Buried Power Rails and Back-side Power Grids: Arm® CPU Power Delivery Network Design Beyond 5nm

Divya Prasad[1], S. S. Teja Nibhanupudi[1,2], Shidhartha Das[1], Odysseas Zografos[3], Bilal Chehab[3], Satadru Sarkar[3], Rogier Baert[3], Alex Robinson[1], Anshul Gupta[3], Alessio Spessot[3], Peter Debacker[3], Diederik Verkest[3], Jaydeep Kulkarni[2], Brian Cline[1] and Saurabh Sinha[1]

[1]Arm Research, email: divya.prasad@arm.com

[2]University of Texas at Austin, Austin, TX, USA, [3]imec, Leuven, Belgium

Abstract— The technology of buried power rails and back-side power delivery has been proposed for future scaling enablement, beyond the 5nm technology node. This paper studies the CPU design implications of power delivery in the context of these technologies. Employing standard VLSI design flows and sign-off techniques, we benchmark the power delivery designs and technology options using the Arm Cortex-A53 CPU at an imec 3nm technology node. DC and AC analyses of the resulting power delivery networks are presented for the various designs with buried power rails (with front-side and back-side power delivery) and compared to conventional designs without buried power rails. It is shown that buried rails with front-side power delivery can improve the worst-case IR drop from 70mV to 42mV (~1.7X reduction); while buried rails with back-side power delivery substantially reduce IR drop to 10mV (a 7X reduction).

I. INTRODUCTION

Relentless scaling of transistors and wires in advanced semiconductor technologies has not only resulted in major process-related challenges but has also imposed severe design challenges in the sub-5nm technology regime. Dimensional scaling of designs, today, has been made possible by (i) Front-End-of-Line (FEOL) and Back-End-of-Line (BEOL) pitch scaling, which worsens short-channel effects in transistors and increases wire/contact resistances; and, (ii) fin depopulation in standard cells, which causes degradation of transistor drive. To enable further area scaling in sub-5nm nodes, a novel approach of burying the power rails into the substrate (Fig. 1–3) has been proposed, which no longer requires reserving two routing tracks for power nets (e.g, V_{DD} or V_{SS}) in the standard cell area [1]. Additionally, these buried power rails (BPRs) can achieve a higher aspect ratio, thus, exhibiting lower resistance than local-level BEOL power rails [2]. However, the new characteristics and design-constraints of BPR require re-design and re-structuring of the power delivery network that connects to it.

In this paper, we analyze and co-optimize different BPR technology configurations with various Power Delivery Network (PDN) designs – aiming to tradeoff IR drop for Performance-Power-Area (PPA) – and benchmark the overall impact using the Arm® Cortex®-A53 CPU. Our main contributions are as follows: (i) CPU PDN re-design for both front-side and back-side BPR configurations, (ii) Design-Technology-Co-Optimization (DTCO) of the back-side network and μTSVs for improved IR drop, (iii) dynamic IR drop and AC analyses, including sensitivity analyses to BPR technology parameters at advanced technology nodes.

II. TECHNOLOGY

Ruthenium (Ru) Buried Power Rails: Non-planar FinFET structures facilitate large portions of Shallow-Trench-Isolation (STI) field between devices, creating space for deep trench, high aspect ratio BPR rails, lowering resistance to beyond regular local BEOL rails. The BPR can be implemented as a module between the STI module and fin reveal. A cavity is etched between fins that can extend in the bulk. A dielectric is deposited in the cavity before metallization. W or Ru are considered to survive the FEOL process thermal budget. High aspect-ratio (≤1:7) Ru BPR has been shown experimentally to have resistances of 50-30Ω/μm and are connected to the fins through a tall via down from the contact layer (Fig. 3). Accessing the rail from a front-side PDN requires a via stack from M1 down to the buried rail, passing through the contact layer which is realized with dedicated power tap cells without active fins (Fig. 5). The free space in tap cells allows resistance optimization by merging vias and adding extra metal tracks to lower the access resistances (≥23Ω) by merging up to 2 vias.

