DEPARTMENT OF MATHEMATICS

# ON TIME STEP SELECTION FOR SOLVING 1D NONLINEAR DIFFUSION EQUATIONS

K. Chen, M. J. Baines and P. K. Sweby

---

Numerical Analysis Report **10/91**

Whiteknights, P. O. Box 220,

Reading RG6 2AX, Berkshire.

UNIVERSITY OF READING

**Abstract**

A useful time step selection strategy is proposed for time stepping schemes of three-level type, in the numerical solution of nonlinear diffusion equations for semiconductor process modelling. The procedure proposed requires the availability of local truncation error estimates. Here it is first applied with the finite difference methods, where error estimates are easily available, and then with a finite element method of Petrov-Galerkin type. In the latter case, we find the local truncation error estimates by using the property of the inverse of diagonally dominant matices. Numerical experiments on a nonlinear semiconductor diffusion model have been carried out and results show that our time step selection strategy is efficient and robust.

**Key words :**

Automatic time step selection, Finite difference methods (FDM's), Petrov-Galerkin methods (FEM's), Predictor corrector schemes, B-splines, Nonlinear parabolic PDE's, Diffusion models.

# Contents

# 1 Introduction

The numerical solution of a nonlinear system of parabolic partial differential equations (PDE's) lies at the centre of semiconductor process modelling simulation problems. Our study here is mainly concerned with the temporal discretization of such systems.

In a typical semiconductor process model, several dopants are present in a silicon base, each with its own initial concentration profile. The purpose of a diffusion simulation is to accurately predict these concentration profiles after a certain time interval TI under a given temperature TE. For each dopant, the redistribution through the diffusion mechanism may be characterized by the two-dimensional nonlinear partial differential equation

$$\frac{\partial C}{\partial t} = \mathrm{Div}\,[D(C)\mathrm{grad}C + ZC\mathrm{grad}\Phi] \tag{1}$$

where $C = C(x, y, t)$ is the dopant concentration, $D$ is the diffusion coefficient, $\Phi = \Phi(C)$ is the electrostatic potential and $Z = \pm 1$ depends on the dopant used. Essentially, the quantities $D$ and $\Phi$ are in general some known nonlinear functions in $C$; refer to Selberherr [30] and Sze [34].

There exist many useful 2D semiconductor process simulators which solve systems of nonlinear PDE's of form (1), such as those of Penumalli [27] and Lorenz et al [21] using finite difference methods and Borucki [6] and Gerodolle [13] using finite element methods, to name just a few. The numerical methods used in most of these simulators for spatial discretizations have been largely conventional in the sense that information on them may be found in the standard literature e.g. Mitchell & Griffiths [24], Mitchell & Wait [25]. To deal with more complicated silicon base structures, it is generally believed to be better to use the flexible finite element approach. However, as we know, standard finite element approximations of the Galerkin type are often oscillatory for equations of non-self-adjoint form (eg. for equation (1) after log transformation of $C$). One remedy is to use a Petrov-Galerkin method instead; refer to Barrett & Morton [5] and Scotney [31] and also the so-called *subdomain methods* of Burnett [4] and Finlayson [12].

In Svoboda [33], a new variant of the Petrov-Galerkin approach is introduced, called the ASWR method, which seeks to solve (1) using quadratic B-splines as trial basis functions and piecewise constants as test functions. The method has since been built into the COMPOSITE code of Lorenz [21] for process modelling; see Lorenz & Svoboda [22]. With the spatial discretization method well established, there remain many choices for time discretization schemes. As is well known, using stable fully implicit time stepping schemes would normally generate a nonlinear system of algebraic equations and it is often difficult and expensive to solve such systems, if not impossible. Refer to Gresho et al [14] and Polak et al [29] for work done in this approach.

Here we shall consider three level semi-implicit schemes, effectively linearizing any nonlinear equations which would be generated otherwise.

To facilitate the discussion, we first consider the one dimensional case. Results on 2D will be reported separately, although most of the analysis will follow from 1D studies. The 1D equation corresponding to (1) may be written as

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left(D(C)\frac{\partial C}{\partial x}\right) \tag{2}$$

where we have dropped the potential term in this context without essential loss of generality. Equations of the form (2) have been studied by King et al [17] and Lacey et al [18] and solved approximately by using standard finite element methods in Ho et al [15] and Douglas & Dupont [11] and by moving finite element methods in Baines et al [2], [3] and [16]. Here we shall solve transformed forms of (2) by both finite difference methods and finite element methods of the ASWR type.

## 2 A scaling logarithmic transformation

The problem of appropriately scaling the main variables is very necessary for semiconductor equations. In Moody & Please [26], suitable non-dimensionalisations are suggested for such equations. However, as is well known, it remains a problem that the dopant level of main interest is many orders of magnitude below the initial dopant level. High

order polynomial approximations to the concentration $C$ often show inevitably oscillatory behaviour.

One way to avoid this problem is to transform the concentration variable into something more amenable to numerical computation, e.g. using the log transformation $f = \log(C)$. With the new variable $f$, accurate approximations are naturally possible with even low order polynomials because $f$ becomes a polynomial function in simple cases (say when $D$ is constant). This particular transformation

$$f = \log(C) \tag{3}$$

has been used in Lorenz & Svoboda [22] and Svoboda [33]. With (3), equation (2) is transformed into

$$\frac{\partial f}{\partial t} = \frac{\partial}{\partial x}\left(D(e^f)\frac{\partial f}{\partial x}\right) + D(e^f)\left(\frac{\partial f}{\partial x}\right)^2 \tag{4}$$

To discuss time stepping schemes for (4), it is convenient to also study the following linear model equation

$$\frac{\partial f}{\partial t} = D\frac{\partial^2 f}{\partial x^2} + V\frac{\partial f}{\partial x} \tag{5}$$

where $D$, $V$ are taken to be constants with $D > 0$. For solving (2), other transformations may be considered; see Please & Sweby [28]. But in our context here, (3) has been found to be very appropriate.

# 3 The spatial discretization

## 3.1 Petrov Galerkin methods

We now introduce the spatial discretization method for solving equation (4). Assume that the equation is defined in [a,b]. Let us place a mesh $\{x_j\}_0^N$ over [a,b] such that

$$\Pi_N : \qquad a = x_0 < x_1 < \cdots < x_N = b \tag{6}$$

Denote by $\xi_j = (x_{j-1} + x_j)/2$ and $h_j = x_j - x_{j-1}$ the mid-points and the local mesh sizes respectively. Define by $S_1$ and $S_2$ two spaces of piecewise polynomials of order $k_1$ and $k_2$ respectively over each interval of $\Pi_N$.

Then for a general operator equation

$$\frac{\partial f}{\partial t} - \mathcal{L}f = r \qquad\qquad x \in [a, b] \tag{7}$$

where $\mathcal{L}$ is an elliptic operator, the Petrov Galerkin method seeks the approximate solution $f_N$ to $f$ in space $S_2$ by insisting on

$$\left\langle \frac{\partial f_N}{\partial t} - \mathcal{L}f_N - r,\ w \right\rangle = 0 \qquad\qquad \forall w \in S_1 \tag{8}$$

where the inner product is defined by $< u, v >= \int_a^b uvdx$. Conventionally the space $S_1$ is called the **test** space and $S_2$ the **trial** space. If the two spaces are identical, the finite element method is the standard Galerkin method. Refer to Barrett & Morton [5] and Scotney [31] for more theoretical studies.

## 3.2 Test and trial spaces of the ASWR method

The two spaces $S_1$ and $S_2$ as defined in §3.1 may be specified by various choices of integers $k_1$ and $k_2$, the order of the polynomial bases involved. Here we shall consider the particular combination of $k_1 = 1$ and $k_2 = 3$ corresponding to spaces of piecewise polynomials of degrees 0 and 2 respectively. This choice was first made in Svoboda [33], motivated by the simplicity of $S_1$ and, in the case of $S_2$, the need to differentiate $C$ twice and the assumption that the solution of (4) behaves like a quadratic function (which is correct in the case of $D$ being constant everywhere in the domain). Note that Petrov-Galerkin methods with the choice $k_1 = 1$ may be referred to as subdomain methods; refer to Burnett [4] and Finlayson [12].

Thus the **test** space $S_1$ may be represented by

$$\text{span}\{\phi_j \mid j = 1, \cdots, N\} \tag{9}$$

where

$$\phi_j(x) = \begin{cases} 1, & x \in (x_{j-1}, x_j], \\ 0, & \text{otherwise.} \end{cases}$$

The specification of the **trial** space follows from De Boor [10, Ch.9]

$$S_2 = \text{span}\{\psi_j \mid j = 1, \cdots, N+2\} \tag{10}$$

$$\psi_j = \begin{cases} (x - x_0)^{j-1}, & \text{for } j = 1, 2, 3, \\ (x - x_{j-3})_+^2, & \text{for } j = 4, \cdots, N+2 \end{cases}$$

where the truncated function is defined by $y_+ = \max\{y, 0\}$ and functions in $S_2$ are assumed to be right continuous in each interval $(x_{j-1}, x_j]$. For a general nonuniform mesh $\Pi_n$, the basis functions can be bad because some of the $\psi_j$'s may be nearly linearly dependent; refer to De Boor [10, Ch.9] for further discussion.

A more stable set of basis functions for $S_2$ can be provided by B-splines. To construct B-spline basis functions representing $S_2$ of (10), we define a set of knot sequence $T_N = \{t_j\}_0^{N+4}$ such that

$$t_0 = t_1 = t_2 \equiv x_0 < t_3 \equiv x_1 < \cdots < x_{N-1} \equiv t_{N+1} < x_N \equiv t_{N+2} = t_{N+3} = t_{N+4} \tag{11}$$

Based on $T_N$, we can construct $n+2$ pieces of B-spline functions of order 3 (refer to De Boor [10, Ch.10])

$$\begin{aligned} B_i &= (t_{i+3} - t_i) G[t_i, t_{i+1}, t_{i+2}, t_{i+3}] \\ &= \begin{cases} \dfrac{(x - t_i)^2}{(t_{i+2} - t_i)(t_{i+1} - t_i)} & \text{for } x \in [t_i, t_{i+1}] \\ \dfrac{(x - t_i)(t_{i+2} - x)}{(t_{i+2} - t_i)(t_{i+2} - t_{i+1})} + \dfrac{(t_{i+3} - x)(x - t_{i+1})}{(t_{i+3} - t_{i+1})(t_{i+2} - t_{i+1})} & \text{for } x \in (t_{i+1}, t_{i+2}] \\ \dfrac{(t_{i+3} - x)^2}{(t_{i+3} - t_{i+2})(t_{i+3} - t_{i+1})} & \text{for } x \in (t_{i+2}, t_{i+3}] \end{cases} \end{aligned} \tag{12}$$

where the function $G(t) = (t - x)_+^2$, $G[\cdot, \cdot, \cdot, \cdot]$ represents the third order divided difference of function $G$ and $i = 0, 1, \cdots, N+1$. Now with the special choice of $T_N$ in (11), the theorem of Curry & Schoenberg [9] ensures that the B-splines $B_i$'s defined in (12) form a stable basis for the **trial** space $S_2$, *i.e.*

$$S_2 = \text{span}\{B_j \mid j = 0, \cdots, N+1\} \tag{13}$$

Note that formulae given in both Lorenz & Svoboda [22] and Svoboda [33] for $B_i$ have misprints.

