

Birefringency: Calculation of Refracted Ray Paths in Biaxial Crystals

Pedro Latorre · Francisco J. Seron · Diego Gutierrez

Received: date / Accepted: date

Abstract The phenomena of birefringency may be observed when light arrives at an anisotropic crystal surface and refracts through it, causing the incident light ray to split into two rays; these become polarized in mutually orthogonal directions, and two images are formed. The principal goal of this paper is the study of the directional issues involved in the behavior of light when refracting through a homogeneous, non participating medium, including both isotropic and anisotropic media (uniaxial and, for the first time, biaxial). The paper focuses on formulating and solving the non-linear algebraic system that is obtained when the refraction process is simulated using the geometric model of Huygens. The main contribution focuses on the case of biaxial media. In the case of uniaxial media, we rely on symbolic calculus techniques to formulate and solve the problem.

This research has been funded by the Spanish Ministry of Science and Technology (project TIN2007-63025) and the Aragón Government (projects OTRI 2009/0411 and CTPP05/09).

Pedro Latorre
Universidad de Zaragoza. María de Luna, 1. 50018 Zaragoza
Spain
Tel.: +34-976762355
Fax: +34-976761914
E-mail: platorre@unizar.es

Francisco J. Seron
Universidad de Zaragoza. María de Luna, 1. 50018 Zaragoza
Spain
Tel.: +34-976761939
Fax: +34-976761914
E-mail: seron@unizar.es

Diego Gutierrez
Universidad de Zaragoza. María de Luna, 1. 50018 Zaragoza
Spain
Tel.: +34-976762354
Fax: +34-976761914
E-mail: diegog@unizar.es

Keywords Computer Graphics · Three-dimensional Graphics and Realism · Simulation of Physical Phenomena · Optics · Physically Based Rendering

1 Introduction

The phenomena of birefringency may be observed when light arrives at an anisotropic crystal surface and refracts through it. Using the ray model for light propagation, birefringency causes two main effects: first, the incident light ray splits into two rays that go through the crystal at two different velocities and with two different directions, and which, with the exception of some specific cases, form two images (see Figure 1); and second, those two rays become polarized in mutually orthogonal directions, even if the incident beam is incoherent. The velocities, propagation directions and polarization -and consequently the distribution of energy between both rays- depend on the optical properties of the crystal and on the propagation direction of the incident ray. Birefringency also depends on wavelength, and in some cases, may cause some color changes that depend on the crystal's orientation (pleochroism or polychroism) [3].

The non-linear nature of the problem leads to a number of challenging issues when it comes to efficiently and robustly solving this problem. Our paper proposes a novel method to solve this biaxial case, describing the optical properties of crystals and the phenomena of birefringency. We also present a set of original methods based on the geometrical construction of the Huygens principle [18] to calculate the solution to ray directions that go through isotropic, uniaxial and biaxial crystals.

The solution for the biaxial media case does not have a closed form. Solving the system has only been

possible by using numerical techniques, which have unexpectedly unveiled a surprisingly rich space of numerical possibilities. For the solution of the non-linear systems a damped Newton method and continuation techniques are used. With these a lookup table is obtained for each of the crystal's surfaces, given their principal refraction indices and the orientations of the interface planes. Each table contains a dense set of directions of refracted rays as a function of the direction of the incident ray. The initial guess for the iterations is obtained from the previous solutions by way of double-pass sampling. Given a direction of incidence, the path of the refracted rays is obtained from the four closest directions in the table.

Finally, we show some images describing our method, and then enumerate the problems that still remain open. We do not attempt to simulate all real materials showing birefringency, but show how to simulate the effect from a more academic perspective instead; however, the results shown have been obtained with a range of indices of refraction similar to those found in real materials such as calcite.



Fig. 1 Light rays split into two by a calcite crystal and forming two images.

2 Previous work

Anisotropic crystals, and particularly birefringency, have received little attention in computer graphics. They have been more extensively studied in the physical optics field [3] [13], although always in a very limited manner; solutions only exist for uniaxial crystals or, in the biaxial case, for very specific directions applicable for instance to polarizers.

In [16], the authors describe methods for performing polarization ray tracing through birefringent media, although they leave the more difficult biaxial case

out. The main contribution, made by Tannenbaum et al. [19], discussed coherency and polarization and presented a matrix based formulation¹. They solved the propagation through an uniaxial crystal and established a set of formulas for calculating those uniaxial propagation directions. Another important contribution comes from Guy et al. [10]. They present an algorithm for rendering faceted colored gemstones in real time, using graphics hardware. Their solution is based on several controlled approximations of the physical phenomena involved when light enters a stone, which allows an implementation based on recent hardware developments. Most crystal-structured materials are optically anisotropic. As in the previously mentioned papers, this work by Guy et al. solves light propagation only through uniaxial crystals, leaving the biaxial case as future work.

More recently, Weidlich and Wilkie [21] derive the complete set of formulas needed to generate physically plausible images of uniaxial crystals. The work contains the complete derivation of the Fresnel coefficients for birefringent transparent materials, as well as for the direction cosines of the extraordinary ray and the Muller matrices necessary to describe polarization effects. This allows computing the interaction of light with such crystals in a form that is useable by graphics applications, especially if a polarization-aware rendering system is being used. However, as the authors state, extending this work to biaxial materials becomes almost intractable, given that the mathematical simplifications used for the derivation of the formulas are no longer applicable. Whilst only a few biaxial crystals exhibit macroscopically noticeable birefringency effects, it is still a challenge to simulate and trace the second extraordinary ray, using the geometric model of Huygens. As we will show, this actually unveils a surprising variety of numerical situations, which were unexpected in advance.

3 Theoretical background

For the sake of clarity, we introduce here the basics of Huygens principle, which will be used and explained in more detail in successive sections, as well as a formal description of the birefringency effect. A complete table with all the necessary definitions can be seen in table 1.

¹ This citation refers to both the two-page version that appeared in the proceedings of SIGGRAPH 94 and the extended version, which appeared on the CD but not in the proceedings

Symbol	Definition
ε	dielectric permittivity
n	refraction index
$\mathbf{v}_c$	velocity of light in vacuum
v_x, v_y, v_z	velocity of wave surface
$\mathbf{v}_o$	velocity of ordinary ray (uniaxial)
$\mathbf{v}_e$	velocity of extraordinary ray (uniaxial)
θ_i, θ_s	angles of incident and refracted rays
(x, y, z)	point on the wave surface
$\mathbf{r}_i$	incident ray direction
$\mathbf{r}_s$	normal of plane of incidence
$\mathbf{r}_o$	ordinary ray direction (uniaxial)
$\mathbf{r}_e$	extraordinary ray direction (uniaxial)
$\mathbf{r}_1, \mathbf{r}_2$	extraordinary ray directions (biaxial)
O, B	points of the wave train located on the plane of incidence
P	equation of plane tangent to a surface
F	equation of the wave surface
T	point of tangency

Table 1 Table of symbols

3.1 Basics of the Huygens principle

Huygens principle can be used to describe effects of wave propagation such as refraction and diffraction. It establishes that each point of an advancing wavefront produces a new disturbance that becomes the source of a new train of waves; so, the whole advancing wave will form a new wavefront formed by the sum of all these secondary waves.

Let us assume a plane wavefront traveling through an isotropic medium, where its speed is perpendicular to this wavefront. When it arrives to an interface with a different medium (see Figure 2), the first point in the wavefront hitting such interface (point O in Figure 2) begins to vibrate at a different speed, transmitting its vibration to the neighboring points in the second medium. The points from O to B will be reached by the wavefront in successive instants t_i ; therefore, the waves originated at points belonging to OB will also produce vibrations in the second medium in successive instants t_i . So finally the new plane wavefront (assuming that the second medium is also isotropic) is formed. Both speed vectors and the normal to the interface plane of separation lay on the same plane.

When the second medium is anisotropic, the wavefront formed by every oscillating point is not a sphere but a two-folded surface as it will be explained later. This means that the wave coming from the first media is split in two different ones, each of them traveling with different speed and direction. Note that in this case speed and normal do not lay on the same plane, depending on the crystal type and orientation instead.

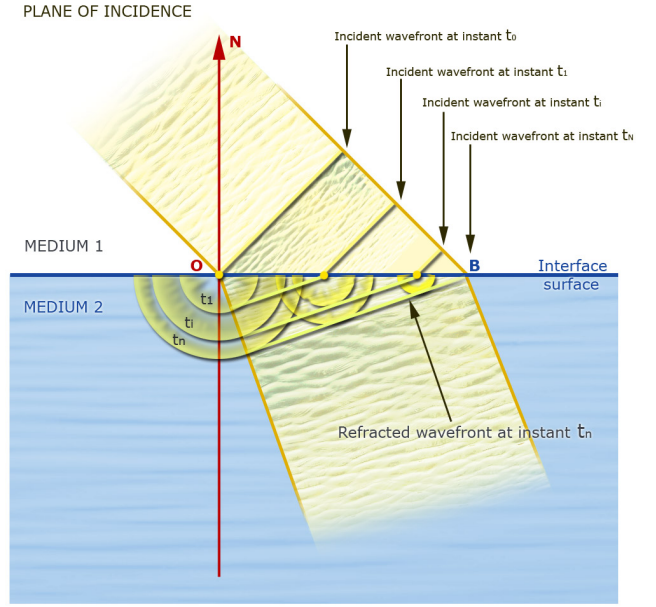


Fig. 2 Explanation of the Huygens principle.

