Saras eVR STile - Demonstration of Next Generation Integrated Passives for Next Generation AI & HPC Power Delivery Network Designs

Bartlet DeProspo, Eelco Bergman
Saras Micro Devices
3200 W Germann Road
Chandler, AZ 85286 USA
Ph: 404-904-4833
Email: Bart.DeProspo@sarasmicro.com

Abstract

High-performance microprocessors and accelerators, integral to data centers, require power delivery networks capable of managing higher currents and power densities while operating at lower voltages. The surge in generative AI applications is reshaping the landscape of semiconductor hardware, necessitating advanced power delivery solutions to meet escalating demands for efficiency and performance. As AI workloads intensify, traditional power delivery methods struggle to provide the necessary energy without compromising spatial constraints on circuit boards.

This paper addresses the critical challenges posed by embedded passive components, specifically capacitors and inductors, while demonstrating an embedded voltage regulation (eVR) STile™. We introduce an innovative approach that leverages embedded passive technology to enhance power efficiency. This paper will demonstrate the integration of capacitors and inductors directly into the package substrate to enable for the first time and subsequently vertical power delivery directly into the package substrate. This demonstration of highly integrated passive components serves to optimize the power delivery network through reduction in the power path length as well as leveraging high performance passive components. By addressing these power delivery challenges, we pave the way for the next generation of semiconductor hardware tailored to the demands of next generation AI applications.

Ensuring optimal power signal integrity is vital for maintaining stable voltage delivery across a wide frequency range. A standard design methodology focuses on minimizing the impedance seen by the load's input nodes, thereby reducing voltage ripple and droop during transient events. STile technology improves performance by embedding high capacitance density capacitors close to the load and combining them with complimentary inductors to, significantly reduce parasitics and optimize space for both VPD and embedded voltage regulation (eVR) solutions. This paper will demonstrate the ability to integrate these passive components specifically designed to support integrated voltage regulation inside of the package. This paper will show configurations of passive components that support a variety of different domains for traditional high performance computing die. This paper will also demonstrate the superior performance capable of being achieved as compared to traditional power delivery network's. These traditional power delivery networks perform the voltage conversion from 48V to POL (Point of Load) voltages in discrete and bulky modules on the periphery of the printed circuit board. These modules combine the use of inductors, capacitors and actives however, are significantly far away from the load that efficiency of the network becomes a problem.

To minimize I^2R losses, efficient power delivery demands higher voltages and low current levels at the package substrate level. While conventional power regulator topologies typically utilize inductors to achieve high conversion efficiencies, integrating inductors onto the die can complicate design and elevate costs. The STile technology provides a groundbreaking solution by enabling the co-location of inductors and capacitors within a customizable three-terminal LC passive module (patent pending) embedded in the core substrate. This integration allows for precise alignment with processor power domains, significantly reducing the need for lateral metal routing. Additionally, the flexible layout of the LC passive modules supports multiple power domains, facilitating interleaved multiphase outputs and innovative configurations such as sharing output capacitors among several inductors. By addressing these design challenges, eVR STile technology significantly enhances power delivery efficiency while meeting the stringent demands of next-generation semiconductor applications.

This paper emphasizes the development of advanced, highly integrated capacitor and inductor solutions that deliver superior performance for next-generation power delivery systems. By leveraging STile technology, we illustrate how embedding capacitors and inductors within the substrate not only enhances power signal integrity and minimizes parasitics but also optimizes space for embedded voltage regulation (eVR). This innovative platform enables the creation of custom eVR designs tailored to specific application requirements, significantly improving efficiency in high-performance computing environments. Through our findings, we highlight the transformative potential of these embedded passive components in addressing the power delivery challenges faced by next generation artificial intelligence chips.

Key words

Capacitor, Heterogeneous Integration, Power Delivery, STile, Saras

I. Introduction

Ensuring optimal power signal integrity is vital for maintaining stable voltage delivery across a wide frequency range. A standard design methodology focuses on minimizing the impedance seen by the load's input nodes, thereby reducing voltage ripple and droop during transient events [1]. Although connecting multiple surface-mount device (SMD) capacitors in parallel can help achieve low impedance, careful consideration of anti-resonance effects due to inductance is essential, as these can generate unwanted peaks in impedance and lead to voltage instability. The inherent inductance associated with traditional SMD configurations can diminish efficiency and complicate power delivery. In contrast, STile technology improves performance by embedding capacitors closer to the load, significantly reducing parasitics and optimizing space for both VPD and embedded voltage regulation (eVR) solutions. This approach not only alleviates traditional design challenges but also maximizes the available PCB area for critical power modules.

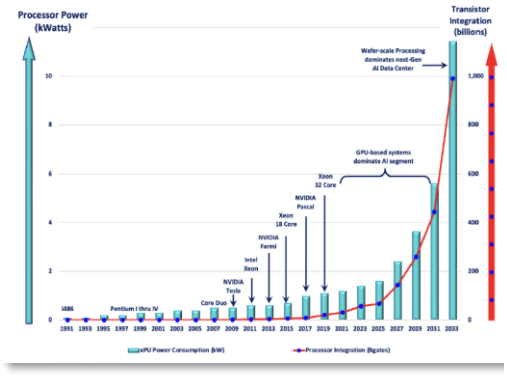

Figure 1. Accelerating Data Center Power Density [3].

