2016 GPU Technology Conference (GTC16) - Dr. Gill Pratt

April 07, 2016

2016 GPU Technology Conference (GTC16) Keynote
San Jose, Calif.
April 7, 2016
Dr. Gill Pratt, CEO of Toyota Research Institute

_____

Intro: NVIDIA Chief Scientist and Senior Vice President of Research Bill Dally:

Every year, 1.2 million people are killed in traffic accidents. Many more are maimed and injured. Artificial intelligence holds a promise to save millions of lives by making our cars smart enough to avoid accidents. At the same time, it promises to offer mobility to people unable to operate a vehicle.

The man at the forefront of this smart car revolution is Gill Pratt. Gill is CEO of Toyota Research Institute, an R&D center focused on AI and robotics. I first met Gill when we were both faculty at MIT. There, he made pioneering contributions to digital systems and led the MIT Leg Lab. We vacationed together at a colleagues resort in Maine.

From MIT, Gill went to Olin College, where he was professor and associate dean of research and faculty affairs. He was later a program manager at DARPA DSO, where he led the DARPA Robotics Challenge, a $100 million competition to develop a disaster response robot. Ladies and gentlemen, please join me in welcoming Dr. Gill Pratt.

Toyota Research Institute CEO Dr. Gill Pratt:

Hi, everybody. It's great to be here. Thank you very much, Bill, for that introduction. I want to start by repeating a little bit of what Bill said, and I want you to keep this in your mind: 1.2 million people.

1.2 million is an incredibly high number. For those of you that care about peace on Earth, 1.2 million people a year far exceeds the number of people killed in war, on average.

And so, the fact that we're willing to tolerate 1.2 million people per year around the world, killed in car accidents, is actually extraordinary. And it's a shame.

So today, I'm going to come back to that number, but I'm going to actually start with a deep technical discussion on two things, and then go into a little bit of an announcement about our new efforts in this area.

Let me start by talking about power. And by power I mean electrical power. And a question that I've been thinking a lot about recently is: How much electrical power should it take to drive an autonomous car?

Well, here are some examples. Now these are somewhat early examples, and modern cars are a little bit better, but current autonomous vehicles take thousands of watts of power to run their systems in perception and

planning.

Here's another solution: Thirty watts of power (picture of human driver).

Wow. A human brain takes 30 Watts — oh, and guess what, it works even while daydreaming. I don't know how many Watts go to daydreaming and how many to driving, but we've all had that experience, haven't we, of arriving home after driving and we've been thinking about something else at the same time and we don't even know how we got there? So how does nature do that? How is it possible that we use so few watts — tens of watts of power — and the autonomous vehicles that we build today use thousands of watts?

When I was at DARPA, before coming to Toyota, I worked on a number of projects both in robotics and neuromorphic computing. And I'm going to give you an example from the world of robotics, from mechanical actuation, that seems to have nothing to do with computation, but you'll see that it does.

So here's a chart of the specific resistance — that's how much thrust it takes for a given amount of weight — to do locomotion versus the speed of that locomotion on the X axis. And what you'll see on this chart, on the top, is a number of robots that take a lot of power to run. Their specific resistance is very high.

Amongst those robots is one called LS3, which is roughly the size and weight of a horse. And another robot on the other side there, called Atlas, which is roughly the size and weight of a very large person. And what you'll see on this chart is that the specific resistance of those two systems ended up being 100 times worse than that of the natural systems that they emulated.

So a person is 100 times more efficient at locomotion than a robot that looks like a person. And a horse is 100 times more efficient than a robot that looks like a horse. And so we began to try to deeply understand why it was that these systems were so inefficient, when we built artificial ones.

And the answer was that nature cares a lot about power efficiency. After all, if you're an animal and you're out in the wild, finding food is the most difficult thing to do. In fact, if finding food isn't hard to do, your species will multiply and get numerous until food is hard to find.

