COO-2094-2

ICR QUARTERLY REPORT

No. 26

August 1, 1970

RECEIVED BY DTIC JAN 19 1971

MASTER

THE UNIVERSITY OF CHICAGO

THE INSTITUTE FOR COMPUTER RESEARCH

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of 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. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

The Institute for Computer Research

The University of Chicago

ICR Quarterly Report

Number 26

August 1, 1970

LEGAL NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Atomic Energy Commission, 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.

Prepared under the auspices of AEC Contract No. AT(11-1)-2094.

Robert L. Ashenhurst, Principal Investigator

Staff of the Institute for Computer Research

Robert Ashenhurst

Clive Church

Lionel Coppleson

William Georgas

John Ginther

Joseph Goguen

Robert Graves

Peter Greene

John Grillos

James Hansley

Richard Herwitt

Bruce Kallick

George Kreisheimer

Glenn Manacher

Richard Miller

John Munn

Daniel Norman

Alex Orden

John Shepherd

Cheng-kiat Teng

Serena Torres

Nancy Trombetta

Cherie Weil

Roman Weil

Victor Yngve

TABLE OF CONTENTS

Section I

- A. Project Support Activity

Section II

- A. Optimal Scheduling Rules for Information Processing Systems
(L. Schrage)
- B. Information Structure Modeling of Organizations (A. Orden)
- C. The Existence and Calculation of Mean Sentence Length for a
Context-Free Stochastic Grammar (S. Soule)
- D. Numerical Experiments on Pair Correlations and on "Thermodynamics"
(R. Miller)
- E. The Fuzzy Tychonoff Theorem (J. Goguen)
- F. Numerical Experiments in Collisionless Systems (R. Miller)

PROJECT SUPPORT ACTIVITY

1. A new control unit and line buffer for the Shepard line printer has been designed and is currently under construction. This has been designed to make use of components which are surplus from the Magnum project, in this way the cost of new hardware has been minimized but a lot of circuit board construction work is involved.
2. One (Z61A) display console of the WS 239A Display System has been adjusted and has been tested using a test unit (the Display Pattern Generator (DPC)) provided with the system. Special methods of testing and adjustment had to be developed, as many sections of the maintenance manual were incorrect, and some of the test procedures given actually damaged components. The analog circuitry in the Z61A display console now seems to be reliable and stable. It has been found, after much study of circuit diagrams, that, by utilizing some features originally provided only for test purposes, information can be stored directly into the Z61A refresh memory, which will make control by Maniac III simpler and faster than anticipated. Design and construction of the necessary control circuits is now in progress.
3. Construction of the Cipher Tape controls has been held up due to priority given to the Shepard printer control.
4. It is understood that the results obtained by Professor Wharton's group of the Chemistry Department with their Molecular Accelerator using the ICR PDP-8 as a portable unit for data-gathering will be published shortly.

Their satisfactory experience with the PDP-8 has led them to order their own PDP-11 for future experiments. The ICR staff is assisting in the design of peripheral interfaces for this machine, and has also written an assembler to permit PDP-11 programs to be assembled from punched cards on Maniac III.

5. The Texas Instruments Digital-to-Analog converter has been linked to the PDP-8 I/O Bus. The Music and Psychology Departments have applications involving this device.
6. The PDP-8 has been linked to the EAI analog computer to permit work with a hybrid system; programs are being written and this will be the subject of a report in due course.
7. An interface between a PDP-8I and a Repco 120 character printer has been designed for Professor Eaton of the Chemistry Department.

OPTIMAL SCHEDULING RULES FOR INFORMATION PROCESSING SYSTEMS

L. Schrage

We try to bring together here for the first time what is known about the optimal scheduling of job streams on single processor systems. Associated with each arriving job is a waiting cost function which specifies the loss incurred by the system for any given length of time that the job spends in the system. It is reasonable to assume that this function is a monotonic nondecreasing function. A decreasing loss function does not make sense if results can be stored at zero cost. A scheduling rule A is said to be optimal if the expected waiting cost per unit time under rule A is lower than that under any other rule.

It is useful to define the following classifications of scheduling rules, waiting cost functions, and processing and interarrival time distributions.

Scheduling Rules

FIFO: First-In-First-Out. The job longest in the system has highest priority.

LIFO: Last-In-First-Out. The most recently arrived job has highest priority of those waiting for processing. Job being processed has highest priority of jobs in system.

Preemptive LIFO: The most recently arrived job has highest priority of jobs in system. Thus, a new arrival always preempts the job being processed.

SPT/C: It is assumed that the waiting cost function for each job i is a linear function with slope c_i . The job with the smallest ratio of processing time over c_i has the highest priority of waiting jobs. The job being processed has highest priority.

ESPT/C: Same as SPT/C except the processing time of each job is not known exactly, only the expected value of the processing time is known.

SRPT: The job with the Shortest Remaining Processing Time has the highest priority of those jobs in system. Thus, if a job arrives which has a shorter processing time than the remaining processing time of the job being processed, the arriving job preempts the job being processed.

ESPR: Same as SRPT except that remaining processing times are not known exactly, only the expected value of the processing time is known.

RR: Round Robin rule. Each job at its turn receives a small amount of processing θ . If the job does not finish in θ it returns to the end of the queue. Fig. 1 indicates the flow of jobs.

For simplicity the limiting case, $\theta \rightarrow 0$, is considered here. In this case, the processor is divided equally among all jobs in system.

FB: Feed Back rule. Each job at its turn receives a small amount of processing θ . If the job does not finish in θ it goes to the end of the next lower priority queue. The processor always chooses its next job from the highest priority nonempty queue. Fig. 2 indicates the flow of jobs. For simplicity, the limiting case $\theta \rightarrow 0$, is considered here. In this case, the processor is always divided equally among the jobs which have received the least amount of processing.

Waiting Cost Functions

Linear: The marginal waiting cost is a constant. The waiting cost is proportional to the time in the system.

Convex: The marginal waiting cost is a nondecreasing function, that is, the second derivative of the waiting cost function is nonnegative.

Concave: The marginal waiting cost is a nonincreasing function, that is, the second derivative of the waiting cost function is nonpositive.

Distributions

GD: General, Dependent. The distribution (of interarrival times or processing times) is of arbitrary form. Times may (not necessarily) be dependent.

G: General. The distribution is of arbitrary form. Times are independent. G. is a special case of GD.

IFR: Increasing Failure Rate. The probability that a job will [complete] (arrive) in the next instant is a nondecreasing function of the [processing already received by the job] (time since last arrival). IFR is a special case of G. The Erlang distribution is an example of an IFR distribution.

DFR: Decreasing Failure Rate. Same as IFR except replace "nondecreasing" by "nonincreasing." DFR is a special case of G. The Hyperexponential distribution is an example of a DFR distribution.

M: Markovian or constant failure rate. Same as IFR except replace "nondecreasing" by "constant." Thus, the probability that [a job being processed will complete] (an arrival will occur) in the next instant is independent of [processing already received] (time since last arrival). M is a special case of IFR and DFR. The Exponential distribution is the only one which is M.

Notation for Classifying Processor Systems

A common convention for denoting a system with type x inter-arrival time distribution, type y processing distribution, and n processors is $x/y/n$. Thus, a system with one processor, exponential interarrival times, and general independent processing times would be denoted as $M/G/1$.

Tabulation of Optimal Scheduling Rules

The amount of information that is available in making scheduling decisions varies from one situation to another. It seems useful to examine three situations of available information:

- (1) Complete information. The exact processing time of each job is known exactly when it arrives.
- (2) Probabilistic information. One can distinguish between jobs according to their expected remaining processing time. The expected processing time remaining is known for every job in the system.

- (3) No information. No information is available with regard to processing times which would cause us to distinguish between jobs.

Another classification of systems is according to whether preemption is allowed.

One can now list the optimal rules according various situations as in Table 1.

Table 1
Optimal Scheduling Rules

Conditions	No Preemption Allowed	Preemption Allowed
Complete information	SPT/C is optimal for linear costs, M/G/1.	SRPT is optimal for identical linear costs, GD/GD/1.
Probabilistic information	ESPT/C is optimal for linear costs, M/G/1.	ESRPT is <u>good</u> but <u>not necessarily optimal</u> for identical linear costs, GD/GD/1.
No information	FIFO is optimal for identical convex costs, G/G/1.	<ol style="list-style-type: none"> 1. FB is optimal for identical linear costs, GD/DFR/1. 2. FIFO is optimal for identical convex costs, GD/IFR/1. 3. Preemptive LIFO is optimal for identical concave costs, GD/M/1. 4. RR is <u>good</u> but <u>not necessarily optimal</u> for identical convex costs, GD/DFR/1.

Goodness of Scheduling Rules

It is of interest to examine the goodness of nonoptimal scheduling rules. Table 2 compares FIFO, RR, FB, SPT/C, and SRPT rules for the case M/M/1, identical loss functions. In this case, the rule which minimizes the mean response time (time-in-system) minimizes the expected loss per unit time. In the tables and figures that follow, $\bar{R}$ denotes mean response time and ρ denotes the load on the system or equivalently, (a) per cent time busy or (b) ratio of mean processing time to mean interarrival time.

One can observe that at high loads, the optimal nonpreemptive, full information rule, SPT/C, can be five times better than the FIFO, RR, and FB rules. SPT/C is almost as good as the optimal preemptive rule, SRPT. This means that in the case of full information and identical linear loss functions, preemption does not offer much improvement.

Discriminatory Features of the Rules

In Figures 3 and 4 we examine the manner in which the various rules discriminate among jobs according to the actual processing time of the jobs. In many cases shorter jobs accumulate cost faster than long jobs while waiting. Thus, one would desire a rule which discriminates in favor of short jobs. Figures 3 and 4 indicate that SRPT discriminates the most in this respect of the five rules examined. Of the no-information rules, FB appears to discriminate the most in favor of short jobs;

, however, it also discriminates very strongly against the infrequent long jobs. FIFO is the least discriminatory. RR discriminates in proportion to the processing time (this is true more generally for the M/G/1 case). It is not obvious from the figures, but SRPT discriminates considerably more than SPT/C for very short jobs.

Table 2

$\bar{R}$ vs. ρ for Various Disciplines,
Processing Times Exponential with Mean 1,
Mean = Standard Deviation
M/M/1, Identical Linear Loss Functions

ρ	FIFO RR FB	SPT/C	SRPT
0.3	1.43	1.36	1.20
0.5	2.00	1.71	1.42
0.7	3.33	2.31	1.87
0.8	5.00	2.88	2.35
0.9	10.0	4.20	3.55
0.95	20.0	6.26	5.54
0.98	50.0	11.28	10.5
0.99	100.	18.50	17.6

Sensitivity of the Scheduling Rules

It is of interest to know how the performance of a scheduling rule is affected by changes in the processing time distribution. A good single measure of the form of a distribution for processing system purposes is the coefficient of variation (ratio of variance to mean squared). IFR distributions have a coefficient of variation less than or equal to 1, and DFR distributions have a coefficient of variation greater than or equal to 1. Figures 5 and 6 examine how the mean response time, $\bar{R}$, under various scheduling disciplines is affected by changes in the coefficient of variation of the processing time distribution. As the coefficient of variation increases, it appears that full information becomes more valuable (SPT/C vs. FIFO) and preemption becomes more valuable (SRPT vs. SPT/C). It appears that FB is very distribution sensitive. It was pointed out in Table 1 that FB is the optimum no-information rule for identical linear loss functions and M/DFR/1; however, in Figures 5 and 6, FB does very poorly as the coefficient of variation increases. The processing time distribution used in Figures 5 and 6 had mass points at .1, 1.0, and 10 with the probabilities of these three processing times adjusted so as to achieve the desired coefficient of variation. For a true DFR distribution, the FB rule should fall between SRPT (the optimal full information rule) and RR (a non-optimal, no-information rule) when the coefficient of variation is greater than 1.

It is worth noting that for every partial information scheduling rule there is a full information rule which dominates the partial information rule in the sense that for any given job stream, every job would finish at least as soon under the full information rule as it would under the partial information rule, and the full information rule uses fewer preemptions. A dominating rule is one which always processes that job that would finish earliest, if the job stream was processed by the partial information rule, of those jobs in system. Thus, for example, the FB scheduling rule is dominated by a preemptive SPT rule, that is, a rule which always processes the job with the shortest processing time of those jobs in system.

Proofs of Optimality

We present here the outlines of the proofs of the various statements in Table 1.

EPT/C and ESPT/C for Linear Costs, $M/G/1$, No Preemption

Aczel [1] and others have shown that for the case $M/G/1$, no preemption, linear costs and processor not idle when there are jobs waiting, the ESPT/C rule is optimal. Thus, it only remains to show that it is nonoptimal to have the processor idle when there are jobs waiting.

Proof of the nonoptimality of enforced idleness relies on the memoryless property of the exponential interarrival time distribution.

An implication of this property is that knowledge of the elapsed time since the last arrival gives no information about the time of arrival of the next job. For example, if the system is IFR/G/1 and a single long job is ready to start processing, it may be worthwhile to hold the processor idle if the time since last arrival is long, in the expectation that a shorter or more costly job will be arriving shortly. There are three situations of enforced idleness to consider:

- 1) The enforced idle period ends before the next arrival,
- 2) The enforced idle period ends at or after the next arrival occurs and
 - a) the new arrival is not the job ending the idle period,
 - b) the new arrival is the first job to be processed after the idle period.

In situation (1) the completion time of the job initiating the busy period can be decreased and all other completion times not increased by starting the initiating job earlier. In situation (2a) the same argument applies. In situation (2b), the memoryless property implies that any job that was previously held back when the processor was available should continue to be held back, and thus never be processed. This certainly cannot be optimal.

It would seem that ESPT/C might also be shown optimal for IFR/G/1.

Optimality of SRPT and ESRPT

Schrage [5] shows the SRPT is optimal for identical costs, GD/GD/1.

Sevcik [6] gives a counterexample to the optimality of ESRT and gives an optimal rule for the case where there are only a finite number of jobs and they are all available at once.

No Information, G/G/1, Identical Convex Costs

"No information" implies that the only way of distinguishing between jobs is by their length of time in system, thus it is never optimal to enforce idleness and wait for another job. Convexity implies that while any two jobs are in system, the older job will always incur cost at a faster rate, therefore we want to give priority to the older job, i. e. the FIFO rule.

No Information, Preemption, Identical Linear Costs, GD/DFR/1

For DFR processing times, the job which has received the least processing time is most likely to complete in its next interval of processing. Since there are no other distinctions between jobs, we also want to process the job most likely to complete in the next instant. The FB rule does just this.

No Information, Preemption, Identical Convex Costs, GD/IFR/1

Cost is always accumulating more rapidly for older jobs, so when selecting the next job to process, the oldest job should be selected. Once processing is started on a job, it always has a higher probability of completing in the next instant of processing than any other job eligible for processing; thus there should be no preemption. This corresponds to the FIFO rule.

No Information, Preemption, Identical Concave Costs, GD/M/1

The latest arrival is always accumulating costs more rapidly than any other job in system. The probability of job completion is independent of which job is being processed, thus we always want to give highest priority to the latest arrival. This is the Preemptive LIFO rule.

No Information, Preemption, Identical Convex Costs, GD/DFR/1

This case would seem to be typical of many computer system job streams. The DFR processing times suggest giving priority to old jobs. The FIFO rule does the latter, the RR rule seems to strike a compromise between the two desires.

REFERENCES

1. Aczel, M.A. "The Effect of Introducing Priorities," Operations Research, VIII,5 (September, 1960).
2. Conway, R.W.; Maxwell, W.L.; and Miller, L.W. Theory of Scheduling, Reading, Mass.: Addison-Wesley Pub. Co., 1967.
3. Greenberger, M. "The Priority Problem and Computer Time-Sharing," Management Science, XII,11 (July, 1966), 888-906.
4. Phipps, T.E. "Machine Repair as a Priority Waiting Line Problem," Operations Research, IV,1 (February, 1956).
5. Schrage, L. "A Proof of the Optimality of the Shortest Remaining Processing Time Discipline," Operations Research, XVI,3 (May-June, 1968), 687-690.
6. Sevcik, K. "Minimizing Expected Cost with Preemption Allowed and Non-Deterministic Service Times," Committee on Information Sciences Report, University of Chicago, July, 1970.

Fig.1

Fig.2

Fig.3

Fig.4

Fig.5

Fig.6

INFORMATION STRUCTURE MODELING OF ORGANIZATIONS

A. Orden

Preface

The phrase "electronic data processing," which has been used for many years to indicate that digital computers provide a wide variety of services to managers, is currently being virtually abandoned in favor of "management information systems." The more recent term--ushered in by the large memory, communication capabilities, and high speed of third generation computers--invokes two perspectives:

1) That with the aid of computers many of the activities in an organization can be more closely coordinated than heretofore, and can thereby be regulated as though they were quasi-automatic.

2) Somewhat antithetical to the trend toward automation (or quasi-automation)--that the quality of the management of an organization can be improved by use of computers to provide a comprehensive data base--a storehouse of information which managers can freely manipulate as background for decision making.

The two attitudes are thought, ideally, to be complementary: high volume, relatively routine activities become computer centered and some functions become computer controlled; and at the same time, largely as a byproduct, information storage in the computer provides middle and upper level managers with a data base for planning and decision making.

II B-2

This idealization is however an oversimplification; it should perhaps be called a fallacy. It does not come to grips with the fact that computer centered information systems are becoming so closely interwoven with the fabric of management that in some areas the determination of computer functions cannot be separated from the determination of the managerial structure of the organization. Some managerial tasks can be formalized and incorporated in computer programs and some cannot. Some types of information which are significant for decision making can be encoded and stored in a computer and others cannot. Thus the determination of the role of the computer requires comprehensive investigation of the structure and functions of management.

Consider a manager who supervises a lower echelon of managers or a group of white collar workers. He partitions areas of information, information handling, and control functions "vertically" and "horizontally" --vertically, in distinguishing functions and information for which his subordinates are responsible from those which he reserves for himself; and horizontally in allocating functions among his staff. In many situations he considers the functions which he delegates to his subordinates to be relatively well structured, and partially programmable, and those which he reserves for himself to be ill structured. However, except for entirely routine clerical steps, he does not assume that the work of his staff can be completely programmed. Instead he recognizes that every managerial activity involves some unstructured problem solving, and requires information which is relevant to that activity.

The development of computer centered management information systems involves similar considerations, and is in some respects more demanding. Systems analysts reconsider some (but not all) previously established vertical and horizontal partitions of responsibilities among managers and their subordinates. They focus primarily on restructuring the horizontal partitioning by finding programmable tasks which can be handled in a computer. But very few functions--as seen by the manager--can be fully automated. The ill-structured aspects of each function remain unprogrammed and--after the subordinate managerial activities have been restructured--the information in the computer must be treated as a "data base" for the subordinate staff even more than for the manager. It is fallacious to aim for symbiosis between the computer and the manager rather than the computer and the manager's subordinates--as though the subordinates are quasi-automata who do not deal with ill-structured problems. The subordinate functions are not so much eliminated as rearranged so that the computer does the more routine parts; and the subordinate staff become more fully concerned with relatively ill-structured matters than they had been.

1. INFORMATION SYSTEMS IN ORGANIZATIONS

The word "system" is used with varied connotations and in many contexts. In order to deal with its ambiguities let us use the terms, "system/concept" and "system," as follows: a system/concept is an identified class of process regularities of some specified set of entities. A system is a system/concept on which there is little or no disagreement.

There have, for example, been system/concepts about motion of the sun, earth, and planets due to Copernicus, Newton, and Einstein, but use of the term, "the solar system," reflects general acceptance of the Newtonian concept.

We imply above that system/concepts and systems are products of the human mind. Beyond this, what it means to say that certain processes and entities are inherently regular is a philosophic matter which we need not pursue. The definitions simply stress two points: (1) the focus of a system/concept is more on processes (or sometimes on relationships) than on objects. For example, "solar system" seems to refer to the sun and planets, but the underlying reason for using the word "system" is the regularity of movement of the specified set of masses with respect to each other. (2) A class of process regularities is identified--a system/concept rarely covers all of the processes in which the specified set of entities is involved.

The process regularities in a system/concept may be natural, as in the case of the solar system, or man-imposed: for example, water flow in a hydroelectric power system. They may be strictly among the selected entities, or among those entities and their environment. The specified entities may be natural, engineered, or abstract. Any or all of these situations may be brought into consideration in system/concepts about organizations. Whatever the approach may be, the first essential is that regularity of a defined process or set of processes be brought under consideration.

In the past, operational regularities of organizations have generally been put into two broad categories: (1) techno-economic processes and associated analysis--associated mainly with areas which involve extensive mechanization--such as manufacturing and transportation; and (2) the self-imposed standardization of procedures which generally characterizes large organizations, i.e., bureaucracy. But no matter how many regularities are found, the underlying human interactions in the formation and maintenance of organizations (leadership, cooperativeness, etc.) have remained ill understood, not well identified, and relatively unpredictable. Consequently, broad system/concepts about organizations have not had wide acceptance, and organizations per se are not generally called, or thought of, as systems.

