Maximizing Covered Area in the Euclidean Plane with Connectivity Constraint

Chien-Chung Huang
DI ENS, École Normale supérieure, Université PSL, Paris, France
CNRS, Paris, France
chien-chung.huang@ens.fr

Mathieu Mari
DI ENS, École Normale supérieure, Université PSL, Paris, France
mathieu.mari@ens.fr

Claire Mathieu
CNRS, Paris, France
clairemmathieu@gmail.com

Joseph S. B. Mitchell
Stony Brook University, Stony Brook, NY 11794, USA
joseph.mitchell@stonybrook.edu

Nabil H. Mustafa
Université Paris-Est, Laboratoire d'Informatique Gaspard-Monge, ESIEE Paris, France
mustafan@esiee.fr

Abstract

Given a set $\mathcal{D}$ of n unit disks in the plane and an integer $k \leq n$, the maximum area connected subset problem asks for a set $\mathcal{D}' \subseteq \mathcal{D}$ of size k maximizing the area of the union of disks in $\mathcal{D}'$, under the constraint that this union is connected. This problem is motivated by wireless router deployment and is a special case of maximizing a submodular function under a connectivity constraint.

We prove that the problem is NP-hard and analyze a greedy algorithm, proving that it is a $\frac{1}{2}$-approximation. We then give a polynomial-time approximation scheme (PTAS) for this problem with resource augmentation, i.e., allowing an additional set of εk unit disks that are not drawn from the input. Additionally, for two special cases of the problem we design a PTAS without resource augmentation.

2012 ACM Subject Classification Theory of computation $\rightarrow$ Design and analysis of algorithms

Keywords and phrases approximation algorithm, submodular function optimisation, unit disk graph, connectivity constraint

Category APPROX

Funding *Joseph S. B. Mitchell*: Partially supported by the National Science Foundation (CCF-1540890), the US-Israel Binational Science Foundation (CCF-1540890), and DARPA (Lagrange).
Nabil H. Mustafa: Supported by the grant ANR SAGA (JCJC-14-CE25-0016-01)

1 Introduction

Maximizing a submodular function[1] under constraints is a classical problem in computer science and operations research [16, 35]. The most commonly studied constraints are cardinality, knapsack and matroids constraints. A natural constraint that has received little attention is the *connectivity* constraint. In this paper, we study the following problem. Given a set $\mathcal{D}$ of n unit disks in the plane and an integer $k \leq n$, compute a set $\mathcal{D}' \subseteq \mathcal{D}$ of size k that maximizes the area of the union of disks in $\mathcal{D}'$, under the constraint that this union is connected. We call this problem *Maximum Area Connected Subset* problem (MACS). Notice that the area covered by the union of a set of disks is a monotone submodular function.

The problem is motivated by wireless router deployment, first introduced in [24]: we need to install a certain number of routers to maximize the number of clients covered while also ensuring that these routers are connected to each other. When the clients are spatially uniformly spread, the number of clients covered is proportional to the area and hence the objective is to maximize the area covered.

Our Contributions

We first analyze a variant of the greedy algorithm and prove that it computes a $\frac{1}{2}$-approximation (Theorem 4); further we show that the analysis of the algorithm is tight. On the other hand, we show that the naive greedy algorithm that adds disks one at a time to maximize the area of the union computes, in the worst-case, a solution that is a $\Omega(k)$-factor smaller than the optimal one.

To improve upon the $\frac{1}{2}$-approximation ratio, we turn to the resource augmentation setting in which the algorithm is allowed to add a few additional disks that are not drawn from the input. We design a PTAS for the resource augmentation version of the problem (Theorem 5) using Arora's shifted dissection technique [1]. The correctness proof hinges on a structural statement (Lemma 21) which shows the existence of a near-optimal solution with $O(\varepsilon k)$ additional disks, and with additional structure that allow it to be computed efficiently by dynamic programming.

As a corollary, for two special cases of the MACS we design a PTAS without resource augmentation: i) when the Euclidean distances are well-approximated by shortest paths in the intersection graph (Corollary 7) and ii) every point of the relevant region of the Euclidean plane is covered by at least one input disk (Corollary 10).

On the negative side, via a reduction from the Rectilinear Steiner Tree problem, we show that MACS is NP-hard (Theorem 3). We leave open the question of whether MACS is APX-hard or admits a PTAS without resource augmentation. Nonetheless, we show that if the goal is to compute MACS for a set of arbitrary quadrilaterals instead of disks, the problem is APX-hard (Theorem 12).

Related work.

Maximising a monotone submodular function under constraint(s) is a subject that has received a large amount of attention over the years. We refer the reader to [3, 7, 13, 16, 23, 25, 35] and the references therein. Our problem can be regarded as maximising a submodular function under a cardinality (knapsack) constraint and a connectivity constraint. Notice that

[1] Given a set X, a function $f : 2^X \to \mathbb{R}$ is *submodular* if given any two subsets $A, B \subseteq X$, $f(A) + f(B) \geq f(A \cap B) + f(A \cup B)$.

the connectivity constraint is central to the difficulty of our problem: without connectivity constraints, MACS admits a PTAS even on the more general case of convex pseudodisks [6]. We also give a short proof of this result for unit-disks in Appendix A; even without connectivity, the problem is still NP-hard[2].

Another motivation for studying the connectivity constraint is related to cancer genome studies. Suppose that a vertex represents an individual protein (and associated gene), an edge represents pairwise interactions, and each vertex has an associated set. Finding the connected subgraph of k genes that is mutated in the largest number of samples is equivalent to the problem of finding the connected subgraph with k nodes that maximizes the cardinality of the union of of the associated sets, see [34].

In the general (non-geometric) setting where a general monotone submodular function is given, a $O(\frac{1}{\sqrt{k}})$-approximation algorithm is given in [24]. Our results show that when the submodular function and the connectivity are induced by a geometric configuration, the approximation ratio can be significantly improved.

We next discuss several related problems where the connectivity constraint is involved. An example is the node-cost budget problem introduced in [31], where the goal is to find a connected set of vertices in a general graph to collect the maximum profit on the vertices while guaranteeing the total cost does not exceed a certain budget. Notice that in this setting the submodular function is a simple additive function of the profits. Another related problem [4] is to assign radii to a given set of points in the plane so that the resulting set of disks is connected, the objective being to minimize the sum of radii.

Khuller et al. [22] study the budgeted connected dominating set problem, where given a general undirected graph, there is a budget k on the number of vertices that can be selected, and the goal is to induce a connected subgraph that dominates as many vertices as possible. It was pointed out to us that via a reduction, their algorithm gives a $\frac{1}{13}\left(1-\frac{1}{e}\right)$-approximate solution for MACS. The authors of [19] consider the problem of selecting k nodes of an input node-weighted graph to form a connected subgraph, with the aim of maximizing or minimizing the selected weight.

We now turn to the geometric setting. A logarithmic-factor approximation algorithm is known [17] for the connected sensor coverage problem, in which one must select a small number (at most k) of sensors in the plane forming a connected communication network and covering the desired region. Here the region covered by each sensor is not necessarily a disk but may be a convex region of the plane (see [14, 21]). Our resource augmentation PTAS relies on ideas used for Euclidean TSP and other geometric problems [1, 28]. A $(1-\varepsilon)$-approximation algorithm in time $n^{O(1/\varepsilon)}$ for the maximum independent set problem on unit disk graphs is known [27]. The authors of [26] present a constant-factor approximation algorithm for several problems on Unit Disk Graphs, including maximum independent set. The maximum independent set problem is NP-hard even for unit disk graphs in the Euclidean plane [8]. When the goal is to cover a specified set of clients (instead of the maximum area) with the minimum number of disks (instead of constraining the number of disks to at most k), and there is no connectivity constraint, the problem is NP-hard [8] but there exists a polynomial-time approximation scheme [20].

[2] The reduction is from Maximum independent set problem that is NP-hard in unit-disk-graphs

2 Formal definitions and our results

The Euclidean distance between two points x and y is denoted by $\|x-y\|$. When there is no confusion, we will refer to a point x in the plane and the unit disk centered at x interchangeably.

▶ **Definition 1.** *Given a finite set S in the plane, the* unit disk intersection graph $\mathrm{UDG}(S)$ *is a graph on S where there is an edge between $x, y \in S$ if and only if $\|x-y\| \leq 2$.*

A set S of points in the plane are *connected* if $\mathrm{UDG}(S)$ is a connected graph.

▶ **Definition 2.** *The Maximum Area Connected Subset (*MACS*) problem is as follows.*
***Input**: a finite set of points $X \subseteq \mathbb{R}^2$ and a non-negative integer k, where $k \leq |X|$.*
***Output**: a subset $S \subseteq X$ of size at most k, such that the unit-disk graph* $\mathrm{UDG}(S)$ *of S is connected.*
***Goal**: maximize the area of the union of the unit disks centered at points of S.*
The optimal solution of MACS *on input (X, k) is denoted by $\boldsymbol{OPT}(X, k)$.*

When the context is clear, we refer to $\mathbf{OPT}\,(X,k)$ as $\mathbf{OPT}$, which is also used to denote the area covered by the optimal solution (observe that $\mathbf{OPT}$ is trivially upper-bounded by πk). Any $S \subseteq X$ with $|S| \leqslant k$ for which $\mathrm{UDG}(S)$ is connected is called a *feasible solution*.

2.1 Results

We state our main results below.

▶ **Theorem 3** (Hardness). MACS *is NP-hard.*

▶ **Theorem 4** (Approximation). MACS *has a* $(1/2)$*-approximation that can be computed in polynomial time (Algorithm 1).*

With resource augmentation, we obtain a $(1-\varepsilon)$-approximation.

▶ **Theorem 5** (Resource augmentation). *Let $\varepsilon > 0$ be fixed. Given a set of points $X \subseteq \mathbb{R}^2$ and a non-negative integer k, there is an algorithm (Algorithm 2) that computes, in time $n^{\mathcal{O}(\varepsilon^{-3})}$, a subset $S \subseteq X$ of size at most k and a set $S_{add} \subseteq \mathbb{R}^2$ of at most εk points, such that* $\mathrm{UDG}\,(S \cup S_{add})$ *is connected, and the area covered by the unit disks centered at S is at least $(1-\varepsilon)\boldsymbol{OPT}(X,k)$.*

In Section B we provide an alternative PTAS, based on the m-guillotine method, which is deterministic, with a somewhat better running time.

Let $d_G(x, y)$ denote the distance of two vertices x and y in graph G. A set X of points in the plane is called α*-well-distributed* if $\mathrm{UDG}(X)$ is an α-spanner [30] for X, namely

▶ **Definition 6.** *Given $\alpha > 0$, a finite set X of points in the plane is called* α-well-distributed *if for all $x, y \in X$, $d_{UDG(X)}\,(x,y) \leqslant \lceil \alpha \cdot \|x-y\| \rceil$.*

▶ **Corollary 7.** MACS *on α-well-distributed inputs (for a fixed α) allows a polynomial-time approximation scheme (Algorithm 3).*

▶ **Definition 8.** *A set X is called* pseudo-convex *if the convex-hull of X is covered by the union of the unit disks centered at points of X.*

▶ **Lemma 9.** *A pseudo-convex set X is* 3.82*-well-distributed.*

▶ **Corollary 10.** MACS *on pseudo-convex inputs admits a polynomial-time approximation scheme.*

In contrast, a similar problem stated with quadrilaterals instead of disks is hard to approximate.

▶ **Definition 11.** *The* QUAD-CONNECTED-COVER *is defined as follows.*
Input*: a set $\mathcal{T}$ of n convex quadrilaterals in the plane, and an integer k.*
Output*: a subset T of $\mathcal{T}$ of size k such that the intersection graph*[3] *of T is connected.*
Goal*: Maximise the area covered by the union of quadrilaterals in T.*

▶ **Theorem 12.** QUAD-CONNECTED-COVER *is APX-hard.*

3 NP-Hardness of MACS

We present a reduction from the RECTILINEAR STEINER TREE (RST) problem, which is NP-hard, to prove that MACS is NP-hard.

> RECTILINEAR STEINER TREE PROBLEM: Given n terminals on a Euclidean plane and a number L, decide whether there exists a tree to connect all the n terminals using horizontal and vertical lines of total length at most L.

The problem is NP-complete [15], even if all terminals have integral coordinates bounded by $V = poly(n)$. In the following, we assume that n is sufficiently large, say $n \geq 1000$.

We next explain the reduction. We start from an instance of RST and define an instance of MACS as follows. For all $0 \leq i, j \leq V$, we define one *cardinal* disk with center at $(n \cdot i, nj)$; $n-4$ *path* disks centered at $\left(in + 2 + (1 + \frac{1}{n-5}) \cdot t, jn\right)$ where $t \in \{0, \dots, n-5\}$ and $n-4$ *path* disks centered at $\left(in, jn + 2 + t(1 + \frac{1}{n-5})\right)$ where $t \in \{0, \dots, n-5\}$. For each terminal (i, j) in the RST instance, we place $n^2/10$ *bonus* disks : the first one centered at $\left(in + \sqrt{2}, jn + \sqrt{2}\right)$ and the remaining centers forming a connected group in $[in+2, (i+1)n-2] \times [jn+2, (j+1)n-2]$ in such a way that each bonus disk is tangent other bonus disks, and can be connected to the first bonus disk. Notice that except the first one, no bonus disk intersects path disks. This defines the set of disks. We set $k = 1 + L(n-3) + n^3/10$. This defines the MACS instance.

See Figure 1 for an illustration. Note that the interior of a cardinal disk is disjoint from all other disks of the instance.

Notice that as the RST instance has all terminals bounded by a rectangular of polynomial size, the above reduction can be done in polynomial time.

Let Z denote the set of the cardinal disk at $(0, 0)$ and the $n-4$ path disks at $\left(2 + t(1 + \frac{1}{n-5}), 0\right)$ where $t \in \{0, \dots, n-5\}$ and let $\mathcal{A}(Z)$ denote the area covered by Z.

▶ **Theorem 13.** *The original* RST *instance has a feasible solution of total length at most L if and only if the derived* MACS *instance has a feasible solution of area of at least $\pi + L \cdot \mathcal{A}(Z) + (\frac{n^3}{10} - \frac{n}{3})\pi$.*

[3] vertex set is T and two quadrilaterals are adjacent if and only if they intersect.

Proof. ($\Rightarrow$) direction is easy to see: We call a set of disks a *segment* if it consists of a cardinal disk and all the $n-4$ path disks between it and one of its four adjacent cardinal disks. Thus the area covered by a segment is exactly $\mathcal{A}(Z)$. Consider a feasible solution for the RST instance, of length exactly L without loss of generality. We root it at an arbitrary integral point, direct it outwards from the root, and view it as a collection of horizontal or vertical directed edges of unit length. In the MACS instance, we take all bonus disks, the cardinal disk associated to the root of the RST solution, and, for each directed edge of the RST solution, all disks of the corresponding segment. The total number of disks is exactly k, and the area covered is at least $\pi + L \cdot \mathcal{A}(Z) + \frac{n^3\pi}{10} - 2n\gamma$, where γ is the size of the overlapped area of the first bonus disk associated with a terminal and the path disk just tangent to the corresponding cardinal disk of the latter. (Recall the first bonus disk can overlap up to two path disks). The distance between two such centers is $h = \sqrt{8-4\sqrt{2}}$. Furthermore, the overlapped area can be expressed as

$$2\arccos\frac{h}{2} - 2\left(\frac{h}{2}\right)\sqrt{1-\left(\frac{h}{2}\right)^2}$$

which is upper-bounded by 0.45. Therefore $2n\gamma \le 0.9n \le n\pi/3$. This gives the proof of one direction.

