

Newton's Colors: Simulating Interference Phenomena in Realistic Image Synthesis

Brian E. Smits and Gary W. Meyer

Department of Computer and Information Science
University of Oregon
Eugene, Oregon 97403

The laws of physics that govern thin film interference are reviewed and the colors in nature that are produced by this mechanism are simulated. An efficient technique for adjusting out of gamut colors is developed and is applied to the problem of synthesizing the highly saturated interference colors. Common examples of thin film interference are reproduced, including the iridescent colors created by multiple film layers.

1 INTRODUCTION

In order to determine the color of an object in a synthetic scene, the spectral composition of the light reaching the observer from that object must be calculated. In nature there are a variety of mechanisms by which a spectral energy distribution can be produced and altered. Simulation of these mechanisms in computer graphics is an example of a physically based approach to realistic image synthesis. In this paper we explore the types of colors that are produced by the modelling of one specific phenomena: interference.

Light striking the surface of a material must be either reflected, transmitted, scattered, or absorbed. In addition, the material itself may emit some light. The reflected light consists of a specular component that is sent off at an angle equal to the angle of incidence, and a diffuse component that goes off in many directions. The path of the transmitted light will be changed due to refraction of the light at the surface of the material. Some portion of the light will be absorbed by the material and turned into heat. Finally, scattering will cause some of the light to be released from the sides of the material.

Interference phenomena are the result of both the reflection and transmittance mechanisms mentioned above. Perhaps the most common example of interference is the colors produced on the surface of oil slicks and soap bubbles. These types of interference phenomena were first studied by Newton and the sequences of colors that are produced by them are called Newton's colors. Interference effects can also be seen on old vases, metal, and peacock feathers. Structural engineers analyze stress by viewing interference patterns in plastic models.

In most cases, these interference effects are caused by very thin films which determine the amount of light at each wavelength that is reflected back to the viewer. As the thickness of the film varies, the spectral composition of the light and therefore the color that gets reflected back to the viewer changes as well. Interference effects have not been modeled in computer graphics so this work enlarges the set of natural phenomena that have been simulated in computer graphics. They are another example of using the physical properties of light and the way it interacts with surfaces to more accurately model the natural world.

2 THE PHYSICS OF INTERFERENCE PHENOMENA

Interference phenomena are the result of the wave-like properties of light. According to the superposition principle, when two waves meet the amplitude of the resultant wave is the sum of the amplitudes of the two component waves. Suppose two waves with the same wavelength and amplitude merge together. If the crest of both waves match up exactly, perfect constructive interference occurs and the resulting wave has twice the amplitude of the two original waves. These two waves are said to have a phase difference of zero. If the crest of one wave matches up with the trough of another, perfect destructive interference results and the waves cancel each other completely. The phase difference in this case is equal to π . Given the phase difference δ of two waves with amplitude A , the resulting amplitude A' is given by

$$A' = 2A \cos\left(\frac{\delta}{2}\right). \quad (1)$$

Because light has wave-like properties, it obeys this rule (Tipler, 1976). The amount of interference can be different for each wavelength producing a change in the color of the light we see.

Interference phenomena are affected by the speed of the light waves through the medium in which they are travelling. Each medium, such as water, air, glass, or oil, has what is called an index of refraction n given by

$$n = \frac{c}{v} \quad (2)$$

where c is the speed of light in a vacuum and v is the speed of light in the medium. A vacuum has an index of refraction $n = 1$ and air is assumed to have the same index for most calculations. Water, however, has an index of refraction $n = 1.33$ showing that light travels at only seventy-five percent of its usual speed through water. The relationship between the speed v , wavelength λ , and frequency f of a wave is given by

$$v = \frac{f}{\lambda}. \quad (3)$$

When a wave enters a new medium, its speed changes. This means that the wavelength must change because, for a given wave, the frequency is fixed. The new wavelength λ' in the new medium is given by

$$\lambda' = \frac{\lambda}{n}. \quad (4)$$

The index of refraction also determines how the wave is reflected back from the boundary. If, at the boundary of two different mediums, the new medium has a higher index of refraction than the old one, when the wave bounces off of the boundary it will be inverted ($\delta = \pi$) with every trough becoming a crest and every crest a trough. A wave will not be inverted if it reflects off a boundary with a medium having a lower index of refraction. This is important because it makes two different kinds of interference effects possible, one with a phase change, and one without.