## 3.3 The ASWR solution

With the test and trial spaces $S_1$ and $S_2$ specified as in §3.2, the ASWR method for the operator equation (7) follows immediately from (8). In particular the ASWR solution $f_N \in S_2$ can be written as

$$f_N = \sum_{i=0}^{N+1} \alpha_i(t) B_i(x) \tag{14}$$

where, for equation (4), the coefficients $\alpha_i$ may be determined by satisfying

$$\left\langle \frac{\partial f_N}{\partial t} - \frac{\partial}{\partial x}\left( D(e^{f_N}) \frac{\partial f_N}{\partial x} \right) - D(e^{f_N}) \left( \frac{\partial f_N}{\partial x} \right)^2 , \phi_j \right\rangle = 0 \qquad j = 1, 2, \cdots, N \tag{15}$$

*i.e.* for $j = 1, 2, \cdots, N$

$$\begin{aligned} \sum_{i=j-1}^{j+1} \left[ \frac{\partial \alpha_i(t)}{\partial t} \int_{t_{j-1}}^{t_j} B_i(x) dx \right. &- \left. \alpha_i(t) \int_{t_{j-1}}^{t_j} \frac{\partial}{\partial x}\left( D(e^{f_N}) \frac{\partial B_i}{\partial x} \right) dx \right] \\ &- \int_{t_{j-1}}^{t_j} D(e^{f_N}) \left( \sum_{i=j-1}^{j+1} \alpha_i(t) \frac{\partial B_i}{\partial x} \right)^2 dx = 0 \end{aligned} \tag{16}$$

where we have used the piecewise properties of B-splines. For the linear problem (5), the above system takes a simpler form

$$\sum_{i=j-1}^{j+1} \left[ \frac{\partial \alpha_i(t)}{\partial t} \int_{t_{j-1}}^{t_j} B_i(x) dx - D\alpha_i(t) \int_{t_{j-1}}^{t_j} \frac{\partial^2 B_i}{\partial x^2} dx - V\alpha_i(t) \int_{t_{j-1}}^{t_j} \frac{\partial B_i}{\partial x} dx \right] = 0 \tag{17}$$

where $j = 1, 2, \cdots, N$. In the above systems, (16) and (17), there are $(N+2)$ unknowns in $N$ equations. For either system, two more equations can be provided from satisfying the boundary conditions at $a$ and $b$. Let us assume that our problem is imposed with the following condition

$$\beta f + \gamma \frac{\partial f}{\partial x} = g \qquad \text{at } a,\ b \tag{18}$$

where $g = g(x)$ is some known function and $\beta$, $\gamma$ are assumed to be known parameters with $\beta\gamma \neq 0$. Then for the ASWR solution (14) to satisfy (18), we need to impose

$$\begin{cases} \sum_{i=0}^{1} \alpha_i(t) \left[ \beta B_i(a) + \gamma \frac{\partial B_i}{\partial x}(a) \right] = g(a) \\ \sum_{i=N}^{N+1} \alpha_i(t) \left[ \beta B_i(b) + \gamma \frac{\partial B_i}{\partial x}(b) \right] = g(b) \end{cases} \tag{19}$$

Thus the semi-discrete ASWR solution to equation (4) with (18) is obtained by solving (16) and (19); while the solution to (5) with (18) is given by solving (17) and (19).

To see the close relationship of ASWR method with the standard Galerkin method, we have taken $h_j \equiv h$ and found that, for the linear problem (5) with the Neumann boundary conditions ($\beta = g = 0$ and $\gamma = 1$), the system corresponding to (17) and (19) can be written as

$$R\frac{d\alpha}{dt} = T\alpha + S\alpha \tag{20}$$

where $\alpha = [\alpha_0(t), \alpha_1(t), \cdots, \alpha_{N+1}(t)]^T$,

$$R = \frac{h}{6}\begin{bmatrix} 5 & 1 & & & & \\ 1 & 4 & 1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & 1 & 4 & 1 & \\ & & & 1 & 5 & \end{bmatrix}, \qquad T = \frac{D}{h}\begin{bmatrix} -1 & 1 & & & & \\ 1 & -2 & 1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & 1 & -2 & 1 & \\ & & & 1 & -1 & \end{bmatrix},$$

$$S = \frac{V}{2}\begin{bmatrix} -1 & 1 & & & \\ -1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 0 & 1 \\ & & & -1 & 1 \end{bmatrix}$$

are all $(N+2)\times(N+2)$ tridiagonal matrices. It can be shown that the matrices $R$, $T$ and $S$ are identical to those obtained from applying the Galerkin method using piecewise linear elements. However the two finite element solutions will be different since one is piecewise linear, the other piecewise quadratic.

# 4 Predictor corrector schemes

## 4.1 Introduction

As discussed in §1, semi-implicit time stepping schemes are easy to implement for solving a nonlinear parabolic PDE. For equation (4), we have analyzed in Chen [7] a class of three-level semi-implicit schemes including those of the following form :

$$\frac{1}{\Delta t}\sum_{\ell=0}^{2}\omega_\ell f_n^{j+\ell} = \frac{D(e^{f_n^{j+1}})}{\Delta x^2}\delta^2\left(\sum_{\ell=0}^{2}\nu_\ell f_n^{j+\ell}\right) + D(e^{f_n^{j+1}})\left(\frac{\bar{\delta} f_n^{j+1}}{2\Delta x}\right)^2 \tag{21}$$

where $f_n^\tau = f(x_n, t_\tau)$, $\sum_{\ell=0}^{2} \omega_\ell = 0$, $\sum_{\ell=0}^{2} \nu_\ell = 1$ and $\bar{\delta}$ and $\delta^2$ denote the first and second order difference operator respectively :

$$\bar{\delta} f_n^j = f_{n+1}^j - f_{n-1}^j \qquad \text{and} \qquad \delta^2 f_n^j = f_{n+1}^j - 2f_n^j + f_{n-1}^j$$

From linear stability analysis and subsequent experiments, it was noted that some simple schemes do have stability restrictions on $\Delta t$ depending on $\Delta x$, such as the Euler scheme (with $\omega_0 = -\omega_2 = -1$, $\omega_1 = 0$ and $\nu_0 = \nu_1 = 0$, $\nu_2 = 1$) and the Crank-Nicolson scheme (with $\omega_0 = -\omega_2 = -1$, $\omega_1 = 0$ and $\nu_0 = \nu_2 = 1/2$, $\nu_1 = 0$). These schemes are therefore not ideal if we are to use very fine meshes.

Another problem associated with three-level (or more generally linear multi-level) schemes is that values involved at the intermediate time levels are also not known and therefore must be somehow provided. This can be easily accomplished in certain cases, *e.g.* if uniform time stepping in $\Delta t$ is permitted. In a more general situation, these intermediate values may be obtained by using explicit methods or methods involving fewer time levels, leading to the so-called predictor-corrector (PC) type schemes.

Predictor-corrector schemes have been extensively studied for solving ordinary differential equations; refer to Lambert [19]. Limited applications to parabolic PDE's can be found in the literature; e.g. see Ames [1] and Douglas & Dupont [11], where the Crank-Nicolson scheme based method was investigated. This kind of work has been focused on solving self-adjoint PDE's of form (1) or (2). See also Strehmel et al [32].

Here we shall study an Euler scheme based PC method throughout this section for simplicity, although other schemes may be analyzed in a similar manner. But we shall investigate the non-self-adjoint PDE (4). Following (21), we can obtain the following PC scheme :

$$\begin{cases} \dfrac{f_n^{j+\theta} - f_n^j}{\theta \Delta t} = \dfrac{D(e^{f_n^j})}{\Delta x^2} \delta^2 f_n^{j+\theta} + D(e^{f_n^j}) \left( \dfrac{\bar{\delta} f_n^j}{2\Delta x} \right)^2 \\ \dfrac{f_n^{j+2} - f_n^j}{2\Delta t} = \dfrac{D(e^{f_n^{j+\theta}})}{\Delta x^2} \delta^2 f_n^{j+2} + D(e^{f_n^{j+\theta}}) \left( \dfrac{\bar{\delta} f_n^{j+\theta}}{2\Delta x} \right)^2 \end{cases} \tag{22}$$

where we have allowed $\theta \neq 1$ for generality.

## 4.2 Linear stability analysis

We now proceed to analyze the stability of (22). For this purpose, we apply (22) to equation (5) and obtain the following

$$\begin{cases} (1-\theta\sigma\delta^2)f_n^{j+\theta} = (1+\theta\omega\bar{\delta})f_n^j \\ (1-2\sigma\delta^2)f_n^{j+2} = 2\omega\bar{\delta}f_n^{j+\theta} + f_n^j \end{cases} \tag{23}$$

where $\sigma = D\Delta t/\Delta x^2$ and $\omega = V\Delta t/(2\Delta x)$. Then we can show that

**THEOREM 1** *(stability) Provided that* $\Delta t \le \frac{D}{V^2}$ *and* $\theta > 0$, *the scheme (22) for (5), i.e. (23) is stable.*

**Proof.** Using the von Neumann analysis, we can obtain the amplification factor (by comparing errors at $t = t_j$ and $t = t_{j+2}$) as

$$\mu = \text{RHS/LHS} \tag{24}$$

where $\text{LHS} = 1 + 8\sigma\sin^2\frac{\xi}{2}$, $\text{RHS} = 1 + \frac{4\omega i\sin\xi - 8\theta\omega^2\sin^2\xi}{1+4\theta\sigma\sin^2\frac{\xi}{2}}$ and $i = \sqrt{-1}$. To find stability conditions, we require $|\mu| < 1$ for all $\Delta t$ with $\Delta x$ fixed, leading to the requirement

$$\begin{aligned} &\left[1 + 4\theta\sigma\sin^2\frac{\xi}{2} - 8\theta\omega^2\sin^2\xi\right]^2 + 16\omega^2\sin^2\xi \\ < &\left[1 + 4\sigma(2+\theta)\sin^2\frac{\xi}{2} + 32\theta\sigma^2\sin^4\frac{\xi}{2}\right]^2 \end{aligned} \tag{25}$$

We may simplify the inequality to

$$4\theta^2\omega^2\sin^2\xi(\sin^2\xi - \sigma\sin^2\frac{\xi}{2}) < \omega^2(\theta-1)\sin^2\xi + \sigma\sin^2\frac{\xi}{2} + R \tag{26}$$

where $R = 64\theta^2\sigma^4\sin^8\frac{\xi}{2} + 16\theta\sigma^3(2+\theta)\sin^6\frac{\xi}{2} + 4\sigma^2(1+2\theta)\sin^4\frac{\xi}{2} \ge 0$ for $\theta \ge 0$. Let $s = \sin^2\frac{\xi}{2}$. Then for $\xi \ne 0,\ \pi$, we get the following sufficient condition for (26) to hold

$$\theta^2(\omega^2 - \sigma)s - \theta^2\omega^2 s^2 < \frac{\sigma}{16\omega^2} + \frac{\theta-1}{4} \tag{27}$$

Now observe that $\omega^2 - \sigma = \frac{\Delta t}{\Delta x^2}\left(\frac{V^2}{4}\Delta t - D\right) \le 0$ because $\Delta t \le \frac{4D}{V^2}$. On the other hand, the right hand side of (27) is non-negative for $\theta > 0$ since $(1-\theta)\Delta t < \Delta t \le D/V^2$. Thus (27) holds and hence the scheme (22) is stable. □

Note that Theorem 1 states that the stability restriction of our predictor-corrector scheme, for a time step $\Delta t$, is actually independent of the spatial discretization; while the stability restriction of the Euler based three-level scheme is known to be dependent on $\Delta x$. This suggests that our predictor-corrector scheme should be more favourable.

