Modular Multiphase Voltage Regulator with Cell Stacking and FinFET Stacking Techniques

Li-Jen Huang*, Yu-His Lin*, Yu-Tse Shih*, Jen-Wei Chang*, Ke-Horng Chen*, Tsung-Heng Tsai*, Ying-Hsi Lin**, Sheng-Hsi Hung†, Philex Ming-Yan Fan††, Xi Zhu†††, Email: khchen@nycu.edu.tw
*Department of Electrical and Computer Engineering, National Chiao Tung University, Hsinchu, Taiwan, **Realtek Semiconductor Corporation, Hsinchu 30010, Taiwan, Department of Electrical Engineering, †National Tsing Hua University, ††National Cheng Kung University, Tainan, Taiwan, †††University of Technology Sydney, Australia

Abstract— A modular multiphase voltage regulator (MMVR) implemented in a 7-nm FinFET process is proposed to address device mismatch challenges in 6V-to-0.5V power conversion for advanced SoCs. Leveraging an optimized FinFET stacking strategy, the design maximizes power density while ensuring effective charge balancing across flying capacitors. The proposed MMVR achieves a peak efficiency of 94.2% and demonstrates strong transient performance, with a 91 mV overshoot and a 67 ns recovery time under a 5 A load step. These results validate that FinFET technologies can simultaneously deliver ultra-high efficiency and fast transient response without suffering from mismatch-induced limitations.

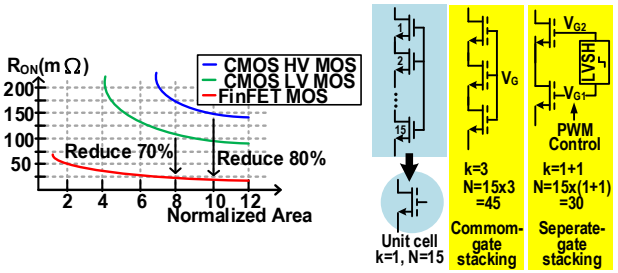
Keywords—Capacitor charge balance, Cell stacking, Fast transient response, FinFET stacking, high efficiency, High power density, Low-power active snubber circuit (LPASC), Threshold clamping circuit (TCC).

I. Introduction

The growing demand for high-performance computing (HPC) and artificial intelligence (AI) has accelerated the adoption of advanced FinFET nodes, enabling reduced R_{ON} and monolithic SoC integration. This transition supports high-frequency operation and direct 6V-to-0.5V conversion, which are critical for AI accelerators and mobile SoCs employing dynamic voltage and frequency scaling (DVFS) for energy-efficient sub-threshold operation. As illustrated in Fig. 1(a), R_{ON} is reduced by more than 70% for the same silicon area compared to conventional CMOS processes, resulting in improved conduction efficiency. Furthermore, monolithic integration facilitates higher switching frequencies, enhancing transient response and allowing for smaller inductors, thereby increasing power density. However, the short-channel effects inherent in FinFET technologies introduce significant device mismatch, leading to severe current imbalance in multiphase converters and non-uniform voltage distribution in hybrid architectures. These issues are exacerbated by the low breakdown voltage of FinFET devices, where mismatch-induced voltage spikes can readily exceed device limits and compromise reliability.

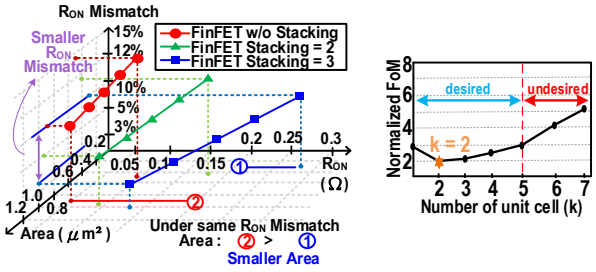
Analog design in FinFET technologies requires stacking a minimum of 15 common-gate devices within a single MOSFET to overcome intrinsic process limitations and ensure device reliability [1]. In addition to these constraints, larger effective device dimensions are necessary to mitigate mismatch. To address this, a modular design approach is adopted, where a 15-device common-gate stack serves as the fundamental unit cell of the power switch, as illustrated in Fig. 1(b). Accordingly, the total stacking number is defined as N=15k, where k is a positive integer representing the number of unit cells.

