A 280 MHz All-Digital Current-Mode IVR with Dual-Observation Current Telemetry and Kalman Filter-Based Fusion for Scalable Ganged Systems

Xuchen Men[1], Renwei Chen[1], Yunxin Wang[1], Boran Zhang[1], Zhoutong Liu[1], Hongru Liu[1],
Xuliang Wang[1], Zhaofu Zhang[2], Jingshu Yu[3], Jianshi Tang[1], Yan Wang[1], Xiaosen Liu[1]
[1]Tsinghua University, BNRist, Beijing, China, [2]Wuhan University, Wuhan, China, [3]Intel Corporation, Hillsboro, Oregon, USA

***Abstract*—This work presents a 280 MHz all-digital current-mode IVR with a dual-observation current telemetry framework for scalable ganged power delivery in 2.5D/3D systems. Conventional current sensing is fundamentally limited by the bandwidth of the analog front-end, while digital sensing achieves high-bandwidth yet suffers from severe switching noise and quantization errors. Model-based approaches provide low-noise and bandwidth-limited estimation, leading to a fundamental trade-off between bandwidth, noise, and accuracy. To overcome this limitation, a dual-observation current telemetry framework is proposed, combining noisy high-bandwidth Vx-based digital sensing with low-noise, bandwidth-limited model prediction. A Kalman filter-based computational engine (KFCE) optimally fuses these complementary observations to generate a minimum-variance current estimate, enabling robust and accurate current telemetry. Fabricated in 65 nm CMOS, the prototype operates up to 280 MHz, achieving >90% current estimation accuracy at 100 MHz and a peak power density of 7.04 A/mm². The proposed architecture enables high-bandwidth and noise-robust current telemetry, providing a scalable solution for high-frequency ganged IVRs in next-generation 2.5D/3D HPC systems.**

***Keywords*—Integrated voltage regulator (IVR), digital current sensors (DCS), model-based current prediction, Kalman filter (KF), ganging, 2.5D/3D-IC**

I. Introduction

Driven by the rapid growth of generative AI (GenAI) and large language models (LLMs), modern 2.5D/3D HPC processors are approaching a fundamental power wall at the kilowatt level. Emerging platforms such as SoW-X [1] enable wafer-scale heterogeneous integration with over 100 billion transistors, significantly expanding the power delivery network (PDN) into an ultra-large-scale distributed system (Fig. 1). In such systems, severe IR drop mismatch across tiles cannot be directly captured by voltage-mode control (VMC) [2], leading to substantial current imbalance and degraded efficiency in ganged operation. Current-mode control (CMC) has therefore been adopted to provide explicit current telemetry for cross-tile coordination and scalable ganging (Fig. 1) [3]. However, conventional implementations rely on analog current sensing, which is fundamentally limited by amplifier bandwidth and scaling constraints even in advanced FinFET technologies (Fig. 2) [4]. While digital current sensing removes this bandwidth bottleneck [5], it remains highly susceptible to switching noise, quantization error, and EMI, necessitating complex calibration and limiting scalability in large-scale PDNs.

This work proposes an all-digital dual-observation current telemetry framework to address this challenge. A Vx-based digital current sensor (Vx-DCS) provides high-bandwidth and noisy measurements, while a model-based current predictor generates low-noise and bandwidth-limited estimates. These limitations lead to a fundamental trade-off between bandwidth, noise, and accuracy, which cannot be resolved by standalone sensing or prediction. A Kalman filter-based computational engine (KFCE) is introduced to optimally fuse these complementary observations, producing a minimum-variance

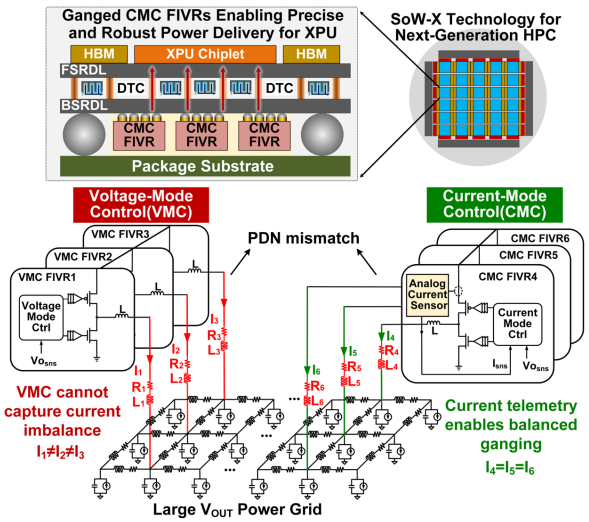

Fig. 1. Ganged CMC FIVRs for robust XPU power delivery, mitigating PDN mismatch and enabling balanced current sharing over VMC.

