UNIVERSITA' DEGLI STUDI DI CATANIA

DOTTORATO DI RICERCA IN ENERGETICA
XXVI CICLO


VIVIANA CHIARELLO

# ANALISYS AND SYNTESYS OF THIN-FILM SOLAR CELLS WITH METALLIC NANOPARTICLES

TESI DI DOTTORATO

Coordinatore: Prof. Luigi Marletta
Tutore: Prof. Salvatore Alfonzetti

# CONTENTS

# Preface

The interaction between electromagnetic radiation and condensed matter is a broad field of study concerning numerous practical applications in everyday life.
The metal nanoparticles have peculiar optical properties that determine a real revolution in the fields of physics, chemistry, materials science and bioscience, due to their ability to increase and to focus the electromagnetic fields in spatial regions smaller than the light wavelengths. This ability is due to the presence of localized surface plasmons (LSP), or collective undulatory excitations of free electrons in the metal particles. The manufacturing techniques bottom-up of plasmonic nanoparticles with high possibilities of control and precision are promising for the future construction of low-cost nanophotonic devices.
The plasmon science is enabling the development of a vast number of applications such as spectroscopy, high sensitivity, the realization of nanometric laser and of ultracompact optical circuities that is expected in the future can serve as an efficient bridge between the circuitry electronics and photonics. The surface plasmon intensity depends on many factors including the wavelength of the incident light and the morphology of the metal surface. The wavelength must be very close to that of the plasma of the metal. For particles of noble metals, such as silver and gold, this can fall in the visible region. This makes the interaction of these metals with the light particularly strong and leads to a highly dispersive permittivity at optical frequencies. In particular, the real part of the permittivity changes sign near the resonance frequency. For metal particles smaller than the skin depth, the plasmon interaction becomes a collective interaction that involves the entire nanoparticle. In practice, more interesting nanoparticles have diameters less than 100 nm. Among the noble metals, silver is the metal that has less absorption and produces more intense resonance effects. The number, the position and the intensity of the plasmon resonance depends on the geometric shape and size of the nanoparticle.
In this doctorate thesis we will see how to analyze numerically the surface plasmons of metal nanoparticles, and to optimally design a solar cell equipped with metal nanoparticles.
The structure of the thesis is the following. In Chapter 1 the finite element method (FEM) is briefly outlined, both in 2D and 3D geometries; in Chapter 2 the FEM analysis of electromagnetic scattering is described by means of the hybrid FEM-RBCI (Robin boundary condition iteration) and of the perfectly matched layer (PML) methods; in Chapter 3, two stochastic optimization methods are described, that is genetic algorithms (Gas) and particle swarm optimization (PSO); in Chapter 4 the photovoltaic conversion principles are recalled; in Chapter 5 the plasmon phenomena are briefly introduced; in Chapter 6 the numerical analysis of light scattering from metal nanoparticle is performed by means of FEM-RBCI method; in Chapter 7 a solar cell, equipped with metal nanoparticles, is optimized from the point of view of efficiency; finally the author's conclusions follow.

# Chapter 1

# The finite element method

## 1.1. The method

The solution of engineering problems makes use of mathematical models to represent physical situations. Within the framework of electrical and electronic engineering the behavior of these models can be described by partial differential equations, the well-known Maxwell's equations, to which are associated the corresponding initial and boundary conditions and the constitutive equations. The complexities of the analysis domains and / or of the constitutive equations make it impossible to obtain exact analytical solutions of these differential equations in most cases. So we often use numerical methods, which provide an estimate of the unknown quantities in a discrete and finite set of points of the domain. The values of these quantities in the other points are obtained by interpolating functions. The most widely used numerical methods are the finite difference method (FDM) and the finite element method (FEM).
The FEM [1-3] was born in the 60s and had a wide circulation as a result of the technological development of electronics and computers. It lends itself well to be implemented in computer programs, and can be used for a wide range of applications. These reasons have led to the success of this method compared to FDM, born before. The latter, although more simple and easy to implement, it is suitable to solve problems that have a high degree of complexity in the domain where there are no inhomogeneities and whose borders are well defined.
The FEM is based on the discretization of the domain of interest by means of a set of subdomains, said finite elements, of finite sizes and simple shapes, interconnected at points called nodes. The unknown values of the scalar field f (function of the spatial coordinates) are obtained from the nodal values $f_n$ (n=1,…,N) through appropriate interpolating functions $\alpha_n$ said form functions (or shape functions), defined within the finite elements:

$$f(x,y,z) = \sum_n f_n \alpha_n (x,y,z) \tag{1.1.1}$$

Substituting this approximate expression in the differential equations that characterize the problem posed in an appropriate integral form, we obtain a system of algebraic equations in the unknown nodal values. These equations are linear or not, depending on the nature of the constitutive laws. The resolution of the system provides the unknown values $f_n$ and then the approximate solution in the entire domain through the use of the shape functions.

## 1.2. Discretization of the domain

The first step of the numerical analysis is the finite element discretization of the domain, namely the creation of a mesh of finite elements, in which it was decided to subdivide the domain. Typical elements used for the solution of electromagnetic problems are the

simplexes, i.e. geometric entities having $N_d + 1$ vertices in a space of $N_d$ dimensions (for example, triangles in a two dimensional space and tetraheda in a three dimensional space).
The choice of the element kind depends on the number of dimensions $N_d$ of the problem and on the shape of the domain. The use of a greater number of elements in a domain increases the accuracy of the solution, but causes an increase of the computational cost in terms of occupation of memory and computing time. A convenient approach is to increase the number of elements only where it is needed, i.e., where the unknown scalar field presents major changes. This can be done based on the experience of the user or automatically by means of algorithms of automatic adaptive meshing.
If we want a better approximation within a given finite element, it is necessary to give the approximate solution a greater number of degrees of freedom, in such a way as to reduce the difference between the approximate solution and the exact one. According to the Weierstrass theorem, given a function f(x), continuous in the closed and limited range [a, b], fixed an arbitrary number $\varepsilon > 0$, there exists a polynomial p(x) such that:

$$\left|f(x) - p(x) < \varepsilon\right| \qquad \forall\, x \in [a,b] \tag{1.1.2}$$

It is clear that every continuous function can be approximated with the desired accuracy by a polynomial of sufficiently high degree. The type of simplest approximation is the linear one, but turns out to be the worst in terms of quality. In fact, the order of the polynomial used in the approximation of the real solution affects the accuracy with which we can evaluate the solutions of differential equations: the higher the grade, the better the approximation.
In Fig. 1.1 it is shown the basic principle used in the FEM method in the one-dimensional case: after having divided the domain of analysis in finite elements (in this case, intervals, typically having different amplitudes), we proceed to approximate the unknown function as in (1.1), choosing as unknowns only the values $f_n$ in the nodes of abscissas $x_n$ on the x-axis. From the solution of a system of algebraic equations we obtain the approximate nodal values of the function f(x), while the values in the internal points are evaluated based on the approximation functions used. As we can deduce from the figure, the approximation would be better if the number of elements were increased or if higher-degree polynomials were chosen to approximate the function from one node to another.

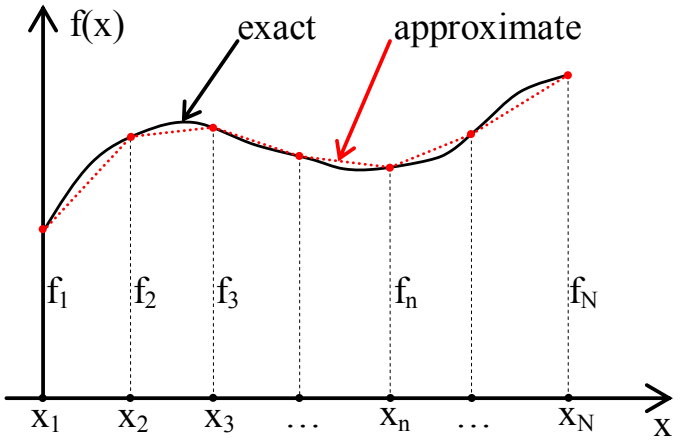



Fig. 0.1. - Exact and approximate 1D FEM solutions.

## 1.3. Scalar finite elements

The electromagnetic problems in two dimensions (2D) can be always traced to scalar problems. This is achieved by assuming as unknown the electric scalar potential; alternatively the variable can be a component (along the z axis) of the magnetic vector potential, or directly of the electric or magnetic field.
The most used finite elements in 2D are the triangle and the quadrangle. Consider a triangular finite element E, defined by three vertices, $P_n \equiv (x_n, y_n)$, n = 0,1,2, where you want to linearly approximate a scalar field f(x, y). We have:

$$f(x,y) = \sum_{n=0}^{2} f_n \alpha_n(x,y) \qquad P \equiv (x,y) \in E \tag{1.3.1}$$

where $f_n$ and $\alpha_n(x,y)$ are the nodal values and the shape functions, respectively. The shape functions $\alpha_n(x,y)$ are non-dimensional linear functions, given by:

$$\alpha_n(x,y) = \frac{x_{n+1}y_{n+2} - x_{n+2}y_{n+1}}{2A} + \frac{y_{n+1} - y_{n+2}}{2A}x + \frac{x_{n+2} - x_{n+1}}{2A}y \tag{1.3.2}$$

They exhibit the property:

$$\alpha_n(x_m, y_m) = \begin{cases} 0 & \text{if } m \neq n \\ 1 & \text{if } m = n \end{cases} \tag{1.3.3}$$

In a domain D discretized with triangular finite elements, the umerical approximation of a scalar field f(x,y) is continuous in D.
It is useful to refer to the generic point P≡(x, y) of a finite element E in a reference frame ξ,η, local to the element (Fig. 1.2 on the left), by means of the coordinate transformation:

$$\begin{cases} x = x_0 + (x_1 - x_0)\xi + (x_2 - x_0)\eta \\ y = y_0 + (y_1 - y_0)\xi + (y_2 - y_0)\eta \end{cases} \tag{1.3.4}$$

whose Jacobian is:

$$J(\xi,\eta) = \begin{vmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial x}{\partial \eta} \\ \frac{\partial y}{\partial \xi} & \frac{\partial y}{\partial \eta} \end{vmatrix} = \begin{vmatrix} x_1 - x_0 & x_2 - x_0 \\ y_1 - y_0 & y_2 - y_0 \end{vmatrix} = 2A \tag{1.3.5}$$

where A is the area of the finite element.
By means of (1.3.4), the finite element E is transformed into the standard triangle T, as shown in Fig. 1.2.
In local coordinates the shape functions simplify to:

$$\alpha_0(\xi,\eta) = 1 - \xi - \eta = \zeta \qquad \alpha_1(\xi,\eta) = \xi \qquad \alpha_2(\xi,\eta) = \eta \tag{1.3.6}$$

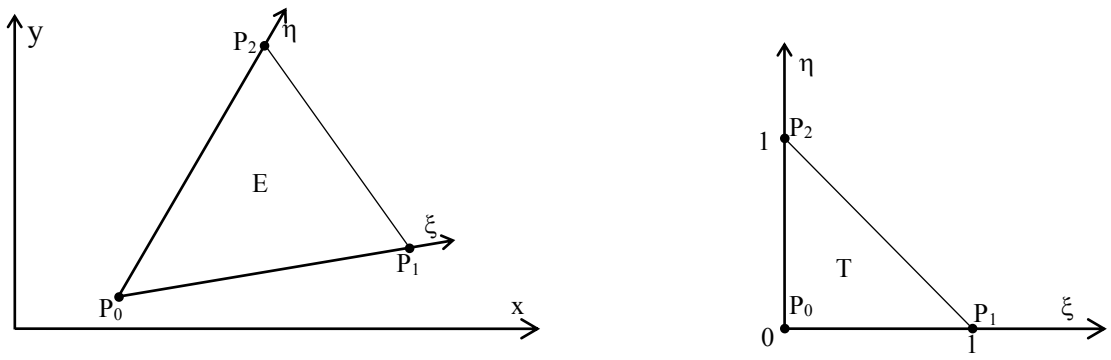



Fig. 0.2. - Triangular finite element of order 1 and standard triangle T.

The derivatives of the shape functions with respect to the absolute Cartesian coordinates are:

$$\begin{cases} \dfrac{\partial \alpha_n}{\partial x} = +\dfrac{y_2 - y_0}{2A}\dfrac{\partial \alpha_n}{\partial \xi} - \dfrac{y_1 - y_0}{2A}\dfrac{\partial \alpha_n}{\partial \eta} \\ \dfrac{\partial \alpha_n}{\partial y} = -\dfrac{x_2 - x_0}{2A}\dfrac{\partial \alpha_n}{\partial \xi} + \dfrac{x_1 - x_0}{2A}\dfrac{\partial \alpha_n}{\partial \eta} \end{cases} \tag{1.3.7}$$

By expressing the shape functions in the standard triangle T with the aid of the third local coordinate $\zeta=1-\xi-\eta$, we have:

$$\begin{cases} \dfrac{\partial \alpha_n(\xi,\eta)}{\partial \xi} = \dfrac{\partial \alpha_n(\xi,\eta,\zeta)}{\partial \xi} - \dfrac{\partial \alpha_n(\xi,\eta,\zeta)}{\partial \zeta} \\ \dfrac{\partial \alpha_n(\xi,\eta)}{\partial \eta} = \dfrac{\partial \alpha_n(\xi,\eta,\zeta)}{\partial \eta} - \dfrac{\partial \alpha_n(\xi,\eta,\zeta)}{\partial \zeta} \end{cases} \tag{1.3.8}$$

The three local coordinates $\xi$, $\eta$, $\zeta$ can be interpreted as areolar coordinates in the element E; we have in fact:

$$\xi = \frac{A_1}{A} \qquad \eta = \frac{A_2}{A} \qquad \zeta = \frac{A_0}{A} \tag{1.3.9}$$

where $A_0$, $A_1$ and $A_2$ are the areas of the triangles obtained by joining the generic point P inside the triangle with its vertices (see Fig. 1.3). It is clear that the sum of the three areolar coordinates is 1, ie: $\xi+\eta+\zeta=1$.

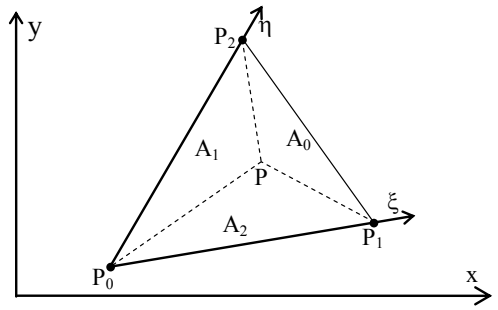



Fig. 0.3. - Areolar coordinates in a triangular finite element.

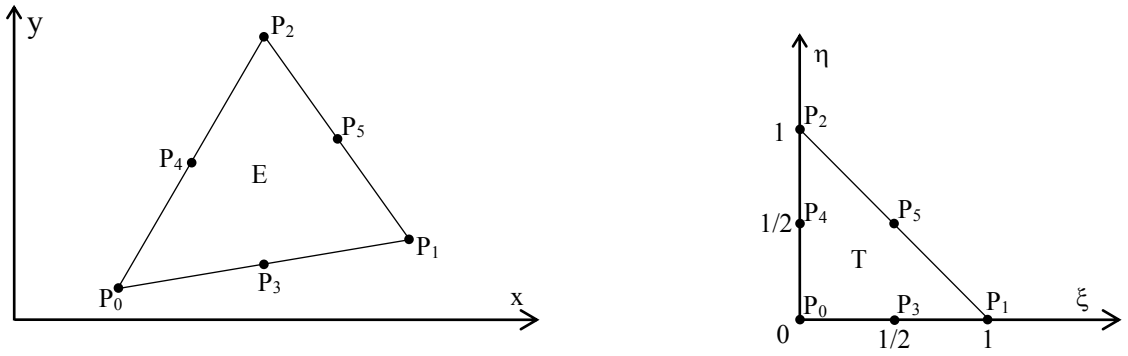



Fig. 0.4. - Triangular finite element of order 2 and standard triangle.

Triangular finite elements of higher order can be used. This choice increases the computational cost and the calculation time in favor of a better accuracy of the solution. For example, a finite element of the second order is shown in Fig. 1.4, with the associated standard triangle T. For this element the shape functions in local coordinates are:

$$\begin{aligned} &\alpha_0 = \zeta(2\zeta-1) \quad \alpha_1 = \xi(2\xi-1) \quad \alpha_2 = \eta\,(2\eta-1) \\ &\alpha_3 = 4\xi\zeta \qquad\quad \alpha_4 = 4\eta\zeta \qquad\quad \alpha_5 = 4\xi\eta \end{aligned} \tag{1.3.10}$$

and the approximation of the scalar field is:

$$f(x,y) = \sum_{n=0}^{5} f_n \alpha_n(x,y) \qquad P \equiv (x,y) \in E \tag{1.3.11}$$

In Fig. 1.5 a quadrangular finite element of order 1 is shown. By defining the local curvilinear coordinates $\xi,\eta$, the shape functions are:

$$\begin{aligned} &\alpha_1 = \tfrac{1}{4}(1-\xi)(1-\eta) \quad \alpha_2 = \tfrac{1}{4}(1+\xi)(1-\eta) \\ &\alpha_3 = \tfrac{1}{4}(1-\xi)(1+\eta) \quad \alpha_4 = \tfrac{1}{4}(1+\xi)(1+\eta) \end{aligned} \tag{1.3.12}$$

and the coordinate transformations are:

$$x = \sum_{n=1}^{4} x_n \alpha_n(\xi,\eta) \qquad y = \sum_{n=1}^{4} y_n \alpha_n(\xi,\eta) \tag{1.3.13}$$

These transformations transform the quadrangle E into the standard quadrangle Q (see Fig. 1.5). Note that the Jacobian of this transformation is not constant (exept for parallelograms), but it is a polynomial of degree 2 in the local coordinates.
The approximation of the scalar field is:

$$f(x,y) = \sum_{n=1}^{4} f_n \alpha_n(x,y) \qquad P \equiv (x,y) \in E \tag{1.3.14}$$

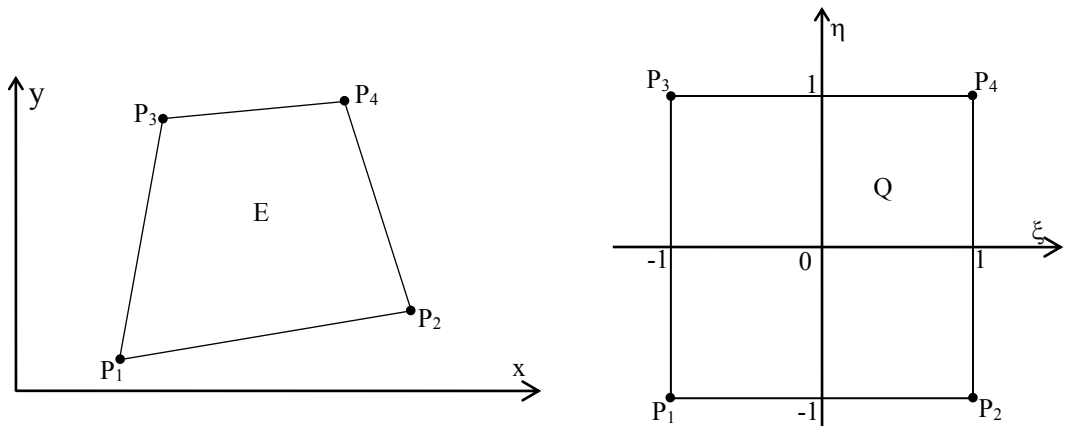



Fig. 0.5. - Quadrangular finite element of order 1 and standard quadrangle Q.

Some electromagnetic problems in three dimensions (3D) can be traced to scalar problems. This is the case of electrostatic and static current density problems formulated in terms of the scalar electric potential v.
The most used finite elements in 3D are the tetrahedron and the hexahedron. Consider a tetrahedral finite element E, defined by four vertices, $P_n \equiv (x_n, y_n)$, n = 0,1,2,3 where you want to linearly approximate a scalar field f(x, y, z). We have:

$$f(x,y,z) = \sum_{n=0}^{3} f_n \alpha_n(x,y,z) \qquad P \equiv (x,y,z) \in E \tag{1.3.15}$$

where $f_n$ and $\alpha_n(x,y,z)$ are the nodal values and the shape functions, respectively. The shape functions $\alpha_n(x,y,z)$ are non-dimensional linear functions:

$$\alpha_n(x,y,z) = a_n x + b_n y + c_n z + d_n \tag{1.3.16}$$

where le coefficients are

$$a_n = \frac{y_{n+1}z_{n+3} + y_{n+2}z_{n+1} + y_{n+3}z_{n+2} - y_{n+1}z_{n+2} - y_{n+2}z_{n+3} - y_{n+3}z_{n+1}}{6V} \tag{1.3.17}$$

$$b_n = \frac{x_{n+1}z_{n+2} + x_{n+2}z_{n+3} + x_{n+3}z_{n+1} - x_{n+1}z_{n+3} - x_{n+2}z_{n+1} - x_{n+3}z_{n+2}}{6V} \tag{1.3.18}$$

$$c_n = \frac{x_{n+1}y_{n+3} + x_{n+2}y_{n+1} + x_{n+3}y_{n+2} - x_{n+1}y_{n+2} - x_{n+2}y_{n+3} - x_{n+3}y_{n+1}}{6V} \tag{1.3.19}$$

$$d_n = + \frac{x_{n+1}y_{n+3}z_{n+2} + x_{n+2}y_{n+1}z_{n+3} + x_{n+3}y_{n+2}z_{n+1}}{6V} + \\ - \frac{x_{n+1}y_{n+2}z_{n+3} + x_{n+2}y_{n+3}z_{n+1} + x_{n+3}y_{n+1}z_{n+2}}{6V} \tag{1.3.120}$$

where V is the volume of the tetrahedron.
They exhibit the property:

$$\alpha_n(x_m, y_m, z_m) = \begin{cases} 0 & \text{if } m \neq n \\ 1 & \text{if } m = n \end{cases} \tag{1.3.21}$$

In a domain D discretized with tetrahedral finite elements, the numerical approximation of a scalar field f(x,y,z) is continuous in D.
It is useful to refer to the generic point P≡(x,y,z) of a finite element E in a reference frame $\xi,\eta,\zeta$, local to the element (Fig. 1.6 on the left), by means of the coordinate transformation:

$$\begin{cases} x = x_0 + (x_1 - x_0)\xi + (x_2 - x_0)\eta + (x_3 - x_0)\,\varsigma \\ y = y_0 + (y_1 - y_0)\xi + (y_2 - y_0)\eta + (y_3 - y_0)\,\varsigma \\ z = z_0 + (z_1 - z_0)\xi + (z_2 - z_0)\eta + (z_3 - z_0)\,\varsigma \end{cases} \tag{1.3.22}$$

whose Jacobian is:

$$J(\xi,\eta) = \begin{vmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial x}{\partial \eta} & \frac{\partial x}{\partial \zeta} \\ \frac{\partial y}{\partial \xi} & \frac{\partial y}{\partial \eta} & \frac{\partial y}{\partial \zeta} \\ \frac{\partial z}{\partial \xi} & \frac{\partial z}{\partial \eta} & \frac{\partial z}{\partial \zeta} \end{vmatrix} = \begin{vmatrix} x_1 - x_0 & x_2 - x_0 & x_3 - x_0 \\ y_1 - y_0 & y_2 - y_0 & y_3 - y_0 \\ z_1 - z_0 & z_2 - z_0 & y_3 - y_0 \end{vmatrix} = 6V \tag{1.3.23}$$

where V is the volume of the tetrahedral finite element.
By means of (1.3.22), the tetrahedral finite element E is transformed into the standard tetrahedron T, as shown in Fig. 1.6. In local coordinates the shape functions simplify to:

$$\begin{aligned} &\alpha_0(\xi,\eta,\zeta) = 1 - \xi - \eta = \psi \qquad && \alpha_1(\xi,\eta,\zeta) = \xi \\ &\alpha_2(\xi,\eta,\zeta) = \eta && \alpha_3(\xi,\eta,\zeta) = \zeta \end{aligned} \tag{1.3.24}$$

and the shape function derivatives with respect to the absolute Cartesian coordinates are:

$$\begin{aligned} \frac{\partial \alpha_n}{\partial x} = &+ \frac{(y_2 - y_0)(z_3 - z_0) - (y_3 - y_0)(z_2 - z_0)}{6V} \frac{\partial \alpha_n}{\partial \xi} + \\ &+ \frac{(y_3 - y_0)(z_1 - z_0) - (y_1 - y_0)(z_3 - z_0)}{6V} \frac{\partial \alpha_n}{\partial \eta} + \\ &+ \frac{(y_1 - y_0)(z_2 - z_0) - (y_2 - y_0)(z_1 - z_0)}{6V} \frac{\partial \alpha_n}{\partial \zeta} \end{aligned} \tag{1.3.25}$$

$$\begin{aligned} \frac{\partial \alpha_n}{\partial y} = &+ \frac{(x_3 - x_0)(z_2 - z_0) - (x_2 - x_0)(z_3 - z_0)}{6V} \frac{\partial \alpha_n}{\partial \xi} + \\ &+ \frac{(x_1 - x_0)(z_3 - z_0) - (x_3 - x_0)(z_1 - z_0)}{6V} \frac{\partial \alpha_n}{\partial \eta} + \\ &+ \frac{(x_1 - x_0)(z_2 - z_0) - (x_2 - x_0)(z_1 - z_0)}{6V} \frac{\partial \alpha_n}{\partial \zeta} \end{aligned} \tag{1.3.26}$$

$$\frac{\partial \alpha_n}{\partial z} = + \frac{(x_2 - x_0)(y_3 - y_0) - (x_3 - x_0)(y_2 - y_0)}{6V} \frac{\partial \alpha_n}{\partial \xi} + \\ + \frac{(x_3 - x_0)(y_1 - y_0) - (x_1 - x_0)(y_3 - y_0)}{6V} \frac{\partial \alpha_n}{\partial \eta} + \\ + \frac{(x_1 - x_0)(y_2 - y_0) - (x_2 - x_0)(y_1 - y_0)}{6V} \frac{\partial \alpha_n}{\partial \zeta} \tag{1.3.27}$$

By expressing the shape functions in the standard triangle T with the aid of the fourth local coordinate $\psi = 1 - \xi - \eta - \zeta$, we have:

$$\begin{cases} \dfrac{\partial \alpha_n(\xi,\eta,\zeta)}{\partial \xi} = \dfrac{\partial \alpha_n(\xi,\eta,\zeta,\psi)}{\partial \xi} - \dfrac{\partial \alpha_n(\xi,\eta,\zeta,\psi)}{\partial \psi} \\ \dfrac{\partial \alpha_n(\xi,\eta,\zeta)}{\partial \eta} = \dfrac{\partial \alpha_n(\xi,\eta,\zeta,\psi)}{\partial \eta} - \dfrac{\partial \alpha_n(\xi,\eta,\zeta,\psi)}{\partial \psi} \\ \dfrac{\partial \alpha_n(\xi,\eta,\zeta)}{\partial \zeta} = \dfrac{\partial \alpha_n(\xi,\eta,\zeta,\psi)}{\partial \zeta} - \dfrac{\partial \alpha_n(\xi,\eta,\zeta,\psi)}{\partial \psi} \end{cases} \tag{1.3.21}$$

The four local coordinates $\xi$, $\eta$, $\zeta$, $\psi$ can be interpreted as volume coordinates in the element E; we have in fact:

$$\xi = \frac{V_1}{V} \qquad \eta = \frac{V_2}{V} \qquad \zeta = \frac{V_3}{V} \qquad \psi = \frac{V_0}{V} \tag{1.3.22}$$

where $V_0$, $V_1$, $V_2$ and $V_3$ are the volumes of the four tetrahedra obtained by joining the generic point P inside the tetrahedron with its four vertices. It is clear that the sum of the four volume coordinates is 1, ie: $\xi+\eta+\zeta+\psi=1$.

