

SOLUTION OF ADVECTION-DOMINATED TRANSPORT
BY EULERIAN-LAGRANGIAN METHODS
USING THE BACKWARDS METHOD OF CHARACTERISTICS

by

António Eugénio de Melo Baptista

Eng. Civil, Academia Militar (Portugal)
(1978)

S.M., Massachusetts Institute of Technology
(1984)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS OF THE
DEGREE OF

DOCTOR OF PHILOSOPHY
IN CIVIL ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

January 1987

© Massachusetts Institute of Technology

Signature of Author _____

Department of Civil Engineering
January 6, 1987

Certified by _____

Professor Keith D. Stolzenbach
Thesis Supervisor

Certified by _____

Dr. E. Eric Adams
Thesis Supervisor

Accepted by _____

Professor Ole S. Madsen
Chairman, Department Committee

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

MAR 23 1987

LIBRARIES

ARCHIVES

SOLUTION OF ADVECTION-DOMINATED TRANSPORT BY
EULERIAN-LAGRANGIAN METHODS USING THE BACKWARDS METHOD OF CHARACTERISTICS

by
António Eugénio de Melo Baptista

Submitted to the Department of Civil Engineering
on January 6, 1987, in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Civil Engineering

ABSTRACT

We provide a systematic analysis of the consistency, stability, convergence and accuracy of the numerical solution of the transport equation by a general Eulerian-Lagrangian Method (ELM). The method involves three basic steps: the backwards tracking of characteristic lines following the flow, the interpolation of concentrations at the feet of these lines, and the solution of dispersion taking such concentrations as initial conditions. The first two steps constitute the Backwards Method of Characteristics (BMC); the third step involves a time-discretization along the characteristic lines, and a spatial discretization of the dispersion operator, both based on conventional techniques (e.g., Euler or Crank-Nicholson for time; finite-elements or finite-differences for space).

The choice of the spatial interpolator is shown to impact the consistency, stability and convergence, as well as the accuracy of the BMC. Most interpolators ensure consistency, but only a few ensure stability, hence convergence; stability criteria are derived from a newly developed generalized Fourier analysis, which can account for non-linearities introduced by quadratic grids. The comparison of formally derived propagation and truncation errors, complemented by numerical experimentation, provides a reference for the choice of the interpolator, given a specific transport problem characterized by prevailing concentration gradients.

The BMC potentiates the use of large time-steps, well above Courant number of order one. In the limiting case of pure advection, optimal accuracy would be obtained for a Δt close to the total time of interest; the presence of dispersion constrains, however, the size of Δt , especially in the case of non-uniform flows. The comparison of the truncation errors for the three basic steps of ELM provides a reference to select Δt , as a function of Δx , of the spatial interpolators and time-discretization schemes, and of the gradients of flow and concentrations.

Thesis Supervisor: Dr. Keith D. Stolzenbach
Title: Associate Professor of Civil Engineering

Thesis Supervisor: Dr. E. Eric Adams
Title: Principal Research Engineer and Lecturer

To my son, Gustavo

CONTENTS

List of Figures	7
List of Tables	9
Acknowledgements	10
Synopsis	11
Motivation and Objectives	12
Research Outline	16
A User's Oriented General Discussion	19
Recommendations for Further Work	29
References	31
Figures	34
Appendixes	39
On the Mathematical Nature of the Transport Equation	40
Review of Numerical Methods for the Solution of the Transport Equation	45
On the Role of Advection and Dispersion	60
Fourier Analysis of the Backwards Method of Characteristics	65
Abstract	66
Introduction	66
Review of the Backwards Method of Characteristics	67
Formal Error Analysis	69
Reference Framework	69
Propagation Errors	75
Individual Components, First Time-Step	75
Individual Components, After N Time Steps	77
Complete Solution, After N Time Steps	79
Verification and Discussion of Error Formulae	79
Comparison with Numerical Experimentation	79
A Brief Look at the Mechanism of Energy Transfer	81
Conclusions	81
References	83
Figures	84
Table	94
Appendix - Derivation of Formulae for Propagation Errors After N Time Steps	95

The Consistency, Stability and Convergence of the Backwards Method of Characteristics	104
Abstracts	105
Introduction	105
Review of the Backwards Method of Characteristics	107
Analysis of Consistency	109
Analysis of Stability	114
Stability Criteria	114
The Stability of the Backwards Method of Characteristics for Specific Interpolators	117
Analysis of Convergence	118
Summary and Conclusions	119
References	122
Figures	123
Tables	133
Appendixes	137
Derivation of the General Form of the Truncation Error for the Backwards Method of Characteristics	137
Derivation of a Stability Criterion for the Case of 3-node Core Elements	141
The Choice of the Interpolator for the Backwards Method of Characteristics	143
Abstracts	144
Introduction	144
Interpolation Strategies: General Options and Implications	147
Global Interpolators	147
Local Interpolators	148
Detailed Comparison of Interpolators of Class C_0	155
Definitions	155
Analysis of Consistency	157
Analysis of Stability	157
Analysis of Accuracy	159
Numerical Damping and Numerical Dispersion	159
Response to Grid Refinement	163
Final Considerations	165
References	168
Figures	170
Tables	205

The Accuracy of Eulerian-Lagrangian Methods	218
Introduction	219
Brief Review of Concepts and Specific Implementations of Eulerian-Lagrangian Methods	221
Reference Algorithm	222
Truncation Errors	224
Accuracy Dependence on the Computational Strategy	227
General Aspects	227
Accuracy Dependence on Δx	228
Accuracy Dependence on Δt	229
Accuracy Dependence on the Interpolator and Time-Discretization	233
Final Considerations	234
References	236
Figures	239
Tables	256
Appendix - Derivation of Truncation Errors	258

LIST OF FIGURES

SYNOPSIS

1. Illustration of the Eulerian-Lagrangian solution of the transport equation
2. Illustration of different tracking approaches
 - (a) Standard
 - (b) Baptista et al. 1984
3. A conceptual approach to the choice of the computational strategy
4. Illustration of scaling, for the problem of a Gauss-hill in a shear flow
- A1. Diagrammatic representation of hyperbolic and parabolic partial differential equations

FOURIER ANALYSIS OF THE BACKWARDS METHOD OF CHARACTERISTICS

1. Illustrative sketch for the Backwards Method of Characteristics
2. Definition of core elements
3. Solution of the reference problem, for $t = T = 9600$ ($\Delta x = 200$; $N = 100$)
4. Amplification factors per time step, as a function of the number of time steps
5. Amplification factors after N time steps, as a function of the dimensionless wavelength

THE CONSISTENCY, STABILITY AND ACCURACY OF THE BACKWARDS METHOD OF CHARACTERISTICS

1. Illustrative sketch for the Backwards Method of Characteristics
2. Definition of the core elements
3. Illustration of the dependence of the convergence of the Backwards Method of characteristics on the choice of the interpolator
4. Amplification factors per time step, for selected interpolators
5. Illustration of the dependence of the accuracy of the Backwards Method of characteristics on the number of time steps required to reach a fixed total time

THE CHOICE OF THE INTERPOLATOR FOR THE BACKWARDS METHOD OF CHARACTERISTICS

1. Illustrative sketch for the Backwards Method of Characteristics
2. Definition of core elements
3. Amplification factors per time step, as a function of the location of the foot of the characteristic line within the core element, and of the dimensionless wavelength
4. The relative importance of amplitude and celerity errors, illustrated for a reference problem
5. Amplification factors, after N time steps, for constant Δt
6. Amplification factors, after N time steps, for Δt randomly chosen in each step
7. Comparison of the accuracy of cubic interpolators, in the context of a reference problem -- what about 5-point interpolators
8. Dependence of accuracy on grid refinement
9. Illustration of the effect of grid refinement on the dependence of accuracy on grid refinement

THE ACCURACY OF EULERIAN-LAGRANGIAN METHODS

1. Illustrative sketch for Eulerian-Lagrangian methods
2. Definition of the reference spatial interpolators
3. Illustration of the effect of poor tracking
4. Example of the choice of scales for a specific problem
5. Dependence of truncation errors on Δx , as a function of Δt
6. Dependence of truncation errors on M , as a function of Δt
7. Dependence of truncation errors on D , as a function of Δt
8. Dependence of truncation errors on Δu , as a function of Δt
9. Dependence of truncation errors on the choice of the spatial interpolator and on the time-discretization scheme
10. Illustration of the strategy for a third-order time-discretization scheme

LIST OF TABLES

FOURIER ANALYSIS OF THE BACKWARDS METHOD OF CHARACTERISTICS

1. Expressions for $r_m(\alpha)$ and $s_m(\alpha)$

THE CONSISTENCY, STABILITY AND CONVERGENCE OF THE BACKWARDS METHOD OF CHARACTERISTICS

1. Definition of selected interpolators
2. Summary of error-propagation formulae

THE CHOICE OF THE INTERPOLATOR FOR THE BACKWARDS METHOD OF CHARACTERISTICS

1. Definition of the alternative interpolators
2. Generation procedure for HL and SP interpolators
3. Summary of the stability criteria for the Backwards method of characteristics
4. Truncation errors for alternative interpolators (general case)
5. Truncation errors for alternative interpolators (case of $\alpha = u \cdot \Delta t / \Delta x$)

THE ACCURACY OF EULERIAN-LAGRANGIAN METHODS

1. Truncation errors
2. Relative importance of different types of errors
 - (a) Quadratic interpolator, Euler time-discretization
 - (b) Quadratic interpolator, Crank-Nicholson time-discretization
 - (c) Quartic interpolator, Euler time-discretization
 - (d) Quartic interpolator, Crank-Nicholson time-discretization

ACKNOWLEDGMENTS

For the past four and a half years, I have been combining my research and graduate studies at Massachusetts Institute of Technology (MIT) with my research work at Laboratorio Nacional de Engenharia Civil (LNEC). This has provided a stressing, but very rich and challenging experience, both personally and professionally.

Time will tell whether and to what extent this and my other thesis (S.M. at MIT, 1984; Research Officer at LNEC, 1986), reports and research papers, published in this period, are of significance to the general progress of numerical modeling, as a science and as an engineering tool. I trust that valid contributions have been given in an area that is broad and complex enough to accommodate years of further research by different teams, and is important enough to justify such research.

A variety of institutions and people have contributed to my work, in different forms, and in different periods. Among the institutions, I gratefully acknowledge:

- Laboratorio Nacional de Engenharia Civil
- Massachusetts Institute of Technology
- NATO Committee on the Challenges of Modern Society
- Commissao INVOTAN
- Luso-American Cultural Commission

Among the people, I would like to show my appreciation to my colleagues at LNEC, and to the faculty, staff and graduate students at the Ralph M. Parsons Laboratory, who provided assistance and motivation in many ways.

The advice and friendship of Eric Adams and Keith Stolzenbach were greatly appreciated; they have strongly influenced my current perception of research, teaching and learning. Paulo Rosman, a fellow graduate student, offered constructive discussion, skilled advice on how to make a personal computer do everything but talking, a joke when needed, and great friendship.

My mother, my wife and my son have, each on his own way, shown their constant love and support. This thesis is dedicated to them, and, in particular, to my son and his future.

The writing of this document benefited from the suggestions of Drs. Donald Harleman and Mike Celia, members of my thesis committee. In the material preparation of the document, I was assisted by my wife, who spent several days and nights drawing and mounting figures and captions, and by the skilled typing of Read Schusky.

**SOLUTION OF ADVECTION-DOMINATED TRANSPORT BY
EULERIAN-LAGRANGIAN METHODS USING THE BACKWARDS METHOD OF CHARACTERISTICS**

SYNOPSIS

1. MOTIVATION AND OBJECTIVES

The numerical solution of the transport equation, describing the fate of a passive scalar in a moving fluid, has been the object of intense research for the past few decades. So much interest concerning an apparently inoffensive (it is even linear...) equation may seem, at first sight, misplaced. However, there is a fundamental difficulty in the solution of the transport equation, which results from the fact that, while advection and dispersion are simultaneous processes, they promote mass transport very differently: in the case of advection, transport is along characteristic lines that follow the flow (in a way that depends only in the past), while for dispersion transport is between the characteristic lines (in a way that depends on both the past and the present). Mathematically (see Appendix A), this means the need to treat simultaneously hyperbolic terms (associated with advection) and parabolic terms (associated with dispersion), a problem that no numerical method has yet fully overcome. With the practical importance of simulating transport mechanisms in fluid environments, this difficulty clearly justifies past and on-going research.

A review of alternative methods for the solution of the transport equation (Appendix B) suggests that they fit into three major categories: Eulerian (EM), Lagrangian (LM) and Eulerian-Lagrangian (ELM). EM, which historically were the first to be introduced and are still very popular, have strong shortcomings in the analysis of transport problems where

advection has a significant role vis à vis dispersion (the case for most natural flows) and where sharp gradients in the flow direction can not be resolved with a reasonable grid size (often the case for pollutant transport near sources or fronts). In turn, LM, which perform extremely well for pure advective transport, run into practical difficulties whenever dispersion has also to be solved, and have hardly been used in the context of realistic problems. ELM combine the best aspects of EM and LM, having the potential to provide accurate solutions for the range of advection-dominated to dispersion-dominated transport problems.

Advection and dispersion are dependent concepts (see qualitative discussion in Appendix C), and distinguishing between the scales of flow variability that contribute to each process may involve some ambiguity (e.g., see [A2], [R1]). In this study, we assume that such distinction was made beforehand, and that the flow characteristics and the dispersion coefficients are available to us.

As discussed in Appendix B, a few research groups have, in the past few years, been very active in the study of ELM; these include Holly and co-workers [H2-H4, K1-K2], Benque, Hauguel and co-workers [B10], Neuman and co-workers [N1-N3], and Baptista and co-workers [A1, B1-B5, K3]; other relevant works in the subject are [B11], [C2], [G1], [H1], [L1] and [V1]. While there are significant differences among the approaches proposed by each group, they all share the basic idea of solving the transport equation in the nodes of a fixed grid, but integrating in such solution past information that is carried along characteristic lines that follow the flow.

The conceptual procedure is illustrated in Figure 1, for 1-D (actual implementations of the procedure have been used also in 2-D, and could be extended to 3-D). The concentrations at the nodes of the computational grid are found, at time n , through a three-step procedure:

- Definition of characteristic lines that start at each grid node, at time n , and follow the flow backwards until time $n-1$ or a boundary is reached.

- Calculation, by interpolation from known nodal concentration values at time $n-1$ (or at boundaries), of the concentration at the feet of the characteristic lines; these concentrations correspond to the concentrations at time n , if advection was the only transport mechanism. More importantly, they are also the correct initial conditions for the transport problem, written between times $n-1$ and n in Lagrangian coordinates.

- Solution of the transport equation in the coordinate system defined by all characteristic lines, taking as initial conditions the concentrations at the feet of these lines.

To implement this procedure, ELM typically split the transport equation, either in its differential or in its time-discretized form, into two sub-equations (advection and dispersion), solving the former through the Backwards Method of Characteristics (BMC), using a choice of interpolators, and the latter by a scheme that combines a finite-difference or finite-element discretization in space with an Euler implicit or a Crank-Nicholson (defined along the characteristic lines) discretization in time. Virtually all methods have been used in connection with some local forward

tracking procedure, to handle gradients that the basic ELM can not handle. Physical, chemical, or biological transport processes (source/sink terms), if present, can be treated within the dispersion step, or as a separate, fourth step.

Results reported by all groups have been extremely promising, suggesting that, for comparable costs, accuracy of ELM should never be worse than for FM, and is significantly better whenever advection is dominant over dispersion and sharp concentration gradients exist in the flow direction.

However, most reported results are based on numerical experimentation, taking as a reference either (a) controlled but relatively simple test problems (transport of instantaneous sources, continuous sources or fronts in 1-D or 2-D uniform or "regularly" non-uniform flows (e.g., instantaneous sources in rigid-body rotation)), or (b) realistic but mostly uncontrolled (accuracy wise) applications (e.g., pollutant transport in coastal waters or in groundwater). This has left significant gaps in the theoretical foundation of ELM, which raise pertinent questions concerning both the relative merits of different ELM, and how accurate and cost-effective any ELM really is in actual applications. Unclear aspects include:

- The consistency, stability and convergence of ELM and, in particular, of the BMC.

- The dependence of ELM accuracy on controlling parameters, related both to the discretization of the domain and to the nature of the transport problem.

- The dependence of ELM accuracy and efficiency on the choice of specific techniques for the tracking of the characteristic lines, for the spatial interpolation of the feet of these lines, and for the time-discretization of the governing equation.

The general objective of our research has been to extend the understanding of these different issues, and use this understanding to improve current modeling ability of the transport equation by ELM. A significant emphasis has been given to the solution of advection, which is commonly accepted as a critical step of the overall procedure. The 1-D transport equation is usually taken as a reference, for formal developments.

2. RESEARCH OUTLINE

The thesis consists of this Synopsis and four self-contained papers, [B6-B9]. In the Synopsis, we motivate the study and introduce its objectives (Section 1), highlight specific contributions from each paper (this section), present a general, user's oriented, discussion of ELM modeling of the transport equation, as we now perceive it (Section 3), and recommend further research (Section 4); background on the mathematical nature of the transport equation, on available numerical methods for the solution of the transport equation, and on the role of advection and dispersion, is presented in Appendixes A, B and C, respectively. Each of the four papers discusses in detail one or a set of closely related specific aspects of the research, as follows:

Fourier Analysis of the Backwards Method of Characteristics [B6]

This is a methods paper, describing the derivation of a novel technique (a generalized Fourier analysis) to predict the propagation in time of amplitude and phase errors. A direct consequence of the derivation is to unveil an internal source of non-linearities, that characterizes the BMC for quadratic and higher-order interpolating core elements, and should also affect conventional finite-element techniques (using higher-than-linear elements).

The technique described in this paper is extensively used by [B7-B8], as a tool for the formal analysis of stability and accuracy of the BMC.

The consistency, stability and convergence of the Backwards Method of Characteristics [B7]

This paper provides a systematic assessment (based on Taylor series and generalized Fourier series analysis of the BMC algorithm, and on Lax equivalence theorem) of the consistency, stability and convergence of the BMC, and their dependence on the selected spatial interpolator for concentrations. It is shown that the BMC is consistent for all interpolators that fit nodal concentrations exactly, and will not, in general, be consistent for those which do not fit these concentrations. Stability, hence convergence, while independent of the Courant number, do depend on the interpolator, and have to be assessed on an individual basis. Quantitative criteria for this assessment are derived from the generalized Fourier analysis of [B6], and used to show that all interpolators based on Lagrange polynomials alone are stable and convergent, but some recently

proposed hybrid interpolators, based on Hermite polynomials with estimated derivatives, are not.

The choice of the interpolator for the Backwards Method of Characteristics

[B8]

The choice of the interpolator to find concentrations at the feet of the characteristic lines has been recognized as crucial for the accuracy of the BMC, and [B7] shows that it also influences more basic properties such as consistency, stability and convergence. This paper provides a systematic comparison of the most promising (or historically important) interpolators, some of which were proposed by the author. The comparison emphasizes the different trade-offs involved in the choice of an interpolator, and provides reference information (amplitude and phase errors, and truncation errors) to allow the reader to make his own choice. While it is emphasized that no optimal choice exists, we suggest that the flexible use of Lagrange polynomials of different orders (defined over quadratic core elements) strikes an attractive balance among cost, convenience, and accuracy. The order of the chosen polynomial should be based on the steepness of concentration gradients.

The accuracy of Eulerian-Lagrangian methods [B9]

This is the only paper that concerns the solution of the full transport equation. It proposes the use of ELM that are flexible in the choice of the spatial interpolator for advection and in the time-discretization scheme for dispersion, and provides the systematic analysis of truncation errors for selected alternative choices of these. This

analysis is used to identify and show the influence of the parameters that control accuracy, which are associated to both the physical problem (concentration gradients, mean velocity, velocity gradients) and to the discretization in space and time of the computational domain (Δx , Δt). Particularly relevant is the fact that, unlike in EM, optimal accuracy, for fixed Δx , does not necessarily correspond to $\Delta t \rightarrow 0$. Criteria to estimate the Δt leading to optimal accuracy are derived, and discussed in the context of the definition of the best computational strategy for ELM, which involves not only the choice of Δx and Δt , but also that of the interpolator and of the time-discretization scheme.

3. A USER'S ORIENTED GENERAL DISCUSSION

In the previous section we identified different individual contributions of this research. We now discuss how these contributions can, as a whole, influence the attitude of modelers towards the solution of actual transport problems, and increase their actual modeling ability.

First, however, we should stress that, in spite of the extensive research that has been done for some decades, the state of the art in the modeling of transport mechanisms is not advanced enough to allow the establishment of a comprehensive user's guide, in the form of unambiguous rules; the success of each individual application depends on the ability of the user to be aware of several available alternatives (sometimes poorly understood in terms of their potentials and limitations), and to critically choose the one(s) that best fit his needs and resources.

We would like also to emphasize that the first and a major concern of

the user should be to characterize the physical problem, through the definition of appropriate scales: in particular, what are the smallest and largest scales of space and time that are of interest, and, within these scales, how large are the concentration gradients, the velocity, and the velocity gradients. The estimation of these scales is a matter of common sense, given a correct perspective of the objectives of the analysis and appropriate information on the environment where transport occurs.

However, formulating the mathematical problem in a way that is unambiguously consistent with these scales is a much more difficult task, as discussed by [A2] and [R1] in the context of so-called filtering techniques.

Our research assumes that the physical problem was correctly stated beforehand, in such a way that not only the proper form of the governing equation is known, but we also know how to quantitatively distinguish between advection and dispersion, and what are the minimum wavelengths and period (L_m and T_m , say) that we want to capture.

From the theory of digital signal processing, we know that the coarsest grid that can resolve all the relevant scales of the transport problem is characterized by $\Delta x = L_m/2$ and $\Delta t = T_m/2$. However, no available numerical method can correctly propagate the L_m wavelength in such a grid; indeed, depending on the numerical technique being used, on the actual Δt to be chosen, and on the time for which L_m will persist as a relevant wavelength (which depends on the importance of dispersion), Δx will often have to be chosen one or more orders of magnitude smaller than L_m , to ensure proper propagation.

If we were not constrained by available resources (in particular, computational time and memory, and round-off errors), any convergent numerical method would be appropriate for the solution of the problem; we would only need to take Δx and Δt "small enough". Because available resources are limited, though, we often want to rationalize our choice of the numerical method and of the discretization in space and time, so as to be as "accurate" as possible, but at the lowest (or, more simply, at a feasible) cost.

For problems where dispersion is dominant over advection, using an EM rather than an ELM, or vice-versa, will hardly affect cost or accuracy, if, within each type of method, equivalent choices are made concerning the representation of space and time. The same is not true, however, when advection is dominant over diffusion. Indeed, in this case, ELM present significant potential advantages by their flexibility in the choice of the space representation (in the advection step) and by their ability to accurately handle very large Courant numbers (i.e., for a fixed grid, to use very large time steps). Whether these potential advantages will be of actual practical interest, i.e., whether better accuracy at a same cost, or the same accuracy at a lower cost, can be achieved, depends on the variability of the concentration and flow fields (ELM will be comparatively more and more efficient as spatial concentration gradients in the flow direction increase), and on the ability of the modeler to adopt the proper computational strategy.

Our work ([B8] and [B9] in particular) indicates that none of the available ELM is "optimal", and suggests that there are advantages in

perceiving them merely as specific implementations of a unique general tool, whose flexible use should be encouraged. While apparently trivial, this notion is of paramount importance for the cost-effective use of ELM, and it raises the question of computational strategy, in the sense of combining the choice of Δx and Δt with the selection of:

- the technique for the tracking of the characteristic lines
- the interpolator to find concentrations at the feet of the characteristic lines
- the time-discretization algorithm
- the spatial discretization of the dispersion operator.

Tracking technique

The choice of the tracking technique plays a key role in the accuracy of ELM, as tracking defines the coordinates along which the problem will be solved. This fact has not, however, been widely recognized, most probably due to the simplicity of the flow fields used in common test problems. As a consequence, most tracking algorithms are rather simplified (e.g., see Figure 2a), and, for non-uniform or unsteady flows, will impose the need to use small time steps, so as to avoid significant errors; this means that the ability of ELM to handle $Cu \gg 1$ is effectively amputated, greatly reducing the attractiveness of these methods.

Recognizing this, and taking advantage of the fact that the tracking is independent of the concentration field, [B2] developed (in 2-D, and allowing for irregular finite element grids) a tracking algorithm whose accuracy can be pre-imposed, in terms of a maximum "closing error"--see

Figure 2b. The algorithm solves the ordinary differential equation that defines each characteristic line by a 4th-order Runge-Kutta scheme, whose time step, $\delta t < \Delta t$, is made as small as necessary to satisfy the imposed accuracy criterion. The algorithm has been shown ([B2], [K3]) to be expensive but feasible, in actual detailed computations of pollutant transport in coastal waters, carried on in small machines (VAX 10-780, μ VAX-II).

While the efficiency of the algorithm proposed by [B2] can be improved, we suggest that its conceptual basis suits ELM extremely well, and that similarly accurate tracking procedures should be used whenever the flow is non-uniform or unsteady, as the only way to actually take advantage of ELM's inherent ability to handle large time steps.

Interpolator

The next option concerns the spatial interpolator to find the concentrations at the feet of the characteristic lines. This option has been widely recognized as crucial for the accuracy of ELM, and, as such, a variety of interpolators have been proposed, as discussed in detail in [B8]. While we recognize that valid alternatives exist, we recommend the combined use of quadratic and/or quartic Lagrange polynomials, defined over quadratic core elements.

Time-discretization algorithm

Because the BMC has the potential to use large time steps, the ELM time discretization scheme should be chosen so as to preserve this

ability, which may be particularly difficult in the case of non-uniform flows ([B9]). The use of a simple Euler implicit discretization, as in [B2], is strongly discouraged. An appropriate reference scheme is the Crank-Nicholson (formulated along the characteristic lines), but a three time-levels scheme (presumably a Richardson extrapolation, formulated along characteristic lines) may also be considered as a way to further increase the allowed Δt --[B9].

Space discretization of the dispersion operator

The space discretization of the dispersion operator is usually not critical for accuracy, but does influence the choice of the interpolator for the advection step, as it determines which types of core element will be convenient to use. We suggest that, as a reference, the use of a discretization scheme (finite element Galerkin, or centered finite differences) that allows a convenient implementation of interpolators with quadratic core elements.

Computational strategy

Given the above general options (which already imply some preliminary selection, not exempt of subjectiveness), how should one define an appropriate computational strategy, for a specific application? While this question have no unique answer, and identifying an optimal strategy may yet be far from the reach of current knowledge, the present research may provide useful insight and some actual guidance in this matter. In particular, we suggest that the conceptual decision-making procedure illustrated in Figure 3 be followed as closely as practically feasible.

This procedure is to be conducted prior to the actual numerical computation; it relies strongly on qualitative and quantitative information generated through this research ([B5-B9]). Formal extensions or judicious extrapolations may, however, become necessary depending on the specific application. We note that the procedure implicitly assumes that:

(a) If some concentration and velocity gradients in space and/or time are not going to be well propagated by the numerical solution, it makes no sense to pretend, at the level of the formulation of the governing equation and establishment of initial and boundary conditions, that this is not so, by keeping the associated space and time scales. We suggest that such scales be explicitly filtered out, to avoid uncontrolled numerical nonlinearities; as shown by [A2] and, especially, [R1], this should lead to a reformulation of the governing equation, but not to a significant change of its mathematical nature.

(b) For an imposed level of resolution of the physical problem (or, similarly, of accuracy of the numerical procedure), the most significant trade-offs in the implementation of ELM are (1) the choice of Δx versus the spatial interpolator for the advection solution, and (2) the choice of Δt versus the time-discretization scheme. While taking larger Δx and higher order interpolators, and larger Δt and higher-order time-discretization schemes, will often pay off in the presence of sharp gradients and advection-dominated problems, the reverse may often be true for strongly diffusion-dominated problems.

Some of the steps of this conceptual procedure are now discussed:

Characterization of the physical problem

The physical problem should, for the sake of the choice of a computational strategy, be characterized in terms of the following main scales:

ℓ_c - a length scale characterizing concentration variability

U - a velocity scale for the mean (in space) flow

ΔU - a velocity scale for flow non-uniformities

ℓ_u - a length scale characterizing flow non-uniformities

D - a scale for the dispersion coefficient

In addition, and assuming that a filtering technique (e.g., [A2] and [R1]) was used to establish the governing equation, we will need to know which were the minimum wavelength, L_m , and period, T_m , that were elected to be explicitly represented in the Fourier description of concentrations and flow.

Figure 4 illustrates the definition of these scales (except L_m and T_m), in the case of the relatively simple problem of the transport of an instantaneous source (a Gauss-hill) in a constant-shear flow. In more complex problems, it may be useful to divide the computational domain into different zones, each characterized by different scales.

Selection of the spatial interpolator and of the time-discretization scheme

We suggest that a quadratic Lagrange interpolator, and a Crank-Nicholson discretization scheme be taken as reference techniques. If necessary or appropriate, each or both these options can be changed, and we

do recommend that this be at least considered unless the cost of the reference technique is insignificant. It should be stressed that changing the interpolator does not necessarily imply changing the time-discretization (or vice-versa), and, indeed, changing each of these at a time is probably the best way to identify the most appropriate computational strategy.

It should also be noticed that, if considerably different concentration gradients are present in different zones of the computational domain, it may be advantageous to use both quadratic and quartic interpolators, the former in the zones of milder gradients, and the latter in the zones of larger gradients (e.g., near sources). In the same perspective, if the time variability of concentration gradients and velocities and velocity gradients is significant, it may be advantageous to change the time-discretization scheme (or the Δt) periodically.

Selection of Δx and Δt

Once both the physical problem and the solution technique are characterized, Δx (loosely taken here as a characteristic nodal spacing, in some relevant space direction(s)) and Δt can be rationally estimated by a proper combination of the information provided in [B6-B9] (or its formal or common-sense extension to multi-dimensions). The approach that we propose is to

(1) Assume Δt as large as possible (consistent with physics), and estimate a "representative" number of time steps, $N = T/\Delta t$, where T is the expected duration for which waves of length L_m remain a significant

component of the whole solution.

(2) Use N and L_m to define Δx , so as to meet some pre-set criterion for the amplitude and phase errors in the propagation of the wave with length L_m (based on plots of amplitude and phase errors, such as those presented in [B8], from the Fourier analysis procedure derived by [B6]). We suggest that the error criterion be quite strict, calling for errors in the order of a few percent, at most. If L_m can not be correctly propagated, this should influence the characterization of the physical problem, rather than being left as an uncontrolled numerical error.

(3) Re-evaluate Δt , by minimizing the total truncation error (i.e., by matching the order of magnitude of the truncation errors due to the interpolation of concentrations at the feet of the characteristic lines and due to the time-discretization). [B9] provide theoretical support for the estimation of the truncation errors, based on Taylor series analysis; the procedure requires the knowledge of e_c , U , ΔU , e_u , and D , as well as Δx .

(4) Re-iterate as necessary (changing Δt will change N , which was used in (2) to estimate Δx).

Irregular grids and/or the presence of different scales of interest, regionally distributed, are recognized difficulties in the implementation of the above procedure. It is tentatively suggested that different zones be considered within the domain, in each of which a different choice of Δx may result, and that Δt be chosen so as to minimize the truncation error in the critical zone (or, if all truncation errors are similar, that the largest Δt be chosen).

Estimation of required resources

The present work does not contribute in any specific way to the estimation of the required resources (computer time and memory). These resources clearly increase as Δx and Δt decrease, in a way that, except for the tracking algorithm, can be found through code-dependent counting of operations and array sizes. Because of the self-control that the tracking algorithm has on its accuracy, resources required by this task will significantly increase with the variability of the flow in space and time, and experience should play a key role in their estimation.

4. RECOMMENDATIONS FOR FURTHER WORK

Each paper contains, as appropriate, recommendations for further research in its specific area of interest. Here, we simply suggest that a critical examination of the conceptual decision-making procedure illustrated in Figure 3 provides a good reference for more general/applied recommendations.

Specifically, steps for further advancement in our ability to solve the transport equation involve:

- A better understanding of how to define the physical problem, and, in particular, how to incorporate our scales of interest in the formulation of the mathematical problem, through a proper distinction between which part of the flow is explicitly represented as advection, and which part contributes to dispersion (in the line of the work by [A2] and [R1]).
- The development and application of ELM models in a wide range of

engineering and environmental problems. None of the presently available ELM codes is, in our perspective, flexible enough to fully explore the potentials of the method, in the broad perspective that we defended in the previous section; however, at least the two-dimensional finite element code ELA ([B2]), and, probably, other available codes, were developed in a way that can easily accommodate the required extensions. Each application should be carefully monitored with regard to: (a) accuracy (this is recognized to involve some subjectiveness in the definition of error measures, and considerable difficulty in the collection of reference data); (b) cost (expressed at least in terms of computational time and memory requirements); (c) practical difficulties in the implementation, and effectiveness of the conceptual decision-making procedure of Figure 3, or similar. The experience by individual users or groups should be periodically compiled, and divulged in the open literature (journals, conference proceedings, specialized forums, etc).

- The further extension of our theoretical understanding of ELM. Areas of particular interest include the evaluation of tracking errors, the verification of the validity in multi-dimensions of concepts developed essentially in 1-D, and the analysis of the effect of grid irregularity. Some work on the last subject was presented by [B5], which, together with on-going research, strongly suggests that, whenever isoparametric mappings are used to perform interpolations, the "improved isoparametric mapping" proposed by [C1] should be used to avoid a strong loss of accuracy.

REFERENCES

- A1 Adams, E. E., R. Kossik, and A. M. Baptista. "Source Representation in a Numerical Transport Model." *Finite Elements in Water Resources: Proceedings of the 6th International Conference* (Sá da Costa et al., Ed.), Springer-Verlag, 1986.
- A2 Aldama-Rodríguez, A. "Theory and Applications of Two- and Three-Scale Filtering Approaches for Turbulent Flow Simulation." Ph.D. Thesis, M.I.T., 1985.
- B1 Baptista, A. M., E. E. Adams, and K. D. Stolzenbach. "The 2-D Unsteady Transport Equation Solved by the Combined Use of the Finite Element Method and the Method of Characteristics", in *5th Int. Conf. on Finite Elements in Water Resources*. Burlington, Vermont, 1984a.
- B2 Baptista, A. M., E. E. Adams, and K. D. Stolzenbach. "Eulerian-Lagrangian Analysis of Pollutant Transport in Shallow Water." MIT R. M. Parsons Laboratory, Technical Report 296, 1984.
- B3 Baptista, A. M., E. E. Adams, and K. D. Stolzenbach. "Comparison of Several Eulerian-Lagrangian Models to Solve the Advection-Diffusion Equation." in *Int. Symp. on Refined Flow Modeling and Turbulence Measurements*. U. Iowa, USA, 1985.
- B4 Baptista, A. M., E. E. Adams, and K. D. Stolzenbach. "Accuracy Analysis of the Backwards Method of Characteristics." in *Finite Elements in Water Resources: Proceedings of the 6th International Conference*. (Sá da Costa et al., Ed.), Springer-Verlag, 1986.
- B5 Baptista, A. M. "Accurate Numerical Modeling of Advection-Dominated Transport of Passive Scalars." LNEC, Lisboa, 1986.
- B6 Baptista, A. M. "Fourier Analysis of the Backwards Method of Characteristics." In Ph.D. Thesis, M.I.T., 1987.
- B7 Baptista, A. M. "The Consistency, Stability, and Convergence of the Backwards Method of Characteristics." In Ph.D. Thesis, M.I.T., 1987.
- B8 Baptista, A. M. "The Choice of the Interpolator for the Backwards Method of Characteristics." In Ph.D. Thesis, M.I.T., 1987.
- B9 Baptista, A. M. "The Accuracy of Eulerian-Lagrangian Methods." In Ph.D. Thesis, M.I.T., 1987.
- B10 Benque, J. P., G. Labadie, and G. Ibler. "A Finite Element Method for Navier-Stokes Equations." *3rd Conf. on Fin. Elem. in Flow Problems*, 1980.
- B11 Branski, J. M. "Higher-Order Spline Schemes for the Advection - Diffusion Equation." Submitted to ASCE, 1986.

- C1 Celia, M. A., and W. G. Gray. "An Improved Isoparametric Transformation for Finite Element Analysis." *International Journal for Numerical Methods in Engineering* 20:1443-1459, 1984.
- C2 Cheng, R. T., V. Casulli, and S. Milford. "Eulerian-Lagrangian Solution of the Convection-Diffusion Equation in Natural Coordinates." *Water Resources Research* 20(7):944-952, 1984.
- G1 Glass, J., and W. Rodi. "A Higher Order Numerical Scheme for Scalar Transport." *Comp. Math. in Appl. Mech. and Eng.* 31:337-358, 1982.
- H1 Hasbani, Y., E. Livne, and M. Bercovier. "Finite Elements and Characteristics Applied to Advection-Diffusion Equations." *Computer and Fluids* 11(2):71-83, 1983.
- H2 Holly, F. M., Jr., and T. Komatsu. "Derivative Approximations in the Two-Point Fourth-Order Method for Pollutant Transport." *Proceedings of the Conference on Frontiers in Hydraulic Engineering, ASCE, M.I.T., 1984.*
- H3 Holly, F. M., Jr., and J. M. Polatera. "Dispersion Simulation in 2-D Tidal Flow." *Journal Hydr. Engrg., ASCE, 1984.*
- H4 Holly, F. M., Jr., and A. Preissmann. "Accurate Calculation of Transport in Two Dimensions." *Journal of the Hydraulics Division, ASCE* 103(HY11):1259-1278, Nov. 1977.
- K1 Komatsu, T., F. M. Holly, Jr., and N. Nakashiki. "Numerical Calculation of Pollutant Transport in Rivers and Coastlines", in *4th Congress, Asian and Pacific Division, IAHR, Chiang Mai, Thailand, 1984.*
- K2 Komatsu, T., F. M. Holly, Jr., N. Nakashiki, and K. Ohgushi. "Numerical Calculation of Pollutant Transport in One and Two Dimensions." *Journal of Hydroscience and Hydraulic Engineering* 3(2):15-30, 1985.
- K3 Kossik, R. F. "Tracing and Modeling Pollutant Transport in Boston Harbor." M.Sc. thesis, MIT, 1984.
- L1 Leith, C. E. "Numerical Simulation of the Earth's Atmosphere." *Methods in Computational Physics* 4:1-28, 1965.
- N1 Neuman, S. P. "An Eulerian-Lagrangian Scheme for the Dispersion Convection Equation Using Conjugate Space-Time Grids." *Journal of Comp. Phys.* 41:270-279, 1981.
- N2 Neuman, S. P. "Adaptive Eulerian-Lagrangian Finite Element Method For Advection-Dispersion." *Int. Journal for Numerical Methods in Engineering* 20:321-337, 1984.
- N3 Neuman, S. P., and S. Sorek. "Eulerian-Lagrangian Methods for Advection-Dispersion." *Finite Elements in Water Resources, (K. P. Mole et al., Ed.)* 4:14.41-14.68, 1982.

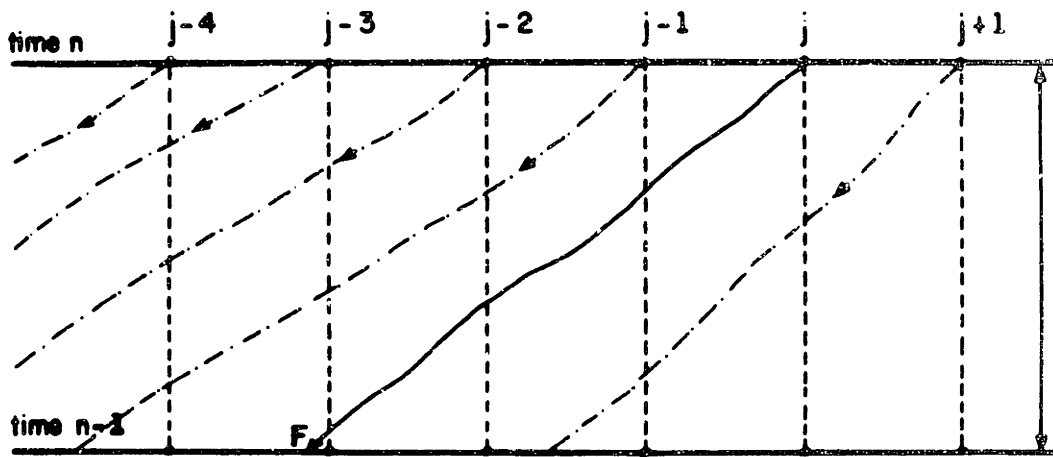
- R1 Rosman, P. C. "Modeling Shallow Water Bodies via Filtering Techniques." Ph.D. Thesis, M.I.T. (in preparation), 1987.
- V1 Varoglu, E., and W. L. Finn. "Space-Time Finite Elements Incorporating Characteristics for the Burgers Equation." *Int. J. Num. Meth. Engrg.* 16:171-184, 1980.

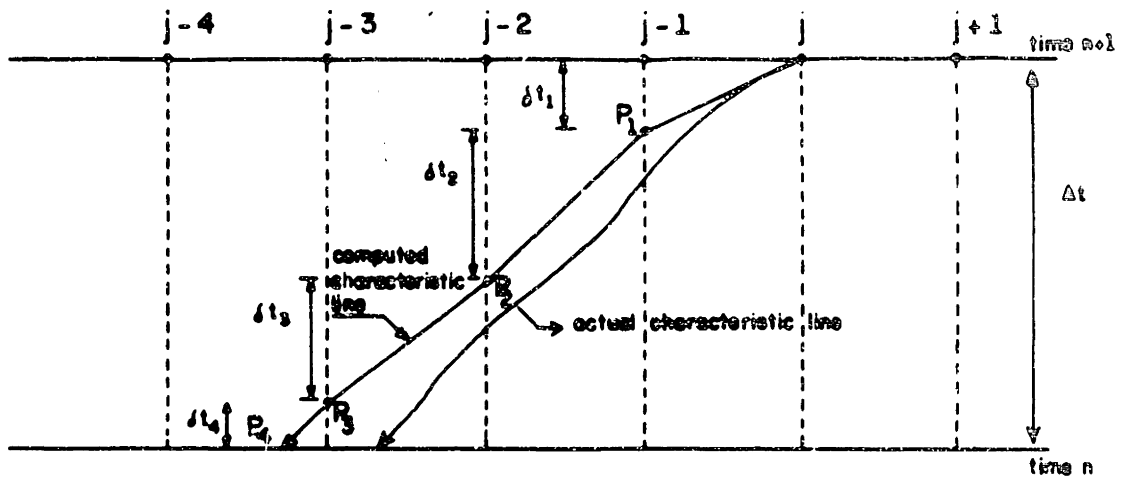
Figure 1

Illustrative sketch for Eulerian-Lagrangian methods

Required steps:

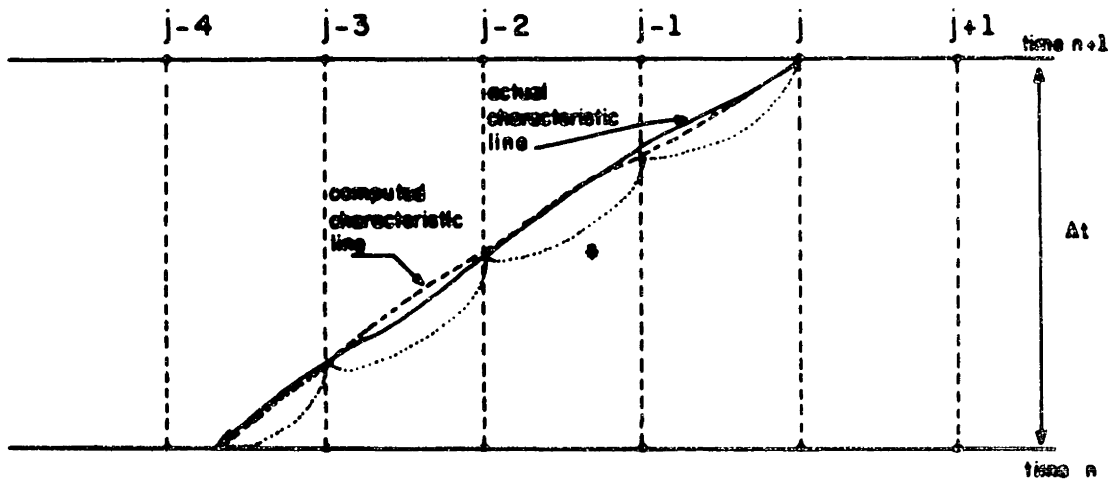
1. Tracking of the characteristic lines. For each node j , a characteristic line is independently defined by the backwards (i.e., between n and $n-1$) solution of an ordinary differential solution of the form $dx_j/dt = u_j$.
2. Interpolation at the feet of the characteristic lines. The concentration at the foot of each characteristic line is found by interpolation from known information on neighboring nodes (time $n-1$).
3. Solution of the transport equation, written in Lagrangian form. This solution involves all nodes simultaneously.





$$\delta t_1 = \frac{x_k - x_{k-1}}{u_k} \quad \delta t_2 = \frac{x_k - x_{k-1}}{u_k} \quad \delta t_3 = \frac{x_k - x_{k-1}}{u_k}$$

$$x(P_4) = x_{k-3} - u_{k-3} \cdot \delta t_4$$



* Accuracy is checked by reverse tracking, and, if necessary, δt is reduced to meet an imposed criterion

Figure 2 Illustration of alternative tracking algorithms: (a) Standard; (b) 4th-order Runge-Kutta, with adjustable time step, δt [B2]

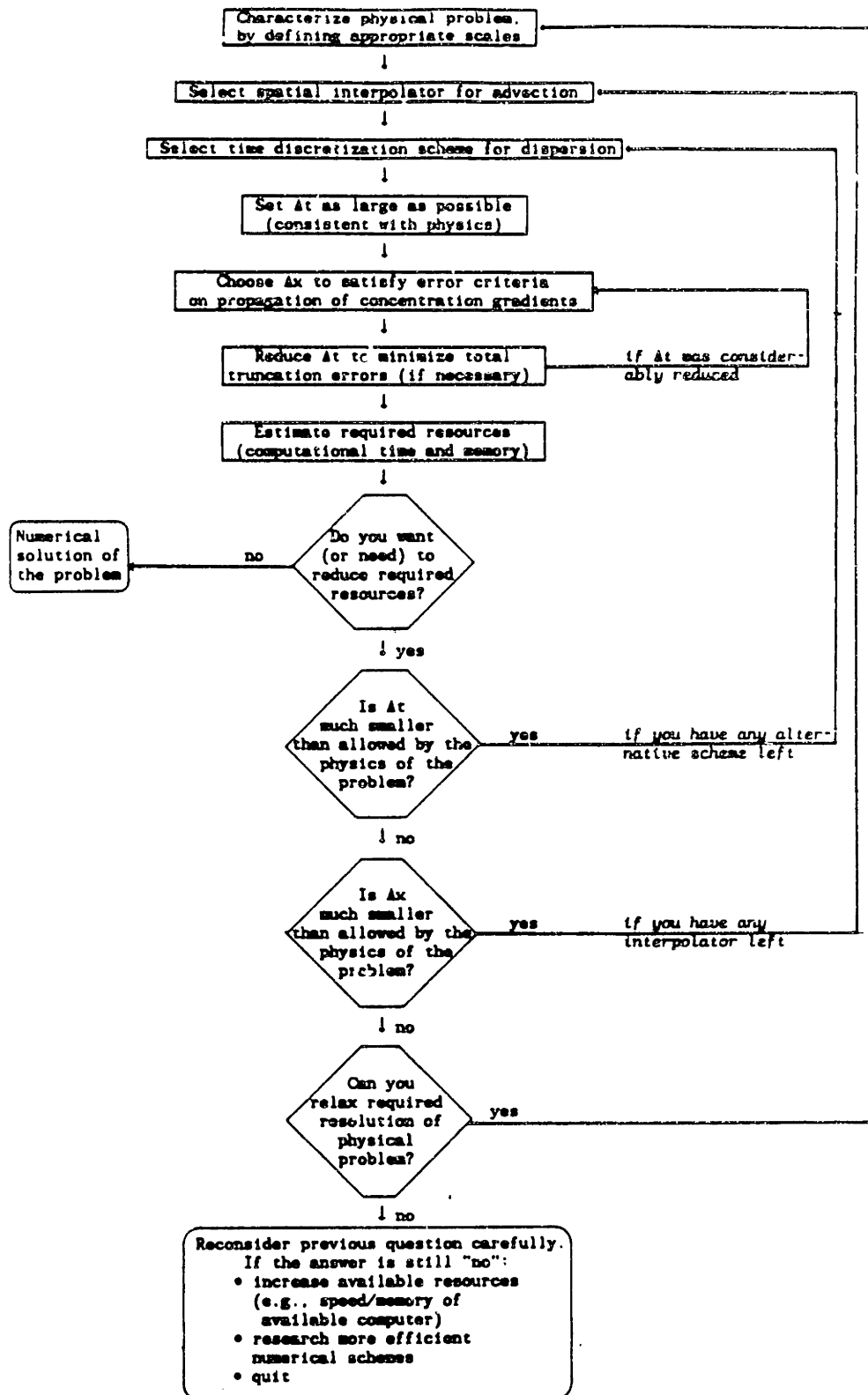


Figure 3 The choice of the computational strategy

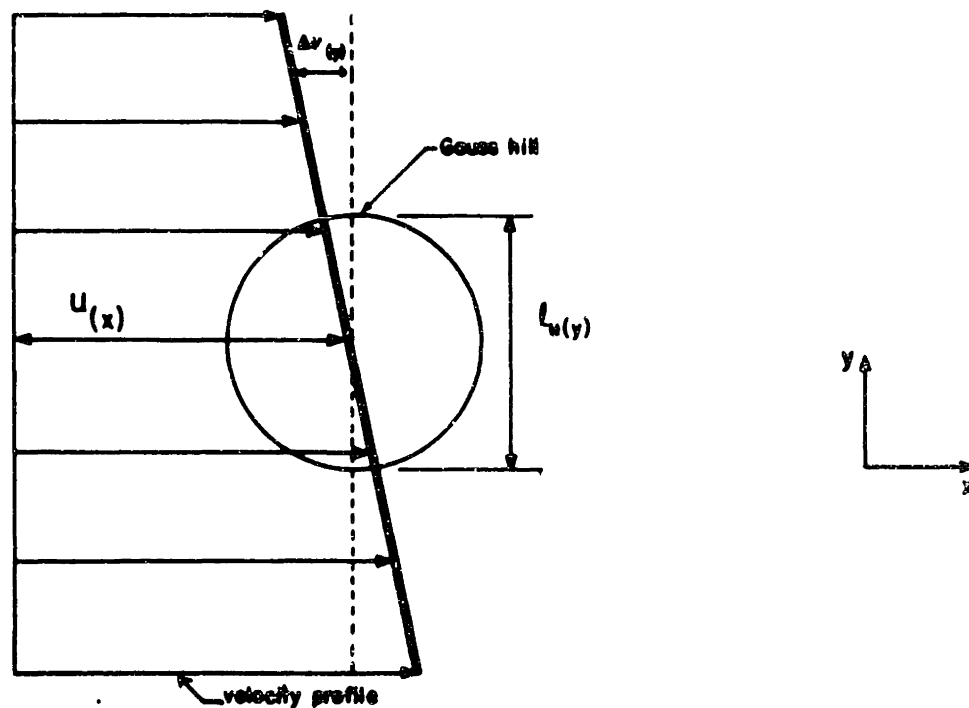
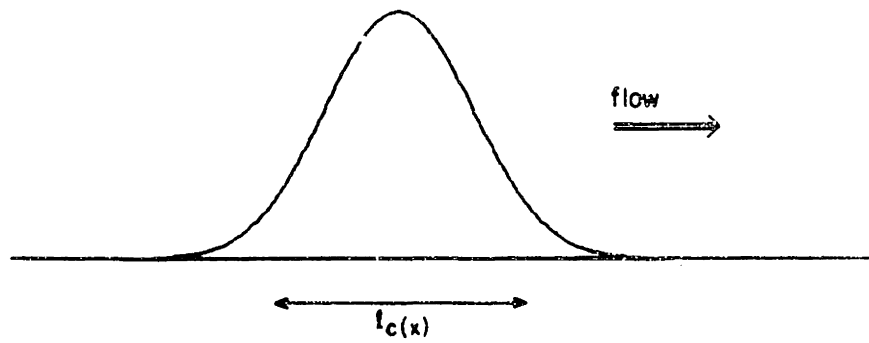


Figure 4 Illustrative scaling for the transport of a Gauss-hill in a shear flow: (a) Concentration profile (side view); (b) Velocity profile (top view)

APPENDIXES¹

¹Adapted from Baptista, A. M., "Accurate Numerical Modeling of Advection-Dominated Transport of Passive Scalars," LNE, Lisboa, 1986.

APPENDIX A

On the Mathematical Nature of the Transport Equation

The detailed analysis of the mathematical nature of the transport equation is beyond the scope of this work. The following brief discussion should, however, be helpful to the understanding of problems arising in the numerical solution of this equation, and is a motivation for Eulerian-Lagrangian methods.

We take as a reference, without loss of conceptual generality, the 1-D partial differential equation

$$\frac{\partial c}{\partial t} + A(x, t) \frac{\partial c}{\partial x} = B(x, t) \frac{\partial^2 c}{\partial x^2} + C(x, t) c + D(x, t) \quad (\text{A.1})$$

where the coefficients $B(x, t)$ and $A(x, t)$ --representing, respectively, dispersion and advection (or advection plus non-uniform dispersion, see [B1])--are of special interest.

This equation is linear, as revealed by the functional dependence of the coefficients A through D^2 . Hence, the extensive body of knowledge

²An n^{th} -order PDE is nonlinear when its coefficients depend on n^{th} -order derivatives of the dependent variable; it is quasi-linear when they depend on m^{th} -order derivatives, with $0 < m < n$; and it is linear when they depend on the independent variables alone.

available on linear PDE applies, and some general properties may be assumed.

In particular, it is easily recognized that Equation A.1 has a considerably different behaviour, depending on whether $B(x,t)$ is or is not null. Indeed, when $B(x,t) \neq 0$, the equation is a second-order parabolic PDE, while it becomes a first-order hyperbolic PDE when $B(x,t) = 0^3$.

Hence, when $B(x,t) \neq 0$, Equation A.1 has a single family of horizontal characteristic lines (or, to be more precise, two coincident--thus necessarily horizontal--families), and is associated with the initial and boundary conditions diagrammatically represented in Figure A.1(a). The function $c(x,t)$ is determined, at any given location of space and time, by all the initial data plus the data on the boundaries which are on or below the relevant characteristic line. Hence, in particular, at any given time, t , solutions at different space locations are all interrelated.

³Second-order linear or quasi linear PDE of the general form

$$a(\cdot) \frac{\partial^2 v}{\partial y^2} + b(\cdot) \frac{\partial^2 v}{\partial y \partial z} + c(\cdot) \frac{\partial^2 v}{\partial z^2} + d(\cdot) \frac{\partial v}{\partial z} + e(\cdot) \frac{\partial v}{\partial z} + f(\cdot) v + g(\cdot) = 0 \quad (A.2)$$

where (\cdot) represents some functional dependence, are classified as hyperbolic if $b^2 > 4ac$, as parabolic if $b^2 = 4ac$, and as elliptic if $b^2 < 4ac$ (e.g., [L1], pp 12-13). The above criterion does not give any useful information for first-order PDE; however, hyperbolic equations are identified as those for which the Cauchy problem is well-posed ([J1], p 42), which is the case for Equation A.1 when $B(x,t) = 0$.

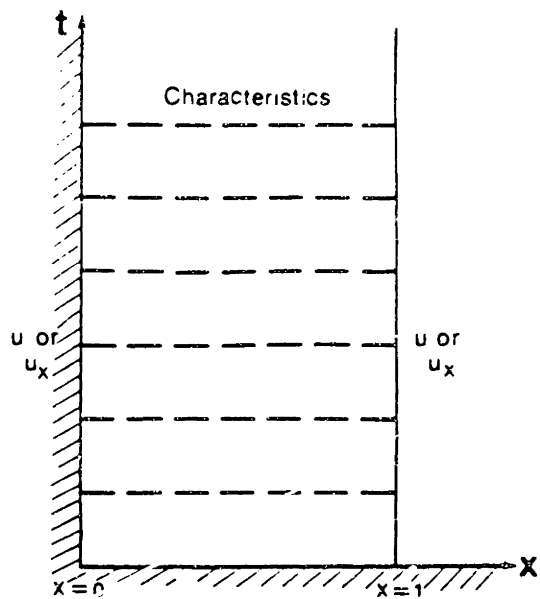
When $B(x,t)=0$, Eq. (A.2) has, again, a single family of characteristic lines. Now, however, this number results from the order of the equation, and not from the coincidence of two families of lines; hence, in particular, the characteristic lines do not have to be (and are not, except in the uninteresting case of $A(x,t) = \infty$) horizontal. The requirements on initial and boundary conditions, and the domain of dependence of the solution are now completely different, as illustrated in Figure A.1(b). In particular, we note that the relevant initial or boundary conditions fully determine the solution along each characteristic line.

The transition between the parabolic and the hyperbolic behavior of the equation is, from the above discussion, discontinuous (associated to a singularity at $B(x,t) = 0$). Although this is formally so, the actual behavior of the solutions is hypothesized to change gradually as dispersion becomes less and less important with regard to advection. This is in agreement with the physics of the transport phenomena, and may justify the frustrating experience of many modellers, in the last decades, while solving numerically the advection-dominated transport equation.

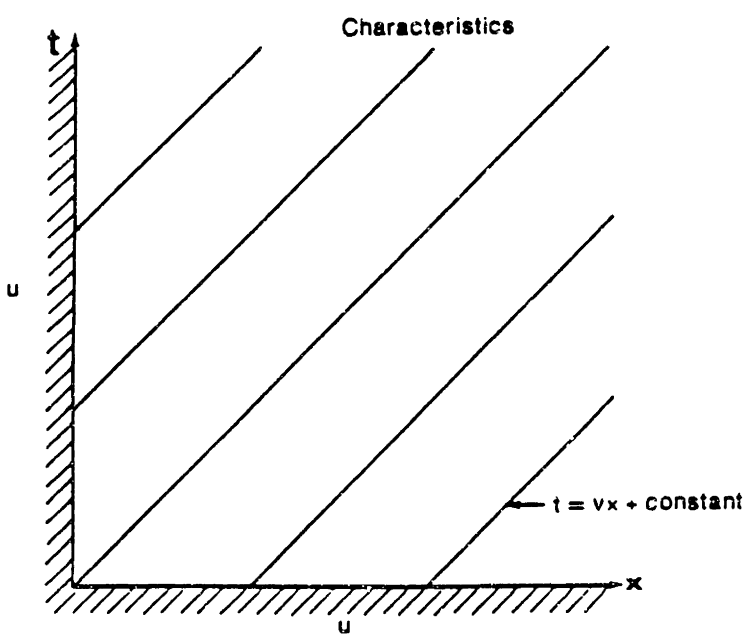
Hence, the idea of splitting the transport into a purely parabolic dispersion equation and a purely hyperbolic advection equation, and solving each by a different method, becomes appealing, and has indeed been increasingly explored (Appendix B).

REFERENCES

- B1 Baptista, A. M., E. E. Adams, K. D. Stolzenbach. "Eulerian-Lagrangian Analysis of Pollutant Transport in Shallow Water." MIT, R. M. Parsons Laboratory, Technical Report 296, 1984.
- J1 Jeffrey, A. "Quasilinear Hyperbolic Systems and Waves." *Research Notes in Mathematics*. Pitman, 1976.
- L1 Lapidus, L., and G. F. Pinder. *Numerical Solution of Partial Differential Equations in Science and Engineering*. J. Wiley & Sons, 1982.



(a) parabolic equation



(b) hyperbolic equation

Figure A.1 Diagrammatic representation of hyperbolic and parabolic partial differential equations

APPENDIX B

Review of Numerical Methods for the Solution of the Transport Equation

B1. Introduction

The transport equation has been solved by several different numerical methods, which may be classified into three broad categories: Eulerian, Lagrangian and Eulerian-Lagrangian.

Eulerian methods (EM) solve the Eulerian form of the transport equation, at the nodes of a fixed grid. This requires the simultaneous solution of hyperbolic (advective) and parabolic (dispersive) operators, which has proved to be a hard task when the former dominates the latter. Indeed, when advection dominates, "centered" EM often generate spurious spatial oscillations (wiggles) while "upwind" and "balanced-dissipation" EM introduce significant numerical damping.

Lagrangian methods (LM) avoid the explicit treatment of hyperbolic operators by solving the Lagrangian form of the transport equation in grids moving with the flow. This approach is potentially very accurate, but is made unattractive or unfeasible in many situations of interest (e.g., continuous sources and complex reversing flows) due to practical difficulties associated with the grid displacement and deformation.

Eulerian-Lagrangian methods (ELM) retain the convenience of a fixed grid, but, at some point of the numerical procedure, a part or the whole transport equation is treated in a Lagrangian form, in order to avoid the explicit treatment of hyperbolic operators. Reported results from ELM are rather promising, showing that wiggles and numerical damping can be greatly reduced, even for very strongly advection-dominated problems.

Discussion in the next paragraphs concentrates on EM (still the most used in engineering practice), and on ELM (which are becoming increasingly used).

B2. Eulerian methods

Methods in this category are typically based on the set-up and solution of a single system of algebraic equations, where both advective and dispersive terms are represented; unknowns are the concentrations at a finite number of fixed locations (nodes) in the computational domain. The transformation of the original differential equation into such a system of algebraic equations is usually achieved using either finite difference methods (FDM) or finite element methods (FEM).

The option between FEM and FDM, while of practical importance, plays a secondary role in what the fundamental difficulties of the accurate solution of the transport equation are concerned. Indeed, as we shall see, each FDM has what can be considered a FEM "equivalent", sharing the same type of fundamental abilities and limitations.

FDM have been used in the solution of the transport equation since the late 1950s. They typically discretize the computational domain through the use of an orthogonal grid (stretching transformations have however been increasingly used to provide some grid refinement or specific shaping). Over each grid element, the differential transport equation is replaced by an algebraic equation, where both the space- and time-derivatives are approximated by finite-differences. The resulting system of algebraic equations is adjusted to take into account the appropriate boundary conditions, and is then solved to give the nodal concentrations.

Initial FDM used centered schemes to approximate both the advection and the dispersion terms. These methods, however, lead often to strong parasitic spatial oscillations (wiggles), specially for large Courant numbers (i.e., often in the range of practically feasible Δx , Δt). In a careful (although too specific) formal analysis in the context of a one-dimensional steady problem, with Dirichlet boundary conditions specified at the two boundaries, [R1] (pp 161-165) showed that wiggles are, in this case, caused by a singularity at low D/u (the numerical equivalent of the singularity at $D=0$ of the behavior of the exact solution, discussed in Appendix A). According to [R1], a perturbation is generated at the outflow boundary, for Peclet numbers, $Pe=u\Delta x/D$, larger than 2, and propagates to the whole domain.

Experience shows, however, that wiggles may have a broader range of origins. A more general statement is that wiggles are the consequence of phase errors at short wavelengths; hence, wiggles will occur whenever such wavelengths are of significance in the true solution (which relates to

insufficient discretization), and are not artificially damped by the numerical algorithm.

