The Panel Is the Computer

Project Vela · Paper v1 · November 2026

Infinite Sunshine. Effortless Intelligence.

www.project-vela.com/paper/

Summary. AI is limited less by ideas than by power: every large data center now waits on the grid. Project Vela puts the computer where the power is. Our satellites carry solar panels that are themselves the computer: each part of the panel makes its own power and sheds its own heat. Our target for the third generation is about **5 billion tokens per second from one 100-tonne Starship launch** on a 7B-class language model. That's roughly the delivered work of **100,000 H100 GPUs**, from about 6 MW of solar power instead of 100+ MW of grid. Our first generation, built from off-the-shelf parts, already measures **on par with an H100 per watt.** This paper sets out the benchmark, the arithmetic, why we believe the targets are reachable, and what still has to be proven.

1.1 The problem: AI runs on power

A 100,000-GPU data center draws well over 100 megawatts once servers and cooling are counted. New sites wait years for grid connections, and they need land and water for cooling. Chips are no longer the scarcest input to AI. Power and permission are.

1.2 The idea: put the computer where the power is

Above the atmosphere, sunlight delivers about 1,361 watts per square metre, and in the right orbit it shines nearly all the time. Most proposals for space data centers launch ordinary servers plus large radiators to shed their heat. We take a different approach: **the solar panel is the computer.** Every part of the panel makes its own power and sheds its own heat, so capacity grows by adding panels, not radiators.

The basic unit is a **compute panel** of about 2.4 × 4.35 m, delivering **~2.9 kW of compute** (nameplate: new cells at standard test temperature). It's made of tiles that each carry their own solar cell and compute. Spare tiles are built in, so a failed tile is switched out, not repaired. A satellite carries many panels.

1.3 Our benchmark: tokens per solar watt

In orbit every system gets the same sunlight, so we measure what counts: **how much useful AI work each watt of sunlight delivers.** We quote it in tokens per second, per watt the solar cells deliver, for an 8B-class language model. We then convert it to the GPU everyone knows.

Generation	Status	What changes	Tokens per solar watt	vs. an NVIDIA H100, per watt
V1	**Building now.** Measured on the bench	Off-the-shelf low-power AI chips on a solar panel	14–18	**On par** (0.7–0.9×)
V2	**In development** with chip suppliers	A custom chip designed for this one job	200–370 (target)	**~10–18×**
V3	**In design**	The panel itself becomes the computer	650–1,300 (target)	**~30–65×**

How we convert to H100s. We use NVIDIA's published MLPerf Inference results, not peak specifications. Eight H100s serving Llama 2 70B deliver about 2.1–2.4 tokens per second per watt of server power. Scaled to an 8B-class model, that's roughly 20 tokens per watt. V1's 14–18 is 0.7–0.9× of that, which we call on par. We compare our solar watts with the GPU's server watts and leave out the ground data center's cooling, which favors the GPU.

1.4 What one launch is designed to do

- **~5 billion tokens per second** on a 7B-class model, in sunlight
- **≈ 100,000 H100 GPUs** of delivered AI work
- **~6 MW of solar power**, where the same work on the ground draws 100+ MW of grid power
- **80 satellites** per 100-tonne Starship launch
- **Mature chip processes:** no EUV, no photonics, no new materials

The arithmetic, so you can check it:

Step	Value	Basis
Panel area per launch	~25,000 m²	80 satellites × ~314 m² each
Solar power per m²	~240 W	Silicon space cells at ~17.6% mid-life (published space-agency data), × 1,361 W/m²
Solar power per launch	**~6 MW**	25,000 m² × 240 W/m²
Tokens per solar watt	~850	Inside the V3 band (650–1,300)
Tokens per second per launch	**~5 billion**	6 MW × ~850

Measured our way (on the MLPerf basis), 5 billion tokens per second is closer to 200,000 H100s. A simpler estimate from peak GPU math, at typical utilization, gives 75,000–125,000. **We quote 100,000 because it holds under both methods.**

These are third-generation design targets, quoted per 100-tonne launch and in sunlight. Nothing at this scale has been built or flown yet.

1.5 Why we believe it: no new physics, no new factories

Building it is hard engineering work, but nothing in the plan is waiting on an invention.

