Early failure prediction by using in-situ monitors: Implementation and application results

A. Benhassain[1,2], F. Cacho[1], V. Huard[1], L. Anghel[2]

[1]STMicroelectronics, Technology R&D, Crolles, France
Phone: + 33(0)438922536, e-mail : sidi-ahmed.benhassain@st.com
[2]TIMA 46, avenue Félix Viallet, 38031 Grenoble, France

***Abstract* – In-situ monitor is a promising strategy to measure timing slacks and to provide pre-error warning prior to any timing violation. In this work, we demonstrate that the usage of in-situ monitors with a feedback loop of voltage regulation is suitable for process and temperature compensation.**
Index Terms — in-situ timing monitors, CMOS reliability, timing margin..

I. INTRODUCTION

With CMOS technology scaling, it becomes more and more difficult to guarantee circuit functionality for all process, voltage, and temperature (PVT) corners. Moreover, circuit wearout degradation lead to additional temporal variation. It results an increase of design margin for reliable systems [1].Adding pessimistic timing margin to guarantee all operating conditions under worse case conditions is no more acceptable due to the huge impact on design costs.

One can report two categories of ageing monitoring techniques. Firstly, we can define standalone sensors utilizing various configurations of ring oscillators [2] and delay chain. Replica paths [3] are a solution to mimic the timing behavior of the original path in combinatory logic. Second, in-situ delay monitors can directly measure the delay degradation of a specific path within the target circuit, this approach is very promising to provide reliable timing information [4]. Delay monitors such as "Razor I" [5] and "Razor II" [6] detect timing errors in actual paths. A local microrollback execution procedure ensures error correction. However, these methods need huge hardware architecture for error recovery. The Adaptive Voltage Scaling (AVS) approach in [7, 8] proposes error correction by using in-situ monitors able to detect timing error and global system action following the error detection.

Another approach consists in detecting timing pre-error instead of timing error by detecting critical transitions [8]. In this case, the in-situ delay monitors can be used as reliability technique to provide alert prior setup violation. This technique is also further combined with global system actions such as AVS or DVFS.

In this paper, an innovative insertion flow of monitor is presented. Two solutions of ISM are discussed and compared. The first one is built with standard cells available in the technology design platform library, named here built-in flow ISM. The second one uses a dedicated custom design, named cell-based ISM. In section III, some benchmarks of strategy of insertion are reported and discussed. Finally, several applications of ISM usage for compensation are presented.

II. ISM INSERTION FLOW

The advantage of ISM located inside digital block is the capability to accurately capture all sources of local physical, environmental and temporal variations. ISMs under investigation are presented in the Fig. 1. The basic idea is to delay the data of a critical path arriving at D in the shadow FF, and to compare it with the regular FF. When Flag signal rises, it means that a violation of the setup time has occurred in the shadow FF and the remaining slack of the data path is close to the timing of the delay element, as defined in the schematic. In this work 3 time windows (TW1=60ps, TW2=100ps, TW3=130ps) have been evaluated. The schematic can be carried out in two different ways: semi-custom (flow-based ISM) or full custom designs (cell-based ISM). In the first one, all schematic elements are issued from the standard cell design platform. Placement and connectivity is performed with scripting during the flow execution. The second one is a new cell dedicated this usage. For that approach, all CAD views of the new cell (functional, physical, timing, etc) need to be developed to be compliant with standard digital flow.

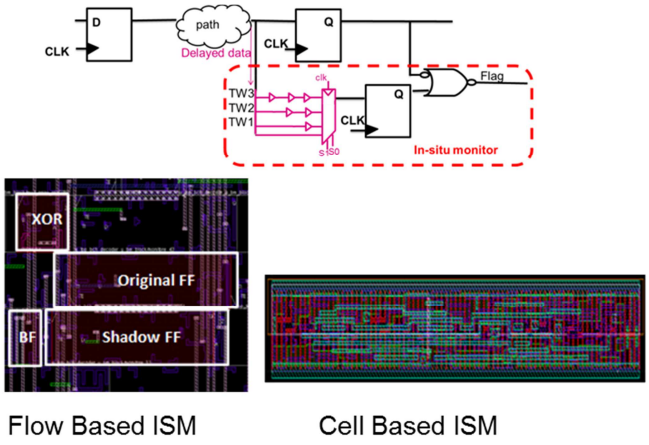

Fig. 1. Schematic and layout of in-situ monitor under investigations. Data arriving at Q is delayed in shadow FF and compared to the regular one. Flow-based ISM is composed of standard cells available in the design platform. Cell-based ISM is a fully customized design.

