0 0 0 0 4 8 0 0 8 3 7

UC-32
UC-34d

LBL-6708

c1

STATISTICAL STUDY OF APPROXIMATIONS TO TWO DIMENSIONAL INVISCID TURBULENCE

Harland M. Glaz
(Ph. D. thesis)

RECEIVED
LAWRENCE
BERKELEY LABORATORY

DEC 12 1977

LIBRARY AND
DOCUMENTS SECTION

September 1977

Prepared for the U. S. Department of Energy
under Contract W-7405-ENG-48

For Reference

Not to be taken from this room

LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY

LBL-6708
c1

LEGAL NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

0 0 0 3 4 8 0 3 8 3 8

TABLE OF CONTENTS

Abstract v

Acknowledgments viii

Introduction 1

1. The Equations of Interest 6

2. The Fourier Mode Model as an Approximation to a Real Physical Flow 21

3. Review of Relevant Results from Ergodic Theory and the Theory of Dynamical Systems 26

4. The Relationship of Statistical Properties to the Existence of Coherent Structures in Two-Dimensional Flows . . 34

5. The Methodology Employed to Numerically Study Ergodicity and Mixing 53

6. Previous Results by Other Authors 69

7. Numerical Results 76

8. Conclusions 87

9. References 90

Table 97

Figure Captions 98

Figures 100

0 0 0 0 4 8 0 0 8 3 9

STATISTICAL STUDY OF APPROXIMATIONS TO TWO DIMENSIONAL INVISCID TURBULENCE

Harland M. Glaz

(Ph. D. Thesis)

Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720

ABSTRACT

A numerical technique is developed for studying the ergodic and mixing hypotheses for the dynamical systems arising from the truncated Fourier transformed two-dimensional inviscid Navier-Stokes equations.

It is commonly assumed a priori in the literature that these dynamical systems are mixing (and, therefore, ergodic); this assumption justifies the interchange of ensemble averages for time averages in the subsequent analysis. However, the phenomenon of macroscopic vortex formation is well documented experimentally; the phase space observables corresponding to large-scale vortex structure would then have time averages different from ensemble averages. This implies nonergodicity. We present two arguments which can explain macroscopic vortices. The first is statistical mechanical and, although it does not theoretically require mixing or ergodicity, the arguments are made more plausible by using these assumptions. The second argument is dynamical in nature; it hypothesizes either a flow invariant set around the macroscopic vortices or that the set of macroscopic vortices is an attractor for the flow. Both cases require nonergodicity of the underlying flow. We show, however, that these arguments are logically flawed in the inviscid case because of violations of the conservation of measure.

In our numerical study of the ergodic hypothesis, we have employed a technique involving the comparison of Cesàro sums with ensemble averages. In particular, we have developed a method for evaluating ensemble averages for a wide class of observables. The mixing hypothesis is tested using similar methods.

The numerical results to be presented illustrate two possible errors in previous numerical work on these questions. First we have observed that energy and enstropy conservation is not sufficient to ensure that numerical trajectories remain close to actual trajectories. Thus, studies using only this criterion for accuracy may be comparing two ensemble averages rather than a time average and an ensemble average. The results so obtained would indicate ergodicity, but are clearly unreliable. Second, the convergence rates of Cesàro sums are strongly dependent on the observables being tested. Previous work has often been restricted to the study of energy and vorticity spectra, but results for these observables are not sufficient to decide the question.

We have applied our techniques to three very small truncations, one of which is known to be nonergodic. In addition, we have studied another dynamical system known a priori to be random and probably mixing. On the basis of the numerical evidence alone, it is possible to conclude that this latter system is mixing, that the nonergodic truncation is indeed nonergodic, and that one of the other truncations is nonergodic. The results for the third truncation are inconclusive as are the mixing tests for all three truncations. However, we show that these questions can probably be settled with further computing effort.

0 0 0 0 4 8 0 0 8 4 0

It is our intention to apply these techniques to much larger truncations. It is shown that this course is feasible for the ergodic question with reasonable computing effort but probably is not for the mixing question (because of the question of computing time). A large truncation would, hopefully, be physically meaningful and it would be possible to study the macroscopic vortex phenomenon directly.

Work performed under the auspices of the U. S. Department of Energy.

0 0 0 0 4 8 0 0 8 4 1

ACKNOWLEDGMENTS

I would like to thank Professor Alexandre Chorin as the main source of inspiration for this work and for his unfailing confidence in its eventual conclusion.

Professor Òle Hald has also taken a genuine interest in this work and has been extremely helpful to me on numerous occasions.

Thanks are also due to Dr. Paul Concus and the Lawrence Berkeley Laboratory for their support and provision of research facilities during this work.

I would also like to thank Professor Alberto Grunbaum, Professor Gary Sod, and Phil Colella for their suggestions and help in the completion of this research.

Finally, June DeLaVergne has done a superb job typing this dissertation and merits my utmost appreciation.

0 0 0 0 4 8 0 0 8 4 2

Introduction

A considerable body of recent work has been focused on two-dimensional, incompressible, inviscid fluid mechanics. In part, this is because of the ease in applying numerical methods relative to the full three-dimensional case. Also, the equations of interest are equivalent to those used to describe a 2-D guiding-center plasma, thereby interesting a new group of workers. Finally, interesting statistical mechanical problems arise, especially due to the existence of an additional quadratic constant of the motion besides energy.

A difficult problem in this work is the construction of dynamical systems with only a finite number of degrees of freedom which faithfully model the governing partial differential equations. The most well-known method is that of finite difference approximations; however, these have the obvious defect of not being able to keep track of the small-scale motions which are critical in a turbulent flow. A different approach is the vortex model in which the fluid is approximated by a large but finite number of point vortices; the analysis proceeds either numerically, or analytically by taking the "thermodynamic limit" as the number of vortices approaches infinity. The technique which will be used in this paper is the numerical solution of a truncation of the Fourier-transformed equations. This method has the advantage of exactly conserving energy and enstrophy (i.e., total vorticity) in the inviscid case except for numerical error in solving the orindary differential equations.

It is well-known and obvious that the study of turbulence uses the concept of randomness in an essential way. Indeed, physical measurements of observables of a turbulent flow involve statistical ideas such as

averaging over ensembles and correlation functions of various kinds. It is, then, natural to extrapolate from this experience and assume that the values of observables in finite-dimensional simulations of the Navier-Stokes equations are distributed as randomly as possible - technically, this is to say that the dynamical system is mixing (for a precise definition, see below).

On the other hand, there is evidence that certain qualitative features persistently recur in 2-D turbulence. We refer in particular to the formation of macroscopic vortex structures. Since the mixing property implies that all states have equal probability, this phenomenon may indicate that the system is not mixing nor even ergodic. However, there is an explanation for large-scale coherence in the face of small-scale randomness. One may argue that the collection of states which correspond to the given large-scale behavior (macroscopic vortices) has large measure relative to the measure of the total state space. This would only be expected to hold in the limit of the state space containing a large number of independent directions (in the case at hand, this corresponds to retaining more and more Fourier modes. In the vortex model, one adds more vortices), and is analogous to the statistical mechanical derivation of equilibrium thermodynamic quantities (i.e., temperature, pressure, and entropy are all well-defined macroscopic quantities even though individual gas molecules behave as randomly as possible).

One of the objectives of the present work is to pursue in detail the ideas and arguments alluded to in the preceding two paragraphs, with a view towards understanding their consequences, both mathematical and physical. As has been noted, the concept of degree of randomness has a

0 0 0 0 4 8 0 0 8 4 3

central place in these arguments. In particular, it would be of great importance to obtain reliable evidence concerning the hypotheses of mixing and ergodicity for the Fourier mode model of 2-D homogeneous turbulence. Therefore, our other objective is (1) to discuss the feasibility and reliability of checking ergodicity and mixing for the class of dynamical systems at hand and (2) to actually investigate this question by numerical means for a few particular truncations. Although these truncations will not be large enough (i.e., not include enough Fourier modes) to be physically meaningful, we believe the results obtained will be useful in assessing the feasibility of a similar study for much larger truncations. It is less clear that the actual conclusions reached can be applied to a larger truncation, but they should be of some use in such an extrapolation, nevertheless.

In Section 1, the equations of interest for the model are derived. In particular, the dynamical system is constructed as a system of nonlinear ordinary differential equations, basic results are proved about this system, the state space is constructed and parametrized, and it is shown that a natural measure exists on this space which is also constructed in detail. This is followed in Section 2 by a discussion of how well the model can be expected to approximate a real physical flow. In addition, the stochastic element of the model is introduced and this is also compared to the situation for a real flow. We also briefly consider an alternative method due to Kraichnan for handling randomness in the Fourier mode model but this is found to be numerically unfeasible and is not considered further in the sequel.

Section 3 summarizes those ideas of ergodic theory and dynamical systems which are relevant here. Proofs of theorems are largely omitted. Ergodicity and mixing are defined, and the ergodic theorem with a few of its corollaries are stated. These results directly lead to the numerical work described in later sections. Some of the important results concerning the Baire category of certain classes of measurable maps are then introduced along with a heuristic discussion of their relevance to the problem at hand.

These concepts are then used in Section 4 where the implications of the mixing hypothesis are derived. This section outlines the possible derivations of coherent large-scale behavior-both the dynamical arguments and the statistical mechanical arguments. These ideas are then extended to the more realistic situation in which a small viscosity has been added to the equations. Also in this section, the vortex model is presented and its statistical mechanical properties are discussed.

Section 5 describes the methods employed to study the degree of randomness of the Fourier mode model. The main technique used is the comparison of Cesàro sums for selected observables with a phase space integration for the same observables. In addition, a simplified version of the 2-D hard point gas problem is introduced. The dynamical system thus obtained is intuitively as random as possible and its use here is to serve as a comparison in the study of rates of convergence in the main problem.

0 0 0 0 4 8 0 0 8 4 4

Previous results by other authors are summarized in Section 6. The numerical details of the present work are discussed in Section 7 along with an error analysis. In Section 7, we also present the results of the numerical work. In Section 8, we discuss the results, present conclusions, and indicate where further work would serve to clarify the situation.

We shall conclude in Section 4 that settling the ergodic hypothesis for the dynamical systems studied here is important in the work of constructing a theoretical explanation for macroscopic vortex formation. Our numerical work will show that it is certainly feasible to study this question numerically for large-scale truncations. This conclusion will be less clear for the mixing hypothesis although we believe it to be correct in principle. The difficulty is that a sufficiently accurate numerical test would require large amounts of computer time.

It is hoped that this thesis will make a contribution in three ways. First, the specific numerical results should be of interest. Second, the discussion of the possible hypotheses in Section 5 and the derivation of their consequences should serve to place the problem at hand in its proper perspective regarding the overall problem of deriving the existence of coherent large-scale structures from first principles. Finally, our study of the reliability of the methods employed will be of value in assessing the feasibility of future numerical studies of the degree of randomness in a dynamical system, both for large scale versions of the truncated Navier-Stokes equations and for similar problems in other fields.

1. The Equations of Interest

Notation: An underline denotes a vector. N = number of dimensions of the physical flow. The spatial coordinates are indicated by $\underline{x} = (x_1, x_2, x_3)$; the velocity and vorticity fields are denoted $\underline{u}$, $\underline{\xi} = \nabla \times \underline{u}$ respectively. In the case N = 2, $\underline{\xi}$ is a scalar and will be so denoted. Summation from 1 to N is implied whenever repeated Greek subscripts are encountered.

The starting point for our investigation is the Navier-Stokes equations:

$$(1.1)\quad \underline{u}_t + (\underline{u} \cdot \nabla)\underline{u} = -\operatorname{grad} p + \nu \nabla^2 \underline{u} + \underline{F}$$

$$(1.2)\quad \operatorname{div} \underline{u} = 0$$

where p is the pressure, $\underline{F}$ is an external force, and ν is the viscosity. Here, $\underline{u}$ is at least two-dimensional but we shall also consider a one-dimensional model equation, Burger's equation:

$$(1.3)\quad u_t + uu_x = \nu u_{xx}$$

Equation (2.1) may also be written in the vorticity transport form in case N = 2:

$$(1.4)\quad \xi_t + (\underline{u} \cdot \nabla)\xi = \nu\nabla^2\xi + \nabla \times \underline{F}$$

in which the pressure term has been eliminated (see [40]). For the most part, $\underline{F} = 0$ in the sequel and this is assumed to be the case in the following derivation.

Expanding $\underline{u}$ and p into Fourier series, we have

$$(1.5)\quad \underline{u}(\underline{x},t) = \Sigma_{\underline{k}} \underline{u}(\underline{k},t) e^{i\underline{k}\cdot\underline{x}} \quad ; \quad p(\underline{x},t) = \Sigma_{\underline{k}}\, p(\underline{k},t) e^{i\underline{k}\cdot\underline{x}} .$$

We proceed as follows to obtain a system of ordinary differential equations. First, substitute (1.5) into (1.1)-(1.2) [or (1.3)]. Then, use (1.2) to eliminate the $p(\underline{k},t)$ terms (analogously, in physical space one can show that $\nabla^2 p = -\nabla[(\underline{u}\cdot\nabla)\underline{u}]$). Finally, equate the coefficients of $e^{i\underline{k}\cdot\underline{x}}$ for each $\underline{k}$ in the resulting expression. The end result is an infinite set of ordinary differential equations in the infinite collection of variables $\{u_\alpha(\underline{k},) : \alpha = 1, \cdots, N; k_i = 0, \pm 1, \pm 2, \cdots \text{ for } i = 1, \cdots, N\}$. It reads

$$(1.6)\quad \left[\frac{\partial}{\partial t} + \nu k^2\right] u_\alpha(\underline{k},t) = -\frac{1}{2} P_{\alpha\beta\gamma}(\underline{k}) \Sigma_{\underline{p}} u_\beta(\underline{p},t) u_\gamma(\underline{k}-\underline{p},t)$$

$$(1.7)\quad k_\alpha u_\alpha(\underline{k},t) = 0$$

where

$$(1.8)\quad P_{\alpha\beta\gamma}(\underline{k}) = k_\beta P_{\alpha\gamma}(\underline{k}) + k_\gamma P_{\alpha\beta}(\underline{k})$$

$$(1.9)\quad P_{\alpha\beta}(\underline{k}) = \delta_{\alpha\beta} - k_\alpha k_\beta / k^2$$

where $k^2 = \underline{k}\cdot\underline{k} = k_1^2 + \cdots + k_N^2$. These equations are valid for $N = 2$ or $N = 3$. The analogue for Burger's equation is simply

$$(1.10)\quad \frac{d}{dt} u(k,t) = \frac{ik}{2} \Sigma_p u(p,t) u(k-p,t).$$

Observe that $\underline{u}(\underline{x},t)$ real implies that

$$(1.11)\quad \underline{u}(-\underline{k},t) = [\underline{u}(\underline{k},t)]^*$$

where z^* denotes the complex conjugate of z. Equation (1.4) also has an analogous Fourier mode representation. If we write

$$(1.12)\quad \xi(\underline{x},t) = \Sigma_{\underline{k}} \xi(\underline{k},t) e^{i\underline{k}\cdot\underline{x}}$$

it follows that

$$(1.13)\quad \left[\frac{\partial}{\partial t} + \nu k^2\right] \xi(\underline{k},t) = \frac{1}{2} \sum_{\underline{p}+\underline{q} = \underline{k}} \left(\frac{1}{q^2} - \frac{1}{p^2}\right) |\underline{p},\underline{q}|\ \xi(\underline{p},t)\ \xi(\underline{q},t)$$

where $|\underline{p},\underline{q}| = p_1 q_2 - p_2 q_1$. In case $\underline{p} = \underline{0}$ or $\underline{q} = \underline{0}$, it is understood that $\left(\frac{1}{\underline{q}^2} - \frac{1}{\underline{p}^2}\right) |\underline{p},\underline{q}| = 0$. As before,

$$(1.14)\quad \xi(-\underline{k},t) = [\xi(\underline{k},t)]^*.$$

For further details concerning the above derivation along with a discussion of the numerical analysis of the system of O.D.E.'s, see [58].

The physical flow will always be assumed to take place inside a cube of side length L with periodic boundary conditions in all directions. For convenience, we take $L = 2\pi$ which implies that the wave numbers above have integer components. It is easy to check that changing L merely scales (1.6) - (1.9) and does not change their substance.

The final step necessary to obtain the dynamical systems to be considered here is to truncate (1.6) - (1.7) [resp. (1.13)] in such a way as to respect the reality condition (1.11) [resp. (1.14)]. That is, we shall set all except a finite number of Fourier modes equal to zero for all time; but if $\underline{u}(\underline{k},t)$ is retained then so must $\underline{u}(-\underline{k},t)$. No further restrictions will be placed on the truncations so that in principle an infinite collection of dynamical systems can be obtained from the equations (1.6) - (1.7) or (1.13). We write the equations for a spherical truncation in which the retained modes are those $\underline{k}$ satisfying $|\underline{k}| < K$ (where $|\underline{k}| = \max |k_i|$) for some K -

$$(1.15)\quad \left[\frac{\partial}{\partial t} + \nu k^2\right] u_\alpha(\underline{k},t) = -\frac{i}{2} P_{\alpha\beta\gamma}(\underline{k}) \sum_{\substack{|\underline{p}|,\ |\underline{q}| < K \\ \underline{p}+\underline{q}=\underline{k}}} u_\beta(\underline{p},t) u_\gamma(\underline{q},t)$$

$$(1.16)\quad k_\alpha u_\alpha(k,t) = 0$$

$$(1.17)\quad \left[\frac{\partial}{\partial t} + \nu k^2\right] \xi(\underline{k},t) = \frac{1}{2} \sum_{\substack{|\underline{p}|,\ |\underline{q}| < K \\ \underline{p}+\underline{q}=\underline{k}}} \left(\frac{1}{q^2} - \frac{1}{p^2}\right) |\underline{p},\underline{q}| \xi(\underline{p},t)\ \xi(\underline{q},t)$$

where there is one equation for each $\underline{k}$ satisfying $|\underline{k}| < K$ and, in (1.15), one equation for each $\alpha = 1, \cdots, N$.

We now list some general results concerning the system (1.15) - (1.16). Analogous results hold for more general truncations as well as for the system (1.17).

Even though the system (1.15) - (1.16) is not Hamiltonian (to be more precise, it has not been shown to be Hamiltonian), a Liouville Theorem still holds. That is,

$$(1.18)\qquad \sum_{|\underline{k}|<K} \frac{\partial}{\partial u_\alpha(\underline{k},t)} \left(\frac{du_\alpha(\underline{k},t)}{dt}\right) = 0.$$

The proof is trivial. The significance of the result is that the dynamical system preserves Lebesgue measure in the phase space consisting of the dynamical variables $\{u_\alpha(k,t)\}$ (see [32] or [35] for a proof of this fact).

Another important result is conservation of energy -

$$(1.19) \qquad \frac{d}{dt} \Sigma_{|\underline{k}| < K} u_\alpha(\underline{k},t)\, u_\alpha(-\underline{k},t) = 0.$$

If we restrict attention to the 2-D case, there is another quadratic constant of the motion called the enstrophy or total vorticity -

$$(1.20) \qquad \frac{d}{dt} \Sigma_{|\underline{k}| < K} k^2\, u_\alpha(\underline{k},t)\, u_\alpha(-\underline{k},t) = 0.$$

The proofs of (1.19) and (1.20) are straightforward

Taking advantage of the relations $k_\alpha u_\alpha(\underline{k},t) = 0$, $u_\alpha(-\underline{k},t) = [u_\alpha(\underline{k},t)]^*$, and $\dot{u}_\alpha(\underline{0},t) = 0$, it is only necessary to retain the independent modes in any truncation (e.g., $u_\alpha(-\underline{k},t)$ and $u_\alpha(\underline{k},t)$ are not independent. The same holds for $u_1(\underline{k},t)$, $u_2(\underline{k},t)$, and $u_3(\underline{k},t)$ for any $\underline{k}$). Note that for any $\alpha,\underline{k}$, the real and imaginary parts of $u_\alpha(\underline{k},t)$ are independent. Say that there are M such independent modes and name them $\underline{\omega} = (\omega_1, \cdots \omega_M)$. Then, (1.15) - (1.16) takes the form

$$(1.21) \qquad \dot{\underline{\omega}} = \underline{F}(\underline{\omega})$$

where each component of $\underline{F}$ is a quadratic polynomial in the ω_i. Note that $\underline{F}$ and $\underline{\omega}$ now have purely real components. Using this notation, the constants of the motion and Liouville's then may be written

$$(1.22) \qquad \sum_{i=1}^{M} \alpha_i \omega_i^2 = E; \quad \sum_{i=1}^{M} \beta_i \omega_i^2 = \Omega, \quad \mathrm{div}(\dot{\underline{\omega}}) = 0$$

where E = total energy, Ω = enstrophy, and the α_i, β_i are easily determined constants. It is easily shown that the α_i, β_i are strictly positive. Observe that for any i, $1 \leq i \leq M$, there exists a j, $1 \leq j \leq M$ such that there exists $\alpha,\underline{k}$ with $u_\alpha(\underline{k},t) = \omega_i(t) + i\omega_j(t)$. This implies that

0 0 0 0 4 8 0 0 8 4 7

for any i, there exists a j such that $\alpha_i = \alpha_j$ and $\beta_i = \beta_j$ - this remark will be useful later. The formulation (1.21) - (1.22) will be used where convenient in the sequel.

As has been already noted, (1.21) has at least two constants of the motion, E and Ω. Therefore, the motion takes place on $S = S_E \cap S_\Omega$ where S_E is the surface of constant energy E and S_Ω is the surface of constant enstrophy Ω. It is trivial to show that S_E and S_Ω must be (differentiable) surfaces of dimension M-1 in $\mathbb{R}^M$ and we will show later that S is also a surface except for possible boundaries and corners. Assuming this for the moment, it is clear that $\dim S = M-2$. For any point $\underline{\omega}_0 \varepsilon S$, let $F_t(\underline{\omega}_0)$ be the solution of (1.21) evaluated at time t with the initial condition given by $\underline{\omega}_0$. From the existence and uniqueness theorems for systems of O.D.E.'s (see [30], for example), one sees that F_t is a well-defined mapping on all of S, i.e., $F_t : S \to S$. The function F_t is called the time t map for the system (1.21). Since solutions depend continuously on data, one sees that F_t is continuous (see [30] for a proof). For the applications in mind, it is important to note that there is an F_t - invariant measure (i.e., if the measure is called μ then $\mu(A) = \mu(F_t A)$ for any measurable set A) for all t defined on S which is absolutely continuous with respect to Lebesgue measure on S. The derivation of this measure for a dynamical system which satisfies Liouville's Theorem, but has only one quadratic conserved quantity (e.g., energy) may be found in [35]. We shall now generalize that proof to the case at hand.

As is customary, we shall refer to the space $\mathbb{R}^M$ with coordinates $\omega_1, \cdots, \omega_M$ as the phase space of our system. Let $dV = d\omega_1 \cdots d\omega_M$ be

M-dimensional Lebesgue measure. Then, $dV = d\Sigma dn_E dn$ where $d\Sigma$ is the (M-2)-dimensional Lebesgue measure on S and dn_E, dn are the differentials associated with two mutually perpendicular unit tangent vectors (to the entire phase space) in the normal space to TS (= tangent space of S). Here, the tangent vector associated with dn_E is in the direction of grad E. Define dn_Ω similiarly and let θ = angle between grad E and grad Ω. Let $V_{E',\Omega'}$ denote that subset of phase space for which $E \leq E'$ and $\Omega \leq \Omega'$. Then, for any function f measurable with respect to dV it follows that

$$\int_{V_{E',\Omega'}} f(P)dV = \int_{V_{E',\Omega'}} f(P)\ d\Sigma dn_E dn$$

$$= \int_{V_{E',\Omega'}} f(P)\ \frac{d\Sigma dn_E dn_\Omega}{\sin\theta}$$

$$= \int_{V_{E',\Omega'}} f(P)\ \frac{d\Sigma\ dE\ d\Omega}{\|\text{grad } E\|\ \|\text{grad } \Omega\| \sin\theta}$$

$$= \int_0^{E'} dE \int_0^{\Omega'} d\Omega \int_{\Sigma_{E,\Omega}} f(P)\ \frac{d\Sigma}{\|\text{grad } E\|\ \|\text{grad } \Omega\| \sin\theta}$$

where $\Sigma_{E,\Omega}$ is the surface of constant energy E and constant enstrophy Ω. We have used here the facts that $dE = \|\text{grad } E\|\ dn_E$, $d\Omega = \|\text{grad } \Omega\|\ dn_\Omega$, and that

$$\begin{pmatrix} dn_E \\ dn_\Omega \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \cos\theta & \sin\theta \end{pmatrix} \begin{pmatrix} dn_E \\ dn \end{pmatrix} .$$

It now follows that

$$\frac{\partial^2}{\partial E \partial \Omega} \int_{V_{E,\Omega}} f(P)dV = \int_{\Sigma_{E,\Omega}} \frac{d\Sigma}{\|\text{grad } E\| \, \|\text{grad } \Omega\| \sin \theta} .$$

The sin θ factor follows from the change of variables formula. It must also be shown that dn can be chosen consistantly on all of S (i.e., that if the vector grad Ω is between grad E and the vector associated with with dn at one point of S, then the same relationship must hold everywhere on S). To see this, it is sufficient to show that $0 < |\theta| < \frac{\pi}{2}$ (since this would imply $\theta \neq 0$, we would have $0 < \theta < \frac{\pi}{2}$ always or $0 > \theta > -\frac{\pi}{2}$ always). The fact that the dot product grad E $\cdot$ grad $\Omega = \Sigma 4\alpha_i \beta_i \omega_i^2$ is a positive definite quadratic form and that $\underline{0} \notin S$ (we always take $E, \Omega > 0$) completes the proof.

We are now in a position to define the invariant measure on S. Let $M \subseteq S$ be measurable and let χ_M denote the characteristic function of M (i.e., $\chi_M(\underline{\omega}) = 1$ if $\underline{\omega} \varepsilon M$ and is 0 otherwise). Then, by (1.22),

$$\lim_{\Delta E, \Delta\Omega \to 0} \frac{1}{\Delta E \Delta \Omega} \int_{\substack{E < E' < E + \Delta E \\ \Omega < \Omega' < \Omega + \Delta\Omega}} \chi_M(P) \, dV$$

is invariant under the flow on S. Hence, if we set

$$(1.23) \quad \mu(M) = \int_M \frac{d\Sigma}{\|\text{grad } E\| \, \|\text{grad } \Omega\| \sin \theta}$$

it follows that μ is F_t - invariant.

We turn now to the computation of $d\Sigma$ which will complete the specification of the measure μ in a numerically usable form. By a parameter map on S we shall mean a diffeomorphism $F: \mathcal{D} \to S$ (i.e.,

F is 1-1, differentiable, and has a differentiable inverse) where $\mathfrak{D} \subseteq \mathbb{R}^{N-2}$ is an open set. Note that such an F cannot be onto. We shall later explicitly construct a covering of S by the images of parameter maps of the following form:

$$(1.24) \quad F(u_1, \cdots, u_{M-2}) = [u_1, \cdots, u_{M-2}, f(\underline{u}), g(\underline{u})]$$

where $\underline{u} = (u_1, \cdots, u_{M-2}) \in \mathfrak{D}$. It is well-known that the surface element $d\Sigma$ is given by $d\Sigma = [\det (g_{ij})]^{1/2} du_1 \cdots du_{M-2}$ where (g_{ij}) is the matrix of coefficients of the Riemmanian metric on S for the parametrization given by F. The computation of the metric (g_{ij}) simply involves finding the relationship between the differentials $d\omega_i$ on S and the differentials du_i on $\mathfrak{D}$ and is straightforward:

$$\sum_{i=1}^{M} d\omega_i^2 = \sum_{i=1}^{M-2} d\omega_i^2 + d\omega_{M-1}\, d\omega_{M-1} + d\omega_M\, d\omega_M$$

$$= \sum_{i=1}^{M-2} du_i^2 + \sum_{i,j=1}^{M-2} \left(\frac{\partial f}{\partial u_i} \frac{\partial f}{\partial u_j} + \frac{\partial g}{\partial u_i} \frac{\partial g}{\partial u_j} \right) du_i du_j .$$

We have used only the chain rule here. It then follows that

$$1.25) \quad g_{ij} = \delta_{ij} + \frac{\partial f}{\partial u_i} \frac{\partial f}{\partial u_j} + \frac{\partial g}{\partial u_i} \frac{\partial g}{\partial u_j} .$$

The determinant turns out to be quite easy to evaluate and we obtain

$$1.26) \quad \det (g_{ij}) = 1 + \sum_{i=1}^{M-2} \left[\left(\frac{\partial f}{\partial u_i} \right)^2 + \left(\frac{\partial g}{\partial u_i} \right)^2 \right] + \sum_{i<j} \begin{vmatrix} f_i & f_j \\ g_i & g_j \end{vmatrix}^2$$

where $f_i = \partial f/\partial u_i$ and $g_i = \partial g/\partial u_i$. We remark that an alternative formula for det (g_{ij}) is derived in [15]. It is

$$(1.27) \quad \det (g_{ij}) = D_1^2 + \cdots + D^2_{\binom{M}{M-2}}$$

where the D_i range over all $\binom{M}{M-2}$ subdeterminants of the Jacobian matrix $\left(\frac{\partial \omega_i}{\partial u_j}\right)_{\substack{i=1,\cdots, M \\ j=1,\cdots, M-2}}$ of order M-2. It is easy to see that (1.27) leads directly to our result (1.26).

