A comparison of the finite difference and finite element methods for heat transfer calculations

A. F. Emery
H. R. Mortazavi
University of Washington
Seattle, Washington

Introduction

Of the many approximate or numerical methods used to solve heat transfer problems, the finite difference and finite element approaches have become the most widely known and used. The finite difference method (FDM) is the older, being introduced in 1928[1], and until recently, probably the better known and more extensively used for both research and production studies. The finite element method (FEM) is newer[2] and has not been used for heat transfer studies to a comparable degree, although it is the dominant method for structural analyses.

Because the FEM has a short history of use for heat transfer, there is considerable confusion about the relative values of the two methods and under which conditions one method is to be preferred. Both methods are based upon minimizing the weighted error in satisfying the first law of thermodynamics. Two common FDM's are: the heat balance FDM (BFDM) [3], based upon minimizing, over a nodal volume with unit weight; the mathematical FDM (MFDM) [4,5], based upon collocating at discrete points, usually regularly spaced upon coordinate lines. The FEM is based on the Galerkin method [6] in which the error is minimized over an elemental volume with a weight function equal to the temperature.

The purpose of this paper is to compare the two methods by describing their bases and their application to some common heat transfer problems. It will be no surprise that neither method is clearly superior or that, in many instances, the choice is quite arbitrary and depends more upon the codes available and upon the personal preference of the analyst than upon any well defined advantages of one method.

Analytical Procedures

The numerical solution of a heat transfer problem by either FDM or FEM, is usually done in the following steps as shown in table 1:

*This work was supported by NASA grant NAG-1-41
TABLE 1
ORDER OF ANALYSES

1. Subdivision of the region into nodes or elements
2. Definition of material properties, initial temperatures and boundary conditions
3. Evaluation of the thermal conductance between nodes and the capacitance of each node
4. Formation of the global conductance and capacitance matrices
5. Imposition of the boundary conditions
6. Solution for the nodal point temperatures and heat fluxes
7. Display of the results

Both methods tend to use these same steps, although the division between the steps may be more or less clearly defined depending upon the method used and upon the organization of the computer code.

A. The Mesh

The process of subdividing the region into elements or nodal volumes, commonly called 'meshing', is usually accomplished by a separate program. Although the FDM mesh was generally created by hand, the FEM programs relied, almost from their inception, upon numerical mesh generators. This difference in approach was probably due to:

1. FDMs usually utilize nodal points which are oriented along coordinate lines. Thus irregular regions are difficult to model automatically and use is often made of hand meshing with special imaginary nodal points. If the region is susceptible to treatment by a different coordinate system, the mathematical problem is usually recast. For example, elliptical or oblique coordinate systems are often used.
2. FEM codes were first used in structural problems where an organized nodal point mesh was clearly evident and an arbitrary mesh could not be used (e.g., the frame and stringer structures of aircraft[7]). By the time the FEM was applied to continuum problems, it had been observed that the solution of the Laplacian gave rise to an acceptable mesh. Because the FEM could be applied to such a wide range of geometries and because the number of elements tended to increase quickly, a need for automatic mesh generation became obvious. As a consequence an emphasis was placed upon automatic mesh generation for the FEM.

Several symposia[8] have been organized solely to discuss sophisticated methods for the FEM and for the FDM used in computational fluid dynamics. Unfortunately, few of the thermal FDM use other than the simplest mesh generators. Lately we find that most codes, whether FDM or FEM tend to use the same mesh generators. Since this mesh generation is also used to identify the regions occupied by the different materials, to define initial temperatures, as well as to prescribe the boundary conditions, the entire mesh generation process is usually referred to as 'PRE-PROCESSING'.
B. Thermal Conductance

The primary difference between the FDM and the FEM lies in the methods used to construct the thermal conductance between the nodes.

1. FDM

In the FDM, the conductance is formed in one of two ways. The simplest approach is illustrated in figure 1 and is based upon the heat balance equation

\[ \int \rho c \frac{\partial T}{\partial t} \, dv = \int q_n \, da + \int q' \, dv \]  

(1)

where the energy conducted through the surface is approximated by the linear relationship

\[ q_{n(i,j)} = q_{ij} = k_{ij} A_{ij} (T_j - T_i) / \ell_{ij} \]  

(2)

The quantity \( A_{ij} / \ell_{ij} \) is a geometric value which, in two dimensions, can be related to the angles of the triangle formed from the nodes[9]. Equivalently it can be found as illustrated in figure 1 by constructing the area \( A_{ij} \) which is perpendicular to the line joining the nodes and located at the mid point of the line. Although this method is difficult to implement, it can be done automatically and also provides a technique for determining the set of closest nodes which surround a node. Unless the nodes are regularly spaced, in which case the BFDM becomes the MFDM, the BFDM is of low accuracy, \( O(\Delta x) \).

The second method, the mathematical FDM (MFDM), is based upon expressing the differential equation

\[ \rho c \frac{\partial T}{\partial t} = \frac{\partial}{\partial x} \left( k_x \frac{\partial T}{\partial x} \right) + \frac{\partial}{\partial y} \left( k_y \frac{\partial T}{\partial y} \right) + q' \]  

(3)

by finite differences using a variety of different numerical approximations to achieve a desired level of accuracy. The method consists of establishing an interpolating function in space, usually a polynomial, and performing the needed differentiation. Another method is to express \( T_j \) in terms of \( T_i \) through a Taylor series expansion. By combining the expressions for the temperatures at different nodal points, i.e., establishing different stencils, figure 2, it is possible to achieve a very high order of accuracy. For example the 5 point stencil is accurate to \( O(\Delta x^2) \), while the 9 point is of \( O(\Delta x^4) \). However, the 13 point is accurate not only to \( O(\Delta x^4) \) but ensures that \( \nabla^4 T \) as well as \( \nabla^2 T - 0 \) is satisfied[10].

