A Low Power System-on-Chip with Memory Stacked on Top of Logic

By *Kristof Blutman*

(Student No. 4236416)

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering

at the

Electronic Instrumentation Laboratory, Faculty Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Netherlands

Supervisors:
Prof. Dr. Ir. Kofi A. A. Makinwa[1], Ir. Ajay Kapoor[2],
Prof. Dr. Jose Pineda de Gyvez[2,3], Dr. Ir. Arnoud van der Wel[2]
[1] Electronic Instrumentation Laboratory, Delft University of Technology
[2] NXP Semiconductors Nederland B.V.
[3] Electronic Systems Group, Eindhoven University of Technology

October, 2014

The undersigned hereby certify that they have read and recommend to the Faculty of Electrical Engineering, Mathematics and Computer Science for acceptance a thesis entitled
A Low Power System-on-Chip with Memory Stacked on Top of Logic
By
Kristof Blutman
in partial fulfillment of the requirements for the degree of
Master of Science in Electrical Engineering

Dated: October 29th, 2014

Chair:

Prof. Dr. Ir. Kofi A. A. Makinwa

Committee Members:

Prof. Dr. Jose Pineda de Gyvez

Dr. Ir. Nick van der Meijs

Ir. Ajay Kapoor

ABSTRACT

This work introduces a low-power Cortex-M0+ based computing platform for battery-powered, embedded applications. Voltage stacking is used to save power by recycling charge through a power domain between 0V and V_{dd}, and a power domain stacked on top of it between V_{dd} and $2V_{dd}$. The technique enables connecting chips of the future directly to the main power source. This increases the power efficiency and the power density of the power delivery scheme. The needed special circuitry components like level shifters and voltage regulators have been designed and integrated into a standard digital SoC flow to demonstrate that voltage stacking can be used for any digital system. For comparison and also functional purposes, the designed system is reconfigurable between the conventional, high throughput, flat mode where all the power- and ground rails are common, and the low power, stacked mode where the power domains are stacked on top of each other. A $1.44\mu m^2$ test chip has been fully designed and is to be fabricated in a 40nm CMOS process to evaluate the concept. Pre tape-out simulations show that the power efficiency of the system improves by 15% from 79.5% in the flat mode to 95% in stacked mode, while running a typical benchmark program in the Cortex-M0+ core at 80MHz clock frequency. The system power density for the same test case improves from $10.5mW/mm^2$ to $34.9mW/mm^2$.

The research has been carried out at NXP Semiconductors in Eindhoven.

CONTENTS

FIGURES

TABLES AND EQUATIONS

ACKNOWLEDGEMENT

This work would not have been possible without the endless support of several people. In the first place I would like to show my appreciation to my supervisors for guiding me day by day on the path from concept to realization; Ajay Kapoor, Professor Jose Pineda de Gyvez and Arnoud van der Wel at NXP Semiconductors and Professor Kofi Makinwa at the Delft University of Technology.

I would also like to thank Dr. Nick van der Meijs for agreeing to be a member of my thesis committee and reviewing my work.

Next, I would like to express my acknowledgment to those at NXP and TU Delft on whom I could count in any case when I got stuck in the project. The list is very long and it would be impossible to mention them all, but among others I am in great dept to Leo Sevat, Juan Diego Escobar, Sebastian Fabrie, Hamed Fatemi, Vibhu Sharma, Surendra Guntur, the whole MCU TP Innovation Center in Eindhoven and MCU-IC in Singapore. Other people who have given me a lot of valuable feedback at NXP are Shawkat Zaman, Ravi Karadi, Iris Bominaar-Silkens, Maarten Vertregt, Johan Verdaasdonk, Jianghai He, Marcello Ganzerli, Robert Rutten and Vladislav Dyachenko, and all others who are not mentioned here. Great ideas came from the members of the Precision Analog Group at TU Delft who have helped me to overcome the hardest problems and provided me advices from a different point of view.

Finally, my special gratitude goes to my parents for bringing me to this world and raising me up with the greatest care, to the rest of my family and my friends for always being there for me when I need them, and to the most special person for her patience and love.

Kristof Blutman

Eindhoven,
22nd October 2014

1 Introduction

The *Industrial Revolution* has turned out to be only the first step in human development. From the second half of the 20th century continuing to this date, we live in another era, often coined as *Information Revolution* [1]. With the invention of the integrated circuit (IC or microchip) by Jack Kilby in 1958 [2], it has become possible– just like the mass-production of goods and mass-generation of *energy* in the Industrial Revolution – to deliver and process *information* on a mass scale.

Since the Information Revolution greatly relies on ICs, its dynamics can be described by *Moore's law* [3], which states that on a microchip the number of devices per area doubles about every two years. There have been similar scaling laws introduced involving various physical quantities, but one particularly interesting is the prediction that the power consumption per unit area remains constant (*Dennard scaling* [4]). This law was valid for applications like computers that required the highest possible throughput with little regard to power, since the advancement of technology nodes ensured extreme power reduction in the so called *"golden days of scaling"* [5].

The reason why *Dennard's law* requires attention is because even though it had been valid since its introduction in 1974, it appears to have broken down in the mid-2000s [6]. One reason for this breakdown is the ignored secondary power components within an integrated circuit, primarily leakage. Leakage is caused by the fact that the MOS transistor threshold voltage cannot keep the pace in scaling with the reduction of supply voltages [7]. The diminishing advantages coming from the advancement of CMOS process do not just affect the traditional applications like high-end CPUs where overheating has become the main bottleneck, but also limit the emerging applications like battery-powered systems [8], where the finite battery capacity is currently the main problem. Since battery technology cannot keep up with the increasing energy requirements, the problem has to be addressed from the microchip side of the system, in the form of a design principle called *low-power design methodologies* [9].

Aiming to reduce dynamic and static power, there have been various schemes introduced on all design levels [5]. Low-power design considerations are employed from the device level including multiple threshold voltage devices [10], thick oxide devices, through the circuit level with reverse body-biasing [11], voltage islands [12] to the system level like dynamic voltage and frequency scaling [13], clock gating and power-shutoff techniques [8]. These levels of optimization can be used simultaneously, which results in extra power savings.
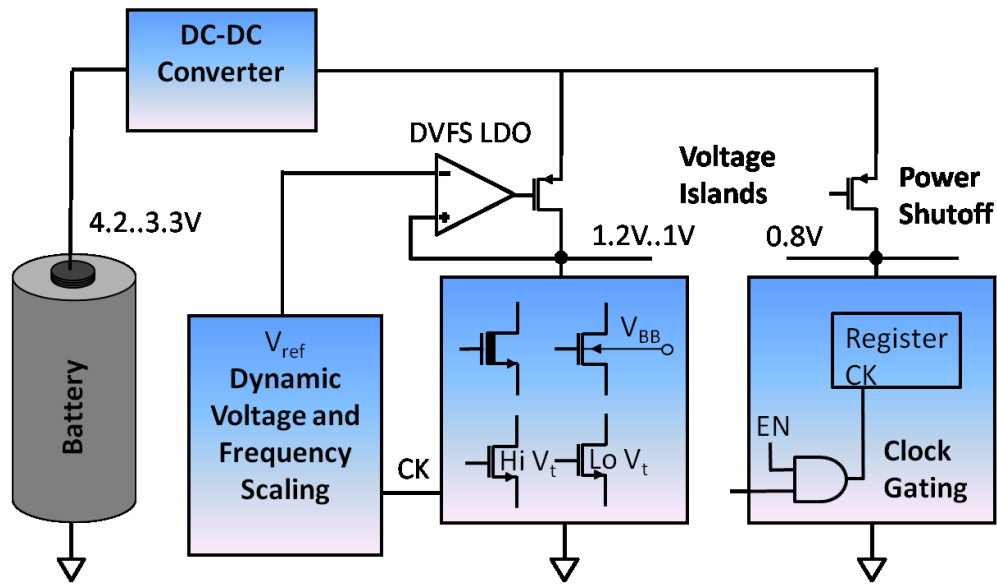
Figure 1-1: Overview of low-power techniques

As it is visible from **Figure 1-1**, there is an abundance of power optimization possibilities already explored. Starting from the power delivery with the battery, which hardly keeps track of the CMOS technology, through the DC-DC converter, where power density and efficiency constraints apply, till the actual architectural-, circuit- and device level within the ASIC. The question arises, whether we can do more to reduce the power. This thesis addresses this problem and proposes a low power design technique that is parallel to the mentioned methods by stacking power domains on top of each other.

1.1 Voltage Stacking and the Charge Recycling Principle

Charge recycling, as its name suggests, is the concept of reusing electric charge dumped during certain phase of operation of an electronic circuit. The electric potential energy of charge which

otherwise would have been wasted, can be re-used during a different phase. There are several ways to benefit from this principle, including leakage management [14], memories [15][16], data converters [17], logic circuits [18], and a lot more [19][20][21].

The voltage stacking method [22] is also a way to exploit charge recycling. The scenario on **Figure 1-2** shows a system in the conventional or flat (left) and the stacked (right) case.
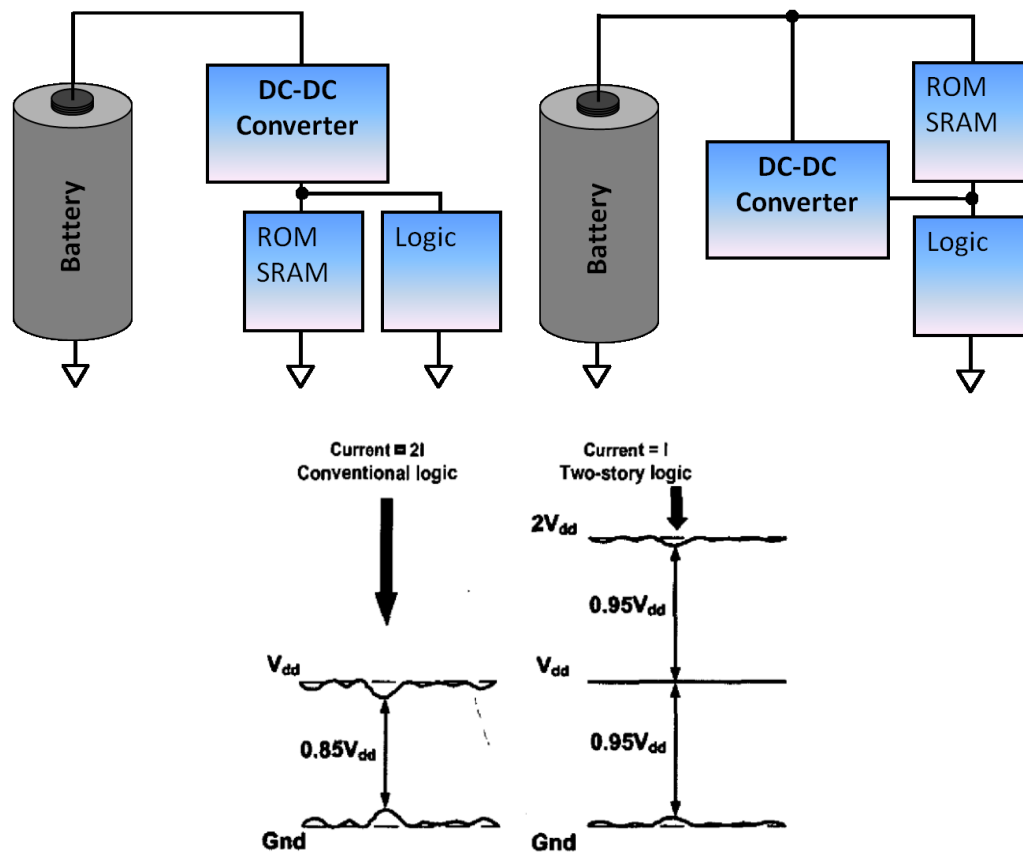
Figure 1-2: Flat (left) and Stacked (right) power delivery [22]

While usually there is a single power and single ground rail that sources and sinks the current, respectively, in principle it is also possible to split the design into two equal parts, and raise the ground and the power voltage of one half part by a supply voltage (V_{DD}) value. Doing so, it has been observed that

1. The power rail of the bottom half has to source exactly the same amount of current that the ground rail of the top half has to sink ($-I_{top,ground} = I_{bot,power}$)

2. The voltages of the mentioned rails are the same (V_{DD}).

Connecting these two nodes will thus satisfy Kirchoff's current and voltage laws (KCL and KVL). The current used by the top domain is re-used by the bottom domain, implementing a simple charge recycling scheme. Of course, here we have ignored the fact that the two stacked power domains will most likely consume a different amount of current, and that the matching will be always approximate ($-I_{top,ground} \approx I_{bot,supply}$). Later we will see how this degrades the efficiency of the charge recycling scheme.

The benefit of voltage stacking is in doubling the voltage and halving the current the system needs. This can relax the requirements set for the power delivery scheme. In most of the cases, voltage converters are necessary to bridge the power source and an integrated circuit. It is enough to think about the way electricity is transported nowadays – from the power station, it is converted up to hundreds of kilovolts to minimize IR losses, to be gradually transformed down to the European standard 230V or U.S. standard 120V. A similar scenario is present with battery-powered system. To maximize the stored energy, batteries cannot scale down their voltages to the order where microchips operate, since then power density would shrink. Instead, their voltage (3.6V for Li-ion batteries) is converted down by some combination of switched-mode and linear power converters to the desired voltage (IC core voltage around 1V). Unlike for electric power lines, though, the conversion is not an AC/AC but a DC/DC conversion, thus it is constrained by the battery that requires high efficiency for a long discharge time. This imposes a challenge for voltage converters that need to deliver higher power each year by rapidly increasing current consumption, but decreasing voltages due to Moore's law [23]. By stacking circuits as many times as necessary to reach the battery voltage, the regulators face relaxed requirements in this aspect.
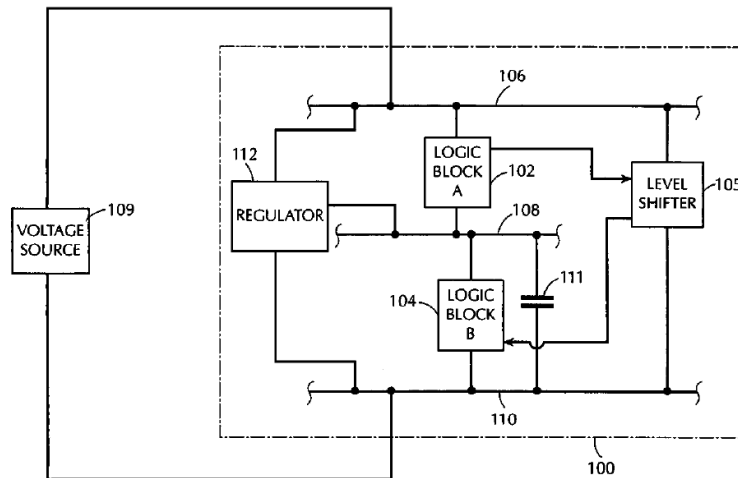
Figure 1-3: Generic Stacked System [24]

Figure 1-3 shows a possible system benefiting from voltage stacking [24]. We can see that the battery (VOLTAGE SOURCE) directly powers the circuit, while the regulator's role has changed from delivering all the power to the circuit to only guarding the rail between the two partitions for the case when there is an imbalance in the current consumption of the top domain (LOGIC BLOCK A) and the bottom domain (LOGIC BLOCK B). This has two important implications.

1. The power requirements of the regulator have decreased since most of the power is delivered directly into the system from the battery, and only the difference in the power consumption of the top- and the bottom domains has to be provided by the regulator. This implies that smaller regulators can be made, or more power can be delivered for the same area in today's increasing power delivery requirements. Thus, the *effective power density* of the regulators has increased.

2. On the other hand, since most of the power comes from the battery, without the need to be regulated, eliminating power losses that otherwise would be present if a regulator would provide all the power. Thus, it can be stated that the major part of the power is converted at 100% efficiency, because it does not pass the voltage regulator. Thus, the *effective efficiency* of the voltage regulator has also increased.

These observations mean that by using stacked circuits, it is possible to deliver power with less area overhead and save power through boosting the efficiency. These two benefits lift some of

the pressure on industry to fulfill Moore's law for reducing sizes which also requires relaxing the power dissipation density, and also reducing the total energy consumed for battery-powered applications.
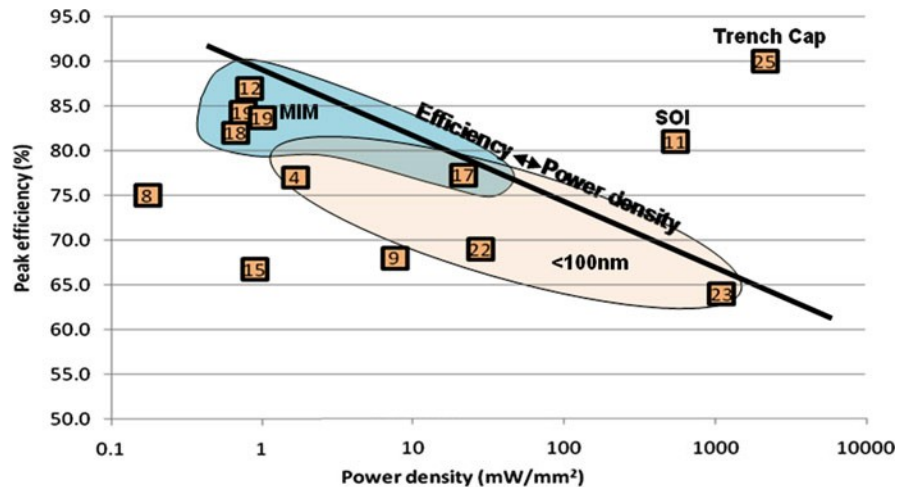
Figure 1-4: Power density and efficiency limits [43]

Consider **Figure 1-4** which shows the tradeoff between power density and power efficiency for various voltage regulators. Conventional voltage regulator design allows the improvement of one at the cost of the other, for a given technology. Improving both these quantities would overcome the technology limitations of power delivery. In the following, the focus will be on low-power battery-based systems. This implies that stacking voltage domains is not just a novel low-power technique that can be applied in parallel to other techniques, but also that it overcomes serious technological limitations. Thus, it is worth investigating and applying it for battery-powered systems.

1.2 Applications

Low power design is a must for various computing platforms from high-end CPUs to low-power energy harvesting sensor nodes. While the former application suffers from thermal limitations, the latter is constrained by the battery capacity. From a systems design perspective, the ASIC designer usually employs a non-conflicting subset of the known low power techniques. Voltage stacking, since it does not reduce the power at the circuit level but rather at the system level,

has the potential of reducing significant power while allowing other low power techniques being applied simultaneously, without conflict. Thus the research described in this thesis can be employed within a wide variety of systems, however, the emphasis will be placed on the emerging applications that suffer from power limitations, .e.g Internet-of-Things (IoT) [26]. The connection of "things" with each other requires a lot of overall computation, which translates into power. Chips like wireless sensor nodes need to operate efficiently in their analog, mixed-signal, RF and digital parts due to the limitation on power sources, often relying on a battery, or, in the even more limited case, harvested energy stored on capacitors [27]. As a conclusion, lifting the power limitations partially by efficient power delivery through voltage stacking could boost the development of IoT applications.

The possible applications of voltage stacked circuits are not restricted to IoT. Though not explored in this research, on the high performance side, multi-core systems can greatly benefit from voltage stacking by placing different cores in different stack domains, where the current balancing could be done from software. Power efficiency is also important in various other applications e.g. mobile devices like smartphones and tablets through automotive or medical electronics to wearable electronics or contactless smartcards.

1.3 Organization of this thesis

The prior art is addressed in Chapter 2. Chapter 3 describes the system level design of the proposed test chip, and Chapter 4 goes into details about the implementation. The preparation for chip measurement is described in Chapter 5, and the conclusions are given in Chapter 6.

2 Prior Art

The main goal of this research project is to apply the known principle of voltage stacking to a realistic system-on-chip. In order to make the proposed test chip useful, it is important to consider the existing low-power solutions present today, examine their capability for reducing power and to investigate ways for improvement. In the following, a thorough review is given in the context of low-power embedded systems. Beginning with microcontroller systems and the CMOS process used to realize them, batteries as source of power are reviewed, followed by a discussion on power delivery through voltage regulators. Finally, the current chapter introduces power delivery through stacked circuits, and concludes with the specifications of the proposed system that improves the state-of-the-art.

Expectations today are that everything is intelligent (it is enough to look at the naming of the emerging applications like smart sensors, smartphones and smart devices in general), and that everything is connected (mobile telephony, Internet-of-things). These requirements and the continuation of Moore's law have driven the ASIC designers towards integrating more and more complex functionality onto silicon chips to enable higher computation throughput and faster, better communication. The fact that *everything* has to be smart and connected has given the space for embedded applications that employ low-power microcontrollers. This is needed energy is limited once so many devices are placed at different locations within an environment.

A microcontroller system typically employs a processor core for computation and peripherals for communication, clock generation, data acquisition, etc. The core is the 'brain' of the system in a sense that it controls the peripherals, the 'body', to behave according to the desired functionality. There are different microcontroller cores for different applications, the difference being mostly the computational performance and power consumption. 32-bit RISC (Reduced Instruction Set Architecture) processor cores are a suitable demonstration vehicle for modeling a typical battery-powered microcontroller system. While 8-bit architectures are still widely used today, the benefits of 32-bit cores have been demonstrated since the latter need shorter

active operation mode and consume less energy to perform the same task [31]. The Cortex-M series from ARM Holdings is an example for 32-bit processors, while 8-bit microcontrollers include the AVR series from Atmel Corporation, or the PIC family from Microchip Technology.
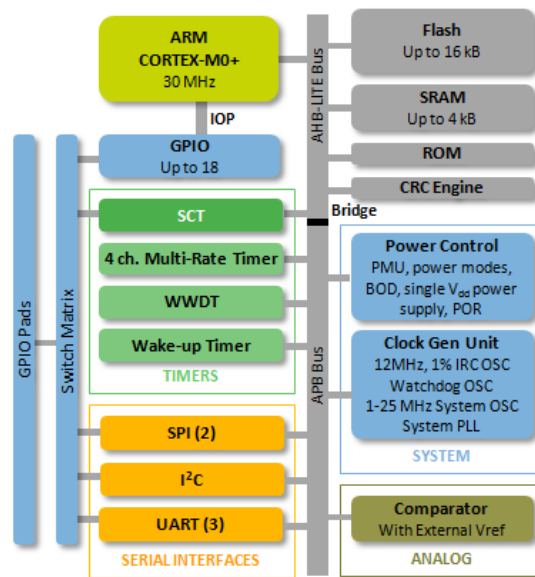
Figure 2-1: LPC800 microcontroller block diagram

In **Figure 2-1** the generic block diagram of the NXP LPC800 series is sketched, where an ARM Cortex-M0+ core is employed [32]. This microchip also features several peripherals which are custom-implemented and are not dependent on the core used. This way the microcontroller can be fit for the application. The LPC800 contains up to 16kB Flash memory and 4kB SRAM, USART, SPI and I2C interfaces for communication, timer circuit, IO modules with switch matrix and clock generation unit. These are either designed by the first party that produces the ASIC, or they are provided by third parties. The LPC800 microchip is intended to be used in environments where power-efficiency is crucial e.g. battery-powered applications.

2.1 CMOS Process

Once the architecture of the microcontroller is chosen, its silicon implementation follows. Though there are emerging alternative technologies, today's digital integrated circuits almost exclusively rely on the CMOS (Complementary Metal-Oxide Semiconductor) technology [33]. The advancement of the CMOS process can be described in the simplest case with a single

quantity, the gate length of the MOSFET devices. With a new technology this dimension shrinks down to 14nm for SRAM [34], dictated by Moore's law. Aggressive scaling of CMOS technology nodes has resulted in dramatically reduced area and power consumption. There have been, however, some unintended side effects. One of the most pressing problems is leakage. With each new process node, the dynamic power consumption is reduced [35], but the total power is more and more dominated by the leakage power [36].
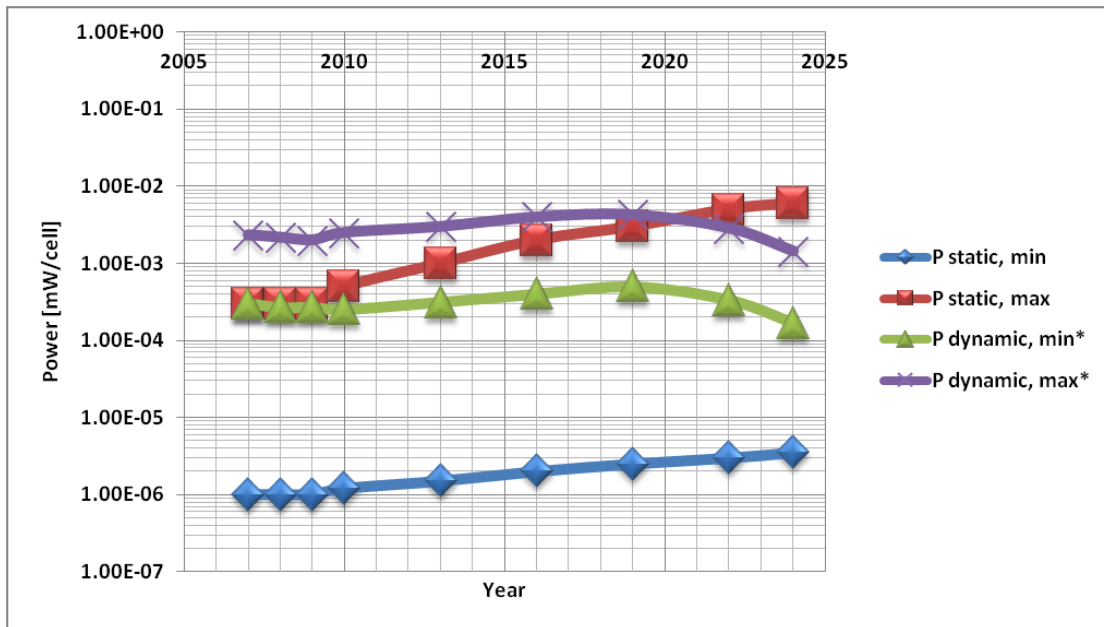
Figure 2-2: Static and dynamic power consumption data and prediction per SRAM cell [35]
(* at max. frequency)

Within one CMOS technology process, there is limited space left for reducing the fundamental power consumption. The dynamic power mostly comes from switching power, where the task performed is the following. A given C capacitance (e.g. MOS capacitor) has to be charged to a voltage V (e.g. supply voltage) within time T (period), through an interconnect resistance R, which can be the wiring and switching resistance added together. Since the circuit charge $Q = CV$ has to come from the power supply which has a constant voltage, the energy consumption will be $E = V_{dd}Q = V_{dd}CV$.

The question might arise, whether this is really the lowest possible energy spent to charge the capacitor. If we assume full charging $V = V_{dd}$, half of the energy is wasted on the interconnect

resistance, irrespective of the value R. Is it possible to increase the efficiency above 50%? To find the minimum energy needed, we can express the energy in terms of charging time T, source current $i_S(t)$ and source voltage $v_s(t)$, then minimize the quantity.

$$E = \int_0^T i_S(t) \cdot v_s(t)dt = \int_0^T \dot{q}_c(t) \cdot \left(\dot{q}_c(t) \cdot R + \frac{q_C(t)}{C}\right)dt$$

Equation 2-1: The energy dissipated when charging a C capacitor in T time with R interconnect resistance

The capacitor charge $q_C(t)$ has known initial and final values, $q_C(0) = 0, q_C(T) = CV$. Thus, the Lagrangian and the Euler-Lagrange equation of this minimization problem are

$$Lagrangian: \mathcal{L}(\dot{q}_c(t), q_C(t)) = \dot{q}_c(t) \cdot \left(\dot{q}_c(t) \cdot R + \frac{q_C(t)}{C}\right)$$

$$Euler - Lagrange\ equation: \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_c(t)}\right) - \frac{\partial \mathcal{L}}{\partial q_C(t)} = 0$$

$$Substitution\ of\ the\ Lagrangian\ yields: \ddot{q}_c(t) \cdot R + \frac{q_C(t)}{C} - \frac{q_C(t)}{C} = 0 \Longrightarrow \ddot{\boldsymbol{q}}_c(\boldsymbol{t}) \cdot \boldsymbol{R} = \boldsymbol{0}$$

Equation 2-2: Calculating the minimum-energy conditions with the Euler-Lagrange method

The solution for $q_C(t)$ in the minimum energy case, that is, when the Euler-Lagrange equation is satisfied, suggests constant source current of $\dot{q}_c(t) = CV/T$, instead of constant charging voltage. Calculating the minimum energy using this solution of the Euler-Lagrange equation yields

$$E_{min} = \int_0^T \frac{CV}{T} \cdot \left(\frac{CV}{T} \cdot R + \frac{\cancel{C}Vt}{T}\frac{1}{\cancel{C}}\right)dt = \left(\frac{2RC}{T} + 1\right)\frac{CV^2}{2}$$

Equation 2-3: Minimum energy required to charge a capacitor

This is the adiabatic charging limit [37] and it is not possible to cross it due to energy conservation. Furthermore, achieving this limit requires higher voltage than V to be present in the circuit. This quantity is thus the ultimate red brick wall of energy consumption for conventional CMOS process. It shows that apart from the useful energy stored on the capacitor, an amount of energy inversely proportional to the charging time will be dissipated on the resistors during a charging event. To compare the minimum energy with the real energy spent

to charge the capacitor, we assume that the charging is slightly incomplete in both cases, and the voltage is

$$V = V_{dd}\left(1 - e^{-\frac{T}{RC}}\right)$$

In this case, the ratio of the real energy versus the minimum energy is

$$\frac{E}{E_{min}} = \frac{V_{dd}CV}{\left(\frac{2RC}{T} + 1\right)\frac{CV^2}{2}} = \frac{2}{\left(\frac{2RC}{T} + 1\right)\left(1 - e^{-\frac{T}{RC}}\right)}$$

Equation 2-4: Energy dissipation compared to the possible minimum

It is important to note that operation speed at or above $T \approx RC$ is not feasible, since the capacitor could only charge up to 63% of the supply voltage. If the frequency is low, i.e. $T \gg RC$, the minimum energy asymptotically approaches the energy stored in the capacitor, $CV^2/2$. It seems thus that half of the energy could be saved, which is expected since the efficiency becomes 100% once only useful work is done ($CV^2/2$). However, at low frequencies, on one hand the performance of the circuit is very low, and on the other hand, the leakage energy per switching event linearly increases, assuming the leakage current to be constant, and the leakage energy being $P_{leakage} \cdot T$. To maximize performance and energy efficiency, thus, it is desirable to operate with T being definitely above, but not by orders of magnitude above, the time constant RC.