For ($\Leftarrow$) direction, assume that a solution S for the MACS instance is given with area at least $\pi + L \cdot \mathcal{A}(Z) + (\frac{n^3}{10} - \frac{n}{3})\pi$. By our construction, we can modify S, while conserving its connectivity and without diminishing covered area, so that the following properties hold.

(i) If any bonus disk corresponding to a terminal is part of S, so is the cardinal disk corresponding to this terminal.

(ii) The path and cardinal disks in S form a tree; furthermore, such a tree consists of a cardinal disk, a set of segments, and at most one sub-segment. (A *sub-segment* is a subset of a segment, so that it induces a connected component.)

We claim that the number B of bonus disks in S is at least $\frac{n^3}{10} - \frac{9n}{10}$. Suppose not. Observe that the covered area of S can be upper-bounded as

$$B\pi + \pi + \frac{L|Z| + \frac{n^3}{10} - B}{|Z|}\mathcal{A}(Z) \tag{1}$$

(Here we ignore the possible intersection of a bonus disk with the path disks. The first term is the area covered by bonus disks; the second term is the area covered by a cardinal disk; the third term is the maximum area that can be covered by segments, and possibly the single sub-segment in S). Now since S is supposed to be a feasible solution in the MACS instance, its covered area should be at least

$$\pi + L \cdot \mathcal{A}(Z) + \left(\frac{n^3}{10} - \frac{n}{3}\right)\pi \tag{2}$$

However,

$$(2) - (1) \quad = \quad \left(\frac{n^3}{10} - B\right)\left(\pi - \frac{A(Z)}{|Z|}\right) - \frac{n\pi}{3} \ge \frac{9n}{10}\left(\pi - \frac{A(Z)}{|Z|}\right) - \frac{n\pi}{3}.$$

Here in order to reach a contradiction (making the last term greater than 0), we need to calculate $\mathcal{A}(Z)$, which is $(n-3)\pi - (n-5)\gamma'$, where γ' is size of the overlapped area of two disks whose centers have distance $1 + \frac{1}{n-5}$. γ' is easily shown to be at most 1.25. Therefore, $0.9n(\pi - \frac{A(Z)}{|Z|}) - \frac{n\pi}{3} \geq 0.9n(\pi - \frac{(n-3)\pi-(n-5)1.25}{n-3}) - \frac{n\pi}{3} \geq 0$, which can be verified when $n \geq 1000$.

So we know that S has at least $\frac{n^3}{10} - \delta$ bonus disks, where $\delta \leq n/10$. Ignoring the possible sub-segment of S, S includes a cardinal disk, L' segments and $\frac{n^3}{10} - \delta$ bonus disks. As a result,

$$k = 1 + L|Z| + \frac{n^3}{10} \geq 1 + L'|Z| + \frac{n^3}{10} - \delta \geq 1 + L'|Z| + \frac{n^3}{10} - \frac{n}{10},$$

implying that $n/10 \geq (L'-L)|Z| = (L'-L)(n-3)$. Thus $L' = L$ and the cardinal disks and path disks of S correspond to a tree of length L in the RST instance. The proof follows. ◀

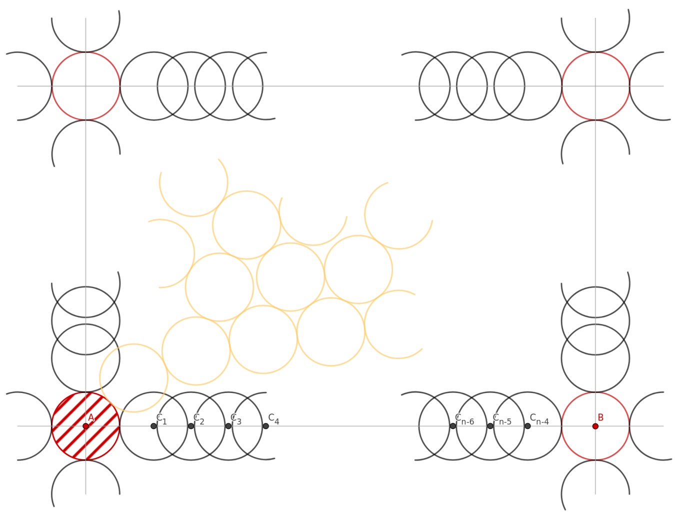


Figure 1 Black, red and orange disks respectively represent *path*,*cardinal* and *bonus* disks. The hatched disk is associated to a terminal node.

4 Proof of Theorem 4: the Two-by-two algorithm

In the section we present a simple (1/2)-approximation for MACS based on a greedy approach, by iteratively adding two unit disks that maximize the additional area covered while maintaining feasibility. Interestingly, the algorithm that adds disks one at a time is not a constant approximation algorithm. See Figure 2 for an example. Moreover, trying all possible sets of s disks, for any $s \geq 3$, in the neighborhood of the current solution does not improve the approximation ratio. This can be seen on Figure 3 where the first disk chosen by the algorithm is not x, but x_s.

Let B_x denote the unit disk centered at $x \in \mathbb{R}^2$ and $B(S) = \bigcup_{x \in S} B_x$ denote the set of points at distance at most one from at least one point in a finite set $S \subset \mathbb{R}^2$ of points. The area covered by a set $C \subset \mathbb{R}^2$ is denoted by $\mathcal{A}(C)$. When $C = B(S)$, its area is simply written as $\mathcal{A}(S)$. Given a graph G, $G[S]$ denotes the subgraph induced by a subset S of vertices. A subset of the vertices of a graph is a *dominating set* if every vertex belongs to

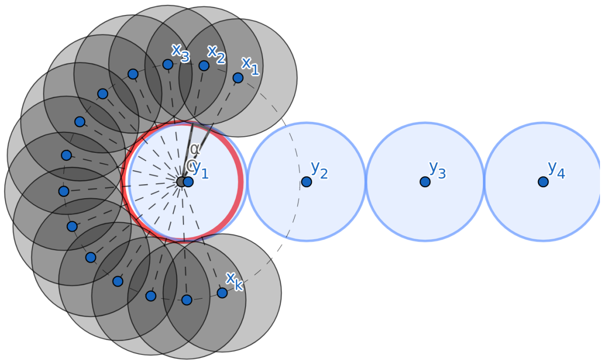


Figure 2 The greedy algorithm that adds only one connected disk maximising the marginal area covered is not a constant factor algorithm. For any $k \geqslant 0$ and $\varepsilon > 0$, consider the above input where $O = (0, 0)$, and $y_i = (2(i-1)+\varepsilon, 0)$ for all i. Then, put all $x_1, \dots, x_k$ evenly spaced (by an angle α) on a circle of radius 2 around O so that none of them intersect y_2. Each light grey regions are covered by only one disk x_i so the marginal gain of adding x_i to any solution is at least the area of one of these regions, say $a > 0$. If ε is chosen such that $\mathcal{A}(B_{y_1} \setminus B_O) < a$, then if the algorithm starts by picking disk O, it will then choose all x_j, so that the area covered by the solution is upper-bounded by the area of a radius 3 disk, 9π, while the optimal solution (disks y_i) has area πk.

the set or is adjacent to some vertex of it.

Algorithm 1: The Two-by-two algorithm for MACS

Input: $X \subseteq \mathbb{R}^2, k \geq 0$, where X is finite and $k \leq |X|$.
Output: a feasible set of size k.

if k *is even* **then**
  $S \leftarrow$ any two intersecting disks of X;
else
  $S \leftarrow$ any one disk of X;
while $|S| \leqslant k-2$ **do**
  $\{x, x'\} \leftarrow \arg\max \{\mathcal{A}(S \cup \{x, x'\}) \colon x, x' \in X,\ S \cup \{x, x'\}$ is feasible $\}$;
  $S \leftarrow S \cup \{x, x'\}$;
return S;

One can find an example similar to Figure 3 to show that optimising the initial choice of the first disk(s) does not improve the approximation ratio.

▶ **Theorem 4** (Approximation). MACS *has a* $(1/2)$*-approximation that can be computed in polynomial time (Algorithm 1).*

We assume here that the input graph is connected. Otherwise, one may consider the largest solution over all connected components. The execution of Algorithm 1 is divided in two phases. An iteration belongs to the first phase as long as the current solution S is not a dominating set in the graph $\mathrm{UDG}(X)$.

During the first phase, in each iteration the area covered increases by at least π. During the second phase, since the current solution is a dominating set, any disk can be added while keeping the solution feasible. Therefore, the algorithm is then a standard greedy algorithm to maximize a submodular function, and the analysis is similar to the proof that Nemhauser's algorithm is a $\left(1 - \frac{1}{e}\right)$-approximation for classic submodular functions, showing that the approximation ratio of Algorithm 1 stays greater than 1/2 during the second phase.

Proof. We first analyze the even case where $k = 2\kappa$, and then we reduce the odd case to the even one. Let $S_\kappa = \{x_1, x_2, \ldots, x_{2\kappa}\}$ be the solution returned by the algorithm. Let $S_i = \{x_1, \ldots, x_{2i}\}$ be the set right before the i-th iteration, and d be the smallest integer such that S_d is a dominating set in UDG(X). If such an integer does not exist, i.e., S_κ is not a dominating set, then set $d = \kappa$.

▷ Claim 14. The area $\mathcal{A}(S_d)$ is at least πd.

Proof. For $i < d$, S_i is not a dominating set. Then there exist two disks y, y' such that $B(S_i) \cap B_y = \emptyset$ and $S \cup \{y, y'\}$ is connected. Adding such a pair increases the area covered by at least $\mathcal{A}(B_y) = \pi$. Since (x_{2i+1}, x_{2i+2}) is chosen to maximize $\mathcal{A}(S_i \cup \{x, x'\})$ among all feasible pairs, $\mathcal{A}(S_{i+1}) \geqslant \mathcal{A}(S_i \cup \{y, y'\}) \geqslant \mathcal{A}(S_i) + \pi$. By induction, $\mathcal{A}(S_d) \geqslant \pi d$. ◀

Note that when $d = \kappa$, Claim 14 immediately implies that $\mathcal{A}(S_\kappa) \geqslant \frac{\mathbf{OPT}}{2}$. Also remark that, regardless of the initial choice, the area covered by the first two disks is at least π. This observation will be useful when we prove the case where k is odd.

▷ Claim 15. For all $d \leqslant i \leqslant \kappa$, $\mathcal{A}(\mathbf{OPT}) \leqslant \mathcal{A}(S_i) + \kappa \cdot (\mathcal{A}(S_{i+1}) - \mathcal{A}(S_i))$.

Proof. It is easy to check that the function $\mathcal{A}(\cdot)$ satisfies the following properties for all $H \subseteq H' \subseteq X$:

positivity : $\mathcal{A}(H) \geqslant 0$.
monotonicity : $\mathcal{A}(H) \leqslant \mathcal{A}(H')$.
submodularity : $\forall H'' \subseteq X,\ \mathcal{A}(H' \cup H'') \leqslant \mathcal{A}(H \cup H'') - \mathcal{A}(H) + \mathcal{A}(H')$.

Let $\mathbf{OPT} = \{y_1, \ldots, y_{2\kappa}\}$. We have for all $d \leqslant i \leqslant \kappa$:

$$\begin{aligned}
\mathcal{A}(\mathbf{OPT}) \leqslant \mathcal{A}(S_i \cup \mathbf{OPT})& \\
&= \mathcal{A}(S_i) + (\mathcal{A}(S_i \cup \{y_1, y_2\}) - \mathcal{A}(S_i)) + \ldots \\
&\qquad + (\mathcal{A}(S_i \cup \{y_1, \ldots, y_{2\kappa}\}) - \mathcal{A}(S_i \cup \{y_1, \ldots, y_{2\kappa-2}\})) \\
&\leqslant \mathcal{A}(S_i) + (\mathcal{A}(S_i \cup \{y_1, y_2\}) - \mathcal{A}(S_i)) + \cdots + (\mathcal{A}(S_i \cup \{y_{2\kappa-1}, y_{2\kappa}\}) - \mathcal{A}(S_i)) \\
&\leqslant \mathcal{A}(S_i) + \kappa \cdot (\mathcal{A}(S_i \cup \{x_{2i+1}, x_{2i+2}\}) - \mathcal{A}(S_i)) \\
&= \mathcal{A}(S_i) + \kappa \cdot (\mathcal{A}(S_{i+1}) - \mathcal{A}(S_i)) .
\end{aligned}$$

The first and the second inequality respectively come from *monotonicity* and *submodularity*, while the third one follows from the fact that for $i \geq d$ (x_{2i+1}, x_{2i+2}) is the pair of disks maximizing $\mathcal{A}(S_i \cup \{x, x'\})$ among **all pairs** (x, x') in X. As S_d is a connected dominating set in X, all pairs (y_{2j-1}, y_{2j}) for $1 \leqslant i \leqslant \kappa$ are considered. ◀

We can now re-write Claim 15 as

$$\text{For all } d \leqslant i \leqslant \kappa : \mathcal{A}(S_{i+1}) \geqslant \left(1 - \frac{1}{\kappa}\right) \mathcal{A}(S_i) + \frac{\mathbf{OPT}}{\kappa}.$$

Combined with Claim 14, simple algebra yields that for $d \leqslant i \leqslant \kappa$, we have

$$\mathcal{A}(S_i) \geqslant \left[1 - \left(1 - \frac{d}{2\kappa}\right)\left(1 - \frac{1}{\kappa}\right)^{i-d}\right] \mathbf{OPT}.$$

Therefore, for $i = \kappa$ we have

$$\mathcal{A}(S) = \mathcal{A}(S_\kappa) \geqslant \left[1 - \left(1 - \frac{d}{2\kappa}\right)\left(1 - \frac{1}{\kappa}\right)^{\kappa - d}\right] \mathbf{OPT} = \left[1 - \frac{1}{2}(1+t)\left(1 - \frac{1}{\kappa}\right)^{\kappa t}\right] \mathbf{OPT}$$

where $t = \dfrac{\kappa - d}{\kappa} \in [0, 1]$. As $1 + x \leqslant e^x$ for all $x \in \mathbb{R}$, we get

$$\mathcal{A}(S) \geqslant \left(1 - \frac{1}{2}(1+t)e^{-t}\right) \mathbf{OPT} \geqslant \left(1 - \frac{1}{2}e^t e^{-t}\right) \mathbf{OPT} = \frac{1}{2}\mathbf{OPT},$$

concluding the proof of the case when k is an even number.

For the odd case $k = 2\kappa - 1$: in the first iteration, instead of adding two disks to S_1, we add a single disk of X to S_1. This is equivalent to adding two copies of the same disk. This iteration belongs to the first phase, and the only properties we used in the first phase is that each iteration adds an area of π, and keeps the solution feasible; these are clearly true for the first iteration even with one disk. ◀

Figure 3 shows a tight example.

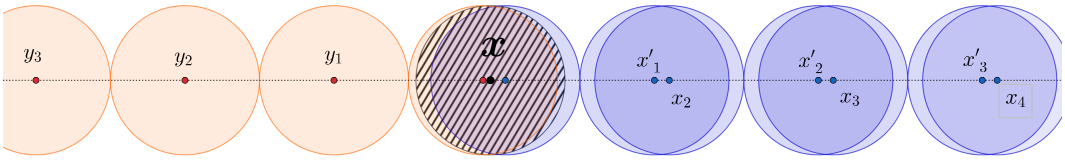


Figure 3 A tight example for Algorithm 1. For any $\varepsilon > 0$, X contains $x = (0, 0)$ (stripe-shaded), $x_i = (2(i-1) + i\varepsilon, 0)$ and $x'_i = ((2+\varepsilon)i, 0)$ for $1 \leqslant i \leqslant k$ (blue) and $y_i = (-2i - \varepsilon/2, 0)$ for $0 \leqslant i \leqslant k$ (orange). Suppose that $k = 1 + 2\kappa$ is odd and the algorithm starts with $S_0 := \{x, x\}$. Then the algorithm will add $\{x_i, x'_i\}$ in iteration i since it covers more additional area than $\{y_0, y_1\}$. The solution returned (blue disks) covers an area of $\pi + \kappa(\pi + f(\varepsilon)) \approx \frac{k}{2}\pi$, for some function $f(\cdot)$ with $\lim_{\varepsilon \to 0} f(\varepsilon) = 0$, while **OPT** (orange disks) covers an area $k\pi$.