The regularities of information processing, and to some extent decision making, which are being identified in the course of computer application development, form a relatively new aspect of the study of organizations. It appears that broad information system/concepts about organizations may now be put alongside bureaucratic system/concepts or techno-economic system/concepts as an indication that major aspects of organizational activity are being brought under consideration.

On a rather narrow basis there have been significant ways in which the term "system" rather than "system/concept" is quite appropriate. In the first place, well before the advent of computers, the word "system" was widely used within organizations to indicate acceptance of regularity

in administrative and data handling procedures, such as accounting systems or personnel record systems, which cut across many or all departments. Secondly, in many organizations computer departments act as quasi-independent service centers which take responsibility for whatever information processing tasks other departments choose to sponsor, which the computer can handle. The set of jobs determines a computer hardware/software system which can be called the computer-based information system of the organization. There is little ambiguity about the jobs and equipment which are involved.

But proceeding to a broader view, the introduction of computers brought prospects not only for extension and unification of routine data handling, but also for regularization and unification in decision making, information retrieval, and complex analysis (e. g. , simulation). The scope of these possibilities has been symbolized, since the early fifties, by a series of catchwords. "Integrated data processing" (first generation) was followed by "total systems" (second generation). "Data bank" and "management information system" are currently popular. "Corporate simulation model" is gaining attention. Under these headings large classes of information processes have been identified and partly incorporated in computer operations. However the computer specialists, the administrative system analysts, and the operating managers of the organization have not formed a mutually agreed view of the scope of this type of development. In the terms expressed at the beginning of this section, developments to

date can more appropriately be described as information "system/concepts" than as information "systems."

The goal of such efforts is the development of information systems which cover: (1) relatively routine administration, (2) formalized decision making -- in particular the implementation of operations research models, and (3) all of the managerial uses of computers, including on-demand production of special reports, simulation, etc. It is quite obvious that agreement on the full scope and content of these classes of process can be achieved only by dealing with the organization as a whole, and within that framework by identifying a comprehensive formal information system of the organization. Accordingly, some organizations have made elaborate arrangements for joint investigation of information processes by teams of computer analysts, administrative procedure analysts, and operating managers.

This paper proposes techniques for succinct representation of all findings about regular information processes in an organization, and for the assembly of such findings in a unified structure. It constitutes a methodology -- suitable for operating managers as well as technical specialists, for the representation of information processes in sufficient detail to exhibit the relationship of proposed new developments in information handling to existing situations, and thus to facilitate planning or assessment of proposed innovations.

2. WORK-CENTERS AND ENVIRONMENT SECTORS

The "structure of an organization" is usually taken to mean the attributes which are represented by an organization chart. Two matters are, in the main, involved:

- (1) The identification of managerial and operational units
- (2) Hierarchy relationships

The division of an organization into discrete units is so closely related to its hierarchic structure that a distinction between (1) and (2) is not usually made. Here, however, rather than "lines of authority," the links between positions are to be viewed as paths along which specific types of information flow. In general terms, each path carries plans and directives from a manager to subordinates, and performance reports in the reverse direction. In the "information structure model" of an organization these loose implications are to be replaced by specific representations of recurring information transfer activities. Recurrent types of information flow which cut across the hierarchy are also to be identified. The underlying point of view is that (a) the managerial and operational units of an organization are the nodes of an information flow network, and (b) the organizational hierarchy is broadly indicative of an important set of flow-lines in that network, but does not, of course, show all of the significant links.

All units which appear on organization charts whether managerial or operational, will for simplicity be called "work-centers" -- they are all

receivers, producers, and transmitters of information. The managerial work-centers are usually identified by job titles such as President, Research Director, or Chief Accountant. It is preferable here to use descriptive phrases such as "General Policy Formulation" (which could represent the president and board of directors), "Research Management," "Quality Control," etc. This implies that for the purpose at hand the form which each work-center may take is of no direct concern; it may be: an individual with a staff, an operating department, a board or committee, the part-time work of an individual who also has other responsibilities; perhaps even a computer.

In the terms of the previous section, the work-centers are the specified entities about which a system/concept is to be formed. Process regularities -- the essential features of systems -- are to be sought with respect to (a) information flow, and (b) information transformation.

The first stage in the construction of an information structure model of an organization is to make up a list of the work-centers. Their locations in the organizational hierarchy are significant. This will, as illustrated below, be indicated by means of hierarchic labels. The labels show, for example, that the Assembly Shop, labeled 1/1/2, is subsidiary to Plant Operation Management, labeled 1/1, which is in turn subsidiary to General Production Management, labeled 1.

1. General production Mgmt.

- 1/1 Plant operation mgmt.
 - 1/1/1 Components shop
 - 1/1/2 Assembly shop
 - 1/1/3 Inspection and packing shop
- 1/2 Production scheduling
- 1/3 Production engineering
- 1/4 Purchasing

2. Marketing Mgmt.

---subsidiary work-centers

3. Financial Mgmt.

---subsidiary work-centers

etc.

The work-centers of a homologous group, e.g. the three shops above, can be expected to have a common pattern of information flow with respect to other parts of the organization. On this basis, such groups can, until quite detailed matters arise, be represented by a single entry, e.g.

1/1/h Shop h, $h = 1, 2, \dots$

Homologous multilevel subdivision, e.g. sales offices in sales regions can be similarly represented:

2/h Region h sales Mgmt., $h = 1, 2, \dots$

2/h/j Sales office j in region h, $j = 1, 2, \dots$

The labels have several roles:

- (1) By indicating location in the organizational hierarchy they act as a frame of reference which facilitates consideration of information processes;

- (2) They provide a useful set of symbols. For example, expressions such as $1/5 \rightarrow 3/2$, or $1/1/h \rightarrow 1/2$, will be used later to symbolize information flows among pairs or groups of work centers.
- (3) In order to form an organizational information structure model in which computer files and programs can be incorporated, it is desirable that all elements be carefully identified. The labeling above, and further aspects which will be introduced later, tend to sharpen the identifications.

In order to identify the types of flow of information which take place between an organization and its environment, the sectors of the environment such as suppliers, customers, tax collectors, etc. must be identified. A partitioning of the environment is implicit in an organization chart; it is to be made explicit here by listing environment sectors as well as work-centers. It is convenient to list each environment sector adjacent to the work-center to which it relates most closely, and to label it by adjoining "/X" to the work-center label, e.g. (in Figure 1)

1/4 Purchasing

followed by

1/4/X Suppliers

A generic list of work centers and environment sectors for a medium-size manufacturing firm -- for use in illustrating further stages of information structure model development -- is shown in Figure 1.

3. INFORMATION FLOW

We now turn to the representation of regular information flows of organizations. (It is to be understood that information "flow" refers here to recurrent discrete events which take place either at regular or at random intervals, for example, the "flow" of monthly expense reports from an accounting department to managerial units.) Information system analysts commonly use network type diagrams to depict existing or proposed information flow patterns. In order, however, to cope with the great variety of information flows which must be considered, a matrix format -- such as that described below -- seems preferable.

The device for identification of information flows among work-centers and environment sectors will be called an "Info-transfer matrix." Figure 2 is a preliminary illustration. The rows and columns of the matrix are named and labeled by listing work-centers and environment sectors on both a horizontal and a vertical axis. The source and destination of each information flow are indicated on the axis. Information flows are identified with regard to source, object, and destination. In order to show the source and destination of a flow, a symbol for the subject is entered in the column which corresponds to the source and the row which corresponds to the destination, e. g. "Sq" in column 1/4/X Suppliers, row 1/4 Purchasing. The subject is named on a line below the matrix:

Supply-price and delivery quotations Sq

A source label, subject symbol, and destination label in the form such

as $1/4/X - Sq \rightarrow 1/4$ provides a convenient symbol for each flow.

A single-source, multiple destination flow, is covered by multiple entry of a subject symbol in a column, e.g. Ps, which appears several times in column 1/2, indicates that product output schedules are sent from the Production Scheduling work-center to Plant Mgmt, to Plant Operations Control, and to Purchasing. Similarly, multiple-source, single-destination flows are to be expressed by multiple entry of a subject symbol in a row.

Like work-centers -- which for the purpose of organization structure representation are identified by names which give a broad indication of function -- the representation of information flow by source-subject-destination, without further detail, will be considered adequate here for information structure modeling. Matters such as medium of transmission (whether as human communication, or in machine form), frequency, or message types, will not be considered. On this score, Fairthorne [1] contends that an information flow has six main attributes: (1) subject (which he calls "designation"), (2) source, (3) destination, (4) message-set, (5) coding, (6) channel. He suggests that various three-way combinations of the six attributes (triads) -- e.g. source/subject/destination, which we use here, and others such as source/channel/destination or channel/message-set/coding -- are "basic communications activities" to which analyses of information processes can properly refer -- without necessarily dealing with all six of the attributes.

An empty Info-transfer matrix on which all of the work-centers of an organization have been placed, but no information flows have as yet been entered, is equivalent to an organization chart. (The labels of the work centers show the hierarchic structure of the organization.) Since the Info-transfer matrix provides a vehicle for representation of information flow as well as organization structure, it may be described as an extension, or generalization, of the organization chart.

An organization chart which represents a branch of a large organization is considered to be a sector of the chart for the entire organization. Similarly, an Info-transfer matrix for a branch of an organization should be viewed as a submatrix of an overall matrix for the organization. The submatrix generalizes the branch organization chart in the same way as the full-matrix generalizes the full organization chart. For example, Figure 3A (based on part of Figure 1) is an organization chart for the Production branch of a firm. Figure 3B is a corresponding Info-transfer matrix, with a sampling of information flows.

The Info-transfer matrix, which emphasizes information flow, rather than the organization chart, which emphasizes lines of authority, will be treated as the basic vehicle for the representation of organization structure in this paper.

4. PROGRAMMES IN ORGANIZATIONS

4.1 Activities and Programme Structures

An identified process which is conducted by a work-center will be called an "activity." An activity which can be modeled or is conducted

under specified rules will be called "structured;" otherwise it is unstructured." A prospective or actual occurrence of an activity will be called an "event."

For the purpose of further development of the information structure model, three types of structured activity will be distinguished:

1. Operation -- automated physical processes, and skilled or unskilled human labor which takes place under schedules and rules of procedure.
2. Information transformation -- activities such as statistical summarization, calculation, and decision-making or collection and organization of information under procedural rules.
3. Information transfer -- activities such as recording, encoding, and transmitting information.

A sequenced set of activities without specification of time or time intervals, will be called a "programme structure," or simply a "programme."* A programme with specific times or time intervals, and with other parameters for events, will be called a "schedule." An activity is to a programme as an event is to a schedule.

A programme structure for production planning, operation, and control for the illustrative manufacturing firm is shown in Figure 4. The

*The British spelling, "programme," is used in order to distinguish its usage here from common uses of "program," as in computer program, or -- where it is synonymous with "plan" -- as in new-product development program.

activities, identified in broad terms, are named and numbered on the left. Work-centers are listed across the top. The activity numbers in the work-center columns show the tasks of the work-centers. Arrows between these entries show the normal order of events which a specific schedule for this managerial/operational process would follow.

An activity/work-center combination will be called an "assignment," and a programme structure in the form shown in Figure 4 will be called an "assignment-profile." Like a flow chart for a computer program it identifies steps and depicts sequential structure; in addition it designates "processors," i. e. work-centers.

Whether each of the activities in a programme is structured or unstructured is left for later consideration. Step 5, "Prepare work-in-progress status reports," might well be done according to definite rules even in a small firm. On the other hand, it is unlikely that step 11, "Review costs," would go beyond the free exercise of judgement by the plant manager even in a large firm.

The following convention has been used in the construction of Figure 4 and is to be applied in general to assignment profiles:

Information transfer activities are not explicitly listed -- they are considered to be implicit in the arrows which connect the assignment entries; instead they are identified explicitly by the entries in the Info-transfer matrix of the organization.

The arrow from step 1 to step 4 in Figure 4, for example, implies that information is transferred from the Production-Scheduling work-center to the Purchasing work-center. This flow, and others which are implicit in Figure 4, have been included in Figure 3B, viz. $1/2 - Lt \rightarrow 1/4, \text{Mfg Lead times}$.

In this way assignment profiles are interlocked with the Info-transfer matrix. As noted earlier, the matrix -- whether all of its entires are shown in a single array, or not -- is to be considered one entity for the entire organization. Thus the basis for information structure modeling of organizations, as conceived in this paper, is an Info-transfer matrix with which diverse programme structures are interlocked.

A second convention, which has been used in Figure 4 and elsewhere in this paper, with regard to the relationship between programme structures and the Info-transfer matrix is:

Interactions between work-centers and environment sectors are identified by entries in the Info-transfer matrix. Processes in the environment are implicit in some of the work-center activities, but they are not explicitly stated on assignment profiles.

This convention need not be strictly followed. In some cases, in order to clarify programme structures, explicit identification of environmental processes on assignment profiles may be desirable.

Structured activities will be discussed further below. Before proceeding, it is worth noting that explicit programmes in which the activities are in general unstructured are not uncommon. Large firms, for example, commonly identify the steps in new-product development (such as the project proposal stage, preliminary cost and market estimation, setting research and engineering budgets, etc.) and the order in which these steps take place. The activities are basically unstructured -- they involve research, innovations in production and marketing, entrepreneurial judgement, and the like. Much of the information flow between work-centers, say between Research Management and Marketing Management, is informal, and cannot be identified in any useful way, but some aspects such as research budgets, target dates, and market forecasts become formalized, and can be identified in the Info-transfer matrix.

4.2 Subprogrammes

"Structured activity," as used above, denotes process regularity in various forms:

Physical process automation

Decision-making on the basis of mathematical models

Procedural rules within work-centers

Computer-based information processing

The structuring of activities -- viewed as the decomposition of an activity into identified subactivities -- is inherent also in the very formation and development of organizations. As organizations grow,

some of the assignments to particular work-centers evolve into "subprogrammes" -- coordinated sets of activities which take place in diverse work-centers.*

The decomposition of an activity generally occurs when parts of it can be structured in one of the forms listed above, e.g., programmed for a computer. The introduction of computers has in fact led to the proliferation of processes which are more appropriately viewed as subprogrammes of some fairly broad programme than as independent programmes. Frequently such subprogrammes include steps (activities) which may be described as fairly well identified, but relatively unstructured.

An assignment profile for a programme at any level in a hierarchy of programmes can be interlocked with the Info-transfer matrix -- on this score no distinction need be made between programmes and subprogrammes. However, when an activity in a programme is replaced by a subprogramme, the information flow entries in the matrix which are associated with that activity must be reviewed.

Activity 5 in Figure 4 identifies the purchasing of materials in the illustrative firm. In a small firm that activity would be relatively unstructured. However in a large firm it typically becomes a subprogramme such as that shown in Figure 5. As before, the arrows between assignment

*The organization of concepts into hierarchies is one of the essential attributes of human thought (Simon [2]). We observe here that as organizations grow, programme structures, as well as authority structures, develop hierarchically.

entries indicate the order of events (to be implemented in specific "schedules"), and imply information transfer activities.

Figure 6 shows the Info-transfer matrix entries which are directly implied by Figure 5. It is a step, introduced only for clarity, toward Figure 7 which includes other information flows which pertain to the purchasing of materials:

Environment flows (suppliers and common carriers)

Managerial reports

Input data required by steps in Figure 5 which are outputs of activities outside the purchasing programme.

As illustrated by Figures 5 and 7, it is convenient to pair an auxiliary matrix with an assignment profile. Each auxiliary matrix thus formed is a family of entries to the overall matrix of the organization. With this in mind, if the flows shown in Figure 7 were copied onto Figure 3B, they would contribute to the formation of a comprehensive Info-transfer matrix for the production branch of the illustrative firm.

REFERENCES

1. Fairthorne, R.A., "Morphology of Information Flow," Jnl. of the Assoc. for Computing Machinery, Vol. 14 (Oct. 1967), pp. 710-719.
2. Simon, H.A., "The Architecture of Complexity," Proc. of the American Philosophical Society 1962 (Vol. 106), pp. 467-482.

Figure 1

Work Centers and Environment Sectors of a Manufacturing Firm

0. Major policy determination
1. Production management
 - 1/1 Plant management
 - 1/1/h Shop h, $h = 1, 2, \dots$
 - 1/1/20 Shipping and receiving
 - 1/1/21 Supply storage and distribution
 - 1/1/22 Plant operations control
 - 1/1/23 Plant accounting
 - 1/1/24 Plant & equipment engineering and maintenance
 - 1/2 Production scheduling
 - 1/3 Production engineering
 - 1/4 Purchasing
 - 1/4/X Suppliers
 - 1/5 Quality control
2. Marketing management
 - 2/1 Sales operations mgmt.
 - 2/1/X Major customers
 - 2/1/h Sales office h, $h = 1, 2, \dots$
 - 2/1/h/X Local customers
 - 2/2 Distribution management
 - 2/2/X Common carriers
 - 2/2/h Regional warehouse h, $h = 1, 2, \dots$
 - 2/2/20 Sales-order processing
 - 2/2/21 Truck fleet
 - 2/3 Sales/production liaison
 - 2/4 Market research
 - 2/5 Advertising
 - 2/5/X Advertising agencies and media
3. Financial management
 - 3/X/1 Stockholders
 - 3/X/2 Banks and other financial institutions
 - 3/1 Accounting
 - 3/2 Budgeting
 - 3/3 Credit control
 - 3/4 Tax management
 - 3/4/X Government agencies
 - 3/5 Employee benefits and pensions
4. Personnel management
 - 4/1 Salaried employee administration
 - 4/2 Hourly employee administration
 - 4/3 Labor union relations
 - 4/3/X Unions and in-plant union agents
 - 4/4 Personnel-record maintenance
 - 4/5 Employee services
5. Research and product development
6. Information system management
 - 6/1 Operations research
 - 6/2 Data systems analysis and programming
 - 6/3 Computer operation

Figure 2

An Info-Transfer Matrix

	Plant Mgmt.	Plant Operations Control	Production Scheduling	Purchasing	Suppliers	Sales- Production Liaison
	1/1	1/1/22	1/2	1/4	1/4/X	2/3
1/1 Plant Mgmt.			Ps			Se
1/1/22 Plant Operations Control			Ps			
1/2 Production Scheduling						Se
1/4 Purchasing			Ps		Sq	
1/4/X Suppliers				Om		
2/3 Sales- Production Liaison						

Short term Sales expectationsSeProduct output schedulesPsSupply price and delivery quotationsSqOrders for materialsOm

Figure 3A

Production Branch of a Manufacturing Organization

Production Branch Info-Transfer Matrix

Mfg. output reports	Mo	(output, labor, material usage)			
product Shipment reports	Sr				
Materials issued for production		Mi			
Shop operation schedules		Se			
Mfg. cost reports			Mc		
Materials received reports	Mr	Mr			
Mfg. Lead times				Li	
Product output schedules and material requirements				Ps	
Production capacity estimates				Pc	
Orders on suppliers					Os
Supply price and delivery quotations					Sp
output requirements for Quality control testing					Qt
Product quality test statistics					Pq
short term requirements to meet Sales and inventory demand					St
product Shipping instructions					Si
Standard cost factors					Sc
Budgets					Bu
Supply status - in inventory and on order					So

Figure 4

An Assignment Profile for Production
Planning, Operation, and Control

Activities	Plant Mgmt.	Shops	Shipping and Receiving	Plant Operations Control	Plant Accounting	Production Scheduling	Purchasing	Distribution Mgmt.	Sales/Production Liaison
	1/1	1/1/6	1/1/20	1/1/22	1/1/23	1/2	1/4	2/2	2/3
1. Determine Mfg. lead time requirements						[1]			
2. Determine short term plant output requirements, based on sales expectations								[2]	
3. Convert output plans into shop operation plans and material requirements						[3]			
4. Determine new orders for materials							[4]		
5. Prepare work-in-progress status reports				[5]					
6. Schedule detailed flow of work in shops				[6]					
7. Perform Mfg. operations		[7]							
8. Route products to warehouses or customers								[8]	
9. Ship products			[9]						
10. Calculate operating costs and variances from standard costs					[10]				
11. Review costs	[11]								

Figure 5

**Subprogramme Assignment Profile
Raw Material Acquisition**

Figure 6

Internal Information Flows Required by
Assignment Profile for Purchasing

	1	1/1	1/1/20	1/1/21	1/2	1/3	1/4	1/4/X	1/5	2/2/X	3/1	5	6/1	6/3
1 Production Mgmt.														
1/1 Plant Mgmt.														
1/1/20 Shipping and Receiving							Os							
1/1/21 Supplies Maintenance and Distribution														Se
1/2 Production Scheduling						Ms								
1/3 Production Engineering							Rs							
1/4 Purchasing				Sh	Us Bz	Ms Cs						Ms Cs	Oq	Se
1/4/X Suppliers														
1/5 Quality Control			Mt											
2/2/X Common Carriers														
3/1 Accounting							Os							Mr
5 Research and Product Development							Rs Os							
6/1 Operations Research					Us Br									
6/3 Computer Operations			Mr				Os					Mr		

raw Material Specifications

MsMs

material Usage Estimates

Us

Buffer stock Recommendations

Bz

Ratings on Suppliers and delivery costs

Rs

recommendation or Choice of Suppliers

CsCs

Delivery Schedules needed for R&D materials

Os

Supply Status: in inventory and on order

Se

Anticipated Shortages

Sh

recommendations on Order Quantities and reorder points

Oq

Orders on Suppliers

Os

Material Received reports

MrMrMr

Materials issued for Testing

Mt

Figure 7

Information Flows of Purchasing Programme

	1	1/1	1/1/20	1/1/21	1/2	1/3	1/4	1/4/X	1/5	2/2/X	3/1	5	6/1	6/3
1 Production Mgmt						Cs							Oq	
1/1 Plant Mgmt				Sh	Us Br	Cs	Os		Tm				Oq	Ss
1/1/20 Shipping and Receiving							Sl			La				
1/1/21 Supplies Maintenance and Distribution							Sl		Tm					So
1/2 Production Scheduling				Sh		Cs			Tm				Oq	Ss
1/3 Production Engineering							Rg	Cs						
1/4 Purchasing				Sh	Us Br	Ms Cs		Cs Sp		Tr		Ms Cs Sp	Oq	Ss
1/4/X Suppliers							Rg Os Sl				Ps			
1/5 Quality Control			Mt											
2/2/X Common Carriers								La			Ps			
3/1 Accounting							Os	B1		B1				Mr
5 Research and Product Development							Rg Os	Cs		La				
6/1 Operations Research				Sh	Us Br		Os							Ss
6/3 Computer Operations			Mr Mt	Ps Mt			Os					Mr		

raw Material SpecificationsMsMsmaterial Usage EstimatesUsBuffer stock RecommendationsBrsupplier CataloguesCsSupplier Price and delivery quotationsSpTransportation RatesTrReviews on Suppliers & delivery costsRgrecommendation or Choice of SuppliersCsCsDelivery Schedules needed for R&D materialsOsPhysical Inventory reportsPlSupply Status - in inventory on orderSsanticipated ShortagesShExpeditingExrecommendations on Order Quantities and reorder pointsOqOrders on SuppliersOsShipping InstructionsSlbills of LadingLaLaMaterial Received reportsMrMrMrMaterials issued for TestingMtMaterials issued for productionMsBills for materials and transportationB1B1Payments for materials and transportationPsTest reports on MaterialsTm

THE EXISTENCE AND CALCULATION OF MEAN SENTENCE
LENGTH FOR A CONTEXT-FREE STOCHASTIC GRAMMAR*

S. Soule

Abstract:

This paper deals with some statistical properties of any context-free, reduced, ϵ -free generative grammar such as that each production rule has some fixed probability of use. The calculation of a necessary and sufficient condition for finite mean derivation length is shown, together with a method for calculating the expected number of occurrences of a given character within a sentence. Both calculations are performed using a pair of real matrices which characterize the statistical properties of the grammar.

