



Are Power Management ICs and MLCCs Becoming the Next Memory Chips?

Roye-Tech Intelligence Center | AI Server Supply Chain Outlook | June 2026

On June 9, Wedbush Securities reported that demand for NVIDIA Grace Blackwell systems, including GB300/B300 platforms, remained exceptionally strong, with customers seeing tougher procurement conditions and longer delivery lead times. Most buyers still look first at GPU availability. Roye-Tech Intelligence Center is watching the next layer: as AI servers move from node-level procurement to rack-scale deployment, power management ICs, MLCCs, aluminum electrolytic capacitors, film capacitors, and SiC/GaN power devices are starting to resemble memory chips one year ago - not in product structure, but in the way AI demand reprices supply.

The conclusion is straightforward. If you are an OEM, server power supplier, module maker, equipment builder, distributor, or spot-market trader, waiting for a formal shortage notice is already too late. The practical response should be built around four actions: preventive production buffers, second-source selection and validation, long-term supply agreements with qualified suppliers, and platform-level AI server power BOM tracking. The rest of this article explains why these actions should begin now.

What to Do Now: Do Not Wait for the Price-Increase Notice

- **Build preventive production buffers:** prioritize high-capacitance MLCCs, high-voltage long-life aluminum electrolytic capacitors, film capacitors, PFC/LLC controllers, isolated gate drivers, and SiC/GaN-related devices instead of blindly stocking every commodity part.
- **Start second-source selection and validation:** for MLCCs, nominal capacitance is not enough. DC bias, temperature characteristics, ESR, package size, lifetime and AVL status must all be checked. For power ICs, driver compatibility, protection behavior, thermal design and digital-control integration matter.
- **Sign long-term supply agreements early:** for parts already designed into customer BOMs and AVL lists, quarterly or semiannual allocation agreements are more valuable than one-off spot quotations. Pricing can float; supply access should not.
- **Create an AI power BOM pool:** track power modules, GPU boards, switch boards, liquid-cooling controls, storage-side power and HVDC-related components by platform, including GB200/GB300, Rubin, Google TPU and custom ASIC server programs.
- **Keep quality and traceability documents ready:** AI server customers are not only buying inventory. Lot consistency, COC, RoHS/REACH, authorized-channel proof and failure-analysis response time can decide whether a part is usable.

1. Memory Has Already Shown the AI Supply-Chain Playbook

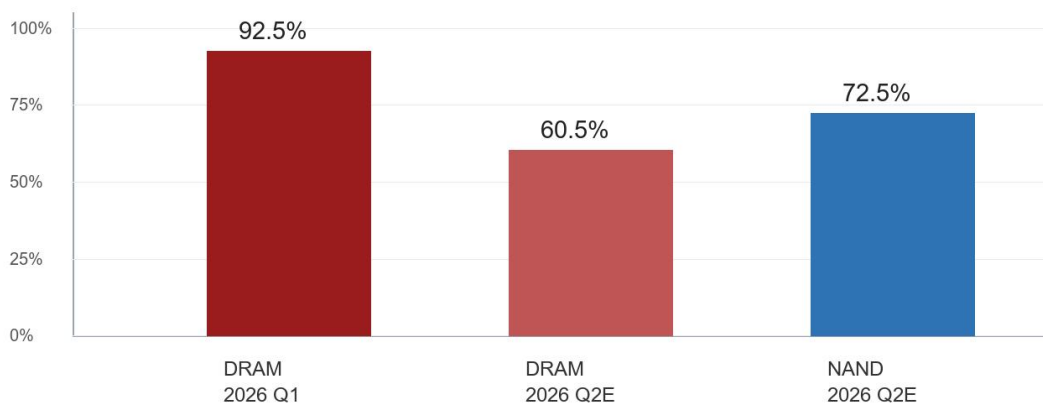
The memory cycle in 2026 has made the transmission mechanism clear. Public TrendForce forecasts indicated that conventional DRAM contract prices rose roughly 90%-95% quarter over quarter in Q1 2026 and were expected to rise another 58%-63% in Q2. NAND Flash contract prices were forecast to increase 70%-75% in Q2. This is not a simple inventory rebound. It is the result of AI infrastructure customers reshuffling capacity allocation, pricing power and shipment priority across the upstream supply chain.



