On the Exploitation of the Inherent Error Resilience of Wireless Systems under Unreliable Silicon

Georgios Karakonstantis[†], Christoph Roth[‡], Christian Benkeser[‡], Andreas Burg[†]
[†]Telecommunications Circuits Lab (TCL), EPFL, Lausanne, VD 1015, Switzerland
[‡]Integrated Systems Lab (IIS), ETH, Zurich, ZH 8092, Switzerland
{georgios.karakonstantis, andreas.burg}@epfl.ch, {rothc, benkeser}@iis.ee.ethz.ch

ABSTRACT

In this paper, we investigate the impact of circuit misbehavior due to parametric variations and voltage scaling on the performance of wireless communication systems. Our study reveals the inherent error resilience of such systems and argues that sufficiently reliable operation can be maintained even in the presence of unreliable circuits and manufacturing defects. We further show how selective application of more robust circuit design techniques is sufficient to deal with high defect rates at low overhead and improve energy efficiency with negligible system performance degradation.

Categories and Subject Descriptors

B.8.2 **[PERFORMANCE AND RELIABILITY]**: Performance Analysis and Design Aids.

General Terms

Algorithm, Design, Reliability.

Keywords

Error-Resiliency, Memory Failures, Wireless Communication Systems, Energy-Efficiency, Reliability, Yield.

1. INTRODUCTION

With the enormous success of wireless communication systems in the last decade, users are asking for ever higher data rates and better quality of service (QoS). However, sophisticated algorithms/systems with increasing percentage of memory components are required in the transceiver IC to meet the throughput requirements of latest wireless communication standards [1]. Unfortunately, this algorithm-complexity increase and especially the exploding memory requirements of the latest communication standards lead to a paramount increase of power consumption, making energy efficiency one of the main challenges in the design of emerging wireless systems.

Though several schemes exist that try to address the increased power consumption, one of the most effective techniques for low power implementation is still considered to be voltage scaling (VS) due to the quadratic dependency of power consumption on voltage [1,2]. However, VS reduces the circuit performance and increases circuit sensitivity to parametric variations that originate from nanometer device sizes and inaccuracies in the delicate fabrication processes [2-5]. Such variations not only lead to delay and memory failures, that could even worsen over time (i.e., due to aging), but also increase the spread in leakage current, making it more difficult to meet today's strict throughput and energy requirements with decent yield. In addition, the shrinking of dimensions to 65nm and below increases layout density, which is of particular importance for area-efficient memory design, but at the same time, reduces the amount of charge required to upset a circuit node and raises the likelihood of having a large number of soft errors on chip [5, 6].

Several approaches exist today that try to address both power consumption and parametric variations simultaneously. However, they often lead to significant area overhead and limit the gains in power consumption since they are based, for instance, on the addition of redundant hardware [2,7]. Interestingly, as the percentage of memory components in wireless systems increases (crucial for supporting the large data load), the overhead of such techniques makes them prohibitive, reducing their viability. For instance, error correction coding (ECC) and novel bit-cell architectures (i.e., 8T) might tackle the high failure probability of traditional 6 transistor (6T) bit-cells (under variations and VS) but can lead to more than 50% power overhead [5-9]. However, although in general purpose processors/systems the overhead of such techniques might still be acceptable due to the equal significance of all data, it might be possible in application-specific systems to depart from the 100% error-free computing paradigm. By accepting dies even with a number of defects or restricting application of robust techniques to only the most critical parts of the system we could improve yield and energy efficiency at no/limited overhead even in the presence of hardware errors (due to VS and/or variations). While several approaches tried to take advantage of such an observation in order to address the issues of power and variations in multimedia systems and individual DSP blocks [2, 8, 9], to the best or our knowledge, no such effort has targeted wireless communication systems so far. Therefore, there is a need to study the impact of hardware errors induced by parametric variations and VS on the performance of such systems, which are ubiquitous components of all today's portable devices. Interestingly, the main characteristic of such systems is that corresponding receivers with sophisticated communication algorithms are able to recover the transmitted data even when the received signal has been heavily distorted by noise and interference due to bad wireless channel conditions. This robustness of such systems motivates the investigation of their inherent resilience against unreliable silicon implementations and raises questions regarding the limits of this error resilience and how it could be improved at low cost.