A stochastic (generative) grammar is a grammar which is used to generate sentences at random from the language it defines. Each production in a stochastic generative grammar has a weight associated with it which determines the likelihood that the production will be used when it is acceptable. In the context-free case, it may be assumed that the weight is a probability.

*Supported in part by the National Science Foundation (Grant GN-412)

In this paper, it is assumed that the weight does not change over time. Further, only reduced ϵ -free context-free grammars are treated.

M. M. Kherts (1968) has shown how to calculate the entropy per character of the string of sentences generated by such a grammar. However, such a calculation is not very meaningful unless there is a finite mean length for the derivation of a sentence. This paper shows an easily calculated necessary and sufficient condition for finite mean derivation length. In addition, there is shown a direct way to calculate the expected number of occurrences per sentence of a given character.

The following stochastic Dyck grammar is an illustration of the kind of system investigated in this paper. The rules are:

$$p_1: S \longrightarrow a$$

$$p_2: S \longrightarrow (S)$$

$$p_3: S \longrightarrow S S$$

where

$$p_1 + p_2 + p_3 = 1 ,$$

and

$$p_1, p_2, p_3 > 0 .$$

A typical sentence, " $\{a\}(a)\{a\}$ ", is produced with probability $p_1^2 p_2^2 p_3$ since the left-to-right derivation of the sentence entails the application of rules 3, 2, 1, 2, 1, in that order. The mean sentence length for this grammar is finite iff

$$p_1 > p_3 ,$$

when the mean sentence length is

$$\frac{p_1 + 2p_2}{p_1 + p_3} .$$

Notation and terminology follow Ginsburg (1966), except as noted. Paradigmatically, $G = (S, T, S_i, R)$ is a context-free stochastic grammar with n nonterminal characters $S_k \in S$, and m terminal characters $T_k \in T$, root (initial) nonterminal S_i , and a set R of production rules. For each nonterminal S_j , there are r_j rules expanding it; and $r_j \neq 0$. A typical rule is

$$p_{ji}: S_j \longrightarrow x_{ji}$$

where

$$1 \leq j \leq n, 1 \leq i \leq r_j ,$$

$$p_{ji} > 0 ,$$

$$x_{ji} \in (S \cup T)^* ,$$

the length of x_{ji} is finite, and

$$\sum_{i=1}^{r_j} p_{ji} = 1 .$$

Any string over $(S \cup T)^*$ containing no S_j is said to be S_j -free. The mean sentence length (MSL) of G is the mean number of terminal characters present in the sentences generated by G . The mean derivation length (MDL) of G is the mean length of the derivations produced by G ,

i. e., the mean number of rule applications performed by G in the generation of a sentence.

If G has finite MDL, then clearly G has finite MSL. However, if G is not ϵ -free, there may be a nonzero probability of infinite derivations in which nonnull productions occur only a finite number of times. Such a situation could lead to finite MSL, but infinite MDL. A grammar with such unhappy characteristics is of little interest, and so this paper states only a criterion for finite MDL. This criterion is equivalent to a criterion for finite MSL for any ϵ -free grammar.

To state the criterion, one needs a few statistics about the form of G . Because G is context-free, the order of the symbols in x_{ji} does not affect the statistics of sentence length, etc. Therefore, a simple count of items in x_{ji} is adequate. Most important:

$\underline{n_{jik}}$ is the number of occurrences of S_k in x_{ji} .

The matrix A associated with a grammar G is defined to be an $n \times n$ matrix with typical element

$$a_{jk} = \sum_{i=1}^{r_j} p_{ji} n_{jik} ;$$

a_{jk} is the expected number of nonterminals S_k produced when S_j is expanded by a single rule application.

The promised criterion for finite MDL hinges on the calculation of the function Z , here defined. Let V be an $n \times n$ real matrix. Then

$Z_j(V)$ is an $n \times n$ matrix with typical element

$$v_{ik} + \begin{cases} \frac{-v_{ik}v_{kj}}{v_{jj}}, & \text{if } i < j \text{ and } v_{jj} \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

(Z_j normally uses elementary row operations to set to zero all elements of V in the column above v_{jj} , but if $v_{jj} = 0$, Z_j leaves V unchanged.)

Now Z may be defined as a simple composition:

$$Z(V) = Z_1 \dots Z_j \dots Z_n(V).$$

Theorem 1: A reduced grammar G with associated matrix A has finite MDL iff all the diagonal elements of $Z(I - A)$ are positive.

The proof will require two lemmata.

Lemma 1: If A is an $n \times n$ matrix of nonnegative numbers, there exists a stochastic ϵ -free context-free generative grammar G with associated matrix A .

Proof: For $1 \leq j \leq n$, let

$$x_j = \left[\sum_{i=1}^n a_{ji} \right].$$

For any $x_j = 0$, set $x_j = 1$.

For $1 \leq i, j, k \leq n$, let

$$p_{ji} = \frac{a_{ji}}{x_j},$$

$$n_{jik} = 0, \text{ if } k \neq i,$$

II C-6

and

$$n_{jii} = x_j.$$

For $1 \leq j \leq n$, let

$$p_{j(n+1)} = 1 - \sum_{i=1}^n p_{ji}.$$

Let G be (S, T, S_1, R) , where

$$S = \{S_j \mid 1 \leq j \leq n\},$$

$$T = \{a\},$$

and for every pair of positive integers

$$(j, i), j \leq n, i \leq n+1,$$

R contains a rule

$$p_{ji}: S_j \longrightarrow a S_i^{n_{jii}}$$

if and only if

$$p_{ji} > 0.$$

Lemma 2: Let G be a reduced grammar with $n > 1$ nonterminals and associated matrix A such that

$$a_{nn} < 1.$$

Let $(I - \hat{A})$ be the first $n-1$ rows and columns of $Z_n(I - A)$. Then there exists a reduced grammar $\hat{G}$ with associated matrix $\hat{A}$ (and so, $n-1$ nonterminals) such that G has finite MDL iff $\hat{G}$ does.

Proof: Where δ_{jk} is the Kronecker delta,

$$Z_n(I - A) = \left[\begin{array}{c|c} \delta_{jk} - (a_{jk} + \frac{a_{jn}a_{nk}}{1 - a_{nn}}) & 0 \\ \hline -a_{nk} & 1 - a_{nn} \end{array} \right]$$

$$\hat{a}_{jk} = a_{jk} + \frac{a_{jn}a_{nk}}{1 - a_{nn}} \geq 0,$$

because

$$a_{jk} \geq 0,$$

and

$$\frac{a_{jn}a_{nk}}{1 - a_{nn}} \geq 0.$$

So by Lemma 1, $\hat{G}$ exists. $\hat{G}$ is reduced if

$$S_j \Rightarrow S_k \quad (\text{in } \hat{G})$$

whenever

$$S_j \Rightarrow S_k \quad (\text{in } G),$$

or

$$S_j \Rightarrow S_n \Rightarrow S_k \quad (\text{in } G).$$

But $S_j \Rightarrow S_k$ (in G) iff $a_{jk} > 0$, and similarly for $\hat{G}$.

So for $j, k < n$,

$$a_{jk} > 0$$

implies $\hat{a}_{jk} > 0$, and so $S_j \Rightarrow S_k$ (in $\hat{G}$). Also,

$$S_j \Rightarrow S_n \Rightarrow S_k \quad (\text{in } G)$$

$$\text{iff } a_{jn}, a_{nk} > 0,$$

$$\text{so } \hat{a}_{jk} > 0, \text{ and } S_j \Rightarrow S_k \quad (\text{in } \hat{G}).$$

Hence, $\hat{G}$ is reduced.

G is context-free, and so the order in which nonterminals are expanded does not affect the MDL of G . Therefore, without loss of generality, let G be operated so that S_n 's are always expanded first. A given S_n , when expanded, will produce an average of a_{nn} new S_n 's, and when all these have been expanded there will be an average of a_{nn}^2 new S_n 's, etc. Hence, the mean number of rule applications it takes to produce an S_n -free string from the original S_n will be

$$\sum_{i=0}^{\infty} a_{nn}^i = \frac{1}{1 - a_{nn}},$$

which is known to converge because by hypothesis $a_{nn} < 1$. Suppose there is an S_n -free string in the derivation, and G expands an S_j within that string. A mean of a_{jn} new S_n 's will be produced, resulting in an expected

$$\frac{a_{jn}}{1 - a_{nn}}$$

rule applications to produce the next S_n -free string in the derivation.

The MDL of G must be greater than the mean number of S_n -free strings per derivation, and less than or equal to the mean number of S_n -free

strings per derivation, times

$$\left(\frac{1}{1 - a_{nn}} \right) \max(a_{1n}, \dots, a_{(n-1)n}).$$

Thus, G has finite MDL iff the mean number of S_n -free strings per derivation is finite.

Given some S_n -free string in a derivation, if G expands S_j , the mean number of new S_k 's in the next S_n -free string is

$$\hat{a}_{jk} = a_{jk} + \frac{a_{jn} a_{nk}}{1 - a_{nn}}.$$

Thus, the mean number of S_n -free strings per derivation is finite iff the MDL of $\hat{G}$ is finite.

Proof of Theorem 1: For a G with one nonterminal ($n = 1$), a derivation consists of expanding $S_1 = S_n$ into an S_n -free string. As in the proof of Lemma 2, this takes a mean of

$$\sum_{i=0}^{\infty} a_{11}^i = \frac{1}{1 - a_{11}},$$

rule applications, which converges to the indicated expression iff

$$a_{11} < 1.$$

But

$$Z(I - A) = [1 - a_{11}],$$

so all the diagonal elements of $Z(I - A)$ are positive iff G has finite MDL.

For the rest of the proof, let G have $n > 1$ nonterminals, and assume the proposition is true for any grammar with fewer than n nonterminals.

Suppose G has finite MDL. Because G is reduced, there is nonzero probability that a derivation will entail the expansion of S_n . Hence, the expansion of S_n into an S_n -free string must take a finite average number of rule applications. In general, that number is

$$\sum_{i=0}^{\infty} a_{nn}^i,$$

which has just been shown to converge, and so $a_{nn} < 1$. Then by Lemma 2, $\hat{G}$ exists. G has finite MDL, so $\hat{G}$ has finite MDL. All elements of the diagonal of $Z(I - \hat{A})$ are therefore positive. The diagonal of $Z(I - A)$ is the diagonal of $Z(I - \hat{A})$ with $1 - a_{nn}$ (which is positive) appended, so all the elements of the diagonal of $Z(I - A)$ are positive.

Suppose all the elements of the diagonal of $Z(I - A)$ are positive; does G have finite MDL? Clearly,

$$1 - a_{nn} > 0,$$

so $\hat{G}$ exists. Then all the elements of $Z(I - \hat{A})$ are positive, and $\hat{G}$ has finite MDL. Hence, G has finite MDL.

A second matrix B (analogous to A) associated with a grammar G is defined to be an $n \times m$ matrix with typical element

$$b_{jk} = \sum_{i=1}^{r_j} p_{ji} m_{jik},$$

where

m_{jik} is the number of occurrences of T_k in x_{ji} .

Thus, b_{jk} is the expected number of terminals T_k produced by a single expansion of S_j .

Theorem 2: Let G be a grammar with associated matrices A and B , and finite MDL. Let c_{jk} be the typical element of the $n \times m$ matrix

$$C = (I - A)^{-1} B.$$

Then c_{jk} is the expected number of occurrences of T_k in a sentence produced by G with root nonterminal S_j . Consequently, the MSL is

$$\sum_{i=1}^m c_{ji}.$$

Proof: $(I - A)^{-1}$ exists because $\det(I - A)$ is positive, because $\det(I - A) =$ the product of all the diagonal elements of $Z(I - A)$, which are all positive, because G has finite MDL.

Let f_k be the expected number of times a nonterminal S_k will be expanded during the derivation of a sentence. Recall that a_{ik} is the expected number of new S_k 's produced in one expansion of S_i . Then, since S_j is the root nonterminal,

$$f_k = \delta_{jk} + \sum_{i=1}^n f_i a_{ik}.$$

Let

$$F = (f_1, \dots, f_1, \dots, f_n).$$

Then

$$\begin{aligned} F &= FA + (\delta_{j1}, \dots, \delta_{j1}, \dots, \delta_{jn}), \\ &= (\delta_{j1}, \dots, \delta_{jn})(I - A)^{-1}. \end{aligned}$$

The expected number of terminals T_k in an entire sentence is

$$\sum_{i=1}^n f_i b_{ik},$$

and so the vector giving the expectancies for all T_k is

$$FB = (\delta_{j1}, \dots, \delta_{jn})(I - A)^{-1}B.$$

But that is just the j -th row of

$$(I - A)^{-1}B.$$

Lastly, the mean sentence length must be simply the sum of the mean numbers of all the (terminal) characters it contains.

Corollary: The MDL of a sentence produced by G with root nonterminal

S_j is the j -th component of

$$(1, \dots, 1)(I - A)^{-1}.$$

Proof: The MDL is

$$\sum_{i=1}^n f_i,$$

where f_i is as developed in the proof of Theorem 2.

Finally, the calculation of $Z(I - A)$ may be regarded as a partial calculation of $(I - A)^{-1}$ by the Gaussian elimination method, so that these calculations may be overlapped for efficiency.

REFERENCES

1. Ginsburg, S. (1966), "The Mathematical Theory of Context-Free Languages," New York, McGraw-Hill.
2. Kherts, M. M. (1968), Entropy of languages generated by automated or context-free grammars with a single-valued deduction, Nauchno-Tekhnicheskaya Informatsia, Ser. 2, No. 1, 29-34; partial translation in Automation Express, 10, 5-7.

NUMERICAL EXPERIMENTS ON PAIR CORRELATIONS
AND ON "THERMODYNAMICS"

R.H. Miller

A "conventional" n-body calculation was constructed to carry out a number of numerical experiments on small stellar systems. Two of these experiments will be reported here, although something of the others will creep in. All these experiments refer to a 32-body system, in which the equations of motion were handled "exactly", with the usual checks on constancy of the ten first integrals of motion. Because the emphasis was on the experiments, the calculation was constructed without particular attention to running speed or minimum storage requirements. Otherwise, the calculation was quite conventional.

I. "Thermodynamics"

An n-body calculation provides one way of attempting the thermodynamic "gedankenexperimente" of using a stellar system as the thermodynamic medium in a Carnot engine. These experiments have not gotten as far as putting the system into a Carnot engine, although routines have been constructed to do that. The work to be reported was undertaken to explore the practical problems that might be encountered. How long would a system have to sit to reach an equilibrium state? How fast could a cylinder

wall move and still carry the system through a sequence of quasi-equilibrium states? Or is an equilibrium state attainable at all? The interaction of a stellar system with a "thermodynamic enclosure" -- even one without moving walls -- can provide a good starting-point.

The system was placed in a box. Whenever a star struck the box, it was caught and thrown back into the box. It was re-projected randomly directed inward, with a randomly selected speed such that its new kinetic energy would be exponentially distributed with a mean value that was called the "temperature" of the box. The star was re-projected from the same point, so there was no change of potential energy. Changes of momentum and of kinetic energy were tallied. Finally, the box was endowed with a "heat capacity", such that the difference of kinetic energy received and given up could cause a change in the "temperature" of the box. The exponential distribution of kinetic energy is appropriate to a Maxwellian velocity distribution. The random velocities of re-projection are satisfactory for all parts of the Carnot cycle, although a specular reflection might be easier to apply to the adiabatic part.

Two sequences of experiments were run. For the first sequence, the box was "cold" -- its temperature was always zero. The second sequence used non-zero temperatures and various heat capacities. The final equilibrium state of the second sequence should define a thermodynamic temperature for the stellar system by the classical definition of temperatures from the equilibrium of two systems in contact, if indeed such

an equilibrium can be reached. It is not clear that an equilibrium can be reached, even by ascribing a negative heat capacity to the box.

The system should quickly readjust itself to something it likes from any initial condition. In fact, initially the particles randomly filled a rectangular parallelepiped in the phase space whose projection onto the configuration space and onto the momentum space was a cube in each space. The cube edge in configuration space was adjustable, but always within the box, while the cube edge in momentum space was chosen to make the ratio of kinetic to potential energy correspond to the virial theorem.

The cold box was typical of most experiments. Since there is an infinite energy store available, arbitrarily large amounts of energy can be transferred to the enclosure. The history of energy transfer to the enclosure in the longest run is shown in Figure 1 (solid curve; the dashed curve shows a shorter run with a different starting condition to illustrate the reproducibility of the experiments). The rate of energy transfer to the enclosure is nearly constant over the duration of the experiment -- it does not diminish or accelerate appreciably over the duration of the run -- even though the total energy transferred to the enclosure was about 4 times the initial kinetic energy present in the system. (The initial kinetic energy was about 250 in the units of the ordinate). It was characteristic of these runs that the virial theorem was approximately satisfied without allowing for "surface pressure" terms.

Thus, near the end of the run, the stellar system was very small and compact, somewhere within the box, with very high kinetic energy. The time-scale in Figure 1 is the time-variable of the differential equations, which is the time-scale of an observer looking at the cluster from outside. In time-units appropriate to the cluster (crossing-times, for example), the rate of energy transfer decreases. The crossing-time (estimated from the virial theorem) was about 1 time-unit at the beginning of the calculation, but was about 1/10 time-unit at the end.

There is no preference for any one star to fly out -- the identity of stars that strike the box seems random. Occasionally, it seems that the same one comes out repeatedly, but the effect is never significant. At the end of the long run, the selection of stars that have struck the enclosure seems to be drawn from the appropriate multinomial distribution.

There was very little change in the average amount of energy carried by a star that struck the enclosure during the calculation; a possible slight trend toward increasing the amount of energy carried is heavily masked in the fluctuations. The average energy delivered to the box was about the same as the average initial kinetic energy. Similarly, the rate at which stars struck the enclosure did not show a detectable trend. The distribution of intervals between times when stars struck the box seemed random -- it looked like a Poisson process with a fixed rate at about 10-12 per initial crossing time. There was no detectable serial correlation of intervals.

The zero-temperature case is interesting because the stars flying out are caught, their kinetic energy removed, and they are merely released where caught. They start falling toward the center of the remaining cluster -- and have very little chance to avoid passing through the most dense part. This arrangement is very efficient at extracting energy from the cluster. It is almost a Maxwell demon.