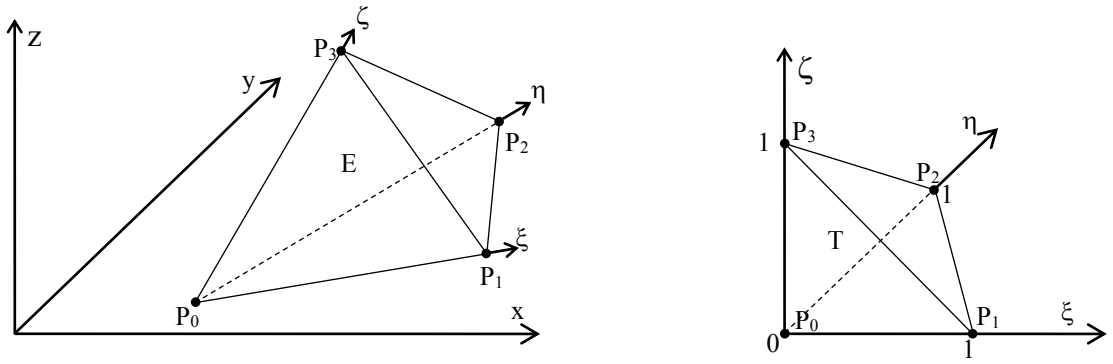



Fig. 0.6. - Tetrahedral finite element of order 1 and standard tetrahedron T.

## 1.4. Vector finite elements

The electromagnetic problems in 3D are very often vector problems, in which the unknowns are vector fields instead of scalar ones. The representation of such vector fields by means of three scalar fields exhibits several drawbacks:

- It imposes the continuity of the three components of the vector at the interface between two adjacent finite elements; this is not true if the two elements are constituted of different materials with different constitutive parameters;
- The imposition of the boundary conditions are difficult;
- The numerical solutions exhibit errors (spurious modes), whose amplitude is not controllable.

For all such reasons, vector finite elements have been devised, named edge elements. In the edge elements the shape functions are of vector kind. The most used edge elements are the triangle and the quadrangle in 2D, and the tetrahedron and hexahedron in 3D. Other elements (more rarely) used are the prism with triangular base and the pyramid with quadrangular base. It is possible to show that a generic finite volume can be subdivided in tetrahedra, but this is not true for hexahedra, prisms and pyramids.

By indicating the four local coordinates $\psi$, $\xi$, $\eta$, $\zeta$ in a tetrahedron by $\zeta_i$, i=0,…,3, the generic edge shape function is:

$$\vec{\alpha}_{ij} = L_{ij}(\zeta_i \nabla\zeta_j - \zeta_j \nabla\zeta_i) \tag{1.4.1}$$

were indices i=1,…3 and j=i+1,…,4 are relative to the beginning and ending nodes of the edge, oriented from node i to node j, as shown in Fig. 1.7. Said $\hat{e}_{ij}$ the versor of the edge ij, it can be shown that:

$$\hat{e}_{ij} \cdot \vec{\alpha}_{kh} = \delta_{ik}\delta_{jh} \tag{1.4.2}$$

where $\delta$ is the Kronecker delta. In other words, the vector shape functions (1.4.1) are interpolating vector functions. These shape functions ensure the continuity of the tangential components of the vector field, but not the normal components. Therefore this kind of finite element is very useful for electromagnetic problems where the unknown field is the electric field or the magnetic one, whose tangential components are continuous across the interface between two different materials.

By numbering the edges of the tetrahedron from 1 to 6 as shown in Fig.1.7, the electric (or magnetic) field inside the tetrahedron is approximated as:

$$\vec{E}(x,y,z) = \sum_{s=1}^{6} E_s \vec{\alpha}_s(x,y,z) \tag{1.4.3}$$

where the six degrees of freedom $E_s$ s=1,…,6 are relative to the mean values of the tangential component of the field along the edge s:

$$E_s = \frac{1}{L_s}\int_{e_s} \vec{E} \cdot \hat{t}\, dl \tag{1.4.4}$$

where $\hat{t}$ is the versor of the edge $e_s$ and $L_s$ its length.

Note that in literature several ways exist to extend the tetrahedral edge element to higher orders.

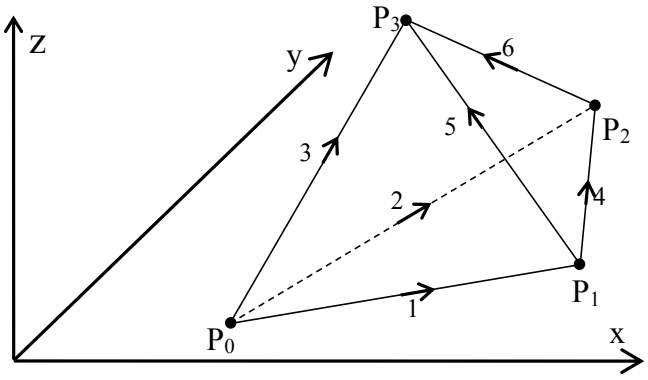



Fig. 0.7. - Tetrahedral edge element.

For a hexahedron of the first order the degrees of freedom are the average values of the tangential component of the vector field along the edges of the element, for a total of 12 unknowns, as shown in Fig. 1.8.
The vector shape functions of a hexahedral finite elements are:

$$
\begin{aligned}
\vec{\alpha}_1 &= \frac{L_1}{8}(1-\eta)(1-\zeta)\nabla\xi \qquad & \vec{\alpha}_7 &= \frac{L_7}{8}(1+\xi)(1-\eta)\nabla\zeta \\
\vec{\alpha}_2 &= \frac{L_2}{8}(1-\xi)(1-\zeta)\nabla\eta \qquad & \vec{\alpha}_8 &= \frac{L_8}{8}(1+\xi)(1-\eta)\nabla\zeta \\
\vec{\alpha}_3 &= \frac{L_5}{8}(1+\xi)(1-\eta)\,\nabla\zeta \qquad & \vec{\alpha}_9 &= \frac{L_9}{8}(1-\eta)(1+\zeta)\nabla\xi \\
\vec{\alpha}_4 &= \frac{L_4}{8}(1+\xi)(1-\zeta)\nabla\eta \qquad & \vec{\alpha}_{10} &= \frac{L_{10}}{8}(1+\xi)(1+\zeta)\nabla\eta \\
\vec{\alpha}_5 &= \frac{L_6}{8}(1+\xi)(1-\eta)\nabla\zeta \qquad & \vec{\alpha}_{11} &= \frac{L_{12}}{8}(1+\xi)(1+\zeta)\nabla\eta \\
\vec{\alpha}_6 &= \frac{L_3}{8}(1+\eta)(1-\zeta)\nabla\xi \qquad & \vec{\alpha}_{12} &= \frac{L_{11}}{8}(1+\eta)(1+\zeta)\nabla\xi
\end{aligned}
\tag{1.4.5}
$$

and the electric (or magnetic) field inside the hexahedron is approximated as:

$$
\vec{E}(x,y,z) = \sum_{s=1}^{12} E_s \vec{\alpha}_s(x,y,z) \tag{1.4.6}
$$

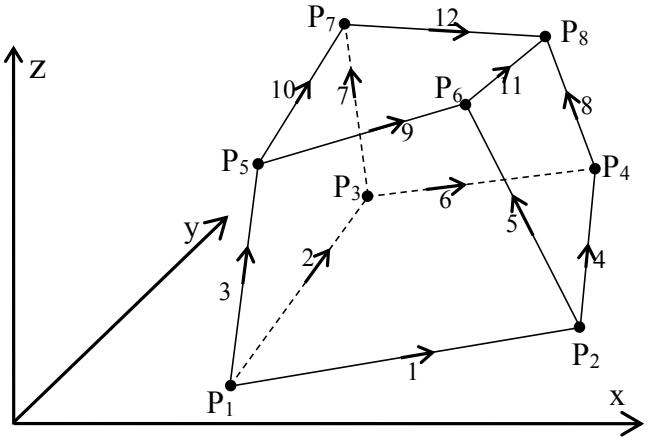



Fig. 0.8. – Hexahedral edge element.

## 1.5. Building of the global algebraic system

The FEM requires that the problem is formulated in terms of a functional,. whose minimization is equivalent to the solution of the partial differential equation with the associated boundary conditions. Consider the partial differential equation:

$$\mathcal{L} f = g \qquad \text{in } D \tag{1.5.1}$$

where $\mathcal{L}$ is a differential linear operator, f is an unknown scalar function and *g* is a known function (source). Assume that the function f satisfies the following Dirichlet and Neumann boundary conditions:

$$\begin{cases} f = f_d & \text{on } \Gamma_{Dir} \quad (\text{Dirichlet}) \\ \dfrac{\partial f}{\partial n} = 0 & \text{on } \Gamma_{Neu} \quad (\text{Neumann}) \end{cases} \tag{1.5.2}$$

where $f_d$ is a known function defined on a subset $\Gamma_{Dir}$ of the boundary $\Gamma$ of the domain D, and $\Gamma_{Meu}$ is such that:

$$\Gamma_{dir} \cup \Gamma_{neu} = \Gamma \qquad \Gamma_{dir} \cap \Gamma_{neu} = \varnothing \tag{1.5.3}$$

It can be shown that the solution of the problem minimizes the functional [2]

$$\begin{cases} \min \ \mathfrak{I}(f) = \dfrac{1}{2}\displaystyle\int_D f \, \mathcal{L} f \, dx dy dz - \int_D g f \, dx dy dz \\ f = f_d \ \text{ on } \Gamma_{dir} \end{cases} \tag{1.5.4}$$

Having discretized the domain D by means of nodal finite elements of a given order, the unknown scalar field f can be represented as a linear combination of shape functions and nodal values:

$$f(x,y,z) \approx \sum_{n=1}^{N_{tot}} f_n \alpha_n(x,y,z) = [f_{tot}]^t [\alpha_{tot}] \tag{1.5.6}$$

where $N_{tot}$ is the total number of nodes in the finite element mesh, $f_n$ are the nodal values and the $\alpha_n$ the relative shape functions. The array $[f_{tot}]$ and $[\alpha_{tot}]$ are the column vectors of the nodal values and of the shape functions, respectively; the superscript t denotes transposition.
Without loss of generality, we can number first the $N_{int}$ internal nodes, after the $N_{neu}$ nodes on the Neumann boundary $\Gamma_{neu}$, and finally the $N_{dir}$ nodes on the Dirichlet boundary $\Gamma_{dir}$. The number of unknowns N is given by:

$$N = N_{int} + N_{neu} \tag{1.5.7}$$

By using the approximation (1.5.6), the functional becomes:

$$\mathfrak{I}(f) \approx \frac{1}{2}[f_{tot}]^t \left( \int_D [\alpha_{tot}] \mathcal{L} [\alpha_{tot}]^t \, dx dy dz \right) [f_{tot}] - [f_{tot}]^t \int_D [\alpha_{tot}] g \, dx dy dz \tag{1.5.8}$$

which can be put in matrix form as:

$$\mathfrak{I}(f) \approx \frac{1}{2}[f_{tot}]^t[A_{tot}][f_{tot}] - [f_{tot}]^t[g_{tot}] \tag{1.5.9}$$

where:
- $[A_{tot}]$ is the matrix of the coefficients, whose generic entry is:

$$a_{mn} = \frac{1}{2}\int_D (\alpha_m \mathcal{L}\alpha_n + \alpha_n \mathcal{L}\alpha_m)\,dx\,dy\,dz \tag{1.5.10}$$

- $[g_{tot}]$ is an array, whose generic entry is:

$$g_m = \int_D \alpha_m g\,dx\,dy\,dz \tag{1.5.11}$$

Now we can partition the arrays $[f_{tot}]$ as

$$[f_{tot}] = \begin{bmatrix} [f] \\ [f_d] \end{bmatrix} \tag{1.5.12}$$

This partition induces similar partitions in the matrix $[A_{tot}]$ and in the array $[g_{tot}]$, so that

$$\mathfrak{I}(f) \approx \frac{1}{2}\begin{bmatrix} [f] \\ [f_d] \end{bmatrix}^t \begin{bmatrix} [A] & [A_d] \\ [A_d]^t & [A_{dd}] \end{bmatrix} \begin{bmatrix} [f] \\ [f_d] \end{bmatrix} - \begin{bmatrix} [f] \\ [f_d] \end{bmatrix}^t \begin{bmatrix} [g] \\ [g_d] \end{bmatrix} \tag{1.5.13}$$

where we have assumed that the matrix $[A_{tot}]$ is symmetric; this is a very common case and is verified if the differential operator $\mathcal{L}$ is self adjoint. By expanding (1.5.13), we get:

$$\begin{aligned} \mathfrak{I}(f) \approx \frac{1}{2}[f]^t[A][f] + \frac{1}{2}[f]^t[A_d][f_d] + \frac{1}{2}[f_d]^t[A_d]^t[f] + \\ + \frac{1}{2}[f_d]^t[A_{dd}][f_d] - [f]^t[g] - [f_d]^t[g_d] \end{aligned} \tag{1.5.14}$$

By noting that

$$[f]^t[A_d][f_d] = [f_d]^t[A_d]^t[f] \tag{1.5.15}$$

we can rewrite (1.5.14) as:

$$\begin{aligned} \mathfrak{I}(f) \approx \frac{1}{2}[f]^t[A][f] + [f]^t\{[A_d][f_d] - [g]\} + \\ + \left\{\frac{1}{2}[f_d]^t[A_{dd}][f_d] - [f_d]^t[g_d]\right\} \end{aligned} \tag{15.16}$$

Note that this expression is a non-homogeneous quadratic form in the nodal unknowns, which, in addition to the quadratic terms, it also contains linear and constant terms.
In order to minimize the functional, we impose that its partial derivatives with respect to the unknown nodal values $f_n$, n=1,...,N are zero:

$$\frac{\partial \Im}{\partial f_n} = 0 \qquad n=1,2,\ldots,N \tag{1.5.17}$$

or in matrix notation:

$$\frac{\partial \Im([f])}{\partial [f]} \approx [A][f] + [A_d][f_d] - [g] = [0] \tag{1.5.18}$$

By setting:

$$[b] = -[A_d][f_d] + [g] \tag{1.5.19}$$

we obtain the global algebraic system of N equations in N unknowns:

$$[A][f] = [b] \tag{1.5.20}$$

Another approach, which leads to the same global system, is the Galerkin method. By this method the partial differential equation (1.5.1) is rewritten in homogeneous form, multiplied by a shape function $\alpha_n$ and then integrated over the whole domain:

$$\int_D (\mathcal{L} f - g)\alpha_n dx dy dz = 0 \qquad n=1,\ldots,N \tag{1.5.21}$$

By substituting to f its approximation (1.5.6), we obtain the global system (1.5.21) again. The Galerkin method is more general than the functional approach, since it can be applied also in the cases in which the functional does not exist, as, for example, in the solution of skin effect problems by means of the integrodifferential Konrad's equation.

## 1.6. Solution of the global system

The methods of solution of linear algebraic equations are classified as direct and iterative solvers.
The direct solvers perform a finite number of operations, which depends on the order of the matrix, in general $O(N^3)$. Such methods obtain an exact solution in exact arithmetic, and no truncation error is introduced.
The iterative solvers are based on repeated corrections of an approximate solution; they introduce a truncation error and, in general it is not possible to foresee a priori the number of steps needed.
The direct methods are used only for systems of limited number of unknowns (some thousands). If the matrix is banded, symmetric and positive defined (SPD) it is possible to solve greater systems (some tens of thousands). Since the FEM linear systems are very big, in general iterative solvers are used.
Consider the linear algebraic system:

$$[A][x]=[b] \tag{1.6.1}$$

where [A] is a square matrix and [b] the known term array. We intend to use an iterative solver to solve (1.6.1) for [x]. Let [y] be an approximate solution, previously computed. We define the error:

$$[e]=[x]-[y] \tag{1.6.2}$$

Since the solution [x] is unknown, also the error [e] is unknown, but it is possible to compute the residual:

$$[r]=[A][e]=[A][x]-[A][y]=[b]-[A][y] \tag{1.6.3}$$

By using the approximate solving method by which we have computed [y], we obtain another approximation [e'] for [e]:

$$[e'] \cong [A]^{-1}[r] \tag{1.6.4}$$

We set:

$$[y']=[y]+[e'] \tag{1.6.5}$$

By repeating this procedure, more good solutions are found. Now it is necessary to specify the algorithm by which to build such approximate solutions. Since this algorithm is applied repeatedly, it needs to be fast in the building of the approximate solutions.
We decompose the matrix [A] as:

$$[A]=[D]-[L]-[U] \tag{1.6.6}$$

where [D] is the diagonal of [A], and –[L] and –[U] are the lower and upper triangular parts of [A]. The linear system is rewritten as:

$$\{[D]-[L]-[U]\}[x]=[b] \tag{1.6.7}$$

and from this we find:

$$[D][x]=\{[L]+[U]\}[x]+[b] \tag{1.6.8}$$

and also:

$$[x]=[D]^{-1}\{[L]+[U]\}[x]+[D]^{-1}[b] \tag{1.6.9}$$

From this we have the following iterative scheme (Gauss-Jacobi) [4]:

$$[y^{(n)}]=[D]^{-1}\{[L]+[U]\}[y^{(n-1)}]+[D]^{-1}[b] \tag{1.6.10}$$

If $[y^{(n)}]=[x]$ at the nth step, then $[y^{(m)}]=[x]$ for all the subsequent steps m>n. This iterative scheme fall within the general case described above. In fact we have:

$$\begin{aligned}[y^{(n)}]&=[D]^{-1}\{[L]+[U]\}[y^{(n-1)}]+[D]^{-1}[b]=\\&=[D]^{-1}\{[D]-[D]+[L]+[U]\}[y^{(n-1)}]+[D]^{-1}[b]=\\&=[y^{(n-1)}]-[D]^{-1}[A][y^{(n-1)}]+[D]^{-1}[b]=\\&=[y^{(n-1)}]+[D]^{-1}[r]\end{aligned} \tag{1.6.11}$$

Obviously the algorithm which finds fast the approximate solutions is the diagonal system having [D] as matrix of coefficients.
Consider now another iterative scheme (Gauss-Seidel) [4]:

$$[y^{(n)}]=\{[D]-[L]\}^{-1}[U][y^{(n-1)}]+\{[D]-[L]\}^{-1}[b] \tag{1.6.12}$$

If $[y^{(n)}]=[x]$ at the nth step, then $[y^{(m)}]=[x]$ for all the subsequent steps m>n. This iterative scheme fall within the general case described above.

$$\begin{aligned}[y^{(n)}]&=\{[D]-[L]\}^{-1}[U][y^{(n-1)}]+\{[D]-[L]\}^{-1}[b]=\\&=\{[D]-[L]\}^{-1}\{[D]-[L]-[D]+[L]+[U]\}[y^{(n-1)}]+\{[D[-[L]\}^{-1}[b]=\\&=[y^{(n-1)}]-\{[D]-[L]\}^{-1}[A][y^{(n-1)}]+\{[D[-[L]\}^{-1}[b]=\\&=[y^{(n-1)}]+\{[D]-[L]\}^{-1}[r]\end{aligned} \tag{1.6.13}$$

The linear system which is solved to obtain the approximate solutions is the system having [D]–[L] as matrix of coefficients.
The two methods are convergent if their iteration matrices, that is $[D]^{-1}([L]+[U])$ for the Gauss-Jacobi solver and $([D]–[L])^{-1}[U]$ for the Gauss-Seidel one, have spectral radii less than 1. It possible to show that the Gauss-Jacobi method converges if the matrix [A] è strictly diagonal dominant, whereas the Gauss-Seidel converges if [A] is SPD. Note that such conditions are sufficient but not necessary.
This two methods in general converge slowly, so that very often another method, called the conjugate gradient (CG) is used [4].
Consider the linear algebraic system (1.6.1) and assume that [A] is a square matrix of order N, symmetric and positive definite. The solution of such a system can be seen as the minimization of the function:

$$g([x]) = \frac{1}{2}[x]^t[A][x] - [x]^t[b] \tag{1.6.14}$$

Having defined the residual as in (1.6.3) in relation to an approximate solution, the conjugate gradient algorithm is described in the following.

0) the counter k is initialized: k=0

1) given the initial approximate solution $[x^{(0)}]$, the initial residual and direction are computed:

$$[r^{(0)}] = [A]\,[x^{(0)}] - [b] \qquad [p^{(0)}] = [r^{(0)}] \tag{1.6.15}$$

2) the function g([x]) is minimized along the straight line through the point $[x^{(n)}]$ and having the direction $[p^{(n)}]$. In other words, the function $g(\xi)=g([x^{(n)}] + \xi\,[p^{(n)}])$ is minimized with respect to the real parameter $\xi$. The minimum is obtained for the parameter value:

$$\xi_n = -\frac{[p^{(n)}]^t[r^{(n)}]}{[p^{(n)}]^t[A][p^{(n)}]} \tag{1.6.16}$$

3) a new approximate solution is found as:

$$[x^{(n+1)}] = [x^{(n)}] + \xi_n[p^{(n)}] \tag{1.6.17}$$

4) and the relative residual is

$$[r^{(n+1)}] = [r^{(n)}] + \xi_n[A][p^{(n)}] \tag{1.6.18}$$

5) a new search direction is set:

$$[p^{(n+1)}] = [r^{(n+1)}] + \beta_n[p^{(n)}] \tag{1.6.19}$$

with:
$$\beta_n = -\frac{[p^{(n)}]^t[A][r^{(n+1)}]}{[p^{(n)}]^t[A][p^{(n)}]} \tag{1.6.20}$$

6) the counter k is increased by 1: k=k+1;

7) the convergence is tested:

$$100\frac{\left|[x^{(n+1)}] - [x^{(n)}]\right|_2}{\left|[x^{(n+1)}]\right|_2} < \varepsilon \tag{1.6.21}$$

where $\varepsilon$ is a user-selected small values;

8) if the test is positive, the algorithm stops; otherwise it goes back to step 2).

Note that the minimization is performed on the quadratic function:

$$
\begin{aligned}
&g(\xi) = g([x^{(n)}] + \xi[p^{(n)}]) = \\
&= \frac{1}{2}\left\{[x^{(n)}] + \xi[p^{(n)}]\right\}^t [A]\left\{[x^{(n)}] + \xi[p^{(n)}]\right\} - \left\{[x^{(n)}] + \xi[p^{(n)}]\right\}^t [b] = \\
&= g([x^{(n)}]) + \frac{1}{2}\xi^2 [p^{(n)}]^t [A][p^{(n)}] + \xi[p^{(n)}]^t [r^{(n)}]
\end{aligned}
\tag{1.6.22}
$$

from which it is easy to find the (1.6.16).
Moreover, from (1.6.16) it follows that:

$$
\xi_n [p^{(n)}]^t [A][p^{(n)}] + [p^{(n)}]^t [r^{(n)}] = 0 \tag{1.6.23}
$$

and also:

$$
[p^{(n)}]^t \left\{\xi_n [A][p^{(n)}] + [r^{(n)}]\right\} = 0 \tag{1.6.24}
$$

and by virtue of (1.6.18):

$$
[p^{(n)}]^t [r^{(n+1)}] = 0 \tag{1.6.24}
$$

that is $[p^{(n)}]$ and $[r^{(n+1)}]$ are orthogonal.
Finally, note that, by imposing that two consecutive residuals are orthogonal, we have:

$$
[r^{(n)}]^t [r^{(n+1)}] = [r^{(n)}]^t [r^{(n)}] + \xi_n [r^{(n)}]^t [A][p^{(n)}] = 0 \tag{1.6.25}
$$

and so:

$$
\xi_n = -\frac{[r^{(n)}]^t [r^{(n)}]}{[r^{(n)}]^t [A][p^{(n)}]} \tag{1.6.26}
$$

which coincides with the (1.6.16). By setting:

$$
[p^{(n)}] = [r^{(n)}] + \beta_{n-1} [p^{(n-1)}] \tag{1.6.27}
$$

we note that the numerators in (1.1.16) and (1.6.26) are the same:

$$
\begin{aligned}
[r^{(n)}]^t [p^{(n)}] &= [r^{(n)}]^t [r^{(n)}] + \beta_{n-1} [r^{(n)}]^t [p^{(n-1)}] = \\
&= [r^{(n)}]^t [r^{(n)}]
\end{aligned}
\tag{1.6.28}
$$

If we impose that two successive directions are conjugate with respect to the matrix [A], that is

$$
[p^{(n+1)}][A][p^{(n)}] = 0 \tag{1.6.29}
$$

one obtains that:

$$
\beta_n = -\frac{[p^{(n)}]^t [A][r^{(n+1)}]}{[p^{(n)}]^t [A][p^{(n)}]} \tag{1.6.30}
$$

By this choice, also the numerators in (1.6.16) e (1.6.20) are the same:

$$
\begin{aligned}
[p^{(n)}]^t[A][p^{(n)}] &= \left\{[r^{(n)}] + \beta_{n-1}[p^{(n-1)}]\right\}^t[A][p^{(n)}] = \\
&= [r^{(n)}]^t[A][p^{(n)}] + \beta_{n-1}[p^{(n-1)}]^t[A][p^{(n)}] = \\
&= [r^{(n)}]^t[A][p^{(n)}]
\end{aligned}
\tag{1.6.31}
$$

## References Chapter 1


[1] O. C. Zienkiewicz and R. I. Taylor, *The Finite Element Method,* McGraw-Hill, Maidenhead, 1991.

[2] P. P. Silvester and R. L. Ferrari, *Finite Elements for Electrical Engineers*, Cambridge University Press, Cambridge, 1990.

[3] J. M. Jin, *The Finite Element Method in Electromagnetics*, Wiley, New York, 1993.

[4] G. H. Golub and C. F. Van Loan, *Matrix Computations*, John Hopkins Univ. Press, London, 1996.

