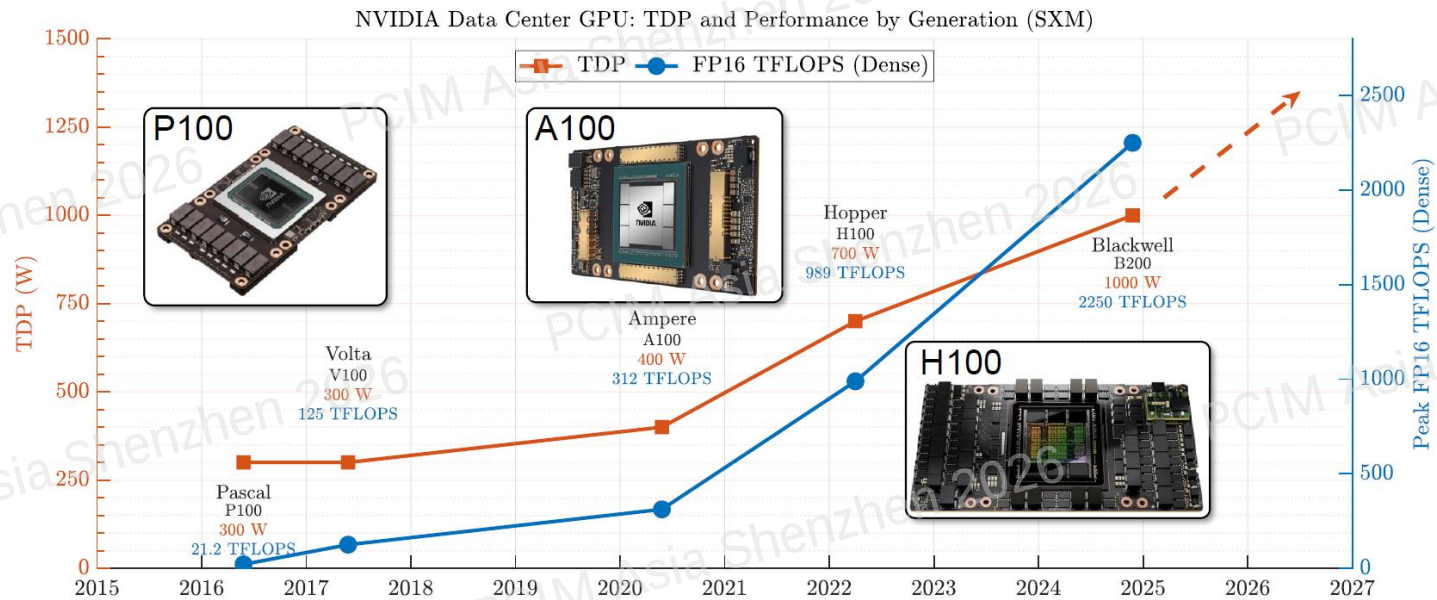


Topology and Magnetism for Vertical Power Delivery

Yenan Chen, Zhejiang University

Power Supply Challenge for GPU

- Exponential increases in power consumption
- Board area limit
- **IC (Moore's Law) vs Power Electronics**



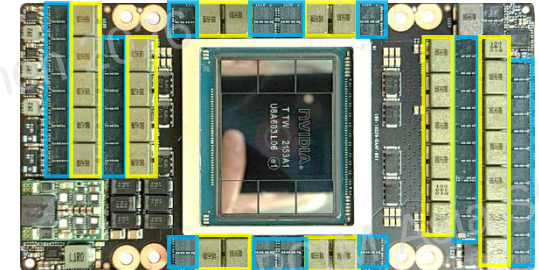
- NVIDIA @ APEC26

2017



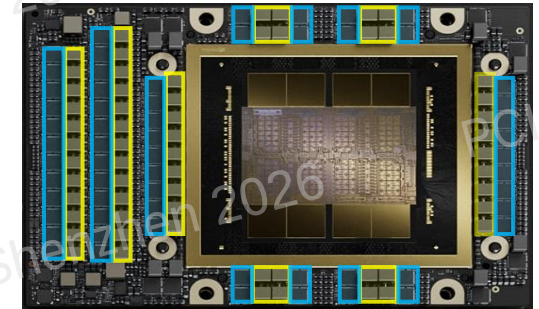
V100~300 W 16-Phase VRM

2022



H100~700 W 60-Phase VRM

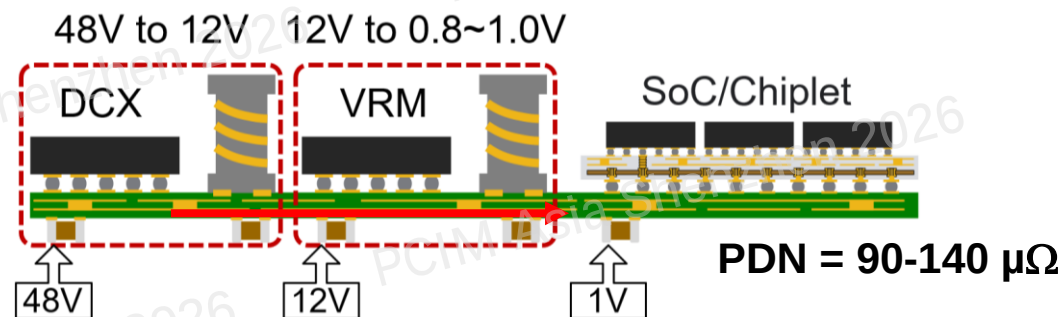
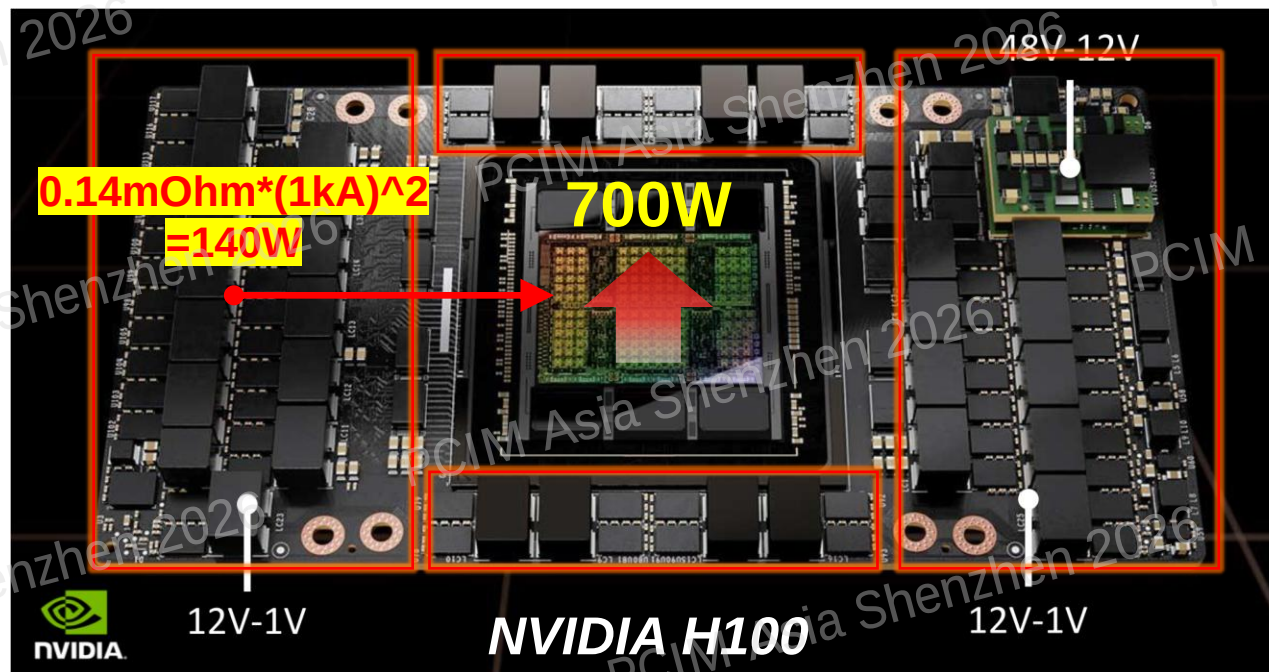
2024