3.2 Refraction on crystals. Birefringency.

Crystalline structures are characterized by the ordered arrangement of atoms in a basic cell that repeats in a three dimensional lattice [9]. This specific arrangement is responsible for the crystal's optical properties, with the inherent symmetries determining its optical isotropy or anisotropy (see Table 2). The lattice determines a reference system that follows the directions of three intersecting edges, the relative dimensions of which are a , b and c . The relative orientation of axes are given by three angles, α , β and γ .

The behavior of light in a medium may be established by using Maxwell's laws for electromagnetic fields [17]. When light arrives onto the surface separating two dielectric media, part of the energy reflects onto the first one, and the remaining energy either goes through the second one or is absorbed. From the point of view of electromagnetism, anisotropy is explained by the directional dependence of the dielectric permittivity tensor ε , which is a simple scalar constant in isotropic media. It may be stated that there is one system of reference in which the permittivity tensor is diagonal; the three mutually orthogonal axes of this orientation are called *principal axes* of the crystal. For simplicity, this orientation will be used throughout this paper. If two of the components of ε are equal, the crystal is *uniaxial*; if $\varepsilon_x \neq \varepsilon_y \neq \varepsilon_z$, the crystal is *biaxial* (see table 2).

The anisotropic propagation may be modeled using the crystal's *wave surface*. This is defined as the envelope of all the points hit by the wavefront at a given

System	Cell shape	Symmetry	Type of anisotropy
Cubic	Cube ($a = b = c$, $\alpha = \beta = \gamma = 90^\circ$)	Four three-fold axis	Optically isotropic
Trigonal	Rhombohedron ($a = b = c$, $\alpha = \beta = \gamma < 120^\circ \neq 90^\circ$) or right prism with 60° rhombic base ($a = b \neq c$, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$)	One three-fold axis	Uniaxial. Dielectric revolution axis parallel to principal symmetry axis
Tetragonal	Square base right prism ($a = b \neq c$, $\alpha = \beta = \gamma = 90^\circ$)	One four-fold axis	
Hexagonal	Right prism with 60° rhombic base ($a = b \neq c$, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$)	One six-fold axis	
Orthorhombic	Rectangular parallelepiped ($a \neq b \neq c$, $\alpha = \beta = \gamma = 90^\circ$)	Three mutually orthogonal two-fold axis	Biaxial. Dielectric principal axis parallel to crystallographic axis
Monoclinic	Oblique parallelepiped ($a \neq b \neq c$, $\alpha = \gamma = 90^\circ$, $\beta \neq 90^\circ$)	One two-fold axis	Biaxial. One principal axis parallel to 2-fold axis; orientation not determined for the other two
Triclinic	Parallelepiped ($a \neq b \neq c$, $\alpha \neq \beta \neq \gamma \neq 90^\circ$)	None	Biaxial. No dielectric - crystallographic axis relationship

Table 2 Crystallographic systems, unit cell shape, type of anisotropy and dielectric - crystallographic axis orientation.

time. In this case, the dimensions of the principal axes of the wave surface coincide with *the principal velocities* [3].

$$v_x = v_c / \sqrt{\varepsilon_x}; v_y = v_c / \sqrt{\varepsilon_y}; v_z = v_c / \sqrt{\varepsilon_z} \quad (1)$$

where v_c is the velocity of light in vacuum.

In an isotropic medium, the wave surface is a sphere; in anisotropic media, a two-fold surface. In biaxial media, the wave surface equations are fourth-order polynomials with even powers only; that is, the surface is symmetrical with respect to the origin. Both surface folds intersect only at four symmetrical points, as can be seen in Figure 3. Note that this intersection does not yield a curve, but only the four points.

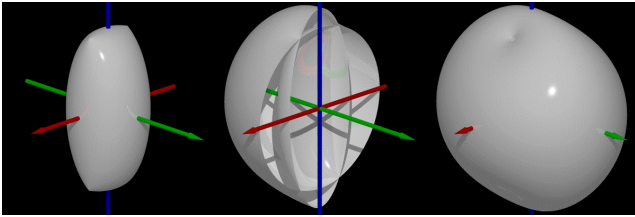


Fig. 3 Wave surface for a biaxial medium. Inner fold (left), vertical section (center) and outer fold (right)

The wave surface equation is [3]:

$$r^2(v_x^2x^2 + v_y^2y^2 + v_z^2z^2) - [v_x^2(v_y^2 + v_z^2)x^2 + v_y^2(v_z^2 + v_x^2)y^2 + v_z^2(v_x^2 + v_y^2)z^2] + v_x^2v_y^2v_z^2 = 0 \quad (2)$$

where $r^2 = x^2 + y^2 + z^2$.

In uniaxial crystals, two of the velocities are equal. The equation may be factorized, resulting in

$$(r^2 - v_0^2)[v_0^2(x^2 + y^2) + v_z^2z^2 - v_0^2v_z^2] = 0 \quad (3)$$

where $v_o = v_x = v_y$.

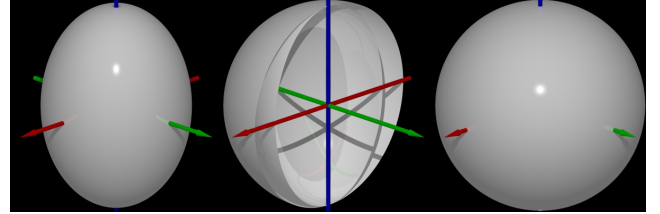


Fig. 4 Positive uniaxial medium wave surface. Inner fold (left), vertical section (center) and outer fold (right)

The first factor corresponds to a spherical fold, and the second to an ellipsoid. One of its principal sections is a circle that coincides with the intersection of both folds. If the spherical fold is located inside the ellipsoid, the uniaxial crystal is called *negative*, and *positive* if the ellipsoid is located inside the sphere. (see Figures 4 and 5). In all equations, (x, y, z) represents a surface point. For a complete and detailed description of all the phenomena involved, we refer the reader to the work by Born and Wolf [3], chapter XV (Optics of Crystals). The authors also describe some applications of anisotropic crystals in optics, such as Nicol prisms and compensators.

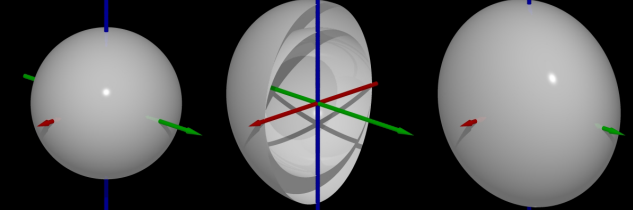


Fig. 5 Negative uniaxial medium wave surface. Inner fold (left), vertical section (center) and outer fold (right)

4 Calculating propagation directions

Our work focuses on formulating and solving the non-linear algebraic system that describes the refraction process for anisotropic media, as explained by the model of Huygens. We will present for the first time a solution for the biaxial case, relying on numerical techniques.

In order to make a complete and clear description of the method, a brief consideration on the isotropic - isotropic media refraction will be presented first. The paper presents and solves the non-linear algebraic system obtained. In the case of uniaxial media the solution has been obtained following symbolic calculus techniques. This is the problem already solved by [19], but it is included here for the sake of clarity and completeness of this paper. Finally, our main contribution, which is the general isotropic - biaxial anisotropic refraction will be introduced. The solution of the resulting system has only been possible by using numerical techniques (damped Newton methods), coupled with pre-computed lookup tables to efficiently find initial values. The numerical techniques have unexpectedly unveiled a surprisingly rich space of numerical possibilities.

One major difference between light propagation in isotropic and anisotropic crystals is that energy and wave propagation directions are not coincident (with the exception of some specific directions). Instead, they depend on the incident ray direction and on the orientation of the crystal's surface with respect to the crystal's principal axis. In this paper we obtain the *ray velocities*, from which wave normals can easily be calculated.

Our method consists of obtaining a lookup table for each crystal surface, given its principal refraction indices and interface plane orientations. Each table contains a dense set of directions of refracted rays as a function of direction of the incident ray, calculated using a damped Newton method and continuation techniques. The initial guess for the iterations is obtained from previous solutions, with double-pass sampling. Given a direction of incidence, the path of refracted rays is calculated by bilinear interpolation from the four closest directions in the table.

4.1 Isotropic - isotropic refraction

When an incident plane wave (see Figure 6) train arrives at a point O on the separation surface between two media at a given time t , this point begins to vibrate, transmitting its vibration to the neighboring points in the second medium. The points from O to B will be hit by the wave in successive instants. After a certain time lapse, the point A_1 in the original wavefront will hit point B . The distance OB may be calculated from $d(OB) = v_1/\sin\theta_i$, where $v_1 = 1/n_1$ is the transmission velocity of the incident ray, the inverse of the refraction index n_1 .

