A Switched-Inductor Integrated Voltage Regulator With Nonlinear Feedback and Network-on-Chip Load in 45 nm SOI

Noah Sturcken, *Student Member, IEEE*, Michele Petracca, Steven Warren, Paolo Mantovani, Luca P. Carloni, *Senior Member, IEEE*, Angel V. Peterchev, *Member, IEEE*, and Kenneth L. Shepard, *Fellow, IEEE*

***Abstract*—A four-phase integrated buck converter in 45 nm silicon-on-insulator (SOI) technology is presented. The controller uses unlatched pulse-width modulation (PWM) with nonlinear gain to provide both stable small-signal dynamics and fast response (∼700 ps) to large input and output transients. This fast control approach reduces the required output capacitance by 5× in comparison to a conventional, latched PWM controller at a similar operating point. The converter switches package-integrated air-core inductors at 80 MHz and delivers 1 A/mm^2 at 83% efficiency and 0.66 conversion ratio. A network-on-chip (NoC) serves as a realistic digital load along with a programmable current source capable of generating load current steps with slew rate of ∼1 A/100 ps for characterization of the control scheme.**

***Index Terms*—DC-DC power conversion, integrated voltage regulator (IVR), voltage regulator (VR), switched inductor, Buck converter, switching-converter, nonlinear feedback, hysteretic control, transient response.**

I. Introduction

A DOMINANT limitation on computational performance in modern microprocessors and systems-on-chip is power consumption. Battery life, energy costs, and maximum operating temperature all impose a power envelope on digital ICs that commonly necessitates throttling computational performance. Consequently, performance-per-watt has become an increasingly important metric. Dynamic voltage and frequency scaling (DVFS) is a technique that has enabled improved performance-per-watt by reducing supply voltages during periods of low computational demand [1], but implementations stand to improve dramatically by reducing the time scales over which the supply voltage is positioned, allowing real-time optimization of power consumption in the presence of workload variability. For the case of chip multiprocessors and heterogeneous systems-on-chip (SoCs), it is natural to divide computational logic into individual voltage-frequency domains, allowing per-core or per-functional-block DVFS [2], [3]. Generally, a DVFS implementation with faster voltage transition times and smaller voltage-frequency domains delivers a more energy-efficient implementation. However, current methods for power supply regulation with board-level voltage regulator modules (VRMs) require tens of microseconds to transition voltages and are too bulky to deliver many independent power supplies in a cost effective manner [4].

External VRMs present two other efficiency challenges. First, I^2R losses in the power distribution network (PDN) are significant when highly scaled voltages are delivered from the board. In a typical PDN (Fig. 1) [5], a resistance from the VRM to the CPU's package of 0.7 mΩ dissipates 7 W of power for 100 W load at 1 V. Second, VRMs require power supply margins that degrade energy efficiency. The high-frequency impedance of the PDN limits the VRM's ability to suppress voltage overshoot in the event of load current transients; consequently, modern VRM specifications stipulate that the supply voltage follow a load-line commonly given as $v_{\mathrm{OUT}} = V_{\mathrm{ZL}} - R_{\mathrm{LL}} i_O$, where v_{OUT} is the processor supply voltage, V_{ZL} is the desired v_{OUT} at zero load, R_{LL} is the desired load-line resistance, and i_O is the load current. Implementation of load-line control reduces the VRM size and cost required to maintain the output voltage within the allowed tolerance during load transients. However, when the system is not operating at maximum power consumption, the load-line is a source of inefficiency as v_{OUT} will be greater than the minimum supply voltage, $V_{\mathrm{MIN}} = V_{\mathrm{ZL}} - R_{\mathrm{LL}} I_{O,\mathrm{MAX}}$, where $I_{O,\mathrm{MAX}}$ is the maximum load current. The wasted power will be $P_{\mathrm{LL}} = i_O R_{\mathrm{LL}} (I_{O,\mathrm{MAX}} - i_O)$. For a typical value for R_{LL} of 1 mΩ [5], a CPU with $I_{O,\mathrm{MAX}}$ of 100 A operating at 50 A and 1 V will waste 2.5 W in the load-line implementation. If the PDN impedance were smaller, the value of R_{LL} and hence the load-line inefficiency could be reduced.

Recent work has explored switch-mode integrated voltage regulators (IVRs) as a means to address these shortcomings in VRMs. In this case, energy is stored on or close to the integrated circuit in capacitors (switched-capacitor converters) or inductors (buck converters). Integrated switched-capacitor converters, taking advantage of high-density integrated capacitors,

Manuscript received November 13, 2011; revised February 17, 2012; accepted March 27, 2012. Date of publication June 04, 2012; date of current version July 19, 2012. This paper was approved by Guest Editor Alvin Loke. This work was supported in part by the Department of Energy under contract DE-EE0002892, by the National Science Foundation under grant ECCS-0903466, and by the C2S2 Center, part of the Focus Center Research Program.

N. Sturcken, S. Warren, and K. L. Shepard are with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA (e-mail: nsturcken@ee.columbia.edu, sbw986@ee.columbia.edu, shepard@ee.columbia.edu).

M. Petracca, P. Mantovani, and L. P. Carloni are with the Department of Computer Science, Columbia University, New York, NY 10027 USA (e-mail: petracca@cs.columbia.edu, paolo@cs.columbia.edu, luca@cs.columbia.edu).

A. V. Peterchev is with the Departments of Psychiatry and Behavioral Sciences, Biomedical Engineering, and Electrical and Computer Engineering, Duke University, Durham, NC 27708 USA (e-mail: angel.peterchev@duke.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSSC.2012.2196316

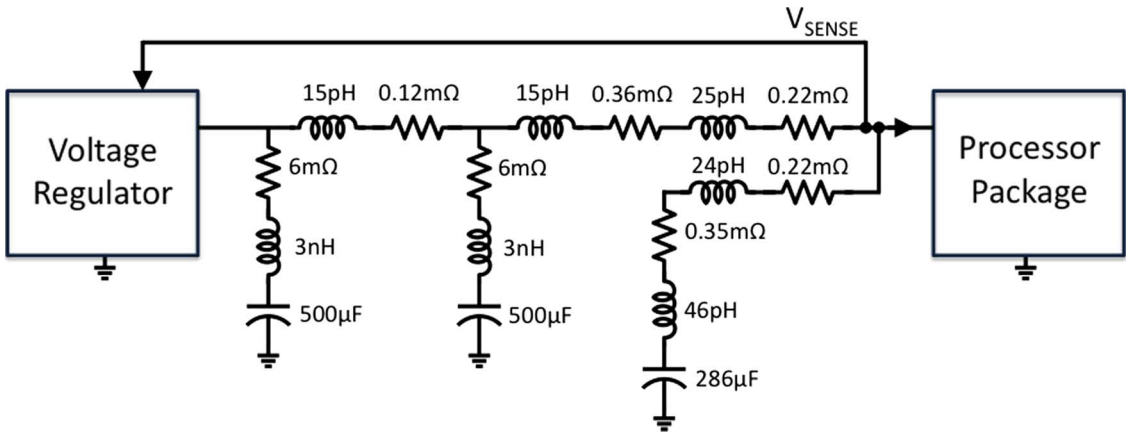

Fig. 1. Power Distribution Network for a modern high-performance microprocessor, from VRM to CPU package [5].

have shown high efficiency at reasonable current densities but have done so only at fixed conversion ratio and without addressing transient requirements [6]–[8]. Meanwhile, integrated buck converters have shown high current densities and efficiencies with a continuous range of conversion ratios but face challenges concerning the integration of high-quality inductors [9]–[16].

Until recently, integrated inductors that offered both low losses and high inductance density were unavailable. Planar spiral or other inductor topologies that can be constructed using the interconnects of a typical CMOS process are too resistive to provide efficient on-chip power conversion at reasonable current densities [16]. The efficient use of surface mount technology (SMT) air-core inductors, which can provide a current density up to $\sim$1.7 A/mm^2 [17], has been successfully demonstrated [9]–[13]. However, the size and discrete nature of these devices hinders the scalability of any IVR incorporating discrete SMT inductors. Fortunately, advances have recently been made in the development of integrated magnetic-core power inductors that are highly scalable and capable of delivering current densities as high as 8 A/mm^2 [18]–[21]. These inductors have been included in IVR prototypes by on-chip integration [14] and chip stacking [15], demonstrating the eventual feasibility of highly scalable and efficient switched-inductor IVRs.

