

Heterogeneous Methodology for Energy Efficient Distribution of On-Chip Power Supplies

Inna Vaisband, *Student Member, IEEE*, and Eby G. Friedman, *Fellow, IEEE*

Abstract—To provide a high quality power delivery system, the power needs to be regulated on-chip with ultra-small locally distributed power efficient converters. Historically, power efficient switching converters require large physical area, while compact linear power supplies exhibit high power conversion losses, which are not ideal for on-chip integration. To exploit the advantages of existing power supplies, a heterogeneous power delivery system is proposed. The power efficiency of the system is shown to be a strong function of the on-chip distribution of the power supplies. The optimal power distribution system with minimum power losses is determined by exhaustively comparing the power efficiency for all possible power supply topologies. A heterogeneous system with ten on-chip voltage domains and an optimal power distribution network has been evaluated, demonstrating up to 93% power efficiency. A power efficient clustering of the on-chip power supplies with linear computational complexity is also proposed. Heterogeneous power delivery systems with up to 100 on-chip voltage domains have been evaluated with power supplies distributed with linear computational complexity. A maximum 1.5% drop in power efficiency from the optimal solution has been observed, yielding a near optimal and high fidelity power supply distribution system.

Index Terms—Linear regulator, power conversion, power distribution, power regulation, power system management, switched-mode power supply.

I. INTRODUCTION

THE delivery of high quality power to the on-chip circuitry with minimum energy loss is a fundamental requirement of all integrated circuits (ICs). To supply sufficient power, a higher unregulated dc voltage is usually stepped down and regulated within the power delivery system [1]. Power conversion and regulation resources should be efficiently managed to supply high quality power with minimum energy losses within multiple on-chip voltage domains [2].

The design complexity of a power delivery system increases with greater requirements on the quality of the power supply, limitations of the passive elements, board and package parasitic impedances, and limited number of I/O pins. In a modern system-on-chip (SoC), the power supplies provide the required voltage for the ICs within the overall system (CPUs, GPUs, hard disks, storage, sensors, and others), as well as the analog and digital circuit blocks within the ICs. A regulated 12 V output

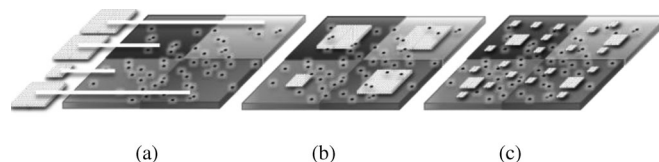


Fig. 1. Power delivery system with four voltage domains: (a) off-chip, (b) integrated on-chip, and (c) distributed point-of-load power supplies for voltage conversion and regulation.

voltage is often derived off-chip from a 48 V battery voltage [1]. The on-chip dc voltage levels are significantly lower and range from a fraction of a volt in low power digital blocks to several volts in input/output buffers, high precision analog blocks, and storage ICs. Furthermore, to effectively exploit the power-delay tradeoff, additional power management techniques such as dynamic voltage scaling (DVS) and dynamic voltage and frequency scaling (DVFS) are employed, further increasing the design complexity of the power delivery system. Thus, to efficiently manage the power delivered to a modern SoC, a methodology to distribute and manage the power supplies is required. To the authors' knowledge, this study is the first contribution presenting a methodology that provides rules for heterogeneous power delivery and performance evaluation of an overall power delivery system.

Traditionally, power is managed off-chip with energy-efficient power converters [see Fig. 1(a)], delivering high quality dc voltage and current to the electrical grid that reliably distributes the on-chip power. The supply voltage, current density, and parasitic impedance, however, scale aggressively with each technology generation, degrading the quality of the power delivered from the off-chip power supplies to the on-chip load circuitry. The power supply in a package (PSiP) approach with partially off-chip yet in package power supplies has recently been considered as an intermediate power supply technology with respect to cost, complexity, and performance [3]. The power is regulated on-chip to lower the parasitic impedance of both the board and package [see Fig. 1(b)]. To fully integrate a power converter on-chip, advanced passive components, packaging technologies, and circuit topologies are essential. Recently, several power converters suitable for on-chip integration have been fabricated [4]–[24]. Based on these power converters, a power supply system with several on-chip power converters can be developed to improve the quality of the power delivered within the ICs.

On-chip power supply integration is an important cornerstone to the power supply design process. A single on-chip power converter is, however, not capable of supplying sufficient, high quality regulated current to the billions of current loads within

Manuscript received June 10, 2012; revised August 31, 2012; accepted October 24, 2012. Date of current version February 15, 2013. Recommended for publication by Associate Editor Y.-F. Liu.

The authors are with the Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627 USA (e-mail: vaisband@ece.rochester.edu; friedman@ece.rochester.edu).

Digital Object Identifier 10.1109/TPEL.2012.2230408

the tens of on-chip voltage domains. To maintain a high quality power supply despite increasing on-chip parasitic impedances, hundreds of ultra-small power converters should ultimately be integrated on-chip, close to the loads within the individual multiple voltage domains [4]–[7]. A distributed point-of-load (POL) power supply system is illustrated in Fig. 1(c).

While the quality of the power supply can be efficiently addressed with a distributed multi-voltage domain system, the limited power efficiency of the on-chip converters is a primary concern for the POL approach. The high power efficiency of the off-chip power converters is traded off for a small area and locally regulated current and voltage. To address the concerns of a POL power supply system, existing power converter topologies are described and compared in Section II. Heterogeneous power delivery is introduced in Section III to both decrease the noise and increase the efficiency of the supplied power. Algorithms to determine how best to distribute the power supplies within a heterogeneous power delivery system and related simulation results are presented, respectively, in Sections IV and V. This paper is summarized in Section VI.

II. POWER CONVERTER TOPOLOGIES

Switching and linear dc–dc converters are the most commonly used topologies for dc–dc conversion and regulation. Historically, a large switching mode power supply (SMPS) is preferred over a compact linear power supply due to the high, ideally 100% power efficiency of an SMPS. With on-chip power converters, strict area constraints are imposed on the dc–dc converters, affecting the choice of power supply topology. Compact switching power converters can potentially be designed at higher switching frequencies. The parasitic impedance in these converters however increases, degrading the power efficiency of the power delivery system. The physical size and power efficiency of switching and linear topologies are discussed, respectively, in Sections II-A and II-B. Some conclusions reviewing the preferable choice of on-chip power supply topology are provided in Section II-C.

A. Switching Converters

A typical SMPS converts an input voltage V_{IN} to an output voltage V_{DD} , supplying the required current I_{DD} to the load circuitry. These converters are operated by a switching signal fed into passive energy storage components through a power MOSFET controlled by a pulse width modulator (PWM). A common step-down SMPS converter operating as a buck converter is shown in Fig. 2. The stored input energy is restored at the output at the required voltage level, maintaining high power efficiency up to a frequency f_s of a few megahertz [25]. The operational mode of a buck converter, output voltage, output current, and transient performance are affected by the output LC filter and controller in the feedback loop, as illustrated in Fig. 2. The on-chip integration of SMPS converters is greatly complicated due to I/O limitations, and constraints related to the physical size of the passive elements [26]. The area required by the passive components to achieve a specific impedance is inversely proportional to the frequency, and can be reduced

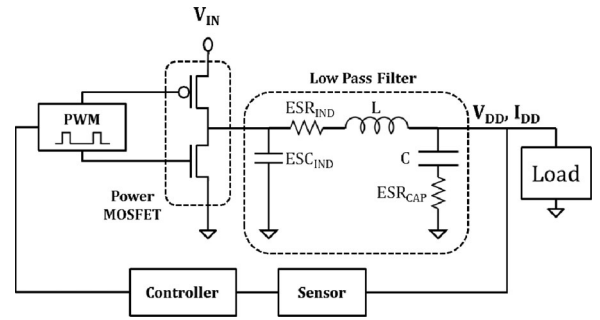


Fig. 2. Buck converter circuit.

in on-chip converters by operating at ultra-high switching frequencies. Conversely, an SMPS operating at a high frequency is more greatly affected by the parasitic impedances, degrading the power efficiency of the converter.

The area of a buck converter is dominated by the size of the passive elements and is

$$A_{\text{Buck}} \approx \frac{L}{L_{\square}} + \frac{C}{C_{\square}} \quad (1)$$

where L_{\square} and C_{\square} are, respectively, the inductance and capacitance per square micrometer of the LC filter. The voltage regulation is a primary concern for POL power delivery. In discontinuous conduction mode (DCM) [5], the current ripple $\gamma_i I_{DD}$ within the inductor L exceeds the output current I_{DD} , and the voltage V_{DD} at the output of a converter becomes load dependent, degrading the quality of the delivered power. To support high-load regulation, the buck converter is assumed in this analysis to be loaded with an output current I_{DD} that exceeds the current ripple ($\gamma_i I_{DD} \leq I_{DD}$), yielding expressions for the inductor and capacitor operating in the continuous conduction mode (CCM) [5]

$$L = \frac{V_{IN} - V_{DD}}{2f_s \gamma_i I_{DD}} \cdot \frac{V_{DD}}{V_{IN}} \quad (2)$$

$$C = \frac{\gamma_i I_{DD}}{8f_s \gamma_v V_{DD}} \quad (3)$$

where $\gamma_v V_{DD}$ is the voltage ripple at the converter output and V_{DD} is the voltage at the load. To satisfy the tight load regulation specifications, the output voltage ripple is assumed to range up to 10% of V_{DD} ($\gamma_v = 0.1$). Substituting (2) and (3) into (1), the area of a buck converter is

$$A_{\text{Buck}} \approx \left(\frac{(V_{IN} - V_{DD}) V_{DD}}{2L_{\square} V_{IN} f_s} \right) \frac{1}{\gamma_i I_{DD}} + \left(\frac{1}{8C_{\square} \gamma_v V_{DD} f_s} \right) \gamma_i I_{DD}. \quad (4)$$

At low values of the current ripple, the area of a buck converter is dominated by the inductor and increases with smaller values of $\gamma_i I_{DD}$. Alternatively, at larger values of $\gamma_i I_{DD}$, the area of a buck converter is dominated by the capacitor size and is proportional to the current ripple. An optimum ripple current $\gamma_{i,\text{OPT}} I_{DD}$, therefore, exists that minimizes the area of a buck converter for a target output voltage ripple $\gamma_v V_{DD}$, and input and output

voltage levels

$$\gamma_{i,\text{OPT}} I_{\text{DD}} = \begin{cases} \gamma_G G \cdot V_{\text{DD}}, & \gamma_G G \cdot V_{\text{DD}} \leq I_{\text{DD}} \\ I_{\text{DD}}, & \gamma_G G \cdot V_{\text{DD}} > I_{\text{DD}}, \end{cases} \quad (5)$$

where $\gamma_G G \cdot V_{\text{DD}} \triangleq 2\sqrt{\gamma_v (1 - V_{\text{DD}}/V_{\text{IN}}) C_{\square}/L_{\square}} \cdot V_{\text{DD}}$, and $\gamma_G G$ is the output conductance ripple and depends upon the technology parameters, converted voltages, and regulation specification. The minimum area of the buck converter is therefore

$$A_{\text{Buck,MIN}} \approx \frac{1}{2f_s} \begin{cases} \sqrt{\frac{1}{\gamma_v} \left(1 - \frac{V_{\text{DD}}}{V_{\text{IN}}}\right) \frac{L_{\square} C_{\square}}{L_{\square} C_{\square}}}, & \gamma_G G \cdot V_{\text{DD}} \leq I_{\text{DD}} \\ \left(\frac{1 - V_{\text{DD}}}{V_{\text{IN}}}\right) \frac{V_{\text{DD}}}{L_{\square} I_{\text{DD}}} + \left(\frac{1}{4\gamma_v}\right) \frac{I_{\text{DD}}}{C_{\square} V_{\text{DD}}} \approx \frac{(1 - V_{\text{DD}}) V_{\text{DD}}}{L_{\square} I_{\text{DD}}}, & \gamma_G G \cdot V_{\text{DD}} > I_{\text{DD}}. \end{cases} \quad (7)$$

