A Techno-Economic Analysis of Space Computing

Thermal mechanics, market sizing, launch economics, and the critical path to technical viability.

Hivemind

Executive Summary

The Opportunity

Deploying a constellation of space computing servers can potentially scale global GPU capacity by 6 – 24x, and annual revenue for HPC and datacenter companies by 2 - 4x. The primary driver is the ability to leverage abundant, free, and stable solar energy, bypassing the land and power constraints currently facing terrestrial datacenters.

Thermal and Orbital Challenges

Unlike Earth, where cooling relies on convection, space requires dissipating heat purely through radiation. This demands complex satellite configurations to manage massive, perpendicular panel systems: solar panels continuously facing the Sun for power generation, radiating panels continuously facing interstellar void to release heat.

Economic and Networking Bottlenecks

Space computing servers are economically unappealing until we achieve a 3x to 13x reduction in starship transportation costs. Furthermore, the limited bandwidth of optical links between satellites cannot match the high-speed physical interconnects of terrestrial datacenters, making the training of large AI models and complex workloads impractical with current architectures.

Key Enabling Technologies

Beyond achieving radical launch cost savings through fully reusable and heavy-payload rockets (like SpaceX Starship), several emerging technologies could fundamentally change the economics:

- **Proton-gated transistors inside GPUs** could reduce the size and weight of space servers by 80% and increase computing power 6x.
- **Heterogeneous computing capabilities** could solve the systems upgrade and maintenance question while also doubling computing power.
- **Superconducting materials** could eliminate major power and cooling requirements, though this research remains in very early stages.

The first half of this research is a direct translation from a French blog by David Louapre (Science etonnante) titled "Peut-on refroidir un data center dans l'espace ?". Starting with section 6, we expanded the analysis with our own computations of how much compute we can really pack-up in orbit, what the blockers are to make space datacenters even feasible putting aside the costs, as well as some thoughts on developing technologies that could make space datacenters a lot more attractive (with caveats). Within this, we weave in a couple comments from engineers and scientists which we found complementary to the analysis in the main post. Finally, we included a brief take on the possible impact on SpaceX and a view on the implications for the startup ecosystem.

A few days ago, SpaceX achieved the largest IPO in history. Following its recent merger with xAI at the beginning of the year, one of SpaceX's main promises to justify a bright future is to deploy data centers in orbit.
As you probably know, one of the main problems with data centers - especially those used for AI models - is their cooling. However, is being in space an advantage for this? Intuitively, we tell ourselves that space is really cold (not far from absolute zero, as we'll see). On the other hand, current cooling fundamentally relies on the use of water... and there isn't exactly an abundance of water in orbit!

So, cooling data centers in space: a stroke of genius or an absurd idea? We're going to try to shed some light on these questions.

1. Cooling a Processor on Earth

Let's start with the basics: how does it work on Earth? When a processor runs, it generates heat. In fact, that's almost all it does, since it exerts no physical action on its environment, all the energy it consumes ends up ultimately dissipating as heat.

Take a good, hefty GPU, like the ones used to run AI models. For a recent model, we're typically looking at a power output of around 1000W. The surface area of the chip itself is small, roughly on the order of a cm^2. If we think in terms of the heat flux that must be evacuated through the surface, we get the enormous figure of 10 megawatts per square meter!

To evacuate heat from a surface, the typical solution is to use convection, which means circulating a fluid (like air or water) over it. Our desktop computers, laptops and tablets use air, for instance.

A fluid's heat removal capacity depends on the fluid's nature, its velocity, and the flow characteristics; it is proportional to the temperature difference between the fluid and the surface being cooled.

For air moving at a good speed, the proportionality coefficient—known as the convection coefficient—is approximately 100 $W/(m^2 \cdot K)$. Note the unit carefully: "watts per square meter per kelvin." If 10 MW/m^2 needs to be dissipated from the chip's surface, a simple calculation shows that, in theory, a temperature difference of at least 100,000 degrees would be required between the surface and the air! Obviously, that is not feasible.

The first step to overcoming this problem is to increase the surface area in contact with the fluid. This is achieved using a heatsink - a piece of thermally conductive metal that contacts the processor surface and spreads the heat to structures offering a large surface area.