1. In orbit, power is the free resource. On the ground, AI is limited by how much grid power a site can get. In orbit, the limit is how much panel you launch. Silicon solar cells are among the most mass-produced things on Earth: the world makes hundreds of gigawatts of them a year. One launch needs about a thousandth of a percent of that.

2. Moving data costs more energy than the math. Most of the energy in an AI answer goes into moving numbers between memory and the processor, not into the arithmetic. The shorter that trip, the less energy each token costs, and we design the whole system around keeping data close to where it's used. The physical floor on energy per operation sits about 100,000 times below where chips work today. The gains we're after come from engineering, not from beating a physical limit.

3. The panel cools itself. A panel in space sheds heat from both faces. When the compute is spread across the same area that collects the sunlight, the panel settles at about 60 °C, well inside what ordinary chips are rated for. There are no pumps, no coolant loops and no separate radiators.

4. It's built on mature chip processes. Our chips use manufacturing processes that have been in volume production for over a decade. They don't need the newest, scarcest factories (EUV lithography), photonics or new materials. The chips for one launch are a fraction of one month's output from a single large factory. The manufacturing steps our later generations need are already in high-volume production in other industries.

1.6 Measured so far

Measured on development hardware, not modeled. November 2026. - **A complete open-source language model** runs with every feed-forward layer on our low-power accelerators, **within 1% of full-precision accuracy** on a standard benchmark. - **Larger models lose even less:** 1.5B to 7B models lose 0.1–0.4% at our precision (software simulation; silicon runs next). - **Serving many requests together changes nothing:** answers are bit-for-bit identical, however the work is batched. - **Under ~1 W per accelerator chip** on language-model layers (an upper bound: below our meter's floor).

1.7 The path: three generations, three programs at once

We're running all three generations in parallel rather than one after another. That's how we intend to move fast. - **V1 proves the system:** real models, real accuracy, real power, built from parts anyone can buy. It already matches the world's leading AI chip per watt, running on sunlight. - **V2 removes what an off-the-shelf chip carries but doesn't need.** A general-purpose part is built for many jobs. A chip built for this one job does it with far less energy. - **V3 rethinks the panel itself.** Every step between sunlight and answer costs energy and weight. V3 is designed to remove most of them.

1.8 Even if we're wrong by 100×

Our first generation, built from today's off-the-shelf parts, already works out to **tens of millions of tokens per second per launch** on paper. If every V2 and V3 gain fell through, a launch of Vela panels would still be a large AI data center that needs no grid, no water and no land.

1.9 What still has to be proven

We'd rather say this plainly: - **Custom-chip efficiency on real language models.** Datasheet figures overstate what chips deliver. We count only what we measure. - **Building the third-generation panel at high yield.** Each manufacturing step exists in volume today; combining them at panel scale hasn't been done yet. - **Solar cell output over years in orbit, at operating temperature.** We plan with conservative, published space-agency figures. - **Launch at full rated payload.** Our figures are per 100-tonne launch, Starship's rated capacity. Flights so far have carried less. - **Ground links and customers** at this scale.

Each item has a test or a study behind it.

1.10 Who we are

Project Vela is developed by a service-disabled veteran-owned small business, SBA-certified, founded by a former U.S. Navy radar technician. We work with chip suppliers, research partners and AI agents as an engineering team, and we publish measured results as we get them.

Want the engineering? We share architecture, test data and the full physics and manufacturing case with researchers, partners and investors under NDA: **info@project-vela.com**. Follow progress at **@ProjectVela** on X.

1.11 Notes

- *Forward-looking figures.* V2 and V3 figures are design targets, not measurements. V1 figures and §1.6 are measured on development hardware. Nothing described here has flown.
- *H100 basis:* MLPerf Inference results for 8 × H100 on Llama 2 70B, at server power, scaled to 8B class.
- *Solar basis:* AM0 sunlight of 1,361 W/m². Silicon space-cell efficiency from published CNES/CEA results (2025).
- *Energy floor:* the thermodynamic minimum energy to erase one bit (Landauer limit) at panel temperature.

Project Vela. Infinite Sunshine. Effortless Intelligence. Our mission is to make useful intelligence accessible to everyone through practical engineering.