As centered FDM are often associated with small damping but significant phase errors at short wavelengths, they should, indeed, promote wiggles, except when physical diffusion is strong enough to smooth out sharp (relative to the grid discretization) gradients.

As a remedy for wiggles in advection-dominated problems with sharp gradients, more recent FDM have used centered finite-differences only for the dispersion terms, replacing the advective derivatives by upwind differences. Upwinding methods do avoid wiggles; however, this is done by very strongly damping short wavelengths; for linear approximations, a numerical diffusion is explicitly introduced (as easily shown by Taylor series expansion), which often overshadows physical diffusion.

[B5] proposed, as an alternative to "brute-force" upwinding, the elimination of wiggles through the controlled addition of (unsteady, non-uniform and non-isotropic) artificial diffusion (Flux-Corrected Transport FD Method) to numerical solutions obtained with centered differences. This and similar techniques lead also to a re-statement of the physical problem, and can loosely be seen as forms of "intelligent," but often relatively expensive, upwinding.

FEM have become popular for the solution of the transport equation since the early 1970s. The computational domain is divided into elements of convenient shape, such as triangles or quadrilaterals. Within each element information is concentrated at nodes, but may be unambiguously

interpolated to the interior using pre-selected interpolation functions. The original partial differential equation is then transformed into a system of ordinary differential equations in time, using a weighted residual method. Numerical integration of this system leads to a system of algebraic equations, whose solution gives the nodal values of the concentration field.

The use of the weighted residual method requires the definition of elementary weighted residuals, resulting from the integration over each element of the errors made in approximating the actual concentration field, weighted by pre-selected weighting functions; the sum over the whole computational domain of the elementary residuals is then forced to be zero, to minimize the approximation errors. Different FEM result from different choices of interpolation and weighting functions. In the early 1970's, most FEM solved the transport equation using the same interpolation and weighting functions; such methods are known as Galerkin-FEM (GA-FEM).

GA-FEM lead to "centered" approximations of the advective terms, and present the same limitation as centered FDM: wiggles are produced when short wavelengths are significant, and are not progressively damped by physical diffusion (Peclet number above a critical value). The increase of the order of the interpolation functions from linear to quadratic seems to have a significant effect on accuracy (e.g. [N4]), but is unable to fully avoid wiggles. Users of GA-FEM (e.g. LEINKULHER 1974) have tried to extend the application of the method to advection-dominated problems with sharp gradients, through the adoption of uniform dispersion coefficients which

are 1 to 2 orders of magnitude larger than the physical ones (which is a rough re-statement of the physical problem).

In the late 1970's several attempts were made to account for the flow direction, i.e., to "upwind" FEM. Petrov-Galerkin FEM (PG-FEM), as presented by [C2], and extended by [H2] and [H3] constitute one such attempt which has been successful in avoiding wiggles. In these methods, the weighting functions are not equal to the interpolation functions, but are obtained from them by a change in shape that increases the relative weight of upstream information in a way that depends on the element geometry and the flow characteristics. Limitations of PG-FEM methods include (a) introduction of numerical damping (close similarity with upwinding FDM); (b) increased computational effort required to generate weighting functions from interpolation functions, at each element and at each time step, and (c) difficulty in handling elements which are not quadrilaterals.

A different upwinding procedure (much in the line of the Flux-Corrected Transport Method in FDM) was proposed in [H7] for 1-D, and was extended to 2-D by [H8] and [K1]. In this procedure, the weighting and interpolation functions are equal, like in standard GA-FEM. However, an artificial anisotropic dispersion term, equivalent to the one that is implicitly introduced by the PG-FEM, is computed and added to each element at each time step. Methods using this procedure have not received a unique designation, but are often referred to as Balanced-Dissipation-FEM (BD-FEM). Results of BD-FEM have been reported as indistinguishable from results obtained with PG-FEM, for a few simple test cases. However, BD-FEM

are much less expensive and are more easily applied to elements of any shape and dimensionality; for complex flows, they should also lead to a more controlled type of upwinding, resulting in enhanced accuracy.

A final comment on Eulerian methods is that none of them can be safely applied for large Courant numbers. As a general rule, explicit methods become unstable for $Cu > 1$ (the classical Courant-Lewy stability criterion), while implicit methods, even if stable, tend to significantly lose accuracy above the same limit.

B3. Eulerian-Lagrangian Methods

We now analyze key aspects of ELM. A distinction will be made between ELM based on the concept of "concentration" (ELM/C), ELM based on the concept of "particle" (ELM/P) and on ELM involving both concepts (ELM/CP). Except for this Appendix, we will refer to ELM/C simply as ELM, throughout the text.

Typically, ELM/C split the transport equation into an advection and a diffusion equations, solving the former by a point-to-point transfer method (e.g., a backwards method of characteristics) and the latter by some conventional global discrete element technique (e.g., finite elements or finite-differences). ELM implementations based on this conceptual approach include those reported by [B1-B2], [B4], [G1], [H1], [H4-H6], [L3], [N1, N3], [K2, K3].

Two major general splitting approaches have been used. The most common of these approaches applies to the time-discretized form of the

transport equation, while the approach suggested by NEUMAN and SOREK 1982 applies to the differential form of the equation. The latter approach is attractive for its formal elegance, potential accuracy, and independence relative to time-discretization schemes; however, when advection is dominant, it apparently generates systematic (although localized) errors in the diffusion step and, therefore its practical advantage over the more conventional former approach is yet to be demonstrated.

The solution of the advection equation by a point-to-point transfer method is based on the fact that the concentration of a particle following the flow remains constant, if advection is the only transport mechanism. Most often, the method takes the form of (a) assigning at each new time step, n , a particle to each node of the computational grid, (b) following each particle backwards along characteristic lines defined by the flow, until reaching time step $n-1$, where concentrations at the foot of each characteristic line are computed, by interpolation between known nodal values, and (c) assigning such concentrations to the corresponding grid nodes at time n . Two major tasks are clearly involved: the particle backtracking and the interpolation to find concentrations at the feet of the characteristic lines.

Very accurate particle tracking algorithms were developed both for simply structured and for complexly structured or unstructured grids (e.g., [B1]). The computational cost of these algorithms increases significantly with the complexity of the grid and of the flow field, but accuracy may be kept excellent within affordable costs.

The interpolation procedure to find the concentrations at the feet of the characteristic lines has proved much harder to handle. A variety of interpolation schemes have been or can be considered; however, even if several of these schemes allow ELM/C to reduce (when compared to EM) the range of dimensionless wavelengths that are affected by significant amplitude and phase errors, no scheme can claim to be free of a "critical" wavelength, which may still be constraining for a number of applications. A comprehensive discussion of the absolute and relative merits and limitations of alternative interpolation schemes is presented in [B3].

Meanwhile, most ELM/C handle accurately the solution of the diffusion equation, by using a conventional centered FD or FE technique. We note that the solution is global (i.e., involves all the grid nodes simultaneously), which implies that large systems of equations must be solved at each time step. The size of these systems has not proved to be a serious problem in 2-D, because they are often nicely banded and symmetric, and because most of the above mentioned ELM/C are implicit, allowing for large time steps (i.e., reducing the number of required solutions of the system of equations). However, in 3-D applications the global solution of the diffusion step will become a major problem in terms of computer costs and memory requirements.

[C1] proposed a ELM/C that is slightly different, conceptually, from the preceding ones. The whole transport equation is written in Lagrangian form, and solved (on a fixed grid) by a backwards method of characteristics in which diffusion is treated as a correction term. The assignment of a particle to each grid node and its backtracking with the flow is shared

with preceding ELM/C; however, the concentration at the feet of the characteristic lines is computed by a weighted-average of the concentrations at points defining a physically based mixing region; the concentration at each of these points is obtained by interpolation between nodal values.

This approach has the merit of providing a non-global solution of the dispersion, which may prove highly valuable in a 3-D context. Also, it allows a natural treatment of non-isotropic dispersion. Reported accuracy and mass preservation characteristics are promising, at the same level of precedent ELM/C. However, restrictions should have to be applied to the maximum allowable time step, to keep on with accuracy, and this may strongly limit the method's efficiency.

ELM/P (e.g., [P1]) are based on a conceptually different approach: particles are introduced in the domain (which was previously discretized in a convenient way) in a number and location related to the initial concentration field; these particles are moved forward with the flow (the flow should represent both "advection" and "diffusion" and is typically described in an Eulerian form); whenever convenient, the number and location of the particles is processed back to the form of concentrations, as to give the instantaneous concentration field.

ELM/P are natural, and physically sound. They inherently avoid the issue of short dimensionless wavelengths, and therefore handle accurately sharp gradients and small (as compared to the grid size) sources of mass, which ELM/C can not do. Also, they are rather versatile, being equally

suited for the analysis of concentrations fields, residual transport and field experiments.

However, ELM/P have some potential problems. Clearly, they are not inherently conservative: mass conservation relies only on accuracy, both requiring that a very large (and sometimes unfeasible) number of particles be tracked, and that a fine support grid be used for the conversion between number and location of particles and concentration. Also, if the "advective" part of the carrying flow may be "easily" found by means of a complementary circulation model, the same is not true for the "diffusive" part (which we will call pseudo-velocities).

Approaches to handle the pseudo-velocities range from purely deterministic to partially statistical methods. [L1], after some manipulation of the theory of diffusion, proposed deterministic pseudo-velocities in the form

$$U_1 = - \frac{K}{C} \frac{\Delta C}{\Delta x_1} \quad (B.1)$$

where U_1 is the pseudo-velocity in the 1-direction, K is a conventional eddy diffusivity coefficient, and C is the concentration, expressed in terms of the number of particles. In this deterministic method, the motion of a single particle is affected by the whole concentration field, i.e., by the positions of the other particles.

Statistical approaches rely on associating the pseudo-velocities to random perturbations of the motion of individual particles. Again, this may be made by resorting to the eddy diffusivity concept, and using it as

to define the statistics of the random motion (e.g., as suggested by [C3]); or in a more fundamental way, by extracting the statistics of the random motion from Eulerian records of the flow ([Z1]).

The ELM/CP proposed by [N2] constitutes a hybrid and very promising novel approach. A ELM/C formulation is used everywhere in the domain, except near gradients too sharp to be handled accurately this way; a ELM/P formulation is adopted in these cases (just for the advection equation). With this approach, most of the computational effort is based on a fixed reference grid; forward tracking of particles is required only in specific regions of time and space, and therefore involve only an affordable number of particles. Principal gray areas for this approach include: mass preservation; efficient and consistent detection of sharp gradients; and accurate procedures for mapping concentrations from particles to the nodes of the fixed grid (we note that this mapping must be performed each time step, previous to the solution of the diffusion equation). All these gray areas may become harder to handle for complex flows than they are for the simple flows that have been used so far to demonstrate the effectiveness of the approach.

ELM overcome in a natural way the limitation on the Courant number referred to for EM, which constitutes one of the fundamental advantages of ELM.

REFERENCES

- B1 Baptista, A. M., E. E. Adams, and K. D. Stolzenbach. "Eulerian-Lagrangian Analysis of Pollutant Transport in Shallow Water." MIT R. M. Parsons Laboratory, Technical Report 296, 1984.
- B2 Baptista, A. M., E. E. Adams, and K. D. Stolzenbach. "Comparison of Several Eulerian-Lagrangian Models to Solve the Advection-Diffusion Equation." in *Int. Symp. on Refined Flow Modeling and Turbulence Measurements*. U. Iowa, USA, 1985.
- B3 Baptista, A. M. "The Choice of the Interpolator for the Backwards Method of Characteristics." In Ph.D. Thesis, M.I.T., 1987.
- B4 Benque, J. P., G. Labadie, and G. Ibler. "A Finite Element Method for Navier-Stokes Equations." *3rd Conf. on Fin. Elem. in Flow Problems*, 1980.
- B5 Book, D. L., J. P. Boris, and K. Hain. "Flux-Corrected Transport II: Generalization of the Method." *J. of Comp. Phys.* 18:248-283, 1975.
- C1 Cheng, R. T., V. Casulli, and S. Milford. "Eulerian-Lagrangian Solution of the Convection-Diffusion Equation in Natural Coordinates." *Water Resources Research* 20(7):944-952, 1984.
- C2 Christie, I., D. F. Griffiths, A. R. Mitchell, and O. C. Zienkiewicz. "Finite Element Methods for Second Order Differential Equations with Significant First Derivatives." *Int. J. Num. Meth. Engrg.* 10:1389-1396, 1976.
- C3 Csanady, G. T. "Turbulent Diffusion in the Environment." D. Reidel, 1973.
- G1 Glass, J., and W. Rodi. "A Higher Order Numerical Scheme for Scalar Transport." *Comp. Math. in Appl. Mech. and Eng.* 31:337-358, 1982.
- H1 Hasbani, Y., E. Livne, and M. Bercovier. "Finite Elements and Characteristics Applied to Advection-Diffusion Equations." *Computer and Fluids* 11(2):71-83, 1983.
- H2 Heinrich, J. C., P. S. Huyakorn, O. C. Zienkiewicz, and A. R. Mitchell. "An 'Upwind' Finite Element Scheme for Two-Dimensional Convective-Transport Equation." *Int. J. Num. Meth. Engrg.* 11:131-143, 1977.
- H3 Heinrich, J. C., and O. C. Zienkiewicz. "Quadratic Finite Element Schemes for Two-Dimensional Convective Transport Problems." *Int. J. Num. Meth. Engrg.* 11:1831-1844, 1977.
- H4 Holly, F. M., Jr., and T. Komatsu. "Derivative Approximations in the Two-Point Fourth-Order Method for Pollutant Transport." *Proceedings of the Conference on Frontiers in Hydraulic Engineering*, ASCE, M.I.T., 1984.

- H5 Holly, F. M., Jr., and J. M. Polatera. "Dispersion Simulation in 2-D Tidal Flow." *Journal Hydr. Engrg.*, ASCE, 1984.
- H6 Holly, F. M., Jr., and A. Preissmann. "Accurate Calculation of Transport in Two Dimensions." *Journal of the Hydraulics Division*, ASCE 103(HY11):1259-1278, Nov. 1977.
- H7 Hughes, T. R. "A Simple Scheme for Developing Upwind Finite Elements." *Int. J. Num. Meth. Engrg.* 12:1359-1365, 1979.
- H8 Hughes, T. R., and A. Brooks. "A Multi-Dimensional Upwind Scheme with No Crosswind Diffusion." *Proc. A.S.M.E. Speciality Conference*, New York, 1979.
- K1 Kelly, D. W., S. Nakazawa, O. C. Zienkiewicz, and J. C. Heinrich. "A Note on Upwinding and Anisotropic-Balancing Dissipation in Finite Element Approximations to Convective Diffusion Problems." *Int. J. Num. Meth. Engrg.* 15:1705-1711, 1980.
- K2 Komatsu, T., F. M. Holly, Jr., and N. Nakashiki. "Numerical Calculation of Pollutant Transport in Rivers and Coastlines", in 4th Congress, Asian and Pacific Division, IAHR, Chiang Mai, Thailand, 1984.
- K3 Komatsu, T., F. M. Holly, Jr., N. Nakashiki, and K. Ohgushi. "Numerical Calculation of Pollutant Transport in One and Two Dimensions." *Journal of Hydrosience and Hydraulic Engineering* 3(2):15-30, 1985.
- L1 Lange, R. in *J. Appl. Meteor.*, 17:320, 1978.
- L2 Leimkuhler, W. F. "A Two-Dimensional Finite Element Dispersion Model." Civil Engineer Thesis, M.I.T., 1974.
- L3 Leith, C. E. "Numerical Simulation of the Earth's Atmosphere." *Methods in Computational Physics* 4:1-28, 1965.
- N1 Neuman, S. P. "An Eulerian-Lagrangian Scheme for the Dispersion Convection Equation Using Conjugate Space-Time Grids." *Journal of Comp. Phys.* 41:270-279, 1981.
- N2 Neuman, S. P. "Adaptive Eulerian-Lagrangian Finite Element Method For Advection-Dispersion." *Int. Journal for Numerical Methods in Engineering* 20:321-337, 1984.
- N3 Neuman, S. P., and S. Sorek. "Eulerian-Lagrangian Methods for Advection-Dispersion." *Finite Elements in Water Resources*, (K. P. Mole et al., Ed.) 4:14.41-14.68, 1982.
- N4 Noronha and Baker. "Solutions of the Reference Test Problems." *Convection-Diffusion Forum*, VI FEWR, LNEC, Lisboa, 1986.

P1 Prickett, T. A., et al. "A Random-Walk Solute Transport Model for Selected Groundwater Quality Evaluations." Bull. 65 , Champaign, Illinois State Water Survey, 1981.

R1 Roache, P. *Computational Fluid Dynamics*. Hermosa Publ., 1982.

Z1 Zanetti, P. "New Monte Carlo Scheme for Simulating Lagrangian Particle Diffusion with Wind Shear Effects." *Appl. Math. Modeling* 8, June 1984.

APPENDIX C

On the Role of Advection and Dispersion

From the derivation of the 2-D (depth-averaged) transport equation-- e.g., see [B1]--it should be clear that advection and dispersion are not staunching mechanisms, the actual meaning of each depending on the time and space scales that we elect to explicitly represent in a given formulation.

To illustrate concepts, and provide insight on the relative importance of these mechanisms, we loosely approximate a continuum host fluid of a passive scalar, in turbulent motion, through an "infinite" set of equal fluid parcels, each of infinitesimal (although larger than the scale of Brownian motion) size. Mass of the scalar is associated to each parcel according to some initial distribution, and we examine the changes that take place as time progresses.

If we look at the problem through the three-dimensional instantaneous form of the transport equation, we identify the transport mechanisms as being advection by the instantaneous flow and molecular diffusion.

Advection redistributes mass through the displacement of fluid parcels, which follow the flow; the original amount of mass in each parcel is strictly preserved. It is useful, at this point, to consider the carrying flow as a superposition of three components: a uniform component, a shear component (associated with the deviations from uniformity that

refer to space and associated time scales larger than what is loosely called the scale of the turbulent fluctuations) and a fluctuating component (associated with the deviations from uniformity within the scale of the turbulent fluctuations). While the uniform component moves parcels in a rigid-body-like way, the non-uniform components rearrange the position of the parcels relative to each other: a deterministic rearrangement over large scales, in the case of the shear component, and a random rearrangement over small scales, in the case of the fluctuating component.

Diffusion, in turn, redistributes mass by actually exchanging it between adjacent parcels, through Brownian motion at the molecular level; no parcel displacements are involved. Mass exchange is set in a way that tends to smooth out existing gradients (i.e., mass goes from parcels with higher content to parcels with lower content), and the exchange rate is proportional to the driving gradients.

For passive scalars, diffusion does not affect advection. However, advection by the non-uniform components of the carrying flow may significantly affect the efficiency of diffusion. Indeed, the relative position of fluid parcels is changed by non-uniform advection, which establishes new parcel neighborhoods: if parcels with high mass content are all brought closer to each other, diffusion becomes less efficient; conversely, if such parcels are spread over larger regions, diffusion becomes more efficient. Typically, non-uniform advection enhances global diffusion, although it may inhibit diffusion locally.

We now take the three-dimensional form of the equation for mean (in the sense of Reynolds averaging) turbulent transport. Both advective and

dispersive mechanisms are, again, present. Now, however, advection is associated only with the mean turbulent flow (uniform and shear components), and the effect of the fluctuating component of the flow is represented as a turbulent diffusion. As pointed out earlier, turbulent diffusion is typically several orders of magnitude more efficient than molecular diffusion, and the latter can therefore be dropped from the governing equation.

Because diffusion implies exchange of mass between fluid parcels, we immediately recognize, from earlier discussion, that the smallest size of the parcels that we can look at has increased: parcels are still small, but they must be larger than the scales associated with turbulent fluctuation.

We lost resolution, but we gained convenience. Indeed, we avoided the explicit representation of the fluctuating component of the flow, which is particularly hard to handle. This approximation may constitute the difference between feasibility and unfeasibility in the modeling of turbulent flow and transport⁴.

⁴To illustrate this statement, we reproduce from [W1] the following reasoning based on figures given by [E1]: for the relatively simple problem of turbulent flow in a pipe, a computer solution revealing the turbulence structure of the flow, at a Reynolds number of 10^7 , would require 10^{22} operations; at the computer speed of 10 microseconds per operation (representative of computers in the seventies), this would require 10^{17} , or 3×10^9 years (about one-fifth of the age of the universe); since we are probably limited by the speed of light to an "ultimate" computer speed of 1 nanosecond per operation, our fastest foreseeable computation would take 3.2×10^4 years (over 500 generations).

Let us now consider the two-dimensional equation for mean (again, in a Reynolds averaging sense) turbulent transport. Advection is associated only with the uniform component and with a part of the shear component of the flow. Indeed, the effect of vertical shear is represented as a dispersion mechanism (vertical-shear dispersion) and added to the turbulent diffusion.

Again, we lose resolution (the horizontal size of the fluid parcels is still constrained only by the scale of the turbulent fluctuations, but the vertical size must be the flow depth) to gain convenience (we avoid the explicit representation of the vertical flow and of the vertical variation of the horizontal flow).

Computational savings related to depth-averaging are much less impressive than those achieved by Reynolds--averaging, but are still significant (may be one to two orders of magnitude in CPU and memory requirements). Although advanced computers already exist that make feasible the computational effort for the solution of three-dimensional mean turbulent transport problems, depth-averaging (or an alternative space averaging) is still often useful or even the only sensible or feasible approach (because of: limitations of the computer actually available; detail that can be achieved in the specification of boundary conditions; accuracy of available numerical solution techniques; etc).

It should be emphasized that the two-dimensional equation for mean turbulent transport assumes that the horizontal plan is represented as a continuum; however, the numerical solution of this equation involves the

discretization of the horizontal plan, through the set-up of a grid. While in some cases (e.g., for finite-element methods) unambiguous interpolation functions hold within each grid element, some space- (and associated time-) scales of the flow are, again, omitted or ill-represented. This further reduces our ability to directly represent advection, and should, again, be compensated by the introduction of an additional dispersion mechanism (see detailed discussion in [A1] and [R1])

Relevant questions are how to evaluate the dispersion coefficient that is actually going to be used in the computations, and how important has dispersion become relative to advection.

Clearly, the answers depend on the specific problem and on the model (dimensionality and form of the equations; solution technique and its spatial refinement) that one elects to use.

REFERENCES

- A1 Aldama-Rodríguez, A. "Theory and Applications of Two-and Three-Scale Filtering Approaches for Turbulent Flow Simulation." Ph.D. Thesis, M.I.T., 1985.
- B1 Baptista, A. M. "Accurate Numerical Modeling of Advection-Dominated Transport of Passive Scalars." LNEC, Lisboa, 1986.
- E1 Emmons, H. W. *Annu. Rev. Fluid Mech.* 2:15-36, 1970.
- R1 Rosman, P. C. "Modeling Shallow Water Bodies via Filtering Techniques." Ph.D. Thesis, M.I.T. (in preparation), 1987.
- W1 White, F. M. *Viscous Fluid Flow*. McGraw Hill, 1974.

FOURIER ANALYSIS OF THE BACKWARDS METHOD OF CHARACTERISTICS

by

António Melo Baptista

Massachusetts Institute of Technology

Cambridge, Massachusetts 02139

January 1987

ABSTRACT

We present a detailed Fourier analysis of the solution of the 1-D advection equation by the Backwards Method of Characteristics (BMC). The analysis shows that, when the grid has different types of nodes (e.g., the case of FE grids, for quadratic or higher order elements), the numerical solution internally generates a mechanism of energy transfer between Fourier components, as a direct consequence of the fact that the error in each node depends on the nodal type. In this case, the standard Von Newman procedure of studying accuracy and stability on the basis of the errors in the first time-step is not valid, as errors become time-dependent. Error formulae that account for time-dependence of error propagation are derived for the case of quadratic core elements, and are used to show that the transfer mechanism has a transitory, although non-negligible, effect.

1. INTRODUCTION

Fourier analysis has played a significant role in the study of the stability and accuracy of numerical methods, since its introduction by Von Newman, circa 1940 (e.g., [R1]). The common assumption behind the analysis is that the amplitude and phase errors in the propagation of the individual Fourier components that constitute the solution are time-independent, i.e., repeat themselves time-step after time-step. It is known that this assumption requires uniform flows and constant nodal spacing; however, the analysis has been applied to grids with quadratic core elements, for instance in the context of finite elements, disregarding the non-uniformity due to the presence of different types of nodes.

In the context of a more general effort on the study of the formal properties of Eulerian-Lagrangian Methods based on the Backwards Method of Characteristics (BMC) [B1-B7], we were faced with the need to investigate the performance of a large number of alternative interpolators used by BMC. Fourier analysis is very well suited to support such investigation, but the

fact that some of the interpolators use quadratic core elements raised a significant problem: for these interpolators, we could not match the results of Fourier analysis, as developed by Von Newman, with those of numerical experimentation, even for standard test problems involving uniform flows and grids.

This study presents the fundamentals for a generalized Fourier analysis of the BMC, for any type of 1-D interpolator that obeys the following conditions:

- The interpolator applies over a core element that has two or three nodes (*linear elements* or *quadratic elements*, respectively); nodes outside the core element are allowed to contribute to the definition of the interpolator.

- The interpolator is of class C_0 , i.e., it enforces the continuity of concentrations, but not of its derivatives, between adjacent elements.

Applications of this analysis are reported by [B4-B6].

2. REVIEW OF THE BACKWARDS METHOD OF CHARACTERISTICS

The advection equation

$$\frac{Dc}{Dt} = \frac{\partial c}{\partial t} + u_1 \frac{\partial c}{\partial x_1} = 0 \quad (1)$$

states that the concentration, $c(x_1, t)$ remains constant along characteristic lines that follow the flow, i.e., obeying the constraint

$$\frac{dx_1}{dt} = u_1(x_1, t) \quad (2)$$

The BMC is a direct numerical application of this statement, involving, per time step, Δt , two basic tasks (Figure 1):

- Tracking backwards, between times n and $n-1$, the characteristic lines whose heads coincide with the nodes of a fixed reference grid, so as to locate their feet.

- Finding the concentration at the foot of each characteristic line, by interpolation from the known nodal concentrations at time $n-1$, and assigning it to the head of the characteristic line.

Given a 1-D uniform grid, and a constant velocity, this procedure can be represented by the algorithm

$$c(j, n) = c(j - \beta, n - 1) \equiv \sum_{p=P_1}^{P_2} \phi_p(\alpha) \cdot c(j - k + p) \quad (3)$$

where (see Figure 2 for reference)

- j - denotes the node where the concentration is to be computed
(global notation)
- n - denotes the instant where the concentration is to be computed
- α - denotes the position of the foot of the characteristic line in a local coordinate system with origin at node $j-k$ (α is associated with the fractional part of the Courant number, $\beta = u \cdot \Delta t / \Delta x$)

P_1, P_2 - denote the extreme nodes of the region that is used to define the interpolator (in local notation)

ϕ_p - are elementary shape functions which, together, define the interpolator

We note the distinction between the region that is used to define the interpolator, and the region where the interpolator is applied--"core element"; all nodes of the core elements contribute the definition of the interpolator, but this may (non-compact interpolators) or may not (compact interpolators) require in addition information from outside nodes. Several interpolators have been used or considered for use in the context of the BMC [B6]. While most of our present analysis is based on the general algorithm described by Equation 3, we will refer occasionally, for illustration purposes, to a selected set of the interpolators considered by [B6].

3. FORMAL ERROR ANALYSIS

3.1. Reference framework

We consider the constant advection of an instantaneous source of a passive scalar in an infinite domain. The exact solution of this problem can be diagrammatically represented, in the spacial window $x \in [0, X]$, as

$$c(x, 0) \rightarrow \boxed{F[\cdot]} \rightarrow c(x, t) = F[c(x, 0)] \quad (4)$$

where both the initial conditions and the exact solution can be conveniently represented in the form of Fourier-series expansions,

$$c(x,0) = \sum_{m=-\infty}^{+\infty} A_m \exp\{i\mu_m x\} \quad (5)$$

$$c(x,t) = \sum_{m=-\infty}^{+\infty} A_m \exp\{i\mu_m (x-ut)\} \quad (6)$$

In the above, $F[\cdot]$ is the transfer function of the governing equation, $i \equiv \sqrt{-1}$, u is the velocity of the carrying flow, $\mu_m = 2\pi m/\lambda$ is the wavenumber of the m^{th} Fourier component, and A_m is the corresponding amplitude.

We now consider the BMC solution of the same problem, which we view as the combination of two basic operations: (a) the sampling of a continuous signal (the initial conditions) in the nodes of the numerical grid, and (b) the numerical propagation of the resulting discrete signal, in accordance with the numerical algorithm. A third operation (the restitution of a continuous signal by interpolation from the propagated discrete signal), has to be performed if non-nodal information is to be made available. Diagrammatically:

$$\text{Sampling: } c(j,0) \rightarrow \boxed{\text{CD}[\cdot]} \rightarrow \bar{c}(j,0) \quad (7)$$

$$\text{Numerical propagation: } \bar{c}(j,0) \rightarrow \boxed{\tilde{\text{F}}[\cdot]} \rightarrow \tilde{c}(j,n) \quad (8)$$

$$\text{Restitution: } \tilde{c}(j,n) \rightarrow \boxed{\text{DC}[\cdot]} \rightarrow \hat{c}(j,n) \quad (9)$$

where, for a uniform space-time grid described by

$$x_j = (j-1)\Delta x \quad j=1,2,\dots,J+1 \text{ (J even)} \quad (10)$$

$$t_n = n\Delta t \quad n=1,2,\dots,N \quad (11)$$

we will have, in general (with $\lambda_m = \mu_m \cdot \Delta x$)

$$\bar{c}(j,0) = \sum_{m=-J/2}^{J/2} B_m \exp\{i\lambda_m(j)\} = \sum_{m=-J/2}^{J/2} \bar{c}_m(j,0) \quad (12)$$

$$\tilde{c}(j,n) = \sum_{m=-J/2}^{J/2} E_m(n) \cdot \exp\{i\lambda_m(j-n\beta)\} = \sum_{m=-J/2}^{J/2} \tilde{c}_m(j,n) \quad (13)$$

$$\hat{c}(x,t) = \sum_{p=P_1}^{P_2} \phi_p(\alpha) \tilde{c}(j-k-p,n) = \sum_{k=-\infty}^{\infty} \sum_{p=P_1}^{P_2} \phi_p(\alpha) \tilde{c}_m(j-k-p,n) \quad (14)$$

Each of the above steps may, and will often, introduce errors, the nature of which is briefly discussed in the remaining of this section. A

detailed analysis of the errors in the propagation step will be provided in Section 3.2.

Errors in the sampling step are due to aliasing, and are independent of the specific numerical technique. The error mechanism is well-known from digital signal processing (e.g., [O1]): because the minimum wavelength that the grid can resolve is $2\Delta x$, the energy of smaller wavelengths is misinterpreted as associated with larger wavelengths, such that:

$$B_m = 0 \quad \text{for } m > J \quad (15)$$

$$B_m \neq A_m \quad \text{for } m \leq J \text{ (s.t. } \bar{c}(j,0) = c(x_j, 0) \text{ for all } j) \quad (16)$$

Hence, while not affecting concentrations at the nodal points, aliasing destroys the ability of the original continuous signal to be recovered, regardless of the interpolation procedure that may be adopted. Clearly, aliasing can be avoided only when the initial conditions have a Fourier representation that is band-limited in the high-frequency range; even in this case, however, practical computational considerations may limit our ability to sufficiently refine the grid.

Errors in the propagation step are directly associated with both the selected numerical algorithm and the grid characteristics. For the BMC, and assuming that one of the available virtually-error-free tracking algorithms (e.g., [B1]) is used, propagation errors are simply due to the interpolation procedure required to find the concentrations at the feet of the characteristic lines. This procedure can be seen as the restitution,

by interpolation, of a continuous signal from a discrete signal (nodal concentrations, at time $n-1$), followed by the sampling of a new discrete signal (concentrations at the feet of the characteristic lines, at time $n-1$, i.e., nodal concentrations, at time n). Diagrammatically

$$\tilde{c}(j,n-1) \rightarrow \boxed{\text{Interpolation}} \rightarrow \dot{c}(x,n) \rightarrow \boxed{\text{Sampling}} \rightarrow \tilde{c}(j,n) \quad (17)$$

The sampling can generate errors only if, through some sort of non-linear mechanism, energy is transferred, during the interpolation, to wavelengths smaller than $2\Delta x$ (which, we recall, had been zeroed at the level of the initial conditions). The potential error mechanism will be called internal aliasing, to indicate that, unlike the aliasing of the original conditions, it depends on the numerical algorithm.

For linear equations, uniform coefficients and uniform grids, propagation errors have typically been studied by examination of the propagation of individual Fourier components in a single time-step. The underlying assumption is that, because non-linear transfer of energy can not occur between Fourier components, errors are time-independent, in which case the coefficient E_m in Equation 13 can be written simply as

$$E_m(n) = \begin{cases} B_m \cdot \left\{ G_m(\beta) \right\}^n & \text{if } m \leq J \\ 0 & \text{if } m > J \end{cases} \quad (18)$$

where $G_m(\beta)$ --typically a complex function--represents the error in any time-step.

We will show, however, that energy transfer between Fourier components can occur even in uniform grids, when different nodal types co-exist (as in the case of BMC based on interpolation functions with quadratic or higher-order core elements). This purely numerical transfer (non-linearity) leads to propagation errors per time step that are both time-dependent and a function of the Fourier representation of the actual initial conditions of the problem under solution.

Given the particular characteristics of the non-linearity, we suggest that a systematic analysis of propagation errors can still be performed on the basis of the propagation of individual Fourier components, by separating the analysis of the errors for each nodal type, according to the model

$$E_m^v(n) = B_m \cdot H_m^v(n) \quad (19)$$

where v denotes the type of node j , and $H_m^v(n)$ is time-dependent, but is independent of the Fourier representation of the actual initial conditions. This approach, which assumes that internal aliasing is negligible, will be described in detail in Section 3.2.

Restitution errors result exclusively from interpolation between nodal values, and if, as assumed in Equation 14, the same interpolator selected to find the concentrations at the feet of the characteristic lines is used, these errors can be described in a way much similar to first-step propagation errors, with β being adjusted so as to describe the position of the point where concentrations are to be found. Because, now,

interpolation is exact at the nodes, restitution errors represent the difference between the accuracy of nodal and non-nodal points, at a given time.

3.2. Propagation errors

3.2.1. Individual components, first time-step

Using the BMC general algorithm, nodal concentrations after the first time-step can be expressed as a function of the nodal values of the initial conditions, in the form

$$\tilde{c}_m(j,1) = \sum_{p=P_1}^{P_2} \phi_p(\alpha) \cdot c_m(j-k-p,0) \quad (20)$$

Substituting Equation 5 into the RHS, and rearranging so as to express $\tilde{c}_m(j,1)$ as a function of $\bar{c}_m(j,1)$, we obtain

$$\tilde{c}_m(j,1) = G_m(\alpha) \cdot \bar{c}_m(j,1) \quad (21)$$

where $G_m(\alpha)$ is given by

$$G_m(\alpha) = \exp\{i\alpha\lambda_m\} \cdot \sum_{p=P_1}^{P_2} \phi_p(\alpha) \cdot \exp\{-ip\lambda_m\} = \exp\{i\alpha\lambda_m\} \cdot g_m(\alpha) \quad (22)$$

$G_m(\alpha)$ will typically be a complex function, and, therefore, the BMC will introduce errors in both the amplitude and the phase of the m^{th}

Fourier component, described, respectively, by

$$|G_m(\alpha)| = |g_m(\alpha)| = \left\{ R_m(\alpha)^2 + I_m(\alpha)^2 \right\}^{1/2} \quad (23)$$

and

$$\arg\{G_m(\alpha)\} = \alpha\lambda_m + \arg\{g_m(\alpha)\} = \alpha\lambda_m + \arctg \left\{ \frac{I_m(\alpha)}{R_m(\alpha)} \right\} \quad (24)$$

with

$$R_m(\alpha) = \operatorname{Re} \{g_m(\alpha)\} = \sum_{p=P_1}^{P_2} \phi_p(\alpha) \cdot \cos(p\lambda_m) \quad (25)$$

$$I_m(\alpha) = \operatorname{Im} \{g_m(\alpha)\} = - \sum_{p=P_1}^{P_2} \phi_p(\alpha) \cdot \sin(p\lambda_m) \quad (26)$$

For interpolation functions with a linear core element, the description of the position of the feet of the characteristic lines is independent of the particular node under consideration ($\alpha \equiv \beta - \operatorname{int}(\beta)$), and, therefore, amplitude and phase errors are equal for all nodes, as required for the numerical solution to remain a single exponential wave with the same wavelength of the initial conditions.

However, when quadratic core elements are used, $\alpha = \beta - \operatorname{int}(\beta) + \xi$, where ξ depends on the type of node associated with the head of the characteristic line; as a consequence, amplitude and phase errors evaluated at, say, the corner nodes will be all equal, but will be different from those associated with middle nodes. Hence, the numerical solution can not be represented by a single exponential wave, a non-linear mechanism of

energy transfer being generated. We note, however, that concentrations at all nodes of a same type still fit a single wave, with the original wavelength.

3.2.2. Individual components, after N time steps

To evaluate propagation errors after N time steps, we will, based on the results of the previous section, distinguish between linear and quadratic core elements.

We consider first the case of linear core elements, and express $\tilde{c}_m(j,n)$ as an appropriate combination of $\bar{c}_m(\gamma,0)$, where γ represent the nodes which concentration at time zero contribute to the concentration of node j at time n, and where $\bar{c}_m(\gamma,0)$ are given by Equation 12. Rearrangement leads to (Appendix A, Section 1),

$$\tilde{c}_m(j,n) = \left\{ G_m(\alpha) \right\}^n \cdot \bar{c}_m(j,n) = H_m(\alpha,n) \cdot \bar{c}_m(j,n) \quad (28)$$

which is clearly consistent with the simple conventional error model described by Equation 18.

For quadratic core elements, we follow the same general idea of expressing $\tilde{c}_m(j,n)$ as a function of appropriate $\bar{c}_m(\gamma,0)$, rearranging to express the latter in terms of $\bar{c}_m(j,n)$, and compare this with $\tilde{c}_m(j,n)$. The procedure is now, however, more cumbersome; it not only has to be taken individually for corner and middle nodes, but have to account for the interaction between the errors of these two types of nodes. An illustration of the procedure for the case of a quadratic Lagrange

interpolator is shown in Appendix A, Section 2, together with the rational for its extension to interpolators using information from more than the three nodes of the core element. Only the result is summarized here

$$\tilde{c}_m(j,n) \approx H_m^v(\alpha,n) \cdot \bar{c}_m(j,n) \quad (29)$$

where v denotes the type of corner node (0 for extremity nodes, ± 1 for middle nodes), and

$$H_m^v(\alpha,n) = \exp\{i\alpha\pi\lambda_m\} \cdot \exp\{i\nu\lambda_m\} \cdot \left\{ p_m^v(\alpha,n) \cdot g_m(\alpha) + q_m^v(\alpha,n) \cdot g_m(\alpha+\delta) \right\} \quad (30)$$

$$p_m^v(\alpha,n) = p_m^v(\alpha,n-1) \cdot s_m(\alpha) + q_m^v(\alpha,n-1) \cdot s_m(\alpha) \quad (31)$$

$$q_m^v(\alpha,n) = p_m^v(\alpha,n-1) \cdot r_m(\alpha) + q_m^v(\alpha,n-1) \cdot r_m(\alpha) \quad (32)$$

$$p_m^v(\alpha,1) = 1 - |v| \quad (33)$$

$$q_m^v(\alpha,1) = |v|$$

Functions $r_m(\alpha)$ and $s_m(\alpha)$ depend on the number actually used by the interpolator, and are summarized in Table 1. In general, $H_m^0(\alpha,n)$ and $H_m^{\pm 1}(\alpha,n)$ will be different from each other, and the ratio $H_m^v(\alpha,n)/H_m^v(\alpha,n-1)$ will not be a constant.

In the derivation of Equations 28 and 30 it is implicitly assumed that no aliasing occurs within any time-step of the numerical procedure. This is strictly true only in the case of linear core elements, and, therefore,

Equation 30 has to be considered an approximation, which will become exact in the absence of interval aliasing.

3.2.3. Complete solution after n time steps

The complete solution of the advection problem is, as indicated earlier, always of the form of Equation 13. To evaluate the coefficient $E_m(n)$ we have, however, to distinguish again between linear and higher-order core elements.

For linear core elements, $E_m(n)$ is simply described by Equations 16 and 20. For quadratic core elements, however, $E_m(n)$ represents the m^{th} Fourier component of the discrete signal

$$\tilde{c}(j,n) \approx \left\{ \sum_{|v|=0}^1 \nabla_v H_m^v(j,n) \right\} \bar{c}(j,n) \quad (36)$$

where

$$\nabla_v = \begin{cases} 1 & \text{if } j \text{ is of type } v \\ 0 & \text{otherwise} \end{cases} \quad (37)$$

4. VERIFICATION AND DISCUSSION OF ERROR FORMULAE

4.1. Comparison with numerical experimentation

Let us first consider the solution of the standard test problem of constant advection of a Gauss-hill, which is governed by Equation 1, with

initial and boundary conditions

$$c(x, t) \rightarrow 0 \quad \text{as } |x| \rightarrow \infty \quad (38)$$

$$c(x, 0) = \frac{1}{\sqrt{2\pi} \sigma} \exp \left\{ -\frac{(x-x_0)^2}{2\sigma^2} \right\} \quad (39)$$

In particular, we set $u=0.5$, $\sigma=264$, $x_0=2000$, and $x \in [0, 13600]$, $t \in [0, 9600]$ (units are irrelevant, as long as consistent). The problem was solved (a) numerically and (b) by Fourier analysing the initial conditions, and propagating the resulting components according to the error formulae derived in the previous section. The 2P-LI2, 3P-LI3 and 5P-HL3 interpolators—see [B6] for definitions—were used as a reference (the last two of these have quadratic core elements).

Numerical and simulated solutions are visually undistinguishable (Figure 3), which illustrates the validity of the generalized Fourier analysis. We note that simulated solutions are in general much cheaper to generate, especially for large numbers of time steps, N (typically one order of magnitude less CPU time, for interpolators with quadratic core elements, if $N > 1000$), which suggests that the error formulae of the previous section can be used, besides as a reference for formal analysis of accuracy and stability, as a cost-saving procedure for numerical experimentation.

4.2. A brief look at the mechanism of energy transfer

The energy transfer between Fourier components, due to the presence of grid nodes of different types, has constituted the motivation for the development of the present generalized Fourier analysis. This mechanism is now recognized, using the derived error formulae, to work towards the uniformization of the errors of the different types of nodes. Indeed, after some number of time steps, which will vary with the interpolator and the location of the foot of the characteristic line within the core element, errors per time step associated with corner and middle nodes become indistinguishable from each other; an exception occurs for $L_m/\Delta x = 4\Delta x$ (i.e., twice the size of the element), where a resonant behavior occurs (e.g., Figure 4). Examination of actual amplification factors for $L_m/\Delta x = 4\Delta x$ suggests that this resonance should be without practical importance, as the energy wavelength is typically quickly dissipated (Figure 5).

5. CONCLUSIONS

This paper provides a reference tool for the formal analysis of the accuracy and stability of the BMC, for any interpolation function of class C_0 . This tool is extensively used by [B5] to study the stability of the BMC, and by [B6] to compare the accuracy of different interpolators.

Some of the results obtained in our present analysis, although derived in the specific context of BMC solutions in uniform grids, can be extrapolated to other situations, and may provide some new perspectives to familiar problems.

In particular, it should be clear that nodes of a different type are not an exclusive property of some BMC. For instance, any FE method based on quadratic and higher-order interpolations will also have different nodal types, and, therefore, will potentially induce the same type of energy transfer identified for the BMC. This may be an unexplored difference between FE and FD methods, and certainly invites further analysis as well as the cautionary revision of previous conclusions on accuracy and stability of FE methods, if based on an oversimplified application of Fourier analysis.

Also, the fact that some BMC induce, for uniform grids, nonlinearities much similar to those due to grid or flow non-uniformity, but that are amenable to systematic quantitative analysis, may be used to provide further insight on the effects of grid and flow non-uniformity.

REFERENCES

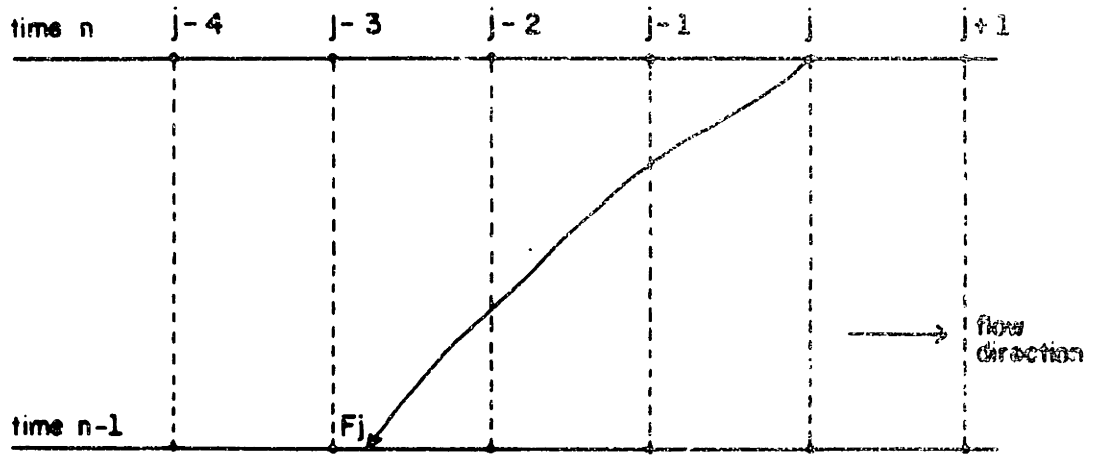
- B1 Baptista, A. M., E. E. Adams, and K. D. Stolzenbach. "Eulerian-Lagrangian Analysis of Pollutant Transport in Shallow Water." MIT R. M. Parsons Laboratory, Technical Report 296, 1984.
- B2 Baptista, A. M., E. E. Adams, and K. D. Stolzenbach. "Comparison of Several Eulerian-Lagrangian Models to Solve the Advection-Diffusion Equation." in *Int. Symp. on Refined Flow Modeling and Turbulence Measurements*. U. Iowa, USA, 1985.
- B3 Baptista, A. M., E. E. Adams, and K. D. Stolzenbach. "Accuracy Analysis of the Backwards Method of Characteristics." in *Finite Elements in Water Resources: Proceedings of the 6th International Conference*. (Sá da Costa et al., Ed.), Springer-Verlag, 1986.
- B4 Baptista, A. M. "Accurate Numerical Modeling of Advection-Dominated Transport of Passive Scalars." LNEC, Lisboa, 1986.
- B5 Baptista, A. M. "The consistency, stability and convergence of the Backwards Method of Characteristics." Ph.D. Thesis, M.I.T., 1987.
- B6 Baptista, A. M. "The Choice of the Interpolator for the Backwards Method of Characteristics." Ph.D. Thesis, M.I.T., 1987.
- B7 Baptista, A. M. "The Accuracy of Eulerian-Lagrangian Methods" Ph.D. Thesis, M.I.T., 1987.
- O1 Oppenheim and Schafer - "Digital Signal Processing". 1975
- R1 Roache, P. *Computational Fluid Dynamics*. Hermosa Publ., 1982.

Figure 1

Illustrative sketch for the Backwards Method of Characteristics

Required steps:

1. Tracking of the characteristic lines. For each node j , a characteristic line is independently defined by the backwards (i.e., between n and $n-1$) solution of an ordinary differential solution of the form $dx_1/dt = u_1$.
2. Interpolation at the feet of the characteristic line. The concentration at the foot of each characteristic line (and, for pure advection, at any other point of the characteristic line, including node j at time n) is found by interpolation from known information on neighboring nodes, at time $n-1$.



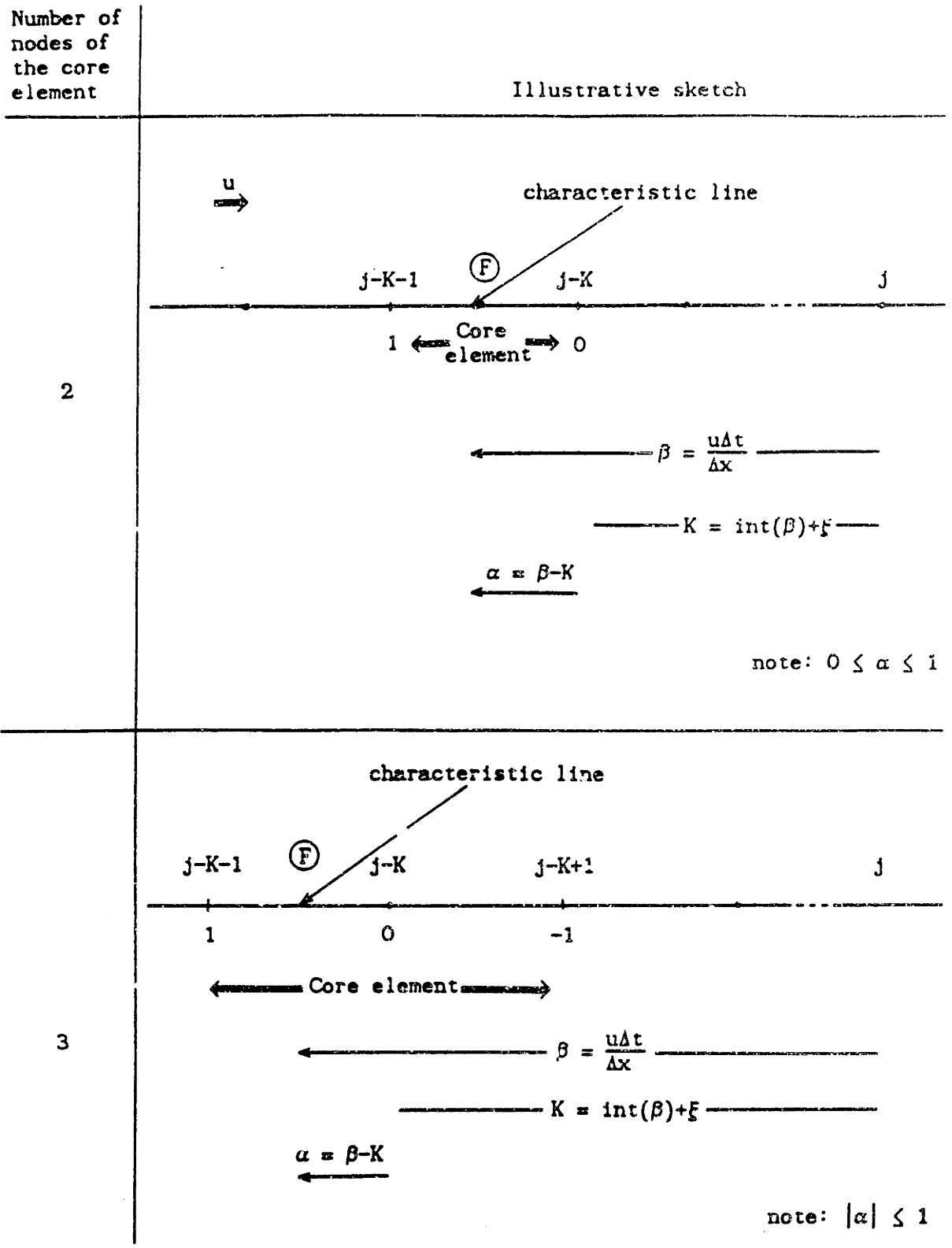


Figure 2 Definition of the core elements

Figure 3

Comparison of the actual numerical solution with the solution obtained
by using error formulae

Reference problem: $\frac{\partial c}{\partial t} + u \frac{\partial c}{\partial x} = 0$
 $c(x, 0) = \exp\left\{-\frac{(x-x_0)^2}{2\sigma^2}\right\}$
 $c(x, t) \rightarrow 0 \quad |x| \rightarrow \infty$

Computational parameters: $\Delta t = 96$

$$\Delta x = 200$$

$$N = T/\Delta t = 100$$

$$v = 0.5$$

$$\sigma = 2.64$$

Legend: 1 - Exact solution

2 - 2P-LI2 (numerical solution)

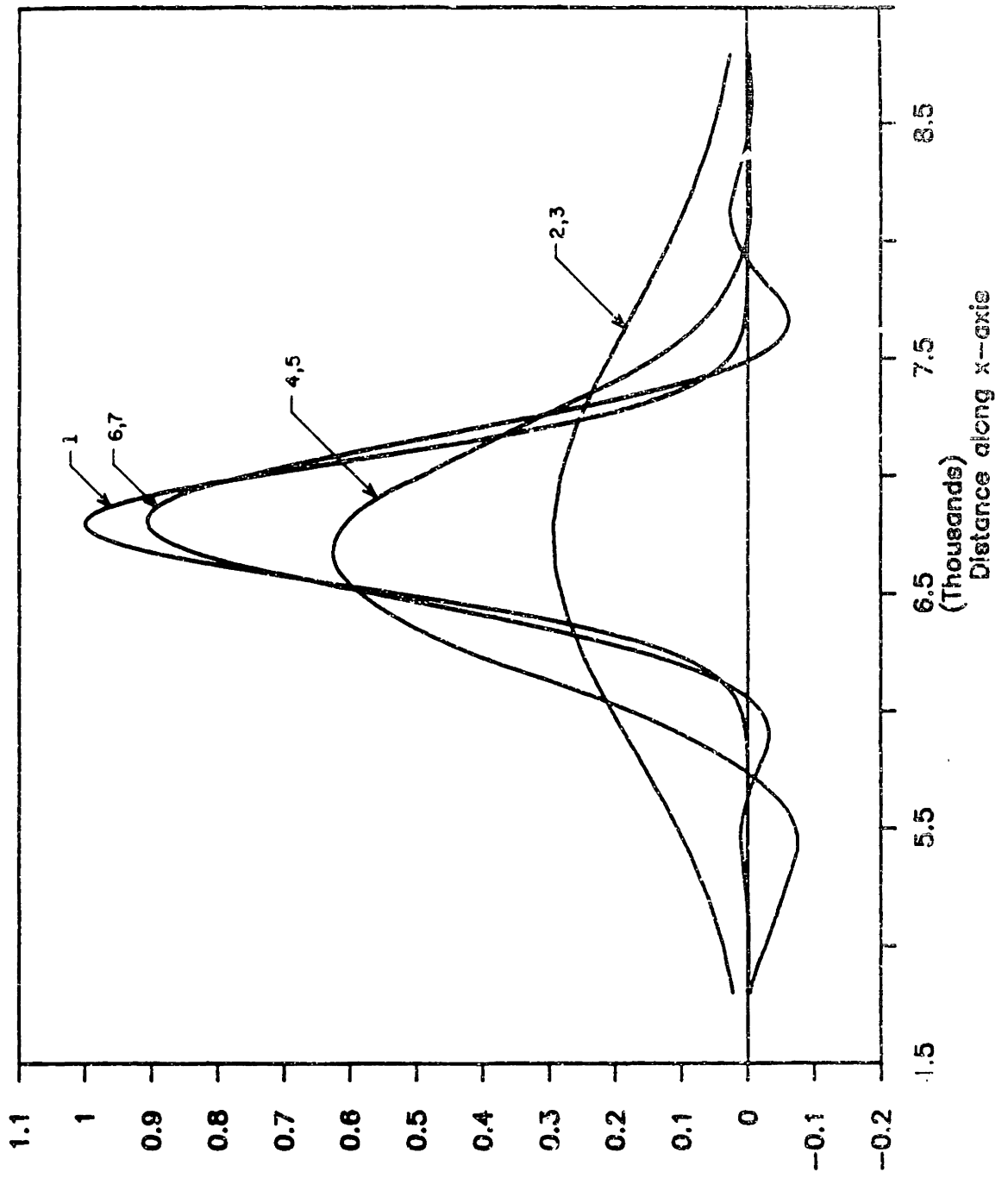
3 - 2P-LI2 (error formulae)

4 - 3P-LI3 (numerical solution)

5 - 3P-LI3 (error formulae)

6 - 5P-HL3 (numerical solution)

7 - 5P-HL3 (error formulae)



Concentration

Figure 4

Amplification factors per time step, as a function of the number of the time steps, for interpolators with quadratic core elements ($\alpha = 0.5$ for middle nodes; $\alpha = -0.5$ for corner nodes)

(a) 3P-LI3, $L_m/\Delta x = 3$

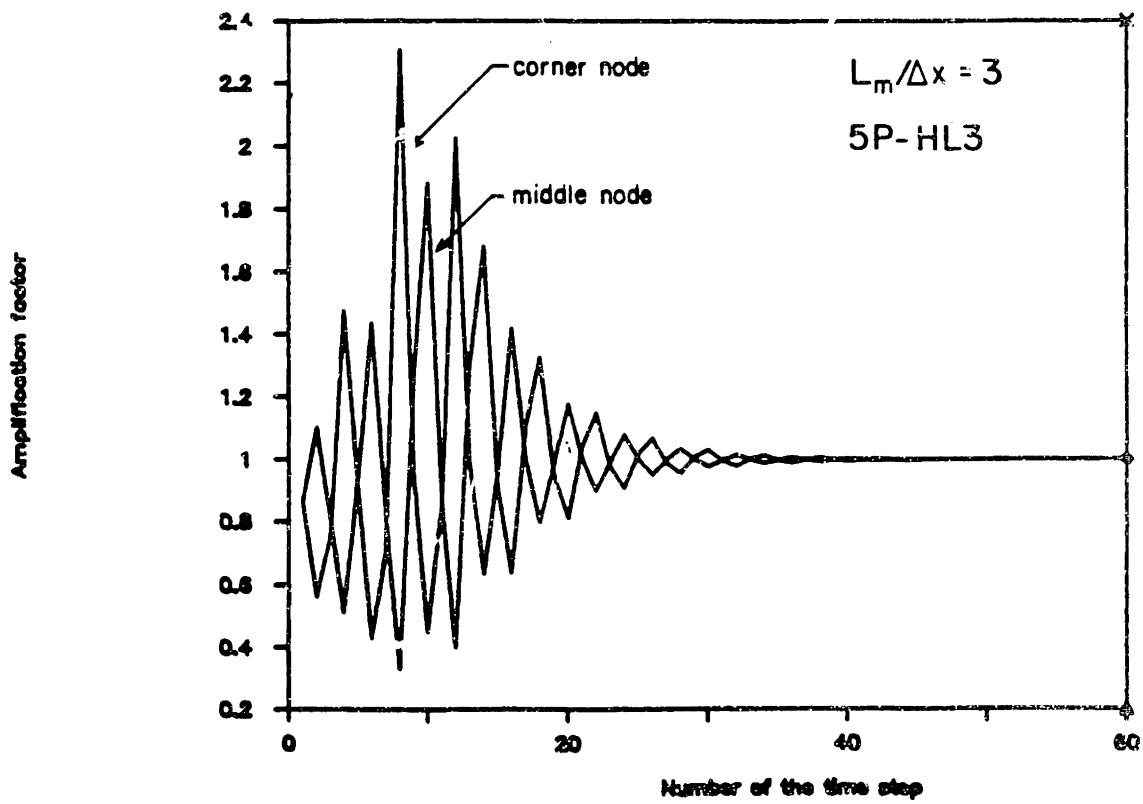
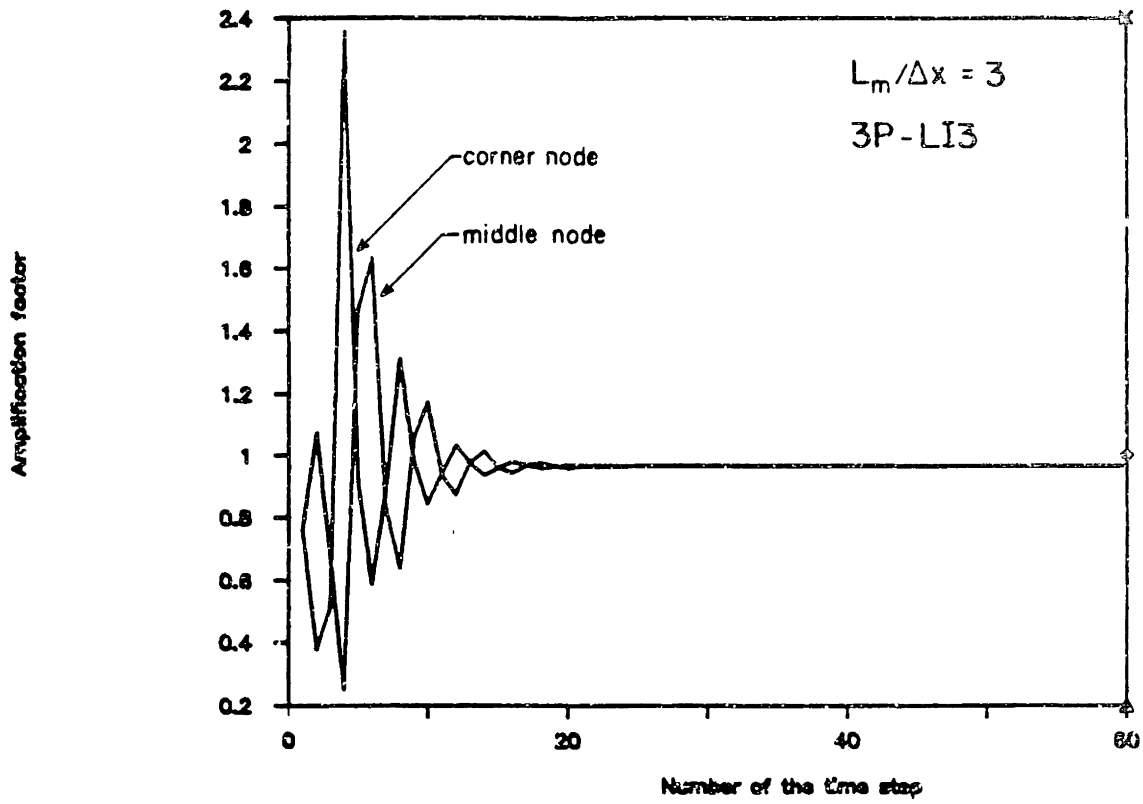
(b) 5P-HL3, $L_m/\Delta x = 3$

