6

Virtual Thermo-Mechanical Prototyping of Microelectronics and Microsystems

A. Wymysłowski[a], G.Q. Zhang[b], W.D. van Driel[b], and L.J. Ernst[c]

[a]*Wroclaw University of Technology, Poland*
[b]*Delft University of Technology, and Philips Semiconductors, The Netherlands*
[c]*Delft University of Technology, The Netherlands*

6.1. INTRODUCTION

The technological trends of microelectronics and microsystems are mainly characterized by miniaturization down to nano-scale, increasing levels of technology and function integration and eco-designing, while the business trends are mainly characterized by cost reduction, short-time-to-market and outsourcing. These trends together lead to increased chances and consequences of failures, increased design complexity, dramatically decreased design margins and increased difficulty to meet quality, robustness, reliability and shorter-time-to-market requirements. Among others, it is found that thermo-mechanical (thermal, mechanical and thermo-mechanical) related failures account for about 65% of total failures in microelectronics. Thermo-mechanical reliability is becoming one of the major concerns for both current and future microelectronics technologies.

Based on the root cause analyses from observed failures of microelectronics during different life cycles, it is clear that most of the thermo-mechanical reliability problems originate from the product/process design phase. However, within electronics industry, microelectronics design and qualification are still largely depending on one's experience. Often, up to 10 cycles (material development/pre-selection, concept designing, building and testing multiple physical prototypes) are needed, with some qualitative support from numerical simulations. Quality, robustness and reliability are usually dealt with after-physical prototyping, wherein reliability qualification testing with duration of 6 months is no exception. Clearly, this experience-based design and qualification method cannot lead to competitive design with short time-to-market, optimized performance, low costs, and guaranteed quality, robustness and reliability. Therefore, there is an urgent need to develop innovative thermo-mechanical design method—virtual thermo-mechanical prototyping method.

For virtual thermo-mechanical prototyping, "accurate and efficient prediction models" and "advanced simulation-based optimization methods" are the two core building

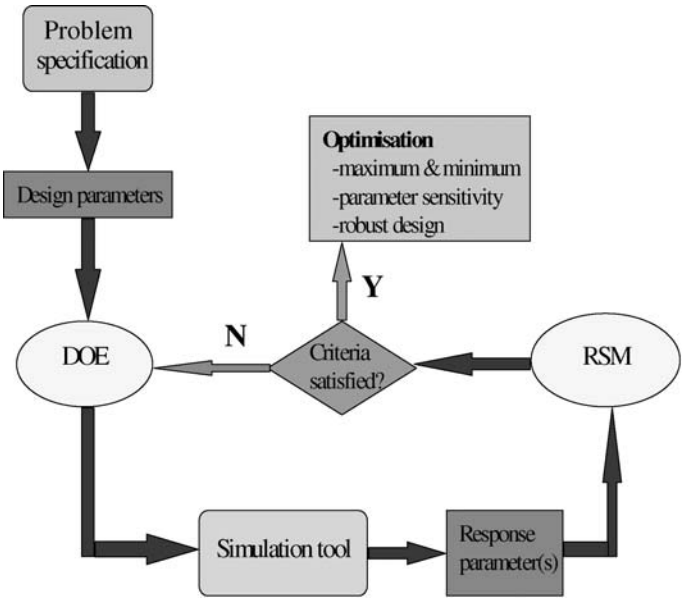


FIGURE 6.1. Typical virtual prototyping diagram.

blocks. By combining these two building blocks in a proper way, one can predict, qualify and optimize the thermo-mechanical behavior and/or trends of microelectronics against the actual requirements prior to major physical prototyping, manufacturing investments and reliability qualification tests. A typical virtual prototyping procedure would consist of the following steps (see Figures 6.1 and 6.2):

- building up a numerical model (such as FEM) of a product, which will be capable of predicting the behavior of the product under given loading conditions,
- carrying out a Design of Experiments (DoE) procedure in order to find out the correlation between the response and input factors (including interactions) and defining their significance in a sense of e.g., mean and variance,
- performing a Response Surface Modeling (RSM) that would approximate the model of the response in a form of a response surface reflecting relationship between the response and significant input factors,
- final stage should be directed towards design optimization, such as finding minimum/maximum, robust design and tolerance design [1,6,39].

6.2. PHYSICAL ASPECTS FOR NUMERICAL SIMULATIONS

In most of the engineer applications we are faced with the field theory problem of continuum domains, such as:

- Thermal fields.
- Mechanical/ structural fields.
- Electromagnetic fields.
- Fluid flow.

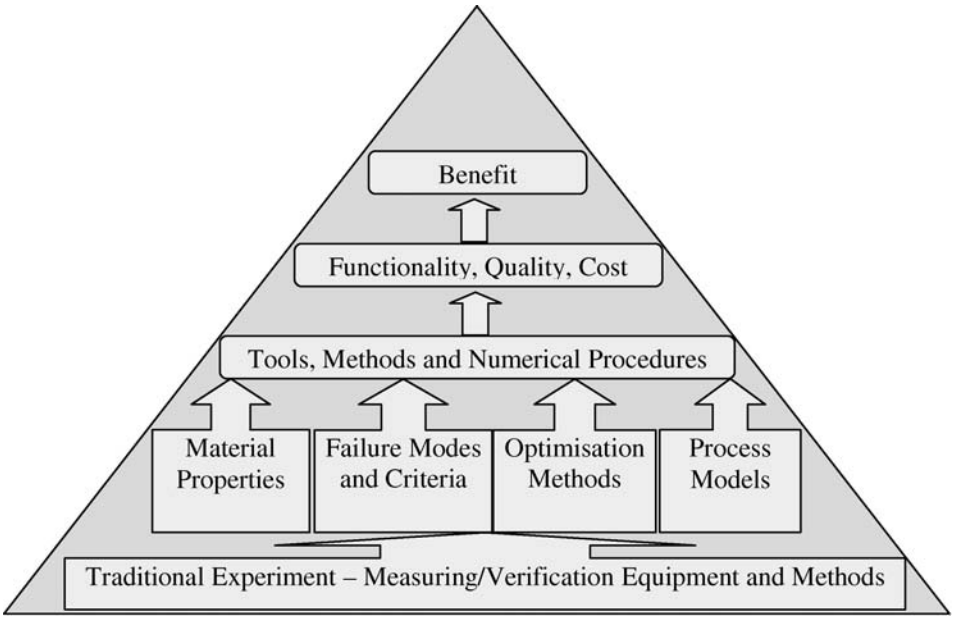


FIGURE 6.2. Advanced virtual prototyping structure.

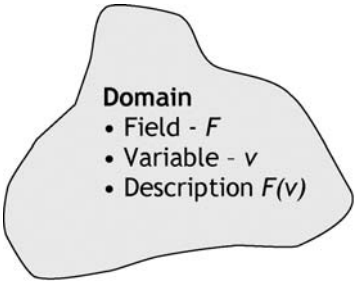


FIGURE 6.3. Continuum domain with defined physical field and its variable.

The continuum can be defined as a part of domain (e.g., solid state, fluid or gas) where the physical phenomenon is defined. The domain is defined by a selected physical field F and its variable v (e.g., thermal field and temperature T), which can receive any finite value in arbitrary point in the domain:

$$v = F(x, y, z). \tag{6.1}$$

The field variable v can receive infinite values in the domain as it is a function of field point coordinates x, y, z, see Figure 6.3.

Most of the problems in engineering practice, concerning field theory, can be described by partial differential equations (PDE). As the analytical solutions of most engineering problems seldom exit, more and more attention is directed towards numerical solutions rather than exact closed-form analytical solutions. The basic problem concerning numerical solution is the accuracy. It is a rule of thumb that numerical simulation is as

precise as the input data (GIGO—garbage in garbage out). There are a number of aspects that would influence accuracy, e.g.,

- Knowledge on physical phenomena.
- Material models and properties.
- Damage models and criteria.
- Product parameters.
- Process details.
- The best engineer practice techniques on: boundary conditions, loads, etc.

6.2.1. Numerical Modeling

Over the years, several numerical analysis methods for solving PDE equations have evolved. The most common ones are: Finite Difference Method (FDM) or Finite Element Method (FEM); the main difference is in representing complex geometrical shapes. Any numerical technique involves dividing the domain of solution into a finite number of simple sub-domains and allows obtaining approximate solutions to a wide variety of engineering problems with boundary-value problems. As an example of the most popular partial differential equations applied in engineering is Poisson's formula for evaluation of electrostatic field distribution:

$$\nabla^2 V(x, y, z) = \frac{q(x, y, z)}{\varepsilon_0}, \tag{6.2}$$

where $q(x, y, z)$ is the space charge in the analyzed region while ε_0 is the vacuum dielectric constant. If space charge is equal to 0, the above formula can be rewritten as follows

$$\nabla^2 V(x, y, z) = 0, \tag{6.3}$$

where ∇^2 stands for the Laplace's operator.

FDM method is based on the assumption that differential equations can be replaced by finite differences. According to the Taylor-series expansion, if we know the function value in point x we can evaluate the function value at point $x + h$ as long as the function is continuous and has derivatives:

$$f(x + h) = f(x) + \frac{h}{1!} f'(x) + \cdots. \tag{6.4}$$

Therefore the first derivate of the function can be replaced by the finite difference as:

$$f'(x) \cong \frac{\Delta f(x)}{h}, \tag{6.5}$$

where the error (for the central difference) will be proportional to

$$error \cong \frac{1}{27} f'''(x) \cdot h^2. \tag{6.6}$$

On the other hand FEM methods use the variational concept and approximations of the solution over the collection of finite elements. It is a numerical method for solving problems

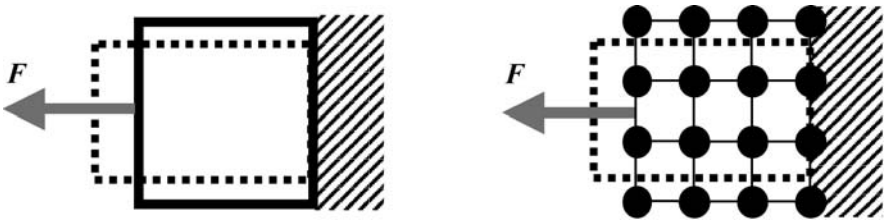


FIGURE 6.4. Problem of mechanical deformation due to force F. Example of the Finite Element approach; nodal points interconnect elements.

of engineering and mathematical physics and is especially useful for problems with complicated geometries, loadings, and material properties where analytical solutions cannot be obtained. Basically FEM method is applied in such applications as: structural analysis, fluid flow, heat transfer, electro-magnetic fields, acoustics, etc. A geometrically similar model consisting of multiple, linked, simplified representations of discrete regions, i.e., finite elements, represents the system. These units (finite elements) are interconnected at points common to two or more elements (nodes or nodal points), boundary lines and surfaces. An example problem of mechanical deformation due to the applied force F is presented in Figure 6.4.

The key problem in FEM method is proper definition of element properties, which in the above case can be defined as a spring constant. The spring constant or stiffness binds the applied force F and nodal displacement d according to the equation:

$$F = k \cdot d. \tag{6.7}$$

In fact, FEM method is based on defining the local $[\boldsymbol{k}]$ and global stiffness $[\boldsymbol{K}]$ matrix for one and number of elements respectively. The goal is to evaluate nodal displacements d_i. In case of one element the following governing equation applies:

$$\begin{aligned} F_1 &= kd_1 - kd_2, \\ F_2 &= -kd_1 + kd_2^2, \end{aligned} \tag{6.8}$$

which can be rewritten in a matrix form as:

$$\begin{bmatrix} k & -k \\ -k & k \end{bmatrix} \begin{Bmatrix} d_1 \\ d_2 \end{Bmatrix} = \begin{Bmatrix} F_1 \\ F_2 \end{Bmatrix}, \tag{6.9}$$

or

$$[\boldsymbol{k}]\{d\} = \{F\}, \tag{6.10}$$

where $[\boldsymbol{k}]$ can be referred as the local stiffness matrix.

In case of two elements the governing equation is given by (see Figure 6.5):

$$\begin{aligned} F_1 &= kd_1 - kd_2, \\ F_2 &= -kd_1 + kd_2 + kd_2 - kd_3, \\ F_3 &= -kd_2 + kd_3, \end{aligned} \tag{6.11}$$

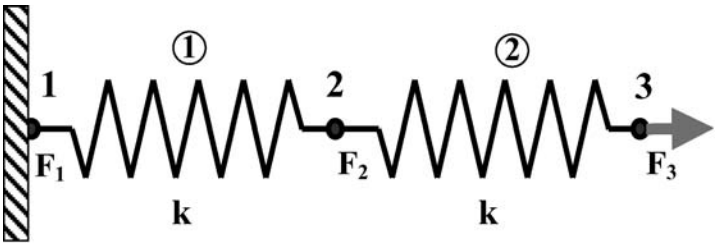


FIGURE 6.5. Global stiffness matrix.

$$\begin{Bmatrix} F_1 \\ F_2 \\ F_3 \end{Bmatrix} = \begin{bmatrix} k & -k & 0 \\ -k & k+k & -k \\ 0 & -k & k \end{bmatrix} \begin{Bmatrix} d_1 \\ d_2 \\ d_3 \end{Bmatrix}$$

FIGURE 6.6. Construction of the global stiffness matrix.

which can be rewritten in a matrix form as:

$$\begin{Bmatrix} F_1 \\ F_2 \\ F_3 \end{Bmatrix} = \begin{bmatrix} k & -k & 0 \\ -k & k+k & -k \\ 0 & -k & k \end{bmatrix} \begin{Bmatrix} d_1 \\ d_2 \\ d_3 \end{Bmatrix}, \tag{6.12}$$

or

$$[\boldsymbol{K}]\{d\} = \{F\}, \tag{6.13}$$

where $[\boldsymbol{K}]$ can be referred as the global stiffness matrix. In fact the global stiffness matrix can be constructed from local stiffness matrices according to the element interconnections. Figure 6.6 shows visualization of the stiffness matrix construction.

Solution methods in FEM are based on finding the equilibrium of minimum potential energy. The total potential energy E_p is defined as the sum of internal strain energy U_i and external U_e potential energy due to external forces $\boldsymbol{F}$:

$$E_p = U_i + U_e, \tag{6.14}$$

where

$$\begin{aligned} U_i &= \frac{1}{2}kx^2, \\ U_e &= -Fx, \end{aligned} \tag{6.15}$$

thus

$$E_p = \frac{1}{2}kx^2 - Fx. \tag{6.16}$$

Therefore the solution can be fund out by applying the variational analysis:

$$\frac{dE_p}{dx} = 0. \tag{6.17}$$

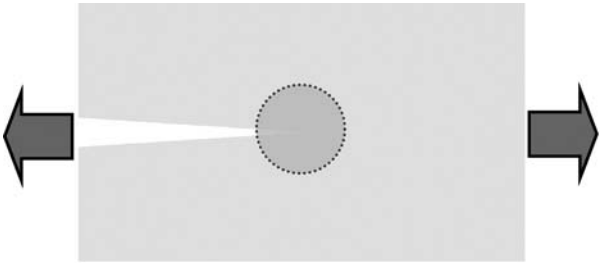


FIGURE 6.7. The singularity problem.

Basic steps involved in the numerical simulations based on FEM method can be summarized as follows [48]:

- Discretize and select element type.
- Select a displacement function.
- Define strain/displacement and stress/strain relationships.
- Derive element stiffness matrix and equations.
- Assemble equations and introduce B.C.'s (boundary conditions).
- Solve for the unknown degrees of freedom.
- Solve for element stresses and strains.
- Interpret the results.

The accuracy of numerical simulations seems to be a very important aspect, which may decide on its application. The main sources of errors are due to [8]:

- Material properties and models.
- FEM method.
- Simplification of multiphysics phenomena.
- Geometry accuracy and details.
- Mesh density and type.
- Time step for transient analysis.

The numerical modeling requires reducing the geometric complexity to the appropriate level of the problem—some features may be omitted. Sometimes the simplification can lead to singularity problems, e.g., concerning stresses at sharp edges due to infinite stress concentration, see Figure 6.7.

The evaluated value at the point depends on e.g., mesh density; however the displacements can be still good. There are a number of tasks that can be taken to avoid or minimize it:

- If the region is not important it can be ignored.
- If the region is important a detailed geometry should be included.
- The region can be modeled using sub-modeling technique.
- Extrapolating the results in the nearby region, e.g., linear.

6.2.2. Material Properties and Models

One of the basic problems of numerical modeling is connected with proper and accurate definition of material properties and models. The most common errors are due to:

- Material non-linearity.
- Thin layers.
- Temperature dependence.
- Process dependence.
- Internal discrepancies.
- Microstructure.
- Anisotropy.

The primary criterion of precise numerical modeling is knowledge on material models, characteristics and properties. Most of the engineer calculations are performed with the assumptions on elastic, linear, isotropic and isothermal behavior of materials. While this assumption can be justified in the macro scale it is not any longer applicable in the micro scale. This refers especially to microsystem and packaging of electronic devices (e.g., thin layers or grain-size). There are number of factors, which determine the description of material properties and can be described according to the following groups, as listed in Figure 6.8:

- Physical: density, melting temperature, etc.
- Mechanical: Young's modulus, shear modulus, Poisson's ratio, etc.
- Thermal: coefficient of thermal expansion, thermal conductivity, etc.
- Electromagnetic: specific resistivity, dielectric constant, etc.
- Acoustic: compression wave velocity, shear wave velocity, bar velocity, etc.

Material behavior will depend on:

- Type: metal, polymer, ceramic, etc.
- State: solid, liquids, gas.
- Form: crystalline, amorphous.
- Model: elastic, plastic, creep.

Constitutive models are very important for accurate stress/strain analysis, especially in case of damage analysis, e.g., fatigue assessment.

$$f(\sigma, \varepsilon, \dot{\varepsilon}, T) = 0, \quad d\sigma = \left(\frac{\partial \sigma}{\partial \varepsilon}\right)_{\dot{\varepsilon},T} d\varepsilon + \left(\frac{\partial \sigma}{\partial \dot{\varepsilon}}\right)_{\varepsilon,T} d\dot{\varepsilon} + \left(\frac{\partial \sigma}{\partial T}\right)_{\varepsilon,\dot{\varepsilon}} dT. \tag{6.18}$$

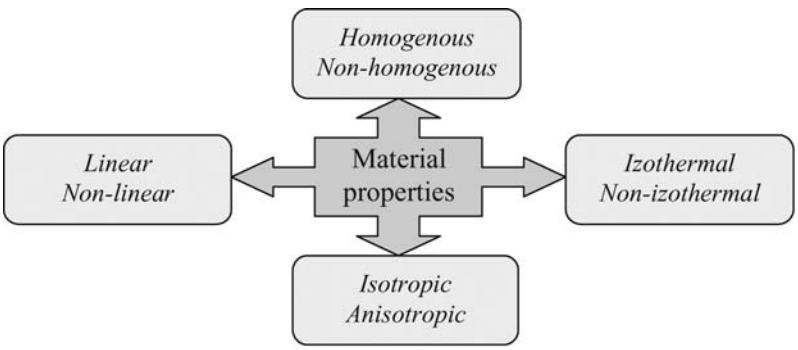


FIGURE 6.8. Material properties.

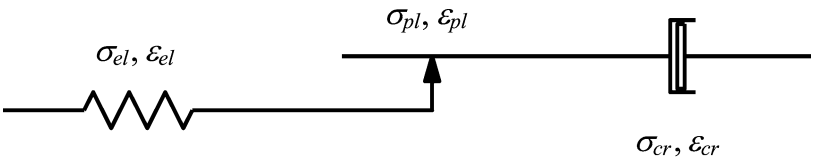


FIGURE 6.9. The combined material model.

- **Partitioned**; non-recoverable: response is modeled by separating time-dependent and time-independent behavior:

$$\varepsilon = \varepsilon_{el} + \varepsilon_{pl} + \varepsilon_{cr}. \tag{6.19}$$

- **Unified**; non-recoverable: response is modeled as a combined time-dependent manner:

$$d\sigma = E(\varepsilon, \dot{\varepsilon}, T), \tag{6.20}$$

where

$$\varepsilon = \varepsilon_{el} + \varepsilon_{inel}(\varepsilon_{pl}, \varepsilon_{cr}). \tag{6.21}$$

Actually partitioned case is applied in most of the applications and fatigue life prediction can be done by applying Miner's rule. In case of unified approach, fatigue life prediction can be achieved through the FEM analysis and is used for e.g., thermal shocks.

Any material exposed to thermo-mechanical loading conditions will deform. The total deformation can be evaluated by combining all three models: elastic, plastic and creep. In order to calculate the total strain under known loading, mathematical models of the corresponding strains should be assumed. According to the material engineering, elastic model is assumed to be linear while the plastic and creep models are nonlinear. Figure 6.9 shows the idea of this kind of partitioning:

$$\varepsilon(\sigma, t) = \varepsilon_{el}(\sigma_{el}) + \varepsilon_{pl}(\sigma_{pl}) + \varepsilon_{cr}(\sigma_{cr}, t). \tag{6.22}$$

In order to evaluate the cycle fatigue a most dominating behavior should be analyzed and the representing damage parameter taken into consideration. In fact, in case of microelectronic applications there are two dominating material behavior:

- Elasto-plasticity.
- Visco-elasticity.

