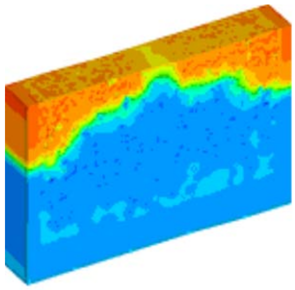
11. Publishable summary: “Reliable and Variability tolerant System-on-a-chip Design in More-Moore Technologies”

Project Facts:

FP7 Project :	European Community funded
Coordination :	IMEC
Website :	www.fp7-reality.eu
Duration :	30 Months
Effort :	382 person-months
Industry :	ARM (UK), ST Microelectronics (Italy)
Start date :	1st January 2008
University :	Glasgow (UK), Bologna (Italy), Leuven (Belgium)
Research Centre :	IMEC (Belgium)

Scope:

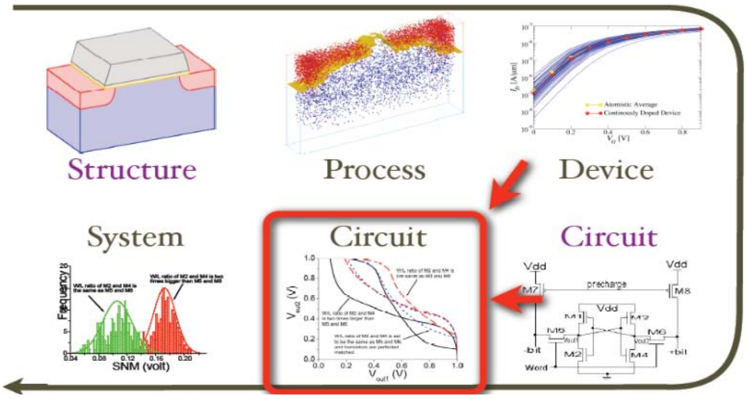
- Scaling beyond the 32 nm technology
- Tackle the increased variability and changing performance of devices from device unto system level.



Random discrete dopants in a 35 nm MOSFET from the present 90 nm technology node.

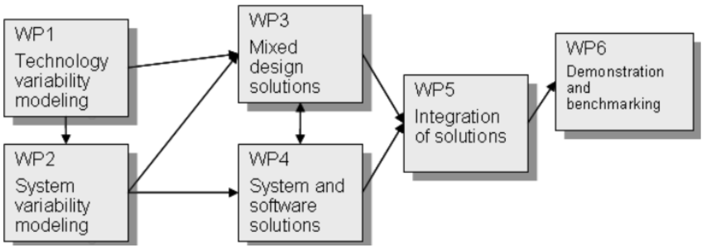
Challenges:

- Increased static variability and static fault rates of devices and interconnects.
- Increased time-dependent dynamic variability and dynamic fault rates.
- Build reliable systems out of unreliable technology while maintaining design productivity.
- Deploy design techniques that allow technology scalable energy efficient SoC systems while guaranteeing real-time performance constraints.



Proposed solution:

- System analysis of performance, power, yield and reliability of manufactured instances across a wide spectrum of operating conditions.
- Generally applicable solution techniques to mitigate the impact of reliability issues of integrated circuits, at component, circuit, and architecture and system design.


WP1
Technology variability modeling
WP2
System variability modeling
WP3
Mixed design solutions
WP4
System and software solutions
WP5
Integration of solutions
WP6
Demonstration and benchmarking

WP1: Device variability and Reliability Models (WP leader: UoG)

Having completed the simulation of variability in the 32nm devices including random discrete dopants, line edge roughness and metal gate granularity, this was used as a starting point for investigating reliability issues due to the trapping of electrons and/or holes in defect states in the gate stack during circuit operation. To each fresh device were added additional fixed charges, randomly within the channel, based on the trap sheet density, and a full 3D simulation was performed. Distributions of threshold voltages and V_T shifts were obtained.

Using the strategy for statistical compact model extraction developed in the project for use with PSP compact models, a statistical library of compact models was created based on the 32nm technology. This can be used for Monte Carlo statistical circuit simulation. An approach based on Principal Component Analysis of the data extracted from the statistical device simulations, which allows correlations between extracted parameters to be maintained, has been developed for on-the-fly calculation of statistical model parameters.

Using the statistical compact models developed for the 45nm technology, including different levels of NBTI/PBTI degradation, a detailed investigation of the effect of statistical variability and reliability on the performance of an SRAM cell was performed.