Thus, in CCM at low current loads ($I_{\text{DD}} < \gamma_G G \cdot V_{\text{DD}}$), the minimum area of a buck converter is dominated by the inductance characteristics and increases with smaller values of I_{DD} . However, for values of I_{DD} larger than $\gamma_G G \cdot V_{\text{DD}}$, the minimum size of a buck converter does not strongly depend on I_{DD} . Alternatively, both the power MOSFET losses and power dissipated in the LC filter are dominant at different frequencies, conversion voltages, and current levels in CCM. The power dissipated in the power MOSFET comprises the MOSFET switching power ($\propto f_s V_{\text{IN}}^2$), and the resistive power ($\propto R_{\text{ON}} I_{\text{DD}}^2$) dissipated by the effective resistor R_{ON} of the MOSFET, yielding

$$P_{\text{Buck,MOS}} = \frac{l_{\text{min}}^2 \cdot f_s V_{\text{IN}}^2}{\mu R_{\text{ON}} (V_{\text{IN}} - V_T)} + \frac{4}{3} R_{\text{ON}} \frac{V_{\text{DD}}}{V_{\text{IN}}} I_{\text{DD}}^2, \quad (9)$$

where l_{min} is the minimum channel length, μ is the MOSFET carrier mobility, and V_T is the threshold voltage [27].

From (9), increasing the effective resistance of the MOSFET reduces the switching power dissipation, while increasing the resistive loss. Thus, an optimum MOSFET resistance $R_{\text{ON}}^{\text{OPT}}$ exists that minimizes the power dissipated in a MOSFET, yielding

$$R_{\text{ON}}^{\text{OPT}} = \sqrt{\frac{3}{4} \frac{l_{\text{min}}^2}{\mu (V_{\text{IN}} - V_T)} \cdot f_s \frac{V_{\text{IN}}}{V_{\text{DD}}} \cdot \frac{V_{\text{IN}}}{I_{\text{DD}}}}, \quad (10)$$

and

$$P_{\text{Buck,MOS}}^{\text{MIN}} = 2I_{\text{DD}} \sqrt{\frac{4}{3} \frac{l_{\text{min}}^2}{\mu (V_{\text{IN}} - V_T)} \cdot f_s V_{\text{IN}} V_{\text{DD}}}. \quad (11)$$

The power dissipated in an LC filter [27] comprises the power losses due to the resistive ESR_{IND} and capacitive ESC_{IND} parasitic impedances of the inductor

$$P_{\text{Buck,IND}} = \frac{4}{3} \text{ESR}_{\text{IND}} \cdot I_{\text{DD}}^2 + \text{ESC}_{\text{IND}} f_s \cdot V_{\text{IN}}^2, \quad (12)$$

and the power losses due to the parasitic resistance of the capacitor ESR_{CAP}

$$P_{\text{Buck,CAP}} = \text{ESR}_{\text{CAP}} (\gamma_i I_{\text{DD}})^2. \quad (13)$$

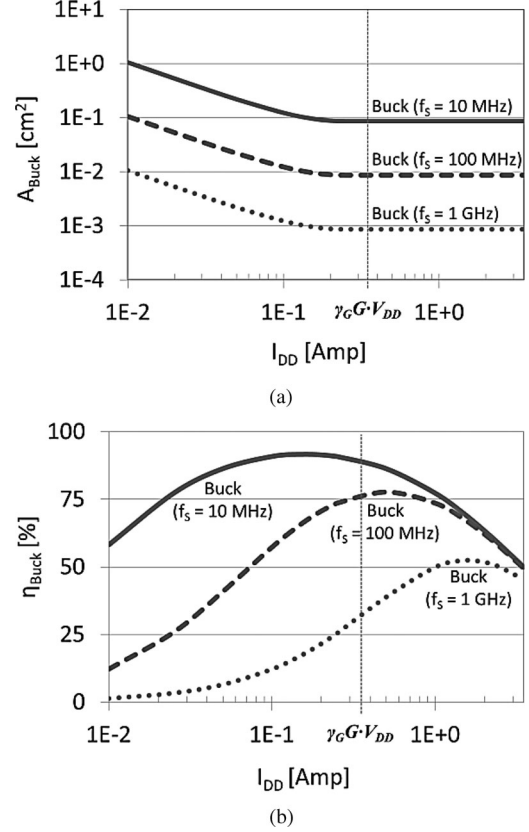


Fig. 3. Buck converter (a) physical area, and (b) power efficiency versus load current for moderate, high, and ultrahigh switching frequencies.

The total power dissipation and power efficiency of the buck converter are, respectively

$$P_{\text{Buck}} = \left(\frac{4}{3} \text{ESR}_{\text{IND}} + \text{ESR}_{\text{CAP}} \right) \cdot I_{\text{DD}}^2 + 2\sqrt{\frac{4}{3} \frac{l_{\text{min}}^2}{\mu (V_{\text{IN}} - V_T)} \cdot f_s V_{\text{IN}} V_{\text{DD}}} \cdot I_{\text{DD}} + \text{ESC}_{\text{IND}} \cdot f_s \cdot V_{\text{IN}}^2 \quad (14)$$

Typical passive component parameters, represented by [28]–[30], and technology parameters [31] are assumed to demonstrate power and area tradeoffs and trends in buck converters. Current load levels from a few milliamperes to several amperes, and the input and output voltages of, respectively, 1 and 0.7 V, are considered. The physical area [see (7)] and power efficiency [see (15), shown at the bottom of the next page] trends are depicted in Fig. 3 for moderate (10 MHz), high (100 MHz), and ultra-high (1 GHz) switching frequencies. At low current loads, the power losses of a buck converter in CCM are dominated by the parasitic capacitance of the inductor ESC_{IND} , decreasing the power efficiency at lower I_{DD} and larger converter size ($A_{\text{Buck}} \propto 1/I_{\text{DD}}$ for $I_{\text{DD}} < \gamma_G G \cdot V_{\text{DD}}$). Alternatively, at high current loads, the power efficiency is dominated by the parasitic resistance of the inductor ESR_{IND} and capacitor ESR_{CAP} , increasing the power losses of a buck converter at higher values of I_{DD} . Thus, a buck converter exhibits a parabolic-shaped

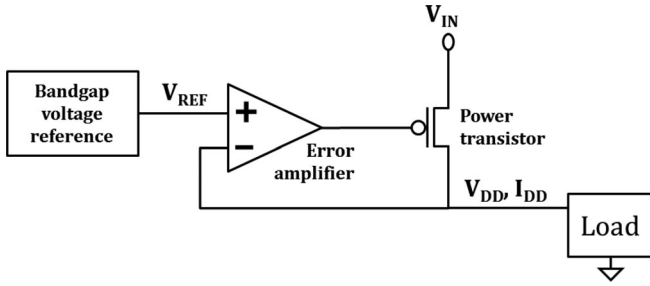


Fig. 4. LDO circuit.

power efficiency with current in CCM, while the physical size of the converter is reduced at higher currents. Therefore, by targeting high switching frequencies, the preferred current load can be determined to convert a voltage with minimum power losses and area for a specific value of switching frequency f_s . For example, as shown in Fig. 3, a preferable current exists for $f_s = 100$ MHz and $f_s = 1$ GHz since the maximum power efficiency is reached at $I_{DD} > \gamma G \cdot V_{DD}$, but not at $f_s = 10$ MHz. The minimum power loss in (15) is proportional to $\sqrt{f_s}$, significantly degrading the power efficiency at high frequencies. Alternatively, the size of the power supply converter is proportional to $1/f_s$, and decreases at higher frequencies, exhibiting an undesirable tradeoff between the power efficiency and physical size of a buck converter.

The high power efficiency of traditional large power converters operating at low frequencies is, therefore, traded off for smaller physical size at ultra-high switching frequencies.

B. Linear Converters

To supply a specific voltage V_{DD} and current I_{DD} to the load circuitry, a linear power supply converts an input dc voltage V_{IN} using a resistive voltage divider controlled by feedback from the output. The primary drawback of a linear topology is the resistive power losses that increase with a larger $V_{IN} - V_{DD}$ voltage drop, which limit the power efficiency to V_{DD}/V_{IN} . Alternatively, linear converters exhibit a relatively small area, an important characteristic for on-chip integration. A low dropout (LDO) dc-dc regulator, depicted in Fig. 4, is a standard linear converter that operates with a low $V_{IN} - V_{DD}$ voltage drop.

The total current supplied by a linear converter comprises the useful LDO current I_{DD} that flows to the load, and the short-circuit current dissipated in the bandgap reference and error amplifier. Power- and area-efficient voltage references have recently been reported [11]–[14]. The total LDO current is, therefore, dominated by the error amplifier and power transistor currents. To mitigate transient voltage peaks while supporting

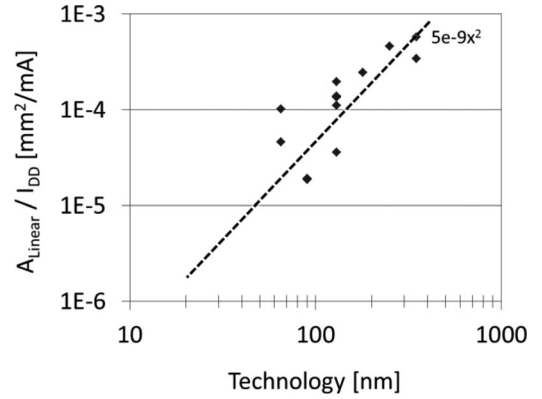


Fig. 5. LDO physical area per 1 mA load.

fast changes in the load current, larger currents should be utilized within the error amplifier, increasing the short-circuit current. Alternatively, to satisfy current load requirements in modern high performance circuits, high currents of up to several amperes are required by the load circuitry. The current flow within an LDO is, therefore, dominated by the load current I_{DD} . In this case, both the area and power dissipation of a linear converter are primarily dictated by the size of the output power transistor and the dissipated power. Thus, the area of an LDO is proportional to the width W of the output transistor, yielding

$$A_{\text{Linear}} \propto \alpha \cdot W l_{\text{min}} = \alpha \cdot \frac{I_{DD} \cdot l_{\text{min}}^2}{\mu C_{\text{OX}} (V_{\text{IN}} - V_T)^2}, \quad (16)$$

where α is the transistor area-to- $W \cdot l_{\text{min}}$ ratio, l_{min} is the minimum channel length, μ is the MOSFET carrier mobility, and C_{OX} is the gate oxide capacitance. To accommodate the effect of the line and load specifications that may significantly affect the physical size of an LDO, a typical area per 1 mA load [4]–[14] (see Fig. 5) is considered for those LDOs with a high current load, exhibiting a parabolic trend of area with minimum technology length ($A_{\text{Linear}}/I_{DD} \propto l_{\text{min}}^2$).