And so the standard state of nature, which we don't really perceive that much because we go to the grocery store, is for food to be scarce. And so energy efficiency is crucial. Well, one thing is not scarce, and that's complexity. Our bodies are incredibly complex. And our muscles have something very complex which is called variable recruitment.

We have thousands and thousands of nerve fibers that go into each one of our muscles, and different muscle fibers turn on depending on what the load is. And so at all times the muscle fibers that are being used are selected so that they're operating at the peak of their efficiency curve. Nature is really extraordinary in how it does that.

This is all to make it possible for us to do what we do without having to eat very much. So it's a general principle — complexity is inexpensive in nature, energy is very expensive in nature. Kind of the opposite of the situation with mechanical systems we manufacture.

When we build something, complexity is very expensive, and energy, well, we can always plug into the wall. That doesn't cost us so much because it's all made from fossil fuels and other very high-density energy sources.

So let me switch now from the mechanical world to the computational world. I also ran a program at DARPA

called Neovision2. This was trying to see if there is any advantage to using neuromorphic techniques to do computer vision. And this program actually ran right before the big growth of deep learning.

Neovision2 found exactly the same result. On this chart, on the right-hand side, are some baseline computer vision systems that we built. And the Y-axis is an axis of performance, and the X-axis is an axis of how many nanojoules per uncompressed bit of visual information were used. How much power was used in the computation?

And so the red and the green on the right-hand side are the baseline. It's a logarithmic scale. And the best of the neuromorphic systems actually was four orders of magnitude lower power consumption than the baseline ones. And they were ones that were architected based on the way that neural visual systems work, the way that the brain works.

So what was the trick here? It was the same thing. In these systems, instead of having complexity being expensive, we imagined the complexity was free. And we unrolled the computation, used a lot of silicon to do the job, and each particular part of the hardware was only used for one purpose.

And when that hardware wasn't being used, because there wasn't particular work to be done, the hardware was turned off. Dark silicon. And so this was much more expensive from a complexity point of view, but it turned out to have orders of magnitude of improvement in energy efficiency. And so we did another program.

This was another program I ran, called SyNAPSE, and this was, again, trying to see what are the advantages of building computers inspired by the nervous system. And so let's take a look at the two different parts here.

In the natural system – brains – complexity is less constrained. We have 10 to the 14th synapses in our brains. At the time that we did this program, the Xbox One system on a chip was the largest chip that was out there, and it had around 5 billion transistors.

So brains were around 20,000 times — if you count synapses and transistors as being roughly equivalent — 20,000 times more complex. But in the natural system, the size, weight, and power are highly constrained; and in the computer, size, weight, and power are much less constrained.

The human brain, as I mentioned, consumes 30 watts. We did a simulation using a supercomputer, one of the biggest ones at the time in the world, which specifically was built for power efficiency, and to run a statistical simulation of the human brain at 1/1000 of real-time took 8 million watts — 8 million is a lot more than 30.

In fact, if you normalize with respect to the real-time nature, you have a factor of 500 million. Weight is also a fun one to look at. That computer weighed 500,000 pounds. A human brain weighs around 3 pounds. So there’s a slight difference there, too.

So, again, the key difference here is, in nature, complexity is almost free – in manmade systems, complexity is very expensive. But the tradeoff is that natural systems can be much more power efficient.

As we begin to think about these systems for cars, can we do the same thing? Can we actually — because we need a lot of computation in cars — can we begin to really get the power down so that these systems will work well?

That particular program ended up with the system called the IBM True North system. It's one way of doing things. This chip was roughly the same size as the Xbox system on a chip, around 5 billion transistors, and it

used only around 50 milliwatts of power per chip.

So that resulted in a rough ratio — again, this is extremely rough — of around 1,000 times worse than the human brain. But that's still much better than the supercomputer that had it before. So I would like to see future systems that explore this kind of territory and go further, even beyond what GPUs can do now.

So what are my conclusions? Well, this is an NVIDIA conference, so of course the first conclusion is buy more hardware. But I'm serious. By buying more hardware and leaving most of it off most of the time, which is my last conclusion, you save total energy. And it can be orders of magnitude of difference.