For a non-linear problem it is usual to express \( \frac{\partial}{\partial x} \left( k_x \frac{\partial T}{\partial x} \right) \) as

\[ \frac{\partial}{\partial x} \left( k_x \frac{\partial T}{\partial x} \right) = k_x \frac{\partial^2 T}{\partial x^2} + \frac{\partial k_x}{\partial x} \frac{\partial T}{\partial x} \]  

(4)

and to evaluate each term separately. Unfortunately this approach does not give a conservative set of equations (in P. D. Lax's sense) and energy is not conserved in steady state problems. As a consequence, the non-linear FDM is usually confined to the 5 point stencil for regular regions and the heat balance method for irregular regions.
2. FEM

The finite element method is based upon the Galerkin approach in which a stationary value of the integral

\[
I = \int_{V} \frac{1}{2} \left[ k_x \left( \frac{\partial T}{\partial x} \right)^2 + k_y \left( \frac{\partial T}{\partial y} \right)^2 - \rho c \left( \frac{\partial T}{\partial t} \right)^2 + 2q \cdot T \right] dv + \int_{A} q \cdot T dA \quad (5)
\]

is sought\[11\]. For steady state problems, the Galerkin and the variational approaches are equivalent and the method, in conjunction with the dual formulation (i.e., the adjoint problem), can be used to establish bounds upon the solution. For transient problems, there is no variational formulation and no bounds can be found.

The procedure consists of approximating the temperature within the element by

\[
T(x,y) = \sum_{L=1}^{I} n_L(x,y) a_L = < n > [ a ]
\]

(6)

where the interpolating functions, \( n(x,y) \), can be chosen at will to satisfy any desired criteria, subject only to the restriction that there be a sufficient number of nodes in the element to determine the coefficients, \( a_L \). Since each element is treated as an entity, it is common to map the element into a simpler shape to expedite the integration, figure 3. Consider that a local set of coordinates, \( \xi, \eta \), are defined and that the mapping

\[
x = \sum_{L=1}^{J} \bar{n}_L(\xi,\eta) b_L = < \bar{n} > [ b ]
\]

(7)

is used where the number of terms \( J \) need not equal the number of terms \( I \) used in the temperature interpolation. We then have

\[
\{ \hat{T} \} = [ A ] \{ a \} , \quad T(x,y) = < n > \quad [ A ]^{-1} \{ \hat{T} \} = < N > \quad \{ \hat{T} \}
\]

(8a)

\[
\{ \hat{\xi} \} = [ B ] \{ b \} \quad \xi = < \bar{n} > \quad [ B ]^{-1} \{ \hat{\xi} \} = < \bar{N} > \quad \{ \hat{\xi} \}
\]

(8b)

Thus for one element the conduction term is of the form

\[
\frac{\partial T}{\partial x} = < \frac{\partial N}{\partial x} > \{ \hat{T} \} = \begin{bmatrix} < \frac{\partial N}{\partial \xi} \frac{\partial \xi}{\partial x} > + < \frac{\partial N}{\partial \eta} \frac{\partial \eta}{\partial x} > \end{bmatrix} \{ \hat{T} \}
\]

(9a)

\[
\frac{\partial}{\partial \{ \xi \} } = \begin{bmatrix} k_x \left( \frac{\partial N}{\partial x} \right)^T \end{bmatrix} < \frac{\partial N}{\partial x} > [ J ] \frac{\partial d\xi d\eta}{\partial x} \{ \hat{T} \}
\]

(9b)

and

\[
\begin{bmatrix} \frac{\partial N}{\partial x} \\ \frac{\partial N}{\partial y} \end{bmatrix} = [ J ]^{-1} \begin{bmatrix} \frac{\partial N}{\partial \xi} \\ \frac{\partial N}{\partial \eta} \end{bmatrix}
\]

(10)

where \( J \) is the Jacobian of the mapping.
Although it is obvious that the spatial interpolation function, $\tilde{N}$, must ensure compatibility of elements (i.e., that an edge common to two elements must be the same when viewed from either element) it is not necessary that the temperatures be compatible. In structural mechanics FEM, incompatible formulations are sometimes used, but almost all thermal problems use compatible temperature functions. When the same number of functions are used to express the mapping and the temperature, the element is termed 'iso-parametric'. Although the isoparametric element is the usual, subparametric elements are often used for heat source problems or when an unusual temperature distribution is desired.

The most common 2-D elements are the 3 and 6 node triangles and the 4 and 8 node quadrilaterals, figure 4. Much of the research in the FEM has been devoted to developing new interpolating functions (i.e., new elements) and identifying their characteristics, particularly with respect to the numerical integration required.

3. Differences

Probably the most striking difference between the FDM and the FEM is the usual lack of continuity of the heat flux in the FEM. Each element is treated separately and if only compatibility of temperature is imposed, the heat flux at the edge common to two elements is not continuous. By contrast, the FDM has a continuous heat flux. Although the effect of lack of continuity of the heat flux is usually not important, it can show up in the form of an oscillating temperature profile where an overshoot in one node is compensated for by an undershoot at a neighboring node in order to minimize the integral over the entire region.

