

Pulse Oximetry: Technical Aspects of Machine Design

JONAS A. POLOGE

Oximetry, in principle, is quite simple. The color of the blood is a function of the oxygen saturation irrespective of the person being tested. The color change with oxygen saturation is due to the optical properties of the hemoglobin molecule and, more specifically, the heme. As the blood deoxygenates, it becomes increasingly less permeable to red light. The tissue then loses its pinkish appearance, taking on a blue tint. An oximeter only needs to measure the "blueness" of the arterial blood, ignoring the patient's natural pigmentation, the venous blood, and any other major absorbers of light between the two halves of the probe, and display this blueness in terms of saturation.

THEORY

An investigation of machine design requires an understanding of the hemoglobin extinction curves, the existence of which is the basis of oximetry. The two curves shown in Figure 1 are for oxyhemoglobin and reduced (or deoxygenated) hemoglobin. The x axis is the wavelength of the light extending from the 450-nm blue region of the spectrum up to 1,000 nm in the near-infrared region. The y axis is the extinction coefficient on a logarithmic scale. Extinction is the light absorption of a unit concentration and path length of a given substance. On a more intuitive level, one can see that oxyhemoglobin is relatively transparent in the red region, from ap-

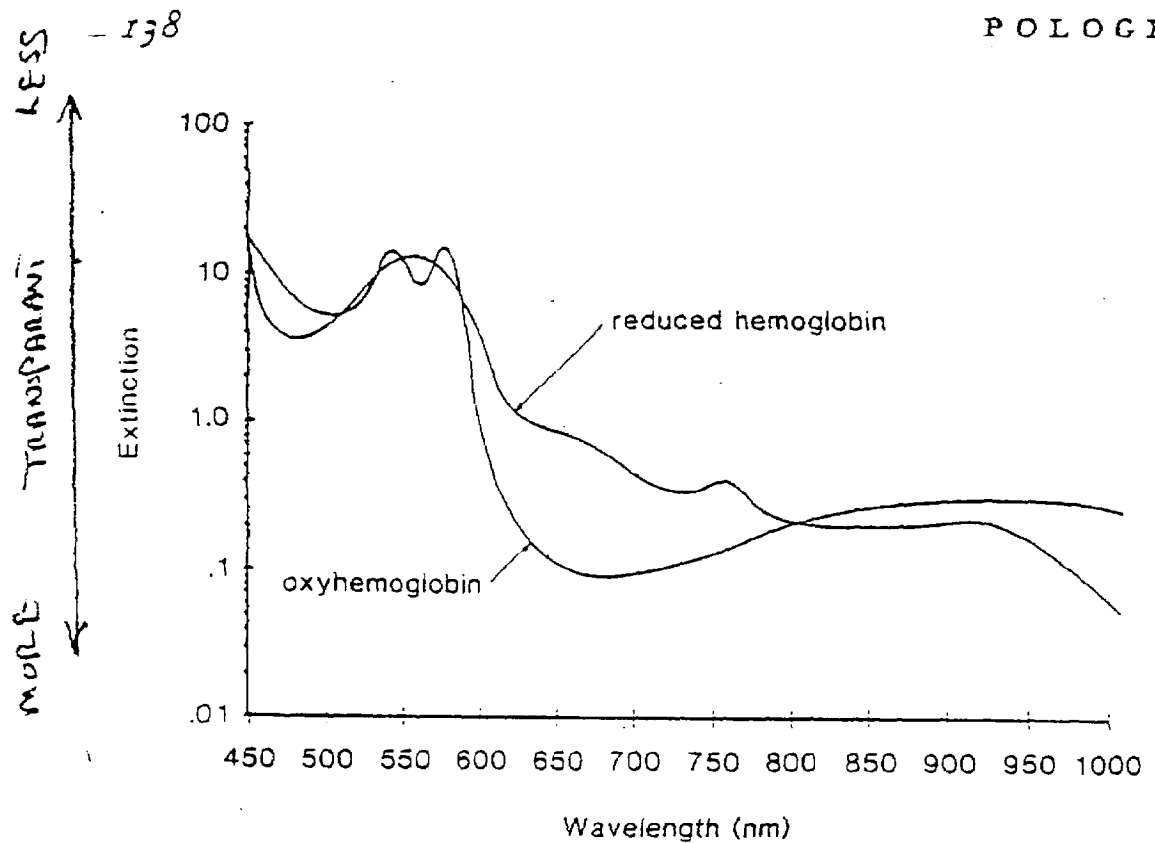


FIG 1. Hemoglobin extinction curves.

proximately 600 nm out to 1000 nm in the infrared, whereas at the shorter wavelengths the hemoglobin of both species is considerably more opaque. Thus, the hemoglobin acts as a filter, allowing only the red and near-infrared light to pass through. This explains why white light incident on one side of the hand emerges only as red light on the opposite side.

