

Hot-Swapping Analysis and Implementation of Series-Stacked Server Power Delivery Architectures

Enver Candan, *Student Member, IEEE*, Derek Heeger, Pradeep S. Shenoy, *Member, IEEE*,
and Robert C. N. Pilawa-Podgurski, *Member, IEEE*

Abstract—Current data center power delivery architectures consist of many cascaded power conversion stages, where the system-level power conversion efficiency is reduced each time the power is processed through the individual stages. Recently, series-stacked power delivery architectures have shown how the overall power conversion can be reduced through architectural changes, reporting above 99% system-level power conversion efficiencies for data centers. In this paper, we contribute to the development of the series-stacked power delivery architectures by addressing the important hot-swapping challenge, without sacrificing the high power conversion efficiency. We analyze the hot-swapped operation of the series-stacked architecture, and experimentally validate it on a testbed that includes four series-connected 12 V, 120 W servers and four custom-designed differential converters with associated circuitry for hot-swapping. The results show that continuous operation of the series-stacked servers can be maintained while a server is hot-swapped without a significant reduction in the high power conversion efficiency.

Index Terms—Bidirectional power flow, dc–dc power converters, differential power processing, high efficiency data center power delivery, hot-swapping.

I. INTRODUCTION

THE continued expansion of IT and cloud services demands a significant growth in both size and capacity of data centers, increasing the need for high energy efficiency. The energy consumption of US data centers has been estimated at 91 billion kWh in 2013 [1] and shows no sign of diminishing. A data center survey projected that the power rating of each IT equipment cabinet will be increased to 50 kW by 2025 [2]. As data processing and storage continue to grow, efficient power conversion for data centers becomes an important research topic.

The IT workhorses in data centers are servers, which operate at a low (typically 12 V [3] or 48 V[4]) dc voltage. Several ac

and dc power delivery architectures are established to provide low dc voltage from the ac grid to the servers. The ac power can be directly rectified and distributed through high (380–400 V) dc voltage within the data center; or the power can be distributed in ac form, and rectified at the rack-level or by the power supply of the individual servers. Regardless of the power delivery architecture, the utility power must pass through cascaded power stages in order to be stepped down to the low dc voltage. Therefore, much work has focused on improving the efficiency of the power stages. Consequently, mid/high-90% efficiencies have been achieved in rectifiers [5]–[7], in both silicon-based [8]–[11] and GaN-based [12]–[14] dc–dc converters, and in uninterruptible power supplies (UPS) [15] for data center applications.

Figure 1 illustrates conventional ac and dc power delivery architectures for data centers. In the ac architectures, as shown in Fig. 1(a)–(c), following the ac power distribution inside the data center, either an ac–dc converter per server performs rectification and step-down to low dc voltage at the server input (as in Fig. 1(a)), or a rack-level ac–dc converter provides dc voltage to bus bars (as in Fig. 1(b) or (c)). Then, each server either directly connects to the dc bus bars (as in Fig. 1(b)), or may require another voltage step-down stage at its input (as in Fig. 1(c)), depending on the bus voltage level. On the other hand, in the dc power delivery architecture as shown in Fig. 1(d), a central rectifier provides high dc voltage which is distributed to the server racks throughout the data center. Then, a dc–dc converter per server steps down the high dc voltage to the low voltage at the server input. A common feature of these power delivery architectures is that the system-level power conversion efficiency is limited by the efficiency of the power converters, since the *full* server power has to be processed in order to be delivered. Because of the correlation between the delivered and processed power in these architectures, the power conversion losses increase as both the rated server power and the number of servers increase.

Recently, series-stacked power delivery architectures have been explored to separate the processed power from the power delivered by processing only the *difference* in power between a group of series-connected loads [16]–[19]. In these architectures, the servers [16], [17] or processors [18], [19] are connected electrically in series and voltage regulation is performed by using differential power processing (DPP) techniques. In [16] and [17], two different DPP architectures were presented and achieved higher than 99% system-level efficiencies while the servers are performing real-life data center operations such as

Manuscript received July 28, 2016; revised October 3, 2016; accepted November 28, 2016. Date of publication December 14, 2016; date of current version May 9, 2017. This work was supported in part by the National Science Foundation under Grant 1509815, in part by the Texas Instruments, and in part by the Google Faculty Research Award. This paper was presented in part at the ECCE Montreal, QC, Canada, September 21, 2015. Recommended for publication by Associate Editor V. Agarwal.

E. Candan and R. C. N. Pilawa-Podgurski are with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: candan2@illinois.edu; pilawa@illinois.edu).

D. Heeger is with Sandia National Labs in Albuquerque, Albuquerque, NM 87123 USA (e-mail: dsheeger@sandia.gov).

P. S. Shenoy is with the Texas Instruments, Dallas, TX 75243 USA (e-mail: pshenoy@ieee.org).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TPEL.2016.2639519

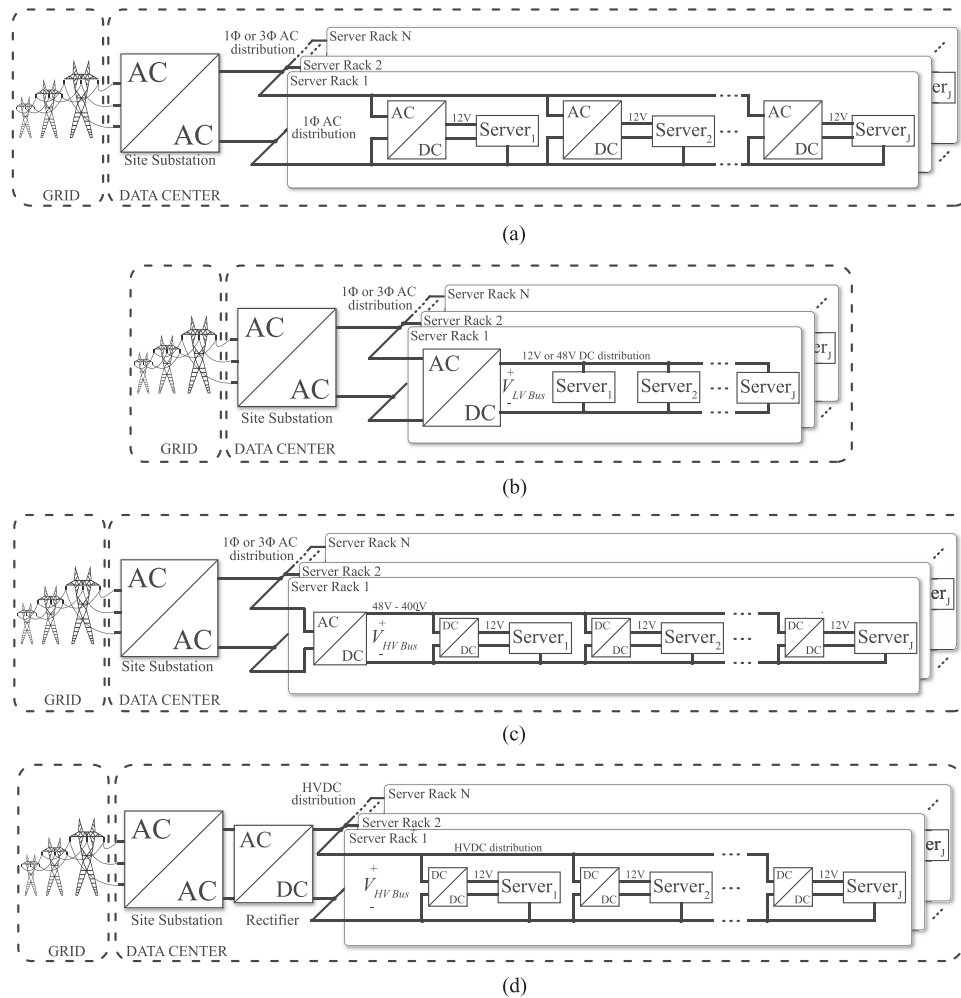


Fig. 1. A few examples of conventional power delivery architectures in data centers. For simplicity, UPS, transformers, and protection equipment are not shown in these figures, but are of course used in practical designs. (a) AC power distribution in data center and server racks. A single power converter provides rectification and step-down at the server input. (b) AC power distribution in data center. The ac power is delivered to the server racks. Then, a rack-level rectification stage provides low dc voltage to the servers through bus bars. (c) AC power distribution in data center. The ac power is delivered to the server racks. Then, a rack-level rectification stage provides an intermediate high dc voltage to the rack. Separate dc–dc converters perform voltage step-down at the server input. (d) DC power distribution in data center. A central rectifying stage provides a high dc voltage which is distributed to the server racks. Then, separate dc–dc converters step-down the high dc voltage at the server input.

web traffic management and computation. In [17], the presented DPP architecture is also compared to a conventional dc power delivery architecture similar to Fig. 1(c), showing up to 40 times reduction in the power conversion losses compared to state-of-the-art solutions during the steady-state operation of the series-stacked servers. Also, in [20] Candan *et al.* provided a reliability assessment of a series-stacked architecture for server power delivery, and showed that the reduction in average processed power may result in similar downtime over a 10 year period in comparison to a conventional architecture.

This paper is based on our previous conference publication [21] in which we provided preliminary experimental hot-swapping in the proposed series-stacked DPP architecture in [17]. In contrast to [17] and [21], this paper derives a mathematical expression for processed power in the proposed architecture, which is, to the best of our knowledge, the first derivation in the literature for power delivery to series-stacked loads

by the proposed architecture. In addition, this paper provides a conceptual explanation of the hot-swapping operation, assisted by the mathematical derivation. Finally, the experimental study in [21] is revised by employing higher power servers and a new DPP hardware that combines a differential converter and associated hot-swapping circuitries on a single board. Through the revised experimental study, we demonstrate the start-up and shutdown of the series-stacked servers, and the hot-swapped operation of a server from the series-stacked architecture while the other servers are still in operation. Start-up sequencing and hot-swapping of malfunctioning servers during maintenance are important practical requirements for data center operations. This work is the first to address these important requirements for the series-stacked architectures.

The remainder of this paper is organized as follows: The background information on series-stacking and the proposed DPP architecture is summarized in Section II. The processed

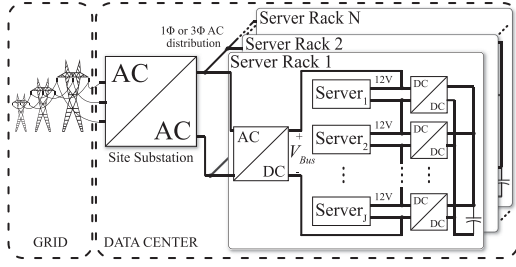


Fig. 2. Server-to-virtual bus DPP architecture for server racks.

power expression for the proposed architecture is derived in Section III. Section IV explains the conceptual hot-swapped operation and practical hot-swapping requirements of the proposed architecture. The DPP hardware to achieve stack initialization and hot-swapping is presented in Section V. In Section VI, the experimental testbed is described and the experimental hot-swapped operation is demonstrated. Finally, Section VII concludes this paper.

II. BACKGROUND ON SERIES-STACKING AND DPP

Low-voltage elements (sources or loads) are commonly connected in series to interface with high dc bus voltage in applications, such as photovoltaic sources, battery systems, and light-emitting diodes. In these applications, the elements can be connected in series because the desired operating voltage of each element is lower than the available or desired dc bus voltage. The large number of low-voltage servers in data centers represents a similar scenario where series connection within the server rack may be beneficial.

Previously, series-stacking and DPP has been proposed for dc systems where a group of low-voltage elements must be connected to a higher voltage dc bus [22] and has been applied to fields where the instantaneous power of the series-stacked elements are changing dynamically, such as photovoltaic sources [23] and digital circuits [24]. Two common topologies for DPP of the series-stacked power delivery architectures are the element-to-element and the element-to-virtual bus architectures, which have been applied to photovoltaic applications [25]–[32], battery chargers [33], [34] and computing loads [16], [17], [35].