Back-Side Power Delivery: To remove the overhead of power tap cells and the whole PDN, altogether, from the front side (which decouples the sharing of wiring resources between signals and power delivery), 3D integration techniques have been proposed to implement a back-side PDN where power is delivered using a small ("micro") Through-Silicon-Via (μTSV) that lands on the buried rail [4]. After processing the front side, the wafer is thinned to 500nm and μTSVs are etched from the back side, using the BPR metal as an etch stop layer (Fig. 4). Thick metal layers (BM1–2) are then created on the back side to distribute power to the C4 bumps.

Standard-cell design: To evaluate these technology choices, 6-track, 1-fin standard cells at an imec 3nm node, with and without BPR, are selected (Fig. 2), allowing iso-area and iso-performance comparisons to quantify the impact of the BPR and PDN strategies. The BEOL stack is defined to align with foundry 3nm projections (Table 1). The local wires (MINT–M3 in Table 2) are implemented in Ru (aspect ratio 1.5).

The IR problem in Advanced Technology Nodes: Owing to fin-depopulation, peak gate current reduces with technology scaling while the minimum resistance of totem (via pillar) to reach the transistor pin from the top of the metal stack is increasing rapidly with technology scaling (Fig. 6). IR drop has become a fundamental bottleneck in high-performance designs at advanced nodes that are characterized by low supply voltages, forcing designers to trade-off signal routing resource to build finer, robust power grids.

III. PDN Design and Implications

Three power-rail technologies are explored in this study (Fig. 7); namely, the traditional Front-Side (FS) PDN, Front-Side with Buried-Power-Rails (FS-BPR) and Back-Side power delivery with Buried-Power-Rails (BS-BPR). The analysis of FS and FS-BPR is performed using a state-of-the-art VLSI design flow (Fig. 8(a)), along with a 13-metal-layer interconnect stack (Table 2) targeting interconnect pitches representative of ~3nm nodes [5] with a single Re-Distribution-Layer (RDL) layer to connect to C4 bumps (placed symmetrically at the four corners of the 150x150µm design, as shown in Fig. 11). M11–M13 are dedicated to power and clock, while the lower layers are shared with signal routing (power grid in Fig. 9 and Table 2). The supply rails reside on MINT for FS scenario, whereas, the supply rail is on MBUR for FS-BPR scenario and three PDNs are designed with decreasing V_{DD}/V_{SS} grid pitch to create a finer power grid for improved IR (Table 3). The large IR drop in local metal layers (MBUR–M4) is resolved in PDN1–2 through finer local grids (Fig. 12). The BS-BPR study utilizes the same front side assumptions, except that, the power delivery is designed using a back-side metal stack that contacts the buried rails through the µTSV (Fig. 8(b), Table 4). A margin of $10\% \cdot V_{DD}$ is allocated ($\Delta V_{DD}+\Delta V_{SS}$) for dynamic analysis. The physical design of the Cortex-A53 CPU is used to model voltage drop scenarios for a range of activity factors ($\alpha_{5\%\text{-}15\%}$) that generally apply across digital CPU design: from low power to high performance.

Front-side (FS) PDN without buried rails: The highly resistive power rails on MINT result in a large resistive drop in the FS PDN1–3 scenarios, with worst-case resistances of 1.5kΩ , ~800Ω, ~700Ω, respectively, in PDN1–3. These high resistances render the design vulnerable to IR hot-spots as depicted in Fig. 14 (a)–(b); for example, there is a large hot-spot around the clock pin where there is dense placement of large clock buffers to satisfy performance requirements of the core. There are many IR failures with PDN1–2 for both α_H and α_L scenarios (defined in Table 5), with minor violations on PDN3 that exceed the allocated dynamic IR margin (10%).