6.2.2.1. Elastic Model The elastic model can be expressed by the following formula:

$$\varepsilon = \frac{1}{E}\sigma, \tag{6.23}$$

where E is a Young's modulus and ε is a strain. It is assumed that elastic model can be applied to strains $\varepsilon < 0.01$. The elastic properties are presented as an ideal string with a stiffness E. The typical characteristics of the elastic model are shown in Figure 6.10.

The main property of the elastic deformation is reversibility. This means that energy accumulated during the loading cycle can be totally recovered during the unloading. The

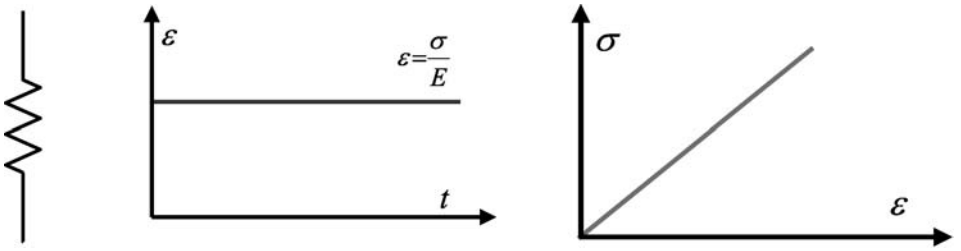


FIGURE 6.10. Characteristics of the elastic model.

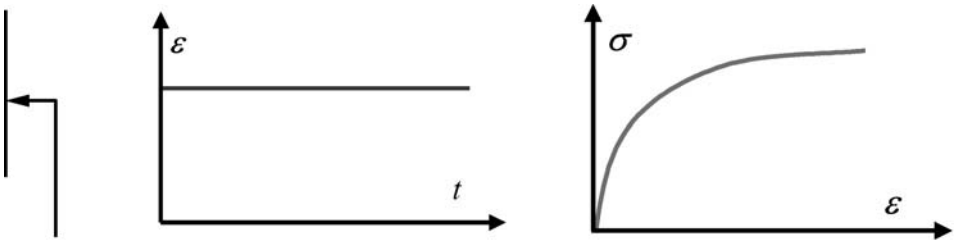


FIGURE 6.11. Characteristics of the plastic model.

same, there is no energy dissipation. The maximum stress $\sigma_{\max}$ can be applied as a representative damage parameter in case of the dominating elastic domain.

6.2.2.2. Plastic Model The plastic model can be represented by the following formula:

$$\varepsilon = \left(\frac{\sigma}{K}\right)^{\frac{1}{n}}, \tag{6.24}$$

where ε is a plastic strain, K is a strength coefficient and n is a strain hardening exponent. Plastic deformation occurs after the material is exposed to stress–strain defined as the yield point. Generally plastic domain appears for strains $\varepsilon = 0.01$–0.6. The typical characteristics of the plastic model are shown in Figure 6.11.

In comparison to elastic model, plastic deformation is not reversible. This means that during the loading cycle some of the accumulated energy is dissipated in the material in a form of a plastic deformation and it cannot be recovered during the unloading. The equivalent accumulated plastic strain $\Delta\varepsilon_{pl}$ can be applied as a representative damage parameter, in case of the dominating plastic domain.

6.2.2.3. Creep Model The creep model can be expressed by the following formula:

$$\sigma = \eta\dot{\varepsilon}, \tag{6.25}$$

where η is a viscosity and $\dot{\varepsilon} = d\varepsilon/dt$ is a rate of a strain change. From a kinematics view, creep is similar to plasticity, except that creep is a function of time. The creep rate depends strongly on temperature and stress:

$$\dot{\varepsilon} = \frac{d\varepsilon}{dt} = f(T, \sigma). \tag{6.26}$$

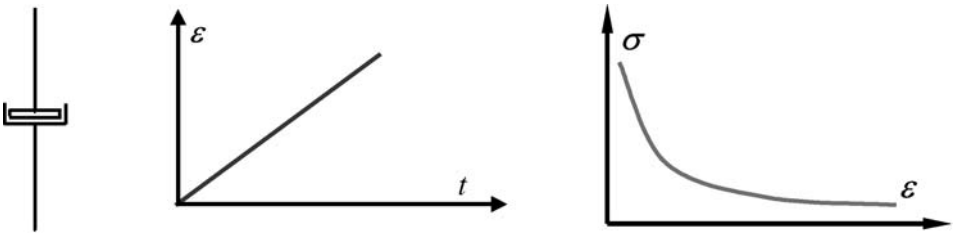


FIGURE 6.12. Characteristics of the creep model.

One of the most popular creep laws is based on time dependent inelastic behavior, described by the steady state power law:

$$\dot{\varepsilon}_{cr} = A\sigma^{\frac{1}{n}} \exp\left(\frac{-Q}{kT}\right), \tag{6.27}$$

where σ corresponds to von Mises' stress, A is a creep constant, k is a universal gas constant, Q is a creep activation energy and n is a creep strain hardening exponent. The typical characteristics of the creep model are shown in Figure 6.12.

In test at high strain, two different phenomena may be observed: cyclic creep and stress relaxation. Cyclic creep is a progressive straining under constant stress, while stress relaxation may occur under constant strain. Similarly to plastic deformation, creep deformation is not reversible as well. This means that during the loading cycle some of the accumulated energy is dissipated in the material and it cannot be recovered during the unloading. The equivalent accumulated inelastic strain energy ΔW_{in} can be applied as a representative damage parameter, in case of the dominating creep domain.

6.2.3. Thermo-Mechanical Related Failures

Typical microelectronic package consists of a number of different materials and interconnections between them. In fact, the list of possible failure modes is very long and precise analysis requires interdisciplinary knowledge including mechanics, physics, chemistry, mathematics, etc. Figure 6.13 shows the outline of a typical microelectronic package, and Figures 6.14 and 6.15 typical thermo-mechanically related failures [54].

During manufacturing and operation of microelectronic packages various thermal and mechanical loading conditions are inherited. This is the major cause of the induced failures represented in Figures 6.14 and 6.15. According to the extensive root cause analysis of failures it was concluded that:

- Thermal, mechanical and thermo-mechanical related failures may account for more than 65% of the total failure rates.
- Thermal, mechanical and thermo-mechanical related failures often originate from the product and process design phase.

Though, the list of possible failures is very long, the basic reason for thermo-mechanical failure is due to the thermal mismatch of CTE (coefficient of thermal expansion), which can be classified as, see Figure 6.16:

- Thermo-elastic.
- Plastic.

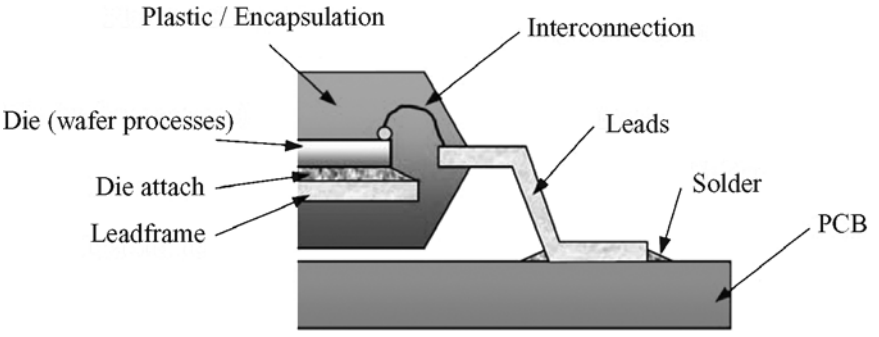


FIGURE 6.13. An example of typical microelectronic package.

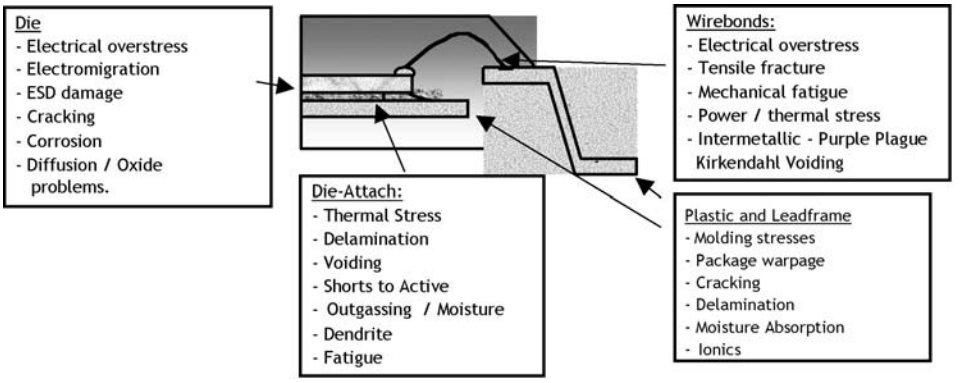


FIGURE 6.14. Failure modes for packaging level I [54].

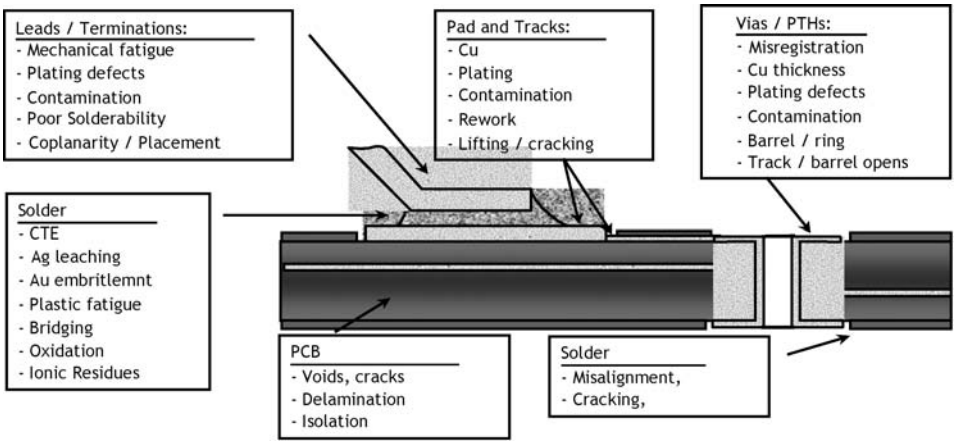


FIGURE 6.15. Failure modes for packaging level II [54].

- Creep.
- Fracture.
- Fatigue.

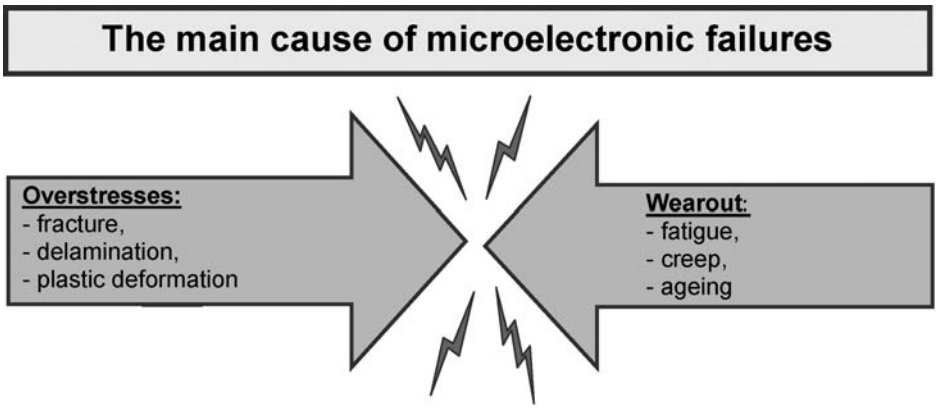


FIGURE 6.16. Thermo-mechanical related failures in microelectronics.

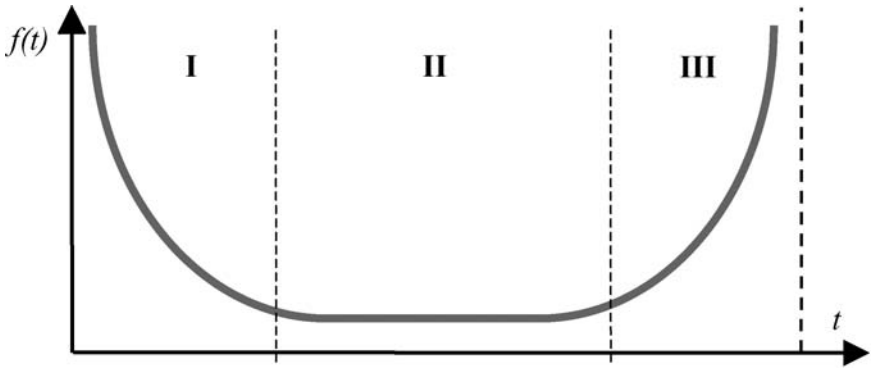


FIGURE 6.17. Failure density distribution curve.

Reliability is major concern in the microelectronic industry. Reliability is defined as *the probability that a product will survive under certain conditions during a certain period of time*. As such, failure is defined as the probability that a product is not functioning as designed. Typical measure to quantify reliability is the failure density function. The failure density function $f(t)$, or "bathtube" curve, can be divided into three parts, see Figure 6.17:

- infant mortality; this due to manufacturing defects, e.g., referred to as quality (I),
- intrinsic failure; this is due to typical user functional usage, most often caused by high-level stresses outside the design specifications, e.g., drop (II),
- wearout; gradual "wearing and tearing" in normal usage over the product life (III).

Data from tests concerning failure analysis are collected in a tabular form and most often are analyzed according to the Weibull distribution function:

$$f(t) = \frac{\beta}{\lambda} \cdot \left(\frac{t - t_0}{\lambda} \right)^{\beta - 1} \cdot e^{-(\frac{t-t_0}{\lambda})^{\beta}}, \tag{6.28}$$

where β is shape, λ is lifetime and t_0 is free lifetime. The above approach allows for reduction of the amount of data that are to be collected through experiment due to extrapolation of the approximated analytical formula:

$$F(t) = 1 - \exp\left[-\left(\frac{t}{\lambda}\right)^{\beta}\right]. \tag{6.29}$$

The product lifetime N_f is defined as:

$$f(63.2\%) \approx \lambda. \tag{6.30}$$

Figure 6.18 shows an example of the Weibull distribution.

6.2.3.1. Example Case Let's assume that we have a simple 2D package model consisting of substrate and silicon chip connected by solder bumps. The package is exposed to temperature change ΔT and the goal is to evaluate the shear stress γ. Figure 6.19 shows a schematic of the example case.

Under the assumptions on elastic material models, the strain due to temperature change is given as:

$$\varepsilon_{Si} = \alpha_{Si}\Delta T, \quad \varepsilon_{FR4} = \alpha_{FR4}\Delta T, \tag{6.31}$$

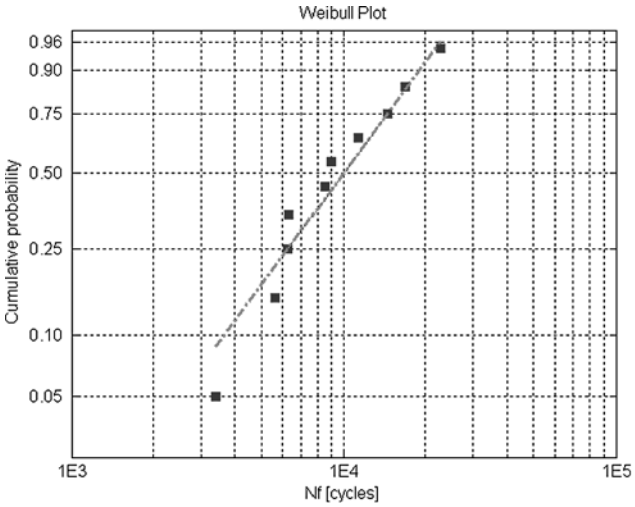


FIGURE 6.18. Weibull density distribution curve.

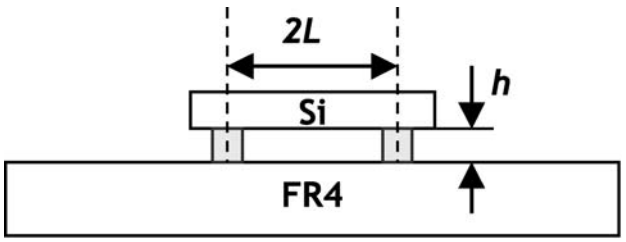


FIGURE 6.19. Thermo-mechanical related failures in microelectronics.

where the elongation at the end of the die:

$$d_{Si} = \varepsilon_{Si} L = \alpha_{Si} \Delta T L, \quad d_{FR4} = \varepsilon_{FR4} L = \alpha_{FR4} \Delta T L. \tag{6.32}$$

The shear strain in the solder can be finally evaluated as follows:

$$\gamma = \frac{d_{FR4} - d_{Si}}{h} = (\alpha_{FR4} - \alpha_{Si}) \left(\frac{L}{h} \right) \Delta T. \tag{6.33}$$

Basing on the formula it is possible to draw the final conclusions concerning the shear strain problem:

- It is proportional to:
  - CTE mismatch $(\alpha_{FR4} - \alpha_{Si})$.
  - Distance to the neutral point (DNP $- L$).
  - Temperature change ΔT.
- It is inversely proportional to height of the solder joint h.

In order to minimize the shear strain it would desirable to diminish the CTE mismatch between the interconnected materials. Unfortunately, in most of the cases it is not feasible and therefore some designing approach is required, which could improve the product reliability.

6.2.4. Designing for Reliability

The ability to assess the reliability of the product is one of the most important tasks of the design engineer. Unfortunately, reliability of the electronic packages is a complex task, which requires interdisciplinary knowledge including e.g., managing. Anyway, most often reliability problem can be presented in a form of the Ishikawa diagram by listing the causes for low reliability, see Figure 6.20.

Despite of the undisputed achievements of the reliability engineering, it is commonly known that electronic components fail. Currently there are two main approaches to reliability assessment

- Statistical: reliability is a function of time.
- Physical: reliability is a function of a component physical state.

Additionally reliability assessment can be considered as, see Figure 6.21:

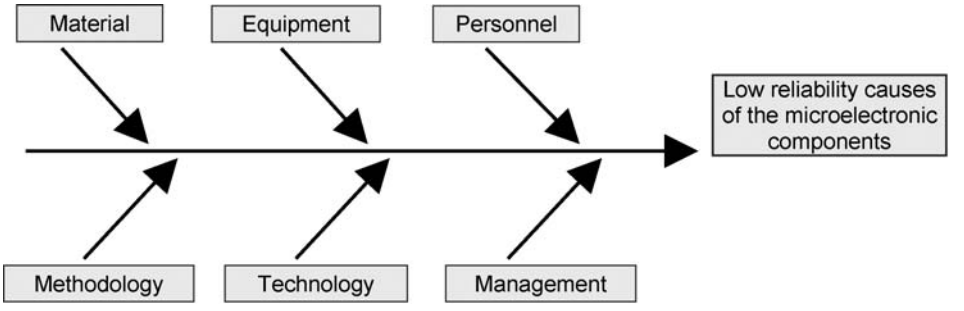


FIGURE 6.20. Ishikawa diagram for the electronic packages.

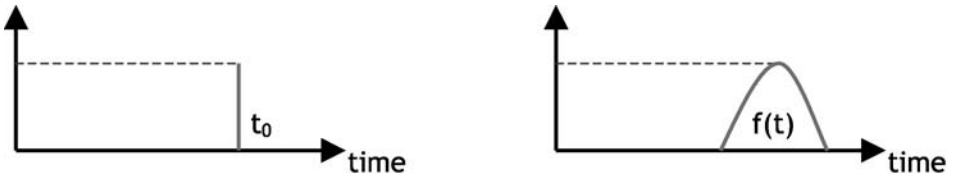


FIGURE 6.21. The difference between deterministic and probabilistic approach.

- Deterministic: input variables are precisely defined and therefore output variable would be precisely defined as well.
- Probabilistic: input variables are treated as random variables and therefore output variable has a random distribution.

Understanding the mechanisms that cause components failure is the key factor to make reliable products. The most common thermo-mechanical failure modes, which are addressed in microelectronic are due to:

- Overstress: fracture.
- Wearout: fatigue.

6.2.4.1. Fracture Fracture is probably the most dominant thermo-mechanical failure mode in microelectronic packages [47]. Fracture can cause:

- Die cracking.
- Underfill cracking.
- Solder joint failure.
- Body cracking.
- And many more.

The basic approach to fracture mechanics is based on linear elastic fracture mechanics (LEFM). The LEFM is based on the observation that near the crack tip in a brittle material, the magnitude of the singular stress field can be controlled by a single parameter. According to LEFM it is possible to evaluate the singular stress near crack tip as:

$$\sigma = \frac{K}{\sqrt{2\pi x}}, \tag{6.34}$$

where x is the distance to the crack tip, K is the stress intensity factor, which is an indication of the stress magnitude near the crack tip. Fracture will occur when:

$$K \geq K_c, \tag{6.35}$$

where K_c is called fracture toughness and is an intrinsic material property that can be measured using a number of techniques. Two distinct fracture mechanisms can be discriminated:

- Brittle fracture: separation of materials along crystallographic planes due to atomic bonds breaking.
- Ductile fracture: initiation, growth and coalescence of micro-voids.