5 Proof of Theorem 5: PTAS with resource augmentation

Let us recall our main result.

▶ **Theorem 5** (Resource augmentation). *Let $\varepsilon > 0$ be fixed. Given a set of points $X \subseteq \mathbb{R}^2$ and a non-negative integer k, there is an algorithm (Algorithm 2) that computes, in time $n^{\mathcal{O}(\varepsilon^{-3})}$, a subset $S \subseteq X$ of size at most k and a set $S_{add} \subseteq \mathbb{R}^2$ of at most εk points, such that* UDG $(S \cup S_{add})$ *is connected, and the area covered by the unit disks centered at S is at least $(1 - \varepsilon)\boldsymbol{OPT}(X, k)$.*

We summarise the high level ideas here and fill in the details in the subsequent sections. Let (X, k) denote an input of MACS and **OPT** be the optimal solution of MACS on input (X, k). When the context is clear **OPT** can also denote the total area covered by the union of the unit disks centered in points of **OPT**.

We start by guessing a bounding box of size $\Theta(k) \times \Theta(k)$ that contains **OPT**. Next, another square of size $L \times L$, where $L = \Theta(k)$, is randomly shifted so that it always contains the bounding box. We remove all disks that are outside the square. That square is then recursively partitioned into smaller squares until they have (large) constant size. The hierarchical dissection induces a grid.

We remove all disks that intersect the lines of the grid. In contrast, we deploy some new, additional disks (X_{add}) in some strategic *portal* positions along the lines and near the boundary of all the smallest squares.

Next, we use dynamic programming to build a solution from the smallest squares upwards. The difficulty lies in having to guarantee the connectivity when combining solutions from smaller squares into larger squares using additional disks, while controlling the time complexity and number of disks added.

The key of our approach lies in Lemma 21, in which we argue that with constant probability, there exists a well-structured near-optimal solution that uses at most εk additional disks.

5.1 The grid.

The first step is to reduce significantly the size of the input by guessing the position of the optimal solution.

▶ **Lemma 16.** *There exists a point $c \in X$ such that $\boldsymbol{OPT}$ is contained in an axis-parallel square, of side length $4k$, centered in c.*

Proof. For c, take any disk in $\mathbf{OPT}$ and recall that $\mathbf{OPT}$ is connected and has at most k disks, so all the disks in $\mathbf{OPT}$ are contained in the square centered at c and with side length $4k$. ◀

Let L' be the sidelength of the box given by the Lemma 16, and set X' be the set of points of X lying inside this box. Let L be the smallest power of 2 greater than $2L'$. The *root square* is defined to be the axis-parallel $L \times L$ square with the same left-bottom corner as the bounding-box.

A *shift* is an non-negative integer a smaller than or equal to $L/2$. We say that the root square is *shifted* by a if it is translated by the vector $(-a, -a)$. Notice that any shifted root square contains the bounding-box.

Given a shifted root square, we can define its *dissection* as a recursive partitioning into smaller squares. The $L \times L$ root square is divided into four squares of size $L/2 \times L/2$. Each of these squares is again divided into four $L/4 \times L/4$ squares, so forth. The process stops when the side length of a square is equal to $L_0 = \Theta(\varepsilon^{-1})$. Let $d = \log(L/L_0) = \mathcal{O}(\log(\varepsilon k))$. We can think of this partitioning as 4-ary tree, where each node at level ℓ corresponds to a $L_0 2^\ell \times L_0 2^\ell$ square and has four children corresponding to four $L_0 2^{\ell-1} \times L_0 2^{\ell-1}$ squares. The root square is at level 0 and the *leaf squares* are at level d. Given two squares of level ℓ and level ℓ', $\ell > \ell'$, we say the former is of higher level than the latter. So the leaf square is the one with the highest level.

This dissection defines a *grid* composed of $2 \cdot (2^d - 1)$ horizontal and vertical lines of length L. We say that a line is at *level* $\ell \in \{1, \ldots, d\}$ if it was added on the grid to divide a square at level $l - 1$ into four squares at level ℓ. There are 2^ℓ horizontal (*resp.* vertical) lines at level ℓ. See figure 4.

On each horizontal line of level $\ell \geqslant 1$, we will place a set of *vertical*—notice the naming asymmetry—*portals* of level ℓ, near which (not exactly on which) we will deploy the portal disks to facilitate the connection of disks on both sides of this line. We define a set of horizontal portals for each vertical line in an analogous manner. Notice that it is possible that a point is both a vertical portal and a horizontal portal. Let $m = \mathcal{O}(\varepsilon^{-1} d)$ be a power of two. Along a line of level ℓ, there are $m2^\ell + 1$ portals evenly spaced so that the distance between two neighboring portals have distance exactly $\frac{L}{m2^\ell}$.

▷ Observation 17. If an horizontal line of level ℓ crosses a vertical line of level greater than or equal to ℓ then the intersection point is a horizontal portal.

We define a set $\mathcal{P}$ of *portal disks* which we position at or near the portals. If a portal (i, j) is on exactly one line of the grid then we add the portal disk (i, j) to $\mathcal{P}$. If a portal (i, j) is at the intersection of two lines of the grid, then *i*) if it is a horizontal portal then we add to $\mathcal{P}$ two portal disks $(i, j + 2)$ and $(i, j - 2)$, and *ii*) if it is a vertical portal then we add to $\mathcal{P}$ two portal disks $(i - 2, j)$ and $(i + 2, j)$.

Given a square C of the dissection, the *potential portal disks* of C, denoted by $\mathcal{P}_C$, are the portal disks on the boundary of C.

▷ Observation 18. For any square, the number of potential portal disks is $\mathcal{O}(m) = \mathcal{O}(\varepsilon^{-1} \log(\varepsilon k))$.

The *border* of a leaf square C, denoted as ∂C, is the set of points in C within distance 1 from C's boundary. The remaining points of C are called the *core* of C, written as *core*(C). A unit disk with its center in C intersects the boundary if and only if its center lies in the border. If two disks are in the core of two different leaf squares, then they do not intersect. We refer to the union of the core of all *leaf squares* as the *core*. In a leaf square $C = [a, b] \times [c, d]$, the set of points formed by the boundary of the square $[a + 2, b - 2] \times [c + 2, d - 2]$ is called the *fence*. We cover the fence of C by *fence disks*, aligned such that each corner of this square is the center of a fence disk. See Figure 5. We denote by $\mathcal{F}$ the set of all fence disks for all leaf squares. The set of portal disks and fence disks form the set of *additional disks* $X_{add} = \mathcal{P} \cup \mathcal{F}$.

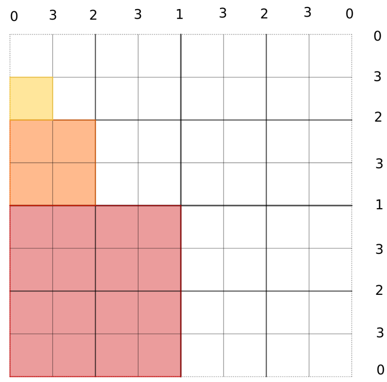


Figure 4 An illustration of the grid with $d = 3$. Numbers on the top and the right are the level of the corresponding lines and the red, orange and yellow are respectively the example of square of the dissection at level $1, 2$ and 3.

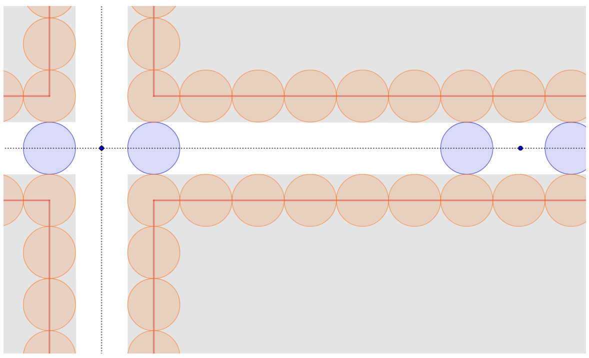


Figure 5 The grey and white area are respectively the *core* and the *border*. Dotted lines are from the grid while the orange lines represent the *fence* and orange disks are the *fence disks*. Blue points are (vertical) *portals* and blue disks are *portal disks*.

5.2 Dynamic program

The algorithm uses dynamic programming. The dynamic programming table is indexed by *configurations*.

▶ **Definition 19.** *A* configuration *is a 5-tuple* $\mathcal{C} = [C, t, t_{add}, P, \sim]$*, where:*
- C *is a square of the dissection.*
- $0 \leq t \leq k$ *is an integer, denoting the number of disks of* S *used by the solution inside* C*.*
- $0 \leq t_{add} \leq \varepsilon k$ *is an integer, denoting the number of additional disks used by the solution inside* C*.*
- $P \subseteq \mathcal{P}_C$ *is a subset of potential portal disks of* C*, those that are used by the solution.*
- $\sim$ *is a planar connectivity relation on* P *(described below), representing the connectivity achieved so far by the part of the solution inside* C*.*

In the following, to facilitate discussion, we will refer to portals disks as simply portals. An equivalence relation $\sim$ on P is a *planar connectivity relation* if each equivalence class has an associated tree with the portals at the leaves, and there exists a planar embedding of those trees inside the square, such that the trees do not intersect.

The content of the dynamic programming table, the *value* of a configuration $\mathcal{C} = [C, t, t_{add}, P, \sim]$, denoted by $\mathcal{A}(\mathcal{C})$, is the maximum area that can be covered by a set $S \subseteq X$ of t disks in $C \cap core$[4], such that there is a set $S_{add} \subseteq X_{add}$ of t_{add} additional disks such that any $p, p' \in P$ with $p \sim p'$ are in the same connected component induced by $S \cup S_{add} \cup P$. We say that p and p' are *connected in* $\mathcal{C}$. If such sets $\{S, S_{add}, P\}$ do not exist for configuration $\mathcal{C}$, the value $\mathcal{A}(\mathcal{C})$ is set to $-\infty$.

We next explain how to fill in the table.

5.2.1 Computing leaf entries of the dynamic programming

We first explain how to fill the entries of the table corresponding to the leaf squares. For each leaf square C, we enumerate
1. all possible subsets $S \subseteq X' \cap core(C)$ of at most k_0 disks, for a parameter $k_0 = \mathcal{O}(\varepsilon^{-3})$ (see Lemma 21).
2. all possible subsets $S_f \subseteq \mathcal{F} \cap C$,
3. all possible subsets $P \subseteq \mathcal{P}_C$, and
4. all possible planar connectivity relations $\sim$ on P.

We say that $(S, S_f, P, \sim)$ is a *guess* in C and that it is *usable* if one of the following two conditions holds:

Case 1. if $P = \emptyset$, then $S \cup S_f$ is connected, otherwise
Case 2. every connected component of $S \cup S_f \cup P$ contains at least one portal disk in P.

Each usable guess (S, S_f, P) in C corresponds to a configuration $\mathcal{C} := [C, |S|, |S_f|, P, \sim]$, where $\sim$ is the planar connectivity relation on P induced by the connected components of $S \cup S_f \cup P$.

Several usable guesses (S, S_f, P) can potentially correspond to the same configuration $\mathcal{C}$. The value of $\mathcal{C}$ is computed[5] as the maximum value $\mathcal{A}(S)$ over all such guesses S.

5.2.2 Computing all entries

It remains to show how to compute the solution of a configuration, say $\mathcal{C} = [C, t, t_{add}, P, \sim]$, for a square C at level ℓ, by combining the solutions $\left[C^i, t^i, t^i_{add}, P^i, \sim^i\right]$ of the four child

[4] Recall that *core* is the union of the $core(C)$ of all leaf squares C.
[5] The area covered by the union of a set of disks is a real number that can be computed exactly. When the desired accuracy is a fixed constant (for instance ε), one can give an approximation of this area with the desired precision in polynomial time.

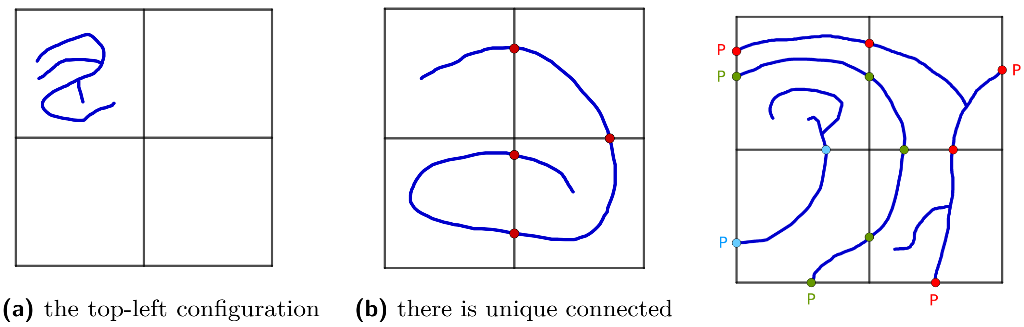
squares C^i, $i = 1, 2, 3, 4$, at level $\ell + 1$. Recall that connectivity relations $\sim^i$ capture the information about connectivity in the squares C^i. Let $P = \{p_0, \ldots, p_s\}$ be the subset of potential portal disks. We define $\sim'$ as the *transitive closure* of all $\sim^i$: $p \sim' p'$ if and only if there exists a sequence of squares $i_1, \ldots, i_s \in \{1, 2, 3, 4\}$ and a sequence of portals $p = p_0, \ldots, p_s = p'$ such that for all $1 \leqslant j \leqslant s$, p_j is a common portal of $P^{i_{j-1}}$ and P^{i_j}. Further, p_{j-1} and p_j must be connected in $\mathcal{C}^{i_j}$. We call $\mathcal{C}$ *empty* if $P = \emptyset$ and $t = 0$, and *closed* if $P = \emptyset$ and $t > 0$.

We now define the notion of compatibility of configurations.

▶ **Definition 20.** *Five configurations* $\left(\mathcal{C}, \mathcal{C}^1, \mathcal{C}^2, \mathcal{C}^3, \mathcal{C}^4\right)$ *with* $\mathcal{C} = [C, t, t_{add}, P, \sim]$ *and* $\mathcal{C}^i = [C^i, t^i, t^i_{add}, P^i, \sim^i]$ *are* compatible *if all the following properties are satisfied.*

1. *all* C^i *have the same level and their union is the square* C*.*
2. $P = \bigcup_{i=1}^4 P^i \cap \partial C$.
3. $\sim$ *is the restriction of the transitive closure* $\sim'$ *of* $\left(\sim^i\right)_{1 \leqslant i \leqslant 4}$ *to* P*.*
4. $t = t^1 + t^2 + t^3 + t^4$ *and* $t \leqslant k$*.*
5. $t_{add} = t^1_{add} + t^2_{add} + t^3_{add} + t^4_{add} + |\bigcup_{i=1}^4 P^i \setminus P|$ *and* $t_{add} \leqslant \varepsilon k$*.*
6. *exactly one of following three conditions holds.*
   (a) $\mathcal{C}^i$, $i \in \{1, 2, 3, 4\}$, *is closed and all* $\mathcal{C}^j$, $j \neq i$ *are empty.*
   (b) $\mathcal{C}$ *is closed and there is exactly one equivalence class for* $\sim'$*.*
   (c) all equivalence classes of $\sim'$ *contain a portal in* P*.*

Remark. By condition 2, the set P of portals used by C is obtained by removing from $\bigcup_{i=1}^4 P^i$ the portals not on the border of C. Notice that these removed portals in $\bigcup_{i=1}^4 P^i \setminus P$ are now counted as additional disks (in condition 5). Condition 6 attempts to capture all possible situations—either we have a single connected component not connected to the "outside world", which is a feasible solution by itself, (see Condition (6a) and Condition (6b)), or we have several connected components, each of which must be further connected to the outside world in a later stage (see Condition (6c)). See Figure 6. Finally, it is easy to see that if all $\sim^i$ satisfy the connectivity relation, then so does $\sim$.