In addition to the choice of the monitor, the insertion flow is of crucial importance. In the objective of developing quantitative results, an industrial framework compliant with STMicroelectronics digital flow design is used. Obviously, the methodology is portable to any other standard or in-house digital flow. The generic approach is illustrated in the Fig. 2. The classical Front-end steps are executed with synthesis and floorplaning. At the end, a gate netlist is provided as input to placement and route tool. After placement and pre clock tree synthesis (CTS), a timing analysis (TA) is performed. For setup functional corner, a decision is made to insert monitor (FF cell sweep for cell-based ISM) and to regenerate connectivity on a sub-set of critical path. It results in a new gate netlist, new timing and power figures, and the flow is normally re-executed: post CTS (hold and setup optimization), route and optimization until the design is timing, power and reliability closed. A certain number of back and forth steps is required to fully satisfy the initial design specification, as shown in figure 2.

For illustration, some timing analyses are presented in the Fig. 3. Based on an initial 5% worst slack selections, ISM are inserted in a sub-set of path. At step 3 (Fig. 2), histogram of paths are reported for an implementation with and without ISM. In the following analysis, delayed paths are not reported.

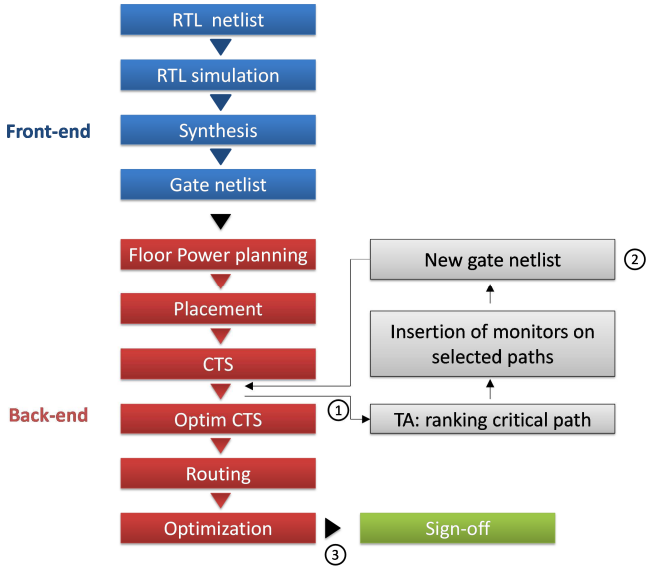

Fig. 2. Flow insertion of in-situ built-in monitors. During Front-End flow, a preliminary Timing Analysis is performed after pre-CTS step. In-situ monitors are inserted in sub-set of critical paths and a new gate netlist is generated. Then the Back-end flow is normally executed with new gate netlist.

A particular attention is paid to be sure that for flow-based ISM the inserted cells are physically the closest possible to the monitored FF. To achieve this objective, timing constraints are adapted to minimize the skew between shadow and regular FF. Moreover the delayed data arriving at shadow FF is not considered as a real path when the place and route tool optimize to fulfill the timing constraint. It means that there is theoretically no timing penalty after ISM insertion expected the one induced by the slightly additional routing resource.

It is important to notice that the delay monitored is in the order of magnitude of the degradation measured on test chips during ageing experiments. It is mandatory to have the highest timing accuracy level during monitor insertion. Whereas the insertion could be possible at synthesis during Front-end, the physical synthesis (Back-end) is able to account for parasitic effects. Thus, timing analysis at this level (post-route) is suitable and relevant to discuss the efficiency of the insertion flow. Benchmarking this methodology on different digital blocks is now reviewed to determine how the insertion flow can cover digital path ageing and establish the performance penalty