In fact, ductile fracture is considered to be dominant mechanism in microelectronics. Figure 6.22 shows the stress–strain relation for both mechanisms.

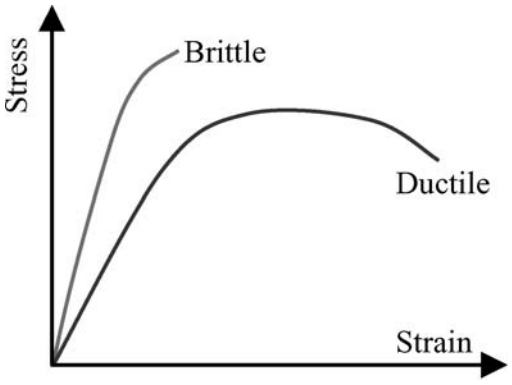


FIGURE 6.22. The difference between deterministic and probabilistic approach.

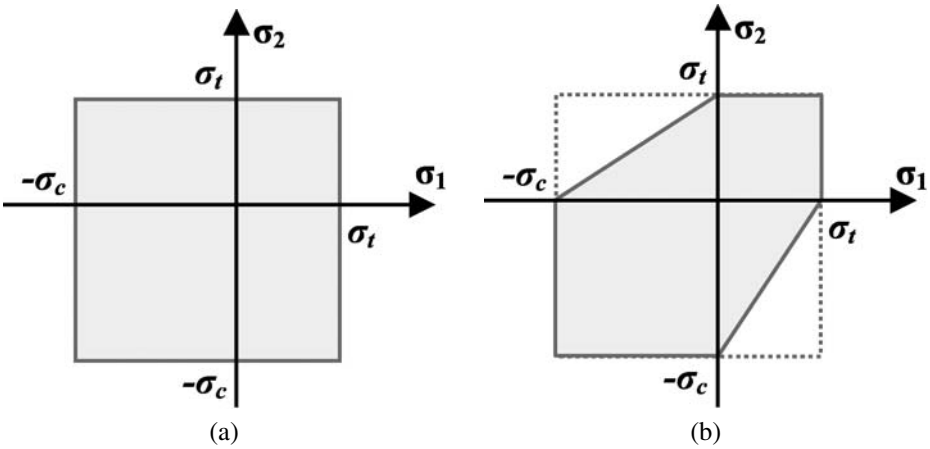


FIGURE 6.23. The fracture criteria for brittle materials: Coulomb (a) and Mohr (b).

One of the basic problems during fracture analysis is due to failure criteria. In fact, it is meant that fracture occurs when the stress level exhibits the material strength. Due to specific behavior of the both materials, failure criterion should be evaluated differently.

Brittle materials (Figure 6.23):

- The maximum stress criterion, also known as the normal stress, Coulomb, or Rankine criterion:

$$\sigma_c < \{\sigma_1, \sigma_2\} < \sigma_t, \tag{6.36}$$

  where σ_t is uniaxial tension strength and σ_c is uniaxial compression strength.
- The Mohr Theory of Failure, also known as the Coulomb-Mohr criterion or internal-friction theory, is based on the Mohr's Circle.

Ductile materials (Figure 6.24):

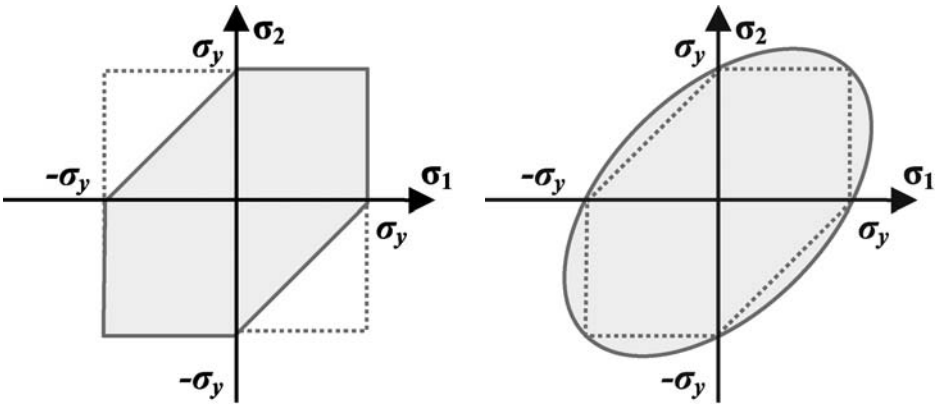


FIGURE 6.24. The fracture criteria for ductile materials: Tresca (a), von Mises (b).

- The maximum shear stress criterion, also known as Tresca's or Guest's criterion:

$$|\sigma_1| \leq \sigma_y,\ |\sigma_2| \leq \sigma_y,\ \text{and}\ |\sigma_1 - \sigma_2| \leq \sigma_y, \tag{6.37}$$

  where σ_y stands for the yield point.
- von Mises criterion, also known as the maximum distortion energy criterion:

$$\frac{1}{2}[(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2] \leq \sigma_y. \tag{6.38}$$

6.2.4.2. Fatigue In practical applications, fracture can occur even when the load amplitude is below the static strength. This is due to fatigue. Fatiguestress is a failure mode caused by the cyclic load, e.g., mechanical, thermal, etc. A commonly method for characterizing the fatigue behavior is the S/N curve, where S stands for load amplitude and N is the number of cycles to failure for the given load amplitude (Figure 6.25).

Thermal cyclic fatigue occurs under repeated applications of thermal stress, resulting from the temperature changes as well as from component materials with different thermal expansion coefficients. The primary reason of the thermal cyclic fatigue failure is fracture. Therefore studying the physics of cracks can bring some light on predicting the fatigue failure of electronic packages (Figure 6.26). Fatigue failure is mainly due to energy (damage) accumulation over cycles, which can be assessed according to the response hysteresis. There are different approaches to fatigue analysis:

- Steady-state crack growth da/dN_f: based on LEFM.
- Stress based σ: based on elastic deformation.
- Strain based ε: based on elastic and plastic deformation.
- Inelastic energy W: based on inelastic energy dissipation.

According to the material model the cyclic fatigue depends on damage parameter and can be considered as LCF (low cycle fatigue) and HCF (high cycle fatigue) domain. The LCF ($N_f < 10^5$) failure is assumed to be the inelastic strain method while the HCF ($N_f > 10^6$) is assumed to be the elastic strain method, where:

- high cycle fatigue for stress $S \ll \sigma_y$ (yield point)
- low-cycle fatigue for stress $S \gg \sigma_y$ (yield point)

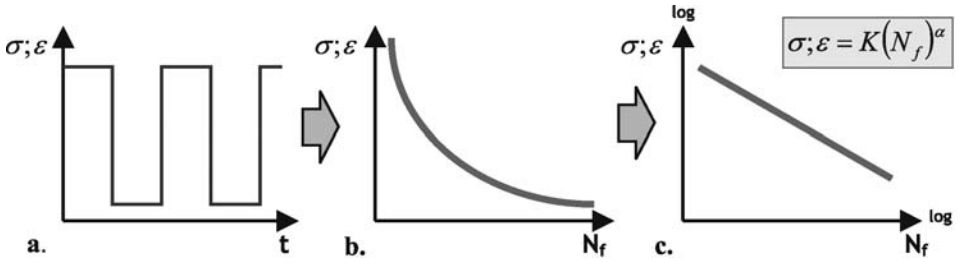


FIGURE 6.25. Failure analysis due to fatigue; cyclic load (a), S/N curve (b), S/N curve in log–log scale (c).

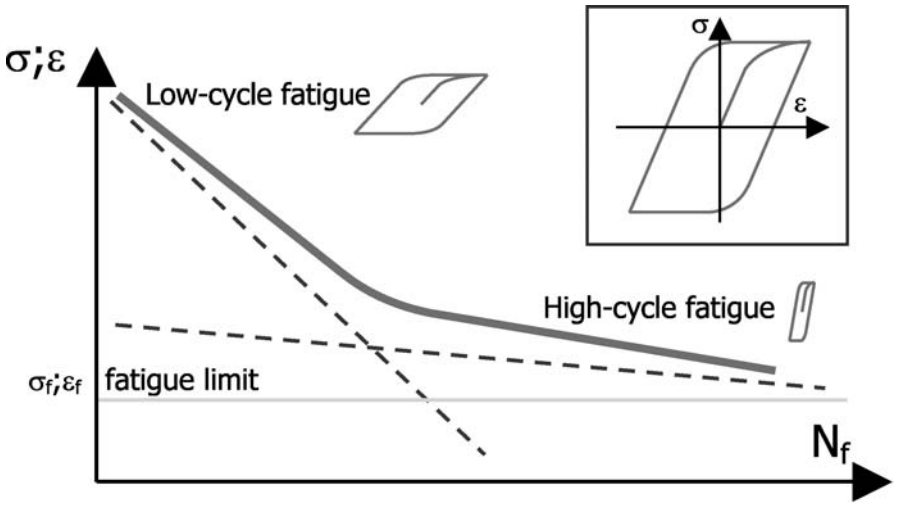


FIGURE 6.26. Cycle fatigue.

6.2.4.2.1. Steady-State Crack Growth. The crack propagation law was introduced basing on the experimental results. According to that, a crack growth rate da/dN shows a linear correlation in the region II, which could be expressed by the following formula, see Figure 6.27:

$$\frac{da}{dN} = C(\Delta K)^n, \tag{6.39}$$

where C is a crack growth rate factor, ΔK is a stress intensity factor, n is an exponent taken from experiments. The stress intensity factor ΔK can be expressed by the formula:

$$\Delta K = \alpha(\sigma_{\max} - \sigma_{\min})\sqrt{\pi a}, \tag{6.40}$$

where α is a proportional coefficient (e.g., $2/\pi$), $\sigma_{\max}$ and $\sigma_{\min}$ are the maximum and minimum stress and a is the crack length.

6.2.4.2.2. Strain Based Approach. In order to determine the thermal fatigue life in case of temperature-dependent elasto-plastic analysis an accumulated equivalent plastic strain

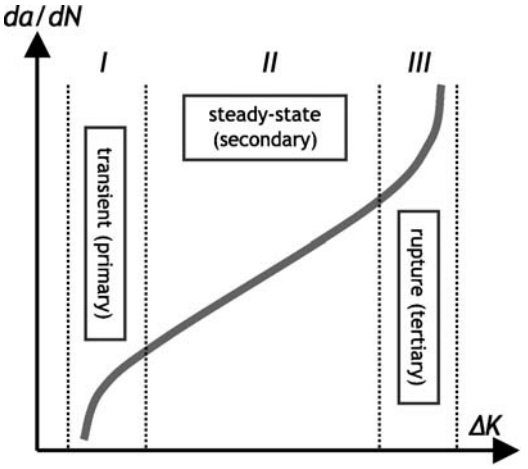


FIGURE 6.27. Dependence of crack growth rate vs. stress intensity.

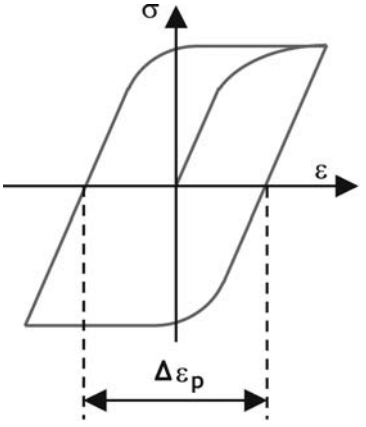


FIGURE 6.28. Equivalent plastic deformation.

range $\Delta\varepsilon_p$ is a good solution (Figure 6.28). The thermal fatigue life can be then estimated using the Coffin–Manson relation:

$$\Delta\varepsilon = \Delta\varepsilon_e + \Delta\varepsilon_p = \frac{\sigma_f}{E}(2N_f)^b + \varepsilon_f(2N_f)^c, \tag{6.41}$$

where σ_f is a fatigue limit and material b is a material constant while c and ε_f are ductility coefficient and exponent. The modified Coffin-Manson model, excluding the elastic component can be rewritten as follows:

$$N_f = \frac{1}{2}\left(\frac{\Delta\varepsilon_p}{2\varepsilon_f}\right)^{\frac{1}{c}}. \tag{6.42}$$

6.2.4.2.3. Inelastic Energy Approach. (Figure 6.29). The fatigue damage begins with the accumulation of damage at a localized region due to cycling loads, which usually leads to

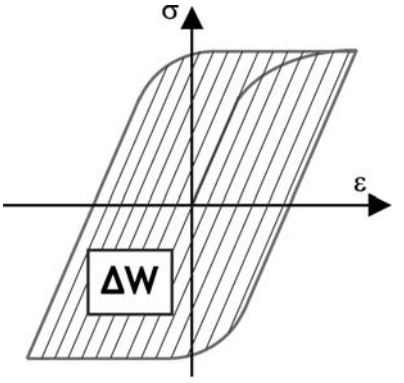


FIGURE 6.29. Inelastic energy dissipation.

cracking. The cracking process is governed by two key phenomena: crack initiation and crack propagation. The cracks usually initiate at the surface where the highly stress regions exist, e.g., notches. Once the crack is initiated it may propagate as a result of further cyclic deformation. So, the expected number of cycles to fatigue failure can be a sum of: number of cycles to crack initiation N_{fi} and number of cycles to complete fracture N_{fp}:

$$N_f = N_{fi} + N_{fp}. \tag{6.43}$$

In time-dependent material models (e.g., visco-elastic) both: the number of cycles to crack initiation N_{fi} and a number of cycles to complete fracture N_{fp} could be estimated from the following equations:

$$N_{fi} = \frac{B}{\Delta W},$$
$$N_{fp} = \frac{1}{C}\int_{a_i}^{a_f} \frac{1}{(\Delta W)^n} da, \tag{6.44}$$

where ΔW is the strain energy per cycle, a is the crack length of the solder joint, a_i is initial crack size and a_f is the crack length at failure, B, C and n are material constants. The inelastic strain energy per cycle ΔW, seems to be an appropriate damage parameter, as it correlates well with the experimental data.

6.3. MATHEMATICAL ASPECTS OF OPTIMIZATION

The first step of virtual prototyping is to create model of the process or product. Most often the procedure for model recognition would be based on controlling the input variable values and observing the output. Unfortunately, for most of the projects the above procedure seems to be very expensive and long lasting. But thanks to mathematical tools the above procedure may become efficient by reducing the required number of experiments. The most popular mathematical tools are:

- DOE (design of experiments): DOE method would be responsible for selecting an optimal set of experiments for recognizing the model of a product or process, which is selecting the most significant factors.

- RSM (response surface modeling): RSM method would be responsible for fitting surface model of the response.

The experiment is understood as planned series of tests. There are a number of problems connected with this task:

- Number of experiments, which are required for model recognition—the goal is to run as less as possible.
- Proper selection of input variables—the goal is to assess the potential power of each of them.
- Induced random distribution of input and noise variables—the goal is to assume their basic random parameters: density distribution, average, variance, etc.
- If the model is linear or non-linear—most often for simplification reasons the model is assumed to be linear, while it is highly non-linear.
- If the model requires multi-response analysis range.
- If variables are continues or step-wise.
- Influence of the noise factors—the goal is to make the final design as less sensitive as possible to noise variables, e.g., temperature.
- Multi-criteria design—the goal is to select the compromised solution in reference to a number of output variables, e.g., quality and price.
- Optimization—the goal is to find out the optimal solution in reference to required level of the response: minimal, maximal or nominal.
- Parameter and tolerance design—the goal is to select the most appropriate values of input variables to achieve optimal and at the same time the most robust design.

6.3.1. Design of Experiments

The DOE method is to answer the question how to design the experiment so it can be done in a minimum number of tests and at the same time it can deliver enough information (Figure 6.30). Unfortunately, the number of tests that are to be done in the experiment grow exponentially with a number of input variables. Therefore, the basic challenge is to design the so-called proper experiment, which could save a lot of time and money. The experiment should include only the "good" tests, which provide appropriate information on the model and to skip the tests that are "overlapped" or not required for model definition [21,32,43].

The main idea behind the design of experiments is controlling the input variables and recording the output signal of a product or process. Most often the input variables are referred to as factors while the output variable as a response. The factors could be continuous or step variables having at least two values. The factor values are called levels and can be divided into:

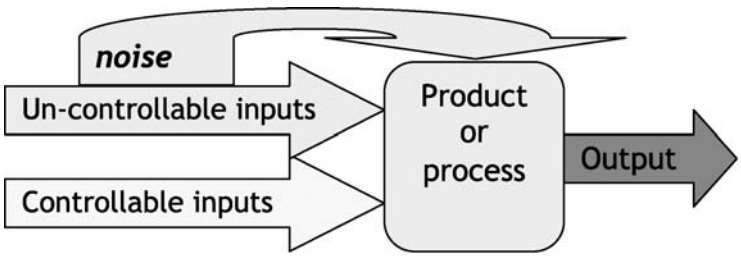


FIGURE 6.30. Selection and description of input and output factors in DOE method.

- Controllable factors, which are easy to identify and control.
- Uncontrollable factors, which are difficult to identify and control, e.g., temperature; the uncontrollable factors can be additionally divided into: internal and external; both of them are referenced as noise factors.

When searching for the adequate experiment some tests are run and the response observed. The most often approach and straightforward would be to evaluate the effect of one input parameter on the response while keeping the values of other factors at constant level. The worst case would be to evaluate the effect of a number of input parameters all at the same time, which is sometimes a common approach. It is possible to change one or more process variables (or factors) in order to observe the effect of changes on one or more response variables. The statistical design of experiments is an efficient procedure for planning experiments so that the data obtained can be analyzed to yield valid and objective conclusions [28,49] (Figure 6.31).

Well-chosen experimental designs maximize the amount of "information" that can be obtained for a given amount of experimental tests. The possible outputs of the experiment can answer the question whether selected factor influences the response mean and/or variance, see Figure 6.32.

A problem arises when response has some statistical distribution due to e.g., noise. Then the basic question refers to the problem of selecting a proper number of tests in order to get the statistical significance of the experiment and appropriate comparison of sample averages. From statistical point of view in such a case in order to have more confidence it is required to increase degrees of freedom for that factor. This can be achieved by increasing the number of tests but possibly getting more degrees of freedom for an error rather than for an analyzed factor. Additionally in most practical cases there would be always some interactions between selected input factors, which should be evaluated as well. Therefore, it seems, that the systematic approach to design of an experiment would be profitable as in cost as in time. There is a number of applications with different objectives where the experimental design can be used:

- *Screening objective*; the primary purpose of the experiment is to select or screen out the few important main effects from the many less important ones; these screening designs are also termed main effects designs.
- *Optimal fitting of a regression model objective*; if the goal is to model a response as a mathematical function of a few continuous factors and a "good" model parameter estimates are required.

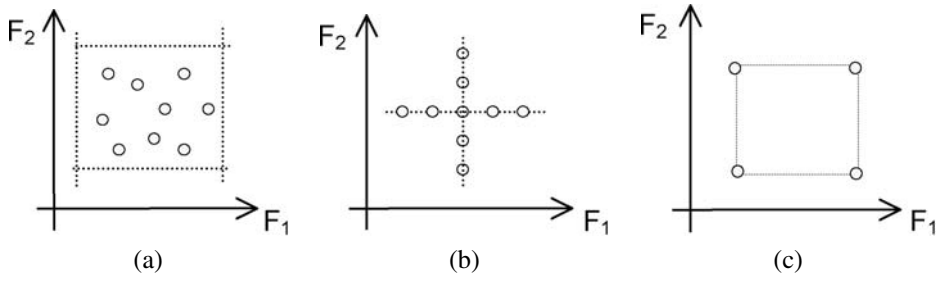


FIGURE 6.31. Different types of experiment designs: best guess (a), one-factor at time (b), statistically designed (c).

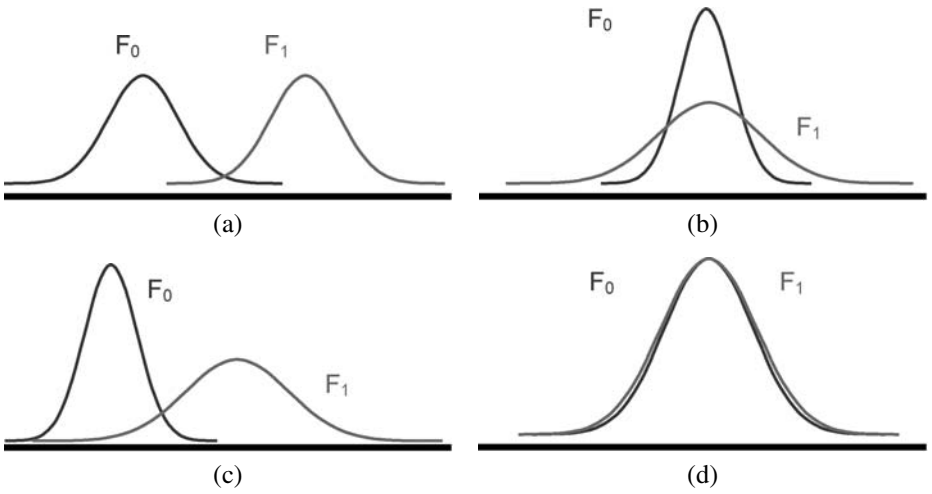


FIGURE 6.32. Possible experiment outputs: factors that shift the mean (a), factors that shift the variability (b), factors that shift both the mean and variability (c), factors that have no affect (d).

