Computer Science Department

TECHNICAL REPORT

On the Piano Movers' Problem:

III. Coordinating the Motion of Several Independent Bodies: The Special Case of Circular Bodies Moving Amidst Polygonal Barriers

by

Jacob T. Schwartz
and
Micha Sharir*

Technical Report No. 52
September 1982

NEW YORK UNIVERSITY

Department of Computer Science
Courant Institute of Mathematical Sciences
251 MERCER STREET, NEW YORK, N.Y. 10012

On the Piano Movers' Problem:

III. Coordinating the Motion of Several Independent Bodies: The Special Case of Circular Bodies Moving Amidst Polygonal Barriers

by

Jacob T. Schwartz
and
Micha Sharir*

Technical Report No. 52
September 1982

* School of Mathematical Sciences, Tel Aviv University

Work on this paper has been supported by ONR Grants N00014-75-C-0571 and N00014-82-K-0381, and by a grant from the U.S.-Israeli Binational Science Foundation.

On the 'Piano Movers' Problem
III. Coordinating the Motion of Several Independent Bodies: The Special Case of Circular Bodies Moving Amidst Polygonal Barriers.(*)

Jacob T. Schwartz
Computer Science Department
Courant Institute of Mathematical Sciences

and

Micha Sharir
School of Mathematical Sciences
Tel Aviv University

September 1982

ABSTRACT: We present an algorithm that solves the following motion-planning problem which arises in robotics: Given several 2-dimensional circular bodies B_1, B_2, ..., and a region bounded by a collection of 'walls', either find a continuous motion connecting two given configurations of these bodies during which they avoid collision with the walls and with each other, or else establish that no such motion exists. This paper continues other studies by the authors on motion-planning algorithms for other kinds of moving objects. The algorithms presented are polynomial in the number of walls for each fixed number of moving circles (for two moving circles the algorithm is shown to run in time $O(n^3)$ if n is the number of walls), but with exponents increasing with the number of moving circles.

(*) Work on this paper has been supported by ONR Grants N00014-75-C-0571 and N00014-82-K-0381, and by a grant from the U.S.-Israeli Binational Science Foundation.

0. Introduction

The 'Piano Movers' problem (see [Re], [LPW], [IKP], [Ud], [SS1], [SS2], [HJW]) is that of finding a continuous motion that will take a given body (or a group of bodies) from a given initial configuration to a desired final configuration, but which is subject to certain geometric constraints during the motion. These constraints forbid the body to come in contact with certain obstacles or 'walls', and, in the case of a coordinated motion of more than one body, also forbid individual bodies to come into contact with each other. In a preceding paper [SS1] we have analyzed the following simplified two dimensional version of this problem: Let B be a given two dimensional polygonal body, and let V be a two dimensional open region bounded by a collection of closed polygonal curves, which we consider to be 'walls'. These walls are allowed to intersect at certain 'corners', and the full collection of walls is not required be connected. The problem solved in [SS1] was: Given two positions and orientations in which the body B does not touch any walls, find a continuous wall-avoiding motion of B between these two positions, or establish that no such motion exists. In a subsequent paper [SS2] we have generalized the problem to the case of an arbitrary number of moving bodies, some of which may be jointed, moving amidst obstacles, provided that the surfaces of the bodies and of the obstacles can be described by algebraic equations. We demonstrated in [SS2] that this general problem can be solved in time polynomial in the number of smooth surfaces of the walls and the bodies, and in the maximal degree of the equations defining them, but exponential in the number of degrees of freedom of the system of bodies. However, even for a fixed number of degrees of freedom, the algorithm presented in [SS2], although polynomial, is of complexity $O(n^e)$, where the exponent e can be quite high. Accordingly, this general algorithm is entirely impractical except possibly for the simplest cases. It therefore remains important to develop more efficient specialized algorithms for specific systems of bodies.

In this paper we consider such a special case, namely that in which the coordinated motion of several disjoint, independent circular bodies B_1, B_2,..., B_k must be planned. We assume that these bodies are free to move inside a two-dimensional region V of the kind described above, subject of course to the constraint that they do not collide with the walls or with each other. In this case, the motion planning problem becomes: Given two configurations of the bodies (where a 'configuration' is defined to be the k-tuple $[X_1,\ldots,X_k]$ of the positions of the centers of B_1 thru B_k in V) in which they touch neither any wall nor each other, construct a continuous coordinated motion of B_1,..., B_k between these two configurations during which the bodies continually avoid collision with the walls and with each other (cf. Fig. 0.1 for an instance of this problem involving two circular bodies).

This problem can be viewed as a simplified prototype of other more realistic problems involving the coordinated motion of several bodies.

We will begin our analysis by considering the special case of two circular bodies, then go on to attack the somewhat more complicated case involving three circular bodies, and finally comment on attacking the general case of k bodies using recursive methods.

One might expect the motion-planning problem for two circles to be harder than the corresponding problem for one polygonal body considered in [SS1]. Indeed, the problem considered in [SS1] involves only three degrees of freedom, whereas the motion of two circular bodies involves four degrees of freedom. It turns out however that we can give a relatively simple solution for the problem of two moving circles, by an algorithm of complexity lower than that of the algorithm presented in [SS1].

As in [SS1], our approach is based on a study of the space FP of all collision-free configurations of B_1 and B_2. The problem is to decompose FP into its connected components. In order to construct such a decomposition, we separate the original 4-dimensional problem into

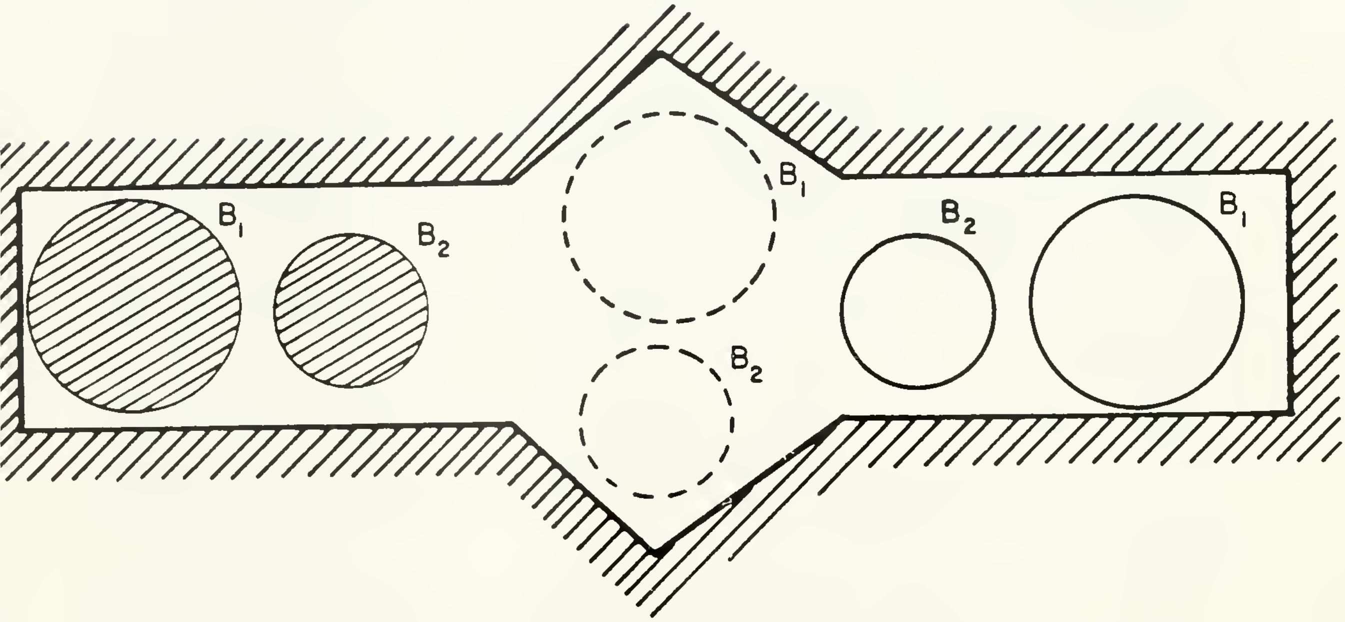

Fig. 0.1. An instance of our case of the piano movers' problem. The shaded circles describe the initial configuration of B_1 and B_2, and the unshaded circles describe the desired final configuration; the intermediate dotted positions describe a possible motion of B_1 and B_2 between the initial and final configurations.

two 2-dimensional subproblems by projecting FP onto a two-dimensional space, each point of which corresponds to a fixed position X_1 of the center of B_1 in V. This leaves B_2 free to move over a subspace A of V that can be decomposed into connected components which can be assigned standard labels. For most positions of the center of B_1 the set of connected components of A changes only slightly and quantitatively if B_1 is moved slightly, but for certain 'critical' positions X of B_1 this set of components changes qualitatively if B_1 moves in the neighborhood of X. These critical positions lie along curves, called 'critical curves', that can be characterized as follows: A critical curve is a locus of points X such that if B_1 is placed in V with its center C_1 at X, then a 'critical' contact between B_1, B_2 and the walls will occur at some position of B_2. More specifically, at such a critical contact, either B_2 touches both B_1 and some wall at diametrically opposite points of B_2, or B_2 touches B_1 and two other walls, etc.

The critical curves introduced in this way divide V into 'noncritical' subregions R. It is then easy to state necessary and sufficient conditions for two configurations, for each of which C_1 lies in the same noncritical subregion R, to be reachable from each other via a continuous coordinated collision-avoiding motion during which C_1 remains within R (see Lemma 1.5). After stating these conditions we go on to study crossings of C_1 from one noncritical subregion R to another R′, and show that to analyze these completely we have only to concern ourselves with finitely many possible types of crossings (see Lemmas 1.8 and 1.9). These observations enable us to reduce our original motion-planning problem to a finite combinatorial problem described by a finite 'connectivity graph' CG which characterizes all possible inter-region crossings, and to show that two given configurations of B_1 and B_2 are reachable from one another by a continuous collision-avoiding motion if and only if two associated vertices in the connectivity graph CG are reachable from one another in CG.

Once this is done we go on to examine the combinatorial reduction of our originally topological problem in more detail, studying the geometric properties of the critical curves mentioned above. These curves are seen to consist of line segments and circular arcs exclusively. This makes straightforward construction of the connectivity graph possible, as demonstrated in an algorithm presented below. The complexity of this algorithm is analyzed and shown to be $O(n^3)$, where n is the number of distinct convex walls appearing in the original problem.

After this treatment of the two circle problem, we go on to study the three-circle case. Here the manifold FP of free configurations of the system is 6-dimensional, but we can project FP onto the two-dimensional space of positions of the center C_1 of B_1. For each such fixed position X_1, the two remaining circles can move in a space $A(X_1)$ bounded by the original polygonal walls and by the fixed circle B_1. The connected components of the corresponding 4-dimensional space $P(X_1)$ of configurations of B_2 and B_3 can be found by the technique used in the case of two circular bodies, and using our analysis of the two-circle case we can give these components standard labels. Accordingly, we can classify points X_1 as being 'critical' if this labeling of the components of $P(X_1)$ changes discontinuously in the neighborhood of X_1, but as 'noncritical' otherwise. Again, critical points are seen to lie on finitely many 'critical curves' which are either straight segments, circular arcs, plus simple algebraic curves of a few additional kinds having degree at most 4, and the connected regions bounded by these curves are regions in which the connectivity of $P(X_1)$ is constant. As in the two-circle case, we then go on to establish 'crossing rules' which describe the manner in which connected components of $P(X_1)$ change as X_1 crosses a critical curve from one noncritical region to another. These crossing rules are used to construct a connectivity graph and to reduce the original problem to a combinatorial search through this graph, in much the same manner as in the two-circle case.

The approach to the three-circle case which has just been outlined is recursive, and this recursive approach can be generalized to the case of arbitrarily many independent circular bodies. We sketch this general recursive approach briefly, but do not give details, as the complexity of these details would increase rapidly with the number of moving circles involved.

The present paper is organised as follows. Section 1 analyses the two-circle problem and reduces it to combinatorial terms. Section 2 contains additional geometric details pertaining to the solution of this problem, and also sketching and analyzing the algorithm corresponding to this solution. Section 3 handles the three-circle problem, and Section 4 discusses the case of arbitrarily many circles.

1. Analysis of the Two-circle Problem: Topological and Geometric Relationships; The Connectivity Graph.

Let B_1 and B_2 be two circular bodies with centers C_1, C_2 and radii r_1, r_2 respectively. We suppose that $r_1 \geqslant r_2$, and assume that the region V in which B_1 and B_2 are free to move is a two-dimensional open region with compact closure, bounded by finitely many polygonal walls which can be partitioned into a disjoint collection of simple polygonal closed curves. We label the various straight line segments (walls) constituting the boundary of V as W,W′, etc. To avoid various minor technicalities we assume that the complement V′ of V is a two-dimensional region (called the wall region) having the same boundary as V. This excludes cases in which the complement of V contains one-dimensional ′slits′ or isolated points. Furthermore, the assumption that the walls separating V′ from V fall into a disjoint union of closed polygonal curves excludes cases in which a boundary point of V is an inner point of two distinct boundary curves. Assume that B_2 is placed in V with C_2 at some point X. X is called an <u>admissible point</u> if when C_2 lies at this position B_2 does not touch or penetrate any wall. Clearly, a point is admissible iff its distance from the nearest wall is at least r_2. (A point satisfying the somewhat more stringent condition that it lies at distance at least r_1 from any wall is said to be <u>admissible for C_1</u>.) It is plain that the set of admissible points is a finite union of closed connected regions bounded by straight-line segments at distance r_2 from the walls bounding V, and by circular arcs at distance r_2 from convex corners formed by these walls. We call a segment at distance r_2 from a wall a <u>displaced wall</u>, and call a circular arc at distance r_2 from a convex wall corner a <u>displaced corner</u>. We will use A to designate the set of all admissible points, and A_1 to designate the set of all points admissible for the center C_1 of the larger circle B_1.

Since $r_1 \geqslant r_2$, the region in which the center of B_1 is free to move (i.e. the set A_1) is a subset of A (having very similar boundaries.) If the center C_1 of B_1 is put at a point X, then the set of positions Y available to C_2 is the set-theoretic difference of A and the circular domain bounded by the circle $\rho(X)$ whose center is X and whose radius is r_1+r_2. Such a pair [X,Y] of points, i.e. a position X of C_1 and a position Y of C_2 in which B_1 and B_2 neither meet each other nor any wall, is called a <u>free configuration</u> of the bodies B_1 and B_2. Similarly, a pair [X,Y] consisting of a position X of C_1 and a position Y of C_2 is called a <u>semi-free configuration</u> if it is a free configuration, or if, at these positions, either B_1 and B_2 touch (but do not cross) each other, or one of these bodies touches, but does not penetrate into, some wall. The set FP of all free configurations of B_1 and B_2 is plainly an open 4-dimensional manifold, and the set SFP of all semi-free configurations is closed. To simplify our analysis we shall, without seriously restricting the problem, assume that the walls are not arranged in such a manner that either of the circles B_1 and B_2 can touch three points on the walls simultaneously, nor can either circle simultaneously touch two points on the walls at diametrically opposed points, nor can either circle simultaneously touch the walls at two points, one of which is a common endpoint of two walls. These last assumptions imply that no two displaced walls or displaced corners are ever tangent. Hence both A_1 and A are, like V, bounded by a finite collection of simple closed curves.

To analyze the irregularly shaped 4-dimensional manifold SFP, it is convenient to project SFP into a more easily graspable space of fewer dimensions. A natural choice in this case is to project SFP onto the two-dimensional region of admissible positions X for C_1, and then consider the set of positions available to B_2 for each such fixed position of C_1. This leads us to the following initial definition and lemma.

<u>Definition 1.1:</u> (a) For each wall edge W, let $\gamma(W)$ denote the displaced wall W (i.e. the locus of all points at distance r_2 from W.)

(b) For each convex corner E between two wall edges W_1 and W_2, let $\gamma(W_1W_2)$ denote the displaced corner E (i.e. the circular arc at distance r_2 from E which connects $\gamma(W_1)$ and $\gamma(W_2)$.)

(c) For each fixed position X of C_1, P(X) will designate the set of all positions Y available to C_2. That is, P(X) is A - $\rho^*(X)$, where $\rho^*(X)$ is the solid disc of radius r_1+r_2 centered at X.

The following elementary geometric lemma is obvious from this definition, and from the fact that the boundary of A is a set of simple closed curves, each made up of straight and circular arcs.

<u>Lemma</u> <u>1.1:</u> (a) For each X, the set P(X) is the disjoint union of a finite collection of open connected planar regions, each of which is bounded by a finite number of straight edges and circular arcs. For each point Y on the boundary of such a region, the configuration [X,Y] is semi-free.

(b) For each pair $K \neq K'$ of connected components of P(X), the boundary of K can meet the boundary of K′ only at a common corner, which is necessarily a point at which $\rho(X)$ is tangent to some displaced wall or corner.

(c) Every straight edge or circular arc in the boundary of a component K of P(X) is either contained in one of the displaced walls or corners, or is an arc of the circle $\rho(X)$. Every corner of K is convex, i.e. the interior angle of K at that corner is less than 180 degrees.

If the exterior boundary E of a component of P(X) does not intersect $\rho(X)$, then E consists of a sequence of displaced walls and corners, and can be obtained by starting at any segment S_1 on E, following S_1 until its intersection with a subsequent displaced wall or corner S_2, and so forth until we have traced out a simple closed curve. If E intersects $\rho(X)$, then its intersection with $\rho(X)$ is a set of arcs of $\rho(X)$ (some of which may degenerate into points) and E consists of these arcs and of various connected displaced wall/corner portions which do not lie interior to $\rho(X)$. These displaced wall and corner portions are also straight line segments and circular arcs, and the boundary E can still be obtained by tracing along these segments and arcs, and along arcs of $\rho(X)$, in the manner just outlined. Note that several separate arcs belonging to a single displaced wall or corner, or to $\rho(X)$, can appear in E. We label the exterior boundary E (and also the component K of P(X) which it bounds) with the circular sequence of displaced walls, corners, and arcs of $\rho(X)$ to which each boundary segment of E belongs, and arrange this sequence in the order in which these boundary segments appear on E as E is traversed with K to its right. The appearance in E of an arc of ρ is indicated simply by including the symbol ρ in the labeling sequence. An example of this is shown in Fig. 1.1.

Definition 1.2: The circular sequence of displaced walls and corners and $\rho(X)$ containing the portions of the exterior boundary of a component K of P(X) arranged in the order of traversal outlined above is called the labeling of K, and is written $\lambda(K)$.

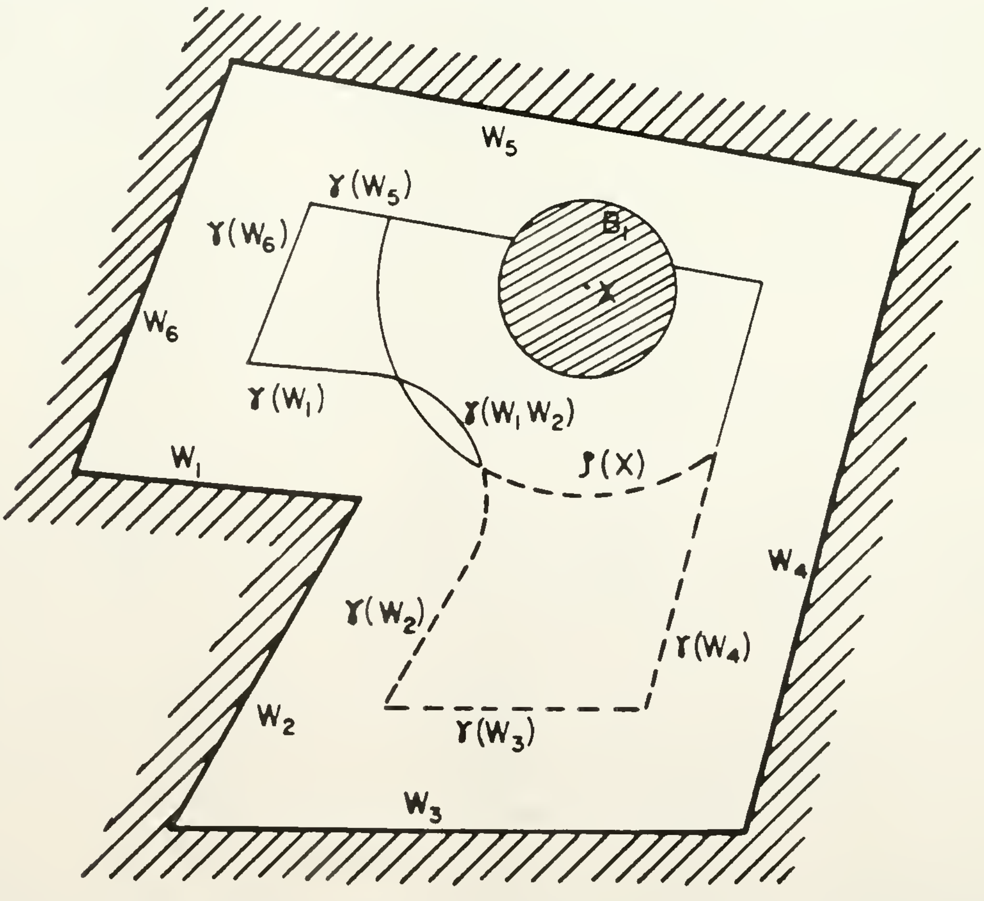

Fig. 1.1. Displaced walls and the circle ρ. The dashed connected component of P(X) has the labeling $[\gamma(W_1W_2),\gamma(W_2),\gamma(W_3),\gamma(W_4),\rho]$.

Remark: A connected component K of P(X) need not be simply-connected. In multiply-connected cases it will be convenient in what follows to ignore the interior boundaries of K and to derive the labeling of K from its exterior boundary only.

Suppose that $\rho(X)$ does not pass through any point at which two displaced walls or corners meet, and moreover that P(X) is not tangent to any displaced wall or corner $\gamma(W)$ (so that, if $\rho(X)$ intersects a curve $C = \gamma(W)$ or $C = \gamma(W_1W_2)$, then $\rho(X)$ and C are transversal at their point(s) of intersection.) Then it is clear that if X undergoes a sufficiently small displacement, the displaced walls and corners which $\rho(X)$ intersects remain exactly the same, also that all points of intersection between $\rho(X)$ and these displaced walls/corners move only slightly, and finally that the arcs into which these points divide $\rho(X)$ and these displaced walls/corners change only slightly. (That is, all these geometric objects depend continuously on X in a sufficiently small neighborhood of such a point.) Hence the number of components of P(X), and also the labeling of these components, remains unchanged when such a point X undergoes a small displacement.

The following definition and lemma capture these basic facts and a few others.

Definition 1.3: A point X is called critical if $\rho(X)$ either passes through some point at which two displaced walls or corners meet, or is tangent to a displaced wall or corner at some point. If X is not critical, it is called noncritical.

Lemma 1.2: (a) The set of critical points is closed, and consists of the union of a finite collection of curves of the three following kinds:

(i) Straight-line segments which lie at distance r_1+2r_2 from a wall W (i.e. segments at distance r_1+r_2 from the displaced wall $\gamma(W)$);

(ii) Circular arcs which lie at distance r_1+2r_2 from a convex corner at which two walls W_1, W_2 meet (i.e. arcs at distance r_1+r_2 from the displaced wall corner $\gamma(W_1W_2)$);

(iii) Circular arcs which lie at distance r_1+r_2 from a convex corner at which two displaced walls or corners meet.

(b) Removal of the closed set of critical points decomposes the set A_1 of admissible locations of the center C_1 of B_1 into a finite number of disjoint connected open regions $R_1, R_2, \ldots$, (which we will call the noncritical regions of the present case of our movers' problem). As X varies in such a region R, both the number of components of P(X), and the labeling of each of these components, remain invariant, and these components remain at positive distance from each other.

Proof: Part (a) is obvious from the definition of critical points. Part (b) is also trivial, and follows from part (a) and from the preceding discussion. Q.E.D.

The following lemma shows that each connected component of P(X) is defined uniquely by the labeling which we have assigned to it (and even by a small portion of this labeling).

Lemma 1.3: Let X be a noncritical point admissible for C_1. Let K,K′ be two connected components of P(X), and suppose that there exist two boundary segments δ_1, δ_2, both of which are portions of a displaced wall, a displaced corner, or of $\rho(X)$, and that δ_1, δ_2 appear as consecutive components in both circular sequences $\lambda(K)$, $\lambda(K')$. Then K = K′. In particular, if $\lambda(K) = \lambda(K')$ (up to a circular shift) then K = K′.

