Morgan Stanley | RESEARCH

IDEA

October 5, 2026 04:01 AM GMT

Global Thematics & US Semiconductors | North America

What Does the Power Shortfall Mean for Semis?

We address key investor feedback from our power call: 1) impact to semis, 2) who we think can get power, and 3) potential upside drivers to our shortfall call. NVDA commentary confirms the power shortfall is real, but we remain convicted NVDA has line-of-sight into chip placement.

AlphaWise α

Stephen C Byrd
Equity Analyst
Stephen.Byrd@morganstanley.com +1 212 761-3865

Joseph Moore
Equity Analyst
Joseph.Moore@morganstanley.com +1 212 761-7516

Brian Nowak, CFA
Equity Analyst
Brian.Nowak@morganstanley.com +1 212 761-3365

Sabrina E Hao
Equity Analyst
Sabrina.Hao@morganstanley.com +1 212 761-3315

Mason Wayne
Research Associate
Mason.Wayne@morganstanley.com +1 212 761-6012

Gregory Gao
Research Associate
Greg.Gao@morganstanley.com +1 212 296-3125

Nikhil Javeri
Research Associate
Nikhil.Javeri1@morganstanley.com +1 212 761-3742

David Arcaro, CFA
Equity Analyst
David.Arcaro@morganstanley.com +1 212 761-1817

Jeremy Herring
Research Associate
Jeremy.Herring@morganstanley.com +1 212 761-4182

Key Takeaways

- Results of our recent deep dive into Powering AI: US data center developers face a >30% power shortfall through 2028.
- The power shortfall is real and affecting semis, but are convicted NVDA/AVGO can make 2027 numbers. Areas of concern: ASICs, memory, optics, other components.
- Capex is becoming more concentrated, with the hyperscaler/neocloud builds set to reach 60-95% of total global DC capacity additions from '26-28.
- We believe large, credit-worthy players will be able to energize (more bullish on hyperscalers), while less credit-worthy or int'l players may face issues.
- What could drive power upside? Faster BTM builds (turbines/fuel cells). Int'l expansion is happening but there's not enough capacity to absorb the US shortfall.

Tech Diffusion

A Morgan Stanley Research Key Theme of 2026

We address the three most pressing investor questions on the power shortfall: 1) impact to semis, 2) who we think can get power, and 3) potential upside drivers that could close the shortfall. Last week, we published Powering AI: Racking Up the Power Shortfall where we sized the US data center power shortfall at 57 GW from 2026-28. We forecast this is 32 GW net of time to power solutions, which translates to a 34% net US power shortfall through 2028. Nvidia confirmed our thesis that there is not enough power to meet demand, and there are more semiconductors being built than there are places to put them. To be clear, in our view, this is not impacting Nvidia and Broadcom in 2027 (*more details below*), but we believe this could affect downstream components, like memory, optical, and other components like power management and analog parts in a bear case. We see a risk that ASICs will lose share in a power-constrained environment relative to GPUs,

investors should be aware that th

Research as only a single factor i
decision.

**For analyst certification and oth
refer to the Disclosure Section, l
report.**



given the lower token production per watt for ASICs. Beyond 2027, as the US land, power, shell (LPS) shortfall worsens (driven by a combination of People, Power and Politics - see recent reports here and here), we see heightened risk across the entire chip ecosystem in terms of securing the necessary data center volumes to place all chip volumes. This shortfall will likely increase the urgency, and premia paid, to secure **(A)** time to power solutions in the US (such as Behind the Meter, or BTM, generation, as well as repurposing the grid access owned by former Bitcoin mining companies - see ways to play this dynamic here) and **(B)** data center growth outside of the US. Regarding this latter dynamic, we see pockets of excellent data center growth, in locations including Korea, Singapore, India, Australia, China, and Europe.

