Differential Power Processing for Ultra-Efficient Data Storage

Ping Wang, *Student Member, IEEE*, Yenan Chen, *Member, IEEE*, Jing Yuan, *Member, IEEE*, Robert C. N. Pilawa-Podgurski, *Member, IEEE*, and Minjie Chen, *Senior Member, IEEE*

***Abstract*—This article presents the hardware, software, and power codesign of an ultra-efficient data storage server with differential power processing (DPP). DPP can reduce the power conversion stress, improve the efficiency, and enhance the functionality of modular power electronics systems. The power inputs of a large number of hard disk drives (HDDs) were connected in series and supported by a multiport ac-coupled differential power processing (MAC-DPP) converter through a multiwinding transformer. Methods for controlling the multi-input multi-output power flow in the multiwinding transformer while avoiding core saturation were investigated. A ten-port MAC-DPP prototype with 700-W/in^3 power density was built to support a 450-W HDD storage system with ten series-stacked voltage domains. The prototype was tested on a 50-HDD server testbench, and the overall system loss is below 1 W (99.77% system efficiency). The server was able to maintain high-speed reading and writing operation of all 50 HDDs against the worst hot-swapping scenarios. A variety of hardware/software configurations and many cloud storage techniques were tested on the fully functioning server. Experimental results show that the energy efficiency of large-scale information systems (CPU/GPU clusters, memory banks, HDD arrays, etc.) can be greatly improved by software, hardware, and power codesign.**

***Index Terms*—Data center, differential power processing (DPP), distributed control, energy-efficient computing, multiport converter, multiwinding transformer.**

I. Introduction

ARTIFICIAL intelligence, cloud computing, and Internet of things applications have stimulated explosive growth in high-performance computing and data center infrastructure. Data centers currently contribute about 2% of the U.S. total electricity [1]. A recent International Data Corporation report estimated that the global datasphere will grow from 33 ZB in 2018 to 175 ZB by 2025 [2]. To keep up with the rapidly growing storage demands, data storage systems, one of the major power demand infrastructure in data centers, need efficient power delivery solutions. High-efficiency and high-power-density power electronics are needed to maximize the storage capacity per unit volume and to support the efficient operation and sustainable development of data storage systems.

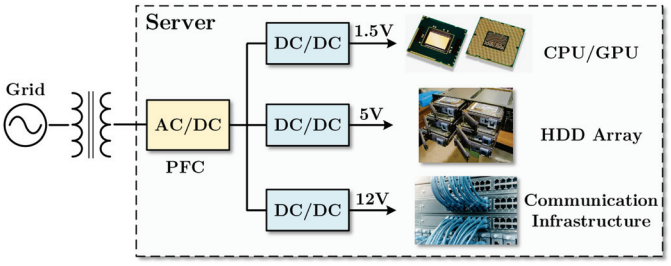


Fig. 1. Conventional power deliver architecture in data centers. Power from the grid is delivered through multiple stages to the low-voltage loads.

The hardware, software, and power architectures in a data storage system are usually designed independently. Storage servers nowadays are still using a classic power delivery architecture developed for the single-server scenario—each server is connected to an ac voltage bus through an ac–dc power factor correction converter followed by multiple dc–dc converters for a variety of information technology (IT) equipment [e.g., 0.8–12 V for CPUs, RAMs, and hard disk drives (HDDs)], as shown in Fig. 1. In this multistage architecture, the overall system efficiency tends to be low, as the full load power is processed sequentially by each stage. It is challenging to design high voltage conversion ratio dc–dc converters with high efficiency and high power density, especially if galvanic isolation is needed [4].

A recent trend in data center power architecture is to distribute 48–54 V dc power on the rack level [5], [6]. A dc voltage bus is created and an uninterruptible power supply is placed on the rack. The dc distribution approach reduces the power conversion stages and improves energy efficiency. Compared to a traditional 12-V intermediate bus architecture, delivering power at 48–54 V dc bus can reduce the conduction loss and leverage the existing 48-V telecom power ecosystem. To deliver power from the 48-V dc voltage bus to low-voltage IT equipment, conventional power architecture employs numerous dc–dc converters with a

Manuscript received May 4, 2020; revised August 11, 2020; accepted August 19, 2020. Date of publication September 7, 2020; date of current version November 20, 2020. This work was supported in part by the Advanced Research Projects Agency-Energy, U.S. Department of Energy, under Award No. DE-AR0000906 in the CIRCUITS program in part by the National Science Foundation CAREER Award No. 1847365, and in part by the Princeton E-ffiliates Partnership Program. This paper was presented at the 2019 IEEE Energy Conversion Congress and Exposition [3]. Recommended for publication by Associate Editor J. He. *(Corresponding author: Minjie Chen.)*

Ping Wang, Yenan Chen, Jing Yuan, and Minjie Chen are with the Department of Electrical Engineering and the Andlinger Center for Energy and the Environment, Princeton University, Princeton, NJ 08540 USA (e-mail: ping.wang@princeton.edu; yenanc@princeton.edu; yua@et.aau.dk; minjie@princeton.edu).

Robert C. N. Pilawa-Podgurski is with the Department of Electrical Engineering and Computer Sciences, University of California Berkeley, Berkeley, CA 94720 USA (e-mail: pilawa@berkeley.edu).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPEL.2020.3022089

0885-8993 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Princeton University. Downloaded on January 24,2021 at 15:53:24 UTC from IEEE Xplore. Restrictions apply.

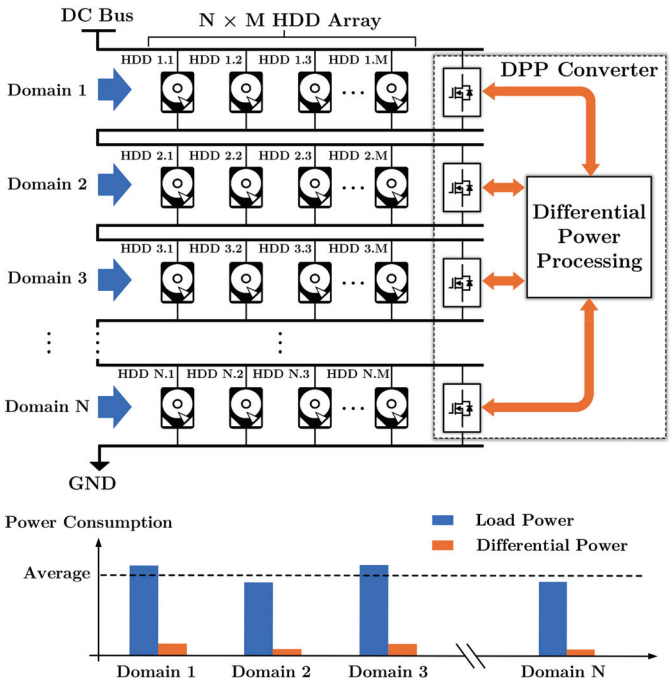


Fig. 2. Data storage server with series-stacked power delivery architecture. It comprises a cluster of $N \times M$ HDDs divided into N series-stacked voltage domains with DPP.

variety of output voltage levels, and full load power needs to be processed by these dc–dc converters. In data storage servers, HDDs and solid-state drives (SSDs) are highly modular with uniform voltage ratings (3.3, 5, or 12 V) and similar power consumption; there are opportunities to adopt series-stacked power delivery with differential power processing (DPP) to realize inherent voltage step down [7].

DPP has been proved effective in a wide range of applications, including solar photovoltaic converters [8]–[13], battery balancers [14]–[16], computers, and servers [17]–[19]. In this article, for the first time, DPP is applied to data storage servers, enabling holistic codesign of hardware, software, and power architectures. Fig. 2 illustrates the key principles of a data storage server with DPP architecture. N voltage domains are connected in series to the dc bus. Each voltage domain supports M HDDs connected in parallel. The HDDs in each voltage domain consume similar load power with little power difference. Thus, the vast majority of power is directly delivered to the loads, and only a small amount of power difference is processed through DPP, yielding significantly reduced power conversion stress and improved energy efficiency. The decrease in processed power of the DPP converter also reduces the converter failure rate, making for more reliable power delivery [18]. The highly uniform load profiles of HDDs and SSDs make DPP attractive in data storage applications.

This article presents the design and implementation of a data storage server with series-stacked DPP. A multiport ac-coupled differential power processing (MAC-DPP) converter is presented to couple all series-stacked voltage domains through a single multiwinding transformer. The proposed isolated fully coupled MAC-DPP architecture features reduced component count, smaller magnetic volume, and lower differential power conversion stages compared to other existing DPP solutions [8]–[19]. Nonisolated fully coupled DPP solutions exist [20], but coupling all ports together through a multiwinding transformer offers the highest modularity and extendability—the DPP architecture can be linearly extended without customizing the design of each port. Other key design considerations of the MAC-DPP architecture, including magnetics, control, and packaging, are also presented.

A 450-W ten-port MAC-DPP prototype was built to support a storage server containing 50 HDDs, which are configured into ten series-stacked voltage domains (5 HDDs×10). High-speed data transfer across different voltage domains was achieved with standard communication protocols (e.g., SAS and SATA). A distributed phase-shift (DPS) control strategy was utilized to route the differential power flow and regulate the voltage of each domain. It was able to maintain the normal operation of the storage server against the worst-case hot-swapping scenario. The storage server was also tested with various storage strategies, including direct storage and many different Redundant Array of Independent Disks (RAID) levels [21]. Experimental results show that the energy efficiency of large-scale information systems can be greatly improved by DPP.

The remainder of this article is structured as follows. Section II compares several different DPP topologies and clarifies their design tradeoffs as well as the advantages of the MAC-DPP architecture. Section III analyzes the fundamental principles of avoiding saturation in the multiwinding transformer. Section IV presents the strategy of controlling multi-input multi-output (MIMO) power flow for voltage regulation. Detailed experimental results are provided in Section V, including the design of a ten-port MAC-DPP prototype and the hardware and software configuration of a 50-HDD storage server testbench. Finally, Section VI concludes this article.

II. MAC-DPP Architecture

Many DPP converter topologies have been proposed. Fig. 3 compares the proposed MAC-DPP architecture against other typical existing DPP solutions. Fig. 3(a) shows a load-to-load DPP architecture, which uses a bidirectional buck–boost circuit to process the differential power between two neighboring loads [10]–[14]. Compared to DPP converters that connect each load to the input dc bus [10]–[12], the load-to-load DPP converter has reduced switch voltage stress ($2V_{\text{load}}$). However, the differential power between two nonadjacent loads has to go through multiple power conversion stages due to the laddered structure. This creates higher power conversion losses and limits the system dynamic performance. Fig. 3(b) shows a resonant ladder switched-capacitor DPP (SC-DPP) topology [9], [19]. The ladder SC-DPP converter can achieve high efficiency and high power density, but during load transient, it can only transfer power between neighboring voltage domains in each switching cycle. If two voltage domains are not directly connected, it

Authorized licensed use limited to: Princeton University. Downloaded on January 24,2021 at 15:53:24 UTC from IEEE Xplore. Restrictions apply.

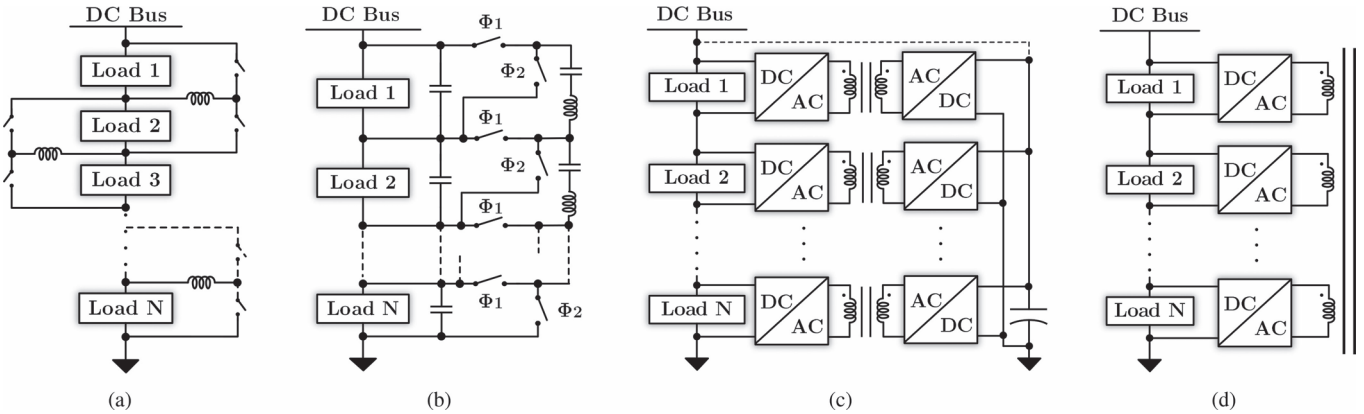


Fig. 3. Circuit diagrams of a few example DPP topologies. (a) Load-to-load DPP. (b) Switched-capacitor DPP. (c) DC-coupled DPP. (d) Proposed MAC-DPP. Besides, the MAC-DPP architecture offers reduced power conversion stress, higher efficiency, smaller magnetic size, and lower component count.