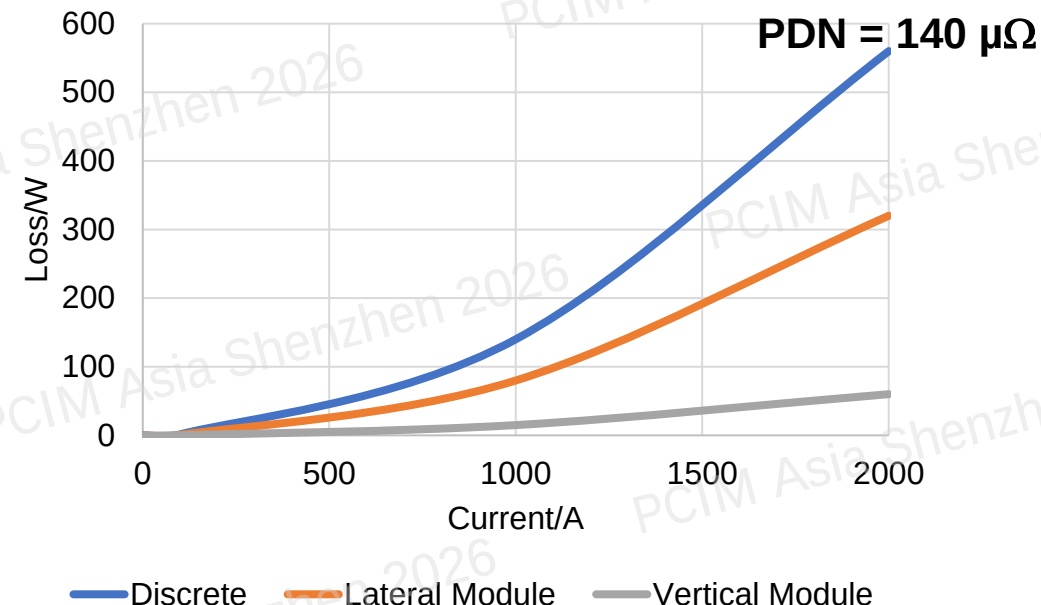
B200~1000 W >60-Phase VRM

Demand for Vertical Power Delivery

No more space for power on the PCB

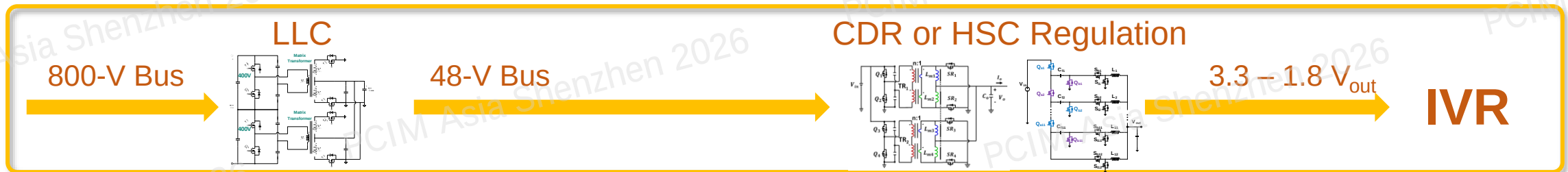
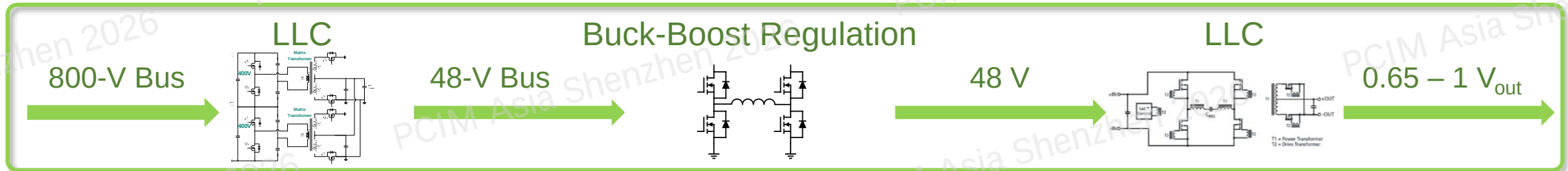
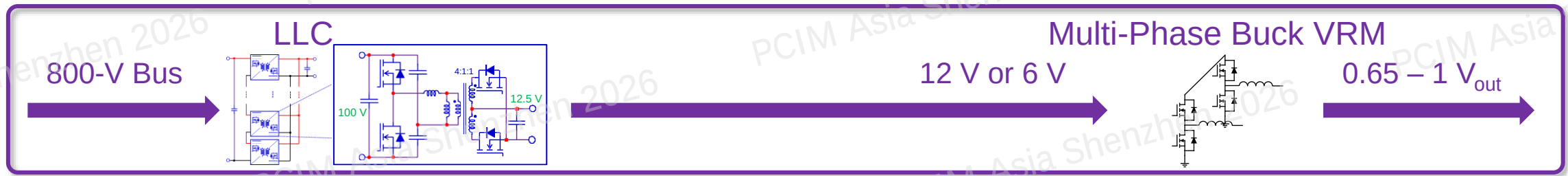
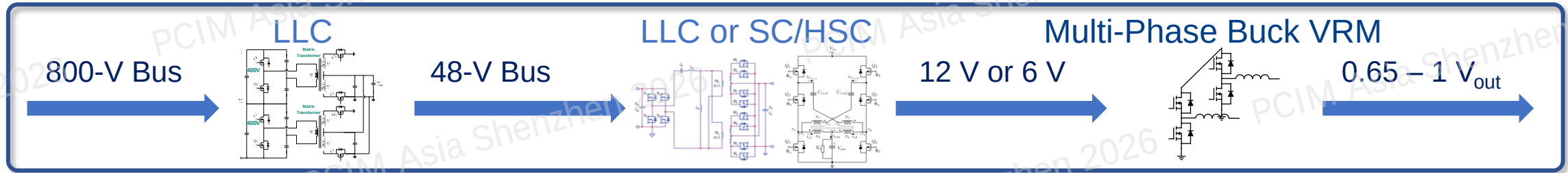


PDN Loss vs. Compute Current[1]



- PCB space constrained
- Power distribution network I^2R losses increase
- GPU fast dynamic > 5 A/ns, need lower parasitic L

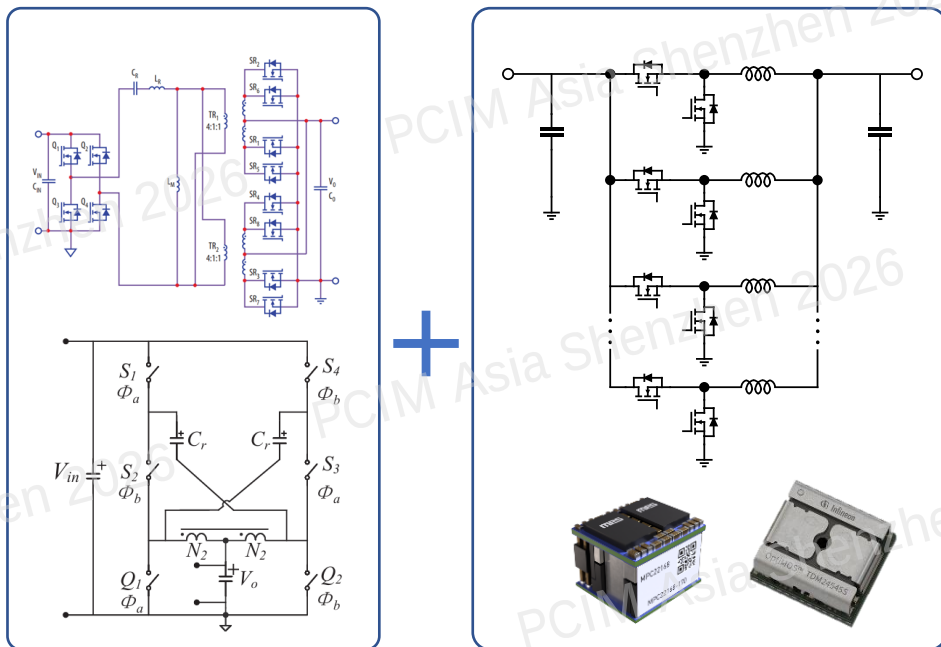
GPU Power Supply: below 800V



- SC: Switched Capacitor
- HSC: Hybrid Switched Capacitor
- CDR: Current Doubler

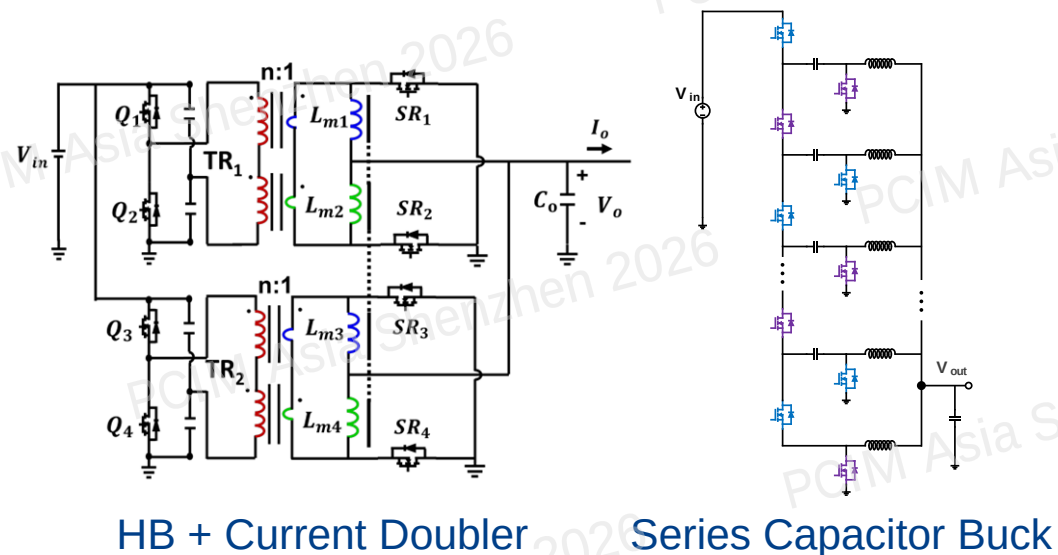
Below 48 V: 2-Stage or 1-Stage

Two-Stage Architecture



- Decoupled Power and Control
- Short current path, low parasitic L
- Distributed electric & thermal stress
- DrMOS and Control 😊

