2

Photons:
Particles of Light

This is the second in a series of lectures about quantum electrodynamics, and since it's clear that none of you were here last time (because I told everyone that they weren't going to understand anything), I'll briefly summarize the first lecture.

We were talking about light. The first important feature about light is that it appears to be particles: when very weak monochromatic light (light of one color) hits a detector, the detector makes equally loud clicks less and less often as the light gets dimmer.

The other important feature about light discussed in the first lecture is partial reflection of monochromatic light. An average of 4% of the photons hitting a *single* surface of glass is reflected. This is already a deep mystery, since it is impossible to predict which photons will bounce back and which will go through. With a *second* surface, the results are strange: instead of the expected reflection of 8% by the two surfaces, the partial reflection can be amplified as high as 16% or turned off, depending on the thickness of the glass.

This strange phenomenon of partial reflection by two surfaces can be explained for intense light by a theory of waves, but the wave theory cannot explain how the detector

makes equally loud clicks as the light gets dimmer. Quantum electrodynamics "resolves" this wave-particle duality by saying that light is made of particles (as Newton originally thought), but the price of this great advancement of science is a retreat by physics to the position of being able to calculate only the *probability* that a photon will hit a detector, without offering a good model of how it actually happens.

In the first lecture I described how physicists calculate the probability that a particular event will happen. They draw some arrows on a piece of paper according to some rules, which go as follows:

—GRAND PRINCIPLE: The probability of an event is equal to the square of the length of an arrow called the "probability amplitude." An arrow of length 0.4, for example, represents a probability of 0.16, or 16%.

—GENERAL RULE for drawing arrows if an event can happen in alternative ways: Draw an arrow for each way, and then combine the arrows ("add" them) by hooking the head of one to the tail of the next. A "final arrow" is then drawn from the tail of the first arrow to the head of the last one. The final arrow is the one whose square gives the probability of the entire event.

There were also some specific rules for drawing arrows in the case of partial reflection by glass (they can be found on pages 26 and 27).

All of the preceding is a review of the first lecture.

What I would like to do now is show you how this model of the world, which is so utterly different from anything you've ever seen before (that perhaps you hope never to see it again), can explain all the simple properties of light that you know: when light reflects off a mirror, the angle of incidence is equal to the angle of reflection; light bends when it goes from air into water; light goes in straight lines;

light can be focused by a lens, and so on. The theory also describes many other properties of light that you are probably not familiar with. In fact, the greatest difficulty I had in preparing these lectures was to resist the temptation to derive all of the things about light that took you so long to learn about in school—such as the behavior of light as it goes past an edge into a shadow (called diffraction)—but since most of you have not carefully observed such phenomena, I won't bother with them. However, I can guarantee you (otherwise, the examples I'm going to show you would be misleading) that *every* phenomenon about light that has been observed in detail can be explained by the theory of quantum electrodynamics, even though I'm going to describe only the simplest and most common phenomena.

We start with a mirror, and the problem of determining how light is reflected from it (see Fig. 19). At S we have a source that emits light of one color at very low intensity (let's use red light again). The source emits one photon at a time. At P, we place a photomultiplier to detect photons. Let's put it at the same height as the source—drawing arrows will be easier if everything is symmetrical. We want to calculate the chance that the detector will make a click after a photon has been emitted by the source. Since it is possible that a photon could go straight across to the detector, let's place a screen at Q to prevent that.

Now, we would expect that all the light that reaches the detector reflects off the middle of the mirror, because that's the place where the angle of incidence equals the angle of reflection. And it seems fairly obvious that the parts of the mirror out near the two ends have as much to do with the reflection as with the price of cheese, right?

Although you might *think* that the parts of the mirror near the two ends have nothing to do with the reflection of the light that goes from the source to the detector, let
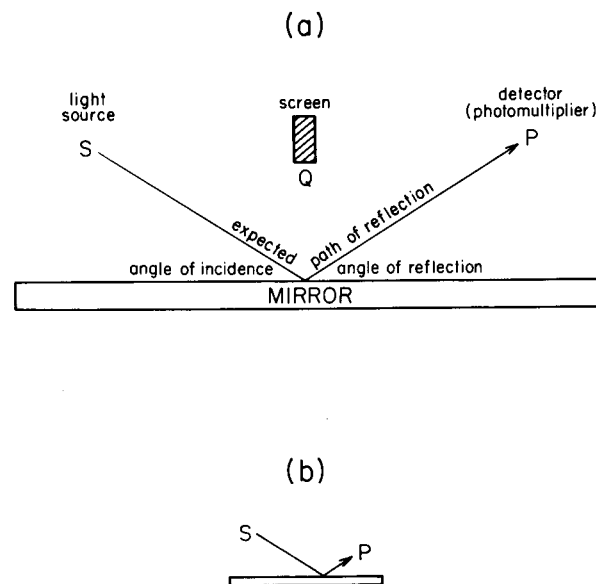
FIGURE 19. *The classical view of the world says that a mirror will reflect light where the angle of incidence is equal to the angle of reflection, even if the source and the detector are at different levels, as in (b).*

us look at what quantum theory has to say. Rule: The probability that a particular event occurs is the square of a final arrow that is found by drawing an arrow for each way the event could happen, and then combining ("adding") the arrows. In the experiment measuring the partial reflection of light by two surfaces, there were two ways a photon could get from the source to the detector. In this experiment, there are millions of ways a photon could go: it could go down to the left-hand part of the mirror at A or B (for example) and bounce up to the detector (see Fig. 20); it could bounce off the part where you think it should, at G; or, it could go down to the right-hand part of the mirror at K or M and bounce up to the detector. You might think
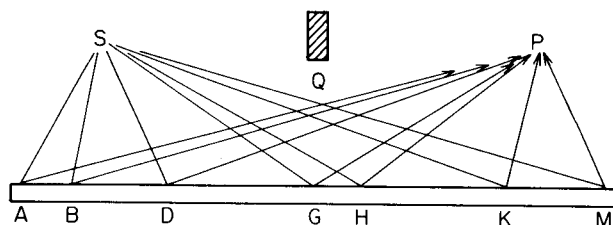
FIGURE 20. *The quantum view of the world says that light has an equal amplitude to reflect from every part of the mirror, from A to M.*

I'm crazy, because for most of the ways I told you a photon could reflect off the mirror, the angles aren't equal. But I'm *not* crazy, because that's the way light really goes! How can that be?

To make this problem easier to understand, let's suppose that the mirror consists of only a long strip from left to right—it's just as well that we forget, for a moment, that the mirror also sticks out from the paper (see Fig. 21). While
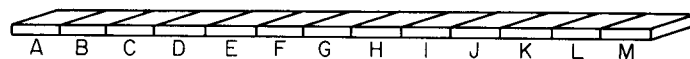

FIGURE 21. *To calculate more easily where the light goes, we shall temporarily consider only a strip of mirror divided into little squares, with one path for each square. This simplification in no way detracts from an accurate analysis of the situation.*

there are, in reality, millions of places where the light could reflect from this strip of mirror, let's make an approximation by temporarily dividing the mirror into a definite number of little squares, and consider only one path for each square—our calculation gets more accurate (but harder to do) as we make the squares smaller and consider more paths.

Now, let's draw a little arrow for each way the light could go in this situation. Each little arrow has a certain length and a certain direction. Let's consider the length first. You might think that the arrow we draw to represent the path that goes to the middle of the mirror, at G, is by far the longest (since there seems to be a very high probability that any photon that gets to the detector must go that way), and the arrows for the paths at the ends of the mirror must be very short. No, no; we should not make such an arbitrary rule. The right rule—what actually happens—is much simpler: a photon that reaches the detector has a nearly equal chance of going on *any* path, so all the little arrows have nearly the same length. (There are, in reality, some very slight variations in length due to the various angles and distances involved, but they are so minor that I am going to ignore them.) So let us say that each little arrow we draw will have an arbitrary standard length—I will make the length very short because there are many of these arrows representing the many ways the light could go (see Fig. 22).