The second ingredient necessary to a discussion of oximetry is the Beer-Lambert-Bouguer law, or more simply, Beer's law. This law states that the total absorption of a system of absorbers is the sum of their independent absorbances, or:

$$A_{\text{total}} = E_1 C_1 L_1 + E_2 C_2 L_2 + \dots + E_n C_n L_n \quad (1)$$

In this equation, A_{total} is the absorbance of a mixture of substances at a specific wavelength; E_n is the extinction of substance n (at that wavelength); C_n is the concentration of substance n ; and L_n is the path length of the light through substance n .

Hewlett-Packard (Palo Alto, CA) took the obvious approach to

solving the problem of noninvasive oximetry. All organic molecules have their own unique extinction curves. In the human body, the various species of hemoglobin and skin pigmentation are the primary absorbers of light. Hewlett-Packard invented an instrument that used eight wavelengths of light ranging from 650 to 1,050 nm. This gave a system of eight linear equations, such as Eq. (1).

Although one can only guess at the inventors' methodology, they probably used a mixture of practical theoretical knowledge and empirical studies to come up with the eight-wavelength instrument. Starting with as few wavelengths as was thought might work, they progressively added one at a time until a good correlation with actual oxygen saturation was obtained. With each additional wavelength there is an additional equation and therefore an additional unknown that might be solved for, assuming one is using a portion of the spectrum that contains the desired information. The substances to be identified must all absorb at the wavelengths used, and each must have its own uniquely different extinction curve over the given spectral region.

The implementation of the Hewlett-Packard concept of oximetry resulted in an oximeter measuring about $19 \times 43 \times 43$ cm and weighing about 17 kg. With a rotating filter wheel containing eight interference filters, it employed fiberoptics to conduct the light to and from the probe. The probe itself clamps over the pinna of the ear and surrounds the head like a lopsided bicycle helmet.

To measure absorption one has to know the intensity of both the incident light and the output light. The difference between these is the light that was absorbed. (There is also scattered light, but that will not be dealt with here.) To accomplish this, Hewlett-Packard separated the light into two distinct paths by means of a beam splitter. One path is through the optics of the probe and then back to the detector, and the other runs directly to the detector.

The Hewlett-Packard oximeter worked fairly well and was the first real introduction of oximetry into clinical practice, primarily in respiratory therapy. Its major problems were size, cost, sensitivity to motion artifact, and a limited patient base on which it worked effectively. If the ear was too darkly pigmented or too poorly per-

fused, it could not obtain a reading. Incidentally, the Hewlett-Packard oximeter did not actually differentiate between venous and arterial blood. Hewlett-Packard eliminated the effects of venous blood by including a heater in the probe to cause arteriolization, thereby bringing the venous blood to arterial saturation levels.

In recent years, the science of oximetry has taken an enormous intuitive and technological leap with the creation and refinement of pulse oximetry. Pulse oximetry is not only a powerful monitoring technology but is also elegant in its inherent simplicity.

According to Beer's law, at least n wavelengths are required to identify any one absorber of light out of a system of n absorbers. Pulse oximeters observe only the arterial blood. Since blood is composed of primarily two absorbers, oxyhemoglobin and reduced hemoglobin, only two wavelengths of light are required. Another advantage of being able to observe specifically arterial blood is the lack of need for a heater to equilibrate arterial and venous saturation, as is done in the Hewlett-Packard device.

The wavelengths used in pulse oximetry are typically around 660 nm in the red region and anywhere from 800 to 1,000 nm in the near-infrared region. For this discussion, the second wavelength will be assumed to be 940 nm, although any wavelength used in the previously stated range could be made to work. Figure 2 shows the change in absorption that would be observed at the two wavelengths as the saturation increases from 0 to 100%. Notice that these two lines result directly from the extinction curves of Figure 1 and Beer's law (assuming the path length is constant and the ratio of oxyhemoglobin to reduced hemoglobin is the variable term). From these two curves a third can be generated with saturation as a function of the ratio (R) of absorbance at 660 nm (A_{660}) to absorbance at 940 nm (A_{940}) (Fig 3). This curve is termed the calibration curve. All of the current oximeters' electronics and programming are designed to obtain this ratio as accurately as possible. R is then simply converted into arterial oxygen saturation using the equation for the calibration curve.