Figure 2 shows a series-stacked architecture with a server-to-virtual bus DPP technique, which is the architecture used in this work. In this architecture, the servers are connected in series to interface with the high dc bus voltage and each server is connected to the primary side of a differential converter (i.e., a bidirectional isolated dc–dc converter). The secondary side of the differential converter is connected to a shared capacitive buffer (here referred as a virtual bus). This configuration enables the bulk server power to flow through the series connection without being processed. In order to guarantee that the server voltages stay within safe limits in case of a power mismatch between the servers, the virtual bus provides an environment where the *difference* in current between the server current and bus current can be instantaneously supplied or stored. In comparison to the conventional systems shown in Fig. 1, where each server’s power con-

verter has to process the *full* server power, the server-to-virtual bus architecture processes only the *difference* in power between servers. This architecture thus separates the total amount of processed power from the power delivered, resulting in considerable reduction of the power conversion losses, in particular when the individual server power consumptions are similar.

Although the power is distributed in ac form within the data center in Fig. 2, neither the proposed architecture nor the discussion in this paper changes if the power is distributed in dc form within the data center.

For a qualitative comparison between various DPP architectures for data center power delivery applications, advantages of the server-to-virtual bus DPP architecture, and experimental results that show the voltage regulation of both the series-stacked servers and the virtual bus during the steady-state operation of the servers, readers are referred to [17].

III. DERIVATION OF THE TOTAL AMOUNT OF PROCESSED POWER IN THE SERVER-TO-VIRTUAL BUS DPP ARCHITECTURE

As stated in Section II, the power delivery nature of the series-stacked and DPP architecture enables the processing of only the difference in power between series-connected servers. The total processed power varies depending on average power consumption of each series-connected server. Moreover, while all servers are operational, lack of a server in the series connection due to a hot-swapping event represents an abnormal power difference for the remaining servers, and thus increases the total power processed. Derivation of the total power processed in the system for a given load distribution scenario is thus an important consideration when evaluating the proposed system. This section derives the total average power processed in the server-to-virtual bus DPP architecture and compares it with the total average power delivered to the servers for various power consumption distributions.

In Fig. 3, a schematic diagram of the server-to-virtual bus DPP architecture is shown that facilitates the discussion. Here, an ideal dc voltage source models the output of an ac–dc converter (or connection to a high-voltage dc bus) which provides power (by V_{Bus} and i_{Bus}) to the series-stacked servers. The terms $i_{S,j}$ and $v_{S,j}$ represent the j th server current and voltage, where the subscript j may refer to any server in the series-stack, and J is the total number of servers (i.e., $j \in \{1, 2, 3, \dots, J\}$) in steady-state operation. The virtual bus capacitor voltage and current are represented by v_{VB} and i_{VB} , respectively. The input and output currents of the dc–dc converters are shown separately as $i_{\Delta,j}$ and $i_{\delta,j}$. Since each dc–dc converter is connected between a server in the series-stack and the virtual bus, the terminal voltages of the j th dc–dc converter are $v_{S,j}$ and v_{VB} .

In the server-to-virtual bus DPP architecture shown in Fig. 3, KVL around the series-stack and KCL at every intermediate node of the series-stack result in

$$V_{\text{Bus}} = \sum_{j=1}^J v_{S,j}, \quad (1)$$

$$i_{\text{Bus}} = i_{S,j} + i_{\Delta,j} \quad \forall j, \quad (2)$$

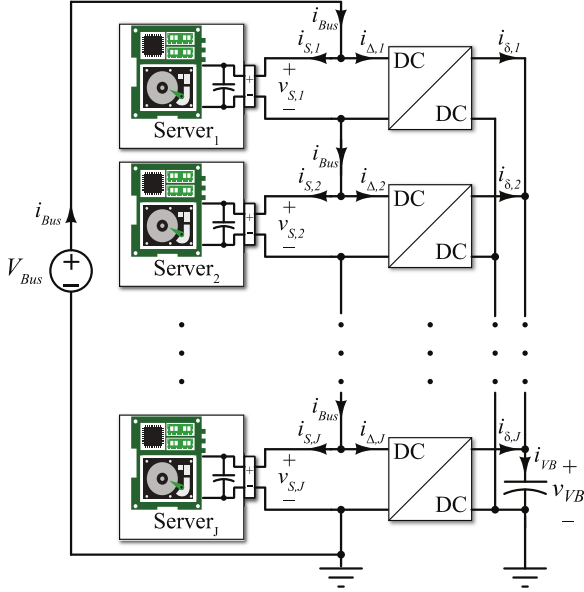


Fig. 3. Server-to-virtual bus DPP architecture.

respectively. Also, since the virtual bus capacitor is connected in parallel with all dc–dc converters, the virtual bus current is given by

$$i_{VB} = \sum_{j=1}^J i_{\delta,j}. \quad (3)$$

Assume that all series-stacked server voltages and the virtual bus voltage are regulated to their nominal steady-state values ($V_{S,nom}$ and $V_{VB,nom}$, respectively) with a general time period. (See [36] for an example of a control algorithm that accomplishes this.) Averaging (1), (2), and (3) in this general time period results in

$$V_{Bus} = \sum_{j=1}^J V_{s,j} = J \times V_{S,nom}, \quad (4)$$

$$I_{Bus} = I_{S,j} + I_{\Delta,j} \quad \forall j, \quad (5)$$

and

$$I_{VB} = \sum_{j=1}^J I_{\delta,j}, \quad (6)$$

respectively. Recall that the virtual bus is a capacitive buffer. As the controller regulates its voltage to a constant value ($V_{VB,nom}$), the average current into the virtual bus becomes zero during the time period. For simplicity of the analysis, assume that the differential converters are ideal, and they provide a one-to-one voltage conversion (i.e., $V_{S,nom} = V_{VB,nom}$, $i_{\Delta,j} = i_{\delta,j}$ and $I_{\Delta,j} = I_{\delta,j}$). Therefore,

$$I_{VB} = \sum_{j=1}^J I_{\delta,j} = 0 \Rightarrow \sum_{j=1}^J I_{\Delta,j} = 0. \quad (7)$$

The node current equations given by (5) are valid for every server and differential converter pair throughout the series-stack, as shown in Fig. 3. The sum of J node current equations gives

$$J \times I_{Bus} = \sum_{j=1}^J (I_{S,j} + I_{\Delta,j}) = \sum_{j=1}^J I_{S,j} + \sum_{j=1}^J I_{\Delta,j}. \quad (8)$$

Given the constraint of (7), (8) simplifies to

$$I_{Bus} = \frac{1}{J} \sum_{j=1}^J I_{S,j} \quad (9)$$

which states that the average bus current equals the mean of time average of the server currents.

The average power processed by the differential converters (i.e., the total power processed in the system) is

$$P_{processed} = \sum_{j=1}^J |P_{\Delta,j}| = \sum_{j=1}^J |V_{s,j} I_{\Delta,j}| \quad (10)$$

where $P_{\Delta,j}$ is the average power processed by the j th differential converter. With a control algorithm that successfully regulates the series-stacked server voltages to their nominal values ($V_{S,nom}$), (10) simplifies to

$$P_{processed} = V_{s,nom} \sum_{j=1}^J |I_{\Delta,j}|. \quad (11)$$

Moreover, using the KCL constraint given by (2), (11) can be rewritten as

$$P_{processed} = V_{s,nom} \sum_{j=1}^J |I_{Bus} - I_{S,j}|. \quad (12)$$

Since the average bus current (I_{Bus}) is equal to the mean of time average of the server currents by (9), the total amount of power processed given by (12) can also be expressed as

$$P_{processed} = V_{s,nom} \sum_{j=1}^J \left| \left(\frac{1}{J} \sum_{j=1}^J I_{S,j} \right) - I_{S,j} \right|. \quad (13)$$

It is illustrative to compare the total power processed with the total power delivered to the servers in the server-to-virtual bus DPP architecture. Assuming ideal converters, the total power delivered to the servers is the sum of all server powers, and it can be expressed as

$$P_{delivered} = \sum_{j=1}^J P_{s,j} = V_{s,nom} \sum_{j=1}^J I_{S,j} \quad (14)$$

where $P_{s,j}$ is the average power consumed by the j th server. Recall that I_{Bus} equals to the mean of time average of the server currents by (9), which implies

$$\min(I_{S,j}) < I_{Bus} < \max(I_{S,j}) \quad \forall j. \quad (15)$$

Since $I_{S,j} > 0$ for all j when all servers are operational, it can be observed that the power processed by (12) is always less than the power delivered by (14). This is an important feature of the series-stacked power delivery architecture.

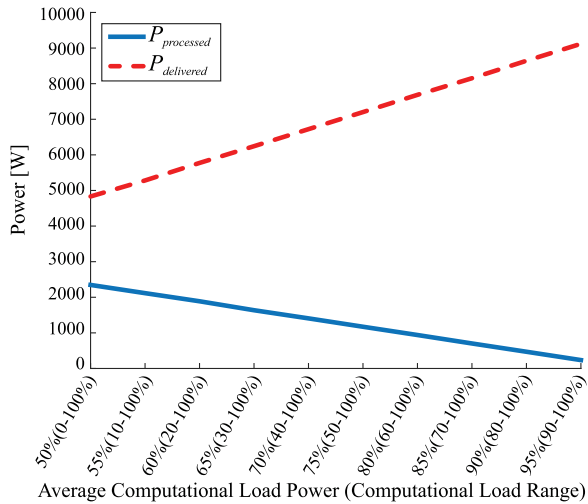


Fig. 4. Statistical comparison between the total processed power and delivered power in the server-to-virtual bus DPP architecture, assuming ideal converters and Gaussian distribution of computational load range.

Case Study I: A statistical case study is performed in order to compare the power processed and power delivered in the server-to-virtual bus DPP architecture using (13) and (14). In this case study, a server rack consisting of 32 servers (each rated at 300 W) is used to illustrate the effect of different mismatch conditions. The average computational load for the server rack is swept from 50% to 95% of the rated power. For every average computational load scenario, the computational load range within the server rack is randomly assigned using Gaussian distribution for 1000 iterations, causing different mismatch conditions. Equations (13) and (14) are used to calculate the total power processed and power delivered at every iteration, and then the results of 1000 iterations are averaged for every scenario. The result is plotted in Fig. 4.

Figure 4 shows the total power processed and delivered in the server-to-virtual bus DPP architecture versus average computational load scenarios (and computational load ranges). As the average computational load is increased from 50% to 95% of the rated power, the computational load range narrows (i.e., less mismatch between the series-stacked servers). As expected, the power delivered increases as the average computational load increases. However, the power processed decreases as the average computational load increases. This is because the average server current is delivered through the dc bus without being processed and the Gaussian nature of the computational load range narrows as the average computational load increases.

IV. HOT-SWAPPING FOR SERIES-STACKED SERVERS

An operation essential for servers in data centers is hot-swapping, which is inserting or removing individual servers while the other servers in the rack are operational. The hot-swapping operation may be required if a server is intentionally removed for maintenance or if a server unexpectedly fails and requires repairs. This section explains the conceptual hot-swapped

operation of the server-to-virtual bus DPP architecture, provides practical requirements to achieve safe implementation of hot-swapping in the series-stacked architectures and also summarizes the high-level system control idea.

A. Hot-Swapped Operation in the Server-to-Virtual Bus DPP Architecture

The *Case Study I* in Section III compared the processed and delivered power in the server-to-virtual bus architecture, while the computational load distribution within the series-stacked servers has a Gaussian nature. Although hot-swapping or a server failure occurs for a fairly low percentage of the operation time, they represent a severe mismatch in terms of the computational load distribution for the series-stacked architecture. Here, the hot-swapped operation in the series-stacked architecture is first conceptually explained through an example server rack that includes six servers. Then, the 32-server rack and two of the average computational load scenarios in *Case Study I* are studied again to show the processed and delivered power under the hot-swapped operation.

Case Study II: Fig. 5 illustrates examples of normal and hot-swapped operation of a six-server rack employed in the server-to-virtual bus DPP architecture. Assuming 300 W rated power and 12 V nominal voltage servers, the annotated currents on the servers shown in Fig. 5(a) represent 95% average computational load and $\pm 5\%$ computational load range within the rack, similar to the one in *Case Study I*. It can be seen in Fig. 5(a) that the bus current equals the average of six server currents by (9), each differential converter injects or rejects the difference in current between its corresponding server current and the bus current by (5), and the total current into the virtual bus capacitor is zero, resulting in 63 W processed power by (10) and 1701 W delivered power by (14).