We have fielded a large volume of questions from investors on how our power shortfall thesis flows through to customers and the supply chain. Below, we address the top three debates right now:

1. What does the power shortfall mean for semis and the rest of the AI value chain?
2. Who is able to get access to power?
3. Is there a path for upside drivers, like international expansion and more BTM coming online, to help narrow the power shortfall more quickly?

1. What does this mean for semis and the rest of the AI value chain?

We remain convicted in Nvidia and Broadcom's ability to meet their 12 month forward outlooks, given 1) conservative management guidance takes into account possible Land-Powered Shell (LPS shortage, 2) the visibility management has into chip placement, 3) geographic expansion outside the US, and 4) NVDA's higher tokens per gigawatt (GW) vs. peers creates a competitive advantage in a world where land, power, and shell is constrained. We are seeing this in real time, where the biggest users of ASICs, which have lower cost per GW, face heightened risk of inadequate Land-Powered Shell access. To some degree, we may have already seen this at Broadcom this year – after guiding for FY27 AI revenue well above $100bn in March, the company's most recent guidance from September was for $115bn, which is much less upside than investors initially anticipated. We think this puts some pressure on AMD as well, but the warrant dynamics with their largest customers do significantly de-risk their estimates in our view.

Nvidia and Broadcom have given 12 month outlooks for ~$800bn of combined AI revenue. We don't think that's at risk due to power limitations today, but it's likely to make upside more difficult. Nvidia and Broadcom have a combined market share in XPUs of ~90%, and given their scale set the tone for much of the AI semis ecosystem. At this point, we do not see power bottlenecks putting 2027 forecasts at risk given the high degree of coordination and planing required to plan GW scale projects. A fully assembled GB200 NVL72 weighs over 3,000lbs and has 600,000 parts, so the level of coordination required between the data centers, semis suppliers, and power supply chain is quite a bit higher than traditional server racks. Data center customers are doing everything they can to avoid the opportunity cost of having assets idle when they could be revenue generating. We think the

management teams at Nvidia and Broadcom have a clear line of sight with respect to both the upstream and downstream inputs to their demand. Recent commentary from both CEOs around their respective earnings illustrates the point.

- *"we want to be sure... It's a judged number based on not just supply of memory wafers, but supply and availability and readiness of power sites"* - Broadcom CEO Hock Tan: Sept 8, 2026
- *"We work with power generator companies...We work with land, power and shell companies all around the world. And that helps prepare all of this computing that's going to be built.... And so, we just have a lot greater visibility now upstream and downstream"* - Nvidia CEO Jensen Huang Aug 26, 2026
- *"we're tracking every single gigawatt of land power shell around the world, literally everything on the planet."* - Nvidia CEO Jensen Huang Sept 10, 2026

We are also not just taking management's word at face value, and continue to hear from our industry contacts that securing LPS is the #1 priority for compute and hyperscaler management teams, and we do see some companies finding more success than others. Timing is also a factor - semiconductor companies recognize revenue when chips go to into the server assembly which can take several weeks to convert into complete systems, which can then take several weeks to be shipped, delivered, and installed.

Where would we be more concerned within semis if we do see a slowdown? If there's chip capacity expected to come online that cannot be deployed at data centers, we could see customers push out delivery or start canceling orders. There could be some bias to hold inventory of GPUs that can't be powered up to avoid losing positions in the queue, but given the annual cadence of new chip releases and limited power, the opportunity cost of using a less performant chip is high and companies should have the lead times to reroute orders to customers who are ready to take delivery. But in the event of a pause, we're likely to see the most disruption at the tail of the bullwhip, and where shortage driven pricing dynamics are supporting revenue and EPS. To us, that looks like memory, optics, and other small dollar content components that are most likely to be over-ordered such as power management and analog parts.

The power risk drives our preference for Nvidia, who has the broadest geographic and customer footprint and whose GPUs produce the greater tokens per watt of power access. Half of Nvidia's business is outside of the traditional hyperscalers, with companies like Neoclouds, sovereigns, and enterprises where their desire for compute fungibility is much higher than hyperscalers, and where they benefit the most from the scale of Nvidia's ecosystem support and full stack offering. Nvidia has 80 cloud partners listed on their website, of which 55 are outside the US (although they may have datacenters in multiple regions).

