Performance Based Tuning of an Inductive Integrated Voltage Regulator Driving a Digital Core against Process and Passive Variations

Venakata Chaitanya Krishna Chekuri, Monodeep Kar, Arvind Singh, Saibal Mukhopadhyay
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, USA
[vchekuri3, mkar3, rathorearvind19]@gatech.edu, saibal@ece.gatech.edu

***Abstract*— This paper presents an auto-tuning method for fully integrated voltage regulators (IVRs) driving digital cores against variations in passive as well as process/temperature of the core. The key contribution is to perform auto-tuning of the coefficients of the feedback loop of the IVR based on the performance of the digital cores. Simulations using a high-frequency IVR Simulink model and digital logic in 45nm CMOS process shows that the proposed performance driven auto-tuning demonstrates potential for up to 12% increase in system performance under inductance and threshold variation.**

Keywords— IVR; Auto tuning; Delay based tuning; Dynamic voltage and frequency scaling (DVFS)

I. INTRODUCTION

Integration of inductive voltage regulators with on-chip/on-package inductors on the same chip as the digital logic cores helps fast recovery from voltage droops (load transient) and a fast transition of the output voltage (reference transient) to support dynamic voltage and frequency scaling (DVFS) [1-3]. In deep nanometer process: aging, temperature and process variations affects performance of the digital circuits [4-5]. As the IVRs are designed in the same process nodes as digital circuits, they also suffer from same variations [2, 6, 8]. In particular, variations in on-chip/on-package passives (inductance and capacitance) causes shift in the IVR's characteristics including transient response to load step and reference step [2, 12]. The variations in the IVR's output response creates additional uncertainty for the digital circuits potentially increasing the error rates. Therefore, it is important to develop auto-tuning mechanisms for IVR to minimize the effect of IVR's variations on the voltage/timing margin or timing error rates of digital cores. The existing post-silicon auto-tuning methods for the IVRs [7-8] involve adjusting the controller transfer function to optimize a tuning cost. However, they are not aware of the process variation in the digital logic.

This paper presents an auto-tuning method for IVR driven by the performance of the digital cores (Fig. 1). We propose to tune the IVR using a cost function that directly captures performance of the digital core with the objective to increase the maximum operating frequency of the digital circuit under

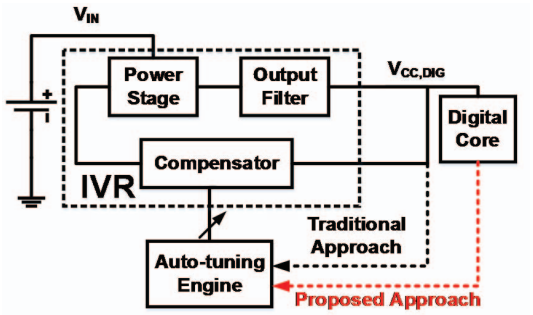

Fig. 1. A system of an inductive IVR and a digital core along with the proposed auto-tuning method.

(a) process variations in cores and passives, and (b) supply noise due to dynamic load transitions, while ensuring stable IVR operation. The proposed tuning method is based on two cost metrics to represent the system performance: (1) the accumulated sum of the delay slack of a digital core with respect to a target frequency and (2) error count during operation with a target operating frequency. To evaluate the tuned systems, we consider a fully integrated inductive IVR (FIVR) with a voltage mode digital PWM control driving a digital core. We design the circuits for on-chip realization of the cost metrics and tuning algorithms.

Considering process variations in both IVR and digital cores, we demonstrate, using simulation in 45nm CMOS, that the proposed tuning not only ensures stable transient response (similar to conventional tuning) but also increases achievable frequency for the digital processor for a target error rate under variation. Using the proposed performance based costs, tuning a digital system with +20% threshold voltage (V_T) and ±20% variation in the inductance value, the proposed tuning method achieves up to 33.98MHz or 12.18% improvement over the base frequency at a fixed error rate (explained in section 4).

