Infinitesimal Machinery

Richard Feynman

Editor's Preface—This speech, delivered in 1983, is a sequel to the famous 1960 speech "There's Plenty of Room at the Bottom" (reprinted in J. Microelectromechanical Systems, *vol. 1, no. 1, pp. 60-66, 1992). It is remarkable in many ways. Feynman anticipates the sacrificial-layer method of making silicon micromotors, the use of electrostatic actuation, and the importance of friction and contact sticking in such devices. He explores the persistent problem of finding meaningful applications for these tiny machines, touching a range of topics along the way. And he looks at the future of computation using a register made of atoms, and quantum-mechanical transitions for computation operations.*

The speech was presented to a large audience of scientists and engineers at the Jet Propulsion Laboratory (JPL). Following a brief introduction by Al Hibbs, Feynman used slides, hand gestures, and sketches on a blackboard to supplement his pithy language. The style was quite informal. Many of the remarks were intended to evoke laughter, and did so; some were directed specifically at the JPL audience (such as the reference to enhancement of pictures). Working from a video recording of the speech, I had the benefit of seeing the slides, the gestures, and the blackboard sketches. The editor's task, in such a case, is to find a way to capture the essence of Feynman's message and spirit with minimal perturbation of his actual words, making changes only where the lack of visual aids, or the normal fits and starts of oral presentation require some adjustment. I was greatly aided in this task by a transcript of the spoken text prepared and orally smoothed by Nora Odendahl. Where necessary, I have added italicized comments to clarify terms or images being described, but have tried to keep these to a minimum. I have also added topic headings to identify major sections, and have moved some text around to improve readability. As you go through this speech, imagine yourself a part of the audience—if you are moved to laughter in spots, that is appropriate; and if you are moved to think anew about the problems discussed, that indeed was Feynman's goal.

Stephen D. Senturia

Introduction of Richard Feynman by Al Hibbs—Welcome to the Feynman lecture on "Infinitesimal Machinery." I have the pleasure of introducing Richard, an old friend and past associate. He was educated at MIT and at Princeton, where he received a Ph.D. in 1942. In the War he was at Los Alamos, where he learned how to pick combination locks—an activity at which he is still quite skillful. He next went to Cornell, where he experimented with swinging hoops. Then, both before and during his time at Caltech, he became an expert in drumming, specializing in complex rhythms, particularly those of South America and recently those of the South Pacific. At Caltech, he learned to decode Mayan hieroglyphs and took up art, becoming quite an accomplished draftsman—specializing in nude women. And he also does jogging.

Richard received the Nobel prize, but I believe it was for physics and not for any of these other accomplishments. He thinks that happened in 1965, although he doesn't remember the exact year. I have never known him to suffer from false modesty, so I believe he really has forgotten which year he got the Nobel prize.

This manuscript is based on a talk given by Richard Feynman on February 23, 1983, at the Jet Propulsion Laboratory, Pasadena CA. It is reprinted with the permission of his estate, Carl Feynman executor.

The author, deceased, was with the California Institute of Technology, Pasadena, CA.

IEEE Log Number 9210135.

WHEN Dick Davies asked me to talk, he didn't tell me the occasion was going to be so elaborate, with TV cameras and everything—he told me I'd be among friends. I didn't realize I had so many friends. I would feel much less uncomfortable if I had more to say. I don't have very much to say—but of course, I'll take a long time to say it.

Revisiting "There's Plenty of Room at the Bottom"

In 1960, about 23 years ago, I gave a talk called "There's Plenty of Room at the Bottom," in which I described the coming technology for making small things. I pointed out what everybody knew: that numbers, information, and computing didn't require any particular size. You could write numbers very small, down to atomic size. (Of course you can't write something much smaller than the size of a single atom.) Therefore, we could store a lot of information in small spaces, and in a little while we'd be able to do so easily. And of course, that's what happened.

I've been asked a number of times to reconsider all the things that I talked about 23 years ago, and to see how the situation has changed. So my talk today could be called "There's Plenty of Room at the Bottom, Revisited."

As I mentioned in the 1960 talk, you could represent a digit by saying it is made of a few atoms. Actually, you'd only have to have to use one atom for each digit, but let's say you make a bit from a bunch of gold atoms, and another bit from a bunch of silver atoms. The gold atoms represent a one, and the silver atoms a zero. Suppose you make the bits into little cubes with a hundred atoms on a side. When you stack the cubes all together, you can write a lot of stuff in a small space. It turns out that all the books in all the world's libraries could have all their information—including pictures using dots down to the resolution of the human eye—stored in a cube 1/120 inch on a side.

That cube would be just about the size you can make out with your eye—about the size of a speck of dirt.

If, however, you used only surfaces rather than the volume of the cubes to store information, and if you simply reduce normal scale by twenty-five thousand times, which was just about possible in those days, then the *Encyclopedia Britannica* could be written on the head of a pin, the Caltech library on one library card, and all the books in the world on thirty-five pages of the *Saturday Evening Post.* I suggested a reduction of twenty-five thousand times just to make the task harder, because due to the limitations of light wavelength, that reduction was about ten times smaller than you could read by means of light. You could, of course, read the information with electron microscopes and electron beams.

Because I had mentioned the possibility of using electron beams and making things still smaller, six or eight years ago someone sent me a picture of a book that he reduced by thirty thousand times. In the picture, there are letters measuring about a tenth of a micron across [*passes the picture around the audience*].

I also talked in the 1960 lecture about small machinery, and was able to suggest no particular use for the small machines. You will see there has been no progress in that respect. And I left as a challenge the goal of making a motor that would measure 1/64 of an inch on a side. At that time, the idea that I proposed was to make a set of hands—like those used in radioactive systems—that followed another set of hands. Only we make these "slave" hands smaller—a quarter of the original hands' size—and then let the slave hands make smaller hands and those make still smaller hands. You're right to laugh—I doubt that that's a sensible technique. At any rate, I wanted to get a motor that couldn't be made directly by hand, so I proposed 1/64 of an inch.

