WHITE PAPER

From 48 V direct to Intel VR12.0: Saving 'Big Data' $500,000 per data center, per year

Written by: Stephen Oliver
Vice President, VI Chip Product Line
July 2012

Abstract

The exponential rise in 'Big Data' generation, processing and storage, from sources such as intensive industrial simulations, medical research and social media sites highlights the growing demand for data center and cloud computing power worldwide. The subsequent task is to maximize energy efficiency, thus saving money and natural energy resources, minimizing pollution and meeting the US Department of Energy's 'Exascale' challenge; perform an ExaFLOP (1 x 10^18 floating point operations per second) of compute workload using only 20 MW of power.

For the power semiconductor industry, the challenge is to provide efficient and 'high quality' power conversion from 480 V 3-phase AC power entering the data center, all the way to 1 V, 100 A+ processors. Here, 'high quality' means providing the optimum voltage and current as dynamically demanded by the processor to operate at peak performance.

This paper describes how Vicor's Factorized Power Architecture® (FPA®) enables the world's first Intel® VR12.0 compliant solution to operate directly from 48 V distribution systems, saving $500,000 per data center in annual running costs.

Introduction

High-volume data processing and data communication tasks require large-scale dedicated, optimized hardware installations – i.e. data centers – designed and operated by companies such as IBM®, Amazon®, Cisco®, Hewlett-Packard®, Google®, Cray® and others. Within the data center infrastructure, classical physics principles dictate that power should be channeled to data processing locations at a high voltage to reduce current and so minimize distribution loss[a].

Compute density (related to the number of processors, memory, input / output functionality) and subsequent power draw is also a key issue. As power per rack increases above 10 kW, loss in traditional 12 V rack distribution becomes excessive, with additional financial- and size-related costs such as larger, more expensive copper bus bars, connectors, etc. For applications approaching 20 kW and above per rack, efficient 48 V distribution is required. An example is the POWER7-based IBM Blue Gene®/Q, which uses 48 V distribution, 80 kW per rack and is the world's highest performance[b] and most efficient[c] supercomputer, achieving 20,132 TFLOPs and 2,026 MFLOPs/W. In addition, high reliability or high 'uptime' servers require an energy storage system (typically lead-acid batteries) to provide back-up power in the event of a main AC-feed interruption.

Historically, processing tasks such as switching and routing in the telecommunications industry were performed by small 30 – 40 W application-specific integrated circuits (ASICs). With the rise of 'triple-play' (voice, video, internet) usage on mobile phones and tablets, voice-only 'telecom' equipment has become 'datacom', with the use of standard computing processors. These systems typically use a 48 V rail but in 'Central Office' equipment, the battery back-up voltage may be much wider, requiring wider input-range converters, further reducing the system efficiency. Energy storage is also related to voltage, so 48 V is again the optimum choice[d].

---

[a] Power = voltage x current. So for the same power, as voltage increases, current decreases.
Distribution loss = current^2 x distribution rail resistance. So, distribution at 48 V is 16 x more efficient than at 12 V.

[b] 'Top500' list June 2012, http://www.top500.org/list/2012/06/100

[c] 'Green500' list November 2011, http://www.green500.org/lists/2011/11/top/list.php

[d] In battery and capacitor technologies, energy storage is related to voltage^2.

Voltage Regulation

Compute workload varies with time. As demand increases, each processor requires more power to maintain performance (MFLOPs). As workload is reduced, the processor may throttle back, moving to idle or sleep states to conserve power. In anticipation of a change in power requirement, the processor sends a serial 'Voltage Identification' (VID) code to the power delivery system. During all steady-state and transient periods, the voltage delivered to the processor must remain within tight, pre-defined limits to maximize performance and minimize the chance of a system crash.

