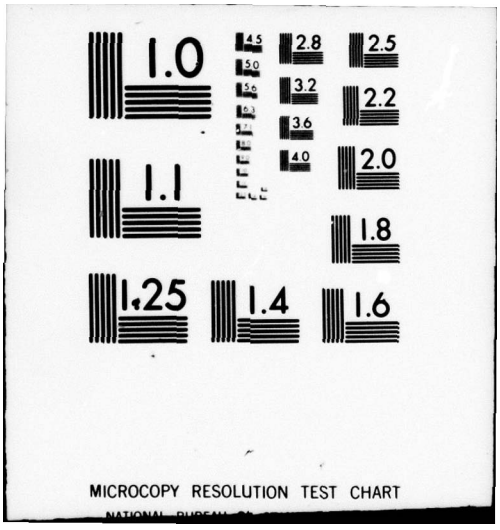
AD-A072 995 AIR FORCE GEOPHYSICS LAB HANSCOM AFB MA F/G 4/1
PHASE SPACE COORDINATES OF LONG LIFETIME TRAPPED RADIATION FROM--ETC(U)
JAN 79 C W DUBS

UNCLASSIFIED AFGL-TR-79-0029 NL

| OF |
AD A
072995

END
DATE
FILMED
9-79
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
MICROCOPY RESOLUTION TEST CHART

LEVEL II

12 B.S.

AFGL-TR-79-0029
ENVIRONMENTAL RESEARCH PAPERS, NO. 655

Phase Space Coordinates of Long Lifetime Trapped Radiation From Størmer Mapping

CHARLES W. DUBS

18 January 1979

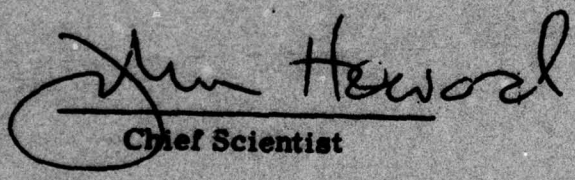
SPACE PHYSICS DIVISION PROJECT 2311
AIR FORCE GEOPHYSICS LABORATORY
HANSCOM AFB, MASSACHUSETTS 01731

AIR FORCE SYSTEMS COMMAND, USAF

This report has been reviewed by the ESD Information Office (OI) and is releasable to the National Technical Information Service (NTIS).

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER



Chief Scientist

Qualified requestors may obtain additional copies from the Defense Documentation Center. All others should apply to the National Technical Information Service.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: AFGL-TR-79-0029, AFGL-ERP-655	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): PHASE SPACE COORDINATES OF LONG LIFETIME TRAPPED RADIATION FROM STORMER MAPPING	5. TYPE OF REPORT & PERIOD COVERED: Scientific. Interim.
	6. PERFORMING ORG. REPORT NUMBER: ERP No. 655
7. AUTHOR(s): Charles W. Dubs	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Air Force Geophysics Laboratory (PHG), Hanscom AFB, Massachusetts 01731	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F, 2311G102
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Geophysics Laboratory (PHG), Hanscom AFB, Massachusetts 01731	12. REPORT DATE: 18 Jan 1979; 13. NUMBER OF PAGES: 27
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

Environmental research papers,

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

'stoermer mapping'

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Trapped radiation	Mapping
Protons	Homoclinic (points)
Electrons	Theorems
Lifetime	Maximum energy
Topology	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

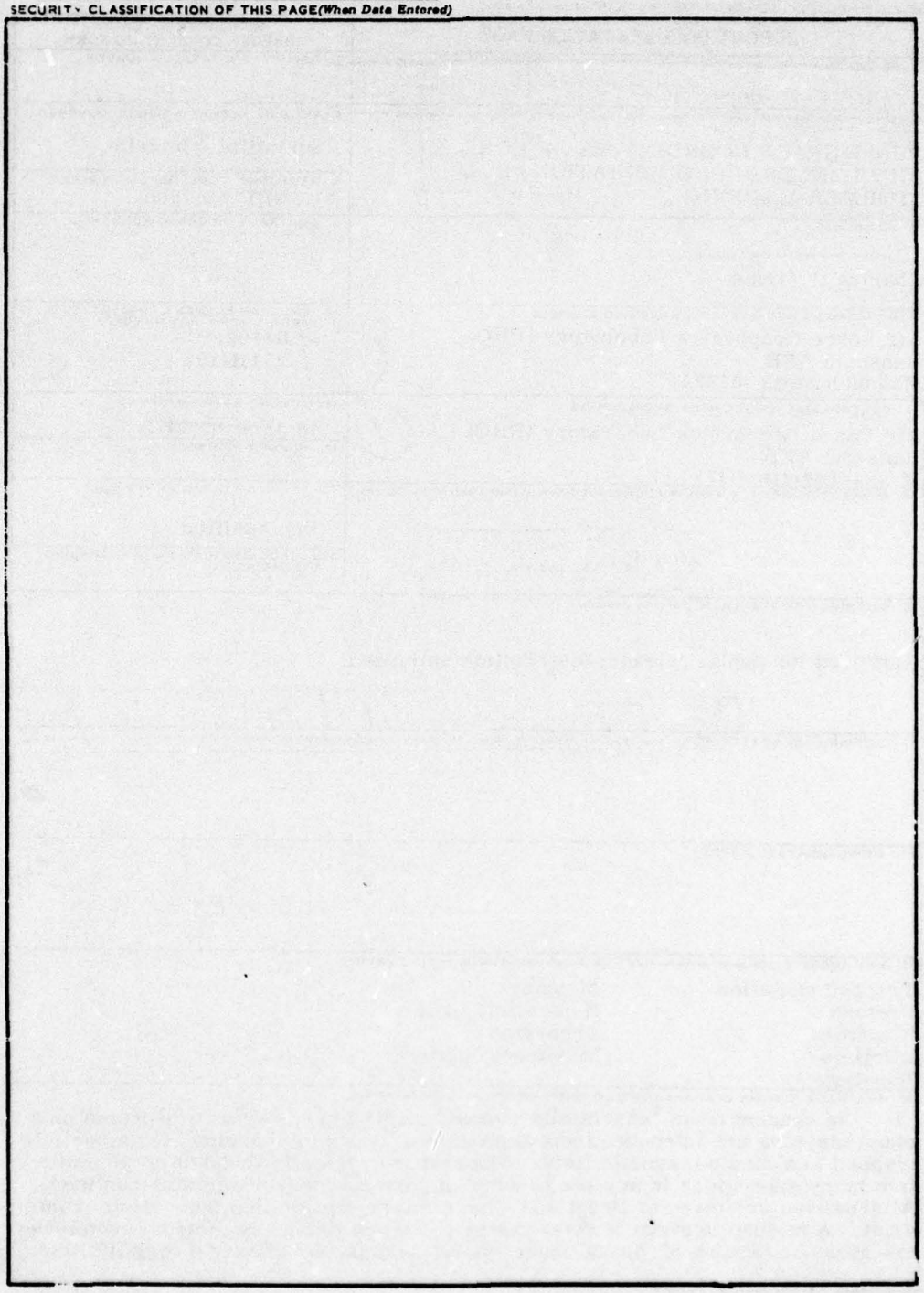
The concept of an "essentially closed" manifold and several theorems on a plane mapping are introduced and applied to a "Stormer mapping" for a particle trapped in a dipole magnetic field. The trapped region is found to be divisible into many subregions in any one of which a particle forever remains confined. Alternatives are given to Dragt and Finn's interpretation that homoclinic points exist. A method is given of determining if they do exist. Numerical examples are given of regions of phase space where protons should have a long lifetime.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

