ARTICLE

Power Pulsating Buffer to meet peak power demands in AI server PSUs without disturbing the grid

Sam Abdel-Rahman, System Architect Server SMPS, Infineon Technologies AG

www.infineon.com

Table of contents

Abstract

The rise of artificial intelligence (AI) has resulted in a significant increase in power demand in data centers. AI server racks will rise to higher power levels reaching 1 MW. Aside from the significant nominal-power rise of the AI PSU, the GPU also draws a higher peak power and generates high load transients, as shown in Figure 1. This high power and high load transient dictates new architectural changes in AC and DC distribution of data center racks to accommodate the significantly higher currents, while also considering the grid power capabilities. As a part of this new architecture, we introduce concepts of energy buffer (also known as Power Pulsation Buffer or PPB) to support AI peak loads without overloading the grid side. This article will also provide an overview of trends and the evolution of PSUs and rack architecture for AI servers.

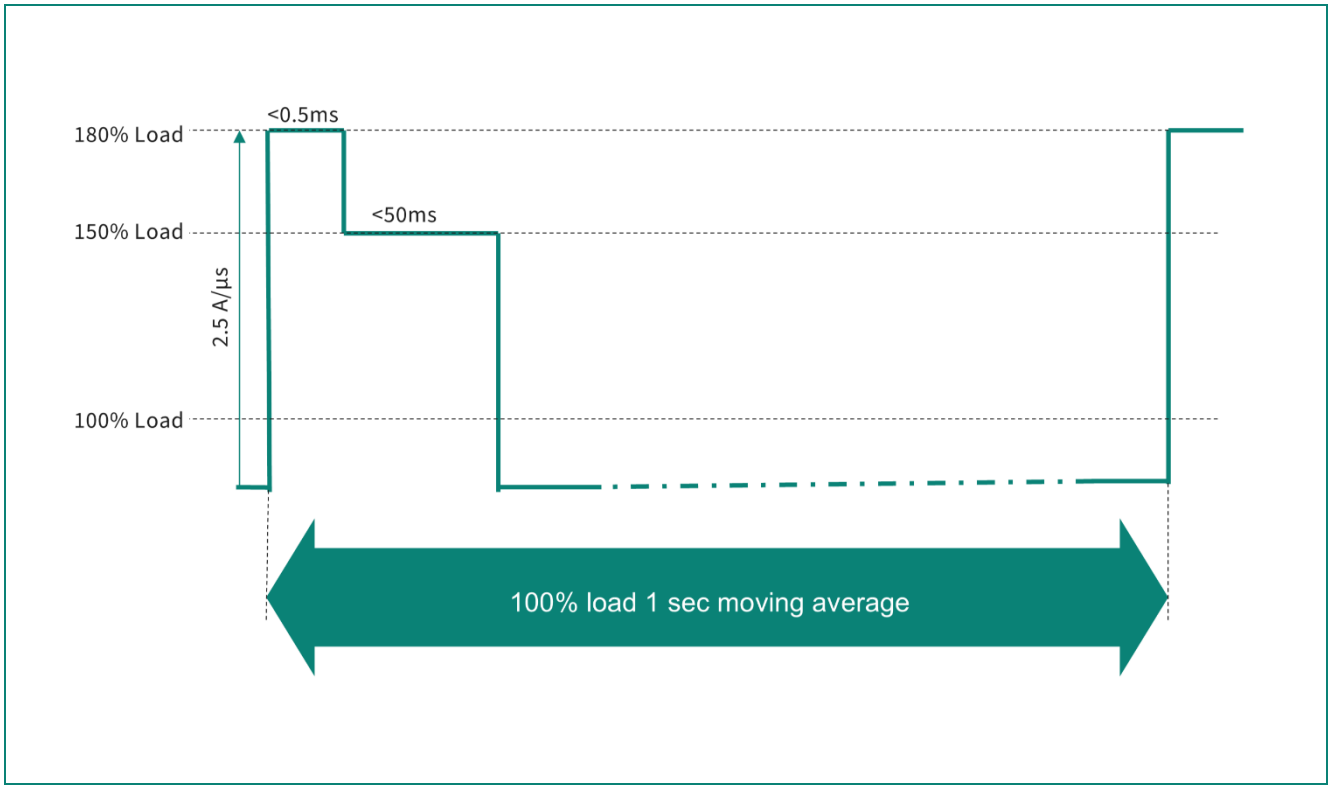

Figure 1 Example peak power profile demanded by AI GPUs

1 Peak power demand in next generation AI server racks

Figure 2 shows the next generation of server rack architecture in this evolution. As the rack power rises above 300 kW, the power components are separated from the IT components in a power sidecar that contains the power supply units (PSUs), the battery backup units (BBUs), and the capacitor backup units (CBUs).

With PSU power reaching up to 30 kW, with 3-phase AC input and high voltage DC (HVDC) output, The power sidecar can extend the rack power up to 1.3 MW. To put this rack power increase in perspective, it is around 10 or even 20 times the rack power of the current generation of server racks.