ARM has used the output of the cell level characterisation from WP5 and WP6 to analyse the difference between the statistical models from the foundry and from UoG available in the project through the variability injectors generated by IMEC.

Deliverable D1.2, due at T0+27, was successfully delivered on time.

WP2: System and circuit characterization and sensitivity analysis (WP leader: IMEC)

The goal of this WP is to develop advanced methodologies and techniques for statistical analysis. The intention is to read the output of transistor level variability, as provided by WP 1, and to propagate this information all the way from the device level to the product level. The WP also targets developing and fully characterizing a limited standard cell library (50-100 cells) for synthesis and analysis based on restricted design rules for use in WP2, WP3, WP4, and WP5.

Commercial EDA solutions (e.g., fast circuit simulators, SSTA tools, power analysis tools, etc) were reused in the flow wherever possible in combination with Monte Carlo-based simulation techniques in order to guarantee the compatibility with existing electronic design simulation/verification tools and easy adoption by engineers trained to these tools. This was not always possible. Even commercial tools show bugs or are simply inappropriate for particular purposes. This can occur in any area, in particular in the areas of electrical simulation, characterization, statistical characterization, and statistical static timing analysis. Depending on case, new methods were developed and implemented, or commercial vendors were asked to revise their products.

New methods were identified in the areas of standard cell characterization (ST’s hybrid flow, imec’s VAM, UoG’s characterizer), memory characterization (ARM’s extreme value theory based memory margining application, imec’s MemoryVAM), statistical timing analysis (ST’s hybrid flow on the digital level, imec’s VAM), as well as system level analysis (VAM).

Also considered in this WP is the strategic aspect of the standardization of the interfaces between different abstraction levels to enable the propagation of variability specific information throughout the design flow. To “lubricate” the flow developed here and applied and integrated in all other work packages, we put in place a standard electronic Information Format (IF) that keeps statistical information and exists parallel to the classical top-down and bottom-up design flows.

All deliverables (D2.3 and D2.4) have been submitted on time.

WP3: Mixed mode countermeasures (WP leader: KUL)

In work package 3, an advanced reliability simulation technique including process variability in the simulation flow was developed. In order to get a computationally efficient solution, a Response Surface Model (RSM) based simulation technique is proposed. In this approach the vast number of computationally intensive simulations needed to perform a Monte-Carlo analysis, is replaced by the evaluation of an analytical model of the circuit. The focus of this work is not to obtain a highly accurate prediction of yield, but to analyze the spatial and temporal reliability of a circuit in a reasonably short time frame. When compared to direct Monte-Carlo yield simulation, examples indicate this method to have a simulation speedup ranging from 1 to 3 orders of magnitude, without sacrifying accuracy. Additionally, weak spot analysis allows to improve the design or to reduce design margins and to gain extra performance.

Also, with the availability of the 32nm compact models, an extra iteration on the KUL SRAM memory has been done. The reliability simulation framework has been applied to one cell of the KUL memory (full memory analysis is performed by the MemoryVAM). A huge amount of minimal sized cells, of which the functionality of every single one is critical demands big design margins. The most vulnerable parts of a memory were identified to be the cells, the sense amplifiers and the timing. Cell stability can be increased by reducing the cell load capacitance. For this reason, divided bit lines are introduced. Additional advantages include possible speed and energy gains. The impact of NBTI, on the performance of one cell has been simulated as a function of cell voltage. Other improvements include sharing sense amplifiers and usage of a configurable timing circuit to reduce variability effects. Some estimations of the effect of BEOL-variability have been done. At this moment, these variations seem to have a limited effect due to the averaging out for the big wires and the small impact of small wires. A fully functional 1KB SRAM memory with a word length of 32bit was build. The obtained speed was 1GHz. Total write energy was 0.69pJ, read energy was a bit higher: 1.07pJ.

WP4: System level countermeasures (WP leader: UNIBO)

The activity of UNIBO in year 3 has been devoted to the porting and optimization of system level policies to the target platform and validation benchmark. We evaluated the capability of the techniques to compensate the variability impact on application performance. The impact of variability on multicore multimedia platform makes hard to get the a certain QoS from the running software because of the speed variations across the cores, which causes a sub-optimal exploitation of the platform parallelism. Moreover, QoS and power consumption vary from platform to platform.