When light with wavelength λ strikes a thin film, such as a soap bubble or an oil film on water, it is reflected back at both surfaces of the film. If the film has an index of refraction greater than the surrounding medium, the wave will be inverted when it is reflected off of the back face of the film. When the ray reflected from the top surface meets the ray reflected from the bottom surface, they interfere with one another (Figure 1). Assume, for the moment, that the light waves travel perpendicular to the surface and the inversion of the wave can be neglected. Then the extra distance or path difference that the light rays travel is equal to two times the thickness t of the film. There are $2t/\lambda'$ waves within the film where λ' is the wavelength of the light modified by the index of refraction of the surface. If the film has a thickness of $t = 0$, then there are no waves within the film and the phase difference is 0. If the film has a thickness of $t = \lambda'/4$ then there are

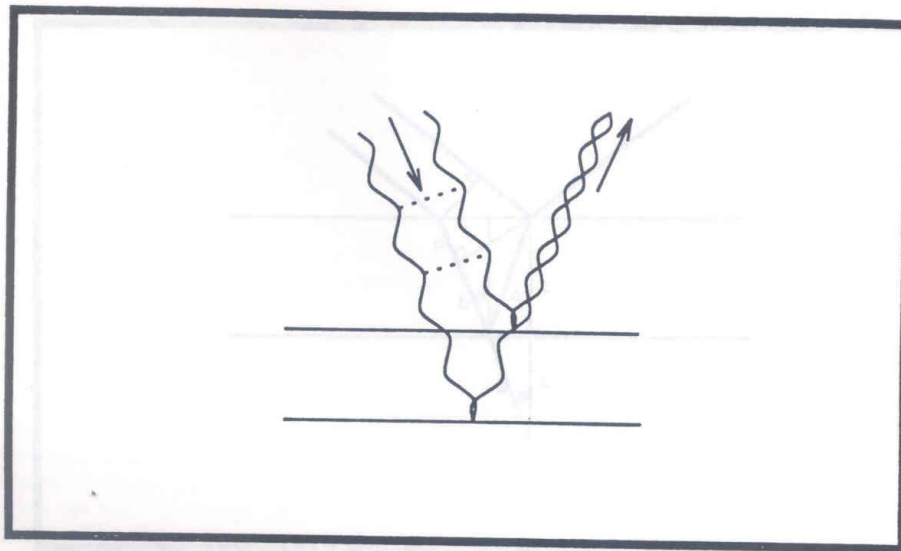


Fig. 1: Interference occurs between light wave that has reflected off the bottom of the film and light wave that has reflected off the top.

$(2 * \lambda' / 4) / \lambda' = 1/2$ waves within the film and the phase difference is π . In addition to the phase difference that is a result of the thickness of the film, there is a one half cycle phase change that occurs when the light reflects off of the back boundary. To accommodate this phase change, a factor of π is added to the phase difference. Given these considerations, the phase difference of light reflecting off of a thin film is

$$\delta = 2\pi \left(\frac{2t}{\lambda'} \right) + \pi \quad (5)$$

and the amplitude A_r of the resulting wave is

$$A_r = 2A \cos \left(\frac{\delta}{2} \right) = 2A \cos \left(\frac{2\pi t n}{\lambda} + \frac{\pi}{2} \right) \quad (6)$$

where A is the amplitude of the light wave reflected off of each surface (Tipler, 1976). For a film with thickness t near zero, the resulting amplitude is close to zero for all wavelengths of light. This causes the top of a soap bubble to appear black, because this is where the bubble is thinnest. This shows that the waves are inverted when they reflect off of the back face and cancel out almost all of the light reflecting off of the front face.

Given this information we can calculate the spectral energy distribution of the light reflecting off of a thin film. We define the reflectance $\rho(\lambda)$ of a film with thickness t and index of refraction n to be

$$\rho(\lambda) = \left[K \cos \left(\frac{2\pi t n}{\lambda} + \frac{\pi}{2} \right) \right]^2 \quad (7)$$

where K is a constant that measures how much light is reflected at each boundary of the film. The expression in brackets is squared because the intensity of a wave is equal to the square of its amplitude.