Because the optics of the earlobe (or finger, or whatever tissue is being monitored) and the optics of the probe do not exactly match the requirements of Beer's law, the calibration curve in currently

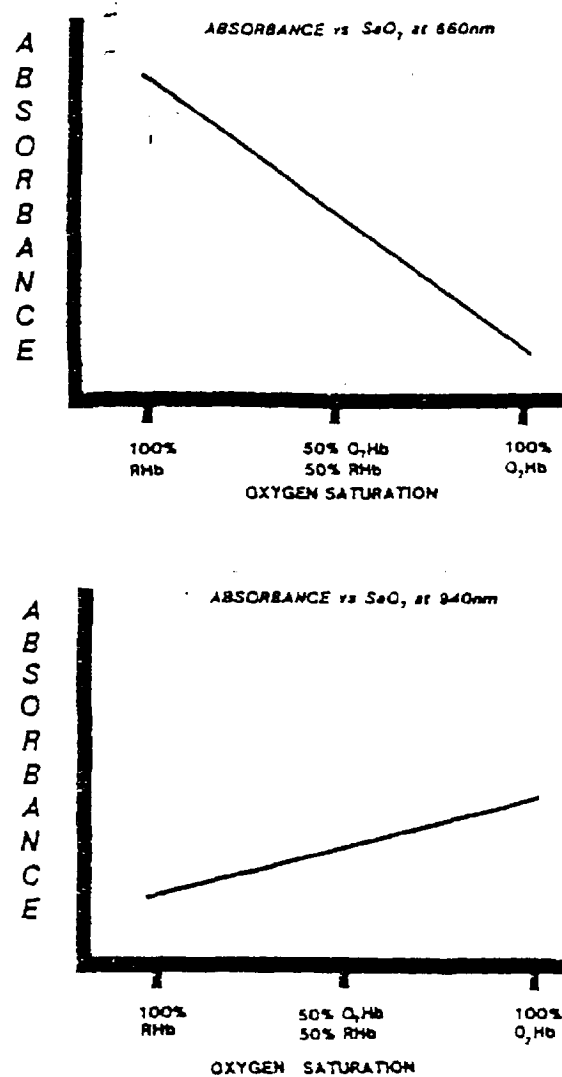


FIG 2. Changes in the absorption of oxygenated (O₂Hb) and reduced hemoglobin (RHb) as a function of arterial oxygen saturation (SaO₂) at 660 and 940 nm.

available oximeters is derived empirically by correlating R to invasive arterial oxygen saturation measurements. One difference between ideal optical conditions and those encountered in vivo in oximetry is the nonhomogeneity of the absorbing system in that the absorbers are confined to red blood cells. Because the cell walls have a different index of refraction than does the plasma, light passing through whole blood is not only absorbed, but also scattered. Although empirical calibration does account for most of the optical problems of in vivo oximetry, it is by no means perfect and will probably always lag behind in vitro measurement of saturation in terms of absolute accuracy.

To understand pulse oximetry one must first understand how

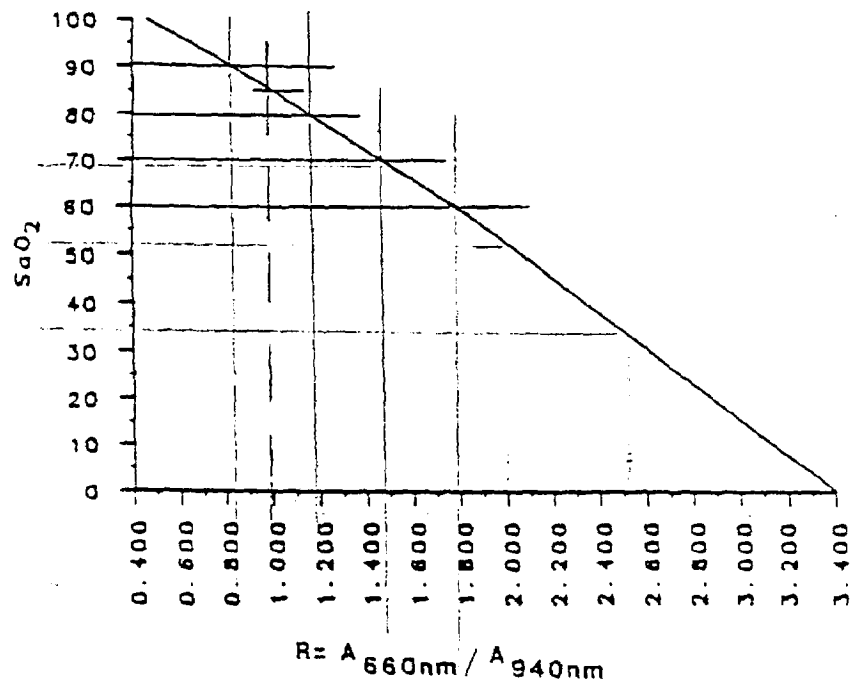


FIG 3. Calibration curve used by the oximeter to calculate arterial oxygen saturation (SaO_2) from the ratio (R) of the light absorbed (A) by the tissue being monitored. (Figure provided courtesy of Ohmeda, 1987.)

absorbance is measured. In reality, it is not. The Hewlett-Packard oximeter was able to measure both the incident light levels and the transmitted levels, allowing direct calculation of absorption. Present-day oximeters look only at the transmitted light: What they measure is more of a relative absorption at the two wavelengths.