FIGURE 22. *Each way the light can go will be represented in our calculation by an arrow of an arbitrary standard length, as shown.*

Although it is safe to assume that the length of all the arrows will be nearly the same, their directions will clearly differ because their timing is different—as you remember from the first lecture, the direction of a particular arrow is determined by the final position of an imaginary stopwatch that times a photon as it moves along that particular path. When a photon goes way off to the left end of the mirror, at A, and then up to the detector, it clearly takes more time than a photon that gets to the detector by reflecting in the middle of the mirror, at G (see Fig. 23). Or,

imagine for a moment that you were in a hurry and had to run from the source over to the mirror and then to the detector. You'd know that it certainly isn't a good idea to go way over to A and then all the way up to the dectector; it would be much faster to touch the mirror somewhere in the middle.
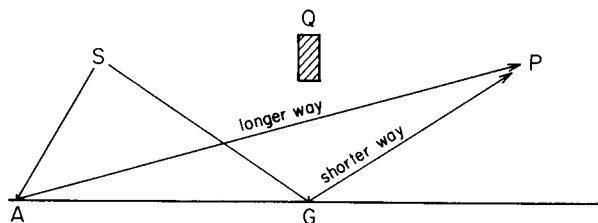

FIGURE 23.   *While the length of each arrow is essentially the same, the direction will be different because the time it takes for a photon to go on each path is different. Clearly, it takes longer to go from S to A to P than from S to G to P.*

To help us calculate the direction of each arrow, I'm going to draw a graph right underneath my sketch of the mirror (see Fig. 24). Directly below each place on the mirror where the light could reflect, I'm going to show, vertically, how much time it would take if the light went that way. The more time it takes, the higher the point will be on the graph. Starting at the left, the time it takes a photon to go on the path that reflects at A is pretty long, so we plot a point pretty high up on the graph. As we move toward the center of the mirror, the time it takes for a photon to go the particular way we're looking at goes down, so we plot each successive point lower than the previous one. After we pass the center of the mirror, the time it takes a photon to go on each successive path gets longer and longer, so we plot our points correspondingly higher and higher. To aid the eye, let's connect the points: they form a symmetrical
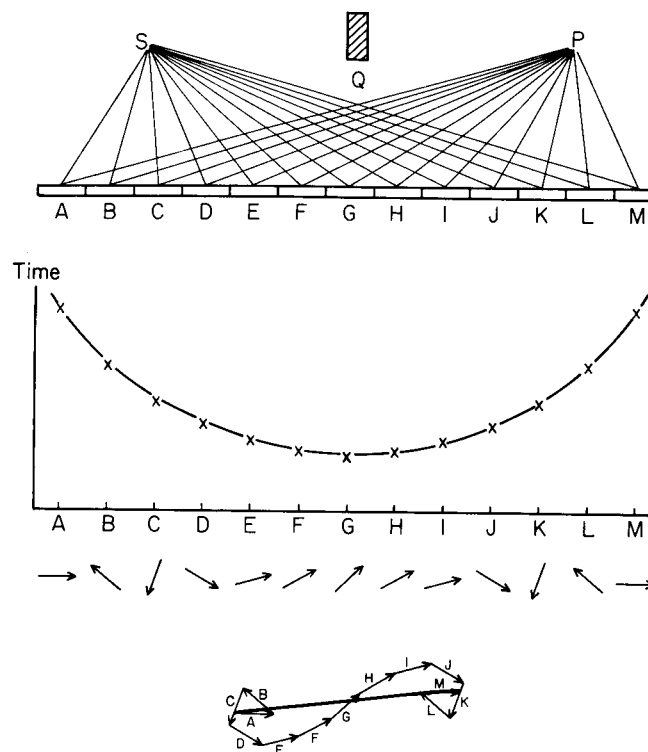
FIGURE 24.   *Each path the light could go (in this simplified situation) is shown at the top, with a point on the graph below it showing the time it takes a photon to go from the source to that point on the mirror, and then to the photomultiplier. Below the graph is the direction of each arrow, and at the bottom is the result of adding all the arrows. It is evident that the major contribution to the final arrow's length is made by arrows E through I, whose directions are nearly the same because the timing of their paths is nearly the same. This also happens to be where the total time is least. It is therefore approximately right to say that light goes where the time is least.*

curve that starts high, goes down, and then goes back up again.

Now, what does that mean for the direction of the little arrows? The direction of a particular arrow corresponds to the amount of time it would take a photon to get from

the source to the detector following that particular path. Let's draw the arrows, starting at the left. Path A takes the most time; its arrow points in some direction (Fig. 24). The arrow for path B points in a different direction because its time is different. At the middle of the mirror, arrows F, G, and H point in nearly the same direction because their times are nearly the same. After passing the center of the mirror, we see that each path on the right side of the mirror corresponds to a path on the left side whose time is exactly the same (this is a consequence of putting the source and the detector at the same height, and path G exactly in the middle). Thus the arrow for path J, for example, has the same direction as the arrow for path D.

Now, let's add the little arrows (Fig. 24). Starting with arrow A, we hook the arrows to each other, head to tail. Now, if we were to take a walk using each little arrow as a step, we wouldn't get very far at the beginning, because the direction from one step to the next is so different. But after a while the arrows begin to point in generally the same direction, and we make some progress. Finally, near the end of our walk, the direction from one step to the next is again quite different, so we stagger about some more.

All we have to do now is draw the final arrow. We simply connect the tail of the first little arrow to the head of the last one, and see how much direct progress we made on our walk (Fig. 24). And behold—we get a sizable final arrow! The theory of quantum electrodynamics predicts that light does, indeed, reflect off the mirror!

Now, let's investigate. What determines how long the final arrow is? We notice a number of things. First, the ends of the mirror are not important: there, the little arrows wander around and don't get anywhere. If I chopped off the ends of the mirror—parts that you instinctively knew I was wasting my time fiddling around with—it would hardly affect the length of the final arrow.

So where is the part of the mirror that gives the final arrow a substantial length? It's the part where the arrows are all pointing in nearly the same direction—because their *time* is almost the *same*. If you look at the graph showing the time for each path (Fig. 24), you see that the time is nearly the same from one path to the next at the bottom of the curve, where the *time* is *least*.

To summarize, where the time is least is also where the time for the nearby paths is nearly the same; that's where the little arrows point in nearly the same direction and add up to a substantial length; that's where the probability of a photon reflecting off a mirror is determined. And that's why, in approximation, we can get away with the crude picture of the world that says that light only goes where the *time* is *least* (and it's easy to prove that where the time is least, the angle of incidence is equal to the angle of reflection, but I don't have the time to show you).

So the theory of quantum electrodynamics gave the right answer—the middle of the mirror is the important part for reflection—but this correct result came out at the expense of believing that light reflects all over the mirror, and having to add a bunch of little arrows together whose sole purpose was to cancel out. All that might seem to you to be a waste of time—some silly game for mathematicians only. After all, it doesn't seem like "real physics" to have something there that only cancels out!

Let's test the idea that there really *is* reflection going on all over the mirror by doing another experiment. First, let's chop off most of the mirror, and leave about a quarter of it, over on the left. We still have a pretty big piece of mirror, but it's in the wrong place. In the previous experiment the arrows on the left side of the mirror were pointing in directions very different from one another because of the large difference in time between neighboring paths (Fig. 24). In this experiment I am going to make a more detailed calculation by taking intervals on that left-hand part of the

mirror that are much closer together—fine enough that there is not much difference in time between neighboring paths (see Fig. 25). With this more detailed picture, we see that some of the arrows point more or less to the right; the others point more or less to the left. If we add *all* the arrows together, we have a bunch of arrows going around in what is essentially a circle, getting nowhere.
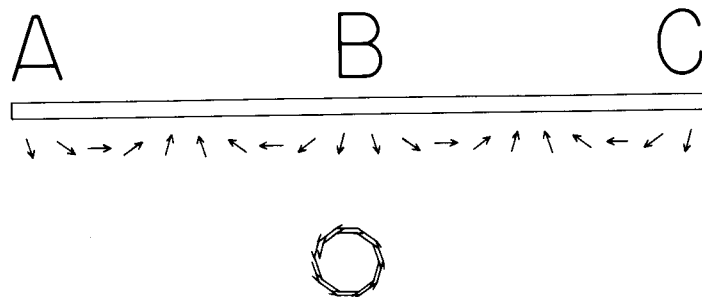

FIGURE 25.   *To test the idea that there is really reflection happening at the ends of the mirror (but it is just cancelling out), we do an experiment with a large piece of mirror that is located in the wrong place for reflection from S to P. This piece of mirror is divided into much smaller sections, so that the timing from one path to the next is not very different. When all the arrows are added, they get nowhere: they go in a circle and add up to nearly nothing.*

But let's suppose we carefully scrape the mirror away in those areas whose arrows have a bias in one direction—let's say, to the left—so that only those places whose arrows point generally the other way remain (see Fig. 26). When we add up only the arrows that point more or less to the right, we get a series of dips and a substantial final arrow—according to the theory, we should now have a strong reflection! And indeed, we do—the theory *is* correct! Such a mirror is called a diffraction grating, and it works like a charm.

