Rising Data Center Power Demands Call for Safer GaN Gates

By Tao Wei, Sr. Applications Manager & Llew Vaughan-Edmunds, Senior Director, Product Management & Marketing, Navitas Semiconductor

By addressing gate fragilities in GaN power transistors, data center power supplies can exceed 80 PLUS efficiency to handle demanding AI workloads. Here's How:

Introduction

The rise of AI is leading to ever bigger data centers and causing greater demands on the power network. As an example, the United States expects the share of power consumed by these data centers to be 8 percent of total power generated by the end of the decade – a 3-fold increase versus 2022.

This increase is, to a highly significant degree, due to the adoption of ever faster processors that are built to handle AI computational models such as deep neural networks. One example is NVIDIA's DGX H100 "Grace Hopper" processor, which consumes 700 W per device. And this figure will only increase for future generations, with Blackwell set to run at 1 kW per device, and Rubin expected to significantly exceed this. Looking at the bigger picture, this trend is set to increase server power demand per cabinet more than twofold, to about 100 kW. Coupled to this is the need for cooling of the processors and the power supplies, which consumes roughly 30-50 percent of the entire data center's power according to the US Department of the Environment (DOE / ASHRAE).

As this trend progresses, increasing power supply efficiency is critically important for data center operators as well as to the net-zero ambitions of governments worldwide.

Keeping Control of Power

For operators, multiple factors influence the cost of the power each facility consumes. These include the sources of energy, with renewables enabling greener operations but at a premium. The features of the power architecture, including backup generators and uninterruptible power supplies (UPS), can also influence how power is purchased and consumed.

Featured in Power Systems Design – January 2025

Within the data center, power usage effectiveness (PUE) is the industry-standard metric used to express the difference between the power used for computing and the total power supplied. This quantifies the effects of non-computing loads such as cooling, lighting, and other systems, as well as the efficiency of the power conversion and distribution infrastructure. The energy consumed by server cooling systems is a major factor and can be almost equal to the power consumed by the computers themselves. Investing in new power supply equipment to improve the conversion and distribution efficiency can reduce the demand for cooling and increase the PUE, ultimately lowering utility costs.

Eco design regulations, including voluntary and mandatory codes, can compel operators to invest in keeping the data center power architecture up to date and purchasing the latest and most efficient power supplies. Effective eco design regulations are ideally created with a technical knowledge of what constitutes the leading edge, as well as the activity of research communities, in order to set realistic targets for the present and a roadmap for the future.

Being in touch with the wishes of legislators, businesses and users, eco design regulations can drive innovation by product developers and investment by enterprises. Ambitious yet achievable targets can move investment from the list of good intentions to the list of actions.

The core standard regulating power supply efficiency is the 80 PLUS certification. This is a voluntary standard, albeit almost universally adhered to by data center operators and with a slight variant that has been mandated for those operating in the EU.

As the most demanding specification in the 80 PLUS series, Titanium raises efficiency targets at 20%, 50%, and 100% load, and introduces a 10%-load target not featured in previous specifications. The table below compares the efficiency requirements of all 80 PLUS regulations for 230 V internal supplies and highlights any differences in the 230 V EU internal supply specification.

	Percent of load			
Specification	10%	20%	50%	100%
80 PLUS Titanium	90%	94%	96%	94% (91%)
80 PLUS Platinum		90% (92%)	94%	91% (90%)
80 PLUS Gold		88% (90%)	92% (92%)	88% (89%)
80 PLUS Silver		85% (87%)	89% (90%)	85% (87%)
80 PLUS Bronze		81% (85%)	85% (88%)	81% (85%)

Figure 1: 80 PLUS efficiency specifications at 230 V internal input, with any differences in the 230 V EU specification shown in blue

Featured in Power Systems Design – January 2025

In addition, 80 PLUS Titanium specifies a power factor of at least 0.95 at lower load levels. Effectively, this calls for compliant power supplies to contain active power-factor correction (PFC). Figure 2 shows the schematic of a power supply suitable for 80 PLUS Titanium, combining active PFC, essential EMI filtering, and an efficient LLC resonant converter with synchronous rectifiers (SR1, SR2).

