Wave nature of light shapes its many properties

The transverse wave theory of light explains the characteristics of polarization, interference, and diffraction.

Thomas V. Higgins, Contributing Editor

Anyone who has spent too much time out in the sun is left with a stinging reminder of the power of solar radiation. Somehow, enough energy reaches us across the vacuum of space to redden our noses. But how? Without a physical medium to transmit the energy, how does it get here? For a long time many scientists thought that such a medium, which they called the "ether," actually existed. After all, if sound needs a medium in which to travel from one place to another, why not light?

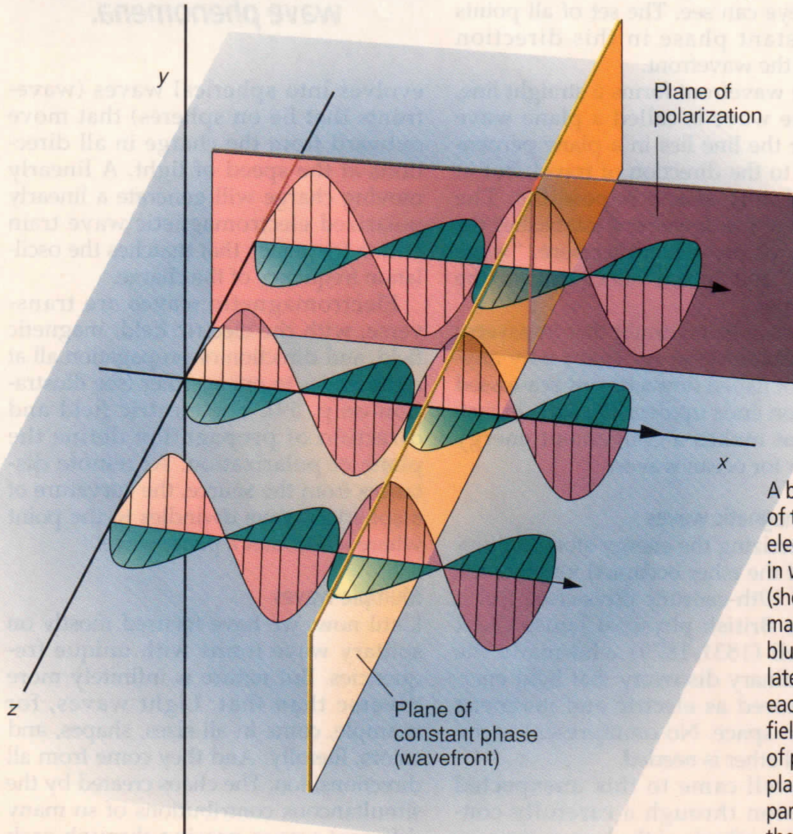
One of the first to propose the existence of an ether was René Descartes (1596-1650), who wrote, "Objects of sight communicate themselves to us only through the fact that they move locally by the intermission of transparent bodies which are between them and us." Later, Robert Hooke (1635-1703) improved on this idea by imagining light energy to be pulsed disturbances traveling through an elastic ether as ever-expanding spheres. Christian Huygens (1629-1695) went even further, postulating that each point on a sphere of disturbance in the ether creates a new, secondary spherical disturbance that propagates outward from that point. This far-reaching concept came to be known as Huygens' principle.

About the same time, however, Isaac Newton (1642-1727) proposed that light itself consisted of tiny particles moving through space and subject to the mechanical forces of matter. Newton's "corpuscular theory" validated the older and simpler belief that light is composed of physical rays that move in straight lines—an effect called rectilinear propagation.

BACK TO BASICS

Huygens' "wave/wavelet" theory could only explain rectilinear propagation through a more complicated geometrical construction of wavefronts, normals, and tangents. But his principle offered a better solution than Newton's theory to two very puzzling subtleties of light: diffraction and double refraction.

Two other peculiar effects, polarization and Newton's rings, gave trouble to Huygens and Newton, respectively. But Huygens' principle, though flawed, was eventually vindicated, and all of these strange effects (except double refraction) are better explained if light is thought of as a wave, particularly a transverse wave.



A beam of light consists of transverse waves of electromagnetic energy in which the electric (shown in pink) and magnetic (shown in blue-green) fields oscillate perpendicularly to each other. The electric field defines the plane of polarization. This plane wave is traveling parallel to the *x* axis at the speed of light.