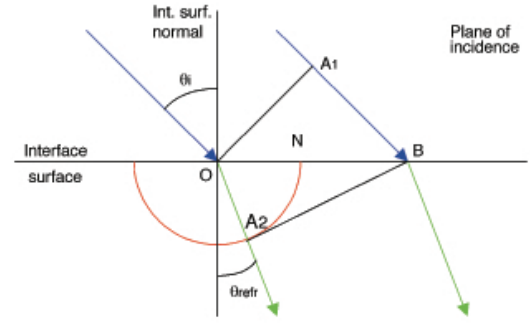


Fig. 6 Geometrical method for calculating propagation direction: Isotropic - isotropic refraction. OA_1 and A_2B represent the wavefronts before and after refraction.

At that time (say $t = 1$ sec.) the swept surface of each secondary emitter will be a hemisphere of radius v_2 , its wave surface. For the same reason, the waves originated in points from O to B will produce hemispheric surfaces with decreasing radius, from v_2 to 0. The resulting total wavefront is again the swept surface of all hemispheres, which results in a plane, the orientation of which will be normal to the refracted rays.

The refracted ray may be calculated from Figure 6. According to the Snell's law of refraction:

$$\frac{\sin \theta_i}{\sin \theta_{refr}} = \frac{v_1}{v_2} = \frac{n_2}{n_1} \quad (4)$$

where θ_i and θ_{refr} are the angles that incident and refracted rays form, respectively, with regard to the surface normal. v_1 , v_2 , n_1 and n_2 are, respectively, the incident and refracted ray velocities and refraction indices. Incident, reflected and refracted rays, as well as the surface normal, all lay in the same plane.

4.2 Isotropic - uniaxial anisotropic refraction

The second medium's wave surface now has two folds, so the incident plane wave will split into two refracted

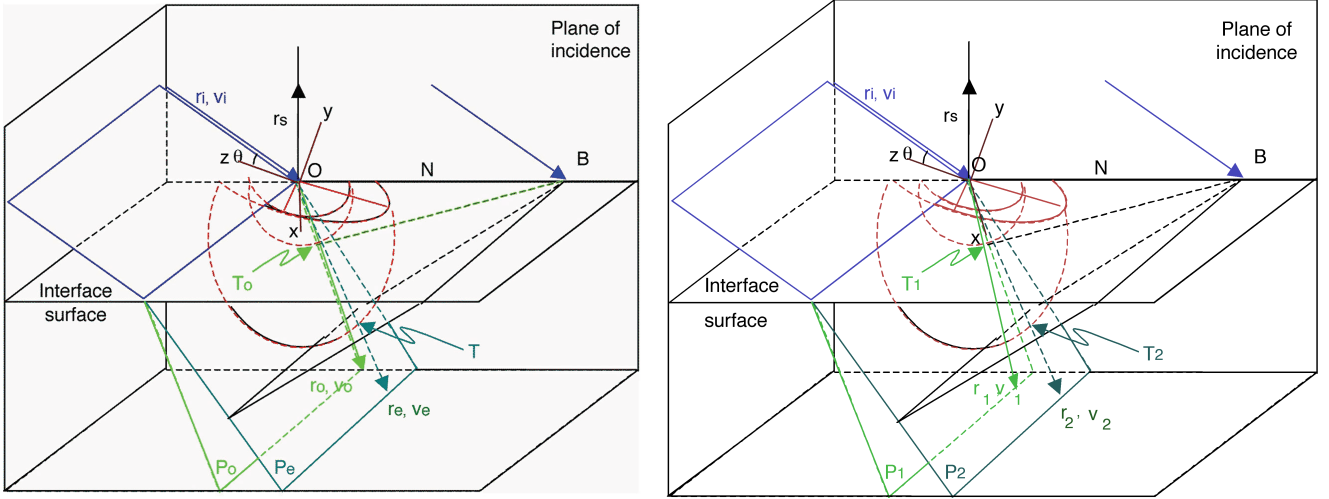


Fig. 7 Geometrical method for calculating propagation directions. Left: Isotropic - uniaxial anisotropic refraction. The incident ray r_i and the ordinary ray r_o belong to the plane of incidence. Right: Isotropic - biaxial anisotropic refraction. Note that in this case the refracted rays r_1, r_2 does not belong to the plane of incidence. For both cases, the separation between the two folds has been exaggerated. (after [19]).

plane waves (see Figure 7, left). The equations of these planes may be calculated by applying the conditions of field continuity to Maxwell laws.

Directions and velocities of refracted rays now depend on the following data:

- The incident ray direction $\mathbf{r}_i(x_i, y_i, z_i)$ and velocity $v_i = 1/n_0$. The first medium is supposed to be isotropic.
- The interface surface orientation, which is supposed to be a plane (specified by its normal $\mathbf{r}_s(x_s, y_s, z_s)$) and the incidence plane normal $\mathbf{r}_i'(x_i', y_i', z_i')$ (calculated as $\mathbf{r}_i' = \mathbf{r}_i \times \mathbf{r}_s$).
- The principal wave velocities of the second medium v_o and v_z .

The orientation of the second medium's wave surface is fixed. Its center coincides with the coordinates origin, and the largest principal axis is made to coincide with the OZ axis (see Figure 7, left). Different orientations may be reduced to this one by applying the appropriate rotations to all vectors.

Following a method similar to the isotropic - isotropic case, the incident plane wave train arriving at point O on the separation surface in a given time t will hit point B $t + \Delta t$ later. The direction and velocity of propagation of the ray which corresponds to the spherical fold (*ordinary ray*, $\mathbf{v}_o$) can be calculated directly using Snell's law. The remaining ray, (*extraordinary ray*, $\mathbf{v}_e$) can be calculated from the ellipsoidal fold ellipsoidal fold as follows:

1. **Determination of $B(x_B, y_B, z_B)$:** point B may be calculated from the three following conditions:

- B is located on the incidence plane
- B is located on the interface plane (the separation plane between the two media)
- $d(OB) = v_i / \sin \theta_i$

2. **Determination of the tangent plane to the ellipsoidal fold:** The equation of a plane tangent to a surface is:

$$P \equiv \left(\frac{\partial F}{\partial x} \right)_T (x - x_T) + \left(\frac{\partial F}{\partial y} \right)_T (y - y_T) + \left(\frac{\partial F}{\partial z} \right)_T (z - z_T) = 0 \quad (5)$$

where $F \equiv x^2/v_z^2 + y^2/v_z^2 + z^2/v_0^2 = 1$ is the equation of the wave surface's elliptical fold, and $T(x_T, y_T, z_T)$ is the point of tangency.

3. **System description:** Planes which
 - are perpendicular to the incidence plane
 - are tangent to the wave surface
 - contain point B

are the refracted plane waves; the vectors OT_o and OT determine the velocities of both rays.

Formulating the system results in:

$$\frac{x_T x_i'}{v_z^2} + \frac{y_T y_i'}{v_z^2} + \frac{z_T z_i'}{v_0^2} = 0 \quad (6)$$

$$\frac{x_T^2}{v_z^2} + \frac{y_T^2}{v_z^2} + \frac{z_T^2}{v_0^2} = 1 \quad (7)$$

$$\frac{x_T x_B}{v_z^2} + \frac{y_T y_B}{v_z^2} + \frac{z_T z_B}{v_0^2} - 1 = 0 \quad (8)$$

where the preceding calculations have been taken into account. As opposed to the rest of cases presented in this paper, this case can be solved analytically (see Appendix A).

4. **Determination of the velocity of the extraordinary ray:** Once the tangency point has been calculated, the extraordinary ray velocity is:

$$v_e = \sqrt{x_T^2 + y_T^2 + z_T^2} \quad (9)$$

The direction of propagation may be calculated by normalizing the $\mathbf{v}_e$ vector.

The method shown above cannot be applied if the direction of incidence coincides with the direction normal to the interface plane. In this particular case, the planes needed to obtain the refracted plane waves must obey the following conditions:

- The planes are parallel to the interface plane (and perpendicular to the direction of incidence)
- The planes are tangent to the wave surface
- The planes contain point B

Solutions may be easily derived from these conditions, resulting in:

$$x_T = \frac{1}{2}\lambda v_z^2 x_i \quad y_T = \frac{1}{2}\lambda v_z^2 y_i \quad z_T = \frac{1}{2}\lambda v_0^2 z_i$$

$$\frac{x_T^2}{v_z^2} + \frac{y_T^2}{v_z^2} + \frac{z_T^2}{v_0^2} = 1 \quad (10)$$

for the extraordinary ray, where λ can take any value. From these, it easily follows:

$$x_T = \frac{v_z^2 x_i}{\sqrt{v_z^2(x_i^2 + y_i^2) + v_0^2 z_i^2}}$$

$$y_T = \frac{v_z^2 y_i}{\sqrt{v_z^2(x_i^2 + y_i^2) + v_0^2 z_i^2}}$$

$$z_T = \frac{v_0^2 z_i}{\sqrt{v_z^2(x_i^2 + y_i^2) + v_0^2 z_i^2}} \quad (11)$$

The ordinary ray follows the direction of incidence.