## 4.3 An error analysis

Let us again consider the linear model equation (5) applied with the predictor corrector (PC) scheme (22). Denote by $F_n^j = F(x_n, t_j)$ the exact solution of (5). Then we can define the local truncation errors (LTE's) and global errors (GE's) respectively by

**LTE's for predictor and corrector steps :**

$$\begin{cases} \tau_p = \dfrac{F_n^{j+\theta} - F_n^j}{\theta \Delta t} - \dfrac{D}{\Delta x^2}\delta^2 F_n^{j+\theta} - \dfrac{V}{2\Delta x}\bar{\delta} F_n^j \\ \tau_c = \dfrac{F_n^{j+2} - F_n^j}{2\Delta t} - \dfrac{D}{\Delta x^2}\delta^2 F_n^{j+2} - \dfrac{V}{2\Delta x}\bar{\delta} F_n^{j+\theta} \end{cases} \tag{28}$$

**GE's for predictor and corrector steps :**

$$\begin{cases} e_p = F_n^{j+\theta} - f_n^{j+\theta} \\ e_c = F_n^{j+2} - f_n^{j+2} \end{cases} \tag{29}$$

Following (23) and (28), we have

$$\begin{cases} (1 - \theta\sigma\delta^2) F_n^{j+\theta} = (1 + \theta\omega\bar{\delta}) F_n^j + \theta \Delta t \tau_p \\ (1 - 2\sigma\delta^2) F_n^{j+2} = 2\omega\bar{\delta} F_n^{j+2} + F_n^j + 2\Delta t \tau_c \end{cases} \tag{30}$$

On subtracting (23) from (30), we obtain the relationship between $\tau$'s and $e$'s

$$\begin{cases} (1 - \theta\sigma\delta^2) e_p = \theta \Delta t \tau_p \\ (1 - 2\sigma\delta^2) e_c = 2\omega\bar{\delta} e_p + 2\Delta t \tau_c \end{cases} \tag{31}$$

where we have assumed that $e_{\text{initial}} = F_n^j - f_n^j = 0$. The GE of one PC step is hence given by substituting the first equation of (31) into the second one

$$e_c = 2\theta\Delta t^2 (1 - 2\sigma\delta^2)^{-1} \frac{\bar{\delta}}{2\Delta x} (1 - \theta\sigma\delta^2)^{-1} \tau_p + 2\Delta t (1 - 2\sigma\delta^2)^{-1} \tau_c \tag{32}$$

To further simplify (32), we use Taylor expansions at $(x_n, t_j)$ in (28) to get

$$\begin{cases} \tau_p = R_p^{(0)} \Delta t + R_p^{(1)} \Delta t^2 + R_p^{(2)} \Delta x^2 \\ \tau_c = R_c^{(0)} \Delta t + R_c^{(1)} \Delta t^2 + R_c^{(2)} \Delta x^2 \end{cases} \tag{33}$$

where

$$\begin{cases} R_p^{(0)} = \frac{\theta}{2}\frac{\partial}{\partial t}\left(V\frac{\partial F}{\partial t} - D\frac{\partial^2 F}{\partial x^2}\right), & R_c^{(0)} = \frac{\partial}{\partial t}\left[(1-\theta)V\frac{\partial F}{\partial x} - D\frac{\partial^2 F}{\partial x^2}\right] \\ R_p^{(1)} = \frac{\theta^2}{2}\frac{\partial^2}{\partial t^2}\left(\frac{1}{3}\frac{\partial F}{\partial t} - D\frac{\partial^2 F}{\partial x^2}\right), & R_c^{(1)} = \frac{\partial^2}{\partial t^2}\left[(\frac{2}{3} - \frac{\theta^2}{2})V\frac{\partial F}{\partial x} - \frac{4D}{3}\frac{\partial^2 F}{\partial x^2}\right] \\ R_p^{(2)} = -\frac{1}{6}\frac{\partial^3}{\partial x^3}\left(\frac{D}{2}\frac{\partial F}{\partial x} + VF\right), & R_c^{(2)} = R_p^{(2)} \end{cases}$$

Now it is easy to see that the dominant term in (32) is the second one. Thus we have shown that the scheme (23) is locally of first order, globally of second order in time, and of second order in space.

## 4.4 Time stepping control

We shall introduce a time stepping strategy for automatic step size selection. To this end, we mainly use $\tau_c$, the dominant part of the underlying global error $e_c$ in (32), for error control.

### 4.4.1 Choice of the predictor step

We first discuss the choice of the predictor step, or equivalently the choice of $\theta$, in an attempt to minimize the error $\tau_c$. Ideally we should consider $\tau_c$ of (33) to be a function of $\theta$ and then proceed to solve the minimization problem under the constraint that $\theta > 0$. But this is not possible since the unknown quantity $\Delta t$ would be required. One way to make an *a priori* choice on $\theta$ is to minimize the coefficient $R_c^{(0)}$ instead. Because $R_c^{(0)}$ is also a discrete function of $x_n$, we shall now look for a least-squares solution to the minimization of $|R_c^{(0)}|$.

Let us rewrite $R_c^{(0)}$ as $a_n - \theta b_n$ at point $x_n$ where $a_n = V^2\frac{\partial^2 F}{\partial x^2} - D^2\frac{\partial^4 F}{\partial x^4}$ and $b_n = V^2\frac{\partial^2 F}{\partial x^2} + DV\frac{\partial^3 F}{\partial x^3}$. Construct a quadratic function of $\theta$ by

$$Q(\theta) = \sum_{n=1}^{N}(a_n - \theta b_n)^2 = \theta^2(\sum_{n=1}^{N} b_n^2) - 2\theta(\sum_{n=1}^{N} a_n b_n) + \sum_{n=1}^{N} a_n^2 \tag{34}$$

The solution of minimizing $Q(\theta)$ is generally given by

$$\theta^* = \sum_{n=1}^{N} a_n b_n \Big/ \sum_{n=1}^{N} b_n^2 \tag{35}$$

However we are only interested in positive $\theta^*$ which is only possible if $\sum_{n=1}^{N} a_n b_n > 0$. Otherwise we have to look for the smallest positive $\theta$ which minimizes $|R_n^{(0)}|$ pointwise or simply fix $\theta$. If the choice of fixed $\theta = 1$ is used, then the time stepping scheme resembles the more conventional three-level scheme as discussed in [7].

### 4.4.2 Choice of the corrector step $\Delta t$

At each time level $t = t_j$, the determination of $\theta$ in §4.4.1 requires the calculation of $R_n^{(0)}$ pointwise and this quantity represents the leading term in the local truncation error. Thus the following choice for $\Delta t$ can be made

$$\Delta t \leq \mathrm{TOL}/\max |R_n^{(0)}| \tag{36}$$

from a prescribed tolerance TOL. Such a choice for $\Delta t$ should of course be subject to the stability constraint (see Theorem 1)

$$\Delta t \leq D/V^2$$

In general, TOL must not be too large and should be comparable to $|R_n^{(2)}|\Delta x^2$.

We remark that that, since the GE's cannot be easily obtained *a priori*, the idea of applying the *Milne's device* to estimate high derivative terms by using the solutions $f_n^j$, $f_n^{j+\theta}$ is unfortunately not applicable; refer to Lambert [19] for the general idea and Gresho [14] for one such application.

# 5 Test examples

To carry out numerical experiments on our proposed scheme, we shall consider the following 1D nonlinear diffusion equation (see King & Please [17])

$$\frac{\partial c}{\partial T} = \frac{\partial}{\partial x}\left(d(c)\frac{\partial c}{\partial x}\right) \tag{37}$$

where the nonlinear diffusion coefficient is given by

$$d(c) = \frac{D_i}{1+\beta}[1 + \beta n_e/n_i]$$

with $n_e = \frac{1}{2}(c + \sqrt{c^2 + 4n_i^2})$ and $D_i$, $\beta$, $n_i$ are positive constants. The model equation has been tested before; see Baines, Birkett & Sweby [2] and Baines, Please & Sweby [3]. Following Moody & Please [26], we may nondimensionalize (37) by using

$$C = c/n_i, \qquad D = d/D_i, \qquad t = TD_i$$

and write the new equation as (2), *i.e.*

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left(D(C)\frac{\partial C}{\partial x}\right) \tag{38}$$

where

$$D(C) = \frac{1}{1+\beta}\left[1 + \frac{\beta}{2}\left(C + \sqrt{C^2+4}\right)\right] \tag{39}$$

Let us suppose that (38) is defined in $[0,1]$, equipped with the Neumann boundary condition

$$\frac{\partial C}{\partial x} = 0 \qquad \text{at } x = 0, \text{ and } 1 \tag{40}$$

Choose the initial profile at $t = 0$ to be the Gaussian distribution

$$C(x,0) = H\exp\left[-S(x-R)^2\right] \tag{41}$$

where the parameters $H$, $S$, $R$ can be chosen to reflect the difficulty of the problem. In process modelling, $H = 100$, $S = 100$ and $R = 0.25$ correspond approximately to the physical specification of an arsenic dose $N = 1.25\times10^{19}$ cm$^{-2}$, straggle $\sigma = 0.05$ $\mu$m, intrinsic carrier concentration $n_i = 10^{18}$ cm$^{-3}$, and intrinsic diffusion coefficient $D_i = 5 \times 10^{-8}$ $\mu m^2/s$, which can be worked out by using the true Gaussian profile $c(x,0) = \frac{N}{\sigma\sqrt{2\pi}}\exp\left[-(x-R)^2/(4\sigma^2)\right]$, where $R$ corresponds to the depth of the implant. Using the transformation (3), equation (38) can be written in the form of (4) *i.e.*

$$\frac{\partial f}{\partial t} = \frac{\partial}{\partial x}\left(D(e^f)\frac{\partial f}{\partial x}\right) + D(e^f)\left(\frac{\partial f}{\partial x}\right)^2 \tag{42}$$

We shall now proceed to solve (42) in this section by finite difference methods and in the next section by the ASWR method.

## 5.1 Adaptive and fixed time stepping

Using scheme (22), we have solved equation (42) (taking $\beta = 100$) for the following two cases

1. H=100, S=100, R=0.25
2. H=110, S=121, R=0.15

We take the number of subdivisions of $[0, 1]$ to be N=100 (or, equivalently, the step size $\Delta x = 10^{-2}$) and set the time T=20,000 seconds (or t=$10^{-3}$) and the tolerance TOL=$10^{-2}$. Using Theorem 1, given the initial profiles, we find that initial stability restrictions on $\Delta t$, independent of $\Delta x$, for the above two cases are respectively $\Delta t \leq 4.5 \times 10^{-5}$ and $\Delta t \leq 2.4 \times 10^{-5}$. This will serve as a guide to choosing the time step size. For our tests from now on, we shall use $\Delta t = 10^{-6}$ for uniform time stepping unless stated otherwise.

In Figs.1-2, we show results from implementing case **1**, where the three curves ("solid", "dotted" and "mixed-dotted") respectively represent the accurate solution, the solution using adaptive time stepping (see §4.4) and the solution using uniform time stepping. While the uniform stepping requires 500 steps to find the approximate solution, the adaptive stepping only needs 108 steps — about one fifth as many steps.

Similar results from running case **2** are shown in Figs.3-4. The test case represents a more difficult problem than case **1**, so a much smaller (scaled) time step of $\Delta t = 10^{-7}$ has to be taken for the uniform stepping case. This means that 1250 steps are needed to acquire the solution at the time $T^* = 5 \times 10^3$ (corresponding to the scaled time $t^* = 2.5 \times 10^{-4}$). In contrast, the adaptive stepping only takes 84 steps. But the uniform solution has become unstable, while the adaptive solution remains stable. More importantly, it is not possible to obtain a stable solution for $T > T^*$ with $\Delta t = 10^{-7}$. However our adaptive stepping does succeed in getting a reasonable solution at $T$, with 132 time steps. Note that Figs.3-4 shows solutions of the two stepping methods for both $T^*$ and $T$.

## 5.2 Alternative calculation of first order derivative terms

The failure of scheme (22) for case **2** with uniform time stepping prompts us to seek further improvements for the scheme. We have found that much improved results can be obtained if we use a *cross-product* differencing instead of a central differencing approximation to all first order derivative terms involved. In detail, we can modify scheme (22) to

$$\begin{cases} \dfrac{f_n^{j+\theta} - f_n^j}{\theta \Delta t} = \dfrac{D(e^{f_n^j})}{\Delta x^2} \delta^2 f_n^{j+\theta} + D(e^{f_n^j}) \left( \dfrac{f_{n+1}^j - f_n^j}{\Delta x} \right) \left( \dfrac{f_n^j - f_{n-1}^j}{\Delta x} \right) \\ \dfrac{f_n^{j+2} - f_n^j}{2\Delta t} = \dfrac{D(e^{f_n^{j+\theta}})}{\Delta x^2} \delta^2 f_n^{j+2} + D(e^{f_n^{j+\theta}}) \left( \dfrac{f_{n+1}^{j+\theta} - f_n^{j+\theta}}{\Delta x} \right) \left( \dfrac{f_n^{j+\theta} - f_{n-1}^{j+\theta}}{\Delta x} \right) \end{cases} \tag{43}$$

The order of approximation remains unchanged after the modification. This can be verified from applying the Taylor's expansion

$$\left( \frac{g_{n+1} - g_n}{\Delta x} \right) \left( \frac{g_n - g_{n-1}}{\Delta x} \right) = \left( \frac{\bar{\delta} g_n}{2\Delta x} \right)^2 - \frac{\Delta x^2}{4} \left( \frac{\partial^2 g}{\partial x^2} \right)^2 + O(\Delta x^3)$$

for any smooth function $g$.

With the modified scheme (43), we have re-calculated the two test cases of §5.1. For case **1**, almost identical results are obtained as those of §5.1. But for the more difficult case **2**, $\Delta t = 10^{-7}$ is acceptable for scheme (43) with uniform time stepping and we present results, in Figs.5-6, from using both the uniform and adaptive step strategy, where the latter takes 119 time steps.

The much improved performance implies that the cross-product approximation is superior to the central differencing and hence should be used whenever possible.

# 6 Application of the ASWR method

We now consider the solution of the nonlinear diffusion equation (42) with conditions (40) and (41) by the ASWR method and, in particular, test the two cases already solved by finite difference methods in §5.1.