II. BACKGROUND AND MOTIVATION

A. Post-silicon Tuning of Digital Circuits

A voltage margin is added to the intrinsic minimum operating voltage of a digital circuit to meet the target frequency under process and dynamic (supply, noise, temperature) variations. To reduce such margin, techniques

This work is supported by Semiconductor Research Corporation (SRC) through Texas Analog Center of Excellence at University of Texas at Dallas (Task ID: 2712.002)

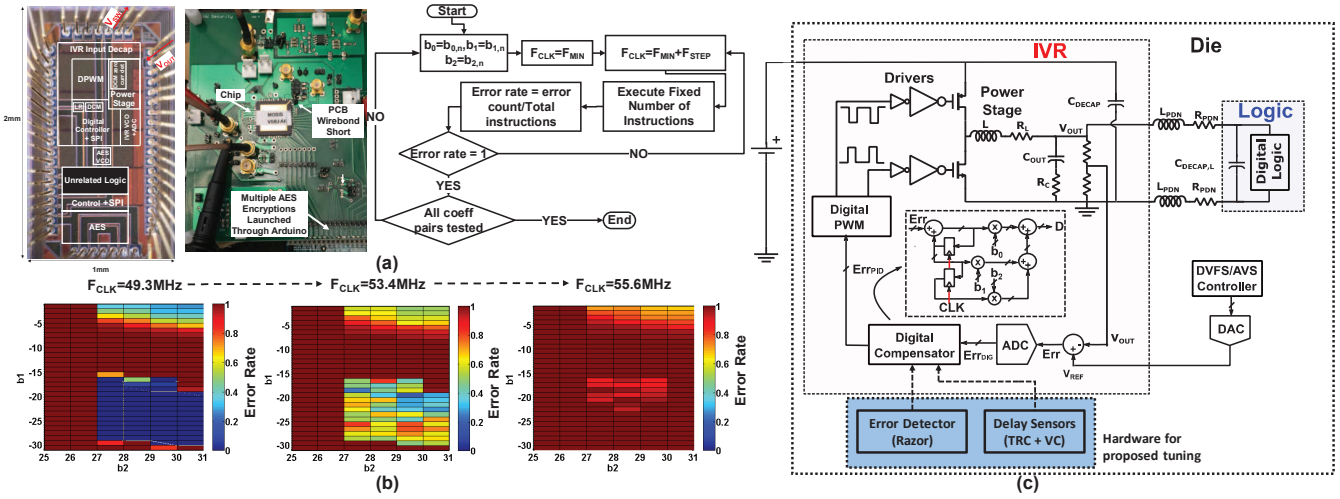

Fig. 2. (a) Chip micrograph, test PCB, and measurement procedure for the 130nm test-chip (b) Measured error rate across different compensator coefficients (b_1 and b_2) for increasing clock frequencies ($V_{CC,AES}$=0.8V) (c) Circuit Diagram of a Fully Integrated Inductive Voltage Regulator with hardware for proposed tuning

such as post-silicon tuning of supply voltage, adaptive circuits with dynamically controlled supply and clock, and error tolerant designs like Razor [10-11] that uses on-chip error-detection/correction to recover from runtime timing errors have been developed. To enable post-silicon and dynamic tuning, critical path replica circuits have been developed to track delay variations of logic circuits [4].

B. *Fully Integrated Voltage Regulator (FIVR)*

An inductive IVR uses power stage composed of a PMOS and NMOS device and an on-die/on-package inductor and an on-chip output capacitor as shown in Fig. 2c (without the proposed hardware for tuning). A high switching frequency (>100MHz) is required to manage ripple with small L/C. A voltage mode PWM control is typically used as controller for these FIVRs. Due to the ease of integration into the advanced process nodes as well as high bandwidth due to high operating clock frequency, digital PID compensators are preferred [2, 12]. A type-III compensator with two zeros is required for compensating the filter double pole as the zero created by the ESR of the output capacitance resides at a high frequency. To improve the loop bandwidth which dictates the response speed from transient, phase shifting of sampling clocks as well as reduced precision multi-sampling [2] have been used.

C. *Impact of IVR Variation on Digital Circuits*

For digital cores which are supplied by an off-chip voltage regulator module (VRM), the supply quality is determined by the local decoupling capacitance as well as impedance of the power distribution network (PDN). For IVRs, as the impedances of the off-chip PDN and decoupling capacitance are eliminated, the transient supply noise is dictated mostly by the control loop. Consequently, as shown by Kar et. al., the performance of digital system is significantly more sensitive to IVR's variations (L & C), compared to off-chip VRM variation [6] . Therefore, tuning the IVR's control loop to mitigate the effect of passive variation, is crucial to enhance the performance of digital circuits under variation.

