Innovative 12 kW Three-level Power Supply for AI Servers Empowered by 400 V SiC MOSFET Technology

Martin Wattenberg, Matthias J. Kasper, Ralf Siemieniec, Gerald Deboy

Infineon Technologies Austria AG, Siemensstr. 2, 9500 Villach, Austria

Corresponding author: Martin Wattenberg, martin.wattenberg@infineon.com
Speaker: Martin Wattenberg, martin.wattenberg@infineon.com

Abstract

This paper presents a 12 kW AC/DC single-phase power supply to meet the growing power demands in data centers driven by training very large AI models. The combination of novel 400V SiC MOSFET device technology and multi-level topology are essential in achieving the desired power density (100 W/in³) and efficiency target ($\eta > 97.5\%$). The introduction of 400 V SiC MOSFET technology bridges the gap between 200 and 600 V super-junction MOSFETs and is characterized by low switching losses and low on-state resistance. In order to keep the load transients of AI chips away from the AC line, a novel control concept involving a power-pulsation buffer circuit is introduced, which acts as an active filter and also provides full control of re-rush currents after line-cycle drop-outs.

1 Introduction

The growth of power consumption of data centers has historically been driven by the increase of power consumption of the CPUs. This historical trend, however, has been tremendously accelerated by the advent of large language models (LLMs) which require massive amounts of parallel GPUs/TPUs to train them. These GPUs/TPUs contain thousands of compute cores compared to just a few in classic CPUs and therefore consume more than twice the amount of power per device. Furthermore, with each new generation, the power consumption of the GPUs/TPUs increases by around 50%. Extrapolating these numbers to the global scale of datacenters shows that the power demand of datacenters of 2% of global electricity consumption today could rise to 7% in 2030, equating to the electricity consumption of India.

In order to cope with the exponentially increasing power demand of hyperscale datacenters, the architecture of these datacenters, as proposed by the consortium of the Open Compute Project, is different compared to classic enterprise datacenters. One of the key differences is that the power supply units (PSUs) are placed in power shelfs at the top and/or bottom of the rack instead of on the motherboards. These compute shelfs house typically six PSUs with a 5+1 redundancy. Typically, three to four power shelfs are placed within one rack. Over the last few years, the power rating of these single-phase PSUs has increased from 3.3 kW to

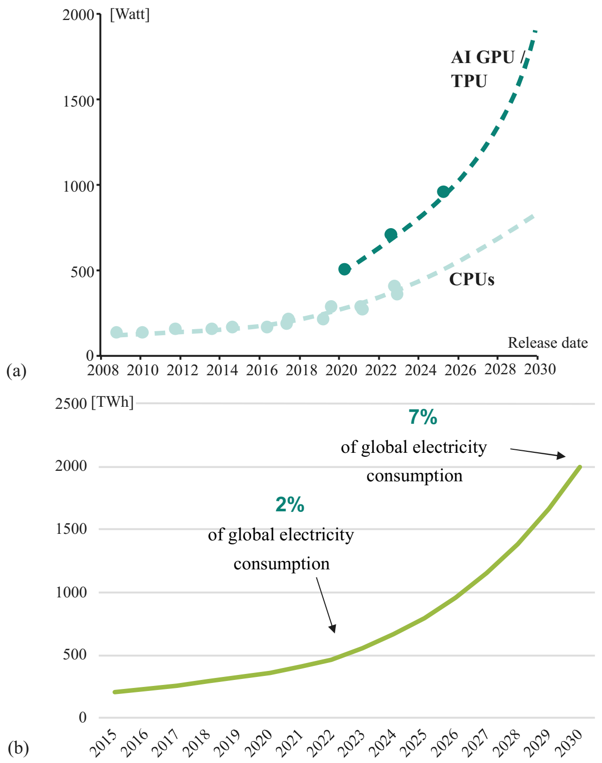

Fig. 1: Growing power demand of AI training GPUs/TPUs (a) and global datacenter electricity consumption (b).

