Towards the Integration of Voltage Regulators in Server Applications

PwrSoC, October 19th, 2018

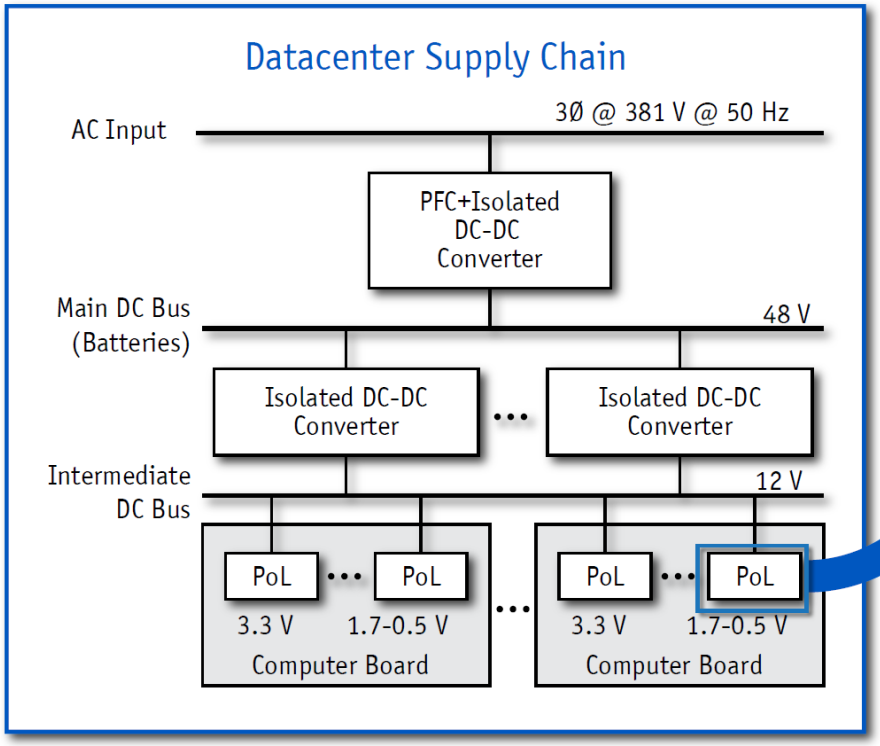
Pedro A. M. Bezerra[1], Florian Krismer[1], Johan W. Kolar[1], Arvind Sridhar[2], Thomas Brunschwiler[2], Thomas Toifl[2]
[1] Power Electronic Systems Laboratory (PES), ETH Zurich, Switzerland
[2] IBM Research, Ruschlikon, Switzerland

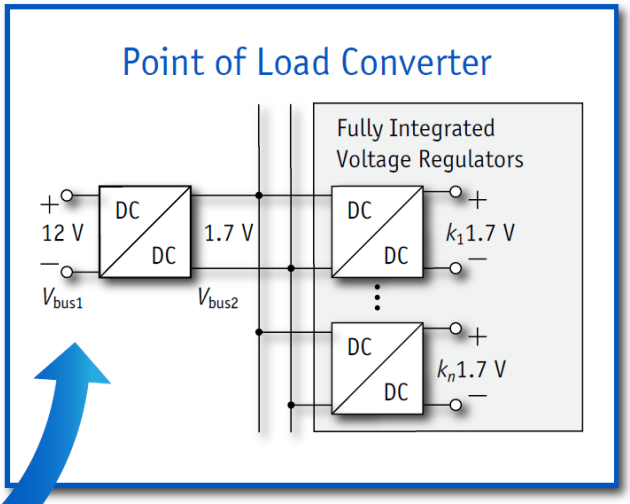
The complete version of the slides are available at the PES-ETH website:
https://www.pes-publications.ee.ethz.ch/publications/conferences/

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

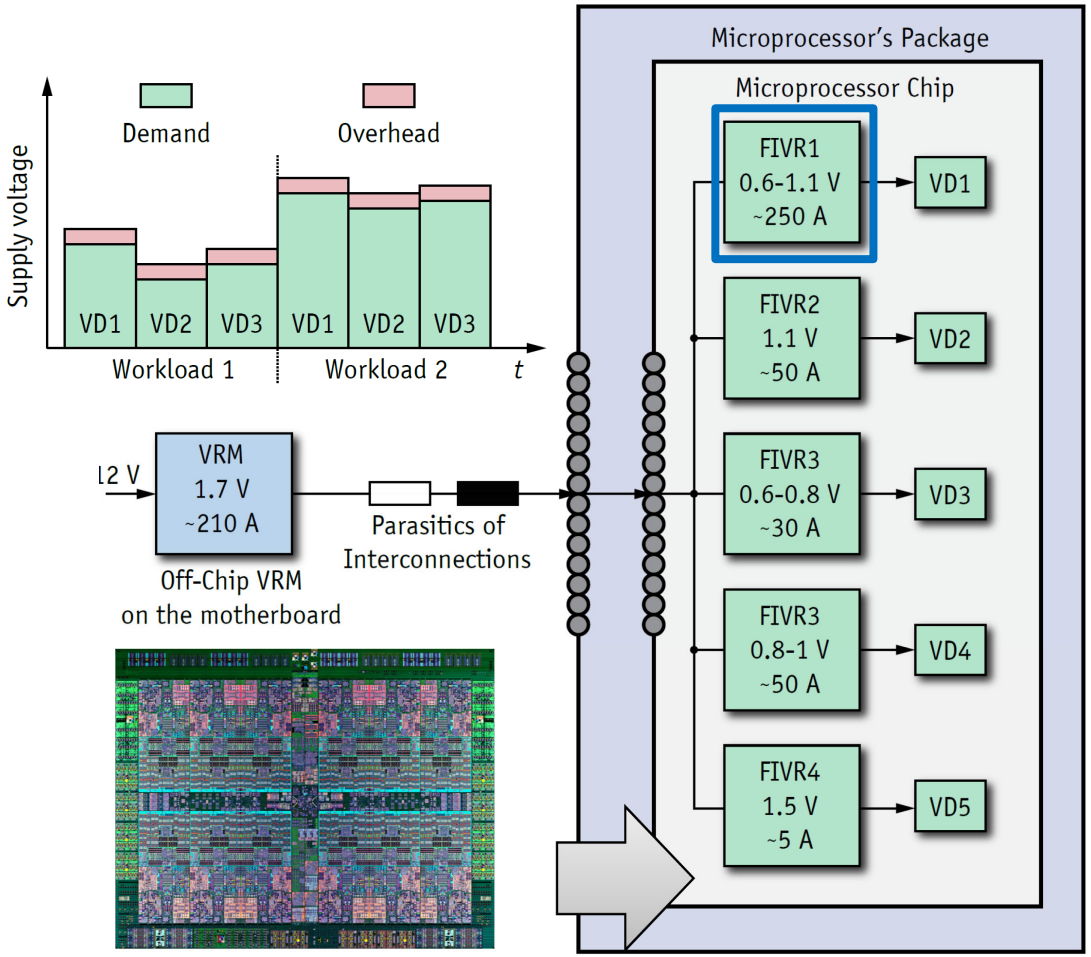




- Power consumption in MW range
- High power demand requires high efficiency!