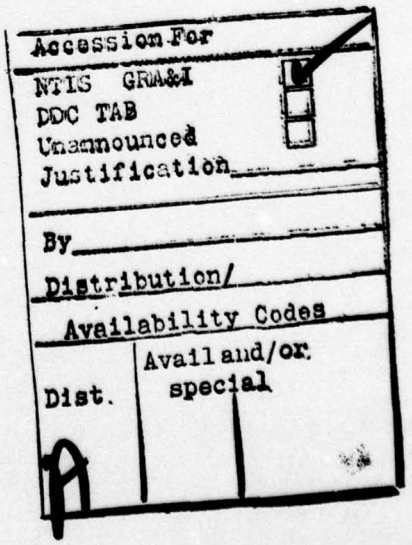
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

409 578



Preface

I thank R. Filz for suggesting this problem, for long interest and encouragement, and for many helpful discussions. I am grateful to Professor Dragt for kindly educating me on many fundamental points, to A. Quesada for many helpful discussions, and to both for reading and criticizing a manuscript of this report. I thank Don Smart for kind help in running the orbit program and M. A. Shea and him for suggested improvements on the manuscript. My thanks also go to M. MacLeod for kind discussions and encouragement.



Contents

Illustrations

Illustrations

Tables

Phase Space Coordinates of Long Lifetime Trapped Radiation From Stφrmer Mapping

1. INTRODUCTION

A knowledge of orbits of charged particles trapped in the geomagnetic field for long periods of time is important for understanding magnetospheric and radiation belt phenomena. For simplicity, only the dipole approximation of the geomagnetic field is considered here. Surprisingly, the problem is still so complicated that it remains largely unsolved. Dragt and Finn[1] have pointed out the difficulties in determining even roughly the trajectory and thus lifetime of charged particles trapped in a dipole field for long times, that is, tens of years. Firstly, there is no analytic solution. Secondly, numerical integration of orbits is not feasible due to a prohibitively large amount of computer time required, excessive truncation and round-off errors. Thirdly, the magnetic moment adiabatic invariant may change by several percent between successive equatorial crossings. No one has been able to show a maximum variation of any of the invariants over a long time. Finally, they find that a magnetic moment series (which they conclude is divergent), when truncated to minimize errors in a certain region, is not accurate enough to give any information after such a long time. They have taken a topological mapping approach,[2] referred to here as the "Stφrmer mapping", which they developed and

(Received for publication 17 January 1979.)

1. Dragt, A.J., and Finn, J.M. (1976) J. Geophys. Res. 81:2327.
2. Dragt, A.J. (1965) Rev. Geophys. 3:255.

utilized in two excellent fundamental papers.[1,3] This approach appears fruitful and is used in the present report.

An important question is, "Are there subregions of the trapped region in phase space to which particles are confined?". The answer is "Yes"; but they have not been well defined. The main purpose of this report is to divide the trapped phase space into confined subregions and to attempt to delineate these regions.

This report is written to be relatively self-contained. Section 2 gives the fundamentals of particle motion in a dipole field, its mapping, and necessary definitions. Much of it is a review taken from Dragt and Finn.[1] Some pertinent theorems are given in Section 3 and some results of numerical checks of some of them with "Cremona" mapping in Section 4. In Section 5, they are applied to the Stǿrmer problem, and integrals of motion and confinement are discussed. Numerical examples of possible confined phase space subregions are given in Section 6, and concluding remarks are made in Section 7.

2. FUNDAMENTALS AND MAPPING

Consider a magnetic dipole of magnitude M in the Z-direction at the origin and a trapped particle of charge q and mass m at the point R, φ, Z in cylindrical coordinates (Figure 1). The Hamiltonian is

$$H_a = \frac{1}{2m}\left\{p_R^2 + p_Z^2 + [qA_\varphi - p_\varphi/R]^2\right\}, \qquad \vec{A} = \hat{\varphi}\,A_\varphi,\ A_\varphi = MR/r_s^3,\ r_s^2 = R^2 + Z^2 . \tag{1}$$

Since this is cyclic in φ, the angular momentum

$$p_\varphi = q\,M\,/L \tag{2}$$

is a constant. For a trapped particle, the constant of integration L is a positive length. Equation (1) then may be regarded as two-dimensional with the square bracket the potential. The latter may be considered as the height of a surface above a plane as a function of two independent variables, R and Z, or r_s and λ, in the plane. The latitude angle is λ; $R = r_s \cos\lambda$; $Z = r_s \sin\lambda$. The relation between these two variables along the floor of the valley of such a surface is called the "thalweg" (German: "valley way"). In this case, the potential vanishes along the floor of the valley and the thalweg is

3. Dragt, A.J., and Finn, J.M. Stability of Charged Particle Motion in a Magnetic Dipole Field: A Preliminary Report of Numerical Studies, unpublished.

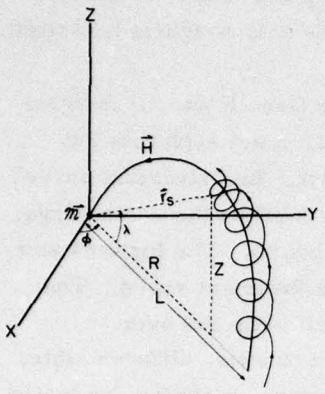


Figure 1. Coordinates of an Orbiting Particle in a Dipole Magnetic Field $\vec{H}$

$$r_s = L \cos^2 \lambda , \tag{3}$$

where L is the equatorial distance from the dipole of the guiding field line about which the particle gyrates. The thalweg is just the equation of this field line.

The variables are normalized to make them dimensionless, lengths normalized to L

$$\rho = R/L$$

$$z = Z/L$$

$$r = r_s/L$$

$$t = t' q M /mL^3$$

$$\dot{\rho} = d\rho/dt$$

$$\dot{z} = dz/dt .$$

Time is t'; t is the normalized time.

Twice the normalized Hamiltonian then becomes

$$2H = W_o^2 = \dot{\rho}^2 + \dot{z}^2 + (\rho/r^3 - 1/\rho)^2 , \tag{4}$$

where W_o is the particle's normalized speed and the parenthesis is the normalized φ component of velocity.