4.3 Isotropic - biaxial anisotropic refraction

The second medium wave surface again has two folds, so the incident plane wave will split into two refracted plane waves in a similar way to isotropic - uniaxial refraction. The equations of these planes may again be calculated by applying the conditions of field continuity to Maxwell laws (see Figure 7, right). Figure 3 shows the wave surface for this case; its intersection with the interface surface is different and the refracted rays r_1 and r_2 do not belong to the plane of incidence.

We will follow a similar reasoning as for the isotropic - uniaxial case, although the equations obtained in this case are quite different. The incident plane wave train arriving onto point O on the separation surface at an instant t will hit the point B one second later. The planes which

- are perpendicular to the plane of incidence
- are tangent to the wave surface
- contain point B

are the refracted plane waves; the vectors OT_1 and OT_2 determine the velocity of both rays.

The wave equation cannot be factorized as in the previous case, so the calculation of both refracted rays cannot be separated. The system to be solved (again using Maple), after factorizing and ordering under decreasing powers is shown in Figure 8, and it is the equivalent to the system defined by Equations 6, 7 and 8.

Directions and velocities of refracted rays, as well as the calculation of point B , are derived from the same data as the isotropic - uniaxial case. The orientation of the second medium's wave surface also coincides with the prior case, with the principal wave velocities now being v_x , v_y and v_z . Different orientations may be reduced to this one by applying the adequate rotations to all vectors.

$$\begin{aligned} & (-2v_x^2 - 2v_y^2)x_T^2 y_T^2 - 2v_x^2 x_T^4 + (-2v_z^2 - 2v_x^2)x_T^2 z_T^2 - 2v_y^2 y_T^4 \\ & + (-2v_y^2 - 2v_z^2)y_T^2 z_T^2 - 2v_z^2 z_T^4 + 2v_x^2 x_B x_T^3 + (v_x^2 y_B + v_y^2 y_B)x_T^2 y_T \\ & + (v_z^2 z_B + v_x^2 z_B)x_T^2 z_T + (v_y^2 x_B + v_x^2 x_B)x_T y_T^2 + (v_x^2 x_B + v_z^2 z_B)x_T z_T^2 \\ & + 2v_y^2 y_B y_T^3 + (v_z^2 z_B + v_y^2 z_B)y_T^2 z_T + (v_z^2 y_B + v_y^2 y_B)y_T z_T^2 + 2v_z^2 z_B z_T^3 \\ & + (v_x^2 v_y^2 + v_x^2 v_z^2)x_T^2 + (v_x^2 v_y^2 + v_y^2 v_z^2)y_T^2 + (v_y^2 v_z^2 + v_x^2 v_z^2)z_T^2 \\ & + (-v_x^2 v_y^2 x_B - v_x^2 v_z^2 z_B)x_T + (-v_y^2 v_x^2 y_B - v_y^2 v_z^2 z_B)y_T \\ & + (-v_z^2 v_x^2 z_B - v_z^2 v_y^2 z_B)z_T = 0 \\ & v_x^2 x_T^4 + (v_x^2 + v_y^2)x_T^2 y_T^2 + (v_z^2 + v_x^2)x_T^2 z_T^2 + v_y^2 y_T^4 + (v_y^2 + v_z^2)y_T^2 z_T^2 \\ & + v_z^2 z_T^4 + (-v_x^2 v_y^2 - v_x^2 v_z^2)x_T^2 + (-v_y^2 v_z^2 - v_x^2 v_y^2)y_T^2 \\ & + (-v_y^2 v_z^2 - v_x^2 v_z^2)z_T^2 + v_x^2 v_y^2 v_z^2 = 0 \\ & 4x'_i v_x^2 x_T^3 + (2y'_i v_y^2 + 2y'_i v_x^2)x_T^2 y_T + (z'_i v_z^2 + z'_i v_x^2)x_T^2 z_T \\ & + (2x'_i v_y^2 + 2x'_i v_x^2)x_T y_T^2 + (2x'_i v_z^2 + 2x'_i v_x^2)x_T z_T^2 + 4y'_i v_y^2 y_T^3 \\ & + (z'_i v_z^2 + z'_i v_y^2)y_T^2 z_T + (2y'_i v_z^2 + 2y'_i v_y^2)y_T z_T^2 + 2z'_i v_z^2 z_T^3 \\ & + (-2x'_i v_x^2 v_y^2 - 2x'_i v_x^2 v_z^2)x_T + (-2y'_i v_y^2 v_x^2 - 2y'_i v_y^2 v_z^2)y_T \\ & + (-z'_i v_z^2 v_y^2 - z'_i v_z^2 v_x^2)z_T = 0 \end{aligned}$$

Fig. 8 System to be solved for the isotropic - biaxial anisotropic refraction. Note that it is equivalent to the system defined by equations 6, 7 and 8.

4.4 Obtaining ray velocities

No analytical solution has yet been found for the system shown in Figure 8. Previous works could only provide an analytical solution for the uniaxial problem, leaving the biaxial problem as unsolved future work. These facts lead us to consider necessary the use of a numerical method to solve the system for the biaxial case. Given

also the strong non-linearity of the problem, the most common numerical methods will fail to converge to a solution.

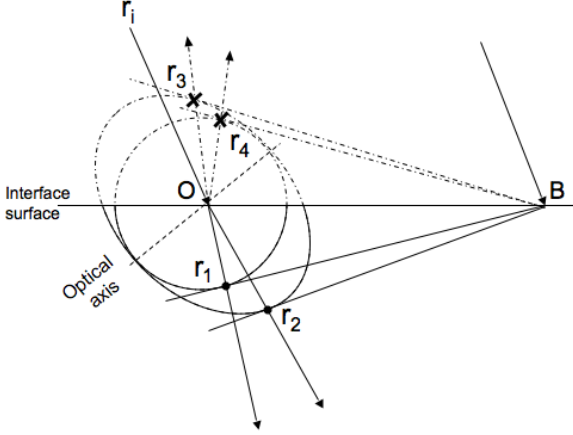


Fig. 9 The four solutions obtained in uniaxial and biaxial cases. Light arrives to interface surface with direction r_i . Directions r_1 and r_2 corresponds to physically correct solutions for refracted rays, while r_3 and r_4 corresponds to solutions back to first media and must be rejected.

A first look at the geometric model (Figure 7, right) shows that, except for one particular direction, four solutions must be expected for each direction of incidence. Two of them are located in the first medium and must be rejected; the other two represent the two tangent points, one on each fold of the wave surface. Figure 9 sketches in a simpler way the election of the two solutions corresponding to the real refracted rays.

Let $F(x) = 0$ be the system to solve, and x the vector of solutions. Iterative numerical methods for nonlinear systems solving [20] [15] have a particularly favorable convergence rate, but for many systems this rapid convergence is only achieved if the initial iterate is chosen from a small subset of the region of attraction. This subset is generally located close to the desired solution x . Thus, in connection with the practical solution of the system, it is most crucial to find an appropriate initial value $x^{(0)}$; this is especially true in this case, because of the proximity of both solutions.

To give an idea of the problem's complexity, given a method (e.g. Newton), a set of data (light velocity in the first medium, the three principal velocities in the second medium and the normal direction to interface plane plus direction of incidence) and an initial solution guess, one of the following cases may occur:

1. The correct solution on the desired fold is found
2. A correct solution on the other fold is found, and must be discarded

3. A geometrically correct but physically impossible solution is found, and must be discarded
4. The iteration diverges, and no solution is found
5. Moreover, since floating point numbers do not constitute a continuum, the iteration may fall within a stationary or periodic sequence before achieving the required approximation to the solution, and fail.

Since the method is deterministic, the only way to (try to) find a solution is to repeat the whole iteration with a different initial guess.

Using Newton's method for solving the system and different initial value techniques, a comprehensive set of attempts has been carried out. The goal was to find a suitable seed which could be derived from data for each given direction of incidence (using Snell's law of refraction to calculate the three theoretical directions of refraction corresponding to isotropic media with index refraction values equal to the three principal biaxial indices). Different linear combinations of these three theoretical directions of refraction were tested as seeds, resulting in failures in different regions.

These attempts suggest that all the methods for calculating the paths of refracted rays for a given set of data -following an independent way- fail. If no adequate initial approximation for starting an iterative method is available, then a homotopy or continuation method often leads to a sufficiently accurate approximation. For that purpose a parameter inherent to the problem (in our case the direction of incidence) is used to transform a non-linear system of n equations into a family of problems $H(\mathbf{r}, w) = 0$, whose solutions $\mathbf{r}^*(w)$ depend continuously on w under certain conditions [5]. In this case the problem has a one-dimensional manifold (a space curve) as its set of solutions.