A heatsink increases the surface area in contact with the fluid

For standard processors, finned structures (like the one shown above) are used. However, professional GPUs employ surfaces featuring mini-channels, which provide a massive increase in surface area—by a factor of at least 100. Within these channels, the heat flux to be dissipated is therefore reduced to "only" about 100,000 W/m². While this is still far too high for air cooling, it becomes feasible with a liquid such as water.

When water flows through mini-channels at sufficient velocity (a few meters per second), it can dissipate 100 times more heat than air—roughly 10,000 W/(m²·K). Since we need to dissipate 100,000 W/m², a 10-degree temperature difference is enough to get the job done. Given that GPUs typically operate around 80°C, this is clearly feasible.

Of course, while this process moves heat away from the processor, the heat itself doesn't simply vanish. It ends up in the fluid and must be permanently removed from the data center. The fluid usually circulates within a closed primary loop; a heat exchanger can then be used to transfer the extracted heat to a secondary loop. This secondary loop might operate by drawing in external water and discharging it either at a higher temperature or as steam—a process known as evaporative cooling.

This is skipping a lot of details, but the bottom line is that it works on Earth: we know how to cool GPU-based data centers.

Now, let’s send our data center into space. How does it work there?

---

2. What About in Space?

As you can see, the sticking point isn't so much the initial cooling phase—moving heat away from the processor—but rather the final heat rejection. In space, there is no water to evaporate! Fortunately, we can rely on another phenomenon: radiative cooling.

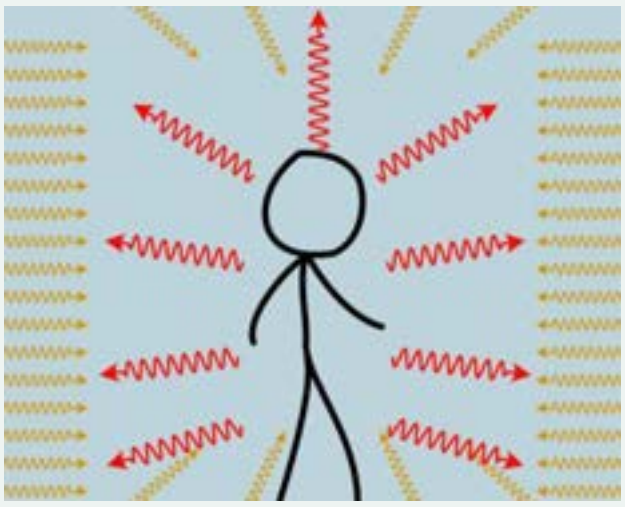
You may be aware that there are three forms of heat transfer: conduction (such as through a heatsink metal), convection (where heat is carried away by a fluid), and radiation (characterized by the emission of energy as electromagnetic waves).

Radiation is governed by the Stefan-Boltzmann law, which states that a body at temperature T radiates a heat flux equal to

$$\Phi = \sigma T4$$

where σ is the Stefan constant and the surface temperature T is raised to the fourth power. An important detail: the temperature must be expressed in kelvins. Given the T^4 dependence, you might think radiation would be incredibly effective for cooling! Yet, in our daily lives on Earth, it is rarely the primary cooling factor—and you’re about to see why.

Imagine standing in a room at 20°C. Your skin surface is at roughly 35°C (or 308 Kelvin), so it radiates energy according to the Stefan-Boltzmann law. Do the math, and it comes out to about 510 W/m². That sounds like a lot, but something is missing: the walls of the room you are in are radiating back at you. They are at 20°C, so they beam about 420 W/m² back your way. If you calculate the difference, that represents a net loss of only about 90 W/m² via radiation (and even then, we are overlooking a lot of factors — clothing, furniture and so on).



A schematic representation of radiative equilibrium in a room: you radiate your heat to the walls, but they radiate theirs back to you

The moral of the story is that cooling by radiation depends on the temperature of the objects surrounding us. So, cooling a terrestrial GPU or data center via radiation is not very efficient.

