Some NP-Hard Polygon Decomposition Problems

JOSEPH O'ROURKE, MEMBER, IEEE, AND KENNETH J. SUPOWIT

Abstract—The inherent computational complexity of polygon decomposition problems is of theoretical interest to researchers in the field of computational geometry and of practical interest to those working in syntactic pattern recognition. Three polygon decomposition problems are shown to be NP-hard and thus unlikely to admit efficient algorithms. The problems are to find minimum decompositions of a polygonal region into (perhaps overlapping) convex, star-shaped, or spiral subsets. We permit the polygonal region to contain holes. The proofs are by transformation from Boolean three-satisfiability, a known NP-complete problem. Several open problems are discussed.

I. Introduction

The COMPUTATIONAL complexity of many polygon decomposition problems is unknown [39]. Here we investigate the problems of decomposing a polygon into either convex, star-shaped, or spiral subsets. Two important qualifications should be noted: we permit our polygons to contain holes, and we permit the pieces of the decomposition to overlap. We establish that all three decomposition problems are NP-hard and so are fundamentally intractable unless P = NP. (For background on the theory of NP-completeness, see Garey and Johnson [13].) If either of these two qualifications are removed, then our proofs no longer hold, and indeed the complexities of the resulting more restrictive problems remain open questions.

There are at least three distinct motivations for examining polygon decomposition problems. The first is for their applications to pattern recognition. A typical strategy for recognizing a shape as a particular member of a library of shapes is to decompose the shape into "primitive" parts, and then compare with the library entries via some type of similarly function. This method is suggested, for example, by Marr and Nishihara [19], who have used the notions of "generalized cones" and "occluding contours" to derive a decomposition; the problem of defining and computing similarity measures is currently being investigated by Shapiro and Haralick [36], [37], and Radig *et al.* [32] among others. Often the primitive parts of the decomposition are restricted to some particular shape-type; Toussaint [39] calls these "component-directed" (as opposed to "procedure-directed") decompositions. Pavlidis has done the pioneering work on decompositions into convex pieces [25]–[27], [28, pp. 236–241], [29]; see also [33], [22]. Recently, Shapiro and Haralick [35] have relaxed the strict convexity requirement for their primitives and obtained very natural decompositions. Maruyama [21] and later Avis and Toussaint [2] have investigated decomposition into star-shaped polygons, and Feng and Pavlidis [9], [30] have studied spiral decompositions. Other specially shaped primitives, such as "monotonic" polygons [39], [31], have also been explored.

A second motivation for polygon decomposition is that certain calculations are difficult for general polygons but easy for certain simple shapes; in such cases, it may be advantageous to decompose a general polygon into simple shapes, perform the computation on each, and combine the results. This is the approach taken by Ahuja *et al.* [1], who decompose nonconvex polygons into convex pieces for their interference/collision detection algorithms.

The final motivation we will discuss comes from the discipline of computational geometry [34], which is concerned with algorithm design and computational complexity for geometric problems. Following Toussaint [39], we will distinguish decompositions according to whether or not they use *Steiner points* (points that are not vertices of the original polygon), and whether they permit overlapping pieces (a *cover*) or require a *partition* into nonoverlapping pieces. Additionally, we will distinguish between input polygons with and without holes. These distinctions together with the variety of shape types that might be used as primitive elements result in a plethora of polygon decomposition problems. We are interested here in investigating the computational complexity of *minimum* decompositions, i.e., decompositions into the fewest possible number of pieces, subject to the various constraints mentioned above.

There are two major results along these lines. The first is Chazelle and Dobkin's $O(n^3)$ algorithm for finding a minimum convex partition, employing Steiner points, of a polygon without holes [6], [7]. Thus despite the similarity of this problem with many NP-complete optimal partitioning problems, it is solvable in polynomial time. The second major result is Masek's proof [20] that the problem of finding a minimum decomposition of a "rectilinear" polygon (one whose edges are aligned with orthogonal coordinate axes) that may contain holes, using Steiner points, into aligned rectangles, is NP-complete. (This problem is called "rectilinear picture compression" in [13, p. 232].)

Manuscript received November 2, 1981; revised April 29, 1982. The work of the first author was supported in part by the National Science Foundation under Grant MCS-8117424.

J. O'Rourke is with the Department of Electrical Engineering and Computer Science, The Johns Hopkins University, Baltimore, MD 21218.

K. J. Supowit is with Hewlett-Packard Laboratories, Computer Research Center, Palo Alto, CA 94304.

Masek's result does not establish the complexity of unrestricted convex coverings because the minimum convex cover of a rectilinear polygon may require the use of nonrectangular pieces [23]. Other interesting results are available, but their relevance is less clear. For example, Chvâtal's "watchman" theorem [8], [10] establishes that at most $\lfloor n/3 \rfloor$ star-shaped pieces are required to cover a simply connected polygon. This provides an upper bound on the size of a minimum decomposition, but seems to be little help in actually constructing a minimum decomposition.

Although several algorithms have been designed for some of the other polygon decomposition problems, none guarantee minimum decompositions and so do not bear directly on the issue of their worst-case computational complexity. In the sections to follow, we will establish the complexity of three problems: minimum decomposition of polygons that may contain holes, using Steiner points, into overlapping convex, star-shaped, or spiral subsets.

II. Problem Definition

In this section notation will be established and the problems precisely defined. A *simply connected polygonal region* is a closed subset of the plane whose boundary is a simple polygon.[1] A *multiply connected polygonal region* is a finite connected union of simply connected polygonal regions. It may contain polygonal holes, but because it is always a closed region, it can have no "point holes." Multiple connection will be assumed unless otherwise noted.

A polygonal region may be specified by a set of circular lists of its vertices, one for each boundary chain of segments in its boundary, with the usual orientation convention: during counterclockwise traversal of an outer boundary chain or clockwise traversal of an inner boundary chain (bounding a hole), the region is towards the left. In order to obtain definite measures for the "size" of a particular input, all vertex coordinates will be restricted to be integers. An example of a polygonal region is shown in Fig. 1.

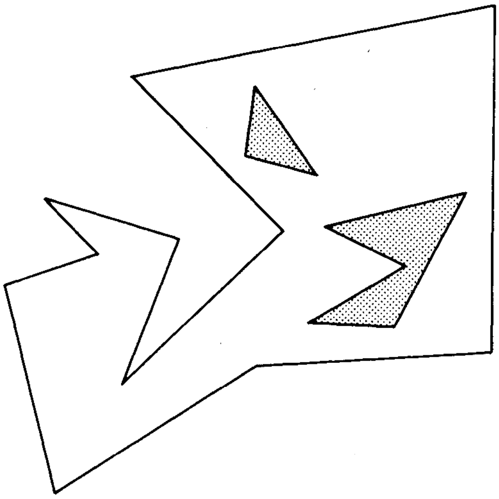


Fig. 1. Multiply connected polygonal region with two holes.