The memory chain moved through three steps. First, hyperscalers and AI infrastructure buyers secured high-end supply early. Second, memory makers shifted capacity and capital expenditure toward HBM, enterprise SSDs and other higher-margin products. Third, even weaker consumer and industrial segments had to accept higher pricing because available capacity had been crowded out. The same pattern is beginning to appear in the server power chain.

AI servers do not only consume more GPUs. They increase rack power density, board-level power conversion complexity, bus filtering requirements, transient-response requirements and thermal reliability thresholds. As long as rack-scale power continues to rise, PMICs and passive components will move from supporting components to delivery-critical components.

Memory has shown the playbook: AI demand reprices upstream supply



Note: Midpoint of public TrendForce forecast ranges; used to illustrate AI-driven upstream price transmission.

Figure 1. Memory pricing shows how AI demand can transmit across upstream component categories.

2. Grace Blackwell Pull-In Means the Power Chain Gets Repriced First

The strength of Grace Blackwell demand is not just a GPU story. It shows that rack-scale AI systems are becoming the purchasing unit of the data center. From GB200 and GB300 to Rubin-class platforms, system complexity keeps rising. Power is no longer a standard accessory inside the server. It is a core subsystem that affects rack stability, energy efficiency, cooling design and shipment timing.

Public reporting on GB200/GB300 NVL72 systems points to a common architecture built around 72 Blackwell GPUs and 36 Grace CPUs per rack. Tom's Hardware reported that Microsoft Azure's GB300 NVL72 cluster used 72 GPUs and 36 Grace CPUs per NVL72 rack, with around 37TB of fast memory. Financial Times previously reported similar GB200 AI rack composition. Once procurement moves to this rack-scale level, component validation and delivery no longer behave like consumer-electronics purchasing.

The power-chain shift has three layers. First, power rating rises as racks move from 100kW-class systems toward multi-100kW systems. Second, efficiency requirements rise because one percentage point of conversion loss becomes a large electricity, thermal and reliability issue at data-center scale. Third, validation barriers rise because customers care more about long-term stability, lot consistency and traceable supply than headline unit price.



Price transmission path in the AI server power chain



Roye-Tech view: not all parts will rise together. The scarcity is in qualified, high-reliability, high-voltage, high-capacitance and high-frequency BOM positions.

Figure 2. AI rack power upgrades transmit into PMICs, power devices and high-reliability passive components.

3. A Rack-Level Estimate: How Much Incremental MLCC and PMIC Demand Appears?

Because NVIDIA and ODMs do not publish full BOMs, exact MLCC and PMIC counts vary by board design, power topology, redundancy scheme, liquid-cooling controls and customer qualification rules. The following is not an official BOM. It is Roye-Tech Intelligence Center's engineering bottom-up estimate based on the mainstream NVL72-class rack architecture, intended to size the order of magnitude of incremental demand.

The model starts from electrical function rather than headline component count:

- **GPU module decoupling network:** each Blackwell GPU complex requires multiple high-current rails. Assuming dozens of power phases across GPU core, HBM, memory I/O and auxiliary rails, and assuming each power phase stacks high-capacitance MLCCs near the load for high-frequency decoupling and transient suppression, a GPU-related module can reasonably reach 3,000-6,000 MLCCs once local regulators, retimers and interconnect devices are included.
- **CPU and memory-side power:** Grace CPU and LPDDR/HBM-adjacent rails require separate decoupling, bulk capacitance and high-frequency bypass networks. The estimate assumes 1,000-2,000 MLCCs per CPU/memory-side module, reflecting lower density than the GPU complex but similar reliability requirements.
- **Power shelf and PSU modules:** moving from 5.5kW-class modules to 18kW-class power modules expands primary-side PFC filtering, secondary-side filtering, auxiliary bias rails, control boards and protection circuits. The MLCC count does not scale linearly with wattage; higher switching frequency, higher transient current and tighter ripple requirements create a steeper component-density curve.
- **Networking, management and cooling controls:** NVLink switching, BMC/control boards, sensors, liquid-cooling controls and high-speed signal conditioning add a meaningful tail of MLCC and PMIC demand that is easy to miss in GPU-only estimates.