# Chapter 2

# Electromagnetic FEM analysis

## 2.1. The Maxwell's equations

In order to solve problems of propagation of electromagnetic wave, it necessary to solve the Maxwell's equations in time-harmonic behavior [1,2]:

$$\mathrm{div}\,\vec{D} = \rho \tag{2.1.1}$$

$$\mathrm{div}\,\vec{B} = 0 \tag{2.1.2}$$

$$\mathrm{rot}\,\vec{E} = -j\omega\vec{B} \tag{2.1.3}$$

$$\mathrm{rot}\,\vec{H} = \vec{J} + j\omega\vec{D} \tag{2.1.4}$$

were $\omega=2\pi f$ is the angular frequency (rad/s) and f the frequency (Hz), $\vec{E}$ is the electric field (V/m), $\vec{H}$ the magnetic field (A/m), $\vec{J}$ the current density (A/m$^2$), $\vec{D}$ the electric induction (C/m$^2$), $\vec{B}$ the magnetic induction (Wb/m$^2$), and $\rho$ the volume charge density (C/m$^3$).
These vector fields are related by the constitutive laws, which describe the media in which they are located:

$$\vec{D} = \varepsilon\vec{E} \tag{2.1.5}$$

$$\vec{B} = \mu\vec{H} \tag{2.1.6}$$

$$\vec{J} = \sigma\vec{E} \tag{2.1.7}$$

where $\varepsilon$ is the electric permittivity (F/m), $\mu$ the magnetic permeability (H/m) and $\sigma$ the electric conductivity (S/m).
In addition to these equations, it is necessary to impose the boundary conditions.
In the solution of the problems of scattering of electromagnetic waves, an incident wave $\vec{E}_{inc}$ (or $\vec{H}_{inc}$), analytically known, irradiates one or more non-homogeneous objects in a surrounding homogeneous unbounded medium, very often the vacuum, characterized by

$$\begin{aligned} \varepsilon_0 &= 8.8541878\cdot 10^{-12}\ \mathrm{F/m} \\ \mu_0 &= 4\pi\cdot 10^{-7}\ \mathrm{H/m} \\ \sigma_0 &= 0\ \ \mathrm{S/m} \end{aligned} \tag{2.1.8}$$

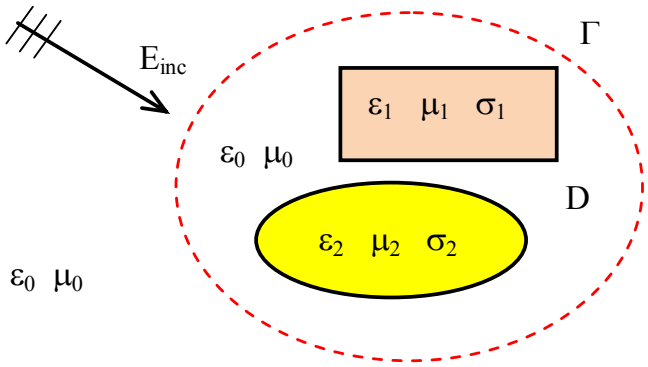



Fig. 2.1. Electromagnetic wave incident on non homogeneous objects.

## 2.2. Scattering of electromagnetic waves

A plane incident wave is given by [2]:

$$\vec{E}_{inc} = E_{max} e^{-jk_0(n_x x + n_y y + n_z z)} \tag{2.2.9}$$

where $n_x$, $n_y$ and $n_z$ are the cosines of the wave propagation direction.
The scattered electromagnetic wave $\vec{E}_{scat}$ (or $\vec{H}_{scat}$), due to the presence of the objects, is superimposed to the incident one to obtain the total electric field:

$$\vec{E} = \vec{E}_{inc} + \vec{E}_{scat} \tag{2.2.10}$$

which satisfies the Helmholtz's equation:

$$\nabla \times \left( \nu_r \nabla \times \vec{E} \right) - k_0^2 \varepsilon'_r \vec{E} = 0 \tag{2.2.11}$$

where $\nu_r$ is the relative reluctivity ($\nu_r = 1/\mu_r$), $\varepsilon'_r$ is the relative complex permittivity given by:

$$\varepsilon'_r = \varepsilon_r - j\frac{\sigma}{\omega \varepsilon_0} \tag{2.2.12}$$

and $k_0$ is the free-space wavenumber:

$$k_0 = \omega \sqrt{\varepsilon_0 \mu_0} \tag{2.2.13}$$

The boundary conditions, to be imposed on the boundaries of the analysis domain, are the following:

- on the perfect conductors (PEC):

$$\hat{n} \times \vec{E} = 0 \qquad (\text{or} \quad \hat{n} \times \text{rot}\, \vec{H} = 0) \tag{2.2.14}$$

- on the symmetry planes of the electric field:

$$\hat{n} \times \vec{E} = 0 \qquad (\text{or} \quad \hat{n} \times \text{rot}\, \vec{H} = 0) \tag{2.2.15}$$

- on the symmetry planes of the electric field:

$$\hat{n}\times \mathrm{rot}\,\vec{E}=0 \qquad (\text{or } \vec{n}\times\vec{H}=0) \tag{2.2.16}$$

- on the far fictitious boundaries:

$$\hat{n}\times\nabla\times\overline{E}_{scat}+jk_0\hat{n}\times(\hat{n}\times\overline{E}_{scat})=0 \tag{2.2.17}$$

In order to use the FEM, the unbounded free space outside the scatterers must be truncated by means of a fictitious truncation boundary $\Gamma_F$. The resulting bounded domain is discretized by tetrahedral edge elements. By applying the Galerkin method to the Helmholtz equation (2.2.11), we have:

$$\iiint_D\left(\nabla\times(\nu_r\nabla\times\vec{E})-k_0^2\varepsilon'_r\vec{E}\right)\cdot\vec{\alpha}_i\,dxdydz=0 \qquad i=1,..,N \tag{2.2.18}$$

where $\vec{\alpha}_i$ is the vector shape function of the i-th edge and N is the number of edges not lying on the Dirichlet boundary. By applying the first Green formula to (2.2.18), we get:

$$\begin{aligned}\iiint_D \nu_r\nabla\times\vec{E}\cdot\nabla\times\vec{\alpha}_i\,dxdydz-k_0^2\iiint_D\varepsilon'_r\vec{E}\cdot\vec{\alpha}_i\,dxdydz=\\ =-\oiint_\Gamma\nu_r\hat{n}\times\nabla\times\vec{E}\cdot\vec{\alpha}_i dS\end{aligned} \qquad i=1,2,..,N \tag{2.2.19}$$

where $\Gamma$ is the boundary of the analysis domain and $\hat{n}$ the outward normal versor to the boundary. By substituting to the electric field its approximation:

$$\vec{E}(x,y,z)=\sum_{j=1}^{N_e}E_j\vec{\alpha}_j(x,y,z) \tag{2.2.20}$$

the following element matrix must be computed:

$$s_{ij}^{(k)}=\iiint_{E_k}\nabla\times\vec{\alpha}_j\cdot\nabla\times\vec{\alpha}_i\,dxdydz \tag{2.121}$$

$$t_{ij}^{(k)}=\iiint_{E_k}\vec{\alpha}_j\cdot\vec{\alpha}_i\,dxdydz \tag{2.2.22}$$

for each finite element $E_k$.
Finally, by imposing the boundary conditions, we obtain the global algebraic system.
Electromagnetic scattering problems can be posed also in 2D. This happens in if the wave is E- or H-polarized along a given direction, say the z axis, and all objects are cylinders aligned along the same axis. In such cases, assumed as unknown the polarized field along the z axis, the problem becomes scalar. In the case of E-polarized wave, the incident wave is:

$$E_{inc}=E_{max}e^{-jk_0(n_x x+n_y y)} \tag{2.2.23}$$

and the Helmholtz's equation becomes:

$$-\frac{\partial}{\partial x}\left(\nu_r \frac{\partial E}{\partial x}\right)-\frac{\partial}{\partial y}\left(\nu_r \frac{\partial E}{\partial x}\right)-k_0^2\varepsilon'_r E=0 \qquad (2.2.24)$$

with the boundary conditions E=0 on the perfect conductors (Dirichlet), $\partial E/\partial n=0$ on the symmetry planes (Neumann) and $\partial E_{scat}/\partial n+jk_0E_{scat}=0$ on the truncation boundary. By discretizing the analysis domain by means of nodal finite elements (triangles and/or quadrangles), the Galerkin method gives:

$$\iint_D \left(\nabla\cdot(\nu_r\nabla E)+k_0^2\varepsilon'_r E\right)\alpha_i\,dxdy=0 \qquad i=1,2,\ldots,N \qquad (2.2.25)$$

where $\alpha_i$ is the scalar shape function of the i-th node and N is the number of nodes not lying in the Dirichlet boundary. By virtue of second Green formula, the (2.2.25) is rewritten as:

$$\iint_D \nu_r\nabla E\cdot\nabla\alpha_i\,dxdy-k_0^2\iint_D \varepsilon'_r E\alpha_i\,dxdy=\oint_\Gamma \nu_r\alpha_i\frac{\partial E}{\partial n}dl \qquad i=1,2,\ldots,N \qquad (2.2.26)$$

where $\Gamma$ is the boundary of the analysis domain D and $\hat{n}$ is the outward normal versor to $\Gamma$. By substituting to the electric field its approximation:

$$E(x,y)=\sum_{j=1}^{N} E_j\alpha_j(x,y) \qquad (2.2.26)$$

and by imposing the boundary conditions, we obtain the global system. In order to derive these equations, we have to compute the following geometrical coefficients [1]:

$$s_{ij}^{(k)}=\iint_{E_k}\nabla\alpha_j\cdot\nabla\alpha_i\,dxdy \qquad (2.2.27)$$

$$t_{ij}^{(k)}=\iint_{E_k}\alpha_j\alpha_i\,dxdy \qquad (2.2.28)$$

for each finite element $E_k$ in the mesh. The coefficients (2.2.27) and (2.2.28) form the stiffness and metric matrices of the finite element, respectively.

## 2.3. The FEM-RBCI method in 2D

Let us consider a set of conducting and/or dielectric objects, infinitely extended in the z-direction, surrounded by an unbounded homogeneous dielectric medium (free space). A given time-harmonic electromagnetic wave $E_{inc}$, E-polarized along the z-axis irradiates these objects, so that a scattered field $E_{scat}$ is excited extending to infinity. The total field E(x,y) (given by $E_{inc}+E_{scat}$ outside the dielectric objects, if any) satisfies the two-dimensional scalar Helmholtz equation

$$\nabla\cdot\nu_r\nabla E+k_0^2\varepsilon_r E=0 \qquad (2.3.1)$$

where $\nu_r$ is the relative magnetic reluctivity, $\varepsilon_r$ the relative electrical permittivity and $k_0$ is the free-space wavenumber $k_0=\omega(\varepsilon_0\mu_0)^{1/2}$, in which $\omega$ is the wave angular frequency and $\mu_0$

and $\varepsilon_0$ are the free-space magnetic permeability and electrical permittivity, respectively. Homogeneous Dirichlet conditions hold on the perfectly conducting scatterer surface $\Gamma_C$, if any. In addition the scattered field $E_{scat}$ satisfies the Sommerfeld radiation condition at infinity. Let us introduce a fictitious boundary $\Gamma_F$ enclosing all the scattering objects. Note that this boundary does not need to be constituted by a single closed curve, but several closed curves can be used, each one, for example, enclosing one scattering object. A Robin boundary condition is initially imposed on such a boundary [3-5]:

$$\Re E = \frac{\partial E}{\partial n} + jk_0 E = \psi \quad \text{on}\,\Gamma_F \tag{2.3.2}$$

where the normal derivative is computed in the outward direction and $\psi$ is a user-selected function of the position on the fictitious boundary (a good initial choice for $\psi$ is given by $\Re E_{inc}$). By applying the Galerkin method to the bounded domain D, delimited by $\Gamma_F$ and $\Gamma_C$ and discretized by means of Lagrangian finite elements, the following algebraic system is obtained:

$$\mathbf{A}\,\mathbf{E} = \mathbf{C}\boldsymbol{\Psi} \tag{2.3.3}$$

where **A** is a square matrix depending on geometry and dielectric materials, **E** is the array of the unknown nodal field values (including that on $\Gamma_F$), $\boldsymbol{\Psi}$ is the array of the nodal values of the right hand-side of the Robin condition on $\Gamma_F$ and **C** is a rectangular matrix.
Consider now the total field outside the surface enclosing the scatterer and enclosed by $\Gamma_F$, with a nonzero distance between them. The total field outside $\Gamma_S$ is given by:

$$E = E_{inc} - \int_{\Gamma_S} \left( G \frac{\partial E}{\partial n'} - E \frac{\partial G}{\partial n'} \right) ds' \tag{2.3.4}$$

where $n'$ is the outward normal to $\Gamma_S$ (toward $\Gamma_F$), and G is the two-dimensional free-space Green's function, given by [2]:

$$G = -\frac{1}{4} j H_0^{(2)}(k_0 r) \tag{2.3.5}$$

where $H_0^{(2)}$ is the Hankel function of the second kind and zero-order and r is the distance between a point on $\Gamma_S$ and a point on $\Gamma_F$. Owing to the fact that $\Gamma_F$ and $\Gamma_S$ are separated by a distance greater than zero, function $\psi$ can then be expressed as:

$$\psi = \Re\phi_{inc} - \int_{\Gamma_M} \left( \frac{\partial E}{\partial n'} \Re G - E \frac{\partial \Re G}{\partial n'} \right) ds' \quad . \tag{2.3.6}$$

In the FEM approximation, this relation is rewritten as:

$$\boldsymbol{\Psi} = \boldsymbol{\Psi}_{inc} + \mathbf{M}\mathbf{E} \tag{2.3.7}$$

where $\boldsymbol{\Psi}_{inc}$ is the vector of the values of the operator $\Re$ applied to the incident field on the nodes of the fictitious boundary and **M** is a rectangular matrix in which null columns appear for the nodes not belonging to the elements external to $\Gamma_S$ and having a side lying on it.

System (2.3.3) will be referred to as the FEM part of the equations, while (2.3.7) will be referred to as the integral part of the equations.
Equations (2.3.3) and (2.3.7) form an algebraic system, which can be efficiently solved with the following iterative scheme:

a) having arbitrarily guessed the vector $\mathbf{\Psi}$,
b) equation (3) is solved for the vector $\mathbf{E}$;
c) another guess for the Robin condition on $\Gamma_F$ is obtained by means of (2.3.7);
d) the procedure is then iterated until convergence takes place.

The convergence is checked by computing the norm of the difference between the solution at the current step with the previous one, and dividing this quantity by the norm of the current solution; when this ratio is less than the convergence tolerance selected by the user the iteration is stopped.
A simple study of the convergence of this iterative procedure can made led by formally relating the initial guess $\mathbf{\Psi}^{(0)}$ to the true values of the Robin condition $\mathbf{\Psi}_t$ by means of the error $\mathbf{\Psi}_e^{(0)}$:

$$\mathbf{\Psi}^{(0)} = \mathbf{\Psi}_t + \mathbf{\Psi}_e^{(0)} \tag{2.3.8}$$

Solving equation (2.3.3) for the vector $\mathbf{E}$, we obtain the field solution at the 0-th step:

$$\mathbf{E}^{(0)} = \mathbf{A}^{-1}\mathbf{C}\mathbf{\Psi}_t + \mathbf{A}^{-1}\mathbf{C}\mathbf{\Psi}_e^{(0)} \tag{2.3.9}$$

in which the first term gives, by definition, the true field solution $\mathbf{E}_t$, whereas the second one represents the error $\mathbf{E}_e^{(0)}$. Starting from this solution, the new guess for the Robin condition on $\Gamma_F$ is computed, whose error is given by:

$$\mathbf{\Psi}_e^{(1)} = \mathbf{P}\mathbf{\Psi}_e^{(0)} \tag{2.3.10}$$

where $\mathbf{P}$ is a square matrix (of order equal to the number of nodes on $\Gamma_F$) given by:

$$\mathbf{P} = \mathbf{M}\mathbf{A}^{-1}\mathbf{C}\,. \tag{2.3.11}$$

By further continuing the procedure, we can generalize (10) for the n-th step:

$$\mathbf{\Psi}_e^{(1)} = \mathbf{P}^n\mathbf{\Psi}_e^{(0)}\,. \tag{2.3.12}$$

From this relation it is easy to understand how the procedure may converge to the true solution for every initial error $\mathbf{\Psi}_e^{(0)}$ on the first guess for $\mathbf{\Psi}$. This happens if and only if the spectral radius $\rho$ of matrix $\mathbf{P}$ is lower than 1. In this case, in fact, $\mathbf{P}^n \to \mathbf{0}$ as $n \to \infty$ and, consequently, $\mathbf{\Psi}_e^{(n)} \to 0$ whatever $\mathbf{\Psi}_e^{(0)}$. If this condition does not apply, divergence may occur.
The spectral radius $\rho$ depends in a complicated way on the distance of $\Gamma_F$ from the scatterers, on the whole FE discretization and on the frequency. Of course it is not possible to check condition $\rho<1$ at the beginning of iteration, since matrix $\mathbf{P}$ is not available.
Fortunately this is not a problem, since by suitably surrounding the scattering objects with two or more layers of finite elements, no divergence has been observed and, on the contrary,

the procedure is rapidly convergent to the true solution. This very attractive behavior of the procedure is lost if one tries to employ Dirichlet (or Neumann) boundary conditions on the fictitious boundary as is successfully done for static and quasi-static field problems for two reasons: equation (2.3.1) may have non vanishing solutions satisfying homogeneous boundary conditions; if there exist an incident field vanishing on $\Gamma_F$ then equation (2.3.4) is singular. On the contrary, with the use of an impedance condition, system (2.3.3) which constitutes the FEM part of the system cannot be singular. The way in which the singularity relative to the integral equation (2.3.6) is avoided is a little more involved. In the global system (2.3.3),(2.3.7) the known term is constituted by the array $\boldsymbol{\Psi}_{inc}$ in equation (2.3.7); it results from the discretization of $\Re E_{inc}$ which vanishes if and only if $E_{inc}$ vanishes, so that this array is zero if, and only if, the incident wave is zero. In other words, the field source is well represented in the formulation.
Finally, some comments are in order about the computing time and memory requirements of the procedure. At first sight one could think that the procedure is too time- consuming. In reality this is not the case if the following points are fully exploited in implementation.
i) Since the FE mesh remains unchanged through the various iteration steps, matrices **A**, **M** and **C** do not change, so that they are computed only once, at the beginning of the procedure and saved for further use.
ii) Equation (2.3.3) may be solved efficiently by means of standard solvers, which exploit the matrix **A** sparsity and symmetry. Specifically, when a direct solver is used, matrix **A** must be decomposed only once (note that in the adaptive ABC described in [8] matrix **A** changes at each step, so that the use of a direct solver may become too time- consuming); when an iterative solver is used the FEM solution at a certain step is used as the initial guess for the next step, reducing the number of solver iterations.
iii) Small extensions of the domain D can be obtained by suitably placing the fictitious boundary near the scattering objects. The distance between the fictitious boundary and the scatterer surface is very short with respect to that necessary to find an acceptable solution by ABC methods. In addition, if the fictitious boundary is constituted by several closed curves, each one enclosing a scattering object, the domain is subdivided into disjoint pieces and, consequently, the global FEM system (2.3.3) is partitionable into independent subsystems, with a reduction in the overall computing time.
iv) The end-iteration test is conveniently restricted to the fictitious boundary, as suggested by (2.3.12), being sure that this assures convergence of the field solution in the domain.
v) A good initial guess for $\boldsymbol{\Psi}^{(0)}$ is $\boldsymbol{\Psi}_{inc}$ since the number of iterations are minimized (note that the selected Robin operator $\Re$ looks like an ABC one).
By implementing all the above items the iterative procedure can be made competitive with respect to other techniques as far as computing time and memory requirements are concerned. In addition the procedure is easy implementable in a pre-existing FEM code for bounded problems. Only one routine has to be developed which calculates the matrix **M** entries (very often a similar routine is already available in the post-processing program). Note that no singularities arise in these calculations since $\Gamma_F$ and $\Gamma_S$ do not have points in common.

## 2.4. The FEM-RBCI method in 3D

Consider a set of conducting and/or dielectric bodies, embedded in free space, lit up by an incident time-harmonic electromagnetic wave. A scattering problem is set up in terms of the total electric field, which satisfies the 3-D vector Helmholtz equation (a time factor $e^{j\omega t}$ has been assumed and suppressed):

$$\nabla\times\left(\mu_r^{-1}\nabla\times\overline{E}\right)-k_0^2\varepsilon_r\overline{E}=0 \tag{2.4.1}$$

where $\mu_r$ is the relative magnetic permeability, $\varepsilon_r$ the relative electrical permittivity and $k_0$ is the free-space wavenumber $k_0=\omega(\varepsilon_0\mu_0)^{1/2}$, in which $\omega$ is the wave angular frequency and $\mu_0$ and $\varepsilon_0$ are the free-space magnetic permeability and electrical permittivity, respectively. Homogeneous Dirichlet conditions ( $\hat{n}\times\overline{E}=0$) hold on the perfectly conducting scatterer surface $\Gamma_C$, if any. In addition the scattered field $E_{scat}$ satisfies the Sommerfeld radiation condition at infinity.

Let us introduce a closed fictitious boundary $\Gamma_F$, strictly enclosing the scatterer, as shown in Fig. 2.1. Note that when the scatterer is composed of several disjoint objects, the boundary can be constituted by several closed surfaces, each one, for example, enclosing a single object. On $\Gamma_F$, a nonhomogeneous Robin boundary condition is assumed [6-7]

$$\Re\overline{E}=\hat{n}\times\nabla\times\overline{E}+jk_0\hat{n}\times(\hat{n}\times\overline{E})=\overline{U} \tag{2.4.2}$$

where $\hat{n}$ is the outward normal to $\Gamma_F$ and $\overline{U}$ is an unknown vector to be determined.

Let us now discretize the bounded domain delimited by $\Gamma_F$ and by $\Gamma_C$ by means of edge elements such as first-order tetrahedra, whose vector shape functions are [1]

$$\vec{\alpha}_i=L_i\left(\zeta_{i1}\nabla\zeta_{i2}-\zeta_{i2}\nabla\zeta_{i1}\right) \tag{2.4.3}$$

where $\zeta_{i1}$ and $\zeta_{i2}$ are the local coordinates relative to the two nodes of the i-th edge and $L_i$ is its length. To simplify the description of the FE formulation and without loss of generality, we number first the edges on from one to, next the interior edges from $N_F$+1 to N, and finally the edges on $\Gamma_C$.

Applying FEM to (2.4.1) inside D with a homogeneous Dirichlet condition on $\Gamma_C$ and a boundary condition (2.4.2) on $\Gamma_F$, a linear algebraic system is obtained

$$\mathbf{AE}=\mathbf{BU} \tag{2.4.4}$$

where **A** is a complex and symmetric matrix, **B** links (2.4.2) with the right hand side of the FEM system, **E** is the array of the expansion coefficients for the electric field , and **U** is the array whose generic entry is given by:

$$U_j=\int_{\Gamma_F}\overline{U}\cdot\vec{\alpha}_j\,dS \tag{2.4.5}$$

Since U is unknown, in order to solve the scattering problem, another equation relating U to E needs to be derived. To this end let us now consider another surface, $\Gamma_M$, lying between the antenna and the fictitious boundary (see Fig. 2.1). At minimum this surface can be selected as coinciding with the scatterer surface itself. The total field outside $\Gamma_M$ can be expressed as:

$$\overline{E}(\bar{r})=\overline{E}_{inc}(\bar{r})+\int_{\Gamma_M}\left[\overline{\overline{G}}(\bar{r},\bar{r}')\cdot\left(\hat{n}'\times\nabla'\times\overline{E}(\bar{r}')\right)+\nabla\times g(\bar{r},\bar{r}')\times\left(\hat{n}'\times\overline{E}(\bar{r}')\right)\right]dS' \tag{2.4.6}$$

where G is the dyadic Green's function

$$\overline{\overline{G}}(\bar{r},\bar{r}')=\left(\overline{\overline{I}}+k_0^{-2}\nabla\nabla\right)g(\bar{r},\bar{r}') \tag{2.4.7}$$

with:

$$g_0(\bar{r},\bar{r}') = \frac{1}{4\pi|\bar{r}-\bar{r}'|} e^{-jk_0|\bar{r}-\bar{r}'|} \tag{2.4.8}$$

Taking the curl of (2.4.6) and performing some manipulations, the following expression is obtained:

$$\nabla\times\bar{E}(\bar{r}) = \nabla\times\bar{E}_{inc}(\bar{r}) + \int_{\Gamma_M} \left[k_0^2\overline{\overline{G}}(\bar{r},\bar{r}')\cdot\left(\hat{n}'\times\bar{E}(\bar{r}')\right) + \nabla g(\bar{r},\bar{r}')\times\left(\hat{n}'\times\nabla'\times\bar{E}(\bar{r}')\right)\right]dS' \tag{2.4.9}$$

Note that in (2.4.9), the curl operator has been put inside the integral sign due to the fact that $\Gamma_F$ and $\Gamma_M$ do not intersect. Substituting (2.4.6) and (2.4.9) in (2.4.2), an integral expression for U is easily obtained. Then taking into account (2.4.5), the following algebraic equation is derived:

$$\mathbf{U} = \mathbf{U}_{inc} + \mathbf{Q}\mathbf{E} \tag{2.4.10}$$

where $\mathbf{U}_{inc}$ is an array whose generic j-th entry (j=1,…$N_F$) is given by

$$\vec{U}_{inc,j} = \int_{\Gamma_F} \Re\vec{E}_{inc}\cdot\vec{\alpha}_j\,dS \tag{2.4.11}$$

and $\mathbf{Q}$ is an $N_F\times N$ rectangular matrix in which null columns appear for the internal edges not involved in the computation. Equations (2.4.4) and (2.4.10) together form the global algebraic system of the FEM-RBCI method, which can be conveniently solved by an iterative scheme as follows:
1) Select an arbitrary first guess for $\mathbf{U}$;
2) Solve equation (2.4.4) for $\mathbf{E}$, by means of a standard conjugate gradient solver (COCG);
3) Obtain an improved guess for $\mathbf{U}$ by means of (2.4.10);
4) If the procedure has converged, stop; otherwise go to 2).
This scheme can be seen as a two-block Gauss-Seidel iterative method:

$$\mathbf{E}^{(n)} = \mathbf{A}^{-1}\mathbf{B}\mathbf{U}^{(n)} \tag{2.4.11}$$

$$\mathbf{U}^{(n+1)} = \mathbf{U}_{inc} + \mathbf{Q}\mathbf{E}^{(n)} \tag{2.4.12}$$

In this way the symmetry and sparsity of matrix $\mathbf{A}$ is fully exploited. Moreover, since this procedure converges in a few iterations (generally 10-20), it also minimizes the number of multiplications of the dense matrix $\mathbf{Q}$ by a vector. It is now clear now that the price paid by FEM-RBCI in meshing some space around the conductors is worth it. In fact if a similar approach were used in FEM-BEM, one would have to invert the sub-matrix relative to the unknowns on the truncation boundary with a computational complexity of $O(N_F^3)$. On the other hand, if an iterative solver for non-symmetric complex matrices were be directly applied to the solution of the FEM-BEM global system the dense BEM sub-matrix would be multiplied by a vector at each solver iteration. Since the number of such iterations is very high (from several hundreds to some thousands for big 3D problems) we can state that the FEM-RBCI global system is cheaper to solve than the FEM-BEM one. Another advantage of FEM-RBCI with respect to FEM-BEM is that it avoids singularities in the integral equation since, as already said, the integration surface $\Gamma_M$ is different from the truncation one $\Gamma_F$.
The FEM-RBCI method can be made computationally more efficient if: a) the fictitious boundary is selected in such a way that very small extensions of the domain D are obtained: in general one or two layers of finite elements can be inserted between the scatterer and the

fictitious boundary; b) the computational complexity of the integral equation (2.4.9) is reduced from $O(N_F^2)$ to $O(N_F \log N_F)$ by means of the well-known fast multipole method (FMM) [9].