- *Robust design objective*; the goal is to perform an experiment, which would allow adjusting the essential factors such that the response would be insensitive to noise and induced variability of the factors.

The choice of an experiment design depends on: number of factors to be investigated and whether an experiment is physical or numerical. In reference to the last condition, the choice of an experiment can be made according to:

- "Corner DOE" experiments: Factorial, CCD.
- Space filling DOE: Latin hypercube, Monte Carlo.

The full-factorial design produces a uniform grid with user-specified density covering the input parameter space. There are other conventional designs, including fractional-factorial. Finally, Latin hypercube sampling (LHS) provides an orthogonal array that randomly samples the entire design space. LHS can be looked upon as a stratified Monte Carlo sampling where the pair-wise correlations can be minimized to a small value (which is essential for uncorrelated parameter estimates) or else set to a desired value. LHS is especially useful in exploring the interior of the parameter space, and for limiting the experiment to a fixed (user specified) number of runs [16,18].

There is a difference between physical and numerical experiment and it seems to be quite important. The basic characteristic of the both can be summarized as follows:

- in computer experimentation noise does not play a role, since running a computer simulation twice yields exactly the same results. Therefore, no information is gained from the repeated simulation of the same design, which is often a rule in classical DOE,
- in physical experiments it is often optimal to have design points lie on the borders of the design region. In computer experiments other parts of the design region are often equally interesting.

A basic drawback of most classical experiments is that they are only applicable for rectangular design regions. Therefore, in case of computer experiments it is almost a rule of thumb to use experiments that are:

- Space filling (test points are evenly spread out throughout the design region).
- Non-collapsing (every test gives bears information on about the influence of the other design parameters on the response).
- Sequential (minimizing the required number of experiments by selecting the minimal initial scheme and then caring out an additional experiments in order to improve the specified criterion on RSM model accuracy).
- Able to handle non-box design regions (in most cases the feasible design region is non-box as points outside this region may have no physical interpretation).

6.3.1.1. Full Factorial Design The most popular scheme of experimental design would be based on orthogonal arrays (OA). Orthogonality means that:

- Factors can be evaluated independently of one another; the effect of one factor does not bother the estimation of the effect of another factor,
- Orthogonal experiments are balanced, which means equal number of samples under various levels of analyzed factors in the selected domain (Figure 6.33).

Using the above experiment, both factor and interaction effect can be estimated. The experiment based on OA can avoid the unnecessary experiments and to keep only the necessary ones. In case of two-level full factorial experiment, the number of required experiment can be evaluated by the following formula:

$$N = 2^f, \tag{6.45}$$

where f is the number of factors (each at two levels). Unfortunately, the number of experiment grows exponentially with the number of factors. The worst situation appears in case of evaluating the factor effects at three levels of each factor. In this case, the number of required experiments can be evaluated as follows:

$$N = 3^f. \tag{6.46}$$

In fact, the full factorial experiment based on OA can be accepted only under certain conditions. The two-level experiment is performed with a few factors, e.g., 4. Otherwise the

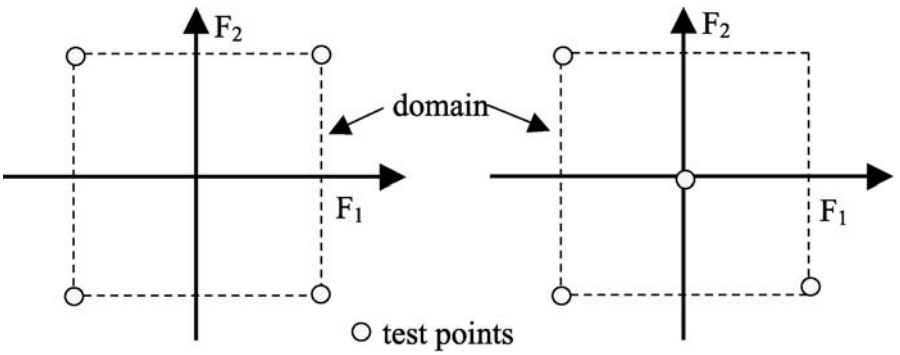


FIGURE 6.33. The orthogonal and non-orthogonal experiment.

TABLE 6.1.
A comparison of a required number of experiments.

Number of factors	Two level experiment	Three level experiment
1	2	3
2	4	9
3	8	27
4	16	81
5	32	243
6	64	729

number of test, included in the experiment may be too large, see Table 6.1. Unfortunately, in most of the engineer tasks the number of factors involved would be greater than 5 and additionally because of the high non-linear effects they should be selected at three levels for each factor.

Nevertheless, the full factorial experiment includes all the main factor effects as well as all possible interactions that are orthogonal to one another. Results of the experiment are most often interpreted according to the analysis of variance (ANOVA).

6.3.1.1.1. ANOVA Analysis. The ANOVA analysis comes from one of the fundamental theorems of statistics, the Central Limit Theorem (CLT). The CLT refers to the population despite of its distribution. The population average and its variance is the only thing that is taken into account. According to the CLT theorem if we select samples from the population than the following tenets are true [30,31]:

- Sample averages tend to be normally distributed regardless of the distribution of the individuals,
- The average of the distribution of sample averages will approach the average of the distribution of the individuals,
- The variance of the sample averages is less than the variance of the distribution of the individuals.

The CLT can be summarized in the sentence (Figure 6.34):
"if the population is sampled than the sample averages have the normal distribution:

$$\overline{E}_{yi} = N(E, \sigma) \tag{6.47}$$

with the following parameters:

$$E = E_y, \quad \sigma = \frac{\sigma_y}{\sqrt{n}}, \tag{6.48}$$

where n is a sample size."

In ANOVA analysis, the most important tenet is that the variance of sample averages σ_y^2 will be equal to the variance of the individuals σ_y^2 divided by the sample size n used to obtain the sample averages. The same is true for the variance estimates:

$$\sigma_{\frac{2}{y}} = \frac{\sigma_y^2}{n} \quad \Rightarrow \quad S_{\frac{2}{y}} = \frac{S_y^2}{n}. \tag{6.49}$$

The above formula can be used to make an estimate of the variance of sample averages by taking individuals variance and dividing it by the sample size. Therefore by comparing the

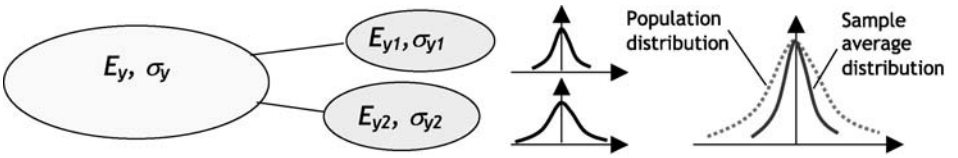


FIGURE 6.34. The concept of CLT.

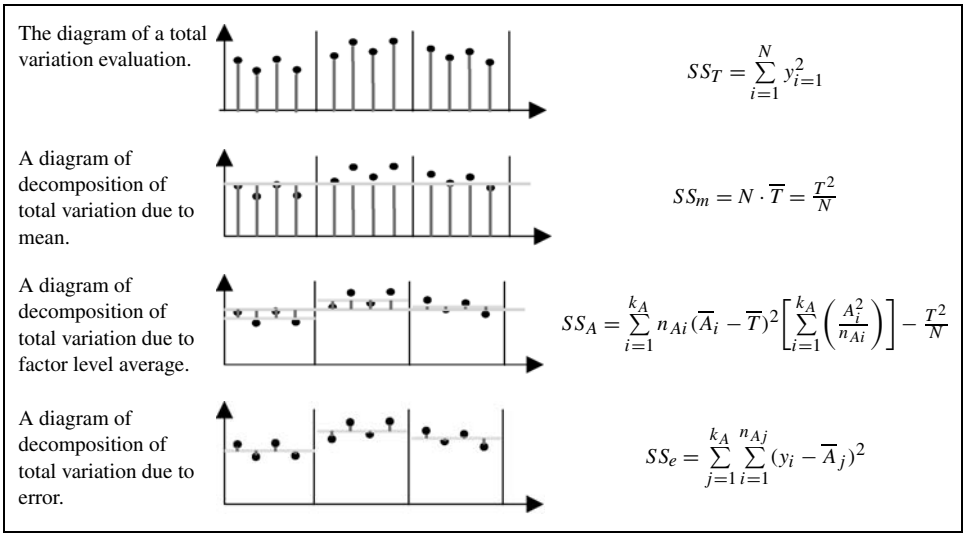


FIGURE 6.35. The concept of decomposition procedure.

variance determined from variation of sample averages and variance of an error determined from the variation of individuals can give a clue whether samples come from the same population. In practice it means that during the experiment we have to change the value of a chosen factor and then read the response and repeat the procedure for every factor depending on the selected design scheme. If the factor affects the response then through the ANOVA analysis it will be clear whether samples come from the same or different populations. The first step in the ANOVA analysis is a separation of different sources of variation in reference to individual tests through the decomposition procedure.

Generally, in case of one factor the decomposition can be presented as in Figure 6.35 and the total variation will sum up to:

$$SS_T = SS_m + SS_A + SS_e. \tag{6.50}$$

The variation due to the mean most often does not affect the overall calculations; therefore it can be excluded from ANOVA analysis:

$$SS_T = \sum_{i=1}^{N} y_{i=1}^2 - SS_m = \sum_{i=1}^{N} y_{i=1}^2 - \frac{T^2}{N}. \tag{6.51}$$

In case of more factors, additionally the variation due to factors interactions should be included, e.g., SS_{A*B}:

$$SS_T = SS_m + SS_A + SS_B + SS_{A*B} + SS_e. \tag{6.52}$$

It is easy to calculate the variance S due to the known variation SS; it is to be divided by degrees of freedom v corresponding to the chosen variation:

$$S = \frac{SS}{V}. \tag{6.53}$$

The main goal is to answer the question whether the change of a selected factor affects the response or not. That can be answered by comparing the factor variance S_A with the error variance S_e:

$$SS_A \leftrightarrow SS_e. \tag{6.54}$$

The comparison is done by analysis of variance (ANOVA). ANOVA analysis is based on F test (ratio of sample variances). This provides a decision at some confidence level whether two samples come from the same population:

$$F = \frac{S_1^2}{S_2^2}. \tag{6.55}$$

6.3.1.2. Fractional Factorial Experiment Design There is a possibility of reducing the required number of experiments but with some cost to pay, which is the lost concerning some interaction effects. Statisticians have developed more effective plans in comparison to the full factorial based on orthogonal array, which are referred to as fractional factorial experiments (FFEs). FFEs use only a portion of the total possible combinations to estimate mainly the main factor effects and some of the interactions, which seem to play a vital role (Figure 6.36). The FFEs are divided according to the part of the full factorial experiment: one-second: ½ FFE, one-fourth: ¼ FFE, etc.

Reduction of the full factorial experiment, e.g., by eighth, it is very tempting, especially from the cost and time perspective. There is a family of FFE matrices, called orthogonal arrays (OA), which can be utilized in various situations. They are referred to as: L8, L9, L16, etc. depending on number of test and number of levels for each factor (Table 6.2).

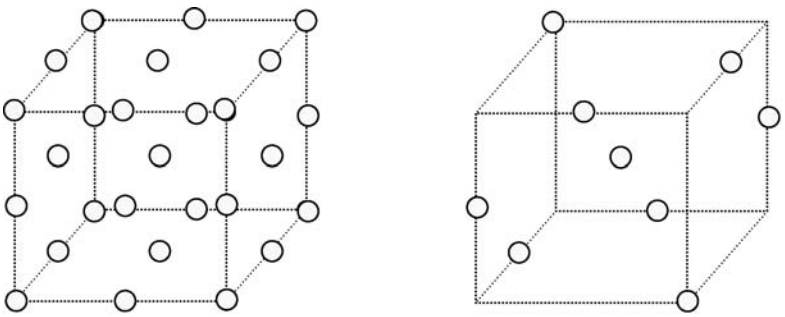


FIGURE 6.36. Example of orthogonal array for full and fractional factorial experiment.

TABLE 6.2.
The most often used orthogonal arrays.

Two-level experiment	Three-level experiment
L8	L9
L12	L18
L16	L27

Due to the fact, that orthogonal arrays based on FFEs have reduced number of tests in comparison to the full factorial experiment, there is a basic question about the resolution of the experiment and how it depends on the number of factors, levels and tests associated to the array. It is up to the expert to decide whether to perform high resolution experiment or low resolution experiment, which is mainly directed towards main factor effects, excluding the interactions. Nevertheless, the real power in using an OA is the ability to evaluate several factors in a minimum number of tests. This is considered an efficient experiment since much can be obtained from a few trials.

In fact the whole OA experiment can be divided into two steps, while the first would be focused on finding out the main factor effects, the second one to recognizing the interaction effects. The first experiment is sometimes refereed to as *screening experiment*. It should include many factors but at few levels of which two are recommended in order to minimize the size of the beginning experiment. In result of the initial experiment many factors will be eliminated from father evaluation and the remaining ones can be investigated with multiple levels without increasing the size of an experiment. The same, both interactions and non-linear effects can be studied more deeply. The above procedure of running two experiments seems very efficient and allows for time and cost reduction.

6.3.1.3. Central Composite Design (CCD) Sometimes, the preliminary experiment would be done with the assumption on linearity of the response. On the other hand, it is almost the rule of thumb that response is non-linear. In order to determine whether response in a promising region is non-linear, e.g., higher order polynomial, the Central Composite Design (CCD) scheme may be planned.

The CCD design contains an embedded factorial or fractional factorial design with center points that are augmented with a group of "star points" that allow estimation of curvature. If the distance from the center of the design space to a factorial point is ± 1 unit for each factor, the distance from the center of the design space to a star point is $\pm\alpha$ with $|\alpha| > 1$. The precise value of α depends on certain properties desired for the design and on the number of factors involved. In case of the CCD, an experiment design would include additionally test points that are within a selected domain region in spite of the orthogonal array experiment scheme (Figure 6.37).

CCD design always contains twice as many star points as factors. The star points represent new extreme values (low and high) for each factor in the design. There is variety of the central composite designs:

- Circumscribed (CCC).
- Inscribed (CCI).
- Face Centered (CCF).

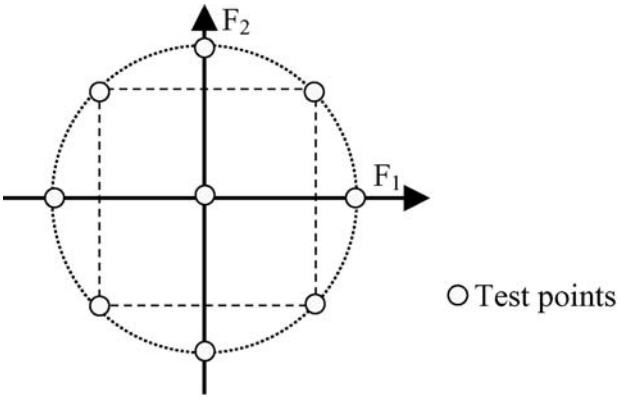


FIGURE 6.37. An idea of the Central Composite Design (CCD).

The value of parameter α depends on so-called rotatability of an experiment. To maintain rotatability, the value of α depends on the number of experimental runs in the factorial portion of the central composite design:

$$\alpha = [\textit{number or factorial runs}]^{1/4}. \tag{6.56}$$

The value of α also depends on whether or not the design is orthogonally blocked. That is whether or not the design is divided into blocks such that the block effects do not affect the estimates of the coefficients in the 2nd order model. Under some circumstances, the value of α allows simultaneously for rotatability and orthogonality of the experiment.

6.3.1.4. D-Optimal Design D-optimal designs are most often provided by a computer algorithm. These types of designs are particularly useful when classical designs do not apply. Unlike standard classical designs such as factorials and fractional factorials, D-optimal design matrices are usually not orthogonal and effect estimates are correlated. These types of designs are always an option regardless of the type of model the experimenter wishes to fit (for example, first order, first order plus some interactions, full quadratic, cubic, etc.) or the objective specified for the experiment (for example, screening, response surface, etc.).

The design is said to be D-optimal if $|\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}|/n^p$ is maximized where $\boldsymbol{X}$ is the expanded design matrix which has n rows (one for each design setting) and p columns (one column for each coefficient to be estimated plus one column for the overall mean). The D-efficiency statistic for comparing designs is as follows:

$$\textit{D-efficiency} = \left(\frac{|\boldsymbol{X}^T\boldsymbol{X}|_{\textit{design}}}{|\boldsymbol{X}^T\boldsymbol{X}|_{\textit{D-optimum}}}\right)^{1/p}. \tag{6.57}$$

Therefore, it compares a design against a D-optimal design—normalized by the size of the matrix in order to compare designs of different sizes. D-optimal designs are straight optimizations based on a chosen optimality criterion and the model that will be fit. The optimality criterion used in generating D-optimal designs is one of maximizing $|\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}|$—the determinant of the information matrix $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}$.

This optimality criterion results in minimizing the generalized variance of the parameter estimates for a pre-specified model. As a result, the "optimality" of a given D-optimal

design is model dependent. That is, the experimenter must specify a model for the design before a computer can generate the specific treatment combinations. Given the total number of treatment runs for an experiment and a specified model, the computer algorithm chooses the optimal set of design runs from a candidate set of possible design treatment runs. This candidate set of treatment runs usually consists of all possible combinations of various factor levels that one wishes to use in the experiment.

6.3.1.5. Latin Hypercube Design The Latin Hypercube (LH) design scheme of an experiment is mostly applied in case of computer experiments. Latin hypercube sampling (LHS) provides an array that randomly samples the entire design space broken down into r^n equal-probability regions (where r is the number of runs, and n is the number of input variables). LHS can be looked upon as a stratified Monte Carlo sampling where the pairwise correlations can be minimized to a small value (which is essential for uncorrelated parameter estimates) or else set to a desired value. Additionally in LHS design experiment in comparison to Monte Carlo approach, the test points in design region do not cluster. LHS is especially useful in exploring the interior of the parameter space, and for limiting the experiment to a fixed (user specified) number of runs. The LH cube experiment can be constructed as follows:

- Selecting the number of tests n that are to be simulated.
- Dividing each factor dimension into n equidistant levels.
- Sampling for each factor n random permutation of the levels.
- Combining permutation of the factors' levels into a simulation scheme.

In case of non-box region, more levels than test points are selected and then randomly generated an LH design (LHD) on the finer level grid. If the LHD is infeasible, the process is repeated while increasing the number of levels. The LHD experiment for a non-box region is referred to as constrained LHD (Figure 6.38).

In practice different LHD schemes exist mainly because there are number of possibilities to assign levels to factor dimensions. It can be done, for instance, uniformly or randomly. Much of the attention is directed towards so called maximum distance simulation scheme for which the minimal distance between test points is maximal. The minimal distance is a measure of the space region fillingness.

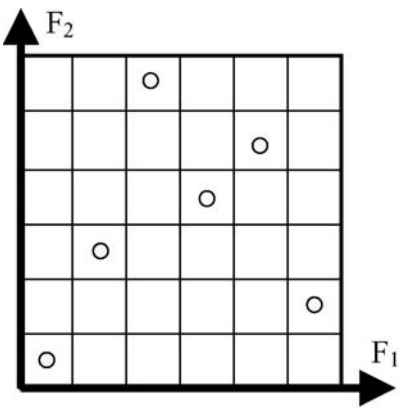


FIGURE 6.38. An idea of the Latin Hypercube experiment.

6.3.2. *Response Surface Modeling*

Thanks to the field theory, material engineering and numerical computing methods, any phenomena, process or a product can be, with some accuracy, described by a mathematical model. The model allows for predicting the behavior of a model under different conditions and can be described by a general formula, which combines the known and unknown variables influencing its behavior:

$$\boldsymbol{Y} = f(\boldsymbol{X}, \boldsymbol{Z}), \tag{6.58}$$

where matrix $\boldsymbol{Y}$ means the output variables, matrix $\boldsymbol{X}$ means the input known and controllable variables and finally matrix $\boldsymbol{Z}$ means the input unknown or uncontrollable variables, which could be referred to as noise variables.

The knowledge on model of a product or process is crucial in virtual prototyping. The model is assumed to be a black box or may require some correction according to the most current knowledge. The first step in model recognition is experiment, which in fact is a planned or unplanned series of tests. The main idea behind that is to elaborate a model, even without a complete understanding of the hidden phenomena, which would allow predicting the behavior of a process or a product within assumed input variable domain. As the response an output is selected, which is supposed to be dependent on input variables, which can be controllable or uncontrollable [22,25].

