RR-75-15

THE APPLICATION OF CATASTROPHE THEORY TO ECOLOGICAL SYSTEMS

Dixon D. Jones

June 1975

Summary

Catastrophe theory is a new field in mathematical topology that allows the formulation of comprehensive qualitative systems models which have previously eluded rigorous mathematical formulation. Because the models have a topological foundation, many seemingly dissimilar phenomena can be related to a common underlying topological structure. The properties of that structure can then be studied in a convenient form and the conclusions related back to the original problem. This paper provides an introduction to catastrophe theory and defines the principal conditions required for its application. The basic properties of bimodality, discontinuity (catastrophe), hysteresis, and divergence are defined and illustrated using the simplest structures of the theory.

The application of catastrophe theory to ecology is illustrated with the spruce budworm system of eastern Canada. With a minimum of descriptive information about the budworm system, a qualitative catastrophe theory model is hypothesized. This model is rich in its ability to provide predictions on the global behavior of the system. To further check and refine the assumptions of this qualitative model, an existing detailed simulation model is analyzed from the perspective of catastrophe theory. The simulation indeed exhibits a basic underlying structure in agreement with the previously hypothesized model. In this instance catastrophe theory provides a consistent framework with which to analyze and interpret the results of the simulation. These interpretations are not at variance with the first rough qualitative model based only on a small set of descriptive information.

The Application of Catastrophe Theory to Ecological Systems*

Dixon D. Jones**

I. Introduction

René Thom [5,6,7] has developed an elegant theory in the mathematical field of topology that allows the formulation of system models which are rich in their ability to capture inherent structure and global, qualitative properties. He has called this the theory of catastrophes. The powerful aspect of topology compared with the traditionally used lines of mathematics is that, by using abstract principles of "topological equivalence," many seemingly dissimilar phenomena can be related to the same underlying topological structure. Properties of that structure can be studied in a mathematically convenient form and the conclusions related back to the original problem. Fortunately, although the theorem proofs and derivations are very abstract and out of reach for most non-specialists, the theorems themselves are often very clear and simply stated.

Thom's principal interest in biology concerns embryology and developmental morphology. His work is illustrated geometrically, as is natural for a topologist. However, geometry in more than three dimensions often strains our intuitive visualization. Among his many fascinating observations is that the pentagonal symmetry of an adult sea anemone is quite simply related to a geometric structure that develops naturally

---

*This paper was presented at the Systems Ecology Conference, Logan, Utah, 20-23 February 1975. It will be published in Simulation in Systems Ecology, George S. Innis, editor, in the Simulation Councils Proceedings series.

**International Institute for Applied Systems Analysis, Laxenburg, Austria, and the Institute of Resource Ecology, University of British Columbia, Vancouver, Canada.

from the bilateral symmetry of the larvae [5].

E.C. Zeeman [8,10,11] has taken the topological structures from Thom's catastrophe theory and applied them to a wide range of dynamic systems. Using these structures, he has developed models for heartbeat and nerve impulse [9]. The latter model is simpler than the Hodgkin-Huxley equations, uses fewer ad hoc equations, and is a better predictor of behavior.

Zeeman has also created a wide range of insightful models for application in the social sciences [2]. His examples cover aggression, economic growth, stock market crashes, the arms race, prison riots, and national war-making policy.

To date there has been very little effort to apply this body of theory to ecology despite the apparent need for models that are global and qualitative. Holling calls for this viewpoint in his development of the concept of ecological resilience [1]. He speaks of perturbations that "flip" a system from one equilibrium region to another. A better comprehension of how and why this happens is critical to understanding and coping with ecological systems. Our level of information and perception often constrains us to qualitative models. Topology and catastrophe theory can provide us with such models that are structurally "robust" -- that accommodate refinement of detailed knowledge. The purpose of this paper is to introduce catastrophe theory to the ecological literature.

Part II contains a non-technical description of catastrophe theory as it pertains to dynamical systems. The basic elements

and properties of the theory are introduced and illustrated with some of the simpler topological structures. In this section, I follow closely the development used by Zeeman; it would be presumptuous to try to outdo his lucid explanatory style.

In Part III a simple, but nontrivial, ecological example is formulated into a model based on catastrophe theory.

In Part IV this model is compared with a detailed simulation model of the same situation. It should be borne in mind that in this instance the simulation model was developed independently of our present purpose and had no a priori connection with catastrophe theory.

II. Elementary Catastrophe Theory

There are four basic system properties of elementary catastrophe structures. Whenever observations reveal one or more of these properties, it would be fruitful to look for others and for an underlying catastrophe topology. If such a structure can be found or hypothesized, the whole body of Thom's theory can be brought to bear.

The basic properties are:

1. Bimodality
2. Discontinuity (catastrophe)
3. Hysteresis (delayed response)
4. Divergence.

These properties refer to selected system behavior. They may not all be feasible under naturally occurring situations, but

we should be forewarned that, if conditions are perturbed, the system may move to a configuration that exhibits additional properties in the above list.

Bimodality refers to situations where observations tend to cluster around two (or more) statistical measures. For example, the weight difference between males and females of a species has strong ecological implications for intraspecific competition and niche separation. Bimodality is a static property and will not be pursued here.

Discontinuity refers to any large change in behavior associated with a small change in some other variable (including time). This discontinuous behavior, or catastrophic jumps, inspired the name for Thom's theory. "Big effects from small causes" is part of the ecological experience and may be applicable here.

Hysteresis occurs when a system has a delayed response to a changing stimulus. Thus a plot of response against stimulus will follow one path when the stimulus increases and another when it decreases. Hysteresis is best illustrated graphically in later examples.

Divergence occurs when nearby starting conditions evolve to widely separated final states. Thom is interested in the divergence of adjacent embryonic cells into separate tissues. Speciation is likely an evolutionary example. In population dynamics initial conditions just above and just below an "extinction threshold" will diverge to very different final states.

We shall return to these properties after constructing illustrations for the simplest catastrophes.

In Zeeman's work with the heartbeat and nerve impulse [9], he began with only three axioms of observed behavior. These were: (1) there exists a stable equilibrium condition, (2) there is a threshold of a stimulating factor that triggers a fast action away from equilibrium, and (3) there is a subsequent return to the original equilibrium. He further subdivided the return phase (3) into (3a), a fast "jump" return (as with the heartbeat), and (3b), a smooth return (as with the nerve).

Starting with these axioms he developed the simplest model possible that could exhibit the necessary dynamic behavior. The important distinction is that he set out to model the overall dynamics rather than (say) the physiochemistry. After developing qualitative dynamic models, he was then able to identify the necessary measurable attributes and variables to transform his models to quantitative ones.

Many applications of catastrophe theory to real situations are at the stage of metaphor or simile. This in itself can be a very useful first step because a large number of global, qualitative characteristics can be carried in a simple, easily understood format. The interested reader is encouraged to read Zeeman's work on heartbeat and nerve impulse [9] as it exhibits the full spectrum of development from the above three axioms to a quantitative, predictive model.

We shall not follow the axiomatic approach, but shall

describe the resulting conditions that are applicable. Many of the vulgar simplifications and the lack of rigor will disturb the pure mathematician, but as we are interested in modelling real, existing situations having a high degree of uncertainty and noise, we shall dismiss discussion of the razor-thin exceptional cases. The rigorous route has been followed by Thom, Zeeman, and others; the final description is consistent with their work even if it avoids some of the precision along the way.

The description "catastrophe" and the property of catastrophic jumps and fast-acting behavior point to the types of situations that are of interest. Namely, the underlying dynamic of our system must be capable of making fast changes. "Fast" is of course only relative and we begin by categorizing the variables of our system into fast variables and slow variables. The greater the separation in the speed, the clearer the distinction between types. If the speeds are more uniformly distributed, the resulting behavior will deviate from the idealized type that we describe here.

To help generalize, the collection of slow variables can be alternatively considered as parameters, external variables, driving variables, inputs, controls or causes, depending upon the context. The fast variables can then be considered as state variables, internal variables, outputs, behavior or effects. Any particular discipline will find some ambiguity in these lists. They are meant to serve only as a guide. The mathematical

ecologist perhaps will be most comfortable with the state variable/parameter combination. Decomposition into fast and slow variables has seldom been used explicitly although processes that are faster or slower than an "ecological time scale" have often been omitted to reduce complexity. Any dynamic variable can be considered as slow if understanding is increased by doing so. We may often begin by holding the slow variables fixed and studying the behavior of this restricted system. Catastrophe theory allows us to translate this constrained behavior into the behavior of the unconstrained system. The distinction between fast and slow variables must be pragmatically defined, but we will find in the example of Part IV that, even when this separation breaks down, we can still gain useful information.

We symbolically represent the collection of fast variables by x and the slow ones by p. The space of fast variables, x, is taken to be of dimension n while the slow variables, p, have dimension k. The entire system then has dimension n+k.

The major requirement for the system is the existence of some function $V(x;p)$, such that when p is held fixed, $V(x;p)$ is minimized as the system evolves. At first glance this appears to be a highly restrictive condition, especially if we are dealing with a system that is poorly understood. The interesting and important feature of this theory is that we never have to know explicitly what this function is, or what it represents. It may be interesting and instructive to look for this function, but its discovery is not a necessity.

V(x;p) can be thought of as a potential function, an energy function, an entropy function, a cost function, or a probability function. (In cases where it is natural to think of a V function that is maximized, we need only replace V by -V to obtain the required minimization.) If these interpretations are objectionable because they imply some internal "purposefulness" for the system, V(x;p) can be thought of as a Lyapunov function for the set of describing equations. There are thus two complementary perspectives available. The first is direct information about the existence of some V(x;p) function. This implies an equilibrium directed trajectory for x. On the other hand, trajectories that are known to evolve to equilibrium states imply the existence of V(x;p).

The effect of the minimization of V(x;p) is that for any fixed p, the system will move to some equilibrium x*. In general there may be more than one such equilibrium. In terms of familiar differential equations, for fixed p, the system evolves according to

$$\frac{dx}{dt} = f(x;p)$$

to some state x* where f(x*;p) = 0. In many applications f(x;p) can be equated with the negative of the gradient of V(x;p) with respect to x.

We are interested in how the various x* equilibrium points change as we move p throughout the range of its k-dimensional space. We call the set of equilibrium points (the points that

satisfy $f(x;p) = 0$) the manifold M_f. In situations of interest M_f is a "k-dimensional" surface. That is, if $k = 1$ (i.e. if we consider only one parameter or slow variable), M_f is a line traversing our $n+1$ dimensional state space. If $k = 2$, M_f is some surface. When $k \geq 2$, we have a higher dimensional "surface" which is not as easy to visualize from common experience.

We want to find the changes in system behavior when we change p, or when the slow variables evolve. To aid this search we construct the projection Π_f of the equilibrium manifold M_f onto the space of p. The projection locates the parameter values corresponding to important features of M_f.

Let us pause and review graphically what we have done thus far. In Fig. 1a is a two-dimensional space with one fast variable x (a population density, say) and one parameter p (e.g. a carrying capacity). With the conditions that we have imposed, we assume for some fixed $p = p_1$ and initial $x = x_1$ that the value of x moves "quickly" to an equilibrium point x_1^* according to some function $\dot{x} = f(x;p_1)$. For another $p = p_2$ and another $x = x_2$ the system goes to x_2^*. The collection of points connecting all equilibrium points is the manifold (line) M_f in Fig. 1b. The projection Π_f is just the p-axis.

The reason for and effect of making the fast/slow distinction in our state variables can be seen in this figure. If p is at p_1, x will be at x_1^*. Now if p moves to p_2, either by external manipulation or by its own dynamic process, x will go from x_1^* to x_2^* along the manifold M_f. In the idealized case where the

ratio of fast to slow speeds is infinite, trajectories would follow the manifold M_f exactly. When the fast/slow separation becomes blurred, trajectories will be displaced from the manifold. However, even then the trajectories will be "organized" around the manifold.

As we noted previously, for any fixed p there can be more than one equilibrium. Fig. 1c shows several trajectories to final equilibria. The complete manifold is shown in Fig. 1d. This manifold differs fundamentally from that in Fig. 1b. First, the sequence of equilibria is broken between T_1 and T_2. The broken line is part of the manifold since it satisfies the condition that $\dot{x} = f(x;p) = 0$. It represents the locus of unstable equilibria separating the upper and lower attracting surfaces. In any real system there will always be a certain amount of noise which will carry x off any such unstable equilibrium. The segment $T_1 - T_2$ then acts as a repellor for trajectories; the solid branches are attractors.

The second major feature of Fig. 1d is that the projective map Π_f is no longer one-to-one onto the p-axis. Between T_1 and T_2 three branches of M_f correspond to the same section of the p-axis. At both T_1 and T_2 the vertical mapping projection coincides tangentially with the manifold M_f. The projections of T_1 and T_2 appear at the parameter values S_1 and S_2. These points are called singularities of the projection Π_f.

The manifold in Fig. 1d schematically represents the first principal type of catastrophe -- the fold catastrophe. The

simplest (lowest degree) polynomial that is equivalent and representative of the fold catastrophe is

$$f(x;p) = -(x^3 - x + p) \quad .$$

The singularities occur whenever

$$\frac{\partial f}{\partial x} = 0 \quad .$$

In the above polynomial case

$$\frac{\partial f}{\partial x} = -3x^2 + 1 = 0 \quad .$$

Substitution gives the singularities at

$$p = \pm\frac{2}{3}\frac{1}{\sqrt{3}} \quad .$$

The fold is the simplest catastrophe. With it we can demonstrate three of the four basic properties that we presented at the start of this section. First, bimodality: this is a consequence of the double equilibria for a portion of the parameter range. Repeated observations can detect the system on the upper attractor at some times and on the lower at others.

To visualize catastrophic jumps, consult the folded manifold in Fig. 2. Initially $p = p_0$ and the system is at A. As we increase p to p_1, the system moves along the manifold to B. But when p crosses beyond the singularity S_1, the system is forced off the manifold at T_1 and makes a catastrophic, "fast" jump to the upper branch at C. Continued increase carries the

system again smoothly along the manifold to D. It is this behavior that gives the theory its name. The significant fact is that the specific form of f(x;p) is not important for this behavior, but only the existence of the fold singularity in the projection onto the slow variable axis.

Hysteresis is easily shown with this same manifold. We begin with the system at D (Fig. 3) and retrace our steps. At C we do not make a jump return to T_1 but rather continue along the manifold to T_2 above the other singularity S_2. Now at this point there is a jump return to the lower attractor and the system proceeds on to A. The property of following a different return path after a reversal of input is called hysteresis.

Why we do not jump from C down to T_1 requires an explanation. When $p = S_1$ both C and T_1 are possible equilibria and therefore both are minima of V(x;p). The value of V at T_1 could actually be less than V at C, but the transition would not be allowed as it would require a temporary increase in V when we move away from the manifold at C. In other words, the system is following the local minimum. The theory can be adapted to systems that seek a global rather than local minimization but the ecological applicability would be limited.

An example of divergence is not possible on the fold catastrophe. To include it we must use a minimum of two slow dimensions and introduce the cusp catastrophe.