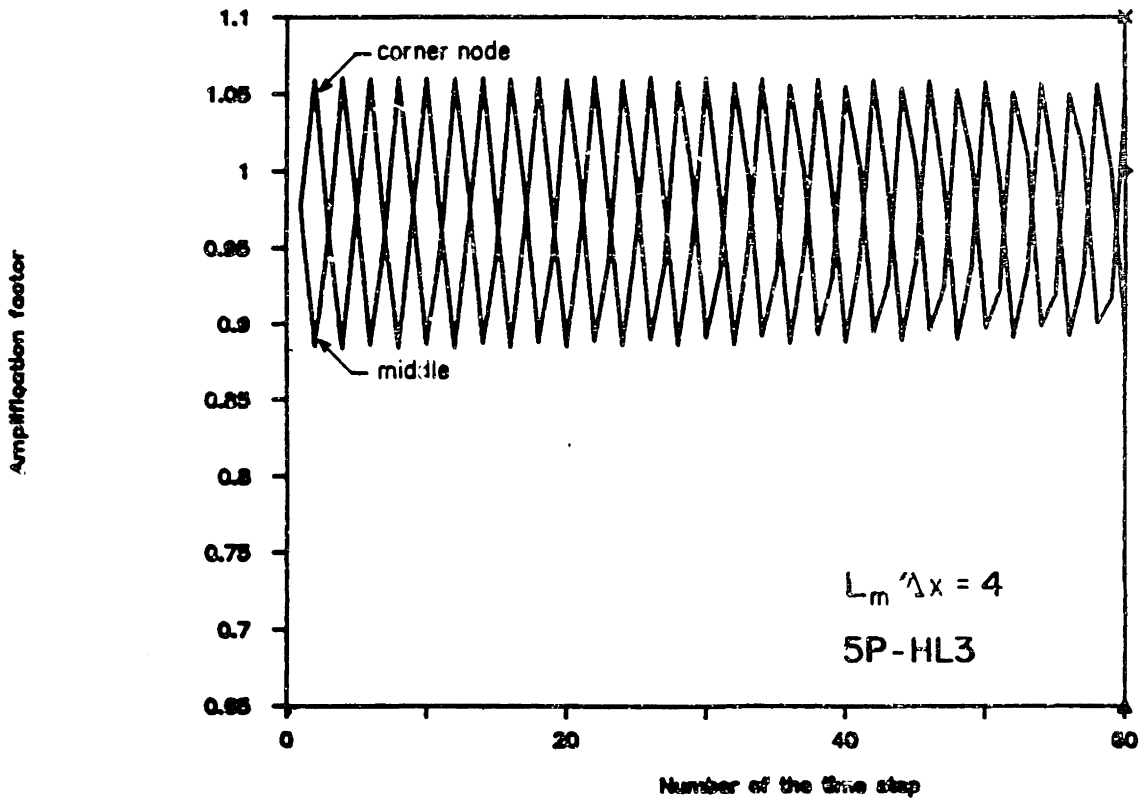
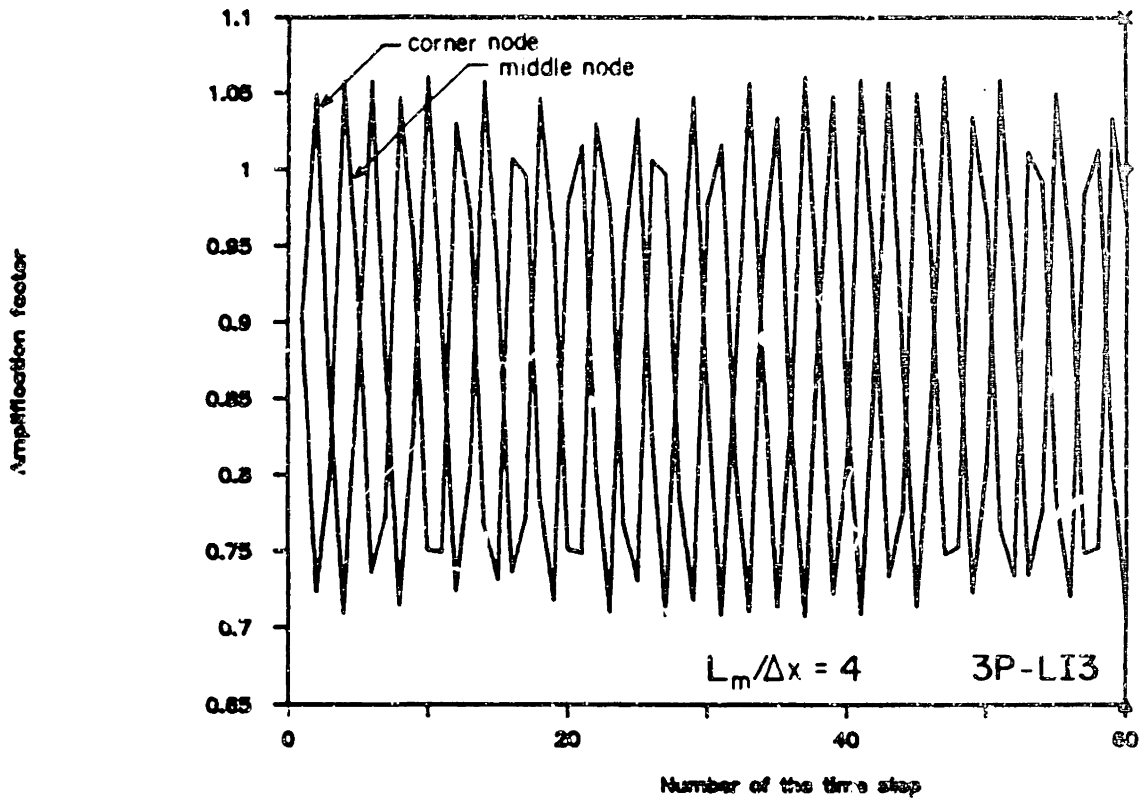
(c) 3P-LI3, $L_m/\Delta x = 4$

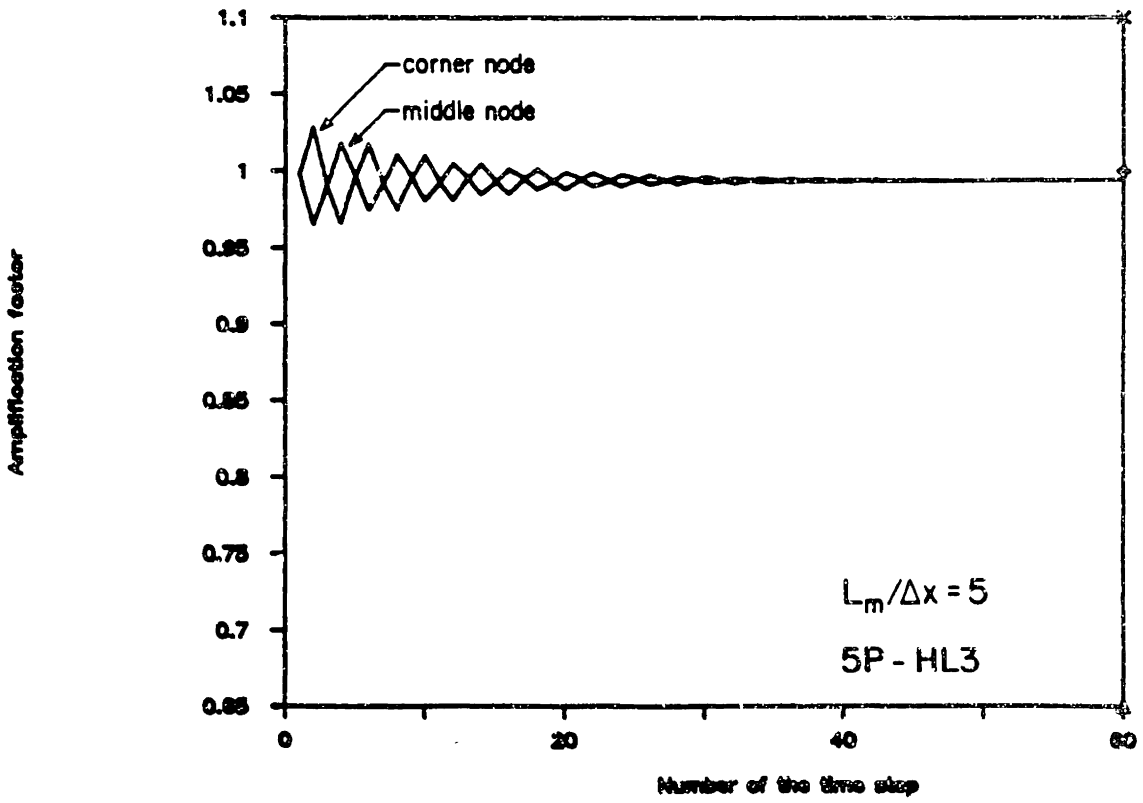
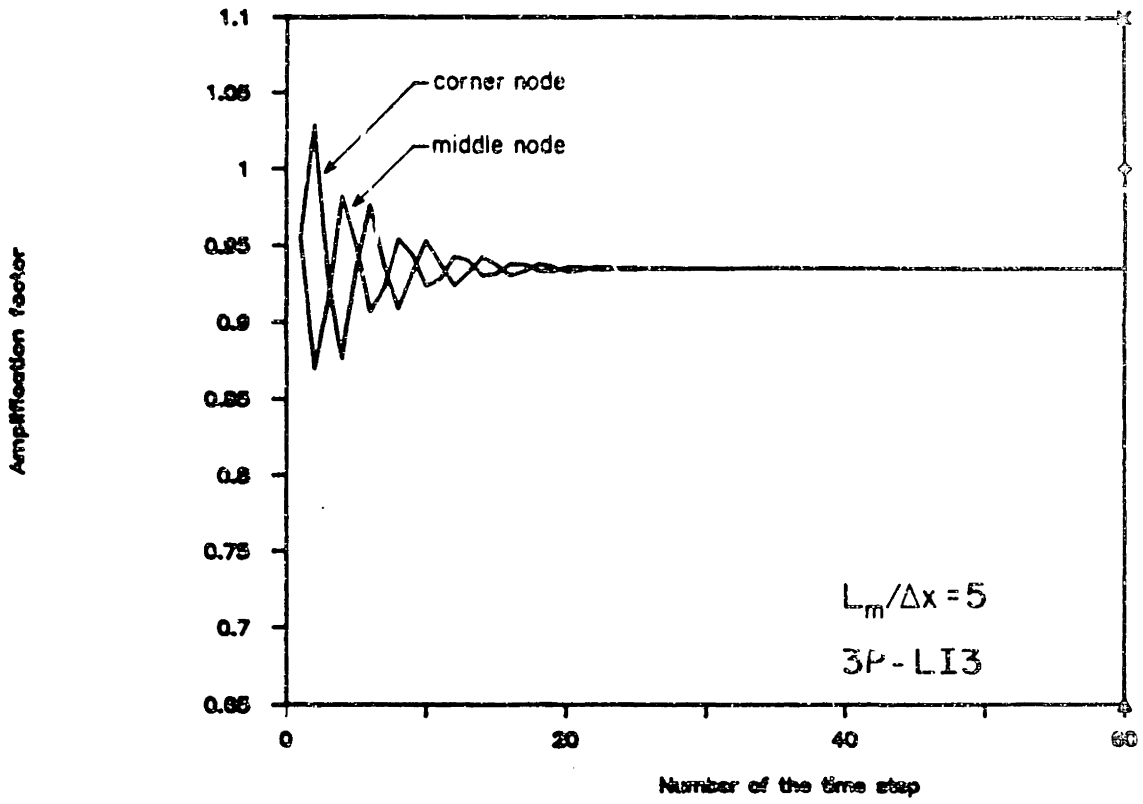
(d) 5P-HL3, $L_m/\Delta x = 4$

(e) 3P-LI3, $L_m/\Delta x = 5$

(f) 5P-HL3, $L_m/\Delta x = 5$







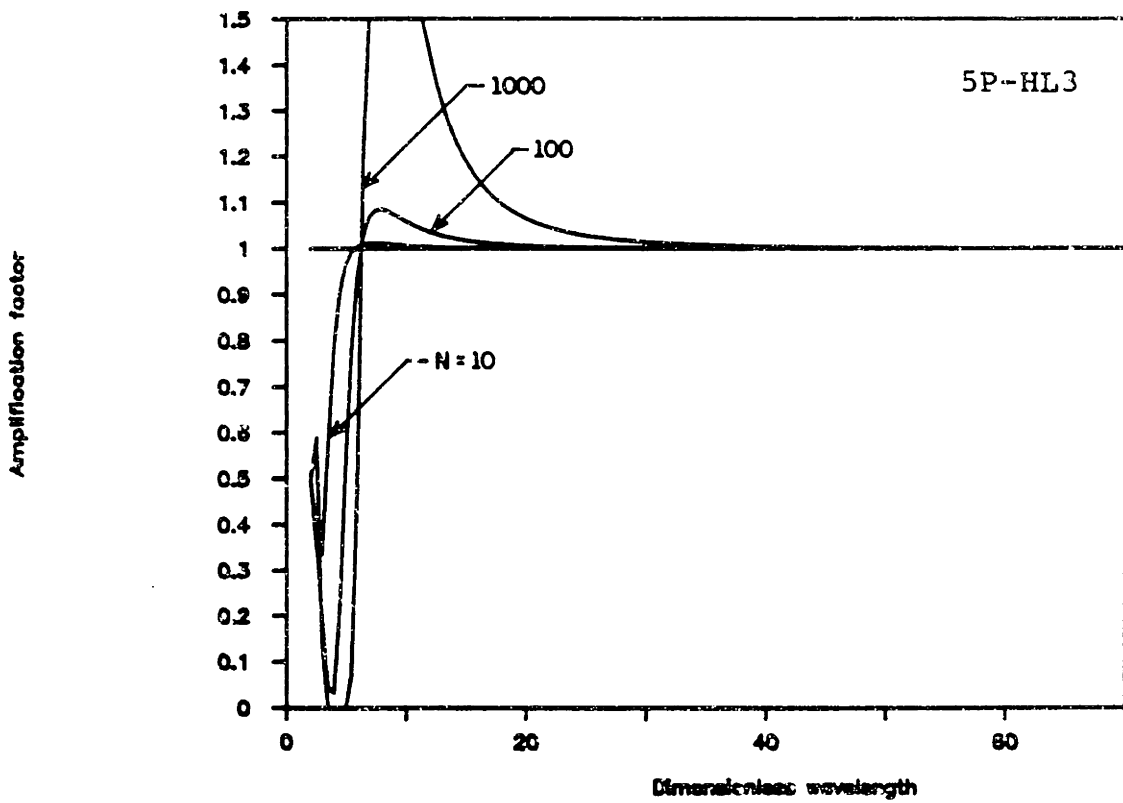
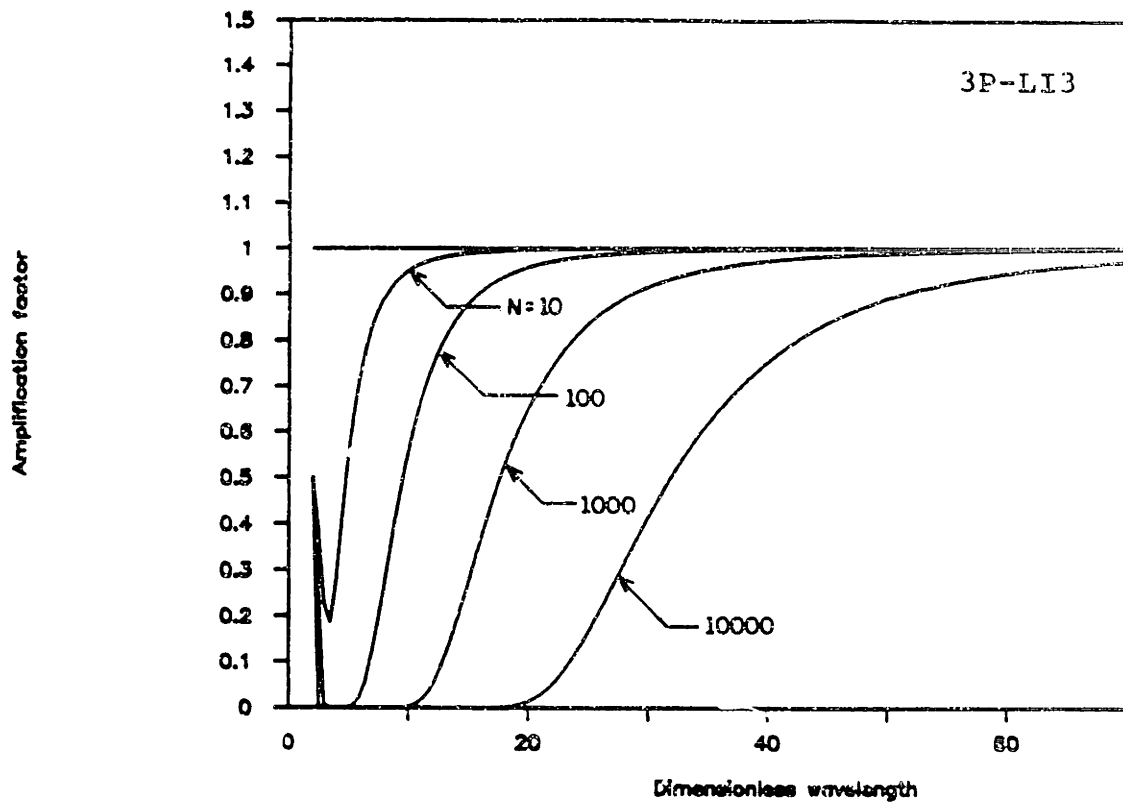


Figure 5 Amplification factors after N time steps, as a function of the dimensionless wavelength

Table 1

Definition of the Functions $r_m(\alpha)$ and $s_m(\alpha)$

B.1 Compact interpolators

$$r_m(\alpha) = \phi_0(\alpha) \exp(\pm i\lambda_m)$$

$$s_m(\alpha) = \phi_{-1}(\alpha) \exp(i\lambda_m) + \phi_1(\alpha) \exp(-i\lambda_m)$$

B.2 Non-compact interpolators (5 nodes)

$$r_m(\alpha) = \begin{cases} [\phi_2(\alpha) + \phi_0(\alpha)] \exp(i\lambda_m) + \phi_{-2}(\alpha) \exp(13\lambda_m) \\ \phi_{-2}(\alpha) \exp(i\lambda_m) + \phi_0(\alpha) \exp(-i\lambda_m) + \phi_{-2}(\alpha) \exp(-13\lambda_m) \end{cases}$$

$$s_m(\alpha) = \phi_{-1}(\alpha) \exp(i\lambda_m) + \phi_1(\alpha) \exp(-i\lambda_m)$$

B.3 Non-compact interpolators (7 nodes)

$$r_m(\alpha) = \begin{cases} \phi_{-2}(\alpha) \exp(13\lambda_m) + [\phi_0(\alpha) + \phi_2(\alpha)] \exp(i\lambda_m) \\ \phi_{-2}(\alpha) \exp(13\lambda_m) + \phi_0(\alpha) \exp(i\lambda_m) + \phi_2(\alpha) \exp(-i\lambda_m) \end{cases}$$

$$s_m(\alpha) = [\phi_{-3}(\alpha) + \phi_3(\alpha)] \exp(13\lambda_m) + [\phi_{-1}(\alpha) + \phi_1(\alpha)] \exp(i\lambda_m)$$

APPENDIX A

Derivation of formulae for propagation errors after n time steps

A.1. - The case of linear core elements

We want to express $\tilde{c}_m(j,n)$ as a function of $\bar{c}_m(j,n)$, i.e.,

$$\tilde{c}_m(j,n) = H_m(\alpha,n) \bar{c}_m(j,n) \quad (\text{A.1})$$

where $H_m(\alpha,n)$ is to be evaluated.

We first use the general expression for the BMC algorithm, Equation 3, to write successively

$$\begin{aligned} \tilde{c}_m(j,n) &= \sum_{p_1=P_1}^{P_2} \phi_{p_1}(\alpha) \bar{c}_m(j-K-p_1, n-1) = \\ &= \sum_{p_1=P_1}^{P_2} \phi_{p_1}(\alpha) \left\{ \sum_{p_2=P_1}^{P_2} \phi_{p_2}(\alpha) \bar{c}_m(j-2K-p_1-p_2) \right\} = \dots = \\ &= \sum_{p_1=P_1}^{P_2} \phi_{p_1}(\alpha) \left\{ \sum_{p_2=P_1}^{P_2} \phi_{p_2}(\alpha) \left\{ \dots \left\{ \sum_{p_n=P_1}^{P_2} \phi_{p_n}(\alpha) \bar{c}_m(j-nK - \sum_{i=1}^n p_i) \right\} \right\} \right\} \end{aligned} \quad (\text{A.2})$$

Now, using the fact that $\alpha = \beta - K$, and expressing \bar{c}_m in Fourier series form, we recognize that

$$\begin{aligned}
 \bar{c}_m \left[j - nK - \sum_{i=1}^n p_i \right] &= B_m \exp \left\{ i\lambda_m \left[j - nK - \sum_{i=1}^n p_i \right] \right\} = \\
 &= B_m \exp \{ i\lambda_m [j - n\beta] \} \exp \left\{ i\lambda_m \left[n\alpha - \sum_{i=1}^n p_i \right] \right\} = \\
 &= \bar{c}_m(j, n) \exp \left\{ i\lambda_m \left[n\alpha - \sum_{i=1}^n p_i \right] \right\} \tag{A.3}
 \end{aligned}$$

Replacing Equation A.3 into Equation A.2, and rearranging, we recover Equation A.1, with

$$\begin{aligned}
 H_m(\alpha, n) &= \sum_{p_1=P_1}^{P_2} \phi_{p_1}(\alpha) \left\{ \sum_{p_2=P_1}^{P_2} \phi_{p_2}(\alpha) \left\{ \dots \left\{ \sum_{p_n=P_1}^{P_2} \phi_{p_n}(\alpha) \cdot \right. \right. \right. \\
 &\quad \left. \left. \left. \cdot \exp \left[i\lambda_m \left[n\alpha - \sum_{i=1}^n p_i \right] \right] \right\} \right\} = \\
 &= G_m(\alpha) \sum_{p_1=P_1}^{P_2} \phi_{p_1}(\alpha) \left\{ \sum_{p_2=P_1}^{P_2} \phi_{p_2}(\alpha) \left\{ \dots \left\{ \sum_{p_{n-1}=P_1}^{P_2} \phi_{p_{n-1}}(\alpha) \cdot \right. \right. \right. \\
 &\quad \left. \left. \left. \cdot \exp \left[i\lambda_m \left[(n-1)\alpha - \sum_{i=1}^{n-1} p_i \right] \right] \right\} \right\} =
 \end{aligned}$$

$$\begin{aligned}
&= \{G_m(\alpha)\}^2 \sum_{p_1=P_1}^{P_2} \phi_{p_1}(\alpha) \left\{ \sum_{p_2=P_1}^{P_2} \phi_{p_2}(\alpha) \left\{ \dots \left\{ \sum_{p_{n-2}=P_1}^{P_2} \phi_{p_{n-2}}(\alpha) \cdot \right. \right. \right. \\
&\quad \left. \left. \left. \cdot \exp \left[i d_m \left[(n-2)\alpha - \sum_{i=1}^{n-2} p_i \right] \right] \right\} \right\} = \\
&= \dots = \{G_m(\alpha)\}^n \tag{A.4}
\end{aligned}$$

where $G_m(\alpha)$ denotes the error in the first time step, and is described by Equation 22. Hence, errors are equal at every time step.

A.2. The case of quadratic core elements

Let us take as a reference the 3P-LI3 interpolator, and the conditions of Figure A.1a, i.e., j is a corner node and $\text{int}(\beta)$ is even.

If the numerical solution had started at time $n-1$, the relationship between $\tilde{c}_m(j,n)$ and $\bar{c}_m(j,n)$ could be found by expressing the former as a function of $\bar{c}_m(\gamma_1, n-1)$ —where γ represents the nodes that contribute to the interpolation at the feet of the characteristic line that starts at (j,n) —and rearranging appropriately. Indeed,

$$\begin{aligned}
\tilde{c}_m(j,n) &= \sum_{p=P_1}^{P_2} \phi_p(\alpha) \bar{c}_m[j-(K+1)-p] = \phi_{-1}(\alpha) \bar{c}(j-K, n-1) + \\
&\quad + \phi_0(\alpha) \bar{c}(j-K-1, n-1) + \phi_1(\alpha) \bar{c}(j-K-2, n-1) = \\
&= \bar{c}(j,n) \exp\{i\lambda_m(\beta-k-1)\} \{\phi_{-1}(\alpha) \exp(i\lambda_m) + \phi_0(\alpha) + \phi_1(\alpha) \exp(-i\lambda_m)\} = \\
&= \bar{c}(j,n) \exp(i\alpha\lambda_m) g_m(\alpha) \tag{A.5}
\end{aligned}$$

Now, assume that the numerical solution had started at time $n-2$. The relationship between $\tilde{c}_m(j,n)$ and $\bar{c}_m(j,n)$ can still be found by expressing $\tilde{c}_m(j,n)$ successively as a function of $\bar{c}_m(\gamma_1, n-1)$ and of $\bar{c}_m(\gamma_2, n-2)$, and rearranging appropriately. This leads to

$$\begin{aligned}
\tilde{c}(j,n) &= \phi_{-1}(\alpha)\bar{c}(j-K,n-1)+\phi_0(\alpha)\bar{c}(j-K-1,n-1)+\phi_1(\alpha)\bar{c}(j-K-2,n-1) = \\
&= \phi_{-1}(\alpha)[\bar{c}(j-2K,n-2)\phi_{-1}(\alpha)+\phi_0(\alpha)\bar{c}(j-2K-1,n-2)+\phi_1(\alpha)\bar{c}(j-2K-2,n-2)] + \\
&+ \phi_0(\alpha)[\bar{c}(j-2K,n-2)\phi_{-1}(\alpha+1)+\phi_0(\alpha+1)\bar{c}(j-2K-1,n-2)+\phi_1(\alpha)\bar{c}(j-2K-2,n-2)] + \\
&+ \phi_1(\alpha)[\bar{c}(j-2K-2,n-2)\phi_{-1}(\alpha)+\phi_0(\alpha)\bar{c}(j-2K-3,n-2)+\phi_1(\alpha)\bar{c}(j-2K-4,n-2)] = \\
&= \bar{c}(j,n)\exp(i2\alpha\lambda_m)\{g_m(\alpha)s_m(\alpha)+g_m(\alpha+1)r_m(\alpha)\} \tag{A.6}
\end{aligned}$$

with

$$\begin{aligned}
s_m(\alpha) &= \phi_{-1}\exp(i\lambda_m)+\phi_1(\alpha)\exp(-i\lambda_m) \\
r_m(\alpha) &= \phi_0(\alpha)\exp(i\lambda_m) \tag{A.7}
\end{aligned}$$

If the solution had started at time $n-3$, we would have, in turn,

$$\begin{aligned}
\tilde{c}(j,n) &= \phi_{-1}(\alpha)\bar{c}(j-K,n-1)+\phi_0(\alpha)\bar{c}(j-K,n-1)+\phi_1(\alpha)\bar{c}(j-K-2,n-1) = \\
&= \phi_{-1}(\alpha)[\phi_{-1}(\alpha)\bar{c}(j-2k,n-2)+\phi_0(\alpha)\bar{c}(j-2k-1,n-2)+\phi_1(\alpha)\bar{c}(j-2k-2,n-2)]+ \\
&\phi_0(\alpha)[\phi_{-1}(\alpha+1)\bar{c}(j-2k,n-2)+\phi_0(\alpha+1)\bar{c}(j-2k-1,n-2)+\phi_1(\alpha)\bar{c}(j-2k-2,n-2)]+ \\
&\phi_1(\alpha)[\phi_{-1}(\alpha)\bar{c}(j-2k-2,n-2)+\phi_0(\alpha)\bar{c}(j-2k-3,n-2)+\phi_1(\alpha)\bar{c}(j-2k-4,n-2)] = \\
&= \phi_{-1}(\alpha)\phi_{-1}(\alpha)[\phi_{-1}(\alpha)\bar{c}(j-3k,n-3)+\phi_0(\alpha)\bar{c}(j-3k-1,n-3)+\phi_1(\alpha)\bar{c}(j-3k-2,n-3)]+ \\
&\phi_{-1}(\alpha)\phi_0(\alpha)[\phi_{-1}(\alpha+1)\bar{c}(j-3k,n-3)+\phi_0(\alpha+1)\bar{c}(j-3k-1,n-3)+\phi_1(\alpha+1)\bar{c}(j-3k-2,n-3)]+ \\
&\phi_{-1}(\alpha)\phi_1(\alpha)[\phi_{-1}(\alpha)\bar{c}(j-3k-2,n-3)+\phi_0(\alpha)\bar{c}(j-3k-3,n-3)+\phi_1(\alpha)\bar{c}(j-3k-4,n-3)] +
\end{aligned}$$

$$\begin{aligned}
& + \phi_0(\alpha)\phi_{-1}(\alpha+1)[\phi_{-1}(\alpha)\bar{c}(j-3k, n-3)+\phi_0(\alpha)\bar{c}(j-3k-1, n-3)+\phi_1(\alpha)\bar{c}(j-3k-2, n-3)]+ \\
& \phi_0(\alpha)\phi_0(\alpha+1)[\phi_{-1}(\alpha+1)\bar{c}(j-3k, n-3)+\phi_0(\alpha+1)\bar{c}(j-3k-1, n-3)+\phi_1(\alpha+1)\bar{c}(j-3k-2, n-3)] \\
& + \\
& \phi_0(\alpha)\phi_1(\alpha+1)[\phi_{-1}(\alpha)\bar{c}(j-3k-2, n-3)+\phi_0(\alpha)\bar{c}(j-3k-3, n-3)+\phi_1(\alpha)\bar{c}(j-3k-4, n-3)] + \\
& + \phi_1(\alpha)\phi_{-1}(\alpha)[\phi_{-1}(\alpha)\bar{c}(j-3k-2, n-3)+\phi_0(\alpha)\bar{c}(j-3k-3, n-3)+\phi_1(\alpha)\bar{c}(j-3k-4, n-3)]+ \\
& \phi_1(\alpha)\phi_0(\alpha)[\phi_{-1}(\alpha+1)\bar{c}(j-3k-2, n-3)+\phi_0(\alpha+1)\bar{c}(j-3k-3, n-3)+\phi_1(\alpha)\bar{c}(j-3k-4, n-3)]+ \\
& \phi_1(\alpha)\phi_1(\alpha)[\phi_{-1}(\alpha)\bar{c}(j-3k-4, n-3)+\phi_0(\alpha)\bar{c}(j-3k-5, n-3)+\phi_1(\alpha)\bar{c}(j-3k-6, n-3)] = \\
& = c(j, n)\exp(i3\alpha\lambda_m)\{g_m(\alpha)[s_m(\alpha)s_m(\alpha)+s_m(\alpha+1)r_m(\alpha)]+ \\
& +g_m(\alpha+1)[r_m(\alpha)s_m(\alpha)+r_m(\alpha+1)r_m(\alpha)]\} \tag{A.8}
\end{aligned}$$

Following the same procedure, and assuming successively that the numerical solution started at times $n-4$, $n-5$, ... , a recurrence formula is seen to prevail, in the form

$$\tilde{c}_m(j, n) = \bar{c}_m(j, n)H_m(j, \alpha, n) \tag{A.9}$$

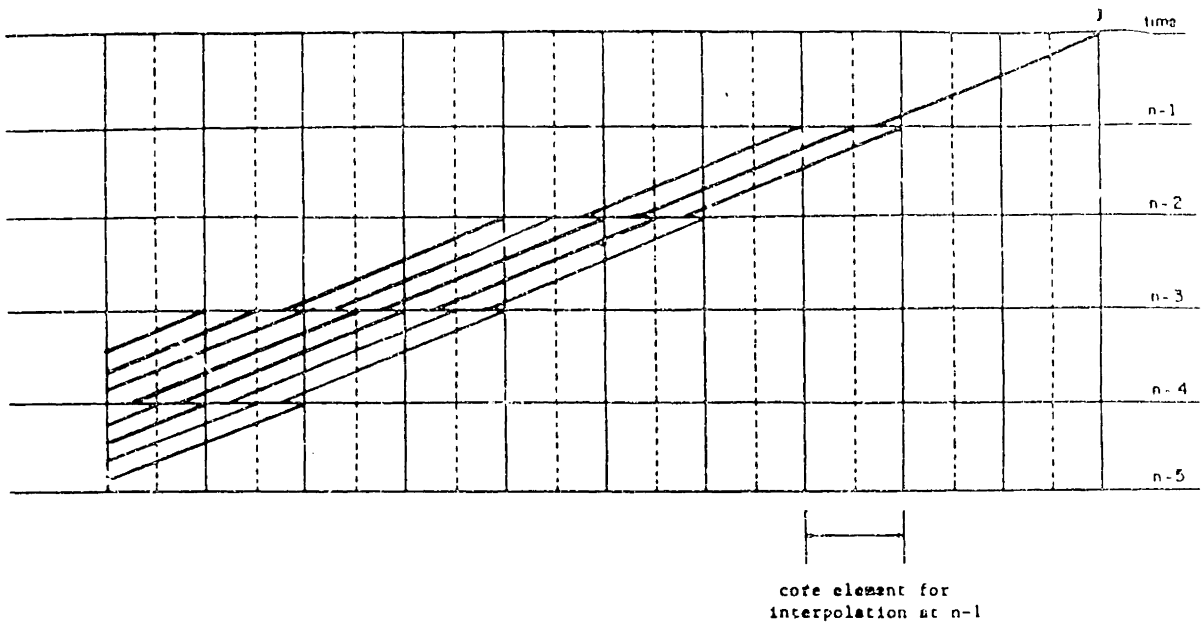
with

$$H_m(j, \alpha, N) = \{p_m(j, \alpha, N)g_m(\alpha)+q_m(j, \alpha, N)g_m(\alpha+1)\}\exp(in\alpha\lambda_m) \tag{A.10}$$

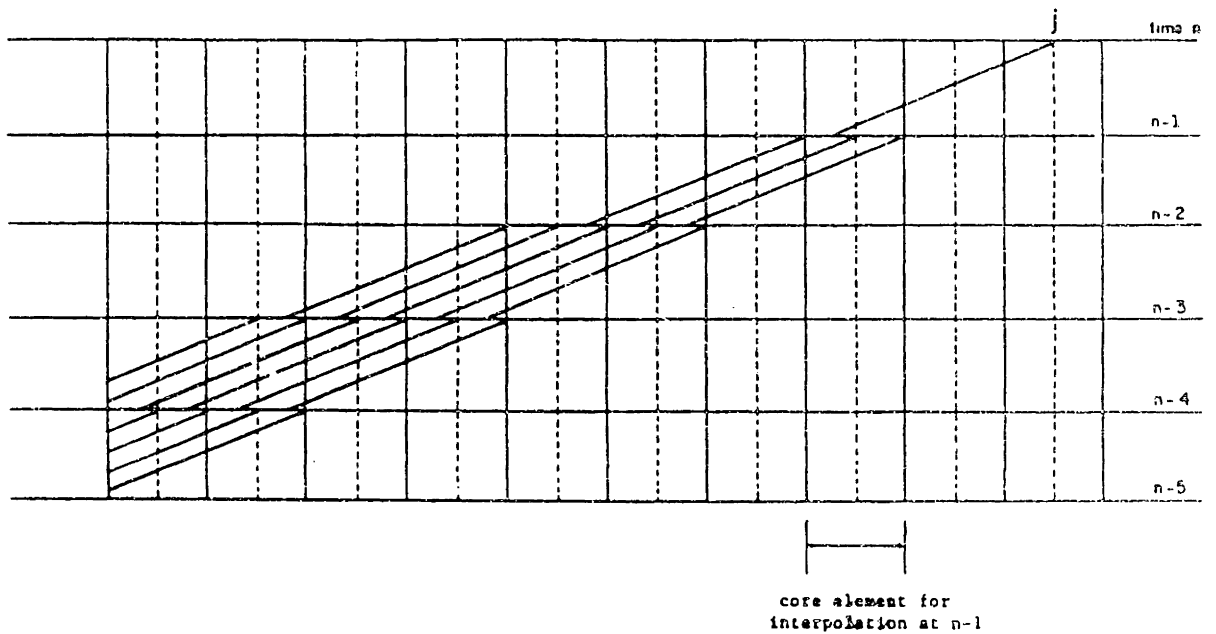
and

n	$p_m(j, \alpha, n)$	$g_m(j, \alpha, N)$
1	1	0
2	$s_m(\alpha)$	$r_m(\alpha)$
3	$s_m(\alpha)s_m(\alpha) + s_m(\alpha+1)r_m(\alpha)$	$r_m(\alpha)s_m(\alpha) + r_m(\alpha+1)r_m(\alpha)$
4	$s_m(\alpha)p_m(j, \alpha, 3) + s_m(\alpha+1)q_m(j, \alpha, 3)$	$r_m(\alpha)p_m(j, \alpha, 3) + r_m(\alpha+1)q_m(j, \alpha, 3)$
5	$s_m(\alpha)p_m(j, \alpha, 4) + s_m(\alpha+1)q_m(j, \alpha, 4)$	$r_m(\alpha)p_m(j, \alpha, 4) + r_m(\alpha+1)q_m(j, \alpha, 4)$
.		
.		
.		

Repeating the procedure for a middle node (and for $\text{int}(\beta)$ odd, for both types of nodes)--Figures A.2b-d--and for different interpolators--Figures A.2a-b--and recollecting the results (omitted here), we obtain the general recurrence formula expressed in the text by Equation 30.

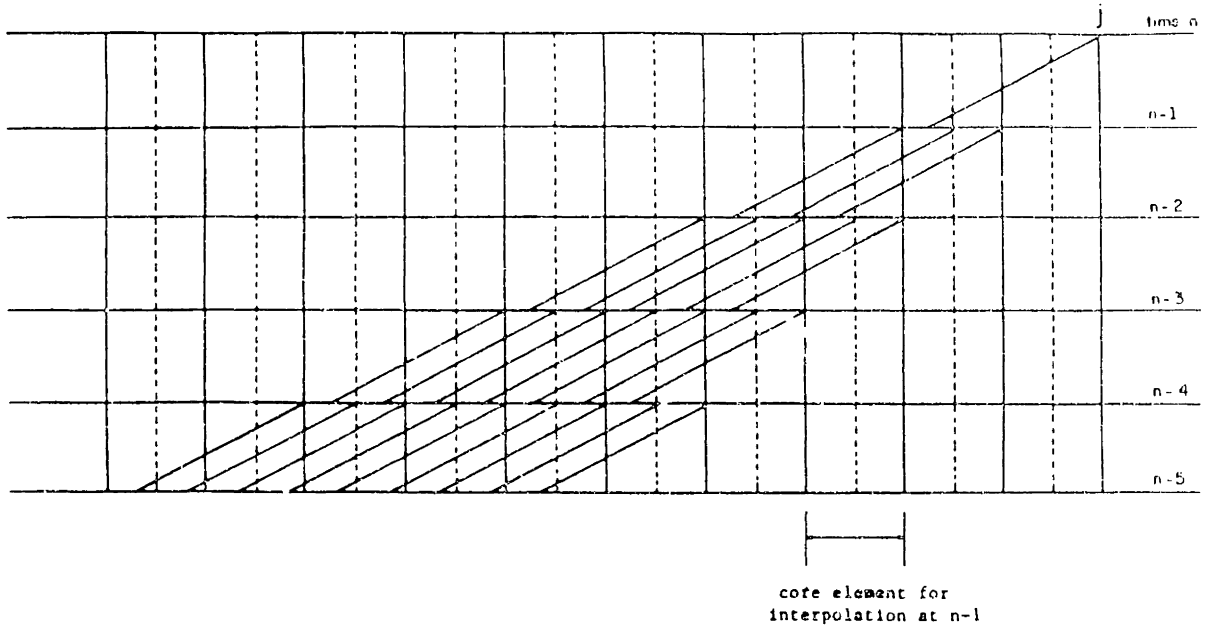


(a) Case of a corner node, with $\text{int}(Cu)$ even

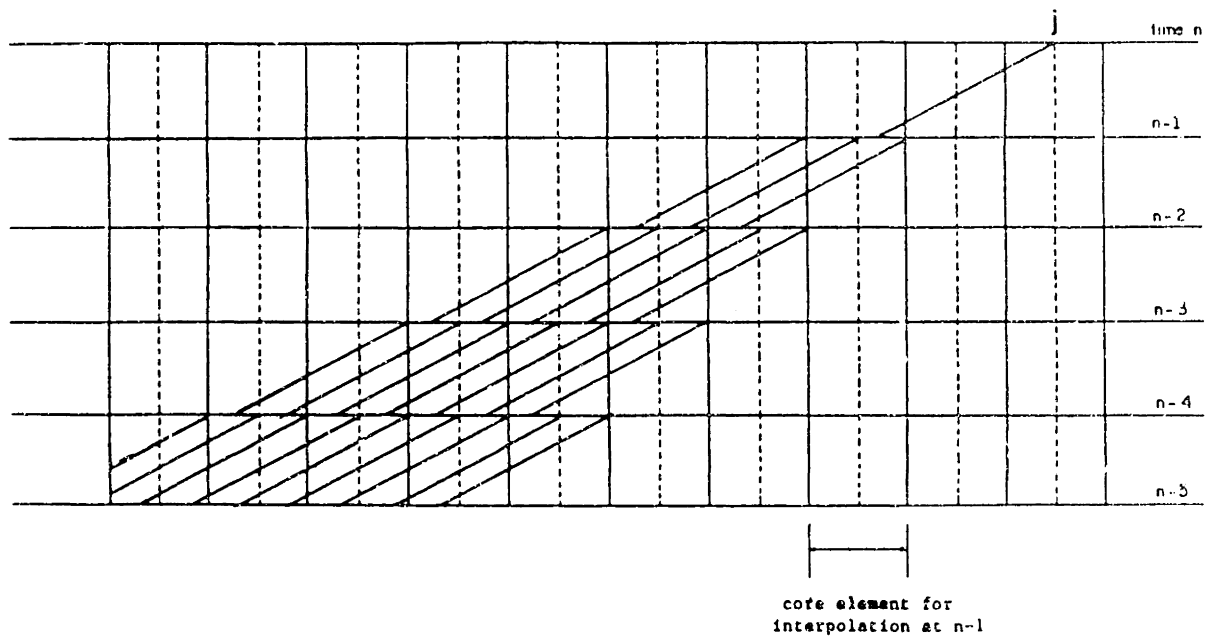


(b) Case of a middle node, with $\text{int}(Cu)$ even

Figure A1 Zone of influence of the concentration at (j, n) , for compact interpolators

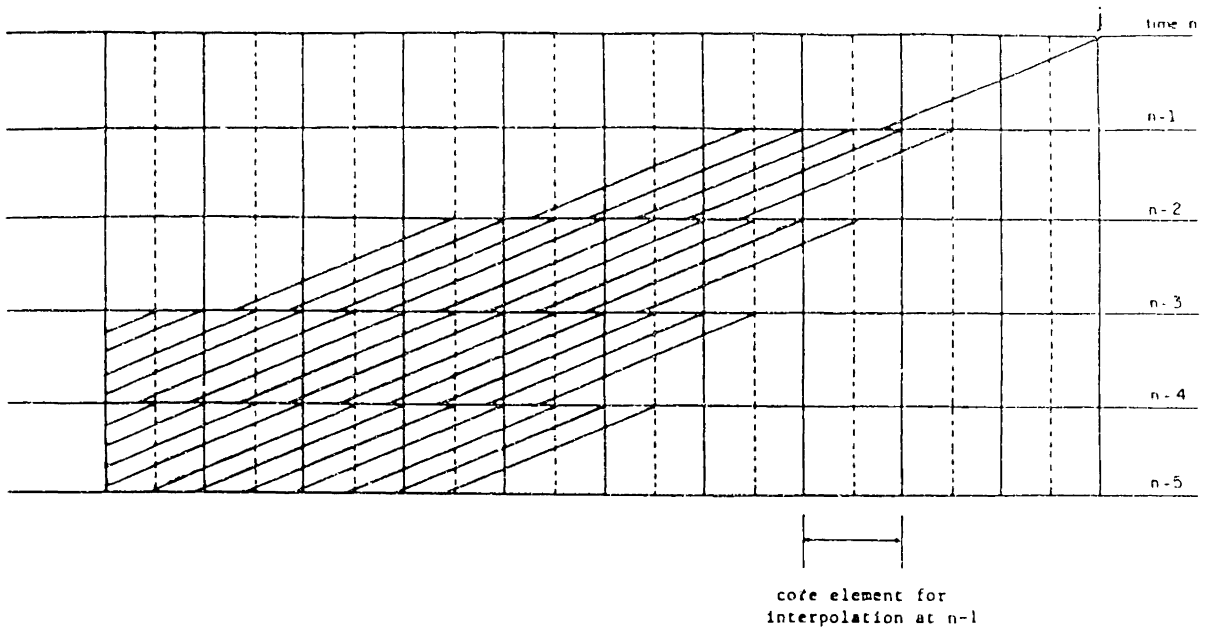


(c) Case of a corner node, with $\text{int}(Cu)$ odd

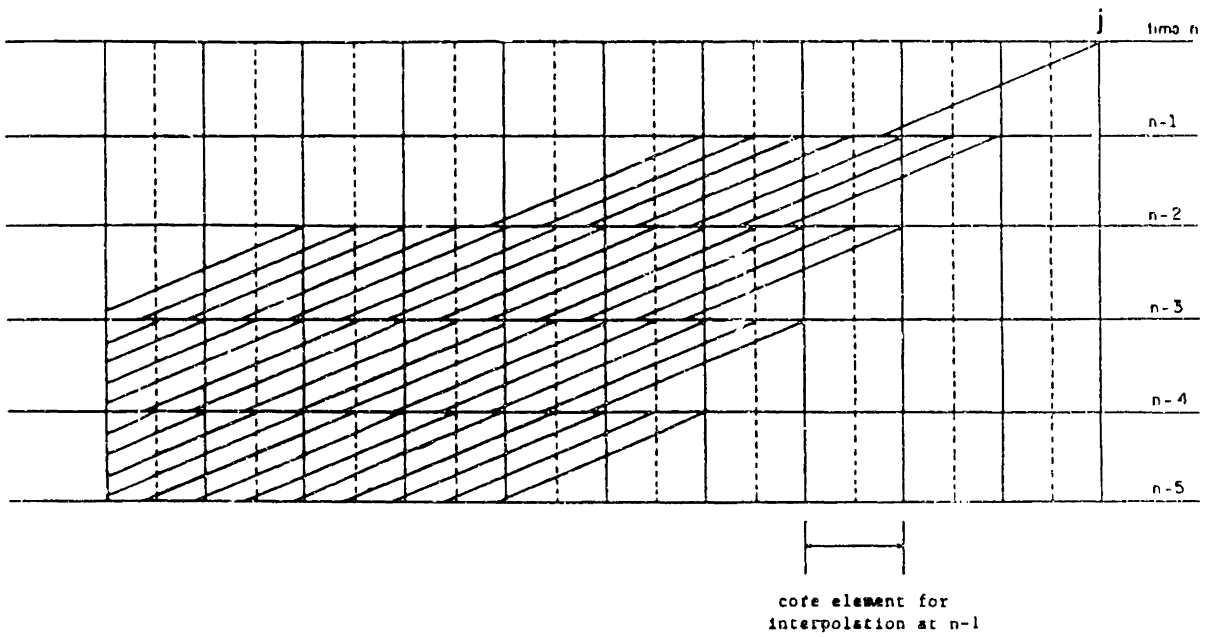


(d) Case of a middle node, with $\text{int}(Cu)$ odd

Figure A1 (cont)



(a) Case of interpolators with five nodes



(b) Case of interpolators with seven nodes

Figure A2 Zone of influence of the concentration at (j, n) , for non-compact interpolators

THE CONSISTENCY, STABILITY, AND CONVERGENCE
OF THE BACKWARDS METHOD OF CHARACTERISTICS

by

António Melo Baptista

Massachusetts Institute of Technology

Cambridge, Massachusetts 02139

January 1987

ABSTRACTS

We present a systematic analysis of the consistency, stability, and convergence of the solution of the linear advection equation by the Backwards Method of Characteristics (BMC), which shows the dependence of these properties on the interpolator selected to find the concentrations at the feet of the characteristic lines. The BMC is unconditionally consistent for interpolators that match nodal concentrations exactly; however, stability, hence convergence, has to be assessed in a interpolator-per-interpolator basis, and criteria, based on the Fourier analysis of a general BMC algorithm, are derived for such evaluation; the application of these criteria indicates that common choices of interpolators lead to stability and convergence. The apparent paradox of the BMC being convergent and improving its accuracy as Δt increases (for fixed Δx and total computational time) is explained through the different dependence of errors on Δt above and below Courant number one.

1. INTRODUCTION

A review of alternative methods for the solution of the transport equation suggests (e.g., see [B4]) that they fit into three very general categories--Eulerian (EM), Lagrangian (LM) and Eulerian-Lagrangian (ELM)--according to the selection of the coordinate system for the equation (Eulerian or Lagrangian) and of the type of computational grid (fixed or moving). While EM (which use Eulerian coordinates and fixed grids) and LM (which use Lagrangian coordinates and moving grids) may be appropriate choices for specific problems, ELM (which use Lagrangian coordinates but fixed grids) are unique in their potential to combine the best aspects of EM and LM. Numerical experimentation (e.g., [B1-B4], [H1-H3], [K1]) has suggested that, indeed, ELM provide accurate solutions for the transport equation in the whole range between advection-dominated problems (where EM have inherent limitations if sharp concentration gradients in the flow

direction are present) and dispersion-dominated problems (while LM experience practical difficulties when dispersion becomes non-negligible).

ELM lack, however, besides tradition, a well-established theoretical background that may (a) validate them as reliable numerical tools, and (b) guide modellers in the choice of computational parameters (Δt and Δx) and of specific techniques to solve each of the sub-equations in which ELM typically divide the transport equation (advection, and dispersion plus internal sinks and sources).

In this study, we deal only with the solution of advection, which has been considered a critical step in the ELM procedure (e.g., [B4]); consistently with common practice, we elect to use the Backwards Method of Characteristics (BMC) as solution technique. It is known that the accuracy of the BMC strongly depends on the interpolator used to find concentrations at the feet of the characteristic lines, and a significant part of recent ELM research has been devoted to find an interpolator for the BMC that ensures an high-level of accuracy at feasible cost. However, two legitimate questions of a more fundamental nature can and have been raised (e.g., [N1]):

- Can the BMC be consistent and stable (hence convergent), knowing that, for fixed grid and total computational time, accuracy tends to improve as Δt increases?

- How do consistency, stability and convergence depend on the choice of the interpolator?

Taking as a reference a general BMC that can accommodate a choice of interpolators of class C_0 (i.e., preserving inter-element continuity of concentrations, but not of its derivatives), we show that the BMC is consistent for all interpolators that preserve the nodal concentrations exactly, but that stability (hence convergence) have to be assessed in an individual basis. Criteria are derived for such assessment. The study relies strongly on the Taylor-series and Fourier analysis of the general algorithm of the BMC, and assumes constant advection in a 1-D uniform grid.

2. REVIEW OF THE BACKWARDS METHOD OF CHARACTERISTICS

The advection equation

$$\frac{Dc}{Dt} = \frac{\partial c}{\partial t} + u_1 \frac{\partial c}{\partial x_1} = 0 \quad (1)$$

states that the concentration, $c(x_1, t)$ remains constant along characteristic lines that follow the flow, i.e., obeying the constraint

$$\frac{dx_1}{dt} = u_1(x_1, t) \quad (2)$$

The BMC is a direct numerical application of this statement, involving, per time step, Δt , two basic tasks (Figure 1):

- Tracking backwards, between times n and $n-1$, the characteristic lines whose heads coincide with the nodes of a fixed reference grid, so as to locate their feet.

- Finding the concentration at the foot of each characteristic line, by interpolation from the known nodal concentrations at time $n-1$, and assigning it to the head of the characteristic line.

Given a 1-D uniform grid, and a constant velocity, this procedure can be represented by the algorithm

$$c(j,n) \equiv c(j-\beta,n-1) \equiv \sum_{p=P_1}^{P_2} \phi_p(\alpha) \cdot c(j-k+p) \quad (3)$$

where (see Figure 2 for reference)

- j - denotes the node where the concentration is to be computed
(global notation)
- n - denotes the instant where the concentration is to be computed
- α - denotes the position of the foot of the characteristic line in a local coordinate system with origin at node $j-k$ (α is associated with the fractional part of the Courant number,
 $\beta = u \cdot \Delta t / \Delta x$)
- P_1, P_2 - denote the extreme nodes of the region that is used to define the interpolator (in local notation)
- ϕ_p - are elementary shape functions which, together, define the interpolator

We note the distinction between the region that is used to define the interpolator, and the region where the interpolator is applied--"core element": all nodes of the core elements contribute the definition of the

interpolator, but this may (non-compact interpolators) or may not (compact interpolators) require in addition information from outside nodes. We will consider, throughout this paper, interpolators with either 2-node or 3-node core elements (which may use information from up to seven nodes).

Several interpolators have been used or considered for use in the context of the BMC. While most of our present analysis is based on the general algorithm described by Equation (3), we will refer occasionally, for illustration purposes, to a selected set of interpolators, defined in Table 1 (note: familiarity with the nomenclature used in this table will be assumed along the text). These and other interpolators are the object of a detailed comparison in a companion paper, concerned with overall performance (including accuracy)--[B6].

3. ANALYSIS OF CONSISTENCY

Expanding each term of the general numerical algorithm of the BMC (Equation 3) in Taylor series around $c(j,n)$, and rearranging, we obtain the local equilibrium statement

$$\frac{\partial c}{\partial t} + u \frac{\partial c}{\partial x} = \epsilon \quad (4)$$

where ϵ denotes the truncation error, and, as shown in Appendix A, can be written in the general form:

$$\epsilon = \frac{(-1)^Q}{(Q+1)!} P(\alpha) \frac{\Delta x^{Q+1}}{\Delta t} \frac{\partial^{Q+1} c}{\partial x^{Q+1}} + \text{H.O.D.} \quad (5)$$

Q is the effective degree of the interpolation function, a concept that we introduce to denote the degree of the highest degree generic polynomial that the interpolator can match exactly. Q can not exceed the actual degree of the interpolator, M; for instance, if a quadratic interpolator is used, $Q \leq 2$. However, Q may be smaller than M (e.g., for interpolators involving Hermite polynomials with derivatives estimated from lower-order Lagrange polynomials--e.g., the case of the 6P-PL2 and 7P-HL3 interpolators).

$P(\alpha)$ represents a polynomial in α of degree $\eta = \max \{M, Q+1\}$; i.e., formally

$$P(\alpha) \equiv \sum_{\rho=0}^{\eta} b_{\rho} \cdot \alpha^{\rho} \equiv \sum_{\rho=0}^{\eta} b_{\rho} \cdot \left[\beta - \text{int}(\beta) + \xi \right]^{\rho} \quad (6)$$

where b_{ρ} are coefficients that depend on the interpolator, and ξ is an auxiliary variable, that may take the values 0 and ± 1 --see Figure 2.

Consistency requires that the truncation error vanishes as $\Delta x, \Delta t \rightarrow 0$, being unconditional only if this happens independently on how Δx and Δt approach zero. The necessary and sufficient condition for the consistency of the BMC is then

$$\lim_{\Delta x, \Delta t \rightarrow 0} P(\alpha) \cdot \frac{\Delta x}{\Delta t}^{Q+1} \equiv \lim_{\Delta x, \Delta t \rightarrow 0} \sum_{\rho=0}^{\eta} b_{\rho} \cdot \left[\beta - \text{int}(\beta) + \xi \right]^{\rho} \cdot \frac{\Delta x}{\Delta t}^{Q+1} = 0 \quad (7)$$

The validity of this condition can not be taken for granted without further analysis. Indeed, due to the meaning of β , both positive and negative powers of Δx and Δt can, in principle, appear in the expression of the truncation error, and therefore, consistency may not be achieved, or may depend on how Δx and Δt approach zero. The three mutually exclusive cases of the limit of $\Delta x/\Delta t$ being zero, finite but non-zero, and infinite are individually considered in the following discussion.

Let Δx go to zero faster than Δt , i.e. $\Delta x/\Delta t \rightarrow 0$. While both β and $\text{int}(\beta)$ tend to infinity as $\Delta x, \Delta t \rightarrow 0$, $\beta - \text{int}(\beta)$ is kept finite (≤ 1) by definition, and, therefore, so is $P(\alpha)$. Hence, in this case, the consistency condition is necessarily satisfied, regardless of the actual interpolator (assuming $Q \geq 0$).

Now, let Δx and Δt go to zero at the same rate. In this case, β and $\text{int}(\beta)$ are held constant and finite, and so is $P(\alpha)$. The consistency condition is necessarily satisfied by all interpolators of effective degree one or above, but is not satisfied by interpolators of effective degree zero or below.

Finally, let Δt go to zero faster than Δx . Now, $\beta \rightarrow 0$, and $\text{int}(\beta)$ becomes zero for $\beta < 1$. Hence,

$$\begin{aligned} \lim_{\Delta x, \Delta t \rightarrow 0} P(\alpha) \cdot \frac{\Delta x^{Q+1}}{\Delta t} &= \lim_{\Delta x, \Delta t \rightarrow 0} \left\{ \frac{\Delta x^{Q+1}}{\Delta t} \cdot \sum_{\rho=0}^{\eta} b_{\rho} \cdot (\beta + \xi)^{\rho} \right\} = \\ &= \lim_{\Delta x, \Delta t \rightarrow 0} \left\{ u \cdot \frac{\Delta x^Q}{\beta} \sum_{\rho=0}^{\eta} b_{\rho} \cdot \sum_{\theta=0}^{\rho} \frac{\rho!}{\theta!(\rho-\theta)!} \beta^{\theta} \cdot \xi^{\rho-\theta} \right\} \quad (8) \end{aligned}$$

and, because both β and Δx tend to zero, all terms in the summation over θ , but those with $\theta = 0$, will also necessarily tend to zero; i.e.,

$$\lim_{\Delta x, \Delta t \rightarrow 0} P(\alpha) \cdot \frac{\Delta x^{Q+1}}{\Delta t} = \lim_{\Delta x, \Delta t \rightarrow 0} \left\{ u \frac{\Delta x^Q}{\beta} \sum_{\rho=0}^{\eta} b_{\rho} \cdot \xi^{\rho} \right\} \equiv P(\xi) \cdot \lim_{\Delta x, \Delta t \rightarrow 0} \left\{ \frac{\Delta x^Q}{\beta} \right\} \quad (9)$$

Hence, when Δt goes to zero faster than Δx , the consistency condition is necessarily satisfied for all interpolators that make $P(\xi)$ zero; for other interpolators, this condition can be satisfied only when the rate of convergence of Δt is slower than that of Δx^{Q+1} (which immediately excludes interpolators of effective degree 0 or below).

Summarizing the above analysis, which covered all possible relative rates of convergence of Δx and Δt , we can now state that:

- If $P(\xi) = 0$, the BMC is unconditionally consistent for all interpolators of effective degree 1 or above.
- If $P(\xi) \neq 0$, the BMC can, at best, be conditionally consistent, with the truncation error vanishing only when $\Delta x^{Q+1}/\Delta t \rightarrow 0$.
- No interpolator with effective degree less than unity can lead to unconditional consistency, regardless of the value of $P(\xi)$.

To understand what $P(\xi) = 0$ actually means as a constraint, we recall, from the definition of the polynomial $P(\alpha)$, that

$$P(\xi) = 0 \iff \xi^{Q+1} = \sum_{p=P_1}^{P_2} \alpha_p^{Q+1} \cdot \phi_p(\xi) \quad (10)$$

We can now recognize that because, by definition, the value of ξ coincides the coordinate of one of the nodes of the core element, a sufficient condition for $P(\xi)$ to be null is that the interpolator be exact at such nodes, or, equivalently, that:

$$\phi_p(\alpha_\tau) = \begin{cases} 1 & \text{if } p = \tau \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

This is, in particular, the case of all Lagrange interpolators, and of all Hermite interpolators with derivatives estimated from Lagrange polynomials of the same degree, which are, therefore, unconditionally consistent. We note that, together, these interpolators represent the largest majority of interpolators of class C_0 ever used in a BMC context.

The condition expressed by Equation 11 is not strictly necessary for $P(\xi)$ to be zero, but other forms of achieving this identity (hence unconditional stability) lack robustness outside the context of uniform grids. For instance, the 7P-HL3, for which the concentration at the middle node of the core element is not necessarily preserved, still leads to $P(\xi) = 0$, but this is due to a symmetry effect that would disappear if a non-uniform grid were used.

Also, we note that, while it makes no sense to develop on purpose interpolators with $Q < 1$ (which are bound to have poor overall accuracy) rounding-off the coefficients of an interpolator may have this effect. The most significant example is the 6P-PL2, which has a theoretical effective degree of 1, but can not represent a constant function within machine precision, hence leading to the unconditional inconsistency of the BMC.

The effect of this inconsistency is illustrated in Figure 3, taking as a reference the problem of the constant advection of a Gauss-hill, which was solved--for fixed total time and time step, but variable Δx --using the 6P-PL2 and the 4P-LR2 interpolators; while for the 4P-LR2 accuracy improves as Δx decreases, for the 6P-PL2 accuracy deteriorates in the region of small Δx .

4. ANALYSIS OF STABILITY

4.1. Stability criteria

Stability requires that errors generated in the arithmetic operations needed to actually apply the numerical algorithm be not amplified by this algorithm. This property is commonly analysed by observing the errors in the propagation of individual Fourier components of the solution: the numerical method will be stable if the amplitudes of these errors converge to a finite limit, as the number of time steps goes to infinity (Δt fixed).

The Fourier analysis of the BMC was addressed in detail by [B5], who derived the general error formulae summarized in Table 2. These formulae show that, for interpolators based on 2-node core elements, errors are time-independent; hence, the requirement for stability is that the amplitude of the error in the first (hence, any) time step do not exceed unity, for any wavelength that can be represented by the grid; i.e.,

$$|G_m(\alpha)| \leq 1, \quad \text{for } m \leq J/2 \quad (12)$$

where $G_m(\alpha)$ represents the error in the propagation of the m^{th} Fourier component, and $J+1$ is the number of grid nodes. Equation 12 is easily recognized as a particular form of the classical Von Newman stability criterion.

For interpolation functions with 3-node core elements, errors are time-dependent, as a consequence of a numerically generated transfer of energy between Fourier components (see [B5]). To establish a formal stability criterion, we have then to require, in this case, that

$$\lim_{N \rightarrow \infty} | H_m^v(\alpha, N) | = C \quad \text{for all } m \leq J/2 \quad (13)$$

where $H_m^v(\alpha, N)$ represents the cumulative error in the propagation of the m^{th} Fourier component of the solution, evaluated after N time steps, and C is some finite value (which will, in general, be zero).

Using well-established knowledge from the theory of iteration, we show in Appendix B that the necessary and sufficient condition for Equation (13) to be obeyed is

$$| \theta_m | \leq 1 \quad \text{for all } m \leq J/2 \quad (14)$$

where $\theta_m(\alpha)$ is the the largest (in modulus) eigenvalue of the matrix

$$\underline{\underline{R}} = \begin{vmatrix} \operatorname{Re}\{s_m(\alpha)\} & -\operatorname{Im}\{s(\alpha)\} & \operatorname{Re}\{s_m(\alpha \pm 1)\} & -\operatorname{Im}\{s_m(\alpha \pm 1)\} \\ \operatorname{Im}\{s_m(\alpha)\} & \operatorname{Re}\{s(\alpha)\} & \operatorname{Im}\{s_m(\alpha \pm 1)\} & \operatorname{Re}\{s_m(\alpha \pm 1)\} \\ \operatorname{Re}\{r_m(\alpha)\} & -\operatorname{Im}\{r(\alpha)\} & \operatorname{Re}\{r_m(\alpha \pm 1)\} & -\operatorname{Im}\{r_m(\alpha \pm 1)\} \\ \operatorname{Im}\{r_m(\alpha)\} & \operatorname{Re}\{r(\alpha)\} & \operatorname{Im}\{r_m(\alpha \pm 1)\} & \operatorname{Re}\{r_m(\alpha \pm 1)\} \end{vmatrix} \quad (15)$$

The functions s_m and r_m are defined in Table 2.