An examination of the output light at each wavelength reveals that it consists of two components. The first component varies with pulsation of the blood. The second is a large, constant light output level; this is the light that passed through the tissues without being absorbed or scattered (Fig 4). These are referred to as the *AC* and *DC components*, respectively. The amplitude of both the DC and AC levels are directly dependent on the incident light intensity. Dividing the AC level by the DC level (at each wavelength) gives a corrected AC level that is no longer a function of the incident intensity. This corrected AC level is a function only of the combined extinction of the two species of hemoglobin and the path length of

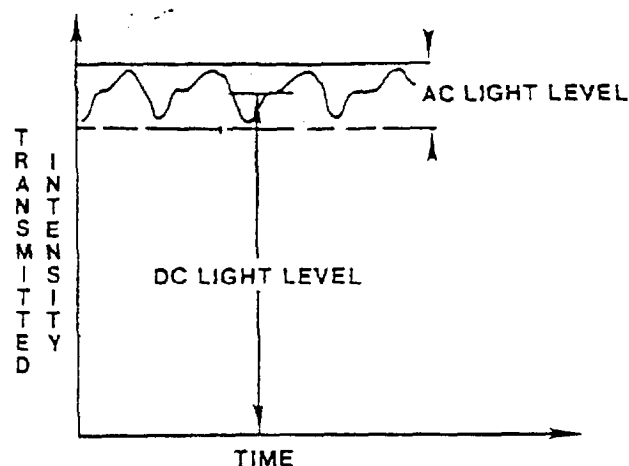


FIG 4. With a constant-intensity light incident on living tissue, the output light consists of a constant light level with a small amount of modulation caused by the pulsating blood.

arterial blood through which the light has passed. The AC light is only a function of the arterial blood since essentially only the arterioles are pulsating in the light's path. The previously described ratio uses the corrected change in AC light levels. Thus:

$$R = (AC_1/DC_1)/(AC_2/DC_2) \quad (2)$$

In this equation, 1 is the red wavelength and 2 is the infrared wavelength.

By using the pulsatile (or AC) light, pulse oximetry effectively ignores the absorbances of venous blood, tissue, and pigmentation. This reduces the problem of saturation measurement in vivo to a two-component, and therefore two-wavelength, system.

DESIGNING THE OXIMETER

Now that the theory of operation has been defined, the design requirements needed to actualize such an instrument can be detailed. The discussion begins with the probe requirements and problems.

The Probe

The probe must provide two narrow-band sources of light at the desired wavelengths. One solution, as described in the discussion of the Hewlett-Packard oximeter, uses a broad-band light source

broken into individual bands of light by interference filters. This solution commands the use of bulky fiberoptic cables to transport the light to and from the oximeter, but it also provides an easy method for obtaining multiple wavelengths of light with very well-defined bandwidths. With only two wavelengths needed, a cheaper, more lightweight technology is available through the use of light-emitting diodes (LEDs). Because these devices are extremely small, $10/1,000$ of an inch square, and need only a minuscule amount of current to drive them, the light sources are an integral part of the probe and require only two or three very thin flexible wires running back to the oximeter to control them.

The choice of LEDs is dependent on the wavelengths desired, the wavelengths available, and the intensities available. A large number of very bright LEDs are available centered at 660 nm in the red region. Brightness is crucial because the light must be able to penetrate the thickest, most darkly pigmented tissue and appear on the opposite side with sufficient intensity to allow for accurate measurement. The LED is truly "the device fantastic." Requiring considerably less power than the average flashlight and generating very little heat, the LED emits a monumental amount of light over an extremely narrow portion of the electromagnetic spectrum. The half-power bandwidth of many red LEDs is less than 20 nm, yet over this region they can emit more than 20 mW of power. By comparison, the sun on a clear day in Boulder, Colorado, strikes the earth with less than 9 mW/cm^2 over the same spectral region.

These LEDs provide sufficient power to barrel through a darkly pigmented, edematous infant foot. However, this same incident light intensity would saturate (or overload) the photo detector if it were placed across a thin, lightly pigmented earlobe. To avoid this difficulty and to keep the transmitted light at the optimal intensity for maximum signal-to-noise ratios, the LED, whose output intensity varies almost linearly with drive current, is simply turned down. The range of usable intensities easily exceeds 100 to 1. To adjust LED intensities in this manner, the oximeter must be "intelligent" enough to recognize and control received light levels by controlling the LED drive current.