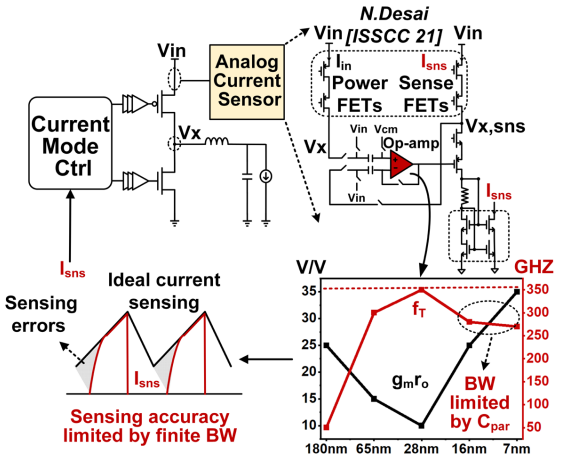

Fig. 2. Fundamental limitations of analog current sensing under technology scaling.

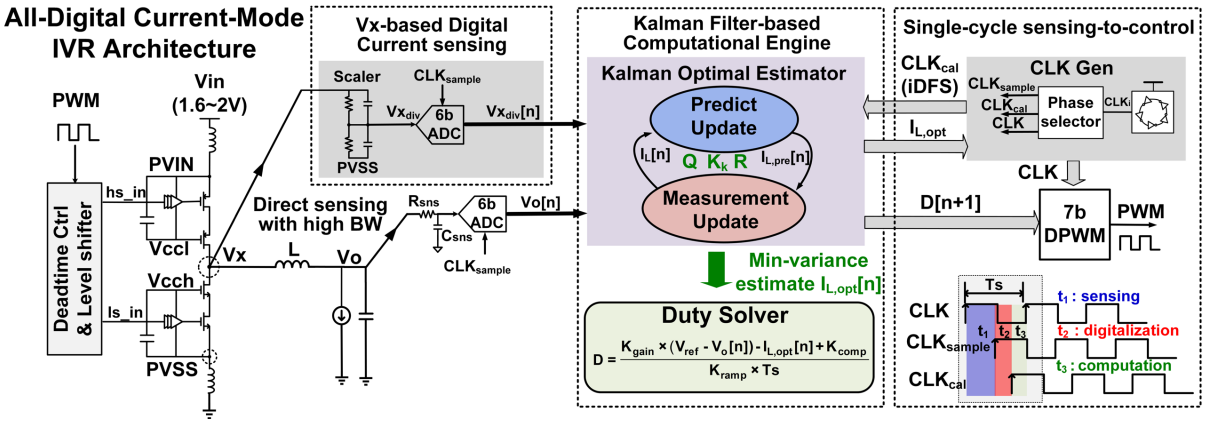

Fig. 3. Proposed all-digital current-mode IVR with Kalman filter-based current telemetry. The Vx signal is directly digitized for high-bandwidth sensing. A KFCE fuses dual observations to generate a minimum-variance current estimate $I_{L,opt}$, enabling high-bandwidth and noise-robust control.