A less formal, but perhaps more intuitive stability criterion can also be derived, based on the empirical observation ([B5]) that, after a finite number of time steps, propagation errors per time step become time-independent, except for $L_m = 4\Delta x$. Indeed, this suggests that stability requires, in a direct extension of Von Newman criterion to the case of 3-node core elements, that:

$$\left| \frac{H_m^D(\alpha, N)}{H_m^D(\alpha, N-1)} \right| \leq 1 \quad \text{for } m \leq J/2 \text{ and } L_m \neq 4\Delta x, \text{ for large } N \quad (16)$$

and

$$\lim_{N \rightarrow \infty} |H_m^D(\alpha, N)| = 0 \quad \text{for } L_m = 4\Delta x \quad (17)$$

The criteria expressed respectively by Equations (14) and (17)-(18) should be equivalent.

4.2. The stability of the BMC for specific interpolators

The application of any of the above stability criteria is, for most interpolators, a very hard--if at all possible--task, when pursued analytically. To overcome this difficulty, we recommend the numerical generation of charts defining the regions of stability as a function of both α and $L_m/\Delta x$, such as those presented in Figure 4a-b. These charts should cover a domain bounded by $2 \leq L_m/\Delta x \leq \ell$ (with ℓ chosen large enough to represent infinity), and by $0 \leq \alpha \leq 1$; we took $\ell = 50$, and computed the amplification factors in a support grid characterized by $\Delta\alpha = 0.02$ and $\Delta L_m = 0.50 \cdot \Delta x$.

Figures 4a-b represent, respectively, the amplification factors (in the first time step) for a non-compact cubic interpolator, 4P-LR2, and for the also non-compact 6P-PL2; both interpolators have 2-node core elements, hence Equation (12) prevails as the proper stability criterion. We note that the BMC is stable for the former interpolator, and unstable for the latter. We also note that, unlike for the 4P-LR2, the amplification factors for the 6P-PL2 are not symmetrical around $\alpha = 0.5$ (a deviation that results directly from the approximations introduced in the evaluation of the coefficients of the 6P-PL2).

Figures 4c-d, in turn, concern the compact quadratic interpolator, 3P-LI3, and the non-compact 5P-HL3 interpolator, constructed from the estimation of derivatives of a cubic Hermite polynomial; both interpolators have 3-node core elements. The figures show, for each interpolator, the amplification factors per time step (after a "large" number of time steps--

40)). We note that (a) the BMC is stable for the 3P-LI3, while it is unstable for the 5P-HL3; (b) except at or near $L_{in}/\Delta x = 4$ --where cumulative amplification is virtually null (results not shown)--amplification factors per time step are, for the 3P-LI3, essentially symmetrical around $\alpha = 0.5$, an indication that energy transfer between Fourier components is approaching the equilibrium; this is still not the case for the 5P-HL3.

5. ANALYSIS OF CONVERGENCE

Lax's well-known equivalence theorem states that, for a linear system of equations and a consistent numerical method, stability is the necessary and sufficient condition for convergence.

We can therefore use our previous analysis of consistency and stability, to conclude that the convergence of the BMC (for the solution of the linear advection equation) depends on the selection of the interpolator, a choice of those providing convergent solutions (e.g., see [B6]). Here, we concentrate in the understanding of an intriguing aspect: if, as suggested by numerical experimentation (e.g., [B1], [N1]), the BMC improves its accuracy as Δt increases (for fixed Δx , and total computational time, T), how can the method be convergent?

Let us take the 4P-LR2 as a reference interpolator; the truncation error of the BMC is, in this case

$$\epsilon = \alpha \cdot (\alpha - 2) \cdot (\alpha^2 - 1) \frac{\Delta x^4}{24\Delta t} \frac{\partial^4 c}{\partial x^4} + \text{H.O.D.} \quad (19)$$

which indicates that for large β (where α is a weak function of Δx , Δt) increasing Δt will indeed improve accuracy, at least in a mean square sense. However, for $\beta < 1$ (in which case $\alpha = u \cdot \Delta t / \Delta x$), we have

$$\epsilon = \frac{u}{24}(u^3 \Delta t^3 - 2u^2 \Delta t^2 \Delta x - u \Delta t \Delta x^2 + 2\Delta x^3) \frac{\partial^4 c}{\partial x^4} + \text{H.O.D.} \quad (20)$$

which indicates that, in this case, *reducing* Δt should improve accuracy, although probably not very efficiently (note that Δx^3 is the dominant coefficient).

Because letting Δt go to zero, with Δx and T fixed, will inevitably bring β to below unity, the dependence of the accuracy on Δt is recognized to be more complex than anticipated by [B1] and [N1], among others. Indeed, reductions of Δt will affect the accuracy of the BMC differently, depending on the value of the Courant number: for $\beta > 1$, increasing Δt will significantly (although only in a least square sense, due to the periodic variation of α) improve accuracy; if $\beta < 1$, decreasing Δt will improve accuracy, but not significantly, as errors are in this region dominated by the spatial discretization--see Figure 5, for illustration. This general behavior had, less formally, been anticipated by [B3], and explains why the BMC can and will in general be convergent, and, still benefit from the use of large Δt .

6. SUMMARY AND CONCLUSIONS

Both common sense and numerical experimentation have long ago recognized that the choice of the interpolator for the BMC strongly affects

its accuracy. This study shows that this choice has an even more fundamental influence, being a determining factor for the consistency, stability and convergence of the method.

Most interpolators of class C_0 pose no threat to consistency. However, not all interpolators that have been considered recently to improve the accuracy of the BMC lead to stability, hence convergence. The Von Newman stability criterion, or its extension to interpolators with 3-node core elements, as proposed in Section 3, provide the necessary tools to investigate the stability of the BMC, for individual interpolators. Application of these tools is illustrated here, but is used in a more systematic way in [B6].

The fear that the BMC could be intrinsically divergent [N1] is not justified, and the ability to improve accuracy by increasing Δt (for a fixed grid and a fixed total computational time, for large Courant numbers) can be considered a definite advantage of the method.

We stress that, although the BMC is an explicit method, it is not, unlike its Eulerian counterparts, subject to stability constraints on the Courant number. This is a direct consequence of the fact that we force interpolations to be performed within the core element that contains the foot, rather than the head, of the relevant characteristic lines.

Some of the formal tools and criteria derived in this paper are interesting on their own, and deserve further research or application. This includes, in particular,

- The general expression derived for the truncation error can be effectively used to compare the accuracy of alternative interpolation functions, and to understand the general dependence of the accuracy of the BMC on controlling parameters--see [B6], [B7].

- The stability criteria derived to account for the time-dependence of the propagation errors can, in principle, be generalized to other numerical methods that use quadratic elements (e.g., FE-Galerkin).

REFERENCES

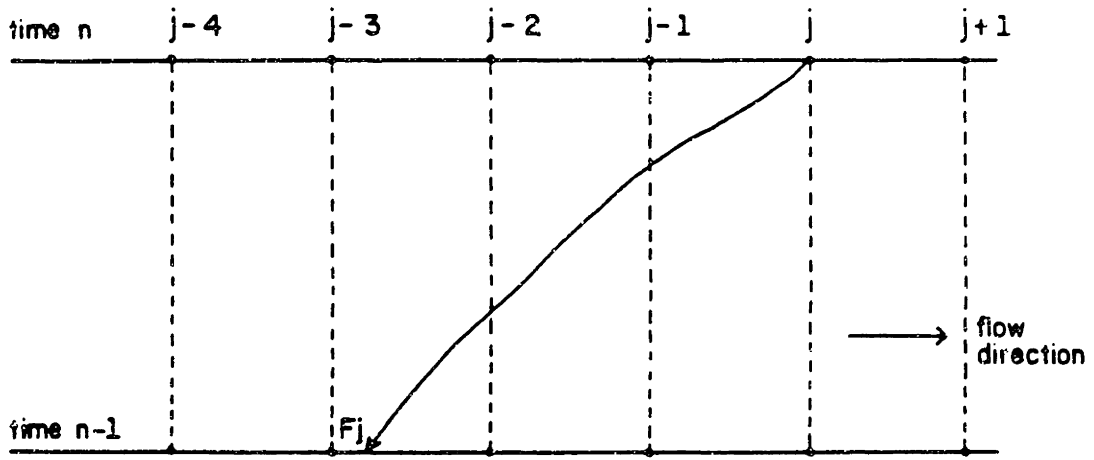
- B1 Baptista, A. M., E. E. Adams, and K. D. Stolzenbach. "Eulerian-Lagrangian Analysis of Pollutant Transport in Shallow Water." MIT R. M. Parsons Laboratory, Technical Report 296, 1984.
- B2 Baptista, A. M., E. E. Adams, and K. D. Stolzenbach. "Comparison of Several Eulerian-Lagrangian Models to Solve the Advection-Diffusion Equation." in *Int. Symp. on Refined Flow Modeling and Turbulence Measurements*. U. Iowa, USA, 1985.
- B3 Baptista, A. M., E. E. Adams, and K. D. Stolzenbach. "Accuracy Analysis of the Backwards Method of Characteristics." in *Finite Elements in Water Resources: Proceedings of the 6th International Conference*. (Sá da Costa et al., Ed.), Springer-Verlag, 1986.
- B4 Baptista, A. M. "Accurate Numerical Modeling of Advection-Dominated Transport of Passive Scalars." LNEC, Lisboa, 1986.
- B5 Baptista, A. M. "Fourier Analysis of the Backwards Method of Characteristics." Ph.D. Thesis, M.I.T., 1987.
- B6 Baptista, A. M. "The Choice of the Interpolator for the Backwards Method of Characteristics." Ph.D. Thesis, M.I.T., 1987.
- B7 Baptista, A. M. "The Accuracy of the Backwards Method of Characteristics." Ph.D. Thesis, M.I.T., 1987.
- F1 Faddeeva, V. N. *Computational Methods of Linear Algebra*. Dover, 1959.
- H1 Holly, F. M., Jr., and T. Komatsu. "Derivative Approximations in the Two-Point Fourth-Order Method for Pollutant Transport." *Proceedings of the Conference on Frontiers in Hydraulic Engineering*, ASCE, M.I.T., 1984.
- H2 Holly, F. M., Jr., and J. M. Polatera. "Dispersion Simulation in 2-D Tidal Flow." *Journal Hydr. Engrg.*, ASCE, 1984.
- H3 Holly, F. M., Jr., and A. Preissmann. "Accurate Calculation of Transport in Two Dimensions." *Journal of the Hydraulics Division*, ASCE 103(HY11):1259-1278, Nov. 1977.
- K1 Komatsu, T., F. M. Holly, Jr., and N. Nakashiki. "Numerical Calculation of Pollutant Transport in Rivers and Coastlines", in *4th Congress, Asian and Pacific Division*, IAHR, Chiang Mai, Thailand, 1984.
- N1 Neuman, S. P. "Adaptive Eulerian-Lagrangian Finite Element Method For Advection-Dispersion." *Int. Journal for Numerical Methods in Engineering* 20:321-337, 1984.

Figure 1

Illustrative sketch for the Backwards Method of Characteristics

Required steps:

1. Tracking of the characteristic lines. For each node j , a characteristic line is independently defined by the backwards (i.e., between n and $n-1$) solution of an ordinary differential solution of the form $dx_i/dt = u_i$.
2. Interpolation at the feet of the characteristic lines. The concentration at the foot of each characteristic line (and, for pure advection, at any other point of the characteristic line, including node j at time n) is found by interpolation from known information on neighboring nodes, at time $n-1$.



Number of nodes of the core element	Illustrative sketch	Value of ξ
2	<p style="text-align: center;"> $\beta = \frac{u \Delta t}{\Delta x}$ $K = \text{int}(\beta) + F$ $\alpha = \beta - K$ </p> <p style="text-align: right;">note: $0 \leq \alpha \leq 1$</p>	$u \geq 0$ $\xi = 0$ $u < 0$ $\xi = -1$
3	<p style="text-align: center;"> $\beta = \frac{u \Delta t}{\Delta x}$ $K = \text{int}(\beta) + F$ $\alpha = \beta - K$ </p> <p style="text-align: right;">note: $\alpha \leq 1$</p>	$u \geq 0$ F between $j-K-1$ and $j-K$: $\xi = 0$ F between $j-K$ and $j-K+1$: $\xi = 1$ $u < 0$ F between $j-K-1$ and $j-K$: $\xi = -1$ F between $j-K$ and $j-K+1$: $\xi = 0$

Figure 2 Definition of the core elements

Figure 3

Illustration of the convergence or divergence of the BMC, as a consequence of the choice of the interpolator

While the 6P-PL2 interpolator is more accurate than the 4P-LK2 interpolator for larger Δx , the inconsistency of the former induces errors to grow at the limit of small Δx . All interpolators considered in this work (but the 6P-PL2) will improve the accuracy of the solution as Δx decreases.

$$\begin{aligned} \text{Reference problem: } & \frac{\partial c}{\partial t} + u \frac{\partial c}{\partial x} = 0 \\ & c(x, 0) = \exp\left\{-\frac{(x-x_0)^2}{2\sigma^2}\right\} \\ & c(x, t) \rightarrow 0 \quad |x| \rightarrow \infty \end{aligned}$$

Computational parameters: $\Delta t = 0.4$

Δx variable

$N = T/\Delta t = 50000$

$u = 0.5$

$\sigma = 264$

L2 error norm

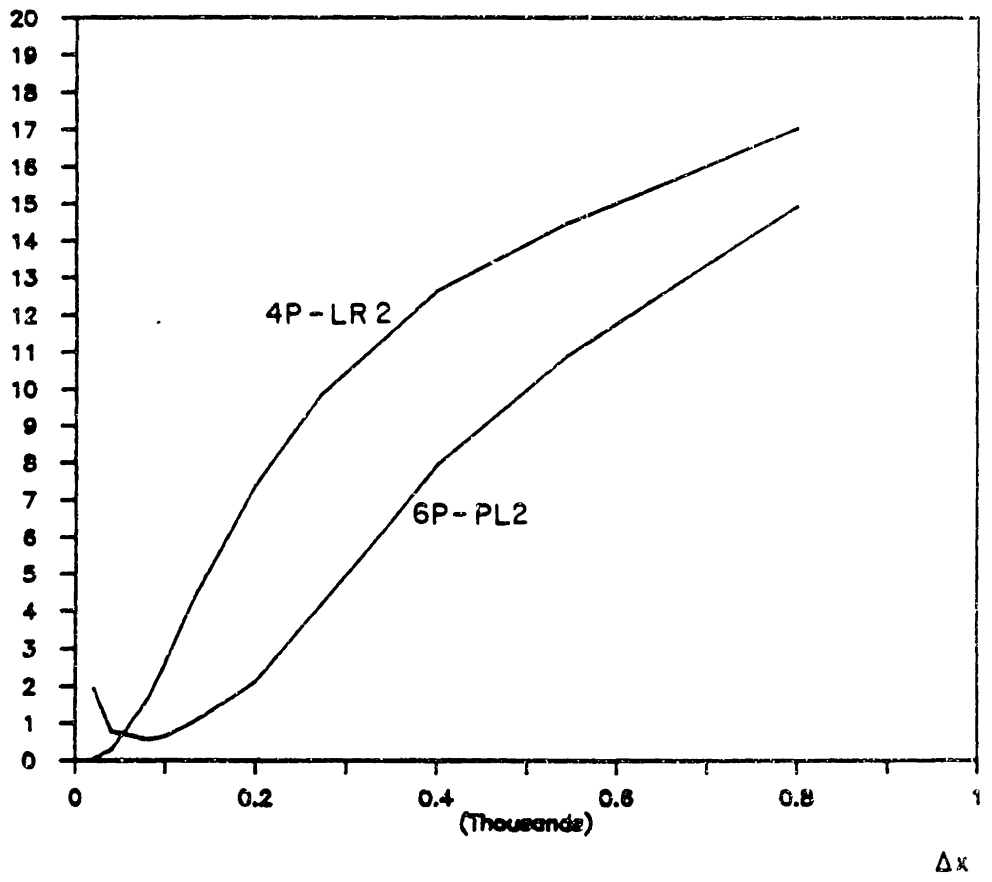


Figure 4

Amplification factors per time step, as a function of the location, α ,
of the foot of the characteristic line within the core element,
and of the dimensionless wavelength

(a) 4P-LR2

(b) 6P-PL2

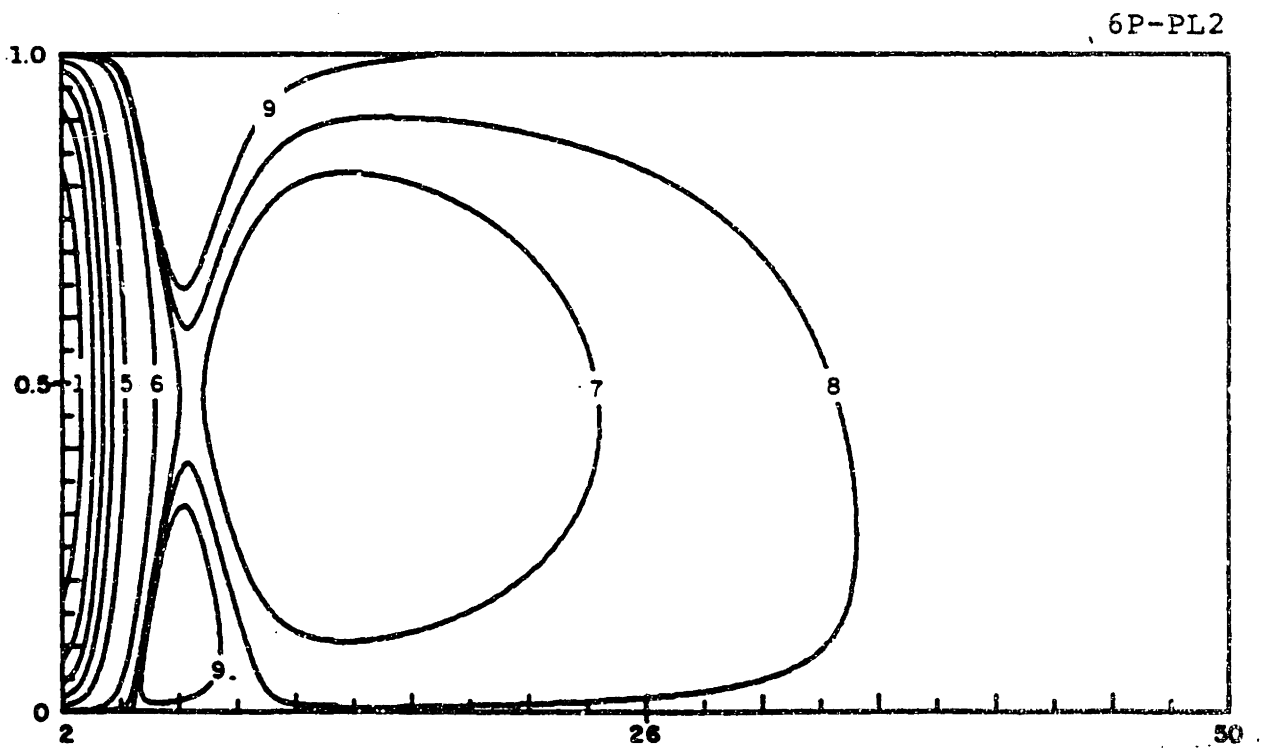
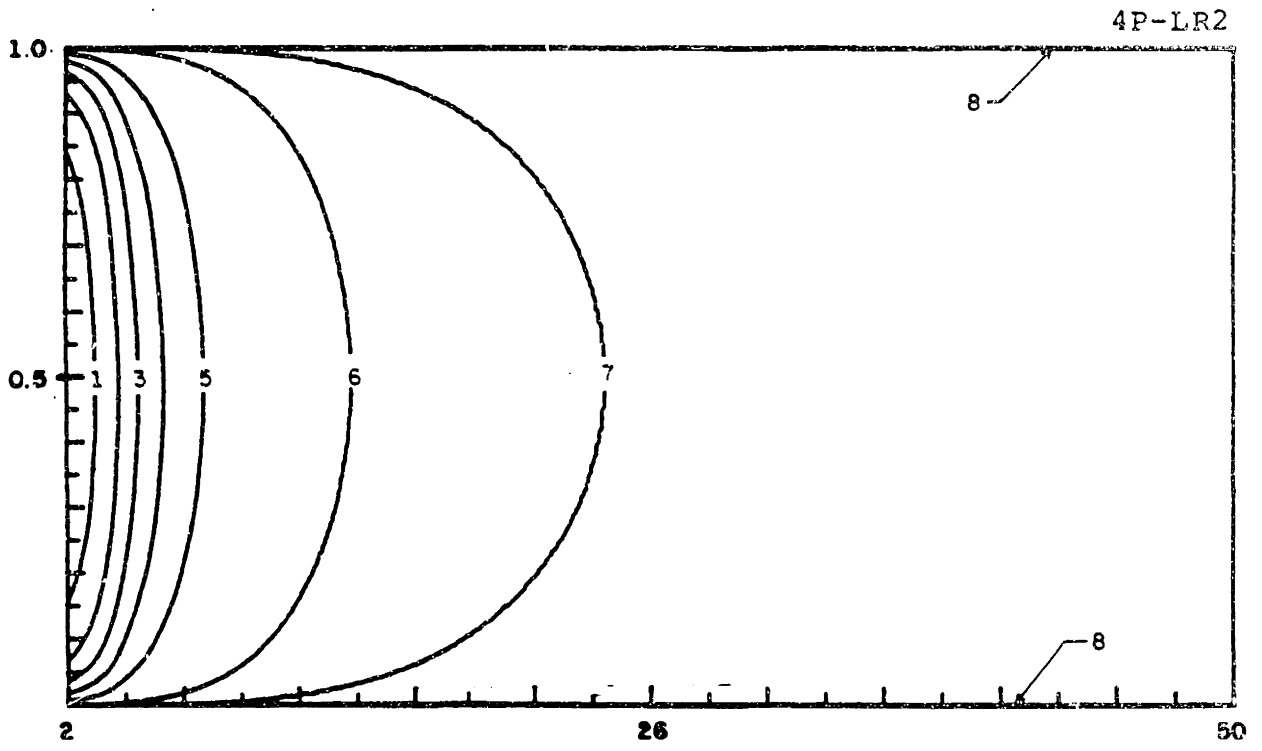
(c) 3P-LI3

(d) 5P-HL3

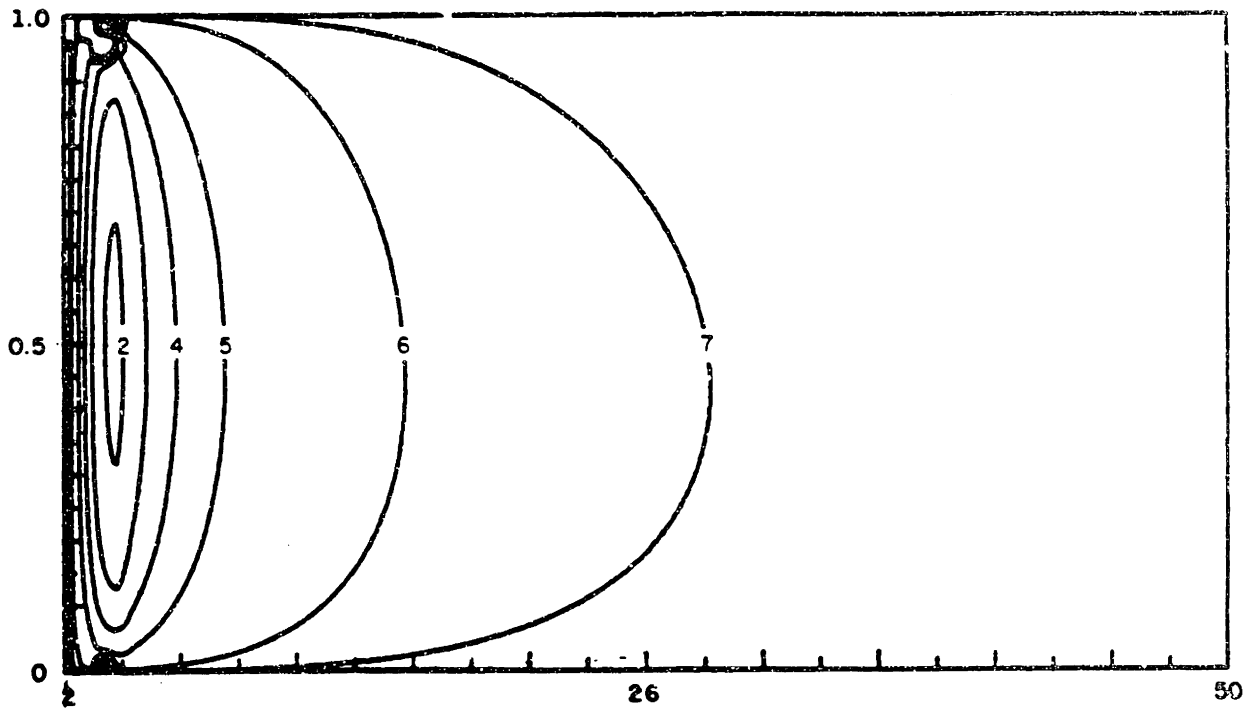
Legend

Contour	Amplification factor
1	0.75
2	0.90
3	0.95
4	0.975
5	0.990
6	0.999
7	0.9999
8	1.0000
9	1.0001
10	1.001
11	1.0025
12	1.005
13	1.01

Note: Irregularities in the contours of the amplification factors
for the 3P-LI3 and 5P-HL3, near $L_m/\Delta x$, result from the
inability of the countouring code to handle the prevailing
gradients



3P-LI3



5P-HL3

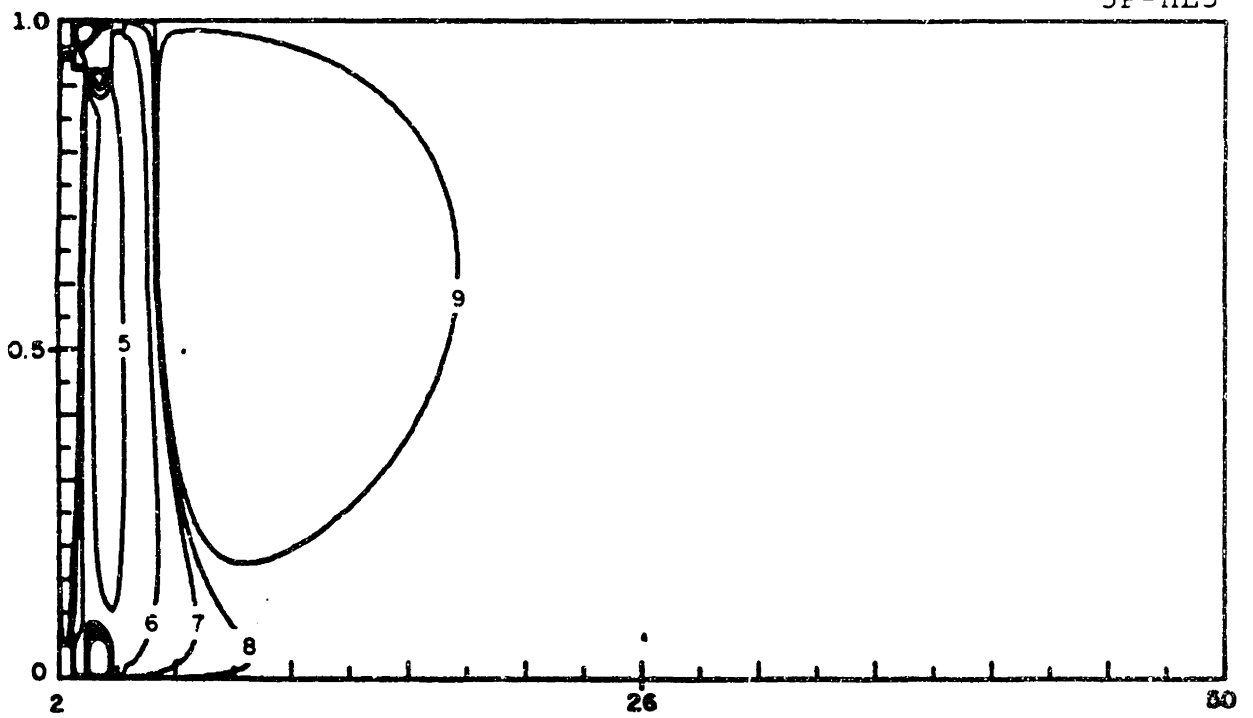


Figure 5

Dependence of accuracy on the number of time steps required to reach a fixed total time

We note that accuracy tends to improve as Δt increases. However, in the limit of large N (small Δt), the accuracy is essentially insensitive to the variation of Δt ; errors define then a plateau, which is a function of Δx .

Reference problem: $\frac{\partial c}{\partial t} + u \frac{\partial c}{\partial x} = 0$
 $c(x, 0) = \exp\left\{-\frac{(x-x_0)^2}{2\sigma^2}\right\}$
 $c(x, t) \rightarrow 0 \quad |x| \rightarrow \infty$

Computational parameters: $\Delta t = \text{variable}$

$$\Delta x = 200$$

$$T = 9600$$

$$u = 0.5$$

$$\sigma = \begin{array}{|l} 132 \\ 264 \\ 528 \end{array}$$

- Legend
- 1: $\sigma/\Delta x = 0.66$
 - 2: $\sigma/\Delta x = 1.32$
 - 3: $\sigma/\Delta x = 2.64$

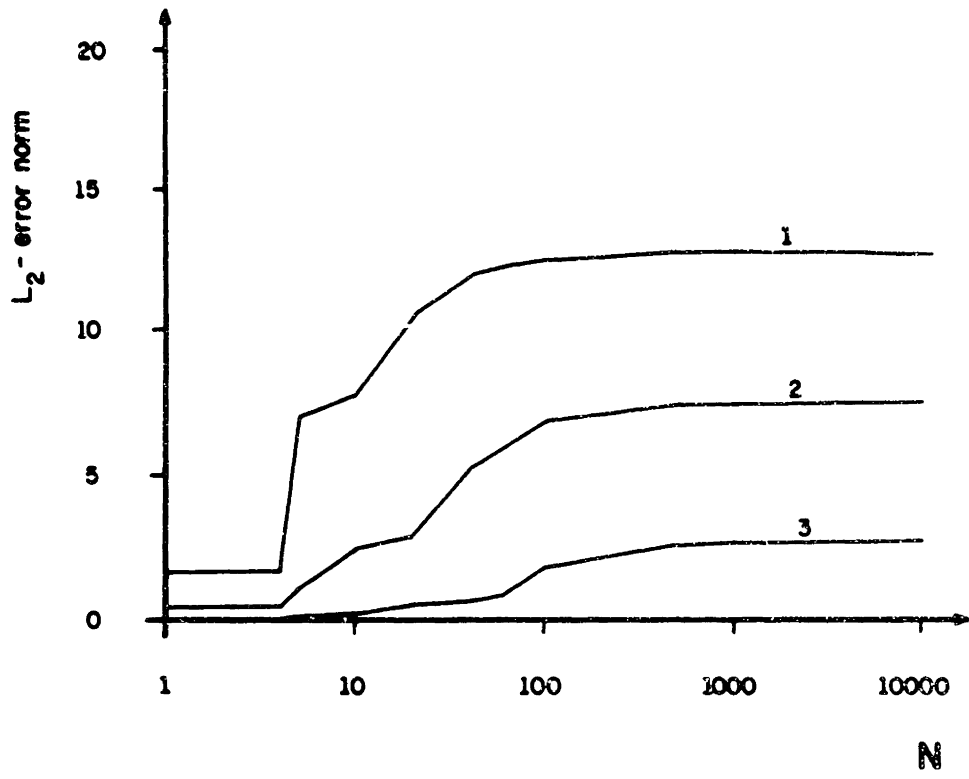


Table 1
Definition of spatial interpolators

3P-LI3

4P-LR2

5P-HL3

6P-PL2

7P-HL3

Scheme	Reference sketch	Definition	Q, M	Comments
3P-LI3		$f(\alpha) = \sum_{p=1}^3 \phi_p(\alpha) \cdot f_p$ $\phi_{-1}(\alpha) = \frac{1}{2}(\alpha^2 - \alpha)$ $\phi_0(\alpha) = 1 - \alpha^2$ $\phi_1(\alpha) = \frac{1}{2}(\alpha^2 + \alpha)$	2, 2	Quadratic Lagrange interpolator Compact, class C ₀ Reference: [B2, B9]
4P-LR2		$f(\alpha) = \sum_{p=1}^4 \phi_p(\alpha) \cdot f_p$ $\phi_{-1}(\alpha) = -\frac{1}{6}(\alpha^3 - 3\alpha^2 + 2\alpha)$ $\phi_0(\alpha) = \frac{1}{6}(3\alpha^3 - 6\alpha^2 - 3\alpha + 6)$ $\phi_1(\alpha) = -\frac{1}{6}(3\alpha^3 - 3\alpha^2 - 6\alpha)$ $\phi_2(\alpha) = \frac{1}{6}(\alpha^3 - \alpha)$	3, 3	
5P-HI3		$f(\alpha) = \sum_{p=1}^5 \phi_p(\alpha) \cdot f_p$ $\phi_{-2}(\alpha) = -\frac{1}{12}\alpha(\alpha^2 - 1)$ $\phi_{-1}(\alpha) = \frac{1}{6}(\alpha^3 + 3\alpha^2 - 4\alpha)$ $\phi_0(\alpha) = 1 - \alpha^2$ $\phi_1(\alpha) = -\frac{1}{6}(\alpha^3 - 3\alpha^2 - 4\alpha)$ $\phi_2(\alpha) = \frac{1}{12}\alpha(\alpha^2 - 1)$	3, 3	Interpolator based on a cubic Hermite polynomial, defined over the core element, and which derivatives at the corner nodes are estimated from concentrations at 5 nodes. Non-compact, class C ₀ Reference: [B6]

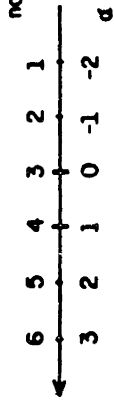
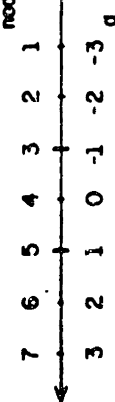
Scheme	Reference sketch	Definition	Q,M	Comments
6P-PL2		$f(\alpha) = \sum_{p=2}^3 \phi_p(\alpha) \cdot f_p$ $\phi_{-2}(\alpha) = 0.01806\alpha^2 - 0.09245\alpha + 0.07439\alpha$ $\phi_{-1}(\alpha) = -0.2570\alpha^2 + 0.8236\alpha - 0.5667\alpha$ $\phi_0(\alpha) = 0.6806\alpha^2 - 1.394\alpha - 0.2869\alpha + 1$ $\phi_1(\alpha) = -0.6806\alpha^2 + 0.6480\alpha^2 + 1.033\alpha$ $\phi_2(\alpha) = 0.2570\alpha^2 + 0.05276\alpha^2 - 0.3097\alpha$ $\phi_3(\alpha) = -0.01806\alpha^2 - 0.03828\alpha^2 + 0.05633\alpha$	1,3	<p>Interpolator based on the average of two pseudo-Hermite polynomials, defined over the core element with the assistance of a fictive middle point. Derivatives at the corner nodes are estimated by weighted average of concentrations at 6 nodes.</p> <p>Non-compact, class C_0</p> <p>Reference: [K1-K2]</p>
7P-HL3		$f(\alpha) = \sum_{p=3}^3 \phi_p(\alpha) \cdot f_p$ $\phi_{-3}(\alpha) = \frac{\tau-1}{8}(\alpha^2 - \alpha^2 - \tau + 1)$ $\phi_{-2}(\alpha) = \frac{\tau-1}{2}(\alpha^2 - \alpha^2 - \tau + 1)$ $\phi_{-1}(\alpha) = \frac{1}{8}(5(1-\tau)\alpha^2 + (7\tau-3)\alpha^2 + (5\tau-9)\alpha + 7(1-\tau))$ $\phi_0(\alpha) = \tau(1-\alpha^2)$ $\phi_1(\alpha) = \frac{1}{8}(5(7-1)\alpha^2 + (7\tau-3)\alpha^2 + (9-5\tau)\alpha + 7(1-\tau))$ $\phi_2(\alpha) = \frac{\tau-1}{2}(\alpha^2 + \alpha^2 - \tau + 1)$ $\phi_3(\alpha) = \frac{1-\tau}{8}(\alpha^2 + \alpha^2 - \tau + 1)$	2,3	<p>Interpolator based on a cubic Hermite polynomial, defined over the core element, and which derivatives at the corner nodes are estimated from concentrations at 7 nodes</p> <p>Non-compact, class C_0</p>

Table 2

Summary of formulae for propagation errors

1. Error in the first time step:

$$C_n(\alpha) = \exp(i\lambda_n \alpha) \cdot \underbrace{\sum_{p=P_1}^{P_2} \phi_p(\alpha) \cdot \exp(i\lambda_n p)}_{g_n(\alpha)} \quad (T1)$$

(note: $\lambda_n = \frac{2\pi \Delta x}{L_n}$)

2. Error after N time steps

A. Interpolators with 2-node core elements

$$H_n(\alpha, N) = (C_n(\alpha))^N \quad (T2)$$

B. Interpolators with 3-node core elements

$$H_n^v(\alpha, N) = (p_n(\alpha, N) \cdot g_n(\alpha) + r_n(\alpha, N) \cdot q_n(\alpha + \ell)) \cdot \exp(iN\alpha\lambda_n) \cdot \exp(i\nu\lambda_n) \quad (T3)$$

with

$$p_n(\alpha, N) = s_n(\alpha)p_n(\alpha, N-1) + s_n(\alpha + \ell) \cdot q_n(\alpha, N-1) \quad (T4)$$

$$q_n(\alpha, N) = r_n(\alpha)p_n(\alpha, N-1) + r_n(\alpha + \ell) \cdot q_n(\alpha, N-1) \quad (T5)$$

$$p_n(\alpha, 1) = 1 - |\nu| \quad (T6)$$

$$g_n(\alpha, 1) = |\nu| \quad (T7)$$

$$\ell = \begin{cases} -1 & \text{if int}(\beta) \text{ odd} \\ 1 & \text{if int}(\beta) \text{ even} \end{cases} \quad (T8)$$

$$\nu = \begin{cases} 0 & \text{for corner nodes} \\ \ell & \text{for middle node} \end{cases} \quad (T9)$$

where $s_n(\alpha)$ and $r_n(\alpha)$ are defined as follows:

B.1 Compact interpolators

$$r_n(\alpha) = \phi_0(\alpha) \exp(\pm i\lambda_n) \quad (T10)$$

$$s_n(\alpha) = \phi_{-1}(\alpha) \exp(i\lambda_n) + \phi_1(\alpha) \exp(-i\lambda_n) \quad (T11)$$

B.2 Non-compact interpolators (5 nodes)

$$r_n(\alpha) = \begin{cases} [\phi_2(\alpha) + \phi_0(\alpha)] \exp(i\lambda_n) + \phi_{-2}(\alpha) \exp(13\lambda_n) \\ \phi_{-2}(\alpha) \exp(1\lambda_n) + \phi_0(\alpha) \exp(-1\lambda_n) + \phi_{-2}(\alpha) \exp(-13\lambda_n) \end{cases} \quad (T12)$$

$$s_n(\alpha) = \phi_{-1}(\alpha) \exp(i\lambda_n) + \phi_1(\alpha) \exp(-i\lambda_n) \quad (T13)$$

B.3 Non-compact interpolators (7 nodes)

$$r_n(\alpha) = \begin{cases} \phi_{-2}(\alpha) \exp(13\lambda_n) + [\phi_0(\alpha) + \phi_2(\alpha)] \exp(i\lambda_n) \\ \phi_{-2}(\alpha) \exp(13\lambda_n) + \phi_0(\alpha) \exp(i\lambda_n) + \phi_2(\alpha) \exp(-1\lambda_n) \end{cases} \quad (T14)$$

$$s_n(\alpha) = [\phi_{-2}(\alpha) + \phi_2(\alpha)] \exp(13\lambda_n) + [\phi_{-1}(\alpha) + \phi_1(\alpha)] \exp(i\lambda_n) \quad (T15)$$

APPENDIX A

Derivation of the general form of the truncation error of the BMC

Let us consider the general BMC algorithm expressed by Equation 3, and expand $c(j,n+1)$ and $c(j-k-\alpha_m)$ in Taylor series around $c(j,n)$. We get, respectively,

$$c(j,n+1) = c(j,n) + \Delta t \frac{\partial c}{\partial t} + \frac{\Delta t^2}{2!} \frac{\partial^2 c}{\partial t^2} + \dots + \frac{\Delta t^s}{s!} \frac{\partial^s c}{\partial t^s} + \dots \quad (A1)$$

and

$$\begin{aligned} c(j-k-\alpha_m,n) = & c(j,n) - (k+\alpha_m) \Delta x \frac{\partial c}{\partial x} + \dots + \\ & + (-1)^s \frac{(k+\alpha_m)^s}{s!} \Delta x^s \frac{\partial^s c}{\partial x^s} + \dots \end{aligned} \quad (A2)$$

where all derivatives are evaluated at (j,n) .

Quadratic and higher-order derivatives in time can be expressed in terms of spatial derivatives of the same order, by trivial manipulation based on the form of the advection equation. Indeed,

$$\frac{\partial^2 c}{\partial t^2} = \frac{\partial}{\partial t} \left[-u \frac{\partial c}{\partial x} \right] = -u \frac{\partial}{\partial x} \left[\frac{\partial c}{\partial t} \right] = u^2 \frac{\partial^2 c}{\partial x^2}$$

⋮

$$\frac{\partial^s c}{\partial t^s} = (-1)^s u^s \frac{\partial^s c}{\partial x^s} \quad (A3)$$

Replacing Equation (A3) into Equations (A1), and, then, Equations (A1) and (A2) into (3), we obtain, after some rearrangement, the local equilibrium (at j,n)

$$\frac{\partial c}{\partial t} + u \frac{\partial c}{\partial x} = \epsilon \quad (\text{A4})$$

where ϵ represents the truncation error of the BMC algorithm, and is expressed as

$$\begin{aligned} \epsilon = & \frac{1}{\Delta t} \left\{ -(1-I_0) + \left[\beta - (k \cdot I_0 + I_1) \right] \cdot \Delta x \cdot \frac{\partial c}{\partial x} - \right. \\ & - \frac{1}{2!} \left[\beta^2 - (k^2 I_0 + 2kI_1 + I_2) \right] \cdot \Delta x^2 \cdot \frac{\partial^2 c}{\partial x^2} + \dots + \\ & \left. + \frac{(-1)^{q+1}}{q!} \left[\beta^q - \sum_{r=0}^q \frac{q!}{r!(q-r)!} I_r k^{(q-r)} \right] \cdot \Delta x^q \frac{\partial^q c}{\partial x^q} + \text{H.O.D} \right\} \quad (\text{A5}) \end{aligned}$$

with

$$I_r = \sum_{p=P_1}^{P_2} p^r \cdot \phi_p(\alpha) \quad (\text{A6})$$

The necessary and sufficient condition for the coefficient affecting the q^{th} derivative in the expression of the truncation error to be null is that

$$\sum_{r=0}^q \frac{q!}{r!(q-r)!} I_r k^{(q-r)} = \beta^q \equiv (k+\alpha)^q \quad (\text{A7})$$

or, equivalently

$$I_r \equiv \sum_{p=P_1}^{P_2} p^r \cdot \phi(\alpha) = \alpha^r \quad \text{for } r \leq q \quad (\text{A8})$$

The interesting consequence is that, for any interpolation function of effective degree Q (i.e., able to exactly represent, within the domain of validity, any polynomial of degree Q), all derivatives of order $r \leq Q$ have zero coefficients, and the expression for the truncation error can be re-written as

$$\begin{aligned} \epsilon = & \frac{(-1)^Q}{(Q+1)!} (\alpha^{Q+1} - I_{Q+1}) \frac{\Delta x^{Q+1}}{\Delta t} \frac{\partial^{Q+1} c}{\partial x^{Q+1}} + \\ & + \sum_{q=Q+2}^{\infty} \left\{ \frac{(-1)^{q+1}}{q!} \left[\sum_{r=Q+1}^q \frac{q!}{r!(q-r)!} k^{q-r} (\alpha^{q-r} - I_r) \right] \right\} \end{aligned} \quad (\text{A9})$$

The proof is easily established. Indeed, if a generic polynomial of degree Q ,

$$Y(\alpha) = \sum_{s=0}^Q a_s \cdot \alpha^s \quad (\text{A10})$$

can be exactly described by the interpolator used in the BMC, we have

$$Y(\alpha) = \sum_{p=P_1}^{P_2} \left\{ \phi_p(\alpha) \cdot \sum_{s=0}^Q a_s \cdot \alpha_p^s \right\} = \sum_{s=0}^Q \left\{ a_s \cdot \sum_{p=P_1}^{P_2} \alpha_p^s \cdot \phi_p(\alpha) \right\} \quad (\text{A11})$$

Now, comparison of Equations (A10) and (A11) shows that, if the latter holds, Equation (A8) has necessary to hold, i.e., the Q^{th} coefficient in Equation (A5) must be zero. However, if Equation (A11) applies for Q , it also applies for $Q-1, Q-2, \dots, 0$, i.e., all coefficients must also be zero for $q < Q-1$, which leads to Equation (A9).

We note that $(\alpha^{Q+1} - I_{Q+1})$ is a polynomial in α of degree $\eta = \max \{M, Q+1\}$, where M is the maximum exponent of α in the interpolation functions ϕ_p ; hence Equation (A9) can be re-written as Equations (5) and (6).

APPENDIX B

Derivation of a stability criterion for the case of 3-node core elements

For stability, we require that the modulus of the cumulative error in the propagation of any Fourier component of the solution, evaluated at time N , be finite in the limit of $N \rightarrow \infty$ (for all α).

From Table 2 (Equation T3), this is equivalent to requiring that the limit of $|p_m(\alpha, N) \cdot g_m(\alpha) + g_m(\alpha, N) \cdot g_m(\alpha + \ell)|$ be finite. Because $g_m(\alpha)$ and $g_m(\alpha + \ell)$ are independent on N , the actual condition for stability is that $|p_m(\alpha, N)|$ and $|g_m(\alpha, N)|$ have finite limits, or, equivalently, that the real and imaginary parts of both $p_m(\alpha, N)$ and $g_m(\alpha, N)$ have finite limits.

Now, from Equations T4 and T5 (Table 2)

$$\begin{bmatrix} \text{Re}\{p_m(\alpha, N)\} \\ \text{Im}\{p_m(\alpha, N)\} \\ \text{Re}\{g_m(\alpha, N)\} \\ \text{Im}\{g_m(\alpha, N)\} \end{bmatrix} = \underline{\underline{R}} \cdot \begin{bmatrix} \text{Re}\{p_m(\alpha, N-1)\} \\ \text{Im}\{p_m(\alpha, N-1)\} \\ \text{Re}\{g_m(\alpha, N-1)\} \\ \text{Im}\{g_m(\alpha, N-1)\} \end{bmatrix} \quad (\text{B.1})$$

where matrix $\underline{\underline{R}}$, described by Equation 15 of the text, is independent of N .

The iteration process described by Equation B.1 is of the same form of iteration processes for the solution of linear systems of equations, which convergence requires (necessary and sufficient condition), for any initial values of p_m and g_m , that all eigenvalues of the matrix R be less than

unity in modulus (e.g., see [F1]). This is then the stability criterion for the BMC, for the case of quadratic core elements.

THE CHOICE OF THE INTERPOLATOR FOR THE BACKWARDS METHOD OF CHARACTERISTICS

by

António Melo Baptista

Massachusetts Institute of Technology

Cambridge, Massachusetts 02139

January 1987

ABSTRACT

This paper compares different interpolators for the solution of the advection equation by the Backwards Method of Characteristics (BMC), taking as major references the stability and accuracy of this method. Interpolators that enforce inter-element continuity of concentrations but not of its derivatives (class C_0) offer over interpolators of higher classes the advantage of not impacting on the cost of other operations required for the solution of the advection or the full transport equation (tracking of characteristic lines, and, if appropriate, solution of dispersion). A choice of C_0 interpolators, using information from both the region of validity of the interpolation and its immediate neighborhood, can match the accuracy of the cubic Hermite interpolator (of class C_1), which has, in the context of the BMC, been taken as a reference for quality; the handling of boundary conditions, and, especially, the extension of these non-compact C_0 interpolators to irregular grids, pose some practical difficulties, which are interpolator-dependent and require further research, but should not limit usefulness.

1. INTRODUCTION

Eulerian-Lagrangian methods (ELM) have been increasingly used for the solution of the advection-dominated transport equation, in the context of engineering and environmental problems. While significant differences exist among available ELM, most split advection from dispersion, solving each of the resulting sub-equations by an appropriate specific technique: typically, dispersion is solved by a centered finite-differences or finite-elements method, and advection is solved by the Backwards Method of Characteristics (BMC).

The solution of advection is very often critical for the accuracy and cost of the overall procedure, in spite of the great attractiveness of the BMC. Introduced by [L1], this method constitutes a direct numerical implementation of the physical meaning of the advection equation, i.e. of the fact that concentrations remain constant along characteristic lines

that follow the flow. The procedure is illustrated in Figure 1, and involves, in each time step, Δt , two complementary tasks:

- The backwards tracking of particles assigned, at time n , to each node, j , of the computational grid, so as to find the position ("foot") of the respective characteristic line at time $n-1$, F_j .

- The interpolation of all nodal concentrations at time n , $c(j,n) \equiv c(F_j,n-1)$, from the known nodal concentrations at time $n-1$, $c(j,n-1)$.

The nature and difficulties of these two tasks are substantially different. The tracking requires, for each grid node, the solution between times n and $n-1$ of an ordinary differential equation of the form

$$\frac{dx}{dt} = u_i(x_i, t) \quad (i=1, \dots, \text{number of dimensions}) \quad (1)$$

and is independent of the actual concentration field. This independence has two important consequences: (a) the time-stepping strategy for the tracking, between n and $n-1$, is unconstrained by the global time-step, Δt ; and (b) the error in the tracking can be evaluated, by looping back to time n and comparing the departure and arrival locations of the particle associated to the characteristic line.

Direct control can therefore be exercised on the tracking accuracy, by adjusting the time-stepping strategy, by trial-and-error, so as to meet some imposed error bound. A 2-D tracking algorithm conceived with this philosophy, and allowing for complex flows and irregular grids, is built in the Eulerian-Lagrangian finite element model ELA ([B2]), which has been

used for the simulation of pollutant transport in coastal waters (e.g., [B2], [K3]); acquired experience indicates that, while expensive when compared to other steps of the solution procedure, the tracking does not compromise feasibility, even if overly strict error bounds are imposed (e.g., of the order of the centimeter, for computational grids with characteristic nodal spacing of the order of one hundred meters).

By contrast, the interpolation to find the concentrations at the feet of the characteristic lines is relatively inexpensive, but: (a) may introduce significant errors if the grid can not, due to practical constraints, be refined as much as it would be necessary to resolve prevailing gradients of concentration in the flow direction; and (b) does not allow for any simple accuracy control, as errors refer to the basic unknown of the problem, concentrations. The search for an accurate interpolator has been the object of intense, but somewhat discoordinated research ([H1-H3], [K1-K2], [B3], [B9]), based mostly on numerical experimentation; by their variety and specificity, results are difficult to apprehend and compare, and a correct perspective of the potentials and limitations of available interpolators is much needed.

This study contributes to such understanding, by providing both a qualitative comparison of relevant categories of interpolators, and a detailed quantitative analysis of interpolators of class C_0 (i.e., those that ensure inter-element continuity, but not inter-element differentiability). Both formal tools (Taylor series and Fourier analysis) and numerical experimentation (based on standard test problems) are used in

the analysis, which takes as a reference the case pure-advection by a constant velocity, in uniform grids.

2. INTERPOLATION STRATEGIES: GENERAL OPTIONS AND IMPLICATIONS

2.1. Global interpolators

An intuitive approach to interpolation is to pass a polynomial over all the nodes of the computational grid, so as to fit exactly their concentrations, and, then, sample the polynomial at the feet of all characteristic lines. This approach has, however, definite shortcomings:

- The definition of the interpolator involves the solution of a system of equations, characterized in general by a matrix that is time-independent and full, and whose size is constrained by the total number of nodes of the computational grid. While the matrix has to be factorized only once (each time step requiring only a backsubstitution), the fullness of the matrix makes the procedure impractical for most grids, in most computers other than main-frames.

- The order of the interpolator depends on the number of computational nodes, and will often be in the hundreds or thousands. As a consequence, and given the form of the terms in the matrix (assuming a 1-D grid, $a_{ij} = x_i^{j-1}$, with j from 1 to the number of nodes), very large and very small terms will co-exist, leading most likely to poor accuracy in the solution of the system of equations. Also, high order polynomials, fit to non-polynomial functions, tend to wiggle between nodes, while approximating non-

polynomial functions, which is a source for numerical instability, or at least, significant phase errors.

An alternative, keeping up with the idea of a global interpolator, is to fit all nodal concentrations only in a least square sense, taking, for instance, a polynomial of specified order as a reference. In this case, storage requirements drop by several orders of magnitude, being determined by the order of the selected polynomial; required CPU time per time step increases, but, as no huge initial matrix factorization is necessary, total CPU time should be decreased drastically. However, such global interpolators can not ensure that nodal concentrations will be exact at the grid nodes, and, therefore, from the formal analysis of [B7], they are likely to lead to inconsistency--i.e., the numerical approximation to the advection equation may not match the exact differential equation as Δx and Δt approach zero.

To the author's knowledge, no attempt has ever been made to use global interpolators in a BMC context. While, historically, this may have been biased by the fact that the BMC was developed to support discrete numerical methods (based on the idea of local expansions of functions and its derivatives and/or integrals), the present analysis suggests that the use of global interpolators is, indeed, unlikely to be generally attractive.

2.2. Local interpolators

The common strategy for interpolation is the piece-wise definition of local interpolators, each valid only within a specified region of the

computational domain (core element). Potential advantages of local interpolators, relative to global interpolators, include much lower expected costs, unconditional consistency, control of stability, and ability to increase the order of interpolation as needed, over the computational domain.

In the following discussion of local interpolators we will always assume that the core element for the interpolation is defined so as to contain the foot of the characteristic line, which has become a standard procedure ([H2-H3], [B1-B7], [N1-N3], etc). The alternative is defining the core element based on the position of the head of the characteristic line ([G1], [H1], [L1]), but the former procedure has the advantage of forcing concentrations to be found always by interpolation (as opposed to extrapolation, which typically leads to instability), without the need to restrict the time step through a Courant-type criterion.

Local interpolators that have been used in a BMC context may, for convenience, be grouped into categories, depending upon whether the interpolator is compact or non-compact, and upon its class. Compact interpolators use information only from nodes within the core element, while non-compact interpolators may use also outside information. The class of an interpolator refers to its differentiability along the boundaries of the core element, with regard to the interpolator of the adjacent core element: C_p , with $p \geq 0$ implies continuity of concentrations and of its p lowest derivatives. Four general categories of interpolators are of interest.

Compact interpolators of class C_n

Compact interpolators of class C_0 , in the form of Lagrangian polynomials, have been used by [B2] and [B9], among others. They constitute a convenient choice, because they use concentrations as unique dependent variables and pose no special problems in the handling of boundary conditions, regions near boundary conditions, or irregular grids.

However, the optimal order for these interpolators is quadratic, as linear polynomials introduce excessive numerical damping (in the form of a numerical diffusion), and cubic and higher order polynomials lead to the instability of the BMC, as a consequence of the progressive increase of the number of nodes of the core element. Even quadratic interpolators have, however, a limited ability to resolve sharp gradients of concentration, and may therefore require the use of grids with very small nodal spacing, leading to unacceptable CPU and memory requirements).

Compact interpolators of class C_1 and above

For compact interpolators of class C_1 and above, the order of the interpolation is increased by requiring further information from each node of a same core element, rather than by extending the core element. This concentrates the information used to construct the interpolator in a closer vicinity of the foot of the characteristic line, and constitutes a both more effective and safer form of improving accuracy; indeed, as a rule, interpolators of class C_{1+} will be more accurate than interpolators of class C_0 of the same order, and will not involve the risk of leading to the instability of the BMC.

However, additional dependent variables (concentration derivatives) are now introduced to the transport problem, which has, in practice, limited the choice within this type of interpolators to cubic Hermite polynomials, of class C_1 ([H1-H2]).

While cubic Hermite polynomials have often been implicitly accepted as a quality standard for BMC interpolators, they have not become a popular choice in engineering practice, and recent research has concentrated in finding equally accurate alternatives ([H3], [K1-K2], [B3-B5]). This is due to the extra work associated with the handling of the additional variable(s)-- c_x in 1-D, c_x , c_y and c_{xy} in 2-D, and c_x , c_y , c_z , c_{xy} , c_{yz} and c_{xz} in 3-D--which, more than the interpolation itself, impacts the cost of the solution of dispersion (by increasing the order and the bandwidth of the matrices that have to be solved), and the cost of the tracking step. To understand the extra cost in the tracking step, let us take as a reference the 1-D case; unlike concentrations, which are kept constant, c_x varies along the characteristic lines, as

$$\frac{\partial}{\partial x} \left(\frac{Dc}{Dt} \right) = 0 \iff \frac{D}{Dt} (c_x) = -u_x \cdot c_x \quad (2)$$

Hence, a new ordinary differential equation is introduced per new variable, and has to be solved along with the equation describing the characteristic line. Even if, as in [H2], cross-derivatives are estimated, rather than treated as dependent variables, the additional computational burden may be unacceptable for two or more dimensions, if the tracking is performed as accurately as it should (we note that [H2] uses a very simple tracking,

without accuracy check; while this does not compromise cost, it seems inconsistent--other than for uniform of very smooth flows--with the accuracy of the interpolator).

Non-compact interpolators of class C_n

Non-compact interpolators of class C_0 aim to increase the order of the interpolation by also using information on concentrations outside the core element, rather than by increasing the size of the core element or by introducing derivatives as additional variables. They are expected to avoid both stability problems and high costs; the price to be paid is the need for special treatment of regions near boundaries (where available information is insufficient), and a non-trivial extension to irregular grids; also, it is expected that, for equal degree of the interpolating polynomial, they will be less accurate than interpolators of class C_1 , because they use information farther away from the foot of the characteristic line.

The simplest form of non-compact interpolators of class C_0 is a Lagrange polynomials, defined over the immediate neighborhood of the core element. Earlier works have used quadratic ([L1]) and cubic ([L2]) polynomials, defined over linear core elements; we will show, however, that the approach becomes more attractive by using quartic or higher order polynomials, defined either over linear (polynomials of odd degree) or quadratic (polynomials of even degree) core elements.

Alternative non-compact interpolators have been proposed based on the use of concentrations outside the core element to estimate the derivatives

required to define, within such element, a cubic pseudo-Hermitian ([H3], [K1-K2]) or Hermitian ([B3]) polynomial. The accuracy of these interpolators has been reported, based mainly on numerical experimentation, as very similar to that of the cubic Hermite interpolator of class C_1 . We will show, however, that most of these interpolators are actually unstable--although weakly (i.e., amplification factors close to unity), and in a way that is independent of the Courant number--and that the observed apparent accuracy is a manifestation of such instability: the amplification of some Fourier components of the solution compensates for the damping of others, for the test problems and computational parameters that have been used. The question of whether these schemes will still be useful in practice, given the fact that the BMC is a numerical method particularly well-adapted to large time steps (hence, typically, small number of time steps), is legitimate, and will be addressed in latter sections.

A third conceivable form of non-compact C_0 interpolators is based on the use of locally-defined splines. To the author's knowledge, this approach is first explored in some detail in this paper. We show, however, that, at least for natural splines, it leads to the same type of apparent high accuracy, hiding a weak instability, that characterizes non-compact interpolators based on the estimation of derivatives.

Non-compact interpolators of class C_1 and above

Just like their C_0 counterparts, non-compact interpolators of class C_1 aim to increase the order of the interpolator by using information on concentrations outside the core element, rather than by increasing the size

of the core element, or by introducing derivatives as additional variables. To do so, however, they elect to enforce the differentiability (but not the value of the derivative) of the interpolator in the boundaries of the core element.

This approach was taken by [B10], who used splines of different order and with different smoothing factors, but, in all cases defined in a global basis; the resulting interpolators are, however, local, in the sense that a different interpolation function applies over each core element. These interpolators are apparently able to match the accuracy of the cubic Hermite interpolators (class C_1), but reported results are insufficiently detailed to allow for a well-based critical analysis.

A definite drawback of these spline interpolators of class C_1+ is that, like in global interpolators, a system of equations with size linked to the total number of nodes of the computational grid has to be solved. For 1-D and straight orthogonal multi-dimensional grids, this system can be built so as to be banded (tridiagonal, for cubic interpolators), and, therefore, is relatively cheap to solve. Extension to multi-dimensional irregular grids is, however, both conceptually non-trivial, and potentially expensive.

Synthesis

A variety of local interpolators can conceivably be used in the context of the BMC. The general implications of the choice of different interpolators were briefly analysed in this section. While no specific category of interpolators can claim to be optimal, we find particularly

attractive the versatility and potential low cost of compact and non-compact interpolators of class C_0 , and will discuss in further detail, in the next sections, their potentials and limitations. In particular, we will show that some of these interpolators can match, in what seems to be an overall more efficient way--given the fact that concentration derivatives have not to be handled as additional dependent variables--the accuracy of the cubic Hermite polynomial (class C_1) proposed by Holly and co-workers [H1-H2].

3. DETAILED COMPARISON OF INTERPOLATORS OF CLASS C_0

3.1. Definitions

This section provides a detailed comparison of the performance of the EMC, for different interpolators of class C_0 , including compact and non-compact Lagrange polynomials, hybrid Hermitian-Lagrangian polynomials, and natural splines. The terms of reference are consistency, stability, and accuracy.

Table 1 defines the selected interpolators, and Table 2 illustrates, when appropriate, their generation procedure for 1-D uniform grids. To simplify the reference to individual interpolators, we will use the notation

$$nP-XXm$$

where

- n is the number of nodes that contribute to define the interpolator;

- XX is an alphanumeric, denoting the type of interpolator: LI for compact Lagrangian; LR for non-compact Lagrangian; HL for hybrid, built from the estimation of the derivatives of a Hermite polynomial; PL for hybrid, built by estimating the derivatives of a pseudo-Hermite polynomial; and SP, for splines;

- m is the number of nodes of the core element.

We note that the BMC algorithm can be written, for all these interpolators, in the general form:

$$c(j,n) \equiv c(j-\beta,n-1) = \sum_{p=P_1}^{P_2} \phi_p(\alpha) \cdot c(j-k+p) \quad (3)$$

where (see Figure 2 for reference)

j - denotes the node where the concentration is to be computed
(global notation)

n - denotes the instant where the concentration is to be computed

α - denotes the position of the foot of the characteristic line in a local coordinate system with origin at node j-k (α is associated the fractional part of the Courant number
($\beta = u \cdot \Delta t / \Delta x$)

P_1, P_2 - denote the extreme nodes of the region that is used to define the interpolator (in local notation)

ϕ_p - are elementary shape functions which, together, define the interpolator

3.2. Analysis of consistency

For interpolators leading to algorithms of the general form of Equation 1, a sufficient condition for the unconditional consistency of the BMC is that the interpolation be exact at all nodes of the core element ([B7]). This trivial condition is, in concept, obeyed by all interpolators of Table 1. In the case of the 6P-PL2, however, round-off errors were introduced by [K1] in the evaluation of the interpolation coefficients, which makes the interpolator inconsistent; indeed, the truncation error becomes of the form

$$\epsilon = (3.7 \alpha - 4.2) \cdot 10^{-4} c + \dots \quad (4)$$

and obviously does not vanish even if $\Delta x, \Delta t$ go to zero. The inconsistency is weak, and not of a fundamental nature; it can be removed, by recalculation of the proper coefficients of the interpolator, if this is found to be otherwise attractive. The results shown in latter sections are for the interpolator as proposed by [K1], hence reflect (and are used to illustrate) the effect of inconsistency.

