Part I

Mechanics and Modeling

Microelectronic devices, optoelectronic devices, and MEMS (micro-electrical-mechanical-systems) devices and their systems are mainly being manufactured with a silicon chip, or chips, in various compact package configurations to satisfy cost and performance requirements. A typical package and assembly is manufactured through many processes (both front-end and back-end), followed by many severe reliability qualification tests. A typical packaging assembly consists of components of different materials with different mechanical and thermal properties, and geometric discontinuities exist in terms of vias, comers, free edges, interfaces, surfaces, and composites. Non-uniform temperature and moisture fields during each process of packaging and assembly, testing, storage, and operation often subject these materials and components to various failure modes, for example, metal line voiding, passivation cracking, die cracking, fine metal line smearing, package swelling, warpage, and delamination. If failure occurs in a process step of manufacturing, and results in the component not being processed, yield can be a big issue, and popcorning during the reflow of plastic packaging onto the board is such an example. System functionality tests performed on these packages and assemblies, in general, will not detect the majority of these failures. This may not cause catastrophic failure but may cause performance degradation and failure before the designed life cycle. Due to the rapid advancement of the related industries, information on detailed failure initiation, growth and performance degradation is not easily available and it is a challenging task for the modeling and simulation community. Closed-form solutions are mostly unavailable due to the many nonlinearities such as material nonlinearities for many polymer and solder materials, geometrical nonlinearity such as membrane deformation in sensors, force nonlinearity such as contact and debonding, contour change for sequential process steps, micro-structural changes such as for annealing, damage development, and mass transport for electromigration and wetting, just to list a very few examples. Serious study is essential for modeling and simulation. Constitutive models for engineering materials are a must to avoid a wrong analysis which can easily result in engineers arriving at wrong conclusions or no conclusions at all. In the first part of the book, we will introduce constitutive models and leave the definition of the basic concepts such as stress, strain, and so on, to Appendix A. Then we will briefly present the finite element methods and some of the advanced features such as sub-modeling, sub-structure, and element birth and death. Material testing for small samples is presented to show the uniqueness of our samples as compared to bulk structures for traditional industries. User-supplied subroutines are also presented with solders as an example. Multi-physics and multi-scale modeling is presented by briefly introducing molecular dynamics (MD), while leaving most of MD to Part IV. Validation tools and their applications for both on-line testing and off-line quality analysis are presented. Fracture mechanics with the focus on the interfacial fracture mechanics is presented and the concurrent engineering approach is presented as a platform for leading research laboratories which are product oriented and have a strong desire to shorten the time-to-market and time-to-profit.

1

Constitutive Models and Finite Element Method

In this chapter, the basic equations of continuum mechanics are introduced. It is assumed that the readers already have a basic prior knowledge of the subject. Therefore, lengthy derivations are omitted, such information being provided in the Appendix. Only equations that are needed later for deriving the mechanics theories are outlined. Most equations are adopted from several famous references listed in the literature of this chapter, including ABAQUS theory manual (Belytschko *et al*., 2000; Hibbit *et al.,* 2008).

1.1 Constitutive Models for Typical Materials

1.1.1 Linear Elasticity

In many engineering applications involving small strains and rotations, the response of the material may be considered to be linearly elastic. The most general way to represent elastic tensor $\boldsymbol{C}$ relation between the stress and strain tensors is given by:

$$\sigma_{ij} = C_{ijkl}\varepsilon_{kl} \quad \boldsymbol{\sigma} = \boldsymbol{C} : \boldsymbol{\varepsilon} \tag{1.1}$$

where C_{ijkl} are components of the 4th-order tensor of elastic moduli. This equation is the generalized Hooke's law which incorporates a fully anisotropic material response.

The strain energy per unit volume, often called the elastic potential, *W,* is generalized to multiaxial states by:

$$W = \int \sigma_{ij} d\varepsilon_{ij} = \frac{1}{2} C_{ijkl}\varepsilon_{ij}\varepsilon_{kl} = \frac{1}{2}\boldsymbol{\varepsilon} : \boldsymbol{C} : \boldsymbol{\varepsilon} \tag{1.2}$$

The stress is then given by:

$$\boldsymbol{\sigma} = \frac{\partial W}{\partial \boldsymbol{\varepsilon}} \tag{1.3}$$

Modeling and Simulation for Microelectronic Packaging Assembly: Manufacturing, Reliability and Testing, First Edition. Sheng Liu and Yong Liu.

which is the tensor equivalent of Equation 1.2. The strain energy is assumed to be positive definite:

$$W = \frac{1}{2} C_{ijkl} \varepsilon_{ij} \varepsilon_{kl} = \frac{1}{2} \boldsymbol{\varepsilon} : C : \boldsymbol{\varepsilon} \geq 0 \tag{1.4}$$

with equality if and only if $\varepsilon_{ij} = 0$ which implies that $\boldsymbol{C}$ is a positive-definite 4th-order tensor. From the symmetries of the stress and strain tensors, the material coefficients have the so called minor symmetries:

$$C_{ijkl} = C_{jikl} = C_{ijlk} \tag{1.5}$$

and from the existence of a strain energy potential Equation 1.2 it follows that:

$$C_{ijkl} = \frac{\partial^2 W}{\partial \varepsilon_{ij} \partial \varepsilon_{kl}} \quad \boldsymbol{C} = \frac{\partial^2 W}{\partial \boldsymbol{\varepsilon} \partial \boldsymbol{\varepsilon}} \tag{1.6}$$

If W is a smooth function of ε, Equation 1.6 implies a property called major symmetry:

$$C_{ijkl} = C_{klij} \tag{1.7}$$

since smoothness implies:

$$\frac{\partial^2 W}{\partial \varepsilon_{ij} \partial \varepsilon_{kl}} = \frac{\partial^2 W}{\partial \varepsilon_{kl} \partial \varepsilon_{ij}} \tag{1.8}$$

The general 4th-order tensor C_{ijkl} has $3^4 = 81$ independent constants. These 81 constants may also be interpreted as arising from the necessity to relate nine components of the complete stress tensor to nine components of the complete strain tensor, that is, $9 \times 9 = 81$. The symmetries of the stress and strain tensors require only that six independent components of stress be related to six independent components of strain. The resulting minor symmetries of the elastic moduli therefore reduce the number of independent constants to $6 \times 6 = 36$. Major symmetry of the moduli, expressed through Equation 1.7 reduces the number of independent elastic constants to $n(n+1)/2 = 21$, for $n = 6$, that is, the number of independent components of a 6×6 matrix.

Considerations of material symmetry further reduce the number of independent material constants. An isotropic material is one which has no preferred orientations or directions in its properties, so that the stress-strain relation is identical when expressed in component form in any rectangular Cartesian coordinate system. The most general constant isotropic 4th-order tensor can be shown to be a linear combination of terms comprised of Kronecker deltas, that is, for an isotropic linearly elastic material:

$$C_{ijkl} = \lambda \delta_{ij} \delta_{kl} + \mu \left(\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk} \right) + \mu' \left(\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk} \right) \tag{1.9}$$

Because of the symmetry of the strain and the associated minor symmetry $C_{ijkl} = C_{ijlk}$ it follows that $\mu' = 0$. Thus Equation 1.9 is written as:

$$C_{ijkl} = \lambda \delta_{ij} \delta_{kl} + \mu \left(\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk} \right) \quad \boldsymbol{C} = \lambda \boldsymbol{I} \otimes \boldsymbol{I} + 2\mu \boldsymbol{I} \tag{1.10}$$

and the two independent material constants λ and μ are called the Lamé constants. The stress strain relation for an isotropic linear elastic material may therefore be written as:

$$\sigma_{ij} = \lambda \varepsilon_{kk} \delta_{ij} + 2\mu \varepsilon_{ij} = C_{jikl} \varepsilon_{kl} \quad \boldsymbol{\sigma} = \lambda trace(\boldsymbol{\varepsilon}) \mathbf{I} + 2\mu \boldsymbol{\varepsilon} \tag{1.11}$$

1.1.2 Elastic-Visco-Plasticity

In classical formulations of elastic-visco-plasticity, the yield criterion is defined through a loading function $F \equiv F(\boldsymbol{\sigma}, \boldsymbol{q})$, where σ denotes the stress state and $\boldsymbol{q}$ denotes the internal variables. As elastic-visco-plastic deformation appears, the stress is permissible outside the closure of the loading surface, that is $F(\boldsymbol{\sigma}, \boldsymbol{q}) > 0$. However, in rate-independent plasticity, $F(\boldsymbol{\sigma}, \boldsymbol{q}) \leq 0$, which is the basic difference between viscoplasticity and rate-independent plasticity.

