Constellations

Episode 214 – COTS Infrastructure, Space-Grade Computers, and Opening New Mission Architectures

Speaker: Edward Ge, co-Founder and CEO, Aethero – 15 minutes

John Gilroy: Welcome to Constellations, the podcast from Kratos. My name is John Gilroy, and I'll be your moderator. Did you know that a graphing calculator has more processing power than most satellites? Our guest today, Ed Ge, co-founder and CEO of Aethero, is working to change all of that by building the next generation of space-grade computers for computationally heavy space missions. Ed is going to give us the details on what it takes to build powerful compute for space, how it benefits the industry, and some of the challenges still ahead. Edward, you ready to rock and roll?

Edward Ge: Absolutely.

John Gilroy: Oh, great, great, great. Let's just start off by getting your hot take on compute technology that today's satellites are currently running. What is the lay of the land?

Edward Ge: Well, right now it's mostly FPGAs and CPUs, or much older processors like at the RAD750. The RAD750, fun fact, is based on the old iMac processor from 1998. So the same processors that were powering your computer back before the whole dot com era and the whole Y2K bug, are the same processors that are powering NASA's Mars rovers like Perseverance on Mars today. Now, the main problem with a lot of these FPGAs and CPUs is that most modern AI... Like everyone talks about AI. AI this, AI that, AI in space, whatever. But most modern AI research happens on CUDA and a lot of these...

And if you're building on CUDA, that means when you want to, let's say, move the latest AI model, the latest and greatest off of an AI paper and move it into space, move it onto one of these older CPU, FPGA-based systems, that means you have to essentially redo a lot of the code, you have to reintegrate it to make sure it works and debug it all over again. And this massively increases the timeline of, let's say, when a new AI model comes out or to when you can actually have it running on your satellite in space. And by the time it's running, it's already obsolete.

John Gilroy: Wow, yeah, that's a challenge there. Usually when we talk about the edge, it's a reference to a point of processing that is as close as possible to the end user. That's why it's the edge. So when we put it on a satellite, it's literally the farthest point away from a human. So how do more advanced onboard systems transform a satellite into an edge device, and how big a difference in capability are we talking about?

Edward Ge: Absolutely massive difference in capability. Because it exists... The ways we run our satellites hasn't changed much since the 1980s. Satellites orbit the Earth, let's say every 90 minutes. Of that 90 minutes, only 15 to 20 minutes they have contact with a ground station. That means only 15, 20 minutes can you actually get data down from the satellite, look at the data, and then send commands back up. And this increases, let's say, the iteration cycle of new commands, new decisions that you can make if you're running the satellite by several orders of magnitude.

If you can, let's say, compress all of this data down, send it down instantaneously, or, let's say, somehow have the satellite process the data, look at the data, and then make the next call autonomously, this essentially, one, enables a satellite to react dynamically and in real time to changing events, let's say a piece of orbital debris coming in on a collision course, or enabling the satellite to conduct operations autonomously without needing human intervention. That saves you a lot of time, that saves you a lot of effort, and that essentially could prevent a multi-million dollar or multi-billion dollar asset from being damaged or destroyed.

John Gilroy: Yeah, that's a good observation. So what are some examples of design decisions that are accelerating innovation and market adoption in orbit compute?

Edward Ge: Moving towards COTS components is a major one. Actually I know for, example, NASA doesn't really like using COTS components that often. The movement towards COTS components is something that's really been pioneered by the new wave of space-age startups that have appeared since the 2010s. But COTS components are surprisingly resilient in space, especially in terms of the displacement damage if you're talking about shorter-term missions. And a lot of these companies, they are going for these shorter-term missions, and I do think there is a trend right now towards more satellites in space that are more disposable lifetimes, but in exchange are cheaper, and what's really also enabling this is the fall in launch prices and the greater accessibility to orbit that we've never seen before with the rise of SpaceX, the rise of Rocket Lab, and other launch providers.

John Gilroy: I never thought of that perspective. It's cheaper, so therefore you can have more disposable type satellites up there. Good. Ed, so as computing capabilities in orbit scale, where do you see the greatest demand emerging? In Earth observation, telecom, servicing, or maybe even government defense?