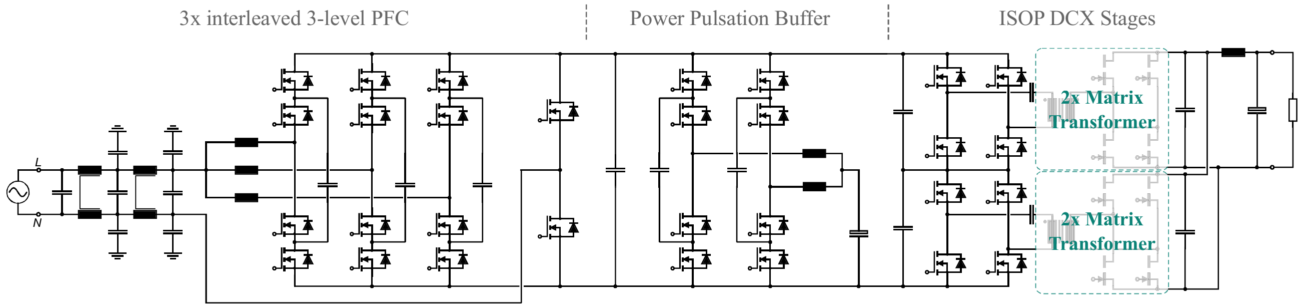

Fig. 2: Schematic overview showing the topology used in the 12 kW demonstrator: Totem-pole PFC with 3x interleaved 3-level flying capacitor bridge legs, power pulsation buffer with 2x interleaved 3-level flying capacitor bridge legs, and unregulated DCX converters in ISOP configuration.

5.5 kW to 8 kW which will be deployed soon. For higher rack-power levels, there is a clear demand for higher PSU power levels, such as 12 kW PSUs as proposed in this paper.

The operation of large amounts of synchronized GPU clusters leads to additional challenges for the power supply infrastructure. When a new training batch is initialized, the power demand of the GPUs rises from 0% load to around 200% for around 500 μs and the step-wise recedes to 100%. These load steps occur every 1-2 s and shall not be seen on the AC line as they create sub-harmonics in the data center power infrastructure and potentially increased cost due to a peak charge component to the total electricity bill [1]. One possible solution for handling load jumps are super-capacitor shelves, which, however, have a delay of a few milliseconds before reacting and additionally take away valuable space inside the compute rack. In the system proposed in this paper, a power pulsation buffer stage is employed which allows to compensate load jumps.

A second challenge arises from the maximal thermal budget and physical limits to air-cooling. As such PSUs typically are limited to a single 40 mm fan, maximal air volume and pressure are limited. This problem is exacerbated by the fact that larger PSUs not only generate more losses in total but also exhibit increased pressure drop due to their extended length, making heat extraction significantly more challenging. On the one hand, if component placement is too restrictive for air to flow, the power capability is penalized by inadequate cooling. On the other hand, if component placement is too generous, the targeted power density might not be achieved. Additionally, real world constraints have to be taken into account. Due to close proximity of several PSUs in one shelve, heat dissipation through the metal housing is limited and top-side cooling through the housing is not viable in steady-state.

This paper presents a comprehensive study of a 12 kW server PSU, targeting the specific challenges and requirements of modern, AI-focused data centers. Thanks to innovations in semiconductors, digital controllers, topologies as well as a holistic design approach that incorporates mechanical and thermal constraints with the optimized choice of topology, it was possible to overcome traditional limits in PSU design.

2 System Concept

In this section, the 12 kW PSU is introduced and the underlying design philosophy is presented. The main specification of the system are listed in **Tab. 1**.

2.1 Design considerations

In order achieve the optimal trade-off between efficiency, power density and thermal design, the 12 kW PSU consists of two 6 kW units with less then 1/2 U height, which are stacked together such that they fit into a 1 U height. Splitting the design into two modules has the added benefit of a wider range of available components, esp. the choice

Tab. 1: Main specifications of the 12 kW PSU.

Parameter	Value
Full power input voltage range	1ϕ, $208-305\,\mathrm{V_{rms}}$
De-rated input voltage range	1ϕ, $180-208\,\mathrm{V_{rms}}$
AC input frequency	$47-63\,\mathrm{Hz}$
Rated output power	$12\,\mathrm{kW}$
Output voltage at full load	$49\,\mathrm{V}$
Output voltage at no load	$50\,\mathrm{V}$
Hold-up time	$20\,\mathrm{ms}$
Physical dimensions (WxHxL)	$68\mathrm{x}40\mathrm{x}705\,\mathrm{mm}^3$

of fuses and relays on the input can otherwise be limited. Fuses on the input of each module can also be rated for a lower current level, thus yielding faster blow time.

The selected topology of a 6 kW unit is shown in **Fig. 2**. The topology consists of three building blocks, namely the PFC, the power pulsation buffer, and the DC/DC stage, which will be described in the following.