In order to obtain a discrete solution, the problem can be subdivided in a sequence of non-linear equations $H(\mathbf{r}, w_k) = 0$ with $k = 1, 2, \dots, K$ and can be solved sequentially, where the solution $\mathbf{r}^{(k)}$ of the k th system of non-linear equations can be used as the initial iterate for solving the $(k+1)$ th system. Then $\mathbf{r}^{(k)}$ is a suitable initial approximation for the iterative determination of $\mathbf{r}^{(k+1)}$.

Therefore, the next logical step is to build -in an orderly way- a lookup table containing a dense enough set of solutions corresponding to a uniformly spaced set of possible directions of incidence for each given set of data. Every preceding solution is used as the initial guess for calculating each refracted ray path for the next direction of incidence; this ensures the closeness of initial guesses. The initial guess for each series of directions of incidence is obtained from the solution corresponding to the normal incidence.

A method for visualizing the results has also been developed (see *Visualization techniques* in Section 5.1). One of these images can be seen in Figure 10, showing mathematically correct solutions (in red) and periodic sequence failures (in green). Regions where no correct solutions were found appear in grey (the wave surface fold).

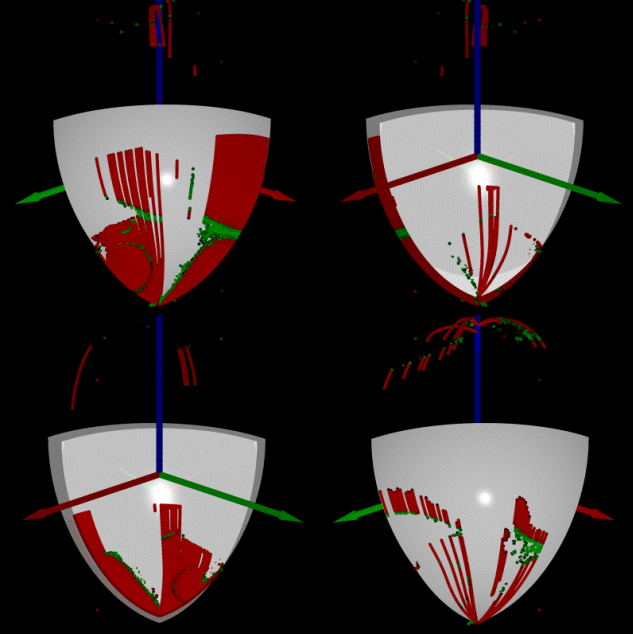


Fig. 10 Distribution of solutions using the Newton method. Upper left, solutions on outer fold coming from initial guesses on inner fold. Upper Right, solutions on inner fold coming from initial guesses on outer fold. Lower left, solutions on inner fold coming from initial guesses on inner fold. Lower right, solutions on outer fold coming from initial guesses on outer fold. Observe the solutions misplaced on the upper hemisphere, which correspond to the first (isotropic) medium.

Subsequent attempts have been carried out to try to build these tables with Newton's method and other different methods for traversing and filling in the table, resulting in better but still incomplete results. Better results were obtained using a double - pass method, filling in the table from the zenith to low angles of incidence and vice versa for each given azimuthal direction (Figure 11). Initial guesses are calculated using the values provided by Snell's law of refraction with minimum and maximum principal values of v ; this choice ensures some proximity to the inner and outer fold solutions, and yields good results for the regions chosen to begin each series: those close to normal incidence (zenith) and those with low angles of incidence.

The Damped Newton method has been used to solve the remaining problems. We choose this method since it provides the best compromise between reliability (con-

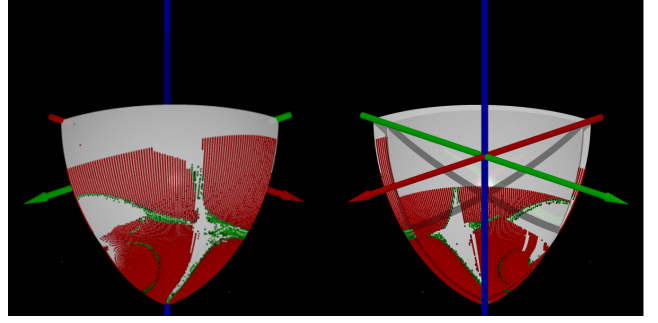


Fig. 11 Distribution of solutions using the Newton method with double - pass initial guess determination.

vergence to the desired solution) and a relative low increase in programming complexity. We found that other variants of the conventional Newton method that focus on efficiency (such as the Modified, or the Quasi-Newton methods) did not compensate the additional programming burden, whilst the Relaxed Newton method does not help near critical points [8]. The Damped Newton Algorithm is central to many computer problems, such as vision (structure from motion, illumination based reconstruction, non-rigid model tracking) [4], image restoration [22], simulation [11], fractal modeling [7] or even computational economics [6]. This method extends the region of attraction of the zeros, which are very small in some regions when using the normal Newton method. As opposed to the usual Newton step, the iteration:

$$x^{(k+1)} := x^{(k)} + \lambda_k \Delta x^{(k)}, \quad 0 < \lambda_k < 1, \quad (12)$$

is executed, where the damping factors λ_k are chosen in such a way that:

$$\|F(x^{(k+1)})\| < \|F(x^{(k)})\|, \quad k = 0, 1, 2, \dots, \quad (13)$$

for some norm $\|\cdot\|$. If $F'(x)$ is regular at $x^{(k)}$ then for a sufficiently small λ_k the inequality (13) must hold [20] [15] [14].

Therefore, the following steps must be carried out to implement the method: first, definition of a norm that allows to choose the damping factors λ_k in such a way that the convergence indicated in equation 13 is guaranteed; second, a strategy for calculation has to be designed; and third, each λ_k must be determined, taking into account the different cases in which problems may arise.

The method has been implemented as follows (for a full description of the implementation the reader can refer to Appendix B): given a set of data -principal crystal indices or velocities and interface plane orientation

with respect to the principal axes- the system is re-oriented in a way that the principal axes coincide with the axes of the coordinates system ($v_x < v_y < v_z$). The lookup table of solutions is calculated for every 1 deg to 0.5 deg (depending on the relative values of the principal refraction indices) in azimuth and zenith angular directions of incidence and using a double - pass filling method, similar to the one previously described for the Newton system solving method.

Given a particular orientation of the crystal and a direction of incidence, the appropriate transformation is applied to make orientation coincide with that of the table, which may be pre-calculated. The four double solutions corresponding to the nearest directions of incidence are found on the table, and precise solutions are calculated by bilinear interpolation; precision is guaranteed provided that table samples are near enough. Finally, an inverse transformation is applied to return the system to its original orientation, and the two refracted directions are used to calculate the path of the refracted rays. The algorithm scheme is shown in Figure 12, where **ExternalGuess** and **InternalGuess** are the seeds for the external and internal folds respectively. The calculation of the temporary table **ProvTable1** is similar to **ProvTable2**, except for the sampling direction for θ (0 deg to 89.5 deg) and the initial values in each series.

5 Results

A lookup table has been calculated and a visualization method for solutions has been developed in order to test the different methods used for calculation of solutions and initial guess choice (in the case of isotropic - biaxial refraction). Additionally, sets of images using similar parameters for the three possible cases (isotropic - isotropic, isotropic - uniaxial, isotropic biaxial) have been calculated.

5.1 Description of the tests

Given a crystal that is characterized by its three principal refraction indices, a table including the pair of solutions corresponding to each (sampled) direction of incidence is needed for each of the crystal's faces (symmetries may reduce the number of tables). Each table will contain $N \times M$ pairs of solutions, N and M being the number of samples for the zenith and azimuth angles of incident directions respectively. The calculation method has been explained in the previous section.

To test the process, a simple experiment has been designed as follows:

```

algorithm calculateLookUpTable( $v_x, v_y, v_z, \nabla\phi, \nabla\theta$ );
begin
  calculateProvTable1;
   $\phi := 0$ ;
  while  $\phi \leq 90$  do
     $\theta := 89.5$ ;
    calculate ExternalGuess and InternalGuess
    using Snell's refraction law;
    while  $\theta \geq 0$  do
      calculate solution1( $\phi, \theta$ , ExternalGuess)
      using damped Newton;
      if solution1 exists then
        place solution1 in right octant;
        if solution1 is external then
          ExternalSol:= solution1
        else InternalSol:= solution1
        endif
      endif;
      calculate solution2( $\phi, \theta$ , InternalGuess)
      using damped Newton;
      if solution2 exists then
        place solution2 in right octant;
        if solution2 is external then
          ExternalSol:= solution2
        else InternalSol:= solution2
        endif
      endif;
      if ExternalSol exists then
        ExternalGuess:= ExternalSol
      endif;
      if InternalSol exists then
        InternalGuess:= InternalSol
      endif;
      save InternalSol and ExternalSol in ProvTable2;
       $\theta := \theta - \nabla\theta$ 
    endwhile;
     $\phi := \phi + \nabla\phi$ 
  endwhile;
  put ProvTable1 and ProvTable2 data in LookUpTable
end

```

Fig. 12 Final method algorithm

- **Data and solutions lookup table:** The table has been calculated using the data and obtaining the results in the following way:
 - **Refraction indices:** In some real-world biaxial materials, two of the principal refraction indices are usually quite similar. That results in very small differences, not allowing a clear comparison of this case with uniaxial materials. On the other hand, real-world materials have relatively high values for all refraction indices. That means that all the rays coming from a given medium, such as air, are thus considerably bent, concentrating in a narrow cone near the direction of the surface normal. Again, this fact makes the biaxial phenomena more difficult to observe than uniaxial phenomena. The values chosen for our simulation are: First medium, isotropic with

$n = 1.0$. Second medium, biaxial, with principal refraction indices $n_x = 1.01$, $n_y = 1.06$, $n_z = 1.15$. These relatively low values still belong to the range of those found in real crystals, whilst being different enough to make the effects more clearly observable.