At the end of my talk, Don Glaser, who won the Nobel prize in physics—that's something that's supposed to be good, right?—said, "You should have asked for a motor 1/200 inch on a side, because 1/64 inch on a side is just about possible by hand." And I said, "Yeah, but if I offered a thousand-dollar prize for a motor 1/200 inch on a side, everybody would say 'Boy, that guy's a cheapskate! Nobody's ever going to do that.'" I didn't believe Glaser, but somebody actually did make the motor by hand!

As a matter of fact, the motor's very interesting, and just for fun, here it is. First look at it directly with your eye, to see how big it is. It's right in the middle of that little circle—it's only the size of a decimal point or a period at the end of a sentence. Mr. McLellan, who made this device for me, arranged it very beautifully, so that it has a magnifier you can attach—but don't look at it through the magnifier until you look at it directly. You'll find you can't see it without the magnifier. Then you can look through the magnifier and turn this knob, which is a little hand generator which makes the juice to turn the motor so you can watch the motor go around [*gives the McLellan motor to the audience to be passed around*].

What We Can Do Today

Now I'd like to talk about what we can do today, as compared to what we were doing in those days. Back then, I was speaking about machinery as well as writing, computers, and information, and although this talk is billed as being about machinery, I'll also discuss computers and information at the end.

My first slide illustrates what can be done today in making small things commercially. This is of course one of the chips that we use in computers, and it represents an area of about three millimeters by four millimeters. Human beings can actually make something on that small a scale, with wires about six microns across (a micron is a millionth of a meter, or a thousandth of a millimeter). The tolerances, dimensions, and separations of some of the wires are controlled to about three microns. This computer chip was manufactured five years ago, and now things have improved so that we can get down to about one-half micron resolution.

These chips are made, as you know, by evaporating successive layers of materials through masks. [*Feynman uses "evaporating" as a generic term for all semiconductor process steps.*] You can create the pattern in a material in several ways. One is to shine light through a mask that has the design that you want, then focus the light very accurately onto a light-sensitive material and use the light to change the material, so that it gets easier to etch or gets less easy to etch. Then you etch the various materials away in stages. You can also deposit one material after another—there's oxide, and silicon, and silicon with materials diffused into it—all arranged in a pattern at that scale. This technology was incredible twenty-three years ago, but that's where we are today.

The real question is, how far can we go? I'll explain to you later why, when it comes to computers, it's always better to get smaller, and everybody's still trying to get smaller. But if light has a finite wavelength, then we're not going to be able to make masks with patterns measuring less than a wavelength. That fact limits us to about a half a micron, which is about possible nowadays, with light, in laboratories. The commercial scale is about twice that big.

So what could we do today, if we were to work as hard as we could in a laboratory—not commercially, but with the greatest effort in the lab? Michael Isaacson from the Laboratory of Submicroscopic Studies (appropriate for us) has made something under the direction of an artist friend of mine named Tom Van Sant. Van Sant is, I believe, the only truly modern artist I know. By truly modern, I mean a man who understands our culture and appreciates our technology and science as well as the character of nature, and incorporates them into the things that he makes.

I would like to show you, in the next slide, a picture by Van Sant. That's art, right? It represents an eye. That's the eyelid and the eyebrow, perhaps, and of course you can recognize the pupil. The interesting thing about this eye is that it's the smallest drawing a human being has ever made. It's a quarter of a micron across—250 millimicrons— and the central spot of the pupil is something like fifteen or twenty millimicrons, which corresponds to about one hundred atoms in diameter. That's the bottom. You're not going to be able to see things being drawn more than one hundred times smaller, because by that time you're at the size of atoms. This picture is as far down as we can make it.

Because I admire Tom Van Sant, I would like to show you some other artwork that he has created. He likes to draw eyes, and the next slide shows another eye by him. This is real art, right? Look at all the colors, the beauty, the light, and so forth—qualities that of course are much more appreciated as art. (Maybe some of you clever JPL guys know what you're looking at, but just keep it to yourselves, eh?)

To get some idea of what you're looking at, we're going to look at that eye from a little bit further back, so you can see some more of the picture's background. The next slide shows it at a different scale. The eye is now smaller, and perhaps you see how the artist has drawn the furrows of the brow, or whatever it is around the eye. The artist now wants to show the eye to us on a still smaller scale, so we can see a little more of the background. So in this next slide, you see the city of Los Angeles covering most of the picture, and the eye is this little speck up in the corner!

Actually, all these pictures of the second eye are LANDSAT pictures of an eye that was made in the desert. You might wonder how someone can make an eye that big—it's two and one-half kilometers across. The way Van Sant made it was to set out twenty-four mirrors, each two feet square, in special locations in the desert. He knew that when the LANDSAT passes back and forth overhead, its eye looks at the land and records information for the picture's pixels. Van Sant used calculations so that the moment the LANDSAT looked at a particular mirror, the sun would be reflecting from the mirror right into the eye of the LANDSAT. The reflection overexposed the pixel, and what would have been a two-foot square mirror instead made a white spot corresponding to an area of several acres. So what you saw in the first picture was a sequence of overexposed pixels on the LANDSAT picture. Now that's the way to make art! As far as I know, this is the largest drawing ever made by man.

If you look again at the original picture, you can see one pixel that didn't come out. When they went back to the desert, they found that the mirror had been knocked off its pedestal, and that there were footprints from a jack rabbit over the surface. So Van Sant lost one pixel.