Thanks to a smart allocation of the workload, it is possible to compensate this impact, obtaining an improvement of the QoS and energy consumption for a given platform as well as an increase of the predictability across many variability affected platforms. System level compensation allows to better guarantee the QoS with respect to a non compensated one. Indeed, a smart workload allocation strategy applied to a variability affected platform, helps reducing the deadline misses and reduces the energy consumption.

During the reporting period, we optimized the workload allocation policies so that they can be applied on-line on a frame-by-frame basis. The optimization was targeted to the reduction of the execution time of the policies, to provide the wanted QoS level with minimum energy, independently form the variability impact on the platform. The techniques have been ported to a relevant industrial case study of a multicore multimedia platform, with a single voltage domain, multiple frequency domain. Since variability causes the cores to be characterized by different speeds, the designer can tune each core to its maximum supported frequency to improve performance. However, now he/she has to handle a heterogeneous platform. This may cause inefficiencies when allocating multitask applications, because the speed heterogeneity lead to a unbalanced allocation. This effect can be compensated by a smart allocation, thanks to which the designer can exploit the advantage of having clocked the system in a heterogeneous way without paying the price of the unbalancing and thus overall leading to a better QoS with respect to a non-compensated platform.

WP5: Design flow, integration, proof of concept (WP leader: ARM)

No summary available. These work package activities have been completed during the earlier reporting period.

WP6: Validation and assessment of results (WP leader: ST)

This WP reported on the analysis, validation and industrial impact for all of the REALITY outcomes.
The validation and benchmarking revolved around a two pronged strategy based on the leverage of an industrially proven embedded microprocessor design provided by ARM, the ARM926, and an advanced

multiprocessor based multimedia accelerator platform from STM, the xSTream platform. Much of the project conclusions are gathered in the reports produced by WP6.

For the first time, full scale 3D simulation of statistical variability associated with metal gate granularity and the corresponding metal work function variations has been carried out to clarify the magnitude of statistical variability in 32 nm CMOS transistors with high-k/metal gate stack. In addition, technology has been developed to simulate the statistical aspects of reliability associated with NBTI/PBTI.

Also for a first time, a full statistical characterization of an ARM core has been achieved. A correlation between the timing, leakage and dynamic power has been demonstrated on local (within die) and non-local (above die) variations. The traditional corner analysis could be benchmarked with innovative statistical analysis techniques. Using the ARM core as driver, REALITY has confirmed that the SRAM components are responsible for more than the half of the variations on critical path timing. Much focus has been placed by both state-of-the-art and EDA vendors on the logic while the variability challenges remain in the memories themselves. For that purpose REALITY has been also first in deploying a holistic statistical characterization flow including SRAM analysis variations at and their evolution over time.

REALITY has also for first time evaluated the impact of process variation in SW level metrics showing process variability is not only a concern for HW but for SW as well. It has concluded that variability affecting multi-core multimedia platforms makes it hard to guarantee a certain QoS from the running application's functionality. The speed variations across the cores cause sub-optimal and platform-dependent parallelism. REALITY has developed an approach to compensate this by using a smart allocation of the workload at run-time, hence also at the SW level, and obtaining an improvement of QoS by 20% and energy consumption by 15% while obtaining better platform predictability.

For that purpose, different circuit design techniques for system adaptation have been investigated, among them Adaptive Body Biasing (ABB). REALITY has shown that even though the possible compensation range in speed up due to ABB is significantly reduced compared to the previous node, it remains still available at 32nm. The technique has been validated on the ALU design of the ARM core using specially characterized commercially available libraries.

The xSTream platform was used as a test case for a system level driver and benchmarking environment to develop and validate multitask sw allocation and scheduling policies in the presence of different kinds of variability 'control and measuring knobs'.

To enable the project to obtain trends and projection on real industrial application retrofitted with such sw control stacks, the system platform needed to be an executable model capable of supporting real-time simulation of applications. The system level validation and benchmarking activities carried out in WP6 have resulted in hundreds of simulation trials were a number of parameters were changed by selecting statistical relevant distributions produced by the IP block variability analysis flows. Results were analyzed in terms of multiple objective metrics keeping power consumption, yield, area and cost into account. The final impact analysis carried out in WP6 did not attempt to exhaustively provide a coverage of all of the advantages and drawbacks; but rather identified and highlighted the concrete impact, for example in terms of adoption of the variability characterization flow, variability aware design techniques, sw and hw countermeasures, when applied to pragmatic product like development conditions measured by a choice of objective industrially relevant metrics applied to pragmatic product like development conditions.