2.1.1 PFC stage

In order to reduce the size of the magnetics in the PFC and to achieve a high efficiency, a three times interleaved, three-level flying capacitor totem-pole PFC topology is chosen. The benefits of the flying capacitor bridge-leg are the introduction of a third voltage level in the switching waveforms, which reduces the voltage applied across the boost inductors. Additionally, due to the interleaved operation of the switches within a bridge-leg, the effective switching frequency of a bridge-leg is twice the switching frequency of the semiconductors. As a consequence, the voltage time-area seen by the boost inductors is reduced by a factor of four compared to two-level bridge-legs, leading to four times lower inductance values. Additionally, by interleaving three flying capacitor bridge-legs, the current stress on the inductors as well as the semiconductors is reduced proportionally. This ultimately allows to employ low profile inductors fitting into the height limitation of 1/2 U. The parallel interleaving the bridge-legs increases also the effective switching frequency seen by the EMI filter and is six times higher than the switching frequency of the individual semiconductors ($f_{\text{sw}} = 80\,\text{kHz}$), i.e. the first harmonic within the EMI band is at 480 kHz, which allows to reduce the size of the EMI filter to fit into the height requirements. An additional advantage of the series and parallel interleaving applied in this PSU is the effect of heat-spreading among multiple inductors and semiconductors which prevent the creation of hot-spots.

2.1.2 Power Pulsation Buffer (PPB)

As data center operators demand the uniterrupted operation of the servers even in the case of line-cycle dropouts (LCDO), all PSUs are required to have a hold-up time, i.e. 20 ms, during which they have to provide the full amount of power (12 kW) to the output. This means that at least 240 J of energy have to be stored inside the PSU, typically within electrolytic capacitors. In order to reduce the size of these capacitors, it is important that during a LCDO a significant portion of the stored energy can be extracted from these capacitors. For this reason, a power pulsation buffer is used, which decouples the energy storage capcitors from the DC-link and allows to deplete them to around half the voltage during LCDO, thus extracting around 75% of the available energy. The PPB circuit can also compensate the twice line-frequency power pulsation from the grid and provide a stable DC-link voltage in all conditions, even during load-jumps. This enables the use of an unregulated DC/DC stage as a third stage. The bridge-legs of the PPB are also designed as three-level flying capacitor bridge-legs to reduce the size of the inductors. The higher effective switching frequency combined with lower inductance for the same current ripple leads to faster transients and higher control band-width.

2.1.3 DC/DC Stage

Due to the pre-regulation of the DC-link voltage by the PPB, an unregulated DC/DC converter stage can be employed. Two series resonant converters (DCXs) operating at the resonant frequency are employed in an input-series output-parallel connection with matrix transformers. This topology offers the advantages of

- load-independent soft-switching given by the magnetizing current,
- resonant current with low harmonic content for reduced AC-losses,
- reduced number of component compared to LLCs by utilizing the leakage inductance for the resonance, and
- self-balancing of the stacked input capacitors due to ISOP configuration [2].

The step-down ratio of the DC/DC stage is selected to be 10:1 such that a DC-link voltage of $V_{\text{DC}} = 500\,\text{V}$ is obtained for the nominal output voltage of $V_{\text{out}} = 50\,\text{V}$. Since all high-frequency devices in the PFC, the PPB, and the DC/DC stage are exposed to half of the DC-link voltage, i.e. 250 V, devices with a blocking voltage of $V_{\text{DS}} = 400\,\text{V}$ are a perfect choice, such as the newly developed CoolSiC™ devices of Infineon.

2.2 400 V Silicon Carbide MOSFETs

Until recently, there were no competitive semiconductor solutions on market with blocking voltages of around 400 V. Scaling-up existing medium-

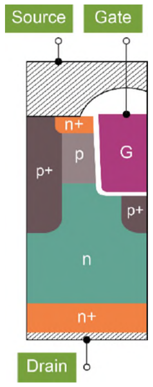

Fig. 3: Cross section of the 400 V SiC MOSFETs.

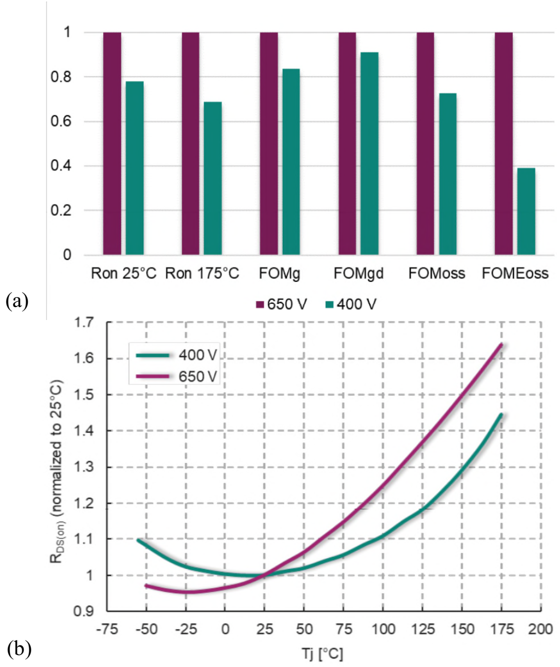

Fig. 4: Figure-of-mertis of the 400 V SiC MOSFETs in comparison to the 650 V counterparts: (a) switching performance related FOMs, and (b) on-state resistance depending on the junction temperature.