D. *Prior Work on Auto-tuning Algorithms*

Auto-tuning process for any VRMs (including IVRs) observes the output behavior of the VRM after perturbing the control loop and adjusts the controller transfer function based on a cost. Existing tuning algorithms for off-chip VRMs perform complex frequency domain measurements. For example Shirazi et. al. in [8] controls the loop phase-margin and unity-gain-frequency (UGF) to achieve a target value. As the frequency domain tuning requires complex computations, they are less suitable for high switching frequency IVRs [12]. Time domain based tuning algorithm tunes the controller by performing simple arithmetic computations on the time domain samples of the IVR output [2, 7]. For example, Kar et. Al. in [2] uses a cost metric which is a summation of aggregated absolute error values, aggregated signed error values and settling time of a load transient to tune the coefficients.

E. *Motivation: Test-chip Measurement*

We perform measurements on an 130nm test-chip with IVR powering an AES (Advanced Encryption Standard) encryption engine to experimentally characterize role of FIVR controller's coefficients on run-time timing error. The FIVR power stage uses two consecutive bondwires with a total of 11.6nH inductance, 3.2nF MIM capacitance and 125MHz switching frequency. The direct form digital controller is sampled at 250MHz frequency and the compensator coefficients are reduced to 6bits. A 128-b AES engine is driven by the IVR and is used as a digital load to the IVR. For a given clock frequency, multiple AES encryption events are executed. Depending on the supply noise and the target clock frequency, there can be timing violations in the AES causing an incorrect encryption. The AES outputs are compared with the golden responses to find out the error rate and this experiment is repeated for different IVR coefficients and increasing clock frequency (F_{CLK}). For a F_{CLK} of 49.5MHz (0.8V FIVR output) multiple FIVR coefficients yield a zero error rate (Fig. 2). As the frequency is increased, the error rate starts to increase for the aforementioned coefficients and becomes dependent on the FIVR coefficients. The measurement shows that timing error rate for a target frequency is dependent on IVR coefficients.

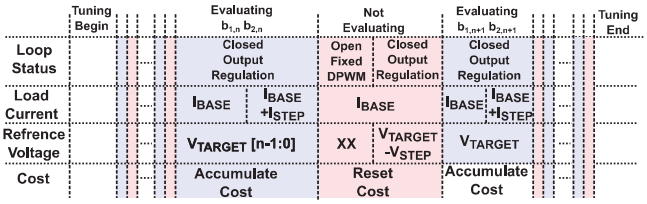

Fig. 3. IVR control flow during the evaluation phase of each coefficient

III. AUTO-TUNING METHODOLOGY

A. Proposed Tuning Methodology

We propose two tuning costs to quantify the performance of a digital system. We show that using these cost metrics enable (1) obtaining a stable response at DC loads and fast recovery from transient droop, and (2) optimize performance of the system under process variation of the digital core. The tuning engine generates different coefficients for different operating conditions of the digital load.

1) Control Flow of Tuning Process: During the tuning process, the system performance is measured for different compensator coefficients. The control flow of the tuning algorithm as well as each evaluation period is elaborated in Fig. 3. Before each coefficient is evaluated, the control loop is opened (power stage driven by a fixed duty cycle) to ensure same initial condition. Unlike [2], before starting evaluation, the control loop is closed with a reference voltage lower than the target voltage. The difference between reference and the target voltage determine the reference step (V_{STEP}). After the loop is closed, the output stability at a base load current (I_{BASE}) is observed and at the middle of the evaluation cycle, a load step (I_{STEP}) is applied. The second half of the evaluation period observes the stability at current I_{STEP}+I_{BASE}. During actual runtime, the total numbers of transient droop events depend on the underlying application. However, as every coefficient goes through the same evaluation period, fairness is ensured during the optimization. The response of the control loop to the transient events during the evaluation period is mapped to the system performance and is captured in the quantified performance. We used two different costs for quantizing the system performance and are discussed in the following section.