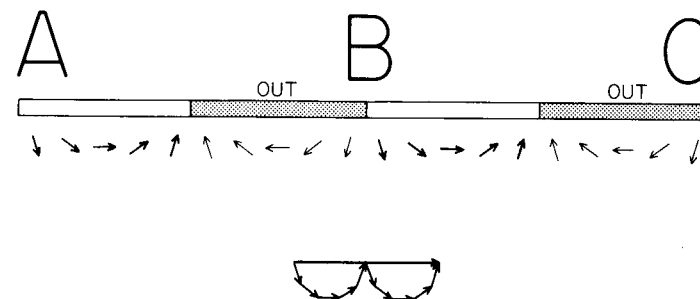Isn't it wonderful—you can take a piece of mirror where

FIGURE 26.   *If only the arrows with a bias in a particular direction—such as to the right—are added, while the others are disregarded (by etching away the mirror in those places), then a substantial amount of light reflects from this piece of mirror located in the wrong place. Such an etched mirror is called a diffraction grating.*

you didn't expect any reflection, scrape away part of it, and it reflects![1]

The particular grating that I just showed you was tailor-made for red light. It wouldn't work for blue light; we would have to make a new grating with the cut-away strips spaced closer together because, as I told you in the first lecture, the stopwatch hand turns around faster when it times a blue photon compared to a red photon. So the cuts that were especially designed for the "red" rate of turning don't fall in the right places for blue light; the arrows get kinked up and the grating doesn't work very well. But as a matter of accident, it happens that if we move the photomultiplier down to a somewhat different angle, the grating made for red light now works for blue light. It's just a

---

[1] The areas of the mirror whose arrows point generally to the left also make a strong reflection (when the areas whose arrows point the other way are erased). It's when both left-biased and right-biased areas reflect together that they cancel out. This is analogous to the case of partial reflection by two surfaces: while either surface will reflect on its own, if the thickness is such that the two surfaces contribute arrows pointing in opposite directions, reflection is cancelled out.

lucky accident, a consequence of the geometry involved (see Fig. 27).

If you shine white light down onto the grating, red light comes out at one place, orange light comes out slightly above it, followed by yellow, green, and blue light—all the
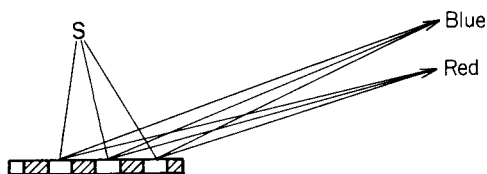

FIGURE 27. *A diffraction grating with grooves at the right distance for red light also works for other colors, if the detector is in a different place. Thus it is possible to see different colors reflecting from a grooved surface—such as a phonograph record—depending on the angle.*

colors of the rainbow. Where there is a series of grooves close together, you can often see colors—for example, when you hold a phonograph record (or better, a videodisc)— under bright light at the correct angles. Perhaps you have seen those wonderful silvery signs (here in sunny California they're often on the backs of cars): when the car moves, you see very bright colors changing from red to blue. Now you know where the colors come from: you're looking at a grating—a mirror that's been scratched in just the right places. The sun is the light source, and your eyes are the detector. I could go on to easily explain how lasers and holograms work, but I know that not everyone has seen these things, and I have too many other things to talk about.[2]

[2] I can't resist telling you about a grating that Nature has made: salt crystals are sodium and chlorine atoms packed in a regular pattern.

So a grating shows that we can't ignore the parts of a mirror that don't seem to be reflecting; if we do some clever things to the mirror, we can demonstrate the reality of the reflections from all parts of the mirror and produce some striking optical phenomena.
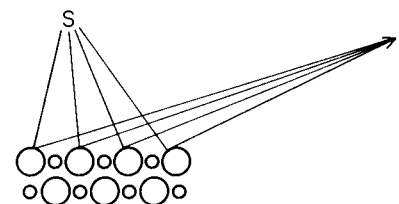

FIGURE 28. *Nature has made many types of diffraction gratings in the form of crystals. A salt crystal reflects X-rays (light for which the imaginary stopwatch hand moves extremely fast—perhaps 10,000 times faster than for visible light) at various angles, from which can be determined the exact arrangement and spacings of the individual atoms.*

More importantly, demonstrating the reality of reflection from *all* parts of the mirror shows that there is an amplitude—an arrow—for *every way* an event can happen. And in order to calculate correctly the probability of an event in different circumstances, we have to add the arrows for *every way* that the event could happen—not just the ways we think are the important ones!

Their alternating pattern, like our grooved surface, acts like a grating when light of the right color (X-rays, in this case) shines on it. By finding the specific locations where a detector picks up a lot of this special reflection (called diffraction), one can determine exactly how far apart the grooves are, and thus how far apart the atoms are (see Fig. 28). It is a beautiful way of determining the structure of all kinds of crystals as well as confirming that X-rays are the same thing as light. Such experiments were first done in 1914. It was very exciting to see, in detail, for the first time how the atoms are packed together in different substances.

Now, I would like to talk about something more familiar than gratings—about light going from air into water. This time, let's put the photomultiplier underwater—we suppose the experimenter can arrange that! The source of light is in the air at S, and the dectector is underwater, at D (see Fig. 29). Once again, we want to calculate the probability that a photon will get from the light source to the detector. To make this calculation, we should consider all
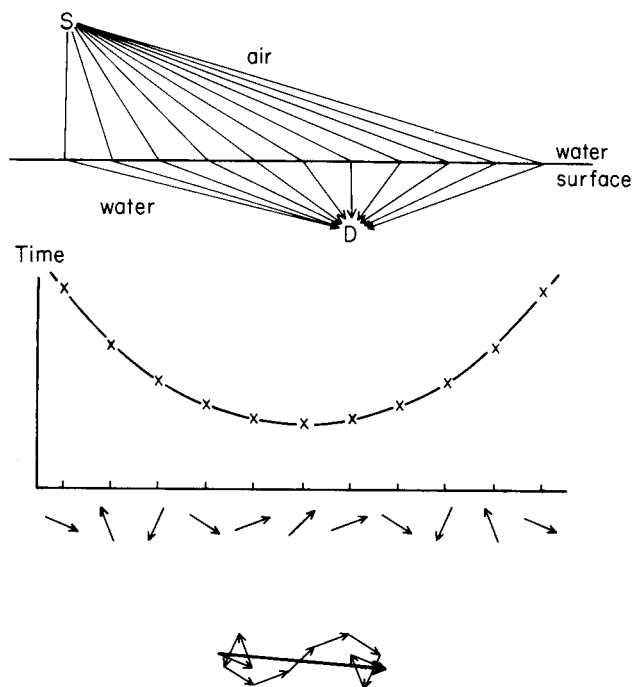

FIGURE 29.   *Quantum theory says that light can go from a source in air to a detector in water in many ways. If the problem is simplified as in the case of the mirror, a graph showing the timing of each path can be drawn, with the direction of each arrow below it. Once again, the major contribution toward the length of the final arrow comes from those paths whose arrows point in nearly the same direction because their timing is nearly the same; once again, this is where the time is least.*

the ways the light could go. Each way the light could go contributes a little arrow and, as in the previous example, all the little arrows have nearly the same length. We can again make a graph of the time it takes a photon to go on each possible path. The graph will be a curve very similar to the one we made for light reflecting off a mirror: it starts up high, goes down, and then back up again; the most important contributions come from the places where the arrows point in nearly the same direction (where the time is nearly the same from one path to the next), which is at the bottom of the curve. That is also where the time is the least, so all we have to do is find out where the time is least.

It turns out that light seems to go slower in water than it does in air (I will explain why in the next lecture), which makes the distance through water more "costly," so to speak, than the distance through air. It's not hard to figure out which path takes the least time: suppose you're the lifeguard, sitting at S, and the beautiful girl is drowning, at D (Fig. 30). You can run on land faster than you can swim in water. The problem is, where do you enter the water in order to reach the drowning victim the fastest? Do you run down to the water at A, and then swim like
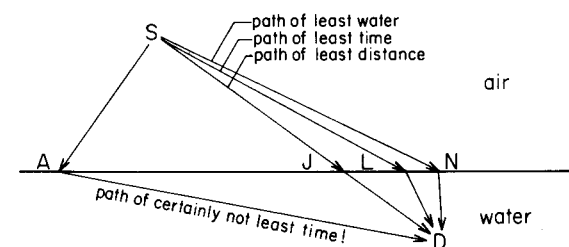

FIGURE 30.   *Finding the path of least time for light is like finding the path of least time for a lifeguard running and then swimming to rescue a drowning victim: the path of least distance has too much water in it; the path of least water has too much land in it; the path of least time is a compromise between the two.*

hell? Of course not. But running directly toward the victim and entering the water at J is not the fastest route, either. While it would be foolish for a lifeguard to analyze and calculate under the circumstances, there is a computable position at which the time is minimum: it's a compromise between taking the direct path, through J, and taking the path with the least water, through N. And so it is with light—the path of least time enters the water at a point between J and N, such as L.

