SIAM J. COMPUT.
Vol. 15, No. 2, May 1986

ON THE COMPLEXITY OF NONCONVEX COVERING*

WOLFGANG MAASS†

Abstract. We study the problem of covering given points in Euclidean space with a minimum number of nonconvex objects of a given type. We concentrate on the one-dimensional case of this problem, whose computational complexity was previously unknown. We define a natural measure for the "degree of nonconvexity" of a nonconvex object. Our results show that for any fixed bound on the degree of nonconvexity of the covering objects the one-dimensional nonconvex covering problem can be solved in polynomial time. On the other hand without such bound on the degree of nonconvexity the one-dimensional nonconvex covering problem is NP-complete. We also consider the capacitated version of the nonconvex covering problem and we exhibit a useful property of minimum coverings by objects whose degree of nonconvexity is low.

Key words. NP-completeness, computational geometry, covering, nonconvexity, polynomial time algorithm, robotics

AMS(MOS) subject classifications. 03D15, 68C25

1. Introduction. In this paper we study the problem of covering given points in Euclidean space with a minimum number of nonconvex objects of a given type. We concentrate on the one-dimensional case of this problem, whose computational complexity was previously unknown. We further restrict our attention to rings—arguably the simplest nonconvex objects (it is not difficult to extend our algorithms to other types of nonconvex objects).

A number of researchers (see the discussion and references in Johnson [6, p. 185]) have shown that the following problem is NP-complete: Decide whether n given points in the Euclidean plane can be covered by k discs of a given radius w. We now look at a nonconvex variation of this problem. In the following, a ring (or annulus) of size $\langle r, w\rangle$ is the set of points that are enclosed by two concentric circles of radius r respectively $r+w$ (r and w will always be nonnegative integers in this paper). If we substitute in the two-dimensional covering problem the discs by rings of given size $\langle r, w\rangle$, the resulting nonconvex covering problem is still in NP. Thus the extension to nonconvex covering objects (rings) does not change the computational complexity of the problem: In *two* dimensions both the convex covering problem (with discs) and the nonconvex covering problem (with rings) are NP-complete.

In contrast to the preceding observation we show in this paper that in the one-dimensional case significant differences arise between the complexity of the convex and the nonconvex covering problem. In the one-dimensional case we assume that n points on a line are given. We assume that the covering rings have their centers on the same line. Thus the intersection of a ring of size $\langle r, w\rangle$ with this line consists of two closed intervals of length w which are separated by an (open) interval of length $2r$ in between. In the following discussion we will refer to such a pair of intervals as a "one-dimensional ring of size $\langle r, w\rangle$". The *one-dimensional ring cover problem* is the problem of computing for n given points on a line the positions of a minimum number of one-dimensional rings of given size $\langle r, w\rangle$ so that all given points are covered. This

* Received by the editors December 29, 1982, and in revised form, July 1, 1984. During the preparation of this paper the author was supported by the Heisenberg Program of the Deutsche Forschungsgemeinschaft, Bonn.

† Department of Mathematics and Computer Science Division, University of California, Berkeley, California 94720. Current address, Department of Mathematics, Statistics and Computer Science, University of Illinois, Chicago, Illinois 60680.

problem contains, for $r=0$ as a special case, the one-dimensional *convex* covering problem. A very simple algorithm (see the beginning of § 3) shows that this one-dimensional convex covering problem can be solved in linear time. In contrast to this we show in § 2 that the one-dimensional ring cover problem is strongly NP-complete. This intractability result comes somewhat unexpectedly insofar as almost all geometrical problems become tractable when they are restricted to one dimension.

In § 3 we close the gap between the previously mentioned two results. We identify the quotient r/w (which may be viewed as a natural measure for the degree of nonconvexity of a ring of size $\langle r, w\rangle$) as the key parameter that determines the complexity of the problem of covering with rings of size $\langle r, w\rangle$. Theorem 3.1 shows that not only for $r/w=0$ (convex case) but also for any fixed bound on this parameter r/w the corresponding one-dimensional ring cover problem can be solved in polynomial time. In § 4 we exhibit in addition a threshold for this parameter r/w (at $r/w=\frac{1}{2}$) where qualitative changes in the structure of minimum coverings by rings of size $\langle r, w\rangle$ take place.

Finally we consider in § 5 a capacitated version of the one-dimensional ring cover problem. We assume here that the number of points that may be "served" by each ring is bounded by a given capacity b. We show that for any fixed bounds on b and r/w this problem is also in P.

It is obvious that our algorithms can be generalized to cases where one covers with other nonconvex one-dimensional objects. We mention further generalizations at the end of § 3.

Finally we would like to mention two possible practical applications of the considered problems. In scheduling theory one may interpret the line as a time axis on which particular time points are given. If resources (for example work shifts) are to be scheduled so that all given time points are covered, one arrives at a one-dimensional covering problem. In certain realistic models where resources are only intermittently available (for example due to lunch breaks for workers or preventive maintenance of machines) this covering problem is nonconvex. For example one covers with one-dimensional rings of size $\langle\frac{1}{2}, 4\rangle$ if every eight-hour work shift is interrupted by a one-hour break in the middle. We refer to Bartholdi III [1] for a further discussion of this application.

There are other possible applications if one interprets the considered line as a line in space. We would like to mention two examples from robotics. In this area one might want to cover given points in space by certain geometrical objects that model the set of points which are reachable by the arm of a robot (for a fixed position of the base of the robot). This set of reachable points may be nonconvex because of imperfections in the robot arm (even for the human arm this set of reachable points forms a ring-like structure). Thus if one wants to compute for a given set of points in space the positions for a minimum number of robot arms so that all points can be reached by some robot arm, one arrives at a (convex or nonconvex) covering problem. Alternatively one might want to compute for one mobile robot a tour where each of a number of given points can be reached by the arm of the robot from some stop of the (base of the) robot. If the goal is to minimize the number of stops for the (base of the) robot, the same covering problem as before arises.

The results of this paper serve as a basis for a series of subsequent papers with Dorit Hochbaum, where we design polynomial time approximation schemes and fast approximation algorithms for one-dimensional and higher-dimensional covering problems [3]–[5].

2. Nonconvex covering in one dimension is NP-complete. We refer to the first section for a definition of the ring cover problem.

THEOREM 2.1. *The one-dimensional ring cover problem is strongly* NP-*complete.*

Proof. We first note that the considered problem is in NP. Here one uses the fact that it is sufficient to consider positions of rings where one of the four endpoints of the (one-dimensional) ring coincides with one of the given points.

In order to show that the considered problem is NP-*complete* we construct a polynomial time computable reduction from the NP-complete problem 3SAT (see Garey and Johnson [2]). The strategy is somewhat similar to the strategy of the reduction from 3SAT to 3-DIMENSIONAL MATCHING. However one has to work harder to construct suitable problem instances in only one dimension.