The rate of energy transfer to the box was about half the total initial kinetic energy in the system per initial crossing time.

The second series of experiments, that with nonzero box temperature, did not tend toward an equilibrium. It tended to have the star system give up energy to the box. With the box temperature colder than the stellar system, this is what would be expected. But with the box hotter than the system, the system still gave up energy to the box. This is not in the direction expected of a system with a definable "specific heat". (See the discussion by Lynden-Bell and Wood, 1968, with which a good deal of this work can be compared.) The general trends are: (1) with the box temperature up to about twice that of the average star, the energy transfer to the box occurred at nearly the rate observed with a cold box. (2) With the box temperature about four times that of the average star, the rate was appreciably diminished. (3) With the hotter box, the fluctuations in the energy transferred were quite large -- fluctuations as great as half the initial kinetic energy of the system were not unusual, in spite of the bound set on the potential energy by

placing the system in a box. A general trend of transferring energy to the box was superposed on the large fluctuations.

Qualitatively, the results of these experiments were anticipated. It was expected that fluctuations in the complexion of a stellar system would lead to localities that were hot, and in which a "heat imbalance" would exist compared to the rest of the system. The questions to be answered by experiment had to do with whether these fluctuations would last long enough to lead to a locally unstable situation. The usual continuum theories (see, e.g., Lynden-Bell and Wood, 1968, Larson 1970) tend to minimize or to ignore these fluctuations. (Formally, they are there, hidden in some of the higher moments of the distribution functions; when the sequence of moment equations is closed by relating high-order moments to lower order moments, the fluctuations are ignored at that order of the moment equations.) Fluctuations should become less important as the number of particles increases, but the amplifying effect of the localized "runaway" makes fluctuations have a much larger effect than expected -- their importance does not diminish like the square root of the particle number, for example. A naive interpretation of these experiments would lead to the conjecture that a stellar system might always be unstable against such "runaways". The importance of localized fluctuations does not seem to have been stressed in discussions applied to stellar systems.

Similarly, with the cold box, the unanticipated result is quantitative. The rate of energy transfer to the box is surprising. It goes much faster than Larson's (1970) discussion would lead you to expect.

II. Pair Correlations

Pair correlations are elusive quantities that are expected in stellar systems, but which have not even been conclusively demonstrated to exist, much less measured. It has been widely accepted, from the earliest days of n-body calculations, that binaries form in the calculations. This implies the existence of pair correlations. But that statement has not been turned into anything quantitative.

In the BBGKY hierarchy, the two-body distribution function can be expressed as the product of two one-body terms plus a pair correlation term: $f^{(2)} = f^{(1)} \cdot f^{(1)} + g^{(2)}$. The term $g^{(2)}$ is the pair correlation. In any real situation, the pair correlation is heavily masked by the product of one-body distribution functions. These experiments were undertaken in hopes of making some statements about the form of $g^{(2)}$; however, it seems that about all that can be done is to show that there is probably something there.

A 32-body system was started from an initial condition in which the angular momentum was accurately zero, so it should reduce to something spherically symmetrical in a rather short time. The initial-state angular momentum was about 10^{-10} of what it would have been had simple random

loadings been used as they were for the other experiments described in this paper. Under spherical symmetry, the two-body terms can depend on 8 invariant quantities and the single-particle terms on 3 each. These quantities will not be listed because the important point is to find convenient physically meaningful combinations of them rather than just to provide a simple enumeration. The game is to find combinations that should behave differently with a nonzero $g^{(2)}$ from what would be expected if it were zero -- ideally, combinations whose behavior should be radically different. This is not particularly simple, and I am not at all sure that I have used anything like the best discriminant.

Taking the binaries as a model, it might be expected that the $g^{(2)}$ term would manifest itself through a tendency in configuration space for stars to be closer than a random distribution and in momentum space for them to be farther apart, but farther apart in a certain specified way. Of course, there could be a kind of "anti-binary" pairing that was the other way around -- the tests used so far would probably not find such a term. The tests that I have been able to devise are reasonably specific as to a model; they are certainly not uniformly powerful against all alternatives.

In configuration space, the two stars of a pair and the centroid of the system form a triangle. The orientation of that triangle is unimportant because of the assumed spherical symmetry, but the lengths of its three sides or the lengths of the two sides that intersect at the

centroid together with their included angle are invariants. In the absence of pair correlations, the cosine of that angle should be uniformly distributed. The test used so far consists in seeing whether the cosine is uniformly distributed.

Over some 15 different experiments, a weak tendency toward an excess of cosines above about 0.9 has shown up. In any one case, there is less than a 1% chance that the distribution is uniform -- in some cases it is very much less than that. A typical case is probably at about the 0.1% level. There is no experiment that does not show an effect at the 1% level. Thus, it seems quite likely that something is there. Further statements are not possible at this time; in particular quantitative measures or possible functional forms for $g^{(2)}$ cannot be obtained from the data as yet.

Some further observations on these experiments:

(1) There is considerable variation from experiment to experiment. The effect shows up clearly in spite of this variation.

(2) A modification of the calculation (described below) to cause an iterative refinement of the first ten integrals of the motion made no noticeable difference so far as this experiment is concerned.

(3) All experiments showed weak correlation through the cosine of the included angle criterion. Particular cases were:

(a) Initially uncorrelated systems developed a correlation.

(b) Systems with initial correlation retain some correlation, if the

initial correlation is moderate.

- (c) Systems with strong initial correlation lose some of the initial correlation, and relax toward about the same degree of correlation as the initially uncorrelated systems achieve.
- (d) Some associated pairings of the strongly correlated systems remain associated for a long time. These initial correlations looked something like binaries.

(4) Very few bound binaries formed, according to a test based on the ratio of kinetic to potential energy for the pair, treated as if it were isolated. This ratio is too large for binding in all but a very few cases. Even pairings that remained associated were not genuine binaries by this test.

(5) Two cases run with the near forces cut off showed the same kind of correlation as the other normal experiments.

(6) One case started with very tight correlation in the initial state weakened the correlation and extended the range of the correlation terms farther and farther as the calculation proceeded.

(7) With 32 stars, there are 496 pairs. The effect shows up with 25-50 pairs having the cosine of the included angle greater than about 0.93, where the expected number is about 17. The distribution would be binomial in the absence of an effect.

(8) An experiment, started with much too tight a correlation in the initial state, developed like a stellar system consisting of binaries --

the binaries remained bound as the system evolved. This, of course, is just what real stellar systems do.

(9) The correlation developed typically within about a crossing-time and persisted thereafter to the end of the calculation. The calculations usually were carried out only to 2 or 2.5 crossing-times.

III. Other Experiments

Two other experiments were run which will be mentioned only briefly.

A) Iterative refinement of the energy and angular momentum.

The first ten integrals of the motion can be conserved to arbitrary precision by iterative refinement. The representative point in the phase space is shifted back onto the hypersurface in which these integrals are conserved after it has drifted off the hypersurface because of inexact integration. The problem, of course, is that the exact point on the hypersurface to which the computed point should be moved is not known. Ambiguity is avoided by using a least-squares criterion. The local gradient of each of the first ten integrals consists of quantities that are readily available: constants, coordinates or velocity of a particle, or components of the force acting on a particle. The least-squares condition can be met by restricting the movement of the computed point to the subspace spanned by the gradients of the integrals. A vector displacement should be a linear combination of these gradients. The tool that accomplishes this is the generalized inverse (since the matrix is not square).

This procedure has been incorporated into the calculation. It turns out to be unnecessary to refine the configuration and velocity centroids. The remaining matrix is $4 \times (6n)$; the computation of the generalized inverse is dominated by the 4 and not by the $6n$, so the amount of computation required to compute the corrections is not prohibitive. One iteration will usually hold the energy to $\pm .01$ (with total energy around 200) and angular momentum components to about $\pm 10^{-4}$. Two iterations typically hold about 10^{-10} in energy and 10^{-13} in angular momentum. Occasionally, extra iterations are required. The calculation was run with a tolerance of 10^{-15} for the sum of squares of the deviations of energy and of the three components of angular momentum. Normally, two iterations sufficed, occasionally three were required. Once, in an extremely close encounter, the calculation was terminated on failure to reach the tolerance in 6 iterations. This happened too infrequently to require remedies.

Experimentally, it is not clear that the refinement improves the calculation appreciably. I have the impression that there are fewer close encounters in the calculation with the integrals better conserved. However, two calculations started from the same initial values, one with and the other without the iterative refinement, soon become two distinct calculations which cannot really be compared with one another. It would be very difficult to strengthen the impression into something like a reasonable experimental result.

B) Parallel calculations comparing two systems.

This sequence of experiments was undertaken as an extension of my earlier work along these lines, as reported at the Thessaloniki conference, for example (Miller 1966). This time, a perturbation calculation is carried out in which the equation of motion for the difference-vector (in the phase-space) is integrated along with the original system. The advantage is that several calculations can be carried along simultaneously at little extra cost.

Three kinds of initial displacements were used: arbitrary, along the trajectory, and perpendicular to the trajectory but in the integral hypersurface. Experimental results show: (1) The projection of the difference-vector is preferentially along the trajectory through the phase space, but the direction-cosine of the difference-vector onto the trajectory swings from positive to negative. (2) The projection of the difference-vector onto the gradient of the energy is essentially random. It drifts about slowly, but never gets very large. (3) Displacements along the trajectory can develop difference-vectors with components not along the trajectory only through numerical errors. They develop such components.

The iterative refinement of the first integrals was developed primarily to use with this series of experiments. The magnitude of the difference-vector grows at about the same rate whether the integrals are conserved or not. The effects of numerical imprecision noted

earlier cannot be avoided merely by using computational methods that conserve the first integrals better.

The calculations reported in this paper were carried out at the Kitt Peak National Observatory. Their hospitality and generosity in making the computer available for this work is gratefully acknowledged.

REFERENCES

1. Larson, R.B. 1970, M.N. 147, 323-37.
2. Lynden-Bell, D., and Wood, R. 1968, M.N. 138, 495-525.
3. Miller, R.H. 1966, I.A.U. Symposium No. 25, Theory of Orbits in the Solar System and in Stellar Systems, G. Contopoulos, Ed. pp. 137-40 (New York: Academic Press).

Figure Caption

Figure 1. Energy transferred to a "cold" enclosure by a 32-particle stellar system. The solid line is one calculation, the dashed line another from similar initial conditions. The initial kinetic energies were about 250 in the units of the ordinate. The time-units of the abscissa are approximately crossing-times of the initial state.

Fig.1

THE FUZZY TYCHONOFF THEOREM

J. A. Goguen

Much of topology can be done in a context where open sets have fuzzy boundaries. One of the more interesting results is the Tychonoff theorem, the validity of which depends on a proper relation between the quality of fuzziness and the number of component spaces. This paper is self-contained and reports research¹ primarily from 1966. We begin by delimiting the spaces we shall use for measuring fuzziness.

Definition: A clog is a complete lattice L with an additional associative binary operation $*$ such that the lattice zero 0 is a zero for $*$, the lattice infinity 1 is an identity for $*$, and the distributive law

$$a * \bigvee_{i \in I} b_i = \bigvee_{i \in I} a * b_i$$

holds for all $a, b_i \in L$, and all index sets I .

Recall that $0 \in L$ is a zero for $*$ iff $0 * a = a * 0 = 0$ for all $a \in L$.

"clog" is an acronym for "complete lattice ordered semigroup", but our definition of "clog" is more restrictive in that the infinity must also

¹ Supported by Office of Naval Research contracts Nonr 222(85) and Nonr 656(08).

be an identity (the word "clom" might be better). Note we do not assume $*$ is commutative. Any completely distributive lattice is a clog with $*$ = $\wedge$, so both the closed unit interval $\mathbb{I} = [0, 1]$ and its subset $\mathbb{I}_k = \{j/k \mid 0 \leq j \leq k\}$ for $k \in \mathbb{P}$ (the set of positive integers) are clogs, and so is any complete Boolean algebra; in particular the classical truth set $2 = \{0, 1\} = \mathbb{I}_2$ is a clog with the usual ordering. In a Boolean algebra one may also consider $a + b = (a' \wedge b) \vee (a \wedge b')$ for $*$. $\mathbb{I}$ is also a clog with $*$ ordinary multiplication, and this is perhaps the most typical example. The set of all subsets of a monoid and the set of all ideals of a ring with unit are clogs, under $*$ "multiplication of complexes", i.e. $A * B = \{a \cdot b \mid a \in A, b \in B\}$. The set $\mathbb{N}$ of nonnegative integers ordered by "is divisible by" with $*$ ordinary multiplication is a clog ($a \vee b$ turns out to be greatest common divisor). It is not hard to check that a product of clogs is a clog, so that $\mathbb{I}^n$ and $(\mathbb{I}_k)^n$ are clogs for any $n \in \mathbb{P}$. 2 is a subclog of any clog with $0 \neq 1$. The set of all open subsets of any topological space forms a complete distributive lattice (for any upper complete lattice is lower complete; in this example $\bigwedge_{i \in I} A_i$ turns out to be $(\bigcap_{i \in I} A_i)^o$; that is, the infimum is the interior of the intersection). If the space is taken to be $\mathbb{I}$ with $a = [0, \frac{1}{2})$, $b_i = (\frac{1}{2} - \frac{1}{i}, 1]$ for $i \in \mathbb{P}$, we obtain in the open subsets of $\mathbb{I}$ a clog in which the dual of the distributive law does not hold. The following is useful later.

Lemma: For a, b elements of a clog L , $a*b \leq a \wedge b$.

Proof: $a \leq 1$, so $1*b \leq (a \vee 1)*b = a*b \vee 1*b$, that is $b = (a*b) \vee b$, and therefore $a*b \leq b$. Similarly $a*b \leq a$. $\square$

Actually, we can consider suitable complete multiplicative categories as truth sets, and essentially all the material of this paper goes through (see [4]).

The "sets with fuzzy boundaries" mentioned above are described precisely by the following (see [5]). The case $L = \mathbb{I}$ was the one first treated by Zadeh [7].

Definition: An L-set A on a set X is a function $A: X \rightarrow L$. X is called the carrier of A , L is called the truth set of A , and for $x \in X$, $A(x)$ is called the degree of membership of x in A .

Although the definition makes sense for any sort of L , we shall always assume L is a clog so as to get a nice algebra of L -sets. Let Y^X denote the set of all functions from X to Y , and let a_X denote the constant function on X with value a , for $a \in Y$. Then the class of all L -sets is a clog under the pointwise lattice and semigroup operations, with zero 0_X and identity 1_X . For the class of L -sets is L^X , which is (isomorphic to) the X -fold product of L with itself, $\prod_{x \in X} L_x$, where $L_x = L$; and any product of clogs is a clog. To be explicit about the operations, let A_i be L -sets on X for $i \in I$; then $\bigvee_{i \in I} A_i$ is the L -set on X defined by

$(\bigvee_{i \in I} A_i)(x) = \bigvee_{i \in I} A_i(x)$ for $x \in X$. If $\mathcal{C} \subseteq L^X$, it will be convenient for us to write $\bigvee \mathcal{C}$ for $\bigvee_{C \in \mathcal{C}} C$. Similarly $A * B$ is defined pointwise by $(A * B)(x) = A(x) * B(x)$, and $A \leq B$ iff $A(x) \leq B(x)$ for all $x \in X$. We say B contains A . It is sometimes convenient to write 1_X for 1_X and 0 for 0_X , for obvious reasons. The reader should think² of $\bigvee_i A_i$ and $\bigwedge_i A_i$ for $A_i \in L^X$ as the union and intersection of the sets A_i .

More generally (as remarked in [2]) if $\mathcal{X}$ is any equational class of algebras and $L \in \mathcal{X}$, then $L^X \in \mathcal{X}$, since L^X is $\prod_{x \in X} L_x$ as above and any equational class is closed under products (see Cohn [1]; or Linton [6] who also treats infinitary operations). For example, if L is a Boolean algebra, so is L^X , and all the laws of Boolean algebra are valid for L -sets; or when L is a clog, the above Lemma applies to L -sets.

We now have all the notation we need to give the main concept of this paper (L is assumed to be a clog).

Definition: An L -topological (or just L -) space is a pair $\langle X, \mathcal{Q} \rangle$ such that X is a set, $\mathcal{Q} \subseteq L^X$, and

$$(1) \mathcal{C} \subseteq \mathcal{Q} \Rightarrow \bigvee \mathcal{C} \in \mathcal{Q}$$

$$(2) A, B \in \mathcal{Q} \Rightarrow A * B \in \mathcal{Q}$$

$$(3) 0_X, 1_X \in \mathcal{Q}.$$

² It can be shown (see [4]) that these definitions are inevitable in the sense that they satisfy the appropriate universal properties in a category of L -sets which is uniquely determined by a set of axioms which are plausible assertions about concepts. The notions of image, inverse image and product are similarly determined.

It should be noted that 2 -topological spaces are just the ordinary topological spaces. Moreover, the axioms imply that $\mathcal{Q}$ is a clog. Elements of $\mathcal{Q}$ are called open sets, and $\mathcal{Q}$ is called an L -topology for X . Any subclog of L^X is an L -space; in fact, the L -spaces on X could be more compactly defined³ as the subclogs of L^X (it is required that the inclusion preserve $\bigvee$ and $*$, but not necessarily $\bigwedge$; more generally, we require the same of a clog homomorphism). Let us give one example now; more will occur later. Let $X = L = \mathbb{I}$, and let $\mathcal{Q}$ be all monotone (nondecreasing) functions to $(0, \frac{1}{2}] \cup \{1\}$, plus the function 0_X . This class is closed under $\bigvee$ and $*$ (in $\mathbb{I}^X$), but not under $\bigwedge$ (in $\mathbb{I}^X$), although $\mathcal{Q}$ has an infimum of its own.

Most of the concepts of ordinary topology generalize. For example, the interior of $B \in L^X$ is $\bigvee \{A \in \mathcal{Q} \mid A \leq B\}$; then infimum in $\mathcal{Q}$ is given by interior of infimum in L^X . $\mathcal{C} \subseteq \mathcal{Q}$ is a cover iff $\bigvee \mathcal{C} = 1_X$ and $\langle X, \mathcal{Q} \rangle$ is compact iff every open cover has a finite subcover. $\mathcal{B} \subseteq \mathcal{Q}$ is a basis iff each $A \in \mathcal{Q}$ is of the form $\bigvee \mathcal{C}$ for some $\mathcal{C} \subseteq \mathcal{B}$, and $\mathcal{S} \subseteq \mathcal{Q}$ is a subbasis iff $\mathcal{S}^* = \{S_1 * \dots * S_n \mid S_i \in \mathcal{S}, n \in \mathbb{P}\}$ is a basis. Any $\mathcal{S} \subseteq L^X$ serves as a subbasis for a unique L -topology on X , said to be generated by $\mathcal{S}$, namely $\mathcal{Q} = \{\bigvee \mathcal{C} \mid \mathcal{C} \subseteq \mathcal{S}^*\} \cup \{0_X, 1_X\}$; the proof of condition (2) for $\mathcal{Q}$ requires a generalization $\bigvee_{i \in I} a_i * \bigvee_{j \in J} b_j = \bigvee_{i \in I \times J} a_i * b_j$ of the distributive law for L ; but this generalization is

³ This observation is valid even when $L = 2$ and seems to shed some light on even the classical case: a topology on X is a subclog of 2^X .

actually equivalent to the original (see [2]). A function $f: X_1 \rightarrow X_2$ is continuous for topologies $\mathcal{Q}_i$ on X_i iff $A_2 \in \mathcal{Q}_2 \Rightarrow f^{-1}(A_2) \in \mathcal{Q}_1$, where $f^{-1}(A_2)(x_1) = A_2(f(x_1))$ for $x_1 \in X_1$ defines the inverse image (see [4]). It is easy to verify that the continuous image of a compact set is compact (the image of $A \in L^{X_1}$, also written $A \leq X_1$, under $f: X_1 \rightarrow X_2$, is defined by $f(A)(x_2) = \bigvee \{A(x_1) \mid f(x_1) = x_2\}$ for $x_2 \in X_2$). Here one uses the more general notion, $\mathcal{C} \subseteq \mathcal{Q}$ covers $A \in L^X$ iff $\bigvee \mathcal{C} \geq A$, so that A is compact iff every cover of A has a finite subcover. The fact that $f^{-1}(f(A)) \geq A$ is also needed.

If A and B are L -sets with carriers X and Y , we define the product $A \times B$ on $X \times Y$ by $(A \times B)(x, y) = A(x) * B(y)$. Arbitrary finite products are defined similarly. If $\langle X_i, \mathcal{Q}_i \rangle$ are L -topological spaces for $i \in I$, we define their product $\prod_{i \in I} \langle X_i, \mathcal{Q}_i \rangle$ to be $\langle X, \mathcal{Q} \rangle$, where $X = \prod_{i \in I} X_i$ is the ordinary set product and $\mathcal{Q}$ is the topology on X generated by the subbase $\mathcal{S} = \{p_i^{-1}(A_i) \mid A_i \in \mathcal{Q}_i, i \in I\}$, where $p_i: X \rightarrow X_i$ is the projection onto the i^{th} coordinate. $\mathcal{Q}$ is the weakest topology such that each p_i is continuous.