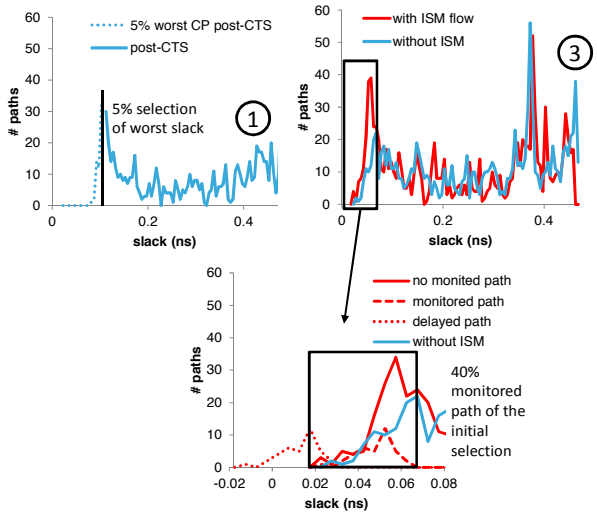

Fig. 3. Timing analysis of BCH results at different step of the flow. A preliminary TA at post-CTS is calculated (step 1). Based on this ranking, 5% worst slack are selected, and ISM are inserted. In the final TA (step 3), slack of monitored and none monitored paths are presented.

III. Benchmark Results

The ISM insertion flow is now applied on different circuits and performance versus ISM covering efficiency is reviewed. The design is synthesized and place-and-routed in 28FDSOI technology with Low-Vt devices. Several circuits are issued from ITC99 benchmark whose characteristics are typical of synthetized circuits. The b19, b15 and b14 have respectively 17, 6 and 10 Kgate after physical implementation. They are respectively composed of 2, 0.8 and 0.4 kFF. In addition industrial customer-related digital block is investigated as well. This block is a Bose, Ray-Choudhary and Hocquenghem (BCH) error correcting code IP consisting of encoder and decoder modules. The IP contains an output signal, Autotest, indicating if error correction is preformed correctly. More details about this circuit can be found in [1].

Different trials of implementation are performed for the same target performance with a large availability of area. Optimization of place and route tool has for first priority, performance, and then area/power. Worst negative slack at post-route step are discussed for all the circuits.

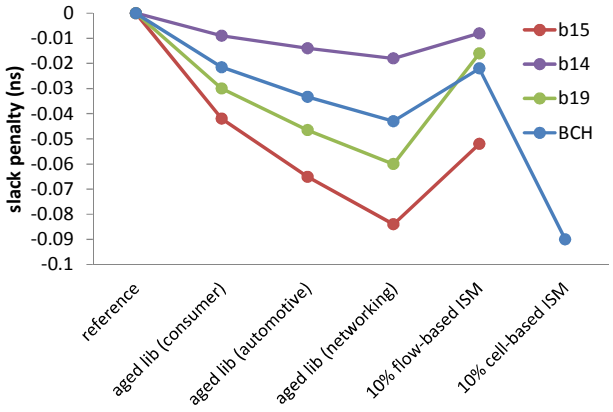

Fig. 4. Relative worst slack penalty for reference implementation, aged library implementation (for different mission profile), 10% flow-based ISM and 10% cell-based ISM.

As depicted in the Fig. 4, performance impact after ISM insertion might depend on circuit. Compared to reference circuit (fresh library without monitors), the implementation with an aged library (consumer, networking or automotive are mission profile dependent) leads to minor penalty. However this guard band enables the circuit to fulfill the timing requirement at the end of life. Concerning the ISM insertion, we chose a sub-set of 10% worst slacks (at step 1 of Fig. 2) to equip with monitors. BCH result shows a 90ps slack degradation for cell-based ISM and less than 20ps for flow-based ISM. The explanation for the penalty of cell-based penalty is the area constraint of the large custom cell. For the sake of clarity, the delayed data arriving at shadow FF for cell-based ISM is not reported in the TA of Fig. 4 for ITC99 benchmark.

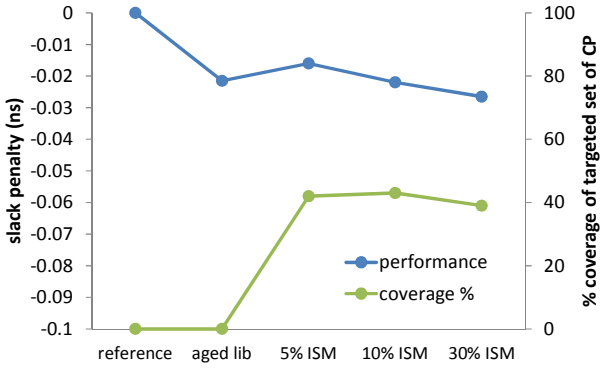

Fig. 5. Slack penalty for different implementations of BCH circuit. Increase of number of ISM leads to a slight timing degradation. The level of coverage (number of critical path monitored on initial critical path targeted to be monitored) remains close to 40%.

