of the elements of $S$. If $(S, V - S)$ is a cut in a graph $G$, then $p(S, V - S)$ is the total profit of edges across the cut. Define the profit $\hat{p}(v)$ of a vertex $v$ to be the sum of the profits of all the edges incident on it. We assume that $\hat{p}(v) \leq k - 1$ for all $v \in V$, for otherwise the given instance is a yes-instance trivially.

**Lemma 18.** *If $(G, B, k, c, p)$ is a yes-instance of* Budgeted Max Cut *then there exists a cut $(S, S - V)$ such that $c(S) \leq B$, $p(S, V - S) \geq k$, and $|\bigcup_{v \in S} S_v| \leq 4k$.*

*Proof.* Since we are given a yes-instance of the problem, there exists a cut $(T, T')$ such that $c(T) \leq B$ and $p(T, T') \geq k$. Call a vertex $v$ of $T$ *redundant* if $p(T - v, T' \cup v) \geq k$. Starting with the cut $(T, T')$, transfer redundant vertices from $T$ to $T'$ and obtain a cut $(S, S')$ such that $S \subset T$ and $S$ does not contain any redundant vertices. Observe that $c(S) \leq B$ and $p(S, S') \geq k$. For any $v \in S$, $\hat{p}(v) \leq k-1$ and $p(S - v, S' \cup v) \leq k - 1$. Therefore $p(S, S') \leq 2k$. For $v \in S$, partition $S_v$ as $I_v \uplus C_v$, where $I_v$ is the set of edges incident on $v$ that lie entirely in $S$ and $C_v$ are the edges that lie across the cut $(S, S')$. Clearly $p(I_v) \leq p(C_v)$, for otherwise, $p(S - v, S' \cup v) > p(S, S')$, a contradiction to the fact that $S$ has no redundant vertices. Therefore $\sum_{v \in S} p(I_v) \leq \sum_{v \in S} p(C_v) \leq 2k$. This yields $\sum_{v \in S} \hat{p}(v) \leq 4k$. Since the profits are at least one, we have $|\bigcup_{v \in S} S_v| \leq 4k$. □

We now use the deterministic algorithm outlined before Theorem 12 with $t = s = 4k$ and a $4k$-perfect hash family by Chen et al. [7]. The run-time then works out to $O(6.4^{4k} \log^2 n \cdot \binom{4k}{k} \cdot 2^k Bmk)$ which simplifies to $O(2^{13.8k} \cdot Bmk \cdot \log^2 n)$.

**Theorem 16.** *Let $(G, B, k, c, p)$ be an instance of* Budgeted Max Cut. *Then we can find a cut $(S, S')$ such that $c(S) \leq B$ and $p(S, S') \geq k$, if one exists, in time $O(2^{13.8k} \cdot Bmk \cdot \log^2 n)$.*

| Unique Coverage (*Parameter: k*) | *Kernel Size* | *Sect.* |
|---|---|---|
| Each element occurs in at most $b$ sets | $(k-1)b$ | 5 |
| Intersection size bounded by $c$ | $k^{c+1}$ | 5.1 |
| General case | $4^k$ | 5.2 |
| Each set of size at most $b$ | $2^{b+k}$ | 5.2 |
| Budgeted Unique Coverage | *Complexity* | *Sect.* |
| Arbitrary weights (parameters $B$ and $k$) | Not FPT (unless P = NP) | 6.1 |
| Integer weights (parameter $B$) | W[1]-hard | 6.1 |
| Integer weights (parameters $B$ and $k$) | FPT | 6.2 |

**Table 1.** Main results in this paper.

## 8 Conclusions

In this paper we studied the parameterized complexity of the Unique Coverage problem and its weighted counterpart Budgeted Unique Coverage. We considered several plausible parameterizations of these problems and showed that except for the standard parameterized version, all other versions are unlikely to be fixed-parameter intractable. We also described problem kernels for several special cases of the standard parameterized version of Unique Coverage and showed that the general version admits a kernel with at most $4^k$ sets. Since Dom et al. [9] have shown that it is impossible to obtain polynomial kernels (modulo some complexity-theoretic assumptions), it is an interesting open problem to obtain sub-exponential kernels for Unique Coverage.

We gave fixed-parameter tractable algorithms for Budgeted Unique Coverage and several of its variants. Our algorithms were based on an application of the well-known method of color-coding in an interesting way. Our randomized algorithms have good running times but the deterministic algorithms make use of either perfect hash families or $(k, s)$-hash families and this introduces large constants in the running times, a common enough phenomenon when derandomizing randomized algorithms using such function families [3]. The main results of this paper are summarized in Table 1.