Consider one fast variable x and two slow ones p and q. The simplest polynomial representation of the cusp catastrophe is

$$\dot{x} = f(x;p,q) = -(x^3 + qx + p) \quad .$$

The manifold is the surface generated by $f(x;p,q) = 0$. The two-dimensional sheet corresponding to this function is embedded in our three-dimensional state space. It is illustrated in Fig. 4.

In this example when $q > 0$, M_f is single sheeted; when $q < 0$, it is triple sheeted. Note that for a fixed negative q, we have the fold catastrophe as a special case. If p goes from p_1 to p_2, the state trajectory follows the manifold smoothly from A until it becomes tangent to the x-axis, at which point there is a catastrophic jump to the lower attractor before continuing to B. A path with fixed positive q (C to D) does not cross a singularity in the projection map Π_f, and thus avoids the fast jump.

The critical feature is the mapping of the manifold onto the space of the parameters (p,q). The outer edges of the folds project down to the curved bifurcation lines. The point where the manifold changes from triple to single sheeted (the origin in this example) is a cusp -- giving the name to this configuration. The entire situation depicted by Fig. 4 we call a cusp catastrophe. It involves the fold singularity discussed above and the cusp singularity where the bifurcation lines join.

To illustrate the fourth property, divergence, consider two nearby states E and F in Fig. 5. If the parameter q is reduced to a negative value, the two states will move steadily to points G and H, respectively. Thus, even though both paths start arbitrarily close, and both experience the same parameter

change, they end up at widely separated final states. The reason, of course, is that their paths take them on either side of the cusp, and EG ends on the upper sheet while FH is on the lower.

Besides divergence there is another important point which is characteristic of the cusp catastrophe but not of the fold. In Fig. 2, movement from A to D is accompanied by a jump at T_1. The only way to return from D to A is as shown in Fig. 3 -- that is, by return jump at T_2. This jump return is also shown on the cusp manifold in Fig. 6. However, there is now a way to obtain a smooth return by going around the cusp on the return from D to A. As a first approximation Zeeman used the fold catastrophe as in Figs. 2 and 3 to model the heartbeat, but was required to use the cusp catastrophe to produce the smooth return of the nerve impulse.

Now that two elementary catastrophes have been described, we are in a position to state Thom's theorem (in a casual but usable manner) and to comment on its fundamental importance.

Recall the conditions that we have set out. The fast variables form a vector x with dimension n. The slow variables (or parameters, or whatever) consist of a space of dimension k, the whole system being of dimension n+k. Next, the fast variables obey a dynamic flow $\dot{x} = f(x;p)$ that places the state somewhere on the manifold M_f given by

$$f(x;p) = 0 \quad .$$

Additionally there is a projection Π_f of M_f onto the space of slow variables.

Theorem (extracted freely from Zeeman)

If $k \leq 5$ and f is generic (an abstract mathematical condition expected in almost all real situations),

then

1) M_f is a manifold with dimension k,
2) The projection Π_f is stable under small perturbations of the function f,
3) Any singularity of Π_f is equivalent to one of a finite number of elementary catastrophes.

The number of elementary catastrophes for each $k \leq 5$ is

k	=	1	2	3	4	5
Number of elementary catastrophes	=	1	2	5	7	11 .

First note that n, the dimension of the fast variables, does not appear anywhere in the theorem. In the simple examples used above, we let n = 1, but we could just as well have let n = 10,000. This feature makes it possible to apply catastrophe models to embryology where there are a staggeringly large number of variables associated with the physical and chemical states in all the cells. The potential for ecology is obvious.

The second point to note is that there are only a finite (and small) number of elementary catastrophes (for $k \leq 5$). In fact, while considering our first example with $k = 1$, we found the only type -- the fold catastrophe. Likewise, when $k = 2$ we found all possible cases with the combination of the fold and the cusp. The classification of catastrophes becomes continuous rather than discrete when k exceeds five. In many practical situations, the assessment of five simultaneously changing parameters will be rich enough.

We have proposed

$$f(x;p) = -(x^3 - x + p) \qquad (k = 1)$$

and

$$f(x;p,q) = -(x^3 + qx + p) \qquad (k = 2)$$

as the simplest representations of the first two catastrophes. But in the phrase of Zeeman, these are the most complicated representations as well. That is, the generated catastrophes are the only ones for $k = 1$ and $k = 2$. By gaining an understanding of the properties of these elementary catastrophes, we will know in advance a great deal about the global properties of any situation that fits the requirements.

Primarily the models of catastrophe theory serve as hypotheses for further testing. The models also show that there can be a sound, deterministic mathematical foundation underlying some perverse phenomena that would otherwise prevent analytic investigation. Finally, for the manager of ecological systems,

catastrophe theory provides a warning that continuous changes and perturbations of a system may lead to very discontinuous outcomes. And recovery may require much more than simply restoring the system to the conditions that prevailed prior to the change.

Most discussion to this point has talked about dynamic systems in general and very little specifically about ecology. In the next section we consider the classic situation of the spruce budworm of eastern Canada. It clearly exhibits catastrophic jumps in abundance. A very qualitative model is proposed using the fold and cusp catastrophes. This model carries with it all the properties that are inherent in that formulation. It suggests some qualitative features that should exist even though they have not yet been recognized. Finally, as this case is of serious economic and managerial importance, we can call on the qualitative features of the cusp catastrophe for preliminary policy.

In Section IV, as a test of our qualitative model, we shall examine a detailed budworm system simulation model constructed previously for quite different purposes. If the simulation has a comparable catastrophe structure, we will be in partial fulfillment of demonstrating a hard example of an ecological system fitting the characteristics of the theory.

III. A Qualitative Ecological Model

The spruce budworm of eastern North America serves as a specific example. Details of this intensively and extensively studied insect can be found in Morris [3]. We present only a very brief summary of the budworm system here, but from a small amount of information, we can propose a richly predictive model of the global dynamic structure of this ecosystem.

In the maritime provinces of eastern Canada the budworm strongly favors balsam fir as a host. Balsam has a large geographic distribution and in many regions well over half the land area is forested with this species. As a first approximation, we consider this as a two-species herbivore/plant system.

The budworm is characterized by long periods (40-70 years) when it is extremely rare, but at a seemingly stable density level. Following this period of low endemic population, the budworm enters an outbreak phase where in three to four years its density increases by upwards of five orders of magnitude. At this population density all of the newly produced foliage, and some of the older, is completely consumed over vast areas of forest. After four or five years of such heavy defoliation, tree mortality becomes nearly complete. The understory is primarily young balsam fir that are effectively immune to budworm attack. They are "released" to grow by the removal of the older parental overstory. Meanwhile, the budworm is faced with an increasing threat of predators, parasites and disease as well as a very diminished food and oviposition resource. The population quite rapidly returns to the endemic state -- the whole outbreak cycle

lasting only seven to fourteen years.

A great deal was known about this cycle in anecdotal form long before any scientific investigations were begun. We hypothesize that the outbreaks (and perhaps the declines) are catastrophe jumps and fit the budworm/forest system to the catastrophe framework. Even at this level of detail some interesting conclusions can be drawn.

Our fast variable is the density of defoliating budworm larvae, NL. We begin with only one slow variable which we call F. This is a qualitative measure of the available foliage in the forest; the exact interpretation is not important at this stage. We recognize that the intrinsic growth rate of NL can be much faster than that of F. We are in a case with $k = 1$, and therefore use the fold catastrophe as our model (Fig. 7).

We begin with the system at point A: The forest is young and the budworm are at the endemic level NL_0. As the forest matures, we move to point B, still with $NL \simeq NL_0$. At T_1 the budworm are forced off the lower equilibrium level and rise quickly to the upper attracting line. The path is not exactly vertical because the forest continues to grow during the two or so years required to reach this upper level. With this time lag imposed by a finite (yearly) generation time, we might expect some over-shoot.

Almost no mention has been made of the dynamic processes that govern the movement of the slow variables on the fast manifold M_f. The functions that describe this slow flow are under no restrictions other than relative speed. However, it is the

specifics of the slow flow that determine the system response -- it is the slow flow that carries the system over or around the catastrophe singularities.

Typically the preponderance of an investigation is centered on the dynamics of the fast variables while only minor consideration is given to the dynamics of the slow ones. This is certainly the case with the spruce budworm. Research on the population dynamics of the insect far outweighs that done on forest response.

Returning to Fig. 7, we propose a reasonable feedback relationship of the budworm on the forest. At high levels of NL, defoliation by the insect decreases F until T_2 is reached, whereupon there is a jump return to the endemic level at C. The accumulated stress on the trees continues to decrease F as the affected trees die. At A the cycle begins anew.

A simple, descriptive feedback dynamic for the forest variable could be

$$\frac{dF}{dt} = r \cdot F \cdot (1 - F/K) - m \cdot NL \cdot F \quad .$$

When NL is small (NL_0), the forest grows toward F = K according to the logistic growth curve. The term $m \cdot NL \cdot F$ is the "mortality" of F due to budworm consumption. When NL is on the upper attractor, we have

$$m \cdot NL > r$$

and F begins an exponential decline. To incorporate the delayed

effect of accumulated stress, we could replace NL by NL(t) + NL(t - τ) to give

$$\frac{dF}{dt} = r \cdot F \cdot (1 - F/K) - m \cdot F \cdot (NL(t) + NL(t - \tau)) \quad .$$

This equation is meant to be only descriptive. Its exact form is not important here. What is required is that F respond to NL so that the flow moves to the right on the lower branch and to the left on the upper. The overriding feature of importance is that M_f is shaped into a fold catastrophe as we have hypothesized.

The folded M_f carries with it some important implications for this system. (Refer to Fig. 8.) First, if the foliage, F, can grow toward an upper asymptote, K, that lies beyond T_1, an outbreak is inevitable. If, however, this level is reduced to K', below T_1 (by tree thinning or logging, for instance), the system is held on the lower attractor at endemic population levels.

With K at K', we are at the stable equilibrium point A'. But even then an addition of a (perhaps small) number of in-migrating budworm is enough to move the population from A' to B. An outbreak is triggered.

If during an outbreak (point C) insecticide control is used (as is the present policy in Canada), the system will move to some new point D. This point will be to the left of C because of the decrease in F due to current and past defoliation. Unless D is below the T_1 - T_2 branch, the system will return again to

the upper attractor. The same displacement at C' will carry NL across the T_1 - T_2 boundary. A collapse will occur even though D' is still to the right of T_1. Note that the earlier in the outbreak, the larger must be the downward displacement to effect control.

Through continual insecticide application, we may be able to hold the system at some point E where intrinsic growth of F is just balanced by defoliation losses due to NL. The consequences of any relaxation of control are obvious. In fact, this situation appears to be what is at present occurring in many parts of eastern Canada.

So far we have used only one slow variable which necessarily requires a jump return to the lower level. As a transition to the next section, we develop an alternate model using the cusp catastrophe. Since the fold is a special case of the cusp, we retain the fast return in our repertoire and add the possibility of a smooth return. We also add the possibility of eliminating the catastrophic jumps altogether.

In the above discussion the variable F was only vaguely defined. We now split this slow variable into its two primary components. The first is the amount of habitat available to the budworm per unit of land area. The normal field measure is in units of branch surface area, SA. There is a nearly monotonic increasing relationship between SA and average tree age, so we have SA increasing with time if we begin with a young forest.

The other slow component is the total amount of foliage

available on each unit of branch area, FT. As a first approximation we assume this to be independent of tree age and largely subject to decrease by defoliation. SA is then a measure of available real estate, and FT is a measure of available food resource.

Fig. 9a shows a possible configuration of a cusp manifold with k = 2. The surface has been rotated 180° and distorted slightly, but it still maintains the features of the cusp.

A typical path begins at A with full foliage and a low branch area. As time passes, SA increases until the trajectory intersects the fold curve at T_1. This corresponds to the point where the trajectory passes out of the cusp region in the SA-FT plane. The system is quickly drawn to the upper attracting sheet. Again we allow for some over-shoot. The path from here has two distinctly different possibilities. In (a), FT is reduced and the resulting food shortage quickly, but smoothly, returns the budworm to endemic levels. The dynamic flow on the manifold is such that the reduced foliage causes increased tree mortality and a lowering of average tree age and branch area. Foliage then recovers as young trees begin to grow into the population.

Path (b) differs from (a) in that tree mortality begins at a lower level of defoliation (higher FT). The flow on the manifold bends sooner than in (a) and is carried back over the cusp, giving a jump return to endemic levels.

We emphasize again that it is the slow flow equations that separate path (a) from (b) and determine when and where the

catastrophic jumps will occur.

In Figs. 9b and c are two alternate configurations of the manifold. In 9b there are two cusps in the (SA, FT) plane. In this case the return to endemic is a jump except for a very narrow path between the cusp points. In 9c the cusp lies outside the region of possible values of SA and FT and thus we have returned to the fold catastrophe.

It is not possible to distinguish between these various manifold configurations simply from the description in the opening summary. These are alternates from which to launch further investigations. In the next section we shall examine a surrogate for the real situation using a previously constructed simulation model of the budworm/forest system. That model was not created with any mind to catastrophe theory. It was part of a program with quite different goals. The simulation was designed to mirror explicitly the functional relationships of the underlying population processes. In other words, the simulation models the biology while catastrophe theory models the dynamics.

As a model, its mathematical structure can be precisely examined. We shall see that this model contains the fundamental features required by catastrophe theory. We will be able to locate the catastrophe manifold M_f and the singularities of the projection Π_f. Even in those places where the fit with theory is not close, the global, structural viewpoint of catastrophe theory will lead to some conclusions and understanding that were not appreciated previously.

IV. The Catastrophes of a Budworm Model

In this section we shall examine the structure of a detailed budworm simulation model to see whether we can identify any of the general catastrophe structure that we have hypothesized. There are several good reasons for taking this approach. First, the simulation incorporates biologically realistic functional relationships that accurately portray the qualitative aspects of the real situation. It is based on an intensive set of data covering more than 25 years of study as well as the collective experience and judgment of many of the principal investigators. There is considerable reason to believe that this is a reasonable approximation to reality.

Second, the simulation was constructed with no intent to "map" it onto catastrophe theory. If we find a correspondence with that theory when there should be none, it will be coincidence and not an unconscious bias of the model.

Third, the simulation is a concrete set of mathematical equations and as such will yield a precise set of characteristics without the distractions of random noise and statistical uncertainty.

Fourth, and not the least compelling, is the following: we have made various claims that when a system displays one or more of the basic catastrophe properties, we have reason to suspect that it fits into the catastrophe theory framework. Once built, the budworm simulation, as a dynamic system itself, exhibits some of these properties independent of the ecological

process it was meant to imitate. It will add credence to those claims if we can show that the simulation, as a dynamic mathematical entity, has a catastrophe structure.

The simulation was the core of a program to develop and test a range of integrated techniques and methodologies for resource management and policy analysis. An outline of that program is in [4]; a detailed monograph is in preparation.

We shall not trace through the inner workings of that model here, but some comment should be made on the state variables used. The budworm generations do not overlap, so the density of any one of its life stages will serve as a single variable for the insect. The density of large larvae (instar III) is chosen for convenience and designated NL. Density is scaled to that used in field measurements -- number of individuals per ten square feet of foliage surface area.