Probably the architecture will have to change, too, but I predict that in the future we're going to see systems with much more silicon, much of it dark most of the time. It gets that power efficiency because we unrolled computation, we avoid multiplexers that steer data around from long distances within the chips themselves, and, as a result, the communication is very low energy, and the system can be very efficient.

So that's my conclusion for the first part of the talk. Let me go ahead to the second part. Parallel autonomy.

There's been a tremendous amount of talk about different levels of autonomy. Level two autonomy, just as a refresher, is one where you are driving the car.

You press a button, say "Drive for me," but at any moment, the car may say "This is too hard for me, I'm reverting control back to you." So at level two, you have to be vigilant all of the time that suddenly the car may turn control back to you.

Level three. What's going on with level three autonomy? There, the car will give you a warning, some reasonable length of time, to reacquaint yourself with the road and reengage. So, for instance, you may be driving down the road, the car sees something up ahead it can't deal with, alarm sounds, and you have between a few seconds and let's say 30 seconds to get back.

Level four autonomy. You can trust the car to do the whole journey on its own. You can go to sleep, you can have a conversation with your friend, you can do whatever you want. So the question here is, do we have to have level four to deal with this so-called handoff problem? The handoff problem is what if the car says "I can't do this anymore," hands it back to you, and you're not actually ready to take control in the time that's available.

Well, it turns out that some of the work I did at DARPA also has something to say about this too. So as Bill mentioned, I ran the DARPA robotics challenge. We had 24 teams from various parts of the world participate in the challenge. The total amount of money of the entire challenge that we spent was around $100 million. There were some multimillion-dollar prizes at the end.

What was very interesting about the challenge — we had a virtual challenge at the beginning when we went from over 100 teams down to a handful of teams. That was done completely in simulation. And it was done in simulation on the Internet in real time. It was actually kind of a first. And we had people and robots interacting with each other throughout the whole world.

Based on the results of the simulation, we gave out some robots that were humanoid robots for some of the teams to use. Other ones were funded from proposals that they sent in. And they built their own machines. And then, other teams, the bottom part of this chart here, had actually built systems on their own and funded it all on their own.

We had a preliminary physical challenge, which was pretty good. And then, we had a final challenge that showed us really how good we could do in this whole thing. Here's what some of the results look like. On the left-hand side, you see a robot. This happened to be one from Japan that did not do so well. On the right, you see the response of a human being to that small catastrophe.

I'm sure a lot of you have heard about the uncanny valley. This is when a robot looks a little bit like a person but not quite, and we get a little bit freaked out. Well, it turned out the uncanny valley was an effect that happened at the robotics challenge, but an even stronger effect was human empathy for the machines that were trying to do these tasks.

People just completely locked in to these machines. And when they fell over and had difficulty, they acted just like you see in this picture here. In fact, I was at a news conference after the MIT robot had fallen and broken one of its arms and a reporter — almost crying — said to me, "Is the MIT robot going to be okay?"

And you know, I sort of stopped and thought about it. And I was like, "Okay. Well, it's a machine. So let's all calm down." And in fact, it got a brand-new arm for the next day.

So as much as we worry about the uncanny valley with regard to technology in our lives, there is another effect of human empathy for machines, or anthropomorphizing of machines that I think is equal if not stronger. And it's actually a quite positive thing.

Of course, some of the robots did great. Here's the winner of the final challenge from the Korean [Advanced] Institute of Science and Technology. And this is what the crowd thought about it. And all it did was go up a flight of stairs. But it was quite an accomplishment.

Scientifically though, we learned a tremendous amount. And the essential part of the DARPA robotics challenge was not about balance or locomotion. It was about people and machines working together across a difficult communications channel. And this has a lot to do with cars, which is the reason that I'm talking about it now.

First, we came up with an idea that we believe that the autonomy that's required in a system is roughly inversely proportional to the Shannon entropy of the communication going back and forth between the two. You know this from your own human experience. If your child is more autonomous, they don't talk to you as much. If they're less autonomous, they're constantly asking questions about what they should do.