For the classic elastic-viscoplastic constitutive model (Figure 1.1, where σ_s is the yield stress), the total strain rate is the sum of its elastic and viscoplastic components:

$$\dot{\boldsymbol{\varepsilon}} = \dot{\boldsymbol{\varepsilon}}^{e} + \dot{\boldsymbol{\varepsilon}}^{vp} \tag{1.12}$$

where the superscript "e" indicates the elastic component, and the superscript "vp" indicates the viscoplastic component.

According to a study by Perzyna (Perzyna, 1971), the viscoplastic strain rate $\dot{\boldsymbol{\varepsilon}}^{vp}$ is defined by a flow rule:

$$\dot{\boldsymbol{\varepsilon}}^{vp} = \gamma \frac{\partial F}{\partial \boldsymbol{\sigma}} = \frac{\langle \boldsymbol{\Phi}(F) \rangle}{\eta} \frac{\partial F}{\partial \boldsymbol{\sigma}} \tag{1.13}$$

where F is a yield function of the material; $\boldsymbol{\Phi}(F)$ is a flow function; $\langle \cdot \rangle$ are MacCauley's brackets, $\langle x \rangle = (x + |x|)/2$; $\dot{\gamma}$ is the viscoplastic flow rate parameter; and η is a given viscoplastic material fluidity parameter.

Considering a mixed strain-hardening model which includes both isotropic hardening and kinematic hardening, the loading yield function is:

$$F(\boldsymbol{\xi}, \bar{e}^{vp}) = \frac{\|\boldsymbol{\xi}\|}{\sqrt{\frac{2}{3}}\kappa(\bar{e}^{vp})} \tag{1.14}$$

where $\kappa(\bar{e}^{vp})$ is a function governing the isotropic expansion of the yield surface. $\boldsymbol{\xi} = \boldsymbol{S} - \boldsymbol{\alpha}$, $\boldsymbol{S}$ is the deviatoric stress tensor, and α refers to a back stress tensor (an internal variable), which defines the translation of the center of the yield surface. The evolution of the back stress tensor is usually defined as $\dot{\boldsymbol{\alpha}} = \frac{2}{3} H' \dot{\gamma} \frac{\partial F}{\partial \boldsymbol{\sigma}}$, where H' is called the kinematic hardening modulus. If $H' = 0$, one gets the isotropic hardening. $\|\boldsymbol{\xi}\| = \sqrt{\boldsymbol{\xi} : \boldsymbol{\xi}}$. $\bar{e}^{vp}$ is the cumulated equivalent viscoplastic strain:

$$\bar{e}^{vp} = \int_0^t \sqrt{\frac{2}{3}} (\dot{\boldsymbol{\varepsilon}}^{vp}(\tau) : \dot{\boldsymbol{\varepsilon}}^{vp}(\tau))^{1/2} \, \mathrm{d}\tau$$

Common choices for the flow function $\boldsymbol{\Phi}(F)$ are exponent-type and power-type, which can be written as:

$$\boldsymbol{\Phi}(F) = e^{cF} - 1 \tag{1.15a}$$

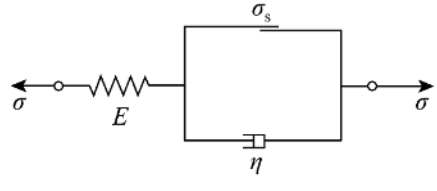

Figure 1.1 One dimensional rheological model

$$\boldsymbol{\Phi}(F) = F^d \tag{1.15b}$$

in which c and d are prescribed constants. In general, the flow function $\boldsymbol{\Phi}(F)$ is the combination of both basic types above.

On the other hand, from Equation 1.12 and Hooke's law, it can be obtained:

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{C} : (\dot{\boldsymbol{\varepsilon}} - \dot{\boldsymbol{\varepsilon}}^{vp}) \tag{1.16}$$

where C is the 4th-order elastic tensor given by $\mathbf{C} = \lambda \mathbf{l} \otimes \mathbf{l} + 2\mu \mathbf{I} \cdot \mathbf{l}$ is the 2nd-order unit tensor given by $\boldsymbol{I} = \delta_{ij}\boldsymbol{e}_i \otimes \boldsymbol{e}_j$; $\boldsymbol{I} = (1/2)\left[\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right]\boldsymbol{e}_i \otimes \boldsymbol{e}_j \otimes \boldsymbol{e}_k \otimes \boldsymbol{e}_l$ is 4th-order unit tensor; where **e** represents basis vector and $\otimes$ denotes the tensor product. λ and μ are the Lame' constants, and μ is the shear modulus. In fact, as the fluidity parameter $\eta \to 0$, stress state outside of the loading surface becomes increasingly penalized and thus $F \to 0$, the viscoplastic Equations 1.13 and 1.16 reduce to the rate-independent plasticity problem. As the fluidity parameter $\eta \to \infty$, $\dot{\gamma} \to 0$, $\dot{\boldsymbol{\alpha}} \to \mathbf{0}$, $\dot{\boldsymbol{\varepsilon}}^{vp} \to \mathbf{0}$, Equations 1.13 and 1.16 collapse to the rate form of linear elasticity.

(i) Elastic-Visco-Plastic Radial Return Evolution (RRE)

A trial (Tr) deviatoric stress is introduced as:

$$\boldsymbol{\xi}_{n+1}^{\mathrm{Tr}} = \boldsymbol{\xi}_n + 2\mu\Delta\boldsymbol{e}_n \tag{1.17}$$

where e is the deviatoric strain tensors.

Considering the viscoplastic evolution problem for strain increment in any given finite time step Δt, the elastic-visco-plastic constitutive law reduces to giving a RRE rule that makes $\boldsymbol{\sigma}_{n+1} \equiv \bar{\boldsymbol{\sigma}}(\boldsymbol{\varepsilon}_n, \boldsymbol{\sigma}_n, \boldsymbol{q}_n, \Delta\boldsymbol{\varepsilon}_n) = \boldsymbol{\sigma}_n + \Delta\boldsymbol{\sigma}_n$ finally be consistent with the loading surface $f = F(\boldsymbol{\xi}, \bar{e}^{vp})$.

Figure 1.2 shows the evolution actions that take place on the π-plane which illustrates a general viscoplastic case with isotropic and kinematic hardening. Note that as the parameter η approaches zero, the yield surface $f_{n+1} = F$ approaches zero, it actually returns to elastic-plastic case. When the kinematic hardening parameter $H' = 0$, this returns to the pure isotropic hardening case (Figure 1.3).