Before we define the desired reduction, we introduce an essential tool for the construction of suitable instances of the one-dimensional ring cover problem. This tool makes it possible to interconnect the coverability properties of various different clusters of points in the constructed instances. Consider a sequence $W_1, \cdots, W_{2k}$ of points on the line that are spaced $2r+w$ apart. For example assume that W_i has the coordinate $i\cdot(2r+w)$. Obviously one can cover all points in this sequence with k rings of size $\langle r, w\rangle$: The ith ring covers W_{2i-1} and W_{2i}. If W_1 does not need to be covered by the considered k rings (because it is already covered by some other ring), we can shift the k rings over a distance $2r+w$ to the right. In this case the ith ring covers W_{2i} and W_{2i+1}. Further the kth ring covers only one point: W_{2k}. Therefore we can use the other w-interval of the kth ring to cover some other point in the neighborhood of W_{2k}. Thus we see that a covering advantage at the beginning W_1 of the sequence causes a chain reaction in the covering of the sequence $W_1, \cdots, W_{2k}$ (the possibility of shifting all k rings to the right), which leads to a covering advantage at the last point W_{2k}: the kth ring has one interval free. In this sense the sequence $W_1, \cdots, W_{2k}$ can transmit covering advantages and therefore we call it a "wire". In the first situation (where the kth ring has to cover W_{2k-1} and W_{2k}), we say that the wire transmits the "signal 0". In the second situation where the kth ring only has to cover the last point W_{2k}, we say that the wire transmits the "signal 1".

So far we have made no use of the nonconvexity of the covering objects. Everything we have said remains true if we cover with (convex) intervals of length $2r+2w$ instead of rings. We now show that the nonconvexity of the covering objects allows us to run several wires in parallel, so that each can transmit a signal 0 or 1 without mutual interference. It turns out that the number of wires that we can run in parallel is proportional to r/w (this is the first indication of the importance of the parameter r/w for the complexity of the ring cover problem). Consider a second wire $V_1, \cdots, V_{2k}$ where point V_i has coordinate $i\cdot(2r+w)+p$. The "phase shift" p that occurs here is some integer with $w<p<2r$. Obviously the wire $V_1, \cdots, V_{2k}$ has the same covering properties as the first wire $W_1, \cdots, W_{2k}$. But in addition the choice of the phase shift p guarantees that no ring can cover two points that belong to different ones of these two wires. This implies that the choice of a covering of one of the two wires has no consequence for the covering of the other wire. In the previously introduced terminology we can say that both wires can transmit a signal 0 or 1 without mutual interference. In the same way we can run d wires in parallel (for any natural number $d \leq r/w$) that transmit signals 0 or 1 without mutual interference. We merely have to choose for the d wires phase shifts $p_1, \cdots, p_d$ so that for any $i \neq j$ we have $w<|p_i-p_j|<2r$.

We now construct the desired reduction from 3SAT. Let F be an arbitrary instance of 3SAT, let $U=\{u_1, \cdots, u_n\}$ be the set of variables in F and let $C=\{c_1, \cdots, c_m\}$ be the set of clauses in F. Each clause c_j is a disjunction of up to three (negated or unnegated) variables from U. The conjunction of all clauses in C yields the considered formula F. In the following we construct in polynomial time a set P of integers (which are interpreted as coordinates for points on a line) and integers r, w, M such that all points in P can be covered by M rings of size $\langle r, w \rangle$ if and only if F is satisfiable. The points in P fall into four classes of components according to their intended function: "truth-setting", "satisfaction testing", "wire" and "wire crossing".

Each truth setting component corresponds to a single variable $u_i \in U$ (we call it the u_i-component for this variable u_i). The points of each u_i-component can be covered in exactly two different ways by minimum coverings by rings of size $\langle r, w \rangle$. In this way each u_i-component forces any minimum covering of all points in P to make a choice between these two possible coverings of the u_i-component. This choice corresponds to the choice between setting $u_i = \text{true}$ or $u_i = \text{false}$ in a truth assignment to all variables in U.

Each satisfaction testing component in P corresponds to a single clause $c_j \in C$ (therefore we call it the c_j-component for this c_j). It is connected by up to three wires to those three or less u_i-components for which u_i or $\bar{u}_i$ occur in the clause c_j. The number M of rings that may be used for a covering of P will be chosen so small that a c_j-component can only be covered if one of the three wires transmits the signal 1 to the c_j-component. In this case the last ring that is used for the covering of this wire can use its "free" w-interval to cover the c_j-component (while its other w-interval covers the last point of that wire). According to this plan we just have to make sure that the wire from a u_i-component to this c_j-component transmits the signal 1 if and only if the chosen covering of the u_i-component corresponds to setting $u_i = \text{true}$ (in case that u_i occurs in c_j), respectively, to setting $u_i = \text{false}$ (in case that $\bar{u}_i$ occurs in c_j).

We set $w = 10$ and $r = 100w \cdot (4m + n)$. According to our outline up to $3m$ wires are needed. We fix a numbering of these wires and we reserve for the kth wire the "track" with phase shift $p_k = 100w \cdot k$. In general all points in P with a coordinate z such that $z \equiv p_k \bmod (2r + w)$ will belong to the kth wire (the only exceptions are points from "crossing components" that will be discussed below). For each u_i-component we reserve a track with phase shift $100w \cdot (3m + i)$. Each point in the u_i-component will have the property that it is within $3w$ of the u_i-track. Finally each c_j-component consists of a single point y such that $y \equiv 100 \cdot (3m + n + j) \bmod (2r + w)$.

We have now assigned to each wire, u_i-component and c_j-component a separate "track". No ring of size $\langle r, w \rangle$ can cover points that belong to two different tracks. Therefore the coverings of the different components are mutually independent, except for those pairs of components where we force a dependency via a wire. Such a wire connecting a u_i-component with a c_j-component begins on the track of the u_i-component (this means that the first points of the wire have the same phase shift as the u_i-component). Then it moves to its assigned track (see the assignment above) and stays on this track until the end, when it moves to the track of the c_j-component. In order to move a wire from one track to another, we use the fact that the points of a wire need not necessarily be spaced $2r + w$ apart. If we choose instead some distance $2r + w + d$ with $d \in [-w, +w]$ between successive points of a wire, the covering properties of the wire remain unchanged. However for $d < 0$ the wire moves towards a track with a smaller phase shift p. If we use this distance several times in the wire, the wire can reach in this way any other track. Similarly if we choose $d > 0$ the wire moves towards a track with a larger phase shift p.

If a wire leaves its assigned track and approaches some track that has been assigned to some other wire, the coverings of both wires may interfere. In order to avoid this we use in these situations a special "crossing component" that allows a wire to cross the track of some other wire without interference.

Figure 1 provides a global picture of the construction for the case of the formula $F \equiv (u_1 \vee u_2) \wedge (\bar{u}_1 \vee u_2) \wedge (u_1 \vee u_3)$. The horizontal dimension of the diagram represents the actual line on which the points of P are located. The vertical dimension of the diagram is used to indicate the different tracks. Of course in reality all these tracks run along the same line (but with different phase shifts).