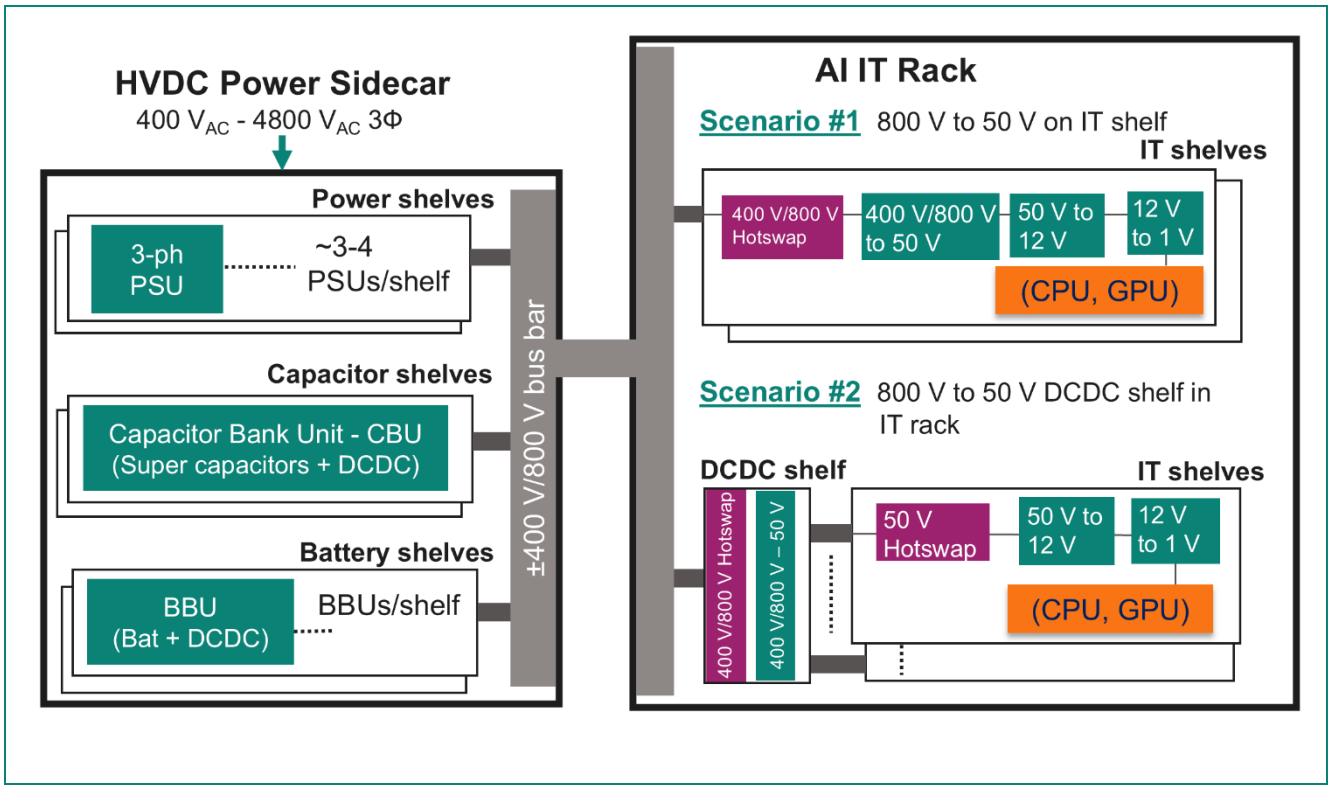

Figure 2 Next generation AI server rack architecture

Considering all GPUs working synchronously in the same IT rack and also with other IT racks in the same datacenter cluster, demanding peak power as shown in Figure 1, we must expect a large current transient on the grid and in the datacenter's front-end power distribution, such as the medium voltage transformers, protection devices, etc.

This overloading of the grid can be illustrated by simulating a PSU input current during the peak power event. Figure 3 shows a PLECs simulation of a traditional two-stage PSU during a peak load event. Here, the energy is first supplied from the bus capacitor causing its voltage to drop, and as the PFC voltage controller tries to recover its output voltage, the controller will command its inner current loop with a higher reference current, causing the line current to overshoot higher than the nominal rated current.

