48V Factorized Power Architecture Slashes Thermal Losses and Outperforms Multiphase Approach when Powering AI

As AI processors demand ever more power at ever lower voltages, conventional power delivery architectures are reaching their limits. This article explores how a 48 V factorized approach can reduce losses while delivering high current directly to the processor.

By Don Gerstle, US-Correspondent, Bodo's Power Systems

There are two critical top-level questions facing generative AI processor power system engineers today.

First, when using $800V_{DC}$ rack power distribution, to what rack or compute tray-level intermediate voltage should $800V_{DC}$ be converted: 48V, 12V or 6V? In which direction or directions will power system architects choose?

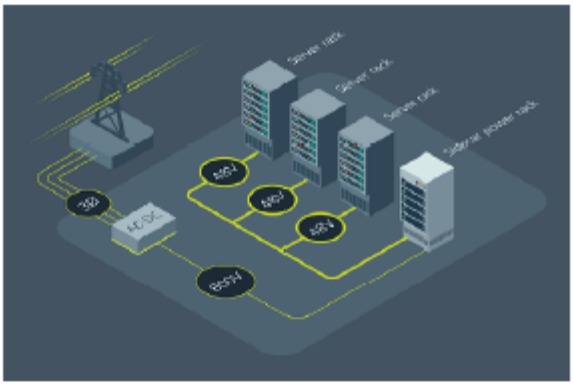


Figure 1: A conceptual illustration of 800VDC distributed to a sidecar rack, with multiple 48VDC feeds to AI server compute trays. This is one direction the AI hardware infrastructure industry might go over the next few years.

Second, what constitutes the most performant vertical power delivery (VPD) point-of-load (PoL) final-stage DC-DC converter architecture? This question is under intense technical debate among AI power system engineers.

AI processor tray high-voltage and low-voltage power architecture options

With respect to the $800V_{DC}$ to intermediate voltage distributed to the processor compute, storage, and network trays, there are several options:

- 800V to 50V (48V nominal) using 16:1 fixed-ratio bus converters
- 800V to 12.5V (12V nominal) using 64:1 fixed-ratio bus converters
- Note 12V intermediate bus and 12V multiphase VRs are rapidly becoming obsolete as suppliers shift to 6V seeking higher current density and improved low-ripple load-step transient response
- 800V to 6.25V (6V nominal) using 128:1 fixed-ratio conversion producing a 6V intermediate bus

There are also several technical solutions for VPD PoL DC-DC converter architectures:

- 48V direct-to-sub-1V using Factorized Power Architecture (FPA™) with high 48:1 current multiplication or "current gain" and high >3A/mm² current density
- 6V to sub-1V (approximately 12:1 regulation ratio) intermediate bus feeding "industry-standard" multiphase TLVR power modules
- 6V to 1.8V/2V to sub-1V using a 3:1 regulation ratio converter followed by a 2:1 regulation ratio converter; this is a potential architecture for new low-voltage, high-frequency multiphase buck regulators called IVRs (integrated voltage regulators)
- 48V to 2V to sub-1V utilizing a 24:1 bus converter followed by an IVR array

Working through the high-voltage and low-voltage power architecture options

There are multiple intertwined technical issues associated with continually-evolving AI processors beyond the dramatically-increasing power requirements, including reduced processor core voltages, tight transient-event "AC" voltage droop specifications, increased numbers of functional power domains operating at different voltage levels, system-level reliability expectations, and three-dimensional compute tray-spacing and cooling constraints. Even with vertical power delivery, rack inter-tray spacing is being gradually reduced, packing more processing performance in each rack (model training FLOPS/m³ and model inferencing tokens/m3), constraining VPD power module height and liquid cooling-based thermal management. AI hardware infrastructure is rapidly becoming power delivery-limited, eclipsing the ambitious compute performance goals set by leading industry system developers.

As the power levels needed by AI processors have surged, hyperscale AI data centers are migrating to $\pm 400V_{DC}$ or $800V_{DC}$ distribution directly to "sidecar" racks adjacent to the compute and scale-up and scale-out networking racks, where this DC voltage is converted to 48V, 12V or 6V (Figure 1). This frees up space in the compute and networking cabinets for 19-inch rack units previously set aside for DC-DC conversion shelves that had been growing larger as power demands inexorably increased. The future-looking end-to-end utility grid-to-AI processor power chain architecture has not yet been nailed down as power electronics engineers debate which versions of these new architectures make the most engineering and economic sense.

A critical challenge is how to efficiently supply adequate power levels into the latest AI processors while complying with the daunting technical obstacles of electrical, thermal and mechanical requirements.

The Vicor 48V-centric Factorized Power Architecture (FPA™), developed over 30 years ago for powering high-current, low-voltage loads, has been continually innovated and improved to meet ever-increasing current density needs at the point of load. Powering AI processors and wafer-scale HyperProcessors is perhaps its most potent expression.

Critically, with FPA, Vicor moves 48V voltage regulation upstream, to provide tightly-regulated 48V directly to the downstream vertical power delivery (VPD) transformer, the VTM™ current multiplier positioned under the processor. This *regulation then transformation* arrangement is the functional opposite of the conventional multiphase *transformation then regulation* approach.

For VPD, 48V FPA brings high voltage, at lower currents as close to the processor as possible, solving the "last inch" challenge and greatly reducing intermediate bus thermal losses. Figure 2 shows the "regulation then transformation" functional partitioning of 48V FPA. In the simple example shown, the current gain (or transformation K factor) of 48 is used to transform 48V to a $1V_{DD}$ processor main power domain rail. The appropriate K factor can be selected from a range of modules to suit the processor voltage and current specifications.