In Fig. 5(b), an example is shown where the third server is swapped out from the six-server rack that is under the same load distribution in Fig. 5(a). Treating the absence of the third server as the third server consuming 0 A, the bus and differential currents can be calculated by (9) and (5), respectively. Note that the differential converter of the third server ensures the flow of the bus current in the series-stack by injecting it to the virtual bus. In order to ensure the virtual bus is regulated, the remaining differential converters share the extra current on the virtual bus, and inject it back to their servers. For this example, the processed and delivered power are 474 and 1422 W, by (10) and (14), respectively.

Figures 5(c)–(f) illustrate a series of examples where the servers are swapped out from the six-server rack one by one. It can be observed that the bus current decreases as more servers are swapped out since the bus current equals the average of six-server currents (again, treating the swapped-out server currents as 0 A). Also, the differential converters of the swapped-out servers inject the bus current to the virtual bus while the remaining differential converters share the total amount of extra current and inject it back to their servers. The processed and delivered power for every example operation, as shown in Fig. 5, are given in Table I.

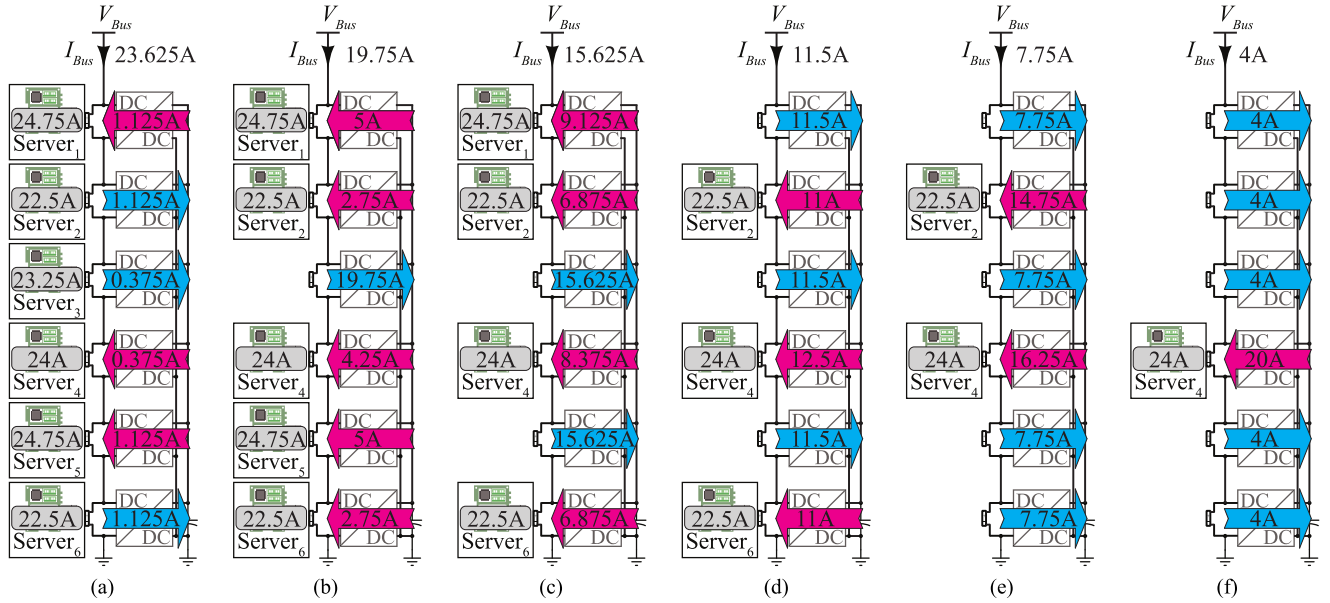


Fig. 5. Examples of normal and hot-swapped operations of a six-server rack. Annotated server currents correspond to a 95% average computational load with $\pm 5\%$ computational load range scenario when all servers are operational. Annotated differential currents show direction and average amount of current flow, assuming ideal converters. (a) All servers are operational. (b) One server is swapped-out. (c) Two servers are swapped-out at the same time. (d) Three servers are swapped-out at the same time. (e) Four servers are swapped-out at the same time. (f) Five servers are swapped-out at the same time.

TABLE I

THE PROCESSED AND DELIVERED POWER IN THE EXAMPLE OPERATIONS OF THE SIX-SERVER RACK IN FIG. 5, CALCULATED BY (10) AND (14)

	Fig. 5(a)	Fig. 5(b)	Fig. 5(c)	Fig. 5(d)	Fig. 5(e)	Fig. 5(f)
$P_{\text{delivered}} [W]$	1701	1422	1125	828	558	288
$P_{\text{processed}} [W]$	63	474	750	828	744	480

Recall from (14) and (12) the expressions for the power delivered and processed in the server-to-virtual bus DPP architecture with ideal differential converters. Since the physical location of a server in the series-stack does not affect (14) and (12), the hot-swapped servers can be lumped on top of the series-stack in order to simplify the expressions of the following equations:

$$P_{\text{delivered}} = V_{s,\text{nom}} \sum_{j=H+1}^J I_{S,j} \quad (16)$$

and

$$P_{\text{processed}} = V_{s,\text{nom}} \left[\sum_{j=1}^H I_{\text{Bus}} + \sum_{j=H+1}^J |I_{\text{Bus}} - I_{S,j}| \right] \quad (17)$$

where H is the number of simultaneously hot-swapped servers (i.e., $I_{S,j} = 0$ for $j \in \{1, 2, \dots, H\}$, and $I_{S,j} > 0$ for $j \in \{H+1, H+2, \dots, J\}$).

In order to further simplify (17), assume that the bus current is less than all remaining server currents during any hot-swap, which means $|I_{\text{Bus}} - I_{S,j}| < 0$ for $j \in \{H+1, H+2, \dots, J\}$. (Note that this is the same situation studied in *Case Study II* and

Fig. 5.) Then, (17) becomes

$$\begin{aligned} P_{\text{processed}} &= V_{s,\text{nom}} \left[\sum_{j=1}^H I_{\text{Bus}} + \sum_{j=H+1}^J (-I_{\text{Bus}} + I_{S,j}) \right] \\ &= V_{s,\text{nom}} \left[H \times I_{\text{Bus}} - (J - H) \times I_{\text{Bus}} + \sum_{j=H+1}^J I_{S,j} \right] \\ &= V_{s,\text{nom}} \left[(2H - J) \times I_{\text{Bus}} + \sum_{j=H+1}^J I_{S,j} \right]. \quad (18) \end{aligned}$$

Recall that the bus current is the mean of time average of the server currents by (9). Since $I_{S,j} = 0$ is valid for all hot-swapped servers, (9) can be restated as

$$I_{\text{Bus}} = \frac{1}{J} \sum_{j=H+1}^J I_{S,j}. \quad (19)$$

Using (19) in (18), the power processed becomes

$$\begin{aligned} P_{\text{processed}} &= V_{s,\text{nom}} \left[(2H - J) \times \frac{1}{J} \sum_{j=H+1}^J I_{S,j} + \sum_{j=H+1}^J I_{S,j} \right] \\ &= V_{s,\text{nom}} \left[\left(\frac{(2H - J)}{J} + 1 \right) \times \sum_{j=H+1}^J I_{S,j} \right] \\ &= V_{s,\text{nom}} \left[\left(\frac{2H}{J} \right) \times \sum_{j=H+1}^J I_{S,j} \right] \\ &= \frac{2H}{J} P_{\text{delivered}}. \quad (20) \end{aligned}$$

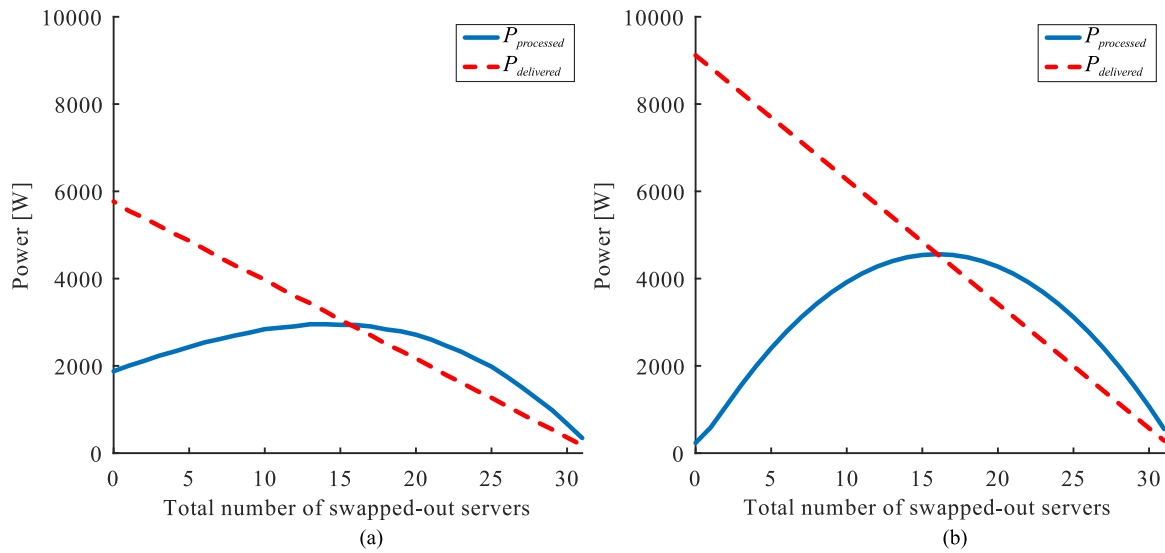


Fig. 6. Comparison of total processed and delivered power in server-to-virtual bus architecture. (a) 60% average computation load and $\pm 40\%$ computational load range. (b) 95% average computation load and $\pm 5\%$ computational load range.

On the other hand, if the bus current is higher than at least one of the remaining server currents during a hot-swap (i.e., $|I_{\text{Bus}} - I_{S,j}| > 0$ for at least one $j \in \{H+1, H+2, \dots, J\}$), further simplification of (17) requires more information about the server currents, and therefore does not produce a closed-form expression.

Equation (20) can be used to relate the power processed to the power delivered in the series-stacked architecture when the bus current is less than all remaining server currents during any hot-swap (i.e., $I_{\text{Bus}} < I_{S,j}$ for $j \in \{H+1, H+2, \dots, J\}$). In such a scenario, it can be observed that due to the $\frac{2H}{J}$ term, the power processed is guaranteed to be less than the power delivered in the server-to-virtual bus DPP architecture unless half of the servers are hot-swapped at the same time. (Note that $P_{\text{processed}}$ and $P_{\text{delivered}}$ in Table I also follow (20).)

Case Study III: In this case study, the hot-swapped operation is applied to the 32-server rack used in *Case Study I*. For 60% and 95% average computational load scenarios in *Case Study I*, (13) and (14) are plotted versus the total number of swapped-out servers in Fig. 6(a) and (b), respectively.

As shown in Fig. 6(a) and (b), unless half of the servers in a series-stack are swapped out at the same time, the power processed in the server-to-virtual bus DPP architecture is less than the power delivered, yielding an efficiency improvement for the series-stacked approach compared to conventional solutions. Although the power processed is more than the power delivered after more than half of the servers are hot-swapped at the same time, the power processed decreases as more servers are hot-swapped since the power processed is related to power delivered by (20).

B. Hot-Swapping Requirements for Series-Stacked Architectures

The hot-swapping implementation in the conventional power delivery architectures is straightforward due to the individual

nature of the power delivery paths. However, in all series-stacked architectures, the hot-swapping implementation becomes challenging since the serially connected servers form the main power delivery path. In order to safely implement hot-swapping in any series-stacked architecture, the system must be capable of isolating the swapped-out server from the series-stack, sinking the bus current through the differential converter during hot-swapping, while also limiting the in-rush current when the server is swapped in.

The hot-swapping operation first requires the complete isolation of the server from the series-stack. This is crucial for the safety of the operators, since each server is at a different voltage level in the series-stack. Moreover, once the server is isolated from the series-stack, the absence of the server in the main current flow path must be detected by the control algorithm and the flow of the bus current in the series-stack must be ensured by the corresponding differential converter. Finally, when the server is plugged in after a hot-swap event, an in-rush current occurs due to the high server input capacitance [37]. In the series-stacked architecture, since such a transient is too rapid to be compensated by the differential converters, the in-rush current tends to flow through the serially connected servers. If unchecked, this may lead to a voltage imbalance across the series-stack, potentially damaging the adjacent servers and differential converters. Therefore, limiting the in-rush current becomes a particularly important requirement in the series-stacked architecture.