The age structure of the trees is contained in twenty-five three-year age groups. Group 25 also holds all ages greater than 75 years. The contribution of branch surface area by each group is summed to give the total surface area SA. This quantity is then scaled between 0 and 1 and shown as SAR in the following figures. Although the time course of SAR will depend on the particular age distribution of trees, we shall consider it as a proper slow state variable in this discussion. SAR, and not the age distribution itself, affects budworm survival.

Balsam fir retains its needles for an average of eight years. In the simulation, foliage is aggregated into new growth and old

(> 1 year), and averaged for all tree age classes. In the following discussion the variable used is the total foliage available (per unit of surface area) in the spring prior to defoliation. We scale this also between 0 and 1 and call it FTS.

A stochastic yearly index of weather was an additional element in the simulation. Weighted measures of accumulated heat units and precipitation were combined into a three-level index: "poor" (w = 1), "average" (w = 2), and "good" (w = 3) weather. A statistical model was constructed to provide a synthetic weather trace comparable to an historical one. We will not consider this element at the moment but will return to it after looking at the model with the weather held constant at its average value (w = 2).

In the following figures NL (the density of larvae) is the fast variable and SAR (surface area) and FTS (foliage) are the slow ones. During the endemic phase the forest is young (SAR < 0.1) and the trees have full foliage (FTS = 1.0). Because the budworm is so rare at this time, very little is known about the population controls in effect at this density. As a pragmatic move, an absolute floor of NL = 10^{-5} was built into the model to represent the endemic density.

The equilibrium manifold M_f was found by holding SAR and FTS fixed and searching for NL values that did not increase or decline through one iteration of the model. A cross-section of the budworm manifold for FTS = 1.0 is shown in Fig. 10. The

"+" traces the attractor surface and the "x" the repellor. This is easily recognized as equivalent to the fold catastrophe. The lower attractor is at NL = 10^{-5}, though it appears as zero on an arithmetic scale. The usual fold points are marked at

$$T_1 \quad : \quad SAR = 0.260$$

and

$$T_2 \quad : \quad SAR = 0.175 \quad .$$

A series of FTS = constant cross-sections was made and assembled into the perspective plot in Fig. 11. Note that the upper-back curved line is the same as the cross-section in Fig. 10. The flat area in the lower left of the FTS - SAR plane is the endemic level NL = 10^{-5}. The upper fold curve (dashed line) goes from T_2 (as before) to the cusp point at

$$FTS = 0.60 \quad ,$$

$$SAR = 0.35 \quad .$$

The shaded area is the projection Π_f of the fold onto the plane of the slow variables.

Fig. 11 clearly shows that for average weather conditions, the budworm manifold is formed into a cusp catastrophe. Given that such is the case, we would expect the system to progress in a manner similar to Fig. 9a. We now follow the course of a typical cycle.

In the absence of budworm the forest will progress to a steady-state age distribution with a surface area value of SAR = 0.65. As this is beyond the fold point T_1 of Fig. 10, at least one outbreak is inevitable. Continued periodic outbreaks require the slow flow to carry the system off the upper surface and onto the lower. We again call attention to the importance of the dynamics of the slow flow.

In Fig. 12 we repeat the fold curve at FTS = 1.0 and project the first few cycles onto the plane of FTS = 1.0. The first cycle begins at point A and moves horizontally to the right. The fast-rising section is still in the plane of FTS = 1.0, but the diagonal decline moves up out of the figure to a point near FTS = 0.3.

Although the vertical rise is rapid, it does not occur at T_1. The reason for this discrepancy will point us to a special consideration for ecological systems. Fig. 13 is exactly like 12 but with NL on a log-scale. The curve does indeed start its ascent at T_1 but, because the fast dynamic is proportional to the population, we find that it is slow when NL is low. We should bear this in mind whenever the fast variable is a species population.

To complete the picture, Fig. 14 shows the same cycles projected onto the plane of SAR = 1.0. When log(NL) is used, the manifold appears as shown -- though not quite a fold, it is very steep. Fig. 15 is a view of the trajectories on the FTS - SAR plane showing the portrait of the slow dynamic.

We find that, even though NL is not particularly "fast" at low levels, the general conclusions of catastrophe theory are still upheld. A second discrepancy with our idealized requirements is the speed with which FTS drops during defoliation. However, this does not eliminate the catastrophe framework as a useful model.

Consider the following: foliage has a stable equilibrium at FTS = 1.0. Periodically it drops quickly to a low level and then slowly recovers to 1.0 again. Can we find an elementary catastrophe in this situation? The answer is yes.

If we temporarily think of FTS as the fast variable and SAR and NL as the slow ones, it is possible to construct the corresponding manifold. A cross-section at SAR = 0.8 is shown in Fig. 16. As NL increases, defoliation carries FTS past T_1 where, due to reduced photosynthetic ability, the foliage is unable to replace its losses, and it drops to the "lower" attracting surface. This surface is not at zero because new foliage is being added as young trees grow into the tree population. Once the budworm density drops below T_2, the foliage recovers again. Thus, a seeming fault in our formulation has led to additional, but consistent, understanding about our system.

The conclusions that we reach from this look at the catastrophe structure of the simulation model are not qualitatively different from those discussed in relation to Fig. 8, which was based only on limited descriptive information. This says a great deal for the potential of the catastrophe viewpoint.

All ecological systems are continually subjected to random perturbations and noise. The global perspective we have presented here makes it possible to qualitatively predict the outcome of shifts in state location. One need only know where the new point is in relation to the fast mainfold. But what of perturbations that change the structure of the manifold itself? The budworm simulation gives us a first look at this problem.

As was mentioned above, the simulation contains an aggregated stochastic weather index. It enters the model dynamics by changing the rate of survival from the larvae to the pupae stage. Relative to the average (w = 2), weather class 1 lowers survival while class 3 promotes it.

First, we look at the manifold for these other two conditions. The case when w = 1 is shown in Fig. 17. Note that the fold is deeper than before and that the cusp has moved outside the region of feasible (SAR, FTS) values. The upper fold curve is the dashed line ending at T_2. The shaded region is the visible portion of the fold projection. Recall that in the absence of the budworm, the flow on the lower attractor will bring the state to FTS = 1.0 and SAR = 0.65. This point is at the intersection of the two heavy arrows. (The actual point is hidden by the fold in this projection.)

If the weather were constantly "bad", it would be possible to place the system at that point and it would remain there, and an outbreak would no longer be inevitable. However, if the budworm density were ever to fluctuate above NL = 10, the repelling

surface would be crossed and an outbreak would occur. Tests with the simulation at constant w = 1 confirm these observations and further show that the slow flow is unable to sustain a sequence of outbreaks.

In Fig. 18 is the manifold when w = 3. It has lost its catastrophe configuration -- and with it the ability to exhibit outbreaks! The system oscillates rapidly with a period of six-seven years. The location of the activity is shown on the FTS - SAR plane in Fig. 19.

To observe what the budworm is doing, we look at the cross-section A - A (Fig. 20). The system is hanging very near the cliff edge, being continually drawn under the manifold by the increase in SAR.

The section of Fig. 19 through B - B is shown in Fig. 21. The fluctuations along the FTS dimension are much larger. We can see from this that, even though we lose the fold in the budworm manifold, the system still operates back and forth on the foliage catastrophe shown in Fig. 16.

In total, the system operates by three different modes. When w = 3, the final state is an oscillation around the FTS fold catastrophe. When w = 2, the budworm cusp catastrophe is added and an approximate limit cycle results. When w = 1, the cusp becomes a fold which is located so as to allow a stationary final state.

What happens when these three forms intermix is not clear a priori. If budworm densities were to rise quickly from endemic

levels whenever a lower fold singularity (T_1) was crossed, we would expect good weather to have a substantial triggering effect. However, because the rise begins so slowly, weather triggering is much diminished. Indeed, simulations with historically realistic weather index sequences produce trajectories that differ only slightly from those with weather held constant at 2.

This last conclusion is at odds with the accepted evidence [3]. Tree ring data have identified a strong association between sequences of "good" weather and outbreaks. The simulation shows only a very weak association. I suggest that intersite migration is the key to resolving this issue.

The initial rise is slow because population growth is proportional to population. However, additions to the population through migration will not in general be highly correlated with the density on site. These population increments are in absolute terms rather than relative ones and will have a great effect in carrying the population up into the faster-moving regions. We therefore expect that fluctuating weather and intersite migration will mix synergistically and affect the outbreak dynamics profoundly. Simulation experiments with a constant level of in-migration confirm this conclusion.

V. Conclusion

The application of catastrophe theory to ecological situations will have to develop as an art before it becomes a science. Experience in both success and failure will help in that transition. As a technique or "style" for analyzing an existing model, it has been very successful. It provides a rigorously based viewpoint from which an organized set of conclusions can be drawn. We have only begun to probe its significance for ecological management. With it, a sound model can often be proposed where none was thought possible. Being global, it addresses issues where new events may take a system to unexplored ground. Being qualitative, it provides guidance as we live and operate in a world of whose nature we are grossly ignorant.

Holling's development of ecological resilience [1] sought generalizations from observed cases. Catastrophe theory provides a model upon which some of those cases can be framed. It satisfies the call for an orientation that is qualitative, structural, and global, and it provides a starting point for generating hypotheses when answers are needed and information scarce.