We now proceed to construct the parametrizations used above and, at the same time, prove that the surface S is (topologically) connected. Regarding the latter objective, it may be noted that the intersection of two-dimensional ellipsoids in $\mathbb{R}^3$ will almost never be connected. In higher dimensions, it is always true that the intersection of two topological spheres (i.e., a surface homoemorphic to a sphere. This would include an ellipsoid but exclude a torus, for example) must be a cross product of topological spheres. However, the degenerate sphere $S^\circ = \{-1, +1\}$ is not eliminated a priori as one of the terms in the product and if it is included the product would have two connected components. Therefore, a proof of connectivity is required. The reason for its inclusion is that a flow on a disconnected surface is trivially nonergodic. This will be clear upon reading Section 3. To begin, we introduce some useful notation. Renumbering the coordinates if necessary, assume that

$$(1.28) \quad \alpha_M/\beta_M \leq \alpha_i/\beta_i \leq \alpha_{M-1}/\beta_{M-1}, \quad \forall i = 1, \cdots, M.$$

Therefore, we may as well assume that

$$(1.29)\quad \alpha_M/\beta_M \leq E/\Omega \leq \alpha_{M-1}/\beta_{M-1}$$

as well, since if not it would follow from (1.22) that S is empty. Note that if one of the extremes occurs in (1.29) then most of the dynamical variables would be forced to be zero. Indeed, let $I = \{i: \alpha_i = \alpha_M$ and $\beta_i = \beta_M\}$ and suppose that $E/\Omega = \alpha_M/\beta_M$. Then, the energy and vorticity surface reduce to

$$(1.30)\quad \Sigma_{i\varepsilon I}\, \alpha_i\omega_i^2 = E, \quad \Sigma_{i\varepsilon I}\, \beta_i\omega_i^2 = \Omega, \ \omega_i = 0 \quad \forall\, i\notin I.$$

Since $\alpha_i/E = \beta_i/\Omega$ for any $i\varepsilon I$, it follows that these two surfaces are identical. It has been previously remarked that the set I contains at least two elements which shows that $S = S_E \cap S_\Omega = S_E$ is connected in this degenerate case. The analysis for the other extreme proceeds simiiiarly. In the numerical work to be presented below, we shall only consider the generic case in which all dynamical variables play a role.

Define

$$(1.31)\quad c_i = \alpha_i\beta_M - \alpha_M\beta_i \ ; \quad d_i = \alpha_{M-1}\beta_i - \alpha_i\beta_{M-1}, \ i = 1, \cdots, M-2$$

and observe that (1.28) implies that $c_i, d_i \geq 0$. It is important to note that $\exists\, j,\ell$ such that $c_j = 0$, $d_\ell = 0$ but that there are no m such that $c_m = d_m = 0$. This follows from (1.19)-(1.22) and the remark following (1.22).

Let us use the coordinates $(\omega_1, \cdots, \omega_{M-2})$ to parametrize S. The subsequent abuse of notation should be obvious. Let $\mathcal{D} = \{(\omega_1, \cdots, \omega_{M-2}) \varepsilon\, \mathbb{R}^{M-2} \mid \exists\, \omega_{M-1}, \omega_M$ with $\underline{\omega}\varepsilon S\}$. If $(\omega_1, \cdots, \omega_{M-2}) \varepsilon\, \mathcal{D}$, then (ω_{M-1}, ω_M) must satisfy the system

$$(1.32) \qquad \alpha_{M-1}\omega_{M-1}^2 + \alpha_M\omega_M^2 = E - E'$$

$$\alpha_{M-1}\omega_{M-1}^2 + \beta_M\omega_M^2 = \Omega-\Omega'$$

where $E' = \sum_{i=1}^{M-2} \alpha_i \omega_i^2 \leq E$ and $\Omega' = \sum_{i=1}^{M-2} \beta_i\omega_i^2 \leq \Omega$. Let

$$(1.33) \quad D = \alpha_{M-1}\beta_M - \alpha_M\beta_{M-1}$$

and observe that $D > 0$ by (1.28). The solution of (1.32) is

$$(1.34) \quad \omega_{M-1}^2 = D^{-1}\,[(\beta_M E - \alpha_M\Omega) + (\alpha_M\Omega' - \beta_M E')]$$

$$\omega_M^2 = D^{-1}\,[(\alpha_{M-1}\Omega - \beta_{M-1}E) + (\beta_{M-1}E' - \alpha_{M-1}\Omega')]$$

which is equivalent to

$$(1.35) \quad \omega_{M-1}^2 = D^{-1}[(\beta_M E - \alpha_M\Omega) - \sum_{i=1}^{M-2} c_i\omega_i^2]$$

$$\omega_M^2 = D^{-1}\,[(\alpha_{M-1}\Omega - \beta_{M-1}E) - \sum_{i=1}^{M-2} d_i\omega_i^2]\ .$$

Since c_i, d_i, $\beta_M E-\alpha_M\Omega$, $\alpha_{M-1}\Omega-\beta_{M-1}E \geq 0$ and necessarily ω_{M-1}^2, $\omega_M^2 \geq 0$, it follows that $(\omega_1, \ldots, \omega_{M-2}) \in D$ implies that

$$(1.36) \qquad \sum_{i=1}^{M-2} c_i\omega_i^2 \leq \beta_M E - \alpha_M \Omega$$

$$\sum_{i=1}^{M-2} d_i\omega_i^2 \leq \alpha_{M-1}\Omega - \beta_{M-1}E$$

$$\sum_{i=1}^{M-2} \alpha_i\omega_i^2 \leq E$$

$$\sum_{i=1}^{M-2} \beta_i\omega_i^2 \leq \Omega$$

and the reverse implication follows easily be reversing the steps in the argument.

Therefore, S may be viewed as being made up of four disjoint copies of $F(\mathbb{D})$ plus boundaries. Here, $\mathbb{D}$ is given by (1.36), F is given by (1.24) and (1.34), and the four separate copies of $F(\mathbb{D})$ arise from taking different signs upon taking square roots in (1.34). That is, S contains one copy of $F(\mathbb{D})$ for each choice of signs in $(\pm\omega_{M-1}, \pm\omega_M)$. If $\partial\mathbb{D}$ = the boundary of $\mathbb{D}$, then it is easy to see that the four copies of $F(\mathbb{D})$ are connected on $F(\partial\mathbb{D}) = \{(\omega_1, \cdots, \omega_{M-2}) \mid \omega_{M-1} = 0 \text{ or } \omega_M = 0\}$. From (2.36), $\mathbb{D} = A \cap B \cap C \cap D$ where A and B are solid (M-2)-dimensional ellipsoids, $C = E^K \times \mathbb{R}^{M-2-K}$, and $D = E^L \times \mathbb{R}^{M-2-L}$ where $E^K(E^L)$ is a solid K(L)-dimensional ellipsoid (recall that some of the $c_i, d_j = 0$ and therefore have no effect in (1.36)). Clearly, all of A,B,C,D are convex and connected which shows that D has the same properties. This completes the proof that S is connected.

Let us note that by simply reversing the roles of the coordinates, we can obtain a new parameter map F' and a new domain $\mathcal{D}'$ of the same form as (1.24) and (1.36) (to see this, recall that $\exists k \neq M$ and $\exists \ell \neq M-1$ such that $a_k = a_M$, $b_k = b_M$, $\alpha_\ell = \alpha_{M-1}$, and $\beta_\ell = \beta_{M-1}$) such that $F(\partial\mathcal{D})$ will be in the interior of $F'(\mathcal{D}')$. This shows that S is, in fact, smooth and that the flow of the dynamical system will not be obstructed by nonsmooth boundaries.

With the parametrization given by (1.28) - (1.36) we are now in a position to explicitly exhibit the measure $d\mu$ which is given formally by (1.23), (1.26). Making use of (1.34), it is straightforward to show that

$$(1.37) \qquad |g_{ij}(\underline{\omega})| = 1 + \sum_{i=1}^{M-2} \left(\frac{c_i^2\omega_i^2}{\omega_{M-1}^2} + \frac{d_i^2\omega_i^2}{\omega_M^2} \right) + \sum_{1\leq i\leq j\leq M-2} (c_i d_j - c_j d_i)^2 \frac{\omega_i^2\omega_j^2}{\omega_{M-1}^2\omega_M^2}$$

$$\|\text{grad } E\|^2 = \sum_{i=1}^{M} 4\alpha_i^2\omega_i^2$$

$$\|\text{grad } \Omega\|^2 = \sum_{i=1}^{M} 4\beta_i^2\omega_i^2$$

$$\sin\theta = 1 - \left[\left(\sum_{i=1}^{M} 4\alpha_i\beta_i\omega_i^2 \right)^2 \Big/ \left(\sum_{i=1}^{M} 4\alpha_i^2\omega_i^2 \right) \cdot \left(\sum_{i=1}^{M} 4\beta_i^2\omega_i^2 \right) \right]$$

where ω_{M-1}, ω_M are given in terms of $\underline{\omega} = (\omega_1, \cdots, \omega_{M-2})$ by equation (1.35). It now follows that

$$(1.38)\quad d\mu(\underline{\omega}) = \left[\sqrt{1 + \frac{A}{\omega_{M-1}^2} + \frac{B}{\omega_M^2} + \frac{AB-C^2}{\omega_{M-1}^2\omega_M^2}} \Big/ 4\sqrt{EF-G^2}\right] d\underline{\omega}$$

where

$$(1.39)\quad A = \sum_{i=1}^{M-2} c_i^2\omega_i^2 \;;\quad B = \sum_{i=1}^{M-2} d_i^2\omega_i^2 \;;\quad C = \sum_{i=1}^{M-2} c_i d_i \omega_i^2$$

$$E = \sum_{i=1}^{M} \alpha_i^2\omega_i^2 \;;\quad F = \sum_{i=1}^{M} \beta_i^2\omega_i^2 \;;\quad G = \sum_{i=1}^{M} \alpha_i\beta_i\omega_i^2$$

The formulae (1.38), (1.39) were used for the numerical computation of $d\mu(\underline{\omega})$. We remark that the number of multiplications necessary for the evaluation is approximately 6M plus two square roots. Thus, it is not an obstacle to the extension of the results reported here to large M. It is trivial to see that $d\mu$ is a positive measure; also, $\sin\theta \neq 0$ implies that $EF-G^2 > 0$ which shows that $d\mu$ is absolutely continuous with respect to Lebesgue measure as claimed.

0 0 0 0 4 8 0 0 6 5 2

2. The Fourier Mode Model as an Approximation to a Real Physical Flow

In this section we will consider how well a truncation of the form (1.15) - (1.16) with $\nu = 0$ can be expected to approximate a real physical flow given by (1.1) - (1.2). We will assume that the Navier-Stokes equations accurately reflect the physics of fluid mechanics. The system (1.17) is equivalent to (1.15) - (1.16) and the discussion applies for it as well. Only turbulent flows are considered so that $0 < \nu << 1$.

There are two difficulties: first, in a real turbulent flow, ν is small but different from zero, and second, not all modes are retained in the truncation. We shall also discuss the use of periodic boundary conditions. Finally, we shall construct the stochastic element of the model as used later in the applications, point out an alternative due to Kraichnan, and discuss both in relation to the stochastic properties of (1.1) - (1.2).

A heuristic argument due to Batchelor (see [2]) indicates that for a 3-D flow the limit for $\nu \to 0$ is singular in the sense that a finite rate of energy dissipation persists in the limit whereas energy is exactly conserved in (1.15) - (1.16) with $\nu = 0$ for any number of spatial dimensions. The argument requires that all Fourier modes be present (i.e., no truncation) so that if his result is correct, it would argue against using (1.15) - (1.16) as a model for three-dimensional turbulence. However, for Burger's equation (1.3) and for the two-dimensional version of (1.1) - (1.2) with periodic boundary conditions it can be shown that solutions of the equations with positive viscosity approach the solutions of the inviscid equations. Furthermore, the convergence

is uniform on compacta for the Fourier transforms of the solutions. (Definition: Let $f_n, f: \mathbb{R}^2 \to \mathbb{R}$ be given. Then, we say that $f_n \to f$ uniformly on compacta if and only if, for every compact set $K \subseteq \mathbb{R}^2$, the restrictions of f_n to K converge uniformly to the restriction of f to K. The same definition holds with $\mathbb{R}$ in place of $\mathbb{R}^2$). Proofs of these results may be found in [31] for Burger's equations and in [17] for the 2-D Navier-Stokes equation with periodic boundary conditions.

The problem with the truncation is not so much a difficulty in theory as it is in practice. In principle, it should be possible to obtain an excellent approximation to (1.1) - (1.2) using (1.15) - (1.16) by retaining all modes with wave numbers less than that used to describe, say, the intermolecular distance for the fluid being studied. This is obviously impractical. The real question is whether the chosen truncation can provide a reasonable approximation. This question is obviously not well-posed: it depends on the definition of "reasonable", the problem under study, and the quantities or measurements desired. We have been unable to resolve this problem, but we hope that the Fourier mode representation of two-dimensional turbulence with zero viscoity will given an accurate account of at least some aspects of turbulence.

The question of boundary conditions is somewhat more complicated. The reason for choosing periodic boundary conditions is that the analysis is simpler than it would have been for, say, Dirichlet boundary conditions. On the other hand, the boundary layer phenomena which is

characteristic of a real physical flow is eliminated. This, in turn, eliminates vorticity creation at the boundaries, see [3], [10], [11], [40] for a discussion of these mechanisms. The advantages of using a periodic cube are twofold: one, the solution can be expanded in a convenient orthogonal system $\{e^{i\underline{k}\cdot\underline{x}}\}$ which possesses an exact and finite multiplication rule. That is, $e^{i\underline{k}\cdot\underline{x}} \cdot e^{i\underline{\ell}\cdot\underline{x}} = e^{i(\underline{k}+\underline{\ell})\cdot\underline{x}}$. For more general boundary conditions and geometries, the resulting expansion of the solution might be in terms of functions, e.g. Bessel functions, either having no multiplication rule at all or a rather complicated one involving infinite sums. This would require a further approximation in the numerical work. The second advantage is that the statistics of the solution will not depend on spatial position if the initial conditions are chosen statistically homogeneous. The use of the homogeneity assumption is both convenient and standard in the turbulence literature. It is hoped that an understanding of this case will be useful in a more general theory.

A point of difficulty in the theory of turbulence is the method of introducing randomness into the mathematical models. Let us suppose that a physical flow is taking place in a region $\mathcal{D}$ and that we have performed a series of physical experiments which are so thorough that we have knowledge of all n-point correlation functions of the variables $\underline{u}(\underline{x}_i), i = 1, \cdots, n$ where n and $\underline{x}_i \in \mathcal{D}$ are arbitrary. Then, a detailed characterization of the flow's stochastic properties would be obtained by constructing a probability space $(\Omega, \mathcal{B}, \lambda)$ such that each $\underline{u}(\underline{x}_i) = \underline{u}(\underline{x}_i, \omega)$, $\omega \in \Omega$ was a random variable and such that the joint distribution of the $\underline{u}(\underline{x}_i, \omega)$ with respect to the measure λ was identical to the

measured correlations. It would then be expected that the velocity fields $\underline{u}(\cdot, \omega)$, ω fixed, would satisfy the Navier-Stokes equations at least almost everywhere with respect to λ. Of course, the space Ω would be complicated and the measure λ would be even more so. In fact, Ω would generally turn out to be a function space whose elements have domain $\mathcal{D}$ and, therefore, would be infinite-dimensional. This viewpoint in which $\underline{u}$ is taken to be a random field is discussed in Chorin's book [10] along with the further development of stochastic integral representations of the velocity and vorticity fields.

Hence, complete knowledge of $(\Omega,\mathcal{B},\lambda)$ for a particular domain $\mathcal{D}$ would be equivalent to knowing everything about statistical properties of solutions of the Navier-Stokes equations in $\mathcal{D}$. Since we do not know either $(\Omega,\mathcal{B},\lambda)$ or the statistical properites of the solutions, it is necessary to make a choice concerning randomness properties of the model, i.e., we must approximate $(\Omega,\mathcal{B},\lambda)$. Our choice is a quite simple one. Since $\nu = 0$, we take (Ω,λ) to be the space (S,μ) constructed in Section 2 with $\mathcal{B}$ = Lebesgue measurable sets in S and μ the F_t-invariant measure. Observe that if $\mathcal{J}$ is the set of velocity fields in $L^2(\mathcal{D})$, then S may be viewed as the projection of $\mathcal{J}$ onto a finite-dimensional subspace which consists of the modes retained in a given truncation. This must be followed by a further projection onto the energy and enstrophy surfaces. It may seem unnatural to expect an approximation of this sort to be physically meaningful. However, a recent paper of Foias [18], indicates, but does not prove, that the measure λ must be concentrated on a finite-dimensional subspace of Ω provided λ is invariant under the flow induced by the Navier-Stokes equations. If $E \subseteq \Omega$ is the given subspace, this

means that $\lambda(X) = \lambda(X \cap E)$, for all $X \in \mathcal{B}$. Note, however, that the subspace of this conjecture need not necessarily have any relationship to our collection of truncated Fourier modes. Of course, the subspace E could be approximated arbitrarily well by a large enough Fourier mode truncation.

In summary then, (S,μ) is the ensemble of flows for our model. That is, initial data is chosen randomly from S with respect to Lebesgue measure and weighted according to $d\mu$. These points then move on S in a completely nonrandom manner which is given by (1.21). Hence, all randomness is relegated to the initial conditions.

We conclude this section by presenting an alternative method, due to Kraichnan, for handling randomness. See [38] for further details. Kraichnan replaces the expression for $\dot{\underline{u}}_\alpha(\underline{x},t)$ in equation (1.6) by

$$(2.1)\quad [\frac{d}{dt} + \nu k^2]u^i_\alpha(\underline{k},t) = -\frac{1}{2}M^{-\frac{1}{2}}P_{\alpha\beta\gamma}(\underline{k})\Sigma_{\underline{p}}[\Sigma_j \phi_{i,j,i-j}u^j_\beta(p,t)u^{i-j}_\gamma(\underline{k}-\underline{p},t)].$$

This expression involves the entire ensemble at once. The ensemble has been approximated by the M points $\underline{u}^1, \cdots, \underline{u}^M$ and the coefficients $\phi_{i,j,i-j}$ are constants and chosen such that inviscid energy and enstrophy conservation are maintained. The manner of selection of these initial points is still undetermined and corresponds to a choice of $(\Omega,\mathcal{B},\lambda)$ as above. Note, however, that the evolution of a single initial point is dependent on the ensemble and the particular realization of the ensemble (i.e., choice of M, $\underline{u}^1, \cdots, \underline{u}^M$). It would be of interest to compare the statistical properties of Kraichnan's model (2.1) with those of the Fourier mode model but it is easy to see that the additional summation (over the index j) in (2.1) makes this numerically infeasible even for small M.

3. Review of Relevant Results from Ergodic Theory and the Theory of Dynamical Systems

In this section we present some definitions and theorems from ergodic theory and dynamical systems which will be used in the following sections. The main references for this section are [26] and [67].

Let $(M,\mathcal{B},\mu)$ denote a measure space such that $\mu(M) < \infty$. We wish to study the asymptotic behavior of measurable transformations on M. Two cases are of interest here. In the first, we let $T: M \to M$ be a measurable map and consider the collection of iterates $\{T^n\}_{n \geq 0}$. In the second case, $T_t: M \to M$ is measurable for each $t \in \mathbb{R}_+$ and we require $T_s \circ T_t = T_{s+t}$. The collection $\{T_t\}_{t \geq 0}$ is a flow on M. These are both cases of the more general situation in which one is given a semi-group G of measurable transformations on M. Here, $G = \mathbb{Z}_+$ and $G = \mathbb{R}_+$

We now define three statistical properties that are of interest in the situation just described.

The first of these is existence of a finite invariant measure. Since $\mu(M) < \infty$, we say that μ is invariant with respect to T or $\{T_t\}$ if for all $A \in \mathcal{B}$, $\mu(A) = \mu(T^n A)$, for all $n \geq 0$ or $u(A) = \mu(T_t A)$, for all $t \geq 0$, respectively. The following result then holds:

<u>Thm</u>. For every measurable integrable function $f: M \to \mathbb{R}$, we have

$$(3.1) \qquad \int_M f(x)d\mu(x) = \int_M f(T^n x)d\mu(x), \ \forall n \geq 0 \qquad (G = \mathbb{Z}_+),$$

$$(3.2) \qquad \int_M f(x)d\mu(x) = \int_M f(T_t x)d\mu(x), \ \forall t \geq 0 \qquad (G = \mathbb{R}_+).$$

<u>Pf</u>. These equations are trivial in case f is a characteristic function and the general results follow by linearity and the fact that any measureable f can be approximated arbitrarily closely in $L^1(M)$ by

some finite sum of characteristic functions. The existence of a finite invariant measure is enough to obtain the following:

Ergodic Thm. Let $f \varepsilon L^1(M,\mu)$ and suppose that μ is a finite invariant measure. Then the following limits exist a.e. (μ):

$$(3.3) \quad \hat{f}(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) \qquad (G = \mathbb{Z}_+)$$

$$(3.4) \quad \hat{f}(x) = \lim_{S\to\infty} \frac{1}{S} \int_0^S f(T_t x)\,dt \quad (G = \mathbb{R}_+)$$

and $\int_M f(x)\,d\mu(x) = \int_M \hat{f}(x)\,d\mu(x)$.

Pf. See [26].

The sum on the right-hand side of (3.3) is called a Cesàro sum. Note that (3.3) and (3.4) need hold only almost everywhere. A related theorem, which is much easier to prove, is the:

Poincaré Recurrence Thm. Let $A \varepsilon \mathcal{B}$ with $\mu(A) > 0$ and suppose that μ is invariant. Then, for a.e. (μ) $x \varepsilon A$ there exists an increasing sequence of positive integers n_i (or positive real numbers t_i) such that $T^{n_i} x \varepsilon A$ $(T_{t_i} x \varepsilon A)$ in case $G = \mathbb{Z}_+$ $(G = \mathbb{R}_+)$.

Pf. See [26] or [67] for the discrete case $G = \mathbb{Z}_+$. The continuous case $G = \mathbb{R}_+$ can be derived as a corollary by letting $T = T_1$. Suppose now that M has a topological structure which is compatible with the measure μ in the sense that every open set has positive measure. Then the above theorem would say that every point of M returns arbitrarily close to itself. This does not imply that $\{T^n\}$ (or $\{T_t\}$) is periodic or even almost periodic since the sequences $\{n_i\}$ and $\{t_i\}$ need not be the same for different points of M.

The second property is ergodicity. A transformation $T: M \to M$ is said to be ergodic if

$$(3.5) \quad \mu(TA-A) = 0 \Rightarrow \mu(A) = 0 \quad \text{or} \quad \mu(A) = \mu(M), \ \forall A \in \mathcal{B}.$$

A flow $\{T_t\}$ is ergodic if and only if each $T_t: M \to M$ is ergodic. The following results hold:

Thm. T is ergodic if and only if every measurable invariant function [i.e., $f(Tx) = f(x)$ a.e.(μ)] is a constant.

Pf. See [26].

Thm. The limit function f in the ergodic theorem is a constant. Indeed,

$$(3.6) \quad \hat{f}(x) = \frac{1}{\mu(M)} \int_M f(x)d\mu(x)$$

Pf. See [26].

We can therefore rewrite (3.3) and (3.4) for the ergodic case:

$$(3.7) \quad \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) = \frac{1}{\mu(M)} \int_M f(x)d\mu(x)$$

$$(3.8) \quad \lim_{S\to\infty} \frac{1}{S} \int_0^S f(T_t x)\, dt = \frac{1}{\mu(M)} \int_M f(x)d\mu(x).$$

Thus, T is ergodic if and only if the time mean is equal to the phase mean for all integrable observables. The converse of the above theorem holds as well; that is, if (3.7), respectively (3.8), holds then T, respectively $\{T_t\}$, is ergodic. Finally, we note the following characterization of ergodicity:

0 0 0 0 4 8 0 0 3 5 6

Thm. A map $T:M \to M$ is ergodic if and only if for all $f,g \in L^2(M,\mu)$

$$(3.9)\quad \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \int_M f(T^j x)g(x)d\mu(x) = \frac{1}{\mu(M)}\int_M f(x)d\mu(x)\cdot\int_M g(x)d\mu(x).$$

For an ergodic flow $\{T_t\}$

$$(3.10)\quad \lim_{S\to\infty} \frac{1}{S}\int_0^S \int_M f(T_t x)g(x)d\mu(x)dt = \frac{1}{\mu(M)}\int_M f(x)d\mu(x)\cdot\int_M g(x)d\mu(x).$$

Pf. (3.9) is derived in [26] and (3.10) follows from (3.9) by taking $T = T_\varepsilon$, $\varepsilon > 0$ and letting $\varepsilon \to 0$. Then, (3.9) becomes an approximating Riemann sum for the integral in (3.10).

Let $f = \chi_F$, $g = \chi_G$ be characteristic functions. That is, $f(x) = 1$ if $x \in F$ and $f(x) = 0$ otherwise. Equation (3.10) becomes

$$(3.11)\quad \lim_{S\to\infty} \frac{1}{S}\int_0^S \mu(T_{-t}F\cap G)\,dt = \frac{\mu(F)\mu(G)}{\mu(M)}$$

where $T_{-t}F = \{x\in M: T_t x \in F\}$. A similiar statement holds for (3.9). This theorem leads to the definition of mixing.

The third property that we shall be using is strong mixing. A transformation T is said to be strong mixing if

$$(3.12)\quad \lim_{n\to\infty}\int_M f(T^n x)g(x)d\mu(x) = \frac{1}{\mu(M)}\int f(x)d\mu(x)\cdot\int g(x)d\mu(x)$$

for all $f,g \in L^2(M,\mu)$. In the case of a flow $\{T_t\}$, (3.12) is replaced by

$$(3.13)\quad \lim_{S\to\infty}\int_M f(T_S x)g(x)d\mu(x) = \frac{1}{\mu(M)}\int f(x)d\mu(x)\cdot\int g(x)d\mu(x).$$

In other words, we have replaced the Cesàro convergence of (3.9), (3.10) by strong convergence. Since strong convergence implies Cesàro convergence we see that mixing implies ergodicity. The converse is false. The analog of (3.11) is

$$(3.14)\quad \lim_{S\to\infty} \mu(T_{-S}F \cap G) = \frac{\mu(F)\mu(G)}{\mu(M)}$$

for any F, G ε $\mathcal{B}$. Finally, we point out a further result. If M is a metric space with metric d, then we say that T is almost periodic if for any $\varepsilon > 0$, there exists an increasing sequence $\{n_i\}$ such that $d(T^{n_i}x,x) < \varepsilon$ for almost every x.

<u>Prop</u>. Let $T:M \to M$ be a measure-preserving transformation. Then, if T is periodic it cannot be ergodic. Also, if T is almost periodic, then T cannot be strong mixing.