TABLE I
COMPARISON OF SEVERAL TYPICAL DPP TOPOLOGIES

Topology	Load-to-load DPP	Switched-capacitor DPP	DC-coupled DPP (half-bridge)	MAC-DPP (half-bridge)
Switch count	$2N-2$	$2N$	$4N$	$2N$
Switch voltage stress	$2V_{load}$	V_{load}	V_{load} or V_{bus}	V_{load}
Magnetic components	$N-1$ inductors	$N-1$ inductors	N two-winding transformers	one N-winding transformer
Power conversion stages	multiple stages	multiple stages	two "dc-ac-dc" stages	one"dc-ac-dc" stage
Port-to-port isolation	nonisolated	nonisolated	galvanically isolated	galvanically isolated
Publication	[10]–[14]	[9], [19]	[10]–[12], [15]–[18]	This article

takes multiple switching cycles to transfer energy from one domain to the other. An alternative DPP approach is to employ multiple isolated dc–dc converters (e.g., flyback, dual active bridge (DAB), etc.) and connect each voltage domain to a virtual dc bus or an input dc bus, as depicted in Fig. 3(c) [10]–[12], [15]–[18]. The dc-coupled DPP architecture can transfer power directly between two arbitrary loads. Compared to laddered-structure based DPP options [see Fig. 3(a) and (b)], this architecture is more scalable and can offer better dynamic performance. However, the dc-coupled DPP topology requires multiple magnetic elements (i.e., transformers) as well as high component count, which increases the cost and total converter size. Moreover, the differential power needs to go through at least two "dc–ac–dc" stages from one port to another, resulting in additional power conversion stress and losses [22].

As shown in Fig. 3(d), the proposed MAC-DPP architecture connects each voltage domain to a multiwinding transformer through a dc–ac unit. The differential power of each voltage domain is coupled to the multiwinding transformer. The dc–ac inverter can be implemented as a half-bridge inverter with a dc blocking capacitor. Other dc–ac inverter circuits, such as full-bridge inverters, or Class-E-based inverters, are also applicable [23]. The power transferred between two different loads is galvanically isolated and is bidirectional. Table I lists the detailed comparison of different DPP architectures. Parameters are calculated assuming half-bridge implementation for all dc–ac units. The advantages of the proposed MAC-DPP architecture include the following.

1) *Fewer "dc–ac–dc" Power Conversion Stages:* The MAC-DPP architecture directly transfers power between two arbitrary ports with one single "dc–ac–dc" conversion stage. Existing DPP solutions usually need two or more "dc–ac–dc" stages when delivering power between two arbitrary loads. The reduced power conversion stress improves the system dynamic performance and reduces the losses.
2) *Reduced Component Count:* In the MAC-DPP architecture, one voltage domain is connected to one dc–ac unit, and n voltage domains only need n dc–ac units, which are reduced by half compared with the dc-coupled DPP architecture. Besides, the MAC-DPP architecture is highly modular. Its component count is among the lowest of the existing DPP options, leading to reduced cost and improved power density.
3) *Smaller Magnetic Size:* Compared to the dc-coupled DPP converter that needs multiple transformers, the MAC-DPP architecture has only one magnetic core. In principle, the magnetic core area of a multiwinding transformer is determined by the highest volt–second per turn of all windings instead of the winding count and is not directly related to the number of windings. In a MAC-DPP architecture with a fully symmetric configuration, each dc–ac unit has an identical voltage rating, and all windings have identical volt–second per turn, which will stay the same as the winding count increases. Therefore, the core area of a multiwinding transformer in the MAC-DPP is roughly the same as that of a two-winding transformer in other

Authorized licensed use limited to: Princeton University. Downloaded on January 24,2021 at 15:53:24 UTC from IEEE Xplore. Restrictions apply.

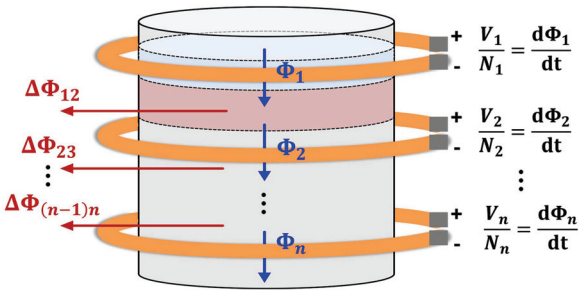


Fig. 4. Magnetic flux in the magnetic core of a multiwinding transformer with a single magnetic linkage. Φ_i is the magnetizing flux and $\Delta\Phi_{ij}$ is the leakage flux.

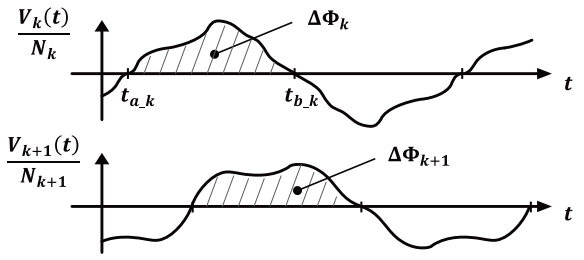


Fig. 5. Waveforms of winding volt-per-turn and peak–peak flux variation.

isolated DPP options. Only the window area increases as the winding count increases. Theoretically, the MAC-DPP architecture can reduce the magnetic core area by n times compared to other isolated DPP implementations (n is the number of series-stacked voltage domains).

Nevertheless, the main purpose of this article is to demonstrate the effectiveness of DPP architecture for ultra-efficient data storage. While a fully coupled MAC-DPP topology is considered as attractive and selected for prototyping, other DPP topologies are also applicable with a variety of tradeoffs.

III. Multiwinding Transformer Design

One challenge of designing a MAC-DPP converter is to build a high-performance miniaturized multiwinding transformer with a single magnetic linkage. A basic requirement is to effectively couple all windings without saturating the magnetic core. In a two-winding transformer, the cross-sectional area of the core is determined by the maximum volt–second per turn in the windings. Here, this rule is extended to the generalized multiwinding cases. Fig. 4 shows the magnetic flux diagram in the magnetic core of the multiwinding transformer. There are two types of magnetic flux in the core: 1) magnetizing flux, which is coupled with each individual winding: Φ_i; and 2) leakage flux, which leaks out through the spacing between two windings: $\Delta\Phi_{ij} = \Phi_i - \Phi_j$. The magnetizing flux of a specific coupled winding is linked to the $V_k(t)/N_k$ (volt per turn) by Faraday's law.

Fig. 5 shows two example arbitrary periodic waveforms of the voltage at two windings. The shaded area (volt–second per turn) is the peak–peak flux variation within one period. The maximum magnetizing flux in the core is

$$\Phi_M^{\max} = \frac{1}{2} \times \max_{k=1,\ldots,n} \{\Delta\Phi_k\} = \frac{1}{2} \times \max_{k=1,\ldots,n} \left\{ \int_{t_{a\_k}}^{t_{b\_k}} \frac{V_k(t)}{N_k} dt \right\}. \quad (1)$$

The maximum leakage flux in the core is

$$\Phi_L^{\max} = \frac{1}{2} \times \max_{k=1,\ldots,n-1} \left\{ \int_{t_{\text{pos}}} \left(\frac{V_k(t)}{N_k} - \frac{V_{k+1}(t)}{N_{k+1}} \right) dt \right\} \quad (2)$$

where t_{pos} represents the time period of the positive integral.

Based on (1) and (2), the maximum flux density in a multiwinding transformer (with a single flux linkage) is located at the spacing between two windings if the winding voltages have opposite phases (i.e., 180° phase shift; assuming equal volt-per-turn amplitudes at all ports). As the phase shift between two winding voltages increases from 0° to 180°, the peak flux density in the spacing area will increase. Therefore, to avoid saturating the core, the minimum core area should be designed for the maximum volt–second per turn, and the spacing distance between two windings should be designed for the opposite-phase case or the maximum phase shift if it is below 180°. Whether a core will saturate or not is independent of the number of windings. A large number of windings driven by different voltage sources can be coupled to a single magnetic linkage without saturating the core, as long as the maximum volt–second per turn does not exceed the designed limit. Extended discussions on saturation and finite-element modeling (FEM) results are presented in Appendix I.

If all windings are driven by square-wave voltage sources with the same volt-per-turn amplitude V_0 and period T, the maximum magnetizing flux in the core is

$$\Phi_{\max} = \frac{1}{2} \int_{\frac{T}{2}} V_0 dt = \frac{1}{4} V_0 T. \quad (3)$$

The maximum magnetizing flux is independent of the number of windings, n, and is only determined by the maximum volt–second per turn (V_0T) of all windings. Accordingly, the minimum core area ($A_{\min}$) of a multiwinding transformer driven by an arbitrary number of square-wave voltage sources with amplitude of V_0 is

$$A_{\min} = \frac{\Phi_{\max}}{B_{\text{sat}}} = \frac{V_0 T}{4 B_{\text{sat}}}. \quad (4)$$

Therefore, coupling many voltage domains with a single linkage multiwinding transformer can significantly reduce the required magnetic core volume of a multiport topology. This is the fundamental reason why the proposed MAC-DPP architecture can achieve much higher power density and better magnetic utilization than other isolated DPP implementations. Compared to nonisolated DPP options without transformers, the MAC-DPP architecture also offers reduced power conversion stress (fewer

Authorized licensed use limited to: Princeton University. Downloaded on January 24,2021 at 15:53:24 UTC from IEEE Xplore. Restrictions apply.

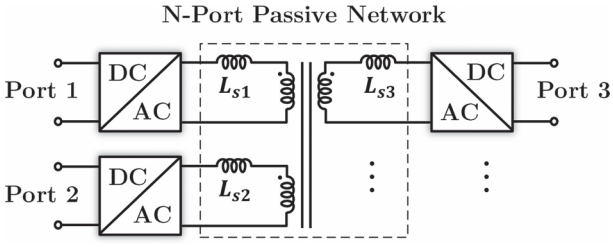


Fig. 6. Photograph of the multiport-ac-coupled converter. Series inductors can be implemented as leak inductors of the multiwinding transformer.

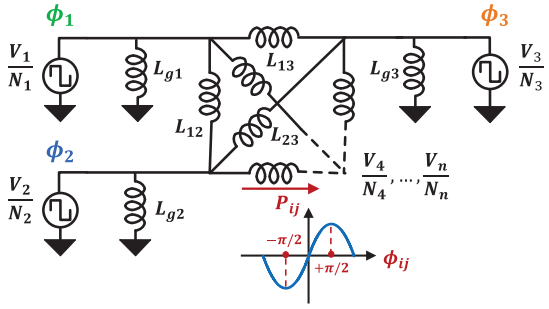


Fig. 7. Equivalent lumped circuit model to analyze the MIMO power flow. The N-port passive network is represented by a delta network, and each dc–ac unit is modeled as a square-wave voltage source.

"dc–ac–dc" stages), lower component voltage rating, higher modularity, and lower component count.

IV. Power Flow Control of the MAC-DPP Converter

Another challenge of designing the MAC-DPP converter is to control the MIMO power flow. As shown in Fig. 6, the MAC-DPP converter is a MIMO system. All ports are bidirectional and are closely coupled with the multiwinding transformer. The multiwinding transformer together with the series inductors is indeed an N-port passive network, whose port voltages and currents are connected by an $N \times N$ impedance matrix

$$\begin{bmatrix} L_{11}+L_{s1} & M_{12} & \cdots & M_{1n} \\ M_{21} & L_{22}+L_{s2} & \cdots & M_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ M_{n1} & M_{n2} & \cdots & L_{nn}+L_{sn} \end{bmatrix}. \tag{5}$$

Here, L_{ii} is the self-inductance of the ith winding, $M_{ij,(i\neq j)}$ is the mutual inductance between windings, and ω is the angular frequency of the system. L_{si} is the series inductance of each winding, which can be either implemented as discrete inductors or the transformer leakage inductance. To analyze the MIMO power flow, the N-port passive network (multiwinding transformer with series inductor) is converted into a delta network, as depicted in Fig. 7. Here, the dc–ac units are implemented as half-bridge or full-bridge circuits, which can be modeled as square-wave voltage sources with normalized voltage amplitudes. Each branch inductor, $L_{ij,(i\neq j)}$, which links the ith and the jth port can be directly obtained from the admittance matrix

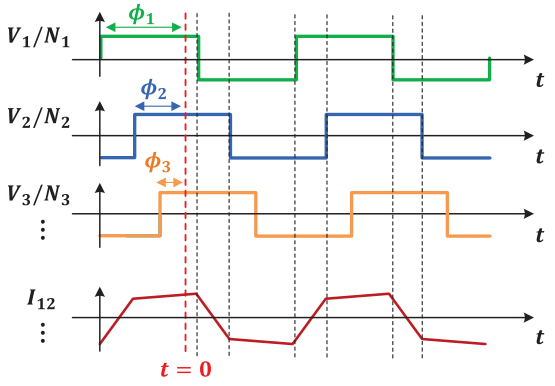


Fig. 8. Example waveforms of normalized port voltages ($\frac{V_1}{N_1} \sim \frac{V_3}{N_3}$) and branch inductor current (I_{12}) with phase-shift modulation.

of the passive network [24]

$$Y = Z^{-1} = \frac{1}{jw}\begin{bmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{nn} \end{bmatrix}, \; L_{ij} = -\frac{1}{N_1 N_2 y_{ij}}. \tag{6}$$

The MIMO power flow can be modulated by adjusting the phase shift at each port (see Figs. 7 and 8). Other power flow modulation methods, such as time-sharing modulation [25], are also applicable. When adjusting the phase shifts, the power flow delivered through each branch inductor (L_{ij}) can be calculated in the same way as that in a DAB converter [26], and the power flow carried by each grounded inductor (L_{gi}) is reactive power, which has no impact on the average power of each port. Thus, the total average power feeds into the passive network from the ith port is

$$P_i = \sum_{j=1}^{n} \frac{V_i V_j}{2\pi f_s N_i N_j L_{ij}} \phi_{ij} \left(1 - \frac{|\phi_{ij}|}{\pi}\right). \tag{7}$$

Open-loop phase-shift modulation is capable of controlling the multiway differential power flow in the steady state, but the system may run into oscillation without feedback control. According to (7), the input average power of one port, P_i (i.e., input differential power in the MAC-DPP system) is related to the phase shifts of all the ports $\{\phi_1, \phi_2, ..., \phi_n\}$. The closely coupled power flow brings challenges to the port voltage regulation, especially in the case where a large number of loads are stacked in series.