Another challenge in the development of switched-inductor IVRs is the integration of decoupling capacitance. While VRMs are able to augment voltage regulation at high frequencies by leveraging large amounts of inexpensive board-level decoupling capacitance, the integrated capacitance required in IVRs comes at much greater expense. In switched-inductor IVRs the dominant constraint on decoupling capacitance is set by the need to suppress voltage overshoot during fast load current transients. Extending the IVR controller bandwidth has the effect of reducing these decoupling capacitance requirements.

Some early switched-inductor IVRs address transient response by employing a multi-phase hysteretic controller to provide nearly instantaneous response to transients, effectively reducing the required output decoupling capacitance [9], [10]. Unfortunately, the closed loop behavior of the multi-phase hysteretic controller is difficult to predict and the loose synchronization of phases produces an under-damped large-signal response. Also, hysteretic controllers do not operate at fixed switching frequency, and can therefore pose challenges when attempting to control EMI. Subsequent work has used more conventional, pulse-width modulation (PWM) controllers and has relied on abundant package-level decoupling capacitance to compensate for increased controller delay [11], [12]. However, the dependence on package-level capacitance increases component and packaging cost and degrades scalability.

In contrast, the interleaved four-phase buck converter presented here, fabricated and tested in 45-nm SOI, employs an unlatched PWM modulator and nonlinear feedback to concurrently provide PWM-like synchronization of multiple phases, linear small-signal dynamics (ensuring stability and load-line regulation), and nearly instantaneous response to large-signal input-voltage and load-current transients without the need for large output decoupling. SMT inductors are employed for this initial implementation but the approaches used here can extend to integrated magnetic-core power inductors. The converter powers a realistic on-chip load composed of four parallel 64-node networks-on-chip (NoCs) along with a programmable current source capable of generating large load current steps for characterization of the controller. In Section II, we discuss the impact of controller bandwidth on the required output capacitance in IVRs, motivating our controller design. Section III describes the design and operation of the proposed control scheme, providing analysis for predicting the controller response. Section IV details the construction and operation of the integrated NoC, and Section V presents experimental results from the IC prototype.

II. Constraints on Output Capacitance

A. Output Voltage Ripple

Candidate capacitor technologies for an IVR include low-inductance discrete ceramic capacitors, on-chip MOS capacitors, and on-chip deep-trench (DT) capacitors, each offering reduced effective series resistance (ESR) and effective series inductance (ESL) relative to the capacitors typically used with VRMs. The high-frequency impedance of low-inductance discrete (LID) capacitors such as land-grid-array or interdigitated capacitors is dominated by ESL with self-resonant frequencies (SRF) around 30 MHz, where $\mathrm{SRF} = 1/2\pi\sqrt{\mathrm{ESL} * C_{\mathrm{OUT}}}$ [22]. In contrast,

the distributed nature of on-chip MOS and DT capacitance results in negligible ESL with a high-frequency impedance dominated by ESR with time constants, $\tau_C = \mathrm{ESR} * C$, around 1 ps for MOS capacitors and 500 ps for DT capacitors, depending on resistance of the on-chip PDN.

With wide impedance variability of candidate IVR capacitor technologies, it is important to use a general model in determining the output voltage ripple and other design parameters that are dependent on the high-frequency output impedance. The total peak-to-peak inductor current ripple is

$$\Delta I_{L,p-p} = \frac{V_{\mathrm{IN}} T_{\mathrm{sw}} D^*(1 - ND^*)}{L_\phi} \tag{1}$$

where V_{IN} is the buck converter input supply voltage, T_{sw} is the switching period, N is the number of phases in a multi-phase converter, $D^* = \mathrm{mod}(D, 1/N)$, and L_ϕ is the filter inductance of each phase [23]. The expression for output voltage ripple, including the effects of ESL, is given by:

$$\Delta V_{\mathrm{OUT},p-p} \approx \frac{\Delta I_{L,p-p}}{C_{\mathrm{OUT}}} \sqrt{\left(\frac{T_{\mathrm{sw}}}{8N} - \frac{N}{8T_{\mathrm{sw}}\mathrm{SRF}^2}\right)^2 + \tau_C^2} \tag{2}$$

using a simple lumped RLC model for the output capacitor.

B. Load-Line Implementation

The low ESR of ceramic capacitors typically requires the output voltage to follow a dynamic load-line [23]

$$v_{\mathrm{OUT}} \rightarrow V_{\mathrm{ZL}} - R_{\mathrm{LL}} I_O \tag{3}$$

where the output impedance is defined as

$$Z_{\mathrm{LL}} = R_{\mathrm{LL}} \frac{1 + s\tau_C}{1 + sR_{\mathrm{LL}} C_{\mathrm{OUT}}}. \tag{4}$$

This remains the case for IVRs that use on-chip MOS or DT capacitance. Typically, the dynamic load-line is implemented by having the controller regulate the output impedance of the converter to R_{LL} until the unity-gain frequency, f_C, at which point C_{OUT}must dominate the output impedance, constraining the output capacitance to

$$C_{\mathrm{OUT}} \geq \frac{1}{2\pi f_C R_{\mathrm{LL}}}. \tag{5}$$

It is desirable to achieve the highest possible f_C in order to reduce the requirement on C_{OUT}. However, a well-accepted guideline for maximum loop-gain bandwidth that avoids instability in closed-loop operation is

$$f_C \leq \alpha N f_{\mathrm{sw}} \tag{6}$$

where f_{sw} is the switching frequency and α is a constant commonly chosen as $<1/6$ [24]. Switching losses can become appreciable at high frequencies, effectively constraining f_{sw}; nevertheless, it has been shown that IVRs can operate efficiently with f_{sw} around 100 MHz [9]–[15]. Combining (5) and (6) produces the constraint on C_{OUT} for load-line regulation with on-chip MOS or DT capacitance

$$C_{\mathrm{OUT}} \geq \frac{1}{2\pi\alpha N f_{\mathrm{sw}} R_{\mathrm{LL}}}. \tag{7}$$

True load-line regulation is not easily achieved with low-inductance discrete capacitors when f_C exceeds the capacitor SRF, which is generally the case for IVRs. ESL will dominate at frequencies above f_C resulting in an appreciable first droop at the onset of a large di_O/dt event [11]. However, dynamic load-line regulation is possible if series resistance is added to the low-inductance discrete capacitors. The discrete capacitance, $C_{\mathrm{OUT,LID}}$, in this case, is accompanied by an added series resistance, $\mathrm{ESR}_{\mathrm{ADD,LID}}$, and an additional on-chip capacitance, $C_{\mathrm{OUT},O-C}$. These values can be chosen according to

$$R_{\mathrm{LL}} = \mathrm{ESR}_{\mathrm{ADD,LID}} + \frac{\tau_{C,\mathrm{LID}}}{C_{\mathrm{OUT,LID}}} \text{ and} \tag{8}$$

$$\frac{1}{C_{\mathrm{OUT,LID}} \left(2\pi \mathrm{SRF}_{\mathrm{LID}}\right)^2 R_{\mathrm{LL}} C_{\mathrm{OUT},O-C}} \leq R_{\mathrm{LL}}. \tag{9}$$

This option will not yield a reduction in the total capacitance; however, it may facilitate a balance between on-chip and off-chip decoupling capacitance that is cost-effective.

C. Load Current Transient Response

While the load-line constraint on output capacitance results in the desired small-signal output impedance, the duty cycle, and hence the controller response, may saturate in the event of a large load-current step, ΔI_O. In this case, the saturated response of the controller is unable to prevent the output voltage from overshooting the load-line; therefore, the output capacitor must provide additional support. The minimum capacitance (critical capacitance) that limits voltage overshoot to ΔV_{OS} during worst-case load-current transients is

$$C_{\mathrm{CRIT}} = \left(\frac{t_L}{2} + \frac{\tau_C^2}{2t_L} + t_d - \tau_I\right) \Bigg/ \left(R_{\mathrm{LL}} + \frac{\Delta V_{\mathrm{OS}}}{\Delta I_O}\right) \quad \text{for } L_\phi > L_{\mathrm{CRIT}}, \text{ and}$$

$$C_{\mathrm{CRIT}} = \left(\tau_C + t_d - \tau_I\right) \Bigg/ \left(R_{\mathrm{LL}} + \frac{\Delta V_{\mathrm{OS}}}{\Delta I_O}\right) \quad \text{for } L_\phi \leq L_{\mathrm{CRIT}} \tag{10}$$

where $L_{\mathrm{CRIT}} = N\tau_C V_L^*/\Delta I_O$, $t_L = L_\phi \Delta I_O / V_L^* N$, $V_L^* = \min(V_{\mathrm{IN}} - v_{\mathrm{OUT}}, v_{\mathrm{OUT}})$ τ_{I} is the load step time constant and t_d is the delay time for the controller to saturate the duty cycle [23]. This expression is applicable to IVRs using on-chip decoupling capacitance, where typically $L_\phi > L_{\mathrm{CRIT}}$. For the case of IVRs using low-inductance discrete capacitors with values selected in accordance with (8) and (9), the critical capacitance can still be determined from (10) if $\tau_{C,\mathrm{LID,TOT}} = \mathrm{ESR}_{\mathrm{ADD,LID}} * C_{\mathrm{OUT,LID}} + \tau_{C,\mathrm{LID}}$ is used for τ_{C}.