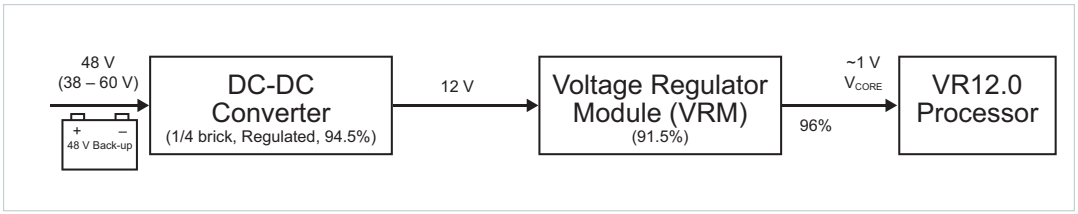
Power requirements for Intel processors are identified in Voltage Regulator (VR) specifications. For the Sandy Bridge and Ivy Bridge processor chip-sets, compliance to the VR12.0 specification is required.

Traditional 12 V Limitations

As power per rack increases with a 48 V rack distribution requirement, traditional 12 V-VR12.0 solutions must be prefaced by a separate 48 V – 12 V conversion stage, as shown in Figure 1, below.

Figure 1.

Traditional high power server architecture showing two stage 48 – 12 – 1 V approach, showing the wide server range 48 V (38 – 60 V) with back-up energy storage at 48 V. The '96%' represents the loss in the motherboard due to the VRM being large and unable to be located adjacent to the processor socket.



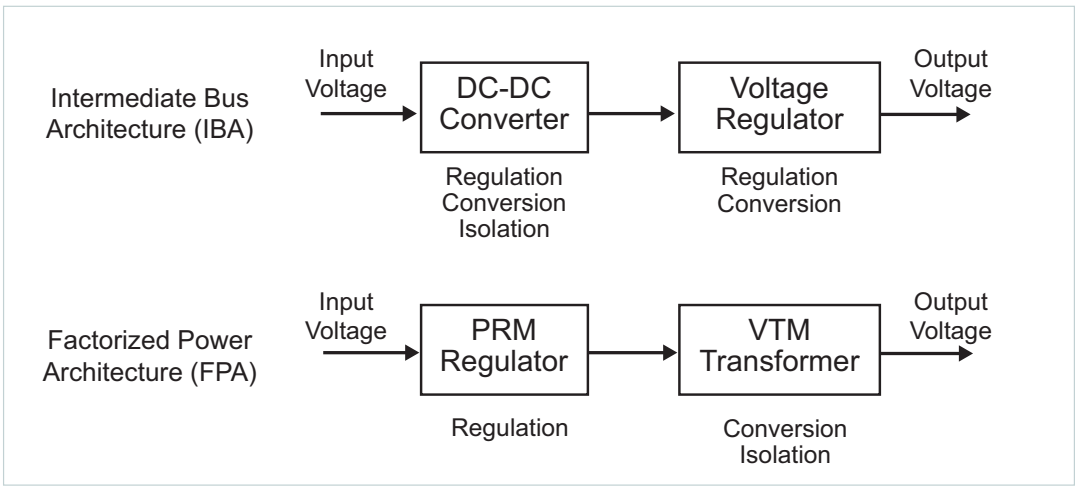
It is critical to consider the power scheme in its entirety. 'Headline' specification claims for DC-DC converters and VRMs ignore losses related to distribution loss on the motherboard and any connector losses. A true approach requires a measurement from the 48 V rail all the way to the processor socket itself, thus taking into account all potential loss elements.

Optimized 48 V Solutions: Factorized Power Architecture® (FPA®)

Factorized Power Architecture (FPA)[e] employs a different approach to power conversion, taking the regulation, isolation and voltage transformation functions of a typical DC-DC converter and separating or 'factorizing' them into individual elements. These individual components (small, high-efficiency regulators and transformer / isolators known as 'VI Chips®') are then arranged in the optimal power architecture.

Figure 2:

In traditional IBA, functions are duplicated thus increasing losses. FPA eliminates the duplication, increasing efficiency.