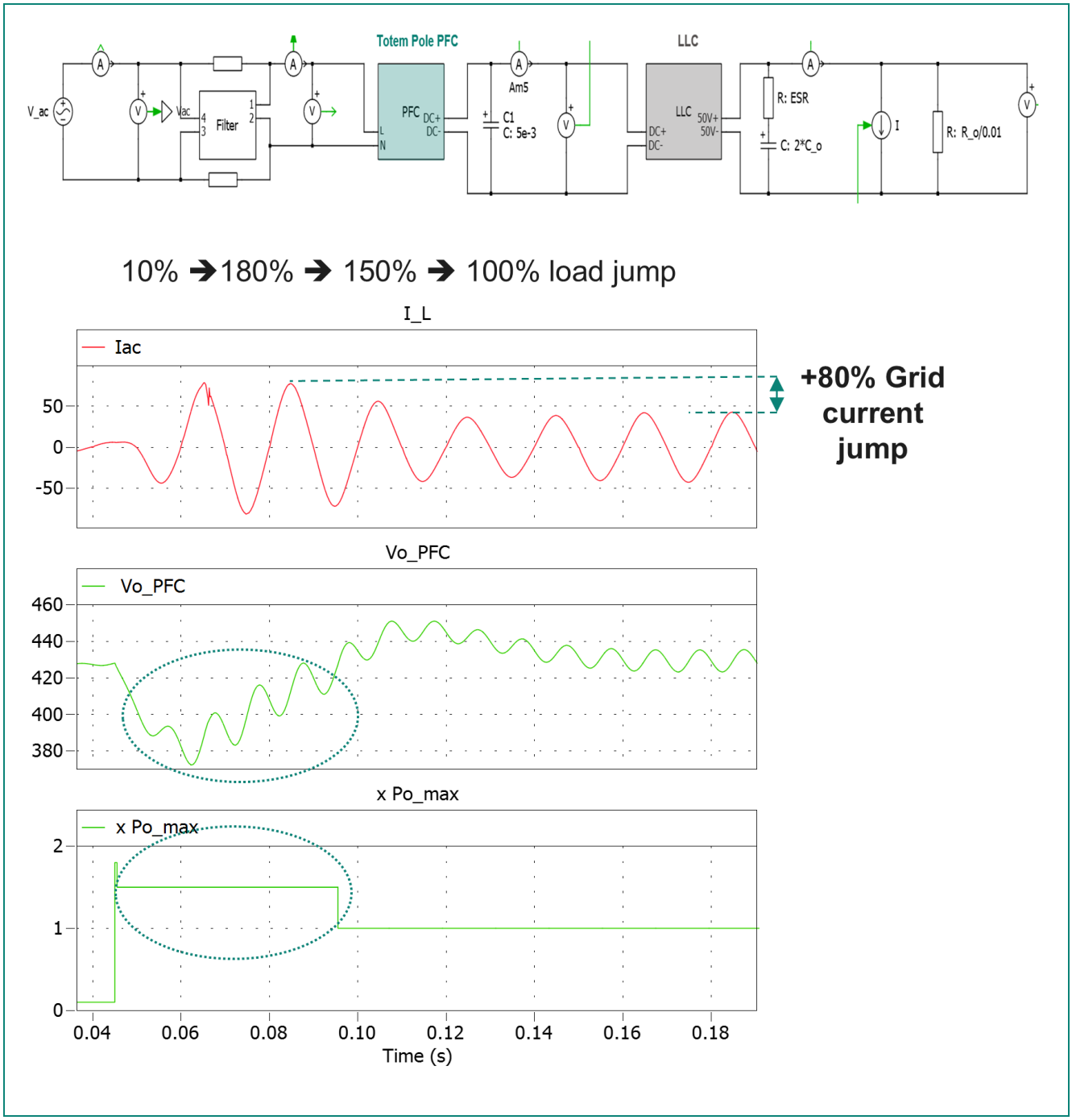

Figure 3 Simulation of a traditional 2-stage PSU during peak load event

To limit grid current transients, the sidecar's input energy must be buffered from the AI peak loads, which is the main function of the CBU shown in Figure 4. A CBU consists of a capacitor bank with a bidirectional DC-DC converter to discharge the capacitors' energy to the DC bus during AI peak load, and then charge the capacitors during the GPU's idle times, which will smooth out the peak power demand from PSUs and therefore the grid side.