voltage 200 V MOSFETs that employ lateral charge-compensation by means of an isolated deep field-plate is not ideal. Similarly, scaling-down existing 600 V SJ MOSFET technologies is problematic. In both cases, the application performance suffers from the compromises originating from the specific device setups, namely large input-, output- and reverse-recovery charges, and a pronounced drop in the output- and miller-capacitance over drain voltage. These properties prevent the devices from common usage in hard-switching half- or full-bridge topologies. The recently introduced 400 V SiC MOSFET devices overcome these limitations and offer improved performance and efficiency, featuring small gate-, output-, and reverse-recovery charges and a highly linear output and Miller capacitance over drain voltage [3].

The general device structure follows the design approach as introduced previously [4], [5]. **Fig. 3** gives a schematic cross section of the general cell concept. The active channel aligns along the a-plane to provide the best channel mobility and lowest interface trap density. The gate oxide is protected by deep p-wells that are connected to the source electrode at the semiconductor surface. As the 2nd trench sidewall does not coincide with this crystal plane, it is not used as an active channel. Instead, the buried p-region is connected to the source electrode along the inactive sidewall. The new 400 V MOSFET benefits from the continuous improvements of the technology that enable a clearly reduced cell pitch, improved channel properties, and improved control over the drift region properties. **Fig. 4(a)** gives a comparison of some key device parameters between the new 400 V and 650 V CoolSiC™ technology. The temperature dependence of the on-resistance stands for one of the key advantages of the new 400 V technology, and increases by only 11% from 25 °C to 100 °C as shown in **Fig. 4(b)**.

2.3 Control scheme

The utilization of the PPB in this system enables the usage of a novel control scheme. In this control scheme the PPB is used to control the output voltage indirectly by regulating the DC-link voltage which is then stepped down with a fixed conversion ratio by the DCX stages (cf. **Fig. 5**). The control-loop implementation of the PPB consists of a cascaded structure with an outer voltage control loop and an inner current control loop. The primary control target is to provide a regulated output voltage at all times also in the events of load jumps, line-cycle drop outs, and in the presence of the inherent power-pulsation coming from the PFC. The PFC control loop is implemented such that it regulates the average voltage of the PPB capacitors $U_{\mathrm{C,PPB,LF}}$ and also consists of a cascaded structure with an inner current control loop. Advantageously, a feed-forward term of the twice line-frequency power pulsation is provided to the PPB controller by the PFC controller.

As the DC-link voltage is always controlled and higher than the AC input voltage, the PFC stage is always able to control the input current. This is

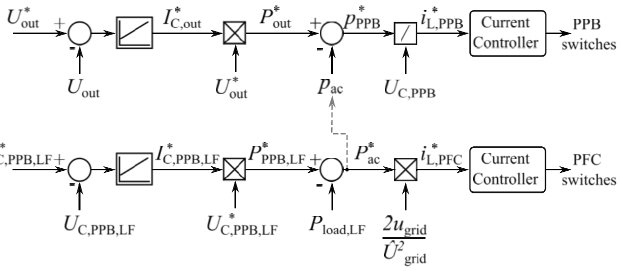

Fig. 5: Implemented control loop design for the power pulsation buffer and the PFC.

especially valuable in case of LCDO where typical PFCs without PPB suffer from the re-rush current phenomenon. This occurs when the AC input voltage returns while the DC-link voltage has dropped due to the power delivery to the load. In this case, the AC voltage can be higher than the DC-link voltage and consequently the PFC can not control the current anymore. This leads to current spikes on the AC grid, which can cause disturbances in data centers, such as

- triggering of fuses,
- triggering of alarms,
- UPS discharge, and/or
- excitation of resonances.