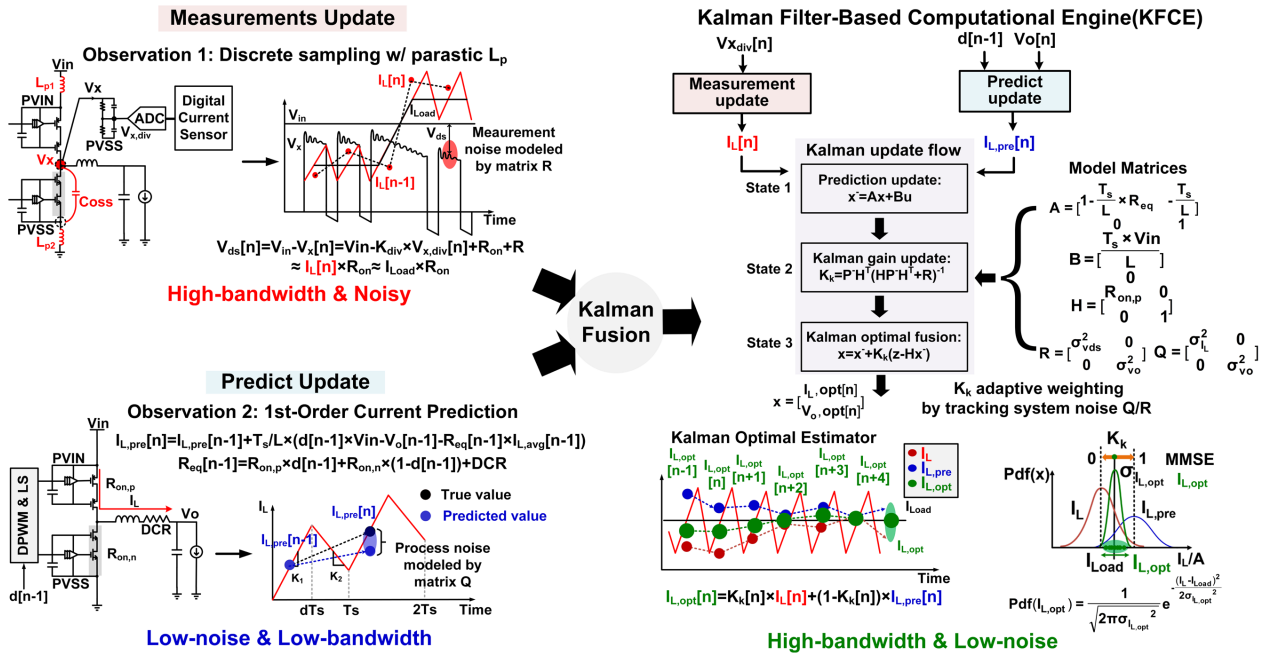

Fig. 4. Proposed dual-observation current telemetry framework with a Kalman filter-based computational engine (KFCE). Noisy high-bandwidth measurements from Vx sensing and low-noise, bandwidth-limited model predictions are optimally fused to obtain a high-bandwidth and low-noise current.

(MMSE) current estimate under both measurement noise and PVT variation. This enables accurate and robust current telemetry without sacrificing bandwidth, supporting scalable ganged operation in large-scale PDNs.

The proposed architecture is implemented in a 65 nm CMOS prototype operating up to 280MHz. Measurement results demonstrate >90% current telemetry accuracy at 100 MHz, a peak power density of 7.04 A/mm², and robust current sharing in ganged operation, validating the effectiveness of the proposed telemetry framework for next-generation 2.5D/3D HPC systems.

II. Dual-Observation Framework With Kalman Fusion

To address the fundamental trade-off between sensing bandwidth and noise, a dual-observation current telemetry framework is proposed, as shown in Fig. 3 and Fig. 4, combining switching-node-based measurement and model-based prediction within a unified estimation framework. Direct sensing from the switching node Vx provides intrinsically high-bandwidth information of the inductor current, yet is severely corrupted by switching noise, quantization error, and EMI. In contrast, model-based prediction leverages discrete-time power-stage dynamics to generate low-noise but bandwidth-limited current estimates, with accuracy constrained by model mismatch and parameter variation. These two observations exhibit complementary error characteristics in both bandwidth and noise domains.

As shown in Fig. 4, the inductor current is modeled by a discrete-time state-space representation characterized by system matrices (A, B, H) as a first-order state equation

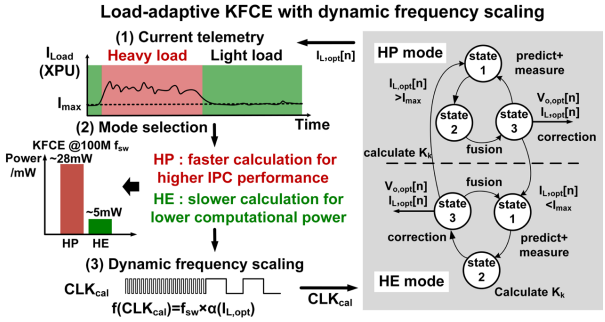

Fig. 5. KFCE with IPC-based DFS enabling load-adaptive computation. The estimated current $I_{L,opt}$ drives mode selection between HP and HE mode, dynamically balancing computational power and performance.