D. Minimum Output Capacitance

The constraints on minimum capacitance as a function of f_{sw} for an IVR that meets the specifications from Table I with conventional voltage mode feedback are shown in Fig. 2. The dominant constraint on minimum output capacitance using either LID or on-chip DT or MOS capacitance for this IVR is load-current transient response. For the case of an IVR with conventional linear feedback, the value of t_d can be approximated as

$$t_{d,\mathrm{lfb}} = \sqrt{\frac{V_{\mathrm{in}} C_{\mathrm{OUT}}}{f_C \Delta I_O}}. \tag{11}$$

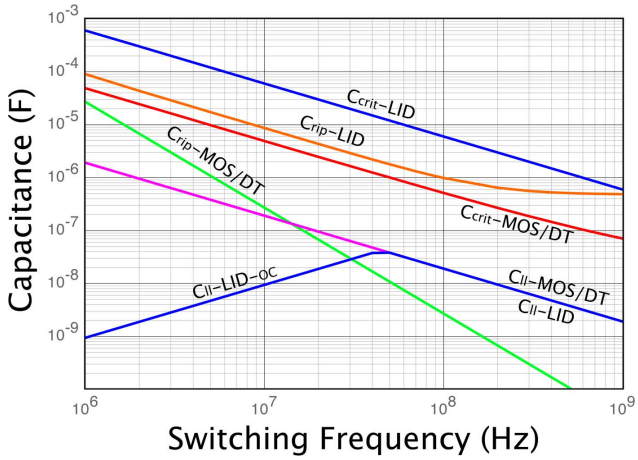

Fig. 2. The minimum capacitance meeting constraints for an IVR with operating parameters defined in Table I. Voltage ripple (C_{rip}), load-line regulation (C_{ll}) and the saturating transient response (C_{crit}) are plotted versus converter switching frequency, f_{sw}, for low-inductance discrete capacitance (LID) and on-chip MOS or DT capacitance.

TABLE I
PROPOSED IVR SPECIFICATIONS

V_{IN}	input voltage	1.8V
$I_{O,MAX}$	max. load current	1.2A
ΔI_O	max. dynamic load step	600mA
τ_I	load step time constant	100ps
ΔV_{OS}	Max. transient overshoot	40mV
R_{OUT}	closed loop output resistance	125mΩ
f_{SW}	switching frequency	80MHz
N	number of phases	4
L	inductance per phase	26nH

For the example used in Fig. 2 with on-chip DT capacitance, t_d dominates the numerator with a value of 154 ns, relative to 6.5 ns and 19 ps for the terms $t_L/2$ and $\tau_C^2/2t_L$ respectively. This result indicates that controller delay is the primary bottleneck in reduction of $C_{\rm OUT}$ for IVRs with conventional feedback controllers. Therefore, control techniques that extend controller bandwidth while maintaining stable operation enable reduction in $C_{\rm OUT}$. Load-current feedforward has been demonstrated as an effective means to extend bandwidth for VRMs [23]. However, in the integrated context, load-current estimation is especially challenging due to the distributed nature of decoupling capacitors, high variability of on-chip resistors and capacitors, and parasitic poles introduced by analog amplifiers at high-frequencies. As a result, we employ a nonlinear, unlatched PWM controller that offers extended controller bandwidth during large load-current transients while maintaining stability.

III. DESIGN OF A NONLINEAR, UNLATCHED PWM CONTROLLER

A. Overview

The proposed control scheme is shown in Fig. 3. A four-phase interleaved buck converter is composed of four identical hardware phases (HPs) along with clock generation circuitry that provides the switching frequency and phase for each of the HPs, $v_{\rm CLK,1-4}$. Within each HP, $v_{\rm CLK}$ is superimposed onto a DC reference voltage, $V_{\rm SET}$, by means of $R_{\rm CLK}$ to create a triangle wave reference input to the controller, $v_{\rm REF}$, that is centered at the desired DC output voltage

$$v_{\rm REF,DC} = V_{\rm SET}\frac{R_{\rm CLK}}{R_{\rm REF}+R_{\rm CLK}} + \frac{V_{\rm in}}{2} * \frac{R_{\rm REF}}{R_{\rm CLK}+R_{\rm REF}} \quad (12)$$

as shown in (12) and Fig. 4. The feedback voltage, $v_{\rm FB}$, is a superposition of the bridge switching node voltage, $v_{\rm BRIDGE}$, at low frequencies and the output voltage, $v_{\rm OUT}$, at high frequencies. The comparison of $v_{\rm REF}$ and $v_{\rm FB}$ at the delay-optimized continuous comparator determines the steady state duty-cycle, D, according to (13)

$$D = \frac{V_{\rm REF} + \frac{1}{2}v_{{\rm ref},p-p}}{V_{\rm in} + v_{{\rm ref},p-p}} \quad (13)$$

The DC output resistance, $R_{\rm OUT}$, of the IVR can be tuned by $R_{H,1}$ and $R_{H,2}$. As the load current increases, the feedback loop will cause the duty cycle to increase, compensating for the increase in voltage drop across the bridge switches. The duty cycle is buffered and drives $R_{H,1}$, which subsequently causes $v_{\rm REF}$ to slightly increase, offsetting the increased voltage drop across the inductor resistance at higher current. This tuning of the DC output resistance follows the equation

$$R_{\rm OUT} = r'_{sl} - r'_{\rm mos}\frac{R_{H,1}+R_{H,2}}{R_{\rm REF}\|R_{\rm CLK}} \quad (14)$$

where $r_{{\rm mos},n}$ and $r_{{\rm mos},p}$ are, respectively, the effective series resistance of NMOS and PMOS bridge switches for an HP and r_{sl} is the effective series resistance of a single inductor, such that

$$r'_{sl} = \frac{r_{sl}}{N} \text{ and} \quad (15)$$

$$r'_{\rm mos} = \frac{D * r_{{\rm mos},p} + (1-D) * r_{{\rm mos},n}}{N}. \quad (16)$$

Accordingly, the DC output voltage will follow the load-line

$$v_{\rm OUT} = DV_{\rm in} - R_{\rm OUT}I_O. \quad (17)$$

B. Large-Signal Behavior

The time constant, $R_{\rm FB}C_{\rm FB}$, is designed to be slightly longer than $R_{\rm CLK}C_{\rm REF}$ such that in steady state, $v_{\rm FB}$ will slew behind $v_{\rm REF}$ as shown in Fig. 4. In the event of a load current step, the resulting $\delta v_{\rm OUT}/\delta t$ across $C_{\rm OUT}$ couples through $C_{\rm FB}$, and causes $v_{\rm FB}$ to cross $v_{\rm REF}$. At this point, the comparators will switch state and the bridge will apply the appropriate voltage at $v_{\rm BRIDGE}$. Each of the HPs responds asynchronously, such that the ensemble exerts the maximum di/dt within a fraction of the switching period. When an HP becomes unsynchronized, the difference between $v_{\rm FB}$ and $v_{\rm REF}$ is larger and the HP's sensitivity to $dv_{\rm OUT}/dt$ is reduced, driving the HP back to proper synchronization. In this manner, the controller simultaneously provides near immediate asynchronous response to load transients and strong synchronization between HPs in steady state.