Some interference phenomena are caused by a film with air on one side and some other medium with a higher index of refraction than the film on the other side. In this case, the light waves are not inverted when they reflect off of either the back boundary or the

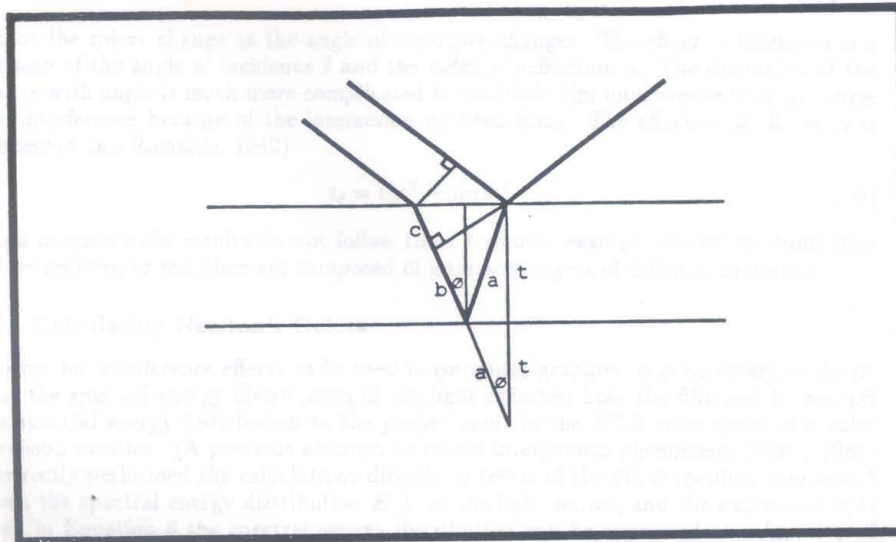


Fig. 2: Angle of view contributes to a difference $a + b$ in path length that is equal to $2t \cos \theta$.

front boundary. The constant π that was added to the previous equation for a film with mediums of lower refractive index on each side should not be used for this case. For this type of interference phenomena, the reflectance $\rho(\lambda)$ for a film with thickness t is defined to be

$$\rho(\lambda) = [K \cos \left(\frac{2\pi t n}{\lambda} \right)]^2 \quad (8)$$

where K again is a measure of how much light is reflected at each boundary of the film.

In both of these types of interference phenomena, the path difference is affected by the angle at which the viewer is looking at the film. When light strikes a film at an angle, the exit point of the wave is different from the entrance point. At the exit point, the wave interferes with a wave that has just reflected after making its initial contact with the film surface. The light travels farther through the film than before, but the path difference ends up being less. This is because the light wave that the exiting ray interferes with has also traveled farther. The light wave that enters the film travels a distance of c before the corresponding point on the other wave reaches the film. The path difference between the two waves is $a + b$ which is equal to $2t \cos \theta$ (Figure 2). For a perpendicular wave, θ is zero and the path difference is twice the thickness. Therefore the effective thickness t in Equation 8 is $t \cos \theta$.

Another interference phenomena occurs when light strikes a surface that is composed of multiple uniform layers of a transparent substance. This multiple film interference produces iridescence which is seen as the bright colors on the feathers of birds (such as peacocks), some insect wings, and opals. Each film of the surface tends to reinforce certain ranges of wavelengths. The more films that are present, the narrower and brighter the range of reflected light. This repeated reinforcement of limited ranges of wavelengths causes the very bright and almost spectral colors that are found in multiple film interference phenomena. The intensity I at each wavelength λ of light is

$$I = \frac{1}{2} A_0 \left(\frac{\sin 2\pi(l+1)t/\lambda}{\sin 2\pi t/\lambda} \right)^2 \quad (9)$$

where A_0 is the amplitude of the light striking the film, l is the number of layers, and t is the thickness of a layer (Anderson and Richards, 1942). One of the properties of iridescence

is that the colors change as the angle of incidence changes. The effective thickness is a function of the angle of incidence θ and the index of refraction n . The derivation of the change with angle is much more complicated for multiple film interference than for single film interference because of the interaction between films. The effective thickness t_θ is (Anderson and Richards, 1942)

$$t_\theta = t(n^2 - \sin^2\theta)^{1/2}. \quad (10)$$

Often in nature the results do not follow these formulae exactly because the films may not be uniform or the films are composed of interlaced layers of different materials.