Our use of $(k, s)$-hash families to derandomize algorithms is perhaps the first application of these hash families outside the domain of coding theory and it suggests that this not-so-well-known class of hash families may be as important as the standard $(n, t, s)$-perfect hash families. It will be interesting to give explicit constructions of $(k, s)$-hash families of size promised by Lemma 12 and explore other applications of our generalization of the color-coding technique.

### Acknowledgments

We thank Srikanth Srinivasan for useful discussions and an anonymous reviewer for comments that helped to improve the presentation.

## References

1. N. Alon, G. Cohen, M. Krivelevich, and S. Litsyn. Generalized hashing and parent-identifying codes. *Journal of Combinatorial Theory Series A*, 104(1):207–215, 2003.

2. N. Alon and J. H. Spencer. *The Probabilistic Method.* Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., 2000.
3. N. Alon, R. Yuster, and U. Zwick. Color-coding. *Journal of the ACM*, 42(4):844–856, 1995.
4. R. Bar-Yehuda, O. Goldreich, and A. Itai. On the time-complexity of broadcast in multi-hop radio networks: An exponential gap between determinism and randomization. *Journal of Computer and System Sciences*, 45:104–126, 1992.
5. A. Barg, G. Cohen, S. Encheva, G. Kabatiansky, and G. Zémor. A hypergraph approach to the identifying parent property: The case of multiple parents. *SIAM Journal on Discrete Mathematics*, 14(3):423–431, 2001.
6. M. Cesati. Perfect Code is W[1]-complete. *Information Processing Letters*, 81(3):163–168, 2002.
7. J. Chen, J. Kneis, S. Lu, D. Mölle, S. Richter, P. Rossmanith, S.-H. Sze, and F. Zhang. Randomized divide-and-conquer: Improved path, matching, and packing algorithms. *SIAM Journal on Computing*, 38(6):2526–2547, 2009.
8. E. D. Demaine, U. Feige, M. Hajiaghayi, and M. R. Salavatipour. Combination can be hard: Approximability of the unique coverage problem. *SIAM Journal on Computing*, 38(4):1464–1483, 2008.
9. M. Dom, D. Lokshtanov, and S. Saurabh. Incompressibility through Colors and IDs. In *Proceedings of 36th International Colloquium of Automata, Languages and Programming (ICALP 2009)*, volume 5555 of *LNCS*. Springer, 2009.
10. R. G. Downey and M. R. Fellows. *Parameterized Complexity.* Springer-Verlag, New York, 1999.
11. T. Erlebach and E. J. van Leeuwen. Approximating geometric coverage problems. In *Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2008)*, pages 1267–1276, 2008.
12. J. Flum and M. Grohe. *Parameterized Complexity Theory.* Texts in Theoretical Computer Science. An EATCS Series. Springer-Verlag, Berlin, 2006.
13. M. R. Garey and D. S. Johnson. *Computers and Intractability.* Freeman, San Francisco, 1981.
14. S. Jukna. *Extremal Combinatorics.* Springer-Verlag, Berlin, 2001.
15. S. Khuller, A. Moss, and J. Naor. The Budgeted Maximum Coverage problem. *Information Processing Letters*, 70(1):39–45, 1999.
16. K. Mehlhorn. On the program size of perfect and universal hash functions. In *Proceedings of the 23th IEEE Symposium on Foundations of Computer Science (FOCS 1982)*, pages 170–175. IEEE, 1982.
17. N. Misra, V. Raman, S. Saurabh, and S. Sikdar. Budgeted unique coverage and color-coding. In *Proceedings of the 4th Computer Science Symposium in Russia (CSR 2009)*, LNCS. Springer, 2009. To appear.
18. H. Moser, V. Raman, and S. Sikdar. The parameterized complexity of the Unique Coverage problem. In *Proceedings of the 17th International Symposium on Algorithms and Computation (ISAAC 2007)*, volume 4835, pages 621–631, 2007.
19. R. Motwani and P. Raghavan. *Randomized Algorithms.* Cambridge University Press, 1995.
20. R. Niedermeier. *Invitation to Fixed-Parameter Algorithms*, volume 31 of *Oxford Lecture Series in Mathematics and its Applications.* Oxford University Press, Oxford, 2006.
21. J. Schmidt and A. Siegel. The spatial complexity of oblivious $k$-probe hash functions. *SIAM Journal on Computing*, 19(5):775–786, 1990.