Proof: By definition, the two curves identified by δ_1 and δ_2 must meet at a corner D of K and at a corner D′ of K′. By the final statement of Lemma 1.2(b) we must have $D \neq D'$, since K and K′ are at positive distance from each other. Since δ_1 and δ_2 are both either straight or circular arcs, they can have at most two points of intersection, which must therefore be D and D′ respectively. It follows from the definition of λ that the curve δ_2 follows the curve δ_1 as we trace the boundary of K (resp. K′) in the vicinity of D (resp. D′) with K (resp. K′) remaining to the right. However, since all the corners of both K, K′ are convex (cf. Lemma 1.1(c)) this can happen at one of the points D,D′ but not at both. This contradiction implies that K = K′. Our second assertion follows immediately, since both sequences $\lambda(K)$, $\lambda(K')$ label the exterior boundary of a region bounded by straight lines and circles, which cannot consist of a single circle since P(X) always lies on the convex (outer) side of any circular portion of its boundary. Thus $\lambda(K)$ and $\lambda(K')$ must be sequences of length at least 2. Q.E.D.

Definition 1.4: Let X be a noncritical point admissible for C_1. Define $\sigma(X)$ to be the set

$$\{\lambda(K) : K \text{ a connected component of } P(X)\}$$

Let $T \varepsilon \sigma(X)$, and let S be a contiguous subsequence of T containing at least two curve labels. (Two such subsequences will be called <u>equivalent</u> at X if they are both subsequences of the same circular label sequence $\lambda(K)$.) We let $\psi(X,S)$ denote the unique connected component K (cf. Lemma 1.3) of P(X) for which $\lambda(K)$ contains S.

The following lemma is an immediate consequence of Lemma 1.3, Definition 1.4, and of the observations made in the paragraph preceding Lemma 1.2.

Lemma 1.4: Let R be a connected open noncritical region of points admissible for C_1. For all $X \varepsilon R$ the sets $\sigma(X)$ are identical, and for each T belonging to such a set $\sigma(X)$, $X \varepsilon R$, the function $\psi(X,T)$ is continuous (in the Hausdorff topology of sets) for $X \varepsilon R$.

Definition 1.5: For each noncritical region R we put $\sigma(R) = \sigma(X)$, where X is a point chosen arbitrarily from R.

As already noted (cf. Lemma 1.2), the critical curves of our problem fall into the two following categories.

Type I: For each wall edge W (resp. for each pair W_1, W_2 of adjacent wall edges) the locus of all points at distance r_1+2r_2 from W (resp. from the corner at which W_1 and W_2 meet) is a type I critical curve.

Type II: Let δ_1,δ_2 be a pair of displaced walls or corners which intersect at a point D, and suppose that D is an admissible point (for C_2). Then the circle at radius r_1+r_2 about D is a type II critical curve.

Note that this category of critical curve also includes the case in which δ_1 and δ_2 meet at a common endpoint at 180 degrees (i.e. when δ_1 is a displaced wall and δ_2 is a displaced endpoint of that wall, or vice versa); the curve β' in Fig. 1.3 is an example of such a curve.)

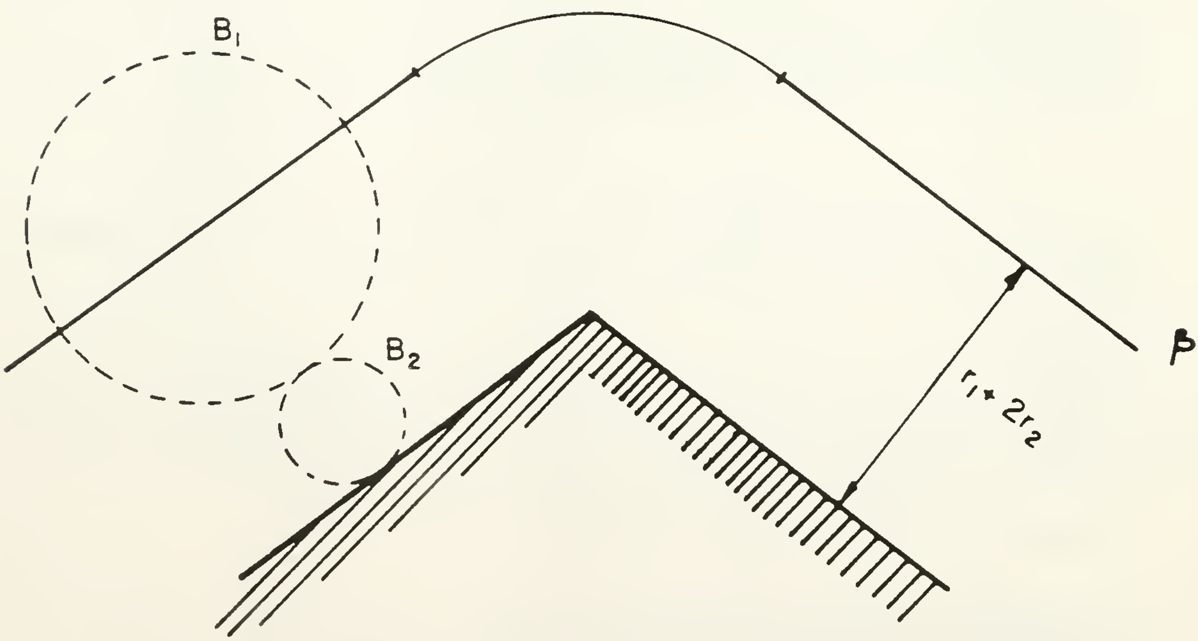

Fig. 1.2. Type I critical curves.

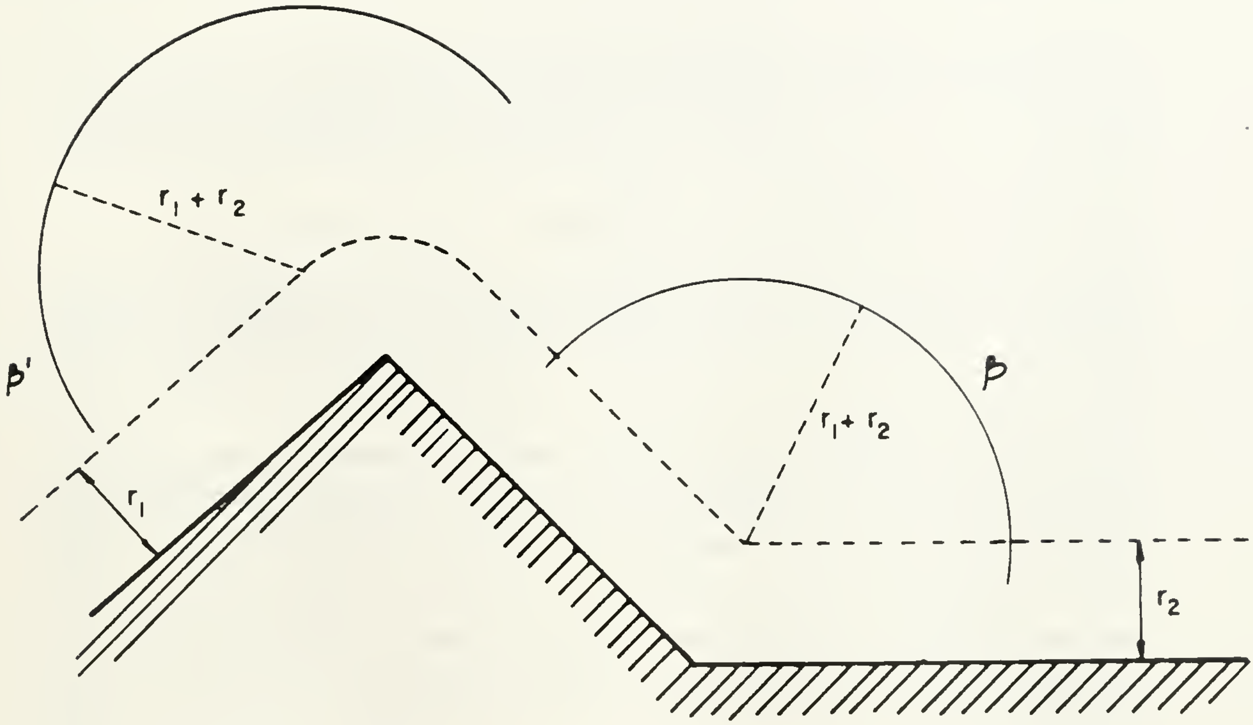

Fig. 1.3. Type II critical curves.

These two types of critical curves will be analyzed in greater detail in the next section, where a way of treating various degenerate cases will also be described.

Lemma 1.5: Let R be a connected open noncritical subregion of the set A_1 of all admissible points for C_1. Suppose that X and X′ are both points in R. Then one can move continuously through FP from a given free configuration [X,Z] to another such configuration [X′,Z′], via a motion during which C_1 remains in R, if and only if the connected component $\kappa(Z,X)$ of P(X) to which Z belongs has a labeling $\lambda(\kappa(Z,X))$ equal to the labeling $\lambda(\kappa(Z',X'))$ of the connected component $\kappa(Z',X')$ of P(X′) to which Z′ belongs.

Proof: It is clear from Lemma 1.4 that $\kappa(U,Y)$ changes continuously as [Y,U] moves continuously through FP with Y remaining in R, so that the label $\lambda(\kappa(U,Y),Y)$ cannot change during such a motion. This proves the 'only if' part of the present lemma.

For the converse, put $\Lambda = \lambda(\kappa(Z,X)) = \lambda(\kappa(Z',X'))$. Take a curve $c(t)$, $0 \leq t \leq 1$, that connects X to X′ in R, and consider the mapping

$$f(t) = \psi(c(t),\Lambda), \qquad 0 \leq t \leq 1,$$

which, by Lemma 1.4, is continuous. Note that the connected component $f(t)$ varies with t only due to the motion of $\rho(c(t))$. Choose some fixed t_0 in the interval $0 \leq t_0 \leq 1$. Then the boundary of the connected component $K = f(t_0)$ of $P(c(t_0))$ must contain a point U at which two displaced walls or corners meet. Indeed, as observed previously, the boundary of K consists entirely of arcs of $\rho(c(t))$ and of such displaced walls and corners, all of which are straight line segments and circular arcs, and K lies on the convex (outer) side of any circular arc of its boundary. Thus the boundary of K must include at least two arcs other than $\rho(c(t_0))$, and hence must have a corner U outside $\rho(c(t_0))$. By Lemma 1.2, any value $c(t)$ for which $\rho(c(t))$ passed through U would lie on a critical curve, contradicting the hypotheses of the present lemma. It follows that this can never happen, i.e. U must be a corner of the boundary of all the connected components $f(t)$, $0 \leq t \leq 1$, and during the motion $c(t)$ of C_1 the circle $\rho(c(t))$ never passes through U. Hence there exists a free position V near U inside $f(t)$ for all $0 \leq t \leq 1$. But then the required motion of the two circles B_1 and B_2 can be constructed as follows:

(a) Move B_2 inside $\kappa(Z,X) = f(0)$ from Z to V;

(b) Move B_1 along the curve $c(t)$ from X to X′;

(c) Move B_2 from V to Z′ along a path inside $\kappa(Z',X') = f(1)$. Q.E.D.

<u>Definition 1.6:</u> Let R be a connected open noncritical region. Then
(a) $C(R)$ is the set of all free configurations $[X,Z]$ such that $X \in R$.
(b) For each pair $\xi = [R,\Lambda]$, where $\Lambda \in \sigma(R)$, we define $C(\xi)$ to be the

set of all $[X,Z] \varepsilon C(R)$ such that Z belongs to the connected component $\psi(X,\Lambda)$.

It is obvious from Lemma 1.5 that the connected components of C(R) are the sets $C(\xi)$ of the form defined in (b).

Next we consider what happens when the center C_1 of B_1 crosses between noncritical regions R_1, R_2 separated by a critical curve. The following simple lemma, taken from [SS1], rules out extreme cases that would otherwise be troublesome.

Lemma 1.6: Let $p(t)=[x(t),z(t)]$ be a continuous curve in the open four-dimensional manifold FP of free configurations of B_1 and B_2. Suppose that the end-points [X,Z], [X′,Z′] of p are specified. Let $\{X_1..X_n\}$ be any finite collection of points in the 2-dimensional space V not containing either X or X′. Then by moving p slightly we can assume that, during the motion described by p, C_1 never passes through any of the points $X_1..X_n$.

Proof: The subset of FP for which C_1 lies at one of the points $X_1...X_n$ is a finite union of submanifolds of dimension 2, and these can never disconnect the four-dimensional manifold FP, even locally. (See [Sc]). Q.E.D.

Remark: A similar argument, based on Sard′s lemma (see [Sc]) shows that by modifying any given free motion very slightly, we can always ensure that the curve x(t) traced out by C_1 during the motion p(t) has a nonvanishing tangent everywhere along its length, and that, given any finite set $\beta_1,..,\beta_n$ of smooth curves in two-dimensional space, we can assume that the tangent to x(t) lies transversal to β_j at any point in which x(t) intersects β_j (see [Sc]). Moreover, we can assume that the position z(t) of C_2 is constant and x(t) is linear in t for all points along p lying in a sufficiently small neighborhood of each such intersection.

These observations imply that in order to characterize the connected components of the four dimensional manifold FP, it is sufficient to analyse what happens as $x(t)$ crosses between regions R_1,R_2 along a line L transveral to a critical curve β separating these two regions, such that L does not pass through any point common to two critical curves. Moreover, we can suppose that B_2 maintains a constant position in the neighborhood of each such crossing.

Lemma 1.7: Suppose that (a portion of) the critical curve β forms part of the boundary of a noncritical region R, and that S is a (subsequence consisting of at least two successive components of an) element of $\sigma(R)$. Put $\xi = [R,S]$. Let $X \varepsilon \beta$, and let $Y_n \varepsilon R$ and $Y_n \to X$. Then the (region-valued) sequence $\psi(Y_n,S)$ converges (in the Hausdorff topology of sets) to a unique closed set, which we denote by $\phi(X,\xi)$, whose interior is contained in P(X) and is a union of connected components of P(X). If $T \varepsilon \sigma(R)$, $T \neq S$, and $\eta = [R,T]$, then $\text{int}(\phi(X,\xi))$ and $\phi(X,\eta)$ are disjoint for each $X \varepsilon \beta$. For each $Z \varepsilon V$ the set $\{X \varepsilon \beta : Z \varepsilon \text{int}(\phi(X,\xi))\}$ is open in β.

Proof: As already noted, the boundary curves of any region P(X) are either straight or circular arcs; this observation holds irrespective of whether X is a critical or noncritical point. Moreover, P(X) lies on the outer (convex) side of each of its circular boundary arcs. As previously, we can orient these arcs by imposing the condition that P(X) lies to the right of each oriented boundary arc. Note that this convention gives a counterclockwise orientation to arcs lying on the circle $\rho(X)$. Moreover, $\rho(X)$ can intersect a boundary arc α of A in at most two points, and if both these intersections actually occur one of them will be uniquely defined as the first point of contact of $\rho(X)$ with α, namely the point p for which the arc on $\rho(X)$ containing p is oriented from p, and the other point q will be uniquely defined as the second point of contact, i.e. for which the corresponding arc of $\rho(X)$ is oriented towards q.

Let X be a critical point and let $Y_n \to X$. Then the first (resp. second) point p_n of contact of $\rho(Y_n)$ with any fixed boundary arc α of A converges to the first (resp. second) point p of contact of $\rho(X)$ with α (these latter two points may coincide). The collection of all these limit points p represent some, but not necessarily all, the points of contact of $\rho(X)$ with the boundary of A, since a finite number of additional tangencies of $\rho(X)$ with that boundary can occur as Y_n approaches the limiting position X. The limit set $\phi(X,\xi)$ of the sequence of components $\psi(Y_n,S)$ is bounded by the limiting arcs to which the boundary arcs of $\psi(Y_n,S)$ converge; note that these are arcs of fixed portions of the boundary of A delimited by limiting positions p of points of intersection of $\rho(Y_n)$ with fixed boundary arcs, plus similarly delimited arcs of $\rho(X)$. Moreover, certain of the arcs appearing in the boundary of $\psi(Y_n,S)$ can shrink to points, while other may be split by new points of tangency of $\rho(X)$ with the boundary of A, as $Y_n \to X$.

It is plain from the foregoing that $\mathrm{int}(\phi(X,\xi))$ is contained in $P(X)$. If $\rho(X)$ is tangent to fixed boundary arcs α of A at points which are not limits of intersections of $\rho(Y_n)$ with such arcs, then $\mathrm{int}(\phi(X,\xi))$ need not be connected. However, since any closed arc in $\mathrm{int}(\phi(X,\xi))$ is disjoint from $\rho(Y_n)$ for all sufficiently large n, $\mathrm{int}(\phi(X,\xi))$ contains every point of any component of $P(X)$ which it intersects, so that $\mathrm{int}(\phi(X,\xi))$ is a union of components of $P(X)$.

Next, let $Z \in \mathrm{int}(\phi(X,\xi))$, so that $Z \in P(X)$, and therefore $Z \in FP$. Then there exists a small disc U about Z which is free of points of the boundary of $\phi(X,\xi)$, and hence free of all points on the boundary of $\psi(Y_n,S)$ for all sufficiently large n. Thus, for large n, U is either wholly contained within $\psi(Y_n,S)$ or wholly outside it. However, the case $U * \psi(Y_n,S) = \{\}$ for all large n is impossible, since $Z \in \phi(X,\xi)$ which is the limiting set of $\psi(Y_n,S)$. It follows that $Z \in \psi(Y_n,S)$ for all large n. If $Z \in \phi(X,\eta)$ also, then since $Z \in FP$ it is part of the interior, not the boundary, of $\phi(X,\eta)$, we can argue in the same way to deduce that $Z \in \psi(Y_n,T)$ for all large n, which is impossible since the sets $\psi(Y_n,S)$ and $\psi(Y_n,T)$ are disjoint.

Next we show that for each $Z \in V$ the set $\beta(Z) = \{X \in \beta : Z \in \mathrm{int}(\phi(X,\xi))\}$ is open in β. For this, let X belong to $\beta(Z)$. By what we have just shown, $[X,Z] \in FP$, and hence for all $Y \in R$ lying in a sufficiently small neighborhood U of X the point Z belongs to some set $\psi(Y,[R,T])$ with T <u>fixed</u>. (Indeed, choose U to be an arcwise connected neighborhood of X in R sufficiently small so that $[Y,Z] \in FP$ for all $Y \in U$. Then $U\times\{Z\}$ is a connected subset of $C(R)$, and so must be contained in one of its connected components.) Therefore $Z \in \phi(X,[R,T])$, and since we have just shown that $\phi(X,[R,T])$ and $\mathrm{int}(\phi(X,[R,S]))$ are disjoint if $R \neq S$, we must have $R = S$. It therefore follows that $Z \in \mathrm{int}(\phi(X',\xi))$ for all $X' \in \beta$ sufficiently near X, so that the set $\beta(Z)$ must be open. Q.E.D.

<u>Lemma 1.8:</u> Suppose that (a portion of) a smooth critical curve β separates two connected noncritical regions R_1 and R_2 and that $R_1+R_2+\beta$ is open. Let S_1 (resp. S_2) be a subsequence (containing at least two components) of an element of $\sigma(R_1)$ (resp. $\sigma(R_2)$. Put $\xi_1 = [R_1,S_1]$ and $\xi_2 = [R_2,S_2]$, and let $C_1 = C(\xi_1)$, $C_2 = C(\xi_2)$. Then the following conditions are equivalent:

<u>Condition A:</u> There exists a point $X \in \beta$ such that the open sets $\mathrm{int}(\phi(X,\xi_1))$, $\mathrm{int}(\phi(X,\xi_2))$ have a non-null intersection.

<u>Condition B:</u> There exists a smooth path $c(t) = [x(t),z(t)] \in FP$ which has the following properties:

(i) $c(0) \in C_1$, $c(1) \in C_2$;

(ii) $x(t) \in R_1+R_2+\beta$ for all $0\leq t\leq 1$;

(iii) $x(t)$ crosses β just once, transversally, when $t=t_0$, $0<t_0<1$, and $z(t)$ is constant for t in the vicinity of t_0.

<u>Proof:</u> Suppose first that there exists a path $[x(t),Z]$ in the open 4-dimensional manifold FP of free configurations of B_1 and B_2 satisfying (i) - (iii) of Condition B. (By Lemma 1.6 we can assume

without loss of generality that Z is constant throughout this whole path.) Let K(t) denote the open connected component of P(x(t)) containing Z. Since $c(0) \in C_1$, it follows from Lemma 1.5 that for $t<t_0$ we have $K(t) = \psi(x(t),S_1)$. Similarly, for $t>t_0$ we have $K(t) = \psi(x(t),S_2)$. Moreover since for $t<t_0$ we have $Z \in \psi(x(t),S_1)$, it follows from Lemma 1.7 that $Z \in \phi(X,\xi_1)$. However, since Z is a free position for C_2, and since the boundary of $\phi(X,\xi_1)$ consists of positions which are semi-free but not free, we must have $Z \in \mathrm{int}(\phi(X,\xi_1))$. Similar reasoning applied to $t>t_0$ shows that Z also lies in $\mathrm{int}(\phi(X,\xi_2))$. Hence these interiors have a non-null intersection, thus establishing Condition A.

Next Suppose that Condition A holds. Let $Z \in P(X)$ be a point in the intersection of the sets $\mathrm{int}(\phi(X,\xi_1))$ and $\mathrm{int}(\phi(X,\xi_2))$. Since $[X,Z] \in FP$ and β is a smooth curve with a nonvanishing tangent (as will be shown below), we can draw a short curve x(t) crossing β at X from R_1 to R_2, such that x(t) satisfies (ii) and (iii) of Condition B, and such that for all t the condition $[x(t),Z] \in FP$ is satisfied. It follows from Definition 1.6 and the remark following it that there exist Λ_1,Λ_2 belonging to $\sigma(R_1),\sigma(R_2)$ respectively such that $[x(t),Z] \in C([R_1,\Lambda_1])$ for $t<t_0$, and $[x(t),Z] \in C([R_2,\Lambda_2])$ for $t>t_0$. By Lemma 1.7 we have $Z \in \phi(X,[R_i,\Lambda_i])$ for $i=1,2$. However, since by Lemma 1.7 the interior of $\phi(X,[R_i,\xi_i])$ is disjoint from $\phi(X,[R_i,\xi_i])$ if $\Lambda_i \neq \xi_i$, we must have $\Lambda_i = \xi_i$ for $i=1,2$. Therefore $c(t) \in C(\xi_1)$ for $t<t_0$ and $c(t) \in C(\xi_2)$ for $t>t_0$, showing that Condition B holds. Q.E.D.

Next we show that if condition A of the preceding Lemma holds for one point lying on a portion β' of β not intersected by any other critical curve, this same condition holds for all points of β'. This fact, closely related to similar assertions derived in [SS1], [SS2], allows us to derive crossing rules for β' without having to be concerned with the particular point at which we cross.

Lemma 1.9: Let the smooth critical curve β separate the two noncritical regions R_1 and R_2. Let β' be a connected open segment of β not intersecting any other critical curve, and suppose that $\beta'+R_1+R_2$ is open. Let S_1, S_2, ξ_1, ξ_2 be defined as in Lemma 1.8. Then the set of $X \varepsilon \beta'$ for which the open sets $\mathrm{int}(\phi(X,\xi_1))$ and $\mathrm{int}(\phi(X,\xi_2))$ have a non-null intersection is either all of β' or is empty.

<u>Proof:</u> Let M be the set of all $Y \varepsilon \beta'$ for which the sets $\mathrm{int}(\phi(Y,\xi_1))$ and $\mathrm{int}(\phi(Y,\xi_2))$ have a point in common. Since M is the union of all sets of the form

$$\{Y \varepsilon \beta : Z \varepsilon \mathrm{int}(\phi(Y,\xi_1))\}*\{Y \varepsilon \beta : Z \varepsilon \mathrm{int}(\phi(Y,\xi_2))\} ,$$

for $Z \varepsilon V$, and since by Lemma 1.7 each of these sets is open, it follows that M is open. Hence we have only to show that M is also closed. Suppose the contrary; then there exists an $X \varepsilon \beta'$ such that $\mathrm{int}(\phi(X,\xi_1))$ and $\mathrm{int}(\phi(X,\xi_2))$ are disjoint but for which there also exists a sequence Y_n of points on β' converging to X such that for all n the sets $\mathrm{int}(\phi(Y_n,\xi_1))$ and $\mathrm{int}(\phi(Y_n,\xi_2))$ intersect each other.