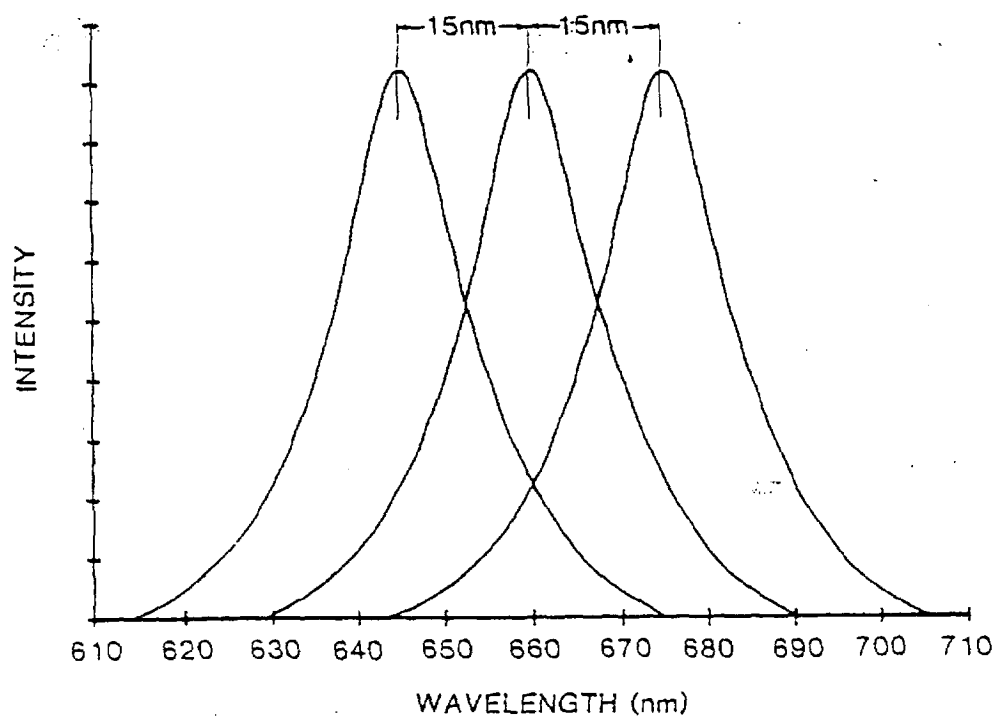


FIG 5. With different light-emitting diodes of the same type from the same manufacturer, the center wavelength of the output light can vary ± 15 nm.

As might be expected from a device so uniquely suited to a specific application, the use of LEDs in oximetry comes complete with its own set of difficulties. The largest problem in the mind of the oximeter designer is the variation in the center wavelength (or peak wavelength) of different LEDs of the same type. Virtually no two LEDs have the same center wavelength. For any given type of LED, the variation in center wavelengths can be as great as ± 15 nm (Fig 5). From the extinction curves shown in Figure 1, it is evident that any shift in wavelength of the light source leaves the oximeter working with a different set of extinction coefficients. This change in extinction appears in Eq. (1) and results in an erroneous calibration curve and therefore an error in the measured oxygen saturation. The change in extinction is most pronounced at the 660 nm wavelength on the reduced hemoglobin curve due to the extremely steep slope of the extinction curve in this vicinity. (Remember that the extinction curves shown display the y axis on a logarithmic scale.)

The lower the oxygen saturation, the more heavily the reduced hemoglobin is weighted and the greater the error in the saturation that is measured with an off-wavelength red LED. This is one reason why the accuracy of oximeters is always worse at lower saturations. It is important clinically to realize that, although the absolute accuracy of saturation measurements does decrease with decreasing saturation, the capability of any given set of oximeter, probe, and patient to trend accurately is unaffected. Further, this inherent inaccuracy only becomes appreciable at reduced saturation levels (i.e., less than approximately 80%), while clinically the absolute accuracy is usually of less importance than raising the patient's saturation. The precise wavelength used in the infrared region is of less concern. The extinction curves in this area are flatter, and any shift in LED center wavelength causes a smaller error than would be generated by the same shift in the red LED.

Since LEDs do vary more than is acceptable, the oximeter manufacturer is left with several possible solutions. First, one could ignore the inaccuracies caused by wavelength shift, and if only one oximeter manufacturer exists, this might work, at least until someone created a more accurate system. Second, one could use only LEDs whose outputs fall within a very narrow range of center wavelengths. This solution is quite expensive in terms of unusable LEDs and the time required to test them. A third and considerably more elegant solution is to compensate mathematically for a number of different LED center wavelengths (maybe in 1-nm steps). A family of calibration curves is generated, one for each center wavelength set of two LEDs. Note that if there are 10 ranges for the red LEDs and three for the infrared LEDs, any infrared LED might be paired with any red LED. Thus, a total of 30 calibration curves are required, one for each set. This methodology makes it essential for the probe to identify itself to the oximeter so that the correct calibration curve may be used for that particular LED set. Although there are many ways to do this, one common technique is to code each probe with a fixed resistor that the oximeter checks when the probe is plugged in. Each different resistance value is matched with a specific calibration curve. The clinician pays for the accuracy of oximetry in terms of probe costs. Each probe must have its LEDs

carefully tested, matched with the proper resistor, and assembled—a technologically sophisticated and rather time-consuming process.