## 6.1 Spline approximation

It is of interest to discuss, at this point, some error analysis theory of the spline approximation. One immediate application of spline approximation is to approximate the initial function $f(x,0)$ in the space $S_2$ (see §3.2). There are a number of methods which can provide good approximations. Extensive discussion can be found in De Boor [10]. Interpolation is an often used technique.

In our context, two interpolation approximations may be used : the quasi-interpolants and the conventional interpolant. Given a smooth function $g \in C^2[a,b]$, we may denote by $Ig$ its interpolation approximation. We seek $Ig$ from $S_2$. The quasi-interpolant can be explicitly constructed by

$$Ig = Qg = \sum_{i=0}^{N+1} \bar{\alpha}_i B_i(x) \tag{44}$$

with

$$\bar{\alpha}_i = g(\eta_i) + \frac{(t_{i+1} - \eta_i)(t_{i+2} - \eta_i)}{2} \frac{d^2 f}{dx^2}(\eta_i)$$

such that

$$\|g - Ig\|_\infty \leq \text{const} \Delta x^2 \omega(g''; \Delta x) \tag{45}$$

where $t_i < \eta_i < t_{i+3}$, $\omega$ denotes the modulus of continuity for function $g$ and $\Delta x = \max\{h_i\}$ (see §3). The approximation $Qg$ is called the quasi-interpolant of $g$ because it approximates $g$ and its derivative $g'$ at optimal orders but does not agree with $g$ at the nodes $\eta_i$. Unfortunately it is not easy to generalize the method to the 2D case. We here prefer to use the conventional interpolation

$$Ig = \sum_{i=0}^{N+1} \alpha_i B_i(x) \tag{46}$$

such that $(Ig)(\xi_j) = g(\xi_j)$ and this leads to a linear system with a tridiagonal matrix

$$\sum_{i=j-1}^{j+1} \alpha_i B_i(\xi_j) = g(\xi_j), \qquad j = 0, 1, 2, \cdots, N, N+1 \tag{47}$$

where $\xi_0 = a$ and $\xi_{N+1} = b$ and $\{\xi_j\}_1^N$ is some strictly monotone sequence in $(a,b)$. When the sequence is chosen to be the mid-points as defined in §3.1, it can be shown that the error bound (45) also holds.

## 6.2 The numerical scheme

In §3.3, we have presented the ASWR method for semi-discrete solution of equation (42). We now combine the method with the specific predictor corrector scheme (22). To simplify the implementation, we shall use the equivalence of piecewise quadratic functions in local and global (B-spline) representations for treating all explicit quantities. More precisely, on the known time level $\tau = j$ or $\tau = j + \theta$, the local representation for $f_N^\tau$ will be used. Since $f_N^\tau \in S_2$, we may expand it, on each interval $I_k = (x_{k-1}, x_k]$, into the Taylor series centred at $\xi_k$

$$P_{N,k}^\tau = f_N^\tau(\xi_k) + (x-\xi_k)\frac{f_N^\tau(x_k) - f_N^\tau(x_{k-1})}{\Delta x} + 2(x-\xi_k)^2\frac{f_N^\tau(x_k) + f_N^\tau(x_{k-1}) - 2f_N^\tau(\xi_k)}{\Delta x^2} \quad (48)$$

Using $P_{N,k}^\tau$ to replace $f_N^\tau$ locally, we can write the full ASWR discretization for (42) (i.e. (4)) as

$$\begin{cases} \sum_{i=k-1}^{k+1} \alpha_i^{j+\theta}\left[\int_{I_k} B_i dx - \theta\Delta t \int_{I_k} \frac{\partial}{\partial x}\left(D(P_{N,k}^j)\frac{\partial B_i}{\partial x}\right) dx\right] \\ = \int_{I_k} P_{N,k}^j dx + \theta\Delta t \int_{I_k} D(P_{N,k}^j)\left(\frac{\partial P_{N,k}}{\partial x}\right)^2 dx, \\ \alpha_0^{j+\theta} = \alpha_1^{j+\theta}, \qquad \alpha_N^{j+\theta} = \alpha_{N+1}^{j+\theta}, \\ \sum_{i=k-1}^{k+1} \alpha_i^{j+2}\left[\int_{I_k} B_i dx - 2\Delta t \int_{I_k} \frac{\partial}{\partial x}\left(D(P_{N,k}^j)\frac{\partial B_i}{\partial x}\right) dx\right] \\ = \int_{I_k} P_{N,k}^j dx + 2\Delta t \int_{I_k} D(P_{N,k}^{j+\theta})\left(\frac{\partial P_{N,k}^{j+\theta}}{\partial x}\right)^2 dx, \\ \alpha_0^{j+2} = \alpha_1^{j+2}, \qquad \alpha_N^{j+2} = \alpha_{N+1}^{j+2}, \end{cases} \quad (49)$$

where $k = 1, \cdots, N$. Evidently the coefficients at the right hand side of (49) involving $D(P)(\frac{\partial P}{\partial x})^2$ will have to be evaluated numerically *e.g.* by using some Gaussian-Legendre quadrature rule. However the remaining integrals can be analytically calculated.

## 6.3 A local truncation error analysis

We now consider the temporal mesh adaption, which requires local truncation error estimates (see §4.4). To carry out a local truncation error analysis for (49), it is necessary to express quantities $\{\alpha_i^\tau\}$ for $\tau = j, j+\theta, j+2$ in terms of the original variable $f$. This

involves the solution of a linear tridiagonal system (see (46) of §6.1). On the other hand, each $\alpha_i^\tau$ is determined by all of $\{f_N^\tau(\xi_k)\}$ since the inverse of a band matrix is full and this global dependence is not suitable for a local error analysis. However interpolation using B-splines always generates a diagonally dominant matrix.

It is known that the elements of the inverse of a diagonally dominant matrix are bounded in an exponentially decaying manner along each row or column. Refer to Concus, Golub & Meurant [8] and Meurant [23]. A simple method to find approximations for such an inverse proceeds as follows. Suppose $A = D - B$, where $D$ is the diagonal of the diagonally dominant matrix $A$. Then $\|D^{-1}B\| < 1$ and $A^{-1} = (I - D^{-1}B)^{-1}D^{-1}$. The first part of $A^{-1}$ may be expanded in a convergent power series $(I - D^{-1}B)^{-1} = \sum_{k=0}^{\infty}(D^{-1}B)^k$. Further

$$A^{-1} = \sum_{k=0}^{\infty}(D^{-1}B)^k D^{-1} \tag{50}$$

may be used to provide approximations by truncating the series.

In our applications, the fact that the inverse of a diagonally dominant matrix may be accurately approximated by a band matrix corresponds to local dependence being dominant. Without using (50), we may compute $A^{-1}$ numerically and find a band approximation from the exact inverse.

Now consider the linear model problem (5) and employ an uniform spatial mesh $\Pi_N$ (see (6)). Then the interpolation relation (47) becomes

$$\begin{bmatrix} 7 & 1 & & & \\ 1 & 6 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 6 & 1 \\ & & & 1 & 7 \end{bmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix} = 8 \begin{pmatrix} f_N(\xi_1) \\ f_N(\xi_2) \\ \vdots \\ f_N(\xi_N) \end{pmatrix} \tag{51}$$

It can be verified that a reasonable approximation for the inverse of the coefficient matrix is

$$\frac{1}{8}\begin{bmatrix} 1.25 & -0.25 & & & \\ -0.25 & 1.5 & -0.25 & & \\ & \ddots & \ddots & \ddots & \\ & & -0.25 & 1.5 & -0.25 \\ & & & -0.25 & 1.25 \end{bmatrix} \tag{52}$$

Therefore for a typical k $(= 2, \cdots, N-1)$, we have the following approximate solution of (51)

$$\alpha_k^\tau = \frac{1}{4}\left[6f_N^\tau(\xi_k) - f_N^\tau(\xi_{k+1}) - f_N^\tau(\xi_{k-1})\right] \tag{53}$$

For equation (5), the local truncation error of the ASWR corrector step can be shown to be the following

$$\tau_A = \frac{1}{12\Delta t}\bar{\delta}_t\left(\alpha_{k-1}^{j+2} + 4\alpha_k^{j+2} + \alpha_{k+1}^{j+2}\right) - \frac{D}{\Delta x^2}\delta^2\alpha_k^{j+2} - \frac{V}{2\Delta x}\bar{\delta}\alpha_k^{j+\theta} \tag{54}$$

where $\bar{\delta}_t\alpha_r^{j+2} = \alpha_r^{j+2} - \alpha_r^j$. Now with the approximation (53), the local truncation error of the ASWR may be expressed in terms of the true solution $F(x,t)$

$$\begin{aligned}\tau_A &= \frac{\bar{\delta}_t}{48\Delta t}\left[22F_k^{j+2} + 2(F_{k+1}^{j+2} + F_{k-1}^{j+2}) - (F_{k+2}^{j+2} + F_{k-2}^{j+2})\right] \\ &- \frac{D}{4\Delta x^2}\left[8(F_{k+1}^{j+2} + F_{k-1}^{j+2}) - (F_{k+2}^{j+2} + F_{k-2}^{j+2}) - 14F_k^{j+2}\right] \\ &- \frac{V}{8\Delta x}\left[6(F_{k+1}^{j+\theta} - F_{k-1}^{j+\theta}) - (F_{k+2}^{j+\theta} - F_{k-2}^{j+\theta})\right]\end{aligned} \tag{55}$$

**THEOREM 2** *(LTE for ASWR) For equation (5), using the ASWR finite element method with the predictor corrector scheme (22), the local truncation error may be estimated by*

$$\tau_A = R_A^{(0)}\Delta t + R_A^{(1)}\Delta t^2 + R_A^{(2)}\Delta x^2 + \cdots \tag{56}$$

*where at* $(\xi_k, t_j)$

$$\begin{aligned}R_A^{(0)} &= \frac{\partial}{\partial t}\left[(1-\theta)V\frac{\partial F}{\partial x} - D\frac{\partial^2 F}{\partial x^2}\right] \\ R_A^{(1)} &= \frac{1}{6}\frac{\partial^2}{\partial t^2}\left[(4-3T^2)V\frac{\partial F}{\partial x} - 4D\frac{\partial^2 F}{\partial x^2}\right] \\ R_A^{(2)} &= -\frac{D}{4}\frac{\partial^4 F}{\partial x^4}\end{aligned}$$

**Proof.** The result follows immediately from the Taylor expansions of (55) at $(\xi_k, t_j)$. Note that $R_A^{(0)} = R_c^{(0)}$ from (33), which is expected from applying the same time stepping scheme. □

With the local error estimates of Theorem 2, automatic selection of temporal step sizes (of $\theta$ and $\Delta t$) follows from the strategy of §4.4. Below we shall present our test results from the solution of system (49) for the two cases of §5.1.

In Figs.7-8 and Figs.9-10, we show the plots of running the test cases of §5.1 respectively, using the ASWR method with our time stepping control, where the "solid" line again denotes the accurate solution at $T = 2\times10^4$. Respectively for the two test cases, the number of time steps taken are 38 and 374. Here we note that for the simpler test case 1 our adaptive ASWR method is faster than the finite difference method but for the more difficult case 2 the converse is true.

Recall from §5.2 that an alternative treatment for the first order derivative terms of the test equation (4) may improve the performance of the finite difference method. To utilize the idea for the ASWR finite element method, we can use the following approximation for the calculation of the right hand sides of (49)

$$\left(\frac{\partial P^{\tau}_{N,k}}{\partial x}\right)^2 = 4\frac{\{f^{\tau}_N(x_{k+1}) - f^{\tau}_N(\xi_k)\}}{\Delta x}\frac{\{f^{\tau}_N(\xi_k) - f^{\tau}_N(x_k)\}}{\Delta x} \tag{57}$$

Using (57) to modify our adaptive method, we have solved the more difficult case 2 with improved results obtained, as shown Figs.11-12, where 322 time steps are required. The improvement is not as dramatic as for the finite difference method (see Figs.5-6). Nevertheless, the incorporation of the *cross product* approximation (57) is still recommended.