By Lemma 1.7, for each $n \geqslant 1$ the sets $\mathrm{int}(\phi(Y_n,\xi_j))$, $j=1,2$, are unions of connected components of $P(Y_n)$. Thus, passing to a subsequence if necessary, we may assume that for each $n \geqslant 1$ both sets $\mathrm{int}(\phi(Y_n,\xi_j))$, $j=1,2$, contain a connected component K_n of $P(Y_n)$ for which $\lambda(K_n)$ is constant. We will prove that these sets K_n converge in the Hausdorff topology of sets to some set $\{D\}$, where D is the intersection of two displaced walls or corners γ_1, γ_2. For this, note that the boundary of all the K_n must contain some fixed corner D at which a certain fixed pair γ_1, γ_2 of displaced walls meet. If the circles $\rho(Y_n)$ intersect both these displaced walls at a sequence of points converging to D, then it is clear that K_n converges to $\{D\}$ in the Hausdorff metric. Suppose therefore that U is a small circular neighborhood of D of radius δ, not intersecting any fixed displaced wall other than γ_1, γ_2, such that either

(a) $\rho(Y_n)$ does not intersect either γ_1 or γ_2 within U, or

(b) $\rho(Y_n)$ intersects one of γ_1, γ_2 (for definiteness, say $_1$) at a sequence of points converging to D, but does not intersect $_2$ within U.

Let p_1 and p_2 be points on γ_1,γ_2 respectively at distance $\delta/2$ from D. In case (a) it is clear that the whole interior of the region bounded by α_1,α_2, and by the circle of radius $\delta/2$ about D, lies outside all the circles $\rho(Y_n)$ and hence belongs to both $\phi(X,\xi_j)$, j=1,2. Similarly, in case (b) the whole part of U lying between α_2 and a circle of the same radius as $\rho(Y_n)$ tangent to α_2 at D and containing α_2 in its exterior, belongs to both $\phi(X,\xi_j)$, j=1,2. Hence in both cases we have

$$\mathrm{int}(\phi(X,\xi_1)) * \mathrm{int}(\phi(X,\xi_2)) \neq \{\},$$

contrary to assumption.

This proves that $K_n \to \{D\}$, and implies that for large n, the circle $\rho(Y_n)$ must intersect both of the arcs γ_1,γ_2, at a sequence of points converging to D, but not identical to D. But then X must lie on the circular type II critical curve β_0 having D as center and radius r_1+r_2. Since $\rho(Y_n)$ does not pass through D, Y_n does not lie on β_0. Hence $\beta \neq \beta_0$, so that X lies on two distinct critical curves, contrary to assumption. This proves that M is closed, and then as noted the lemma follows immediately. Q.E.D.

<u>Corollary</u>: Let R_1,R_2,β be as in Lemma 1.9, and let $T_i \in \sigma(R_i)$, i=1,2. Let $X_1 \in \beta$. Then $\mathrm{int}(\phi(X_1,[R_1,T_1]))$ and $\mathrm{int}(\phi(X_1,[R_2,T_2]))$ have a non-null intersection if and only if there exists an intersecting pair $S = W_1W_2$ of r_2-displaced walls or corners, through which intersection $\rho_{12}(X_1)$ does not pass, which bound an open angle $\alpha < \pi$ (not containing any other r_2-displaced wall) and having the following property: For any (and, equivalently, for every) $X_1' \in R_j$, j=1,2, there exists $\varepsilon > 0$ such that all points interior to α and lying within distance $< \varepsilon$ from its apex belong to a component of $P(X_1')$ whose external boundary has the label T_j.

<u>Proof</u>: First suppose that such a pair S exists, and let X_2 lie in α and be near enough to the apex of α to be disjoint from $\rho_{12}(X_1')$ for all X_1' in a small neighborhood U of X_1. Then it is clear that any

sufficiently small neighborhood V of X_2 will be included in $\psi(X_j',[R_j,T_j])$ if $X_j' \varepsilon U$ and $X_j' \varepsilon R_j$, j=1,2. Hence V is a subset of both $int(\phi(X_1,[R_1,T_1]))$ and $int(\phi(X_1,[R_2,T_2]))$, proving that these sets have a nonempty intersection.

Conversely, suppose that these sets have a nonenpty intersection. Then, by Lemma 1.7, they have a connected component K of $P(X_1)$ in common. Since K lies exterior to each of its bounding circles, some corner of K must lie off $\rho_{12}(X_1)$. At this corner, two displaced walls or corners W_1 and W_2 must meet and bound an angle $\alpha < \pi$; all points X_2 interior to this angle and close enough to its apex belong to K. Take such a point X_2; then it is clear that for X_1' sufficiently close to X_1, $X_2 \varepsilon \psi(X_1',[R_j,T_j])$ if $X_1' \varepsilon R_j$, j=1,2; hence X_2 remains in $\psi(X_1'',[R_j,T_j])$ as long as X_1'' can be connected to X_1' by a path in R_j for which the circle $\rho_{12}(X_1'')$ does not pass through X_2. If X_2 lies near enough to the apex of α, which is a point disjoint from $\rho_{12}(Y_1)$ for all noncritical Y_1, the X_1'' having this property will approximate the whole of R_j. Q.E.D.

As in [SS1], we can now define a finite graph, called the <u>connectivity graph</u> for the case of two independent circular bodies, whose edges describe the way in which the components of the sets C(R) connect as we cross between adjacent noncritical regions R.

<u>Definition 1.7:</u> The connectivity graph CG of an instance of our case of the movers' problem is an undirected graph whose nodes are all pairs of the form [R,T] where R is some connected noncritical region (bounded by critical curves) and where T is an equivalence class of labels which are all subsequences (having length at least 2) of the same element of $\sigma(R)$ (i.e. they all label the same connected component of P(X), for each X ε R). The graph CG contains an edge connecting $[R_1,T_1]$ and $[R_2,T_2]$ if and only if the following conditions hold:

(1) R_1 and R_2 are adjacent and meet along a critical curve β.

(2) For some one of the open connected portions β' of β contained in

the common boundary of R_1 and R_2 and not intersecting any other critical curve, and for some (and hence every) point $X \in \beta'$ there exist $S_1 \in T_1$, $S_2 \in T_2$ such that the sets $\mathrm{int}(\phi(X,[R_1,S_1]))$ and $\mathrm{int}(\phi(X,[R_2,S_2]))$ have a non-null intersection.

We can now state the main result of this section.

Theorem 1.1: There exists a continuous motion c of B_1 and B_2 through the space FP of free configurations from an initial configuration $[X_1,Y_1]$ to a final configuration $[X_2,Y_2]$ if and only if the vertices $[R_1,T_1]$ and $[R_2,T_2]$ of the connectivity graph CG introduced above can be connected by a path in CG, where R_1, R_2 are the noncritical regions containing X_1, X_2 respectively, and where T_1 (resp. T_2) is the marking of the connected component in $P(X_1)$ (resp. $P(X_2)$) containing Y_1 (resp. Y_2).

Remark: We assume here that neither X_1 nor X_2 lies on a critical curve. If either X_1 or X_2 lies on such a curve, we first move X_1 (or X_2) slightly into a noncritical region, and then apply the above theorem.

Proof: Suppose that there exists a path connecting $[R_1,T_1]$ to $[R_2,T_2]$ in CG. Let $[R,T]$, $[R',T']$ be two adjacent nodes along that path. Then Lemma 1.5 implies that $R \neq R'$, and Lemma 1.8 implies that there exists a short path in FP connecting points in $C([R,T])$ to points in $C([R',T'])$. Since by Lemma 1.5 any two points in $C([R,T])$ can be connected to each other by a path in FP, one can construct a path in FP connecting $[X_1,Y_1]$ to $[X_2,Y_2]$ by an appropriate concatenation of 'crossing paths' between two domains $C(\upsilon_1)$, $C(\upsilon_2)$, and of internal paths within such domains.

Conversely, if there exists a path $p(t) = [x(t),z(t)]$ in FP connecting the two configurations $[X_1,Y_1]$ and $[X_2,Y_2]$, then we can assume, using Lemma 1.6 and its corollary, that this path is such that $x(t)$ crosses critical curves only finitely many times, transversally, avoiding intersections between critical curves, and that $z(t)$ is constant near each such crossing. Lemma 1.8 and the definition of CG then imply that by tracing the domains $C(\upsilon)$ through which p passes, one obtains a path in CG connecting $[R_1,T_1]$ and $[R_2,T_2]$. Q.E.D.

2. Additional Geometric and Algorithmic Details.

In this section we will study the critical curves and their associated crossing rules in more detail. As will be shown below, the crossing patterns that can arise are quite similar to those described in [SS1] for a single polygonal body. Specifically, assuming that no two critical curves coincide, exactly one of the three following crossing patterns can arise as we cross a critical curve θ at a point X not lying on any other critical curve (in what follows R_1 and R_2 are the regions lying on the two sides of θ near X, and for specificity we assume that P(X) for $X \tau R_1$ contains at least as many components as P(X) for $X \tau R_2$):

(i) One component of P(X) may shrink to a point, and then disappear, in which case $\xi(R_1)$ consists of all labels in $\xi(R_2)$ plus an extra label marking this component.

(ii) Two connected components of P(X) may join each other at a point as X approaches θ, and then merge with each other as θ is crossed. In this case, $\xi(R_1)$ consists of all the labels in $\xi(R_2)$, plus the labels of the two components that merge as we cross into R_2, less the label of the component into which these two components merge.

(iii) The labeling T_1 of one component of P(X) may change to another labeling T_2 as θ is crossed. In this case, $\xi(R_1)$ and $\xi(R_2)$ differ by just two components, one of which appears in $\xi(R_1)$, the other in $\xi(R_2)$.

The crossing rules then assume the following simple and general form: Connect $[R_1,T]$ to $[R_2,T]$ for each $T \tau \xi(R_1)*\xi(R_2)$, and connect each $[R_1,T_1]$, $T_1 \tau \xi(R_1)-\xi(R_2)$, to each $[R_2,T_2]$, $T_2 \tau \xi(R_2)-\xi(R_1)$.

Next we give additional details concerning the structure of the various critical curves decribed in the section 1, state 'crossing rules' for each type of curve as special cases of the general crossing rule just given, and explain how to deal with degenerate cases in which several critical curves become coincident.

As noted earlier, the critical curves fall into the following two categories.

Type I curves: These consist of concatenated sequences of line segments and circular arcs at distance $d = r_1+2r_2$ from a convex wall section W. Ignoring the exceptional case (treated below) in which there exist two parallel walls exactly 2d apart, we can easily see what happens as we cross such a curve θ at a point X not lying on any other critical curve. Specifically (see Fig. 2.1), when we cross at X from R_1 to R_2, there appears exactly one new position Z at which the stationary boundary curve $\zeta(W)$ and the moving curve $\kappa(X)$ touch each other. If B_1 is placed with its center at a point $X' \tau R_1$ near X, these two curves will not meet, whereas if B_1 is placed with its center at X'' IN R_2 near X, these two curves will intersect each other. This implies that there exists a neighborhood N of Z such that $N*P(X')$ is connected, whereas $N*P(X'')$ is not. Hence when we cross θ at X from R_1 to R_2, either a connected component of $P(X')$ splits into two separate components, or an interior boundary curve of some connected component K

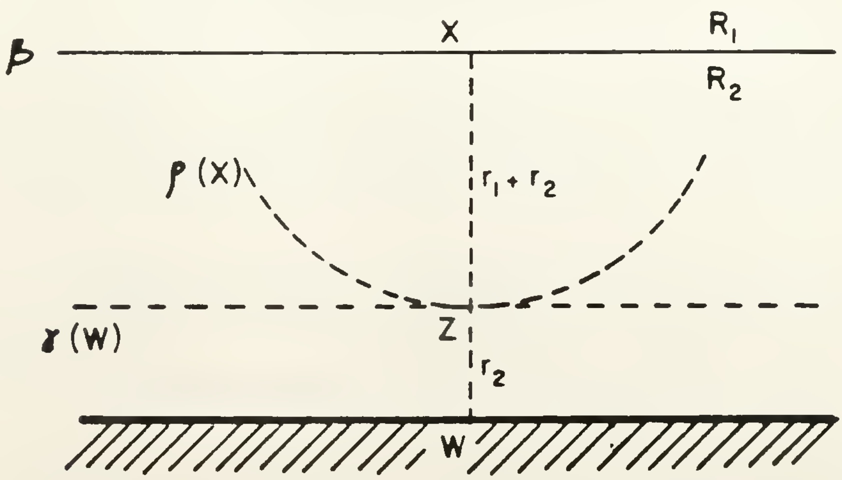

Fig 2.1. Crossing a type I critical curve.

of $P(X')$ comes to touch another boundary curve (interior or exterior) of the same K. To ease the statement of the crossing rule that applies in this case, we find it convenient to represent the node $[R,T]$ of CG corresponding to a connected component K of $P(X)$ for any $X \tau R$ as follows: T is represented as a collection of circular lists, each having the form $[t_1,\ldots,t_m]$, where each t_i is either a wall section or is B_1. Each such list describes a connected (interior or exterior) portion ζ^* of the boundary of K, specifically by listing the wall sections which contain the boundary curve segments comprising ζ^*, which we arrange in the order in which they appear as ζ^* is traversed with K to the right.

Keeping this convention in mind, let $[R_1,T_1]$ be the node designating the component K, where first we suppose that X lies on the R_1-side of θ. Then let X cross θ from the R_1-side, and suppose that the point Z of contact between two boundaries that appears during the crossing lies on the boundary curve segments CS_1 (which is a portion of $\zeta(W)$) and CS_2 (which is a portion of $\kappa(X)$). Two cases can occur:

(i) The labels of CS_1 and CS_2 may appear as two components of the same circular list in T_1, say as t_1 and t_i respectively in the list $L = [t_1,\ldots,t_m]$. It is easily checked that in this case K always splits into two subcomponents, and to represent this topological fact combinatorially we split L into two (circular) sublists $L_1 = [t_1,\ldots,t_i]$ and $L_2 = [t_i,\ldots,t_m,t_1]$. If L labeled the exterior boundary of K, then L_1 and L_2 are labelings for the exterior boundaries of the two new components. On the other hand, if L labeled an interior portion of the boundary of K, then one of the new lists, say L_1, labels the exterior boundary of a new component, whereas the exterior boundary of the second component is the same as the exterior boundary of K itself. (Note: some additional geometric analysis, whose details we leave to the reader, will be needed to assign the remaining interior boundary portions to one or another of the two new components.) Overall, we obtain two collections of circular lists T_2, T_2', defining two nodes $[R_2,T_2]$ and $[R_2,T_2']$ belonging to CG, and we connect $[R_1,T_1]$

to both these nodes. As usual, we also connect $[R_1,T]$ to $[R_2,T]$ for all other T appearing in connectivity graph nodes $[R_1,T]$.

(ii) The labels of CS_1 and CS_2 may appear as components of two different circular lists L_1 and L_2 in T_1. In this case it is easily seen that K is not split across θ, and that only its labeling changes due to the merging of two portions of its boundary into one. In this case we simply merge the two lists L_1, L_2 into one circular list L by re-linking CS_1 in L_1 to CS_2 in L_2 and vice versa. Replacing the two circular lists L_1 and L_2 in T_1 by the single list L gives us a new collection T_2 and $[R_2,T_2]$ τ CG. We then link $[R_1,T_1]$ to $[R_2,T_2]$, and also connect all other nodes $[R_1,T]$ τ CG to $[R_2,T]$ τ CG.

<u>Type II Curves:</u> These require a somewhat different treatment than curves of type I. Recall that a type II curve is a circular arc θ of radius r_1+r_2 centered at a corner point Z at which two boundary curves $\zeta(W_1)$, $\zeta(W_2)$ intersect. Here we distinguish between the two following subcases:

(a) The interior angle at Z (between $\zeta(W_1)$ and $\zeta(W_2)$) is less than 180^7 but greater than 90^0.

The type I critical curves γ_1, γ_2 that touch β partition it into three segments β_1, β_2, β_3 (see Fig. 2.2(a)). If we cross the portion β_3 of β at some point X, then the point at which $\rho(X)$ intersects $\gamma(W_2)$ approaches Z, coincides with Z on β_3, and in R_3' $\rho(X)$ intersects $\gamma(W_1)$ rather than $\gamma(W_2)$. In this case, the corresponding component of P(X) simply changes its labeling. More specifically, let $[R_3,T]$ ε CG be the node designating K. Then one of the lists in T contains three consecutive labels B_1,W_2 and W_1. By removing W_2 from this list we obtain a new T' such that $[R_3',T']$ ε CG describes K on the R_3'-side of β_3. We thus connect $[R_3,T]$ to $[R_3',T']$ and also connect every other $[R_3,T'']$ ε CG to $[R_3',T'']$ ε CG. A completely symmetric situation arises as we cross β_1, with the triple W_2,W_1,B_1 of consecutive list entries on the R_1-side of β_1 replaced by the pair W_2,B_1 on the R_1'-side of β_1.

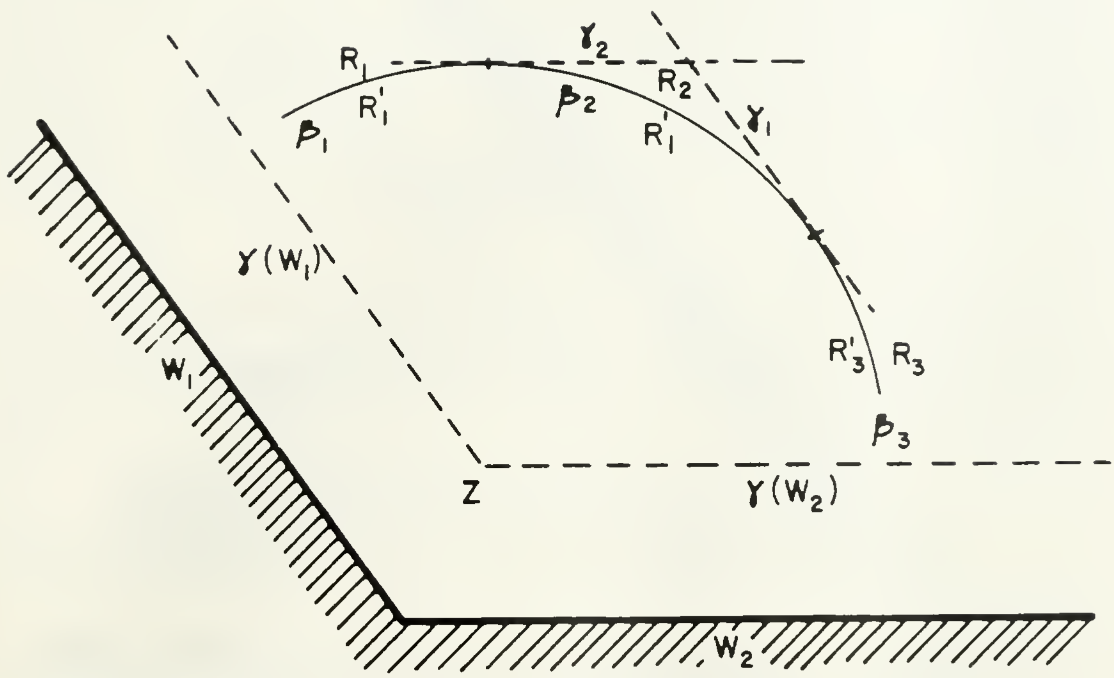

Fig. 2.2(a) Crossing a type II critical curve (obtuse case)

However, if we cross the curve portion β_2, the situation is quite different: As the center X of B_1 crosses β_2 from R_2 to R_2' a small connected component K of P(X) about Z shrinks to the single point Z, and then disappears. In this case we simply do not connect the node $[R_2,T]$ in CG to any of the nodes of R_2', but connect every other node $[R_2,T']$ to the corresponding node $[R_2',T']$.

(b) The interior angle at Z (between $\gamma(W_1)$ and $\gamma(W_2)$) is less than or equal to 90 degrees. This case is really a special case of case (a) just considered. Here only the curve section β_2 appears, and the appropriate crossing rule already seen in (a), i.e. that which applies when a small component about Z shrinks to the point Z and then disappears, applies in this case too.

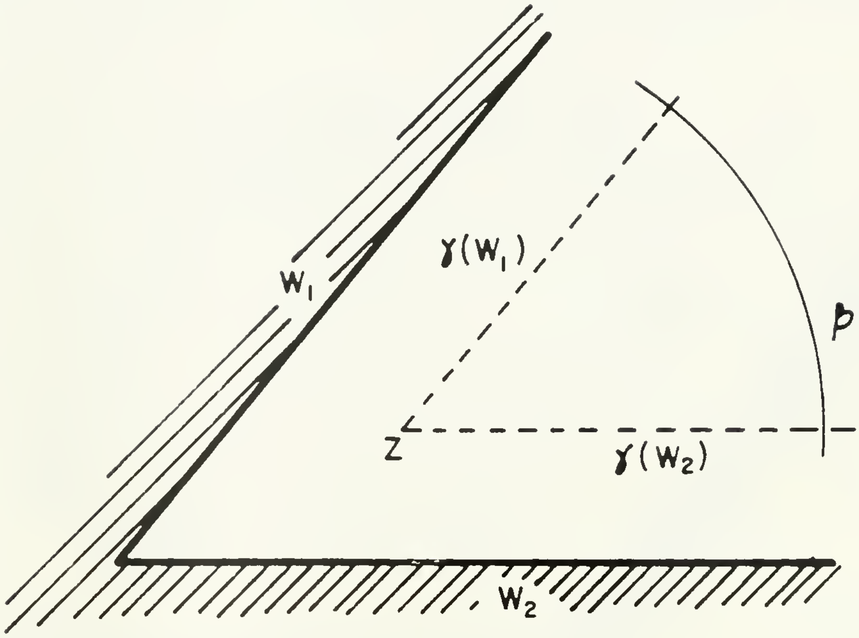

Fig. 2.2(b). Crossing a type II critical curve (acute case).

Finally we need to consider various extreme cases ignored in the preceding discussion. Since all our critical curves are straight segments and circular arcs, they can have non-isolated intersection only if they coincide. Two circles can only coincide if they have the same center and radius, so that no two critical circular arcs can coincide. Hence the only coincidence of critical curves than can arise is that of two type I straight segments overlapping each other. Plainly this can happen only if there exist two parallel walls at distance $2(r_1+2r_2)$ apart.

The crossing rules applicable in this case are easily derived by imagining such a pair of walls to be shifted by some randomly chosen infinitesimal amount. This splits the corresponding coincident critical line into two lines separated by an infinitesimal distance; and then any crossing of the original line amounts to crossing both these infinitely close split lines. An infinitely thin strip region will then appear between the two lines split in this way from a single critical line. After this splitting, all coincidences of critical curves will have been removed, and then the crossing rules stated above will apply. Then the crossing rule applying in the case of coincident critical curves derive from those of the infinitely displaced case by treating connections across the coincident curves as successive connections across the two infinitely separated critical curves introduced by this 'splitting' procedure.

Having provided all these geometrical details concerning critical curves and their associated crossing rules we now summarize the techniques described so far in this paper in a crude sketch of an algorithm which solves the motion-planning problem for two circular bodies.

Algorithm TWOCIRCLES:

The algorithm consists of a preprocessing phase followed by a path-planning phase. The preprocessing phase consists of the following steps:

(a) An input step which reads in the radii r_1, r_2 and the structure of the walls. This structure is also checked for consistency, and the region V is identified.

(b) The wall edges and convex corners are displaced by distances r_2, r_1 and r_1+2r_2 to form respectively the collections of boundary arcs of A, the boundary arcs of A_1 and the type I critical curves. All intersections of r_2-displaced walls and corners are computed, and the circular type II curve of radius r_1+r_2 about each of these points is generated.

(c) All intersection points of pairs of curves, each of which is either a boundary arc of A_1, a type I critical curve or a type II curve, are computed. For each curve β of this family we sort the points of intersection of β with other such curves by the order in which they appear along β. This will partition β into open subsegments, each of which is not intersected by any other critical curve or boundary arc.

Let Σ denote the collection of these segments. Each $\beta' \in \Sigma$ is stored as a directed curve with which we also associate two virtual noncritical regions $r(\beta')$ and $\ell(\beta')$ designating the noncritical region lying immediately to the right (resp. left) side of the directed β'. Then for each intersection point that has been computed we sort the segments of Σ that have X as an endpoint by the counterclockwise order in which they emerge from or enter into X.