The Størmer mapping of the particle's trajectory is as follows. Every time the particle crosses the equatorial plane, z = 0, the value of $\dot{\rho}$ is plotted as a function of ρ. Let $P = \rho, \dot{\rho}$ be one such point. The forward mapping of P, MP, is defined as the point in the $\rho, \dot{\rho}$ plane when the particle next comes to z = 0, roughly half of a bounce period later. MP may be mapped again to give M^2P, and so on.

The backward mapping of P, $M^{-1}P$ is the point in the $\rho, \dot{\rho}$ plane when the particle just previously traversed the equatorial plane, $M^{-2}P$ is the next previous traversal, and so on.

For any mapping in a plane, the set of points M^nP for fixed P and all integral values of n constitutes an "invariant set" of points, that is, a set such that the mapping of any point in the set gives a point in the same set. An "invariant curve" is a curve such that the mapping of any point on it gives a point on the same curve. It usually contains an infinite number of invariant sets of points. The forward and backward mapping of any curve connecting P and MP is an invariant curve. The latter would in general be very complicated, crossing itself over and over.

The Störmer mapping is single-valued, invertible, continuous, differentiable, area-preserving, has fixed points, and is bounded. "Cremona" mapping,[1] a mathematical two-dimensional quadratic mapping, has these same properties except that it is unbounded. These properties of the Störmer mapping deserve explanation. There is a unique trajectory through every point in phase space. A phase point of an orbit maps to only one point, so the mapping is single-valued. Conversely, a mapped point corresponds only to one phase point of an orbit or of its mirror image in the equatorial plane. So knowledge of the mapping is equivalent to knowledge of the orbit. By going backward or forward in time, a point may be mapped backward or forward; the mapping is thus invertible. As any point P approaches P_1, MP approaches MP_1, so the mapping is continuous. Except for $P_o = \rho_o, \dot{\rho}_o$, the point from which a particle goes to the origin (to the dipole), any point MP may be differentiated with respect to the coordinates of P. The area inside of any closed curve equals the area inside of the mapping of that curve, that is, the mapping is area-preserving.

For any mapping in a plane, if MP = P, then P is a "fixed point" of M. It is an invariant set containing only one point and corresponds to a periodic orbit. Generally, M and its powers have three kinds of fixed points; hyperbolic, reflection hyperbolic, and elliptic. The names relate to the smoothest invariant curves close to the fixed point.[1,3]

In the Störmer mapping, for $W_o = 0.1$, there are[3] about 20 hyperbolic fixed points of M, h_i, $i = 1, 2, \ldots$, on the ρ axis at $\rho > \rho_o$. The number increases rapidly as W_o decreases, but remains finite.[4] The two hyperbolic fixed points of M, h_1 and h_2, at the largest values of ρ are shown in Figure 2. The curve B is the equation

$$W_o^2 = \dot{\rho}^2 + (1/\rho^2 - 1/\rho)^2 , \tag{5}$$

obtained from Eq. (4) with $z = \dot{z} = 0$. It is the boundary of the physically meaningful region and corresponds to orbits in the equatorial plane.

4. Dragt, A.J. private communication.

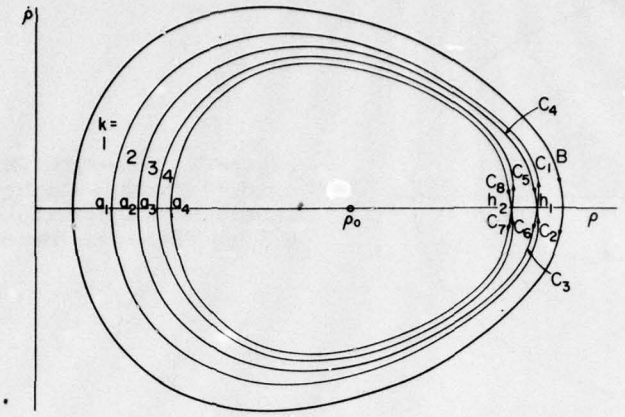


Figure 2. Mapping, M, of the Bound Stǿrmer Problem. B is the boundary of the physical region; the C_j's are the manifolds of the hyperbolic fixed points h_1 and h_2, the two closest to B. Values of k denote regions between the first parts of pairs of (double or homoclinic) manifolds, a_k the intersection points of the manifolds with the ρ axis, and ρ_0 the ρ coordinate of the orbit to the dipole. (ρ_0 is slightly negative.) The diagram, from calculated data of Dragt and Finn[1,3] is nearly to scale for $W_0 \simeq 0.1$, except that the manifolds are distorted slightly near h_1 and h_2 to separate them clearly

In a plane mapping, four invariant curves, for example C_j in Figure 2, called "manifolds" ("separatrices" by some) emanate from every hyperbolic fixed point, h. Two "unstable" manifolds, arrows pointing away from h, start from h in opposite directions along one straight line, and two "stable" manifolds, arrows pointing toward h, start from h in opposite directions along another straight line. Each arrow indicates the direction of forward mapping of a point and thus the stability of the manifold. In this report, "manifold" is used to mean only such an unstable or stable manifold.

In the Stǿrmer mapping, for both positive and negative particles, points move with time in the direction of the arrows on the manifolds C_j (Figure 2). An unstable (j odd) and a stable (j even) manifold from each h_i come together. The pair either merge together smoothly as shown in Figure 2 to make what is called here a "double manifold", or, as concluded by Dragt and Finn[1] and sketched in Figure 3 (or possibly as in Figure 4), the pair cross each other at nonzero angles at what are called "homoclinic points", making what here is called a pair of "homoclinic manifolds" with "homoclinic loops", loops of nonzero area between homoclinic points.

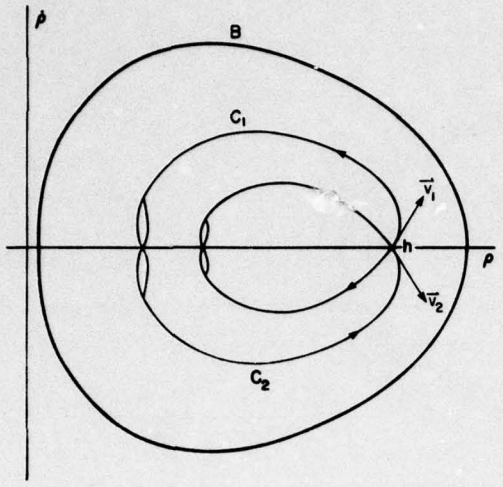


Figure 3. Schematic Diagram of an a priori Possible Configuration of Manifolds of a Hyperbolic Fixed Point h When There are Homoclinic Points