From the dynamic energy E_{dyn} of one switching event, it is possible to calculate the average energy with a new quantity called switching activity, denoted with α. This quantifies what percentage of the clock period the device is switching. For a clock signal, for example, it takes the value $\alpha = 1$, while for a constant signal, $\alpha = 0$. The average energy consumption is then αE_{dyn}. The relation between average power consumption and average energy per switching event is straightforward, the former can be derived from the latter multiplying it with the clock frequency $f = 1/T$ which yields the familiar formula $\alpha f E_{dyn}$. For CMOS circuits, the average dynamic energy is $V_{dd}CV \approx CV_{dd}{}^2$, and the power becomes $\alpha f CV_{dd}{}^2$. Thus the energy of one switching event can be made directly proportional to switching power, and the same limitation holds for both pointed out in **Equation 2-4**.

We can conclude from the above findings that *fundamentally*, it is not possible to reduce the dynamic switching power consumption of a given e.g. MOSFET device for a normal operation frequency by more than a factor of 2, and when optimizing the overall energy per switching of one device, it is heavily compromised with the static power coming primarily from leakage. This limited headroom for power reduction calls for new approaches. To find an alternative to the conventional low power techniques, we need to explore other mechanism where the power is lost, examining the power delivery from the battery to the circuit.

2.2 Battery

One primary reason for the necessity of low power design is the fact that battery technology could not keep up with the CMOS process scaling in terms of power density. This has been a problem starting as early as the beginning of the1990s [38]. This is due to the fact that battery science relies on chemical reactions, which are in turn limited by physical laws. Aiming to come up with a new battery technology, novel chemical reactions must be employed, while keeping the energy density in balance with the reliability and safety of the device as well as the fabrication costs [39]. It is difficult to explore the design space of the batteries for optimizing both the energy density and the reliability, since the chemical reactions are governed by quantum mechanics and cannot be engineered in the sense integrated circuits scale [40]. In selecting the optimal chemical reaction, battery-powered autonomous devices favor compact, low weight and -volume solutions that are rechargeable. In the **Figure 2-3** such batteries are considered. If one observes the year of invention of these batteries, and keeps in mind that these are state-of-the-art solutions, one can have insight into at how slower pace battery technology advances compared to CMOS process. It can be seen from **Figure 2-3 (b)** that the direction of battery research has been focused at increasing the energy density (Wh/L) and energy capacity (Wh/kg). In some cases this has resulted in increasing output voltage like in the case of Lithium-ion batteries. While low-voltage, high-energy-density batteries like Nickel metal hydride and Zinc-Air exist, they have several practical disadvantages that make it very difficult to apply them in today's embedded devices [41]. While the former has corrosion problems, and is built from heavy materials, the latter is not fully rechargeable as it needs replaceable

electrodes. So even though they are expensive and use toxic materials, Lithium-ion batteries are dominating.
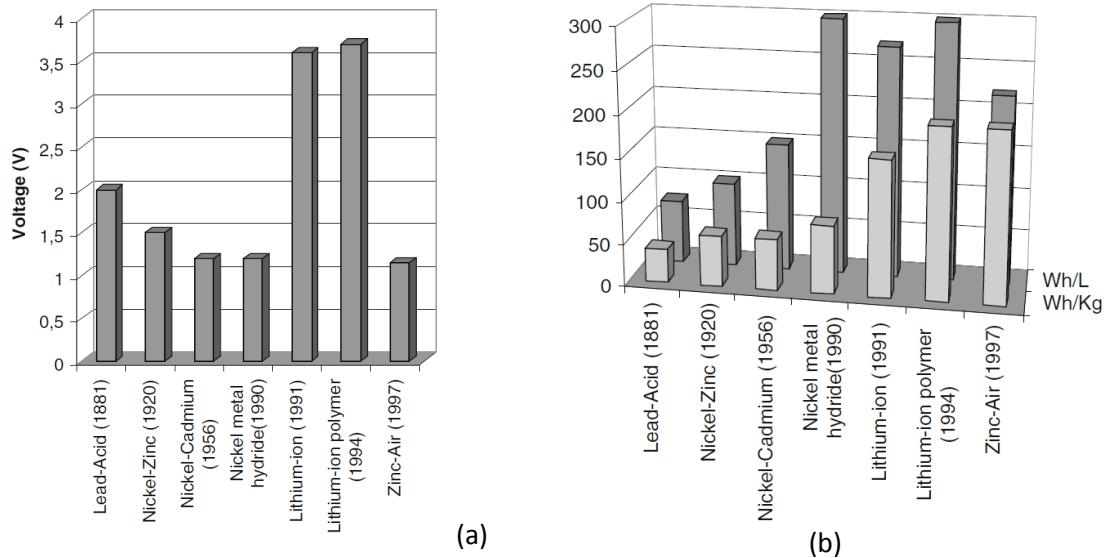
Figure 2-3: Battery voltages vs. power density [42]

To maximize the discharge cycle, batteries often are operated by a Battery Management System (BMS, [44]) which requires accurate modeling of the battery lifetime. Thus, next to finding the optimal battery and implementing an efficient battery management scheme, the modeling of batteries also imposes a challenge. There have been several battery models proposed which account for more and more parameters and quantities such as temperature effects and capacity fading [45]. In the following short analysis, we will stick to the simplest ones.

Let us assume that we have a battery which serves as a power source for our system. The battery delivers power at approximately constant output voltage V_{bat}. If we assume the load consumes the rated current level, the rated discharge time will be

$$T = \frac{Q}{I}$$

Equation 2-5: Rated battery discharge time
T – rated discharge time, I – rated current, Q – rated capacity

Now assume that we have a resistance R representing the load to the battery. We would like to increase the discharge time of the battery. One option would be to increase the capacity Q. We have to keep in mind, however, that the energy density of batteries is a limited number, so our battery size would grow which contradicts today's trend of miniaturization. If we take the capacity as a constant, the second parameter to adjust is the load current. This is determined by the load resistance, which in our case is R. Reducing the load would mean that we either decrease the size of our system, or employ the conventional low power techniques. The latter however has its own limitations, as was proven in Section 2.1. It seems that despite our efforts, we are unable to increase the discharge time. There is, however, a third parameter which can be adjusted, and that is the battery voltage V_{bat}. Increasing it by adding new materials like the Lithium-ion solutions can increase the power density. For most of the portable applications however (smartphones, tablets, etc.) the voltage that the load needs (typically around 1V) is much lower than that of the battery provides (3.6V for Lithium-ion batteries). Thus, by increasing the energy density and reducing the discharge time for a same sized battery through increasing the battery voltage, the output must be regulated down to the appropriate voltage level by a voltage regulator, with certain losses. But this still has the advantage of extending the time the system can operate from the battery, especially if we consider non-ideal discharging effects. By decreasing the load current lower than the rated current, the effective capacity of the battery increases. This phenomenon is described by Peukert's law [45]:

$$t = \frac{T}{\left(\dfrac{i}{I}\right)^{k}}$$

Equation 2-6: Peukert's law
t – discharge time, i – current, k – Peukert Exponent (1<)

This equation tells us that the slower the pace a battery is drained, the longer it lasts. A 2X current reduction this way extends the discharge time by approximately 130% ($k = 1.2$), compared to the original 100%.

So far increasing the battery voltage has demonstrated beneficial properties. The drawback is the limited efficiency with which the high battery voltage can be regulated down to the level the system requires. A possible solution to this problem is to keep up with the battery voltage

and stack the system in terms of supply voltage, as mentioned in the Introduction Chapter. This means connecting the circuit directly to the battery, a technique known as Passive Voltage Scaling (PVS) [46]. Coming back to our example of R load, this means splitting it into two parallel $2R$ parts and connecting them in series instead of parallel. We need double the battery voltage and half the load current in this case. A voltage regulator now does not need to process all the power of the battery, instead it only needs to regulate the node between the two circuit stacks. This is a certain charge recycling scheme as the current used by the load on the top is also used by the one on the bottom. There will be a detailed discussion on stacked circuits in Section 2.4, but first let us turn our attention to the voltage regulators which are necessary for our system, even in the stacked case.

2.3 On-chip Voltage Regulation

A low-power system like the LPC800 in the **Figure 2-1** requires efficient power delivery, where switched-mode power supplies have a significant advantage over linear regulators [47]. On the other hand, for a battery-powered autonomous system, the drive to integrate whole systems onto silicon chips drives the voltage regulators to be fully integrated themselves, fully on-chip. For power-efficient switched-mode converters, this results considerable area overhead compared to linear regulators. Applications determine whether power or area constrains are dominating, so various systems employ either or both the switched-mode and linear regulators [48].

On-chip DC-DC converters, opposed to external DC-DC converters, face difficulties in using an inductor due to the low achievable quality factor that limits the performance. This makes the buck and boost converters employed in discrete DC-DC converters a less popular choice for on-chip power converter architectures. To reach the desired efficiency, either external, discrete components are utilized [49] or a second die is used to accommodate the inductor in a System-in-Package [50]. Instead of using an inductor, designers often turn their attention towards purely capacitor-based converters [51]. Switched capacitor DC-DC converters, as the name suggests, require only switches and capacitors on-chip. Since these circuit components are native to CMOS technology, switched-capacitor voltage regulators are widely employed.
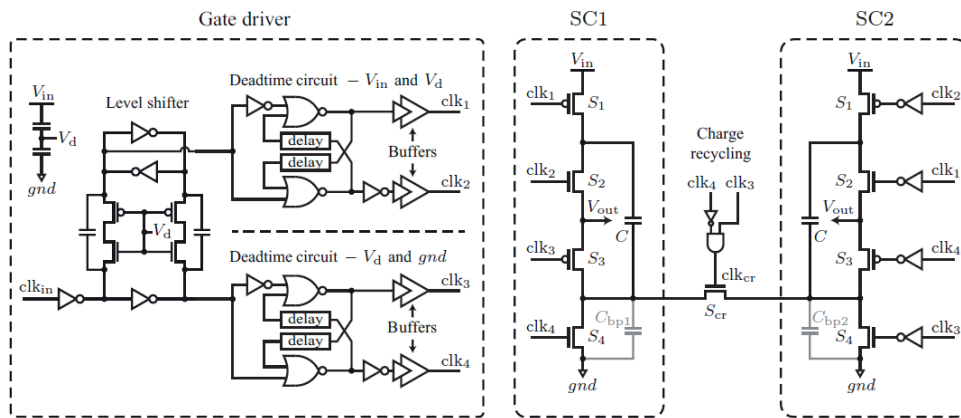
Figure 2-4: A commonly used 2:1 conversion ratio topology [55]

Several practical implementations of switched-capacitor DC-DC converters have been reported [53][55][58]. One of the simplest topologies, is the voltage halving converter with 2:1 conversion ratio, requiring only two capacitors and two phases. It is worth noting that there are several other architectures with voltage conversion ratios of 3:1, 3:2 or the combination of these [53]. Since this research project utilizes the 2:1 conversion ratio, the focus will be placed on those converters.

The relevance of 2:1 switched capacitor converters to this project comes from the possibility of using them to provide the intermediate voltage for stacked circuits, as it is highlighted in [59]. Stacking circuits can boost the efficiency of the power delivery scheme, as it will be explained in Section 2.4.

So far power-efficient voltage regulators have been discussed. In most practical applications however, due to area constraints, typically a linear regulator is used to provide the supply voltage for the chip core [60]. Even though the efficiency of a linear regulator is usually inferior to that of its switched-capacitor counterpart, even for switched-mode power supplies often a small linear regulator is used to correctly bias the circuit and ensure proper start-up [61].
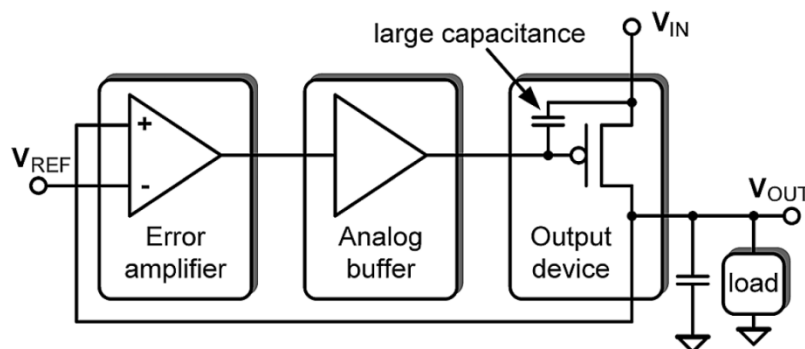
Figure 2-5: Conventional linear regulator [61]

The schematic of a conventional linear regulator is depicted in **Figure 2-5**. The circuit regulates its output to match the voltage of the V_{REF} input. It is basically a control system with the reference signal being the input voltage V_{REF}, the plant being the load circuit, the error signal the output of an operational amplifier, and the actuator a large power transistor. When there is a deviation from the desired value in the power rail of the plant, the error amplifier increases the error signal at the output, which in turn controls the power transistor to source more (or less) current. There are a couple of issues to overcome, for example the stability of the loop. This includes placing capacitors at the output and between the gate and the source of the power transistor. The figures of merits are different from switched-mode power supplies in the sense that linear regulators typically occupy very small area, thus the power density is very high, and their efficiency is limited by the formula $\eta < V_{out}/V_{in}$. The latter can be proven considering charge conservation within the regulator – the same current that flows in flows out through the output in the ideal case. The ideal case efficiency formula then becomes $\eta = P_{out}/P_{in} = IV_{out}/IV_{in}$. In a realistic case, the output current is little bit less than the input current, since the operational amplifiers require bias currents to provide sufficient gain, thus the real efficiency is the current efficiency and the well-known merits of an operational amplifier – loop bandwidth, settling error, etc.

While the linear regulation scheme discussed so far is capable of sourcing current for a typical load circuit, some applications require not just current sourcing, but also current sinking ability. This is true for stacked circuits, where the mismatch in the current consumption between the top and the bottom power domain can be both positive and negative. Thus, it is beneficial to

consider two linear regulators connected in parallel to the load, with one sourcing the current with a PMOS transistor and the other one sinking it with a power NMOS, as shown in the **Figure 2-6** [62]. The comparison can be made for these push-pull regulators as well, as is done with stacked circuits.
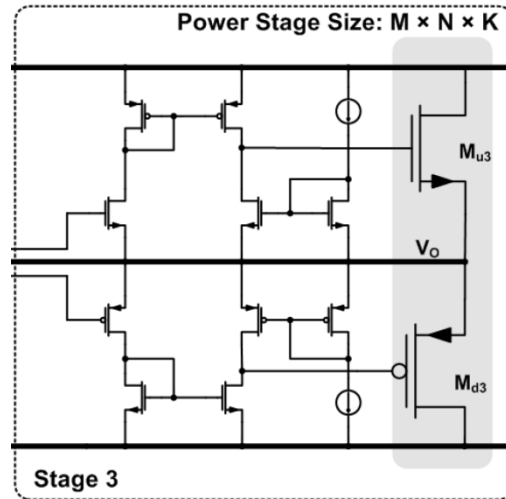
Figure 2-6: Push-pull linear regulator

So far we have seen that though a switched-mode on-chip power supply is power-efficient, it requires large chip area, and that a linear power supply is area-efficient but not power-efficient. Thus, it would be beneficial to either boost the efficiency of the LDO-s, or to organize the system in such a way that it is enough to integrate a smaller switched-capacitor converter on-chip. In the following Section, it will be argued that stacked circuits are able to fulfill both of these two requirements.

2.4 Stacked Circuits

The utilization of stacked circuits as a power delivery method for implicit voltage down-conversion is a relatively new technique. The idea can be summarized as follows. A system is partitioned into power domains where level shifters are used to provide interface for the cross-domain signals. This application of voltage islands within an integrated circuit is a well-known technique [63]. Stacked circuits can be regarded as a special application of implementing voltage islands. While in the conventional case, these power domains share the same ground

rail and separate their power rails, for stacked circuits conventionally opposite polairty rails are connected. The ground node of one of the domains is shared with the supply node of the other power domain. In this way, the current used by one domain is re-used by another domain. This charge and current recycling mechanism is responsible for the power efficiency and power density boost in stacked circuits. The reason for this is the following. If the stacked domains consume similar order of current, then that current directly comes from the external source without on-chip regulation, and is formally provided at 100% efficiency, while only the fraction of the total current is sourced from the on-chip regulator. In turn, the regulator also can be made smaller than it would be required for the conventional case. This way, even if a switched-mode voltage converter is used, it can be a fraction of the size that would be normally the case [64]. If a linear regulator is used, on the other hand, then the efficiency boost helps to keep the overall power efficiency high, compared to the traditional scenarios where all the power has to go through the less efficient linear power supply.

In the following different stacked systems will be considered from literature based on this power saving property. A stacked system that employs an efficient 2:1 voltage regulator is depicted in the **Figure 2-7**. This system can be expected to serve as an adequate demonstration of the voltage stacking concept, since it employs an efficient, while at the same time small area switched-capacitor DC-DC converter, which has its efficiency boosted . In literature, theoretical predictions claim that such a system can reach very high, over 90% efficiency on-chip, even in the case where one stack domain consumes 50% more power than the other [65].
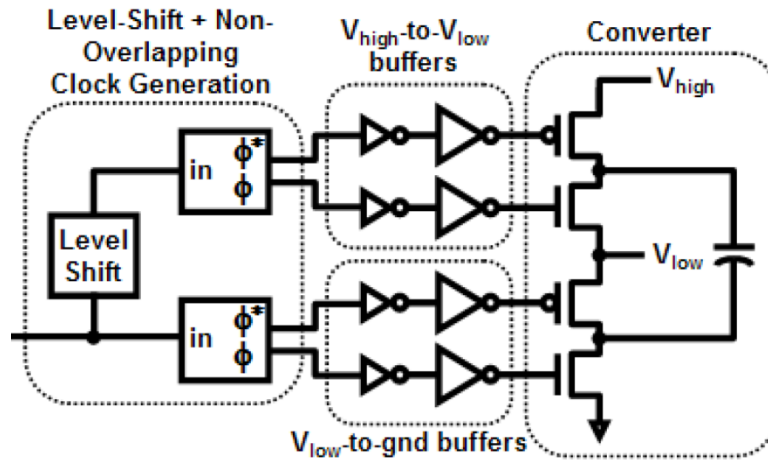
Figure 2-7: Stacked system with switched-capacitor DC-DC converter [59]

Another implementation of stacked circuits uses push-pull LDO-s to regulate the middle node of three power domains employing multipliers with different test vectors [67]. For well-matching test vectors, the configuration containing two stacked domains reaches a system power delivery efficiency over 90%, while the LDO efficiency in the conventional case would be ultimately limited below 50%. This is achieved with a stacked (also called push-pull) LDO with replica bias.

For the system described above, it is important to emphasize, however, that the large efficiency increase only holds for deliberately matched power domains – in realistic applications, one cannot rely on high matching for entire operation periods. Also, it is possible that the total energy for a period of time equals for two domains, but they are not being consumed concurrently. For this problem a large tank capacitor can be used at the middle node, as is done in [71]. There LDO-s are utilized but with a more complicated control scheme. Once the top circuit is active, an appropriate LDO opens and stores the charge dumped to the ground rail of the top domain, on a tank capacitor. When the bottom circuit becomes active, another LDO opens and uses the tank capacitor to provide the required charge. If the tank capacitor becomes 'full' or 'empty', that is, the charge stored on it would significantly alter the voltage level following the $Q = CV$ formula, other LDO-s turn on to source (sink) current directly from (to) the main power (ground) rail. With this scheme, the power delivery system can support loads which are active at different points of time, but still have a matching performance. If the

power consumption is greatly unequal, the efficiency will drop significantly. One last aspect worth mentioning that [71]is the only work that guarantees constant voltage over all the system in case of a decreasing battery voltage.

Stacked circuits with either switched-mode or linear regulators have been discussed, but it is possible to combine the two types of regulators into one common power delivery scheme. Such a system has been implemented in [66] for stacked IO drivers. Though the authors used the switched-capacitor regulator and the linear regulator separately, this scheme gives opportunity to combine them in a single operation. In the following, only a hypothetical scenario is described which could not be tested as the test chip was not at this research project's disposal.

The voltage regulation scheme can mainly rely on the switched-capacitor regulator. If the load is unexpectedly high, a push-pull linear regulator activates and the middle rail gets pulled back to the nominal range if the voltage crosses a delta value. This is ensured by the voltage margin Δ in the reference voltage of the error amplifiers. This way, the efficiency still can be kept high since the load is mostly small enough to be handled by a switched-capacitor converter, while situations where the operation of the circuit is in danger due supply drop, are handled with the less-efficient linear regulator.

It is well known that SRAM cells scale less than logic in supply voltage reduction, and tend to be the main contributor to standby power due to high leakage. As a final application mentioned here, using stacked circuits with tank capacitors can also be utilized for standby modes of SRAM memory cells [72]. Since it is difficult to provide low standby voltages for memories in sleep mode, stacking two instances by connecting one ground rail with another power rail of an SRAM matrix gives a simple way to avoid using DC-DC converters and dramatically reduce leakage power. Since the leakage power is not processed by any kind of voltage regulator, the efficiency reached 98% as there was some IR drop on the switches. Comparing this to an LDO, the leakage power saving curve is of higher order as a function of supply voltage.

2.4.1 Level Shifters

It is important to note that the level shifters used to transfer data between stacked domains impose a power and delay overhead on the charge recycling system. **Figure 2-8** shows the level shifters used in [67]. There are three parts of these cells that can be distinguished. The transmitter side is composed of two inverters which drive the devices in the 'channel'. The channel itself has two parts. An AC path with large MOS capacitors ensures high speed and low power operation, while the DC path is responsible for correct start-up conditions and that the latch in the receiving side always is in the state corresponding to that of the transmitter side. This way, there can be no errors scenarios occuring e.g. the input is logical 1 and the output is logical 0. The third part of the circuit is the receiving part which is an inverter-based latch structure. Its state is controlled by the transmitting inverters through the AC and DC path of the channel. While this is not the only level shifter used in stacked circuits, most of the implementations follow the same approach that is shown here [68][69][70].
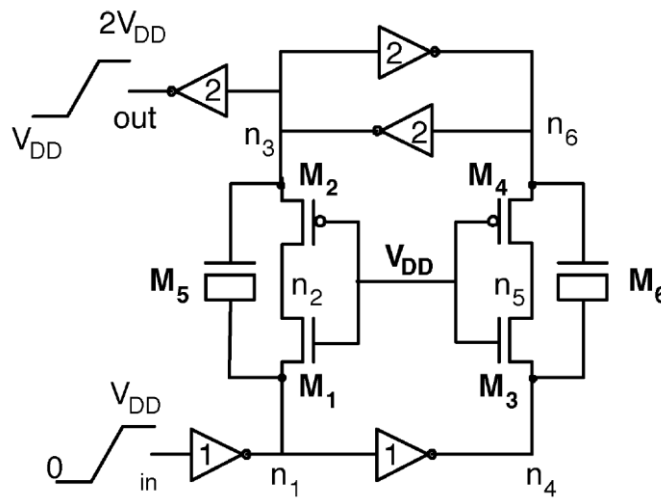
Figure 2-8: Level shifter employed with stacked circuits [67]

To make stacked circuits applicable for low-power digital systems, several problems have to be solved for the level shifters reported in the literature. One of these is the capacitor that is very often employed in these cells. To achieve the highest capacitance density, MOS capacitors are employed. They typically take large area to achieve the desired performance. The overhead in area can reach even 40X ratio compared to the other transistors in the cell [67]. Also, the use of

capacitors is not efficient considering supply variation. The idea is to have zero voltage change over them, but if the power rails have high current spikes, the IR drop can have an influence on the transition and the level shifter cell might even fail to work. Furthermore, with the decreasing process nodes, large gate area means high leakage due to electron tunneling through the oxide. Multi-threshold voltage devices [73] should be avoided as the design becomes very difficult to transfer between process nodes or technologies. Apart from the mentioned problems, the level shifter cell should fit within a digital row, and should be possible to flip it along either the power or ground rails to maximize the density of these cells.

2.5 Research Contributions of this Case Study Chip

This section describes the current project's scope within the state-of-the-art. The main goal of this research is to apply the known principle of voltage stacking to a realistic system-on-chip. To the best of the author's knowledge, this is the first silicon implementation of a stacked microcontroller system together with an efficient switched-capacitor voltage regulator.

We consider a system where two power domains with the nominal CMOS process supply voltage (V_{dd}), are stacked upon each other. This way, apart from the level shifters, the cells still operate at the same supply conditions like in a conventional system. It should be noted, however, that the supply noise, IR drop and ground bounce issues have improved, since the ground current of the top domain passes to the supply of the bottom domain directly, avoiding long interconnect wires that are external to the chip. The system also uses half the current compared to a conventional implementation, which reduces the IR drop by a half.

The design incorporates a whole Cortex-M0+ based microcontroller system with memory and peripherals, and an on-chip switched-capacitor voltage regulator. It is important to emphasize that this project targeted a test chip that represents a system widely used in industry. This demonstrator proves the concept of stacked circuits in a realistic application that makes the improvements directly applicable in current microcontroller ASICs. The application to a realistic system makes the current work unique because prior art usually dealt with less realistic system designs [67][71]. By making voltage stacking universal in this way, it can be applied to most of

the present and future systems, let them be battery-powered sensor chips or high-complexity microprocessors.

The designed integrated circuit is also the first one to employ a switched capacitor voltage regulator together with a stacked microcontroller system. This allows the power delivery efficiency to remain high even in the case when the current consumption of the domains does not match very well. As an example, for a 65% efficiency voltage regulator (see Section 2.4), under circumstances that one power domain consumes twice the current compared to the other one, the efficiency still can reach 85%. Though it is an external component in this project, an LDO can also be employed on-chip as voltage regulator in this system, which can act as a watchdog if the switched-capacitor DC-DC converter fails to provide the appropriate voltage. The application of an LDO also could limit the supply noise [74].

The third novelty of this thesis work is that the designed chip can be operated not just in the stacked mode, but also in the conventional, 'flat' mode. In the latter case, all the power rails are at V_{DD}, and all the ground rails are at ground voltage. This gives the freedom to choose between low-power stacked mode and high-performance flat mode. It also enables integrating the chip in systems with a $2V_{DD}$ supply or in systems where only V_{DD} supply is available. Stacking and de-stacking the circuit can have couple of benefits, e.g. when switching from state retention operation to active mode and vice versa. One example can be a smartcard which, when connected to a terminal, power requirements are not so important, but once removed from the contact, the system can go in stacked mode where leakage and dynamic currents are recycled, enabling low power operation.

3 Proposed System-on-Chip

Since the main goal of this research project is to implement the concept of voltage stacking into a realistic microcontroller system realized with a scalable process, it is not enough to only look at the implementation details of the specific 40nm low power CMOS process used here. Instead, a generic, transferable system level design should precede the actual silicon realization. This way, the system level design and the actual implementation have been separated. The former can be used to keep the ideas proposed in this thesis transferable into different technologies and requirements, while the latter is concerned only about the current test chip implementation details. In this chapter, the system level design is described.

The approach followed here consists of analyzing the state-of-the-art stacked circuit implementations, and enhance them with novel architectural choices to create a novel design flow for voltage stacking. A study has been made for the basic building blocks needed for such a system in Chapter 2. Based on the findings, the design space exploration can be started.

3.1 Digital System Level Design

The system level design starts at the concept depicted in **Figure 3-1**. The lower part is the conventional digital power domain ('bottom' power domain) that operates between 0V and V_{dd} voltages. It contains roughly one half of the microcontroller system in terms of power. The other half of the system is also digital but it is organized into the *'top'* domain which operates on top of the *'bottom'* domain in terms of voltage, e.g. between V_{dd} and $2V_{dd}$ gournd and supply voltages. Since the ASIC must be functional in the case when only a $2V_{dd}$ supply is provided, it is necessary to generate the V_{dd} supply. This means sourcing current for the bottom power domain when there is a lack of charge on the intermediate node, and sinking current from the top domain when there is an excess charge. This is why a third building block, the voltage regulator has been included. This power delivery block must be capable of two-way charge transfer or in other words, both sourcing and sinking current. In a similar way that

power interfacing is needed for the various supply and ground voltages, the signals also need a path to travel between the power domains. This functionality is covered by a signal interface block that includes level shifters and direct signal connections.
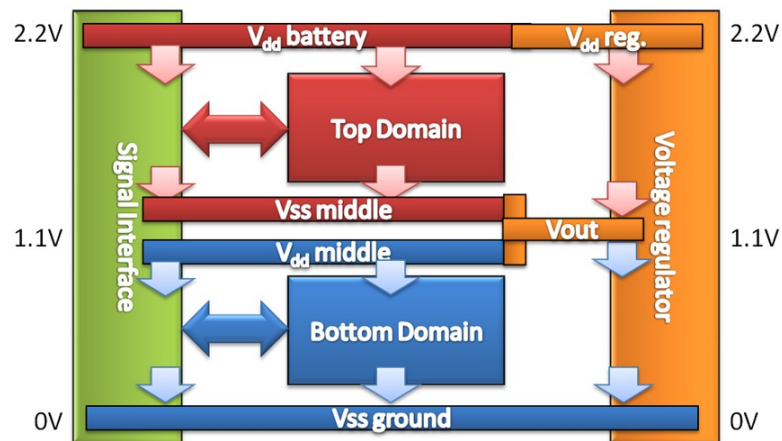
Figure 3-1: Stacked System Block Diagram

The novelty of this research project from the system design perspective is twofold. One is that the top and the bottom domain here represent a real system used in various applications, unlike in literature [67][71]. The complexity coming with such a complicated system takes a large part in the proposed novelty of this research project. Another innovation is reconfigurability. That is, a stacked circuit should be possible to be re-configured to work in the conventional, single-supply, *flat* mode. While there is only one work [57] that addresses this problem, the solution there was to apply a strong ARM latch as level shifter which employs thick-oxide, non-scalable devices. The current research aimed to create a scalable solution that consumes lower power.