(d) Next we construct all noncritical regions. This is done by tracing the boundary of each such region R so that R lies on the right side of each traced arc. To achieve this we pick some curve segment $\beta \in \Sigma$ and follow it in some direction d (which is either 'forward' if β is followed in the direction associated with it, or 'backward' otherwise). For each $\beta \in \Sigma$ and direction d we let REG(β,d) denote either $r(\beta)$ if d='forward' or $\ell(\beta)$ if d='backward'. Let X be the endpoint of β in the direction d, and let β' be the next element of Σ appearing in the circular list of curves about X. We then continue to follow β' in the direction pointing away from X. We repeat this procedure until β is encountered once more, thereby enclosing a noncritical region R which can be defined as an equivalence class containing all sides REG(β',d') for curves and directions [β',d'] encountered in the above tracing. This tracing procedure is repeated as long as there remain curves $\beta \in \Sigma$ and directions d such that β has not yet been traced in the direction d. After all noncritical regions have been constructed, the information gathered enables us to construct all pairs of noncritical curves that are adjacent to one another, together with the critical curve separating each such pair.

(e) Next we construct the connectivity graph CG. This can be done in several ways. One crude technique would be to pick an arbitrary point X in each noncritical region R, compute $\sigma(X)$ by considering the collection of all r_2-displaced walls and corners and the circle $\rho(X)$, and applying to this collection the same tracing procedure used above, thereby partitioning P(X) into its connected components and obtaining the label of each of these components. This yields us the sets $\sigma(R)$, for all noncritical regions R, from which we can generate all nodes of

CG. Also, assuming that no two critical curves coincide, we can use the general crossing rules established earlier to find all edges of CG in a straightforward way, namely: For each pair of adjacent noncritical regions R, R' we connect the nodes $[R,S]$ and $[R',S]$ of CG for each $S \in \sigma(R)*\sigma(R')$, and connect each node $[R,S]$, $S \in \sigma(R)-\sigma(R')$, to each node $[R',S']$, $S' \in \sigma(R')-\sigma(R)$. A more efficient technique is to calculate $\sigma(R)$ in the manner just outlined for only one noncritical region R, and then proceed to other noncritical regions by successive crossings of critical curve segments. Each time we cross from a region R_1 to another R_2 across a common boundary segment β, we can calculate $\sigma(R_2)$ from $\sigma(R_1)$ by applying the crossing rule associated with β. (If A_1 is not connected, this has to be repeated for each connected component of A_1.)

This step ends the preprocessing phase, which is independent of the specific positions of B_1 and B_2. The next phase makes use of the preprocessed connectivity graph to search for a path between two specified configurations of the two circles, as follows:

(f) Let $[X,Y]$, $[X',Y']$ be the initial and final configurations of the two circles. We begin by finding the noncritical regions R, R' containing the points X, X' respectively. This is an instance of the planar point-location problem for regions bounded by straight lines and circular arcs, for which efficient procedures are known (cf. [Pr]). For our purpose it even suffices to locate the regions R, R' by the following simpler technique: Pass a ray from X, find all its intersections with critical curves, and take the intersection nearest to X. This gives us a critical curve section β bounding R, and also determines the side of β on which R lies; this information defines R uniquely. R' can be found in similar fashion. Next we find the label S (resp. S') of the connected component of $P(X)$ (resp. $P(X')$) containing Y (resp. Y'). (Again, this is done by applying an appropriate point-location procedure, or even by using a 'brute force' technique similar to that noted above for locating the regions R, R'.)

(g) Next we search for a path π in CG between $[R,S]$ and $[R',S']$. If no such path exists the required continuous motion between $[X,Y]$ and $[X',Y']$ also cannot exist; the algorithm reports this and halts.

(h) If such a path π exists we use it to construct the required continuous motion between the specified configurations as in the proofs of Lemma 1.5 and Theorem 1.1. Specifically, the required motion consists of a sequence of submotions, each of which has one of the following forms:

(i) B_2 is held fixed while B_1 moves along a straight line for a short distance so that its center crosses a boundary between two noncritical regions.

(ii) B_1 is held fixed while B_2 moves to some corner of A.

(iii) B_2 is held fixed while B_1 moves so that its center moves within some noncritical region R from a point near one of its boundary arcs (or, initially, from X) to a point near another such arc (or, finally, to X′).

If carefully implemented, the complexity of the algorithm just sketched is $O(n^3)$. This follows from the fact that the total number of displaced walls and corners and of critical curves is $O(n)$. Since all of these curves are straight or circular arcs, it follows that they can intersect in at most $O(n^2)$ points; consequently there are at most $O(n^2)$ possible noncritical regions, and for each of these regions R the set $\sigma(R)$ can contain at most n labels. The size of CG is therefore $O(n^3)$, and careful implementation of the steps outlined above allows one to construct and search through this graph in total time $O(n^3)$.

An Example:

We conclude our description of the two-circle movers′ problem by an example which involves two circles moving through the inside V of a regular pentagon. The sizes of the circles are chosen so that when B_1 is placed nearly touching the midpoint of one edge AB of the pentagon, B_2 can barely fit in the space between B_1 and the two edges CD and DE of the pentagon (see Fig. 2.3(a)). The instance of the problem that

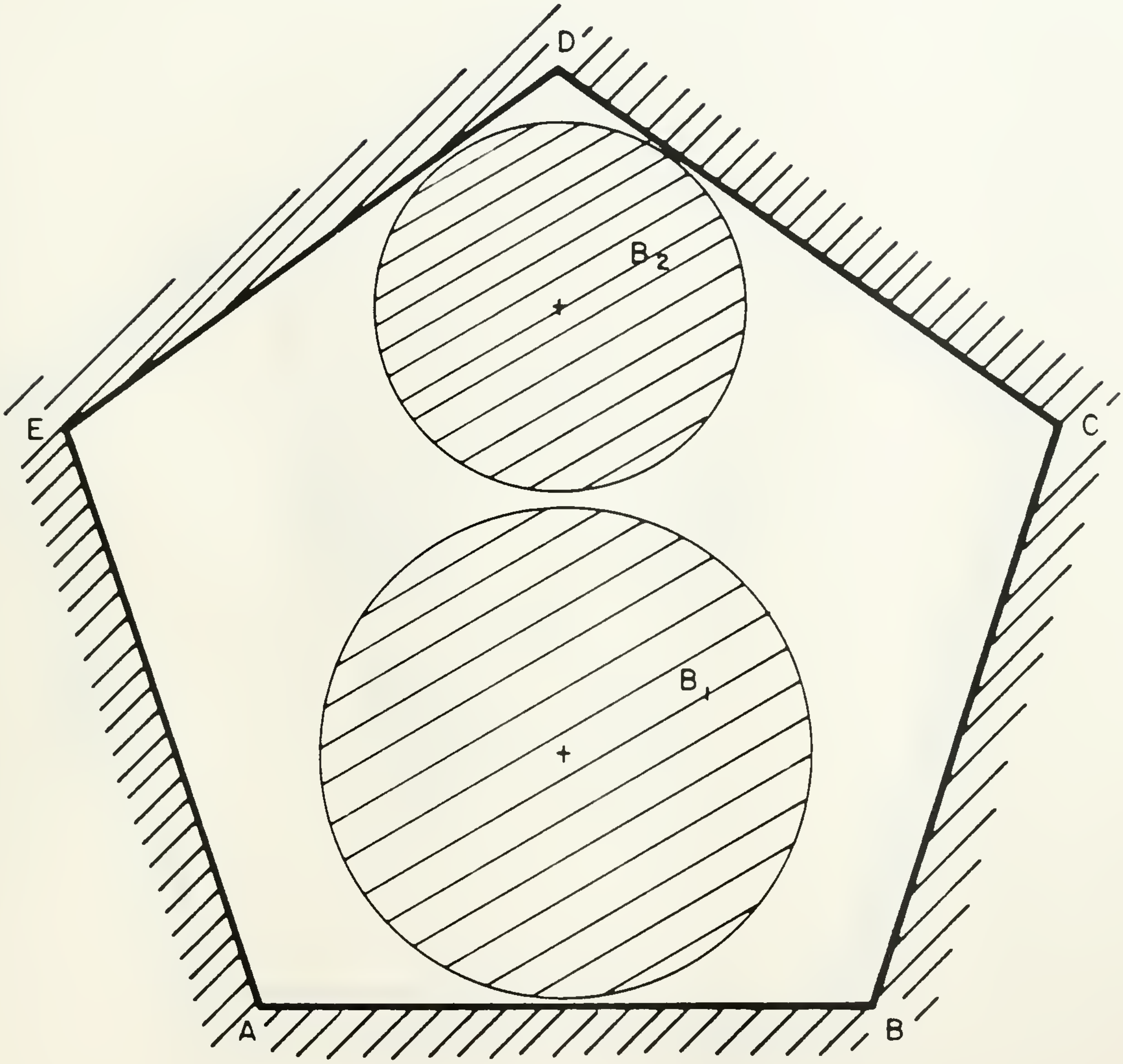

Fig. 2.3(a). An Example of the 2-circle Movers' Problem: Initial Positions.

we wish to solve is to move the circles from the positions shown in Fig. 2.3(a) to those shown in Fig. 2.3(b).

As is demonstrated by our algorithm, this task can be accomplished by the following sequence of motions: (i) Move B_1 from its initial position parallel to AB until it almost touches both AB and BC; (ii) Move B_2 parallel to DE until it almost touches both DE and AE; (iii) Move B_1 parallel to BC until it almost touches both BC and CD; (iv) Move B_2 parallel to AE until it almost touches both AE and AB; (v) Move

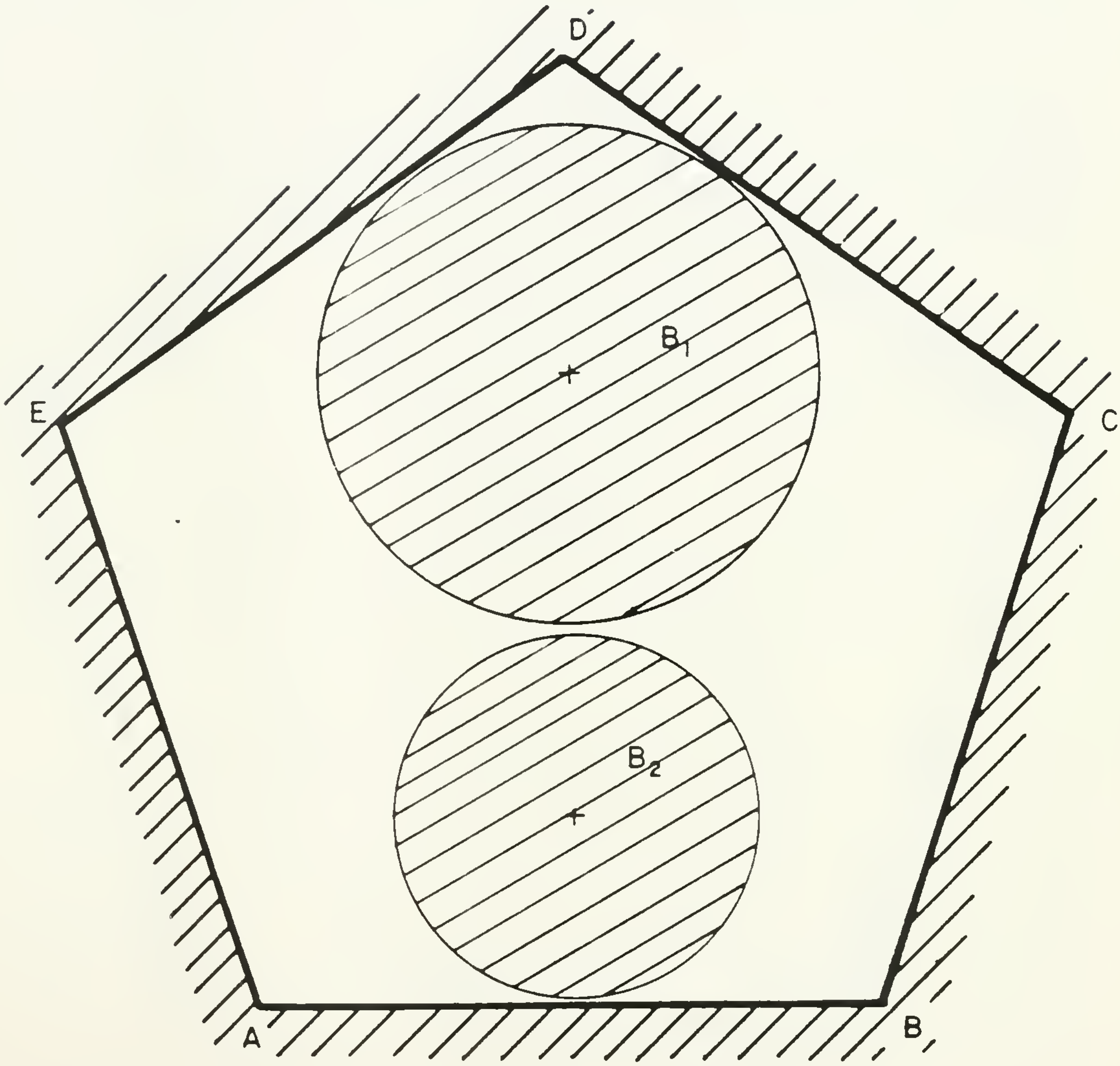

Fig. 2.3(b). Final Positions of the Circles in the Example.

B_1 parallel to CD until it reaches its target position; (vi) Move B_2 parallel to AB until it reaches its target position. Furthermore, again as shown by our algorithm, there is no shorter sequence of motions of B_1 and B_2 which accomplishes the required motion.

Remark: This example can be easily generalized to yield examples in which two circles move inside a regular (2k+1)-gon. To rotate them around the polygon (never moving two circles simultaneously) will require a sequence of O(k) alternating motions of each of the circles.

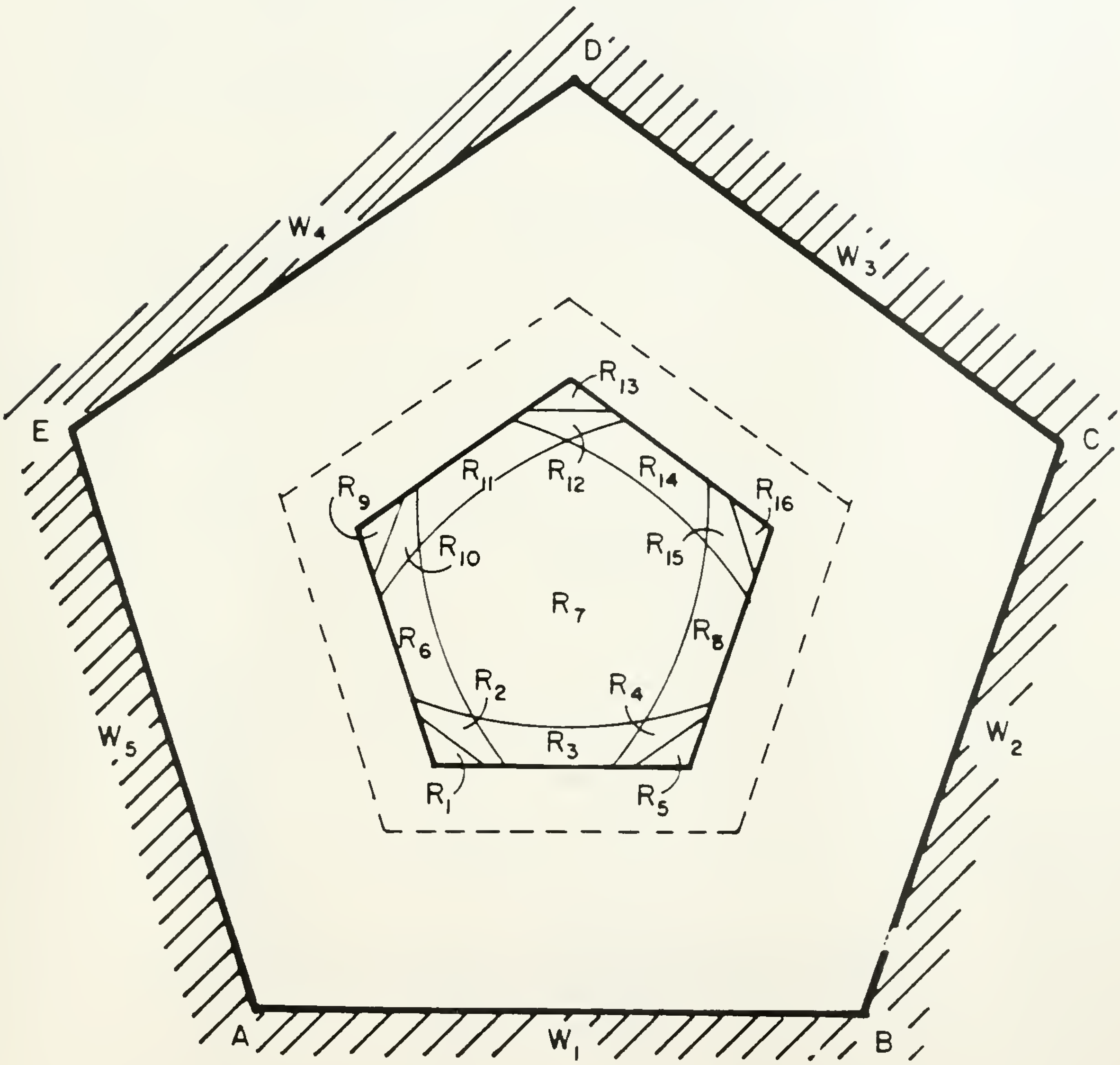

Fig. 2.4. The critical curves and noncritical regions of our example.

To apply our technique to this instance of the two-circle problem, we first displace each side of the pentagon into V by the amount r_2, thereby obtaining the boundary arcs of the region A of B_2-admissible positions, shown as dashed lines in Fig. 2.4. Let us denote the r_2-displacement of the side W_j of the pentagon by γ_j, j=1,..,5. Next we displace the sides of our pentagon into V by the amount r_1, to obtain the boundary of the region A_1 of B_1-admissible positions. Within this region we draw the type I critical curves, i.e. the sides of the pentagon displaced by the amount r_1+2r_2 into V, and the type II critical curves, i.e. the circular arcs of radius r_1+r_2 about the corners of A. These critical curves partition A_1 into 16 noncritical regions. The table below lists the characteristics of each of these noncritical regions.

noncritical region	characteristic	corresponding node(s) in connectivity graph
R_1	$\{[\rho\gamma_4\gamma_3\gamma_2]\}$	n_1
R_2	$\{[\rho\gamma_4\gamma_3], [\rho\gamma_3\gamma_2]\}$	n_2^a, n_2^b
R_3	$\{[\rho\gamma_4\gamma_3]\}$	n_3
R_4	$\{[\rho\gamma_5\gamma_4], [\rho\gamma_4\gamma_3]\}$	n_4^a, n_4^b
R_5	$\{[\rho\gamma_5\gamma_4\gamma_3]\}$	n_5
R_6	$\{[\rho\gamma_3\gamma_2]\}$	n_6
R_7	$\{\}$	
R_8	$\{[\rho\gamma_5\gamma_4]\}$	n_8
R_9	$\{[\rho\gamma_3\gamma_2\gamma_1]\}$	n_9
R_{10}	$\{[\rho\gamma_3\gamma_2], [\rho\gamma_2\gamma_1]\}$	n_{10}^a, n_{10}^b
R_{11}	$\{[\rho\gamma_2\gamma_1]\}$	n_{11}
R_{12}	$\{[\rho\gamma_2\gamma_1], [\rho\gamma_1\gamma_5]\}$	n_{12}^a, n_{12}^b
R_{13}	$\{[\rho\gamma_2\gamma_1\gamma_5]\}$	n_{13}
R_{14}	$\{[\rho\gamma_1\gamma_5]\}$	n_{14}
R_{15}	$\{[\rho\gamma_1\gamma_5], [\rho\gamma_5\gamma_4]\}$	n_{15}^a, n_{15}^b
R_{16}	$\{[\rho\gamma_1\gamma_5\gamma_4]\}$	n_{16}

Note that the center region R_7 consists of positions of B_1 for which no free position of B_2 exists. The crossing rules in this example are derived using the principles outlined above, and are depicted in the connectivity graph CG shown in Fig. 2.5. It is also easy to check that the initial (resp. final) configuration of the two circles (shown in Fig. 2.3) belongs to the cell of FP represented by the node n_3 (resp. n_{13}) of CG. Since CG is connected, there exists a continuous motion of the two circles between the initial and the final configurations. In fact, CG contains just two paths between n_3 and n_{13}, corresponding to motions of the circles in which the line

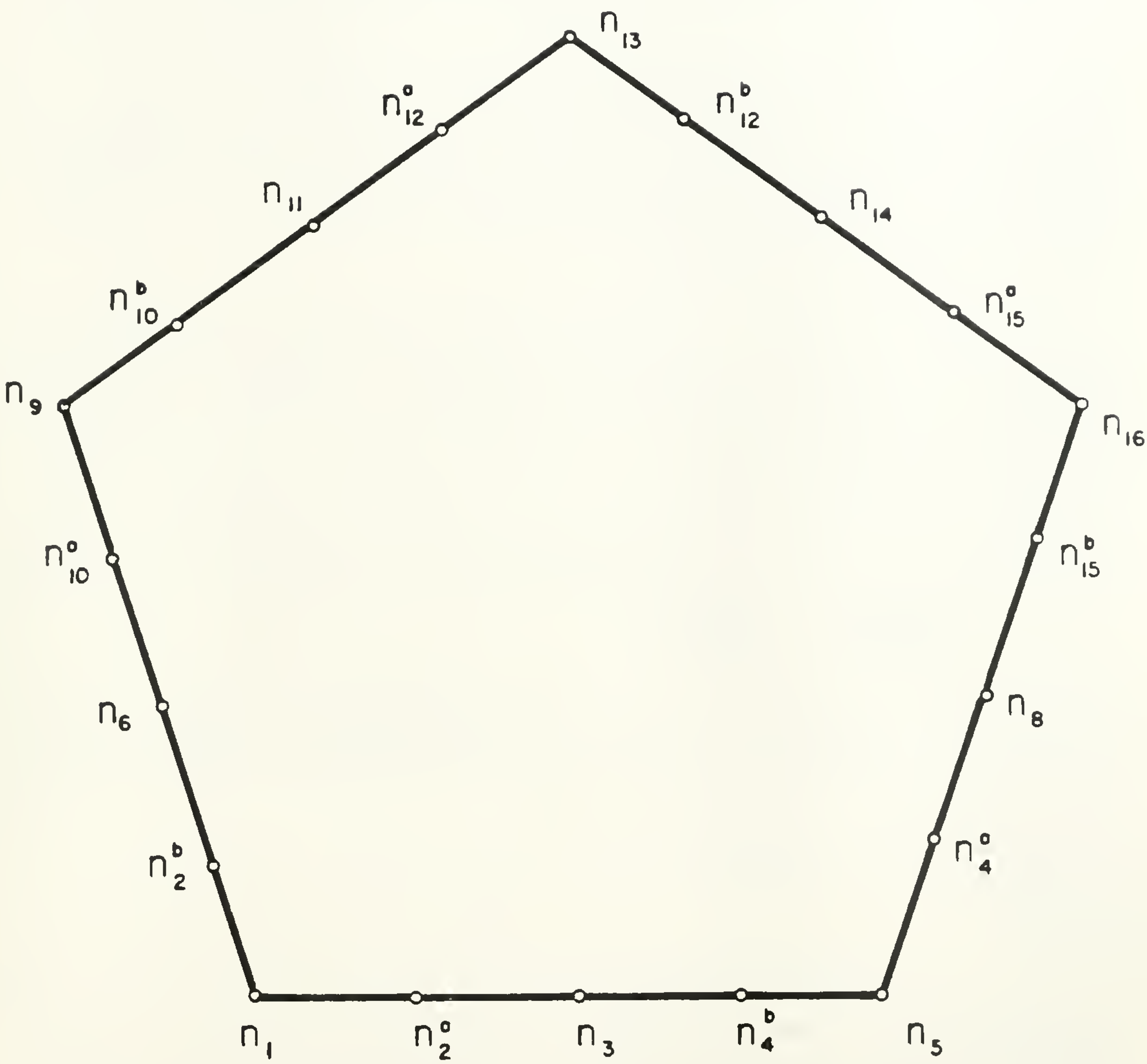

Fig. 2.5. The Connectivity Graph of our Example.

connecting their centers rotates in a clockwise or counterclockwise direction. The actual motion of the two circles can be reconstructed from these paths in the manner described in Theorem 1.1. One of these motions is easily seen to coincide with the motion described above, and the other is simply a 'mirror image' of this motion.