Single-Stage Architecture

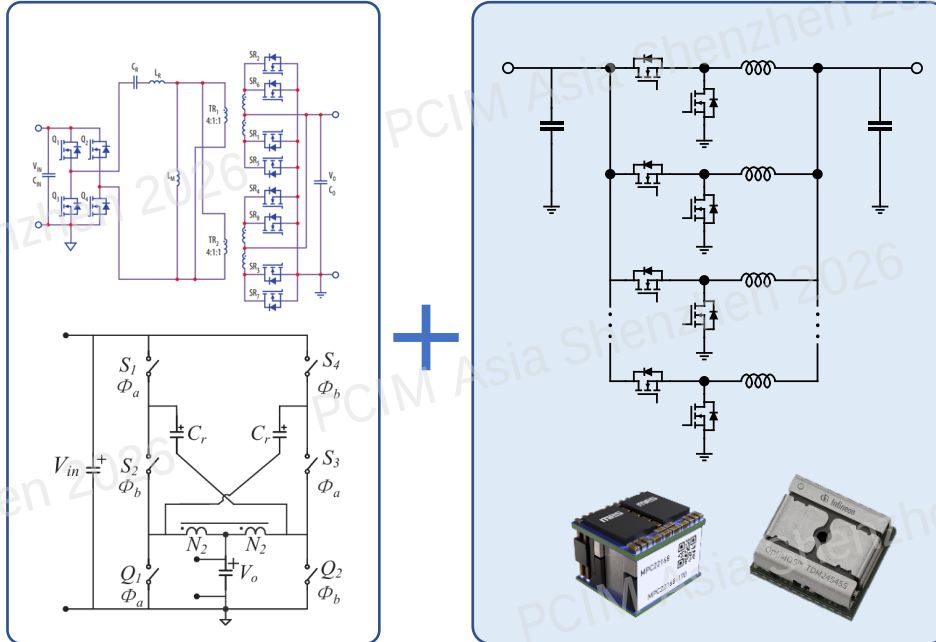


HB + Current Doubler Series Capacitor Buck

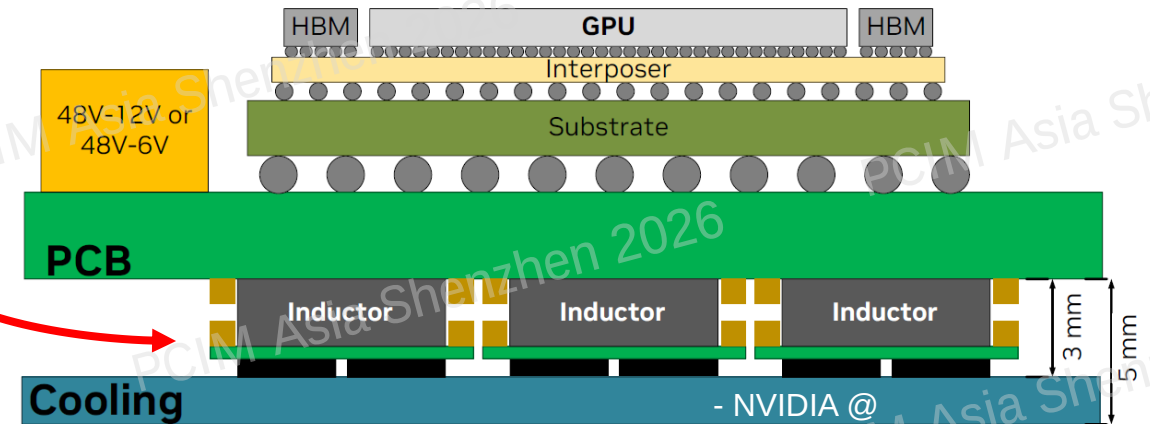
- Fewer components
- Long current path, high parasitic L
- Concentrated electric & thermal stress
- No package integration

Topology for VPD

Two-Stage Architecture



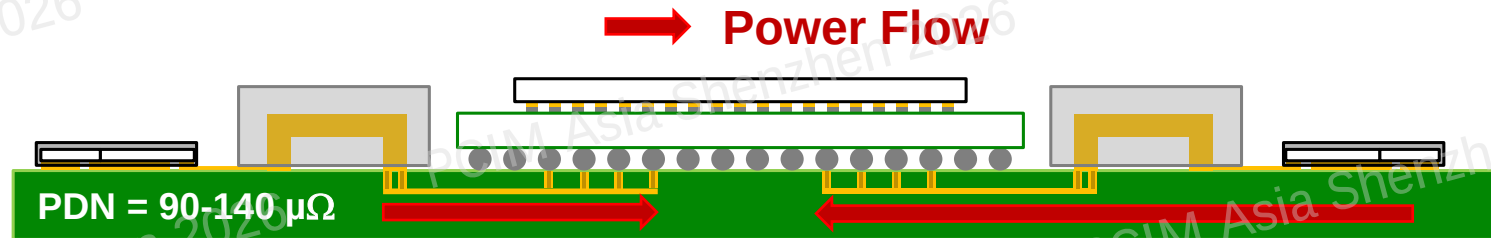
- Decoupled Power and Control
- Short current path, low parasitic L
- Distributed electric & thermal stress
- DrMOS and Control ☺



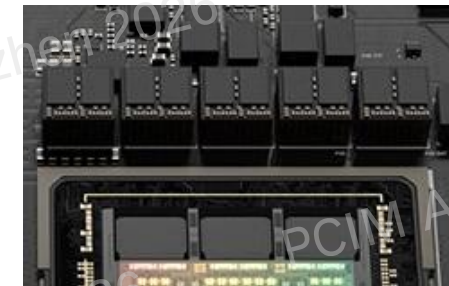
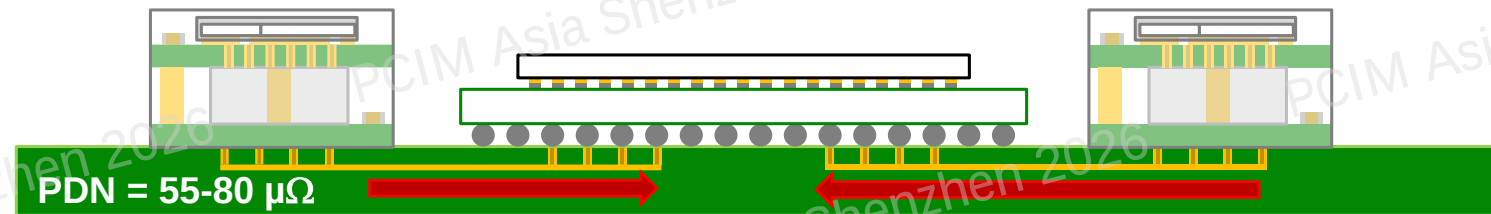
- Limited backside space (height): < 5mm
- What can we put on the back ?
 - Only the voltage regulator module (VRM)
 - Reduce PDN loss between VRM and GPU

From Lateral to Vertical

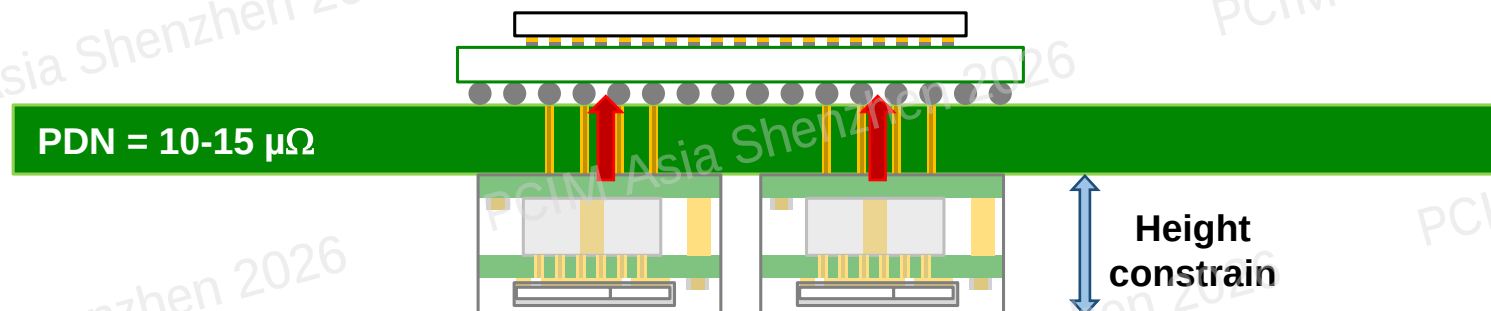
- Discrete



- Lateral Module



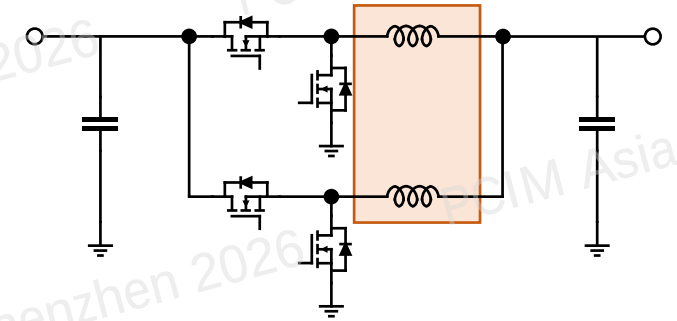
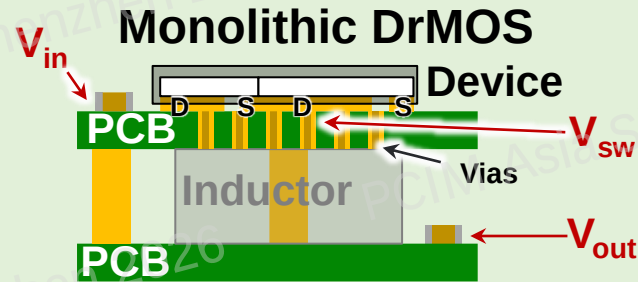
- Vertical Module



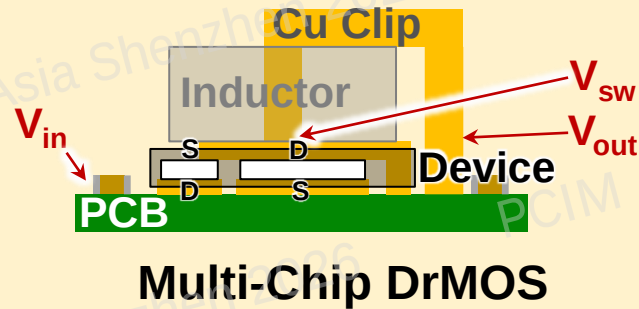
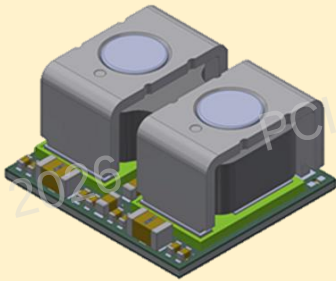
16-Phase VPD Module, $V_{in} = 6.75 \text{ V}$
55.5 x 26.5 x 6.5 mm

VPD Module Structure

Device
On Top
Module

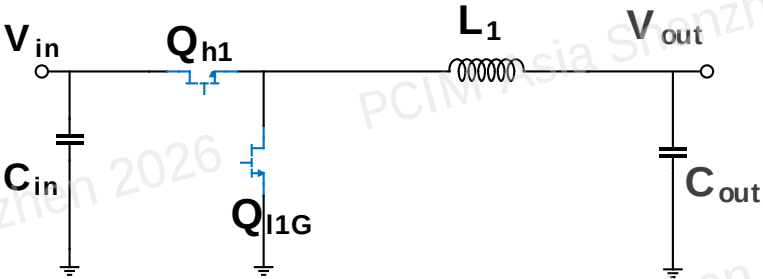


Inductor
On Top
Module



- Height and volume limited by ***the output inductors***
- Need higher f_s !
- But switching loss ?