3.1.1 Reconfigurable Power Domains

The benefits of stacked voltage domains have been demonstrated in Chapter 2. The intention to make voltage stacking available for practical applications has also been emphasized. However, there might be considerations that would favor the conventional voltage domain organization instead of stacking, in some circumstances. When a battery-powered system is being charged, performance is more important than power consumption, thus the overhead caused by the level shifters in the stacked system would cause a bottleneck for that operation

mode. Also taking the high effort to design a microcontroller ASIC in the stacked way would make it impossible to integrate that chip into a system where only a single V_{dd} supply is present. That is why the end user should be given the flexibility to switch between the *stacked* and the *flat* mode depending on the requirements, or use only one of them exclusively. This way there will be no necessity to design two separate ASICs, saving considerable design effort costs. In order to ensure these requirements, the current test chip was designed so that it can both be operated in the stacked and the flat mode. The state diagram of the possible transitions is given in **Figure 3-2**.
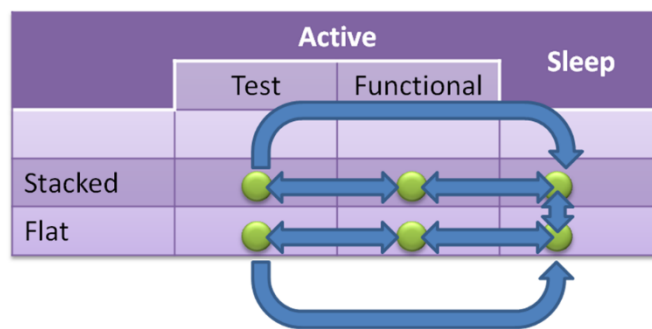
Figure 3-2: Stacking/de-stacking state machine

It is not possible to switch to stacked mode from flat mode and vice versa while the core is active. The signals traversing through level shifters would be subject to voltage spikes that reduce the noise margin and could cause malfunction of the system. To change the stacking configuration, the system first has to be brought into a sleep mode, which in this context means the complete stop of the system clock. During sleep mode, the memory and the latches will retain their state while the voltages are ramped up/down.

The concept of system level stacking has been illustrated in **Figure 3-1**. It can be seen that in stacked mode, the bottom domain operates between ground voltage and V_{dd}. The top domain containing memory controllers and the memories themselves is between V_{dd} and $2V_{dd}$ voltages. Thus, the ground node of the top domain is connected to the power node of the bottom domain. This mode is the core of the innovation proposed in this work. The benefit is that most of the current can be supplied at a single $2V_{dd}$ supply voltage to the chip, reducing the power delivery overhead. In the stacked mode, the level shifters are enabled to interface between the

different power domains. The on-chip voltage regulator provides the necessary current for the ground node of the top domain and supply node of the bottom domain at V_{dd} voltage, which is needed in case there is a mismatch in the current consumption of the power domains. Apart from this difference current, most of the current is sourced directly for the chip.

The physical state of the system during the *flat* mode is depicted in **Figure 3-3**. The top power domain is disconnected from the 2.2V and 1.1V supply, and connected instead to the bottom power rails. The flat mode is the conventional operation mode of the chip, with both power domains sharing the same power and ground rails, just like in most of the microcontrollers. This mode serves as a reference to compare the benefits of the stacked mode to conventional power delivery. The chip needs a single V_{dd} supply voltage, which is provided either externally or by the voltage regulator. The level shifters are bypassed in this mode, and the power domain edges are connected through buffers. All the power traverses through the on-chip voltage regulator, yielding lower efficiency. The circuit, however, can operate faster since the bypass buffers included in the signal interface block have lower delay than the level shifters.
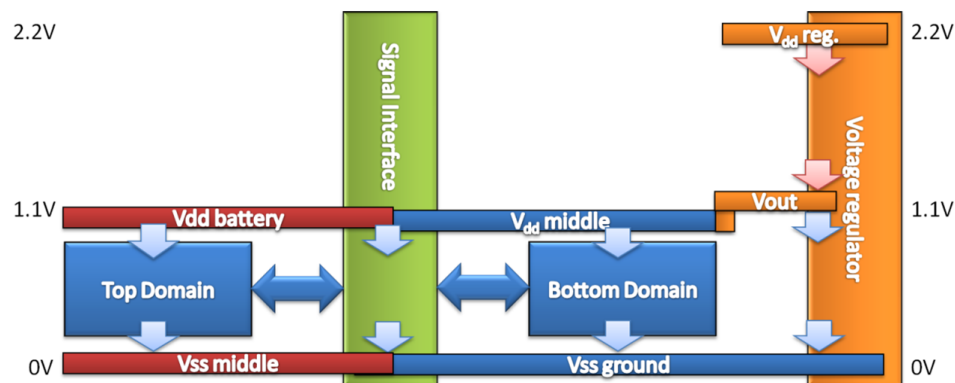
Figure 3-3: Reconfiguration into conventional (flat) mode

The most important aspect to ensure during the change of operation mode is the proper signal interfacing. On one hand, during stacked mode the signals must be level-shifted in such a way that they are shifted by a whole V_{dd}. This is a difficult requirement to meet as we will see it in Chapter 4. On the other hand, in flat mode we want to ensure that the signals have a bypass mechanism to avoid the slow level shifters and it are directly connected to the other power

domain. Furthermore, we want to separate the mentioned two functionalities by multiplexing between the two alternative signal paths.
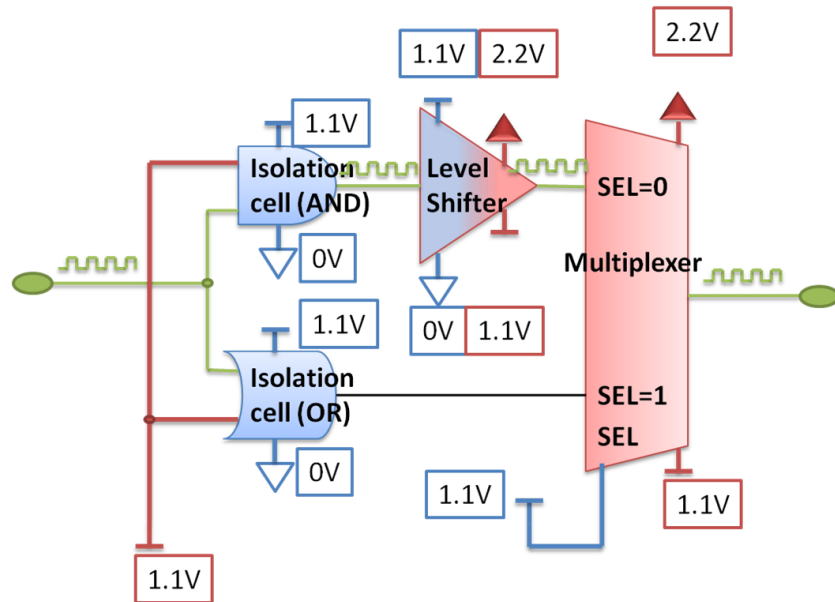
Figure 3-4: Level shifter bypass scheme in stacked mode
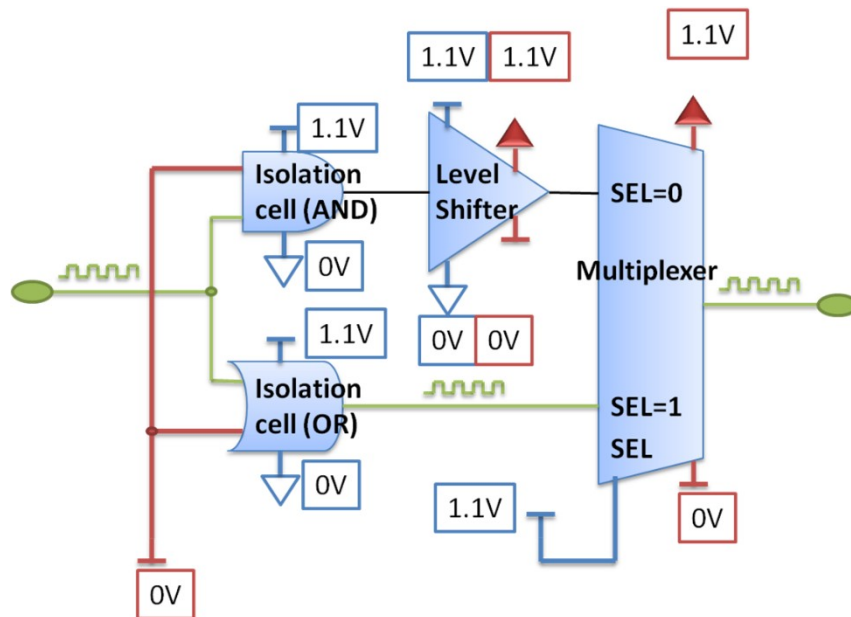
Figure 3-5: Level shifter bypass scheme in flat mode

More complex the overhead from this signal interface block, the less profitable voltage stacking becomes as a design choice. This is why the design is focused on simplicity. No external control signals have been used. Instead, the information encoded in the power and ground rails, has been used, in the way it is depicted in **Figure 3-4** and **Figure 3-5**. The figures show a bottom-to-top signal interface block in different operation modes. The circuit consists of a level shifter which translates the signal between the voltage levels of bottom and top domain in stacked mode, and a direct wired connection i.e. a bypass path. A multiplexer and two isolation cells, an AND- and an OR gate activate one of these paths depending on the operation mode – in stacked mode, the level shifter input and output is selected, while in flat mode, the bypass path. The selection is done without using external control signals. To control the isolation cells and the multiplexer, the voltage information of the power domain on the opposite side is used, i.e. the power rail of the bottom power domain for the multiplexer and the ground rail of the top power domain for the isolation cells. For example, if the circuit is in stacked mode, the multiplexer receives a 1.1V bottom supply voltage on its select input. This equals to a ground voltage for the top power domain, so the level shifter data input (since SEL=0) will be selected. On the other hand, when the circuit is in flat mode, the select signal is still at 1.1V, because the supply voltage of the bottom domain is not changing. However, a 1.1V signal now equals a logical one value since the top domain has been de-stacked and is between 0V and 1.1V voltages. Thus, the bypass path is selected. The isolation cells work in a similar way. In stacked mode, their control signal comes from the ground rail of the top power domain which is 1.1V, and equals a logical '1' in the bottom power domain (0-1.1V). This way, the OR isolation cell will have its output forced into 1.1V, while the AND isolation cell is sensitive to the input signal level and buffers the signal for the level shifter input. The level shifter receives the input signal and converts it to the desired voltage levels. In flat mode, on the other hand, the ground voltage of the top power domain that controls the two isolation cells is 0V, which is a logical '0' in the bottom power domain (0-1.1V in both stacked and flat mode). A logical '0' will force the output of the AND isolation cell to 0V while it enables the OR isolation cell to bypass the signal to the multiplexer. To summarize, the signal interfacing block makes sure that the signal reaches from one power domain to the other one over the two operatinig modes.

3.1.2   Microcontroller Architecture

As mentioned in the beginning of the current Section 3.1, one of the key proposals in this work
is the complexity of the stacked domains. For that purpose, a microcontroller system has been
constructed to reflect and represent a realistic application. The core of the system is a Cortex-
M0+ core licensed from ARM. The Cortex-M0+ is the most energy efficient ARM processor
available to the date of this work [80]. It uses the ARM Thumb instruction set and employs a 2-
stage core pipeline.
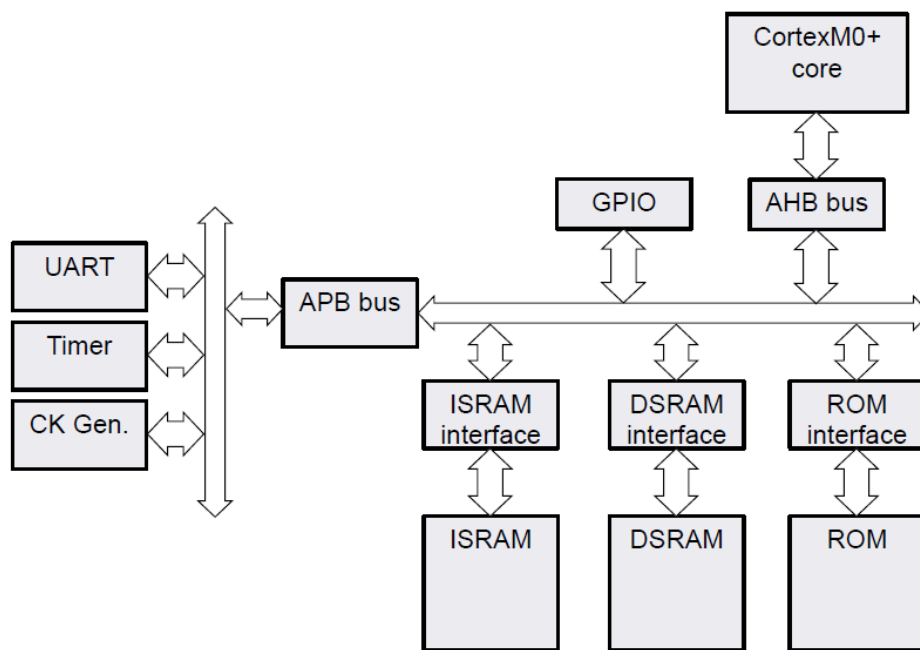
Figure 3-6: Block diagram of the proposed system

The IPs communicate with each other through an interconnect system organized based on the
Advanced Microcontroller Bus Architecture (AMBA) [81]. The core directly connects to a high
performance version of AMBA called Advanced High-performance Bus (AHB), which provides
fast connection between the processor, memories, separate peripheral bus (Advanced
Peripheral Bus, APB) and GPIO bank. There is only one master in the system, so the AHB-Lite
variant is used.

There are three memories present in the system. A ROM stores the bootloader program that
initializes the system upon each reset. Next to the ROM, there are two SRAMs. One is for

instruction storage (ISRAM) that can be programmed through the serial wire interface of the Cortex core, while the other one is for data storage (DSRAM).

The peripherals are connected to the APB which in turn is connected to the AHB. The APB is a low bandwidth version of the AMBA system, mostly designed for simple tasks like configuring the peripherals. It has low complexity to reduce power consumption. There are several peripherals connecting to the APB. One of the IPs used is the clock generation unit which is simplified to only distribute the internal clock to the main clock, the serial wire clock and the test clock. There is also a UART transceiver and a timer circuit that can generate interrupts for a wide range of periods.

The IPs that are selected here reflect the architecture of a typical microcontroller system from the NXP LPC family. This way the final system will be an adequate demonstration vehicle to prove the concept of stacked circuits in realistic applications.

3.2 Stacked Voltage Domain Partitioning

Voltage stacking requires the implementation of voltage islands which are stacked on top of each other. Furthermore, to have high power delivery efficiency, these voltage stacks should have matching power consumption, as pointed out in **Equation 3-5**. Another constraint is to have a small number of signals traveling between power domains since the level shifters that are needed impose power, timing and area overhead.

It has been observed that though prior art mentions mostly two power domains stacked on the top of each other [71], stacking three power domains has also been proposed [67]. The current work was limited to accommodate only two stacked power domains for design partitioning reasons. Having to create three or more power domains that have similar power consumption is considerably more challenging than separating the design into two parts. The power matching and the number of level shifters have to be carefully balanced and there is a sweet spot with respect to the expected power consumption of the circuit. This sweet spot heavily depends on the implementation, and it is only possible to approximately match the power consumption at the system design stage.

After deciding for the two-way partitioning, the next step is to analyze the possibilities within the system for dividing it into two parts. We consider the architecture shown in **Figure 3-6**. As an ad hoc method, the best way of partitioning would be to select a block that has small enough number of connections to the rest of the system, while having considerable power consumption share of the total. To estimate the power breakdown of the system, synthesis has been performed in a 40nm low-power process. The power estimation has been performed through activity annotation of the netlist nodes based on simulation data.
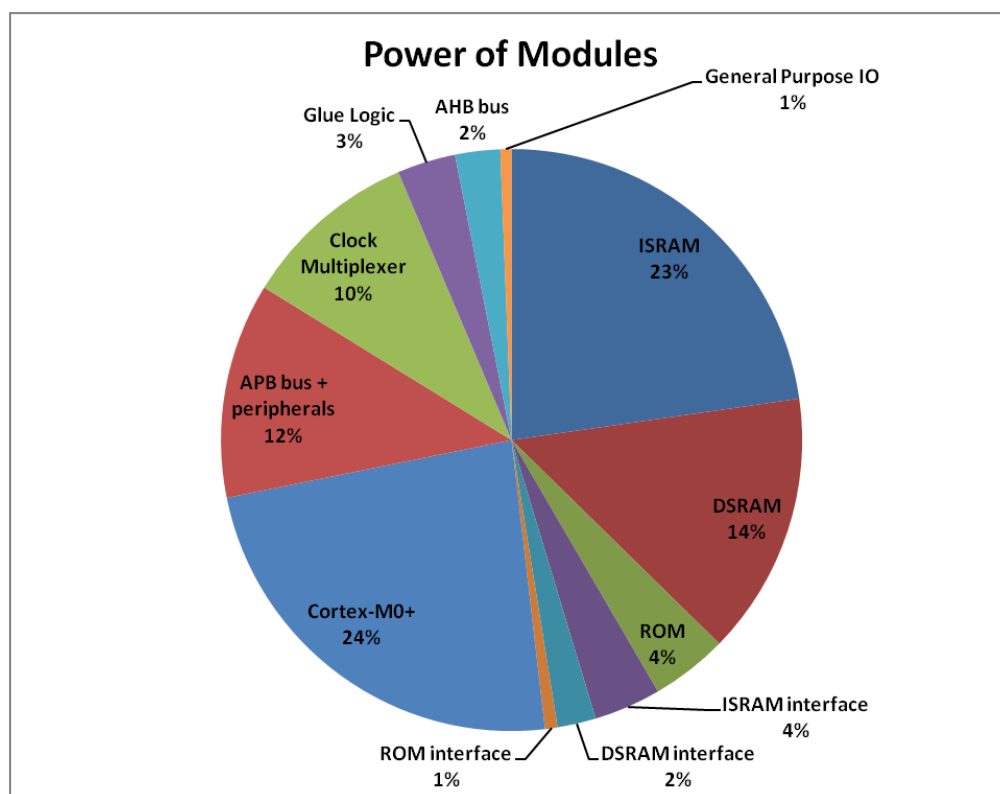
Figure 3-7: Power consumption distribution of the system at 100MHz with ISRAM active

In **Figure 3-7** it is possible to see that the memory blocks together with their controllers consume roughly half of the total power of the system. This ratio is, however, subject to variations based on the status registers and the program that is executed. In principle almost all the peripherals could be turned off for power saving purposes, while the processor might run either a *while (1)* loop or a high intensity benchmark program. In the current power analysis a matrix multiplication program has been selected as the reference for partitioning, which was

executed from the ISRAM. It can be concluded that dividing the system into a domain with the memories and their controller interface is going to provide a good power matching given the program that is executed is the same and the power figures do not change significantly during the placement and routing stages.
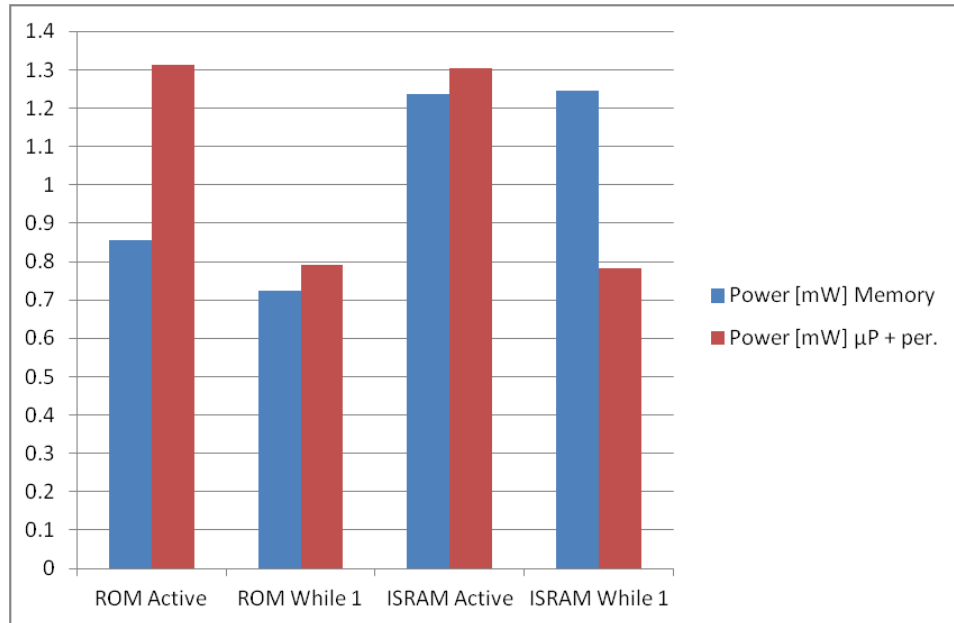
Figure 3-8: Power estimation of the two stacked domains

In this research project it has been decided to partition the design into two blocks along the memory AHB interface. Thus in the up domain are the ROM, the instruction SRAM and the data SRAM, each with their AHB bus controller interfaces. During the power estimation, four testbenches have been used – for the ROM and the instruction SRAM, one high and one low activity program has been executed. The power estimation results can be seen in **Figure 3-8**. It can be seen that the power consumption of the memory domain on the top of the stack largely depends on whether the ISRAM or the ROM is active, while the microprocessor power consumption depends more on how computation-intensive the program is. For the low-activity testbench a simple *while (1)* loop has been implemented, for the high activity a matrix multiplication algorithm has been analyzed.
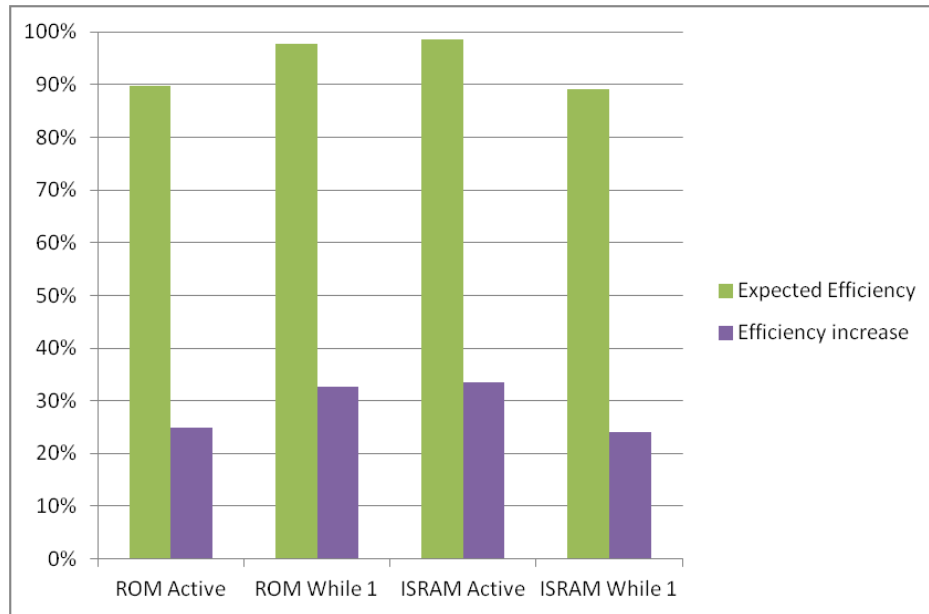
Figure 3-9: Power estimation based on the power matching and assuming 65% DC-DC conversion efficiency

Knowing the power matching that can be achieved, the estimated efficiency can be calculated using **Equation 3-5**. The DC-DC converter efficiency was assumed to be 65%. The results of the efficiency estimation are shown in **Figure 3-9**. It can be seen that though better matching implies better power efficiency increase, in general it is not required to have very good matching for considerable (25% <) increase in the total efficiency. The second remark is that the numbers in this calculation are optimistic, since the power mismatch can be spread within time, which decreases the efficiency. Also, the power consumption overhead of level shifters is not counted in this model. Furthermore, the efficiency of the DC-DC converter is not constant over the range of the loads presented here. This latter problem is addressed in Section 3.3.

3.3 Power Delivery

In Section 2.3 and 2.4 a couple of possible voltage regulators and stacked power delivery methods have been presented. The task from the system level perspective is to analyze these solutions and to propose an adequate architecture. In order to proceed, some basic considerations about the specifications of the stacked voltage domain system will be given. To

keep the first test chip of this research project at the proof-of-concept level, the simpler options were favored at the design choices. The following constraints have been made:

- The standard cells will be operated at nominal voltage. That is 1.1V V_{dd} for a 40nm process, and the substrate NMOS (PMOS) devices are tied to the ground (power) rail. This is a typical situation for microcontroller systems.

- The voltages for the bottom domain are 0V for ground and 1.1V for power, while in the top domain are 1.1V for ground and 2.2V for power.

- The power rails will be split for each domain, and connected externally with each other as needed.

- The chip in stacked operation mode takes a 2.2V external power source. The voltage regulator may provide the 1.1V core supplies, thus it is a 2:1 ratio converter. The IO pads may be powered externally since their performance is not relevant to the core.

- For the reason that a linear regulator is limited to 50% power efficiency due to the 2:1 conversion step, a switched mode power supply is proposed. More precisely, due to the non-conventional methods needed for the integration of an inductor, the switched capacitor regulators are favored. External linear regulators are implemented off-chip and can provide voltages between 0.6V and 1.2V during the measurement.

3.3.1   System Level Power Delivery

Interleaving is a commonly used technique for voltage regulators [55]. The key benefit is the ripple reduction of the output node. The ripple reduction improves the supply quality of the regulated power rail and increases the efficiency of switched-capacitor converters. Interleaving is commonly done by generating the regulator clock signals by a ring oscillator and distributing the nodes to the converters in a way that they are uniformly delayed with respect to each other within one clock period. In this test chip an external clock is provided, and the regulators have been chained one after another so that the delay between the stages is a couple of inverters plus interconnect delay.
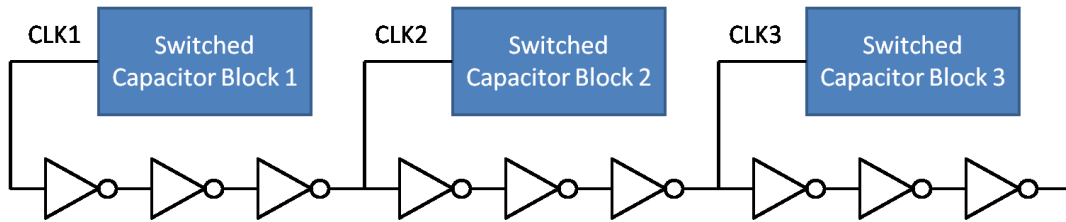
Figure 3-10: Simple interleaving scheme of regulator blocks

To increase the flexibility of the voltage regulation, it has been decided to include two separate interleaved clock chains. This way there are four different configurations for the number of active regulator blocks. Chain 1 contains 6 regulator blocks while chain 2 contains 10 stages, and either 6, 14 or 20 blocks are active, and also there is the possibility to disable all the regulators by keeping both the clock inputs constant. The benefit of this scheme is that the peak efficiency curve of the DC-DC converter can be shifted as deemed necessary without modifying the clock frequency thus introducing extra switching losses. In the following the choice of switched capacitor topology follows.

3.3.2 Switched-Capacitor Topology

In the design of switched-capacitor DC-DC converters, the operation has to be carefully analyzed, and advanced techniques have to be used for high efficiency and power density [52]. In modeling switched-capacitor regulators, the usual approach is to enumerate the losses. The first loss component is due to an effective output impedance, which will determine the *conduction loss $I_{load}^2 R_{out}$.*
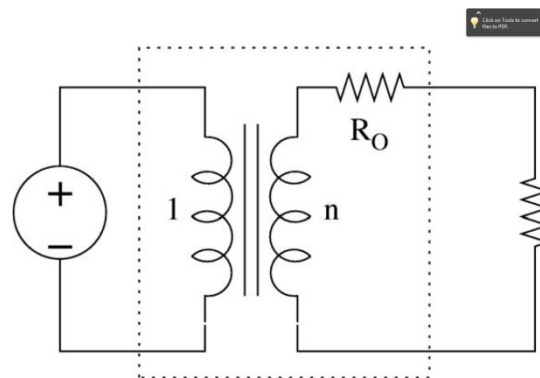
Figure 3-11: Model of a DC-DC converter [52]

The output impedance is further divided into two limit cases for easy calculation. One asymptotic limit is the *Slow Switching Limit.* It accounts for the loss that occurs when a capacitor is suddenly connected to a node with a different voltage causing a discharge event that dissipates $q\Delta V/2$ energy in each clock cycle. This would happen even for ideal switches. The model assumes that charge transfer is immediate (Dirac-delta current function in calculations). The slow-switching output impedance is inversely proportional to the switching frequency and also depends on the topology of the converter.

$$\eta_{SC} = \frac{V_{out}I_{load}}{V_{out}I_{load} + P_{sw} + R_{out}I_{load}^2}$$

Equation 3-1: Efficiency approximation of a switched-mode DC-DC converter
V_{out} - output voltage, I_{load} - output current, R_{out} – output impedance, P_{sw} - switching loss

The other limit case of the output impedance, the *Fast Switching Limit* accounts for the finite resistance of the switches and interconnect. Here, the currents are modeled as constant rather than as a Dirac-delta (constant current function in calculations). The fast-switching output impedance is not related to the switching frequency but it is dependent instead on the topology of the converter, namely, the connection of switches and the net resistance of that topology. In reality, a mix of these two limits occurs in the form of an exponential function, rather than a Dirac-delta and a constant current behavior. For calculation purposes, however, the mentioned calculation method provides an easy way to optimize the output impedance of the switched-capacitor DC-DC converter. It is not required to have knowledge about the internal node voltages and currents, only the topology and the switch/capacitor values and the switching frequency are important. There is a second loss component, the *switching losses* which are present even when the regulator is not loaded ($I_{load} = 0$). This loss component comes from driving the parasitic capacitance of the components like large switches of the converter, and is proportional to the frequency. The generic formula for the switching losses always follow the scheme $1/2\, f_{sw}V_{DD}^2 C_{par}$. The third loss component is the static loss, e.g. leakage of the employed transistors. Considering all the loss components, the efficiency and the power density can be directly optimized.