Another phenomenon of light that I would like to mention briefly is the mirage. When you're driving along a road that is very hot, you can sometimes see what looks like water on the road. What you're really seeing is the sky, and when you normally see sky on the road, it's because the road has puddles of water on it (partial reflection of light by a single surface). But how can you see sky on the road when there's no water there? What you need to know is that light goes slower through cooler air than through warmer air, and for a mirage to be seen, the observer must be in the cooler air that is above the hot air next to the road surface (see Fig. 31). How it is possible to look *down* and see the sky can be understood by finding the path of least time. I'll let you play with that one at home—it's fun to think about, and pretty easy to figure out.
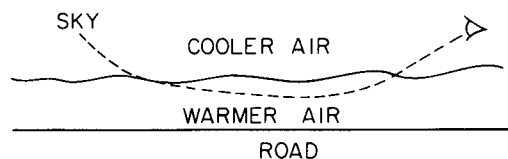
FIGURE 31. *Finding the path of least time explains how a mirage works. Light goes faster through warm air than through cool air. Some of the sky appears to be on the road because some of the light from the sky reaches the eye by coming up from the road. The only other time sky appears to be on the road is when water is reflecting it, and thus a mirage appears to be water.*

In the examples I showed you of light reflecting off a mirror and of light going through air and then water, I was making an approximation: for the sake of simplicity, I drew the various ways the light could go as double straight lines—two straight lines that form an angle. But we don't have to *assume* that light goes in straight lines when it is in a uniform material like air or water; even *that* is explainable by the general principle of quantum theory: the probability of an event is found by adding arrows for *all* the ways the event could happen.

So for our next example, I'm going to show you how, by adding little arrows, it can appear that light goes in a straight line. Let's put a source and a photomultiplier at S and P, respectively (see Fig. 32), and look at *all* the ways
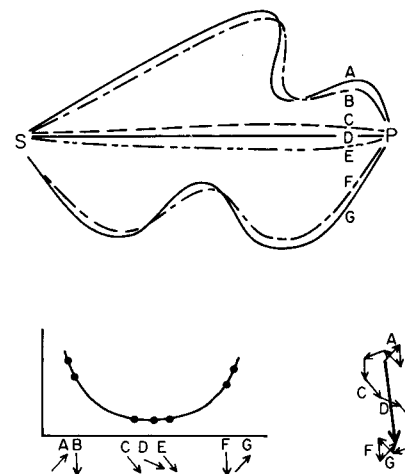
FIGURE 32. *Quantum theory can be used to show why light appears to travel in straight lines. When all possible paths are considered, each crooked path has a nearby path of considerably less distance and therefore much less time (and a substantially different direction for the arrow). Only the paths near the straight-line path at D have arrows pointing in nearly the same direction, because their timings are nearly the same. Only such arrows are important, because it is from them that we accumulate a large final arrow.*

the light could go—in all sorts of crooked paths—to get from the source to the detector. Then we draw a little arrow for each path, and we're learning our lesson well!

For each crooked path, such as path A, there's a nearby path that's a little bit straighter and distinctly shorter—that is, it takes much less time. But where the paths become nearly straight—at C, for example—a nearby, straighter path has nearly the same time. That's where the arrows add up rather than cancel out; that's where the light goes.

It is important to note that the single arrow that represents the straight-line path, through D (Fig. 32), is not enough to account for the probability that light gets from the source to the detector. The nearby, nearly straight paths—through C and E, for example—also make important contributions. So light doesn't *really* travel only in a straight line; it "smells" the neighboring paths around it, and uses a small core of nearby space. (In the same way, a mirror has to have enough size to reflect normally: if the mirror is too small for the core of neighboring paths, the light scatters in many directions, no matter where you put the mirror.)

Let's investigate this core of light more closely by putting a source at S, a photomultiplier at P, and a pair of blocks between them to keep the paths of light from wandering too far away (see Fig. 33). Now, let's put a second photomultiplier at Q, below P, and assume again, for the sake of simplicity, that the light can get from S to Q only by paths of double straight lines. Now, what happens? When the gap between the blocks is wide enough to allow many neighboring paths to P and to Q, the arrows for the paths to P add up (because all the paths to P take nearly the same time), while the paths to Q cancel out (because those paths have a sizable difference in time). Thus the photomultiplier at Q doesn't click.

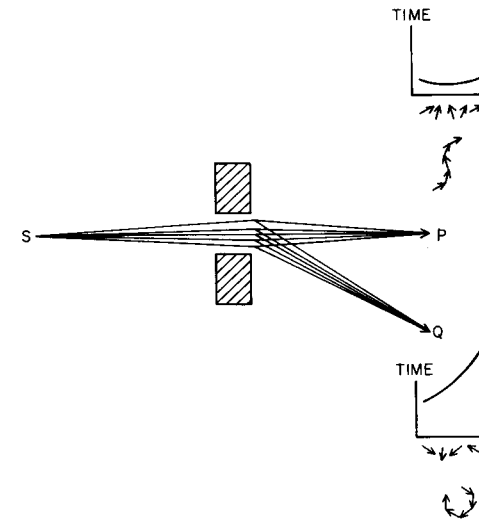But as we push the blocks closer together, at a certain

FIGURE 33. *Light travels in not just the straight-line path, but in the nearby paths as well. When two blocks are separated enough to allow for these nearby paths, the photons proceed normally to P, and hardly ever go to Q.*

point, the detector at Q starts clicking! When the gap is nearly closed and there are only a few neighboring paths, the arrows to Q *also* add up, because there is hardly any difference in time between them, either (see Fig. 34). Of course, both final arrows are small, so there's not much light either way through such a small hole, but the detector at Q clicks almost as much as the one at P! So when you try to squeeze light too much to make sure it's going in only a straight line, it refuses to cooperate and begins to spread out.[3]

[3] This is an example of the "uncertainty principle": there is a kind of "complementarity" between knowledge of where the light goes between the blocks and where it goes afterwards—precise knowledge of both is impossible. I would like to put the uncertainty principle in its historical place: When the revolutionary ideas of quantum physics were first coming
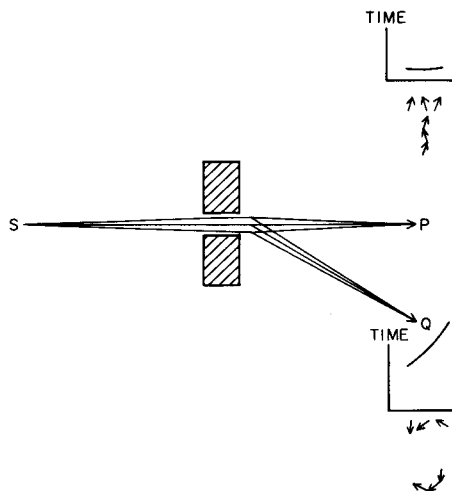
FIGURE 34. *When light is restricted so much that only a few paths are possible, the light that is able to get through the narrow slit goes to Q almost as much as to P, because there are not enough arrows representing the paths to Q to cancel each other out.*

So the idea that light goes in a straight line is a convenient approximation to describe what happens in the world that is familiar to us; it's similar to the crude approximation that says when light reflects off a mirror, the angle of incidence is equal to the angle of reflection.

Just as we were able to do a clever trick to make light reflect off a mirror at many angles, we can do a similar

out, people still tried to understand them in terms of old-fashioned ideas (such as, light goes in straight lines). But at a certain point the old-fashioned ideas would begin to fail, so a warning was developed that said, in effect, "Your old-fashioned ideas are no damn good when . . ." If you get rid of all the old-fashioned ideas and instead use the ideas that I'm explaining in these lectures—adding *arrows* for all the ways an event can happen—there is no need for an uncertainty principle!

trick to get light to go from one point to another in many ways.

First, to simplify the situation, I'm going to draw a vertical dashed line (see Fig. 35) between the light source and the detector (the line means nothing; it's just an artificial line)
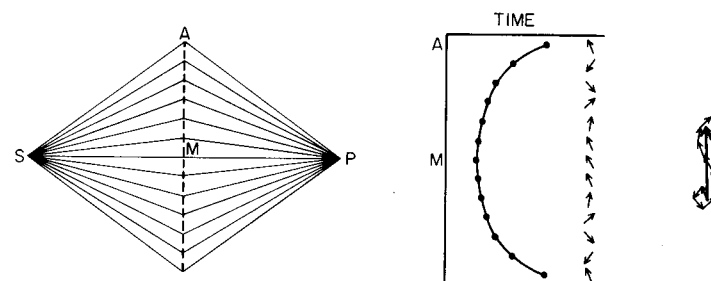

FIGURE 35. *Analysis of all possible paths from S to P is simplified to include only double straight lines (in a single plane). The effect is the same as in the more complicated, real case: there is a time curve with a minimum, where most of the contribution to the final arrow is made.*

and say that the only paths we're going to look at are double straight lines. The graph that shows the time for each path looks the same as in the case of the mirror (but I'll draw it sideways, this time): the curve starts at A, at the top, and then it comes in, because the paths in the middle are shorter and take less time. Finally, the curve goes back out again.