Transverse waves

If you have sailed the high seas, you'll remember the stomach-churning experience of transverse waves. If the waves are high enough, and the boat is dead in the water, the boat bobs up and down as the swells move past it at a constant velocity. One instant you're on top of the world and the next moment you're in a valley of water.

The time that elapses between one crest and another is called the period, *T*, of the wave. The number of crests that pass by per unit of time establishes the frequency, ν, which equals $1/T$. The distance between wave crests defines the wavelength, λ. The maximum displacement of the wave above or below the equilibrium point (the water level in calm seas) specifies the amplitude, a. And the product of wavelength and frequency determines the speed of the wave, v.

The important feature to remember here is that the boat moves at right angles to the wave. This defines a transverse wave. And since the water under the boat only moves up and down along a line, the wave is said to be linearly polarized. But if the water just moves vertically, then what is it that moves horizontally? The answer is: all the points of constant phase.

The concept of phase enables us to keep track of where we are on a wave. If, for example, our boat could somehow surf on one of the wave crests, it would be at a point of constant phase. Consequently, it would stop oscillating vertically and move sideways with the wave at a steady speed.

While surfing along with the wave, we might discover that we are traveling on a seemingly infinite train of crests extending as far as the eye can see. If we looked to the side, we might see that the crest we are riding on also extends as far as the eye can see. The set of all points of constant phase in this direction defines the wavefront.

If the wavefront forms a straight line, then the wave is called a plane wave because the line lies in a plane perpendicular to the direction of travel. But in general, any shape is possible. The velocity of the wavefront establishes the velocity of the wave; therefore, "wave velocity" and "phase velocity" mean the same thing.

There can be no doubt that transverse waves contain energy. That's why anything not nailed down on our sea-tossed craft soon ends up over the side. In fact, our boat makes a convenient energy detector for ocean waves.

Electromagnetic waves

Characterizing the energy-storage properties of the ether occupied much of the time of 19th-century physicists. But it was the British physicist James Clerk Maxwell (1831-1879) who made the extraordinary discovery that light energy is stored as electric and magnetic fields in space. No omnipresent metaphysical ether is needed.

Maxwell came to this unexpected conclusion through a carefully constructed synthesis of the known theories of electricity and magnetism. He combined these concepts into four famous simultaneous equations that describe electric and magnetic fields in space and time. Light waves, according to Maxwell, are electromagnetic disturbances that travel through space as a kind of waltz performed between the electric and magnetic fields. The waltz begins, for example, when an electric charge vibrates up and down.

The continuously accelerating charge creates changing electric and magnetic fields around it, which in turn induce changing magnetic and electric fields around themselves, and so on. In this way, the energy propagates through space at the speed of light. Potential energy is stored in a capacitive form via the permittivity of space, while "inertial" energy is stored in an inductive form via the permeability of space.

Electromagnetic energy quickly

> *It is the principle of superposition that provides the key to a deeper understanding of all wave phenomena.*

evolves into spherical waves (wavefronts that lie on spheres) that move outward from the charge in all directions at the speed of light. A linearly moving charge will generate a linearly polarized electromagnetic wave train with a frequency that matches the oscillation frequency of the charge.

Electromagnetic waves are transverse, with the electric field, magnetic field, and direction of propagation all at right angles to one another (see illustration on p. 59). The electric field and direction of propagation define the plane of polarization. At remote distances from the source, the curvature of a spherical wave dwindles to the point where it becomes a plane wave.

Multiple waves

Until now, we have focused mostly on solitary wave trains with unique frequencies. But nature is infinitely more diverse than that. Light waves, for example, come in all sizes, shapes, and colors, literally. And they come from all directions, too. The chaos created by the simultaneous contributions of so many different waves passing through each