3.3. Analysis of stability

Unlike consistency, stability proves to be a major factor of distinction among interpolators. In [B7], we provide the theoretical basis

for the analysis of stability of BMC with interpolators of class C_0 (assuming uniform flows and grids). Quite conventionally, we define stability as the ability of the numerical method to propagate without amplification all Fourier components with wavelength above $2\Delta x$; to assess this property we derived criteria that extend the conventional Von Newman criterion to quadratic core elements. These criteria account for the fact that interpolators with quadratic or higher order core elements internally generate energy transfer between Fourier components, until an equilibrium is reached (which often requires several tens to a few hundreds time steps). Only after this equilibrium is reached is the amplification factor per time-step time-independent.

Figure 3 shows amplification factors per time step, as a function of α^- and the dimensionless wavelength $L_m/\Delta x$, for the different interpolators of Table 1. For interpolators with linear core elements, these factors refer to any time step, while for quadratic core elements they refer to the equilibrium state (and were actually computed as the ratio between cumulative errors after 400 and 399 time steps). For the 4P-LI4, the only of the considered interpolators that has a higher than quadratic (cubic) core element, the amplification factors shown correspond to the first time step.

Taking those figures as a reference, we find that:

- For Lagrangian interpolators, with linear or quadratic core elements, the BMC is stable, regardless of the actual order of the

polynomial. By contrast, instability arises for cubic compact Lagrangian polynomials (we note that Figure 3 can only suggest the instability of the 4P-LI4, as it shows amplification factors in the first time step, rather than at a time where energy transfer between Fourier components has stopped; however, numerical experimentation confirms the instability).

- From the hybrid Hermitian-Lagrangian polynomials (H1, PL), only the 4P-HL2 leads to stability. The instability associated with the other interpolators is however very weak, in the sense of amplification factors very close to unity. A significant difference among these interpolators is the size and the location of the zones of the $\alpha - L_m/\Delta x$ plane that induce amplification; those zones have a minimal extent (and concentrate in the smaller wavelengths) for the 8P-PL2, but affect most of the plane in the other cases. The round-off errors in the calculation of the 6P-PL2, discussed in Section 4.2), are showing up strongly in the symmetry of the amplifying factors.

- The two interpolators based on natural splines lead also to instability, this being stronger than for HP and PL interpolators, in the sense that the amplification factors are larger (but still close to unity).

3.4. Analysis of accuracy

3.4.1. Numerical damping and numerical dispersion

Numerical damping and numerical dispersion result, respectively, from the amplitude and celerity (or phase) errors in the propagation of individual Fourier components of the solution. No numerical method is able

to avoid both types of errors, and the balance between them is determinant for accuracy.

We note, to avoid common misinterpretation, that both amplitude and celerity errors can independently lead both to the reduction of peak concentrations, and to a wiggly behavior of the solution (resulting, in particular, in negative concentrations). However, amplitude errors are responsible for much more significant peak reductions, and the wiggles that they generate, unlike those due to celerity errors, are quickly damped and preserve symmetry.

Figure 4 both illustrates the above aspects and suggests that, typically, amplitude errors strongly dominate celerity errors in BMC solutions, for centered interpolators (i.e., where the core element is centered within the region that contributes to the definition of the interpolator); this is not necessarily so for non-centered interpolators, such as the 3P-LR2 used by [L1] (not considered in this study), for which celerity errors may become quite significant.

The general dominance of amplitude errors for centered interpolators is a result of the fact that the regions of the $\alpha-L_m/\Delta x$ plane that induce significant celerity errors typically overlap with those that have larger amplitude damping. Indeed, any wave that tends to travel at a speed significantly different from the flow velocity will be quickly damped, contributing little to the numerical solution.

Because of this dominance, a meaningful comparison among the interpolators of Table 1 can be performed only on the basis of their

amplification factors, already displayed in Figure 3 (see [B5] for similar information on celerity errors). Rather conventionally, all interpolators tend to handle better larger than shorter wavelengths, and the difference among interpolators can be evaluated by examining how well they preserve the latter.

Such preservation is a function of both the amplification factors per time step, and of the number of time steps, N . This is illustrated in Figure 5, which display the actual amplifications of different interpolators after 10, 100, 1000 and 10000 time steps ($\alpha = 0.5$).

For alternative unstable interpolators, the effect of the instability may become significant after a very different number of time steps, as illustrated by the point in time where the amplification factors for short wavelengths start blowing up in Figure 5 (note that this point depends, for each interpolator, on α); spline interpolators are, in particular, seen to be of rather limited practical value. In the actual numerical solution of a given problem, the blow-up of amplification factors for individual Fourier components may or may not be of practical significance, depending on their magnitude relative to the dominant components, and on the range of α 's that, as a combination of the characteristics of the flow and of the grid, will most frequently be called upon in the interpolations

On the other extreme, after one thousand time steps, the linear Lagrangian interpolator significantly damps out even wavelengths fifty times

longer than Δx . The solution will look nicely smooth, but it will likely be unacceptably diffused.

In general, the ability to handle short wavelengths improves as the order of the interpolator increases, improvements being more dramatic for low-order interpolators: e.g., the differences between the amplification factors for the 2P-LI2 and the 3P-LI3 are much easier to be accepted as practically relevant than those between the amplification factors for the 6P-LR2 and the 8P-LR2.

It is particularly interesting to compare cubic interpolators of class C_0 (4P-LR2, 4P-HL2, 4P-SP2, 5P-HL3, 5P-SP3, 6P-PL2 and 8P-PL2), among themselves and with the cubic Hermite interpolator of class C_1 . To assist on this comparison, we use as a reference the test problem of constant advection of a Gauss-hill, for which solutions are presented in Figure 6, for a choice of interpolators.

We observe a clear overall best performance of the C_0 interpolators involving the largest number of nodes, with interpolators with five or more nodes actually outperforming the cubic Hermite interpolator of class C_1 , 2P-HI2. From the charts of amplification factors shown earlier, this outperformance is, however, recognized to be somewhat artificial and mostly uncontrolled, as it results from the compensation of the damping of some wavelengths by the amplification of others; for a different problem, or for the same problem at a latter time, the amplification effect may become dominant, and lead to poor accuracy and instability. This emphasizes the risks of the comparison of numerical methods based only on numerical experimentation, which, necessary as it may be, is unable to systematically

identify critical conditions, and may be misleading regarding general conclusions.

Also interesting is to compare interpolators based on the same number of nodes. We take the case of interpolators with five nodes (5P-LR3, 5P-HL3, and 5P-SP3). The amplifying factors for the 5P-LR3, which is a quartic interpolator, are significantly different from those of the cubic interpolators 5P-HL3 and 5P-SP3 essentially by the fact that deviations from unity lead allways to damping, rather than to amplification. This suggests a superior reliability of the 5P-LR3, especially for long-term calculations (and according to on-going research, to non-uniform grids), but may cost it a somewhat poorer performance for short-term calculations (e.g., Figure 7). We note, however, that even in this case the 5P-LR3 has an accuracy comparable to that of the 2P-HI2.

3.4.2. Response to grid refinement

We recall from [B7] that C_0 interpolators lead to truncation errors of the form:

$$\epsilon = P(\alpha) \cdot \frac{\Delta x}{\Delta t}^{Q+1} \frac{\partial^{Q+1} c}{\partial x^{Q+1}} + \text{H.O.D.} \quad (5)$$

where Q denotes the effective order of the interpolator, i.e., the degree of the highest degree function that the interpolator can exactly fit (Q is at most equal to the actual order of the interpolator, M , being smaller whenever a lower order polynomial is involved in the definition of the interpolator—e.g., cases of the 6P-PL2 and the 4P-HL2). $P(\alpha)$ is an

interpolator-dependent polynomial in α , of degree $\tau_i = \max \{N, Q+1\}$; if the Courant number is large, α and, hence, $P(\alpha)$, are essentially independent of Δx , Δt ; if Courant number is below unity, though, $P(\alpha)$ is an explicit function of Δx , Δt . The consequence, in terms of the dependence of truncation errors on Δt and Δx can be examined, for selected interpolators, in Tables 4 and 5.

The effective order of the interpolator is seen to play a key role in accuracy. Indeed, from this expression, we observe that, given the relationship between moments and derivatives, the Q^{th} moment of the concentration distribution is the highest moment that can a priori be expected to be exactly propagated by the BMC, regardless of Δx (assuming that no aliasing is introduced). Also, the largest expected rate of convergence of truncation errors to zero, is, for fixed Δt , characterized by Δx^{Q+1} , and occurs for large Courant number (for small Courant numbers, $\alpha = u \cdot \Delta t / \Delta x$, and, therefore, the convergence rate should be reduced).

These aspects are illustrated by considering again the problem of the constant advection of a Gauss-hill, which we solve with a 3P-LI3 interpolator ($Q = 2$); the time step and the number of time steps is kept fixed, but we let the grid spacing vary. Results are presented in Figure 8, in the form of error norms concerning global accuracy and preservation of moments up to the 3rd. We note that, while both the preservation of the third moment and the global accuracy strongly depend on Δx (and global errors, as measured by the L2-error norm, do decrease, in the zone of larger Courant numbers, roughly as Δx^3), moments of up to order 2

(representing, respectively, mass, mean displacement and spreading) are preserved in an essentially exact way. Extrapolations to non-uniform grids and/or flows should be considered with caution. On-going investigation tentatively shows that both the global accuracy and the preservation of individual moments are sensitive to non-uniformity--e.g., Figure 9--but different interpolators and different types of non-uniformity affect accuracy differently [B5].

To contrast non-compact Lagrange interpolators with other types of non-compact interpolators, we take 5-node interpolators as a reference to note that, while for the 5P-LR3 the information from outside the core element was used primarily to increase (relative to the 3P-LI3) the order of the interpolation to quartic, for the 5P-HL3 and 5P-SP3 that information was used in part to reduce $P(\alpha)$ and in part to increase the order of the interpolation to cubic. For an unspecified Δx , neither of the choices is necessary better; however, the 5P-LR3 is expected to become progressively more accurate than its counterparts as the Courant number increases, and as Δx decreases.

4. FINAL CONSIDERATIONS

None of the C_0 interpolators considered in detail in the previous section can be recognized as optimal, in a general sense. However, given a specific problem (characterized in particular by the gradients of the concentration field--which may vary in time, especially if diffusion plays a significant role--and by the magnitude of the carrying flow), and a numerical grid (often constrained by available computational resources),

some interpolators or combination of interpolators, present important advantages over others, which this study may help recognize.

Among compact interpolators, the 3P-LI3 will often be the best option, as lower order interpolators (2P-LI2) introduce significant numerical damping, and higher order interpolators (e.g., 4P-LI4) are unstable. While the accuracy of the 3P-LI3 may be appropriate in most of the spacial domain of most problems of practical interest, concentrations in localized zones of steep gradients (e.g., vicinity of sources, edge of advancing fronts, etc) may result unacceptably deformed.

In these zones, we may consider the use of a more accurate interpolator. The C_1 cubic Hermite interpolator, 2P-HI2, is a legitimate option, but impacts the cost not only of the interpolation procedure, but, especially, that of the tracking and (when appropriate) of the solution of diffusion; furthermore, using the 2P-HI2 in only a localized region in space is unnatural, as concentration derivatives are, for this interpolator, handled as additional dependent variables.

By contrast, non-compact interpolators are natural and effective in providing local accuracy improvement. If, as we recommend and are assuming here, the 3P-LI3 is taken as the basic interpolator, non-compact interpolators sharing its quadratic core element are particularly convenient for implementational purposes; if the 2P-LI2 is taken as the reference, non-compact interpolators with linear core elements will be preferable.

Among non-compact interpolators, the 5P-LR3 presents a number of advantages that may suggest it as a coherent first-choice: conceptual simplicity, minimal amount of required outside information, unconditional stability, and accuracy comparable to the 2P-HI2. A variety of options do exist, though, and common-sense use of the accuracy information contained in this paper may help making specific decisions for specific transport problems.

Extension of non-compact interpolators to multi-dimensional uniform grids poses no special problem; for instance, in 2-D, a double sweep technique, with m interpolations being carried first in one direction, and one final interpolation being carried in the orthogonal direction, as been successfully used [H3,K1,B3-B5]. The extension to irregular grids, of any dimensionality, is more challenging: in addition to potential loss of accuracy (which is shared by compact interpolators, and by any Eulerian numerical method for solution of the transport equation), it involves, in different scales for different interpolators, some ambiguity in the definition of the interpolator, and risk of unattractive costs. Preliminary results and conceptualization suggest [B5] that the relative attractiveness of the 5P-LR3 interpolator may result strengthened, but further research and actual implementation and application of this and other interpolators in industrial codes is deemed necessary to clarify the relevant issues.

REFERENCES

- A1 Adams E.E., R. Kossik, and A. M. Baptista. "Source Representation in a Numerical Transport Model." *Finite Elements in Water Resources: Proceedings of the 6th International Conference* (Sá da Costa et al., Ed.). Springer-Verlag, 1986.
- A2 Aldama-Rodríguez, A. "Theory and Applications of Two- and Three-Scale Filtering Approaches for Turbulent Flow Simulation." Ph.D. Thesis, M.I.T., 1985.
- B1 Baptista, A. M., E. E. Adams, and K. D. Stolzenbach. "The 2-D Unsteady Transport Equation Solved by the Combined Use of the Finite Element Method and the Method of Characteristics", in *5th Int. Conf. on Finite Elements in Water Resources*. Burlington, Vermont, 1984a.
- B2 Baptista, A. M., E. E. Adams, and K. D. Stolzenbach. "Eulerian-Lagrangian Analysis of Pollutant Transport in Shallow Water." MIT R. M. Parsons Laboratory, Technical Report 296, 1984.
- B3 Baptista, A. M., E. E. Adams, and K. D. Stolzenbach. "Comparison of Several Eulerian-Lagrangian Models to Solve the Advection-Diffusion Equation." in *Int. Symp. on Refined Flow Modeling and Turbulence Measurements*. U. Iowa, USA, 1985.
- B4 Baptista, A. M., E. E. Adams, and K. D. Stolzenbach. "Accuracy Analysis of the Backwards Method of Characteristics." in *Finite Elements in Water Resources: Proceedings of the 6th International Conference*. (Sá da Costa et al., Ed.), Springer-Verlag, 1986.
- B5 Baptista, A. M. "Accurate Numerical Modeling of Advection-Dominated Transport of Passive Scalars." LNEC, Lisbon, 1986.
- B6 Baptista, A. M. "Fourier Analysis of the Backwards Method of Characteristics." Ph.D. Thesis, M.I.T., 1987.
- B7 Baptista, A. M. "The Consistency, Stability, and Convergence of the Backwards Method of Characteristics." Ph.D. Thesis, M.I.T., 1987.
- B8 Baptista, A. M. "The accuracy of Eulerian-Lagrangian Methods" Ph.D. Thesis, M.I.T., 1987.
- B9 Benque, J. P., G. Labadie, and G. Ibler. "A Finite Element Method for Navier-Stokes Equations." *3rd Conf. on Fin. Elem. in Flow Problems*, 1980.
- B10 Branski, J. M. "Higher-Order Spline Schemes for the Advection - Diffusion Equation." Submitted to ASCE, 1986.

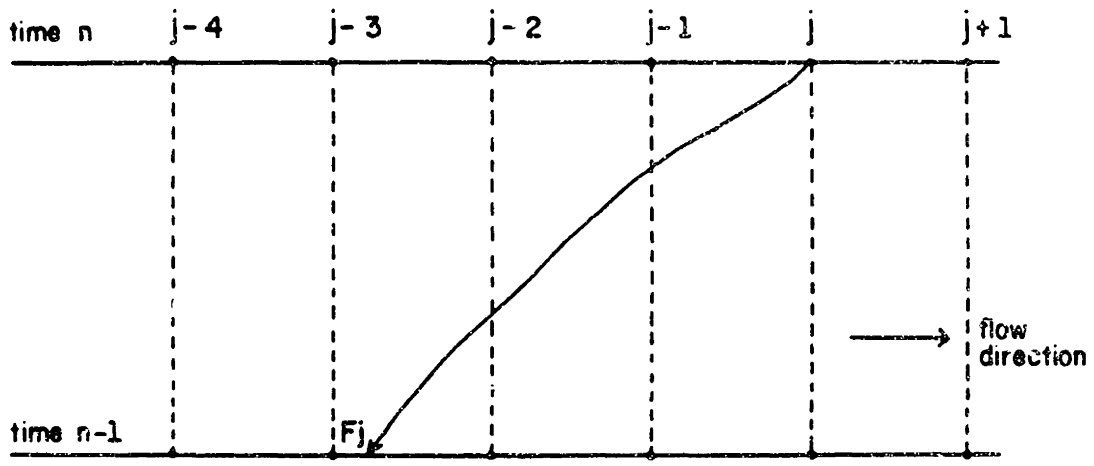
- C1 Celia, M. A., and W. G. Gray. "An Improved Isoparametric Transformation for Finite Element Analysis." *International Journal for Numerical Methods in Engineering* 20:1443-1459, 1984.
- C2 Cheng, R. T., V. Casulli, and S. Milford. "Eulerian-Lagrangian Solution of the Convection-Diffusion Equation in Natural Coordinates." *Water Resources Research* 20(7):944-952, 1984.
- G1 Glass, J., and W. Rodi. "A Higher Order Numerical Scheme for Scalar Transport." *Comp. Math. in Appl. Mech. and Eng.* 31:337-358, 1982.
- H3 Holly, F. M., Jr., and T. Komatsu. "Derivative Approximations in the Two-Point Fourth-Order Method for Pollutant Transport." *Proceedings of the Conference on Frontiers in Hydraulic Engineering*, ASCE, M.I.T., 1984.
- H2 Holly, F. M., Jr., and J. M. Polatera. "Dispersion Simulation in 2-D Tidal Flow." *Journal Hydr. Engrg.*, ASCE, 1984.
- H1 Holly, F. M., Jr., and A. Preissmann. "Accurate Calculation of Transport in Two Dimensions." *Journal of the Hydraulics Division*, ASCE 103(HY11):1259-1278, Nov. 1977.
- K1 Komatsu, T., F. M. Holly, Jr., and N. Nakashiki. "Numerical Calculation of Pollutant Transport in Rivers and Coastlines", in 4th Congress, Asian and Pacific Division, IAHR, Chiang Mai, Thailand, 1984.
- K2 Komatsu, T., F. M. Holly, Jr., N. Nakashiki, and K. Ohgushi. "Numerical Calculation of Pollutant Transport in One and Two Dimensions." *Journal of Hydroscience and Hydraulic Engineering* 3(2):15-30, 1985.
- K3 Kossik, R. F. "Tracing and Modeling Pollutant Transport in Boston Harbor." M.Sc. thesis, MIT, 1984.
- L1 Leith, C. E. "Numerical Simulation of the Earth's Atmosphere." *Methods in Computational Physics* 4:1-28, 1965.
- L2 Laboratoire National d'Hydraulique - Internal reports, 1970-1980
- N1 Neuman, S. P. "An Eulerian-Lagrangian Scheme for the Dispersion Convection Equation Using Conjugate Space-Time Grids." *Journal of Comp. Phys.* 41:270-279, 1981.

Figure 1

Illustrative sketch for the Backwards Method of Characteristics

Required steps:

1. Tracking of the characteristic lines. For each node j , a characteristic line is independently defined by the backwards (i.e., between n and $n-1$) solution of an ordinary differential solution of the form $dx_1/dt = u_1$.
2. Interpolation at the feet of the characteristic lines. The concentration at the foot of each characteristic line (and, for pure advection, at any other point of the characteristic line, including node j at time n) is found by interpolation from known information on neighboring nodes, at time $n-1$.



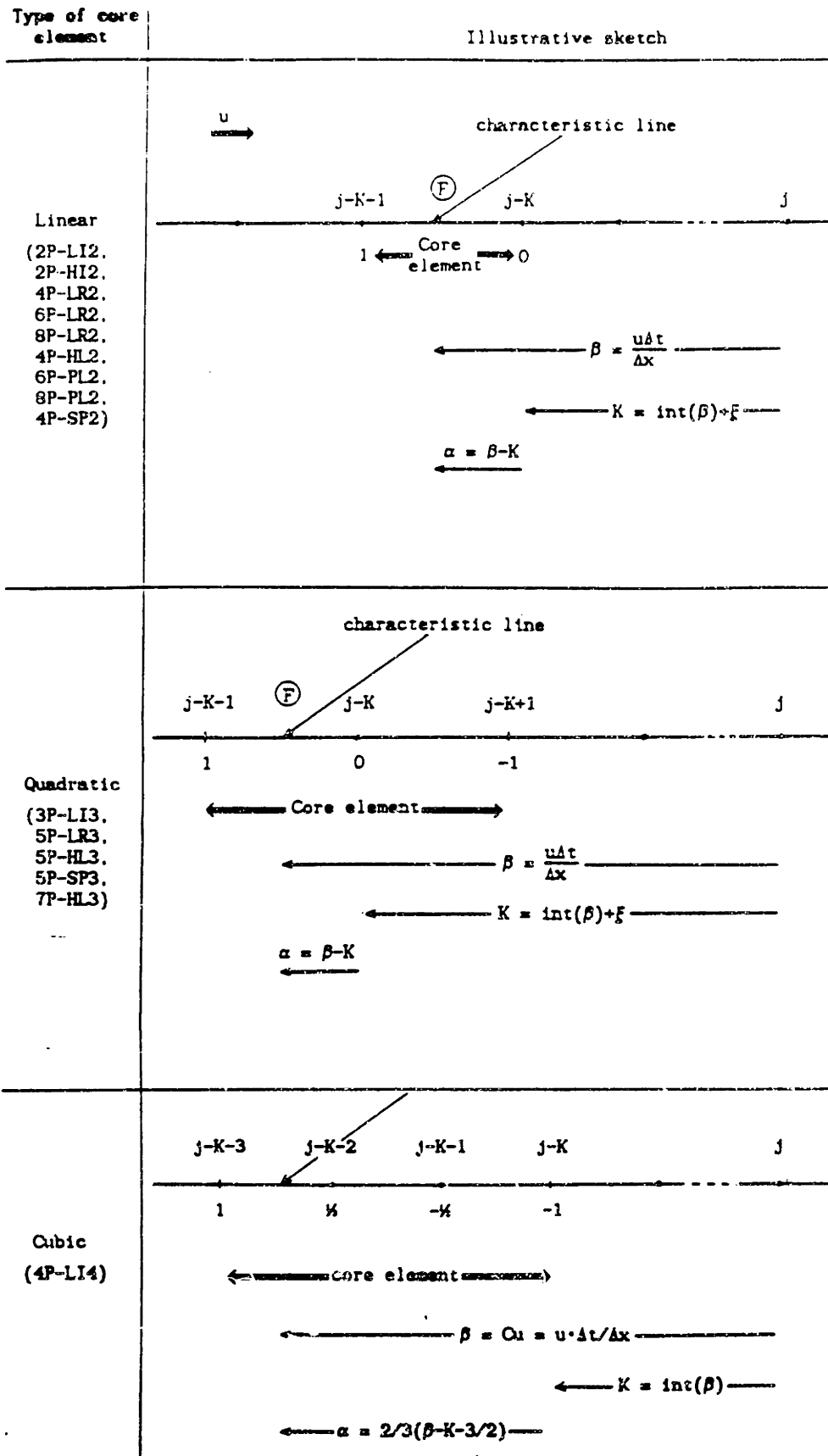


Figure 2 Definition of the core elements

Figure 3

Amplification factors per time step, as a function of the location, α , of the foot of the characteristic line within the core element, and of the dimensionless wavelength

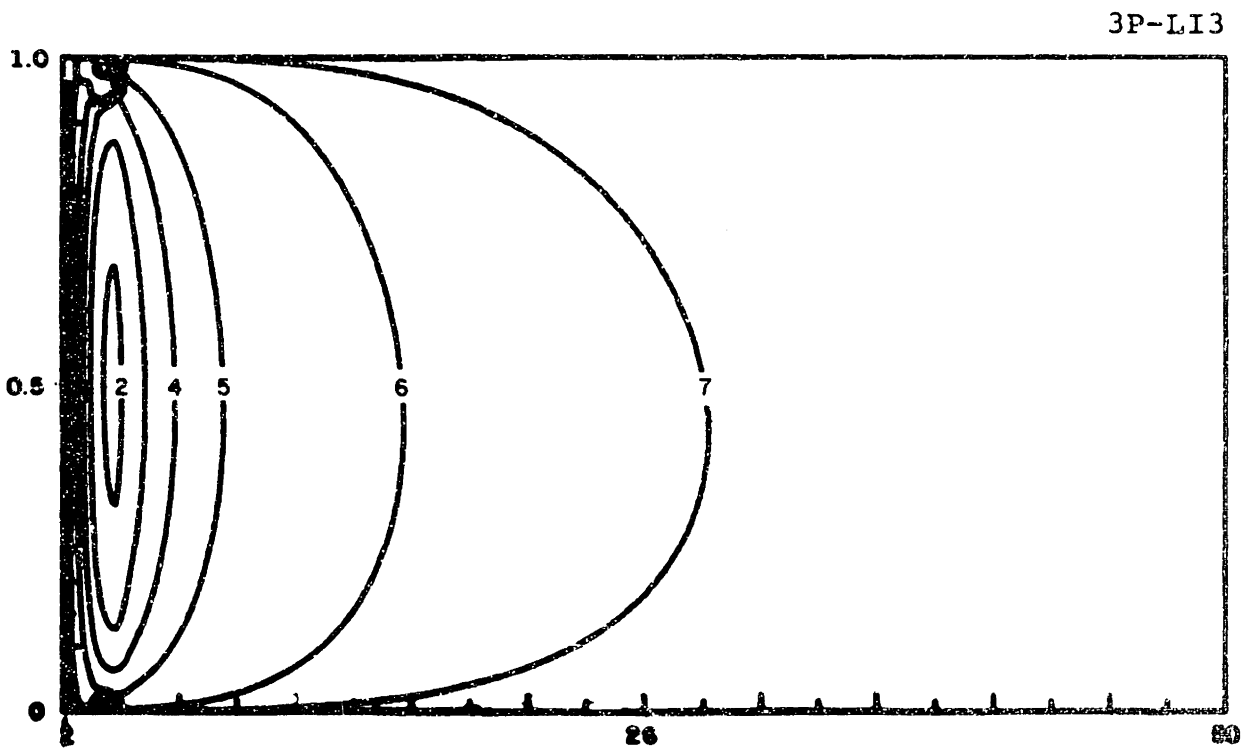
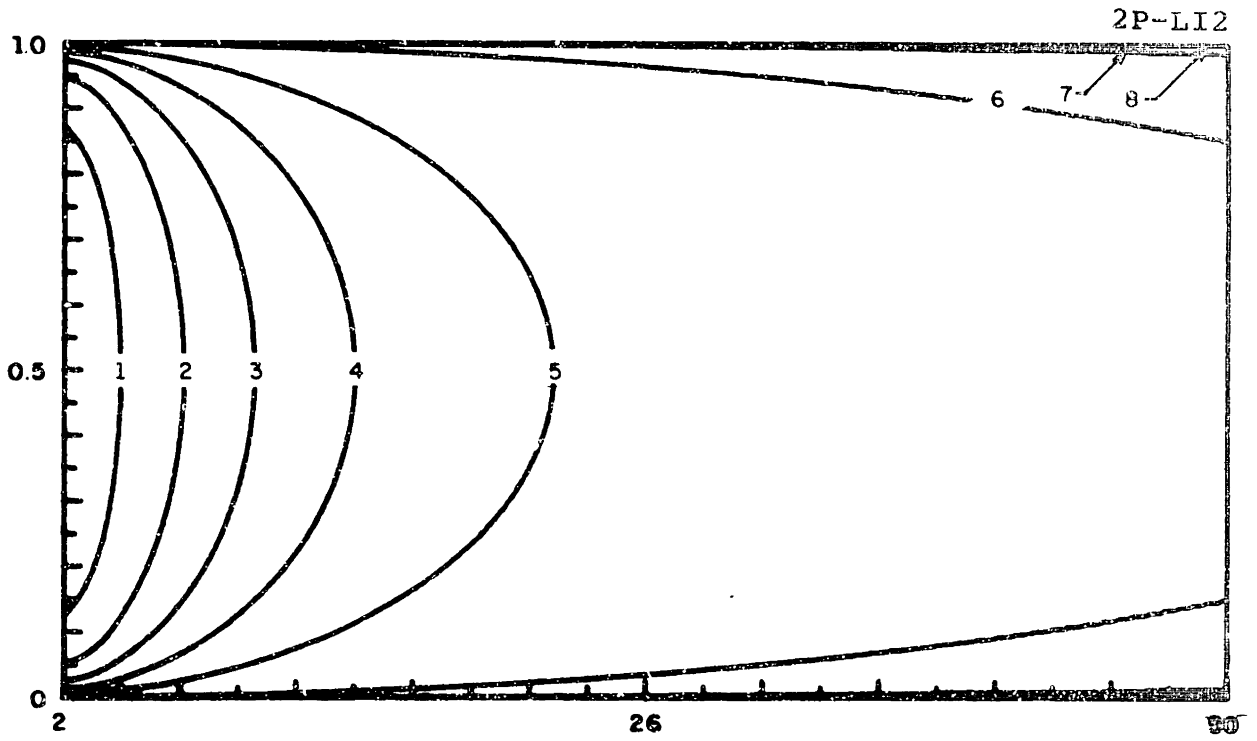
- (a) 2P-LI2 (b) 3P-LI3 (c) 4P-LI4 (d) 4P-LR2 (e) 5P-LR3
 (f) 6P-LR2 (g) 8P-LR2 (h) 6P-PL2 (i) 8P-PL2 (j) 4P-SP2
 (k) 5P-SP3 (l) 4P-HL2 (m) 5P-HL3

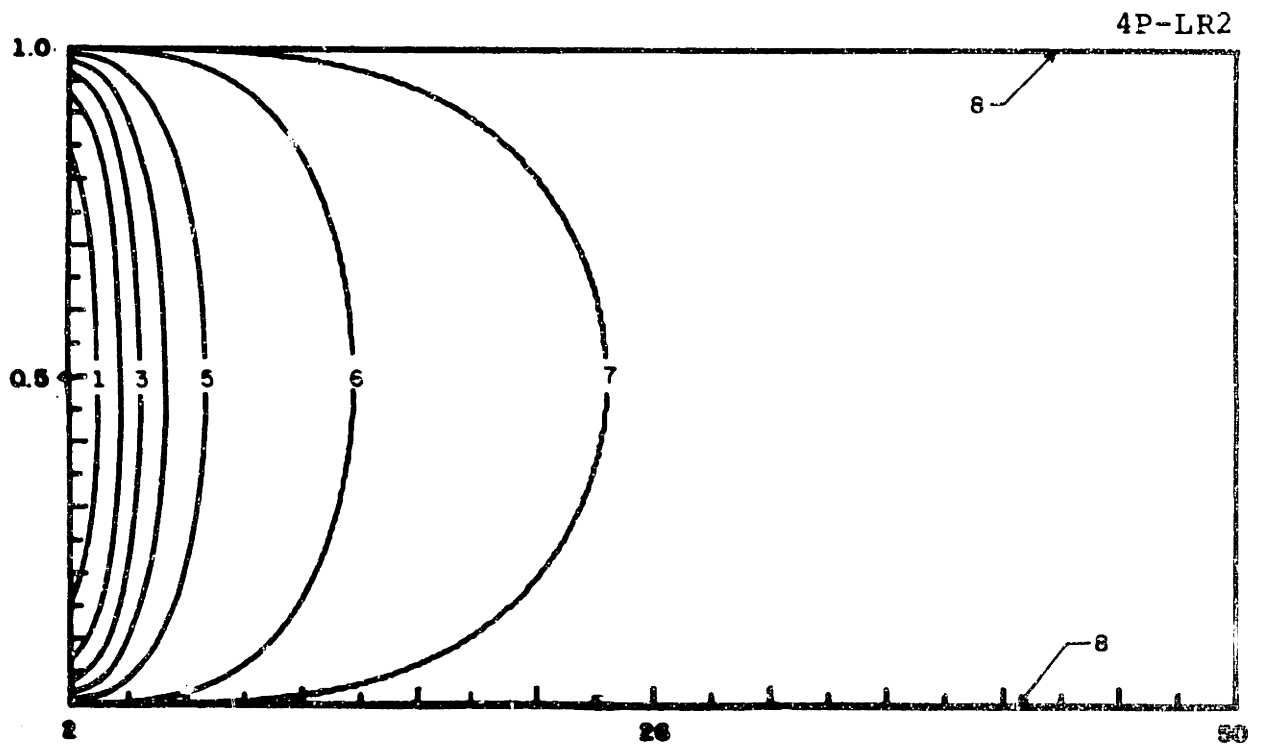
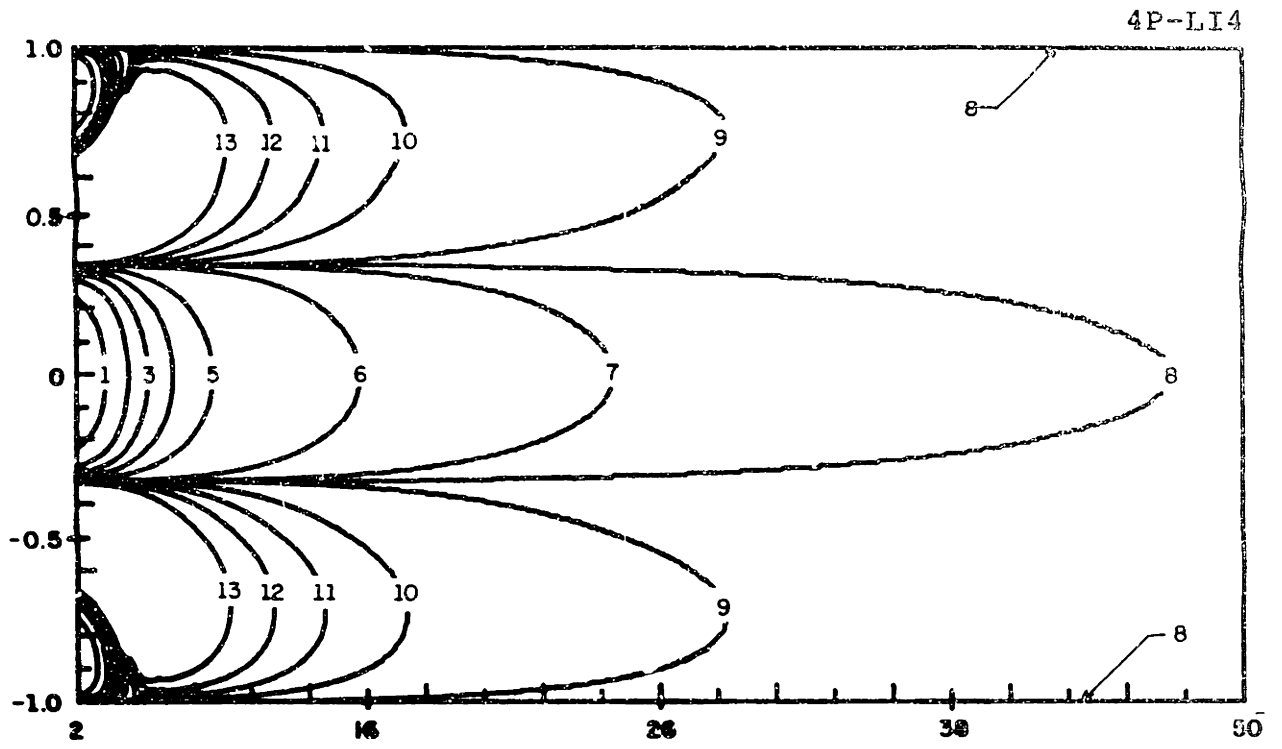
Amplification factors per time step, for interpolators with quadratic core elements, where computed as the ratio between the cumulative amplification factors after 400 and 399 time steps.

Legend

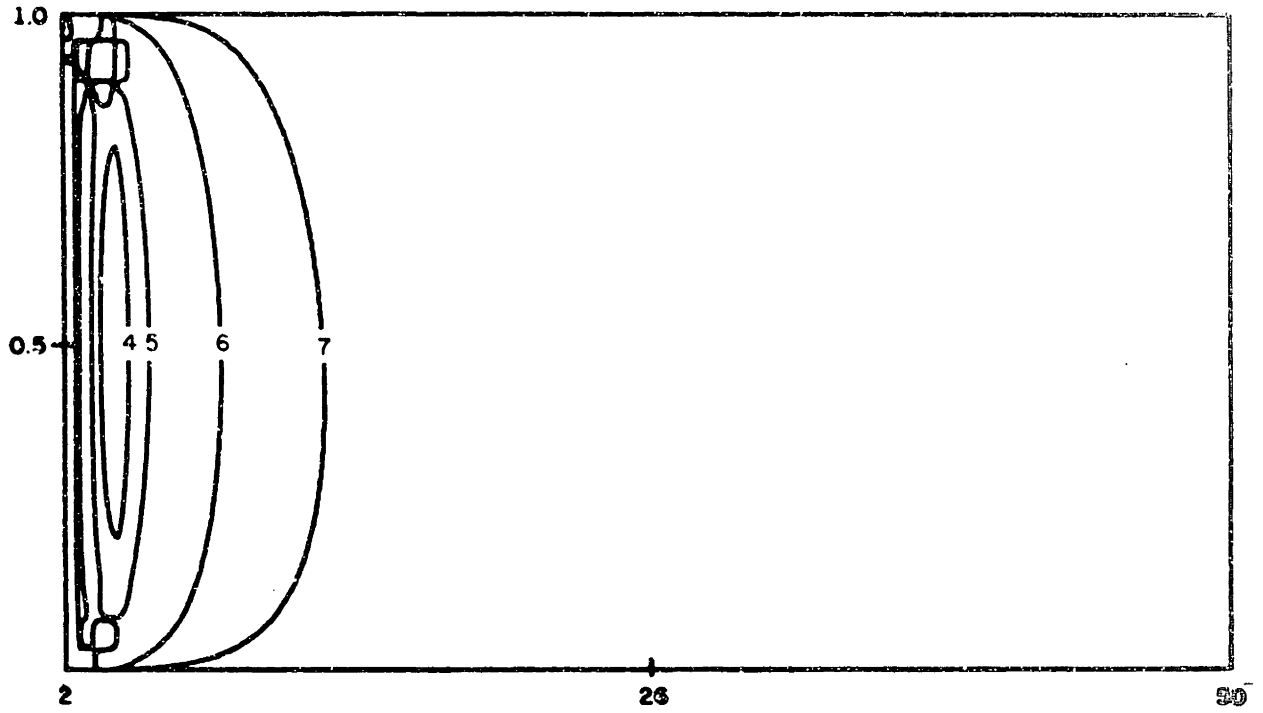
Contour	Amplification factor
1	0.75
2	0.90
3	0.95
4	0.99
5	0.999
6	0.9999
7	1.0000
8	1.0001
9	1.0010
10	1.0025
11	1.0050
12	1.0100

Note: Irregular behavior of contours near $L/\Delta x = 4$, for interpolators based on quadratic core elements, reflects limitations of the contouring code to handle prevailing gradients

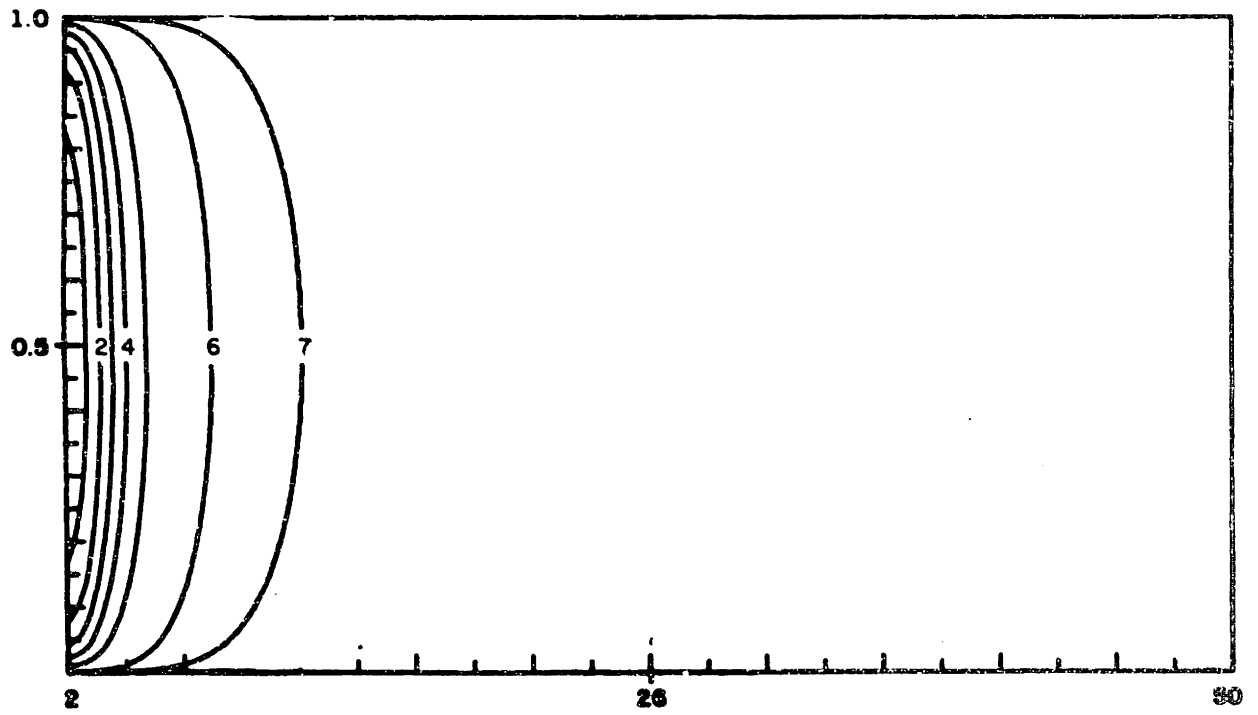




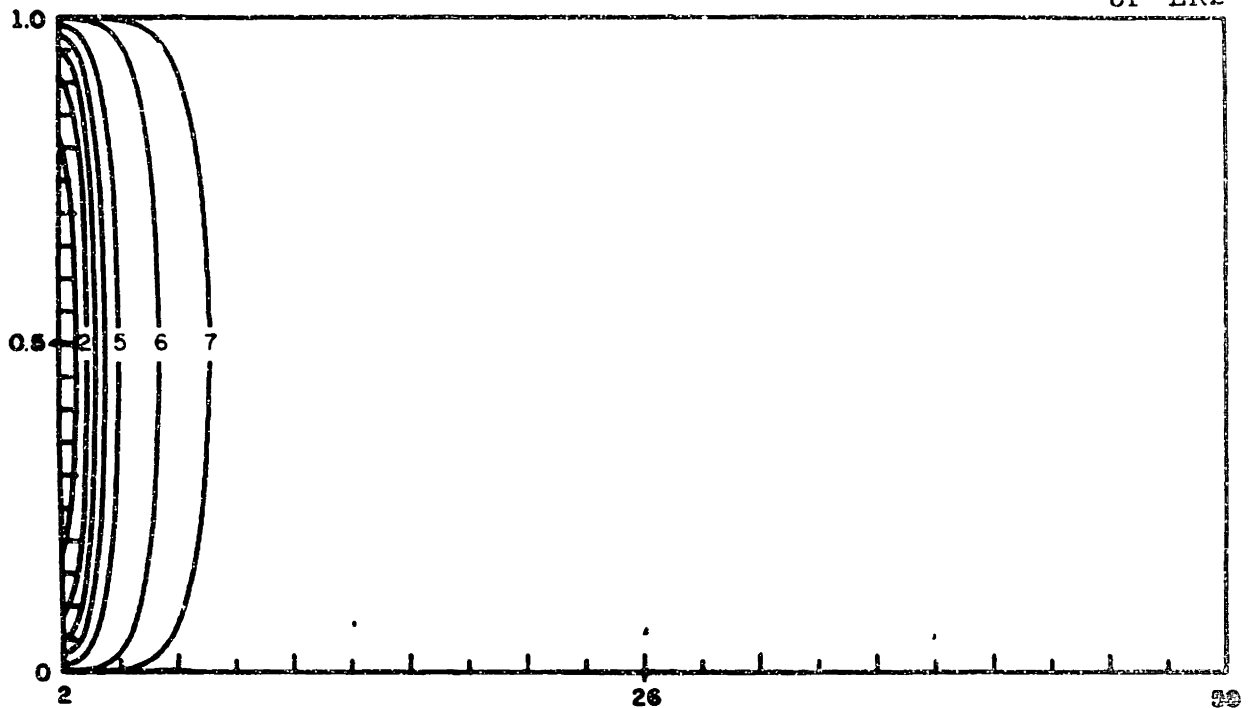
5P-LR3



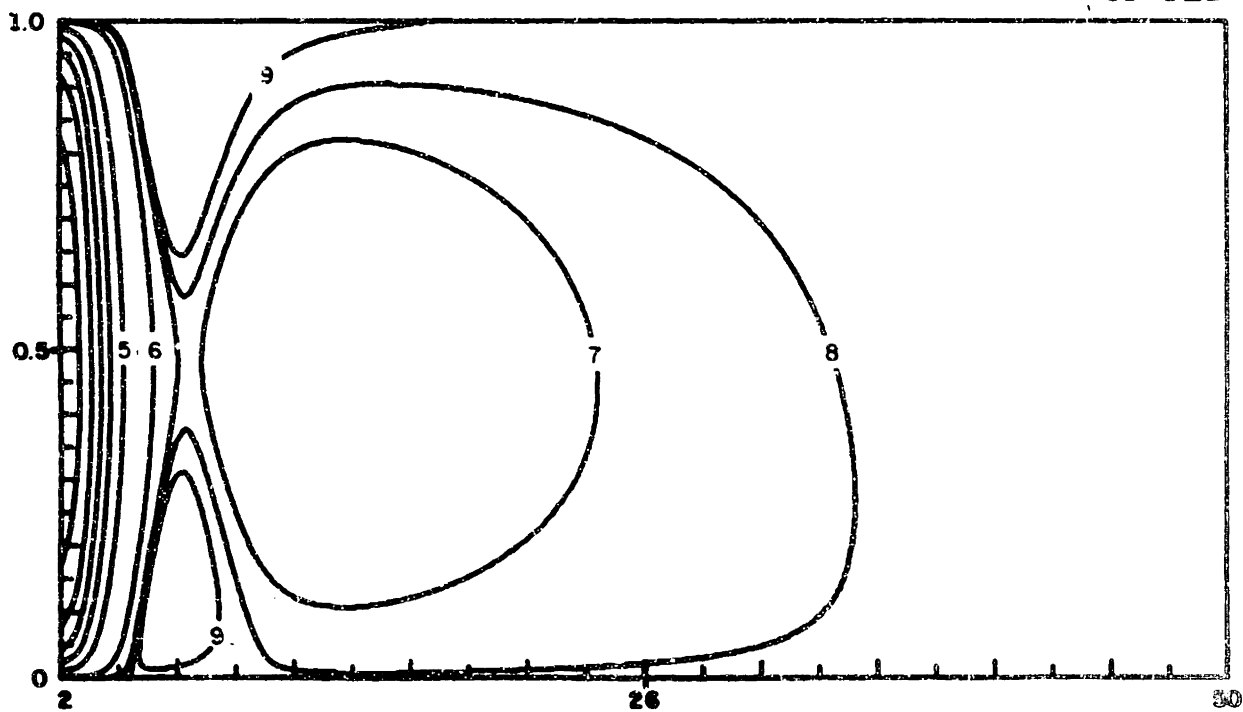
6P-LR2



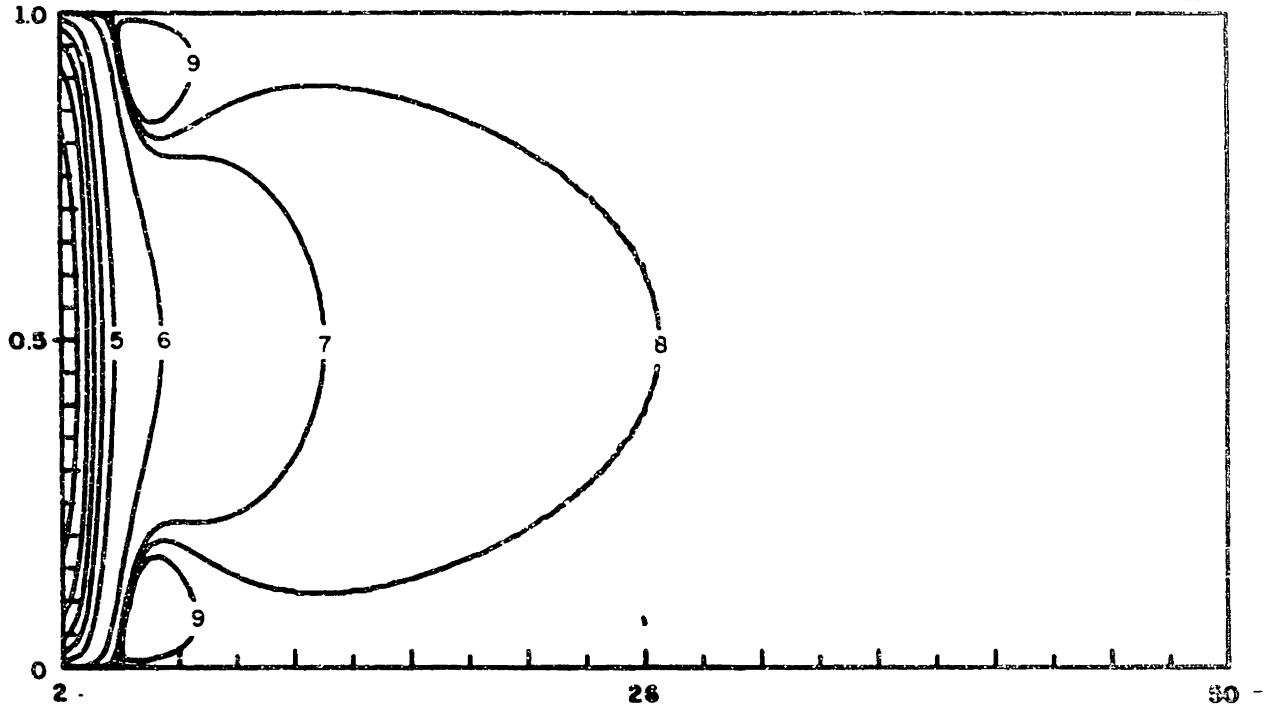
8P-LR2



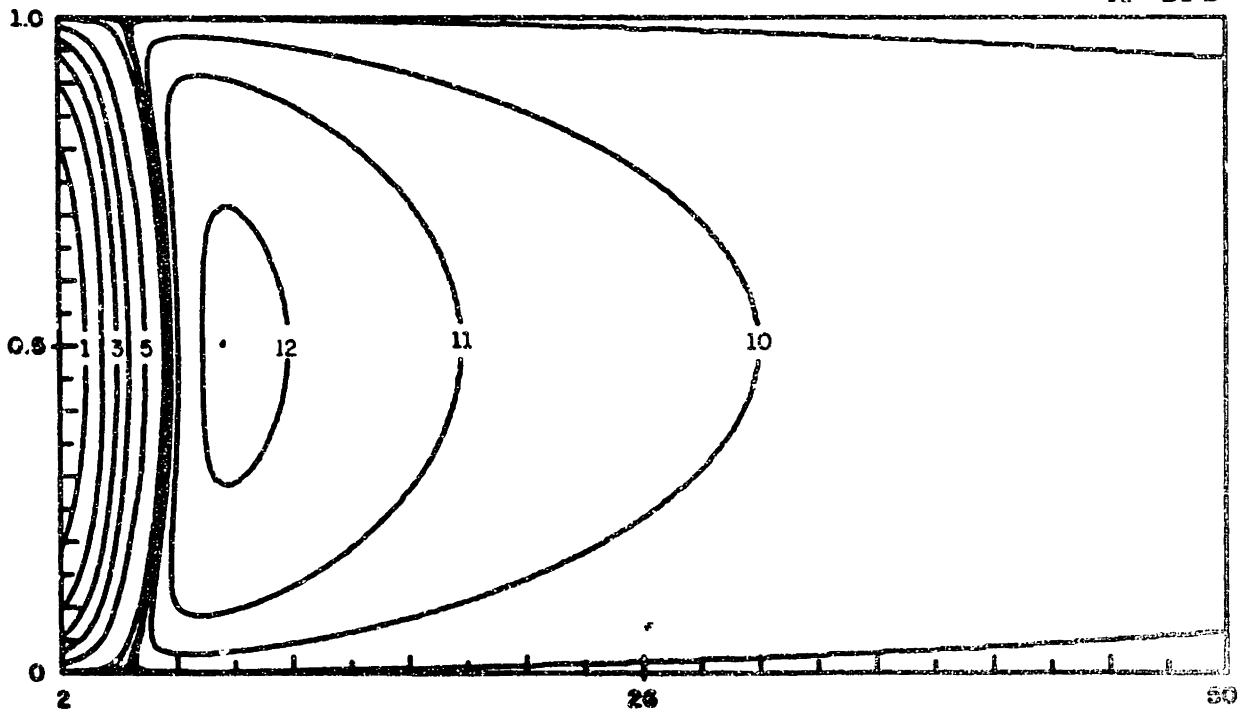
6P-PL2

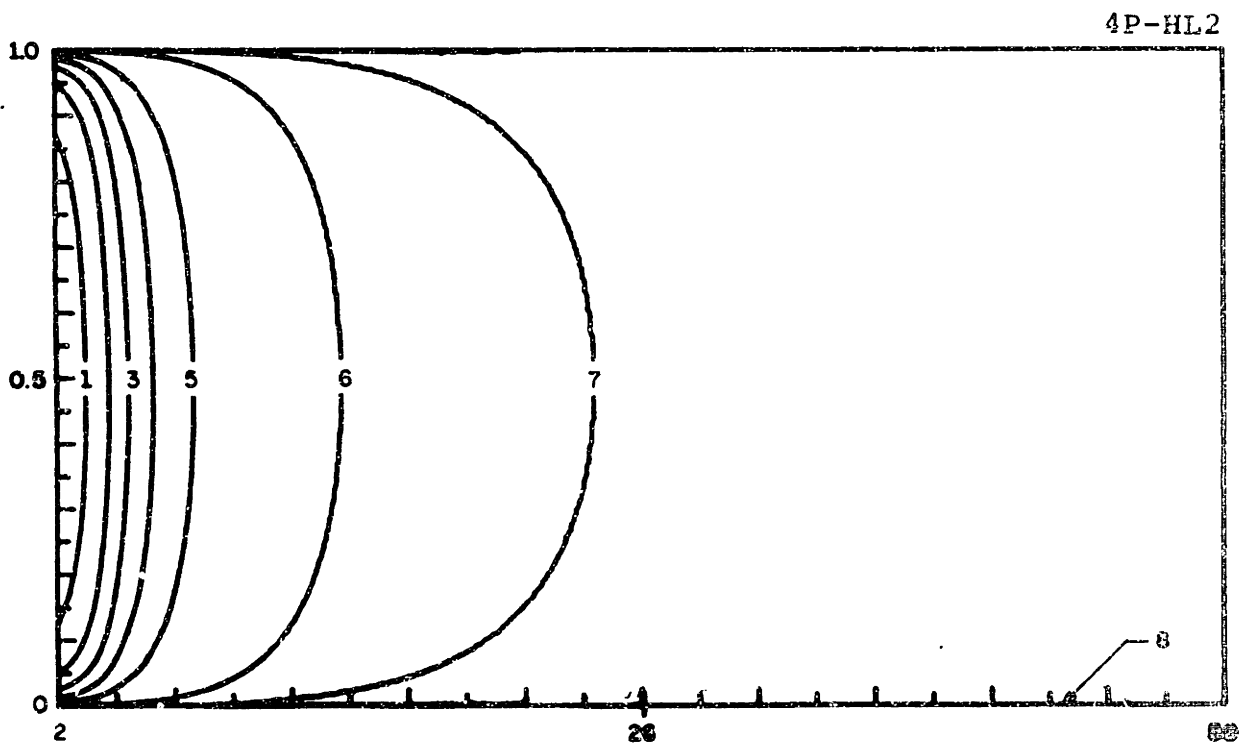
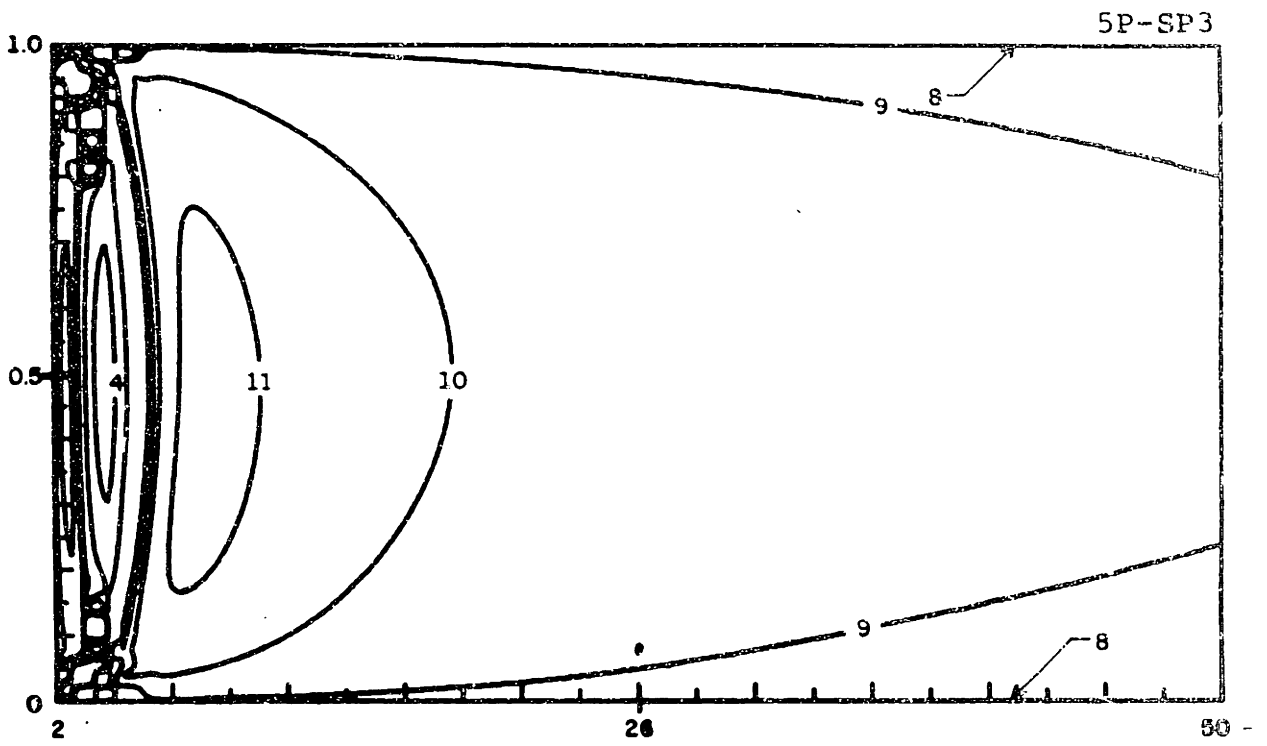


8F-PL2



4P-SP2





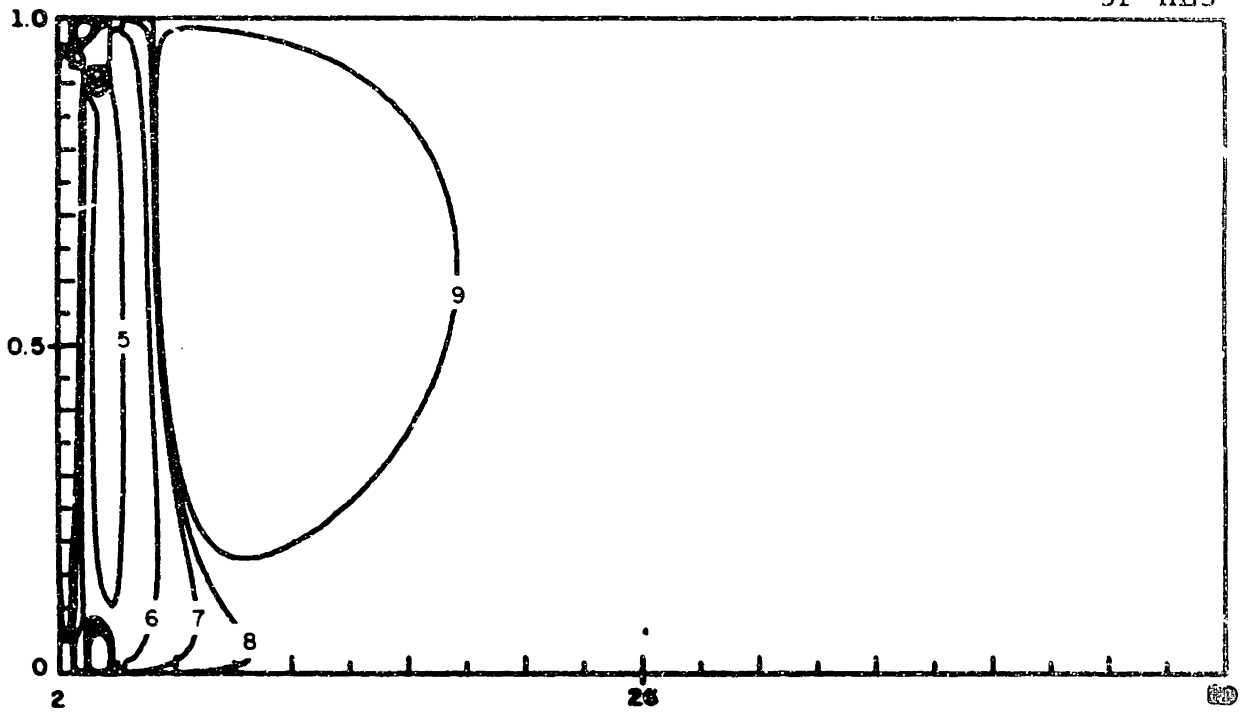


Figure 4

Illustration of the relative importance of amplitude and phase errors

- (a) 2P-LI2 (b) 3P-LI3 (c) 4P-LR2 (d) 5P-LR3
 (e) 6P-LR2 (f) 8P-LR2 (g) 6P-PL2 (h) 8P-PL2
 (i) 4P-SP2 (j) 5P-SP3 (k) 4P-HL2 (l) 5P-HL3
 (m) 7P-HL3

Reference problem: $\frac{\partial c}{\partial t} + u \frac{\partial c}{\partial x} = 0$
 $c(x, 0) = \exp\left\{-\frac{(x-x_0)^2}{2\sigma^2}\right\}$
 $c(x, t) \rightarrow 0 \quad |x| \rightarrow \infty$