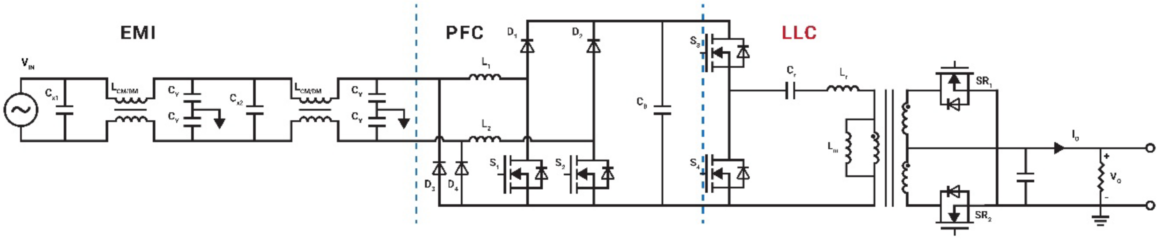

Figure 2: Two-stage server-rack power supply with active PFC and LLC resonant converter

Beyond the Limits of Silicon

While eco design targets drive progress on efficiency, any practicable solution must also comply with standardized industry form factors. The Open Compute Project consortium has defined the OCP as well as the two Common Redundant Power Supply specifications (CRPS185, CRPS265). These specifications simplify deployment and maintenance by enabling modularity and flexibility, as well as easing redundancy.

Each of these three standards sets the same width and height (40 mm x 73.5 mm) for rack power supplies but varies in length between the CRPS (185 mm and 265 mm) and OCP (up to 700 mm) form factors.

As server power demands increase towards 100 kW per cabinet and, ultimately, even higher, the challenge for power supply designers is to deliver more power and raise efficiency, within the pre-defined form factor.

The CRPS specifications standardize two different output voltages, at 12 V and 54 V. In the case of a 12V CRPS, the output current of, say, a 1.5 kW converter would be 125 A. To achieve this using the LLC topology of figure 2, Li et al(1) proposed an optimum number of six output sets. In this design, three sets of two synchronous rectifiers are connected in parallel (figure 3) to minimize the overall footprint.

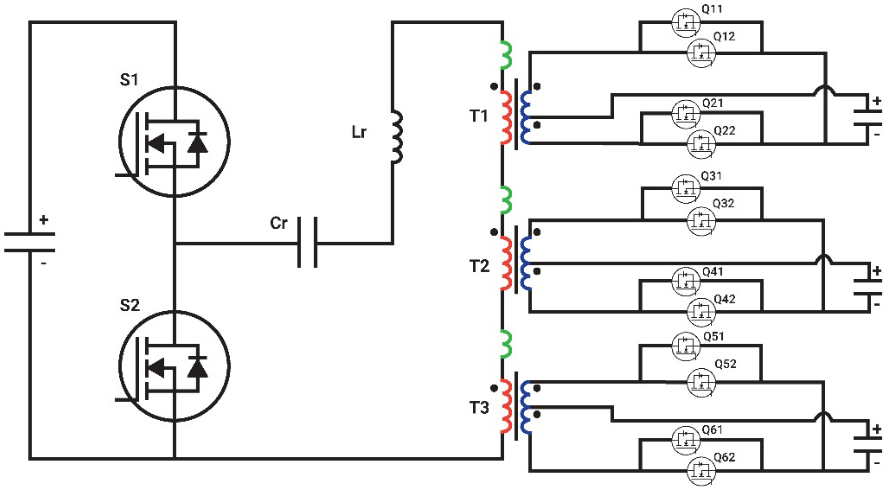

Figure 3: Multi-output CRPS LLC converter with silicon MOSFETs

Building a multi-output LLC converter using silicon transistors calls for a delicate balance between minimizing transformer winding losses, switching and conduction losses in the synchronous rectifiers, and power-supply termination losses. While raising the switching frequency is desirable to allow for smaller magnetic components, termination losses are increased. Using multiple transformers can help avoid termination losses, at the expense of a larger footprint.

Li presented a 1.5 kW server PSU operating at 600 kHz using small planar transformers embedded in the PCB metallization to meet the CRPS form factor. The elevated switching frequency compares with 100-150 kHz typical of CRPS LLC converters in the market today and is well above the maximum practical limit for silicon power MOSFETs. To overcome this and meet 80PLUS Titanium efficiency, the synchronous rectifiers are gallium-nitride (GaN) high electron mobility transistors (HEMTs).