One way to control the closely coupled power flow in a MIMO system is to decouple the control loop either with an inverse matrix [27], [28] or using iterative algorithms (e.g., Newton–Raphson method [29], [30]) to solve the nonlinear power flow equations. The port phases are modulated by a central controller. However, these methods have heavy computational demands, making it challenging to meet the dynamic requirements for fast load transients. In addition, they are less scalable to large-scale DPP systems of numerous series-stacked loads. A simplified decoupling method was proposed in [30] and [31], where the power flow equations are linearized, assuming that the phase shift of each port is close to zero. However, the strictly restricted

Authorized licensed use limited to: Princeton University. Downloaded on January 24,2021 at 15:53:24 UTC from IEEE Xplore. Restrictions apply.

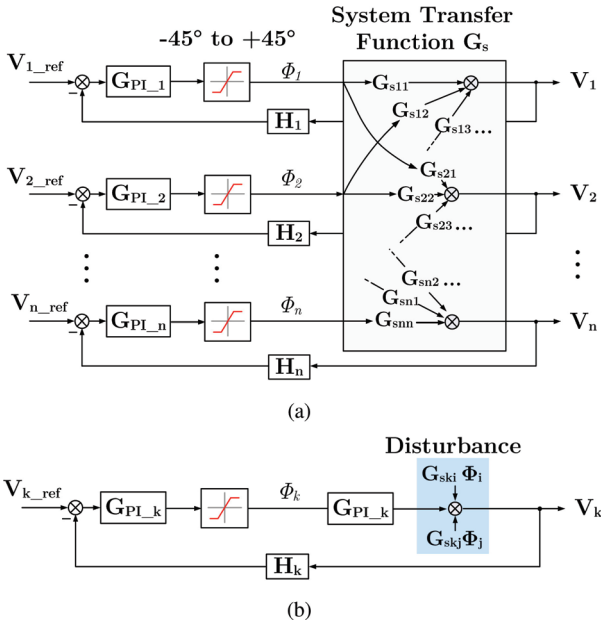


Fig. 9. (a) Block diagrams of the DPS control strategy. (b) Equivalent individual control loop for each port.

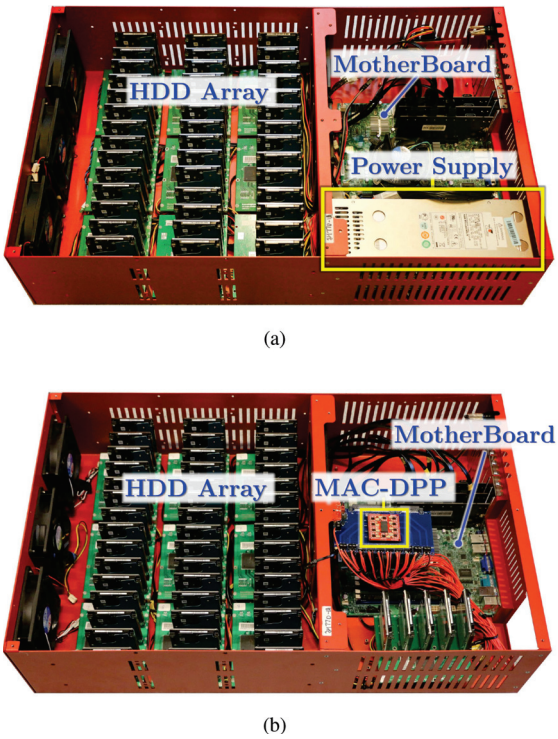


Fig. 10. Photographs of the Backblaze server (a) with the original ac–dc power supply and (b) after replacing the power supply with MAC-DPP converter. The power and communication circuitry are reconfigured.

phase shift places a limit on the maximum power rating of the converter. In addition, the applicable phase-shift range was not specified in these methods, which may push the system out of the stable operation region.

A DPS control strategy as proposed in [32] was adopted to regulate the port voltage. DPS control is simple, effective, and scalable. It fits particularly well to large-scale ac-coupled multiport architectures. Fig. 9 illustrates the principles of the DPS control. Each port utilizes a voltage feedback loop to adjust its own phase (ϕ_i) based on the locally measured port voltage (V_i). As plotted in Fig. 7, the power flow (P_{ij}) through any branch inductor (L_{ij}) is monotonous to the phase difference (ϕ_{ij}) in the range of $[-\frac{\pi}{2}, +\frac{\pi}{2}]$. Therefore, the total input power (P_i) at the ith port is also monotonous to its own phase (ϕ_i), if all the port phases are within the range of $[-\frac{\pi}{4}, +\frac{\pi}{4}]$, which is the applicable phase-shift range for applying DPS control without oscillation.

The stability of the DPS control framework is studied by analyzing the system transfer functions, as illustrated in Fig. 9. Wang *et al.* [32] presented a systematic approach to modeling the MAC-DPP converter with an arbitrary number of ports. The modeling approach accurately captures the impacts of power losses and derives the system transfer function matrix ($\boldsymbol{G_s}$) that describes the dynamics from any control phase shift (ϕ_i) to port voltage (V_j). The nondiagonal elements ($G_{sij(i\neq j)}$) of the transfer function matrix reflects the interactions between different control loops. In the DPS control, the interactions between different feedback loops are considered as disturbances, so the coupled control system can be simplified as multiple standalone feedback control loop at each port, as shown in Fig. 9(b). Based on the derived system transfer function, the loop gain of individual control loop is

$$G_{Li}(s) = G_{\text{PI\_}i}(s) \times G_{sii}(s) \times H_i(s). \tag{8}$$

Here, $G_{\text{PI\_}i}(s)$ is the PI controller parameters. $G_{sii}(s)$ is the diagonal elements of the system transfer function matrix. $H_i(s)$ is the transfer function of the sampling circuitry. The explicitly derived loop gain can be used to analyze the dynamic performance of the system. Through designing the phase margin of each control loop, the oscillation caused by interactions between different ports is minimized. The DPS control is highly modular and scalable and can support large-scale MAC-DPP systems with numerous series voltage domains.

V. Prototype Data Storage Server With DPP

This section presents the details of a MAC-DPP-supported data storage server, including the power stage design, the data communication infrastructure, and the software configuration of the testbench. A Backblaze 4U 45 Drive Storage Pod is selected as the base model for the server. The original server comprised an Intel i3-2100 3.10-GHz CPU, a Supermicro MBD-X9SCM-F mother board, 8-GB RAMs, and 45 2.5-in 320-GB HDDs (TOSHIBA MQ01ABD032V). After modification, the original power supply in the server was replaced with a MAC-DPP converter, and the 45 HDDs were extended to 50 HDDs. The power and communication configuration of the SATA-to-PCIe extension card was modified to enable data transfer across different voltage domains. Fig. 10(a) shows an annotated photograph of the Backblaze server with an original ac–dc power supply, and Fig. 10(b) shows the same Backblaze server after modification, where it is now powered by an ultra-efficient and miniaturized ten-port 450-W MAC-DPP power converter. The HDD server testbench was tested with a variety of data center tasks to

Authorized licensed use limited to: Princeton University. Downloaded on January 24,2021 at 15:53:24 UTC from IEEE Xplore. Restrictions apply.

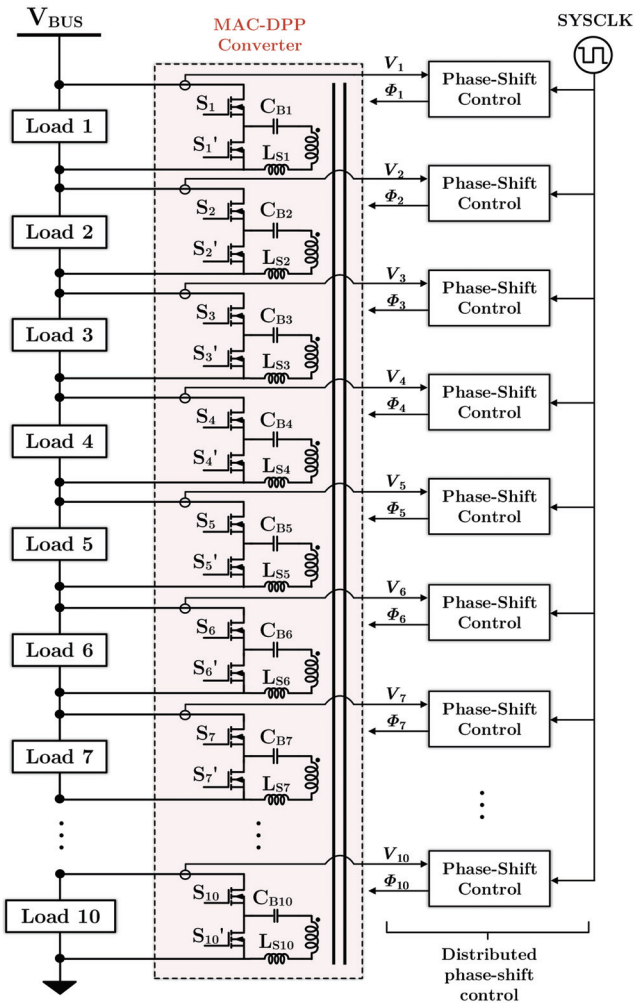


Fig. 11. Topology of a ten-port MAC-DPP converter with dc–ac units implemented as half-bridge circuits.

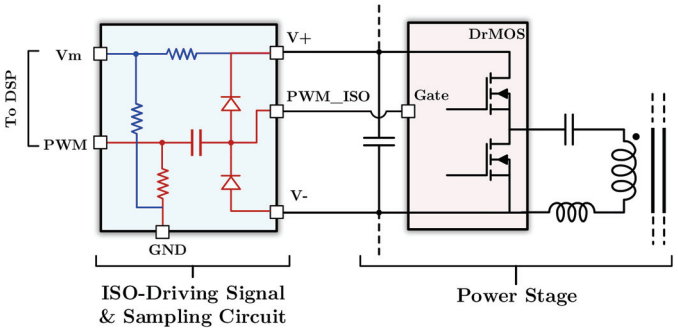


Fig. 12. Modular isolated PWM driving circuit (in red) and voltage sampling circuit (in blue) at each port.

TABLE II
BILL OF MATERIAL OF THE MAC-DPP CONVERTER

Device & Symbol	Component Description
Half-Bridge Switch, $S_1 \sim S_{10}$	DrMOS, CSD95377Q4M
Blocking Capacitor, $C_{B1} \sim C_{B10}$	Murata X5R, 100 μF × 3
Series Inductor, $L_{s1} \sim L_{s10}$	Coilcraft SLC7649, 100 nH
Port Voltage, $V_1 \sim V_{10}$	5 V
Switching Frequency, f_{sw}	100 kHz
Transformer Core	Ferroxcube, ELP18-3C95
Main Power Board Winding	2 oz, single turn × 4
Bottom Cover Winding	2 oz, single turn × 6

validate the applicability of the MAC-DPP prototype. It was also tested in various storage modes to systematically analyze the performance of the MAC-DPP converter and provide guidelines for hardware, software, and power architecture codesign.

A. DPP Power Stage for the Storage Server

This subsection introduces the design of the DPP power stage. Fig. 11 shows the circuit topology of the ten-port MAC-DPP prototype. The dc–ac units are implemented as half-bridge circuits with dc blocking capacitors, and all ports are ac-coupled to a ten-winding transformer. The port-to-port power delivery of this converter is the same as that of a DAB converter with a 1:1 conversion ratio. It offers the lowest power conversion stress and can realize soft switching across the full operation range [33]. The 50-V dc bus is split into ten series-stacked 5-V voltage domains to support 50 2.5-in HDDs. The DPS control units are implemented as standalone phase-shift modules synchronized by a system clock. The voltage sampling circuits and isolated pulsewidth modulation (PWM) signal circuits are designed as scalable modules, as depicted in Fig. 12. In each driving and sampling module, a bootstrapping circuit (annotated in red) is utilized to create a dc bias voltage on the capacitor and generate an isolated PWM signal referred to the floating negative node ($V-$). The voltage sampling circuit (in blue) uses a resistive divider to scale down the positive node voltage ($V+$) and sends it back to the controller. The driving and sampling circuit together with the DPS module can be further integrated into the half-bridge power stage, enabling fully integrated modular building blocks for the MAC-DPP architecture.

Tradeoffs are needed to balance the cost, size, efficiency, power density, and other design targets. Multiobjective optimization is an effective way to select the parameters of a sophisticated system to meet multiple design targets [34], [35]. Based on a detailed loss analysis as presented in Appendix II, switching at a higher frequency can improve the MAC-DPP converter's light-load efficiency, but may reduce the maximum power that can be delivered from port to port. The switching frequency of this prototype was selected as 100 kHz. Other key design parameters of the prototype are listed in Table II.

Fig. 13 shows the top and side views of the MAC-DPP prototype. To create symmetric winding paths, the ten-winding transformer is placed in the middle, surrounded by the ten ports. The driving, sampling circuit, and the power stage are all included. The prototype is 40 mm×35 mm in area, 7.56 mm in height, and the total volume is only 10.58 cm^3 (0.64 in^3).

Fig. 14 shows the 3-D assembly view of the ten-winding printed circuit board (PCB) planar transformer. Two PCB boards are stacked and integrated with an ELP18/10 magnetic core, whose effective core area is 39.5 mm^2. To avoid saturation, the core area is selected as two times of the minimum core

Authorized licensed use limited to: Princeton University. Downloaded on January 24,2021 at 15:53:24 UTC from IEEE Xplore. Restrictions apply.