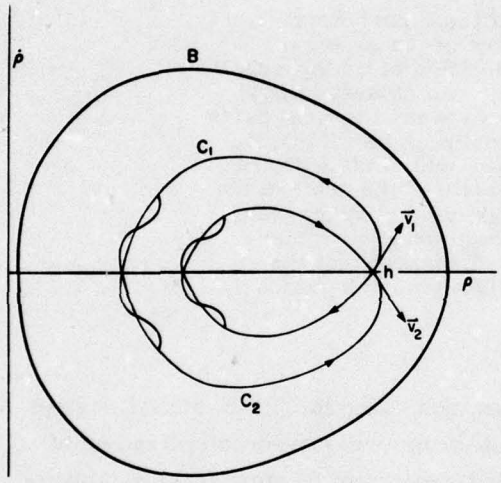


Figure 4. The Same as Figure 3: Another Possible Configuration

If two manifolds from different hyperbolic fixed points cross, the term "heteroclinic" is used instead of "homoclinic". Consider two merging or crossing manifolds. Suppose one moves equal distances forward on the unstable manifold and backward on the stable manifold from their starts at the hyperbolic fixed point(s) to the same point or to points nearest each other. In the latter case, the two points are then moved along the manifolds to the nearest homoclinic or heteroclinic point. The part of each manifold moved along and not retraced is called here the "first part" of the manifold. In general, C_j may be paired with C_{j+1} or with C_{j+1+2n}, n a nonzero integer. The length of the first part of C_j depends on which C_J is considered in the C_j, C_J pair.

In Figure 2 or 3 or 4 for the pair C_1, C_2, for example, the first part of C_1 is the part from h_1 to a_1, $\rho \geq 0$, and the first part of C_2 is the part from h_1 to a_1, $\rho \leq 0$. (For the pair C_1, C_4, the length of the first part of C_1 is longer than it is for the pair C_1, C_2.) Thus far, numerical plots of the manifolds of h_i's from

accurate numerical integration show them to be indistinguishable from double manifolds. In either case, the first part of these manifolds divides the bounded area within B into smaller regions, $k = 2i - 1$ and $k = 2i$ as indicated in Figure 2. If there are homoclinic points, other regions (see Section 5.1) are defined which are more complicated. If the first part of a manifold goes counterclockwise in the direction of the arrow (C_1, C_2, C_5, C_6, and so on), the right-hand side of the manifold is defined as outside of it and the left-hand side as inside. The right-hand side of B or of a manifold whose first parts go clockwise in the direction of the arrow (C_3, C_4, C_7, C_8 and so on) is defined as inside and the left-hand side as outside of that curve.

Every point on B stays on and moves continuously along B with time in the direction of the arrow. Every point within B maps to a point within B.

For a plane mapping in general, if a point is on a manifold, the whole invariant set containing that point is on the manifold. Suppose that we continuously move along a manifold, in the direction of the arrow if unstable, in the opposite direction of the arrow if stable. Suppose then that parts of the manifold reapproach the hyperbolic fixed point from which it stems. If the point of closest approach reapproaches this fixed point without limit, then the manifold is called here "essentially closed". Each of a pair of homoclinic manifolds (from the same fixed point) as well as a double manifold is always essentially closed.

3. THEOREMS

Some useful theorems for a plane, single-valued mapping M follow. Theorems 1, 3, and 4 apply only to manifolds of a hyperbolic fixed point h. Theorems 2 and 7 also apply to manifolds of two hyperbolic fixed points.

Theorem 1. A manifold cannot intersect, merge with, or become tangent to itself.

In general, this theorem does not hold for an invariant curve which is not a manifold.

Theorem 2. Two stable (unstable) manifolds can have no point in common other than a common fixed point from which both stem.

Theorem 3. If a stable and unstable manifold from the same fixed point have a point in common other than h, then either both manifolds are essentially closed homoclinic manifolds or form a double manifold.

Theorem 4. Any point outside (inside) of an essentially closed manifold cannot be mapped to a point inside (outside). That is, the points of an invariant set are all outside, all on, or all inside of each essentially closed manifold.

Corollary 1. A point outside (inside) of both of a pair of homoclinic manifolds can be mapped only to points outside (inside) of both of them.

The proofs of these four theorems are given elsewhere.[5] Theorem 1 was found to have been discovered previously by Poincaré. Diagrams in the literature generally adhere to Theorem 2; so it apparently has been discovered previously also, probably by Poincaré. Theorem 3 is rather trivial. Theorem 4 appears useful and new.

In the next three theorems and corollary, the mapping is assumed to be single-valued, area-preserving, and bounded (by a boundary B). (Instead of area-preserving, it is sufficient to assume that a nonzero finite area is always mapped to a nonzero finite area.) In the next two theorems, let A_o be any region with nonzero area within B, and let $A_n \equiv M^n A_o$, n integral, be the nth mapping of A_o. Also, n is chosen positive.

Theorem 5. There exists no value of n for which A_n and A_{n+1} (A_{-n} and A_{-n-1}) intersect but A_{n-1} and A_n (A_{-n+1} and A_{-n}) do not.

Proof: Suppose the contrary. M^{-1} of a point in the intersection of A_n and A_{n+1} is simultaneously a point in A_{n-1} and A_n, a contradiction. Similarly with negative subscripts, taking M of a point, q.e.d..

Theorem 6. There exists a value N of n such that A_N intersects A_o, A_{N+1} intersects A_1, and so on, but none of the different A_n's intersect one another for $1 \leq n \leq N$. Similarly for negative subscripts.

Proof: From the assumptions, there must exist a value of n such that A_n intersects A_i, $0 \leq i < n$. For N the smallest such value of n, M^{-1} of a point in the intersection must simultaneously be in both A_{N-1} and A_{i-1}. This contradicts the definition of N unless $i = 0$. M of the intersection of A_o and A_N must lie in both A_1 and A_{N+1}, and so on. Similarly for negative subscripts, q.e.d.. Making A_o arbitrarily small leads to the following corrolary.

Corollary 2. The continued mapping of any point yields points repeatedly without limit arbitrarily close to the original point.

Theorem 7. Every region of a homoclinic loop must be cut repeatedly without limit by one or both manifolds so that the area of every uncut region approaches zero.

Proof: By Theorems 1 and 6, every homoclinic loop must be intersected repeatedly by mappings of that loop. But since the loop is formed by homoclinic manifolds, the boundaries of the mapping of the loop are also (mapped) parts of the same manifolds. If there remained any nonzero uncut area, all of its mappings would also have to be uncut and therefore non-overlapping. This is impossible, since the sum of their areas would be infinite, q.e.d..

5. Dubs, C.W. Cremona Mapping, unpublished.

Eq. (3.3) of Reference 3 leads to what here will be called

Theorem DF. In the Stǿrmer mapping, every stable (unstable) manifold of h_i (on the ρ axis) is the mirror image through the ρ axis of an unstable (a stable) manifold.

