JOURNAL OF MATERIALS SCIENCE 16 (1981) 3122–3130

Efficient numerical modelling of hydrogen diffusion with trapping

PETER G. THOMAS, ELLEN J. STERN
Faculty of Mathematics, The Open University, Milton Keynes, Bucks, UK

In view of the growing interest both in numerical solutions of the McNabb and Foster model for hydrogen diffusion with trapping, and the computational scheme proposed by Caskey and Pillinger, a more efficient, simpler scheme is presented which more closely replaces this model. Further techniques are presented which reduce the cost of the computations yet increase the accuracy of the results. Using a formulation better adapted to analysis, and with readily comprehensible variables, bounds are given on the numerical mesh sizes which ensure the stability of the scheme. For cases where these bounds are overly restrictive, alternative numerical tests are suggested. Finally, driving equations are given on which an approximation to the McNabb and Foster model in which the concentration of traps varies with time may be validly based.

1. Introduction

In their seminal paper "A New Analysis of the Diffusion of Hydrogen in Iron and Ferritic Steels", McNabb and Foster [1] published in 1963 the mathematical model now commonly referred to as "hydrogen diffusion with trapping". The driving equations of the model form a coupled, non-linear, second-order system of two partial differential equations, involving four parameters. Even with the simplest non-trivial boundary and initial conditions, the model has not, so far, been solved analytically except in a few limiting cases. In the present state of experimental expertise, moreover, at most two of the parameters may be amenable to individual physical determination, together with the ratio of the remaining two. However, McNabb and Foster [1] showed convincingly in their elegant paper that the model "does appear to explain many anomalies that have been reported in the literature" (concerning diffusion of hydrogen in iron and ferritic steels), and this has led to considerable interest in numerical solutions of the model.

The first reported numerical solution to the McNabb and Foster model in one finite space dimension was published by Caskey and Pillinger [2] in 1975; this appears now to be regarded as standard, and is quoted as such for example by Frank *et al.* [3], who show how the same computational scheme may be adapted to one of the many possible modifications of the basic model. We present here a computationally more efficient scheme, requiring less computer storage, and describe how the run-time of the relevant computations may be reduced from that required for the Caskey and Pillinger scheme whilst retaining the same order of accuracy. We also indicate how better accuracy can be obtained with little increase in computational complexity.

2. The McNabb and Foster model in one finite space dimension

The model under consideration here (and also, it is to be inferred, in the paper by Caskey and Pillinger [2]), refers physically to diffusion of hydrogen through a metal plate of uniform thickness a, with plane end faces, under a constant concentration K of hydrogen at the input face, zero at the output face. No account is taken of possible "edge effects" where the plate is clamped into the containing vessel, so that the plate is most usefully thought of as thin (say, a metal foil), with a small compared to the dimensions that characterize the cross-section parallel to the end faces. This cross-section is assumed to be simply connected, but no other specific assumptions are

 0022–2461/81/113122–9$02.90/0

made as to its shape; in view of the simplifying assumptions of the model, however, a simple shape such as a circle or rectangle would seem the most appropriate.

The McNabb and Foster model, hereafter referred to as the MF model, is a system of six equations. Denoting physical distance and time by X, T, where $0 \leqslant X \leqslant a$ and $T \geqslant 0$, let: $C(X, T)$ denote the number of atoms per unit volume (of metal) of diffusing hydrogen; N be the (constant) number of traps per unit volume of metal; $v(X, T)$ be the fraction of these traps occupied by a captured hydrogen atom; D be the diffusion coefficient (in $\mathrm{cm}^2\ \mathrm{sec}^{-1}$); κ be a parameter (of dimension L^3T^{-1}) such that the number of hydrogen atoms captured per second by traps in a volume element δV is $\kappa CN(1-v)\delta V$; and ρ (in sec^{-1}) be the reciprocal of the mean time-of-stay of a captured atom in a trap. Then the driving equations are

$$\frac{\partial C}{\partial T} + N\frac{\partial v}{\partial T} = D\frac{\partial^2 C}{\partial X^2} \tag{1}$$

and

$$\frac{\partial v}{\partial T} = \kappa C(1-v) - \rho v \tag{2}$$

and both are obeyed on $(0, a) \times \mathbb{R}^+$.

The boundary conditions for C correspond to

$$C(0, T) = K, \qquad T \geqslant 0, \tag{3}$$

and

$$C(a, T) = 0, \qquad T \geqslant 0, \tag{4}$$

where K is a constant. The initial conditions (corresponding, in the physical model, to vacuum in both the input and output chambers for $T < 0$) are

$$C(X, 0) = 0 \qquad (0 < X \leqslant a) \tag{5}$$

and

$$v(X, 0) = 0 \qquad (0 < X < a). \tag{6}$$

Clearly these six equations define a well-posed problem, given the assumption that the parameters N, D, κ, ρ, K are all constant. Caskey and Pillinger [2], however, and Frank *et al.* [3], add two further equation giving boundary conditions for v, impossible to justify or even comprehend either for the physical model or for a mathematical system devoid of any X-derivative (let alone a second X-derivative) of v.