The point about the two different eyes is this: that Van Sant wanted to make an eye much bigger than a normal eye, and the eye in the desert was 100 000 times bigger than a normal eye. The first eye, the tiny one, was 100 000 times smaller than a normal eye. So you get an idea of what the scale is. We're talking about going down to that small level, which is like the difference in scale between the two-and-one-half-kilometer desert object and our own eye. Also amusing to think about, even though it has nothing to do with going small, but rather with going big—what happens if you go to the next eye, 100 000 times bigger? Then the eye's scale is very close to the rings of Saturn, with the pupil in the middle.

I wanted to use these pictures to tell us about scale and also to show us what, at the present time, is the ultimate limit of our actual ability to construct small things. And that summarizes how we stand today, as compared to how the situation looked when I finished my talk in 1960. We see that computers are well on their way to small scale, even though there are limitations. But I would like to discuss something else—small machines.

Small Machines—How to Make Them

By a machine, I mean things that have movable parts you can control, that have wheels and stuff inside. You can turn the movable parts; they are actual objects. As far as I can tell, this interest of mine in small machines is a misguided one, or more correctly, the suggestion in the lecture "Plenty of Room at the Bottom" that soon we would have small machines was certainly a misguided prediction. The only small machine we have is the one that I've passed around to you, the one that Mr. McLellan made by hand.

There is no use for these machines, so I still don't understand why I'm fascinated by the question of making small machines with movable and controllable parts. Therefore I just want to tell you some ideas and considerations about the machines. Any attempt to make out that this is anything but a game—well, let's leave it the way it is: I'm fascinated and I don't know why.

Every once in a while I try to find a use. I know there's already been a lot of laughter in the audience— just save it for the uses that I'm going to suggest for some of these devices, okay?

But the first question is, how can we make small machines? Let's say I'm talking about very small machines, with something like ten microns (that's a hundredth of a millimeter) for the size of a rotor. That's forty times smaller than the motor I passed around—it's invisible, it's so small.

I would like to shock you by stating that I believe that with today's technology we can easily—I say *easily*—construct motors one fortieth of this size on each dimension. That's sixty-four thousand times smaller than the size of McLellan's motor. And in fact, with our present technology, we can make thousands of these motors at a time, all separately controllable. Why do you want to make them? I told you there's going to be lots of laughter, but just for fun, I'll suggest how to do it—it's very easy.

It's just like the way we put those evaporated layers down, and made all kinds of structures. We keep making the structures a little thicker by adding a few more layers. We arrange the layers so that you can dissolve away a layer supporting some mechanical piece, and loosen the piece. The stuff that you evaporate would be such that it could be dissolved, or boiled away, or evaporated out. And it could be that you build this stuff up in a matrix, and build other things on it, and then other stuff over it. Let's call the material "soft wax," although it's not going to be wax. You put the wax down, and with a mask you put some silicon lumps that are not connected to anything, some more wax, some more wax, and then silicon dioxide or something. You melt out or evaporate the wax, and then you're left with loose pieces of silicon. The way I described it, that piece would fall somewhere, but you have other structures that hold it down. It does seem to me perfectly obvious that with today's technology, if you wanted to, you could make something one-fortieth the size of McLellan's motor.

When I gave the talk called "Plenty of Room at the Bottom," I offered a thousand-dollar prize for the motor—I was single at the time. In fact, there was some consternation at home, because I got married after that, and had forgotten all about the prize. When I was getting married, I explained my financial position to my future wife, and she thought that it was bad, but not so bad. About three or four days after we came back from the honeymoon, with a lot of clearing of my throat I explained to her that I had to pay a thousand dollars that I had forgotten about—that I had promised if somebody made a small motor. So she didn't trust me too much for a while.

Because I am now married, and have a daughter who likes horses, and a son in college, I cannot offer a thousand dollars to motivate you to make movable engines even forty times smaller. But Mr. McLellan himself said that the thousand dollars didn't make any difference—he got interested in the challenge.

Of course, if we had these movable parts, we could move them and turn them with electrostatic forces. The wires would run in from the edges. We've seen how to make controllable wires—we can make computers, a perfect example of accurate control. So there would be no reason why, at the present time, we couldn't make little rotors and other little things turn.

Small Machines—How to Use Them

What use would such things be? Now it gets embarrassing. I tried very hard to think of a use that sounded sensible—or semisensible—you'll have to judge. If you had a closed area and a half wheel that you turned underneath, you could open and shut a hole to let light through or shut it out. And so you have light valves. But because these tiny valves could be placed all over an area, you could make a gate that would let through patterns of light. You could quickly change these patterns by means of electrical voltages, so that you could make a series of pictures. Or, you could use the valves to control an intense source of light and project pictures that vary rapidly—television pictures. I don't think projecting television pictures has any use, though, except to sell more television pictures or something like that. I don't consider that a use—advertising toilet paper.

At first I couldn't think of much more than that, but there are a number of possibilities. For example, if you had little rollers on a surface, you could clean off dirt whenever it fell, and could keep the surface clean all the time.

Then you might think of using these devices—if they had needles sticking out—as a drill, for grinding a surface. That's a very bad idea, as far as I can tell, for several reasons. First, it turns out that materials are too hard when they are dimensioned at this small scale. You find that everything is very stiff, and the grinder has a heck of a job trying to grind anything. There's an awful lot of force, and the grinder would probably grind down its own face before it ground anything else. Also, this particular idea doesn't use the individualization that is possible with small machines—you can individually localize which one is turning which way. If I make all the small machines do grinding, I've done nothing I can't do with a big grinding wheel. What's nice about these machines—if they're worth anything—is that you can wire them to move different parts differently at different times.

One application, although I don't know how to use it, would be to test the circuits in a computer that is being manufactured. It would be nice if we could go in and make contacts at different places inside the circuit. The right way to do that is to design ahead of time places where you could make contacts and bring them out. But if you forgot to design ahead, it would be convenient to have a face with prongs that you could bring up. The small machines would move their little prongs out to touch and make contact in different places.