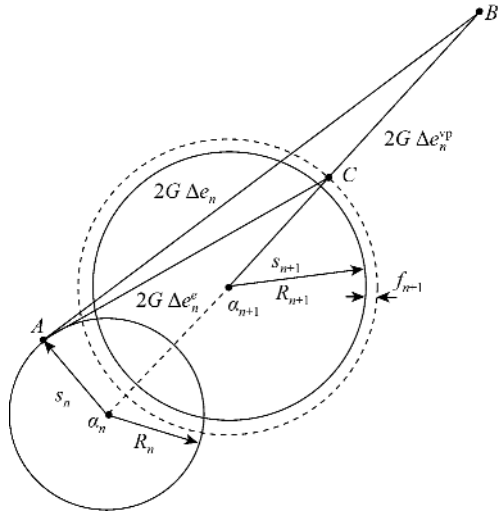

Figure 1.2 Illustration of radial return evolution with both kinematic and isotropic hardening with the radius $R = \sqrt{\frac{2}{3}}\kappa(\bar{e}^{vp})$

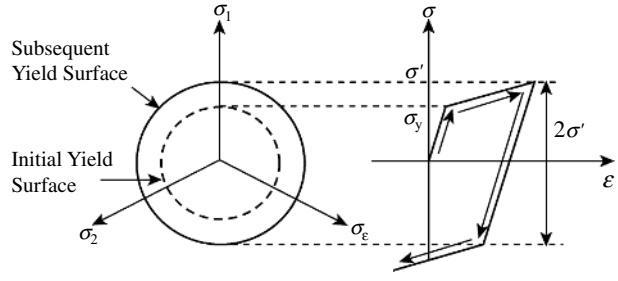

Figure 1.3 Isotropic hardening plasticity

If $F\left(\xi_{n+1}^{\mathrm{Tr}},\ \kappa_n\right) \leq 0$, that is the elastic deformation, one has:

$$\boldsymbol{\sigma}_{n+1} = \boldsymbol{\sigma} = K\Delta\boldsymbol{\varepsilon}_n : (\boldsymbol{I}\otimes\boldsymbol{I}) + 2\mu\Delta\boldsymbol{e}_n + \boldsymbol{\sigma}_n \tag{1.18}$$

where K is the bulk modulus. This is the elastic constitutive equation in incremental form.

If $F\left(\boldsymbol{\xi}_{n+1}^{\mathrm{Tr}},\ \kappa_n\right) > 0$, that is the viscoplastic deformation, one has:

$$\boldsymbol{\sigma}_{n+1} = \bar{\boldsymbol{\sigma}} = K\boldsymbol{\varepsilon}_{n+1} : (\boldsymbol{I}\otimes\boldsymbol{I}) + \boldsymbol{S}_{n+1} \tag{1.19}$$

On the other hand, one can obtain the following counterpart of the implicit backward-Euler difference scheme and viscoplastic RRE consistency equations:

$$\begin{cases} \dot{\boldsymbol{\varepsilon}}^{vp} = \dot{\gamma}\hat{\boldsymbol{n}} = \dfrac{\langle\boldsymbol{\Phi}(F)\rangle}{\eta}\hat{\boldsymbol{n}} \\ \boldsymbol{\varepsilon}_{n+1}^{vp} = \boldsymbol{\varepsilon}_n^{vp} + [\dot{\gamma}\Delta t]\hat{\boldsymbol{n}}_{n+1} \\ \boldsymbol{\alpha}_{n+1} = \boldsymbol{\alpha}_n + \dfrac{2}{3}H'[\dot{\gamma}\Delta t]\hat{\boldsymbol{n}}_{n+1} \\ \bar{e}_{n+1}^{vp} = \bar{e}_n^{vp} + \sqrt{\dfrac{2}{3}}[\dot{\gamma}\Delta t] \\ \hat{\boldsymbol{n}}_{n+1} = \dfrac{1}{\left\|\xi_{n+1}^{\mathrm{Tr}}\right\|}\xi_{n+1}^{\mathrm{Tr}} \end{cases} \tag{1.20}$$

along with the condition:

$$\left\|\boldsymbol{\xi}_{n+1}\right\| = \left\|\boldsymbol{\xi}_{n+1}^{\mathrm{Tr}}\right\| - 2\mu[\dot{\gamma}\Delta t]\left(1 + \frac{H'}{3\mu}\right) \tag{1.21}$$

In addition, note that:

$$\boldsymbol{S}_{n+1} = \left\|\boldsymbol{\xi}_{n+1}\right\|\hat{\boldsymbol{n}}_{n+1} + \boldsymbol{\alpha}_{n+1} \tag{1.22}$$

Equations 1.14, 1.15, 1.17 and 1.22 solve the viscoplastic RRE consistency equation:

$$\Re(\dot{\gamma}\Delta t) \equiv \left\|\xi_{n+1}^{\mathrm{Tr}}\right\| - \sqrt{\frac{2}{3}}\kappa\left(\bar{e}_{n+1}^{vp}\right)\cdot\hbar(\dot{\gamma}\Delta t) - 2\mu[\dot{\gamma}\Delta t]\left(1 + \frac{H'}{3\mu}\right) = 0 \tag{1.23}$$

where:

$$\begin{cases} \hbar(\dot{\gamma}\Delta t) = 1 + \dfrac{1}{c}\ln\left(1 + \dfrac{\eta}{\Delta t}[\dot{\gamma}\Delta t]\right), & \text{if} \quad \Phi(F) \quad \text{follows} \quad \text{Equation 1.15a} \\ \hbar(\dot{\gamma}\Delta t) = 1 + \left(\dfrac{\eta}{\Delta t}[\dot{\gamma}\Delta t]\right)^{\frac{1}{d}}, & \text{if} \quad \Phi(F) \quad \text{follows} \quad \text{Equation 1.15b} \end{cases}$$

The solution of Equation 1.23 from which the values of $|\dot{\gamma}\Delta t|$ are determined, can be effectively solved by the local Newton iteration procedure.

(ii) Elastic-Visco-Plastic Consistent Tangent Operator (CTO)

The CTO is defined in a 4th-order tensor (Simo and Taylor, 1985):

$$\boldsymbol{C}_{n+1} = \frac{\partial \bar{\boldsymbol{\sigma}}}{\partial \Delta \boldsymbol{\varepsilon}_n} = \frac{\partial \boldsymbol{\sigma}_{n+1}}{\partial \Delta \boldsymbol{\varepsilon}_n} \tag{1.24}$$

which depends on the particular algorithm $\delta\boldsymbol{\varepsilon}_n \rightarrow \boldsymbol{\sigma}_{n+1}(\boldsymbol{\varepsilon}_n, \boldsymbol{\sigma}_n, \boldsymbol{q}_n, \Delta\boldsymbol{\varepsilon}_n)$ chosen. Neglecting the deriving process, the CTO takes the form:

$$\boldsymbol{C}_{n+1} = \boldsymbol{C}^{vp}_{n+1} = K\boldsymbol{I}\otimes\boldsymbol{I} + 2\mu\beta\left(\boldsymbol{I} - \frac{1}{3}\boldsymbol{I}\otimes\boldsymbol{I}\right) - 2\mu(\beta - \vartheta)\hat{\boldsymbol{n}}_{n+1}\otimes\hat{\boldsymbol{n}}_{n+1} \tag{1.25}$$

where:

$$\beta = 1 - \frac{2\mu[\dot{\gamma}\Delta t]}{\left\|\xi^{\mathrm{Tr}}_{n+1}\right\|}, \quad \vartheta = 1 - \frac{1}{\frac{1}{3\mu}\kappa'\left(\bar{e}^{vp}_{n+1}\right)\cdot\hbar(\dot{\gamma}\Delta t) + \frac{1}{2\mu}\sqrt{\frac{2}{3}}\kappa\left(\bar{e}^{vp}_{n+1}\right)\cdot\hbar'(\dot{\gamma}\Delta t) + 1 + \frac{H'}{3\mu}}$$

where:

$$\begin{cases} \hbar'(\dot{\gamma}\Delta t) = \dfrac{\dfrac{h}{\Delta t}}{c\left(1 + \dfrac{\eta}{\Delta t}[\dot{\gamma}\Delta t]\right)}, & \text{if } \Phi(F) \quad \text{follows} \quad \text{Equation 1.15a} \\ \hbar'(\dot{\gamma}\Delta t) = \dfrac{\dfrac{\eta}{\Delta t}}{d}\left(\dfrac{\eta}{\Delta t}[\dot{\gamma}\Delta t]\right)^{\frac{1}{d}-1}, & \text{if } \Phi(F) \quad \text{follows} \quad \text{Equation 1.15b} \end{cases}.$$

It should be indicated that when $\boldsymbol{\sigma}_{n+1} = \bar{\boldsymbol{\sigma}}(\boldsymbol{\varepsilon}_n, \boldsymbol{\sigma}_n, \boldsymbol{q}_n, \Delta\boldsymbol{\varepsilon}_n)$ is elastic, one has $\boldsymbol{C}_{n+1} = \boldsymbol{C}^{vp}_{n+1} = \boldsymbol{C}$.