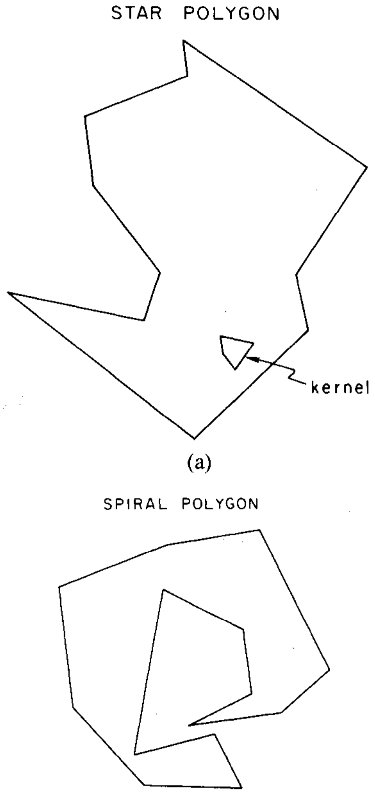


Fig. 2. (a) Star-shaped polygonal region. Every point in nonempty kernel can "see" entire boundary. (b) Spiral region with single chain of contiguous reflex vertices.

Decompositions into three particular shape types will be considered: convex, star-shaped, and spiral. A *convex* polygonal region is a polygonal region that is convex in the usual sense: any two internal points may be connected by a straight line segment that is a subset of the region. A *star-shaped* polygonal region is one whose kernel is nonempty. The kernel of a region is the set of points within the region from which the entire boundary is visible; thus a region is star-shaped if there is at least one point that can "see" every point on the boundary (see Fig. 2(a)). A *spiral* polygonal region is a simply connected polygonal region whose single boundary chain has at most one concave subchain (see Fig. 2(b)). Note that a convex polygonal region is always both star-shaped (its kernel is equal to itself) and spiral (it has no concave subchains). Minimum decompositions of Fig. 1 into the three shape types are shown in Fig. 3(a)–(c).

Letting σ vary over shape properties, and using Σ to represent the set of three shapes,

$$\Sigma = \{\text{convex, star-shaped, spiral}\},$$

we will speak of "σ-subsets" and "σ-covers" for $\sigma \in \Sigma$. The problem we are concerned with in this paper will be called the "minimum σ-cover of a multiply connected

[1] A *simply connected set* S has the property that every closed curve in S can be contracted in S to a point.

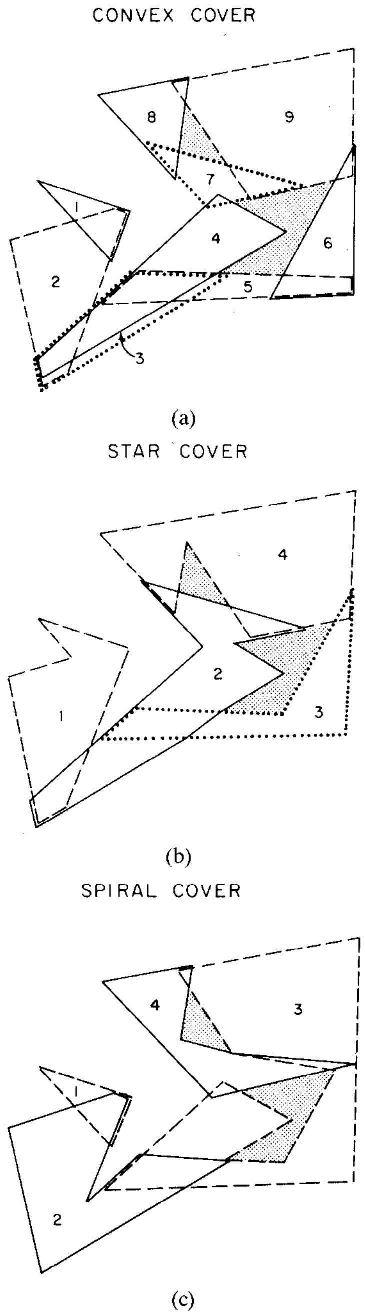


Fig. 3. Minimal decompositions of polygonal region of Fig. 1. (a) Cover by convex pieces. (b) Cover by star-shaped pieces. (c) Cover by spiral pieces.

polygon" problem, or MσCMP, and can be specified as follows (after the style of Garey and Johnson [13]).

Minimum σ-Cover of a Multiply Connected Polygon (MσCMP)

Instance: A set of lists of integer-coordinate vertices representing a polygonal region P, and a positive integer bound K.

Question: Is there a decomposition of P into K or fewer σ-subsets of P, i.e., do σ-subsets $S_1, S_2, \cdots, S_k$ with $k \leqslant K$ exist such that $S_1 \cup S_2 \cup \cdots \cup S_k = P$?

For $\sigma \in \Sigma$, this problem will be shown to be NP-hard by transformation from the Boolean three-satisfiability (3SAT) problem [16], [13, pp. 48–50]:

Boolean Three-Satisfiability (3SAT)

Instance: A set $U = \{u_1, u_2, \cdots, u_n\}$ of Boolean variables and a collection $C = \{c_1, c_2, \cdots, c_m\}$ of clauses over U such that each $c_i \in C$ is a disjunction of precisely three literals.

Question: Is there a satisfying truth assignment for C, i.e., is there a truth assignment to the n variables in U such that the conjunctive normal form $c_1 \cdot c_2 \cdots c_m$ is true?

The proof proceeds along lines similar to proofs of NP-completeness published recently by Fowler *et al.* on the "box-cover" problem [11], [12] and by Supowit on point and disk coverage problems [38].

III. Transformation from 3SAT

The usual first step in a proof of NP-completeness is to show that the problem is a member of the class of NP problems, that is, solvable via a nondeterministic algorithm in polynomial time [13, pp. 27–32]. Often this is easy, merely requiring a demonstration that a solution "guessed" by a nondeterministic program can be checked in polynomial time. With the integer-lattice geometric objects used in our constructions, however, it is unclear how to establish this. In the absence of proofs that the three problems under consideration are members of NP, the arguments presented will establish that the problems are NP-hard rather than NP-complete.

We will now show that 3SAT is polynomially transformable to MσCMP for each $\sigma \in \Sigma$. The three proofs will proceed in parallel, as their overall structure is the same. The goal is to accept an instance of 3SAT as input and construct, in polynomial time, a polygonal region that has a σ-decomposition into a certain number K or fewer σ-subsets iff the given set of clauses is satisfiable. As with other 3SAT transformations [11]–[13], [16], [38], the construction forces a truth assignment with n "truth-setting components" that simulate the Boolean variables, and ensures satisfaction with m "clause components" that correspond to the disjunctive clauses.