4. CREMONA MAPPING

As a check on the first four theorems, especially Theorem 4 and Corollary 1, Dragt and Finn's Cremona mapping,[1] M_c, a mathematical two dimensional quadratic mapping, was studied numerically in some detail.[5] With the eigenvalue $\lambda = 3$, the homoclinic manifolds were calculated with M_c^n and computer plotted (28 × 30 in.) to n = 35, starting at a distance of about 10^{-5} from the hyperbolic fixed point h. To check Theorem 4, a homoclinic point P and 39 other points on the unstable manifold, evenly spaced in the linear region (close to h) between P and MP, generated 40 invariant sets of points. The 39 sets were all found to be outside of the stable manifold. (One point had to be mapped out to n = 91.) To check Corollary 1, several points inside of the first parts of the manifolds were mapped. All but two of the invariant sets were outside of both manifolds. One of the two was 14.5 and the other 17 percent of the way from the first part of the manifolds to the elliptic fixed point inside of them. Both were mapped to $n = \pm 32,768,000$ and the invariant sets found to be wholly inside of both manifolds. The eight manifolds (four from a reflection hyperbolic and four from a hyperbolic fixed point) for the Cremona mapping with $\lambda = -3$ were also computer plotted with a large scale. The four unstable (stable) manifolds repeatedly cross the four stable (unstable) manifolds. In conclusion, no inconsistency could be found with Theorems 1 through 4 and Corollary 1.

5. STØRMER PROBLEM

Theorems such as those in Section 3 along with previous knowledge can give further insight into the Stǿrmer problem. Such an improved understanding of some aspects of this problem is the purpose of this section.

5.1 Application of Theorems

All of the theorems and corollaries in Section 3 apply to the Stǿrmer mapping. The proof of Theorem 4 requires the existence of a point outside of the manifold whose forward and backward mapping once are also outside. This can be a point on B.

All of the manifolds of the points h_i are essentially closed. (See Figures 2, 3, and 4.) If there are no homoclinic points, by Theorem 4, an orbiting particle which maps to a particular region k moves so that it forever maps only to that region k. If there are homoclinic points, other regions are also defined.* The region inside of B is expected to be divided into many non-overlapping regions $u_{a;b}$ ($s_{a;b}$), a = i, j, ..., b = m, n, ..., each of which is inside of unstable (stable) manifolds C_i, C_j, ..., and is outside of unstable (stable) manifolds C_m, C_n, ... The following argument shows this indeed to be the case. From Theorem 2 and the fact that the manifolds are essentially closed, C_3 is inside of C_1, C_7 is inside of C_5, and so on. (It can be shown that $C_{1+4i+\epsilon}$ and $C_{3+4i+\epsilon}$ both reapproach h_{1+i} along the start of both $C_{2+4i-\epsilon}$ and $C_{4+4i-\epsilon}$, i a non-negative integer, ϵ = 0 or 1.) A priori, an unstable manifold of h_i (or both unstable manifolds) can conceivably include (one or more) r_s, j > i, and their unstable manifolds and conceivably $\rho_o, \dot{\rho}_o$ in one of its homoclinic loops making them <u>outside</u> of the unstable manifold of h_i without violating the theorems. However, the extreme smallness of the upper limit of the area of such a homoclinic loop as determined from computation,[1] precludes this possibility.** Therefore, also C_5 is inside of C_3, C_9 is inside of C_7, and so on. Region u_1 or $u_{B;1}$ is defined as the region inside of B and outside of C_1; region u_i or $u_{2i-3;\,2i-1}$ is defined as the region inside of C_{2i-3} and outside of C_{2i-1}, i > 1. The last (innermost) region is defined as all of the region inside of the last unstable manifold (except for any part** which may already be included in another u_i). Similarly for stable manifolds, C_4 is inside of C_2, and so on. Region s_1 or $s_{B;2}$ is defined as the region inside of B and outside of C_2; region s_j or $s_{2j-2;\,2j}$ is defined as the region inside of C_{2j-2} and outside of C_{2j}, j > 1. From Theorem DF, regions u_i and s_i are mirror images of each other through the ρ axis. None of the u_i (s_j) regions overlap one another, although a u_i (s_j) region may overlap one or more s_j (u_i) regions, and the sum of the u_i (s_j) regions is just the region inside of B. From Theorem 4, a particle that maps to a point in region u_i (s_j) forever stays in the region of phase space that maps to that region. From Corollary 1, a particle that maps to region us_{ij} forever stays in the region of phase space that maps to us_{ij}, where the latter is the intersection of u_i and s_j. In general, us_{ij} may be the sum of many separate regions. It may be that no us_{ij}, $i \neq j$, region exists. It may be that $u_i = s_i = us_{ii}$. (See the last paragraph of this subsection.) With homoclinic points on these manifolds, the trouble is that each manifold may possibly wander outside and inside of the first part of it in a complicated way. No one has yet been able to set limits on how close to or far from B a manifold and thus a region u_i (s_j) can extend. Since

*To visualize a priori possible configurations of the manifolds, it is helpful to think of the first parts of manifolds as rigid and the other parts as elastic. The latter may then be bent, distorted, stretched, and thus moved in any way that does not violate the theorems.

**Although it is unlikely, a homoclinic loop could conceivably include a small neighborhood of $\rho_o, \dot{\rho}_o$.

in general the mirror point is closer to the dipole the closer a mapped point is to $\rho_o, \dot{\rho}_o$ (the farther from B), how low an altitude a geomagnetically trapped particle can come, and thus its lifetime, cannot (at least as yet) be determined without doubt, even if it is known to which us_{ij} region it maps. Some help is obtained from a consideration of integrals of motion.

When the mapping is the phase plane as in the Størmer mapping, Corollary 2 is just what is known as Poincaré's theorem. Theorems 6 and 7 and Corollary 2 indicate the possibility of ergodicity and may aid in leading to its existence.

The Cremona mapping[5] indicated (but did not prove) that there exist no points on or outside of one of a pair of homoclinic manifolds that are inside the other manifold of the pair. It may be, therefore, that the region inside of loops between homoclinic points, which is apparently inside one and outside of the other of the homoclinic manifolds, is actually all inside or (more probably) outside of both manifolds.