## 2.5. The perfectly matched layer (PML)

In dealing with electromagnetic problems in open boundaries, it is necessary to truncate the computational domain by means of a fictitious truncation boundary. The key question is how to do this without introducing significant errors in the model of the device to be tested. Some electromagnetic field problems by their own nature are confined in a specific region of space. Others present solutions that decay quickly in space, making irrelevant the error introduced by the truncation, if the computational domain is large enough. The problems regarding wave propagations, whose solutions oscillate and typically decay slowly, are the most complicated to truncate. Methods such as the simple truncation of the domain (using homogeneous Dirichlet or Neumann conditions) or the coordinate transformations would cause reflection invalidating the solution found.
In 1994 Berenger presents a new method [10]: instead of imposing a boundary condition that absorbs the wave, he uses a layer of artificial material that does not reflect the incident fields on it. The relative dielectric constant and the relative magnetic permeability of the material are both anisotropic. This material is a sort of picture frame for domain along the direction in which you want to simulate the open space (see Fig. 2.2); on the external surface of the adsorbing layer a Dirichlet boundary condition is imposed. When the electromagnetic field meets the absorbing layer, the amplitude of the wave decays exponentially; thus, in the case where the wave reaches the Dirichlet condition and is reflected, it must again pass through the absorbing layer, resulting largely attenuated and therefore negligible. This method takes the name of Perfectly Matched Layer (PML).

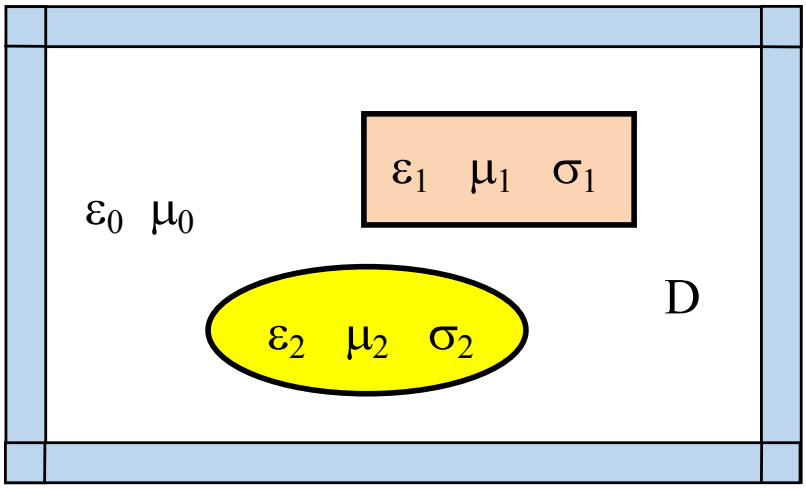



Fig. 2.2. – Scattering objects enclosed by a PML.

## References Chapter 2


[1] P. P. Silvester and R. L. Ferrari, *Finite Elements for Electrical Engineers*, Cambridge University Press, Cambridge, 1990.

[2] J. M. Jin, *The Finite Element Method in Electromagnetics*, Wiley, New York, 1993.

[3] S. Alfonzetti, G. Borzì, and N. Salerno, "Iteratively-improved Robin boundary conditions for the finite element solution of scattering problems in unbounded domains", *Int. J. Num. Meth. Engng*, vol. 42, pp. 601-629, 1998.

[4] S. Alfonzetti, G. Borzì, and N. Salerno, "Accelerating the Robin iteration procedure by means of GMRES", *Compel*, vol. 17, pp. 49-54, 1998.

[5] S. Alfonzetti, G. Borzì, and N. Salerno, "An iterative solution to scattering from cavity-backed apertures in a perfectly conducting wedge", *IEEE Trans. Magn*., vol. 34, pp. 2704-2707, 1998.

[6] S. Alfonzetti, B. Azzerboni, and G. Borzì, "Numerical computation of antenna parameters by means of RBCI", *Electromagnetics*, vol. 22, pp. 381-392, 2002.

[7] S. Alfonzetti, and G. Borzì, "Finite element solution to electromagnetic scattering problems by means of the Robin boundary condition iteration method", *IEEE Trans. Ant. Prop*., vol. 50, pp. 132-140, 2002.

[8] Y. Li and Z. Cendes, "High-accuracy absorbing boundary conditions", IEEE Trans. Magn., vol. 31, pp. 1524-1529, 1995.

[9] N. Engheta, W. D. Murphy, V. Rokhlin, and M. Vassiliou, "The Fast Multipole Method for Electromagnetic Scattering Computation," *IEEE Trans. Ant. Prop.*, vol. 40, pp. 634-641, 1992.

[10] J. P. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves", *J. Comput. Phys*., vol. 114, pp. 185-200, 1994.

# Chapter 3

# Stochastic Optimization

## 3.1. Generalities

The aim of an optimization problem is to find the value of some design parameters in order to minimize (maximize) a given quantity, called objective function. Moreover, the optimization problem can be subject to some restrictions (constraints) on the parameter ranges allowed.
In industrial applications, the optimized design is often problematic because of the simultaneous occurrence of many conflicting objectives. Besides, in optimised design it is often preferred to have a wide range of solutions to choose from, taking into account further design factors (cost, feasibility, …), instead of only considering the best one. There are different methods to solve this kind of problem, such as optimizing a single multi-objective function obtained by a weighted sum of the objectives or finding multiple Pareto-optimal solutions.
Many multi-objectives evolutionary algorithms exist in the literature but they can be extremely expensive. This is especially harmful in the design of electromagnetic devices, where each estimation of the objective function calls for a numerical solution of the electromagnetic problem by means of the Finite Element Method (FEM).
In this thesis, two stochastic optimization algorithms are presented: the Genetic Algorithms (GAs) and the Particle Swarm Optimization (PSO). Moreover, the Pattern Search (PS) deterministic algorithm is used to further improve the optima obtained by GAs and PSO.
This chapter is structured as follows. In section 3.2 the concepts of single-objective and multi-objective optimization are introduced and an overview of well-known evolutionary algorithms is presented. In sections 3.3 and 3.4 the Gas and PSO are outlined, respectively, whereas in section 3.5 the PS is briefly described.

## 3.2. Single- and multi-objective optimization

The goal of an optimization problem [1-2] can be formulated as follows: find the combination of some design parameters (independent variables) which minimize a given quantity, possibly subject to some restrictions on the parameter ranges allowed. The quantity to be optimized is termed the objective function; the parameters are called control or decision variables and their values may be changed in the search for the optimum; the restrictions on allowed parameter values are known as constraints. The general optimization problem can be stated mathematically as:

$$\begin{cases} \text{minimize } f(\mathbf{x}) & \mathbf{x} = [x_1, x_2, \ldots, x_N]_t \\ \\ \text{subject to} & c_i(\mathbf{x}) = 0, \quad i = 1, 2, \ldots, M' \\ & c_i(\mathbf{x}) \geq 0, \quad i = m'+1, \ldots, M. \end{cases} \tag{3.2.1}$$

where $f(\mathbf{x})$ is the objective function, $\mathbf{x}$ is the column vector of the N independent variables, and $c_i(\mathbf{x})$ is the set of constraint functions. Constraint equations of the form $c_i(\mathbf{x})=0$ are termed equality constraints, and those of the form $c_i(\mathbf{x}) \geq 0$ are inequality constraints. Taken together, $f(\mathbf{x})$ and $c_i(\mathbf{x})$ are known as the problem functions.
There are many optimization algorithms available to the computational scientist, but many methods are only appropriate for some types of problems. It is therefore important to be able to recognize the characteristics of a problem in order to identify an appropriate solution technique. Within each class of problems there are different minimization methods, varying in computational requirements, convergence properties, and so on.
The goal of optimization is to find global optimum $\mathbf{x}^*$ of the objective function $f(\mathbf{x})$, i.e. for a minimization problem,

$$f(\mathbf{x}^*) \leq f(\mathbf{y}) \qquad \forall \mathbf{y} \in V(\mathbf{x}), \mathbf{y} \neq \mathbf{x}^* \; (1.2) \tag{3.2.2}$$

where $V(\mathbf{x})$ is the set of feasible values of the control variables $\mathbf{x}$. Obviously, for an unconstrained problem $V(\mathbf{x})$ is infinitely large.
A point $\mathbf{y}^*$ is a strong local minimum of $f(\mathbf{x})$ if

$$f(\mathbf{y}^*) < f(\mathbf{y}) \qquad \forall \mathbf{y} \in N(\mathbf{y}, \eta), \mathbf{y} \neq \mathbf{y}^* \tag{3.2.3}$$

where $N(\mathbf{y},\eta)$ is defined as the set of feasible points contained in the neighborhood of $\mathbf{y}^*$, i.e., within some arbitrarily small distance $\eta$ from $\mathbf{y}^*$. For $\mathbf{y}^*$ to be a weak local minimum (maximum) only an inequality need be satisfied

$$f(\mathbf{y}^*) \leq f(\mathbf{y}) \qquad \forall \mathbf{y} \in N(\mathbf{y}, \eta), \mathbf{y} \neq \mathbf{y}^* \tag{3.2.4}$$

The different types of stationary points for an unconstrained univariate function are shown in **Error! Reference source not found.**. The situation is slightly more complex for constrained optimization problem, as shown in **Error! Reference source not found.**, where the presence of a constraint boundary (in the form of a simple bound on the permitted values of the control variable) can cause the global minimum to be an extreme value, an extremum (i.e., an endpoint), rather than a true stationary point.

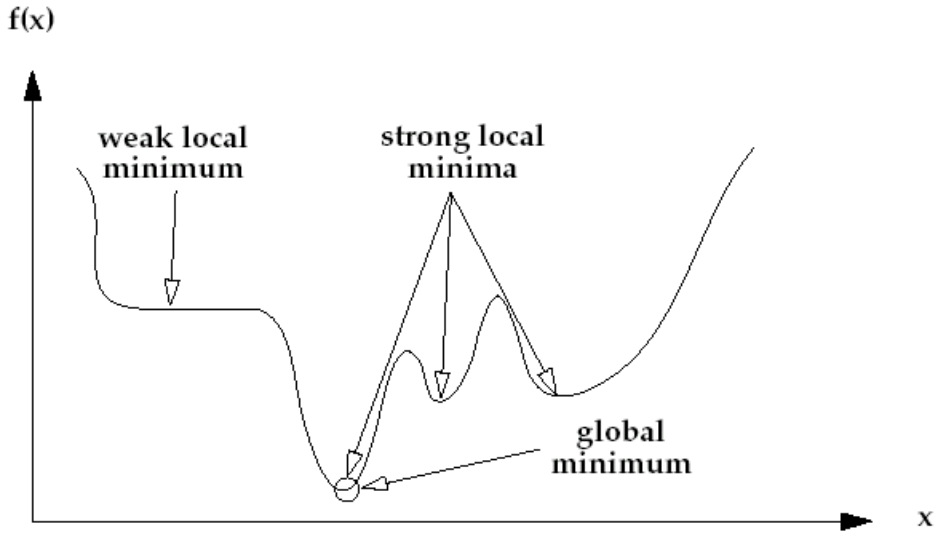



Fig. 3.1. - Types of minima for unconstrained optimization problems.

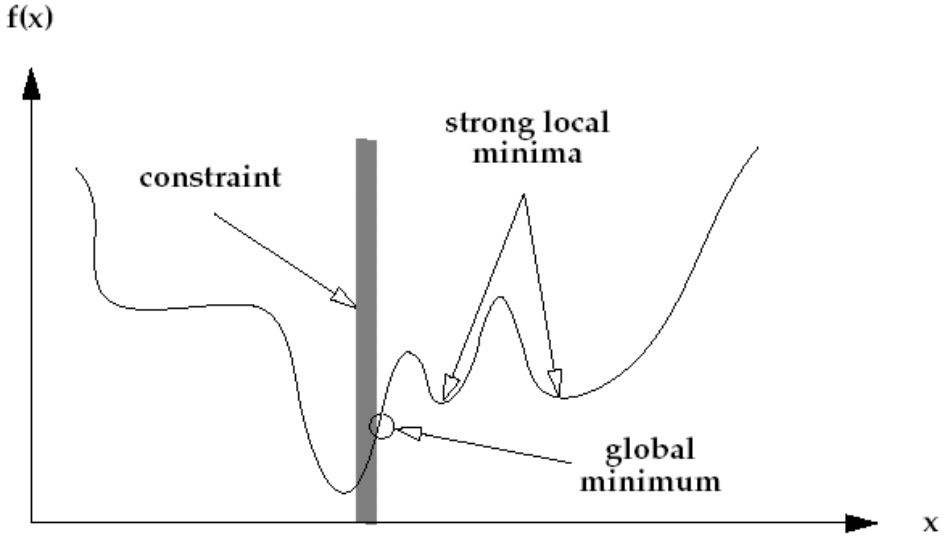



Fig. 3.2. - Types of minima for constrained optimization problems.

Real-world optimization problems usually require to reach many goals. In fact, in industrial applications, optimized design is often problematic because of the simultaneous occurrence of many conflicting targets. There are different methods to solve this kind of problem, such as optimizing a single multi-objective function obtained by a weighted sum of targets or finding multiple Pareto-optimal solutions. Whichever method one adopts, real-world optimization problems usually exhibit multiple optima: i.e. it is necessary to find several optima of a single multimodal function or a set of Pareto optimal solutions.
In the former case, deterministic methods do not perform well, while standard stochastic methods tend to find the best single global optimum. To explore a multimodal function correctly, the evolutionary algorithms (EAs) must maintain population diversity: several methods have been developed to adapt standard EAs for multimodal function optimization, such as the Niching Genetic Algorithm (NGA. Better results could be obtained by performing the recognition of subpopulations by means of NGAs, and then identifying the best point in every niche by means of a deterministic zero-order method. In practice this approach is nontrivial: the identification of niches and the attribution of each individual (a point in the

search space) to its niche is very hard due to the unknown behaviour of the objective function. In order to give good results, they require a crucial *a priori* specification of a dissimilarity measure, corresponding to the "niche radius".
In the latter case, the multiobjective optimization problems are solved by finding a set of solutions, generally denoted as Pareto-optimal, that can be considered equivalent in the absence of information concerning the relevance of each objective relative to the others. Often the search space can be too large and too complex to be solved by exact methods, thus efficient optimization strategies are required that are able to deal with both difficulties. Evolutionary algorithms possess several characteristics that are desirable for this kind of problem, in fact multiple individuals can search for multiple solution in parallel.
Unfortunately, both methods can be extremely expensive, because the exploitation of different solutions requires a great number of evaluations of the objective function. This is especially harmful in electromagnetic problems when each estimation of the objective function calls for a numerical solution by means of the Finite Element Method (FEM).

## 3.3. Genetic algorithms (GAs)

GAs, originally developed by Holland, are general purpose stochastic search strategies, based on the metaphor of natural evolution [3], [4].
The design variables are coded into fixed-length strings, called chromosomes (or genotypes); each position in the string is called a gene and the possible values of each gene are called alleles. In standard GAs, chromosomes are bit strings, obtained with or without Gray encoding [5].
A population (or generation) contains a finite number P of chromosomes. Each chromosome has an assigned fitness that measures its ability to survive and produce offspring: fitness is calculated by means of an objective function that plays the role of the environment in which the population evolves. In the optimization of electromagnetic devices with FEM, the fitness estimate involves a full FEM analysis.
The initial population is randomly created. A new population is generated by means of three main operators: selection, crossover and mutation. The selection operator stochastically chooses chromosomes to become parents and reproduce according to their fitness. The crossover operator exchanges a portion of the binary representations between two parents, randomly chosen, in order to generate new child strings. The mutation operator randomly flips a bit in an individual's bit string representation. Crossover and mutation are applied with probabilities $P_c$ and $P_m$, respectively, with $P_m$ less than $P_c$. Both operators introduce changes into new chromosomes, which replace existing ones in the new population. The chromosomes that survive will be those which have proven to be most fit. The algorithm stops when the optimum is found or the maximum number of generations ($N_g$) is reached.
To use GAs in optimization, it is essential to settle the configuration of many parameters. chromosome length ($N_c$), population size (P) and number of generations ($N_g$) are heuristically determined and are strictly dependent on the optimization problem. Selection, crossover and mutation are also dependent on the problem, yet less rigidly: the way in which each operator is implemented and their rate of application influence the evolution of the optimization and its chance of success. GA parameters are usually selected following heuristic criteria; a typical configuration is suggested by Carroll in [6]:

*Representation*: binary (not Gray encoding)

*Selection*: tournament selection with elitism

*Crossover type*: uniform crossover

*Crossover probability*: $P_c = 0.5$

*Mutation probability*: $P_m = 1/N_c$.

Searching for a global optimum of the objective function often involves a tradeoff between two apparently conflicting items: the judicious and robust sampling of the design variables (exploration) and the improvement of a good solution to reach valley bottom (exploitation). Our work aimed to find a choice of the GA parameters that generally leads to good results for any kind of optimization problem.
Starting from Carroll's configuration, binary representation without Gray encoding and the well-established "tournament selection" [7], [8] with elitism were maintained. Two crossover schemes were examined: two-point crossover and uniform crossover. The first is implemented by choosing two crossover points at random, whereas in the second, for each parent chromosome gene, a random number is generated and if it is less than the crossover probability $P_c$, this position becomes a crossover point.

Furthermore, sharing the ideas put forward in [9] and [10], variable crossover and mutation rates were implemented, in order to improve the GA search by assuring a good exploration at the beginning of evolution, and more and more exploitation capability while optimization goes on. Simple linear variations were used. Several optimizations of some mathematical functions were performed to tune the starting and ending values of $P_c$ and $P_m$. The following GA parameter configuration is therefore proposed [11]:

*Representation*: binary (not Gray encoding)

*Selection*: tournament selection with elitism

*Crossover type*: two-point crossover

*Crossover probability at generation k*: $P_c^{(k)} = 0.3 + 0.4\frac{k-1}{N_g - 1}$

*Mutation probability at generation k*: $P_m^{(k)} = 0.05 - 0.04\frac{k-1}{N_g - 1}$

The performance of this parameter configuration was measured using a set of four De Jong's mathematical functions which are typically used for GA benchmarking [12]. Function f1 is a unimodal quadratic function in three dimensions, which has only one minimum given by zero. Function f2 is the classical two-dimensional Rosenbrock's function, with a local minimum and a global one, given by zero. Function f3 is a discontinuous step function in five dimensions, which exhibits a global minimum given by -25. Function f5 is the 'foxhole' function in two dimensions with a global minimum of 0.998 and 24 local minima.
Having fixed the population size and the number of generations to P=100 and $N_g$=50, respectively, the best parameter configuration for each objective function was found by considering two kinds of crossover (two-point and uniform crossover), 9 crossover probability values (from 0.1 to 0.9) and 9 mutation probability values (from 0.01 to 0.09). 50 optimizations were performed for each parameter configuration, giving a total of 8100 optimizations. Table 1 shows the best configuration for each De Jong's function.
The proposed GA parameter configuration was compared with the Carroll and best ones. Table II shows this comparison. Three performance aspects were considered: number of successes (NS), accuracy of optimum (BO) and convergence speed (CS) [13]. NS is simply the number of times that the optimum was found (with a tolerance of $10^{-3}$) and verifies the ability of the GAs to reach the optimum solution. BO is the optimum value found by the GAs. CS is the average number of generations required to find the optimum. The average optimum (AO) value and worse optimum (WO) (out of the 50 optimizations performed) also provide

TABLE 1
BEST PARAMETER CONFIGURATION FOR SOME DE JONG'S FUNCTIONS

| | f1 | f2 | f3 | f5 |
|---|---|---|---|---|
| crossover type | uniform | 2-point | uniform | 2-point |
| crossover probability | 0.1 | 0.4 | 0.5 | 0.9 |
| mutation probability | 0.01 | 0.05 | 0.01 | 0.05 |

meaningful information. For functions f1 and f3 the GAs invariably find the optimum with every parameter configuration (Table 2). For multi-minima functions f2 and f5, the proposed GA parameter configuration exhibits performance which is always in between that of the Carroll and best configurations, closer to the best one with regard to NS and AO; CS is also very similar. Note that, if the GA parameters are not suitably tuned for the specific function, the percentage of success may be below 50%, as shown in Table 2 for the Carroll configuration.

TABLE 2
COMPARING GA PARAMETER CONFIGURATIONS

| | | NS | BO | WO | AO | CS |
|---|---|---|---|---|---|---|
| f1 | Carroll | 50 | 0.0 | $3.0\ 10^{-4}$ | $1.4\ 10^{-4}$ | 25 |
| | universal | 50 | 0.0 | $2.1\ 10^{-3}$ | $2.1\ 10^{-4}$ | 34 |
| | best | 50 | 0.0 | $3.0\ 10^{-4}$ | $1.4\ 10^{-4}$ | 25 |
| f2 | Carroll | 22 | 0.0 | 0.106 | $6.769\ 10^{-3}$ | 22 |
| | universal | 32 | 0.0 | 0.082 | $3.708\ 10^{-3}$ | 25 |
| | best | 38 | 0.0 | 0.016 | $1.075\ 10^{-3}$ | 26 |
| f3 | Carroll | 50 | –25.0 | –25.0 | –25.0 | 8 |
| | universal | 50 | –25.0 | –25.0 | –25.0 | 8 |
| | best | 50 | –25.0 | –25.0 | –25.0 | 8 |
| f5 | Carroll | 23 | 0.998 | 4.172 | 1.138 | 14 |
| | universal | 41 | 0.998 | 1.421 | 1.044 | 16 |
| | best | 44 | 0.998 | 1.207 | 1.023 | 18 |

## 3.4. Particle swarm optimization (PSO)

Particle Swarm Optimization (PSO) is a relatively new family of algorithms which can be used to find optimal (or near optimal) solutions to numerical and combinatorial problems. It is easily implemented (the core of the algorithm can be written in a few lines of code) and has proven both very effective and quick when applied to a diverse set of optimization problems.
PSO was originally developed by Kennedy and Eberhart in 1995 [14], taking inspiration both from the related field of evolutionary algorithms and in artificial life methodologies.
Animal social behavior, such as that seen in flocks, schools, or herds, has always attracted many researchers, interested in discovering the underlying rules which enable, as an example, large numbers of birds to flock synchronously, often scattering and regrouping, and suddenly changing direction. Many models of this flocking behavior were also used to create computer simulations, such as those by Heppner and Grenander [15], and by Reynolds [16].
A general way to search for a solution of a problem is to determine an objective, or cost function which describes the problem and to optimize it. Thus the field of function optimization is of wide interest. Optimizing a function f(x) means either minimizing (or maximizing) it.
From an optimization point of view, it is straightforward to see the particles' flight as a trajectory in the solution space. In this way the PSO algorithm can be used to minimize a generic function in and $N_d$-dimensional space.
The standard algorithm of the PSO is the following one.

1: **procedure** PSO
2: Initialize particles with random positions and velocities.
3: Set particles' *pbests* to their current positions.
4: Calculate particles' fitness and set *gbest*.
5: **for** T generations **do**
6: Update particles' velocities.
7: Update particles' positions.
8: Recalculate particles' fitness.
9: Update particles' *pbest* and *gbest*.
10: **end for**
11: **end procedure**

Consider a search space of d dimensions. Then $\mathbf{x}_i=(x_{i1},\ldots, x_{id})$ denotes the position of the i-th particle of the swarm (i=1,…,N), and $\mathbf{p}_i=(p_{i1},\ldots,p_{id})$ denotes the best position it has ever visited. The index of the best particle in the population (the one which has visited the global best position) is represented by the symbol g. At each time step t in the simulation, the velocity $\mathbf{v}_i=(v_{i1},\ldots,v_{id})$ of the i-th particle, is adjusted along each axis j according to the following equation:

$$v_{ij}(t+1) = v_{ij}(t+1) + \xi_p\left(p_{ij}(t) - x_{ij}(t)\right) + \xi_g\left(p_{gj}(t) - x_{ij}(t)\right) \tag{3.4.1}$$

where $\xi_p$ and $\xi_g$ are random numbers uniformly distributed in $[0, p_{incr}]$ and $[0, g_{incr}]$, respectively, $p_{incr}$ and $g_{incr}$ being the cognitive and social acceleration coefficients. Moreover, the velocity of the particle can be constricted to stay in a fixed range:

$$-V_{max} \le v_{ij}(t+1) \le V_{max} \tag{3.4.2}$$

In this way the likelihood of particles leaving the search space is reduced, although indirectly, by limiting the maximum distance a particle will cover in a single step, instead of restricting the values of $\mathbf{x}_i$. The new position of a particle is calculated using:

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \mathbf{v}_i(t+1) \tag{3.4.2}$$

The personal best position of each particle is updated using:

$$\mathbf{p}_i(t+1) = \begin{cases} \mathbf{p}_i(t) & \text{if } f(\mathbf{x}_i(t+1)) \geq f(\mathbf{p}_i(t)) \\ \mathbf{x}_i(t+1) & \text{if } f(\mathbf{x}_i(t+1)) < f(\mathbf{p}_i(t)) \end{cases} \tag{3.4.3}$$

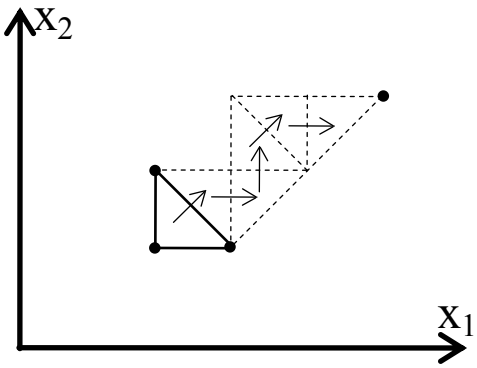
while the global best index is defined as:

$$g = \arg\min f(\mathbf{p}_i(t+1)) \tag{3.4.4}$$

An essential feature of the PSO algorithm is the way in which the local and global best positions, $\mathbf{p}_i$ and $\mathbf{p}_g$, and their respective acceleration coefficients, are involved in velocity updates. Conceptually, $\mathbf{p}_i$ (also known as $\mathbf{p}_{best}$) resembles the particle's autobiographical memory, i.e. its own previous experience, and the velocity adjustment associated with it is a kind of simple nostalgia, as it leads the particle to return in the position where it obtained its best evaluation. On the other hand, $\mathbf{p}_g$ ($g_{best}$) is a sort of group knowledge, a common standard which every single particle seeks to attain.
The overall effect is such that when particles find a good position, they begin to look nearby for even better solutions, but, on the other hand, they continue to explore a wider area, likely avoiding premature convergence on local optima and realizing a good balance between exploration of the whole search space and exploitation of known good areas [17].