Computational parameters: $\Delta t = 192$

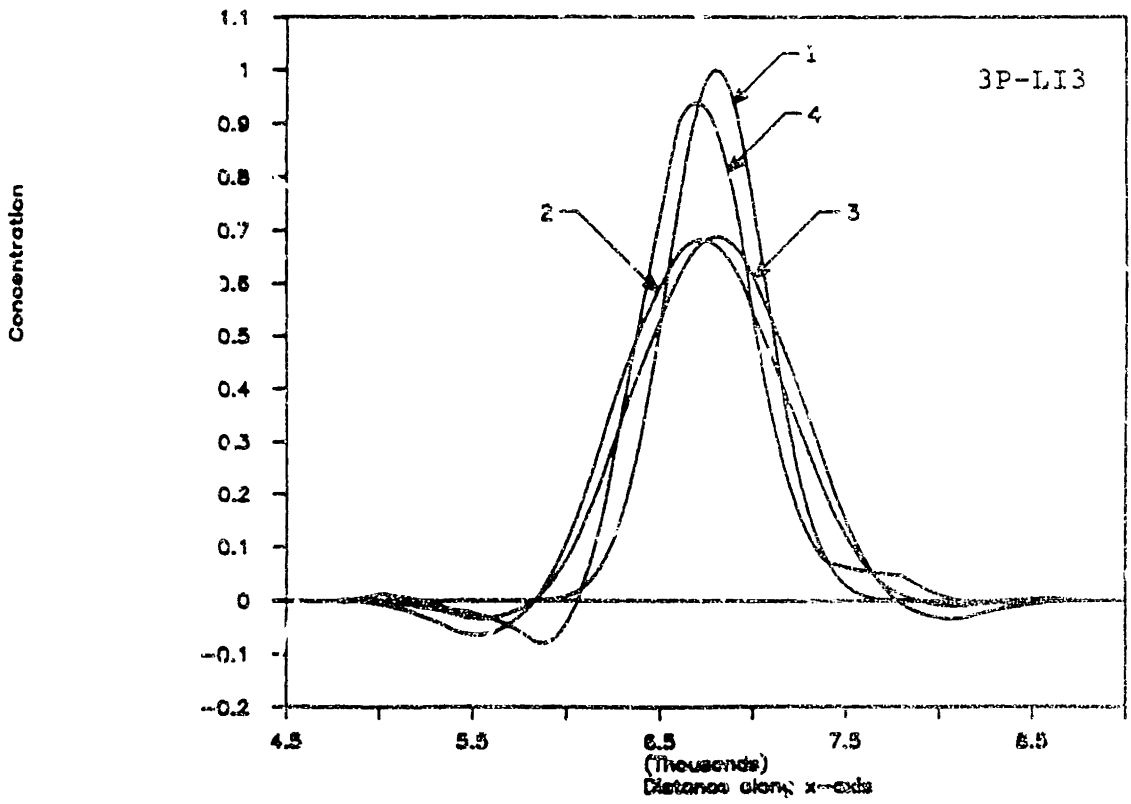
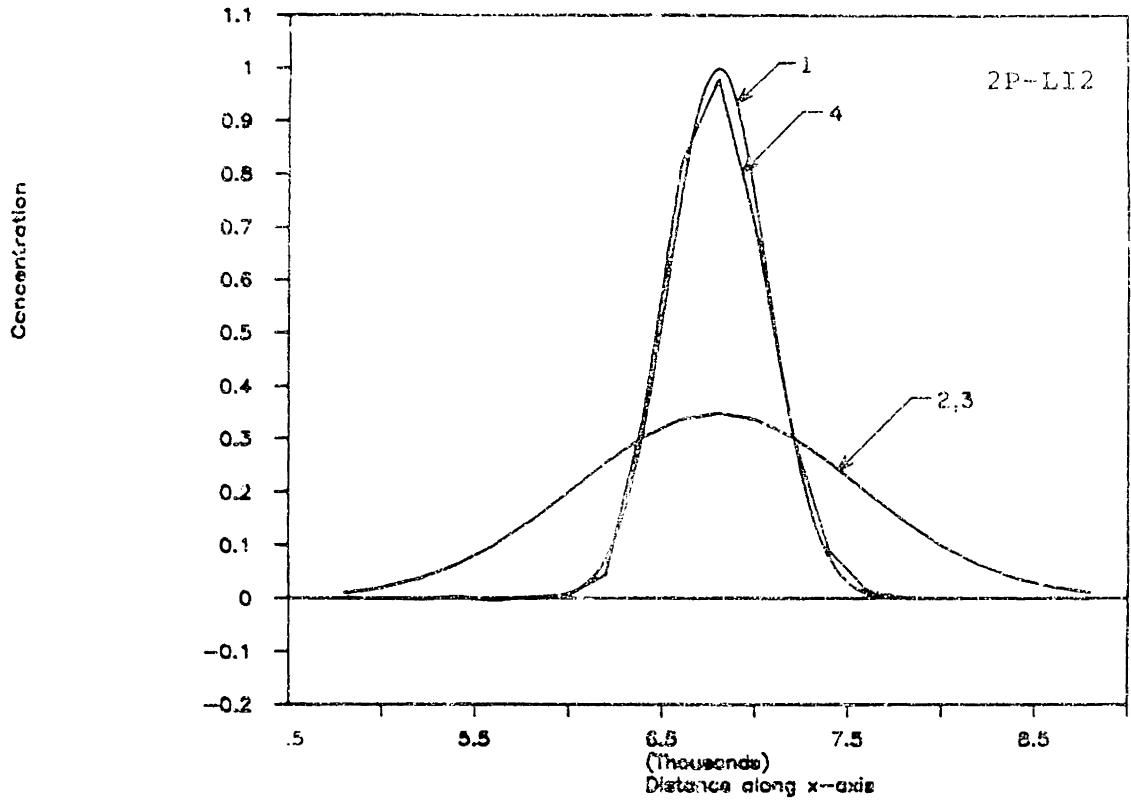
$$\Delta x = 200$$

$$N = 50$$

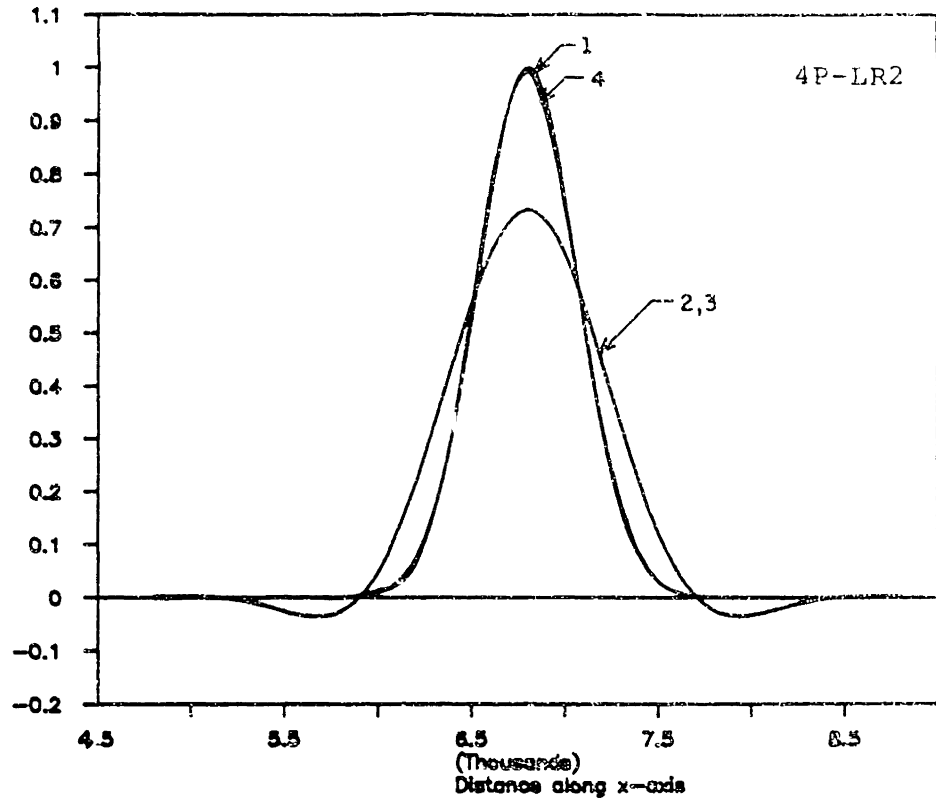
$$u = 0.5$$

$$\sigma = 264$$

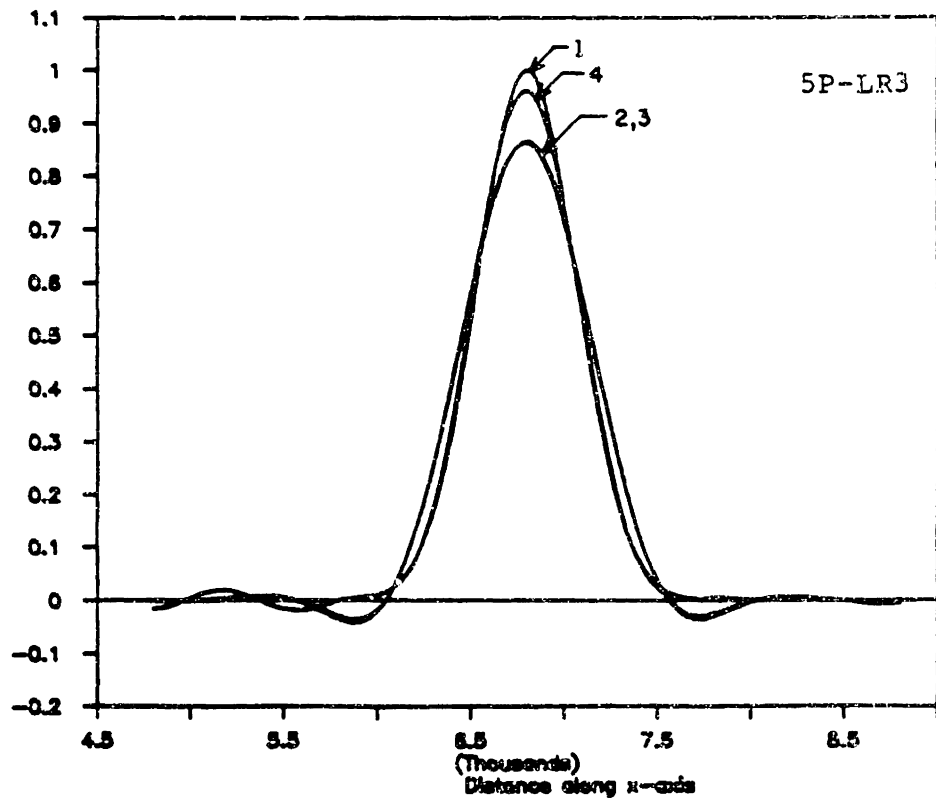
- Legend:
- 1 Exact solution
 - 2 Numerical solution
 - 3 Numerical solution, purged from phase errors
 - 4 Numerical solution, purged from amplitude errors

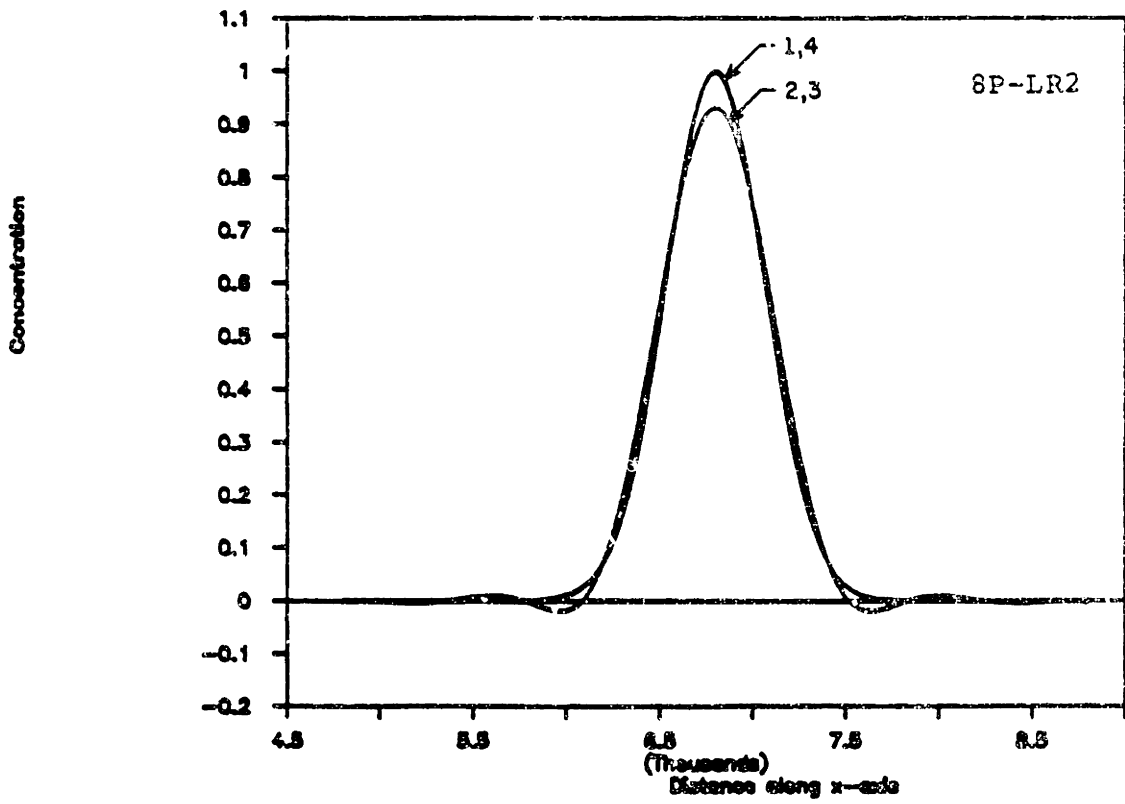
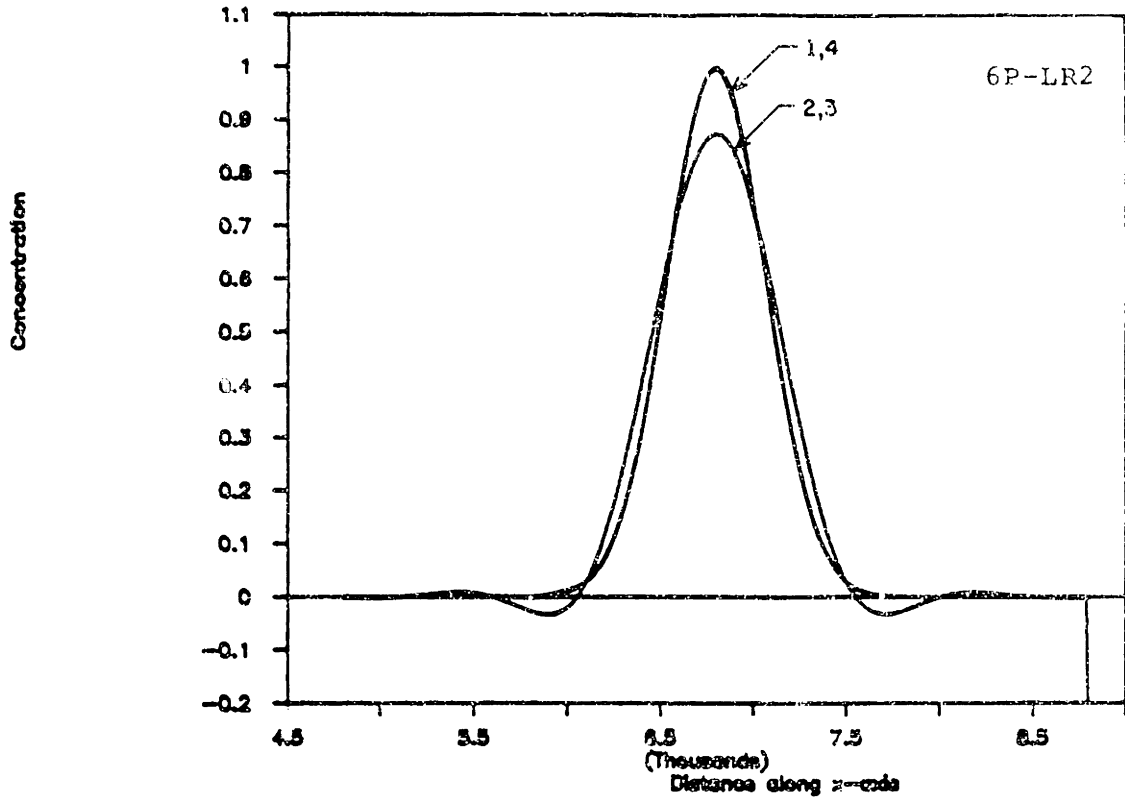


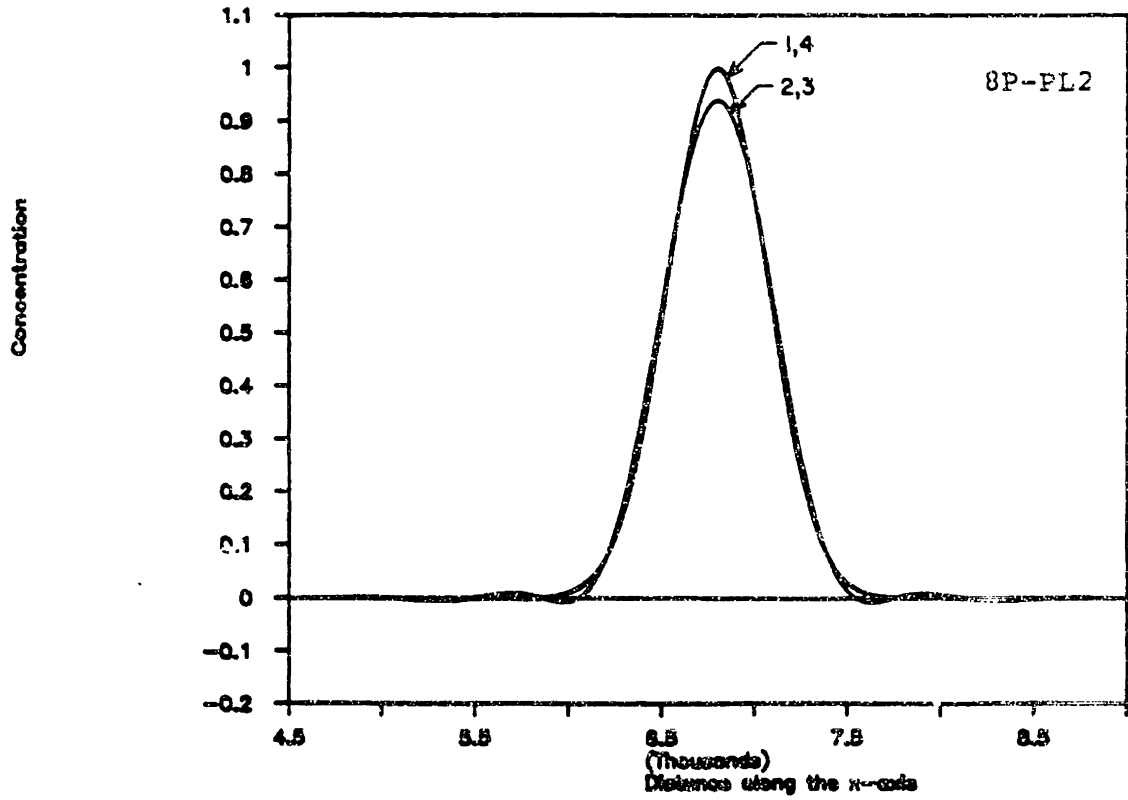
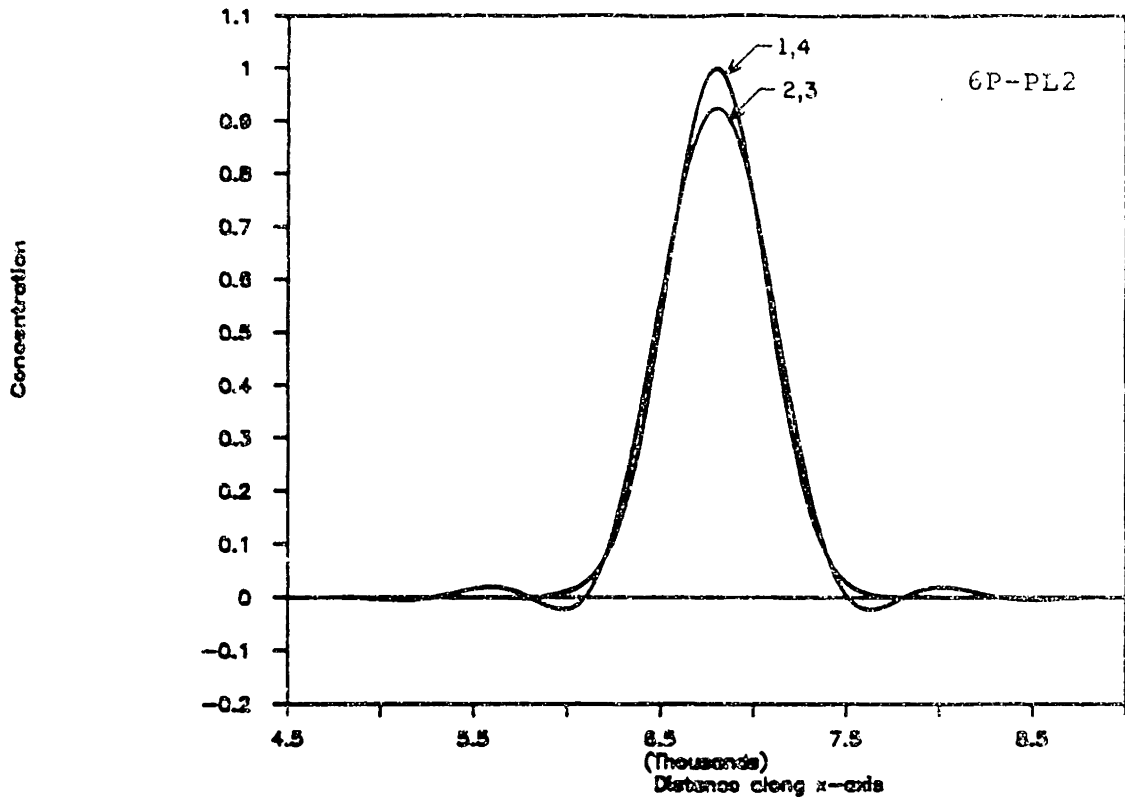
Concentration



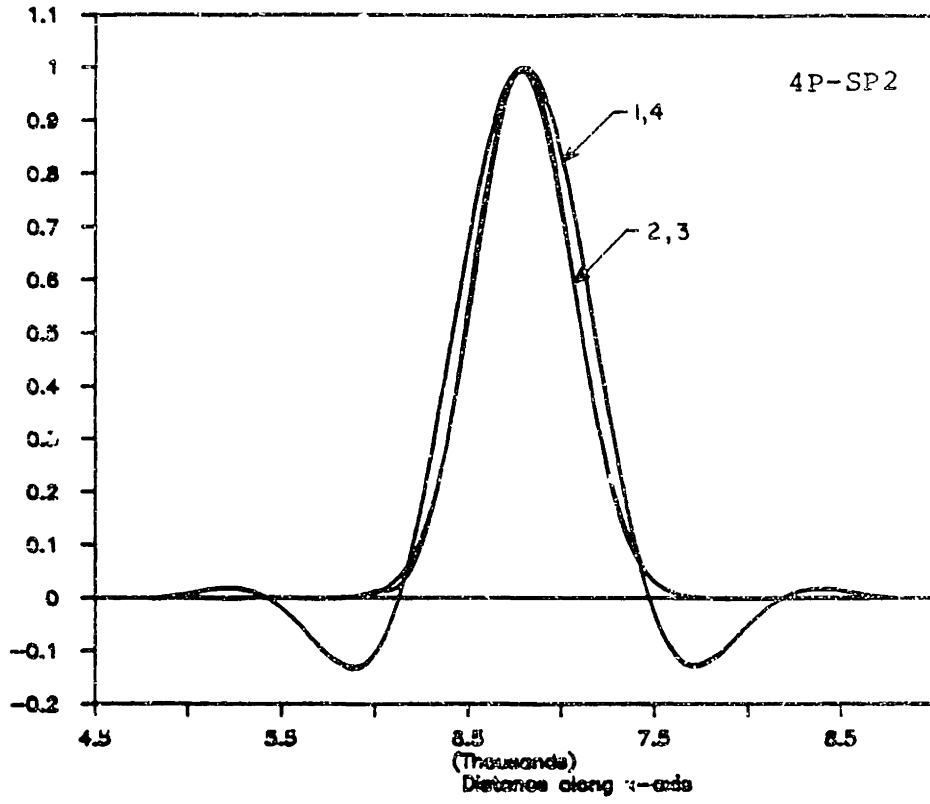
Concentrations



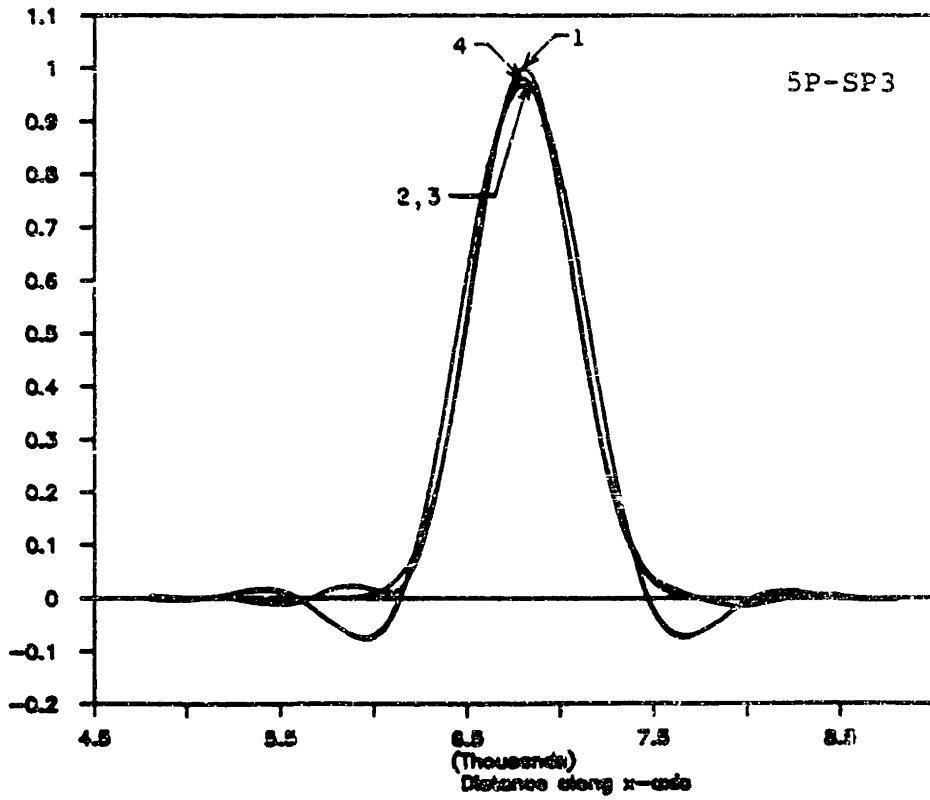




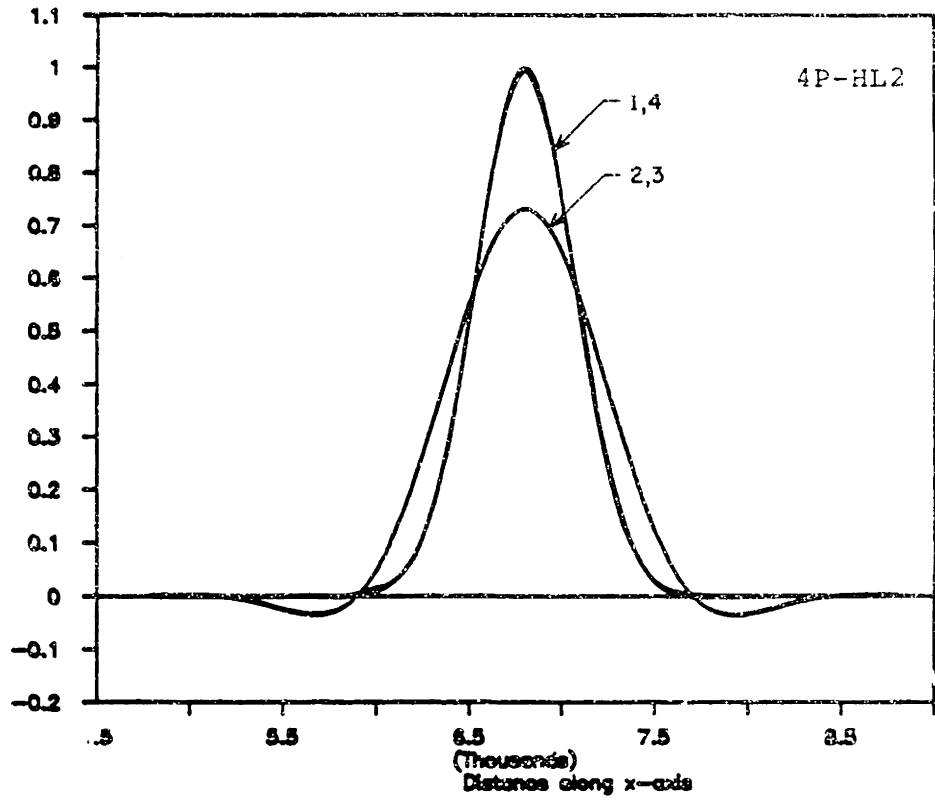
Concentration



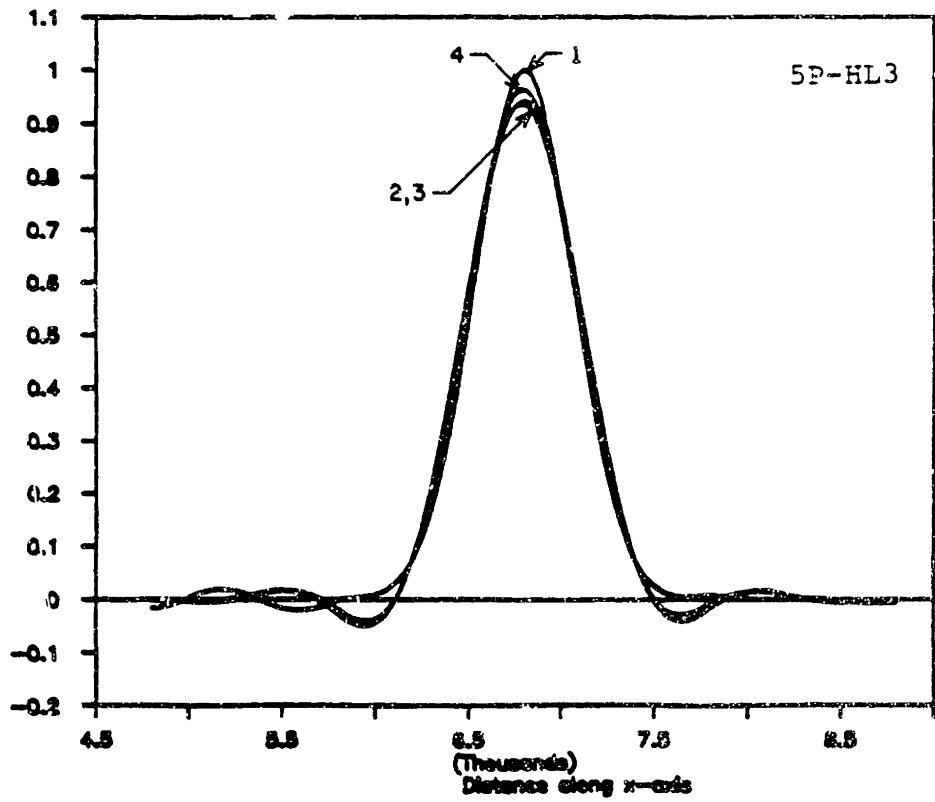
Concentration



Concentration



Concentration



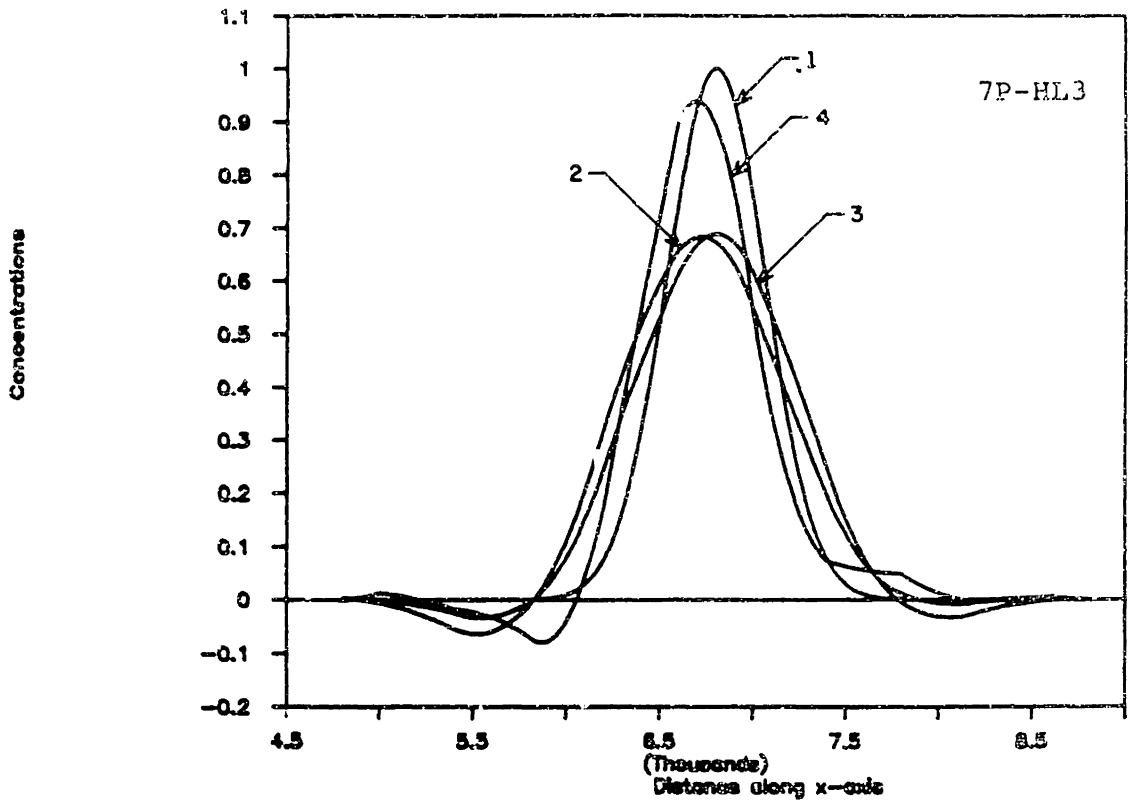


Figure 5

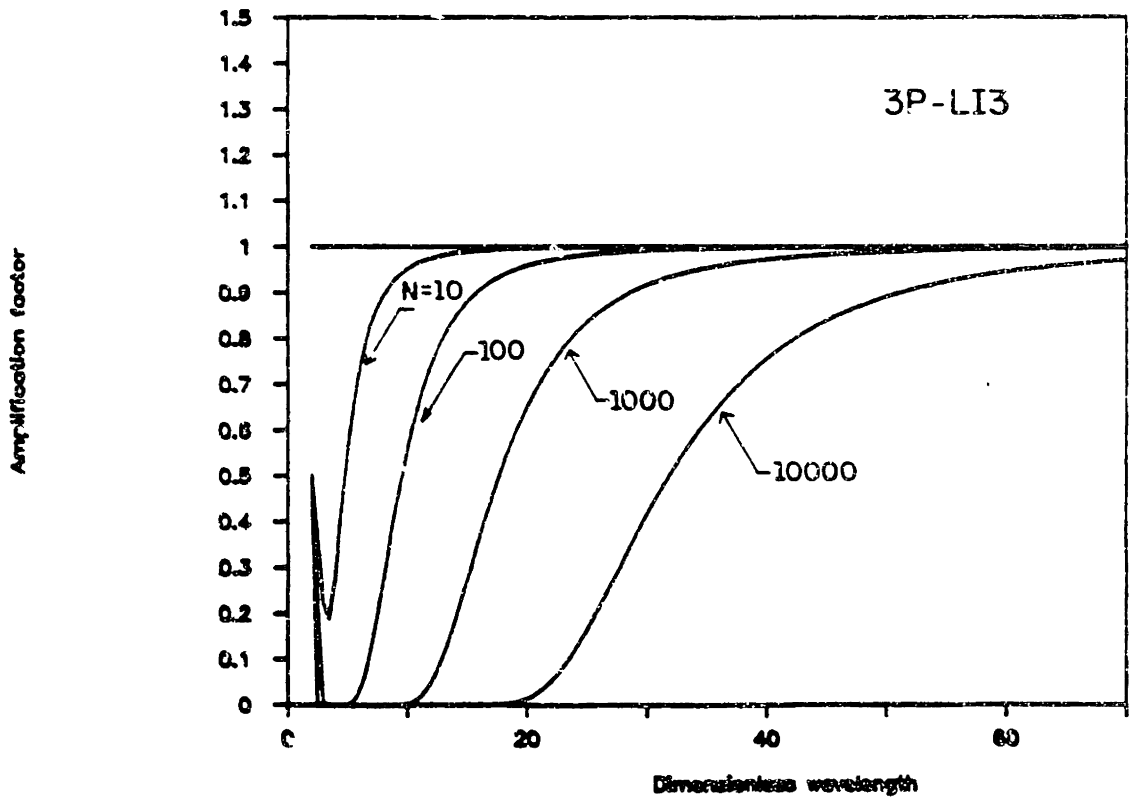
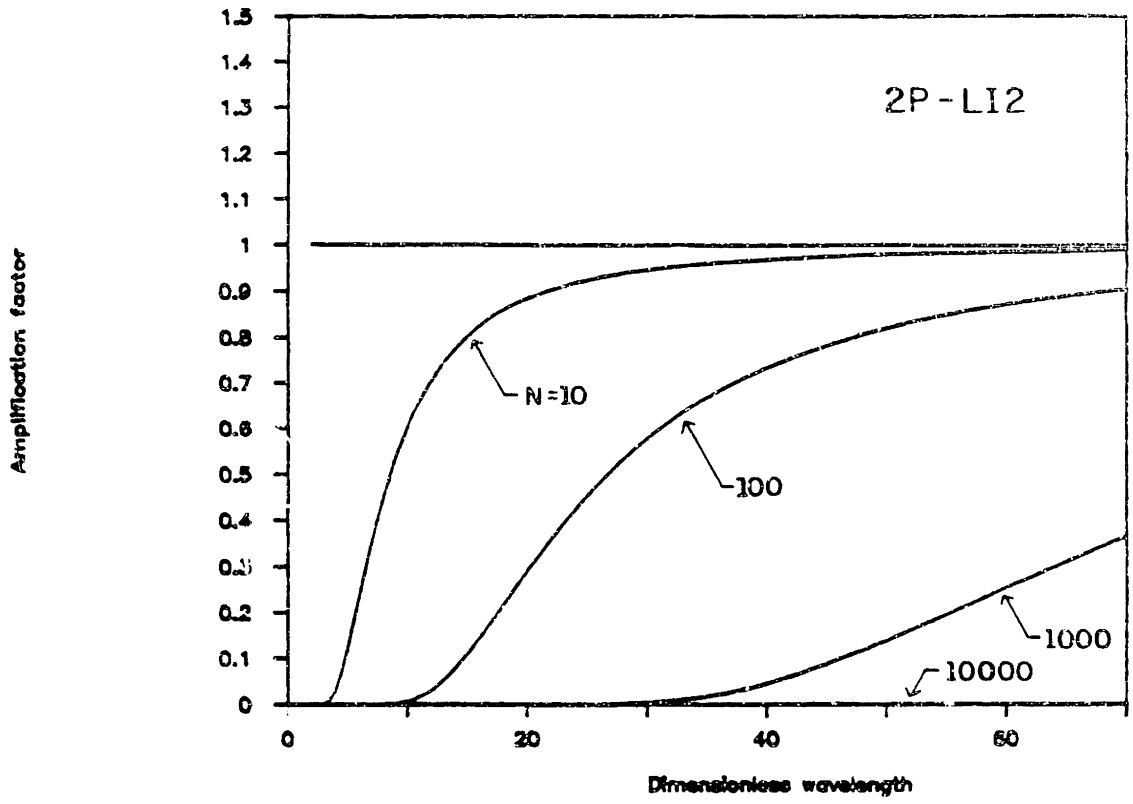
Amplification factors after N time steps, as a function of the
dimensionless wavelength ($\alpha = 0.5$)

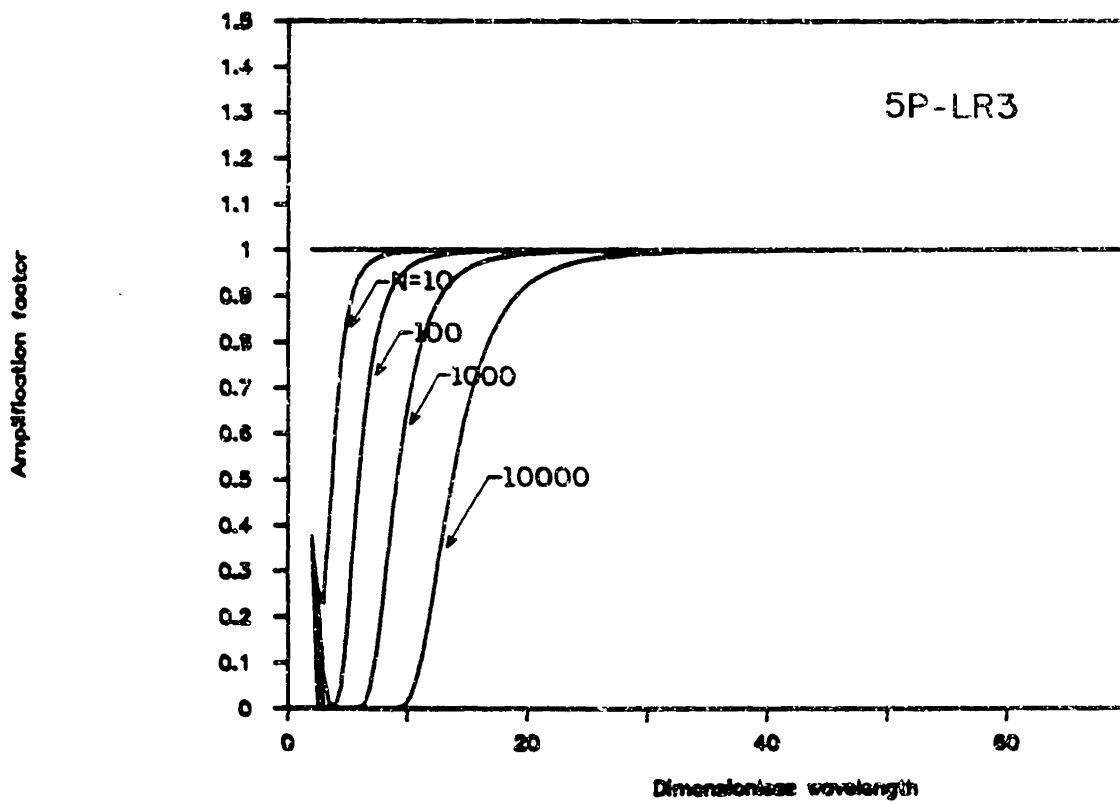
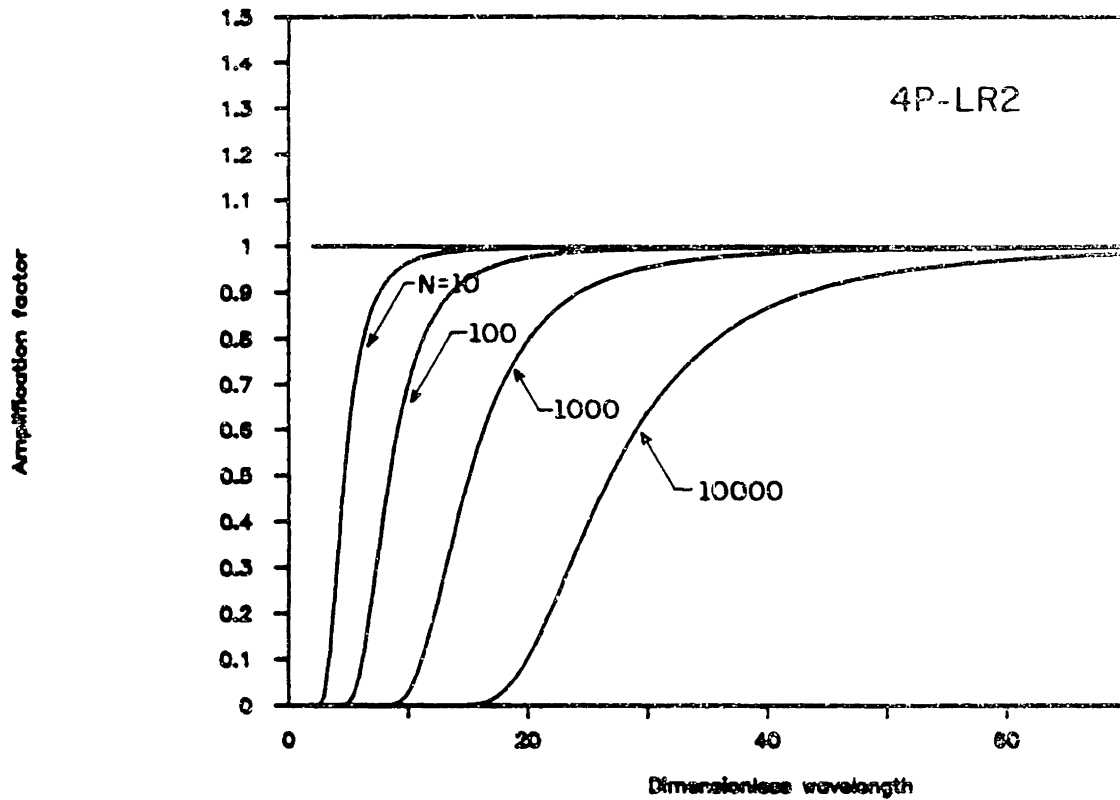
- (a) 2P-LI2 (b) 3P-LI3 (c) 4P-LR2 (d) 5P-LR3
(e) 6P-LR2 (f) 8P-LR2 (g) 6P-PL2 (h) 8P-PL2
(i) 4P-SP2 (j) 5P-SP3 (k) 4P-HL2 (l) 5P-HL3
(m) 7P-HL3

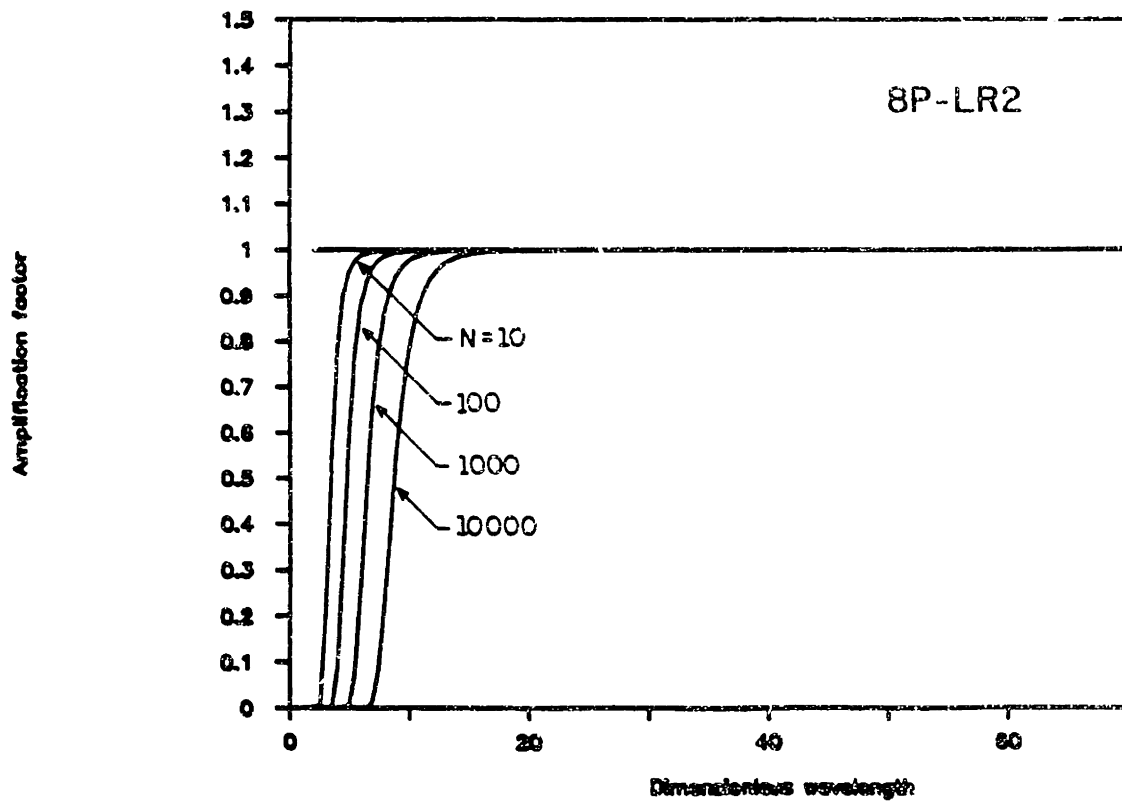
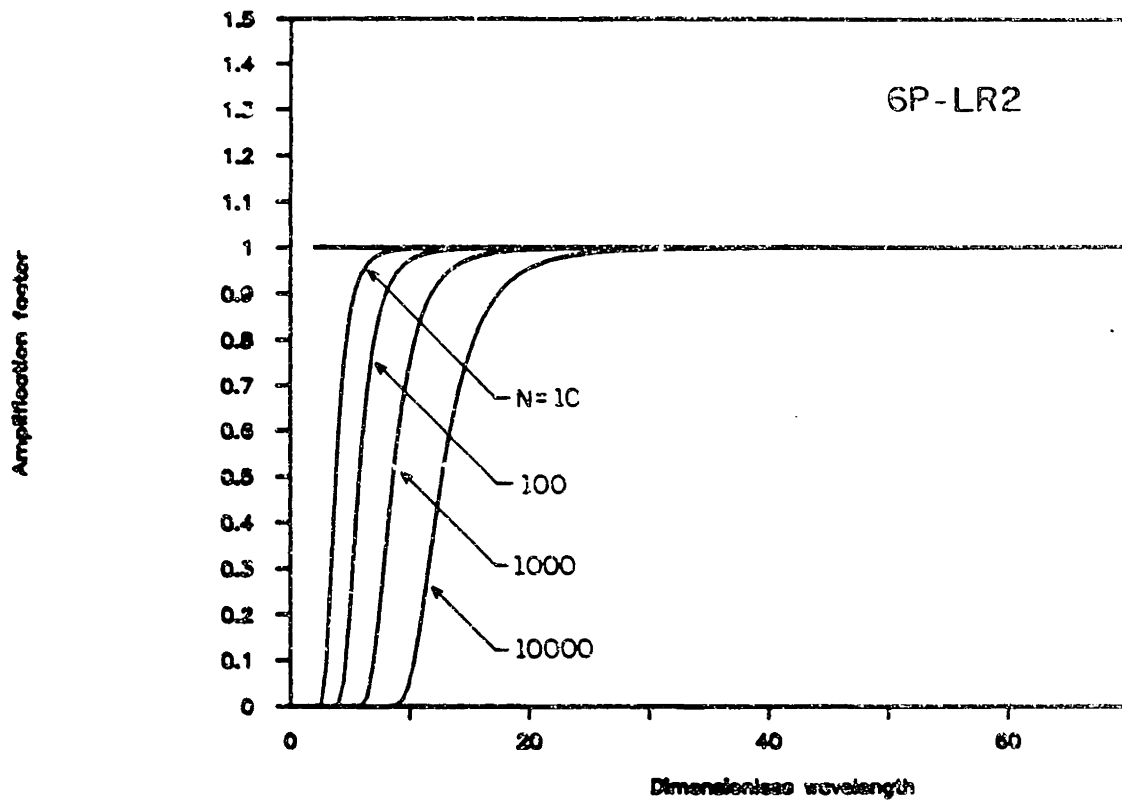
Reference problem: $\frac{\partial c}{\partial t} + u \frac{\partial c}{\partial x} = 0$
 $c(x, 0) = \exp\left\{-\frac{(x-x_0)^2}{2\sigma^2}\right\}$
 $c(x, t) \rightarrow 0 \quad |x| \rightarrow \infty$

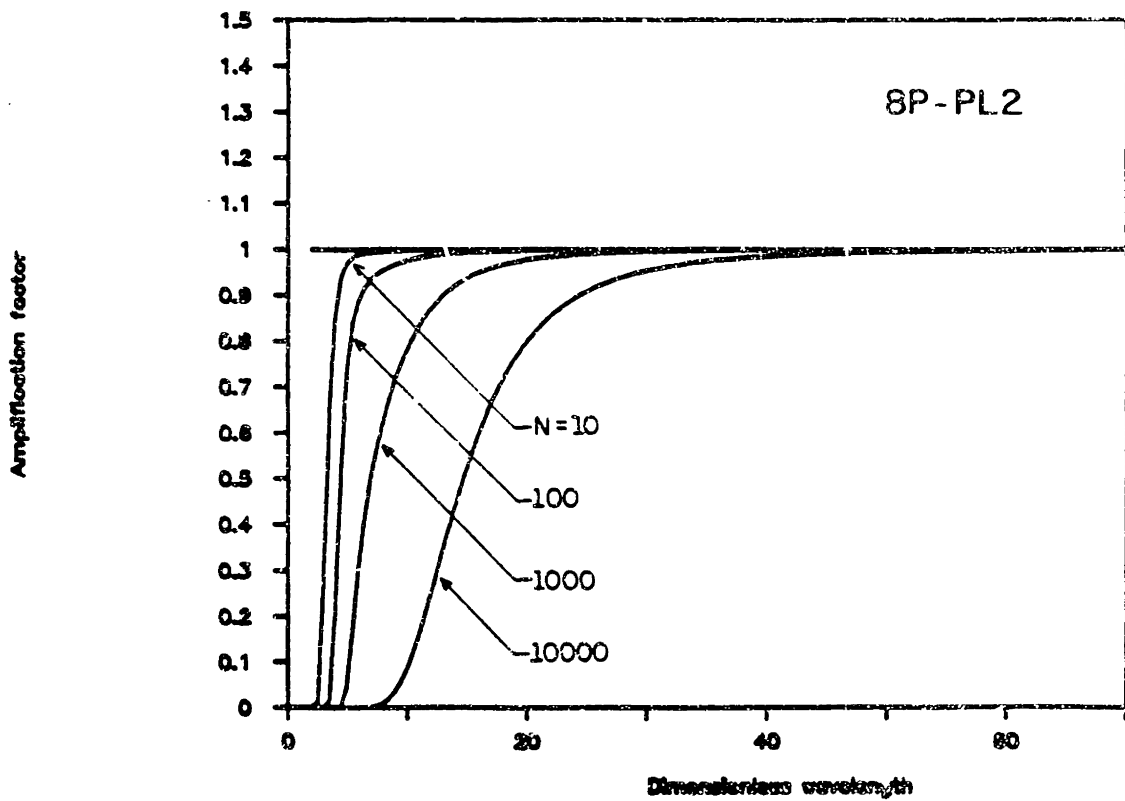
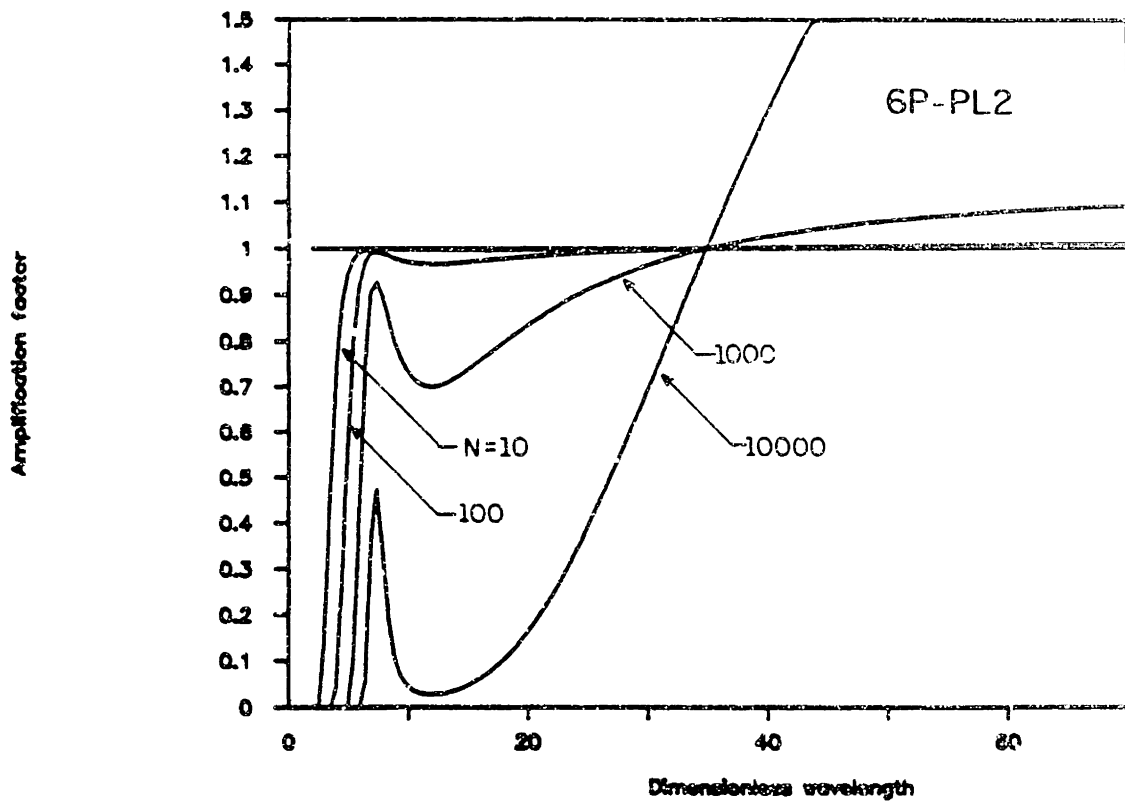
Computational parameters: $\alpha = 0.5$

N variable

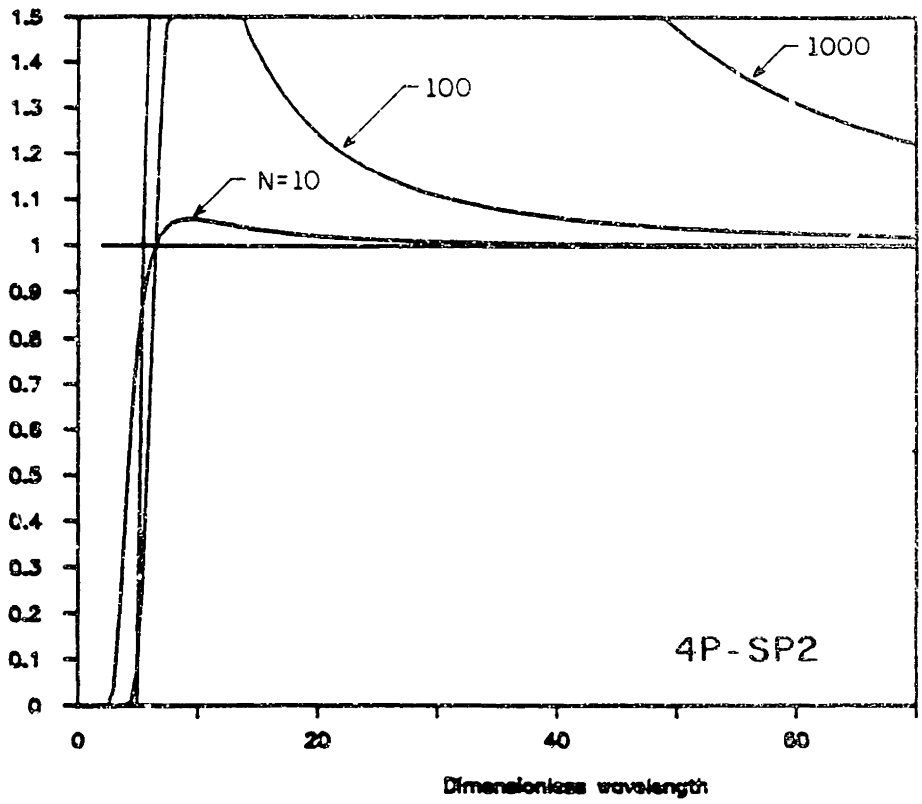




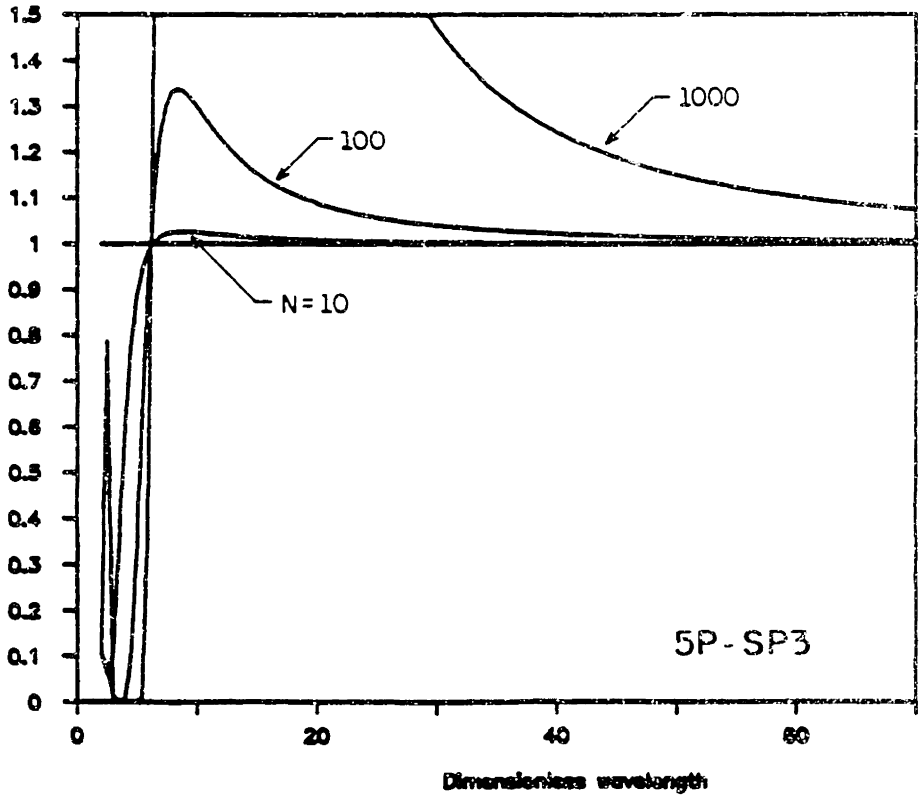


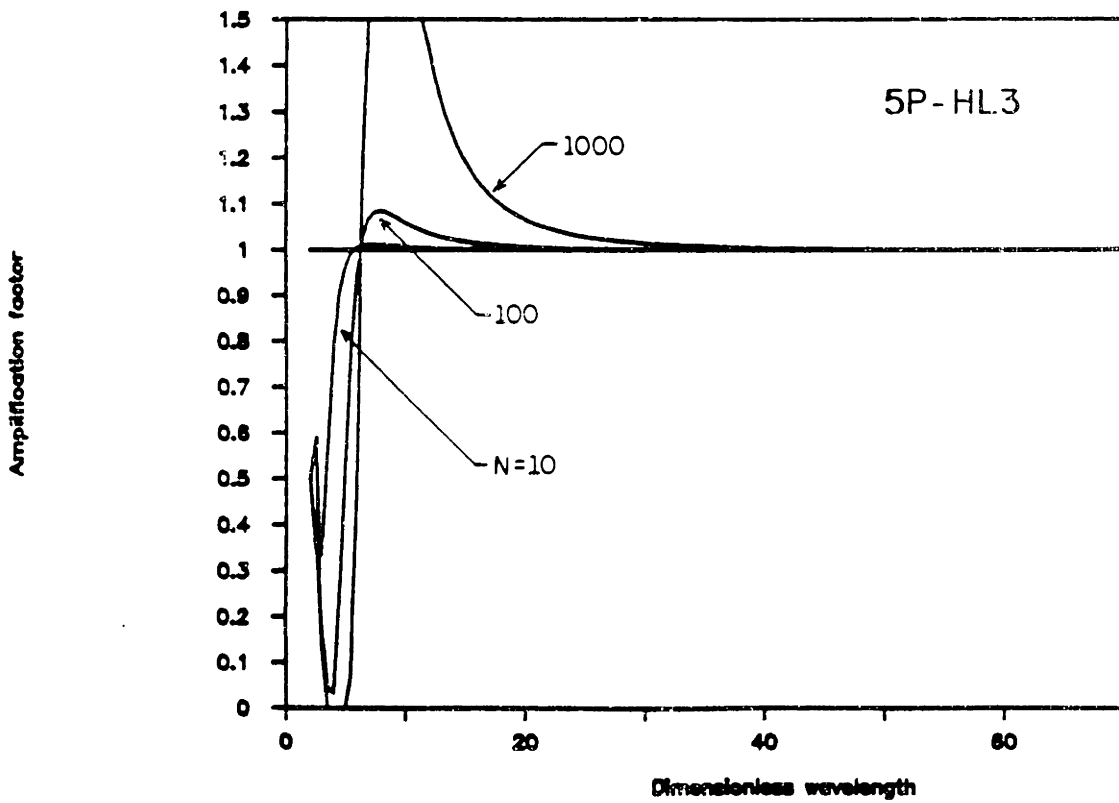
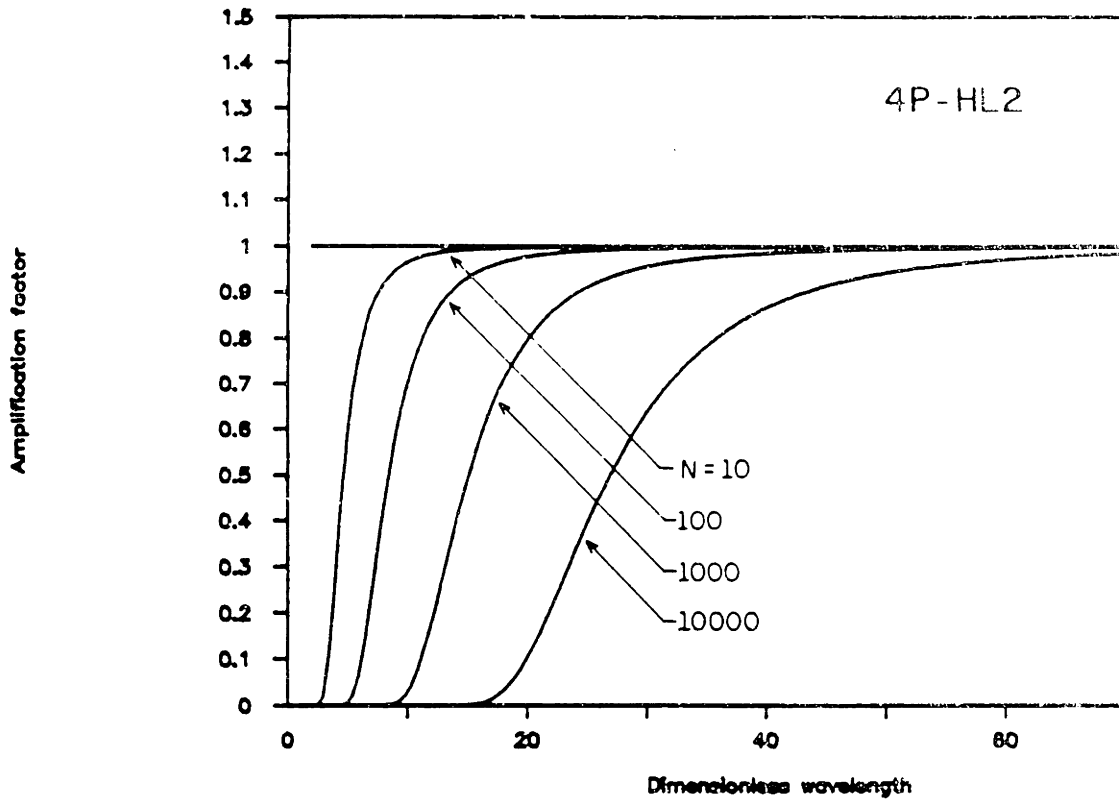