C. Capacitance Matrix

In the FDM, the capacity term, $\rho c \partial T/ \partial t$ is represented by the simple lumped term

$$\left( \rho_i c_i V_i \right) \frac{\partial T_i}{\partial t}$$

where $V_i$ is the volume surrounding the node $i$. In the FEM, the use of the interpolating function gives rise to the consistent capacitance matrix

$$[C] \left[ \frac{\partial T}{\partial t} \right] = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix} \{T\}$$

in which the storage at each node is related to that at every other node in the element. This difference in the two capacitance matrices is responsible for much of the difference in the solutions to transient problems and is discussed further in the EXAMPLE section. For 3 node triangular elements with regular node spacing, the FDM and FEM have identical conductance and capacitance matrices. Other FEM elements give different matrices, usually as indicated by the sign of the off-diagonal terms. In the FEM the integration is usually accomplished through Gaussian integration and the need to evaluate the properties and the Jacobian at each point contributes to the longer execution times of the FEM.
D. Assemblage of the Global Matrix

1. FDM

In the FDM, each node is treated in turn, and the conductances are entered into the global matrix as they are determined. The process is simple and straightforward and requires very little computer time. Unless an FEM preprocessor mesh is used, FDM codes generally make use of a separate matrix which identifies the nodes which interact, the corresponding value of the term $A_{ii}/\ell_{ii}$ and the material numbers to permit rapid determination of the average conductance $(k_i+k_j)/2$.

2. FEM

In the FEM, the element matrix is first evaluated, then assembled into the global matrix. If the global matrix cannot be stored in core, it is necessary to search through the elements to treat only those containing the nodes whose equations are currently in core. This assembly process tends to be reasonably time consuming and may contribute to the reduced speed of the FEM.

3. Comparison

Table 2 lists some typical execution times for FDM and FEM solutions. In general, the FEM assembly costs rise quickly as the number of nodes per element is increased. Figure 5[12] shows some comparable results for a structural problem and the extra expense of the assembling is easily seen. (The figure also shows that for some variables the FDM is superior while for others the FEM is. Because of this, it is difficult to define one method as being the best for all parameters of a problem.)

<table>
<thead>
<tr>
<th>n</th>
<th>L-W Explicit</th>
<th>C-N Linear</th>
<th></th>
<th>FDM Quadratic</th>
<th>FEM Cubic</th>
<th>Special Cubic</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.8</td>
<td>1</td>
<td>1.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.4</td>
<td>1.5</td>
<td>5.4</td>
<td>6.5</td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3.7</td>
<td>3.5</td>
<td>32</td>
<td>36</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td>12</td>
<td>24</td>
<td>249</td>
<td>183</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td>125</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td>540</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>107</td>
<td>64</td>
<td>1624</td>
<td>1880</td>
<td>1500</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td>4350</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>760</td>
<td>407</td>
<td>12200</td>
<td>14700</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n = number of elements  
L-W = Lax Wendroff  
C-N = Crank Nicolson
E. Boundary Conditions and Irregular Meshes

Although the greatest fundamental difference between the FDM and the FEM lies in the formulation of the conductances and capacitances, the greatest practical difference lies in their treatment of boundary conditions and irregularly shaped regions.

1. FDM

Specified temperature nodes are treated by ignoring the equation for the node. Boundary flux conditions are difficult to treat for irregular meshes unless one uses the simple heat balance formulation (BFDM) with its low order of accuracy. Irregular meshes have been studied for years with emphasis on using triangular surface elements[13]. Since many analysts tend to define their nodes along coordinate lines, it is most common to make use of imaginary nodes, figure 6, special differencing equations to achieve the desired accuracy[14], arbitrary grids[15], or even no nodes, but only cells[16]. In general, the advanced mesh generators used in fluid computations and the automatic high order boundary gradient differencing methods are not used in production FD methods. Consequently they suffer from reduced accuracy at irregular boundaries. On the other hand, research FDM's almost always maintain high accuracy at all types of boundaries.

2. FEM

Irregular regions are handled in the same manner as any other region. The accuracy is limited only by the ability of the mesh generator to represent the boundaries. This ability is probably the greatest strength of the FEM.

Specified temperature nodes are handled in two ways. In the first, the equation is deleted from the matrix and the right hand sides of all affected equations are modified. Another way is to replace the diagonal coefficient by a very large number, L, and the right hand side by L*T[17]. Although faster and simpler than the first method, the accuracy of the computer may limit the effectiveness of this approach.

Prescribed heat flux boundary conditions are handled better by FDM than by FEM. Because in the FEM each element is treated separately, the accuracy is limited by the interpolating accuracy of the element shape functions. By contrast, the FDM analyst can easily modify a difference equation to achieve the desired accuracy. The FEM analyst must call for a boundary element which is different from the interior element, figure 4, and may have some difficulty in merging this higher order element with the rest of the mesh unless a special mesh generator or slide lines are used.