5.2 Integrals of Motion

Figure 5 shows schematically the outer-most fixed points of M and M^2 and the manifolds of the hyperbolic ones for W_o near 0. 1. The points with odd half integral subscripts are fixed points of M^2 and not of M. Elliptic fixed points are labelled e. Farther from B (closer to $\rho_o, \dot{\rho}_o$) than shown, there are reflection hyperbolic instead of elliptic fixed points. However, they become elliptic as W_o is decreased.[3]

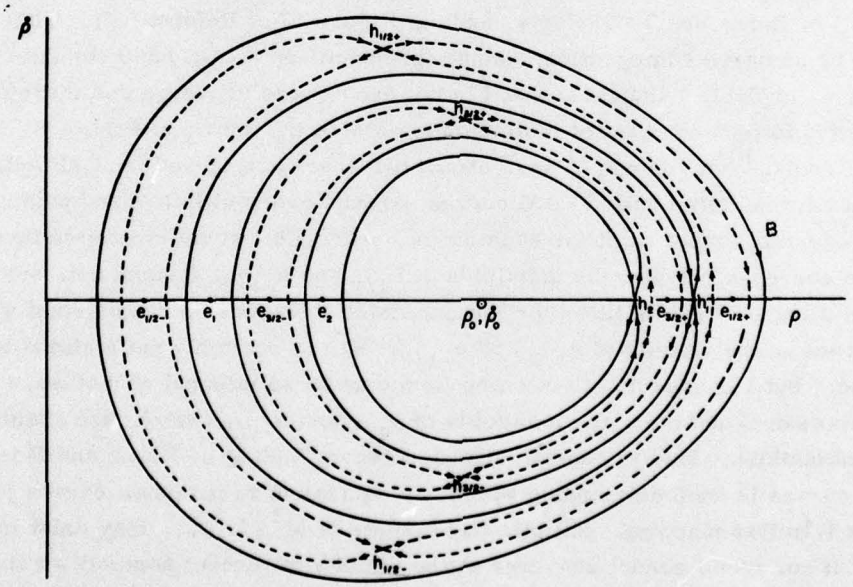


Figure 5. Schematic Diagram (not to scale) of the Outermost Fixed Points and Their Manifolds of M, Integral Subscripts, and of M^2, Odd Half Integral Subscripts

Contopoulos and Vlahos[1,6,7] have developed the Hamiltonian in a power series in dipolar coordinates, q_1, q_2, and their canonical momenta p_1, p_2.

$$q_1 = z/r^3, \quad q_2 = r^3/\rho^2 - 1 . \tag{6}$$

Each term is homogeneous in these four coordinates, the first term being quadratic. The series converges[7] for $|q_1| < 3\sqrt{3}/16 = 0.325$. $q_1 = \sin\lambda/(1+q_2)^2 \cos^4\lambda$, where λ is the latitude angle. Neglecting q_2, this implies $|\lambda| < 35°$. Also required for convergence is $|q_2| < 1$, but this is always the case for geomagnetically trapped particles; $|q_2| \lesssim W_o$. Even for $W_o = 0.1$, which is larger than that for Van Allen Belt particles, $|q_2| < 0.13$.

In the process of reducing H to normal form, Dragt and Finn[1,7] have developed another similar but independent series I. No analytic determination of its convergence has been made; it may diverge everywhere as concluded by Dragt and Finn,[1] or it may converge in some region. It seems reasonable to expect it to converge in the same region as that (in last paragraph) for which the series for H converges. Figures V-30 through V-37 of Reference 7 indicate a radius of convergence in q_1 of about 0.325 for I. The quadratic terms, the cubic terms, and so on, of the Poisson bracket of I with the series for H vanish. I must therefore be a third (in addition to H and p_ϕ) integral of motion in the region in which it and the series for H converge. The values of maximum $|q_1|$ for some orbits are shown in Table 1. The orbits were started at the values of $\rho, \dot{\rho}$ given in the first five and the ninth rows of Table 1 of Reference 3 and (for a_1 and a_2) Figure 13 of Reference 1. They were obtained by numerical integration, using a program[8] on a CDC 6600 computer. The value for a_2 in Table 1 indicates that I is not expected to converge and therefore is not expected to be a constant of motion on manifolds C_3 and C_4 of Figure 2. Kolmogorov,[9] Arnold,[10] and Moser[11] have shown that there are closed (roughly elliptical) invariant curves, often called KAM curves, around every elliptic fixed point.[7] An integral of motion must exist for each curve. Although I would otherwise be expected to converge between the manifolds of $h_{1/2}$ and $h_{-1/2}$, it must not, since constant I is a locus of points like B or the manifolds in Figure 5, inconsistent with the KAM curves in the vicinity of $e_{1/2}$ and $e_{-1/2}$. So I is certainly not a global integral of motion. But I is expected to converge and thus be an integral of motion in the region between B and the outer manifolds of $h_{1/2}$ and $h_{-1/2}$ (which then should be double manifolds). This is consistent with a recent finding by Braun and Moser[12] of KAM curves in another mapping which are equivalent to invariant curves just inside of B in this mapping. Elliptic fixed points of M^3, M^4, ... may exist in this region. If so, then I cannot converge in the vicinity of (except possibly at) those points either.

(Because of the large number of references cited above, they will not be listed here. See References on Page 25.)

Table 1. Values of Maximum $|q_1|$ for Orbits Passing Through Various Points in Figures 2 and 5 for $W_o = 0.1$

Point	Max $\|q_1\|$
$h_{1/2}$	0.1917
$h_{-1/2}$	0.1917
$e_{1/2}$	0.1927
$e_{-1/2}$	0.1927
a_1	0.252
h_1	0.3057
e_1	0.325
a_2	0.392
h_2	0.5251

For orbits passing through points inside of C_3 and C_4, the maximum value of $|\lambda|$, nearly the magnitude of the mirror latitude, λ_m, is larger than 35°, so I is not expected to exist, much less remain a constant of motion, over all of the orbit. However, since q_1 approaches zero at the equator, I may converge, be analytic, and be constant on the part near the equator of every orbit. Then I would be a certain constant until $|\lambda|$ increases to more than ~35°; when $|\lambda|$ decreases to less than ~35°, it would again become constant, but in general a different constant value. A numerical example that indicates this is shown in Figure V-32 of Reference 7. If the mapping of the orbit is in a subregion containing an elliptic fixed point, the constant would have to change, unless the orbit is periodic, since the locus of constant I differs from the invariant curves surrounding an elliptic fixed point. In other subregions, for example between adjacent solid and dashed curves in Figure 5, invariant curves may exist for which I may be a correct expansion of the integral of motion for q_1 sufficiently small. Since periodic orbits return to the equator with the same value of phase coordinates with which they left, of course the value of I (as any function only of phase coordinates) must also return to the same value.