In **Figure 3-12** an example is given for a voltage regulator efficiency curve using **Equation 3-1** and values $P_{sw} = 0.33mW$, $R_{out} = 33\Omega$, $V_{out} = 1V$. Qualitatively it can be stated that if these three key parameters do not change, there is an optimum load current where the regulator performs in the most power-efficient way.
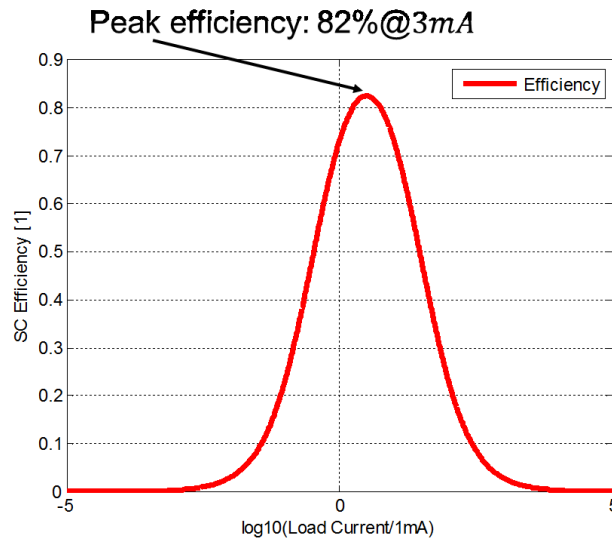
Figure 3-12: Theoretical Efficiency curve of a Voltage Regulator

A possible topology of a switched-capacitor converter that meets the voltage level requirements outlined before, that is, a 2:1 converter [55] is depicted in **Figure 3-13**. For each clock signal, the fly capacitors are either connected between the input and the output, or between the output and the ground. If there was a voltage ripple ΔV deviation from the nominal output value in one phase, the fly capacitor experiences $-\Delta V$ voltage change in the next phase, causing a discharge event that regulates the output back towards the nominal voltage.
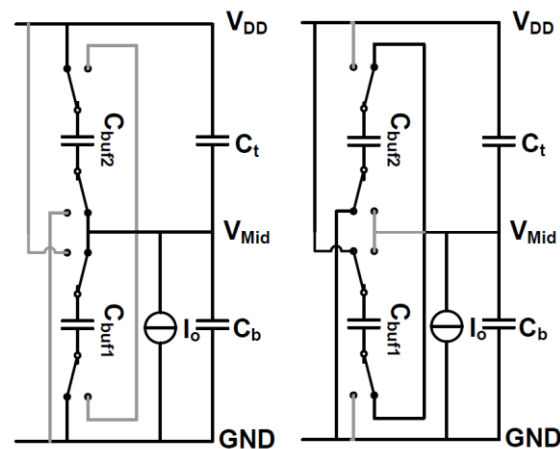
Figure 3-13: A commonly used 2:1 conversion ratio topology [56]

For the 2:1 topology, the following considerations must be made in a practical design. Since the switches are operating at different voltages, level shifter is needed to transfer the clock signal to the upper domain. Another consideration in the design is that the switches should not turn on at the same time since that causes extra loss in the form of short-circuit current. To avoid this situation, non-overlapping clock generators are also employed in the design. Taking these precautions, the final and most important component to be optimized is the fly capacitor itself. If a conventional bulk CMOS process is used [58], the highest capacitance density that can be achieved is with a MOS capacitor, instead with a fringe capacitor. The latter has low parasitic loss, however, a MOS capacitor suffers from considerable parasitic capacitance towards the substrate, which is typically at the ground voltage, 0V. This can reach even 5-10% parasitic capacitance if the intended MOS capacitance is taken 100%, depending on the CMOS process [58]. One possible workaround is to implement the capacitor within an n-well or a triple p-well. If the well is connected to the same potential as the drain and the source, the aforementioned parasitic capacitance will be shorted by an interconnect wire, and only the well-to-substrate capacitance will cause losses. This is typically only 2-3% of the intended capacitance, which is a significant improvement. This is another trick not just to minimize the parasitic capacitance, but to use it to boost the regulator. By connecting the drain-source-well terminals to the higher voltage, the charge dumped upon the bottom plate will be directed from the input to the output, yielding the efficiency of a linear regulator ($\eta < V_{out}/V_{in}$). These techniques help to

keep the efficiency relatively high. However, even after careful considerations, only efficiencies not exceeding 70-80% can be achieved with conventional bulk CMOS process [58].

A possible solution to the efficiency limitations is to use a silicon-on-insulator (SOI) technology with high quality factor deep-trench capacitors. These devices have very high capacitance density due to their 3D geometry, as well as low parasitic capacitance since they do not rely on a planar MOS structure. Switched-capacitor converters designed with SOI technology can achieve even up to 90% peak efficiency [59]. The drawback is that SOI is still not as widely used and it is a rather expensive technology. In the future, this is expected to change with the advancement of process nodes and the willingness of the conservative semiconductor industry to change.

3.3.3 Analysis of Power Savings through Voltage Stacking

So far it has been qualitatively stated that stacked circuits can spare battery power for an autonomous system. It is also important, however, to quantitatively evaluate the power savings that we can expect. The power efficiency of stacked circuits can be calculated with the generic formula $\eta_{total} = P_{out}/P_{in}$. Expressing this formula in more detail yields

$$\eta_{total} = \frac{P_{out}}{P_{in}} = \frac{P_{stack} + P_{reg}}{\dfrac{P_{stack}}{\eta_{stack}} + \dfrac{P_{reg}}{\eta_{reg}}}$$

Equation 3-2: Generic power efficiency formula

The output power can be separated into two parts. The component P_{stack} represents the power dissipated by the current that is flowing through all the stacked domains, the common mode power, while P_{reg} is the difference in power betwen the stacks that must be delivered by a voltage regulator. While the DC-DC converter employed as such ultimately will have a less-than-one η_{reg} efficiency value, the common power P_{stack} does not need power processing, and $\eta_{stack} = 1$.

$$\eta_{total} = \frac{P_{stack} + P_{reg}}{P_{stack} + \dfrac{P_{reg}}{\eta_{reg}}}$$

Equation 3-3: Power efficiency of stacked circuits

The power efficiency is a good figure of merit to approximate the energy efficiency if average power values $\langle P \rangle = \frac{1}{T}\int_T P(t)dt$ are used for a certain measurement time T, however, it is not entirely accurate. To get energy efficiency from power efficiency, one must be very careful. There are examples in literature where the efficiency of stacked circuits is measured real-time, which is the so-called "running" efficiency [57]. The following inequality holds:

$$\eta_{E,total} = \frac{\langle P_{out} \rangle}{\langle P_{in} \rangle} = \frac{\langle P_{stack} \rangle + \langle P_{reg} \rangle}{\langle P_{stack} \rangle + \langle \frac{P_{reg}(t)}{\eta_{reg}(t)} \rangle} < \frac{\langle P_{stack} \rangle + \langle P_{reg} \rangle}{\langle P_{stack} \rangle + \frac{\langle P_{reg} \rangle}{\langle \eta_{reg} \rangle}}$$

$$\frac{\frac{1}{T}\int_T P_{stack}(t)dt + \frac{1}{T}\int_T P_{reg}(t)dt}{\frac{1}{T}\int_T P_{stack}(t)dt + \frac{1}{T}\int_T \frac{P_{reg}(t)}{\eta_{reg}(t)}dt} < \frac{\frac{1}{T}\int_T P_{stack}(t)dt + \frac{1}{T}\int_T P_{reg}(t)dt}{\frac{1}{T}\int_T P_{stack}(t)dt + \frac{\frac{1}{T}\int_T P_{reg}(t)dt}{\frac{1}{T}\int_T \eta_{reg}(t)dt}}$$

Equation 3-4: Energy efficiency ("running" efficiency) of stacked circuits

The above formula shows that efficiency measurement of a complex load that has its instantaneous power changing rapidly will always yield worse results than the measurements commonly employed in literature for standalone voltage regulators involving a constant load, for the same amount of energy processed. In other words, a regulator that has its load current varying greatly over a certain time will have lower efficiency than the same regulator for the same time and same average load, but less variation. This is especially true if we consider that for a larger load variation, we might tend to differ from the peak efficiency point more than for a steady load. Thus, expecting the efficiency boost from stacking to be as significant as **Equation 3-3** suggests is not feasible.

If we accept that our approximation is optimistic, we can consider an example. If, for simplicity, we assume that the voltages remain equal and constant over the power domains, we can use the notation $\Phi = P_{reg}/P_{stack} = I_{reg}/(2I_{stack})$, meaning that we are using the current mismatch to approximate the power mismatch. This way **Equation 3-3** becomes the following:

$$\eta_{total} = \frac{1 + \Phi}{1 + \dfrac{\Phi}{\eta_{reg}}}$$

Equation 3-5: Simplified power efficiency formula of stacked circuits

Let us consider a typical example here. We can assume that the on-chip regulator achieves 65% power efficiency, which is a realistic number [58]. The second assumption is that in the non-ideal case, one of the stacked domains consumes e.g. twice the current of the other one, which means $\Phi = |I_1 - I_2|/\min \ (I_1, I_2) = 0.5$. In this latter case evaluating **Equation 3-5** yields $\eta_{total} = 84.8\%$. It can be observed that with far from ideal conditions, the optimistic estimation predicts almost 20% efficiency boost. If we also consider the circuitry overhead for stacking like level shifters, and keep in mind that we are calculating an optimistic value, it is still safe to state that stacked circuits still offer a double digit efficiency boost in percentage for on-chip power delivery, in a realistic scenario.

For the more general case, the on-chip regulator can be assumed to take any efficiency value, and any load current. If the load current comes from the mismatch current of two stacked domains where their common current is fixed, the total stacked system efficiency can be calculated. In **Figure 3-14** the 2D plot shows the stacked system efficiency as a function of switched capacitor (SC) converter efficiency and the mismatch current. It is a graphical representation of **Equation 3-5**. The black line, on the other hand, is the projection of the example voltage regulator efficiency curve from **Figure 3-12**, which is the graphical representation of **Equation 3-1**. This graph shows what voltage stacking provides and how it is limited with the actual DC-DC converter that is used. The interesting part would be to make **Figure 3-14** based on actual measurements of the stacked system and directly demonstrate the power saving capabilities.
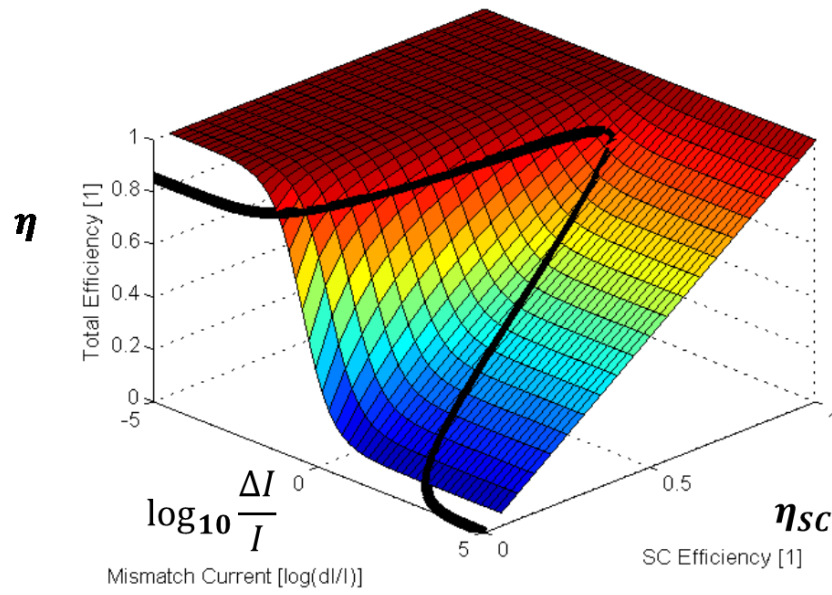
Figure 3-14: The Efficiency Boost from Stacking and the limitation for the Voltage Regulator
$$I = 1mA$$

4 Test Chip Implementation

This chapter describes the test chip realization starting from the system level design in Chapter 3 to the actual photo mask file (GDSII) required for the fabrication. Building on NXP's conventional SoC design flow, several enhancements are necessary to implement the proposed stacked system. The overhead is due to the additional steps needed in the design cycle of a stacked SoC. This way the IC designer has a clear view on this low power option and can decide if it is worth the additional investment. Due to this and to simplify the scope of the project, the flow additions have been kept as simple as possible with the used EDA tools and technology.

4.1 Level Shifter for Stacked Logic

In order to achieve the benefits that voltage stacking offers, the functionality of the system has to be ensured by meeting the signal integrity criteria for all signals. This is also true for the boundary of the power domains. Since in the stacked case the power domains have different voltage levels, the signals are not compatible with each other and have to be translated to the voltage levels of the receiving power domain. This translation is especially difficult for high power- and ground voltage differences, as it is the case in this research project.

	Bottom Power Domain	Top Power Domain
Ground Voltage	0V	1.1V
Supply Voltage	1.1V	2.2V
Ideal Signal Waveform	$V_{SS,gnd} + v(t) = v(t)$	$V_{SS,mid} + v(t) = 1.1V + v(t)$

Table 4-1: Level Shifter Specification

The level translation of a signal is a affine transform in the form $v(t) \mapsto a \cdot v(t) + b$. Since the voltage headroom between ground and supply does not change in the system, the gain value $a = 1$. The offset b takes the value of $V_{dd} = 1.1V$, as can be seen in **Table 4-1**. This ideal functionality of the required level shifter can be described by an ideal 1.1V voltage source.

4.1.1 Architecture

Since an ideal voltage source would be an active element, it needs to take energy from a power source. The only power source in the design are the power rails. For a real circuit that imitates the behavior of an ideal voltage source, this power could be delivered in the form of current. However, static bias currents cannot flow in the level shifters since low standby power has to be achieved. The only possibility is to deliver the current in the form of pulses upon signal transition, which is the case for digital level shifters in the prior art [67]. These pulses must be fast spikes and their current level cannot be controlled. Due to this the voltage levels on the receiver side will not be accurate and have to be regenerated by a latch.

A second way to approximate the behavior of a voltage source is using a capacitor. From the equation $dV = I \cdot dt/C$ follows that for sufficiently fast signal transitions and large C, the voltage varies little. Capacitors have implementation difficulties, however. If MOS capacitors are used, they have increasing leakage and often require a hot well to minimize parasitic capacitance, while MIM capacitors require blockage over a wider metal area which makes congestion in digital signal routing. Even if the capacitor was ideal, its efficiency is very dependent on the supply conditions since inequalities cause additional charging-discharging event and higher power consumption.

The third way of coupling the receiver to the transmitter is using devices with negative threshold voltage (always-on devices) for the up level shifter and over-V_{dd} threshold voltage devices for down level shifter. These are however not available in a modern CMOS process.
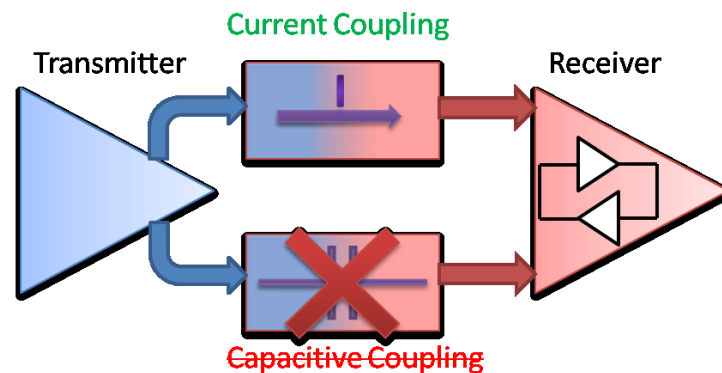
Figure 4-1: Level Shifting Scheme

For the considerations above it has been decided to use the current coupling method exclusively at the design of this level shifter. Current mode level shifters have the disadvantage that the receiver side regeneration imposes high power- and delay overhead, and that transistor drive strength has to be very different between the transmitter and the receiver circuits. To ensure the transistor drive strength, hot wells have to be used which increase the area. To overcome the power-, speed- and area requirements, reliability and yield would have to be sacrificed.

The key figure of the level shifter described here is that it is possible to integrate it into a digital system with standard cell rows. Thus no special devices like thick-oxide transistors or multiple threshold voltage transistors are used. The former requirement imposes a challenge that none of the devices should have a higher-than-V_{dd} DC voltage over any terminals for degradation reasons. Furthermore, the level shifters must have low leakage in both stacked and flat mode.

4.1.2   Schematic

The designed schematic of the up level shifter is depicted in **Figure 4-2**. The nodes $V_{ss,gnd}$ and $V_{dd,mid}$ are the ground and power rails of the bottom power domain with 0V and 1.1V, respectively. The nodes $V_{ss,mid}$ and $V_{dd,bat}$ are, on the other hand, the ground and power rails of the top power domain with 1.1V and 2.2V, respectively. Inverters *I1...I3* and devices *M1...M8* are all implemented with high threshold voltage transistors from a 40nm CMOS process. *M1...M6* devices are placed in hot wells while the rest of the devices have a constant substrate voltage. This is because the former devices must keep their drive strength throughout the operation. The latter devices in the bottom domain are implemented following the conventions in bulk CMOS process, while in the top domain the NMOS devices are placed in a triple p-well biased to $V_{ss,mid}$ voltage of 1.1V. In **Figure 4-2** only the up level shifter is shown, but the down level shifter is designed following the same convention. Swapping the NMOS and PMOS transistors of the up level shifter and mirroring the power and ground rails with respect to voltage yields the down level shifter schematic. Due to PMOS-NMOS asymmetry, however, the sizing had to be adjusted for the proper drive strength.
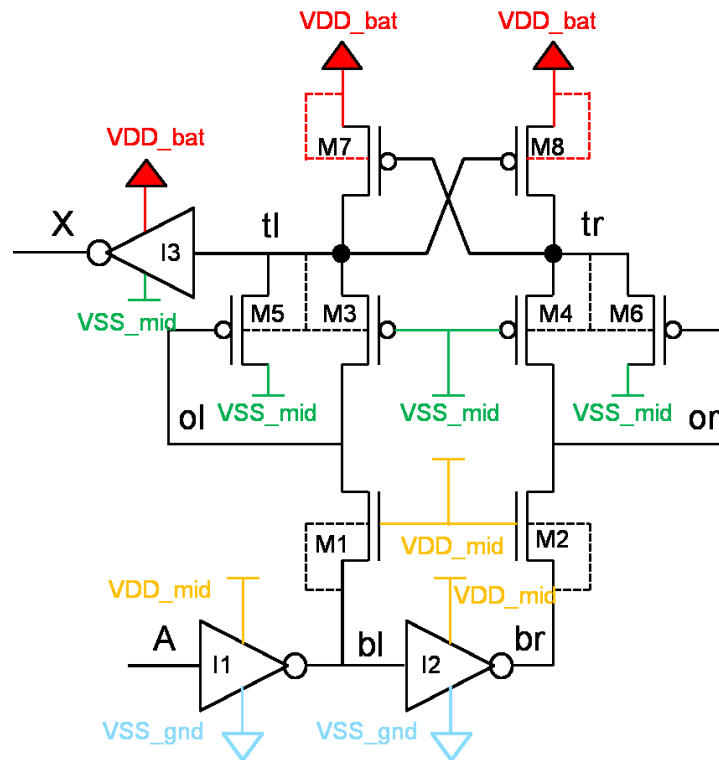
Figure 4-2: Up Level Shifter Schematic

The operation of the up level shifter is as follows. Let us suppose that the input and output of the cell, node **A** and node **X** are at logic **'0'** value. This means node **bl** takes logic **'1'** *value* at 1.1V and **br** takes logic **'0'** at *0V*. **M1** is off, while **M2** is on and pulls down the node called **or** to *0V*, the same voltage as node **br**. As a result, node **tr** is pulled down to 1.1V through **M6** and the opposite node of the PMOS latch, **tl** is at 2.2V. Finally, since **M3** is on, node **ol** is at 2.2V. When the input **A** is driven from **'0'** to **'1'**, both nodes **bl** and **br** change value. **M2** turns off, leaving node **or** floating and **M1** turns on, pulling down node **ol** to 0V. Since **ol** is at 0V, and **tl** is at 2.2V, **M5** opens and **tl** gets pulled down to 1.1V. Due to the latching event, node **tr** is pulled up to 2.2V and node **or** follows it through **M4.** The output **X** changes from **'0'** to **'1'** due to the transition of node **tl**.

We can see that when node **tl** is changing value in the described scenario, there is short circuit current flowing between **M6** and **M8**. This is the price that is paid for enabling fast transition of node **tl** through **M5.** Without **M5** and **M6** transistors, the level shifter cell would need a lot bigger **M1...M4** devices, and would burn considerable more power and would take longer time

to resolve its value. *M5* and *M6* are there thus to compensate for the lack of capacitors in this level shifter, which is normally present in the prior art cells. Also the sizing of the devices is important. While *M3...M8* are all minimum size devices, *I1* and *I2* employ 6X size NMOS devices, while *M1* and *M2* are both 4X sized. Throughout the operation, all the DC values are at maximum 1.1V. During transition, however, there can be higher voltage spikes than the nominal 1.1V value and this has to be addressed from the reliability point of view. The verification in this regard is described in Section 4.1.4.

4.1.3   Layout

The layout of the up level shifter can be seen in **Figure 4-3**. The cell requires only 5 standard cell rows if placed in pairs, with the second being mirrored to the $V_{ss,mid}$ rail. This was necessary because the deep n-well has a minimum dimension design rule. Also the n-wells that touch the deep n-well must have a minimum distance from the n-wells that are not connected to the deep n-well. This makes the layout less compact and scattered, but still compatible with the digital flow.
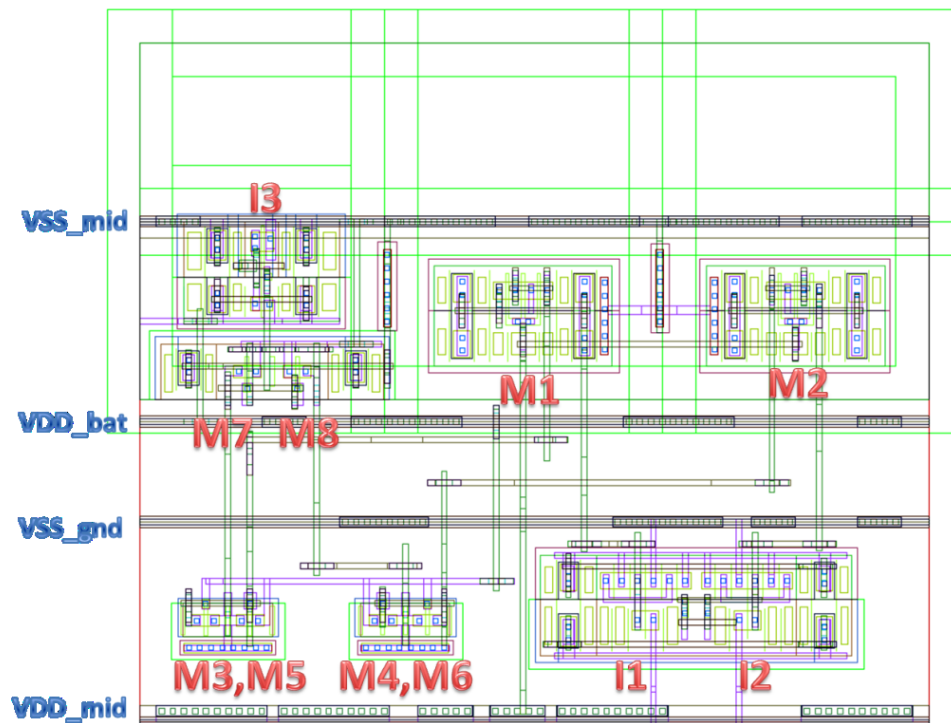
Figure 4-3: Layout of the Up Level Shifter

There is no gap required to be left between the level shifter cells during placement. *I1* and *I2* are inverters in the bottom power domain, so they are placed the conventional way, with NMOS being embedded in the bulk and the PMOS in an n-well, biased to 1.1V. *M1* and *M2* NMOS transistors both require separate p-well for substrate biasing.

They have their own p-well separated from the rest of the area over the deep n-well as shown in **Figure 4-3**. The triple p-wells are tied to the source of *M1* and *M2,* respectively. spared Area savings have been achieved by *M3, M5*, and *M4, M6* devices having their n-well shared since in the schematic their substrate connection is made to their source. The minimum distance required between these two n-wells and the deep n-well increased the number of rows from 3 to 5, and thus is a key limiting factor in terms of area. *M7* and *M8* form the PMOS latch, and they are not needed to have hot well. They are connected to the n-well ring around the deep n-well, which is biased to 2.2V. Finally, the top domain inverter *I3* is placed in the deep n-well, with a separate triple p-well reserved for the NMOS transistor.

The layout has considerable area overhead due to the deep n-well. In an SOI process where the bulk can be isolated, this overhead would significantly reduce. Also the hot wells could be implemented in an easier way. The current design, however, has demonstrated that level shifter cell for digital design flow integration can be created in bulk CMOS process as well. Though there is high parasitic capacitance added due to the long wires, the level shifter remains functional over the corners in the simulations, as shown in the next section.

4.1.4 Verification

The level shifter functionality and power-delay figures have been evaluated over various PVT corners and process variation analysis using the Monte Carlo method. In **Figure 4-4** the waveform for a 100MHz clock signal input and output at the typical corner can be seen. The propagation delay of the cell is significantly higher for high-to-low transition than for low-to-high transition. The reason can be found in the schematic topology, as described earlier. The short-circuit current between M6 and M8 in **Figure 4-2** increases the resolution time. The second property of the output signal is that its slew rate is lower than for the input signal. This is because even though the receiver side consists of lot weaker devices than the transmitter side to decrease the time spent in metastability, the resolution of the output still takes

relatively long. This limitation is technology dependent. The important figure here is the ratio in the drive strength of the devices. If that number can be made higher, the latch resolution time becomes smaller.
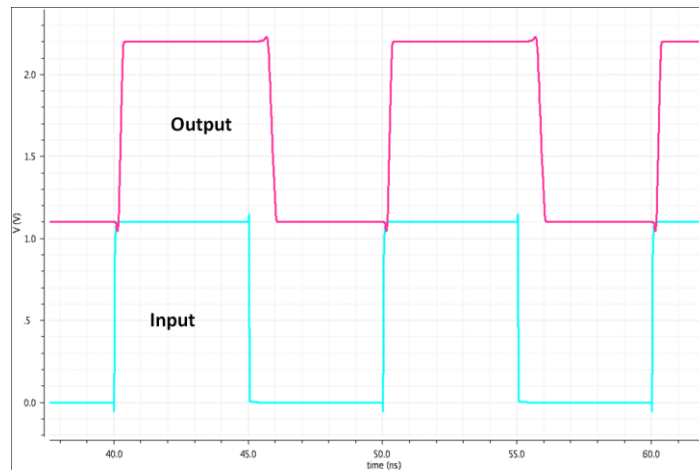
Figure 4-4: Level Shifter layout-annotated simulation for typical conditions

To ensure proper operation of the level shifter, the timing and the power consumption has been analyzed in different voltage conditions. The results for worst case corner (slow NMOS, fast PMOS) and worst case temperature (-40°C) can be seen in **Figure 4-5**. The cell fails only if either the input or the output supply voltage drops to 0.9V In the test chip the IR drop is minimized so it is not expected that the supply voltage would drop so significantly. The 5% rule for maximum supply deviation has been kept in mind for power integrity. Nevertheless, the 10% supply deviation corners, the 0.99V-1.21V corner is passed. The rise delay mostly remains within 0.6ns and the fall delay within 2ns. The power consumption for the simulated 100MHz clock signal was around 11μW, which corresponds to about 110fJ/cycle figure of merit. For the 140 up level shifters this would correspond to 1.6mW power consumption, but we have to keep in mind that not all the level shifters are active at all times, and the power report for the chip is done under typical conditions. Hence are the power values reported in **Table 4-3** significantly smaller.
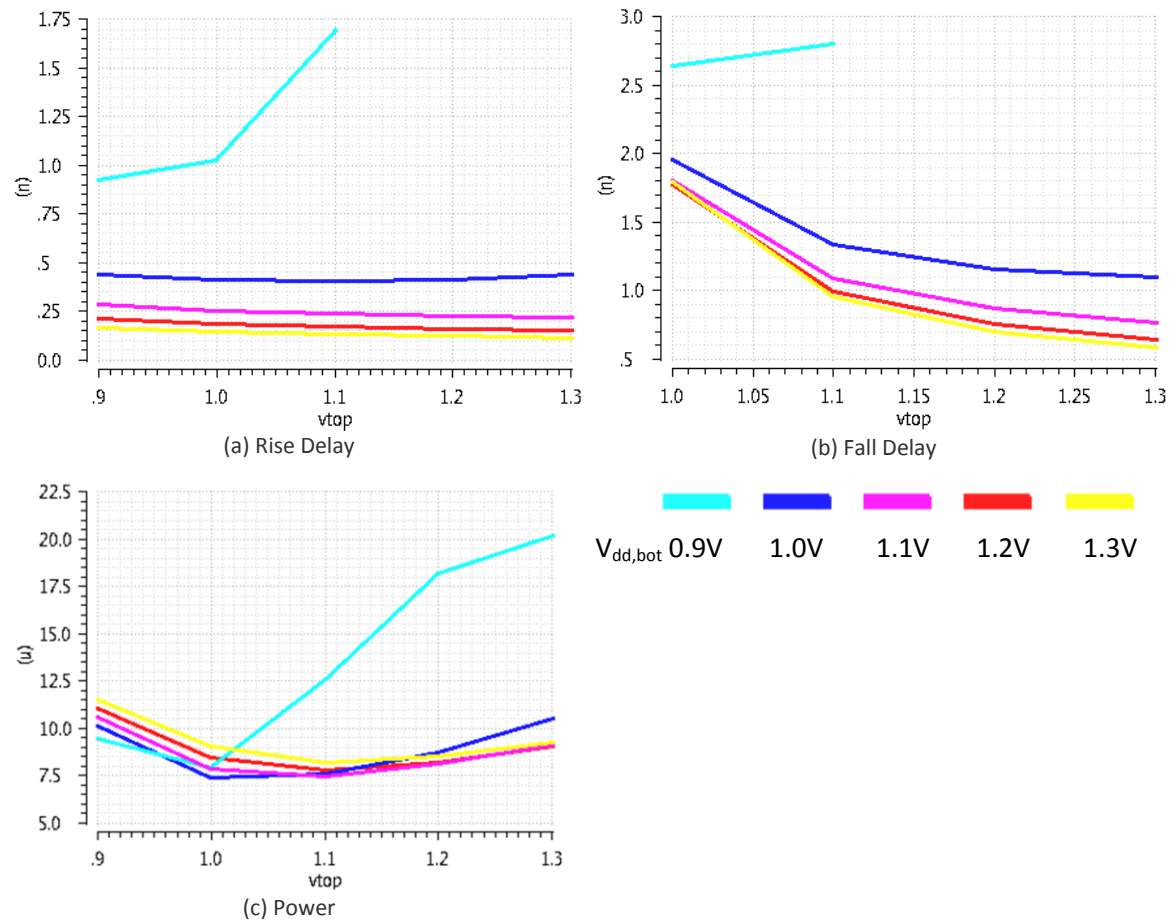
Figure 4-5: Timing and Power over different supply voltages

The results of the Monte Carlo analysis performed in the worst temperature- (-40°C) and supply (0.99V $V_{dd,bot}$ 1.21V $V_{dd,top}$) corner is in **Figure 4-6**. The samples were taken over 200 points. For the rise and the fall delay, 0.565ns and 1.903ns mean values were obtained. The mean power was below 11μW. All the plots have a long tail in the positive direction, which implies that certain process variation schemes can have high impact on the level shifter.