McNabb and Foster [1] propose two dimensionless forms of their model in both of which $X = ax$, $T = (a^2/D)t$ (where $x \in [0, 1]$, $t \geqslant 0$ represent reduced distance and time) and $C(X, T)$ is replaced by $Ku(x, t)$. In the first of these, $v(X, T)$ is replaced by $v(x, t)$, yielding what we shall now call the MFC system:

$$\left.\begin{aligned} \frac{\partial u}{\partial t} + \beta\frac{\partial v}{\partial t} &= \frac{\partial^2 u}{\partial x^2} \qquad (7)\\ \frac{\partial v}{\partial t} &= vu - \mu v - \nu uv \qquad (8) \end{aligned}\right\} \quad (x, t) \in (0, 1) \times \mathbb{R}^+$$

$$u(0, t) = 1 \qquad t \geqslant 0 \tag{9}$$

$$u(1, t) = 0 \qquad t \geqslant 0 \tag{10}$$

$$u(x, 0) = 0 \qquad 0 < x \leqslant 1 \tag{11}$$

$$v(x, 0) = 0 \qquad 0 < x < 1, \tag{12}$$

where $\beta = N/K$, the number of traps per input hydrogen atom, $\mu = (a^2/D)\rho$, and $\nu = (a^2/D)\kappa K$, are dimensionless parameters.

For reasons that we shall enlarge on below, this is certainly the easier dimensionless form to work with. Published numerical work to date, however, has concentrated on the second form, in which $\beta v(X, T)$ is replaced by $w(x, t)$, so that the parameter β does not appear, but the parameter

$$\lambda = \beta\nu \tag{13}$$
$$= \frac{a^2}{D}\kappa N$$

is introduced. Equations 3 and 4 thus become

$$\frac{\partial u}{\partial t} + \frac{\partial w}{\partial t} = \frac{\partial^2 u}{\partial x^2} \tag{14}$$

and

$$\frac{\partial w}{\partial t} = \lambda u - \mu w - \nu uw \tag{15}$$

on the same domain as above, and w replaces v in Equation 12. We defer detailed consideration of why the first of these dimensionless formulations is preferable for numerical work to later paragraphs. It is clear, however, from $\lambda = \beta\nu$, firstly, that two of the three parameters β, λ and ν must be set independently, and secondly, that neither formulation restricts the choice of the pair on which attention will be focussed. Since the MF model takes account of three time constants

(a^2/D, $1/\rho$ and $1/K\kappa$) involved in hydrogen diffusion, there is a good argument in favour of selecting β, μ and either λ or ν for the determining parameters; these dimensionless quantities reflect the physical parameters N, ρ and κ with the greatest fidelity.

The second dimensionless formulation is often preferred on physical grounds, because w is a dimensionless scaling of the concentration of trapped hydrogen, and so is a quantity of the "same kind" as the dimensionless scaling of the concentration of free hydrogen, u. There are, however, obstacles to the physical interpretation even of the second formulation. Of the four parameters involved in the two formulations, β alone has a direct physical meaning; the other three are all products of a^2/D with some quantity of dimension $(\text{time})^{-1}$, and so depend, for their physical interpretation, on the choice both of the metal and of the thickness, a, of the metal specimen.

The MF system thus presents two quite distinct problems: one of efficient and reliable numerical solutions of a non-linear multiparameter system, the other of physical interpretation of the computed output. We have opted, in this paper, not to prejudge the choices the reader may wish to make, whether of D, of a, or of the determining parameters. In a later paper and in the context, also, of more complicated MF models, we hope to present detailed physical interpretations based on actual ranges of values of D and of a. Here we are only concerned with presenting an efficient method of arriving at numerical values whose reliability can be estimated, so we have chosen specific parameter values primarily to illustrate the points we make in the analysis of our scheme.

3. Numerical solution of the MFC system

Before detailing our scheme, and to avoid confusion due to the unusual use, in this context, of the word "convergence" by Caskey and Pillinger [2], we briefly review the chief factors to be looked for in any finite difference replacement of a system of differential equations. The first is consistency: that is, the difference equations must converge to the differential equations as the mesh constants h, k tend to 0 (not necessarily independently; in particular, we shall require the mesh ratio r to remain fixed).

Given consistency, one asks for stability and convergence. Stability is independent of the differential equations, and requires that the difference between the theoretical and computed solutions to the numerical scheme remain bounded as $n \to \infty$. Convergence, on the other hand, is independent of the computed values, and requires that the theoretical solutions to the numerical scheme tend to the theoretical solution of the differential equations as the mesh is refined in a prescribed manner.

Given all these, one asks for accuracy and, beyond that, for a scheme that is efficient, i.e. uses the minimum computing time and storage.