There are multiple methods that can be used to limit the in-rush current into the server when it is connected. Most methods operate by producing a high-resistance path to allow the server capacitance to charge at reduced currents, and then connecting a low-resistance path. One simple solution is to have a connector with staggering pin lengths [38], where the high-resistance paths connect first, followed by the low-resistance paths. The use of positive temperature coefficient thermistors to reduce the in-

rush current is possible but can be unreliable due to the slow response time. Digitally controlled discrete MOSFETs can be used to electronically enable and disable the high- and low-resistance paths. There are also custom ICs made for hot-swapping that sense current and operate a MOSFET in the linear region if the in-rush current is too high [39]. Although such methods can successfully limit the in-rush current, they are not capable of providing a complete isolation since off-the-shelf solutions only consists of one transistor to limit the in-rush current at the high-side or low-side terminal. A more advanced power circuitry is thus needed for the series-stacked architecture, where the negative supply terminal can be isolated as well. A hot-swapping circuitry that achieves the complete isolation of the server and in-rush current limitation during swap-in is described in Section V.

C. Bidirectional Hysteresis Control for the Server-to-Virtual Bus DPP Architecture

As can be seen through the mathematical derivation of the processed power in Section III and *Case Study II*, the differential converters must provide the difference between the bus current and the server current, while assuring that the virtual bus is a safe energy reservoir for instantaneous power mismatch. Each differential converter is thus responsible for regulating both its input (i.e., virtual bus) and output (i.e., server) voltage, by injecting or rejecting current to its server or the virtual bus. A control algorithm that achieves the voltage regulation of the series-stacked servers and the virtual bus was explained and validated in simulation in [36] for both normal and hot-swapped operations. The control algorithm was experimentally validated for the normal operation of the server-to-virtual bus DPP architecture in [17]. Also, the preliminary experimental results that show the hot-swapped operation in [21] use the same control algorithm explained in [36]. The details of the control algorithm are beyond the scope of this paper; however, the control objectives and idea are summarized here.

The control objectives can be summarized as follows: First, each differential converter is responsible for its input and output voltage regulation by injecting current to its server or the virtual bus. The control idea thus needs to determine the current flow direction dynamically. In addition, during normal operation where the load distribution between the servers is balanced, the expected mismatch between the servers is much less than the mismatch during a hot-swapped operation. The differential converter should efficiently process the differential power between the servers at both light-load (during normal operation) and full-load (during hot-swapped operation) conditions. Moreover, a control algorithm that does not require communication between the differential converters and uses only voltage feedback to regulate both the series-stacked server voltages and the virtual bus voltage simplifies the overall system implementation.

A distributed bidirectional hysteresis control algorithm that uses only voltage feedback for regulating the series-stacked server voltages and the virtual bus voltage can be summarized as follows: In order to determine the differential current need of each voltage domain, in every sampling time each differential converter measures both its input and output voltage, then

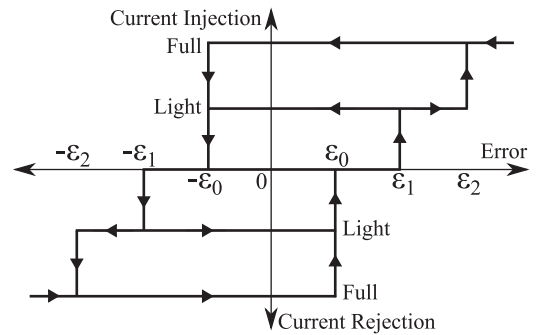


Fig. 7. Bidirectional hysteresis shape. Note that the bidirectional hysteresis shape is generic and independently valid for each series-stacked server and the virtual bus.

subtracts these values from a reference voltage value in order to calculate the errors in its input and output voltage. Depending on the magnitude and direction of the errors, the current injection or rejection need of any of the voltage domains (servers and the virtual bus) can be determined by referring to the bidirectional hysteresis shape given in Fig. 7. For example, when the voltage error of any voltage domain is less than $\pm\epsilon_2$, current injection or rejection in light-load mode is sufficient to regulate the voltage domain within $\pm\epsilon_0$ range. However, if the voltage error of any voltage domain is more than $\pm\epsilon_2$, the differential converter needs to operate in full-load mode until the voltage domain is regulated within $\pm\epsilon_0$ range. Also, the differential converters are kept OFF to maximize power delivery efficiency when their corresponding voltage domains are within $\pm\epsilon_1$ range. Lastly, based on the current injection or rejection need of both voltage domains of a differential converter, a decision about current flow direction and magnitude is made at every sampling time. Readers are referred to [36] for a more in-depth analysis and explanation of this control algorithm, and to [17] for its experimental validation during normal operation of the series-stacked servers.

V. PROTOTYPE DPP HARDWARE FOR THE SERVER-TO-VIRTUAL BUS DPP ARCHITECTURE

The schematic of the prototype DPP hardware for the server-to-virtual bus architecture is depicted in detail in Fig. 8. This hardware has three terminals: a server terminal (*Server+* and *Server-* in Fig. 8), a virtual bus terminal (*Virtual Bus+* and *Virtual Bus-* in Fig. 8), and a series-stack terminal (*Series-Stack+* and *Series-Stack-* in Fig. 8) for interconnecting to the series-stack. These three terminals separate the hardware into two stages: the interface stage and the differential converter stage. The interface stage between the server and the series-stack terminals holds the stack initialization circuitry and also is responsible for the hot-swapping operation. The differential converter stage is placed between the series-stack and virtual bus terminals and responsible for the server and virtual bus voltage regulation. This section provides the implementation details of the initialization circuitry, hot-swapping circuitry, and differential converter.

TABLE III
KEY COMPONENTS OF THE HOT-SWAPPING CIRCUITRY

Hot-swapping switches	TI CSD18540Q5
Digital isolator	TI ISO724x series
Isolated power supply	TI DCP010505BP
LDO	TI LP2985
R_{limit}	2.8 Ω

is completed, v_{Stack} is actually divided between R_{limit} and the high-resistance in parallel with the input capacitor. Although R_{limit} is small, the voltage drop across it still creates a small current spike as M_1 turns ON, which affects the other series-stacked voltages if not managed. The turn-on transient of M_1 is thus decelerated by increasing the gate resistance R_{G1} as well. An additional discrete capacitor ($C_{G1} = 1 \mu\text{F}$ in Fig. 8) is added to the gate of M_1 to further decelerate the turn-on transient of M_1 without increasing R_{G1} beyond 100 k Ω , as excessive gate resistance can cause the transistor to operate in the linear region, owing to the inherent gate leakage of the device. The turn-off transient of M_1 and M_3 , on the other hand, must be as rapid as possible to quickly isolate the malfunctioning server from the series-stack during a swap-out event. Ultrafast Schottky diodes are added to the gate driving circuit of M_1 and M_3 in order to bypass the highly resistive R_{G1} and R_{G3} , while their gate capacitors are discharged during a swap-out event.

The goal of this work was to design hot-swapping circuitry that did not reduce the efficiency noticeably. The TI CSD16570Q5B was found to have the lowest on-state resistance at 0.59 m Ω with a 5 V gate signal. In order to achieve lower on-state resistance, five MOSFETs are paralleled in the hot-swapping circuitry. The additional components of the hot-swapping circuitry are listed in Table III. Note that the hot-swapping switches used in this implementation are rated at 60 V since the dc bus voltage in the experimental testbed described in Section VI-A is 48 V.

C. Differential Converter

The differential converter stage depicted in Fig. 8 is a dual active bridge (DAB) dc–dc converter which offers bidirectional power flow with a symmetric design at both sides of the transformer when the input (server) and output (virtual bus) voltages are nominally the same [12], [40], [41]. The DAB converter is implemented with off-the-shelf discrete components that are common in many server power supply designs such as a power stage that employs a high-side and a low-side MOSFET with integrated gate driver circuitry for each half bridge in the converter. Digital isolators are used as level shifters in order to transfer necessary digital signals to the different levels of the series-stack. The key components of the differential converter are listed in Table IV.

Each DAB converter in this work is designed to be able to sink or source 120 W at 12 V from the series-stack terminal depending on the power flow direction since the experimental work in Section VI employs 120 W servers. The switching frequency is 200 kHz and simple phase-shift modulation technique is used to determine the power flow direction and output power. The

TABLE IV
KEY COMPONENTS OF THE DIFFERENTIAL CONVERTER

Power stage	TI CSD95372BQ5MC
Digital isolators	TI ISO724x series
LDOs	TI LP298x series
Transformer	Coilcraft SMT PL160 \times 2
Inductor	Coilcraft SLC1480
C_{in} and C_{out} , ceramic	TDK 10 μF 16 V X5R \times 6
C_{in} and C_{out} , aluminum	Panasonic 1 mF 16 V SMD \times 2

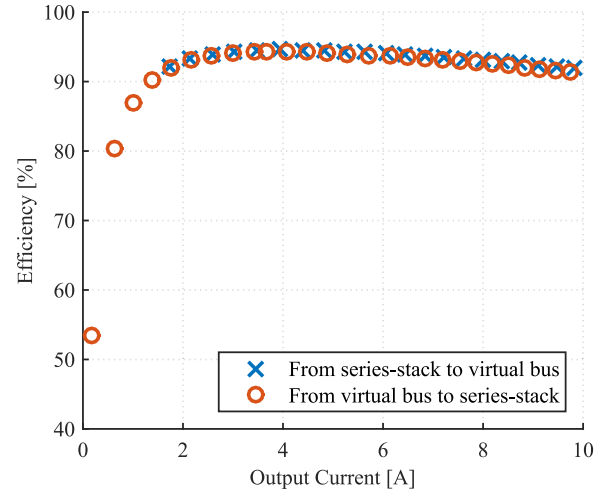


Fig. 9. Measured efficiency versus output current of the differential converter. All necessary auxiliary logic voltages are generated by on-board linear regulators; however, an off-board microcontroller is used for converter control. The efficiency measurement thus contains all converter-related losses except microcontroller power consumption.

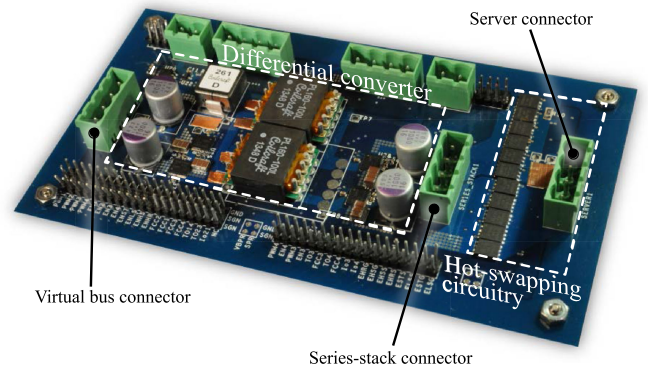


Fig. 10. Annotated photograph of the prototype hardware.

measured efficiency of the DAB converter prototype is plotted in Fig. 9 for both power flow directions. As shown in Fig. 9, due to the symmetric design of the converter, almost identical efficiency curves for both power flow directions are achieved with a peak at 95% around 40 W. An annotated photograph of the prototype hardware is given in Fig. 10.

VI. EXPERIMENTAL WORK

In order to validate hot-swapping operation in the server-to-virtual bus DPP architecture, an experimental study is

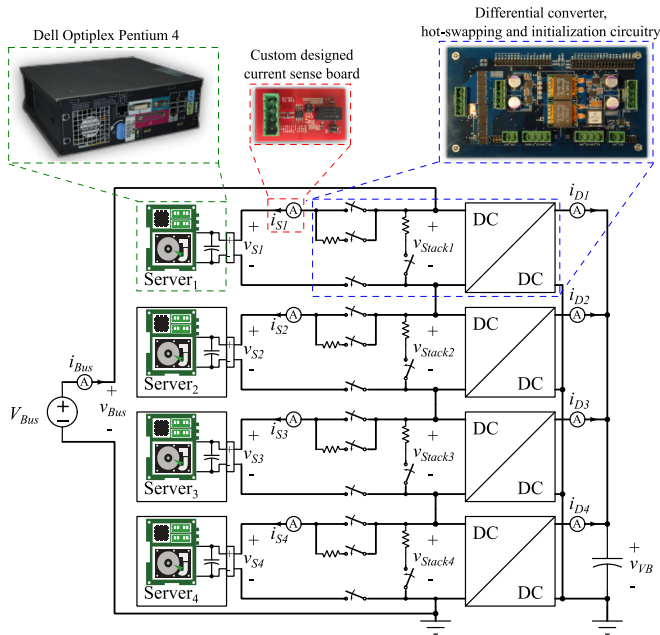


Fig. 11. Annotated schematic of the testbed.

performed. This section explains the details of the experimental setup and provides the results.