F. Solving the Equations - Steady State Problems

Although sparse matrix solvers have become popular recently[18], most FDM and FEM codes use some form of elimination (usually Gauss or Cholesky) and rely upon a banded matrix or skyline approach to achieve fast solution times[19]. For highly non-linear problems, iterative solutions may be used, especially by FDM codes.
1. FDM

FDM codes normally use direct reduction or iteration. Because of the way in which the global matrix is assembled, iteration is a common procedure and, when used, the global matrix is not created. If reduction is used, only enough equations to treat the band are kept in core and they are shifted immediately after each nodal equation is reduced (the wave front method). Although iteration is an appropriate technique, convergence is difficult to determine and because iterative acceleration factors are often unknown, many users are reluctant to employ iterative techniques. Fortunately most heat transfer problems can be treated by using the SOR method and values of the over relaxation factor can be quickly approximated. For problems which are not symmetric or do not have Young's property A[20], the SSOR technique[21] is easy to implement and convergence is very rapid. Non-linear problems are ideally suited to the iterative method and rarely increase the solution time by an appreciable amount. Considerable research is still being conducted into improving the rate of convergence and treating highly non-linear problems[22].

2. FEM

Most FEM codes use a reduction process with a subsequent back substitution. Even with a skyline procedure, care should be taken to minimize the band width. To do this, most FEM codes will use a band width minimizing program after the mesh generation[23], although this may complicate the identification of the nodal point positions within the code. Non-linear problems are treated in two ways:

1. The matrices are reformed and solved for each iteration.

   \[ K^{(n)}T^{(n)} = R^{(n)} \]  

2. A residual is found, and an increment, \( \Delta T \) determined from

   \[ \{e\}^{(n)} = \{R\}^{(n)} - [K]\{T\}^{(n-1)} \]

   \[ \Delta T^{(n)} = [K]^{-1}\{e\}^{(n)} \]  

   (13b)

It is common to use a mixture of these two methods by applying the boundary loads in increments, calculating a new \( K \) for each increment, but iterating within the increment using a constant \( K \) until the residual is sufficiently small[24]. Method 1 requires excessive computer time and may not converge. Method 2 requires that the original conductance matrix be stored in order to calculate the residual. Either method is expensive, and for highly non-linear problems, the iterative FDM appears to be preferred. Suendermann [25] suggests that the ratio of the execution times for hydrodynamic problems is 100 to 27b^2/2+18b for implicit methods and 100 to 23bi for explicit methods where b=band width and i the number of FEM iterations.

In both methods it is often desirable to know the amount of heat added to a constant temperature node. In the FDM, the calculation of \( Q_i \) is simply done by calculating the conductances and evaluating the heat balance. In the FEM, this may be done by using the original \( K \) matrix if it is stored. A simpler way
is to permit a specified temperature node, \( T_i \), to float by connecting it to a constant temperature, \( T_c \), through a very high conductance, \( K_H \), and evaluating

\[
Q_i = K_H (T_c - T_i)
\]

(14)

Although much simpler, the size of \( K_H \) required may be so large that some computers may not accurately calculate the correct inverse of the global conductance matrix. In this case the energy balance will have to be computed for the node under question.

G. Solving the Equations - Transient Problems

In general, both the FDM and the FEM simply step along in time, re-evaluating \( K \) and \( C \) at each time step, if needed. If the time steps are large or the problem is highly non-linear, then iterations may be needed within each time step[26]. If the non-linearities may be expressed analytically a Runge-Kutta technique may be applied to the resulting non-linear equations[27]. For very non-linear problems, Gear's method may be used [28]. Regardless of why the matrices must be re-evaluated (i.e., changing boundary conditions or non-linear properties) Young's method [29] can be used to modify the matrix and the solution with a minimum increase in time, although the code structure may have to be changed considerably.

H. Graphical Display

Besides printed output, the most common output is a spatial plot of a single variable or contour plots. Such plots of the temperature are relatively easy to generate for either the FDM or the FEM if a mesh generator was used. If the FDM was based upon the heat balance approach, then it is necessary to create a set of minimum size triangular elements to perform the contour plotting [30], especially if contour smoothing is to be used [31]. Table 3 illustrates the times necessary to compute a set of triangular elements connecting 418 nodes. If the region is complete, the element calculation is rapid. If there is one or more internal voids, then the establishment of the mesh may be very expensive, as indicated by the third table entry.

<table>
<thead>
<tr>
<th>Contour Plotting Times</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Temperature contours</td>
</tr>
<tr>
<td>2. Temperature contours-establishing a mesh</td>
</tr>
<tr>
<td>3. Temperature contours-establishing a mesh with an internal void</td>
</tr>
</tbody>
</table>

If the contours are to be based upon other than linear interpolation or if the heat fluxes between elements are to be plotted, the FEM is much simpler to use since the element interpolation functions may be used, both to give a higher order curve fit and to interpolate to establish the values of \( q_x \) and \( q_y \) at the nodal points.
In general the FEM output is much better suited than the FDM output to graphical display, particularly for curved surfaces. To permit rapid detection of input errors and for rapid analysis, both a plane and an isometric (with hidden line capability) contour plotter should be available.

EXAMPLES

We present some typical examples of the use of the two methods.

A. Distributed Heat Sources

Consider a one dimensional problem, 0≤x≤1, with a distributed source strength of q=q^m. The problem was treated with an FDM using two different boundary conditions and the FEM source term and the FEM utilizing 4 different elements[32]. Figure 7 illustrates the error in the temperature at the insulated surface. The solid line is for both the MFDM with a virtual point and for the BFDM. The dashed line is for the MFDM using the zero heat flux boundary condition and shows the underprediction of the temperature because of the lack of a source at the last node. The FEM results are substantially better for two reasons. Firstly, the source is not lumped at a single node, but is distributed between nodes, thus permitting a better representation of the spatially varying heat source (this is clearly shown by the equivalence of the results for the FDM using the FEM source term and the FEM.) Secondly, the higher accuracy of the FEM temperature interpolation permits consideration of temperature variations higher than the linear. The special cubic element of Tocher[33] provides for continuity of the temperature and the heat flux at the interface between elements. Over a range of source functions, 0≤m≤5, this special cubic was not found to be any better than the quadratic or the normal cubic element, neither of which ensure continuity of the heat flux.