<u>Pf</u>. The first result is trivial. To obtain the second result, let F,G be two sets in $\mathcal{B}$ with nonzero measure such that $\overline{F} \cap \overline{G}$ is empty, where $\overline{F}$ = closure of F. Then, if $\varepsilon < \frac{1}{2} d(\overline{F},\overline{G})$ one sees that the left-hand side of (3.14) is zero for arbitrarily large values of S.

We shall now set down two results which show that an arbitrary measure-preserving transformation is generically ergodic but not mixing in the sense of Baire category.

Let $G_1 = G_1(M,\mathcal{B},\mu)$ be the group of (measure) automorphisms of M. I.e., T is in G_1 if T is measurable, invertible, and measure-preserving. We endow G_1 with the weak topology which is characterized by the following convergence criterion: $T_j \to T$ if and only if $\mu(T_jE \,\Delta\, TE) \to 0$, for all E ε $\mathcal{B}$. Here $A \,\Delta\, B$ is the symmetric difference $(A-B) \cup (B-A)$.

0 0 0 0 4 8 0 0 8 5 7

Define $G_2 = G_2\ (M,\mathcal{B},\mu)$ to be the group of measure-preserving homeomorphisms of M with the topology given by the metric:

$$(3.17)\quad d(T,S) = \max_{x\varepsilon M}\ (\max|Tx-Sx|,\ \max_{x\varepsilon M}|T^{-1}x-S^{-1}x|),\ \forall T,S\ \varepsilon G_2.$$

The norm $|\ \ |$ is that induced by the Riemannian metric on M.

Let X be a topological space and let $U \subseteq X$ be an open set. A set $A \subseteq X$ is dense in U if A has non-empty intersection with every open set of U. A is said to be dense if it is dense in X. A is said to be nowhere dense if there is no open set U such that A is dense in U. Also, A is said to be of the first category if it can be represented as a countable union of nowhere dense sets; otherwise, it is of the second category. Finally, A is said to be a G_δ set if it is the countable intersection of open sets. See [59] for further discussion of these concepts.

<u>Thm</u>. In G_1, the set of all strongly mixing transformations is a set of the 1st category. The set of all ergodic transformations is a dense G_δ set.

<u>Pf</u>. See [26].

<u>Oxtoby-Ulam Thm.</u> If the dimension of M is at least two then the set of nonergodic automorphisms is a set of the 1st category in G_2.

<u>Pf</u>. The case in which M is a cube in $\mathbb{R}^n$ and μ = Lebesgue measure can be found in [55]. The proof therein can be adapted to the case where M is a manifold. The original proof is given in [60].

We shall now consider our dynamical system (1.21)-(1.22) with the associated flow maps $F_t:S \to S$. We observe that for t fixed, $F_t\ \varepsilon\ G_1 \cap G_2$. The conclusions of these theorems is that it is likely in the sense of category that each F_t is ergodic but not mixing. This does not imply

that our flow $\{F_t\}$ is necessarily ergodic and not mixing because we have significant a priori information concerning F_t: the right-hand side of (1.21) is an analytic function on S and each F_t will therefore be smooth, see [27]. That is, each $F_t \in H$ where H is the subgroup of G_1 and G_2 consisting of the smooth transformations and H may itself be of 1st category in G_1 and G_2 (for example). Also, it is well-known that, e.g., the unit interval with the usual topology and Lebesgue measure contains subsets with measure one and of the first category (see [59]). Extrapolating to G_1 or G_2, it is possible that the nonergodic automorphisms, even though they form a set of the first category, have positive measure for some physically interesting measure on G_1 or G_2.

We now introduce some concepts from the theory of dynamical systems. A good reference for further details and examples is [30]. Let X be a differentiable surface in R^m for some m with the induced topology. X may or may not have a measure attached to it. We consider now the two cases of the iterates $\{T^n\}_{n \geq 0}$ of a differentiable map $T:X \to X$ and a flow $\{T_t\}_{t > 0}$ which is a collection of differentiable maps $T_t:X \to X$ such that $T_t \circ T_s = T_{s+t}$ for all $s,t, \in \mathbb{R}_+$. The main example here is the case where the flow $\{T_t\}$ is induced by the time t maps of a system of first-order differential equations, $\dot{x} = F(x)$, on X as in Section 1. An example of a transformation T would be $T = T_1$, the time one map of a flow. Note that we do not require these maps to be onto (whereas in ergodic theory, one only considers the measure-preserving case). Both the ordered pair $(X, \{T^n\}_{n \geq 0})$ or the ordered pair $(X, \{T_t\}_{t \geq 0})$ are called dynamical systems. We will now define an attractor for a dynamical system. A set $A \subseteq X$ is a local attractor for $(X, \{T^n\}_{n \geq 0})$ if there exists on open set U such that $A \subset U \subseteq X$ and for any $x \in U$, $T^n x \to A$ as $n \to \infty$ in the sense

of the topology of X. If the open set U can be taken to be the entire space X, then we call A a global attractor. The definitions for the case of a flow on X are analogous. We emphasize that the set A need not have any particular structure. The attractors can be a single point, a finite number of points, a submanifold of X, the cross product of a submanifold of X and a set homeomorphic to a Cantor set (see [59] for a definition). The latter is a so-called "strange attractor." Observe that if X is endowed with a measure which is absolutely continuous with respect to Lebesgue measure then all of these examples of attractors have measure zero.

Thm. Let μ be a measure on X which is absolutely continuous with respect to Lebesgue measure and suppose that $\{T^n\}_{n \geq 0}$ (or $\{T_t\}_{t \geq 0}$) preserves this measure. Then the dynamical system cannot possess an attractor of measure zero.

Pf. By definition, open sets in X have positive Lebesgue measure, hence positive μ measure. The result then follows immediately.
This result essentially extends to arbitrary attractors except that one must exclude situations in which $A \subset U$ but $\mu(A) = \mu(U)$ (this case, of course, is not of great interest). This theorem will be crucial in the discussion in Section 4 below.

The other idea that will be useful from dynamical systems is that of an invariant set. A set $A \subseteq X$ is said to be invariant for T if $x \in A$ implies $T x \in A$ (hence, $T^n x \in A$, $\forall n \geq 0$) with a similar definition for a flow.

4. The Relationship of Statistical Properties to the Existence of Coherent Structures in Two-Dimensional Flows

The main purpose of this section is to tie together the notions of ergodic theory and dynamical systems introduced in Section 3 with the actual behavior of physical flows. In particular, we shall develop physical implications of mixing or nonmixing. Unless stated otherwise, we shall assume that the truncation of (1.1)-(1.2) used to derive (1.21) contains enough modes to be physically significant.

A critical observable in two-dimensional turbulence is the vorticity distribution. There is evidence to the effect that macroscopic (i.e., large-scale) vortices form as the velocity field develops from arbitrary initial values according to the Navier-Stokes equations. See [13] and [47] for a numerical study. The question of interest here is how does this phenomenon manifest itself in the model system (1.21)-(1.22)? We shall investigate two possible mechanisms which may explain this behavior. The first is dynamical in nature and the second is statistical-mechanical. In the first, it shall be proposed that the system (1.21) contains some dynamical mechanism which leads arbitrary initial states into states which correspond to macroscopic vortices. However, the conservation of μ-measure on S makes such an explanation quite complicated. This objection is not present in the viscid case of (1.1) (i.e., $\nu > 0$) which corresponds to a real physical flow. In the second mechanism we propose that the overwhelming majority (in the sense of μ-measure) of states on S correspond to physical states with macroscopic vortices. This is meant to be in analogy to the derivation of thermodynamic quantities such as temperature from the statistical mechanics of ensembles of gas particles.

In this case, it might not make any difference whether or not (1.21) was mixing because one would observe macroscopic vortex formation in either case. However, if the system were not mixing, then there could be a set of small positive measure on S which never evolved into macroscopic vortices. This situation could not occur if (1.21) were mixing (ergodic is sufficient) because then time means would equal phase means for almost every initial state. In contrast, it is to be noted that the dynamical explanations to be presented will require some degree of nonrandomness for the system (1.21) - certainly, nonergodicity.

In summary, the problem to be considered in this section is the theoretical justification for large-scale coherent behavior of a physical system in the face of small-scale randomness. As has been noted, the solution of this problem is well understood (at least in its general outline, if not in specific details) for a large system of gas molecules. It will be obvious throughout this section that a solution of the problem for the system (1.21) - (1.22) will of necessity be much more complicated. Indeed, little is known about the statistical mechanics (or the ergodic theory) of this system beyond the level of conjecture. As a result, our discussion as well will contain few results but several conjectures. Let us outline the reasons for this situation at the outset-details will follow later. First, (1.21) has two quadratic constants of the motion instead of just one as in gas dynamics. This forces some of the basic ideas of statistical mechanics to be inoperable - e.g., equipartition of energy. Second, the macroscopic observable that we are interested in - vortex structure - is much more complex then temperature, pressure, etc. The reason is that the latter are single

real-valued functions on the state space; on the other hand, it is not at all clear how to represent large-scale vortex structure as a finite collection of such functions, even approximately. Third, in gas dynamics, the phase space is given by the physical degrees of freedom of each of the molecules. For the case at hand, the degrees of freedom are independent Fourier modes - this severely limits ones physical intuition.

In outline, we shall first develop the vortex model as a coarse approximation to the Navier-Stokes equations and correlate macroscopic vortex formation in this model to so-called "negative temperature" states. This has led historically to attempts to define "negative temperatures" in the Fourier mode model and we shall present the statistical mechanical arguments used along with a rather critical appraisal of their validity. We shall next turn to the dynamical ideas of attractors and invariant sets and their application to (1.21) - (1.22). This will lead to a discussion of these concepts in case (1.21) were modified to include viscosity and external forces. Finally, we shall summarize the arguments concerning ergodicity and mixing on the one hand, and coherent large-scale structures on the other hand.

The idea of the vortex model is to replace the continuous distribution of velocity by a discrete set of point vortices. Thus, the model explicitly assumes that the vorticity distribution is the most important observable in a two-dimensional flow. We refer to [3] and [20] for details of the following construction. We proceed as follows: since div $\underline{u} = 0$, there exists a stream function $\psi(\underline{x})$ such that

$$(4.1) \quad \underline{u} = (\partial\psi/\partial y, \; - \partial\psi/\partial x).$$

By the definition of vorticity, it follows that

$$(4.2) \quad \Delta\psi \equiv \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} = -\xi.$$

A single point vortex located at $\underline{x}_0 \in \mathbb{D}$ ($\mathbb{D}$ is the fluid domain) is defined by the vorticity field

$$(4.3) \quad \xi(\underline{x}-\underline{x}_0) = \kappa_0\delta(\underline{x}-\underline{x}_0)$$

where $\kappa_0 \equiv$ strength of the vortex $= \int_{\mathbb{D}} \xi d\underline{x}$ and δ is the delta function. Note that the sign of κ_0 indicates the orientation of the vortex. A stream function for a single point vortex is now easily obtained:

$$(4.4) \quad \psi(\underline{x}-\underline{x}_0) = -\Delta^{-1}\kappa_0\,\delta(\underline{x}-\underline{x}_0) = -\frac{\kappa_0}{2\pi}\log|\underline{x}-\underline{x}_0|.$$

Suppose now that there are M point vortices located at the points $\underline{x}_j$, $j = 1,\cdots, M$, and that there are no other contributions to the vorticity field. Assume further that each vortex retains its structure in time, i.e., that they remain point vortices. Then, since the total vorticity of a fluid element is conserved according to Kelvin's Thm., we see that each κ_j remains constant. We assume that the point vortices move with the fluid. Hence, the velocity of the jth vortex, denoted by $\underline{u}_j$, will be a superposition of the velocity fields generated by all of the other vortices (point vortices do not self-interact). That is,

$$(4.5) \quad d\underline{x}_i/dt = \Sigma_{j\neq i}\,\underline{u}_j, \qquad i = 1,\cdots, M.$$

This defines the vortex model, i.e., one considers an initial distribution of point vortices in $\mathbb{D}$ and studies their subsequent evolution using (4.5).

An important feature of the vortex model is that it is a Hamiltonian system. To see this, define

$$(4.6)\quad q_i = \sqrt{|\kappa_i|}x_i, \qquad P_i = \sqrt{|\kappa_i|}\,\mathrm{sgn}\,(\kappa_i)\, y_i, \qquad i = 1,\cdots, M$$

where $\underline{x}_i = (x_i, y_i)$ and sgn (A) denotes the sign of A. Now define

$$(4.7)\quad H(\underline{q},\underline{p}) = -\frac{1}{2\pi}\,\Sigma_{i<j}\,\kappa_i\kappa_j \log r_{ij}$$

where $\underline{q} = (q_1, \cdots, q_M)$, $\underline{p} = (p_1, \cdots, p_M)$, and $r_{ij} = |\underline{x}_i - \underline{x}_j|$. Using (4.5), one easily verifies that

$$(4.8)\quad dH/dt = 0$$

$$(4.9)\quad dq_i/dt = \partial H/\partial p_i \;\; ; \;\; dp_i/dt = -\,\partial H/\partial q_i, \quad i = 1,\cdots, M.$$

Together, (4.8) - (4.9) define a Hamiltonian system with "energy" H and canonical coordinates (q_i, p_i). Note that H is not necessarily positive and that the configuration space of the system coincides with the plane of motion. The system (4.8) and (4.9) also has two linear constants of the motion corresponding to x- and y- momentum, and an additional quadratic integral which measures the average dispersion of the vortices.

The system (4.5) - (4.9) may also be interpreted in terms of two-dimensional plasma physics. Let the coordinates $\underline{x}_j$ denote the positions of very long and thin rods aligned parallel to an external magnetic field directed perpendicular to the plane of motion which react via the Coulomb potential. In the limit of infinite length and infinite thinness, the rods will remain so aligned, the Hamiltonian function describes the motion of the rods according to a "guiding-center drift," and H may be interpreted as the total potential energy due to the Coulomb interactions. [33] and [51] may be consulted for further details. Also,

0 0 0 0 4 8 0 0 8 6 1

the form of H is somewhat more complicated if boundaries are taken into account; these modifications are taken into account in [63].

We shall now present the statistical mechanical argument leading to negative temperature states, following [10]. Many of the ideas originated in [56]. See also [11], [33], [34], [50], [51], [52], [63], [69], and [70]. We require that the region $\mathcal{D} \subseteq \mathbb{R}^2$ be bounded. This implies that the phase space, $\mathcal{D} \times \cdots \mathcal{D} \times M$ times, has finite volume which is important below and is the main difference between this system and a typical gas model. This assumption entails modifications in the definition of H to take the boundaries into account (as above in the plasma interpretation)- see [63] for details. The system remains Hamiltonian and of the same form as (5.7) so the following argument still applies.

For simplicity, assume that $|\kappa_i| = 1$, for all i. We now assume that M is large and divide $\mathcal{D}$ into subregions $\omega_1, \cdots, \omega_N$ of equal area which are large enough so that interactions between vortices in different regions are negligible compared to interactions between vortices in the same region. Hence, we write

$$(4.10) \quad E = \sum_{i=1}^{N} \varepsilon_i$$

where E = total energy and ε_i = energy in ω_i. Now, if N is large and $\{n_i\}$ is the most likely distribution of vortices among the regions $\{\omega_i\}$, then it is well-known that

$$(4.11) \quad n_i = \text{constant} \times \exp(-\beta\varepsilon_i)$$

where β is interpreted as being inversely proportional to the temperature. See [32] for a proof of (4.11). There are two cases to consider: $\beta > 0$ and $\beta < 0$ (the latter possibility arises because the total phase

space volume is finite. See [56] where this is worked out in detail). To proceed further, we must now make assumptions about the organization of the vortices within the regions ω_i (analogously, in the classical statistical mechanics of an interacting gas, similar assumptions must be made regarding the interaction potential). So, assume first that the vortices are organized into L macroscopic vortices each of a definite sign and each having a density of $|\xi|$ decreasing from some center, where $L \ll N$. Thus, each region ω_i is assumed to contain vortices all of the same sign. If the regions are chosen small enough so that they have diameter less than unity, then it follows that for any pair of vortices in ω_i of strengths κ_ℓ, κ_m that $r_{\ell m} < 1$ which implies that $\log r_{\ell m} < 0$. Since $\mathrm{sgn}\,(\kappa_\ell) = \mathrm{sgn}\,(\kappa_m)$, (4.7) shows that the energy ε_i must be positive. Inserting this result into (4.11), our hypothesis concerning the density of $|\xi|$ (which is equivalent to the distribution $\{n_i\}$) implies that $\beta < 0$. That is, $\beta < 0$ is compatible with the formation of macroscopic vortices and by a similar argument, $\beta > 0$ is not. These macroscopic vortices are the "negative temperature" states previously referred to. Needless to say, the above argument is not a rigorous proof. We remark that by using the equilibrium Gibbs ensemble (4.11) in the derivation of the above dynamical result concerning the Hamiltonian system (4.5) - (4.9) we were implictly substituting phase means for the physically relevant time means, i.e., we assumed that (4.5) - (4.9) is at least ergodic.

Before leaving the subject of point vortices, let us point out that the ideas have proven quite useful recently in the numerical modelling of two-dimensional turbulent flows. The methods were developed in

Chorin's paper [7] and have been studied further in [8], [10], [11], [12], and [66]. The basic idea is to use "cut-off" point vortices in the sense that the stream function is given by (4.4) except for a small area near the vortex where the singularity is smoothed out.

We now show how negative temperatures might arise for the Fourier mode model (1.15) - (1.16). The basic ideas presented here are in [39] and we refer also to [57]. Further references to the statistical mechanics of the Fourier mode model and to negative temperature states are [14], [16], [20], [20], [21], [23], [37], [48], [49], [51], [64], [65], [68], [70].

Recall that $\Sigma_{\underline{k}}\ u_\alpha(\underline{k},t)\ u_\alpha(-\underline{k},t)$ and $\Sigma_{\underline{k}}\ \underline{k}^2 u_\alpha(\underline{k},t)\ u_\alpha(-\underline{k},t)$ are constants of the motion for (1.15) - (1.16). Therefore, given any $\beta,\gamma,\varepsilon\ \mathbb{R}$,

$$(4.12) \qquad \eta \equiv \Sigma_{\underline{k}}\ (\beta\underline{k}^2 + \gamma)\ u_\alpha(\underline{k})\ u_\alpha(-\underline{k})$$

is also a constant of the motion. Now, according to the theorem on equipartition of energy (see [32] and [57]),

$$(4.13) \quad < u_\alpha(\underline{k},t)\ \frac{\partial\eta}{\partial u_\alpha(\underline{k},t)} > = C$$

where C is a constant depending on the total energy, $< >$ denotes an ensemble average, and no summation takes place. The ensemble involved here is the uniform distribution over the surface of constant η. Performing the differentiation in (4.13) we obtain

$$(4.14) \quad < u_\alpha(\underline{k},t)\ u_\alpha(-\underline{k},t) > = C/(\beta\underline{k}^2 + \gamma).$$

Define the energy spectrum $E(k) = \Sigma_{|\underline{k}| = k}\ |\underline{u}(\underline{k})|^2$, so that

$E = \int_0^\infty E(k)dk$; we get $E(\underline{k}) \propto k < u_\alpha(\underline{k})\ u_\alpha(-\underline{k}) >$ and (4.14) leads to

$$(4.15) \quad E(k) = ak/(k^2 + b)$$

where a,b are suitable constants. It can easily be shown that for any realizable choice of total energy and enstrophy, a and b can be found such that (4.15) is a realizable energy spectrum - see [20] for a proof. Equation (4.15) will shortly lead us to negative temperature states in the Fourier mode model. We remark, however, that it is not obvious how to derive (4.15) from the vortex model, which is the heuristic model for negative temperatures.

The number b is taken to be inversely proportional to the temperature of the equilibrium ensemble (4.15). As for β in the vortex model, there is no reason that b cannot be less than zero. Of course, the question now arises whether or not the value of b has anything to do with the dynamics of (1.15) - (1.16)? In particular, does the sign of b have any relation to the formation or lack of formation of macroscopic vortices? There has been one numerical study devoted entirely to this problem, [20]. The methods employed were designed to investigate the equilibrium energy spectra and compare it to (4.15). The conclusion was that the sign of b has no dynamical significance. However, since more subtle observables connected to the vorticity distribution were not tabulated, this result cannot be conclusive. The precise results are as follows: with an initial ensemble in which b is large it is found that the energy spectrum satisfies $E(k) \propto k$ asymptotically and with b small initially the spectrum $E(k) \propto k^{-1}$ was obtained. This correlates exactly with (4.15). However, deviations from these values are noted at high wave numbers - precisely those which should relax to equilibrium values

fastest. If it indeed is the case that (4.15) is not a dynamic equilibrium (i.e., time means of the energy spectrum are not given by (4.15)), this would argue against mixing on surfaces of constant η (however, see the following paragraph which shows that this is probably not a significant question to begin with). Related results are reported in [16]. Here, the authors introduced a small viscosity into the equations and were able to construct two initial ensembles approaching distinct equilibrium spectra neither of which resemble (4.15). Two possible explanations present themselves. First, the inviscid equations may differ markedly in their behavior from the viscous case although this seems unlikely in view of the discussion in Section 2. Second, (4.15) may not hold for all initial ensembles.

Let us reconsider the derivation embodied in (4.12) - (4.14). Note that (1.15) - (1.16) has two quadratic constants of motion whereas (4.12) combines these into one "energy". Consider the constants of the motion E, Ω, and η. Since $\eta' = \Sigma_{\underline{k}}(\phi \underline{k}^2 + \theta)|\underline{u}(\underline{k})|^2$ is also a constant of motion for any choice of $\phi, \theta \in \mathbb{R}$, we conclude that there exists an isolating constant of motion on surfaces of constant η. Since η' constitutes a nonconstant measurable invariant function, the system (1.15) cannot be mixing on a surface of constant η. This implies that the ensembles (4.15) cannot be a statistical equilibrium. However, (4.15) might still be a dynamical equilibrium.

The two isolating constants, E and Ω, are of equal importance a priori. The correct surface to consider is the intersection of the surface of constant energy with the surface of constant enstrophy. On this surface, the analog of the equipartition theorem will presumably

be much more complicated than (4.14) and so the same will apply to the energy spectrum (4.15). We propose, therefore, that the system (1.15) - (1.16) should be viewed as having two thermodynamical quantities associated with it, one for each of the two isolating quadratic constants of the motion. There is no reason to suppose that the statistical equilibria will be of the form (4.15). However, a and b are two candidates for independent temperatures. So, (4.15) is a possible equilibrium spectrum and the results of [20] indicate that it may be valid for part of the range. In the above derivations ergodicity is assumed for the underlying dynamical system. However, this assumption may not be logically necessary to explain coherent large-scale behavior. To see this, let $V \subseteq S$ denote those states which correspond to macroscopic vortices. By considering the formation of macroscopic vortices as a thermodynamic phenomenon or a statistical equilibrium, we are saying that $\mu(S)-\mu(V)$ is overwhelmingly small in the "thermodynamic limit", i.e., in the limit as the number of retained modes in the truncation (1.15) - (1.16) approaches infinity. The only knowledge necessary to transfer this result from the phase ensemble in which they are derived to the physically observed time averages is that S-V is not an exceptional subset of V regarding the dynamics of (1.15), i.e., we must know that initial data in S-V does not stay in S-V a long time and that S-V does not attract data from outside of S-V. Knowing this, one is then justified in claiming that a statistical equilibrium exists for the time ensemble as well as the phase ensemble. Of course, the assumption of mixing on S for (1.15) implies all this and more; the point is that even ergodicity is a stronger assumption then necessary.

As mentioned above, the problem of computing thermodynamic limits and equilibrium ensembles in the event that the underlying dynamical system has two isolating constants of the motion is difficult. Indeed, it is not clear that such limits even exist. We shall now consider two dynamical explanations of the macroscopic vortex phenomenon. As before, let $V \subseteq S$ be the set corresponding to macroscopic vortices in physical space. Then, one may hypothesize

(I) V is an attractor for the dynamical system $\{F_t : S \to S\}$ constructed in Section 2. Alternatively, one might only suppose that V is a local attractor.

(II) V is contained in a nontrivial (explanation below) open set $W \subseteq S$ such that $\mu(W) > 0$ and W is F_t-invariant. Section 3 contains definitions of the terms used here.

In view of the last theorem proved in Section 3, we may elimiante (I) immediately (it has already been shown that μ is absolutely continuous with respect to Lebesgue measure on S. See Section 1). Nevertheless, in the presence of viscosity, the motion is no longer restricted to S but takes place in the entire phase space. Also, the origin of phase space is a global attractor. This is clearly a simple situation from the dynamical systems point of view. It is, however, possible that "metastable" attractors may exist for the system if ν is small enough. That is, the flow $\{F_t\}$ may take arbitrary initial data through the metastable attractor, and the physical behavior characteristic of the attractor persists for a long time relative to the time scale of the problem. Finally, the velocity field decays due to the attraction of the origin. This type of attractor may be of interest in the study of

the so-called inertial range, i.e., those wave-numbers intermediate between the very high modes where viscosity is predominant and the very low ones which describe the larger scales of the motion. See [10] and [39].

A more complicated situation might arise if one were to introduce a time-dependent random stirring force designed to conserve energy and enstrophy. The analogue of (1.15) would be

$$(4.16)\quad \left[\frac{\partial}{\partial t} + \nu \underline{k}^2\right] u_\alpha(\underline{k},t)$$

$$= -\frac{i}{2} P_{\alpha\beta\gamma}(\underline{k}) \sum_{\substack{|\underline{p}|,|\underline{q}| < K \\ \underline{p}+\underline{q} = \underline{k}}} u_\beta(\underline{p},t) u_\alpha(\underline{q},t) + f_\alpha(\underline{k},t)$$

where the components of the external force $\{f_\alpha(\underline{k},t): \alpha = 1,\cdots, N; |\underline{k}| < K\}$ are chosen from an ensemble which conserves energy and enstrophy, but not necessarily in the trivial manner $f_\alpha(\underline{k},t) = \nu \underline{k}^2 u_\alpha(\underline{k},t)$ which would lead us back to (1.15). The system (4.16) may well be a more realistic simulation of a real physical flow than (1.15). The analogue of the Liouville Thm. for (1.15) is

$$(4.17)\quad \sum_{|\underline{k}|<K} \frac{\partial}{\partial u_\alpha(\underline{k},t)} (\dot{u}_\alpha(\underline{k},t))$$

$$= -\nu N \sum_{|\underline{k}| < K} \underline{k}^2 + \sum_{|\underline{k}| < K} \frac{\partial}{\partial u_\alpha(\underline{k},t)} (f_\alpha(\underline{k},t)).$$

This shows that the flow (4.16) does not necessarily conserve the natural measure in phase space and, therefore, the same must apply to S. In the second summation on the right-hand side, the $\{f_\alpha(\underline{k},t)\}$ are chosen from

an ensemble which depends on the instantaneous value of the $\{u_\alpha(k,t)\}$ (this dependence is necessary in order to conserve energy and enstrophy). Thus, our analysis of ergodicity and mixing would not apply to (4.16) because these concepts require an a priori preserved measure. On the other hand, we note that attractors are a distinct possibility for the system (4.16).

It may well be possible to extend this construction one step further and choose the components of the external force from an ensemble which satisifies Liouville's Thm. To see how this development would proceed, we exhibit the system of equations that the $\{f_\alpha(\underline{k},t)\}$ must satisfy in order to belong to the previous ensemble. Conservation of energy requires

$$(4.18) \quad \Sigma_{\underline{k}}(-\nu\underline{k}^2 u_\alpha(\underline{k},t) + f_\alpha(\underline{k},t))\, u_\alpha(-\underline{k},t) = 0$$

and conservation of enstrophy requires

$$(4.19) \quad \Sigma_{\underline{k}}\, \underline{k}^2(-\nu\underline{k}^2 u_\alpha(\underline{k},t) + f_\alpha(\underline{k},t)\,)\, u_\alpha(-\underline{k},t) = 0$$

These two equations follow immediately from (4.16), (1.19), and (1.20). Thus, the ensemble is determined by these four linear constraints (the real and imaginary parts of (4.18) and (4.19) in the $\{f_\alpha(k,t)\}$). In order that Lebesgue measure be preserved in phase space, we must have

$$(4.20) \quad \Sigma_{\underline{k}} \frac{\partial}{\partial u_\alpha(\underline{k},t)} (f_\alpha(\underline{k},t)) = \nu\, N\, \Sigma_{\underline{k}} \underline{k}^2.$$

One way of insuring this is to simply fix a time-independent external force (which is a function of the $\{u_\alpha(\underline{k},t)\}$ once and for all which satisfies (4.18) - (4.20). There are enough degrees of freedom remaining

after satisfying (4.18) - (4.19) to do so. However, the external force would then no longer be entirely random. Once constructed, the induced measure on S for (4.16) would be found in the same way as in Section 1 and the analysis of Sections 3 and 5 would be identical. However, this construction will not be pursued further in this paper.