Building on the knowledge gained with this design, Navitas has produced a multi-kilowatt CRPS reference design with a single 54 V output and wirewound transformer operating at 300 kHz. The chosen frequency allows small component sizes, meeting the CRPS 185 mm form factor, and the transformer windings are Litz wire that counteracts skin effects to minimize I^2R losses. As before, GaN power transistors are needed to ensure energy-efficient switching at a frequency unsuited to silicon MOSFETs. So far, the team has scaled this design up to 4.5 kW, equivalent to 83 A, with a single transformer and output circuit.

GaN with Gate Protection

Designing with GaN devices requires care to avoid exposing the fragile gate structure to excessively high or low voltages. The gate voltage range of GaN MOSFETs is -10 V – 7 V, and threshold voltage of 1 V – 2 V; this compares with the gate-voltage range of -10 V – 20 V and 2 V – 5 V threshold of typical silicon MOSFETs. Unwanted effects such as voltage spikes and ringing can damage the GaN transistor gate, potentially accelerating device failure. These can arise when low-side transistors are turned off, causing negative spikes in the gate-source voltage (V_{GS}) that exceed the device's specified maximum. Also, the combined effects of gate-loop inductance and high di/dt can cause severe high-side and low-side V_{GS} ringing. In addition, there is a high risk of shoot-through currents.

Integrating an optimized gate driver in the same package as the GaN HEMT helps designers properly control the V_{GS} to avoid these risks. Implementing the driver in the same package as the GaN HEMT gives control over the inductance and resistance values between the driver and the gate and significantly reduces gate-loop inductance (figure 4), preventing excessive peak voltages. Leveraging protection features also integrated in the package, designers can take a "digital in, power out" approach to simplify power supply design with faster switching, higher efficiency, and superior power density.

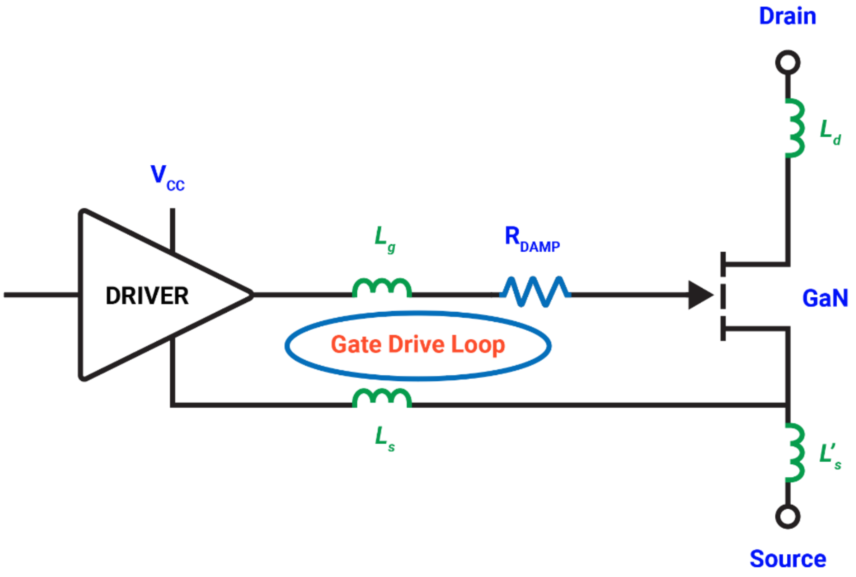

Figure 4: GaNSafe™ integrated power devices contain an optimized driver circuit that protects the gate against excessive spurious voltages

Featured in Power Systems Design – January 2025

Addressing this challenge lets designers safely take advantage of GaN technology to build high-power resonant LLC converters with high efficiency and compact dimensions suitable for OCP and CRPS power supplies. Combining GaNSafe™ ICs with GeneSiC™ MOSFETs, which enable bridgeless totem-pole power-factor correction (PFC) to operate efficiently in continuous conduction mode, Navitas demonstrated a 4.5 kW CRPS185 module (figure 5). The power density of this unit is 137 W/in^3, which cannot be achieved using silicon technology alone.