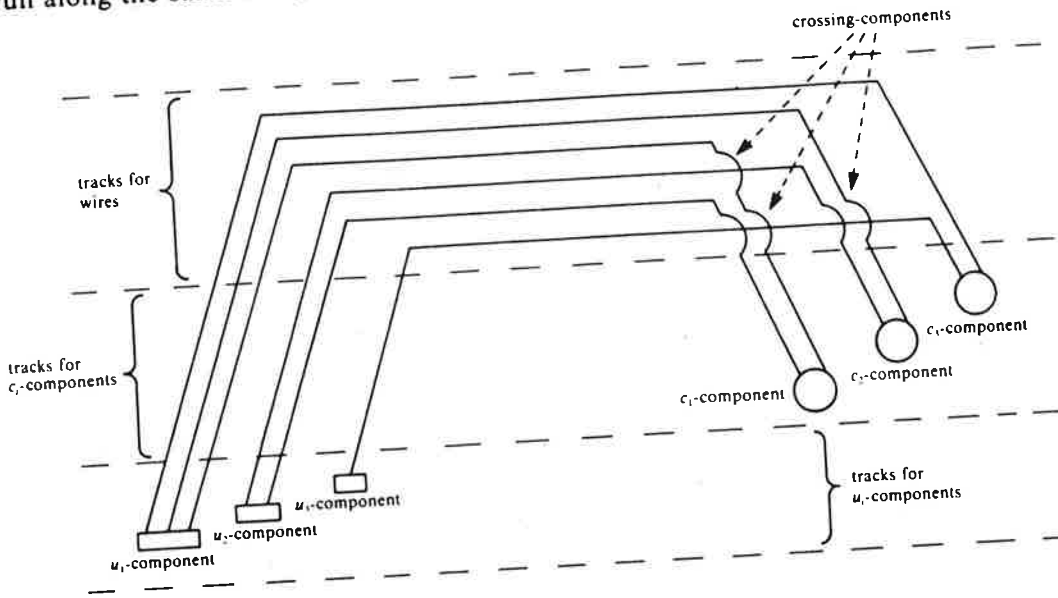
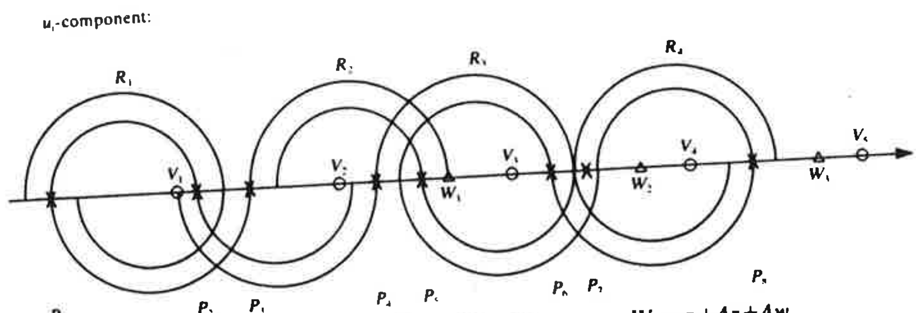


FIG. 1

We will now look at the design of u_i-components, c_j-components and crossing components in more detail.

Figure 2 shows a u_i-component for the simple case where u_i or $\bar{u}_i$ occur only in two clauses, say u_i occurs in c_1 and $\bar{u}_i$ in c_2. The u_i-component consists of the points $P_1, \cdots, P_8$, whose coordinates are also given in Fig. 2. The figure shows in addition



$P_1 = z$
$P_2 = z + 2r$
$P_3 = z + 2r + 2w$
$P_4 = z + 4r + w$
$P_5 = z + 4r + 3w$
$P_6 = z + 6r + 2w$
$P_7 = z + 6r + \frac{7}{2}w$
$P_8 = z + 8r + 4w$

$V_1 = z + 2r - w$
$V_2 = z + 4r - \frac{1}{2}w$
$V_3 = z + 6r + \frac{1}{2}w$
$V_4 = z + 8r + \frac{3}{2}w$
$V_5 = z + 10r + \frac{5}{2}w$

$W_1 = z + 4r + 4w$
$W_2 = z + 6r + \frac{11}{2}w$
$W_3 = z + 8r + \frac{13}{2}w$

FIG. 2

the initial segments of two wires that are attached to this u_i-component: Wire $W_1, W_2, \cdots$ leads to the c_1-component and wire $V_1, V_2, \cdots$ leads to the c_2-component. These two wires are not attached in the same way to the u_i-component. By assumption variable u_i occurs positively in c_1, therefore, the wire to the c_1-component has to transmit the signal 1 to the c_1-component if and only if the chosen minimum covering of the u_i-component corresponds to setting $u_i =$ true. This means that the rings of the corresponding minimum covering of the u_i-component have to be able to cover in addition also the first point W_1 of this wire. Similarly only a covering of the u_i-component that corresponds to setting $u_i =$ false should be able to cover in addition V_1 (whereas it cannot cover in addition W_1), since u_i occurs negatively in c_2. In the total number M of allowed rings in the covering exactly four rings would be allocated to the covering of the points $P_1, \cdots, P_8$ of the u_i-component of Fig. 2. The previously mentioned two different minimum coverings of the u_i-component (that correspond to setting $u_i =$ true respectively $u_i =$ false) are indicated in the upper respectively lower half of Fig. 2. The coordinate z of the point P_1 satisfies $z \equiv 100w \cdot (3m+i) \bmod (2r+w)$ (according to our earlier assignment of tracks).

In order to verify that the u_i-component in Fig. 2 has the desired properties, we consider any covering of the points $P_1, \cdots, P_8$ by four rings R_1, R_2, R_3, R_4 (numbered according to their location from left to right) of size $\langle r, w\rangle$. Two w-intervals from two different ones of these rings are needed to cover the points P_2 and P_3 because $w < |P_2 - P_3| < 2r$. The same fact holds for the pairs P_4, P_5 and P_6, P_7. Together this implies that P_1 is covered by the left interval of R_1 and P_8 is covered by the right interval of R_4. Further P_2, P_3 (P_4, P_5; P_6, P_7) are covered by the right interval of R_1 (R_2; R_3) together with the left interval of R_2 (R_3; R_4).

It is obvious that no ring R_i can cover with one w-interval both a point P_j and a point from $\{V_2, \cdots, V_5\}$ or $\{W_2, W_3, W_4\}$ (because all these wire points have distance bigger than w from every point P_j).