• Topology

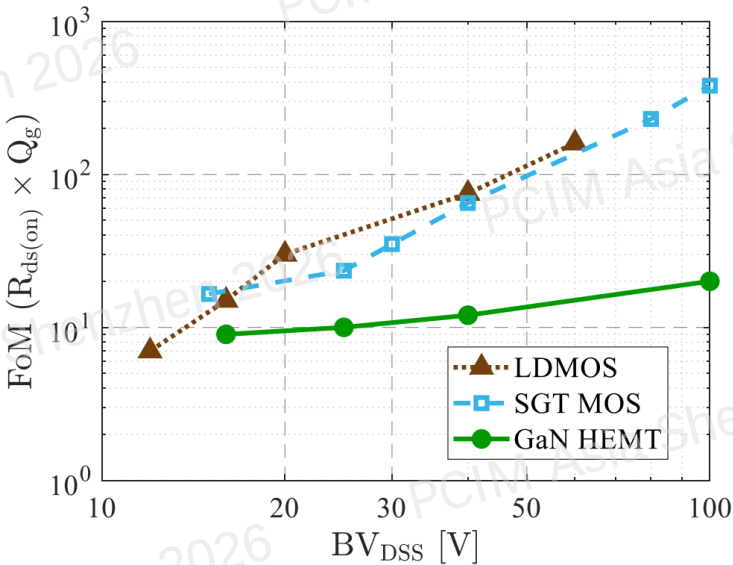


Part No.	BV [†] (V)	R _{ds(on)} (mΩ)	Q _g [‡] (nC)	Q _{oss} ^{††} (nC)	L _s [§] (pH)	L _d ^{§§} (pH)	V _{sd} (V)	Size (mm)
EPC2067	40	1.3	17.1	18	~0	~0	2	3.25 × 2.85
EPC2066	40	0.8	25	30	~0	~0	2	6.05 × 2.3
EPC2055	40	3	6.6	7	~0	~0	1.9	2.5 × 1.5
IQE004	15	0.47	32	27	100	250	0.76	3.3 × 3.3

GaN

Si-MOS

[†] Breakdown Voltage.
[‡] Q_g @ V_{gs} = 5 V.
^{††} Q_{oss} @ V_{ds} = 7.5 V.
[§] Source package parasitic inductance.
^{§§} Drain package parasitic inductance.

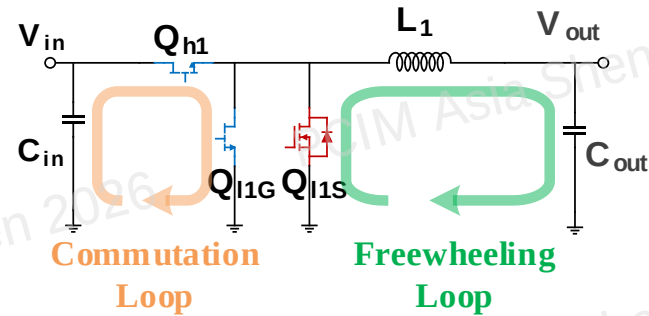


• LV-GaN:

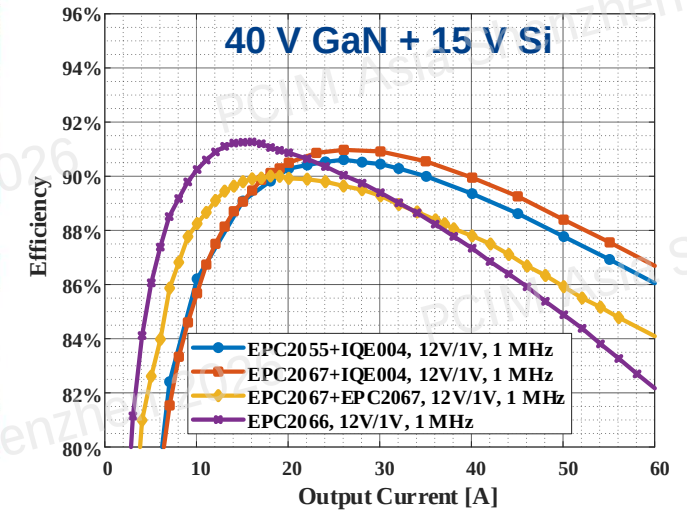
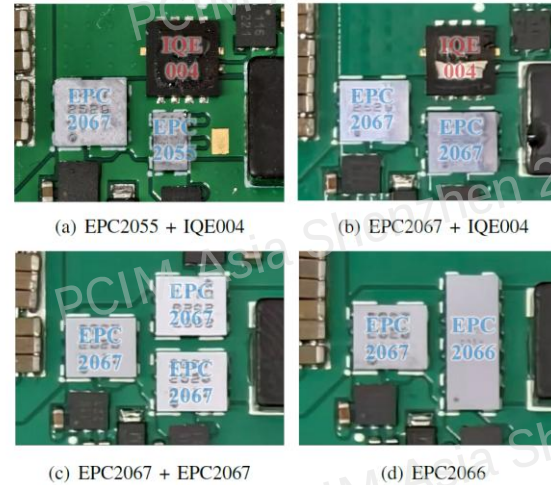
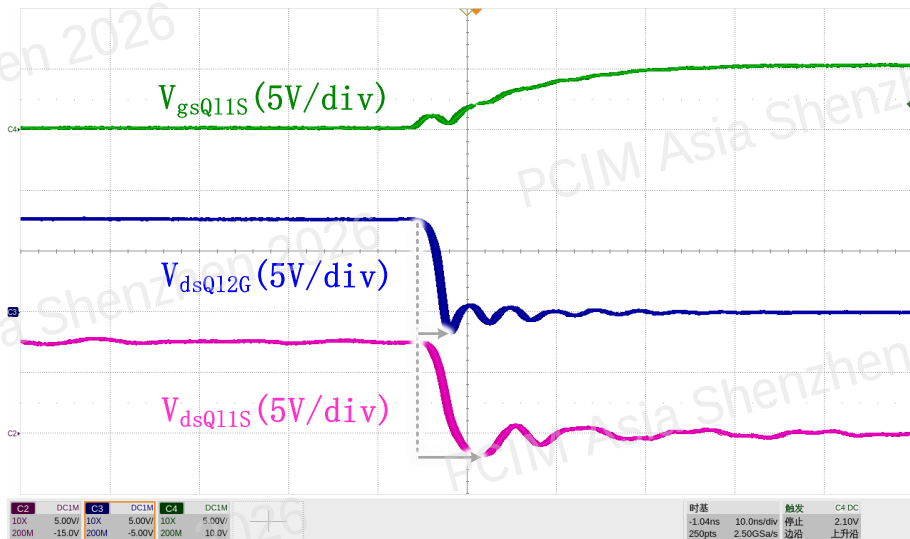
- Better FoM and lower package inductance
- No low voltage & low R_{ds(on)} GaN (1 year ago)
- No Q_{rr}, but higher V_{sd} (deadtime loss)

Hybrid GaN-Si VRM

- Topology



- Faster Switching Speed with GaN

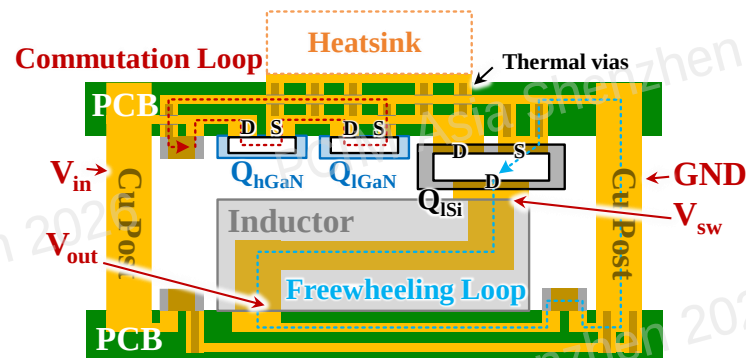


- High-Side: hard-switching, low conduction loss, only
- Low-side: ZVS, high conduction loss, GaN + Si in parallel
- GaN + GaN: low inductance commutation loop
- Si-MOS: low resistance freewheeling loop (large duty)