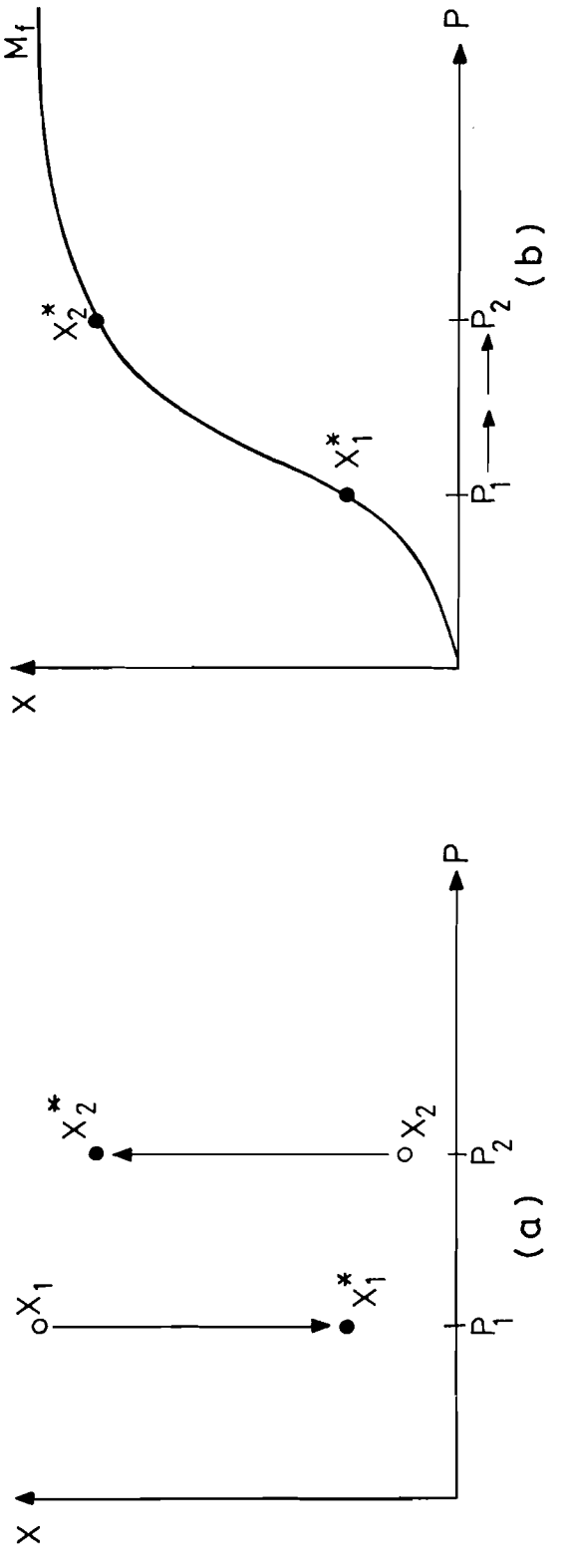

FIGURE 1

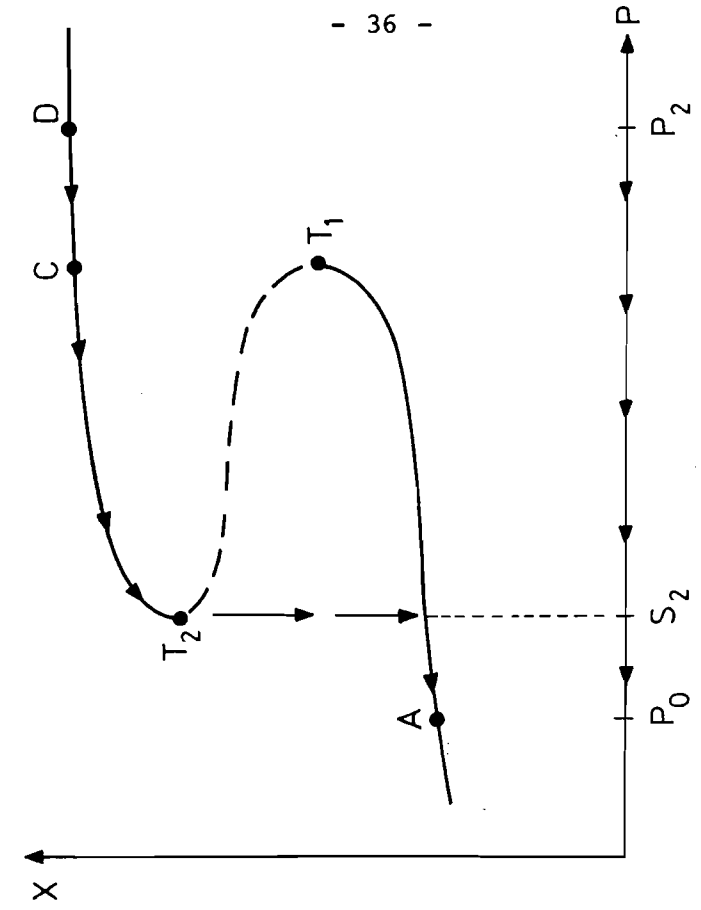

FIGURE 3

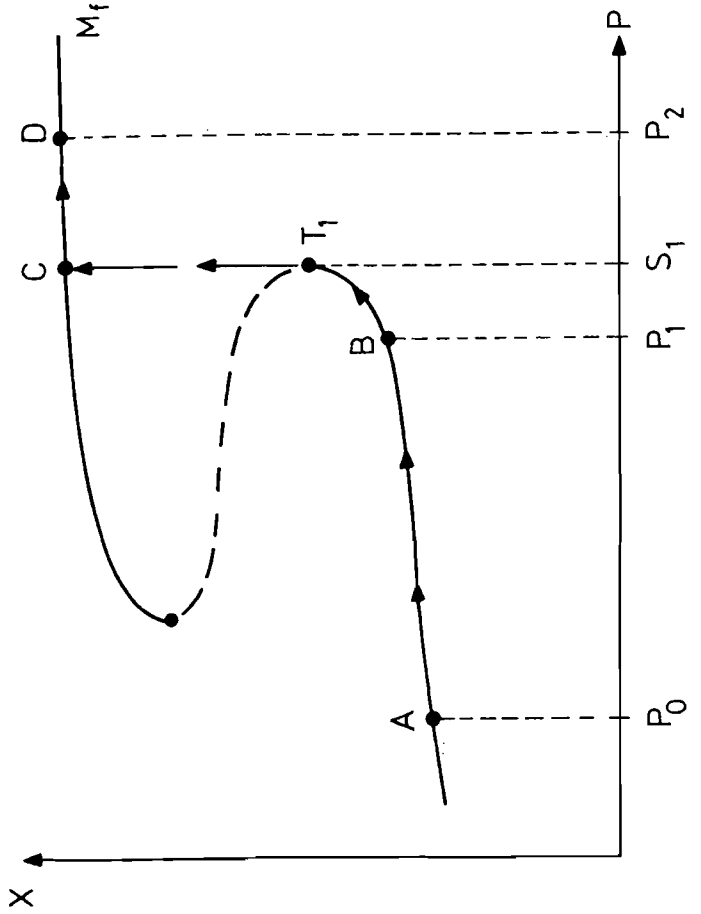

FIGURE 2

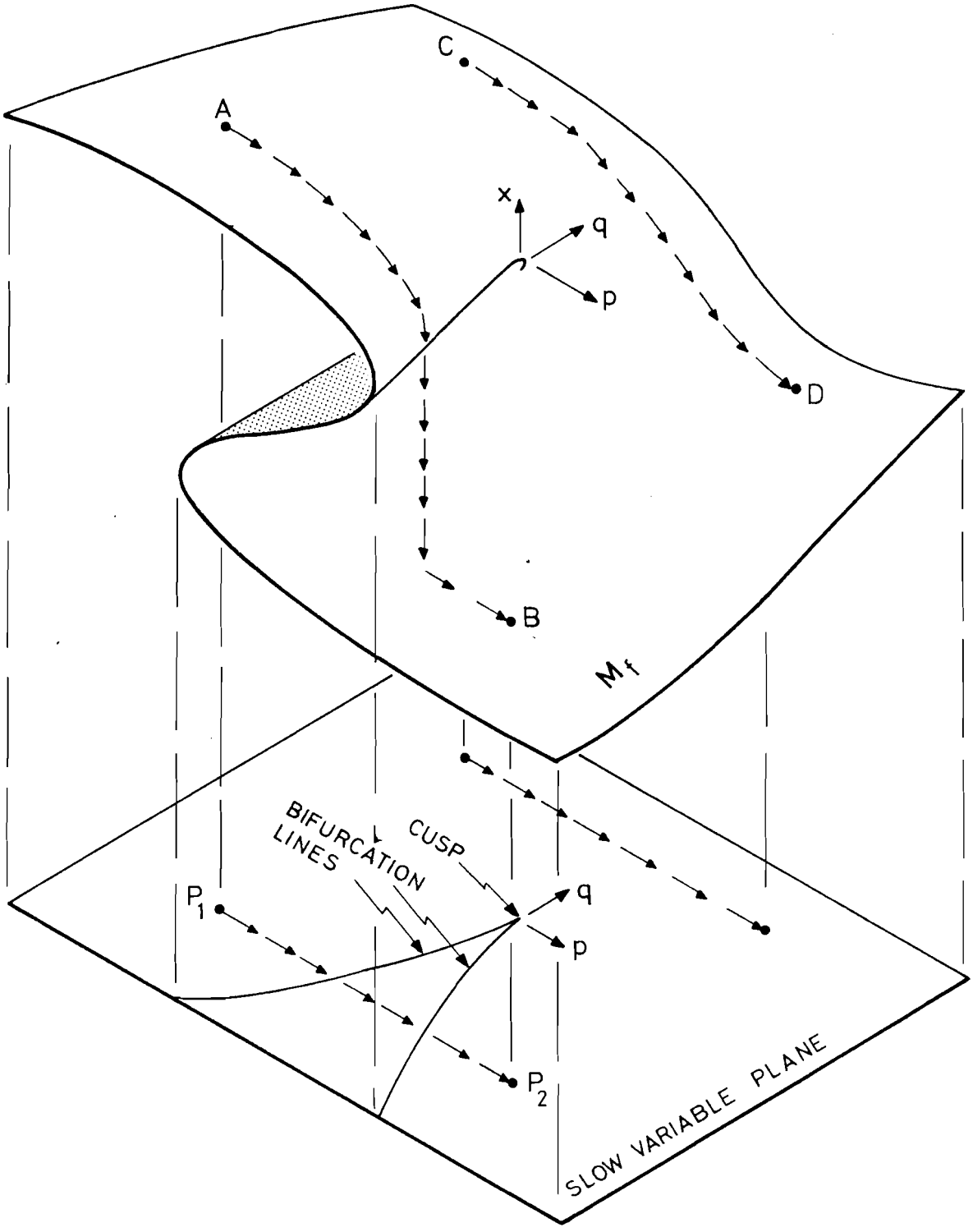
C
A
x
q
p
D
B
M_f
BIFURCATION LINES
CUSP
P_1
q
p
P_2
SLOW VARIABLE PLANE