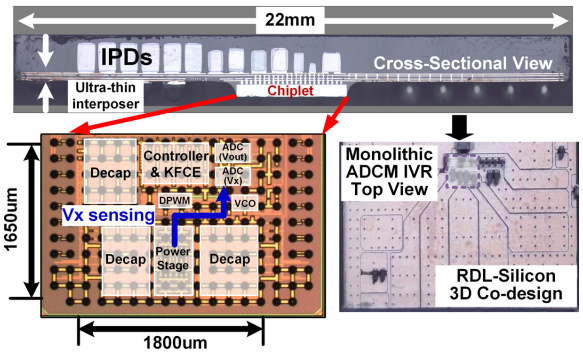

Fig. 6. Fabricated ADCM IVR chiplet and 2.5D/3D heterogeneously integrated IPDs with co-designed vertical power delivery.

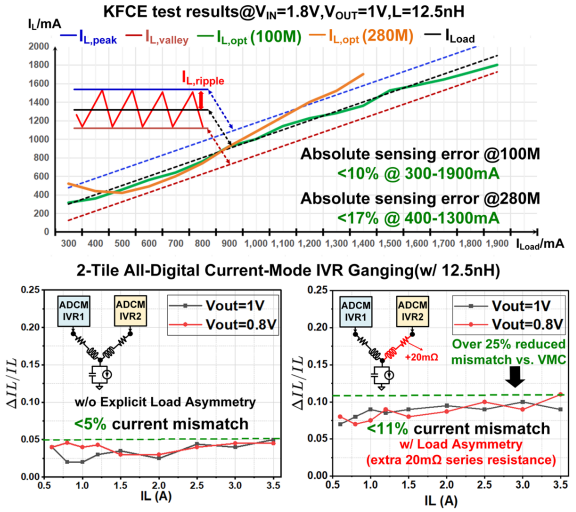

Fig. 7. Measured KFCE current-sensing results and current-sharing performance of the two-tile ganged ADCM IVR chiplet.

derived from Vx-DCS and model-based prediction, where process uncertainty—including model mismatch and load disturbance—is captured by process noise Q. Meanwhile, the Vx-DCS provides a direct but noisy observation of the switching-node behavior, modeled as an equation with measurement noise R dominated by switching noise and quantization effects.

Based on this formulation, a Kalman filter-based computational engine (KFCE) shown in Fig. 4 is employed to recursively estimate the discrete inductor current $I_{L,opt}$. The Kalman gain K_k is adaptively determined by the relative covariance of process and measurement noise (Q and R), enabling optimal weighting between the two observation paths under time-varying conditions. As a result, the estimated current $I_{L,opt}[n]$ corresponds to the minimum-variance estimate, simultaneously suppressing measurement noise while compensating for model bias. Unlike fixed-coefficient filtering approaches that impose a static trade-off between bandwidth and noise, the proposed Kalman-based fusion dynamically adapts to both operating conditions and uncertainty levels, achieving high-bandwidth and low-noise current estimation without sacrificing robustness. This formulation provides a unified framework for integrating sensing and modeling in digital current telemetry.