What about using these things for tools? After all, you could drill holes. But drilling holes has the same problem—the materials are hard, so you'll have to drill holes in soft material.

Well, maybe we can use these tools for constructing those silicon devices. We have a nifty way of doing it now, by evaporating layers, and you might say, "Don't bother me." You're probably right, but I'd like to suggest something that may or may not be a good idea.

Suppose we use the small machines as adjustable masks for controlling the evaporation process. If I could open and close these masks mechanically, and if I had a source of some sort of atoms behind, then I could evaporate those atoms through the holes. Then I could change the hole—by changing the voltages—in order to change the mask and put a new one on for the next layer.

At the present time, it is a painstaking job to draw all the masks for all the different layers—very, very carefully—and then to line the masks up to be projected. When you're finished with one layer you take that layer off and put it in a bath with etch in it; then you put the next layer on, adjust it, go crazy, evaporate, and so on. And that way, we can make four to five layers. If we try to make four hundred layers, too many errors accumulate; it's very, very difficult, and it takes entirely too long.

Is it possible that we could make the surfaces quickly? The key is to put the mask next to the device, not to project it by light. Then we don't have the limitations of light. So you put this machine right up against the silicon, open and close holes, and let stuff come through. Right away you see the problem. The back end of this machine is going to accumulate goop that's evaporating against it, and everything is going to get stuck.

Well then, you haven't thought it through. You should have a thicker machine with tubes and pipes that brings in chemicals. Tubes with controllable valves—all very tiny. What I want is to build in three dimensions by squirting the various substances from different holes that are electrically controlled, and by rapidly working my way back and doing layer after layer, I make a three-dimensional pattern.

Notice that the silicon devices are all two-dimensional. We've gone very far in the development of computing devices, in building these two-dimensional things. They're essentially flat; they have at most three or four layers. Everyone who works with computing machinery has learned to appreciate Rent's law, which says how many wires you need to make how many connections to how many devices. The number of wires goes up as the 2.5 power of the number of devices. If you think a while, you'll find that's a little bit too big for a surface—you can put so many devices on a surface, but you can't get the wires out. In other words, after a while this two-dimensional circuit becomes all wires and no devices, practically.

If you've ever tried to trace lines in two dimensions to make a circuit, you can see that if you're only allowed one or two levels of crossover, the circuit's going to be a mess to design. But if you have three-dimensional space available, so that you can have connections up and down to the transistors, in depth as well as horizontally, then the entire design problem of the wires and everything else becomes very easy. In fact, there's more than enough space. There's no doubt in my mind that the ultimate development of computing machines will end up with the development of a technology—I don't mean my technology, with my crazy machines—but *some* technology for building up three-dimensional circuits, instead of just two-dimensional circuits. That is to say, thick layers, with many, many layers—hundreds and hundreds of them.

So we have to go to three dimensions somehow, maybe with tubes and valves controlled at small scale by machines. Of course, if this did turn out to be useful, then we'd have to make the machines, and they would have to be three-dimensional, too. So we'd have to use the machines to make more machines.

The particular machines I have described so far were just loose pieces that were moving in place—drills, valves, and so forth that only operate in place. Another interesting idea might be to move something over a surface or from one place to another. For example, you could build the same idea that we talked about before, but the things—the little bars or something—are in slots, and they can slide or move all over the surface. Maybe there's some kind of T-shaped slot they come to, and then they can go

up and down. Instead of trying to leave the parts in one place, maybe we can move them around on rollers, or simply have them slide.

Electrostatic Actuation

Now how do you pull them along? That's not very hard—I'll give you a design for pulling. [*At the blackboard, Feynman draws a rectangular block with a set of alternating electrodes creating a path for the block.*] If you had, for example, any object like a dielectric that could only move in a slot, and you wanted to move the object, then if you had electrodes arranged along the slot, and if you made one of them plus, and another one minus, the field that's generated pulls the dielectric along. When this piece gets to a new location, you change the voltages so that you're always pulling, and these dielectrics go like those wonderful things that they have in the department store. You stick something in the tube, and it goes whshhhht! to where it has to go.

There is another way, perhaps, of building the silicon circuits using these sliding devices. I have decided this new way is no good, but I'll describe it anyway. You have a supply of parts, and a sliding device goes over, picks up a part, carries it to the right place, and puts it in—the sliding devices assemble everything. These devices are all moving, of course, under the electrical control of computer stuff below them, under their surfaces. But this method is not very good compared to the present evaporation technique, because there's one very serious problem. That is, after you put a piece in, you want to make electrical contacts with the other pieces, but it's very difficult to make good contacts. You can't just put them next to each other—there's no contact. You've got to electrodeposit something or use some such method, but once you start talking about electrochemically depositing something to seal the contact, you might as well make the whole thing the other way by evaporation.

Another question is whether you should use AC or DC to do the pulling: you could work it either way. You could also do the same thing to generate rotations of parts by arranging electrostatic systems for pulling things around a central point. The forces that will move these parts are not big enough to bend anything very much; things are very stiff at this dimensional scale.

If you talk about rotating something, the problem of viscosity becomes fairly important. You'll be somewhat disappointed to discover that if you left the air at normal air pressure in a small hole ten microns big, and then tried to turn something, you'd be able to do it in milliseconds, but not faster. That would be okay for a lot of applications, but it's only milliseconds. The time would be in microseconds, if it weren't for viscous losses.

I enjoy thinking about these things, and you can't stop, no matter how ridiculous things get, so you keep on going. At first, the devices weren't moving—they were in place. Now they can slide back and forth on the surface. Next come the tiny, free-swimming machines.