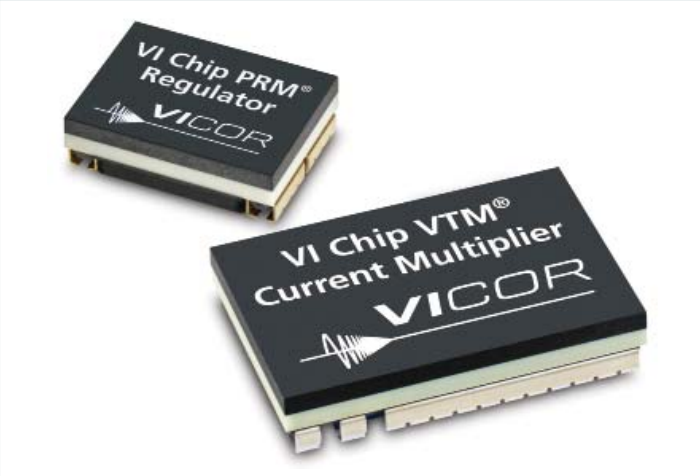


[e] For additional information, please refer to white paper 'FPA 101' available at www.vicorpower.com

VI Chip PRM® regulators use a non-isolated buck-boost topology. The PRM accepts a varying DC input and creates a tightly regulated, adjustable DC output (the 'factorized bus' (V_F)) which feeds into a downstream VTM® transformer. The VTM is a fixed-ratio DC-DC transformer using a 'Sine Amplitude Converter®' (SAC®), which down-converts V_F directly to the processor's core voltage V_{CORE}. Thanks to zero-voltage (ZVS) and zero-current (ZCS) MHz-switching techniques, high-efficiency and high-power densities are achieved, with the PRM[f] up to 97% peak and more than 1,000 W/in^3, and the VTM more than 94% peak and over 100 A/in^2.

Figure 3.

VI Chip PRM and VTM power components used in FPA



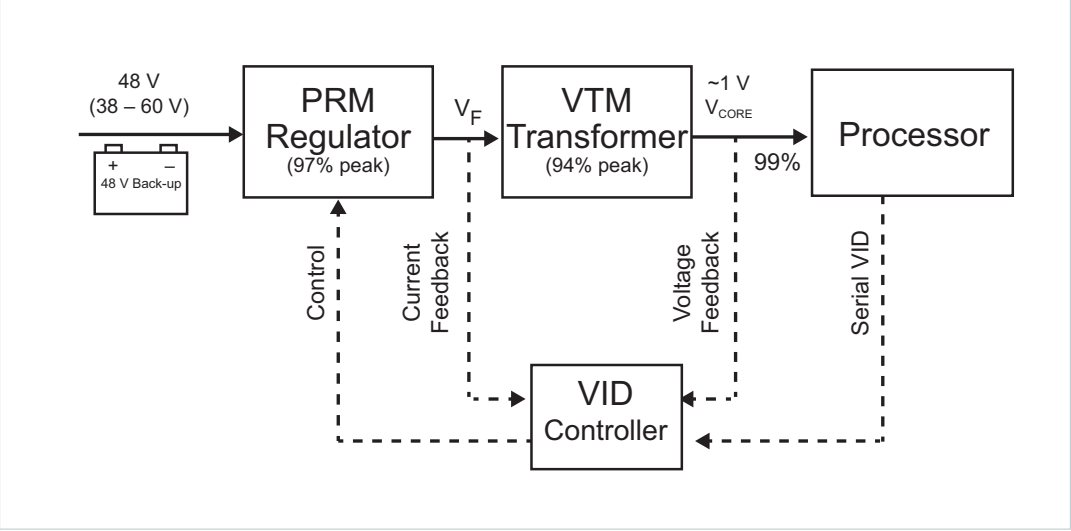
With FPA, the power system is architected to maintain high efficiency 48 V distribution along the entire path to the motherboard and to use the PRM and VTM VI Chips adjacent to the processor socket. The result is a highly efficient, small power system with a proven record in high performance systems including the 48 –1 V Blue Gene/Q system referenced earlier.

World's First 48 V-VR12.0 Solution

For Intel processor systems, the VI Chips form a 'pure' powertrain which is accompanied by a separate VID controller, which acts as a translator between the digital processor VID and the FPA powertrain, which, in turn uses the optimal fast analog control loop to provide accurate processor core voltage (V_{CORE}).