That pulse oximetry is possible at all is only an accident of evolution and semiconductor technology. The electromagnetic spectrum extends over 10^{14} nm. LEDs are only available from 560 nm in the green to 950 nm in the near-infrared region, and then only at roughly twenty distinctly different wavelengths. By accident, the window of low absorption in the hemoglobin extinction curves extends from about 600 to 1,000 nm. In this region, 660 nm is the single best wavelength to use for two reasons. The oxyhemoglobin extinction curve is relatively flat, so that errors due to small shifts in center wavelength are minimal, and the separation between the two extinction curves is large, yielding detectable changes in absorption with small changes in oxygen saturation. By coincidence, 660 nm is one of the wavelengths at which LEDs are manufactured. In the selection of the second wavelength, one criterion is essential. The extinction curves must be as different as possible from their position at the first wavelength, perhaps even inverted, giving a large change in absorption ratio (R) for any change in saturation. Obliging, the extinction curves cross at 804 nm, and are once again nicely separated at 940 nm (where LEDs are also manufactured).

One commonly asked question is why the isobestic point at 804 nm is not used for the second wavelength. The use of an isobestic point for pulse oximetry gains nothing, since using a point at which the extinction curves coincide does not increase the change in R with any change in saturation. Therefore, using 804 nm for the second wavelength instead of 940 nm would actually reduce the accuracy of the oximeter.

Selection of a transducer to convert the transmitted light into an electronic signal is straightforward. The silicone photodiode has a large dynamic range (it has an output linearly proportional to the incident light level over a range of 10 decades of light intensity) and low noise levels, is lightweight and small, and responds to light through the visible region up to about 1,100 nm. The photodiode is universally used as the means of light detection in pulse oximetry.

The photodiode does not “know” one wavelength of light from

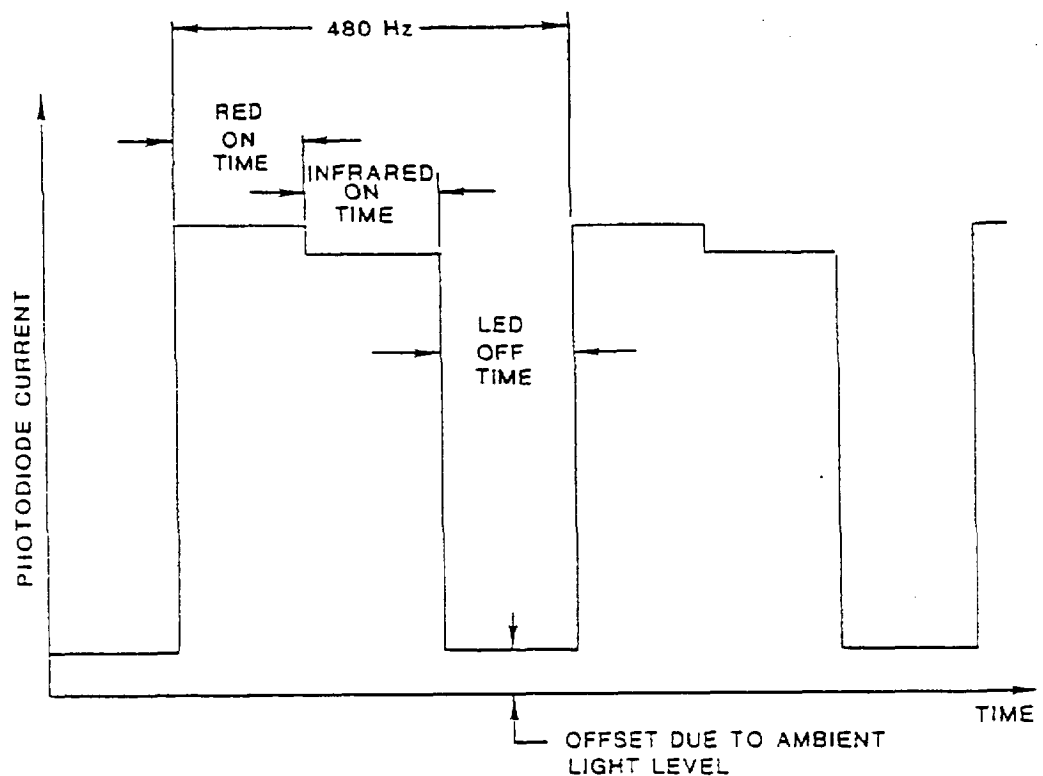


FIG 6. Output of the photo detector on the Ohmeda 3700 pulse oximeter. (LED = light-emitting diode.)

another. Consequently, if both the red and infrared LEDs were on at the same time, the output current of the photodiode would be proportional to the sum of the two incident intensities. The oximeter needs to know the light output at each wavelength. To obtain an output that is a function of only one LED at a time, the oximeter alternately pulses the LEDs on and off—first turning on one LED, then turning it off and the other on, then turning them both off. On the Ohmeda Biox 3700 oximeter (Boulder, CO) this cycle is repeated at a rate of 480 times per second (480 Hz). This cycling allows the oximeter to know which LED the photodiode is watching at any instant in time. The output of the photodiode is shown in Figure 6.