# 7 Conclusions

In this report we have investigated the problem of automatic time step selection for a predictor-corrector scheme applied to both the finite difference and finite element Petrov-Galerkin methods for a process modelling diffusion model. The strategy for the hybrid predictor-corrector method (using 'off-step' time levels) seeks a nearly optimal predictor step in order to minimize the one step local truncation error and then selects the corrector step via controlling the local truncation error. For the finite element method of the ASWR type, an appropriate approximation has been found to estimate the local truncation error and further to apply the time stepping strategy designed for the finite difference method. Numerical experiments on solving the 1D nonlinear semiconductor diffusion equations are

carried out, which demonstrate the efficiency of our proposed adaptive strategy. The systematic development presented here serves us as a guide in investigating the 2D nonlinear diffusion equations. Some new results have been obtained from our 2D study and further details will be reported soon.

## Acknowledgement

Financial support from the Department of Trade and Industry (DTI) through the Science and Engineering Research Council (SERC) is gratefully acknowledged.

## References

[1] W. F. AMES, *Numerical methods of partial differential equations*, 2nd ed., Academic, 1977.

[2] M. J. BAINES, N. R. C. BIRKETT & P. K. SWEBY, "Nonlinear diffusion in process modelling", *Int. J. Numer. Model. : Elec. Net. Dev. Fields*, **3**, 79-90, 1990.

[3] M. J. BAINES, C. P. PLEASE & P. K. SWEBY, "Numerical solution of dopant diffusion equations", in *Simulation of Semiconductor Devices and Processes*, **2**, eds. K. Board & D. R. J. Owen, 272-286. (Pineridge Press, 1986).

[4] D. S. BURNETT, *Finite element analysis : from concepts to applications*, Addison-Wesley, 1987.

[5] J. W. BARRETT & K. W. MORTON, "Optimal Petrov-Galerkin methods through approximate symmetrization", *IMA J. Numer. Anal.*, **1**, 469-487, 1981.

[6] L. BORUCKI, H. H. HANSEN & K. VARAHRAMYAN, "FEDSS - a 2D semiconductor fabrication process simulator", *IBM J. Res. Dev.*, **29**(3), 217-328, 1985.

[7] K. CHEN, "Analysis and experiments of a set of three-level semi-implicit finite difference schemes for nonlinear parabolic equations", NA Report **13/90**, Dept of Mathematics, University of Reading, England, 1990.

[8] P. CONCUS, G. H. GOLUB & G. MEURANT, "Block preconditioning for the conjugate gradient method", *SIAM J. Sci. Stat. Comput.*, **6**(1), 220-252, 1985.

[9] H. B. CURRY & I. J. SCHOENBERG, "On Polya frequency functions. IV : the fundamental functions and their limits", *J. Anal. Math.*, **17**, 71-107, 1966.

[10] C. DE BOOR, *A practical guide to splines*, Springer-Verlag, 1978.

[11] J. DOUGLAS & T. DUPONT, "Galerkin methods for parabolic equations", *SIAM J. Numer. Anal.*, **7**(4), 575-626, 1970.

[12] B. A. FINLAYSON, *The method of weighted residuals and variational principles*, Academic, 1972.

[13] A. GERODOLLE, S. MARTIN & A. MARROCCO, "Finite element method applied to 2D MOS process simulation and point defect diffusion : program TITAN", in : Proc. NASECODE IV, 287-292, Boole Press, Dublin, 1985

[14] P. M. GRESHO, R. L. LEE & R. L. SANI, "On the time dependent solution of the incompressible Navies Stokes equations in two and three dimensions", in : *Recent Advances in Numerical Methods in Fluids*, **1**, 27-79, eds. C. Taylor & K. Morgan, Pineridge Press, 1980.

[15] G. P. HO, J. D. PLUMMER, S. E. HANSEN & R. W. DUTTON, "VSLI — process modelling — Suprem III", *IEEE Trans. Elec. Dev.*, **ED-27**, 1438-1453, 1983.

[16] I. W. JOHNSON, A. J. WATHEN & M. J. BAINES, "Moving finite element methods for evolutionary problems. II. applications", *J. Compt. Phys.*, **79**(2), 270-297, 1988.

[17] J. KING & C. PLEASE, "Diffusion of dopant in cystalline silicon", *IMA J. Appl. Math.*, **37**, 185-197, 1986.

[18] A. LACEY, J. OCKENDON & A. B. TAYLER, "'Waiting time' solutions of a non-linear diffusion equation", *SIAM J. Numer. Anal.*, **42**, 1252-1264, 1982.

[19] J. D. LAMBERT, *Computational methods in ordinary differential equations*, Wiley, 1973.

[20] M. E. LAW & R. W. DUTTON, "Verification of analytic point defect models using SUPREM-IV", *IEEE Trans. CAD*, **CAD-7**(2), 181-190, 1988.

[21] J. LORENZ, J. PELKA, H. RYSSEL, A. SACH, A. SEIDEL & M. SVOBODA, "COMPOSITE – a complete modeling program of silicon technology", *IEEE Trans. Elec. Dev.*, **ED-32**(10), 1977-1986, 1985.

[22] J. LORENZ & M. SVOBODA, "ASWR - method for the simulation of dopant redistribution in silicon", in *Simulation of Semiconductor Devices and Processes*, **3**, eds. G. Daccarani et al., 243-254, Tecnoprint, 1988.

[23] G. MEURANT, "A review on the inverse of symmetric tridiagonal and block tridiagonal matrices", to appear in *SIAM J. Appl. Math.*, 1991.

[24] A. R. MITCHELL & D. F. GRIFFITHS, *The Finite Difference Method in Partial Differential Equations*, Wiley, 1980.

[25] A. MITCHELL & R. WAIT, *The Finite Element Method in Partial Differential Equations*, Wiley, 1977.

[26] R. O. MOODY & C. P. PLEASE, "A numerical simulation of one-dimensional dopant diffusion in oxidising silicon', *Numerical Analysis Report*, **3/87**, Department of Mathematics, University of Reading, UK, 1987.

[27] B. R. PENUMALLI, "A comprehensive two-dimensional VLSI process simulation program, BICEPS", *IEEE Tran. Elec. Dev.*, **ED-30**(9), p.986-992, 1983.

[28] C. P. PLEASE & P. K. SWEBY, "A transformation to assist the numerical solution of diffusion equations', *Numerical Analysis Report*, **5/86**, Department of Mathematics, University of Reading, UK, 1986.

[29] S. J. POLAK, W.H.A. SCHILDERS & P. MARKOWICH, "Semiconductor device modelling from the numerical point of view", *Int. J. Numer. Meth. Engng.*, **24**, 763-838, 1987.

[30] S. SELBERHERR, *Analysis and simulation of semiconductor devices*, Springer-verlag, 1984.

[31] B. W. SCOTNEY, *An analysis of the Petrov-Galerkin finite element method*, Ph.D thesis, Dept of Mathematics, University of Reading, UK, 1985.

[32] K. STREHMEL, W. H. HUNDSDORFER, R. WEINER & A. ARNOLD, "The linearly implicit Euler method for quasi-linear parabolic differential equations", *Report* **NM-R9002**, Department of Mathematics, Centre for Mathematics and Computer Science (CWI), Amsterdam, Netherlands, Feb., 1990.

[33] M. SVOBODA, "A Petrov-Galerkin procedure for simulation of solid-state diffusion in semiconductor fabrications", Ph.D. thesis, Faculty of Mathematics, Ludwig Maximillians University, Munich, Germany, 1988.

[34] S. M. SZE (ed.), *VLSI technology*, second edition, McGraw-Hill, 1988.