2.1 Calculating Newton's Colors

In order for interference effects to be used in computer graphics, it is necessary to determine the spectral energy distribution of the light reflected from the film and to convert this spectral energy distribution to the proper point in the *RGB* color space of a color television monitor. (A previous attempt to model interference phenomena (Watt, 1989) incorrectly performed the calculations directly in terms of the *RGB* monitor primaries.) Given the spectral energy distribution $E(\lambda)$ of the light source, and the expression $\rho(\lambda)$ given in Equation 8 the spectral energy distribution can be expressed as a function of t the thickness, n the index of refraction, and λ the wavelength. Using the matching functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ for CIE *XYZ* space, we can convert this spectral energy distribution into CIE *XYZ* tristimulus values and then into into the *RGB* space of the monitor (Meyer and Greenberg, 1987).

For some ranges of interference colors, the *RGB* tristimulus values become negative. This occurs where the film is relatively thin, because in some cases only very narrow bands of wavelengths have a high intensity. These ranges are closer to monochromatic colors and cannot be displayed on a monitor. The process of deciding what color to display for these out of gamut colors will be covered in Section 3.

A table of *RGB* values for interference effects with a phase change and one for effects without the phase change is precomputed and is indexed based upon the thickness of the film and the index of refraction. This is done so that the expensive processes of calculating the color need only be done once for each thickness. Note from Equation 7 that the table can be precomputed for an index of refraction of 1 (Nassau, 1980). When a film with thickness t and index of refraction n is seen at an angle θ , the product t' of the thickness and the index of refraction becomes

$$t' = tn \cos \theta \quad (11)$$

Once the product t' is determined, the color for that thickness is looked up in the table. This makes the rendering much more efficient.

3 OUT OF GAMUT COLORS

The gamut of a color device is the set of colors that it can produce. The primaries used in the device determine its gamut, and when the primaries of two devices differ, the gamuts do not match up exactly. This gamut mismatch problem causes ranges of color on one device that cannot be reproduced on the other. These colors are called out of gamut colors.

Gamut mismatch occurs in several different circumstances. One of the situations in which gamut mismatch is a major problem is between color monitors and color printers. The vivid, highly saturated colors that a monitor can display are impossible to reproduce using a color printer because color printers use different primaries and a different process for creating color. This problem makes the final printed results look much different than the display on the monitor. Gamut mismatch also occurs between monitors, causing dissimilarity between the same picture seen on different monitors. This is because the

red, green, and blue phosphors vary slightly between monitors. The gamut mismatch that happened while modeling interference phenomena was a third mismatch, that between nature and the monitor. Interference effects produce colors that are almost spectral and that lie outside of the gamut of almost all reproduction devices.

3.1 A Solution to Out of Gamut Colors

In all of these gamut mismatches, adjustments need to be made to the colors that are out of the destination gamut. The color that is chosen as a replacement for the out of gamut color must be in the destination gamut and should be as close as possible to the original color. One potential solution is to clip the colors outside the destination gamut to the boundary of the gamut. This involves finding a color on the boundary of the gamut that is close to the original out of gamut color. Another idea is to compress the input gamut and all colors in the image down to fit within the destination gamut. This involves scaling the colors down by some amount that is determined by the input gamut.

In a recent paper (Gentile and Allebach, 1989) thirteen different techniques for dealing with out of gamut colors were compared. All but one of these techniques (the control) were performed in the perceptually uniform $L^*u^*v^*$ space. The control was done in RGB space and used the Euclidean distance formula to select the closest color to the out of gamut color. Both clipping and compression techniques were tested, keeping various subsets of lightness, saturation, and hue constant while performing the color adjustment. Considering only techniques which did no analysis on the picture prior to clipping or compression, the best techniques were clipping with a constant lightness, or clipping with constant lightness and hue. For this work, we have chosen to use the clipping with constant lightness and hue approach.