Amplification factor



Amplification factor





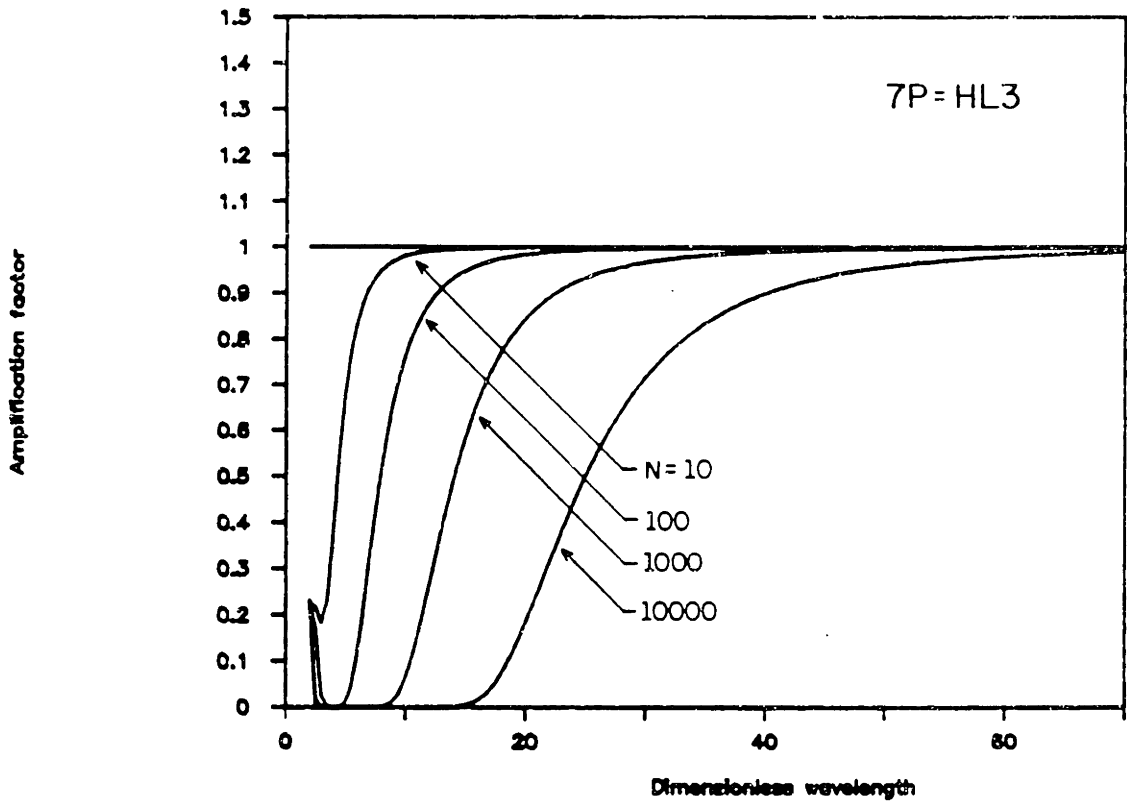


Figure 6

Comparison of the accuracy of alternative cubic interpolators

Reference problem: $\frac{\partial c}{\partial t} + u \frac{\partial c}{\partial x} = 0$
 $c(x, 0) = \exp\left\{-\frac{(x-x_0)^2}{2\sigma^2}\right\}$
 $c(x, t) \rightarrow 0 \quad |x| \rightarrow \infty$

Computational parameters: $\Delta t = 96$
 $\Delta x = 200$
 $N = 100$
 $u = 0.5$
 $\sigma = 264$

- Legend:
- 1 Exact solution
 - 2 2P-HI2 (reference C_1 interpolator)
 - 3 4P-LR2
 - 4 6P-PL2
 - 5 8P-PL2
 - 6 4P-SP2
 - 7 5P-SP3
 - 9 4P-HL2
 - 10 5P-HL3

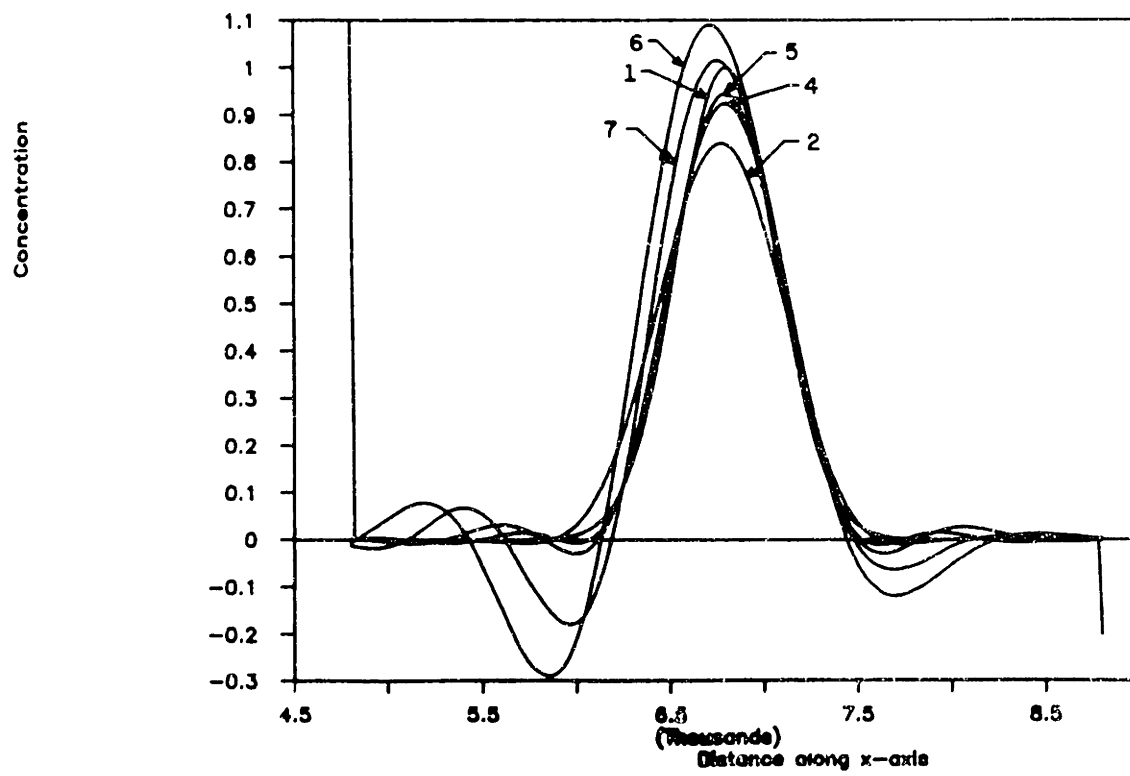
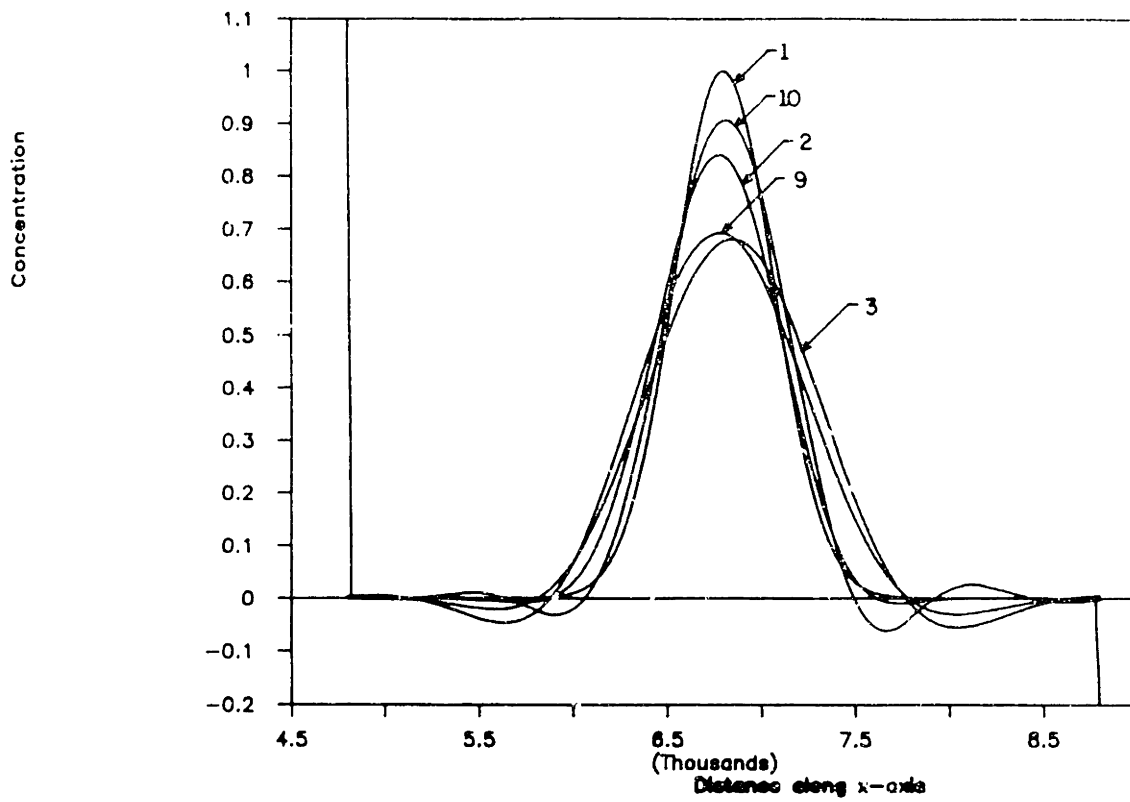


Figure 7

Comparison of the accuracy of alternative interpolators requiring
information from five nodes

Reference problem: $\frac{\partial c}{\partial t} + u \frac{\partial c}{\partial x} = 0$
 $c(x, 0) = \exp\left\{-\frac{(x-x_0)^2}{2\sigma^2}\right\}$
 $c(x, t) \rightarrow 0 \quad |x| \rightarrow \infty$

Computational parameters: $\Delta t = 96$

$$\Delta x = 200$$

$$N = 100$$

$$u = 0.5$$

$$\sigma = 264$$

- Legend:
- 1 Exact solution
 - 2 2P-HI2 (reference C_1 interpolator)
 - 3 5P-LR3
 - 4 5P-SP3
 - 5 5P-HL3

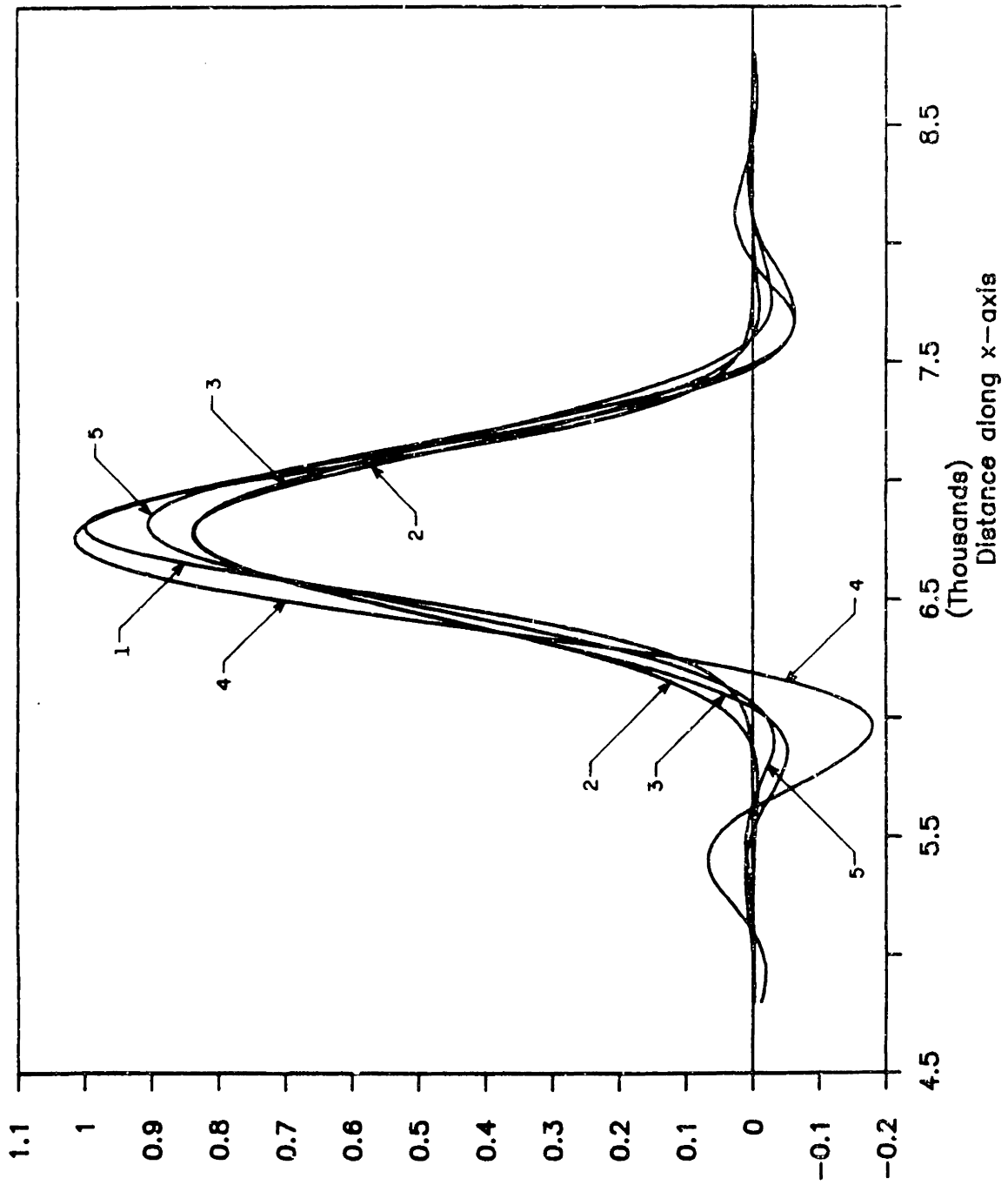


Figure 8

Dependence of accuracy on the grid refinement (3P-LI3 interpolator)

- (a) L2 error norm
- (b) Mass ratio
- (c) 2nd-moment ratio
- (d) 3rd-moment ratio

Reference problem: $\frac{\partial c}{\partial t} + u \frac{\partial c}{\partial x} = 0$
 $c(x, 0) = \exp\left\{-\frac{(x-x_0)^2}{2\sigma^2}\right\}$
 $c(x, t) \rightarrow 0 \quad |x| \rightarrow \infty$

Computational parameters: $\Delta t = 96$

$$\Delta x = 200$$

$$N = 100$$

$$u = 0.5$$

$$\sigma = 264$$

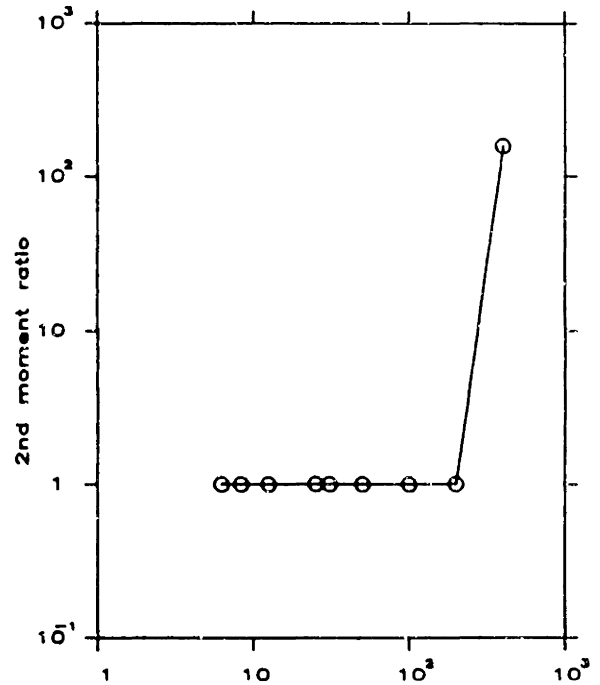
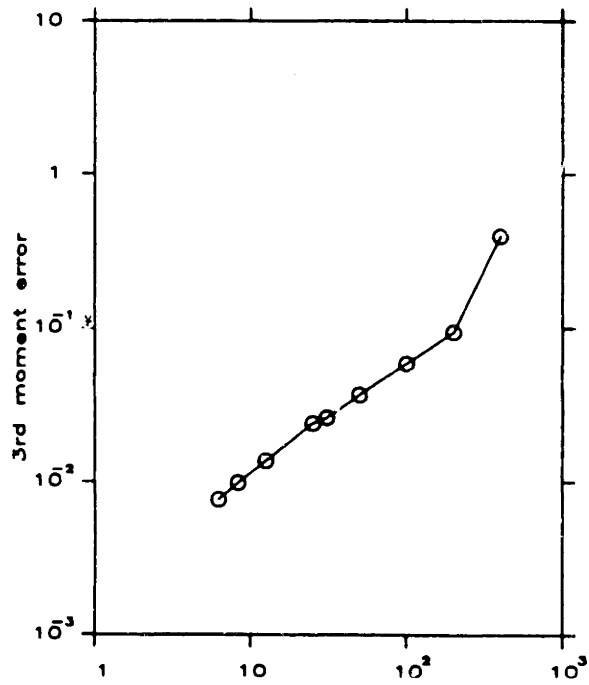
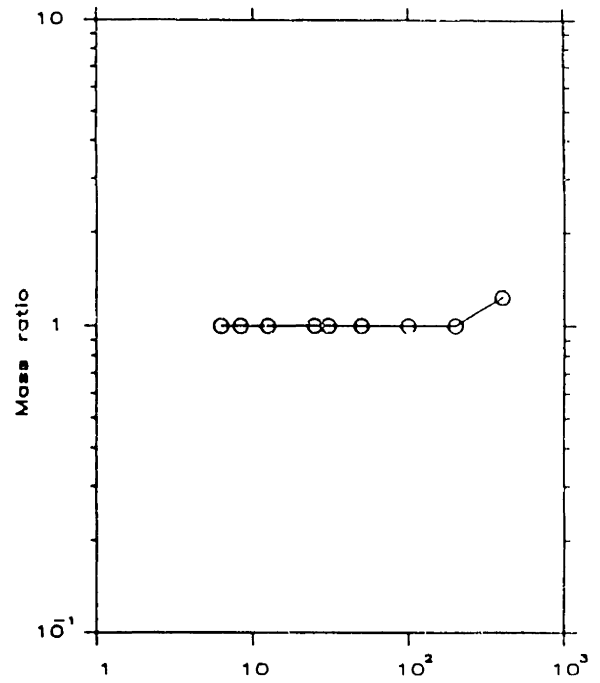
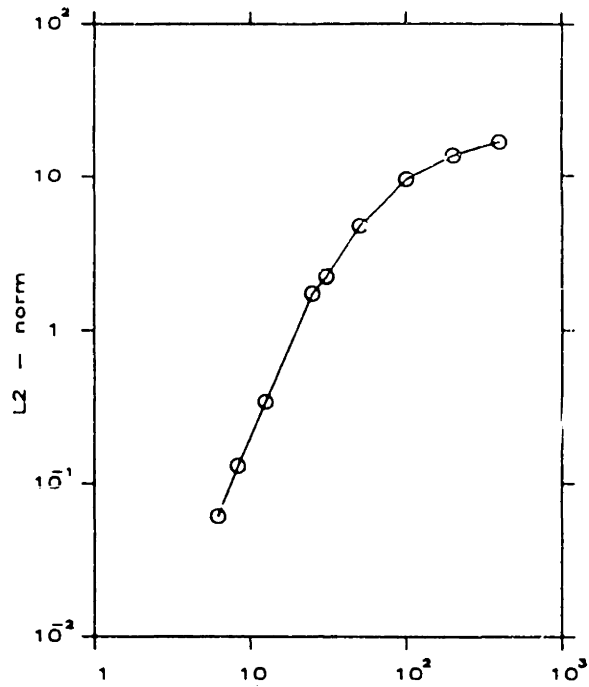


Figure 9

The effect of grid non-uniformity (3P-LI3 interpolator)

(a) L2 error norm

(b) 2nd-moment ratio

Reference problem: $\frac{\partial c}{\partial t} + u \frac{\partial c}{\partial x} = 0$

$$c(x, 0) = \exp\left\{-\frac{(x-x_0)^2}{2\sigma^2}\right\}$$

$$c(x, t) \rightarrow 0 \quad |x| \rightarrow \infty$$

Computational parameters: $\Delta t = 96$

$$N = 100$$

$$u = 0.5$$

$$\sigma = 264$$

Grid characteristics: — Uniform grid, $\Delta x = 200$

+/* Step grid

$$\Delta x_j \equiv x_j - x_{j-1} = \begin{cases} \Delta x_1 & \text{if } j \leq J/2 \\ s\Delta x_1 & \text{otherwise} \end{cases}$$

$$\text{with } ((1+s)\Delta x_1 \equiv 200)$$

$$(+ s=5 \quad * s=2)$$

o Alternate grid

$$\Delta x_j \equiv x_j - x_{j-1} = \begin{cases} \Delta x_1 & \text{if } j \text{ even} \\ 2 \Delta x_1 & \text{if } j \text{ odd} \end{cases}$$

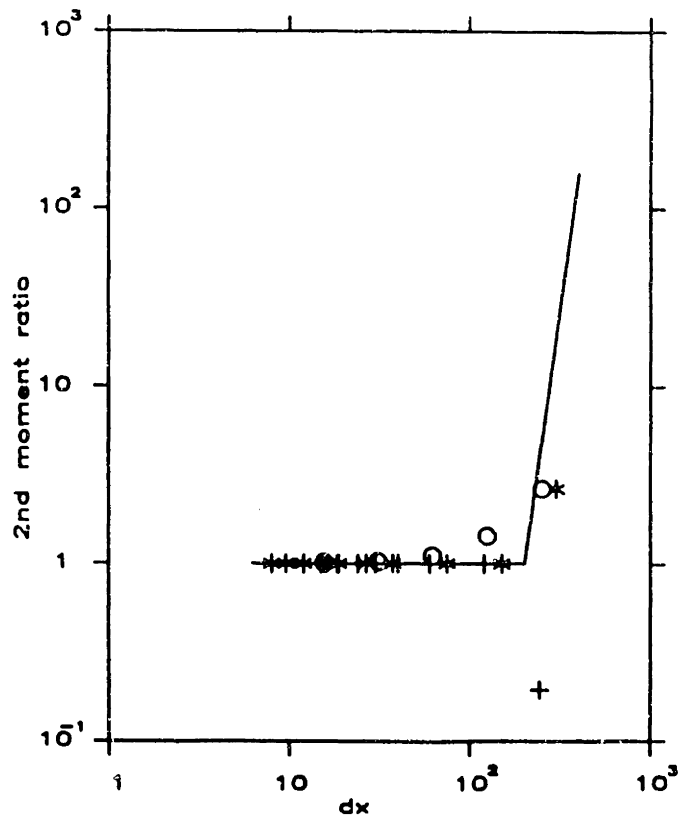
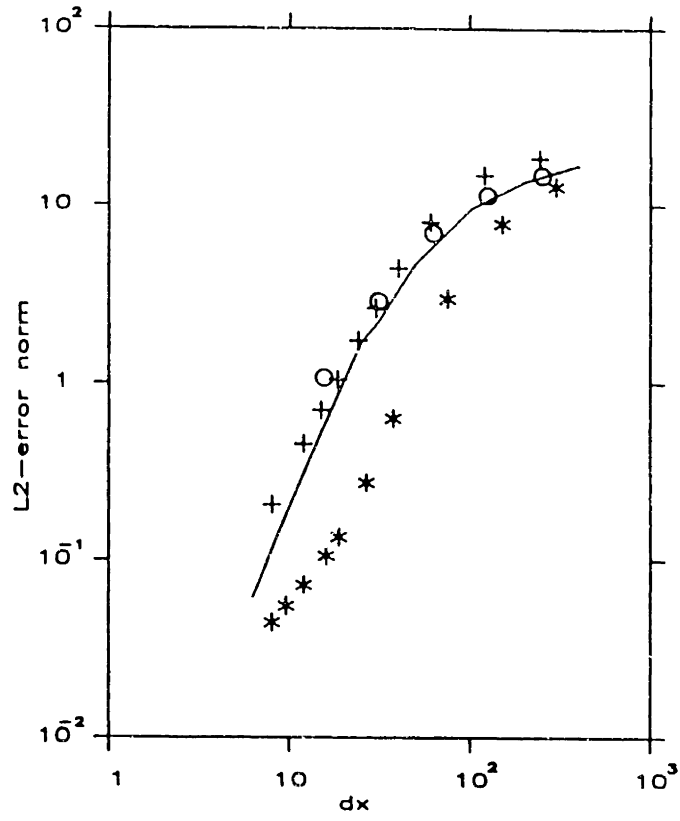


Table 1

Definition of spatial interpolators

2P-HI2

2P-LI2

3P-LI3

4P-LI4

4P-LR2

5P-LR3

6P-LR2

8P-LR2

6P-PL2

8P-PL2

4P-HL2

5P-HL3

7P-HL3

4P-SP2

5P-SP3

Scheme	Reference sketch	Definition	Q, M	Comments
2P-HI2		$f(\alpha) = \sum_{p=0}^1 \phi_p(\alpha) \cdot f_p + \psi_p(\alpha) \cdot \left. \frac{df}{d\alpha} \right _p$ $\phi_0(\alpha) = \frac{1}{4}(\alpha^2 - 3\alpha + 2)$ $\phi_1(\alpha) = \frac{1}{4}(-\alpha^2 + 3\alpha + 2)$ $\psi_0(\alpha) = \frac{1}{4}(\alpha^2 - \alpha^2 - \alpha + 1)$ $\psi_1(\alpha) = \frac{1}{4}(\alpha^2 + \alpha^2 - \alpha - 1)$	3,3	Cubic Hermite interpolator Compact, class C_1 Reference: [H1-HI2]
2P-LI2		$f(\alpha) = \sum_{p=0}^1 \phi_p(\alpha) \cdot f_p$ $\phi_0(\alpha) = 1 - \alpha$ $\phi_1(\alpha) = \alpha$	1,1	Linear Lagrange interpolator Compact, class C_0
3P-LI3		$f(\alpha) = \sum_{p=-1}^1 \phi_p(\alpha) \cdot f_p$ $\phi_{-1}(\alpha) = \frac{1}{2}(\alpha^2 - \alpha)$ $\phi_0(\alpha) = 1 - \alpha^2$ $\phi_1(\alpha) = \frac{1}{2}(\alpha^2 + \alpha)$	2,2	Quadratic Lagrange interpolator Compact, class C_0 Reference: [B2, B9]
4P-LI4		$f(\alpha) = \sum_{p=-1}^1 \phi_p(\alpha) \cdot f_p$ $\phi_{-1}(\alpha) = \frac{1}{16}(-9\alpha^3 + 9\alpha^2 + \alpha - 1)$ $\phi_{-1/3}(\alpha) = \frac{1}{16}(27\alpha^3 - 9\alpha^2 - 27\alpha + 9)$ $\phi_{1/3}(\alpha) = \frac{1}{16}(-27\alpha^3 - 9\alpha^2 + 27\alpha + 9)$ $\phi_1(\alpha) = \frac{1}{16}(9\alpha^3 + 9\alpha^2 - \alpha - 1)$	3,3	Cubic Lagrange interpolator Compact, class C_0

Scheme	Reference sketch	Definition	Q, M	Comments
4P-LR2		$f(\alpha) = \sum_{p=-1}^2 \phi_p(\alpha) \cdot f_p$ $\phi_{-1}(\alpha) = -\frac{1}{6}(\alpha^2 - 3\alpha^2 + 2\alpha)$ $\phi_0(\alpha) = \frac{1}{6}(3\alpha^2 - 6\alpha^2 - 3\alpha + 6)$ $\phi_1(\alpha) = -\frac{1}{6}(3\alpha^2 - 3\alpha^2 - 6\alpha)$ $\phi_2(\alpha) = \frac{1}{6}(\alpha^2 - \alpha)$	3, 3	
5P-LR3		$f(\alpha) = \sum_{p=-2}^2 \phi_p(\alpha) \cdot f_p$ $\phi_{-2}(\alpha) = \frac{1}{24}(\alpha^4 - 2\alpha^3 - \alpha^2 + 2\alpha)$ $\phi_{-1}(\alpha) = -\frac{1}{6}(\alpha^4 - \alpha^3 - 4\alpha^2 + 4\alpha)$ $\phi_0(\alpha) = \frac{1}{4}(\alpha^4 - 5\alpha^2 + 4)$ $\phi_1(\alpha) = -\frac{1}{6}(\alpha^4 + \alpha^3 - 4\alpha^2 - 4\alpha)$ $\phi_2(\alpha) = \frac{1}{24}(\alpha^4 + 2\alpha^3 - \alpha^2 - 2\alpha)$	4, 4	

Scheme	Reference sketch	Definition	Q, M	Comments
6P-LR2		$f(\alpha) = \sum_{p=-2}^3 \phi_p(\alpha) \cdot f_p$ $\phi_{-2}(\alpha) = -\frac{1}{120}\alpha(\alpha^2-1)(\alpha-2)(\alpha-3)$ $\phi_{-1}(\alpha) = \frac{1}{24}\alpha(\alpha-1)(\alpha^2-4)(\alpha-3)$ $\phi_0(\alpha) = -\frac{1}{12}(\alpha^2-1)(\alpha^2-4)(\alpha-3)$ $\phi_1(\alpha) = \frac{1}{12}\alpha(\alpha+1)(\alpha^2-4)(\alpha-3)$ $\phi_2(\alpha) = -\frac{1}{24}\alpha(\alpha^2-1)(\alpha+2)(\alpha-3)$ $\phi_3(\alpha) = \frac{1}{120}\alpha(\alpha^2-1)(\alpha^2-4)$	5,5	
8P-LR2		$f(\alpha) = \sum_{p=-3}^4 \phi_p(\alpha) \cdot f_p$ $\phi_{-3}(\alpha) = \frac{\alpha(\alpha^2-1)(\alpha^2-4)(\alpha-3)(\alpha-4)}{5040}$ $\phi_{-2}(\alpha) = \frac{\alpha(\alpha^2-1)(\alpha-2)(\alpha^2-9)(\alpha-4)}{720}$ $\phi_{-1}(\alpha) = -\frac{\alpha(\alpha-1)(\alpha^2-4)(\alpha^2-9)(\alpha-4)}{240}$ $\phi_0(\alpha) = \frac{(\alpha^2-1)(\alpha^2-4)(\alpha^2-9)(\alpha-4)}{144}$ $\phi_1(\alpha) = \frac{\alpha(\alpha+1)(\alpha^2-4)(\alpha^2-9)(\alpha-4)}{144}$ $\phi_2(\alpha) = \frac{\alpha(\alpha^2-1)(\alpha+2)(\alpha^2-9)(\alpha-4)}{240}$ $\phi_3(\alpha) = -\frac{\alpha(\alpha^2-1)(\alpha^2-4)(\alpha+3)(\alpha-4)}{720}$ $\phi_4(\alpha) = \frac{\alpha(\alpha^2-1)(\alpha^2-4)(\alpha^2-9)}{5040}$	7,7	

Scheme	Reference sketch	Definition	Q, M	Comments
6P-PL2		$f(\alpha) = \sum_{p=2}^3 \phi_p(\alpha) \cdot f_p$ $\phi_{-2}(\alpha) = 0.01806\alpha^2 - 0.09245\alpha + 0.07439\alpha$ $\phi_{-1}(\alpha) = -0.2570\alpha^2 + 0.8236\alpha - 0.5667\alpha$ $\phi_0(\alpha) = 0.6806\alpha^2 - 1.394\alpha + 0.2869\alpha + 1$ $\phi_1(\alpha) = -0.6806\alpha^2 + 0.6480\alpha^2 + 1.033\alpha$ $\phi_2(\alpha) = 0.2570\alpha^2 + 0.05276\alpha^2 - 0.3097\alpha$ $\phi_3(\alpha) = -0.01806\alpha^2 - 0.03828\alpha^2 + 0.05633\alpha$	1, 3	<p>Interpolator based on the average of two pseudo-Hermite polynomials, defined over the core element with the assistance of a fictive middle point. Derivatives at the corner nodes are estimated by weighted average of concentrations at 6 nodes.</p> <p>Non-compact, class C_0</p> <p>Reference: [K1-K2]</p>
8P-PL2		$f(\alpha) = \sum_{p=3}^4 \phi_p(\alpha) \cdot f_p$ $\phi_{-3}(\alpha) = \phi_4(\alpha) = \frac{1}{66}\alpha^2 - \frac{1}{66}\alpha$ $\phi_{-2}(\alpha) = \frac{1}{54}\alpha^3 - \frac{115}{792}\alpha^2 + \frac{301}{2376}\alpha$ $\phi_{-1}(\alpha) = -\frac{7}{27}\alpha^3 + \frac{713}{792}\alpha^2 - \frac{1523}{2376}\alpha$ $\phi_0(\alpha) = \frac{37}{54}\alpha^3 - \frac{569}{396}\alpha^2 - \frac{296}{1188}\alpha + 1$ $\phi_1(\alpha) = -\frac{37}{54}\alpha^3 + \frac{245}{396}\alpha^2 + \frac{1267}{1188}\alpha$ $\phi_2(\alpha) = \frac{7}{27}\alpha^3 + \frac{97}{792}\alpha^2 - \frac{907}{2376}\alpha$ $\phi_3(\alpha) = -\frac{1}{54}\alpha^3 - \frac{71}{792}\alpha^2 + \frac{257}{2376}\alpha$	3, 3	<p>Interpolator based on the average of two pseudo-Hermite polynomials, defined over the core element with the assistance of a fictive middle point. Derivatives at the corner nodes are estimated by weighted average of concentrations at 8 nodes.</p> <p>Non-compact, class C_0</p> <p>Reference: [H3]</p>
4P-HL2		$f(\alpha) = \sum_{p=1}^2 \phi_p(\alpha) \cdot f_p$ $\phi_{-1}(\alpha) = \frac{1}{4}(\alpha^2 - \alpha)$ $\phi_0(\alpha) = \frac{1}{4}(\alpha^2 + 3\alpha - 4)$ $\phi_1(\alpha) = \frac{1}{4}(\alpha^2 - 5\alpha)$ $\phi_2(\alpha) = \frac{1}{4}(\alpha^2 - \alpha)$	2, 3	<p>Interpolator based on a cubic Hermite polynomial, defined over the core element, and which derivatives at the corner nodes are estimated from concentrations at 4 nodes</p> <p>Non-compact, class C_0</p>

Scheme	Reference sketch	Definition	Q, M	Comments
5P-HL3		$f(\alpha) = \sum_{p=-2}^2 \phi_p(\alpha) \cdot f_p$ $\phi_{-2}(\alpha) = -\frac{1}{12}\alpha(\alpha^2-1)$ $\phi_{-1}(\alpha) = \frac{1}{6}(\alpha^2+3\alpha^2-4\alpha)$ $\phi_0(\alpha) = 1-\alpha^2$ $\phi_1(\alpha) = -\frac{1}{6}(\alpha^2-3\alpha^2-4\alpha)$ $\phi_2(\alpha) = \frac{1}{12}\alpha(\alpha^2-1)$	3, 3	<p>Interpolator based on a cubic Hermite polynomial, defined over the core element, and which derivatives at the corner nodes are estimated from concentrations at 5 nodes.</p> <p>Non-compact, class C_0</p>
7P-HL3		$f(\alpha) = \sum_{p=-3}^3 \phi_p(\alpha) \cdot f_p$ $\phi_{-3}(\alpha) = \frac{7-1}{8}(\alpha^2-\alpha^2-\alpha+1)$ $\phi_{-2}(\alpha) = \frac{7-1}{2}(\alpha^2-\alpha^2-\alpha+1)$ $\phi_{-1}(\alpha) = \frac{1}{8}(5(1-\tau)\alpha^2+(7\tau-3)\alpha^2+(5\tau-9)\alpha+7(1-\tau))$ $\phi_0(\alpha) = 7(1-\alpha^2)$ $\phi_1(\alpha) = \frac{1}{8}(5(1-\tau)\alpha^2+(7\tau-3)\alpha^2+(9-5\tau)\alpha+7(1-\tau))$ $\phi_2(\alpha) = \frac{7-1}{2}(\alpha^2+\alpha^2-\alpha-1)$ $\phi_3(\alpha) = \frac{7-1}{8}(\alpha^2+\alpha^2-\alpha-1)$	2, 3	<p>Interpolator based on a cubic Hermite polynomial, defined over the core element, and which derivatives at the corner nodes are estimated from concentrations at 7 nodes</p> <p>Non-compact, class C_0</p>

Scheme	Reference sketch	Definition	Q,M	Comments
4P-SP2		$f(\alpha) = \sum_{p=0}^2 \phi_p(\alpha) \cdot f_p$ $\phi_{-1}(\alpha) = -\frac{1}{15}(5\alpha^2 - 12\alpha^2 + 7\alpha)$ $\phi_0(\alpha) = \frac{1}{15}(15\alpha^2 - 27\alpha^2 - 3\alpha + 15)$ $\phi_1(\alpha) = -\frac{1}{15}(15\alpha^2 - 18\alpha^2 - 12\alpha)$ $\phi_2(\alpha) = \frac{1}{15}(5\alpha^2 - 3\alpha^2 - 2\alpha)$	3,3	<p>Interpolator based on locally defined natural splines, valid within the core element but built from concentrations at 4 nodes. Second derivatives are forced to be zero at -1 and 2</p> <p>Non-compact. class C₀</p>
5P-SP3		$f(\alpha) = \sum_{p=-2}^2 \phi_p(\alpha) \cdot f_p$ <p>if $\alpha \leq 0$</p> $\phi_{-2}(\alpha) = \frac{1}{56}(-19\alpha^2 - 12\alpha^2 + 7\alpha)$ $\phi_{-1}(\alpha) = \frac{1}{56}(58\alpha^2 + 72\alpha^2 - 42\alpha)$ $\phi_0(\alpha) = \frac{1}{56}(-64\alpha^2 - 120\alpha^2 + 56)$ $\phi_1(\alpha) = \frac{1}{56}(30\alpha^2 + 72\alpha^2 + 42\alpha)$ $\phi_2(\alpha) = \frac{1}{56}(-5\alpha^2 - 12\alpha^2 - 7\alpha)$ <p>if $\alpha > 0$</p> $\phi_2(\alpha) = \frac{1}{56}(19\alpha^2 - 12\alpha^2 - 7\alpha)$ $\phi_1(\alpha) = \frac{1}{56}(-58\alpha^2 + 72\alpha^2 + 42\alpha)$ $\phi_0(\alpha) = \frac{1}{56}(64\alpha^2 - 120\alpha^2 + 56)$ $\phi_{-1}(\alpha) = \frac{1}{56}(-30\alpha^2 + 72\alpha^2 - 42\alpha)$ $\phi_{-2}(\alpha) = \frac{1}{56}(5\alpha^2 - 12\alpha^2 + 7\alpha)$	3,3	<p>Interpolator based on locally defined natural splines, valid within the core element but built from concentrations at 5 nodes. Second derivatives are forced to be zero at -2 and 2</p> <p>Non-compact. class C₀</p>

Table 2

Illustration of the Generation Procedure for Hybrid Interpolators
Based on the Estimation of Derivatives of a Core Cubic Hermite Polynomial

(a) 4P-HL2

(b) 5P-HL3

(c) 7P-HL3

Basic Hermite polynomial

$$c_j^{n+1} = r_0(\alpha)c_{j-k}^n + r_1(\alpha)c_{j-k-1}^n + s_0(\alpha)\left.\frac{dc}{d\alpha}\right|_{j-k} + s_1(\alpha)\left.\frac{dc}{d\alpha}\right|_{j-k-1}^n$$

with

$$r_0(\alpha) = 2\alpha^3 - 3\alpha^2 + 1$$

$$r_1(\alpha) = -2\alpha^3 + 3\alpha^2$$

$$s_0(\alpha) = \alpha^3 - 2\alpha^2 + \alpha$$

$$s_1(\alpha) = \alpha^3 - \alpha^2$$

Estimation of derivatives

$$\left.\frac{dc}{d\alpha}\right|_{j-k}^n = \frac{1}{2} \left\{ \left.\frac{dP_1(\alpha)}{d\alpha}\right|_{\alpha=0} + \left.\frac{dP_2(\alpha)}{d\alpha}\right|_{\alpha=0} \right\}$$

$$\left.\frac{dc}{d\alpha}\right|_{j-k-1}^n = \frac{1}{2} \left\{ \left.\frac{dP_1(\alpha)}{d\alpha}\right|_{\alpha=1} + \left.\frac{dP_2(\alpha)}{d\alpha}\right|_{\alpha=1} \right\}$$

where

$$P_1(\alpha) = \frac{1}{2}(\alpha^2 + \alpha)c_{j-k-1}^n + (1 - \alpha^2)c_{j-k}^n + \frac{1}{2}(\alpha^2 - \alpha)c_{j-k+1}^n$$

$$P_2(\alpha) = \frac{1}{2}\alpha(\alpha-1)c_{j-k-2}^n - \alpha(\alpha-2)c_{j-k-1}^n + \frac{1}{2}(\alpha-1)(\alpha-2)c_{j-k}^n$$

Basic Hermite polynomial

$$c_j^{n+1} = r_{-1}(\alpha)c_{j-k+1}^n + r_1(\alpha)c_{j-k-1}^n + s_{-1}(\alpha) \left. \frac{dc}{d\alpha} \right|_{j-k+1}^n + s_1(\alpha) \left. \frac{dc}{d\alpha} \right|_{j-k-1}^n$$

with

$$r_{-1}(\alpha) = 0.25(\alpha^3 - 3\alpha + 2)$$

$$r_1(\alpha) = 0.25(-\alpha^3 + 3\alpha + 2)$$

$$s_{-1}(\alpha) = 0.25(\alpha^3 - \alpha^2 - \alpha + 1)$$

$$s_1(\alpha) = 0.25(\alpha^3 + \alpha^2 - \alpha - 1)$$

Estimation of derivatives

$$\left. \frac{dc}{d\alpha} \right|_{j-k+1}^n = \frac{1}{2} \left\{ \left. \frac{dP_1(\alpha)}{d\alpha} \right|_{\alpha=-1} + \left. \frac{dP_2(\alpha)}{d\alpha} \right|_{\alpha=-1} \right\}$$

$$\left. \frac{dc}{d\alpha} \right|_{j-k-1}^n = \frac{1}{2} \left\{ \left. \frac{dP_1(\alpha)}{d\alpha} \right|_{\alpha=1} + \left. \frac{dP_2(\alpha)}{d\alpha} \right|_{\alpha=1} \right\}$$

where

$$P_1(\alpha) = -\frac{1}{6}(\alpha^3 - 3\alpha^2 + 2\alpha)c_{j-k+1} + \frac{1}{2}(\alpha^3 - 2\alpha^2 - \alpha + 2)c_{j-k} - \frac{1}{2}(\alpha^3 - \alpha^2 - 2\alpha)c_{j-k-1} + \frac{1}{6}(\alpha^3 - \alpha)c_{j-k-2}$$

$$P_2(\alpha) = -\frac{1}{6}(\alpha^3 - \alpha)c_{j-k+2} + \frac{1}{2}(\alpha^3 + \alpha^2 - 2\alpha)c_{j-k+1} + \frac{1}{2}(-\alpha^3 - 2\alpha^2 + \alpha + 2)c_{j-k} + \frac{1}{6}(\alpha^3 + 3\alpha^2 + 2\alpha)c_{j-k-1}$$

Basic Hermite polynomial

$$c_j^{n+1} = r_{-1}(\alpha)c_{j-k+1}^n + r_1(\alpha)c_{j-k-1}^n + s_{-1}(\alpha)\frac{dc}{d\alpha}\Big|_{j-k+1}^n + s_1(\alpha)\frac{dc}{d\alpha}\Big|_{j-k-1}^n$$

with

$$r_1(\alpha) = 0.25(\alpha^3 - 3\alpha + 2)$$

$$r_2(\alpha) = 0.25(-\alpha^3 + 3\alpha + 2)$$

$$s_1(\alpha) = 0.25(\alpha^3 - \alpha^2 - \alpha + 1)$$

$$s_2(\alpha) = 0.25(\alpha^3 + \alpha^2 - \alpha - 1)$$

Estimation of derivatives

$$\frac{dc}{d\alpha}\Big|_{j-k+1}^n = \tau \frac{dP_2(\alpha)}{d\alpha}\Big|_{\alpha=-1} + (1-\tau) \frac{dP_1(\alpha)}{d\alpha}\Big|_{\alpha=-1}$$

$$\frac{dc}{d\alpha}\Big|_{j-k-1}^n = \tau \frac{dP_2(\alpha)}{d\alpha}\Big|_{\alpha=1} + (1-\tau) \frac{dP_1(\alpha)}{d\alpha}\Big|_{\alpha=1}$$

where

$$P_1(\alpha) = \frac{(\alpha+2)(\alpha+1)}{2}c_{j-k+3} - (\alpha+3)(\alpha+1)c_{j-k+2} + \frac{(\alpha+3)(\alpha+2)}{2}c_{j-k+1}$$

$$P_2(\alpha) = \frac{\alpha(\alpha-1)}{2}c_{j-k+1} - (\alpha+1)(\alpha-1)c_{j-k} + \frac{\alpha(\alpha+1)}{2}c_{j-k-1}$$

$$P_3(\alpha) = \frac{(\alpha-2)(\alpha-3)}{2}c_{j-k-1} - (\alpha-1)(\alpha-3)c_{j-k-2} + \frac{(\alpha-1)(\alpha-2)}{2}c_{j-k-3}$$

Table 5

Truncation Errors for Alternative Interpolation Schemes
($\alpha = \beta = u\Delta t/\Delta x$)

<u>Scheme</u>	<u>Truncation error</u>
2P-LI2	$\frac{u}{2}(\Delta x - u\Delta t)\frac{\partial^2 c}{\partial x^2} + \text{higher order derivatives (HOD)}$
3P-LI3	$\frac{u}{6}(u^2\Delta t^2 - \Delta x^2)\frac{\partial^3 c}{\partial x^3} + \text{HOD}$
4P-LI4	$-\frac{3}{128}(9u^4\Delta t^3 + 10u^2\Delta t\Delta x^2 - \frac{\Delta x^4}{\Delta t})\frac{\partial^4 c}{\partial x^4} + \text{HOD}$
4P-LR2	$\frac{u}{24}(u^3\Delta t^3 - 2u^2\Delta t^2 - u\Delta t\Delta x^2 + 2\Delta x^3)\frac{\partial^4 c}{\partial x^4} + \text{HOD}$
5P-LR3	$-\frac{u}{120}(u^4\Delta t^4 - 5u^2\Delta t^2\Delta x^2 - 4\Delta x^4)\frac{\partial^5 c}{\partial x^5} + \text{HOD}$
6P-LR2	$(\dots)\frac{\partial^5 c}{\partial x^5} + \text{HOD}$
8P-LR2	$(\dots)\frac{\partial^8 c}{\partial x^8} + \text{HOD}$
2P-HI2	$-\frac{u}{24}(u^2\Delta t^3 - 2u\Delta t^2\Delta x + \Delta t\Delta x^2)\frac{\partial^4 c}{\partial x^4} + \text{HOD}$
4P-HL2	$\frac{u}{12}(2u^2\Delta t^2 - 3u\Delta t\Delta x + \Delta x^2)\frac{\partial^3 c}{\partial x^3} + \text{HOD}$
5P-HL3	$-\frac{u}{24}(u^2\Delta t^3 + \Delta x^2)\frac{\partial^4 c}{\partial x^4} + \text{HOD}$
7P-HL3	$\frac{u}{6}(u^2\Delta t^2 - \Delta x^2)\frac{\partial^3 c}{\partial x^3} + \text{HOD}$
6P-PL2	$(\dots)c + (\dots)\frac{\partial c}{\partial x} + (\dots)\frac{\partial^2 c}{\partial x^2} + \text{HOD}$
8P-PL2	$-\frac{u}{24}(u^2\Delta t^3 - 2u\Delta t^2\Delta x + \Delta t\Delta x^3)\frac{\partial^4 c}{\partial x^4} + \text{HOD}$

Table 4

Truncation Errors for Alternative Interpolation Schemes
(general case)

Scheme	Truncation error
2P-LI2	$\alpha(1-\alpha)\frac{\Delta x^2}{2\Delta t}\frac{\partial^2 c}{\partial x^2} + \text{higher order derivatives (HOD)}$
3P-LI3	$\alpha(\alpha^2-1)\frac{\Delta x^3}{6\Delta t}\frac{\partial^3 c}{\partial x^3} + \text{HOD}$
4P-LI4	$-\frac{3}{128}(9\alpha^4+10\alpha^2-1)\frac{\Delta x^4}{\Delta t}\frac{\partial^4 c}{\partial x^4} + \text{HOD}$
4P-LR2	$\alpha(\alpha-2)(\alpha^2+1)\frac{\Delta x^4}{24\Delta t}\frac{\partial^4 c}{\partial x^4} + \text{HOD}$
5P-LR3	$-\alpha(\alpha^2-1)(\alpha^2-4)\frac{\Delta x^5}{120\Delta t}\frac{\partial^5 c}{\partial x^5} + \text{HOD}$
6P-LR2	$(\dots)\frac{\Delta x^6}{\Delta t}\frac{\partial^6 c}{\partial x^6} + \text{HOD}$
8P-LR2	$(\dots)\frac{\Delta x^8}{\Delta t}\frac{\partial^8 c}{\partial x^8} + \text{HOD}$
2P-HI2	$(\alpha^2-\alpha)^2\frac{\Delta x^4}{24\Delta t}\frac{\partial^4 c}{\partial x^4} + \text{HOD}$
4P-HL2	$-\alpha(\alpha-1)(2\alpha-1)\frac{\Delta x^3}{12\Delta t}\frac{\partial^3 c}{\partial x^3} + \text{HOD}$
5P-HL3	$-\alpha^2(\alpha^2+1)\frac{\Delta x^4}{24\Delta t}\frac{\partial^4 c}{\partial x^4} + \text{HOD}$
7P-HL3	$-\alpha(1-\alpha^2)\frac{\Delta x^3}{6\Delta t}\frac{\partial^3 c}{\partial x^3} + \text{HOD}$
6P-PL2	$(3.7\alpha-4.2)\alpha\cdot 10^{-4}c+\frac{\Delta x}{\Delta t}[\alpha^2+(3.7K+3.5)\alpha+(0.9-4.2K)]\alpha\cdot 10^{-4}\frac{\partial c}{\partial x} +$ $+\frac{\Delta x^2}{2\Delta t}[\alpha^2(2K-1443.8)+\alpha(3.7K^2+7K+320.1)+(-4.2K^2+18K-314.3)]\alpha\cdot 10^{-4}\cdot$ $\frac{\partial^2 c}{\partial x^2} + \text{HOD}$
8P-PL2	$(\alpha^2-\alpha)^2\frac{\Delta x^4}{24\Delta t}\frac{\partial^4 c}{\partial x^4} + \text{HOD}$

THE ACCURACY OF EULERIAN-LAGRANGIAN METHODS

by

António Melo Baptista

Massachusetts Institute of Technology

Cambridge, Massachusetts 02139

January 1987

1. INTRODUCTION

The numerical solution of the transport equation, describing the fate of a passive scalar in a moving fluid, has been the object of intense research for the past few decades. A review of available solution methods [B2, B5] suggests that they fit into three major categories: Eulerian (EM), Lagrangian (LM) and Eulerian-Lagrangian (ELM). EM, which historically were the first to be introduced and are still very popular, have strong shortcomings in the analysis of transport problems where advection has a significant role vis à vis dispersion (the case for most natural flows) and where sharp gradients in the flow direction can not be resolved with a reasonable grid size (often the case for pollutant transport near sources or fronts). In turn, LM, which perform extremely well for pure advective transport, run into practical difficulties whenever dispersion has also to be solved, and have hardly been used in the context of realistic problems. ELM combine the best aspects of EM and LM, having the potential to provide accurate solutions for the range of advection-dominated to dispersion-dominated transport problems.

Several research groups have, in the past few years, attempted to explore the potential of ELM (e.g., Holly and co-workers [H2-H4, K1-K2], Benque, Hauguel and co-workers [B10], Newman and co-workers [N1-N3], and Baptista and co-workers [A1, B1-B5, K3]; other relevant works in the subject include [B1], [C2], [G1], [H1], [L1] and [V1]). Because ELM are relatively new, and lack tradition, the emphasis of most of the research has been on developing solution strategies, observing the resulting errors (mostly through numerical experimentation), identifying possible causes of

error, and suggesting a fix-up.

This trial-and-error procedure has generated a choice of specific ELM, and has pin-pointed some of their potentials and limitations; however, the "broad picture" is still missing, while it is already clear that the ability to recognize and explore such picture is, in ELM even more than in other methods, fundamental for efficiency and accuracy.

In particular, given their very "physical" approach to the solution of the transport equation, and the inherent decoupling of the full problem in fundamentally different steps (tracking of characteristic lines, interpolation to find the initial conditions associated to each these lines, and solution of dispersion plus internal sources and sinks along these lines), ELM have great flexibility in combining, both in time and in space, different procedures to solve specific tasks; an example of this flexibility is the choice of the interpolator to find the concentrations at the feet of the characteristic lines, which may vary in space (in such a way to replace or complement local grid refinement), or in time (e.g., to accomodate changes in the characteristic gradients of concentration in the flow direction). Also, due to the use of the BMC for the solution of the advection equation, ELM can handle accurately a much broader range of time step/space step ratios than conventional EM, and, indeed, the increase of the time step may (depending on the relative role of dispersion, mean advection, and differential advection) lead to improved accuracy.

In this study, we seek basic insight on the dependences and constraints of ELM accuracy, so as to provide modellers with a reliable conceptual framework for their decisions on computational strategy. The

analysis is based on the derivation and examination of the truncation errors of a 1-D algorithm that is flexible to accommodate different choices for time-discretization and for spatial interpolation at the feet of the characteristic lines.

2. BRIEF REVIEW OF CONCEPTS AND SPECIFIC IMPLEMENTATIONS OF ELM

The conceptual approach behind most ELM is illustrated in Figure 1, for 1-D (actual implementations of the procedure have been used also in 2-D, and could be extended to 3-D). The concentrations at the nodes of the computational grid are found, at time n , through a three-step procedure:

- Definition of characteristic lines that start at each grid node, j , at time n , and follow the flow backwards until time $n-1$ or a boundary is reached.

- Calculation, by interpolation from known nodal concentration values at time $n-1$ (or at boundaries), of the concentration at the feet, ξ , of the characteristic lines; these concentrations would correspond to the concentrations at time n , if advection was the only transport mechanism. More importantly, they are also the correct initial conditions for the transport problem, written in Lagrangian coordinates between times $n-1$ and n .

- Solution of the transport equation in the coordinate system defined by all characteristic lines, taking as initial conditions the concentrations at the feet of these lines.

To implement this procedure, ELM typically split the transport

equation, written either in its differential form (e.g., [N1-N3]) or, more commonly, in its time-discretized form, into two sub-equations (advection and dispersion). Advection is most often solved by the the Backwards Method of Characteristics BMC (which implements the the first step of the previously described procedure), using a choice of interpolators, reviewed in detail by [B7]; a Forward Method of Characteristics was alternatively used by [B10], but, to our knowledge, does not present any specific advantage, and may lead to practical problems linked to the uncontrolled deformation of the grid. Dispersion is typically solved by combining a finite-difference (e.g., [H2-H4], [K1-K2], [L1], [G1], [B10]) or finite-element (e.g. [B2], [B9], [H1]) discretization in space with an Euler implicit or a Crank-Nicholson--defined along the characteristic lines--discretization in time (e.g., [B2] and [H3], respectively. Virtually all methods have been used in connection with some local forward tracking procedure, to handle gradients that the basic ELM can not handle (e.g., near sources [H4, K1] or fronts [N1]). Physical, chemical, or biological transport processes (source/sink terms), if present, can be treated within the dispersion step, or as a separate, fourth step.

3. REFERENCE ALGORITHM

The ELM taken for reference in this study follows closely the conceptual procedure described in the previous section, being flexible to accomodate two alternative interpolators and two alternative time-discretization schemes. The governing equation is assumed to be of the form

$$\frac{Dc}{Dt} = \frac{\partial c}{\partial t} + u(x,t) \frac{\partial c}{\partial x} = L \quad (1)$$

where L represents the dispersion operator, of the form $D \cdot \partial^2 c / \partial x^2$. The description of the techniques selected to perform each of the steps of the solution follows (a 1-D uniform grid is assumed, as in Figure 1).

Tracking of the characteristic lines

The tracking of the characteristic lines, within each global time step Δt , is performed by the backwards (i.e., from $n+1$ to n) solution, for each node of the grid, of the ordinary differential equation

$$\frac{dx}{dt} = -u(x,t) \quad (2)$$

with initial condition $x = x_j$. The algorithm is a 2nd order Runge-Kutta scheme, with time step $\delta t < \Delta t$; once ξ , at time n , is reached, Equation (2) is solved in reverse direction, so to find the position of the starting point, i.e., of node j at time $n+1$; let us denote the result by $x_{j,n}$. If $x_j - x_{j,n}$ is larger than a user-imposed tolerance, the time-step is adjusted by trial and error until the closure error becomes acceptable.

Interpolation of concentrations at the feet of the characteristic lines

To find the concentration at the feet of the characteristic lines, we use either a compact quadratic or a non-compact quartic Lagrange interpolator, both of which were studied in detail by [B7]. Their definition is presented in Figure 2.

Solution of dispersion

(A) Time-discretization

The discretization in time of the transport equation, written in Lagrange coordinates, takes the general form

$$\frac{c^{n+1} - c^{\bar{f}}}{\Delta t} = A \cdot L^{n+1} + B \cdot L^{\bar{f}} \quad (3)$$

where A and B are chosen alternatively as

A=1, B=0 Euler implicit

A=0.5, B=0.5 Crank-Nicholson

The space discretization of the dispersion operator resorts to centered differences, e.g.,

$$L^{n+1} \approx \frac{D}{\Delta x^2} (c_{j+1}^{n+1} - 2c_j^{n+1} + c_{j-1}^{n+1}) \quad (4)$$

4. TRUNCATION ERRORS

Each of the steps of the procedure described in the previous section may, and, in general, will, introduce numerical errors. However, errors in the tracking of the characteristic lines have rather unique properties, as a consequence of the fact that this step of the procedure is independent of the actual concentration field; in particular, errors can be kept below some pre-imposed threshold (defined in the form of maximum acceptable

distance between j and j') by reducing the tracking time-step, δt , without affecting the global time-step of the solution, Δt .