Two stacking approaches for these unit cells are considered: common-gate stacking and separate-gate stacking. Fig. 2(a) compares conventional long-channel, non-stacked designs with the proposed stacked common-gate structure. The common-gate stacking technique effectively reduces R_{ON} mismatch while improving area efficiency, power density, and performance consistency. To quantitatively evaluate these trade-offs, a normalized figure of merit (FoM) is defined as FoM = (R_{ON} * Q_{OSS} * ($R_{ON\_Mismatch}$)) / Area, as shown in Fig. 2(b). The results indicate that a stacking configuration of N=30 (i.e., 15×2) achieves the optimal balance among conduction loss, switching loss, mismatch, and silicon area under low-voltage operation.



(a) (b)

Fig. 1. (a) Advantage of FinFET device scaling characteristics. (b) Unit cell illustration.



(a) (b)

Fig. 2. (a) Correlation of R_{ON} Mismatch, R_{ON}, and device area. (b) FoM comparison versus the power switch number.

In addition to mitigating intrinsic power-switch mismatch, maintaining impedance symmetry across all conduction paths in a hybrid architecture is essential to avoid current imbalance. To address these challenges, this work proposes an MMVR based on an optimized FinFET stacking strategy. By employing a carefully selected number of common-gate stacked devices, the proposed MMVR effectively suppresses mismatch while achieving accurate impedance matching across hybrid conduction paths.

The remainder of this paper is organized as follows. Section II introduces the proposed MMVR with optimized FinFET stacking. Section III presents the circuit implementation. Section IV provides measurement results and performance comparisons, and Section V concludes the paper.

II. Proposed MMVR Architecture

A. *Phase operation of proposed MMVR*

Fig. 3 illustrates the hybrid power stage of the proposed MMVR, employing a triple cell-stacking architecture. The three stacked cells share four key DC nodes—V_A, V_B, V_C, and V_{OUT}—which distribute voltage stress and enable direct high-voltage regulation using low-breakdown 7-nm FinFET

devices. This structure inherently supports self-balancing of the flying capacitors (C_F). In addition, the cell-stacking architecture is highly scalable, allowing additional cells to be incorporated to increase current-driving capability and further reduce capacitor mismatch.

Fig. 4 illustrates the four operating phases of the proposed MMVR, comprising three interleaved steady-state phases—φ_1, φ_2, φ_a, and a full-on phase φ_U. This interleaved operation ensures balanced charge distribution, reduced V_{OUT} ripple, and equal inductor current sharing (i.e., $I_L/4$ in each flying-capacitor path). During φ_1 (energizing phase), C_{F1}-C_{F3} are charged, V_X is set to $V_{IN}/4$, and the inductor is magnetized. In the subsequent φ_a (de-energizing phase), the inductor voltage is clamped to $-V_{OUT}$, releasing the stored energy to the output. In φ_2, C_{F1}-C_{F3} are discharged, and the inductor is demagnetized by pulling V_X to ground. Under large load transients, the MMVR turns into φ_U to turn on all CS simultaneously, enabling parallel inductor energizing. This mode maximizes the current slew rate, allowing rapid response to abrupt load changes while maintaining stable output regulation.

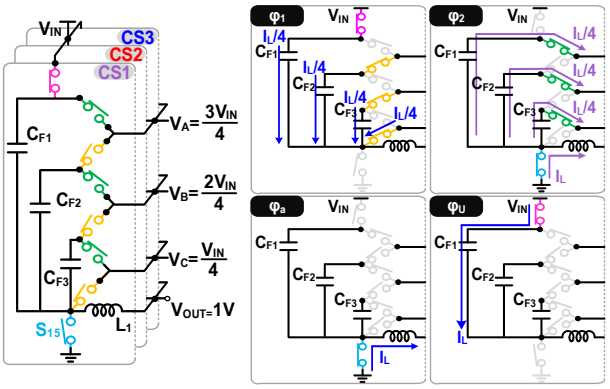


Fig. 3. Architecture of proposed MMVR with triple cell-stacking blocks.

Fig. 4. Three-phase interleaving operation and full-on transient mode of the proposed MMVR.

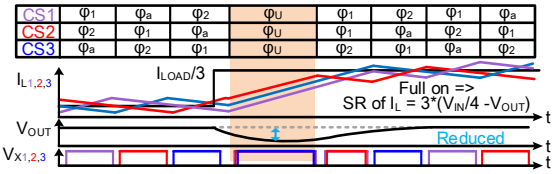


Fig. 5. Timing diagram of the proposed MMVR.