So because we were emulating disasters, we put a box in between the people and the robots that emulated the degradation in communication that happened during a disaster. So the people are on the left-hand side of the screen. They're inside of an isolated room. The robots are on the right-hand side of the screen out in the field. The idea is that it's too dangerous for people to go very close to where the robots are working.

And in between, we put this network degradation device. Now, some of the time, we actually cut the network traffic off completely. We had blackouts that were up to 30 seconds long. And somehow, the robot had to keep operating in order to do well in the challenge even though it was cut off from the human beings on the left.

How did it do that? And the answer was models. The robot had a model of what the human operator would have it do. And the human operators had a model of what was going on inside of the robot's field and what was going on out in the environment. And both of those models were being simulated in real time.

The robot continued to act as if it was acting on behalf of the person. And the person continued to have this kind of simulation-based situational awareness of what was going on on the other side. And this technique of using

simulation in a model-based controller on both sides was highly effective in allowing these machines to work despite interruptions of communication on both sides.

So this is one model of autonomy. If you take it to the extreme, you end up with this. This was another program that I had run called Autonomous Robotic Manipulation. And in this case, we had a robot changing a tire. This was from the Jet Propulsion Lab.

And it's doing this entirely on its own. It's using machine vision in order to perceive the environment without any fiducials or any other hints and some force-controlled and torque-controlled arms to take the tire off. And it even manages to wiggle the tire back on to the studs, which — if any of you have done that, that's a relatively difficult task to do.

So you can do it with no communication at all. But the question is, is this the only model that we have of autonomy, of a commander giving a command to a robot, whether it's a car or a physical machine different than a car and the autonomy then doing the next thing in series?

And the answer is no. There are actually many different ways that autonomy can be used. So let me go through these. Number one is the one that I've been talking about, series autonomy, things that are in series with each other.

So we have a commander. And we have a subordinate. It doesn't matter if these are two human beings or whether it's a person and a robot. And the commander gives the subordinate a command. Change the tire.

And the subordinate goes and executes that in interaction with the environment. And then, of course, the commander gets to see the result one way or the other and then issue new commands. If you think about it in terms of a parent and a child playing golf, this would be the parent tells the child, "Hit the ball over there."

And this is what I show in the picture over on the right. And hopefully, the child manages to do it somewhat well. And then, they go, and they do it again and again and again. Now, there are other modes of autonomy.

So one way that two agents can act — one somewhat autonomous from the other — is, by example, the interleaved method, a pilot and a co-pilot. And they take turns. And one says, "You have control." And then, the one says back to the other one, "Now you have control."

And if you think about the handoff problem, that's very much the kind of thing that goes on in a car. A driver will drive for a while, engage the autonomy in the car. Then, the car drives for a while by itself. And then, perhaps because of an inability to see very well, the car gives a handoff back to the user. "You have control now."

But it turns out there's a third mode. And this is at the bottom of the screen here. And it's called parallel autonomy. It's an exact dual to series autonomy, which I talked about at the beginning. And here, you have the parent and the child working together in parallel. [Their control efforts sum]. They both have their hands on the golf club.

And the words that I use here is "I'll help you swing." And they're both trying to act at the same time. And the child kind of learns from the parent because they can sort of feel what it's like to actually do the task.

And in fact, if you think about this loop of teaching, the ultimate goal of the parent is to use less and less control effort, fewer and fewer forces as the learning occurs, until the child can completely do it on their own.

But this dual to series autonomy, of parallel autonomy, is an idea that actually we studied at DARPA and that I'm moving right now into what we're doing with cars. So let me show you the DARPA work first.

This is a rather extraordinary thing. We have a human being who is a quadriplegic. And she has a brain-computer interface implanted in her skull, which is looking at the motor system, the motor neurons, and plugged into that as part of her brain.

In this particular case, the patient had had this implant for a number of years. And it had begun to go bad because the immune system reacts to that and begins to coat the electrodes in a way that the signal-to-noise ratio goes down.