The number of ISM inserted and it performance impacts is discussed in the Fig. 5. For a selection of the most 5% of the critical data path, the performance penalty is only 15ps, and 40% of the initial selection is covered by monitors. A classic approach would consist to select the CP according to an absolute delay window and not in a number of CP criteria. Thus, the distribution of the CP and sub-CP histogram is important to analyze when using this approach. The violation hazards of a path due to the induced ageing failures are a function of its remaining slack. However, for the ISM flow, we use the number of inserted ISM as the metric under investigation.

IV. Example of Applications

After discussing the strategy and the benchmark of insertion ISM flow, some experimental results are now reviewed. Dedicated digital block are developed where on 10% of critical paths custom cell-based ISM has been inserted. We have investigated designs in 28nm, Low Power (LP) and Fully Depleted SOI (FDSOI) developed at STMicroelectronics. Digital block studied is the BCH block mentioned in previous section.

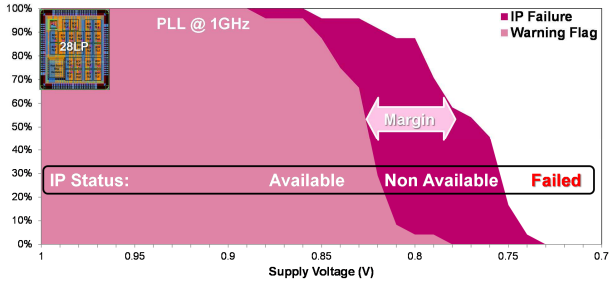

Fig. 6. Management of multicore architecture using ISM. At fixed 1GHz clock, when decreasing supply voltage, a warning Flag appears earlier before the IP failure. The 18 core safety margins are in a 100mV supply voltage range.

In the first application ISM are inserted in the architecture to manage the variability under optimum power budget. Major challenge in multicore architecture is to cope with inter-core dispersion. Indeed, local process dispersion leads to variation of speed and thus power consumption of all cores. To tackle this dispersion, an additional margin in the voltage stack needs to be used. It is not a trivial task to establish this margin because it is deeply influenced by the process centering and dispersion of manufacturability. Alternative approach is to insert ISM and to use their Flag as a warning to be considered as inputs of margin capabilities. As depicted in Fig. 6, 18 BCH cores are implemented in LP technology with ISM without any feedback loop. Under constant 1GHz clock frequency, when supply voltage is decreased, a first Flag monitor occurs at 0.99V, corresponding to a 1% of voltage decrease. At that point, the operating functionality is still correct. While supply voltage continue to decrease, more and more Flags occur on different cores and a first failure is reported (setup violation) at 0.85V. Interestingly, the V_{MIN} (minimal voltage sustaining to maintain functionality at a given PLL clock) distribution for all 18 cores, depends on the application execution of all cores and their ageing experience. To optimize the choice of voltage stack in multicore architecture, the strategy would be to monitor the first Flag of each core instead of using a conservative extra margin covering intra-core dispersion.

The second application focuses on monitoring the Flag number. The Flag number is the indicator for circuit speed and used for local variations along with aging aware voltage adaptation. An important measure campaign is performed on

300 dies using 3 time windows (TW1, TW2, TW3) and 5 workloads (Workload A, B, C, D, E). This workloads are patterns containing 0, 1,2,3,4 errors respectively. Fig. 7 shows the Flag count when decreasing the voltage until the Autotest signal fail using TW1 for different workloads. When modifying the pattern, the activity is modified and it results a strong modification on CP ranking. A direct consequence is that $V_{MIN\_AUTOTEST}$ (the supply voltage before Autotest signal fail) and V_{MIN_Flag} (the supply voltage when the count is starting) vary strongly with the workload ($V_{MIN\_AUTOTEST}$ ~ 0.8V and V_{MIN_Flag}~ 0.84 for Workload A, $V_{MIN\_AUTOTEST}$~ 0.825V and V_{MIN_Flag} ~ 0.825V for Workload D)