Front-side PDN with Buried Rails (FS-BPR): While buried rails (MBUR) exhibit lower resistance (≥30Ω□µm) than the MINT rails, connecting or "tapping" to the buried rails is a challenge. In the FS-BPR case, BPR "taps" are made using dedicated tap cells that are carefully placed across the design, balancing block-level implications. Frequent tap cells are desirable for low IR drop, but can also impose block-level placement limitations and impact performance. For example, a tap-cell frequency of $24 \cdot CPP$ in columns (as in [4]), limits the use of large buffers (≥X12) and complex logic cells, rendering the realization of a compute intensive core with stringent PPA very challenging. Instead, a checkerboard scheme is designed (Fig. 10 (b)) to alleviate this problem. The dynamic IR drop is improved with BPR compared to the FS scenario (Fig. 14 (b)); however, there persists a large resistive drop on local metal-layers (Fig. 12). Despite reducing the V_{DD}/V_{SS} tap cell pitch by 0.5x, there are still IR violations in α_H scenario. Doubling the M4 power straps in PDN3 allows further reduction in the M3 resistance that dominates the PDN resistance (Fig. 12), with minimal overhead on wiring resource (since M4 in the PDN is only a small segment/strap). The tradeoff between IR drop and performance by using a finer power grid is depicted in Fig. 13.

Back-side PDN with Buried Rails (BS-BPR): Back-side power delivery removes wiring resource contention between power nets and signal nets but requires tight-pitch µTSV contacts for every buried rail, increasing the process complexity. We stagger the contacts to relieve the µTSV pitch and quantify IR drop (Fig. 15). The aggressive µTSV pitch shows tremendous promise towards addressing the IR drop problem (Fig. 14(d), 16). However, a sweep of µTSV pitch with various MBUR aspect ratios (Fig. 17) shows that even the BS-BPR poses an IR drop challenge at a µTSV pitch beyond 1µm, owing to long distances traversed on the buried rails that are more resistive than BM1–2.

Performance versus IR drop: The frequency versus power dissipation (at iso-block-area) for the FS, FS-BPR, BS-BPR scenarios (Fig. 18–19) shows the PPA benefit of the BS-BPR over FS and FS-BPR. Additionally, the BS-BPR shows both lower energy (mW/GHz) and lower IR drop (Fig. 19). A robust front-side power grid incurs an energy loss to improve IR drop, particularly in FS-BPR due to the overhead of tap cells; however, the back-side power delivery decouples this trade-off.

AC analysis with chip/package/PCB model: Low PDN resistance in BS-BPR lowers the IR-drop significantly; however, the lower resistance provides reduced damping effect during power-supply resonant oscillations. These worst-case oscillations occur when periodic activity in the cores align with the natural resonance frequencies of the chip-package-PCB dynamical system, causing large power-supply noise [6, 7]. Fig. 20 shows the peak PDN impedance for PDN3 for the FS, FS-BPR and BS-BPR designs under various multi-core configurations and Fig. 21 shows the equivalent PCB, package and die *RLC* model used for the AC analyses. We assume a single-switching core while the remaining cores are idle to provide useful decoupling capacitance (decap) to minimize noise. A combination of reduced capacitance and reduced damping resistance in the BS-BPR (50mΩ PDN resistance) design, causes peak impedance at resonance to be higher than the FS-BPR and FS configurations which have relatively higher effective resistances of 105mΩ and 108mΩ, respectively. Increased decap at higher core counts mitigates the impact of the low damping resistance (16 cores and 32 cores), thereby reducing the impedance peaks for all PDN configurations.