## 3.5. Pattern search (PS).

One of the best known deterministic algorithms is the Pattern Search (PS). It was conceived by Hooke and Jeeves in 1961 [18] and is part of the so-called zero-order methods, for which you need to calculate the value of the single objective function. Key features of this method are ease of implementation and speed of convergence.
The parameters of the algorithm are: the starting point, the step with which it must move to the next solution and the minimum step with which move from one point to another (tolerance and stopping criterion). Launched the execution, the calculation program performs a search based on points arranged at the vertices of a simplex. The first simplex is built around the starting point indicated during setting and by the values of the objective function at the vertices of the simplex are calculated. Subsequently, the simplex is reflected along one of its faces (in agreement with the step provided in the initial setting), the objective function is evaluated in the vertex of the new element, and this value is compared with the previous one. The algorithm proceeds by computing the function values for all elements adjacent to the starting simplex and finally it moves to the node that has the best value. This process is done until it finds a better solution than the current element. If this does not occur, the optimizer refines the search grid halving the step. After a series of iterations, if the optimizer takes less displacement compared to the tolerance, it is assumed that convergence has been reached.




.

Fig. 0.3. – Searching for the optimum in a function of two variables.

Fig. 2.2 shows a possible path of the algorithm that switch from one configuration to the next. Being a deterministic algorithm, this method may not converge to the optimum if the objective function has its several extreme points. In cases where the change of the objective function is unknown, the choice of the PS as the optimizer is not very happy. However, because of the simplicity of the method, it is not advisable to exclude a priori the use of this procedure. A frequently used approach is to repeat the simulation several times using different starting points. If, during the various tests, the PS converges to the same point, it can be concluded that the objective function has only one minimum, and it is found by the algorithm. Otherwise, it is not guaranteed that the optimum is between the points found by the PS during the various simulations and it is advisable the use of another optimization algorithm.

## References Chapter 3


[1] S. Russenschuck, "Synthesis, inverse problems and optimization in computational electromagnetics," *Int. J. Numer. Modelling: Electronic Networks, Devices and Fields*, vol. 9, pp. 45-57, January-April 1996.

[2] P. G. Alotto, et alii, "Stochastic algorithms in electromagnetic optimization," *IEEE Trans. Magn.*, vol. 34, pp. 3674-3684, Sept. 1998.

[3] J. H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, 1975.

[4] D. E. Goldberg, "*Genetic Algorithms in Search, Optimization & Machine Learning,*" Addison Wesley, 1989.

[5] Y. Yokose, V. Cingoski, K. Kaneda, and H. Yamashita, "Perfomance comparison between Gray coded and binary coded genetic algorithms for inverse shape optimization of magnetic devices," *Jap.-Bulg.-Maced. J. Seminar Appl. Electrom. Mech.*, Sofia, 14-15 Sept. 1998.

[6] D. L. Carroll, "D. L. Carroll's FORTRAN Genetic Algorithm Driver," *http://www.staff.uiuc.edu/~carroll/ga.html*, ver. 1.6.4, 1997.

[7] D. E. Goldberg, and K. Deb, "A comparative analysis of selection schemes used in genetic algorithms," *Foundations of Genetic Algorithms*, vol. 1, pp.69-93, 1991.

[8] D. Thierens, and D. E. Goldberg, "Convergence models of genetic algorithm selection schemes," in *Parallel Problem Solving from Nature-PPSN III,* Y. Davidor, H.P. Schwefel, and R. Manner (Eds.), Springer-Verlag, Berlin, pp. 119-129, 1994.

[9] L. B. Booker, "Improving search in genetic algorithms," in *Genetic Algorithms and Simulated Annealing*, L. Davis (Ed.), Pitman Publishing, London, pp.61-73, 1987.

[10] D. T. Pham, and D. Karaboga, "Some variable mutation rate strategies for genetic algorithms," *Intelligent System Laboratory*, University of Wales College of Cardiff, Cardiff, UK, personal comm., 1996.

[11] S. Alfonzetti, E. Dilettoso, N. Salerno, "A proposal for a universal parameter configuration for genetic algorithm optimization of electromagnetic devices," *IEEE Transactions on Magnetics*, vol. 37, no. 5, pp.3208-3211, 2001.

[12] K. A. De Jong, *An Analysis of the Behavior of a Class of Genetic Adaptive Systems*, Ph.D. Thesis, Dept. Comp. Commun. Sciences, University of Michigan, Ann Arbor, 1975.

[13] J.S. Chun, H.K. Jung, S. Y. Hahn, "A study on comparisons of optimi-zation performances between immune algorithm and other heuristic algorithms," *IEEE Trans. Magn.*, vol. 34, pp. 2972-2975, Sept. 1998.

[14] J. Kennedy and R. C. Eberhart, "Particle swarm optimization", *IEEE International Conference on Neural Networks,* Piscataway, Nov, 27- Dec. 1, 1995

[15] F. Heppner and U. Grenander, A stochastic nonlinear model for coordinated bird flocks, *The Ubiquity of Chaos,* S. Krusna (ed.), AAAS Publications, Washington, pp. 233–238, 1990.

[16] C. W. Reynolds, "Flocks, herds, and schools: A distributed behavioral model, in computer graphics", *SIGGRAPH '87 Conference*, pp. 25–34, 1987.

[17] E. Ozcan and C. Mohan, "Analysis of a simple particle swarm optimization system," *Intell. Eng. Syst. Through Artif. Neural Networks*, vol 8, pp. 253–258, 1998.

[18] R. Hooke and T. A. Jeeves, "Direct Search Solution of Numerical and Statistical Problems", *Journal of the ACM*, vol. 8, no 2, pp. 212-229, 1961.

# Chapter 4

# The Photovoltaic Conversion

## 4.1. Functioning principle of a photovoltaic cell

Photovoltaic systems convert solar energy into electricity. The term "photos" comes from the Greek "phos", which means light, and "Volt" is derived from Alessandro Volta (1745-1827), who was the first to study the electrolyte phenomenon. Commonly the term "PV" is used with the meaning of "solar cell". Photovoltaic systems can be simple energy-supplying systems for small calculators and wristwatches, or more advanced systems which can provide electricity for the operation of hydraulic pumps, communication systems, lighting systems for homes and many other applications. In most cases the supply of electricity through photovoltaic systems is the most economical solution.
The photoelectric effect, that is the ability of certain materials to convert solar energy into electrical energy, is known since 1839, thanks to the experience gained by the French physicist Edmond Becquerel (1820-1891). He presented to the Academy of Sciences of Paris his "notes on the electrical effects under the influence of sunlight." This discovery occurred randomly, while performing some experiments on an electrolytic cell in which two platinum electrodes were immersed.
The first commercial silicon solar cell was built at Bell Labs in 1954 (Person, Fuller and Chapin). At the end of the 50s, due to the high cost of this new technology, the development of photovoltaic devices was mainly due to the research in the field of space programs, for which it was necessary to have a reliable and inexhaustible source of electricity. Currently we are seeing a rapid spread of photovoltaic technology for terrestrial applications, such as for isolated user powering and systems installed on buildings and connected to a pre-existing network.
The fundamental physical phenomenon on which is based the operation of a photovoltaic device is the photoelectric effect: it is characterized by the emission of electrons from the surface of a conductor or semiconductor material, when it is struck by an electromagnetic radiation (for example the light). In this way it is possible to convert the energy of the solar radiation into electrical energy in direct current.
The basic component of a photovoltaic system is the photovoltaic cell. A standard photovoltaic cell, generally a 125 × 125 mm square with a thickness between 0.25 to 0.35 mm, is usually capable of producing about 1.5 W of power under normal conditions, ie when it is at a temperature of 25 ° C and is subjected to a power density of the radiation of 1000 W/m². The output power from a photovoltaic device under normal conditions is called peak power (WP) and is used as a reference.
More cells assembled and connected between them in a single structure form the photovoltaic module. It is constituted by the series connection of 36 cells, and delivers an output power of 50 W, approximately. Currently, especially for architectural demands, manufacturers put on the market modules consisting of a much higher number of cells so that the output power can reach up to 200 W. According to the voltage required to supply of the electrical devices, multiple modules can be connected in series in a "string". The electric power required

determines the number of strings to be connected in parallel to achieve a photovoltaic generator.
The transfer of energy from the photovoltaic system to the users occurs through additional devices, necessary to transform and adapt the direct current produced by the modules to the needs of end users. The set of such devices is called BOS (Balance Of System). An essential component of BOS is the inverter, a device that converts the direct current output from the PV array into alternating current. The conversion of solar radiation into electric current takes place in the photovoltaic cell, a device consisting of a thin wafer of semiconductor material suitably treated. The material most frequently used for the construction of such devices is the monocrystalline silicon, polycrystalline and amorphous.

## 4.2. The silicon structure

The Silicon (Si) belongs to the group IV of the periodic table of the elements and is a semiconductor material. The silicon atom has the first two orbitals filled and the outer orbital contains 4 electrons of the 8 ones needed to fill the orbital. These electrons are called valence electrons and may participate to interactions with other atoms. In a crystalline structure, the silicon atom forms four valence bonds with neighbouring atoms, thus completing the outer orbital (see Fig. 4.1).
The difference of the potential energy between the valence band (VB) and the conduction band (CB) of the electrons in a material (insulator, semiconductor or conductor) is called "Energy Gap" (EG) and its value is an intrinsic property of the material. This value represents the minimum amount of energy that must be supplied to the electron so that it can move from VB to CB. When an electron makes this passage, it leaves an empty orbital in VB. The electrical behavior of such empty orbital can be mathematically described as that of a particle of charge equal and opposite (ie positive) to that of an electron. Such a particle takes the name of hole.

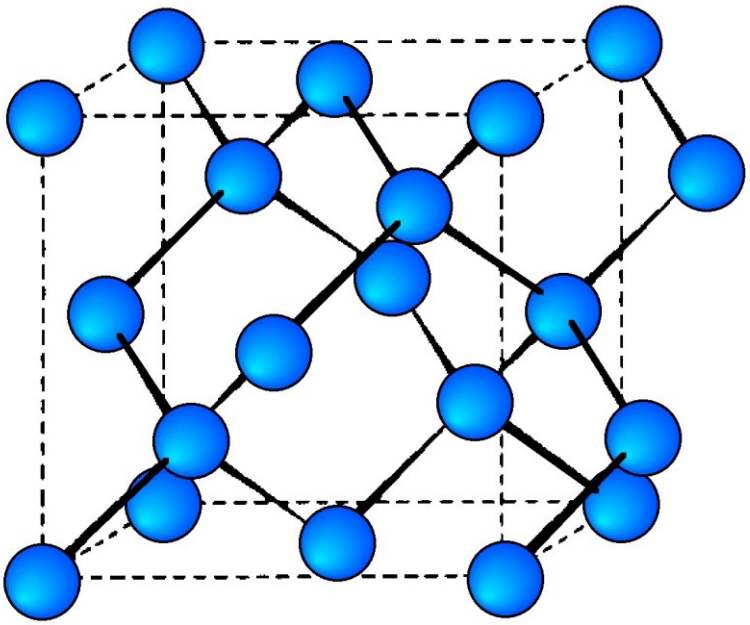


Fig. 4.1. – Crystal structure of silicon.

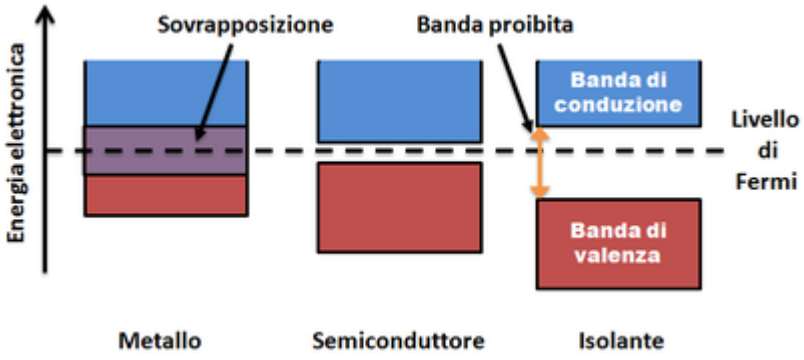



Fig. 4.2. – Energy gaps for metals, semiconductors and insulators.

In the case of an insulating material the EG is very high, making the probability of passage of the electron from the VB to BC very close to zero. For a semiconductor material, the energy required has a lower value than the previous case and the probability of passage of electrons is always finite and different from zero (at a temperature different from absolute zero). In metals the two bands are overlapped and the electrons can move easily from one energy level to another making the material a good conductor. In the case of silicon, the value of EG is equal to about 1.12 eV (electron volts). Figure 4.2 shows the three cases described above. The energy required to overcome the band gap can be provided to electrons by thermal excitation or by absorption of photons of appropriate energy.
Unlike what happens in metals, in the semiconductors the motion of the charges is not only due to the applied electric field, but also presents a so-called diffusion current. It is determined by the motion of electric charges generated by a gradient of concentration of electrons and holes. The expression of the current in a semiconductor can therefore be written analytically in the following way:

$$I = I_n + I_p = qA\mu_n nE + qA\mu_p pE + qAD_n \frac{dn}{dx} - qAD_p \frac{dp}{dx} \tag{4.2.1}$$

where:

- A is the cross section of the semiconductor;
- q is the electron charge ($q=1.602\cdot10^{-19}$ C);
- n is electron density ($m^{-3}$);
- p is hole density ($m^{-3}$);
- E is the electric field (V/m);
- $\mu_n$ is the electron mobility;
- $\mu_p$ is the hole mobility;
- $D_n$ is the electron diffusion constant ($m^2/Vs$);
- $D_p$ is the hole diffusion constant ($m^2/Vs$).

When a luminous flux invests the silicon crystal lattice, a number of electrons is excited and passes in CB, thus creating an equal number of holes. The process described takes the name of generation of electron-hole pairs. The process of recombination occurs when an electron occupies a hole, returning a part of the energy possessed in the form of heat. To exploit the electricity it is necessary to create a coherent motion of electrons (and holes), or a current, by means of an electric field internal to the cell. This field is obtained by putting in contact two semiconductor materials with charge excesses of opposite sign.

## 4.3. Semiconductor doping

The silicon crystals may be treated by means of physical or chemical processes, by inserting inside the crystalline structure some impurities, that is atoms of other elements. These treatments are called doping. Some silicon atoms are replaced with atoms of the group V of the table of the elements (typically phosphorus, P), said donors, or with atoms of group III (typically boron, B), said acceptors.
In the first case an electron is introduced in the outer orbital, so that this exhibits an electron in excess with respect the number needed to complete the same orbital. This electron is weakly bound (fraction of eV) and therefore requires a modest energy to jump in CB; materials whose conductivity is mainly due to negative charges are called n-type. In the second case, instead, the concentration of holes increases; such materials are called p-type.

## 4.4. The p-n junction

A photovoltaic cell is constituted by the coupling of a p-type doped semiconductor and an n-type one (p-n junction). Through the contact surface of the two semiconductors, some electrons pass from the n-type material to the p-type one, while some holes moving in the opposite direction. The n-type material thus acquires a weak positive charge, while that of the p-type becomes slightly negative. At the interface between the two materials, therefore, an electric field is generated, directed from the n-type material to the p-type one, to which a potential difference $V_e$ is associated. This layer, called the depletion region, prevents any further spread of the charge carriers in both directions.
If the junction is hit by a light radiation some electron-hole pairs can be generated. This process occurs if the energy possessed by the photon E = hf (where h is Planck's constant and f is the frequency of the photon) is higher than the EG. In this case, due to the electric field present therein, the electron is pushed towards the n-type material and the hole toward the p-type material, generating an accumulation of charge carriers in the two doped zones. If the two materials are connected by a conducting wire, the equilibrium is re-established by means of a flow of electrons from the p-type semiconductor to p-type one. The absorption of light radiation causes in this way a continuous electrical current in the wire (Fig. 4.3).

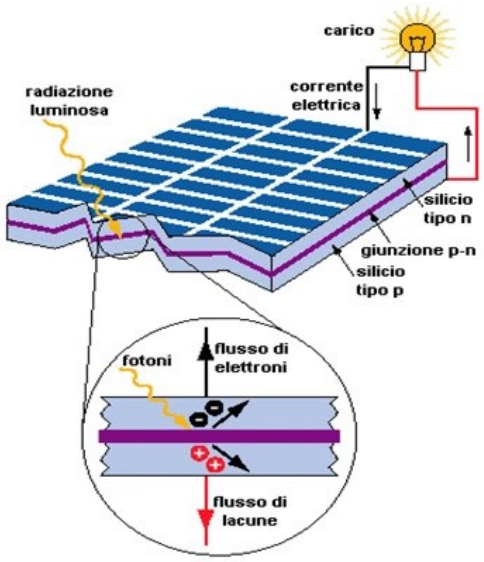



Fig. 4.3 – A p-n junction radiated by the light.

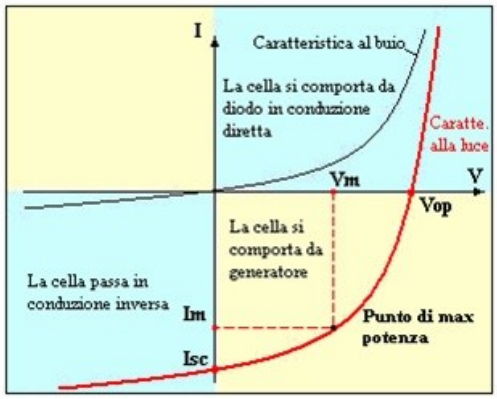



Fig. 4.4. - Characteristic tension-current of a PV cell.

## 4.5. Electrical characterization of a photovoltaic cell

A photovoltaic cell exposed to solar radiation behaves as a current generator. Its operation can be described by the voltage-current characteristic curve, as shown in fig 4.4.
If subjected to an external voltage V, the behavior of a photovoltaic cell is similar to that of a semiconductor diode: if $V<V_e$, there is no passage of current; if V tends to $V_e$, the device becomes a good conductor. If one changes the sign of the voltage, an extremely modest current flows. In the case of excitation by voltage V, the current through the cell is that of a diode in direct conduction:

$$I_D = I_0\left(e^{qV/NKT} - 1\right) \tag{4.5.1}$$

where:
- q is the electron charge;
- K is the Boltzmann constant ($1.38\cdot10^{-23}$ J/°K);
- T i the absolute temperature (°K);
- $I_0$ is a constant which depends on the semiconductors characteristics;
- N is a coefficient between 1 and 2 which depends on the generation and recombination processes in the spatial charge region (for an ideal diode N=1).

The analytical expression of $I_0$ is:

$$I_0 = A_0 T^3 e^{-E_G/KT} \tag{4.5.2}$$

where $A_0$ is a constant which depends on the semiconductor utilized.
When a cell is radiated by photons of frequency $f>E_G/h$, the p-n junction becomes a source of pairs electron-hole.
In open circuit operations, the voltage across the cell reaches a maximum value $V_0$, while the current of the device is zero; In short-circuit operations, the current is maximum and is called $I_{cc}$. In the presence of an external load, the current $I_{cc}$ decreases by an amount equal to $I_D$ in the opposite direction to that generated by the photovoltaic process. In this case, in fact, the cell behaves like a diode to which a voltage is applied.
If we choose by convention that the photocurrent is positive, the current $I_D$ is negative. The equivalent circuit of the cell is shown in Fig. 4.5.

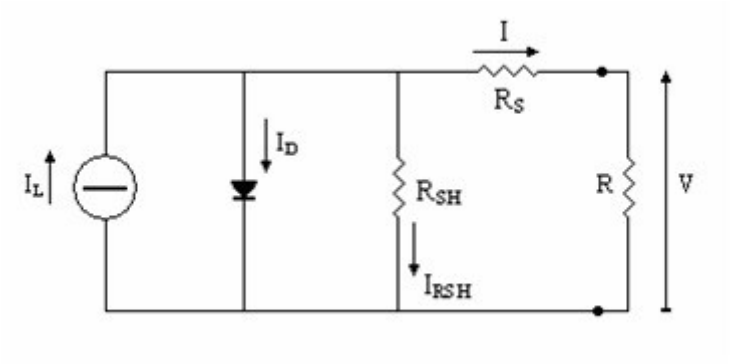



Fig. 4.5. - Circuit model of a photovoltaic cell.

The current $I_L$ is the generated by the light, the intensity of which is proportional to the number of photons with frequency f> $E_G$/h; the current $I_D$ is the one that passes through the junction of the cell, while the current I is flowing on the external load, that is the current which we need to know for practical purposes.
The $R_S$ is the parasitic resistance of the cell and includes the resistance of the two layers of material that form the cell and the resistance of the contacts. The resistance $R_{SH}$, said shunt resistance, represents those losses due to leakage currents occurring within the cell.
The characteristic equation of the cell illuminated thus becomes:

$$I = I_L - I_D - I_{R_{SH}} = I_L - I_0\left(e^{q(V+R_S I)/NKT} - 1\right) - \frac{V + R_S I}{R_{SH}} \qquad (4.5.3)$$

By multiplying (4.5.3) to the voltage, it is possible to derive the power generated, whose graph is shown in Fig. 4.6 (dashed curve in red).
If, as often happens, $R_S$ and $G_{SH}$ (= 1 / $R_{SH}$) are negligible, V coincides with the difference of potential $V_R$ that the cell transmits to the load.
If V vanishes, the current $I_{CC}$ attains the maximum current value that the cell can deliver for a given illumination, and is given by the following equation:

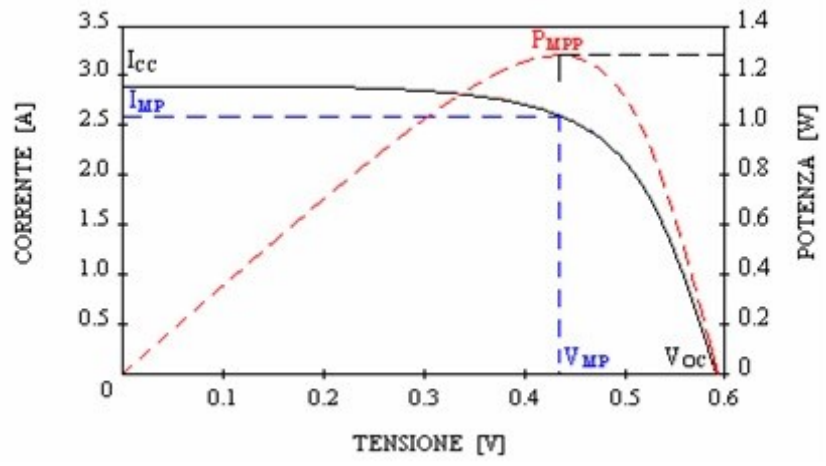



Fig. 4.6. – Power generated

$$I_{cc}(t) = I_L(t) - I_0(t)\left(e^{qR_S I_{cc}(t)/NKT(t)} - 1\right) - \frac{R_S I_{cc}(t)}{R_{SH}} \tag{4.5.4}$$

In normal conditions $R_S << R_{SH}$, then the third term of equation (4.5.4) can be neglected; moreover, since the exponent of the exponential much less than one, it can be approximated by the Taylor series stopped at the first order as $e^x \approx 1 + x$. Then, equation (4.5.4) can be solved for $I_{cc}$, by giving:

$$I_{cc}(t) \cong \frac{NKT(t)}{NKT(t) + qR_S I_0} I_L(t) \tag{4.5.6}$$

From (4.5.6) it is possible to obtain $I_L$:

$$I_L(t) \cong \left(1 + \frac{qR_S I_0}{NKT(t)}\right) I_{cc}(t) \tag{4.5.7}$$

The second term in parentheses is negligible because $R_S$ is small and the current $I_0$ has a value of $\approx 1.5 \cdot 10^{-10}$ A at a temperature of 300 °K and with an $E_G$ equal to 1.1 eV; then we have:

$$I_L(t) \cong I_{cc}(t) \tag{4.5.8}$$

We can therefore say that the short-circuit current $I_{cc}$ is proportional to the irradiation.
The potential difference that occurs at the ends of the photovoltaic cell, in the open circuit condition, is indicated with $V_O$. Its analytical expression can be derived from that of the current I, placing it equal to zero and neglecting the resistance $R_S$ and $R_{SH}$. We have:

$$V_0 = NV_T \ln \frac{I_L + I_0}{I_0} \tag{4.5.9}$$

The main variables that affect the characteristic of a photovoltaic cell are three: the intensity of the solar radiation, the temperature and the cell area. The intensity of the short-circuit current, as already stated, varies proportionally to the intensity of the radiation.
On the contrary, the intensity of the solar radiation does not have a significant effect on the value of the open-circuit voltage $V_0$; for this reason $V_0$ attains values close to the maximum even at low values of the solar radiation. The open circuit voltage between the cases of maximum and minimum value of radiation varies between 0.50-0.60 V (see fig 4.7).
The only way to avoid the presence of voltage at the terminals of a photovoltaic generator consists in the total obscuring of the capture surface.
As the temperature of the cell rises, the open-circuit voltage $V_0$ decreases of about 2.3 mV/°C and, jointly, the short-circuit current $I_{cc}$ increases of about 0.2%/°C, as shown in Fig .4.8.

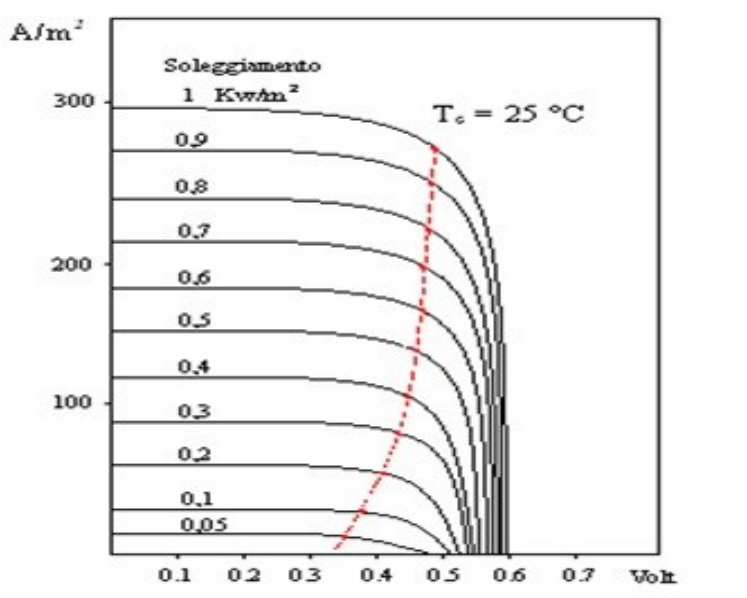



Fig. 4.7. – Characteristic curves for several values of the incident radiation.