(a) Rise Delay

(b) Fall Delay
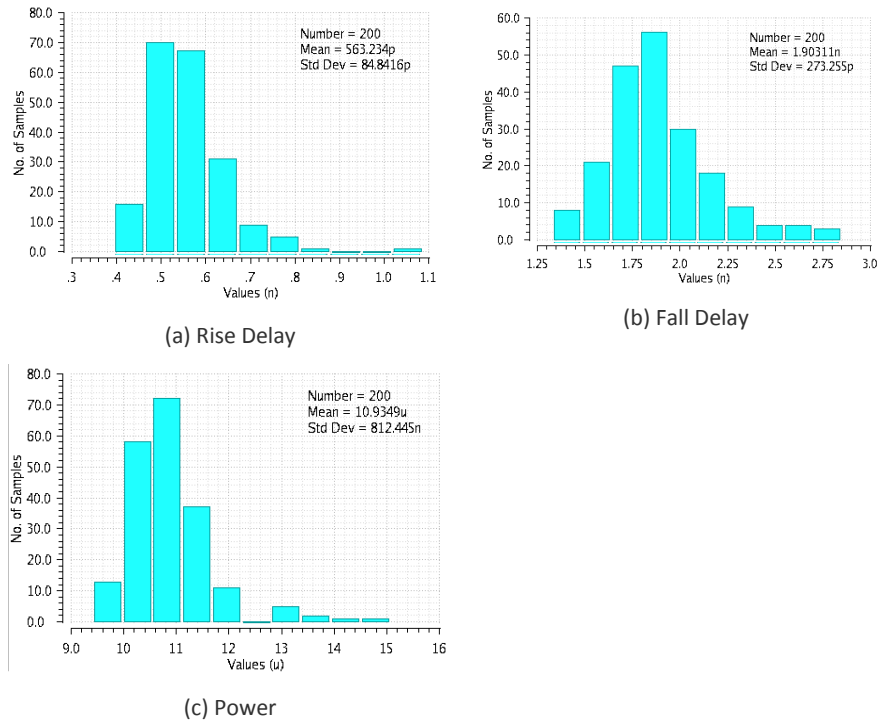
(c) Power

Figure 4-6: Monte Carlo Analysis of Power and Timing

Since the level shifter employs hot wells, the triple p-wells and n-wells are no longer biased to a constant voltage, but change their value between 0V-1.1V in the former case, and 1.1V-2.2V in the latter case. This creates a 2.2V voltage headroom possibility for latch-up events. To mitigate this problem, substrate connections as close to the devices as was possible were placed in both the hot triple p-well and the hot n-well, as well as the bulk, the DC-biased n-well, deep n-well and triple p-well.
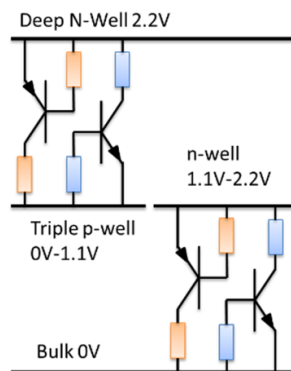
Figure 4-7: Latch-up consideration for the up- and down level shifters

Next to the latch-up, there is another reliability concern. The 40nm CMOS devices used in the process are designed for 1.1V nominal supply voltage. Though they can operate at higher voltages as well, their lifetime will decrease under extreme voltage conditions. In both the up- and the down level shifter cell the voltage without any switching event is within the nominal value, but temporarily there are spikes occurring that can reach even 50% excess over 1.1V at the gate-source terminals (**Figure 4-8**).

(a) The critical devices
(b) The voltage spikes during signal transition
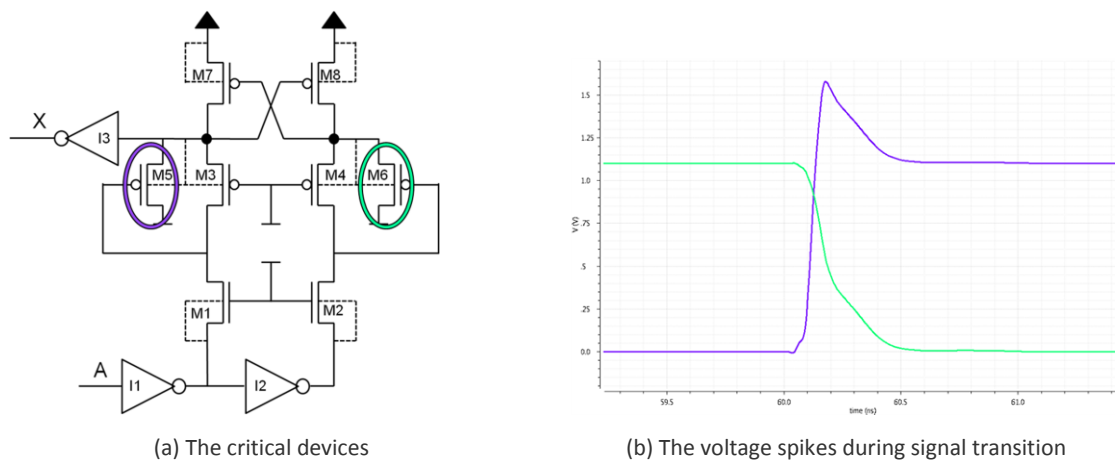
Figure 4-8: Voltage spikes over M5 device in the up level shifter

For such conditions it is important to verify that the devices affected will work for at least couple of years in the case of a test prototype, and several decades in case of a product. A device failure event of 10% saturation current change was chosen, which already makes differences in the circuit timing behavior. Calculations and simulation have been performed taking various degradation mechanisms into account. The first degradation mechanism considered was Negative Bias Temperature Instability (NBTI), which occurs in p-channel devices due to their negative gate-source and gate-substrate voltage. The calculations found over 20 years lifetime at 50 °C for continuous operation at a 100MHz clock signal, which is the most switching-intensive signal the level shifter is intended to process. The lifetime value reduces to 68 days for 150 °C. The second mechanism considered in calculations was Time-Dependent Dielectric Breakdown (TDDB). Due to the small device area, the lifetime was proven to be over 1000 years if only affected by TDDB. The final degradation mechanism considered was Hot Carrier Injection (HCI). Since it is a complex phenomenon dependent on couple of device voltages, simulation with reliability EDA tool RelXpert has been done. The results showed small

impact with respect to HCI, the lifetime was over 100 years. Thus, it was found that the dominant degradation mechanism is NBTI.

The final level shifter layout complies with the design rules of the 40nm CMOS process used in this test chip implementation. Similarly, the LVS checks also succeeded, and the parasitic extracted netlist has been used for all the simulations described above. The level shifter cell has been also fully characterized for various PVT corners and integrated into the digital flow described in Section 4.2.

4.2 Enhancement of Standard Digital Flow for Stacked Logic

The execution of the digital design flow starts with defining the pinning assignment. Since conventional IO pads are used, it has been decided that the core of the digital IO pads will be placed in the *bottom* domain. These IO pads should cover the basic functionality for the system like clock and reset signals, serial wire, general-purpose IO (GPIO), Design for Testability (DFT) options. The IO pad and power domain configuration can be seen in the **Figure 4-9**. The different colors correspond to different power domains. The pads with white filling are the already mentioned digital IO pads. The blue and orange pads are the power and ground connections for the *bottom* and the *top* domain, respectively. The violet pads provide the regulator with two clock signals and input-output power rails. Finally, the pads colored in green provide the external (2.2V) and core side (1.1V) supply for the IO pads. To interface between the two digital power domains, an array of level shifters is placed on the boundary of the two power domains.

The concept outlined in **Figure 4-9** has been followed throughout the implementation. There are two principal stages in the digital design flow – the synthesis stage and the layout stage. During the synthesis, the RTL level digital design is translated to the circuit level by covering the functionality with standard cells, yielding a circuit-level netlist. Then the netlist is placed and routed in the layout stage, together with the integration of macros and custom designed blocks, to define the final photo mask.
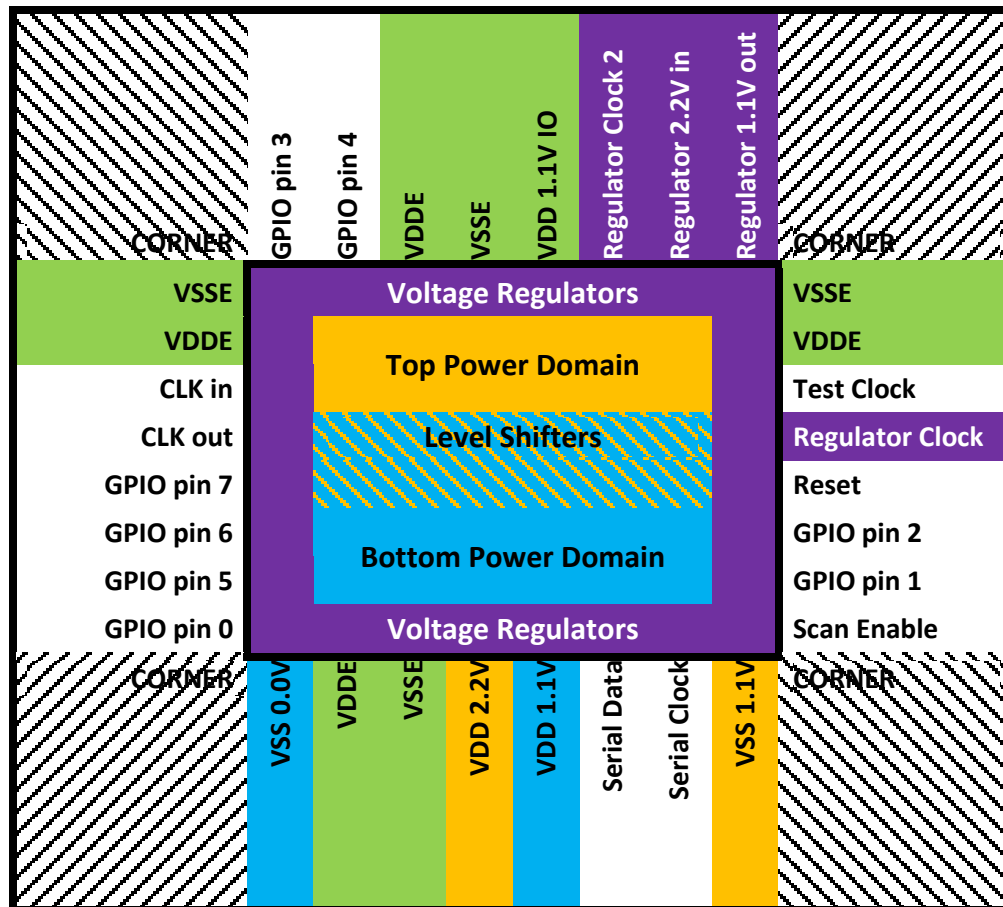
Figure 4-9: IO pad assignment

4.2.1   Synthesis

The synthesis was performed for a 40nm low power CMOS process with high-threshold voltage standard cell library. Three key corners were used to analyze the timing, a worst-case corner with slow devices, -10% supply voltages and 125°C temperature, a best-case corner with fast devices, +10% supply voltage and -40°C temperature, and the nominal corner where normal device speed and supply voltage was assumed at room temperature, 25°C. The IO pads were provided, while the ROM and the SRAM macros were generated using EDA tools. The full-custom blocks are the level shifter and voltage regulator. While the latter is isolated from the rest of the system and is not used during synthesis, the former needed to be characterized for the synthesis. The integration of various blocks within the front-end stage can be seen in **Figure 4-10**. The logic synthesis phase takes as input the timing libraries (.LIB), the layout abstract files (.LEF), the memory timing information and the IO pads. Similar timing information is required

for the custom-designed level shifters, and are shown as the orange colored boxes next to the timing library (LIB) and the abstract (LEF) steps.
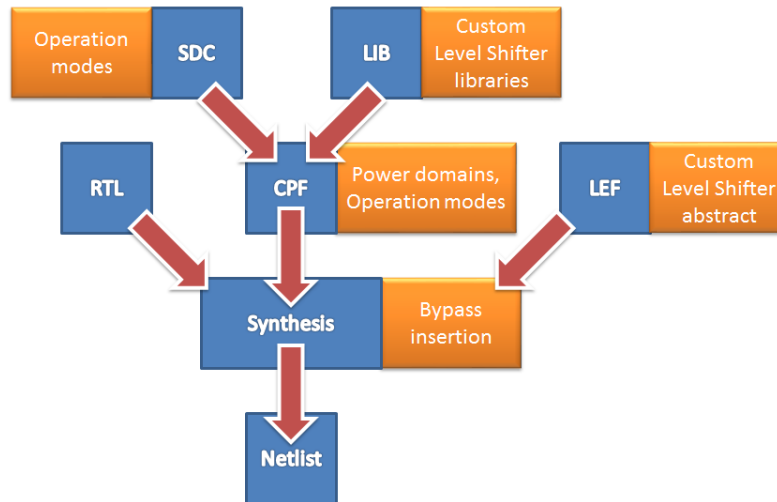
Figure 4-10: Front-end implementation steps. Enhancements are indicated with orange boxes.

Since it is difficult to describe signals shown in **Figure 3-4** and **Figure 3-5** that are from different power domains, constraints are written in the SDC file to force the control inputs of the isolation cells and the multiplexer into fixed logical values for either stacked or flat operation. This way, correct timing analysis can be ensured for both flat and stack modes. In the stacked mode the clock of the memories, for example, will propagate through the level shifter, while in flat mode, through the bypass path. The synthesis tool can then ensure proper operation in the operation modes, and optimize for both of them. The different clock frequencies for different operation modes can be seen in **Table 4-2**. In addition the active stacked and active flat modes, test modes have been implemented where the values of the flops in the design can be read out through scan paths. The setup uncertainty of the clock was set to be 0.3ns, 3% of the clock period, while the latency was set to 10ps.

	Active	Test
Stacked	80MHz	10MHz
Flat	100MHz	10MHz

Table 4-2: Clock frequencies used in the operating modes.

The power intent can be described with a CPF (Common Power Format) file. The file format can include the elements of low power techniques like power domains, level shifters, power nets in the early stage of the design. This way, it is possible to optimize the timing and power of the ASIC keeping the power intent in mind. In this research project the CPF was used to describe the power domains, the level shifter insertion rules and the operation modes. The visualization of the CPF power intent is depicted in the **Figure 4-11**. The bottom power domain is the default for all the blocks that are not specified to be in another power domain. The level shifters and the bypass logic in that sense are formally in the default power domain, while physically they are at the boundary between top and bottom power domain. Although the isolation cells and the multiplexers are shown as part of the power intent, they were inserted into the netlist by a custom script during the synthesis and are not described in the CPF file.
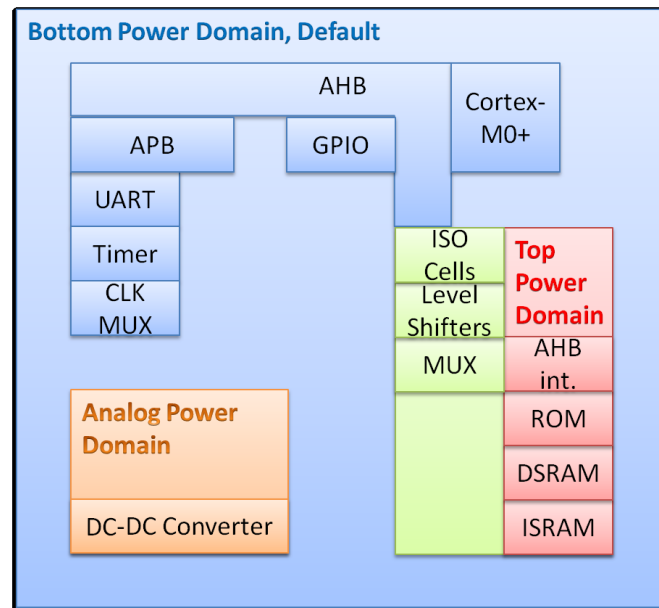
Figure 4-11: Power Intent of the System described in CPF

After the RTL level netlist has been read in, the power intent has been described and the layout abstract has been processed, the logic synthesis tool produces the netlist. It is important to minimize the dynamic power of the bottom power domain since the core and the peripheral power consumption is critical from the power matching perspective. Low power directives like dynamic and static power minimization and clock gating have been applied. Next to the

standard steps, there has been a custom step included in the synthesis – the bypass insertion for the level shifters.

Figure 4-12: Precedence of bypass insertion during synthesis

The sequence of the bypass insertion has to be preserved during synthesis, as depicted in **Figure 4-12**. The bypass cells, that is, the multiplexers, isolation cells and tie cells cannot be inserted before the definition of the power domains, otherwise the EDA tool will falsely assign them to either of the power domains. At the same time, the level shifter insertion converts the domain crossings into locked nets which cannot be modified, thus the bypass cells must be inserted and their connections to the domain crossing signal have to be made. However, the bypass cell signals cannot cross the power domains, otherwise in the next step, the level shifter insertion will be applied on the bypass crossings as well. Thus the cross-domain bypass connections must come after the level shifter insertion. This way the sequence is unique and no different order can be used.

The results of the synthesis state are reported in **Table 4-3**. The design includes about 20-25k gates. As expected, most of the complexity comes from the processor core that includes about 1/2 of the total gates. The peripherals on the APB bus accounts to 1/3 of the gates, while the memory interface and various other modules occupy the remaining 1/6. The cell area is dominated by the memories, which is about a ratio of half. Also half of the power consumption is due to the memories, while another half is due to the core and the peripherals.

Module	Cells	Cell Area [μm^2]	Total Power [mW]
Memories and interface	1410	65505	1.170
Memory Side Bypass Logic	*389*	*515*	*0.018*
Cortex-M0+ core	9722	21789	0.574
APB bus peripherals	6705	17429	0.291
Up Level Shifters	*140*	*8767*	*0.030*
Down Level Shifters	*102*	*7017*	*0.048*
Clock Multiplexer	569	1339	0.239
General Purpose IO	407	950	0.015
AHB bus	531	843	0.061
Glue Logic	92	219	0.079
Core Side Bypass Logic	*420*	*500*	*0.008*
Glue Logic 2	59	98	0.006
Reset Control	25	70	0.002
Pin Multiplexer	36	44	0.000
TOTAL - top domain*	**1799**	**66020**	**1.188**
TOTAL - bottom domain*	**18566**	**43281**	**1.274**
TOTAL	**20607**	**125085**	**2.539**
TOTAL Level shifters & Bypass	***1051***	***16799***	***0.104***
Overhead	***5.1%***	***13.4%***	***4.1%***

Table 4-3: Design complexity and area estimation based on synthesis
(*without level shifter power)

An important figure of merit is the overhead coming from the signal interface circuits necessary for stacking. These cells are the level shifters and the bypass logic (denoted in green in **Table 4-3**). It can be seen that the overhead mostly comes in cell area, which is about 13%, while the cell count is about 5% of the total number of gates. The additional power consumption coming from the signal interface is less than 5%.

The power overhead will decrease the power reduction advantage of stacking. This can be quantified in the following way:

$$P_{in,flat} = \frac{P_{out,flat}}{\eta_{SC}}; \; P_{in,stacked} = \frac{P_{out,stacked}}{\eta_{stacked}}$$

$$\eta_{stacked} = \frac{2 + \frac{\Delta I}{I}}{2 + \frac{\Delta I}{I}\frac{1}{\eta_{SC}}}; \; P_{out,stacked} = 2.539 mW;$$

$$P_{out,flat} = 2.539mW - 0.104mW = 2.435mW$$

Equation 4-1: The definition of power efficiency and the delivered power at the output of the power delivery

Calculating with a 65% efficiency for the switched capacitor regulator, the input power from the battery of the flat configuration is,

$$P_{in,flat} = \frac{2.435mW}{0.65} = 3.74mW$$

Equation 4-2: Example input power of flat system

The current mismatch can be approximated from the top and the bottom power numbers. Then the input power for the stacked mode is

$$\frac{\Delta I}{I} \approx \frac{\Delta P}{P} = \frac{1.274mW - 1.188mW}{1.188mW} = 0.072$$

$$P_{in,stacked} = 2.539mW \frac{2 + 0.072\frac{1}{0.65}}{2 + 0.072} = 2.58mW$$

Equation 4-3: Example input power of stacked system

We can see that even with the $0.104mW$ additional power that is consumed by the interface logic, we have saved more than $1mW$ by stacking the system, if we assume 65% percent power delivery efficiency. The power savings is due to the increased efficiency of the power delivery system. If the switched capacitor efficiency is higher, for example 85%, then the power saving reduces to $0.3mW$. The 85% efficiency value, however, is difficult to achieve for on-chip bulk CMOS switched capacitor converters.

4.2.2 Layout, Integration

The digital back-end flow consists of defining the chip floorplan, placing the macros and the standard cells, synthesizing the clock tree and routing the power and the signal nets. The result of the floorplanning and the placement can be seen in **Figure 4-14**. The bottom power domain consisting of the Cortex-M0+ core and the peripherals are placed on the lower half of the central area, while the top power domain with the ROM, ISRAM and DSRAM are placed in the upper part. The two power domains are connected through the signal interface with level shifters and bypass cells, as can be seen from the placement. The central area with the core

and the memories take 170100µm^2, while the level shifter area of 28350µm^2 gives 1/6 overhead. This is somewhat higher than expected from synthesis (**Table 4-3**). The central area is surrounded by four voltage regulator banks. There are two interleaved chains of the regulators, the north-west-south bank with 14 regulators for chain 1 and the east bank with 6 instances for chain 2. The regulators are all active to provide power in the conventional flat mode. They occupy a gross area of 368800µm^2, which is more than double of the central area. It is worth noting, however, that here thick oxide capacitors have been used for low leakage. Using thin oxide capacitors roughly halves this area. In stacked mode, only chain 2 needs to be active to regulate the middle node of the system. This reduces the required area to 100200µm^2, an area reduction in converter area of more than 3 times.
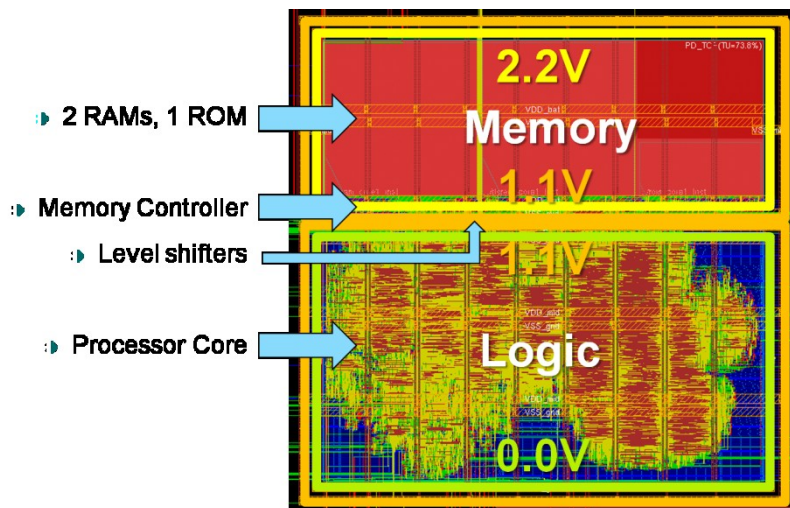
Figure 4-13: The Layout Concept

This means that the system power density will increase, even though the DC-DC converter power density is the same for both modes. The regulator has been sized to deliver 4.5mA current at 1.1V when both chains are active, which corresponds to a 12mW/mm^2 power density. If that power is delivered in stacked mode, the effective power density of the system can become higher than the regulator power density, since only chain 2 is needed to provide the mismatch current between the stacks. That way, the power density becomes 45mW/mm^2. This can be measured in standalone mode when the regulator is loaded externally and the rest of the system is powered down.

	Stacked Area [μm^2]	Flat Area [μm^2]
µP & Memories & Peripherals	170100	141750
Level Shifters and Bypass	*28350*	*-*
Voltage Regulators	*100200*	*368800*
TOTAL AREA	**298650**	**410570**

Table 4-4: Area Comparison between stacked and flat design
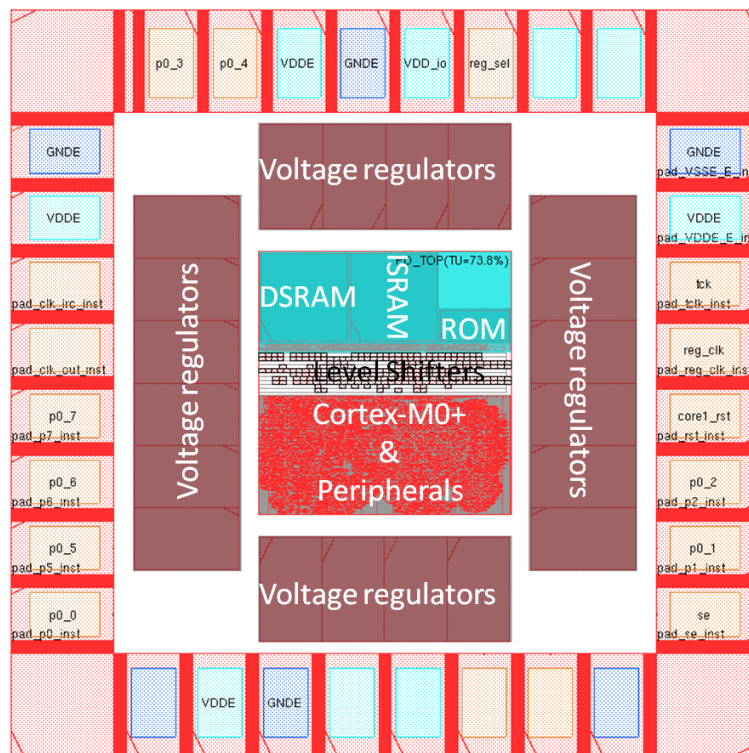
Figure 4-14: Chip Floorplan and Placement

In the synthesis step, the clock network was assumed to be ideal. The reason behind this is that even though a clock network might work in the front-end simulations, the placement and routing can greatly impact the timing of the design. That is why the clock tree synthesis, instead of being performed along with the rest of the synthesis steps, is typically executed after the placement. As for the rest of the digital flow, this step is also mostly automated. However, there are plenty of parameters that can be specified to reach an optimal timing-power-area tradeoff. In this research project the timing was not very constrained, however, the power of the clock network had to be minimized as much as possible to achieve best possible power

matching between the power domains. Due to these considerations, the clock buffers were constrained to be up to 8x drive strength, while the clock uncertainty was relaxed to be 20% of the period, 2ns. The maximum allowed transition time of the clock buffers was 350ps, which is 3-4% of the clock period. Based on these considerations, the setup and hold timing slack reported in **Table 4-6** and **Table 5-2** has been achieved based on sign-off signal integrity timing analysis. The clock tree power for the stacked mode and nominal conditions was between 0.40mW and 0.82mW depending on the code that was executed.