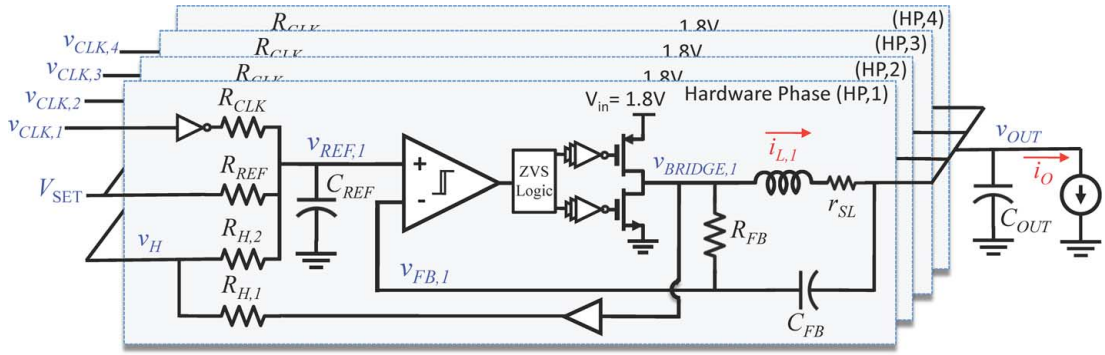

Fig. 3. Diagram of the proposed nonlinear, unlatched PWM control scheme.

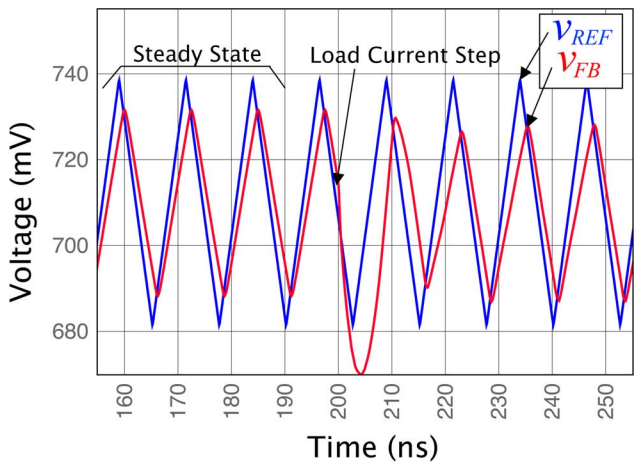

Fig. 4. Simulation of v_{REF} and v_{FB} during steady state and load transient.

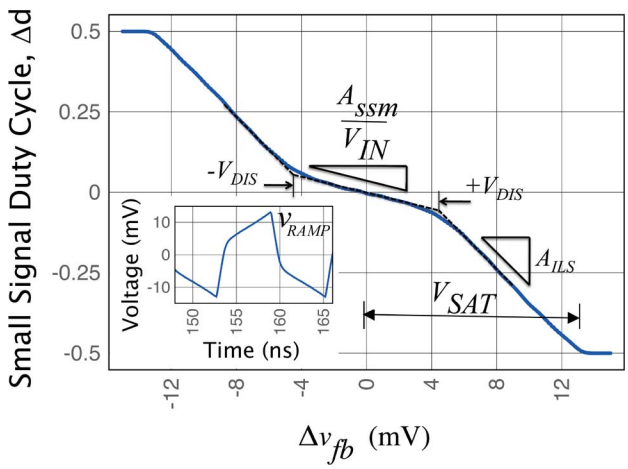

Fig. 5. Small signal change in duty cycle, Δd, as a function of small signal deviation from steady state v_{fb} with $\mathrm{D} = 0.5$. Inset: effective PWM ramp v_{ramp}.

C. Small Signal Dynamics

The small-signal dynamics can be determined using a combination of conventional linear circuit analysis and circuit averaging, if we assume that the frequency content of a small-signal perturbation, Δv_{FB}, is sufficiently below Nf_{sw} for averaging to be valid. The small-signal, steady state gain, A_{SSM}, of the comparator stage is similar to a conventional PWM modulator with the exception that both v_{REF} and v_{FB} have large signal components at f_{sw} in steady state (see Fig. 4), and, hence, the effective PWM ramp signal is $v_{\mathrm{RAMP}} = v_{\mathrm{REF}} - v_{\mathrm{FB}}$ as shown in Fig. 5 inset. A_{SSM} is inversely proportional to the slope of v_{RAMP} where it intersects Δv_{FB}. Fig. 5 shows the feedback gain, the small signal change in the duty cycle, Δd, as a function of Δv_{FB}. The discontinuity in the feedback gain occurs at V_{DIS}, which is approximated as

$$V_{\mathrm{DIS}} = V_{\mathrm{in}} \left(\frac{1}{2R_{\mathrm{CLK}}C_{\mathrm{REF}}} + \frac{1}{2R_{\mathrm{FB}}C_{\mathrm{FB}}} \right) \times \left(\frac{T_{\mathrm{sw}}}{4} \left(1 - \frac{R_{\mathrm{CLK}}C_{\mathrm{REF}}}{R_{\mathrm{FB}}C_{\mathrm{FB}}} \right) + t_{\mathrm{cd}} \right) \quad (18)$$

where t_{cd} accounts for circuit delay through the continuous comparator, ZVS logic and bridge switches. When $|\Delta v_{\mathrm{FB}}| < V_{\mathrm{DIS}}$ the gain through the comparator is linear and approximated as

$$A_{\mathrm{SSM}} \approx \frac{4 * f_{\mathrm{sw}}}{\frac{1}{R_{\mathrm{CLK}}C_{\mathrm{REF}}} + \frac{1}{R_{\mathrm{FB}}C_{\mathrm{FB}}}} \quad (19)$$

for larger deviations, $|\Delta v_{\mathrm{FB}}| > V_{\mathrm{DIS}}$, the gain through the comparator is non-linear and increasing, which provides improved transient response. The instantaneous gain for $|\Delta v_{\mathrm{FB}}| > V_{\mathrm{DIS}}$ is

$$A_{\mathrm{ILS}} \approx \frac{4 * f_{\mathrm{sw}}}{\frac{1}{R_{\mathrm{CLK}}C_{\mathrm{REF}}} - \frac{1}{R_{\mathrm{FB}}C_{\mathrm{FB}}}}. \quad (20)$$

The remainder of the loop transfer function can be determined with linear circuit analysis; the small signal model, transfer functions and output impedance are shown in Figs. 6 and 7. Comparing the open-loop and closed-loop output impedances, we see that the controller regulates the output to a dynamic load-line. Assuming the output capacitor is implemented with on-chip MOS capacitance, the ESR zero occurs above 100 GHz, beyond the range of Fig. 7.

D. Test Chip

The proposed control scheme achieves high feedback bandwidth using a combination of unlatched PWM modulation, nonlinear feedback gain, and high linear feedback bandwidth relative to the effective switching frequency ($f_C \sim Nf_{\mathrm{sw}}/5$). Controllers with such features can be sensitive to noise and/or prone to chaotic behavior, which can cause unpredictable switching,

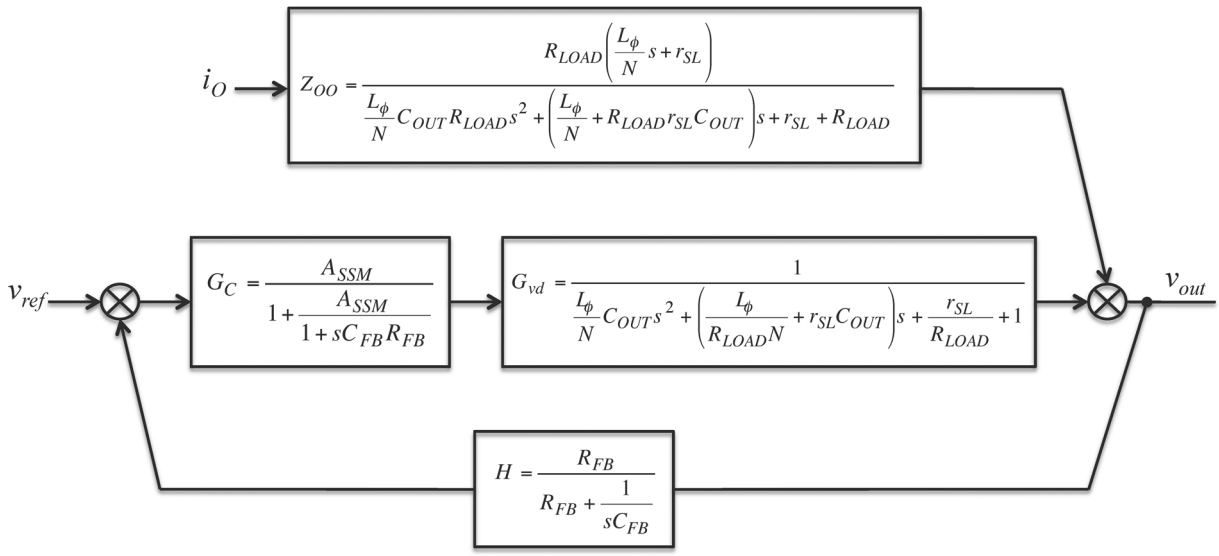

Fig. 6. Small-signal model of control loop.