Turning now, to the finite-difference replacement of the MFC system, let h denote the x-step, k the time-step, and $r = k/h^2$ the mesh ratio. Further, let u_m^n, v_m^n denote $u(mh, nk)$, $v(mh, nk)$, where $m = 0, 1 \ldots, M = 1/h$ and $n = 0, 1 \ldots$. Approximating in the usual way, $(\partial y/\partial t)|_{x=mh}$ by $(y_m^{n+1} - y_m^n)/k$, $(y = \mathrm{u}, v)$ and $(\partial^2 u/\partial x^2)|_{t=nk}$ by $(u_{m-1}^n - 2u_m^n + u_{m+1}^n)/h^2$, a direct replacement of the MFC system leads to a computational system of extremely restricted stability, in which r must be confined to $(0, \rho]$ with $\rho < \frac{1}{2}$. To attempt to extend the range of admissible values of r, one uses the device first introduced by Crank and Nicolson [4], of "averaging" the right-hand sides of Equations 7 and 8 over the time levels $(n+1)k$ and nk. Caskey and Pillinger [2] went further, introducing an extra parameter θ "that allows admixture both forward and backward in time"; this, it turned out, meant that $(\partial^2 u/\partial x^2)$ was replaced by

$$[\theta(u_{m-1}^{n+1} - 2u_m^{n+1} + u_{m+1}^{n+1}) + (1-\theta)(u_{m-1}^n - 2u_m^n + u_{m+1}^n)]/h^2,$$

where $0 \leqslant \theta \leqslant 1$. It is, however, easy to show that with $\theta = \frac{1}{2}$ the local truncation error for the replacement of Equation 4 is $0(k^2)$, but that it deteriorates to $0(k)$ if $\theta \neq \frac{1}{2}$ is used, needlessly leading to a substantial loss of accuracy. For this reason, we do not introduce θ as a parameter to our scheme, and use the approach of Crank and Nicolson [4] unamended.

The replacements of Equations 7 and 8 are thus

$$u_m^{n+1} - u_m^n + \beta(v_m^{n+1} - v_m^n) = \frac{r}{2}(u_{m-1}^{n+1} - 2u_m^{n+1} + u_{m+1}^{n+1}) + \frac{r}{2}(u_{m-1}^n - 2u_m^n + u_{m+1}^n) \qquad (16)$$

and

$$v_m^{n+1} - v_m^n = \frac{k}{2}(\nu u_m^{n+1} - \mu v_m^{n+1} - \nu u_m^{n+1} v_m^{n+1}) + \frac{k}{2}(\nu u_m^n - \mu v_m^n - \nu u_m^n v_m^n) \quad (17)$$

for $m \in \mathscr{M} = \{1 \ldots M-1\}$, and $n \geqslant 0$.

Now, the main computational difficulty of the MF system resides in the non-linearity of Equation 2. Given any boundary and initial conditions, the system driven by Equations 16 and 17 would normally be solved as follows:

At time level $(n+1)k$, u_m^{n+1} is replaced by u_m^n on the right-hand side of Equation 17, and Equation 17 is solved for v_m^{n+1}; this approximation to v_m^{n+1} is substituted into Equation 16 to yield an approximation to u_m^{n+1}; this value of u_m^{n+1} is substituted into Equation 17 and the process repeated until requisite degrees of agreement are attained between successive iterates both of u_m^{n+1} and of v_m^{n+1}. This iterative process is extremely time-consuming and expensive; to avoid it one linearizes Equation 17. Accordingly, replacing $u_m^{n+1} v_m^{n+1}$ by the product of the first-order Taylor expansions, Equation 17 becomes

$$v_m^{n+1} - v_m^n = \frac{k}{2}[\nu u_m^{n+1} - \mu v_m^{n+1} - \nu(u_m^{n+1} v_m^n + u_m^n v_m^{n+1})] + \frac{k}{2}(\nu u_m^n - \mu v_m^n) \quad (18)$$

after simplification. This is the most direct and trouble-free linearization (it corresponds exactly to that used by Caskey and Pillinger [2]).

The boundary conditions on u are replaced by

$$u_0^n = 1, u_M^n = 0 \qquad (n = 0, 1, \ldots) \quad (19)$$

and the initial conditions on u and v by

$$u_m^0 = v_m^0 = 0 \qquad (m \in \mathscr{M}). \quad (20)$$

(Note that there are no boundary conditions on v!)

Presented in a manner that may facilitate comparison with the Caskey and Pillinger [2] scheme, Equations 16 and 18 yield (for $m \in \mathscr{M}$ and $n \geqslant 0$) the relations

$$-\frac{r}{2} u_{m-1}^{n+1} + K_m u_m^{n+1} - \frac{r}{2} u_{m+1}^{n+1} = C_m, \quad (21)$$

and

$$v_m^{n+1} = R_m u_m^{n+1} + S_m \quad (22)$$

where

$$R_m = z_m / d_m,$$

$$d_m = 2 + k(\mu + \nu u_m^n),$$

$$z_m = k\nu(1 - v_m^n),$$

$$K_m = 1 + r + \beta R_m,$$

$$C_m = \frac{r}{2}(u_{m-1}^n + u_{m+1}^n) + (2 - K_m) u_m^n + \frac{2k\beta\mu v_m^n}{d_m},$$

and

$$S_m = \frac{(2 - k\mu) v_m^n + k\nu u_m^n}{d_m}$$

Using Equation 19, Equation 21 may be solved by the recurrence relation

$$u_m^{n+1} = A_m u_{m-1}^{n+1} + B_m \qquad (m = 1 \ldots M) \quad (23)$$

whence Equation 22 may be evaluated directly. Substituting $A_{n+1} u_m^{n+1} + B_{m+1}$ for u_{m+1}^{n+1} in Equation 21 yields

$$A_m = \frac{r}{2K_m - rA_{m+1}}, \qquad B_m = \frac{2C_m + rB_{m+1}}{2K_m - rA_{m+1}} \qquad (m \in \mathscr{M}) \quad (24)$$

with, by Equation 19, $A_M = B_M = 0$.

4. Consistency, stability and accuracy

The scheme is certainly consistent; the actual local truncation errors (which we shall need to refer to in the discussion of accuracy) are

$$\frac{1}{12}\left(k^2 \frac{\partial^4 u}{\partial x^2 \partial t^2} - h^2 \frac{\partial^4 u}{\partial x^4}\right) \quad (25)$$

for the replacement of Equation 1, and

$$-\frac{k^2}{12}\left[\nu \frac{\partial^2 u}{\partial t^2} - \mu \frac{\partial^2 v}{\partial t^2} - \nu\left(v \frac{\partial^2 u}{\partial t^2} + u \frac{\partial^2 v}{\partial t^2}\right)\right] \quad (26)$$

for the replacement of Equation 2, the replacement of the initial and boundary conditions being error-free.

It is not possible to discuss stability and convergence independently for a non-linear system; nor is there a method, to date, of ascertaining necessary and sufficient conditions for the convergence of the replacement of a non-linear

system with an unknown analytic solution. This brings us, however, to the chief advantage of the MFC formulation, namely that both $u(x, t)$ and $v(x, t)$ must lie in $[0, 1]$ for all $(x, t) \in (0, 1) \times \mathbb{R}_0^+$. Consequently, it is possible, in place of convergence, to make the minimal demand on the numerical scheme, that

$$u_m^n \text{ and } v_m^n \text{ also lie in } [0, 1] \text{ for all } m \in \mathscr{M} \text{ and } n \geqslant 0. \tag{27}$$

Assuming, first, that our scheme satisfies Condition 27, and supposing $\beta > 0$, application of the von Neumann method [5] of stability analysis shows that the scheme is unrestrictedly stable if $\nu = 0$, and that, for $\nu < 0$, local stability is guaranteed if

$$k\nu\,[u_m^{n+1} - u_m^n - \beta(v_m^{n+1} - v_m^n)] \leqslant 4 \tag{28}$$

for $m \in \mathscr{M}$. Given Condition 27, this condition will certainly hold if

$$k \leqslant \frac{4}{\nu(1+\beta)}. \tag{29}$$

Examination of the scheme then shows that Condition 27 can be guaranteed if

$$0 < r \leqslant 1, \tag{30}$$

$$k(\nu - \mu) \leqslant 2 \tag{31}$$

and

$$1 - r + \frac{\frac{1}{2}k\beta\nu(\frac{1}{2}k\mu - 1)}{(1 + \frac{1}{2}k\mu)^2} > -\epsilon/M \tag{32}$$

where ϵ is the working tolerance.

Given $r \leqslant 1$ and Condition 32, but not Condition 31, both u_m^n and v_m^n may exceed 1, though this would not necessarily be evident if only the relative flux is printed out; u_1^n and v_1^n, in particular, should be tested in this case, to guard against persistent error.

Given $r \leqslant 1$ and Condition 31, but not Condition 32, u_m^n may be less than 0 for some $m < M - 1$. The intrusion of ϵ, here, into a condition for a convergence-related property, is caused by the substitution of 0 for several very small quantities into the true condition in which the right-hand side is 0; this is of negligible importance. Far more serious is the fact that Condition 32 is the most restrictive of the conditions we have found, and that testing $u_m^n \geqslant 0$ for a range of values of m is extremely expensive.

The argument leading to the results just reported may be outlined as follows:

For $m \in \mathscr{M}$ and $\beta > 0$, if $c_2 \geqslant c_1 \geqslant 0$ and $E_1 = 1 + r + c_1 \leqslant K_m \leqslant 1 + r + c_2 = E_2$, it follows that

$$F_m \leqslant A_m \leqslant G_m \tag{33}$$

where

$$G_m = \frac{r}{s} \cdot \frac{1 - (r/s)^{2(M-m)}}{1 - (r/s)^{2(M-m+1)}}$$

and $s = E_1 + [(E_1)^2 - r^2]^{1/2}$; F_m is similarly defined with $\sigma = E_2 + [(E_2)^2 - r^2]^{1/2}$ replacing s. If, further

$$P \leqslant C_m \leqslant Q \tag{34}$$

then

$$P_m \leqslant B_m \leqslant Q_m \tag{35}$$

where

$$Q_m = \left(\frac{2Q}{s - r}\right) \cdot \left[\frac{1 - (r/s)^{M-m}}{1 + (r/s)^{M-m+1}}\right]$$

and P, σ replace Q, s, respectively, in the definition of P_m. From these results it follows that, if

$$Q \leqslant Q^* = (1 + c_1)\left\{1 + \frac{\left(\frac{r}{s}\right)^{M-1}\left(1 + \frac{r}{s}\right)}{\left[1 - \left(\frac{r}{s}\right)^{M-1}\right]\left[1 - \left(\frac{r}{s}\right)^{M}\right]}\right\} \tag{36}$$

and

$$P \geqslant P^* = -(1 + c_2)\left\{\frac{\left(\frac{r}{\sigma}\right)^{M-1}\left(1 + \frac{r}{\sigma}\right)}{\left[1 - \left(\frac{r}{\sigma}\right)^{M-1}\right]\left[1 - \left(\frac{r}{\sigma}\right)^{M}\right]}\right\} \tag{37}$$

then $u_m^{n+1} \in [0, 1]$ for all $m \in \mathscr{M}$ and $n \geqslant 0$.

Application of Equations 36 and 37 to the special case, $\nu = 0$, of unrestricted stability, leads to Condition 30. Since Condition 27 must hold, for sufficiently small h and k, if the scheme is convergent, and convergence is clearly impossible if we cannot guarantee Condition 27 in the $\nu = 0$ case (when v remains identically zero), we assume that r satisfies Condition 30 for the remainder of the argument.

With $\beta\nu > 0$, account now needs to be taken of the dependence of v_m^n on u_m^n; that is, $C_m - \frac{1}{2}r(u_{m-1}^n + u_m^n)$ must be expressed as a function, f say, of u_m^n alone in order to determine the bounds, P and Q, on C_m. This leads to a step by step argument. It is easily verified that $u_m^1 \in (0, 1)$ and $v_m^1 > 0$, by the results already established; $v_m^1 \leqslant 1$ if Condition 31 is satisfied. To ensure $u_m^2 \geqslant 0$, $f'(u_m^1)$ must be non-negative for all $m \in \mathscr{M}$, leading to Condition 32, and Condition 31 then ensures that none of the u_m^2 or v_m^2 exceed 1. These conditions are then shown to ensure that $u_m^2 \geqslant u_m^1$ and $v_m^2 \geqslant v_m^1$, so that the same conditions apply at each subsequent time step.

5. Accuracy and efficiency

One way of improving the accuracy of a finite difference scheme is to reduce the mesh spacings h and k, keeping the mesh ratio r fixed. With $r = k/h^2$, halving h thus means quartering k, and the amount of computation can rise dramatically if $1/h$ is already large. An alternative is to use Richardson's method of deferred approach to the limit [5]. Here we assume that the local truncation error, i.e. the difference between the analytical solution u and the finite difference solution U, can be expressed as a power series in one of the mesh constants.

Applied to our scheme, for which the local truncation error is $0(h^2 + k^2)$, this gives

$$u = U + e_1h^2 + e_2h^4 \ldots + e_jh^{2j} \ldots, \quad (38)$$

where $e_1, e_2 \ldots e_j$ are the coefficients of the expansion and we assume

$$|e_j| \geqslant |e_{j+1}| \qquad (j \geqslant 1). \quad (39)$$

If the finite difference scheme is solved using $h = 2H$ as the step length in the x-direction, then

$$u = U_1 + 4e_1H^2 + 16e_2H^4 \ldots \quad . \quad (40)$$

repeating the computation with $h = H$ and r unchanged gives

$$u = U_2 + e_1H^2 + e_2H^4 \ldots \quad . \quad (41)$$

Elimination of e_1 between these two equations leads to

$$u = \tfrac{1}{3}(4U_2 - U_1) + 0(H^4), \quad (42)$$

so that, if Condition 39 is justified,

$$\tfrac{1}{3}(4U_2 - U_1) \quad (43)$$

is a better approximation to the true solution u. Depending on the size of the derivatives in Equations 25 and 26, the use of this method can make quite small values of M acceptable.

A further improvement in accuracy is obtained if the singularity, at (0, 0), of the MFC system is taken into account. Since the numerical solution of any system can, at best, tend only to the theoretical solution, the perturbation caused by the discontinuity of the boundary at the origin can be appreciably reduced if $u(0, 0)$ is replaced by the expected value, $\frac{1}{2}$, of a Fourier series approximating u, at the point singularity (see also Crandall [6]).

To compare the efficiency of our scheme with that of Caskey and Pillinger [2] we wrote a special program, taking the same care there as in our own program to avoid repeated accessing of the same element of a vector as far as practicable. Trial runs using coarse mesh spacings of $h = 1/20$, $k = 1/800$ indicated that the run-time for our scheme is just under 80% of the Caskey and Pillinger [2] scheme. Six vectors were needed for our scheme, compared with eight for the Caskey and Pillinger scheme.

Now, whilst a 20% saving in run-time and a concomitant saving in storage are certainly useful, they still do relatively little to bring down the expense of each run if mesh spacings such as the $h = 1/100$ and $k = 1/20\,000$ of Caskey and Pillinger [2] have to be used. Use of Richardson's method leads to a major improvement in efficiency [5].

6. Behaviour of the scheme in practice

Figs 1 and 2 give examples of curves obtained using our scheme together with the deferred approach to the limit method and the device of setting $u(0, 0)$ to $\frac{1}{2}$ (relative flux = (output flux at time t)/equilibrium flux). Fig. 3 shows the corresponding curves for u and v when equilibrium (steady state) is attained. Although in Fig. 3 the v_{eq} curves appear to have been plotted on [0, 1] they are, in fact, plotted on [0.002, 0.998]. Note the considerable distortion that would result from the Caskey and Pillinger condition $w(0, t) = \beta v(0, t) = 0$. To discuss all these, we begin by reporting our tests of the validity of using the Richardson method.

The sizes of the derivatives in Equations 25 and 26 are much influenced by the values of the parameters β, μ and ν, and the laws governing these dependences appear to be complicated. So it is

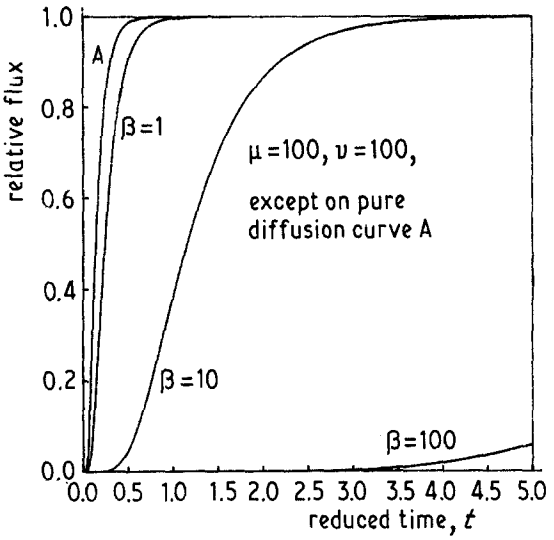

Figure 1 Relative flux against reduced time.

vital to test Condition 39 as there may be regions of the parameter space in which the increased contribution from the e_2 term in Equation 38, occasioned by the use of Richardson's method, outweights the elimination of the e_1 term. Two straightforward test methods are available.

The first, which can be used at any arbitrarily fixed point of the parameter space, consists of running the program with, for example, $M = 20$, 40 and 80 in turn (keeping r constant), and comparing the results of applying Equation 43 to the output from the first two with the output from the third. This is an expensive test, as the run-time for $M = 80$ is over four times the total required to run first with $M = 20$ then with $M = 40$ and finally applying Equation 43 to the results.

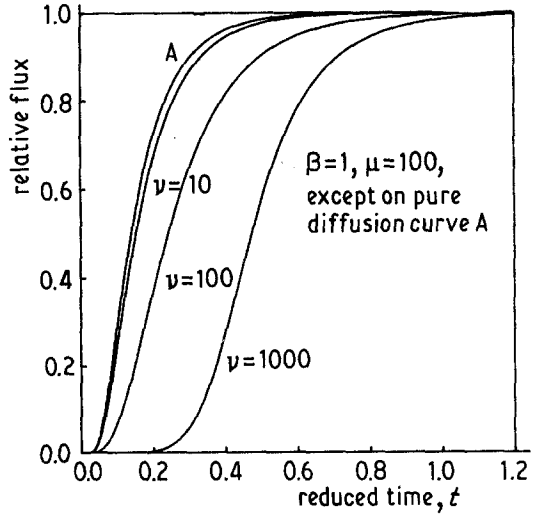

Figure 2 Relative flux against reduced time.

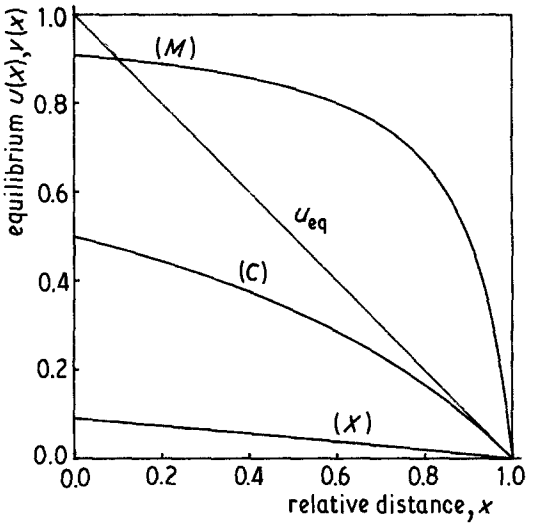

Figure 3 Equilibrium $u(x)$ and $v(x)$ against relative distance. For v_{eq} curves $\beta > 0$, $\mu = 100$, $\nu = 10(X)$, 100 (C), 1000 (M).

The second method combines a test of Condition 39 with a numerical test of convergence. This is relatively inexpensive, but can only be used on or near bounding planes of the parameter space.

To test convergence numerically rather than theoretically, one or more of the parameters can be set to zero (or allowed to approach 0), so that the system of differential equations reduces to an allied linear problem which can be solved analytically. The seemingly most attractive test, therefore, would be to run the program with $\beta = \mu = \nu = 0$, and compare the output with the relative flux,

$$1 + 2 \sum_{n=1}^{\infty} (-1)^n \exp(-n^2 \pi^2 t) \qquad (44)$$

for the corresponding pure diffusion system. Unfortunately this is an insensitive test, of debatable value even as a program "de-bugging" device.

Fortunately, a simple and much more sensitive test is available. For, with $\nu = 0$, but both $\beta > 0$ and $\mu > 0$, a case that corresponds in the MFC formulation to non-zero input, non-zero trap concentration and release rate, but "non-trapping" traps, it is easily proved that v must remain identically zero and the MFC system again reduces to its corresponding pure diffusion problem. It should be noted that this test is not available in the dimensionless formulation driven by Equations 14 and 15; on the other hand, the complicated test made by Caskey and Pillinger [2], with $v = 0$ and $\lambda\mu > 0$, corresponding there to zero input, is not available in the MFC formulation,

where it would require infinite β; the simplicity of Equation 44 compared to the analytical solution for the $\lambda\mu > 0$, $\nu = 0$ case is, we contend, yet another reason for preferring the MFC formulation.

Calculating Equation 44 to the same working tolerance, $\epsilon = 5 \times 10^{-5}$, to which we calculate the relative flux in our programs, we find, independently of the positive values chosen for β and μ, no deviation exceeding 4×10^{-5} between the output from our program and the analytical solution, provided that r is kept to $(0, 1]$. This is illustrated in Fig. 4; v remains indistinguishable from 0 to the maximum accuracy of the computer. Referring back to a point made earlier, it is worth adding here that the deviation graphs shown bear scant resemblance to those obtained if $u(0, 0)$ is kept at 1. An experiment with $r = 1/2$ revealed, in place of the well-behaved curve shown in Fig. 4, a curve that oscillated rapidly, with peaks of the order of $\pm 7 \times 10^{-5}$ for $t < 0.12$, and still of $\pm 3 \times 10^{-5}$ near $t = 0.35$ (where the experiment was terminated). The MFC system, it seems, is a further example of the high sensitivity to coarse mesh spacings reported by Crandall [6] for parabolic equations with 0–1 singularity at $(0, 0)$.

With the validity of Equation 39 established, we next explored the necessity, in practice, of the three conditions to guarantee that all u_m^n and v_m^n lie in $[0, 1]$.

First, in the $\nu = 0$ test already described, we tested r values up to 2. Now, as Fig. 4 shows, accuracy decreases as r increases: with $r = 2$ the maximum deviation from Equation 44 does, however, exceed 3ϵ, and the corresponding curve could not be included in Fig. 4. There are thus two valid reasons for restricting r to $(0, 1]$: either to guarantee a convergence-related condition or to maintain an acceptable order of accuracy.

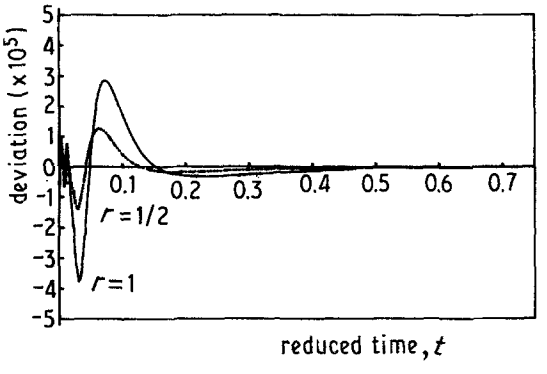

Figure 4 Deviation, for r values shown, of relative flux for $\nu = 0$ and $\beta\mu > 0$ from pure diffusion.

The points on the three curves with $\mu = \nu = 100$ in Fig. 1 were all calculated using $r = \frac{1}{2}$ and $M = 20$ and 40. Only for the $\beta = 1$ curve is Condition 32 satisfied for both M values; for the $\beta = 10$ curve, Condition 32 is satisfied only when $M = 40$; and for the $\beta = 100$ curve, Condition 32 is not satisfied at all. To reduce testing costs, we decided to examine the equilibrium values of u and v, arguing that any persistent error must perturb these, rather than test $u_m^n \geqslant 0$, We found agreement, well within our working tolerance, with the theoretical values

$$u_{eq}(x) = 1 - x \tag{45}$$

and

$$v_{eq}(x) = \frac{\nu u_{eq}(x)}{\mu + \nu u_{eq}(x)} \tag{46}$$

It is interesting to note about the formula for v_{eq} firstly that it is independent of the value of $\beta < 0$, secondly that is is invariant under simultaneous magnification of μ and ν by the same factor; since the ratio ν: $\mu = K\kappa$: ρ, this observation may be relevant to the very brief discussion by Caskey and Pillinger [2] of their Fig. 3c, but this is difficult to ascertain in view of the lack of information concerning the interpretation of their expression $u_0 + w_0$.

We have illustrated the formula for v_{eq}, in Fig 3, for all the sets of parameter values occurring in Figs 1 and 2; since $|v_{corr}(i/20) - v_{eq}(i/20)|$ was of the order of 1×10^{-6} or less for $i = 1 \ldots 19$ (where $v_{corr}(x)$ is the value obtained from the final v vectors at x after applying the Richardson method to the $h = 1/20$ and $h = 1/40$ runs, and $v_{eq}(x)$ is obtained from the formula above), all 19 computed equilibrium points do lie on the relevant curve for each set of parameters with $\beta \leqslant 10$ (the $\beta = 10$ run in Fig. 1 was terminated before reaching equilibrium).

Finally, with reference to stability, since Condition 29 is a sufficient, but not necessary, condition for Condition 28, we included in the $r = \frac{1}{2}$, $M = 20$ run for the $\nu = 1000$ curve of Fig. 2, a direct test of Condition 28; the upper bound of 4 was nowhere exceeded.

7. Adaptability of the scheme

As an instance of the ease with which our scheme can be adapted, we detail here how it may be altered to replace the dimensionless form driven by Equations 14 and 15 in Equation 22:

w replaces v;
z_m becomes $k(\lambda - \nu w_m^n)$;
β disappears from the expression for K_m;
the final term of the expression for C becomes $2k\mu w_m^n/d_m$;
w replaces v, λ replaces ν in the numerator of S_m.

More valuably, we have successfully used the scheme to investigate MF systems with more complicated, time-dependent boundary conditions. The results of these investigations will be detailed in a subsequent paper; it is, however, worth reporting here that these MF systems yield extremely sensitive tests of (numerical) convergence if any one of the parameters β, μ, ν is allowed to approach 0, the other two parameters remaining fixed and positive. The results of these tests strongly suggest that the scheme is convergent, and show that Condition 39 holds on all three bounding planes of the parameter space.

Finally, we have successfully extended our scheme to MF systems with x-dependent N. In view, however, of the assertion that "... it is possible to incorporate a time variation of N ... into the computer program" in the paper by Frank *et al.* [3], we should mention that, whilst this is certainly possible, it is by no means acceptable using our scheme or any other. With a time-dependent N the argument that led McNabb and Foster to Equation 1 now leads to a new model. In this, N must have a continuous time-differentiable time derivative, and Equations 7 and 8 become

$$\frac{\partial u}{\partial t} + \beta\frac{\partial v}{\partial t} + v\frac{\partial \beta}{\partial t} = \frac{\partial^2 u}{\partial x^2} \qquad (47)$$

and

$$\beta\frac{\partial v}{\partial t} + v\frac{\partial \beta}{\partial t} = \nu\beta u - \mu\beta v - \nu\beta uv. \qquad (48)$$

A third driving equation is evidently needed if β is not a known function of t.

The finite-difference replacement of each time-dependent-β model presents its own problems, depending on the assumptions made about β, and it is no longer suitable to speak of a mere "extension" of our scheme. Our experience to date has shown that great care needs to be exercised to keep the local truncation error sufficiently small, and further analysis of the resulting scheme becomes so difficult that it is best replaced by tests, using the appropriate MF (i.e. constant β) system as a test-bed.

8. Conclusions

Using a dimensionless formulation of the McNabb and Foster model that is better adapted to numerical work than the formulation used in previously published studies, we have been able to:

(1) Present a more efficient scheme, which, using simple techniques, yields more accurate results using approximately 20% of the computer time, and using less than one-third the storage of the method recommended by Caskey and Pillinger [2];

(2) Present bounds on computation parameters which guarantee results with a measure of reliability;

(3) Suggest a variety of tests that may be incorporated into the program where the above-mentioned bounds are not practicable.

The scheme may be adapted to a variety of conditions, and we have presented driving equations that may be used if the trap concentration depends on time.

Acknowledgements

This work was carried out whilst E. J. Stern was in receipt of an Open University Research Studentship. We wish also to thank Dr D. F. Mayers of Oxford University and Dr. R. M. Bromilow of the Open University for valuable discussions, and the Oxford University Computing Service for allowing us the use of their facilities.

References

1. A. MCNABB and P. K. FOSTER, *Trans. Metal. Soc.* **227** (1963) 618.
2. G. R. CASKEY Jr and W. L. PILLINGER, *Metall. Trans. A* **6A** (1975) 467.
3. R. C. FRANK, C. W. WERT and H. K. BIRNBAUM, *ibid.* **10A** (1979) 1627.
4. J. CRANK and P. NICOLSON, *Proc. Camb. Phil. Soc. Math. Phys. Sci.* **43** (1947) 50.
5. L. FOX, "Numerical Solution of Ordinary and Partial Differential Equations" (Pergamon Press, Oxford, 1962).
6. S. H. CRANDALL, *J. Ass. Comput. Mach.* **2** (1955) 42.

Received 5 December 1980 and accepted 27 April 1981.