The above consistent tangent operator based constitutive relation is a unified material constitutive model that has been applied to various nonlinear problems in computational mechanics. It has been proven to be very powerful in dealing with various highly nonlinear and large deformation problems in engineering.

1.2 Finite Element Method

The finite element method is a numerical analysis technique for obtaining approximate solutions to a wide variety of engineering problems.

1.2.1 Basic Finite Element Equations

We begin with the equilibrium statement, written as the virtual work principle:

$$\int_{\mathrm{V}} \boldsymbol{\sigma} : \delta \boldsymbol{D} \mathrm{d}V = \int_{\mathrm{s}} \boldsymbol{t}^{\mathrm{T}} \cdot \delta \boldsymbol{v} \mathrm{d}S + \int_{\mathrm{V}} \boldsymbol{f}^{\mathrm{T}} \cdot \delta \boldsymbol{v} \mathrm{d}V \tag{1.26}$$

the left-hand side of this equation, the internal virtual work rate term is replaced with the integral over the reference volume of the virtual work rate per reference volume defined by any conjugate pairing of stress and strain:

$$\int_{V^0} \boldsymbol{\tau}^{\mathrm{c}} : \delta \boldsymbol{\varepsilon} dV^0 = \int_{s} \boldsymbol{t}^{\mathrm{T}} \cdot \delta \boldsymbol{v} \mathrm{d}S + \int_{\mathrm{V}} \boldsymbol{f}^{\mathrm{T}} \cdot \delta \boldsymbol{v} \mathrm{d}V \tag{1.27}$$

where $\boldsymbol{\tau}^{\mathrm{c}}$ and ε are any conjugate pairing of material stress and strain measures. The particular choice of ε depends on the individual element.

The finite element interpolator can be written in general as:

$$\boldsymbol{u} = N_{\mathrm{N}} u^{\mathrm{N}} \tag{1.28}$$

where N_{N} are interpolation functions that depend on some material coordinate system, u^{N} are nodal variables, and the summation convention is adopted for the upper case subscripts and superscripts that indicate nodal variables.

The virtual field, $\delta \boldsymbol{v}$, must be compatible with all kinematic constraints. Introducing the above interpolation constrains the displacement to have a certain spatial variation, so $\delta \boldsymbol{v}$ must also have the same spatial form:

$$\delta \boldsymbol{v} = \boldsymbol{N}_{\mathrm{N}} \delta v^{\mathrm{N}} \tag{1.29}$$

The continuum variational statement Equation 1.27 is, thus, approximated by a variation over the finite set δv.

Now δε is the virtual rate of material strain associated with $\delta \boldsymbol{v}$, and because it is a rate form, it must be linear in $\delta \boldsymbol{v}$. Hence, the interpolation assumption gives:

$$\delta \boldsymbol{\varepsilon} = \boldsymbol{\beta}_{\mathrm{N}} \delta v^{\mathrm{N}} \tag{1.30}$$

where β_{N} is a matrix that depends, in general, on the current position, $\boldsymbol{x}$, of the material point being considered. The matrix β_{N} that defines the strain variation from the variations of the kinematic variables is derivable immediately from the interpolation functions once the particular strain measure to be used is defined.

Without loss of generality we can write $\boldsymbol{\beta}_{\mathrm{N}} = \boldsymbol{\beta}_{\mathrm{N}}(\boldsymbol{x}, \boldsymbol{N}_{\mathrm{N}})$, and, with this notation, the equilibrium equation is approximated as:

$$\delta v^{\mathrm{N}} \int_{V^0} \boldsymbol{\beta}_{\mathrm{N}} : \boldsymbol{\tau}^{\mathrm{c}} \mathrm{d}V^0 = \delta \boldsymbol{v}^{\mathrm{N}} \left[\int_{\mathrm{s}} N_{\mathrm{N}}^{\mathrm{T}} \cdot \boldsymbol{t} \mathrm{d}S + \int_{\mathrm{V}} N_{\mathrm{N}}^{\mathrm{T}} \cdot \boldsymbol{f} \mathrm{d}V \right] \tag{1.31}$$

Since the δv^{N} are independent variables, we can choose each one to be nonzero and all others zero in turn, to arrive at a system of nonlinear equilibrium equations:

$$\int_{V^0} \boldsymbol{\beta}_{\mathrm{N}} : \boldsymbol{\tau}^{\mathrm{c}} \mathrm{d}V^0 = \int_s \boldsymbol{N}_{\mathrm{N}}^{\mathrm{T}} \cdot \boldsymbol{t} \mathrm{d}S + \int_{\mathrm{V}} \boldsymbol{N}_{\mathrm{N}}^{\mathrm{T}} \cdot \boldsymbol{f} \mathrm{d}V \tag{1.32}$$

This system of equations forms the basis for the standard assumed displacement finite element analysis procedure and is of the form:

$$F^{\mathrm{N}}(u^{\mathrm{M}}) = 0 \tag{1.33}$$

as discussed above. The above equations are valid for static and dynamic analysis if the body force is assumed to contain the inertia contribution. In dynamic analysis, however, the inertia contribution is more commonly considered separately, leading to the equation:

$$M^{\mathrm{NM}} \ddot{u}^{\mathrm{M}} + F^{\mathrm{N}}(u^{\mathrm{M}}) = 0 \tag{1.34}$$

For the Newton algorithm, or for the linear perturbation procedure we need the Jacobian of the finite element equilibrium equations. To develop the Jacobian, we begin by taking the variation of Equation 1.26, giving:

$$\begin{aligned} \int_{V^0} (\mathrm{d}\boldsymbol{\tau}^{\mathrm{c}} : \delta\boldsymbol{\varepsilon} + \boldsymbol{\tau}^{\mathrm{c}} : d\delta\boldsymbol{\varepsilon}) \mathrm{d}V^0 - \int_{\mathrm{s}} d\boldsymbol{t}^{\mathrm{T}} \cdot \delta\boldsymbol{v} \mathrm{d}S - \int_{\mathrm{s}} \boldsymbol{t}^{\mathrm{T}} \cdot \delta\boldsymbol{v} \mathrm{d}A_{\mathrm{r}} \frac{1}{A_{\mathrm{r}}} \mathrm{d}S \\ - \int_{\mathrm{V}} \mathrm{d}\boldsymbol{f}^{\mathrm{T}} \cdot \delta\boldsymbol{v} \mathrm{d}V - \int_{\mathrm{V}} \boldsymbol{f}^{\mathrm{T}} \cdot \delta\boldsymbol{v} J \frac{1}{J} \mathrm{d}V = 0 \end{aligned} \tag{1.35}$$

where $d()$ represents the linear variation of the quantity () with respect to the basic variables which are the degrees of freedom of the finite element model. In the above expression, $J = \left|\mathrm{d}V/\mathrm{d}V^0\right|$ is the volume change between the reference and the current volume occupied by a piece of the structure and, likewise, $A_{\mathrm{r}} = \left|\mathrm{d}S/\mathrm{d}S^0\right|$ is the surface area ratio between the reference and the current configuration. The Jacobian matrix is obtained by restricting the above variation, allowing variations in the nodal variables, u^{N}, only. Let such a restricted variation be indicated by $\partial_{\mathrm{N}} = \partial/\partial u^{\mathrm{N}}$. Examining Equation 1.35 term by term with this in mind, we proceed as follows. The first term contains $\mathrm{d}\boldsymbol{\tau}^{\mathrm{c}}$. We now assume that the constitutive theory allows us to write:

$$\mathrm{d}\boldsymbol{\tau}^{\mathrm{c}} = \boldsymbol{H} : \mathrm{d}\boldsymbol{\varepsilon} + \boldsymbol{g} \tag{1.36}$$

where $\boldsymbol{H}$ and $\boldsymbol{g}$ are defined in terms of the current state, direction of straining, and so on, and on the kinematic assumptions used to form the generalized strains. From the choice of generalized strain measure and interpolation function:

$$\partial_{\mathrm{N}}\boldsymbol{\varepsilon} = \frac{\partial \boldsymbol{\varepsilon}}{\partial u^{\mathrm{N}}} = \boldsymbol{\beta}_{\mathrm{N}} \tag{1.37}$$

from the above constitutive assumption:

$$\partial_{\mathrm{N}}\boldsymbol{\tau}^{\mathrm{c}} = \boldsymbol{H} : \boldsymbol{\beta}_{\mathrm{N}} \tag{1.38}$$

Now, since $\delta\boldsymbol{\varepsilon}$ is the first variation of ε with respect to nodal variables:

$$\delta\boldsymbol{\varepsilon} = \partial_{\mathrm{M}}\boldsymbol{\varepsilon}\delta u^{\mathrm{M}} = \boldsymbol{\beta}_{\mathrm{M}}\delta u^{\mathrm{M}} \tag{1.39}$$

thus, the first term in Equation 1.35 of the Jacobian matrix can be written as:

$$\int_{V^0} \boldsymbol{\beta}_{\mathrm{M}} : \boldsymbol{H} : \boldsymbol{\beta}_{\mathrm{N}} \mathrm{d}V^0 \tag{1.40}$$

which is the usual "small-displacement stiffness matrix", except that, since the strain measure ε is always nonlinear in displacement, the β_{N} in this term is a function of displacement.

The second term in Equation 1.35 is:

$$\int_{V^0} \boldsymbol{\tau}^{\mathrm{c}} : \mathrm{d}\delta\boldsymbol{\varepsilon} \mathrm{d}V^0$$

which can be rewritten as:

$$\int_{V^0} \boldsymbol{\tau}^{\mathrm{c}} : \partial_{\mathrm{N}}\delta\boldsymbol{\varepsilon} \mathrm{d}V^0$$

and further leads to:

$$\int_{V^0} \boldsymbol{\tau}^{\mathrm{c}} : \partial_{\mathrm{N}}\boldsymbol{\beta}_{\mathrm{M}} \mathrm{d}V^0$$

where this term contributes to the Jacobian and is the "initial stress matrix."

The external load rate terms in Equation 1.35 are considered next. In general, these load vectors can be written as:

$$\boldsymbol{t} = \boldsymbol{t}(\lambda, \boldsymbol{x}) \text{ and } \boldsymbol{f} = \boldsymbol{f}(\lambda, \boldsymbol{x}) \tag{1.41}$$

where λ represents the externally prescribed loading parameters. Whether the load depends on position or not, it is up to the particular load type, but common types of loading such as pressure, or centrifugal load do depend on position—for example, if $\boldsymbol{t}$ is caused by pressure on the surface, $\boldsymbol{t}$ depends on the pressure magnitude, on the direction of the normal to the surface, and on the current surface area: the latter two are functions of the current position of points on the surface. The variation of the load vector with nodal variables can then be written symbolically as:

$$\partial_{\mathrm{N}}\boldsymbol{t} + \boldsymbol{t}\frac{1}{A_{\mathrm{r}}}\partial_{\mathrm{N}}A_{\mathrm{r}} = \boldsymbol{Q}_{\mathrm{N}}^{\mathrm{S}} \tag{1.42}$$

$$\partial_{\mathrm{N}}\boldsymbol{f} + \boldsymbol{f}\frac{1}{J}\partial_{\mathrm{N}}J = \boldsymbol{Q}_{\mathrm{N}}^{\mathrm{V}} \tag{1.43}$$

and then:

$$\delta\boldsymbol{v} = \boldsymbol{N}_{\mathrm{M}}\delta v^{\mathrm{M}} \tag{1.44}$$

where $\boldsymbol{N}_{\mathrm{M}}$ is obtained directly from the interpolation functions, we can write the Jacobian terms pertaining to the last four terms of Equation 1.35 as:

$$-\int_{\mathrm{s}} N_{\mathrm{M}}^{\mathrm{T}} \cdot Q_{\mathrm{N}}^{\mathrm{S}} \mathrm{d}S - \int_{\mathrm{V}} N_{\mathrm{M}}^{\mathrm{T}} \cdot Q_{\mathrm{N}}^{\mathrm{V}} \mathrm{d}V$$

these are commonly called the "load stiffness matrix." The actual form of the load stiffness is very much dependent on the type of load being considered.

The complete Jacobian matrix is then:

$$K_{\mathrm{MN}} = \int_{V^0} \boldsymbol{\beta}_{\mathrm{M}} : \boldsymbol{H} : \boldsymbol{\beta}_{\mathrm{N}} \mathrm{d}V^0 + \int_{V^0} \boldsymbol{\tau}^{\mathrm{c}} : \partial_{\mathrm{N}} \boldsymbol{\beta}_F \mathrm{d}V^0 - \int_{\mathrm{s}} N_{\mathrm{M}}^{\mathrm{T}} : Q_{\mathrm{N}}^{\mathrm{S}} \mathrm{d}S - \int_{\mathrm{V}} N_{\mathrm{M}}^{\mathrm{T}} \cdot \boldsymbol{Q}_{\mathrm{N}}^{\mathrm{V}} \mathrm{d}V \tag{1.45}$$

With the advances of various kinds of commercial finite element software, most of procedures described above have been automated. The challenges arise when the general purpose finite element software is used to solve specific problems such as in microelectronics with advanced analysis techniques and methods.

1.2.2 Nonlinear Solution Methods

The finite element modeling usually relates to nonlinear problems and can involve from a few to thousands of variables. In terms of these variables the equilibrium equations obtained by discretizing the virtual work equation can be written symbolically as:

$$F^{\mathrm{N}}(u^{\mathrm{M}}) = 0 \tag{1.46}$$

where F^{N} is the force component conjugate to the N^{th} variable in the problem and u^{M} is the value of the M^{th} variable. The basic problem is to solve Equation 1.46 for the u^{M} throughout the history of interest.

Many of the problems are nonlinear and history-dependent, so the solution must be developed by a series of "small" increments. Two issues arise: how the discrete equilibrium statement Equation 1.46 is to be solved at each increment, and how the increment size is chosen.