III. All-Digital Current-Mode IVR With IPC-Based Dynamic Frequency Scaling

Fig. 3 shows the architecture of the all-digital current-mode (ADCM) IVR integrating high-bandwidth Vx-based sensing, Kalman filter-based computational engine, and single-cycle sensing-to-control timing, where the proposed KFCE enables minimum-variance current estimation $I_{L,opt}[n]$ within each switching cycle. However, the effective estimation bandwidth and computational overhead of the KFCE are inherently determined by its update rate, making a fixed-frequency implementation fundamentally mismatched to time-varying load dynamics. Under heavy-load or fast transient conditions, insufficient update frequency degrades tracking capability, while under light-load conditions, excessive updates lead to unnecessary power consumption without significant accuracy improvement. To address this limitation, a load-adaptive computation framework is introduced, as illustrated in Fig. 5, where the $I_{L,opt}$ update rate is dynamically adjusted according to the instantaneous operating condition inferred from the estimated current $I_{L,opt}$. During heavy-load operation, a high-performance mode is activated, in which the $I_{L,opt}$ update loop runs at an elevated frequency to enhance response speed and reduce estimation latency, enabling accurate tracking of rapid current fluctuations. Conversely, under light-load conditions, the system transitions to a high-efficiency mode with reduced update rate, leveraging the intrinsic noise-suppression capability of the Kalman estimator to maintain accuracy while significantly lowering computational power. This adaptive operation reduces the KFCE power consumption at 100MHz from 28mW in HP mode to 5mW in HE mode, achieving more than 5× power reduction while preserving accurate current estimation. This adaptive behavior is realized through an IPC-based dynamic frequency scaling (iDFS) clock generation scheme that modulates the computation rate in alignment with load conditions, effectively implementing dynamic bandwidth scaling of the KFCE.

Furthermore, this mechanism is tightly integrated with the single-cycle sensing-to-control loop (Fig. 3), where sensing, digitalization, and computation are time-aligned within each switching period, ensuring that the estimated current $I_{L,opt}$ is immediately available for duty control. As a result, the

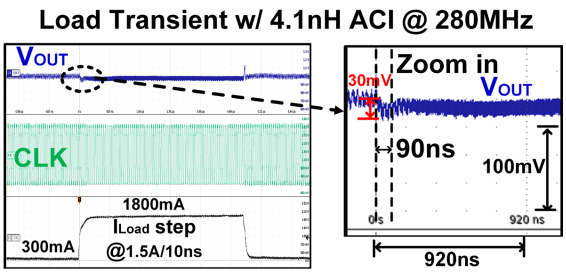

Fig. 8. Measured transient response at 280MHz using a 4.1nH air-core inductor.

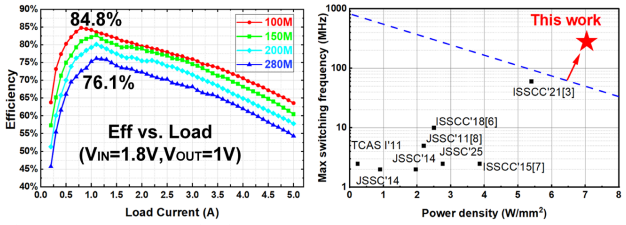

Fig. 9. Efficiency vs. load current at 100–280MHz (1V output) using a 12.5nH IPD inductor and comparison with prior CMC IVRs.

proposed approach achieves a balance between high-bandwidth transient tracking and energy-efficient operation, while preserving real-time control capability.

IV. Measurement Results

As shown in Fig. 6, the proposed all-digital current-mode IVR is fabricated in 65-nm CMOS with a 0.71 mm² active area. The prototype supports F_{sw} up to 280MHz with package-integrated inductors. Measured KFCE current sensing results under different F_{sw} are shown in Fig. 7, the proposed scheme achieves <10% absolute sensing error from 0.3-1.9 A at 100 MHz and achieving <17% error from 0.4-1.3 A at 280MHz. These results validate that the proposed KFCE effectively suppresses switching noise while preserving high-bandwidth current tracking.

In two-chip ganged operation, the current mismatch remains below 5% under symmetric loading and within 11% under a 20 mΩ load mismatch, while conventional voltage-mode control is expected to exhibit significantly larger imbalance. This corresponds to over 25% reduction in current mismatch compared to VMC-based ganging, confirming that the proposed telemetry effectively compensates PDN-induced asymmetry and enables robust current sharing. As shown in Fig. 8, under a 1.5 A/10 ns load step at 280MHz with a 4.1 nH air-core inductor, the output voltage droop is limited to 30 mV with a recovery time of 90 ns, demonstrating fast transient response enabled by high F_{sw} and high-bandwidth current telemetry.

The IVR delivers up to 5 A load current as shown in Fig. 9, achieving a peak efficiency of 84.8% at 100 MHz and 76.1% at 280MHz, corresponding to a peak power density of 7.04 A/mm². These results validate that the fully digital architecture sustains high-frequency operation without sacrificing efficiency or power density.