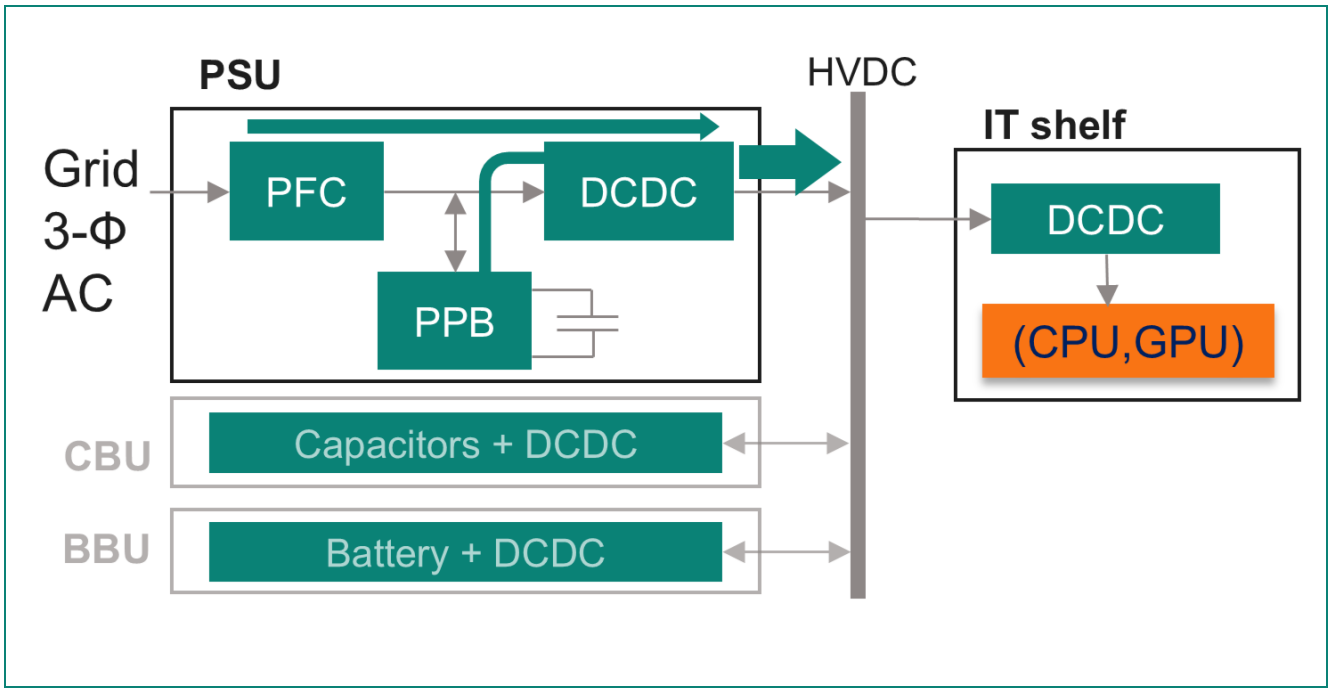

Figure 4 Simulation of a traditional 2-stage PSU during peak load event

Although CBUs can offer energy storage needed for peak shaving, they need separate shelves, and consume valuable space in the rack. Additionally, they may need to be compatible with, and to communicate with the PSU shelves, to manage rack power. Alternatively, energy buffer circuits can be integrated inside the PSU to offer the same functionality as the CBU as shown in Figure 4, allowing the rack to eliminate or reduce the need of extra capacitor shelves. Energy buffer circuits can increase the energy utilization in the PSU bulk capacitor by allowing it to discharge deeper while maintaining regulation of the bus voltage without affecting the DC-DC stage design, as detailed in the following section.

2 Power Pulsation Buffer (PPB) to support GPU peak load profiles

As GPUs draw a higher peak power and generate high load transients, the Power Pulsation Buffer functions as an efficient, high-density energy storage between the PFC and the DC-DC stages, and can support peak loads without overloading the upstream grid. PPBs can be integrated inside single-phase or three-phase PSUs, and can have different implementations, some of which are shown in Figure 5. The PPB bridge circuit can be 2-level or 3-level, and they can also be in series or in parallel. These bridge switches can be rated 440 V, 650 V, or 1200 V, designed with SiC, depending on the circuit type and the bus voltage.