Subsystem	Reference Architecture	Estimate Assumption	MLCC Demand per Rack
GPU / accelerator modules	72 GPUs	Power phases + decoupling network + HBM/interconnect support; approx. 3,000-6,000 MLCCs per GPU-related module	216k-432k MLCCs
Grace CPU / memory side	36 Grace CPUs plus LPDDR/HBM-related power	CPU rails + memory-side decoupling + local regulators; approx. 1,000-2,000 MLCCs per module	36k-72k MLCCs
Power modules / power shelf	Approx. 18-36 high-power PSU modules	Primary-side PFC, secondary-side filtering, auxiliary rails and protection/control boards; approx. 5,000-8,000 MLCCs per PSU module	90k-288k MLCCs
Switching, management, cooling controls	NVLink/networking/BMC/sensors/liquid cooling	Approx. 30k-80k MLCCs per rack	30k-80k MLCCs
Rack-level total	NVL72-class AI rack	Roye-Tech integrated estimate	Approx. 370k-870k MLCCs

A similar rack-level approach suggests that PMICs, controllers, drivers, power stages, current-sense ICs, isolation devices and protection components are measured in thousands per rack, not dozens. GPU modules, CPU/memory power, switch boards, PSU modules, fan/liquid-cooling control and monitoring circuits all consume power-management content. For an NVL72-class rack, Roye-Tech estimates roughly 5,800-11,600 power-management-related ICs and power-stage devices per rack.

Incremental AI Racks	Incremental MLCC Demand	Incremental PMIC / Power-Stage Demand
10,000 racks	Approx. 3.7-8.7 billion MLCCs	Approx. 58-116 million devices
30,000 racks	Approx. 11.1-26.1 billion MLCCs	Approx. 174-348 million devices
50,000 racks	Approx. 18.5-43.5 billion MLCCs	Approx. 290-580 million devices

The point is not precision to the last component. The point is where the demand comes from. AI servers are not ordinary servers with a faster GPU installed. A rack concentrates dozens of GPUs, dozens of CPUs, large pools of high-speed memory, switching silicon and high-power PSU modules. Even a modest increase in component count per subsystem becomes a real incremental demand pool once multiplied by rack volume.

4. Will MLCCs Follow the Memory Price Path?

Not exactly, and that distinction matters. Memory is a classic oligopoly: a small number of suppliers control a standardized, capacity-driven market with clear contract-price discovery. MLCCs and PMICs, by contrast, are fragmented markets. Suppliers are numerous, specifications are highly segmented, and price behavior is usually far less synchronized than DRAM or NAND.

The twist is that AI servers can make a fragmented passive-components market behave like an oligopoly at the high end. Roye-Tech calls this a spec-driven oligopoly. Globally, many vendors can manufacture MLCCs. Far fewer can consistently supply the high-capacitance MLCCs that survive NVL72-class operating conditions: elevated temperature, ultra-low ESR, stable DC-bias characteristics, tight lot consistency and long qualification histories with server ODMs. In the most demanding positions, the realistic supplier field narrows toward high-end production lines from Murata and Samsung Electro-Mechanics (SEMCO), with selected Japanese, Taiwanese and domestic alternatives competing only after qualification.



This is why the shortage call should be precise. We are not arguing for a broad-based MLCC shortage. The more likely pattern is a coexistence of structural deflation in low-end and overbuilt commodity part numbers, and absolute shortage in high-end qualified part numbers. For distributors, the spread between these two markets is the opportunity: generic inventory may remain cheap, while qualified high-reliability MLCCs tied to AI server BOMs can trade with allocation, lead-time premiums and spot scarcity.