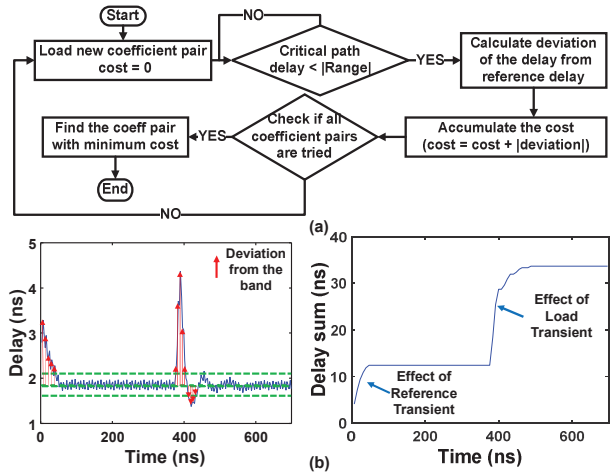

Fig. 4. (a) Control flow of the delay-sum based cost (b) An example delay response for 0.7V V_{CC} at nominal V_T corner and the cost profile for the corresponding response

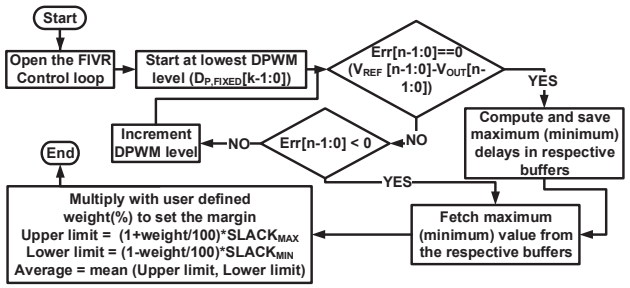

Fig. 5. The proposed open loop test for characterizing the variation in critical path delay for steady state IVR variation

2) Delay-sum Based Tuning: Instead of accumulating the absolute IVR output error as part of tuning cost in [2], we propose to accumulate the absolute delay slack between the critical path delay of the combinational logic and the target clock period. Using a delay-slack based cost captures the same effect as the IVR output errors and ensures rejection of unstable coefficients, but also helps fine-tuning the response for process variation in the core. The outcome of minimizing the proposed delay-slack based cost selects a set of coefficients where across the evaluation period, the logic delay stays closest to the target delay. Fig. 4 shows the control flow for delay based tuning. The effect of the initial reference transient and load transient as well as the effect of IVR steady state output voltage ripple is captured in the delay of the logic. Although transient responses can be tuned by changing the compensator coefficients, the steady state ripple at the IVR output does not get affected by coefficient tuning. We propose

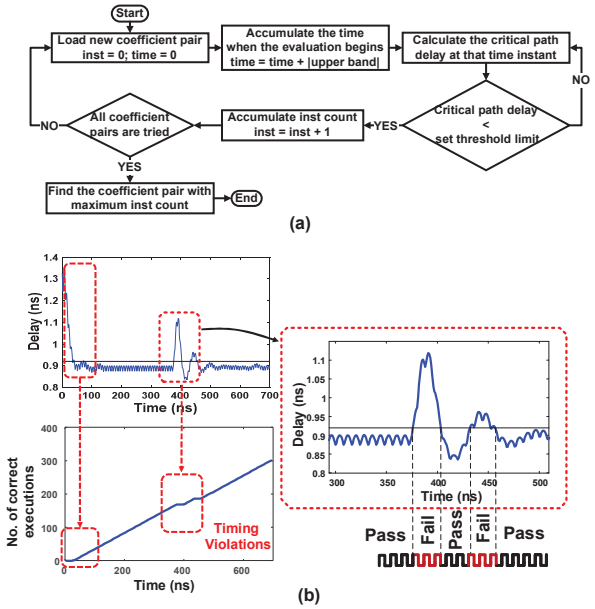

Fig. 6. (a) Control flow of the error-count based cost (b) An example delay response for 1V V_{CC} at nominal V_T corner and corresponding cost calculation (number of correct instructions executed against time is shown)

to use a band around the delay slack and any digitized slack value within that band is neglected during accumulation to eliminate the effect of supply ripple. The band is determined using the test shown in Fig. 5. To determine the bands, the IVR control loop is opened, DPWM is driven by a fixed input and the digitized error is observed (Fig. 5). DPWM resolution is generally set higher than ADC resolution to avoid limit cycling, so for multiple $D_{P,FIXED}$ values Err_{DIG} will be zero. When the control loop is closed and the output is regulated, the $D_{P,FIXED}$ can settle to any of these values at a steady state condition. The minimum and the maximum digitized slack is calculated for these $D_{P,FIXED}$ levels, which represent the variation in the logic delay at a steady state of the IVR and cannot be tuned by changing coefficients. To account for this effect, the maximum and the minimum delay values should be multiplied with a factor to account for small fluctuations at the output voltage. The reference clock frequency can be chosen from the mean of all the delay samples collected during the open loop test. To integrate proposed cost into existing hardware we propose to use a tunable replica circuit (TRC) followed by a Vernier delay chain (VC), which acts as a time-to-digital (TDC) converter, to quantify the critical path delay [10]. The absolute value of the digitized delay is aggregated over the evaluation period of one coefficient and minimized across different sets of coefficients to obtain the optimum coefficients. Depending on the critical path obtained during synthesis, selectable fixed length portions of the cells are chosen to mimic the critical path [4]. A series of inverters are appended at the end to fine tune the TRC for delay tracking.