The ratio $A_{\text{Linear}}/I_{DD} = 5 \cdot 10^{-6}$ mm²/mA corresponds to the 28 nm technology node considered in Fig. 6. Typical 28-nm CMOS technology parameters [31], and input and load voltages are assumed in this analysis to demonstrate the need for a large power transistor to supply high current to the load (see Fig. 6). The size of the linear converter ranges from $60 \times 60 \mu\text{m}^2$ for $I_{DD} = 0.5$ A to $150 \times 150 \mu\text{m}^2$ for $I_{DD} = 3.5$ A (see Fig. 6), which can be further reduced with technology scaling ($A_{\text{Linear}} \propto l_{\text{min}}^2$) and advanced design solutions [4]–[24]. The current can, therefore, be supplied to the load with an LDO that is orders of magnitude smaller than a corresponding buck converter.

$$\begin{aligned} \eta_{\text{Buck}} &= \frac{P_{\text{Load}}}{P_{\text{Load}} + P_{\text{Buck}}} \\ &= \frac{I_{DD} V_{DD}}{\left(\frac{4}{3} \text{ESR}_{\text{IND}} + \text{ESR}_{\text{CAP}}\right) \cdot I_{DD}^2 + \left[V_{DD} + 2\sqrt{\frac{4}{3} \frac{l_{\text{min}}^2}{\mu (V_{\text{IN}} - V_T)}} \cdot f_s V_{\text{IN}} V_{DD}\right] \cdot I_{DD} + \text{ESC}_{\text{IND}} \cdot f_s \cdot V_{\text{IN}}^2} \end{aligned} \quad (15)$$

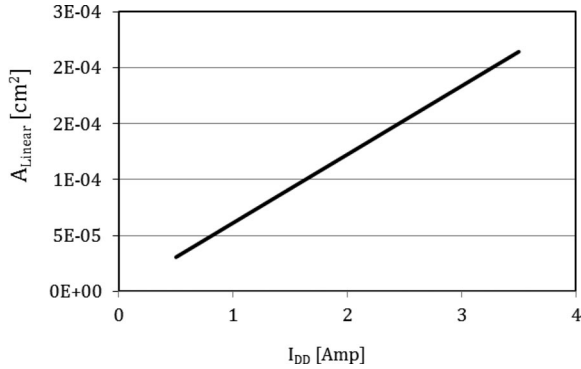
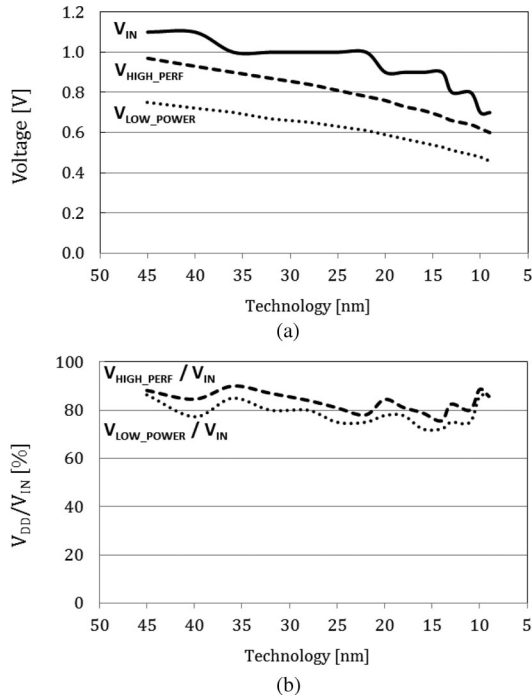


Fig. 6. LDO area for typical current loads.

Fig. 7. Trends in typical (a) high performance ($V_{\text{HIGH_PERF}}$), low power ($V_{\text{LOW_POWER_IC}}$), and internal core primary (V_{IN}) voltage supplies, and (b) voltage conversion ratios ($V_{\text{HIGH_PERF}}/V_{\text{IN}}$) and ($V_{\text{LOW_POWER_IC}}/V_{\text{IN}}$) (ITRS 2011).

The power dissipation of an LDO is

$$P_{\text{Linear}} \approx (V_{\text{IN}} - V_{\text{DD}}) I_{\text{DD}}. \quad (17)$$

Thus, the power loss in a linear converter increases with a higher $V_{\text{IN}} - V_{\text{DD}}$ drop, degrading the power efficiency of the converter. Recent supply voltage trends are illustrated in Fig. 7 for the internal core primary voltage V_{IN} , and typical high and low V_{DD} levels [31], yielding efficiency bounds within the 70% to 90% range of the $V_{\text{DD}}/V_{\text{IN}}$ ratio shown in Fig. 7. Thus, a moderate LDO power efficiency $\eta_{\text{Linear}} = V_{\text{DD}}/V_{\text{IN}}$ of at least 70% can be predicted.

Sub/near threshold computing is a promising technique to reduce the power consumed by an IC [32], [33]. To provide a stable supply voltage at sub/near threshold levels, tunable low noise voltage regulation below 0.5 V is required. A conventional analog LDO, however, fails to operate at these low voltages. A

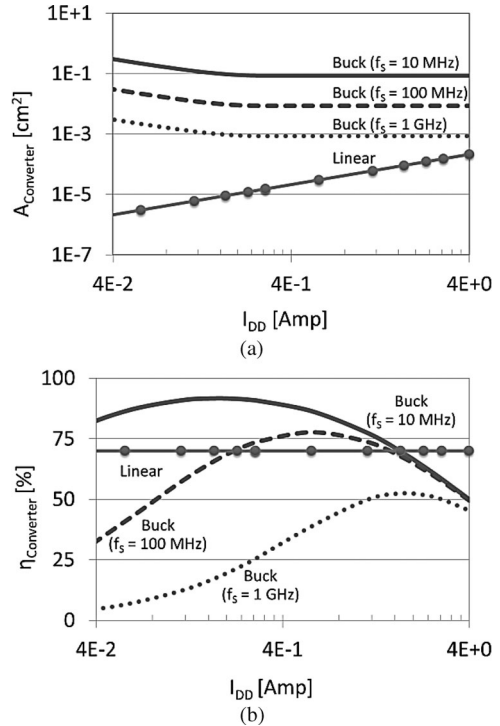


Fig. 8. LDO and buck converter (a) physical area, and (b) power efficiency for moderate, high, and ultra-high switching frequencies.

digital LDO can be used to suppress the analog nature of a conventional LDO [34], [35].

C. Comparison of Power Supply Topologies

The physical area and power efficiency of an LDO and a buck converter are shown in Fig. 8. Buck converters that are more power efficient than an alternative LDO can operate at lower switching frequencies. These buck converters are, however, inappropriate for on-chip power conversion due to the large physical size and technology constraints of the passive elements that make on-chip integration even more difficult. Alternatively, compact buck converters can operate at high switching frequencies. These buck converters, however, exhibit a lower power efficiency and are, therefore, less effective for on-chip integration. Thus, to deliver high-quality power to the load circuitry under typical area constraints, on-chip linear regulators should be considered. The moderate power efficiency of an LDO becomes a significant constraint when the power consumption at the load increases. For example, converting 2 V into 1 V while delivering $1 \mu\text{A}$ to the current load results in a 50% power efficiency and $1 \mu\text{W}$ power loss that can possibly be absorbed by the power delivery system. Alternatively, converting 1.25 V into 1 V while delivering 1 mA to the current load results in 80% power efficiency and a significant $250 \mu\text{W}$ power loss that is difficult to mitigate. Thus, linear regulators are preferable to switching power supplies, mainly for small input–output voltage differences. A heterogeneous power delivery system that efficiently exploits the power and area characteristics of linear and switching converters is desirable to enhance the power supply quality and efficiency while satisfying on-chip area constraints.

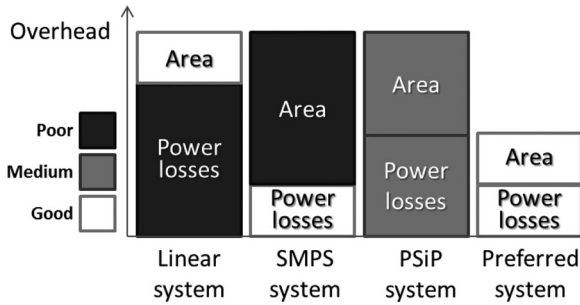


Fig. 9. Power and area overhead of a linear, SMPS, PSiP, and preferred power conversion system.

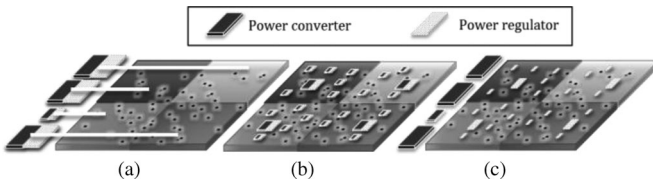


Fig. 10. Power delivery system with four voltage domains, utilizing (a) off-chip power supplies, (b) distributed POL power supplies, and (c) a heterogeneous system with off-chip converters and on-chip regulators.

III. HETEROGENEOUS POWER DELIVERY SYSTEM

Both linear and switching power regulators are characterized by an undesirable power-area tradeoff, exhibiting either high power in compact linear regulators or a large area in power-efficient SMPS, as depicted in Fig. 9. Thus, the overhead of a power delivery system composed of only switching or linear regulators is significant. Several power delivery solutions exist that exhibit intermediate power losses and area as compared to either linear or traditional SMPS systems. For example, in a PSiP system, lower power losses as compared to a linear system, and smaller area as compared to a traditional off-chip SMPS system, are traded off for greater design complexity. A desirable power delivery system minimizes power losses while satisfying on-chip area constraints, yielding both high power efficiency and small area, as depicted in Fig. 9.

To exploit the advantages of switching and linear converters, a heterogeneous power delivery system is considered that converts the power in off-chip switching power supplies and regulates the on-chip power with compact linear power supplies, minimizing LDO voltage drops and on-chip power losses. In a heterogeneous power delivery system, the area overhead is primarily constrained by the compact LDOs that regulate the on-chip power, while the power overhead is dictated by the power-efficient switching converters. Power conversion is, therefore, decoupled from power regulation, lowering the power and area overhead of the overall power delivery system. A heterogeneous power delivery system moderates the drawbacks and exploits the advantages of the historically power efficient power supplies that both convert and regulate the power off-chip with more recent trends for area efficient distributed power supplies that both convert and regulate the power on-chip. Off-chip, on-chip distributed, and heterogeneous power delivery topologies are illustrated in Fig. 10.