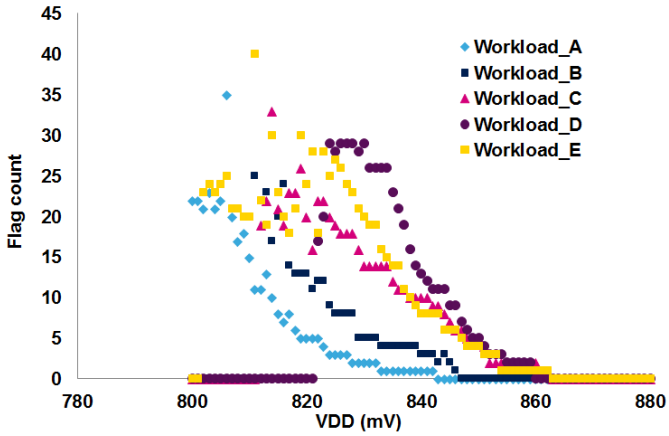

Fig. 7: Evolution of Flag number with VDD using time window 1 (TW1) for different workloads: A, B, C, D (0, 1,2,3,4 errors injected).

In order to demonstrate the robustness of ISM, it is important to test them under various conditions. For that propose, various temperature change have been exercised on BCH IP. Figure 8 shows the result of V_{MIN} variation under 30°C and 125°C for both Autotest and Flag signals. As depicted, degradation by 250 mV of $V_{MIN\_AUTOTEST}$ is observed when 125°C is applied, confirming the ability of ISM to capture local variations induced by temperature change.

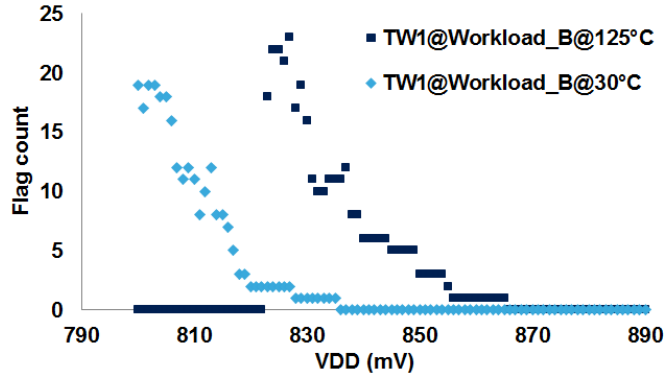

Fig. 8: Evolution of Flag number with VDD using TW1 and workload B under 30°C (magenta) and 125°C (blue) temperatures

V. Conclusion

A flow of in-situ monitor is developed and applied to different circuits. Two types of monitors are compared and discussed: cell-based and flow-based approach. Performance penalty and area overhead of ISM is slightly small. This additional margin provided by Flag signal is more accurate than the additional voltage stack margin to account for ageing degradation. The coverage path statistics, number of critical path monitored on desired critical path to be monitored, is around 40%. This approach is suitable for dynamic management of ageing because at long-term, the probability to activate one path from critical path selection is high. Some applications of adaptive regulation are illustrated, this scheme is promising for process compensation and temperature change.

References

[1] V. Huard, "Adaptative wear out management with in-situ management" IRPS 2014
[2] X.Wang ,"Path-RO: a novel on-chip critical path delay measurement under process variation IEEE ACM(2008)
[3] S.Wang "Representative Critical Reliability Paths for low-cost and accurate on-chip aging evaluation" IEEE/ICCAD (2012)
[4] Saliva.M ,"Digial circuits reliability with in-situ monitors in 28nm fully depleted SOI " IEEE/DATE (2015)
[5] S. Das et al.,"A Self-Tuning DVS Processor Using Delay-Error Detection and Correction" IEEE J. Solid-State Circuits, Apr. (2006)
[6] D. Blaauw et al.,"RazorII: In Situ Error Detection and Correction for PVT and SER Tolerance" IEEE J. Solid-State Circuits,Jan. (2009)
[7] K.A.Bowman "Energy-efficient and Metastability-Immune Resilient Circuits for Dynamic Variation Tolerance" IEEE Journal of Solid-State Circuits
[8] M. Wirnshofer "A Variation-Aware Adaptive Voltage Scaling Technique Based on In-situ Delay monitoring" IEEE/DDECS (2012)