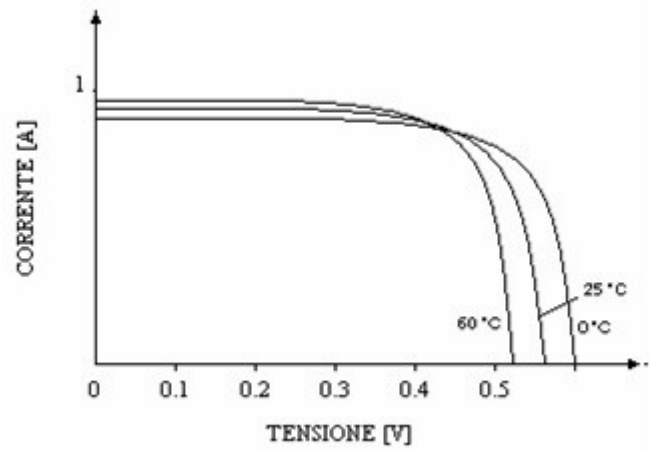



Fig. 4.8. – Characteristic curves for several values of the temperature.

## 4.6. Efficiency of a photovoltaic cell

Performance or efficiency of a PV is the module the ratio, expressed in percentage, between the captured and transformed energy, with respect to the total incident one on the surface of the module. It is thus a parameter of the quality or performance of the module itself. As in all energy conversion systems, the efficiency of the photovoltaic module is always less than unity (or 100%) due to unavoidable losses in the real systems. The main reasons for the losses are listed below:

- inefficiency of penetration of photons inside the cell: not all photons that radiate the cell penetrate inside it, given that in part they are reflected from the cell surface and in part they strike the metal grid of the contacts;
- inefficiency of the conversion of the photon energy into energy of the not electron-hole pairs: in order to break the bond between electron and nucleus a well determined amount of energy is required and not all the incident photons carry enough energy;
- inefficiency of conversion of the energy of the electron-hole pairs into electrical energy: not all the electron-hole pairs generated are collected by the electric field of the junction

and are sent to the external load, since in the path from the generation point to the junction they can meet charges of opposite sign, and then recombine;

- inefficiency due to the presence of parasitic resistances: the charges generated and collected in the depletion zone must be sent outside; the harvesting operation is accomplished by the metal contacts, placed on the front and back of the cell; even if during the manufacture an alloy process is performed between the silicon and the aluminum of the contacts, a certain resistance at the interface remains, which causes a dissipation and hence a reduction of the power transferred to the load.

In the case of polycrystalline silicon cells, the efficiency is further decreased due to the resistance that the electrons meet at the boundary between a crystal and another, and even more in the case of amorphous silicon cells, due to the random orientation of individual atoms.

## References Chapter 4


[1] A. W. Blakers and M. A. Green. "20% efficiency silicon solar cells," *Applied Physics Letters*, vol. 48, no. 3, pp. 215-217, 1986.

[2] M. A. Green, "Silicon solar cells: evolution, high-efficiency design and efficiency enhancements", *Semicond. Sci. Techno.*, vol. 8, pp.1-12, 1993.

[3] A. W. Blakers and M. A. Green, "20% efficiency silicon solar cells", *Applied Physics Letters*, vol. 48, no. 3, pp. 215-217, 1986.

[4] M. A. Green. "Silicon solar cells: evolution, high-efficiency design and efficiency enhancements", *Semicond. Sci. Techno*., vol. 8, pp. 1-12, 1993.

[5] P. G. Borden, R Walsh, and R. D. Nasby, "The V-groove silicon solar cell", *16th Photovoltaic Specialists Conference*, vol. 1, pages 574-577, 1982.

[6] J. Nelson, *The Physics of Solar Cells*, Imperial College Press, London, UK, 2003.

[7] M. A. Green, *Solar cells: operating principles, technology, and system applications*, Prentice-Hall, 1982.

[8] P Faine, S. R Kurtz, C Riordan, and J. M. Olson, "The influence of spectral solar irradiance variations on the performance of selected single-junction and multijunction solar cells". *Solar Cells*, vol. 31, no. 3, pp. 259-278, 1991.

[9] S. Sze, *Physics of Semiconductor Devices*, 2nd ed., John Wiley and Sons, New York, 1981.

[10] W. Shockley and H- J. Queisser, "Detailed Balance Limit of Efficiency of p-n Junction Solar Cells", *Journal of Applied Physics*, vol. 32, no. 3, pp. 510-519, 1961.

[11] M. A. Green, *Silicon solar cells: advanced principles and practice*", Bridge printery, Sydney, 1995.

[12] M. A. Green, J. Zhao, A. Wang, and S. R. Wenham, "Very high efficiency silicon solar cells-science and technology". *IEEE Transactions on Electron Devices*, vol. 46, no. 10, pp. 1940-1947, 1999.

# Chapter 5

# Plasmonic Resonances

## 5.1. Response models of the metals

The behavior of the materials is strongly connected to the frequency of the radiation that passes through them. An example is given by the metals that at lower frequencies of the visible spectrum, that is in the infrared and in the microwave bands, they act as reflective materials and can be used as coatings for waveguides and resonators. As the frequency increases, instead, the radiation penetrates the metal, which behaves like a dielectric material, resulting in a dissipation and an altering of the responses of the devices, according to the materials used. Indeed, we can consider the different response to the passage of the ultraviolet radiation through the sodium (alkali metal), which behaves as a transparent material, that is with a reduced dissipation, and the gold (noble metal), that, due to the band transitions, generates a high dissipation [1,2].
As it is well known, Maxwell's equations are a useful tool for the study of the macroscopic electromagnetic phenomena. In order to study the interaction with a material medium, we introduce a parameter, the electric permittivity ε, which in the frequency domain becomes a complex function, whose imaginary part represents the absorption of the medium. The permittivity, therefore, describes the connection between the radiation and the medium; it depends on the polarization of the medium to the passage of an electromagnetic wave, by generating a field which influences the radiation itself.
By subdividing the charge density ρ and the current density J in internal and external ones [3], in such a way that the external quantities condition the system, whereas the external quantities respond to the external excitations, we can write:

$$\nabla \cdot \vec{D} = \rho_{ext} \tag{5.1.1}$$

where $\vec{D}$ is the electric displacement (or electric induction) and $\rho_{ext}$ is the external charge density. The relation which links the electric field E to electric displacement D is:

$$\vec{D} = \varepsilon_0 \vec{E} + \vec{P} \tag{5.1.2}$$

where $\varepsilon_0$ is the permittivity (or dielectric constant) of the free space (or vacuum) and P is the polarization vector which describes the dipole moment per unit volume and is generated by the alignment of the microscopic dipoles in the medium.
Assuming that the material is isotropic, linear and non magnetic, we can write the following relations:

$$\vec{D} = \varepsilon_0 \varepsilon_r \vec{E} \tag{5.1.3}$$

$$\vec{P} = \varepsilon_0 \chi \vec{E} \tag{5.1.4}$$

where $\varepsilon_r$ is the relative permittivity and $\chi$ is the dielectric susceptibility, which are connected by the relation:

$$\varepsilon_r = 1 + \chi \tag{5.1.5}$$

Finally, we introduce the electrical conductivity $\sigma$ in the linear dependence of J from E:

$$\vec{J} = \sigma \vec{E} \tag{5.1.6}$$

Finally, turning from the time domain to the angular frequency domain, one derives the relationship between the permittivity and the electric conductivity:

$$\varepsilon_r = 1 + j\frac{\sigma}{\varepsilon_0 \omega} \tag{5.1.7}$$

Assuming also that the response is localized (then K = 0) the dielectric function we can consider depending only on ω: ε (ω); approximation valid as long as the wavelength of the radiation is at least an order of magnitude higher than the typical dimensions of the system, such as the free path average of the electrons or the lattice.
Recalling that the dielectric function is in general a function complex, we can write:

$$\varepsilon_r(\omega) = \varepsilon_1(\omega) + j\varepsilon_2(\omega) \tag{5.1.8}$$

The imaginary part is a measurement of the dissipation and, therefore, is related to the absorption of medium.
Consider now the Maxwell's equation in the absence of external sources:

$$\nabla \times \vec{E} = -\frac{\partial}{\partial t}\vec{B} \tag{5.1.9}$$

$$\nabla \times \vec{H} = \frac{\partial}{\partial t}\vec{D} \tag{5.1.10}$$

From these equations we derive:

$$\nabla \times \nabla \times \vec{E} = -\mu_0 \frac{\partial^2}{\partial t^2}\vec{D} \tag{5.1.11}$$

$$\vec{K}(\vec{K} \cdot \vec{E}) - K^2 \vec{E} = -\varepsilon_r \frac{\omega^2}{c^2}\vec{E} \tag{5.1.12}$$

where K is the wavevector. This equation allows us to highlight a particular aspect of the behavior of the electromagnetic radiation, which depends on the polarization direction of the electric field. If the wave is transverse the scalar product $\vec{K} \cdot \vec{E}$ is zero and we get:

$$K^2 = \varepsilon_r \frac{\omega^2}{c^2} \tag{5.1.13}$$

In the case of longitudinal waves we have:

$$\varepsilon_r = 0 \tag{5.1.14}$$

which is of fundamental importance in the response of metals at high frequencies, since it allows us to develop a model, called plasmon, which has interesting applications [1].

## 5.2. The Drude model

As already mentioned, the Drude model allows us to study, within certain limits, the optical response of the material media at the passage of electromagnetic waves. It is based on the assumption that in the medium there is a gas of free non-interacting electrons, which moves under the influence of the field generated by the radiation, and in the presence of a viscous friction, generated by collisions with fixed positive ions [4]. From the Newton's second law:

$$m\frac{d}{dt}\langle\vec{v}\rangle = q\vec{E} - \gamma\langle\vec{v}\rangle \tag{5.2.1}$$

where m is the effective mass, which take into account the effects which we have previously excluded before, $\langle\vec{v}\rangle$ is the average speed of the electrons and a $\gamma$ is coefficient equal to $m/\tau$, where $\tau$ is the relaxation time, that is, the time that elapses between two clashes.
Considering that the current density may be expressed as:

$$\vec{J} = nq\langle\vec{v}\rangle \tag{5.2.2}$$

where n is the density of electrons per unit volume, and that in stationary conditions from (5.2.1) the average speed is

$$\langle\vec{v}\rangle = \frac{q}{\gamma}\vec{E} = \frac{q\tau}{m}\vec{E} \tag{5.2.3}$$

we obtain

$$\vec{J} = n\frac{q^2\tau}{m}\vec{E} \tag{5.2.4}$$

which is the Ohm's law, the conductivity $\sigma$ being:

$$\sigma = n\frac{q^2\tau}{m} \tag{5.2.3}$$

Now we write the motion equation as:

$$m\ddot{\vec{x}} + m\gamma\dot{\vec{x}} = -e\vec{E} \tag{5.2.4}$$

in which e is the electron charge and $\gamma = 1/\tau$. If a time-harmonic behavior is assumed:

$$\vec{E}(t) = \vec{E}_0 e^{j\omega t} \qquad \vec{x}(t) = \vec{x}_0 e^{j\omega t} \tag{5.2.5}$$

The solution of the motion equation (5.2.4) is:

$$\vec{x}_0 = \frac{-e}{m(-\omega^2 + j\omega\gamma)}\vec{E}_0 \tag{5.2.6}$$

By considering that the polarization P is:

$$\vec{P}_0 = -ne\vec{x}_0 = \frac{ne^2}{m(-\omega^2 + j\omega\gamma)}\vec{E}_0 \tag{5.2.7}$$

we finally obtain:

$$\vec{D}_0 = \varepsilon_0\vec{E}_0 + \vec{P}_0 = \left(\varepsilon_0 + \frac{ne^2}{m(-\omega^2 + j\omega\gamma)}\right)\vec{E}_0 \tag{5.2.8}$$

which can be written as:

$$\vec{D}_0 = \varepsilon_0\left(1 + \frac{\omega_p^2}{-\omega^2 + j\omega\gamma}\right)\vec{E}_0 \tag{5.2.9}$$

where

$$\omega_p = \sqrt{\frac{ne^2}{m\varepsilon_0}} \tag{5.2.10}$$

We can derive the relative permittivity from (5.2.9):

$$\varepsilon_r(\omega) = 1 - \frac{\omega_p^2}{\omega^2 - j\omega\gamma} \tag{5.2.11}$$

which gives the following real and imaginary parts:

$$\varepsilon_1(\omega) = 1 - \frac{\omega_p^2\tau^2}{1 + \omega^2\tau^2} \tag{5.2.12}$$

$$\varepsilon_2(\omega) = -\frac{\omega_p^2\tau}{\omega(1 + \omega^2\tau^2)} \tag{5.2.13}$$

Consider the response of the metals within the limit of $\omega < \omega_p$, where for very low frequencies we have a strong absorption, the permittivity being predominantly imaginary ($\omega\tau << 1$), for intermediate frequencies we observe an almost total reflection ($1 \leq \omega\tau \leq \omega_p\tau$) and for frequencies to the limit we have a behavior of total transparency, the permittivity being

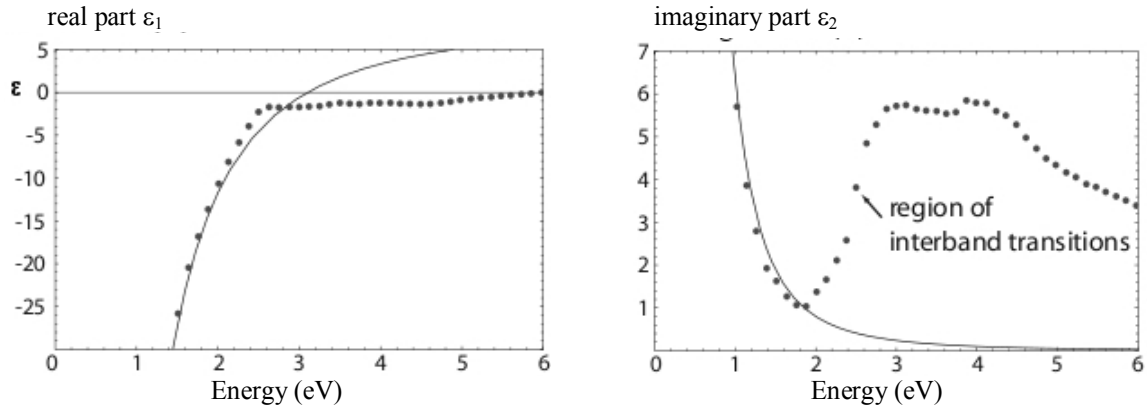



Fig. 5.1. - Behaviors of noble metals (dashed lines) with respect to the theoretical Drude model (solid line): the two behaviors diverge at around 2 eV, corresponding to the gap of the band transitions.

predominantly real ($\omega\tau >> 1$). We recall that in the latter regime the noble metals instead show a different behavior given by an increase of the absorption due to the jump of the band of the electrons, as already mentioned (see Fig. 5.1).
To overcome the different behavior of the permittivity of metals for frequencies comparable to the interband transitions [5], it is possible to still use the Drude model, rewriting the equation (5.2.4) in the following way:

$$m\ddot{\vec{x}} + m\gamma\dot{\vec{x}} + m\omega_0^2\vec{x} = -e\vec{E} \tag{5.2.14}$$

in which the additional term expresses the contribution of a fixed electron with resonant frequency $\omega_0$. From this equation we can obtain the polarization vector as we did for (5.2.4). We note, however, that the equation (5.2.14) is not only for one but for a given number pf electrons, each of which contributes to the effect of polarization providing a set of terms called Lorentz oscillators [6]. The addition of such terms contributes to the modification of the permittivity (5.2.11).

## 5.3. Volume plasmons

In the limit $\omega\tau >> 1$, where metals have a dielectric behavior, the permittivity is predominantly real and we can write:

$$\varepsilon_1(\omega) = 1 - \frac{\omega_p^2\tau^2}{1+\omega^2\tau^2} \approx 1 - \frac{\omega_p^2}{\omega^2} \tag{5.3.1}$$

We note immediately that if $\omega = \omega_p$, the permittivity $\varepsilon$ vanishes and under these conditions it is possible verify that Maxwell's equations allow longitudinal waves! Consider then a layer of electrons in two dimensions which oscillates in the third direction (denoted by z) parallel to a layer of positive charges (see Fig. 5.2).
The surface charge density is $\rho = \pm neu_z$, with n number of charges, while the 5.1.2 becomes:

$$\vec{E} = -\frac{\vec{P}}{\varepsilon_0} = \frac{neu_z}{\varepsilon_0}\hat{k} \tag{5.3.2}$$

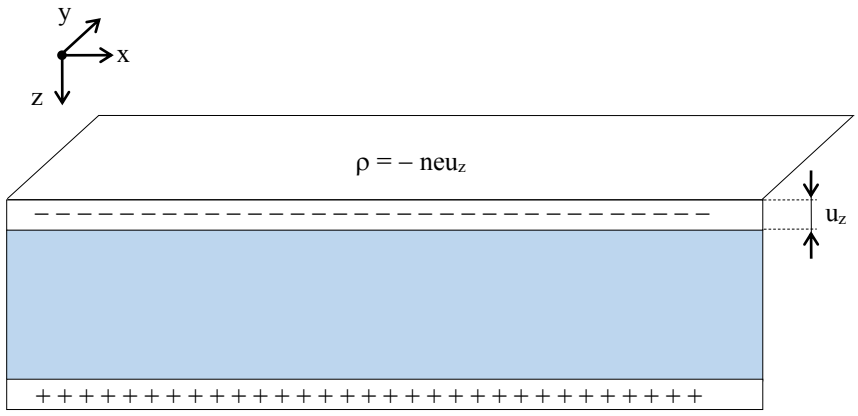



Fig. 5.2. - Charges which oscillate longitudinally in a metal.

Considering that a force F=qE acts on the charges, we can write the motion equation as follows:

$$n m \ddot{u}_z = -\frac{n e^2 u_z}{\varepsilon_0} \tag{5.3.3}$$

or equivalently:

$$\ddot{u}_z + \omega_p^2 u_z = 0 \tag{5.3.4}$$

Therefore we can identify the physical meaning of $\omega_p$ as the oscillation frequency of the longitudinal motions of the electron sea (which we suppose to move in phase). The quantum of such oscillation is the volume plasmon. The plasma frequency, for many metals, is in the order of 5↔15 eV, that is in the range the ultraviolet band.

## 5.3. Surface plasmons

In the previous section we have studied the behavior of metals on the basis of the response generated by the electron plasma in it content. With this approach we have obtained the permittivity and verified the existence of volume plasmonic quasi-particles, related to the quantized oscillatory motion of the plasma within of the metal. Let's see what happens on the surface of a metal and a dielectric in certain conditions. When an electromagnetic wave hitting this contact surface gives rise to a permittivity, whose real part is negative for the metal and absolutely greater than that of the dielectric. What happens is that coherent longitudinal oscillations of the plasma of free metal electrons take place and propagate in parallel to the contact surface. The associated electromagnetic field has a maximum near its interface and an exponentially decaying intensity in the direction perpendicular to the motion of such a wave, then towards the inside of the metal and the dielectric respectively. This electromagnetic waves connected to the electron plasma of the metal, which propagate along the interface and fade away from it, are called surface plasmon polaritons (SPPs).

## References Chapter 5


[1] S. A. Maier, *Plasmonics, Fundamentals and Applications*, Springer, 2007.

[2] J. D. Jackson, *Classical Electrodynamics*, Wiley, New York, 1975.

[3] M. P. Marder, *Condensed Matter Physics*, John Wiley & Sons, Inc., New York, 2000.

[4] P. Drude, "Zur Elektronentheorie der Metalle", 1900. *Ann. Phys*., 1:566–613.

[5] C. Kittel, *Introduction to Solid State Physics*, John Wiley & Sons, Inc., NewYork, 1996. 7th edition.

[6] A. Vial, A. S. Grimault, D. Macías, D. Barchiesi, and M. L. de la Chapelle,"Improved analytical fit of gold dispersion: Application to the modeling of extinction spectra with a finite-difference time-domain method", *Phys. Rev. B*, 71:085416 , 2005.

[7] C. Hägglund, B. Kasemo, "Nanoparticle Plasmonics for 2D-Photovoltaics: Mechanisms, Optimization and Limits", *Optics Express*, vol. 17, no. 14, pp. 11944-11957, 2009.

[8] E. Hutter and J. H. Fendler, "Exploitation of Localized Surface Plasmon Resonance", *Adv. Mater.*, vol. 16, pp. 1685–1706, 2004.

[9] H. A. Atwater and A. Polman, "Plasmonics for improved photovoltaic devices", *Nat. Mater*, vol. 9, pp. 205-213, 2010.

[10] M. I. Alonso, K. Wakita, J. Pascual, M. Garriga, and N. Yamamoto, "Optical functions and electronic structure of CuInSe2, CuGaSe2, CuInS2, and CuGaS2", *Phys. Rev. B,* vol. 63, 2001.

[13] C. Hägglund, M. Zäch, G. Petersson, and B. Kasemo, "Electromagnetic coupling of light into a silicon solar cell by nanodisk plasmons", *Appl. Phys. Lett*., vol. 92, 053110, 2008

[14] X. Huang, Z. Zeng, L. Zhong, D. Wu and F. Yan, "Optical Absorption Enhancement Effects of Silver Nanodisk Arrays in the Application of Silicon Solar Cell", *Journal of Electronic Science and Technology*, vol. 9, no. 1, pp. 35-4, 2011.

[17] D. E. Aspnes, "Optical Properties of Si," in *Properties of Crystalline Silicon*, R. Hull (ed), IEE INSPEC, London UK, 1999.

# Chapter 6

# Numerical analysis of light scattering from metallic nanoparticles

## 6.1. Analysis of Plasmons in Metallic Nanoparticles by FEM-RBCI

The FEM-RBCI method can be applied to the computation of plasmon oscillations induced by an incident monochromatic electromagnetic wave on one or more metallic nanoparticles of arbitrary shapes, embedded in free space. In a large range of particle sizes the metal can be simply modelled by means of a complex relative electric permittivity (Drude model) [5], given by:

$$\varepsilon_r = -\frac{\omega_p^2 - (\omega^2 + \nu^2)}{\omega^2 + \nu^2} - j\frac{\nu\omega_p^2}{\omega(\omega^2 + \nu^2)} \tag{6.1.1}$$

where $\omega_p$ is the plasma frequency of the free electrons and $\nu$ is the relaxation frequency, the values of which are experimentally determined [6]. Note that at the optical frequencies of interest, the real part of the relative electric permittivity is negative. Assuming a unitary relative magnetic permeability, the FEM analysis is easily performed.

By applying the FEM-RBCI method to the analysis of the scattering of an incident wave from a single particle, a single fictitious boundary $\Gamma_F$ may be selected homologously to the particle surface $\Gamma_P$ at a mean distance of $\lambda/20$ - $\lambda/10$ from it, $\lambda$ being the wavelength in free space (see Fig. 6,1). The same surface $\Gamma_P$ may be used as integration surface $\Gamma_M$.

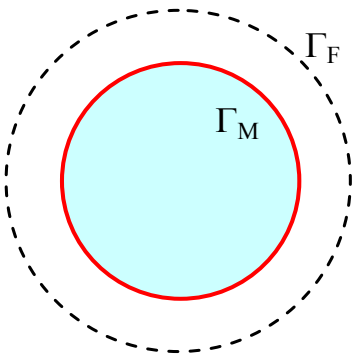



Fig. 6.1. - A single nanoparticle enclosed by a fictitious boundary.

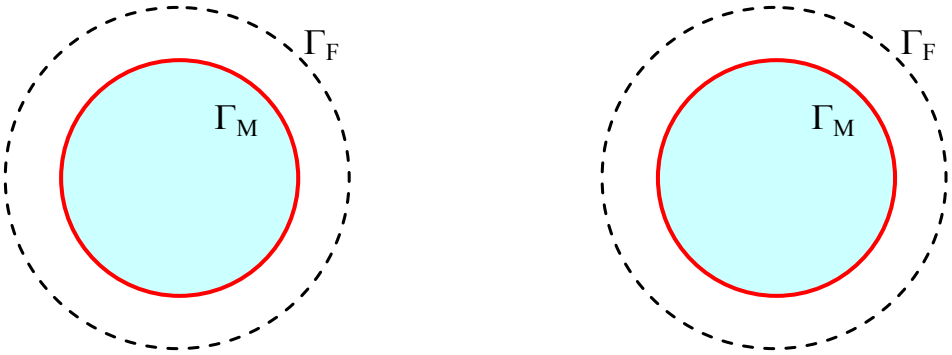



Fig. 6.2. - A couple of nanoparticles enclosed by two fictitious boundaries.

This simple strategy can be used also for systems of several particles which are placed at great distance from each others (see Fig. 6.2). For particles which are very near from each others, a single integration surface is conveniently selected which includes the particles (in some parts may be coincident with their surfaces) (see Fig. 6.3); analogously the fictitious boundary is selected as a single closed surface at a distance of about $\lambda/20$ - $\lambda/10$ from $\Gamma_M$. The same considerations apply to 2D analyses, provided that $\Gamma_F$ and $\Gamma_M$ are closed curves.
In postprocessing very often the calculation is required of some integral quantities, such as the absorption, scattering and extinction cross sections, respectively $\sigma_{abs}$, $\sigma_{scat}$ and $\sigma_{ext}$ [5]. Having solved the problem by means of the FEM-RBCI method, these calculations are straightforward. In fact the fictitious boundary may be used to calculate the following integrals:

$$\sigma_{abs} = -\frac{1}{\left|\overline{S}_{inc}\right|}\iint_{\Gamma_F} \mathrm{Re}\left\{\overline{E}\times\overline{H}^*\right\}\cdot\hat{n}\,dS \tag{6.1.2}$$

$$\sigma_{scat} = \frac{1}{\left|\overline{S}_{inc}\right|}\iint_{\Gamma_F} \mathrm{Re}\left\{\overline{E}_{scat}\times\overline{H}^*_{scat}\right\}\cdot\hat{n}\,dS \tag{6.1.3}$$

$$\sigma_{ext} = \frac{1}{\left|\overline{S}_{inc}\right|}\iint_{\Gamma_F} \mathrm{Re}\left\{\overline{E}_{inc}\times\overline{H}^*_{scat} + \overline{E}_{scat}\times\overline{H}^*_{inc}\right\}\cdot\hat{n}\,dS = \sigma_{abs} + \sigma_{scat} \tag{6.1.4}$$

where $\overline{S}_{inc} = \overline{E}_{inc}\times\overline{H}^*_{inc}$ is the Poynting vector of the incident field, very often a linearly-polarized plane wave.

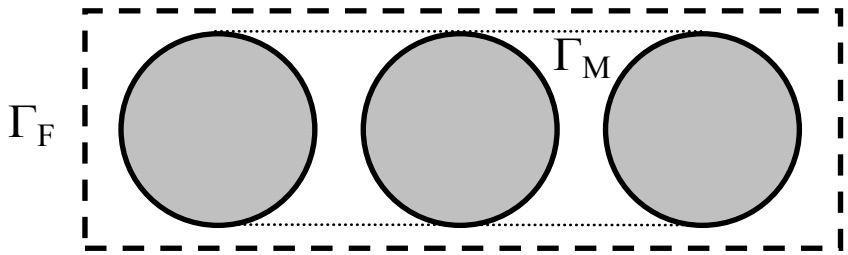



Fig. 6.3. - Selection of the truncation boundary $\Gamma_F$ and of the integration surface $\Gamma_M$ in a system of three adjacent nanoparticles.