Consider a heterogeneous power delivery system with N_L on-chip LDOs and N_S off-chip SMPSs that deliver power to N voltage domains $\{(V_{DD}^{(i)}, I_{DD}^{(i)})\}_{i=1}^N$ with an operating voltage $V_{DD}^{(i)}$ and current $I_{DD}^{(i)}$. To supply the required voltages, $V_{DD}^{(i)} \neq V_{DD}^{(j)} \forall i \neq j$, the number of on-chip power supplies N_L should be equal to or greater than the number of voltage domains $N \leq N_L$. Alternatively, each SMPS drives one or more LDOs, yielding the relation, $N_S \leq N_L$. The effect of the number of on-chip power regulators and off-chip power converters, and the distribution of the on-chip power supplies in a heterogeneous power delivery system is described, respectively, in Sections III-A, III-B, and III-C.

A. Number of On-Chip Power Regulators

The area of an LDO is proportional to the current load [see (16)], and the power efficiency is primarily dictated by the current load and voltage drop V_{DroP} across the power transistor within the LDO [see (17)]. Thus, a single LDO that provides a specific current and voltage to a load consumes approximately the same area and dissipates similar power as numerous LDOs providing the same total current and voltage to a load. Consider K on-chip distributed LDOs to maintain a regulated voltage V_{DD} and load current I_{DD} within a specific voltage domain (V_{DD}, I_{DD}) . Let I_i ($i = 1, \dots, K$) be a local current load supplied by a single LDO within the domain, such that $\sum I_i = I_{DD}$. The LDO area A_i is linearly proportional to the supply current I_i [see (16)], $A_i = \alpha I_i$. The K LDOs form a distributed on-chip power regulation system with a total size, $A \equiv \sum A_i = \alpha \sum I_i = \alpha I_{DD}$. Thus, the total area of the distributed regulation system does not depend on K , the number of LDOs. To maximize the power efficiency of a system, all of the LDOs operate at the minimum voltage drop V_{DroP} , exhibiting a total power loss $V_{DroP} \cdot \sum I_i = V_{DroP} \cdot I_{DD}$ which is independent of K . Alternatively, the distance between an LDO and a current load is reduced at higher values of K , decreasing the on-chip voltage drops and increasing the quality of the supplied power.

B. Number of Off-Chip Power Converters

Intuitively, the number of off-chip voltage levels increases with the larger number of off-chip converters, increasing the granularity of the voltage levels supplied to the on-chip regulators and lowering the voltage drop across the hundreds of ultra-small regulators distributed on-chip. To minimize the voltage drop across an on-chip linear regulator, each off-chip SMPS converter should drive a single on-chip LDO. In practice, however, the number of power converters that can be placed off-chip is limited. Thus, each off-chip SMPS supplies power to several on-chip LDOs within an SMPS cluster. As a result, the voltage drop across the on-chip regulators is greater, degrading the overall power efficiency of the system. The upper and lower bounds of the power efficiency of a heterogeneous system for a specific number of SMPS are described in this section.

Given N voltage domains $\{(V_{DD}^{(i)}, I_{DD}^{(i)})\}_{i=1}^N$ sorted by the supply voltages $V_{DD}^{(i)} < V_{DD}^{(j)} \forall i < j$, $N_L = K \cdot N$ linear power supplies should be distributed on-chip to deliver high quality

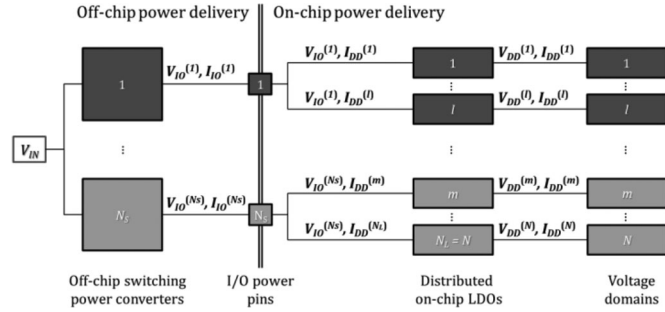


Fig. 11. Model of heterogeneous power delivery system with N_S off-chip switching converters, N_L on-chip linear regulators, and N on-chip voltage domains.

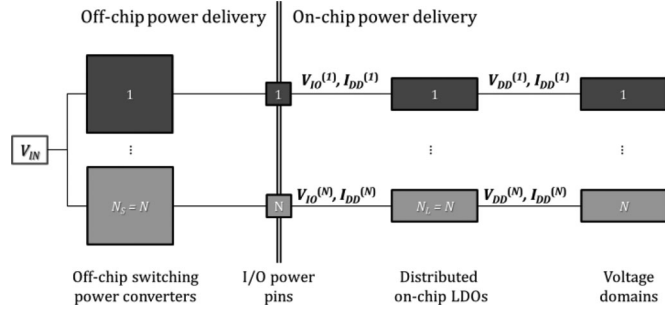


Fig. 12. Heterogeneous power delivery system with an equal number of off-chip switching converters, on-chip linear regulators, and on-chip voltage domains ($N_S = N_L = N$).

power to the load circuitry. To explore the area–power efficiency tradeoff in a heterogeneous power delivery system, a single linear regulator is assumed capable of providing sufficient high-quality current within a voltage domain, yielding $K = 1$ and $N_L = N$. The voltage supplied by an LDO to a voltage domain cannot be stepped up by an LDO. The output voltage of each SMPS is, therefore, higher than the voltage within the individual voltage domains, increasing the voltage drop across the LDOs within an SMPS cluster, degrading power efficiency.

An expression for determining the optimal LDO clustering within the SMPS clusters is presented later. Consider N_S switching power supplies to convert the off-chip input voltage V_{IN} feeding N_S voltage and current levels $\{(V_{IO}^{(i)}, I_{IO}^{(i)})\}_{i=1}^{N_S}$ into the input/output (I/O) power pins, as shown in Fig. 11.

To increase the power efficiency of a heterogeneous power delivery system, the voltage drops across the distributed on-chip LDOs should be reduced. The granularity of the converted voltage levels supplied on-chip increases with additional off-chip SMPS converters, reducing the power losses within the on-chip LDOs. At the limit, $N_S = N_L$ switching power converters are placed off-chip, providing voltages $\{V_{IO}^{(i)}\}_{i=1}^N$ at the I/O power pins, as shown in Fig. 12. In the configuration shown in Fig. 12, the on-chip LDOs operate with a minimum output voltage drop V_T , yielding

$$V_{IO}^{(i)} = V_{DD}^{(i)} + V_T, i = 1, \dots, N, \quad (18)$$

where V_T is the voltage threshold of the output transistor within the LDO. Assuming ideal power efficiency of the off-chip

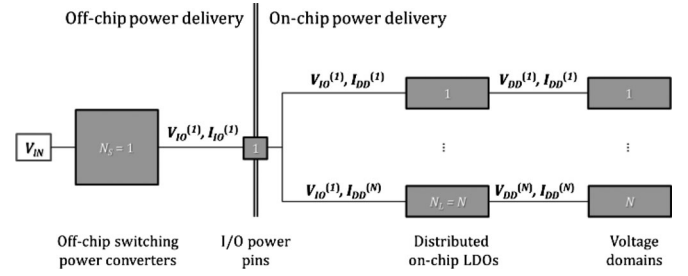


Fig. 13. Heterogeneous power delivery system with a single off-chip switching converter, and an equal number of on-chip linear regulators and on-chip voltage domains ($N_S = 1, N_L = N$).

SMPS, the power efficiency of a system with the maximum number of SMPS converters ($N_S = N_L$) is

$$\begin{aligned} \eta_{N_S = N_L = N} &= \frac{P_{\text{Load}}}{P_{\text{IN}}} = \frac{\sum_{i=1}^N V_{\text{DD}}^{(i)} I_{\text{DD}}^{(i)}}{\sum_{i=1}^{N_S} V_{\text{IO}}^{(i)} I_{\text{IO}}^{(i)}} \\ &= \frac{\sum_{i=1}^N V_{\text{DD}}^{(i)} I_{\text{DD}}^{(i)}}{\sum_{i=1}^N (V_{\text{DD}}^{(i)} + V_T) I_{\text{DD}}^{(i)}}. \end{aligned} \quad (19)$$

In this case, the power efficiency is only limited by the threshold voltage of the transistor, and exhibits a high power efficiency for low V_T devices.

Area and I/O power pin constraints exist, however, that limit the number of off-chip power supplies, degrading the overall power efficiency. Let $N_{S, \text{MAX}}$ be the maximum number of off-chip switching power converters in a heterogeneous power delivery system. The worst case power efficiency scenario where $N_{S, \text{MAX}} = 1$ is illustrated in Fig. 13.

To minimize the voltage drop across the on-chip LDOs for $N_S = 1$, the off-chip SMPS produces a voltage $V_{IO}^{(1)}$ that is higher than the maximum domain voltage by one threshold voltage V_T

$$V_{IO}^{(1)} = \max \left\{ V_{\text{DD}}^{(i)} \right\}_{i=1}^N + V_T, \quad (20)$$

exhibiting a power efficiency

$$\begin{aligned} \eta_{N_L = N, N_S = 1} &= \frac{P_{\text{Load}}}{P_{\text{IN}}} = \frac{\sum_{i=1}^N V_{\text{DD}}^{(i)} I_{\text{DD}}^{(i)}}{V_{\text{IO}}^{(1)} I_{\text{IO}}^{(1)}} \\ &= \frac{\sum_{i=1}^N V_{\text{DD}}^{(i)} I_{\text{DD}}^{(i)}}{\max_i \left\{ V_{\text{DD}}^{(i)} + V_T \right\} \sum_{i=1}^N I_{\text{DD}}^{(i)}}. \end{aligned} \quad (21)$$

In a system with a single off-chip SMPS, the power loss within each domain, in addition to the V_T drop, is determined by the difference between the domain voltage and maximum voltage in the system. Those voltage domains with lower voltages exhibit greater power losses, significantly degrading the power efficiency of a heterogeneous system. The upper and lower bounds of the power efficiency of a heterogeneous system under the $N_S \leq N_{S, \text{MAX}}$ constraint are given, respectively, by (19) and

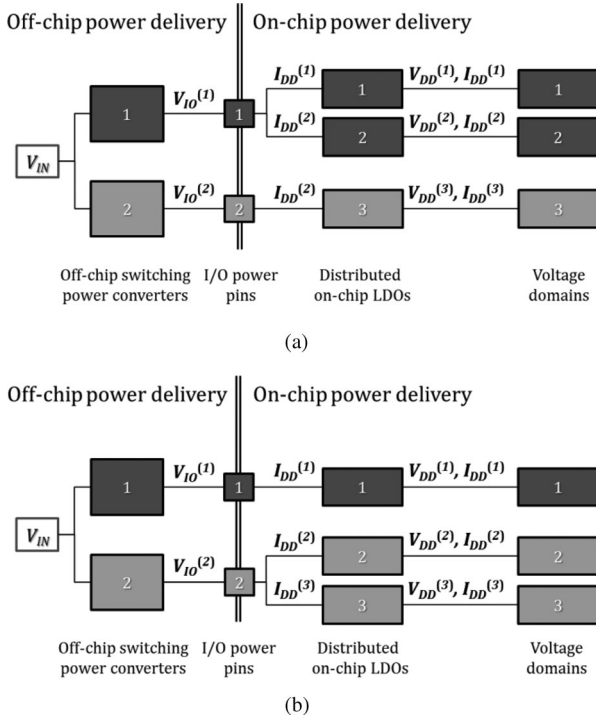


Fig. 14. Power supply clusterings for a heterogeneous power delivery system with $N_S = 2$, $N_L = N = 3$: (a) $\{K_0 = 0, K_1 = 2, K_2 = 3\}$ and (b) $\{K_0 = 0, K_1 = 1, K_2 = 3\}$.