Edward Ge: Now that's a good question because it's actually one I've put a lot of thought into, and I have a contrarian thesis compared to most edge computing companies. When one imagines edge computing in space, the most often cited use cases, let's say you have a bunch of Earth observation satellites collecting images of the Earth, and these Earth observation satellites have to compress the data before they send it down to the ground so there's cost savings. But a lot of the times, let's say, most commercial adopters of Earth observation don't really have a preference when it comes to latency, when it comes to time latency.

But when a farmer, for example, does not care if his image of his fields arrives every week, every month, or hell, every quarter. But the Department of Defense cares down to the very last second because they have lives on the line, they have billions of dollars in assets on the line. And being able to see what enemy is doing, being able to see what the rivals are doing can essentially tip the balance between either the US or our rivals winning in the next great power conflict. And as a result I think the greatest demand actually is coming from the government and defense side of things. And any edge computing company, any company that wants to build autonomy in space that is ignoring defense does so at their own peril.

John Gilroy: Yeah, edge computing I think is important, Defense Department. So I think the defense and national security users often use early adopters of edge computing. So what kinds of national security missions or threat responses could onboard processing enable that weren't feasible before?

Edward Ge: That's a really good question. And to be honest... do you know the Strategic Defense Initiative, the Reagan Star Wars program? So people often say Golden Dome is the spiritual successor of that, and it is, because a lot of the technologies needed for, let's say, advanced missile defense for boost space interceptors, that cannot happen without onboard compute, that cannot happen without onboard autonomy. If you cannot see the missile coming at you in real time and move a satellite in position to intercept it, you do not have the ability to prevent missiles from being targeting the US homeland. And it's also worth noting that a lot of advanced use cases like ISAM, autonomous RPO, and even defense use cases like non-cooperative RPO, all depend very heavily on your satellite being able to react in real time, because it comes down to being able to see what the enemy is doing before they know what they're doing and being able to react before they even can get the data down from this satellite and make a response.

John Gilroy: So Ed, what sorts of new techniques and strategies enable these powerful processors to operate reliably for these multi-year missions?

Edward Ge: Got it, yeah, that's a good question. And honestly, it really comes down to hardware and software hardening, because a lot of companies have this idea that, "Hey, you can just coat the chip in some radiation shielding and it's going to essentially survive for long periods of time in space." But the thing is, most displacement damage is surprisingly easy to mitigate against. The difficult part comes down to the software. You need hardened software architectures that can withstand bit flips, that can withstand latch ups, that can take a punch and keep going. Because when something happens in space... If something happens to your computer on Earth, you can turn it off and turn it back on again. You can't exactly do the same thing to a satellite, and you can't really reinstall a hard drive on the satellite the same way you can do on Earth. When you send something up, it has to work perfectly that one time under any potential stresses, and that's where a mix of both hardware and software mitigation come into play. And that's a huge part of what we're doing at Aethero.

John Gilroy: So how does enabling onboard training and not just inference fundamentally change what autonomous spacecraft can do?

Edward Ge: Well, so one fun example I'd like to point to is during the war in Ukraine, the Russians figured out that the Ukrainians were using satellite imagery to spot their vehicle movements or tank movements. So what the Russians did was they started disguising their tanks as buildings. Now, if you, let's say, were running AI computer vision model on the satellite, and you couldn't retrain the model, you wouldn't pick up the new buildings disguised as the tank building amalgamation. But if you had onboard training, you could essentially feed the new updated data that you were collecting back into its onboard model and enable the satellite to essentially respond dynamically to what the enemy, to what the target, the subject of interest is doing on the ground. And really onboard training is also critical, because it means the longer the satellite is in orbit, the smarter it becomes, and nothing on Earth is static. Stuff will change, environments will change, cities will change, people will change, vehicles will change. And if you don't have the onboard training, you're essentially holding yourself... You're putting yourself at risk of obsolescence whenever... Due to terrestrial-based factors.

John Gilroy: Ed, earlier you mentioned latency. For latency-sensitive applications like a disaster response or planetary defense, what difference does in-orbit compute make versus ground-based processing? Latency there, huh?