Figure 1. Finite difference methods using the ASWR time stepping

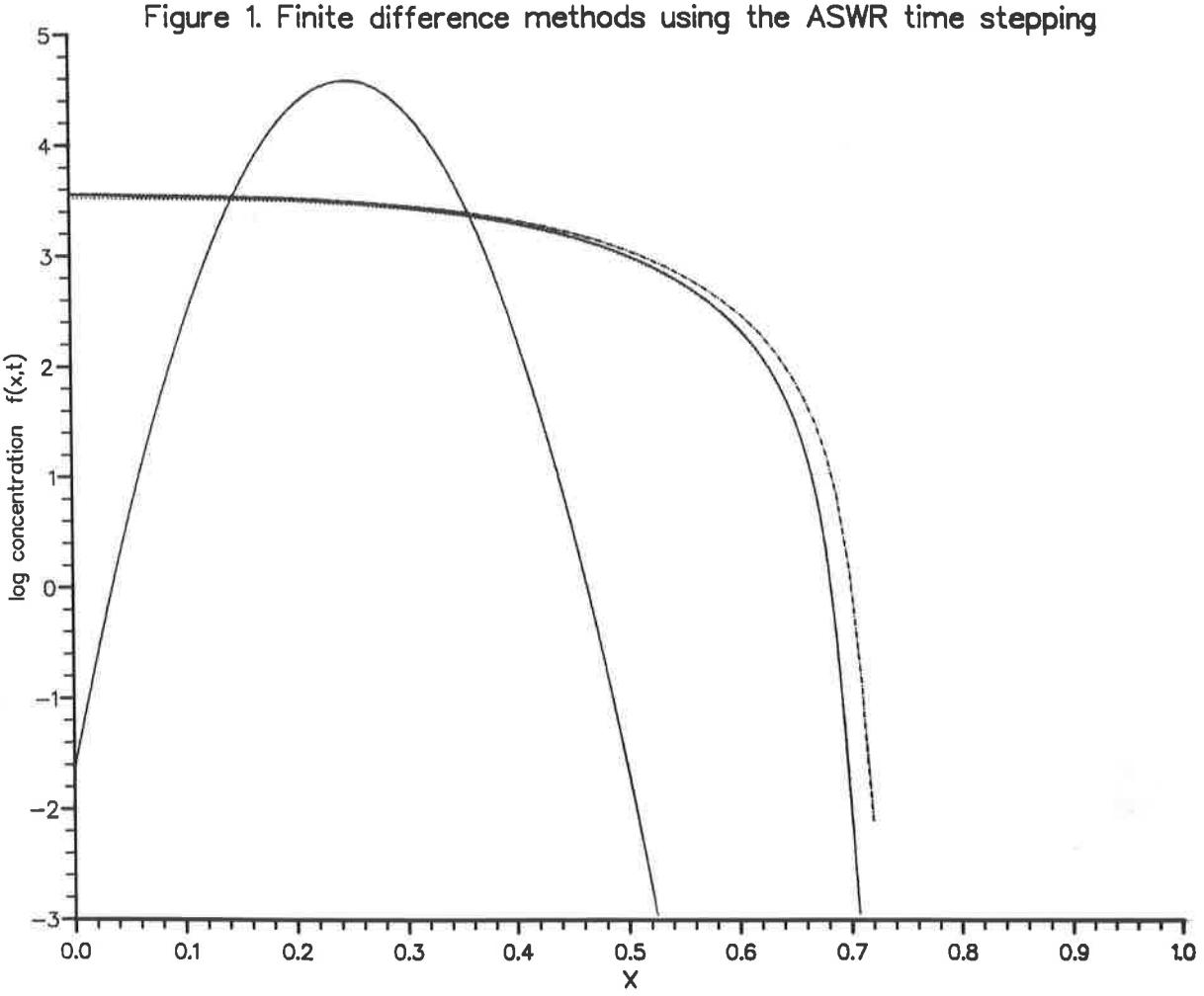


Figure 2. Finite difference methods using the ASWR time stepping

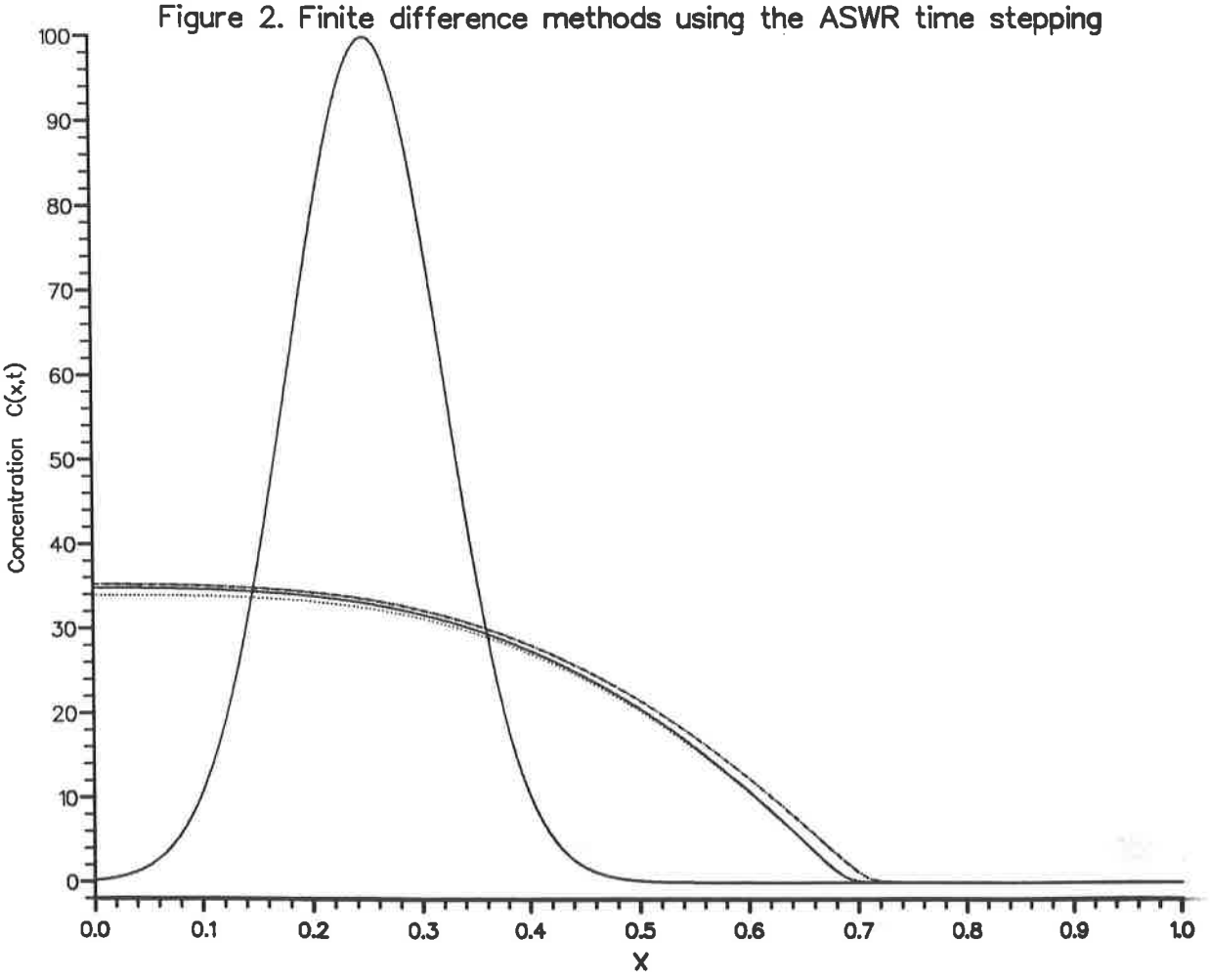

Figure 3. Finite difference methods using the ASWR time stepping

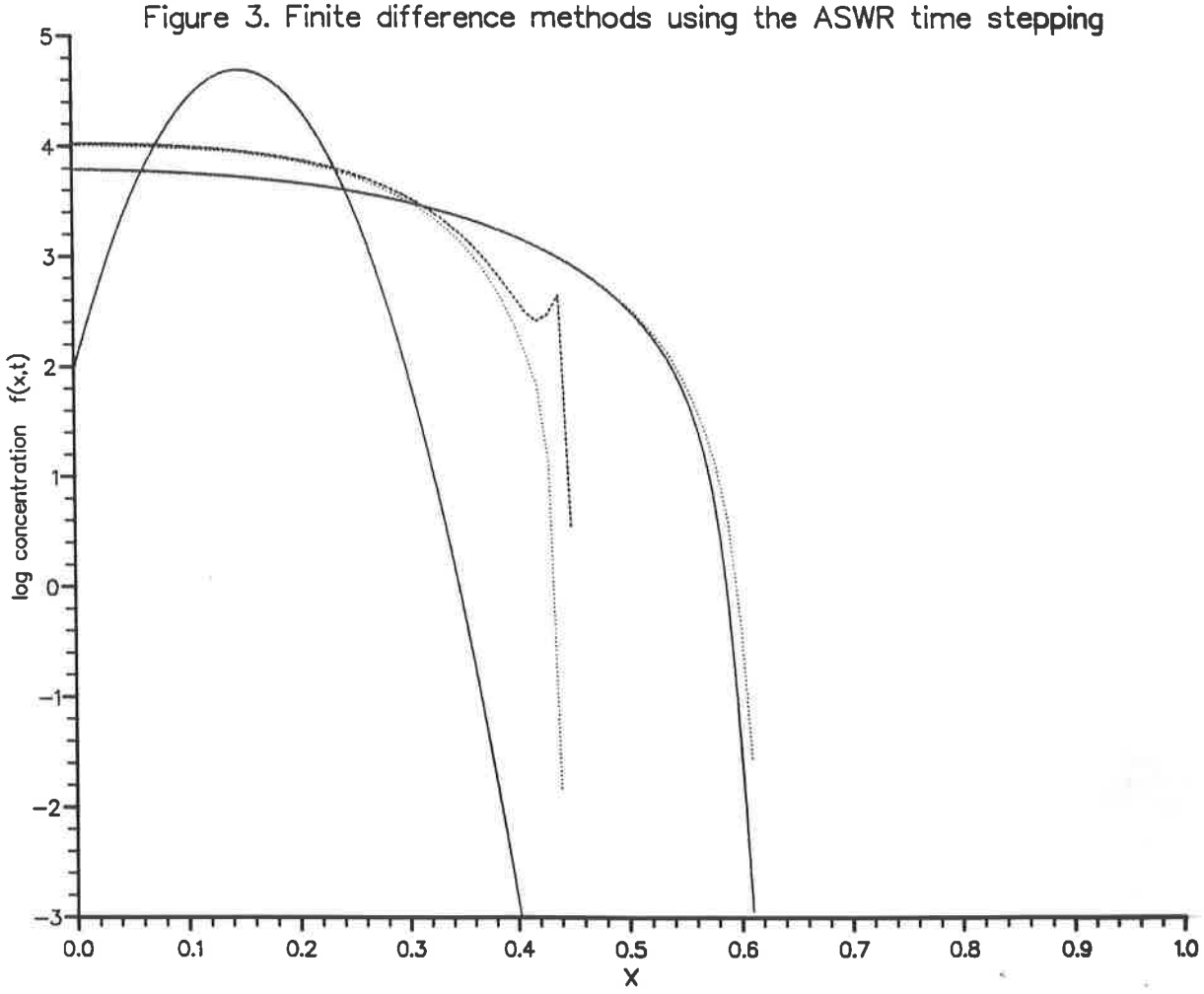


Figure 4. Finite difference methods using the ASWR time stepping

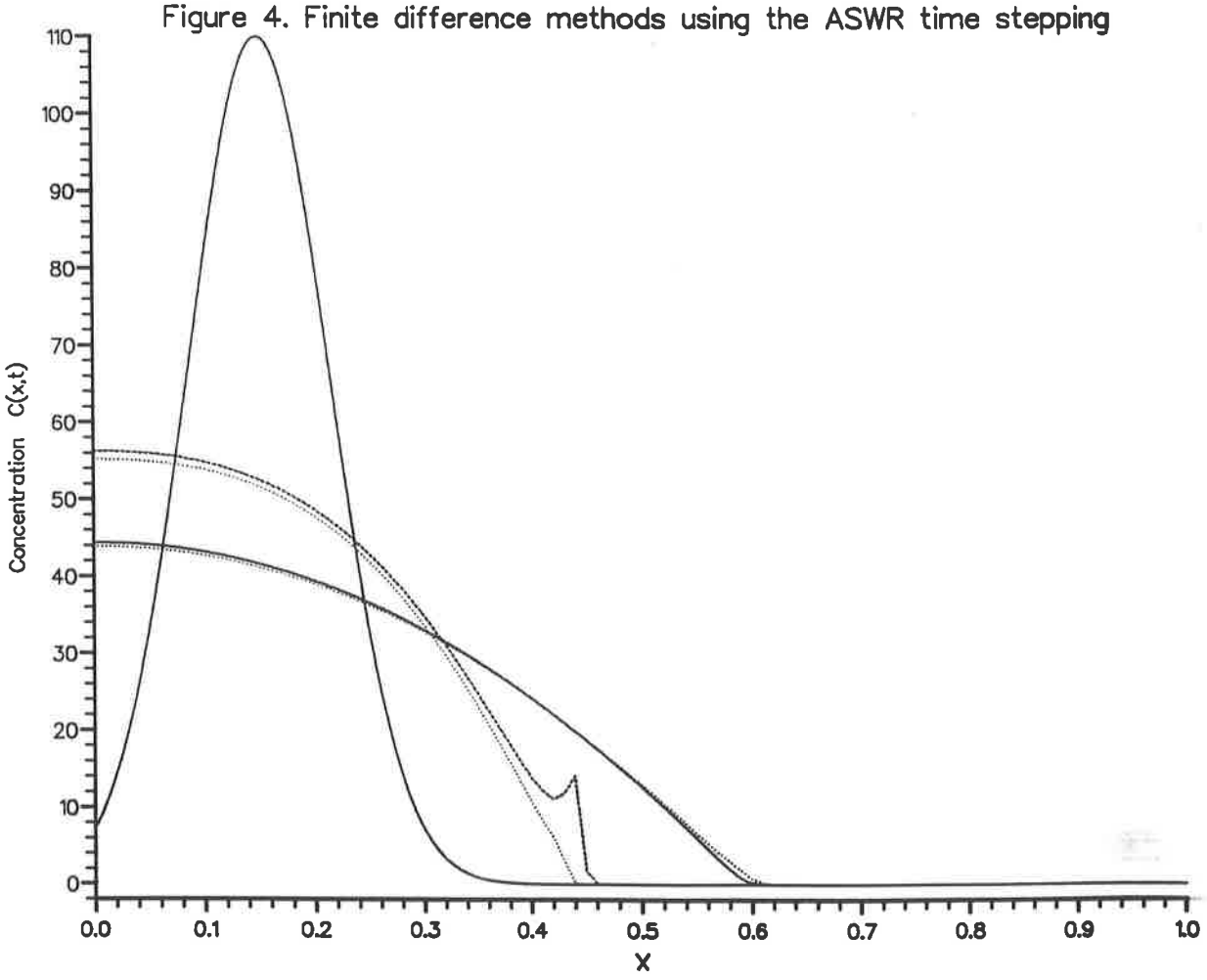