Definition: Let α be a cardinal. We say that the identity 1 in a clog L is α -isolated iff whenever $|I| \leq \alpha$, $a_i \in L$ for $i \in I$ and $a_i < 1$, then $\bigvee_{i \in I} a_i < 1$.

Here $|I|$ denotes the cardinality of I . For example, 1 is α -isolated in Π_k for every α (let us say that 1 is absolutely isolated), and 1 is

α -isolated in $\mathbb{I}$ for every finite α (let us say 1 is finitely isolated), but 1 is not ω -isolated in $\mathbb{I}$ (where ω is the cardinality of $\mathbb{N}$; i.e., countable). Also, 1 is α -isolated in $\mathbb{I}^J$ iff $\alpha < |J|$ and $\alpha < \omega$, and in $(\mathbb{I}_k)^J$ iff $\alpha < |J|$.

Theorem 1: Let 1 be α -isolated in L and let $|I| \leq \alpha$. Then the product $\prod_{i \in I} \langle X_i, \mathcal{Q}_i \rangle$ of compact L -topological spaces $\langle X_i, \mathcal{Q}_i \rangle$ is compact.

The following converse to Theorem 1 shows it is, in a sense, the best possible result.

Theorem 2: If 1 is not α -isolated in L , then there is a collection $\langle X_i, \mathcal{Q}_i \rangle$ for $i \in I$ and $|I| = \alpha$ of compact L -spaces such that the product $\langle X, \mathcal{Q} \rangle$ is noncompact.

Proof: Non- α -isolation gives I with $|I| = \alpha$ and $a_i < 1$ for $i \in I$ such that $\bigvee_i a_i = 1$. Let $X_i = \mathbb{N}$ for $i \in I$, and let $\mathcal{Q}_i = \{0, (a_i)_{[n]}, 1\}^*$, where a_S , for $a \in L$ and $S \subseteq X_i$, denotes the function equal to a on S and 0 on $X_i - S$, where $[n] = \{0, 1, \dots, n-1\}$, and where $\mathcal{S}^*$ indicates L -topology generated by $\mathcal{S}$ as subbasis. Then $\langle \mathbb{N}, \mathcal{Q}_i \rangle$ is compact, as every $A \in \mathcal{Q}_i$ except 1 is contained in $(a_i)_{\mathbb{N}}$, and $(a_i)_{\mathbb{N}} < 1$, so that $\mathcal{C} \subseteq \mathcal{Q}_i$ is a cover iff $1 \in \mathcal{C}$. Therefore every cover has $\{1\}$ as a subcover.

But $\langle X, \mathcal{Q} \rangle = \prod_{i \in I} \langle X_i, \mathcal{Q}_i \rangle$ is noncompact. Let a_{in} denote the open L -set $p_i^{-1}((a_i)_{[n]}) \in \mathcal{Q}$, for $i \in I$ and $n \in \mathbb{N}$; it is given for $x \in X$ by the formula $a_{in}(x) = a_i$ if $x_i < n$, and $a_{in}(x) = 0$ otherwise. $\{a_{in} \mid i \in I, n \in \mathbb{N}\}$

is an open cover with no finite subcover. First, $\bigvee_n a_{i,n} = (a_i)_X$ since $\bigvee_n (a_i)_{[n]} = (a_i)_\mathbb{N}$ in $\langle X_i, \mathcal{Q}_i \rangle$, and p_i^{-1} preserves suprema⁴, so that $\bigvee_{i,n} a_{i,n} = \bigvee_i (\bigvee_n a_{i,n}) = \bigvee_i (a_i)_X = 1$. Second, no finite subcollection $\{a_{i,n} \mid i,n \in J\}$, $|J| < \omega$, can cover, since $\bigvee_{\langle i,n \rangle \in J} a_{i,n}(x) = 0$ for any x with $x_i \geq \bigvee \{n \mid \langle i,n \rangle \in J\}$ for any i , and infinitely many such x always exist. $\square$

Thus we can state the Fuzzy Tychonoff Theorem as "Every product of $\mathcal{Q}$ compact L -spaces is compact iff 1 is $\mathcal{Q}$ -isolated in L ", but we prefer to break the result into two pieces because the techniques of proof are so different. Before turning to the proof of Theorem 1, let us discuss some examples. Since $\mathbb{I}$ is finitely but not ω -isolated, finite products of compact $\mathbb{I}$ -spaces are compact, but countable or larger products need not be. For $L = \mathbb{I}^2$, Theorem 2 provides an example of two compact L -spaces whose product is noncompact. In fact for this case, we do not need X_1 and X_2 to be countable, but can take them to be the one point set $\{0\}$. Since the classical truth set 2 has 1 absolutely isolated, Theorem 1 gives the usual Tychonoff theorem, every product of compact 2 -spaces is compact. This strong result also holds for $L = \mathbb{I}_k$ for any $k \in \mathbb{P}$. However, there are two compact 2^2 -spaces whose product is noncompact; these spaces cannot be taken with only finitely many points.

⁴That inverse images preserve suprema can be shown by a small calculation; it also follows from the general category theoretic properties of the constructions; see [4].

These examples cover what are probably the most important truth sets.

We do not have closed sets⁵ available for proving Theorem 1, and we should try to avoid arguments involving points, as the fuzzification of a point is an L-set. We therefore proceed by first generalizing Alexander's theorem on subbases and compactness as in Kelley [5]. This result requires Zorn's lemma, which we state first. A subset $\{B_i | i \in I\}$ of a partially ordered set $\mathcal{F}$ is a chain iff for each $i, j \in I$, either $B_i \leq B_j$ or $B_j \leq B_i$. $B \in \mathcal{F}$ is an upper bound for a subset $\{B_i | i \in I\}$ of $\mathcal{F}$ iff $B \geq B_i$ for all $i \in I$. An element M of $\mathcal{F}$ is maximal iff for no $B \in \mathcal{F}$ is it true that $M < B$.

Zorn's Lemma: If each chain in a nonempty partially ordered set has an upper bound, then the set has a maximal element.

Theorem 3: If $\mathcal{B}$ is a subbase for $\langle X, \mathcal{A} \rangle$ and every cover of X by sets in $\mathcal{B}$ has a finite subcover, then $\langle X, \mathcal{A} \rangle$ is compact.

⁵ Closed sets can be had if L has an involutory endomorphism satisfying the generalized de Morgan laws; $(\bigvee_i a_i)' = \bigwedge_i a_i'$ and $(\bigwedge_i a_i)' = \bigvee_i a_i'$. For we can define A to be closed iff A' is open. However, it is in general false that $A \vee A' = 1_X$ and $A \wedge A' = 0_X$, so that this notion of closed will not in general be as useful as in the Boolean case. In most of the truth sets we considered as examples, a suitable operation is $a' = 1 - a$. In a Boolean algebra ordinary compliment will do, and closed sets will be more natural. However, not all clogs admit such endomorphisms.

Proof: Let us say $\mathcal{C} \subseteq \mathcal{Q}$ has the finite union property (abbreviated FUP) iff $\forall \mathcal{C}_0 < 1_X$ for all $\mathcal{C}_0 \subseteq \mathcal{C}$ such that $|\mathcal{C}_0| < \omega$. Then $\langle X, \mathcal{Q} \rangle$ is compact iff no $\mathcal{C} \subseteq \mathcal{Q}$ with FUP is a cover. Let $\mathcal{C}$ have FUP, and let $\mathcal{F} = \{\mathcal{B} \subseteq \mathcal{Q} \mid \mathcal{C} \subseteq \mathcal{B} \text{ and } \mathcal{B} \text{ has FUP}\}$; we now prove from Zorn's Lemma that $\mathcal{F}$, ordered by inclusion, has a maximal member, to be denoted $\tilde{\mathcal{C}}$. $\mathcal{F}$ is nonempty since it contains $\mathcal{C}$. Let $\mathcal{B}_i \in \mathcal{F}$ for $i \in I$ be a chain. Then clearly $\mathcal{C} \subseteq \bigcup_i \mathcal{B}_i$, and it remains to show that $\bigcup_i \mathcal{B}_i$ has FUP. Let $\mathcal{B}_0 \subseteq \bigcup_i \mathcal{B}_i$ be finite; then each element of $\mathcal{B}_0$ appears first in some $\mathcal{B}_i$, and therefore all of $\mathcal{B}_0$ appears in the largest, say $\mathcal{B}_m$, of this finite set of $\mathcal{B}_i$. Since $\mathcal{B}_m$ has FUP, $\forall \mathcal{B}_0 < 1_X$. Therefore $\bigcup_i \mathcal{B}_i$ is an upper bound for the chain $\mathcal{B}_i$, and $\tilde{\mathcal{C}}$ exists.

If $\mathcal{C} \notin \tilde{\mathcal{C}}$ but $\mathcal{C} \in \mathcal{Q}$, then no open set containing $\mathcal{C}$ belongs to $\tilde{\mathcal{C}}$. For $\mathcal{C} \notin \tilde{\mathcal{C}}$ iff $\exists A_1, \dots, A_m \in \mathcal{Q}$ such that $\mathcal{C} \vee \bigvee_{i=1}^m A_i = X$, because of maximality. Then $D \geq \mathcal{C}$ would imply $D \vee \bigvee_{i=1}^m A_i = X$, so we must have $D \in \tilde{\mathcal{C}}$ also if $D \in \mathcal{Q}$.

If $\mathcal{C}, D \notin \tilde{\mathcal{C}}$ but $\mathcal{C}, D \in \mathcal{Q}$, then $\mathcal{C} * D \notin \tilde{\mathcal{C}}$. For $\mathcal{C} \vee A = X$ and $D \vee B = X$, where $A = \bigvee_{i=1}^m A_i$ and $B = \bigvee_{i=1}^n B_i$, so that $(\mathcal{C} * D) \vee A \vee B = X$, and this implies $\mathcal{C} * D \notin \tilde{\mathcal{C}}$. For $(\mathcal{C} \vee A) * (D \vee B) = X * X = X$ gives $(\mathcal{C} \vee A) * D \vee (\mathcal{C} \vee A) * B = (\mathcal{C} * D) \vee (A * D) \vee (\mathcal{C} * B) \vee (A * B) = X$, by distributivity.

By the Lemma (after the definition clos_g), each of the last three terms

is $\leq A$ or $\leq B$, so that $X \leq (C * D) \forall A \forall B$. This result extends to $C_1, \dots, C_n \notin \tilde{\mathcal{C}}$ and $C_1, \dots, C_n \in \mathcal{A}$ imply $C_1 * \dots * C_n \notin \tilde{\mathcal{C}}$.

It now follows that if $C_1, \dots, C_n \in \tilde{\mathcal{C}}$ but are open, and $A \geq C_1 * \dots * C_n$ and $A \in \mathcal{A}$, then $A \in \tilde{\mathcal{C}}$. The contrapositive of this says if $A \in \tilde{\mathcal{C}}$ and $C_1 * \dots * C_n \leq A$ for $C_i \in \mathcal{A}$, then some $C_i \in \tilde{\mathcal{C}}$, and this is the form we use.

Now assume that $\mathcal{S}$ is a subbasis for $\langle X, \mathcal{A} \rangle$ such that each subfamily of $\mathcal{S}$ with FUP is a noncover, and let $\mathcal{C} \subseteq \mathcal{A}$ have FUP. We have to show that $\tilde{\mathcal{C}}$ is a noncover. By the preceding, there is a maximal $\tilde{\mathcal{C}} \supseteq \mathcal{C}$ and having FUP. It will suffice to show that $\tilde{\mathcal{C}}$ is not a cover.

$\mathcal{S} \cap \tilde{\mathcal{C}} \neq \emptyset$ and $\mathcal{S} \cap \tilde{\mathcal{C}}$ has FUP, so $\mathcal{S} \cap \tilde{\mathcal{C}}$ is not a cover. It will therefore suffice to prove that $\bigvee \tilde{\mathcal{C}} \leq \bigvee \mathcal{S} \cap \tilde{\mathcal{C}}$. Since $\mathcal{S}$ is a subbase, each $A \in \tilde{\mathcal{C}}$ is of the form $\bigvee_{i \in I} S_{i1} * \dots * S_{in_i}$, for $n_i \in \mathbb{P}$. Then $S_{i1} * \dots * S_{in_i} \leq A$ for all $i \in I$, and by the observation above, we must have $S_{ij(i)} \in \mathcal{S} \cap \tilde{\mathcal{C}}$ for some $j(i) \in \mathbb{P}$ for each $i \in I$. It then follows that $A = \bigvee_i S_{i1} * \dots * S_{in_i} \leq \bigvee_i S_{ij(i)}$, for $S_{ij(i)} \in \mathcal{S} \cap \tilde{\mathcal{C}}$, that is, $\bigvee \tilde{\mathcal{C}} \leq \bigvee \mathcal{S} \cap \tilde{\mathcal{C}}$. $\square$

Proof of Theorem 1: We have $\langle X_i, \mathcal{A}_i \rangle$ compact for $i \in I$, $|I| \leq \alpha$, and $\langle X, \mathcal{A} \rangle = \prod_{i \in I} \langle X_i, \mathcal{A}_i \rangle$, where $\mathcal{A}$ has the subbasis $\mathcal{S} = \{p_i^{-1}(A_i) \mid A_i \in \mathcal{A}_i, i \in I\}$. By Theorem 3, it suffices to show that each $\mathcal{C} \subseteq \mathcal{S}$ with FUP is a noncover. Let $\mathcal{C} \subseteq \mathcal{S}$ with FUP be given, and

for $i \in I$ let $\mathcal{C}_i = \{A \in \mathcal{A}_i \mid p_i^{-1}(A) \in \mathcal{C}\}$. Then each $\mathcal{C}_i$ has FUP;
 for if $\bigvee_{j=1}^n A_j = 1_{X_i}$, where $p_i^{-1}(A_j) \in \mathcal{C}$, then $\bigvee_{j=1}^n p_i^{-1}(A_j) = 1_X$,
 since $p_i^{-1}(1_{X_i}) = 1_X$ and p_i^{-1} preserves $\bigvee$. Therefore $\mathcal{C}_i$ is a noncover
 for each $i \in I$, and there exists $x_i \in X_i$ such that $(\bigvee \mathcal{C}_i)(x_i) = a_i < 1$.
 Now let $x = \langle x_i \rangle \in X$ and $\mathcal{C}'_i = \{p_i^{-1}(A) \in \mathcal{C} \mid A \in \mathcal{A}_i\}$. Then $\mathcal{C} \in \mathcal{B}$
 implies $\mathcal{C} = \bigvee_i \mathcal{C}'_i$; and $(\bigvee \mathcal{C}_i)(x_i) = a_i$ implies $(\bigvee \mathcal{C}'_i)(x) = a_i$, since
 $(\bigvee \mathcal{C}'_i)(x) = \bigvee \{p_i^{-1}(A)(x) \mid A \in \mathcal{A}_i \text{ and } p_i^{-1}(A) \in \mathcal{C}\} = \bigvee \{A(p_i(x)) \mid p_i^{-1}(A) \in \mathcal{C}$
 and $A \in \mathcal{A}_i\} = (\bigvee \mathcal{C}_i)(x_i)$. Therefore $(\bigvee \mathcal{C})(x) = \bigvee_{i \in I} ((\bigvee \mathcal{C}'_i)(x)) = \bigvee_i a_i < 1$,
 since 1 is α -isolated in L , each $a_i < 1$, and $|I| \leq \alpha$. $\square$

APPENDIX

Those already acquainted with categorical algebra may find these further remarks on matters in footnotes and matters in general, of some interest. First, there is a category $\underline{\text{Sets}}(L)$ of L -sets, for any partially ordered set L : its objects are L -sets; a morphism $f: A \rightarrow B$, where $A: X \rightarrow L$ and $B: Y \rightarrow L$, is a (set) function $\tilde{f}: X \rightarrow Y$ such that $A(x) \leq B(\tilde{f}(x))$ for all $x \in X$; and composition is just composition of the underlying set functions. It has been shown that a category $\underline{C}$ is equivalent to $\underline{\text{Sets}}(L)$ for some completely distributive lattice L iff $\underline{C}$ satisfies certain simple axioms. It has also been shown that these axioms are plausible assertions about inexact concepts (this argument is philosophy rather than mathematics); see [4]. It is fairly easy to show that when L is a completely distributive

lattice, $\underline{\underline{\text{Sets}}}(\mathbf{L})$ has images, inverse images, unions, intersections, and products in the usual categorical sense, and that these universal constructions are given by the formulas given in this paper; $\underline{\underline{\text{Sets}}}(\mathbf{L})$ also has exponentials (and is therefore cartesian closed) for these $\mathbf{L}$. These results extend in various ways to weaker $\mathbf{L}$. For example, if $\mathbf{L}$ is a clog, $\underline{\underline{\text{Sets}}}(\mathbf{L})$ is a multiplicative (monoidal) category with the product using $*$ as multiplication, and this operation has a right adjoint; thus $\underline{\underline{\text{Sets}}}(\mathbf{L})$ is closed monoidal category, and in particular there is a distributive law for multiplication over the usual (categorical) coproduct. That is, $\underline{\underline{\text{Sets}}}(\mathbf{L})$ an instance of the categorical generalization of a clog, a distributive multiplicative category. Such categories can be used a truth sets for a more general species of fuzzy sets.

There is also a category $\underline{\underline{\text{Clog}}}$ of clogs with clog morphisms. This category is complete and even has an exponential. For each clog $\mathbf{L}$, there is a functor $E_{\mathbf{L}}: \underline{\underline{\text{Sets}}} \rightarrow \underline{\underline{\text{Clog}}}^{\text{op}}$, assigning to a set X the clog $\mathbf{L}^X$, and to a function $f: X \rightarrow Y$ the clog morphism $f^{-1}: \mathbf{L}^Y \rightarrow \mathbf{L}^X$.

The category $\underline{\underline{\text{Top}}}(\mathbf{L})$ of $\mathbf{L}$ -topological spaces with continuous maps can be described as the full subcategory of the comma category $\underline{\underline{\text{Clog}}}^{\text{op}}/E_{\mathbf{L}}$ whose objects are just the subclog inclusions. This category has for products the construction we have given, provided $\mathbf{L}$ is a completely distributive lattice. When $\mathbf{L}$ is a clog, our product renders $\underline{\underline{\text{Top}}}(\mathbf{L})$ another multiplicative category. This paper can be described as

determining the extent of closure under multiplication of the subcategory of compact L -spaces, as a function of L . Many results extend to L a distributive multiplicative category.

REFERENCES

- [1] Cohn, P.M., Universal Algebra, Harper and Row, New York, 1965.
- [2] Goguen, J.A., "L-fuzzy Sets", Jnl. Math. Anal. and Applics. 18, 145-174, 1967.
- [3] Goguen, J.A., "The Logic of Inexact Concepts," Synthese, 19, 325-373, 1969.
- [4] Goguen, J.A., "Representing Inexact Concepts," Institute for Computer Research Quarterly Report, No. 20, IIA1-III19, 1969.
- [5] Kelley, General Topology, van Nostrand, Princeton, New Jersey, 1955.
- [6] Linton, F., "Some Aspects of Equational Categories," Proc. Conf. on Categorical Algebra, La Jolla 1965, Springer-Verlag, New York, 1966.
- [7] Zadeh, L.A., "Fuzzy Sets," Inform. and Control 8, 338-353, 1965.

NUMERICAL EXPERIMENTS IN COLLISIONLESS SYSTEMS

R. H. Miller

Gravitational n-body calculations handling 10^5 particles have become routine in the last three or four years. The plasma physicists have also run calculations with comparable numbers of particles--you will hear of these from Dawson this afternoon. In the stellar dynamics area, two groups have been making these calculations--our own group with Kevin Prendergast and William Quirk, and the competition, the group that formed when Roger Hockney and Frank Hohl merged (and later parted ways). Our calculation is described in Miller and Prendergast (1968); theirs is best described by Hohl and Hockney (1969), and some very thorough discussions of computational speed of force calculations by Hockney (1970).

The interpretation of these calculations is beset by the same kinds of difficulties that trouble any numerical experiments--the astronomical value results from the questions asked and the kinds of experiments performed. And, all too often, the experimenter's prior prejudices come through unaffected by any experimental evidence. The astronomical interpretations of the results have been

published elsewhere (Hockney and Hohl, 1969; Miller, Prendergast, and Quirk, 1970; Quirk, 1970).

In this report, I want to take a different tack: this is supposed to be a meeting of experts in n-body calculations, so I want to concentrate on difficulties with the calculations. In particular, the emphasis will be on attempts to convince you that the bits running around inside those nice, big computers bear some relationship to the physics of stellar systems. It is not ipso facto evident that they do: the mere fact that the experimenter intends his calculation to relate to some kind of system in the sky does not assure any similarity.