The problem with all of the techniques investigated in Gentile and Allebach (1989) is that they operate in $L^*u^*v^*$ space. $L^*u^*v^*$ space is not computationally efficient for large pictures because of the non-linear transformation that is required for every color in the image. What is needed is an algorithm which will work in CIE XYZ space and have similar properties to the constant lightness and hue $L^*u^*v^*$ space algorithms. By holding lightness and hue constant, only the saturation of the color changes. The color is being pulled in towards the achromatic colors that lie on the diagonal line between black and white in the color space. In this paper, we will refer to this line as the neutral diagonal. In CIE XYZ space lightness is a function of Y alone, so we could pull back in towards the neutral diagonal while keeping Y constant (Figure 3). This will desaturate the color and bring it within the gamut of the output device. In CIE XYZ space, colors with constant hue do not lie on a straight line from the neutral diagonal to the edge of the gamut, but are curved somewhat (Meyer and Greenberg, 1987). No linear transform will make the lines of constant hue straight. Because the constant hue lines are curved, the hue will not remain constant as the out of gamut color is brought back in towards the neutral diagonal. The change in hue, however, should not be significant enough to cause problems.

The gamut that will be used in this case as the destination gamut is the RGB space of a color television monitor. The input gamut is the calculated interference colors, but it could just as easily be an input device such as a scanner, or the RGB space of a different monitor. Some ranges of the interference colors and the entire spectrum are the worst case for an input gamut because they are spectral colors. The techniques used here to deal with out of gamut colors should be applicable to any type of gamut mismatch. The following algorithm is designed to work on a gamut that has the shape of a parallelepiped such as an RGB monitor gamut in CIE XYZ space.

3.2 The Clipping Algorithm

The clipping algorithm is composed of three steps. Given a color to be reproduced, it first checks to see if the color needs to be clipped. Next it determines the direction of the