A. Testbed

A flexible and modular laboratory testbed is developed for the server-to-virtual bus DPP architecture. The servers are 4 Dell Optiplex SX280 workstations with Pentium 4 CPU and 80 GB Western Digital magnetic hard drive running Linux operating systems. These workstations have a single 12 V motherboard input and are rated for 120 W peak power.¹ A dc power supply (HP 6674A) provides the 48 V dc bus to the four-server rack. The testbed is grounded to earth at the negative terminal of V_{Bus} and v_{VB} ; therefore in this work all voltages are positive with respect to earth. We note that various other grounding options such as positive-earthed, mid-raid earthed or indirectly earthed through the ac mains input rectifier may exist in data centers. While the proposed architecture is independent of ground location, care must be taken in the safety isolation of the floating servers, as well as the implementation of hot-swapping circuitry with respect to polarity and blocking capabilities of the switches. Four hardware prototypes with the above-mentioned functionalities are built for the four-server testbed.

A single off-board microcontroller (TI C2000 Piccolo F28069) samples all series-stack voltages, runs the control algorithm to generate the pulse width modulation (PWM) signals for four differential converters, and manages the enable/disable signals for all interface circuitries. A 32 mF discrete capacitor is used as the virtual bus capacitor, and doubled 16 AWG copper

¹Although these workstations consume lower power than typically encountered servers in data centers, their terminal characteristics in response to computational loads are similar to the high-end servers. The workstations used in this experimental work serve as scaled down versions of the costly servers and enable us to validate the hot-swapping concept at reasonable cost.

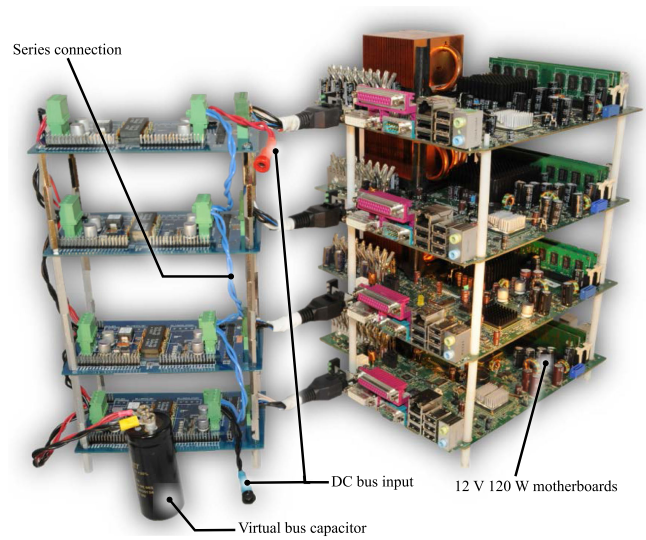


Fig. 12. Annotated picture of the testbed.

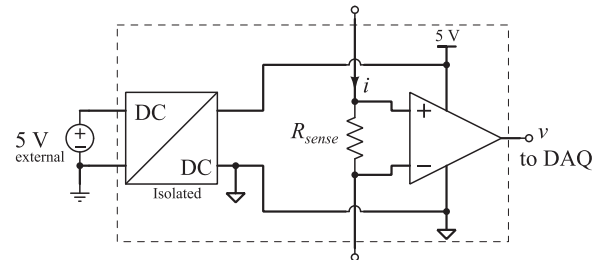


Fig. 13. Schematic of the custom designed current sense board.

TABLE V
KEY COMPONENTS OF THE CUSTOM DESIGN CURRENT SENSE BOARD

Current sense resistor	Stackpole electronics CSNL1206FT3L00 (3 mΩ, 1%)
Current shunt monitor	Analog devices AD8210
Isolated power supply	CUI PQM1-S5-S5-M

wire is used for all interconnections in the testbed. An annotated schematic and photograph of the testbed are given in Figs. 11 and 12, respectively.

B. Measurement System

A data acquisition unit from National Instruments is used to simultaneously sample the annotated signals in Fig. 11: server voltages (v_{S1} – v_{S4}) and currents (i_{S1} – i_{S4}), series-stack voltages (v_{Stack1} – v_{Stack4}), bus voltage (v_{Bus}) and current (i_{Bus}), differential currents (i_{D1} – i_{D4}), and virtual bus voltage (v_{VB}), at 5000 samples/s. All voltages are measured directly, however, for each current measurement, a custom design current sense board is used.

In Fig. 13, the schematic of the custom design current sense board that consists of a 3 mΩ high power current sense resistor, a high common-mode voltage current shunt monitor, and an isolated dc–dc converter with a single regulated output (see Table V for part numbers of the custom design current sense

board) is shown. All current sense boards are energized with a separate 5 V dc power supply regulated through an on-board isolated dc-dc converter. The voltage output of each current shunt monitor is carefully calibrated by an Agilent 34410A 6 1/2 digit digital multimeter at each corresponding common mode voltage in order to accurately capture the very high efficiencies of the series-stacked system.

C. Efficiency and Power Loss Calculations

The system-level efficiency is calculated as follows. The instantaneous input power to the system is calculated by the multiplication of the measured instantaneous bus current and voltage as follows:

$$p_{in} = v_{Bus} \times i_{Bus}. \quad (21)$$

Each server's instantaneous power is calculated by multiplying the measured server current and voltage, and the total instantaneous output power is the sum of each server's power consumption by

$$p_{out} = \sum_{j=1}^4 v_{S,j} \times i_{S,j}. \quad (22)$$

The difference between (21) and (22) is the power loss in the system by

$$p_{loss} = p_{in} - p_{out} \quad (23)$$

and it can be grouped into four categories as follows:

- (1) Measurement loss ($p_{loss,meas.}$): The current sense boards used to measure the server and differential currents are placed inside the series-stacked system, as shown in Fig. 11. The measurement loss due to the sense resistors is captured by

$$p_{loss,meas.} = \underbrace{\sum_{j=1}^4 R_{sense} \times i_{S,j}^2}_{\text{due to server current}} + \underbrace{\sum_{j=1}^4 R_{sense} \times i_{D,j}^2}_{\text{due to differential current}} \quad (24)$$

where R_{sense} is 3 m Ω . In this paper, R_{sense} is assumed to be constant regardless of the conducted current since the current range (i.e., 0–12 A) is well below the rated power of the sense resistor (i.e., 1 W).

- (2) Hot-swapping circuitry loss ($p_{loss,HS}$): The hot-swapping circuitry, as shown in Fig. 8, comprises transistors that cause conduction loss due to their on-state resistance. In this paper, it is referred to as *hot-swapping circuitry loss* and calculated by

$$p_{loss,HS} = \sum_{j=1}^4 (v_{Stack,j} - v_{Out,j}) \times i_{S,j} \quad (25)$$

where $v_{Out,j}$ is the voltage at the server terminal of j th prototype DPP hardware, and calculated by

$$v_{Out,j} = v_{S,j} + i_{S,j} \times R_{sense} \quad \forall j. \quad (26)$$

- (3) Cabling loss ($p_{loss,cabling}$): As mentioned in Section VI-A, the dc bus and hardware prototypes are connected to each

other with doubled 16 AWG copper wire in order to form the series-stack connection. The total voltage drop in the series-stack connection can be captured by

$$v_{drop} = v_{Bus} - \sum_{j=1}^4 v_{S,j}$$

which causes conduction loss in the series connection. In this paper, it is referred to as *cabling loss* and calculated by

$$p_{loss,cabling} = v_{drop} \times i_{Bus} \quad (27)$$

since the current in the series-stack connection is the same as the bus current.

- (4) Power conversion loss ($p_{loss,conv.}$): As mentioned before, the differential converters in Fig. 11 are bidirectional dc-dc converters. The measurement of power loss due to power processing in a differential converter thus requires instant detection of power flow. Instead, in this paper, all remaining power loss in the system is lumped together in power conversion loss, which is calculated by

$$p_{loss,conv.} = p_{loss} - p_{loss,meas.} - p_{loss,HS} - p_{loss,cabling}. \quad (28)$$

In order to calculate the efficiency of the series-stacked system, the instantaneous power calculations given by (21)–(28) are averaged over a time interval

$$P = \frac{1}{T \times f_s} \sum_1^{T \times f_s} p \quad (29)$$

where T is the length of the time interval, f_s is the sampling rate of the data acquisition unit, and P is the average value of the instantaneous power of interest p .

Finally, the system-level efficiency that includes power conversion loss, hot-swapping circuitry loss, and cabling loss but excludes the loss due to the current sensing is calculated by

$$\eta_{sys} = 1 - \frac{P_{loss,sys}}{P_{in}} \quad (30)$$

where

$$P_{loss,sys} = P_{loss,HS} + P_{loss,cabling} + P_{loss,conv.}$$

On the other hand, the system-level *power conversion efficiency* (i.e., the power processing efficiency of all differential converters) is calculated by

$$\eta_{conv.} = 1 - \frac{P_{loss,conv.}}{P_{in}}. \quad (31)$$

D. Test Scenario

A 500-s test is executed on the testbed in order to validate the hot-swapping concept on the server-to-virtual bus DPP architecture. In this paper, the standard Linux “stress” utility [42] is used as computational load on servers in order to replicate a real-world computation scenario. The test starts with the initialization of the series-stacked converters by using the shunt resistors as described before. Then, four servers are connected to the series-stack by using the hot-swapping circuitry. As the

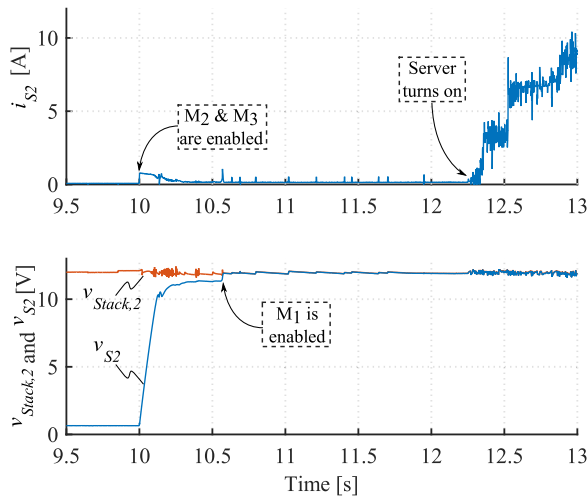


Fig. 14. Measured instantaneous data showing swap-in transient of the second server.

operating system on the servers initializes, an Ethernet connection is established in order to start the stress test for 300 s. After the first two minutes of the stress test, one of the servers is swapped out and kept isolated from the series-stacked architecture for a minute, while the other servers are still continuing the stress test. The swapped-out server is then swapped in to the series-stack and reinitialized, and the stress test continues for 2 more minutes. Following the conclusion of the stress test, shutdown commands are sent to the servers. The shunt resistors are then connected back to the series-stack nodes in order to keep voltage balanced throughout the stack until the dc bus is disconnected.