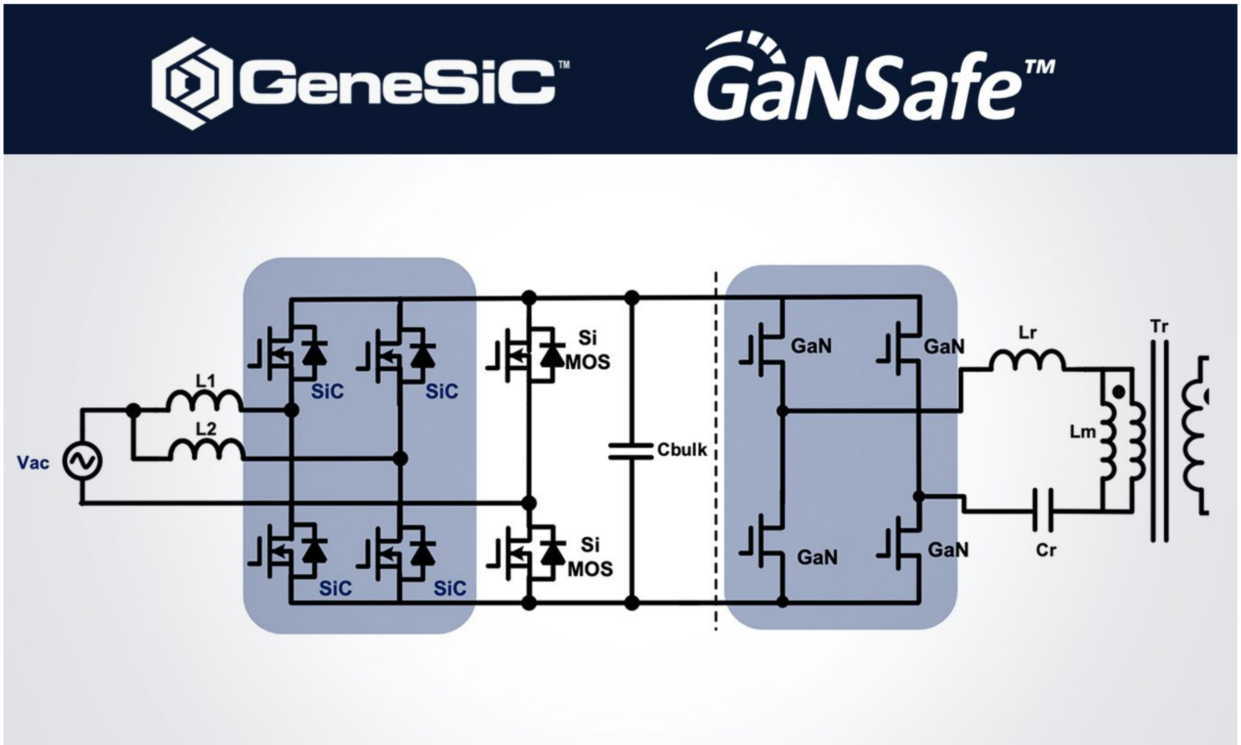

Figure 5: Combining GaN and SiC MOSFETs to create a 4.5 kW CRPS185 reference design with an industry-leading 137 W/in^3 power density

Conclusion

The influx of AI workloads is driving the power demands of large data centers, with projections indicating a rise to 60 kW – 100 kW and beyond for AI-based servers.

Delivering the required power for these servers within the established industry-standard form factors, while fulfilling today's most stringent eco design specifications calls for power supplies operating beyond the limits of ordinary silicon transistors.

In this context, only silicon carbide and gallium nitride technologies can meet these ever-advancing needs.

However, care is needed when working with GaN to avoid exposing the fragile gate structure to excessively high or low voltages. As we have shown, this is achievable and by addressing this it is possible to deliver ever greater power densities and meet the needs of the most demanding of AI data center operator.

References

1. High-Frequency High-Efficiency LLC Module with Planar Matrix Transformer for CRPS Application Using GaN Power IC, PCIM Europe 2023. https://navitassemi.com/wp-content/uploads/2023/06/High-Frequency-High-Efficiency-LLC-Module-with-Planar-Matrix-Transformer-for-CRPS-Application-Using-GaN-Power-IC-paper.pdf

Featured in Power Systems Design – January 2025