A. Truth-Setting Components

The truth-setting components are polygonal regions of a repetitive pattern that have just two distinct minimum σ-decompositions. One of the minimum decompositions is associated with the truth assignment TRUE and the other with FALSE. The basic patterns for $\sigma \in \Sigma$ are shown in Fig. 4. Fig. 4(a) can be minimally covered either with vertical (convex) rectangles or horizontal rectangles. (The basic idea here is also used by Masek [20].) Fig. 4(b) serves as the basic pattern for both star-shaped and spiral decompositions: in both cases, there are exactly two minimum covers. There are certain distinguished points associated with each pattern, labeled with integers in the figures. These points are not part of the construction; they are used in the proof that the polygon constructed has the appropriate cover iff

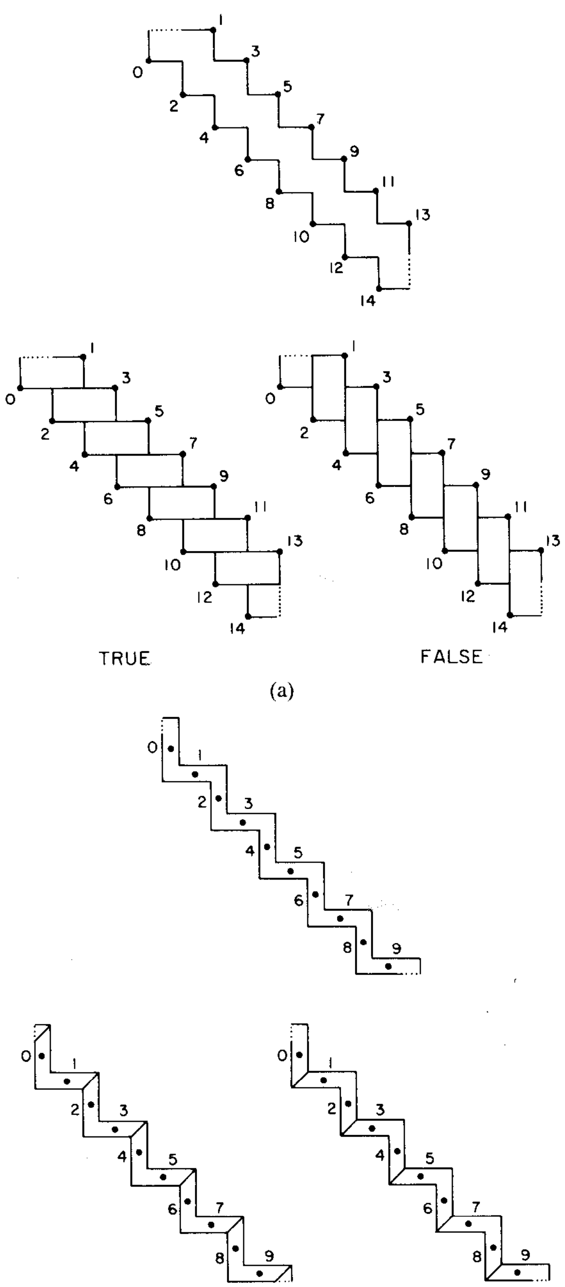


Fig. 4. Basic truth-setting patterns. (a) Convex truth-setting pattern can be minimally covered either by vertically oriented rectangles or by horizontally oriented ones. (b) Star-shaped and spiral truth-setting pattern can be minimally covered either by pieces centered at inner corners or by pieces centered at outer corners.

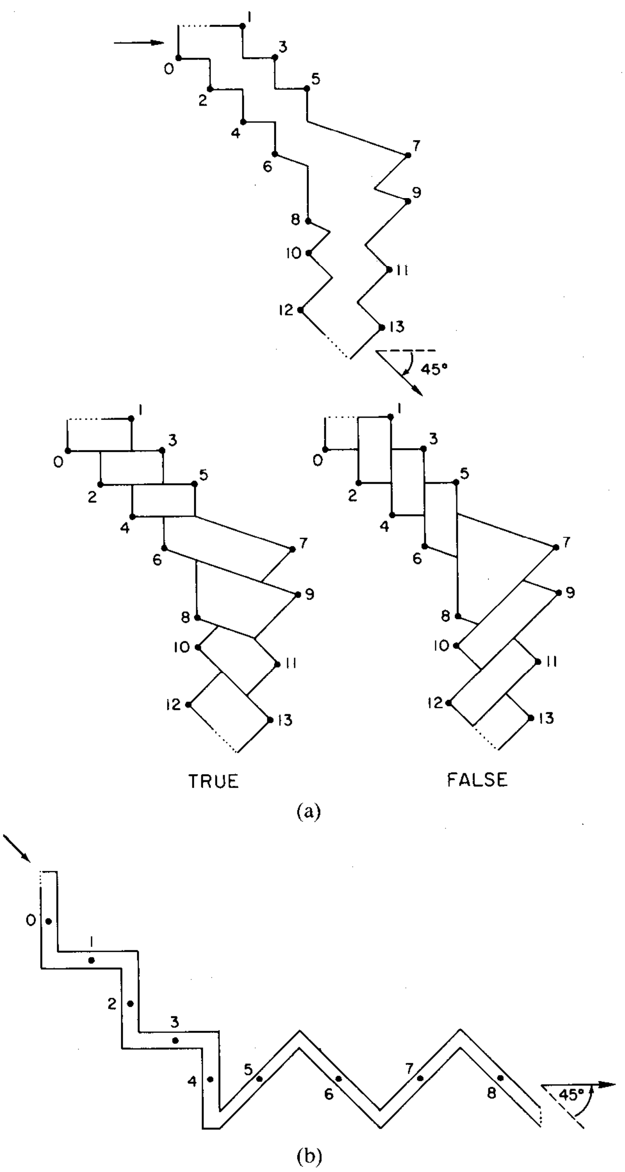


Fig. 5. Variable loop bends. (a) Arm of convex variable loop bending 45 deg and two minimum covers. (b) Arm of star-shaped or spiral variable loop bending 45 deg.

the clause is satisfiable. Note that two of the distinguished points can be covered by one σ-polygon iff the two points are consecutive in the numerical sequence.

The truth-setting patterns are bent (see Proposition 2 in the following) to form closed loops, called *variable loops*. There will be one such loop per Boolean variable u_k in the final construction.

By these remarks, each minimum σ-decomposition for a variable loop contains exactly $r_k/2$ elements, where r_k is the number of distinguished points in the variable loop corresponding to u_k. We call such a decomposition TRUE if it contains distinguished points i and $i + 1$ for all even i (taken modulo r_k), and we call it FALSE if it contains distinguished points j and $j + 1$ for all odd j. Define the bound K used in the definition of the MσCMP problem as equal to $1/2\Sigma_{k=1}^{n} r_k$.

The main properties of variable loops are stated somewhat informally as follows (the proofs, being straightforward, are either omitted or sketched).

Proposition 1: Each minimum σ-decomposition is either a TRUE decomposition or a FALSE decomposition.