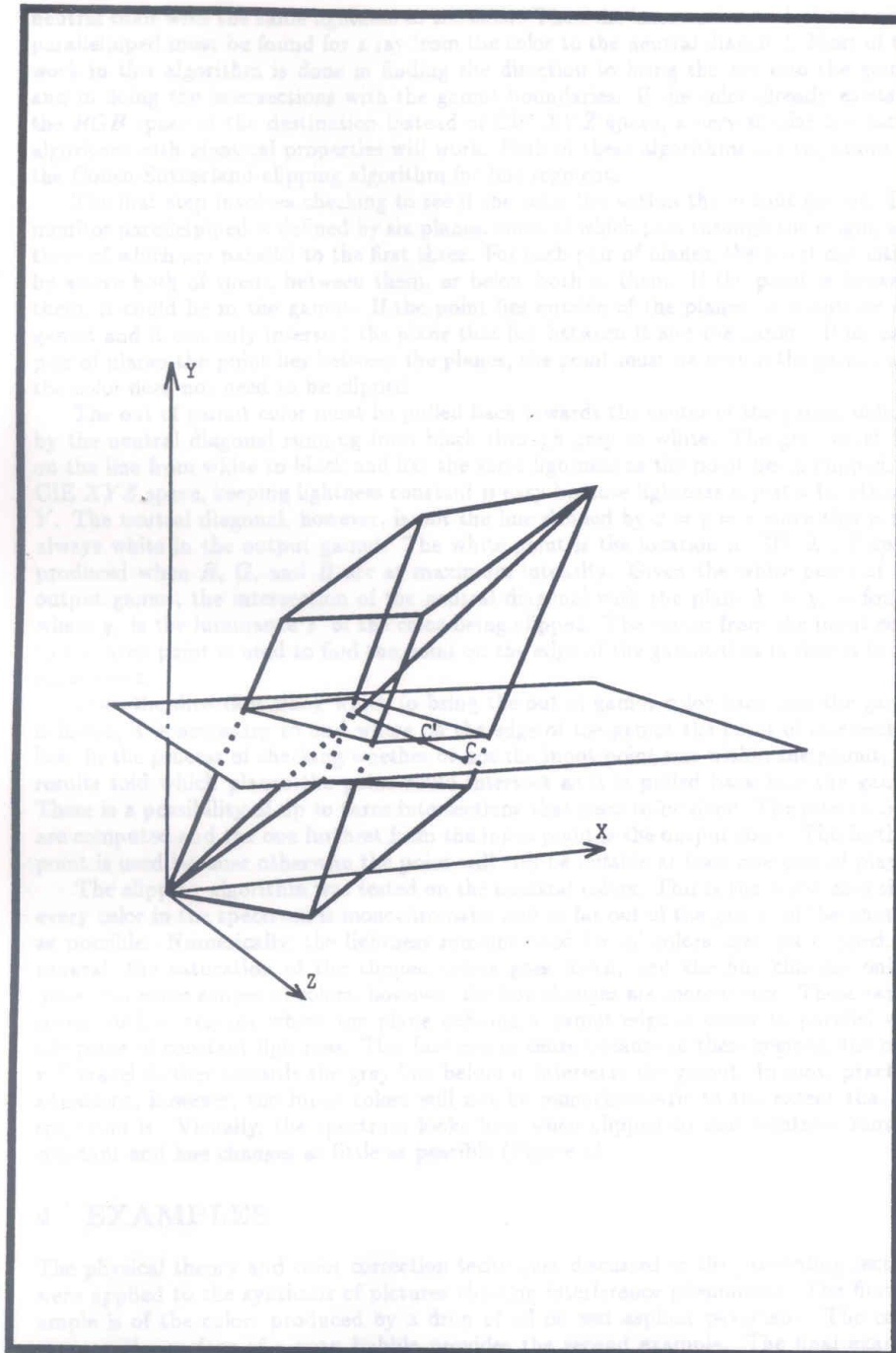


Fig. 3: Out of gamut color is approximated by following a line of constant hue and lightness in towards the neutral diagonal.

neutral color with the same lightness as the color. Then the intersection with the monitor parallelepiped must be found for a ray from the color to the neutral diagonal. Most of the work in this algorithm is done in finding the direction to bring the ray into the gamut and in doing the intersections with the gamut boundaries. If the color already exists in the *RGB* space of the destination instead of CIE *XYZ* space, a very similar but faster algorithm with identical properties will work. Both of these algorithms are variations on the Cohen-Sutherland clipping algorithm for line segments.

The first step involves checking to see if the color lies within the output gamut. The monitor parallelepiped is defined by six planes, three of which pass through the origin, and three of which are parallel to the first three. For each pair of planes, the point can either be above both of them, between them, or below both of them. If the point is between them, it could lie in the gamut. If the point lies outside of the planes, it is outside the gamut and it can only intersect the plane that lies between it and the gamut. If for each pair of planes the point lies between the planes, the point must be within the gamut and the color does not need to be clipped.

The out of gamut color must be pulled back towards the center of the gamut defined by the neutral diagonal running from black through grey to white. The grey point lies on the line from white to black and has the same lightness as the point being clipped. In CIE *XYZ* space, keeping lightness constant is easy because lightness is just a function of *Y*. The neutral diagonal, however, is not the line defined by $x = y = z$ since this is not always white in the output gamut. The white point is the location in CIE *XYZ* space produced when *R*, *G*, and *B* are at maximum intensity. Given the white point of the output gamut, the intersection of the neutral diagonal with the plane $Y = y_p$ is found, where y_p is the luminance *Y* of the color being clipped. The vector from the input color to this grey point is used to find the point on the edge of the gamut that is closest to the input color.

Once the direction along which to bring the out of gamut color back into the gamut is found, it is necessary to find where on the edge of the gamut the point of intersection lies. In the process of checking whether or not the input point was within the gamut, the results told which planes the point could intersect as it is pulled back into the gamut. There is a possibility of up to three intersections that need to be done. The intersections are computed and the one furthest from the input point is the output color. The furthest point is used because otherwise the point will still be outside at least one pair of planes.

The clipping algorithm was tested on the spectral colors. This is the worst case since every color in the spectrum is monochromatic and as far out of the gamut of the monitor as possible. Numerically, the lightness remains fixed for all colors that get clipped. In general, the saturation of the clipped colors goes down, and the hue changes only a little. For some ranges of colors, however, the hue changes are more severe. These ranges correspond to regions where the plane defining a gamut edge is closer to parallel with the plane of constant lightness. This fact makes sense because in these regions, the color will travel farther towards the grey line before it intersects the gamut. In most practical situations, however, the input colors will not be monochromatic to the extent that the spectrum is. Visually, the spectrum looks best when clipped so that lightness remains constant and hue changes as little as possible (Figure 4).

4 EXAMPLES

The physical theory and color correction techniques discussed in the preceding sections were applied to the synthesis of pictures showing interference phenomena. The first example is of the colors produced by a drop of oil on wet asphalt pavement. The colors seen on the surface of a soap bubble provides the second example. The final example is of the iridescent colors exhibited by the wings of the *Morpho* butterfly. In all three cases the nearly spectral colors that were produced by the interference effects provided an opportunity to evaluate the effectiveness of the color correction techniques.