But in space, things are different. There is nothing radiating heat onto you; radiation represents a net loss. To be precise, there is the so-called "3 K" background radiation—that remnant of the Big Bang bathing the Universe in electromagnetic waves corresponding to the radiation emitted by a body at a temperature of 3 kelvins. This is why the "temperature of space" is generally considered to be 3 degrees above absolute zero, or around -270°C. So yes, space is very cold from a radiation standpoint, but on the other hand, there is no air or water at that temperature to carry heat away via convection. We can therefore rely only on radiation. Is that enough to cool a data center?

---

3. How Do We Evacuate Heat in Space?

Let’s try to work out the numbers. The first thing to realize is that the term "space-based data centers" is misleading. When we hear it, we naturally tend to picture huge, box-like buildings orbiting the Earth.

Definitely not what a space datacenter will be

But as you’ll see from our calculations, that is not how it will happen at all.

Let's look at the numbers. Instead of a complete data center, let's just consider a single large computing unit. Let's take an NVIDIA GB300 NVL 72, a large rack containing 72 GPUs. It's about the size of a closet, so we can imagine fitting it into a satellite.

GB300 NVL 72 next to Jensen Huang, Nvidia’s CEO

This unit consumes a maximum of about 150 kW, which will entirely end up as heat. This means we will constantly have to evacuate up to 150 kW from our satellite, and do so purely by radiation.

Since radiated power is proportional to the radiating surface area, how much surface do we need to do this?

Let's be optimistic and assume the evacuation surface is at 80°C (the typical operating temperature for processors). The Stefan-Boltzmann law gives us a radiated power of 880 W/m² at this temperature (assuming a perfect radiator; in practice, it will be slightly less because surfaces aren't perfectly emissive).

It will therefore take a surface area of 150,000 / 880 ≈ 170 m² to hope to evacuate this heat into space via radiation. Imagine that for this purpose we deploy panels that can radiate on both sides, and giving a bit of leeway, we will put a discount and say 100 m².

You would have to deploy 100 m² of radiant panels around our satellite to hope to evacuate the heat from this single computing unit containing 72 GPUs. It's not an absurdly crazy idea, but it's still not that simple! Especially for something we manage to do very well on Earth using convection.

From a strict cooling standpoint, being in space is far from an advantage.

---

4. Where Should We Put the Satellite?

So far, we've stayed on a very theoretical level, but there are still plenty of practical questions. For instance, how should we position our satellite and its radiating panels? A crucial point is that these panels must not receive the Sun's rays. Indeed, at Earth's distance (outside the atmosphere), the solar flux is about 1300 W/m². We must therefore avoid at all costs having this sunlight fall on our radiant panels because they would then turn into absorbing panels!

One solution would be to try and stay in the Earth's shadow, but besides the fact that it's difficult to stay there permanently, this would deprive our satellite of the solar energy it needs to operate. Yes indeed, in addition to the radiating panels, we will also need to include solar panels!

As we mentioned, our computing unit requires about 150 kW of electrical power at its peak. A good solar panel in space, kept perpendicular to the sun, can produce about 300 W/m². That leaves us needing 500 m² of solar panels to power the beast! (By comparison, the ISS has about 2500 m²). The design of our satellite is starting to get complex.

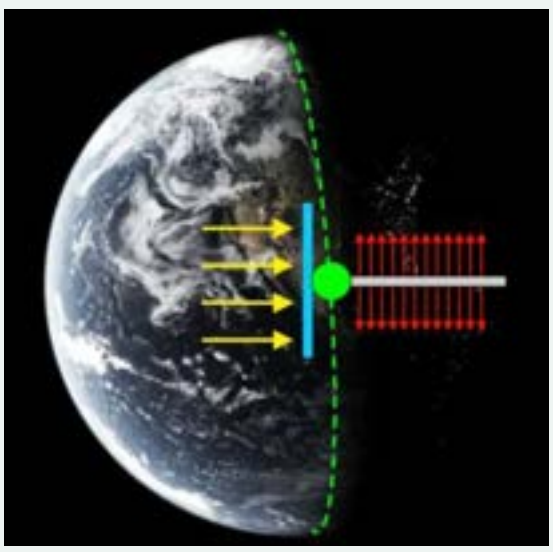
To summarize the problem: we must fit 500 m² of solar panels that must constantly face the sun, as well as at least 100 m² of radiating panels, which—on the contrary—must always face the immense void of interstellar space. Technically, it's possible, but we'll have to be a bit clever.