Why going for IVRs in Modern Microprocessor Applications?

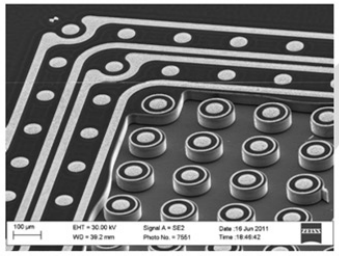
- Allow for considerably energy savings
  - Dynamic Voltage and Frequency Scaling (DVFS)
  - Reduced number of interconnects to the microprocessor package
- Size reduction
- Reliability improvement
- Allow the use of modern CMOS Technologies for power switches
- Faster response to load and reference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



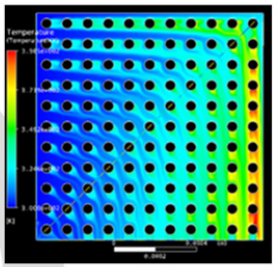
CarrICool Project (FP7-ICT-619488)

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

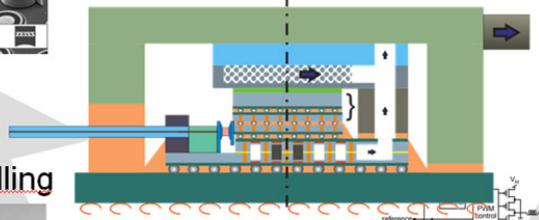
WP5: Interposer platform



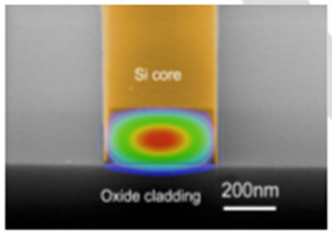
WP2: Heat removal



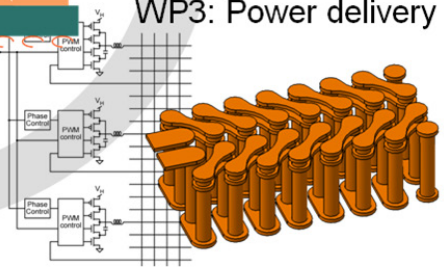
WP6: Demonstrator



WP4: Optical signalling



WP3: Power delivery



►Partners in the power management working package:

Tyndall National Institute

ipdia

Fraunhofer IZM

Considered System Specifications and Target Achievements

▶ Specifications

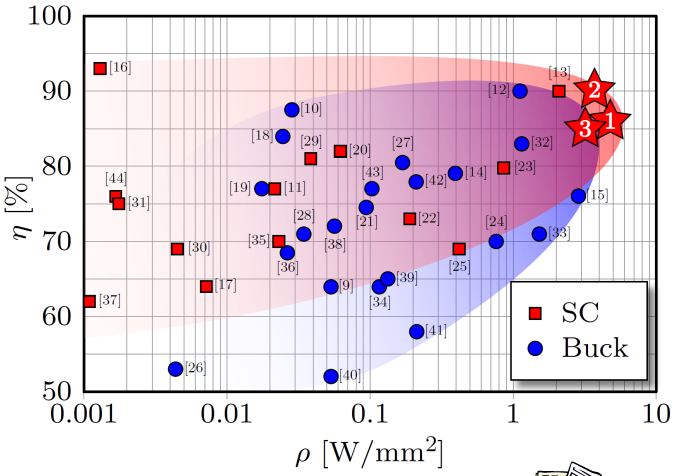
V_{in} = 1.6 V
$V_{out,nom}$ = 0.8 V
I_{out} = 1 A
P_{out} = 0.8 W
$\Delta V_{out,max} = 1\% \cdot V_{out}$

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

▶ Target achievements

- ▶ Overall efficiency η > 90%
- ▶ Overall power density ρ > 1 W/mm²
- ▶ Chip power density ρ > 20 W/mm²
  - Only 1% of the microprocessor area is allowed for power management

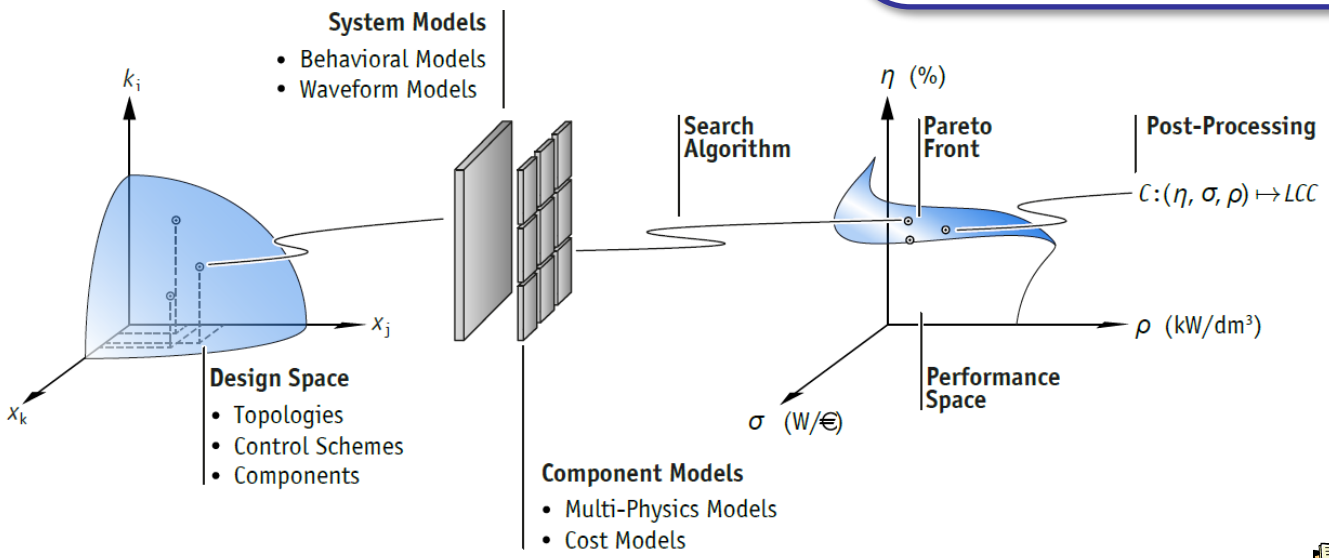
▶ Beyond the state of the art!