(a) the top-left configuration is closed while other configurations are empty.

(b) there is unique connected component independent from the outside world.

(c) Each connected component contains a portal in P.

Figure 6 Illustration of cases (a)-(b)-(c) of point 6. in Definition 20

5.3 Structural Lemma

Let a be a shift chosen uniformly at random in $\left\{0, \frac{L}{2}\right\}$. We consider the grid associated to this shift and the set of additional disks on this grid as defined in the previous section. The following lemma is essential to our main theorem. Recall that $\mathcal{P}$ denotes the set of portal disks and $\mathcal{F}$ the set of fence disks.

▶ **Lemma 21** (Structural Lemma). *Given a fixed parameter $\varepsilon > 0$, there exists a subset $S \subseteq core$ of input disks and a set $S_{add} \subseteq \mathcal{P} \cup \mathcal{F}$ of* additional disks*, such that with probability at least* $1/3$*,*
(i) *(feasibility) $|S| \leqslant k$ and $S \cup S_{add}$ is connected,*
(ii) *(bounded resource augmentation) $|S_{add}| \leqslant \varepsilon k$,*
(iii) *(near-optimality) $\mathcal{A}(S) \geqslant (1-\varepsilon)\boldsymbol{OPT}$,*
(iv) *(bounded local size) For each leaf square C, $|C \cap S| = \mathcal{O}(\varepsilon^{-3})$.*

Our dynamic programming aims at finding a solution satisfying all conditions of this Structural Lemma. We show that such a solution can be computed in time $n^{\mathcal{O}(\varepsilon^{-3})}$. The *bounded local size* property ensures that we can try all possible configurations in the leaf squares in polynomial time. We also prove that for any square, the number of different planar connectivity relations is upper-bounded by the *Catalan number* of the number of potential portal disks of the square. It follows from Observation 18 that this number is polynomially bounded.

5.4 Proof of the structural Lemma

We construct S and S_{add} from $\mathbf{OPT}$ in two steps. In the first step, we build sets S' and S_{add} that satisfy properties *(i)-(iii)*; and in the second step, we construct $S \subseteq S'$ by removing some disks from S' so as to satisfy property *(iv)* while maintaining the validity of the three first properties.

5.4.1 Part 1 : Construction of the set of additional disks

Fix any shift, consider its associated grid and dissection and the corresponding set of additional disks $X_{add} = \mathcal{P} \cup \mathcal{F}$. Let S' be the union of disks in $\mathbf{OPT}$ that are located in the core of a leaf square of the dissection, namely

$$S' = \mathbf{OPT} \cap core.$$

Observe that S' might be disconnected since we have removed from $\mathbf{OPT}$ all the disks that were intersecting the grid. Letting *border* denote $\bigcup_{C \text{ is leaf}} \partial(C)$, we show how to replace the set of input disks $\mathbf{OPT} \cap border$ by a subset $S_{add} \subseteq \mathcal{F} \cup \mathcal{P}$ of additional disks.

Each leaf square $[a, b] \times [c, d]$ has an associated fence that is the boundary of the square $[a+2, b-2] \times [c+2, d-2]$. For each vertical (*resp.* horizontal) portal disk (x, y), we define a *connection line*, which is $\{x\} \times [y-2, y+2]$ (*resp.* $[x-2, x+2] \times \{y\}$). The set of fences and connection lines naturally partition the set of points which are at distance at most 2 from the lines of the grid into a set of *rectangles* $\mathcal{R}$. See Figure 7.

Notice that all connections and fences are covered by the union of additional disks. Given a rectangle $R \in \mathcal{R}$, we define $\text{disk}(R) \subseteq X_{add}$ as the minimal set of additional disks that contain R.

We construct S_{add} as the union of $\text{disk}(R)$, over all rectangles R that intersect a disk $x \in \mathbf{OPT} \cap border$.

$$S_{add} = \bigcup \{\text{disk}(R) \colon R \in \mathcal{R}, \exists x \in \mathbf{OPT} \cap border \text{ such that } B_x \cap R \neq \emptyset\}$$

Notice that each disk $x \in \mathbf{OPT} \cap border$ intersects at most two rectangles. Furthermore, such a disk does not intersect with any fence and can intersect at most one connection line.

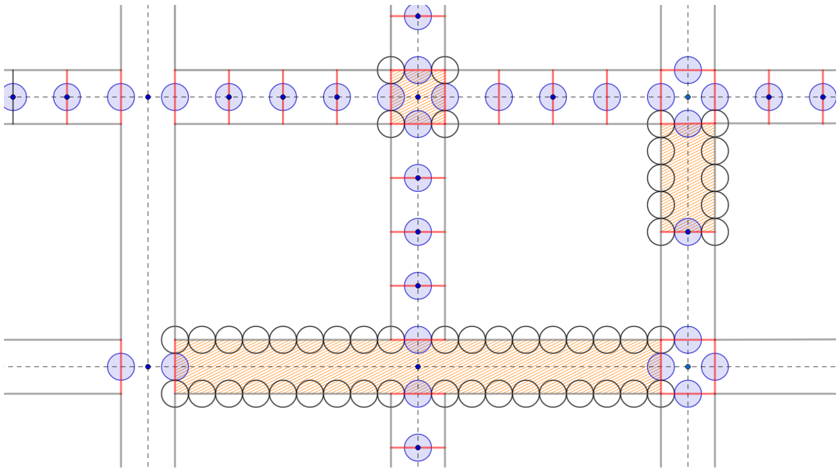


Figure 7 Dotted lines are the grid lines. The bottom and top horizontal lines have respectively level 8 and 10, and the vertical lines from left to right have level $5, 10$ and 9. Grey continuous line are the *fence*, and the red ones, the *connection lines*. Points and blue disks are *portals* and *portal disks*. Striped orange areas illustrate some rectangles $R \in \mathcal{R}$, and other disks are *fence disks* of the corresponding sets $\mathrm{disk}(R)$.

▷ Claim 22. Sets S' and S_{add} are such that $S' \cup S_{add}$ is connected, S' has size at most k and with probability at least $1/3$: $|S_{add}| \leqslant \mathcal{O}(\varepsilon k)$ and $\mathcal{A}(S') \geqslant (1 - \mathcal{O}(\varepsilon))\mathbf{OPT}$.

The argument is similar to the one of Arora [1]. We first upper-bound the expectation of $|S_{add}|$ and $\mathcal{A}(\mathbf{OPT}) - \mathcal{A}(S')$, and then use Markov's inequality. To bound the expectation of $|S_{add}|$, we observe that the number of additional disks added in S_{add} for each disk in $\mathbf{OPT}$ intersecting a line at level ℓ is $\mathcal{O}(L/(m2^\ell))$ while the probability that a disk intersects a line at level ℓ is $\mathcal{O}(2^\ell/L)$.

Proof. (Claim 22)

Clearly $|S'| \leqslant |\mathbf{OPT}| \leqslant k$. We now prove that $S' \cup S_{add}$ is connected. Suppose that there exists a disk $x \in \mathbf{OPT} \cap \mathit{border}$ such that $\mathbf{OPT} \setminus \{x\}$ is split into several connected components. We know that x intersects only one rectangle $R_1 \in \mathcal{R}$ or two rectangles $R_1, R_2 \in \mathcal{R}$. Since $\mathbf{OPT}$ is connected, and B_x is contained in the set $U = R_1$ or $U = R_1 \cup R_2$, each connected component intersects the boundary of U. Then, B_x intersects a disk in $\mathrm{disk}(R_1)$ or $\mathrm{disk}(R_2)$. Therefore, $\mathbf{OPT} \setminus \{x\} \cup \mathrm{disk}(R_1) \cup \mathrm{disk}(R_2)$ is connected. By doing so for each $x \in \mathbf{OPT} \cap \mathit{border}$, it follows that $S' \cup S_{add}$ is connected.

It remains to show that, under a uniform random shift a, with probability at least one third we have $|S_{add}| \leqslant \mathcal{O}(\varepsilon k)$ **and** $\mathcal{A}(S') = \mathcal{A}(\mathbf{OPT} \cap \mathit{core}) \geqslant (1 - \mathcal{O}(\varepsilon))\mathbf{OPT}$. The proof is very similar to Arora's approach, we first upper-bound the expectation of $|S_{add}|$ and $\mathcal{A}(\mathbf{OPT}) - \mathcal{A}(S')$, and then use *Markov inequality* to conclude.

We first upper-bound the expected number of additional disks. For each $x \in \mathbf{OPT}$ intersecting a line at level ℓ, we have added at most two sets of additional disks associated to rectangles with side length smaller than the distance between two consecutive portals of this line. It follows that $\mathcal{O}(L/(m2^\ell))$ additional disks have been added to S_{add} for each disk in $\mathbf{OPT}$ intersecting a line of level ℓ. This can be observed in Figure 8. Moreover, the probability that a disk intersects a line at level ℓ is $\mathcal{O}(2^\ell/L)$. Then,

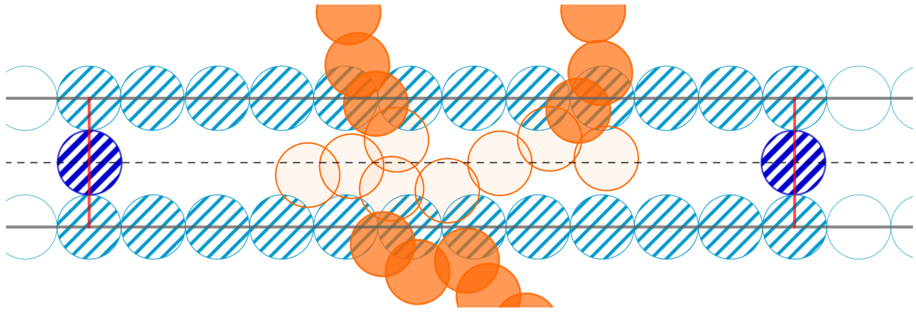


Figure 8 **OPT** is represented by orange disks. Disks of **OPT** that intersect the grid (dotted line) are replaced by additional disks (striped blue disks). This operation maintains the connectivity of the set.

$$\begin{aligned}\mathbb{E}(|S_{add}|) &\leqslant \sum_{x\in\mathbf{OPT}}\sum_{\ell=0}^{d-1} P(x \text{ intersects exactly one line at level } \ell)\mathcal{O}(\frac{L}{m2^\ell})\\ &= \sum_{x\in\mathbf{OPT}}\sum_{\ell=0}^{d-1}\mathcal{O}(\frac{2^\ell}{L}\cdot\frac{L}{m2^\ell}) = \mathcal{O}(\frac{dk}{m}) = \mathcal{O}(\varepsilon k)\end{aligned}$$

We now upper-bound the expectation of $\mathcal{A}(\mathbf{OPT}) - \mathcal{A}(S')$. First we have $\mathcal{A}(\mathbf{OPT}) - \mathcal{A}(S') \leqslant \mathcal{A}(\mathbf{OPT} \cap border)$, and the probability that a point $p \in B(\mathbf{OPT})$ is in $B(\mathbf{OPT} \cap border)$ is smaller that p is at distance 2 from the lines of the grid. Therefore

$$\begin{aligned}\mathbb{E}\left(\mathcal{A}(\mathbf{OPT}) - \mathcal{A}(S')\right) &\leqslant \mathbb{E}(\mathcal{A}(\mathbf{OPT} \cap border))\\ &\leqslant \int_{p\in B(\mathbf{OPT})} P(p \text{ is at distance at most 4 from the grid})\mathrm{d}p\\ &\leqslant \int_{p\in B(\mathbf{OPT})} 2\cdot\frac{4}{L_0}\mathrm{d}p\\ &\leqslant \frac{8\cdot\mathbf{OPT}}{L_0} = \mathcal{O}(\varepsilon\mathbf{OPT})\end{aligned}$$

By choosing the constant properly in the big O notation and using the Markov inequality, we can show that the probability of $|S_{add}| > O(\varepsilon k)$ and the probability of $A(OPT) - A(S) > O(\varepsilon OPT)$ are both upper bounded by $\frac{1}{3}$. Thus, by a union bound, we conclude the proof. ◀

5.4.2 Part 2 : Sparsification of S'

The sets $S' \cup S_{add}$ obtained so far may not satisfy the last property (*bounded local size*). In this section, we show how to remove some disks from S' to guarantee this property while still maintaining the other required properties in Lemma 21.

Suppose that there exists a leaf square C such that $S'_C := S' \cap C$ has size greater than $k_0 := (1+\beta^{-1})L_0^2 = \mathcal{O}(\varepsilon^{-3})$, where $\beta = \min\{\varepsilon/12, 1\}$. Then the core of C is "overcrowded" and we show how to construct a non-overcrowded subset maintaining connectivity while losing only an $\varepsilon/2$-th fraction of the covered area.

Define a set S to be initially equal to S'. Consider each overcrowded leaf square C one by one, and define $S_C = S \cap C$. Start with an empty set H and for each disk $x \in S_C$, add

x in H if $\mathcal{A}(H \cup \{x\}) - \mathcal{A}(H) \geqslant \beta$. Define $\overline{H} = S_C \setminus H$ as the complement of H and then apply Claim 23 to $G = UDG(S \cup S_{add})$ and $D = S \cup S_{add} \setminus \overline{H}$ to define $D' \subseteq \overline{H}$. Finally update S to $(S \setminus \overline{H}) \cup D'$.

▷ Claim 23. Let $G = (V, E)$ be a connected graph and D a dominating set with μ connected components. There exists a subset $D' \subseteq V \setminus D$ of size at most $2(\mu - 1)$ such that $G[D \cup D']$ is connected.

Proof. Let H and H' be two connected components in D that minimize $d_G(H, H')$. Then, $d_G(H, H') \leqslant 3$. Indeed, if $d_G(H, H') \geqslant 4$, then there exists a vertex x on a shortest path from H to H' that is not dominated by D. This implies that we can find two vertices that connect H and H'. We repeat this operation until there is only one connected component. This requires at most $2(\mu - 1)$ vertices. ◀

The following claim, together with Claim 22 ensures that sets S and S_{add} built in Part 1 and Part 2 satisfy the expected properties of our structural Lemma.

▷ Claim 24. The constructed sets S and S_{add} satisfy
(i) $S \cup S_{add}$ is connected,
(ii) for each leaf square C, $|S \cap C| \leqslant k_0$, and
(iii) $\mathcal{A}(S) \geqslant (1 - \varepsilon/2)\mathcal{A}(S')$.

This Claim might not be true if the radius of disks considered are arbitrary. The proof of this fact follows from geometrical observations about unit disks.