On the left-hand side is a robotic arm — it turns out it's the same one that we had used on the tire-changing task — that is directly being controlled by the best encoding system that we could think of with her brain in order to do a task. What is the task? Pick up the three-dimensional cube and try to place it in the white box.

On the left, she's trying to do it on her own. She can't do it. On the right, we have a parallel autonomy system. We have the equivalent of the parent with the child, saying, "Here, I'm going to help you do the task." And that parallel autonomy system is watching what she's doing, and it's inferring her intent by watching what she's trying to do with the arm, both X, Y, and Z; and also open and close of the gripper.

And every time that she's having trouble, it actually gives her a little bit of a shove, saying, "I think you're trying to do this," and helps by adding torques and forces to the arm and controlling the gripper. And of course, what you'll see is that it's in fact quite successful. It was successful 100 percent of the time, which is really remarkable.

But the most remarkable thing about this experiment — and this, by the way, was done with both CMU and the University of Pittsburgh — is that she didn't know when we were turning on the assistance and when we were turning it off.

In fact, she felt that she was responsible for doing it on her own roughly independently of when we turned it on and when we turned it off. She felt a sense of agency that was maintained throughout the whole test. And so it's kind of an interesting way of doing things, different than giving the robot a command, saying, "Pick up the box for me." It's having the robot help her do the task herself.

So, how does this apply to autonomous driving? Well, I came to Toyota — this is a picture of the president of Toyota, Mr. Akio Toyoda, and myself. And I learned a whole bunch of things about the company. The first thing that I learned was that Akio has a sense of humor.

And this is a picture of me when I had much more hair, doing a brake job, and I had shared this with some lower level folks in the company, but he decided he really wanted to use it during the introduction of the company. And what's really great, if you look carefully at the picture on the right — you can barely see — it's a Toyota that I'm fixing at the time. So, that was good.

But more seriously, he told me what his priorities were. Number one, safety. Remember that number we talked about at the beginning. It is incredibly important to our company and I'm sure the other car companies, as well. Number two, the environment. We care tremendously about climate change. We have a hydrogen fuel cell project that many of you have heard about, and we have an incredibly liberal IP policy with regard to fuel cells.

Number three, mobility for all. It's incredibly important to us that people can have the independence, the dignity, the autonomy themselves of being able to move around in the world, and to have high quality of life regardless of their infirmity, regardless of their age. We think that's tremendously important.

So, if you are too young to drive, if you are too old to drive, if you are sick, if you're tired, any of these reasons, we want to help you to feel the independence and the quality of life benefits that are true in mobility.

And last and also important is that President Toyoda loves to drive. In fact, he's a race car driver. And his office is filled with race car stuff. And he maintains — and I think rightly so — that the fun of driving, that sort of thrill that you get that this machine is amplifying your natural evolved desire to be mobile — should be maintained. We shouldn't give this up. And so, as we think about autonomy in cars, he said to me, "Let's not forget the thrill of driving."

Okay. So, I met the president. I then couldn't sleep the next night. Because I began to think about how hard this job was that I had taken on. And these were the numbers that really got to me. I learned that Toyota makes 10 million cars per year. I calculated in my head, well, okay; order of magnitude, each one lasts around 10 years, so that means there's around 100 million Toyotas in service around the world.

Well, each one of them is driven how much? It's not 100,000 miles, it's not 1,000 miles. Order of magnitude, 10,000 miles per year. Multiply those numbers out, we have Toyotas driving around the world a total of a trillion miles every year, a trillion miles, 10 to the 12th.

Well, I knew that the very best of the autonomous car R&D places had tested cars in the millions of miles. This was a million times more. And furthermore, I knew from history that any car manufacturer that has only a few defects that cause accidents each year, can actually cause an existential crisis to the company. And this is true for several companies in history.

And so, how do I match these things up? Is it really possible to build an autonomous car that will only have a handful, if not fewer, mistakes that either cause an injury or a death, over a trillion miles of operation? It's incredibly, incredibly difficult.