Fig. 5 shows the timing diagram of the proposed MMVR. During steady state, the three CS blocks operate sequentially in φ_1–φ_a–φ_2 order, producing balanced inductor currents and a low-ripple output voltage. When a load transient occurs, the converter enters φ_U, resulting in an effective current slew rate of $3 \times (V_{IN}/4 - V_{OUT})/L$. Once the transient subsides, the system automatically returns to its steady-state interleaving sequence to maintain high efficiency.

B. MMVR with optimized stacked FinFET

The proposed MMVR leverages an optimized FinFET stacking (FS) strategy and a closed-loop control system, as illustrated in Fig. 6. This architecture utilizes a hybrid design to ensure high efficiency and reliability in advanced 7-nm FinFET nodes. By optimizing the stacking count (N=15k) based on current stress, voltage stress, and charge balance, the design achieves superior performance. In Fig. 7, the blue switch in FS1 experiences peak current stress up to $(7/4)\times I_L$, requiring k=1 for low R_{ON} to minimize conduction loss. While the green switch in FS3 encounters overshoot from V_{IN} parasitics during fast C_{F1} charging, a separate-gate stacking of two unit cells (k=1+1) is required to ensure safe blocking. To ensure charge balance is maintained by resistance-matched conduction paths of flying capacitors (C_F), k=2 is designed for the yellow switch in FS2 for a symmetric path of charge and discharge of C_{F2}, while k=5 is designed for the pink switch in FS3. This path-matching strategy keeps voltages of C_{FS} stable, improving regulation, phase symmetry, and overall reliability.

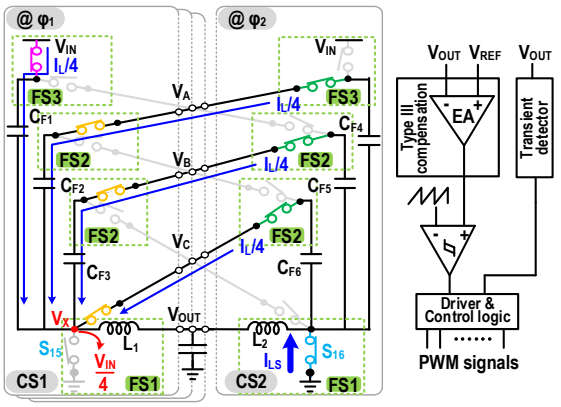


Fig. 6. Proposed hybrid architecture with cell and FinFET stacking.

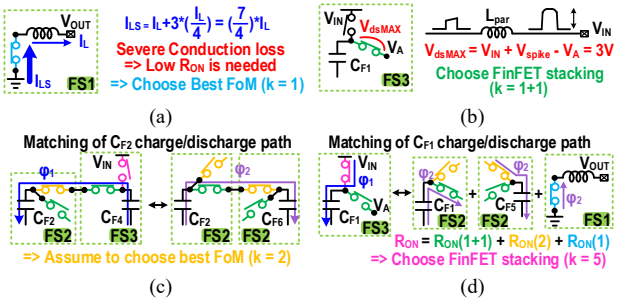


Fig. 7. FinFET stacking in hybrid converter. (a) FS1: conduction loss stress. (b) FS3: overshoot from V_{IN} parasitics. (c) C_{F2}: symmetric charging/discharging path. (d) C_{F1}: symmetric charging/discharging path.

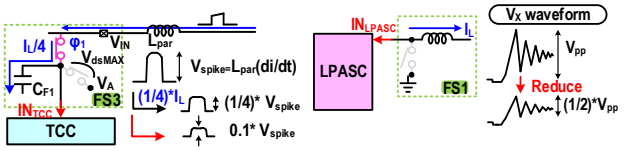
III. Low Power and Robust Protection Circuit

In addition, because FinFET devices are highly sensitive to over-voltage stress. Although the hybrid power stage reduce input current to $I_L/4$ and thus the parasitic-induced V_{spike} is reduced to $V_{spike}/4$. Nevertheless, FS3 remains vulnerable to V_{IN}-path ringing, which can generate overshoot beyond the safe operating limit. As shown in Fig. 8(a), a proposed threshold clamping circuit (TCC) for FS3 is introduced to actively suppress residual V_{spike}. To avoid excessive overshoot and ringing without incurring additional FinFET stacking overhead, a proposed low-power active snubber circuit (LPASC) is integrated in Fig. 8(b). The LPASC dynamically absorbs parasitic energy and damps oscillations, thereby reducing voltage excursions and maintaining FS1 conduction efficiency with low R_{ON}.