Finally assume that (like in the top half of Fig. 2) the point P_2 is covered by R_1. This implies that P_3 is covered by the left interval of R_2. Therefore the right interval of R_2 does not cover P_4 because $|P_3 - P_4| = 2r - w < 2r$. Thus P_4 is covered by R_3. From this we conclude that R_3 does not cover P_7 (since $|P_4 - P_7| > 2r + 2w$). Therefore R_3 covers P_6 and R_4 covers P_7. Altogether we see that the initial assumption that P_2 is covered by R_1 forces a structure of the covering where all points P_j are covered by the same rings as in the top half of Fig. 2. Further we see that in this case V_1 cannot be covered by $R_1, \cdots, R_4$: R_1 cannot cover V_1 because R_1 covers P_1 and $|P_1 - V_1| < 2r$; R_2 cannot cover V_1 because R_2 covers P_3 and $|V_1 - P_3| > w$.

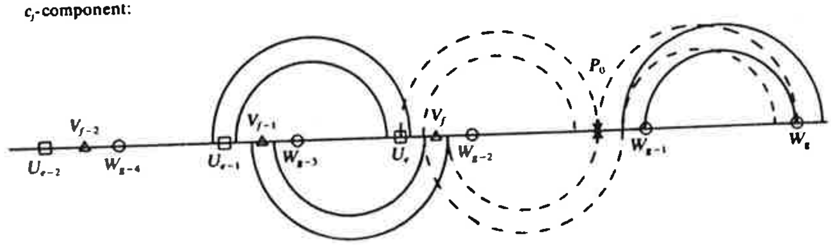
Analogously one shows that in the case where R_1 covers P_3 (instead of P_2), all points P_i are covered by those rings that cover them in the bottom half of Fig. 2. In this case W_1 cannot be covered by $R_1, \cdots, R_4$ (the argument is the same as for V_1 in the previous case).

In the general case more wires may have to be attached to a u_i-component. Then, instead of just three pairs (P_2, P_3), (P_4, P_5), (P_6, P_7) one has to use a correspondingly larger number of pairs (P_{2k}, P_{2k+1}) with $|P_{2k} - P_{2k+1}| = 2w$ and $|P_{2k} - P_{2k+2}| = 2r + w$ (a few of these distances are changed slightly as described below). For all wires that lead to c_j-components such that u_i occurs positively in c_j one positions the first point W of such wire to the right of a pair of points (P_{2k}, P_{2k+1}) (like point W_1 in Fig. 2). One moves in this case the point P_{2k+3} over a distance $w/2$ to the left (like point P_7 in Fig. 2) from its previously indicated position. This small shift ensures that no ring covers both W and P_{2k+3}. Similarly for all wires that lead to a c_j-component such that $\bar{u}_i$ occurs in c_j, one positions the first point V of such wires to the left of a pair of points

(P_{2k}, P_{2k+1}). In this case one shifts the point $P_{2(k-1)}$ over distance $w/2$ to the right of its previously assigned position. Because of this shift no ring can cover both $P_{2(k-1)}$ and V. In order to avoid unexpected interferences between the locations where wires are attached to the u_i-component, one uses in the general case only every fourth pair (P_{2k}, P_{2k+1}) for attaching a wire to the u_i-component (by placing the first point of that wire to the left respectively to the right of this pair). Note that in order to save space we had to ignore this rule in Fig. 2.

Figure 3 shows a c_j-component in full detail. The component itself consists only of one point P_0 at a coordinate z such that $z \equiv 100w \cdot (3m+n+j) \bmod (2r+w)$. The end segments of up to three wires $\tilde{U}=(U_1, \cdots, U_e)$, $V=(V_1, \cdots, V_f)$, $W=(W_1, \cdots, W_g)$ are attached to the c_j-component. In the number M of rings that we allow for covering all points in P, no extra ring is allocated for the covering of any c_j-component. Therefore point P_0 can only be covered if the signal 1 is transmitted to the c_j-component through at least one of the wires $\tilde{U}$, V, W. More precisely the lengths e, f, g of these wires are even numbers and $e/2, f/2, g/2$ rings are allocated in M for the covering of these wires. Therefore if and only if the first point U_1 (V_1, W_1) of such a wire is covered by some other ring (necessarily a ring from the covering of the u_i-component to which this wire is attached), the last one of the rings that are allocated to this wire has to cover only its last point U_e (V_f, W_g). This last ring can cover then with its other w-interval the point P_0 of the c_j-component. In Fig. 3 we have indicated with broken lines the position of the last ring of each wire in the case where this wire transmits the signal 1 to the c_j-component. Figure 3 also shows (with solid lines) the case where the last ring that is allocated to a wire has to cover its last *two* points (this means that the corresponding wire transmits the signal 0 to the c_j-component). It is obvious from the coordinates that are given in Fig. 3 that no ring can cover points that belong to different wires.

c_j-component:



$P_0 = z$

$U_{e-2} = z-6r-4w$
$U_{e-1} = z-4r-3w$
$U_e = z-2r-2w$

$V_{f-2} = z-6r-\frac{5}{2}w$
$V_{f-1} = z-4r-\frac{3}{2}w$
$V_f = z-2r-\frac{1}{2}w$

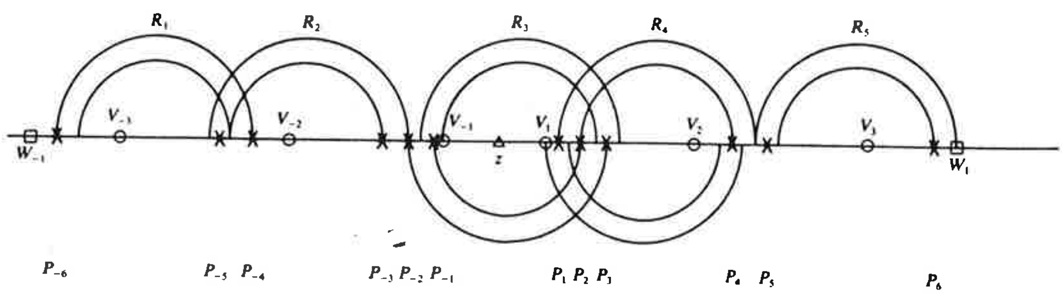
$W_{g-4} = z-6r-w$
$W_{g-3} = z-4r$
$W_{g-2} = z-2r+w$
$W_{g-1} = z+2w$
$W_g = z+2r+2w$

FIG. 3

Figure 4 shows the design of a "crossing component" for the crossing of two wires without interference. Such components are needed because a wire that connects a u_i-component with a c_j-component may have to cross other wires on its way to or from its regular assigned track (see Fig. 1). Figure 4 shows the component for the crossing of wires V and W (of these wires only the segments $V_{-3}, \cdots, V_3$ and W_{-1}, W_1 are indicated in the diagram). The crossing component consists of 12 points $P_{-6}, \cdots, P_6$. The left and the right half of the crossing component are drawn symmetrically (this will simplify the verification). The coordinate z of the middle of the component depends on the track that has been assigned to the wire which is crossed by another wire with the help of the crossing component.