2. Who is able to get access to power, and who is at risk of being power-constrained?

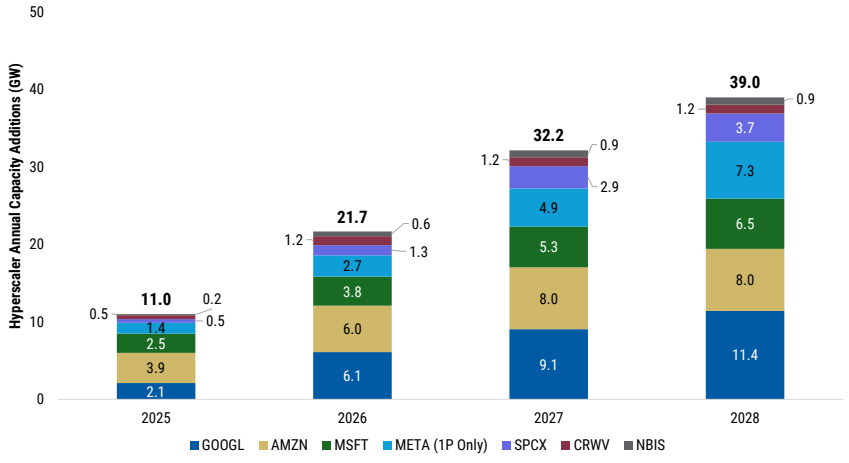
We believe we're in an environment where data center energization is being concentrated across a handful of large, credit-worthy players. Feedback we have

received is that companies with credit support, revenue visibility and large volume needs – like the hyperscalers – are getting access to power more quickly, leaving smaller, less credit-worthy, or international players at risk. Categories of companies potentially at risk of inadequate US power access include smaller neoclouds lacking strong credit backing, non-US players, and chips that have a lower production of tokens per watt of power access (such as ASICs). Our Internet team led by Brian Nowak forecasts capacity additions from the hyperscalers and neoclouds (AMZN, GOOGL, MSFT, META, SPCX, CRWV, NBIS) at 60-95% of our forecasted global data center capacity coming online, after taking into account the US power shortfall (Exhibit 1; Exhibit 2). We believe this could potentially "squeeze out" or put pressure on other players over time. We would also note that **1)** we're hearing Amazon is well-positioned to get access to power and **2)** GOOGL is adding the most incremental capacity in '27/'28 and thus could mathematically face more challenges in securing power for the compute they are buying.

Exhibit 1: We forecast the hyperscalers and neoclouds representing ~60-95% of the global data center capacity we expect to come online each year.

Global Data Center Power Demand				
Year	**2026**	**2027**	**2028**	**2029**
Incremental Global DC Demand (GW)	32	56	86	124
(-) US Power Shortfall	(9)	(5)	(18)	(35)
Global DC Energized (GW)	**23**	**51**	**68**	**89**
Hyperscaler/Neocloud Capacity Additions (MSe)	22	32	39	
Hyperscaler/Neocloud % Global Capacity Additions	***94%***	***64%***	***57%***	
*Includes AMZN, GOOGL, MSFT, META, SPCX, CRWV, NBIS				

Exhibit 2: Our Internet team forecasts the hyperscalers/neoclouds bringing on 32 GW in '27 and 39 GW in '28, with Google leading in terms of incremental capacity brought online.



3. Is there a path for upside drivers, like international expansion and more BTM generation, to help narrow the power shortfall more quickly?