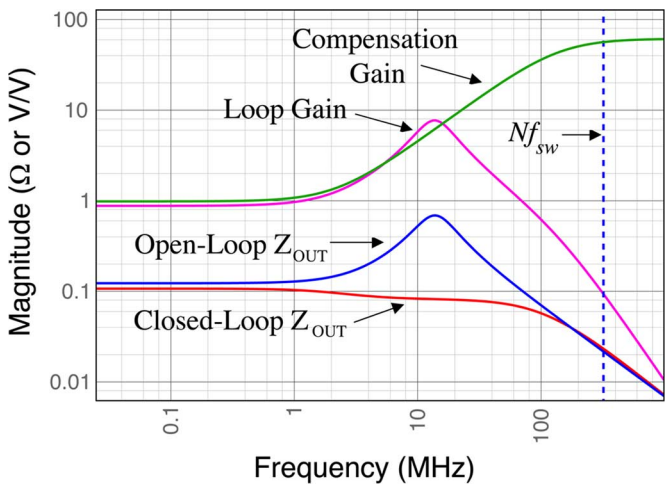

Fig. 7. Small signal transfer functions and output impedance of proposed control scheme.

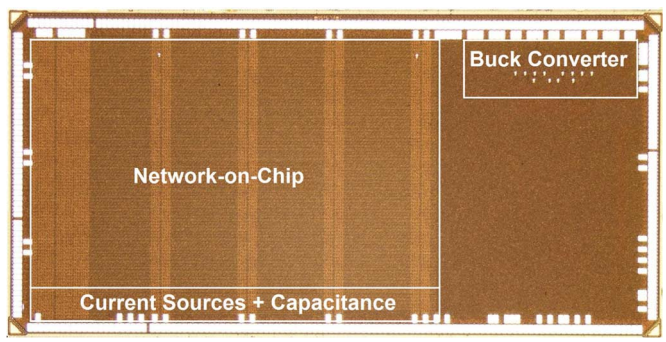

Fig. 8. Photo of test chip.

potentially degrading efficiency and output voltage regulation [23]. Extensive modeling and simulation of the proposed controller was conducted with Matlab and Spectre to verify stability and the absence of bifurcations and strange attractors in the converter operation [25]. Unfortunately, other factors such as inductor or device mismatch may upset the balance between HPs and cause multiple switching, thus, a four-phase buck converter was designed and fabricated on a test chip in a 45 nm SOI process to experimentally verify proper converter operation. The converter provides a regulated supply voltage to a digital load in the form of four 64-tile networks-on-chip (NoC) and a programmable current source capable of generating load-current steps of 1 A with slew rates of ~1 A/100 ps. An image of the chip is shown in Fig. 8 with dimensions of 3 mm by 6 mm. The converter occupies 0.75 mm^2 including all input and output decoupling capacitance (0.32 mm^2excluding these capacitors). It operates with a switching frequency $f_{sw} = 80$ MHz and voltage ripple < 1 mV. The down-converter supports a continuous range of conversion ratios from a 1.5 V supply with a load current as high as 1.25 A. The bridge switches are thick-oxide floating body FETs where the widths have been optimized for 80 MHz switching and 300 mA per phase. A discretely programmable dead-time can be added to the NMOS turn-on transition, allowing zero voltage switching (ZVS) when v_{BRIDGE} transitions from high to low. The continuous comparators have an adjustable hysteresis ranging from 5 mV to 30 mV to prevent chatter. An independent 1 V supply powers the control circuitry and is isolated from the bridge power supply to prevent switching noise form disturbing the controller.

Four 26 nH, SMT-0402 air-core inductors are integrated on top of the chip by bondwire connections as shown in Fig. 9. The inductance value is chosen to limit current ripple such that the converter efficiently operates in continuous conduction mode at f_{sw} of 80 MHz and i_O of 500 mA. The total controller delay during a worst-case load transient is ~700 ps according to simulation, 325 ps for $v_{FB}$to cross v_{REF}, 160 ps for the comparators to switch, and 200 ps for the digital delay through ZVS logic and bridge buffers. With this short delay time, C_{CRIT} required for the specifications in Table I is only 20 nF according to (10). An IVR with the same power train and f_{sw} using a conventional feedback controller with latched PWM modulator would require $C_{OUT} > 100$ nF. The total C_{OUT} on the test chip is ~23 nF,

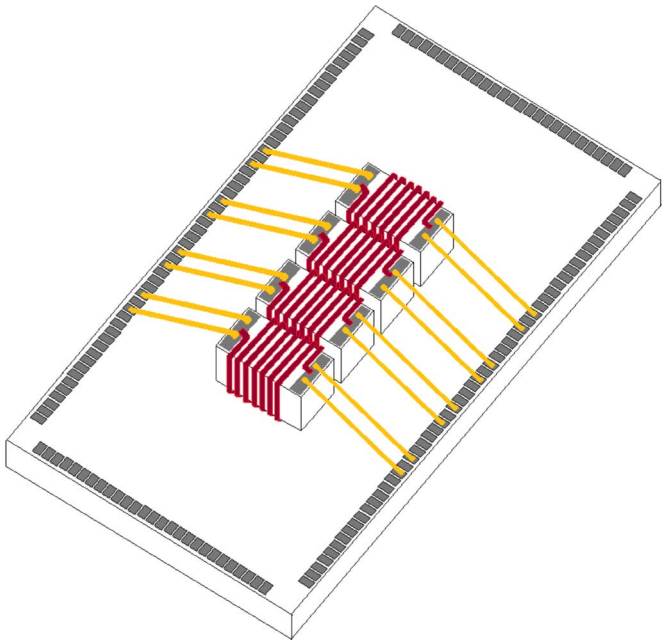

Fig. 9. Illustration of SMT inductor integration by bondwire connections.

including explicit MOS capacitors and non-switching gate capacitance from the digital load.

IV. Network-on-Chip as a Regulated Load

Four independent 64-tile NoCs serve as a realistic digital load for the IVR; the NoC provides a highly scalable platform for exploring granular power distributions given the ease with which traffic patterns can be used to modulate load currents and transients. NoCs are becoming the basic interconnect infrastructure for complex SoCs. Since communication plays a key role in SoCs and given the very strict energy and performance requirements imposed on NoCs, recent designs have reserved a separate voltage-clock domain for the NoC alone [2].

In future SoCs, NoCs will be required to support an increasing number of traffic classes and communication protocols. Adding virtual channels (VCs) to a NoC helps to avoid deadlock and optimize the bandwidth of the physical channels in exchange for a more complex design of the routers. Another, possibly alternative, approach is to build multiple parallel physical networks (multiplanes, MPs) with smaller channels and simpler router organizations. Yoon *et al.* compared the two approaches from a power-performance point of view and concluded that while VCs guarantee higher performance then MPs, MPs are more flexible and better suit applications that have a limited power budget [26]. We organized the NoC in this chip as MPs because they are easier to implement and they better represent an architecture designed with power being the primary concern. Further details of the NoC are provided in the Appendix.

V. Experimental Results

The measured response of the test chip to a load current step from 0.6 A to 1.2 A in ~100 ps is shown in Fig. 10. The simulated behavior is determined from a time-domain Matlab model that is able to capture the nonlinear behavior of the control loop.

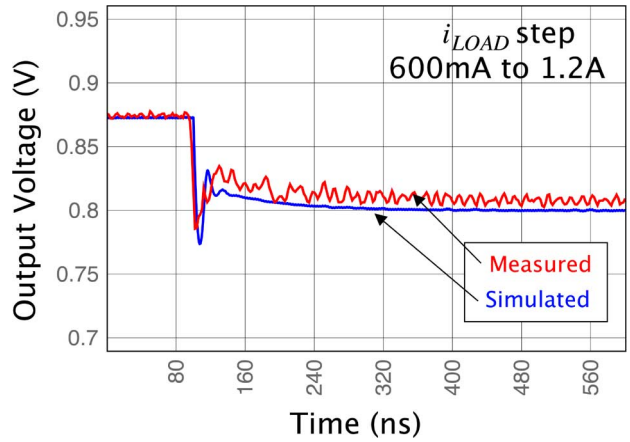

Fig. 10. Measured and simulated load-current transient step response.