FIGURE 4

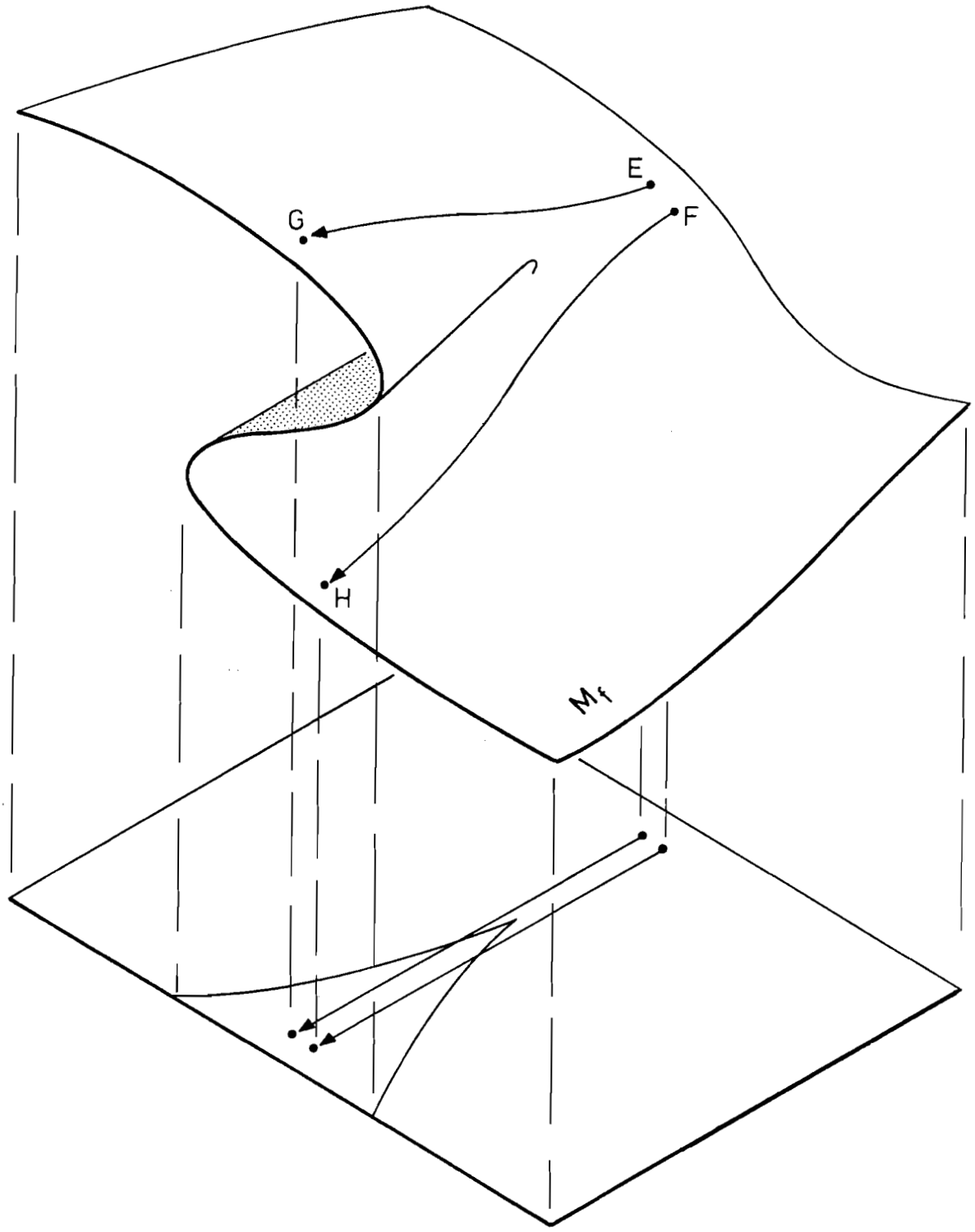

FIGURE 5

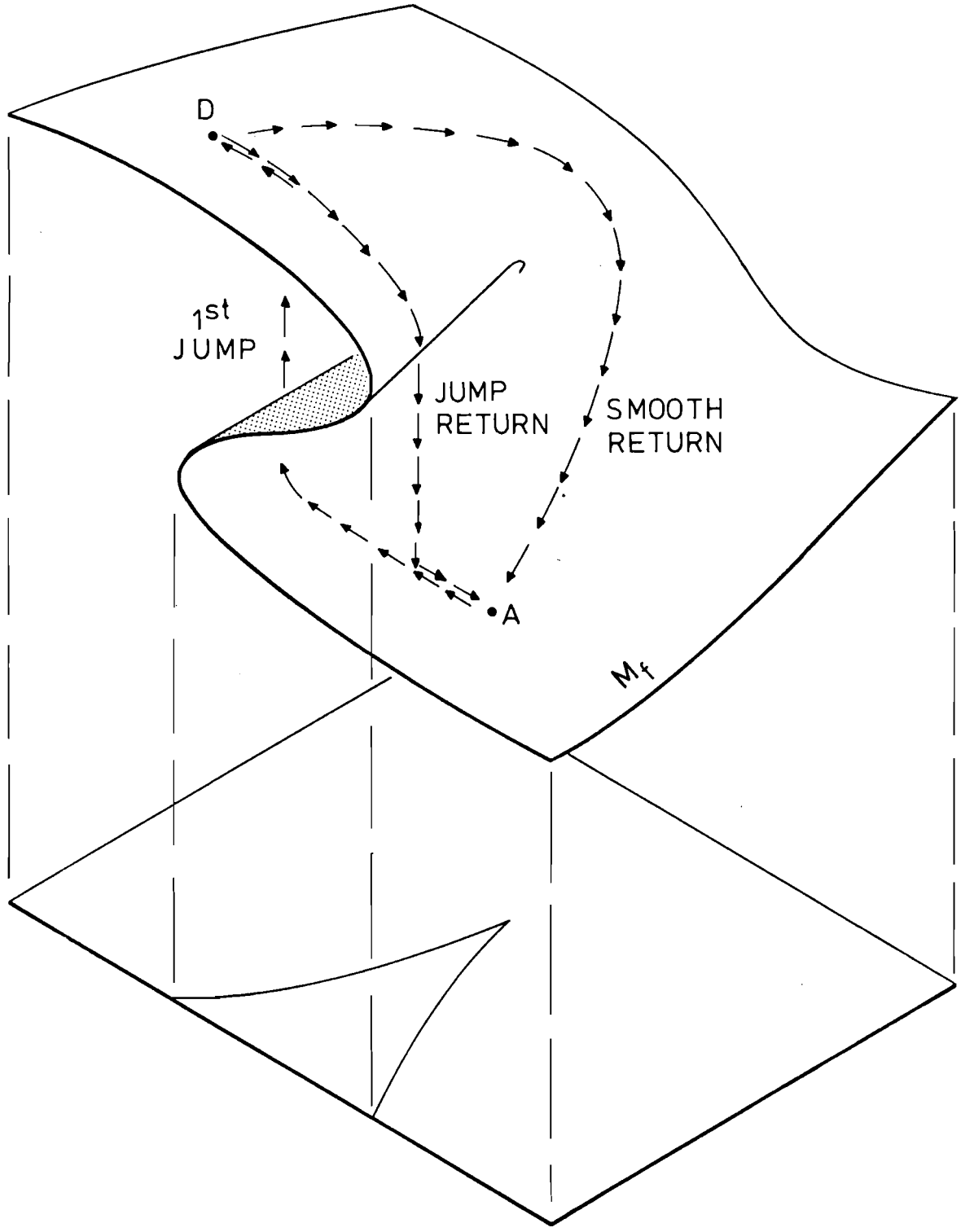

FIGURE 6

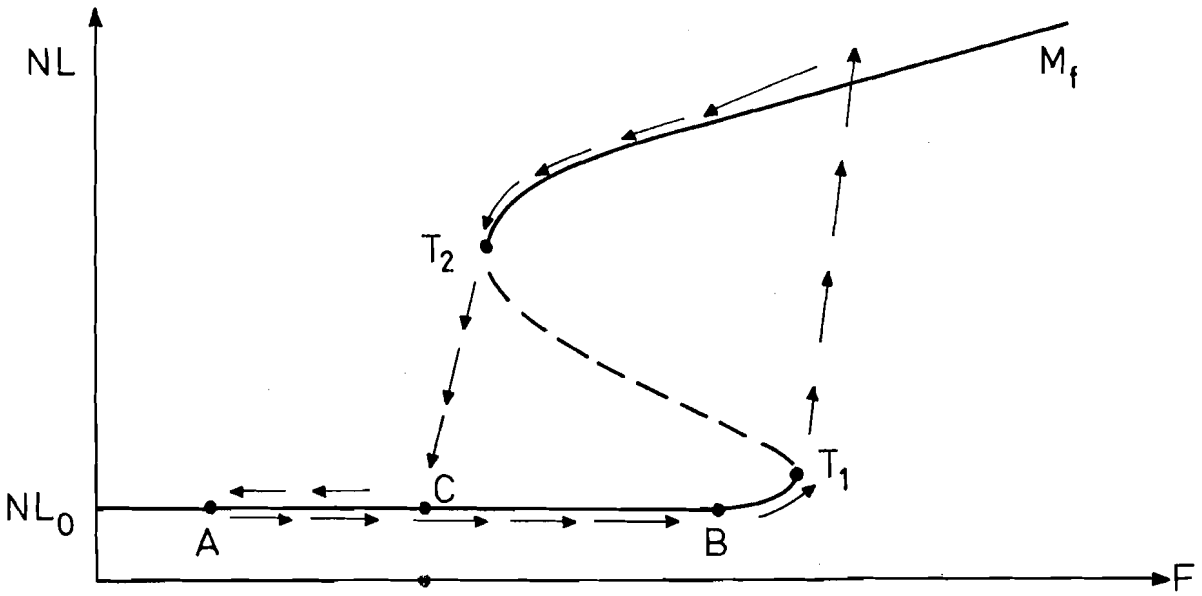

FIGURE 7

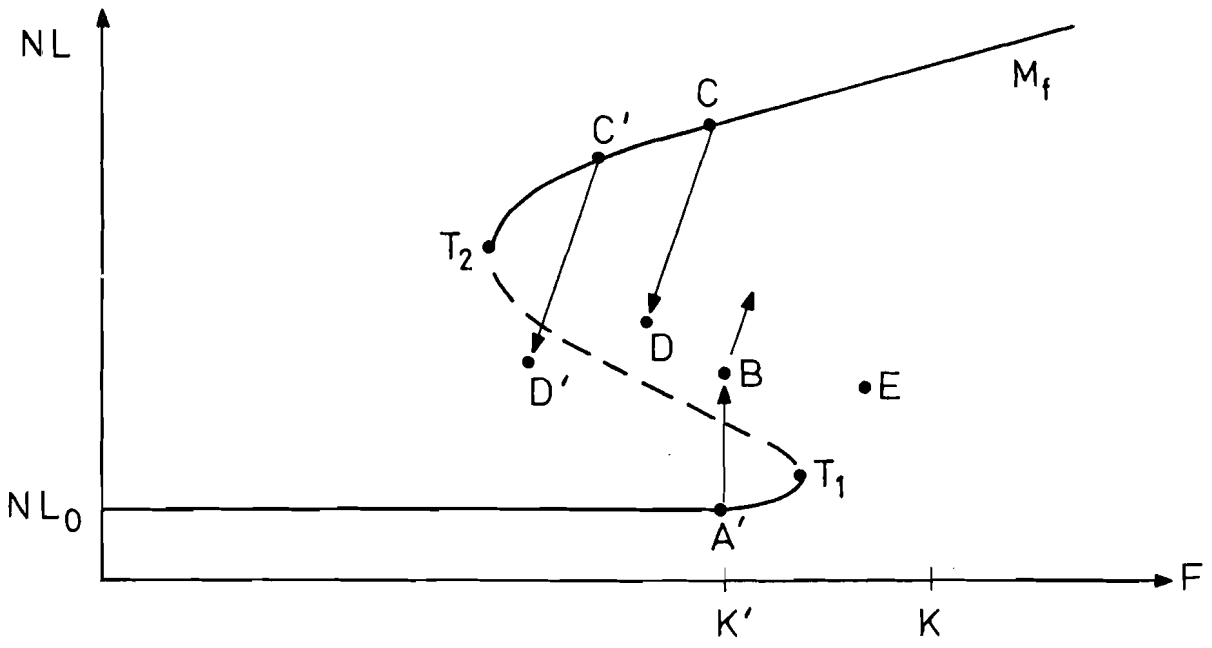

FIGURE 8

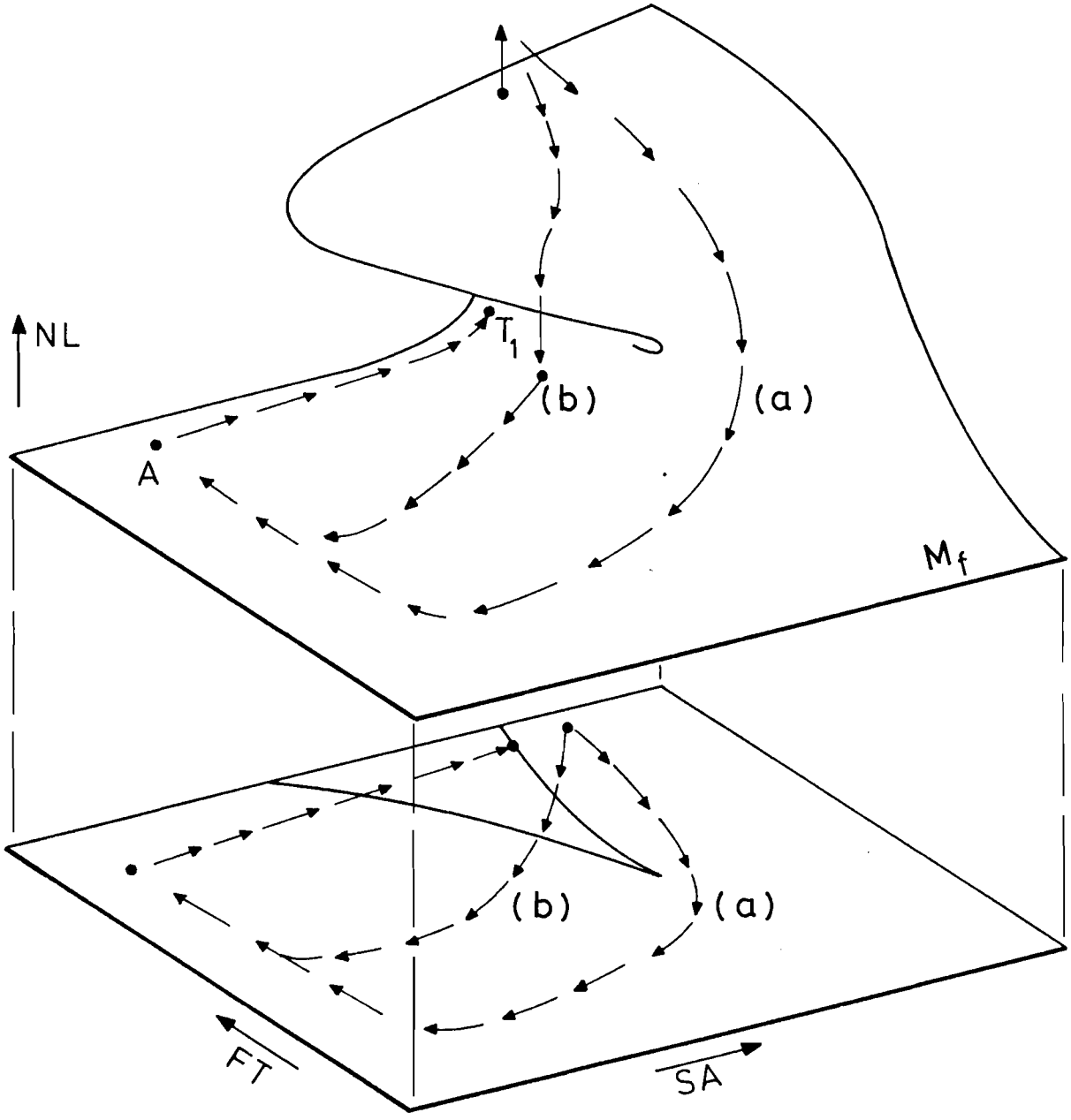

FIGURE 9a

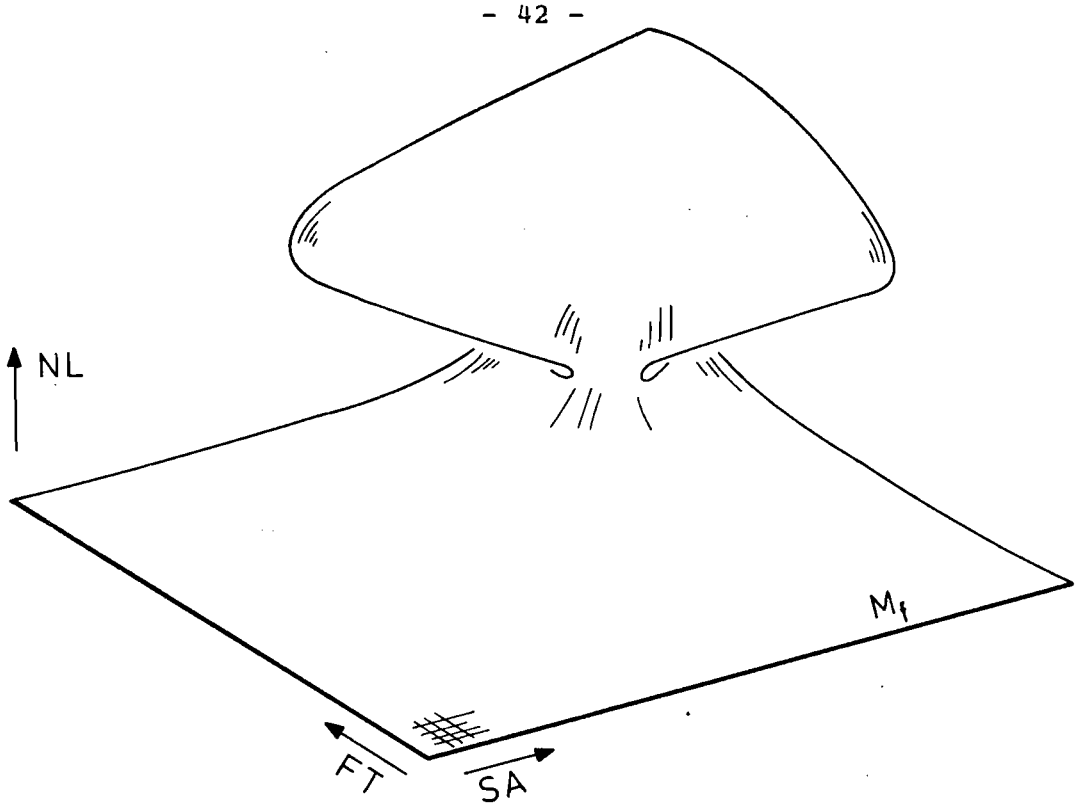

FIGURE 9b

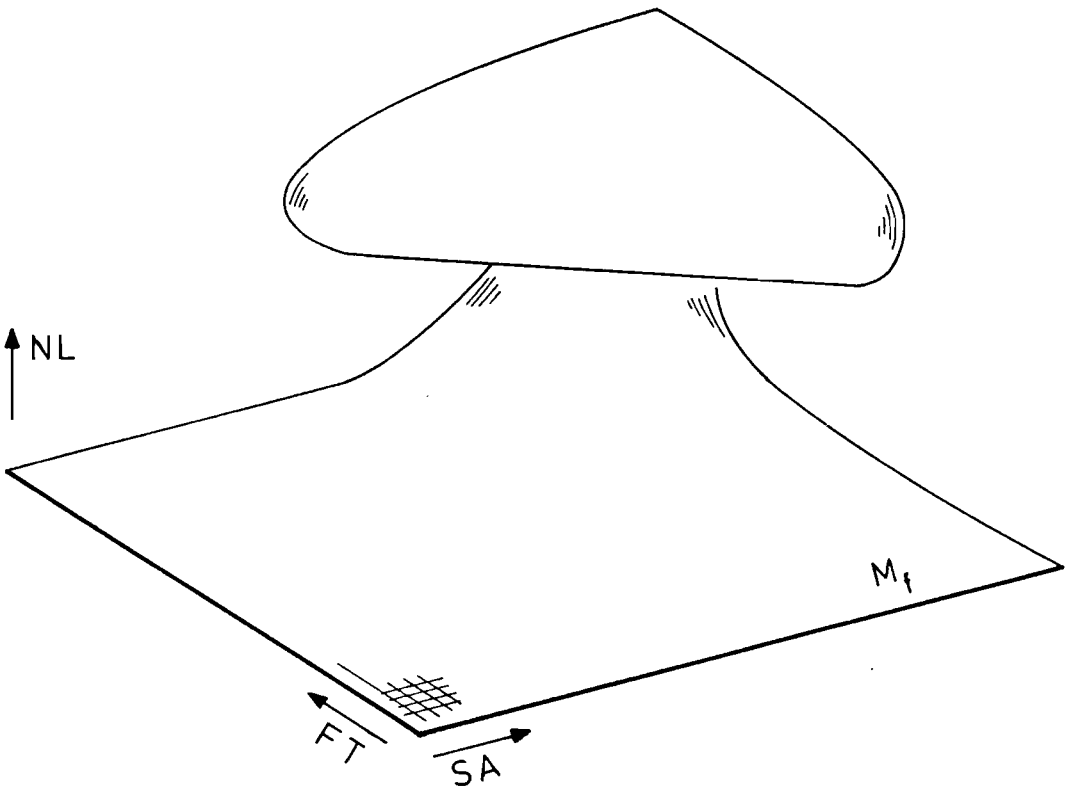

FIGURE 9c

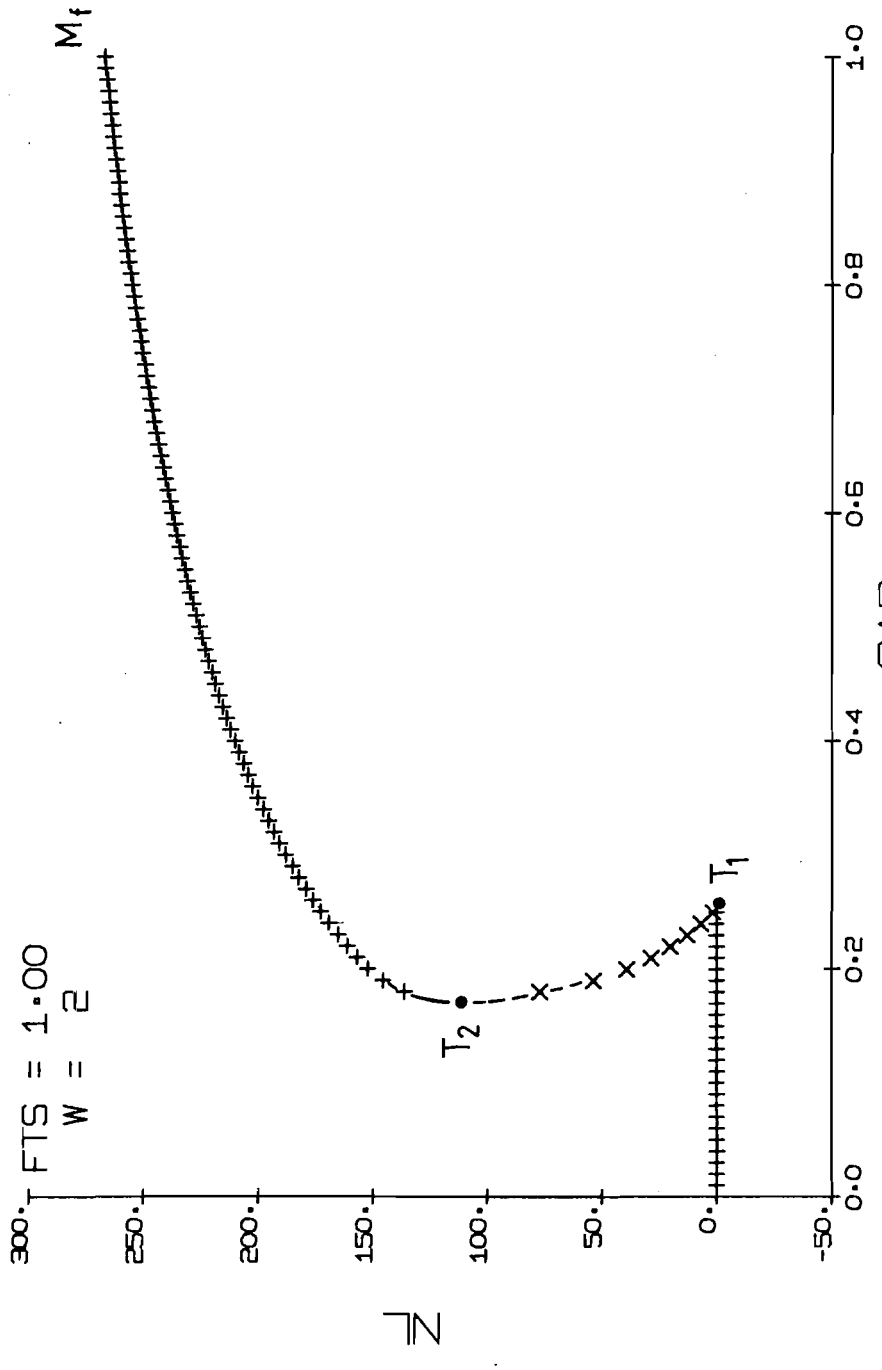
FTS = 1.00
W = 2
M_f
T_2
T_1
NL
SAR
300.
250.
200.
150.
100.
50.
0.
-50.
0.0
0.2
0.4
0.6
0.8
1.0