Since the surface $\Gamma_F$ is made of plane triangular patches, resulting from the tetrahedral finite element discretization of the domain internal to this boundary, the calculation of the first integral can be performed as

$$\sigma_{abs} = -\frac{1}{\left|\overline{S}_{inc}\right|}\frac{1}{\omega\mu_0}\sum_k\sum_{m=1}^{N-1}\sum_{n=m+1}^{N}\left[\mathrm{Im}\{E_m\}\mathrm{Re}\{E_n\}-\mathrm{Im}\{E_n\}\mathrm{Re}\{E_m\}\right]q_{mn}^{(k)} \tag{6.1.5}$$

where index k refer to the k-th triangular face $S_k$ on $\Gamma_F$, indices m and n refer respectively to the m-th and n-th edges of the tetrahedral finite element relative to $S_k$, N is the number of edges of the finite element (N=6 for first-order tetrahedra) and $q_{mn}$ are geometrical scalar values given by:

$$q_{mn}^{(k)} = \iint_{S_k}(\overline{\alpha}_m \times \mathrm{rot}\,\overline{\alpha}_n - \overline{\alpha}_m \times \mathrm{rot}\,\overline{\alpha}_n)\cdot\hat{n}\,dS \tag{6.1.6}$$

Formula (6.1.5) applies also for the numerical computation of the $\sigma_{scat}$ cross section, provided that the total field be substituted by the scattered field.
The same formula (6.1.5) applies also for 2-D problems, provided that now $S_k$ is the k-th segment (coincident with a side of a finite element), N is the number of the nodes of the element and

$$q_{mn}^{(k)} = \int_{S_k}(\alpha_m \mathrm{grad}\,\alpha_n - \alpha_n \mathrm{grad}\,\alpha_m)\cdot\hat{n}\,ds \tag{6.1,7}$$

where $\alpha_m$ and $\alpha_n$ are nodal shape functions. Of course other surfaces (curves) which include all the particles can be used in the equations (6.1.2)-(6.1.4) instead of $\Gamma_F$, as for example the integration surface $\Gamma_M$. To obtain a greater accuracy, one can compute the above cross sections by employing both the surfaces $\Gamma_F$ and $\Gamma_M$ and assume the mean of such values.

## 6.2. Numerical results

The first system analyzed is a 2-D one in which a silver nanocylinder of circular cross section of radius R= 50 nm, having its axis coincident with the z-axis, is lit up by a plane wave of wavelength $\lambda$= 413 nm, E-polarized along the z axis and travelling toward the positive x-axis: $\overline{E}_{inc} = E_0 e^{-jk_0 x}\hat{z}$, with $E_0$=1 V/m. According to the data reported in [6] the relative electric permittivity of the metal was assumed to be $\varepsilon_r$= –5.173125 –j0.2275 whereas the relative magnetic permeability was set to $\mu_r$=1. In order to observe the field around the nanocylinder, the fictitious boundary was selected rather far from the cylinder surface (at a distance of 75 nm), whereas the cylinder surface itself was selected as integration surface. For symmetry reasons the analysis can be conveniently restricted to half the xy plane (y>0), by imposing a homogeneous Neumann boundary condition on the x-axis. The bounded domain so obtained was discretized by means of 906 triangular finite elements of the second order (256 lie in the cylinder and 650 outside) and 1881 nodes. An end-iteration tolerance of 0.01 per cent was set for the FEM-RBCI iterations and 0.0001 for the COGC solver. Five iteration steps were needed to obtain convergence. Fig. 6.4 reports the contours of the

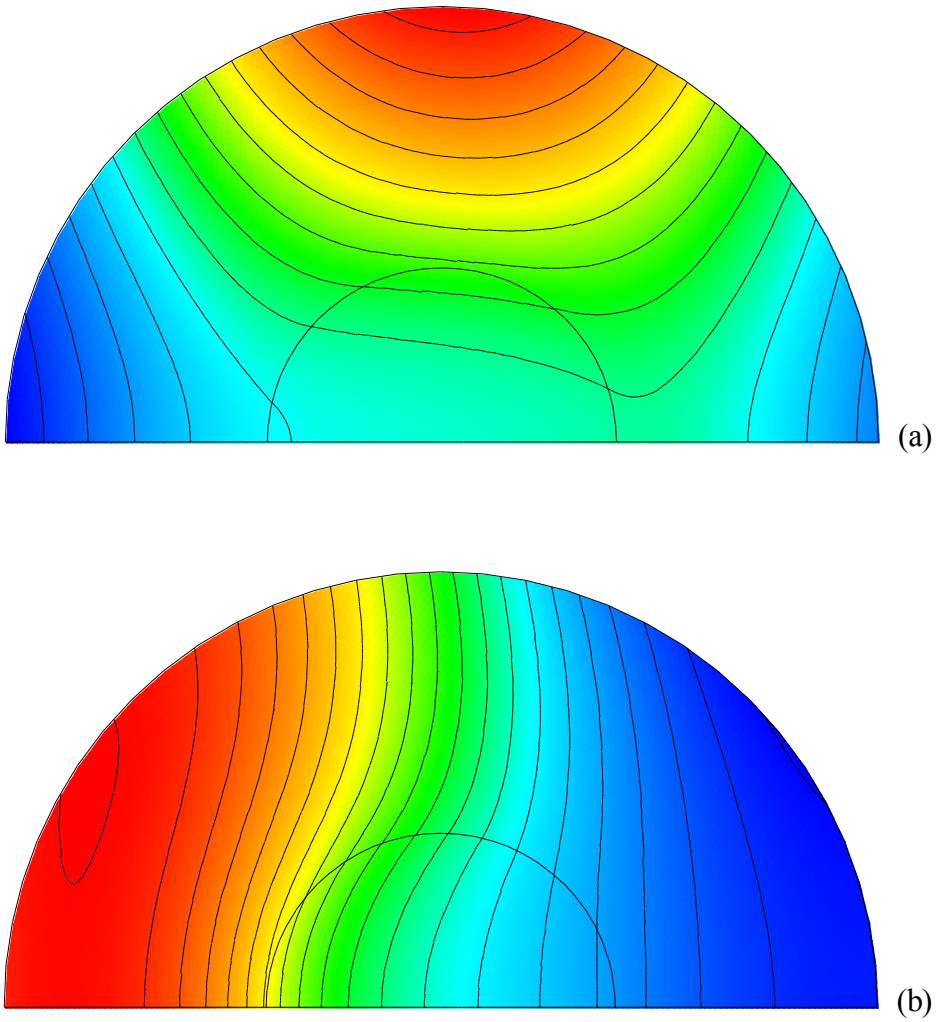



Fig. 6.4. - Contours of the real (a) and imaginary(b) parts of the electric field (1st system)

real and imaginary parts the total electric field. This solution was compared with the analytical one: a mean difference of 0.096 per cent was estimated. Starting from the numerical solution, the cross sections per unit length were evaluated by employing both $\Gamma_F$ and $\Gamma_M$ curves to obtain: $\sigma_{abs}$= 3.856 nm, $\sigma_{scat}$= 171.5 nm, $\sigma_{ext}$= 175.4 nm by using $\Gamma_F$, and $\sigma_{abs}$= 3.937 nm, $\sigma_{scat}$= 171.0 nm, $\sigma_{ext}$= 175.0 nm by using $\Gamma_M$. A good agreement can be pointed out.

A second system analyzed is constituted by a pair of nanocylinders each of which is the same as that above; the system is illuminated by the same wave of the previous analysis. The centres of the two cylinders are $C_1$=(0,R+d/2) and $C_2$=(0, –R–d/2), where d=5 nm is the gap between their surfaces. The integration surface was selected as the envelope of two circumferences of radius $R_M$=52.25 nm centred in $C_1$ and $C_2$; analogously the fictitious boundary was selected as the envelope of two circumferences of equal radius $R_F$=125 nm, centred in $C_1$ and $C_2$. Also in this case the analysis was restricted to the y>0 half plane. The mesh is formed by 1546 second-order triangles (512 inside the nanocylinder) and 3173 nodes. Fig. 6.5 reports the contours of the magnitude of the electrical field. The cross sections are evaluated as: $\sigma_{abs}$= 6.690 nm, $\sigma_{scat}$= 366.0 nm, $\sigma_{ext}$= 372.7 nm by using $\Gamma_F$, and $\sigma_{abs}$= 6.797 nm, $\sigma_{scat}$= 365.4 nm, $\sigma_{ext}$= 372.2 nm by using $\Gamma_M$.

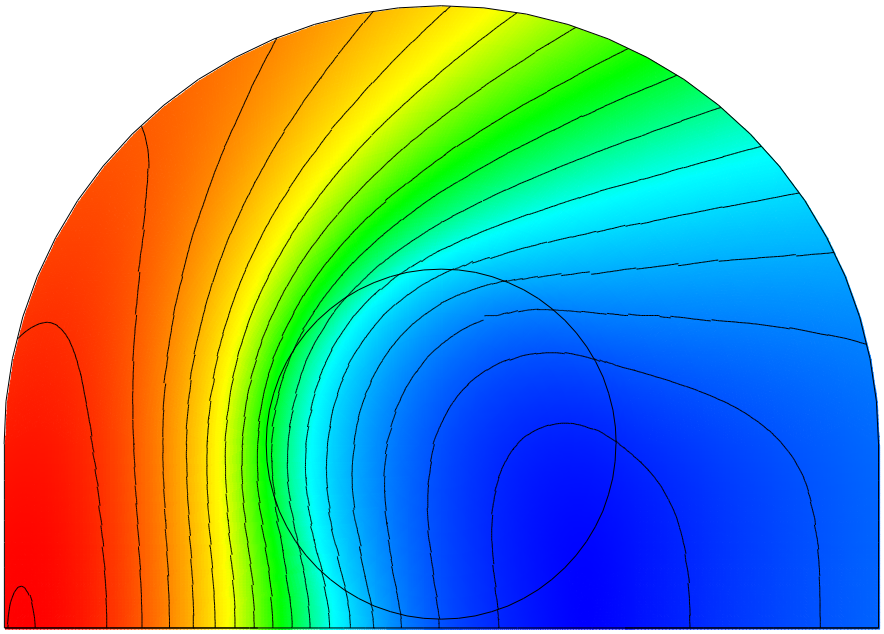


Fig. 6.5. - Contours of the magnitude of the electric field (2nd system)

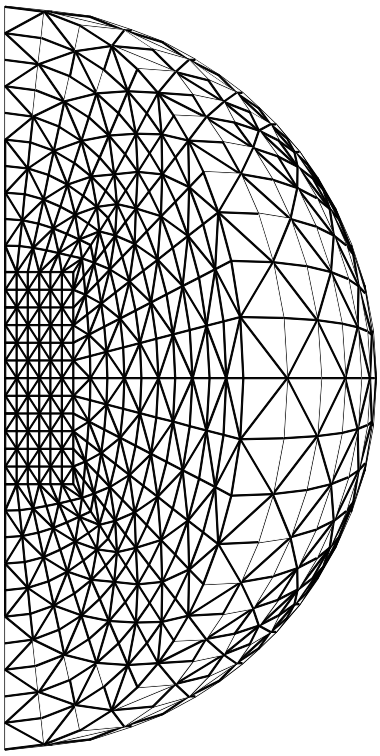


Figure 6.6. - Tetrahedral mesh.

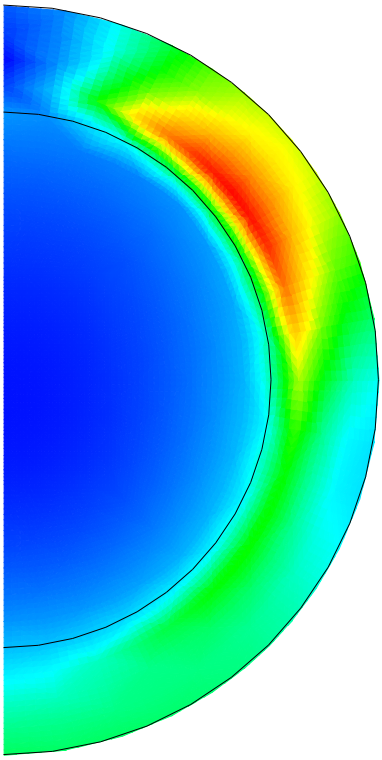


Figure 6.7. - Electrical field modulus in the xz plane.

The third example is a spherical gold nanoparticle of radius R= 100 nm, centered in the origin, lit up by a plane wave $\overline{E}_{inc} = E_0 e^{-jk_0 z}\hat{x}$, of wavelength $\lambda$ = 500 nm, E-polarized along the x-axis, proceeding toward the positive z-axis. The relative electric permittivity of the metal was assumed to be $\varepsilon_r = -2.56066 - j3.60402$ whereas the relative magnetic permeability was set to $\mu_r=1$. Due to symmetry reasons the analysis can be restricted to a quarter of the space (x>0, y>0) by imposing homogeneous Dirichlet and Neumann boundary condition on the yz and xz planes, respectively. A spherical fictitious boundary $\Gamma_F$ is selected homologously to the particle surface, having a radius of $R_F$ = 140 nm. The domain was discretized by means of 12960 tetrahedral edge elements of the first order (8640 lie in the nanoparticle), 16846 edges and 3047 nodes. The surface of the nanoparticle (432 triangular patches) is used as integration surface $\Gamma_M$. Fig. 6.6 shows the FE mesh, whereas Fig 6.7 reports the modulus of the total electrical field in the xz plane. The solution was compared with the analytical one: a mean difference of 6.1 per cent was estimated. The cross sections are evaluated as: $\sigma_{abs}= 5.291\cdot10^{-14}$ m$^2$, $\sigma_{scat}= 5.497\cdot10^{-14}$ m$^2$, $\sigma_{ext}= 10.79\cdot10^{-14}$ m$^2$ by using the surface $\Gamma_F$.

The fourth system analyzed is constituted by a single nanoring made of gold. The dimensions of the particle are: outer radius R = 60 nm, height h = 40 nm, thickness s = 14 nm (see Fig. 6.8). A Cartesian reference frame is set having the origin in the centre of the nanoring and the z-axis coincident with that of the particle. The particle is embedded in air. An electromagnetic plane wave lights the particle; its wavelength is $\lambda$= 1215 nm. The wave is E-polarized along the x-axis and travels toward the positive z-axis: $\overline{E}_{inc} = E_0 e^{-jk_0 z}\hat{x}$, with $E_0$=1 V/m. According to the data reported in [9] the relative electric permittivity of the metal was assumed to be $\varepsilon_r = -66.218525 - j57015$ whereas the relative magnetic permeability was set to $\mu_r$=1. In order to observe the field around the nanocylinder, the truncation boundary was selected at a distance of 10 nm from the ring surface, whereas the nanoparticle surface itself was selected as integration surface. For symmetry reasons the analysis can be conveniently restricted to a quarter of the system (domain x>0 y>0), by imposing homogeneous Neumann and Dirichlet boundary conditions on the xz- and yz-planes, respectively. The bounded domain so obtained was discretized by means of 23040 tetrahedral edge elements of the first order (5120 lie in the nanoring), 30440 edges and 5577 nodes (see Fig. 6.9). An end-iteration tolerance of 0.01 per cent was set for the FEM-RBCI

iterations and 0.0001 for the COGC solver. Sept iteration steps were needed to obtain convergence. Fig. 6.10 reports the contours of the real and imaginary parts of the total electric field on the xz plane.

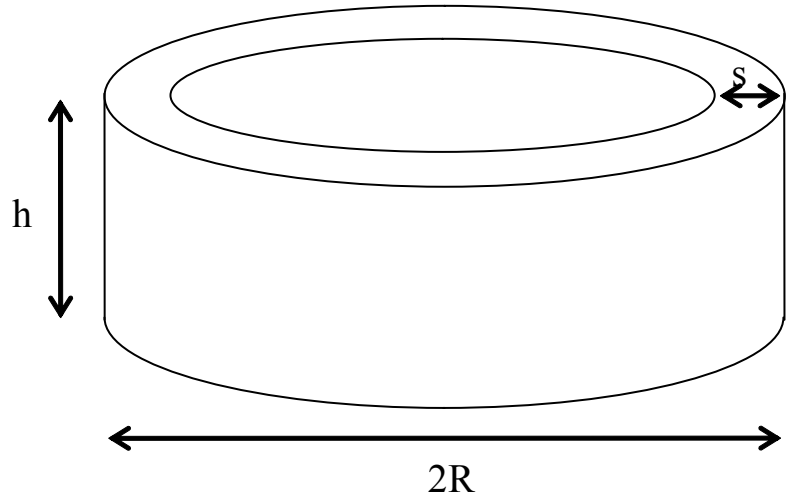



Fig. 6.8. - Geometry of the gold nanoring analyzed.

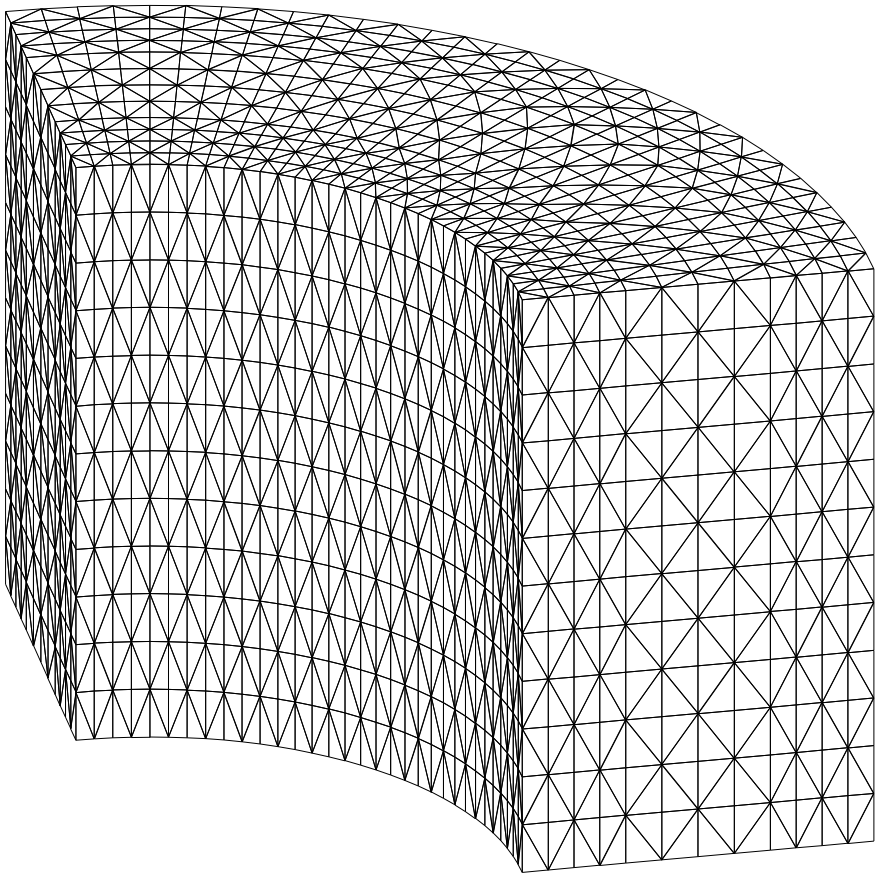


Fig. 6.9. - Finite element mesh of the nanoring.

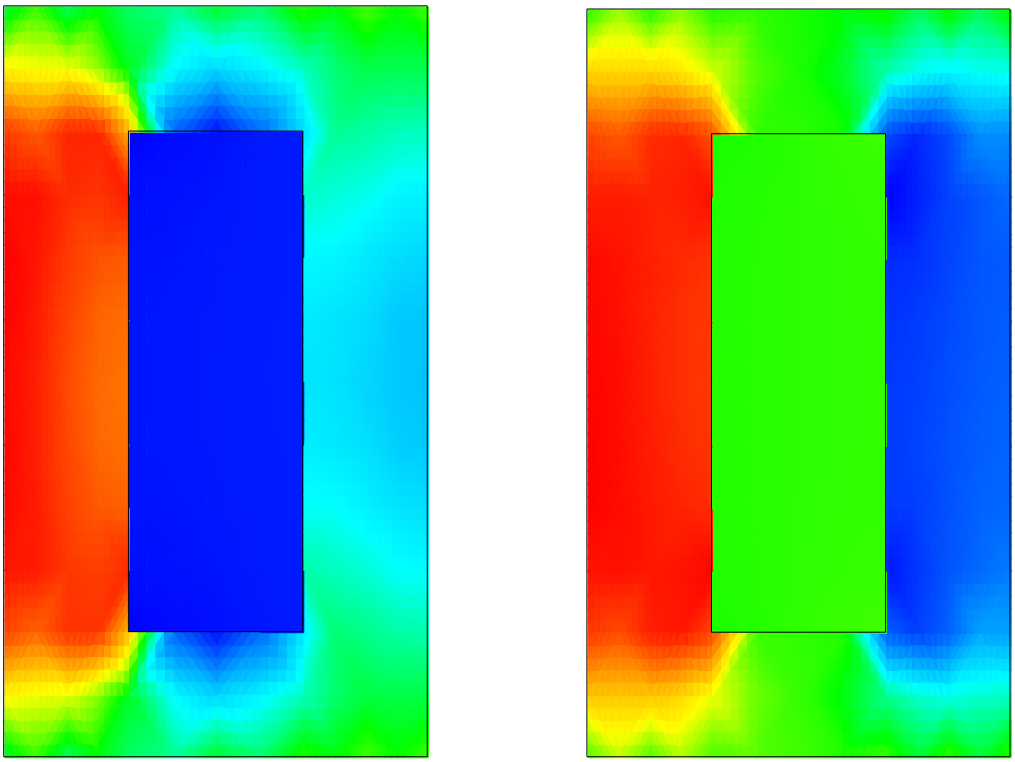


Fig. 6.10. - Real and imaginary part of E.

## References Chapter 6


[1] S. Alfonzetti, G. Borzì, and N. Salerno, "FEM Analysis of unbounded electro-magnetic scattering by the Robin iteration procedure", *Electronic Letters*, vol. 32, pp. 1768-1769, 1996.

[2] S. Alfonzetti, G. Borzì, and N. Salerno, "Iteratively-Improved Robin boundary conditions for the finite element solution of scattering problems in unbounded domains", *Int. J. for Numerical Methods in Engineering,* vol. 42, pp. 601-629, 1998.

[3] S. Alfonzetti and G. Borzì, "Accuracy of the Robin boundary condition iteration method for the finite element solution of scattering problems", *Int. J. of Numerical Modelling: Electronic Networks, Devices and Fields*, vol. 13, pp. 217-231, vol. 2000.

[4] S. Alfonzetti and G. Borzì, "Finite element solution to electromagnetic scattering problems by means of the Robin boundary condition iteration method", *IEEE Trans. Ant. Prop.,* vol. 50, pp. 132-140, 2002.

[5] C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles*, Wiley, 1983.

[6] P.B. Johnson and R. W. Cristy, "Optical constants of the noble metals", *Phys. Rev. B,* vol. 6, pp. 4370-4379, 1972.

[7] G. Aiello, S. Alfonzetti, G. Borzì, and N. Salerno, "An overview of the ELFIN code for finite element research in electrical engineering", in *Software for Electrical Engineering Analysis and Design*, A Konrad & C. A. Brebbia (ed.), WIT Press, Southampton (U.K.), 1999.

# Chapter 7

# Optimization of a solar cell with metallic nanoparticles

## 7.1. Generality

The study of the efficiency of solar cells made of photovoltaic (PV) layers is very important to the aim of reduce the use of fossil fuels. In particular the capacity of very thin layers to capture the sunlight is of great interest in order to reduce the amount of expensive material. Unfortunately, the standard materials of today commercially available solar cells do not allow such a thickness reduction without a great decaying of light absorption.
In the solar energy conversion, light absorption is mainly connected to a low number of electrons contributing to the absorption cross section over the spectral range of interest. Theoretically, the smallest amount of material needed for total absorption has been estimated to be a film of thickness of about 10 nm, whereas the today's thin solar cells have thickness of about 1 μm.
Recently several researchers have demonstrated that the insertion of metallic (in particular aurum, silver and copper) nanoparticles within or near PV layers may improve significantly the efficiency of thin solar cells. This effect is due to plasmon resonances [1-6], which lead to increased electron-hole generation in the close PV layer.
In the following we optimize the geometry of a thin solar cell by employing an FEM (Finite Element Method) code to compute the light scattering from the solar cell and suitable genetic algorithms (GAs) for the optimization of the cell geometry.

## 7.2. 3D FEM analysis of light scattering from solar cells

Consider the thin film solar cell depicted in Fig. 7.1, in which hemi-ellipsoidal metallic nanoparticles, having semi-axes a, b and c, with a=b>c, are placed near a PV layer, having a thickness t. The nanoparticle centers are regularly placed at the nodes of a rectangular grid, exhibiting the same grid step 2d along the x- and y-axis. To simplify the optimization the nanoparticles are assumed to have a given volume $V=2\pi abc/3$, so that only one degree of freedom specifies their shape.
For the sake of simplicity, we assume that this system is radiated by a monochromatic electromagnetic plane wave, E-polarized along the x-axis (a time factor $\exp(j\omega t)$ is understood):

$$\overline{E} = E_{max} \exp(jk_0 z)\,\hat{x} \tag{7.2.1}$$

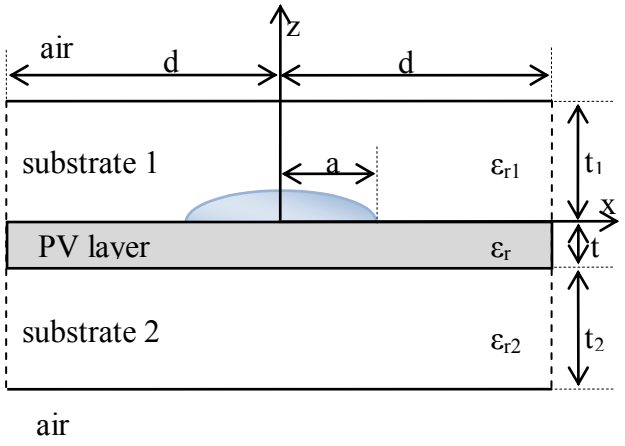



Fig. 7.1. - Three-layer solar cell with a grid of hemi-ellipsoidal metallic nanoparticles.

where $E_{max}$ is the maximum value of the electrical field, $\hat{x}$ is the versor of the x-axis and $k_0$ is the free-space wavenumber, given by:

$$k_0 = \omega\sqrt{\varepsilon_0\mu_0} \quad , \tag{7.2.2}$$

in which $\omega$ is the angular frequency and $\mu_0$ and $\varepsilon_0$ are the free-space magnetic permeability and electric permittivity, respectively.
For this electromagnetic scattering problem the Helmholtz vector equation holds:

$$\nabla\times\left(\mu_r^{-1}\nabla\times\overline{E}\right)-k_0^2\varepsilon_r\overline{E}=0 \tag{7.2.3}$$

where $\mu_r$ and $\varepsilon_r$ are the relative magnetic permeability and electrical permittivity, respectively
At optical frequencies the metallic nanoparticles give rise to plasmons oscillations, which are taken into account by modeling the metal by means of a complex relative electric permittivity (Drude model) [1], given by:

$$\varepsilon_r = -\frac{\omega_p^2-(\omega^2+\nu^2)}{\omega^2+\nu^2}-j\frac{\nu\,\omega_p^2}{\omega(\omega^2+\nu^2)} \tag{7.2.4}$$

where $\omega_p$ is the plasma frequency of the free electrons and $\nu$ is the relaxation frequency, the values of which are experimentally determined [8]. Note that at the optical frequencies of interest, the real part of the relative electric permittivity is negative.
The active PV layer, is made of semiconductor material, namely the copper indium selenide ($CuInSe_2$ or CIS), which is electrically modeled by a complex electrical permittivity which varies with the wavelength; these values are obtained from tabulated data in [9].
A substrate having a fixed thickness $t_1$=250 nm and exhibiting a real relative electric permittivity $\varepsilon_{r1}\approx Re\{\varepsilon_r\}$ at the wavelength $\lambda$=900 nm is inserted on the top of the PV layer to minimize the reflection losses at the interface air-solar cell. Analogously another substrate of fixed thickness $t_2$=250 nm and real relative electric permittivity $\varepsilon_{r2}<\varepsilon_{r1}$ is inserted behind the PV layer.
For the nanoparticle metal, the PV layer and the substrate materials a unitary relative magnetic permeability is assumed.