One solution is to put our satellite on the trajectory of what is known as the "terminator." It sounds a bit scary, but this term simply refers to the day/night dividing line on Earth.

Earth terminator line, sourced from Astronomy Stack Exchange

The terminator is a moving circle on the Earth's surface, and if a satellite manages to continuously orbit directly above the terminator, it is guaranteed to always have sun exposure for its solar panels; whereas any other orbit forces it to be periodically deprived of the Sun by the Earth's shadow. If, in addition, we place the radiating panels orthogonally, facing "up" and "down" (i.e., perpendicular to the plane of the Earth's orbit), we get exactly the configuration we need.

With this arrangement, the solar panels permanently see only the Sun, while the radiating panels see only the interstellar void at 3 Kelvin.



Non-binding illustration: the satellite orbits above the terminator, the solar panel (in blue) is oriented perpendicularly to the sun's rays. The radiating panel (in gray) dissipates its heat perpendicularly to the orbital plane.

In reality, the net balance isn't quite so simple. Indeed, our satellite will have to be in a rather low orbit (Low Earth Orbit, less than 1000 km from the surface), and therefore the radiating panels will be partially exposed to Earth's radiation!

Indeed, Earth is at about 15°C on average, so it radiates infrared heat as well. Even if well-oriented, the apparent size of the Earth is so massive in low orbit that the radiating panels will inevitably capture a portion of it, which will complicate their thermal balance even further.

---

5. Is All of This Really Worth It?

If we put it all together, we can see that it's technically feasible, and quite complex. Just the energy balance alone is barely favorable, and that is after many simplifications regarding the details of moving the heat to the panels (which will require a fluid, a pump to power...). So even being optimistic, this is shaping up to be quite an engineering headache.

What we see, as explained earlier, is that the image of a data center as a big hangar in space is misleading. To achieve computing capacity analogous to a terrestrial data center, we are talking instead about a constellation of satellites (each a Space Server), each carrying a few dozen to a few hundred GPUs. This is in essence the operating architecture that Elon Musk has in mind, and likely for the very reasons we have just highlighted.

Further in section 7, we discuss other technical challenges to make that constellation of Space Servers work, among which maintenance (impossible today, and there are various space debris orbiting around that will inevitably strike and damage the panels.), recycling (impossible today), communication between computing units (can’t do fiber optics).

Of course, the real major reason for moving into space is to benefit from abundant, free, and stable solar energy. At 150 kW of power, the same computing unit on Earth would consume over 1 million kWh every year, resulting in a bill (at current prices) of about $150,000. Is this gain enough to justify the effort?

However, let’s not forget that we must also add the fact that these satellites will have to be launched, a cost that Earth Datacenters do not face. SpaceX's Falcon 9 launches and puts satellites into orbit for about $3,000/kg. If our satellite weighs a few tons (I estimated ~6-8 metric tons based on the typical weight of radiant panels, solar panels, and of a GB300 NVL 72), today that seems economically prohibitive: $18 million per Space Server and we would need at least several hundreds of thousands of such Space Servers (more on that in the next section) to deliver a compelling Space Datacenter.

We can see that even considering the talent of SpaceX's engineering teams and the vast capital markets that firms can tap into, the techno-economic equation is currently quite hard to justify: a Space Datacenter is both harder to manufacture and harder to transport and service than an Earth Datacenter.

Can the size of the opportunity justify a staggering spending to make Space Datacenters technologically feasible and economically viable?

---

6. Sizing the Space Computing Market

Let's assume that the dimensions estimated above for a multi-satellites datacenter are within the right ballpark: one GB300 NVL 72 in space (a Space Server) needs 500m2 of solar panels to power it and 100m2 of radiant panels to cool it down, with these two groups of panels essentially on top of each other or at a 90 degree angle, with a large enough gap left between them.