Now, let's have some fun. Let's "fool the light," so that *all* the paths take exactly the same amount of time. How can we do this? How can we make the shortest path, through M, take exactly the same time as the longest path, through A?

Well, light goes slower in water than it does in air; it also goes slower in glass (which is much easier to handle!). So, if we put in just the right thickness of glass on the shortest path, through M, we can make the time for that path exactly

the same as for the path through A. The paths next to M, which are just a little longer, won't need quite as much glass (see Fig. 36). The nearer we get to A, the less glass we have to put in to slow up the light. By carefully calculating and putting in just the right thickness of glass to compensate for the time along each path, we can make all the times the same. When we draw the arrows for each way the light
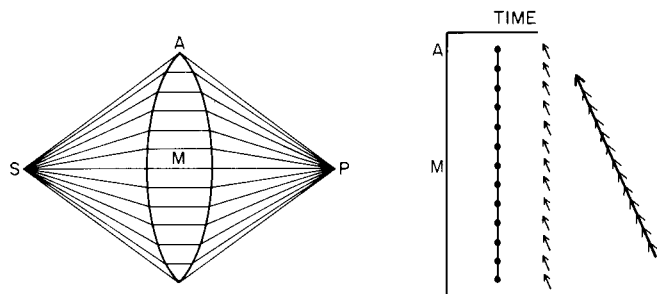

FIGURE 36. *A "trick" can be played on Nature by slowing down the light that takes shorter paths: glass of just the right thickness is inserted so that all the paths will take exactly the same time. This causes all of the arrows to point in the same direction, and to produce a whopping final arrow—lots of light! Such a piece of glass made to greatly increase the probability of light getting from a source to a single point is called a focusing lens.*

could go, we find we have succeeded in straightening them all out—and there are, in reality, *millions* of tiny arrows—so the net result is a sensationally large, unexpectedly enormous final arrow! Of course you know what I'm describing; it's a focusing lens. By arranging things so that all the times are equal, we can focus light—we can make the probability very high that light will arrive at a particular point, and very low that it will arrive anywhere else.

I have used these examples to show you how the theory of quantum electrodynamics, which looks at first like an absurd idea with no causality, no mechanism, and nothing

real to it, produces effects that you are familiar with: light bouncing off a mirror, light bending when it goes from air into water, and light focused by a lens. It also produces other effects that you may or may not have seen, such as the diffraction grating and a number of other things. In fact, the theory continues to be successful at explaining *every* phenomenon of light.

I have shown you with examples how to calculate the probability of an event that can happen in *alternative ways*: we draw an arrow for each way the event can happen, and add the arrows. "Adding arrows" means the arrows are placed head to tail and a "final arrow" is drawn. The square of the resulting final arrow represents the probability of the event.

In order to give you a fuller flavor of quantum theory, I would now like to show you how physicists calculate the probability of compound events—events that can be broken down into a series of steps, or events that consist of a number of things happening independently.

An example of a compound event can be demonstrated by modifying our first experiment, in which we aimed some red photons at a single surface of glass to measure partial reflection. Instead of putting the photomultiplier at A (see Fig. 37), let's put in a screen with a hole in it to let the photons that reach point A go through. Then let's put in a sheet of glass at B, and place the photomultiplier at C. How do we figure out the probability that a photon will get from the source to C?

We can think of this event as a sequence of two steps. Step 1: a photon goes from the source to point A, reflecting off the single surface of glass. Step 2: the photon goes from point A to the photomultiplier at C, reflecting off the sheet of glass at B. Each step has a final arrow—an *"amplitude"* (I'm going to use the words interchangeably)—that can be calculated according to the rules we know so far. The am-

plitude for the first step has a length of 0.2 (whose square is 0.04, the probability of reflection by a single surface of glass), and is turned at some angle—let's say, 2 o'clock (Fig. 37).

To calculate the amplitude for the second step, we temporarily put the light source at A and aim the photons at the layer of glass above. We draw arrows for the front and back surface reflections and add them—let's say we end up with a final arrow with a length of 0.3, and turned toward 5 o'clock.
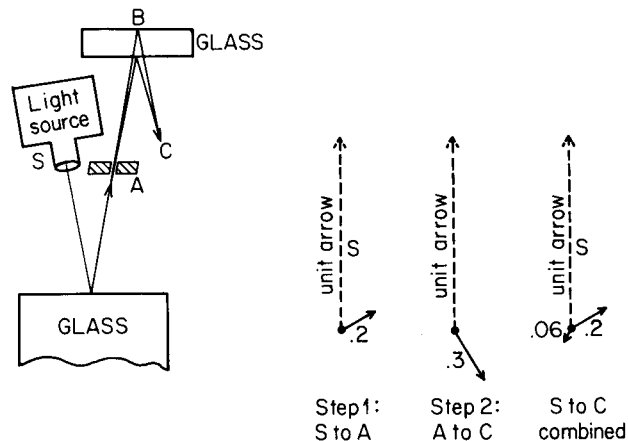

FIGURE 37. *A compound event can be analyzed as a succession of steps. In this example, the path of a photon going from S to C can be divided into two steps: 1) a photon gets from S to A, and 2) the photon gets from A to C. Each step can be analyzed separately to produce an arrow that can be regarded in a new way: as a unit arrow (an arrow of length 1 pointed at 12 o'clock) that has gone through a shrink and turn. In this example, the shrink and turn for Step 1 are 0.2 and 2 o'clock; the shrink and turn for Step 2 are 0.3 and 5 o'clock. To get the amplitude for the two steps in succession, we shrink and turn in succession: the unit arrow is shrunk and turned to produce an arrow of length 0.2 turned to 2 o'clock, which itself is shrunk and turned (as if it were the unit arrow) by 0.3 and 5 o'clock to produce an arrow of length 0.06 and turned to 7 o'clock. This process of successive shrinking and turning is called "multiplying" arrows.*

Now, how do we combine the two arrows to draw the amplitude for the entire event? We look at each arrow in a new way: as instructions for a *shrink* and *turn*.

In this example, the first amplitude has a length of 0.2 and is turned toward 2 o'clock. If we begin with a "unit arrow"—an arrow of length 1 pointed straight up—we can *shrink* this unit arrow from 1 down to 0.2, and *turn* it from 12 o'clock to 2 o'clock. The amplitude for the second step can be thought of as shrinking the unit arrow from 1 to 0.3 and turning it from 12 o'clock to 5 o'clock.

Now, to combine the amplitudes for both steps, we shrink and turn *in succession*. First, we shrink the unit arrow from 1 to 0.2 and turn it from 12 to 2 o'clock; then we shrink the arrow further, from 0.2 down to three-tenths of that, and turn it by the amount from 12 to 5—that is, we turn it from 2 o'clock to 7 o'clock. The resulting arrow has a length of 0.06 and is pointed toward 7 o'clock. It represents a probability of 0.06 squared, or 0.0036.

Observing the arrows carefully, we see that the result of shrinking and turning two arrows in succession is the same as adding their angles (2 o'clock + 5 o'clock) and multiplying their lengths (0.2 * 0.3). To understand why we add the angles is easy: the angle of an arrow is determined by the amount of turning by the imaginary stopwatch hand. So the total amount of turning for the two steps in succession is simply the sum of the turning for the first step plus the additional turning for the second step.

Why we call this process "multiplying arrows" takes a bit more explanation, but it's interesting. Let's look at multiplication, for a moment, from the point of view of the Greeks (this has nothing to do with the lecture). The Greeks wanted to use numbers that were not necessarily integers, so they represented numbers with lines. Any number can be expressed as a *transformation* of the unit line—by expanding it or shrinking it. For example, if Line A is the

unit line (see Fig. 38), then line B represents 2 and line C represents 3.