These calculations deal with particles moving on a plane under $1/r^2$ forces. The ability to handle large numbers of particles comes from the fact that forces are computed only at a restricted set of points--if needed, forces are obtained for intermediate locations by interpolation rules. This avoids the need to compute forces between particle pairs, and makes the amount of computation necessary to obtain the forces independent of the number of particles. Of course, the main computational difficulty in gravitational n-body problems comes from the need to handle close encounters--or the possible formation of binaries. In these calculations, the forces are cut off for close encounters, thus sidestepping that problem. The justification

ultimately lies in the observation that we know how to correct stellar dynamical calculations for the divergences at close encounters, but the long-range effects are the feature that defies theoretical treatment.

There is nothing magic about two dimensions. Three dimensional calculations have been made (Miller and Alton, 1968). Storage requirements and the amount of computation increase explosively in going from two to three dimensions, and the problem of displaying the results in a comprehensible manner is much more difficult in three dimensions. There are so many challenging problems in two dimensions that there is little urge to go on to three. We just haven't gotten to three dimensions yet.

The early models used periodic configuration spaces, but both groups now have gotten rid of the periodic replications. A rather coarse grid for computing the forces remains.

The astronomically interesting problems so far attacked using these computer models are (1) the "Jeans instability"--gravitational collapse of an initially uniform plane system, (2) attempts at static self-consistent models, and (3) spiral patterns.

II F-4

If you concentrate on the spiral patterns, you are led to the other two, and the spirals are most interesting, so we'll follow that route today.

The spiral patterns obtained do some of the things that you think spirals should do. For example, individual mass points ("gas" particles) can be seen to move through the spiral features (Figures 1 & 2). The usual galactic parameters can be computed-- the density of "stars" and of "gas" separately (Figure 3), the epicyclic frequencies and the angular velocity of the pattern (Figure 4), and the mean circular and radial velocity components (Figure 5). So far everything looks very nice and reasonable--but a closer look at Figure 5 shows that the "stars" have half the local circular velocity. Aside from complicating the definitions of epicyclic frequencies, forcing care to use gradients of the forces rather than "Oort constants" determined from the mean velocities, this means that an equilibrium model must be pressure-supported, or must have very large velocity dispersions. This is the first clear sign of trouble. While some real galaxies have significant differences between gas velocities and star velocities (Code, 1967), our own galaxy, for example, shows very little difference between velocities determined from 21-cm observations and those determined optically from the stellar

motions. Scaled to our own galaxy, the pressure support required to make up this difference would require velocity dispersions in excess of 100 km/sec, quite different from the 30-40 observed.

Both groups have obtained spiral patterns. Neither has obtained spirals in a pure stellar case, and the artifices used are interesting in themselves. Hockney and Hohl have placed their system in a background potential--essentially a Schmidt model. They then get spirals that live as long as the dollars last to explore them. Our group used a two-component system in which the component that looked like "gas" showed a spiral pattern. With both groups, the reason that spirals could not be obtained in a pure stellar case was that the pure stellar cases got too "hot"--the velocity dispersions or pressure became very large. Figure 6 shows a curve of mean velocities determined for a purely stellar system back in the days when we were first trying to produce spirals in stellar systems by cooling them to make them move with the circular velocity. Figure 6 gave the condition just before that kind of "cooling" was attempted. Again, the local standard of rest has half the local circular velocity. This condition of "hot" populations is present in Hohl's calculations, too. Figure 7, which appeared in an unpublished report Hohl was kind enough to send, shows Q , the ratio of the actual velocity dispersion in the computer model to Toomre's (1964) velocity dispersion required

II F-6

to stabilize the system against axisymmetric modes. Figure 7 is one of Hohl's more extreme cases, but it shows that these models can get very "hot." Toomre (1964) calculated that the velocity dispersions observed in our galaxy are, if anything, just a little bit too small for comfort under the stability criterion.

It is tempting to speculate that we may have constructed models that were a little bit too "cool" and underwent a gravitational collapse. But is the price a stellar system must pay for having violated the stability criterion (somewhere, not even everywhere) a subsequent heating to Q of at least 2, and more likely 3? If so, it is difficult to see how any galaxy could be as cool as our own.

Here, then, is the major problem that I want to stress. Why are these systems so hot? No one has yet succeeded in building a "cool" stellar model. And no one has seen a reasonable looking spiral in a "hot" population. We have gotten around this by introducing a second population that is strongly "cooled" all the way through the calculation; a spiral pattern appeared in one of these experiments. Hohl has placed a stellar system in a background potential--a Schmidt (1965) model of the galaxy--and then finds that spirals can form and persist. But even then, Hohl finds Q to be $1\frac{1}{2}$ to 2 or more--so even these models are "hot". Hohl is now undertaking experiments at "weak cooling" in which the velocity dispersion is very gently reduced

taking care not to get too close to the Toomre stability criterion.

Hohl's experiments using a fixed background potential, and some of our own along the same lines, leave the impression that the systems do not build up this high velocity dispersion merely to have a way of providing a potential field that will not collapse under its own weight. Toomre's models (1963, 1964) can be stabilized without such a large velocity dispersion, and should provide a noncollapsing potential. But are nonaxisymmetric models harder to stabilize? Ng's (1967) axisymmetric models are just as hot, but they were constructed in such a way that it was practically assured at the outset that they would turn out to be hot.

The "gas" population seems to do more than merely to provide a population with small velocity dispersion that could feel irregularities in the potential to show up a spiral pattern. There is a cooperation between the two populations to make the spiral pattern (Miller, Prendergast, and Quirk 1970; Quirk 1970). So the dissipative character may be necessary; but Hohl's model in a background is not dissipative.

This rather aimless discussion is included to point out the difficulty of really coming to grips with the essential problem: Why are the stars so hot? Almost any discussion of these models comes back to that fundamental question. Sometimes the point comes up

in discussion that we shouldn't worry about needing a two-population system of "gas" and "stars" in order to build a model that shows a spiral pattern, since real galaxies show their spiral patterns in gas and in the hot young stars that have been created out of the gas but cannot live long enough to leave the neighborhood of the gas. That is fine for rationalizing a need for two populations; but real galaxies do not have stellar populations that are anywhere nearly this hot. So, even with two populations, we don't have a good model of a real galaxy. With one of the populations replaced by an "equivalent gravitational potential," the remaining population is uncomfortably hot.

It is not clear where the difficulty lies--but as experimentalists, we must adopt the viewpoint that it most likely lies in our experiments. Unfortunately, computer experiments cannot be used to prove that systems must reach this "hot" condition--the result need not follow from all initial conditions. If the argument is made probabilistically, that we have somehow sampled the parameter space of possible initial states, a careful look will convince you that very few initial conditions have actually been tried. And even then, the search stopped once spirals were found. It was much more exciting to explore the properties of those spirals. The only way to prove that "hot" systems

are not necessarily the only final state for a computer model is to build a "cool" model. But no one has yet built a "cool" static self-consistent model.

The answer may lie in studies of the stability of purely "stellar" systems. There are, of course, some features of the computer models that make the experiments seem strained or artificial. In a sense, neither the usual theoretical model nor the computer model properly represents a real stellar system. Each is an approximation that overlooks certain essential features. The approximations are complementary. The theoretical (continuum) models overlook the grainy structure of a stellar system, while the computer models are much grainier than actual stellar systems. It appears that a "grainy" fluctuation in the force field tends to destabilize a stellar system, so continuum models may underestimate the difficulty of achieving stable systems, while computer models may exaggerate it. Computer models have other features that stellar systems lack. The discrete number representation in a computer makes it impossible to define an infinitesimal perturbation --any change is finite (although it may be quite small). This restricts the methods available for studying stability. Distinctions (such as between instability and failure to be static-self-consistent), which are clearly defined for the continuum theoretical models may not

be meaningful in computer experiments. The distinction might be meaningful for a stellar system and still be difficult or impossible to apply operationally because an instability cannot be recognized.

The importance of the question, quite aside from a natural desire to make our models be as close to nature as possible, lies in the interpretations that are otherwise made of things like Toomre's stability criterion and the observed velocity dispersions. They figure heavily in the arguments of Lin, Yuan, and Shu (1969), for example, in discussions of whether spiral galaxies that we observe are stable configurations.

Some attempts to construct static self-consistent models will be taken up next, then a fairly long discussion of some numerical properties of the Miller-Prendergast (1968) model as a preliminary to a different way of undertaking the construction of self-consistent models. Along the way, the advertised features of our model as providing an exact handling of the collision-free Boltzmann, or Vlasov, equation will be pointed out.

II. SELF-CONSISTENT MODEL

The models that we have used as initial conditions for various calculations have not been designed to be static self-consistent models. Invariably, the models have evolved through spectacular collapses into systems that were largely pressure-supported rather than being "cool" in the sense of having small velocity dispersions.

The "hot" condition does not seem to be necessary for disk galaxies, but the velocity dispersions of our evolved computer models are substantially greater (perhaps twice or more) than required for stability, and are disconcertingly large for comparison with real stellar systems. A static self-consistent model that can be maintained in the computer with smaller velocity dispersions is essential for studying actual galaxies, particularly if the galaxies are to be built without resorting to several populations.

This work was undertaken to see whether a "cool" static self-consistent model could be maintained in the computer. A successful attempt would show that there is nothing intrinsic to plane stellar systems that requires unusually large velocity dispersions in order to remain stable. The model was tested for its stability limits and to see what kinds of instabilities might develop. A complication in

such a study is that the onset and early growth of an instability cannot be readily detected. Rather, the instability must grow to finite amplitude to be detected--unless we can devise a more sensitive indicator. A genuine instability is difficult to distinguish from mere failure to construct a truly static self-consistent model. Three-dimensional modes, such as bending, are not allowed.

a) The Model

The model chosen was essentially "Model I" of Toomre (1963). All of the standard galactic parameters can be easily worked out for this model, and the stability criterion of Toomre (1964) can be applied as well. This model has, in obvious notation:

$$\rho(\varpi, z) = \mu(\varpi) \delta(z) = \frac{M b}{2\pi} (\varpi^2 + b^2)^{-3/2} \delta(z), \quad (1)$$

for which the corresponding potential is

$$V(\varpi, z) = -G M [\varpi^2 + (b + |z|)^2]^{-1/2}. \quad (2)$$

Here b is a scale parameter that sets the dimensions of the system. If all particles have unit mass, M is just the number of particles.

In the initial conditions generated for a machine calculation, a number of particles was chosen (125,000), and a scale parameter b . With $b = 13$, the number of particles per location drops below $1/2$ at a radius $\varpi = 80$. Velocities were chosen for the particles at each configuration space location to make a Gaussian (Schwarzschild) velocity distribution whose mean is the local circular velocity and with a dispersion whose axis ratios conformed to the usual ones computed from the Oort constants (see e.g. Chandrasekhar 1960, Sec. 4.3) with the actual root-mean-square velocities being a multiple, T , of a convenient form that is large at the center and decreases outward. The actual velocity of each particle was generated by a pseudorandom number generator, to meet these conditions.

The velocity dispersion that results with Toomre's stability criterion is large enough to require allowance for pressure-support. This is done by reducing the local circular velocity according to the usual hydrodynamical equations of stellar dynamics (see, e.g. Chandrasekhar 1960, Sec. 4.8 (iv)). It can happen (at T about 1.5 that of the Toomre condition), that the system becomes wholly pressure supported in the center--there is then no rotation in those parts of the system that are pressure-supported.

b) Modification to the Actual Force Law

The force between pairs of particles, used in the calculation, is derivable from a potential

$$\varphi(x, y) = \text{const.} [(x-x')^2 + (y-y')^2 + a^2]^{-\frac{1}{2}}. \quad (3)$$

This modification was used to avoid computational troubles from near encounters. It is convenient to think of this potential in a 3-dimensional space. The particles move in a plane, at $z = \text{constant}$, and the force on a test particle is measured a units above that plane, at $z = a$. Only those force components that lie in the plane $z = a$ enter the problem--the test particles is not permitted to leave the plane. For most calculations $a = 3$. This change, which seems small, produces a surprisingly large change in the actual force field near the center of a configuration like that of Eq. (1). Corrected expressions for the parameters involving the force (rotational velocity, Oort constants, epicyclic frequency, and so on) can be obtained from the usual expressions by substituting $c = b + a$ in place of b wherever it appears. Thus all the model parameters can be exactly corrected for the revised force law.

A remarkable feature of this force law is that the same notion can be used to make the appropriate modification to the stability calculation.

Toomre has also noticed that this force law permits the stability calculation to be carried through. The result is interesting by itself because it separates long-range from short-range contributions to the predicted stability conditions. Toomre's calculation proceeds from a linearized collision-free Boltzmann equation (Eq. 43 of Toomre 1964) in which the only term affected by the modified force law (assuming that the model representing the unperturbed system is properly self-consistent with the revised force-law) is that containing the force due to the perturbed density distribution. The force for each Fourier mode is obtainable from the usual solution to the Poisson equation,

$$\psi'(x, z) = \text{Const.} \frac{2\pi G}{k} e^{ikx} e^{-k|z|}, \quad (4)$$

which indicates that the replacement $G \rightarrow G e^{-ka}$ wherever G appears will properly account for the revised force law. Toomre's calculation now follows, yielding a modified dispersion relation

$$\frac{k^2 a}{G \mu} = 2\pi \beta \frac{1 - e^{-y} I_0(y)}{\sqrt{y}} e^{-\beta y}, \quad (5)$$

where

$$y = \frac{\sigma_u^2 k_n^2}{\kappa^2}, \quad \beta = \frac{\kappa a}{\sigma_u}, \quad (6)$$

k_n is the wavenumber of the neutrally stable mode, σ_u is the velocity dispersion in the radial direction, and κ is the epicyclic frequency. The system is stable if $\kappa^2 a / G\mu$ is greater than the righthand side of Eq. (5). Let $K(\beta)$ be the maximum value attained by the righthand side of Eq. (5) for any value of $y (\geq 0)$ and the given value of β . Since the lefthand side contains (local) quantities that are completely determined by the model, this relation can be solved for β anywhere in the system. An interesting feature is that $K(\beta)$ attains an asymptotic value $2\pi/e = 2.31145\dots$ as $\beta \rightarrow \infty$ and any value of $\kappa^2 a / G\mu$ greater than this can be stabilized with any value of σ_u including $\sigma_u = 0$, by taking β infinite. Thus, with this modification to the force law (near cutoff in the forces), there are situations in which rotation alone is sufficient to stabilize the system--a result that appears quite different from Toomre's. (The distinction from the results of Lynden-Bell (1962) and Lee (1967) concerning rotational stabilization should be noted: these both refer to three-dimensional mass-distributions.)

In the model designed for the machine calculation, with $(G\Omega)^{\frac{1}{2}} = 3.0165C$ (a value that sets the maximum force to 3.5), the rotation should have been sufficient to stabilize the system inside $\omega \sim 26$. The actual velocity dispersion (in the radial direction) that is required for stability is shown, in units appropriate to the model, in Figure 8. With reasonable values for the velocity dispersions, the greatest instability should occur around $\omega \approx 70$ in these models--fairly far out where the density is already quite low. Furthermore, the velocity dispersions required to stabilize are reduced by at least a factor of two below those expected without allowing for the near cutoff of the forces.

c) Experimental Results

In the machine representation, all physical quantities are integers. To avoid sharp discontinuities in the initial density and velocity fields, we rounded by adding a pseudorandom number, uniform in $(0,1)$ to the "real" value before truncating. The rounding is unbiased, but has twice the variance of the usual rounding rule. The variance acts like a velocity dispersion (with a circular velocity ellipse), and can contribute to the stabilization of the system through

random motions. In fact, this term alone should be sufficient to stabilize the systems with the modified force law.

Roundoff "noise" in the force-values is expected to be destabilizing, in contrast to the stabilizing influence of velocity dispersions. This can be seen by multiplying the linearized Boltzmann equations appropriate to different phase space points, and assuming that cross terms correlating forces to velocities, velocities to positions, and forces to coordinates can be neglected. The term with force covariances enters with the opposite sign to those with velocity variances. In physical terms, fluctuations in the force seem directly to induce fluctuations in the part of the distribution function that depends upon the velocities, broadens the velocity dependence. The effect of a velocity spread behaves differently, since it is already included in the distribution function. Quantitative estimates of the effect are complicated by correlations in the "fluctuations" of force-values at different spatial locations. The force rounding remained with the 'add one-half' rule, to preserve the reversibility that we feel is an important feature of our model.

Two series of experiments were run with these configurations. The experiments started from an initial load produced by the methods just described, then followed the evolution of the system for a few

integration steps. Normally, ten integration steps were sufficient to tell how the system was going to behave.

Each series of experiments was a sequence of runs with different values of T , the parameter that controls the velocity dispersion in the initial load. The two series of experiments were run with different numbers of configuration space locations per periodic cell length. The number of lattice points per periodic length was increased by the square of the factor by which the timestep is reduced, so force-values were unaltered.

The results of each experiment were observed in step-by-step density plots of the configuration and through values of the force tabulated for certain control points in the system. Some runs, duplicated with different sequences of pseudorandom numbers used in the initial loading routine, gave results that were qualitatively similar. The extent of quantitative difference indicates the reliability of the numerical results. The similarities of two systems starting with different pseudorandom number sequences were comparable with the retention of gross symmetry within a given system. Thus, these results are not peculiar to the particular set of pseudorandom numbers used.

The experimental results may be qualitatively described as follows. First, the systems run with our usual 256×256 grid appeared

not to be static self-consistent models. They tended to rearrange themselves rather quickly, although those with larger velocity dispersions did so less rapidly. This probably means that discrete models are not adequately approximated by models designed according to continuous density distributions, although the departure is less troublesome for "hotter" systems. There were no instabilities that were recognizable as such.

Second, the experiments with "reduced time-step" (more points per configuration space lattice), were more nearly static. Runs were made with $1/5$ to $1/31$ of the normal time-step, with T about 1.3 times the value required for stabilization in the absence of "quantization noise." All these runs behaved about the same way, so $1/5$ was used thereafter. A run with $T = 0$ showed a clumping similar to that seen earlier with the plane gravitational instability--very irregular patterns leading to large density variations characterized by fairly short length scales. These seem to be genuine instabilities, as distinguished from mere failure to have achieved a static self-consistent model, although the distinction is not clear-cut. The instability was strongest near the center, where the system should have been rotationally stabilized.

A sequence of runs was made with $\underline{T} = 0, 0.1, 0.2, 0.3, 0.4$ and 0.5 . Stabilization should be achieved (without "quantization noise") at $\underline{T} = 0.267$ for the form of velocity dispersion used. The velocity dispersion was largest at the center and fell off uniformly toward the outside; the precise analytic form is not important. The system becomes completely pressure-supported near the center at $\underline{T} = 0.457$. In general, larger values of $\underline{T}$ lead to systems that are more nearly static, self-consistent, and stable, as expected. At $\underline{T} = 0.1$, an instability is still present but it grows less rapidly. Values of $\underline{T}$ greater than 0.2 do not produce recognizable instabilities, although the systems still rearrange themselves in search of an equilibrium configuration. The rearrangement is slowed with larger values of $\underline{T}$. There does not seem to be a recognizable boundary between genuine instabilities (if that is what they are) and rearrangements. There is rather a uniform trend toward slower rearrangement and longer characteristic lengths with larger values of $\underline{T}$ (larger velocity dispersions).

Hockney and Hohl (1969) have carried out similar experiments with a different model. Their low-velocity-dispersion configurations form into patterns characterized by short length scales; configurations with increased velocity dispersion change more slowly and form into

patterns with larger length scales. Their treatment of the forces effectively introduces a near-cutoff that must influence the analytic stability criterion in some way like that of subsection b; their treatment of the forces also introduces local departures from the analytic form that are equivalent to the stochastic contribution noted above. It is difficult to distinguish an instability from a failure of self-consistency in their models, just as it is in ours. The picture that emerges from their published results seems similar.

d) Discussion

These experimental results seem to disagree with expectations on two counts. (1) Something happens, even though "quantization noise" resulting from the discrete allowed velocities alone should be sufficient to stabilize the system. As noted earlier, "quantization noise" in the forces tends to destabilize the configuration; it works against the effects of the velocity dispersion, and could make the system unstable. There is not a clear-cut stability threshold to compare with theory. (2) That "something" is strongest near the center, a place where rotation alone should be adequate for stabilization.

There are several points at which these experiments fail to meet the conditions of the stability calculations. First, the system

is discrete, so the continuous model is an approximation. It is difficult to estimate the effect of the approximation and even more difficult to do the calculation for a discrete system (but see Section 4.) Second, the onset of the instability is not detected. The instability (if indeed that is what it is) may well have started where theory indicated and later have spread to the center, where we were able to observe it. Third, the configuration is no longer axisymmetric at the time that it is clear that something has happened. We cannot say whether the unstable mode(s) that started the whole process were axisymmetric; the final state that we observe is not. Fourth, the initial state is not strictly axisymmetric. Here again, we do not know how important the "perturbations" of the computer model are.

Our earlier experiences in trying to produce "cool" models led to the conjecture that the "hot" condition resulted from a rearrangement following a gravitational instability which developed at some place that was too "cool." The response to this instability could be a local "heating." The fact that our self-consistent equilibrium systems were so far beyond the stability threshold is disturbing. It suggests that the penalty for violating the stability criterion might be rather severe; that it could lead to "heating" far in excess of that just sufficient to lift the instability, and that this "heating" could extend over the entire system even if it arose in response to a localized instability.