Turning now to our second hypothesis, we construct an F_t - invariant set $X \subseteq S$ for the system (1.15) - (1.16) which is invariant for the untruncated system (1.6) - (1.7) as well. Its elements are the Fourier transforms of vortices in physical space. Explicitly, consider

$$(4.21) \qquad u_1(\underline{k},t) = ik_2\psi(\underline{k},t), \qquad u_2(\underline{k},t) = -\, ik_1\psi(\underline{k},t)$$

where the stream function is given by

$$(4.22) \qquad \psi(\underline{k},t) = C \exp(iak_1 + ibk_2)/H(\underline{k},t).$$

C,a,b are constants and $X = \{\underline{u}$: the stream function ψ is of the form (4.22) with $H(\underline{k},t)$ arbitrary$\}$, see [6]. X is F_t-invariant because of the arbitrariness of the denominators $H(\underline{k},t)$. Clearly, $\mu(X) = 0$ and each member of X is a vortex. By itself, this may not be a significant result because $\mu(X) = 0$; however, it becomes significant if the following conjecture were known to hold: X is contained in an invariant set W of positive measure. If this were indeed the case, we would have a set of positive measure whose elements correspond closely to vortices in physical space and which stay in this correspondence under the influence of the flow $\{F_t\}$. Some results have been obtained concerning the existence of such a set W given X as above in the case of certain Hamiltonian systems and also some reversible systems. See [53] and

[54]. The Fourier mode model is not known to be Hamiltonian but it is reversible. However, it is difficult to establish the assumptions in these theorems for the Fourier model.

Suppose now that $V \subseteq W$ is as in (II) above. Of course, it is trivial to see that $W = S$ satisfies (II). This case is clearly of no interest and what we require is that W be physically significant. This requirement presumably implies that $\mu(S-W) > 0$ in addition to having $\mu(W) > 0$. The problem remaining is how to treat the F_t-invariant set $S-W$ which, in our model, is of equal a priori importance to the set W relative to the measure μ. So, in the preceding example, we would have an invariant set of vortices and another invariant set of (perhaps) nonvortices, both of positive μ-measure. Clearly, one of our main assumptions would have to be dispensed with if such a set W existed. The possibilities are (1) vortex formation is not an intrinsic property of 2-D turbulence or (2) the state space $(S,\mathcal{B},\mu)$ is an invalid approximation for two-dimensional turbulent flows (see Section 2 for a further discussion of the latter property). Once again, the existence of W as above can also have a thermodynamic explanation; that is, as the number of retained modes approaches infinity, it may happen that $\mu(S-W) \to 0$ (S and W both depend implicitly on the number of modes). This would allow the "unphysical" states of $S-W$ to have μ-measure 0 in the limit of all modes being present.

We now try to summarize the relationship between statistical properties of the flow $\{F_t\}$ such as ergodicity and mixing, and the ideas sketched in this section. Our basic hypothesis is that large-scale vortex formation occurs in two-dimensional turbulent flows and should,

therefore, be a feature of our model, at least for large-scale truncations. We have also assumed that $(S,\mathcal{B},\mu)$ with the dynamical system $\{F_t:S \to S\}$ is a valid model for fully-developed inviscid turbulence. We have noted two classes of explanations for the vortex phenomena: statistical mechanical and dynamical. The latter view the flow $\{F_t\}$ as being non-random to some extent - at least it is not mixing nor even ergodic. However, the dynamical explanations are logically flawed in the inviscid case because of violations of conservation of μ-measure (and the invariant set hypothesis is consistent only in conjunction with a thermodynamic argument). On account of this problem, it is not clear what ergodicity or nonergodicity of the flow $\{F_t\}$ would imply about turbulence or vortex structure. Still, one assumes intuitively that nonergodicity would imply the existence of some further significant constraint on the statistics of turbulent flows while ergodicity would tend to imply the opposite result. As we have already noted, the statistical mechanical explanations do not technically require that the flow be mixing or ergodic. However, the assumption of mixing makes the derivations much more plausible because it allows interchange of phase averages and time averages. Thus, it appears unlikely that well-defined temperatures or similar phenomena could exist if the underlying flow were not at least ergodic.

Our conclusion is that the vortex structure of two-dimensional turbulence can be explained with or without the assumption of mixing or ergodicity for the system (1.20) - (1.21). Nevertheless, the validity of these assumptions deserve to be investigated. First, nonergodicity would imply the existence of a further isolating constant of the motion

besides energy and enstrophy (see Section 3). Although this integral need only be measurable, it may still have physical significance. Second, a firm answer to the question may point the way in deciding between thermodynamic and dynamical explanations of macroscopic phenomena such as vortex coalescence in turbulence theory.

Suppose now that the dynamical system (1.20) - (1.21) is nonergodic and suppose further that a constant of the motion, η, has been found explicitly. It then becomes relevant to study the statistical properties on the surface $S' = S \cap S_\eta$ where S_η = surface of constant η. We would then have dim $S' = M-3$ where M is the dimension of phase space. This study would be particularly relevant in case the existence of η were not sufficient to describe the phenomena of interest - coherent large-scale vortex structures. These comments apply as well to the situation in which a finite number of constants of the motion were found, say $\eta_1, \cdots, \eta_\ell$, and dim $S' = M-\ell$. In other words, it is possible that S is partitioned into surfaces of low codimension on each of which the restriction of the flow $\{F_t\}$ is mixing, without even being ergodic on all of S. In this event, we might say that the flow $\{F_t\}$ is essentially random, especially if the integrals $\eta_1, \cdots, \eta_\ell$ were not of great physical significance. Thus, it may be premature to draw drastic conclusions from the finding that (1.20) is not ergodic. We should note that in case η is merely measurable, the partition of S indicated above would also only be measurable and not necessarily a "surface," thereby invalidating much of the analysis of Sections 1,3, and 5. However, $\{F_t\}$ is a differentiable flow so it seems unlikely that such an η would not be at least piecewise continuous.

As a final note, we shall now compare an aspect of three-dimensional turbulence with two-dimensional turbulence. In the three-dimensional case, the circulation around any closed curve is a constant of the motion-hence, there are an infinite number of constants of the motion, one for each curve. Nevertheless, it is generally agreed that three-dimensional turbulence is random because there is an obvious mechanism by which these constants become "coarse-grained" throughout the fluid domain. Namely, the stretching of vortex tubes causes each element to become thinner and longer. Since the fluid domain is finite, this forces the tube to randomly weave its way through the fluid and, eventually, the distribution of any observable taken along the curve in question becomes indistinguishable from the uniform distribution in phase space - this is precisely the criterion for mixing (see [57]). Analogously, in two-dimensional flow there are also an infinite collection of constants of the motion; these are the total vorticity of each fluid element, one constant for each element. The difference is that in this case it is not clear whether or not there exists a physical mechanism by which these elements can become coarse-grained - there is no vortex stretching in two dimensions. It should be pointed out that we have been discussing the untruncated system (1.1) - (1.2) and that none of the integrals discussed here survive the truncation in the Fourier mode model. See [3]. Also, it is not clear what is meant by mixing or ergodicity in an infinite-dimensional space until the latter is assigned a flow invariant measure.

5. The Methodology Employed to Numerically Study Ergodicity and Mixing

In this section, we shall develop the numerical methods used in studying the statistical properties of the dynamical system (1.21) - (1.22). These methods are applicable to a truncation of any size leading to (1.21) - (1.22). The numerical results to be presented in Section 7 treat only very small truncations and will serve to test the feasibility of applying the methods to a larger truncation.

The study of the ergodicity and mixing properties of a dynamical system can be quite simple when the solution is known explicitly. For example, if the solution is almost periodic then the system cannot be mixing - we will present an example of this in Section 6. However, the solution of (1.21) is not known explicitly, in general. Therefore, it will be necessary to work with an approximate solution of (1.21); furthermore, this approximate solution will be too complex to be explictly solved for all time and will not be of any use in the manner just indicated. As a result, we shall be unable to work directly with the definitions of ergodicity and mixing; instead, we shall apply their consequences as described in Section 3, especially equations (3.8) and (3.13).

Recall that a transformation or flow is ergodic if and only if every measurable invariant function is a constant. In particular, the existence of a nontrivial constant of the motion proves that a dynamical system cannot be ergodic. Let M be the number of independent variables retained in (1.21) and suppose that

$$(5.1) \quad \eta(\underline{\omega}) = \sum_{i=1}^{M} \gamma_i \omega_i^2$$

is a constant of the motion for the flow on the surface S induced by (1.21), i.e., $\dot{\eta}(\underline{\omega}) = 0$ identically on S. Since $\dot{\eta}(\underline{\omega})$ is a homogeneous cubic polynomial, this will yield a system of linear equations in the unknowns $\{\gamma_i\}$, one equation for each term of the form $\omega_\ell \omega_m \omega_n$ which appears in $\dot{\eta}(\omega)$. Conversely, one may hypothesize an integral of the motion of the form (5.1) with as yet undetermined coefficients $\{\gamma_i\}$, derive the system of linear equations, and see whether or not they have a nonzero solution. If so, an integral of the motion will be obtained, and this will be a new integral once one has checked that η is independent of E and Ω. Clearly, this method is applicable to any hypothetical polynomial constant of the motion. A few brief computations will convince the reader that as inhomogeneous terms are included, or cross terms such as $\omega_i \omega_j$, $i \neq j$, are included, or as the degree of the polynomial increases, the probability of finding a new constant of the motion is small because the number of linear equations will be much larger than the number of unknown coefficients. Nevertheless, see [24] and Section 6 for a successful application of this technique.

We shall now describe how we have used equations (3.8) and (3.13) as a numerical method for the study of the ergodicity and mixing properties of the system (1.21). Recall that $\{F_t\}$ is ergodic if and only if (3.8) holds for every measurable f and for almost every $\underline{\omega} \in S$ and that $\{F_t\}$ is mixing if and only if (3.13) holds for every choice of measurable f and g. Thus, to prove ergodicity, it is necessary to test an infinite collection of functions on a set of full measure in S. On

the other hand, to prove nonergodicity, it is necessary to find one function such that (3.8) is false on a set of positive measure in S. Analogous statements hold for mixing and (3.13). In addition, both (3.8) and (3.13) involve an infinite limit in time. Therefore, it is necessary to have a solution for all time in order to prove either ergodicity or nonergodicity. Since we will be using an approximate solution, we will only have solutions for a finite time, for a finite number of points, and for a finite number of measurable functions. As a consequence, our numerical methods will not be able to prove any results but will only be able to provide evidence. Note: There are theoretical grounds for believing that sampling a finite number of points is sufficient to obtain a proof of ergodicity. However, the argument, which is in [41] and [60], is nonconstructive so that one has no way of knowing which are the correct points to choose.

Let $\rho(\underline{\omega})$ denote the density function of the measure μ on S, i.e., $d\mu(\underline{\omega}) = \rho(\underline{\omega})\, d\underline{\omega}$. We shall need to describe how the following approximations were carried out:

$$(5.2) \qquad \frac{1}{T}\int_0^T f(F_t\underline{\omega})dt \sim \frac{1}{L}\sum_{i=1}^{L} f(F_{t_i}\underline{\omega})\Delta t_i$$

$$(5.3) \qquad \int_S f(\underline{\omega})d\mu(\underline{\omega}) \sim \frac{1}{N}\sum_{i=1}^{N} f(\underline{\omega}_i)\rho(\underline{\omega}_i)$$

$$(5.4) \qquad \int_S f(F_t\underline{\omega})g(\underline{\omega})d\mu(\underline{\omega}) \sim \frac{1}{N}\sum_{i=1}^{N} f(F_t\underline{\omega}_i)g(\underline{\omega}_i)\rho(\omega_i), \text{ t fixed.}$$

The approximation (5.2) will be straightforward after we have described the ordinary differential equation solver used for (1.21). We used a Monte Carlo routine for the approximations (5.3) and (5.4) which will be presented in detail. In addition, our choice of final integration time

T, the measurable test functions f and g, and the values of E and Ω will also be discussed.

The system (1.21) was integrated using a readily available routine known as Gear's method. Complete documentation may be found in [29] and a theoretical analysis in [22]. This routine contains two subroutines, one for stiff problems and the other for nonstiff problems. Since (1.21) is nonlinear, it is not known a priori what percentage of the time that $F_t(\underline{\omega})$ will lie in a region of S which is stiff for $\dot{\underline{\omega}} = \underline{F}(\underline{\omega})$. We tested several points with both routines (obtaining nearly identical results) and became convinced that the problem is rarely stiff. Therefore, we have used the more efficient nonstiff routine for the bulk of the calculations. The latter is an Adams-Moulton, i.e., predictor-corrector, method of variable order. The user inputs the maximum allowable local truncation error (i.e., the magnitude of the difference between the actual solution and the approximate solution for one time step) and the routine automatically sets the order, the number of corrector iterations, and the time step with optimization of computing effort in mind. Of course, the global truncation error is a complicated function of the local truncation error and the problem being solved and cannot be calculated explicitly. Thus, accuracy for long integration times must be insured by other means. Suppose that a local truncation error of 10^{-K} has been chosen (K varied between 7 and 10 for the truncations actually studied). The total integration time T was then chosen as follows: several tests were made with identical initial data solved using local truncation errors of 10^{-K} and $10^{-(K+1)}$. T was then taken to be the time at which the solutions begin to differ in the third

significant figure. In this way, we were able to insure three significant figures in the solution for the entire calculation; also, we thereby obtained an average accuracy of six or seven significant figures over the range of the integration. It is of interest to note that energy and enstrophy are conserved almost exactly for the values of local truncation error and total integration time chosen. Indeed, this would still be the case even if T were increased by an order of magnitude. Therefore, conservation of E and Ω is not a sufficient criterion in selecting T. There seems little reason to doubt that the method just described is adequate to insure accuracy of solutions to the differential equations.

A difficulty which must be dealt with is that the statistical properties of the system (1.21) - (1.22) may be dependent on the choice of E and Ω. It was impractical to numerically test a wide range of choices and we did not do so. In practice, only one value of E and Ω was selected for each truncation, and the choice was made on purely pragmatic grounds. Specifically, small values of E and Ω lead to large (i.e., efficient) step sizes in the Gear routine while large values for E and Ω lead to small step sizes. So, E and Ω were chosen to be small enough so that the computing effort was within reason but large enough so that (5.2) would be well approximated. Also, the relative sizes of E and Ω must also be selected. Of course, (1.29) must be satisfied; to facilitate the Monte Carlo approximations in (5.3) and (5.4), it is best to choose E/Ω somewhere in the middle of the range allowed by (1.29). In this way, $\mu(S)$ is approximately maximized because $\mu(S)$ approaches zero as E/Ω approaches either of its extreme

values. Now, if $\mu(S)$ were too close to zero then (5.2) - (5.4) would be forced to hold to several significant figures whether or not (1.21) - (1.22) is ergodic, unless the functions f,g are chosen to insure a wide total variation over S. It was thought best to avoid this purely numerical difficulty in the manner indicated.

We now consider the approximation (5.2). Clearly, the right-hand side is simply an approximating Riemann sum for the left-hand side. The Δt_i are given by the (variable) step sizes of the Gear method. Even though this is the simplest possible choice of numerical integration routine, it was felt that the small step sizes used (between 10^{-2} and 10^{-1}) combined with the simplicitly of the functions to be integrated (see below) insured sufficient accuracy. However, we are actually integrating $f \circ F$ and not simply f; the former may have a much larger variation than the latter even for very simple choices of f. In view of the fact that the flow $\{F_t\}$ is smooth (see Section 3), we hope that this is not a significant factor. Finally, observe that there is a numerical tradeoff between the final integration time T and the average step size Δt arising from the choice of E and Ω. That is small (respectively large) values of E and Ω lead to large (small) T and large (small) Δt. Of course, large T and small Δt are desirable for a good approximation in (5.2) whereas large Δt is desirable in order to lessen the computing effort.

We turn now to the Monte Carlo evaluation of (5.3). Because of the complicated topological structure of S (see equation (1.36)), we have been unable to devise a method to uniformly sample points in the parameter domain $\mathbb{D}$ of S. The usual method of enclosing $\mathbb{D}$ in a simpler

region is prohibitively inefficient. Our solution to this problem is to choose a subset $E \subseteq \mathbb{D}$ which can be sampled uniformly and then to evaluate only those measurable functions which vanish on the complement of the image of E in S of the parameter map. We now construct E and describe the sampling algorithm. The notation is that of Section 1. Define

$$(5.5) \quad \lambda_i = c_i/(\beta_M E - \alpha_M \Omega); \quad \mu_i = d_i/(\alpha_{M-1}\Omega - \beta_{M-1}E)$$

$$(5.6) \quad \zeta_i = \max(\lambda_i, \mu_i)$$

for $i = 1, \cdots, M-2$. Then,

$$(5.7) \quad \sum_{i=1}^{M-2} \zeta_i \, \omega_i^2 \leq 1$$

describes an (M-2)-ellipsoid inscribed in D. More generally,

$$(5.8) \quad \sum_{i=1}^{M-2} \zeta_i \omega_i^2 \leq R, \; R \geq 1$$

will describe an (M-2)-ellipsoid which is partially interior and partially exterior to $\mathbb{D}$. Clearly, there exists an $R_{crit} > 1$ such that (5.8) describes an (M-2)-ellipsoid which superscribes $\mathbb{D}$. We have chosen $R = R_c$ with $1 < R_c < R_{crit}$ in such a way as to optimize a numerical tradeoff which we now describe. For values of R close to 1, almost every sampled point will lie in $\mathbb{D}$ and can be used in the approximation (5.3). However, it then follows that $\mu(E)/\mu(S)$ is so small that the the trajectory of a point $\underline{\omega}_0 \varepsilon S$ will almost never lie in E (we view $E \subseteq S$ by abuse of notation). Since the sample functions are chosen to

vanish outside of E, this would cause the approximation (5.2) to be statistically meaningless. Suppose next that $R \sim R_{crit}$. In this event, the trajectories of (1.21) will be nearly contained in E but the sample points for the integration of (5.3) will seldom lie in D. Since evaluations of (5.2) and (5.4) are several orders of magnitude more costly per point than evaluations of (5.3), we have chosen R_c in such a way that on the average well over 99 per cent of a trajectory lies in E. In this way, the information obtained through the integration of the system is not artifically lost. Of course, this requires $R_c \sim R_{crit}$ and the price must be paid in the approximation (5.3).

Next, we demonstrate a procedure for uniformly sampling points from a solid (M-2)-ellipsoid. See also [36] and [61]. First, we describe the "polar method" which yields two independent normally distributed random variables. It consists of four steps:

(I) Generate two independent variables U_1 and U_2 uniformly distributed on (0,1) and set $V_1 = 2U_1-1$, $V_2 = 2U_2-1$.

(II) Set $Z = V_1^2 + V_2^2$.

(III) If $Z \geq 1$, return to Step I.

(IV) Define

$$X_1 = V_1 \sqrt{-\frac{2\ln Z}{Z}} \quad , \quad X_2 = V_2 \sqrt{-\frac{2\ln Z}{Z}}$$

The proof that X_1, X_2 are independent and normally distributed is straightforward - see [36]. Second, to obtain M-2 such indpendent normally distributed random variables, it is necessary to repeat the above procedure (M-2)/2 times (recall that M is even). Denote X_1, X_2, and Z above by X_1^i, X_2^i, and Z^i for $i = 1, \cdots, (M-2)/2$. If we set

$$X_i \leftarrow X_i \Big/ \sqrt{\sum_{i=1}^{(M-2)/2} (-2\ln Z^i)}$$

then it follows that $\sum_{i=1}^{M-2} X_i^2 = 1$. See [36] for a proof that the random points $(X_1, \cdots, X_{M-2})$ are uniformly distributed on the unit (M-2)-sphere. Third, we set

$$X_i \leftarrow (R_c/\sqrt{\zeta_i})\, X_i$$

and obtain sample points uniformly distributed on the surface of the (M-2) - ellipsoid (5.8) with $R = R_c$. Finally, to obtain a uniform distribution on the solid (M-2) - ellipsoid, we set

$$X_i \leftarrow \sigma X_i$$

where σ, $0 \leq \sigma \leq 1$, is a random variable distributed in such a way as to weight the volume of the ellipsoid according to radial distance. The construction of σ is straightforward and is omitted.

In summary, points are uniformly sampled from $E \subseteq \mathfrak{B}$ and the corresponding point on S is then constructed from (1.24) and (1.35). The signs of ω_{M-1}, ω_M are chosen randomly with equal probabilities for each case. The density $\rho(\underline{\omega})$ is then evaluated using (1.38) and (1.39). Thus, the approximation (5.3) is a Monte Carlo integration of f on the set E. Since f vanishes on the complement of E, this is all that is required.

The approximation (5.4) was handled in the same way that (5.3) was. Once again, g was chosen to vanish outside of E. The integration was performed once at t = 0 and then ten times during the integration at t = 0.1 T, 0.2 T, ···, T. Of course, it is a trivial matter to save the

value of $g(\underline{\omega})$ during an integration as well as to save the partial sums between integrations of different initial data.

We have already noted that the functions f and g used in the integrations were chosen to vanish outside of E. Hence, they are merely measurable, and not even continuous. Denote $f(\underline{\omega}) = \chi_E(\underline{\omega})h(\underline{\omega})$ where χ_E is the characteristic function of E. The functions h that were tested are:

$$(5.9) \quad h(\underline{\omega}) = \omega_i^{\ell}, \quad \ell = 0,1, \cdots, 4; \quad i = 1, \cdots, M$$

$$(5.10) \quad h(\underline{\omega}) = |\omega_i^{\ell}|, \quad \ell = 1,3; \quad i = 1, \cdots, M$$

$$(5.11) \quad h(\underline{\omega}) = |\omega_i \omega_j|, \quad 1 \leq i, j \leq M$$

$$(5.12) \quad h(\underline{\omega}) = \omega_i^2 \omega_j^2, \quad 1 \leq i, j \leq M$$

$$(5.13) \quad h(\underline{\omega}) = \begin{cases} 100 & \text{if} \quad |\omega_i| < 0.01 \\ 1/|\omega_i| & \text{if} \quad |\omega_i| \geq 0.01 \end{cases}, \quad i = 1, \cdots, M$$

$$(5.14) \quad h(\underline{\omega}) = \begin{cases} 0 & \text{if} \quad \omega_i \leq 0 \\ 1 & \text{if} \quad \omega_i > 0 \end{cases}, \quad i = 1, \cdots, M$$

$$(5.15) \quad h(\underline{\omega}) = \begin{cases} -1 & \text{if} \quad |\omega_i| \leq C \\ +1 & \text{if} \quad |\omega_i| \geq C \end{cases}, \quad i = 1, \cdots, M$$

where C is a suitably chosen constant (different for each truncation)

0 0 0 0 4 8 0 0 8 7 3

$$(5.16) \quad h(\underline{\omega}) = \begin{cases} 0 & \text{if } \omega_i < 0 \text{ and } \omega_j < 0 \\ 1 & \text{otherwise} \end{cases}, \quad 1 \leq i, j \leq M$$

$$(5.17) \quad h(\underline{\omega}) = \omega_i' - [\omega_i'], \quad i = 1, \cdots, M$$

where $\omega_i' = 5\zeta_i^{1/2}\omega_i + 5$ and [ ] denotes the greatest integer function.

$$(5.18) \quad h(\underline{\omega}) = P_j(\underline{\omega}), \quad j = 1, \cdots, 8$$

where the P_j are special polynomial functions of the coordinates ω_i. These will be presented in Section 6.

By the linearity of the integral, we may test any linear combination of the functions given by (5.9) - (5.18) without any additional computing effort.

It will be noted that (5.9) - (5.18) are all rather simple functions. Nevertheless, (5.9) already includes the energy and vorticity spectra which have been the object of much of the previous numerical work on the ergodicity problem for (1.21) - see Section 6. Also, since the truncations chosen for the numerical study are small, it is not clear that more complex functions would be of any greater physical interest than those in (5.9) - (5.18). Of course, this will not be the case for large-scale truncations and the choice of functions to be integrated will have to be given much greater thought before similar tests are undertaken for such truncations. In particular, it would be of great interest to construct observables which measure the extent of vortex coalescence - see Section 4.

It remains to discuss our handling of the constant $\mu(S)$ which occurs in equations (3.8) and (3.13). This number is not known because of the

complexity of the surface S. Furthermore, it cannot be calculated approximately because we cannot sample over the entire surface S. However, observe that

$$(5.19)\qquad \frac{1}{\mu(S)}\int_S f(\underline{\omega})\,d\mu(\underline{\omega}) = \frac{1}{\mu(S)}\int_E f(\underline{\omega})\,d\mu(\underline{\omega})$$

$$= \left(\frac{\mu(E)}{\mu(S)}\right)\frac{1}{\mu(E)}\int_E f(\underline{\omega})\,d\mu(\underline{\omega})$$

since f vanishes outside of E. We make the approximation

$$(5.20)\qquad \mu(E) \sim \frac{1}{N}\sum_{i=1}^{N} \rho(\underline{\omega}_i)$$

where the $\underline{\omega}_i$ are the Monte Carlo points used in (5.3). The equations (5.3) and (5.20) lead to

$$(5.21)\qquad \frac{1}{\mu(S)}\int_S f(\underline{\omega})\,d\mu(\underline{\omega}) \sim \left(\frac{\mu(E)}{\mu(S)}\right) \times \sum_{i=1}^{N} f(\underline{\omega}_i)\rho(\underline{\omega}_i) \Big/ \sum_{i=1}^{N} \rho(\underline{\omega}_i).$$

Consider equation (3.8). We wish to test whether or not the ratio

$$(5.23)\qquad \frac{1}{T}\int_0^T f(F_t\underline{\omega})\,dt \Big/ \frac{1}{\mu(S)}\int_S f(\underline{\omega})\,d\mu(\underline{\omega})$$

approaches unity as $T \to \infty$ for the functions f in (5.9) - (5.18). Using the approximations, we are testing whether or not

$$(5.24)\qquad \frac{1}{L}\sum_{i=1}^{L} f(F_{t_i}\underline{\omega})\Delta t_i \Big/ \frac{\sum_{i=1}^{N} f(\underline{\omega}_i)\rho(\omega_i)}{\sum_{i=1}^{N} \rho(\omega_i)} \to \frac{\mu(E)}{\mu(S)}$$

as $T \to \infty$. Similiarly, with equation (3.13) we wish to test whether

or not the ratio

$$(5.25)\qquad \frac{1}{\mu(S)}\int_S f(F_t\underline{\omega})g(\underline{\omega})\,d\mu(\underline{\omega}) \Big/ \frac{1}{\mu(S)}\int_S f(\underline{\omega})\,d\mu(\underline{\omega}) \cdot \frac{1}{\mu(S)}\int_S g(\underline{\omega})\,d\mu(\underline{\omega})$$

approaches unity as $t \to \infty$. In terms of the approximations this becomes

$$(5.26)\qquad \frac{\sum_{i=1}^{N} f(F_t\underline{\omega}_i)g(\underline{\omega}_i)\rho(\underline{\omega}_i)}{\sum_{i=1}^{N} \rho(\underline{\omega}_i)} \Big/ \frac{\sum_{i=1}^{N} f(\underline{\omega}_i)\rho(\underline{\omega}_i)}{\sum_{i=1}^{N} \rho(\underline{\omega}_i)} \cdot \frac{\sum_{i=1}^{N} g(\underline{\omega}_i)\rho(\underline{\omega}_i)}{\sum_{i=1}^{N} \rho(\underline{\omega}_i)}$$

approaching $\mu(E)/\mu(S)$ as $t \to \infty$. Even though this ratio is not known exactly, it is a constant for all functions f chosen. Thus, if we can find two functions for which the ratios of the form (5.24) differ significantly, we would have evidence of nonergodicity of the flow. Similiarly for mixing and equation (5.26). In practice, $\mu(E)/\mu(S)$ is very close to unity because of our choice of the region E. Note also that the Lebesgue measure of the parameter domain $\mathbb{D}$ occurs as a factor in both $\mu(E)$ and $\mu(S)$. However, it cancels in the ratio $\mu(E)/\mu(S)$ and is of no concern in the computations.