Mobile Microrobots

What about the free-swimming machine? The purpose is no doubt for entertainment. It's entertaining because you have control—it's like a new game. Nobody figured when they first designed computers that there would be video games. So I have the imagination to realize what the game here is: You get this little machine you can control from the outside, and it has a sword. The machine gets in the water with a paramecium, and you try to stab it.

How are we going to make this game? The first problem is energy supply. Another one is controlling the device. And if you wanted to find out how the paramecium looks to the device, you might want to get some information out.

The energy supply is, I think, fairly easy. At first it looks very difficult because the device is free-swimming, but there are many ways to put energy into the device through electrical induction. You could use either electrical or magnetic fields that vary slowly, generating EMFs inside.

Another way, of course, is to use chemicals from the environment. This method would use a kind of battery, but not as small as the device. The whole environment would be used—the liquid surrounding the device would be the source of a chemical reaction by which you could generate power. Or you could use electromagnetic radiation. With this method you would shine the light on the device to send the signal, or use lower frequencies that go through water—well, not much goes through water but light.

The same methods can be used for control. Once you have a way to get energy in—by electrical induction, for example—it's very easy to put digits or bits on the energy signal to control what the machine is going to do. And the same idea could be used to send signals out. I shouldn't be telling people at JPL how to communicate with things that are difficult to get at or are far away—this is far away because it's so small. You'll figure out a way to send the signals out and get them back again—and enhance the pictures at the end.

It's very curious that what looks obvious is impossible. That is, how are you going to propel yourself through the liquid? Well, you all know how to do that—you have a tail that swishes. But it turns out that if this is a tiny machine a few microns long, the size of a paramecium, then the liquid, in proportion, is enormously viscous. It's like living in a thick honey. And you can try swimming in thick honey, but you have to learn a new technique. It turns out that the only way you can swim in thick honey is to have a kind of an "S" shaped fin. Twisting the shape pushes it forward. It has to be like a piece of a screw, so that as you turn it, it unscrews out of the thick liquid, so to speak. Now, how do we drive the screw?

You always think that there aren't any wheels in biology, and you say, "Why not?" Then you realize that a wheel is a separate part that moves. It's hard to lubricate, it's hard to get new blood in there, and so forth. So we have our parts all connected together—no loose pieces. Bacteria, however, have flagella with corkscrew twists and have cilia that also go around in a type of corkscrew turn. As a matter of fact, the flagellum is the one place in biology where we really do have a movable, separable part. At the end of the flagellum on the back is a kind of a disc, a surface with proteins and enzymes. What happens is a complicated enzyme reaction in which ATP, the energy source, comes up and combines, producing a rotational distortion [*here, Feynman is using his hands to simulate a molecule changing shape and experiencing a net rotation*]; when the ATP releases, the rotation stays, and then another ATP comes, and so forth. It just goes around like a ratchet. And it's connected through a tube to the spiral flagellum that's on the outside.

Twenty years ago when I gave my talk, my friend Al Hibbs, who introduced me today, suggested a use of small devices in medicine. Suppose we could make free-swimming little gadgets like this. You might say, "Oh, that's the size of cells—great. If you've got trouble with your liver, you just put new liver cells in." But twenty years ago, I was talking about somewhat bigger machines. And he said, "Well, swallow the surgeon." The machine is a surgeon—it has tools and controls in it. It goes over to the place where you've got plaque in your blood vessel and it hacks away the plaque.

So we have the idea of making small devices that would go into the biological system in order to control what to cut and to get into places that we can't ordinarily reach. Actually, this idea isn't so bad, and if we back off from the craziness of making such tiny things, and ask about a device that is more practical today, I think it is worth considering having autonomous machines—that is, machines that are sort of robots. I would tether the machines with thin wires—swallowing wires isn't much. It's a little bit discouraging to think of swallowing those long tubes with the optics fibers and everyting else that would have to go down so the guy can watch the inside of your duodenum. But with just the little wires, you could make the device go everywhere, and you could still control it.

Even the wires are really unnecessary, because you could control the machine from the outside by changing magnetic fields or electric induction. And then we don't have to make the motors, engines, or devices so very tiny as I'm talking about, but a reasonable size. Now it's not as crazily small as I would like—a centimeter or one half of a centimeter—depending on what you want to do the first few times, the scale will get smaller as we go along, but it'll start that way. It doesn't seem impossible to me that you could watch the machine with X-rays or NMR and steer it until it gets where you want. Then you send a signal to start cutting. You watch it and control it from the outside, but you don't have to have all these pipes, and you aren't so limited as to where you can get this machine to go. It goes around corners and backs up.

I think that Hibbs's "swallowable surgeon" is not such a bad idea, but it isn't quite appropriate to the tiny machines, the "infinitesimal machines." It's something that should be appropriate for small machines on the way to the infinitesimal machines.

Making Precise Things from Imprecise Tools

These machines have a general problem, and that's the refinement of precision. If you built a machine of a certain size, and you said, "Well, next year I want to build one of a smaller size," then you would have a problem: you've only got a certain accuracy in dimensions. The next question is, "How do you make the smaller one when you've only got that much accuracy?" It gets worse. You might say, "I'll use this machine to make the smaller one," but if this machine has wobbly bearings and sloppy pins, how does it make an accurate, beautiful, smaller machine?

As soon as you ask that question, you realize it's a very interesting question. Human beings came onto the earth, and at the beginning of our history, we found sticks and stone—bent sticks and roundish funny stones, nothing very accurate. And here we are today, with beautifully accurate machines—you can cut and measure some very accurate distances.

How do you get started? How do you get something accurate from nothing? Well, all machinists know what you do. In the case of large machinery, you take the stones, or whatever, and rub them against each other in every which way, until one grinds against the other. If you did that with one pair of stones, they'd get to a position at which, no matter where you put them, they would fit. They would have perfectly matched concave and convex spherical surfaces.