T. M. Andersen, *PhD Thesis*, 2015

IVR Design and Optimization

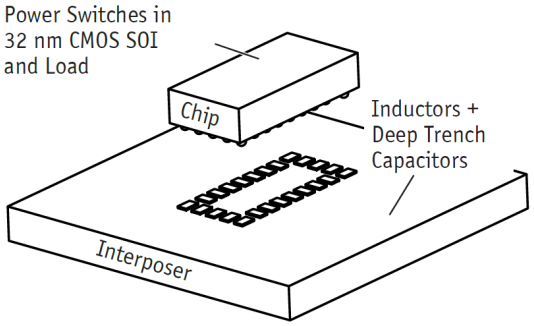
- **Challenge:** IVR design and optimization
  - No standard design tool is common to the partners
  - Expensive for prototyping
- **Methodology:** Virtual prototyping and multi-objective optimization



R. Burkart , *PhD Thesis*, 2016

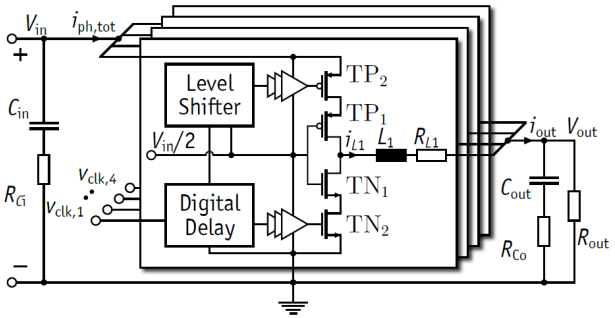
Integration Level and Considered Topology

▶ 2.5D integration level



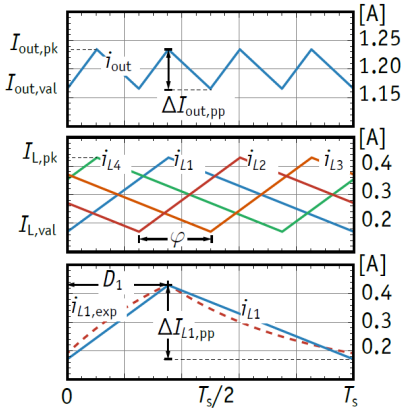
- Better quality factor integrated passives compared to 2D and 3D approaches
- Take advantage of high FOM deep sub-micron transistors

▶ 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

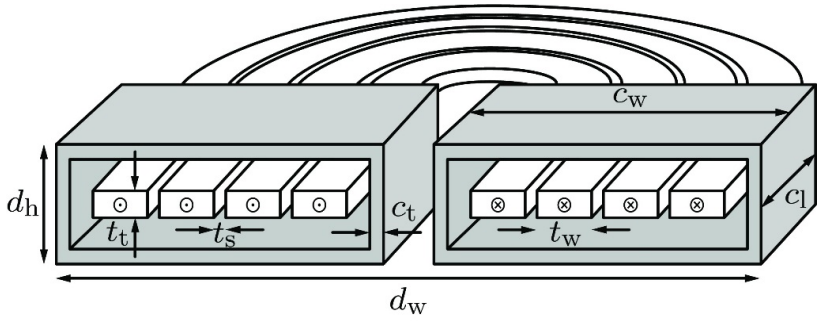
▶ Main waveforms



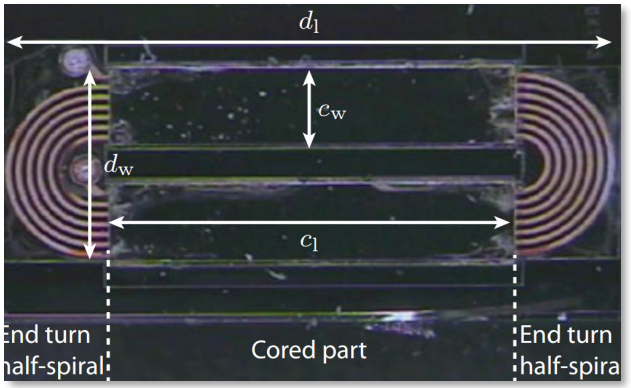
- Output and input current ripple reductions

Racetrack Inductors with Core Material

▶ Dimensions description



▶ Microscopic view



▶ Input

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

Racetrack inductor with core material model

▶ Output

- Inductance (L)
- Copper Losses (P_w)
- Core Losses (P_c)
- Footprint Area (A_L)

▶ Thin-film with magnetic material

- ▶ $Ni_{45}Fe_{55}$

▶ Loss Model considers

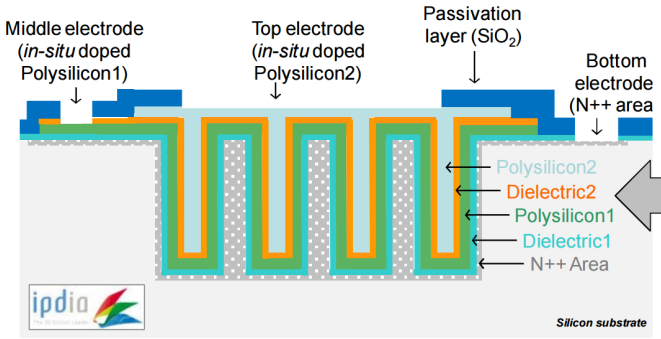
- ▶ Dc and ac copper losses
- ▶ Eddy currents and hysteresis core losses

Andersen et al., *IEEE Trans. on Power Electronics*, 2013

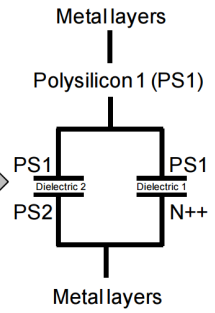
Deep-Trench Capacitors

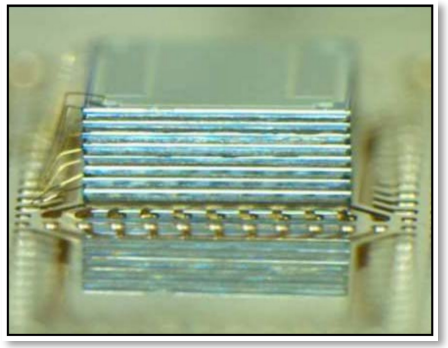
▶ 3D structures in silicon - Tripods

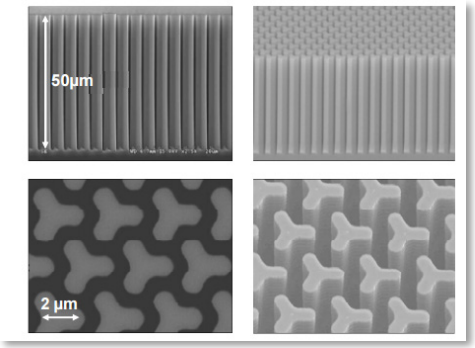
Simplified MIMIM architecture from the PICS3™



Related schematic



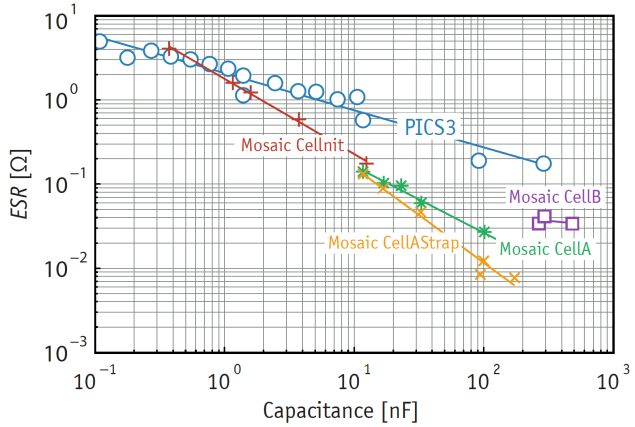




▶ PICS3: Passive Integration Connective Substrate

- ▶ Capacitance density up to 250 nF/mm² with high capacitance stability vs. temperature

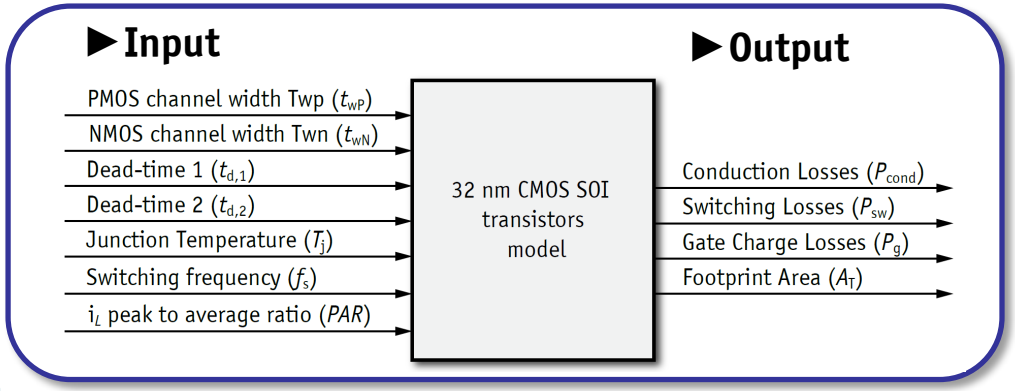
▶ ESR vs. capacitance extracted from experimental data



Lallemand et al., *EMPC*, 2013

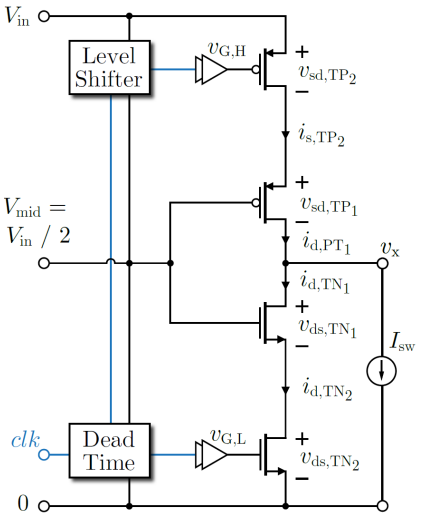
Power Transistor Model for Stacked Configuration

- **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



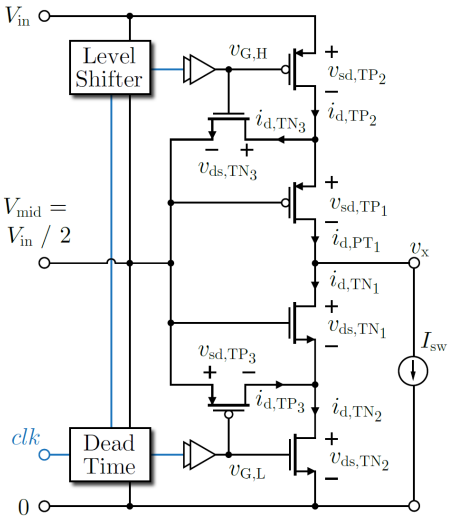
Considered Power Stages

▶ Conventional CMOS HB



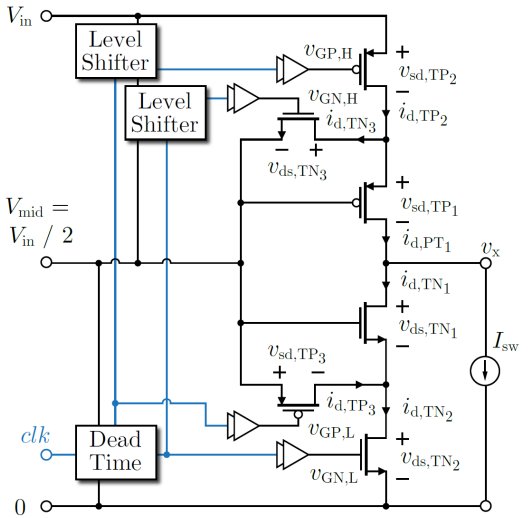
Gate drivers: 2
Power devices: 4
Level Shifters: 1
Independent gate signals: 2

▶ CMOS ANPC



Gate drivers: 2
Power devices: 6
Level Shifters: 1
Independent gate signals: 2

▶ Proposed CMOS ANPC



Gate drivers: 4
Power devices: 6
Level Shifters: 2
Independent gate signals: 2

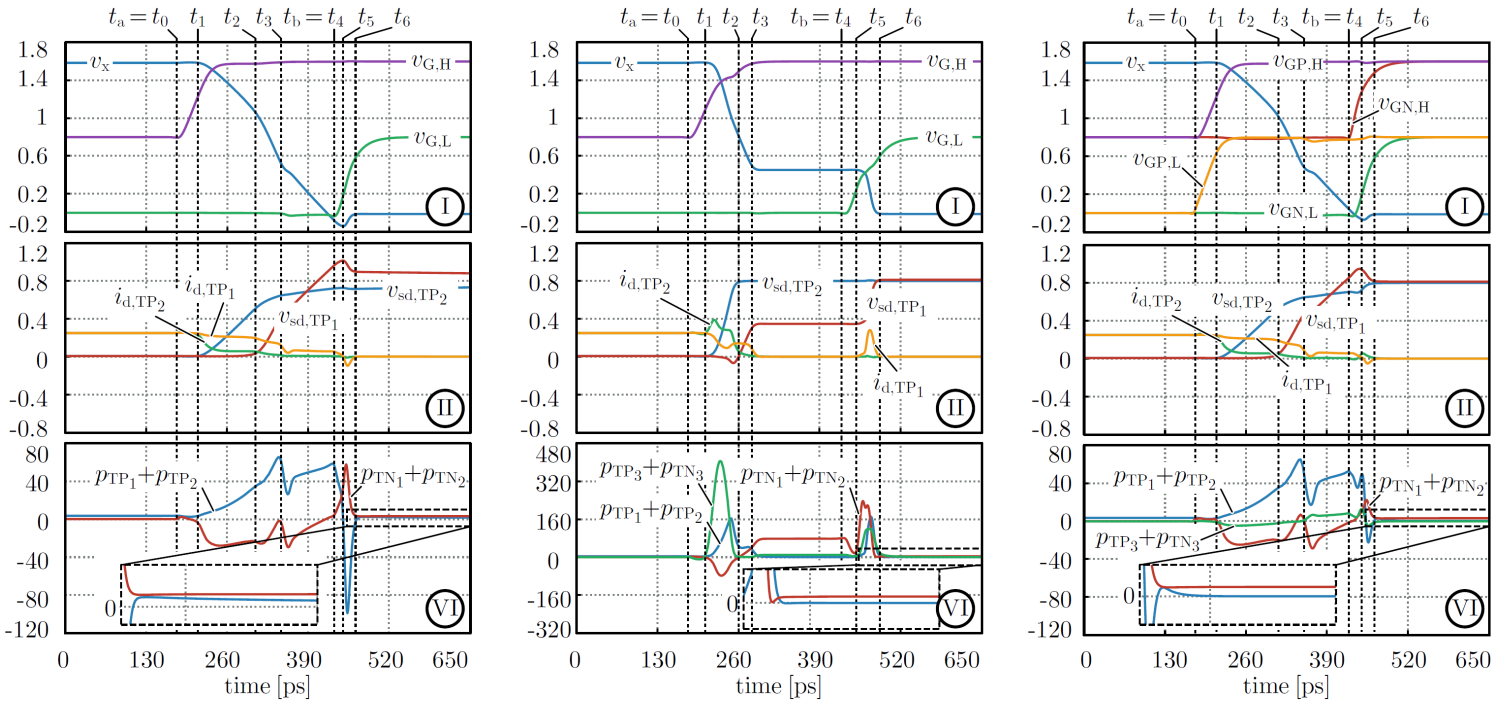
▶ **Problem of unequal voltage distribution during the switching transients and steady-state.**

▶ **Clamping switches are added to assure voltage balance among the transistors**

Bezerra et al., *COMPEL*, 2017

Considered Power Stages

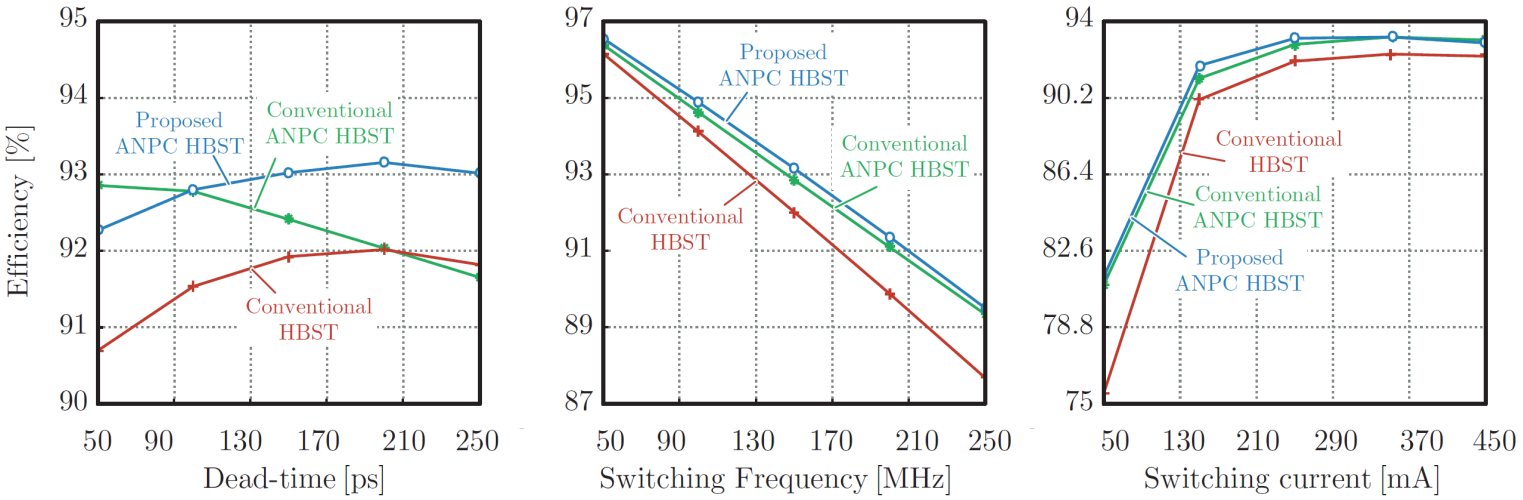
- Conventional CMOS HB
- CMOS ANPC
- Proposed CMOS ANPC



- 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

cadence

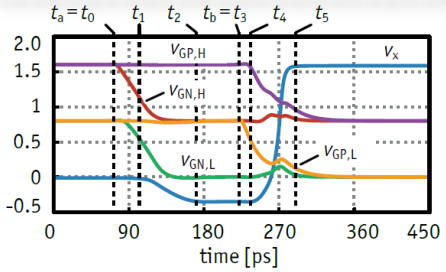
Global design variables:
- Topology
- Modulation Scheme
- Design space

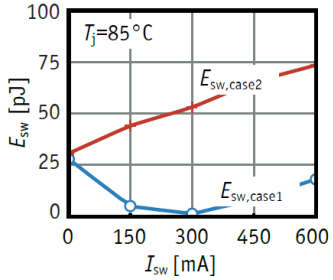
Schematics of the characterization setup and load the switches' design space

Simulation of the half-bridge for one switching cycle

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

Matlab





Automatic calculation of the distinguishing time intervals

- Scan of the waveforms
- Identification of the switching intervals

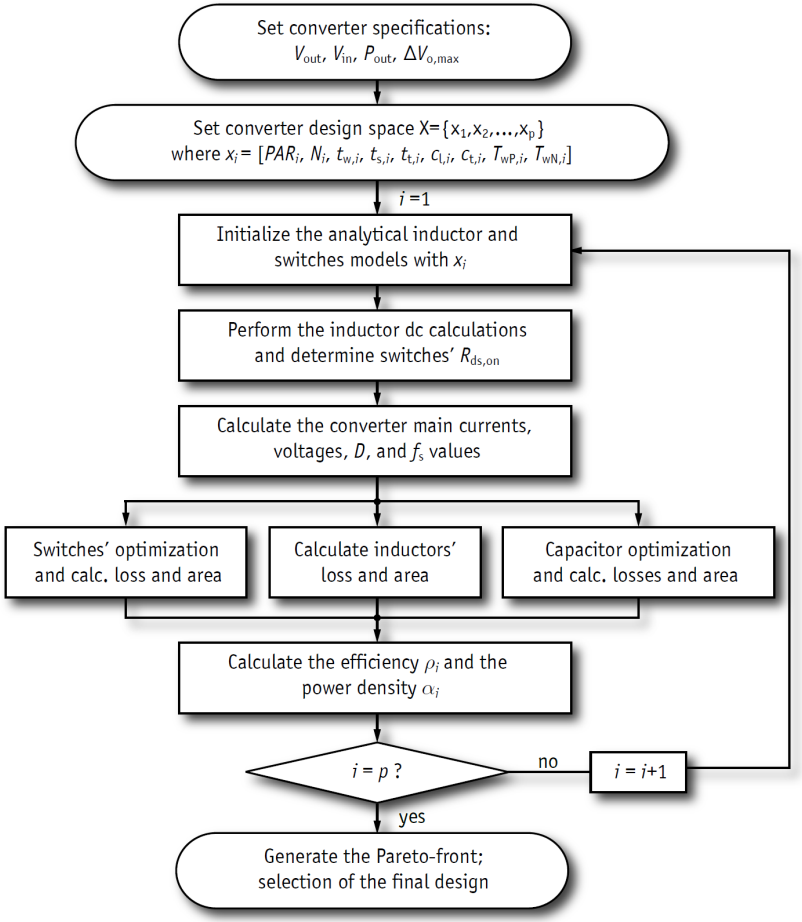
Calculation of the transistors' energy losses and resistances

Generation of the interpolation matices and fitting functions

Dead-time optimization for switching losses minimization

To the main loop

Optimization Procedure



▶ Considered design space

▶ Inductor

Sym.	Description	Range	Step	Unit
N	Number of turns	1... 5	1	
t_w	Winding width	10... 1400	10 or 500	µm
t_t	Winding thickness	10... 50	20	µm
t_s	Winding spacing	10... 50	20	µm
c_t	Core thickness	1... 10	3	µm
c_l	Core length	1... 10	3	mm

▶ Power Switches

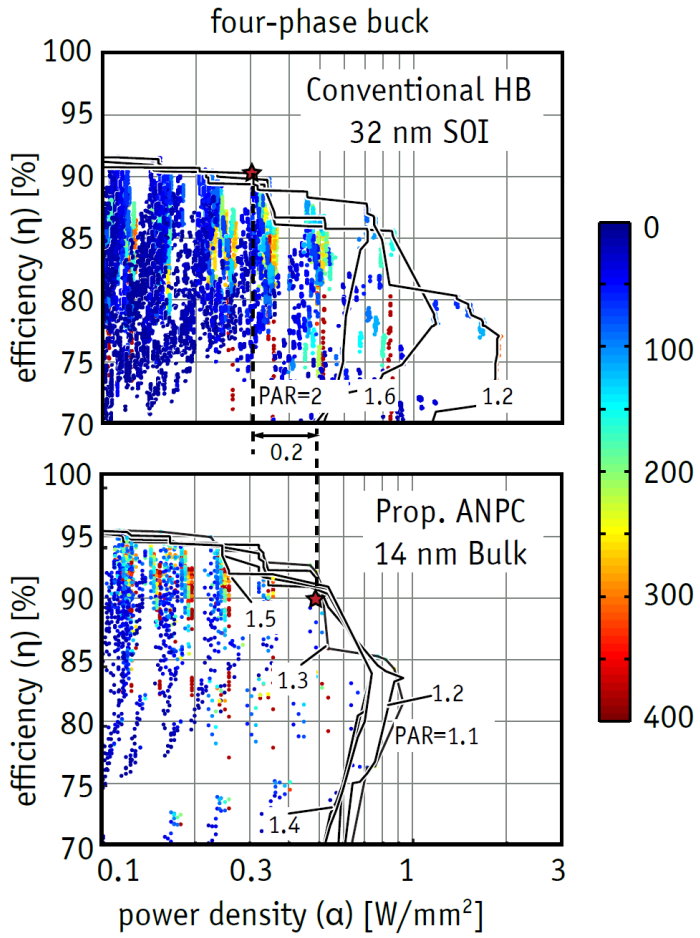
Sym.	Description	Range	Step	Unit
T_{wP}	P channel width	5... 15	5	mm
T_{wP}	N channel width	5... 15	5	mm
$t_{d,1}$	Dead-time 1	20... 120	50	ps
$t_{d,2}$	Dead-time 2	10... 50	50	ps

▶ Capacitor

Sym.	Description	Range	Step	Unit
C_{out}	Output capacitance	0.1... 500	10.2	nF

▶ *Only one transistor size was used for the 14 nm IVR*

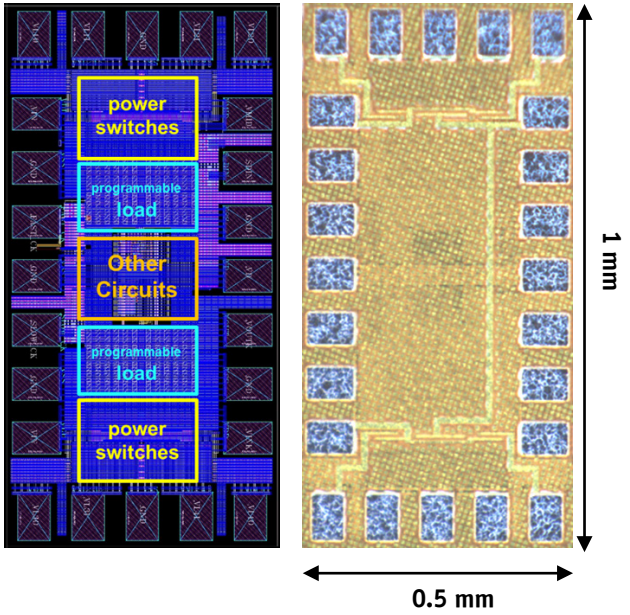
Performance Comparison Between IVRs

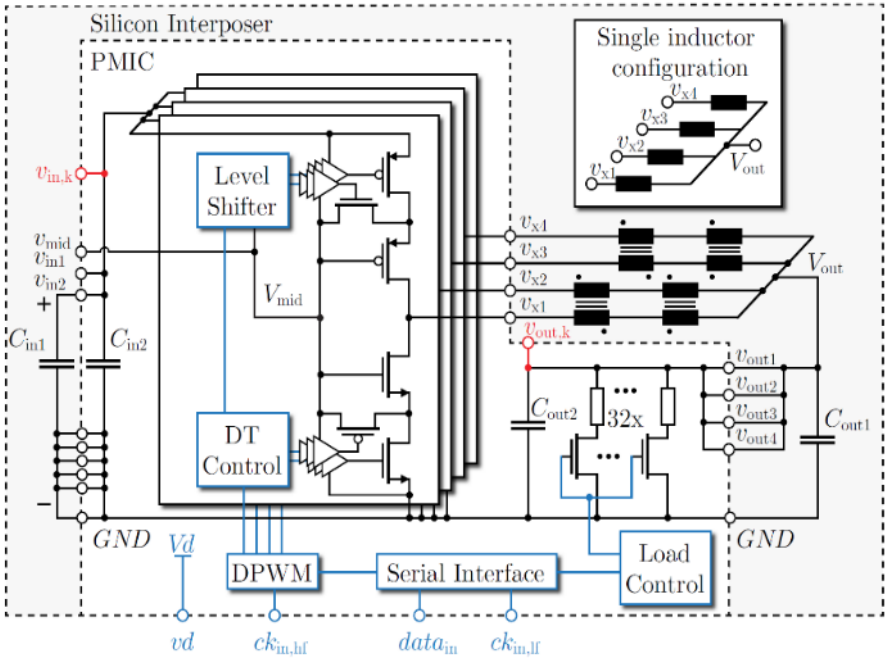


- 90% efficiency achievable with 0.2 W/mm² more power density using the proposed ANPC HB and 14 nm technology
  - Switches and power stage are more efficient for high frequency operation
- Selected inductance and switching frequency at 90% efficiency:
  - 32 nm Conventional HB:
    - 51 nH @70 MHz
  - 14 nm Proposed ANPC HB:
    - 16 nH @160 MHz

Implementation: PMIC

- PMIC in 14 nm CMOS process
- Four-phase interleaved ANPC buck



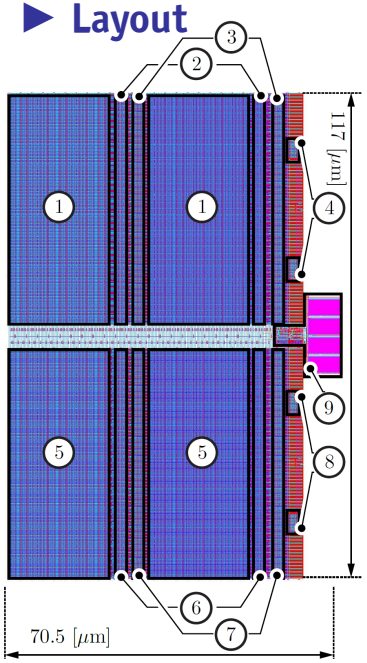


- Versatile open loop converter for better PMIC characterization
- Compatible with single and coupled inductors

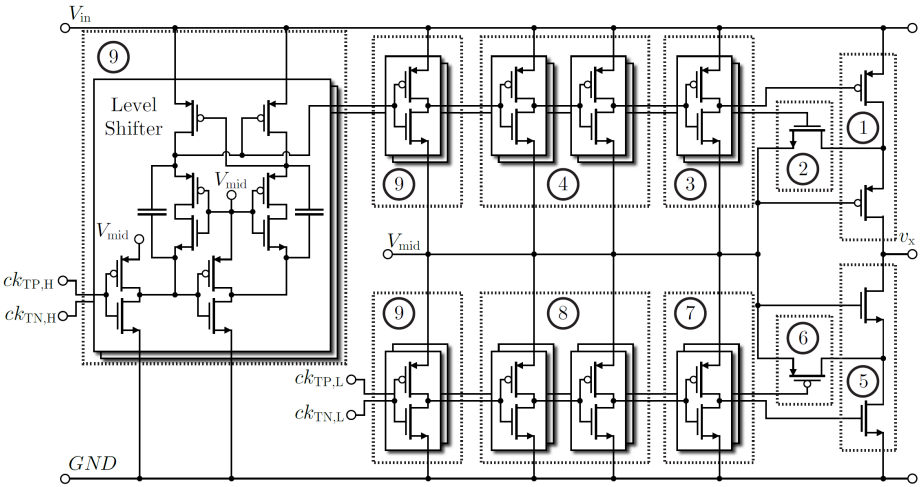
Implementation: Active Power Stage

GLOBALFOUNDRIES

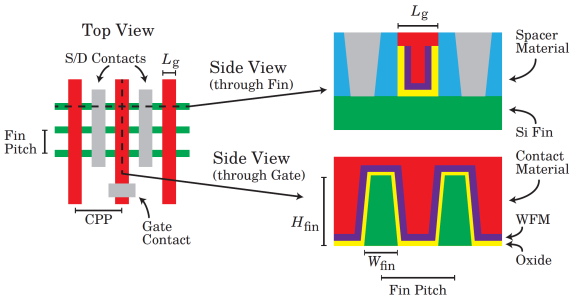
▶ Layout



▶ Schematics



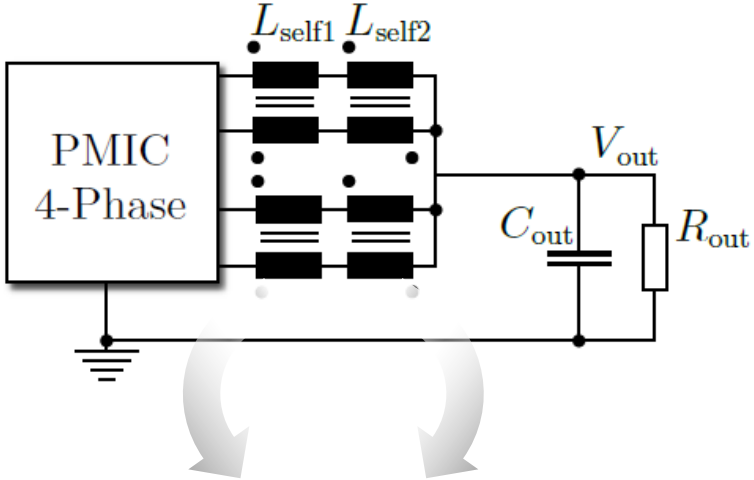
▶ Finfet concept

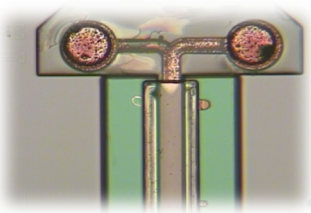


- ▶ Globalfoundrie's 14 nm Bulk CMOS
- ▶ Finfet 3D transistors optimized for digital circuits
- ▶ Design uses exclusively low voltage devices

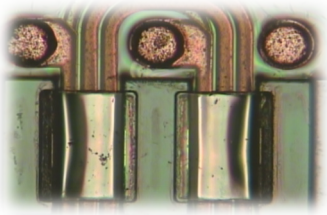
Implementation: Inductors

▶ Coupled inductors with magnetic core





▶ Strip-line



▶ Racetrack

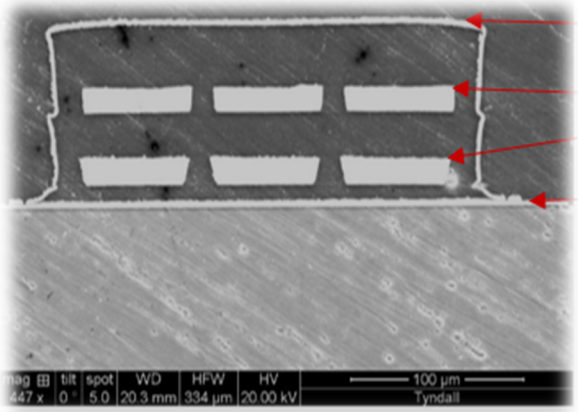
▶ 150 MHz design

L_{Self2}	= 14.4 nH
k_2	= - 0.95 (k_{max})
L_{Self1}	= 1.6 nH
k_1	= 0.95 (k_{max})

▶ 100 MHz design

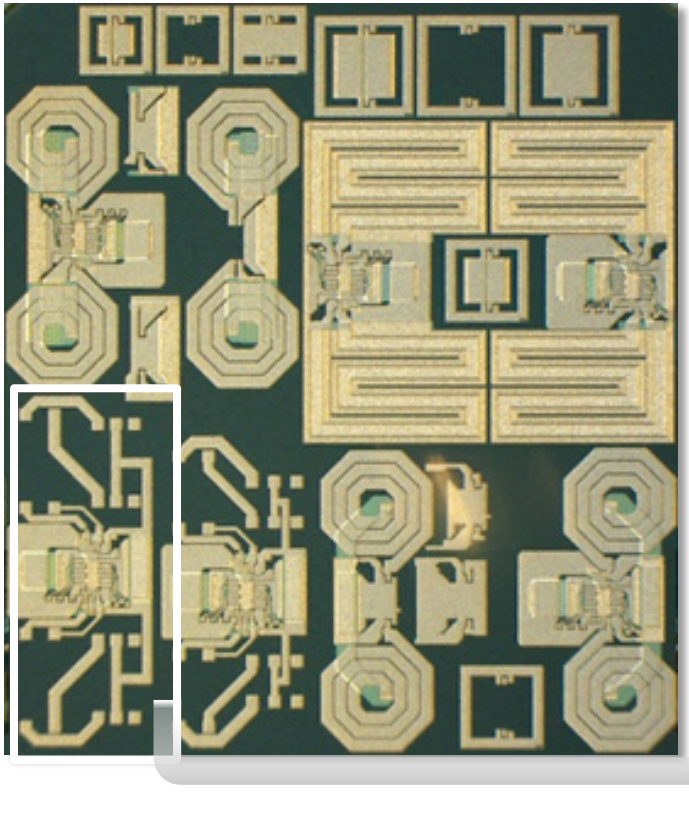
L_{Self2}	= 21.6 nH
k_2	= - 0.95 (k_{max})
L_{Self1}	= 2.4 nH
k_1	= 0.95 (k_{max})

▶ Cross section view

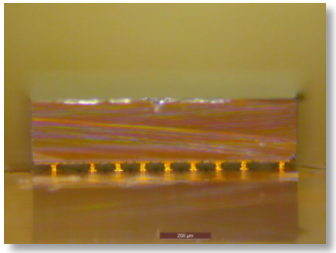


Implementation: All-silicon-based Demonstrators

▶ Interposer

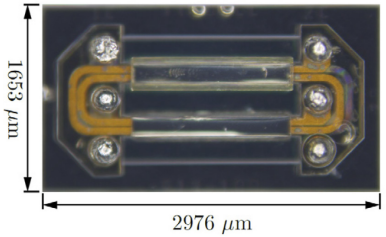


▶ 14 nm PMIC

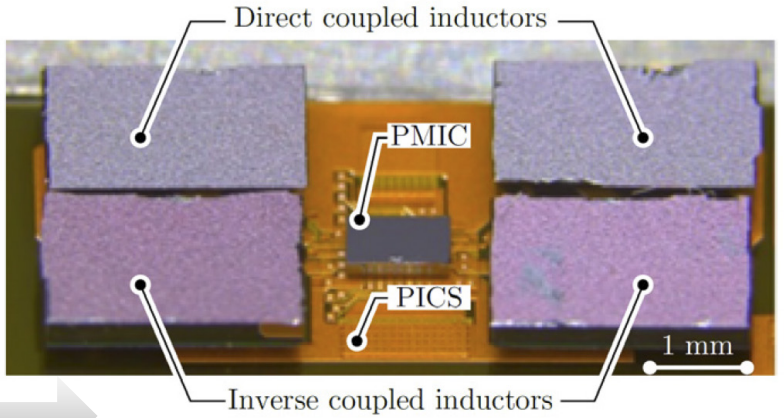


GLOBALFOUNDRIES

▶ Coupled inductor (Inv.)



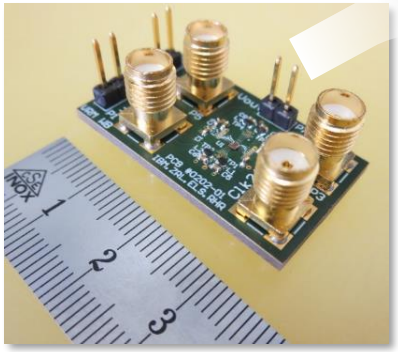
Tyndall National Institute



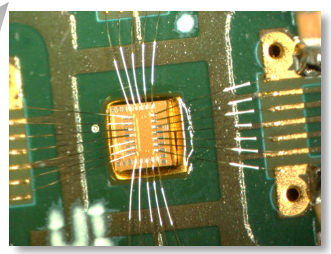
Fraunhofer IZM

PCB-based demonstrators and Experimental Results

▶ Demonstrator

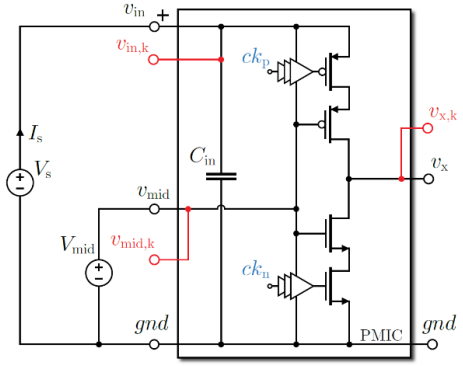


▶ Wire-bonded PMIC

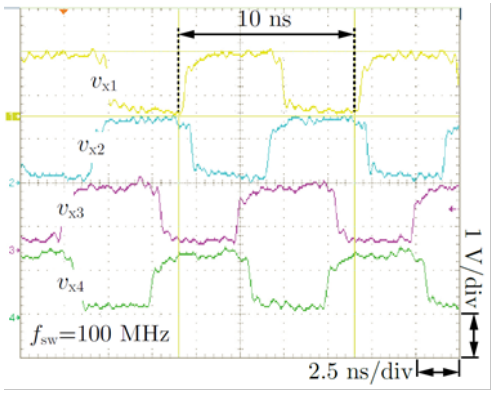


▶ Passive devices

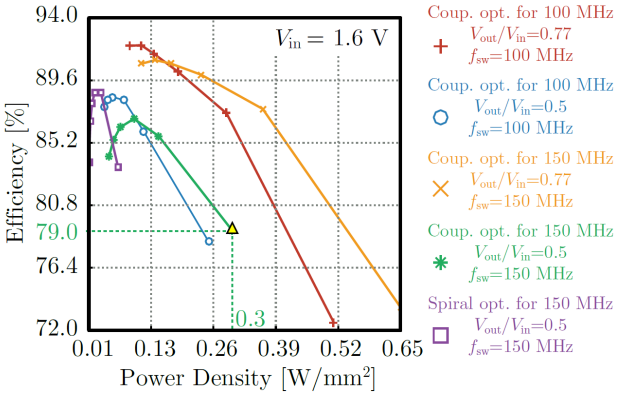
▶ Switching phase



▶ Switching node



▶ Efficiency estimation



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!**