3. Coordinating the motion of 3 independent circular bodies

Having now treated the case of 2 independent circles, we go on to study motion-planning algorithms for 3 independent circles. The approach adopted will illustrate a more general recursive technique for successive elimination of degrees of freedom, which can be applied to other motion planning problems.

Denote the three circular bodies with whose motion we are concerned as B_1, B_2, B_3, and suppose that their respective radii are $r_1 \geq r_2 \geq r_3$. These bodies are constrained to move in a 2-dimensional region V bounded by polygonal walls and obstacles, as described in the preceding sections, and must avoid collision with the walls and with each other. For i=1,2,3, let A_1 denote the open subset of V consisting of all admissible positions for the center C_i of B_i (that is, all positions X whose distance from any of the walls is greater than r_i). We will refer to positions in A_i as B_i-admissible positions. We make simplifying assumptions concerning V that are similar to those made for the two-circle case, namely we assume that the boundary of each set A_i (which consists of walls and corners displaced into V by distance r_i) consists of a collection of disjoint simple closed curves.

A triple $[X_1,X_2,X_3]$ of points $X_j \in A_j$, j=1,..,3, is called a free configuration if when each circle B_i is placed with its center C_i at X_i, no circle touches any wall or any other circle. Similarly, a semi-free configuration $[X_1,X_2,X_3]$ is a configuration at which zero or more contacts (but no penetrations) between circles and walls or between two circles occur. As before, FP denote the space of all free configurations and SFP the space of all semi-free configurations. Plainly FP is an open 6-dimensional manifold, whereas SFP is a 6-dimensional manifold with boundary. As usual, our aim is to partition FP into its connected components.

As in the two-circle case, we attack this problem by projecting FP into a subspace of fewer dimensions. Specifically, let $P(X_1)$ denote the 4-dimensional subspace consisting of all free positions of the centers of B_2 and B_3, when B_1 is fixed with its center at X_1. Then each such projected set $P(X_1)$ needs to be partitioned into its connected components, and the dependence of these components on X_1 must be studied. We can use the preceding results concerning the coordinated motion of two circles to obtain the desired partitioning of $P(X_1)$. For this, note that each position of B_1 with its center C_1 at X_1 can be viewed as defining an additional barrier for the collision-free motion of B_2 and B_3. Although this new barrier is not polygonal, our analysis of the motion of two circles applies even when some of the barriers are displaced polygonal, rather than simple polygonal, curves. Since B_1 can be viewed as the single point X_1 displaced by distance r_1, this remark applies to the case at hand. Hence, for each fixed value of X_1, decomposition of $P(X_1)$ into connected components can proceed as follows.

(i) Displace all walls (including B_1) by distance r_3. The subset of V lying outside these displaced walls is the set $A_3'(X_1)$ consisting of all B_3-admissible positions in the presence of B_1 with its center at X_1.

(ii) Displace all walls (including B_1) by distance r_2, to obtain a set $A_2'(X_1)$ having analogous meaning.

(iii) Draw all B_2-critical curves. These are either walls (or the circular periphery of B_1) displaced by distance r_2+2r_3, or are corners of $A_3'(X_1)$ displaced by distance r_2+r_3. These curves partition $A_2'(X_1)$ into connected open <u>noncritical regions</u>. Suppose that B_2 is constrained to move with its center C_2 remaining inside such a region R. Then the space $Q(X_1,X_2)$ of all free positions in $A_3'(X)$ of the center of B_3 decomposes into connected components K whose labelings $\lambda(K)$ (as defined in Definition 1.2) remain invariant throughout R. Furthermore, the component of $Q(X_1,X_2)$ having a given labeling (which, by Lemma 1.3, characterizes this component uniquely) varies continuously with $X_2 \in R$.

(iv) Next, we construct the connectivity graph $CG(X_1)$: Its nodes are of the form $[R,\lambda(K)]$, where R is a B_2-noncritical region and where $\lambda(K)$ is a labeling of some connected component of $Q(X_1,X_2)$ for any (hence every) $X_2 \in R$. An edge connects $[R,\Lambda]$ to $[R',\Lambda']$ iff R and R′ are adjacent regions having a portion β' of some B_2-critical curve as part of their common boundary and if X_2 can cross β' from R to R′ in a way allowing a B_3-position X_3 to move continuously along with X_2 from a free position in some component K′ of $Q(X_1,X_2)$ for which $\lambda(K) = \Lambda$, $X_2 \in R$, to a free position in some component K′ of $Q(X_1,X_2)$ with $X_2 \in R'$, $\lambda(K') = \Lambda'$. As shown earlier, with each critical curve segment β' of this sort there is associated a fixed crossing rule which is independent of the particular point on β' at which B_2 crosses from R to R′.

It follows from the preceding results that the number of connected components of the open manifold $P(X_1)$ is the same as the number of connected components of the graph $CG(X_1)$. Moreover, each component C of $CG(X_1)$ defines the following connected component $\mu(X_1,C)$ of $P(X_1)$:

$$\mu(X_1,C) = \{[X_2,X_3] : [R,\Lambda] \in C,\ X_2 \in R,\ X_3 \in \psi_{X_1}(X_2,\Lambda)\}^- * P(X_1)$$

where $\psi_{X_1}(X_2,\Lambda)$ denotes the connected component of $Q(X_1,X_2)$ whose label is T. That is, the connected components of the finite graph $CG(X_1)$ can serve as discrete labels for the connected components of $P(X_1)$.

Of course, the noncritical regions R appearing in the preceding discussion depend on the position X_1 of the center of B_1, and hence they cannot be used directly to achieve a discrete labeling of the components of $P(X_1)$. However, since (by arguments analogous to Lemma 1.3) each such R is labeled uniquely by the circular sequence of displaced wall and critical curve sections constituting its boundary, and since (as will be shown below) only finitely many such sequences are possible, the regions R can themselves be given discrete labels. In what follows, we will label each B_2-noncritical region in this manner. Accordingly, given a position X_1 for the center of B_1, we can let $\tau(X_1,L)$ denote the B_2-noncritical region R labeled by L in the above sense.

The next step is to study the way in which $P(X_1)$ and $CG(X_1)$ depend on X_1. Adapting the strategy used in what has gone before, we proceed to define a collection of <u>B_1-critical curves</u> which collectively constitute the locus of all points X_1 such that if B_1 is placed with its center at X_1, then some discontinuity in the structure of $CG(X_1)$ can occur even if B_1 is moved only slightly. Such a discontinuity can only result if one of the following combinatorial events occurs.

(i) The collection of labeled B_2-noncritical regions changes; that is, either one B_2-noncritical region splits into several subregions (or vice versa); or one such region shrinks to a point and then disappears, or the labeling of some non-critical region R changes. The latter situation can arise either when a boundary edge of R splits into subsegments, or when such an edge shrinks to a point and then disappears.

(ii) The set of labels belonging to the collection of connected components of $Q(X_1,X_2)$ associated with each of the points X_2 belonging to some noncritical region R changes; that is, either one or more of the components of this set splits into subcomponents (or vice versa), or one component shrinks to a point and then disappears, or the labeling of one such component changes, again either because a boundary edge of $Q(X_1,X_2)$ splits into subedges, or because an edge of $Q(X_1,X_2)$ shrinks to a point and then disappears.

(iii) The structure of the B_2-noncritical regions and of the B_3-connected components associated with them remains unchanged, but the graph $CG(X_1)$ changes due to the appearance or disappearance of one or more edges in it.

If none of the combinatorial changes (i), (ii), (iii) listed above occur, then each B_2-noncritical region having a given labeling varies continuously with X_1, and for X_2 moving continuously within one of those noncritical regions, the connected component of $Q(X_1,X_2)$ having a given labeling varies continuously with X_1 and X_2. Indeed, suppose that the first assertion is false; then one could obtain a sequence X_{1n}

$\to X_1$, where X_1 is a point at which none of the above combinatorial changes occurs, such that for some labeling L, the regions $\tau(X_{1n},L)$ converge to some set R^* which is different from $\tau(X_1,L)$. However, the boundary of $\tau(X_{1n},L)$ converges to a closed curve which is a concatenation of B_2-critical curves and bounding displaced walls and which encloses a region of B_2-noncritical points, whose label must be L. Thus both R^* and $\tau(X_1,L)$ are labeled L, which contradicts the fact that a given label attaches to just one noncritical region. The second assertion follows by similar arguments.

Note that a discontinuous event of one of these three types can only occur in consequence of the motion of some geometric element which appears in the analysis of the decomposition of $P(X_1)$ into its components and which moves with X_1. It is easy to enumerate all such elements, which are (a) the circle $\rho_{12}(X_1)$ of radius r_1+r_2 and the circle $\rho_{13}(X_1)$ of radius r_1+r_3 about X_1. These act as moving 'displaced walls', which limit the motion of the centers of B_2 and B_3 respectively. (b) the circle $\rho_{123}(X_1)$ of radius $r_1+r_2+2r_3$ about X_1. This appears as a moving type I critical curve (for the center of B_2) in the analysis of $P(X_1)$ into its components. It is the 'wall' B_1 displaced by the amount r_2+2r_3. (c) all loci of points p obtained by taking the intersection of the circle of radius r_1+r_3 about X_1 with a fixed wall or corner displaced by the amount r_3, and then by displacing such an intersection point by the amount r_2+r_3 in any direction. These are the moving type II critical curves (for the center of B_2) generated by the intersection of the displaced wall B_1 with a displaced fixed wall (cf. Fig. 3.1 for a display of all these moving curves).

We call all of these moving curves <u>curves</u> (either boundaries or critical curves) <u>induced by</u> B_1.

Suppose that the center of B_1 moves slightly in a neighborhood N of some point X_1. Then the curves induced by B_1 and listed above move with it, and the points of their intersection with the other fixed displaced walls and critical curves change. As long as these changes are slight and quantitative, the combinatorial structure of the

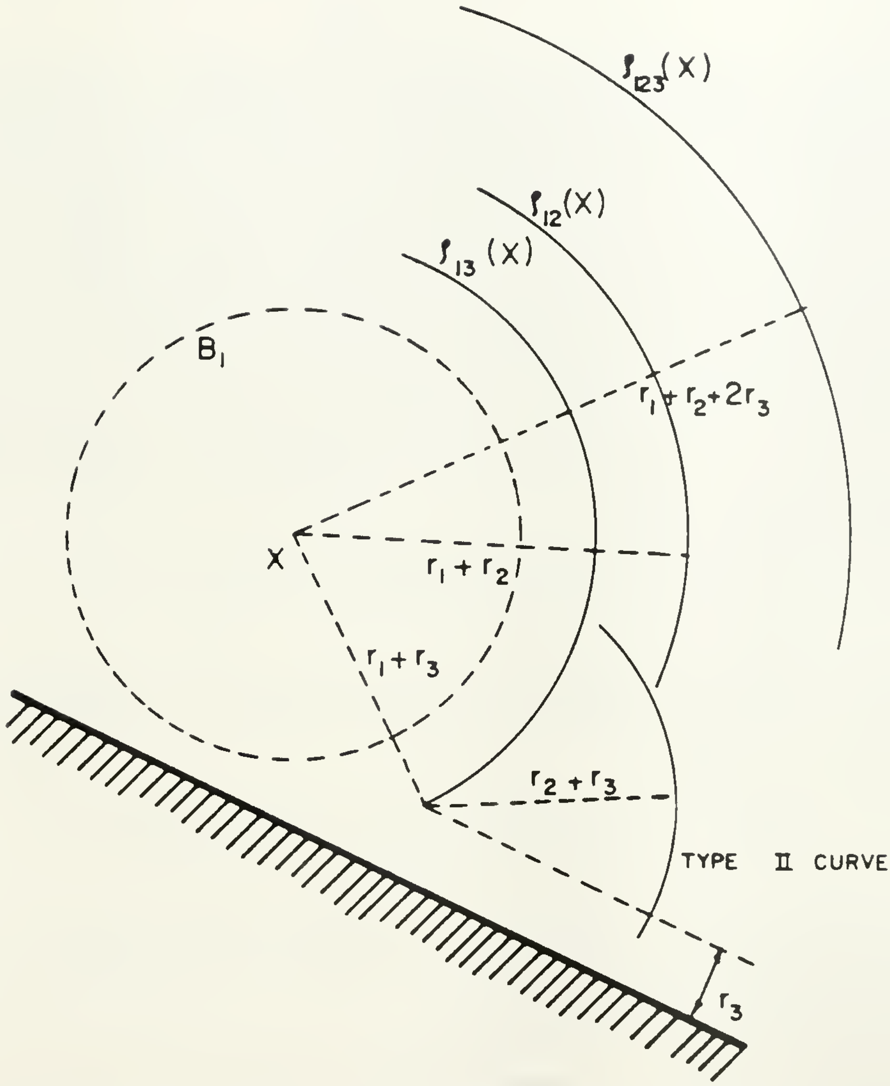

Fig. 3.1. Critical Curves and Displaced Walls Induced by B_1.

components of $P(X_1)$ will remain constant in N, and consequently X_1 will be a noncritical point. Thus, for X_1 to be critical the pattern of such intersections has to change qualitatively, and plainly this can happen only when either one of elements moving with X_1 becomes tangent to a displaced wall or critical curve, or when three such elements, at least one of which moves with X_1, meet at a point.

To assess the implications of this remark, we will begin by considering cases in which $\rho_{13}(X_1)$ becomes tangent to some displaced wall or critical curve, or passes through the intersection of two other displaced walls or critical curves. Since $\rho_{13}(X_1)$ does not count as a B_2-critical curve (see the definition of these curves for the two-circle problem, immediately preceding Lemma 1.5), this can never correspond to a discontinuity of the first type (i) listed above. Moreover, the only tangencies or triple intersections involving $\rho_{13}(X_1)$ that can cause a discontinuity of type (ii) are those in which $\rho_{13}(X_1)$ is tangent to a curve or passes through a corner which can form part of the boundary of a component of $Q(X_1,X_2)$, uniformly for all X_2 in some small open set, which is to say, tangent to an r_3-displaced wall or corner, or passes through the intersection of two r_3-displaced walls or corners. The locus of points X_1 at which this happens consists of the following curves:

(1) walls and corners of V displaced by r_1+2r_3;

(2) circles of radius r_1+r_3 about intersection points of r_3-displaced walls or corners.

These are our first two types of critical curves.

Next we consider extreme configurations of type (iii); but we will show that they do not exist. For this, suppose that X_1 is a point at which no critical configuration of type (i) or (ii) occurs. Then $\rho_{13}(X_1)$ does not pass through the intersection E_0 of any two r_3-displaced walls or corners (if it did, some $Q(X_1',X_2)$ would change at $X_1' = X_1$, for every X_2 not within distance r_2+r_3 of E_0), and is not tangent to any r_3-displaced wall or corner (for similar reasons). Hence there exists a neighborhood U of X_1 such that for $X_1' \in U$, $\rho_{13}(X_1')$ is not tangent to any r_3-displaced wall or corner and does not pass through the intersection of any two such walls. We can also take U small enough so that no type (i) or (ii) critical configuration occurs in U. Suppose that an edge connecting some cell $[L^a,T^a]$ to $[L^b,T^b]$ in $CG(X_1')$ exists for some $X_1' \in U$ but disappears as we pass through X_1. By the Corollary to Lemma 1.9, there exists a corner $E(X_1')$ at which two r_3-displaced walls (one of which may be the circle $\rho_{13}(X_1')$ meet and bound an angle α such that for $X_2 \in L^a$ (resp. $X_2 \in L^b$), the points of α lying near enough to its apex belong to a region whose exterior boundary has the label T^a (resp. T^b). Let X_1'' move continuously from X_1' to X_1 along a curve in U. Then the B_2-noncritical regions labeled L^a, L^b, and their bounding curves, will vary continuously, neither dividing into separate subparts, nor shrinking to points. Take points X_2^a, X_2^b lying in L^a, L^b respectively and varying continuously as those regions and their boundaries vary with X_1''. Then there is a uniquely defined continuously varying point of intersection $E(X_1'')$ of the two r_3-displaced walls which initially intersect at $E(X_1')$ (one of these may be $\rho_{13}(X_1'')$). None of the continuously varying circles $\rho_{23}(X_2^a)$ or $\rho_{23}(X_2^b)$ pass through this point, since if either did X_2^a or X_2^b would by definition lie on a type II B_2-critical curve, rather than in the noncritical region L^a or L^b. Moreover, since no type (ii) critical configuration occurs in U, the boundaries of the components of $Q(X_1'',X_2^a)$ and $Q(X_1'',X_2^b)$ retain fixed labelings as X_1'', X_2^a, X_2^b vary continuously, and the boundary curves of these components vary continuously. It follows that the angle $\alpha = \alpha(X_1'')$ formed by the two r_3-displaced walls intersecting at the continuously varying point $E(X_1'')$ also varies continuously, and that all points in this angle and sufficiently near its apex remain in a component of $Q(X_1'',X_2^a)$ (resp. of $Q(X_1'',X_2^b)$) whose boundary labeling remains fixed as X_1'' and X_2^a (resp. X_2^b) vary continuously. This implies that for $X_1''=X_1$ the points in the angle $\alpha(X_1)$ sufficiently near its apex belong to a component of $Q(X_1,X_2^a)$ (resp. $Q(X_1,X_2^b)$) with labeling T^a (resp. T^b). Used in its converse direction, the corollary to Lemma 1.9 now shows that an edge of $CG(X_1)$ does connect $[L^a,T^a]$ to $[L^b,T^b]$, contrary to assumption. This proves that configurations of type (iii) are impossible at points X_1 for which configurations of type (i) and (ii) do not occur.

Finally, we consider the more complex case of configurations of type (i). Here we need to consider all possible tangencies and triple intersections involving B_1-induced curves which can influence the structure or the labeling of B_2-noncritical regions. It is helpful to list all such interactions in a systematic table first, and then to give a more detailed description of the B_1-critical curves corresponding to each table entry. (Note however that some of the interactions appearing in the table shown can never actually arise because the geometric constraints that they impose are self-contradictory; these cases will be disposed below).

Table 3.1 below is organized as follows: Each row has at most five entries: a serial number for convenient reference, two or three entries designating the nature of the curves involved in the critical configuration (two entries designate a tangency, whereas three entries designate a triple intersection), and a number referencing a paragraph in the detailed list of critical curves following this table, in which every type of curve is discussed. Each curve involved in a tangency or intersection is represented in the table by a mnemonic symbol which can be either 'bd', designating a boundary, i.e. a displaced wall limiting the motion of B_2, or 'I', designating a type I critical curve for B_2, or 'II', designating a type II critical curve for B_2. This mnemonic symbol always appears either by itself, designating a curve which does not depend on B_1, or is followed by '(B1)', designating a B_1-induced curve.

Serial No.	1st Curve	2nd Curve	3rd Curve	Critical Curve No.
1	bd (B1)	bd		(3)
2	I (B1)	bd		(4)
3	II (B1)	bd		(20)
4	bd (B1)	I		(4)
5	I (B1)	I		(5)
6	II (B1)	I		(21)
7	bd (B1)	II		(6)
8	I (B1)	II		(7)
9	II (B1)	II		(22)
10	bd (B1)	II (B1)		-
11	I (B1)	II (B1)		-
12	II (B1)	II (B1)		-
13	bd (B1)	bd	bd	(8)
14	bd (B1)	bd	I	(10)
15	bd (B1)	bd	II	(12)
16	bd (B1)	I	I	(14)
17	bd (B1)	I	II	(16)
18	bd (B1)	II	II	(18)
19	I (B1)	bd	bd	(9)
20	I (B1)	bd	I	(11)
21	I (B1)	bd	II	(13)
22	I (B1)	I	I	(15)
23	I (B1)	I	II	(17)
24	I (B1)	II	II	(19)
25	II (B1)	bd	bd	(23)
26	II (B1)	bd	I	(24)
27	II (B1)	bd	II	(26)
28	II (B1)	I	I	(25)
29	II (B1)	I	II	(27)
30	II (B1)	II	II	(28)
31	bd (B1)	II (B1)	bd	(29)
32	bd (B1)	II (B1)	I	(30)
33	bd (B1)	II (B1)	II	(35)
34	I (B1)	II (B1)	bd	(31)
35	I (B1)	II (B1)	I	(32)
36	I (B1)	II (B1)	II	(36)
37	II (B1)	II (B1)	bd	(33)
38	II (B1)	II (B1)	I	(34)
39	II (B1)	II (B1)	II	(37)
40	bd (B1)	II (B1)	II (B1)	(38)
41	I (B1)	II (B1)	II (B1)	-
42	II (B1)	II (B1)	II (B1)	-

Table 3.1. Critical Interactions of Curves.

Concerning this table, note the following:

(i) Since curves of type bd(B1) and type I(B1) always remain at the same distance from each other, no interactions between these curves are possible.

(ii) Case (10) is impossible, since it would require a circle of radius r_2+r_3 about a point on a circle of radius r_1+r_3 about X_1 to be tangent to a circle of radius r_1+r_2 about X_1.

(iii) Case (11) describes an interaction which always takes place, namely that in which a circle of radius r_2+r_3 about a point on a circle of radius r_1+r_3 about X_1 is tangent to a circle of radius $r_1+r_2+2r_3$. Hence this condition does not generate any B_1-critical curve.

(iv) Case (12) is impossible because two B_1-induced type II B_2-critical curves cannot be tangent to each other at a free or semi-free position, because these curves are circular arcs of the same radius whose centers lie on the circle $\mu_{13}(X_1)$, so that they can be tangent to each other only at a point interior to that circle, which is not a free position for either center of B_2 or B_3.

(v) Case (41) is impossible since it would require some point to be at distance r_2+r_3 from two distinct points on the circle of radius r_1+r_3 about X_1, and also to be at distance $r_1+r_2+2r_3$ from X_1, which is plainly impossible.

(vi) Case (42) is impossible since it would require some point to be at the same distance r_2+r_3 from three distinct points on the circle of radius r_1+r_3 about X_1, contradicting the fact that two circles can intersect in at most two points.

Now that we have listed all possible critical interactions, we go on to describe them in more detail. For this, note that $\mu_{12}(X_1)$ and $\mu_{123}(X_1)$ are respectively the only displaced wall and the only type I critical curve generated by B_1 and affecting the motion of B_2. A first group of B_1-critical curves at which extreme configurations of type (i) arise is obtained by considering situations in which either $\rho_{12}(X_1)$ or $\rho_{123}(X_1)$ is tangent to another B_2-boundary or critical curve, or when $\rho_{12}(X_1)$ or $\rho_{123}(X_1)$ passes through an intersection of two B_2-boundary or critical curves. The resulting B_1-critical curves are:

(3) (resp. (4)) $\rho_{12}(X_1)$ (resp. $\rho_{123}(X_1)$) is tangent to a B_2-boundary curve: The X_1-loci at which this happens are the walls and corners of V displaced by r_1+2r_2. (resp. $r_1+2r_2+2r_3$).

(5) $\rho_{123}(X_1)$ is tangent to a type I B_2-critical curve: The X_1-loci at which this happens are the walls and corners of V displaced by $r_1+2r_2+4r_3$.

(6) (resp. (7)) $\rho_{12}(X_1)$ (resp. $\rho_{123}(X_1)$) is tangent to a type II B_2-critical curve: The X_1-loci at which this happens are the circles at radius $r_1+2r_2+r_3$ (resp. $r_1+2r_2+3r_3$) about intersection points of r_3-displaced walls and corners.

(8) (resp. (9)) $\rho_{12}(X_1)$ (resp. $\rho_{123}(X_1)$) passes thru an intersection point of two B_2-boundary curves: The X_1-loci at which this happens are the circles at radius r_1+r_2 (resp. $r_1+r_2+2r_3$) about the intersections of two r_2-displaced walls or corners.

(10) (resp. (11)) $\rho_{12}(X_1)$ (resp. $\rho_{123}(X_1)$) passes thru an intersection point of a B_2-boundary curve with a type I B_2-critical curve: The X_1-loci at which this happens are the circles at radius r_1+r_2 (resp. $r_1+r_2+2r_3$) about the intersections of an r_2-displaced wall or corner with an (r_2+2r_3)-displaced wall or corner.

(12) (resp. (13)) $\rho_{12}(X_1)$ (resp. $\rho_{123}(X_1)$) passes thru an intersection point of a B_2-boundary curve with a type II B_2-critical curve: The X_1-loci at which this happens are the circles at radius r_1+r_2 (resp. $r_1+r_2+2r_3$) about the intersections of an r_2-displaced wall or corner W with the circle at radius r_2+r_3 about an intersection of two r_3-displaced walls and corners W′,W" (cf. Fig. 3.2).