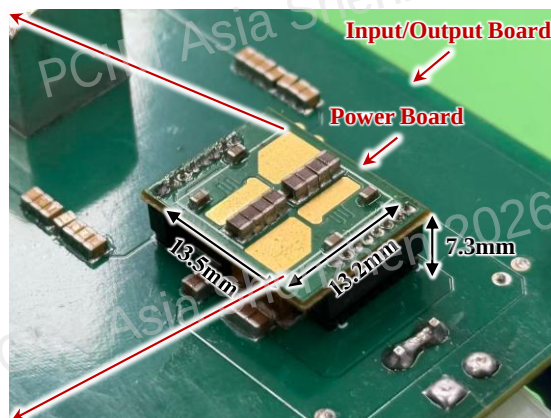
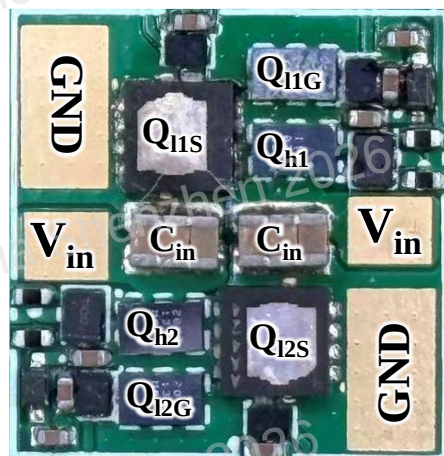
Y. Wu, W. Fan, Y. Chen and D. Xu, "A Two-Phase 140-A/1-MHz GaN-Silicon Hybrid Voltage Regulator Module for Vertical Power Delivery," IEEE TPEL 2026

Hybrid GaN-Si VPD Module

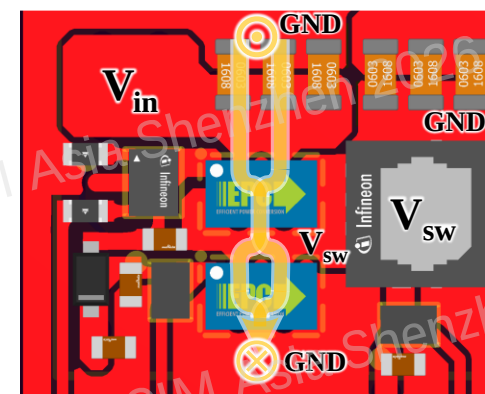
- Module Structure



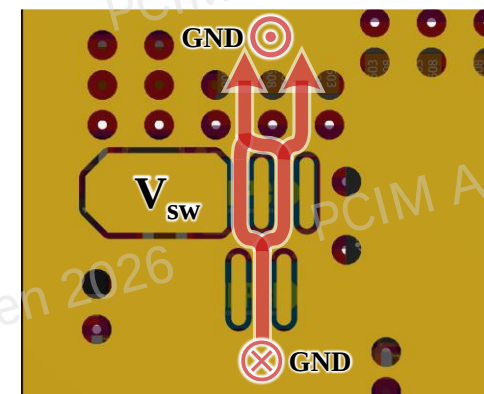
- Design Result



- Commutation Loop $L_{loop} = 0.26 \text{ nH}$

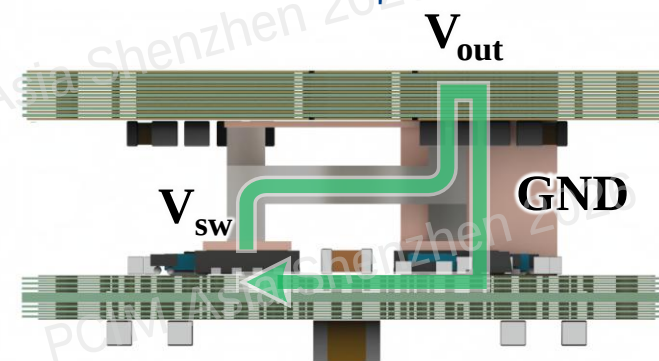


1st Layer

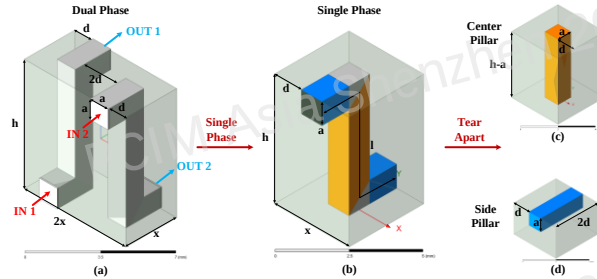


2nd Layer

- Freewheeling Loop $R_{loop} = 0.18 \text{ m}\Omega$



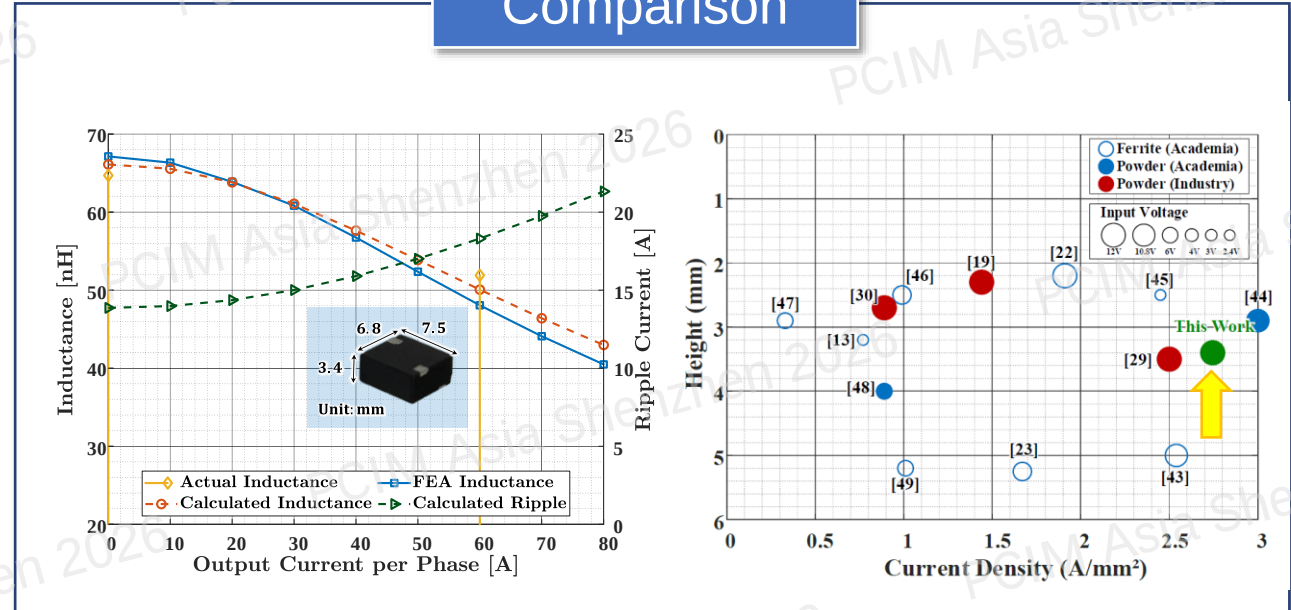
2-Phase Vertical Inductor



Target	Value
$R_{dc,max}/m\Omega$	0.15
I_{set}/A	60
L_{set}/nH	50
Design Parameter	Value
a/mm	1.2
d/mm	1.1
h/mm	7.5
Core Material	POCO XF series $\mu_r(0) = 50$
Test Results [†]	Value
$R_{DC}/m\Omega$	0.17
L_{0A}/nH	64.7
L_{60A}/nH	53.8
Coupling Coefficient	< 0.05

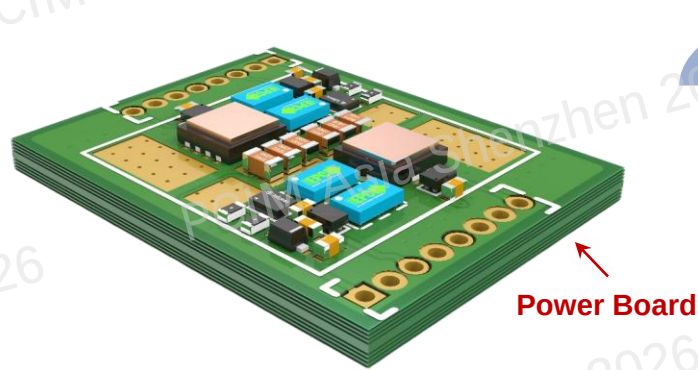
[†] Test current bias is added by Microtest 6220, inductance is measured by Microtest 6330, DCR is measured by Hioki RM3548.

Comparison

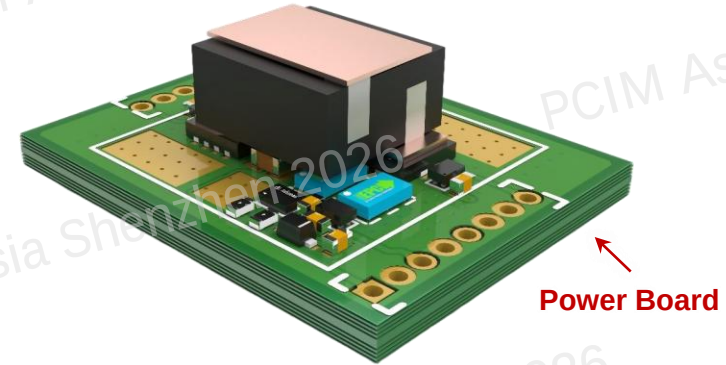


- ☺ Alloy powder core: soft saturation, high magnetic flux density
- ☺ High peak current, high current density
- ☺ Cu-Fe Co-sintering, no air gap
- ☹ Low coupling coefficient