- **Interface plane orientation:** Coincident with the XY plane. This choice has two advantages: first, azimuth direction sampling is reduced by half due to symmetry; second, solutions corresponding to normal incidence may be calculated directly and used as initial guesses to begin the calculations.

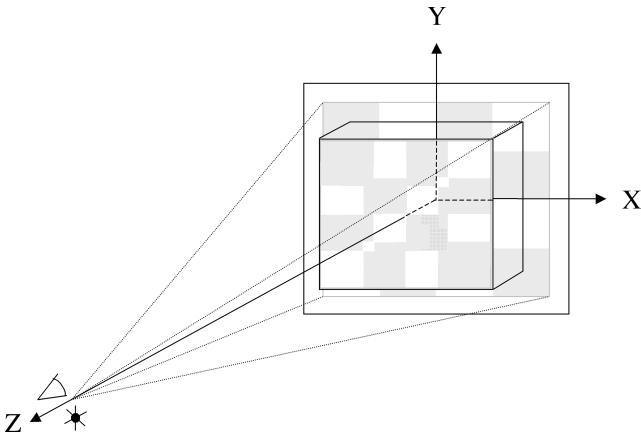


Fig. 13 Diagram of the experiment.

Tables may be calculated in a similar way for any other orientation. The program has been developed to allow calculations for this general case.

- **Sampling distribution of directions of incidence:** the directions of incidence are determined by two polar angles: azimuth (ϕ), measured from the OX axis in the $z = 0$ plane, and zenith (θ), measured from the OZ axis. The directions of incidence have been sampled at each half degree in the zenith direction and at each degree in the azimuth direction. Taking symmetry into account, only an octant is sampled ($0^\circ \leq \phi \leq 90^\circ$, $0^\circ \leq \theta \leq 90^\circ$). In any other case, and depending on symmetry, the zenith sampling range must extend to two or four octants corresponding to the first medium.
- **Number of iterations:** 256. We have empirically found that, once a suitable numerical method is found, it converges in a much lesser number of iterations, so 256 is a safe limit. The color uniformity in Figure 14 reflects this fact.
- **Solutions:** The three components of the tangent point have been obtained for each solution. Two

additional fields have been included: number of iterations and type of result; the latter encodes information about the method used and the validity and type of convergence.

- **Visualization techniques:** Different sets of images have been calculated to visualize and analyze the results of the different methods tested. These images display the following information:
 - **System of reference:** The three coordinate axes XX' , YY' and ZZ' are drawn in red, green and blue, respectively.
 - **Wave surface:** The wave surface is drawn in grey, only in the octant corresponding to refraction directions.
 - **Solutions representation:** Valid solutions are visualized as small red spheres; solutions falling in a periodic iteration are represented as green spheres; hue varies from lighter to darker green depending on the number of iterations before becoming periodic (0 to 255). Obviously, the solution is not represented if the iteration diverges.
 - **Points of view (position):** Solutions are found on both folds of the wave surface, and cannot be seen in one image only. For that reason, images are presented in pairs under symmetric points of view, in such a way that each image shows the inner or outer fold and solutions.

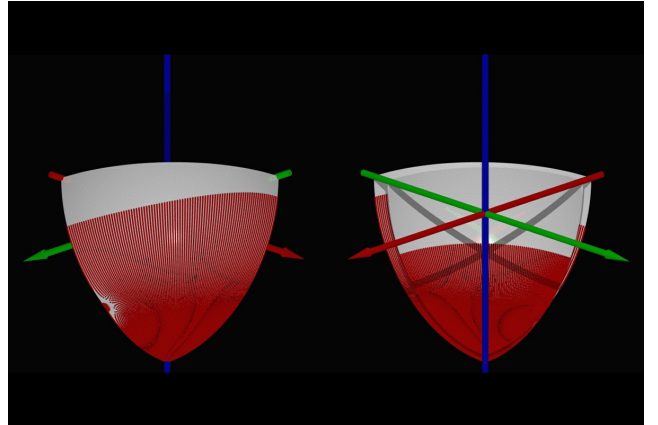


Fig. 14 Distribution of solutions using Damped Newton method.

Images obtained following this method help to explain the behavior of the solving methods:

- If all solutions are drawn (Figure 14), the general behavior of the method, as well as regions with problems and validity of solutions (which must be located on the wave surface) can be observed.
- If only solutions corresponding to a given set of directions of incidence, e.g. those corresponding

to a fixed value of ϕ (Figure 15) are drawn, the path described by the solutions as the direction of incidence changes may be observed.

- A particular pair of solutions may also be represented in order to make a detailed study.

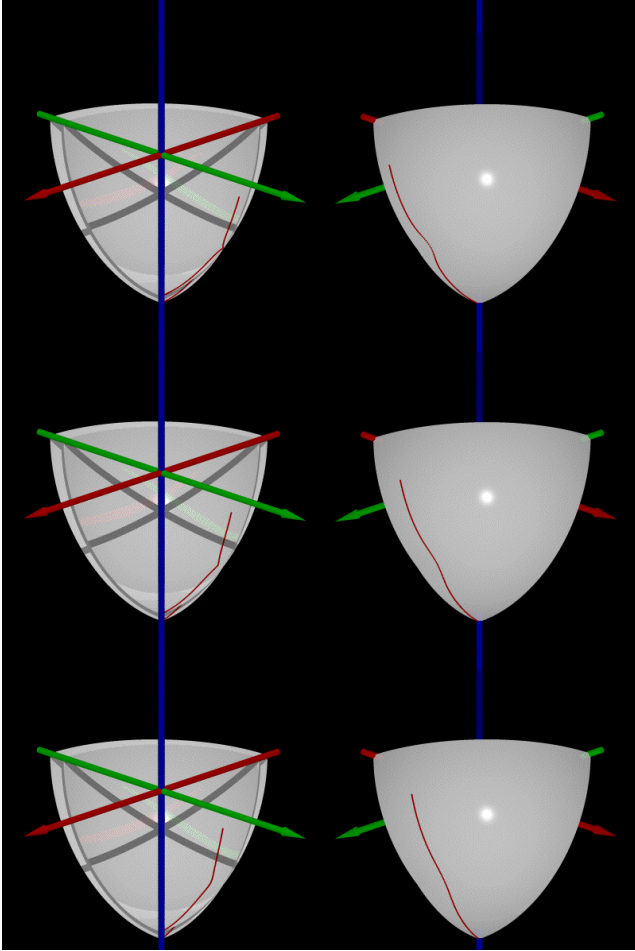


Fig. 15 Solutions corresponding to directions of incidence $\phi = 5^\circ$, $\phi = 10^\circ$ and $\phi = 15^\circ$.

5.2 Images

Figure 16 shows a real picture of a calcite uniaxial crystal, showing the visual deformations caused over a checkered pattern. The pattern was displayed on a computer monitor for better visibility of the effect. In order to visualize the results of the simulations presented in the paper, and to provide direct comparison with the real effect shown in Figure 16, two series of images have been obtained. The first set (Figure 17) displays three images which correspond to refraction on isotropic, uniaxial and biaxial crystals. Both refracted

rays receive the same energy ($k_1 = k_2 = 0.5$). The second set (Figure 18) shows three images which correspond to non-uniform energy distribution on uniaxial and biaxial crystals.

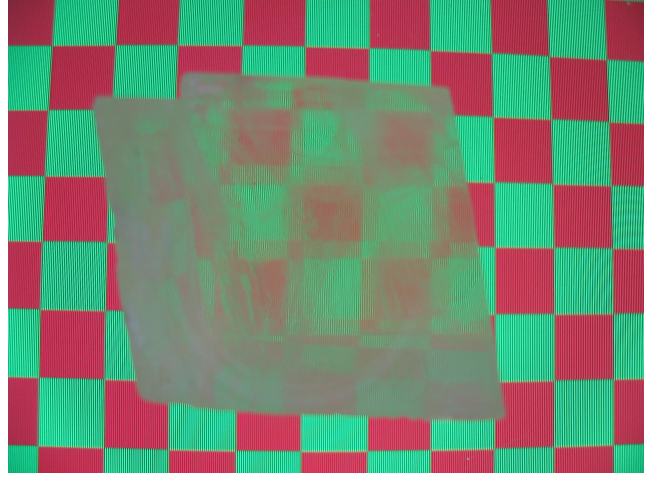


Fig. 16 Deformations caused by a calcite uniaxial crystal over a checkered pattern.