IV. Conclusions

A thorough PDN design study and associated DC and AC analyses of the Cortex-A53 CPU at the 3nm node, targeting buried power rails and back-side PDN, is presented in this paper. We found that meeting IR drop margins at the 3nm node is extremely challenging using regular, front-side PDNs (FS) but the margins could be achieved with buried rails (FS-BPR and BS-BPR). FS-BPR improved the IR drop by 40% compared to FS with PDN3 ($10\% \cdot V_{DD}$ to $6\% \cdot V_{DD}$). Furthermore, BS-BPR improved IR drop to $\sim 1\% \cdot V_{DD}$ with tight µTSV pitch. BS-BPR with low PDN resistance shows worse peak impedance at resonance, that can be mitigated with on-chip decap in large die SoCs.

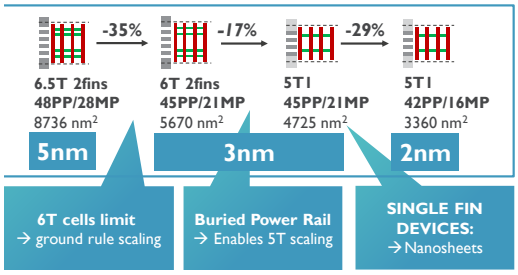

Fig. 1. Imec Technology scaling roadmap to iN3. BPR enables transition from 6T to 5T for 1 fin or nanosheet devices to reduce area by 17% without pitch scaling.

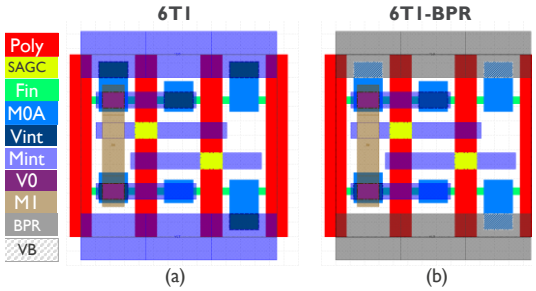

Fig. 2. iN6 Standard cells used: 6T 1fin cells with and without BPR. Cell architectures are identical except for the power rail.

Layer	Ω // Ω/µm
BPR(MBUR)	50–30
VBPR(VBUR)	72–23
Mint/M1/M3	900
M2/M4	214
µTSV	5

Table 1. BPR, to local rail TSV, and backside line resistances.

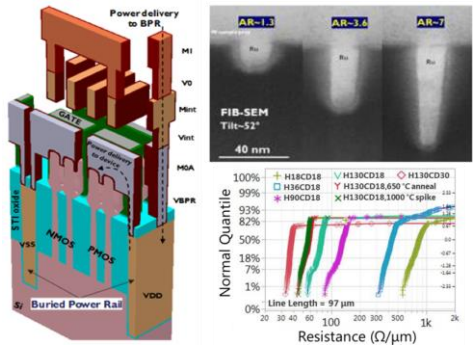

Fig. 3. (left) BPR contacting from fin and from M1 (right) TEM/SEM image of Ru BPR and measured line resistances down to 35Ω.

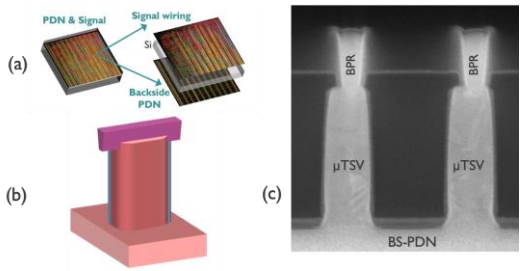

Fig. 4. BS-PDN: (a) conceptual diagram (b) µTSV down to 90nm CD, landing on BPR through a 500nm thin wafer (c) BPR connected to BS-PDN with a µTSV.

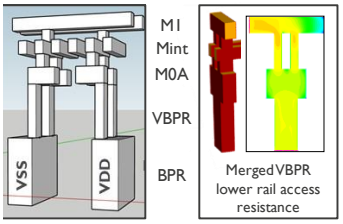

Fig. 5. Power Tap strategy. High aspect ratio single VBPR have a fairly high resistance. Merging multiple vias in a tapcell to a larger rectangular via lowers the VBPR resistance from 72Ω to 23Ω.