(21), yielding

$$\begin{aligned} \frac{\sum_{i=1}^N V_{DD}^{(i)} I_{DD}^{(i)}}{\max_i \left\{ V_{DD}^{(i)} + V_T \right\} \sum_{i=1}^N I_{DD}^{(i)}} &\leq \eta_{N_S, \text{MAX}} \\ &\leq \frac{\sum_{i=1}^N V_{DD}^{(i)} I_{DD}^{(i)}}{\sum_{i=1}^N \left(V_{DD}^{(i)} + V_T \right) I_{DD}^{(i)}}. \end{aligned} \quad (22)$$

Thus, the power efficiency of a heterogeneous system is a strong function of the number of off-chip power converters.

C. Power Supply Clusters

In a practical heterogeneous power delivery system, the number of off-chip SMPS converters is smaller than the number of on-chip LDO regulators ($N_{S, \text{MAX}} < N_L$). Thus, several options exist to distribute the on-chip LDOs within SMPS clusters. Two possible clusterings are illustrated in Fig. 14 for a heterogeneous system with two SMPS and three LDOs.

The power efficiency of a general heterogeneous power delivery system, as illustrated in Fig. 11, under the $N_S \leq N_{S, \text{MAX}}$ constraint is

$$\begin{aligned} \eta_{N_S, \text{MAX}} &= \frac{P_{\text{Load}}}{P_{\text{IN}}} = \frac{\sum_{i=1}^N V_{DD}^{(i)} I_{DD}^{(i)}}{\sum_{i=1}^{N_{S, \text{MAX}}} V_{IO}^{(i)} I_{IO}^{(i)}} \\ &= \frac{\sum_{i=1}^N V_{DD}^{(i)} I_{DD}^{(i)}}{\sum_{i=1}^{N_{S, \text{MAX}}} V_{IO}^{(i)} \left(\sum_{j=K_{i-1}+1}^{K_i} I_{DD}^{(j)} \right)}, \end{aligned} \quad (23)$$

where $\{K_i\}_{i=1}^{N_{S, \text{MAX}}}$ is the power supply clustering, $K_i - K_{i-1}$ is the number of LDO regulators driven by the i th SMPS converter, and $K_0 = 0$. For example, the power supply clustering in the heterogeneous power delivery system shown in Fig. 14 can be described by $\{K_0 = 0, K_1 = 2, K_2 = 3\}$ [see Fig. 14(a)] and $\{K_0 = 0, K_1 = 1, K_2 = 3\}$ [see Fig. 14(b)]. In the configuration shown in Fig. 14(a), the first SMPS cluster contains two LDOs ($K_1 - K_0 = 2$) that regulate voltage domains 1 and 2, and the second SMPS cluster contains an additional single LDO ($K_2 - K_1 = 1$) that regulates the third voltage domain. Alternatively, in the configuration shown in Fig. 14(b), the first SMPS cluster contains a single LDO ($K_1 - K_0 = 1$), while the other two LDOs ($K_2 - K_1 = 2$) are distributed into the second SMPS cluster. To maximize the power efficiency $\eta_{N_S, \text{MAX}}$ of the proposed heterogeneous power system under the $N_S \leq N_{S, \text{MAX}}$ constraint, the input voltage for each SMPS cluster $V_{IO}^{(i)}$ that minimizes the voltage drops across the LDOs within that cluster is

$$V_{IO}^{(i)} = \max \left\{ V_{DD}^{(j)} \right\}_{j=K_{i-1}+1}^{K_i} + V_T, i = 1, \dots, N_{S, \text{MAX}}. \quad (24)$$

The minimum power efficiency of a heterogeneous system with distributed power supplies $\{K_i\}$

$$\begin{aligned} \eta_{N_S, \text{MAX}} &= \\ &= \frac{\sum_{i=1}^N V_{DD}^{(i)} I_{DD}^{(i)}}{\sum_{i=1}^{N_{S, \text{MAX}}} \left(\max \left\{ V_{DD}^{(j)} \right\}_{j=K_{i-1}+1}^{K_i} + V_T \right) \cdot \sum_{j=K_{i-1}+1}^{K_i} I_{DD}^{(j)}}, \end{aligned} \quad (25)$$

is strongly dependent on the power supply clustering $\{K_i\}_{i=1}^{N_{S, \text{MAX}}}$.

The effect of the power supply clustering on a heterogeneous power delivery system is illustrated in Fig. 14(a) and (b) for different power efficiencies yielding, respectively, (26) and (27) as shown at the bottom of the page.

$$\eta_{N_S, \text{MAX}}^{(a)} = \frac{V_{DD}^{(1)} I_{DD}^{(1)} + V_{DD}^{(2)} I_{DD}^{(2)} + V_{DD}^{(3)} I_{DD}^{(3)}}{\left(\max \left\{ V_{DD}^{(1)}, V_{DD}^{(2)} \right\} + V_T \right) \cdot \left(I_{DD}^{(1)} + I_{DD}^{(2)} \right) + \left(V_{DD}^{(3)} + V_T \right) \cdot I_{DD}^{(3)}} \quad (26)$$

$$\eta_{N_S, \text{MAX}}^{(b)} = \frac{V_{DD}^{(1)} I_{DD}^{(1)} + V_{DD}^{(2)} I_{DD}^{(2)} + V_{DD}^{(3)} I_{DD}^{(3)}}{\left(V_{DD}^{(1)} + V_T \right) \cdot I_{DD}^{(1)} + \left(\max \left\{ V_{DD}^{(2)}, V_{DD}^{(3)} \right\} + V_T \right) \cdot \left(I_{DD}^{(2)} + I_{DD}^{(3)} \right)} \neq \eta_{N_S, \text{MAX}}^{(a)} \quad (27)$$

For each SMPS converter, the voltage drop across the driven LDOs increases with a wider range of voltages included within that SMPS cluster, increasing the overall power dissipation. Intuitively, for any power supply clustering, adding a voltage domain with a specific voltage in an SMPS cluster that includes a similar voltage range results in a lower voltage drop and power loss than including the same voltage domain in an SMPS cluster with a significantly different range of voltages. Thus, the choice of power clustering directly affects the efficiency of the power delivery system. To minimize power losses in a heterogeneous power delivery system, a power distribution network with a higher $\eta_{N_S, \text{MAX}}$ is preferred.

The power efficiency of a heterogeneous system is also a strong function of the current distribution, which is not necessarily equally distributed to the individual voltage domains. Optimizing the power efficiency of a heterogeneous system based on the current distribution within the voltage domains requires additional assumptions regarding the behavior and specifications of the currents. The purpose here is to provide a framework for a power delivery methodology and specific rules for efficiently delivering power.

IV. ALGORITHMS FOR ENERGY-EFFICIENT POWER SUPPLY CLUSTERING

The power efficiency of a heterogeneous power delivery system depends upon the distribution of the power supply resources. Given a power supply system with N voltage domains and a limited number of off-chip switching power converters $N_{S, \text{MAX}}$, the clustering of the $N_L \geq N_{S, \text{MAX}}$ on-chip linear regulators into $N_{S, \text{MAX}}$ SMPS clusters $K^{\text{OPT}} = \{K_i\}_{i=1}^{N_{S, \text{MAX}}}$ that minimizes power losses should be determined. The optimal solution with minimum power losses can be obtained by exhaustively comparing the power efficiency $\eta_{N_S, \text{MAX}}$ [see (25)] for all possible clusterings, and choosing the configuration with the maximum efficiency $\eta_{N_S, \text{MAX}}^{\text{OPT}}$

$$\eta_{N_S, \text{MAX}}^{\text{OPT}} = \max_{\substack{\text{all}\{K_i\} \\ \text{distributions}}} \{ \eta_{N_S, \text{MAX}} \} = \max_{\substack{\text{all}\{K_i\} \\ \text{distributions}}} \left\{ \frac{\sum_{i=1}^N V_{\text{DD}}^{(i)} I_{\text{DD}}^{(i)}}{\sum_{i=1}^{N_{S, \text{MAX}}} \left(\max_{j=K_{i-1}+1}^{K_i} \{V_{\text{DD}}^{(j)}\} + V_T \right) \cdot \sum_{j=K_{i-1}+1}^{K_i} I_{\text{DD}}^{(j)}} \right\}. \quad (28)$$

The number of possible clusterings $\{K_i\}$, however, grows exponentially with $N_{S, \text{MAX}}$, producing a computationally infeasible solution. To efficiently determine the preferable power supply clusters, alternative computationally efficient solutions are required. Binary and linear near-optimal power supply clusterings are described, respectively, in Sections IV-A and IV-B.

A. Binary Power Supply Clustering

Intuitively, to reduce the voltage drop across the on-chip LDOs, LDOs that regulate the voltage domains with a small difference in voltage levels should be assembled into a voltage

List_of_Clusters = **binary_power_supply_clustering** (sorted supply voltages $\{V_{\text{DD}}^{(i)}\}_{i=1..N}$)

1. *Next_Cluster_to_Distribute* = $(V_{\text{DD}}^{(i)})_{i=1..N}$
2. *List_of_Clusters* = $\{ \text{Next_Cluster_to_Distribute} \}$
3. (*New_Low_Cluster*, *New_High_Cluster*) = **distribute_a_cluster**(*Next_Cluster_to_Distribute*)
4. *List_of_Clusters* += *New_Low_Cluster* + *New_High_Cluster* – *Next_Cluster_to_Distribute*
5. If number of clusters in *List_of_Clusters* < $N_{S, \text{MAX}}$
 - 5.1 Find *Cluster* in *List_of_Clusters* such that $(\max\{\text{Cluster}\} - \min\{\text{Cluster}\})$ is maximal
 - 5.2 *Next_Cluster_to_Distribute* = *Cluster*
 - 5.3 Return to 3.

(*New_Low_Cluster*, *New_High_Cluster*) = **distribute_a_cluster**(*Next_Cluster_to_Distribute*)

1. $V_{\text{Mean}} = \frac{1}{2} (\min\{\text{Next_Cluster_to_Distribute}\} + \max\{\text{Next_Cluster_to_Distribute}\})$
2. *New_Low_Cluster* = $\{V_{\text{DD}}^{(i)} \in \text{Next_Cluster_to_Distribute} \mid V_{\text{DD}}^{(i)} \leq V_{\text{Mean}}\}$
3. *New_High_Cluster* = $\{V_{\text{DD}}^{(i)} \in \text{Next_Cluster_to_Distribute} \mid V_{\text{DD}}^{(i)} > V_{\text{Mean}}\}$

Fig. 15. Algorithm for a binary power supply clustering.