While more data center capacity is shifting abroad, we don't believe there's enough capacity in ROW to absorb the US's net power shortfall (32 GW from '26-28). Companies are increasingly looking for capacity abroad, with Nvidia recently

noting that Australia is a region "with excess energy" where they stood up 2 GW for 2027. Given these latest data points, we move our US share of global compute to 55% from 60% previously (*more details below*). While we expect international expansion to continue, we don't see it as a solution to the US power problem. Conversations with our European data center analyst Emmet Kelly indicates limited room for additional spillover in Europe, given power constraints and the fact that planning permission is slow (see more here). While there could be potential for expansion into the Middle East, we believe this is difficult given geopolitical instability (see example of AWS data center being hit by Iran strikes). Our Asia data center team's forecast (led by Mayank Maheshwari) bakes in 14 GW of US hyperscaler demand spilling over into Asia through 2030 – and we still forecast a US power shortfall in spite of that – so any incremental spillover would require tapping into energy sources like coal. While there is a willingness to look abroad to build data centers, we don't believe there's enough capacity available to absorb the US power shortfall.

We do see potential upside to our forecast from behind the meter generation... Net of our "time to power" solutions (BTM Turbines, Bloom Energy Fuel Cells, Nuclear capacity, and PSP Site Conversion - details here), we project that US data center developers will experience a 32 GW power shortfall through 2028, equal to 34% of power demanded. In these scenarios, we forecast 11-49 GW of BTM turbines/ engines coming online, with 19 GW in our base case. *How do we get there?* 11 GW represents the firm, contracted BTM deals we see today, while 49 GW represents total possible production given factory capacity (keep in mind that labor availability, time required for complex engineering work, and permitting challenges limit this growth potential significantly). Our base case nets out announced projects; rather displaying *additional, behind the meter* deals signed through 2028. Our Utilities team, led by Dave Arcaro, can track 25-30 GW of manufacturer orders from 2026-28 and believes power developers can begin moving closer to adding 15 GW or more of BTM capacity per year beginning in 2028, narrowing the gap between OEM supply and annual operational capacity. Between our bull and base case, if 30 GW of BTM power successfully comes online from 2026-28, that implies a 21 GW shortfall.

In addition, there could be potential upside to Bloom Energy fuel cells – we currently forecast 6GW of fuels cells from 2026-28, but there could be upside to this, given recent reports of site expansion (see team's thoughts here). Bloom Energy is a key beneficiary of this "time to power" shortfall, in our view. In our view, the urgency to secure power being expressed by AI players is much greater than appreciated. For any "time to power" provider that has a viable path to providing data center developers with power in 2027 and 2028, we believe such providers will be able to enter into lucrative agreements.

...however, People (labor availability) and Politics (growing pushback against data centers, and AI more broadly) are key driving forces in the opposite direction. Even if we can increase capacity of behind the meter generation coming online, people are still a bottleneck. CSIS estimates that by 2030 the US requires 140K+ incremental skilled trades people beyond current capacity in a scenario like the Second Industrial Revolution. The existing labor base would likely support 10-20 GW of annual gas additions, considering both large power plants and onsite

projects. However, companies are heavily investing in their labor forces to grow these numbers, such that it could be possible to reach 50+ GW of annual gas additions by the end of the decade. We've also written extensively about the impact politics, moratoriums, and public perception are having on the build out (see more here, here).

Our US power shortfall remains at 57GW, and moves to 32 GW net of time to power solutions. With this recent update, we increase both our historical and projected non-AI Data Center demand, aligning with our Data Center team. We simultaneously revise our expectation of US Compute share down to 55% from 60%; these changes effectively balance out, leaving our total 2026-2028 power shortfall unchanged. As a result of increased global DC power demand projections, we find a 57GW power shortfall before time to power solutions and a net shortfall of 32GW through 2028, 34% of total demanded power. This shortfall is driven by the increasingly power-hungry rack-level GPU architectures (NVL72 or 144 versus extant 8U systems).