4.1 Oil Film on Water

When a drop of oil falls onto a surface it often forms a pattern of concentric circles. On a completely smooth and level surface the oil would spread slowly, staying thicker in the middle and becoming very thin at the edges. It forms a mound on the surface with the thickness of the oil being a function of distance from the center. Based on these observations and an illustration from Minnaert (1954), the thickness was modeled as an exponential function of the square of the distance from the center point. For an oil spill centered at the origin, this normal distribution is

$$t = e^{-k_1(x^2+y^2)}, \quad (12)$$

where k_1 is a constant that determines how quickly the oil diminishes to a negligible thickness.

This method is a good approximation to an oil slick on a perfectly smooth surface. Oil, however, is most commonly seen on pavement and does not spread out evenly due to the roughness and slope of this surface. Using a turbulence function from Perlin (1985), the distance from the center can be randomly perturbed at each point in a way that keeps the oil slick roughly concentric. The resulting formula is as follows

$$t = e^{-k_1(x^2+y^2+k_2\text{turb}(x,y))} \quad (13)$$

where k_2 is a constant that affects how irregular the spreading of the oil is. The turbulence function returns a self-similar pattern of perturbation which gives a good visual impression of a turbulent flow (Perlin, 1985). This formula approximates how each area of the oil slick spreads out at different rate. Further distortions could be done to t to model the effects of a sloped surface or other irregularities, but this formula serves as a good representation for a basic oil slick (Figure 5). For a point on the surface, the thickness t is computed based on the distance from the center and the value of the turbulence function at that point. The thickness t is then changed based on the angle between the viewer and the surface. The resulting t is used as an index into a precomputed table of colors and the color in *RGB* space is finally determined for that point.

In order to get a good picture, the slick needs to lie on an appropriate surface. The most familiar surface for an oil slick is pavement, but pavement is a very complex surface because of all of its irregularities. To get a realistic effect, a surface that had an irregular scale of bumps and an irregular amount of roughness to it had to be approximated. This was done by using two separate calls to a noise function to determine the scale and the roughness of the final texture. After doing this distortion, very fine bumps were mapped over the result to make the surface appear even rougher.

4.2 Soap Films

A soap film is another example of interference phenomena. The bright colors that reflect off of a soap bubble are caused by the very thin film of the bubble. As with the oil slick, the light changes phase when it strikes the back face of the film. The proof of the phase change can be seen near the tops of bubbles and soap films. Just before they pop, the tops of the bubbles look black because this is where they are the thinnest. The blackness is caused by almost complete destructive interference resulting from a phase difference of π due to almost identical path lengths.

Soap films are affected by gravity. The water in the film is being pulled down by gravity and up by the surface tension in the film. In still air the film is thicker near the bottom than the top, and therefore the color at each point on the film is a function of its height. This relationship causes bands of color to appear on a soap film after it has stabilized. In modeling soap bubbles, height was normalized to a zero to one range, the result from a call to the turbulence function was added to the height, and the thickness was taken as being directly proportional to the height. The turbulence approximates

the effects of air currents and other irregularities that influence the thickness of the film. Using a simple linear function for the thickness is probably incorrect, but because of the large amounts of turbulence that are added, the inaccuracy is not noticeable (Figure 6).

4.3 Iridescence

Iridescence effects are found on the wings of *Morpho* butterflies. The vanes of this butterfly's wings have twelve horizontal mullions with the spacing and thickness necessary to create a multiple film interference effect (Anderson and Richards, 1942). The surface reflectance that results has a peak at about 400nm. Because of the thickness of the mullions, the effective thickness t_θ of the film is a function of the angle θ between the incident ray and the surface normal, of the spacing between the mullions d_1 , of the thickness of the mullions d_2 , and of the refractive index of the mullions n which is 1.53 for *Morpho* butterflies (Anderson and Richards, 1942)

$$t_\theta = d_1 \cos \theta + d_2(n^2 - \sin^2 \theta)^{\frac{1}{2}}. \quad (14)$$

An iridescent surface was modeled based on the colors of the butterfly wings. A table referenced by angle was generated for $d_1 = 150$ nm and $d_2 = 50$ nm and used to determine the color at each point. The surface was made to look wavy to show the colors that occur with varying angle (Figure 7). The mullions vary in thickness between 50 nm and 150 nm, so the actual colors reflected from the wings of the butterfly are a blend of multiple film interference colors resulting from the variations in the thicknesses of the mullions (Anderson and Richards, 1942).