We first note that five rings $R_1, \cdots, R_5$ can be positioned in such a way that they cover $P_{-6}, \cdots, P_6$ and in addition a given one of the two points V_{-1}, V_1 and also a given one of the two points W_{-1}, W_1. The top half of Fig. 4 shows the positions of five rings $R_1, \cdots, R_5$ that cover $P_{-6}, \cdots, P_6$, W_1 and V_{-1}. The bottom half of Fig. 4 indicates a way of shifting R_3, R_4 so that together with the (unchanged) rings R_1, R_2, R_5 the five rings together now cover $P_{-6}, \cdots, P_6$, W_1 and V_1. In the case where W_{-1} has to be covered (instead of W_1) we use the fact that the component is symmetrical with respect to z. With the help of reflection at z we get from Fig. 4 the positions of five rings that cover $P_{-6}, \cdots, P_6$, W_{-1} and V_1 (respectively V_{-1}).

crossing component:



$P_1 = z + r - \frac{1}{2}w$ $\quad V_1 = z + r - w$ $\quad W_1 = z + 5r + \frac{7}{2}w$

$P_2 = z + r + \frac{1}{2}w$ $\quad V_2 = z + 3r - w$ $\quad W_2 = z + 7r + \frac{11}{2}w$

$P_3 = z + r + \frac{3}{2}w$ $\quad V_3 = z + 5r$ $\quad$ (this point is not shown)

$P_4 = z + 3r + \frac{1}{2}w$ $\quad V_4 = z + 7r$

$P_5 = z + 3r + 2w$

$P_6 = z + 5r + \frac{5}{2}w$

If $z + d$ is the coordinate of point P_i, then $z - d$ is the coordinate of point P_{-i} (analogously for the points V_{-i} and W_{-i}).

FIG. 4

One can see from the coordinates of the points in Fig. 4 that a ring that covers any of the points $P_{-6}, \cdots, P_6$ can reach no point that belongs to wire V or W except possibly some of the points V_{-1}, V_1, W_{-1}, W_1. Further in the total number M of rings that are allowed for the covering of all points in P only five rings are allocated for each crossing component. If six or more rings are used to cover $P_{-6}, \cdots, P_6$ then these rings may cover simultaneously all points in the set $\{V_{-1}, V_1, W_{-1}, W_1\}$. But by the design of the other components it is then impossible to cover with $M-6$ rings all the remaining points in P.

We now show that if any five rings $R_1, \cdots, R_5$ (numbered according to their position from left to right) cover all points $P_{-6}, \cdots, P_6$ then either W_{-1} or W_1 and either V_{-1} or V_1 are not covered by these rings. Obviously all the groups $\{P_{-5}, P_{-4}\}$, $\{P_{-3}, P_{-2}, P_{-1}\}$, $\{P_1, P_2, P_3\}$, $\{P_4, P_5\}$ contain two points whose distance d lies strictly between w and $2r$. Therefore for each of these groups at least two different rings must participate in the covering of this group. This implies that each of these groups is covered by precisely the same rings as in Fig. 4.

Assume for a contradiction that both V_{-1} and V_1 are covered by $R_1, \cdots, R_5$ (besides $P_{-6}, \cdots, P_6$). If R_2 covers V_{-1} and R_4 covers V_1 then R_3 has to cover both P_{-3} and P_3, although $|P_{-3} - P_3| > 2r + 2w$. Thus we may assume that R_3 covers V_{-1} (the case where R_3 covers V_1 is symmetrical). Then the location of the left end of R_3 is at some coordinate $\geqq z-r$. Therefore the other interval of R_3 does not cover P_2. Further V_1 can only be covered by R_4. Therefore the left end of R_4 is located at some coordinate $\leqq r-w$. This implies that R_4 does not cover P_2. Thus P_2 remains uncovered, a contradiction.

Finally assume for a contradiction that both W_{-1} and W_1 are covered by $R_1, \cdots, R_5$ (besides $P_{-6}, \cdots, P_6$). Our preceding consideration implies that only R_1 can cover W_{-1} and that only R_5 can cover W_1. Therefore R_1 does not cover P_{-4} and R_5 does not cover P_4. Thus R_2 covers P_{-4} and R_4 covers P_4. This implies that R_2 does not cover P_{-3} and R_4 does not cover P_3. But R_3 cannot cover P_{-3} and P_3 since $|P_{-3}-P_3|>2r+2w$. Thus either P_{-3} or P_3 remains uncovered, a contradiction.

We have specified for each occurring component c the number M_c of rings that are allocated for this component among the M rings. M is then defined as the sum of these M_c (over all components c). P is defined as the union of all components (note that the number of components is polynomial in m and n). The preceding arguments imply that the given formula F is satisfiable if and only if all points in P can be covered by M rings of size $\langle r, w\rangle$ (for r and w as defined before).

Remark 2.2. In some sense it is easier to reduce PLANAR 3SAT instead of 3SAT to the considered problem (see Lichtenstein [7]): no crossing components are needed in this case. On the other hand one then has less control over the structure of the (planar) graph that has to be represented. This fact makes an explicit description of this variation of the proof very difficult.

3. Covering with rings of bounded degree of nonconvexity is in P. In order to demonstrate why covering with nonconvex objects is more difficult than covering with convex objects, we first give a simple algorithm for covering with convex objects in one dimension (we cover with one-dimensional rings of size $\langle r, w\rangle$ where $r/w=0$). In this algorithm one places intervals of length $2w$ successively so that their left endpoint coincides with the leftmost one of the given points that is not yet covered.

If one covers with nonconvex rings, one has several choices among positions of rings that cover the leftmost point that is not yet covered. One can either place this ring far to the right (so that its right end reaches as far as possible) or one can place it more to the left (so that the left end of the right w-interval covers additional points). In this way the number of reasonable choices for placing the first m rings grows exponentially in m. Therefore we use a different approach in the following polynomial time algorithm.

THEOREM 3.1. *There is an algorithm that computes for n given points on the line and a given ring size $\langle r, w\rangle$ a covering of the given points by a minimum number of rings of size $\langle r, w\rangle$ in $O(n^{O(\lceil r/w\rceil)})$ steps (respectively in $O(n)$ steps if $r/w=0$).*

Proof. Assume n points on the line and a ring size $\langle r, w\rangle$ with $r>0$ are given. The algorithm relies on the following definition.

DEFINITION 3.2. Consider an arrangement B of rings of size $\langle r, w\rangle$ on the line, where the leftmost ring has its center at C_L and the rightmost ring has its center at C_R. We call B a *block* if every given point in the interval (C_L+r+w, C_R-r-w) is covered by some ring in B.

LEMMA 3.3. *Consider any covering C of all given points by rings of size $\langle r, w\rangle$. Let B be a subset of rings from C. Let C_L be the center of the leftmost ring in B and let C_R be the center of the rightmost ring in B. Assume that every ring in C whose center is located in the interval (C_L, C_R) belongs to B. Then B is a block.*

The proof of Lemma 3.3 follows immediately from the definition of a block.