FIGURE 10

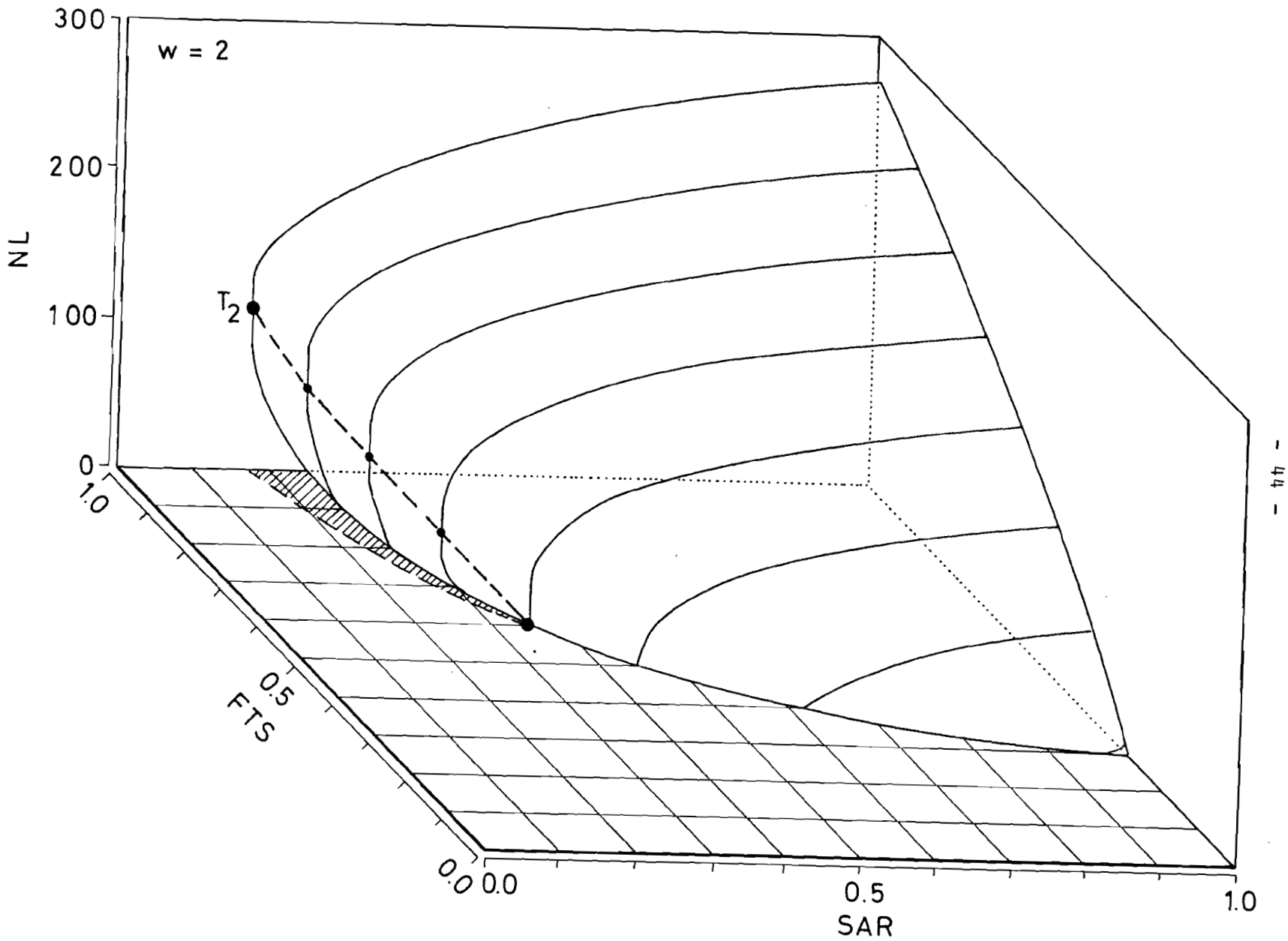

FIGURE 11

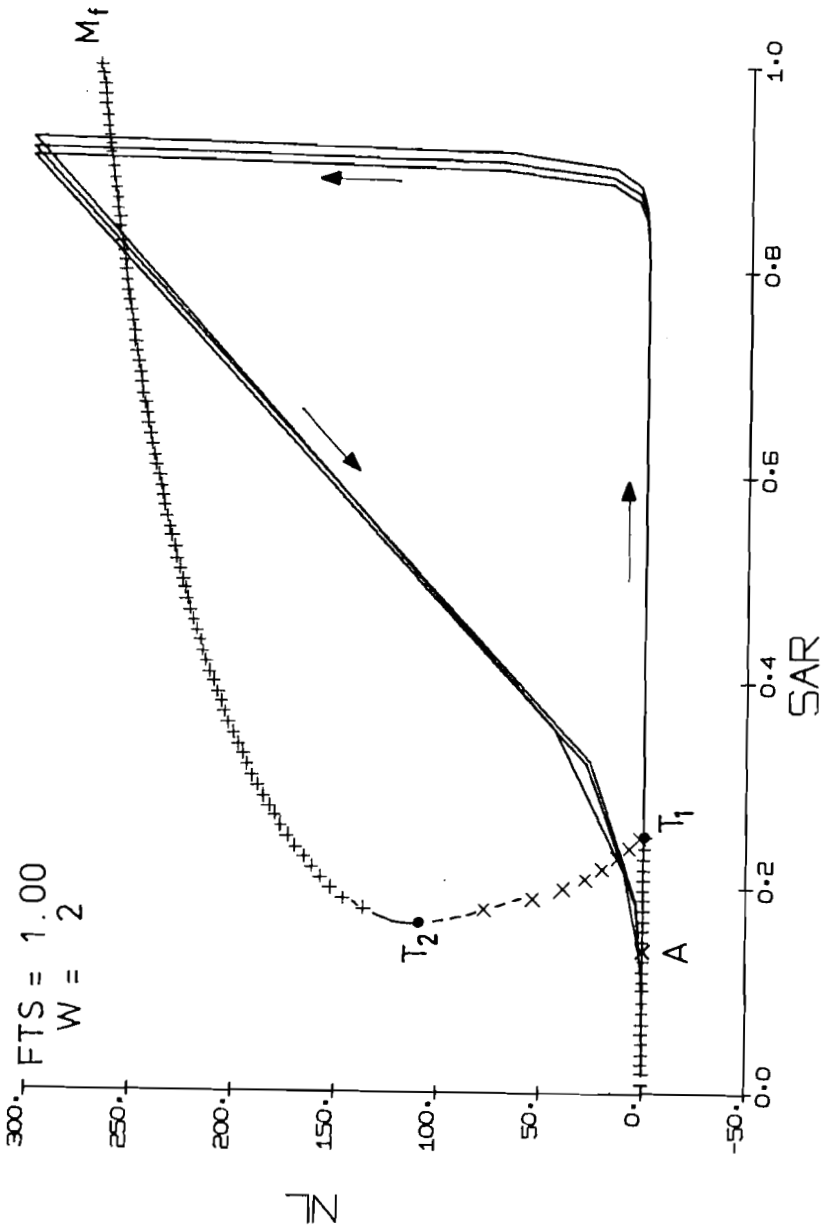
FTS = 1.00
W = 2
M_f
T_1
T_2
A
SAR
N
300.
250.
200.
150.
100.
50.
0.
-50.
0.0
0.2
0.4
0.6
0.8
1.0