Now, how do we multiply 3 times 2? We apply the transformations *in succession*: starting with line A as the unit line, we expand it 2 times and then 3 times (or 3 times and then 2 times—the order doesn't make any difference). The result is line D, whose length represents 6. What about multiplying 1/3 times 1/2? Taking line D to be the unit line, now, we shrink it to 1/2 (line C) and then to 1/3 of that. The result is line A, which represents 1/6.
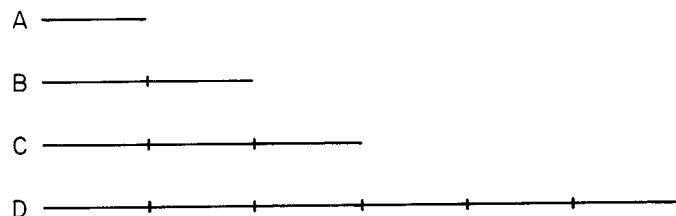

FIGURE 38. *We can express any number as a transformation of the unit line through expansion or shrinkage. If A is the unit line, then B represents 2 (expansion), and C represents 3 (expansion). Multiplying lines is achieved through successive transformations. For example, multiplying 3 by 2 means that the unit line is expanded 3 times and then 2 times, producing the answer, an expansion of 6 (line D). If D is the unit line, then line C represents 1/2 (shrinkage), line B represents 1/3 (shrinkage), and multiplying 1/2 by 1/3 means the unit line D is shrunk to 1/2, and then to 1/3 of that, producing the answer, a shrinkage to 1/6 (line A).*

Multiplying arrows works the same way (see Fig. 39). We apply transformations to the unit arrow in succession—it just happens that the transformation of an *arrow* involves *two* operations, a shrink and turn. To multiply arrow V times arrow W, we shrink and turn the unit arrow by the prescribed amounts for V, and then shrink it and turn it the amounts prescribed for W—again, the order doesn't make any difference. So multiplying arrows follows the
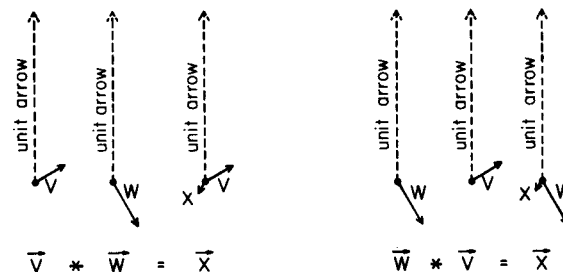
FIGURE 39. *Mathematicians found that multiplying arrows can also be expressed as successive transformations (for our purposes, successive shrinks and turns) of the unit arrow. As in normal multiplication, the order is not important: the answer, arrow X, can be obtained by multiplying arrow V by arrow W or arrow W by arrow V.*

same rule of successive transformations that work for regular numbers.[4]

[4] Mathematicians have tried to find all the objects one could possibly find that obey the rules of algebra (A + B = B + A, A * B = B * A, and so on). The rules were originally made for positive integers, used for counting things like apples or people. Numbers were improved with the invention of zero, fractions, irrational numbers—numbers that cannot be expressed as a ratio of two integers—and negative numbers, and continued to obey the original rules of algebra. Some of the numbers that mathematicians invented posed difficulties for people at first—the idea of half a person was difficult to imagine—but today, there's no difficulty at all: nobody has any moral qualms or discomforting gory feelings when they hear that there is an average of 3.2 people per square mile in some regions. They don't try to imagine the 0.2 people; rather, they know what 3.2 means: if they multiply 3.2 by 10, they get 32. Thus, some things that satisfy the rules of algebra can be interesting to mathematicians even though they don't always represent a real situation. Arrows on a plane can be "added" by putting the head of one arrow on the tail of another, or "multiplied" by successive turns and shrinks. Since these arrows obey the same rules of algebra as regular numbers, mathematicians call them numbers. But to distinguish them from ordinary numbers, they're called "complex numbers." For those of you who have studied mathematics enough to have come to complex numbers, I could have said, "the probability of an event is the absolute square of a complex number. When an event can happen in alternative ways, you add the complex numbers; when it can happen only as a succession of steps, you multiply the complex numbers." Although it may sound more impressive that way, I have not said any more than I did before—I just used a different language.

Let's go back to the first experiment from the first lecture—partial reflection by a single surface—with this idea of successive steps in mind (see Fig. 40). We can divide the path of reflection into three steps: 1) the light goes from the source down to the glass, 2) it is reflected by the glass, and 3) it goes from the glass up to the detector. Each step can be considered as a certain amount of shrinking and turning of the unit arrow.
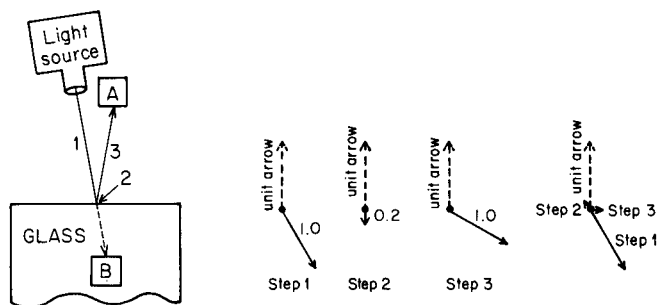

FIGURE 40. *Reflection by a single surface can be divided into three steps, each with a shrink and/or turn of the unit arrow. The net result, an arrow of length 0.2 pointed in some direction, is the same as before, but our method of analysis is more detailed now.*

You'll remember that in the first lecture, we did not consider *all* of the ways the light could reflect off the glass, which requires drawing and adding lots and lots of little tiny arrows. In order to avoid all that detail, I gave the impression that the light goes down to a particular point on the surface of the glass—that it doesn't spread out. When light goes from one point to another, it does, in reality, spread out (unless it's fooled by a lens), and there is some shrinkage of the unit arrow associated with that. For the moment, however, I would like to stick to the simplified view that light does *not* spread out, and so it is ap-

propriate to disregard this shrinkage. It is also appropriate to assume that since the light doesn't spread out, every photon that leaves the source ends up at either A or B.

So: in the first step there is no shrinking, but there is turning—it corresponds to the amount of turning by the imaginary stopwatch hand as it times the photon going from the source to the front surface of the glass. In this example, the arrow for the first step ends up with a length of 1 at some angle—let's say, 5 o'clock.

The second step is the reflection of the photon by the glass. Here, there is a sizable shrink—from 1 to 0.2—and half a turn. (These numbers seem arbitrary now: they depend upon whether the light is reflected by glass or some other material. In the third lecture, I'll explain them, too!) Thus the second step is represented by an amplitude of length 0.2 and a direction of 6 o'clock (half a turn).

The last step is the photon going from the glass up to the detector. Here, as in the first step, there is no shrinking, but there is turning—let's say this distance is slightly shorter than in step 1, and the arrow points toward 4 o'clock.

We now "multiply" arrows 1, 2, and 3 in succession (add the angles, and multiply the lengths). The net effect of the three steps—1) turning, 2) a shrink and half a turn, and 3) turning—is the same as in the first lecture: the turning from steps 1 and 3—(5 o'clock plus 4 o'clock) is the same amount of turning that we got then when we let the stopwatch run for the whole distance (9 o'clock); the extra half turn from step 2 makes the arrow point in the direction opposite the stopwatch hand, as it did in the first lecture, and the shrinking to 0.2 in the second step leaves an arrow whose square represents the 4% partial reflection observed for a single surface.

In this experiment, there is a question we didn't look at in the first lecture: what about the photons that go to B— the ones that are transmitted by the surface of the glass?

The amplitude for a photon to arrive at B must have a length near 0.98, since 0.98 * 0.98 = 0.9604, which is close enough to 96%. This amplitude can also be analyzed by breaking it down into steps (see Fig. 41).

The first step is the same as for the path to A—the photon goes from the light source down to the glass—the unit arrow is turned toward 5 o'clock.

The second step is the photon passing through the surface of the glass: there is no turning associated with transmission, just a little bit of shrinking—to 0.98.
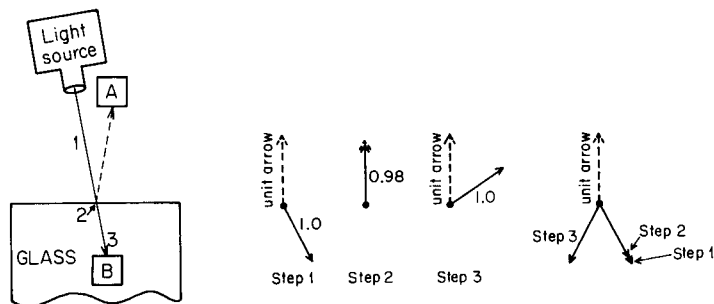

FIGURE 41.   *Transmission by a single surface can also be divided into three steps, with a shrink and/or turn for each step. An arrow of length 0.98 has a square of about 0.96, representing a probabilty of transmission of 96% (which, combined with the 4% probability of reflection, accounts for 100% of the light).*

The third step—the photon going through the interior of the glass—involves additional turning and no shrinking.