To minimize I^2R losses, efficient power delivery demands higher voltages and low current levels at the package substrate level [2]. While conventional power regulator topologies typically utilize inductors to achieve high conversion efficiencies, integrating inductors into the die can complicate design and elevate costs. The STile technology provides a groundbreaking solution by enabling the co-location of inductors and capacitors within a customizable three-terminal LC passive module (patent pending) embedded in the core substrate. This integration allows for precise alignment with processor power domains, significantly reducing the need for lateral metal routing. Additionally, the flexible layout of the LC passive modules supports multiple power domains, facilitating interleaved multiphase outputs and innovative configurations such as sharing output capacitors among several inductors. By addressing these design challenges, eVR STile technology significantly enhances power delivery efficiency while meeting the stringent demands of next generation semiconductor applications.

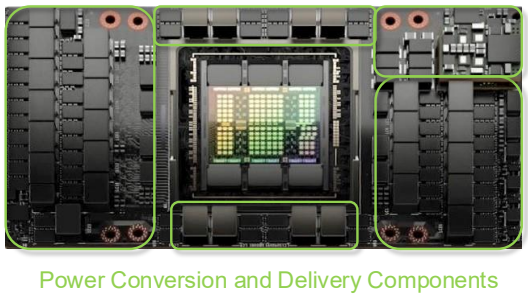

Figure 2. Top down view of H100 nVidia GPU which power delivery network components highlighted [4].

Examining AI data center and power density requirements we can see in Figure 1 that as the data centers have been supporting more artificial intelligence applications the total power and power density requirements have increased exponentially. It is expected that board level power will exceed 10 kW which is representative of a 3-4x increase from where we are at today [3]. This explosive increase in power demand and power density requires innovation in the chip-package level power delivery network to support these

next generation of AI server spaces.

II. Saras eVR STile™ IPD

A. Current Power Delivery Approach

Power Delivery networks today focus highly on leveraging on-chip circuits, surface-mounted components, and large VRM (voltage regulation modules) on the PCB (printed circuit board). All of these different components work together to provide thousands of AMPS of current to the SOC for AI and high-performance computing applications. In figure 2 we examine the H100 nVidia GPU. In this figure we specifically highlight board level voltage regulator modules. These modules are tasked with taking the board input voltage (48V) and performing a single or dual stage power conversion down to the operating voltage of the core. Today's core voltages are <0.7V for the most advanced technology nodes. After the voltage has been converted to the point of load voltage the power still needs to pass through a series of capacitor networks for decoupling. These capacitors are utilized in surface mount and land/die side configurations.

These power delivery networks introduce a variety of losses into the network. The power path length is still significantly long and causes PDN designers to have extremely tight tolerances to maintain effective and reliable power delivery at the point of load. As the amount of current that needs to traverse these networks continue to increase exponentially with each subsequent chip generation the resulting I^2R routing losses continue to pose significantly challenges. These losses manifest heat generation inside of the package which leads to degradation in device performance. Closer examination of the discrete capacitor networks also highlights the need for new capacitor technologies. Surface mount components continue to be placed further from the point of load which means the total number of components and amount of capacitance needed continues to increase over time. As the silicon area continues to occupy more total package area the need for new power delivery network architecture is needed to support the next generation of advanced computer chips.

B. Saras eVR STile™

To address these conventional power delivery and discrete component challenges, Saras has developed an integrated eVR (embeddable voltage regulator) STile™ that offers the following key benefits, the first, fully customizable inductor and capacitor domains to support a variety of voltage regulation, decoupling and filtering functions needed. As multiple power rails with different voltages are required the eVR STile™ can be customized and tailored to support a variety of on-chip power requirements. Second, components that are designed to be embedded. Current discrete capacitor components are not designed to be embedded inside of the package substrate. Currently silicon DTC (deep trench capacitors) have challenges being embedded from thickness and material incompatibility. As DTC's are thin and made of silicon, whereas AI and HPC package substrates require thick organic cores for mechanical rigidity and large body sizes to support large monolithic SOC and chiplet architectures. Other capacitor technologies such as MLCC's can't be fabricated to tolerance requirements tight enough