Both sets of images have been calculated as follows:

- **Geometric parameters of the camera and the scene elements:** The camera has been placed on the ZZ' axis, facing the coordinates origin. A checkered texture surface has been placed on the XY plane; the refracting plate has been placed right over this surface.
The only point light source is placed in the same position as the camera.
- **Rendering:** Images have been calculated using a simplified and adapted ray tracer, with perspective projection. Aliasing has been reduced using super-sampling techniques (4 rays/pixel).
- **Calculation of directions of refracted rays**
 - **Biaxial medium:** Directions of refraction are determined using a lookup table, which has been pre-calculated following the method and values previously described.
 - **Uniaxial medium:** Directions of ordinary and extraordinary rays are determined following the method previously described. Refraction indices are set to $n = 1.0$ (first medium) and $n_o = 1.06, n_e = 1.15$ (second medium).
 - **Isotropic medium:** In order to compare the results visually, an image has been calculated using $n = 1.0$ (first medium) and $n_r = 1.15$ (second medium).

Once the directions of the refracted rays have been calculated, the ray paths are followed using standard ray tracing techniques, thus determining their

intersection with the checkered texture plane and returning the resulting RGB color. No other reflections or internal refractions are included.

- **Energy distribution:** Energy and polarization issues have been left out of this work. Anyway, to allow a simple visualization of the directional behavior of the phenomena we have assumed a simple approximation where the energy transmitted by each ray E_1 and E_2 is obtained as $E_T = k_1 E_1 + k_2 E_2$, instead of using variable Fresnel terms.

5.3 Explanation of images

The following observations may be formulated after comparing the images corresponding to isotropic, uniaxial and biaxial media:

- Double refraction (see Figure 16), which is characteristic of anisotropic media, may be observed on the areas of separation between red and green squares, which acquire an intermediate color instead of a clean separation, as in isotropic media (see Figure 17, center and right and Figure 18).
- Separation between both rays (in biaxial and uniaxial media) depends on the zenith angle of incidence, which is zero in the normal direction of incidence (see Figure 17, center and right and Figure 18).
- The patterns formed near the corners of the squares in the biaxial medium is different from the uniaxial medium (see Figure 17, center and right and Figure 18). This is caused by the difference in the principal refraction indices (three different values in biaxial, two in uniaxial).

6 Conclusions and future work

The main goal of this paper is to study the directional issues of light behavior when ideally refracting from a homogeneous, non participating and isotropic medium to a homogeneous, non participating and anisotropic (uniaxial or biaxial) medium.

The main contribution has focused on formulating and solving the non-linear algebraic system that is obtained when the refraction process is simulated using the geometric model of Huygens for anisotropic media, and solving for the first time the biaxial case. This has been possible only using numerical techniques. The use of pre-computed tabled data should save computation times, thus being a good step towards real time simulation of the birefringency effect. We believe that the results shown are practical for the computer graphics community, in that the implementation is robust and feasible, and computational times acceptable.

We have additionally found a special case (which is never found in real materials, but can be reproduced in laboratory settings) when the direction of incidence coincides with the direction normal to the interface plane. For that case the system to solve is different. The solution for that specific case has also been obtained (see [12]) but it is not included here due to its minimal importance in the simulation of real materials. A second particular case for biaxial materials is given when the direction of the propagation of the wave inside the crystal coincides with its optical axis [3]. This case is known as *conical refraction* and its specific solution can also be found in [12].

The last open problems needed to generate physically-accurate images of biaxial crystals are including light polarization and taking into account the distribution of energy through Fresnel coefficients. As stated in [21], this represents a formidable problem for which no current solution has been found in literature.

A Analytical solution of the isotropic - uniaxial anisotropic refraction

The system given by equations 6, 7 and 8 has two different analytic solutions:

$$\begin{aligned} x_T &= \frac{v_z^2 (2v_0^2 y'_i E_{16} - y'_i z_B E_{17} + z'_i y_B E_{17})}{2v_0^2 (y'_i x_B - y_B x'_i) E_{16}} \\ y_T &= -\frac{v_z^2 (2v_0^2 x'_i E_{16} - x'_i z_B E_{17} + z'_i x_B E_{17})}{2v_0^2 (y'_i x_B - y_B x'_i) E_{16}} \\ z_T &= \frac{E_{17}}{2E_{16}} \end{aligned} \quad (14)$$

and

$$\begin{aligned} x_T &= \frac{v_z^2 (2v_0^2 y'_i E_{16} - y'_i z_B E_{15} + z'_i y_B E_{15})}{2v_0^2 (y'_i x_B - y_B x'_i) E_{16}} \\ y_T &= -\frac{v_z^2 (2v_0^2 x'_i E_{16} - x'_i z_B E_{15} + z'_i x_B E_{15})}{2v_0^2 (y'_i x_B - y_B x'_i) E_{16}} \\ z_T &= \frac{E_{15}}{2E_{16}} \end{aligned} \quad (15)$$

where the following substitutions have been made:

$$\begin{aligned} E_1 &= x_B y_B x'_i y'_i v_0^2 \\ E_2 &= y_i'^2 x_B^2 v_0^2 \\ E_3 &= y_i'^2 z_B^2 v_z^2 \\ E_4 &= y_i' z'_i z_B y_B v_z^2 \\ E_5 &= z_i'^2 y_B^2 v_z^2 \\ E_6 &= z_i'^2 x_B^2 v_z^2 \\ E_7 &= x_i' z'_i x_B z_B v_z^2 \\ E_8 &= x_i'^2 z_B^2 v_z^2 \\ E_9 &= y_B^2 x_i'^2 v_0^2 \end{aligned}$$

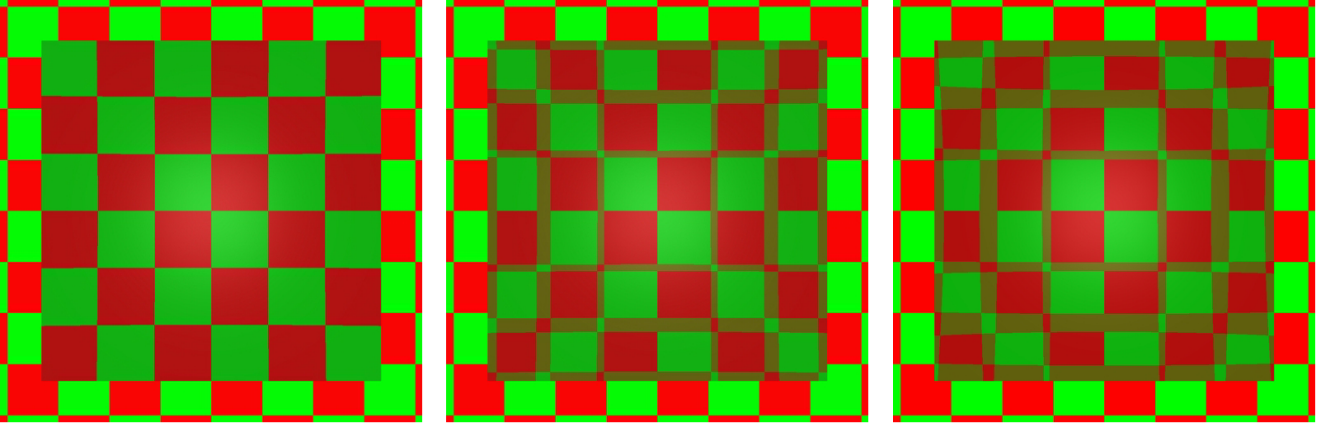


Fig. 17 Final images with uniform energy distribution. From left to right, isotropic, uniaxial and biaxial crystals over a checkered board. Notice how there is no distortion in the isotropic case, while the distortions caused by the biaxial crystal are more pronounced and seemingly irregular than in the uniaxial case.