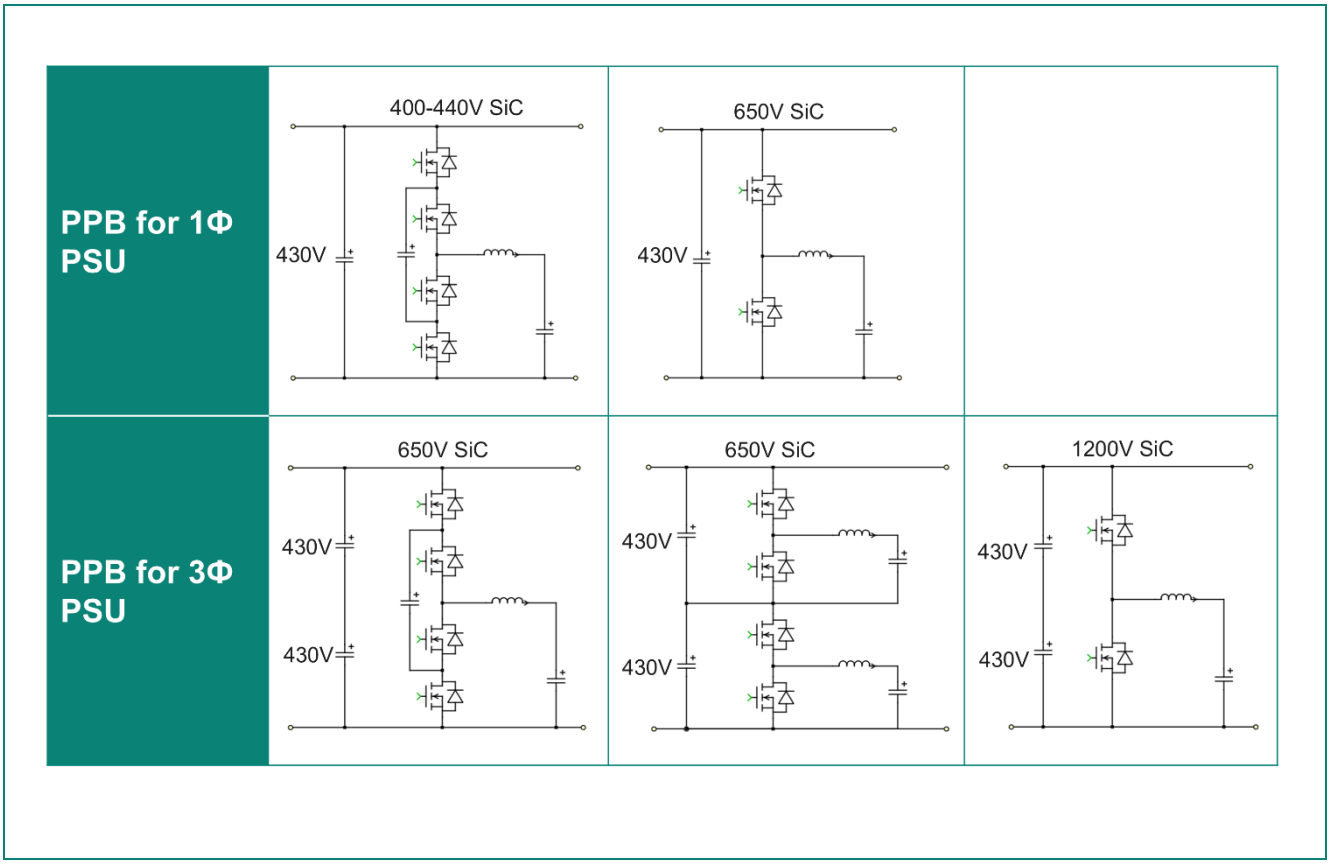

Figure 5 Different implementations of Power Pulsating Buffer in 1-ph and 3-ph PSU

Figure 6 shows an example 3-ph HVDC PSU with interleaved 3-level PPB. It shows a PLECs simulation of a peak-power event; with and without the PPB. From the figure, it is clear how the PPB provides energy during the peak-load event while maintaining the grid power limited to 110% of the PSU rated power.