In constructing the polygonal region P, it will be necessary to cross variable loops over one another and "bend" them 45 deg, without these modifications affecting the truth of Proposition 1 for any variable loop. The ability to bend a variable loop effectively gives us what is sometimes called an "inverter" in other NP-completeness constructions [20].

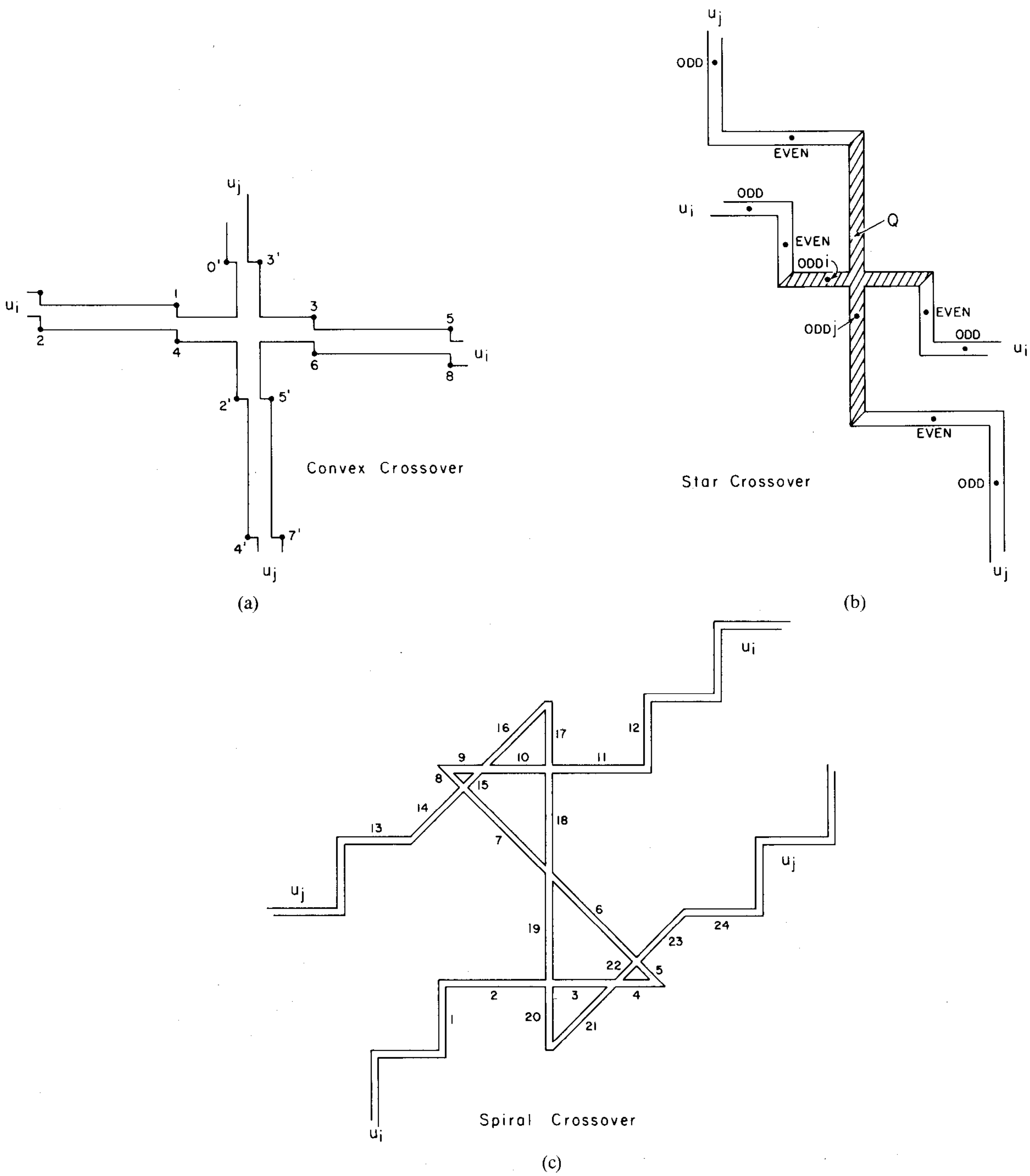


Fig. 6. Variable loop crossovers. Arms of two different variable loops can cross without interfering with their coverage properties. (a) Convex variable loop crossover. None of distinguished points for variable loop u_i (unprimed) can see points for u_j (primed). (b) Star-shaped variable loop crossover. Crossover is arranged so only odd distinguished points (i and j) can be covered by "cross star" such as Q. (c) Spiral variable loop crossover. Twenty-four uninterrupted straight segments in crossover region are numbered for reference.

Proposition 2: The variable loops may bend 45 deg without affecting their properties.

Proof: Fig. 5(a) shows the construction for convex variable loops. Note that it is not difficult to make all vertices have integer coordinates. Fig. 5(b) shows the simpler construction used for star-shaped and spiral variable loops. □

Proposition 3: Two variable loops may cross over one another without affecting their independent coverage properties. More precisely, if two unconnected variable loops require k_1 and k_2 σ-pieces in a minimum decomposition, then one can cross over the other in such a manner that the resulting (connected) polygonal region requires $k_1 + k_2$ σ-pieces in a minimum decomposition, and without altering the type of coverage (TRUE/FALSE) within either variable loop.

Proof: The constructions are shown in Fig. 6. In the convex case, Fig. 6(a), none of the distinguished points in the u_i variable loop is visible to any of the u_j distinguished points. Since any convex piece in a minimum covering must cover at least two distinguished points, and two mutually invisible points cannot be covered by one convex piece, no possibility of interference exists between the two loops.

The star-shaped crossover (Fig. 6(b)), while simple in form, requires a more complicated argument than does the convex case to show that it preserves the desired properties. The star-shaped case would not be difficult were it not for the possibility that distinguished points i and j may be covered by a single star-shaped polygon Q (shown in hatched lines). We can, however, arrange the crossovers to ensure that at each crossover in the complete construction, i and j are both odd. We will argue that this arrangement ensures that each covering that contains $\leqslant K$ pieces (if there are any) is a TRUE / FALSE covering.

Define a *cross star* to be a star-shaped polygon containing at least one distinguished point of each of two variable loops, e.g., the polygon Q in Fig. 6(b). Since the points i and j are both odd at every crossover, each cross star contains exactly two distinguished points, both of which are odd. Since no star-shaped polygon can contain more than one even distinguished point, and since there are K even distinguished points, every cover must contain K stars to cover the even distinguished points. Therefore, if a cover also includes one or more cross stars, then it must have more than K elements. A similar argument was used in [38, pp. 89–93].

The spiral crossover (Fig. 6(c)) is a more complicated form. Our claim that it preserves the desired properties is based on an exhaustive computer search; we have been unable to construct a deductive proof.