Proof. (claim 24)

For (i), we just need to argue that for each leaf square C, after $\overline{H}$ is defined, $S \cup S_{add} \setminus \overline{H}$ is a dominating set in $UDG(S \cup S_{add})$ (then the proof follows from Claim 23). Indeed if a disk x is in $\overline{H}$ then it means that $\mathcal{A}(H \cup \{x\}) - \mathcal{A}(H) < \beta \leqslant 1$. In particular, it implies that there exists a disk in $H \subseteq S \cup S_{add} \setminus \overline{H}$ that intersects x.

For (ii), observe that the size of $S \cap C$ is the sum of the size of the corresponding sets H and D' built during the "sparsification" of C. Since all disks in H increases the area covered by at least β and are contained in a square of area L_0^2, the number of disks in H is upper-bounded by $\beta^{-1}L_0^2$. Moreover, each connected component of $S \cup S_{add} \setminus \overline{H}$ had a disk contained in C so that the number μ of connected component is upper-bounded by $L_0^2/\pi < L_0^2/2$. Therefore $|D'| < L_0^2$. Finally $|H \cup D'| < (1 + \beta^{-1})L_0^2 = k_0$.

For (iii), we start by observing that the union $B(S')$ of disks in S' is contained in the set $B^+(S)$, which is defined as

$$B^+(S) := \{z \in \mathbb{R}^2 \mid \exists x \in S \text{ such that } ||z - x|| \leqslant 1 + \beta\}$$

Indeed, if there exists a point p covered by a disk x in S' but at distance at least $1 + \beta$ from any disk of S then adding x to S would increase the area covered by S by more that β. Therefore, we have the following inclusion

$$B(S) \subseteq B(S') \subseteq B^+(S), \tag{3}$$

and if the following geometrical claim holds, our proof of (iii) will be complete.

▷ Claim 25. $\mathcal{A}(B(S)) \geqslant (1 - \varepsilon/2)\mathcal{A}(B^+(S))$

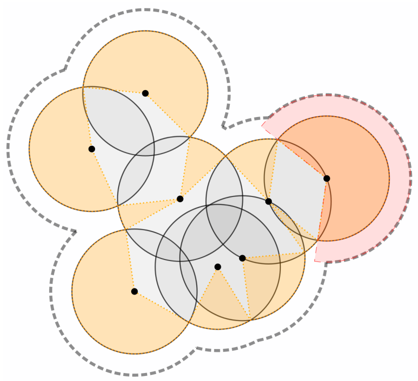


Figure 9 S consists of grey disks. The boundary of $B^+(S)$ is the dotted curve. Circular sectors are in orange while the red one represents a circular sector in $B^+(S)$.

The result follows from the fact that $B(S)$ is a union of unit-disks. See Figure 9. The boundary of $B(S)$ is made of *circular arcs* and each of these arcs is associated with a *circular sector* θ_i. Circular sectors intersect with other circular sectors only on the extreme points of their corresponding arcs, thus $\mathcal{A}(\cup_i \theta_i) = \sum_i \mathcal{A}(\theta_i)$.

We can associate with each circular sector θ_i (of a disk of radius 1) its "dilation" θ_i^+ which corresponds to the same circular sector in a disk of radius $1+\beta$. We have $\mathcal{A}(\theta_i^+) = (1+\beta)^2\mathcal{A}(\theta_i)$ and can see that $B^+(S) \setminus B(S) \subseteq \bigcup_i (\theta_i^+ \setminus \theta_i)$. Then

$$\begin{aligned}
\mathcal{A}(B^+(S)) - \mathcal{A}(B(S)) = \mathcal{A}(B^+(S) \setminus B(S)) &= \mathcal{A}\left(\bigcup_i (\theta_i^+ \setminus \theta_i)\right) \\
&\leqslant \sum_i \mathcal{A}(\theta_i^+ \setminus \theta_i) = \sum_i \mathcal{A}(\theta_i^+) - \mathcal{A}(\theta_i) \\
&\leqslant \sum_i (1+\beta)^2 \mathcal{A}(\theta_i) - \mathcal{A}(\theta_i) \\
&\leqslant \sum_i 3\beta \mathcal{A}(\theta_i) = 3\beta\mathcal{A}\left(\bigcup_i \theta_i\right) \leqslant 3\beta\mathcal{A}(B(S))
\end{aligned}$$

Therefore, $\mathcal{A}(B(S)) \geqslant \dfrac{\mathcal{A}(B^+(S))}{1+3\beta} \geqslant (1-\varepsilon/2)\mathcal{A}(B^+(S))$. This concludes the proofs of Claims 25 and 24. ◀

5.5 The algorithm

Algorithm 2: PTAS for MACS with resource augmentation

Input: X, k, ε.
Output: a real number $maxi \geqslant (1-\varepsilon)$**OPT**.

forall $c \in X$ **do**
  let $\mathcal{B}'$ be the $4k \times 4k$ square centered at c;
  $X' \leftarrow X \cap \mathcal{B}'$;
  $L \leftarrow$ the smallest power of 2 such that $L \geqslant 8k$;
  forall *shift* $a \in \{0, \dots, L/2\}$ **do**
    Create a table *tab*;
    foreach *configuration* $\mathcal{C}$ **do**
      $tab[\mathcal{C}] \leftarrow -\infty$;
    `/* Initialization */`
    foreach $\mathcal{C}$ *at level* d *(leaf square)* **do**
      $tab[\mathcal{C}] \leftarrow \max\{\mathcal{A}(S) : (S, S_f, P)$ is usable and corresponds to $\mathcal{C}\}$;
    `/* Fusion */`
    foreach *level* $0 \leqslant i \leqslant d-1$ *in decreasing order* **do**
      foreach *configuration* $\mathcal{C}$ *at level* i **do**
        $tab[\mathcal{C}] \leftarrow \max\left\{\sum_{i=1}^{4} tab[\mathcal{C}^i] : (\mathcal{C}, \mathcal{C}^1, \mathcal{C}^2, \mathcal{C}^3, \mathcal{C}^4) \text{ are compatible}\right\}$;

return $maxi = \max_{\substack{\text{configuration } \mathcal{C} \\ \text{for root square}}} tab\,[\mathcal{C}]$;

Notice that since the root square has no potential portals (portals are only placed on lines at level at least 1), any configuration that corresponds to the root square has only one connected component. We can easily add information in the table so that the algorithm also outputs the corresponding sets S and S_{add}.

Notice that Algorithm 2 tries all possible shift a. Our structural Lemma 21 ensures that there exists at least one shift such that the output satisfies all expected properties of Theorem 5.

▶ **Theorem 26.** *Algorithm 2 has a running time* $n^{\mathcal{O}(\varepsilon^{-3})}$.

The key ingredient in order to prove that our algorithm is polynomial follows from Observation 18. We show that the number of connectivity relations of a set of $\mathcal{O}(m)$ portals corresponds to its *Catalan number* which is polynomial when $m = \mathcal{O}(\varepsilon^{-1}\log(\varepsilon k))$.

Proof. (Theorem 26)

Size of tab. There exists 4^i squares at level i so the total of squares is $\sum_{i=0}^{d} 4^i = \mathcal{O}(4^{d+1})$. For any square C, the number of potential portal disks is at most $4m$. To see this, observe that if C is of level i, it is of size $L/2^i \times L/2^i$. Furthermore, it is surrounded by lines of level at most i and two adjacent portals on such a line has distance $\Omega(L/(m2^i))$. Therefore, the number of possible sets $P \subseteq \mathcal{P}_C$ is 2^{4m}, and for each set P of size r the total number of planar connectivity relations is equal to the r-th *Catalan number* : $P(r) = \dfrac{1}{r-1}\dbinom{2r}{r} = \mathcal{O}\left(\dfrac{1}{m-1}\dbinom{8m}{m}\right)$ and then by Stirling formula we get $P(r) = \mathcal{O}(4^{4m})$. To see that $P(r)$ is the r-th Catalan number, we check that it satisfies the same recurrence relation :

$$P(r) = \sum_{k=1}^{r} P(k-1) \cdot P(r-k) \tag{4}$$

with $P(0) = 1$. Indeed, if k denotes the index of the first portal p_k that is on the connected component of the r-th portal disk p_r, then the portal disk p_i with $1 \leqslant i \leqslant k-1$ cannot be equivalent to a portal p_j disk with $k \leqslant j \leqslant n$, and then the equivalence relation can be restricted to the set $\{p_i, 1 \leqslant i \leqslant k-1\}$ and there are $P(k-1)$ possible distinct choices. Next observe that since p_n and p_k are connected (i.e. $p_n \sim p_k$), it is enough to count the number of different equivalence relations in $\{p_j, k+1 \leqslant j \leqslant r\}$, which is $P(r-k)$. Finally, observe that k can be from 1 to r ($k = r$ means that p_r is alone in its connected component.) We thus concludes (4). Therefore, creating *tab* in line 6 can be done in time $\mathcal{O}(4^{d+1}\varepsilon k^2 8^{4m}) = k^{\mathcal{O}(1/\varepsilon)}$.

Initialization There exists 4^d leaf squares and for each of them, we try all possible guesses. This can be done in time $n^{\mathcal{O}(\varepsilon^{-3})}$.

Fusion Trying all possible combinations can be done in time $k^{\mathcal{O}(1/\varepsilon)}$

◀

6 A PTAS for well-distributed inputs

Let us recall the definition of well-distributed input.

▶ **Definition 6.** *Given $\alpha > 0$, a finite set X of points in the plane is called* α-well-distributed *if for all $x, y \in X$, $d_{UDG(X)}(x, y) \leqslant \lceil \alpha \cdot \|x - y\| \rceil$.*

Here $\lceil \cdot \rceil$ is the ceiling function. This ensures that the right-hand side is always at least one. Notice that a well-distributed set is necessarily connected.

One intuitive view of a well-distributed input is to look at the shape of the "holes" of the input, that are the different connected components of the complement of the union of the input disks in the plane. The assumption of well-distribution means that these holes are roughly *fat*.

One particular interesting case arises when there is no hole at all. We call these sets *pseudo-convex*, and we prove that this is a particular case of well-distributed inputs.

▶ **Definition 8.** *A set X is called* pseudo-convex *if the convex-hull of X is covered by the union of the unit disks centered at points of X.*

▶ **Lemma 9.** *A pseudo-convex set X is 3.82-well-distributed.*

Proof. (Lemma 9) Let X be a pseudo-convex set, G its unit-disk-graph, and x and y be any two disks in X at distance $L = \|x - y\|$. We show that $d_G(x, y) \leqslant \lceil \alpha L \rceil$ where $\alpha = 12/\pi < 3.82$.

If $L < 2$ then the two unit disks associated to x and y overlap so that $d_G(x, y) = 1 \leqslant \lceil \alpha L \rceil$. Otherwise suppose that $L \geqslant 2$. Since X is pseudo-convex, it is connected and any point in the line segment $[x, y]$ is covered by a disk in X. Let $S = \{z \in X \mid B_z \cap [x, y] \neq \emptyset, \|x - z\| > 2 \text{ and } \|y - z\| > 2\}$ and let I be any maximal independent set in $S \cup \{x, y\}$. Since S is at distance at least 2 from x and y, we deduce that $x, y \in I$ and all disks in $I \setminus \{x, y\}$ are inside a $L \times 4$ rectangle and then $|I| \leqslant 4L/\pi$. Since I is maximal, it is a dominating set in S. Therefore, claim 23 implies that there exists a connected subset $D \subseteq X$ such that $I \subseteq D$ and $|D| \leqslant 3|I| - 2 \leqslant 12L/\pi - 2$. We deduce that $d_G(x, y) \leqslant (12L/\pi - 2) + 1 \leqslant \lceil \alpha L \rceil$. ◀

Our Corollary 7 states that the restriction of MACS to well-distributed inputs admits a PTAS. The algorithm works as follows. Given a parameter $0 < \varepsilon \leqslant 1/2$, and an input (X, k) of MACS, we run Algorithm 2 on input (X, k', ε') for suitable values k' and ε' specified below. Next, we transform the set of additional disks obtained into a set of input disks that has roughly the same size while maintaining the connectivity of the solution. See Lemma 27 and Algorithm 3 for details. This algorithm naturally applies to pseudo-convex inputs (Corollary 10).

▶ **Lemma 27.** *Given an α-well-distributed input X and two finite sets $S \subseteq X$ and $S_{add} \subseteq \mathbb{R}^2$ such that $UDG(S \cup S_{add})$ is connected, there exists a set $S' \subseteq X$ of size at most $(22\alpha+4)|S_{add}|$ such that $UDG(S \cup S')$ is connected. Moreover, such a set can be computed in polynomial time.*

In the previous lemma, the set S_{add} is not supposed to be a set of additional disks as defined in Section 5.

Proof. (Lemma 27) Let us use the same notation as in the statement of Lemma 27. We prove how to build S' from S_{add} such that $|S'| \leqslant (22\alpha + 4)|S_{add}|$ while preserving connectivity.

Let Y be a connected component of S_{add}. We prove that we can find a set $Y' \subseteq X$ of input disks such that $|Y'| \leqslant (4 + 22\alpha)|Y|$ and $(S_{add} \setminus Y) \cup (S \cup Y')$ is connected. Removing Y might split the solution into several connected components $F_1, \ldots, F_s$. For each connected component F_i, pick one disk x_i in $F_i \cap X$ that intersects Y.

Step 1. Each additional disk y in Y is adjacent to at most 6 disks x_i. We can connect the corresponding connected component by using 20α disks of the input. Indeed, any two x_i and x_j adjacent to y has a Euclidean distance at most 4. Since X is well-distributed their distance in $UDG(X)$ is at most 4α. Then, we can find $\lceil 4\alpha - 1\rceil$ disks in X which connect x_i and x_j. In order to connect all the x_i that are adjacent to y, it is sufficient to repeat this operation 5 times, which asks at most 20α disks. We can perform this operation for each additional disk that was not already considered. Then, in total for this first step we need to use at most $20\alpha|Y|$ disks.

Step 2. During step 1, we may have connected some disks x_i, so that the number of connected components has decreased. The number of connected components is $s' \leqslant s$, each of them corresponds to a disk x_i, and without loss of generality we can assume that the corresponding indexes are such that $1 \leqslant i \leqslant s'$. Let T be a spanning tree on $UDG(Y)$. Without loss of generality, we can suppose that indexes i are such that the sequence $(x_1, \ldots, x_{s'})$ correspond to a T transversal. Note that after step 1, each x_i can be associated to a different y in Y. Then, we reconnect each x_i to x_{i+1} for $1 \leqslant i \leqslant s - 1$. If x_i and x_{i+1} are respectively associated to y_i and y_{i+1}, then $||x_i - x_{i+1}|| \leqslant 2 + 2d_T(y_i, y_{i+1})$ and thus $d_{UDG(X)}(x_i, x_{i+1}) \leqslant \lceil \alpha(2 + 2d_T(y_i, y_{i+1})\rceil$. Then, we can find $\lceil \alpha(2 + 2d_T(y_i, y_{i+1})\rceil - 1$ disks in X to connect x_i and x_{i+1}. In order to connect all x_i we need to use at most

$$\sum_{i=1}^{s'-1} \lceil \alpha(2 + 2d_T(y_i, y_{i+1})\rceil - 1 \leqslant 2(s' - 1)\alpha + 2\sum_{i=1}^{s'-1} d_T(y_i, y_{i+1})$$

input disks. Since the order corresponds to a T transversal, each edge is visited at most twice and then $\sum_{i=1}^{s'-1} d_T(y_i, y_{i+1}) \leqslant 2(|Y| - 1)$. Therefore the total number of disks that were added during this second step is bounded by $|Y|(4 + 2\alpha)$.