LEMMA 3.4. *For any block B the subset of the n given points that are covered by B can be characterized with the help of at most* $8\lceil r/w\rceil+6$ *of the given points (independent of the size of B).*

Proof of Lemma 3.4. Let C_L be the leftmost and C_R be the rightmost center of rings in B. Then all of the given points in (C_L+r+w, C_R-r-w) and none of the given points in $(-\infty, C_L-r-w)$ or $(C_R+r+w, +\infty)$ are covered by B. Within the interval $[C_L-r-w, C_L+r+w]$ the block B defines a set of at most $2\lceil r/w\rceil+1$ disjoint intervals of length $\geqq w$ so that in each such interval all given points are covered by B and any two such intervals are separated by intervals of uncovered points. Each of these up to $2\lceil r/w\rceil+1$ intervals can be characterized by the leftmost one and the rightmost one of the given points that it covers. Together this requires up to $2\cdot(2\lceil r/w\rceil+1)$ points. We need the same amount of information to characterize the right end of B. Finally we need two more points to describe the endpoints of the largest interval in the interior of B where all given points are covered.

To motivate our algorithm we consider any minimum covering C of all given points. We partition C into two blocks B_1 and B_2 where B_1 consists of the 2^m leftmost rings in C and m is maximal such that $2^m<|C|$. In the same way we partition each of the blocks B_1, B_2 into two blocks of about half its size. After at most $\lceil\log_2|C|\rceil$ iterations of this step we have broken down C into its "atoms": single rings. The following dynamic programming algorithm reverses the described process: we look at all possible ways of concatenating two smaller blocks so that they yield one larger block. We can do this in polynomial time because by Lemma 3.4 there exist at most $n^{8\lceil r/w\rceil+6}$ different subsets S of the set of all n given points so that some block of rings covers precisely the points in S.

ALGORITHM. We develop a table where we record for blocks of increasing lengths the subset of the n given points which is covered by each block. In addition we record for each block the number of rings that it uses and the locations of the centers of its leftmost and its rightmost ring. Thus each entry in the table requires at most $O((8\lceil r/w\rceil+9)\cdot\log n)$ bits. We also set up a list that allows us to check in $O((8\lceil r/w\rceil+9)\cdot\log n)$ steps whether a candidate entry for the table already appears in the table.

In the first row of the table we record for all blocks of length 1 the described data. It is sufficient to consider here only rings that are positioned in such a way that one of their four endpoints coincides with one of the given points.

In each subsequent row we record the described data for each block that arises as the union of two blocks from previous rows (unless we get an entry that appears already in the table).

After we have written $\lceil\log_2 n\rceil$ rows we give as output the first entry in the table where all n given points are covered such that no covering of all points with fewer rings has been recorded in the table.

To justify the algorithm we note that one can always shift a ring—without changing the set of given points that it covers—until one of its endpoints coincides with one of the given points. The correctness of the algorithm follows then from our preceding observations.

There are at most $O(n^{8\lceil r/w\rceil+9})$ entries in the table. Thus one has to check for at most $O(n^{16\lceil r/w\rceil+18})$ pairs of previously recorded entries whether they yield a new entry in the table. For each of these pairs one needs at most $O(\lceil r/w\rceil\cdot\log n)$ steps to check whether the union of the corresponding blocks yields a new block whose characteristic data do not yet appear in the table (and to compute the characteristic data of the new block). In this way we arrive at an upper bound of $O(n^{16\lceil r/w\rceil+18}\cdot\lceil r/w\rceil\cdot\log n)$ steps for the algorithm.

Remark 3.5. Theorem 3.1 shows that for any fixed bound on the degree of nonconvexity r/w of the covering rings the one-dimensional ring cover problem is in P. Easy variations of the proof show that this remains true if in addition certain parts of the line are "forbidden" as centers of rings. Further one can associate different costs with having centers of rings at different locations and then compute in polynomial time a covering of minimum cost. One can also extend these algorithms to the case where besides points on a line also certain whole intervals have to be covered.

In another possible extension one might assume that k different ring sizes $\langle r_1, w_1\rangle, \cdots, \langle r_k, w_k\rangle$ are given, where rings from any of these sizes may be used for a minimum covering. Also one can associate different costs with different ring sizes and compute a covering of minimum cost in polynomial time. Notice that this variation includes the case where rings of a certain size may be placed not only with their centers *on* the line but also at a number of different distances from the line (in our terminology each distance gives rise to a different ring size when we consider the intersection of such a two-dimensional ring with the considered line). In this extension the ratio $\max\{r_i | i \leqq k\}/\min\{w_i | i \leqq k\}$ appears in the degree of the polynomial time bound in place of r/w.

4. A property of minimum covers by rings of low nonconvexity. If one covers given points on the line by a minimum number of intervals (i.e. rings with $r/w = 0$), one can assume without loss of generality that the leftmost interval of the covering is positioned with its left end at the leftmost given point. Because of this property one can compute in one dimension minimum covers by convex objects in linear time (see the beginning of § 3). Unfortunately this property does not hold for minimum covers by rings of size $\langle r, w\rangle$ for any $r/w > 0$. We show in this section that nevertheless a more general property holds for rings with ratio $r/w \leqq \frac{1}{2}$ (and not for rings with any bigger ratio). The property says that for rings with ratio $r/w \leqq \frac{1}{2}$ one can assume without loss of generality that the leftmost ring of a minimum cover is positioned at one of *two* canonical positions, both of which are easy to compute. Thus one can answer certain questions about the position of the first ring of a minimum cover without computing a minimum cover. One can further use this structural property of minimum covers to design a fast approximation algorithm for covering with rings of ratio $r/w \leqq \frac{1}{2}$ (see [3] and [4]).

LEMMA 4.1. *The following implication holds if and only if $r/w \leqq \frac{1}{2}$: If there exists a minimum cover of given points by rings of size $\langle r, w\rangle$ where the leftmost ring has one of the given points in the gap between its two w-intervals, then there also exists a minimum cover by rings of size $\langle r, w\rangle$ where the leftmost ring is positioned with its left end at the leftmost given point.*

Proof. We first give a counterexample for the case $r/w > \frac{1}{2}$. We assume that three points a, b, c are given. We choose the distance d between b and c such that $w < d < 2r$ (this is possible if and only if $r/w > \frac{1}{2}$). The locations of the points a, b, c are indicated in Fig. 5. The two rings in the upper half of Fig. 5 form a minimum cover. The leftmost ring has point b in its gap. On the other hand if we position the leftmost ring of a

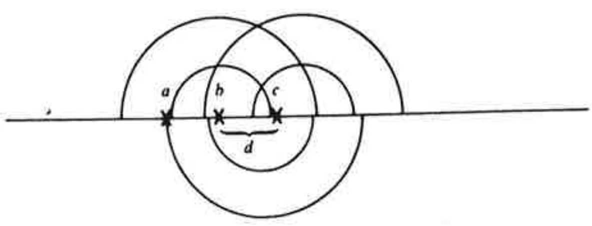


FIG. 5

covering with its left end at the leftmost given point a (as in the bottom half of Fig. 5) we need two more rings to cover b and c.