A. Threshold Clamping Circuit

As illustrated in Fig. 9(a), the proposed TCC functions as a threshold-based suppression network. Fig. 9(b) shows the operation waveform. When the input node IN_{TCC} exceeds the reverse breakdown voltage of the Zener diode, the Zener conducts current from IN_{TCC} through M_0. Once M_0 conducts, V_1 is asserted, thereby turning on M_2 and pulling V_4 low. The

activation of M_2 subsequently turns on M_3 and M_4, asserting V_2. Meanwhile, these actions simultaneously activate M_5, providing faster response and lower clamping voltage. As a result, M_5 pulls IN_{TCC} low while M_1 also turns on to prevent charging of C_1 at V_0. The TCC remains active until the disturbance at node IN_{TCC} subsides.

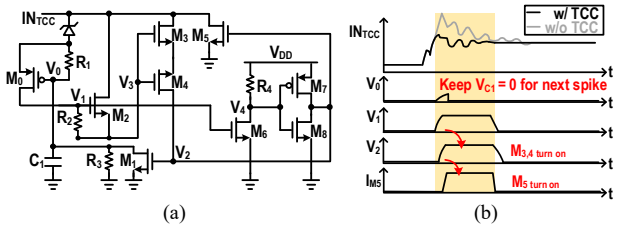


(a) (b)

Fig. 8. (a) Proposed TCC for FS3. (b) Proposed LPASC for FS1.

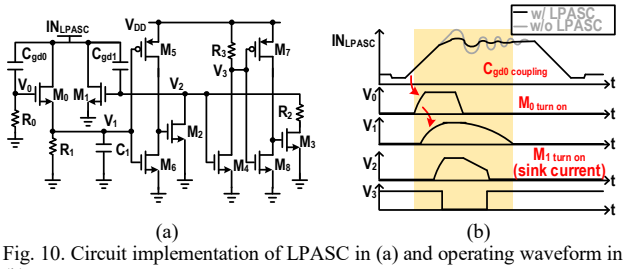
B. *Low-Power Active Snubber Circuit*

The proposed LPASC in Fig. 10 utilizes an active snubber topology. As IN_{LPASC} rises, the coupling through C_{gd0} increases V_0, turning on M_0. This asserts V_1 and switches off M_2. Concurrently, energy transferred via C_{gd1} charges V_2, activating M_1 to sink current from IN_{LPASC}. This triggers M_4 and M_3, allowing V_2 to discharge through R_2 and M_3. Once M_1 and M_4 deactivate, the circuit enters a low-power sleep mode to optimize efficiency.



(a) (b)

Fig. 9. Circuit implementation of TCC in (a) and operating waveform in (b).



(a) (b)

Fig. 10. Circuit implementation of LPASC in (a) and operating waveform in (b).

IV. Experimental Result

Fig. 11 shows the proposed MMVR converter chips fabricated in a 7-nm FinFET process. The die occupies an active silicon area of 640 × 1040 μm², including the power stage, TCC, LPASC, gate drivers, and bootstrap networks, demonstrating the high integration density achievable in advanced FinFET technology.

Fig. 12 shows the steady-state waveform of the MMVR. Fig. 12 (a) shows the effectiveness of the optimized stacking strategy through the switching node voltage (V_x) waveforms. The V_X follows $V_x = V_{IN} - V_{CF1}$, allowing indirect observation of the flying-capacitor voltage (V_{CF1}) behavior. With the optimized stacking configuration, the voltage imbalance across V_{CF1} is substantially mitigated—only 11 and 8mV V_{CF} mismatch is measured. These results confirm that the proposed FinFET stacking optimization effectively enhances charge balance and voltage uniformity across the modular cell-stacking network. Fig. 12(b) shows the steady state inductor current waveform. This configuration effectively partitions the total load current into four parallel branches, with each conducting $I_L/4$, thereby mitigating parasitic-induced voltage spikes (V_{spike}). Under operating conditions of V_{IN} = 6 V, V_{OUT} = 0.5 V, and I_{Load} = 3 A, each inductor carries approximately 0.667 A, resulting in a measured output voltage ripple of only 8 mV.

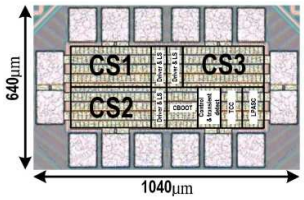
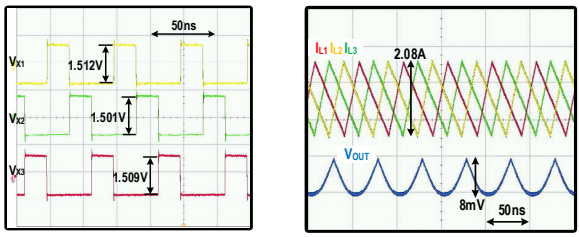


Fig. 11. Die micrograph.