3) Error-count Based Tuning: For error tolerant systems such as Razor, the voltage margin is aggressively reduced to a point of first failure (PoFF) which allows a higher frequency of operation under a given supply voltage. However, total numbers of instructions correctly executed is dependent of the number of timing error detected by the error-detecting-latch [10-11]. Once a timing error is detected, the instruction is replayed for multiple cycles till no further error is detected, leading to performance (throughput) loss. Hence, reducing he error rate, by reducing the supply noise variations is crucial to improve effective throughput. Fig. 6 illustrates the concept of error count based tuning of IVR's coefficients. If the voltage-margin is aggressively set, each time a large load transient occurs, there will be timing failure till the system recovers from the droop. The first droop (Fig. 6) depends on the value of output capacitance and the ESR of the output capacitance, is mostly insensitive to the values of the compensator

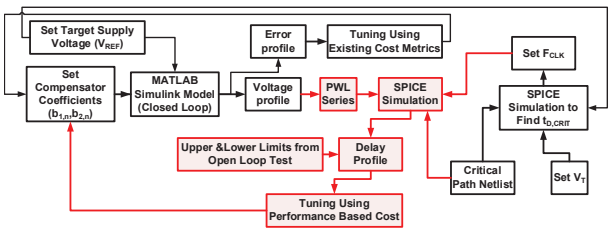

Fig. 7. Simulation infrastructure

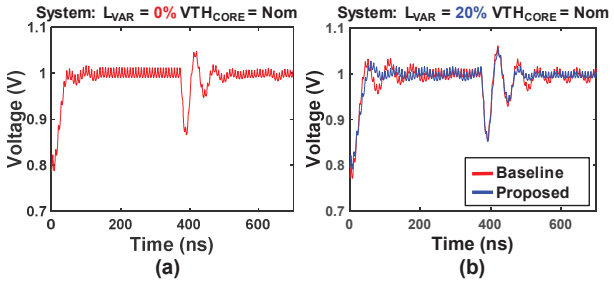

Fig. 8. Voltage responses of IVR against passive variation, before and after tuning using existing cost (a) At no L variation, (b) At 20% L variation

coefficients and hence the tuning process. However, the performance is also dependent on the droop settling time and the second droop (Fig. 6) which can be tuned using IVR's control loop. Note, the delay based metric penalizes the coefficients if the logic delay is both lower and higher than the target delay, whereas the error count based metric penalizes the coefficients only if the logic delay is more than the target delay. Error-count based cost can be easily incorporated in IVRs powering digital engines with an error-detection circuit. For example, the cost can be computed by accumulating the number of error events detected by a Razor latch [11] over the evaluation period. For the evaluation purpose, we set the target frequency as the upper threshold of the band found during the open loop test. This ensures that no errors are detected during the steady state operation of the IVR.

IV. Simulation Results

Fig. 7 shows the simulation setup for the analysis. A time-domain model of the IVR in MATLAB Simulink is used performing transient simulations. We use an IVR with 1.2V input, 6nH inductance, 50mΩ ESR, 10nF capacitance, 125MHz switching frequency with 250MHz sampling frequency. Each coefficient is represented using a 7-bit signed integer. An 8-bit ADC digitizes the difference between the reference and the output voltage. The compensator output is fed to a DPWM with 10-bit resolution (7.8ps resolution). The ADC and the compensator operate at 250MHz clock frequency whereas the DPWM operates at 125MHz. Each coefficient is evaluated for 700ns i.e. 88 IVR switching cycles. We performed experiments at two output levels: 1V and 0.7V with V_{STEP} as 0.15V and 0.1V respectively. An I_{BASE} of 10mA