Most of the contemporary engineer tasks are directed towards proper model recognition. The traditional method is mostly based on physical experiments while the contemporary ones are more directed towards numerical experimentation. Nevertheless both methods are based on an experiment, no mater if it is physical or numerical. The method that allows for response model fitting is referenced as Response Surface Modeling (RSM) (Figure 6.39). It has been almost a rule that in many publications, the RSM is interchangeably referenced as RSA, which stands for Response Surface Analysis. The RSM method allows elaborating the mathematical model, which describes behavior of a product or process due to changes of selected essential factors. Having the model it is possible to design a contour plot of the response as a function of selected factors and selecting their most appropriate values. In the simplest cases RSM method does not require using even a computer but most often it does, especially in cases of non-linear interpolation or multi-domain problems. The benefit of RSM method is that it allows for direct application of the following advanced designing tools:

- Optimization.
- Robust design.
- Tolerance design.

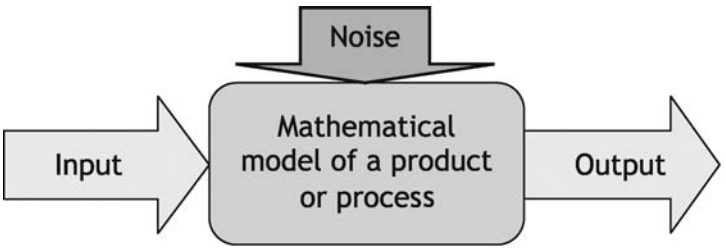


FIGURE 6.39. The idea of RSM method.

Therefore, RSM method is meant to save a lot of simulation runs in numerical prototyping or experiments in traditional prototyping. The method in most cases is refereed to as especially devoted DOE scheme for the response design. Once one knows the primary variables (factors) that affect the responses of interest, a number of additional objectives of the response design may be pursued. These include [11,19,50]:

- Hitting a target: This is a frequently encountered goal for an experiment. One might try out different settings until the desired target is "hit" consistently. Rather than experimenting in an ad hoc manner until we happen to find a setup that hits the target, one can fit a model estimated from a small experiment and use this model to determine the necessary adjustments to hit the target.
- Maximizing or minimizing a response: Many processes are being run at sub-optimal settings, even though each factor has been optimized individually over time. Finding settings that increase yield or decrease the amount of scrap and rework represent opportunities for substantial financial gain.
- Reducing variation: A product may be affected by high internal variation. Excessive variation can result from many causes: lack of having or following standards or due to certain hard-to-control inputs that affect the critical output characteristics. When this latter situation is the case, one may experiment with these hard-to-control factors, looking for a region where the surface is flatter and the process is easier to manage.
- Making a process robust: An item designed and made under controlled conditions will be later tested in the hands of the customer and may prove susceptible to failure modes not seen in the lab or thought of by design, e.g., operation under extremes of external temperature. Designing an item so that it is robust for a special experimental effort. It is possible in the lab to determine the critical components affecting its performance.
- Seeking multiple goals: A product or process seldom has just one desirable output characteristic. There are usually several, and they are often interrelated so that improving one will cause a deterioration of another. Any product is a trade-off between these various desirable final characteristics. This is done by either constructing some weighted objective function (desirability function) and optimizing it, or examining contour plots of evaluated responses.

The choice of an experiment design scheme in RSM method would depend on the behavior of a response function (model), which reflects the real life phenomena and other requirements:

- Model linearity.
- Experiment rotability.

In case of a model linearity, the possible behaviors of responses as functions of factor settings can include linear, quadratic or cubic. If a response behaves as linear, the design matrix to quantify that behavior need only contain factors with two levels and can be referenced by factorial and fractional factorial designs. If a response behaves as quadratic, the minimum number of levels required for a factor to quantify that behavior equals to three. It may be assumed that adding centre points to a two-level design would satisfy that requirement, but the arrangement of the treatments in such a matrix confounds all quadratic effects with each other. While a two-level design with centre points cannot estimate individual pure quadratic effects, it can detect them effectively. A solution to creating a design

matrix that permits the estimation of simple curvature would be to use a three-level factorial design. Finally, in more complex cases including cubic function, the design matrix must contain at least four levels of each factor to characterize the behavior of a response adequately [9,15,23].

Basic goal of RSM method is evaluation of a process or product model with an output being a good-fitting mathematical function with high predictive power, and to have good estimates of coefficients in that function with maximal accuracy. The output from process modeling is a fitted mathematical function whether based on approximation or interpolation of the data points. For a given data set the most common response would be given by a formula:

$$y = (\mathbf{X}), \tag{6.59}$$

or simply by a matrix in case of numerical evaluation. In fact, there are a number of mathematical methods, which allow for response approximation or interpolation. Additionally, in case of advanced prototyping methods as e.g., sequential procedure, the ability of approximation or interpolation error estimation would be the additional benefit and key factor for reducing number of "needed" experiments [2,37,40].

There is a distinct difference between the traditional laboratory experiments and numerical experiments, which is noise. The numerical experiments tend to give the same results while the laboratory experiments are error/noise biased. The above requires different approach to RSM model of the response:

- Laboratory experiments: approximation.
- Numerical experiments: interpolation.

The application of advanced prototyping procedures is more difficult in case of traditional experiments due to approximation method. The both approaches require different RSM models and the same different evaluation method of the model error estimation (Figure 6.40).

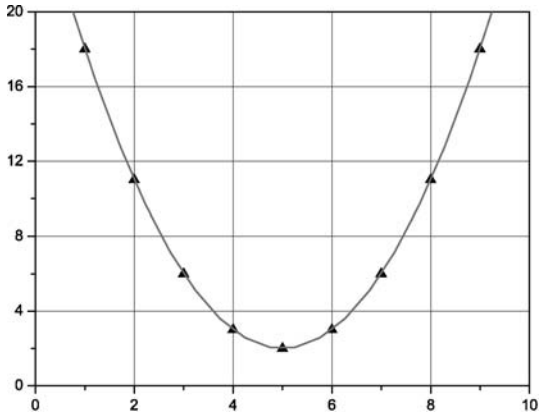
6.3.2.1. Polynomial Model For the most of response surfaces, the approximation functions are polynomials mainly because of simplicity. The most widely used are the low-order polynomials, e.g., first or second. For low curvature, a first-order polynomial can be used while for significant curvature, a second-order polynomial including all two-factor interactions.

In case of the first-order polynomial and low curvature the response surface is described as follows:

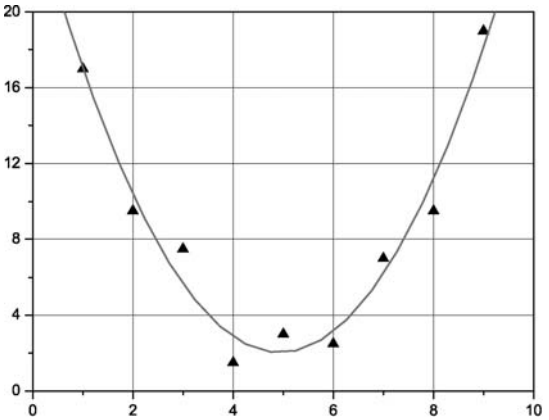
$$\hat{y} = \beta_0 + \sum_{i=1}^{k} \beta_i x_i, \tag{6.60}$$

where β's are coefficients. The above equation can be rewritten in the matrix form as:

$$\mathbf{Y} = \mathbf{BX} + \mathbf{E}, \tag{6.61}$$



(a)



(b)

FIGURE 6.40. The difference between interpolation (a) and approximation (b).

where

$$\mathbf{Y} = \begin{Bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{Bmatrix}, \quad \mathbf{B} = \begin{Bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{Bmatrix}, \quad \mathbf{X} = [x_1 \quad x_2 \quad \cdots \quad x_k], \quad \mathbf{E} = \begin{Bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{Bmatrix}. \tag{6.62}$$

The unbiased estimator $\boldsymbol{b}$ of the coefficient vector $\boldsymbol{B}$ is obtained using the least square error method:

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}. \tag{6.63}$$

The variance-covariance matrix of the $\boldsymbol{b}$ is obtained as follows:

$$\operatorname{cov}(b_i, b_j) = C_{ij} = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}, \tag{6.64}$$

where the σ is an error of $\boldsymbol{Y}$. The estimated value of σ is obtained as follows:

$$\sigma^2 = \frac{\mathbf{Y}^T\mathbf{Y} - b^T\mathbf{X}^T\mathbf{Y}}{k-1}. \tag{6.65}$$

For the cases of quadratic polynomials and significant curvature, the response surface is described as follows:

$$\hat{y} = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j. \tag{6.66}$$

The coefficients β of the above polynomials are usually determined, as described above, by least squares regression analysis and fitting the response surface approximations to existing data. Higher order polynomials are used seldom because in case of more advanced response surfaces rather different approximation functions and methods are used.

6.3.2.2. Spline Model Polynomials are the approximating functions of choice when a smooth function is to be approximated locally otherwise the degree n of the approximating polynomial may have to be chosen unacceptably large. The alternative is to subdivide the interval $[a \ldots b]$ of approximation into sufficiently small intervals $[\xi_j \ldots \xi_{j+1}]$ where:

$$a = \xi_1 < \cdots < \xi_{l+1} = b, \tag{6.67}$$

so that, on each such interval, a polynomial p_j of relatively low degree can provide a good approximation of the function f. This can even be done in such a way that the polynomial pieces blend smoothly, e.g., so that the resulting patched or composite function $s(x)$ that equals $p_j(x)$ for a chosen interval $x \in [\xi_j \ldots \xi_{j+1}]$, for all j, has several continuous derivatives. Any such smooth piecewise polynomial functions are called a spline.

There are two commonly used ways to represent a polynomial spline, the pp-form and the B-form. While a spline in pp-form is often referred to as a piecewise polynomial, then B-form is often referred to as a spline. This reflects in the fact that piecewise polynomials and (polynomial) splines are just two different views of the same thing.

6.3.2.3. Stochastical Model The stochastical modeling approach is based on considering the deterministic response $y(\boldsymbol{X})$ as a realization of a stochastic process, which means that an error B is replaced by another term $Z(\boldsymbol{X})$ representing a random process. For example, computer analysis is deterministic and not subjected to a measurement error therefore the usual uncertainty derived from least-squares residuals have no meaning, therefore the response model can be treated as a combination of a polynomial model and additional factor refereeing to the deviation from the assumed model [20,26]:

$$\hat{y}(x) = \sum_{i=1}^{k} \beta_i f_i(x) + \varepsilon(x), \tag{6.68}$$

where $\varepsilon(x)$ is the systematic deviation from the assumed model, see Figure 6.41.

In fact function $\varepsilon(x)$ representing the realization of a stochastic process is assumed to have zero mean and covariance V between two inputs u and v given by:

$$V(u, v) = \sigma^2 R(u, v) \tag{6.69}$$

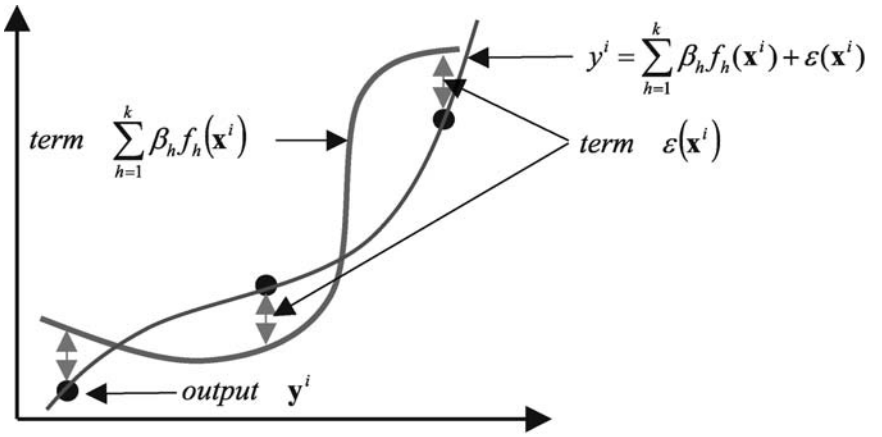


FIGURE 6.41. The concept of the stochastical model.

between $\varepsilon(u)$ and $Z(v)$, where σ^2 is the process variance and $R(u, v)$ is a correlation. The covariance structure of Z relates to the smoothness of the fitting surface. For a smooth response, a covariance function with some derivatives might be adequate, whereas an irregular response might call for a function with no derivatives. The fitting procedure can be viewed as two stage problem:

- Calculation of the generalized least-squares predictor.
- Interpolation of the residuals at the design points as if there were no regression.

One of the most popular methods for such a stochastic model interpolation is *kriging*. Kriging is extremely flexible due to the wide range of correlation functions *R(u,v)*, which may be chosen. Depending on the choice of a correlation function, kriging can either result in exact interpolation of the data points or smooth interpolation, providing an inexact interpolation. It is worth noticing that kriging is different than fitting splines and in fact it is believed even better than splines. One of the most crucial aspects of Kriging method is a problem of the interpolation error assessment. Once the interpolation error is estimated it is possible to locate a position of an additional experiment point, as required by the iterative approach procedure, which may improve the RSM model accuracy.

6.3.2.4. Radial Basis Function Model The name Radial Basis Function (RBF) comes from the function properties. Radial function is defined as function which value depends only on a distance of its argument from the center. There is a number of different radial functions:

- Multiquadratic.
- Gaussa.
- etc.

One of the most popular RBF method is based on multiquadratic functions of the form:

$$\hat{y}(x) = \sum_{j=1}^{N} C_j \varphi_j(x), \tag{6.70}$$

where matrix C_j is evaluated as:

$$A_{ij}C_j = y_i, \tag{6.71}$$

and

$$A_{ij} = \varphi_j(x_i). \tag{6.72}$$

In a standard method function $\varphi_j(x)$ is given by:

$$\varphi_j(x) = \sqrt{\|x - x_j\|^2 + h^2}. \tag{6.73}$$

In fact, this type of radial function was primarily used to approximate geographical surfaces and only recently has been adopted for optimization. In fact, approximation fitting based on RBF functions is nowadays believed to be as good as the interpolation fitting based on stochastical method.

6.3.3. Advanced Approach to Virtual Prototyping

The virtual prototyping is based on numerical whether uncorrelated (traditional) or correlated (advanced), in order to achieve usually optimal or sub-optimal designs. Numerical experiments are usually selected according to the knowledge and experience of an expert or defined according to selected experimentally/statistically orientated methodologies. The traditional virtual prototyping is based on a sequence of uncorrelated sequence of tests and procedures (Figure 6.42). The basic steps applied in a traditional virtual experiment would be based on [5]:

- Capturing the simulation sequence necessary to be able to simulate the desired design attributes, examples here are stress simulation, fatigue life simulation, thermal simulation.
- DOE (design of experiments) in order to scatter the simulations in the region of interest, according to the selected scheme, e.g., orthogonal, random and so forth.
- RSM (response surface modeling) in order to interpolate/approximate the model of the response by a mathematical model.
- Optimization, in order to find out the required response: minimum, maximum or nominal value.

Apart from the advantages well documented in the literature, there are some recognized drawbacks of the traditional virtual prototyping. The most essential ones are:

- The inability to automate the capturing and federation of the necessary simulation sequences.

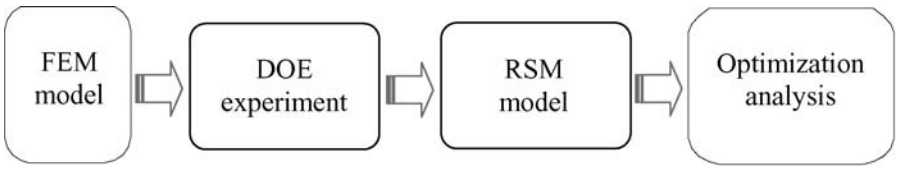


FIGURE 6.42. A simplified schematic diagram of the traditional virtual prototyping procedure.

- Efficiency, due to the required number of simulations, which grow exponentially with the number of input factors (variables).
- Quality, due to the reliability/credibility of the response model, which would be in turn used for an optimization and robust design.

The first is a limiting factor since the engineer will have to perform a lot of error prune manual work. Beyond, most often the engineer would have to decide on the compromise between the both as it seems unrealistic to e.g., improve the quality of the response model and at the same time reducing the number of tests in the chosen experiment [10,36].

In contrast to traditional methods the advanced virtual prototyping methods are based on a sequence of correlated tests and procedures. The advanced virtual prototyping methods provide the ability to capture simulation processes and further automate the running of the simulation programs. Secondly it allows for saving the total number of required experiments in order to achieve the reliable model of the response by improving the quality of the response model at a fraction of the number of experiments compared to the traditional methodologies. Figure 6.43 shows a schematic diagram of the advanced prototyping method.

First, a screening experiment is performed in order to deduct the main essential factors. Secondly, the post-screening experiment is used in order to fit the model of the response, which will be used for optimization.

The advanced scheme can be summarized by the following steps [44,45]:

- Building up numerical FEM model of a product or process capturing the physics of the problem and simulated all design critical attributes.
- Capturing and automating the design process.
- Carrying out a screening experiment based on orthogonal DOE scheme procedure in order to find out the correlation between the response and input factors (including interactions) and defining their significance in a sense of e.g., mean and variance.
- Selecting the most essential/significant input factors and adding additional experiment tests according to the elaborated modified LH design scheme.
- Interpolating/approximating the initial RSM model of the response in a form of a response surface reflecting relationship between the response and the most significant factors.
- Implementing the iterative approach in order to improve the final model of the response by sequential adding additional experiment points basing on the estimation of the interpolation error.
- Running up an optimization and sensitivity analysis to find out the best levels of the essential factors due to the expected response in the evaluated region.

Nowadays more and more companies are applying the advanced prototyping approach in order to design more reliable and high quality products. Though the current

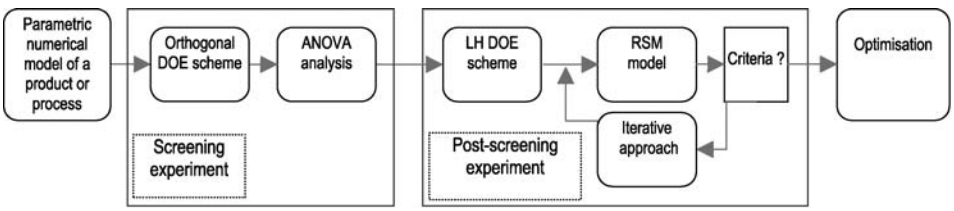


FIGURE 6.43. The advanced prototyping procedure.

procedures of DOE and RSM are broadly used worldwide, is has been a quite of an effort to adopt them for numerical prototyping in the area of electronic packages and assemblies. This is mainly due to increasing level of packaging and higher functionality of products and partly due to scale change (micro to nano). The above is due to: high non-linearity, multi-domain, multi-response, multi-factor interactions and high number of input variables. A few solutions were suggested to solve those problems: smart and sequential DOE and RSM methods, knowledge base support, etc. Nevertheless, the main goal is to reduce the number of required experiments and to improve reliability of the applied procedures.

6.3.3.1. Sequential RSM The sequential RSM (SRSM) optimization method was introduced mainly in reference to computer-aided design of experiments but in fact can be used with some modifications to traditional experiments as well [12,13,33]. The main idea of sequential method is to find a good fitting curve to the optimum of an unknown function of several variables in a minimum number of function evaluations–experiments. This can be achieved by sequential exploring of the domain of interest. The idea behind the sequential approach is to perform an iterative procedure to improve the RSM model of the response by adding additional one at a time points of experiment (Figure 6.44). The method is based on a two-step approach:

- Initial experiment based on space filing DOE scheme, e.g., LH.
- Iterative experiment based on RSM model and interpolation error estimation, e.g., Kriging model.

At the first stage an initial number of experiments is selected using e.g., Latin Hypercube method and then if required additional experiments are performed, which would improve the model accuracy. At each stage an interpolating function, derived from e.g., stochastic RSM model (Kriging) of the objective function, is set up, and this is used to determine the location of the next function evaluation and afterwards the model improvement. This process continues until agreement is reached between the optimum interpolating function value and the true value of the objective function. A balance between exploring unknown regions and optimizing the function in known regions is struck by means of a weighting factor, which varies as new data are accumulated.

The main problem is a choice of initial tests N to explore the region of interest. If N is too large, then more function evaluations will be carried out than are required for a good appreciation of the general form of the objective function. If N is too small, then the exploration of the region of interest will be insufficient and possible optimum locations may be overlooked. Obviously, the number of data points needed to "explore" the region effectively depends on a structure of the function, which is not known a priori. Data points are inserted into the region of interest one at a time and the parameters of the interpolating

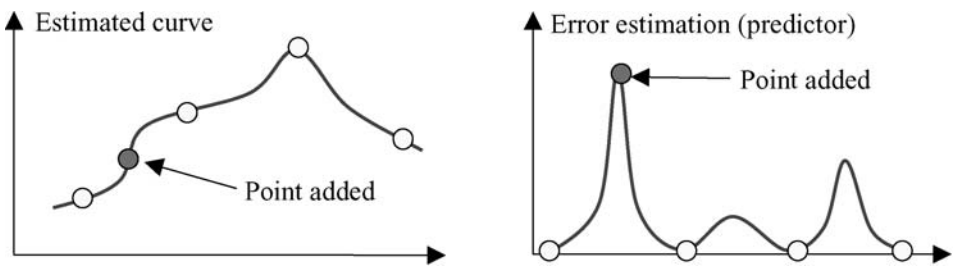


FIGURE 6.44. The advanced prototyping procedure.