FIGURE 12

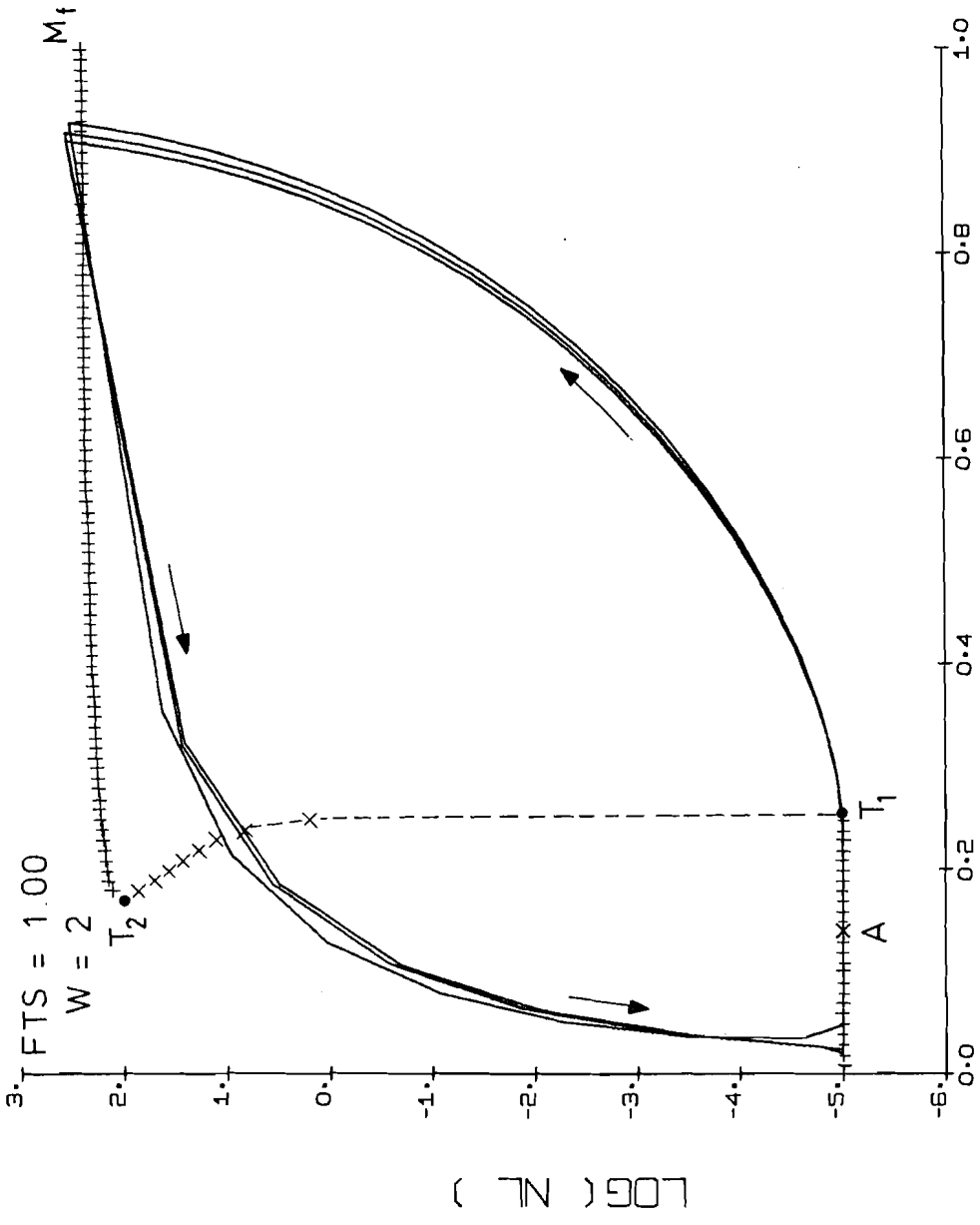

FIGURE 13

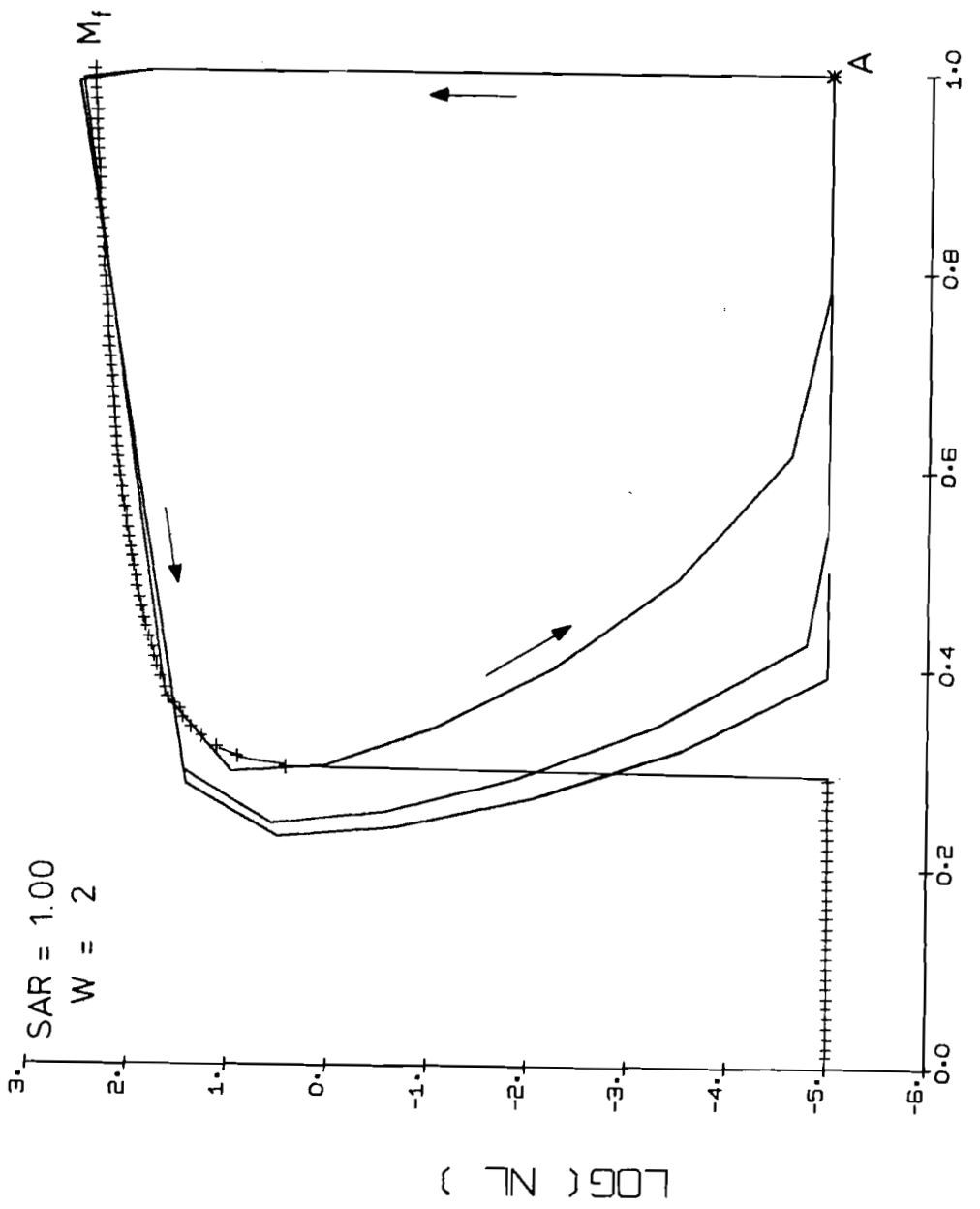

FIGURE 14

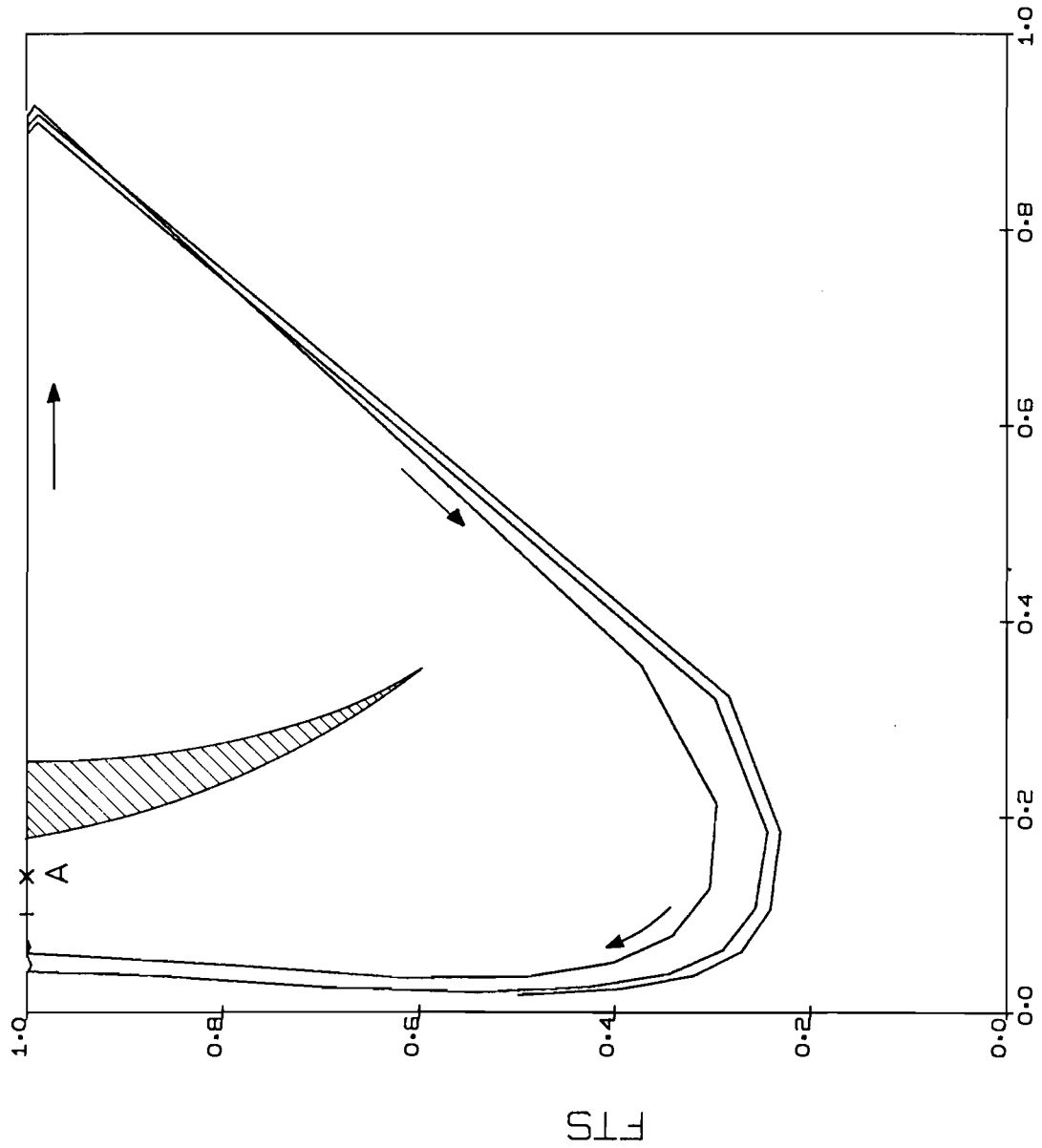

FIGURE 15

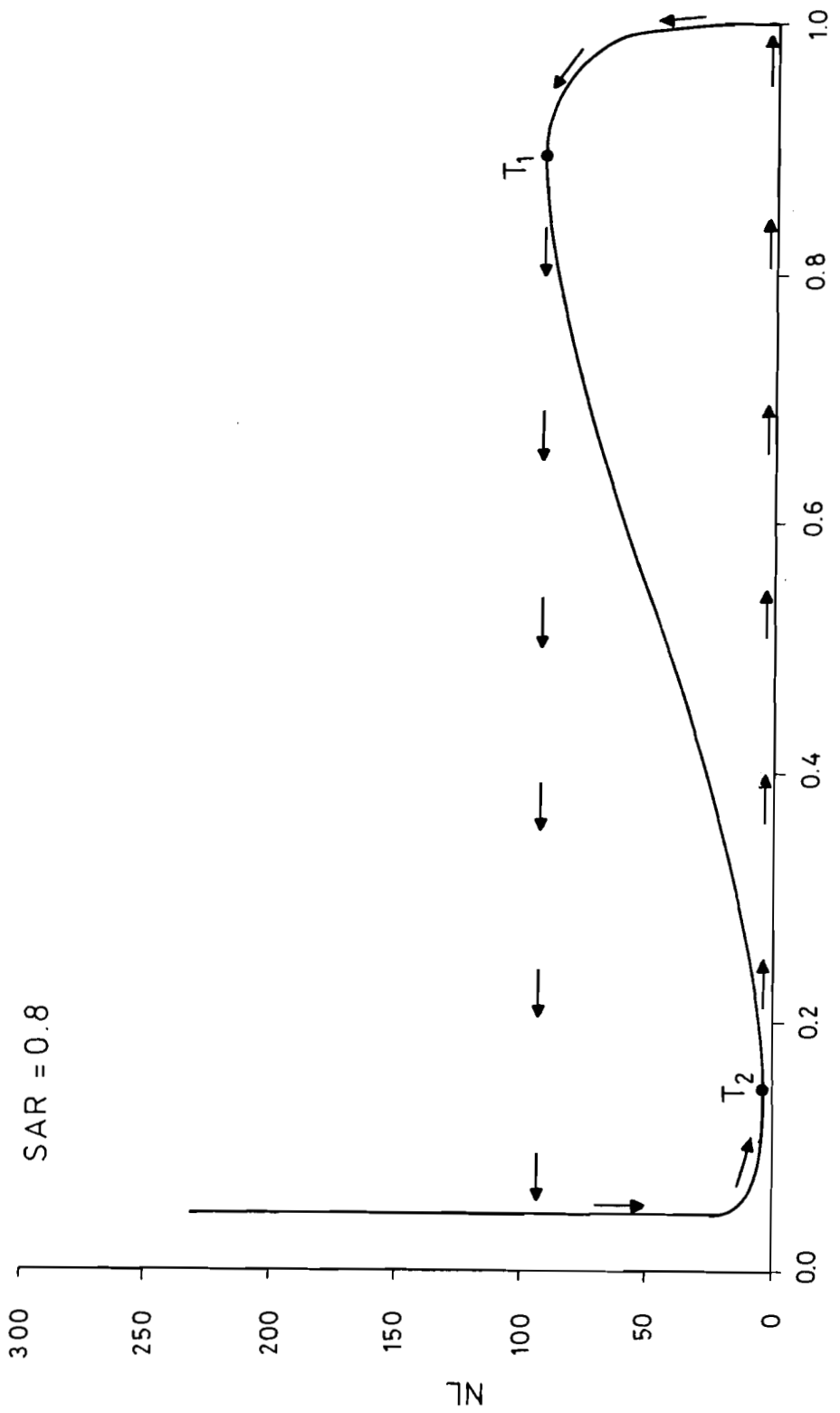

FIGURE 16

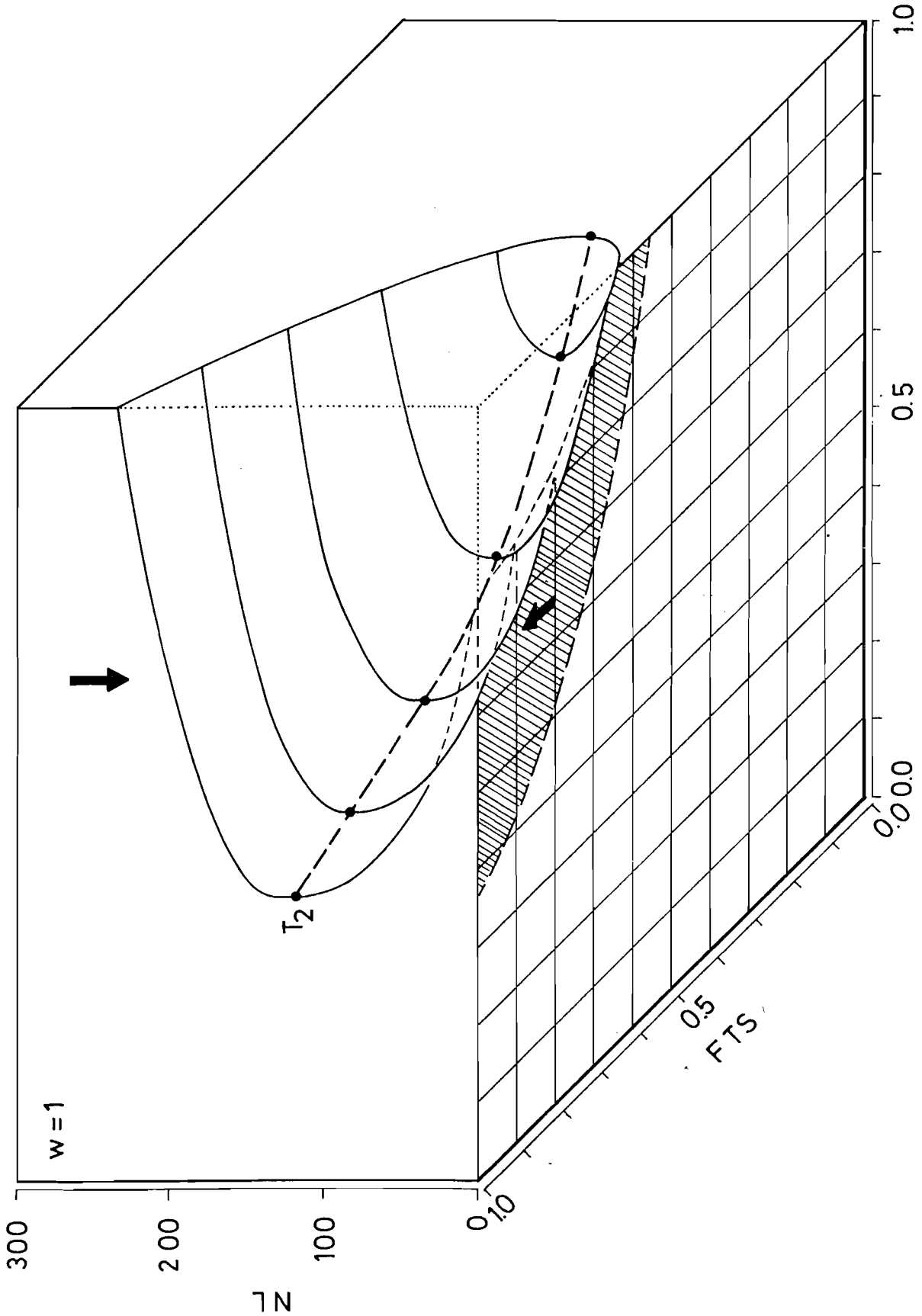

FIGURE 17

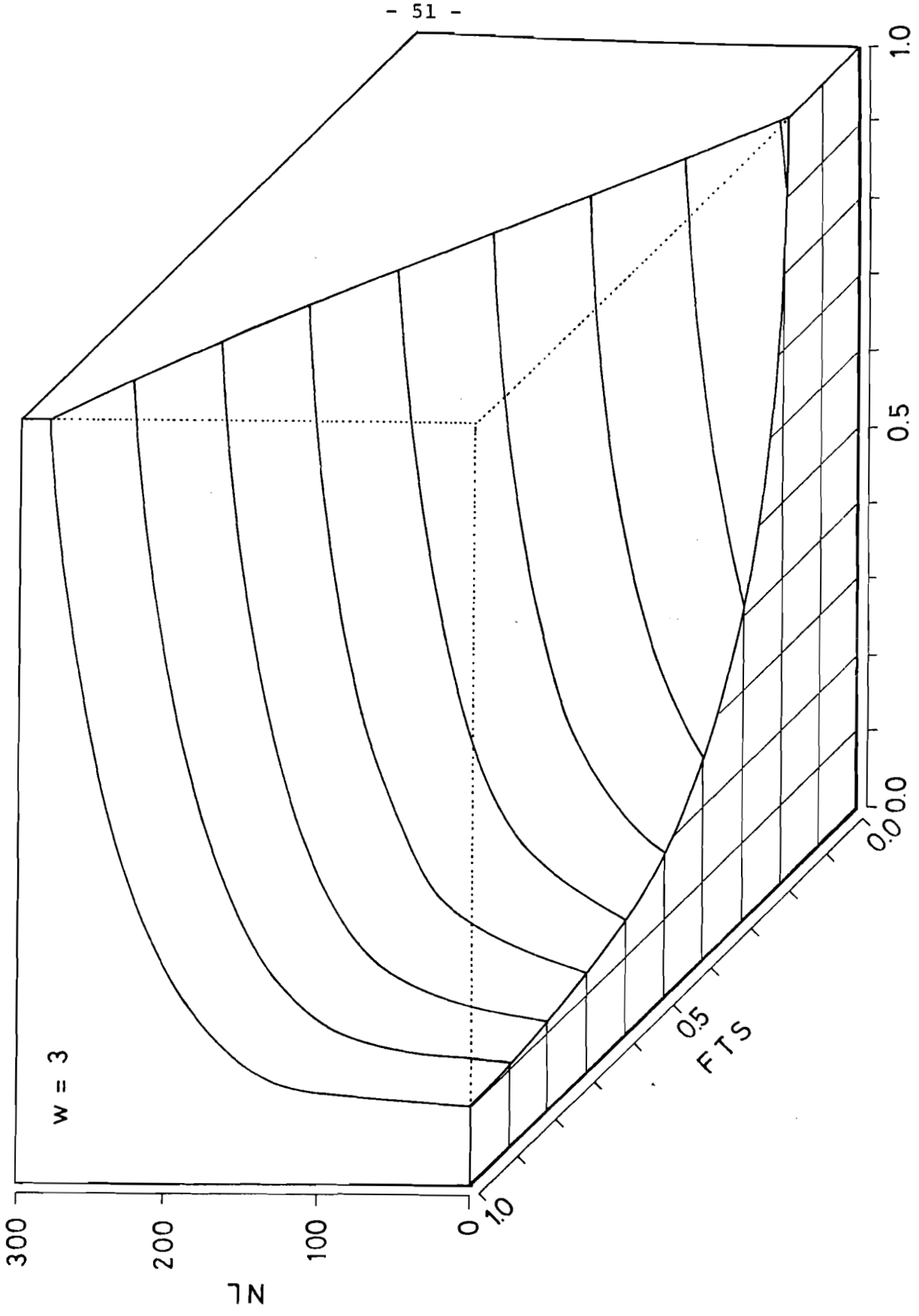

FIGURE 18

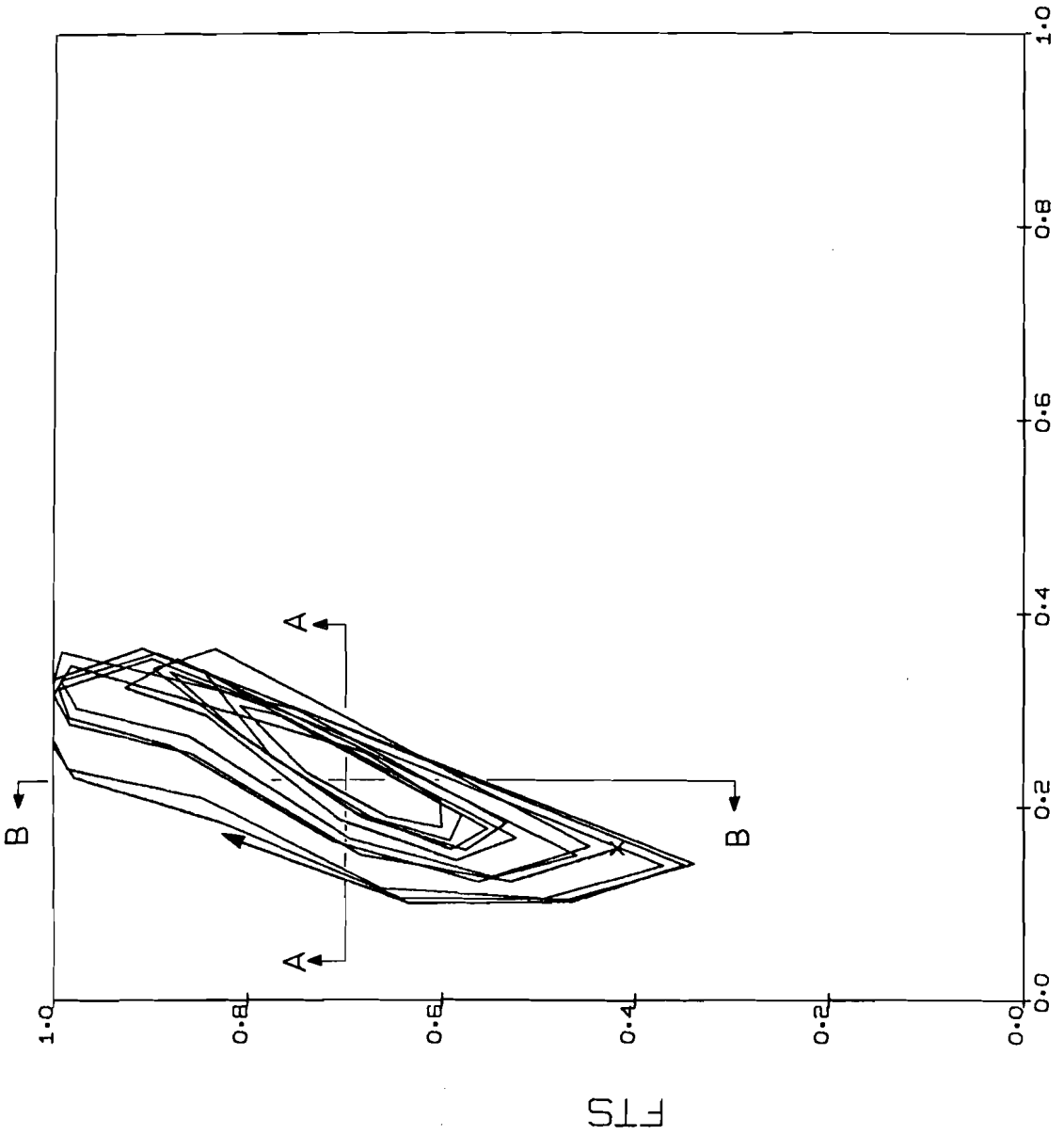

FIGURE 19

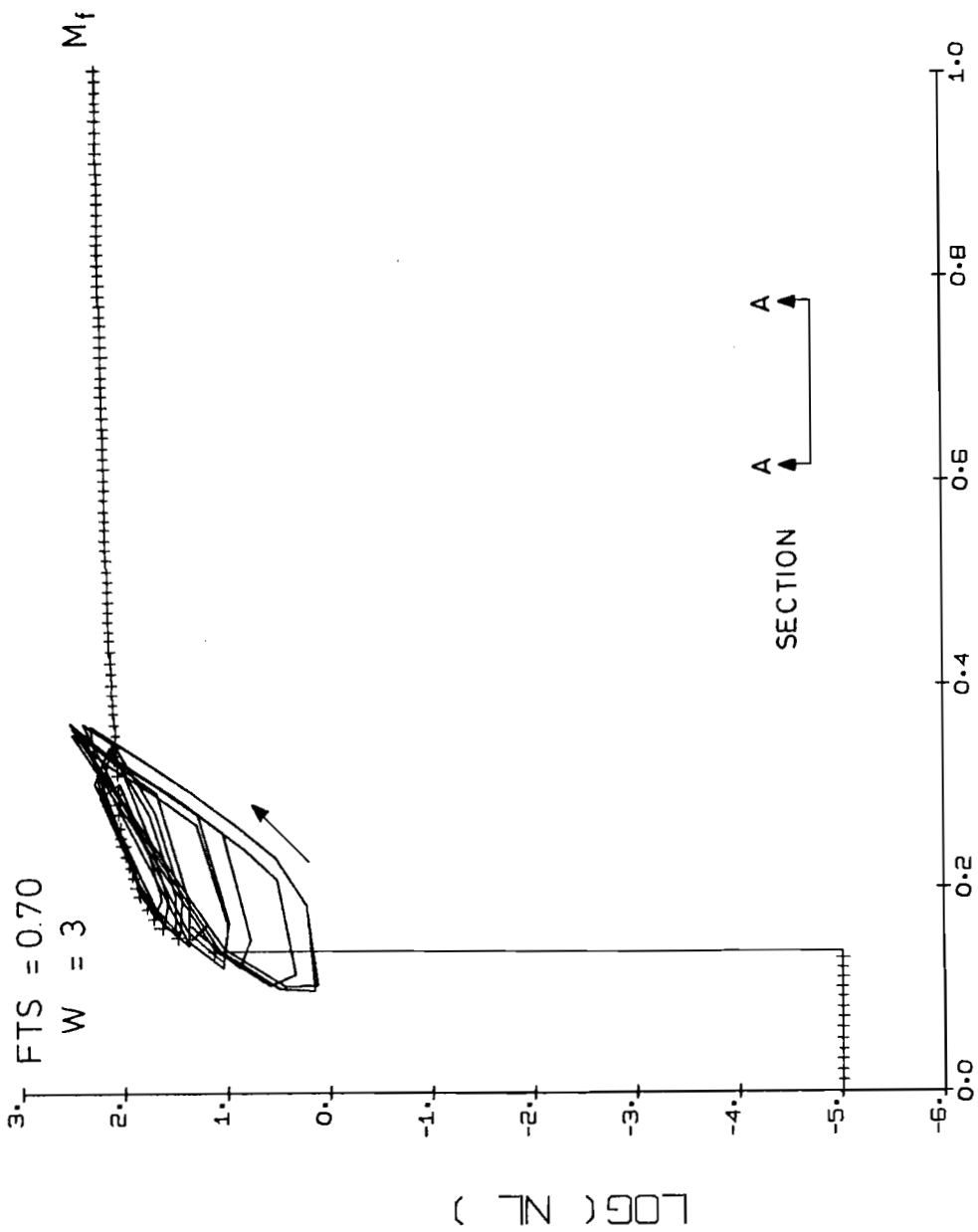

FIGURE 20

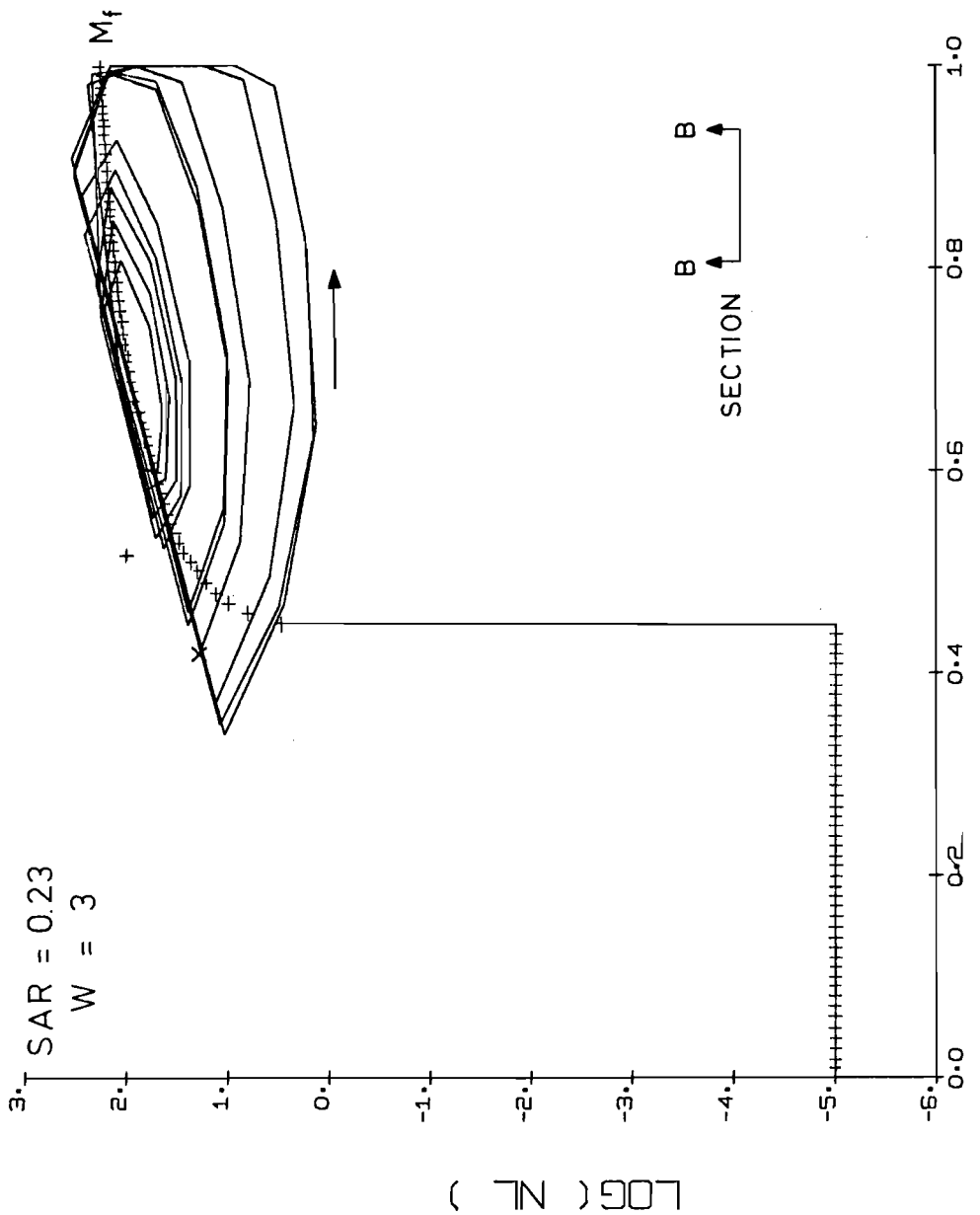

FIGURE 21

Table of Symbols

x	A general vector for variables with a fast dynamic behavior.
p,q	Parameters or slow variables.
n	The dimension of x (the number of components).
k	The number of parameters under consideration.
x^*	An equilibrium of x for fixed p.
$\dot{x}$	The time derivative of x $(=\frac{dx}{dt})$.
$f(x;p)$	The dynamic function that determines the behavior of x for fixed p. $(\dot{x} = f(x;p))$.
$V(x;p)$	An unspecified function, minimized as x approaches x^*.
M_f	Catastrophe manifold defined by $f(x;p) = 0$.
Π_f	The projection of M_f onto the parameter space.
T_1, T_2	Locations of folds on the manifold M_f.
S_1, S_2	Locations of the projections of T_1, T_2 in the parameter space.
NL	The population density of budworm larvae.
NL_0	The endemic level of NL.
F	A general measure of forest resource availability.
K	The upper asymptote of F.
r	"Intrinsic growth rate" of F.
m	"Mortality rate" of F due to budworm.
τ	Time delay for mortality of F.
SA	A unit of branch surface area.
SAR	SA scaled between 0 and 1.