Let's assume a compact design, aka a sandwich with top layer being the solar panels, middle layer being the server core, and bottom layer being the radiant panels. With such design, the projected surface area effectively occupied by the Space Server is that of its solar panels. Keeping a rectangular shape for these solar panels like that of the ISS, one gets a shape of 40m x 13m (note that if we assume a T-bone design, the answer remains the same).

At a typical Low Earth Orbit of ~400km, these satellites can be spread across a perimeter of ~42,000 km. If we were to perfectly align satellites one after the other without leaving any distance between them, we would be able to put ~3 million Space Servers on Low Earth Orbit. However, these satellites need to move through the day to maintain their perfect exposure to sun on one side and night on the other side so some distance must be kept between them. Let's assume that 3-10m, a third of the width of a given satellite to a full satellite width, is a reasonable approximation. 600k-2 million Space Servers on Low Earth Orbit appears to be a more realistic estimate of what is feasible. Since each server here is assumed to have 72 GPUs, that means 43-145million GPUs can be put into low orbit into Space Datacenters. That is about 6 - 24x the current annual volume of datacenter GPUs sold by Nvidia, AMD and the likes.

6 - 24x is quite an expansion. In dollar terms, that corresponds to $1.3 trillion to $4.3 trillion in hardware revenue. Taking a server life of 6 years results in Space Datacenters doubling to quadrupling of the current high-performance computing hardware market. The corresponding datacenter business would at least double and up to triple the current global datacenter market, if we assume that the relation between datacenter costs and datacenter revenues does not change.

However, with today's costs of sending satellites to space, that would imply at least $10 trillion in transportation costs alone, or 10 to 20 years of gross revenue for this industry sector.

Per section 5, we would need to see a 3 - 13x compression in transportation costs for the financials on the Space version to reach cost parity with the Earth version. A 13x compression would bring transportation costs to ~$200/kg. SpaceX starship has an aspirational target of $67-100/kg, essentially driven by making the vehicle fully reusable and increasing its size and payload capacity, and we can now easily understand where that target comes from.

> If we assume the company aims for a 50% to 70% gross margin (based on their IPO roadshow deck), that gets us at $200/kg if SpaceX transports Space Servers on behalf of other clients

> If SpaceX transports its own Space Servers, which makes more sense as a corporate strategy, then it gives them a 2 – 3x buffer on their aspirational starship targets to still achieve an economically viable Space Datacenter, and it would still achieve a 94% cost reduction from where we are today

> - At $200/kg, the cost of sending all these Space Servers to space drops to 1 to 2 years of gross revenue for this industry sector (~$800 billions), which would be much more likely to get financed through both debt and equity capital markets (for comparison, that would represent 4x Amazon, 8x Oracle, 20x Meta's outstanding corporate debt as of Q1 2026 based on the EDGAR filings)

So, if Space Computing transportation costs were drastically compressed, and they at least are trending in the right direction, then we would be contemplating one of the largest market opportunities as well as one of the coolest engineering challenges in history.

What are the technical blockers getting in the way?

What technologies could make it even more economically compelling?

---

7. What Technology Challenges Need to Be Solved for Space Datacenters to Be Achievable?

All the following are essentially non-starters, regardless of the economics of setting up a datacenter in space. Some of these we know how to solve; some remain at a research stage.

Level 1 Challenges: We know how to solve

Cosmic radiation requires the use of shielded GPUs, memory and telecom equipment.

- Shield everything, it costs somewhat more but also further weighs down the rocket's payload.

- Of broader note: anything that increases the weight of a Server in space relative to on earth ends up requiring further transportation costs' reduction. For instance, with SpaceX starship, if they indeed achieve their target of $67-100/kg, then as long as the additional weight of a Space Server to make it space-suitable less than double the server weight, it would be fine.

Level 2 Challenges: We have reasonable-to-good ideas on how to solve

Hardware maintenance in space to repair damage from micrometeorites and debris, and to replace faulty parts.

- Industrial robots seem like the natural answer, but these specific robots do not exist yet. There are hurdles across hardware (fine motor skills, zero-G low-K designs), middleware (governance, orchestration), and software (training data, maintenance) to name a few.