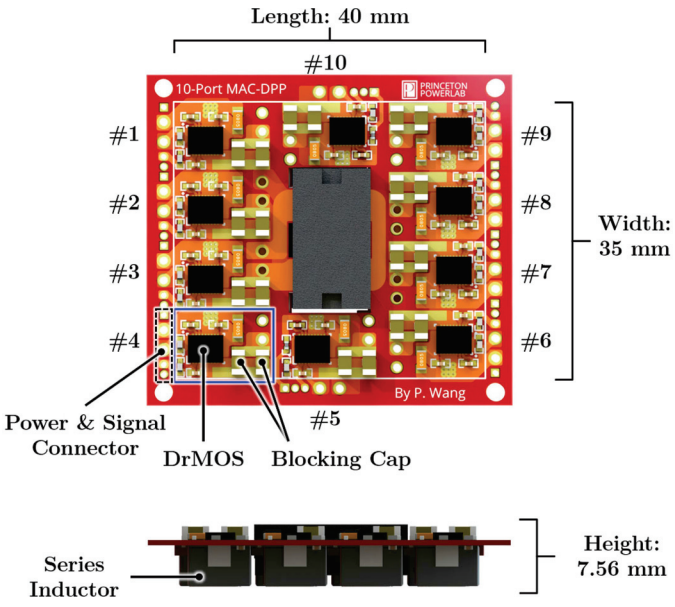


Fig. 13. Annotated top and side views of the ten-port MAC-DPP prototype. The prototype is 40 mm × 35 mm in area and 7.56 mm in height.

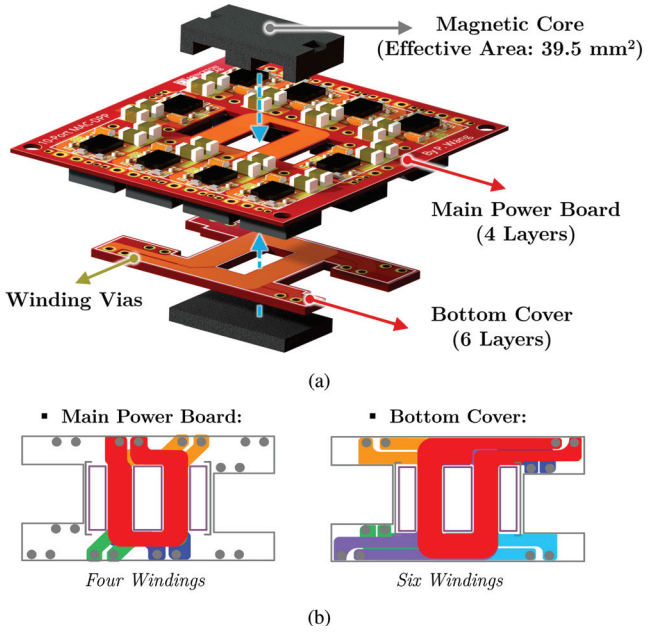


Fig. 14. (a) 3-D assembly view of the stacked PCB planar magnetics. (b) Winding patterns on main power board (four layers) and bottom cover (six layers).

area calculated from (4). This area is comparable to that of a two-winding transformer with the same volt–seconds per turn. Since the additional window area is negligible, the MAC-DPP prototype reduces the magnetic volume by ten times compared to a ten-port dc-coupled DPP converter. Fig. 14(b) shows the PCB patterns of the ten windings. Each winding consists of one single turn in one PCB layer. The main power board comprises four windings, while the bottom cover comprises six windings,

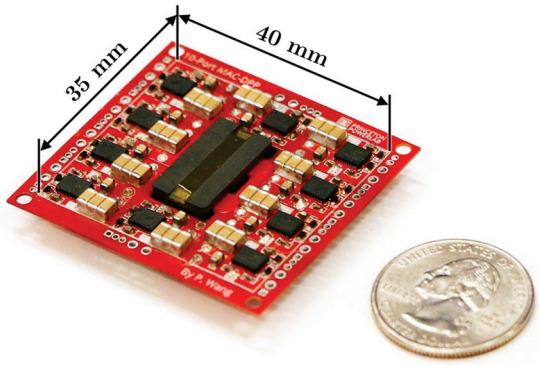


Fig. 15. 450-W ten-port MAC-DPP prototype and a U.S. quarter. The peak system efficiency is >99%, and the peak converter efficiency is >96%.

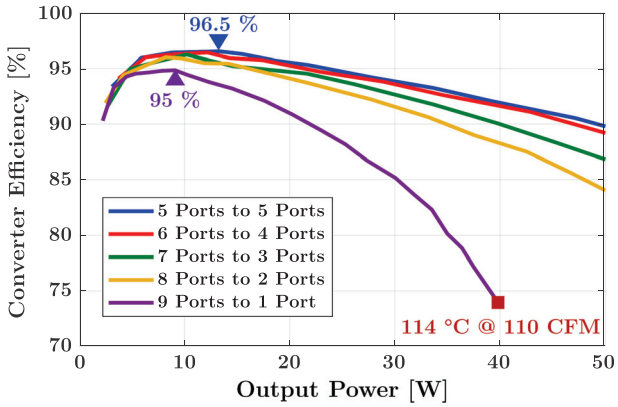


Fig. 16. Port-to-port power converter efficiency in different cases. When delivering 40 W from nine ports to one port, the hot-spot temperature of the output port reached 114 °C under 110 CFM airflow.

which are connected vertically to the main power board through vias. The copper thickness of the PCB is 2 oz.

Since all windings are single-turn PCB windings and the core has high permeability, the magnetic field distribution within the core can be approximated as 1-D. Many models can capture the high-frequency skin and proximity effects in 1-D planar magnetics and provide guidance to the geometry design. For example, Chen *et al.* [36] presents a systematical approach to modeling the impedance and current distribution in multiwinding planar magnetics, which can be used as a guideline to design the windings in the multiwinding transformer.

Fig. 15 shows the MAC-DPP prototype in comparison with a U.S. quarter. The MAC-DPP prototype is a ten-port dc–dc converter, and all ten ports are bidirectional ports. Fig. 16 shows the measured efficiency of the converter under a variety of different power delivery scenarios. In the test, each port is connected to a 5-V dc source/load and switching at 100 kHz. A few ports are connected in parallel as input ports, and other ports are in parallel as output ports. The entire MAC-DPP converter functions equivalently as a one-to-one converter. When delivering power from nine ports to one port, current concentrates at one port. Since conduction loss increases quadratically as current increases, the nine-port-to-one-port scenario dissipates large loss at one

Authorized licensed use limited to: Princeton University. Downloaded on January 24,2021 at 15:53:24 UTC from IEEE Xplore. Restrictions apply.

port, yielding the lowest efficiency. The five-port-to-five-port case has the highest efficiency because the power conversion stress is well distributed. The peak port-to-port conversion efficiency is 96.5% when delivering power from five ports to five ports. The peak efficiency in the worst power delivery scenario (nine-port-to-one-port) is still maintained above 95%. Limited by the concentrated heat at one port, the MAC-DPP prototype can deliver a maximum of 40-W power from nine ports to one port when the hot-spot temperature of the output port reaches 114 °C under 110-CFM airflow. Appendix II presents a detailed loss analysis of the MAC-DPP prototype. Two key figures of merits are defined to evaluate the DPP performance.

1) *System power rating:* The MAC-DPP converter is designed for a DPP system with ten series-stacked voltage domains. The system power rating is defined as the maximum overall load power that the DPP system can support for the desired application, which is different from the actual power processed by the power converter. In a DPP system, the load power P_i at each voltage domain changes between $[0, P_{\max}]$. The differential power that the MAC-DPP converter needs to process in the ith domain is

$$\Delta P_i = \left| P_i - \frac{\sum_{i=1}^{10} P_i}{10} \right|. \tag{9}$$

The maximum differential power at one port is reached if nine voltage domains have no load, while the remaining one operates at full load ($P_{\max}$) or if one voltage domain has no load and the other nine are operating at full load. In this case, the maximum differential power that the MAC-DPP converter needs to deliver from nine ports to one port is $\frac{9}{10}P_{\max}$, which is 40 W according to Fig. 16. As a result, the maximum power of each voltage domain, $P_{\max}$, is approximately 45 W, and the maximum load power that the ten-port MAC-DPP converter can support is 450 W. The power density of the MAC-DPP converter is 700 W/in^3.

2) *System efficiency:* The system efficiency of the MAC-DPP system is defined as the overall load power of all voltage domains divided by the input power from the dc bus

$$\eta_{\text{sys}} = \frac{\sum_{i=1}^{10} P_i}{P_{\text{input}}} = 1 - \frac{P_{\text{loss}}}{P_{\text{input}}}. \tag{10}$$

P_{loss} is the power loss resulting from DPP. In a DPP system, the processed differential power is a small portion of the total load power, so only a small amount of power loss is generated and the system efficiency of a DPP converter can be much higher than the converter efficiency. Define the ratio between the total processed differential power and the total load power as: $r = \sum_{i=1}^{10} \Delta P_i / \sum_{i=1}^{10} P_i$. The generated power loss of the MAC-DPP converter can be calculated as

$$P_{\text{loss}} = r \cdot \sum_{i=1}^{10} P_i \cdot (1 - \eta_{\text{con}}) \tag{11}$$

where η_{con} is the converter efficiency of the MAC-DPP prototype. Based on the converter efficiency in Fig. 16 and (10) and (11), the system efficiency when the server is working at 450-W full load is estimated in Fig. 17.

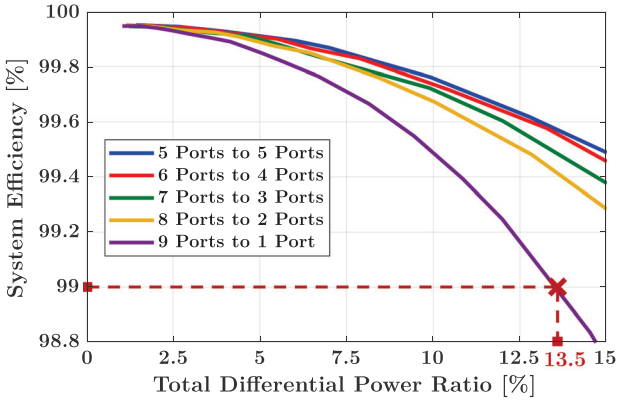


Fig. 17. System power conversion efficiency (total load power: 450 W).

A well-designed storage server usually has uniformly allocated storage tasks among many HDDs. Each HDD has similar reading/writing power consumption. On a series-stacked HDD array (in Fig. 2), many HDDs are connected in parallel in one voltage domain. The power demands of different voltage domains are usually very close to each other with a very low differential power ratio. Therefore, as shown in Fig. 17, the MAC-DPP prototype can maintain over 99% system efficiency of a 450-W data storage server if the differential power ratio is below 13.5%, which covers most of the operation conditions of the storage server. Compared to the conventional 50–5-V dc–dc power delivery solutions for HDDs, the proposed MAC-DPP converter can achieve extremely high system efficiency with very small converter size and can significantly improve the storage capacity per unit volume in storage servers.

B. Data Link Infrastructure for the Data Storage Server

Figs. 18 and 19 shows the detailed implementation of the high-speed data link infrastructure across series-stacked voltage domains. The data link infrastructure comprises three layers. The 50 HDDs are divided into ten groups, and each group contains 5 2.5-in HDDs in parallel on a SATA III port multiplier, namely, backplane board. Ten backplanes in different voltage domains transfer data to the SATA-to-PCIe extension card through isolated differential signals with dc blocking capacitors. Indeed, the SATA/SAS protocol signal is differential. By simply removing the common ground wires and adding blocking capacitors to the SATA/SAS differential signal links, the isolated signal transfer across voltage domains is achieved without major modification to standard communication protocols and existing wiring configuration, as shown in Fig. 19. At Layer 2, a group of SATA-to-PCIe extension cards is placed on the same voltage domain. They are directly connected to the mother board through PCIe Express slots. The three-layer data link infrastructure is scalable to large-scale data storage systems with numerous stacked voltage domains.

Fig. 20 demonstrates the experimental setup for the HDD read/write speed test of the isolated SATA communication based

Authorized licensed use limited to: Princeton University. Downloaded on January 24,2021 at 15:53:24 UTC from IEEE Xplore. Restrictions apply.

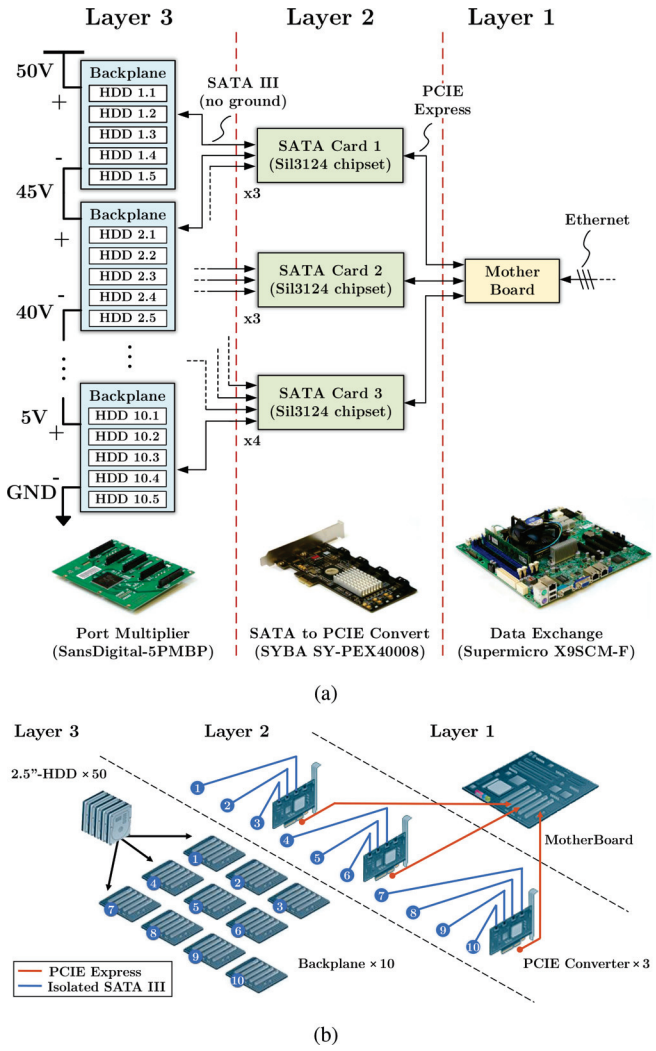


Fig. 18. Data link infrastructure of the series-stacked HDD server testbench. (a) Three-layer data link block diagram. (b) Component connection diagram.