When comparing 48V FPA versus industry-standard 6V multiphase power architecture (Figure 3) for a conceptual 3KW TDP AI processor, power losses are cut significantly using 48V FPA.

A longer-term trend is the migration to wafer-scale and panel-scale AI processors built on glass-core substrates. The panel area sizes projected by advanced packaging eco-

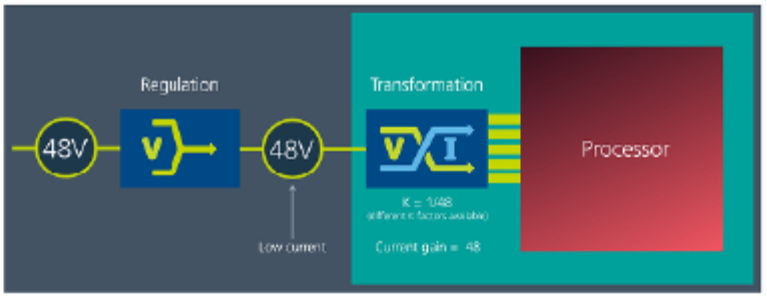


Figure 2: The Vicor Factorized Power Architecture brings 48V to the point-of-load current multiplier, which transforms this relatively high voltage level to the sub-1V high current levels required by contemporary AI processors.

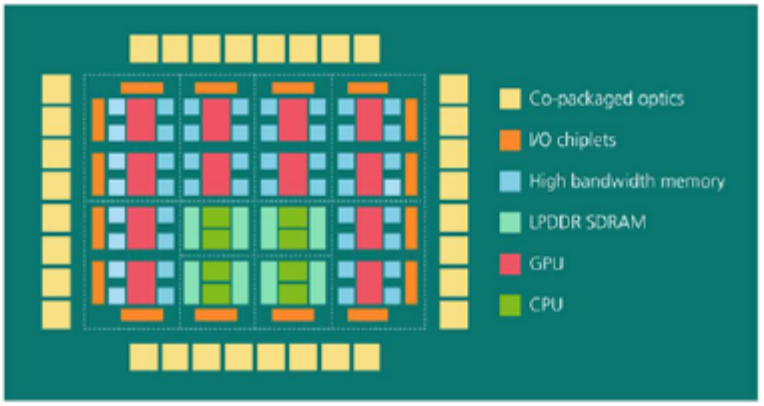
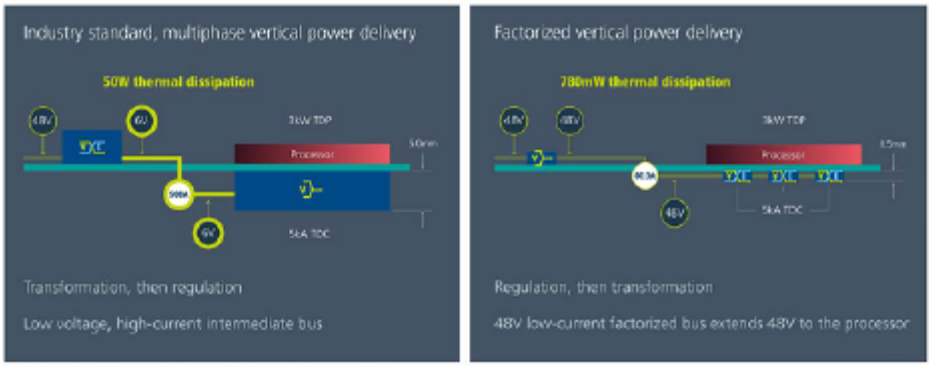


Figure 4: A conceptual late-decade 200 x 200mm panel-scale AI HyperProcessor with a mix of GPU and CPU chiplets, along with supporting memory and I/O chiplets, as well as high-bandwidth co-packaged optics (CPO) modules. These next-generation panel-scale HyperProcessors will require vertical power delivery, and 48V to the points-of-load to manage the thermal energy associated with the tens of thousands of amps they will continuously pull.

system developers range from 200 x 200mm all the way to 600 x 600mm. Figure 4 shows a conceptual panel-scale AI processor with 12 GPUs, eight CPUs, and high-bandwidth co-package optics.

This trend towards panel-scale chiplet-based AI processing engines is driven by the relentless pursuit of higher rack-level compute density. These large panels will support hundreds of reticle-sized chiplets, blowing past the dozens of reticle-sized chiplets supported by existing CoWoS and EMIB package types. The power demands projected for these wafer-level and panel-level AI processors necessitate vertical power delivery and will greatly benefit from 48V FPA, using high current density and high current gain, low-profile thermally-agile power modules. High-performance VPD requires both current density and current gain. Vicor 48V-centric Factorized Power Architecture with current multiplication reduces losses and has significant advantages in that it has both of these key attributes of high current density and high current gain.

Special thanks to Maury Wood, VP of Strategic Marketing, Vicor Corporation, for his support and consultation in writing this article.



Both approaches assume 200µΩ bus resistance. Simplified models

Figure 3: Industry-standard 6V multiphase VPD versus 48V FPA VPD with current multiplication. 48V FPA reduces intermediate bus thermal losses from 50W to 780mW when applied to an example 3KW TDP / 5KA TDC AI processor.

www.vicorpower.com