Dragt and Finn show in Figures 16 and 17 of Reference 1 that the value of I_6^T, the terms of I through the sixth degree, varies with i when calculated at M^i of a point quite close to the first parts of C_1 and C_2 in Figure 2 for $W_o = 0.1$. This variation: (1) is appreciably larger[4] than $|I_n|$, n close to 6, where I_n is the nth degree (polynomial) term of I, and (2) is much larger than the error in calculation.[1]

One possibility, the interpretation of Dragt and Finn,[1] is that I does not converge, and this variation is due to the existence of homoclinic points. If this is correct, the computer should also be accurate enough to show directly the existence of homoclinic points. A method of doing this is outlined in Appendix A. A second possibility is that I does not converge at all points of the orbit, so takes on different values at different equatorial crossings; the manifolds contain no homoclinic points; an integral of motion different from I exists for the manifolds.† These two possibilities are consistent with the expectation in the next to the last paragraph. A third possibility is that, in spite of that expectation, I converges; the variation is due to I converging slowly enough to be from $I - I_6$. For a point P in some regions, including neighborhoods of hyperbolic fixed points, the position of M^iP for large $|i|$ is critically dependent on the position of P and the accuracy of mapping. Instead of the existence of homoclinic points, the closeness of P to h and small errors in mapping may possibly account for the "transitions" found by Dragt and Finn[1] (next to last paragraph of their Section 6). The closer that h_i is to $\rho_o, 0$, the more integration steps are required and thus the more magnified is the error in calculating the position of M^iP. Possibly this error accounts for the interpretation of increasing homoclinic angles and area between oscillations the closer h_i is to $\rho_o, 0$ in the last paragraph of Section 6 in Reference 1. In conclusion, the existence of homoclinic points in this Stǿrmer mapping seems not yet to be well established, except perhaps in the neighborhood of $\rho_o, \dot{\rho}_o$.[12]

If the manifolds of h_i are homoclinic, each is probably limited to a narrow oval distorted annulus. Even in the Cremona mapping,[5] with only one hyperbolic and one elliptic fixed point of M_c, the manifolds extend a maximum of about 43 percent of the way from the first parts of the manifolds to the elliptic fixed point. With the many hyperbolic fixed points of M that exist in the Stǿrmer mapping, the width of the extent of each manifold should be a much smaller fraction of the way from B to $\rho_o, \dot{\rho}_o$.

---

†Although unrealistic, to show how such an integral of motion could exist, suppose that b_i is an integral of motion satisfying the equation

$$\left[\frac{r}{b_i} - \frac{1+c\cos\theta}{1+e\cos\theta} - \delta_1\delta\frac{\sqrt{1-\cos\theta}}{\delta^2+1-\cos\theta}\right]\left[\frac{r'}{fb_i} - \frac{1+c\cos\theta'}{1+e\cos\theta'} + \delta_1\delta\frac{\sqrt{1-\cos\theta'}}{\delta^2+1-\cos\theta'}\right] = 0 ,$$

$$x - x_1 = r\cos\theta , \; y = r\sin\theta , \; x' = r'\cos\theta' , \; y = r'\sin\theta' ,$$

$$x' = x - x_1 - \frac{1+c}{1+e}(1-f)b_i, \; 0 < e < c < f < 1, \; 0 < \delta < c - e ,$$

$0 < \delta_1 < 1$. The locus of points satisfying this equation looks qualitatively like C_1, C_2, C_3, and C_4 in Figure 2.

In the vicinity of a hyperbolic fixed point, h, an analytic transformation to variables ξ and η can be made[1, 13] such that $\xi\eta$ is a constant, giving hyperbola-like invariant curves. Thus, except for the two invariant sets per pair of manifolds containing the homoclinic points shown in Figures 3 and 4 which are only on the manifolds, each invariant set of homoclinic points is simultaneously on three or more invariant curves: one or more such distorted hyperbolae in the vicinity of h and the two manifolds.

6. A FEW NUMERICAL EXAMPLES

In this section, only the case of $W_o = 0.1$ is considered, except in the last paragraph. Section 5 indicated that each trapped particle is confined to a region of phase space that maps to one of a number of distorted concentric annular regions. To see the implications of such confinement, some parameters are calculated here for protons, assuming that the manifolds of h_1 and h_2 are double manifolds or remain quite close to the first parts of the manifolds.

From Figure 13 of Reference 1, $\rho = 0.933$ at a_1 and $\rho = 0.944$ at a_2 in Figure 2. By integrating proton orbits numerically,[8] starting at these points and at h_1 and h_2, using values of ρ from Reference 1 and Table 1 of Reference 3, the minimum values of r/L in Table 2 are found. The minimum value in region 1 is near h_1, that in region 2 is near a_2, and that in region 3 was then assumed to be the minimum value near h_2.

Table 2. Parameters for Three Regions in Figure 2 for $W_o = 0.1$

Region k	Min r/L	Min L	Max T
1	0.872	1.165 R_e	3.51 Gev
2	0.860	1..181	3.39
3	0.827	1.228	3.08

Assuming that the proton makes no collisions more than 100 km above the earth and taking the earth to be spherical with a radius of $R_e = 6371$ km, the minimum value of L for the proton not to scatter is (6471/6371)/(min r/L) in R_e. This gives the third column of Table 2. From Eqs. (1) and (4) and the normalization, the particle momentum is

13. Moser, J. (1956) Commun. Pure Appl. Math. 9:673.

$$p = q M W_o / L^2 . \tag{7}$$

The kinetic energy of the protons in Gev is

$$T = 0.938 \left([(p/mc)^2 + 1]^{1/2} - 1 \right) . \tag{8}$$

From these equations and the minimum values of L, taking $M = 7.98 \times 10^{10}$ Gauss km^3, the maximum kinetic energy for a proton confined to each region in Figure 2 to have a long lifetime is given in Table 2. The values would be 0.11 to 0.13 Gev lower if a minimum altitude of 200 km were assumed instead of 100 km. The solid curve in Figure 6 is T plotted against L from Eqs. (7) and (8). The three points in Table 2 are at the intersections of this curve with the dashed curves. To summarize, a proton in a dipole magnetic field with $W_o = 0.1$ whose parameters satisfy Eq. (7) and the limit in Table 2 for the region where $\rho, \dot{\rho}$ are when it crosses the equatorial plane will have a very large lifetime.

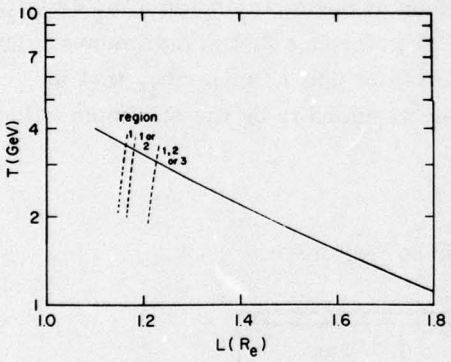


Figure 6. The Kinetic Energy of Protons for $W_o = 0.1$ as a Function of the Guiding Center Field Line Equatorial Distance in Earth Radii (solid curve). The dashed curves are sketches of the lower limits of T and L of long lifetime protons in the region to the right of the dashed curve

The (solid) curve in Figure 6 drops lower and lower the more W_o is decreased. Assuming that the corresponding manifolds remain double, corresponding limiting points on each such curve can be connected to give regions in T, L space for which particles are confined. Three such regions are shown schematically by dashed lines in Figure 6. Any other double manifolds, KAM-like curves, or limits to homoclinic manifolds farther inside of B will delineate other regions and thus dashed curves in Figure 6. A proton with T, L to the right of the left-most dashed line if it maps to the region marked in Figure 6 will have a long lifetime. In general, a proton that maps to a certain region k has a shorter and shorter lifetime the farther the T, L point is to the left of the left limiting dashed curve for that region in Figure 6. Since there are many elliptic fixed points of M, M^2, and probably of higher powers of M

with KAM curves around each one, there are many smaller parts of the above-mentioned regions to which particles must be confined.