SETUP ANALYSIS						
		Corner				
Operation Mode		P	V	T	Cap	Slack [ns]
Functional	Stacked 80MHz	ss	0.99V	125°C	MAX	0.335
		tt	1.1V	25°C	MAX	3.446
	Flat 100MHz	ss	0.99V	125°C	MAX	0.17
		tt	1.1V	25°C	MAX	3.803
Test	Stacked 10MHz	ss	0.99V	125°C	MAX	87.452
		tt	1.1V	25°C	MAX	92.739
	Flat 10MHz	ss	0.99V	125°C	MAX	89.452
		tt	1.1V	25°C	MAX	94.756
Worst Slack						0.17

Table 4-5: Setup analysis results

For the worst setup corner and worst hold corner, the slack time remains over 100ps, while for typical corner the setup slack is over 3ns and the hold slack is 258ps.

HOLD ANALYSIS						
		Corner				**Slack [ns]**
Operation Mode		**P**	**V**	**T**	**Cap**	
Functional	Stacked 80MHz	tt	1.1V	25°C	MAX	0.258
		ff	1.21V	-40°C	MAX	0.167
		ff	1.21V	-40°C	MIN	0.106
	Flat 100MHz	tt	1.1V	25°C	MAX	0.98
		ff	1.21V	-40°C	MAX	0.893
		ff	1.21V	-40°C	MIN	0.887
Test	Stacked 10MHz	tt	1.1V	25°C	MAX	0.33
		ff	1.21V	-40°C	MAX	0.194
		ff	1.21V	-40°C	MIN	0.195
	Flat 10MHz	tt	1.1V	25°C	MAX	1.664
		ff	1.21V	-40°C	MAX	1.643
		ff	1.21V	-40°C	MIN	1.644
Worst Slack						**0.106**

Table 4-6: Hold analysis results

During the power routing, it has been decided to separate all the power nets. Keeping separate rails enables measuring the currents and calculating the efficiency values individually. Furthermore, it enables operating the system in the flat mode where all the digital supply rails are at V_{dd} and the digital ground rails at 0V. To fulfill these requirements, the following rails were proposed:

- **Bottom Power Domain:** VSS ground (0V), VDD middle node (V_{dd})
- **Top Power Domain:** VSS middle node (V_{dd}), VDD battery ($2V_{dd}$)
- **Regulator Domain:** VSS ground (0V), V_{out} regulator (V_{dd}), VDD regulator ($2V_{dd}$)
- **IO Domain:** VSSE (external 0V), VDD IO (core side V_{dd}), VDDE (external $3V_{dd}$)

These rails have separate IO pads each so that externally they can be connected into any topology. The first power ring has been implemented over the central area with the top and the bottom power domain nets, shown in **Figure 4-15**.
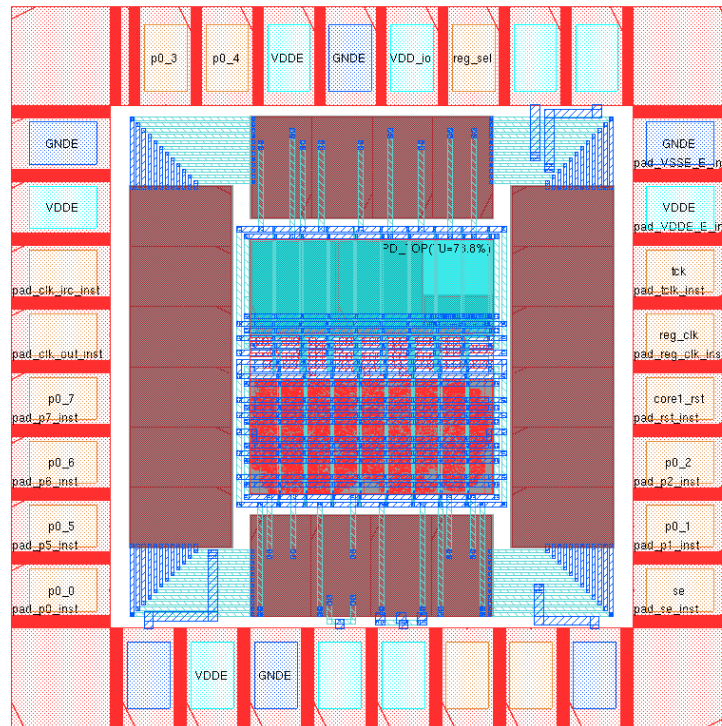
Figure 4-15: Power Routing

The second ring is routed through the voltage regulators, and contains, in addition to the core area supply nets, the voltage regulator supply nets. Finally, the third power ring is in the IO ring, which includes the core side and IO side power nets. This triple ring structure ensures low IR drops during operation.

Before finalizing the design, the actual "stacking" of the system is yet to be done. Currently the substrate of NMOS devices and n-well of PMOS devices in the standard cell rows are connected to the power and ground net of the power domain. Since bulk CMOS process is used, this would cause a short between the NMOS p-type substrate of the top domain, which has to be biased to 1.1V, and that of the bottom power domain, which has to be biased to 0V. Separation of NMOS substrates is necessary. Since triple well is possible in the current process, a deep n-well can isolate the p-substrate of the top domain from the rest of the bulk, as depicted in **Figure 4-16**.
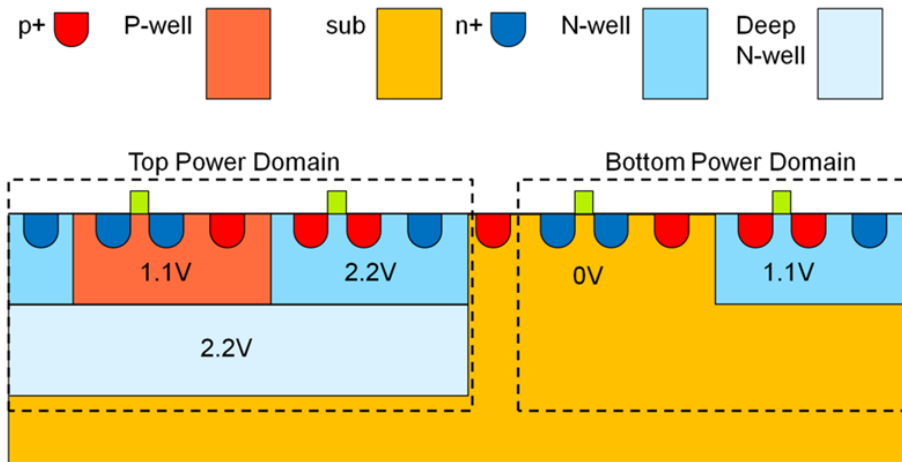
Figure 4-16: Deep n-well placement in the stacked system

The deep n-well is biased to 2.2V and makes connection with the n-well of the PMOS devices, which is also biased to 2.2V. Furthermore, an n-well guard ring is added at the boundary of the deep n-well to fully isolate the triple p-well from the bulk. This way the NMOS p-type substrate in the top domain can be biased to 1.1V and the bulk can be biased to 0V where the bottom power domain NMOS devices are placed at. Latch-up analysis has also been done of this scheme. The first step in latch-up analysis is to identify each BJTs that can contribute to a possible latch-up event. To simplify this search, the sketch of the different doping regions in **Figure 4-17** can be of help.
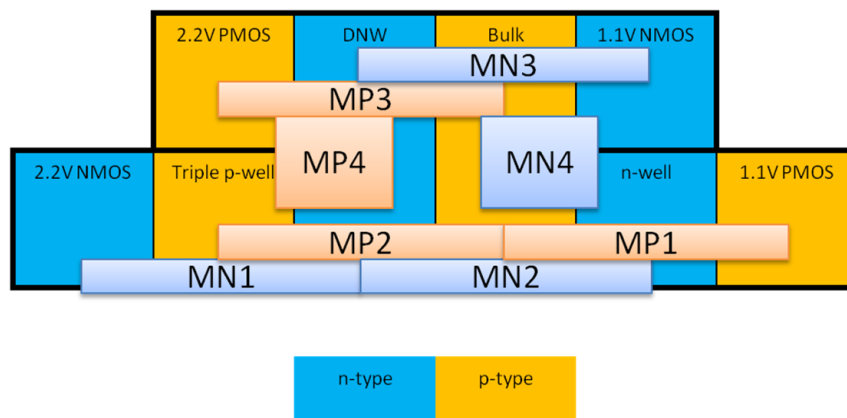
Figure 4-17: Systematic search of latch-up BJTs using the doping region graph

Three alternating vertices of p- and n-type regions results in a parasitic BJT, as it has been drawn. From **Figure 4-17** it can be seen that the doping region graph is symmetric, which means that for each PNP transistor there is a corresponding NPN transistor and vice versa. This simplifies the search. In total eight different BJTs have been considered. Once the devices have been identified, they can be paired with each other to see if they can cause latch-up. Latch-up can form between an NMOS and a PMOS device in the case when they share two terminals. In **Figure 4-18** the resistors leading to the power and ground connections have been colored. Those BJT pairs have to be analyzed which are opposite type and thus have different color resistors connecting to two shared nodes. For example, a classical latch-up scenario occurs with *MP1* and *MN4*. There is no stacking needed for this scheme to be present in any digital ASIC. The mentioned NMOS-PMOS pair shares the bulk and the n-well nodes, and their resistors can be united to make a latch-up circuit.
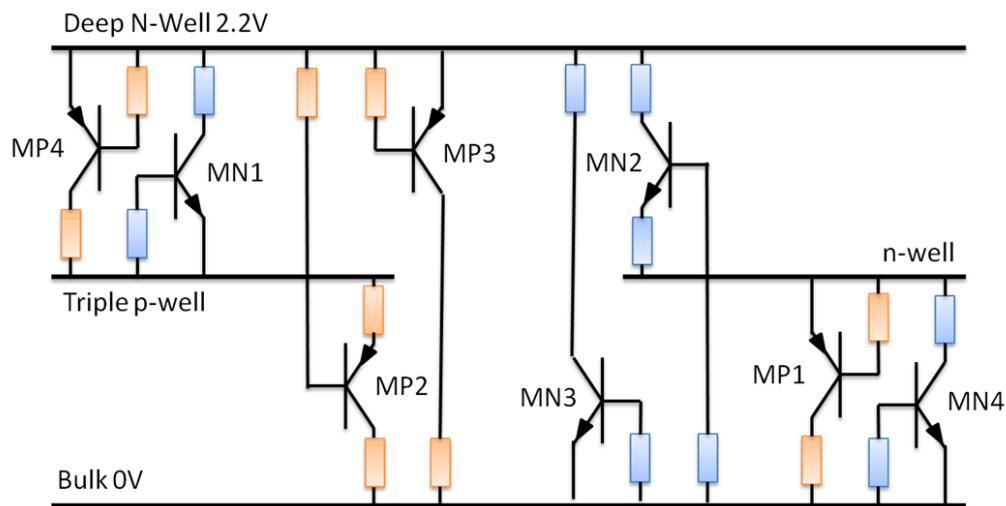
Figure 4-18: Systematic search of possible latch-up scheme among the parasitic BJTs

In a similar way a couple of other devices can be paired. In **Table 4-7** there is a summary of the possible latch-up events. Apart from the classical case, there is couple of possibility for latch-up, but with an increased voltage headroom. Since these latch-up events are heavily dependent on the deep n-well and the bulk series resistance, special attention has to be paid for placing guard rings for both in the system.

MP1-MN4, MN1-MP4	Classical latch-up situation
MN3-MP3	Classical latch-up situation with 2.2V supply
	Deep n-well and bulk guard ring has to be placed
MN2-MP2	No latch-up
MN1-MP2, MP1-MN2	No latch-up
MN2-MP3, MP2-MN3	Latch-up possible, but less likely
	resistance to triple p-well/n-well for MP2-MN3/MN2-MP3
	Deep n-well and bulk guard ring has to be placed

Table 4-7: Summary of latch-up situations

The deep n-well insertion was done manually after the digital back-end stages, together with the final steps e.g. tiling. The final layout of the test chip can be seen in **Figure 4-19**. The silicon area used is1.44μm^2.
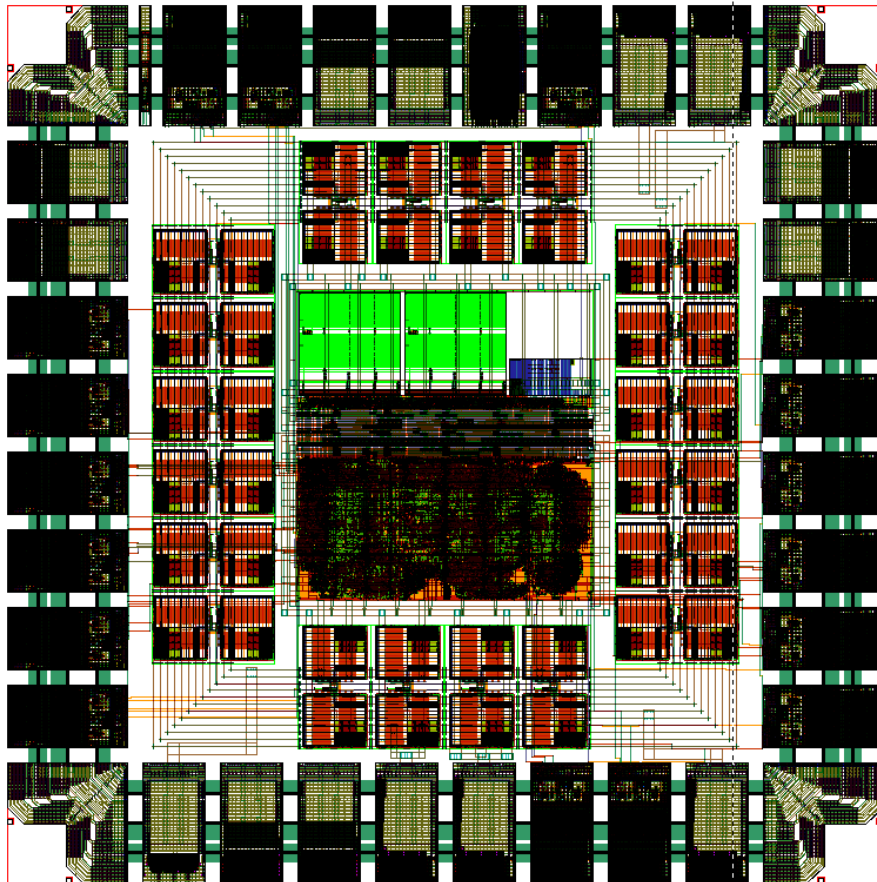
Figure 4-19: Final Layout of the Test Chip

4.2.3   Verification

Through every stage in the digital design flow, verification is necessary to ensure that the system works after executing a particular design stage. There are different types of verification methods used at different stages. One of the first steps is functional verification, where the design is checked whether it behaves the way intended at the system level. In our case it means simulating the RTL level code with a testbench stimulus that covers most of the functionality. The functionality verified is the following:

1. The system can work in flat and stacked modes
2. The processor can run various programs correctly from the both ROM or the ISRAM
3. The ISRAM can be programmed externally through serial wire interface
4. The UART interface, the timer module and the General Purpose IO work properly and can generate interrupt to the Cortex-M0+ core
5. The system generates a clock output which is a replica of the main system clock

The stacked/flat operation was verified only after the netlist was synthesized with the level shifters and bypass circuits inserted. Point 2 was verified by executing various test programs from ROM and ISRAM, as it was mentioned in Section 3.1.1. Both memories were tested with a *while (1)* loop, and with two high activity algorithms – a matrix multiplication and an image FIR filtering program. These test programs can be executed as described next. The bootloader sequence in the ROM is executed following the system start-up. Here, it is possible to control the behavior of this program with external pins the way it is shown in the flowchart of **Figure 4-20**. If GPIO pin 1 is set to **'0'**, the program checks if the ISRAM is programmed by looking for a special mark word at a given ISRAM memory address. If it finds the mark word there, then it assumes that the ISRAM is ready to execute the code, and the PC pointer is updated to the ISRAM location. If the mark is not there, the processor goes into a *while (1)* loop and it only recovers from it upon a reset. This way it is possible to program the ISRAM through the serial wire interface then reset the core and pass the PC pointer from the ROM to the ISRAM, as it can be seen in **Figure 4-21**. It is also possible to execute a test program from the ROM itself, for this GPIO pin 1 has to be set to **'1'**. In this case either a matrix multiplication algorithm (GPIO pin 5 **'0'**) or an FFT (GPIO pin **'1'**) is executed.
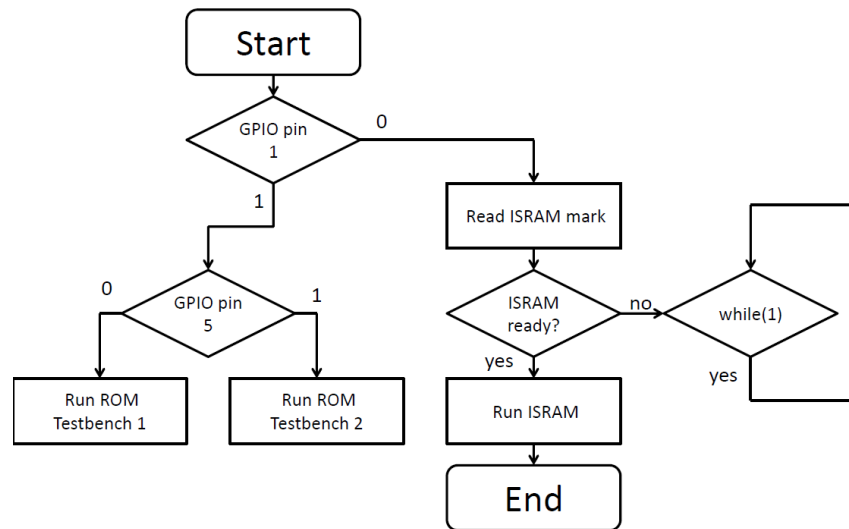
Figure 4-20: Program flow chart of the bootloader sequence

The various ways of executing the different programs from different memories is crucial for balancing the power of various blocks within the design, as was suggested in **Figure 3-8**. The four possible outcomes in **Figure 4-20** will correspond to four different switching activities. In addition, the ISRAM code can be freely chosen, further stretching the possibilities.
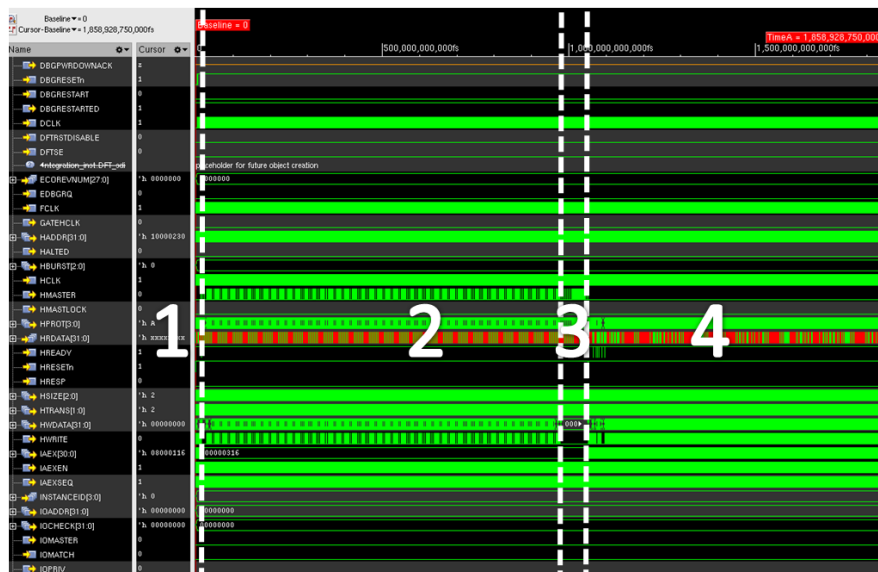
Figure 4-21: The core signals of the processor during ISRAM programming check.
1: startup 2: ISRAM programming 3: verifying part of written code 4: ISRAM code execution

The UART, timer and GPIO need to be enabled and configured through the program that the processor executes. In the testbench used, the UART is connected in a loopback test setup. Single characters are sent out through the transmitter, periodically. The UART receiver stores the received values in a shift register. The next character to be sent by the transmitter is always the last value that has been read in. This can be overwritten by a new value. This way the UART is going to circulate the same value until the program does not write a new value for transmitting, in the receiver buffer. The waveform of this test can be seen in **Figure 4-22**. An 8-bit counter's output character sequence is used to test the UART transceiver. After the Start bit, which is a *'0'*, comes the LSB and so on. First a double repetition of 0, then triple repetition of 1, 2, 3 numbers can be seen. While repeating the 3, an external interrupt arrives on *p0_4* GPIO pin. For each interrupt the UART reacts by sending out a special character.
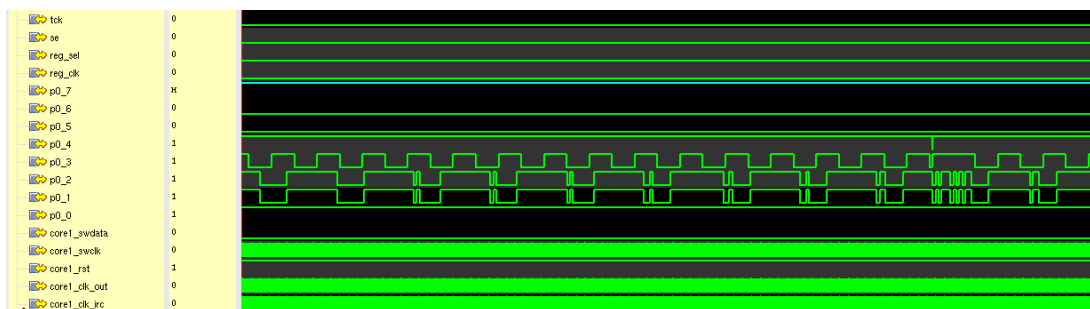
Figure 4-22: Test Chip Digital Pins. UART, timer, GPIO and clock out test signals. GPIO pin configuration **p0_0**: '0' – ISRAM code '1' – ROM code **p0_1**: UART TX **p0_2**: UART RX **p0_3**: Timer LED signal **p0_4**: external interrupt **p0_5**: '0' – Testbench 1 '1' – Testbench 2

The timer circuit is contrlling an external LED to blink periodically, at GPIO pin *p0_3*. For simulation purposes, the frequency has been accelerated by a factor of 1000. In the final bootloader there will be a 1 second long period for the LED blinking. Also for external interrupt at pin *p0_4*, the timer LED signal changes polarity, which is shown again on **Figure 4-22**. Finally, the output clock signal, *core1_clk_out* is shown that is sent out from the chip to prove that the clock signal is active.

The tests performed above must be reproducible with the synthesized netlist and also the final netlist that is generated upon layout finishing. To perform the netlist simulations with accurate timing values, the nets have to be annotated with an approximation of their capacitance and their drivers must provide their drive strength information. This information is stored in the

SDF file format. Using the netlist with the SDF file it is possible to make a simulation with better timing values. After the synthesis, the clock signals are assumed to be ideal even in the SDF since the clock tree is not yet synthesized. In the final, layout-annotated netlist every signal is subject to non-ideal delays, including the clock signals.

An important task to perform during netlist simulation is the verification of the characterized level shifters and the overall signal interfacing block with the bypass circuitry included. It is not enough to check the program execution, but also it is necessary to see all the signals in the signal interface block to determine if the signal was traveling through the level shifter or the bypass circuitry. Such a test can be seen in **Figure 4-23**. The waveform shows the path of the ROM clock signal from the AHB (*rom_clk_CORE* signal) to the ROM IP itself (*rom_clk_MEM* signal). The clock signal arrives in both operating modes to the destined location, but through different paths. The *tie0* and *tie1* signals control whether it goes through the level shifter (*rom_clk_CORE_LS* is the level shifter input and *rom_clk_MEM_LS* is the level shifter output signal), or the bypass path (*rom_clk_0_MEM_LS_bypass_in* signal).
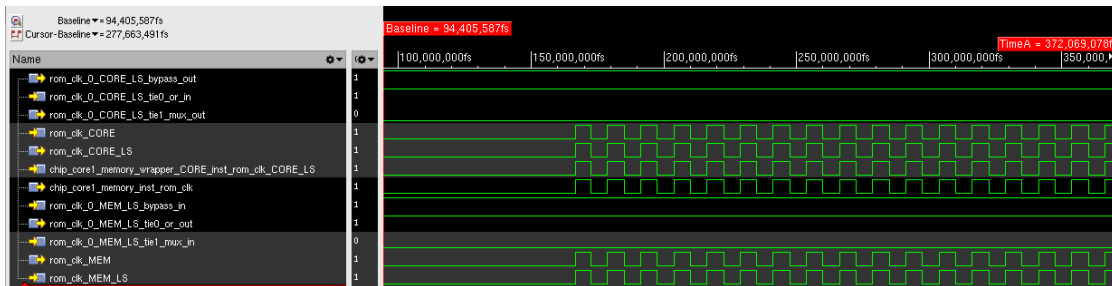
Figure 4-23-a: The ROM clock signal in stacked mode at 80MHz
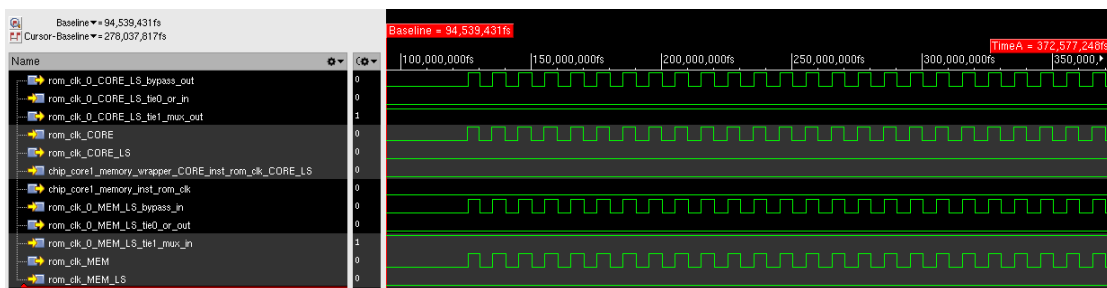
Figure 4-23-b: The ROM clock signal in flat mode at 100MHz

Once the netlist simulation has been performed, there is another type of verification – the power reporting. The power numbers are especially important in this project. The problem is that the CMOS dynamic power consumption strongly depends on the switching activity of a gate. If the gate is not switching, there is no dynamic charge transfer and thus no dynamic power. This is why it is very important to know the switching activity of each circuit node. The switching activity data can be collected from the netlist simulations, and also can be used during synthesis from the behavioral simulation. Using the TCF file that stores the switching activity of each net, accurate power reports can be generated.

(a) Synthesis Power estimation [mW]   (b) Layout-Annotated Power estimation [mW]
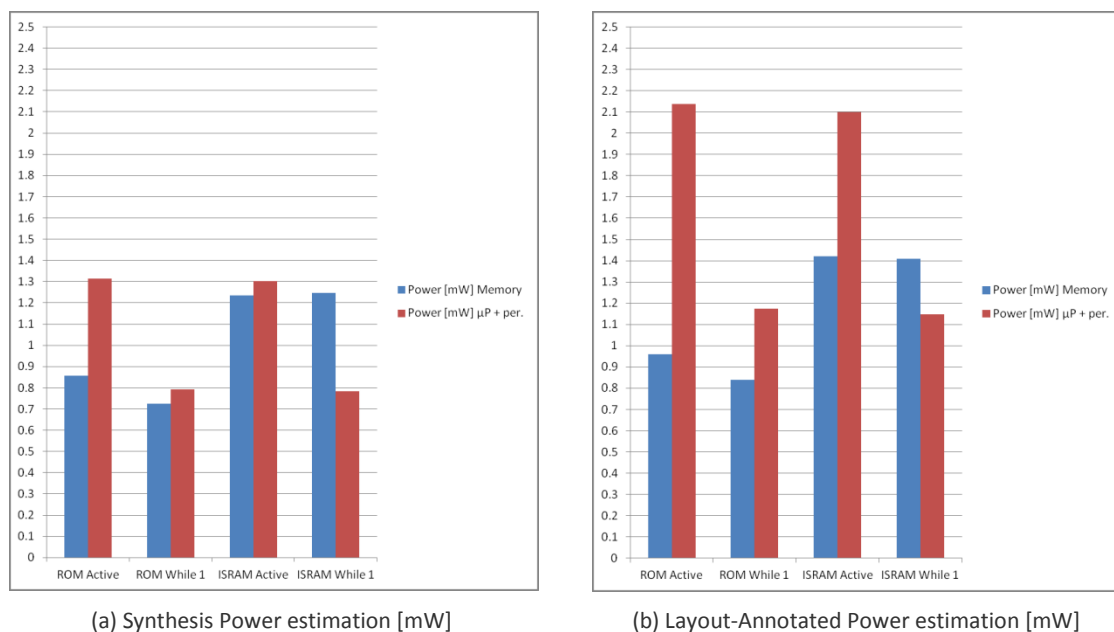
Figure 4-24: Power Consumption over different test programs

The functional verification and the power reports still do not give insight into the physical connections of the design or the proper power and ground net connections within the netlist. This is why there are specialized EDA tools to check the power intent in the CPF file and compare it with the netlist that has been generated. For this purpose Cadence Encounter Conformal Low Power software has been used.

The Design Rules Check (DRC) of the final layout was done with Cadence Physical Verification System, and so was the Layout Versus Schematic (LVS) performed.