B. Transient Temperatures - 1-D

Consider the one-dimensional slab with an insulated back surface and a front surface whose temperature is suddenly reduced to zero. Figure 8 illustrates the results obtained using several FD methods and several FEM elements[32]. In the FEM we write,

\[ T(x,y,t) = \sum N(x,y) \{ \hat{T}(t) \} \]

\[ T_n+1 = T_n + [\Delta_t \bar{T} + (1-\alpha)\hat{T}] \Delta t \]

\[ [\alpha K + \frac{C}{\Delta_t}] \{ \hat{T} \} = (1-\alpha)[C] \{ \hat{T} \} + \frac{C}{\Delta t} \{ \hat{T} \} + \alpha \{ R \} \]

or

\[ [\alpha K + \frac{C}{\Delta t}] \{ \hat{T} \} = \left[ -(1-\alpha)K + \frac{C}{\Delta t} \right] \{ \hat{T} \} + \alpha \{ R \} + (1-\alpha) \{ R \} \]
One may also use a combined space-time interpolant [34]

\[ T(x,y,t) = \sum N_i(x,y,t) \hat{T}_i \] (18)

but this method does not yet appear to be in regular use for production codes. In figure 8, the use of equation 16 with \( T(x,0)=0 \) is referred to as 'without \text{T}'. If \( T(x,0) \neq 0 \) or if equation 17 is used, the solution is referred to as 'with \text{T}'. The relatively poor FEM results are due to the use of the consistent capacitance matrix and the impulsive start. Figure 9 compares the FEM with consistent matrices and the FDM with the lumped capacitance[35]. Because the FEM overpredicts the eigenvalues (i.e., yields an overly stiff system) while the FDM underpredicts, the FEM results show an early time overshoot. This can be corrected by lumping the capacitance terms [36,37] or by expressing \( \hat{T} \) by using a different basis which is orthogonal to the shape functions for \( T[38,39] \). The effect of lumping is shown most clearly by examining the element eigenvalues for the 8 node quadrilateral as illustrated in figure 10. Lumping may also be effected by using a reduced order of integration, which softens the transient response, but care must be taken to avoid creating singular capacitance matrices[40]. The best results have been found when using the same order of integration as for the conductance. For non-linear problems, one point integration has been found to be successful[41]. Lumping has the further advantage of simplifying the algorithm and reducing the execution time and has been extensively used for structural problems[42]. Figure 11 shows the error in the effective thermal diffusivity for several different methods. Cubic Hermite elements[43] are seen to be the best, in agreement with the results of figure 8.

Another solution is to permit the discontinuous change to take place over several time increments, a technique commonly used in hydrodynamic shock analyses. Figure 12 shows that while this smooths the results, it does not reduce the lag in the response which is due to the excessive dissipation of the FDM and lumped FEM solutions. A better way is to evaluate the right hand side at the half time intervals or by a consistent time interpolation[44]. As illustrated, this gives the best performance. Table 2 lists the execution times needed for the different methods for an equal number of time steps. For non-linear problems in which the matrices must be reformed and re-solved, the FEM is substantially slower than the FDM, particularly when the SSOR method is used.


Consider a two-dimensional plate as shown in figure 13. Because of the sharp edge of the insulated splitter plate, the heat flux at the tip is infinite and of the form

\[ q = \kappa_1 f_1(\theta) / \sqrt{r} + \kappa_2 f_2(\theta) \sqrt{r} \] (19)

This problem was treated by using the MFDM and the FEM.
1. FDM

Figure 13 illustrates the results obtained with different mesh sizes using a 5 point stencil. Substantial errors in the temperature and in the heat fluxes, figure 14, were found and even a reduction from $\Delta x = \Delta y = 1/4$ to $\Delta x = 1/20$ did not yield convergence, even for mesh points further than $2\Delta x$ from the singularity.

2. FEM

Figures 13 and 14 also show the comparable results found by using different elements and sizes. Of the three elements used, the quadratic is seen to give the best results. Table 4 lists the execution times. Although the quadratic element times are long, the times necessary to produce an accuracy in the temperature comparable to that of the linear element or the MFDM are approximately equal. On the other hand, the accuracy in evaluating the heat flux is much better than that of the FDM, with the result that the comparable execution times are much less.

<table>
<thead>
<tr>
<th>number of nodes</th>
<th>$\Delta x (= \Delta y)$</th>
<th>FEM linear</th>
<th>quadratic</th>
<th>cubic</th>
<th>FDM SOR</th>
<th>FDM SSOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>1/4</td>
<td>1</td>
<td>1</td>
<td></td>
<td>0.8</td>
<td>1.1</td>
</tr>
<tr>
<td>45</td>
<td>1/8</td>
<td>3.6</td>
<td>10.4</td>
<td>110</td>
<td>4.5</td>
<td>3</td>
</tr>
<tr>
<td>66</td>
<td>1/10</td>
<td>6.0</td>
<td></td>
<td></td>
<td>8.1</td>
<td>4.3</td>
</tr>
<tr>
<td>153</td>
<td>1/16</td>
<td>20.5</td>
<td>121</td>
<td>3097</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>231</td>
<td>1/20</td>
<td>39.1</td>
<td>274</td>
<td></td>
<td>53</td>
<td>31</td>
</tr>
<tr>
<td>361</td>
<td>1/32</td>
<td>172</td>
<td>1531</td>
<td></td>
<td>207</td>
<td>183</td>
</tr>
</tbody>
</table>