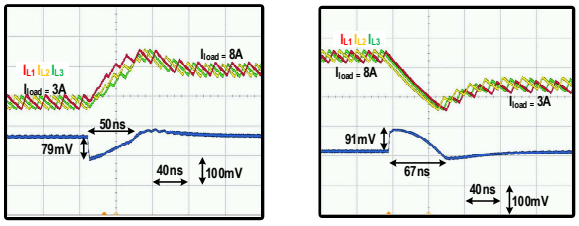


(a) (b)

Fig. 12. Steady state waveform of MMVR. (a) Waveforms of switching node $V_{X1} - V_{X3}$. (b) Waveforms of inductor current I_{L1}-I_{L3}.

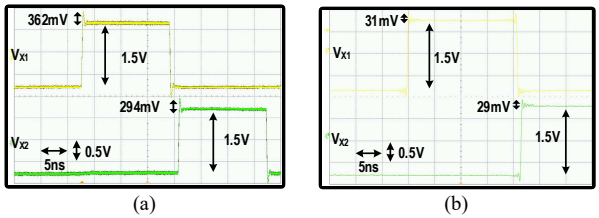
Fig. 13 illustrates the load transient response without protection circuits. For a 3A-to-8A load transient, the converter performs a 79mV undershoot and 50ns recovery time thanks to the high switching frequency and the full-on mode for load transient. Under a 8A-to-3A load transient, the converter performs a 91mV overshoot and 67ns recovery time, showing still great transient performance.

Fig. 14 compares the measured V_X node waveform without and with the proposed protection circuits enabled. Without the proposed protection circuit, the V_{spike} of V_X nodes may rise to 362mV and 294mV, potentially breaking down the devices. With the help of the protection circuit, the V_{spike} on the switching nodes has then been reduced to 31mV and 29mV, respectively. This shows the great effectiveness of the proposed protection circuit.



(a) (b)

Fig. 13. Load transient waveform of MMVR (a) during light to heavy load transient and (b) during heavy to light load transient.



(a) (b)

Fig. 14. Switching node voltage stress (a) without and (b) with TCC and LPASC enabled.

Fig. 15 shows the measured efficiency of the proposed MMVR under different output voltages, with V_{IN} = 6 V and

F_{SW} = 20 MHz. The converter delivers up to 8 A output current, achieving a peak efficiency of 94.2% at V_{OUT} = 0.5 V. At light loads, flying capacitor imbalance has minimal impact on efficiency, but as the load current increases, charge redistribution and capacitor mismatch begin to degrade efficiency. This highlights the importance of robust protection techniques to suppress imbalance and sustain performance at heavy load conditions.

As shown in Fig. 16, performance compares between the die with and without the well-protection circuit enabled. Experimental results confirm that the device with the active protection scheme enabled achieves the best overall performance by combining optimized stacking with active protection circuits, effectively suppressing voltage overshoot and minimizing capacitor imbalance, leading to a consistent efficiency improvement across the entire load range.

As benchmarked in Fig. 17, the proposed MMVR not only achieves high peak efficiency but also delivers the highest current density among reported integrated converters. Compared with prior designs constrained by either limited efficiency or low current capability, the proposed MMVR defines a new performance benchmark by simultaneously enhancing both efficiency and current density. These results validate that the proposed hierarchical stacking architecture, augmented by well-protection circuits and transient enhancement mechanisms, provides a scalable, reliable, and energy-efficient solution for next-generation high-current, high-power AI-server applications.

Table I summarizes the performance of the proposed MMVR and provides a comprehensive comparison with state-of-the-art integrated converters. Fabricated in a 7-nm FinFET process, this work achieves the highest die area current density of 8 A/mm^2 and power density of 242 mW/mm^3 among the referenced designs. By operating at a high switching frequency of 20 MHz, the converter significantly reduces passive component sizes while maintaining a peak efficiency of 94.2% at a 3 A load. Notably, the proposed architecture exhibits superior transient handling, with a recovery time of only 67 ns for a 5 A load step, which is significantly faster than prior hybrid or multilevel topologies.