The experiments reported here neither confirm nor disprove this conjecture. The cause cannot easily be tied in with instabilities because it is operationally difficult to identify instabilities as such. The picture evidently can be tied to a failure to be self-consistent as well as to instabilities. Systems that started with low velocity dispersions (small values of $\underline{T}$) rather quickly developed "clumps" of "hot" stars. Those with higher velocity dispersions did not change so rapidly, but did rearrange themselves, doubtless with some resultant "heating." None of these models was run long enough to tell how "hot" the system would have looked if it had reached an equilibrium state, but it seems likely that any one of them would have become as "hot" as our other typical models. A "cool" stable equilibrium was not reached, but this does not prove that none exists.

III. NUMERICAL CONSIDERATIONS

This section will, in the interests of definiteness, be restricted to consideration of the discrete model of Miller and Prendergast (1968). A general description may be found in that reference. The model brings an unusual viewpoint to numerical experiments: the "data" representing the physical system are modified to conform to the model. The physics is expressed in certain invariance properties which are exactly incorporated into the model. An exactly calculable model results--one that is pleasing to work with because of the feeling of precision that it affords.

The transition from the physics to the model is the principal

subject of this section. Not all of the physical invariance properties can be built into a calculation; a matter of taste is involved in the selection of those deemed most important. We feel that the model is close to the physics because it incorporates the physically essential features exactly. The point of view might be expressed by stating that the physics does not live in the differential equations--they are at best one way of describing what is going on.

The invariance properties are incorporated through a simple finite-difference scheme that can be made to represent a discrete phase space by starting it with integer values and thereafter giving it only integer values to work with. The size of the interval between allowed discrete locations is equivalent to the roundoff in a conventional calculation; a matter of some concern is precisely what effects result from intentionally making roundoff and truncation error as large as seems reasonable to be in keeping with the spirit of the problem.

The problem in a discrete phase space may be approached from either of two viewpoints. In the first, which is adopted here, the discrete representation may be regarded as a model approximating the dynamics of continuous systems. The model is capable of exact representation in the computer. All the approximations are made between the physics and the model. The advantage is that the approximations occur at a

well-defined stage of the argument, and once made, leave an exactly calculable system. The investigations of numerical properties then describe the extent to which the model does or does not describe a physical situation. The other viewpoint is to argue, from the outset, that the mechanics appropriate to a discrete system is being developed, with less concern for its possible applicability in the real world. The mechanics of a discrete representation is an interesting problem which will not be developed here; Greenspan (1970) seems to have made some progress along these lines.

This model yields an exact Liouville theorem and is reversible. Because these properties summarize the essential physics of stellar dynamics (in the relaxation-free limit), this scheme is close to the basic physics of the problem. There is a large number of integrals of the motion; the occupancy of any phase location, as the motion maps the phase space into itself, is an integral of the motion. The motion is also exactly reversible (apart from questions as to how the force-values $\underline{f}^{(n)}$ turned out to be integers; this is discussed later). A quantity that looks like the conventional energy integral can be formulated (cf. Greenspan 1970); but it differs in important respects from the usual energy integral. To be more precise, it does not seem to be possible to derive the forces from a scalar potential in an exact way.

Throughout this section, the present calculation is contrasted with conventional gravitational n-body calculations. The term "conventional", as applied to gravitational n-body calculations, means a calculation which, in spirit, attempts to treat everything exactly to the precision allowed by the computer. The forces are usually obtained by explicit summation of contributions due to each pair of particles. The calculations of Aarseth (1963, 1966), Standish (1968), and Wielen (1967, 1968) are conventional in this sense. As a practical matter, it has not proved possible to handle more than about 250 particles in such calculations.

The features setting this model apart from conventional gravitational n-body calculations are quantitative and not qualitative. Other representations are discrete, although with much finer steps than we use. Other integration schemes are finite-difference, although usually the ignored terms are of higher order than in this calculation. The conventional calculations have at least one exact integral (shared by this calculation)--the number of particles. Thus, the working out of numerical details for this calculation indicates how they might similarly be worked out for other calculations. The distinction is a matter of degree, not of kind.

There are certainly some features that are handled much better in the conventional calculations than in this one. Angular momentum conservation is such a feature--in our calculation, angular momentum is expected to be conserved only in a statistical sense. But these disadvantages are more than offset by the ability to handle large numbers of particles.

a) The Finite-Difference Scheme

1. A norm for the state vectors

Discussions of the stability of the finite-difference scheme require a norm that does not change as the step size is altered. This seems to be most conveniently done in the following way. Define a (dimensionless) state vector

$$\xi^{(n)} = \begin{pmatrix} \chi^{(n)} \\ u^{(n-\frac{1}{2})} \end{pmatrix} \quad (7)$$

which, according to our difference-scheme obeys

$$\begin{aligned} \xi^{(n)} &= A \xi^{(n-1)} + B f^{(n-1)} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \xi^{(n-1)} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} f^{(n-1)} \\ &= A^n \xi^{(0)} + \sum_{k=1}^n A^{n-k} B f^{(n-k)}, \end{aligned} \quad (8)$$

the usual iterated discrete-time relations. Although the system may easily be generalized to unequal time-steps, that is not necessary for this calculation. The integer $\underline{x}^{(n)}$ represents a physical length $\underline{X}^{(n)} = \underline{x}^{(n)} \Delta \ell$, where $\Delta \ell$ is the (dimensioned) spacing of the allowed discrete locations. This, in turn, may be related to a dimensionless number representing displacements in units of some (dimensioned) normalization-length $\underline{L}$:

$$\mathbb{X}^{(n)} = \underline{X}^{(n)} / \underline{L} = x^{(n)} \frac{\Delta \ell}{L} . \quad (9)$$

The same thing can be done with the velocities and the forces by introducing the time-step Δt and a normalization time, $\underline{T}$:

$$\begin{aligned} \mathbb{U}^{(n-\frac{1}{2})} &= u^{(n-\frac{1}{2})} \frac{\Delta \ell}{L} \frac{T}{\Delta t} \\ \mathbb{F}^{(n)} &= f^{(n)} \frac{\Delta \ell}{L} \left(\frac{T}{\Delta t} \right)^2 . \end{aligned} \quad (10)$$

Then a new state-vector can be defined:

$$\Xi^{(n)} = \begin{pmatrix} \mathbb{X}^{(n)} \\ \mathbb{U}^{(n-\frac{1}{2})} \end{pmatrix} = N \xi^{(n)} = \frac{\Delta \ell}{L} \begin{pmatrix} 1 & 0 \\ 0 & T/\Delta t \end{pmatrix} \xi^{(n)} , \quad (11)$$

which can be used to define a norm:

$$\|\xi^{(n)}\|^2 = (\Xi^{(n)})^T (\Xi^{(n)}). \quad (12)$$

The superscript T denotes transpose. The independence of step-size works as follows: suppose Δl is decreased. Then $\underline{x}$ must increase to represent the same physical length (requiring both the old and the new $\underline{x}$ to be integers places some constraints on the way Δl is reduced; these only enter the discussion later). But the values of the elements of Ξ are not changed, so the norm is unchanged.

The separation of $\underline{f}^{(n)}$ into an inhomogeneous term allows the particle interactions to be introduced in a way that permits the large matrix that advances the positions and velocities of all particles to be block-decomposed in the manner of Eq. (7).

2. Stability of the Finite-Difference Scheme

Stability is usually defined as the uniform boundedness of the difference between the computed solution (represented by $\xi^{(n)}$) and the exact solution to the system of differential equations, $v(n\Delta t)$, starting from the same initial condition ($\xi^{(0)} = v(0)$) and requiring that the two kinds of solution be

compared at corresponding times. In this sense, a necessary and sufficient condition for stability (see, for example, Wendorff 1966), is that the norm of $\|A\| \leq 1 + c \Delta t + O(\Delta t^2)$.

In fact, for this case,

$$\|A\| = \sup \frac{\|A\xi\|}{\|\xi\|} = 1 + \frac{1}{2} \left(\frac{\Delta t}{T} \right) + O \left(\frac{\Delta t}{T} \right)^2. \quad (13)$$

An interesting feature of this calculation is that $(\Delta \ell / L)$ cancels out in this definition of $\|A\|$ through $\|\xi\|$. This is to be expected because uniformity requires that noninteger $\underline{x}$'s and $\underline{u}$'s be used--and there is nothing about the difference scheme, per se, that requires the $\underline{x}$'s and $\underline{u}$'s to be integers (or rational). The discrete phase space notion is tacked on to the difference scheme, and is not essential to it.

Intermediate degrees of reduction of time-step can also be investigated. Let the normalization length, $\underline{L}$, be the spatial lattice interval of the original calculation. Then, let $m \Delta \ell = \underline{L}$, so that increasing $\underline{m}$ refines the space-lattice. Similarly, let $q \Delta t = \underline{T}$, and let the calculation run for $\underline{\tau}$ time-steps of the initial calculation, so $\tau = q \underline{\tau}$. After $\underline{\tau}$ time-steps of the original calculation, $\underline{A}^{\underline{\tau}}$ is involved, giving

$$\|A^n\|^2 \rightarrow 1 + \frac{\tau^2}{2} + \frac{\tau}{2} \sqrt{4 + \tau^2} \leq 1 + \tau + \tau^2. \quad (13a)$$

The $\underline{m}$ and $\underline{q}$ dependence has dropped out, but the norm stays bounded for finite τ . This argument does not make use of any continuity properties of $\underline{m}$, $\underline{q}$, $\Delta \ell$, or Δt .

b) The Discrete Phase Space

The nice features of this calculation, such as reversibility and the special kind of Liouville theorem (particles initially together remain together), derive from the discrete phase space. That, in turn, comes from the property of the difference scheme that the $\underline{x}$'s and $\underline{u}$'s will remain integers if the $\underline{f}$'s are all integers. It now remains to investigate how this modification affects the stability and accuracy of the calculation. It may be stressed again that this is a modification of the data to force it into the mold of a given model, such that the model can be treated exactly.

There is nothing unusual about selecting an initial condition such that all the $\underline{x}$'s and $\underline{u}$'s happen to be integers. The effect of the calculation with integers appears through the inhomogeneous term of Eq. (7) - the term involving $\underline{f}^{(n)}$.

Let $\xi^{(n)}$ be the vector representing the system with integer forces $\underline{f}^{(n)}$; let $\zeta^{(n)}$ be a corresponding vector representing a

system with "exact" forces, $\psi^{(n)}$. Take the two initially equal: $\xi^{(0)} = \zeta^{(0)}$. Then the norm of the difference-vector describes the additional effect of approximations in the force-values:

$$\begin{aligned} D^2 &= \|\xi^{(n)} - \zeta^{(n)}\|^2 = \sum_{\lambda=1}^n \sum_{s=1}^n (NA^{\lambda-s}B)^T (NA^{s-1}B) \varphi^{(n-s)} \varphi^{(n-\lambda)} \\ &= \left(\frac{\Delta \ell}{L}\right)^2 \sum_{\lambda, s=1}^n \left[\lambda s + (\tau/\Delta t)^2 \right] \varphi^{(n-s)} \varphi^{(n-\lambda)}, \end{aligned} \quad (14)$$

where $\varphi^{(k)} = f^{(k)} - \psi^{(k)}$ is the difference between the approximate and the "exact" forces at step k .

The series (14) can be summed under a variety of circumstances. Only two extreme cases will be given, with indications of the important features in more realistic situations. The first of these is (a) completely uncorrelated: $\varphi^{(j)} \varphi^{(k)} = \langle \varphi^2 \rangle \delta_{j,k}$. Then

$$\begin{aligned} D_A^2 &= \left(\frac{\Delta \ell}{L}\right)^2 \langle \varphi^2 \rangle \left[\frac{n(n+1)(2n+1)}{6} + n(\tau/\Delta t)^2 \right] \\ &\approx \left(\frac{\Delta \ell}{L}\right)^2 \langle \varphi^2 \rangle \left[\frac{n^3}{3} + n(\tau/\Delta t)^2 \right]. \end{aligned} \quad (15a)$$

The second is (b) completely correlated: $\varphi^{(j)} \varphi^{(k)} = \langle \varphi^2 \rangle$ independently of j, k . Then

$$\begin{aligned} \mathcal{D}_s^2 &= \left(\frac{\Delta \ell}{L}\right)^2 \langle \varphi^2 \rangle \left[\left(\frac{n(n+1)}{2}\right)^2 + n^2 \left(\tau/\Delta t\right)^2 \right] \\ &\approx \left(\frac{\Delta \ell}{L}\right)^2 \langle \varphi^2 \rangle \left[\frac{n^4}{4} + n^2 \left(\tau/\Delta t\right)^2 \right]. \end{aligned} \tag{15b}$$

This discussion is independent of the cause of the variation in the forces; in particular it describes the usual roundoff in computations ^{as well as} as well as our more drastic roundoff to produce integer force-values. Case (a) describes situations in which there is no memory in the roundoff from one evaluation of the forces to the next. Most roundoff situations are assumed to be of this kind. A more realistic description results from admitting that rounding in one direction on one integration step is likely to be followed by another rounding in the same direction on the next integration step. Although case (b) is much more stringent than that situation, it contains the essential features. The fourth power of n is typical of those cases with "memory" even if the "memory" lasts over a small fraction of the total duration of the calculation.

The approximate expressions in (15a) and (15b) can be used to investigate the effect of refining the spatial grid and the

time step. The two cases yield:

$$D_A^2 \approx \frac{g^3}{m^2} \left[\frac{\tau^3}{3} + \tau \right] \langle \varphi^2 \rangle \quad (16a)$$

and

$$D_B^2 \approx \frac{g^4}{m^2} \left[\frac{\tau^4}{4} + \tau^2 \right] \langle \varphi^2 \rangle. \quad (16b)$$

With the correlation "memory" extending over shorter times, the dependence of D^2 is very similar to (16b). For a "correlation" that dies out linearly with time, going to zero at τ_c ,

$$D_c^2 \approx \frac{g^4}{m^2} \tau_c \left[\frac{\tau^3}{3} + \tau \right] \langle \varphi^2 \rangle. \quad (16c)$$

The g^4/m^2 dependence remains, as might be expected. In this expression, the correlation lasts for the same time-interval independently of the number of integration steps required to cover that time interval. The partial memory is a reasonably good picture of the rounding of force values in an actual calculation. Particles that don't move much from one integration step to the next, in force fields that don't change much from one integration

step to the next, are likely to round in the same direction in successive integration steps. But after a while, the particle should have moved into a region where the rounding is substantially independent of the earlier history. In a system containing many "particles" with quite different motions, no one "correlation" history can describe the situation--so some kind of "average particle" or even a "worst-case particle" is described by $\mathcal{U}_c$. This matter of rounding the force-values to integers is essentially the only point at which our calculation differs in spirit from other n-body calculations. The correlation-term arising in our calculation is matched by similar terms in more conventional calculations. The difference depends on the degree to which the difference in force-values at neighboring particle locations exceed the roundoff interval between successive allowed force-values and on the kind of difference-scheme used.

An interesting aspect of $\mathcal{D}^2$ in (16b) and (16c) is that it is natural to consider alterations in which the time-step and spatial interval are changed keeping $\underline{q}^2/\underline{m}$ constant; this keeps the forces constant and is the simplest change that does not affect the mass of the "particles." Under such a change, a buildup of $\mathcal{D}^2$ with "uncorrelated" force roundoff terms would go to zero in the limit

$q \rightarrow \infty$, but that contribution to $\mathcal{D}^2$ due to the "correlated" force roundoff is not affected. The limiting process must be carried out in such a way that the spatial lattice is refined faster than the square of the refinement of the time-step to make $\mathcal{D}^2$ go to zero.

The "correlated" contribution to $\mathcal{D}^2$ can be avoided by using rounding rules that eliminate the correlation between successive values of the force. The force-values can be rounded by a rule like entier ($\underline{F} + \underline{\text{random}}$), where random produces a "pseudorandom number" in the interval (0,1), rather than by the usual entier ($\underline{F} + 1/2$), for example. This makes the correlation go to zero at the cost of doubled variance. But it can also lose the reversibility of the calculation.

The uncertainty introduced through the round-off can be quite large. Taking as examples, $\langle \varphi^2 \rangle = 1/12$ from the usual rounding, with $\underline{q} = \underline{m} = 1$ and $\tau = 10$, (16a) gives $\sqrt{\mathcal{D}^2} \approx 6$ spatial units, and (16b) gives ≈ 16 . In conventional calculations where the lattice spacing is much finer, these terms are less harmful.

c) Reversibility

Much has been written on the reversibility of machine calculations (see, for example, Buneman, 1967). Some of the

discussions may leave the impression that finding a suitable reversible algorithm is the main problem. Reversible algorithms may be generated easily by deriving them from variational principles, but the resulting algorithms may be implicit or otherwise ill-suited to machine calculations. Even so, a reversible algorithm may not run reversibly if a different numerical value is computed at some point of the calculation as the system passes in the forward and reverse directions. It is instructive to see how the performance of our difference scheme matches expectations. The formalism that is developed in the course of doing this clarifies the essential features of reversibility of machine calculations.

Reversibility can be discussed in much the same way as the growth of roundoff errors. The same process that causes a calculation to be irreversible could cause two separate ("parallel") calculations to differ if they started from identical initial conditions and ran forward together. Such "parallel" calculations were used to study reversibility in n-body calculations earlier (Miller, 1964).

Two calculations, using our difference scheme, with integer values of $\underline{x}$, $\underline{u}$, and $\underline{f}$, can differ only if the value of $\underline{f}$ at some part of the system has been rounded differently in the two cases.

Let the two systems be described at the n th step by vectors $\xi^{(n)}$, $\eta^{(n)}$, with $\xi^{(0)} = \eta^{(0)}$. The forces governing the system described by ξ are the same as those described by η , but each is to be computed separately. Most of the time, the forces in the two systems will round identically: $f_{\xi}^{(n)} = f_{\eta}^{(n)}$. But once in a while--presumably very rarely, the two will differ somewhere. As soon as this happens, the two systems will rapidly become quite different. The different rounding was the irreversible process. The notion of reversibility implies that it was somehow beyond the experimenter's control. The number of separate integrations that were successfully done before this happened might be used as a measure of the "reversibility."

This definition of reversibility may seem too stringent--it could be argued that all that is required is that the reversed system return to a place "close to" the original starting point (Aarseth, 1966). The definition used here does not make explicit use of the norm that would be required to give meaning to "close to." This definition also seems preferable because it is free of some of the complications of interpretation that go with explicit evaluations of $\|\xi^{(n)} - \xi^{(0)}\|$ (See, for example, Miller, 1964).

Using the language of statistics (with substantial reservations about the language, but the comparison of two systems already implies an ensemble of sorts), we are led to inquire about the probability for the two force-values to be different. Let $\underline{F}$ be the ("exact") force that would be computed by a perfect process, and let the computed value be $(\underline{F} + \delta \underline{F})$. There is somewhere a roundoff boundary $\underline{R}$. The probability for an error in the rounded force is the sum of the probabilities for making a rounding error in either direction. Near one rounding boundary, this is $\mathcal{E}(|\delta F|)$ (the "mean absolute deviation"). If $\mathcal{E}(|\delta F|)$ is essentially independent of $\underline{F}$, the total probability of making a rounding error (in either direction) summed over all possible rounded results is approximately $\mathcal{E} = \mathcal{E}(|\delta F|)$. This can be obtained by considering that the probability for rounding incorrectly between $\underline{R}$ and $\underline{R} + 1$ is $\mathcal{E}(|\delta F|) \times p(\underline{R})$, where $p(\underline{R})$ is the probability to get $\underline{R}$ as a rounded value. When a sum over all possible $\underline{R}$'s is evaluated, $\mathcal{E}(|\delta F|)$ factors out and $\sum p(\underline{R}) = 1$ since $p(\underline{R})$ is a probability. This estimate of the chance of making an error in rounding is based on the notion of an underlying "reasonable" probability distribution governing the force-values. That is an oversimplification, but it seems to provide the only reasonable way to estimate the chance of

making a rounding error, and thus of determining how good (or how bad) the calculation looks from the standpoint of reversibility.

The probability of making a rounding error on one particle in one integration step is $\epsilon \approx \mathcal{E}(|\delta F|)$. In the entire system with many particles and many integration steps, one error somewhere, at some time, will throw the system off the track. This is a "go-no-go" process that is described by a binomial distribution characterized by ϵ . With w the total number of tries, the probability that no error has been made is $(1 - \epsilon)^w \approx 1 - w\epsilon$. When $w \approx 1/\epsilon$, there is a good chance that something irreversible has happened. Then w is the desired "reversibility measure."

In the large n -body calculation, force values are computed for each possible location in the space-lattice. The number of force-evaluations per integration step is twice the number of occupied lattice points (for 2 components of force in a 2-dimensional calculation).