The positive result for $r/w \leqq 1/2$ follows from the following observation. Let $R_1, \cdots, R_k$ be any minimum cover by rings of size $\langle r, w\rangle$ with $r/w \leqq \frac{1}{2}$ (we assume that the rings are numbered from left to right). Assume that one of the given points lies in the inner disc I_1 (of radius r) of the first ring R_1. Let $\tilde{R}_1$ be a ring that is positioned with its left end at the leftmost given point and let $\tilde{I}_1$ be its inner disc (of radius r). We define point L as the left end of $\tilde{I}_1$. We consider two cases.

Case 1. The left end of ring R_2 is at point L or to the left of L. In this case we continue the cover that was started by $\tilde{R}_1$ with a ring $\tilde{R}_2$ that is positioned with its left end at L. Since $2r \leqq w$ the rings $\tilde{R}_1$ and $\tilde{R}_2$ together cover all points from the leftmost given point until as far as $2r + 2w$ to the right of point L. It is obvious that all points that are covered by R_1 or R_2 fall into this interval.

Case 2. Otherwise. By assumption one of the given points lies in I_1 and without loss of generality this point is covered by R_2. Since this point is left of the right end of $\tilde{I}_1$, the left end of R_2 lies inside of $\tilde{I}_1$. Therefore all given points in $\tilde{I}_1$ are covered by R_2 because $2r \leqq w$. This implies that all given points that are covered by R_1 or R_2 are also covered by $\tilde{R}_1$ or R_2.

THEOREM 4.2. *Assume that points on a line and a ring size $\langle r, w\rangle$ with $r/w \leqq \frac{1}{2}$ are given. Then there is a minimum cover of these points by rings of size $\langle r, w\rangle$ where the leftmost ring R of this cover has at least one of the following two properties*:

(1) *The left end of R coincides with the leftmost given point.*

(2) *R is in the rightmost possible position where it covers the leftmost given point and has none of the given points in its gap.*

Proof. Let S be the leftmost ring of a minimum cover. If one of the given points falls into the gap of S the claim follows from Lemma 4.1. Otherwise the rightmost possible ring R that covers the leftmost given point and has none of the given points in its gap covers all points that are covered by S.

5. The capacitated ring cover problem. We now consider a capacitated version of the ring cover problem. We assume that in addition to the previously considered input data a natural number b is given, which we interpret as the "capacity" of a ring. In addition to a covering we now also have to assign to each of the given points one of the rings that cover this point (one says that the assigned ring "serves" this point). This assignment has to be arranged in such a way that no ring has to serve more than b points. The goal is again to minimize the number of rings that are used.

The capacitated version appears to be of interest for both of the possible applications that were described in the introduction. It also appears to be of some mathematical interest because the algorithm from the previous section does not readily extend to the capacitated problem. The reason for this difficulty is the fact that the degree of the polynomial time bound of the algorithm from Theorem 3.1 is proportional to the number of points that are needed to characterize which of the given points are covered by a block. In an extension of this algorithm to the capacitated case one also has to record for every block which of the points that it covers are served by rings in this block. If for example $n/10$ points lie at the fringe of a block, this may require up to $n/10$ data. Therefore even for a fixed bound on b and on r/w the resulting algorithm is no longer polynomial in the number n of given points.

We show below that there exist among all minimum solutions of the capacitated ring cover problem certain "normal" solutions. Normal solutions are characterized by the fact that they can be decomposed into blocks of a particular simple structure which we call b-blocks (see Definition 5.4). The number of data that are needed to characterize the set of points that are served by the rings of a b-block is independent of n. Since there exist minimum solutions that are in addition normal, it is sufficient to record in the table of a dynamic programming algorithm only those sets of given points that are served by a b-block. In this way we arrive again at a polynomial time algorithm.

We now show that one can "normalize" any given solution to the capacitated ring cover problem without increasing the number of rings that are used. This normalization process consists of two steps. First we minimize the number of rings that serve points in both of their w-intervals. Then we change positions of rings and the assignment of points for those rings that serve now only points in one of their w-intervals in order to minimize the overlap of their "service areas". This second step of the normalization process uses the same method as the proof of the following result for the convex case.

THEOREM 5.1. *There is an algorithm that computes for n given points on a line, a given interval length d and a given capacity b in $O(n)$ steps the positions of a minimum number of intervals of length d together with an assignment of each given point to some interval that covers this point such that no interval serves more than b points.*

Proof. We extend the simple algorithm from the beginning of § 3. We place successively the next interval so that its left end coincides with the leftmost point that is not yet served. We assign to this interval the b leftmost points that it covers (there may be less than b points).

One shows by induction on n that this algorithm uses the minimum number of intervals. For the induction step consider any minimum solution C. It is possible to change the position of the leftmost interval in C and the assignment of points to this interval so that this interval serves the same points as the first interval that is placed by the algorithm. We can then apply the induction hypothesis to the remaining points, where we use the (previously slightly altered) rest of C for comparison.

The following is the desired result for the nonconvex case.

THEOREM 5.2. *There is an algorithm that computes for n given points on the line, a given ring size $\langle r, w\rangle$ and a given capacity b in $O(n^{O(\lceil r/w \rceil b^5)})$ steps a minimum solution to the capacitated ring cover problem.*

Proof. According to the outline at the beginning of this section we first show that among the minimum solutions of the considered problem there exist certain "normal" ones. Let C be a solution of the considered problem. In the first step of the normalization process we minimize the number of rings that serve points in both of their intervals. Thus let $\tilde{C}$ be the result of replacing—without increasing the number of rings that are used—the maximum possible number of rings in C that serve points in both of their intervals by rings that serve points in only one of their intervals (we change the assignment of points accordingly). Of course this has to be done in such a way that $\tilde{C}$ is also a solution of the considered problem.

LEMMA 5.3. *For every real number c there are in $\tilde{C}$ less than b^4+3b rings with center in $[c, c+w]$ that serve points in both of their intervals.*

Proof of Lemma 5.3. Assume for a contradiction that there are in $\tilde{C}$ at least b^4+3b rings with center in $[c, c+w]$ that serve points in both of their intervals. Each such ring R defines a triple of numbers $\langle x_R, y_R, z_R\rangle$ which are the numbers of points that ring R serves in each of the three intervals $[c-r-w, c-r]$, $(c-r, c-r+w] \cap (c-r, c+r)$, $[c+r, c+r+w]$. Since the numbers x_R, y_R, z_R range from 0 to b there are less than b^3+3 different triples of numbers that occur. Thus at least b of these rings have the same triple $\langle x, y, z\rangle$. We show that these b rings can be replaced by b rings that serve points in only one of their intervals. We position x rings with the left end at $c-r-w$ and assign to them those bx points that lie in $[c-r-w, c-r]$ and which were served before by the b replaced rings. Analogously we position y, z, $b-x-y-z$ rings with the left end at $c-r$, $c+r$, $c+r+w$ respectively and we assign to them those points in the corresponding intervals $(c-r, c-r+w]\cap(c-r, c+r)$, $[c+r, c+r+w]$, $(c+r+w, c+r+2w]$ that were served before by the b replaced rings.