Exhibit 3: Net of "time to power" solutions, we project that US data center developers will experience a 32 gigawatt shortfall through 2028—this is a result of intensifying challenges with respect to People, Power and Politics

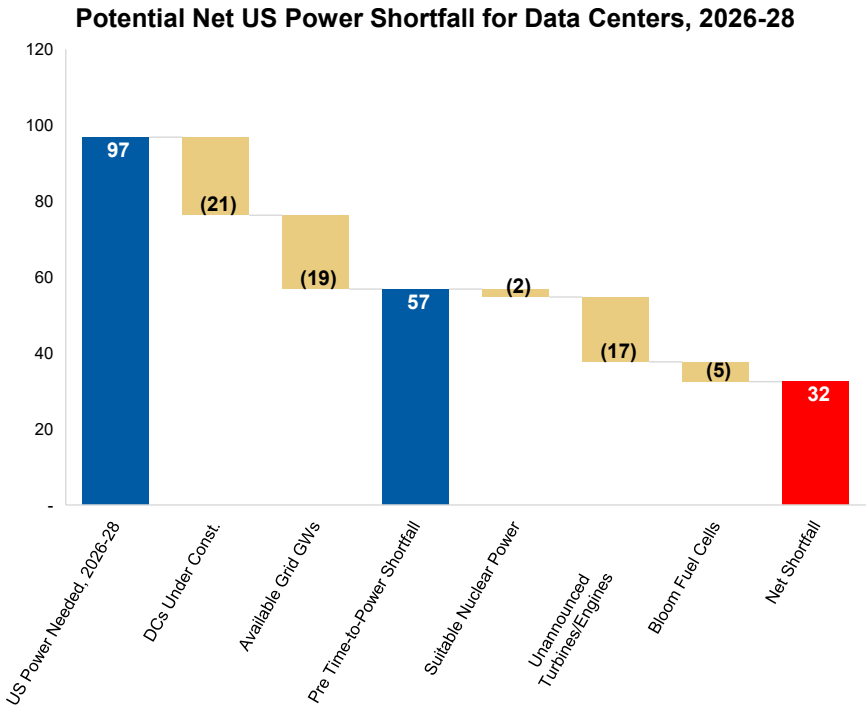


Exhibit 4: Net of innovative "time to power" solutions, we project that US data center developers will face a 32-42GW Shortfall through 2028, 34-43% of demanded power. Note that the PSP sites line is not included in reducing the power shortfall but are included in either "Data Centers Under Construction", "Utility Grid Access", or "Addt'l BTM Turbines/Engines.

Total US Data Center Power Demand (GW), 2026-28	**97**
Less Data Centers Under Construction	(21)
Less Utility Grid Access	(19)
Power Shortfall Before New "Time to Power" Solutions	**57**

	Low End	Midpoint	High End	Probability of Success
Solution #1: Addt'l BTM Turbines/Engines	11	19	49	90%
Solution #2: Bloom Energy Fuel Cells	5	6	8	90%
Solution #3: Site DC at Op'l Nuclear Plant	1	3	8	75%
Solution #4: Convert PSP Sites (included in turbines or DCs under construction)	8	13	19	90%
Probability-Weighted "Time to Power" Solutions through 2028	15	24	57	
Net Shortfall Through 2028	**(42)**	**(32)**	**None**	
Shortfall as a % of 2026-28 US DC Deployments	*43%*	*34%*	*0%*	

We believe this outcome is most likely

Exhibit 5: We project a 9GW Net Shortfall in 2026, scaling to a cumulative 68GWs in 2029.