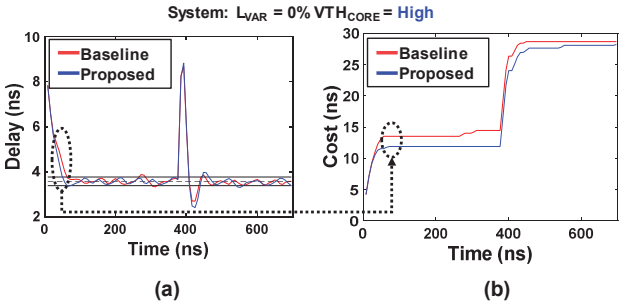

Fig. 9. Example tuning on a system under only process variation in the digital logic (no L variation) using delay-sum based cost (a) at high V_T and (b) corresponding costs against time

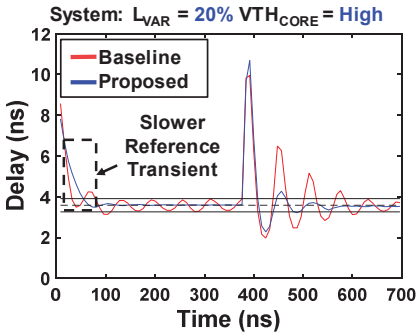

Fig. 10. Improvement in the delay profile by using the delay-sum based cost with both process variation and passive variation at high V_T and high L

and an I_{STEP} of 100mA are chosen. We selected an exhaustive range for the direct form coefficients. The critical path of the digital logic is emulated as an open chain of 100 standard cell inverters in 45nm CMOS technology and simulated using SPICE. To demonstrate the advantage of the proposed tuning methodology, we use ±20% variation in the V_T of the digital core and filter inductance (L) of the IVR at constant V_{CC}.

A. IVR Tuning Using Existing Algorithm

The IVR is first tuned against variation in passives. The accumulated absolute value of the IVR error samples (V_{REF} - V_{OUT}, Fig 2) during the evaluation period, is used as cost. The compensator coefficient pair obtained for the design with no passive and process variations using the existing auto-tuning algorithm [2], is considered as baseline coefficient pair ($C_{IVR}L_N$). Fig 8 shows the response of the baseline FIVR with $C_{IVR}L_N$ and the response for a FIVR with +20% variation in the inductance value using the same coefficient. After tuning, an updated coefficient is obtained ($C_{IVR}L_H$) and the response with +20% L variation improves both in terms of DC load stability as well as transient response.

B. IVR Tuning Using Proposed Algorithm

1) Delay-sum based cost: Fig. 9 illustrates tuning a baseline FIVR using delay-sum based metric can reduce effect of process variation in the digital core. First, we tune the system at nominal V_T at 0.7V supply voltage (optimum coefficients $C_{SYS}L_NVT_N$). Next, we consider the digital core has moved to a high V_T corner. The delay-sum based tuning results in a new coefficient ($C_{SYS}L_NVT_H$). Note as L is not

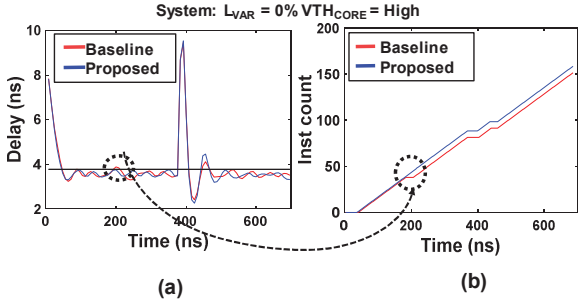

Fig. 11. Example tuning on a system under only process variation in the digital logic (no L variation) using the error-count based cost (a) the delay profile before and after tuning and (b) the corresponding cost against time before and after tuning

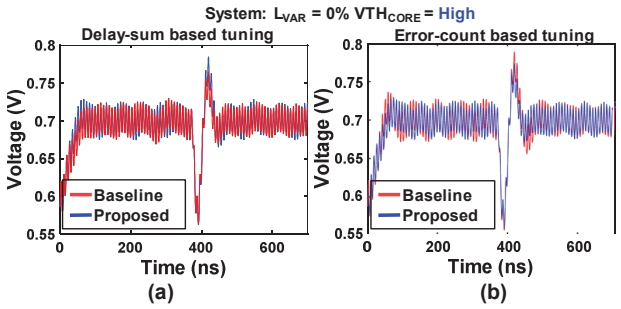

Fig. 12. Voltage responses of IVR against process variation show improvement when using (a) Delay-sum based cost, (b) Error-count based cost