Figure 4.

Information from the processor (VID instructions) and feedback on current and voltage are fed to the VID control IC. The IC sends a control signal to the PRM regulator which controls the powertrain.



[f] VI Chip components used in the VR12.0 example are PRM48DH480T250A03 and VTM48EF012T130A01, available from www.vicorpower.com

To demonstrate VR12.0 compliance, a Voltage Regulator Test Board (VTRB) was created as shown in Figure 5 and configured to support a 145 W, 'socket R' processor.

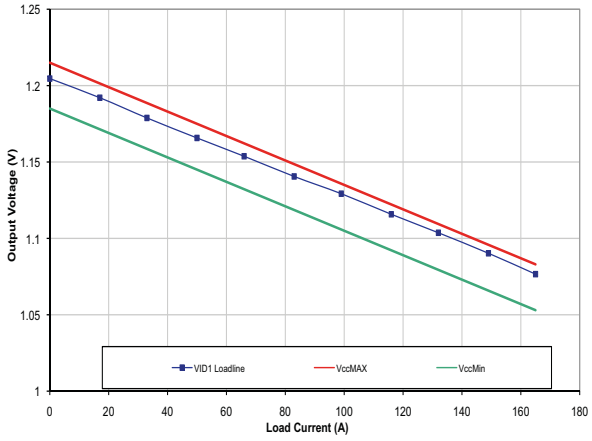
Figure 5.

VTRB with close-up showing PRM and VTM modules adjacent to processor socket. No separate 48 – 12 V converter is required for V_{CORE}. Note also that no bulky electrolytic capacitors are present in the FPA design.

A Voltage Test Tool (VTT) is inserted into the processor socket which emulates processor behavior to characterize the performance of the powertrain. The performance is recorded and compared to the VR12.0 specifications using an automated spreadsheet. A typical 'Load Line' result, in which V_{CORE} must remain within tight limits while processor current increases, is shown in Figure 6.

Figure 6:

As processor current demand increases; power cannot be instantly increased, so the core voltage is allowed to droop in relation to the load. This is known as a 'load line' and the core voltage must be maintained within tight limits to ensure processor stability and performance. The FPA result (blue line) is shown in the load increase from zero to 165 A.



Figures 7 and 8 show examples of voltage and current oscilloscope traces during power state transitions.

Figure 7.

Load transient waveforms (Power State 1 to Power State 0 transition (16 to 147 A) shown at two different time scales. The FPA system has a clean, stable response within 5 µs. In the scope images, yellow = current step, magenta = load voltage measured in VTT with only SMT ceramic caps. The FPA system does not use traditional large, unreliable electrolytic capacitors.