Edward Ge: Well, the thing is ground-based processing can't do much if you can't stream the data down 24/7. Because if your satellite only has contact with the ground, say for 15 to 20 minutes, ground-based compute, that's still a massive window of every orbit where you can't get the data down, and you can't get the data down, there's nothing to run the compute on. It's also worth mentioning that, yes, we could link it to whole Earth and ground stations. But a lot of ground stations, particularly for popular spectrums like S-band and X-band are very limited in terms of geographical distribution. So even if you had, let's say, a bunch of Ka-band ground stations around the world, your satellites might not be compatible with that. And it's also worth noting that even if you stream all the data down to the ground station, there's still a ton of latency in getting the data down. The satellites are limited by transmit power, by transmit frequency and duty cycles, and ultimately your satellite... Ultimately there will be a delay of seconds or minutes or hours and getting a response time back up there.

John Gilroy: Ed, we're at the SmallSat conference in lovely downtown Salt Lake City and a lot of vendors here, and usually part of a conversation is the financial aspect of it. So I've got to go in a little bit of this with you. So at what point does compute and space become more cost-effective than repeatedly downlinking raw data like we just talked about?

Edward Ge: Well, that depends on the use case. And quite frankly, I try to look at it less from the aspect of how much money you can save and more of the aspect of what capabilities can our customers now offer to the Department of Defense that makes them far more competitive to other contracts than you've ever had before? One thing I've discovered while working with large space companies that they aren't particularly cost-sensitive. Yes, cost is a huge factor when how we select vendors, cost is a huge factor, and everyone wants to be profitable. But the thing is, if you can offer them a capability that has been never, ever been done before, and let them offer that capability to their end customer, be it the DOD, be it a commercial user on the round, they're going to go for it, because space is a very competitive environment. Look at all of the Earth observation companies out there. And you look at it and start to realize, well, if this Earth observation, if you can give one of them a huge leg up, they will take it.

John Gilroy: We mentioned the Department of Defense here earlier. It equates with regulations and the regulatory world. So I got to ask you a regulatory question. So what are the regulatory hurdles for deploying high compute modules in orbit? So what are they now and how are they evolving?

Edward Ge: Well, it's one the fields where it's so far ahead that there is not much of a regulatory environment. The FCC is tracking down quite heavily when it comes to orbital debris, and for good cause, right? No one wants Kessler syndrome to happen and knock us out of low Earth orbit for the next decade. But in terms of actually the payloads on the satellites, there isn't that much just yet in terms of regulatory overhead. And that's good. It gives us room to innovate, it gives us room to experiment and let's us figure out what works and what doesn't work before every single satellite in orbit is AI-powered.

John Gilroy: Well, we talked about the latency going to ground. Let's go in the other direction. So let's look toward beyond LEO applications, lunar service applications, cis-lunar logistics, deep space exploration, really, yeah. So how might onboard compute technologies open new mission architectures?

Edward Ge: Oh, absolutely. So for light speed, for light to go from the moon back to Earth, that's 1.5 seconds. For light to go from Mars back to Earth, that's 40 minutes. For light to go from the outer planets back to Earth, that's a couple of hours. And if you have a probe to one of these deep space destinations, you essentially are very much handicapped in what you can do in terms of mission capabilities. Because imagine, let's say, you're trying to control a robot, but all of the data you're getting is two hours old, and every command you send you that robot is you have to wait two hours for that robot to respond and actually execute that command. That severely will limit the amount of tasks you can do. It severely limits the risks you can take because you can't respond to anything in real time.

And it also severely limits the type of payloads and sensors that you can have on that robot. When it comes to lunar surface operations, it's also quite fascinating, this essentially could enable operations on like... Let's take the lunar far side, where you don't have a line of sight to Earth, where you can't access, you can't communicate that to Earth. So is communications on Mars where you have blackouts, sandstorms or blackouts due, let's say, another celestial body being between the Earth and Mars. And essentially our goals enable these capabilities, have never been done before, and eliminate the problem of communication blackouts entirely.

John Gilroy: Edward, I think you've really given our audience a real good handle on compute for space. I'd like to thank our guest, Edward Ge, co-founder and CEO of Aethero.