There are 24 uninterrupted segments in the critical region of Fig. 6(c), 22 of which are covered by four spirals from the standard TRUE and FALSE coverings of the participating variable loops. The four TRUE / FALSE combinations are shown in Fig. 7(a)–(d). To establish that these coverings are minimum, a list of all *maximal* spirals (those that cannot be extended by a segment on either end) was compiled. No spirals including more than seven segments exist, and the number of distinct 7-, 6-, 5-, 4-, and 3-segment spirals is 8, 12, 4, 26, and 8, respectively. No two- or one-segment spirals can play a role in a minimum covering because that would require the other three spirals to cover $\geqslant 20$ segments, and no three seven-segment spirals exist that overlap on as few as a single segment, nor any two seven-segment and a six-segment spiral that share no segments. Thus there are just 58 spirals from which to choose four to cover $\geqslant 22$ segments.

A computer search found six minimum coverings: the four already discussed (Fig. 7(a)–(d), and two others, shown in Fig. 7(e) and (f). Fig. 7(e) is just a minor variant of Fig. 7(a), and similarly Fig. 7(f) is a variant of Fig. 7(d); neither affects the propagation of TRUE and FALSE " values" through the participating variable loops. Therefore, the spiral crossover preserves the independent coverage properties of each loop. □

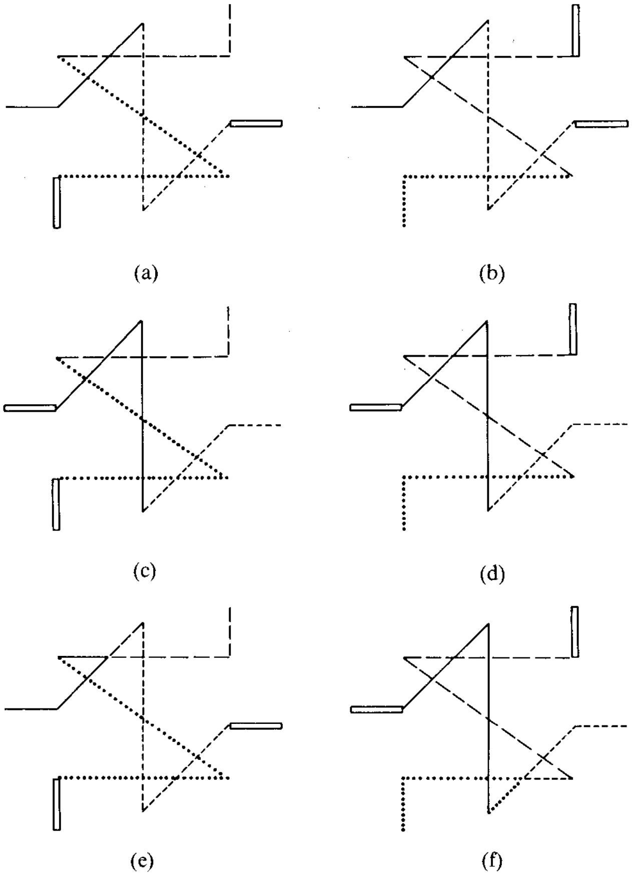
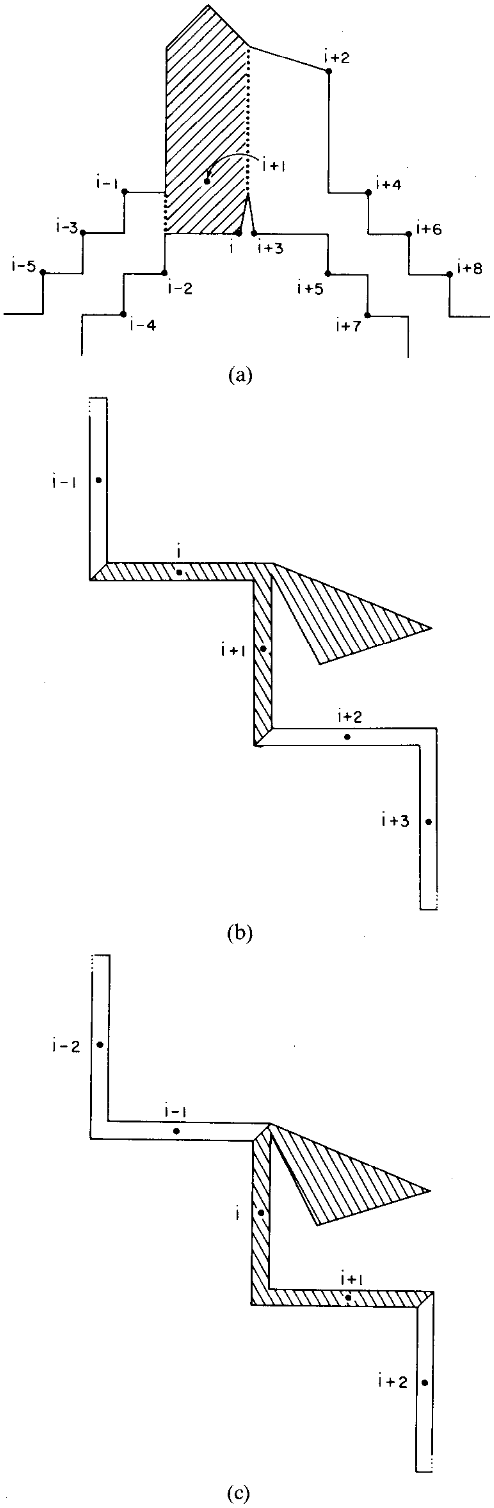


Fig. 7. Minimum covers of spiral crossover region. Four TRUE / FALSE covers (a)–(d) cover 22 of 24 segments; boxed segments remain uncovered. Covers (a) and (d) have two equivalent variants, (e) and (f), respectively.

B. Clause Junctions

If a particular convex variable loop is covered in a convex minimum cover by, say, vertical pieces, then small areas near the ends of the vertical pieces exist that could be covered "free" (without increasing the number of pieces) by extending the vertical pieces towards it, but that cannot be covered free if horizontal pieces were used (see Fig. 8(a)). Similar statements may be made about star-shaped and spiral minimum covers (see Fig 8(b) and (c)). This is the key idea in the formation of the clause junctions.

The heart of a clause junction is an isosceles triangle whose equal sides both slope at 45 deg. (The shape of this triangle is not critical, but it is easier to keep to integer coordinates if its sides slope at 45 deg.) Arms of three different variable loops are brought to the junction, one for each side of the triangle. Suppose the clause represented by the junction is $c = \bar{u}_i + u_j + \bar{u}_k$. Then variable loop j is arranged so that a setting of TRUE will permit the triangle to be covered free, and variable loops i and k are placed so that a setting of FALSE will result in free coverage. The result is that the triangle at the clause junction will be covered free if and only if the clause is satisfied by the truth assignment established by the truth-setting components.