We now describe our choice of the problems that were numerically studied. This choice was motivated by our lack of a priori knowledge concerning the expected convergence rates of (5.25) and (5.26) in the mixing case. On the other hand, we had no a priori knowledge of the amount of divergence to be expected in the nonergodic case. In particular, if the maximum deviation from the ratio $\mu(E)/\mu(S)$ in (5.24) is less than that expected from purely numerical error, then the ratio (5.24)

would indicate that the flow is ergodic even though it was known otherwise. Therefore, it was thought best to choose a wide variety of problems containing various amounts of a priori information concerning ergodicity and mixing.

Three different truncations were selected for the study. The first was a ten mode (therefore, twenty independent variables) truncation of the system (2.15) - (2.16). The second was a five mode truncation of equation (2.17) and the third was a six mode truncation of (2.17). Furthermore, the five mode truncation was known in advance to be nonergodic - see Section 6. We had no a priori information for either the six or ten mode truncations.

For each of the three truncations, we constructed a new dynamical system as follows: equation (1.21) was solved as above with Gear's method except that an unrealistically high local truncation error was selected. The effect of this is that the solution begins to differ from the actual solution in the first significant figure at integration times about one-tenth or one-fifth of the total integration time. Thus, the approximations (5.2) and (5.4) will have less dynamical significance for long integration times and will take on more of the character of a Monte Carlo integration of the type (5.3). We therefore expect the ratios (5.24) and (5.26) to exhibit a somewhat greater degree of convergence for these systems than they would for the original systems.

For the purpose of comparison, we have also tested a system which, intuitively, is as random as possible. The system may be viewed as a simplified model for a "hard point gas" or simply as a "computer game." It was introduced in [9].

Let S denote the unit 2L-dimensional sphere with coordinates $(a_1, \cdots, a_L, b_1, \cdots, b_L)$. A transformation $T:S \to S$ is constructed as follows: given $\underline{p}$ in S consider in succession all integers $1 \leq \ell \leq L$. Given ℓ, pick ℓ' at random between 1 and L with all integers having equal probability of being chosen. Given ℓ and ℓ', consider (a_ℓ, b_ℓ) and $(a_{\ell'}, b_{\ell'})$ to be the velocities of two particles undergoing an "elastic collision." That is, let ξ be a random variable equidistributed between 0 and 1, define

$$\underline{e} = (\cos 2\pi\xi, \sin 2\pi\xi),$$

let

$$\underline{V} = (a_\ell - a_{\ell'}, b_\ell - b_{\ell'}),$$

and set

$$T(a_\ell, b_\ell) = (a_\ell, b_\ell) - (\underline{V} \cdot \underline{e})\, \underline{e}$$

$$T(a_{\ell'}, b_{\ell'}) = (a_{\ell'}, b_{\ell'}) + (\underline{V} \cdot \underline{e})\, \underline{e}.$$

T preserves Lebesgue measure and "energy" (i.e., T maps S into S). In addition, T preserves the total momentum

$$A = \sum_{\ell=1}^{L} a_\ell \quad ; \quad B = \sum_{\ell=1}^{L} b_\ell .$$

On the average, each pair (a_ℓ, b_ℓ) undergoes two "collisions" each time T is applied.

We have taken $A = 0$ and $B = 0$ in the numerical work. Thus, the flow of T takes place on the great circle C given by $A = 0$, $B = 0$. The flow is believed to be mixing, although we cannot prove it.

The dynamical system given by T differs substantially from the system (1.21) - (1.22). First, T is a discrete transformation whereas (1.21) defines a continuous flow. Second, due to the absence of the spatial coordinates from phase space, T has no natural time scale. The numerical analysis of (3.8) and (3.13) proceeds just as it did for the flow $\{F_t\}$ of (1.21). However, there are two simplifications. First, points can be uniformly sampled from C using the algorithm presented above. Second, the density function $\rho(\underline{\omega})$ is a constant on C: the analogue of (2.23) for the conserved measure on C is

$$(5.27) \qquad \mu(M) = \int_M \frac{d\Sigma}{\|\text{grad } E\|}$$

where $M \subseteq C$ is measurable, $E = \sum_{\ell=1}^{L} (a_\ell^2 + b_\ell^2)$ is the only conserved quadratic quantity, and $d\Sigma$ is a surface element on C. Since $d\Sigma$ is already being uniformly sampled and $\|\text{grad } E\|$ is constant on C, it follows that $\rho(\underline{\omega})$ is also constant on C. If we define a new measure on C by $\nu(M) = \alpha\mu(M)$ and suppose that μ is T-invariant then $\nu(TM) = \alpha\mu(TM) = \alpha\mu(M) = \mu(M)$ and ν is also T-invariant. Hence, there is no harm in taking $\rho(\underline{\omega}) \equiv 1$ on C.

We have studied the two models L = 4 and L = 8 in the numerical work.

0 0 0 0 4 8 0 5 8 7 6

6. Previous Results by Other Authors

A recent study of the one-dimensional two particle hard point gas system is reported in [5]. See also [19] and [71]. The two particles are constrained to move in a one-dimensional box and they move with a uniform velocity except for instantaneous collisions with each other and the two walls. Both the total phase space in which velocity and position coordinates are retained and the velocity phase space (position coordinates not included) are considered. Thus, the energy surfaces are three-dimensional and one-dimensional, respectively. The resulting dynamical systems are viewed as taking place on these two energy surfaces. In case the two particles are assigned equal mass, it is easy to construct additional analytic constants of the motion besides energy; hence, neither system is ergodic. Using analytic methods, Casati and Ford show in [5] that neither system is ergodic for a countably infinite dense collection of mass ratios. On the other hand, numerical results are presented which provide strong evidence that the systems are both ergodic and mixing for all other possible mass ratios. The numerical methods take full advantage of the low dimensionality of phase space and the simplicity of the system. Consequently, the techniques involved are not as time-consuming as our method of comparing time means to phase means. The main reason for outlining these results is to point out the dependence of the statistical properties of the system upon a parameter, the mass ratio. The ratio of energy to enstrophy is a similar parameter for the system (1.21) - (1.22); if the ergodic properties of (1.21) were known to have a discontinuous dependence on E/Ω such as observed in [5] then our analysis would need to be much more refined. In particular, it would be necessary to study several E/Ω ratios for each truncation.

Further studies relevant to the numerical analysis of statistical properties of other dynamical systems are [3] and [28].

A prototype for this thesis is [9] where the Fourier-transformed truncated Burgers equation (1.10) is studied using the same methods as presented in Section 5. In addition, the computer game as described in Section 5 is introduced and results for it are compared to those obtained for the Burgers equation. The truncation used is very small-scale; only four independent modes are retained. Similarly, the four particle computer game is studied. In the comparison of time means with phase means, only quadratic functions of the coordinates are considered. The numerical results are impressive: the Cesàro sums for the computer game converge rapidly to ensemble averages whereas it is clear that no convergence takes place at all for the truncated Burgers equation (1.10). Thus, it is reasonably certain from these computer tests alone that the system (1.10) is not ergodic for the truncation considered.

It was later shown by Hald in [25] that equation (1.10) has an extra analytic constant of the motion. It is given by

$$(6.1) \quad T = \sum_{\substack{j+k+\ell = 0 \\ j,k,\ell \neq 0}} u(j,t)\, u(k,t)\, u(\ell,t).$$

This fact was also known to Orszag (private communication). Of course, this verifies the numerical results obtained by Chorin in [9] since the existence of the extra constant of the motion proves that the system (1.10) cannot be ergodic on surfaces of constant energy. We note that the Burgers equation, the two-dimensional Navier-Stokes equation, and the three-dimensional Navier-Stokes equation are now known to each have an extra

0 0 0 0 4 8 0 6 8 7 7

analytic constant of the motion besides energy. These are T, enstrophy, and helicity (= $\underline{u}\cdot\underline{\xi}$) respectively. However, T is a cubic whereas enstrophy and helicity are quadratic.

Also in [25], a simple truncation of (1.10) is studied analytically in detail. Only two independent modes are retained, $k_1 = 1$ and $k_2 = 2$. Hence, the phase space is four-dimensional. It is shown that the motion on the intersection of the surface of constant energy with the surface of constant T is sometimes periodic, in general ergodic, but never mixing. The latter result follows by showing that the motion is almost periodic.

We turn now to papers concerned with the two-dimensional Navier-Stokes equations. Using the analytic techniques briefly sketched in the beginning of Section 5., Hald has constructed extra analytic constants of the motion besides energy and estrophy for four different truncations. This work appears in [24]. He works with the vorticity representation (1.13) of the equations. The independent modes for the four models are

(6.2) (I) "Fundamental Triad Interaction" - any three modes which satisfy

$\underline{k} + \underline{p} + \underline{q} = 0.$

(II) (0,1), (2,1), (2,0), (2,-1).

(III) (1,1), (2,1), (3,0), (2,-1), (1,-1).

(IV) (0,4), (1,3), (1,1), (3,1), (4,0), (3,-1), (1,-1), (1,-3).

These models are shown explicitly to contain three, six, eight, and four independent isolating integrals of the motion (including energy and enstrophy) respectively. Hence, all are nonergodic on the intersection of the surfaces of constant energy and constant enstrophy. The system (III) was chosen for numerical study in this paper; also, the six-mode system used in our numerical work was obtained from (III) by the addition of

one other mode. The eight constants of motion for (III) are given explicitly by

$$(6.3) \qquad \gamma_1 = 35|\xi_1|^2 + 8|\xi_4|^2, \quad \gamma_2 = 35|\xi_5|^2 + 8|\xi_2|^2,$$

$$\gamma_3 = 35|\xi_3|^2 + 27|\xi_2|^2 + 27|\xi_4|^2,$$

$$I_1 = \xi_1\xi_2^* + \xi_4^*\xi_5, \quad I_2 = 35\,\xi_1\xi_5^* - 8\xi_2\xi_4^*,$$

$$I_3 = \mathrm{Im}(\xi_1\xi_3^*\xi_4 + \xi_2^*\xi_3\xi_5^*).$$

Here, $\xi_1 \equiv \xi_1(1,1), \cdots, \xi_5 \equiv \xi_5(1,-1)$ denotes the Fourier-transformed vorticity field. It may be shown that $E = (\gamma_1 + \gamma_2)/140 + \gamma_3/630$ and that $\Omega = (\gamma_1 + \gamma_2 + \gamma_3)/35$. The polynomials in (6.3) are identical to the $P_j(\underline{\omega})$ in (5.18); of course, I_1 and I_2 each contain two real polynomial constants of the motion.

In view of this work, the question arises as to how far these results can be generalized to larger truncations. The relative ease in locating the constants of the motion in the above four systems is due to the simplicity of the resulting dynamical systems. Indeed, the six-mode system studied here is considerably more complex than any of the systems listed in (6.2) and our ten-mode system is far more complex. Lee [45] argues that Hald's truncations are "special" and that the isolating integrals must disappear for more general truncations. Also, Lee [44] presents an argument, based on coupling triad interactions, that the only quadratic constants of the motion for a sufficiently general truncation are energy and enstrophy. This does not contradict Hald's result in (6.21) because

the extra integral constructed there is a cubic. On the other hand, Lee only considers quadratics without cross terms (i.e., in our notation, if $\omega_i\omega_j$ appears then $i = j$) and does not consider cubics or higher order expressions at all. Yet, as already indicated, Hald has constructed a cubic isolating integral for the fundamental triad interaction. See also [42] concerning Lee's work.

We have already discussed in Section 4 the numerical results due to Fox and Orszag [20] and Deem and Zabusky [16]. These papers study truncations large enough to be of physical significance and the numerical work accordingly concentrates on determining the dynamical evolution of the energy spectrum and deciding whether or not these relax to the predicted equilibrium ensemble deduced in equation (4.15). More complicated observables of the system are not considered. As we have noted, the actual results are possibly contradictory (a positive viscosity is used in [16] so that a literal comparison cannot be made) and certainly not conclusive. Nevertheless, both papers conclude that time averages of the energy spectrum converge to an equilibrium with the form of the equilibrium dependent on the E/Ω ratio. See also [14] concerning these two papers.

In addition to [9], the numerical work which is the most closely related to ours is that of Basdevant and Sadourney [1]. The truncation of equation (1.21) which they use is relatively large-scale; the 16×16 lowest Fourier modes are retained in the computation. To compute phase means, they evaluate the asymptotic limit of the integrals involved as the number of modes approaches infinity. They test these values against a Monte Carlo evaluation of the integrals for the smaller 8×8 problem to ensure validity. The integration of the system of differential equations

is done with the leap frog scheme (see [22]) with a time average of odd and even solutions every 100 time steps. This is a second order scheme and the integration is performed for approximately 10^5 time steps. Hence, as the local truncation error is $O(h^2)$, we may infer that the step size could have been no greater than approximately 10^{-3} which yields a total integration time of about 10^2. The author observed a relative damping of order 10^{-8} per time step in the values for energy and enstrophy. The only observables for which convergence is tested are the quadratics associated with the energy spectrum. Their numerical results yield a time-averaged spectrum which is virtually indistinguishable from their computed phase ensemble. Of course, one cannot conclude that (1.21) - (1.22) is ergodic from this information alone. Nevertheless, the result is significant because it allows the computation of the energy and vorticity spectra in the phase ensemble followed by a substitution of phase means for time means.

The observed dissipation of energy and enstrophy in [1] closely parallels the corresponding figures in our calculations and is certainly acceptable. Still, we question the accuracy of the calculations. As noted in Section 5, we have found it necessary to use a local truncation error no larger than 10^{-7} in order to obtain accurate trajectories for total integration times of order 10 or 100. These times correspond to $\sim 10^3$ time steps in our calculations. Assuming a step size of order 10^{-3} in [1], we see that the local truncation error is somewhat larger than 10^{-7} and the number of steps is larger by two orders of magnitude. In addition, the Gear routine uses a sixth order method on average whereas the leap-frog scheme is only second order. The combination of these factors

0 0 4 3 4 8 0 0 8 7 9

leads us to conclude that the numerical integrations in [1] must leave the real trajectory approximately every 10^3 time steps or less. If this is indeed the case, the authors of [1] are really comparing two phase space integrations and their results are hardly surprising. However, our objections are conjectural since we do not know the time step used and have not tested their difference scheme on our problem. Moreover, it is conceivable albeit unlikely that the 16×16 problem is more well-behaved numerically than the small truncations that we ahve studied.

7. Numerical Results

We have studied three dynamical systems of the form (1.21) using the techniques described in Section 5. The parameters describing these three systems are as follows where M is the number of retained modes, $k_1, \cdots, k_M$ are the independent modes retained, E = energy, Ω = total squared vorticity, T is the final integration time, and ε is the maximum allowable local truncation error:

(7.1) (I) $M = 10$,

$k_1 = (0,2)$, $k_2 = (0,1)$, $k_3 = (1,2)$, $k_4 = (1,1)$, $k_5 = (1,0)$,

$k_6 = (1,-2)$, $k_7 = (1,-2)$, $k_8 = (2,1)$, $k_9 = (2,0)$, $k_{10} = (2,-1)$,

$E = 3.0$, $\Omega = 10.0$,

$T = 30.0$, $\varepsilon = 10^{-10}$.

(II) $M = 5$,

$k_1 = (1,1)$, $k_2 = (2,1)$, $k_3 = (2,-1)$, $k_4 = (1,-1)$, $k_5 = (3,0)$,

$E = 2.0$, $\Omega = 5.0$,

$T = 50.0$, $\varepsilon = 10^{-7}$.

(III) $M = 6$

$k_1 = (1,1)$, $k_2 = (2,1)$, $k_3 = (2,-1)$,

$k_4 = (1,-1)$, $k_5 = (1,0)$, $k_6 = (3,0)$,

$E = 1.0$, $\Omega = 5.0$,

$T = 45.0$, $\varepsilon = 10^{-8}$.

As noted previously, (I) arises from a truncation of the Navier-Stokes equations (1.1) - (1.2) whereas (II) and (III) arise from truncations of the vorticity transport equation (1.4). Observe that the number of independent variables is twice the value of M.

We have also studied the dynamical systems (I'), (II'), and (III') which are identical to (I), (II), and (III) except that the local truncation error ε is larger. The value $\varepsilon = 10^{-3}$ was used in all three cases.

Two models of the "computer game" described in Section 5 were also studied. These may be described as follows:

(7.2) (IV) $M = 4$,

$E = 1.0$, $A = B = 0$, $T = 5000$.

(V) $M = 8$,

$E = 1.0$, $A = B = 0$. $T = 5000$.

Here, M is the number of particles, E is the energy of the system, A and B are the linear momenta, and T is the total number of time steps which were observed. Again, the number of independent variables is twice the value of M. Note that it is not possible to correlate a time step with a physically meaningful time increment.

All calculations were performed on the CDC 7600 machine at Lawrence Berkeley Laboratory using the FTN4 (OPT = 2) Fortran compiler. In Table I, we list some basic data concerning each of the problems (I) - (V). The first four columns refer to a single sample point integrated from $t = 0$ to $t = T$. The fifth column refers to a sample of one hundred points integrated from $t = 0$ to $t = T$. The last two columns consider a sample of 1000 Monte Carlo evaluations. The computer processing times indicated include time spent on various computations not directly related to the integrations or the Monte Carlo routines. Thus T_1 and T_2 should be multiplied by some fraction, about 3/4. The exact multiple is not known. The figures in Table I are based on a random sample; however, the values would not differ significantly if another sample were used.

From the second column of Table I, we see that the Gear routine found Problem II to be significantly more difficult than either I or III

in the sense that it was forced to make the step sizes relatively small. Recall also that the local truncation error for Problem II is 10^{-7} which is the largest of the three. Additionally, the system of differential equations to be solved in Problem I is considerably more complex than those in Problem II and III.

Conservation of energy and enstrophy is clearly satisfactory for all the unprimed problems as the third and fourth columns of Table I indicate. Again, the worst behsvior is exhibited by Problem II. Note that the computer games (IV) and (V) conserve all three known integrals almost exactly. Problem I behaves similarly in this regard.

The last column of the table indicates how many points it was necessary to sample in order to obtain 1000 points in the region E given by equation (5.8). It would be dangerous to try and draw general conclusions from the comparison of these figures as there was no well-defined algorithm for selecting the relative size of E in the surface S.

The numbers T_1 and T_2 from the table give some timing figures. It can be seen that the Monte Carlo evaluations are reasonably inexpensive (despite the large values of N in the last column). On the other hand, the integration of a single point using the Gear routine requires over one C.P. second.

For each of the problems (I) - (V), we have performed 10^3 integrations; thus, $N = 10^3$ in equation (5.4). Each of the Monte Carlo evaluations involved 10^5 sample points; thus $N = 10^5$ in equation (5.3). These computations required just under two hours of computing time. We note that the Cesàro sums (5.2) were not computed in all of the 10^3 integrations as this would have significantly increased the computation

time while little useful information would have been obtained.

Let us comment briefly on the accuracy of the Monte Carlo computations. Since the exact values for all of the integrals which were evaluated are not known, we cannot make any definitive claims. However, these values were known a priori in some special cases. In particular,

$\int_S \omega_i d\mu(\underline{\omega}) = 0$ for any choice of i and for each of the problems (I) - (V). The largest error observed for the approximation (5.3) is 1.18×10^{-2} for all i and all problems. Typically, the error is less than 10^{-3} and errors below 10^{-4} are not unusual. We have also noted that these errors for Problem V are about an order of magnitude less than for the other problems (including Problem IV). It is also known that

$\int_S \omega_i^2 d\mu(\underline{\omega}) = 0.125$ for Problem IV and 0.0625 for Problem V. The largest error observed for these integrals is less than 4×10^{-4}. Next, with $h(\underline{\omega})$ as in (5.14), we know that the corresponding integral must have the value 0.5 by symmetry (for all problems). The largest error observed here is less than 5×10^{-3}.

On the basis of these figures, we will assume that the approximations (5.3) are valid to two significant figures. We note that these error estimates would be somewhat worse if $N = 10^4$ in (5.3) and that they would be considerably worse for $N = 10^3$. This, of course, will be important in evaluating the results for the approximation (5.4).

We first consider the results for Problems I', II', III' in which the local truncation was selected at an unrealistically high level for the integration times used. The results were disappointing. In all cases, the behavior of the ratios (5.24) and (5.26) were worse than the

corresponding behavior for the unprimed problems in the sense that there is less convergence apparent. This occurred despite the fact that the primed problems were deliberately chosen to be somewhat more random than the unprimed problems and should, therefore, exhibit faster convergence for (5.24) and (5.26). We conjecture that these results can be explained by the fact that the primed problems did not conserve energy and enstrophy to a significant extent and that this factor outweighed the a priori randomness of the models. In any event, we shall not present the detailed numerical results for the primed problems.

We shall now present the results concerning the ratio (5.24) in detail for three sample points, denoted Q_1, Q_2, and Q_3, for each problem (I) - (V). First, we have observed experimentally that $1 - (\mu(E)/\mu(S)) < 10^{-3}$ in the sense that it is unusual for $F_t\underline{\omega} \notin E$ for more than one or two time steps in a typical integration; further, one usually finds $F_t\underline{\omega} \in E$ during the entire run. Thus, we shall take $\mu(E)/\mu(S) = 1$ in the sequel, for simplicity. Let $A'(f)$ denote the left-hand side of (5.24) and define

$$(7.3) \qquad A_1(f) = |1 - A'(f)|.$$

Of course, $A_1(f)$ is also a function of time; it measures the rate of convergence of the Cesàro sum for f in the sense that $A_1(f) \to 0$ as $t \to \infty$ if and only if the Cesàro sum approaches the ensemble average. Denoting $f(\underline{\omega}) = \chi_E(\underline{\omega})\, h(\underline{\omega})$, we have studied $A_1(f)$ for all choices of $h(\underline{\omega})$ listed in (5.9) - (5.18). However, we shall avoid listing results for those $h(\underline{\omega})$ for which $\int_S f(\underline{\omega})\, d\mu(\underline{\omega}) = 0$ since the numerical error in these cases is far greater

than the phenomena that we are interested in. We now list those functions $h(\underline{\omega})$ that were useful:

(7.4) $h_{1,i}(\underline{\omega}) = |\omega_i|$, $h_{2,i}(\underline{\omega}) = \omega_i^2$, $h_{3,i}(\underline{\omega}) = |\omega_i^3|$,

$h_{3,i}(\underline{\omega}) = \omega_i^4$, $h_{5,i,j}(\underline{\omega}) = |\omega_i\omega_j|$, $h_{6,i,j}(\underline{\omega}) = \omega_i^2\omega_j^2$,

$h_{7,i}(\underline{\omega}) = h(\underline{\omega})$, h as in (5.13),

$h_{8,i}(\underline{\omega}) = h(\underline{\omega})$, h as in (5.14),

$h_{9,i}(\underline{\omega}) = h(\underline{\omega})$, h as in (5.15),

$h_{10,i,j}(\underline{\omega}) = h(\underline{\omega})$, h as in (5.16),

$h_{11,i}(\underline{\omega}) = h(\underline{\omega})$, h as in (5.17),

$h_{12,i}(\underline{\omega}) = h(\underline{\omega})$, h as in (5.18).

The index i usually ranges from 1 to M (= dimension of phase space). However, for $h_{12,i}$, i only takes the values 1,2, and 3 as the other five polynomials have integrals close to zero; also, for the function $h_{5,i,j}$, $h_{6,i,j}$, and $h_{7,i,j}$, we did not compute Cesàro sums for the full range of indices - only 21 choices of (i,j) were made. We now define

(7.5) $A_2(h_k) = (\#I_k)^{-1} \ \Sigma_{i\varepsilon I_k} A_1(h_{k,i})$

for k = 1, ..., 12. Here, I_k is the set of all indices for which $h_{k,i}$ (or $h_{k,i,j}$) was studied and $\#I_k$ is the number of elements in I_k. For example, $\#I_1 = M$, $\#I_{12} = 3$ and $\#I_5 = 21$ for all problems. Thus, $A_2(f)$ represents an average rate of convergence over a class of similar functions. Finally, we define

$$(7.6) \qquad A_3 = \left(\sum_{k=1}^{12} \#I_k \right)^{-1} \sum_{k=1}^{12} (\#I_k) \cdot A_2(h_k)$$

which is a weighted average over the $A_2(h_k)$. Thus, A_3, which is a function only of time and the problem under study, represents an average convergence rate over all the functions studied.

Figure 1 is a graph of the composite averages A_3 versus time. $A_3(Q_1)$, $A_3(Q_2)$, and $A_3(Q_3)$ are depicted in Fig. 1a, 1b, and 1c, respectively. Each figure contains graphs of each of the five problems under study (needless to say, the Q_i are functions of the problem). We make the convention, which will remain in force for all of the figures, that one hundred time steps (in Problems IV and V) equals one second (in Problems I, II, and III). This has no significance besides enabling us to draw all five graphs in one figure.

Figures 2-5 contain graphs of $A_2(h_k)$ (in (a) of each figure) and $A_1(h_{k,M})$ (in (b) of each figure) for k = 2,3,4,6,8,12 for Problems I - IV, respectively (except k = 12 is not graphed for Problem IV because it was not computed). Only results for the point Q_1 are presented. Observe that the scales vary from figure to figure. Graphs of this type for Problem V would have been similar to those for Problem IV. These figures are meant to illustrate how the functions $A_1(f)$ and $A_2(f)$ depend on f. Also, by comparing $A_2(h_k)$ with $A_1(h_{k,M})$ within one of the figures, one sees immediately that $A_1(h_{k,i})$ varies considerably with i.

In Fig. 6, we graph $A_2(h_3)$ and $A_1(h_{3,M})$ for Problem I. This figure illustrates the dependence of $A_1(f)$ and $A_2(f)$ upon the initial point. It is clear that, at least for A_1, this dependence may be considerable.

It should be borne in mind when reading these graphs that each of $h_{12,i}$ for $i = 1,2,3$ is an integral of the motion for Problem II but not for the other problems.

The most striking feature of Fig. 1 is that the graph of the composite average A_3 for Problem II is virtually a horizontal line for all three points. This property extends to the averages A_2 in Fig. 3a and, to a large extent, in the individual Cesàro sums exhibited in Fig. 3b. Of course, it was known in advance that Problem II was not ergodic and, therefore, we knew that we would not see convergence to zero for these averages. Still, the lack of oscillation in these curves comes as a surprise and is quite a contrast relative to the other problems. It implies that the Cesàro sums studied are converging rapidly (but not to zero, of course). For Problem III, we see convergence initially for all three points in Fig. I followed either by slight convergence or slight divergence depending on the point. Thus, Fig. I is evidence for non-ergodicity of Problem III, although this is not conclusive. The nearly horizontal curves of Fig. 4 are stronger evidence in this regard. Problem I is the hardest to draw any conclusions about. On the one hand, we see rapid initial convergence for all three points in Fig. 1; but, on the other hand, this is followed by a gradual leveling off in all three cases. It would have been of use here to have had a somewhat larger final integration time. For the most part, Fig. 2 bears out this problem of interpretation even further. Although it is clear from Fig. 1 that Problem I is more random than either Problem II on Problem III, we do not believe that there is sufficient evidence to say that Problem I is ergodic.

Finally, we turn to Problems IV and V. As Fig. 1 illustrates, the behavior of A_3 for these two problems is essentially the same. It is apparent that the oscillations in the Cesàro sums have already been damped out prior to time step 500 and that convergence to zero is rapid. Furthermore, convergence is still taking place at time step 5000. These conclusions are further verified in Fig. 5. This is strong numerical evidence that Problems IV and V are ergodic.