This unwanted behavior can be prevented with the proposed system implementation.
When considering high power-density PSUs, the thermal limit is ultimately the most critical limit to respect. Not only is the electrical performance affected by heat but also lifetime of the components, esp. wet-electrolytic capacitors quickly degrades at elevated temperatures. Power loss at full load, loss distribution across the components as well as their placement inside an airflow through the PSU are all closely linked. While this problem has always required attention from the design engineer, the problem is significantly exacerbated on modern high-power designs. On the one side, supplies get longer to account somewhat for the increased power of a PSU, but this leads to an increased pressure drop the fan needs to overcome. On the other side, because supplies get longer, not all components are provided directly with cool air from the intake side but may sit downstream in the airflow and receive pre-heated air. Traditionally, PSUs in the data center could maintain operation within the thermal specifications with a single fan. Even though more fans, e.g. on the input and output side, are possible, they occupy valuable space and take power themselves to operate, thus degrading the PSU efficiency.

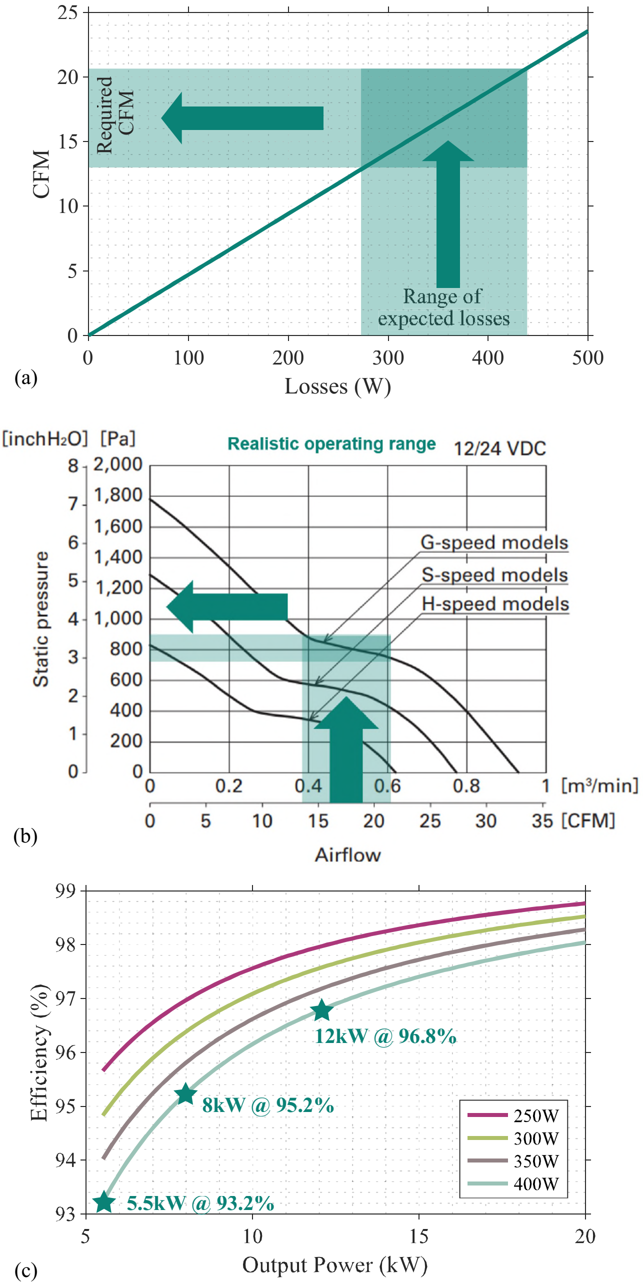

Fig. 6: Thermal calculations to estimate a reasonable loss budget with only one fan in the system.

2.4 Thermal optimization

In order to estimate the possible loss budget that can be dissipated by a single fan, following basic calculations are performed. At first, based on the heat capacity of air, the inlet and outlet temperatures, it is straightforward to derive the relation between the required air-flow in cubic feet per meter (CFM) and the maximum amount of losses that

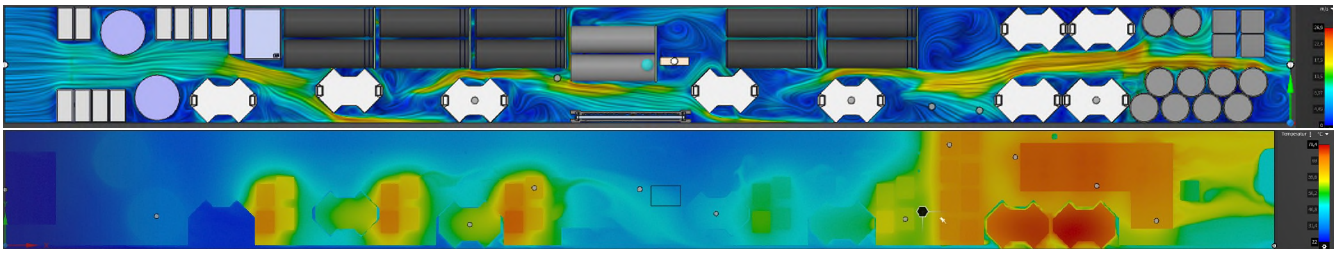

Fig. 7: Airspeed and peak temperature from CFD simulation at the top and bottom respectively. Placement of components was chosen to increase effective surface. Predicted hotspot (secondary sync. rec) was moved close to the outlet to prevent spill-over effects.