5 CONCLUSION

Modelling interference phenomena is another example of applying the laws of physics to the synthesis of realistic images. Interference effects have not been accurately simulated in computer graphics, and by doing so we have enlarged the set of available techniques for synthesizing color and for modelling the natural world. Three specific examples of naturally occurring interference effects were simulated and images depicting these phenomena were produced. Because very pure colors are created by the interference of light waves, an efficient color correction technique for out of gamut colors was also devised. This technique is applicable to other situations such as color printing where there is a need to adjust colors.

References

- Anderson, Thomas F. and Richards, A. Glenn, Jr. (1942) An Electron Microscope Study of Some Structural Colors of Insects. *Journal of Applied Physics*, 13, 748-758.
- Gentile, Ronald S. and Allebach, Jan P. (1989) A Comparison of Techniques for Color Gamut Mismatch Compensation. in *Human Vision, Visual Processing, and Digital Display (SPIE Proceedings Series)*, 1077, 342-354.
- Meyer, Gary W. and Greenberg, Donald P. (1987) Perceptual Color Spaces for Computer Graphics. In *Color and the Computer* edited by H. John Durrett. Academic Press, Boston.
- Minnaert, M. (1954) *The Nature of Light and Color in the Open Air*. Dover Publications, New York.
- Nassau, Kurt (1983) *The Physics and Chemistry of Color: The Fifteen Causes of Color*. John Wiley and Sons, New York.
- Perlin, Ken (1985) An Image Synthesizer. *Computer Graphics*, 19, 287-296.
- Tipler, Paul A. (1976) *Physics*. Worth Publishers, New York.
- Watt, Alan (1989) *Fundamentals of Three-Dimensional Computer Graphics*. Addison-Wesley, Wokingham.

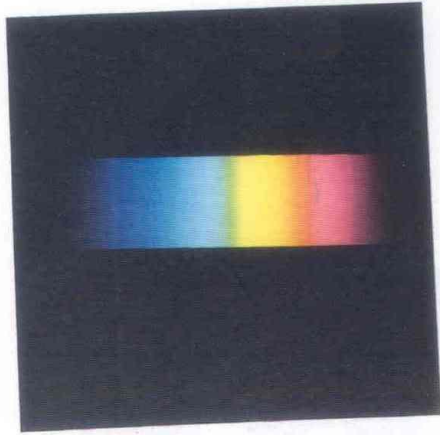


Fig. 4: Result of applying out of gamut color approximation scheme to the spectral colors.



Fig. 5: Simulation of the interference colors produced by an oil slick on wet asphalt pavement.

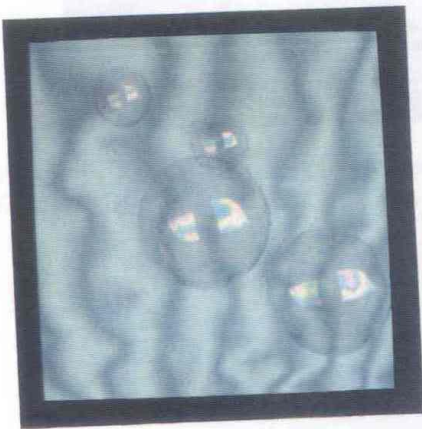


Fig. 6: Simulation of the interference colors produced on the surface of a soap bubble.

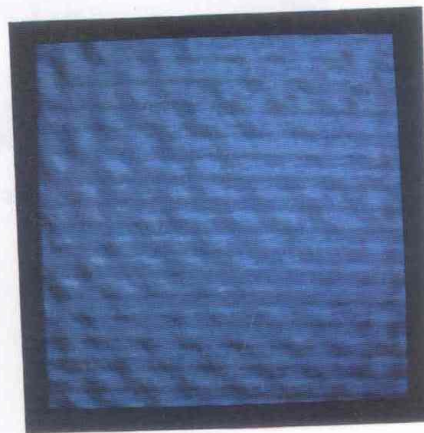


Fig. 7: Simulation of the iridescent blue color found on the wings of the *Morpho* butterfly.