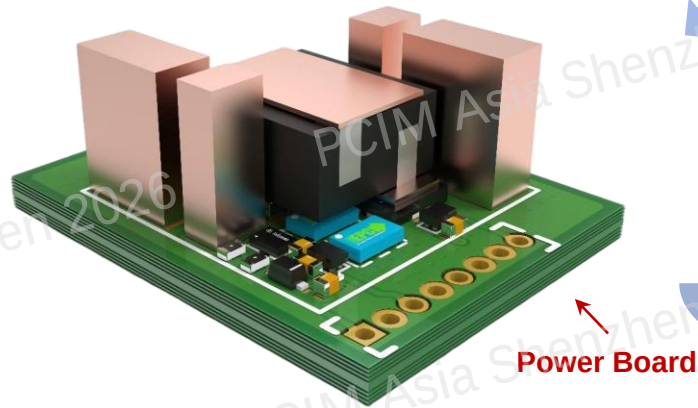
VPD Module Assembly



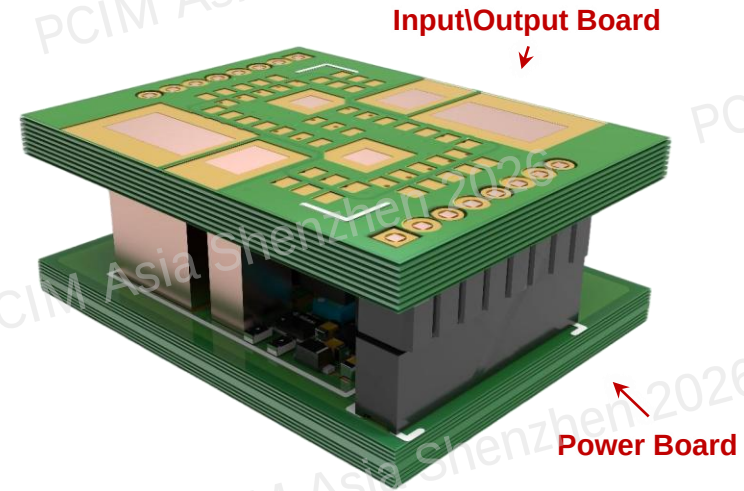
a) Soldering Power Devices



b) Soldering Inductor



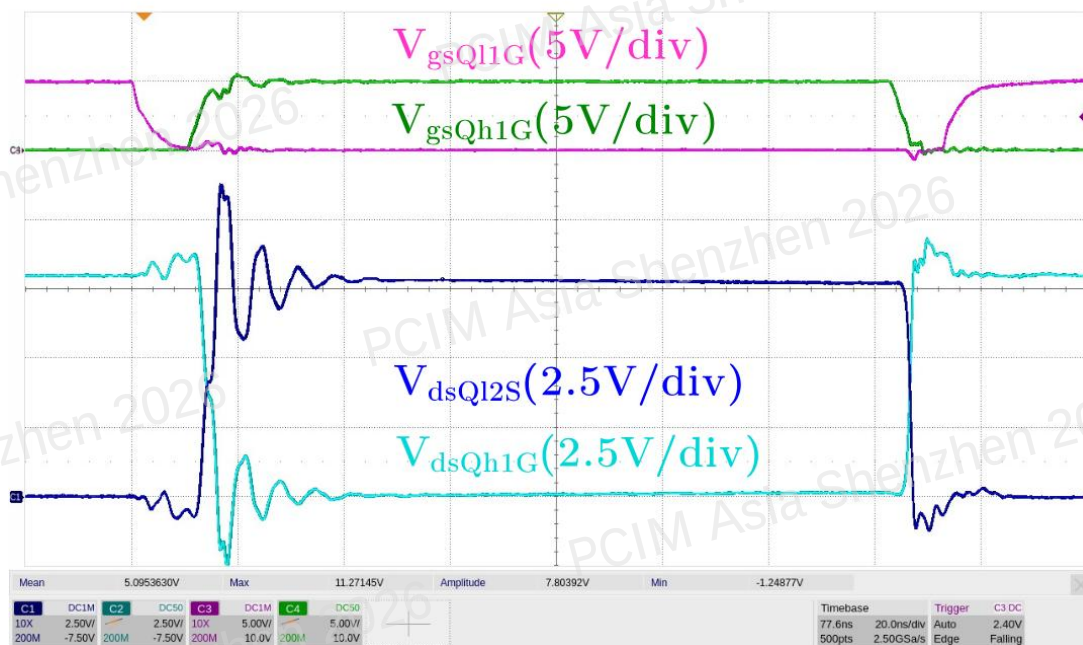
c) Soldering Copper Post



d) Soldering Input\Output Board

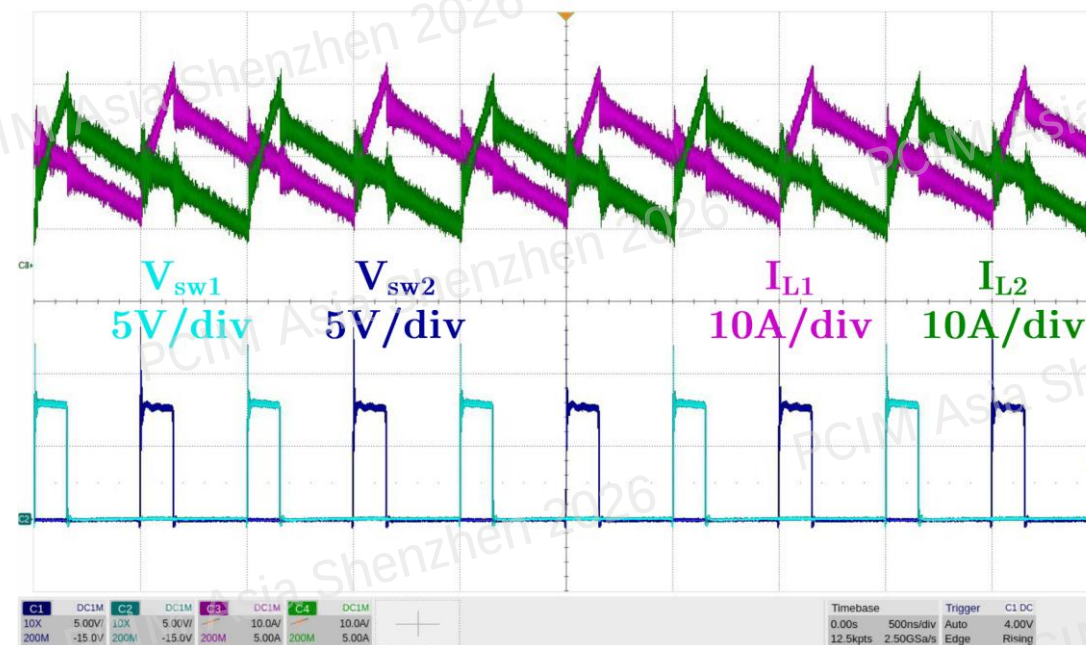
Experiment Waveform

- V_{ds} Waveform 8V/1V 1MHz



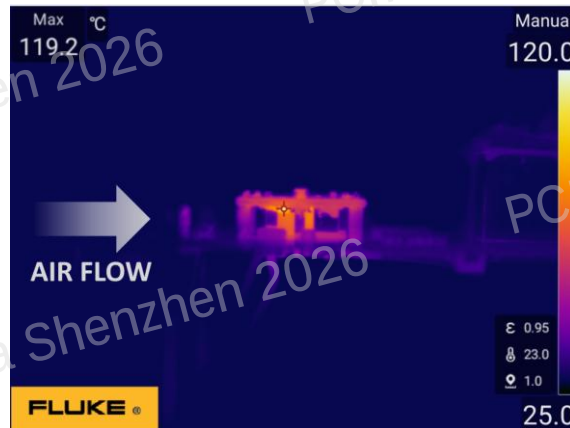
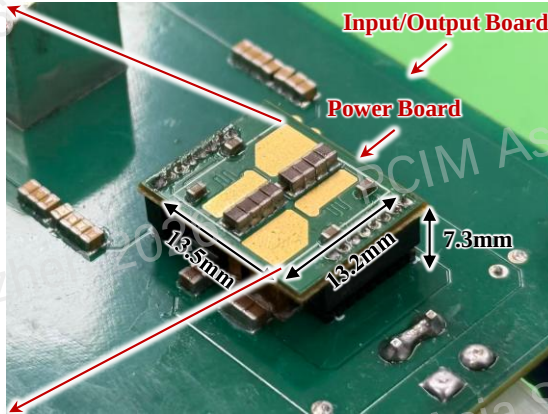
Fast commutation < 10 ns
Low voltage spikes