So, I turned to these lessons that I had learned at DARPA, and in particular I looked at this question of series autonomy versus parallel autonomy. And in series autonomy, the nomenclature that's used in the car industry is, this is the chauffeur mode. Think of a person that drives you around, whether it's Uber or a taxi, anything like that.

Well, the chauffeur has to drive 100 percent of the time. If the chauffeur is autonomy that's built by the car company, they're 100 percent liable. It has to handle any driving at any time, no matter what, unless the route is particularly constrained. And you can't deploy it until it works perfectly.

And what that means unfortunately is that number of 1.2 million, which is an extraordinarily high number, you can't have an impact on that until you're completely done developing to that level of perfection.

It ignores the abilities that the driver in the car, who now is turned into a passenger. And of course, in terms of being fun to drive, is it actually fun to drive anymore, or are you now a passive passenger in a car even if you're in the front seat? Worst of all, this hand-off problem — if it's not perfect and occasionally does have to hand things off to you — becomes a very, very difficult issue.

And so, that's the world of the chauffeur mode that we're talking about.

Now, for many years, cars have had these assistant systems that we think about like a guardian angel. It's actually parallel autonomy. It's a system that's watching out for you, trying to prevent you from having an accident. The guardian angel intervenes only when you're about to get into trouble.

Anti-lock brakes are a great, simple example of parallel autonomy. You're having a skid because of the ice, you want to keep pressing down on the brake pedal. The guardian angel says, "No, that's not the right thing to do," and relieves some of the pressure inside of the brake system, puts it back on, and effectively pumps the brakes for you even though you may be giving a different command.

Well, what's it like in guardian angel, if we took that idea of anti-lock brakes, put it on steroids and say, "What if the car actually would intervene whenever it thought that you might get into an accident, and temporarily took control and then handed it back to you?"

Well, first of all, if you're a normal driver and aren't particularly being reckless, we would imagine far less than 1 percent of the time is when the guardian angel would have to intervene, only if an accident is imminent. The liability is mostly with the driver, only occasionally with the autonomy and the manufacturer.

The standard of how good the system needs to be is actually not the level of perfection the chauffeur needs. Your anti-lock brakes don't guarantee you won't have an accident. What they do guarantee is that they won't make things worse. And so, "do no harm" is the correct standard when thinking about these guardian angel systems.

How about saving lives? Do we have to wait until the system is perfect? Well, the answer is no. In fact, anti-lock brakes have saved many lives now. Stability control has saved lives. There's something called automatic emergency braking, which is when the car stops if it's about to hit something in front. And Toyota is actually the leading manufacturer, and all of our U.S. cars, no matter how inexpensive, by the end of 2017 will have automatic emergency braking.

So, these systems can save lives now. You don't have to wait until you're perfect and can drive on the road 100 percent of the time.

How about the fun of the car? If you want to have fun in your car, and the guardian angel is protecting you, you can actually push the car further to its performance limits and still feel safer. Now, do we want you do that? Probably not. Okay. But the truth is, it certainly doesn't take away from the fun of the car, and potentially could add to it too.

And in terms of the hand-off problem, is there ever a time where you turn control over to the autonomy and say, "I expect you to do this from now on," and then are surprised because it says, "Alert, I can't do this anymore. Now it's your turn"? That never happens. You're trying to drive all of the time, and it's just assisting you, and occasionally assisting you a lot by saying, "Watch out, you really don't want to do that." And so, this we feel is a very valid approach.

Now, the right way to think about this is that they're orthogonal. It's actually two different kind of dual ways of looking at the problem. One in parallel, one in series. One guardian angel, one chauffeur. And I put up on this chart here a number of technologies, as they've advanced with time for improving autonomy, both in series and in parallel.

And many companies are sort of fixated on the chauffeur model for doing this and have various debates on "Do I go straight to level four, and do I sort of hop up vertically all the way, or do I do an incremental approach on the

vertical axis?" What we're trying to say is that there's this horizontal axis that people have been working on a long time, that can also be pushed much further, and that's the guardian angel approach, as well.