But I don't want spherical surfaces—I want flat surfaces. So then you take three stones and grind them in pairs, so that everybody fits with everybody else. It's painstaking and it takes time, but after a while, sure enough, you've got nice flat surfaces. Someday, when you're on a camping trip, and everything gets boring, pick up some stones. Not too hard—something that can grind away a little bit, such as consolidated or weak sandstones. I used to do this all the time when I was a kid in Boston. I'd go to work at MIT and on the way pick up two lumps of snow, hard snow that was pushed up by the snowplow and refrozen. I'd grind the snow all the way till I got to MIT, then I could see my beautiful spherical surfaces.

Or, for example, let's say you were making screws to make a lathe. If the screw has irregularities, you could use a nut that's breakable; you would take the nut apart and turn it backwards. If you ran the screw back and forth through the nut, both reversed and straight, soon you would have a perfect screw and a perfect nut, more accurate than the pieces you started with. So it's possible.

I don't think any of these things would work very well with the small machines. Turning things over and reversing and grinding them is so much work, and is so difficult with the hard materials, that I'm not really quite sure how to get increased precision at the very small level.

One way, which isn't very satisfactory, would be to use the electrostatic dielectric push-pull mechanism. If this device were fairly crude in shape, and contained some kind of a point or tooth that was used for a grinder or a marker, you could control the position of the tooth by changing the voltage rather smoothly. You could move it a small fraction of its own irregularity, although you wouldn't really know exactly what that fraction was. I don't know that we're getting much precision this way, but I do think it's possible to make things finer out of things that are cruder.

If you go down far enough in scale, the problem is gone. If I can make something one-half of a percent correct, and the size of the thing is only one hundred atoms wide, then I've got one hundred and not one hundred and one atoms in it, and every part becomes identical. With the finite number of atoms in a small object, at a certain stage, objects can only differ by one atom. That's a finite percentage, and so if you can get reasonably close to the right dimensions, the small objects will be exactly the same.

I thought about casting, which is a good process. You ought to be able to manufacture things at this scale by casting. We don't know of any limitation—except atomic limitations—to casting accurate figures by making molds for figures that match the originals. We know that already, because we can make replicas of all kinds of biological things by using silicone or acetate castings. The electron microscope pictures that you see are often not of the actual object, but of the casting that you've made. The casting can be done down to any reasonable dimension.

One always looks at biology as a kind of a guide, even though it never invents the wheel, and even though we don't make flapping wings for airplanes because we thought of a better way. That is, biology is a guide, but not a perfect guide. If you are having trouble making smooth-looking movable things out of rather hard materials, you might make sacs of liquid that have electric fields in them and can change their shapes. Of course, you would then be imitating cells we already know about. There are probably some materials that can change their shape under electric fields. Let's say that the viscosity depends on the electric field, and so by applying pressure, and then weakening the material in different places with electric fields, the material would move and bend in various ways. I think it's possible to get motion that way.

Friction and Sticking

Now we ask, "What does happen differently with small things?" First of all, we can make them in very great numbers. The amount of material you need for the machines is very tiny, so that you can make billions of them for any normal weight of any material. No cost for materials—all the cost is in manufacturing and arranging the materials. But special problems occur when things get small—or what look like problems, and might turn out to be advantages if you knew how to design for them.

One problem is that things stick together by molecular attraction. Now friction becomes a difficulty. If you were to have two tungsten parts, perfectly clean, next to each other, they would bind and jam. The atoms simply pull together as if the two parts were one piece. The friction is enormous, and you will never be able to move the parts. Therefore you've got to have oxide layers or other layers in between the materials as a type of lubricant—you have to be very careful about that or everything will stick.

On the other hand, if you get still smaller, nothing is going to stick unless it's built out of one piece. Because of the Brownian motion, the parts are always shaking; if you put them together and a part were to get stuck, it would shake until it found a way to move around. So now you have an advantage.

At the end of it all, I keep getting frustrated in thinking about these small machines. I want somebody to think of a good use, so that the future will really have these machines in it. Of course, if the machines turn out to be any good, we'll also have to make the machines, and that will be very interesting to try to do.

Computing with Atoms

Now we're going to talk about small, small computing. I'm taking the point of view of 1983 rather than of 1960, and will talk about what is going to happen, or which way we should go.

Let's ask, what do we need to do to have a computer? We need numbers, and we need to manipulate the numbers and calculate an answer. So we have to be able to write the numbers.

How small can a number be? If you have N digits, you know the special way of writing them with base two numbers, that is, with ones and zeros. Now we're going to go way down to the bottom—atoms! Remember that we have to obey quantum-mechanical laws, if we are talking about atoms. And each of these atoms is going to be in one of two states—actually, atoms can be in a lot of states, but let's take a simple counting scheme that has either ones or zeros. Let's say that an atom can be in a state of spin up or of spin down, or say that an ammonia molecule is either in the lowest or the next lowest state, or suppose various other kinds of two-state systems. When an atom is in the excited state—a spin up—let's call it a "one"; a "zero" will correspond to spin down. Hereafter when I say a one, I mean an atom in an excited state. So to write a number takes no more atoms than there are digits, and that's really nothing!

Reversible Gates

Now what about operations—computing something with the numbers? It is known that if you can only do a few operations of the right kind, then by compounding the operations again and again in various combinations, you can do anything you want with numbers.

The usual way of discussing this fact is to have these numbers as voltages on a wire instead of states in an atom, so we'll start with the usual way. [*Feynman draws a two-input AND gate at the blackboard.*] We would have a device with two input wires A and B, and one output wire. If a wire has a voltage on it, I call it a "one"; if it has zero voltage, it's a "zero." For this particular device, if both wires are ones, then the output turns to one. If either wire is one, but not both, or if neither is one, the output stays at zero—that's called an AND gate. It's easy to make an electric transistor circuit that will do the AND gate function.