Another important feature of the results is the large disparity between Problems I, II, III and Problems IV, V. We believe that this can be explained by the lack of a time scale in the last two problems and the fact that the trajectories for the first three problems must be continuous. It is also conceivable that the error in the approximation (5.2), which is zero for the last two problems, plays a role but this isn't likely. In any event, it is invalid to declare Problem I to be nonergodic simply because it doesn't converge as "fast" as Problems IV or V.

We now present the numerical results for the ratio (5.26) which bears on the question of mixing. The functions f and g to be considered are given by (7.4); the reason for excluding the others is the same as given above. In analogy to A_1, we define

$$(7.7) \qquad B_1(f,g) = |1 - B'(f,g)|$$

where $B'(f,g)$ is the ratio (5.26) for f and g. Let us define four composite averages B_2, B_3, B_4, B_5. B_2 is to be the average over those (f,g) for which one of the pair is simply χ_E. B_3 encompasses those (f,g) for which f = g or f,g are both functions of the same coordinate

(e.g., $f(\underline{\omega}) = |\omega_5|$ and $g(\underline{\omega}) = \omega_5^2$). B_4 consists of the cross correlations where f,g are functions of different coordinates. B_5 is a weighted average of B_3 and B_4.

Figure 7 consists of the graphs for the B_k, $k = 2,\cdots 5$ for each of the Problems I-V in turn. Figure 8 contains graphs for five samples of the ratios (5.26) where the following notation is used:

$$(7.8) \quad B_1(1) = B_1(f,g) : \; f = \omega_M^4, \; g = \chi_E(\underline{\omega})$$

$$B_1(2) = B_1(f,g): \; f = P_1(\underline{\omega}), \; g = \chi_E(\underline{\omega})$$

$$B_1(3) = B_1(f,g): \; f = \omega_M^2, \; g = \omega_M^2$$

$$B_1(4) = B_1(f,g): \; f = g \text{ as in } (5.17), \; i = M$$

$$B_1(5) = B_1(f,g): \; f = \omega_{M-1}^2, \; g = \omega_M^2 .$$

Since the trajectories $\{F_t\underline{\omega} = 0 \le t \le T\}$ lie almost entirely within E, it is approximately true that χ_E is the identity function. Hence, B_2 and $B_1(1)$, $B_1(2)$ measure how well the flow $\{F_t\}$ conserves μ-measure on S to a good approximation and do not test the mixing hypothesis to a significant degree. As μ-measure is known to be conserved, we can use the results for these functions in determining the error inherent in the approximmation (5.4) with $N = 10^3$. In particular, any graph which crosses below B_2, $B_1(1)$, or $B_1(2)$ indicates convergence for the ratio or average being tested; a graph which stays significantly above would indicate lack of convergence.

On the basis of Fig. 7 alone, it is hard to draw any conclusions concerning Problems I, II, III. Although B_3, B_4, and B_5 remain above B_2 in the entire range $0 \le t \le T$, the difference in amplitudes is not

significant. On the other hand, one could not expect even a known mixing transformation to exhibit faster convergence than that shown in Fig. 7d and 7e. Thus, it is tempting to conclude that Problems IV and V are mixing on the basis of the numerical evidence. This conclusion is buttressed in Figs. (8d) and (8e) in which the amplitudes and oscillations of the four graphs are similar; this indicates that whatever nonconvergence is present is due solely to the approximation error in (5.4). We now turn to Figs. 8a, 8b, and 8c. In all three cases, two of the mixing ratios (i.e., $f,g \neq \chi_E$) lie well above the baseline given by $B_1(1)$ and $B_1(2)$. Furthermore, neither of these curves gives any indication that it would converge to zero if the final integration time were increased. This is evidence for nonmixing in Problems I, II, and III. Observe that this evidence is of virtually equal weight for the three problems.

In conclusion, the numerical results provide strong evidence for the ergodic hypothesis and some evidence for the mixing hypothesis in Problems IV and V. Problem II is clearly nonergodic on the basis of the numerical results alone. Also, Problem III is almost surely nonergodic. We cannot say one way or the other for Problem I without a larger integration time. All of Problems I, II, and III appear to be nonmixing. However, this result is flawed by the low value of $N = 10^3$ in the approximation (5.4); recall that $N = 10^4$ was necessary in the integration for (5.3) and there is every reason to suspect that (5.4) would require even more sample points.

8. Conclusions

We now discuss the feasibility of extending the techniques of Section 5 to the study of larger scale truncations, in view of the numerical results presented in Section 7.

The feature of our results that causes the greatest difficulty of interpretation is the dependence of our convergence graphs for the ratios (5.24) and (5.26) upon the function being studied. In the case of (5.24), the dependence upon the initial point is also important. We have seen that this dependence is highly nontrivial. For example, consider the sequence of functions $|\omega_i|\chi_E$, $\omega_i^2\,\chi_E$, $|\omega_i^4|\chi_E$. These graphs all have the same qualitative features; however, their amplitudes form a monotonically increasing sequence. It is tempting to say that merely by increasing our final integration time, we could have conclusively settled the convergence issue for these four ratios. This is probably true, but the question then arises what would happen for the functions $|\omega_i|^5\chi_E$, $\omega_i^6\chi_E$, etc? Presumably longer and longer integration times would be required. In other words, for a numerical experiment involving a finite integration time, however large, it is impossible to obtain conclusive results about convergence for an infinite set of functions. A similar analysis applies to the variation with respect to initial data.

This type of limitation upon the interpretation of results was to be expected a priori since our technique involves replacing infinite limits by finite procedures. Despite these limitations, we have been able to conclude on the basis of the numerical evidence that Problem I is considerably more random than either Problems II or III, that Problem II is definitely not ergodic, and that Problem III is probably

not ergodic. Furthermore, a strong case has been presented that Problems IV and V are mixing. Also, these questions could probably have been settled for Problem I if a somewhat longer integration time had been used (which could easily have been arranged by using double precision arithmetic in the differential equation solver).

Perhaps the key question in extending these results to larger problems is how do the convergence rates depend on the size of the truncation? If they become much slower than those observed for Problem I, the extension of our methods is certainly infeasible. As indicated in Section 6, the numerical results of Basdevant and Sadourney [1] indicate that the opposite may take place. In any event, this factor will have to be thoroughly tested.

We have not observed any reason why our Monte Carlo techniques for the evaluation of the approximation (5.3) cannot be extended to larger problems. The same holds for the computation of the density function $\rho(\underline{\omega})$ and for the construction of the region $E \subseteq S$. Of course, the computing effort will be greater but not insuperably so.

Thus, there is every reason to believe that interesting results can be obtained concerning the ergodicity hypothesis by using our numerical techniques on larger truncations. We cannot say the same for the study of the mixing hypothesis. It appears from our results that the interpretation of the numerical figures arising from the ratio (5.26) will be much more difficult than those from (5.24). For example, consider the similarity in Figs. 7 and 8 between the graphs for Problems I, II, and III. It is quite possible that these difficulties would have disappeared had we taken $N = 10^5$ or at least $N = 10^4$ in the approximation

(5.4). This would have been just barely possible in terms of the computing effort involved; for a very large problem, it unfortunately would not be so. Nevertheless, we see no difficulties in principle in extending our techniques.

As we have just observed, one can never numerically test for convergence of Cesàro sums of all possible functions in a finite time. This, however, should not present a barrier to investigating the physical phenomena in which we are interested, in particular, macroscopic vortex formation. What is required is the construction of a finite collection of observables which, at least approximately, describes the qualitative features of the vorticity field. With these in hand, one could then perform the numerical tests as in Section 7 and decide whether or not the relevant Cesàro sums converge to ensemble averages. The answer to this question would, of course, be significant in formulating a theoretical explanation for macroscopic vortex formation.

We have remarked in Section 4 that a dynamical system can be exhibit highly random behavior and even support a thermodynamic analysis without being even ergodic. Thus, it is not necessary from a physical point of view to definitively decide the ergodic hypothesis for a particular dynamical system. What is important is the careful study of those observables which define the phenomena of interest. We believe the numerical techniques presented here can be useful in such a study.

9. References

1. C. Basdevant and R. Sadourney, "Ergodic Properties of Inviscid Truncated Models of Two-Dimensional Incompressible Flows," J. Fluid Mech., 69, (1975), pp. 673-688.
2. G. K. Batchelor, The Theory of Homogeneous Turbulence, (1960), Cambridge Univ. Press, London.
3. G. K. Batchelor, An Introduction to Fluid Dynamics, (1967), Cambridge Univ. Press, London.
4. G. Benettin, L. Galgani, and J.-M. Strelcyn, "Kolmogorov Entropy and Numerical Experiments," Phys. Rev. A, 14, (1976), pp. 2338-2345.
5. G. Casati and J. Ford, "Computer Study of Ergodicity and Mixing in a Two-Particle, Hard Point Gas System," J. Comp. Phys., 20, (1976), pp. 97-109.
6. A. J. Chorin, "Computational Aspects of the Turbulence Problem," Proc. 2nd Int. Conf. Num. Meth. Fluid Mech., (1970), Springer-Verlag, New York.
7. A. J. Chorin, "Numerical Study of Slightly Viscous Flow," J. Fluid Mech., 57, (1973), pp. 785-796.
8. A. J. Chorin and P. S. Bernard," Discretization of a Vortex Sheet with an Example of Roll-Up," J. Comp. Phys., 13, (1973), pp. 423-429.
9. A. J. Chorin, "Numerical Experiments with a Truncated Spectral Representation of a Random Flow," unpublished.
10. A. J. Chorin, Lectures on Turbulence Theory, (1976), Publish or Perish, Boston.
11. A. J. Chorin, Numerical Methods in Statistical Hydrodynamics, (1977), Les Presses de l'Université de Montréal, Montreal.

0 0 0 0 4 8 0 5 8 8 7

12. J. P. Christiansen, "Numerical Simulation of Hydrodynamics by the Method of Point Vortices," J. Comp. Phys., 13, (1973), pp. 363-379.

13. J. P. Christiansen and N. J. Zabusky, "Instability, Coalescence and Fission of Finite-Area Vortex Structures," J. Fluid Mech., 61, (1973). pp. 219-243.

14. I. Cook and J. B. Taylor, "Stationary States of Two-Dimensional Turbulence," Phys. Rev. Lett., 28, (1972), pp. 82-84.

15. R. Courant, Differential and Integral Calculus, Vol. II., (1936), Wiley-Interscience, New York.

16. G. S. Deem and N. J. Zabusky, "Ergodic Boundary in Numerical Simulations of Two-Dimensional Turbulence," Phys. Rev. Lett., 27, (1971), pp. 391-399.

17. D. G. Ebin and J. Marsden, "Groups of Diffeomorphisms and the Motion of an Incompressible Fluid," Ann. of Math., 92, (1970), pp. 102-163.

18. C. Foias, "Ergodic Problems in Functional Spaces Related to the Navier-Stokes Equations," Proc. Int. Conf. Funct. Anal.rel. Topics, (1969), pp. 290-303.

19. J. Ford, "The Statistical Mechanics of Classical Analytic Dynamics," Fundamental Problems in Statistical Mechanics, Vol. 3, (1975), North-Holland Pub. Co., Amsterdam.

20. D. G. Fox and S. A. Orszag, "Inviscid Dynamics of Two-Dimensional Turbulence," Phys. of Fluids, 16, (1973), pp. 169-171.

21. D. Fyfe and D. Montgomery, "High-Beta Turbulence in Two-Dimensional Magnetohydrodynamics," J. Plasma Phys., 16, (1976), pp. 181-191.

22. C. W. Gear, Numerical Initial Value Problems in Ordinary Differential Equations, (1971), Prentice-Hall, Englewood Cliffs, N.J.

23. H. Glaz, "Two Attempts at Modeling Two-Dimensional Turbulence," to be published.

24. O. Hald, "Constants of Motion in Models of Two-Dimensional Turbulence," Phys. of Fluids, 19, (1976), pp. 914-915.

25. O. Hald, "On Chorin's Conjecture for the Spectral Model in Turbulence," preprint.

26. P. R. Halmos, Lectures on Ergodic Theory, (1956), Chelsea Publishing Co., New York.

27. P. Hartmann, Ordinary Differential Equations, (1973), John Wiley & Sons, Inc., Baltimore.

28. M. Henon, "Numerical Study of Quadratic Area-Preserving Mappings," Quant. Appl. Math., 27, (1969) pp. 291-312.

29. A. C. Hindmarsh, Gear: Ordinary Differential Equation System Solver, (1974), Lawrence Livermore Lab., UCID-30001, Rev. 3.

30. M. Hirsch and S. Smale, Differential Equations, Dynamical Systems, and Linear Algebra, (1974), Academic Press, New York.

31. E. Hopf, "Statistical Hydromechanics and Functional Calculus," J. Rat. Mech. Anal., 1, (1952), pp. 87-123.

32. K. Huang, Statistical Mechanics, (1963), John Wiley & Sons, New York.

33. G. Joyce and D. Montgomery, "Negative Temperature States for the Two-Dimensional Guiding-Centre Plasma," J. Plasma Phys., 10, (1973), pp. 107-121.

34. G. Joyce, D. Montgomery, and M. Emery, "Electric Field Correlations in the Guiding-Centre Plasma," (1973), preprint.

35. A. I. Khinchin, Mathematical Foundations of Statistical Mechanics, (1949), Dover, New York.

36. D. E. Knuth, The Art of Computer Programming, Vol. 2: Seminumerical Algorithms, (1969), Addison-Wesley, Reading, Mass.

37. T. Koga, "A Kinetic Theory of Turbulence in an Incompressible Fluid," (1968), Dept. of Aeorospace Engineering, Polytechnic Inst. of Brooklyn, Report 68-5.

38. R. H. Kraichnan, "Dynamics of Nonlinear Stochastic Systems," J. Math. Phys., 2, (1961), pp. 124-148.

39. R. H. Kraichnan, "Inertial Ranges in Two-Dimensional Turbulence," Phys. of Fluids, 10, (1967), pp. 1417-1423.

40. L. D. Landau and E. M. Lifschitz, Fluid Mechanics, (1959), Pergammon Press, Oxford.

41. P. D. Lax, "Approximation of Measure Preserving Transformations," Comm. Pure and Appl. Math., 24, (1971), pp. 133-135.

42. J. Lee, "Statistical Mechanical Approaches to Fluid Turbulence," J. Math. Phys., 15, (1974), pp. 1571-1586.

43. J. Lee, "The Triad-Interaction Representation of Homogeneous Turbulence," J. Math. Phys., 16, (1975), pp, 1359-1366.

44. J. Lee, "Isolating Constants of Motion for the Homogeneous Turbulence of Two and Three Dimensions," J. Math. Phys., 16, (1975), pp. 1367-1373.

45. J. Lee, "How Many Isolating Constants of Motion in 2-D Turbulence?," preprint.

46. D. K. Lilly, "Numerical Simulation of Two-Dimensional Turbulence," Phys. of Fluids Supp.II,(1969), pp. II 240-II 249.

47. R. K. C. Lo and L. Ting, "Studies of the Merging of Vortices," Phys. of Fluids, 19, (1976), pp. 912-913.

48. T. S. Lundgren and Y. B. Pointin, "Turbulent Self-Diffusion," Phys. of Fluids, 19, (1976), pp. 355-358.

49. D. Merlini and K. I. Golden, "On the Cut-Off in Two-Dimensional Guiding Center Plasma," Phys. Lett., 60A, (1977), pp. 209-211.

50. D. Montgomery, "Two-Dimensional Vortex Motion and 'Negative Temperatures'," Phys. Lett., 39A, (1972), pp. 7-8.

51. D. Montgomery and G. Joyce, "Statistical Mechanics of 'Negative Temperature' States," (1973), preprint.

52. D. Montgomery and F. Tappert, "Conductivity of a Two-Dimensional Guiding Center Plasma," Phys. of Fluids, 15, pp. 683-687.

53. J. Moser, Lectures on Hamiltonian Systems, (1967), Memoir No. 81, A.M.S., Providence, R.I.

54. J. Moser, Stable and Random Motions in Dynamical Systems., (1973), Princeton Univ. Press, No. 77, Annals of Math. Studies, Princeton.

55. J. Moser, E. Phillips, and S. Varadhan, Ergodic Theory: A Seminar, (1975), Courant Inst. Lecture Notes, New York.

56. L. Onsager, "Statistical Hydrodynamics," Nuovo Cimento, Supp.A1, Vol. VI, Series IX, (1949), pp. 279-287.

57. S. A. Orszag, "Analytical Theories of Turbulence," J. Fluid Mech., 41, (1970), pp. 363-386.

58. S. A. Orszag, "Numerical Simulations of Incompressible Flows within Simple Boundaries. I. Galerkin (spectral) Representations," Stud. in Appl. Math., L, (1971), pp. 293-327.

59. J. C. Oxtoby, Measure and Category, (1970), Springer-Verlag, New York.

60. J. C. Oxtoby and S. M. Ulam, "Measure-Preserving Homeomorphisms and Metrical Transitivity," (1941), Ann. of Math., 42, pp. 874-920.

61. R. Paley and N. Weiner, Fourier Transforms in the Complex Domain, (1934), A.M.S., New York.

62. D. Ruelle and F. Takens, "On the Nature of Turbulence," Comm. Math. Phys., 20, (1971), pp. 167-192.

63. C. E. Seyler, Jr., "Thermodynamics of Two-Dimensional Plasmas or Discrete Line Vortex Fluids," Phys. of Fluids, 19, (1976), pp. 1336-1341.

64. C. E. Seyler, Jr., "Partition Function for a Two-Dimensional Plasma in the Random Phase Approximation," (1973), preprint.

65. C. E. Seyler, Jr., Y. Salu, D. Montegomery, and G. Knorr, "Two-Dimensional Turbulence in Inviscid Fluids or Guiding Center Plasma," (1975), preprint.

66. A. I. Shestakov, Numerical Solution of the Navier-Stokes Equations at High Reynolds Numbers, (1975), Lawrence Livermore Lab., UCRL-51894.

67. Ya. G. Sinai, Introduction to Ergodic Theory, (1976), Princeton Univ. Press, Princeton.

68. J. B. Taylor and B. McNamara, "Plasma Diffusion in Two-Dimensions," Phys. of Fluids, 14, (1971), pp. 1492-1499.

69. J. B. Taylor, "Negative Temperatures in Two-Dimensional Vortex Motion," Phys. Lett., 40A, (1972), pp. 1-2.

70. G. Vahala and D. Montgomery, "Kinetic Theory of a Two-Dimensional Magnetized Plasma," J. Plasma Phys., 6, pp. 425-439.

71. W. W. Wood, "Computer Studies on Fluid Systems of Hard-Core Particles," Fundamental Problems in Statistical Mechanics, 3, (1975), North-Holland Pub. Co., Amsterdam.

0 0 0 0 4 8 0 0 8 9 0

TABLE I.

Problem	T	STEPS	ΔE	$\Delta\Omega$	T_1	T_2	N
I	30	915	10^{-10}	10^{-9}	119	2.62	8019
I'	30	232	10^{-2}	10^{-2}	20	—	—
II	50	2153	10^{-5}	10^{-5}	128	1.07	5909
II'	50	1054	0.05	0.2	44	—	—
III	45	1599	10^{-8}	10^{-7}	132	1.60	8018
III'	45	739	10^{-2}	0.08	36	—	—
IV	5000	—	10^{-10}	$10^{-13(*)}$	44	0.31	—
V	5000	—	10^{-10}	$10^{-13(*)}$	96	0.55	—

T = final integration time.

STEPS = number of time steps used in the integration from $t = 0$ to $t = T$.

ΔE, $\Delta\Omega$ are the actual errors in the values for E, Ω at $t = T$.

T_1 = C.P. seconds needed to integrate 100 points from $t = 0$ to $t = T$.

T_2 = C.P. seconds needed to evaluate 1000 Monte Carlo points.

N = # of random points needed to obtain 1000 points satisfying equation (5.8).

(*) figures refer to the error in the linear momenta, not enstrophy.

FIGURE CAPTIONS

Fig. 1. Composite averages A_3 (see pg. 82) for each of the problems I-V, plotted as a function of time.

1a - Point Q_1 for each problem.

1b - Point Q_2 for each problem.

1c - Point Q_3 for each problem.

Fig. 2. The averages $A_2(h_k)$ (see pg. 81) and the ratios for individual Cesàro sums $A_1(h_{k,20})$ (see pg. 80) for k = 2,3,4,6,9,12, plotted as a function of time for Point Q_1, Problem I.

2a - $A_2(h_k)$ for each k.

2b - $A_1(h_{k,20})$ for each k.

Fig. 3. Same as Fig. 2 except graphs refer to Problem II.

Fig. 4. Same as Fig. 2 except graphs refer to Problem III.

Fig. 5. Same as Fig. 2 except graphs refer to Problem IV and k = 12 is not plotted.

Fig. 6. Comparison of convergence rates by point for Problem I.

6a - $A_2(h_3)$ plotted as a function of time for each of the initial points Q_1,Q_2,Q_3.

6b - $A_1(h_{3,20})$ plotted as a function of time for each of the initial points Q_1,Q_2,Q_3.

Fig. 7. The composite averages B_2,B_3,B_4,B_5 (see pg. 84) plotted as a function of time.

7a - Problem I.

7b - Problem II.

7c - Problem III.

7d - Problem IV.

7e - Problem V.

0 0 0 0 4 8 0 9 8 9 1

Fig. 8. The ratios (5.26), denoted by $B_1(k)$, $k = 1,\cdots,5$ (see pg. 85) plotted as a function of time.

8a - Problem I.

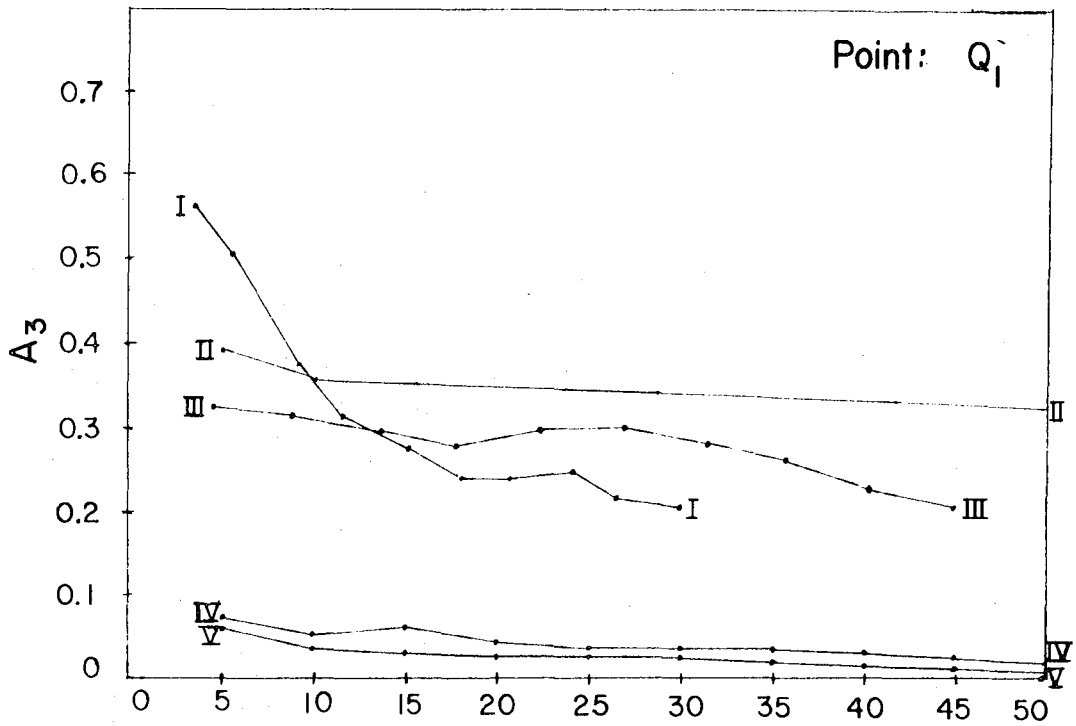
8b - Problem II.

8c - Problem III.

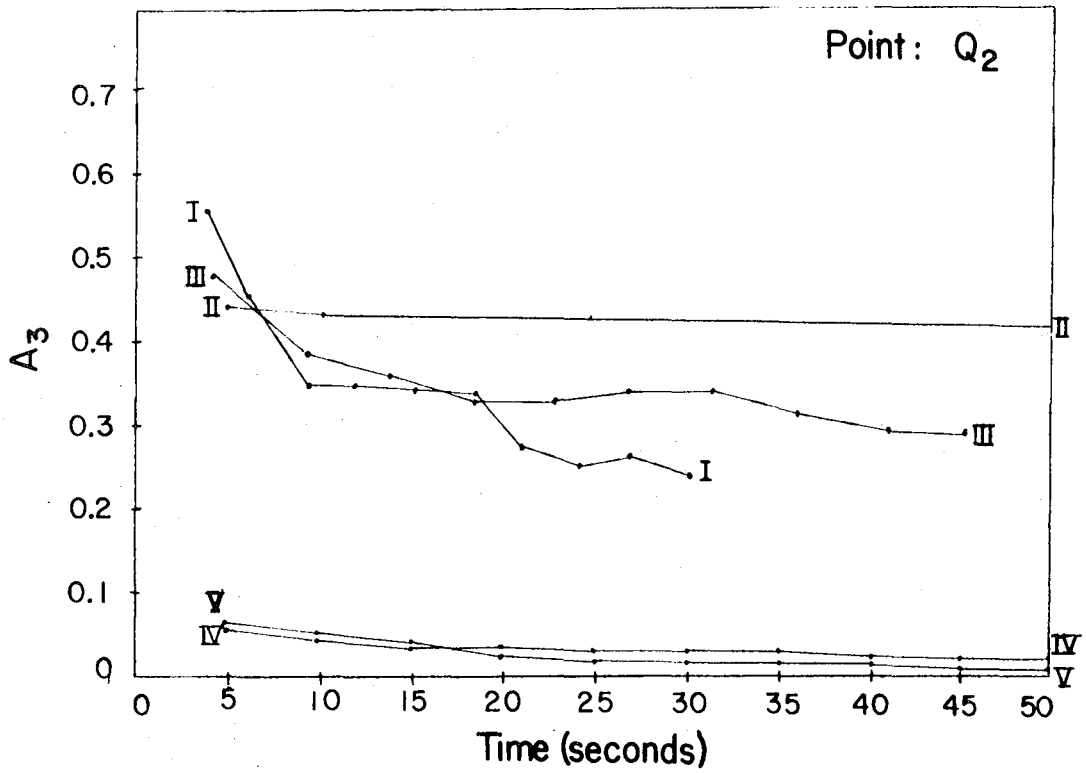
8d - Problem IV, $k = 2$ not included.

8e - Problem V, $k = 2$ not included.