Newton's method is generally used as a numerical technique for solving the nonlinear equilibrium equations. The motivation for this choice is primarily the convergence rate obtained by using Newton's method compared to the convergence rates exhibited by alternate methods, usually modified Newton or quasi-Newton methods for the types of nonlinear problems. The basic algorithm of Newton's method is as follows. Assume that, after an iteration i, an approximation u_i^{M}, to the solution has been obtained. Let c_{i+1}^{M} be the difference between this solution and the exact solution to the discrete equilibrium Equation 1.46. This means that:

$$F^{\mathrm{N}}\left(u_i^{\mathrm{M}} + c_{i+1}^{\mathrm{M}}\right) = 0 \tag{1.47}$$

Expanding the left-hand side of this equation in a Taylor series about the approximate solution then gives:

$$F^{\mathrm{N}}(u_i^{\mathrm{M}}) + \frac{\partial F^{\mathrm{N}}}{\partial u^{\mathrm{P}}}\left(u_i^{\mathrm{M}}\right) c_{i+1}^{\mathrm{P}} + \frac{\partial^2 F^{\mathrm{N}}}{\partial u^{\mathrm{P}} \partial u^{\mathrm{Q}}}\left(u_i^{\mathrm{M}}\right) c_{i+1}^{\mathrm{P}} c_{i+1}^{\mathrm{Q}} + \cdots = 0 \tag{1.48}$$

If u_i^M is a close approximation to the solution, the magnitude of each c_{i+1}^M will be small, and so all but the first two terms above can be neglected giving a linear system of equations:

$$K_i^{\mathrm{NP}} c_{i+1}^{\mathrm{P}} = -F_i^{\mathrm{N}} \tag{1.49}$$

where:

$$K_i^{\mathrm{NP}} = \frac{\partial F^{\mathrm{N}}}{\partial u^{\mathrm{P}}}\left(u_i^{\mathrm{M}}\right) \tag{1.50}$$

is the Jacobian matrix and:

$$F_i^{\mathrm{N}} = F^{\mathrm{N}}\left(u_i^{\mathrm{M}}\right) \tag{1.51}$$

the next approximation to the solution is then:

$$u_{i+1}^{\mathrm{M}} = u_i^{\mathrm{M}} + c_{i+1}^{\mathrm{M}} \tag{1.52}$$

and the iteration continues.

Convergence of Newton's method is best measured by ensuring that all entries in F_i^{N} and all entries in c_{i+1}^{N} are sufficiently small. Both these criteria are checked by default in a solution.

Newton's method is usually avoided in large finite element codes, apparently for two reasons. First, the complete Jacobian matrix is sometimes difficult to formulate; and for some problems it can be impossible to obtain this matrix in closed form, so it must be calculated numerically—an expensive (and not always reliable) process. Secondly, the method is expensive per iteration, because the Jacobian must be formed and solved in each iteration. The most commonly used alternative to Newton is the modified Newton method, in which the Jacobian in Equation 1.50 is recalculated only occasionally (or not at all, as in the initial strain method of simple contained plasticity problems). This method is attractive for mildly nonlinear problems involving softening behavior (such as contained plasticity with monotonic straining) but is not suitable for severely nonlinear cases.

Another alternative is the quasi-Newton method, in which Equation 1.50 is symbolically rewritten:

$$c_{i+1}^{\mathrm{P}} = -\left[K_i^{\mathrm{NP}}\right]^{-1} F_i^{\mathrm{N}} \tag{1.53}$$

and the inverse Jacobian is obtained by an iteration process.

There are a wide range of quasi-Newton methods. The more appropriate methods for structural applications appear to be reasonably well behaved in all but the most extremely nonlinear cases—the trade-off is that more iterations are required to converge, compared to Newton. While the savings in forming and solving the Jacobian might seem large, the savings might be offset by the additional arithmetic involved in the residual evaluations (that is, in calculating the F_i), and in the cascading vector transformations associated with the quasi-Newton iterations. Thus, for some practical cases quasi-Newton methods are more economic than full Newton, but in other cases they are more expensive.

When any iterative algorithm is applied to a history-dependent problem, the intermediate, non-converged solutions obtained during the iteration process are usually not on the actual solution path; thus, the integration of history-dependent variables must be performed completely over the increment at each iteration, and not obtained as the sum of integrations associated with each Newton iteration, c_i. This is done by assuming that the basic nodal variables, u, vary linearly over the increment, so that:

$$u(\tau) = \left(1 - \frac{\tau}{\Delta t}\right) u(t) + \frac{\tau}{\Delta t} u(t + \Delta t) \tag{1.54}$$

where $0 \leq \tau \leq \Delta t$ represents "time" during the increment. Then, for any history-dependent variable, $g(t)$, we compute:

$$g(t + \Delta t) = g(t) + \int_t^{t+\Delta t} \frac{\mathrm{d}g}{\mathrm{d}\tau}(\tau)\mathrm{d}\tau \tag{1.55}$$

at each iteration.

The issue of choosing suitable time steps is a difficult problem to resolve. First of all, the considerations are quite different in static, dynamic, or diffusion cases. It is always necessary to model the response as a function of time to some acceptable level of accuracy. In the case of dynamic or diffusion problems, time is a physical dimension for the problem and the time stepping scheme must provide suitable steps to allow accurate modeling in this dimension. Even if the problem is linear, this accuracy requirement imposes restrictions on the choice of time step. In contrast, most static problems have no imposed time scale, and the only criterion involved in time step choice is

accuracy in modeling nonlinear effects. In dynamic and diffusion problems it is exceptional to encounter discontinuities in the time history, because inertia or viscous effects provide smoothing in the solution. However, in static cases sharp discontinuities (such as bifurcations caused by buckling) are common. Softening systems, or unconstrained systems, require special consideration in static cases but are handled naturally in dynamic or diffusion cases. Thus, the considerations upon which time step choice is made are quite different for the three different problem classes.

Both "automatic" time step choice and direct user control for all classes of problems are provided. Direct user control can be useful in cases where the problem behavior is well understood (as might occur when the user is carrying out a series of parameter studies) or in cases where the automatic algorithms do not handle the problem well. However, the automatic schemes are based on extensive experience with a wide range of problems. Therefore, it generally can provide a reliable approach.

One other ingredient in this algorithm is that a minimum increment size is specified, which prevents excessive computation in cases where buckling, limit load, or some modeling error causes the solution to stall. This control is handled internally, with user override if needed. Several other controls are built into the algorithm; for example, it will cut back the increment size if an element inverts due to excessively large geometry changes. These detailed controls are based on empirical testing.

In dynamic analysis when implicit integration is used, the automatic time stepping is based on the concept of half-step residuals. The basic idea is that the time stepping operator defines the velocities and accelerations at the end of the step $(t+\Delta t)$ in terms of displacement at the end of the step and conditions at the beginning of the step. Equilibrium is then established at $(t+\Delta t)$ which ensures an equilibrium solution at the end of each time step and, thus, at the beginning and end of any individual time step. However, these equilibrium solutions do not guarantee equilibrium throughout the step. The time step control is based on measuring the equilibrium error (the force residuals) at some point during the time step, by using the integration operator, together with the solution obtained at $(t+\Delta t)$, to interpolate within the time step. The evaluation is performed at the half step $(t+\Delta t/2)$. If the maximum entry in this residual vector—the maximum "half-step residual"—is greater than a user-specified tolerance, the time step is considered to be too big and is reduced by an appropriate factor. If the maximum half-step residual is sufficiently below the user-specified tolerance, the time step can be increased by an appropriate factor for the next increment. Otherwise, the time step is deemed adequate. The algorithm is somewhat more complicated at traumatic events such as impact. Here, the time step can also be adjusted based on the magnitude of residuals in the first or second iteration following such events. Clearly, if these residuals are several orders of magnitude greater than those permitted, convergence is unlikely and the time step is altered immediately to avoid unproductive iteration.

1.2.3 Advanced Modeling Techniques in Finite Element Analysis

(i) Sub-Modeling

The submodeling technique is used to study a local part of a model with a refined mesh based on interpolation of the solution from an initial, relatively coarse, global model; it is most useful when it is necessary to obtain an accurate, detailed solution in a local region and the detailed modeling of that local region has negligible effect on the overall solution and can use a combination of linear and nonlinear procedures.

The submodel is run as a separate analysis from the global analysis. The only link between the submodel and the global model is the transfer of the time-dependent values of variables saved in the global analysis to the relevant boundary nodes of the submodel (Figure 1.4). This transfer is accomplished by saving the results from the global model in the results, then reading these results into the submodel analysis. Since the submodel is a separate analysis, submodeling can be used to any number of levels; a submodel can be used as the global model for a subsequent submodel. The results from the global model are interpolated onto the nodes on the appropriate parts of the boundary of the submodel. Thus, the response at the boundary of the local region is defined by the solution for the global model. The driven nodes and any loads applied to the local region determine the solution in the submodel.