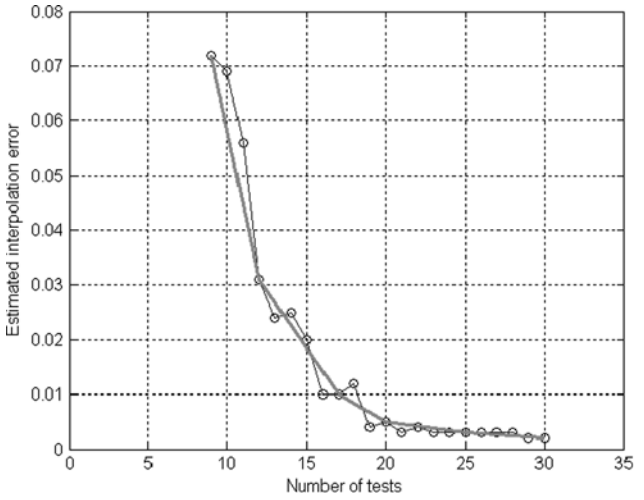


FIGURE 6.45. Example of the estimated interpolation error as a function of tests.

function are constantly updated. To define a position of a new data point, a balance is struck between the twin criteria of "exploration" and "optimization." In the early stage ("exploration") a new point is positioned as far as possible away from all existing points. In the later stages ("optimization") new points tend to concentrate around optimum values of the interpolating function. The balance between these twin criteria is defined by means of a weighting factor, which depends not only on the current number of data points but also on the apparent structure of the objective function (Figure 6.45).

The sequential approach to exploring unknown functions for optimization has the advantage of not depending on an estimate of the number of initial data points required to explore fully the region of interest. The shape of the function itself determines, to a certain extent, the point at which the "exploration" gives way to "optimization."

6.3.3.1.1. The Kriging Error Estimation. Application of the sequential approach requires one of the most crucial aspects of Kriging method, which is an assessment of the interpolation error. The idea of the sequential approach algorithm is to use the interpolation curve (predictor) together with the knowledge about the accuracy of the interpolation (standard error) in order to iteratively add points in the design space in those locations where the *expected improvement* of the objective function is the highest, as mentioned earlier, the function $\varepsilon(\mathbf{x}^i)$ representing the realization of a stochastic process is assumed to have zero mean and covariance V between two inputs u and v are given by [51–53]:

$$V(u, v) = \sigma^2 R(u, v) \tag{6.74}$$

between $\varepsilon(u)$ and $\varepsilon(v)$, where σ^2 is the process variance and $R(u, v)$ is a correlation. The covariance structure of ε relates to the smoothness of the approximating surface. For a smooth response, a covariance function with some derivatives might be adequate, whereas an irregular response might call for a function with no derivatives. The fitting procedure can be viewed as two stage problem:

- calculation of the generalized least-squares predictor,
- interpolation of the residuals at the design points as if there were no regression.

As in computer simulation, being deterministic by its nature, the error is totally due to modeling error and not to e.g., measurement error or inner or outer noise, then it is justified to treat the error $\boldsymbol{\varepsilon}^i$ as a continuous function of $\mathbf{x}^i$:

$$\boldsymbol{\varepsilon}^i = \boldsymbol{\varepsilon}^i(\mathbf{x}^i). \tag{6.75}$$

As the error is the continuous function then the errors could be considered as correlated by the distance function between the points. If points are close, then the relevant errors should also be comparable. This means high correlation. Therefore it can be assumed that the correlation between errors would be related to the distance between the corresponding points. As the distance function can be used a special weighted distance formula, which in comparison to the Euclidean distance does not weights all the variables equally:

$$d(\mathbf{x}^i, \mathbf{x}^j) = \sum_{h=1}^{k} \Theta_h |x_h^i - x_h^j|^{p_h}, \tag{6.76}$$

where $\Theta \geq 0$ and $p_h \in [1, 2]$. Using this distance function, the correlation between the errors can be defined as follows:

$$corr[\varepsilon(\mathbf{x}^i), \varepsilon(\mathbf{x}^j)] = \frac{1}{e^{d(\mathbf{x}^i, \mathbf{x}^j)}}. \tag{6.77}$$

The so defined correlation function has obvious properties, which means that in case of small distance the correlation is high while in case of large distance the correlation will approach zero. The values of the correlation function define the correlation matrix $\mathbf{R}$ of the order $n \times n$, which has practical meaning in the final response model definition:

$$r_{i,j} = corr[\varepsilon(\mathbf{x}^i), \varepsilon(\mathbf{x}^j)], \tag{6.78}$$

$$\mathbf{R} = \begin{pmatrix} r_{1,1} & \cdots & r_{1,n} \\ \vdots & & \vdots \\ r_{n,1} & \cdots & r_{n,n} \end{pmatrix}, \tag{6.79}$$

where the values of matrix $\mathbf{R}$ depend on parameters (θ_h, p_h). Thanks to the so defined correlation function and the correlation matrix $\mathbf{R}$ it is possible to get a simple linear regression model and avoid a quite complicated functional form of the response. The evaluation of the so defined stochastic model has a very important virtue, which allows replacing the regression terms by the constant value μ:

$$\mu = \sum_{h=1}^{k} \beta_h f_h(\mathbf{x}^i), \quad i = 1, \ldots, n, \tag{6.80}$$

and the same the stochastic model of the response can be rewritten as follows:

$$y(\mathbf{x}^i) = \mu + \varepsilon(\mathbf{x}^i), \quad \varepsilon(\mathbf{x}^i) \to N(0, \delta^2). \tag{6.81}$$

Therefore in order to define the stochastic model of the response it is required to estimate $2k + 2$ parameters: μ, δ^2, $\theta_1, \ldots \theta_2$, $p_1, \ldots p_k$. This task can be achieved by maximizing the likelihood function F of the sample, which is defined as follows:

$$F = \frac{1}{(2\pi)^{n/2}(\delta^2)^{n/2}|\mathbf{R}|^{1/2}} \frac{1}{e^{\frac{(\mathbf{y}-\mathbf{1}\mu)'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{1}\mu)}{2\delta^2}}}, \tag{6.82}$$

where $\mathbf{1}$ denotes the n-vector of ones and $\mathbf{y}$ denotes the n-vector of observed function values:

$$\mathbf{y} = \left(y^1, y^2, \ldots, y^n\right)'. \tag{6.83}$$

The estimators of parameters μ and δ^2 that maximize the likelihood function are given in a closed form by:

$$\hat{\mu} = \frac{\mathbf{1}'\mathbf{R}^{-1}\mathbf{y}}{\mathbf{1}'\mathbf{R}^{-1}\mathbf{1}}, \tag{6.84}$$

$$\hat{\delta}^2 = \frac{(\mathbf{y} - \mathbf{1}\hat{\mu})'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu})}{n}. \tag{6.85}$$

By substituting the above in the likelihood function ones gets the so-called "concentrated likelihood function," which depends only on parameters (θ_h, p_h):

$$L = L(\Theta_h, p_h). \tag{6.86}$$

Optimization of this function gives finally the estimates of parameters (θ_h, p_h) and hence the estimate of the correlation matrix $\mathbf{R}$. Finally it is possible to evaluate the estimates of μ and δ^2:

$$(\hat{\Theta}_h, \hat{p}_h) = \max_{(\Theta_h, p_h)} L(\Theta_h, p_h). \tag{6.87}$$

The best linear unbiased estimator of the response value y at point $\mathbf{x}^*$ is defined as:

$$y(\mathbf{x}^*) = \hat{\mu} + \mathbf{r}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu}), \tag{6.88}$$

where the $\mathbf{r}$ is the n-vector matrix given as follows:

$$r_i(\mathbf{x}^*) = corr[\varepsilon(\mathbf{x}^*), \varepsilon(\mathbf{x}^i)]. \tag{6.89}$$

It is very important to assess the estimation of the prediction accuracy at point $\mathbf{x}^*$, which can be evaluated as the mean squared error $s^2(\mathbf{x}^*)$ as follows:

$$s^2(\mathbf{x}^*) = E\left\{[\hat{y}(\mathbf{x}^*) - y(x^*)]^2\right\} = \hat{\delta}^2\left[1 - \mathbf{r}'\mathbf{R}\mathbf{r} + \frac{(1 - \mathbf{1}'\mathbf{R}^{-1}\mathbf{r}')}{\mathbf{1}'\mathbf{R}^{-1}\mathbf{1}}\right]. \tag{6.90}$$

Most often it would most convenient to work with the square root of the mean squared error $s(\mathbf{x})$ instead:

$$s(\mathbf{x}) = \sqrt{s^2(\mathbf{x})}. \tag{6.91}$$

This provides a standard error or estimated standard deviation for measuring uncertainty in the response prediction.

The parameter Θ can be treated as a measuring factor of the importance of the variable x_h, which can be related to statement that even small values of:

$$d_h^{i,j} = \left| x_h^i - x_h^j \right| \tag{6.92}$$

may lead to large differences in the function values at $\mathbf{x}^i$ and $\mathbf{x}^j$. Therefore from the statistical sense it can be stated:

- If parameter Θ_h is large then small values of $d_h^{i,j}$ would lead to high distance d value and hence low correlation *corr*.
- If parameter Θ_h is small then small values of $d_h^{i,j}$ would lead to small distance d value and hence high correlation *corr*.

Anyway, one of the most crucial factors of Kriging method is estimation of Θ parameter. It can be done according to the calculus of variations and finding the extreme of the defined parametric functional $F(\Theta, p)$ as a total error of the interpolation, which can be defined as:

$$F(\Theta, p) = \int_{\mathbf{x}'}^{\mathbf{x}''} s(\mathbf{x}, \Theta, p) d\mathbf{x}, \tag{6.93}$$

or in case of sampled data it can be evaluated numerically as:

$$F(\Theta, p) = \sum_{\mathbf{x}'}^{\mathbf{x}''} s(\mathbf{x}, \Theta, p) \Delta\mathbf{x}, \tag{6.94}$$

where $\mathbf{x}$ is a vector of the prediction points, the $s(\mathbf{x})$ is the square root of the mean squared error of the interpolation:

$$s(\mathbf{x}, \Theta, p) = \sqrt{\hat{\delta}^2 \left[1 - \mathbf{r}'\mathbf{R}\mathbf{r} + \frac{(1 - \mathbf{1}'\mathbf{R}^{-1}\mathbf{r}')}{\mathbf{1}'\mathbf{R}^{-1}\mathbf{1}} \right]}. \tag{6.95}$$

According to the previous considerations it can be noticed that both matrix $\mathbf{R}$ and vector $\mathbf{r}$ are the functions of θ and p while the estimator of δ^2 is a function of the correlation matrix $\mathbf{R}$. As the parameter p is most often selected at value 2, in fact $p \in [1..2]$, therefore the best estimator of θ can be found as the value that minimizes the functional $F(\Theta, p = 2)$:

$$\hat{\Theta} = \min_{\Theta} F(\Theta, p = 2). \tag{6.96}$$

6.3.3.2. Multiresponse Analysis In many experimental situations, a number of responses are measured for each setting of a group of design variables. The goal of the multiresponse analysis would be to find out the optimal solution due to a few responses whether a real one or most often the compromised [4,7,14].

There are some works which stress the importance of analyzing multiresponse by means of multivariate techniques that take into account interrelations among the responses [24]. For example, the models that represent the responses may have several parameters in common. It would, therefore, make sense to combine information from all

the responses to estimate these parameters. Some problems connected with multiresponse analysis are:

- Multiresponse estimation.
- Multiresponse design of experiments.
- Multiresponse optimization.

An example of multiresponse estimation may be one proposed by Box and Draper, who developed a method of estimating the parameter vector B in a general multiresponse model. The responses are assumed to be normally distributed and to have a variance-covariance matrix Σ considered constant for the various runs within the experimental region. Box and Draper used a Bayesian argument by considering a non-informative prior distribution for B and Σ. Estimates of the elements of B are obtained by maximizing the marginal posterior density of B. This method is known as the Box-Draper estimation criterion. It applies to linear as well non-linear models. There are some drawbacks of the above method and one of them is that in case of exact linear relationships among the responses, the Box-Draper estimation criterion can lead to meaningless results [41,42].

Multiresponse optimization applies to, in fact, most of the engineer or research problems. This means, that conditions that are optimal for one response may be far from optimal or even physically impractical for the other responses. RSM methods introduce some solution for such multiresponse optimization:

- Graphical, based on superimposing response contours and visual searching for a common region where the responses achieve near optimal values or to find a location of a "compromised" optimum. Unfortunately, this procedure is difficult and almost impossible to apply when the umber of responses is greater than three.
- Another approach is based on an assumption that each response function undergoes a certain transformation into a desirability function ϕ such that $0 \leq \phi \leq 1$. The choice of transformation depends on a subjective judgment concerning the importance or desirability of the corresponding response values. A measure of the overall desirability of the responses is obtained by combining the individual desirability functions through the use of geometric mean.
- Advanced approach is based on a procedure for the simultaneous optimization of responses that are represented by linear multiresponse model. A distance function is chosen that measures the overall closeness of the response functions to achieving their respective optimal values at the same set of operating conditions. Optimum operating conditions are then derived by minimizing this distance function over the experimental region.

Unfortunately, multiresponse analysis is not as developed as its single-response counterpart. It is still relatively new, and its utility has yet to be fully appreciated. This is mainly attributed to the fact that it requires advanced numerical procedures. A lot depends as well on an expert knowledge, which can be helpful in defining so-called objective function for a number of responses.

6.3.4. Designing for Quality

Design for quality is the final stage of virtual prototyping. In the simplest case, it can be only devoted to optimization as finding the optimal factor/parameter values in a sense of the expected output, whether maximum, minimum or a nominal value. In general, it should

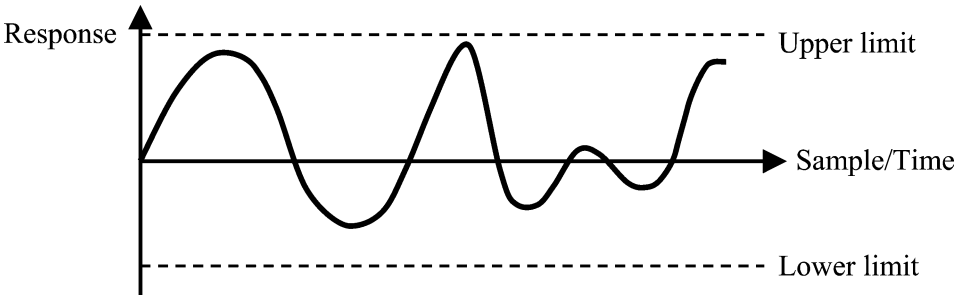


FIGURE 6.46. The scatter of the response due to time, noises, operation conditions or production phase.

include additionally such stages as robust and/or tolerance design [3,29,35]. If we have a product or a process and a selected output variable referred to as a response, then its value will differ from product to product or along timescale due to:

- Noises.
- Operation conditions.
- Production phase caused by induced random distribution of geometrical dimensions or material properties.
- Material or process parameters.

Taking into account all the above, the selected product sample could be described by random distribution with an average and variance response (Figure 6.46). As long as its average is close to the expected one and variance is low, the loss function would be small but otherwise the loss function will be high. In fact the quality of a product can be measured by such parameters as: nominal, minimal or maximal value. Additional to that, tolerances of selected factors to acceptable levels should be defined accordingly (Table 6.3). The more sample responses are within the stated limits the higher quality of a product and lower loss [17,27].

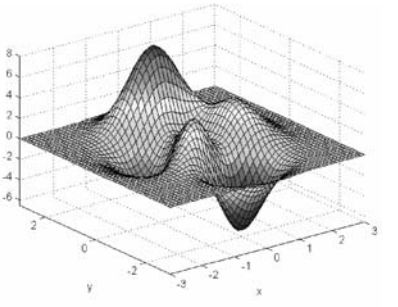
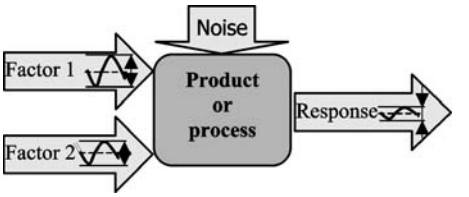
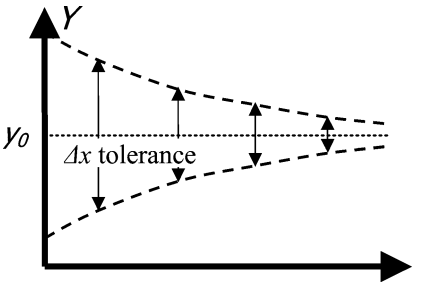
In the last few decades the quality design has been revolutionized by the innovative approach elaborated by Genichi Taguchi. At first he wrote a book devoted to experiment design and a few years later a following book on the signal to noise rate. Nevertheless his main idea was based on introducing the loss function in the experiment design. The goal was to improve the quality of the process or products so as the lost caused by the need of having them mended or improved was as low as possible. The loss does not only refer to the company profit but primarily to the society by e.g., environment pollution, noise, client complaint and so on. Therefore the higher quality then the so defined loss function value is lower. In comparison with other theoreticians, Taguchi prefers to refer to quality lost rather than the quality itself. The typical loss function L would be defined as [34]:

$$L = k\lfloor S^2 + (\overline{y} - m)^2 \rfloor, \tag{6.97}$$

where k is a constant, S is a variance of the output signal sample, m is a designed average while y is an average value of the output signal sample. Figure 6.47 shows a typical sketch of the loss function.

Though Taguchi method is very handy and does not require high knowledge on statistics, in fact it is directed towards engineers. It allows for designing such a product or process that would be satisfying to the client and at the same time reduce the costs of the

TABLE 6.3.
Design for quality characteristic.

Optimization	• The objective is to find the operating conditions or factor levels X that would optimize the system response y. • The common problem of the optimization is how to distinguish the local optimum from the global one.	
Robust design	• The aim is to make a product or process less sensitive (more robust) in the face of variation over which we have little or no control. • The robust design can be based on the Monte Carlo approach.	
Tolerance design	• The tolerance design can be performed if the robust design is not enough. • The goal is to tighten up the tolerances so as the response can be set up in the acceptable range and e.g., balanced quality vs. cost.	

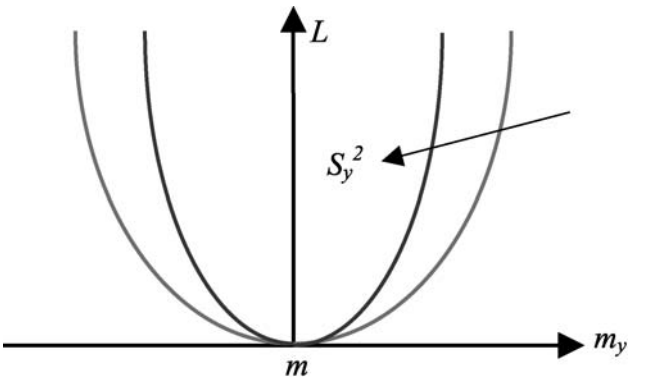


FIGURE 6.47. A sketch of a typical loss function.

company. Primarily it is based on orthogonal DOE scheme and ANOVA analysis. Though, no questionable simplicity of Taguchi approach to quality design there are some drawbacks

of which the ability of analyzing highly non-linear responses and multiple responses are the most crucial ones. In such a case, a more precise analysis would be required including advanced approach, especially in case of virtual prototyping. The method could be based on sequential DOE and RSM, which lead to more accurate product or process model in the whole domain, and can be used for optimization, robust and tolerance analysis.

As mentioned already the main reason for product or process scatter is due to controllable and uncontrollable factors. The primary task would be the recognition of controllable and uncontrollable or difficult to control factors and defining their power. In fact the lower number of uncontrollable factors the better quality of a product can be achieved, which is depicted by lower loss function value.

Most often the question would be not only how to improve the product quality but additionally how to do it with the lower cost e.g., by changing construction or adjust process parameters. This task can be performed during the designing and implementation phase. For that a product design would be required and implementation of the Quality Function Deployment QFD method. The QFD method refers to proper adjusting of the product functionality and quality so as the expectation of the clients could be met. The next phase would be designing and testing for the best input variable values and defining their quality. The goal is to achieve such a design that would be the least sensitive to outer and inter noises. If this goal is achieved than we can go to the cost reduction. At least we can be able to improve the quality and keeping the cost at the same level, which is desirable as well.

6.4. APPLICATION CASE

Numerical prototyping method can be used to predict, evaluate, optimize, and eventually qualify the thermo-mechanical behavior of electronic packages against the actual package requirements prior to major physical prototyping and manufacturing investments. Any electronic packaging is strongly non-linear, including: material non-linearity's, such as visco-plasticity, creep and/or elasto-plastic behavior, geometric non-linearity's, such as large deformation, boundary non-linearity's as edge and contacting effects. Reliable and efficient FEM-based thermo-mechanical prediction models can only be obtained if such non-linearity is taken into account.

In order to present the basics of virtual prototyping an example case is demonstrated. As could be expected there are a number of possible approaches to solve a certain problem, primarily depending on a final goal but additionally on a possibility of experimental verification, available software tools, time of evaluation, expected improvement, experience of a person and design cost.