(1a)



(1b)

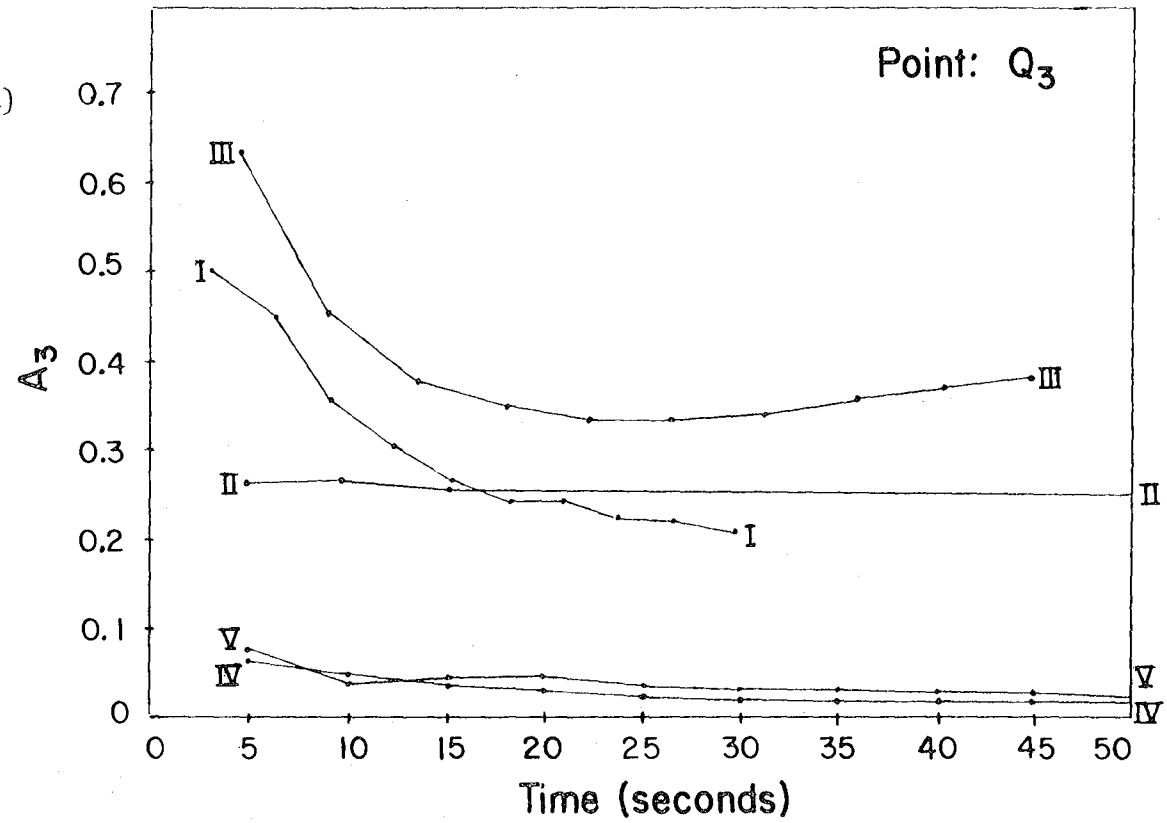


XBL 7710-10039

Fig. 1

0 0 0 0 4 8 0 0 8 9 2

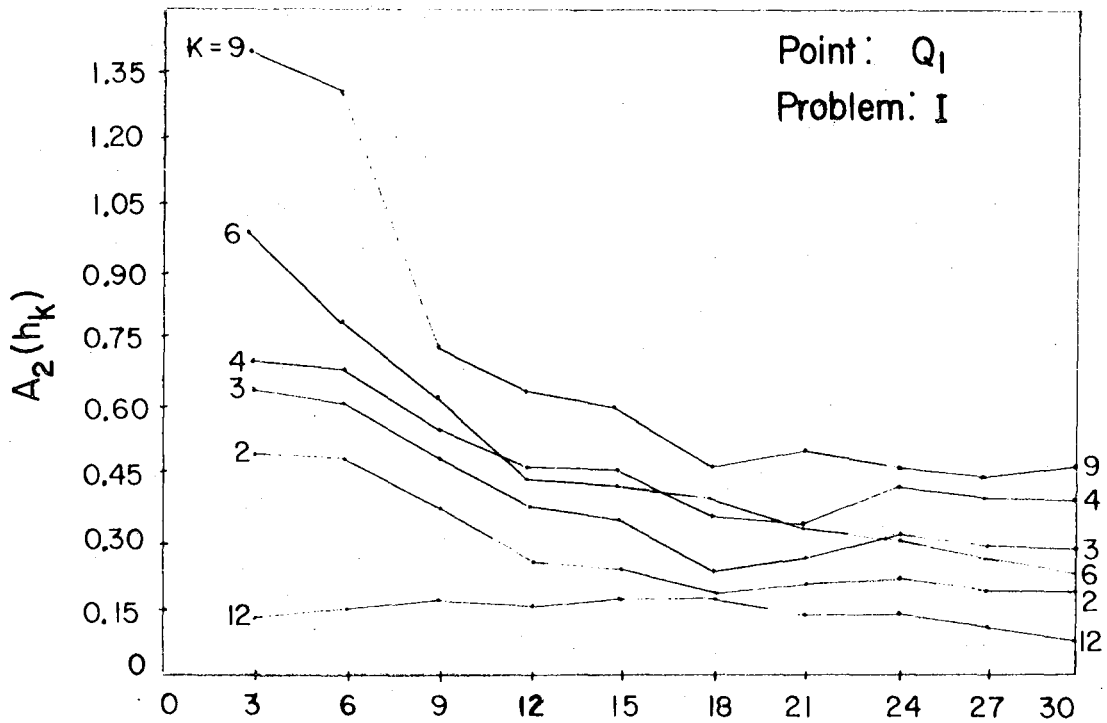
(1c)



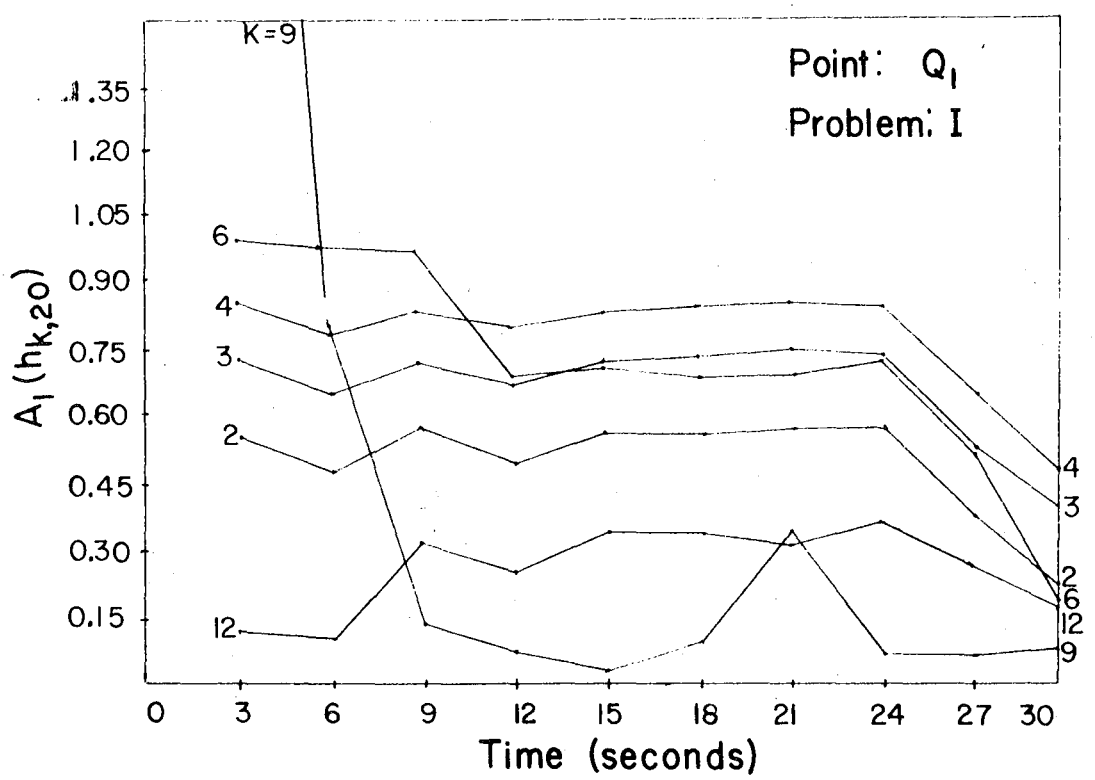
XBL 7710-10038

Fig. 1 (cont.)

(2a)



(2b)

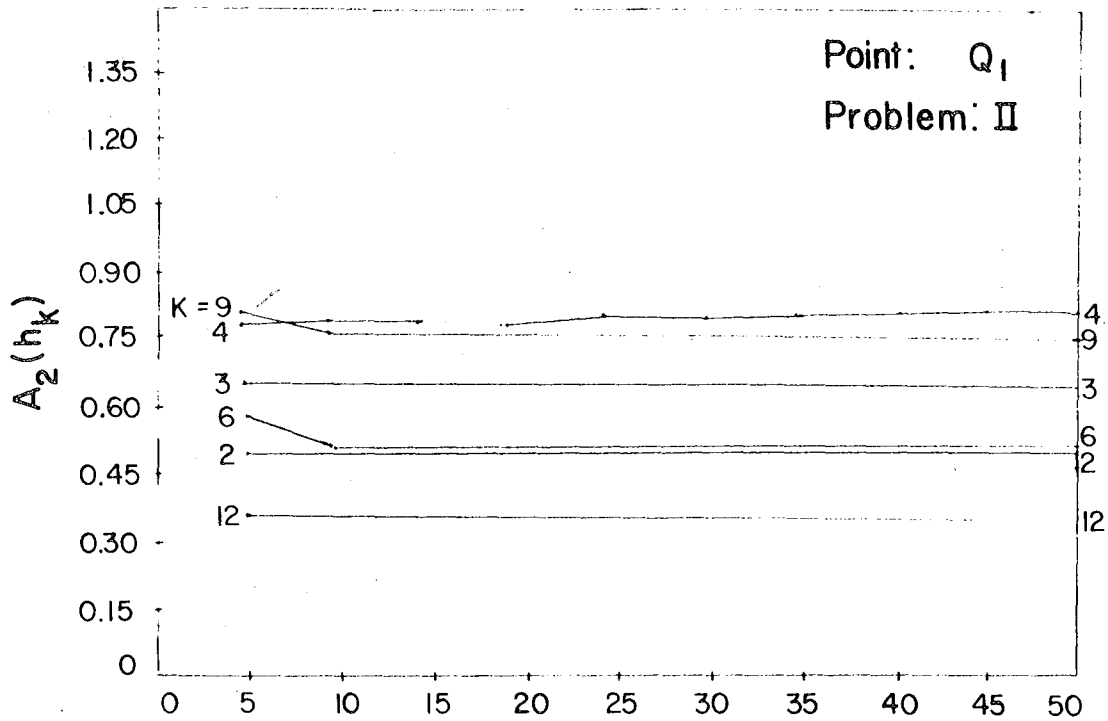


XBL 7710-10037

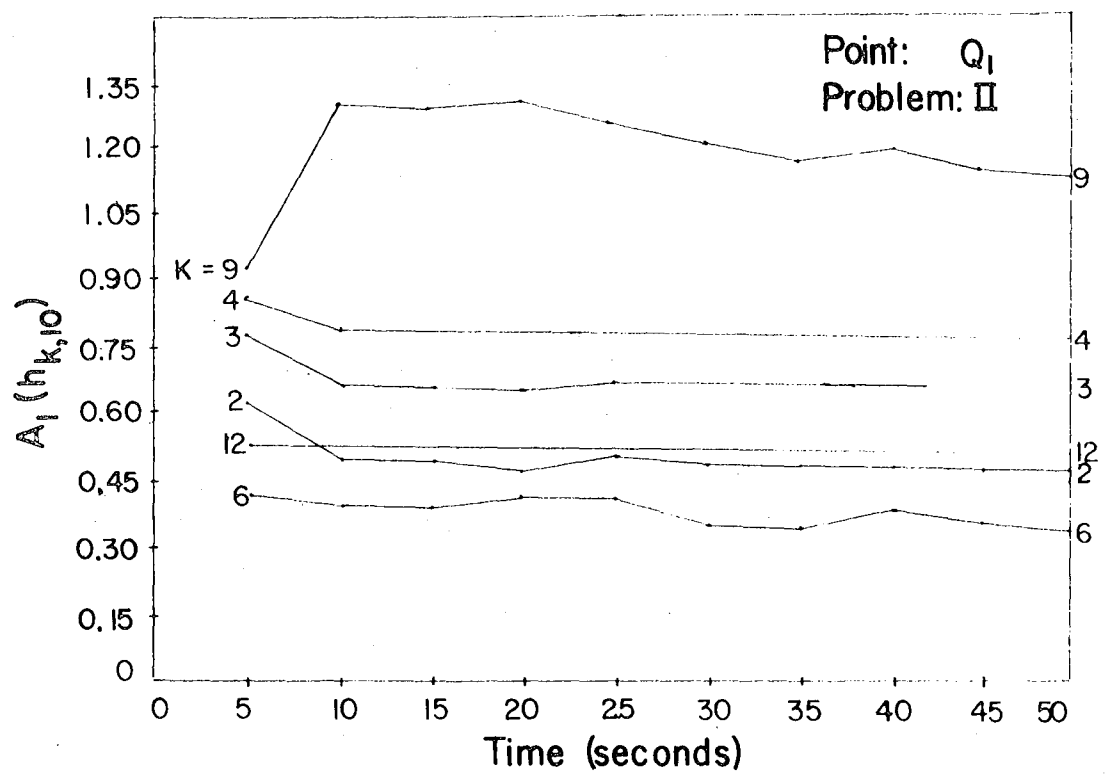
Fig. 2

0 0 0 0 4 8 0 0 8 9 3

(3a)



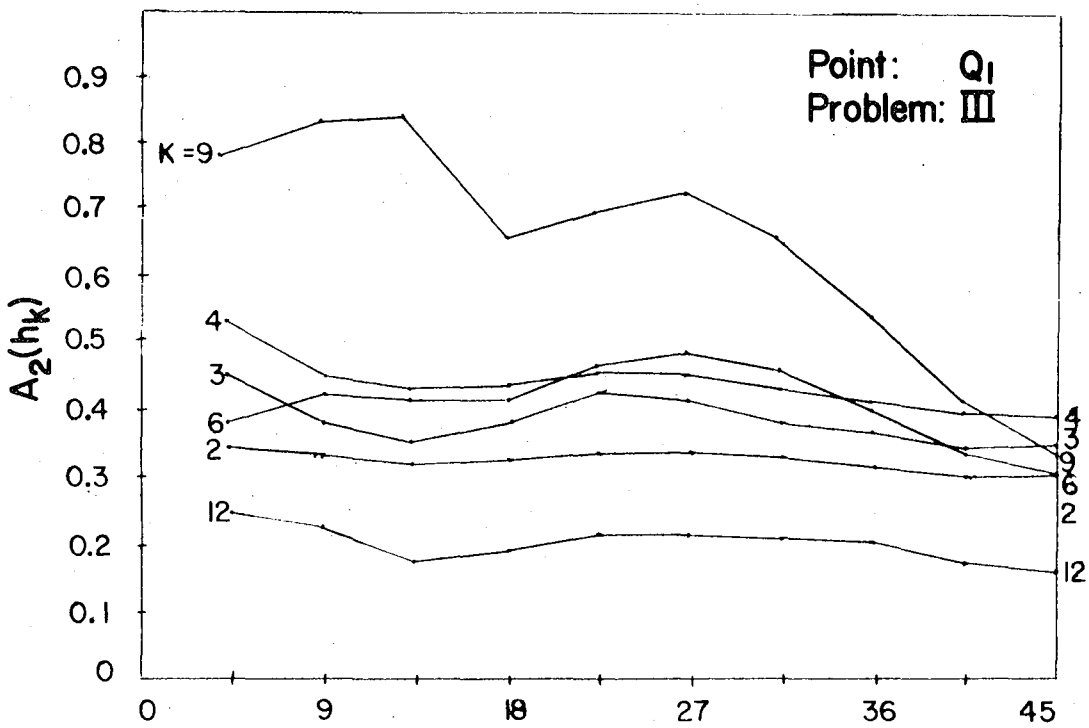
(3b)



XBL 7710-10036

Fig. 3

(4a)



(4b)

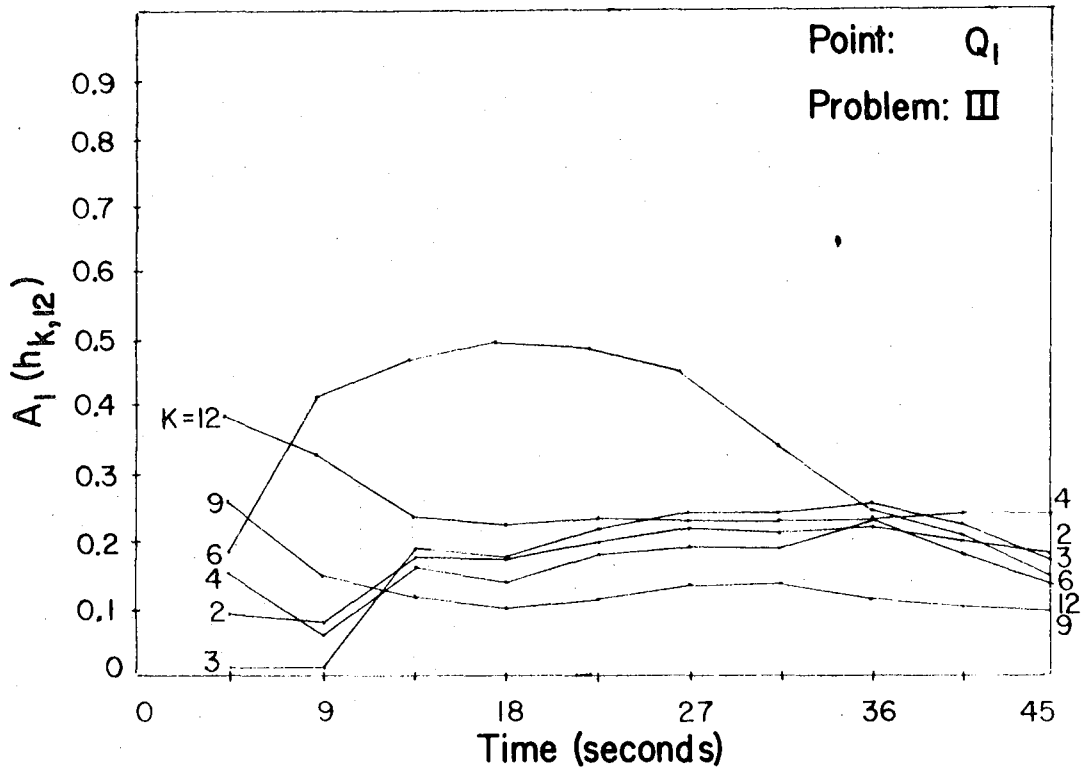
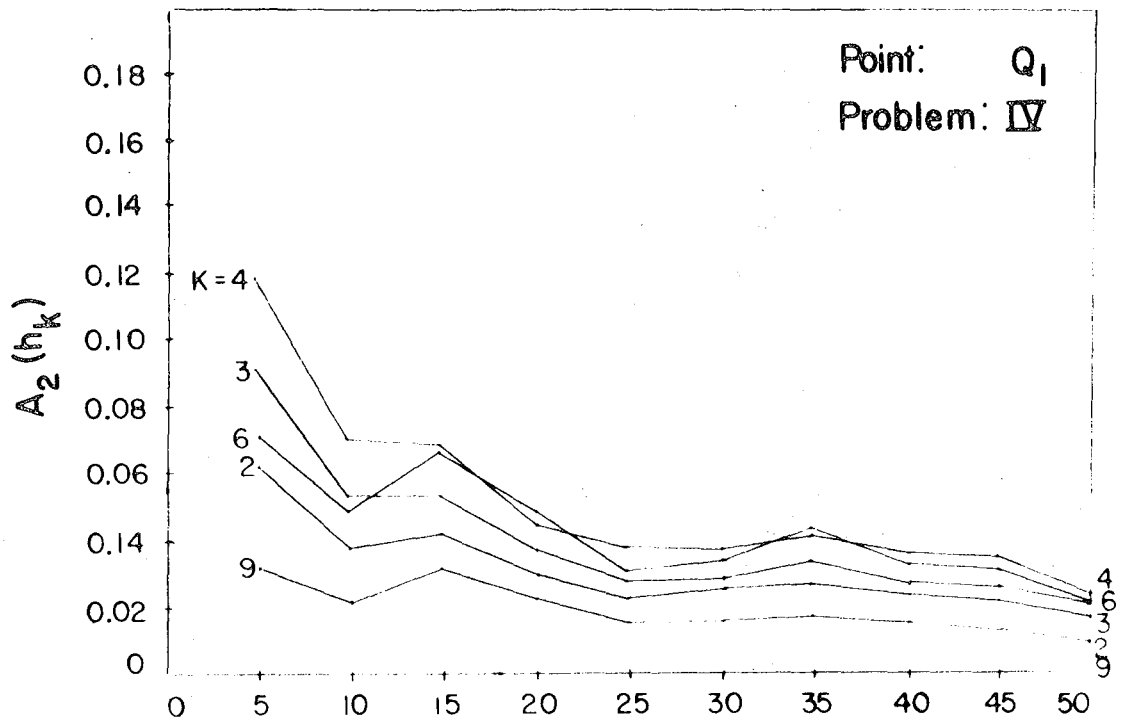


Fig. 4

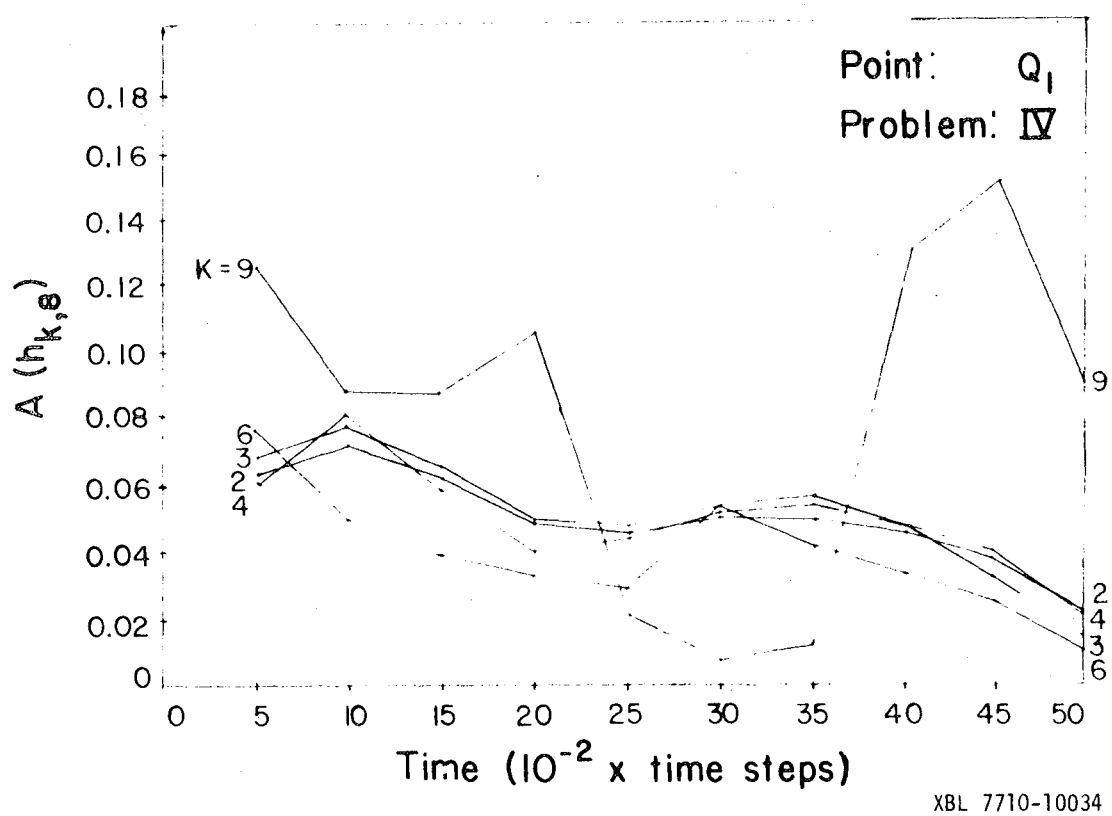
XBL 7710-10035

0 0 0 0 4 8 0 0 8 9 4

(5a)



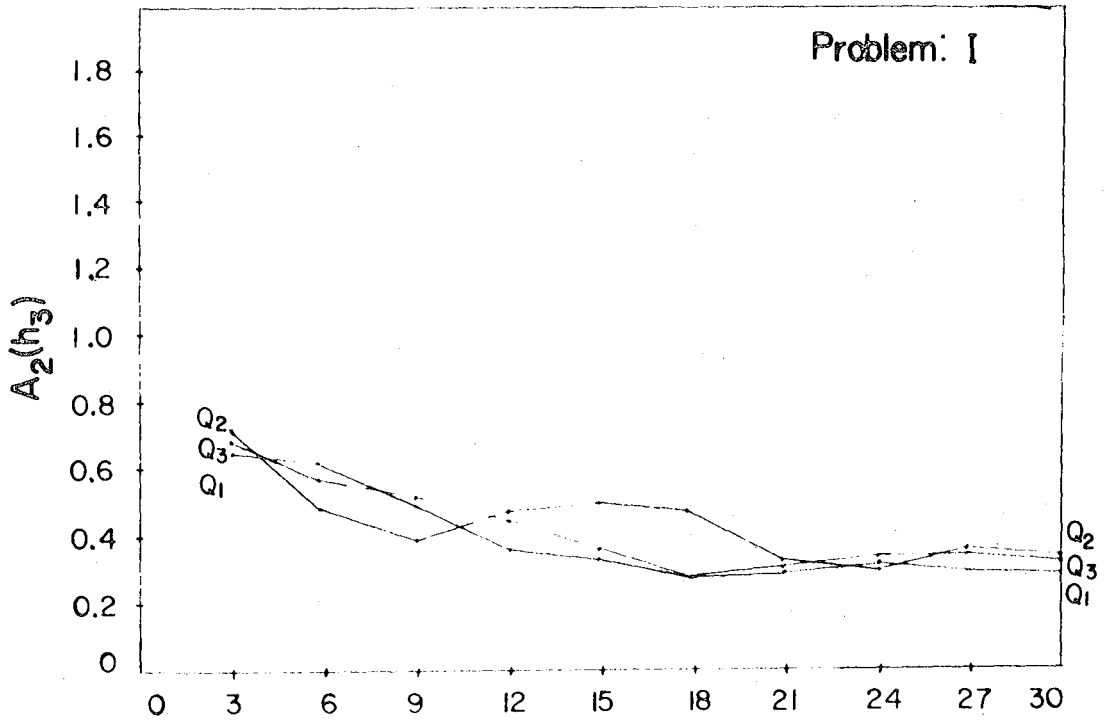
(5b)



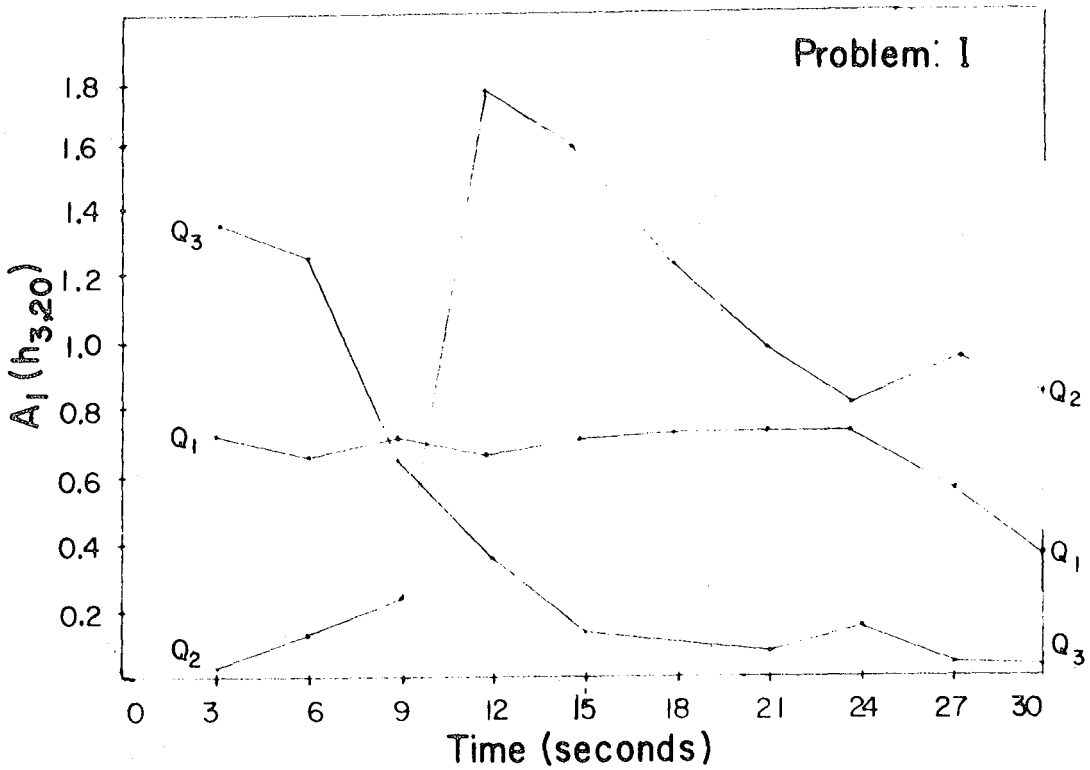
XBL 7710-10034

Fig. 5

(6a)