There are devices that do other things, such as a little device that does NOT—if the input wire is a one, the output is a zero; if the input wire is a zero, the output is one. Some people have fun trying to pick one combination with which they can do everything, for example, a NAND gate that is a combination of NOT and AND—it is zero when both input wires are ones, and one when either or both inputs are not ones. By arranging and wiring NAND gates together in the correct manner, you can do any operation. There are a lot of questions about branchings and so forth, but that's all been worked out. I want to discuss what happens if we try to do this process with atoms.

First, we can't use classical mechanics or classical ideas about wires and circuits. We have atoms, and we have to use quantum mechanics. Well, I love quantum mechanics. So, the question is, can you design a machine that computes and that works by quantum-mechanical laws of physics—directly on the atoms—instead of by classical laws.

We find that we can't make an AND gate, we can't make a NAND gate, and we can't make any of the gates that people used to say you could make everything out of. You see immediately why I can't make an AND gate. I've only got one wire out and two in, so I can't go backwards. If I know that the answer is zero, I can't tell what the two inputs were. It's an irreversible process. I have to emphasize this fact because atomic physics is reversible, as you all know, microscopically reversible. When I write the laws of how things behave at the atomic scale, I have to use reversible laws. Therefore, I have to have reversible gates.

Bennett from IBM, Fredkin, and later Toffoli investigated whether, with gates that are reversible, you can do everything. And it turns out, wonderfully true, that the irreversibility is not essential for computation. It just happens to be the way we designed the circuits.

It's possible to make a gate reversible in the following cheesy way, which works perfectly. [*Feynman now draws a block with two inputs, A and B, and three outputs.*] Let's suppose that two wires came in here, but we also keep the problem at the output. So we have three outputs: the A that we put in, the B that we put in, and the answer. Well, of course, if you know the A and the B along with the answer, it isn't hard to figure out where the answer came from.

The trouble is that the process still isn't quite reversible, because you have two pieces of information at the input, that is, two atoms, and three pieces of information at the output. It's like a new atom came from somewhere. So I'll have to have a third atom at the input [*he adds a third input line, labeled C*]. We can characterize what happens as follows:

Unless A and B are both one, do nothing. Just pass A, B, and C through to the output. If A and B are both one, they still pass through as A and B, but C, whatever it is, changes to NOT C. I call this a "controlled, controlled, NOT" gate.

Now this gate is completely reversible, because if A and B are not both ones, everything passes through either way, while if A and B are both ones on the input side, they are both ones on the output side too. So if you go through the gate forward with A and B as ones, you get NOT C from C, and when you go backward with NOT C at the output, you get C back again at the input. That is, you do a NOT twice, and the circuit, or atom is back to itself, so it's reversible. And it turns out, as Toffoli has pointed out, that this circuit would enable me to do any logical operation.

So how do we represent a calculation? Let's say that we have invented a method whereby choosing any three atoms from a set of N would enable us to make an interaction converting them from a state of ones and zeros to a new state of ones and zeros. It turns out, from the mathematical standpoint, that we would have a sort of matrix, called M. Matrix M converts one of the eight possible combination states of three atoms to another combination state of the three atoms, and it's a matrix whose square is equal to one, a so-called unitary matrix. The thing you want to calculate can be written as a product of a whole string of matrices like M—millions of them, maybe, but each one involves only three atoms at a time.

I must emphasize, that, in my previous example with AND gates and wires, the wires that carried the answer after the operation were new ones. But the situation is simpler here. After my matrix operates, it's the same register—the same atoms—that contain the answer. I have the input represented by N atoms, and then I'm going to change them, change them, change them, three atoms at a time, until I finally get the output.

The Electron as Calculating Engine

It's not hard to write down the matrix in terms of interactions between the atoms. In other words, in principle, you can invent a kind of coupling among the atoms that you turn on to make the calculation. But the question is, how do you make the succession of three-atom transformations go bup-bup-bup-bup-bup in a row? It turns out to be rather easy—the idea is very simple. [*Feynman draws a row of small circles, and points often to various circles in the row through the following discussion.*]

You can have a whole lot of spots, such as atoms on which an electron can sit, in a long chain. If you put an electron on one spot, then in a classical world it would have a certain chance of jumping to another spot. In quantum mechanics, you would say it has a certain amplitude to get there. Of course, it's all complex numbers and fancy business, but what happens is that the Schrödinger function diffuses: the amplitude defined in different places wanders around. Maybe the electron comes down to the end, and maybe it comes back and just wanders around. In other words, there's some amplitude that the electron jumped to here and jumped to there. When you square the answer, it represents a probability that the elelctron has jumped all the way along.

As you all know, this row of sites is a wire. That's the way electrons go through a wire—they jump from site to site. Assume it's a long wire. I want to arrange the Hamiltonian of the world—the connections between sites—so that an electron will have zero amplitude to get from one site to the next because of a barrier, and it can only cross the barrier if it interacts with the atoms [of the registers] that are keeping track of the answer. [*In response to a question following the lecture, Feynman did write out a typical term in such a Hamiltonian using an atom-transforming matrix M positioned between electron creation and annihilation operators on adjacent sites.*]

That is, the idea is to make the coupling so that the electron has no amplitude to go from site to site, unless it disturbs the N atoms by multiplying by the matrix $M2$, in this case, or by $M1$ or $M3$ in these other cases. If the electron started at one end, and went right along and came out at the other end, we would know that it had made the succession of operations $M1$, $M2$, $M3$, $M4$, $M5$—the whole set, just what you wanted.

But wait a minute—electrons don't go like that! They have a certain amplitude to go forward, then they come back, and then they go forward. If the electron goes forward, say, from here to there, and does the operation $M2$ along the way, then if the electron goes backwards, it has to do the operation $M2$ again.