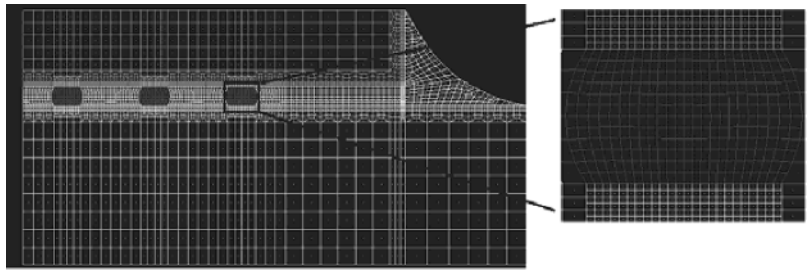

Figure 1.4 Relevant boundary nodes of the submodel

The global model in a submodeling analysis must define the submodel boundary response with sufficient accuracy. It is the user's responsibility to ensure that any particular use of the submodeling technique provides physically meaningful results. In general, the solution at the boundary of the submodel must not be altered significantly by the different local modeling.

(ii) Sub-Structure Modeling

Substructures are collections of elements from which the internal degrees of freedom have been eliminated. Retained nodes and degrees of freedom are those that will be recognized externally at the usage level (when the substructure is used in an analysis), and they are defined during generation of the substructure. System matrices (stiffness, mass) are small as a result of substructuring. Subsequent to the creation of the substructure, only the retained degrees of freedom and the associated reduced stiffness (and mass) matrix are used in the analysis until it is necessary to recover the solution internal to the substructure.

Substructuring can isolate possible changes outside substructures to save time during re-analysis. During the design process large portions of the structure will often remain unchanged; these portions can be isolated in a substructure to save the computational effort involved in forming the stiffness of that part of the structure.

In a problem with local nonlinearities, such as a model that includes interfaces with possible separation or contact, the iterations to resolve these local nonlinearities can be made on a very much reduced number of degrees of freedom if the substructure capability is used to condense the model down to just those degrees of freedom involved in the local nonlinearity.

Substructuring provides a systematic approach to complex analyses. The design process often begins with independent analyses of naturally occurring substructures. Therefore, it is efficient to perform the final design analysis with the use of substructure data obtained during these independent analyses.

Many practical structures are so large and complex that a finite element model of the complete structure places excessive demands on available computational resources. Such a large linear problem can be solved by building the model, substructure by substructure, and stacking these level by level until the whole structure is complete and then recovering the displacements and stresses locally, as required.

(iii) Adaptive Mesh Generation

Adaptive meshing is a tool that makes it possible to maintain a high-quality mesh throughout an analysis, even when large deformation or loss of material occurs, by allowing the mesh to move independently of the material. Adaptive meshing does not alter the topology (elements and connectivity) of the mesh, which implies some limitations on the ability of this method to maintain a high-quality mesh upon extreme deformation. Adaptive re-meshing is typically used for accuracy control, although it can also be used for distortion control in some situations. The adaptive re-meshing process involves the iterative generation of multiple dissimilar meshes to determine a single, optimized mesh that is used

throughout an analysis. The goal of adaptive re-meshing is to obtain a solution that satisfies mesh discretization error indicator targets that you set, while minimizing the number of elements and, hence, the cost of the analysis.

(iv) Element Removal and Reactivation

Specified elements are removed from the model in a general analysis step. Just prior to the removal step, the forces/fluxes on that the region to be removed is exerting on the remaining part of the model at the nodes on the boundary between them, are stored. These forces are ramped down to zero during the removal step; therefore, the effect of the removed region on the rest of the model is completely absent only at the end of the removal step. The forces are ramped down gradually to ensure that element removal has a smooth effect on the model.

Care must be taken in removing elements in transient procedures. The nodal flux that the removed elements apply at the boundary with the rest of the model is ramped down over the step. In transient heat transfer or fully coupled temperature-displacement analysis if the fluxes are high and the step is long, this ramping down may have the effect of cooling down or heating up the rest of the body. In dynamic analysis if the forces are high and the step is long, kinetic energy can be imparted to the remaining portion of the model. This problem can be avoided by removing the elements in a very short transient step prior to the rest of the analysis. This step can be done in a single increment.

Two distinct types of reactivation are provided for stress/displacement elements: strain-free reactivation and reactivation with strain. Strain-free reactivation resets the initial configuration; reactivation with strain does not.

Although elements cannot be created within an analysis, a similar effect can be achieved by creating elements in the model definition, removing them in the first step, and subsequently reactivating them.

When stress/displacement elements are reactivated in a strain-free state, they become fully active immediately at the moment of reactivation. They are reset to an "annealed" state in the configuration in which they lie at the start of the reactivation step. This configuration depends on whether a small- or large-displacement analysis is being conducted. Alternatively, reactivation in a non virgin state can be specified, as described below.

Since these elements are reactivated in a virgin state that is, with zero stress, they exert zero nodal forces on the rest of the model. This result allows reactivation to be done immediately, without an adverse effect on the smoothness of the solution.

After reactivation the strains and the deformation gradients are based on the displacements subsequent to the moment of reactivation, rather than on their total displacements. Thus, the current configuration at the start of the reactivation step is the new initial configuration for the element.

This kind of reactivation usually is used to model the creation of an undeformed and unstrained region of the model that is sharing a boundary with another, possibly stressed, deformed region. For example, in tunnel excavation an unstressed tunnel liner is added to line the walls of an already deformed tunnel.

(v) Multi-physics Coupling Analysis

A coupled-field analysis is a combination of analyses from different engineering disciplines (physics fields) that interact to solve a global engineering problem. Hence, we often refer to a coupled-field analysis as a multi-physics analysis. When the input of one field analysis depends on the results from another analysis, the analyses are coupled.

Some analyses can have one-way coupling. For example, in a thermal stress problem, the temperature field introduces thermal strains in the structural field, but the structural strains generally do not affect the temperature distribution. Thus, there is no need to iterate between the two field solutions. More complicated cases involve two-way coupling. A piezoelectric analysis in an MEMS structure, for example, handles the interaction between the structural and electric fields: it solves the voltage distribution due to applied displacements, or vice versa. In a fluid-structure interaction problem, the fluid pressure causes the structure to deform, which in turn causes the fluid solution to change. This problem requires iterations between the two physics fields for convergence.

The coupling between the fields can be accomplished by either direct or indirect (load transfer) coupling. Coupling across fields can be complicated because different fields may be the solver for different types of analyses during a simulation. For example, in an induction heating problem,

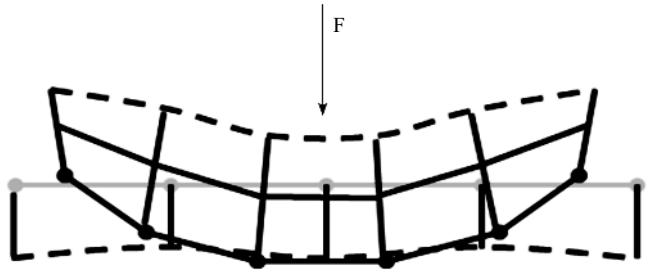

Figure 1.5 Contact pair with interpenetration

a harmonic electromagnetic analysis calculates Joule heating, which is used in a transient thermal analysis to predict a time-dependent temperature solution. The induction heating problem is complicated further by the fact that the material properties in both physics simulations depend highly on temperature.

(vi) Contact Mechanics

When two separate surfaces touch each other such that they become mutually tangent, they are said to be in contact. In the common physical sense, surfaces that are in contact have these characteristics:

- They do not interpenetrate.
- They can transmit compressive normal forces and tangential friction forces.
- They often do not transmit tensile normal forces. They are therefore free to separate and move away from each other.