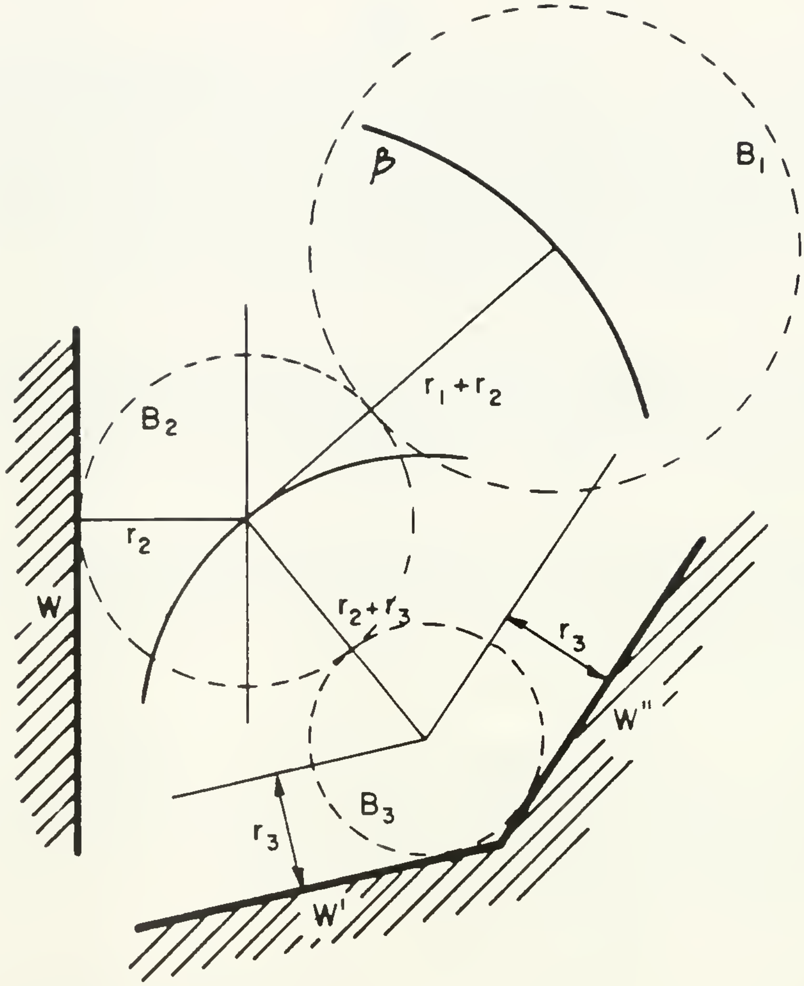

Fig. 3.2 A Type (12) Critical Curve β, and Corresponding Critical Positions of the Three Circles.

(14) (resp. (15)) $\rho_{12}(X_1)$ (resp. $\rho_{123}(X_1)$) passes thru an intersection point of two type I B_2-critical curves: The X_1-loci at which this happens are the circles at radius r_1+r_2 (resp. $r_1+r_2+2r_3$) about the intersections of two (r_2+2r_3)-displaced walls or corners.

(16) (resp. (17)) $\rho_{12}(X_1)$ (resp. $\rho_{123}(X_1)$) passes thru an intersection point of a type I B_2-critical curve and a type II B_2-critical curve: The X_1-loci at which this happens are the circles at radius r_1+r_2 (resp. $r_1+r_2+2r_3$) about the intersections of a (r_2+2r_3)-displaced wall or corner with the circle at radius (r_2+r_3)

about an intersection of two r_3-displaced walls or corners. (These cases are very similar to cases (12),(13) above.)

(18) (resp. (19)) $\rho_{12}(X_1)$ (resp. $\rho_{123}(X_1)$) passes thru an intersection point of two type II B_2-critical curves: The X_1-loci at which this happens are the circles at radius r_1+r_2 (resp. $r_1+r_2+2r_3$) about the intersection points of two circles at radius r_2+r_3, each about an intersection of two r_3-displaced walls or corners (cf. Fig. 3.3).

Note that all these curves are either displaced walls, displaced corners, or circles about one or another center.

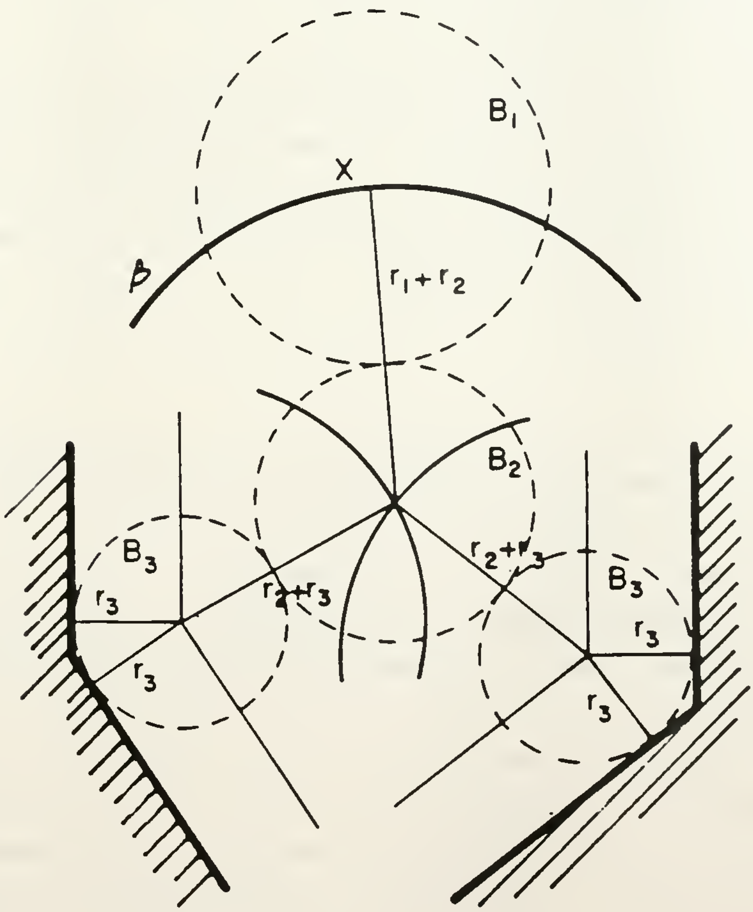

Fig. 3.3 A Type (18) Critical Curve β and Corresponding Critical Positions of the Three Circles.

So far we have considered extreme configurations of type (i) arising from some critical contact of either a B_2-boundary or a B_2-type I critical curve induced by B_1 with other B_2-boundary or critical curves. A second group of B_1-critical curves arise from critical intersections of a B_2-type II critical curve induced by B_1 with another B_2-boundary or critical curve. The possible cases of this kind are as follows:

(20) (resp. (21)) A B_1-induced type II B_2-critical curve is tangent to a B_2-boundary (resp. a type I B_2-critical curve): The X_1-locus for this event is a circular arc of radius r_1+r_3 about an intersection point of an r_3-displaced wall or corner with another $(2r_2+r_3)$-displaced (resp. $(2r_2+3r_3)$-displaced) wall or corner.

(22) A B_1-induced type II B_2-critical curve is tangent to a type II B_2-critical curve: The X_1-locus for this event is a circular arc of radius r_1+r_3 about the intersection of an r_3-displaced wall or corner with the circle of radius $2r_2+2r_3$ whose center is the intersection of two r_3-displaced walls or corners.

(23) (resp. (24),(25)) A B_1-induced type II B_2-critical curve passes thru an intersection of two B_2-boundaries (resp. a B_2-boundary with a type I B_2-critical curve, two type I B_2-critical curves): The X_1-locus for this event is a circular arc of radius r_1+r_3 about the intersection of an r_3-displaced wall or corner with the circle of radius r_2+r_3 whose center is the intersection of two r_2-displaced (resp. one r_2-displaced and one (r_2+2r_3)-displaced, two (r_2+2r_3)-displaced) walls or corners.

(26) (resp. (27)) A B_1-induced type II B_2-critical curve passes thru an intersection of a B_2-boundary (resp. a type I B_2-critical curve) and a B_2-type II critical curve: The X_1-locus for this event is a circular arc of radius r_1+r_3 about the intersection of an r_3-displaced wall with the circle of radius r_2+r_3 whose center is the intersection of an r_2-displaced wall or corner (resp. an (r_2+2r_3)-displaced wall or corner) with the circle whose center is the intersection of two r_3-displaced walls.

(28) A B_1-induced type II B_2-critical curve passes thru an intersection of two type II B_2-critical curves: The X_1-locus for this event is a circular arc of radius r_1+r_3 about the intersection of an r_3-displaced wall with the circle of radius r_2+r_3 whose center is the intersection of two circles, each of which is of radius r_2+r_3 and has its center at an intersection of two r_3-displaced walls.

Note that all these critical curves lie on circles of radius r_1+r_3 about one or another center.

A final group of B_2-critical curves is obtained by considering situations in which two or more B_2-critical curves, both induced by B_1, have a critical contact. As already noted in the remarks immediately following Table 3.1, these critical intersections can only arise from interaction of a B_1-induced type II critical curve with another B_1-induced boundary or critical curve and/or with stationary B_2-boundary or critical curves. Moreover, a B_1-induced type II B_2-critical curve can never be tangent to a B_1-induced B_2-boundary, and is always tangent to a type I B_2-critical curve (see Fig. 3.4). Hence these potentially critical tangencies do not generate any B_1-critical point.

The remaining cases yield the following B_1-critical curves:

(29) (resp. 30) A B_1-induced type II B_2-critical curve passes through the intersection of a B_1-induced B_2-boundary and a stationary B_2-boundary (resp. a type I B_2-critical curve): The points X_1 at which this occurs lie along the locus traversed by the vertex A of the triangle ABC, defined by the conditions $|AB| = r_1+r_3$, $|AC| = r_1+r_2$, $|BC| = r_2+r_3$, as the triangle moves in such a way that B glides along an r_3-displaced wall or corner and C glides along an r_2-displaced (resp. an (r_2+2r_3)-displaced) wall or corner.

(31) (resp. (32)) A B_1-induced type II B_2-critical curve passes through the intersection of a B_1-induced type I B_2-critical curve and a stationary B_2-boundary (resp. a type I B_2-critical curve): same as in

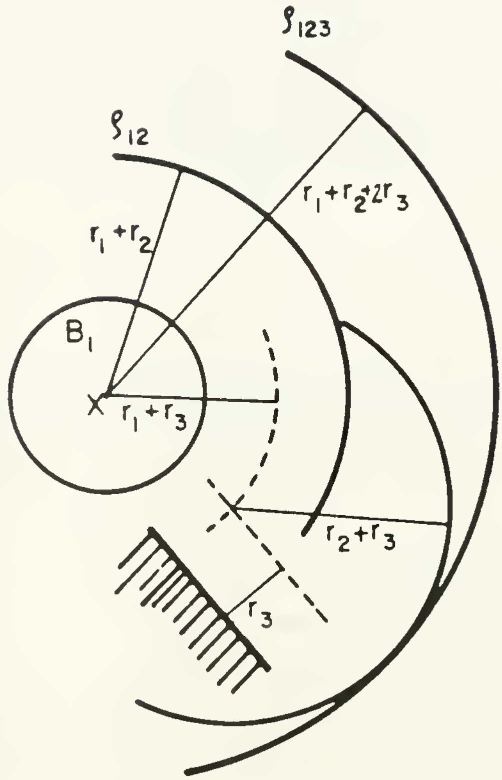

Fig. 3.4. Persistent and Impossible Touches Between B_1-induced B_2-critical Curves

cases (29), (30), except that the triangle ABC is now defined by the conditions $|AB| = r_1+r_3$, $|BC| = r_2+r_3$, $|AC| = r_1+r_2+2r_3$, and therefore degenerates into a straight line segment.

(33) (resp. (34)) Two B_1-induced type II B_2-critical curves and a stationary B_2-boundary (resp. a type I B_2-critical curve) have a common intersection point: Consider a quadrangle ABCD, defined so that $|AB| = |AD| = r_1+r_3$, $|BC| = |CD| = r_2+r_3$, and assume it to be hinged at its vertices. Then the curves of type (33) (resp. (34)) are loci of points traversed by the vertex A as the quadrangle ABCD moves in such a way that the vertices B and D glide along r_3-displaced walls or corners

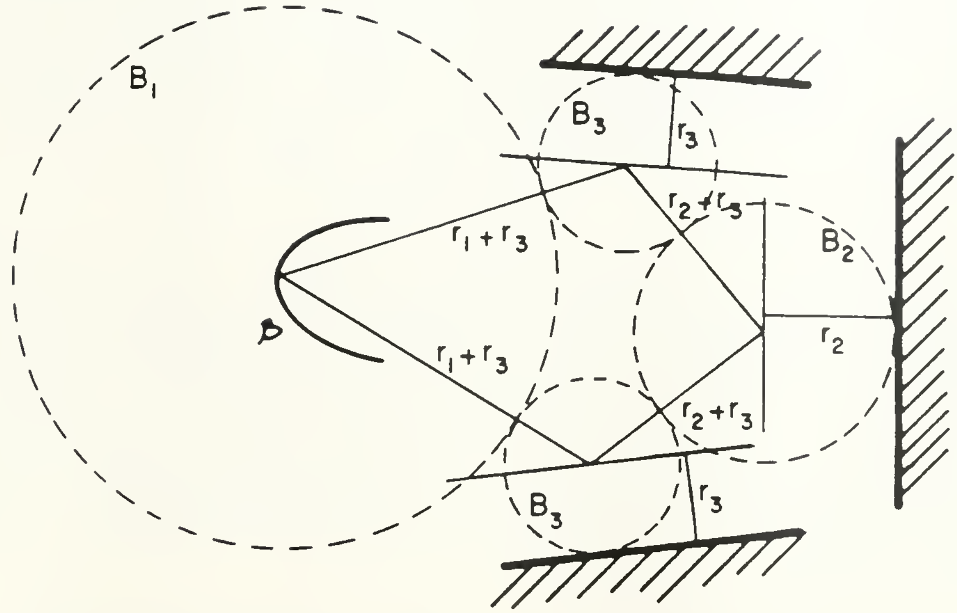

Fig. 3.5. A Type (33) Critical Curve $\mathscr{E}$, and Corresponding Critical Positions of the Three Circles.

and the vertex C glides along an r_2-displaced (resp. (r_2+2r_3)-displaced) wall or corner (cf. Fig. 3.5).

(35) A B_1-induced type II B_2-critical curve passes through the intersection of a B_1-induced B_2-boundary and a stationary type II B_2-critical curve: Let ABC be the triangle appearing in the definition of the critical curves of types (29) and (30). Critical curves of type (35) are then loci of points traversed by the vertex A as ABC moves with B gliding on an r_3-displaced wall or corner and C gliding along the circle of radius r_2+r_3 whose center is an intersection point of two r_3-displaced walls or corners.

(36) A B_1-induced type II B_2-critical curve passes through the intersection of a B_1-induced type I B_2-critical curve and a stationary type II B_2-critical curve: same as in case (35), except that the triangle ABC is replaced by the degenerate triangle used in cases (31) and (32).

(37) Two B_1-induced type II B_2-critical curves and a stationary type II B_2-critical curve have a common intersection point: same as in the cases (33) and (34), except that the vertex C of the hinged quadrangle ABCD glides along the circle of radius r_2+r_3 whose center is the intersection of two r_3-displaced walls or corners.

(38) Two B_1-induced type II B_2-critical curves intersect a B_1-induced B_2-boundary curve at a common point: Let ABCD be the quadrangle whose sides are as defined in cases (33) and (34), but assume this time that this quadrangle is rigid and that its diagonal satisfies $|AC| = r_1+r_2$. Then critical curves of type (38) are the loci of points traversed by the vertex A as ABCD moves so that its vertex B slides along one r_3-displaced wall or corner, and D slides along another such displaced wall or corner.

All critical curves of classes (29)-(38) are relatively simple algebraic curves. Curves of types (29), (30), (31), (32), (35), (36) and (38) are all 'glissetes' (cf. [Lo]) traversed by one vertex of a triangle or rigid quadrangle as its other vertices traverse a straight line or circle. Curves of types (33), (34) and (37) are produced by one vertex of a hinged quadrangle as its other three vertices slide along lines or circles.

We have now listed all the curves in the space of the variable X_1 along which there can occur configurations causing discontinuities in $CG(X_1)$. These curves partition the space A_1 of B_1-admissible positions into finitely many connected regions, which we will call <u>B_1-noncritical regions</u>. Our next task is to show that these regions possess properties analogous to those of noncritical regions of other cases of the Movers' problem studied previously. The following lemma accomplishes this.

<u>Lemma 3.1:</u> Let R be a B_1-noncritical region R. Then $CG(X_1)$ is constant for $X_1 \in R$. Moreover, for each connected component C of $CG(X_1)$, the corresponding connected component $\mu(X_1,C)$ of $P(X_1)$ depends continuously (in the Hausdorff metric) on $X_1 \in R$. Moreover, for it to be possible to

move continuously in FP from one free configuration $[X_1,X_2,X_3] \in FP$, $X_1 \in R$, to another such configuration $[X_1',X_2',X_3'] \in FP$, $X_1' \in R$ in such a way that the center of B_1 remains in R throughout the motion, it is necessary and sufficient that the connected component of $P(X_1)$ containing $[X_2,X_3]$ should have the same label as the connected component of $P(X_1')$ containing $[X_2',X_3']$.

<u>Proof:</u> That $CG(X_1)$ remains constant for $X_1 \in R$ follows from the definition of B_1-noncritical regions, because any change in $CG(X_1)$ must occur at points lying on some B_1-critical curve. We can also re-establish this fact, together with the continuity of μ, as follows. Let $X_1 \in R$ be a noncritical point, and let C be a connected component of $CG(X_1)$. Then when the center of B_1 varies in a sufficiently small neighborhood of X_1, all B_2-boundaries and critical curves, as well as their points of intersection, will vary continuously with B_1. It follows that the number of B_2-noncritical regions onto which C projects remains unchanged, and that for each $[L,T] \in C$ the B_2-noncritical region $\tau(X_1,L)$ remains nonempty and varies continuously with X_1. Let $[L,T] \in C$, and suppose that X_1 varies in some small neighborhood U contained in R. Then it is easily seen that there exists an open nonempty set N which is contained in each of the B_2-noncritical regions $\tau(X_1,L)$, for $X_1 \in U$. Let us fix the center of B_2 at some $X_2 \in N$. Then the connected component $K(X_1) = \psi_{X_1}(X_2,T)$ of $Q(X_1,X_2)$ will depend continuously on $X_1 \in U$. Indeed, each boundary arc of $K(X_1)$ lies either along an r_3-displaced wall or corner, or along $\rho_{13}(X_1)$, or along the circle $\rho_{23}(X_2)$ of radius r_2+r_3 about X_2. Moreover, for each $X_1 \in U$, $\rho_{23}(X_2)$ is not tangent to any of the other boundary curves, and does not pass through the intersection of two of these curves, because then X_2 would lie on a B_2-critical curve, contrary to assumption. Similarly, for each $X_1 \in U$, none of the other boundary curves of $K(X_1)$ becomes tangent to another such curve, and no three of them pass through the same point, since any of these conditions would imply that X_1 lies on a type (1) or type (2) B_1-critical curve, contrary to assumption. It follows that each boundary arc of $K(X_1)$ and its endpoints depend continuously on X_1, for $X_1 \in U$, so that $K(X_1)$ itself varies continuously with $X_1 \in U$. Finally, if the center of B_2 is placed

on a B_2-critical curve β at some point X_2 which does not lie on any other B_2-critical curve, then moving X_1 slightly will either leave β unmoved, or move it slightly so that if X_2 is kept on that curve it will still not meet any other B_2-critical curve, and moreover the crossing rules at X_2 will remain unchanged (see the corollary to Lemma 1.9, and also condition (B) of Lemma 1.8). All this implies that $CG(X_1)$ does not change in a sufficiently small neighborhood of X_1, and that $\mu(X_1,C)$ varies continuously with $X_1 \in R$.

The 'only if' part of the final assertion of the lemma is obvious, due to the continuity of the maps μ just established. For the 'if' part, suppose that $[X_1,X_2,X_3]$, $[X_1',X_2',X_3'] \in FP$ with $X_1,X_1' \in R$ are such that the connected component K of $P(X_1)$ and the connected component K' of $P(X_1')$ are both labeled with C, where C is some connected component of $CG(X_1) = CG(X_1')$. By the definition of this labeling it follows (excluding cases where X_2 or X_2' lie on a B_2-criticial curve) that there exist two nodes $[S,T]$, $[S',T']$ of C such that $X_2 \in \tau(X_1,S)$, $X_2' \in \tau(X_1',S')$, and such that X_3 (resp. X_3') lies in $\psi_{X_1}(X_2,T)$ (resp. $\psi_{X_1'}(X_2',T')$). Since $[S,T]$ and $[S',T']$ are connected to each other in the 2-circle connectivity graph $CG(X_1')$, we can move B_2 and B_3 continuously (keeping B_1 fixed with its center at X_1') in a collision-free motion from the configuration $[X_2',X_3']$ to any configuration $[X_2'',X_3'']$ such that $X_2'' \in \tau(X_1',S)$ and $X_3'' \in \psi_{X_1'}(X_2'',T)$. Let $c(t)$, $0 \leq t \leq 1$ be any continuous motion of the center of B_1 from X_1 to X_1' which is fully contained in R. For each $0 \leq t \leq 1$, let $\gamma_1(t)$, $\gamma_2(t)$ be two boundary arcs of $\tau(c(t),S)$ meeting at a corner $D(t)$. Since $c(t)$ is always noncritical, there clearly exists a continuous function $d(t) \in \tau(c(t),S)$, $t \in [0,1]$, which can be effectively constructed by choosing $d(t)$ inside an angular neighborhood of $D(t)$ lying between the curves $\gamma_1(t)$ and $\gamma_2(t)$ and sufficiently near $D(t)$. Similarly, one can construct another continuous function $e(t) \in \psi_{c(t)}(d(t),T)$, $t \in [0,1]$. Put $U = d(0)$, $V = e(0)$, $U' = d(1)$ and $V' = e(1)$. Then the desired motion of the three bodies can be constructed as follows:

(i) Move B_2 and B_3 so that their centers lie on U,V respectively. Since B_1 does not move, this is an instance of intra-region motion of two circles, which is possible because X_2 and U both lie in the same B_2-noncritical region $\tau(X_1,S)$, and because $X_3 \in \psi_{X_1}(X_2,T)$ and $V \in \psi_{X_1}(U,T)$ (see Lemma 1.5).

(ii) Move the centers of B_1, B_2 and B_3 from $[X_1,U,V]$ to $[X_1',U',V']$ along the curves c(t), d(t) and e(t) respectively.

(iii) Move B_2 and B_3 so that their centers move from $[U',V']$ to $[X_2',X_3']$. As already noted, since B_1 does not move, this is a two circle motion-planning operation, and the required motion is possible because the marking [S,T] of $[U',V']$ and the marking $[S',T']$ of $[X_2',X_3']$ belong to the same connected component C of $CG(X_1')$ (see Theorem 1.1). Q.E.D.

<u>Remark:</u> The functions d(t) and e(t) used in the preceding proof can in many (but not all) cases be taken to be constant positions, as indicated by the proof of Lemma 1.5. However, there do exist B_2-noncritical regions whose entire area moves with B_1 (e.g. regions bounded between the two circular arcs $\rho_{12}(X_1)$ and $\rho_{123}(X_1)$ about the center X_1 of B_1), for which we must move the circles B_2 and B_3 simultaneously with B_1 in step (ii) of the preceding proof.

The following definition is analogous to Definition 1.4.

<u>Definition 3.1:</u> Given a B_1-noncritical region R, we let $\sigma(R)$ denote the set of all components of $CG(X_1)$, where X_1 is any point in R.

<u>Corollary 3.1:</u> Let $\Gamma(R)$ denote the subset of all free configurations $[X_1,X_2,X_3]$ satisfying $X_1 \in R$. Then $\Gamma(R)$ is the disjoint union of its connected components $\Gamma([R,C])$, for $C \in \sigma(R)$, where $\Gamma([R,C])$ is defined as the set of all configurations $[X_1,X_2,X_3]$ such that $[X_2,X_3] \in \mu(X_1,C)$.

<u>Proof:</u> Immediate from the proof of Lemma 3.1.