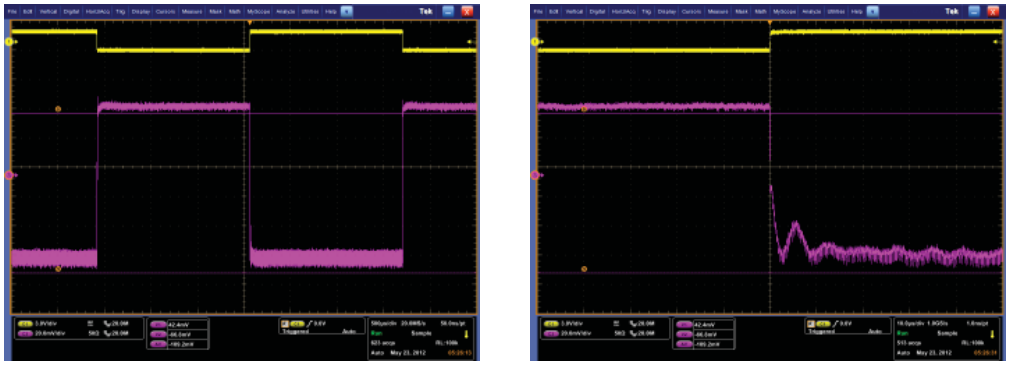


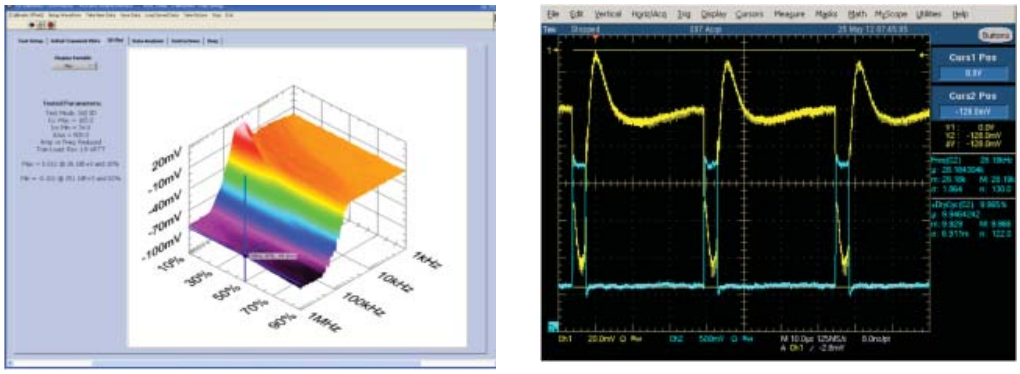
Figure 8:

Shows the core voltage response to a VID instruction (Dvid_ps0_121A_up (1.05 V to 1.07 V)). The magenta trace = V_{CORE} (output voltage), cyan = CSO (chip select), green = alert. The results show a stable response in only 2 µs.

Additionally, more stringent tests, using 'sweeps' of processor frequency, load power and VID commands are also performed. '3D Matrix' plots are created, with an example in Figure 9.

Figure 9:

3D Matrix result at Intel Dupont. Chart axes are: X = load (processor) switching frequency (kHz), y = Load switching duty cycle (%), Z axis is maximum excursion above set point.



Savings

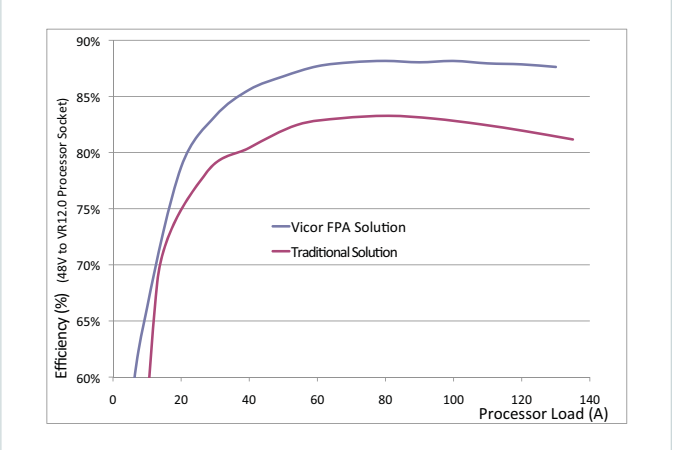
A complete evaluation of the powertrains showed significant size and efficiency savings for FPA vs. the traditional IBA system.

Server motherboard real estate (PCB area) is expensive. A reduction in powertrain size means an increase in space available for more compute functionality (processors, memory, input / output functions). The FPA solution uses 50% less motherboard than IBA, while in addition, eliminates the off-board DC-DC converter; in total, a 3 x reduction in size.

In power terms, the FPA solution is >5% points more efficient from 60% to 100% processor load as shown in Figure. 10. This is a significant improvement in VR12.0 systems, representing a 10 W or 30% reduction in power loss per processor.

Figure 10.

48 V to VR12.0 socket efficiency (accounts for conversion and distribution losses).



Adjusting for duty cycle / usage rate (85%) and air-conditioning costs (+70%), the final value per processor is 14.5 W saved.

A new-build data center typically uses 30,000 processors. Using a $0.13/kWhr price for electricity, 14.5 W quickly becomes $500,000 reduction per year in data center running costs (equivalent to 2,300 imported barrels of oil). The annual saving means that in less than three years, the VI Chip VR12.0 powertrain completely pays for itself.