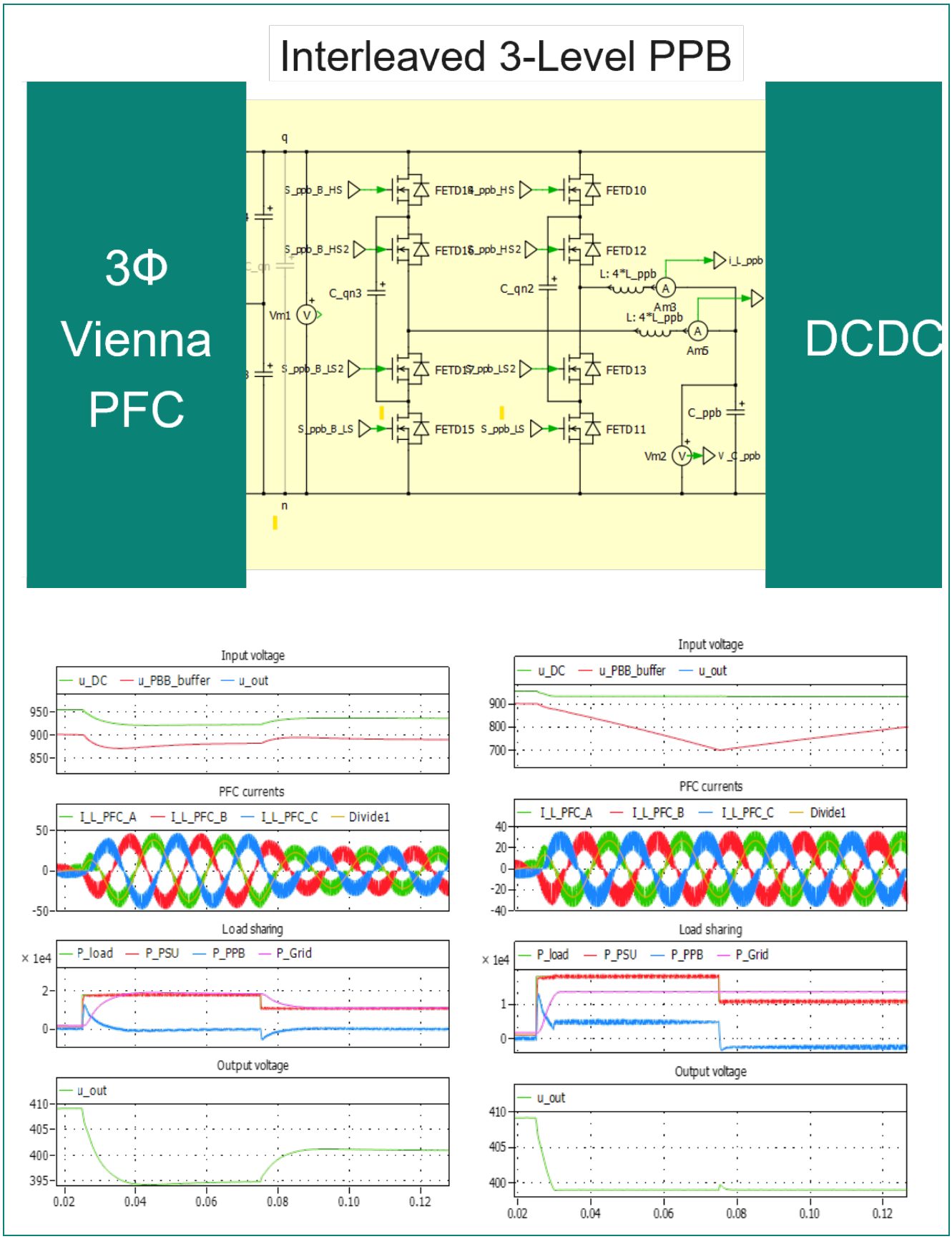

Figure 6 Simulation of peak load event: Without PPB (left) and with PPB (right) in 3-ph HVDC PSU

3 Control and operation of PPB

PPB can be operated continuously during normal operation, exchanging energy between the storage capacitor and the bus capacitor. Alternatively, it can be idle in normal operation and only operating on-demand during peak power events. Table 1 shows a comparison of both these PPB operation options.

Table 1 Comparison of PPB operation options

PPB on-demand operation	PPB continuous operation
– Good utilization of stored energy – Can support load jumps – Active only during hold-up and peak power ➔ no additional losses of PPB – Components sized for short event operation – No grid frequency cancellation ➔ DCX not possible, LLC must regulate small range of Vin fluctuation – Used for third harmonic injection in 3-ph PFC – Stable input and output voltages – Controlled re-rush current	– Good utilization of stored energy – Can support load jumps – Active all the time ➔ additional losses of PPB operation – Components need to be sized for steady state operation – Grid frequency cancellation ➔ allows for optimum DCX with fixed frequency and high efficiency – Grid frequency cancellation is more suitable for 1-ph PSUs, but cannot be used for third harmonic injection in 3-ph PFC – Stable input and output voltages – Controlled re-rush current

Learn more about how Power Pulsation Buffer can improve SMPS performance via Infineon's PSU reference designs at www.infineon.com/aipsu.

This article was originally published in Bodo’s Power and has been reproduced here with permission

Published by
Infineon Technologies AG
Am Campeon 1-15, 85579 Neubiberg
Germany

Public

Version: V1.0_EN
Date: 12/2025

Stay connected!

Scan QR code and explore offering
www.infineon.com

Please note!
This Document is for information purposes only and any information given herein shall in no event be regarded as a warranty, guarantee or description of any functionality, conditions and/or quality of our products or any suitability for a particular purpose. With regard to the technical specifications of our products, we kindly ask you to refer to the relevant product data sheets provided by us. Our customers and their technical departments are required to evaluate the suitability of our products for the intended application.

We reserve the right to change this document and/or the information given herein at any time.

Additional information
For further information on technologies, our products, the application of our products, delivery terms and conditions and/or prices, please contact your nearest Infineon Technologies office (www.infineon.com).

Warnings
Due to technical requirements, our products may contain dangerous substances. For information on the types in question, please contact your nearest Infineon Technologies office.

Except as otherwise explicitly approved by us in a written document signed by authorized representatives of Infineon Technologies, our products may not be used in any life-endangering applications, including but not limited to medical, nuclear, military, life-critical or any other applications where a failure of the product or any consequences of the use thereof can result in personal injury.