E. Results

When the dc bus is first applied to the series-stack at the beginning of the test, all servers are isolated from the series-stack and the shunt resistors are connected between the series-stack terminals of the DPP hardware. The applied bus voltage is thus equally divided between the shunt resistors, allowing linear regulators on DPP hardware to provide logic voltages to the digital isolators and gate drivers. The differential converters are then enabled to regulate both their input and output voltages to 12 V. The shunt resistors are disconnected a few seconds after the control algorithm initializes as explained in Section V-A. The servers are then simultaneously connected to the series-stack 10 s into the experiment by using the hot-swapping circuitry. This operation (previously explained in detail in Section V-B) is shown in Fig. 14 through the measured and annotated current and voltage of the second server as an example of the swap-in transient from the experiment. At 10 s, the high-resistance path of the hot-swapping interface (given in Fig. 8) is enabled by turning ON M_2 and M_3 . As can be seen in Fig. 14, the server current is limited to less than 1 A by the linear region operation of M_2 , causing a linear increase of the server voltage (v_{S2}). Around the 10.7 s mark, the low-resistance path of the hot-swapping interface is slowly enabled through M_1 , causing a small current increase due to the voltage drop across R_{limit} in

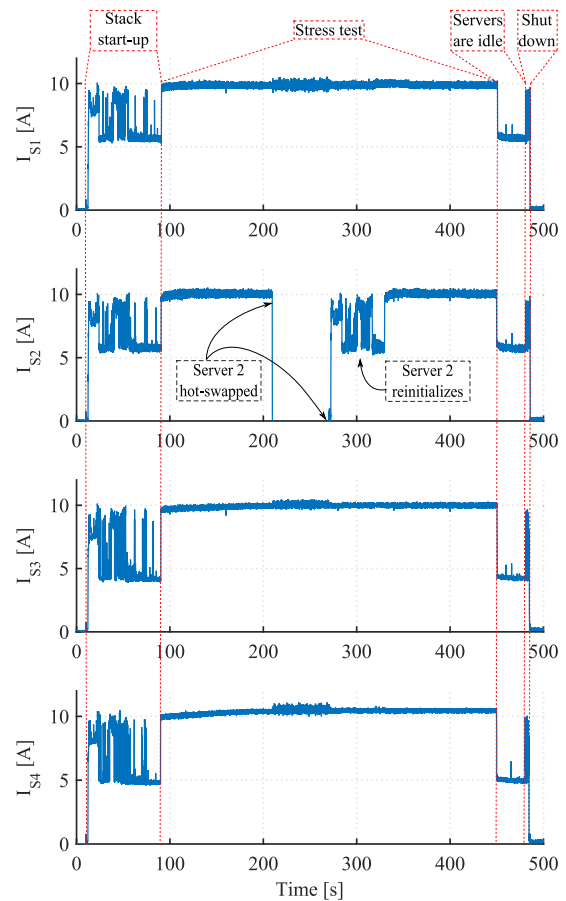


Fig. 15. Measured server currents. (Measured data is 10 ms window averaged for better illustration of the entire test on a single plot.) Major events of the experiment are annotated on the plot. Note the absence of the second server current when it is hot-swapped.

Fig. 8. The server turns ON around the 12.3 s mark. Although this mechanism is demonstrated for only the second server here, similar behavior is observed in every server in the series-stack.

Figures 15 and 16 show the measured and 10 ms window averaged server currents and voltages, and Fig. 17 shows the measured and 10 ms window averaged virtual bus voltage during the entire test. Following the swap-in of all servers at the 10th second, the initialization of the servers starts at approximately 12 s and takes approximately 80 s. During this time interval, the system-level efficiency is measured as 97.2%. The computation test is started on the servers at the 90th second. During the first 2 minutes of the test, all four servers are executing the computation test and the system-level efficiency is measured as 98.4%. Then around the 210th second, the second server is swapped out from the series-stack, and is kept isolated from the series-stack for approximately 1 min. As shown in Figs. 15 and 16, the differential converters are able to regulate the operating server voltages and the virtual bus voltage while the second server's current and voltage are zero. During this time interval, the system-level efficiency decreases to 95% because the differential converter of the second server processes the full bus current and acts as a dc voltage sink by regulating $v_{Stack,2}$. The second server is swapped in to the series-stack and reinitialized around the 270th second, while the other servers are still executing the computa-

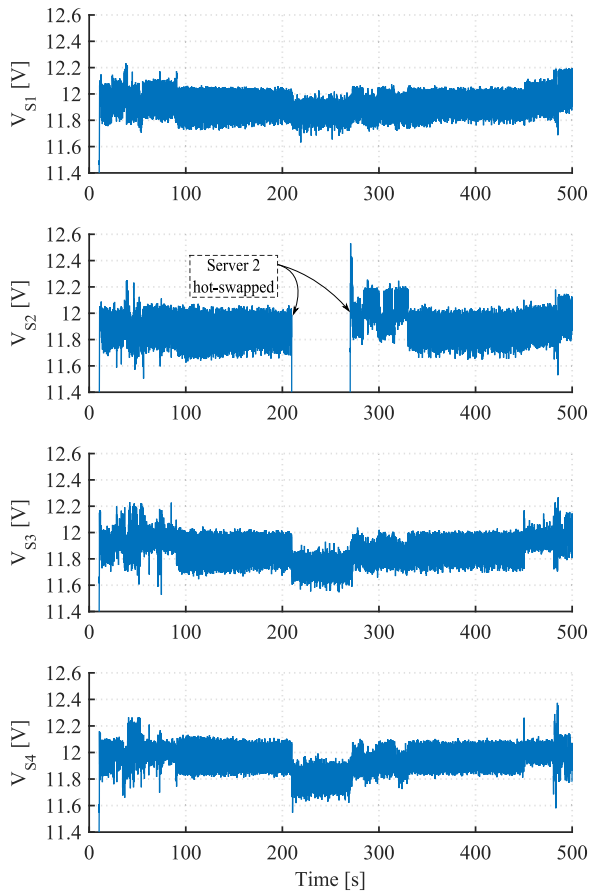


Fig. 16. Measured server voltages. (Measured data is 10 ms window averaged for better illustration of the entire test on a single plot.) Note the absence of the second server voltage when it is hot-swapped.

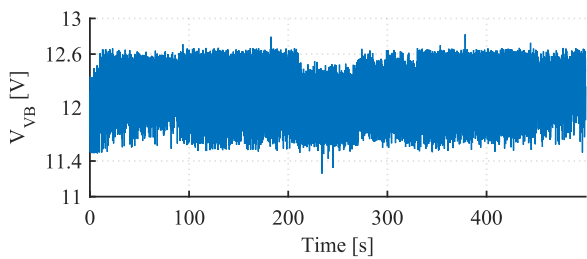


Fig. 17. Measured virtual bus voltage that shows its successful regulation. (Measured data is 10 ms window averaged for better illustration of the entire test on a single plot.)

tion test. During this time interval, the system-level efficiency is calculated as 97.9%. After the second server's reinitialization is completed, the computation test continues on all four servers for 2 more minutes and the system-level efficiency during the last minute of the test is calculated as 98.3%. After the stress test is completed, the servers are kept in their idle state before the shutdown command is executed around the 480th second. During this time interval the system-level efficiency is 96.9%. The server currents immediately go to zero, however, the server voltages are regulated to 12 V until all servers are isolated from the series-stack by using the hot-swapping circuitry. The series-stacked system then returns to its initial state by reactivating the

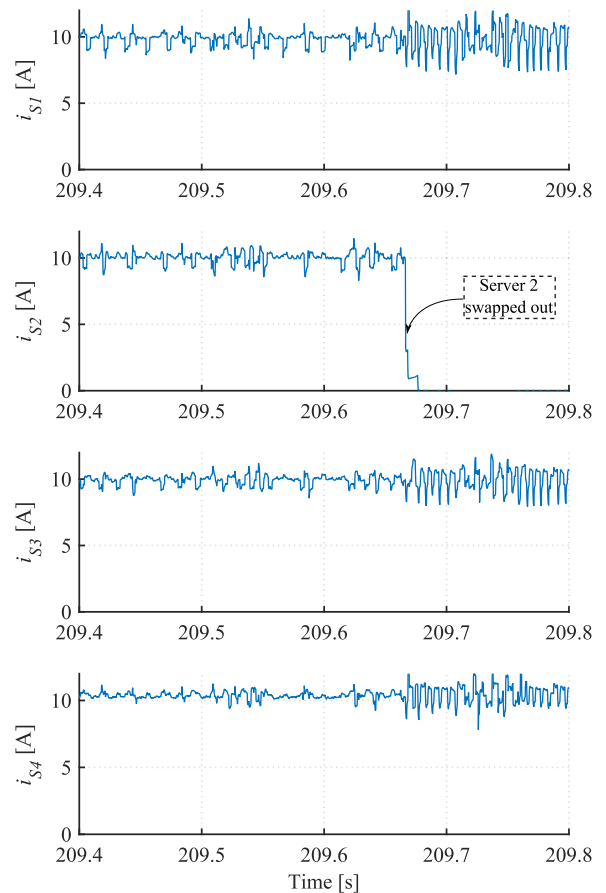


Fig. 18. Instantaneous server currents during swap out.

shunt resistors to allow safe voltage transient as the dc bus is disconnected from the series-stack. A breakdown of the average input and output powers, the efficiency, and the average power loss are given in Table VI for the entire test.

Figures 18 and 19 plot the instantaneous server currents and voltages during the swap-out of the second server around the 210th second. In order to demonstrate and explain the operation of the bidirectional hysteresis algorithm, V_{Stack2} is plotted along with V_{S2} in Fig. 19, and also the instantaneous differential currents into the virtual bus node are plotted in Fig. 20. Before the second server is swapped out around the 210th second, all server voltages are regulated to a hysteresis band (as shown in Fig. 19), while the differential converters are operating in light-load mode by bidirectional hysteresis control (as shown in Fig. 20). Right after the second server is swapped out, the light-load operation of the second differential converter is not sufficient to regulate V_{Stack2} within the same hysteresis band as before. The second differential converter thus switches to full-load mode with increased hysteresis bands, while the other differential converters are still able to regulate their server voltages while mostly maintaining their light-load operation mode as before the swap-out occurred. Note that the frequency of the server voltage ripples slightly increase during the hot-swapped operation. This indicates that the differential converters turn ON and OFF more often than the normal operation, which aligns with increased average power loss during the hot-swapped operation.

TABLE VI
BREAKDOWN OF THE AVERAGE INPUT AND OUTPUT POWERS, THE EFFICIENCY, AND THE AVERAGE POWER LOSS DURING THE EXPERIMENT

	Stack Start-up	Stress Test	Server 2 Hot-Swapping	Server 2 Start-up	Stress Test	Shut Down
Time interval [s]	$0 < t < 90$	$90 < t < 210$	$210 < t < 270$	$270 < t < 330$	$330 < t < 450$	$450 < t < 500$
$\langle P_{in} \rangle$ [W]	264.9	481.9	378.8	452.5	488.5	192.6
$\langle P_{out} \rangle$ [W]	257.1	472.8	358.6	441.7	479.2	186.4
$\langle P_{loss, meas.} \rangle$ [W]	0.5	1.2	1.2	1.1	1.2	0.3
$\langle P_{loss, HS} \rangle$ [W]	1.1	1.6	1.1	1.5	1.6	0.5
$\langle P_{loss, cabling} \rangle$ [W]	0.9	2.4	1.6	2.2	2.5	0.5
$\langle P_{loss, conv.} \rangle$ [W]	5.3	3.9	16.3	6.0	4.0	4.9
$\langle P_{loss, sys} \rangle$ [W]	7.3	7.9	19.0	9.7	8.1	5.9
$\langle \eta_{sys} \rangle$ [%]	97.2	98.4	95.0	97.9	98.3	96.9
$\langle \eta_{conv.} \rangle$ [%]	98.0	99.2	95.7	98.7	99.2	97.5

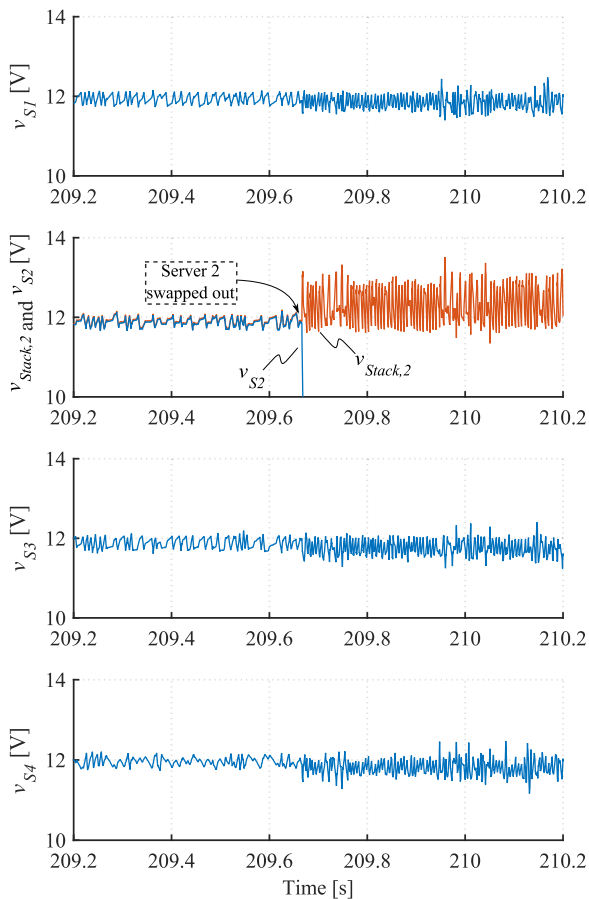


Fig. 19. Instantaneous server voltages during swap out.