D. Singular Problems - Lagrangian Variables

As described above, problems with concentrated heat sources or with singular heat fluxes can rarely be treated with standard FDM or FEM programs since the interpolating functions are incapable of adequately representing the singular temperature field. In both methods, the interpolating functions can be expanded to include the singular behavior. If the strength of the singularity is known, the extra term simply serves as an additional heat source. If the strength is unknown, the FDM and FEM must be modified. For the FDM, let $\hat{T}$ be the FD approximation to the field equation and $B_T$ be the FD approximation to the boundary condition. Then by expressing the temperature as

$$T = \hat{T} \text{(smooth)} + K_\perp S \quad (20a)$$

we find

$$\hat{T} \neq \nabla^2 T + K_\perp (\nabla S - \nabla^2 S) \quad (20b)$$

$$B_T \neq \beta T + K_\perp (\beta S - \beta S) \quad (20c)$$
It is thus apparent that the effect of the singularities can be considered by adding the pseudo heat source and boundary heat fluxes.

For the FEM, the appropriate formulation is\cite{45}:

\[ T = \langle n \rangle \{ \hat{T} \} + K_1 (S - \langle N \rangle \{ \hat{S} \}) \]  

(21)

In this solution the value of $K_1$ is considered to be a Lagrangian variable which is determined by differentiating the integral to yield extra equations of the form:

\[ \frac{\partial I}{\partial \{ \hat{T} \}} = [K] \{ \hat{T} \} + K_1 (S_x - \langle N_x \rangle \{ \hat{S} \}) \{ N_x \} \]  

(22a)

\[ \frac{\partial I}{\partial K_1} = (S_x - \langle N_x \rangle \{ \hat{S} \}) \{ N_x \} \{ \hat{T} \} + (S_x - \langle N_x \rangle \{ \hat{S} \})^2 K_1 \]  

(22b)

Figure 13 illustrates a problem in which the magnitude of the singularity is unknown. In the FDM, the value of $K_1$ is found by applying the field equation to the closest boundary points, in addition to using the expanded boundary conditions. When using the FEM, it is not necessary that all elements contain the singular terms, only those near the singularity. However it has been found important\cite{46} that a smooth transition between the singular elements and the regular elements be provided by establishing transition elements in which the temperature is of the form

\[ T(x,y) = \langle N \rangle \{ \hat{T} \} + f(x,y) K_1 (S - \langle N \rangle \{ \hat{S} \}) \]  

(23)

in which the function $f(x,y)$ has the value of 1 along edges common with the singular elements and 0 along edges common with the regular mesh. As seen from figure 15, relatively coarse meshes for both the FDM and the FEM are sufficient to determine $q_x$ with good accuracy if the singular term is included, but even increasing the number of elements by a factor of 16 is insufficient if it is omitted.

Table 5 gives a comparison of the values of the temperature at point $P$ as determined by several different FDM and FEM solutions and the equivalence of the two singular methods is apparent. Table 6 lists the values of the singular strengths, $K_1$ and $K_2$. Although $K_2$ varies considerably, $K_1$ (which is the dominant singularity) is quite constant. If $K_2$ were to be computed more exactly, then it would be necessary to include the next term in the singular series\cite{47}. The superiority of the singular approach is evident.


<table>
<thead>
<tr>
<th>Δx</th>
<th>1st order triangle</th>
<th>2nd order</th>
<th>1st order Singular triangle</th>
<th>1st order quadrilateral</th>
<th>2nd order</th>
<th>2nd order Singular quadrilateral</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/2</td>
<td>0.1429</td>
<td>0.1176</td>
<td>0.170</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/4</td>
<td>0.1591</td>
<td>0.1737</td>
<td>0.1579</td>
<td>0.083</td>
<td>0.133</td>
<td></td>
</tr>
<tr>
<td>1/6</td>
<td>0.1659</td>
<td>0.1771</td>
<td>0.1674</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/8</td>
<td>0.1697</td>
<td>0.1784</td>
<td>0.1717</td>
<td>0.181</td>
<td></td>
<td>0.151</td>
</tr>
<tr>
<td>1/10</td>
<td>0.1721</td>
<td></td>
<td>0.1741</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/12</td>
<td>0.1738</td>
<td></td>
<td>0.1756</td>
<td>0.166</td>
<td>0.178</td>
<td>0.1824</td>
</tr>
<tr>
<td>1/16</td>
<td>0.1759</td>
<td></td>
<td>0.1775</td>
<td>0.173</td>
<td>0.180</td>
<td>0.1827</td>
</tr>
<tr>
<td>1/20</td>
<td>0.1772</td>
<td></td>
<td>0.176</td>
<td>0.176</td>
<td>0.181</td>
<td>0.1827</td>
</tr>
<tr>
<td>1/24</td>
<td>0.1781</td>
<td></td>
<td>0.1793</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Δx</th>
<th>number of singular elements</th>
<th>FEM K₁</th>
<th>FEM K₂</th>
<th>FDM K₁</th>
<th>FDM K₂</th>
<th>FDM Tp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/2</td>
<td>2(all)</td>
<td>1.285</td>
<td>0.091</td>
<td>0.170</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/4</td>
<td>2</td>
<td>1.276</td>
<td>0.067</td>
<td>0.179</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1.234</td>
<td>-0.032</td>
<td>0.179</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8(all)</td>
<td>1.224</td>
<td>-0.045</td>
<td>0.178</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/8</td>
<td>2</td>
<td>1.275</td>
<td>-0.056</td>
<td>0.182</td>
<td>1.1325</td>
<td>-0.492</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1.232</td>
<td>-0.048</td>
<td>0.182</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>1.212</td>
<td>-0.077</td>
<td>0.181</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>1.203</td>
<td>-0.086</td>
<td>0.181</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>32(all)</td>
<td>1.195</td>
<td>-0.094</td>
<td>0.181</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Under some conditions, the analytical form of the singular terms may be so complex that the integration required for the FEM may be difficult to perform with acceptable accuracy. In this case, the singular FDM, which satisfies only at the nodal points, is a more useful method.