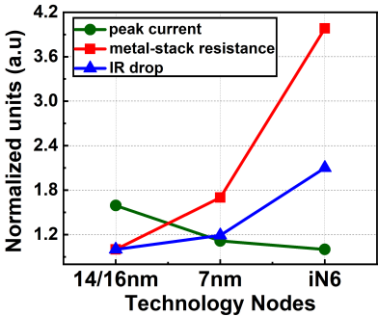

Fig. 6. Current, metal stack resistance and IR drop scaling trends.

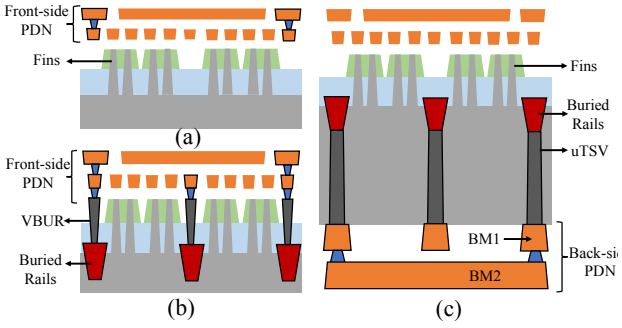

Fig. 7. Cross-section of (a) regular front-side (FS) PDN, (b) front-side with buried power rail (FS-BPR) PDN and (c) BS-BPR PDN.

Layer	Pitch	PDN
MBUR	64 nm	Stripe
MINT	22 nm	Totem
M1-M2	30-22 nm	Totem
M3	30 nm	Stripe
M4	48 nm	Stripe
M5-M6	48 nm	Totem
M7-M9	80 nm	Grid
M10-M11	160 nm	Grid
M12-M13	320 nm	Grid
RDL	3.2 µm	Grid
C4 bumps	150 µm	Grid

Table 2. Front-side PDN design.

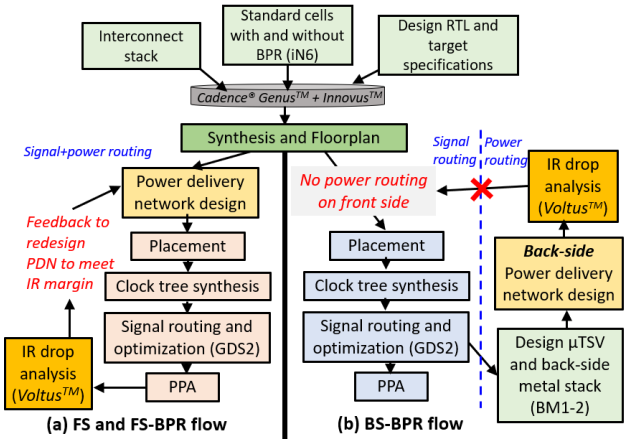

Fig. 8. Design benchmarking and power delivery analysis flow.

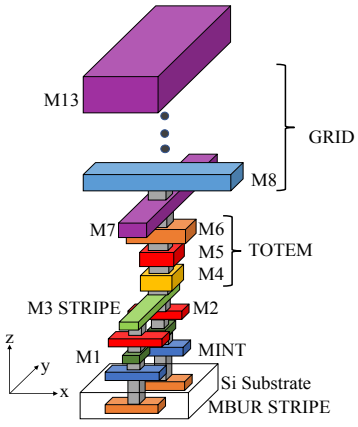

Fig. 9 Metal orientation for front-side (FS) power delivery network (PDN).

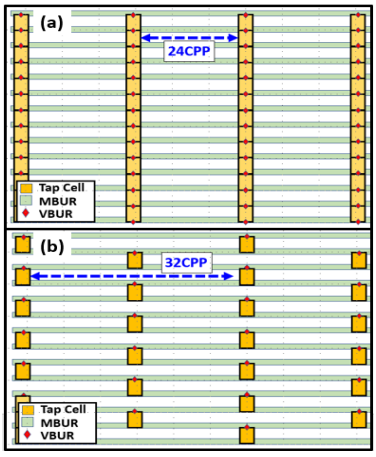

Fig. 10 (a) column tap-cell from [4] vs (b) checkerboard tap-cell configuration in FS-PDN used in this study.