6.4.1. Problem Description

One of the latest developments in packaging technology is the so-called exposed pad family, for instance QFN (Quad Flat Non-lead) package. An exposed pad package is a package composed of an Integrated Circuit (IC) attached to an exposed pad and in a later stage encapsulated with an epoxy moulding compound. It has been introduced into the semi-conductor market as a thin, cost effective, thermal and high frequency package solution. The exposed pad is a metal plate that is located on the bottom of the package. Exposed pads on the top of the package are less common but they exist. Many variations exist; exposed pads are found on many packages types. Mature package types with gull

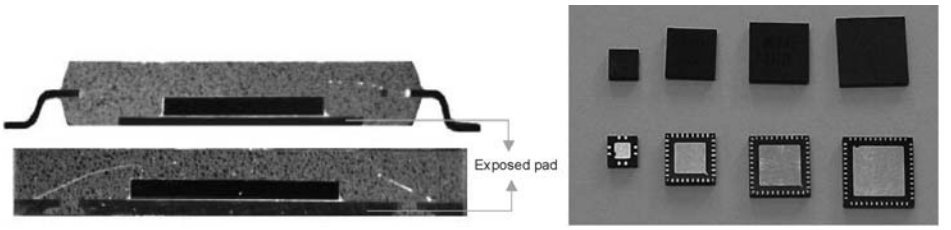


FIGURE 6.48. Examples of exposed pad packages; cross-section (left) and 3D view (right).

wing leads, such as TSSOP, offer exposed pads as an optional configuration. The exposed pad is a standard feature for QFN packages. For the leaded packages with a gull wing lead, exposed pad products are made using leadframes with a 'deep downset' paddle which is exposed to the outside of the package after the mold process. Figure 6.48 shows examples of exposed pad packages. Exposed pad packages are now on the market and exhibit a number of advantageous over other packages. Because of their size, price and performance the packages typically can be found in mobile phones and laptop computers. The presented prototyping method is applied to a typical application of a microelectronic exposed pad package. For these packages, it is vital to optimize the thickness of the leadframe toward the following restrictions [38,46]:

- Costs: minimal material (copper) will be beneficial.
- Reliability: during processing a vertical die-crack phenomenon may occur which is strongly related to the thickness of the leadframe.

6.4.2. Numerical Approach to QFN Package Design

Exposed pad packages have a very simple construction, see Figure 6.49. There is a flat leadframe, which consist of a square or rectangular diepad. The diepad in surrounded by small island of copper (which could be called leads), which are used to create to contact from the die to these islands by means of a gold wire. The die is glued on the diepad. And moulding compound in the box form shape covers the complete construction.

In order to design the parametric numerical model of the selected case, following assumptions were applied: 2D model with axi-symmetric elements were found to match well with the reality, the displacement in the x direction was fixed along the symmetry axis, the node at the left bottom corner was fixed in both the x and y directions, the single-crystal silicon die was modeled as a linear elastic material, the leadframe material was modeled as an ideally elasto-plastic material, Young's modulus and the yield stress were temperature dependent, molding compound was modeled as linear elastic material where both the Young's modulus E and the coefficient of thermal expansion were temperature dependent, the Poisson's ratio for the compound was estimated as 0.33. As the output two stresses at the top and bottom of the silicon-die surface were selected, which seemed to induce the failure of vertical die cracking after soldering and moulding process. Therefore it was necessary to include the process dependent numerical model, which is presented in Figure 6.50. The geometrical design values and the design space are listed in Table 6.4.

The phenomenon of vertical die-crack is related to the stress levels in the chip. Allowable silicon stress levels are provided by tensile and compression tests:

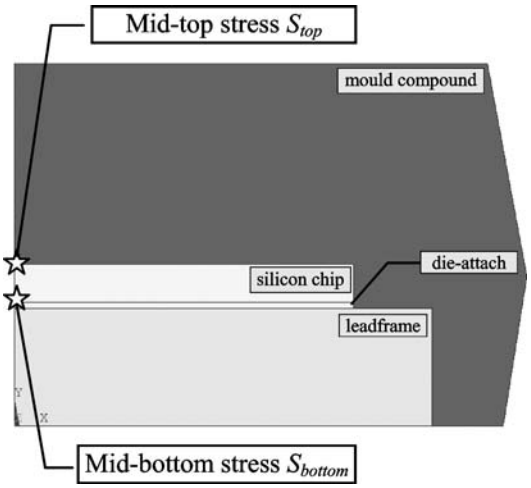


FIGURE 6.49. The parametric numerical model of a QFN product.

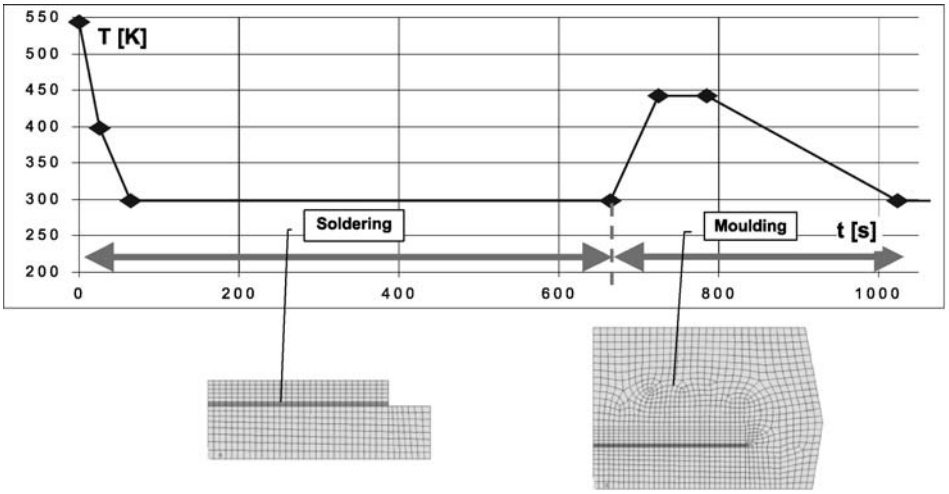


FIGURE 6.50. The major process specification, including: soldering and moulding.

TABLE 6.4.
The design materials, corresponding models and parameter values.

Material	Model	Parameters and symbols	Nominal values (mm)	Screening experiment	Post-screening experiment (mm)
Compound	linear-elastic, with:	length	6.5	None	None
	$E = f(T), a = f(T)$	thickness	2.3	None	None
Silicon die	linear-elastic	length L_{chip}	4.3	3.87–4.73	3.3–5.3
		thickness h_{chip}	0.240	0.216–0.264	None
Leadframe	elasto-plastic, with:	length $L_{leadframe}$	5.3	4.77–5.83	None
	$E = f(T), Ys = f(T)$	thickness $h_{leadframe}$	0.6	0.54–0.66	0.2–1.0
Solder	visco-plastic, with:	length	4.3	None	None
	$E = f(T), Ys = f(T)$	thickness h_{solder}	0.05	0.045–0.055	None

- In tensile direction the maximum stress is approximately +150 MPa, but strongly depends on the surface treatment.
- In compressive direction the maximum stress is approximately −500 MPa.

In contrast to traditional methods the presented approach is based on an advanced sequential approach. First, a screening experiment is performed in order to deduct the main parameters responsible for the maximum stress levels at the top and bottom side of the chip. Secondly, the post-screening experiment is used to reduce these stress levels while changing the significant geometrical parameters.

6.4.2.1. Screening Experiment Table 6.5 shows the results of the screening experiment. The whole experiment was based on a two level fractional factorial orthogonal experiment according to the L16 scheme (Figure 6.51).

According to the presented results and corresponding ANOVA (analysis of variance) analysis it was concluded that there are three dominating parameters: length of the chip L_{chip}, thickness of the leadframe $h_{leadframe}$ and thickness of the chip h_{chip}. Finally, it was decided that for the next stage only two: chip length L_{chip} and leadframe thickness

TABLE 6.5.
The screening experiment and results.

No	L_{chip}	h_{chip}	h_{solder}	$h_{leadframe}$	$L_{leadframe}$	S_{top} (MPa)	S_{bottom} (MPa)
1	1	1	1	1	1	−32.54	−3.85
2	1	1	1	2	2	−28.13	−14.66
3	1	1	2	1	2	−31.47	−5.16
4	1	1	2	2	2	−26.89	−15.34
5	1	2	1	1	2	−32.05	8.53
6	1	2	1	2	1	−27.12	−2.49
7	1	2	2	1	1	−31.57	9.02
8	1	2	2	2	2	−25.55	−4.07
9	2	1	1	1	2	−40.26	−5.82
10	2	1	1	2	1	−35.69	−17.36
11	2	1	2	1	1	−39.15	−4.15
12	2	1	2	2	2	−33.30	−19.01
13	2	2	1	1	1	−41.05	13.31
14	2	2	1	2	2	−34.42	−3.43
15	2	2	2	1	2	−39.03	11.25
16	2	2	2	2	1	−32.96	−2.49

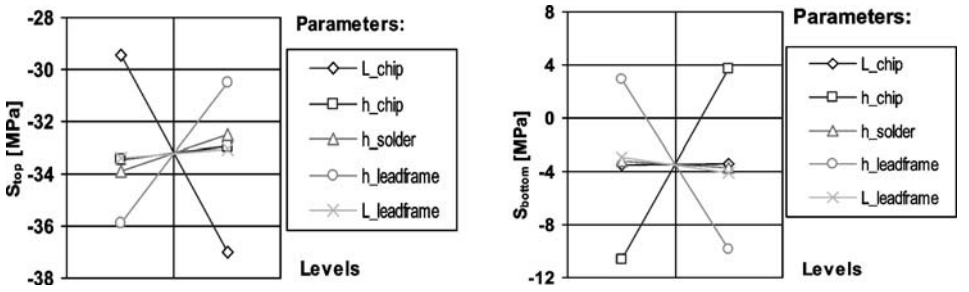


FIGURE 6.51. Main factor effects for S_{top} and S_{bottom} stress.

$h_{leadframe}$ out of three essential parameters were selected why the third parameter: chip thickness h_{chip} was set at the nominal value.

6.4.2.2. Post-Screening Experiment Table 6.6 and Figure 6.52 show the generated LH scheme and results of the sequential approach for the selected design parameters. The ex-

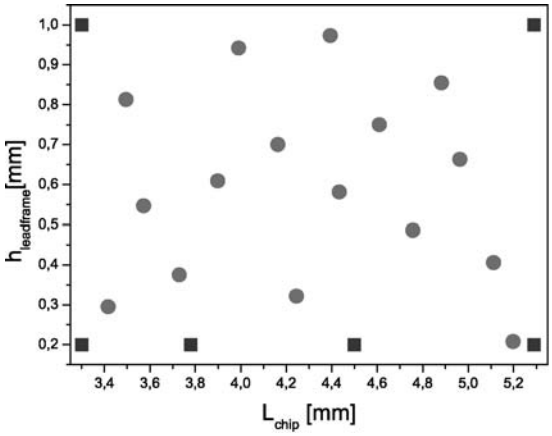
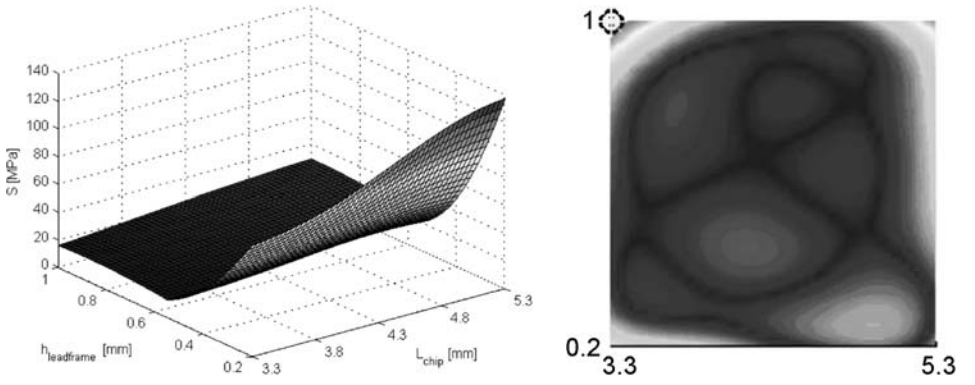


FIGURE 6.52. Selected LH scheme, where: ● initial experiments and ■ additional experiments.

TABLE 6.6.
The sequential experiment and results.

Test No	L_{chip} (mm)	$h_{leadframe}$ (mm)	S_{top} (MPa)	S_{bottom} (MPa)
Initial tests				
1	3.417	0.295	−54.03	56.74
2	3.495	0.813	−20.21	−15.11
3	3.573	0.547	−28.40	1.02
4	3.730	0.375	−45.00	33.99
5	3.898	0.609	−28.67	−4.70
6	3.991	0.942	−22.64	−20.50
7	4.163	0.700	−28.03	−11.73
8	4.245	0.322	−60.97	60.10
9	4.394	0.973	−25.34	−23.22
10	4.433	0.581	−34.58	−1.74
11	4.609	0.750	−30.33	−16.00
12	4.757	0.486	−43.98	14.17
13	4.882	0.855	−30.44	−22.22
14	4.964	0.663	−35.60	−10.76
15	5.112	0.406	−57.24	38.22
16	5.198	0.208	−112.92	149.96
Sequential tests				
17	3.30	1.00	−16.72	−18.39
18	3.30	0.20	−72.89	95.01
19	3.78	0.20	−83.76	109.95
20	5.29	1.00	−31.51	−29.54
21	5.29	0.20	−117.68	158.24
22	4.50	0.20	−99.68	132.43



(a) (b)

FIGURE 6.53. The combined stress (a) and the evaluated interpolation error (b).

periment was based on LH scheme with 16 initial tests. Basing on the interpolation error estimation additional 6 tests were done, which were to improve the overall accuracy of the final RSM model of the response.

The post-screening experiment was extended in order to include the multi-response, which seems to be important in daily engineering practice. For the presented case, in order to minimize the overall stress, two selected stresses S_{top} and S_{bottom} were combined by the following formula and then analyzed, which is presented in Figure 6.53.

$$S = 0.5(|S_{top}| + |S_{bottom}|). \tag{6.98}$$

That in turn allowed for optimization of the selected design parameters: chip length L_{chip} and leadframe thickness $h_{leadframe}$.

In order to verify the accuracy of the evaluated RSM model there was a comparison made between the predicted results and the simulation results. In order to do that, a few additional numerical experiments were done and compared with the prediction given by the RSM model. The experiment points were selected randomly within the domains of L_{chip} and $h_{leadframe}$. The verification results for the combined stress are presented in Figure 6.54. The performed verification experiment confirmed high accuracy of the RSM model, which is lower than 2% that allows for the next step of prototyping, which is optimization.

6.4.2.3. Optimization The final step of the prototyping is always devoted to optimization. In the simplest case optimization is only directed towards finding the optimal factor/parameter values in a sense of the expected output, whether maximum, minimum or a nominal value. Figure 6.55 shows the results of this optimization step.

According to the results presented in figure the expected solution for the analyzed case can be found by minimizing the objective function, which is achieved for the following factor/parameter values:

- F1: $L_{chip} = 3.3$ (mm),
- F2: $h_{leadframe} = 0.55$ (mm).

In more general case optimization should include as well sensitivity analysis due to the induced scatter of the factor/parameter values. In fact, the scatter is quite difficult to be defined a priori but it could be described by the normal density distribution. Sensitivity analysis can be achieved by Monte Carlo analysis. In case of physical experiments,

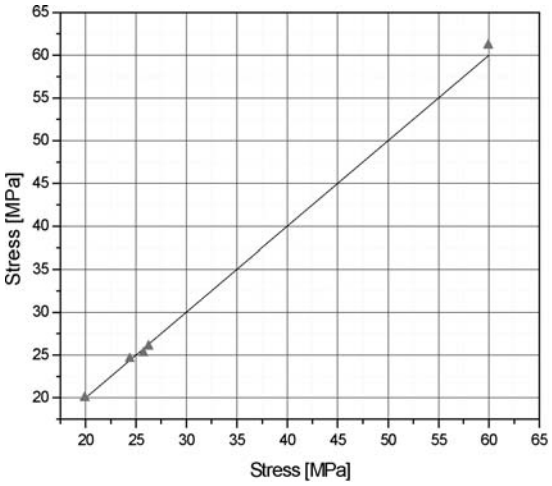


FIGURE 6.54. The results of the verification experiment.

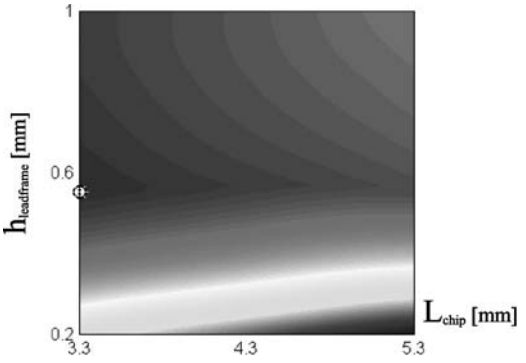


FIGURE 6.55. The optimal stress value S search for the combined stress.

Monte Carlo analysis is not feasible as it requires a large number of experiments, e.g., more than 1000. Nevertheless, in case of numerical experiments, Monte Carlo analysis can be reduced to RSM model of the defined response. Unfortunately, the whole optimization procedure requires precise RSM model, which is difficult to be precisely defined unless the whole prototyping procedure is supported by the expert knowledge.

For the current case, the Monte Carlo analysis of the defined response is given in Figure 6.56. The optimal solution can found by minimizing the both: objective and sensitivity function. The optimal solution is then located at the following parameter values:

- F1: $L_{chip} = 3.3$ (mm),
- F2: $h_{leadframe} = 0.60$ (mm).

The main difference between the optimal solution without and with sensitivity analysis is due to leadframe thickness $h_{leadframe}$ (factor F2). The optimization results with the sensitivity analysis can be interpreted as a need of changing the value of leadframe thickness $h_{leadframe}$ by 0.05 (mm) from 0.55 (mm) to 0.60 (mm). The above change is required

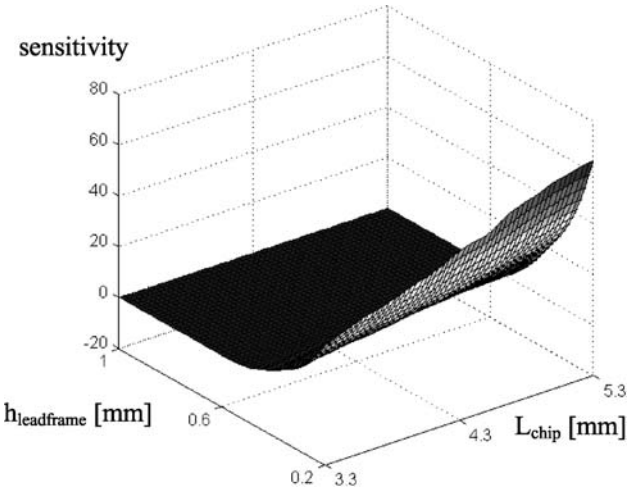


FIGURE 6.56. The sensitivity graph for the combined stress (Gaussian distribution with 1000 Monte Carlo samples).

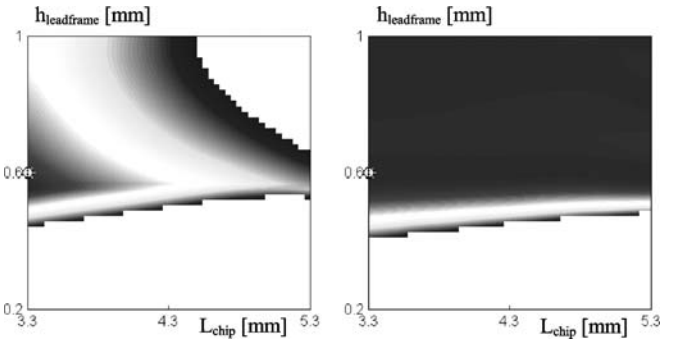


FIGURE 6.57. The optimal values search for the combined stress and sensitivity analysis with 10% tolerances.

when assuming that the input factors/parameters have the Gaussian distribution with the defined nominal value (Figure 6.57).

6.5. CONCLUSION AND CHALLENGES

This chapter highlights our major research and development results and the state-of-the-art methodology of virtual prototyping of microelectronics. Focus is on the method of virtual thermo-mechanical (thermal, mechanical and thermo-mechanical) prototyping. The results of virtual thermo-mechanical prototyping can be used to predict, qualify and optimize the thermo-mechanical behavior and/or trends of microelectronics against the actual requirements prior to major physical prototyping, manufacturing investments and reliability qualification tests.

One should also notice, that traditional experiments and tests will continue to play an important role in the content of virtual prototyping. First, they are needed in providing inputs for modeling, such as characterizing material and their interface behavior (mater-

ial properties, damage initiation, evolution and failure criteria). Secondly, very often, the correctness and accuracy of the developed simulation models and virtual prototyping results need to be verified via experiments for the whole range of the design spaces, and by covering all the critical processes.