Part IID.1. Annualized Cumulative Shortfall

Year	2026	2027	2028	2029
Power Needed	22	51	97	160
US DCs Under Construction	(7)	(17)	(21)	(22)
Available/Contractable US Grid Capacity	(3)	(7)	(19)	(36)
Shortfall Before Time to Power Solutions	**12**	**28**	**57**	**102**
Suitable Nuclear Power	(0)	(1)	(2)	(3)
Bloom Fuel Cell Pontential	(1)	(2)	(5)	(9)
Unannounced BTM Turbines/Engines	(2)	(10)	(17)	(23)
Crypto Sites Available (counted in DCs Under Construction or Turbines)	(2)	(6)	(12)	(18)
Est. Midpoint Net Shortfall (Midpoint)	**9**	**15**	**32**	**68**

Part IID.2. Annualized Year-By-Year Shortfall

Year	2026	2027	2028	2029
Power Needed	22	30	45	63
US DCs Under Construction	(7)	(10)	(4)	(1)
Available US Grid Capacity	(3)	(4)	(13)	(17)
Shortfall Before Time to Power Solutions	**12**	**16**	**29**	**45**
Suitable Nuclear Power	(0)	(1)	(1)	(1)
Bloom Fuel Cell Pontential	(1)	(2)	(3)	(4)
Unannounced BTM Turbines/Engines	(2)	(8)	(7)	(6)
Crypto Sites Available (counted in DCs Under Construction or Turbines)	(2)	(4)	(6)	(6)
Est. Midpoint Net Shortfall (Midpoint)	**9**	**5**	**18**	**35**

Exhibits

Exhibit 6: Many CEOs – particularly across the tech supply chain – have stressed the importance of the power shortfall and LPS availability.

Key Recent Quotes on Power Shortfall
"But as we look forward the bottleneck is no longer just in a single chip or a single technology or a single capability. ***The bottleneck moves to power.*** And more specifically the entire power domain from grid to site to core." - On Semiconductor President of Power Solutions Group, Achyut Shah (Sept 16)
"There is no question ***land, power and shell is a problem*** ." - NVIDIA CEO, Jensen Huang (Sept 10)
"I'm not saying supply chips and memory are not constraints. They are constraints. But at least they are constraints we can work with and are more deterministic. ***What is less deterministic is a power site*** ...It's also construction. Construction includes local permitting." - Broadcom CEO, Hock Tan (Sept 10)
"There will be a ***significant electricity shortfall next year*** ." - Elon Musk stressing the importance of power at the G20 tech ministers meeting (Sept 1)
"The fact that data center capacity is now measured in gigawatts rather than flops and tops, underscores one of the most defining challenges of the AI era. ***Power availability has become the primary constraint to further AI progress*** ." - CEO Analog Devices, Vincent Roche (Aug 19)
"But the ***biggest limitation today is power and cooling*** ...There is one challenge we all need to overcome, not just for our industry, but for our society and our planet, and that is power. Every model, every workload, every agent depends on power because at its core" - CEO HPE, Antonio Neri (June 16)

Important note regarding economic sanctions. This report references jurisdictions which may be the subject of economic sanctions. Readers are solely responsible for ensuring that their investment activities are carried out in compliance with applicable laws.

Valuation Methodology and Risks

Broadcom Inc. (AVGO.O)

We value AVGO at 28X CY2027e ModelWare EPS of $18.04. This is roughly 26x non-GAAP EPS of $19.63, which is broadly in-line or below their AI peer group.

Risks to Upside

- Stronger AI revenues
- Faster recovery in core semis businesses
- Realized VMware synergies

Risks to Downside

- Lose networking share to Nvidia (Mellanox)
- ASIC chips are uncompetitive; lose customers to competitors
- Execution on VMware acquisition

NVIDIA Corp. (NVDA.O)

~20 our MW CY27 EPS estimate of $15.01, in-line with the broader market and a discount to compute semis peers (AMD/AVGO/INTC) as high marketshare and gross margins leave limited levers for multiple expansion in the near term.

Risks to Upside

- Growth in inference and training propel data center revenue
- Gaming sales accelerate as GPU based AI PCs gain traction
- Nvidia can recapture lost revenue in China

Risks to Downside

- AI end markets don't materialize as expected, customers sharply reduce GPU purchases
- AMD reemerges as a viable GPU competitor
- Cloud customers outside of Google are able to develop competitive custom hardware