7. CONCLUDING REMARKS

The basic principles of mapping in a plane are reviewed. The concept of an "*essentially closed*" *manifold*, an invariant curve containing a hyperbolic fixed point approaching closure, is introduced. Four theorems for a plane, single-valued mapping are stated. The last one (an invariant set of points cannot exist on both sides of an essentially closed manifold) appears to be new and useful. Three more theorems for the case of the mapping also being area-preserving and bounded are stated and proved. The last two indicate but do not prove the existence of ergodicity. Accurate computer calculation and plotting of two (unbounded) "Cremona" mappings yield results wholly consistent with the first four theorems. One of these mappings also indicates the possibility that no points exist inside one and outside of the other of two homoclinic manifolds.

The theorems are applied to the Størmer mapping, M, of De Vogelaere and Dragt for a particle trapped in a dipole magnetic field. The trapped region is found to be divisible into many sub-regions into any one of which a particle is forever confined. If there are no homoclinic points, the boundaries of these regions may be accurately determined relatively easily. If there are homoclinic points, the boundaries are unknown. Reason indicates, though, the existence of at least two KAM-like curves surrounding $\rho_o, \dot{\rho}_o$, the point from which a particle goes to the dipole, between each pair of hyperbolic fixed points of M (at $\dot{\rho} = o$) and their manifolds. It also indicates that each of these manifolds will be contained in a narrow distorted oval annular ring around $\rho_o, \dot{\rho}_o$.

Dragt and Finn's I is discussed, a series in dipolar coordinates, which must be an integral of motion in the Størmer case if it and a similar series for the Hamiltonian converge. They give a reason for its non-convergence and the existence of homoclinic points. If their conclusions are correct, it should be possible to find homoclinic points directly. A method of doing so is outlined and should be carried out. Two other possible interpretations of Dragt and Finn's results are given; both of them include the non-existence of homoclinic points on the outermost manifolds. I cannot be a global integral of motion. However, reason indicates that it should converge near the boundary B and therefore be an integral of motion for particles mirroring near the equator. This is consistent with a recent finding of Braun and Moser that KAM-like curves exist in the region near B. In some regions, I may be a valid expansion of an integral of motion which also hold in other regions, for example, on the KAM-like curves mentioned in the last paragraph above. There

must be other (regional) integrals of motion, for example for closed invariant (KAM) curves around the elliptic fixed points.

In the dipole approximation, the results of this work apply to geomagnetically trapped electrons, protons, and heavier particles. Some numerical results are given for the upper limits of kinetic energy T and lower limits of guiding field line distance L for each region of confinement for trapped protons to avoid atmospheric scattering. They assume the normalized speed $W_o = 0.1$ and that the manifolds of the outer two hyperbolic fixed points of M (at $\dot{p} = o$), h_1 and h_2, remain very close to the first part of the manifolds. Protons within these limits then must have long lifetimes. The upper limits of T and lower limits of L for long lifetime for $W_o < 0.1$, the case for geomagnetically trapped protons, are also given qualitatively and how to determine them quantitatively. For $W_o << 0.1$, the results should reduce to those from adiabatic theory since there are so many hyperbolic fixed points of M.

The basic properties of trapping in a dipole field, for example, whether or not homoclinic points exist, probably carry over to the real geomagnetic field. Though more complicated, the effects of non-dipole parts of the geomagnetic field and perhaps of electric fields on the lifetime of radiation belt particles should be investigated.

References

1. Dragt, A.J., and Finn, J.M. (1976) J. Geophys. Res. 81:2327.
2. Dragt, A.J. (1965) Rev. Geophys. 3:255.
3. Dragt, A.J., and Finn, J.M. Stability of Charged Particle Motion in a Magnetic Dipole Field: A Preliminary Report of Numerical Studies, unpublished.
4. Dragt, A.J. private communication.
5. Dubs, C.W. Cremona Mapping, unpublished.
6. Contopoulos, G., and Vlahos, L. (1975) J. Math. Phys. 16:1469.
7. Finn, J.M. (1974) Ph.D. thesis, Univ. of Md., College Park, Md.
8. McCracken, K.G., Rao, U.R., and Shea, M.A. (1962) M.I.T. Tech. Rep. 77, NYO-2760. Shea, M.A., Smart, D.F., and Carmichal, H. (1976) AFGL-TR-76-0115. Modified for a dipole field by D.F. Smart.
9. Kolmogorov, A.N. (1954) Dokl. Akad. Nauk. 98(No. 4):527.
10. Arnold, V.I. (1963) Usp. Math. Nauk. 18(No. 5):13.
11. Moser, J. (1962) Nachr. Akad. Wiss. Göttinger (No. 1).
12. Braun, M. Mathematical Remarks on the Van Allen Radiation Belt: A Survey of Old and New Results, to be published.
13. Moser, J. (1956) Commun. Pure Appl. Math. 9:673.

Appendix A

Outline of a Method for Testing Directly for Homoclinic Points

Determine quite accurately the location of an h_i and its two eigenvectors, $\vec{v}_1$ for the unstable manifolds and $\vec{v}_2$ for the stable manifolds. Choose three evenly spaced points P_1, P_2, and P_3 on $\vec{v}_1$, all on the same side of and close enough to h_i to be accurately on one of the unstable manifolds. The distance between adjacent points is chosen much smaller than that between P_j and MP_j. Then map the points to $P_j' = M^n P_j$, $j = 1, 2, 3$, n large enough for the points to return to the other side of and close to h_i if its manifolds were double manifolds. If there are no homoclinic points, P_j' for all three values of j will be accurately on $\vec{v}_2$. If there are homoclinic points, P_j' for the three values of j will be on a line (close to and) parallel to $\vec{v}_1$ or will not be close to h_i. In the latter case, $P_j'' = M^{-n} P_j$ should be calculated to be sure that it is on $\vec{v}_1$ on the same side of and closer to h_i than P_j. To avoid the latter case, P_2 could be chosen also to be one of the points of the invariant set which includes a, the point (other than h_i) at $\dot{\rho} = o$. With this choice, the criterion of whether or not the slope of the manifold is infinite at a may also be used. If $a = M^i P_2$, then the slope of the line formed by $M^i P_j$ is determined. If it deviates from the vertical, the same method except starting on $\vec{v}_2$ and changing the sign of i could be repeated to see if a slope equal in magnitude but opposite in sign is obtained.

Printed by
United States Air Force
Hanscom AFB, Mass. 01731