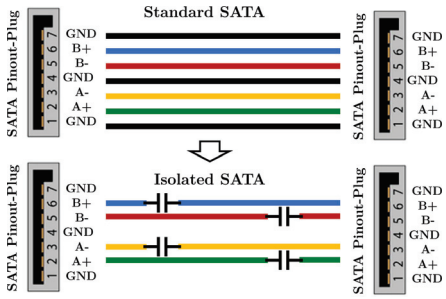


Fig. 19. Isolated SATA wiring pattern of the modified Backblaze storage server. The three ground wires are removed, and the four differential signals are capacitive isolated. Note the SATA extension cards selected in this prototype have internal isolation capacitors. No external capacitors are needed.

on a disk drive benchmark tool, CrystalDiskMark V6.0. Ten 2.5-in HDDs are connected in series to a 50-V dc bus. In this experiment, one HDD was swapped from an isolated voltage domain to a ground-referenced voltage domain, and the reading and writing speed were compared. As listed in Table III, both the sequential read/write speed and 4-KB random read/write speed are nearly the same in two different SATA connections. The

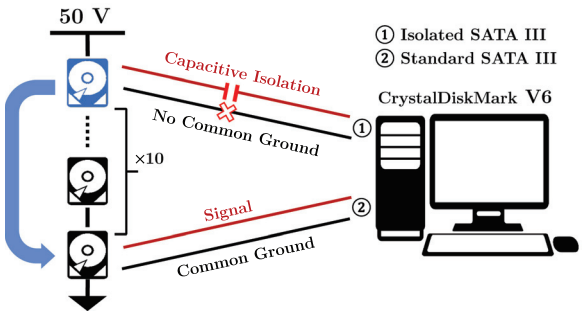


Fig. 20. Experimental setup for the HDD read/write speed comparison between isolated SATA and standard SATA communication. Ten 2.5-in HDDs are in series to a 50-V dc bus. The same HDD was swapped from the first voltage domain (isolated SATA) to the last domain (standard SATA) to test the read/write speed in sequential and 4-kB random mode. The speed were tested using the disk drive benchmark tool, CrystalDiskMark V6.0.

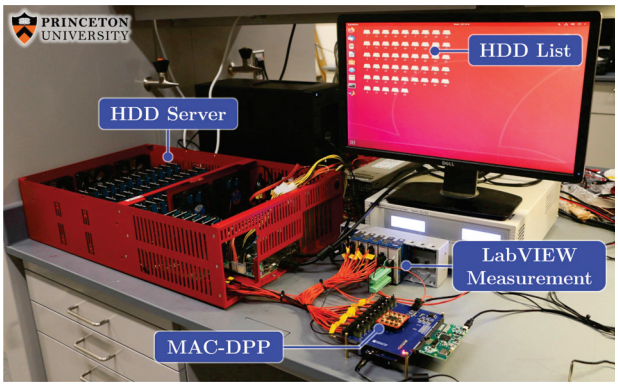


Fig. 21. Side view of the HDD server testbench with the MAC-DPP converter.

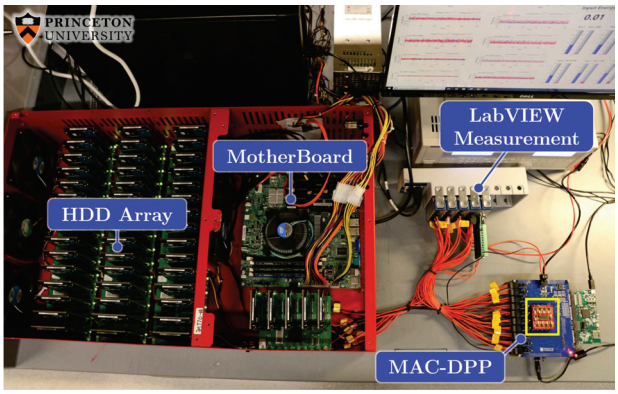


Fig. 22. Top view of the HDD server testbench with the MAC-DPP converter.

results indicate that the bottleneck of SATA transmission speed is the read/write speed of mechanical HDDs and is independent of whether the SATA connection is grounded or not. In applications where a high data rate is needed, the isolated SATA transmission can also be replaced with optic fibers, which are by nature isolated, and can offer higher communication bandwidth.

C. Complete Function Test for the Data Storage Server

Figs. 21 and 22 shows the 50-HDD storage server testbench with a LabVIEW monitoring system. A Linux-based operating

Authorized licensed use limited to: Princeton University. Downloaded on January 24,2021 at 15:53:24 UTC from IEEE Xplore. Restrictions apply.

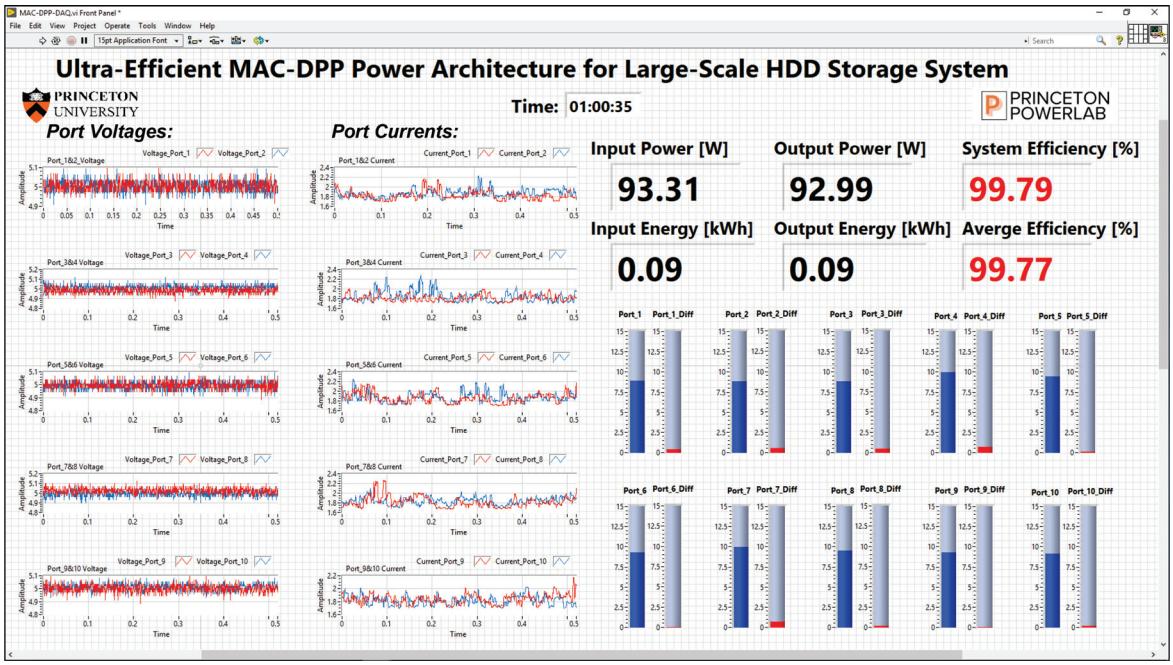


Fig. 23. LabVIEW real-time monitoring system. It measures and records the voltage and current waveforms of all ten series-stacked domains and calculates the system efficiency in real time. In this example, the input power is 93.31 W, the load power is 92.99 W, and the system efficiency is 99.79%.[1]

TABLE III
HDD READ/WRITE SPEED COMPARISON OF ISOLATED SATA AND STANDARD SATA LINK

	Reading (MB/s)		Writing (MB/s)	
	Sequential	4KB Random	Sequential	4KB Random
Isolated	104.0	1.037	104.1	1.036
Standard	104.3	0.987	104.1	1.055

TABLE IV
LONG-TERM RANDOM READ/WRITE TESTING RESULTS

Elapsed Time	Input Energy	Load Energy	System Efficiency
60 min	333.801 kJ	333.031 kJ	99.77 %

system (Ubuntu) is installed to manage the reading, writing, and hot-swapping functions. A dc voltage source (QPX-600D) is utilized as the 50-V dc bus.

A LabVIEW system was set up to monitor the power consumption of the HDD server testbench. The monitoring system utilizes an NI-compactDAQ (cDAQ-9178) together with extendable analog input modules (NI9221 and NI9227) to simultaneously sample the voltages and currents of all the ten voltage domains as well as the input voltage and current of the dc bus. The sampling rate of each voltage or current sampling channel is 1600 Samples/s (the sampling period is about 620 μs), and the sampled voltage and current were calibrated by a Keysight Digital Multimeter (34401A). In the LabVIEW console shown in Fig. 23, the voltage and current of ten voltage domains are monitored in real time, including the voltage ripple, load power, and differential power of each voltage domain as well as system efficiency, etc. The LabVIEW monitoring system is also capable of recording the system dynamic response when hot-swapping HDDs.

An HDD usually has two operating states: 1) reading or writing, each HDD used in this hardware setup consumes about 2.8 W to drive the motor; and 2) idling, each HDD in the hardware setup consumes about 0.7 W to maintain active. In data centers, the reading/writing operation of each HDD is commanded by external software requests. To validate the MAC-DPP architecture on the HDD server with typical data center tasks, a random reading/writing program was created, in which each HDD has a 20% probability to perform reading/writing tasks and 80% probability to stay idling at any time instant. Fig. 24 shows the measured voltage and current waveforms of the ten voltage domains under the random reading/writing test. The average power of each voltage domain is about 9 W, consisting of the random HDD load power and the power consumption of the backplane board. Due to the random reading/writing tasks, the load currents were fluctuating continuously, but the voltages of all the domains were maintained stably at 5 V. The random reading/writing task was run for 1 h, during which the accumulated input and load energy was recorded, as listed in Table IV. The total input energy from the dc bus was 333.801 kJ, while the total load energy (including energy consumptions of HDDs and backplanes) was 333.031 kJ, so the average system efficiency was as high as 99.77%. The testing results show that the MAC-DPP converter can feed power to the ten voltage domains with very high system efficiency.

Maintaining a dc voltage within a narrow ripple range is of great importance for the robust operation of HDDs. A typical

[1][Online]. Available: https://www.youtube.com/watch?v=EIbYMvVjsWg

Authorized licensed use limited to: Princeton University. Downloaded on January 24,2021 at 15:53:24 UTC from IEEE Xplore. Restrictions apply.

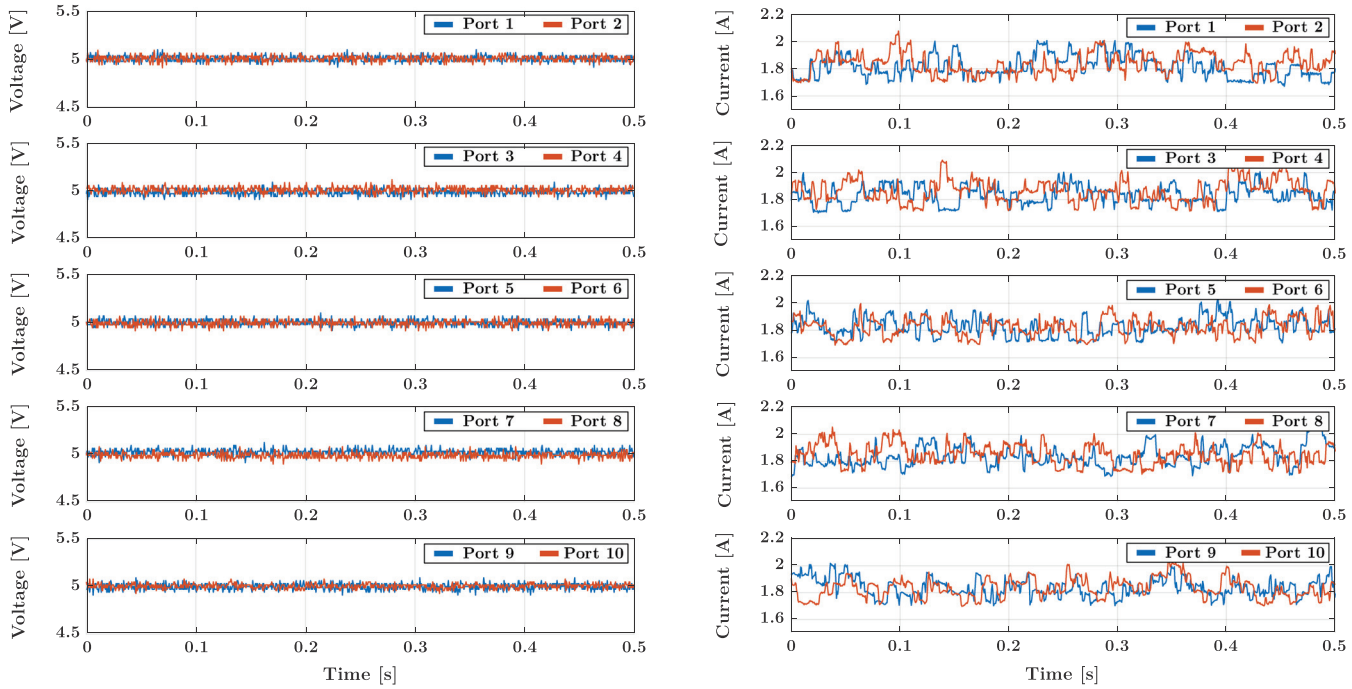


Fig. 24. Experimental waveforms of all voltage domains at random reading/writing test measured by LabVIEW. (a) Voltage waveforms. (b) Current waveforms.