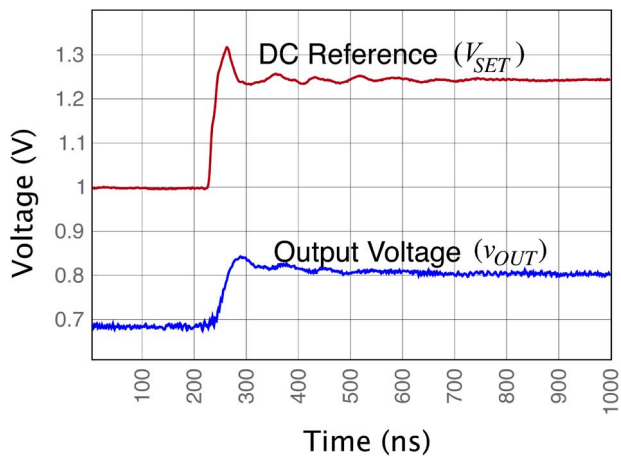

Fig. 11. Measured input step-up response.

The output voltage, $v_{\rm OUT}$, follows the load-line with $R_{\rm LL}$ of 125 mΩ, so that if the converter were scaled to deliver 100 A, $R_{\rm LL}$would scale to 1.25 mΩ. $V_{\rm OUT}$ overshoots the load-line by only ~30 mV and closely matches simulated results with the exception of some ringing that occurs after the step. This ringing is attributed to oscillation between $C_{\rm OUT}$ and the bondwire inductance on the ground return of the load. The estimated resonant frequency of this series LC, 75 MHz, is the same as the frequency of ringing in Fig. 10. Fig. 11 shows the input step-up response, where we see a settling time for $v_{\rm OUT}$ of ~70 ns.

In order to verify the controller switching stability and noise immunity in closed-loop operation, efficiency was measured while the converter operated in open- and closed-loop with the same operating conditions. The open-loop configuration bypasses the comparator to directly drive the bridge with a fixed duty cycle, producing a $v_{\rm OUT}$ of 1 V with i_O of 1 A at $f_{\rm sw}$ of 80 MHz. The converter was subsequently configured to deliver the same output voltage and current at 80 MHz $f_{\rm sw}$ in a closed-loop configuration. In both open- and closed-loop configuration, the efficiency was 78% and the spectral content of the output voltage peaked at ~320 MHz, which is the expected effective switching frequency $Nf_{\rm sw}$.

The converter efficiency (Fig. 12) is hindered by the relatively high $r'_{\rm SL}$ of 120 mΩ, which is dominated by bond wire

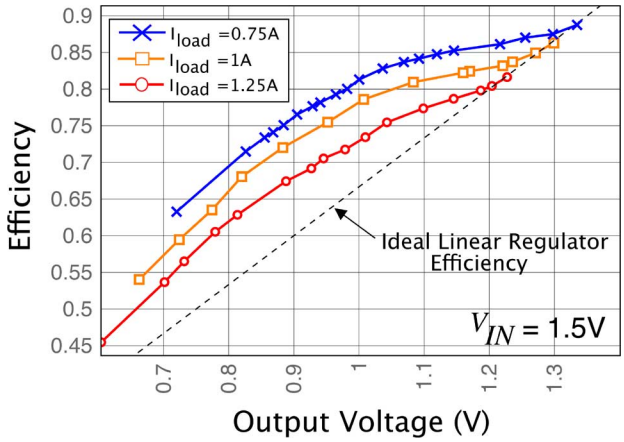

Fig. 12. IVR efficiency as a function of output voltage, v_{OUT}.

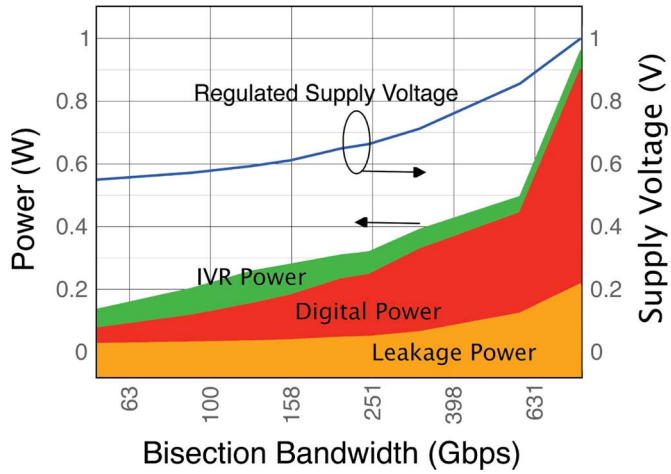

Fig. 13. NoC power consumption as a function of bandwidth (operating frequency).

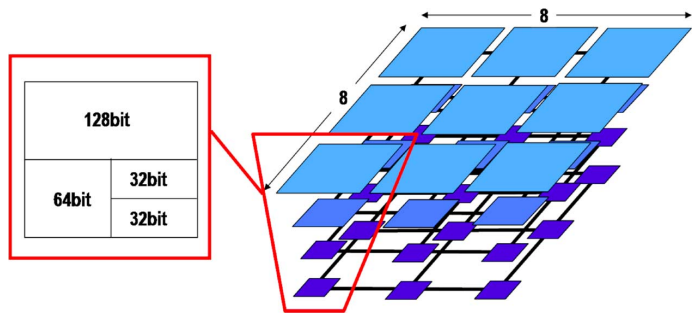

Fig. 14. Network-on-chip architecture.

resistance. The efficiency for $v_{\mathrm{OUT}} < 900$ mV is further adversely impacted by an ESD diode at the v_{BRIDGE} node that turns on with decreasing v_{OUT}. Converter efficiency could be improved by removing the ESD diode and using an alternative packaging strategy that reduces r'_{SL} as demonstrated in [15]. The data in Figs. 10–12 was taken from a single unit. However the efficiency of four units were measured as a check, exhibiting variation that was below the noise of the measurement, each achieving an efficiency of 83% at a current density of 1 A/mm^2 (2.35 A/mm^2 if decoupling capacitor area is not considered) and a 0.66 conversion ratio. The proposed control scheme allows for $\sim 5\times$ reduction in the output capacitance relative to an IVR with conventional control scheme. This corresponds to $\sim 2.2\times$ improvement in total current density for the IVR implementation described here, assuming C_{OUT} is implemented with on-chip MOS capacitance. Fig. 13 shows a breakdown of the test chip's power consumption with scaled NoC supply voltage and frequency (bandwidth). Fig. 13 also illustrates the dramatic decrease in the power consumption of the system when the power supply of the NoC is scaled; even when considering the inefficiency of the IVR, this serves as evidence for the potential power savings achieved with DVFS.

VI. Conclusion

We demonstrate a four-phase integrated buck converter with a novel control scheme that uses an unlatched PWM modulator and nonlinear feedback. The proposed controller provides predictable small-signal dynamics along with fast response to input and load-current steps, which facilitates a $2.2\times$ improvement in current density. Combined with recent developments in inductive energy storage [18]–[21], such a converter could enable implementation of integrated power conversion on a large scale.

Appendix

The NoC has four independent planes, each organized as an eight-by-eight 2D-mesh NoCs (Fig. 14). Each plane supports a different data parallelism: 128, 64, 32, and 32 bits, respectively. Each plane has an independent global clock, and all planes share the common power supply provided by the IVR. In aggregate the entire NoC has 256 routers and a bisection bandwidth of 2 kbit/T, where T is the clock period. For instance, when at the same time all the NoCs run at a clock frequency of 500 MHz (T = 2 ns) the bisection bandwidth is 1 Tbit/s.

All planes adopt traditional wormhole flow control and XY dimension-order routing, which is proven to be simple to implement and deadlock-free for 2D mesh networks. The 2D-Mesh topology is achieved using five-by-five routers (Fig. 15), where four I/O ports are attached to neighbor routers, and the fifth port is used for traffic injection/ejection. The router is a traditional input-queued router. A five-by-five crossbar connects each input to every output, and a simple per-output distributed round-robin arbitration solves the contention when multiple input packets request to be forwarded towards the same output port.