There are mainly three important success factors for virtual thermo-mechanical prototyping. The first is to develop accurate and efficient simulation models to predict the thermo-mechanical behavior and/or trends of microelectronics. Accuracy is essential to capture the response of microelectronics correctly under manufacturing, testing and usage conditions. It can be either qualitative or quantitative, depending on the application requirements. Efficiency is needed due to the fact that it is expensive (if it is not impossible) to predict the complicated responses of microelectronics covering the whole design space and whole life cycle, and to conduct global design optimization, such as finding the maximum/minimum, robust designing, parameter sensitivity. In order to do that, one should pay attention to the following aspects:

Product/process inputs
Without reliable product/process inputs, the simulation models cannot be reliable. Two types of product/process inputs are required. One is the design parameter, another is the deviation parameter. The feasible design parameters and the associated design spaces are the starting point for modeling. The process design parameters, for example, determine the actual loading, the boundary conditions, damage initiation and evolution, partially the geometry and the material properties. The deviation parameters are fixed design parameters without given design space. However, they do have inherited scatters compared with the desired normal design values. Knowing the probabilistic distribution of these deviation parameters is important for designing for product/process robustness. Acquisition of reliable design and deviation parameters is not a trivial task, due to the facts that the real inputs, especially the deviation parameters, have strong probabilistic character and it is difficult, time and money consuming to extract the real data via in situ measurements and observations.

Tests and experiments
There are mainly two types of tests used for microelectronics, namely, the functionality test and the reliability qualification test. For functionality test, due to increased design complexity and technology and function integration, new test strategy and methods are needed to achieve maximum test coverage with minimum costs. For reliability qualification test, the essential is the correlation between the accelerated reliability tests with reliability qualification specifications, and between the real application conditions with reliability qualification specifications. The vital issue here is to achieve qualitative and quantitative matching for failure mechanisms and failure criteria.

Three types of experiments are widely used in supporting thermo-mechanical simulation, namely, experiments for material and interface characterization; for damage and failure criteria extraction; and for simulation model verification. Developing experimental methods and tools with sufficiently accurate resolution, correlating the experimental conditions with the real loading history and constraints, and designing samples representing the real product configurations, are the obvious difficulties in microelectronics.

Multiscale mechanics
The major challenge for the fundamental mechanics knowledge (both theoretic and experimental) in microelectronics is the multiscale nature of microelectronics in both geometric (from nano to millimeters) and time (from nanosecond to years) domains.

First, much effort should be spent on developing non-continuum (or enriched continuum) mechanics capable of simulating the behavior of microelectronics with micron, deep-submicron and nano dimensions, in order to capture the strong size effects inherited in microelectronics. These size effects are often related with microstructures and their evolution, various gradient effects (chemical, electrical, thermal and mechanical) and surface effect. For example, metals with a grain size of around 10 nanometers can be as much as seven times harder and tougher than their ordinary counterparts with grain sizes in the micrometer range. Tremendous size effects like this are known to play a significant role in miniaturization, and the implementation of these effects in future design processes is a necessary prerequisite to make optimal use of materials and structures at the nano scale. Presently the required knowledge in this area is substantially under-developed. The product/process behavior at nano scales cannot be predicted by simply applying the conventional macro-scale based approaches, such as continuum mechanics and thermal management, because they do not include any peculiarities of the small-scale structure of materials, but merely represent an average behavior. For this reason, they are not directly applicable for current and future product/process.

Secondly, a bridge should be developed to close the gap between non-continuum (or enriched continuum) theories, simulation tools and results of atomistic scale with the continuum theories, simulation tools and results of macro-scale. So that it will be possible to conduct multi-scale modeling, such as integrated process modeling starting from wafer processing, packaging to systems levels.

Advanced simulation tools

FEM is a well-established technique for predicting thermo-mechanical behavior of product/process. It has made significant progress especially during the last 20 years due to the rapid development of computer hardware and software. However, the commercially available FEM tools are not specifically developed for applications and needs of microelectronics. To make the FEM tools suitable for applications in microelectronics, the following issues deserve special attentions:

- Developing efficient and robust algorithms and solvers. In many applications, in order to obtain accurate results, the complicated geometric effects with high aspect ratio, nonlinear, time and temperature dependent material behavior and the complicated process history should all be considered. Despite the rapid development of computer hardware, days are still needed for the results of a single run. Therefore, virtual prototyping and qualification cannot be efficiently conducted without more efficient and robust algorithms and solvers.
- Developing efficient and reliable stochastic simulation methods. Since the design and response parameters are all probabilistic in nature, deterministic modeling and results alone cannot lead to the optimal thermo-mechanical solutions for business. Very often, robust designs are targeted, and failure probability is required. Therefore, efficient and reliable stochastic simulation methods and tools should be further developed.

Multi-physics experiments and modeling

Multi-physics modeling is another challenge of microelectronics. Microelectronics is strongly multi-discipline and multi-process. Most of the time, it is not possible to predict the behavior of microelectronics correctly by covering only one single discipline and single process. The design and qualification of microelectronics should base on integrated

understanding and solutions covering all the major involved disciplines (electric, mechanics, physics, chemistry, metallurgy, etc.) and processes (IC, packaging, assembly, testing, etc.). Taking CMP process as one example, both mechanical and chemical simulations are needed. Many important failure modes and mechanisms, such as humidity and moisture related failures, electrical overstress, stress corrosion, various cracking and fractures, MEMS striction, void formation, combined diffusion, undesired intermetallic growth, material aging, electro/thermal/stress/chemical migrations, etc. are results of strong multidisciplinary interactions. From process modeling aspect, it is well known that stress/strain induced from IC process will have impact on packaging, and packaging processes might have significant impact on the design and reliability of board level assembly as well. As the ever-increasing application of SiP, wherein different technologies (such as IC, packaging and assembly technologies), different functionalities (such as electrical, optimal, mechanical, etc.), different scales (from deep-submicron to mm) and different discipline (electric, mechanical, optimal, chemical, etc.) are strongly interact with each other, multi-physics simulation and experimental capabilities is essential part of the virtual prototyping.

Material and interface behavior

Reliable models to describe the process dependent behavior (properties, damage initiation, evolution and failure criteria) of materials and their interfaces are essential for not only predictive modeling, but also material development, pre-selection and process optimization. The ultimate aim for characterization and modeling of material and interface behavior is to develop chemical/metallurgy/physics based material design rules to tailor and manipulate material properties according to specific application needs. Several issues requires special research attention:

- For the "macro-scale" application in microelectronics, damage models and failure criteria are vital for reliable failure predictions. Knowing stress/strain distributions alone is not sufficient. The materials used for microelectronics are usually size and processes (time, stress, temperature, constraint, etc.) dependent. However, it is not easy to obtain quantitatively reliable damage models describing the size and process dependent damage initiation, evolution and failures, using the available theories and experimental techniques. Yet another difficulty is the characterization and modeling multi-damage problems, where different failure modes occur simultaneously or consequentially.
- Presently, the characterization and modeling of material behavior are usually based on the partitioning of constitutive models that describe the material properties with damage models that describe the damage and failures of materials. This practice cannot meet the need of microelectronics with nano-scale and strong multi-disciplinary interaction. Strictly speaking, material properties are always linked with damage initiation, evolution and failures. In the scale of microelectronics, the conventional constitutive models should be integrated with damage models, to form a law that describes and governs the total behavior of materials. From computational point of view, the macroscopic equations of physics of failures and the kinetic equations of the microstructural transformation (including micro-damages) should be solved simultaneously. From experimental point of view, practical identification techniques for evolution of microstructures and failure should be further developed.
- Interface strengths and interfacial failures are probably the most prevalent and pervasive issues in the electronic industry. In particular, as more organic materials being used in various types of microelectronic and Microsystems, and the ongoing trends

of miniaturization towards nano-scale, the interfaces between polymer and metal and between polymer and other adjacent materials are becoming more critical. From Integrated Circuits (ICs) to their packages and to the electronic systems made of these packages, numerous material interfaces exist. The quality, robustness, and reliability of these electronic devices depend, to a large extent, on the adhesion and durability of these interfaces. Debonding or delamination of these interfaces often results in the malfunction or failures of these electronic devices. Although interfacial adhesion had been studied for decades by numerous researchers, a few studies dealt with the adhesion issue from a multi-disciplinary viewpoint. The vast majority of studies focused on either the chemical, or the physical or the mechanical aspect alone. Because of such compartmentalized approaches, no effective methodologies, models and tools are available currently for the prediction of interfacial strength in microelectronics and Microsystems, and industry is still heavily depending on trial-error method for determining the interfacial strength. This situation is becoming even more critical due to mainly the ongoing trends of miniaturization towards nano-scale, which adds the size effect as an extra dimension to the existing scientific challenges. Thus, it is important that a generic framework for prediction of interface strengths incorporating the combined effects of and interactions among physical, chemical and mechanical bonding be developed.

The second important success factor for virtual prototyping is to develop advanced simulation based optimization method. Despite many progresses, the currently available simulation-based optimization methods are still not always reliable and rather expensive to deal with design optimization with requirements of

- Strong nonlinear responses.
- Multi-objective targets.
- Multi-level constraints.
- Large numbers of design parameters.
- Combination of continuous with discrete design parameters.

This chapter presents some development results of smart DOE algorithms in conjunction with advanced RSM methods and software. The focus is to obtain more accurate RSM with less DOE. However, many questions, such as, infill sampling criteria for multi-objective optimization, efficiency and accuracy of cross validation scheme, multiple constraints, convergence properties, Gaussian distribution assumption, etc. should be further investigated.

The third important success factor for virtual prototyping is the way to integrate the simulation models with optimization method, wherein accuracy correlation is essential. There are two types of accuracy specifications for outputs. First is the accuracy of the developed simulation models, second is the accuracy of the developed RMS. Beside that, due to the fact that all the design parameters, in principal, are statistic in nature, the accuracy of these design parameters will have important impact on the accuracy of all the outputs. Successful implementation of virtual prototyping needs reliable correlation method to interlink the accuracy specifications between the modeling results and RSM results, with predefined error criteria. Research is ongoing to make hybrid- description of the correlations of different types of errors, i.e., both mathematically and experience-based, and to make hybrid description of the error specification and control for different types of errors in both FEM models and RSM. That is to say, either from given modeling error to the resulting RSM accuracy specification, or from given RSM error to the modeling accuracy specification.

6.6. LIST OF ACRONYMS

ANOVA	Analysis of Variance
CCD	Central Composite Design
CLT	Central Limit Theorem
CTE	Coefficient of Thermal Expansion
DNP	Distance from Neutral Point
DOE	Design of Experiments
FDM	Finite Difference Method
FEM	Finite Element Method
FFE	Fractional Factorial Experiment
GIGO	Garbage In Garbage Out
HCF	High Cycle Fatigue
IC	Integrated Circuits
LCF	Low Cycle Fatigue
LEFM	Linear Elastic Fracture Mechanics
LH	Latin Hypercube
LHD	Latin Hypercube Design
LHS	Latin Hypercube Sampling
MEMS	Micro Electro-Mechanical Systems
MLH	Modified Latin Hypercube
OA	Orthogonal Arrays
PCB	Printed Circuit Board
PCP	Product Creation Process
PDE	Partial Differential Equation
QFD	Quality Function Deployment
QFN	Quad Flat Non-Lead
RBF	Radial Basis Function
RSA	Response Surface Analysis
RSM	Response Surface Model
SRSM	Sequential Response Surface Modeling
VP	Virtual Prototyping

ACKNOWLEDGMENTS

The work presented in this book chapter is part of the project results of MEVIPRO, financed by the EC in the 5th European Research Program.

REFERENCES

1. G.Q. Zhang, A. Tay, and L.J. Ernst, Virtual thermo-mechanical prototyping of electronic packaging—Bottlenecks and solutions of damaging modeling, 3rd Electronic Packaging Technology Conference (EPTC), Singapore, 2000.
2. A.R. Conn and Ph.L. Toint, An algorithm using quadratic interpolation for unconstrained derivative free optimization, in G. di Pillo and F. Giannes, Eds., Nonlinear Optimization and Applications, Plenum Publishing, 1996, pp. 27–47.
3. M.J.D. Powell, A direct search optimization method that models the objective and constraint functions by linear interpolation, Presentation at the SIAM Conference, Virginia, 1996.

4. V.V. Toropov, Multipoint approximation method in optimization problems with expensive function values, in A. Sydow, Ed., Computational System Analysis, Elsevier, 1992, pp. 207–212.
5. D. Belavic, K.P. Friedel, A. Wymyslowski, and M. Santo-Zarnik, Virtual prototyping of the ceramic pressure sensor, Proceedings of the 3nd International Conference on Benefiting from Thermal and Mechanical Simulation in (Micro)- Electronics, Paris, 2002, pp. 38–44.
6. G.Q. Zhang, J. Bisschop, and P. Maessen, Virtual thermo-mechanical prototyping of microelectronics products—towards optimised designing in reliability, Advancing of Microelectronics, 28(1) (2001).
7. V.V. Toropov, A.A. Filatov, and A.A. Polynkine, Multiparameter structural optimization using FEM and multipoint explicit approximations, Structural Optimization, 6, pp. 7–14 (1993).
8. V. Adams and A. Askenazi, Building Better Products with Finite Element Analysis, OnWord Press, 1999.
9. R.H. Myers, Response surface methodology—current status and future directions, Journal of Quality Technology, 31(1), pp. 30–44 (1999).
10. P. Stehouwer and D. Hertog, Simulation-based design optimisation: methodology and applications, Engineering Design Optimization Proceedings of 1st ASMO UK, 1999.
11. I.M. Edwards and A. Jutan, Optimization and control using response surface methods, Computers Chem. Engng, 21(4), pp. 441–453 (1997).
12. G.G. Wang and Z. Dong, Design optimization of a complex mechanical system using adaptive response surface method, Transaction of the CSME, 24(1B), pp. 295–306 (2000).
13. I.P. Schagen, Sequential exploration of unknown multi-dimensional functions as an aid to optimization, IMA Journal of Numerical Analysis, 4, pp. 337–347 (1984).
14. V.V. Toropov, A.A. Filatov, and A.A. Polynkin, Multiparameter structural optimization using FEM and multipoint explicit approximations, Structural Optimization, 6, pp. 7–14 (1993).
15. R.H. Myers, A.I. Khuri, and W.H. Carter, Response surface methodology: 1996–1988, Technometrics, 31(2), pp. 137–157 (1989).
16. T.W. Simpson, J. Peplinski, P.N. Koch, and J.K. Allen, On the use of statistics in design and the implications for deterministic computer experiments, Proceedings of DETC'97, 1997 ASME Design Engineering Technical Conferences, Sacramento, California, 1997.
17. M.C. Li and J. Chen, Determining process mean for machining while unbalanced tolerance design occurs, Journal of Industrial Technology, 17(1), pp. 2–6 (2001).
18. A.E. Raferty and D. Madigan, Bayesian model for linear regression models, Journal of the American Statistical Association, pp. 179–191 (1998).
19. M.D. Profirescu, G. Dima, B. Govoreanu, and O. Mitrea, HEMT Optimization by Advanced RSM Models, University Politehnica of Bucharest, Romania.
20. D.M. Ghiocel, Stochastic field models for approximating highly nonlinear random responses, Proceedings of the 15th ASCE Engineering Mechanics, Columbia University, New York, 2002.
21. L. Trocine and L.C. Malone, Finding important independent variables through screening designs: a comparison of methods, Proceedings of 2000 Winter Simulation Conference, 2000, pp. 749–754.
22. A. Todoroki, Teach Yourself response Surface Methodology, Tokyo Institute of Technology, Tokyo, Japan.
23. A. Wu, K.Y. Wu, R.M.M. Chen, and Y. Shen, Parallel optimal statistical design method with response surface modelling using genetic algorithms, IEE Proc. Circuit Devices Syst., 145(1), pp. 7–12 (1998).
24. T.H. Smith, B.E. Goodlin, and D.S. Boning, A statistical analysis of single and multiple response surface modeling, IEEE Transactions on Semiconductor Manufacturing, 12(4), pp. 419–430 (1999).
25. A.G. Greenwood, L.P. Rees, and F.C. Siochi, An investigation of the behavior of simulation response surfaces, European Journal of Operational Research, 110, pp. 282–313 (1998).
26. J. Sacks, S.B. Schiller, and W.J. Welch, Design for computer experiments, Technometrics, 31(1), pp. 41–47 (1989).
27. P.C. Benjamin, M. Erraguntla, and R.J. Mayer, Using simulation for robust system design, Simulation, 65(2), pp. 116–128 (1995).
28. T.J. Green and R.G. Launsby, Using DOE to reduce costs and improve the quality of microelectronic manufacturing processes, The International Journal of Microcircuits and Electronic Packaging, 18(3), pp. 290–296 (1995).
29. R. Khattree, Robust parameter design: a response surface approach, Journal of Quality Technology, 28(2), pp. 187–198 (1996).
30. P.J. Ross, Taguchi Techniques for Quality Engineering, McGraw-Hill Book Company, 1988.
31. W.W. Hines and D.C. Montgomery, Probability and Statistics in Engineering and Management Science, John Wiley & Sons.
32. M.R. Beauregard, R.J. Mikulak, and B.A. Olson, Experimenting for Breakthrough Improvement, Resource Engineering, 1989.

33. D.R. Jones, M. Schonlau, and W.J. Welch, Efficient global optimization of expensive black-box functions, Journal of Global Optimization, 13, pp. 455–492 (1998).
34. G. Taguchi, System of Experimental Design: Engineering Methods to Optimize Quality and Minimise Costs, UNIPUB/Kraus International Publications, 1987.
35. J.R. Koshel, Enhancement of the downhill simplex method of optimisation, International Optical Design Conference, 2002.
36. R.O. Bowden and J.D. Hall, Simulation optimisation research and development, Simulation Conference Proceedings, Vol. 2, 1998, pp. 693–1698.
37. M.J.D. Powell, A direct search optimisation method that models the objective and constraint functions by linear interpolation, Proceedings SIAM Conference, 1996, pp. xx-yy.
38. W.D. Van Driel, G.Q. Zhang, J.H.J. Janssen, and L.J. Ernst, Response surface modelling for non-linear packaging stresses, Journal of Electronic Packaging, 125(4), pp. 490–497 (2003).
39. G.Q. Zhang, The challenges of virtual prototyping and qualification for future microelectronics, J. Microelectronics Reliability, 43, pp. 1777–1785 (2003).
40. A.R. Conn and Ph.L. Toint, An algorithm using quadratic interpolation for unconstrained derivative free optimisation, in G. di Pillo and F. Giannes, Eds., Nonlinear Optimization and Applications, Plenum Publishing, 1996, pp. 27–47.
41. V.V. Toropov, Multipoint approximation method in optimisation problems with expensive function values, in A. Sydow, Ed., Computational System Analysis, Elsevier, 1992, pp. 207–212.
42. V.V. Toropov, A.A. Filatov, and A.A. Polynkine, Multiparameter structural optimisation using FEM and multipoint explicit approximations, Structural Optimisation, 6, pp. 7–14 (1993).
43. D.C. Montgomery, Design and Analysis of Experiments, John Wiley & Sons Inc., 2005.
44. W.D. Van Driel, J. Van de Peer, N. Tzannetakis, A. Wymyslowski, and G.Q. Zhang, Advanced numerical prototyping methods in modern engineering applications, Proceedings of the 5th International EuroSimE Conference, 2004, pp. 211–218.
45. A. Wymyslowski, W.D. Van Driel, G.Q. Zhang, J. Van de Peer, and N. Tzannetakis, Smart and sequential approach to numerical prototyping in micro-electronic application, JMEP, 2(1), pp. 1–7 (2005).
46. R. Van den Boomen and M.C. Seegers, Leadframe materials, Internal Philips Report, 1996.
47. J.D. Wu, C.Y. Huang, and C.C. Liao, Fracture strength characterization and failure analysis of silicon dies, Microelectronics Reliability, 43, pp. 269–277 (2003).
48. O.C. Zienkiewicz and R.L. Taylor, The Finite Element Method, Volumes 1–3, Butterworth-Heinemann, London, 2000.
49. D.C. Montgomery, Design and Analysis of Experiments, John Wiley & Sons Inc., 2005.
50. R.H. Myers and D.C. Montgomery, Response Surface Methodology: Process and Product Optimisation Using Designed Experiments, John Wiley & Sons Inc.
51. M. Sasena, M. Parkinson, P. Goovaerts, P. Papalambros, and M. Reed, Adaptive experimental design applied to an ergonomics testing procedure, Proceedings of DETC'02, Montreal, Canada, October 2002.
52. D.R. Jones, M. Schonlau, and W.J. Welch, Efficient Global Optimization of Expensive Black-Box Functions, Kluwer Academic Publishers, Netherlands, 1998.
53. J. Sacks, W.J. Welch, W.J. Mitchell, and H.P. Wynn, Design and analysis of computer experiments, Statistical Science, 4(4), pp. 409–435 (1989).
54. O.F. Slattery, G. Kelly, and J. Greer, Benefiting from thermal and mechanical simulation in microelectronics, in G.Q. Zhang, L.J. Ernst, and O. de Saint Leger, Eds., Thermal and Mechanical Problems in Microelectronics, Kluwer Academic Publishers, 2000, pp. 17–26.