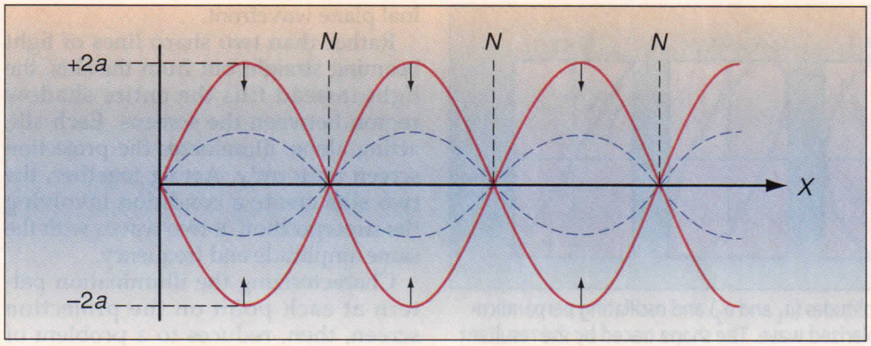


FIGURE 1. When two identical sine waves of amplitude *a* are traveling in opposite directions, the combined effect is a standing wave. Standing waves do not propagate, but instead oscillate at the antinodes with an amplitude between 2*a* and −2*a* (indicated by arrows). The resultant amplitude at the nodes, N, is always zero.

point in space would ordinarily leave us hopelessly confused. But thankfully, the principle of superposition simplifies the situation by allowing each wave's contribution to be considered separately from the others.

Any two sine waves of the same frequency and wavelength will combine as another sine wave whose phase and amplitude differ by a predictable amount. In fact, any number of sine waves with the same frequency and wavelength will add up to a single sine wave in the same predetermined way, although the math becomes a little more involved.

An especially useful superposition of waves for electro- optical applications involves two waves of the same amplitude and frequency traveling in opposite directions. The sum of these two waves is called a standing wave because it doesn't actually go anywhere (see Fig. 1). The resultant wave pattern has a maximum amplitude twice that of the constituent waves, but all of the energy is consumed by a sinusoidally changing amplitude between stationary nodes of zero amplitude. The antinodes define positions of maximum amplitude.

A true standing wave cannot transmit energy unless there is some mechanism for energy loss, and all real-world standing waves suffer some kind of energy loss. In these circumstances they are called stationary waves. Laser cavities, for instance, create stationary light waves because at least one of the mirrors reflects less than 100% of the energy back on itself.

When the amplitude, phase, and frequency of two or more waves differ, the combined effect produces a complex wave pattern. Such waves form a variety of periodic patterns depending on the magnitude of frequency and phase differences (see Fig. 2). Unscrambling these waveforms into their sinusoidal constituents defines the essence of Fourier analysis.

The superposition of two collateral transverse waves whose planes of polarization are perpendicular to each other leads to some interesting and useful effects. If the frequencies match, then the resultant amplitude traces either an ellipse or a line, depending on the phase difference between the two waves (see Fig. 3). The straight line portrays linear polarization, while the ellipses indicate elliptical polarization. The direction in which these shapes are traced as the wave moves through space also depends on the phase difference.

If the amplitudes of the two orthogonal waves are equal and the phase difference is $\pm\pi/2$, then the resultant wave becomes circularly polarized. The amplitude of such a wave remains constant as it traces a corkscrew path through space.

Interference

In 1801, Thomas Young (1773-1829) performed a convincing experiment that demonstrated the principle of superposi-

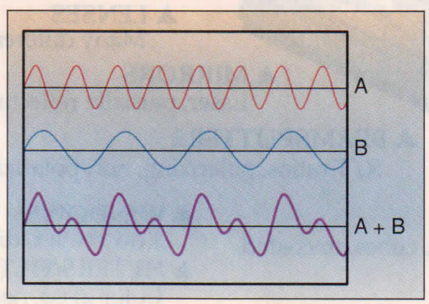


FIGURE 2. Complex wave patterns result when two or more waves having different frequencies are brought together. The complexity varies tremendously depending on the magnitude of the frequency difference.

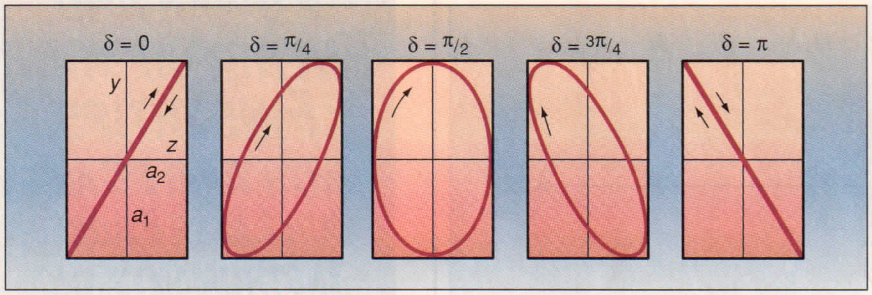


FIGURE 3. Two transverse waves with different amplitudes (a_1 and a_2) and oscillating perpendicularly to each other will combine as an elliptically polarized wave. The shape traced by the resultant amplitude as the wave moves through space depends on the phase difference between the constituent waves ($\delta = \alpha_1 - \alpha_2$).

tion and the wave nature of light. Young's famous experiment actually demonstrated two wave effects of light, interference and diffraction.

Our version of Young's experimental setup consists of just two parallel opaque screens: one has two narrow slits in it and the second serves as a projection screen (see Fig. 4). The distance between the screens, D, usually exceeds the distance between the slits, d, by a factor of about 1000 or more.

Monochromatic light (light of a single frequency) enters from the left as a plane wave, striking the screen with the slits at a right angle to its surface. If the wavelength of the light is significantly longer than the slits are wide, then light passing through the two slits will diffract into wave patterns that diverge from the slits with cylindrical wavefronts. The slits, in effect, obey Huygens' principle by acting as secondary sources from their positions on the original plane wavefront.

Rather than two sharp lines of light beaming straight out from the slits, the light instead fills the entire shadow region between the screens. Each slit, acting alone, illuminates the projection screen uniformly. Acting together, the two slits create a condition involving the superposition of two waves with the same amplitude and frequency.

Characterizing the illumination pattern at each point on the projection screen, then, reduces to a problem of finding the superposition of two waves whose phases differ by an amount dependent on the relative distance from the slits. A little trigonometry tells us that for a very large D and a small d this relative distance can be expressed as dy/D, where y is the distance of the point from the center of the projection screen. Therefore, the phase difference between the two waves is $\delta = 2\pi dy/\lambda D$.

The superposition of two sine waves tells us that the peak resultant amplitude is $A = 2a\cos(\delta/2)$. To relate this to the intensity that we would actually see projected on the screen, we square the amplitude to get the energy density:

$$I \approx A^2 = 4a^2\cos^2(\pi dy/\lambda D)$$

This shows that the intensity pattern on the projection screen will appear as equally spaced bright and dark lines called fringes whose peaks and valleys are governed by the value of $\cos^2$ as it varies between ±1 (see Fig. 4). Bright fringes appear centered at the points where the wave amplitudes reinforce each other ($\cos^2 = \pm 1$), which is called constructive interference. Dark fringes appear where the amplitudes nullify each other ($\cos^2 = 0$), which is called destructive interference.

The distance between adjacent bright lines (or adjacent dark lines), Δy, is directly proportional to the wavelength and screen distance and inversely proportional to the slit spacing ($\Delta y = \lambda D/d$). Therefore, if we know D and d, a direct measurement of Δy will give us the wavelength.

Interferometers

Young's double-slit experiment has become a milestone of experimental physics, but the apparatus itself also identifies a basic type of interferometer. Interferometers split waveforms into two or more parts and recombine them in a variety of ways for an endless array of practical applications. Optical interferometers are especially useful for making exacting measurements in metrology, optics, spectrometry, and

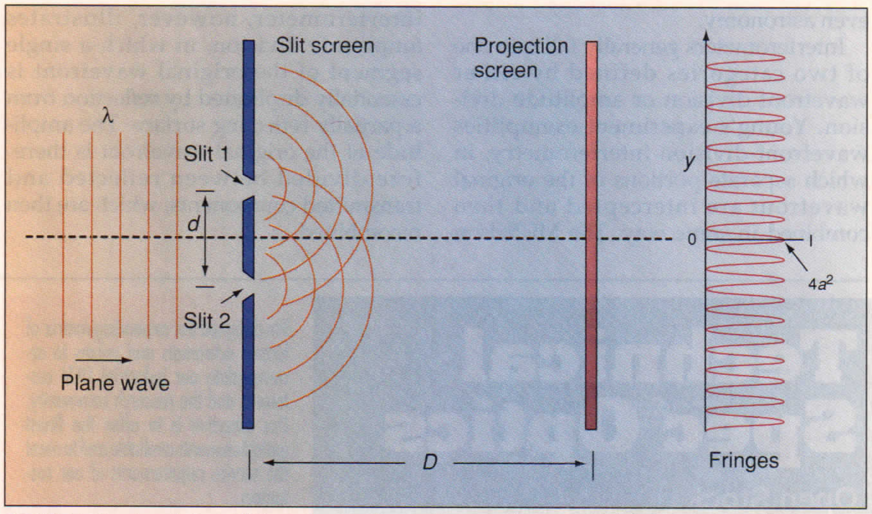


FIGURE 4. Overhead view of Young's experiment shows a plane wave hitting a screen with two very narrow slits in it (left). The diffracted light from the slits interferes, forming fringes on the projection screen (right). The fringes have a sinusoidal intensity pattern with a maximum intensity of $4a^2$.

even astronomy.

Interferometers generally fall into one of two categories defined by either wavefront division or amplitude division. Young's experiment exemplifies wavefront-division interferometry, in which separate portions of the original wavefront are intercepted and then combined in some way. The Michelson interferometer, however, illustrates amplitude division, in which a single segment of the original wavefront is essentially duplicated by reflection from a partially reflecting surface. The amplitude of the original wavefront is therefore divided between reflected and transmitted components, which are then recombined.

The American physicist Albert Michelson (1852-1931) introduced this instrument in 1881, and its contribution to physics is legendary. The basic design consists of a partial reflector (beamsplitter) and two totally reflective mirrors at right angles to each other (see Fig. 5). A plane wave of monochromatic light enters the instrument from the left, striking the beamsplitter first. For a 50% beamsplitter, half the light reflects toward one mirror, while the other half passes through the beamsplitter toward the other mirror.

When the planes of the mirrors are perpendicular and the plane of the beamsplitter is 45° to the second mirror, the back-reflected light recombines at the beamsplitter to form two sets of collateral waves: one set that returns to the source and one set that heads toward

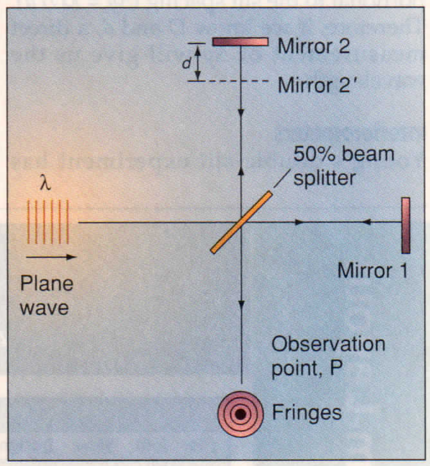


FIGURE 5. At optical wavelengths, Michelson's interferometer can measure minute dimensional changes in either leg (such as M2 to M2′) because each concentric fringe seen from point P represents λ/2. Movement of as little as λ/50 is observable with this instrument.

the observation point, P. Interference between the collateral waves will form concentric circular fringes observable from P. Should the position of one of the mirrors change, the circular system of fringes will collapse or expand, depending on which way the mirror moves. Each circular fringe represents $\lambda/2$ of movement, which makes the instrument a very sensitive one at optical wavelengths.

Michelson's original interferometer has inspired a number of specialized variations that enjoy a broad range of applications in modern electro-optics. Chief among these are the Twyman-Green and Mach-Zehnder interferometers. Other designs exploit the interference effects caused by the multiple reflections of waves from two surfaces or more. Some of these, such as the Fabry-Perot interferometer, play pivotal roles in electro-optics. And all of them rely on the principle of superposition.

Huygens, Fresnel, and diffraction

The diffraction of light, first noted by Francesco Grimaldi (1618-1663), proves that we really can see around corners. Therefore, the principle of rectilinear propagation cannot always be true. Young's double-slit experiment clearly shows this. But Young's apparatus, as described here, is a heuristic idealization, because slit widths that are significantly narrower than a wavelength of light cannot pass enough light to be of practical use. Apertures that do pass enough light span dimensions of several wavelengths. Instead of the uniform intensity distribution described earlier, the intensity pattern formed by these wider apertures is very different, and its explanation demands a more rigorous application of Huygens' principle.

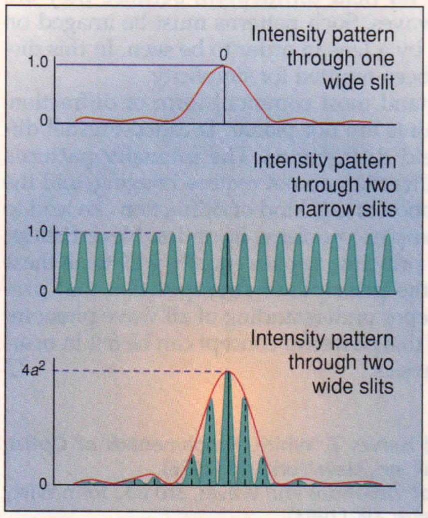


FIGURE 6. The diffraction of a plane wave through two slits whose separation is three times the width of each slit looks like the bottom intensity plot. It is the product of the interference of two narrow slits (middle curve) with the diffraction of one wide slit (top curve).

The first successful description of the diffraction pattern of a single aperture was worked out by Augustin-Jean Fresnel (1788-1827) around 1818 using Huygens' principle. Like Young before him, Fresnel applied the principle of superposition, but with a twist. He assumed that the diffracted wavefront was the resultant superposition (interference) of all the Huygens wavelets emanating from the entire width of the aperture.

Using this insight, Fresnel was able to accurately calculate the diffraction pattern of a single slit illuminated by a plane monochromatic wave (see Fig. 6, top curve). But he accomplished much more than that, because the Huygens-Fresnel principle can successfully describe the diffraction of any number of slits or apertures in various circumstances.

For example, when two slits of finite extension are used in Young's experiment, the interference pattern looks like the one shown in the bottom curve of Fig. 6. The intimate connection between diffraction and interference reveals itself if we compare the three curves in the figure. The pattern depicted in the bottom curve is a superposition of the interference pattern of two narrow slits (middle curve) with the diffraction pattern of a single wide slit (top curve). Diffraction appears to modulate the intensity of the interference fringes. An array of equidistant slits defines a diffraction grating, one of the most versatile and powerful tools of electro-optics. And the most important feature of a diffraction grating lies with its capacity to separate various wave frequencies in space.

This ability can be understood from what we already know about the interference pattern in Young's experiment with two slits. The distance between the fringes is directly proportional to the wavelength; therefore, longer wavelengths will generate fringes with wider spacings. Consequently, if the light passing through the two slits is composed of two or more wavelengths, then two or more interference patterns with different spacings will appear on the projection screen. The ultimate expression of this property occurs when sunlight encounters a diffraction grating. The resulting diffraction/interference pattern appears to the eye as a rainbow of color.

All of the intensity patterns described here illustrate Fraunhofer diffraction (or far-field diffraction), because they are formed from plane waves. Such patterns must be imaged on the projection screen by a lens in order to be seen. In this discussion, the lens has been omitted for simplicity.

The most general (and most common) form of diffraction, in which the wavefronts are not planar, is called Fresnel diffraction (or near-field diffraction). The intensity patterns formed by Fresnel diffraction do not require imaging, and the mathematical description of this kind of diffraction can lead to some of the most complex problems in optics. Nevertheless, the Huygens-Fresnel principle applies equally well to all these situations. But it is the principle of superposition that provides the key to a deeper understanding of all wave phenomena. The influence of this powerful concept can be felt in nearly every branch of physics. □

FURTHER READING

Francis A. Jenkins and Harvey E. White, *Fundamentals of Optics*, 4th ed., McGraw-Hill Inc., New York, NY (1976).

H. J. Pain, *The Physics of Vibrations and Waves*, 3rd ed., John Wiley & Sons Ltd., Chichester, UK (1983).

Richard P. Feynman et al., *The Feynman Lectures on Physics*, Addison-Wesley Publishing Co., Menlo Park, CA (1963).

David Halliday and Robert Resnick, *Fundamentals of Physics*, John Wiley & Sons Inc., New York, NY (1970).

Frank L. Pedrotti and Leno S. Pedrotti, *Introduction to Optics*, Prentice-Hall Inc., Englewood Cliffs, NJ (1987).

Max Born and Emil Wolf, *Principles of Optics*, 6th ed., Pergamon Press Ltd., Oxford, UK (1980).

Michael I. Sobel, *Light*, The University of Chicago Press, Chicago, IL (1987).

Jed Z. Buchwald, *The Rise of the Wave Theory of Light*, The University of Chicago Press, Chicago, IL (1989).

Bulk reprints of all LFWorld *articles can be ordered from June Bozarth, reprints manager, PennWell Publishing Co., at (800) 331-4463 or (918) 835-3161.*

YOUR RATING, PLEASE...
Is this article of value to you? Please circle appropriate number on the Reader Service Card.
YES—404 NO—405