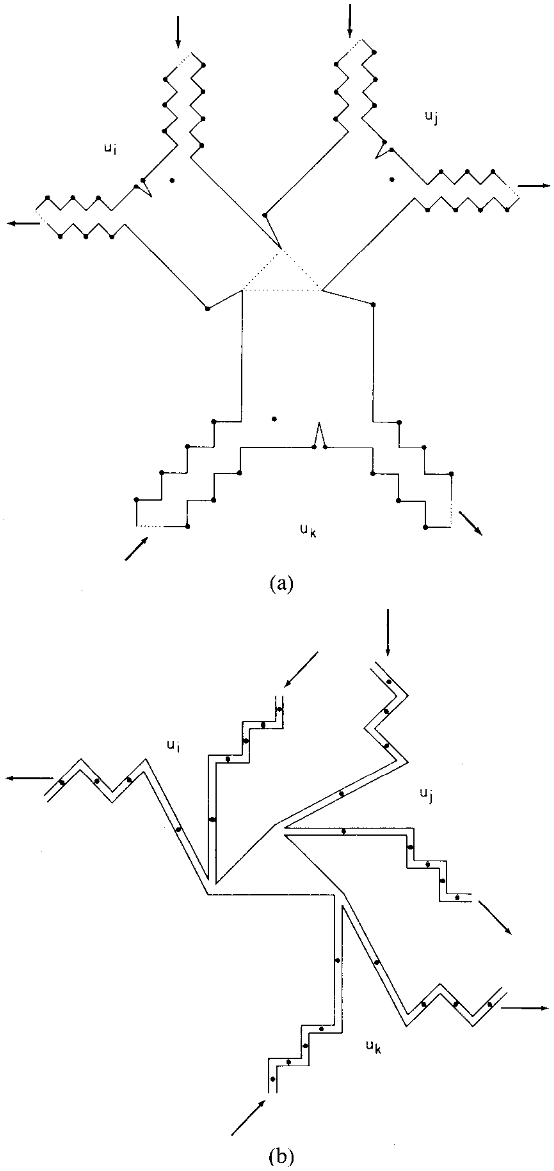


(a)

(b)

(c)

Fig. 8. Triangular regions covered "free." In each of (a)–(c), if loop represents an uncomplemented variable and i is even, then triangular region is covered free by TRUE covering. (a) Convex case. (b) Star-shaped case. (c) Spiral case.



(a)

(b)

Fig. 9. Clause junctions. Central isosceles triangle can be covered free if and only if the corresponding clause is satisfied. Note that no two distinguished points belonging to distinct chains can be covered by σ-polygon. (a) Convex clause junction. (b) Star-shaped and spiral clause junction.

The details of clause junction construction are shown in Fig 9(a) and (b). The three important claims concerning the clause junctions are contained in the following proposition.

Proposition 4: The σ clause junctions as illustrated in Fig. 9 possess the following properties:

1) all vertex coordinates are integers;
2) the central triangle is covered free iff the clause is satisfied;
3) the junction does not affect the independent coverage properties of the participating variable loops.

Proof: That all vertex coordinates are integers follows from the method of constructing the bends and the use of 45 deg angles in the clause triangles. Fig. 8(a)–(c) establish that the central triangle can be covered free if the clause is satisfied. On the other hand, if the clause is not satisfied, then the construction of the clause junction prevents any piece covering a distinguished point of one of the variable

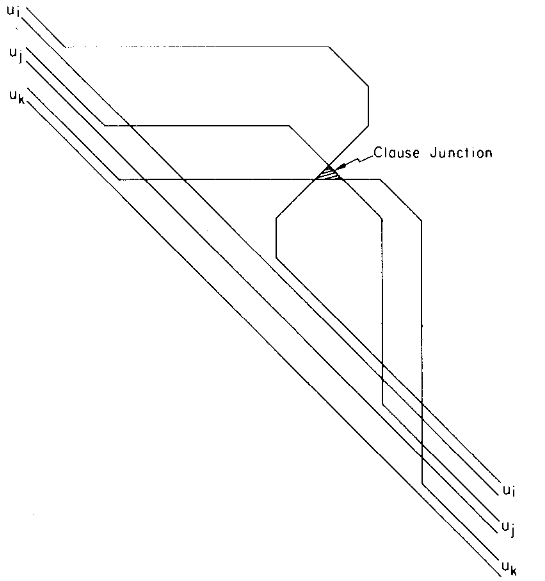


Fig. 10. General arrangement of variable loops and sample clause junction.

loops to also cover the central triangle. Finally, no two distinguished points belonging to different variable loops can be covered by a single piece. Since every piece in a minimum cover must include two distinguished points, there is no interference in their independent coverage properties. □

C. Complete Construction of Polygonal Region

The overall structure of the polygonal region constructed for a given instance of 3SAT consists of n variable loops arranged in parallel slanting columns, one for each Boolean variable in U, with m clause junctions placed to the right, one for each clause in C. Arms of the three variable loops corresponding to the three literals that participate in a clause are brought across the other loops to the right, bent in 45 deg increments until they are oriented properly for the chosen triangle side and according to their complemented/uncomplemented status in the clause, brought to the clause triangle as illustrated previously, and finally returned to their proper slanting columns. A schematic example is shown in Fig. 10.

Although the details are complicated, the entire construction can clearly be performed mechanically using the bend, crossover, and clause junction patterns shown previously. The construction requires no more than $O(m)$ bends, $O(mn)$ crossovers, and $O(m)$ clause junctions, so the execution time of the entire procedure is polynomial bounded by $O(mn)$. Note again that since all of the patterns use integer coordinates, the vertices of the final polygonal region will all have integer coordinates. These observations imply the following proposition.

Proposition 5: The construction of the polygonal region requires only polynomial time.

Recall that the bound K was defined to be half the total number of distinguished points. Our argument to this point has shown the following.

Proposition 6: A given set of clauses is satisfiable iff there is a σ-decomposition of the constructed polygonal region into K pieces.

We may finally state our main result.

Theorem: The problem MσCMP, for each $\sigma \in \Sigma =$ {convex, star-shaped, spiral}, is NP-hard.

Proof: Propositions 5 and 6 establish that 3SAT is polynomial transformable to MσCMP. Since 3SAT is known to be NP-complete, MσCMP is NP-hard for each $\sigma \in \Sigma$. □

Corollary: MσCMP *without Steiner points* is NP-hard for each $\sigma \in \Sigma$.

Proof: No Steiner points are needed in any of our constructions. □

IV. Discussion

Our results answer only a small portion of the complexity questions on minimum polygon decompositions. For example, because the construction of the previous section results in a multiply connected polygonal region, our proofs say nothing about decompositions of simply connected regions. In addition, aside from the many variants discussed in the Introduction, there is an entire class of problems involving *sum / difference* decompositions. In these decompositions, primitive components may be subtracted (set difference) as well as added (set union). For example, if each C_i is a convex polygon, the expression $P = C_1 + C_2 - C_3 + C_4 - C_5 - C_6$ represents a convex sum/difference decomposition. Very often, minimum sum/difference decompositions result in both fewer and more "meaningful" pieces; see [3], [39, p. 1341], [6, p. 97].