TABLE I. Comparison with the State-of-the-Art

	JSSC'11 [8]	ISSCC'15 [7]	ISSCC'21 [3]	ISSCC'18 [6]	This work
Technology	0.35μm	0.35μm	22nm	65nm	**65nm**
Current Sensor	Analog	Analog	Analog	Analog	**Digital**
Regulation	Peak & Valley Current-Mode	Hysteretic & Quasi Current-Mode	Peak Current-Mode	Time-Domain Current-Mode	**Digital KF Current-Mode**
V_{IN} (V)	2.7-4.2	2.7-4.5	1.6-2	1.8	**1.6-2**
V_{OUT} (V)	0.5-2.6	2	0.7-1.2	0.15-1.69	**0.8-1.2**
F_{SW} (MHz)	5	1	30-60	10	**100-280**
L (nH)	1000	4700	2-5.5	220	**4.1-28**
I_{max} (A)	0.6	1	1.2	0.6	**5**
Peak Efficiency	91%@5MHz	95.5%@1MHz	89.1%@30M	92%@10M	**84.8%@100M 76.1%@280M**
Power Density (W/mm²)	2.2	3.86	5.4	2.5	**7.04**
Current Sensing Accuracy	NA	NA	81%@100M	NA	**>90%@100M >83%@280M**
Current Sharing Accuracy	NA	NA	1.2%@60M	NA	**<5%@100M**
Recovery Time	6us	4.8us	20us	3.5us	**90ns**

V. Conclusions

This work presents a 280MHz all-digital current-mode IVR based on a dual-observation current telemetry framework. By combining Vx-based digital sensing with model-based prediction through Kalman filter-based fusion, the proposed architecture overcomes the bandwidth limitations of conventional current sensing approaches. The prototype demonstrates >90% current estimation accuracy, 7.04 A/mm² peak power density, and robust current sharing in ganged operation. These results establish a scalable all-digital current telemetry and control paradigm, enabling high-frequency and large-scale power delivery for next-generation 2.5D/3D HPC.

Acknowledgement

This work was supported by the National Key Research and Development Program of China 2024YFF1400200 and Grant BNR2026RC01013. *Corresponding Author: Xiaosen Liu.*

References

[1] P. -C. Shih et al., "SoW-X: A Novel System-on-Wafer Technology for Next Generation AI Server Application," *IEEE 75th Electronic Components and Technology Conference (ECTC)*, Dallas, TX, USA, 2025, pp. 1-6.

[2] S. Kim et al., "14.9 A Monolithic 10.5W/mm² 600 MHz Top-Metal and C4 Planar Spiral Inductor-Based Integrated Buck Voltage Regulator on 16nm-Class CMOS," *ISSCC*, pp. 270-272, 2024.

[3] N. Desai et al., "17.4 Peak-Current-Controlled Ganged Integrated High-Frequency Buck Voltage Regulators in 22nm CMOS for Robust Cross-Tile Current Sharing," *ISSCC*, pp. 262-264, 2021.

[4] A. L. S. Loke *et al.*, "Analog/Mixed-Signal Design Challenges in 7-nm CMOS and Beyond," *IEEE Custom Integrated Circuits Conference (CICC)*, Austin, TX, USA, 2019, pp. 1-8

[5] M. P. Chan and P. K. T. Mok, "A Monolithic Digital Ripple-Based Adaptive-Off-Time DC-DC Converter With a Digital Inductor Current Sensor," *JSSC*, vol. 49, no. 8, pp. 1837-1847, Aug. 2014.

[6] J. -G. Kanget al., "A 10MHz time-domain-controlled current-mode buck converter with 8.5% to 93% switching duty cycle," *ISSCC*, 2018, pp. 424-426.

[7] S. -H. Lee et al., "12.1 A 0.518mm2 quasi-current-mode hysteretic buck DC-DC converter with 3μs load transient response in 0.35μm BCDMOS," *ISSCC*, 2015, pp. 1-3, 2015.

[8] M. Du et al., "A 5-MHz 91% Peak-Power-Efficiency Buck Regulator With Auto-Selectable Peak- and Valley-Current Control," *JSSC*, vol. 46, no. 8, pp. 1928-1939, Aug. 2011.