5. PMICs and Power Devices: Tightness Starts with Qualified High-End Parts

The PMIC shortage logic is less about total wafer capacity and more about qualified-design scarcity. AI server power systems use PFC controllers, LLC controllers, digital power controllers, isolated gate drivers, current-sense ICs, supervisors, hot-swap/eFuse devices, protection ICs, thermal-monitoring devices and sequencers. The unit price of each IC may not look large, but once a device is qualified into a customer BOM, replacement cost becomes high.

For power devices, SiC and GaN benefit from efficiency and power-density pressure. As AI rack power rises, power systems must maintain high efficiency under higher loads while keeping thermals manageable. SiC is better positioned in high-voltage, high-efficiency, high-reliability stages. GaN continues to penetrate high-frequency, high-density and board-level conversion stages. Long term, new capacity may reduce pricing for standard devices. Short term, high-voltage, high-reliability and stable-supply parts still have structural premium potential.

The market symptom will not always be simple 'out of stock.' More often, the original part is constrained, the alternative is not yet qualified, the spot market has inventory but traceability is incomplete, or the available lot cannot meet end-customer documentation requirements. For distributors, value shifts from quoting inventory to managing alternatives, validation documents, lead-time combinations and supply risk.

6. Aluminum Electrolytic and Film Capacitors: Less Hype, Earlier Impact

Aluminum electrolytic and film capacitors are directly exposed to power-density upgrades. Aluminum electrolytics are judged by ripple current, lifetime, temperature and capacitance. Film capacitors are judged by voltage rating, loss, size, temperature rise and long-term stability. They do not trade on a single visible price curve like memory, but project-specific tightness can appear quickly when certain sizes, voltage ratings, lifetime grades and qualified brands are locked by AI power programs.

Upstream raw-material inflation can amplify the pressure. Copper, aluminum, silver paste, resin, ceramic powder, nickel, tin and tantalum price movements tend to pass through passive components with a lag. Competitive commodity specifications may absorb part of the cost pressure, but high-reliability and long-lead-time positions are more likely to pass costs on to customers.

Roye-Tech expects AI-power-related aluminum electrolytic and film capacitors to show three features in the second half of 2026: longer lead times from leading brands, more forward orders from customers, and higher spot premiums for qualified part numbers.

Channel Checks & Supply Chain Whispers

- **Tier 1 Japanese passive vendor:** procurement contacts indicate early lead-time stretch on selected high-voltage, long-life aluminum electrolytic capacitors used in AI PSU input and DC-link positions. The issue is not catalog availability; it is qualified series allocation and production-slot priority.
- **Taiwanese power supply maker:** engineering teams are flagging allocation risk on selected 1200V SiC MOSFETs from US and European suppliers during new high-efficiency PSU platform validation.



Samples are available, but production lots with stable date codes and full traceability are becoming harder to secure.

- **Spot market:** inquiries for high-end isolated gate drivers and digital power controllers with original COC documentation have risen noticeably. Buyers are less interested in lowest price and more focused on traceability, date code, and whether the lot can pass end-customer audit.

7. Component Categories to Watch First

Category	Key Specifications	Price / Supply View
High-cap MLCCs	High capacitance, low ESR, stable DC bias, high temperature reliability	Tightness first, price later; qualified part numbers carry premium
Aluminum electrolytics	High voltage, high capacitance, long lifetime, high ripple current	Lead times likely extend; pricing trends upward moderately
Film capacitors	PFC, DC link, snubber and EMI positions	Project-specific tightness; voltage and form factor matter
SiC / GaN devices	High-efficiency, high-frequency, high-density power stages	High-end parts retain premium; standard parts diverge
Power control / driver ICs	PFC, LLC, digital power, isolated drivers, protection and monitoring	Qualified parts tight; substitution cycle is long
Connectors and HV protection	HVDC, busbar and energy-storage cabinet positions	Demand rises as 800VDC validation advances

8. Supply Chain Mitigation Strategies & Tech Shifts

A rigorous view also needs the devil's advocate. Buyers are not passive. Hyperscalers, ODMs and power-supply makers will use several mitigation tools that can delay or soften a shortage cycle.