Table I summarizes the current state of knowledge concerning minimum polygons decompositions. As one would expect, the more restrictive the class of polygons, and the more specialized the shape of the pieces, the more likely it is that the problem can be solved in polynomial time. Several outstanding open problems are indicated in the table, but the boundary between P and NP for polygon decomposition problems is gradually being clarified. What is sorely needed, however, is a collection of fast near-optimal algorithms for those problems that have been shown to be NP-hard.

Finally, some observations on three-dimensional *polyhedron* decompositions will be mentioned. In analogy to the two-dimensional case, define a *multiply connected polyhedral region* to be a finite union of simple polyhedra. In topological terms, such an object is a three-dimensional connected simplicial complex whose surface may have genus greater than zero [14, pp. 95–96]. Trivially, the results of this paper imply that the minimum decomposition of these regions into convex or star-shaped regions is NP-hard: the reduction is the same as for the two-dimensional cases, except that the region constructed in the reduction has a constant (say zero) for each point in the third dimension. (There is no obvious counterpart to a spiral polygon in

TABLE I
COMPLEXITY OF POLYGON DECOMPOSITION PROBLEMS

Decomposition Type	Class of Polygons	Shape of Components	Complexity
partition	multiply connected	convex	NP-complete [18]
		all others	?
partition	simply connected	convex	$O(n^3)$ [6]
		all others	?
partition	rectilinear with point holes	rectangular	NP-complete [18]
partition	simply connected rectilinear	rectangular	$O(n^3)$ [24]
cover	multiply connected	convex	NP-hard
		star-shaped	NP-hard
		spiral	NP-hard
		all others	?
cover	simply connected	all	?
cover	multiply connected rectilinear	rectangular	NP-complete [20]
cover	"convex" rectilinear	rectangular	polynomial [4]
sum/difference	all	all	?

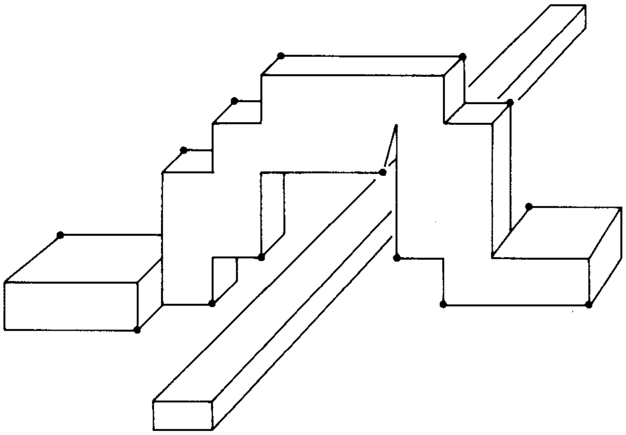


Fig. 11. Avoiding crossover in three dimensions in convex case.

three-dimensions.) Now define a *partition* of a polyhedral or polygonal region as a decomposition in which the composing pieces are pairwise nonoverlapping, that is, the intersection of any pair has measure zero. As discussed in the Introduction, Chazelle and Dobkin have shown that in the two-dimensional case, minimum convex partitions of simply connected polygons can be found in polynomial time [7], [6]. We conjecture, however, that minimum convex partition of three-dimensional multiply connected polyhedral regions is NP-hard. To support this conjecture, note that in our proof for two-dimensional convex decomposition, the only place where we needed overlap between the composing polygons was at crossovers. In three dimensions, however, it seems that crossovers can be avoided by one variable loop going up and over another variable loop, while still preserving the desired properties, as suggested by Fig. 11. We must consider this a conjecture, rather than a result, since a considerable amount of computation is still necessary in order to verify that the distinguished points as indicated in the figure satisfy the desired property that two of them are mutually visible if and only if they are consecutive. If the conjecture turns out indeed to be true, then it would raise the following intriguing open question: is this problem, as opposed to Chazelle and Dobkin's hard because of the added dimension, or because the regions are multiply connected rather than simply connected? Perhaps neither is the case: both of these added features may be necessary to make it hard. If the hardness is due to the added dimension, then the multiply connected convex partition problem would be the first problem of which we are aware that is polynomial in two dimensions but NP-hard in three.

Final Note: During final revisions we learned that Lingas [18] has proven that two- and three-dimensional convex partition is NP-hard, using a result of Lichtenstein [17]. Related results are discussed in [15].

ACKNOWLEDGMENT

We thank Godfried Toussaint for his help in preparing this paper, and an anonymous referee for several useful comments, including the observation that our constructions do not require Steiner points.

REFERENCES

[1] N. Ahuja, R. T. Chien, R. Yen, and N. Birdwell, "Interference detection and collision avoidance among three dimensional objects," in *Proc. 1st Annu. Nat. Conf. Artificial Intelligence*, Stanford, CA, 1980, pp. 44–48.

[2] D. Avis and G. T. Toussaint, "An efficient algorithm for decomposing a polygon into star-shaped polygons," *Pattern Recognition*, vol. 13, pp. 295–398, 1981.

[3] B. G. Batchelor, "Hierarchical shape description based upon convex hulls of concavities," *J. Cybern.*, vol. 10, pp. 205–210, 1980.

[4] S. Chaiken, D. J. Kleitman, M. Saks, and J. Shearer, "Covering regions by rectangles," *SIAM J. Algebraic Discrete Methods*, vol. 2, pp. 394–410, Dec. 1981.

[5] B. M. Chazelle, "Convex decomposition of polyhedra," in *Proc. 13th ACM Symp. Theory of Computing*, Milwaukee, WI, 1981, pp. 70–79.

[6] ——, "Computational geometry and convexity," Ph.D. dissertation, Carnegie-Mellon Univ., Tech. Rep. CMU-CS-80-150, 1980.

[7] B. M. Chazelle and D. Dobkin, "Decomposing a polygon into its convex parts," *Proc. 11th ACM Symp. Theory of Computing*, Atlanta, GA, pp. 38–48, 1979.

[8] V. Chvàtal, "A combinatorial theorem in plane geometry," *J. Combinatorial Theory B*, vol. 18, pp. 39–41, 1975.

[9] H-Y. F. Feng and T. Pavlidis, "Decomposition of polygons into simpler components: feature generation for syntactic pattern recognition," *IEEE Trans. Comput.*, vol. C-24, pp. 636–650, June 1975.

[10] S. Fisk, "A short proof of Chvàtal's watchman theorem," *J. Combinatorial Theory B*, vol. 24, p. 374, 1978.