Figure 5. Finite difference methods using the ASWR time stepping

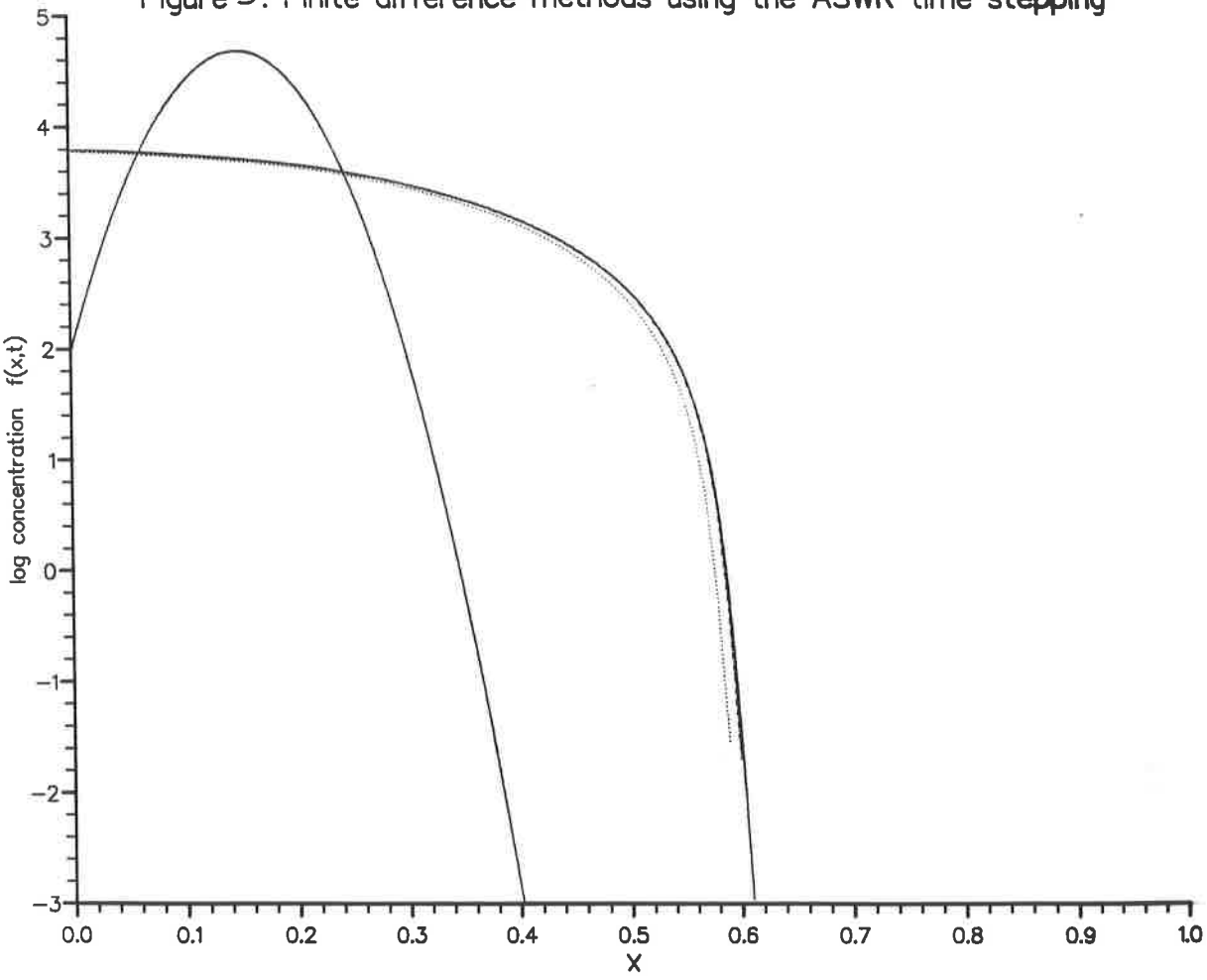


Figure 6. Finite difference methods using the ASWR time stepping

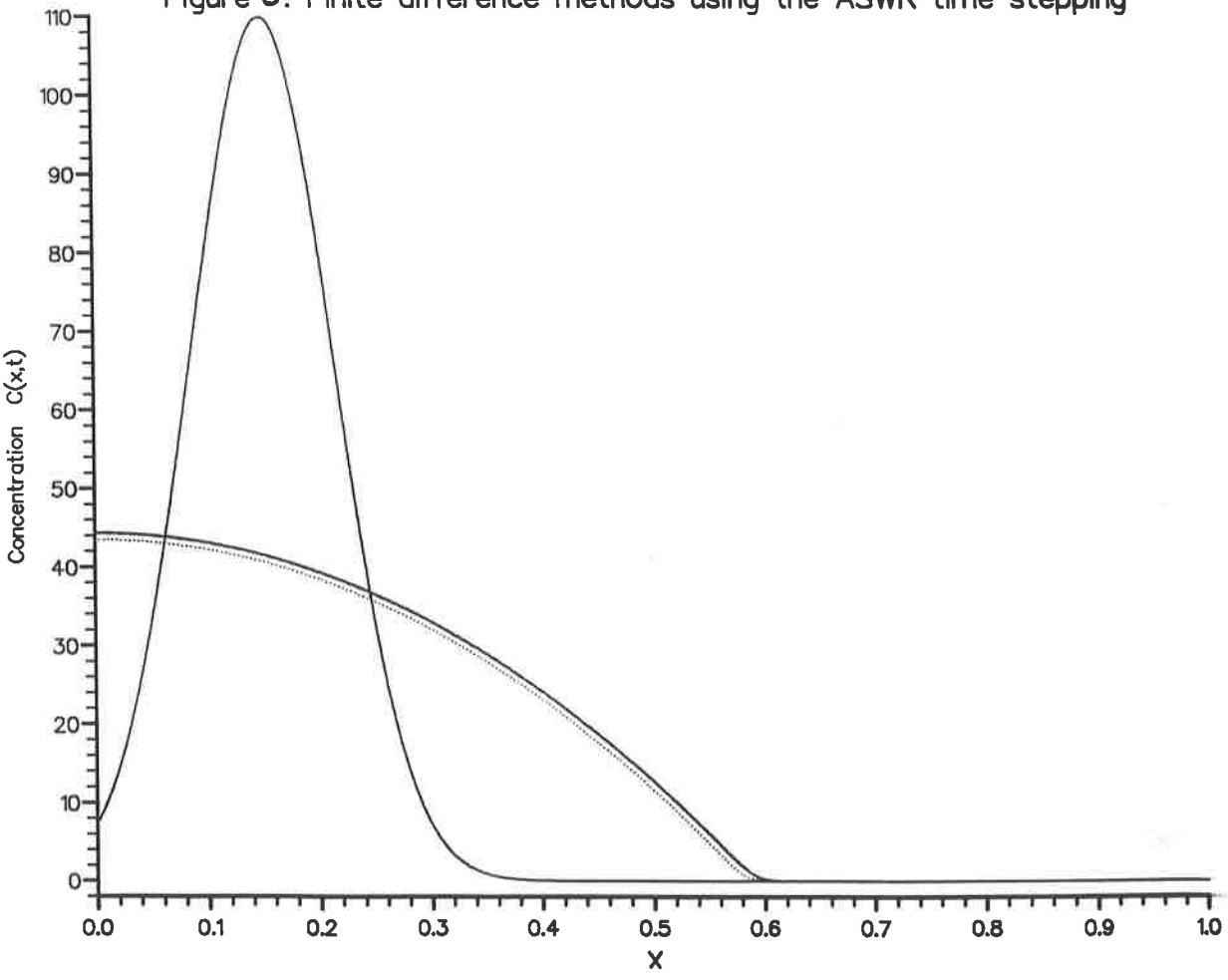

Figure 7. ASWR finite element methods using the ASWR time stepping

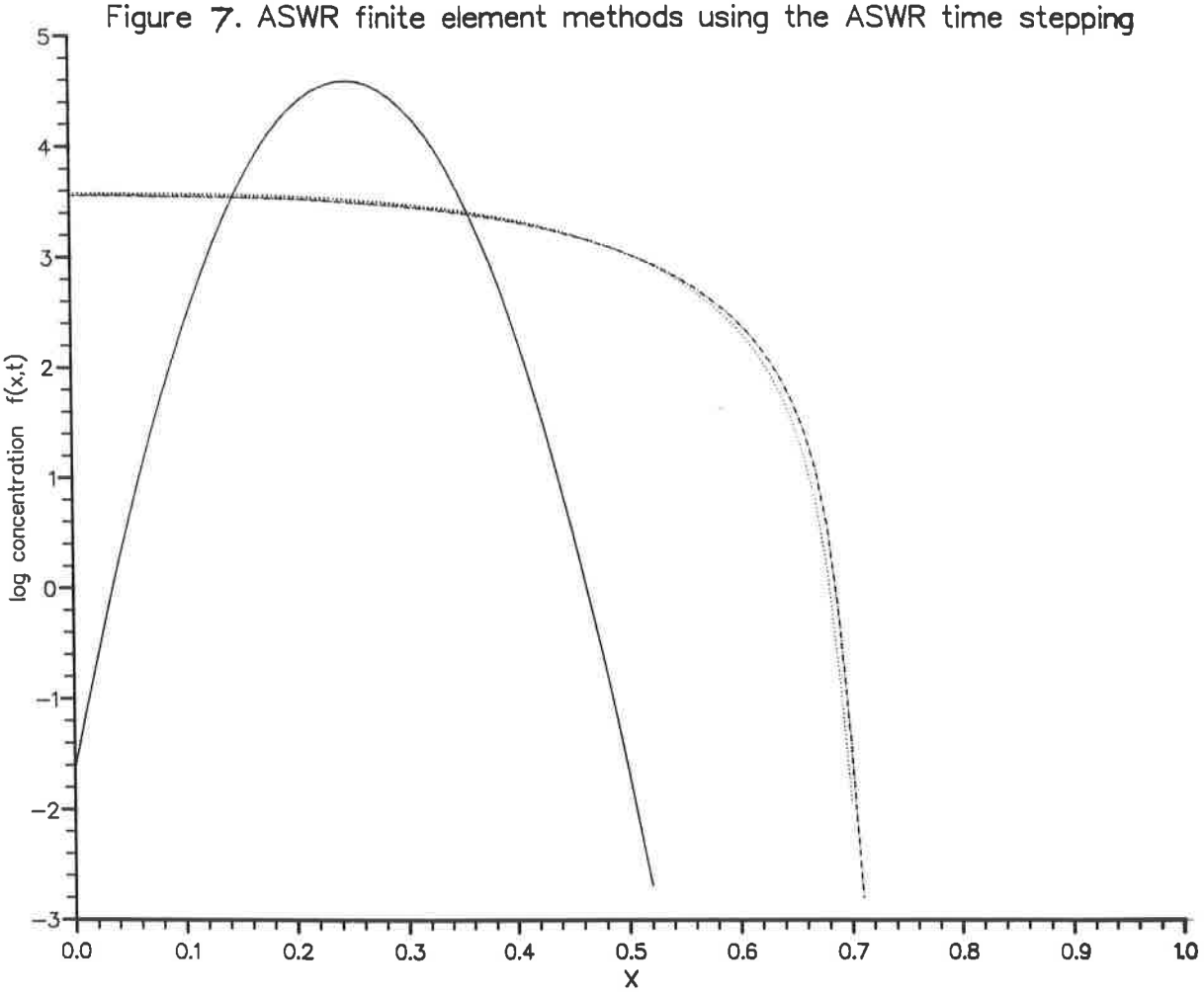


Figure 8. ASWR finite element methods using the ASWR time stepping