4.3 Stacked Power Delivery System

The stacked power delivery system requires a voltage regulator that provides the voltage in stacked mode for the ground node of the top domain and the supply node of the bottom domain, while in flat mode for the supply nodes of both top and bottom power domains. It has been decided to use a voltage halving switched capacitor topology depicted on **Figure 3-13**. The maximum output power in stacked mode should be no more than 1.5mW and in flat mode maximum 4mW, according the power estimation of the laid out system in **Figure 4-24**. The target peak efficiency considering the limitations from technology is set at 65%. That means that on schematic level the efficiency should reach 90%, while after layout 80%. This requires careful selection of the building blocks, especially the capacitor which dominates in the switching losses. At the start-up phase, as well as during operation, there also should not be any voltage spike that can reduce the lifetime of the devices. It is also important to take into account the inductance of the bonding wires and the PCB interconnect for the power delivery, since these can degrade the performance significantly. All the considerations, however, start with defining the key parameters of the regulator.

4.3.1 Architecture

The switching frequency has to be adequately chosen for the switch sizes (thus on-resistance) and for the capacitor size. The area limitations coming from the chip floorplan in **Figure 4-14** also have to be met. In order to calculate these values, the well-known method of slow- and fast switching limit output impedance calculations are used. The slow switching limit can be derived by setting the input voltage to zero and placing an ideal voltage source at the output of the converter. This way, the current that flows due to this test voltage will be proportional to the output impedance. This is a generalization of Ohm's law for multi-phase circuits. From Tellegen's theorem considering the charge flow through the capacitors and expressing the output voltage as function of output current, the slow switching limit output impedance R_{SSL} can be derived. The fast switching limit output impedance R_{FSL} comes from assuming constant current flow through the switches and summing up the power losses, then dividing it with the output current. In **Equation 4-4** the formula from literature has been used for the calculations. For further details and proof of these formulas, please refer to [54].

(a)
$$a_C^1 = \frac{1}{2}, a_C^2 = -\frac{1}{2}; \quad R_{SSL} = \sum_{j=1}^{2} \frac{\left(a_C^j\right)^2}{2Cf_{sw}} = \frac{1}{4Cf_{sw}}$$

$$a_r^1 = \begin{bmatrix} 0 & -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}, a_r^2 = \begin{bmatrix} -\frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}, R = \begin{bmatrix} R_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & R_4 \end{bmatrix}, D_j = \frac{1}{2};$$

(b)
$$R_{FSL} = \sum_{j=1}^{2} \frac{a_r^j \cdot R \cdot \left(a_r^j\right)^T}{D_j} = 2R$$

(c)
$$R_{SSL} \approx R_{FSL} \Longrightarrow R_{sw} = \frac{1}{8Cf_{sw}}$$

Equation 4-4: Calculating the switch resistances of the 2:1 SC converter

The main limitation in the design is the silicon area, which in turn is determined mostly by the capacitor size. In the current technology an accumulation PMOS capacitor with 20pF capacitance occupies a rectangle of 70µm by 85µm. This means that two of these capacitors can be placed within one voltage regulator block for a total of 20 voltage regulator blocks, which makes the total capacitance 40·20=800pF. The switching frequency can be freely chosen, but for switching loss considerations it is maximized at 25MHz. This corresponds to a slow-switching resistance value R_{SSL} of 12.5Ω. Due to theoretical observations, the optimum R_{FSL} value for a given R_{SSL} in order not to limit the output impedance and on the other hand not to impose large switching losses, occurs when the two output impedance components are equal. From **Equation 4-4** the optimum on-resistance of the switches is found to be half the R_{SSL} value, 6.25Ω. The total output impedance can be approximated in the way it is done in **Equation 4-5**.

$$R_{out} \cong \sqrt{R_{SSL}^2 + R_{FSL}^2} = \sqrt{2} \cdot 12.5\Omega = 17.7\Omega$$

Equation 4-5: Output Impedance of the proposed DC-DC Converter

For the maximum output current of about 5mA, the output impedance calculated contributes with an IR drop of 89mV. This is 8.1% deviation from the supply, which is too high for the goal of 5%. Due to this it has been decided to size the switches double the size with 12.5Ω on-resistance, to leave the possibility of an operating frequency of 50MHz open. This would

increase the switching losses and reduce the efficiency, but would reduce the supply variation as well to 45mV, which is about 4.1% of the full-scale supply.

4.3.2 <u>Schematic</u>

The gate level implementation of the voltage halving topology is straightforward and is known from literature [55]. The designed schematic is depicted in the **Figure 4-25**. The clock signal is split into two branches along the power domains. The branch belonging to the top power domain is level shifted by an up level shifter described in Section 4.1, while the clock path in the bottom power domain employs a delay cell to account for the delay of the level shifter. After these stages, both clock signals are further split into two branches by a non-overlapping signal generator. This is necessary to prevent short-circuit currents between the switches of the SC converter. Once the four clock signals are generated, they are buffered through switch drive inverters with FO4 delay scheme and finally arrive at the gate of the switch transistors. The switches are implemented by two large inverters that connect their output either to their power or ground node. The inputs are split into PMOS and NMOS inputs and are controlled by the two non-overlapping waveforms that have been generated. In *phase 1*, when the input clock signal is high, the positive terminal of the flying capacitor is connected to the $V_{dd,bat}$ rail through the top inverter's PMOS switch, and the negative terminal to the V_{out} rail through the bottom inverter's PMOS switch. In *phase 2,* the input clock is low and the fly capacitor is connected to V_{out} and $V_{ss,gnd}$ rails through the NMOS switches.

Since the regulator is 20 times interleaved, the switch on-resistance was chosen to be 20 times the overall switch resistance, that is, 250Ω. The transistor size was found by measuring the current for a 1.1V (-1.1V) gate-source and low drain-source voltage for the NMOS (PMOS) devices, and calculating the on-resistance. The resistance was dependent on the drain-source voltage given, so the maximum expected output ripple, 100mV was used as test voltage. Throughout the voltage regulator, except for the capacitor, the same devices have been used as in the standard cells in the digital flow.
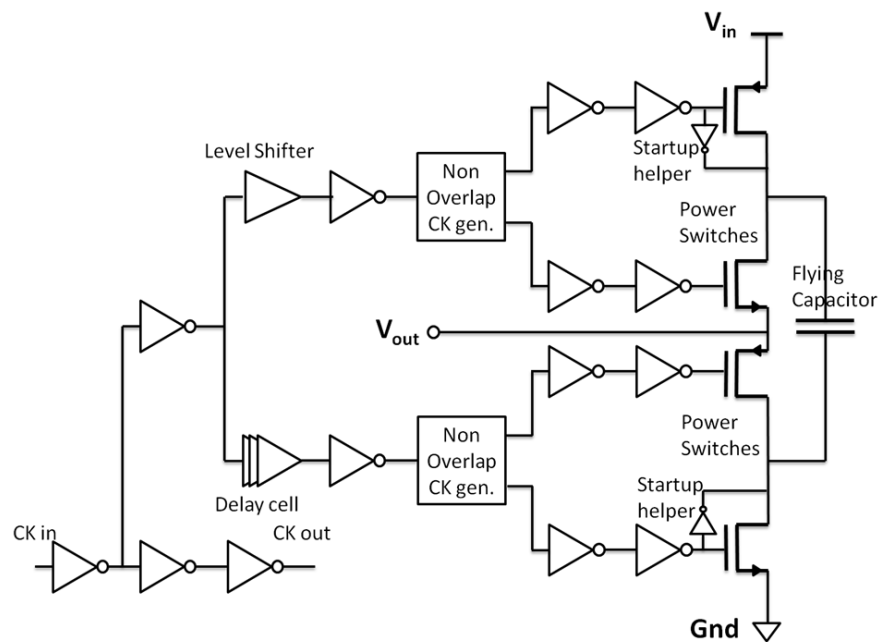
Figure 4-25: DC-DC Converter Schematic Design

To prevent voltage spikes over the fly capacitor during a supply start-up event, small inverters have been placed in parallel to the capacitor switches to pull the positive and negative terminals of the capacitor to a known voltage. These inverters have been sized to be strong enough to charge the parasitic capacitance of the floating capacitor nodes, while weak enough not to degrade the efficiency significantly.

The final and most critical component of the DC-DC converter is the fly capacitor. For higher capacitance-, thus power density, MOS capacitors are usually used. On the other hand, MIM capacitors have less parasitics and can allow higher efficiencies, but they require large silicon area. In this research project accumulation PMOS capacitors have been used, which consists of a PMOS structure placed within a triple p-well instead of an n-well. To minimize the parasitic bottom-plate capacitance, the triple p-well is connected to its enclosing deep n-well, since the deep n-well to bulk capacitance is smaller than the triple p-well to deep n-well. Furthermore, the drain and the source are also connected to the triple p-well. Since an accumulation PMOS capacitor is used, the drain-source-substrate terminal should be the positive terminal, while the gate the negative terminal. This motivates the choice behind the accumulation PMOS capacitor. Since the parasitic capacitance towards the bulk is most significant at the triple well

substrate connection of the capacitor, the charge stored on this capacitance should be used as optimally as possible. If the substrate was connected to the negative terminal of the capacitor, the charge would then be transferred between V_{out} and the ground, thus it would be completely lost in each cycle with no useful work (0% efficiency). The bottom plate capacitance would be discharged in this case in *phase 2* of the operation. On the other hand, since the parasitic capacitance is placed at the positive terminal, the charge is now transferred between V_{in} and V_{out}, and is not totally wasted. Since the charge is conserved and only the voltage is changed from 2.2Vto 1.1V, this parasitic charge is still delivered with up to 50% efficiency. This way, the efficiency of the converter can be maximized, despite the unavoidable parasitics of the flying capacitor.

4.3.3   Layout

The layout diagram is shown in **Figure 4-26**. The top power domain is placed within a deep n-well, while a second, hot deep n-well is used for the flying capacitor (deep n-wells are denoted with *T3*). The capacitor switches are directly connected to the in- and output power stripes, the ground and the capacitor positive and negative terminal. The fly capacitor is split into two 20pF instances.
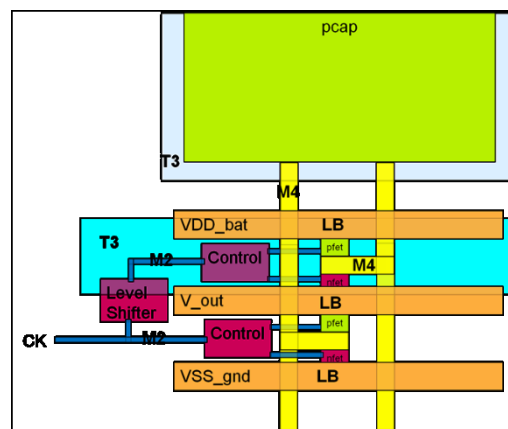
Figure 4-26: Voltage Regulator Layout Plan

The layout of the control logic and the switches can be seen in **Figure 4-27**. The clock signal enters at the bottom left and is split between the level shifter and the bottom domain control logic. The deep n-well has been placed according to the plans in **Figure 4-26**. Along the control logic, deep n-well connections have been placed to mitigate latch-up issues. Just like in the case

of the level shifter, the layout area here was limited by the deep n-well design rules. Unlike in the case of the level shifter, however, the power is not determined by the increased dimensions, but rather the flying capacitor and the control logic internal power.
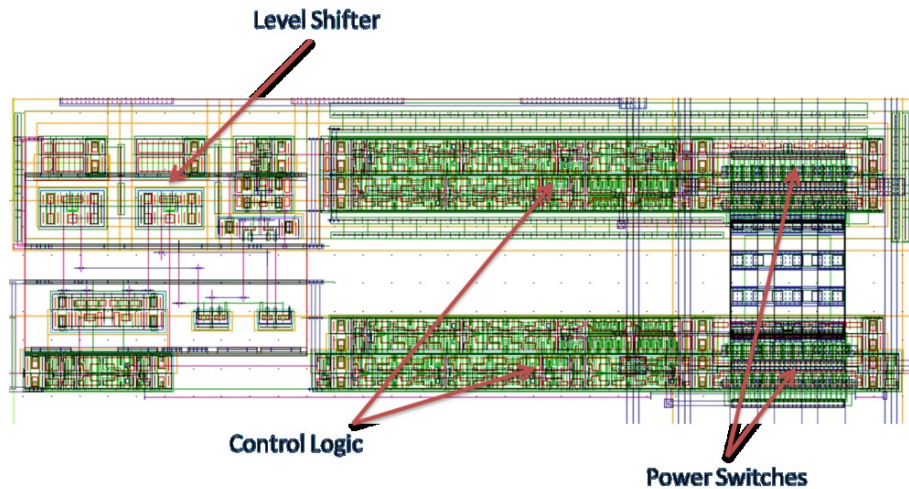
Figure 4-27: Voltage Regulator Logic Layout

The full top-level layout of one DC-DC converter block is shown in **Figure 4-28**. The logic occupies only $360\mu m^2$, which is slightly over 2.2% of the total area ($16800\mu m^2$). The flying capacitors, on the other hand, account for 71% of the DC-DC converter area. They are partitioned into an array of 8 row by 100 column for a total aspect ratio of $88\mu m$ by $70\mu m$. One capacitor device has a W/L ratio of 80:6 since the accumulation mode channel square resistance between drain and source is a lot higher than the resistance of the polycrystalline silicon gate. The capacitors are surrounded by a guard ring that provides connection to the bulk to avoid latch-up possibilities. To minimize the series resistance, the connection between the switches and the capacitor has been constrained to 10 squares per metal layer.

The layout has been designed so that routing tools can access the IO pads from the core. Above the third metal layer, the signals can be freely routed except for the control logic area, while below that it is still possible to route through the gap between two neighboring regulator cells. Due to the sufficient distance kept between the edge of the top level layout and the capacitors, the voltage regulator macros can be placed right next to each other, which also automatically

ensure the continuity of the power rails and the clock signal. At the corners of the regulator ring though, as it is shown in **Figure 4-15**, manual power routing had to be done.
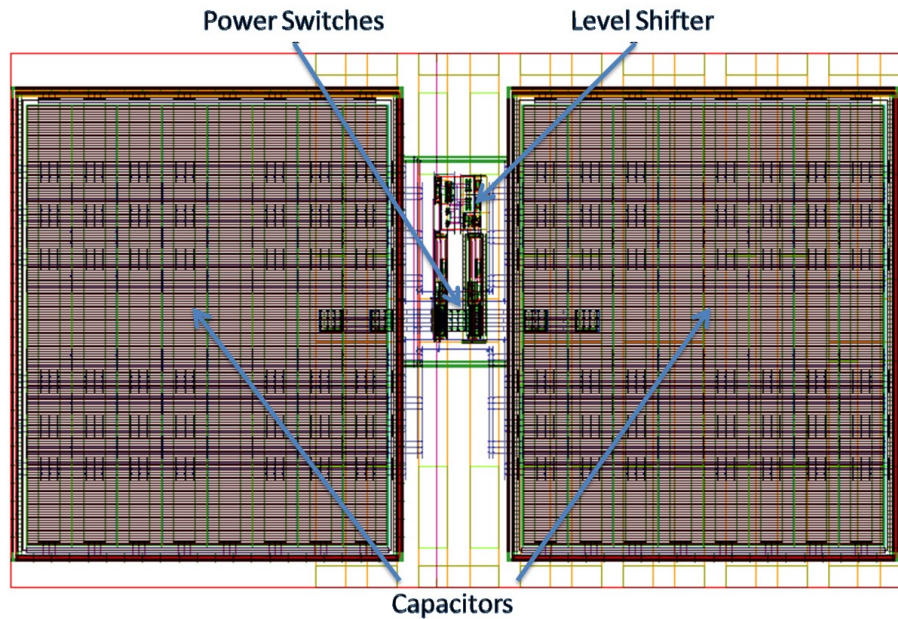
Figure 4-28: Voltage Regulator Top Level Layout

4.3.4   Verification

Among the most important properties of a voltage regulator are the efficiency and the power density values. **Figure 4-29** represents the efficiency over various load conditions, for 25MHz clock signal. The simulation has been performed on the extracted SPICE netlist of the designed regulator. The load of the regulator was chosen to be an ideal current source to make sure that the correct load current is sourced. The efficiency curve can be analyzed based on its shape. For low output current values, the switching losses dominate since even the unloaded regulator consumes some power, which makes the efficiency zero. The efficiency is thus low for small output current because the switching losses do not depend on the output current, are more or less constant in that respect. They depend, however, on the switching frequency, thus varying it will distort the efficiency plot. Increasing the load current to the other extremes will increase the $I_{out}{}^2 R_{out}$ conduction losses, and the efficiency will again start to decrease. This is because the conduction losses are quadratic in the output current, while the output power only increases linearly with the increase in the output current.
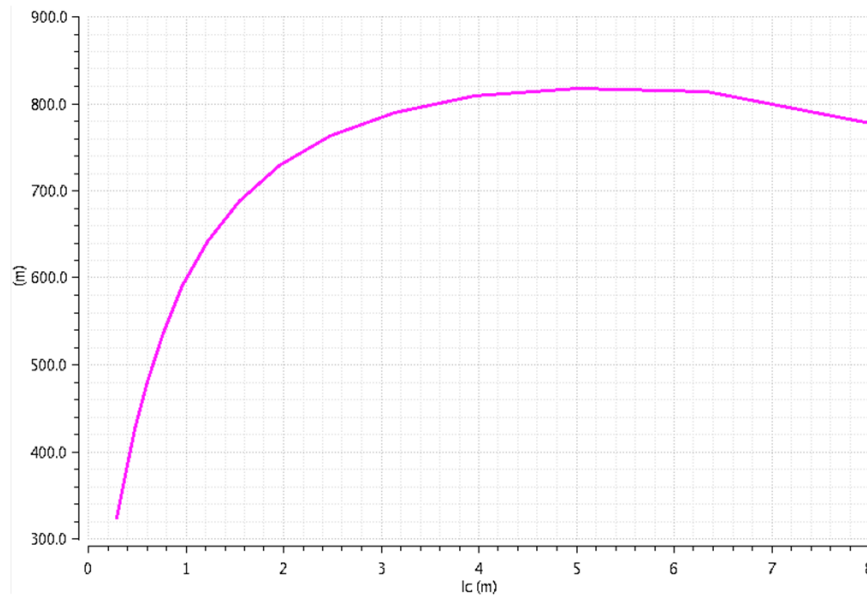
Figure 4-29: Efficiency Curve of the Voltage Regulator based on Layout-Annotated Simulation

It can be stated that the regulator peak efficiency meets the 80% requirement outlined before, at about 5mA load current. However, the peak might not be feasible to reach since the output voltage rail suffering from IR drop will have bigger ripple, and will deviate more from the nominal value. In **Figure 4-30** the output waveform for various current load values is plotted. It can be seen that the 10% allowed supply variation is fulfilled up to about 4mA output current. In the real test chip measurement thus the droop must be taken into consideration and a safer point on the efficiency curve should be selected. Nevertheless, close to 80% efficiency in the layout simulations can be obtained for loads between 3mA and 4mA.
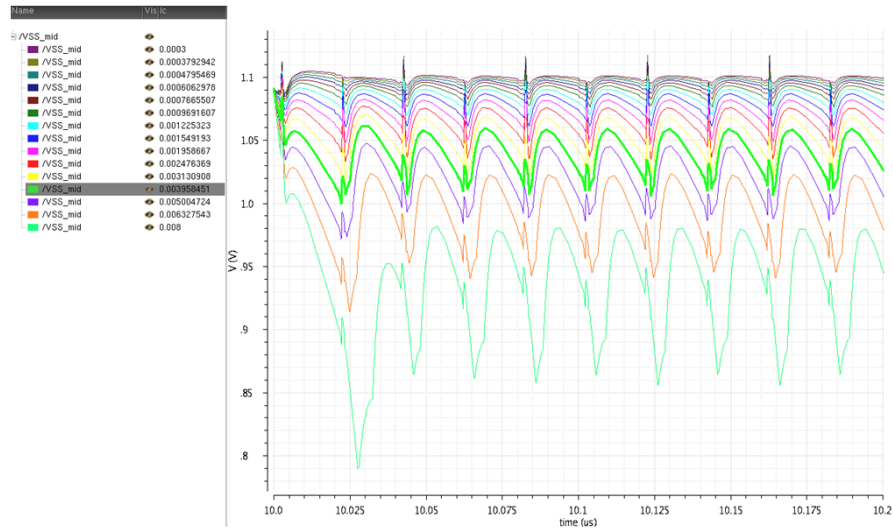
Figure 4-30: Output voltage waveform for various load currents

Not just the output voltage waveform, but all the important nodes of the regulator have been analyzed. One important consideration is that during start-up, the devices do not suffer from overvoltage for a longer time that can degrade their performance. Another function that must be checked is whether the non-overlapping clock signals have the correct delay and arrive in the intended sequence to avoid short circuit current. For these verification steps, the waveforms of **Figure 4-31** have been used.

(a)

(b)
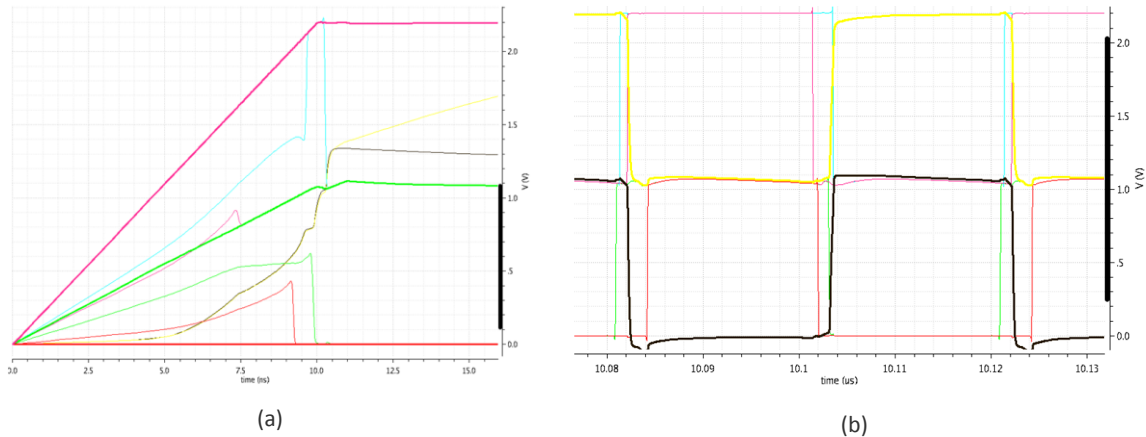
Figure 4-31: Voltage regulator start-up signals (a) and non-overlapping clock signals (b)

As was done for the digital back-end and for the level shifter, latch-up considerations also have been taken into account for the voltage regulator. In **Figure 4-32** the parasitic BJTs of the hot deep n-well are shown.
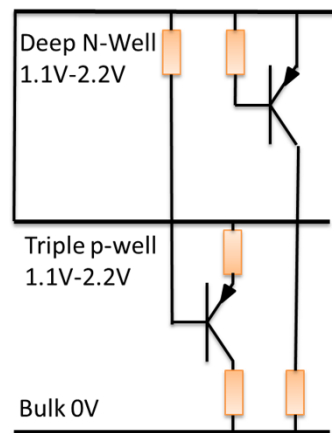
Figure 4-32: Parasitic BJTs of the flying capacitors

Though there is no latch-up situation, it must be ensured that the deep n-well does not change from 2.2V to 1.1V faster than the triple p-well, and that the triple p-well does not change faster from 1.1V to 2.2V than the deep n-well. In these cases the parasitic pnp transistor formed between the triple well, the deep n-well and the bulk might open. Stopping this behavior can be achieved by placing a ring of deep n-well contacts around the flying capacitor, as was done in this research project.

5 Chip Measurement Preparation

The proposed system is to be submitted for fabrication, which will most likely take couple of months. In that time it is important to make a ready plan for the test measurements and the various verification schemes. Once this is summarized, it is possible to make testing algorithms for different operation levels of the chip like logic functionality and power delivery. The planned test setup and the future work are also briefly described in this chapter.

5.1 Test Preparation

During the test preparation it is important to identify and separate parts of the system that have possibility of failure, and propose testing methods to filter out which block is failing. Such a proposal can be seen in **Table 5-1**.

Failure			Test	Indicator
Specific	**Block**	**Behavior**		
ISRAM side	**A:** ULS, DLS, Serial Wire **B:** UB, DB, Serial Wire **C:** MUX, ISRAM memory, controller	**A,B:** Stuck **C:** Any	Write and read from ISRAM through serial wire in stacked and flat mode.	<table><tr><td></td><td>A</td><td>B</td><td>C</td></tr><tr><td>Stacked</td><td>X</td><td>OK</td><td>X</td></tr><tr><td>Flat</td><td>OK</td><td>X</td><td>X</td></tr></table>
DSRAM side	**A:** ULS, DLS **B:** UB, DB **C:** MUX, DSRAM memory, controller	**A,B:** Stuck **C:** Any	Write and read from DSRAM through custom ISRAM code in stacked and flat mode.	
ROM side	**A:** ULS, DLS **B:** UB, DB **C:** MUX, ROM memory, controller	**A,B:** Stuck **C:** Any	Execute ROM code in stacked and flat mode.	<table><tr><td></td><td>A</td><td>B</td><td>C</td></tr><tr><td>Stacked</td><td>X</td><td>OK</td><td>X</td></tr><tr><td>Flat</td><td>OK</td><td>X</td><td>X</td></tr></table>
SC DC-DC	**A:** Reg. clock/IO **B:** ULS **C:** Capacitor, Switches **D:** non-overlapping clock generator, switch control	**A:** Stuck **B:** Stuck **C:** Short **D:** Timing Failure	Power up only the DC-DC converter and measure the current in.	Current **A:** Low **B:** High/half cycle **C:** High **D:** High spikes
AHB / APB	**A:** UART **B:** Timer **C:** GPIO Int. **D:** Clock out **E:** other	Incorrect Behavior	Execute proc. peripheral test program.	GPIO signals in stacked and flat mode
Core	**A:** Serial Wire **B:** AHB/ APB controller, Cortex-M0+ basic	Incorrect Behavior	**A:** Serial Wire program. **B:** Proc. program.	**A:** Serial Wire data **B:** Chip not working at all

Table 5-1: The key failure possibilities and testing methods
ULS: Up Level Shifter DLS: Down Level Shifter UB: Up Bypass DB: Down Bypass MUX: Bypass multiplexer

Since they involve special circuitry, the most critical blocks are the signal interface modules with level shifters and bypass circuitry. If neither the level shifter nor the bypass work, the core cannot access the ROM and the SRAMs, and no program can be executed. If the more

important part of the standard logic fails like the core or the AHB, the system is also not working. However this is less likely since they only consists of standard cells with standard voltage conditions.

For the chip testing and troubleshooting, it is important to keep the number of possible configurations at maximum. For example, it must be ensured that the clock frequencies and supply voltages can be adjusted, the power connections modified and different program executed. The key options can be found in **Table 5-2**.

OPTION		CONFIGURATION			
Program		ISRAM active	ISRAM while 1	ROM active	ROM while 1
Operation Mode		stacked	flat		
SC Converter	**Status**	on	powered down		
	Clock freq.	100kHz ... 50MHz			
	Chains	chain 1 active	chain 2 active	chain 1&2 active	all disabled
IO supply	**Core side**	connected to 1.1V supply	externally supplied		
	IO side	connected to 2.2V supply	externally supplied		
System Clock		1MHz ... (80/100)MHz			
System Supply		0.8V(logic)/1V(memory) ... 1.1V	powered down		

Table 5-2: Various configuration of the Chip

The different programs implemented are further described in Section 4.2.3, just as the flat/stack operation modes. The stacked and flat operation modes can be further customized by configuring the various blocks in the system. The core side IO power rails, for example, can be connected to the bottom domain, or supplied externally. The output efficiency curve of the SC converter can be shifted by activating either of the two regulator chains (**Figure 3-10**). Another option is the possibility to either disable the SC converter by stopping its input clock, or disconnecting it from the system entirely by connecting its input and output to ground. The

flexibility with the IO side power rails is a third possibility. They can be either supplied separately or connected to the 2.2V supply, counting it in as output power of the system.

5.1.1 <u>Functionality Test</u>

The possible failures and the various test modes are the basic building blocks of a test algorithm used for verifying the functionality of the test chip. The test algorithm follows a step-by-step approach. The chip first should be tested in the supposedly most failure-proof way. That is why the test sequence starts with the flat mode where level shifters are bypassed. Also, the IO ring is powered externally and the voltage regulator is powered down. This way the test chip needs a single 1.1V supply for the core, one 1.1V supply for the IO core side and another 2.2V supply for the external IO supply. The code to be executed from the ROM is the *while(1)* program.

If the chip is not working in this basic setup, then one assumption will be that there is some kind of error in the logic, since no custom-designed cells have been used in this configuration. In this case the serial wire interface can be used to access registers and memory addresses within the system and troubleshoot. Through the serial wire, the ISRAM also can be programmed and the system can be rebooted to execute custom code. Thus with the serial wire, basic memory test can be performed like writing and reading back from the ISRAM or DSRAM, or reading from the ROM. If not the same value has been read back from the ISRAM or DSRAM as was written, or there is wrong code in the ROM compared to the designed bootloader sequence, then these errors should be analyzed. If there is only localized error involving one of the memories, then it can be assumed with high probability that the memory or its controller in question have failure. Further tests can reveal the nature of this failure and possible plans to avoid it can be proposed. As an example, if there is a stuck at '0'/'1' in the SRAM, then the program execution can be modified by placing a jump command to avoid that memory location. If the ISRAM memory controller fails, which is less likely, then the solution largely depends on the severity of the error. In some cases e.g. when only part of the memory can be addressed, the operation can be ensured by constraining the memory location of the program. If the ROM controller fails, then there is much less freedom to make changes. The same applies if there is failure either in the core or the AHB. These cases can be analyzed by

studying the contents of the registers. Though not utilized on the test PCB board, a scan path has also been implemented on the chip to be used in case none of the other methods yield results. In this latter case the scan tester must be integrated into the test setup by using a new PCB board.