TABLE I
POWER SUPPLY CLUSTERING FOR A HETEROGENEOUS POWER DELIVERY SYSTEM WITH $N_S = 3$, $N_L = N = 4$, AND V DOMAINS (1 V, 1 A), (1.49 V, 1 A), (1.51 V, 1 A) AND (2 V, 1 A). (A) $\{K_0 = 0, K_1 = 1, K_2 = 3, K_3 = 4\}$ AND (B) $\{K_0 = 0, K_1 = 2, K_2 = 3, K_3 = 4\}$

| | Power supply clustering $\{K_0, K_1, K_2, K_3\}$ | SMPS output voltages $\{V_{\text{IO}}^{(1)}, V_{\text{IO}}^{(2)}, V_{\text{IO}}^{(3)}\}$ [Volts] | Power efficiency [%] |
|---------------------------------|--|--|----------------------|
| Binary power supply clustering | {0, 1, 3, 4} | $\{1 + V_T, 1.51 + V_T, 2 + V_T\}$ | 91 |
| Optimal power supply clustering | {0, 2, 3, 4} | $\{1.49 + V_T, 1.51 + V_T, 2 + V_T\}$ | 85 |

cluster driven by the same SMPS, minimizing the voltage range within each cluster. A binary power supply clustering, based on a greedy algorithm, identifies in each step the voltage cluster with the widest voltage range and distributes the LDOs into two separate clusters. Pseudocode of the algorithm is provided in Fig. 15.

The algorithm produces a set of $N_{S, \text{MAX}}$ SMPS voltage clusters *List_of_Clusters* with a binary clustering of power supplies. The third step is executed $N_{S, \text{MAX}}$ times, yielding an algorithm that exhibits linear complexity $O(N_{S, \text{MAX}})$ with the number of switching converters.

B. Linear Power Supply Clustering

The primary weakness of the binary power supply clustering is the greedy nature of the algorithm. The number of voltage clusters $N_{S, \text{MAX}}$ is only considered when the algorithm is terminated, reducing the power efficiency of the overall power delivery system. Consider a heterogeneous power delivery system with three switching converters and four LDO regulators that supply power to four voltage domains. The voltage and current levels within the voltage domains are (1 V, 1 A), (1.49 V, 1 A), (1.51 V, 1 A), and (2 V, 1 A). The optimal and binary power supply clusterings, SMPS output voltages, and power efficiency are summarized in Table I, exhibiting, respectively, 91% and 85% power efficiency for $V_T = 0.2$ V [from (25)].

Alternatively, a linear power supply clustering produces a topology by linearly distributing the LDOs within $N_{S, \text{MAX}}$

```

List_of_Clusters = linear_power_supply_clustering (sorted supply voltages  $\{V_{DD}^{(i)}\}_{i=1..N}$ )
1. List_of_Clusters =  $\{()\}_{i=1..N}$  %  $N_{S,MAX}$  empty clusters
2. Cluster_Range =  $(\max\{V_{DD}^{(i)}\}_{i=1..N} - \min\{V_{DD}^{(i)}\}_{i=1..N}) / N_{S,MAX}$ 
3. For each  $V_{DD} \in \{V_{DD}^{(i)}\}_{i=1..N}$ 
    3.1  $k = \lfloor (V_{DD} - \min\{V_{DD}^{(i)}\}_{i=1..N}) / \text{Cluster\_Range} \rfloor + 1$ 
    3.2 Add  $V_{DD}$  to the  $k^{\text{th}}$  cluster in List_of_Clusters
4. If number of non-empty clusters in List_of_Clusters <  $N_{S,MAX}$ 
    4.1 Find Cluster in List_of_Clusters such that  $(\max\{\text{Cluster}\} - \min\{\text{Cluster}\})$  is maximum
    4.2 Next_Cluster_to_Distribute = Cluster
    4.3  $(\text{New\_Low\_Cluster}, \text{New\_High\_Cluster}) = \text{distribute\_a\_cluster}(\text{Next\_Cluster\_to\_Distribute})$ 
    4.4 List_of_Clusters += New_Low_Cluster + New_High_Cluster - Next_Cluster_to_Distribute
    4.5 Return to 4.

```

Fig. 16. Algorithm for a linear power supply clustering.

voltage clusters, as described by the algorithm represented by the pseudocode provided in Fig. 16.

If less than $N_{S,MAX}$ SMPS voltage clusters are produced within steps 1 through 3 in the linear power supply clustering algorithm, the linearly generated clusters are distributed into additional clusters using a binary algorithm. This algorithm produces a set of $N_{S,MAX}$ SMPS voltage clusters *List_of_Clusters* with a linear power supply clustering. In the worst case, the third and fourth steps are executed, respectively, N and $N_{S,MAX}$ times, yielding an algorithm complexity that is linear with the number of voltage domains, $O(N)$.

In modern ICs, advanced power techniques, such as DVS and DVFS, are often employed [36]. To apply the proposed binary or linear algorithm in a heterogeneous system with dynamically changing voltage levels, the average voltage level for each voltage domain is used. Once the power supply clusters are determined with either the binary or linear algorithm based on average domain voltage levels, the maximum voltage level within each SMPS cluster determines the SMPS output voltage. The drop in efficiency of the proposed algorithms is likely to be higher in a power system with dynamically changing voltage levels as compared to a power system with fixed operating conditions.

V. SIMULATION RESULTS

The optimal exhaustive power delivery network and the near-optimal solutions described in Section IV have been implemented in MATLAB. To compare the power efficiency of the near-optimal and optimal power delivery networks, a heterogeneous power delivery system with a small number of voltage domains is considered in Section V-A due to the computational complexity of the exhaustive optimal algorithm. To evaluate the power efficiency of the proposed linear, binary, and hybrid clusterings, heterogeneous power delivery systems with a larger number of voltage domains are considered in Section V-B

A. Power Efficiency in Optimal and Near-Optimal Power Delivery Networks

The exhaustive algorithm determines the most power efficient clustering by comparing the power efficiency of all the possible clusterings. The efficiency of the optimal power net-

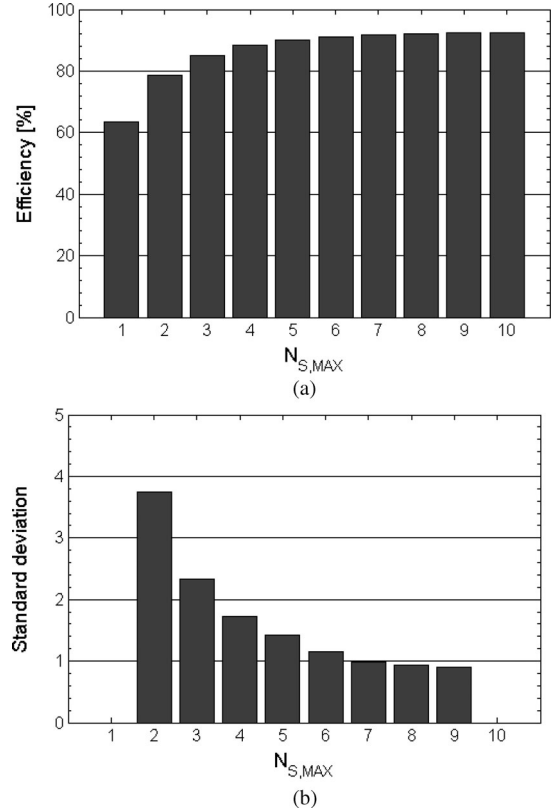


Fig. 17. Heterogeneous power delivery system (a) average efficiency, and (b) standard deviation using an exhaustive power supply clustering algorithm.

work produced by the exhaustive algorithm is compared in this section to the power efficiency of the near-optimal clustering algorithms. To estimate the power efficiency of the optimal power supply clusters, a heterogeneous power delivery system S_1 with ten voltage domains ($N = 10$) and ten on-chip linear regulators ($N_L = 10$) is considered. The maximum number of off-chip switching converters is evaluated for one to ten converters ($1 \leq N_{S,MAX} \leq 10$). A voltage threshold of $V_T = 0.1$ V, and domain voltages and currents of, respectively, 0.5 to 2 V and 0.5 to 3.5 A, are considered. Simulation results are sampled for 100 iterations. The power efficiency of a heterogeneous power delivery system with the power supply clusters, determined by an exhaustive analysis, is presented in Fig. 17(a). A power efficiency above 80% is demonstrated for $N_{S,MAX} \geq 2$, and a maximum 93% power efficiency is achieved for $N_{S,MAX} = N$. Thus, the power efficiency of a heterogeneous power delivery system with an optimal power clustering exhibits a reasonable power efficiency of 80%, using only two off-chip switching converters. The efficiency increases rapidly with additional off-chip converters.

Based on the Monte Carlo integration technique [37], the average error in the efficiency is bounded by σ_M / \sqrt{M} , where σ_M is the standard deviation of a power efficiency sample and M is the number of samples. The standard deviation of the power efficiency is shown in Fig. 17(b) for $2 \leq N_{S,MAX} \leq 9$. Values of σ_M range from 3.7 for $N_{S,MAX} = 2$ to 0.9 for $N_{S,MAX} = 9$, bounding the power efficiency error for $M = 100$

by, respectively, 0.37% to 0.09%. Power supply clustering for $N_{S,MAX} = 1$ and $N_{S,MAX} = N$ is explicit, yielding no error in the power efficiency.

To evaluate the power efficiency of the near-optimal power supply clustering topologies described in Section IV, algorithms for binary and linear power supply clusterings have also been implemented in MATLAB. The same heterogeneous system S_1 is considered for both linear and binary distributed power supplies. For a heterogeneous system with a single off-chip SMPS converter ($N_{S,MAX} = 1$) or the maximum number of off-chip SMPS converters ($N_{S,MAX} = N_L$), the linear, binary, and optimal clusterings of the on-chip LDO regulators are identical. For $N_{S,MAX} = 1$, all of the LDOs are driven by a single SMPS converter, while for $N_{S,MAX} = N_L$, each LDO is driven by a different SMPS converter. Thus, the power efficiency of a heterogeneous system with $N_{S,MAX} = 1$ or $N_{S,MAX} = N$ is optimal with either the linear or binary power supply clustering. Alternatively for $N_{S,MAX} < N_L$, the linear and binary clusterings of the power supplies may differ from the exhaustive optimal solution, exhibiting a lower than optimal power efficiency. Due to the uniform nature of the linear approach, the linear clustering of the on-chip LDO regulators within the off-chip SMPS converters exhibits near-optimal efficiency for power delivery systems with near uniformly distributed domain voltages. Alternatively, for a power delivery system with domain voltages that exhibit significant deviation from a uniform distribution, the power efficiency with the binary power supply clustering may be higher than with the linear clustering. This behavior is due to the greedy nature of the binary approach that iteratively identifies the on-chip power supply cluster with the lowest power efficiency and splits the cluster, increasing the overall efficiency of the system. To demonstrate the power efficiency of the binary and linear clusterings, the reduction in efficiency with both the binary and linear power supply clusterings is simulated for two different power profiles, exhibiting a maximum 4% drop in power efficiency. The optimal solution with zero reduction in power efficiency is demonstrated for both power profiles in Fig. 18 for $N_{S,MAX} = 1$ and $N_{S,MAX} = N_L$. In the first power profile, the voltage levels are assumed to be randomly distributed between 0.5 and 2 V, yielding an average power efficiency generated from over 100 iterations, as depicted in Fig. 18(a). In this case, for $1 < N_{S,MAX} < N_L$, the exhaustive optimal solution produces a power supply that is uniformly distributed, and the linear power supply clustering yields a higher power efficiency.