varying, the IVR-only tuning would have resulted in original coefficents. Fig. 8 shows that for the high-VT core, re-tuning the IVR with $C_{SYS}L_NVT_H$ provides smaller delay variation during the reference transient (compared to the original coefficients $C_{SYS}L_NVT_N$), and causes the steady state delay variation to stay within the delay bands. The results show that delay-based tuning of IVR helps improve performance of digital core. To understand the effect of the process and passive variations we perform tuning on three individual systems, one with L only variation, one with V_T only variation and last with both. Fig. 10 illustrates delay profiles for systems with high VT and high L. Each system is tuned to a different coefficient to reduce the delay variation especially under second droop. Note the tuned coefficients in Fig. 10 reduces second droop but causes slower reference transient.

2) Error count based cost: The same experiments were performed using an error count based cost function as illustrated in Fig. 11. Due to the higher voltage-delay sensitivity of the high-V_T system, the coefficient $C_{SYS}L_NVT_N$ causes delay violations (at ~200ns) while tuned coefficients ($C_{SYS}L_NVT_H$) eliminates the violation and reduces error-count based cost. Note, an error based cost is sensitive only to the duration of the delay violation, not the exact value of the delay slack. Hence, this tuning may result in more than one "optimum coefficient".

C. Impact of Performance-based IVR Tuning

In this section, we evaluate the impact of proposed IVR tuning on the system performance by estimating the error rate of the digital core at different target frequencies with tuned IVR coefficients. We apply different amounts of variations in

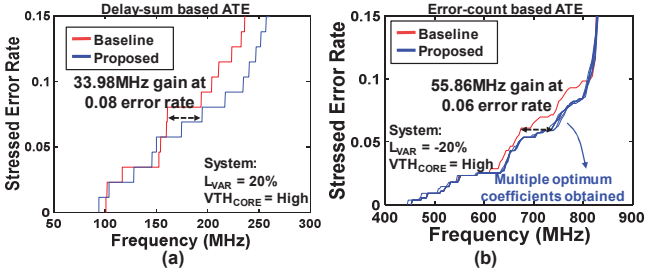

Fig. 13. Maximum frequency gain for a given stressed (SER) rate achieved using (a) delay-based sum cost (b) error count based cost.

TABLE I. IMPROVEMENT IN FREQUENCY FOR GIVEN ERROR RATE WITH PROPOSED METRICS

Delay-Sum Based Tuning								Error-Count Based Tuning							
L (%)	VTH	V_{DD} = 1V			V_{DD} = 0.7V			L (%)	VTH	V_{DD} = 1V			V_{DD} = 0.7V		
		Error Rate	Freq Gain (%)	Gain (MHz)	Error Rate	Freq Gain (%)	Gain (MHz)			Error Rate	Freq Gain (%)	Gain (MHz)	Error Rate	Freq Gain (%)	Gain (MHz)
0	Low	0	0	0	0.064	1.96	16.87	0	Low	0.093	0.75	10.58	0.056	2.71	23.332
0	Nom	0	0	0	0.1	2.31	12.63	0	Nom	0	0	0	0.068	**3.73**	20.23
0	High	0.094	1.063	9.083	0.1	**9.45**	26.36	0	High	0	0	0	0.022	2.29	6.42
20	Low	0.006	2.102	29.67	0.064	**4.71**	40.37	20	Low	0.026	1.53	21.66	0.002	3.09	26.46
20	Nom	0.012	2.457	27.57	0.1	**7.21**	39.33	20	Nom	0.028	1.66	18.61	0.081	**4.37**	23.81
20	High	0.007	**3.069**	26.09	0.08	**12.18**	33.98	20	High	0.076	2.03	17.23	0.011	2.78	7.76
-20	Low	0.088	1.42	2.01	0.051	2.23	18.99	-20	Low	0.061	**3.85**	54.29	0.056	**3.58**	30.57
-20	Nom	0.025	0.82	9.24	0.061	**3.81**	20.52	-20	Nom	0.062	**5.48**	61.6	0.064	**6.30**	33.90
-20	High	0.061	1.32	11.25	0.045	2.53	6.91	-20	High	0.06	**6.57**	55.86	0.1	2.71	7.39