This scheme solves another troublesome problem: ambient light noise. Any light incident on the detector not originating at the LEDs is considered noise. This noise would artificially boost the DC levels as defined in Eq. (2), once again introducing error into

the system. The sources of ambient light are obvious: the sun, room lights, operating theater lights, phototherapy lights. What is not so obvious is how the light reaches the detector. After all, the detector is supposed to be totally covered by the tissue being measured. In addition, the clinician may have wrapped the tissue and probe to further block ambient light. Tissue, however, is a highly light-scattering medium. When light enters the tissue, it heads off in all directions, including, for example, down the axis of the finger to the probe detector. By having a short interval when both LEDs are off, the oximeter can then measure the ambient light level and subtract it from the levels obtained when the LEDs are on. Then, just as two telephone conversations can be chopped up and sent over the same telephone line and then reconstructed and sent on to their final destinations, the output of the photodiode is reconstructed, or demultiplexed back, to the signal from each individual LED for further processing.

Patient Signal

The oximeter must work with a signal—the light transmitted by the LEDs—that is attenuated by the tissue, bone, cartilage, venous blood, and pulsating arterial blood and finally received and converted into current by the photodiode. As explained earlier, the signal has an AC and a DC component. The DC component is easily controlled by adjusting the LEDs' output intensities. Thus, the oximeter can set the DC component to an easily and accurately measured level. But what of the AC component? An increase in the LED output increases both the DC and AC levels, but how large is the AC component? It varies enormously, from less than 0.01% to greater than 10% of the DC level. This percentage is referred to as the *percent modulation*. The sources of this variability include the vascularity and perfusion of the tissue being measured. One case in which the perfusion is extremely low but under which, given the right circumstances, the oximeter has been seen to function accurately is in cardiac bypass surgery, when the peristaltic pump in the heart-lung machine is supplying the pulse to the patient's hemodynamic system.

To get a feeling for what this range of AC levels means, consider

a car radio for which the volume on any station could range from 1,000 to 1 and you have no idea what level the volume will be when you change stations. Fortunately, this is taken care of automatically and all stations come in at about the same volume, alleviating the listener of the task of adjusting the volume each time the station is changed. The oximeter also has automatic gain control to get the AC signal into a usable range, but there is a problem with this. Once again the radio analogy serves us well. When a weak station is picked up, the output volume remains unchanged but the percentage of the sound that is static, or noise, increases dramatically. In any electronic system there is noise. The smaller the signal, the greater the gain required and the more evident the noise becomes. In reference to the car radio, one finds the increase in noise makes it increasingly difficult to discern the notes. In the case of the oximeter, noise degrades the accuracy and stability of the readings. Thus, it behooves the clinician to find and maintain a probe site with adequate vascularization and perfusion.

Filters

The signal information of concern to the oximeter consists only of transmitted light modulated by the pulsating blood. Because this pulsation is generated by the beating heart, one can reasonably assume that it will not exceed 5 Hz (i.e., 300 beats/min). A Fourier transform of the plethysmographic waveform reveals that the vast majority of the information is contained in the first 10 Hz. Knowing this and knowing that electronic noise is prevalent in all electronic systems, often at high frequencies, the oximeter designer will usually limit the signal that the oximeter has to work with to be less than about 10 Hz through the use of a low-pass filter. Electronic noise can be generated by the 60-Hz power lines, electrosurgical equipment, and any other electronic equipment in the vicinity. Especially of concern are high-current devices such as motors.

Once the signals are properly amplified and filtered, they are converted to digital values by an analog-to-digital converter and fed into the microprocessor, where the ratio in Eq. (2) is computed. R is then converted into an oxygen saturation value using the calibration curve.

Although the AC and DC levels for each channel are the only data the microprocessor receives, it can be programmed to do quite a bit with them. If the light levels are too high, then no tissue is between the emitter and detector in the probe and the oximeter can display the information as "PROBE OFF PATIENT." If gains are at maximum and the LEDs are being driven as hard as possible but the signal is still too small to "see," then the oximeter can display "INSUFFICIENT LIGHT," indicating that the tissue in the probe is causing too much attenuation. This attenuation could be due to misalignment of the probe halves, dark fingernail polish, or too much tissue to read through, along with a number of other possibilities. The AC data from one channel can also be used to calculate the pulse rate by locating the peaks in the AC waveform. Some inherent difficulties in discerning pulse rate will be discussed later.