Next we consider the 'crossing rules' that apply as X_1 crosses some critical curve β between two adjacent B_1-noncritical regions. Using an appropriate modification of Lemma 1.6 as in the two-circle case, we see that we need only consider the case in which X_1 crosses β transversally along some short straight line segment, at a point which does not lie on any other B_1-critical curve, and in which X_2 and X_3 are fixed during that motion. The following definition and lemma generalizes Lemma 1.7:

Definition 3.2: Let the critical curve β be part of the boundary of a B_1-noncritical region R, and let $C \in \sigma(R)$. Put $\zeta = [R,C]$. Then $\nu(X,\zeta)$ is the set of all limit points of sequences $[Y_n,Z_n]$ such that $[Y_n,Z_n] \in \mu(X_n,C)$, taken over all sequences $X_n \in R$ such that $X_n \to X$.

Lemma 3.2: Suppose that (a portion of) the critical curve β forms part of the boundary of a B_1-noncritical region R, and that $C \in \sigma(R)$. Put $\zeta = [R,C]$. Then the interior of the set $\nu(X,\zeta)$ is contained in $P(X)$ and is a union of connected components of $P(X)$, and the boundary of $\nu(X,\zeta)$ is contained in SFP-P(X). If $C' \in \sigma(R)$, $C' \neq C$, and $\zeta' = [R,C']$, then $\mathrm{int}(\nu(X,\zeta))$ and $\nu(X,\zeta')$ are disjoint for each $X \in \beta$. Moreover, for each $Z_1,Z_2 \in V$ the set $\{X \in \beta : [Z_1,Z_2] \in \mathrm{int}(\nu(X,\zeta))\}$ is open in β.

Proof: Let $[Y,Z]$ be an interior point of $\nu(X,\zeta)$. We claim that $[X,Y,Z] \in$ FP. Indeed, since $\nu(X,\zeta)$ is a set of limits of sequences of free configurations, it follows that $\nu(X,\zeta)$ must be contained in SFP. Suppose that $[X,Y,Z]$ belonged to SFP-FP. Then in this configuration at least one of the circles B_2, B_3 must touch another circle or a wall, for otherwise B_1 would have to touch some wall, so that X would lie on the boundary of A_1, contrary to assumption. It is then plain that every neighborhood of $[Y,Z]$ contains configurations which are not semi-free, which contradicts the fact that a small neighborhood of $[Y,Z]$ is contained in $\nu(X,\zeta)$, thus in SFP.

Next let $[Y,Z]$ be a point on the boundary of $\nu(X,\zeta)$. We claim that $[X,Y,Z] \in$ SFP-FP. Indeed, suppose to the contrary that $[X,Y,Z] \in$ FP. Since $[Y,Z] \in \nu(X,\zeta)$, $[X,Y,Z]$ is the limit of a sequence of (free) configurations $[X_n,Y_n,Z_n]$, where $X_n \in R$, $Y_n \in \tau(X_n,L)$ and $Z_n \in \psi_{X_n}(Y_n,T)$, for some $[L,T] \in C$. Since FP is open there exists a fixed neighborhood U of $[0,0]$ such that the sets $[X_n,Y_n,Z_n] + \{0\}\times U$ are contained in FP for all sufficiently large n. It follows that the set $[Y_n,Z_n] + U$ is contained in $\mu(X_n,C)$ for all sufficiently large n, and consequently the set $[Y,Z] + U$ is contained in $\nu(X,\zeta)$, contradicting our assumption that $[Y,Z]$ is a boundary point of $\nu(X,\zeta)$.

It follows that any connected component K of P(X) which intersects $\mathrm{int}(\nu(X,\zeta))$ is contained in $\nu(X,\zeta)$, for otherwise there would exist a point p in K belonging to the boundary of $\nu(X,\zeta)$, and hence $\{X\}\times p$ would belong to SFP−FP, contradicting the fact that $\{X\}\times P(X)$ is a subset of FP. This proves that $\mathrm{int}(\nu(X,\zeta))$ is a union of connected components of P(X).

Next let $C' \in \sigma(R)$, $C' \neq C$, and let $\zeta' = [R,C']$. Suppose that there exists a point $[Y,Z] \in \mathrm{int}(\nu(X,\zeta))$ which also belongs to $\nu(X,\zeta')$. By the foregoing, $[X,Y,Z] \in$ FP, so that there exists a connected neighborhood U of $[Y,Z]$ such that $\{X_n\}\times U$ is contained in FP for all sufficiently large n. Thus for large n the set U is contained in a single connected component of $P(X_n)$, and hence is either contained in $\mu(X_n,C)$ or else lies wholly outside it. However, the second case is impossible for large n because then $[Y,Z]$ could not be a limit of points in $\mu(X_n,C)$. It follows that $[Y,Z] \in \mu(X_n,C)$ for all large n. However, since $[X,Y,Z]$ is a free configuration and $[Y,Z] \in \nu(X,\zeta')$ we must have $[Y,Z] \in \mathrm{int}(\nu(X,\zeta'))$, so that, arguing as above, we can deduce that $[Y,Z] \in \mu(X_n,C')$ for all large n, which is impossible, since the sets $\mu(X_n,C)$ and $\mu(X_n,C')$ are disjoint for $C \neq C'$.

Next we show that for each $Y,Z \in V$ the set

$$D(Y,Z) = \{X \in \beta : [Y,Z] \in \mathrm{int}(\nu(X,\zeta))\}$$

is open in β. For this, let X belong to D(Y,Z), i.e. $[Y,Z] \in \mathrm{int}(\nu(X,\zeta))$. As above, $\zeta = [R,C]$. By what we have just shown, $[X,Y,Z] \in$ FP, and hence for all $X' \in R$ lying in a sufficiently small neighborhood U' of X the point $[Y,Z]$ belongs to some set $\mu(X',C')$ with C' fixed. (Indeed, choose U' to be a neighborhood of X whose intersection $U = U'*R$ with R is arcwise connected, and let U' be sufficiently small so that $[X',Y,Z] \in FP$ for all $X' \in U$. Then $U\times\{[Y,Z]\}$ is a connected subset of $\Gamma(R)$, and so must be contained in a single one of the connected components of $\Gamma(R)$). Passing to the limit X through points in U we find that $[Y,Z] \in \nu(X,[R,C'])$, and since we have just shown that $\nu(X,[R,C'])$ and $\mathrm{int}(\nu(X,[R,C]))$ are disjoint if $C \neq C'$, we must have $C = C'$. In the same way, passing to limits $X' \in \beta$ through sequences of points in U, we see that $[Y,Z] \in \mathrm{int}(\nu(X',\zeta))$ for all $X' \in \beta$ sufficiently near X, so that the set $D(Y,Z)$ must be open. Q.E.D.

Next we generalize Lemma 1.8 to the three-circle case. The formulation and proof of the generalized lemma turn out to be almost identical to the material presented in Section 1. For completeness sake we state the generalized lemma, to wit

Lemma 3.3: Suppose that (a portion of) a smooth critical curve β separates two connected B_1-noncritical regions R_1 and R_2 and that $R_1+R_2+\beta$ is open. Let $C_1 \in \sigma(R_1)$, and $C_2 \in \sigma(R_2)$. Put $\zeta_1 = [R_1,C_1]$ and $\zeta_2 = [R_2,C_2]$, and let $\Gamma_1 = \Gamma(\zeta_1)$, $\Gamma_2 = \Gamma(\zeta_2)$. Then the following conditions are equivalent:

Condition A: There exists a point $X \in \beta$ such that the open sets $\mathrm{int}(\nu(X,\zeta_1))$, $\mathrm{int}(\nu(X,\zeta_2))$ have a non-null intersection.

Condition B: There exists a smooth path $c(t) = [x(t),y(t),z(t)] \in FP$ which has the following properties:

(i) $c(0) \in \Gamma_1$, $c(1) \in \Gamma_2$;

(ii) $x(t) \in R_1+R_2+\beta$ for all $0 \leq t \leq 1$;

(iii) $x(t)$ crosses β just once, transversally, when $t=t_0$, $0<t_0<1$, and $y(t)$ and $z(t)$ are constant for t in the vicinity of t_0.

Proof: Essentially identical to that of Lemma 1.8, and therefore omitted. Q.E.D.

Next we generalize Lemma 1.9 to the three-circle case, i.e. we show that if condition A of the preceding lemma holds for one point lying on a portion β' of β not intersected by any other B_1-critical curve, this same condition holds for all points of β'.

Lemma 3.4: Let the smooth B_1-critical curve β separate two B_1-noncritical regions R_1 and R_2. Let β' be a connected open segment of β not intersecting any other B_1-critical curve, and suppose that $R_1+R_2+\beta'$ is open. Let C_1, C_2, ζ_1, ζ_2 be defined as in the preceding lemma. Then the set of $X \in \beta'$ for which the open sets $\mathrm{int}(\nu(X,\zeta_1))$ and $\mathrm{int}(\nu(X,\zeta_2))$ have a non-null intersection is either all of β' or is empty.

Proof: Let M be the set of all $X \in \beta'$ for which the sets $\mathrm{int}(\nu(X,\zeta_1))$ and $\mathrm{int}(\nu(X,\zeta_2))$ have a point in common. Since M is the union of sets of the form

$$\{X \in \beta' : [Y,Z] \in \mathrm{int}(\nu(X,\zeta_1))\} * \{X \in \beta : [Y,Z] \in \mathrm{int}(\nu(X,\zeta_2))\} ,$$

taken over $Y,Z \in V$, and since by Lemma 3.2 each of these sets is open, it follows that M is open relative to β'. Hence we have only to show that M is also closed relative to β'. Suppose the contrary; then there exists an $X \in \beta'$ such that $\mathrm{int}(\nu(X,\zeta_1))$ and $\mathrm{int}(\nu(X,\zeta_2))$ are disjoint but for which there also exists a sequence X_n of points on β' converging to X such that for all n the sets $\mathrm{int}(\nu(X_n,\zeta_1))$ and $\mathrm{int}(\nu(X_n,\zeta_2))$ intersect each other.

By Lemma 3.2, for each $n \geq 1$ the sets $\mathrm{int}(\nu(X_n,\zeta_j))$, $j=1,2$, are unions of connected components of $P(X_n)$. Thus, passing to a subsequence if necessary, we can assume that for each $n \geq 1$ both sets $\mathrm{int}(\nu(X_n,\zeta_j))$, $j=1,2$, contain a connected component K_n of $P(X_n)$ whose label $\lambda(K_n)$ is constant. To study the limiting behavior of the sets K_n as $n \to \infty$, we begin by observing that for any limit point $[Y,Z]$ of a sequence of points in K_n, the point $[X,Y,Z]$ must belong to SFP-FP. Indeed, if $[X,Y,Z]$ were free there would exist a small neighborhood U of $[Y,Z]$ such that for all sufficiently large n the sets $\{X_n\}\times U$ would be wholly contained in FP. Hence for large n the set U would either be wholly contained in K_n, or would lie wholly outside K_n. However the second alternative cannot hold for all large n, since if it did $[Y,Z]$ would not be a limit of points in K_n. Thus if $[X,Y,Z]$ is free, it follows that $[Y,Z] \in \mathrm{int}(\nu(X_n,\zeta_j))$, $j=1,2$, for all sufficiently large n. On the other hand, if $[X,Y,Z]$ is free, it is easily seen that $[Y,Z]$ must belong to the interior of $\nu(X,[R_1,C_1'])$ for some $C_1' \in \sigma(R_1)$, and also to the interior of $\nu(X,[R_2,C_2'])$ for some $C_2' \in \sigma(R_2)$. But then it follows from the final statement of Lemma 3.2 that $[Y,Z]$ belongs to the interior of $\nu(X_n,[R_j,C_j'])$ for all large n, so that (using Lemma 3.2 once more) $C_j=C_j'$ for $j=1,2$. However, this contradicts our initial assumption that the sets $\mathrm{int}(\nu(X,\zeta_j))$, $j=1,2$, do not intersect each other.

Thus as asserted each point $[Y,Z]$ in the set K of limit points of sequences of points in K_n is such that $[X,Y,Z]$ is in SFP-FP. Next we analyze the possible forms of such a K. For simplicity, let us assume that each of the sets K_n is just a single cell of FP, corresponding to a single node $[L,T] \in CG(X_n)$, for each $n \geq 1$. Projecting K_n and K into the subspace A_2 of positions of the center of B_2, we obtain a corresponding sequence of B_2-noncritical regions $\tau(X_n,L)$, and plainly the set of all limit points of sequences of points $Y_n \in \tau(X_n,L)$ coincides with the projection J of K into A_2. J is clearly a connected set, and since the boundary of each of the regions $\tau(X_n,L)$ consists of some fixed finite number of straight and circular arcs J must have one of the three following forms:

I. J consists of a single point Y.

II. J consists of some arcs, but has empty interior.

III. J has a nonempty interior J_0.

As noted, each of the regions $\tau(X_n,L)$ is bounded by a fixed sequence of B_2-boundary arcs and critical curves; some of these curves may move with X_n, but each such moving curve is a circular arc, either of radius r_1+r_2 about X_n, or of radius $r_1+r_2+2r_3$ about X_n, or of radius r_2+r_3 about an intersection of an r_3-displaced wall with the circle of radius r_1+r_3 about X_n. Each of these moving curves converges as $X_n \to X$ to a limit curve which is contained in a circle of one of these three radii centered at X.

Consider case I first. It is clear that in this case all the boundary arcs of $\tau(X_n,L)$ must converge to the single point Y as $X_n \to X$. This can only happen if either the limiting positions of at least three curves containing boundary arcs of $\tau(X_n,L)$ pass through the point Y, or, if the boundary of $\tau(X_n,L)$ consists of just two arcs, if the limiting positions of the two curves containing these arcs become tangent to each other at Y. Moreover, in either case, for each $n \geq 1$, this triple intersection or tangency does not occur at the region $\tau(X_n,L)$. However, points X for which such a triple intersection or tangency of B_2-boundary or critical curves occurs must lie on the B_1-critical curve β_1 that is defined by this geometric condition. Thus $X \in \beta_1$, while none of the points X_n belong to β_1. But then X belongs to the two distinct critical curves β', β_1, contrary to assumption.

Next consider case II. Let $\beta_1(X_n)$ and $\beta_2(X_n)$ be two adjacent boundary arcs of $\tau(X_n,L)$, meeting at a corner D_n of that region. Suppose that as $n\to\infty$ D_n converges to some point D. Then D lies on the limiting positions β_1 and β_2 of the two curves (lines or circles) containing $\beta_1(X_n)$ and $\beta_2(X_n)$. If these two limiting curves do not overlap each other, then either they become tangent to each other, or else there must exist a third boundary arc of $\tau(X_n,L)$ which converges to D, for otherwise a small angular neighborhood of D between β_1 and β_2 would be wholly contained in J, contrary to our assumption. If either of these two conditions occurs then, arguing as in case I above, we deduce that X must lie on another B_1-critical curve which does not contain the sequence X_n, which is impossible by our assumptions. Next consider the case in which the two limiting curves β_1 and β_2 overlap each other (and in which this overlap does not occur in any of the regions $\tau(X_n,L)$). Such an overlap cannot occur between two straight boundary segments, because these segments do not move as $X_n \to X$. Moreover, among the various B_2-boundary and critical curves there exists just one circle (i.e. $\rho_{12}(X)$) of radius r_1+r_2 and just one circle (i.e. $\rho_{123}(X)$) of radius $r_1+r_2+2r_3$, which plainly cannot overlap with any other curve. It follows that the only possible overlaps are between two type II B_2-critical curves β_1,β_2, at least one of which moves with X_n. Suppose first that β_2 is a stationary circular arc of radius r_2+r_3 about the intersection D of two r_3-displaced walls or corners γ_1,γ_2, and that β_1 is a circular arc of radius r_2+r_3 about the intersection E_n of an r_3-displaced wall or corner γ_3 with the circle $\rho_{13}(X_n)$ of radius r_1+r_3 about X_n. It follows that E_n is distinct from E for all n, but that the sequence $\{E_n\}$ converges to E as $n \to \infty$. Hence X lies on the type (2) B_1-critical curve β_0, which is a circular arc of radius r_1+r_3 about E, but none of the points X_n lie on β_0, which gives a contradiction as in the preceding case I. Suppose next that both β_1 and β_2 are circular arcs of radius r_2+r_3 about two points $E_n \neq E_n'$ in which the circle $\rho_{13}(X_n)$ intersects two r_3-displaced walls or corners γ_1,γ_2 respectively. If $\gamma_1 = \gamma_2$ then, since E_n and E_n' must approach each other as $n \to \infty$, it follows that X must lie on the type (1) B_1-critical curve at distance r_1+2r_3 from the wall or corner corresponding to γ_1. On the other hand, if $\gamma_1 \neq \gamma_2$ then E_n and E_n' must both converge to a point E at which γ_1 and γ_2 intersect each other, so that X lies on the type (2) B_1-critical circle at radius r_1+r_3 about E. In either case it follows that X lies on a B_1-critical curve other than β, leading to a contradiction as before.

Finally, consider case III, in which J has a nonempty interior J_0. If we eliminate the points lying on a finite number of curves, the remaining $Y \in J_0$ do not lie on any B_2-boundary or critical curve when B_1 is placed at X, and for sufficiently large n no such Y will belong to any B_2-boundary or critical curve when B_1 is placed at X_n. Thus, when B_1 is placed at either X_n or X and B_2 is placed at Y, B_1 and B_2 do not touch either a wall or each other.

Since Y belongs to $\tau(X_n,L)$ for all large n, the sets $M_n^* = \{Z : [Y,Z] \varepsilon K_n\}$ are open and nonempty for all large n. Focus attention on a single connected component M_n of M_n^*. Each of the arcs bounding M_n is either an r_3-displaced wall or corner, or a circular arc of radius r_1+r_3 about X_n, or a circular arc of radius r_2+r_3 about Y, and M_n lies on the outside (convex) side of each circular arc of its boundary. It follows, as in the proofs of Lemmas 1.7 and 1.9, that as $X_n \to X$, M_n converges to a set M which either consists of a single intersection point at which two r_3-displaced walls or corners (one of which may be B_2) come together with $\rho_{13}(X)$, or else M has a nonempty interior. However, the second alternative is impossible, since for each Z in the interior of M the configuration $[X,Y,Z] \varepsilon K$ would plainly belong to FP, contrary to the fact, established above, that K is contained in SFP-FP. Hence M is a singleton $\{D\}$, for some intersection point D of two r_3-displaced walls and corners (one of which may be B_2) which also lies on $\rho_{13}(X)$. By choosing Y ε J_0 for which no intersection D of $\rho_{23}(Y)$ with a displaced wall or corner lies at distance r_1+r_3 from X, we can rule out the case in which D is the intersection of an r_3-displaced wall or corner with the r_3-displaced B_2. It therefore follows that X lies on the type (2) B_1-critical curve β_0 which is a circular arc of radius r_1+r_3 about the corner D, but that none of the points X_n lie on β_0, giving us a final contradiction.

All this proves that M is closed, completing the proof of the present lemma. Q.E.D.

As in Section 1, we can now define the connectivity graph for the case of three circular moving objects, as follows:

<u>Definition 3.3:</u> The connectivity graph CG_2 of an instance of the three-circle movers' problem is an undirected graph whose nodes are all pairs of the form [R,C] where R is some connected B_1-noncritical region (bounded by B_1-critical curves) and where C ε $\sigma(R)$. The graph CG_2 contains an edge connecting [R,C] and [R′,C′] if and only if the following conditions hold:

(1) R and R′ are adjacent and meet along a B_1-critical curve β.

(2) For some one of the open connected portions β' of β contained in the common boundary of R and R′ and not intersecting any other B_1-critical curve, and for some (and hence every) point $X \in \beta'$ the sets $\mathrm{int}(\nu(X,[R,C]))$ and $\mathrm{int}(\nu(X,[R',C']))$ have a non-null intersection.

We can now state the main result of this section.

Theorem 3.1: There exists a continuous motion c of B_1, B_2 and B_3 through the space FP of free configurations from an initial configuration $[X,Y,Z]$ to a final configuration $[X',Y',Z']$ if and only if the vertices $[R,C]$ and $[R',C']$ of the connectivity graph CG_2 introduced above can be connected by a path in CG_2, where R, R′ are the B_1-noncritical regions containing X, X′ respectively, and where C (resp. C′) is the label of the connected component in P(X) (resp. P(X′)) containing $[Y,Z]$ (resp. $[Y',Z']$).

Proof: The proof is an obvious modification of the proof of Theorem 1.1, and is therefore omitted here. We note that the present theorem follows from Lemmas 3.1-3.4 in much the same way that Theorem 1.1 follows from Lemmas 1.5-1.9. Q.E.D.

To conclude this section we will say a few words about the crossing rules associated with the various kinds of B_1-critical curves β listed above. As in the two-circle case, each of these rules describes a situation falling into one of the following three categories: either

(a) One or more components of $P(X_1)$ split, each into two components, as X_1 crosses β, which then pull away from each other as we move into the region lying on the other side of β. Conversely, various pairs of components may make contact and fuse together as β is crossed. A related possibility is that two otherwise disjoint portions ('lobes') of a single connected component should make contact as β is crossed, or conversely that a portion of one connected component should thin down to a point and then separate, but without connectivity being lost. Each of these latter cases represents a situation in which one component of $P(X_1)$ changes its label as β is crossed. In addition to these structural changes, some other components of $P(X_1)$ can change their labels at such a crossing, simply because some of the B_2-noncritical regions onto which these components project may change their labels, due to the appearance of a new boundary edge, or to the disappearance of an existing boundary edge (see below for details); or

(b) One or more components of $P(X_1)$ shrink, each to a point, and then disappear as X_1 crosses β (or vice versa); as before, some additional components of $P(X_1)$ may change their labels as β is crossed, for reasons similar to those mentioned above; or

(c) The number of connected components of $P(X_1)$ does not change at such a crossing, but some components change their labels as β is crossed.

Earlier, in preparing to describe the 38 kinds of B_1-critical curves, we noted that each critical position X_1 of the center of B_1 is associated either with a change in the collection of B_2-noncritical regions of $P(X_1)$, or (in the case of type (1) and type (2) B_1-critical curves β) with a change in the collection of B_3-components of $Q(X_1,X_2)$ which occurs uniformly for an entire region of positions X_2 of the circle B_2. Moreover, the combinatorial descriptors that have been associated with components of $P(X_1)$ are sets, namely sets of pairs [L,T] comprising a single component of the connectivity graph $CG(X_1)$; here L describes some B_2-noncritical region of the set $A_2'(X_1)$ of all positions admissible for B_2 if the center of B_1 is placed at X_1, and T is an edge sequence which describes a connected component of $Q(X_1,X_2)$ for each X_2 in the region described by L.

Suppose first that β is not of type (1) or type (2), so that when X_1 crosses β the set of B_2-noncritical regions will change. This change occurs because when X_1 lies on β there occurs either a tangency between two B_2-boundary or critical curves, or a triple intersection of three such curves. Moreover, as we cross β at X_1 from one of the

B_1-noncritical regions R adjacent to β to the region R′ lying on the other side of β, one of the following phenomena will occur: Either

(a) Some B_2-noncritical region splits into two subregions which then pull away from each other (or conversely two such regions meet at a point and then fuse into one another); or

(b) Some B_2-noncritical region shrinks to a point and then either disappears or is replaced by another newly appearing B_2-noncritical region (or conversely some new B_2-noncritical region appears); or

(c) The label of some B_2-noncritical region changes.

It is important to realize that these changes may not always mean that the collection of connected components of $P(X_1)$ will change. Indeed, in order to determine the effect on the structure of $P(X_1)$ of such changes in the structure of B_2-noncritical regions, one first needs to analyze the manner in which the intersection of $P(X_1)$ with $U \times V$ changes, where U is a small neighborhood of the point Y at which the critical tangency or triple intersection of B_2-boundary or critical curves takes place. To see in more detail what this analysis will involve, suppose first that the crossing pattern at X_1 falls into category (a) above, i.e. that a B_2-noncritical region L splits into two subregions L_1, L_2 locally at Y. This will cause each cell $[L,T] \in CG(X')$, for X′ lying on one side of β, to split into two subcells $[L_1,T]$ and $[L_2,T]$ as X′ crosses β at X_1, and there will exist no edge linking these two cells directly, since the two noncritical regions L_1, L_2 will not be adjacent. However this does not necessarily imply that these cells have become disconnected from each other in $P(X_1)$, since it may still be the case that one can cross from positions in the subcell described by $[L_1,T]$ to positions in the subcell described by $[L_2,T]$ by passing through other cells, and in particular there may exist a strictly 'local' connection through a cell which projects onto the B_2-noncritical region that has just appeared between L_1 and L_2 and separated them. To find the cases in which this observation applies, one needs to analyze the geometric details of the neighborhood of the point at which L_1 and L_2 have pulled apart. (Note however that even if such a local analysis rules out relatively direct, local connections between cells $[L_1,T]$, $[L_2,T]$ of $CG(X')$, these cells may still be connected globally via some longer path in $CG(X')$. If this is the case, the structure of $P(X')$ for X' near X_1 will not change: only the way in which we label components by sets of pairs $[L,T]$ will change.)