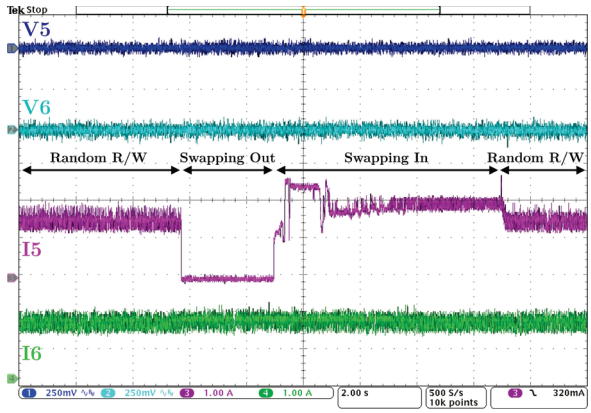


Fig. 25. Transient response when hot swapping an entire voltage domain (removing five HDDs from port #5) of the HDD server testbench. Voltage measurements are ac coupled, and current measurements are dc coupled.

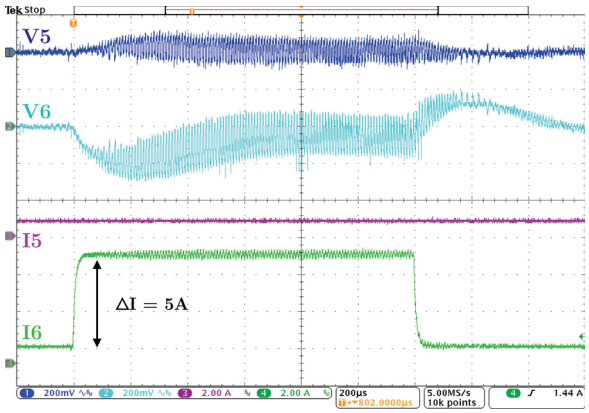


Fig. 26. Transient response of a 25-W step load change at port #6. The settling time is 0.5 ms, and the voltage overshoot is less than 250 mV. Voltage measurements are ac coupled, and current measurements are dc coupled.

requirement for 2.5-in HDDs is to regulate the voltage within 5% of the nominal value (250 mV out of 5 V). In data centers, to avoid interrupting the normal operation, HDDs are usually removed or replaced, while the server systems are still running (i.e., hot swapping). Hot swapping induces large load current transient, bringing challenges to voltage regulation. In the random reading/writing experiment, a worst-case hot-swapping test was performed, where an entire voltage domain (five HDDs and one backplane) was abruptly pulled out and plugged in. In this scenario, the differential power change at one port reaches the maximum, resulting in the largest voltage fluctuation during the transient. DPS control regulates the voltage of the ten voltage domains. Fig. 25 shows the measured port voltage and load current waveforms at the fifth and sixth voltage domains during the hot-swapping test. A 2.2-mF electrolytic capacitor was included at each port, and the fifth domain was hot swapped, while the HDDs in other voltage domains were kept performing the random reading/writing task. During hot swapping, the voltage transition was very smooth. The fluctuation is almost negligible. Fig. 25 also shows that the current variation during swapping in is higher than that of swapping out, because of the current overshoot caused by the motor spinning up when

Authorized licensed use limited to: Princeton University. Downloaded on January 24,2021 at 15:53:24 UTC from IEEE Xplore. Restrictions apply.

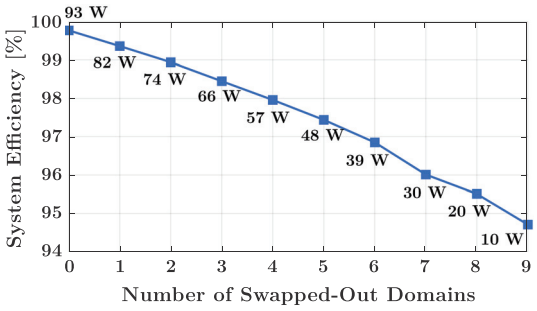
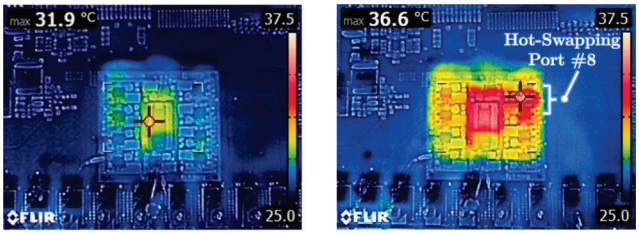


Fig. 27. Measured system efficiency when different number of voltage domains were swapped out. The average overall load power is annotated aside each data point. The system efficiency drops as more HDDs were removed.



(a) (b)

Fig. 28. Thermal images of the MAC-DPP prototype in (a) balanced load and (b) hot swapping an entire voltage domain. The thermal images were measured at 25 °C ambient temperature after the testbench running for 10 min without forced air flow.

swapping in. The behavior indicates that the transient performance of a DPP system on an HDD server should be designed for the case of hot swapping. A soft starting circuit can also be implemented to meet higher requirements on HDD voltage ripple.

Benefiting from the control strategy to support hot swapping, the DPP system is robust against device failure. By connecting a protection device in series with the loads in each voltage domain, which fails as open (e.g., a fuse or a current limiting device), the challenge of managing a failure condition is translated into a managing a hot-swapping transient—the voltage domain, which has a fault condition, is removed from the series stack and the power is instantly redistributed.

Since the MAC-DPP prototype is designed to support 45-W peak power at each voltage domain, the transient response of the prototype was also tested in an extreme case with 25-W load step change in one voltage domain (i.e., 56% of full load step change). In the test, each series-stacked voltage domain was connected to an electronic load. All the load currents were kept at 1 A except for the current at port #6, which was stepped up from 1 to 6 A and then returned back to 1 A, as shown in Fig. 26. The MAC-DPP converter can successfully limit the overshoot of the "hot-swapping" port voltage to 250 mV with only 0.5-ms settling time, fulfilling the 5% voltage ripple requirements. Fig. 26 also indicates that the load step change in one port induces voltage fluctuation on other ports (e.g., V_5), but they can also be effectively controlled by the DPS control strategy. These hot-swapping experiments verified that the designed MAC-DPP prototype is capable of maintaining a smooth operation of the HDD server against the worst-case hot-swapping scenarios.

Hot swapping leads to unbalanced load power, yielding reduced system efficiency. As more voltage domains are swapped out, the power mismatch between different voltage domains usually increases. Fig. 27 shows the measured system efficiency in the random reading/writing test when different numbers of voltage domains were swapped out. The overall load power decreased as more voltage domains were removed, and the system efficiency also dropped. In the worst case, where nine voltage domains were out, the system efficiency dropped to 94.7%. Under this circumstance, power was delivered to the load bypassing nine voltage domains. The lowest efficiency, 94.7%, is still comparable to that of the state-of-the-art 10:1 dc–dc converters. A DPP solution can offer much higher efficiency than dc–dc converters in most cases.

Fig. 28 shows the thermal images of the MAC-DPP converter operating in different load conditions. Both thermal images were taken after the testbench running for over 10 min. The experiment is performed under 25 °C ambient temperature with no forced airflow. At the beginning, when all HDDs were doing the same random reading/writing tasks, the load power was very balanced with only a small amount of differential power to be processed by the MAC-DPP converter. The temperature distribution on the MAC-DPP converter was uniform, and little hot spot could be observed. The transformer is the hottest component due to core loss. When all five HDDs of an entire voltage domain were removed, the hot-swapping port delivered about 9-W differential power to the other nine ports. Since the current at the hot-swapping port was roughly the summation of currents of all other nine ports, its loss was much higher than others. A significant temperature rise was observed at the hot-swapping port (port #8 in this case), as shown in Fig. 28(b). In this worst case, the temperature of the MAC-DPP converter was still maintained lower than 40 °C without forced air cooling.

Fig. 29 compares the system efficiency and power density of the MAC-DPP prototype with many state-of-the-art commercial 48–5-V dc–dc converters. Benefiting from the DPP architecture and the single "dc–ac–dc" power delivery path, the MAC-DPP prototype can support a 450-W HDD server with about 1 W of loss (99.77% system efficiency), reducing the power loss by 10× compared to most of the commercial products. By employing the MAC-DPP topology, the prototype has a smaller overall magnetic volume and lower component count compared to many other DPP topologies. The MAC-DPP converter is miniaturized with a power density above 700 W/in^3, which is higher than most commercial products. The voltage sampling circuit and isolated driving signal circuit are all included in the MAC-DPP prototype and are considered in volume calculation. The microcontroller (TI F28379D) is off-board and is not included in the power density calculation.

Authorized licensed use limited to: Princeton University. Downloaded on January 24,2021 at 15:53:24 UTC from IEEE Xplore. Restrictions apply.

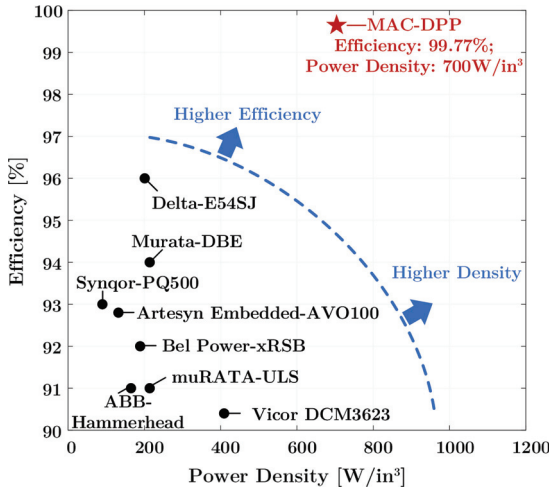


Fig. 29. Comparison of the ten-port MAC-DPP prototype with many state-of-the-art commercial 48–5-V dc–dc converters. The MAC-DPP converter achieves over 10× power loss reduction compared with most of industry products with top-ranking power density. This comparison is based on the DPP system efficiency. The port-to-port converter efficiency is shown in Fig. 16. The size of the microcontroller is not included in the volume calculation.

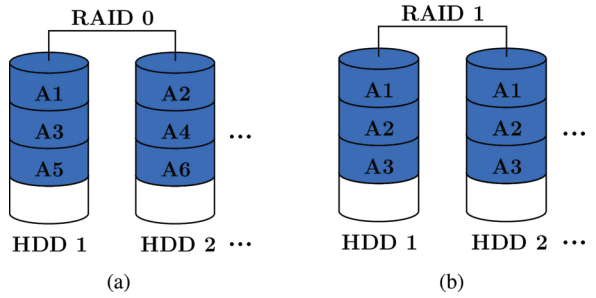


Fig. 30. Two different RAID levels. (a) RAID 0 (striped volume). (b) RAID 1 (mirrored volume) [21].

D. Software, Hardware, and Power Architecture Codesign

The performance of the DPP system is closely related to the load power variation between series-stacked voltage domains. In data centers, hardware infrastructure and software algorithms will have an impact on the power consumption, thus influencing the performance of power converters. There are opportunities to investigate software, hardware, and power codesign of large-scale computing systems in data centers, such as CPU/GPU clusters, memory banks, and HDD arrays.

RAID is a popular data storage architecture adopted in commercial cloud storage HDD arrays [21]. It combines multiple HDDs into one or more logical units in order to improve storage reliability or storage speed. Fig. 30 demonstrates two typical RAID configurations: (a) RAID 0, where the data are divided into multiple parts (namely striped) and written into multiple disks in parallel; there is no redundancy of data, but the storage speed is improved; and (b) RAID 1, where the data are duplicated and stored in multiple disks (namely mirror); the storage speed is the same as for a single disk, but the storage reliability is improved due to the data redundancy. Other RAID levels such as RAID 5 (striped with parity check), RAID 10 (striped and mirrored), etc., are extensions of these two RAID levels.

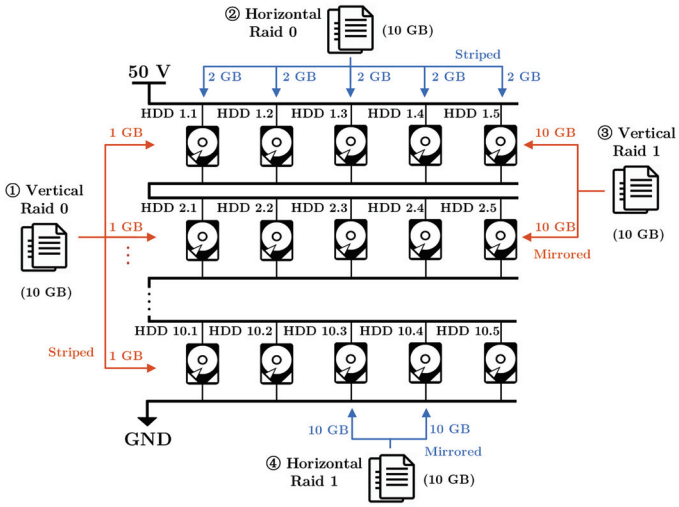


Fig. 31. Implementation of different RAID levels on the 10 × 5 HDD array. HDDs can be vertically or horizontally grouped together into RAID systems.