the L, and VTH and perform both of the proposed tuning for each system to obtain the optimal coefficients. Next, we use the optimal coefficients for the IVR, perform transient simulation of the system by applying load transient during evaluation, and estimate the error rate of digital core versus frequency. As during evaluation, the system is tested with higher than normal transient events within a time window, we call the error rate as the stressed error rate (SER). Fig. 13 shows SER versus frequency for a nominal L and high VT system using coefficients for IVR-only ($C_{SYS}L_NVT_N$) and proposed tuning ($C_{SYS}L_NV_{TH}$). We observe that delay-sum based and error-count based tuning improves frequency by 33.98MHz (@SER of 0.08) and 55.86MHz (@SER=0.06) respectively. Table I shows the percentage improvement in frequency for SER of 0.1. We observe that improvement in frequency for a given error rate is higher for higher VT at lower supply voltage. This is attributed due to higher sensitivity of the delay to supply voltage noise which brings out the advantage of performance based tuning.

V. CONCLUSION

This paper presents a performance based auto-tuning algorithm to tune a system of an IVR driving a digital core. We demonstrate that performance-based IVR tuning ensures a stable response with fast recovery from transient events under variations in the passives. More importantly, in the proposed approach, by tuning the IVR coefficients we can enhance the digital system performance considering process variation in the digital core and in the passives, which is beyond the capability of the existing IVR tuning methods. In conclusion, we show that the tuning of any IVR should be performed using quantifiable performance of the entire system instead of only using the performance of the IVR.

REFERENCES

[1] N. Sturcken et al., "A Switched-Inductor Integrated Voltage Regulator With Nonlinear Feedback and Network-on-Chip Load in 45 nm SOI," in IEEE JSSC, vol. 47, no. 8, pp. 1935-1945, Aug. 2012.

[2] M. Kar et al., "An integrated inductive VR with a 250MHz all-digital multisampled compensator and on-chip auto-tuning of coefficients in 130nm CMOS," ESSCIRC, Lausanne, 2016, pp. 453-456.

[3] E. A. Burton et al., "FIVR — Fully integrated voltage regulators on 4th generation Intel® Core™ SoCs," 2014 IEEE APEC, Fort Worth, TX, 2014, pp. 432-439.

[4] M. Cho et al., "Postsilicon Voltage Guard-Band Reduction in a 22 nm Graphics Execution Core Using Adaptive Voltage Scaling and Dynamic Power Gating," in IEEE JSSC, vol. 52, no. 1, pp. 50-63, Jan. 2017.

[5] K. Bowman et al., "Circuit techniques for dynamic variation tolerance," 2009 46th ACM/IEEE DAC, San Francisco, CA, 2009, pp. 4-7.

[6] M. Kar et al., "Impact of process variation in inductive Integrated Voltage Regulator on delay and power of digital circuits," 2014 IEEE/ACM ISLPED, La Jolla, CA, 2014, pp. 227-232.

[7] J. A. Abu Qahouq et al., "Online Closed-Loop Autotuning Digital Controller for Switching Power Converters," in IEEE Trans. Ind. Electron, vol. 60, no. 5, pp. 1747-1758, May 2013.

[8] M. Shirazi et al., "An Autotuning Digital Controller for DC–DC Power Converters Based on Online Frequency-Response Measurement," in IEEE Trans. Power Electron., vol. 24, no. 11, pp. 2578-2588, Nov. 2009.

[9] C. Huang and P. K. T. Mok, "An 84.7% Efficiency 100-MHz Package Bondwire-Based Fully Integrated Buck Converter With Precise DCM Operation and Enhanced Light-Load Efficiency," in IEEE JSSC, vol. 48, no. 11, pp. 2595-2607, Nov. 2013.

[10] S. Das et al., "RazorII: In Situ Error Detection and Correction for PVT and SER Tolerance," in IEEE JSSC, vol. 44, no. 1, pp. 32-48, Jan. 2009.

[11] D. Ernst et al., "Razor: a low-power pipeline based on circuit-level timing speculation," Proceedings. 36th Annual IEEE/ACM International Symposium on Microarchitecture, 2003. MICRO-36., 2003, pp. 7-18.

[12] M. Kar et al., "An All-Digital Fully Integrated Inductive Buck Regulator With A 250-MHz Multi-Sampled Compensator and a Lightweight Auto-Tuner in 130-nm CMOS," in IEEE JSSC, vol. 52, no. 7, pp. 1825-1835, July 2017.