One of the interesting features of the FEM is that it is often possible to distort an element to produce the desired singularity[48]. Referring the splitter plate problem, if an 8 node quadrilateral is used and the mid side nodes are shifted to the quarter points as indicated in figure 16, the element automatically includes a square root singularity. The figure compares these quarter point element results with those of the singular FDM and the comparison is excellent.
E. Radiation Problems

Regions with internal voids which have strongly radiating boundaries pose a problem, particularly for the FEM, even if the boundary conditions are linearized, since all of the nodes interact simultaneously and the resultant band width is very large and portions of the matrix are dense. One approach is to consider the problem as two problems, the radiation problem, and the conduction problem. The radiation problem is solved separately ensuring radiative equilibrium among the nodes. The radiative heat flux is then considered as a known heat flux and applied to the right hand side of equation 5. Because the boundary condition is so non-linear, it often proves necessary to use a strong under-relaxation factor to limit the flux change to assure convergence and avoid the overshoot observed with impulsively changed boundary conditions as indicated by the transient problem. If the rest of the problem is linear, this approach is satisfactory. If the rest of the problem is highly non-linear, an iterative FDM appears to work more efficiently.

F. Phase Changes

Probably the most difficult problem to treat effectively by either method is that of a transient phase change. Three methods, figure 17, appear to be the most common. In the first, the interface motion is computed on the basis of the conservation equations, just as is done for fluid shock calculations[49]. If an implicit time solution is used, this method requires the solution of non-linear algebraic equations[50,51]. The second method is to assume that the phase change occurs over a small temperature range, T. The latent heat is then approximated by a large specific heat value, or the enthalpy may be used as the primary dependent variable. An iterative solution is often needed to ensure that the interface is maintained at the correct fusion temperature and special care must be taken in evaluating the capacity[51,52]. In a third method, based upon the use of the enthalpy and the temperature, the interface position is not explicitly determined. In this method an artificial specific heat is not used. This method has not been fully developed and some problems have been noted if the enthalpy interpolation is other than a step function[53]. In its present form, this method may be better suited for use with the BFDM.

Figure 18 compares some typical results of the first two methods using FDM and FEM. Comini's solution[51] used Lee's time integration and gave slightly greater solidification depths than the other methods. The FDM interface was slightly in error at the earliest time because of the lumped capacitance, although it quickly gave correct values. The FDM enthalpy solution gave essentially identical results.
G. Conclusions

Having used both FDM and FEM for more than 2 decades to solve a variety of thermal problems, ranging from simple 1-D transient cases to 3-D singular biological problems, we have drawn the following general conclusions.

1. The mathematical FDM appears to be best suited to research problems, especially ones for which the analyst wishes to ensure that the boundary conditions are treated with high order accurate schemes or ones in which special algorithms are used at specified nodal points.

2. The heat balance FDM appears to be best for:
   1. Highly non-linear problems, for which iterative solutions are efficient.
   2. Problems in which the continuity of the heat flux is important.
   3. Multi-dimensional problems involving change of phase.

3. The FEM is best suited to:
   1. Irregular regions for which the automatic mesh generation and a library of highly accurate elements permit good modelling of the region and the consequent temperature profile.
   2. Mildly non-linear problems for which the iterations are few.
   3. Problems for which graphical output is important.
   4. Problems in which special temperature profiles are desired, since these may be easily obtained with special elements.
   5. Problems involving singular temperature fields or concentrated heat sources.
   6. Problems in which different approximations are to be used in different regions or problems which involve the joining of several parts.

Although each method may appear to be best for a particular class of problems, we have also reached the rather general conclusions that:

1. The analyst should be knowledgeable about both methods, at least to the extent that their general characteristics are understood.

2. Because thermal FEM elements are still being developed, their characteristics are not generally known. The analyst should experiment with such elements until their behavior, singly or in concert with other elements, is clearly understood[54]. In particular, the performance of any one element is not intuitively obvious and may not be representative of a 'similar element' for another problem[55].

3. Except for very special problems, either approach is satisfactory and which method used depends more upon the availability and familiarity of codes (and pre- and post-processors) than upon any intrinsic differences between the two methods.
We recognize that this last conclusion is rather fuzzy, but since the results obtained with either method will vary as the stencil or the element is changed, we have found that nothing will compensate for the analyst’s insight into the detailed characteristics of the method used. Since such insight is normally developed only by exercising a program on a variety of problems and since it is rare that any one person is comfortably familiar with more than one method, it is not surprising that there is a considerable tendency to continue using a familiar code, even if some shortcomings exist.