The MAC-DPP system was tested together with different storage architectures. RAID 0 and RAID 1 levels were applied, and a 10-GB file chunk was utilized as a testing sample. Fig. 31 shows the implementation of four different RAID levels on the 10 × 5 HDD array. The following five modes were tested.

1) *Vertical RAID 0:* The 10-GB file chunk was striped into ten HDDs across ten voltage domains. Each HDD was written into 1-GB file chunk.
2) *Horizontal RAID 0:* The 10-GB file chunk was striped into five HDDs within one voltage domain. Each HDD was written into 2-GB file chunk.
3) *Vertical RAID 1:* The 10-GB file chunk was mirrored into two HDDs across two voltage domains. Each HDD was written into 10-GB file chunk.
4) *Horizontal RAID 1:* The 10-GB file chunk was mirrored into two HDDs within one voltage domain. Each HDD was written into 10-GB file chunk.
5) *Direct storage:* The 10-GB file chunk was directly written into one single HDD.

A systematic performance analysis of the HDD server is performed. Time consumption, system efficiency, and energy consumption of the HDD array when writing the 10-GB file sample under different storage strategies were measured in LabVIEW, and the experimental results are shown in Fig. 32. As indicated by the results, RAID 0 offers faster transmission speed due to the mechanism of parallel storage. Although RAID 1 needs higher HDD energy consumption, it provides higher storage redundancy. Fig. 32(b) shows that vertical RAID 0 has the highest system efficiency. Horizontal RAID 1 is the least efficient. This is because the load distribution of vertical RAID 0 is the most balanced across different voltage domains, but horizontal RAID 0 has the most unbalanced load distribution. The difference of system efficiency in different HDD storage architecture will be more distinct in larger HDD arrays with more HDDs included in the storage tasks. Due to the limited

Authorized licensed use limited to: Princeton University. Downloaded on January 24,2021 at 15:53:24 UTC from IEEE Xplore. Restrictions apply.

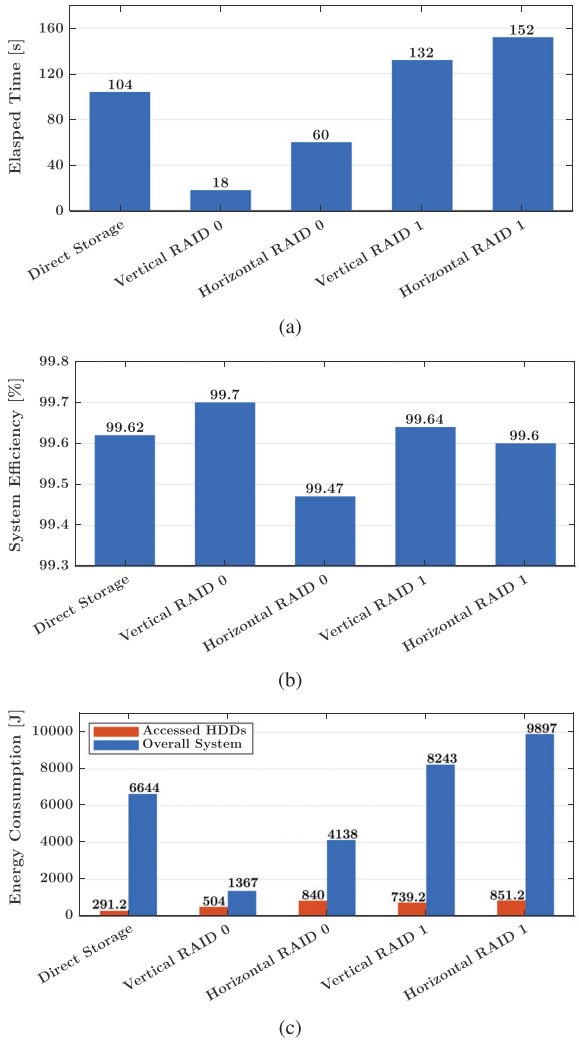


Fig. 32. Experimental results of writing test under different storage architectures. HDD server performance was analyzed in multiple aspects including (a) time consumption, (b) system efficiency, and (c) energy consumption of the overall system (including working/idling HDDs and backplanes), or just the HDDs accessed by the writing test.

bandwidth, the advantages of parallel storage speed were not completely exploited. Because of these nonideal factors involved in the test, a more rigorous study is needed to fully reveal the advantages and disadvantages of grouping HDDs in different ways. However, it can still be distinctly concluded from the results that vertical RAID modes have higher system efficiency and lower energy consumption compared with the horizontal counterparts due to more balanced power distribution among different voltage domains. It suggests that storage algorithm and storage architecture in data centers can be optimized to allocate storage tasks more balanced across different voltage domains, creating a more balanced load power, and thus greatly improving the overall performance of the system.

VI. Conclusion

This article presented the design and implementation of the first data storage server supported by series-stacked DPP.

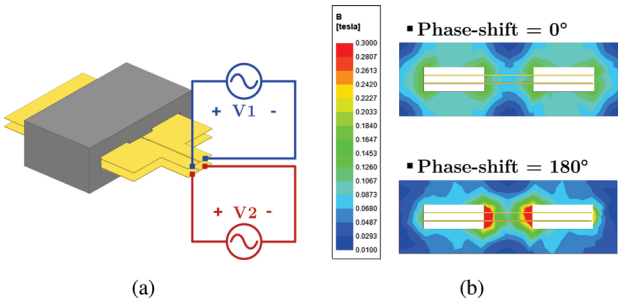


Fig. 33. (a) FEM simulation setup: two windings are driven by two sinusoidal voltage sources of different phase shifts. (b) Simulated magnetic flux density inside the core at the phase shift of 0° and 180°, respectively.

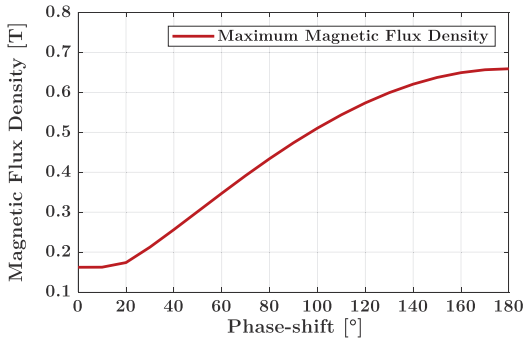


Fig. 34. Maximum magnetic flux density in the spacing between two adjacent windings when sweeping the voltage phase shift from 0° to 180°.

A MAC-DPP architecture was developed to offer reduced component count, a single "dc–ac–dc" power conversion stage, and the smallest magnetic size. The multiwinding transformer was implemented as a closely coupled PCB planar transformer. A DPS control strategy was implemented for the MAC-DPP converter. A 450-W ten-port MAC-DPP converter was designed and tested in a 50-HDD data storage server testbench. The HDD server can maintain normal reading/writing operation against the worst hot-swapping scenario for the HDDs. The storage server was also tested in an extreme case when 25-W load was hot swapped at one port. The transient response of the MAC-DPP system meets the requirements of typical HDDs, and the system efficiency for a 450 W storage server remains above 99% for a majority of operating conditions. The storage server was also tested with various HDD storage modes including direct storage and different RAID levels. Experimental results showed that the performance of large-scale modular information systems can be greatly improved by software, hardware, and power architecture codesign.

Appendix I
FEM Analysis of the Multiwinding Transformer

Fig. 33(a) shows an example transformer simulated in ANSYS Maxwell to validate the design guidelines with FEM. This transformer has a ferrite planar core (ELP18/10 with $\mu_r = 1000$). Each winding has one single turn. Two sinusoidal voltage sources (2.5-V amplitude, 100 kHz) were connected to the two windings. Fig. 33(b) shows the simulated magnetic flux

Authorized licensed use limited to: Princeton University. Downloaded on January 24,2021 at 15:53:24 UTC from IEEE Xplore. Restrictions apply.

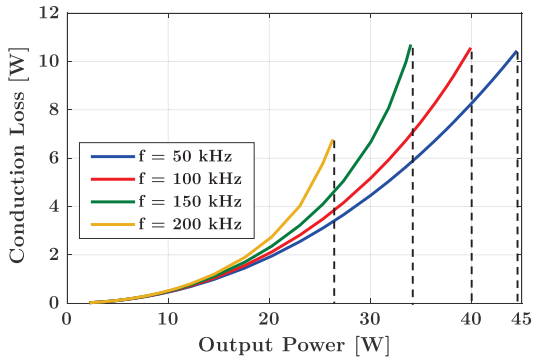


Fig. 35. Estimated conduction loss when delivering power from nine ports to one port at different switching frequencies.

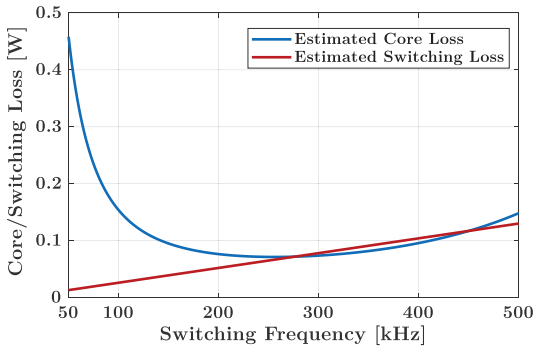


Fig. 36. Estimated core loss and switching loss as a function of the switching frequency from 50 to 200 kHz. Gate drive loss is not included.

density inside the core with different phase shifts. If two voltage sources are in phase, the magnetic flux density inside the core is relatively uniform, and the maximum flux density ($B_{\max}$) is low. When the phase shift increases to 180°, the two voltage sources have exactly opposite phases, and the magnetic flux concentrates at the spacing between two windings, leading to a high peak flux density that might saturate the core. Fig. 34 shows the maximum flux density of the spacing area between two windings when sweeping the phase shift from 0° to 180°. The $B_{\max}$ increases as the phase shift increases, indicating that the spacing between two windings should be designed for the 180° phase shift or the maximum phase shift if it is below 180°. The voltage applied to the winding terminals set the boundary conditions needed to be solved for the magnetic flux density in the core.

As a result, to avoid saturating a voltage-source-driven planar transformer with multiple windings, the minimum cross-sectional area of the core is determined by the maximum volt–second per turn of the windings, and the minimum spacing between two windings is determined by the maximum phase shift between them.

APPENDIX II
MAC-DPP LOSS ANALYSIS

The performance of the MAC-DPP converter is directly related to the operating conditions. The power loss consists of core loss, conduction loss, and switching loss. Figs. 35–37 perform

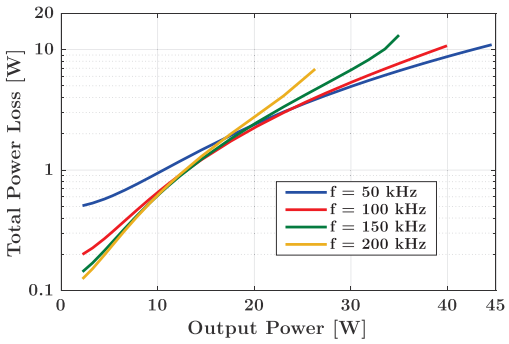


Fig. 37. Estimated total power loss of the MAC-DPP prototype when delivering power from nine ports to one port at different frequencies. The total power loss includes conduction loss, core loss, and switching loss.

a loss analysis for the MAC-DPP converter when delivering power from 9 ports to 1 port under different operating conditions. The core loss is calculated by the Steinmetz's equation with the fitted coefficient from the Ferroxcube-3C95 datasheet. The root-mean-square (rms) current of each conduction path is calculated based on the output load current and phase shift between input and output.

Based on (7), when outputting the same amount of power, the phase shift between the input and output ports increases as the switching frequency increases, leading to higher rms current and higher conduction loss, as shown in Fig. 35. When operating at 200 kHz, the maximum output power of the MAC-DPP converter is determined by the phase shift. It delivers 26.3 W from nine ports to one port at 90° phase shift. When the switching frequency is 150, 100, and 50 kHz, the maximum power that the MAC-DPP converter can deliver is 34, 40, and 44.5 W, respectively, limited by the maximum allowable component temperature (assume that the temperature limit is reached when the conduction loss reaches the same value as that of the experiment with 114 °C temperature in Fig. 16).

Fig. 36 shows the estimated core loss and switching loss as a function of the switching frequency. Fig. 37 shows the estimated full system loss at different frequencies. The core loss and switching loss dominate the system loss at light load. The conduction loss dominates the system loss at heavy load.

ACKNOWLEDGMENT

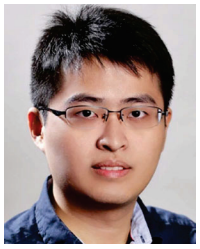
The views and opinions of the authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

REFERENCES

[1] A. Shehabi *et al.*, "United States Data Center energy usage report," Lawrence Berkeley Nat. Lab., Berkeley, CA, USA, Tech. Rep. LBNL-1005775, 2016.
[2] D. Reinsel, J. Gantz, and J. Rydning, *The Digitization of the World: From Edge to Core*. Framingham, MA, USA: Int. Data Corp., 2018.
[3] P. Wang, Y. Chen, P. Kushima, Y. Elasser, M. Liu, and M. Chen, "A 99.7% efficient 300 W hard disk drive storage server with multiport ac-coupled differential power processing (MAC-DPP) architecture," in *Proc. IEEE Energy Convers. Cong. Expo.*, Sep. 2019, pp. 5124–5131.

Authorized licensed use limited to: Princeton University. Downloaded on January 24,2021 at 15:53:24 UTC from IEEE Xplore. Restrictions apply.