We adopted Ack-Nack as link-level flow control between adjacent routers. In order to implement this protocol, we added two signals to the data bus that carries the flits. One signal validates the flit at a given clock cycle, while the other wire transports back-pressure information. Back-pressure is a way for the downstream router to signal congestion to the upstream router. Under congestion the input queue of a router tends to fill up, and when it is finally full the flit currently in flight on the link cannot be stored properly. The upstream router must then maintain the old flit on the output port in such a way that it can be correctly received by the downstream router once congestion is resolved. Under persistent congestion, since no new flits can be forwarded towards the busy output port, the input queue occupation in the upstream router might grow as well, and might

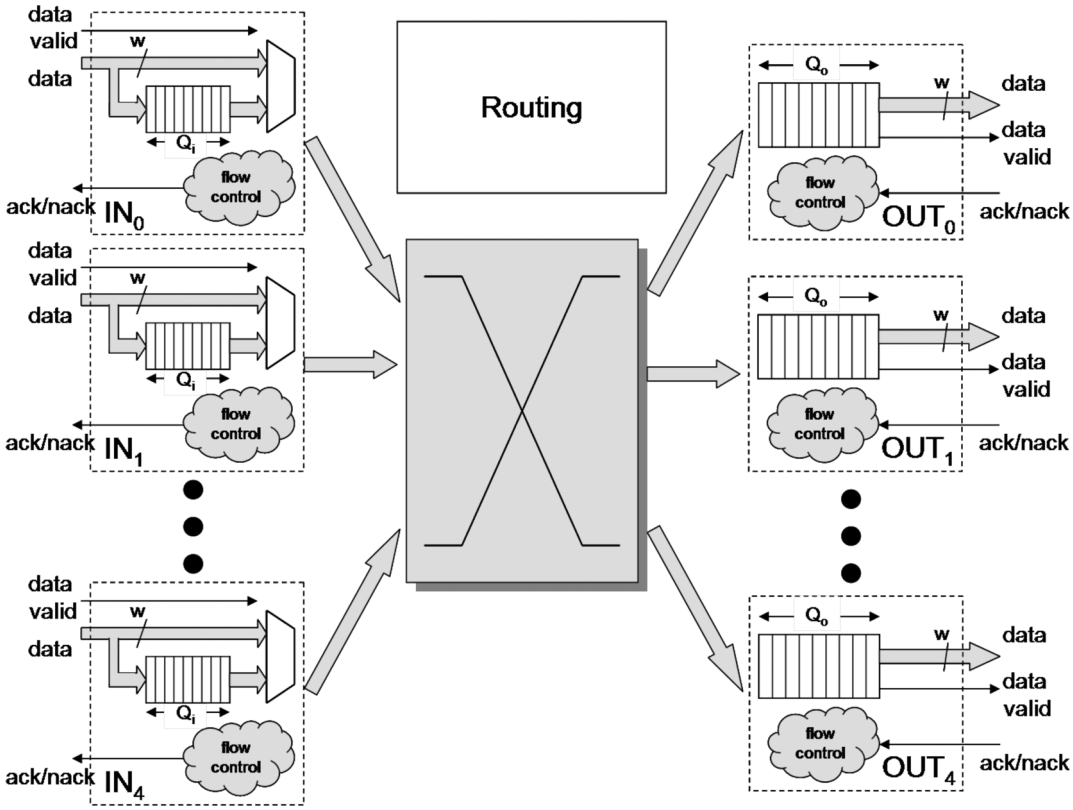

Fig. 15. Network-on-chip router micro-architecture.

require the back-pressure to be propagated backward, up to the traffic source when necessary.

We used a constant depth of $Q_i = 2$ flits for all the input queues, in order to fit the desired topology in the form factor of the chip. Every router has a synchronous output, i.e., $Q_o = 1$. As shown in Fig. 15, we adopted bypassable input queue so that the zero-load latency of traversing one router is one clock cycle. Under no congestion, the incoming flit bypasses the input queue and is routed and stored directly in the appropriate output register. Only under congestion the input queue is used to store the incoming flits until congestion is resolved. We also installe relay-stations (RS) on the links between adjacent routers. RSs are synchronous flow-control aware repeaters, which on one side increase the modularity of the design and facilitate timing closure during layout, while on the other side act as distributed buffers, expanding the capacity of the router input queues, thus alleviating congestion [27]. Our layout is very regular, but under less regular NoCs, RS promise to fix timing exceptions in a very flexible way, without requiring change to the queue sizing within the routers or the network topology. The traffic injected at each router is generated according to externally programmable parameters, supporting four synthetic random traffic patterns: uniform, tornado, transpose and hot-spot. We obtained all the results presented in this paper by averaging across different traffic patterns and traffic injection rates.

REFERENCES

[1] P. Macken, M. Degrauwe, M. V. Paemel, and H. Oguey, "A voltage reduction technique for digital systems," in *IEEE ISSCC Dig.*, 1990, pp. 238–239.

[2] J. Howard, S. Dighe, S. R. Vangal, G. Ruhl, N. Borkar, S. Jain, V. Erraguntla, M. Konow, M. Riepen, M. Gries, G. Droege, T. Lund-Larsen, S. Steibl, S. Borkar, V. K. De, and R. Van Der Wijngaart, "A 48-core IA-32 processor in 45 nm CMOS using on-die message-passing and DVFS for performance and power scaling," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 173–183, Jan. 2011.

[3] W. Kim, D. M. Brooks, and G. Wei, "A fully-integrated 3-level DC/DC converter for nanosecond-scale DVS with fast shunt regulation," in *IEEE ISSCC Dig.*, 2011, pp. 268–270.

[4] W. Kim, M. S. Gupta, G. Wei, and D. M. Brooks, "System level analysis of fast, per-core DVFS using on-chip switching regulators," in *IEEE 14th Int. Symp. High Performance Computer Architecture, HPCA 2008*, Feb. 16–20, 2008, pp. 123–134.

[5] Voltage Regulator-Down. Sep. 2009, (VRD) 11.1, Intel Corp..

[6] G. Patounakis, Y. W. Li, and K. L. Shepard, "A fully integrated on-chip DC-DC conversion and power management system," *IEEE J. Solid-State Circuits*, vol. 39, no. 3, pp. 443–451, Mar. 2004.

[7] H. P. Le, S. R. Sanders, and E. Alon, "Design techniques for fully integrated switched-capacitor DC-DC converters," *IEEE J. Solid-State Circuits*, vol. 46, no. 9, pp. 2120–2131, Sept. 2011.

[8] L. Chang, R. K. Montoye, B. L. Ji, A. J. Weger, K. G. Stawiasz, and R. H. Dennard, "A fully-integrated switched-capacitor 2:1 voltage converter with regulation capability and 90% efficiency at 2.3 A/mm^2," in *2010 IEEE Symp. VLSI Circuits (VLSIC) Dig.*, Jun. 2010, pp. 55–56.

[9] G. Schrom, P. Hazucha, J. Hahn, D. S. Gardner, B. A. Bloechel, G. Dermer, S. G. Narendra, T. Karnik, and V. K. De, "A 480-MHz, multiphase interleaved buck DC-DC converter with hysteretic control," in *2004 IEEE 35th Annual Power Electronics Specialists Conf., PESC'04*, 2004, vol. 6, pp. 4702–4707.

[10] P. Hazucha, G. Schrom, H. Jaehong, B. A. Bloechel, P. Hack, G. E. Dermer, S. Narendra, D. Gardner, T. Karnik, V. K. De, and S. Borkar, "A 233-MHz 80%–87% efficient four-phase DC-DC converter utilizing air-core inductors on package," *IEEE J. Solid-State Circuits*, vol. 40, no. 4, pp. 838–845, Apr. 2005.

[11] G. Schrom, P. Hazucha, F. Paillet, D. J. Rennie, S. T. Moon, D. S. Gardner, T. Karnik, P. Sun, T. T. Nguyen, M. J. Hill, K. Radhakrishnan, and T. Memioglu, "A 100 MHz eight-phase buck converter delivering 12 A in 25 mm^2 using air-core inductors," in *Proc. 22nd Annu. IEEE Applied Power Electronics Conf., APEC 2007*, pp. 727–730.