- Current Waveform 8V/1V 1MHz

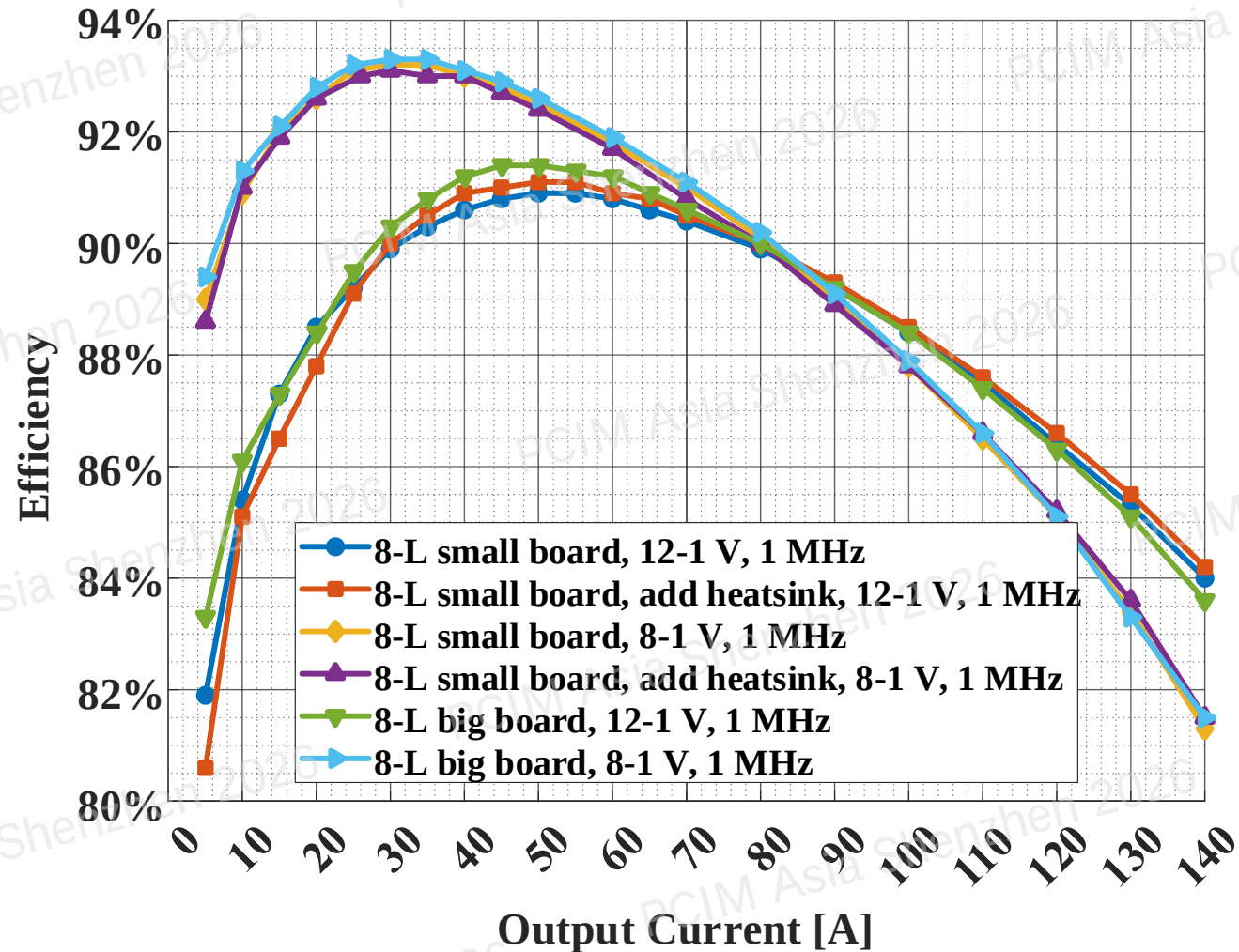


$L_{pss} = 50 \text{ nH}$

Efficiency and Thermal



12 V / 1 V, 1 MHz, 140 A, without heatsink
2 × 32.9 CFM fan cooling from the left side



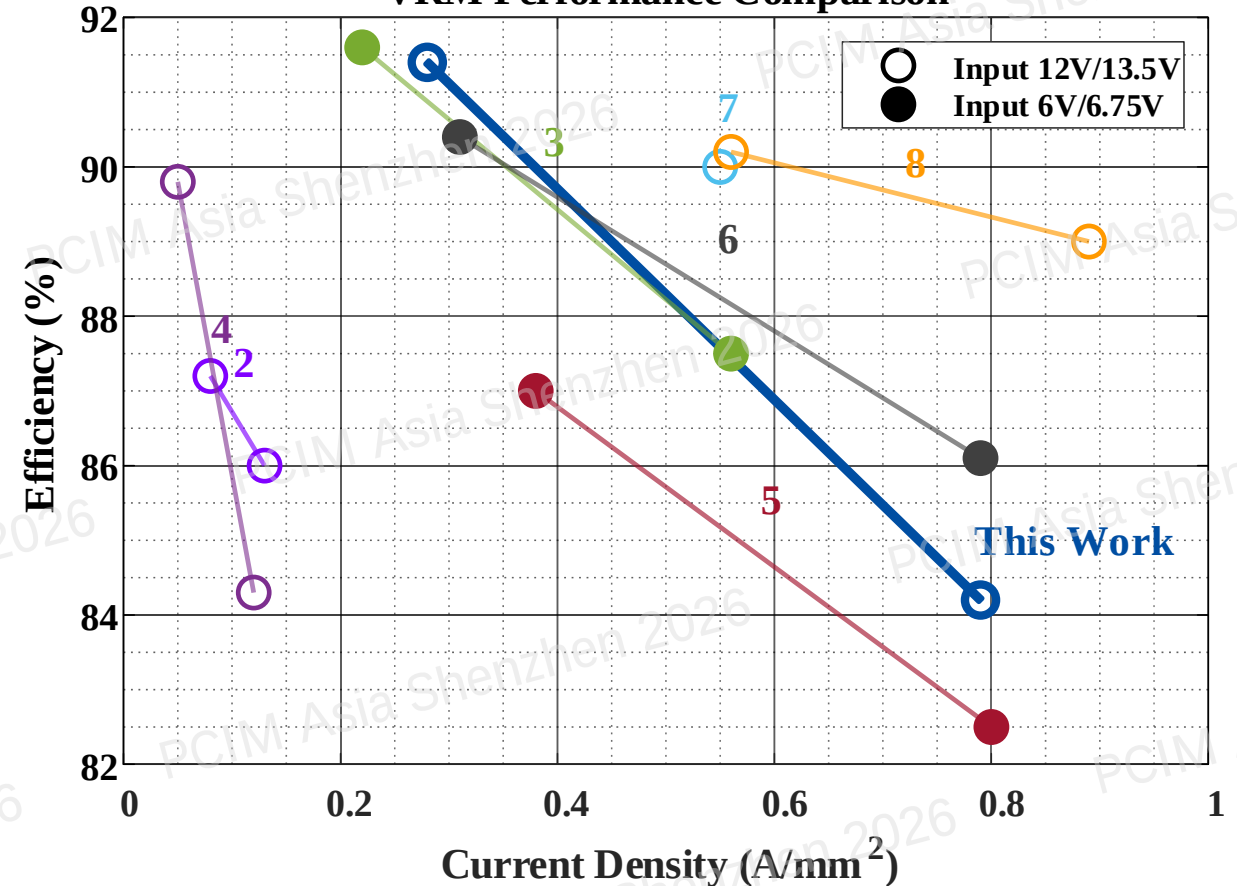
• Academia

2. S. Ji, D. Reusch, and F. C. Lee, "High-Frequency High Power Density 3-D Integrated Gallium-Nitride-Based Point of Load Module Design," IEEE Transactions on Power Electronics, 2013
3. Y. Yang, K. Zhang, M. Xu, J. Sun, J. Sun, and C. Wang, "Design for Ultra-High Power Density VRM Based on Six-Phase Coupled Inductors With Small Leakage Inductances," IEEE Transactions on Power Electronics, 2025
4. L. Yu, S. Zhao, X. Sun, W. Mu, X. Song, S. You, Y. Hao, and J. Zhang, "Symmetrical GaN-Based Vertical Module With Low Interference Integration and High Thermal Performance for PoL Converters," IEEE Transactions on Industrial Electronics, 2025

• Industry

5. S. Zou, J. Wu, S. Saini, Y. Hang, L. Lu, and J. Tang, "Vertical Power Delivery Module Design and Integration for 850A AI ASIC," in 2026 APEC
6. H. Gan, S. Jiang, S. Teng, S. Yamamoto, V. Chivukula, B. Edwards, C. Chung, J. Chen, M. Mohideen, G. Sizikov, and X. Li, "Vertical Power Delivery for 1000 Amps Machine Learning ASICs," in 2024 APEC
7. RRV29830 Renesas
8. TDM22545D Infineon

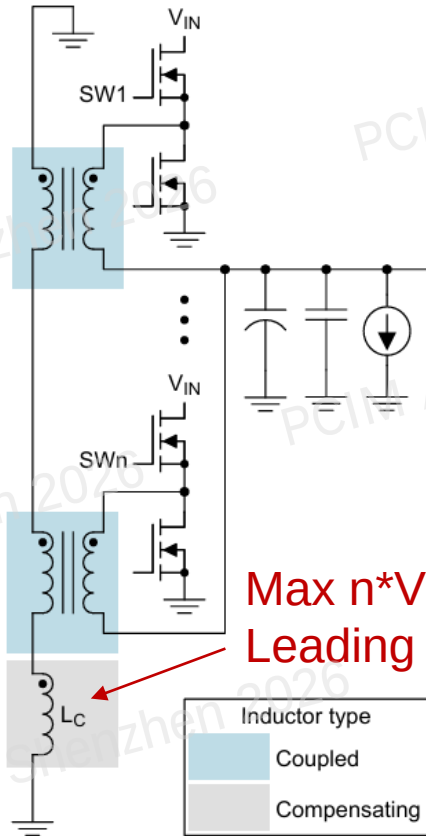
VRM Performance Comparison



- Use discrete device, achieve similar performance to industry module with integrated package
- Collaborating with device company for DrGaN

Transient Response: TLVR

- Topology



- Pros

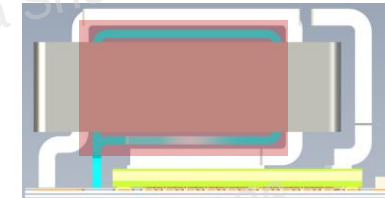
- Same steady state inductance while lower transient inductance
- Reduce C_{out} by ~40% [1][2]

- Cons

- Thicker Coil, thicker Inductor, bad for VPD



Conventional TLVR Inductor

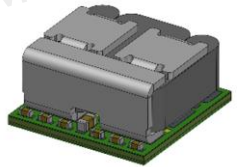


Customized TLVR Inductor

- Secondary coil space

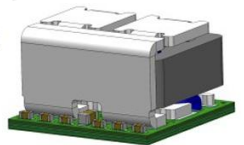
- Lower Efficiency, RRV29380
w/o tlvr $\eta=90\%$, w/ tlvr $\eta=87\%$ [3]