We proved that there exists a subset $Y' \subseteq X$ of size at most $(4 + 22\alpha)|Y|$ such that $(S \cup S_{add} \setminus Y) \cup Y'$ is connected. By doing so for each connected component of S_{add}, we get the result claimed. ◀

Algorithm 3: PTAS for MACS for well-distributed inputs

Input: X an α-well-distributed input, $k \geqslant 0$, $\varepsilon > 0$.
Output: A feasible solution to MACS(X, k).

Choose $\varepsilon' > 0$ and $k' \leqslant k$ such that $(1 - \varepsilon')(1 - 10(22\alpha + 4)\varepsilon') \geqslant (1 - \varepsilon)$ and $k'(1 + (22\alpha + 4)\varepsilon') = k$;
Let S, S_{add} be the solution of Algorithm 2 on input (X, k', ε') ;
Let S' be the set obtained from S_{add} by Lemma 27;
return $S \cup S'$;

Since $\varepsilon' = \Theta(\varepsilon/\alpha)$, the previous algorithm runs in polynomial time when ε and α are fixed constants.

▷ Claim 28. The solution returned by Algorithm 3 on input (X, k, ε) is a feasible solution to MACS(X, k) and covers an area at least $(1 - \varepsilon)\mathbf{OPT}(X, k)$.

In order to prove this result we need to state the following "stability" property over optimal solutions.

▶ **Lemma 29.** *Let $\eta < 1/2$. Then: $\boldsymbol{OPT}(X, k) \geqslant (1 - 10\eta) \cdot \boldsymbol{OPT}(X, k(1 + \eta))$.*

Proof. (Lemma 29) Let X be a set of points of the plane, k a positive integer and $\eta \leqslant 1/2$ a parameter. We prove a stronger result. Given any solution feasible solution S to MACS$(X, k(1 + \eta))$, there exists a subset S' of S that is a feasible solution to MACS(X, k) with value at least $(1 - 10\eta)\mathcal{A}(S)$. Obviously Lemma 29 follows when S is optimal. If $\mathcal{A}(S) \geqslant k/3$, then remove ηk disks from S without disconnecting S. For instance, consider a spanning tree on $UDG(S)$ and remove the nodes from the leaves to the root until you reach the desired size. Let S' denote the subset obtained.

$$\mathcal{A}(S') \geqslant \mathcal{A}(S) - \eta\pi k \geqslant (1 - 3\pi\eta)\mathcal{A}(S) \geqslant (1 - 10\eta)\mathcal{A}(S)$$

If $\mathcal{A}(S) < k/3$, let I be a maximal independent set in S. We have $|I|\pi = \mathcal{A}(I) \leqslant \mathcal{A}(S) < \dfrac{k}{3}$. According to claim 23, there exists a connected dominating set $I \subseteq D \subseteq S$ in S of size at most $3|I| - 2 < k/\pi < \dfrac{k}{3}$. Consider a set $H \subseteq S \setminus D$ of size $k - |D| > \frac{2}{3}k$ built by greedily adding a disk $h \in S \setminus (D \cup H)$ maximising the marginal area $\mathcal{A}(D \cup H \cup \{h\}) - \mathcal{A}(D \cup H)$. Since D is a connected dominating set, the set $S' := D \cup H$ is connected. Since all disk where added greedily in H, for all $H \in S \setminus S'$, we have

$$\mathcal{A}(S' \cup \{h\}) - \mathcal{A}(S')) \leqslant \frac{\mathcal{A}(S) - \mathcal{A}(D)}{|H|} \leqslant \frac{2\mathcal{A}(S)}{k}$$

.

By submodularity, we deduce that $\mathcal{A}(S) - \mathcal{A}(S') \leqslant \eta k \cdot \dfrac{2\mathcal{A}(S)}{3k}$. That implies $\mathcal{A}(S') \geqslant (1 - \frac{3}{2}\eta)\mathcal{A}(S)$. This concludes the proof of lemma 29. ◀

Remark that this proof is constructive and it is easy to check that finding S' from any given set S can be done in polynomial time.

Proof. (Claim 28) The solution output by Algorithm 2 on input (X, k', ε') verifies the following properties: $S \cup S_{add}$ is connected, the size of S and S_{add} are respectively upper-bounded by k' and $\varepsilon' k'$ and $\mathcal{A}(S) \geqslant (1 - \varepsilon')\mathbf{OPT}(X, k')$. Therefore, the set S' given by Lemma 27 has size at most $(22\alpha + 4)|S_{add}| \leqslant (22\alpha + 4)\varepsilon' k'$, and then $|S \cup S'| \leqslant k' + (22\alpha + 4)\varepsilon' k' \leqslant (1 + (22\alpha + 4)\varepsilon')k' = k$. Since $S \cup S'$ is connected, this set is a feasible solution to MACS(X, k).

Finally, from Lemma 29 with parameter $\eta = (22\alpha + 4)\varepsilon'$, we get that the area covered by this solution is

$$\begin{aligned}\mathcal{A}(S \cup S') &\geqslant \mathcal{A}(S) \geqslant (1-\varepsilon')\mathbf{OPT}(X,k') \geqslant (1-\varepsilon')(1-10\eta)\mathbf{OPT}(X,k'(1+\eta))\\ &\geqslant (1-\varepsilon')(1-10(22\alpha+4)\varepsilon')\mathbf{OPT}(X,k'(1+(22\alpha+4)\varepsilon'))\\ &\geqslant (1-\varepsilon)\mathbf{OPT}(X,k)\end{aligned}$$

which concludes the proof. ◀

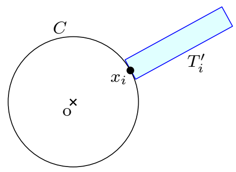
7 APX-hardness of Quad-Connected-Cover

▶ **Theorem 12.** Quad-connected-cover *is APX-hard.*

The reduction will be from the following problem.

3-set-cover. Given a set X of n elements, and its subsets $\mathcal{S} = \{S_1, \ldots, S_m\}$ such that $|S_i| \leq 3$ for $i = 1, \ldots, m$, compute a minimum size subset of $\mathcal{S}$ that covers X. 3-set-cover is APX-hard (due to the fact that minimum vertex cover on graphs with maximum degree 3 is APX-hard).

Proof. (Theorem 12) The proof is by a reduction from 3-set-cover to quad-connected-cover. In particular, given a set $X = \{x_1, \ldots, x_n\}$ and subsets $\S = \{S_1, \ldots, S_m\}$, we show how to construct, in polynomial time and for any parameter $\varepsilon < \frac{1}{6}$, a $(1+\varepsilon)$-approximation to 3-set-cover from a $\left(1 - \frac{\varepsilon}{6}\right)$-approximation to the quad-connected-cover.



Map the n points of X to n points placed uniformly on a circle C of unit area centered at the origin o; we will use the notation x_i for these points as well. Our set $\mathcal{T}$ will consist of convex quadrilaterals of two types:

- **center-quads.** These are, for each set $S_j \in \S$, the quadrilateral $T_j = \text{convexhull}\,(S_j \cup \{o\})$.
- **side-quads.** For each element x_i, let T_i' be the rectangle with width $\frac{1}{2n}$, length $4n$, containing x_i and tangent to C

Note that every pair of center-quads intersect (namely, at o), no two side-quads intersect, and T_j intersects T_i' if and only if $x_i \in S_j$. The area of the union of the center-quads is at most 1, and the area of each side-quad is 2.

Let s be the size of an optimal set-cover for X and $\S$. Let $\mathcal{T}'$ be a $(1 - \frac{\varepsilon}{6})$-approximate solution to the quad-connected-cover problem on the set $\{T_1, \ldots, T_m, T_1', \ldots, T_n'\}$ with $k = n + s$. Observe that to maintain connectivity of the intersection graph of $\mathcal{T}'$, if a point x_i is covered by a side-quad of $\mathcal{T}'$, it must also be covered by some center-quad of $\mathcal{T}'$, as a side-quad only intersects center-quads.

One possible solution consists of picking the s center-quads of the set-cover, and all the n side-quads to get the total area of at least $2n$; in particular, an optimal solution has value at least $2n$. Thus the area of the union of the quadrilaterals in $\mathcal{T}'$ is at least $\left(1 - \frac{\varepsilon}{6}\right) \cdot 2n$. This implies that $\mathcal{T}'$ leaves at most $\frac{\varepsilon n}{6}$ elements of X uncovered by center-quads; otherwise at least $\frac{\varepsilon n}{6} + 1$ side-quads are not picked, and so the area covered by $\mathcal{T}'$ can only be $1 + 2\left(n - \frac{\varepsilon n}{6} - 1\right) \leq \left(1 - \frac{\varepsilon}{6}\right) \cdot 2n - 1$. Thus, out of the $n + s$ quadrilaterals in $\mathcal{T}'$, at least $n - \frac{\varepsilon n}{6}$ side-quads are present, and at most $(n + s) - \left(n - \frac{\varepsilon n}{6}\right) = s + \frac{\varepsilon n}{6}$ center-quads are present. Thus one can pick arbitrarily one set for each uncovered point to construct a set cover for X of size at most $\left(s + \frac{\varepsilon n}{6}\right) + \frac{\varepsilon n}{6} \leq s + \frac{2\varepsilon}{6} \cdot 3s \leq (1 + \varepsilon) \cdot s$, where the first inequality follows from the fact that $s \geq \frac{n}{3}$. This completes the proof.

◀

One can imagine that finding a more specific reduction from APX-hard geometric covering problems in [18] for instance suggests that the problem QUAD-CONNECTED-COVER remains APX-hard even when the quadrilaterals are triangles with area arbitrarily close to one.

References

1 Sanjeev Arora. Polynomial time approximation schemes for euclidean traveling salesman and other geometric problems. *J. ACM*, 45(5):753–782, September 1998. URL: `http://doi.acm.org/10.1145/290179.290180`, `doi:10.1145/290179.290180`.

2 Yair Bartal, Lee-Ad Gottlieb, and Robert Krauthgamer. The traveling salesman problem: low-dimensionality implies a polynomial time approximation scheme. In Howard J. Karloff and Toniann Pitassi, editors, *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 663–672. ACM, 2012. URL: `http://doi.acm.org/10.1145/2213977.2214038`, `doi:10.1145/2213977.2214038`.

3 Gruia Călinescu, Chandra Chekuri, Martin Pál, and Jan Vondrák. Maximizing a monotone submodular function subject to a matroid constraint. *SIAM J. Comput.*, 40(6):1740–1766, 2011.

4 Erin W. Chambers, Sándor P. Fekete, Hella-Franziska Hoffmann, Dimitri Marinakis, Joseph S.B. Mitchell, Venkatesh Srinivasan, Ulrike Stege, and Sue Whitesides. Connecting a set of circles with minimum sum of radii. *Computational Geometry*, 68(1-3):62–76, January 1991. special issue in memory of Ferran Hurtado.

5 T.-H. Hubert Chan and Shaofeng H.-C. Jiang. Reducing curse of dimensionality: Improved PTAS for TSP (with neighborhoods) in doubling metrics. In Robert Krauthgamer, editor, *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 754–765. SIAM, 2016. URL: `http://dx.doi.org/10.1137/1.9781611974331.ch54`, `doi:10.1137/1.9781611974331.ch54`.

6 Steven Chaplick, Minati De, Alexander Ravsky, and Joachim Spoerhase. Approximation schemes for geometric coverage problems. In *ESA*, volume 112 of *LIPIcs*, pages 17:1–17:15. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2018.

7 Chandra Chekuri, Jan Vondrák, and Rico Zenklusen. Submodular function maximization via the multilinear relaxation and contention resolution schemes. *SIAM J. Comput.*, 43(6):1831–1879, 2014.

8 Brent N. Clark, Charles J. Colbourn, and David S. Johnson. Unit disk graphs. *Discrete Math.*, 86(1-3):165–177, January 1991. URL: `http://dx.doi.org/10.1016/0012-365X(90)90358-O`, `doi:10.1016/0012-365X(90)90358-O`.

9 Adrian Dumitrescu and Csaba D. Tóth. Constant-factor approximation for TSP with disks. *CoRR*, abs/1506.07903, 2015. URL: `http://arxiv.org/abs/1506.07903`.

10 Khaled M. Elbassioni, Aleksei V. Fishkin, and René Sitters. On approximating the TSP with intersecting neighborhoods. In Tetsuo Asano, editor, *Algorithms and Computation, 17th International Symposium, ISAAC 2006, Kolkata, India, December 18-20, 2006, Proceedings*, volume 4288 of *Lecture Notes in Computer Science*, pages 213–222. Springer, 2006. URL: `http://dx.doi.org/10.1007/11940128_23`, `doi:10.1007/11940128_23`.

11 Khaled M. Elbassioni, Aleksei V. Fishkin, and René Sitters. Approximation algorithms for the euclidean traveling salesman problem with discrete and continuous neighborhoods.

Int. J. Comput. Geometry Appl., 19(2):173–193, 2009. URL: `http://dx.doi.org/10.1142/S0218195909002897`, `doi:10.1142/S0218195909002897`.

12 Corinne Feremans and Alexander Grigoriev. Approximation schemes for the generalized geometric problems with geographic clustering. In *(Informal) Proceedings of the 21st European Workshop on Computational Geometry, Eindhoven, The Netherlands, March 9-11, 2005*, pages 101–102. Technische Universiteit Eindhoven, 2005. URL: `http://www.win.tue.nl/EWCG2005/Proceedings/26.pdf`.

13 Yuval Filmus and Justin Ward. Monotone submodular maximization over a matroid via non-oblivious local search. *SIAM J. Comput.*, 43(2):514–542, 2014. URL: `https://doi.org/10.1137/130920277`, `doi:10.1137/130920277`.

14 Stefan Funke, Alex Kesselman, Fabian Kuhn, Zvi Lotker, and Michael Segal. Improved approximation algorithms for connected sensor cover. *Wirel. Netw.*, 13(2):153–164, April 2007. URL: `http://dx.doi.org/10.1007/s11276-006-3724-9`, `doi:10.1007/s11276-006-3724-9`.

15 Michael M. Garey. and David S. Johnson. The rectilinear steiner tree problem is NP-complete. *SIAM Journal on Applied Mathematics*, 32(4):826–834, 1977. URL: `https://doi.org/10.1137/0132071`, `arXiv:https://doi.org/10.1137/0132071`, `doi:10.1137/0132071`.

16 L. A. Wolsey G.L. Nemhauser and M.L. Fisher. An analysis of approximations for maximizing submodular set functions - i. *Mathematical Programming*, 14(1):265–294, 1978.

17 Himanshu Gupta, Zongheng Zhou, Samir R. Das, and Quinyi Gu. Connected sensor cover: Self-organization of sensor networks for efficient query execution. *IEEE/ACM Trans. Netw.*, 14(1):55–67, February 2006. URL: `http://dx.doi.org/10.1109/TNET.2005.863478`, `doi:10.1109/TNET.2005.863478`.

18 Sariel Har-Peled. Being fat and friendly is not enough. *CoRR*, abs/0908.2369, 2009. URL: `http://arxiv.org/abs/0908.2369`, `arXiv:0908.2369`.

19 Dorit S. Hochbaum and Anu Pathria. Node-optimal connected k-subgraphs, 1994.

20 Harry B. Hunt, III, Madhav V. Marathe, Venkatesh Radhakrishnan, S. S. Ravi, Daniel J. Rosenkrantz, and Richard E. Stearns. NC-approximation schemes for NP- and PSPACE-hard problems for geometric graphs, 1997.

21 Koushik Kar and Suman Banerjee. Node placement for connected coverage in sensor networks, 2003.

22 Samir Khuller, Manish Purohit, and Kanthi K. Sarpatwar. Analyzing the optimal neighborhood: Algorithms for budgeted and partial connected dominating set problems. In *Proceedings of the Twenty-fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '14, pages 1702–1713, Philadelphia, PA, USA, 2014. Society for Industrial and Applied Mathematics. URL: `http://dl.acm.org/citation.cfm?id=2634074.2634197`.

