Powering AI Processors – 48 V Vertical Power Delivery Changes the Game

AI accelerators are pushing power requirements beyond what many legacy power delivery architectures were designed for. Bodo's Power Systems asked Maury Wood, VP of Strategic Marketing, Vicor, about the future of conventional multiphase voltage regulators, 48 V designs, cooling and more.

By Alfred Vollmer, Editor in Chief, Bodo's Power Systems



Alfred Vollmer (AV): What has changed in AI workloads in recent years which makes power delivery such a challenge today?

Maury Wood: Training and inferencing workloads have become more computationally intense as the number of model parameters has grown with frontier LLMs. The shift towards agentic AI and an increasing focus on token generation throughput increase the need for process orchestration. These factors have ballooned AI algorithmic compute workloads.

Additionally, AI data center operators continually seek higher rack compute density and higher power utilization effectiveness. These expectations drive towards higher end-to-end power delivery efficiency, which means decreased thermal losses along each segment of the power chain. The importance of thermal loss reduction has put pressure on power delivery architectures, evident in recent AI power electronics innovations.

AV: Conventional multiphase voltage regulator designs have served the industry well for years. Where do you see their biggest technical limitations when supporting today's high-current AI GPUs?

Maury Wood: Multiphase TLVR power architectures require an increasing number of current delivery phases to source the current demands of recently announced AI processors. With a maximum continuous output current of about 65 amps per phase, five dozen phases are need to supply the 3,800 A, a typical example of upcoming GPU-based AI processors. In a vertical power delivery arrangement, the required multiphase power modules must coexist with the decoupling capacitors on the back-side of the accelerator PCB. This design challenge is strongly tied to the low current density – measured in A/mm2 – of competitive multiphase power modules. The height, i.e. the thickness, of multiphase power modules is an impediment to the primary goal of higher rack-level compute density, which requires decreased compute sled vertical spacing.

AV: Vicor has been advocating 48 V power delivery and FPA, Factorized Power Architecture, for some time. Can you explain how this architecture differs from traditional board-level power distribution schemes, and why it delivers measurable improvements in efficiency and power density, particularly with consideration to AI data centers?

Maury Wood: Traditional multiphase TLVR power architectures convert 48 V to 6 V, then convert 6V to 0.6 VDD; this 6 V intermediate bus carries high levels of current, due to the relatively low voltage across the thick copper PCB planes. The unique 48 V Factorized Power Architecture (FPA) flips on its head the multiphase "transformation then regulation" power chain arrangement. By contrast, 48 V FPA uses a "regulation then transformation" functional arrangement (48 V to 48 V regulator, then 48:1 current multiplication) to bring 48 V as close as possible to the point-of-load (PoL). The thermal dissipation savings associated with bringing 48 V to the vertically-placed current multiplier compared to bringing 6 V to the PoL multiphase TLVR buck regulators is about a factor of 64.

Because AI processors consume about 50 % of total AI datacenter power, this approximate 50 W thermal reduction multiplied by the tens of thousands of compute trays in contemporary AI data centers results in a very significant energy and cost savings, on the order of 1 % or more in terms of top-level data center PUE. In a 1 GW AI data center, 1 % is 10 MW, significantly moving industry-wide imperatives in the right direction.

AV: Vertical power delivery, VPD, is receiving significant industry attention. What are the most important system-level benefits of VPD?

Maury Wood: VPD reduces the conduction distance from the final-stage DC/DC converter to the AI processor, reducing the vertical PDN impedance, and thus the vertical bus thermal dissipation, what we call the "last centimeter" problem.

Thin, low-thermal-impedance VPD power modules enable consideration of back-side liquid cooling and support the goal of increased compute density at the rack-level. In order to co-exist with essential PCB back-side decoupling capacitors, VPD power modules should ideally offer high output current, high current density, and high levels of current multiplication (or current gain).

AV: As processor power continues to increase toward multi-kilowatt levels, what new challenges emerge in power integrity and design, and how should power engineers begin preparing for these next-generation systems?

Maury Wood: AI algorithmic workloads are notoriously dynamic, with transitions from the idle processing state to fully active processing state in which dozens of functional logic units are energized within microseconds. The active power delivery architecture must possess high transient response combined with an exceptionally agile control loop to limit output voltage droop to within the tolerance limits specified for the processor.

AV: Data center operators ultimately care about total cost of ownership as much or more than electrical efficiency. How do improvements in power conversion efficiency translate into tangible savings in cooling infrastructure, rack density and operational costs at the AI factory scale?

Maury Wood: Any thermal power reductions in AI compute racks propagate backwards to the power shelf sidecar racks, the ±400 V or 800 V cabling, the three-phase AC/DC conversion subsystem, the utility step-down transformers, all the way back to the very high-voltage utility grid. The 1 % PUE improvement possible with 48 V FPA and the associated lowered I2R losses positively impacts the sizing of the liquid cooling infrastructure.

AV: Many organisations already have established 12 V and 6 V-based AI hardware infrastructure. What practical considerations should engineers evaluate when transitioning to 48 V AI accelerator power architectures, and what barriers still exist to broader industry adoption?

Maury Wood: A 6 V input allows the use of lower voltage MOSFETs in the PoL multiphase DC/DC converters, marginally increasing conversion efficiency. However, this conversion efficiency increase is more than offset by the jump in I2R heat generated on the 6 V intermediate bus – 6 V doubles the current on this bus compared to 12 V, and when squared, the heat on this bus quadruples. From the Vicor perspective, this is running in the wrong direction with respect to achieving the highest efficiency power delivery to AI processors.

Specifically, a 48 V factorized bus reduces heat generation on this segment by a factor of 64 compared to a 6 V intermediate bus without impacting DC/DC conversion efficiency. This is the fundamental value proposition of 48 V factorized power architecture. Powering AI processors is an excellent application to realize the benefits of 48 V FPA.

AV: Looking ahead over the next five years, where do you believe the biggest innovations in AI power delivery will occur? Are we approaching another architectural shift beyond today's 48 V approaches, or is there still significant headroom within current technologies?

Maury Wood: The use of 48 V current multipliers using a 48:1 current gain or "K" factor (similar to a transformer turns ratio) avoids the use of a 6 V intermediate bus, improving end-to-end power delivery efficiency by up to 10%. End-to-end means 800 V rack distribution to 1 VDD or lower PoL; therefore, this makes a very significant impact on data-center-level power utilization effectiveness.

A longer term trend is the migration to wafer-scale and panel-scale AI processors build on glass core substrates. The panel area sizes projected by advanced packaging ecosystem developers range from 200 mm by 200 mm all the way to 600 mm by 600 mm. These large panels will support hundreds of reticle-sized chiplets, blowing past the half dozen 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 48 V factorized power architecture, using high-current-density and high-current-gain, low-profile thermally-agile power modules.

www.vicorpower.com