[11] R. J. Fowler, M. S. Paterson, and S. L. Tanimoto, "Optimal packing and covering in the plane are NP-complete," *Inform. Processing Lett.*, vol. 12, pp. 133–137, Apr. 1981.

[12] ——, "The complexity of covering and packing in the plane and related intersection graph problems," Dep. Comp. Sci., Univ. of Washington, Tech. Rep. 80-05-02, May 1980.

[13] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness.* San Francisco, CA: Freeman, 1979.

[14] P. J. Giblin, *Graphs, Surfaces and Homology.* London: Chapman and Hall, 1977.

[15] D. S. Johnson, "The NP-completeness column," *J. Algorithms*, vol. 3, pp. 182–195, June 1982.

[16] R. M. Karp, "Reducibility among combinatorial problems," in *Complexity of Computer Computations*, R. E. Miller and J. W. Thatcher, Eds. New York: Plenum, 1972, pp. 85–103.

[17] D. Lichtenstein, "Planar formula and their uses," *SIAM J. Comput.*, vol. 11, pp. 329–343, May 1982.

[18] A. Lingas, "The power of non-rectilinear holes," in *Proc. 9th Colloquium Automata, Languages, and Programming*, Aarhus, 1982.

[19] D. Marr and H. K. Nishihara, "Representation and recognition of the spatial organization of three dimensional shapes," Mass. Inst. Technol., Cambridge, Tech. Rep. AIM 416, May 1977, especially pp. 18–19.

[20] W. J. Masek, "Some NP-complete set covering problems," unpublished manuscript, Aug. 1979; quoted in [13, p. 232].

[21] K. Maruyama, "A study of visual shape perception," Dep. Comput. Sci., Univ. of Illinois, Tech. Rep. UIUCDCS-R-72-533, Oct. 1972.

[22] J. O'Rourke, "Polygon decomposition and switching function minimization," *Comput. Graphics and Image Processing*, vol. 18, pp. 382–391, Apr. 1982.

[23] ——, "The complexity of computing minimum convex covers for polygons," in *Proc. 20th Allerton Conf.*, Allerton, IL, Oct. 1982.

[24] L. Pagli, E. Lodi, F. Luccio, C. Mugnai, and W. Lipski, "On two dimensional data organization 2," *Fundamenta Informaticae*, vol. 2, pp. 211–226, 1979.

[25] T. Pavlidis, "Analysis of set patterns," *Pattern Recognition*, vol. 1, pp. 165–178, 1968.

[26] ——, "Representation of figures by labelled graphs," *Pattern Recognition*, vol. 4, pp. 5–17, 1972.

[27] ——, "Structural pattern recognition: Primitives and juxtaposition relations," in *Frontiers of Pattern Recognition*, S. Watanabe, Ed. New York: Academic, pp. 421–451, 1972.

[28] ——, *Structural Pattern Recognition*, Berlin-Heidleberg-New York: Springer-Verlag, 1977, pp. 236–241.

[29] ——, "Survey: A review of algorithms for shape analysis," *Comput. Graphics and Image Processing*, vol. 7, pp. 243–258, 1978.

[30] T. Pavlidis and H-Y. F. Feng, "Shape discrimination," in *Syntactic Pattern Recognition, Applications*, K. S. Fu, Ed. New York: Springer-Verlag, 1977, pp. 125–145.

[31] F. P. Preparata and K. J. Supowit, "Testing a simple polygon for monotonicity," *Inform. Processing Lett.*, vol. 12, pp. 161–164, Aug. 1981.

[32] B. Radig, R. Kraasch, and W. Zach, "Matching symbolic descriptions for 3D reconstruction of simple moving objects," in *Proc. 5th Int. Conf. Pattern Recognition*, Dec. 1980, pp. 1081–1084.

[33] B. Schacter, "Decomposition of polygons into convex sets," *IEEE Trans. Comput.*, vol. C-27, pp. 636–650, 1975.

[34] M. I. Shamos, "Computational geometry," Ph.D. dissertation, Yale University, New Haven, CT, 1978.

[35] L. G. Shapiro and R. M. Haralick, "Decomposition of two-dimensional shapes by graph-theoretic clustering," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-1, pp. 10–20, Jan. 1979.

[36] ——, "Structural descriptions and inexact matching," Virginia Polytechnic Inst., Tech. Rep. #CS7901-R, Nov. 1979.

[37] ——, "Algorithms for inexact matching," in *Proc. 5th Int. Conf. Pattern Recognition*, Dec. 1980, pp. 202–207.

[38] K. J. Supowit, "Topics in computational geometry," Ph.D. dissertation, Univ. of Illinois, Urbana-Champaign, Apr. 1981.

[39] G. T. Toussaint, "Pattern recognition and geometrical complexity," in *Proc. 5th Int. Conf. Pattern Recognition*, Dec. 1980, pp. 1324–1347.

Error Probabilities for Simple Substitution Ciphers

ANDREA SGARRO

Abstract—**Unlike recent works by Blom and Dunham on simple substitution ciphers, papers, we do not consider equivocations (conditional entropies given the cryptogram) but rather the probability that the enemy makes an error when he tries to decipher the cryptogram or to identify the key by means of optimal identification procedures. This approach is suggested by the usual approach to coding problems taken in Shannon theory, where one evaluates error probabilities with respect to optimal encoding–decoding procedures. The main results are asymptotic; the same relevant parameters are obtained as in Blom or Dunham.**

I. Introduction

IN THIS PAPER we consider simple substitution ciphers. A stationary memoryless source outputs letters taken from a finite alphabet $\mathcal{A} = \{a_1, a_2, \cdots, a_s\}$, $s \geq 2$, according to a probability distribution $P = \{p_1, p_2, \cdots, p_s\}$. The resulting flow of information is enciphered to protect it from successful wire tapping: a permutation (one-to-one transformation) of $\mathcal{A}$, called the *key*, is chosen by the transmitter and communicated to the legitimate user via a secure special channel. Prior to (noiseless) transmission over the insecure normal channel, each source letter is changed using the key; the enemy intercepts the enciphered string and tries to recover the original information without knowledge of the key.

This rough sketch of simple substitution ciphers is made mathematically more tractable by some additional specifications. Attention is paid to a source string of length n (the *message*) and to the corresponding enciphered string (the *cryptogram*). The key is chosen randomly with uniform distribution out of the $s!$ keys which are *a priori* possible; the random message and the random key are assumed to be independent. Rather pessimistically, the enemy is as-

Manuscript received November 10, 1981; revised May 11, 1982. This work is part of the research program of the Gruppo Nazionale per l'Informatica Matematica of the Italian Consiglio Nazionale delle Ricerche. This paper was presented in part at the IEEE Symposium, Les Arcs, France, June 1982.

The author is with the Istituto di Matematica and Istituto di Elettrotecnica ed Elettronica, Università di Trieste, Italy.

0018-9448/83/0300-0190$01.00 ©1983 IEEE