In the second power profile, the voltage levels are assumed to be normally distributed within each of the [0.5, 1.5), [1.5, 1.8), and [1.8, 2] ranges, prioritizing the mean value of the groups. Due to the nonuniform clustered nature of the voltage domain profile, for a heterogeneous system with three off-chip SMPS converters, intuitively, the on-chip LDO regulators should be nonuniformly distributed into three clusters covering the ranges [0.5, 1.5), [1.5, 1.8), and [1.8, 2]. In this case, a system with uniformly distributed clusters with voltage ranges [0.5, 1), [1, 1.5), [1.5, 2) is less power efficient. This heterogeneous system is, therefore, more suitable for a binary power supply clustering rather than a linear power supply clustering. The average power efficiency for the second power profile, generated from over 100

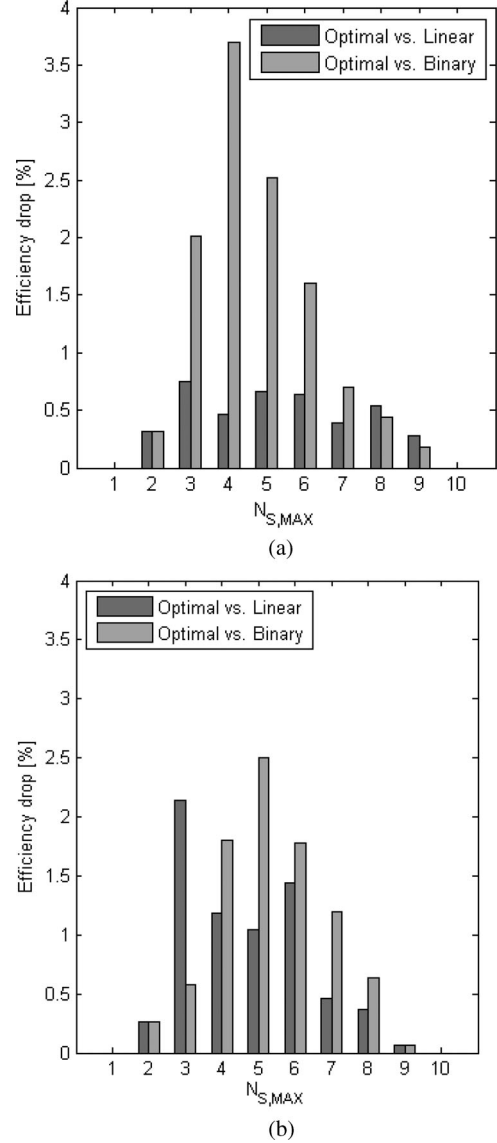


Fig. 18. Decrease in linear and binary power efficiency from the optimal power efficiency for (a) randomly distributed voltage levels, and (b) voltage levels grouped within three voltage ranges.

iterations, is depicted in Fig. 18(b). In this case, specifically for $N_{S,MAX} = 3$, the optimal solution produces three nonuniform SMPS clusters, covering the three ranges, [0.5, 1.5), [1.5, 1.8), and [1.8, 2]. The binary power supply clustering with $N_{S,MAX} = 3$ also produces three SMPS clusters with voltage ranges, [0.5, 1.25), [1.25, 1.625), and [1.625, 2], exhibiting a higher power efficiency than the efficiency of the linear power supply clustering. Based on a Monte Carlo integration technique, the error in estimating the drop in power efficiency, illustrated in Fig. 18, is smaller than 0.63% for all values of $N_{S,MAX}$.

Due to the greedy nature of the binary power supply clustering, the binary algorithm is better for those voltage domain levels grouped near specific voltage levels. Alternatively, the number of SMPS clusters $N_{S,MAX}$ is only considered at the termination of the binary algorithm, potentially reducing the effectiveness of the binary clustering algorithm in those systems with

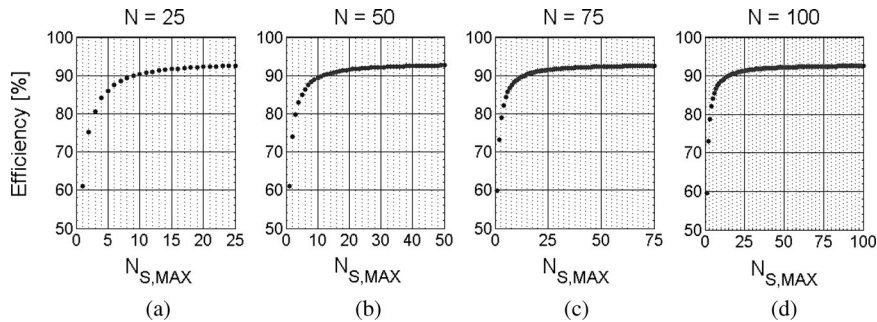


Fig. 19. Linear and binary power efficiency for (a) 25, (b) 50, (c) 75, and (d) 100 V domains.

uniformly distributed voltage domains. As expected, for most values of $N_{S,MAX}$ and power supply specifications, the drop in power efficiency for the linear power supply clustering algorithm is lower than with the binary approach. However, the second power profile that forms three nonuniform voltage groups is better addressed by the binary power supply clustering algorithm, producing a more efficient heterogeneous power delivery system for $N_{S,MAX} = 3$. Thus, a heterogeneous power delivery system with a higher power efficiency is usually produced with a linear power supply clustering. However, for certain power profiles, a binary power supply clustering is preferable.

To increase the power efficiency of a heterogeneous power delivery system, a combined hybrid approach should be employed. The power efficiency should be evaluated with both the binary and linear algorithms, and the configuration with the higher power efficiency should be employed. Analyzing the results depicted in Fig. 18 based on this combined hybrid approach, the drop in power efficiency from the optimal solution is reduced to 1.5%, yielding a computationally efficient, $O(N_{S,MAX} + N)$ complexity, near-optimal, and high fidelity power supply clustering.

B. Power Efficiency With Binary, Linear, and Hybrid Clusterings

The power efficiency of a heterogeneous power delivery system S_2 with 25, 50, 75, and 100 V domains is presented in Fig. 19, exhibiting a maximum power efficiency of 93% for $N_{S,MAX} = N$. A reasonable on-chip power efficiency of 79% is, therefore, achievable using only a small number ($N_{S,MAX} > 2$) of switching converters when the on-chip power supplies are distributed using a combined hybrid, binary, and linear clustering algorithm.

The power efficiency exhibits a similar behavior for 25, 50, 75, and 100 V domains, as shown in Fig. 19. For a specific number of voltage domains N and on-chip LDO regulators ($N_L = N$), the number of LDOs within each SMPS cluster decreases with a larger number of off-chip SMPS converters ($1 \leq N_{S,MAX} \leq N$). As a result, the maximum voltage drop across the on-chip LDOs is less, decreasing the losses within the power delivery system.

The power efficiency, illustrated in Fig. 19, increases rapidly with a larger number of off-chip converters and saturates for $N_{S,MAX} > \frac{1}{2}N$ for 25, 50, 75, and 100 V domains. To avoid the redundancy of the off-chip power supplies, the power efficiency

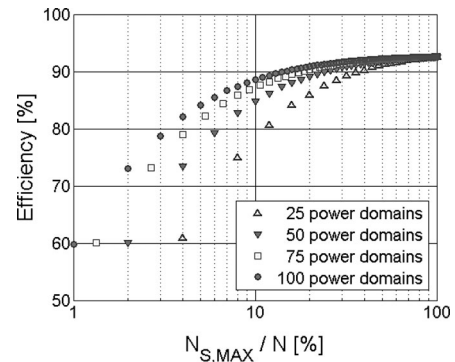


Fig. 20. Linear and binary power efficiency versus the $N_{S,MAX}/N$ ratio for 25, 50, 75, and 100 V domains.

trend as a function of the $N_{S,MAX}/N$ ratio, shown in Fig. 20, considers a heterogeneous system with 25, 50, 75, and 100 V domains. Targeting a specific ratio between the number of off-chip power converters and on-chip voltage domains, the overall power efficiency of the system increases with a larger number of voltage domains. For example, for a heterogeneous power system with 25 V domains and a limited number of off-chip SMPS converters ($N_{S,MAX} = 8\% \cdot N$), a moderate power efficiency of 75% is noted. Alternatively, in a power delivery system with 100 V domains and the same $N_{S,MAX}/N$ ratio, a higher power efficiency of 84% is observed. The efficiency of a heterogeneous system with a different number of voltage domains decreases with higher $N_{S,MAX}/N$ ratios, becoming insignificant for $N_{S,MAX}/N > 50\%$. The current load at the output of an LDO is assumed to be the total current consumed by the voltage domain, and is, therefore, significantly greater than the quiescent current of an LDO. Alternatively, by clustering hundreds of ultra-small power regulators, lower LDO output currents are expected. As a result, the effect of the quiescent current on the LDO power efficiency increases, degrading the maximum efficiency of the overall heterogeneous power delivery system. In addition, the drop in efficiency of the hybrid, binary, and linear clustering algorithms increases in heterogeneous systems with dynamically changing voltage and current specifications, lowering the overall power efficiency of the system. Thus, a heterogeneous power delivery system with a more complex distributed power supply system and accurate specifications should be considered as future work in the development of a robust power delivery methodology.

VI. SUMMARY

On-chip power integration is necessary for delivering high-quality power to modern high performance circuits. With on-chip power supplies, new design challenges have arisen that require advanced circuit design solutions. A power delivery topology is, therefore, required that minimizes power conversion and regulation losses while satisfying specific design constraints. Accurate, computationally efficient methods to distribute on-chip power supplies are essential.

The tradeoff between power efficiency and area for switching and linear power supplies is discussed in this paper. To convert power with minimum power losses while avoiding area consuming on-chip passive components, power efficient SMPSs should be placed off-chip. In addition, area efficient LDOs should be employed on-chip to regulate and deliver the converted power to the load circuitry, reducing power losses from the low voltages dropped across the LDOs. Thus, to maintain high quality on-chip power delivery, the power conversion and regulation operations should be decoupled. Based on this decoupling principle of power conversion and regulation, a heterogeneous power delivery system is proposed in this paper. To optimize a specific heterogeneous system given the number of voltage domains and off-chip SMPS, clustering of the on-chip LDOs within the SMPS clusters should maximize the power efficiency of the overall system.

An exhaustive solution that produces the on-chip power supply clusterings with the highest power efficiency is, however, computationally inefficient. Thus, computationally efficient binary and linear algorithms for determining a near-optimal heterogeneous power supply clustering are presented, exhibiting, when combined, a drop in power efficiency of less than 1.5% from the optimal solution. A power efficiency above 80% is demonstrated for power delivery systems with more than two off-chip switching converters. Power efficiency is also shown to increase rapidly with additional SMPS converters and voltage domains, saturating when the number of off-chip converters exceeds 50% of the voltage domains. A hybrid methodology for linear and binary heterogeneous power delivery should, therefore, be employed to determine the preferred number of voltage domains and the efficient clustering of the on-chip power supplies in large scale systems. A heterogeneous integrated power delivery system is shown to be a power efficient alternative to existing topologies that employ either switching or linear on-chip power supplies.