23 Ariel Kulik, Hadas Shachnai, and Tami Tamir. Maximizing submodular set functions subject to multiple linear constraints. In *SODA*, pages 545–554. SIAM, 2009.

24 Tung-Wei Kuo, Kate Ching-Ju Lin, and Ming-Jer Tsai. Maximizing submodular set function with connectivity constraint: Theory and application to networks. *IEEE/ACM Trans. Netw.*, 23(2):533–546, April 2015. URL: `http://dx.doi.org/10.1109/TNET.2014.2301816`, `doi:10.1109/TNET.2014.2301816`.

25 Jon Lee, Maxim Sviridenko, and Jan Vondrák. Submodular maximization over multiple matroids via generalized exchange properties. *Math. Oper. Res.*, 35(4):795–806, 2010.

26 M. V. Marathe, H. Breu, H. B. Hunt III, S. S. Ravi, and D. J. Rosenkrantz. Simple heuristics for unit disk graphs. *NETWORKS*, 1995.

27 Tomomi Matsui. Approximation algorithms for maximum independent set problems and fractional coloring problems on unit disk graphs. In Jin Akiyama, Mikio Kano, and Masatsugu Urabe, editors, *Discrete and Computational Geometry*, pages 194–200, Berlin, Heidelberg, 2000. Springer Berlin Heidelberg.

28 Joseph S. B. Mitchell. Guillotine subdivisions approximate polygonal subdivisions: A simple polynomial-time approximation scheme for geometric tsp, k-mst, and related prob-

lems. *SIAM J. Comput.*, 28(4):1298–1309, March 1999. URL: `https://doi.org/10.1137/S0097539796309764`, `doi:10.1137/S0097539796309764`.

29 Joseph S. B. Mitchell. Guillotine subdivisions approximate polygonal subdivisions: A simple polynomial-time approximation scheme for geometric TSP, k-MST, and related problems. *SIAM J. Comput.*, 28:1298–1309, 1999.

30 Giri Narasimhan and Michiel Smid. *Geometric Spanner Networks*. Cambridge University Press, New York, NY, USA, 2007.

31 Yuval Rabani and Gabriel Scalosub. Bicriteria approximation tradeoff for the node-cost budget problem. *ACM Trans. Algorithms*, 5(2):19:1–19:14, 2009.

32 Sophie Theresa Spirkl. The guillotine approach for TSP with neighborhoods revisited. *CoRR*, abs/1312.0378, 2013. URL: `http://arxiv.org/abs/1312.0378`.

33 Fred van Nijnatten, René Sitters, Gerhard J. Woeginger, Alexander Wolff, and Mark de Berg. The traveling salesman problem under squared euclidean distances. In Jean-Yves Marion and Thomas Schwentick, editors, *27th International Symposium on Theoretical Aspects of Computer Science, STACS 2010, March 4-6, 2010, Nancy, France*, volume 5 of *LIPIcs*, pages 239–250. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2010. URL: `http://dx.doi.org/10.4230/LIPIcs.STACS.2010.2458`, `doi:10.4230/LIPIcs.STACS.2010.2458`.

34 Fabio Vandin, Eli Upfal, and Benjamin J. Raphael. Algorithms for detecting significantly mutated pathways in cancer. *Journal of Computational Biology*, 18(3):507–522, 2011.

35 L. Wolsey. Maximising real-valued submodular functions: primal and dual heuristics for location problems. *Mathematics of Operations Research*, 7(3):410–425, 1982.

A A simple PTAS when connectivity is not required

We explain here how to obtain a PTAS when connectivity is not required. Given a set X of n unit-disks in the plane and an integer k, let us denote by OPT a subset $S \subseteq X$ of size at most k, that maximizes the value $\mathcal{A}(S)$. We prove that for any $\varepsilon > 0$, there exists a algorithm that finds in time $n^{O_\varepsilon(1)}$ a set S of size at most k such that $\mathcal{A}(S) \geqslant (1-\varepsilon)\mathcal{A}(OPT)$.

The principle is the following. We start with a bounding box of size $O(n) \times O(n)$ that contains all input disks and partition it into a set of squares of constant size $\varepsilon^{-1} \times \varepsilon^{-1}$ similarly as in Section 5.1. Then we remove all disks that intersect a line of the grid to get a input X' where any two disks from distinct square can not overlap anymore . We can show using the same shifting argument (see Claim 22) that there exists at least one position of the grid, among $O(n)$ possible choices such that $\mathcal{A}(OPT \cap X') \geqslant (1 - O(\varepsilon))\mathcal{A}(OPT)$.

Now, we run a dynamic program to compute an solution of size at most k in X' that is near-optimal. Let $C_1, \dots, C_{n'}$ be the set of squares induced by the grid, arbitrarily ordered. Notice that $n' = O(n^2)$. Each DP-cell is described by two integers $i \leqslant n'$ and $k' \leqslant k$. The value $A(i, k')$ associated with a cell (i, k') corresponds to the maximum area that can be covered by a subset $S \subseteq X'$ of size k' that contains only disks from squares C_j with $j \leqslant i$ and such that $|S \cap C_j| \leqslant \varepsilon^{-3}$. We initialize $A(i, 0) = A(0, k') = 0$ for all i, k' and we have $A(i+1, k') = \max\left\{A(i, k' - k'') + \mathcal{A}(S'') \mid S'' \subseteq X' \cap C_{i+1}, |S''| = k'' \leqslant \varepsilon^{-3}\right\}$. We output the value $A(n', k)$, that can be computed in time $O(n^{1/\varepsilon^3} \cdot n \cdot k)$. The near-optimality of the solution returned is guaranteed by following "structural" lemma.

▶ **Lemma 30.** *Given an input (X, k), let $OPT \subseteq X$ be at set of size k maximising the area covered, and consider a grid chosen uniformly at random. Then, there exists a set $OPT' \subseteq OPT$ such that*

1. *no disks in OPT' intersect the grid*
2. *for all square C we have $|OPT' \cap C| \leqslant \varepsilon^{-3}$*
3. *with probability at least $1/3$, we have $\mathcal{A}(OPT') \geqslant (1 - O(\varepsilon))\mathcal{A}(OPT)$*

The proof is almost identical as the proof of Lemma 24. We show that if a set $X_C = OPT \cap C$ of unit-disks lies into a square of size $\varepsilon^{-1} \times \varepsilon^{-1}$, then there exists a subset $X'_C \subseteq X_C$ of size at most ε^{-3} that covers an area $\mathcal{A}(X'_C) \geqslant (1 - O(\varepsilon))\mathcal{A}(X_C)$. This set can even be built using a simple greedy procedure : starting with $X'_C = \emptyset$, add in X'_C the unpicked disk that maximises the uncovered area until $|X'_C| \geqslant \varepsilon^{-3}$. At the end, any disk $x \in X_C \setminus X'_C$ can only increase the area by at most $\mathcal{A}(X'_C \cup \{x\}) - \mathcal{A}(X'_C) \leqslant \varepsilon^{-2}/(\pi \cdot \varepsilon^{-3}) = O(\varepsilon)$. In particular, any point of the plane covered by X_C is at distance at most $1 + \beta$ from a point in X'_C, where $\beta = O(\varepsilon)$. It follows from Claim 25 that $\mathcal{A}(X'_C) \geqslant (1 - O(\varepsilon))\mathcal{A}(X_C)$.

B An alternative proof of Theorem 5 with m-guillotine subdivisions

Let $\mathcal{R} = \{R_1, \ldots, R_n\}$ be a set of n connected regions in the plane. We assume that the union $\bigcup_i R_i$ of the regions is connected. Given a positive integer k, our goal is to compute a subset $\mathcal{R}^* \subseteq \mathcal{R}$ of cardinality $|\mathcal{R}^*| \leq k$ such that the union $\bigcup_{R_i \in \mathcal{R}^*} R_i$ is connected and has maximum possible area. This problem is the *Maximum Area Connected Subset* (MACS) problem.

We begin with the case in which $\mathcal{R}$ is a set of n unit-radius disks, specified by the set, $X \subseteq \Re^2$, of their center points. We let $\mathbf{OPT}(X, k)$ denote the area of an optimal solution of MACS on the set of unit-radius disks centered at X.

In the MACS *with augmentation* problem, we allow a small number (at most εk) of additional disks to be computed, which, together with the subset of disks chosen, yield a connected union. We now provide a proof of the following theorem, which shows that the MACS with augmentation, on a set of unit-radius disks, has a PTAS.

▶ **Theorem 31.** *Let $\varepsilon > 0$ be fixed. Given a set $X \subseteq \Re^2$ of points and a positive integer k, there is a deterministic algorithm that computes, in time $n^{O(\varepsilon^{-1})}$, a subset $S \subseteq X$ of size at most k and a set $S_{add} \subseteq \Re^2$ of at most εk points, such that UDG($S \cup S_{add}$) is connected, and the area covered by the unit disks centered at S is at least $(1 - \varepsilon)\mathbf{OPT}(X, k)$.*

Let $X^* \subseteq X$ be an optimal (for MACS) subset of $k = |X^*|$ centers of unit-radius disks, the union of which is connected and of maximum possible area $\mathbf{OPT}(X, k)$. Let $Z = \{(x, y) : x = i/2, y = j/2$, for some integers $i, j\}$ be the set of points in the plane having half-integral coordinates. Our algorithm will select augmentation disks centered at points $S_{add} \subset Z$.

Let $\mathcal{Q} = \{Q_1, \ldots, Q_k\}$ be the set of k axis-aligned bounding squares of the unit-radius disks centered at the points X^*. Let E be the set of $4k$ axis-parallel line segments (each of length 2) that bound the squares $\mathcal{Q}$. The union of the segments E is connected, since the union of the (equal-size) squares $\mathcal{Q}$ is connected.

As we know from the standard m-guillotine structure theorem (Theorem 3.3 of [29]), the edge set E can be made to be m-guillotine[6], for any positive integer m, by the addition of m-spans (horizontal/vertical line segments) whose total length is at most $O(k/m)$; with the appropriate choice of $m = \Theta(1/\varepsilon)$, the total length of all added m-spans is thus at most εk. Further, the m-spans can be chosen to lie along horizontal/vertical cut lines that pass through edges of squares in $\mathcal{Q}$. (This follows immediately from the proof of the existence of a favorable cut [29], using the fact that the edge set E is axis-parallel, yielding piecewise-constant cost functions f and g.)

[6] In Section B.2 we review the definition of m-guillotine from [29].

First, if an m-span has length less than 2, then any square that the m-span intersects is already among the squares that intersect the cut that are accounted for among the first m or last m edges crossed by the cut. Thus, the total number of squares intersected by the cut is at most $2m$, and we can afford to ignore this short m-span, since our goal is to have $O(m)$ information specified across a partitioning cut. Thus, we can assume that all m-span bridges are of length at least 2.

Now, associated with each (remaining) horizontal/vertical m-span segment, ab, we define an *m-span rectangle*, which is axis-parallel, centered on ab, of width 2; i.e., if ab is vertical, with $a = (x_a, y_a)$ and $b = (x_b = x_a, y_b)$, the corresponding m-span rectangle is $[x_a - 1, x_a + 1] \times [y_a, y_b]$. It is readily seen that the m-span rectangle associated with ab is covered by a set of $O(|ab|)$ unit-radius disks with centers at half-integral points Z; thus, the set of all m-span rectangles is covered by a set of $O(\varepsilon k)$ augmentation disks, centered at points of Z. Refer to Figure 10. The purpose of the m-span rectangle is to allow us to decouple the subproblems on each side of a cut: any unit-radius disk centered at a point of X to the right of a vertical cut contributes nothing to the union of disks centered at points left of the cut, unless it is one of the $O(m)$ specified disks crossing the cut, since, if it crosses an m-span segment on the cut, the m-span rectangle fully covers it, so that the augmentation disks fully cover it as well.

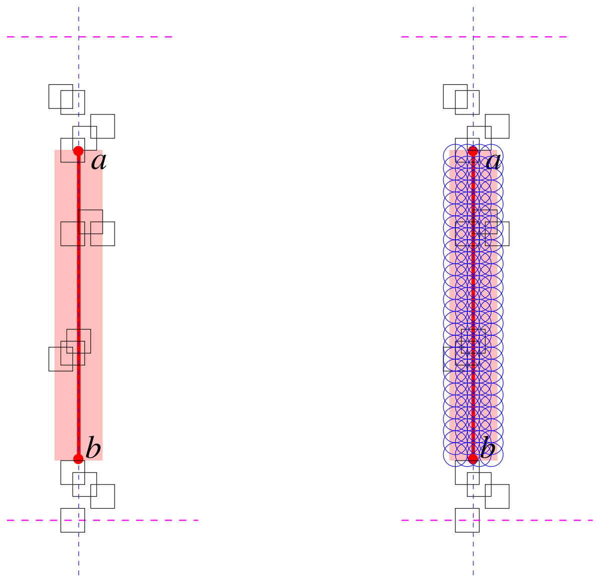


Figure 10 Left: A vertical cut, the m-span segment ab (for $m = 5$), and the associated m-span rectangle. Right: The set of augmentation unit-radius disks centered at half-integral points of Z within the m-span rectangle; the augmentation disks cover the m-span rectangle.

Let U be a finite set of unit-radius disks centered at points of $X \cup Z$. We now define the notion of the set U being "m-guillotine". An axis-parallel cut line ℓ is *m-good with respect to the set U of unit-radius disks and an axis-aligned rectangle window W* if (1) $\ell \cap W$ intersects at most $3m$ disks of U that are centered at points of X; and (2) U includes all unit-radius disks centered at points of Z that lie within the m-span rectangle associated with the m-span of ℓ with respect to the edge set E of segments bounding the axis-aligned bounding squares of unit-radius disks of U that are centered at points of X. An m-good cut has a succinct specification of those disks of U that are intersected by the cut: $O(m)$ disks of U centered at points of X, together with a single m-span segment (rectangle), which specifies the set of all augmentation disks (centered at half-integral points Z) within U that intersect the cut.

We say that a set U of unit-radius disks centered at points of $X \cup Z$ *satisfies the m-guillotine property with respect to (axis-aligned) rectangle* W if either (1) no disk of U lies (completely) inside W; or (2) there exists an axis-parallel cut line ℓ that is m-good with respect to U and W, such that ℓ splits W into W_1 and W_2, and, recursively, U satisfies the m-guillotine property with respect to W_1 and with respect to W_2. We say that U *satisfies the m-guillotine property* if U satisfies the m-guillotine property with respect to the axis-aligned bounding rectangle of U.

The following key structural lemma allows us to prove our main claim, since it shows that an arbitrary set of input disks (e.g., the optimal set, centered at points $X^* \subseteq X$) can be converted to a covering set of disks that satisfies the m-guillotine property, with only a small (factor $(1 + O(1/m))$) increase in the total number of disks.

▶ **Lemma 32.** *For any positive integer m and finite set U of k unit-radius disks centered at points of X, there exists a set U' of unit-radius disks centered at points of $X \cup Z$ such that U' satisfies the m-guillotine property, the union of the disks U' covers the union of the disks U, and $|U'| \leq (1 + O(1/m))k$.*

Proof. As noted above, the standard m-guillotine structure theorem (Theorem 3.3 of [29]) implies that the edge set E, of edges bounding the bounding squares of the k unit-radius disks U, can be made to be m-guillotine through the addition of m-spans (horizontal/vertical line segments) whose total length is at most $O(k/m)$.