[4] M. H. Ahmed, C. Fei, F. C. Lee, and Q. Li, "Single-stage high-efficiency 48/1 V sigma converter with integrated magnetics," *IEEE Trans. Ind. Electron.*, vol. 67, no. 1, pp. 192–202, Jan. 2020.
[5] S. Jiang, S. Saggini, C. Nan, X. Li, C. Chung, and M. Yazdani, "Switched tank converters," *IEEE Trans. Power Electron.*, vol. 34, no. 6, pp. 5048–5062, Jun. 2019.
[6] Z. Ye, Y. Lei, and R. C. N. Pilawa-Podgurski, "The cascaded resonant converter: A hybrid switched-capacitor topology with high power density and efficiency," *IEEE Trans. Power Electron.*, vol. 35, no. 5, pp. 4946–4958, May 2020.
[7] R. C. N. Pilawa-Podgurski and D. J. Perreault, "Submodule integrated distributed maximum power point tracking for solar photovoltaic applications," *IEEE Trans. Power Electron.*, vol. 28, no. 6, pp. 2957–2967, Jun. 2013.
[8] A. H. Chang, A. Avestruz, and S. B. Leeb, "Capacitor-less photovoltaic cell-level power balancing using diffusion charge redistribution," *IEEE Trans. Power Electron.*, vol. 30, no. 2, pp. 537–546, Feb. 2015.
[9] J. T. Stauth, M. D. Seeman, and K. Kesarwani, "Resonant switched-capacitor converters for sub-module distributed photovoltaic power management," *IEEE Trans. Power Electron.*, vol. 28, no. 3, pp. 1189–1198, Mar. 2013.
[10] 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.
[11] P. S. Shenoy, K. A. Kim, B. B. Johnson, and P. T. 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.
[12] K. A. Kim, P. S. Shenoy, and P. T. Krein, "Converter rating analysis for photovoltaic differential power processing systems," *IEEE Trans. Power Electron.*, vol. 30, no. 4, pp. 1987–1997, Apr. 2015.
[13] C. Liu, D. Li, Y. Zheng, and B. Lehman, "Modular differential power processing (mDPP)," in *Proc. IEEE Workshop Control Model. Power Electron.*, Stanford, CA, USA, 2017, pp. 1–7.
[14] G. L. Brainard, "Non-dissipative battery charger equalizer," U.S. Patent 5 479 083, Dec. 1995.
[15] A. M. Imtiaz and F. H. Khan, "Time shared flyback converter based regenerative cell balancing technique for series connected Li-ion battery strings," *IEEE Trans. Power Electron.*, vol. 28, no. 12, pp. 5960–5975, Dec. 2013.
[16] M. Evzelman, M. M. Ur Rehman, K. Hathaway, R. Zane, D. Costinett, and D. Maksimovic, "Active balancing system for electric vehicles with incorporated low-voltage bus," *IEEE Trans. Power Electron.*, vol. 31, no. 11, pp. 7887–7895, Nov. 2016.
[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] E. Candan, A. Stillwell, and R. C. N. Pilawa-Podgurski, "A reliability assessment of series-stacked servers with server-to-bus differential power processing," *Proc. IEEE Int. Telecom. Energy Conf.*, Austin, TX, USA, 2016, pp. 1–7.
[19] 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. Cong. Expo.*, Montreal, QC, 2015, pp. 563–570.
[20] S. K. Dam and V. John, "A modular fast cell-to-cell battery voltage equalizer," *IEEE Trans. Power Electron.*, vol. 35, no. 9, pp. 9443–9461, Sep. 2020.
[21] P. M. Chen, E. K. Lee, G. A. Gibson, R. H. Katz, and D. A. Patterson, "RAID: High-performance, reliable secondary storage," *ACM Comput. Surv.*, vol. 26, no. 2, pp. 145–185, 1994.
[22] P. Wang and M. Chen, "Towards power FPGA: Architecture, modeling and control of multiport power converters," in *Proc. IEEE Workshop Control Model. Power Electron.*, Jun. 2018, pp. 1–8.
[23] M. Liu, P. Wang, Y. Guan, and M. Chen, "A 13.56 MHz multiport-wireless-coupled (MWC) battery balancer with high frequency online electrochemical impedance spectroscopy," in *Proc. IEEE Energy Convers. Cong. Expo.*, Baltimore, MD, USA, 2019, pp. 537–544.
[24] R. W. Erickson and D. Maksimovic, "A multiple-winding magnetics model having directly measurable parameters," in *Proc. 29th Annu. IEEE Power Electron. Spec. Conf.*, 1998, vol. 2, pp. 1472–1478.
[25] Y. Chen, P. Wang, H. Li, and M. Chen, "Power flow control in multi-active-bridge converters: Theories and applications," in *Proc. IEEE Appl. Power Electron. Conf. Expo.*, Anaheim, CA, USA, 2019, pp. 1500–1507.
[26] R. W. A. A. De Doncker, D. M. Divan, and M. H. Kheraluwala, "A three-phase soft-switched high-power-density dc/dc converter for high-power applications," *IEEE Trans. Ind. Appl.*, vol. 27, no. 1, pp. 63–73, Jan./Feb. 1991.
[27] C. Zhao, S. D. Round, and J. W. Kolar, "An isolated three-port bidirectional dc-dc converter with decoupled power flow management," *IEEE Trans. Power Electron.*, vol. 23, no. 5, pp. 2443–2453, Sep. 2008.
[28] S. Falcones, R. Ayyanar, and X. Mao, "A dc-dc multiport-converter based solid-state transformer integrating distributed generation and storage," *IEEE Trans. Power Electron.*, vol. 28, no. 5, pp. 2192–2203, May 2013.
[29] L. Ortega, P. Zumel, C. Fernández, J. López-López, A. Lázaro, and A. Barrado, "Power distribution algorithm and steady state operation analysis of a modular multi-active bridge converter," *IEEE Trans. Transp. Electrific.*, to be published.
[30] C. Gu, Z. Zheng, L. Xu, K. Wang, and Y. Li, "Modeling and control of a multiport power electronic transformer (PET) for electric traction applications," *IEEE Trans. Power Electron.*, vol. 31, no. 2, pp. 915–927, Feb. 2016.
[31] G. Buticchi, L. F. Costa, D. Barater, M. Liserre, and E. D. Amarillo, "A quadruple active bridge converter for the storage integration on the more electric aircraft," *IEEE Trans. Power Electron.*, vol. 33, no. 9, pp. 8174–8186, Sep. 2018.
[32] P. Wang, Y. Chen, Y. Elasser, and M. Chen, "Small signal model for very-large-scale multi-active-bridge differential power processing (MAB-DPP) architecture," in *Proc. IEEE Workshop Control Model. Power Electron.*, Toronto, ON, Canada, 2019, pp. 1–8.
[33] M. N. Kheraluwala, R. W. Gascoigne, D. M. Divan, and E. D. Baumann, "Performance characterization of a high-power dual active bridge dc-to-dc converter," *IEEE Trans. Ind. Electron.*, vol. 28, no. 6, pp. 1294–1301, Nov./Dec. 1992.
[34] J. W. Kolar, J. Biela, and J. Minibock, "Exploring the Pareto front of multi-objective single-phase PFC rectifier design optimization—99.2% efficiency vs. 7 kW/din^3 power density," in *Proc. IEEE Int. Power Electron. Motion Control Conf.*, Wuhan, China, 2009, pp. 1-21.
[35] R. Bosshard and J. W. Kolar, "Multi-objective optimization of 50 kW/85 kHz IPT system for public transport," *IEEE Trans. Emerg. Sel. Topics Power Electron.*, vol. 4, no. 4, pp. 1370–1382, Dec. 2016.
[36] M. Chen, M. Araghchini, K. K. Afridi, J. H. Lang, C. R. Sullivan, and D. J. Perreault, "A systematic approach to modeling impedances and current distribution in planar magnetics," *IEEE Trans. Power Electron.*, vol. 31, no. 1, pp. 560–580, Jan. 2016.



Ping Wang (Student Member, IEEE) received the B.S. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2017, and the M.A. degree in electrical engineering in 2019 from Princeton University, Princeton, NJ, USA, where he is currently working toward the Ph.D. degree.

His research interests include high-efficiency/high-density power converters, multiport dc–dc converters, and high-performance power electronics design for data center applications.

Mr. Wang received the National Scholarship, in 2014 and 2016, while he was with Shanghai Jiao Tong University. At Princeton University, he received the First Place Award of the IEEE ECCE Best Student Project Demonstration and the First Place Award from the Innovation Forum of Princeton University in 2019.

Yenan Chen (Member, IEEE) received the bachelor's and Ph.D. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 2010 and 2018, respectively.

Since 2018, he has been a Postdoctoral Research Associate with the Department of Electrical Engineering, Princeton University, Princeton, NJ, USA. His research interests include high-frequency power converters, advanced power electronics architecture, grid-interface power electronics, and renewable energy systems. He holds three issued Chinese patents.

Dr. Chen received the Outstanding Presentation Award at the 2019 IEEE Applied Power Electronics Conference and Exposition and the First Place Award from the Innovation Forum of Princeton University in 2019.

Authorized licensed use limited to: Princeton University. Downloaded on January 24,2021 at 15:53:24 UTC from IEEE Xplore. Restrictions apply.

Jing Yuan (Member, IEEE) received the B.S. degree from the Shenyang University of Technology, Shenyang, China, in 2013, and two S.M. degrees from the China University of Petroleum (East China), Qingdao, China, and from the Khalifa University, Abu Dhabi, United Arab Emirates, in 2017. He is currently working toward the Ph.D. degree with the Department of Energy Technology, Aalborg University, Aalborg, Denmark.

From September 2019 to January 2020, he was a Visiting Student Research Collaborator with the Department of Electrical Engineering, and the Andlinger Center for Energy and the Environment, Princeton University, Princeton, NJ, USA. His research interests include high-performance power converters, high-frequency power electronics and grid-connected system design.

Mr. Yuan received the First Place Award of the IEEE ECCE Best Student Project Demonstration and the Young Professionals & Student Award at the 2019 IEEE International Conference on Compatibility, Power Electronics and Power Engineering.

Robert C. N. Pilawa-Podgurski (Member, IEEE) was born in Hedemora, Sweden. He received dual B.S. degrees in physics, and electrical engineering and computer science, the M.Eng. degree in electrical engineering and computer science, and the Ph.D. degree in electrical engineering, from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2005, 2007, and 2012, respectively.

He is currently an Associate Professor with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA. Previously, he was an Associate Professor of Electrical and Computer Engineering with the University of Illinois Urbana–Champaign (UIUC), Champaign, IL, USA. 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 served as Student Activities Chair for IEEE Energy Conversion Congress and Exposition Conference, in 2016 and 2017, and as a Technical Co-Chair for the 4th IEEE Workshop on Wide Bandgap Power Devices and Applications, in 2016. From 2014 to 2016, he served as the Award Chair for the IEEE Power Electronics Society (PELS) Technical Committee 6—High Performance and Emerging Technologies, where he is currently the Vice-Chair. From 2016 to 2019, he served as Chair of the IEEE PELS Technical Committee 2—Power Conversion Systems and Components. From 2014 to 2019, he was an Associate Editor for the IEEE TRANSACTIONS ON POWER ELECTRONICS and the IEEE JOURNAL OF EMERGING AND SELECTED TOPICS IN POWER ELECTRONICS. Since 2017, he has served on the Power Management Subcommittee of the IEEE International Solid-State Circuits Conference. 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. In 2015, he received the Air Force Office of Scientific Research Young Investigator Award, the UIUC Dean's Award for Excellence in Research in 2016, the UIUC Campus Distinguished Promotion Award in 2017, and the UIUC ECE Ronald W. Pratt Faculty Outstanding Teaching Award in 2017. He is the recipient of the IEEE Education Society Mac E. Van Valkenburg Award, in 2018, for outstanding contributions to teaching unusually early in his career. He is the co-author of ten IEEE prize papers.

Minjie Chen (Senior Member, IEEE) received the B.S. degree from Tsinghua University, Beijing, China, in 2009, and the S.M., E.E., and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2012, 2014, and 2015, respectively.

He was a Postdoctoral Research Associate with MIT, in 2016. Since 2017, he has been with the Department of Electrical Engineering and the Andlinger Center for Energy and the Environment, Princeton University, Princeton, NJ, USA, where he leads the Princeton Power Electronics Research Lab. He holds five issued U.S. patents. His research interests include high-frequency power electronics, advanced power electronics architectures, power magnetics, machine learning, and the design of high-performance power electronics for emerging and important applications.

Dr. Chen was the recipient of the Prize Paper Awards of the IEEE TRANSACTIONS ON POWER ELECTRONICS in 2017 and 2018, the National Science Foundation CAREER Award, the Dimitris N. Chorafas Award for Outstanding MIT EECS Ph.D. Thesis, The Outstanding Reviewer Award from the IEEE TRANSACTIONS ON POWER ELECTRONICS, two IEEE Energy Conversion Congress and Exposition (ECCE) Best Demonstration Awards, multiple IEEE APEC Outstanding Presentation Awards, a Siebel Energy Institute Research Award, and the First Place Award from the Innovation Forum of Princeton University. He was honored by the Princeton Engineering Commendation List for Outstanding Teaching. He is an Associate Editor for the IEEE TRANSACTIONS ON POWER ELECTRONICS, and the IEEE JOURNAL OF EMERGING AND SELECTED TOPICS IN POWER ELECTRONICS, an Associate Technical Program Committee Chair of the IEEE ECCE in 2019, and the Technical Program Committee Chair of the IEEE International Conference on DC Microgrids in 2021.

Authorized licensed use limited to: Princeton University. Downloaded on January 24,2021 at 15:53:24 UTC from IEEE Xplore. Restrictions apply.