The net result is an arrow of length 0.98 turned in some direction, whose square represents the probability that a photon will arrive at B—96%.

Now let's look at partial reflection by two surfaces again. Reflection from the front surface is the same as for a single surface, so the three steps for front surface reflection are the same as we saw a moment ago (Fig. 40).

Reflection from the back surface can be broken down into seven steps (see Fig. 42). It involves turning equal to the total amount of turning of the stopwatch hand timing a photon over the entire distance (steps 1, 3, 5, and 7), shrinking to 0.2 (step 4), and two shrinks to 0.98 (steps 2 and 6). The resulting arrow ends up in the same direction as before, but the length is about 0.192 (0.98 * 0.2 * 0.98), which I approximated as 0.2 in the first lecture.

FIGURE 42.   *Reflection from the back surface of a layer of glass can be divided into seven steps. Steps 1, 3, 5, and 7 involve turning only; steps 2 and 6 involve shrinks to 0.98, and step 4 involves a shrink to 0.2. The result is an arrow of length 0.192—which was approximated as 0.2 in the first lecture—turned at an angle that corresponds to the total amount of turning by the imaginary stopwatch hand.*
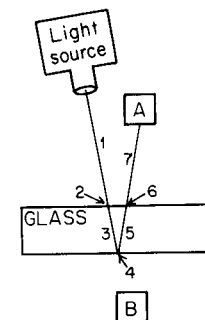
In summary, here are the rules for reflection and transmission of light by glass: 1) reflection from air back to air (off a front surface) involves a shrink to 0.2 and half a turn; 2) reflection from glass back to glass (off a back surface) also involves a shrink to 0.2, but no turning; and 3) transmission from air to glass or from glass to air involves a shrink to 0.98 and no turning in either case.

Perhaps it is too much of a good thing, but I cannot resist showing you a cute further example of how things work and are analyzed by these rules of successive steps. Let us move the detector to a location below the glass, and consider something we didn't talk about in the first lecture—the probability of *transmission* by two surfaces of glass (see Fig. 43).

Of course you know the answer: the probability of a

photon to arrive at B is simply 100% minus the probability
to arrive at A, which we worked out beforehand. Thus, if
we found the chance to arrive at A is 7%, the chance to
arrive at B must be 93%. And as the chance for A varies
from zero through 8% to 16% (due to the different thick-
nesses of glass), the chance for B changes from 100%
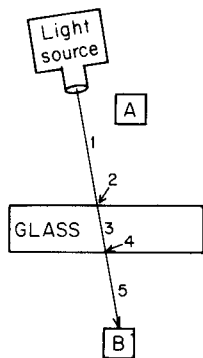through 92% to 84%.

FIGURE 43. *Transmission by two surfaces
can be broken down into five steps. Step 2
shrinks the unit arrow to 0.98, step 4 shrinks
the 0.98 arrow to 0.98 of that (about 0.96);
steps 1, 3, and 5 involve turning only. The
resulting arrow of length 0.96 has a square
of about 0.92, representing a probability of
transmission by two surfaces of 92% (which
corresponds to the expected 8% reflection,
which is right only "twice a day"). When the
thickness of the layer is right to produce a prob-
ability of 16% reflection, with a 92% proba-
bility of transmission, 108% of the light is
accounted for! Something is wrong with this
analysis!*

That is the right answer, but we are expecting to calculate
all probabilities by squaring a final arrow. How do we cal-
culate the amplitude arrow for transmission by a layer of
glass, and how does it manage to vary in length so appro-
priately as to fit with the length for A in each case, so the
probability for A and the probability for B always add up
to exactly 100%? Let us look a little into the details.

For a photon to go from the source to the detector below
the glass, at B, five steps are involved. Let's shrink and turn
the unit arrow as we go along.

The first three steps are the same as in the previous
example: the photon goes from the source to the glass
(turning, no shrinking); the photon is transmitted by the

front surface (no turning, shrinking to 0.98); the photon
goes through the glass (turning, no shrinking).

The fourth step—the photon passes through the back
surface of the glass—is the same as the second step, as far
as shrinks and turns go: no turns, but a shrinkage to 0.98
of the 0.98, so the arrow now has a length of 0.96.

Finally, the photon goes through the air again, down to
the detector—that means more turning, but no further
shrinking. The result is an arrow of length 0.96, pointing
in some direction determined by the successive turnings of
the stopwatch hand.

An arrow whose length is 0.96 represents a probability
of about 92% (0.96 squared), which means an average of
92 photons reach B out of every 100 that leave the source.
That also means that 8% of the photons are reflected by
the two surfaces and reach A. But we found out in the first
lecture that an 8% reflection by two surfaces is only right
sometimes ("twice a day")—that in reality, the reflection by
two surfaces fluctuates in a cycle from zero to 16% as the
thickness of the layer steadily increases. What happens
when the glass is just the right thickness to make a partial
reflection of 16%? For every 100 photons that leave the
source, 16 arrive at A and 92 arrive at B, which means
108% of the light has been accounted for—horrifying!
Something is wrong.

We neglected to consider *all* the ways the light could get
to B! For instance, it could bounce off the back surface, go
up through the glass as if it were going to A, but then
reflect off the front surface, back down toward B (see Fig.
44). This path takes nine steps. Let's see what happens
successively to the unit arrow as the light goes through each
step (don't worry; it only shrinks and turns!).

First step—photon goes through the air—turning; no
shrinking. Second step—photon passes through the glass—
no turning, but shrinking to 0.98. Third step—photon goes

through the glass—turning; no shrinking. Fourth step—reflection off the back surface—no turning, but shrinking to 0.2 of 0.98, or 0.196. Fifth step—photon goes back up through the glass—turning; no shrinking. Sixth step—photon bounces off front surface (it's really a "back" surface, because the photon stays *inside* the glass)—no turning, but shrinking to 0.2 of 0.196, or 0.0392. Seventh step—photon
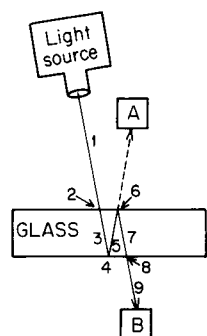

FIGURE 44. *Another way that light could be transmitted by two surfaces must be considered in order to make the calculation more accurate. This path involves two shrinks of 0.98 (steps 2 and 8) and two shrinks of 0.2 (steps 4 and 6), resulting in an arrow of length 0.0384 (rounded off to 0.04).*

goes back down through glass—more turning; no shrinking. Eighth step—photon passes through back surface—no turning, but shrinking to 0.98 of 0.0392, or 0.0384. Finally, the ninth step—photon goes through air to detector—turning; no shrinking.

The result of all this shrinking and turning is an amplitude of length 0.0384—call it 0.04, for all practical purposes—and turned at an angle that corresponds to the total amount of turning by the stopwatch as it times the photon going through this longer path. This arrow represents a *second* way that light can get from the source to B. Now we have two alternatives, so we must *add* the two arrows—the arrow for the more direct path, whose length is 0.96, and the arrow for the longer way, whose length is 0.04—to make the final arrow.

The two arrows are usually not in the same direction, because changing the thickness of the glass changes the relative direction of the 0.04 arrow to the 0.96 arrow. But look how nicely things work out: the extra turns made by the stopwatch timing a photon during steps 3 and 5 (on its way to A) are exactly equal to the extra turns it makes timing a photon during steps 5 and 7 (on its way to B). That means when the two reflection arrows are cancelling each other to make a final arrow representing zero reflection, the arrows for transmission are reinforcing each other to make an arrow of length 0.96 + 0.04, or 1—when the probability of reflection is zero, the probability of transmission is 100% (see Fig. 45). And when the arrows for reflection are rein-
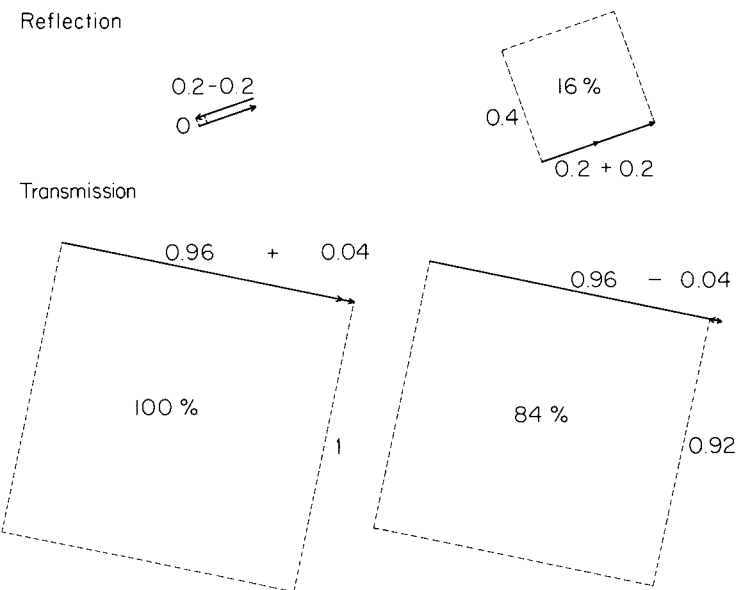