The instantaneous server currents and voltages during the swap-in of the second server around the 270th second are also illustrated in Figs. 21 and 22, respectively. Starting from the 269th second, the voltage and current of the second server increase in a controlled manner, similar to the demonstration in Fig. 14. Following the 270th second, the second server initializes; however, the second differential converter still remains in full-load hysteresis mode as can be seen in Fig. 23 since the second server's power consumption during initialization is quite different from that of the remaining servers that continue the stress test.

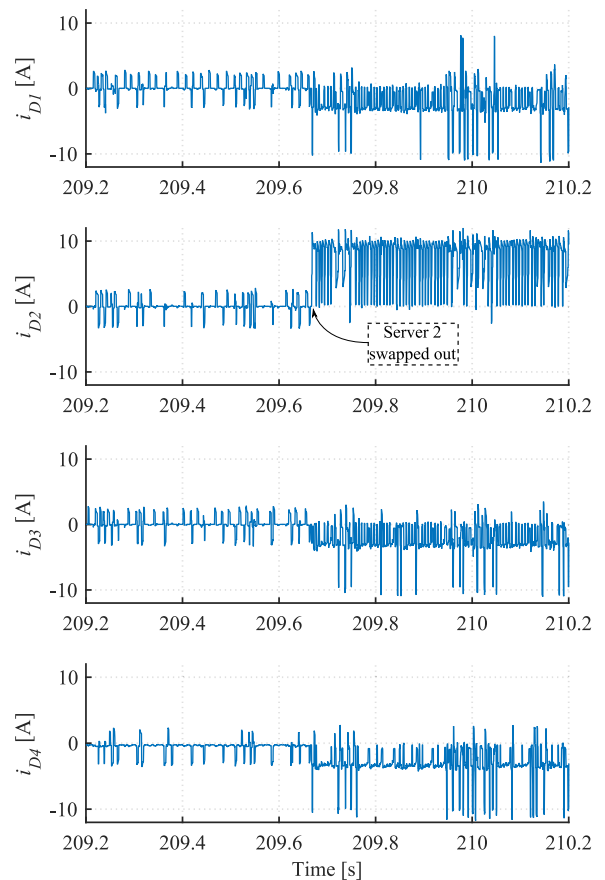


Fig. 20. Instantaneous differential currents into the virtual bus node during swap out.

The series-stacked architecture separates the power processed from the power delivered. By processing only the difference power throughout the experiment, a power conversion efficiency that is always higher than the efficiency of the differential converter is achieved. When the servers are almost equally loaded during the stress test (i.e., $90 < t < 210$ and $330 < t < 450$), the differential converters are processing only insignificant amounts of power in the system, yielding above 99% power conversion efficiency. Further note that by leveraging the hysteresis mode control, the differential converters are kept OFF whenever needed in order to avoid the instantaneous power conversion, as

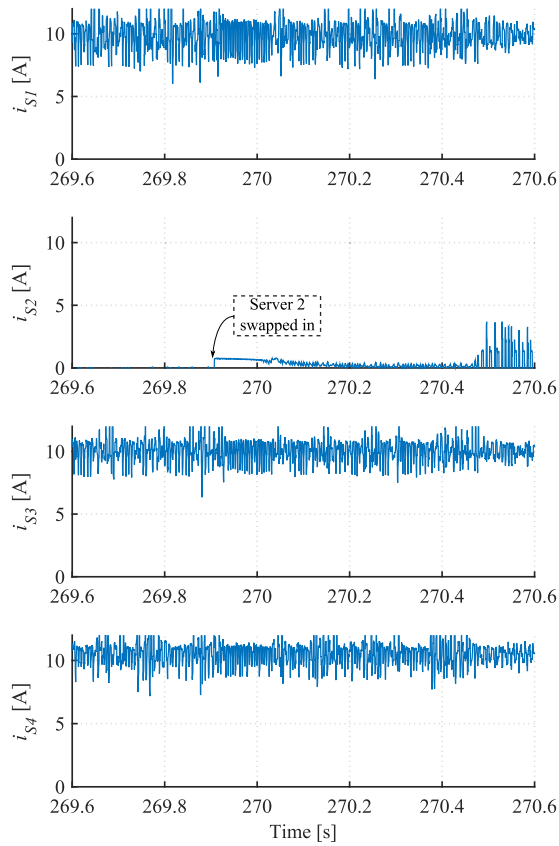


Fig. 21. Instantaneous server currents during swap in.

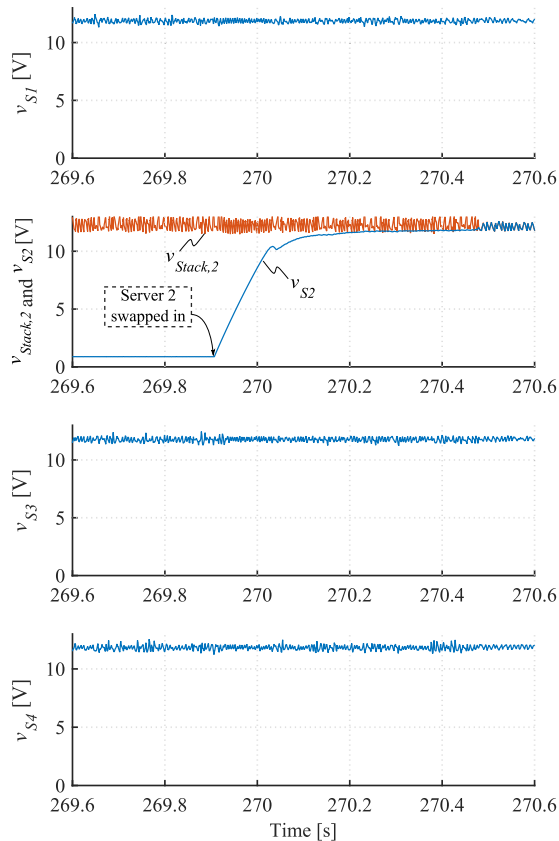


Fig. 22. Instantaneous server voltages during swap in.

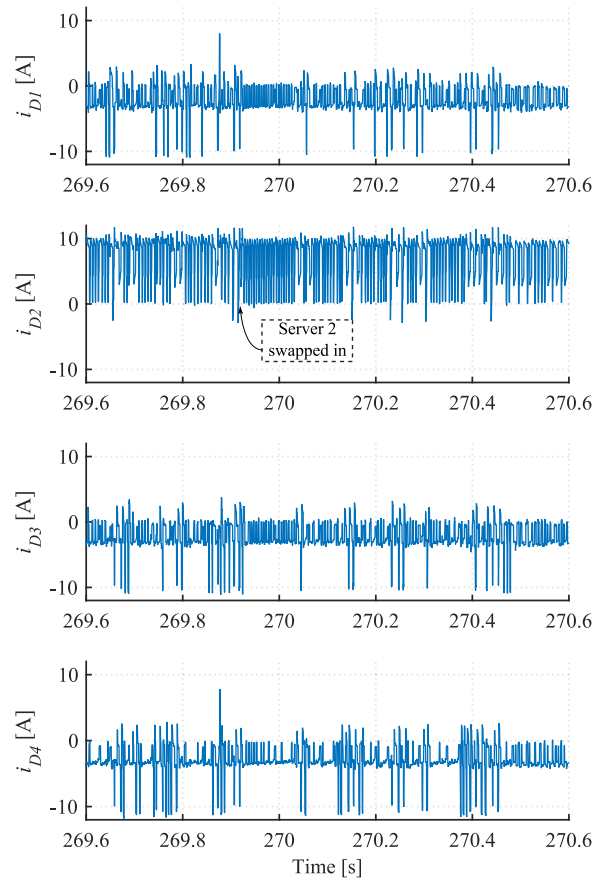
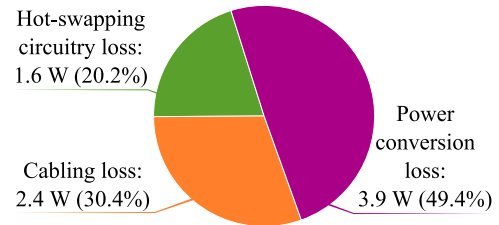


Fig. 23. Instantaneous differential currents into the virtual bus node during swap in.

Fig. 24. Pie chart of $P_{\text{loss,sys}}$ distribution during the stress test $90 < t < 210$. During this time interval, an average of 472.8 W was delivered to the servers.

shown in the first 400 ms of Fig. 20. The average power loss distribution while the series-stacked servers are executing the stress test between 90 and 210 s is also demonstrated in a pie chart in Fig. 24. Here, the power conversion losses are reduced to almost half of the overall losses in the system, while the other half is shared as conduction loss between the hot-swapping circuit and cabling.

VII. CONCLUSION AND FUTURE WORK

In this paper, we have explained the details of the hot-swapping operation in the series-stacked server power delivery architecture. The presented experimental results are the first successful demonstration of hot-swapped operation in series-stacked servers. The practical challenges of the series-stacked system (i.e., the server initialization, hot-swapping, and isolation

of the servers) are discussed and addressed in the experimental work with up to 99.2% power conversion efficiency.

Future work includes unregulated dc bus operation of the series-stacked architecture employing a UPS at the dc bus terminal, an improved control algorithm to achieve the operation of the server-to-virtual bus DPP architecture after a loss of power in the rack, and a server-to-bus DPP architecture in which the secondary side of the differential converters is connected to the dc bus instead of a virtual bus capacitor.