And so these dashed diagonal lines here — the double diagonal lines are meant to show there's a whole lot more work that we can do, in fact, on both of these things. We intend to do both. We are trying to improve both safety and also access. And if you're talking about a person who really can't drive, the guardian angel is not going to make the car drivable. And so to truly have access we need chauffeur also.

And so Toyota intends to develop both of these things, but we believe that there's a lot of reasons to push on the parallel axis as a sort of second approach to the vertical axis. We're not making claims about doing one will get you to the other — in fact, it's true that perception and planning on either one of these axes does help the other — but we're saying that both are important and that horizontal axis is a tremendously valuable method of approach.

So let me summarize this part here. Toyota Research Institute has three autonomy-related goals. Number one, our goal is to improve safety. Number two, chauffeur mode is important as well, in order to improve access to cars.

There's a goal that I haven't talked about, that also harkens back to the DARPA work and that is we believe the technology for mobility outdoors can also be used for mobility indoors, and, in fact, Toyota can diversify itself and apply its strength in design and manufacturing and sales and support to aging society and robotics that is inside the home as well.

So that's what we're doing, in general, at TRI. We have some other efforts as well. But these are the three main ones that relate to autonomy.

But if you've been paying attention, I actually haven't answered the trillion-mile question. How are we going to handle reliability at a trillion miles? And we have several approaches that we're taking. But one of them that we think is very important is simulation. And this is actually an area where we're collaborating with NVIDIA, and so I want to talk about it here.

Why do we use simulation? The picture that you see here is an incredible six degree of freedom, multiple football field size simulator. It actually has a car inside of that dome, and you sit inside of this thing, and it slews around like an XY plotter, as well as tilting. And, of course, it has an incredibly beautiful graphics display in front of it that gives you an immersive experience of actually being in a car.

So both from an inertial sense and also a visual sense, you can practice driving in this. If you think about guardian angel mode, the car is going to have to push back at you. You're turning the wheel to try to switch lanes, and it's a bad idea. The car warns you with beepers and lights. But in the end, you're still trying to do it. The guardian angel would grab the wheel and pull it back. How are you going to respond?

How are you going to respond if you're doing that while the car is having a skid and you're getting these inertial cues that are funny? We can only test that through simulation. We can't have the danger of doing that in the real world.

Second, this is incredibly important, and I believe all of the companies that are working on autonomous driving do it. I saw some books for sale out in the hall on writing clean software. Well, regression testing matters a lot. And so we can use simulation to play back all the logs, and I believe all the companies are doing it. But we need accelerators for that. And working with GPUs is a very important part of it.

And then the final thing, and this isn't unique to TRI, is using simulation to accelerate and amplify physical testing. A million miles, two million miles, three million miles of physical driving is not a substitute and not adequate for the trillion-mile reliability that we need to show. And so we're planning to do that as well, and NVIDIA has been a wonderful partner with us in trying to push this thing forward.

Let me talk about the last part of my talk here. It's actually on people. We have a little bit of an announcement to make. Here's what TRI's plan was initially. We planned to have 150 people in Palo Alto working on this new guardian angel approach. It turned out that a team in Japan had been working for many years on the chauffeur approach, as well.

And in Cambridge, near MIT, our plan is to have around 50 people, and their focus is on simulation. Now, we have people on both coasts that are actually working on stuff having to do with the other side, so this isn't absolute, but that's the general focus. The news is that we're adding a third site.

And this is in Ann Arbor, Michigan, right near the university. And like the Cambridge facility it will have approximately 50 people, and this site will actually take a lot of the work that's been done in Japan and some that has been done at one of the Toyota centers in Ann Arbor before and accelerate it much more, and that will concentrate mostly on chauffeur.

Why did we pick Ann Arbor? First of all, and most importantly, the University of Michigan is right there. It's an incredible gem and has produced tremendous work in artificial intelligence. Right next to the university is a place called M City, and it's the mobility transformation center, and is a place Toyota has funded before. And it is, if you want to think about it, a physical simulator for doing autonomous driving. And so it's a great place to run experiments.