- **Waiver process:** when a multi-million-dollar rack is blocked by a few dollars of components, hyperscalers can accelerate engineering waivers, release second-source parts, or temporarily accept slightly lower-spec alternatives if system-level reliability remains within risk limits.
- **Architecture shift:** changes in power topology, TLVR inductor adoption, higher-integration power modules and wafer-level packaged power devices may reduce dependence on some traditional discrete passives over the next two product generations.
- **Friction costs remain:** these mitigation tools do not eliminate short-term opportunity. Qualification gaps, customer audit cycles, firmware/control-loop tuning, thermal revalidation and documentation checks create a validation vacuum. That vacuum is where qualified inventory, traceable lots and credible substitutes command real trading value.

9. Price Outlook: Structural Repricing, Not Universal Inflation

Roye-Tech Intelligence Center's view for the second half of 2026 through 2027 is as follows:

- Commodity parts will not all enter an uncontrolled price spike. Low-voltage, small-capacitance and consumer-grade parts remain relatively well supplied.
- AI server, power-module and HVDC-related specifications will stay structurally tight. Tightness will appear first in lead times and allocation, then in spot pricing.
- Customer-qualified part numbers will be worth more than parameter-matched alternatives. AVL status itself becomes part of the price.
- Memory cost inflation will raise total server BOM pressure. That may make server makers more aggressive in securing availability for other critical components, including power ICs and passives.



- If Rubin, Google TPU, custom ASIC servers and 800VDC architectures continue to ramp in 2027, the power-chain repricing cycle could last longer than the market currently expects.

10. Recommendations for OEMs and Distributors

OEMs should not focus only on GPU lead times. Once AI servers move into volume deployment, bottlenecks often appear in lower-visibility materials: a capacitor that fails qualification, a driver IC whose lead time suddenly extends, or a high-voltage connector that gets allocated to hyperscale customers first. Any of these can delay rack shipment.

OEMs should act on three priorities now. First, identify high-risk parts against the next two quarters of production and build safety stock. Second, prepare at least two qualified alternatives for critical MLCCs, capacitors and power ICs and start sample validation early. Third, sign long-term supply agreements for validated part numbers, clearly defining lead time, allocation and quality responsibility.

Distributors and spot-market traders should also change posture. This AI hardware cycle will not reward simple one-off price quoting. Customers need help managing critical part numbers, alternatives, qualification documents, lead-time risk and inventory timing. Whoever builds the AI power BOM map earlier will have more leverage in the next shortage cycle.

Conclusion: The Power Chain Is Becoming AI Hardware's Second Price Anchor

The memory price cycle has already shown that AI demand does not stop at GPUs. It moves through the data-center hardware stack and reprices upstream components. Strong Grace Blackwell demand is only the opening signal. The next area to watch is the power chain behind higher rack density: PMICs, MLCCs, aluminum electrolytics, film capacitors, SiC/GaN devices and HVDC-related components.

Roye-Tech's core view is clear: the server power chain is likely to follow the path memory has already taken, entering a cycle where lead time, qualification, allocation and high-end specification premiums matter more than generic spot availability. Not every part will rise. But the parts that truly serve AI servers will become increasingly tight.

Public Information Referenced

- Wedbush Securities' June 9 market view on NVIDIA Grace Blackwell demand and delivery conditions.
- TrendForce public forecasts for 2026 DRAM and NAND Flash contract pricing.
- Tom's Hardware reporting on GB300 shipment cadence, GB200 demand and AI-driven memory/storage pricing.
- Financial Times reporting on NVIDIA GB200 AI rack supply-chain and shipment progress.
- Tom's Hardware reporting on Microsoft Azure's GB300 NVL72 cluster, including 72 GPUs, 36 Grace CPUs and approximately 37TB of fast memory per rack.
- TechRadar analysis on 800VDC as an emerging power architecture for next-generation high-density data centers.
- MarketWatch, Business Insider and ITPro reporting on AI-driven memory cost inflation and downstream price transmission.