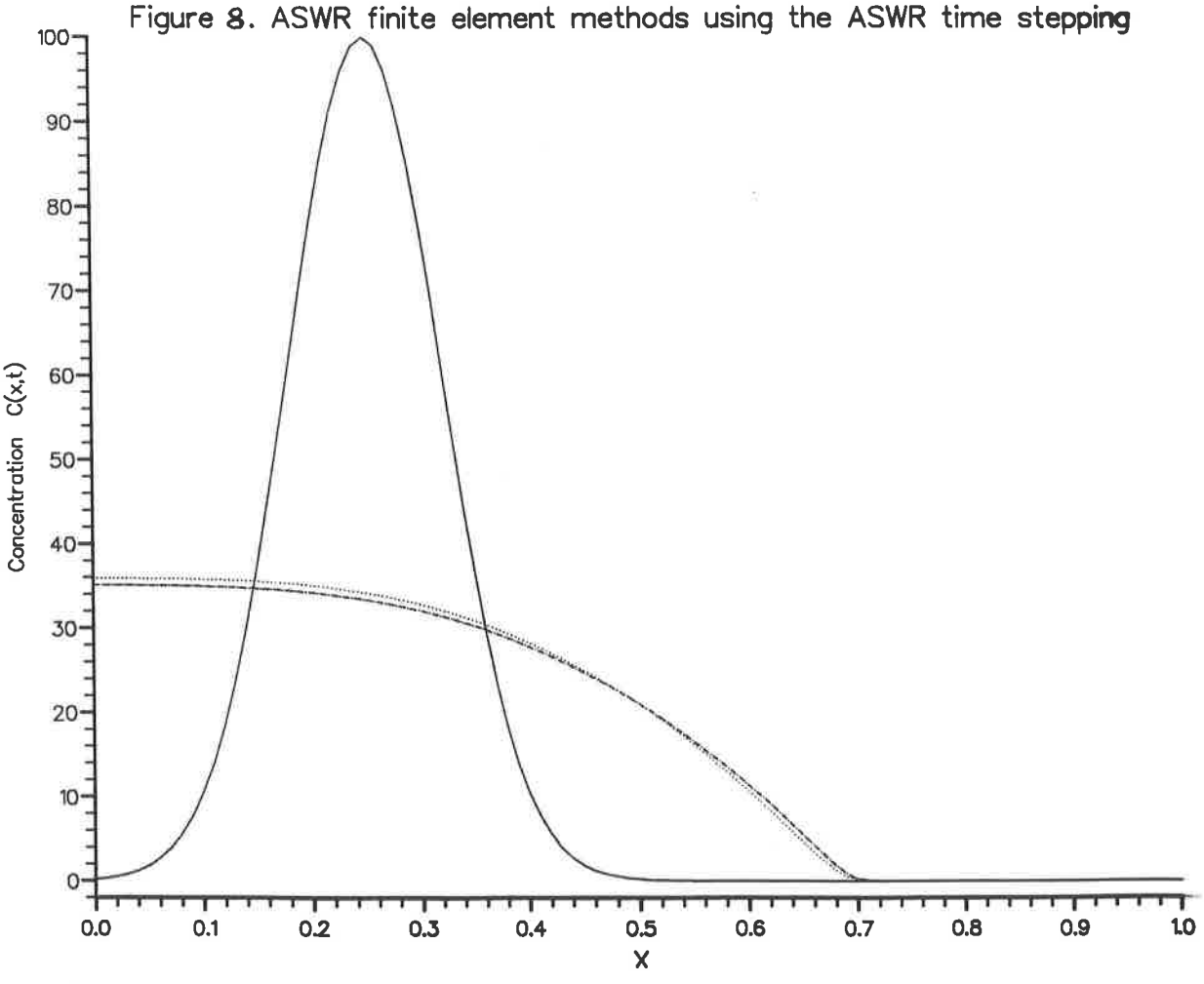

Figure 9. ASWR finite element methods using the ASWR time stepping

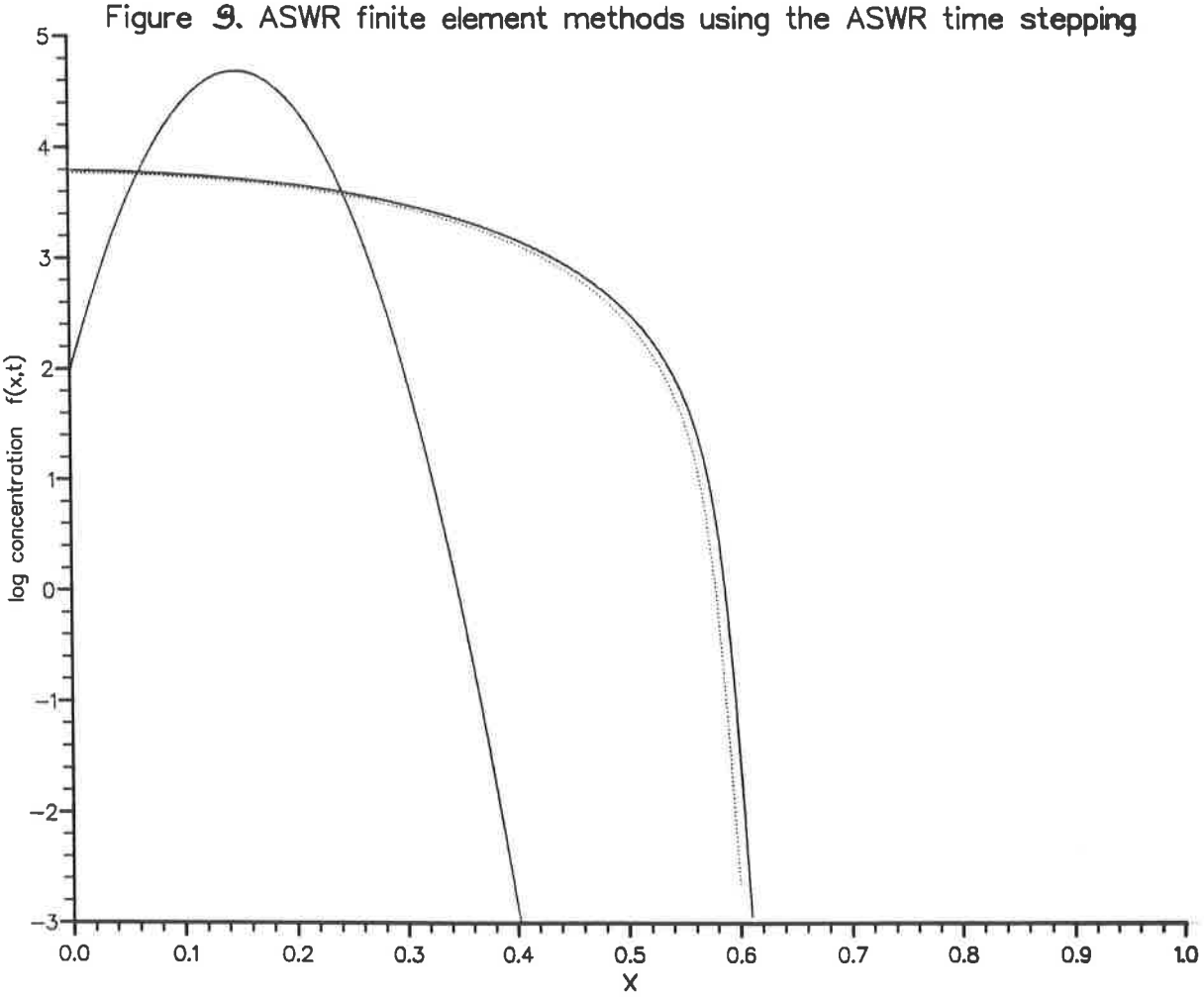


Figure 10. ASWR finite element methods using the ASWR time stepping

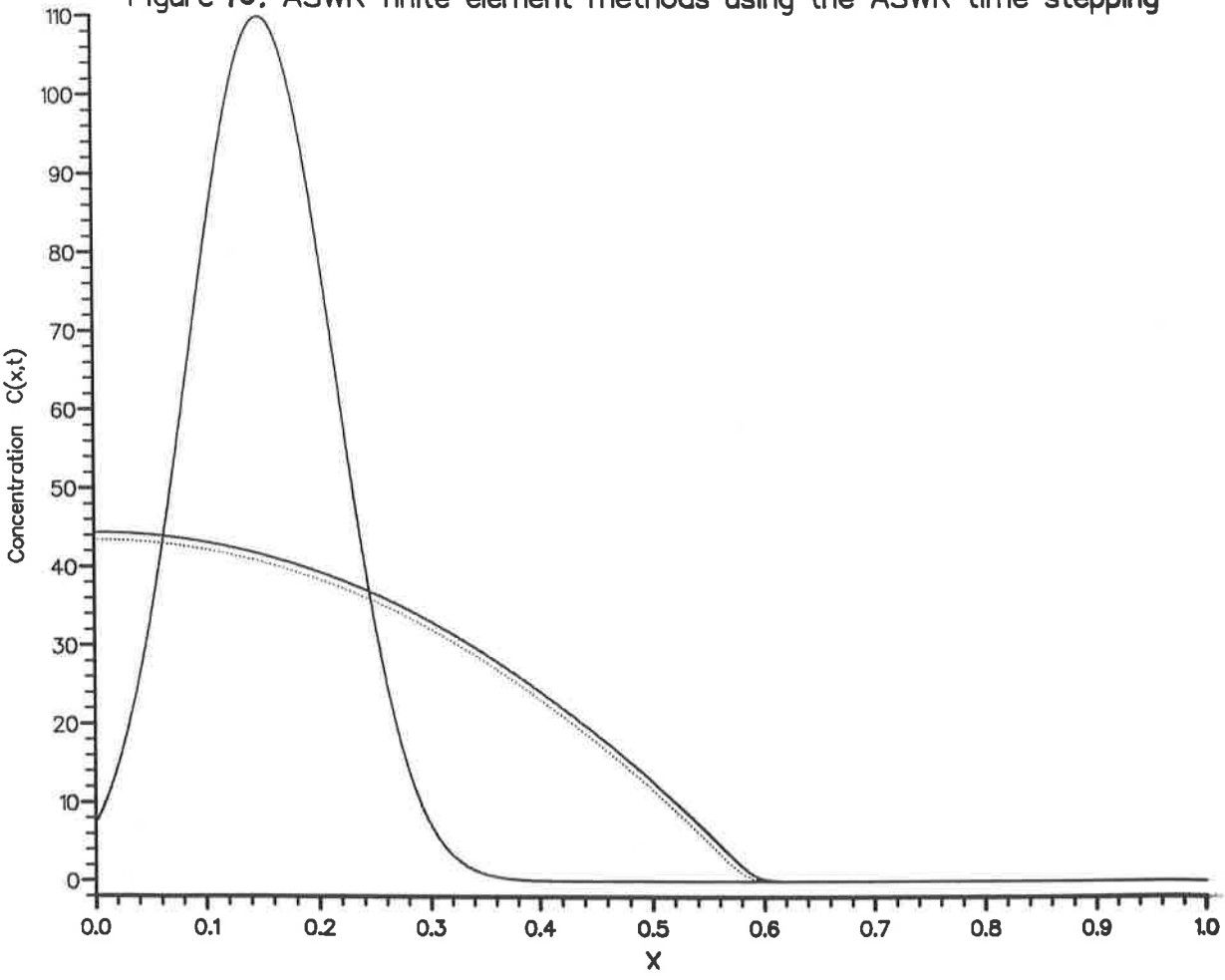

Figure 11. ASWR finite element methods using the ASWR time stepping

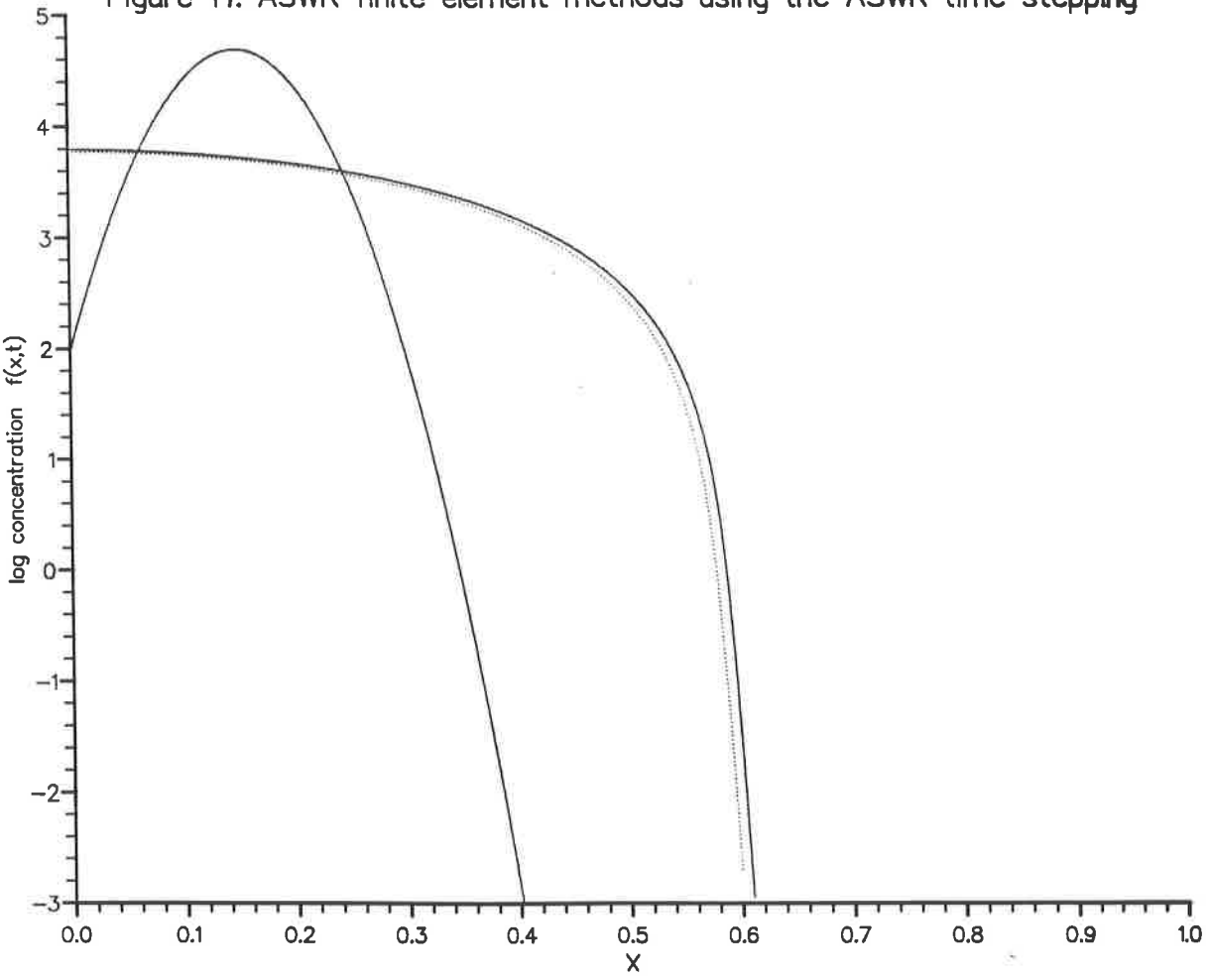


Figure 12. ASWR finite element methods using the ASWR time stepping