PDN configuration	Layers	V_{DD}–V_{SS} Pitch
FS / FS-BPR ***PDN1***	M1-M13	32 CPP
FS/ FS-BPR ***PDN2***	M1-M13	16 CPP
FS/ FS-BPR ***PDN3***	M1-M3 M4 M5-M13	16 CPP 8 CPP 16 CPP

Table 3. Front-side PDN configurations. CPP = 45nm at 3nm.

PDN	Layers	CD	Pitch
BS-BPR ***PDN4***	µTSV BM1 BM2	60 nm 250 nm 1 µm	500 nm 500 nm 2 µm
BS-BPR ***PDN5***	µTSV BM1 BM2	60 nm 375 nm 1 µm	750 nm 750 nm 2 µm
BS-BPR ***PDN6***	µTSV BM1–2	60nm 1 µm	2 µm 2 µm

Table 4. Back-side PDN configurations. CPP = 45nm at 3nm.

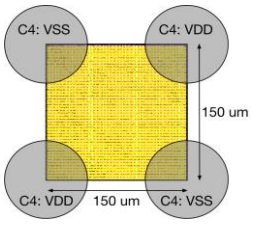

Fig. 11 C4 Bump configuration.

	α_L	α_H
Data α	5%	10%
Clock α	100%	100%

Table 5. Dynamic vector-less IR drop analysis setup with high and low activity.

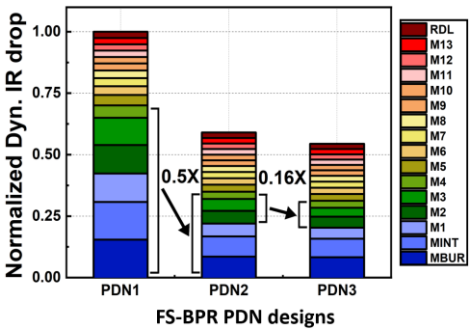

Fig. 12. FS-BPR layer-based IR drop distribution for PDN1–3 depicting the contribution of local metal-levels and targeted improvements in PDN2–3.

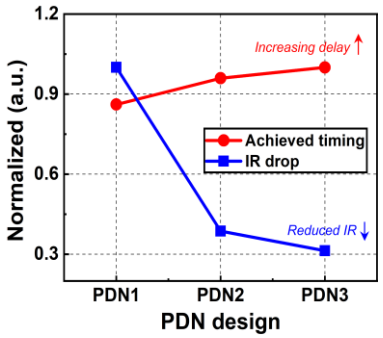

Fig. 13. Performance vs IR drop for FS PDN. A robust PDN limits frequency.

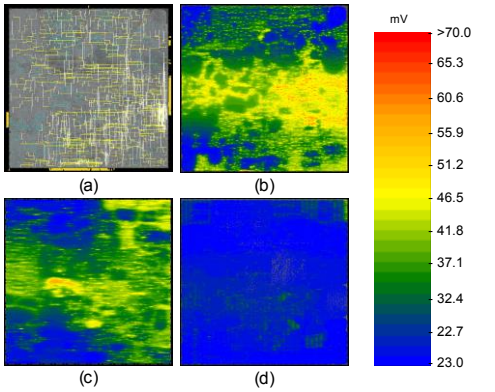

Fig. 14. (a) FS PDN die-image with clock tree (b) FS PDN dynamic IR heat-map, (c) FS-BPR PDN dynamic IR heat-map and (d) BS-BPR PDN dynamic IR heat map.