The possibility of this substitution contradicts the definition of $\tilde{C}$.

It may still occur that for example two rings R_1 and R_2 in $\tilde{C}$ serve only points in their left interval and R_1 is left of R_2, but R_1 serves some points that lie to the right of points that are served by R_2. Such overlap can make the description of the set of points that are served by a block arbitrarily long. Therefore we consider now the subset S of the n given points that are served in $\tilde{C}$ by rings that serve points in only one of their intervals. Say there are k such rings in $\tilde{C}$. We apply to the points in S the algorithm of Theorem 5.1, where we use intervals of length w. By Theorem 5.1 the algorithm uses exactly k such intervals. We now interpret each such interval as the left interval of a ring of size $\langle r, w\rangle$. After we have changed in this way those rings in $\tilde{C}$ that serve points in only one of their intervals, we call the resulting new covering of all n points C'. By construction C' uses no more rings than C. Further the properties of the algorithm from Theorem 5.1 guarantee that:

(I) If R_1 and R_2 are two rings in C' that serve points in only one of their intervals, then both rings are positioned with the left end at the leftmost point which they serve and if R_1 is positioned left of R_2, then all points that are served by R_1 lie to the left of every point that is served by R_2.

In addition C' has the same rings as $\tilde{C}$ that serve points in both of their intervals. Thus C' retains the property that was proved in Lemma 5.3 for $\tilde{C}$. Thus we have:

(II) For every real number c there are in C' less than b^4+3b rings with center in $[c, c+w]$ that serve points in both of their intervals.

We call a solution C' with properties (I) and (II) a *normal* solution. The preceding construction shows that there always exists a minimum solution that is in addition normal.

As in Theorem 3.1, the key for the dynamic programming algorithm is the definition of a relatively small class of "building blocks" from which one can build via concatenation an optimal solution.

DEFINITION 5.4. Consider an arrangement B of rings of size $\langle r, w\rangle$ such that no ring in B serves more than b points. Let $C_L(C_R)$ be the center of the leftmost (rightmost) ring in B. We call B a *b-block* if we have for $f(b, r/w)=(2\ulcorner r/w\urcorner+2)\cdot(b^5+3b^2)+b$

i) every point in (C_L+r+w, C_R-r-w) is served by a ring from B,

ii) at most $f(b, r/w)$ points in $[C_L-r-w, C_L+r+w]\cap[C_L-r-w, C_R-r-w)$ are not served by a ring from B,

iii) at most $f(b, r/w)$ points in $[C_R-r-w, C_R+r+w]$ are served by a ring from B.

It is obvious that the relevant properties of such b-block can be described with at most $2\cdot f(b, r/w)+3$ data (each of which is essentially a number between 1 and n). Besides C_L, C_R and the number of rings that are used in the b-block this description includes those $2\cdot f(b, r/w)$ points that are mentioned in part ii) respectively iii) of the definition.

LEMMA 5.5. *Let C be a normal solution of the considered problem. Consider a collection B of rings from C where the leftmost ring in B has its center at C_L, the rightmost center in B has its center at C_R and every ring in C that has its center in (C_L, C_R) belongs to B. Then B is a b-block.*

Proof of Lemma 5.5. Property i) is obvious. For property ii) we first consider those points in $I=[C_L-r-w, C_L+r+w]\cap[C_L-r-w, C_R-r-w)$ which are served by rings in C that serve points in only one of their intervals. Except for possibly the first one, these rings have their centers in (C_L, C_R) and thus they belong to B (we use part (I) of the normality definition). This gives rise to at most b points in I that are not served by rings from B. Next we consider those points in I that are served by rings of C that serve points in both of their intervals. Unless their centers are in $[C_L-2r-2w, C_L]$, these rings necessarily belong to B. By part (II) of the normality definition there are at most $((2r+2w)/w)\cdot(b^4+3b)$ rings in C that have their centers in $[C_L-2r-2w, C_L]$ and serve points in both of their intervals. These rings serve at most $(2\ulcorner r/w\urcorner+2)\cdot(b^5+3b^2)$ points. Property iii) is verified analogously.

Similarly as before it is sufficient to consider in the following algorithm only positions of rings where one of the given points coincides with one of the four endpoints of the ring.

ALGORITHM. In the first row of the table we write down the covering properties of all b-blocks that consists of one ring. In each subsequent row we list the characteristic data (consisting of $O(f(b, r/w)\cdot\log n)$ bits) for each new b-block which we get by taking the union of two b-blocks from previous rows.

After we have written down $\ulcorner \log_2 n\urcorner$ rows we output the first b-block in the table that serves all n given points and such that no other b-block in the table serves all n points with fewer rings.

The correctness of the algorithm follows from the previous observations. In particular some minimum solution that is in addition normal will appear in the table. This ensures that the output is a minumum solution.

For the time analysis we note that there are at most $O(n^{2\cdot f(b,r/w)+3})$ entries in the table. Thus we consider at most $O(n^{4\cdot f(b,r/w)+6})$ pairs of b-blocks during the algorithm. For each pair we need at most $O(\log n\cdot f(b, r/w))$ steps to check whether its union forms a b-block whose characteristic data do not yet appear in the table. This leads to an upper bound of $O(n^{4\cdot f(b,r/w)+7}\cdot f(b, r/w))$ steps for the algorithm.

Remark 5.6. The proof of Theorem 2.1 implies that already for a fixed capacity $b\geqq 3$ the one-dimensional capacitated ring cover problem is NP-complete.

REFERENCES

[1] J. J. BARTHOLDI III, *A guaranteed-accuracy round-off algorithm for cyclic scheduling and set covering*, Oper. Res., 29 (1981), pp. 501–510.

[2] M. R. GAREY AND D. S. JOHNSON, *Computers and Intractability*, Freeman, San Francisco, 1979.

[3] D. HOCHBAUM AND W. MAASS, *Approximation schemes for covering and packing problems in robotics and VLSI*, Proc. STACS 84, Lecture Notes in Computer Science, 166, Springer, Berlin, 1984.

[4] ——, *Fast approximation algorithms for a nonconvex covering problem*, submitted.

[5] ——, *Approximation algorithms for covering and packing problems in image processing and VLSI*, J. Assoc. Comput. Mach. (1985).

[6] D. S. JOHNSON, *The NP-completeness column: an ongoing guide*, J. Algorithms, 3 (1982), pp. 182–195.

[7] D. LICHTENSTEIN, *Planar formulae and their uses*, this Journal, 11 (1982), pp. 329–343.