Now, each recursive (axis-parallel) cut, within a rectangular window W, in the associated m-guillotine hierarchy gives rise, potentially, to an m-span segment ab, which has an associated m-span rectangle, which is covered completely by the $O(|ab|)$ unit-radius augmentation disks centered at half-integral points Z within the m-span rectangle (refer to Figure 10); these augmentation disks are included in the set U'. Any disk of U that lies interior to W and has its bounding square crossed by the m-span segment ab must lie fully within the associated m-span rectangle; thus, it is covered by the augmentation disks and is not included in the set U'. All other disks of U are included in U'; thus, the union of disks in U' covers the union of disks in U. Further, the cut now intersects at most $O(m)$ disks of U', and the resulting set U', consisting of a subset of disks of U and a set of $O(k/m)$ augmentation disks, has the m-guillotine property and has cardinality $|U'| \leq (1 + O(1/m))k$. ◀

To complete the proof of Theorem 31, we now provide a dynamic programming algorithm to compute, for given positive integers k, k', and m, and an input set of points X for which UDG(X) is connected, a set U of unit-radius disks centered at points of $X \cup Z$ such that (i) U satisfies the m-guillotine property, (ii) U has at most k disks centered at X and at most k' centered at points of Z, and (iii) the union of the disks of U has the maximum possible area among all sets of disks satisfying (i) and (ii). The application of this algorithm, with $k' = (1 + O(1/m))k$, yields the claimed PTAS, since we know, by Lemma 32 applied to an MACS-optimal set U of k disks, that, among the m-guillotine sets of disks over which the dynamic program optimizes, there is such a set that includes a MACS-optimal set of k disks.

The dynamic program proceeds in much the same way that similar algorithms are used to compute optimal m-guillotine subdivisions for TSP and other problems [29, 32, 5, 10, 11, 9, 2, 33, 12]. Subproblems will be specified by axis-aligned rectangles, W, the coordinates of which come from the left/right/top/bottom coordinates of the n input disks; specifically, we let $x_1 \leq x_2 \leq \cdots \leq x_{2n}$ and $y_1 \leq y_2 \leq \cdots \leq y_{2n}$ denotes the sorted coordinates. The optimization of a subproblem is to select an axis-parallel cut, partitioning the rectangular window into two, along with the $O(m)$ data associated with the cut, including the $O(m)$

unit-radius disks centered at points of X that intersect the cut, the connection requirements ($O(1)$, for fixed m) for the two new subproblems, and the defining coordinates of an m-span rectangle (if any), which succinctly encodes the set of augmentation disks centered at half-integral points of Z within the m-span rectangle. Overall, the approximation to the MACS with augmentation will be given by the optimal solution of a subproblem associated with a root rectangle, W_0; there are only $O(n^4)$ possible choices of W_0, and these include the axis-aligned bounding box of an (exact) MACS-optimal set of k disks.

A subproblem is specified by a rectangle $W \subseteq W_0$, with coordinates among the x_i's and y_j's, together with a specification of certain *boundary information*, $\mathcal{B}$, that gives the information necessary to describe how the solution inside W interfaces with the solution outside of the window W. This information $\mathcal{B}$ includes the following:

(a) For each of the four sides of W, we specify at most $2m$ unit-radius disks, centered at input points X, that intersect the side. Additionally, each side can have one "bridge" (m-span) segment specified, which defines the associated m-span rectangle, with the corresponding set of all unit-radius (augmentation) disks centered at points of Z that lie within the rectangle. There are $n^{O(m)}$ choices for this information.

(b) We specify a required "connectivity constraint" within W. In particular, we indicate which subsets of the $O(m)$ boundary elements (disks centered at points of X, and clusters of augmentation disks covering the at most four m-span rectangles on the boundary of W) are required to be connected within W. (If there are no boundary elements associated with W (meaning that $W = W_0$ is one of the choices of a root rectangle), then the connectivity constraint is simply that all disks within W must have a connected union.) Since the number of different partitions of the $O(m)$ boundary elements is purely a function of m, considered to be a constant, there are only a constant number of choices of these connectivity constraints.

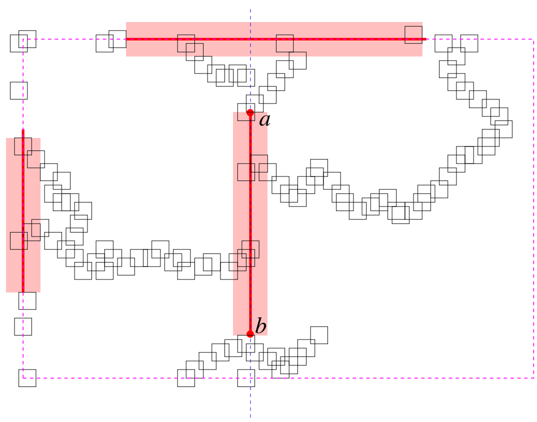


Figure 11 A subproblem in the dynamic program that optimizes over sets of disks that satisfy the m-guillotine property.

Let $f(W, \mathcal{B}, k, k')$ denote the value of a subproblem, the maximum area of W intersected with the union of the disks in a set U of unit-radius disks that satisfy the following properties: (a) U satisfied the m-guillotine property with respect to W; (b) U consists of k unit-radius disks centered at input points X and at most k' unit-radius (augmentation) disks centered at points of Z; and, (c) U satisfies the specified boundary information $\mathcal{B}$, including the required connectivity. In order to tabulate values of f, we build up the solutions bottom-up, as usual, starting with subproblems that are trivial, and tabulating values corresponding to window

W, defined by rectangle $[x_i, x_{i'}] \times [y_j, y_{j'}]$ with $i < i'$ and $j < j'$, in order of increasing values of $(i' - i)$ and $(j' - j)$, with each choice of boundary information, k, and k'. The value $f(W, \mathcal{B}, k, k')$ is computed recursively:

$$f(W, \mathcal{B}, k, k') = \max_{\xi, \mathcal{B}_\xi} \{ f(W_1, \mathcal{B}_1, k_1, k_1') + f(W_2, \mathcal{B}_2, k_2, k_2') \},$$

where ξ is an axis-parallel cut (at one of the discrete coordinates x_i, y_j), $\mathcal{B}_\xi$ is the boundary information across the cut ξ (including specification of k_ξ unit-radius disks, centered at points of X, that cross ξ and (possibly) specification of an m-span rectangle that contains k'_ξ half-integral points of Z where augmentation disks are centered), W_1 and W_2 are the subrectangles of W obtained when making cut ξ, $\mathcal{B}_1$ and $\mathcal{B}_2$ are boundary information consistent with $\mathcal{B}$ and $\mathcal{B}_\xi$, k_1 and k_2 satisfy $k = k_1 + k_2 + k_\xi$, and k_1' and k_2' satisfy $k' = k_1' + k_2' + k_\xi'$.

The overall solution to the problem is given by $f(W_0, \mathcal{B}_0, k, k')$ for W_0 chosen to be one of the $O(n^4)$ possible root rectangles, $\mathcal{B}_0$ specifying no crossed disks or augmentation disks for the boundary of W_0 (so that all disks are interior to W_0), k equal to the input parameter of the MACS instance, $k' = ck/m$ for a fixed constant c, and connectivity specifying that the union of all disks interior to W_0 must be connected.

The number of subproblems is $n^{O(m)}$, and the evaluation of $f(W, \mathcal{B}, k, k')$ for any one subproblem requires time $n^{O(m)}$ to optimize over all choices of cuts and boundary specifications. Thus, the overall running time is $n^{O(m)}$, which is $n^{O(1/\varepsilon)}$, with the choice of $m = O(1/\varepsilon)$.

B.1 Extensions

B.1.1 Cases in Which No Augmentation Is Needed

In the case of regions $\mathcal{R}$ that are unit-radius disks, we used augmentation disks (centered at points of Z within an m-span rectangle) in order to decouple the subproblems on each side of a cut, doing so with a specification of $O(1)$ information (the coordinates of the rectangle). We were able to afford to add the augmentation disks, staying within $O(k/m) = O(\varepsilon k)$ such disks, because the total length of all m-span segments ab needed to convert the edge set of an arbitrary set of k bounding squares to be m-guillotine is only $O(k/m)$, implying that only $O(k/m)$ augmentation disks are needed to make any connected set of disks satisfy the m-guillotine property.

If the set of input unit-radius disks is such that for any potential m-span segment ab there is a path in the UDG(X) from a to b of length at most $O(|ab|)$, then a shortest path from a to b within the UDG(X) can serve to decouple the subproblems across a cut, in place of the augmentation disks within an m-span rectangle. This property is implied by the assumption that X is "α-well-distributed", for a fixed α [30].

B.1.2 More General Regions

For a given set $\mathcal{R} = \{R_1, \ldots, R_n\}$ of n connected regions in the plane, with a connected union $\bigcup_i R_i$, we now assume that each region R_i has an axis-aligned bounding box, $BB(R_i)$, whose aspect ratio is at most ρ; i.e., the ratio of the length of the longer side of $BB(R_i)$ to the length of the shorter side of $BB(R_i)$ is at most ρ. Further, we assume that the sizes of the regions R_i are all about the same, within a constant factor; more precisely, we assume that the ratio $diam(R_i)/diam(R_j)$ is bounded by a constant.

Then, if we allow at most (εk) augmentation disks of size $\Theta(\max_i diam(R_i))$, then the PTAS we described for regions that are unit-radius disks generalizes immediately to this case. In particular, the set E of edges of bounding boxes of the regions R_i can be made to be m-guillotine with the addition of m-span segments of lengths totalling $O((1/m)\sum_i diam(R_i)) = O((k/m)\max_i diam(R_i))$. Each of the m-span segments ab yields an m-span rectangle, of width $\max_i diam(R_i)$, centered on it, which can be covered by $O(|ab|/\max_i diam(R_i))$ augmentation disks of size $\max_i diam(R_i)$. Thus, using only $O(\varepsilon k)$ augmentation disks, an optimal set of regions, with area-maximizing connected union, can be made to have an m-guillotine property. The rest of the dynamic programming algorithm goes through to optimize the area of a set of regions whose union is connected.

B.2 useful definitions and facts

We review some definitions and facts from [29]. Let G be an embedding of a planar graph, and let L denote the total Euclidean length of its edges, E. We can assume (without loss of generality) that G is restricted to the unit square, B (i.e., $E \subset int(B)$).

Consider an axis-aligned rectangle W (a *window*) with $W \subseteq B$ and with corners at grid points within the N-by-N grid centered on a vertex c_0 (for one choice of possible c_0). Rectangle W will correspond to a subproblem in a dynamic programming algorithm. Let ℓ be an axis-parallel line, through grid points, intersecting W. We refer to ℓ as a *cut* for W. We will refer to a *root window*, W_0, which is a window that is hypothesized to be the minimal enclosing bounding box of an optimal solution (e.g., of the disks in an optimal solution to MACS). All windows W of interest will then be subwindows of W_0. Clearly, there are only a polynomial number ($O(n^4)$) of possible choices for W_0; we can afford to try each one.

The intersection, $\ell \cap (E \cap int(W))$, of a cut ℓ with $E \cap int(W)$ (the restriction of E to the window W) consists of a (possibly empty) set of subsegments (possibly singleton points) of ℓ. Let ξ be the number of endpoints of such subsegments along ℓ, and let the points be denoted by $p_1, \ldots, p_\xi$, in order along ℓ. For a positive integer m, we define the *m-span*, $\sigma_m(\ell)$, of ℓ (with respect to W) as follows. If $\xi \leq 2(m-1)$, then $\sigma_m(\ell) = \emptyset$; otherwise, $\sigma_m(\ell)$ is defined to be the (possibly zero-length) line segment, $p_m p_{\xi-m+1}$, joining the mth endpoint, p_m, with the mth-from-the-last endpoints, $p_{\xi-m+1}$. Line ℓ is an *m-good cut with respect to W and E* if $\sigma_m(\ell) \subseteq E$. (In particular, if $\xi \leq 2(m-1)$, then ℓ is trivially an m-good cut.)

We now say that E satisfies the *m-guillotine property with respect to window W* if either (1) no edge of E lies (completely) interior to W; or (2) there exists a cut ℓ, that is m-good with respect to W and E, such that ℓ splits W into W_1 and W_2, and, recursively, E satisfies the m-guillotine property with respect to both W_1 and W_2.

We say that a point $p \in W$ is *m-dark with respect to horizontal cuts of W* if the vertical rays going upwards/downwards from p each cross at least m edges of E before reaching the boundary of W. As in [29], the length of the m-dark portion of a cut is the "chargeable" length of the cut that is chargeable to the lengths of the m layers of E on each side of the cut that become "exposed" after the cut.

Given an edge set E of a connected planar graph G, if E is not already satisfying the m-guillotine property with respect to W_0, then the following lemma of [29] (reproduced here for completeness) shows that there exists a "favorable cut" for which we can afford to charge off (to the edges of E) the construction of any m-span that must be added to E in order to make the cut m-good with respect to W and E. A cut ℓ is said to be *favorable* if its chargeable length (i.e., length of its m-dark portion) is at least as large as the cost of ℓ ($|\sigma_m(\ell)|$).

▶ **Lemma 33.** *[from [29]] For any G and any window W, there is a favorable cut.*

We show that there must be a favorable cut that is either horizontal or vertical.

Let $f(x)$ denote the "cost" of the vertical line, ℓ_x, through x, where "cost" means the sum of the lengths of the m-span; thus, $f(x) = |\sigma_m(\ell_x)|$.

Then, $A_x = \int_0^1 f(x)dx$ is simply the area, $A_x^{(m)} = \int_0^1 |\sigma_m(\ell_x)|dx$, of the ($x$-monotone) region $R_x^{(m)}$ of points of B that are m-dark with respect to horizontal cuts. Similarly, define $g(y)$ to be the cost of the horizontal line through y, and let $A_y = \int_0^1 g(y)dy$.

Assume, without loss of generality, that $A_x \geq A_y$. We claim that there exists a horizontal favorable cut; i.e., we claim that there exists a horizontal cut, ℓ, such that its chargeable length (i.e., length of its m-dark portion) is at least as large as the cost of ℓ ($|\sigma_m(\ell)|$). To see this, note that A_x can be computed by switching the order of integration, "slicing" horizontally, rather than vertically; i.e., $A_x = \int_0^1 h(y)$ where $h(y)$ is the chargeable length of the horizontal line through y. Thus, since $A_x \geq A_y$, we get that $\int_0^1 h(y)dy \geq \int_0^1 g(y)dy \geq 0$. Thus, it cannot be that for all values of $y \in [0,1]$, $h(y) < g(y)$, so there exists a $y = y^*$ for which $h(y^*) \geq g(y^*)$. The horizontal line through this y^* is a cut satisfying the claim of the lemma. (If, instead, we had $A_x \leq A_y$, then we would get a *vertical* cut satisfying the claim.)

The charging scheme assigns a charge to the edges of E of total amount equal to (roughly) $1/m$th of the length of E. We therefore have shown the following structure theorem:

▶ **Theorem 34.** *Let G be an embedded connected planar graph, with edge set E consisting of line segments of total length L. Let $\mathcal{R}$ be a set of disjoint fat regions and assume that $E \cap P_i \neq \emptyset$ for every $P_i \in \mathcal{R}$. Let W_0 be the axis-aligned bounding box of E. Then, for any positive integers m and M, there exists an edge set $E' \supseteq E$ that obeys the (m, M)-guillotine property with respect to window W_0 and regions $\mathcal{R}_{W_0}$ and for which the length of E' is at most $L + O(\frac{1}{m})L + O(\frac{1}{M})\lambda(\mathcal{R}_{W_0})$, where $\lambda(\mathcal{R}_{W_0})$ is the sum of the diameters of the regions $\mathcal{R}_{W_0}$.*