[12] G. Schrom *et al.*, "A 60 MHz 50 W fine-grain package integrated VR powering a CPU from 3.3 V," *APEC—Special Presentation*, 2010, Palm Springs, CA.
[13] N. Sturcken, M. Petracca, S. Warren, L. P. Carloni, A. V. Peterchev, and K. L. Shepard, "An integrated four-phase buck converter delivering 1 A/mm^2 with 700 ps controller delay and network-on-chip load in 45-nm SOI," in *Proc. IEEE Custom Integrated Circuits Conf. (CICC)*, 2011, pp. 1–4.
[14] J. T. DiBene *et al.*, "A 400 Amp fully integrated silicon voltage regulator with in-die magnetically coupled embedded inductors," *APEC—Special Presentation*, Feb. 25, 2010, Palm Springs, CA.
[15] N. Sturcken *et al.*, "A 2.5D integrated voltage regulator using coupled-magnetic-core inductors on silicon interposer delivering 10.8A/mm^2," in *2012 IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2012, pp. 400–402.
[16] J. Wibben and R. Harjani, "A high efficiency DC-DC converter using 2 nH on-chip inductors," in *2007 IEEE Symp. VLSI Circuits Dig.*, pp. 22–23.
[17] Chip Inductors 0402PA Series Datasheet, Coilcraft Inc..
[18] D. S. Gardner, G. Schrom, F. Paillet, B. Jamieson, T. Karnik, and S. Borkar, "Review of on-chip inductor structures with magnetic films," *IEEE Trans. Magn.*, vol. 45, no. 10, pp. 4760–4766, Oct. 2009.
[19] P. R. Morrow, C. M. Park, H. W. Koertzen, and J. T. DiBene, "Design and fabrication of on-chip coupled inductors integrated with magnetic material for voltage regulators," *IEEE Trans. Magn.*, vol. 47, no. 6, pp. 1678–1686, Jun. 2011.
[20] N. Wang *et al.*, "Integrated on-chip inductors with electroplated magnetic yokes (invited)," *J. Appl. Phys.*, vol. 111, 2012, 07E732.
[21] N. Sturcken *et al.*, "Design of coupled power inductors with crossed anisotropy magnetic core for integrated power conversion," in *Proc. 27th Annual IEEE Applied Power Electronics Conf. Expo., APEC 2012*, 2012, pp. 417–423.
[22] Low Inductance Capacitors LGA Series Datasheet, AVX Corp.
[23] A. V. Peterchev and S. R. Sanders, "Load-line regulation with estimated load-current feedforward: Application to microprocessor voltage regulators," *IEEE Trans. Power Electron.*, vol. 21, no. 6, pp. 1704–1717, Nov. 2006.
[24] P.-L. Wong, F. C. Lee, P. Xu, and K. Yao, "Critical inductance in voltage regulator modules," *IEEE Trans. Power Electron.*, vol. 17, no. 4, pp. 485–492, Jul. 2002.
[25] J. H. B. Deane and D. C. Hamill, "Analysis, simulation and experimental study of chaos in the buck converter," in *21st Annual IEEE Power Electronics Specialists Conf., 1990. PESC'90 Record*, pp. 491–498.
[26] Y. J. Yoon, N. Concer, M. Petracca, and L. P. Carloni, "Virtual channels vs. multiple physical networks: A comparative analysis," in *Proc. 47th ACM/IEEE Design Automation Conf. (DAC 2010)*, pp. 162–165.
[27] N. Concer, M. Petracca, and L. P. Carloni, "Distributed flit-buffer flow control for networks-on-chip," in *Proc. 6th Int. Conf. Hardware/Software Codesign & System Synthesis (CODES+ISSS)*, 2008.

Noah Sturcken (S'08) received the B.S. degree *summa cum laude* in electrical and computer engineering from Cornell University, Ithaca, NY, in 2008 and the M.S. degree in electrical engineering from Columbia University, New York, NY, in 2009. He is currently working toward the Ph.D. degree at Columbia University.

He has held several internships, most recently at AMD's Research and Development Laboratory. His research interests include integrated magnetic materials for power conversion, switched mode power supplies, control theory and mixed signal circuit design.

Mr. Sturcken is a member of Eta Kappa Nu.

Michele Petracca received the M.S. and Ph.D. degrees in electronic engineering from Politecnico di Torino, Italy. He was a visiting Ph.D. Student and then a Post-Doctorate Research Scientist in the Computer Science Department of Columbia University in the City of New York.

He is currently with Cadence Design Systems. His interests include networking, on-chip communication infrastructures, system-level design, hardware-software interaction and power-management in embedded systems.

Steven Warren received the B.S. degree in electrical engineering from Tufts University, Medford, MA, in 2008 and the M.S. degree from Columbia University, New York, NY in 2010. He is currently working toward the Ph.D. degree at Columbia University.

He has interned at MIT's Lincoln Laboratory, where he helped design radar downconverters. Additionally, he consulted for the microbial fuel cell startup, Trophos Energy, where he designed potentiometry equipment. He is currently interested in integrated bioelectronic sensors.

Paolo Mantovani received the B.S. at Alma Mater Studiorum—Universita' di Bologna in 2008 and the M.S. in electronic engineering at Politecnico di Torino in 2010. He also completed the piano diploma program at the Conservatory of Music L. Campiani in Mantova in 2007.

He is a first year Ph.D. student in computer science at Columbia University, where he works in the system level design research group, under Professor Luca Carloni.

Luca P. Carloni (S'95–M'04–SM'09) received the Laurea degree (*summa cum laude*) in electrical engineering from the Università di Bologna, Bologna, Italy, in 1995, and the M.S. and Ph.D. degrees in electrical engineering and computer sciences from the University of California, Berkeley, in 1997 and 2004, respectively.

He is currently an Associate Professor with the Department of Computer Science, Columbia University, New York, NY. He has authored over 90 publications and is the holds one patent. His current research interests include design methodologies and tools for integrated circuits and systems, distributed embedded systems design, and design of high-performance computer systems.

Dr. Carloni received the Faculty Early Career Development (CAREER) Award from the National Science Foundation in 2006, was selected as an Alfred P. Sloan Research Fellow in 2008, and received the ONR Young Investigator Award in 2010. He is the recipient of the 2002 Demetri Angelakos Memorial Achievement Award for "recognition of altruistic attitude toward fellow graduate students." In 2002, one of his papers was selected for the "Best of ICCAD," a collection of the best IEEE International Conference on Computer-Aided Design papers of the past 20 years. He is a Senior Member of the Association for Computing Machinery (ACM).

Angel V. Peterchev (S'96–M'05) received the A.B. degree in physics and engineering sciences from Harvard University in 1999 and the M.S. and Ph.D. degrees in electrical engineering from the University of California at Berkeley in 2002 and 2005, respectively. He completed postdoctoral training in transcranial magnetic stimulation at Columbia University.

He is presently Assistant Professor in the Departments of Psychiatry and Behavioral Sciences, Biomedical Engineering, and Electrical and Computer Engineering at Duke University. His primary research interests are in the design and modeling of devices and application paradigms for transcranial brain stimulation.

Dr. Peterchev has also published on digital control in power electronics and on architecture and control strategies for microprocessor voltage regulators. Dr. Peterchev has authored over 20 journal papers, three book chapters, and multiple conference papers and abstracts.

Kenneth L. Shepard (M'91–SM'03–F'08) received the B.S.E. degree from Princeton University, Princeton, NJ, in 1987 and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1988 and 1992, respectively.

From 1992 to 1997, he was a Research Staff Member and Manager with the VLSI Design Department, IBM T. J. Watson Research Center, Yorktown Heights, NY, where he was responsible for the design methodology for IBM's G4 S/390 microprocessors. Since 1997, he has been with Columbia University, New York, where he is now Professor of Electrical Engineering and Biomedical Engineering. He also was Chief Technology Officer of CadMOS Design Technology, San Jose, CA, until its acquisition by Cadence Design Systems in 2001. His current research interests include power electronics, carbon-based devices and circuits, and CMOS bioelectronics.

Dr. Shepard was Technical Program Chair and General Chair for the 2002 and 2003 International Conference on Computer Design, respectively. He has served on the Program Committees for IEDM, ISSCC, VLSI Symposium, ICCAD, DAC, ISCAS, ISQED, GLS-VLSI, TAU, and ICCD. He received the Fannie and John Hertz Foundation Doctoral Thesis Prize in 1992, a National Science Foundation CAREER Award in 1998, and the 1999 Distinguished Faculty Teaching Award from the Columbia Engineering School Alumni Association. He has been an Associate Editor of IEEE TRANSACTIONS ON VERY LARGE-SCALE INTEGRATION (VLSI) SYSTEMS and is currently an Associate Editor for the IEEE JOURNAL OF SOLID-STATE CIRCUITS and IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS.