Towards the Integration of Voltage Regulators in Server Applications

PwrSoC, October 19th, 2018

Pedro A. M. Bezerra¹, Florian Krismer¹, Johan W. Kolar¹, Arvind Sridhar², Thomas Brunschwiler², Thomas Toifl²

¹ Power Electronic Systems Laboratory (PES), ETH Zurich, Switzerland
² IBM Research, Ruschlikon, Switzerland

The complete version of the slides are available at the PES-ETH website: https://www.pes-publications.ee.ethz.ch/publications/conferences/
Outline

► Target application
► Motivation to use IVRs
► System specifications and target achievements
► Components’ model
► Optimization procedure
► Preliminary experimental results
► Conclusion and outlook
Point of Load Conversion for Server Applications

Datacenter Supply Chain

- AC Input: 3Ø @ 381 V @ 50 Hz
- Main DC Bus (Batteries): 48 V
- Intermediate DC Bus: 12 V
- PoL: 3.3 V and 1.7-0.5 V

Point of Load Converter

- Fully Integrated Voltage Regulators
- DC-DC Converters

Power consumption in MW range
High power demand requires high efficiency!
Why going for IVRs in Modern Microprocessor Applications?

► Allow for considerably energy savings
  ► Dynamic Voltage and Frequency Scaling (DVFS)
  ► Reduced number of interconnects to the microprocessor package
► Size reduction
► Reliability improvement
► Allow the use of modern CMOS Technologies for power switches
► Faster response to load and reference voltage transients
CarrICool Project (FP7-ICT-619488)

Multi-functional interposer platform that provides scalable cooling, granular chip-level power delivery and optical signaling required for scale-up systems

WP5: Interposer platform

WP4: Optical signalling

WP6: Demonstrator

WP2: Heat removal

WP3: Power delivery

► Partners in the power management working package:

Tyndall National Institute

ipd ia

Fraunhofer IZM

ETH Zürich

IBM
Considered System Specifications and Target Achievements

► Specifications

\[
\begin{align*}
V_{\text{in}} &= 1.6 \text{ V} \\
V_{\text{out,nom}} &= 0.8 \text{ V} \\
I_{\text{out}} &= 1 \text{ A} \\
P_{\text{out}} &= 0.8 \text{ W} \\
\Delta V_{\text{out,max}} &= 1\% \cdot V_{\text{out}}
\end{align*}
\]

► Specifications taken from the most power consuming voltage domain
► Power is scaled down

► Target achievements

► Overall efficiency $\eta > 90\%$
► Overall power density $\rho > 1 \text{ W/mm}^2$
► Chip power density $\rho > 20 \text{ W/mm}^2$
  • Only 1% of the microprocessor area is allowed for power management

► Beyond the state of the art!
IVR Design and Optimization

**Challenge:** IVR design and optimization
- No standard design tool is common to the partners
- Expensive for prototyping

**Methodology:** Virtual prototyping and multi-objective optimization
Integration Level and Considered Topology

► 2.5D integration level

- Better quality factor integrated passives compared to 2D and 3D approaches
- Take advantage of high FOM deep sub-micron transistors

► Four-phase interleaved buck

- Better heat and loss distribution among the components
- Allow for phase shedding at low load operation
- Stacked configuration supports the relatively high input voltage

► Main waveforms

- Output and input current ripple reductions
Racetrack Inductors with Core Material

Dimensions description

Input
- Number of turns (n)
- Winding width (t_w)
- Winding thickness (t_t)
- Winding spacing (t_s)
- Core thickness (c_t)
- Core length (c_l)
- Switching frequency (f_s)
- $i_i$ peak to average ratio (PAR)

Racetrack inductor with core material model

Output
- Inductance (L)
- Copper Losses ($P_c$)
- Core Losses ($P_L$)
- Footprint Area ($A_f$)

Microscopic view

Thin-film with magnetic material
- $Ni_{45}Fe_{55}$

Loss Model considers
- Dc and ac copper losses
- Eddy currents and hysteresis core losses

Andersen et al., IEEE Trans. on Power Electronics, 2013
Deep-Trench Capacitors

- **3D structures in silicon - Tripods**

- **PICS3: Passive Integration Connective Substrate**
  - Capacitance density up to 250 nF/mm² with high capacitance stability vs. temperature
  - ESR vs. capacitance extracted from experimental data

Lallemand et al., EMPC, 2013
Power Transistor Model for Stacked Configuration

- Cadence transient simulations demand high computational efforts
  - Based on measurements results
  - Too long simulation time
- Necessity of accurate and simplified loss modeling for optimization
  - Semiconductor losses dependent on the transistors channel width ($T_{wp}$, $T_{wn}$ - channel length fixed by design rules), dead times ($t_{d,1}$, $t_{d,2}$), and chip temperature
  - Low computational effort
- Cadence based simplified transistors model!
  - Represents the most significant source of losses
  - Uses a discrete number of cadence simulations and a multivariable interpolation algorithm
Considered Power Stages

► Conventional CMOS HB

► CMOS ANPC

► Proposed CMOS ANPC

Problem of unequal voltage distribution during the switching transients and steady-state.

Clamping switches are added to assure voltage balance among the transistors

Bezerra et al., COMPEL, 2017
Considered Power Stages

- **Conventional CMOS HB**
- **CMOS ANPC**
- **Proposed CMOS ANPC**

Unlike the conventional CMOS ANPC, the proposed bridge maintains the clamping switches off during the entire dead-time period assuring soft-switching.
Considered Power Stages

- Conventional CMOS HB
- CMOS ANPC
- Proposed CMOS ANPC

- Due to the voltage balance and less losses during the hard switching event up to 1% efficiency can be saved using the proposed bridge at 150 MHz
- The efficiency improvements of using the proposed CMOS ANPC increase with frequency compared to the conventional approach
Pre-optimization Loop of the Power Switches

Global design variables:
- Topology
- Modulation Scheme
- Design space

Schematics of the characterization setup and load the switches’ design space

Simulation of the half-bridge for one switching cycle

Automatic calculation of the distinguishing time intervals

Calculation of the transistors’ energy losses and resistances

Generation of the interpolation matrices and fitting functions

Dead-time optimization for switching losses minimization

To the main loop

- Cadence
  - Sweeps of $T_w$, $D_r$, $I_{sw}$, $T_j$
  - 1000s simulations for each HB
  - Saving all switches current and voltages
  - Big data generation (> 50GB)

- Matlab
  - Scan of the waveforms
  - Identification of the switching intervals
Optimization Procedure

Considered design space

Inductor

<table>
<thead>
<tr>
<th>Sym.</th>
<th>Description</th>
<th>Range</th>
<th>Step</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Number of turns</td>
<td>1… 5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$t_w$</td>
<td>Winding width</td>
<td>10… 1400</td>
<td>10 or 500 µm</td>
<td></td>
</tr>
<tr>
<td>$t_t$</td>
<td>Winding thickness</td>
<td>10… 50</td>
<td>20 µm</td>
<td></td>
</tr>
<tr>
<td>$t_s$</td>
<td>Winding spacing</td>
<td>10… 50</td>
<td>20 µm</td>
<td></td>
</tr>
<tr>
<td>$c_t$</td>
<td>Core thickness</td>
<td>1… 10</td>
<td>3 µm</td>
<td></td>
</tr>
<tr>
<td>$c_l$</td>
<td>Core length</td>
<td>1… 10</td>
<td>3 mm</td>
<td></td>
</tr>
</tbody>
</table>

Power Switches

<table>
<thead>
<tr>
<th>Sym.</th>
<th>Description</th>
<th>Range</th>
<th>Step</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{wp}$</td>
<td>P channel width</td>
<td>5… 15</td>
<td>5</td>
<td>mm</td>
</tr>
<tr>
<td>$T_{np}$</td>
<td>N channel width</td>
<td>5… 15</td>
<td>5</td>
<td>mm</td>
</tr>
<tr>
<td>$t_{d,1}$</td>
<td>Dead-time 1</td>
<td>20… 120</td>
<td>50</td>
<td>ps</td>
</tr>
<tr>
<td>$t_{d,2}$</td>
<td>Dead-time 2</td>
<td>10… 50</td>
<td>50</td>
<td>ps</td>
</tr>
</tbody>
</table>

Capacitor

<table>
<thead>
<tr>
<th>Sym.</th>
<th>Description</th>
<th>Range</th>
<th>Step</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{out}$</td>
<td>Output capacitance</td>
<td>0.1… 500</td>
<td>10.2</td>
<td>nF</td>
</tr>
</tbody>
</table>

*Only one transistor size was used for the 14 nm IVR*
Performance Comparison Between IVRs

- 90% efficiency achievable with 0.2 W/mm² more power density using the proposed ANPC HB and 14 nm technology
  - Switches and power stage are more efficient for high frequency operation

- Selected inductance and switching frequency at 90% efficiency:
  - 32 nm Conventional HB:
    - 51 nH @70 MHz
  - 14 nm Proposed ANPC HB:
    - 16 nH @160 MHz
Implementation: PMIC

- PMIC in 14 nm CMOS process
- Four-phase interleaved ANPC buck
- Versatile open loop converter for better PMIC characterization
- Compatible with single and coupled inductors
Implementation: Active Power Stage

- Globalfoundries’ 14 nm Bulk CMOS
- Finfet 3D transistors optimized for digital circuits
- Design uses exclusively low voltage devices

Layout

Schematics

Finfet concept
Implementation: Inductors

- Coupled inductors with magnetic core

  ![Coupled Inductors Diagram]

- 150 MHz design

  \[
  \begin{align*}
  L_{Self2} &= 14.4 \text{ nH} \\
  k_2 &= -0.95 \ (k_{max}) \\
  L_{Self1} &= 1.6 \text{ nH} \\
  k_1 &= 0.95 \ (k_{max})
  \end{align*}
  \]

- 100 MHz design

  \[
  \begin{align*}
  L_{Self2} &= 21.6 \text{ nH} \\
  k_2 &= -0.95 \ (k_{max}) \\
  L_{Self1} &= 2.4 \text{ nH} \\
  k_1 &= 0.95 \ (k_{max})
  \end{align*}
  \]

- Cross section view

  ![Cross Section View]

- Strip-line

- Racetrack
Implementation: All-silicon-based Demonstrators

- Interposer
- 14 nm PMIC
- Coupled inductor (Inv.)

[Images of electronic components and logos from ETH Zurich, IBM, and other institutions]
PCB-based demonstrators and Experimental Results

- **Demonstrator**
- **Wire-bonded PMIC**
- **Passive devices**

- **Switching phase**
- **Switching node**
- **Efficiency estimation**

Coil Craft 36nH

Efficiency [%] vs Power Density [W/mm²] for:
- Coup. opt. for 100 MHz
- V_{in}/V_{ref}=0.77, f_{sw}=100 MHz
- Coup. opt. for 100 MHz
- V_{in}/V_{ref}=0.5, f_{sw}=100 MHz
- Coup. opt. for 150 MHz
- V_{in}/V_{ref}=0.77, f_{sw}=150 MHz
- Coup. opt. for 150 MHz
- V_{in}/V_{ref}=0.5, f_{sw}=150 MHz
- Spiral opt. for 150 MHz
- V_{in}/V_{ref}=0.5, f_{sw}=150 MHz

V_{in} = 1.6 V

ETH Zürich IBM
Conclusions

► Extraction of the switches’ loss models could be transposed for different Technologies
► Migration from 32nm SOI to 14 nm Bulk allows for high efficient and dense designs
► Interposer-based 2.5D integration allows the use of different components’ process
► Better understanding of the switches’ switching behavior allows for improvement in efficiency and reliability of the power stage

Outlook

► Fully Characterization of the designed demonstrators
  ► Losses characterization of the individual devices and full systems
    • Requires accurate temperature measurements
    • Voltage probes embedded to the chip
► Testing of the designed closed-loop IVRs
► Thank you for your attention!