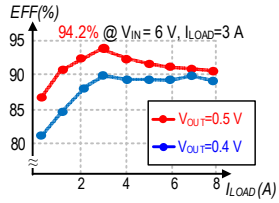


Fig. 15. Measured efficiency of the proposed MMVR at different output voltages.

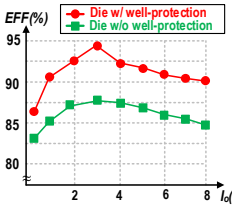


Fig. 16. Efficiency comparison with and without the well-protection enabled.

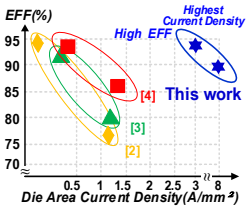


Fig. 17. Benchmark of efficiency versus current density compared with prior works.

V. Conclusion

This paper presents a 7-nm FinFET modular multiphase voltage regulator (MMVR) designed for high-density power conversion. By utilizing a modular FinFET stacking strategy (N=15k), the design effectively overcomes device mismatch and process limitations inherent in advanced nodes. The triple cell-stacking architecture ensures flying capacitor auto-balancing while enabling direct 6V-to-0.5V regulation. Experimental results demonstrate a peak efficiency of 94.2% and a record-high power density of 242 mW/mm^3. With an ultra-fast recovery time of 67 ns under a 5 A load step, this work provides a scalable and reliable power management solution for next-generation AI and HPC applications.

TABLE I. Performance Summary and Comparison

Design	[2] ISSCC' 23	[3] ISSCC' 25	[4] ISSCC' 19	This Work
Topology	Dual-path SC	SI	4L FCML	MMVR
Technology (nm)	180	180	22	7
Input Voltage (V)	9-16	12	5	6
Output Voltage (V)	0.6-1.6	1-1.8	1.8	0.4-0.5
Max. Output Current (A)	5	6	10	8
Flying Capacitor (F)	10µ x 4	1µ x 2	13.2 x 2µ	1µ x 9
Inductor (H)	680n	600n x 3	10n	8n x 3
Output Capacitor (F)	22µ	10µ	18µ	10µ
Switching Frequency (Hz)	1M	2M	5M	20M
Peak Efficiency (%)	94.5 @ 0.8A	92.9 @ 2.25A	93.8 @ 3A	94.2 @ 3A
Die Area Current Density (A / mm²)	1.294	1	1.45	8
Power Density (mW / mm³)	27.1	153	198	242
Load Step (ΔI_{LOAD} / T_{EDGE})	2A / 200n	3.5A / 20n	6A / (NA)	5A / 20n
Overs/Undershoot (mV)	125 / 128	92 / 190	(NA) / 170	91 / 79
Recovery Time	32 µs	NA	3 µs	67 ns
Max. SR	(V_X - V_{OUT}) / L	3(V_X - V_{OUT}) / L	(V_X - V_{OUT}) / L	3(V_X - V_{OUT}) / L

Acknowledgement

The authors would like to thank Ying-Hsi Lin, Vice President of Realtek Semiconductor Corp., Prof. Win-San Vince Khwa, Dr. Meng-Fan (Marvin) Chang, and the Taiwan Semiconductor Manufacturing Company (TSMC) Joint Development Project (JDP) for their valuable support, technical discussions, and collaboration.

References

[1] C. -E. Hsia et al., "Current Mirrors with Tapered Stacked-Gates for Area Saving or Noise Improvement in 3nm FinFET Process," *2024 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits)*, Honolulu, HI, USA, 2024, pp. 1-2

[2] X. Yang et al., "11.8 A 5A 94.5% Peak Efficiency 9~16V-to-1V Dual-Path Series-Capacitor Converter with Full Duty Range and Low V.A Metric," *IEEE International Solid-State Circuits Conference (ISSCC)*, pp. 196-198, 2023.

[3] J. Yang, R. P. Martins and M. Huang, "21.3 A Segmented-Interlacing Multi-Phase Hybrid Converter with Inherently Auto-Balanced ILs and Boosted IL Slew Rate During Load Transients," *IEEE International Solid-State Circuits Conference (ISSCC)*, pp. 378-380, 2025.

[4] C. Schaef et al., "8.1 A 93.8% Peak Efficiency, 5V-Input, 10A Max ILOAD Flying Capacitor Multilevel Converter in 22nm CMOS Featuring Wide Output Voltage Range and Flying Capacitor Precharging," *IEEE International Solid-State Circuits Conference (ISSCC)*, pp. 146-148, 2019.