- Of broader note that if we know how to do that, then we know how to efficiently upgrade Space Servers. However, upgrades come with other challenges, as we discuss further down.

Keeping latency within enterprise "standards"

- Altitude is not the problem: at 600 km, the round trip to the satellite only costs a few milliseconds. The problem is geometry. A satellite in low orbit speeds along at ~7.5 km/s and only flies over a given point for a few minutes per pass: no fixed site has continuous coverage. On a "terminator" orbit (sun-synchronous dawn-dusk), the ground track stays glued to the day/night border, so we cannot park the Space Servers over a population basin.

- To serve a market continuously, we need an entire constellation to be in place. Specifically, we need enough Space Servers such that any point on the globe has constant access to at least one Space Server.

    - From a given location on Earth, one typically gets a visibility cone of ~16 degrees, which yields an orbital arc length of ~3850km over which one server covers one location on earth. 12 Space Servers would thus strictly be able to cover a given location continuously.

    - As a task gets handed off from one server exiting the visibility cone to the next server entering the visibility cone, the latency incurred by that hand-off needs to stay below ~2ms (inter-GPU & inter-rack communication). This implies the two servers need to be within 300km of one another. 133 Space Servers would thus strictly be able to serve a given location continuously.

- Thus, a minimum viable product for a Space Datacenter would have ~10k GPUs, which is a legit hyperscale datacenter, and a ~$2.5B transportation bill. As we discuss in the next two bullet points, there are steep challenges for such MVP to serve advanced workloads to clients.

Training very large models

- For large models whose parameters cannot fit on a single GPU, the computation must be distributed across several GPUs using tensor parallelism. This technique, essential for training very large models, requires GPUs to exchange parameters almost instantaneously between each layer of the network. Laser latency prohibits this approach. Models would have to be partitioned in a suboptimal way, which would saturate the lasers' bandwidth during parameter updates.

- Gradient descent training of a large neural network in orbit with distributed partitioning and parameter exchanges over laser links is not just an engineering challenge; it seems to contradict the mathematical prerequisites of computation synchronization for training today's large AI models.

- We could solve that by focusing Space Datacenters on inference and training models capable of fitting within a single GPU (or ASIC, FPGA). Earth Datacenters aren’t going away.

- We could also solve that by moving to a new category of algorithm architectures that are less parameter hungry or more memory sparse, which is an active field of academic research notably at top US universities.

Level 3 Challenges: We don’t yet have good ideas on how to solve

Delivering total bandwidth for a Space Datacenter on par with its Earth counterpart

- In pure latency, a vacuum such as in space is faster than fiber. The problem is therefore not the speed of the link but reconstituting the total bandwidth of a terrestrial cluster in orbit.

- A large data center is hundreds or even thousands of racks connected by an internal network with an aggregate bandwidth counted in terabytes per second. It's this interconnection density that allows serving the largest models at a very large scale and high throughput.

- In orbit, you would either have to stack hundreds of racks on a single platform (a nightmare for energy, cooling, and transportation), assemble on-orbit (a steep but a-priori not impossible robotic challenge), or link several satellites via free-space optical links, which peak at a few hundred Gb/s, whereas the NVLink of a single cabinet reaches tens of TB/s.

Upgrading hardware

- Today, a GB300 rack draws ~130 to 140 kW. The Rubin Ultra ("Kyber" rack NVL576) arrives in 2027 with ~600 kW announced by NVIDIA; and the next generation (Feynman, ~2028) is aiming for a megawatt per rack. Each of these new generations would require increasing the solar panels and the radiant panels by 5x and 10x respectively from our numbers above. It may not be mechanically feasible to put all this added surface lengthwise for each Space Server. Instead, we may have to increase their width, which will reduce the number of servers. At the limit, that will cause a stagnation in Space Datacenter compute

- The cadence of setting up Space Servers is much slower than the cadence of rack upgrades, so we either must assume that:

  - Rack upgrades will drastically slow down by the time Space Servers are economically viable, or

  - Space Servers will be able to mix and match different types of racks and still work together well. This is another active field of academic research notably at top US universities.