Recently, the American center for mobility has been proposed, and this is at the Willow Run site, which is quite a large site, and we think that that will be wonderful, too, and we look forward, assuming that the plans go through, to using that site, as well.

And then finally we actually have some very large facilities in Ann Arbor right now and other parts of Michigan, and we think that it's really good to invest even further. The people who are going to lead the efforts in autonomy there are two professors from the University of Michigan. One is Professor Ryan Eustice, who will work primarily on mapping and localization. And the other one is Professor Ed Olson, who will work primarily on perception.

These are two incredibly strong professors in robotics and AI who have been working on automobile autonomy for some time. Those of you that have been following this stuff know they used to work for Ford. Now, often in this industry, there is a lot of news where people care a whole lot about the supposed soap opera of people moving from one company to the other. Well, the answer is this happens all the time and is actually not a big deal.

Our company, TRI, is brand new, and of course we're going to have people coming to us from many different parts of the industry. Here's another guy that used to work for Ford. And I'm flattered by you taking pictures — that's my dad back in 1961. This was six months before I was born. So he used to work for Ford, too.

The general idea here, though, and what I want to emphasize the most is regardless of where people come from — we have some people who used to work at Google, now some people that used to work at Ford — we think that co-opetition in this industry is absolutely key. The autonomous car field right now is incredibly hot. Deals

are being made for billions of dollars. It's just amazing. And it's important to remember that we don't have to always work alone.

So this is a picture from 1950. It was when Eiji Toyoda, who would go on to be the president of the company, visited the Rouge Plant for Ford. Again, back in 1950. And he was amazed by what he saw about the assembly line there. And this in turn led to the further refinement and incredible results of the Toyota Production System.

So coopetition I think is actually the goal here. Our great hope is for constructive competition and also collaboration between all the car manufacturers, the IT companies, different governments, which have a lot of work to do to figure out what the right rules should be, and also hardware manufacturers like NVIDIA and others.

So I'm going to wrap up a little early here, but let me ask the question. Why should we cooperate? After all, it's not the American way. We're all about competition and think that will be best.

Well, the reason, of course, for both safety and also for traffic, since traffic's not anybody's friend, is that 1.2 million people per year demand nothing less. And so I want to make it clear that, as aggressively as we're pursuing this, we believe that safety and traffic are things that are of concern in general to society, and we want to work on them together.

Let me conclude with going back to this incredible experiment. It was one of the last things I did at DARPA. And it involved, again, a patient with a brain/computer interface. Let me set up the slide here just a little bit.

You're going to see a picture showing just the top of her head, as well as the electrode assembly that comes out of the top. You don't have to worry. It's not going to make you faint or anything like that. But pay attention on the lower-right-hand part of the picture.

This time, instead of controlling the simple test task of picking up cubes and putting them on squares, she's going to do an actually important task for mobility and other action within the home. She's going to open a door.

And the way she's going to do it is not entirely on her own because you saw how ridiculously ineffective she was just trying to move that cube. She's going to be doing it by having a guardian angel working in parallel with her, watching what she's trying to do and giving a little bit of assistance in parallel to help her do the task.

Here we go. Again, this is work that was done in collaboration with Carnegie Mellon University and the University of Pittsburgh.

Same type of arm as before — this is actually a difficult task because you have to have motion in one axis even though it's constrained in the other axis by the door itself.

Again, you can see in the lower-right-hand part her head and the electrode assembly coming out of it. And as she is successful in opening the door, she is incredibly happy. And you see her head move a little bit as she laughs.

So in conclusion, I think that parallel autonomy has tremendous promise in augmenting the series autonomy that we seem to be fixated on in this field. I think that a lot of the techniques that have been talked about at the conference here, most importantly deep learning, can be brought to bear at this problem.

And the future is incredibly bright. The work that we're doing, all of us, is incredibly important. Most importantly, I want to say thank you to NVIDIA, and thank you all very much.