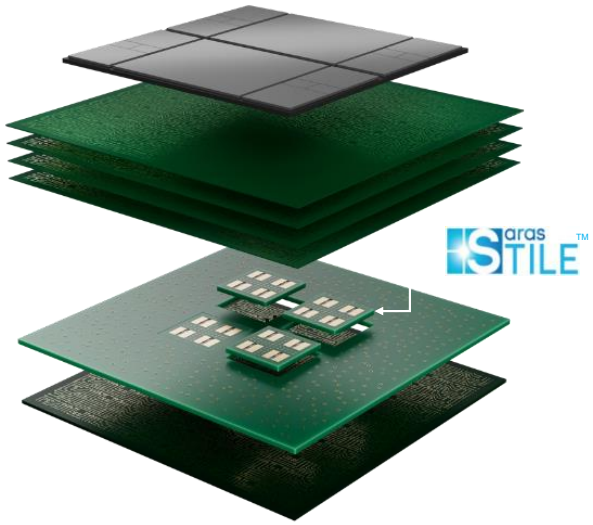

Figure 3. 3D exploded view of the eVR STile showing the die level (grey) the substrate RDL (green) the organic core with Saras eVR STile

and are incompatible with substrate redistribution layer build up processes that causes them to significantly degrade. Saras eVR STile™ will demonstrate a customizable capacitor platform that is designed to be embedded into a variety of package substrate materials with performance stability and superiority to those currently available. Figure 4 captures and highlights the technology advantage of this capacitor technology as compared to the other technologies that could be embedded into the package substrate. The eVR STile™ platform also demonstrates customizable inductors that can be tailored and/or matched to the capacitor technology from a frequency and performance requirements perspective. Current inductor technology is ideal for being integrated on chip in the upper metal levels of the BEOL (back-end-of-line) or they are extremely bulky and incompatible with being embedded in the package substrate. Saras inductor technology enables power delivery network designs a fully customizable embedded power module for high current device architectures. By leveraging both these discrete technologies into a single integrated platform, it provides our

third benefit of significantly reducing the power path length to be vertical through the package substrate. This minimizes the I^2R losses and relaxes the tolerances on the design side for the power delivery network. Saras eVR STile™ also can be co-integrated with actives to provide additional functionality inside the package substrate if needed.

Figure 3 exemplifies this next generation of power delivery leveraging the eVR STile™ technology. In this expanded view we show how the power delivery components sit directly in the die shadow to enable vertical power delivery with the lowest losses. This integration works not only for monolithic SOC integration, but it also works with chiplet integration architectures as well. For chiplet architectures it is easy for one to envision that each chiplet would have a subsequently matched STile™ to enable quick IP reuse and to minimize design cycles in the future.

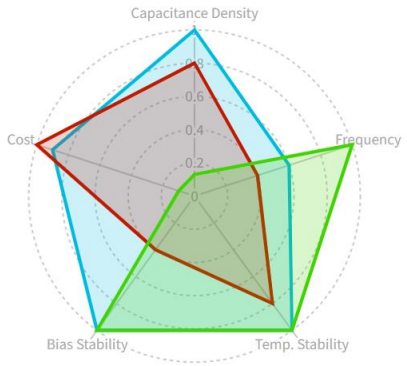

Figure 4. Examination of capacitor technology. Red is representative of silicon DTC's, Green represents MLCC's (under operating bias and temperatures) and Blue highlights the Saras Capacitor technology.

STile™ also provides customization independent to the passive components. STile™ can be customized to include TTV (through tile vias) that enable the signals to be routed vertically into and out of the package substrate without having to route those signal traces around the perimeter of the package. Furthermore, STile technology is built using similar organic materials and processes that are compatible to the RDL build up. This enables the organic package to be still matched closely with the silicon interposer and die's from a CTE perspective. This technology can also be customized and utilized inside of the PCB to enable vertical power delivery and for small power platforms that can be used for advancement of discrete power modules as well.

III. Conclusion

This paper examines a new paradigm in power delivery by presenting for the first time Saras eVR STile™. The eVR STile™ shows customizable and advanced capacitor and inductor solutions capable of support advanced power delivery network architectures. This technology is designed to be embedded into the package substrate. This enables vertical power delivery which reduces the number of discrete components and voltage regulation modules while also reducing the losses incurred from long power delivery network path lengths.

Acknowledgment

The author would like to acknowledge the diligence and work ethic of the Sars Micro Devices engineering team. The author would also like to acknowledge our customers and supply chain partners that have helped us push the limits of what is traditionally possible from a design, equipment and material innovation perspective.

References

[1] J. Kim *et al*., "Chiplet/Interposer Co-Design for Power Delivery Network Optimization in Heterogeneous 2.5-D ICs," in *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 11, no. 12, pp. 2148-2157, Dec. 2021, doi: 10.1109/TCPMT.2021.3113664.K. Radhakrishnan, Power Delivery for Heterogeneous Systems, Connecting Heterogeneous Systems Summit, September 2021.

[2] W. Liu, G. Chen, Q. Ding and J. Jiang, "A Comprehensive Methodology for Optimizing Power Integrity of High-Performance IC Packages," *2023 IEEE 32nd Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*, Milpitas, CA, USA, 2023, pp. 1-3, doi: 10.1109/EPEPS58208.2023.10314865.

[3] [1]Emilio and Emilio, "Navitas Semiconductor Unveils High-Efficiency 4.5 kW Power Supply for AI Data Centers - Power Electronics News," Power Electronics News, Aug. 08, 2024. https://www.powerelectronicsnews.com/navitas-semiconductor-unveils-high-efficiency-4-5-kw-power-supply-for-ai-data-centers/ (accessed Feb. 14, 2025).

[4] "NVIDIA H100 Tensor Core GPU Architecture Overview," *nVidia*, 2023. Web.