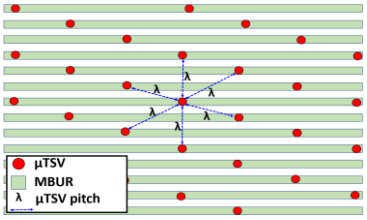

Fig. 15 Back-side µTSV landing on MBUR; BM1–2 are routed above µTSVs.

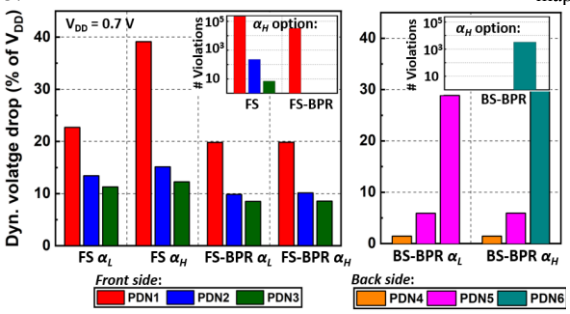

Fig. 16 Worst case dynamic voltage drop for different PDN and technology configurations.

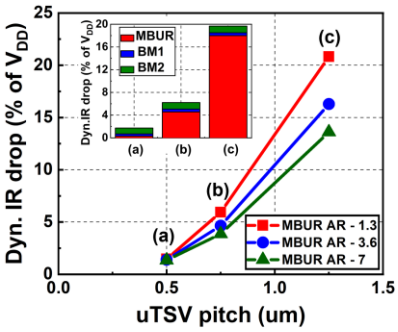

Fig. 17. Voltage drop vs. µTSV pitch.

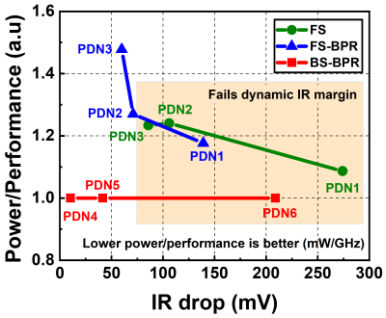

Fig. 18. Energy versus dynamic IR drop.

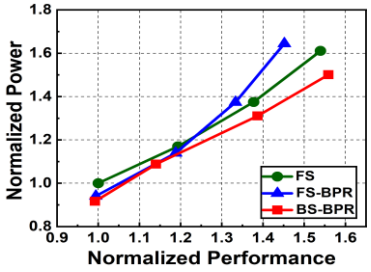

Fig. 19. Performance vs Power for 3 cases.

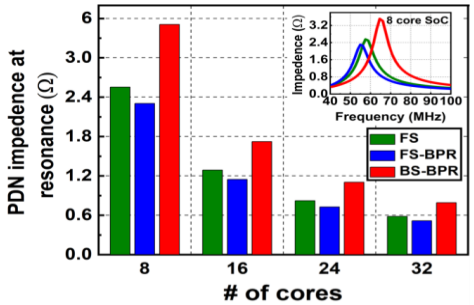

Fig. 20. Peak impedance at resonance for FS, FS-BPR and BS-BPR under multi-core configurations.

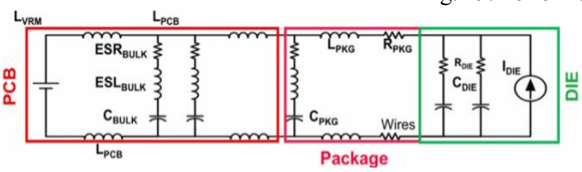

Fig. 21 Equivalent circuit model for PCB, package (PKG) and die [6]. VRM: Voltage regulator module

REFERENCES

[1] B. Chava, et al., SPIE 2018
[2] S. M. Salahuddin et al., EDL 2019
[3] A. Gupta, et al., IITC 2018
[4] B. Chava, et al., SPIE 2019
[5] S.M. Sherazi, et al., SPIE 2019
[6] S. Das, et al., ISLPED 2015
[7] K. Chang, et al., ISLPED 2017