FT	Total foliage (needles) per unit of SA.
FTS	FT scaled between 0 and 1.
w	An index of weather.

References

[1] HOLLING, C.S.
Resilience and Stability of Ecological Systems
Annual Review of Ecology and Systematics vol.4 1973

[2] ISNARD, C.A. ZEEMAN, E.C.
Some Models from Catastrophe Theory in the Social Sciences in Use of Models in the Social Sciences
ed. L. Collins
Tavistock London 1974

[3] MORRIS, R.F.
The Dynamics of Epidemic Spruce Budworm Populations
Memoirs of the Entomological Society of Canada vol.31 1963

[4] Project Status Report: Ecology and Environment Project
International Institute for Applied Systems Analysis
SR-74-2-EC 1974

[5] THOM, R.
A Global Dynamical Scheme for Vertebrate Embryology in
Lectures on Maths in the Life Sciences 5
The American Mathematical Society Providence, Rhode Island
1973

[6] THOM, R.
Stabilité structurelle et morphogénèse
Benjamin New York 1972

[7] THOM, R.
Topological Models in Biology in
Towards a Theoretical Biology 3
Edinburgh University Press Edinburgh 1970

[8] ZEEMAN, E.C.
Applications of Catastrophe Theory
International Conference on Manifolds Tokyo University
1973 (in press)

[9] ZEEMAN, E.C.
Differential Equations for the Heartbeat and Nerve Impulse in Towards a Theoretical Biology 4
Edinburgh University Press Edinburgh 1972

[10] ZEEMAN, E.C.
Geometry of Catastrophes
Times Literary Supplement 10th December 1971

[11] ZEEMAN, E.C.
Levels of Structure in Catastrophe Theory
Invited Address. International Congress of Mathematicians
Vancouver, Canada 1974