(6b)

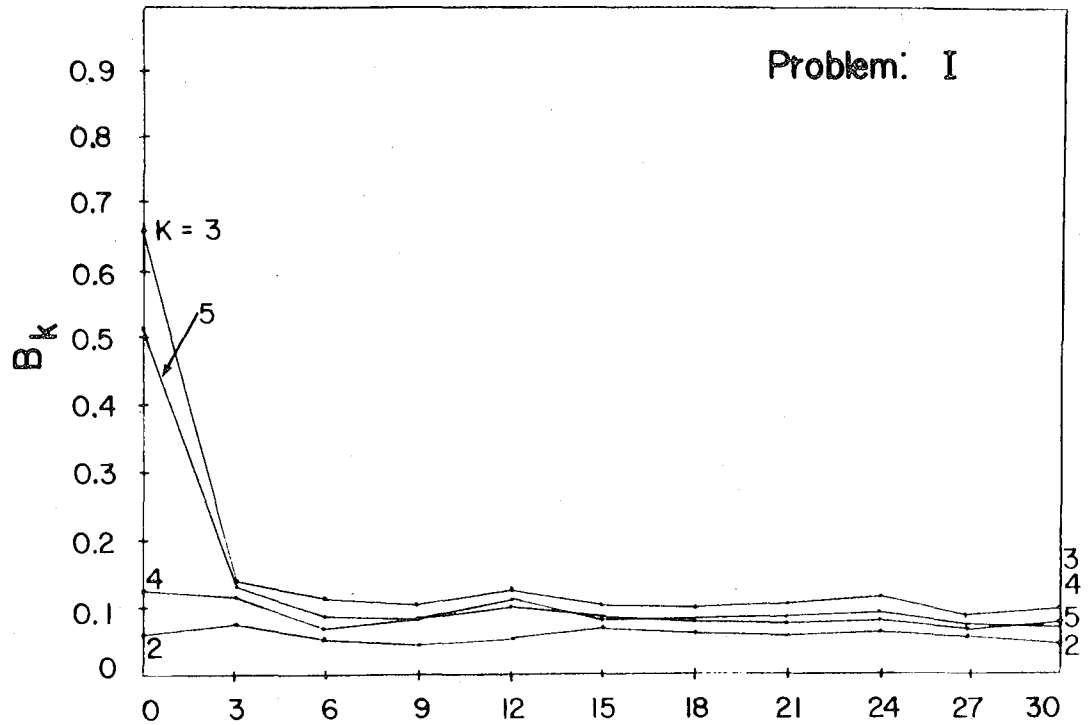


XBL 7710-10033

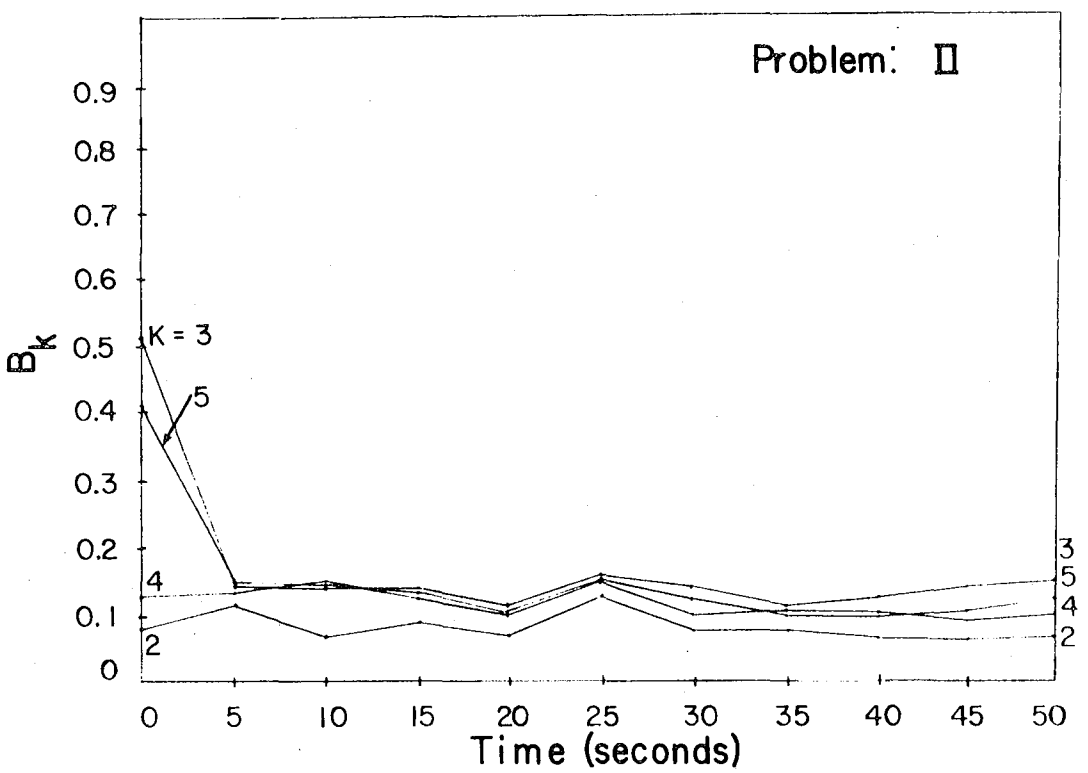
Fig. 6

0 0 0 0 4 8 0 0 8 9 5

(7a)



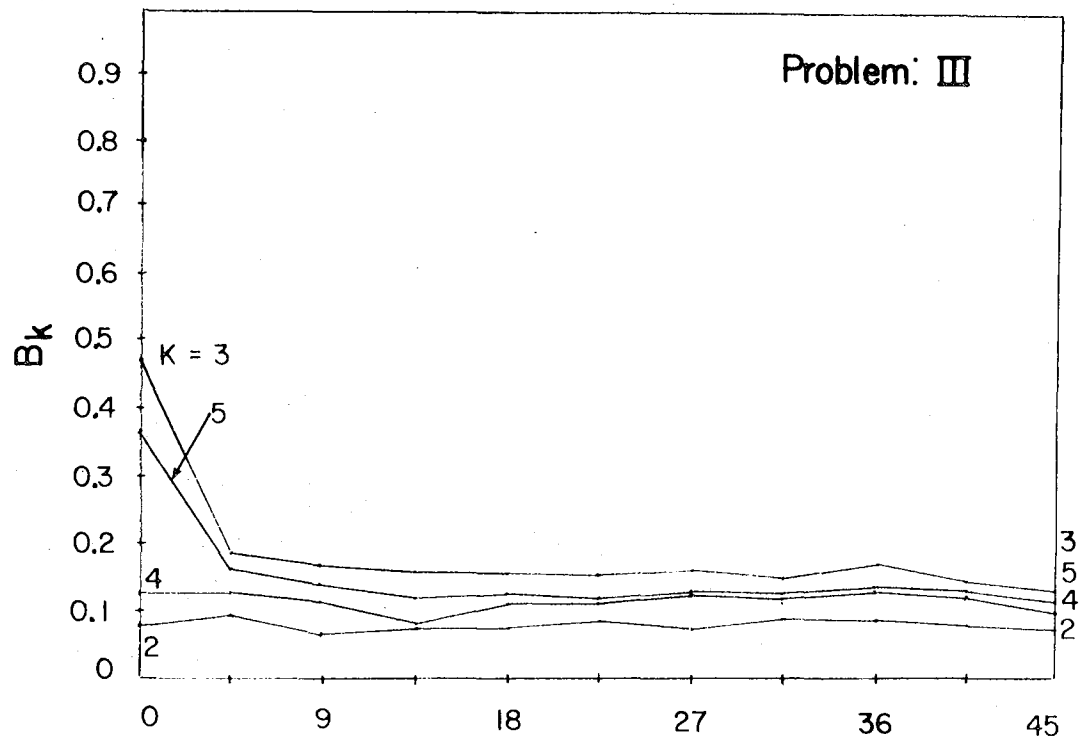
(7b)



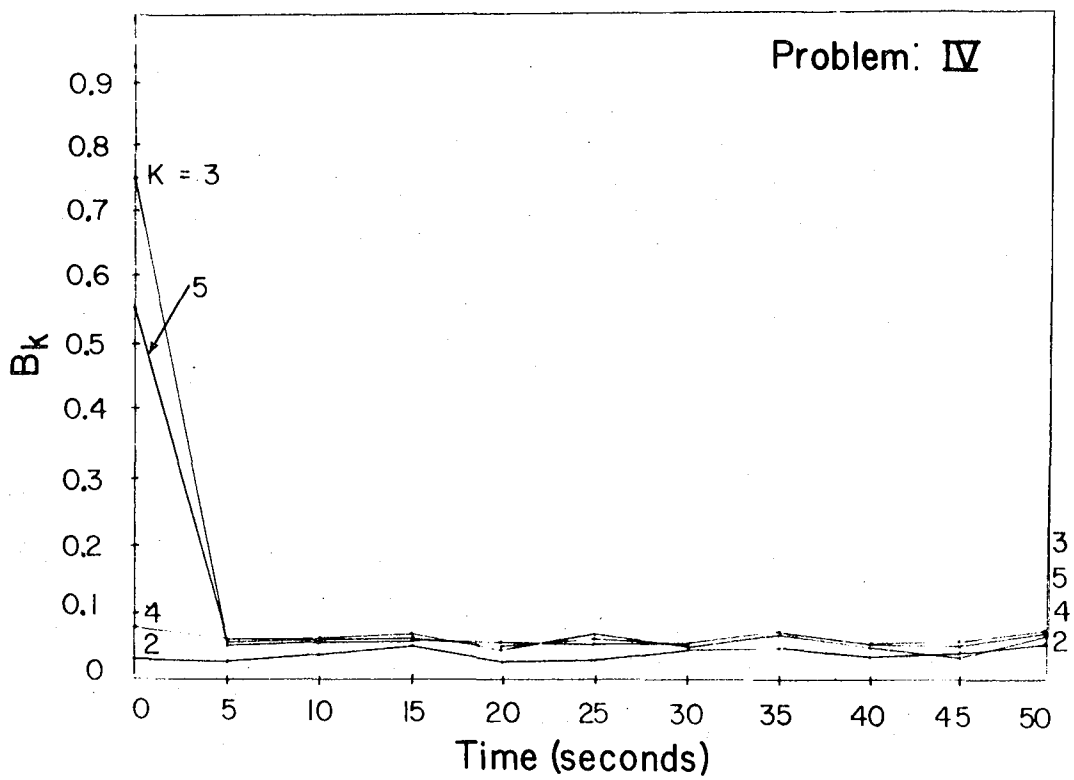
XBL 7710-10032

Fig. 7

(7c)



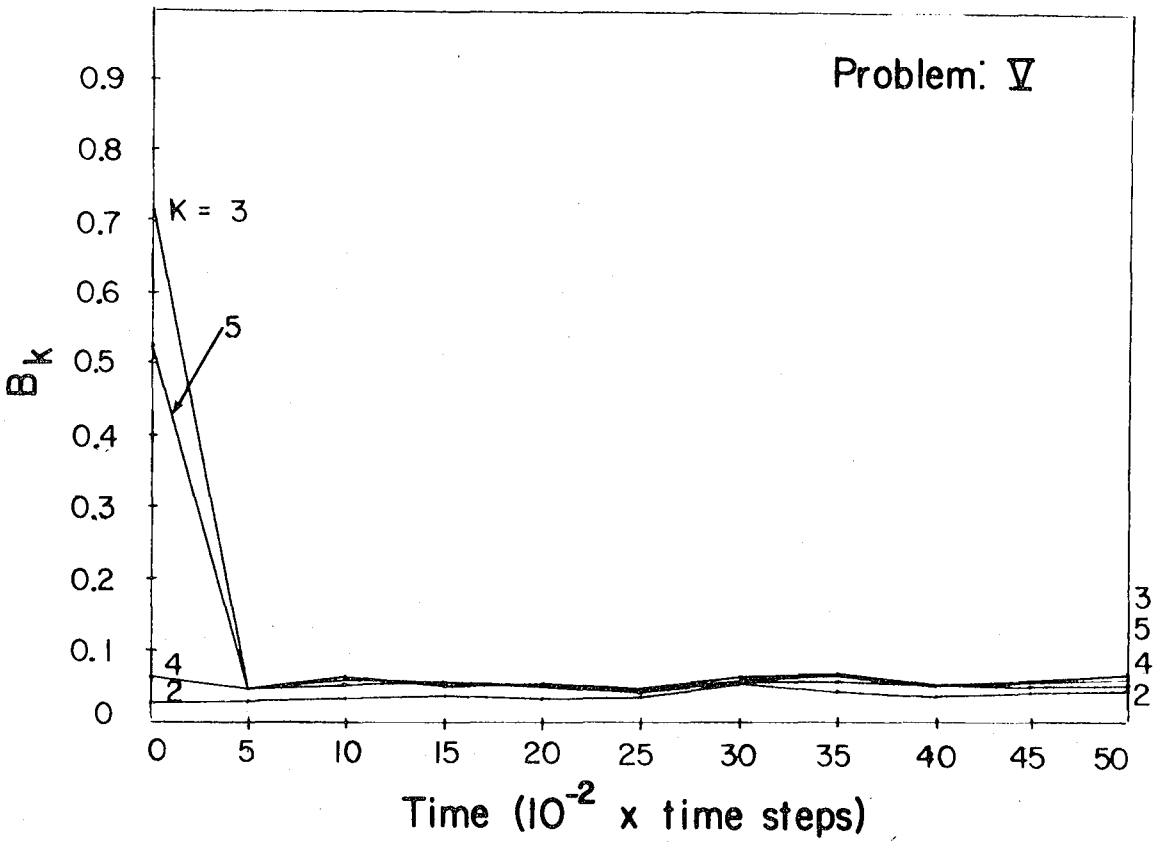
(7d)



XBL 7710-10031

Fig. 7 (cont.)

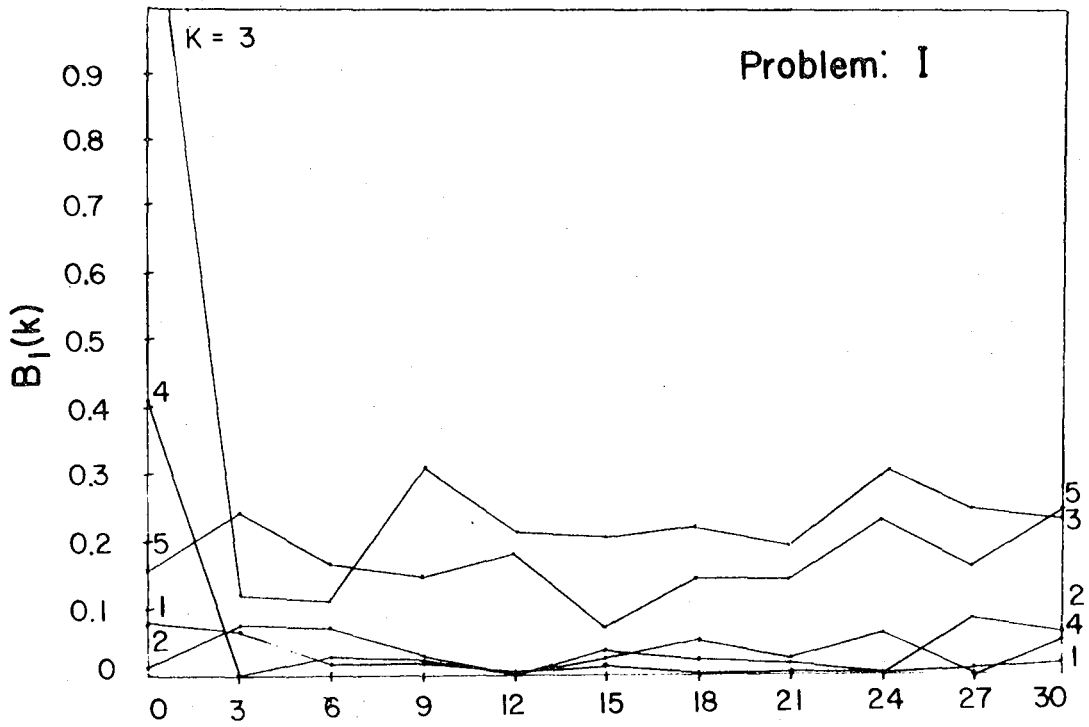
(7e)



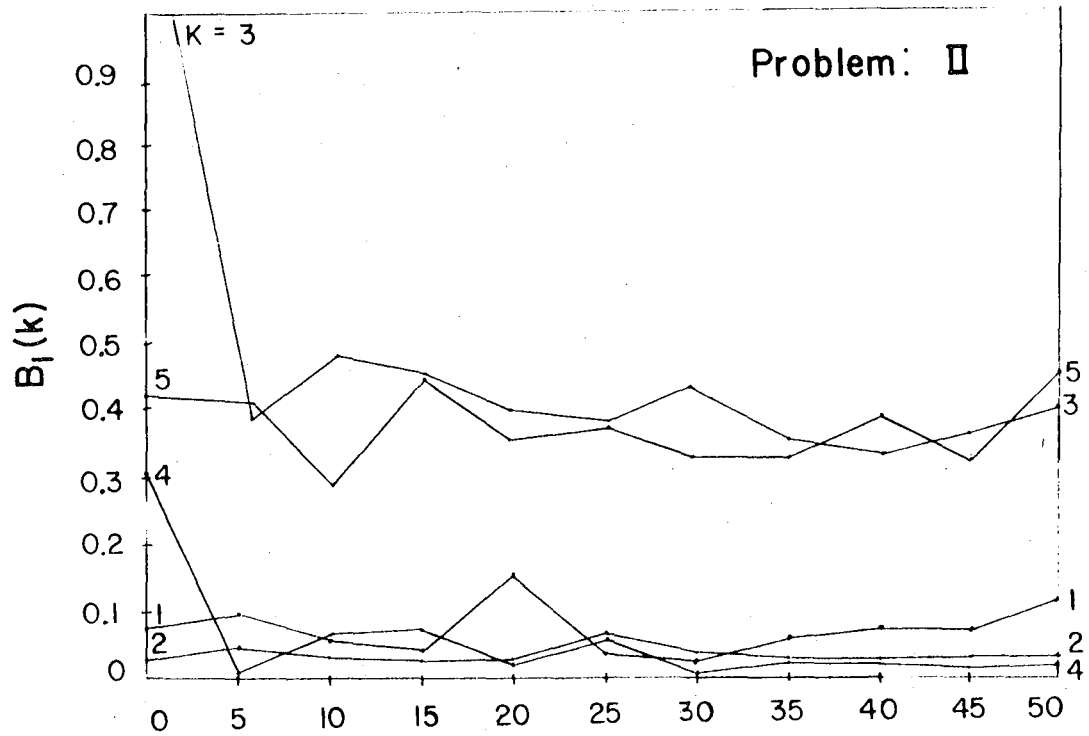
XBL 7710-10030

Fig. 7 (cont.)

(8a)



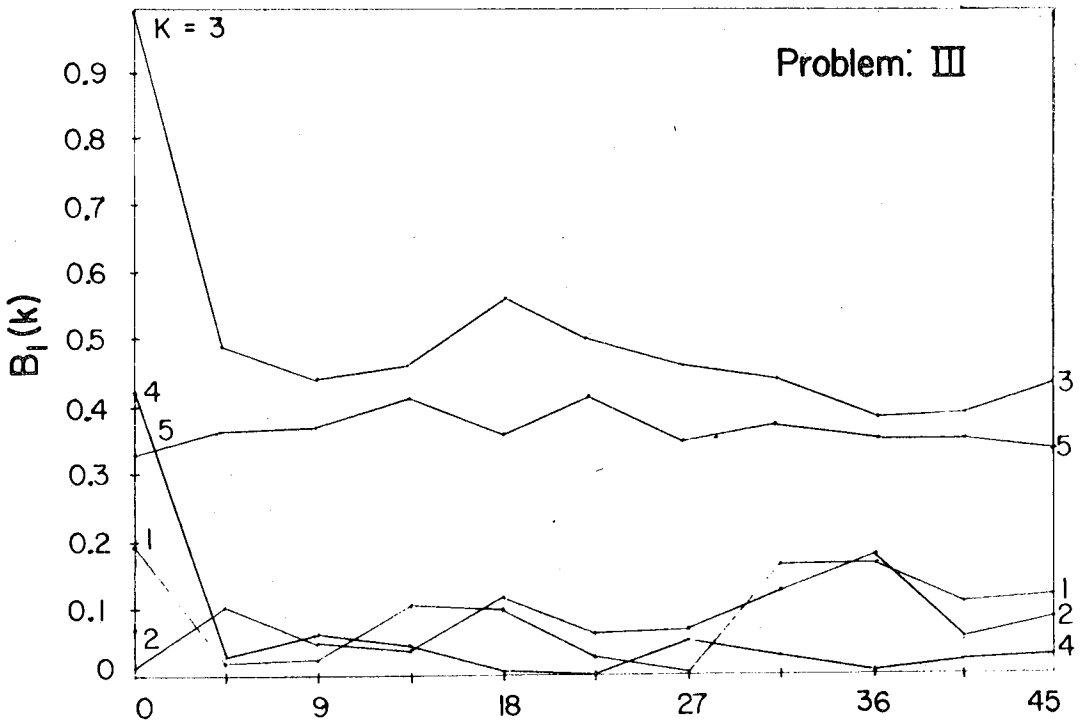
(8b)



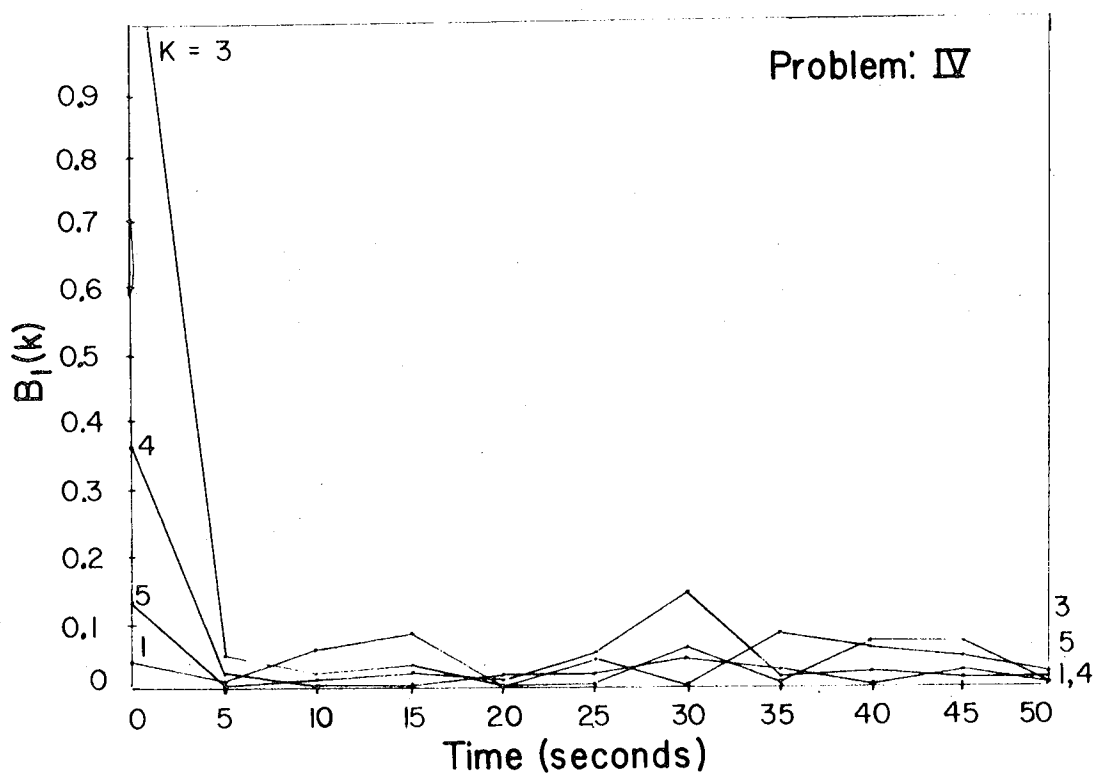
XBL 7710-10029

Fig. 8

(8c)



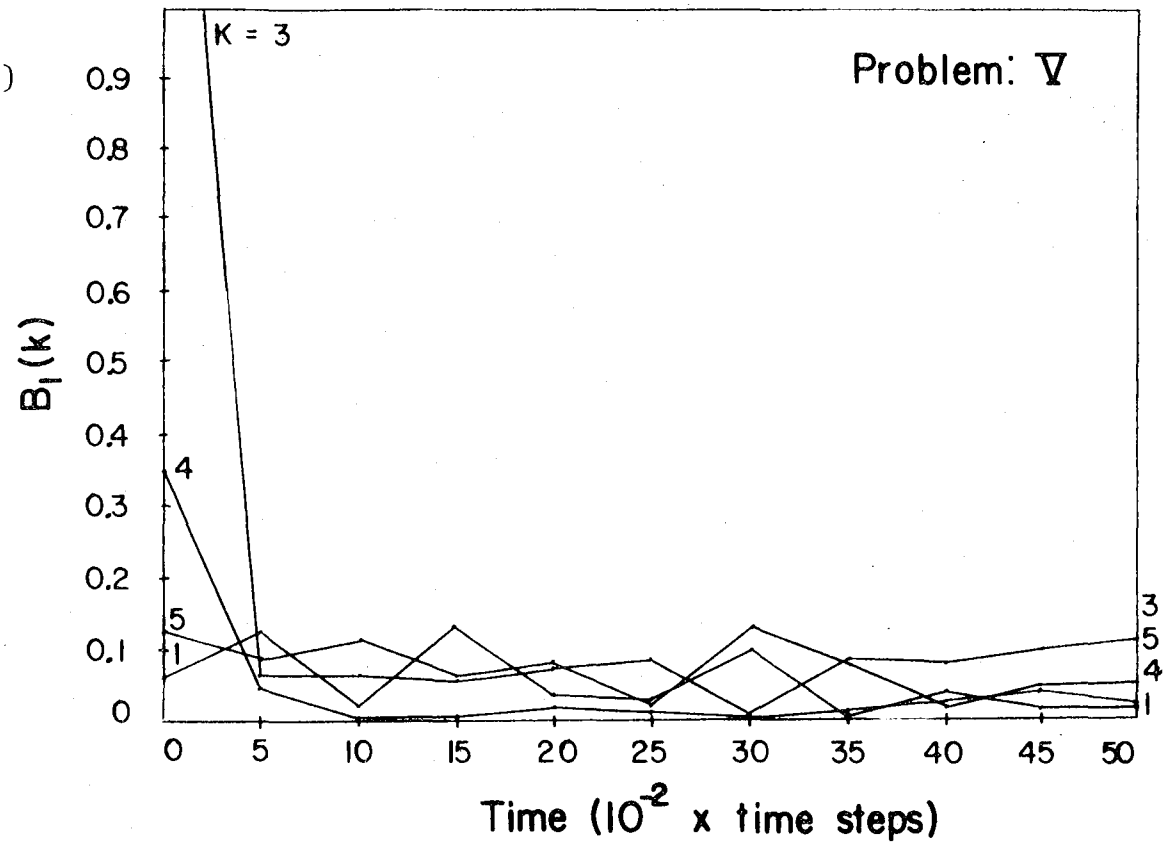
(8d)



XBL 7710-10028

Fig. 3 (cont.)

(8e)



XBL 7710-10027

Fig. 8 (cont.)

This report was done with support from the United States Energy Research and Development Administration. Any conclusions or opinions expressed in this report represent solely those of the author(s) and not necessarily those of The Regents of the University of California, the Lawrence Berkeley Laboratory or the United States Energy Research and Development Administration.

0 0 0 0 4 5 0 2 2 9 4