8. How Do We Leverage Technology to Make Space Expansion More Economically Compelling and Faster?

Level 1 Ideas: Seemingly low-hanging fruit that will not work

[NO] Leveraging outer space's temperature to cool the Space DC, thereby not having to rely on radiant panels, and saving a lot of weight

- Cold doesn't get transported, only heat does, and it will always have to be dissipated by radiation on the "space side". There is no cold in space, since there is no matter. The vacuum is not "cold", the gas is, and it is too sparsely dense to be of interest at altitude, and absent in space.

[NO] Leveraging Peltier modules for cooling, to greatly reduce the size of the radiant panels to be used. Leveraging small nuclear reactors for energy, instead of solar panels, to reduce the weight of the system as well.

- The Peltier effect only moves heat around and it does it with very low efficiency

- Reactors also need cooling. So, we could reduce the surface of solar panels, but then we would need to increase the surface of the radiant panels, and these are much heavier than solar panels

[MAYBE BUT] Recycling the heat generated by the server to power it and thus reduce the size of the solar panels

- Using the produced heat and transforming it into usable electricity would save on solar panels, but the weight of the corresponding additional equipment is another drag. The gains are not large enough to justify it (5% - 16% on an Earth system with what is currently considered best-in-class solar-boosted organic Rankine cycle)

Level 2 Ideas: The tech exists, but the products don't exist yet

[YES BUT] Updating transistor technology inside the GPUs to proton-gated rather than atom-gated. This would result in a substantial reduction in energy consumption and thus heat produced, as well as in smaller form factors and thus lower weight and higher attainable size of Space Datacenters.

- There already are prototypes for cards and slots demonstrating 90% compression in form factor and energy consumption while remaining compatible with existing AI libraries
- With enough capital injected, we could have a line of sight to ~ 5 years for a meaningful production deployment
- This is not a space-specific solution; Earth data centers would also benefit from this. Chip producers, DC operators, and hyperscalers would opt to update the hardware stack in their Earth DCs first, as it is operationally much easier.

[YES BUT] Enabling generalized heterogeneous computing with substantial kernel optimization so that we can do a lot more with existing cards and simplify the operational overhead of maintaining and upgrading constellations of satellites. This would result in better economics for Space Datacenters, an overall higher quality product, and faster and higher availability

- There already are working prototypes for significant kernel optimization pushing the average utilization rate for cards to 60% instead of 30%. There are early research efforts suggesting we may be able to reach an average of 80% utilization rate
- There are early working prototypes for restricted generalized heterogeneous computing, allowing to mix a few types of GPUs from the same chip designer

Much like upgrading transistor technology, as mentioned above, heterogeneous computing benefits both Earth and Space datacenters and will be deployed on Earth first because it is operationally easier.

Level 3 Ideas: The tech doesn’t exist yet

[MAYBE] Designing a computing unit solution based on superconducting materials. Theoretically, space could provide the low temperature necessary to drop below the critical temperature, and thus, would limit power and cooling needs.

- While this could be the most promising technology development for space datacenters, the research is still very early. Thus,I think it is fair to say that we currently have no line of sight to a timeline.

Conclusion: Where does this leave us?

- At the pace at which Earth DCs are consuming energy and land, we need a paradigm shift.
- The shift can come from either moving to Space and/or making revolutionary advances in hardware, algorithms, robotics and materials
- There is a lot of research that needs to succeed to enable these paradigm shifts – and much of this research is going to take place at the best labs and startups
- SpaceX has its work cut out, and markets may be more patient with them than anyone would reasonably expect

References & Footnotes

Peut-on refroidir un data center dans l'espace ?

Orbital Data Centers: Spacecraft Constraints and Economic Viability

Space Launch Economics Analysis | SpaceNexus

Rice researchers turn wasted data center heat into clean power | Rice News | News and Media Relations | Rice University

SpaceX_IPO_Roadshow_Final.pdf

The Only AI Moat is Hardware. And Compute is the Upper Bound for... | by Murat Onen | Medium

CMOS-Compatible Protonic Resistive Devices | MIT Technology Licensing Office