For the rest of this work, we will assume that the tolerance criterion is set restrictive enough for tracking to be "exact," or, more precisely, to have errors that are negligible when compared to errors in the other steps. We note that this is not so much an approximation to allow for formal error analysis; rather, it should (although apparently is not--e.g., [H4], [B10]) be clear that this is a necessary condition to give reliability to the all procedure: if the characteristic lines are poorly tracked, none of the steps that follow can be expected to provide any consistent correction--Figure 2.

To understand the nature of the errors in the remaining steps of the numerical procedure, it is useful to examine the truncation errors of the numerical algorithm. These were derived in Appendix A, and can be expressed as

$$\epsilon \equiv \left\{ \frac{\partial c}{\partial t} + u(x, t) \frac{\partial c}{\partial x} - L \right\} = \epsilon_{int} + \epsilon_{dif}^{sd} + \underbrace{\epsilon_{dif}^{tdm} + \epsilon_{dif}^{tdf}}_{\epsilon_{dif}^{td}} \quad (5)$$

where

ϵ_{int} - is the error in the interpolation of concentrations at the feet of characteristic lines

ϵ_{dif}^{sd} - is the error due to the space discretization of the dispersion operator

ϵ_{dif}^{td} - is the error due to the time-discretization of the dispersion

operator along the characteristic lines; it may be conveniently divided into two components: ϵ_{dif}^{tdm} , associated with the time-discretization along parallel characteristic lines defined by a mean flow; and ϵ_{dif}^{tdf} accounting for the fact that, due to flow non-uniformity, the characteristic lines may come closer or further away from each other as time progresses.

Table 1 summarizes the expressions for these errors, for the different choices of interpolators and time-discretization schemes considered in the previous section, and Table 2 summarizes, for the choice of a quadratic interpolator and an Euler implicit time-discretization, the ratios between the different types of errors (within a same time step).

No individual type of error can be identified *a priori* as dominant over the others. Indeed each type of error is affected differently by the characteristics of the transport problem (specifically, concentration derivatives in space, flow velocity and respective derivatives in space and time), and by the adopted space and time discretizations. Both the absolute accuracy and the relative importance of each error are seen to be controlled by dimensionless parameters, which can conveniently be classified in two groups:

Controlling parameters that are independent of the concentration field

Courant number, $Cu = u \cdot \Delta t / \Delta x$

Dispersion number, $Di = D \cdot \Delta t / \Delta x^2$

Differential Courant number, $Cu_f = \Delta u \cdot \Delta t / \Delta x$

Peclet number, $Pe = \frac{Cu}{Di} = \frac{u \cdot \Delta x}{D}$

$$\text{Differential Peclet number, } Pe_f = \frac{\Delta u \cdot \Delta x}{D}$$

We note that only the first three parameters are independent; Peclet numbers were kept for convenience. The magnitude of the spacial variability of the flow was scaled by Δu ; for realistic problems, u and D will also have to be interpreted as scales, rather than as local values.

Controlling parameters that depend on the concentration field

These parameters concern the ratios between concentration derivatives of different orders, weighted, when $\epsilon_{\text{dif}}^{\text{tdf}}$ is involved, by velocities and/or velocity derivatives. Their actual number and form depend on the choice of the interpolator and of the space- and time-discretization schemes. To minimize their number, we have to scale space and time derivatives (although the latter will have typically a secondary effect); the following scaling form is suggested

$$\frac{\partial^p}{\partial x^p} \approx \frac{1}{\ell^p} \equiv \frac{1}{(N\Delta x)^p} \quad (6)$$

$$\frac{\partial^p}{\partial t^p} \approx \frac{1}{\tau^p} \equiv \frac{1}{(N\Delta t)^p} \quad (7)$$

5. ACCURACY DEPENDENCE ON THE COMPUTATIONAL STRATEGY

5.1. General aspects

For a given transport problem, the accuracy and the cost of the reference ELM will depend on the computational strategy, which includes

both the choice of the space and the time discretization, and the choice of the interpolator and the time-discretization scheme. This section provides conceptual support for a rational decision-making process, by analysing the dependence of accuracy on the relevant choices; often we will use as a reference, for illustration purposes, the problem of the transport of an instantaneous source, as characterized in Figure 4.

5.2. Accuracy dependence on Δx

The effect of Δx is reflected on global accuracy through the errors associated with the interpolation at the feet of the characteristic lines and with the spatial discretization of the dispersion operator. In both cases, accuracy improves as Δx decreases, but errors will in general have significantly different magnitudes, and vanish at different rates. Interpolation errors will dominate for large Pe (or, similarly, in the case of $Cu \gg 1$, for small Dispersion numbers). The meaning of "small" and "large" (e.g., see Table 2) is interpolator-dependent; for most combinations of realistic transport problems and feasible grids, interpolation errors will be dominant, although the use of a quartic, rather than a quadratic, interpolator for advection brings these two types of errors closer to each other--which suggests improved efficiency of the overall solution procedure.

For a fixed Δx , accuracy is better for problems with smoother gradients than for problems with steeper gradients, a well-known behavior. In particular, dispersion has a smoothing effect on the gradients, as time progresses, and therefore the increase of D , although increasing ϵ_{dif}^{sd} with regard to ϵ_{int} , contributes ultimately to reduce the absolute value of both

ϵ_{dif}^{sd} and ϵ_{int} .

5.3. Accuracy dependence on Δt

The dependence of the accuracy on Δt is more complex than that on Δx . Indeed, depending on the relative roles of advection, differential advection and dispersion, decreasing Δt may either improve or deteriorate accuracy. This is analysed starting with the case of pure advection, and evolving towards the full transport equation.

Pure advection

In this case, the only errors come from the interpolation at the feet of the characteristic lines. Within each time step, interpolation errors depend on Δt only through the location of the feet of the characteristic lines within the core element, characterized by α . For large Cu , α is a weak function of Cu , depending only on its fractional part, and therefore errors per time step are, in this case, virtually independent of Δt (for u and Δx fixed). As a consequence, to solve the transport equation between 0 and T , we should look for the minimum number of interpolations (one, in the limit, if feasible), hence for the largest possible Δt .

Choosing Δt small, so as to decrease Cu , a must in EM, is then a generally poor option in the BMC solution of the advection equation. We should stress, however, that the improvement of accuracy as Δt increases is legitimate, in the sense that the numerical solution does not blow up in the limit of small Δt : it rather tends to an accuracy plateau that depends on Δx and its ability to resolve prevailing concentration gradients.

Indeed, as suggested by the form of ϵ_{int} for $Cu = \alpha$ (Table 1), and formally shown by [B7], the BMC is, for the interpolators considered in the present work, unconditionally consistent, stable and convergent.

The effect of dispersion

We recognized before that dispersion has the effect of smoothing gradients, and, therefore, should ultimately improve accuracy, but that it also introduces additional errors, associated with both the space and time discretization of the dispersion operator. The errors in the time-discretization have particular significance, as they decrease with the time step Δt , and, therefore, may reduce or eliminate the ability of the BMC to accurately handle large Δt .

The time-step that, at a given point in time, minimizes overall errors (optimal time-step) is the one than can bring ϵ_{int} and ϵ_{dif}^{td} as close together as possible. For instance, in the case of the quadratic interpolator and the Euler implicit time-discretization, simple combination of the expression of these errors indicates the optimal time-step is, in order of magnitude

$$\Delta t_{opt} \approx \left[\frac{\alpha(\alpha^2-1)}{3 \cdot D^2} \frac{\frac{\partial^3 c}{\partial x^3}}{\left| \Delta x \frac{\partial^4 c}{\partial x^4} - \frac{\Delta x}{D} \left\{ \frac{\partial^2 u}{\partial x^2} \frac{\partial c}{\partial x} + 2 \frac{\partial u}{\partial x} \frac{\partial^2 c}{\partial x^2} \right\} \right|} \right]^{0.5} \Delta x^2 \quad (8)$$

Expressions of this form can be similarly derived for other interpolators and time-discretization schemes (based on Table 1), and, given a specific problem, they can be simplified by proper scaling

derivatives, so as to help the selection of Δt as a function of Δx ; e.g., using the scales of Section 4, Equation 8 becomes

$$\Delta t_{\text{opt}} \approx \left\{ \frac{\alpha(\alpha^2-1)}{3D^2} \frac{M}{1-M_0 Pe_f} \right\}^{1/2} \cdot \Delta x^2 \quad (9)$$

where now M and Δu have to be specified consistently with the physical problem (e.g., Figure 4).

Plots showing the dependence of the total truncation error on Δt may however be considerably more informative. Examples of these plots, for specific choices of (u , D , Δu , M , and Δx) are shown in Figures 5 to 9. They may be interpreted, for instance, as referring to the problem of the transport of a Gauss-hill in a 1-D flow. Figures 5 to 8 assume the use of a quadratic interpolator and a Crank-Nicholson time-discretization; Figure 9 refers to alternative choices of interpolators and time-discretization schemes.

We observe that:

- The optimal Δt is, in a large range of situations, such that $Cu > 1$ (in contrast with what happens in EM). The optimal Δt varies significantly, though, with all controlling parameters identified in Section 4; in particular, for a given problem and grid, it may vary as time progresses, if the characteristic concentration gradients change (e.g., by effect of dispersion) or if the role of differential advection relative to mean advection and dispersion changes (e.g., as in a tidal flow).

- Large velocity gradients may reduce the optimal time-step very

significantly (e.g., Figure 8). This reduction is larger for larger Pe_f and for smoother concentration gradients, i.e., larger M , in which case ϵ_{dif}^{tdf} is dominant over ϵ_{dif}^{tdm} --e.g., see Table 2. The overall accuracy of the solution shows no dependence on Pe_f in the region where interpolation errors are dominant, but it consistently deteriorates with increasing Pe_f in the zone where time-discretization errors are dominant; the limit between these zones, for a given problem and grid, determines the optimal time step, and depends on all Cu , Di , Pe , Cu_f , Pe_f and M (e.g., see Table 2).

- Smooth concentration gradients may lead to either larger optimal time-steps (e.g., Figure 6) than those for steeper gradients, or to no optimal time-step (accuracy improving all the way down to the minimal possible Δt) depending on the choice of the interpolator and the time-discretization. However, errors are, for smoother gradients, consistently much smaller, not only in the sense of the minimal achievable error, but also for fixed Δt . This suggests that, when gradients are smooth (e.g., due to the long term effect of dispersion), the use of an optimal time step may be of secondary interest, for practical purposes.

- The dependence of the accuracy on the time step is also relaxed (in the sense of a large range of Δt providing similar accuracy) when the errors associated with the spacial discretization play a significant role; this is typically associated with small Cu_f , with small Cu , and with Di numbers in some intermediate range (e.g., see Table 2).

5.4. Accuracy dependence on the interpolator and time-discretization

Taking as a reference some choice of interpolator and time-discretization (a quadratic interpolator and the Euler time-discretization scheme, say) the optimal time step will increase if the order of the time-discretization is increased (e.g., Crank-Nicholson instead of Euler), keeping fixed the interpolator, and will decrease if the order of the interpolator is increased (e.g., quartic instead of quadratic), keeping the time-discretization scheme fixed; in both cases, optimal accuracy will improve, but the trade-off in costs should be considerably different.

If the order of both the interpolator and the time-discretization are changed the optimal time may either decrease or increase; the choice of a quartic interpolator and a Crank-Nicholson discretization tends to decrease the optimal time step, suggesting that interpolation errors are being helped further than time-discretization errors (not surprising, considering that the order of the interpolator is increased by two, while that of the time-discretization is increased only by one).

For uniform flows, i.e., parallel characteristic lines, it would be conceivable to use for dispersion (as we did for the tracking) a time-step smaller than Δt (dt , say), so as to take larger benefit of the ability of the BMC to handle large time-steps--i.e., to extend the optimal time step. However, flow non-uniformity prevents this strategy, which would be effective only if the solution of dispersion could be done following the characteristic lines, hence requiring interpolations at each dt (which would then necessarily coincide with Δt).

Hence, for a fixed interpolator, the ability to increase the optimal

time-step requires increasing the order of the time-discretization scheme, which involves a trade-off in cost. The difference in cost between the Euler and Crank-Nicholson schemes is minor, and we do recommend that the latter be used in a systematic basis. Further increasing the order of the time-discretization scheme may or may not be cost-efficient, however; indeed, 3rd order accuracy in time would require considerable additional work (including tracking), as illustrated in Figure 10.

6. FINAL CONSIDERATIONS

Remarkable characteristics of ELM, as compared to more conventional EM, include the ability of the former (a) to flexibly choose between different spatial interpolators for advection (without having to change the grid discretization, an/or increase the bandwidth of matrices required for the solution of dispersion), and between different time-discretization schemes, and (b) to use large time-steps, well beyond $Cu > 1$, conditional only to the time-discretization adopted to solve dispersion along the characteristic lines that follow the flow.

The efficient use of ELM is, however, more complex than that of EM: while ELM will in general be at least as accurate as EM that use similar space and time-discretizations for the dispersion operator, the accuracy of the former can often be vastly improved by appropriate decisions on the computational strategy, and, in particular, on the choice of Δt and on the local (in time and/or space) increase of the order of the interpolator for advection and/or the time-discretization.

The analysis of the truncation errors of ELM, for alternative

interpolators and time-discretization schemes, is of considerable interest in providing a reference for the different choices involved, as shown in this study. Such analysis has necessarily to include an assessment of the effect of the flow non-uniformity, which plays a key role in the optimal time-step, and, for larger time-steps, on the actual accuracy of ELM.

To our knowledge, this study provides the first conceptual "model" to guide the choice of the ELM computational strategy. Priority steps to improve such model include its extension to 2-D (flow non-uniformity is more complex in multi-dimensions), and its extensive application in connection with industrial codes and actual engineering problems. The success of this application is expected to require familiarity with the trade-offs involved, common sense in the choice of the scales characterizing the physical problem, and ingenuity in handling problems involving multiple scales and/or irregular grids.

REFERENCES

- A1 Adams, E. E., R. Kossik, and A. M. Baptista. "Source Representation in a Numerical Transport Model." *Finite Elements in Water Resources: Proceedings of the 6th International Conference* (Sá da Costa et al., Ed.), Springer-Verlag, 1986.
- A2 Aldama-Rodríguez, A. "Theory and Applications of Two- and Three-Scale Filtering Approaches for Turbulent Flow Simulation." Ph.D. Thesis, M.I.T., 1985.
- B1 Baptista, A. M., E. E. Adams, and K. D. Stolzenbach. "The 2-D Unsteady Transport Equation Solved by the Combined Use of the Finite Element Method and the Method of Characteristics", in *5th Int. Conf. on Finite Elements in Water Resources*. Burlington, Vermont, 1984a.
- B2 Baptista, A. M., E. E. Adams, and K. D. Stolzenbach. "Eulerian-Lagrangian Analysis of Pollutant Transport in Shallow Water." MIT R. M. Parsons Laboratory, Technical Report 296, 1984.
- B3 Baptista, A. M., E. E. Adams, and K. D. Stolzenbach. "Comparison of Several Eulerian-Lagrangian Models to Solve the Advection-Diffusion Equation." in *Int. Symp. on Refined Flow Modeling and Turbulence Measurements*. U. Iowa, USA, 1985.
- B4 Baptista, A. M., E. E. Adams, and K. D. Stolzenbach. "Accuracy Analysis of the Backwards Method of Characteristics." in *Finite Elements in Water Resources: Proceedings of the 6th International Conference*. (Sá da Costa et al., Ed.), Springer-Verlag, 1986.
- B5 Baptista, A. M. "Accurate Numerical Modeling of Advection-Dominated Transport of Passive Scalars." LNEC, Lisboa, 1986.
- B6 Baptista, A. M. "Fourier Analysis of the Backwards Method of Characteristics." Ph.D. Thesis, M.I.T., 1987.
- B7 Baptista, A. M. "The Consistency, Stability, and Convergence of the Backwards Method of Characteristics." Ph.D. Thesis, M.I.T., 1987.
- B8 Baptista, A. M. "The Choice of the Interpolator for the Backwards Method of Characteristics." Ph.D. Thesis, M.I.T., 1987.
- B9 Benque, J. P., G. Labadie, and G. Ibler. "A Finite Element Method for Navier-Stokes Equations." *3rd Conf. on Fin. Elem. in Flow Problems*, 1980.
- B10 Branski, J. M. "Higher-Order Spline Schemes for the Advection - Diffusion Equation." Submitted to ASCE, 1986.
- C1 Celia, M. A., and W. G. Gray. "An Improved Isoparametric

- Transformation for Finite Element Analysis." *International Journal for Numerical Methods in Engineering* 20:1443-1459, 1984.
- C2 Cheng, R. T., V. Casulli, and S. Milford. "Eulerian-Lagrangian Solution of the Convection-Diffusion Equation in Natural Coordinates." *Water Resources Research* 20(7):944-952, 1984.
- G1 Glass, J., and W. Rodi. "A Higher Order Numerical Scheme for Scalar Transport." *Comp. Math. in Appl. Mech. and Eng.* 31:337-358, 1982.
- H1 Hasbani, Y., E. Livne, and M. Bercovier. "Finite Elements and Characteristics Applied to Advection-Diffusion Equations." *Computer and Fluids* 11(2):71-83, 1983.
- H2 Holly, F. M., Jr., and T. Komatsu. "Derivative Approximations in the Two-Point Fourth-Order Method for Pollutant Transport." *Proceedings of the Conference on Frontiers in Hydraulic Engineering*, ASCE, M.I.T., 1984.
- H3 Holly, F. M., Jr., and J. M. Polatera. "Dispersion Simulation in 2-D Tidal Flow." *Journal Hydr. Engrg.*, ASCE, 1984.
- H4 Holly, F. M., Jr., and A. Preissmann. "Accurate Calculation of Transport in Two Dimensions." *Journal of the Hydraulics Division*, ASCE - 103(HY11):1259-1278, Nov. 1977.
- K1 Komatsu, T., F. M. Holly, Jr., and N. Nakashiki. "Numerical Calculation of Pollutant Transport in Rivers and Coastlines", in *4th Congress, Asian and Pacific Division, IAHR*, Chiang Mai, Thailand, 1984.
- K2 Komatsu, T., F. M. Holly, Jr., N. Nakashiki, and K. Ohgushi. "Numerical Calculation of Pollutant Transport in One and Two Dimensions." *Journal of Hydroscience and Hydraulic Engineering* 3(2):15-30, 1985.
- K3 Kossik, R. F. "Tracing and Modeling Pollutant Transport in Boston Harbor." M.Sc. thesis, MIT, 1984.
- L1 Leith, C. E. "Numerical Simulation of the Earth's Atmosphere." *Methods in Computational Physics* 4:1-28, 1965.
- N1 Neuman, S. P. "An Eulerian-Lagrangian Scheme for the Dispersion Convection Equation Using Conjugate Space-Time Grids." *Journal of Comp. Phys.* 41:270-279, 1981.
- N2 Neuman, S. P. "Adaptive Eulerian-Lagrangian Finite Element Method For Advection-Dispersion." *Int. Journal for Numerical Methods in Engineering* 20:321-337, 1984.
- N3 Neuman, S. P., and S. Sorek. "Eulerian-Lagrangian Methods for Advection-Dispersion." *Finite Elements in Water Resources*, (K. P. Mole et al., Ed.) 4:14.41-14.68, 1982.

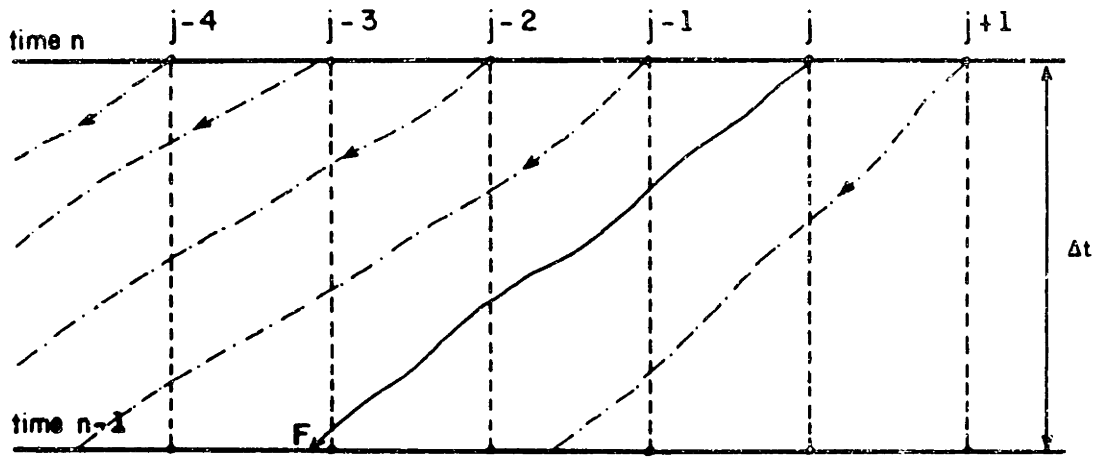
- R1 Rosman, P. C. "Modeling Shallow Water Bodies via Filtering Techniques." Ph.D. Thesis, M.I.T. (in preparation), 1987.
- V1 Varoglu, E., and W. L. Finn. "Space-Time Finite Elements Incorporating Characteristics for the Burgers Equation." *Int. J. Num. Meth. Engrg.* 16:171-184, 1980.

Figure 1

Illustrative sketch for Eulerian-Lagrangian methods

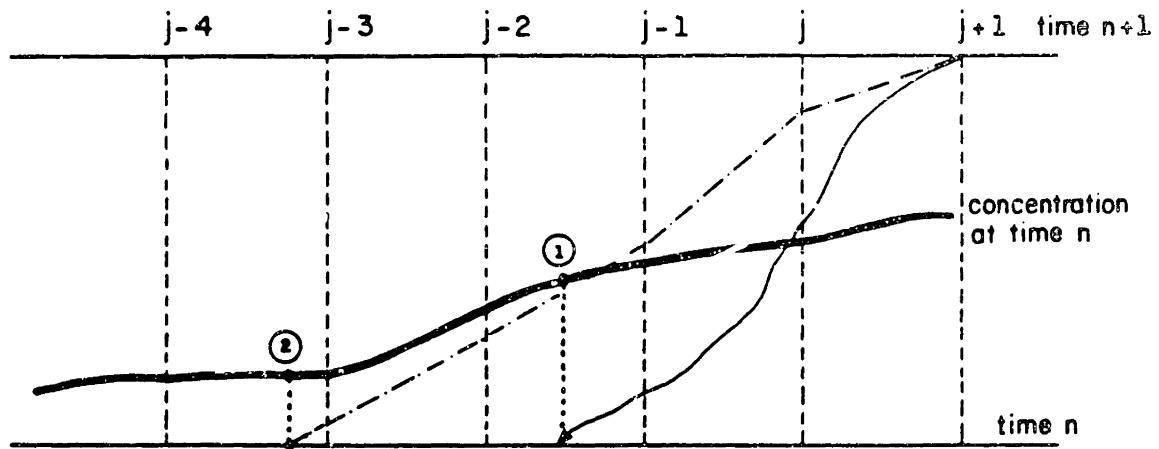
Required steps:

1. Tracking of the characteristic lines. For each node j , a characteristic line is independently defined by the backwards (i.e., between n and $n-1$) solution of an ordinary differential solution of the form $dx_1/dt = u_1$.
2. Interpolation at the feet of the characteristic lines. The concentration at the foot of each characteristic line is found by interpolation from known information on neighboring nodes (time $n-1$).
3. Solution of the transport equation, written in Lagrangian form. This solution involves all nodes simultaneously.



Scheme	Reference sketch	Definition	Q, M	Comments
3P-LI3		$f(\alpha) = \sum_{p=0}^1 \phi_p(\alpha) \cdot f_p$ $\phi_{-1}(\alpha) = \frac{1}{2}(\alpha^2 - \alpha)$ $\phi_0(\alpha) = 1 - \alpha^2$ $\phi_1(\alpha) = \frac{1}{2}(\alpha^2 + \alpha)$	2, 2	Quadratic Lagrange interpolator Compact, class C ₀ Reference: [B2, B9]
5P-LR3		$f(\alpha) = \sum_{p=0}^2 \phi_p(\alpha) \cdot f_p$ $\phi_{-2}(\alpha) = \frac{1}{24}(\alpha^4 - 2\alpha^2 - \alpha^2 + 2\alpha)$ $\phi_{-1}(\alpha) = -\frac{1}{6}(\alpha^4 - \alpha^2 - 4\alpha^2 + 4\alpha)$ $\phi_0(\alpha) = \frac{1}{4}(\alpha^4 - 5\alpha^2 + 4)$ $\phi_1(\alpha) = -\frac{1}{6}(\alpha^4 + \alpha^2 - 4\alpha^2 - 4\alpha)$ $\phi_2(\alpha) = \frac{1}{24}(\alpha^4 + 2\alpha^2 - \alpha^2 - 2\alpha)$	4, 4	

Figure 2 Definition of spatial interpolators



- 1 Actual concentration at the foot of the characteristic line
- 2 Concentration obtained due to poor tracking (assuming exact interpolation)

Figure 3 Illustration of the effect of poor tracking

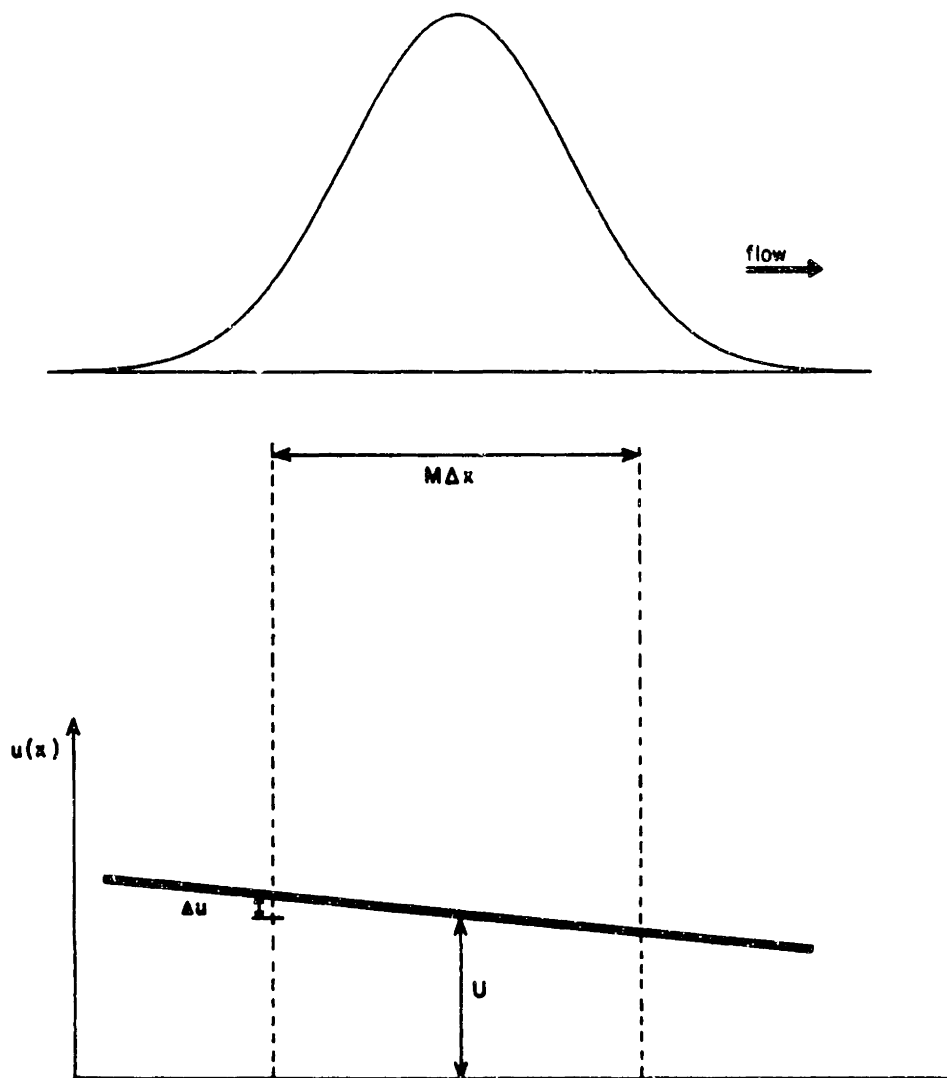


Figure 4 Example of the choice of scales for a specific problem

Figure 5

Dependence of errors on the nodal spacing, as a function of the time step

Computational parameters:

$$\Delta x = \begin{cases} 25 \\ 50 \\ 100 \end{cases} \quad \begin{array}{l} M = 500/\Delta x \\ Cu = \Delta t/\Delta x \\ Cu_f = 0.1 \Delta t/\Delta x \\ Di = \Delta t/\Delta x^2 \\ Pe = \Delta x \\ Pe_f = 0.1\Delta x \end{array}$$

variable Δt

$$T = 11000$$
$$u = 1$$
$$\Delta u = 0.1$$
$$D = 1$$

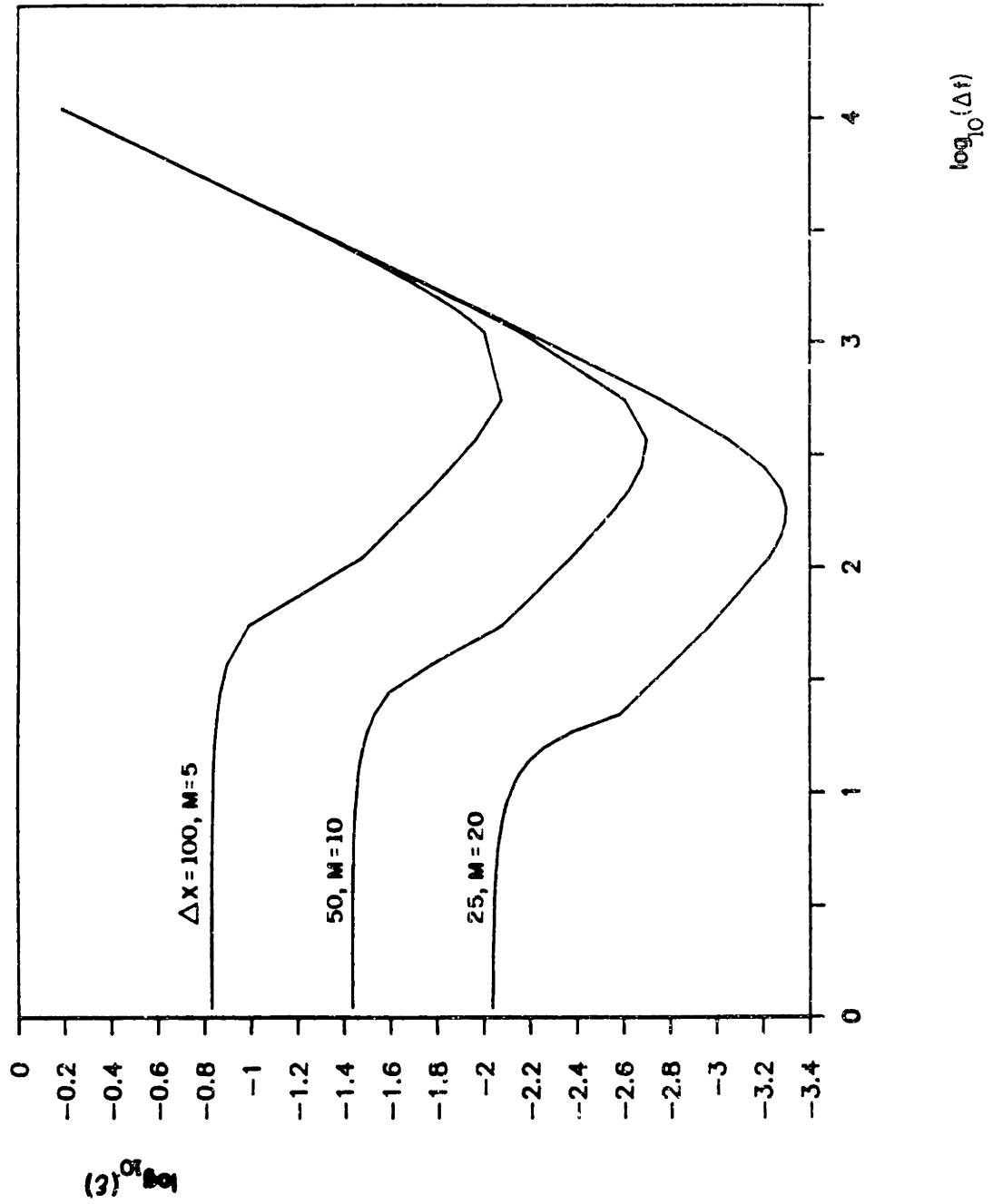


Figure 6

Dependence of errors on the dimensionless length scale for concentrations,
as a function of the time step

Computational parameters:

$$\Delta x = 100$$

Δt variable

$$T = 11000$$

$$u = 1$$

$$\Delta u = 0.1$$

$$D = 1$$

$$M = \begin{cases} 5 \\ 25 \\ 50 \end{cases}$$

$$Cu = \Delta t / 100$$

$$Cu_f = \Delta t / 1000$$

$$Dl = \Delta t / 10000$$

$$Pe = 100$$

$$Pe_f = 10$$

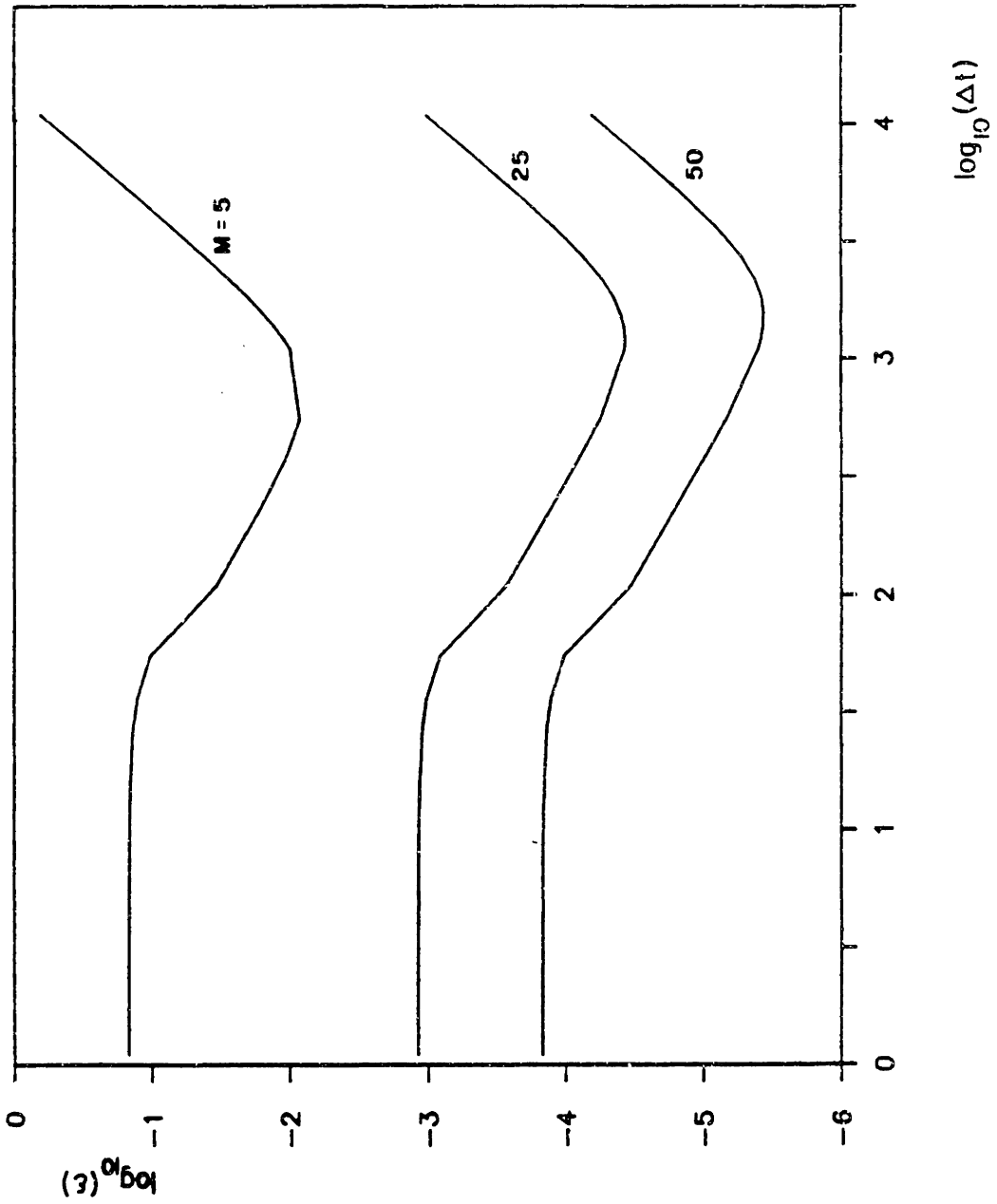


Figure 7

Dependence of errors on the dispersion coefficient,
as a function of the time step

Computational parameters:

$$\Delta x = 100$$

$$M = 5$$

$$\Delta t \text{ variable}$$

$$Cu = \Delta t/100$$

$$T = 11000$$

$$Cu_f = \Delta t/1000$$

$$u = 1$$

$$Di = D\Delta t/10000$$

$$\Delta u = 0.1$$

$$Pe = 100/D$$

$$D = \begin{cases} 1 \\ 0.1 \\ 0.01 \end{cases}$$

$$Pe_f = 10/D$$

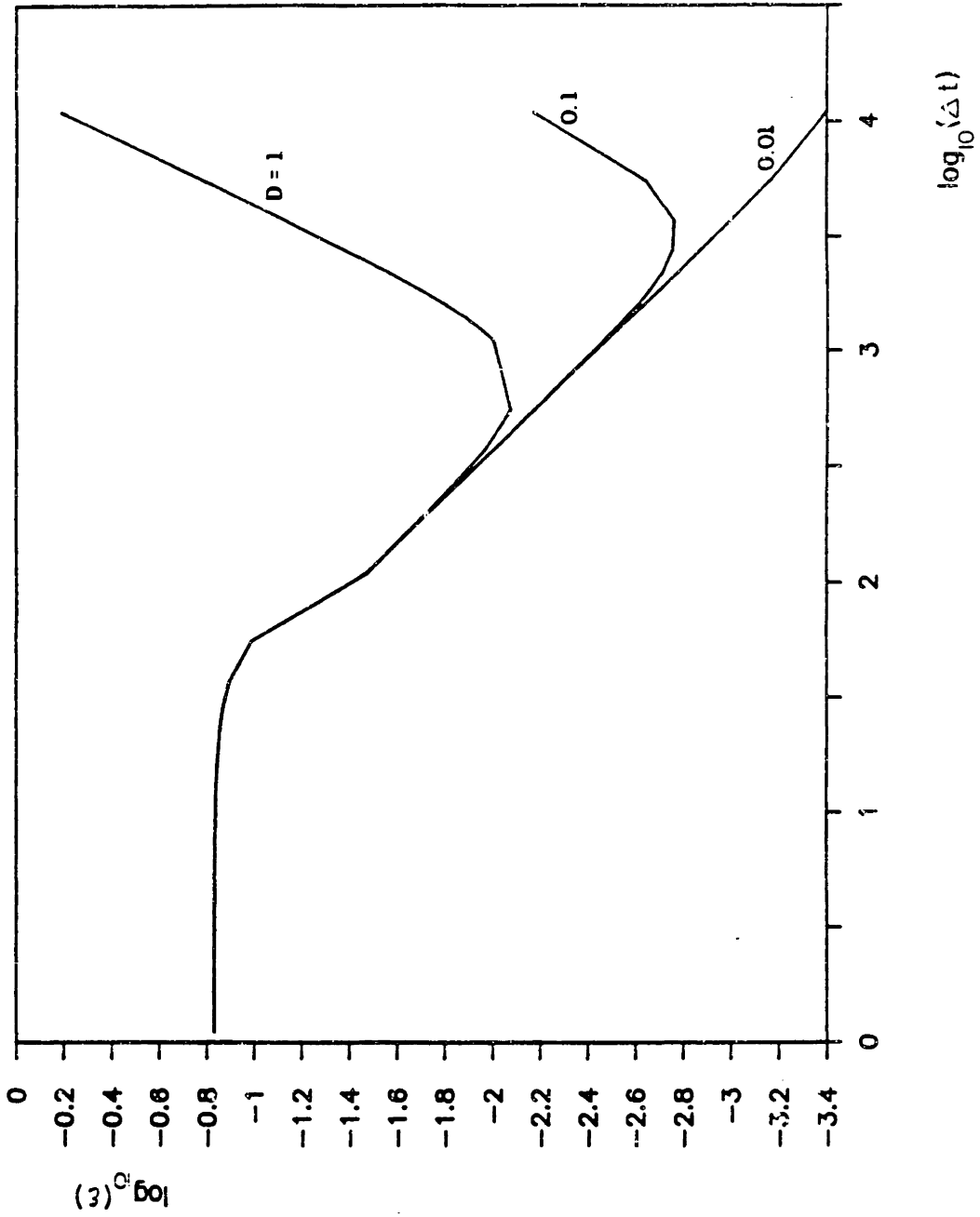


Figure 8

Dependence of errors on the flow non-uniformity,
as a function of the time step

Computational parameters:

$$\Delta x = 100$$

$$M = 5$$

$$\Delta t \text{ variable}$$

$$Cu = \Delta t/100$$

$$T = 11000$$

$$Cu_f = \Delta t \cdot \Delta t/1000$$

$$u = 1$$

$$Di = \Delta t/10000$$

$$\Delta u = \begin{cases} 0.1 \\ 0.01 \\ 0.001 \end{cases}$$

$$Pe = 100/D$$

$$Pe_f = 100\Delta u/D$$

$$D = 1$$

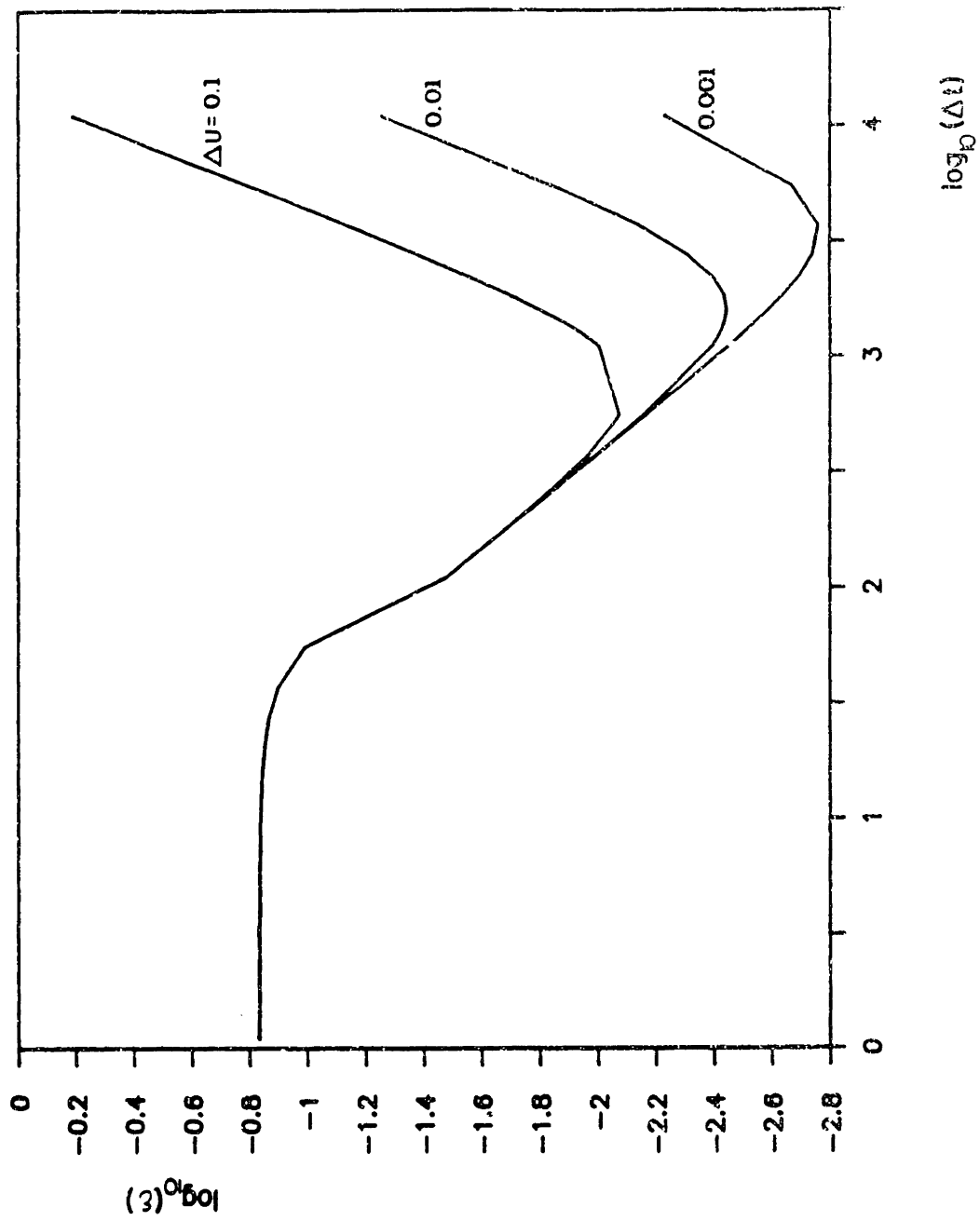


Figure 9

Dependence of errors on the spatial interpolator
and on the time-discretization scheme

Computational parameters:

$$\Delta x = 100$$

$$M = 5$$

$$\Delta t \text{ variable}$$

$$Cu = \Delta t/100$$

$$T = 11000$$

$$Cu_f = \Delta t/1000$$

$$u = 1$$

$$Di = \Delta t/10000$$

$$\Delta u = 0.1$$

$$Pe = 100$$

$$D = 1$$

$$Pe_f = 10/D$$

Legend

- 1: Quadratic interpolator; Euler time-discretization
- 2: Quadratic interpolator; Crank-Nicholson time-discretization
- 3: Quartic interpolator; Euler time-discretization
- 4: Quartic interpolator; Crank-Nicholson time-discretization

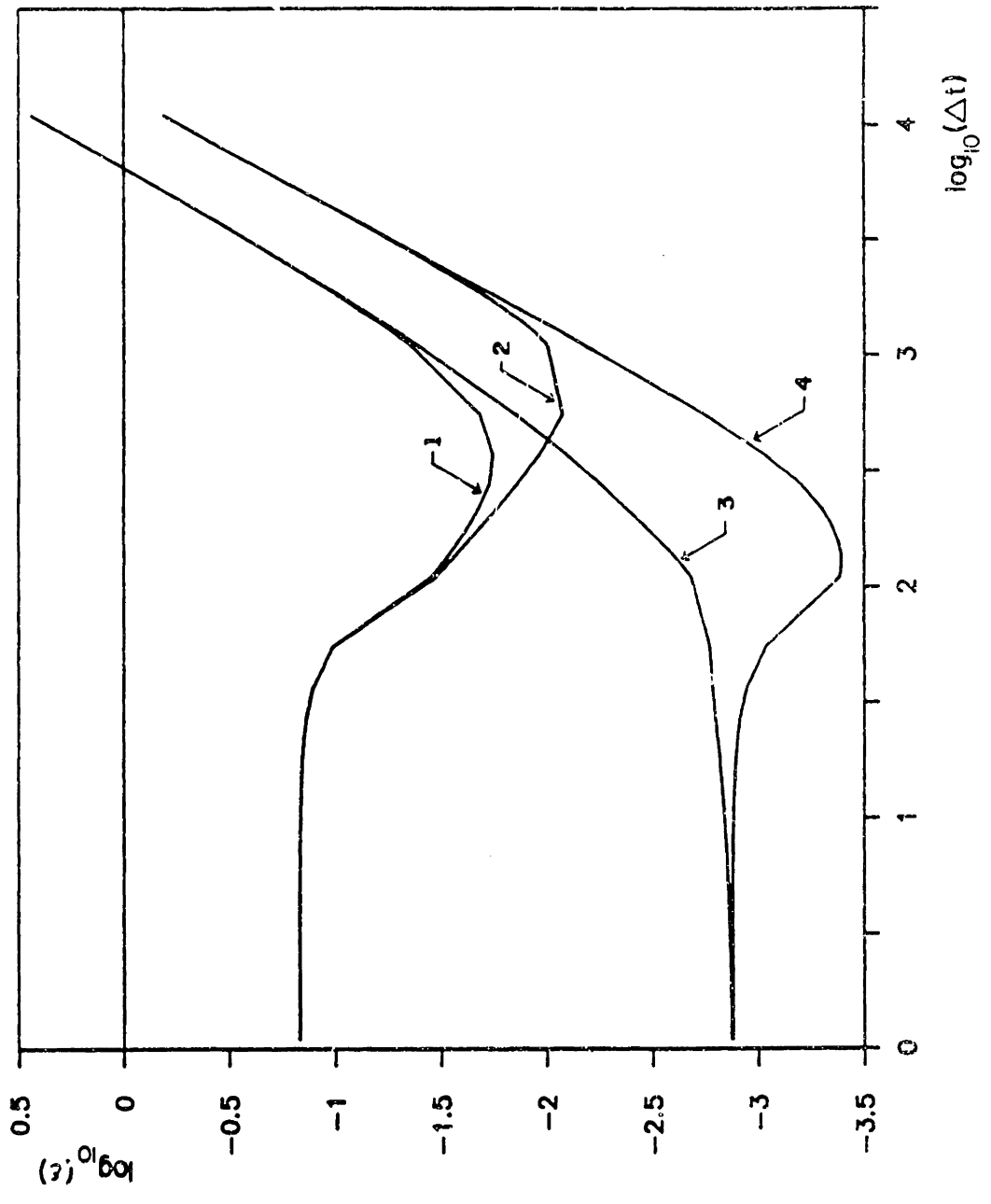
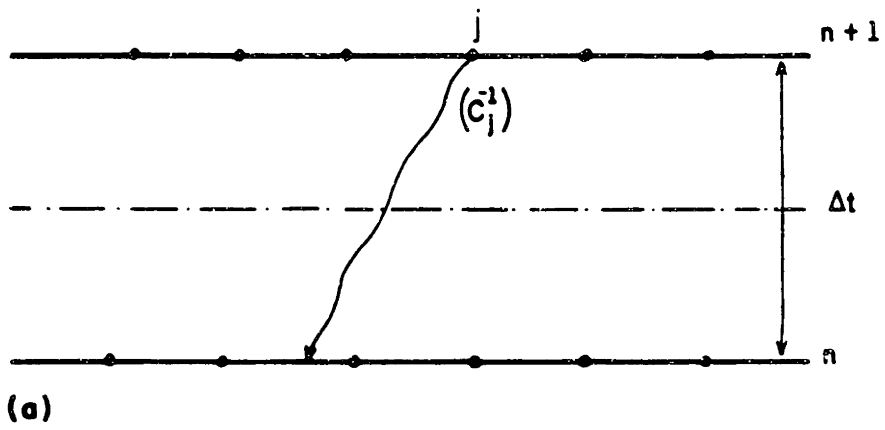


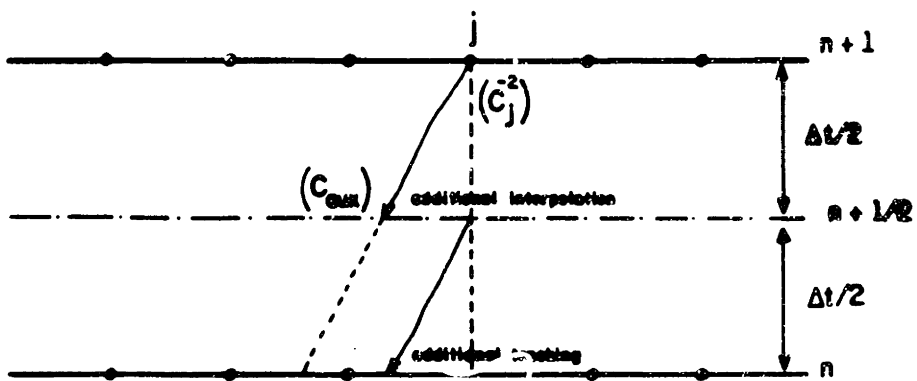
Figure 10

Illustration of the strategy for a third-order time-discretization scheme

- Step 1: Use Crank-Nicholson with time-step Δt to find \bar{c}_j^1 --see (a)
- Step 2: Use Crank-Nicholson with time-step Δt to find \bar{c}_j^2 (C_{aux} must be found, which involve additional tracking, interpolation, and solution of the set of linear algebraic equations associated with spatial discretization of the dispersion operator)--see (b)
- Step 3: Use a Richardson extrapolation to improve the estimate of c_j (from \bar{c}_j^1 and \bar{c}_j^2)--not illustrated



(a)



(b)

Table 1

Truncation Errors

Computational strategy	Quadratic interpolator: Euler (implicit) discretization	Quartic interpolator: Euler (implicit) discretization	Quadratic interpolator: Crank-Nicholson discretization	Quartic interpolator: Crank-Nicholson discretization
int	$\alpha(\alpha^2-1)\frac{\Delta x^2}{6\Delta t}\frac{\partial^2 c}{\partial x^2}$	$-\alpha(\alpha^2-1)(\alpha^2-4)\frac{\Delta x^5}{120\Delta t}\frac{\partial^5 c}{\partial x^5}$	$\alpha(\alpha^2-1)\frac{\Delta x^3}{6\Delta t}\frac{\partial^3 c}{\partial x^3}$	$-\alpha(\alpha^2-1)(\alpha^2-4)\frac{\Delta}{12}\frac{\partial^5 c}{\partial x^5}$
ad lif	$\frac{D^2\Delta t}{2}\frac{\partial^4 c}{\partial x^4}$			$\frac{D^3\Delta t^2}{12}\frac{\partial^6 c}{\partial x^6}$
ad lif	$-\frac{D\Delta t}{2}\left[\frac{\partial^2 u}{\partial x^2}\frac{\partial c}{\partial x} + 2\frac{\partial u}{\partial x}\frac{\partial^2 c}{\partial x^2}\right]$			$-\frac{D^2\Delta t^2}{12}\left[K_1\frac{\partial c}{\partial x} + K_2\frac{\partial^2 c}{\partial x^2} + K_3\frac{\partial^3 c}{\partial x^3} + K_4\frac{\partial^4 c}{\partial x^4}\right]$ with $K_1 = \frac{\partial^3 u}{\partial x^3\partial t} + 3\frac{\partial u}{\partial x}\frac{\partial^2 u}{\partial x^2} + u\frac{\partial^3 u}{\partial x^3} + D\frac{\partial^4 u}{\partial x^4}$ $K_2 = 2\frac{\partial^2 u}{\partial x^2\partial t} + u\frac{\partial^2 u}{\partial x^2} + 3\frac{\partial u}{\partial x}\frac{\partial u}{\partial x} + 4D\frac{\partial^3 u}{\partial x^3}$ $K_3 = -u\frac{\partial u}{\partial x} + 7D\frac{\partial^2 u}{\partial x^2}$ $K_4 = 6D\frac{\partial u}{\partial x}$
ad lif				$\frac{D\Delta x^2}{12}\frac{\partial^4 c}{\partial x^4}$

Table 2

Relative Importance of Different Types of Errors
(for a quadratic interpolator and an Euler implicit time-discretization)

Error type	Dimensionless controlling parameters	Condition for the dominance of the first type of error over the second
ϵ_{int}^{sd} , ϵ_{dif}^{sd}	$Cu > \alpha: Di = \frac{D\Delta t}{\Delta x^2}$	$Di < 2\alpha(\alpha^2-1) \left \frac{c^{(3)}}{\Delta x c^{(4)}} \right $
	$Cu = \alpha: Pe = \frac{u\Delta x}{D}$	$Pe > \frac{1}{2 1-Cu } \left \frac{\Delta x \cdot c^{(4)}}{c^{(3)}} \right $
	$Cu = \frac{u\Delta t}{\Delta x}$	
ϵ_{dif}^{tdm} , ϵ_{dif}^{sd}	Di	$Di > \frac{1}{6}$
ϵ_{dif}^{tdf} , ϵ_{dif}^{sd}	$Cu_f = \frac{Au\Delta t}{\Delta x}$	$Cu_f > \frac{1}{6} \frac{\Delta x \cdot Au \cdot c^{(4)}}{u^{(2)}c^{(1)} + 2u^{(1)}c^{(2)}}$
ϵ_{jut}^{tdm} , ϵ_{dif}^{tdm}	$Cu > \alpha: Di$	$Di^2 < \frac{ \alpha(\alpha^2-1) }{3} \left \frac{c^{(3)}}{\Delta x c^{(4)}} \right $
	$Cu = \alpha: Pe, Cu, Di$	$Pe > \frac{1}{3 \left \frac{1}{Di} - Cu \cdot Pe \right } \left \frac{\Delta x \cdot c^{(4)}}{c^{(3)}} \right $
ϵ_{dif}^{tdf} , ϵ_{dif}^{tdm}	$Pe_f = \frac{Au\Delta x}{D}$	$Pe_f > \left \frac{Au\Delta x c^{(4)}}{u^{(2)}c^{(1)} + u^{(1)}c^{(2)}} \right $
ϵ_{int}^{tdf} , ϵ_{dif}^{tdf}	$Cu > \alpha: Cu_f, Di$	$Di \cdot Cu_f < \frac{ \alpha(\alpha^2-1) }{3} \left \frac{Au c^{(3)}}{u^{(2)}c^{(1)} + 2u^{(1)}c^{(2)}} \right $
	$Cu = \alpha: Pe, Cu, Cu_f$	$Pe > \frac{3Cu_f}{ 1-Cu^3 } \left \frac{u^{(2)}c^{(1)} + 2u^{(1)}c^{(2)}}{uc^{(4)}} \right $

note: $v^{(n)} = \frac{\partial^n v}{\partial x^n}$ with $v = c$ or $v = u$

APPENDIX A

Derivation of truncation errors

We consider first the general time-discretization scheme, characterized by Equation (3). Expanding individual terms around the foot of the characteristic line, ξ , we obtain

$$c^{n+1} = c^{\xi} + \Delta t \left. \frac{Dc}{Dt} \right|_{\xi} + \frac{\Delta t^2}{2!} \left. \frac{D^2c}{Dt^2} \right|_{\xi} + \frac{\Delta t^3}{3!} \left. \frac{D^3c}{Dt^3} \right|_{\xi} + \text{H.O.D.} \quad (\text{A1})$$

$$L^{n+1} = L^{\xi} + \Delta t \left. \frac{DL}{Dt} \right|_{\xi} + \frac{\Delta t^2}{2!} \left. \frac{D^2L}{Dt^2} \right|_{\xi} + \frac{\Delta t^3}{3!} \left. \frac{D^3L}{Dt^3} \right|_{\xi} + \text{H.O.D.} \quad (\text{A2})$$

which can be combined so as to obtain the local equilibrium

$$\begin{aligned} \left. \frac{Dc}{Dt} \right|_{\xi} = L \left|_{\xi} - \Delta t \left\{ \frac{1}{2} \left(\frac{D^2c}{Dt^2} - \frac{DL}{Dt} \right) + \left(\frac{1}{2} - A \right) \frac{DL}{Dt} \right\} - \\ - \Delta t^2 \left\{ \frac{1}{3!} \left(\frac{D^3c}{Dt^3} - \frac{D^2c}{Dt^2} \right) + \left(\frac{1}{6} - \frac{A}{2} \right) \frac{D^2L}{Dt^2} \right\} + \text{H.O.D.} \end{aligned} \quad (\text{A3})$$

Now, from [B7],

$$\left. \frac{Dc}{Dt} \right|_{\xi} = \left. \frac{Dc}{Dt} \right|_n - \epsilon_{int} \quad (\text{A4})$$

with

$$\epsilon_{int} = \begin{cases} \alpha \cdot (\alpha^2 - 1) \cdot \frac{\Delta x^3}{6\Delta t} \frac{\partial^3 c}{\partial x^3} + \text{H.O.D.} & \text{Quadratic} \\ & \text{interpolator} \\ -\alpha \cdot (\alpha^2 - 1) \cdot (\alpha^2 - 4) \frac{\Delta x^5}{120\Delta t} \frac{\partial^5 c}{\partial x^5} + \text{H.O.D.} & \text{Quartic} \\ & \text{interpolator} \end{cases} \quad (\text{A5})$$

Similarly,

$$L^{\xi} = L^n + \epsilon_{L,int} \quad (\text{A6})$$

Hence, the local equilibrium at ξ expressed by Equation * can be rewritten at (j,n) as

$$\frac{Dc}{Dt} = L + \epsilon_{int} + \epsilon_{L,int} + \begin{cases} \frac{\Delta t}{2} \frac{D}{Dt} (L + \epsilon_{int} + \epsilon_{L,int}) + \text{H.O.D.} & \text{Euler} \\ \frac{\Delta t^2}{12} \frac{D^2}{Dt^2} (L + \epsilon_{int} + \epsilon_{L,int}) + \text{H.O.D.} & \text{Crank-} \\ & \text{Nicholson} \end{cases} \quad (\text{A7})$$

i.e.,

$$\frac{Dc}{Dt} = L + \epsilon_{int} + \epsilon_{dif}^{td} + \text{H.O.D.} \quad (\text{A8})$$

with

$$\epsilon_{dif}^{td} = \begin{cases} \frac{\Delta t}{2} \frac{DL}{Dt} + \text{H.O.D.} & \text{Euler} \\ \frac{\Delta t^2}{12} \frac{D^2 L}{Dt^2} + \text{H.O.D.} & \text{Crank-Nicholson} \end{cases} \quad (\text{A9})$$

We now express the time-derivatives of L as a function of space derivatives of L and c , using the form of the governing equation for transport to assist in the transformation, but making no assumption on the

characteristics of the flow field (i.e., a general non-uniform, unsteady flow is considered). This leads to

$$\epsilon_{dif}^{td} = \epsilon_{dif}^{tdm} + \epsilon_{dif}^{tdf} + \text{H.O.D.} = \begin{cases} \frac{D\Delta t}{2} \frac{\partial^2 L}{\partial x^2} - \\ \frac{D^2 \Delta t^2}{12} \frac{\partial^2 L}{\partial x^2} - \end{cases}$$

$$\begin{cases} - \frac{D\Delta t}{2} \left[u_{xx} \frac{\partial c}{\partial x} + 2u_x \frac{\partial^2 c}{\partial x^2} \right] + \text{H.O.D.} & \text{Euler} \\ - \frac{D^2 \Delta t^2}{12} \left[k_1 \frac{\partial c}{\partial x} + k_2 \frac{\partial^2 c}{\partial x^2} + k_3 \frac{\partial^3 c}{\partial x^3} + k_4 \frac{\partial^4 c}{\partial x^4} \right] + \text{H.O.D.} & \text{Crank-} \\ & \text{Nicholson} \end{cases} \quad (\text{A10})$$

where k_1, \dots, k_4 represent products of concentration derivatives by velocities or velocity derivatives (see Table 1 for definitions). We note that ϵ_{dif}^{tdm} represents the result of the above derivation for the case of uniform u , and ϵ_{dif}^{tdf} represents a correction due to non-uniformities.

Now, we consider the spatial discretization of the operator L (Equation 4 in text), to obtain

$$L = D \frac{\partial^2 c}{\partial x^2} = L + \epsilon_{dif}^{sd} = L + \frac{D\Delta x^2}{12} \frac{\partial^4 c}{\partial x^4} + \text{H.O.D.} \quad (\text{A11})$$

Finally, collecting the contribution of the different errors, we obtain Equation 5 in the text.