If the serial wire is unable to access the memories, then another possible source of error can be the bypass circuitry. To deactivate it, the system can be tested by switching into stacked mode. If there is no difference in the behavior between stacked and flat mode, then, as **Table 5-1** suggests, there probably no error in the signal interface, thus the level shifters and the bypass. However if the system either works in stacked (flat) mode but not in flat (stacked), then it is likely that the bypass circuit (level shifter circuit) is not working. In that case the system can only be operated in one working mode.

The DC-DC converter can be operated and measured independently from the rest of the system, so next to the flat mode testing for the digital part, the voltage regulator standalone testing is another default mode that has less chance of failure. In case one of the chains is not working, the other chain can be used. If none of the chains are working, then the input and output voltage rails can be analyzed under different test conditions to determine the nature of the error. The causes can be identified according to **Table 5-1**.

Finally, if both the DC-DC converter and the digital system work, they can be connected so that the former generates power to the latter, and the power measurements can be executed. It is generally expected that the efficiency will vary based on the test chip sample, while the functionality should be the same for all the samples since the cells are robust with respect to the process spread and are tested extensively.

5.1.2   Power Delivery Measurement

To prove the power savings that stacked circuits can achieve, it is important to obtain knowledge about the power consumption of the individual blocks. This needs various voltage- and current measurements throughout the system, which must be performed in real time. Inserting a real-time on-chip current measurement system would impose considerable

overhead. Due to this reason it has been decided to split the power and ground rails and perform the voltage and current measurements externally, as depicted in **Figure 5-1**.
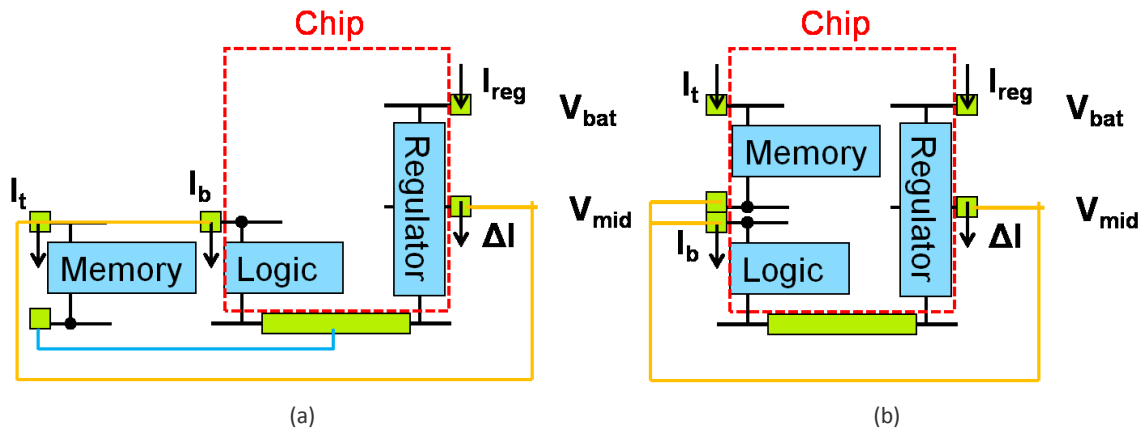
Figure 5-1: Quantities to be measured in flat (a) and stacked (b) mode

As it has been mentioned before, there are two operating modes for the test chip, the flat mode and the stacked mode. In the flat mode, the current of both the top and bottom power domain are coming from the SC regulator, thus the power measurement setup is simple. To get the total efficiency, the product of the output current and voltage of the regulator has to be divided by the product of the input current and voltage. In the stacked mode, on the other hand, all the current measurements must be separated to identify the input and output power and conclude with the regulator's efficiency and the total efficiency.

For the measurements in stacked mode, only three current and two voltage quantities are needed to compute the efficiency, while for flat mode two current and two voltage values. In the measurement setup it is possible to measure four different current values. This way different measurement methods are availablefor the same quantity, thus it is possible to mitigate measurement errors by comparing the two – theoretically equivalent – results. The formulas for the efficiency calculations can be found in **Table 5-3**.

	Regulator Efficiency	Total Efficiency
Stacked	$$\frac{|\Delta I| \cdot V_{mid}}{I_{reg} \cdot V_{bat}} = \frac{|I_t - I_b| \cdot V_{mid}}{I_{reg} \cdot V_{bat}}$$	$$\frac{\min\{I_t, I_b\} \cdot V_{bat} + |\Delta I| \cdot V_{mid}}{\min\{I_t, I_b\} \cdot V_{bat} + I_{reg} \cdot V_{bat}}$$ $$= \frac{\min\{I_t, I_b\} \cdot V_{bat} + |I_t - I_b| \cdot V_{mid}}{\min\{I_t, I_b\} \cdot V_{bat} + I_{reg} \cdot V_{bat}}$$
Flat	$$\frac{|\Delta I| \cdot V_{mid}}{I_{reg} \cdot V_{bat}} = \frac{|I_t + I_b| \cdot V_{mid}}{I_{reg} \cdot V_{bat}}$$	$$\frac{|\Delta I| \cdot V_{mid}}{I_{reg} \cdot V_{bat}} = \frac{|I_t + I_b| \cdot V_{mid}}{I_{reg} \cdot V_{bat}}$$

Table 5-3: Regulator- and Total Efficiency for Stacked and Flat mode

The voltage regulator power density can be calculated by dividing the output power with the area of the active chain of the regulator blocks (chain 1 and/or chain 2). This is 20 times regulator area for flat mode since both chain 1 and 2 are active, and 6 for stacked mode since only chain 2 is active. For flat mode, at about 3.4mA load, this corresponds to 10.5mW/mm^2 power density based on data from **Figure 4-30**, considering that the voltage drop reduces the output voltage to about 1.04V. The efficiency for this load value is 79.5% from **Figure 4-29**. The power density and efficiency numbers will increase if the system is in stacked mode. If we assume the load conditions of **Figure 4-24 (b)** with ISRAM active testbench, the current mismatch of **Equation 5-1** will be valid. The power density becomes 34.9mW/mm^2 and the total efficiency 95%, if stacking is used.

$$\Phi = \frac{\Delta I}{2I} \approx \frac{2.1mW - 1.4mW}{2 \cdot 1.4mW} = 0.25;$$

Equation 5-1: Calculation of mismatch current to stack current ratio

	Power Density**	Total Efficiency
Flat	$$\frac{3.4mA \cdot 1.04V}{20 \cdot 0.0168mm^2} = \mathbf{10.5} \frac{mW}{mm^2}$$	**79.5%**
Stacked*	$$\frac{(1.4 + 2.1)mW}{6 \cdot 0.0168mm^2} = \mathbf{34.9} \frac{mW}{mm^2}$$	$$\frac{1 + \Phi}{1 + \frac{\Phi}{\eta_{reg}}} = \frac{1 + 0.25}{1 + \frac{0.25}{79\%}} = \mathbf{95}\%$$

Table 5-4: Efficiency and power density estimations based on the final design
(Results from **Figure 4-24**, **Figure 4-29** and **Figure 4-30** have been used)
*Assuming ideal 1.1V supply without droop. With 60mV droop power density reduces to 33mW/mm^2
**The total output power for flat and stacked mode is about 3.5mW, with 50µW tolerance

6 Conclusion

In this research project a low-power System-on-Chip has been designed that uses the charge recycling principle by stacking power domains on top of each other. The concept of saving power through voltage stacking in a realistic application has been proven for the first time in this work, to the best of the author's knowledge. The prototype chip is ready for fabrication and the measurement results will follow.

The voltage regulator that provides the supply voltages for the system is no longer the only source of power, but changes its role into supporting the direct power connection that has become the main power source. The result is decreased power and area overhead in the power delivery system. This technique can be combined with other low-power schemes which usually act on the power load, rather than the power delivery.

One of the main conclusions that can be drawn is that the power consumption matching of the stacked voltage domains will impose a challenge for future stacked systems. The ad hoc partitioning method followed here must be extended into a systematic approach where the power matching is actively optimized during the design steps. Similarly, the timing-, power- and area penalty of level shifting circuitry should be overcome by minimizing the domain crossings in parallel to the power matching. Finally, the voltage regulator has to be carefully designed to have high efficiency at the expected load conditions.

In this work only two stacked domains were implemented. To reach the commonly used battery voltages, three domains are required, which is treated here as future reference. This configuration will increase the complexity of the overall system, the level shifters and the voltage regulator, so it has to be issued in later works.

REFERENCES

[1]   Petritz, Richard, "View from the top: Semiconductors and the second industrial revolution: The first was essentially an oil-based revolution, while today's information revolution uses the fundamental technology of semiconductors," *Potentials, IEEE* , vol.1, no.Fall, pp.38,39, Third Quarter 1982

[2]   Kilby, J. S. "*Miniaturized electronic circuits* [US Patent No. 3,138, 743]." *Solid-State Circuits Society Newsletter, IEEE* 12.2 (2007): 44-54.

[3]   Mack, C.A, "Fifty Years of Moore's Law," *Semiconductor Manufacturing, IEEE Transactions on* , vol.24, no.2, pp.202,207, May 2011

[4]   Dennard, R.H.; Gaensslen, F.H.; Yu, Hwa-Nien; LEO RIDEOVT, V.; Bassous, Ernest; Leblanc, Andre R., "Design of ion-implanted MOSFET's with very small physical dimensions," *Solid-State Circuits Society Newsletter, IEEE* , vol.12, no.1, pp.38,50, Winter 2007

[5]   Kapoor, A; Groot, C.; Pique, G.V.; Fatemi, H.; Echeverri, J.; Sevat, L.; Vertregt, M.; Meijer, M.; Sharma, V.; Yu Pu; de Gyvez, J.P., "Digital Systems Power Management for High Performance Mixed Signal Platforms," *Circuits and Systems I: Regular Papers, IEEE Transactions on* , vol.61, no.4, pp.961,975, April 2014

[6]   Bohr, M., "A 30 Year Retrospective on Dennard's MOSFET Scaling Paper," *Solid-State Circuits Society Newsletter, IEEE* , vol.12, no.1, pp.11,13, Winter 2007

[7]   Yo-Sheng Lin; Chung-Cheng Wu; Chih-Sheng Chang; Rong-Ping Yang; Wei-Ming Chen; Jhon-Jhy Liaw; Diaz, C.H., "Leakage scaling in deep submicron CMOS for SoC," *Electron Devices, IEEE Transactions on* , vol.49, no.6, pp.1034,1041, Jun 2002

[8]   Deepaksubramanyan, B.S.; Nunez, A, "Analysis of subthreshold leakage reduction in CMOS digital circuits," *Circuits and Systems, 2007. MWSCAS 2007. 50th Midwest Symposium on* , vol., no., pp.1400,1404, 5-8 Aug. 2007

[9]   Lahiri, K.; Raghunathan, A; Dey, S.; Panigrahi, D., "Battery-driven system design: a new frontier in low power design," *Design Automation Conference, 2002. Proceedings of ASP-DAC 2002. 7th Asia and South Pacific and the 15th International Conference on VLSI Design. Proceedings.* , vol., no., pp.261,267, 2002

[10]   Srivastav, M.; Rao, S. S S P; Bhatnagar, H., "Power reduction technique using multi-Vt libraries," *System-on-Chip for Real-Time Applications, 2005. Proceedings. Fifth International Workshop on* , vol., no., pp.363,367, 20-24 July 2005

[11]   Heungjun Jeon; Yong-Bin Kim; Minsu Choi, "Standby Leakage Power Reduction Technique for Nanoscale CMOS VLSI Systems," *Instrumentation and Measurement, IEEE Transactions on* , vol.59, no.5, pp.1127,1133, May 2010

[12] Lackey, D.E.; Zuchowski, P.S.; Bednar, T.R.; Stout, D.W.; Gould, S.W.; Cohn, J.M., "Managing power and performance for system-on-chip designs using Voltage Islands," *Computer Aided Design, 2002. ICCAD 2002. IEEE/ACM International Conference on* , vol., no., pp.195,202, 10-14 Nov. 2002

[13] Semeraro, G.; Magklis, G.; Balasubramonian, R.; Albonesi, D.H.; Dwarkadas, S.; Scott, M.L., "Energy-efficient processor design using multiple clock domains with dynamic voltage and frequency scaling," *High-Performance Computer Architecture, 2002. Proceedings. Eighth International Symposium on* , vol., no., pp.29,40, 2-6 Feb. 2002

[14] Pakbaznia, E.; Fallah, F.; Pedram, M., "Charge recycling in MTCMOS circuits: concept and analysis," *Design Automation Conference, 2006 43rd ACM/IEEE* , vol., no., pp.97,102, 0-0 0

[15] Keejong Kim; Mahmoodi, H.; Roy, K., "A Low-Power SRAM Using Bit-Line Charge-Recycling," *Solid-State Circuits, IEEE Journal of* , vol.43, no.2, pp.446,459, Feb. 2008

[16] Byung-Do Yang; Lee-Sup Kim, "A low-power ROM using charge recycling and charge sharing techniques," *Solid-State Circuits, IEEE Journal of* , vol.38, no.4, pp.641,653, Apr 2003

[17] Ginsburg, B.P.; Chandrakasan, AP., "An energy-efficient charge recycling approach for a SAR converter with capacitive DAC,"*Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on* , vol., no., pp.184,187 Vol. 1, 23-26 May 2005

[18] Abbasian, A; Rasouli, S.H.; Afzali-Kusha, A; Nourani, M., "No-race charge recycling complementary pass transistor logic (NCRCPL) for low power applications," *Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on* , vol.5, no., pp.V-289,V-292 vol.5, 25-28 May 2003

[19] Yong Liu; Ping-Hsuan Hsieh; Seongwon Kim; Jae-sun Seo; Montoye, R.; Chang, L.; Tierno, J.; Friedman, D., "A 0.1pJ/b 5-to-10Gb/s charge-recycling stacked low-power I/O for on-chip signaling in 45nm CMOS SOI," *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2013 IEEE International* , vol., no., pp.400,401, 17-21 Feb. 2013

[20] Taejoong Song; Kim, S.; Kyutae Lim; Laskar, J., "Power analysis of asynchronous design using charge recycling and push-pull level converter," *Circuits and Systems (MWSCAS), 2010 53rd IEEE International Midwest Symposium on* , vol., no., pp.1270,1273, 1-4 Aug. 2010

[21] Jeongwon Cha; Taejoong Song; Changhyuk Cho; Minsik Ahn; Chang-Ho Lee; Laskar, J., "A Low-Power CMOS Antenna-Switch Driver Using Shared-Charge Recycling Charge Pump," *Microwave Theory and Techniques, IEEE Transactions on* , vol.58, no.12, pp.3626,3633, Dec. 2010

[22] Jie Gu; Kim, C.H., "Multi-story power delivery for supply noise reduction and low voltage operation," *Low Power Electronics and Design, 2005. ISLPED '05. Proceedings of the 2005 International Symposium on* , vol., no., pp.192,197, 8-10 Aug. 2005

[23] Salem, L.G.; Mercier, P.P., "4.6 An 85%-efficiency fully integrated 15-ratio recursive switched-capacitor DC-DC converter with 0.1-to-2.2V output voltage range," *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014 IEEE International* , vol., no., pp.88,89, 9-13 Feb. 2014

[24] Rajapandian, S.; Shepard, S, "Charge-recycling voltage domains for energy-efficient low-voltage operation of digital CMOS circuits," U.S. Patent 7329968 B2, Feb. 12, 2008.

[25] Pique, G.V., "A 41-phase switched-capacitor power converter with 3.8mV output ripple and 81% efficiency in baseline 90nm CMOS," *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2012 IEEE International* , vol., no., pp.98,100, 19-23 Feb. 2012

[26] Lu Tan; Neng Wang, "Future internet: The Internet of Things," *Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on* , vol.5, no., pp.V5-376,V5-380, 20-22 Aug. 2010

[27] Hao Gao; Matters-Kammerer, M.K.; Harpe, P.; Milosevic, D.; van Roermund, A; Linnartz, J.-P.; Baltus, P.G.M., "A 60-GHz energy harvesting module with on-chip antenna and switch for co-integration with ULP radios in 65-nm CMOS with fully wireless mm-wave power transfer measurement," *Circuits and Systems (ISCAS), 2014 IEEE International Symposium on* , vol., no., pp.1640,1643, 1-5 June 2014

[28] Yao Yingbiao; Zhang Jianwu; Zhao Danying, "Survey on microprocessor architecture and development trends," *Communication Technology, 2008. ICCT 2008. 11th IEEE International Conference on* , vol., no., pp.297,300, 10-12 Nov. 2008

[29] Blem, E.; Menon, J.; Sankaralingam, K., "Power struggles: Revisiting the RISC vs. CISC debate on contemporary ARM and x86 architectures," *High Performance Computer Architecture (HPCA2013), 2013 IEEE 19th International Symposium on* , vol., no., pp.1,12, 23-27 Feb. 2013

[30] Tanenbaum, Andrew S., Todd Austin, and B. R. Chandavarkar. *Structured computer organization*. Pearson, 2013.

[31] ARM Cortex-M official website. http://www.arm.com/products/processors/cortex-m/

[32] NXP LPC800. http://www.nxp.com/products/microcontrollers/cortex_m0_m0/lpc800/

[33] Frank, D.J.; Dennard, R.H.; Nowak, E.; Solomon, P.M.; Yuan Taur; Hen-Sum Philip Wong, "Device scaling limits of Si MOSFETs and their application dependencies," *Proceedings of the IEEE* , vol.89, no.3, pp.259,288, Mar 2001

[34] Taejoong Song; Woojin Rim; Jonghoon Jung; Giyong Yang; Jaeho Park; Sunghyun Park; Kang-Hyun Baek; Sanghoon Baek; Sang-Kyu Oh; Jinsuk Jung; Sungbong Kim; Gyuhong Kim; Jintae Kim; Youngkeun Lee; Kee Sup Kim; Sang-Pil Sim; Jong Shik Yoon; Kyu-Myung Choi, "13.2 A 14nm FinFET 128Mb 6T SRAM with VMIN-enhancement techniques for low-power applications," *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014 IEEE International* , vol., no., pp.232,233, 9-13 Feb. 2014

[35] ITRS 2011 (International Technology Roadmap for Semiconductors) [Online]. Available: http://www.itrs.net/

[36] Roy, K.; Mukhopadhyay, S.; Mahmoodi-Meimand, H., "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits," *Proceedings of the IEEE* , vol.91, no.2, pp.305,327, Feb 2003

[37] Paul, S.; Schlaffer, AM.; Nossek, J.A, "Optimal charging of capacitors," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on* , vol.47, no.7, pp.1009,1016, Jul 2000

[38] Chandrakasan, AP.; Sheng, S.; Brodersen, R.W., "Low-power CMOS digital design," *Solid-State Circuits, IEEE Journal of* , vol.27, no.4, pp.473,484, Apr 1992

[39] Goebel, Stefan, "An Overview of Battery Development placed in a historical Context and future Aspects," *Telecommunication - Energy Special Conference (TELESCON), 2009 4th International Conference on* , vol., no., pp.1,4, 10-13 May 2009

[40] Sousa, R.; Ribeiro, J.F.; Sousa, J.A; Goncalves, L.M.; Correia, J.H., "All-solid-state batteries: An overview for bio applications,"*Bioengineering (ENBENG), 2013 IEEE 3rd Portuguese Meeting in* , vol., no., pp.1,4, 20-23 Feb. 2013

[41] Electropaedia, Online: http://www.mpoweruk.com/specifications/comparisons.pdf

[42] Piqué, G. V.; Alarcón, E., "CMOS Integrated Switching Power Converters," Springer, 2011.

[43] Piqué, G. Villar; Bergveld, H. J., "State-of-the-Art of Integrated Switching Power Converters," Analog Circuits Design, pp. 259-281, Springer, 2012.

[44] Chin-Long Wey; Ping-Chang Jui, "A unitized charging and discharging smart battery management system," *Connected Vehicles and Expo (ICCVE), 2013 International Conference on* , vol., no., pp.903,909, 2-6 Dec. 2013

[45] Rao, R.; Vrudhula, S.; Rakhmatov, D.N., "Battery modeling for energy aware system design," *Computer* , vol.36, no.12, pp.77,87, Dec. 2003

[46] Youngjin Cho; Younghyun Kim; Yongsoo Joo; Kyungsoo Lee; Naehyuck Chang, "Simultaneous optimization of battery-aware voltage regulator scheduling with dynamic voltage and frequency scaling," *Low Power Electronics and Design (ISLPED), 2008 ACM/IEEE International Symposium on* , vol., no., pp.309,314, 11-13 Aug. 2008

[47] Wens, M.; Steyaert, M., "A fully-integrated 130nm CMOS DC-DC step-down converter, regulated by a constant on/off-time control system," *Solid-State Circuits Conference, 2008. ESSCIRC 2008. 34th European* , vol., no., pp.62,65, 15-19 Sept. 2008

[48] Patounakis, G.; Li, Y.W.; Shepard, Kenneth L., "A fully integrated on-chip DC-DC conversion and power management system," *Solid-State Circuits, IEEE Journal of* , vol.39, no.3, pp.443,451, March 2004

[49] Hazucha, P.; Schrom, G.; Hahn, J.; Bloechel, B.A; Hack, P.; Dermer, G.E.; Narendra, S.; Gardner, D.; Karnik, T.; De, V.; Borkar, S., "A 233-MHz 80%-87% efficient four-phase DC-DC converter utilizing air-core inductors on package," Solid-State Circuits, IEEE Journal of , vol.40, no.4, pp.838,845, April 2005

[50] Onizuka, K.; Inagaki, K.; Kawaguchi, H.; Takamiya, M.; Sakurai, T., "Stacked-Chip Implementation of On-Chip Buck Converter for Distributed Power Supply System in SiPs," IEEE Journal of Solid-State Circuits, vol.42, no.11, pp.2404,2410, Nov. 2007

[51] Maksimovic, D.; Dhar, S., "Switched-capacitor DC-DC converters for low-power on-chip applications," *Power Electronics Specialists Conference, 1999. PESC 99. 30th Annual IEEE* , vol.1, no., pp.54,59 vol.1, Aug 1999

[52] Seeman, M.D.; Sanders, S.R., "Analysis and Optimization of Switched-Capacitor DC-DC Converters," *Computers in Power Electronics, 2006. COMPEL '06. IEEE Workshops on* , vol., no., pp.216,224, 16-19 July 2006

[53] Seeman, M.D.; Sanders, S.R.; Rabaey, J.M., "An Ultra-Low-Power Power Management IC for Wireless Sensor Nodes," *Custom Integrated Circuits Conference, 2007. CICC '07. IEEE* , vol., no., pp.567,570, 16-19 Sept. 2007

[54] Seeman, M.D., "*A Design Methodology for Switched-Capacitor DC-DC Converters,*" Ph.D. dissertation, University of California at Berkeley, 21 May 2009

[55] Andersen, T.M.; Krismer, F.; Kolar, J.W.; Toifl, T.; Menolfi, C.; Kull, L.; Morf, T.; Kossel, M.; Brandli, M.; Buchmann, P.; Francese, P.A, "A 4.6W/mm^2 power density 86% efficiency on-chip switched capacitor DC-DC converter in 32 nm SOI CMOS,"*Applied Power Electronics Conference and Exposition (APEC), 2013 Twenty-Eighth Annual IEEE* , vol., no., pp.692,699, 17-21 March 2013

[56] Liu, C. *Voltage Regulation of CMOS Stacked Digital Circuits*. Technical Report from Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands, August 30, 2010.

[57] Rajapandian, S.; Zheng Xu; Shepard, Kenneth L., "Implicit DC-DC downconversion through charge-recycling," *Solid-State Circuits, IEEE Journal of* , vol.40, no.4, pp.846,852, April 2005

[58] Meyvaert, H.; Van Breussegem, T.; Steyaert, M., "A 1.65W fully integrated 90nm Bulk CMOS Intrinsic Charge Recycling capacitive DC-DC converter: Design & techniques for high power density," *Energy Conversion Congress and Exposition (ECCE), 2011 IEEE* , vol., no., pp.3234,3241, 17-22 Sept. 2011

[59] Chang, L.; Montoye, R.K.; Ji, B.L.; Weger, AJ.; Stawiasz, K.G.; Dennard, R.H., "A fully-integrated switched-capacitor 2:1 voltage converter with regulation capability and 90% efficiency at 2.3A/mm^2," *VLSI Circuits (VLSIC), 2010 IEEE Symposium on* , vol., no., pp.55,56, 16-18 June 2010

[60] Endoh, T.; Sunaga, K.; Sakuraba, Hiroshi; Masuoka, F., "An on-chip 96.5% current efficiency CMOS linear regulator using a flexible control technique of output current," *Solid-State Circuits, IEEE Journal of* , vol.36, no.1, pp.34,39, Jan 2001

[61] Hazucha, P.; Sung Tae Moon; Schrom, G.; Paillet, F.; Gardner, D.; Rajapandian, S.; Karnik, T., "High Voltage Tolerant Linear Regulator With Fast Digital Control for Biasing of Integrated DC-DC Converters," *Solid-State Circuits, IEEE Journal of* , vol.42, no.1, pp.66,73, Jan. 2007

[62] Wang, Yikai; Ma, Dongsheng, "Ultra-fast on-chip load-current adaptive linear regulator for switch mode power supply load transient enhancement," *Applied Power Electronics Conference and Exposition (APEC), 2013 Twenty-Eighth Annual IEEE* , vol., no., pp.1366,1369, 17-21 March 2013

[63] Lackey, D.E.; Zuchowski, P.S.; Bednar, T.R.; Stout, D.W.; Gould, S.W.; Cohn, J.M., "Managing power and performance for system-on-chip designs using Voltage Islands," *Computer Aided Design, 2002. ICCAD 2002. IEEE/ACM International Conference on* , vol., no., pp.195,202, 10-14 Nov. 2002

[64] Shenoy, P. S.; Krein, P. T., "Differential Power Processing for DC Systems," *Power Electronics, IEEE Transactions on* , vol.28, no.4, pp.1795,1806, April 2013

[65] Mazumdar, K.; Stan, M.R., "Charge recycling on-chip DC-DC conversion for near-threshold operation," *Subthreshold Microelectronics Conference (SubVT), 2012 IEEE* , vol., no., pp.1,3, 9-10 Oct. 2012

[66] Yong Liu; Ping-Hsuan Hsieh; Seongwon Kim; Jae-sun Seo; Montoye, R.; Chang, L.; Tierno, J.; Friedman, D., "A 0.1pJ/b 5-to-10Gb/s charge-recycling stacked low-power I/O for on-chip signaling in 45nm CMOS SOI," *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2013 IEEE International* , vol., no., pp.400,401, 17-21 Feb. 2013

[67] Rajapandian, S.; Shepard, Kenneth L.; Hazucha, P.; Karnik, T., "High-voltage power delivery through charge recycling," *Solid-State Circuits, IEEE Journal of* , vol.41, no.6, pp.1400,1410, June 2006

[68] Junhua Liu; Le Ye; Zhixin Deng; Jinshu Zhao; Huailin Liao, "A 1.8V to 10V CMOS level shifter for RFID transponders," *Solid-State and Integrated Circuit Technology (ICSICT), 2010 10th IEEE International Conference on* , vol., no., pp.491,493, 1-4 Nov. 2010

[69] Moghe, Y.; Lehmann, T.; Piessens, T., "Nanosecond Delay Floating High Voltage Level Shifters in a 0.35 μ m HV-CMOS Technology," *Solid-State Circuits, IEEE Journal of* , vol.46, no.2, pp.485,497, Feb. 2011

[70] Serneels, B.; Steyaert, M.; Dehaene, W., "A High speed, Low Voltage to High Voltage Level Shifter in Standard 1.2V 0.13μm CMOS," *Electronics, Circuits and Systems, 2006. ICECS '06. 13th IEEE International Conference on* , vol., no., pp.668,671, 10-13 Dec. 2006

[71] Ueda, K.; Morishita, F.; Okura, S.; Okamura, L.; Yoshihara, T.; Arimoto, K., "Low-Power On-Chip Charge-Recycling DC-DC Conversion Circuit and System," *Solid-State Circuits, IEEE Journal of* , vol.48, no.11, pp.2608,2617, Nov. 2013

[72] Cabe, AC.; Zhenyu Qi; Stan, M.R., "Stacking SRAM banks for ultra low power standby mode operation," *Design Automation Conference (DAC), 2010 47th ACM/IEEE* , vol., no., pp.699,704, 13-18 June 2010

[73] Mohammadi, B.; Rodrigues, J.N., "A 65 nm single stage 28 fJ/cycle 0.12 to 1.2V level-shifter," *Circuits and Systems (ISCAS), 2014 IEEE International Symposium on* , vol., no., pp.990,993, 1-5 June 2014

[74] Alon, E.; Horowitz, M., "Integrated Regulation for Energy-Efficient Digital Circuits," *Solid-State Circuits, IEEE Journal of* , vol.43, no.8, pp.1795,1807, Aug. 2008

[75] Kurd, N.; Chowdhury, M.; Burton, E.; Thomas, T.P.; Mozak, C.; Boswell, B.; Lal, M.; Deval, A; Douglas, J.; Elassal, M.; Nalamalpu, A; Wilson, T.M.; Merten, M.; Chennupaty, S.; Gomes, W.; Kumar, R., "5.9 Haswell: A family of IA 22nm processors," *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014 IEEE International* , vol., no., pp.112,113, 9-13 Feb. 2014

[76] M. R. Garey and D. S. Johnson. A guide to the theory of NP-completeness. Freeman, San Francisco, 1979.

[77] Andreev, Konstantin and Räcke, Harald, (2004). "Balanced Graph Partitioning". Proceedings of the sixteenth annual ACM symposium on Parallelism in algorithms and architectures (Barcelona, Spain): 120–124.

[78] NXP LPC81XM Datasheet http://www.nxp.com/documents/data_sheet/LPC81XM.pdf

[79] CoreMark, EEMBC http://www.eembc.org/coremark/

[80] ARM Cortex-M0+ http://www.arm.com/products/processors/cortex-m/cortex-m0plus.php

[81] ARM AMBA http://www.arm.com/products/system-ip/amba