FIGURE 45. *Nature always makes sure 100% of the light is accounted for. When the thickness is right for the transmission arrows to accumulate, the arrows for reflection oppose each other; when the arrows for reflection accumulate, the arrows for transmission oppose each other.*

forcing each other to make an amplitude of 0.4, the arrows for transmission are going against each other, making an amplitude of length 0.96 − 0.04, or 0.92—when reflection is calculated to be 16%, transmission is calculated to be 84% (0.92 squared). You see how clever Nature is with Her rules to make sure that we always come out with 100% of the photons accounted for![5]

Finally, before I go, I would like to tell you that there is an extension to the rule that tells us when to multiply arrows: arrows are to be multiplied not only for an event that consists of a succession of steps, but also for an event that consists of a number of things happening concomitantly—independently and possibly simultaneously. For example, suppose we have two sources, X and Y, and two detectors, A and B (see Fig. 47), and we want to calculate the prob-
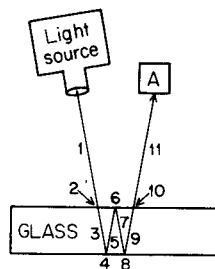
FIGURE 46. *Yet other ways the light could reflect should be considered for a more accurate calculation. In this figure, shrinks of 0.98 occur at steps 2 and 10; shrinks of 0.2 occur at steps 4, 6, and 8. The result is an arrow with a length of about 0.008, which is another alternative for reflection, and should therefore be added to the other arrows which represent reflection (0.2 for the front surface and 0.192 for the back surface).*

[5] You'll notice that we changed 0.0384 to 0.04 and used 84% as the square of 0.92, in order to make 100% of the light accounted for. But when *everything* is added together, 0.0384 and 84% don't have to be rounded off—all the little bits and pieces of arrows (representing all the ways the light could go) compensate for each other and keep the answer correct. For those of you who like this sort of thing, here is an example of another way that the light could go from the light source to the detector at A—a series of three reflections (and two transmissions), resulting in a final arrow of length 0.98 * 0.2 * 0.2 * 0.2 * 0.98, or about 0.008—a very tiny arrow (see Fig. 46). To make a complete calculation of partial reflection by two surfaces, you would have to add in that small arrow, plus an even smaller one that represents five reflections, and so on.

ability for the following event: after X and Y each lose a photon, A and B each gain a photon.

In this example, the photons travel through space to get to the detectors—they are neither reflected nor transmitted—so now is a good time for me to stop disregarding the fact that light spreads out as it goes along. I now present
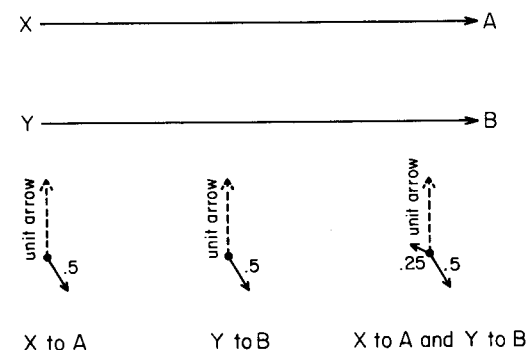
FIGURE 47. *If one of the ways a particular event can happen depends on a number of things happening independently, the amplitude for this way is calculated by multiplying the arrows of the independent things. In this case, the final event is: after sources X and Y each lose a photon, photomultipliers A and B make a click. One way this event could happen is that a photon could go from X to A and a photon could go from Y to B (two independent things). To calculate the probability for this "first way," the arrows for each independent thing—X to A and Y to B—are multiplied to produce the amplitude for this particular way. (Analysis continued in Fig. 48.)*

you with the *complete rule* for monochromatic light travelling from one point to another through space—there is nothing approximate here, and no simplification. This is all there is to know about monochromatic light going through space (disregarding polarization): the *angle* of the arrow depends on the imaginary stopwatch hand, which rotates a certain number of times per inch (depending on the color of the photon); the *length* of the arrow is inversely proportional

to the distance the light goes—in other words, the arrow shrinks as the light goes along.[6]

Let's suppose the arrow for X to A is 0.5 in length and is pointing toward 5 o'clock, as is the arrow for Y to B (Fig. 47). Multiplying one arrow by the other, we get a final arrow of length 0.25, pointed at 10 o'clock.
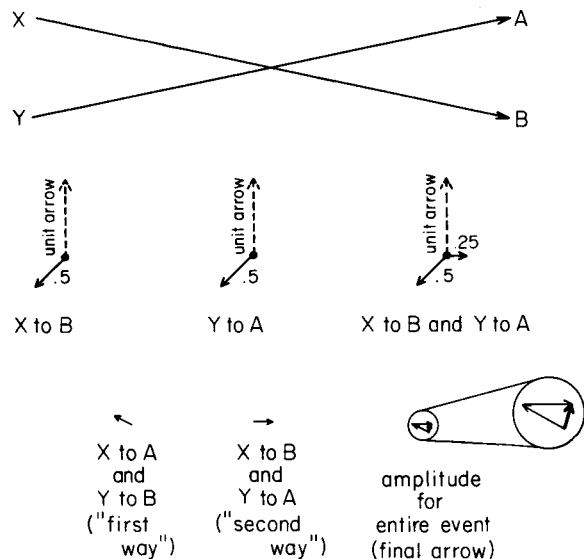

FIGURE 48.   *The other way the event described in Figure 47 could happen— a photon goes from X to B and a photon goes from Y to A—also depends on two independent things happening, so the amplitude for this "second way" is also calculated by multiplying the arrows of the independent things. The "first way" and "second way" arrows are ultimately added together, resulting in the final arrow for the event. The probability of an event is always represented by a single final arrow—no matter how many arrows were drawn, multiplied, and added to achieve it.*

---

[6] This rule checks out with what they teach in school—the amount of light transmitted over a distance varies inversely as the square of the distance—because an arrow that shrinks to half its original size has a square one-fourth as big.

But wait! There is another way this event could happen: the photon from X could go to B, and the photon from Y could go to A. Each of these subevents has an amplitude, and these arrows must also be drawn and multiplied to produce an amplitude for this particular way the event could happen (see Fig. 48). Since the amount of shrinkage over distance is very small compared to the amount of turning, the arrows from X to B and Y to A have essentially the same length as the other arrows, 0.5, but their turning is quite different: the stopwatch hand rotates 36,000 times per inch for red light, so even a tiny difference in distance results in a substantial difference in timing.

The amplitudes for each way the event could happen are added to produce the final arrow. Since their lengths are essentially the same, it is possible for the arrows to cancel each other out if their directions are opposed to each other. The relative directions of the two arrows can be changed by changing the distance between the sources or the detectors: simply moving the detectors apart or together a little bit can make the probability of the event amplify or completely cancel out, just as in the case of partial reflection by two surfaces:[7]

In this example, arrows were multiplied and then added to produce a final arrow (the amplitude for the event), whose square is the probability of the event. It is to be emphasized that no matter how many arrows we draw, add, or multiply, our objective is to calculate a *single final arrow for the event.* Mistakes are often made by physics students at first because they do not keep this important point in mind. They work for so long analyzing events involving a single photon that they begin to think that the arrow is

---

[7] This phenomenon, called the Hanbury-Brown-Twiss effect, has been used to distinguish between a single source and a double source of radio waves in deep space, even when the two sources are extremely close together.

somehow associated with the photon. But these arrows are probability amplitudes, that give, when squared, the *probability* of a complete event.[8]

In the next lecture I will begin the process of simplifying and explaining the properties of matter—to explain where the shrinking to 0.2 comes from, why light appears to go slower through glass or water than through air, and so on— because I have been cheating so far: the photons don't really bounce off the surface of the glass; they interact with the electrons *inside* the glass. I'll show you how photons do nothing but go from one electron to another, and how reflection and transmission are really the result of an electron picking up a photon, "scratching its head," so to speak, and emitting a *new* photon. This simplification of everything we have talked about so far is very pretty.

---

[8] Keeping this principle in mind should help the student avoid being confused by things such as the "reduction of a wave packet" and similar magic.