can be dissipated. It is reasonable to assume heat transfer to not be perfect and outlet airflow to exhibit thermal variations. **Fig. 6** shows the range of required CFM based on a range of expected power loss at full load. As a first-order approximation, the thermal power P can be calculated based on the allowed temperature rise ΔT, the heat capacity of air C_ρ. and the mass flow q_m:

$$P = \Delta T \cdot C_\rho \cdot q_m. \tag{1}$$

Given one of the strongest fans in the market, a realistic target for the air-flow can be between 12 and 22 CFM, which equates to a maximal allowable pressure range of 700 to 900 Pa. On the one hand if the back-pressure is too low, heat transfer could be enhanced with increased surface area, e.g. via pin-fin heatsinks. On the other hand, if the back-pressure is too high, not enough air passes through to stay within the temperature envelope. The chosen operation point is right in the stall-region of the fans operation profile. Marginally lower back-pressure (10 to 15%) can increase the transferred air volume easily by 33%, e.g. from 15 to 20 CFM. Thus, slightly more generous spacing between components with carefully planned channels for the air to flow can have a dramatic effect and temperature, efficiency (lower I^2R-losses) and life-time. Based on that, a maximum loss budget of 300 to 400 W can be derived which can be translated into minimum full load efficiency numbers for different output power levels. For a 12 kW a minimum full-load efficiency of 96.8% is required for an assumed loss budget of 400 W.

Based on a virtual design, thermal and CFD calculations have been carried out and are shown in **Fig. 7**. The air intake as well as AC input are on the left and the outlet as well as DC output are on the right side. Components were placed carefully, to allow for as little obstruction as possible while simultaneously guiding the air-flow to components that need to dissipate heat and enhance heat transfer.

3 Hardware Demonstrator

Based on the previous design considerations, a hardware demonstrator has been built up (cf. **Fig. 8**). The components were optimized by means of Pareto optimizations such that the highest efficiency is obtained within the stringent height limitation of each 6 kW module. The height constraint was a predefined limitation for the Pareto optimization that in particular affected the choice of core sizes for the magnetic components and the size of capacitors. The remaining degrees of freedom were, among others, the choice of winding configurations, switching frequencies, and semiconductor chip sizes. The selected values of the main components in the system are listed in **Tab. 2**.

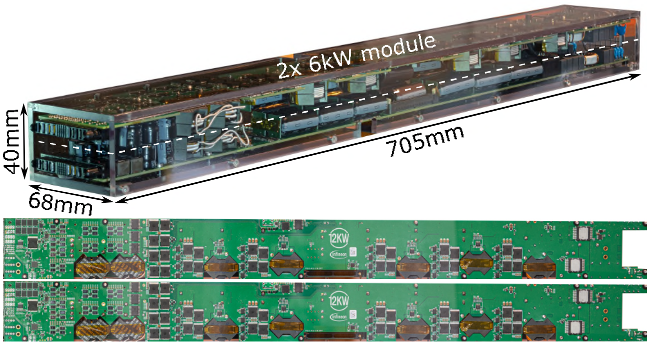

Fig. 8: 12 kW demonstrator with 2 6 kW modules stacked in acrylic housing.

The calculated efficiency of the entire PSU as well as the individual stages at $V_{\mathrm{in,rms}}$ = 277 V is shown in **Fig. 9**.

The system is practically controlled by one MCU placed on the primary side, which drives the three interleaved 3L PFC stages, the two interleaved 3L PPB stages and the primariy and secondary side DCX. No secondary side sync. rec. controllers are needed as the DCX operates in resonance with no phase phase shift between primary and secondary side. This allows to finely tune switching patterns and minimize deadtime losses. Propagation delays are reasonably closely matched by using the same coreless transformer (CT) isolation technology on all signals.

A small secondary side MCU feeds back information about the output voltage and current to the primary side across the isolation barrier. This MCU could also interface with the PMbus and provide the necessary information to the rack controller.

Tab. 2: Main components of the 12 kW PSU technology demonstrator .