Similarly, if the crossing at X_1 is of category (b), it may or may not allow one or more components of $P(X_1)$ to shrink and disappear. Some cases in which component disappearance is impossible will be revealed by local analysis of $CG(X')$, for X' near X_1, near the critical position Y of B_2 at which the disappearing B_2-noncritical region L vanishes. In particular, if such analysis shows that every pair $[L,T]$ is necessarily connected in $CG(X_1)$ to a pair $[L',T]$, where L' is a noncritical region adjacent to L which causes L to disappear by 'swallowing' it, then the set of components of $P(X_1)$ will not change even though L disappears; components will simply be renamed.

Finally, if the crossing at X_1 is of category (c), then the structure of the set of connected components of $P(X)$ will not change, though of course the sets labeling its components will generally change in this case also.

Note that the above considerations imply that when a B_1-critical curve β is crossed, several components of $P(X_1)$ (all of which contain cells $[L,T]$ which project onto the same B_2-noncritical region L) may simultaneously split, each into two subcomponents. Similarly, several components $[L,T]$ may shrink and disappear simultaneously if L shrinks and disappears. Additionally, the tangency or triple intersection of B_2-boundary or critical curves which occurs when X_1 comes to lie on a B_1-critical curve β will generally affect the labeling of all B_2-noncritical regions adjacent to the point Y at which a tangency or triple intersection occurs, usually by the appearance or disappearance of one of the boundary edges of these regions. This will cause changes in the labels of all connected components of $P(X_1)$ which contain cells $[L,T]$, for B_2-noncritical regions L adjacent to the point Y, above and beyond component relabelings which result from the splits or disappearances of some of these components.

These general principles underlie the information summarized in Table 3.2 below, which classifies the crossing rules that can be associated with each of the 38 different types of B_1-critical curves listed above. The crossing rules are labeled (a), (b) or (c), corresponding to the three possible changes in $P(X_1)$ listed above. As explained above, a renaming change (c) can always occur if a change of type (a) or (b) is possible.

Type of Critical Curve	Associated Crossing Rules
1	(a) or (c)
2	(b) or (c)
3	(a) or (c)
4	(c)
5	(c)
6	(a) or (c)
7	(c)
8	(b) or (c)
9	(a) or (c)
10	(a) or (c)
11	(c)
12	(b) or (c)
13	(c)
14	(c)
15	(c)
16	(c)
17	(c)
18	(c)
19	(c)
20	(a) or (c)
21	(c)
22	(c)
23	(b) or (c)
24	(c)
25	(c)
26	(c)
27	(c)
28	(c)
29	(b) or (c)
30	(c)
31	(c)
32	(c)
33	(c)
34	(c)
35	(c)
36	(c)
37	(c)
38	(c)

Table 3.2. Critical Curves and their Crossing Rules

To illustrate the statements concerning crossings represented in this table, we will consider a few representative types of B_1-critical curves, and analyze their associated crossing rules in more detail.

First consider crossing a B_1-critical curve of type (6). Recall that such a curve β consists of points X for which the circle γ_1 of radius r_1+r_2 about X becomes tangent to the circle γ_2 of radius r_2+r_3 about an intersection point D of two r_3-displaced walls or corners. Fig. 3.6 shows the structure of B_2-noncritical regions in the neighborhood of the point Y of tangency, for two positions X of the center of B_1, one on either side of β. Note that when X crosses β the B_2-noncritical region L splits into two subregions L_1 and L_2. Moreover, the characteristic sets of L, L_1 and L_2 will all contain a label T, representing the component of $Q(X,X_2)$ containing positions of the center of B_3 which lie in an angular neighborhood near D, for $X_2 \in$ L, L_1 or L_2; however, for $X_2 \in L'$ this component of $Q(X,X_2)$ disappears, so that the label T does not appear in the characteristic of L'. Since

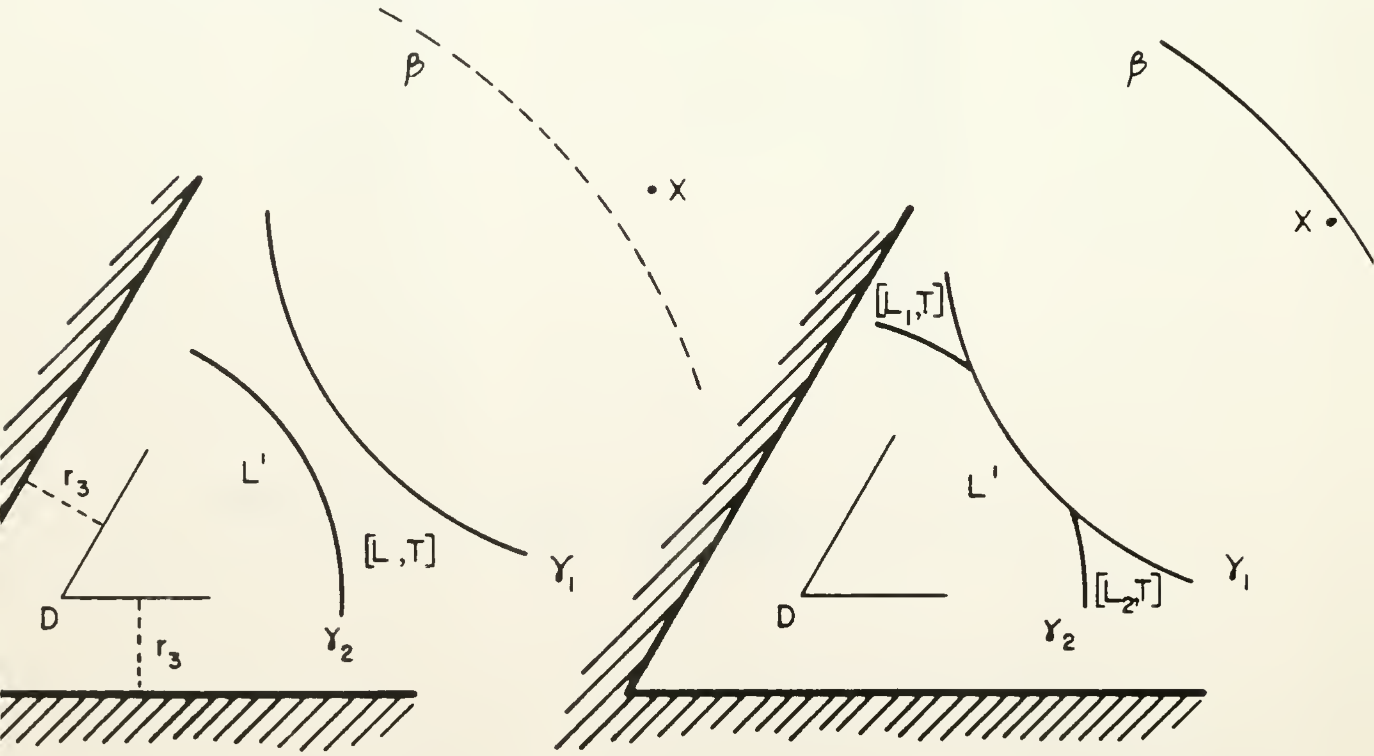

Fig. 3.6. Crossing a Type (6) B_1-critical Curve β; The Structure of B_2-noncritical Regions Near Critical Point of Tangency.

the center of B_2 cannot cross γ_1, it follows that if X lies on the inner side of β, there is no way for B_2 and B_3 to cross from the cell labeled $[L_1,T]$ to the cell labeled $[L_2,T]$, with the center of B_2 remaining near Y. On the other hand, if X lies on the outer side of β, the two cells $[L_1,T]$, $[L_2,T]$ merge into a cell labeled [L,T]. It follows that, unless B_2 and B_3 can cross from $[L_1,T]$ to $[L_2,T]$ along a path involving more global motions, e.g. by moving B_2 around B_1, the crossing rule associated with β is of type (a). Of course, as noted above, it is possible that this locally type (a) crossing really is of type (c) because of global connections ignored by the purely local analysis just given.

Next consider the situation which occurs when we cross a B_1-critical curve β of type (12). Recall that such a curve is the locus of points X for which the circle γ_1 of radius r_1+r_2 about X passes through the intersection point Y of an r_2-displaced wall or corner γ_2 and a circle γ_3 of radius r_2+r_3 about an intersection point D of two r_3-displaced walls or corners. Fig. 3.7 shows the B_2-noncritical regions in the vicinity of Y, for X lying on the inner side of β, and also for X lying on the outer side of β. Note that as X crosses β from its outer side to its inner side, a small B_2-noncritical region L shrinks to the point Y and then disappears. Moreover, the characteristic set of L contains a label T corresponding to the connected component of $Q(X,X_2)$, for $X_2 \in L$, which includes all admissible positions for the center of B_3 which lie in a neighborhood of D. Finally, T does not appear in the (sole) B_2-noncritical region L′ adjacent to L, since when B_2 moves into L′ the component of $Q(X,X_2)$ labeled T will itself shrink to a point and disappear. It follows that in this case the crossing rule at β is of type (b), since the cell of P(X) labeled [L,T] shrinks to a point and disappears as β is crossed, and since in the example [L,T] has no connections to other cells of P(X). However, if the geometry of the walls generating γ_2 and γ_3 were different, so that when B_1 is placed at X and B_2 is placed at Y, B_3 is not 'stuck' at D, but can still move freely near D, then the crossing rule would be of type (c).

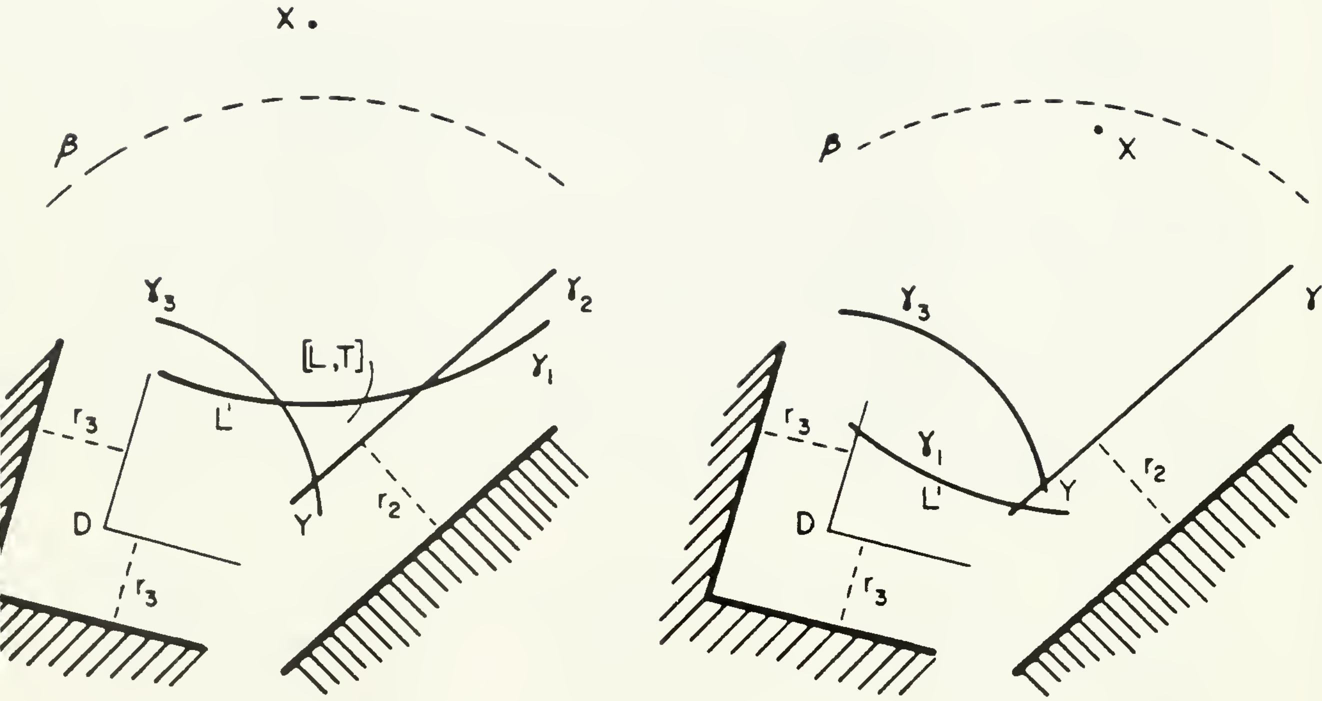

Fig. 3.7. Crossing a Type (12) B_1-critical Curve β; The Structure of B_2-noncritical Regions Near Critical Point of Intersection.

Finally, consider the situation which occurs when we cross a B_1-critical curve β of type (13). Such a curve is defined in much the same way as type (12) critical curves, except that the circle γ_1 about $X \in \beta$ now has radius $r_1+r_2+2r_3$. Fig. 3.8 shows the B_2-noncritical regions for two positions of X, one on either side of β, in the vicinity of the point Y in which the three curves γ_j, $j=1,..,3$, intersect when $X \in \beta$. Note that as X crosses β, the small B_2-noncritical region L shrinks to a point and then disappears as in the preceding case, but this time it has a neighboring region L_1 whose characteristic set also contains the label T, and in the graph CG(X) an edge connects [L,T] to $[L_1,T]$ for X lying on the outer side of β. Thus the disappearance of L does not affect the overall structure of P(X), but only causes a change in the sets which label certain of the components of P(X). More specifically, the component of CG(X) which contains the pairs [L,T] and $[L_1,T]$ will no longer contain [L,T] once β has been crossed from outside to inside, and furthermore, the labeling of the B_2-noncritical regions L_1,L_2 and L_3 will change due to the

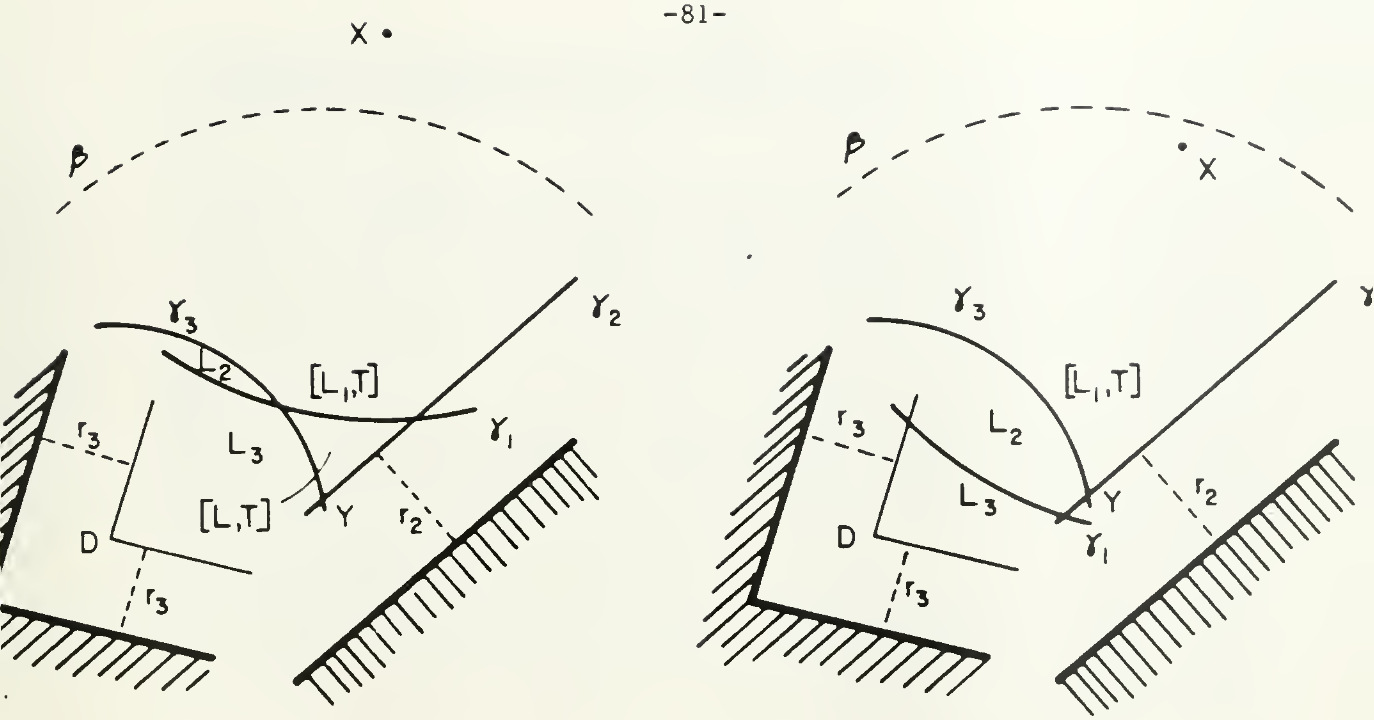

Fig. 3.8. Crossing a Type (13) B_1-critical Curve β; The Structure of B_2-noncritical Regions Near Critical Point of Intersection.

disappearance or addition of a boundary curve in each of them. Therefore the crossing rule applicable to curves of this kind is necessarily of type (c).

The crossing rules which apply to the other 35 types of B_1-critical curves can be derived in much the same way; a summary has been given in Table 3.2 above. We leave it to the reader to work out these rules along the lines of the few examples we have developed in fuller detail.

Once these crossing rules have been derived, we can develop an algorithm for solving the movers' problem for three circles in much the same way as was done in Section 2 for the two-circle case. (Note that the algorithm of Section 2 will be used as a subroutine in the new algorithm to compute the characteristic of some initial B_1-noncritical region.)

The complexity of the algorithm for the 3-circle movers' problem that we have described is bounded below by the size of the connectivity graph which it has to search. We can give the following very crude estimate of this size. First note that the total number of B_1-boundary and critical curves is $O(n^5)$ (there can exist $O(n^5)$ critical curves of type (28) in the worst case). Since these are all algebraic curves of some fixed low degree, they will intersect in at most $O(n^{10})$ points, so that the number of B_1-noncritical regions is also $O(n^{10})$. For each such noncritical region R, the number of connected components of P(X) for any $X \in R$ is bounded by the size of the (two-circle) connectivity graph CG(X) which, as shown in Section 2, can have at most $O(n^3)$ nodes. Thus the size of the three-circle connectivity graph is at most $O(n^{13})$, and, using techniques similar to those outlined in Section 2 for the 2-circle problem, one can construct this graph, and then search it in time no more than $O(n^{13})$.

4. The Case of Arbitrarily Many Circular Bodies.

The treatment of three moving circles in the preceding section used the solution of the two-circle problem repeatedly, both in order to obtain labels for the connected components of P(X), and to construct a path between two specified configurations when such a path exists. This suggests a recursive approach to the motion-planning problem for an arbitrary number of circular bodies moving amidst polygonal barriers. In this section we will give a brief account of the general recursive approach that we propose, omitting most detail, and using informal arguments mainly. Note however that, as can be seen comparing the analysis required in the three-circle case to that which suffices in the two-circle case, the complexity of detail that can be expected to appear in a full treatment by the method to be sketched will increase rapidly with the number of circles.

Let $B_1,\ldots,B_k$ be k circular bodies with centers $C_1,\ldots,C_k$ and radii $r_1 \geq \ldots \geq r_k$ respectively. We assume that these circles are free to move in a polygonal region V, but that none of these circles may touch or penetrate any wall or other circle. Then the space FP of <u>free configurations</u> $[X_1,\ldots,X_k]$ of the centers of these circles forms a 2k-dimensional open manifold, and our problem is to decompose this space into its connected components.

To achieve such a decomposition we can project FP into the 2-dimensional space A_1 of the positions available for the center C_1 of B_1. For each such fixed position X_1 consider the 'fiber' space $P(X_1)$ of all configurations $[X_2,\ldots,X_k]$ of the centers of the remaining circles such that $[X_1,\ldots,X_k] \in FP$. We can decompose $P(X_1)$ into its components by noting that it represents the space of free configurations of the remaining k-1 circles confined to move in the space $V(X_1)$ obtained by adding B_1 as an additional barrier in V. Although B_1 is not polygonal, it can be regarded as a displaced point, and the methods for handling k-1 circles can be adapted to handle displaced walls of this form. Thus, using the algorithm for k-1 circles, we can compute the corresponding connectivity graph $CG(X_1)$, and use each of its connected components to label a corresponding connected component of $P(X_1)$ in a 1-1 manner.

We then divide the points X_1 into critical and noncritical points, where X_1 is critical if the connected components of $P(X_1)$ change discontinuously as X_1 is moved slightly; otherwise X_1 is noncritical. One can show that the critical points lie on finitely many critical curves (although their number increases exponentially with k), which partition A_1 into finitely many noncritical regions. The next step is to generalize Lemmas 3.1-3.4 to this case. That is, one must first show that there exists a a continuous motion between two configurations $[X_1,\ldots,X_k]$ and $[X_1',\ldots,X_k']$, such that X_1,X_1' both belong to some noncritical subregion R, and such that during that motion the first circle moves with its center remaining in R, if and only if the label of the connected component of $P(X_1)$ containing $[X_2,\ldots,X_k]$ and of the component of $P(X_1')$ containing $[X_2',\ldots,X_k']$ are identical. Then one wants to show that for each label C of a connected component $\mu(X,C)$ of $P(X)$ for X in some noncritical region R, the set $\mu(X,C)$ and its interior vary continuously (in the Hausdorff topology of sets) with $X \in R$, and admits a continuous extension to the closure of R. Continuing in analogy with the treatment of the two- and three-circle cases, the next aim is to show that a continuous motion in which the center of B_1 crosses a critical curve β separating between two noncritical regions R, R′ can take place if and only if the initial (resp. final) configuration $[X_1,\ldots,X_k]$ (resp. $[X_1',\ldots,X_k']$) are such that the labels C (resp. C′) of the connected component of $P(X_1)$ (resp. $P(X_1')$ containing $[X_2,\ldots,X_k]$ (resp. $[X_2',\ldots,X_k']$) have the property that the limits of $\mu(X,C)$ (resp. $\mu(X,C'))$ as X approaches the common boundary β from the R-side (resp. the R′-side) have overlapping interiors. Moreover, one wants to show that this condition does not depend on the particular point on β at which B_1 crosses from R to R′.

All these results would enable us to define a finite connectivity graph, in much the same way as was done in the three-circle case, and to reduce the problem to a combinatorial path-searching through that graph; this would yield a recursive solution to the k-circle problem.

R E F E R E N C E S

[HJW] J.E. Hopcroft, D.A. Joseph and S.H. Whitesides, On the movement of Robot Arms in 2-Dimensional Bounded Regions, Tech. Rept. TR 82-486, Comp. Sci. Dept., Cornell University, March 1982.

[IKP] M.B. Ignat'yev, F.M. Kulakov and A.M. Pokrovskiy, Robot Manipulator Control Algorithms, Rep. No. JPRS 59717 NTIS, Springfield, Va. 1973.

[Lo] E.H. Lockwood and A. Prag, Book of Curves, Cambridge University Press, 1961.

[LPW] T. Lozano-Perez and M. Wesley, An Algorithm for Planning Collision-Free Paths Among Polyhedral Obstacles, CACM 22 (1979) pp 560 - 570

[Re] J. Reif, Complexity of the Mover's Problem and generalizations, Proc. 20th Symp. on the Foundations of Computer Science (1979) pp 421 - 427.

[Sc] J.T. Schwartz, Differential Geometry and Topology, Gordon and Breach, New York 1968.

[SS1] J.T. Schwartz and M. Sharir, On the Piano Movers' Problem: I. The Special Case of a Rigid Polygonal Body Moving Amidst Polygonal Barriers, to appear in Comm. Pure Appl. Math.

[SS2] J.T. Schwartz and M. Sharir, On the Piano Movers' Problem: II. General Techniques for Computing Topological Properties of Real Algebraic Manifolds, to appear in Adv. Appl. Math.

[Ud] S. Udupa, Collision Detection and Avoidance in Computer-controlled Manipulators, Ph.D. Dissertation, California Inst. of Technology, Passadena, California 1977.