For convenience, assume that all locations are occupied. In our early calculations, there were $(128)^2 = 2^{14}$ locations. The force calculation provided nearly full single register precision for the IBM 360 series with floating point numbers scaled to about 16 for the largest magnitude, giving $\delta F \approx 2^{-19}$. This leads to $w = 2 \times (\text{no. integration steps}) \times 2^{14} \approx 2^{19}$, or about 16 steps to an

irreversible error. This is to be compared with our early experiments (Miller and Prendergast 1968) in which an exact fourfold symmetry was held for about 20 steps. The symmetry test can easily be seen to relate to the same criterion, since symmetrically disposed points were treated differently in the force calculation.

The reversibility of the calculation will be affected by changes in the time-step or in the lattice-spacing. According to Eq. (10), the values of the integer force scale according to $\underline{m}/q^2$, and in a floating point calculation, $\epsilon \propto \bar{\epsilon}(|\delta F|)$ will scale by the same amount. As $\underline{m}$ increases, the number of occupied lattice points will increase, but it cannot exceed the number of particles. Since the number of particles is a few times the number of allowed lattice points, (2-10 times the 2-dimensional calculations; with 3-dimensional calculations this ratio may be substantially higher) this won't make much difference. Then

$$\begin{aligned} w \epsilon &\rightarrow \epsilon \propto \frac{m}{q^2} \epsilon_0 \text{ [number of occupied lattice points]} \\ &\approx \propto \frac{m}{q^2} \epsilon_0 \text{ [Min (no. allowed locations, no. particles)]}. \end{aligned} \quad (17)$$

The number of integration steps (at the original time-step) is γ and ϵ_0 is that "probability" of error in one step at one location. But if the problem is scaled to keep $\underline{q}^2/\underline{m}$ constant, then the reversibility deteriorates with increasing $\underline{q}$. A rule for scaling that simultaneously improves both $\underline{D}^2$ and reversibility seems to be impossible.

Random rounding, suggested in the previous section as a means to reduce $\underline{D}^2$ due to quantization, would destroy reversibility. This provides a way of testing for effects of reversibility in a given numerical experiment. However, actual "random-number generators" are not strictly random; it is possible to construct one that runs reversibly in the sense that it can recover any previous random output. With a "reversible pseudorandom number generator" it would be possible to design a reversible calculation with "random-number" rounding. This violates the notion that the "random" element is outside the experimenter's control.

d) Liouville Theorem

A second advantage of the difference-scheme is that occupancy of phase locations is conserved in the motion. Again, this results from the fact that integers are used as allowable

position, velocity, and force values. A further feature assuring this conservation is that the forces are computed at each of the allowed configuration space locations. All particles that occupy a given configuration-space location experience the same force. With the exactness that integer arithmetic provides, this assures that all particles in a given phase location go to the same new location. Similarly, there is no way in which particles from two different phase space locations can wind up together. This is true even if there are errors in computing the forces, because a table of rounded integer forces is constructed and used. This method of handling the forces is the feature of the calculation that permits large numbers of particles to be handled.

For calculations that are formulated differently from this one, the degree of phase conservation in the Liouville theorem can be discussed along lines of the reversibility argument of the previous section. The essential parameter is the chance that two particles, initially in the same phase space location, will wind up at different locations. With our difference-scheme, this is the chance that the two particles will be found to have different forces acting on them. Our calculation is set up so that this cannot happen

because forces are rounded before being entered in the table; even rounding-by adding a random number before entering in the table would not destroy the Liouville theorem property of our calculation.

Just as rounding with random numbers provides an experimental way for testing reversibility effects, a modification of the present calculation can be devised that permits testing for effects resulting from violations of the Liouville theorem. This would simply involve rounding the force-values separately each time the force associated with a certain configuration space point is fetched. While the test for reversibility with random rounding can be carried out without affecting the behavior of the system with respect to the Liouville theorem, the reverse is not true. If the experiment is set up to test the effects of violating the Liouville theorem, reversibility will also be violated unless a "reversible pseudorandom number generator" is used.

For experiments on spiral structure, we have intentionally violated the Liouville theorem by rearranging the velocity distributions for the "gas" at a given configuration space location. Because the process was dissipative as well as being in violation of the Liouville theorem, it was not an adequate test of the effects of violating the Liouville theorem alone.

e) Integration Accuracy

The calculation approximates dynamics to the extent that our difference scheme leaves out terms $\mathcal{O}(\Delta t^3)$. In this sense it again differs from other, more sophisticated, schemes only quantitatively. Some calculations have been run in which the ignored terms get dangerously large. Because the coefficient of the Δt^3 term includes $(\vec{u} \cdot \nabla) \vec{p}$ (interpreted in terms of the physical problem), the situation is worst for high velocity particles in regions where the force gradients are large. We try to attach physical interpretations to experimental results only where the integration criteria indicate that the integration is proceeding safely. The integration improves as the problem is refined by increasing $\underline{m}$ and $\underline{q}$, if $\underline{q}^2/\underline{m}$ is held constant.

Optimization of the calculation requires choosing the physical parameters in such a way that the truncation errors (errors involved in the finite difference scheme as approximating a derivative) in the worst case of about one lattice spacing after one integration step. This is larger than is customarily allowed, because the lattice spacing is large. The higher order ignored terms must be investigated; they may even be larger. We have partially avoided the problem by using a near-cutoff in the force law $((\underline{a}^2 + \underline{x}^2 + \underline{y}^2)^{-1})$

which suppresses high derivatives of the forces. Even so, we have had to look into the problem.

f) Forces

The discussion so far might apply to any kind of nonrelaxing system. The relation to stellar dynamics--or to the gravitational n-body problem--arises through the forces used. These force-values are computed just as if they were not to be rounded to integers, and then they are, in fact, so rounded. It is the fact that the forces need not be determined to high precision and can thus be obtained from the Poisson equation at a fairly small number of locations that lets our model handle very large numbers of particles. The time required for the force calculation is independent of the number of particles. It is important, however, that the force calculation take proper account of large numbers of particles at substantial distances. This is done by calculating in floating point, and only rounding to integer values at the very last step. The force calculation is the only part of the "integration" that requires floating-point arithmetic.

The computational efficiency resulting from the use of discrete representations is obtained at the price of using a bounded space, which in turn means introducing models for the density distribution beyond the boundaries of the system. These models

always seem somewhat artificial. Isolated systems may be represented by using model density distributions that contain wide regions of zero density (Miller, 1968; Hohl and Hockney, 1969). Suppose that an infinite system is broken up into allowed and forbidden regions with boundaries such that, if a particle strays across the boundary into a forbidden region, that particle is deleted from the system. If there is no vector separation of points wholly within one allowed region that, suitably displaced, can connect points of two different allowed regions, then a force law can be constructed such that particles behave as if they interact only with other particles of the same allowed region. With squares or cubes of edge $\underline{L}$, this means an empty space of width $\underline{L}$ between image squares (or cubes). Larger spacings, even infinite ones, are permissible. The Fourier transform, with an empty space of width $\underline{L}$, provides the most efficient way to compute forces for isolated systems. The few particles that spill over the boundaries into the forbidden regions are easily handled.

We will not go into further details at this point. There are many subtleties and complications. The force calculation as used by Miller, Prendergast, and Quirk was described by Miller (1968); later developments have mostly been in the use of more

efficient computational methods for various parts of that calculation, and in the elimination of periodically replicated neighbors. While those parts of the report are superseded, many of the subtleties skipped over here are discussed there. Hockney and Hohl have done much more than we have to achieve computational efficiency. Their work is fully described by Hockney (1970), and by Hohl and Hockney (1969).

g) Significance

The particularly simple difference scheme used in the large n -body calculation makes it easy to investigate the numerical and computational properties of the system, even though restricting the "particles" to a discrete phase space as coarse as the one used here forces consideration of some aspects of the difference schemes that are usually ignored. The appearance of spatial lattices as well as time steps makes this look a little bit like the problem that arises in discussing partial differential equations. Indeed, we are solving a partial differential equation by integrating along its characteristics--the individual particle orbits.

The finite difference scheme is stable as an initial-value calculation. Modifications due to the discrete phase space and discrete force-values can be introduced into the inhomogeneous

term in the difference scheme and explicitly evaluated. The features of the system that give rise to reversibility and to the special form of the Liouville Theorem appropriate to this problem can be isolated and tested experimentally. We have not yet conducted such experiments and do not know how important these features are in the representation of real stellar systems, but consideration of how they might be tested experimentally leads to the impression that they are probably not very important in the sense that they would not lead to experimentally detectable effects.

The idea underlying this calculation--that of modifying the data to fit the calculation, rather than the other way around--may have wide applicability in physical calculations. The description has been given in detail to show that the approximations can be fairly thoroughly analyzed.

The notions of discrete spaces can be directly applied to several kinds of calculations. Prendergast has devised a similar calculation for hydrodynamics--the system is made to relax to an assumed form for a distribution function after each integration step so that what were "particles" in the n -body calculation become "streams." Prendergast reports very effective handling of shock-like discontinuities. Quirk has recently been modifying the programs for handling the "gas" in our calculations to use these ideas.

IV. A THOUGHT-EXPERIMENT

Static self-consistent models should be straightforward to construct in the discrete phase space if they are undertaken directly rather than by analogy to continuous models as described in section 2. These discrete static self-consistent models are instructive, although they may not be very interesting to construct in the computer.

In the discrete phase space, a point moves over an integer lattice according to the rules of the "game" described by Miller and Prendergast (1968). If the model were truly static and self-consistent, the force at any allowed configuration space point would be the same at one integration step as at any other.

A realistic "galaxy" model should contain one isolated region of nonzero forces (periodicity is unimportant for this discussion) in an infinite expanse where the forces are zero. Orbits that are restricted to the region in configuration space where the forces are nonzero occupy a bounded portion of the phase space--there is a maximum velocity that a particle may have and still be in a bound orbit. Any allowed phase space point is either on a periodic orbit or on an escape orbit (it is on a unique orbit, and the number of points in the region containing bounded orbits is finite). This is the discrete analogue of Hopf's first theorem (See, e.g., Contopoulos, 1966). The static

self-consistent model can be constructed of these periodic orbits.

Consider a periodic orbit containing one particle at each phase space point along the orbit. At any integration step, the particle occupying any one of these points will move to another point along the orbit, but the original point will be filled by a particle that has moved up from the previous point. The integration process looks like a huge game of "musical chairs," with just as many "players" as "chairs." After the integration step, the entire system looks exactly as it did before the step.

The problem of constructing static self-consistent models consists in: (1) Design a "reasonable" force-field. (2) Find all the periodic orbits in that force field. (3) Find a subset of the periodic orbits that, when projected to configuration space, will produce the original force field. If there is no such subset, pick a new force field and start over. The whole process sounds trivial-- but it is not at all trivial except in some particularly simple cases.

Such a system, once constructed would not be very interesting. It would sit forever and never change. As an example, suppose there were a ring of particles--the discrete analogue of a set of particles at the vertices of a regular polygon, rotating in the plane.

In the discrete phase space, this could be mimicked by two-dimensional harmonic oscillator orbits, chosen to be periodic in one rotation around the origin (same periodicity in x and in y ; two-dimensional harmonic oscillator orbits tend to make Lissajous figures with periods not 1:1 in the discrete phase space). The forces produced by the ring of particles, in the neighborhood of the ring itself, will look like harmonic oscillator forces. Adjust the force constant to balance things.

This ring would, if not disturbed, circulate forever. Yet it represents a system that is known to be unstable in continuous space. This system is stable although "cold" in the sense of having no random velocities. It is not a very interesting galaxy model. The interest attaches to the response of the system to a perturbation-- but this system cannot be perturbed by arbitrarily small amounts. The least perturbation that can be given to it is to remove one of the particles to another orbit. If this were done, the ring would probably disintegrate to a chaotic motion very quickly.

The same thought-experiment can be carried out with the kind of model attempted earlier. Suppose we had succeeded in constructing a static self-consistent model. It would circulate forever, each particle on its own periodic orbit. But then remove one particle somewhere, either leaving a hole or replacing the particle elsewhere in some

other orbit--preferably periodic so that it will stay around for a while. The two anomalies will move around the phase space until they happen to reach locations at which they conspire to make the force value round differently than it would in the absence of the perturbation. More particles will be disturbed by this change. Thereafter, the system is no longer static and self-consistent; it will take off and can change drastically. The argument is exactly like the reversibility argument of Sec. 3c.

Without actually doing the experiment, we cannot say whether the system that would eventually result would be "cool" or "hot," or whether it would be approximately axisymmetric. Some points are clear, however. First, it is simple to estimate how long, on the average, a particle or a hole would have to move around the system before it would, in a probabilistic sense, cause a computed force value to round differently. For the model tested, this is 10-15 integration steps. It is this long because the force near a single particle is much smaller in magnitude than the interval between successive allowed values of the force, and because the force field around a particle effectively extends only over a very small spatial region. The system might not ever notice that it had been perturbed. This, of course, is just what is meant, operationally,

by an "infinitesimal" perturbation. But the system response is nonlinear, and it is the nonlinear response that causes the trouble. A second point is that the model described in section 2 is hard to make self-consistent because of a design error--it had particles in regions of zero force. These must be on escape orbits. This is not serious with periodically replicated systems.

This discussion makes it clear that there is a smallest perturbation that can be given to the system and a smallest perturbation to which the system can respond. Because of the discrete number representation in computers, any model in a computer shares these attributes. But with the very fine discrete phase space allowed with full-precision computer numbers, the number of particles in the self-consistent model must be very large. It is straightforward to estimate how the number of particles scales with the number of points allowed in each coordinate direction and in each velocity component as the spatial lattice and time-steps are separately refined. The present calculation is just crude enough to build somewhat realistic static self-consistent models out of the number of particles that it can handle. But we have not yet developed an algorithm that will generate such models at will. Initial steps in this direction have produced some very interesting results, showing that a phase space point may be on a

periodic orbit while its neighbors are on escape orbits--some regions have small fractions of periodic orbits (Figure 9).

There are no nice continuity properties to the phase space, of course; there is not a closed surface inside which all orbits are periodic and outside which all orbits are of the escape variety.

Some orbits require several thousand steps to repeat periodically; some have extraordinary shapes (Figures 10 and 11). The book-keeping required to keep track of even these few periodic orbits seems prohibitive.

The notion of such self-consistent models can be extended to models that move in regular ways. A pattern that translates itself across the plane without change of shape is one such possibility --patterns that appear to rotate seem more difficult.

A system containing many "holes" should be able to respond to smaller perturbations. This might provide a useful route to attack the self-consistency problem posed in this report.

A truly static self-consistent model seems difficult to build. An attempt to do so is very instructive, and provides considerable insight into the way the computer models behave.

V. CONCLUSIONS AND ACKNOWLEDGEMENTS

The search for reasons why the computer models might be so hot has led through some interesting investigations. At the end of all these investigations, we still don't know whether it is possible to build a static self-consistent model that is cool. However, along the lines of the "thought-experiment" of Section 4, it seems quite likely that one can be constructed. Whether it would be stable (whatever that means) to small disturbances is unclear, but the suspicion is strong that, if the system felt the disturbance at all, it might well be quite unstable. That means that its response to as small a disturbance as it could feel would be a major rearrangement that might result in its being as hot as all the other models.

A related question is why spiral patterns do not last any longer than they do. Hohl's last quite long--for reasons that are as mysterious as why ours don't last very long. The obvious reasons might be (1) that there aren't enough particles to build a long-lived spiral, or (2) that the spirals dissipate because of numerical effects. After all this, there is no clear-cut basis to choose between the two alternatives, but there is no a priori reason why numerical effects should work against the maintenance of spirals instead of helping to maintain them.

It is a pleasure to acknowledge the assistance of many people who have participated in this program. The bulk of the calculations were carried out at the Goddard Institute for Space Studies in New York through the courtesy of Dr. Robert Jastrow, Director. My colleagues on this project, Kevin Prendergast and William J. Quirk of Columbia University have been full partners in the work at all stages. Many thanks are due to Frank Hohl for generously taking time to give me rather complete descriptions of what he and Roger Hockney have been doing--I hope that the descriptions of his work in this report do him justice. Hohl has essentially gotten as far in this work as we have, a fact that is obscured by the heavy emphasis on our work in this report.

FIGURE CAPTIONS

- Figure 1. Particles moving through spiral feature, from Miller, Prendergast, and Quirk (1970). Complete pattern at beginning is shown. Time between frames is about $1/150$ pattern rotation time.
- Figure 2. Continuation of sequence of Figure 1.
- Figure 3. Galactic parameters from model. Density of "Gas" and of "Stars." From Quirk (1970).
- Figure 4. Galactic parameters from model. Epicyclic frequencies and the angular velocity of gas. The pattern angular velocity is shown. From Quirk (1970). It is conceivable that the feature setting barred spirals apart from normal spirals might be the absence of an inner Lindblad resonance, as in this figure.
- Figure 5. Galactic parameters from model. Local circular velocity and local standards of rest for "gas" and for "stars." From Quirk (1970). The circumferential mean velocity of the "stars" is only half the local circular velocity, implying predominance of pressure-support.

Figure 6. Galactic parameters from a different model. "Star" velocities compared to local circular velocity from a pure "star" case run some time ago.

Figure 7. Q , the ratio of the actual velocity dispersion in the model to that required for stability by the Toomre (1964) criterion. Hohl's data, courtesy of Frank Hohl.

Figure 8. Velocity dispersion in the radial direction required to stabilize configuration against axisymmetric modes. The lower curve is for the modified force law, according to Eq. (5); the upper curve is for the same configuration with $1/r^2$ forces. The upper curve would reach 2.91 at the center.

Figure 9. Search for periodic orbits in a static self-consistent model. All periodic orbits having a point in the plane $y=0, u=0$. The number of steps in the periodic orbit is shown. Blanks are escape orbits. Some orbits have more than one point in this plane.

Figure 10. Search for periodic orbits in a static self-consistent model. A well-behaved periodic orbit. Track of orbit over the configuration space.

Figure 11. Search for periodic orbits in a static self-consistent model. An extraordinary periodic orbit. This orbit passes through the origin, and changes the sign of its angular momentum. Plot of the track of the orbit over the configuration space.

REFERENCES

1. Aarseth, S.J., 1963, Mon. Not. Roy. Astron. Soc. 126, 223.
2. Aarseth, S.J., 1966, Mon. Not. Roy. Astron. Soc. 132, 35.
3. Buneman, O., 1967, J. Comp. Phys. 1, 517.
4. Chandrasekhar, S., 1960, Principles of Stellar Dynamics (New York: Dover Publications, Inc.).
5. Code, A.D., 1967, Astron. J. 72, 789-90 (A).
6. Contopoulos, G., 1966, "Problems of Stellar Dynamics," pp. 169-257 in Relativity Theory and Astrophysics, J. Ehlers, Ed., Vol. 1 (Providence, R.I.: American Mathematical Society).
7. Greenspan, D., 1970, to appear in Studies in Applied Mathematics, published by the Mathematical Association of America.
8. Hockney, R.W., 1970, Methods in Computational Physics, Vol. 9, pp. 136-212, B. Adler, S. Fernbach, and M. Rotenberg, Ed's. (New York: Academic Press).
9. Hockney, R.W., and Hohl, F., 1969, A. J. 74, 1102.
10. Hohl, F., and Hockney, R.W., 1969, J. Comp. Phys. 4, 306.
11. Lee, E.P., 1967, Ap. J. 148, 185.
12. Lin, C.C., Yuan, C., and Shu, F.H., 1969, Ap. J. 155, 721.
13. Lynden-Bell, D., 1962, M.N.R.A.S. 144, 279.
14. Miller, R.H., 1964, Ap. J. 140, 250.
15. Miller, R.H., 1968, "Forces for the N-Body Calculation," ICR Quarterly Report No. 17 (May 1, 1968), Sec. III-A, Institute for Computer Research, University of Chicago (Unpublished).

16. Miller, R.H. and Alton, N. 1968, "Three Dimensional N-Body Calculation," ICR Quarterly Report, No. 18 (August 1, 1968), Institute for Computer Research, University of Chicago (Unpublished).
17. Miller, R.H., and Prendergast, K.H., 1968, Ap. J. 151, 699-709.
18. Miller, R.H., Prendergast, K.H. and Quirk, W.J., 1970, Ap. J. (to be published).
19. Ng, E., 1967, Ap. J. 150, 787.
20. Quirk, William J., 1970, Ph.D. Thesis, Columbia University, (Unpublished).
21. Schmidt, M., 1965, Galactic Structure, pp. 513-530. A. Blaauw and M. Schmidt, Eds. (Chicago: University of Chicago Press) Vol. V, of Stars and Stellar Systems, B. M. Middlehurst and G. Kuiper, Eds.
22. Standish, E.M., 1968, Thesis, Yale University (Unpublished).
23. Toomre, A., 1963, Ap. J. 138, 385.
24. Toomre, A., 1964, Ap. J., 139, 1217.
25. Wielen, R., 1967, Ver. Astron. Rechen-Institut Heidelberg, Nr. 19.
26. Wielen, R., 1968, Bull. Astronomique (3), 3, 127.
27. Wendorff, B., 1966, Theoretical Numerical Analysis (Academic Press, New York).

Fig.1

Fig.2

Fig. 3

Fig. 4

Fig.5

Fig.6

Fig.7

Fig.8

[illegible]

Fig. 9

Fig.10

Fig.11