Signal Processing

In the land of oximetry, noise comes in many forms. Electronic and light noise have already been mentioned, but a much more devious form of noise is motion artifact. This is noise generated by patient motion, thereby varying the path length of the light, but not by arterial pulsation. The quantity of motion required to disturb the signal is extremely small. If the probe were placed across the finger, motion could be caused by even the slightest flexing of the hand. Shivering can generate such large quantities of motion artifact as to render the actual signal undetectable. Coughing can send pressure waves through the venous system many times larger than the arterial waveform. This motion interferes because the oximeter sees a pulse of hemoglobin, at venous saturation levels, that perturbs the calculations.

As mentioned earlier, noise is often filtered out quite easily, however, its removal requires that the noise differ considerably from the signal. The difficulty with motion artifact is that it is often at frequencies well within normal physiological ranges. Figure 7 shows some of the "normal" signals an oximeter might have to accept and one generated by shivering. Clearly "normal" includes quite a range of frequencies, waveshapes, and amplitudes. To try to program an oximeter to unmistakably discern a valid signal from motion or

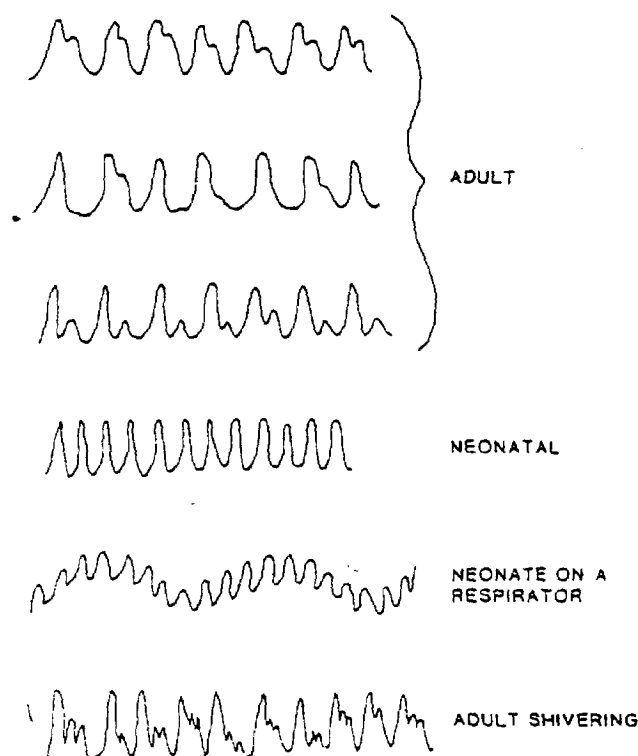


FIG 7. The variability of input waveforms received by the pulse oximeter.

noise by either analog or digital means would be an astronomical feat. However, judging by the ease with which a person can separate the motion from the signal when viewing the waveform, it is hard to see why this task would be difficult for a computer to master. Without spending many pages on a technical treatise of why this task is so difficult and perhaps still not convincing anyone, a more intuitive explanation will be presented. Humans are the world's most sophisticated signal processors ever created. Two capabilities which we take for granted that no computer has ever been able to match except in limited ways are reading cursive writing and understanding the spoken word. This is not due to a lack of effort. The benefits and profitability of a computer that could accept voice data entry from anyone have been and continue to be powerful motivators for some of the finest minds and computer laboratories in the world.

With this rather negative picture, what then can be done about motion artifact? One thing seems painfully obvious. If the human mind is such a fine signal processor, present it with the data and let

it do the work. This is what was done on the Ohmeda Biox 3700 oximeter with the plethysmographic waveform display. The clinician can always see when the input data are signal and when they are noise and can then choose to accept or reject the corresponding saturation or pulse rate data. This technique assures the clinician of the capability to take readings only on arterial pulsation signal data, but it does nothing to stabilize the saturation readings during noisy conditions. For this condition the oximeter designers employ techniques such as averaging the saturation data over time in the hopes that the noise condition will be transient. Another technique is to have the oximeter observe the input data over time. If the data are truly from arterial pressure variations, they will start to show a pattern in terms of the changes in amplitude of the signal from moment to moment. If large variations in the signal amplitude suddenly occur, then the oximeter can weight this portion of the input data very low in a weighted averaging scheme. Of course, as one extends the averaging time used, one trades the response time of the device for stability.

Any motion artifact that is consistent and of notable duration has a fascinating effect on pulse oximetry. The calculated saturation tends toward approximately 85%. Motion has a tendency to introduce a signal of approximately the same amplitude into both channels (the red and the infrared). Furthermore, the absolute amplitude of the motion-induced signal tends to be quite large. If the DC levels are close to the same, then the R takes on a value close to 1. Converting an R of 1 to a saturation value using the calibration curve in Figure 3 yields about 85%. This is a somewhat unfortunate consequence of the physics of oximetry, since it can lead the untrained to false conclusions about the patient's condition.