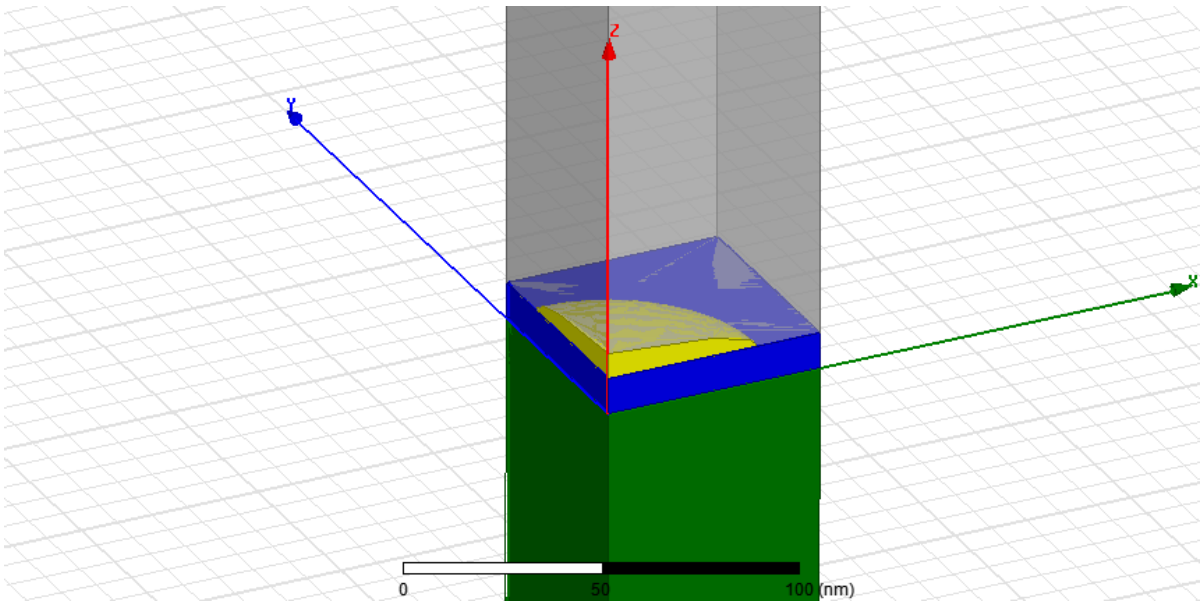



Fig. 7.2. – Domain of the FEM analysis.

For symmetry reasons the analysis can be restricted to the square domain $0 \le x \le d$, $0 \le y \le d$, by imposing homogeneous Dirichlet boundary conditions on the x=0 and x=d planes and homogeneous Neumann ones on the y=0 and y=d planes (see Fig. 7.2). On the z direction the domain is truncated by means of two thin layers of air and PMLs (Perfectly Matched Layers) [7], not shown in Fig. 7.1, placed over and under the cell.
By discretizing the analysis domain by means of tetrahedral edge elements, the finite element analysis is easily performed [10-11]. In postprocessing the following fluxes of the Poynting vector (powers) are computed:

$$W_i = \frac{1}{2}\iint_{\Gamma_1} \mathrm{Re}\left\{\overline{E} \times \overline{H}^*\right\} \cdot \hat{n}_1 \, dS \tag{7.2.5}$$

$$W_o = \frac{1}{2}\iint_{\Gamma_2} \mathrm{Re}\left\{\overline{E} \times \overline{H}^*\right\} \cdot \hat{n}_2 \, dS \tag{7.2.6}$$

where $\Gamma_1$ and $\Gamma_2$ are the two air-PML plane interfaces lying over and under the cell, respectively, $\hat{n}_1$ and $\hat{n}_2$ are the normal versors to the above surfaces ($\hat{n}_1 = \hat{n}_2 = -\hat{z}$, with $\hat{z}$ the versor of the z-axis). Since the surfaces $\Gamma_1$ and $\Gamma_2$ are made of plane triangular patches, resulting from the tetrahedral discretization, the contribution of the generic patch $T_k$ to the integrals (4) and (5) can be computed as

$$w_k = \frac{1}{2\omega\mu_0} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left[\mathrm{Re}\{E_i\}\mathrm{Im}\{E_j\} - \mathrm{Re}\{E_j\}\mathrm{Im}\{E_i\}\right] q_{ij}^{(k)} \tag{7.2.7}$$

where index k refers to the k-th triangular patch $T_k$ on $\Gamma_1$ or $\Gamma_2$, indices i and j refer to the i-th and j-th edges, respectively, of the tetrahedral element relative to $T_k$, N is the number of edges of the finite element (N=6 for first-order tetrahedra), $E_i$ and $E_j$ are the mean values of the electric fields along the i-th and j-th edges, and $q_{ij}$ are geometrical scalar values given by:

$$q_{ij}^{(k)} = \iint_{T_k} (\overline{\alpha}_i \times \nabla \times \overline{\alpha}_j - \overline{\alpha}_j \times \nabla \times \overline{\alpha}_i) \cdot \hat{n}_k \, dS \tag{7.2.8}$$

in which $\overline{\alpha}_i$ and $\overline{\alpha}_j$ are the vector form functions of the i-th and j-th tetrahedron edges, respectively, and $\hat{n}_k$ is the normal versor of the tetrahedron face $T_k$.
Afterwards the Joule losses in the metallic particle are computed:

$$W_j = \frac{1}{2} \iiint_V \sigma \left|\overline{E}\right|^2 dx dy dz \tag{7.2.9}$$

where V is the volume of the nanoparticle and $\sigma$ the metal conductivity. The contribution $w_h$ of the h-th tetrahedron $E_h$ of the nanoparticle to the Joule loss $W_j$ is computed numerically as:

$$w_h = \frac{1}{2} \sigma \sum_{i=1}^{N} \sum_{j=1}^{N} \left[ \mathrm{Re}\{E_i\} \mathrm{Re}\{E_j\} + \mathrm{Im}\{E_i\} \mathrm{Im}\{E_j\} \right] t_{ij}^{(h)} \tag{7.2.10}$$

where $t_{ij}$ is the generic entry of the metric matrix:

$$t_{ij}^{(h)} = \iiint_{E_h} \overline{\alpha}_i \cdot \overline{\alpha}_j \, dx dy dz \tag{7.2.11}$$

The following non dimensional quantity is then evaluated:

$$\eta(a, d, t) = \frac{W_i - W_o - W_j}{S_{inc} d^2} \tag{7.2.12}$$

where $S_{inc}$ is the modulus of the Poynting vector of the incident electromagnetic wave:

$$\overline{S}_{inc} = \frac{1}{2} \overline{E}_{inc} \times \overline{H}_{inc}^{*} = -\frac{E_{max}^2}{2Z_0} \hat{z} \tag{7.2.13}$$

in which $Z_0$ is the impedance of free space.
The meaning of $\eta$ (here considered as a function of the geometrical parameters a, d and t) is that of representing the power absorbed by the PV layer, normalized by the power of the incident wave in the square of area $d^2$. Obviously, a fraction of the power absorbed by the PV layer is not useful to the solar energy conversion, because it is transformed into heat. However several experimental studies have shown that this fraction should be small, so that in the following we assume that all the power absorbed by the PV layer will be useful for the photocurrent conversion, as well.

## 7.3. Optimization by GAs

By employing the quantity defined in (7.2.12) as objective function to be maximized, a stochastic optimization is started by assuming the following data:

| | |
|---|---|
| *light wavelength:* | $\lambda = 900$ nm |
| *nanoparticle metal*: | silver |
| *nanoparticle volume* | $V = 25600$ nm$^3$ |
| *metal permittivity*: | $\varepsilon_r = -40.6 - j0.51$ |
| *metal permeability*: | $\mu_r = 1.00$ |
| *PV layer permittivity*: | $\varepsilon_r = 8.65 - j2.68$ |
| *PV layer permeability*: | $\mu_r = 1.00$ |
| *substrate-1 permittivity*: | $\varepsilon_{r1} = 9.00$ |
| *substrate-1 permeability*: | $\mu_{r1} = 1.00$ |
| *substrate-2 permittivity*: | $\varepsilon_{r2} = 2.25$ |
| *substrate-2 permeability*: | $\mu_{r2} = 1.00$ |

The geometrical parameters to be optimized are assumed to vary in the following ranges:

$$5 \text{ nm} < a < 45 \text{ nm}$$

$$50 \text{ nm} < d < 500 \text{ nm}$$

$$10 \text{ nm} < t < 100 \text{ nm}$$

The stochastic optimization is pursued by means of Genetic Algorithms, which are search algorithms which simulate the random evolution of populations of biological entities. These well-known algorithms have proven to be very efficient in optimizing electromagnetic devices [12-17] both for low and high frequency applications. In this paper suitable GAs are used [14], which have already shown their ability in the optimization of mathematical test functions and electromagnetic devices [15-17]. Their main characteristics are the following:

- *population size:*

$$P=30$$

- *number of generations:*

$$N_g=30$$

- *binary lengths:*

$$N_a=6 \qquad N_d=7 \qquad N_t=7$$

- *selection*:

tournament (no elitism)

- *crossover type*:

two-point crossover

- *crossover probability at generation k*:

$$P_c^{(k)} = 0.3 + 0.4\frac{k-1}{N_g - 1}$$

- *mutation probability at generation k*:

$$P_m^{(k)} = 0.05 - 0.04\frac{k-1}{N_g - 1}$$

As far as representation is concerned, the three geometric variables a, d, and t, were coded into binary strings of six, seven and seven bits, respectively, giving a total of N=20 bits.
The reproduction process, which randomly creates a new generation from the old one, was chosen by tournament selection with a shuffling technique to choose random pairs.
The crossover process, by means of which individuals exchange chromosomes from one generation to the other, was two-point crossover with a probability $P_C$ linearly varying from 0.3 to 0.7 while optimization goes on.
The mutation process, by means of which some random flips in the chromosomes of an individual are made, was employed with a probability $P_M$ linearly decreasing from 0.05 to 0.01 while optimization proceeds.
The evolution was halted after 30 generations, reaching an optimal function value $\eta_{opt}$=0.69, in relation to the following parameter configuration (values rounded to 0.5 nm):

$$a = 23\,\text{nm}, \quad d = 50\ \text{nm}, \quad t = 14\ \text{nm} \tag{7.3.1}$$

In Fig. 7.3 the best (maximum) and mean values of the objective function are plotted through the various generations: it shows the good convergence property of the Genetic Algorithms employed.

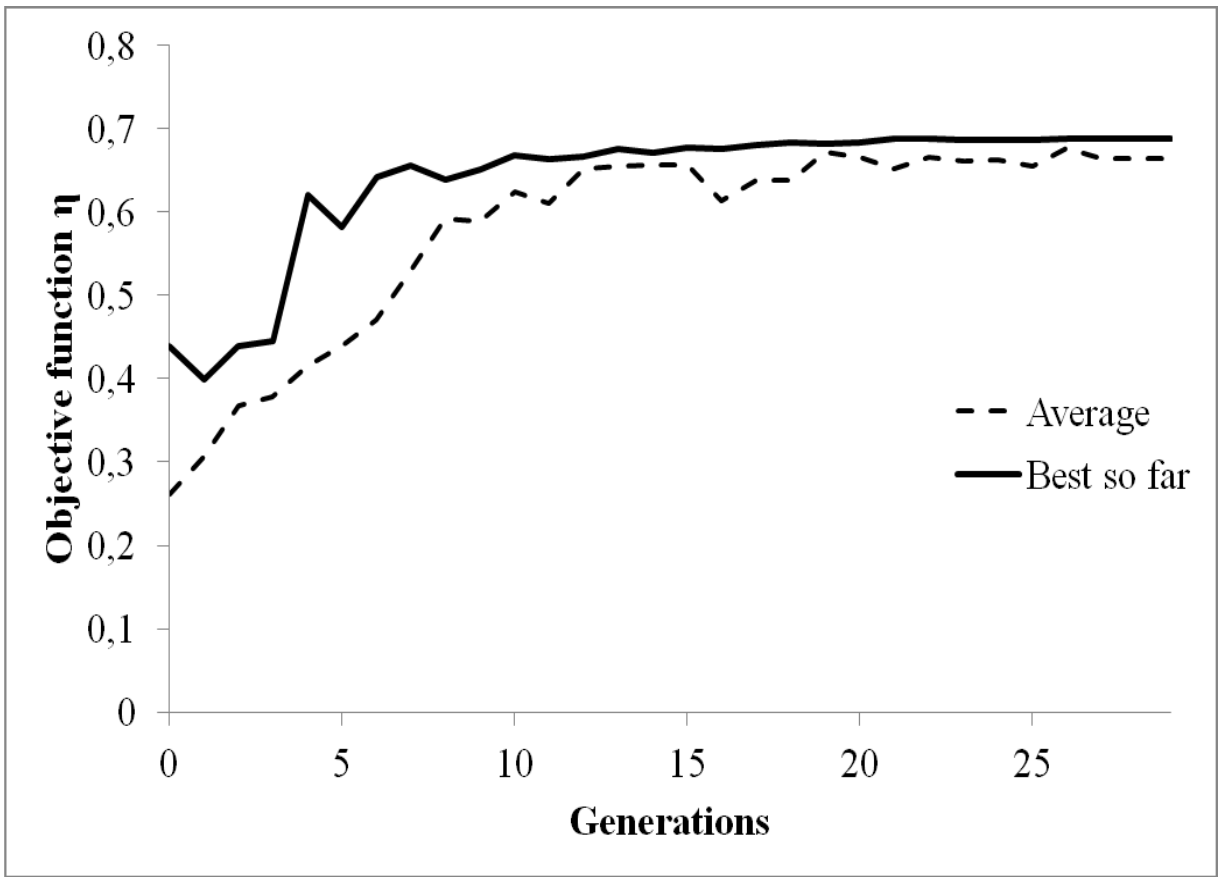



Fig. 7.3. - Best and average objective function η values over GA generations.

Fig. 7.4 shows the behavior of objective function η vs. the wavelength for the best configuration (7.3.1) of the geometrical parameters. The solid line refers to the function η defined in (7.2.12); note that its maximum value is about 0.69 and is obtained at a wavelength very close to 900 nm.
In the same figure the dotted line represents the function

$$\eta_0(t) = \frac{W_i - W_o}{S_{inc} d^2} \tag{7.3.2}$$

which has the same meaning of η but it is evaluated for the same solar cell in the absence of the metallic nanoparticles. The dashed line gives the difference $\Delta\eta=\eta-\eta_0$ and represents the absorption gain induced by the nanoparticles: its maximum is about 0.59 at 908 nm.

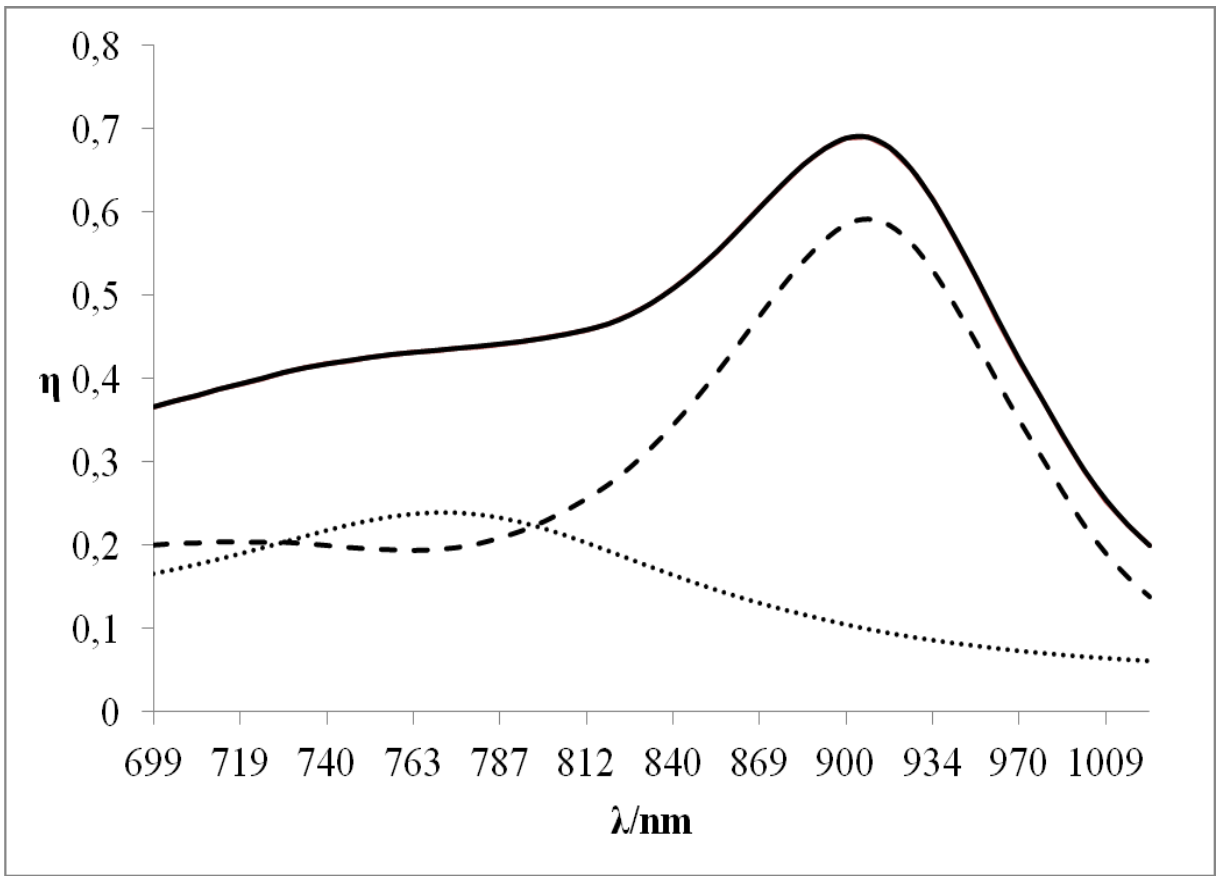



Fig. 7.4. - Behavior of the objective function η (solid line) vs the wavelength λfor the best configuration of the geometrical parameters. The dotted line represent the function $\eta_0$ evaluated in the absence of the nanoparticles. The dashed line gives the difference $\Delta\eta = \eta - \eta_0$.

## References Chapter 7


[1] C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles*, Wiley, 1983

[2] C. Hägglund, M. Zäch, G. Petersson, and and B. Kasemo, "Electro-magnetic coupling of light into a silicon solar cell by nanodisk plasmons", *Applied Physics Letters,* vol. 92, 053110, 2008.

[3] C. Hägglund and B. Kasemo, "Nanoparticle plasmonics for 2D-photovoltaics: mechanisms, optimization, and limits", *Optics Express*, vol. 17, pp. 11944-11957, 2009.

[4] C. Hägglund and S. P. Apell "Plasmonic near-field absorbers for ultrathin solar cells", *Physical Chemistry Letters*, vol. 3, pp. 1275-1285, 2012.

[5] S. H. Lim, W. Mar, P. Matheu, D. Derkacs, E. T. Yu "Photocurrent spectroscopy of optical absorption enhancement in silicon photodiodes via scattering from surface plasmon polaritons in gold nanoparticles", *Journal of Applied Physics*, vol. 101, no. 10, pp. 104309, 2007.

[6] P. Matheu, S. H. Lim, D. Derkacs, C. McPheeters, E. T. Yu "Metal and dielectric nanoparticle scattering for improved absorption in photovoltaic devices", *Applied Physics Letters*, vol. 93, 13108, 2008.

[7] J.-P. Berenger, "A perfectly matched layer for the absorption of electromagnetic wave", *Journal of Computational Physics*, vol. 114, pp. 185-200, 1994.

[8] P.B. Johnson and R. W. Cristy, "Optical constants of the noble metals", *Phys. Rev. B*, vol. 6, pp. 4370-4379, 1972.

[9] M. I. Alonso, K. Wakita, J. Pascual, M. Garriga, and N. Yamamoto, "Optical functions and electronic structure of $CuInSe_2$, $CuGaSe_2$, $CuInS_2$, and $CuGaS_2$", *Phys. Rev. B*, vol. 63, pp. 075203_1-075203_13, 2001.

[10] S. Alfonzetti, G. Borzì, "Finite element solution to electromagnetic scattering problems by means of the Robin boundary condition iteration method," *IEEE Trans. Ant. Prop.*, vol. 50, no. 2, pp. 132-140. Febr. 2002.

[11] G. Aiello, S. Alfonzetti, V. Brancaforte, V. Chiarello, N. Salerno, "Applying FEM-RBCI to the analysis of plasmons in metallic nanoparticles", *Int. J. of Appl. Electromagn. Mech. (IJAEM),* vol. 39, no. 1-4, pp. 13-20, 2012.

[12] S. Russenschuck, "Synthesis, inverse problems and optimization in computational electromagnetics," *Int. J. Numer. Modelling: Electronic Networks, Devices and Fields*, vol. 9, pp. 45-57, January-April 1996.

[13] P. G. Alotto, et alii, "Stochastic algorithms in electromagnetic optimization," *IEEE Transactions on Magnetics*, vol. 34, pp. 3674-3684, Sept. 1998.

[14] S. Alfonzetti, E. Dilettoso, N. Salerno, "A proposal for a universal parameter configuration for genetic algorithm optimization of electromagnetic devices," *IEEE Trans. Magn.*, vol. 37, n. 5, pp. 3208-3211, Sept. 2001.

[15] S. Alfonzetti, F. Dughiero, N. Salerno, "Stochastic optimization of an induction heating system by means of DBCI," *COMPEL*, vol. 19, no. 2, pp. 569-575, 2000.

[16] G. Aiello, S. Alfonzetti, N. Salerno, "Stochastic optimization of an electromagnetic actuator by means of Dirichlet boundary condition iteration," *IEEE Trans. Magn.*, vol. 36, pp. 1110-1113, 2000.

[17] S. Alfonzetti, G. Borzì, E. Dilettoso, N. Salerno, "Stochastic optimization of a patch antenna", *ACES Journal*, vol. 23, no. 3, pp. 237-242, 2008.

[18] G. Aiello, S. Alfonzetti, G. Borzì, and N. Salerno, "An overview of the ELFIN code for finite element research in electrical engineering", in *Software for Electrical Engineering Analysis and Design*, A Konrad & C. A. Brebbia (ed.), WIT Press, Southampton (U.K.), 1999.

# Conclusions

In this thesis we have shown that the FEM-RBCI method can be successfully applied to the analysis of the scattering of time-harmonic electromagnetic waves at optical frequencies from metallic nanoparticles of arbitrary shape.
Moreover, the optimization of a thin solar cell with metallic nanoparticles has been performed by means of Genetic Algorithms and the Finite Element Method. The goal was to design a solar cell which shows good performances in terms of sunlight absorption.
Suitable Genetic Algorithms with varying crossover and mutation probabilities have been employed. The optimum was reached in about half the time required by the standard procedure. The optimized solar cell performs well in the sunlight frequency bandwidth.
The scientific publications connected to the research activity of this thesis are listed in the following.


[1] G. Aiello, S. Alfonzetti, V. Brancaforte, V. Chiarello, N. Salerno, "Applying FEM-RBCI to the Analysis of Plasmons in Metallic Nanoparticles", *International Journal of Applied Electromagnetics and Mechanics (IJAEM),* vol. 39, n. 1-4, p. 13-20, 2012.

[2] G. Aiello, S. Alfonzetti, V. Brancaforte, V. Chiarello, N. Salerno, "Applying FEM-RBCI to the Analysis of Plasmons in Metallic Nanoparticles", *15th International Symposium on Applied Electromagnetics and Mechanics (ISEM)* , Napoli, Italy, Sept. 6-9, 2011.

[3] G. Aiello, S. Alfonzetti, G. Borzì, V. Chiarello, N. Salerno, "Plasmon Analysis of Systems of Metallic Nanorings by means of FEM-RBCI", *16th IEEE Mediterranean Electrotechnical Conference (MELECON),* Medina Yasmine Hammamet, Tunisia, March, 25-28, 2012.

[4] G. Aiello, S. Alfonzetti, V. Chiarello, N. Salerno "Solar Cell Optimization by means of Metallic Nanodisks", *17th International Symposium on Theoretical Electrical Engineering (ISTET)*, Pilsen, Czech Republic, June 24-26, 2013.

[5] G. Aiello, S. Alfonzetti, G. Borzì, V. Chiarello, M. Cimino, N. Salerno, "Optimization of a Thin Film Solar Cell with Metallic Nanoparticles", *19th Conference on the Computation of Electromagnetic Fields (COMPUMAG),* Budapest, Hungary, June 30 – July 4, 2013.

[6] G. Aiello, S. Alfonzetti, V. Chiarello, N. Salerno, "Analisi di risonanze plasmoniche in nanoparticelle metalliche mediante il metodo FEM-RBCI", *XXVIII Riunione Annuale dei Ricercatori di Elettrotecnica (ET2012),* Catania, Italy, June 20-22, 2012.

[7] G. Aiello, S. Alfonzetti, G. Borzì. V. Chiarello, N. Salerno, "Ottimizzazione di una cella solare a film sottile mediante nanoparticelle metalliche", *XXIX Riunione Annuale dei Ricercatori di Elettrotecnica (ET2013),* Padua, Italy, June 19-21, 2013.