REFERENCES

- [1] E. Salman and E. G. Friedman, *High Performance Integrated Circuit Design*. New York: McGraw-Hill, 2012.
- [2] V. Kursun and E. G. Friedman, *Multi-Voltage CMOS Circuit Design*. Hoboken, NJ: Wiley, 2006.
- [3] F. Waldron, J. Slowey, A. Alderman, B. Narveson, and S. C. O'Mathuna, "Technology roadmapping for power supply in package (PSiP) and power supply on chip (PwrSoC)," in *Proc. IEEE Appl. Power Electron. Conf. Expos.*, Feb. 2010, pp. 525–532.
- [4] S. Kose, S. Pinzon, B. McDermott, S. Tam, and E. G. Friedman, "Active filter based hybrid on-chip DC–DC converters for point-of-load voltage regulation," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. PP, no. 99, pp. 1–12, Nov. 2012.
- [5] S. Kose and E. G. Friedman, "Distributed power network co-design with on-chip power supplies and decoupling capacitors," in *Proc. ACM/IEEE Int. Workshop Syst. Level Interconnect Prediction*, Jun. 2011, pp. 1–5.
- [6] S. Kose and E. G. Friedman, "Distributed on-chip power delivery," *IEEE J. Emerging Select. Topics Circuits Syst.*, vol. 2, no. 4, pp. 704–713, Dec. 2012.
- [7] S. Kose and E. G. Friedman, "Simultaneous co-design of distributed on-chip power supplies and decoupling capacitors," in *Proc. IEEE Int. SoC Conf.*, Sep. 2010, pp. 15–18.
- [8] S. Kose, S. Pinzon, B. McDermott, S. Tam, and E. G. Friedman, "An area efficient on-chip hybrid voltage regulator," in *Proc. IEEE Int. Symp. Qual. Electron. Design*, Mar. 2012, pp. 398–403.
- [9] V. Kursun, S. G. Narendra, V. K. De, and E. G. Friedman, "Analysis of buck converters for on-chip integration with a dual supply voltage micro-processor," *IEEE Trans. Very Large Scale Integr. Circuits*, vol. 11, no. 3, pp. 514–522, Jun. 2003.
- [10] V. Kursun, S. G. Narendra, V. K. De, and E. G. Friedman, "High input voltage step-down DC–DC converters for integration in a low voltage CMOS process," in *Proc. IEEE Int. Symp. Qual. Electron. Design*, Mar. 2004, pp. 517–521.
- [11] M. Al-Shyoukh, H. Lee, and R. Perez, "A transient-enhanced low-quiescent current low-dropout regulator with buffer impedance attenuation," *IEEE J. Solid-State Circuits*, vol. 42, no. 8, pp. 1732–1742, Aug. 2007.
- [12] T. Y. Man, K. N. Leung, C. Y. Leung, P. K. T. Mok, and M. Chan, "Development of single-transistor-control LDO based on flipped voltage follower for SoC," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 55, no. 5, pp. 1392–1401, Jun. 2008.
- [13] P. Hazucha, T. Karnik, B. A. Bloechel, C. Parsons, D. Finan, and S. Borkar, "Area-efficient linear regulator with ultra-fast load regulation," *IEEE J. Solid-State Circuits*, vol. 40, no. 4, pp. 933–940, Apr. 2005.
- [14] J. Guo and K. N. Leung, "A 6- μ W chip-area-efficient output-capacitorless LDO in 90-nm CMOS technology," *IEEE J. Solid-State Circuits*, vol. 45, no. 9, pp. 1896–1905, Sep. 2010.
- [15] M. Wens and M. S. J. Steyaert, "A fully integrated CMOS 800-mW four-phase semiconstant ON/OFF-time step-down converter," *IEEE Trans. Power Electron.*, vol. 26, no. 2, pp. 326–333, Feb. 2011.
- [16] H. Nam, Y. Ahn, and J. Roh, "5-V buck converter using 3.3-V standard CMOS process with adaptive power transistor driver increasing efficiency and maximum load capacity," *IEEE Trans. Power Electron.*, vol. 27, no. 1, pp. 463–471, Jan. 2012.
- [17] L. Wang, Y. Pei, X. Yang, Y. Qin, and Z. Wang, "Improving light and intermediate load efficiencies of buck converters with planar nonlinear inductors and variable on time control," *IEEE Trans. Power Electron.*, vol. 27, no. 1, pp. 342–353, Jan. 2012.
- [18] W. Yan, W. Li, and R. Liu, "A noise-shaped buck DC–DC converter with improved light-load efficiency and fast transient response," *IEEE Trans. Power Electron.*, vol. 26, no. 12, pp. 3908–3924, Dec. 2011.
- [19] H. Jia, J. Lu, X. Wang, K. Padmanabhan, and Z. J. Shen, "Integration of a monolithic buck converter power IC and bondwire inductors with ferrite epoxy glob cores," *IEEE Trans. Power Electron.*, vol. 26, no. 6, pp. 1627–1630, Jun. 2011.
- [20] Y.-H. Lee, S.-C. Huang, S.-W. Wang, W.-C. Wu, P.-C. Huang, H.-H. Ho, Y.-T. Lai, and K.-H. Chen, "Power-tracking embedded buck–boost converter with fast dynamic voltage scaling for the SoC system," *IEEE Trans. Power Electron.*, vol. 27, no. 3, pp. 1271–1282, Mar. 2012.
- [21] Y. Ahn, H. Nam, and J. Roh, "A 50-MHz fully integrated low-swing buck converter using packaging inductors," *IEEE Trans. Power Electron.*, vol. 27, no. 10, pp. 4347–4356, Oct. 2012.
- [22] M. Bathily, B. Allard, and F. Hasbani, "A 200-MHz integrated buck converter with resonant gate drivers for an RF power amplifier," *IEEE Trans. Power Electron.*, vol. 27, no. 2, pp. 610–613, Feb. 2012.
- [23] Y. Ramadass, A. Fayed, and A. Chandrakasan, "A fully-integrated switched-capacitor step-down DC–DC converter with digital capacitance modulation in 45 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 45, no. 12, pp. 2557–2565, Dec. 2010.
- [24] 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, Sep. 2011.
- [25] R. W. Erickson and D. Maksimovic, *Fundamentals of Power Electronics*. Norwell, MA: Kluwer, 2001.
- [26] C. O'Mathuna, N. Wang, S. Kulkarni, and S. Roy, "Review of integrated magnetics for power supply on chip (PwrSoC)," *IEEE Trans. Power Electron.*, vol. 27, no. 1, pp. 4799–4816, Nov. 2012.

- [27] F. Wei and A. Fayed, "A feasibility study of high-frequency buck regulators in nanometer CMOS technologies," in *Proc. IEEE Dallas Circuits Syst. Workshop*, Oct. 2009, pp. 1–4.
- [28] D. Lu and C. P. Wong, *Materials for Advanced Packaging*. New York: Springer, 2008.
- [29] Johanson Technology. Integrated Passive Components. (2012). [Online]. Available: http://www.johansontechnology.com/images/stories/catalog/JTI_CAT_2012.pdf
- [30] Vishay. Passive Components. (2009). [Online]. Available: <http://www.vishay.com/power-ics/step-down-regulators>
- [31] International Technology Roadmap for Semiconductors. (2011). [Online]. Available: www.itrs.net/Links/2011ITRS/2011Tables/PIDS_2011Tables.xlsx
- [32] S. Bhunia and S. Mukhopadhyay, *Low-Power Variation-Tolerant Design in Nanometer Silicon*. New York: Springer, 2011.
- [33] R. G. Dreslinski, M. Wieckowski, D. Blaauw, D. Sylvester, and T. Mudge, "Near-threshold computing: Reclaiming Moore's law through energy efficient integrated circuits," *Proc. IEEE*, vol. 98, no. 2, pp. 253–266, Feb. 2010.
- [34] K. Ishida, Y. Ryu, Xin Zhang, Po-Hung Chen, K. Watanabe, M. Takamiya, and T. Sakurai, "0.5-V input digital LDO with 98.7% current efficiency and 2.7- μ A quiescent current in 65 nm CMOS," in *Proc. IEEE Custom Integrated Circuits Conf.*, Sep. 2010, pp. 1–4.
- [35] K. Hirairi, Y. Okuma, H. Fuketa, T. Yasufuku, M. Takamiya, M. Nomura, H. Shinohara, and T. Sakurai, "13% power reduction in 16b integer unit in 40 nm CMOS by adaptive power supply voltage control with parity-based error prediction and detection (PEPD) and fully integrated digital LDO," in *Proc. IEEE Int. Solid-State Circuits Conf.*, Feb. 2012, pp. 486–488.
- [36] R. Urgaonkar, U. C. Kozat, K. Igarashi, and M. J. Neely, "Dynamic resource allocation and power management in virtualized data centers," in *Proc. IEEE Network Oper. Manage. Symp.*, Apr. 2010, pp. 479–486.
- [37] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York: Springer, 1999.



Eby G. Friedman (S'80–M'81–SM'90–F'00) received the B.S. degree from Lafayette College, Easton, PA, in 1979, and the M.S. and Ph.D. degrees from the University of California, Irvine, in 1981 and 1989, respectively, all in electrical engineering.

From 1979 to 1991, he was with Hughes Aircraft Company, rising to the position of manager of the Signal Processing Design and Test Department, responsible for the design and test of high-performance digital and analog ICs. He has been with the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, since 1991, where he is currently a Distinguished Professor, and the Director of the High Performance VLSI/IC Design and Analysis Laboratory. He is also a Visiting Professor with the Technion—Israel Institute of Technology, Haifa, Israel. He is the author of about 400 papers and book chapters, a dozen, and the author or editor of 15 books in the fields of high speed and low power CMOS design techniques, high speed interconnect, and the theory and application of synchronous clock and power distribution networks. His current research and teaching interests include high performance synchronous digital and mixed-signal microelectronic design and analysis with application to high speed portable processors and low power wireless communications.

Dr. Friedman is the Regional Editor of the *Journal of Circuits, Systems, and Computers*, a Member of the editorial boards of the *Analog Integrated Circuits and Signal Processing*, *Microelectronics Journal*, *Journal of Low Power Electronics*, and *Journal of Signal Processing Systems*, Chair of the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS steering committee, and a member of the technical program committee of a number of conferences. He previously was the Editor-in-Chief of the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, a member of the editorial board of the PROCEEDINGS OF THE IEEE and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II: ANALOG AND DIGITAL SIGNAL PROCESSING, a Member of the Circuits and Systems (CAS) Society Board of Governors, and Program and Technical chair of several IEEE conferences. He was a recipient of the University of Rochester Graduate Teaching Award and a College of Engineering Teaching Excellence Award. He is a Senior Fulbright Fellow.



Inna Vaisband (S'12) received the B.Sc. degree in computer engineering and the M.Sc. degree in electrical engineering from the Technion—Israel Institute of Technology, Haifa, Israel, in 2006 and 2009, respectively. She is currently working toward the Ph.D. degree in electrical engineering from the University of Rochester, Rochester, NY, under the supervision of Prof. E. G. Friedman.

Between 2003 and 2009, she held a variety of software and hardware R&D positions at Tower Semiconductor Ltd., G-Connect Ltd., and IBM Ltd., all in Israel, and a Visiting Researcher Position at the Stanford, CA in 2012. Her current research interests include the analysis and design of high performance integrated circuits, analog design, and on-chip power delivery.