A recent round robin test of FDM and FEM[56] applied to a typical mixed boundary condition, transient problem (fig. 19) showed that either method was satisfactory and that the apparent value of either depended more upon the pre- and post-processors available than upon the intrinsic characteristics or accuracies of the methods. On the other hand, for some conditions, particularly anisotropic problems, the results of reference 57 indicate that the FDM may be more accurate, but when there was no cross coupling, the FEM was superior.

Readers should consult references 58 and 59 for more information about the FDM and references 60 and 61 for detailed insight into the FEM. Current work in the FDM is concentrated on improving the accuracy and execution times, with special emphasis on thermal network correction methods[62]. Extensions of the FEM to treat combined convection-conduction problems and to merge the thermal FEM with the structural FEM, taking the different mesh requirements and time responses into consideration, are also being studied[63]. In addition, some recent work[64,65] has shown that elliptical PDE’s and conduction-convection problems can be accurately solved by combining collocation at the Gaussian points with the FEM elements. Finally, we should note that the development of small-scale personal computers, which are slow but may be dedicated to specific tasks, can be expected to have profound influences upon the structure of the computer codes, the use of interactive execution, and the specific algorithms used[66]. In this context, it should be recognized that the substructuring method, which is commonly used on these small computers, cannot be used for transient problems.
**NOMENCLATURE**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Area</td>
</tr>
<tr>
<td>B</td>
<td>FDM boundary condition operator</td>
</tr>
<tr>
<td>c</td>
<td>Specific heat capacity</td>
</tr>
<tr>
<td>C</td>
<td>Capacitance matrix</td>
</tr>
<tr>
<td>f</td>
<td>Transition function</td>
</tr>
<tr>
<td>I</td>
<td>Integral</td>
</tr>
<tr>
<td>J</td>
<td>Jacobian of transformation</td>
</tr>
<tr>
<td>k</td>
<td>Thermal conductivity</td>
</tr>
<tr>
<td>K</td>
<td>Thermal conductance matrix</td>
</tr>
<tr>
<td>l</td>
<td>Distance between nodes</td>
</tr>
<tr>
<td>n</td>
<td>Interpolating function</td>
</tr>
<tr>
<td>N</td>
<td>Shape function</td>
</tr>
<tr>
<td>q'</td>
<td>Generated heat density</td>
</tr>
<tr>
<td>q_n</td>
<td>Boundary heat flux</td>
</tr>
<tr>
<td>q_x</td>
<td>Directional heat flux</td>
</tr>
<tr>
<td>Q_i</td>
<td>Net heat input to a node</td>
</tr>
<tr>
<td>R</td>
<td>Right hand vector</td>
</tr>
<tr>
<td>S</td>
<td>Singular function</td>
</tr>
<tr>
<td>t</td>
<td>Time</td>
</tr>
<tr>
<td>T</td>
<td>Temperature</td>
</tr>
<tr>
<td>*T</td>
<td>Time derivative of temperature</td>
</tr>
<tr>
<td>\bar{T}</td>
<td>Smooth temperature distribution</td>
</tr>
<tr>
<td>\alpha</td>
<td>Weighting parameter</td>
</tr>
<tr>
<td>\beta</td>
<td>Boundary condition operator</td>
</tr>
<tr>
<td>\epsilon</td>
<td>Error</td>
</tr>
<tr>
<td>\xi, \eta</td>
<td>Coordinates of unit square</td>
</tr>
<tr>
<td>\rho</td>
<td>Density</td>
</tr>
<tr>
<td>\Delta</td>
<td>Change</td>
</tr>
<tr>
<td>\hat{\gamma}</td>
<td>Nodal values</td>
</tr>
<tr>
<td>\Theta</td>
<td>Approximation to the Laplacian</td>
</tr>
</tbody>
</table>
REFERENCES

56. Emery, A.F., Workshop on FE and FD Methods, Sandia Laboratories, Livermore, Calif., 1979
Figure 1.— Finite difference mesh for the heat balance method.

Figure 2.— Finite difference nodal point arrangements.
Figure 3.- Mapping an element on a unit square (showing the Gaussian integration points).

Figure 4.- Common two dimensional finite elements.
Figure 5. - Error (%) and computer time (sec) for a clamped shell using an FEM and an FDM [12].

Figure 6. - Treatment of irregular boundaries.
Figure 7.- Temperature error at the insulated surface for a source strength $= X^m$.

$T(\text{exact}) = 0.1429$

Figure 8.- Temperature error for $\alpha T/L^2 = 0.0625$ and $X/L = 0.5$.

$T(\text{exact}) = 0.1573$
Figure 9. - Temperature at the insulated surface using an FDM and an FEM with 2 node linear (L-2) and 3 node linear (L-3) elements.

Figure 10. - Eigenvectors for the eight node element showing the effect of lumping.
Figure 11.- Error in the numerical diffusivity for different methods [43].

Figure 12.- Transient temperature response for a slab using \( \tilde{T}(x,0) = 0 \) with ramps of 0 and 1 \( \Delta t \) and a consistent method using a modified initial surface temperature (TOLD = 50).
Figure 13.- Temperature at point P for the singular problem.

Figure 14.- Heat flux at point Q for the singular problem.
Figure 15. Values of the heat flux along the insulating splitter plate.

Figure 16. Calculation of the time dependent singularity strength for the problem of figure 13.
Figure 17.— Enthalpy and interface models for phase change calculations.

Figure 18.— Solidification front position for a plane problem.
Figure 19. - Comparison of FDM and FEM results for a mixed boundary condition problem.