Parameter	Value
Totem-Pole PFC Stage	
Boost inductors L_{B}	60 μH, RM12LP N97 ferrite, 21 turns, 150x71 μm litz
LF switches	600 V CoolMOS™ 7 mΩ_{typ} IPDQ60R007CM8
HF switches	400 V CoolSiC™ 36 mΩ_{typ} IMT40R036M2H
In-phase i-sensor	Isolated Hall hybrid, 10 MHz$_{\text{typ.}}$, TLE4978
Sw. Freq. f_{sw}	80 kHz (per MOSFET)
EMI filter	2-stage π filter
PPB Stage	
Switches	400 V CoolSiC™ 25 mΩ_{typ} IMT40R025M2H
In-phase i-sensor	Isolated Hall hybrid, 10 MHz$_{\text{typ.}}$, TLE4978
Inductors L_{PPB}	40 μH, RM12LP N97 ferrite, 18 turns, 180x71 μm litz
Sw. Freq. f_{sw}	120 kHz
Capacitors	450 V, 6x 180 μF
DC/DC Stage	
Primary Switches	400 V CoolSiC™ 15 mΩ_{typ} IMT40R015M2H
Secondary switches	80 V CoolGaN™ SG HEMTs 1.8 mΩ_{typ} IGC025S08S1
Output i-sensor	Isolated TMR, 2.5 MHz$_{\text{min.}}$, TLE5571
Transformer	RM12LP core, N49 ferrite, 5 turns prim., 595x40 μm litz, 2 turns sec., 2380x40 μm litz, $L_{\text{m}} = 40\,\mu\text{H}$
Sw. Freq. f_{sw}	425 kHz

4 Experimental Results

At the time of writing this paper, first preliminary tests have been carried out, validating the proposed novel control concept at reduced power and voltage levels in hardware. The shown waveforms are not necessarily representative of the final, optimized implementation but already show feasibility in practice.

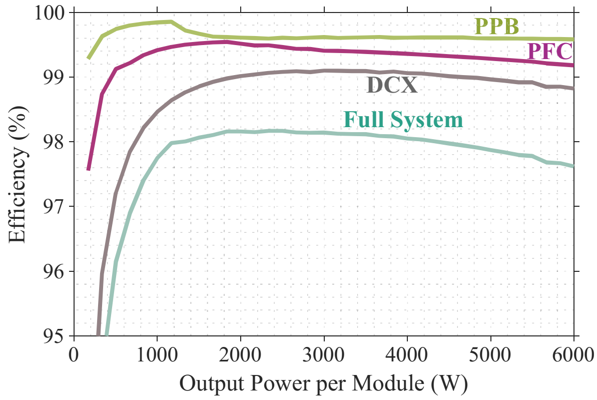

Fig. 9: Calculated efficiency of the entire PSU and the individual converter stages in dependence of the output power for 277 V_{rms} input.

The waveforms at the top of **Fig. 10** show the PFC input current as a total sum I_{ac} and per phase, i.e. I_{ph1} to I_{ph3} along with the AC input voltage V_{ac} measured right at the input terminals before the EMI fitler. Furthermore, the DC-link voltage V_{dc} and power pulsation buffer voltage V_{ppb} are shown. Here, the PPB is running at a fixed duty cycle of 90%. In the center of **Fig. 10** the ripple compensation via PPB is enabled dynamically at run-time. It can be see how the 100 Hz ripple on V_{dc} is almost entirely eliminated while the ripple on V_{ppb} increases. As intended, this behaviour has no impact on the AC input current. Lastly, at the bottom of **Fig. 10**, the differential output voltages of each of the two stacked DCX stages along with the primary side transformer current are shown. The combined output current on the secondary side was 40 A. Achieving full ZVS transition on all switching transients was achived with a deadtime of 100 ns. Deadtime will shorten when operating the DCX at the targeted voltage which increases magnetizing current, thus accelerates the transitions. No active balancing between the upper and lower DCX is carried out or needed due to the self-balancing effect of the ISOP configuration [2]. Additionally, the transformers were hand-wound prototypes which exhibit some variation in stray and main inductance. Professionally wound transformers should improved matching even further.