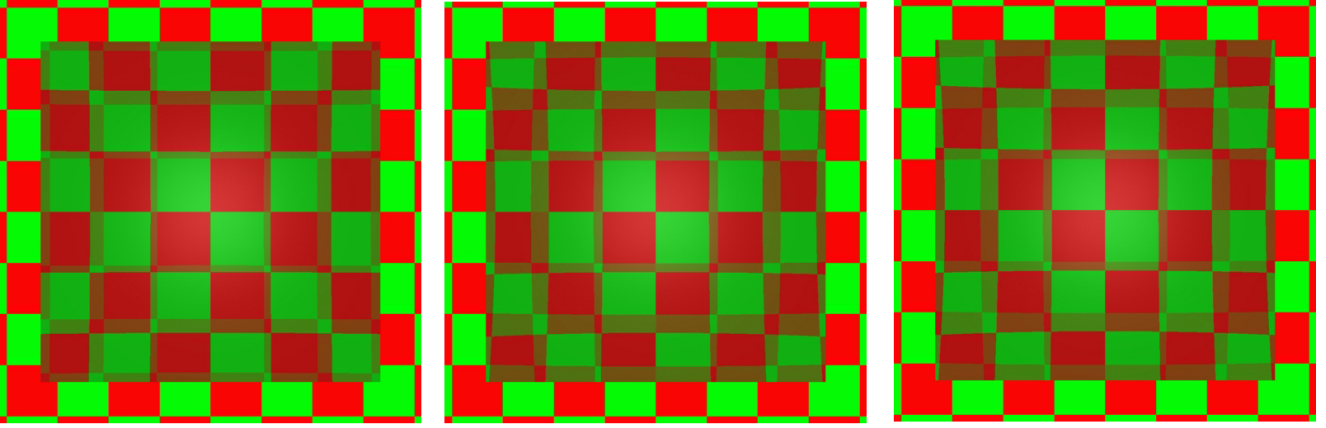


Fig. 18 Final images with non-uniform energy distribution. From left to right, uniaxial ($k_2 = 0.7$), biaxial ($k_1 = 0.4$) and biaxial ($k_2 = 0.6$)

$$\begin{aligned}
 E_{10} &= -v_0^2 (x_B y'_i - y_B x'_i) \\
 &\quad \sqrt{E_{18} - E_6 + v_z^4 z_i'^2 - E_5 + 2E_4 - E_3 - E_2 + 2E_1} \\
 E_{11} &= y'_i z'_i y_B v_z^2 v_0^2 \\
 E_{12} &= y_i'^2 v_0^2 v_z^2 z_B \\
 E_{13} &= x_B x'_i z'_i v_z^2 v_0^2 \\
 E_{14} &= z_B v_0^2 v_z^2 x_i'^2 \\
 E_{15} &= 2E_{14} - 2E_{13} + 2E_{12} - 2E_{11} - 2E_{10} \\
 E_{16} &= E_9 + E_8 - 2E_7 + E_6 + E_5 - 2E_4 + E_3 + E_2 - 2E_1 \\
 E_{17} &= 2E_{14} - 2E_{13} + 2E_{12} - 2E_{11} + 2E_{10} \\
 E_{18} &= y_i'^2 v_0^2 v_z^2 - E_9 - E_8 + 2E_7 + v_0^2 v_z^2 x_i'^2
 \end{aligned} \tag{16}$$

One of the solutions must be rejected, as it corresponds to a first medium point and has no physical meaning. The right solution may be determined by taking into account the direction of incidence and the interface plane normal. All the above calculations have been performed in Maple. This tool also allows to obtain the C (and other programming languages) code automatically [1] [2].

B Implementation of the numerical method

1. **Election of the norm:** The function

$$f = \frac{1}{2} F \cdot F \tag{17}$$

has been chosen due to the fact that the Newton step Δx is a *descent direction* for f :

$$\nabla f \cdot \Delta x = (F \cdot J) \cdot (-J^{-1} \cdot F) = -F \cdot F < 0 \tag{18}$$

where J is the Jacobian matrix.

2. **Calculation strategy:** The full Newton step $\lambda = 1$ is first attempted, because once the iteration is close enough to the solution quadratic convergence is achieved. However, the adequate reduction of f with the proposed step is checked at each iteration. If not suitable, *backtracking* along the Newton direction is performed until an acceptable step is found; this is guaranteed because the Newton step is always a descent direction for f . The only occasional failure may be caused by falling into a local minimum of f , and the only remedy is to try a different starting point.
3. **Election of step λ_k .** It is not sufficient to simply require that $f(x_{k+1}) < f(x_k)$. This criterion can fail to converge to a minimum of f in one of two ways. First, a sequence of steps satisfying this criterion with f decreasing too slowly relative to the step lengths may appear.

A way to solve the problem is to require the average rate of decrease of f to be at least some fraction α of the initial rate of decrease $\nabla f \cdot \Delta x$:

$$f(x_{k+1}) \leq f(x_k) + \alpha \nabla f \cdot (x_{k+1} - x_k) \quad (19)$$

where $0 < \alpha < 1$; a good choice may be $\alpha = 10^{-4}$.

Second, a sequence where the step lengths are too small relative to the initial rate of decrease of f may also appear. The solution now is to require the rate of decrease of f to be greater than some cutoff value.

4. **Backtracking algorithm strategy.** For each iteration, a $g(\lambda)$ function is defined:

$$g(\lambda) \equiv f(x_k + \lambda \Delta x) \quad (20)$$

so that

$$g'(\lambda) = \nabla f \cdot \Delta x \quad (21)$$

If backtracking is needed, then g is determined with the most up to date information available, and λ is chosen to minimize g .

At the starting point $g(0)$ and $g'(0)$ are known. The first step is always the Newton full step $\lambda = 1$ obtaining $g(1)$. If this step is not acceptable, $g(\lambda)$ may be obtained from a quadratic approximation $g(\lambda) \approx a\lambda^2 + b\lambda + c$. Taking this equation and its derivative in $\lambda = 0$ and the equation in $\lambda = 1$, $c = g(0)$, $b = g'(0)$ and $a = g(1) - g(0) - g'(0)$ are obtained, and

$$g(\lambda) \approx [g(1) - g(0) - g'(0)]\lambda^2 + g'(0)\lambda + g(0) \quad (22)$$

Taking the derivative, a minimum is found when

$$\lambda = -\frac{g'(0)}{2[g(1) - g(0) - g'(0)]} \quad (23)$$

It may be shown that $\lambda \simeq \frac{1}{2}$ for small values of α . Moreover, $\lambda_{min} = 0.1$ is imposed to avoid too small values of λ .

On second and subsequent backtracks (if needed) g may be calculated as a cubic in λ using the previous and second most previous values ($g(\lambda_1)$) and ($g(\lambda_2)$).

Acknowledgements The authors wish to thank Fermin Gomez, Jorge Lopez-Moreno and Jorge Jimenez for their help putting the paper together, as well as the anonymous reviewers for their keen comments on previous drafts.

References

1. B. W. Char and others. *Maple V. Maple Library Reference Manual*. Springer Verlag, New York, 1992.
2. B. W. Char and others. *Maple V. Maple V Language Reference Manual*. Springer Verlag, New York, 1992.
3. Max Born and Emil Wolf. *Principles of Optics, 7th Edition*. Cambridge University Press, New York, 1999.
4. A. M. Buchanan and Andrew W. Fitzgibbon. Damped newton algorithms for matrix factorization with missing data. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 316–322, 2005.
5. P. Denfhard. *Numerical Methods for Non-linear Problems*. Springer, 2004.
6. S. P. Dirkse and M. C. Ferris. A pathsearch damped newton method for computing general equilibria. *Annals of Operation Research*, 68(2):211–232, 1996.
7. B. I. Epureanu and H. S. Greenside. Fractal basins of attraction associated with a damped newtons method. *Siam Review*, 40(1):1–8, 1998.
8. Carlos A. Felippa. Nonlinear finite element methods course (chapter 21). <http://www.colorado.edu/engineering/cas/courses.d/NFEM.d/>, October 2010.
9. Peter Gay. *The crystalline state: An introduction*. Longman Group Limited, 1972.
10. S. Guy and C. Soler. Graphics gems revisited. fast and physically based rendering of gemstones. *ACM Transaction on Graphics*, 23(3):231–238, 2004.
11. S. Kaufmann and A. Montalvo. The application of a modified damped newton method for the simulation of vapor-liquid stagewise processes. *Computers and Chemical Engineering*, 7(2):93–103, 1983.
12. Pedro M. Latorre Andrés. *Modelos físicos de iluminación: Simulaciones por computador*. PhD thesis, Universidad de Zaragoza, Zaragoza, Spain, 2001.
13. A. N. Matveev. *Optics*. Mir Publishers, Moscow, 1988.
14. J. Ortega and W. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, 1970.
15. William H. Press et al. *Numerical recipes in C. The Art of Scientific Computing, 2nd. Edition*. Cambridge University Press, New York, 1995.
16. R.A. Chipman S.C. McClain. Polarization ray tracing in anisotropic optically active media. *Proc. SPIE Int. Soc. Opt. Eng.*, 1746:107–118, 1992.
17. J.M. Stone. *Radiation and Optics: an Introduction to Classical Theory*. McGraw-Hill, New York, 1963.
18. G. Szivessy. *Handbuch der Physik*, chapter Kristalloptik, page 840. H. Geiger and K. Scheel, Berlin, 1928.
19. David C. Tannenbaum, Peter Tannenbaum, and Michael J. Wozny. Polarization and birefringency considerations in rendering. In *Computer Graphics. Proceedings, Annual Conference Series*, pages 221–222. ACM SIGGRAPH, 1994.
20. C. W. Ueberhuber. *Numerical Computation. Methods, Software, and Analysis*. Springer, Berlin, 1997.
21. Andrea Weidlich and Alexander Wilkie. Realistic rendering of birefringency in uniaxial crystals. *ACM Transactions on Graphics*, 27(1):6:1–6:12, March 2008.
22. Dokkyun Yi. Damped newtons method for image restoration. *Trends in Mathematics*, 9(1):109–113, june 2006.