Bad luck? No! $M2$ is designed to be a reversible operation. If you do it twice, you don't do anything; it undoes what it did before. It's like a zipper that somebody's trying to pull up, but the person doesn't zip very well, and zips it up and down. Nevertheless, wherever the zipper is at, it's zipped up correctly to that particular point. Even though the person unzips it partly and zips it up again, it's always right, so that when it's finished at the end, and the Talon fastener is at the top, the zipper has completed the correct operations.

So if we find the electron at the far end, the calculation is finished and correct. You just wait, and when you see it, quickly take it away and put it in your pocket so it doesn't back up. With an electric field, that's easy.

It turns out that this idea is quite sound. The idea is very interesting to analyze, to see what a computer's limitations are. Although this computer is not one we can build easily, it has got everything defined in it. Everything is written: the Hamiltonian, the details. You can study the limitations of this machine, with regard to speed, with regard to heat, with regard to how many elements you need to do a calculation, and so on. And the results are rather interesting.

Heat in a Quantum Computer

With regard to heat: everybody knows that computers generate a lot of heat. When you make computers smaller, all the heat that's generated is packed into a small space, and you have all kinds of cooling problems. That is due to bad design. Bennett first demonstrated that you can do reversible computing—that is, if you use reversible gates, the amount of energy needed to operate the gates is essentially indefinitely small if you wait long enough, and allow the electrons to go slowly through the computer. If you weren't in such a hurry, and if you used ideal reversible gates—like Carnot's reversible cycle (I know everything has a little friction, but this is idealized)—then the amount of heat is zero! That is, essentially zero, in the limit—it only depends on the losses due to imperfections.

Furthermore, if you have ordinary reversible gates, and you try to drag the thing through as quickly as you can, then the amount of energy lost at each fundamental operation is one kT of energy per gate, or per decision, at most! If you went slower, and gave yourself more time, the loss would be proportionately lower.

And how much kT do we use per decision now? 10^{10} kT! So we can gain a factor of 10^{10} without a tremendous loss of speed, I think. The problem is, of course, that it depends on the size that you're going to make the computer.

If computers were made smaller, we could make them very much more efficient. It hadn't been realized previous to Bennett's work that there was, essentially, no heat requirement to operate a computer if you weren't in such a hurry. I have also analyzed this model, and get the same results as Bennett with a slight modification, or improvement.

If this device is made perfectly, then the computer could work ballistically. That is, you could have this chain of electron sites and start the electrons off with a momentum, and they simply coast through and come out the other end. The thing is done—whshshshsht! You're finished, just like shooting an electron through a perfect wire.

If you have a certain energy available to the electron, it has a certain speed—there's a relation between the energy and the speed. If I call this energy that the electron has kT, although it isn't necessarily a thermal energy, then there's a velocity that goes with it, v_T, which is the maximum speed at which the electron goes through the machine. And when you do it that way, there are no losses. This is the ideal case; the electron just coasts through. At the other end, you take the electron that had a lot of energy, you take that energy out, you store it, and get it ready for shooting in the next electron. No losses! There are no kT losses in an idealized computer—none at all.

In practice, of course, you would not have a perfect machine, just as a Carnot cycle doesn't work exactly. You have to have some friction. So let's put in some friction.

Suppose that I have irregularities in the coupling here and there—that the machine isn't perfect. We know what happens, because we study that in the theory of metals. Due to the irregularities in the positions or couplings, the electrons do what we call "scattering." They head to the right, if I started them to the right, but they bounce and come back. And they may hit another irregularity and bounce the other way. They don't go straight through. They rattle around due to scattering, and you might guess that they'll never get through. But if you put a little electric field pulling the electrons, then although they bounce, they try again, try again, and make their way through. And all you have is, effectively, a resistance. It's as if my wire had a resistance, instead of being a perfect conductor.

One way to characterize this situation is to say that there's a certain chance of scattering—a certain chance to be sent back at each irregularity. Maybe one chance in a hundred, say. That means if I did a computation at each site, I'd have to pass a hundred sites before I got one average scattering. So you're sending electrons through with a velocity v_T that corresponds to this energy kT. You can write the loss per scattering in terms of free energy if you want, but the entropy loss per scattering is really the irreversible loss, and note that it's the loss **per scattering, not per calculation step** [*heavily emphasized, by writing the words on the blackboard*]. The better you make the computer, the more steps you're going to get per scattering, and, in effect, the less loss per calculation step.

The entropy loss per scattering is one of those famous $\log_2$ numbers—let me guess it is Boltzmann's constant, k, or some such unit, for each scattering if you drive the electron as quickly as you can for the energy that you've got.

If you take your time, though, and drive the electron through with an average speed, which I call the drift speed, v_D (compared to the thermal speed at which it would ordinarily be jostling back and forth), then you get a decrease in the amount of entropy you need. If you go slow enough, when there's scattering, the electron has a certain energy and it goes forward-backward-forward-bounce-bounce and comes to some energy based on the temperature. The electron then has a certain velocity—thermal velocity—for going back and forth. It's not the velocity at which the electron is getting through the machine, because it's wasting its time going back and forth. But it turns out that the amount of entropy you lose every time you have 100% scattering is simply a fraction of k—the ratio of the velocity that you actually make the electron drift compared to how fast you could make it drift. [*Feynman writes on the board the formula:* $k(v_D/v_T)$.]

If you drag the electron, the moment you start dragging it you get losses from the resistance—you make a current. In energy terms, you lose only a kT of energy for each scattering, not for each calculation, and you can make the loss smaller proportionally as you're willing to wait longer than the ideal maximum speed. Therefore, with good design in future computers, heat is not going to be a real problem. The key is that those computers ultimately have to be designed—or should be designed—with reversible gates.

We have a long way to go in that direction—a factor of 10^{10}. And so, I'm just suggesting to you that you start chipping away at the exponent.

Thank you very much.