4mm



Standard Module

5mm



TLVR Module*

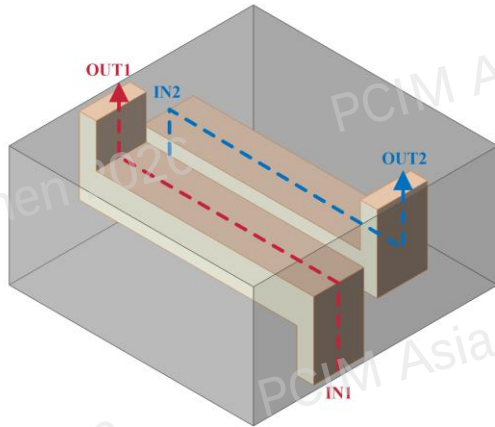
[1] Introduction to the Trans-Inductor Voltage Regulator (TLVR), TI, 2024

[2] Transformer-Based Voltage Regulators with Flexible TLVR Structure for Extremely Fast Dynamic Response, ADI, 2023

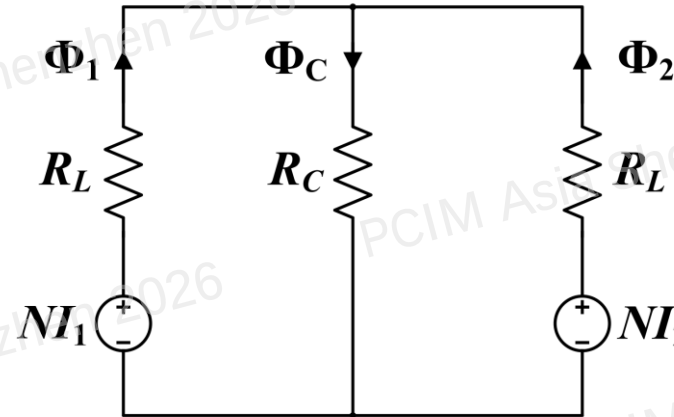
[3] FLEX, https://flex.com/products/power-modules/series/rrv-series?order=asc&order_by=part_number&order_type=primary&tableOnly=true&series=Power%20Modules&meta=series_ref:RRV

2-Phase Vertical Coupled Inductor

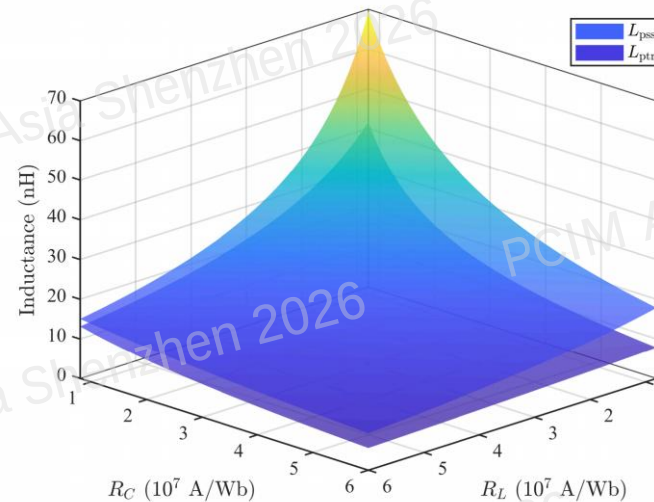
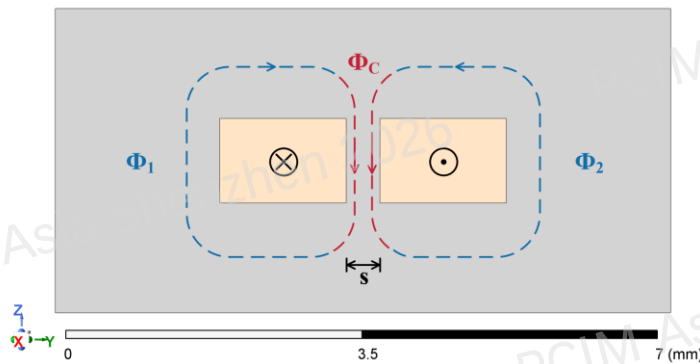
- Coupled Structure



- Reluctance Model

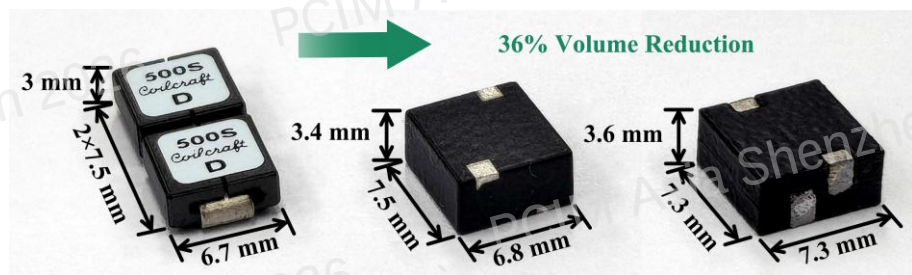


$$\begin{cases} L_{pss} = \frac{N^2}{R_L + \frac{1-2D}{1-D} R_C} \\ L_{ptr} = \frac{N^2}{R_L + 2R_C} \end{cases}$$



- Negative coupling by opposite current directions
- Reducing the winding spacing “s” increases R_C
- Transient inductance L_{ptr} can be reduced

Design Result

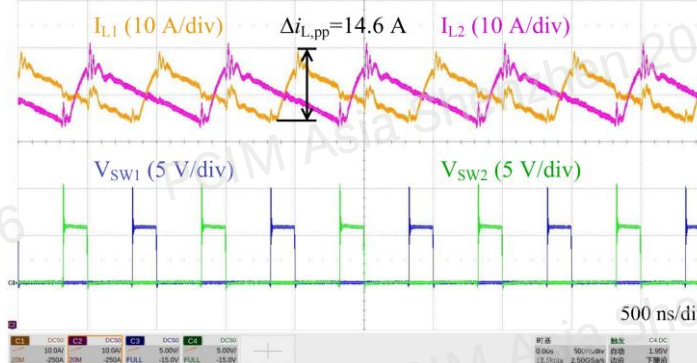


Parameter	Discrete	Uncoupled	Coupled
Material	ferrite	alloy powder	alloy powder
Total volume / mm ³	301.5	173.4	191.8
L_{ss} @ 0 A / nH	50	64.7	50.8
L_{tr} @ 0 A / nH	50	≈ 64.7	29.8
L_{ss} @ 60 A / nH	saturated	53.8	42.8
DCR / mΩ	0.12	0.17	0.11

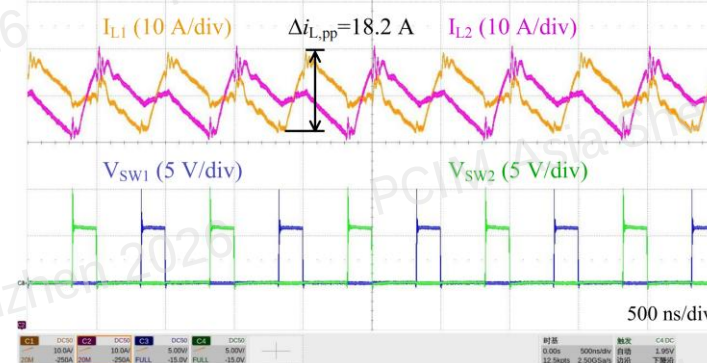
- 2-Phase Coupled: size reduced by 36%, DCR = 0.11 mΩ
- Negative coupling: high L_{ss} for steady state, low L_{tr} for transient

Steady-State Current Waveform

2-Phase Non-Coupled

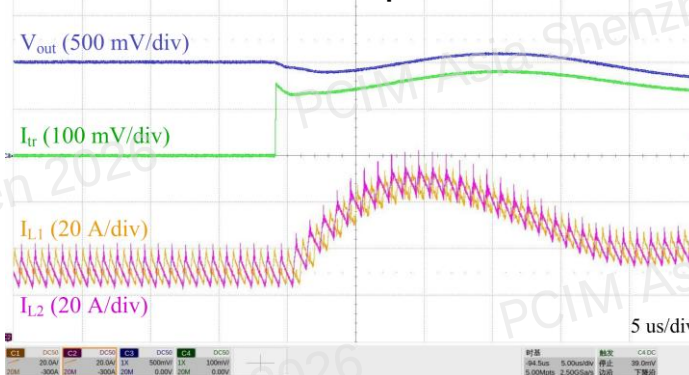


2-Phase Coupled

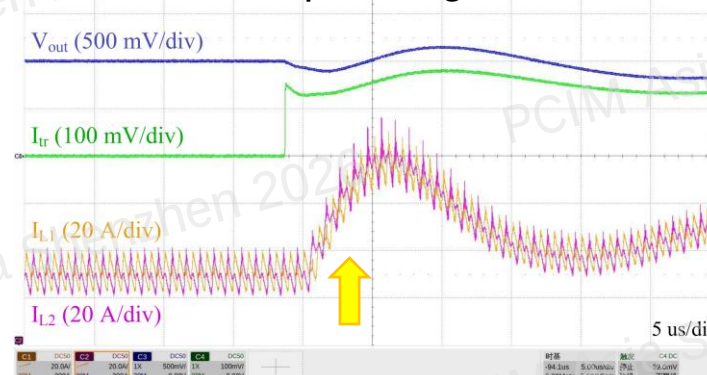


Load Transient @ 1415 A/μs

2-Phase Non-Coupled



2-Phase Coupled, higher di/dt



- **Height constrain requires simple topology and package integration for VDP: Buck, DrMOS, DrGaN in the future**
- **Alloy powder offers advantages over Ferrite for high dc current and high frequency VRM applications**
- **TLVR & Coupled Inductor needed for transient response**