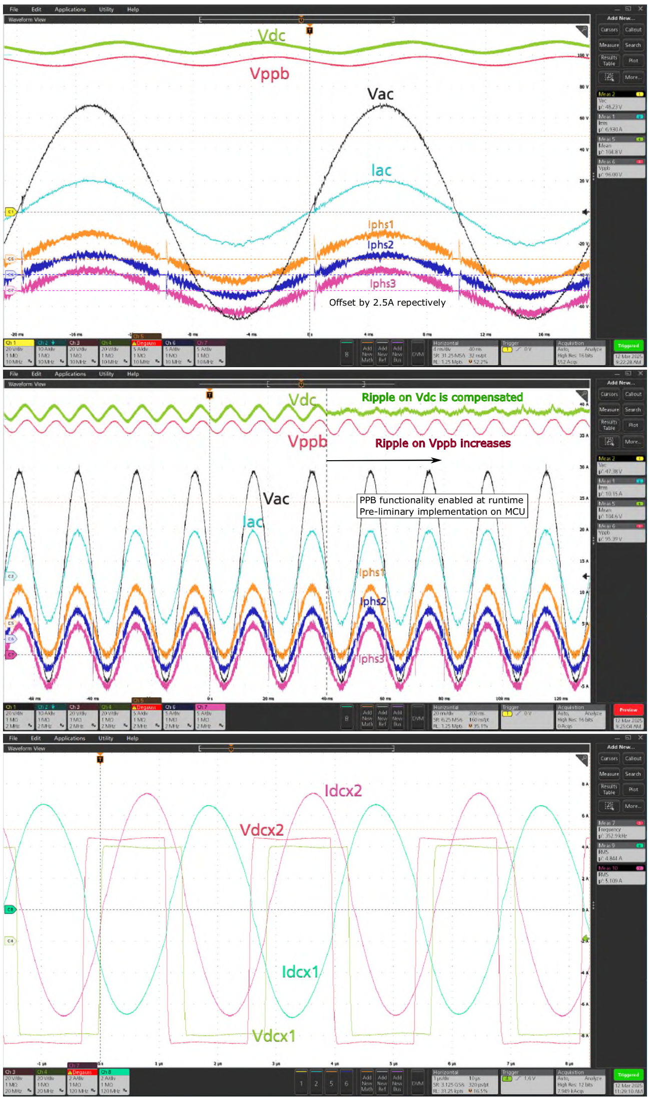

Fig. 10: Preliminary measurement results. Top: Interleaved PFC without PPB control. Center: PPB enabled at runtime. Bottom: stacked DCX voltage and current

5 Conclusion and Outlook

The training of large language models for AI with trillions of parameters poses a massive challenge for data centers, as the power demand per server rack is rising with unprecedented speed. Hence, new solutions are needed in all power conversion stages along the entire power flow.

As a consequence, future AC/DC server power supplies will have power ratings significantly above the 5.5 kW currently specified by the Open Compute Project. In order to fit higher power levels in the existing height of 1 U (40 mm), multi-level topologies are a key enabler for highest efficiency at unrivalled power density. The newly developed 400 V CoolSiC™ MOSFETs are perfectly suited for this type of topology by offering improved figure-of-merits compared to their higher voltage counterparts and an ultra-flat on-state resistance versus junction temperature.

A novel control concept is proposed, which utilizes the power pulsation buffer to regulate the output voltage indirectly by controlling the DC-link voltage. This offers several advantages, both in terms of performance improvements and low disturbances to the AC grid, namely

- reduction of bulk capacitor size,
- usage of unregulated DC/DC stage (DCX) instead of LLC,
- peak-shaving of load power to reduce stress on AC grid, and
- controlled grid currents after line-cycle drop-out.

One of the biggest challenges in designing such a power supply is the thermal management due to the limited capability to extract losses with only a single fan. Hence, the decision was made to split the system into two 6 kW units with low heights that are stacked together. This provides a better means of heatspreading among the boards and facilitates the air-flow through the system.

The 12 kW power supply will enable rack power levels of up to 240 kW (four shelves with 5+1 redundancy). At this power level also the 48 V bus bars will reach their limits, which marks a natural transition point to 3-phase power supplies with high-voltage power distribution in the future.

References

[1] M. Dabbagh et al, “Shaving data center power demand peaks through energy storage and workload shifting control,” *IEEE Transactions on Cloud Computing*, 2019.

[2] M. Kasper, D. Bortis, and J. W. Kolar, “Scaling and Balancing of Multi-cell Converters,” in *IPEC*, 2014, pp. 2079–2086.

[3] R. Siemieniec et al, “New 400V SiC MOSFET and its use in Multi-Level Applications,” *ECCE Europe*, 2024.

[4] D. Peters et al, “The New CoolSiC™ Trench MOSFET Technology for Low Gate Oxide Stress and High Performance,” *PCIM*, 2017.

[5] R. Siemieniec et al, “650 V SiC Trench MOSFET for High-Efficiency Power Supplies,” *PCIM*, 2017.