Contact is a changing-status nonlinearity. That is, the stiffness of the system depends on the contact status, whether parts are touching or separated. Physical contacting bodies do not interpenetrate. Therefore, the finite element program must establish a relationship between the two surfaces to prevent them from passing through each other in the analysis. Figure 1.5 shows the bad contact model with interpenetration. When the program prevents interpenetration, we say that it enforces contact compatibility. Normally, good finite element software should offer several different contact algorithms to enforce compatibility at the contact interface. There are typically three types of algorithms: Augmented Lagrangian, Pure Penalty, and Normal Lagrange Multiplier. Penalty-based methods formulate contact as $[K]\{x\}$, so that there is a concept of contact stiffness[K] and some allowable penetration$\{x\}$. Normal Lagrange solves contact pressure as a DOF directly, so that there is no contact stiffness or penetration, although the solver selection becomes limited because of the unique formulation. Friction describes the tangential behavior between two moving parts. In addition, contact with friction for the tangential or sliding behavior of two contacting bodies is very important in engineering. With friction defined, parts can only slide relative to one another if the tangential force exceeds the product of the normal force and coefficient of friction. Therefore, contact with friction is a function that the advanced finite element should include and carefully consider in the modeling and simulation.

1.2.4 Finite Element Applications in Semiconductor Packaging Modeling

Figure 1.6 shows the flow chart of the solving process of the finite element in the fundamental problems in semiconductor packaging.

Figure 1.7 shows the modeling and simulation mapping for the application of finite elements to various areas of the semiconductor industry. This mapping figure has presented the major functions and roles of modeling and simulation in semiconductor engineering.

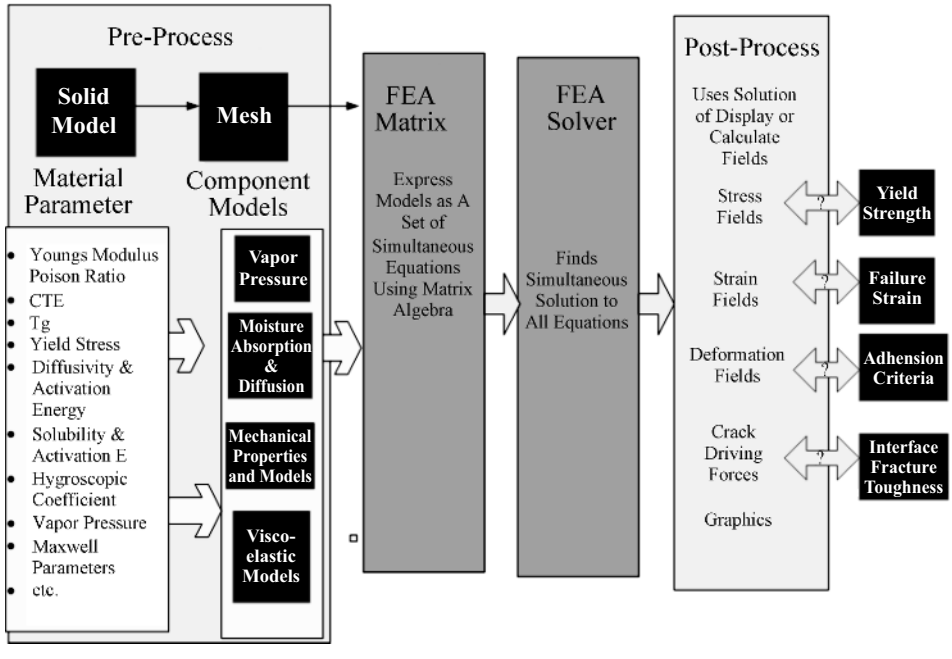

Figure 1.6 Finite element flow chart for the semiconductor packaging modeling

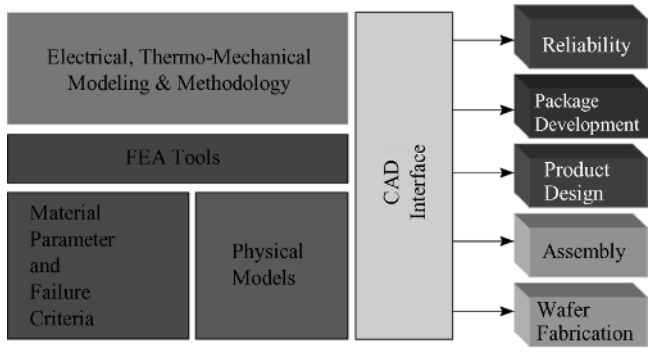

Figure 1.7 Basic mapping of the modeling and simulation in semiconductor industry

1.3 Chapter Summary

In this chapter, we summarized the basic constitutive equations for typical materials which include the linear elasticity, the elastic-visco-plasticity, and the related radial return evolution as well as the consistent tangent operator based on the theory of continuum mechanics. The elastic-visco-plastic consistent tangent operator based constitutive relation is a unified material constitutive model that has been applied to various nonlinear problems in computational mechanics. Regular elasticity and plasticity are its special limit cases. It has been proven to be very powerful in dealing with various highly nonlinear and large deformation problems in engineering. Then the finite element method was briefly presented starting from

the virtual work principle and shape functions. Nonlinear iteration solutions based on Newton's theory were also discussed. Advanced modeling techniques such as sub-modeling, sub-structure modeling, adaptive mesh generation, element removal and reactivation, multi-physics coupling analysis, and contact mechanics are presented. Finally, the flow chart of the solving process of the finite element for semiconductor packaging was presented and discussed. Modeling and simulation mapping for the application of finite elements to various areas of the semiconductor industry was also provided.

References

Belytschko, T., Liu, W.K. and Moran, B. (2000) *Nonlinear Finite Elements for Continua and Structures*, John Wiley and Sons Ltd, New York.

Fung, Y.C. (1965) *Foundation of Solid Mechanics*, Prentice-Hall, New Jersey.

Gittus, J. (1975) *Viscoelasticity and Creep Fracture in Solids*, John Wiley-Halsted Press, New York.

Hibbit, H.D., Karlsson B.I. and Sorensen. *ABAQUS Theory Manual*, Version 6.8.

Hill, R. (1982) *Mathematical Theory of Plasticity*, Pergamon Press, Oxford.

Khan, A. and Huang, S. (1995) *Contimuum Theory of Plasticity*, Wiley, New York.

Liu, S. (1992) Damage Mechanics of Cross-ply Laminates Resulting from Transverse Concentrated Loads, Ph.D. dissertation, Mechanical Engineering Department, Stanford University, CA, Aug. 1992.

Liu, Y. and Antes, H. (1999) An improved implicit algorithm for elastic viscoplastic boundary element method, *Zeitschrift für Angewandte Mathematik und Mechanik* **79**:317–333.

Maniatty, A.M. and Liu, Y. (2003) Stabilized finite element method for viscoplastic flow: formulation with state variable evolution, *International Journal of Numerical Methods in Engineering* **56**:185–209.

Paulino, G.H. and Liu, Y. (2000) Implicit consistent and continuum tangent operators in elastoplastic boundary element formulations, *Computational Methods in Applied Mechanics and Engineering* **190**:2157–2179.

Perzyna, P. (1971) Thermodynamic theory of viscoplasticity, *Advances in Applied Mechanics* **11**:313–354. Academic Press, New York.

Simo, J.C. and Taylor, R.L. (1985) Consistent tangent operators for rate-independent elastoplasticity, *Computer Methods in Applied Mechanics and Engineering* **48**:101–118.

Simo, J.C. and Hughes, T.J.R. (1997) *Computational Inelasticity*, Springer, New York.

Timoshenko, S.P. and Goodier, J.N. (1970) *Theory of Elasticity*, McGraw-Hill, New York.

Zhang, G.Q., van Driel, W.D. and Fan, X.J. (2006) *Mechanics of Microelectronics*, Springer, New York.