REFERENCES

- [1] J. Whitney and P. Delforge, "Data center efficiency assessment," Natural Resources Def. Council, New York, NY, USA, Tech. Rep. IP:14-08-a, Aug. 2014.
- [2] Data center 2025: Exploring the possibilities. Emerson Network Power, Columbus, OH, USA, 2014. [Online]. Available: http://www.emersonnetworkpower.com/en-US/Latest-Thinking/Data-Center-2025/Documents/002401_DataCenter2025Report_HR_INTERACTIVE.PDF
- [3] *Intel Server Board S2600GZ and GL Technical Product Specification, 2nd ed.*, Intel, Santa Clara, CA, USA, Feb. 2014. [Online]. Available: http://download.intel.com/support/motherboards/server/sb/s2600gzgl_tps_r2_1.pdf
- [4] *Sun Fire V120 and Netra 120 Server User's Guide*, Sun Microsystems, Santa Clara, CA, USA, Dec. 2001. [Online]. Available: <https://docs.oracle.com/cd/E19088-01/v120.srvr/816-2090-10/816-2090-10.pdf>
- [5] F. Xu, B. Guo, L. M. Tolbert, F. Wang, and B. J. Blalock, "An all-SiC three-phase buck rectifier for high-efficiency data center power supplies," *IEEE Trans. Ind. Appl.*, vol. 49, no. 6, pp. 2662–2673, Nov. 2013.
- [6] F. Xu, B. Guo, Z. Xu, L. M. Tolbert, F. Wang, and B. J. Blalock, "Paralleled three-phase current-source rectifiers for high-efficiency power supply applications," *IEEE Trans. Ind. Appl.*, vol. 51, no. 3, pp. 2388–2397, May 2015.
- [7] L. Schrittwieser, J. W. Kolar, and T. B. Soeiro, "99% efficient three-phase buck-type sic MOSFET PFC rectifier minimizing life cycle cost in DC data centers," in *Proc. IEEE Int. Telecommun. Energy Conf.*, Oct. 2016, pp. 1–8.
- [8] C. Zhao, X. Wu, P. Meng, and Z. Qian, "Optimum design consideration and implementation of a novel synchronous rectified soft-switched phase-shift full-bridge converter for low-output-voltage high-output-current applications," *IEEE Trans. Power Electron.*, vol. 24, no. 2, pp. 388–397, Feb. 2009.
- [9] J. Biela, U. Badstuebner, and J. W. Kolar, "Design of a 5-kw, 1-u, 10-kw/dm³ resonant DC-DC converter for telecom applications," *IEEE Trans. Power Electron.*, vol. 24, no. 7, pp. 1701–1710, Jul. 2009.
- [10] M. Salato and P. Makrum, "350v to 12v DC "yeaman topology" power system," in *Proc. Int. Conf. Energy Aware Comput.*, Dec. 2010, pp. 1–4.
- [11] Y. D. Kim, K. M. Cho, D. Y. Kim, and G. W. Moon, "Wide-range ZVS phase-shift full-bridge converter with reduced conduction loss caused by circulating current," *IEEE Trans. Power Electron.*, vol. 28, no. 7, pp. 3308–3316, Jul. 2013.
- [12] D. Costinett, H. Nguyen, R. Zane, and D. Maksimovic, "GaN-FET based dual active bridge DC-DC converter," in *Proc. 26th Annu. IEEE Appl. Power Electron. Conf. Expo.*, Mar. 2011, pp. 1425–1432.
- [13] W. Zhang *et al.*, "Wide bandgap power devices based high efficiency power converters for data center application," in *Proc. IEEE Workshop Wide Bandgap Power Devices Appl.*, Oct. 2014, pp. 121–126.
- [14] D. Reusch and J. Strydom, "Evaluation of gallium nitride transistors in high frequency resonant and soft-switching DC-DC converters," *IEEE Trans. Power Electron.*, vol. 30, no. 9, pp. 5151–5158, Sep. 2015.
- [15] Netsure 4015 30kw 400v dc power system. Emerson Network Power, Columbus, OH, USA. [Online]. Available: <http://www.emersonnetworkpower.com/documentation/en-us/brands/netsure/documents/netsure-4015-system.pdf>
- [16] J. McClurg, R. Pilawa-Podgurski, and P. Shenoy, "A series-stacked architecture for high-efficiency data center power delivery," in *Proc. IEEE Energy Convers. Congr. Expo.*, Sep. 2014, pp. 170–177.
- [17] E. Candan, P. S. Shenoy, and R. C. N. Pilawa-Podgurski, "A series-stacked power delivery architecture with isolated differential power conversion for data centers," *IEEE Trans. Power Electron.*, vol. 31, no. 5, pp. 3690–3703, May 2016.
- [18] A. Stillwell and R. C. N. Pilawa-Podgurski, "A resonant switched-capacitor converter with GaN transistors for series-stacked processors with 99.8% power delivery efficiency," in *Proc. IEEE Energy Convers. Congr. Expo.*, Sep. 2015, pp. 563–570.
- [19] D. Das and P. T. Krein, "Voltage regulation of a series-stacked system of processors by differential power processing," in *Proc. IEEE 16th Workshop Control Model. Power Electron.*, Jul. 2015, pp. 1–7.
- [20] E. Candan, A. Stillwell, and R. C. N. Pilawa-Podgurski, "A reliability assessment of series-stacked servers with server-to-bus differential power processing," in *Proc. IEEE Int. Telecommun. Energy Conf.*, Oct. 2016, pp. 1–7.
- [21] E. Candan, D. Heeger, P. S. Shenoy, and R. C. N. Pilawa-Podgurski, "A series-stacked power delivery architecture with hot-swapping for high-efficiency data centers," in *Proc. IEEE Energy Convers. Congr. Expo.*, Sep. 2015, pp. 571–578.
- [22] P. S. Shenoy and P. T. Krein, "Differential power processing for dc systems," *IEEE Trans. Power Electron.*, vol. 28, no. 4, pp. 1795–1806, Apr. 2013.
- [23] P. Shenoy, K. Kim, B. Johnson, and P. Krein, "Differential power processing for increased energy production and reliability of photovoltaic systems," *IEEE Trans. Power Electron.*, vol. 28, no. 6, pp. 2968–2979, Jun. 2013.
- [24] S. Rajapandian, K. L. Shepard, P. Hazucha, and T. Karnik, "High-voltage power delivery through charge recycling," *IEEE J. Solid-State Circuits*, vol. 41, no. 6, pp. 1400–1410, Jun. 2006.
- [25] S. Qin and R. C. Pilawa-Podgurski, "Sub-module differential power processing for photovoltaic applications," in *Proc. 28th Ann. IEEE Appl. Power Electron. Conf. Expo.*, Mar. 2013, pp. 101–108.
- [26] C. Olalla, D. Clement, M. Rodriguez, and D. Maksimovic, "Architectures and control of submodule integrated DC-DC converters for photovoltaic applications," *IEEE Trans. Power Electron.*, vol. 28, no. 6, pp. 2980–2997, Jun. 2013.
- [27] Y. Levron, D. Clement, D. Maksimovic, and C. Olalla, "Nonlinear control design for the photovoltaic isolated-port architecture with submodule integrated converters," in *Proc. IEEE Energy Convers. Congr. Expo.*, Sep. 2013, pp. 2398–2405.
- [28] C. Olalla, C. Deline, and D. Maksimovic, "Performance of mismatched PV systems with submodule integrated converters," *IEEE J. Photovolt.*, vol. 4, no. 1, pp. 396–404, Jan. 2014.
- [29] C. Olalla, C. Deline, D. Clement, Y. Levron, M. Rodriguez, and D. Maksimovic, "Performance of power-limited differential power processing architectures in mismatched PV systems," *IEEE Trans. Power Electron.*, vol. 30, no. 2, pp. 618–631, Feb. 2015.
- [30] S. Qin, S. T. Cady, A. D. Domnguez-Garca, and R. C. N. Pilawa-Podgurski, "A distributed approach to maximum power point tracking for photovoltaic submodule differential power processing," *IEEE Trans. Power Electron.*, vol. 30, no. 4, pp. 2024–2040, Apr. 2015.
- [31] R. Bell and R. C. N. Pilawa-Podgurski, "Decoupled and distributed maximum power point tracking of series-connected photovoltaic submodules using differential power processing," *IEEE J. Emerg. Sel. Top. Power Electron.*, vol. 3, no. 4, pp. 881–891, Dec. 2015.
- [32] S. Qin, C. B. Barth, and R. C. N. Pilawa-Podgurski, "Enhancing microinverter energy capture with submodule differential power processing," *IEEE Trans. Power Electron.*, vol. 31, no. 5, pp. 3575–3585, May 2016.
- [33] M. Einhorn *et al.*, "A current equalization method for serially connected battery cells using a single power converter for each cell," *IEEE Trans. Veh. Technol.*, vol. 60, no. 9, pp. 4227–4237, Nov. 2011.
- [34] M. M. U. Rehman, F. Zhang, M. Evzelman, R. Zane, and D. Maksimovic, "Control of a series-input, parallel-output cell balancing system for electric vehicle battery packs," in *Proc. IEEE 16th Workshop Control Model. Power Electron.*, Jul. 2015, pp. 1–7.
- [35] C. Schaefer and J. T. Stauth, "Efficient voltage regulation for microprocessor cores stacked in vertical voltage domains," *IEEE Trans. Power Electron.*, vol. 31, no. 2, pp. 1795–1808, Feb. 2016.
- [36] E. Candan, P. S. Shenoy, and R. Pilawa-Podgurski, "A distributed bi-directional hysteresis control algorithm for server-to-virtual bus differential power processing," in *Proc. IEEE 16th Workshop Control Model. Power Electron.*, Jul. 2015, pp. 1–8.
- [37] B. Davies, "Analysis of inrush currents for DC powered IT equipment," in *Proc. IEEE 33th Int. Telecommun. Energy Conf.*, Oct. 2011, pp. 1–4.
- [38] D. Olson, "Hot swap power management," 2001, Texas Instruments, Dallas, TX, USA.
- [39] *Texas Instruments, Dallas, TX, USA*, "Tps2490 positive high-voltage power-limiting hotswap controller," Datasheet, Jul. 2012.

- [40] M. Kheraluwala, R. Gascoigne, D. Divan, and E. Baumann, "Performance characterization of a high-power dual active bridge DC-to-DC converter," *IEEE Trans. Ind. Appl.*, vol. 28, no. 6, pp. 1294–1301, Nov. 1992.
- [41] F. Krismer and J. Kolar, "Accurate power loss model derivation of a high-current dual active bridge converter for an automotive application," *IEEE Trans. Ind. Electron.*, vol. 57, no. 3, pp. 881–891, Mar. 2010.
- [42] A. Waterland, "Stress POSIX workload generator," 2014. [Online]. Available: <http://people.seas.harvard.edu/~apw/stress>



Enver Candan (S'12) received the dual B.S. degrees in control engineering and in electrical engineering from Istanbul Technical University, Istanbul, Turkey, both in 2012. He received the M.S. degree in electrical engineering in 2014 under the Fulbright Scholarship from the University of Illinois at Urbana-Champaign, Urbana, IL, USA, where he is currently working toward the Ph.D. degree in high efficiency dc–dc power conversion.

In the summer of 2015, he was a hardware engineering intern at Google Inc. His research interests include data center power delivery systems and isolated dc–dc converters.



Derek Heeger (S'14) received the B.S. degree in electrical engineering from Rose-Hulman Institute of Technology, Terre Haute, IN, USA, in 2014, and the M.S. degree in electrical engineering from the University of Illinois, Urbana-Champaign, Champaign, IL, USA, in 2016, under the National Physical Science Consortium fellowship.

He is currently an Engineer in the Sandia National Labs, where he is working on the battery management systems. His research interests include battery management systems, compact and efficient power electronics, and switched-capacitor converters.



Pradeep S. Shenoy (S'06–M'12) received the B.S. degree in electrical engineering from the Illinois Institute of Technology, Chicago, IL, USA, in 2007, and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois, Urbana-Champaign, Urbana, IL, in 2010 and 2012, respectively.

In 2008, he participated in the National Science Foundations East Asia and Pacific Summer Institutes Program conducting research at Tsinghua University, Beijing, China. He was an intern in the Caterpillars Electric Power Division in 2005 and in the Texas Instruments Systems and Applications R&D Lab in 2011. In 2012, he joined Texas Instruments Kilby Labs.

Dr. Shenoy received the Camras scholarship in 2003–2007 and a Foreign Language and Area Studies fellowship in 2009–2010. He was the Vice Chair of the IEEE Power Engineering Society and Power Electronics Society Joint Student Chapter at the University of Illinois in 2009–2010 and the Co-Director of the 2011 IEEE Power and Energy Conference at Illinois. He currently serves in the IEEE Power Electronics Society as the Regions 1-7 Chair and as an Ad Com Member-at-Large. He received the Illinois International Graduate Achievement Award in 2010 and the Jack Kilby Award for innovation in 2015. He was a finalist for the Lemelson-MIT Illinois Student Prize for innovation in 2012.



Robert C. N. Pilawa-Podgurski (S'06–M'11) received the dual B.S. degrees in physics, electrical engineering, and computer science in 2005, the M.Eng. degree in electrical engineering and computer science (EECS) in 2007, and the Ph.D. degree in electrical engineering in 2012, all from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA.

He is currently an Assistant Professor in the Electrical and Computer Engineering Department, University of Illinois at Urbana-Champaign (UIUC), Champaign, IL, USA, and is affiliated with the Power and Energy Systems group. He performs research in the area of power electronics. His research interests include renewable energy applications, electric vehicles, energy harvesting, CMOS power management, high-density and high-efficiency power converters, and advanced control of power converters.

Dr. Pilawa-Podgurski has been an Associate Editor for the IEEE TRANSACTIONS ON POWER ELECTRONICS, and for the IEEE JOURNAL OF EMERGING AND SELECTED TOPICS IN POWER ELECTRONICS, SINCE 2014. He received the Chorafas Award for outstanding MIT EECS Master's thesis, the Google Faculty Research Award in 2013, and the 2014 Richard M. Bass Outstanding Young Power Electronics Engineer Award of the IEEE Power Electronics Society, for outstanding contributions to the field of power electronics before the age of 35. In 2015, he received the Air Force Office of Scientific Research Young Investigator Award, and in 2016, the UIUC Deans Award for Excellence in Research. He is a co-author of six IEEE prize papers.