To this end, in this paper we investigate the impact of hardware defects/errors induced for example by VS and parametric variations on the performance and yield of wireless communication systems using the latest high-speed packet access evolution of the 3G mobile cellular standard HSPA+ [10]. Our study focuses on a large and power hungry memory required for the hybrid automatic repeat request (HARQ) block that is critical for the correct and high throughput operation of the overall system. Our contributions can be summarized as follows:

- Develop a system-level fault simulation approach for capturing the effects of errors on the system performance and relating the

DAC 2012, June 3-7, 2012, San Francisco, California, USA.

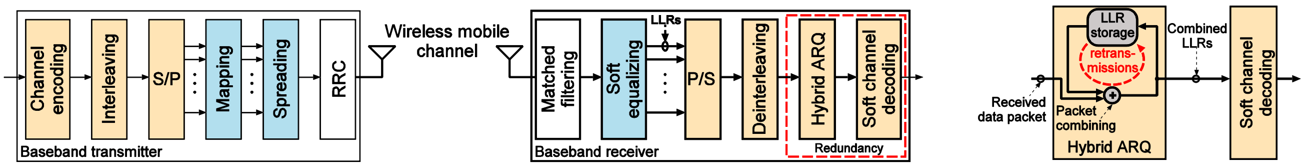

Figure 1: (a) Wireless Communication System (HSPA), (b) Principle of HARQ Operation

results to the yield in a meaningful way.

- Exploit the resilience limits of communication systems moving away from the 100% reliable computing paradigm. Interestingly, we find that the system is able to operate correctly even in the presence of hardware errors, but as such errors increase beyond a critical rate (making them comparable to channel-induced errors) the system throughput deteriorates significantly. This finding allows us to actually accept dies with up-to a specific number of defects leading either to a better yield or enabling energy reduction (since such dies may operate at low voltages).
- Explore low-overhead techniques that can improve robustness and facilitate aggressive VS, thus improving energy efficiency under very high defect rates. Specifically, we show how selective application of robust circuit techniques, such as 8T cells, only on some critical parts of the system (that are identified by our study) can reduce the overhead of conventional conservative robustness techniques that aim at restoring 100% reliable operation while ensuring minimum system throughput at high yield loss under high defect rates.

The rest of this paper is organized as follows. In Section 2 we briefly present the basic characteristics of a modern HSPA+ communication system which serves as an excellent and commercially relevant test vehicle for our study. Section 3 discusses the various failure mechanisms and their impact on memory cells/arrays and yield. Section 4 presents our approach for studying the impact of errors on throughput and yield. Section 5 then reveals the resilience limits of communication systems to hardware defects and discusses the achieved yield improvement. Section 6 proposes low overhead techniques for improving the robustness of the system. Conclusions are drawn in Section 7.

2. ERROR RESILIENCE OF WIRELESS SYSTEMS

In the following we briefly summarize an HSPA+ system as specified by the 3GPP standard suite [10], which serves as a challenging vehicle to verify our findings and to demonstrate the effectiveness of our proposed low-overhead techniques for modern wireless communication systems. Arguing that a communication system is designed to cope with noisy data, we will then highlight the inherent error resilience of such a system and exploit its characteristics that may help in tolerating hardware induced errors.

2.1. HSPA+ System Model

HSPA+ is based on code division multiple access (CDMA), a channel access method where a single wireless transmission channel is simultaneously shared by several users. A simplified block diagram of an HSPA+ baseband transmitter and receiver separated by a noisy mobile channel is shown in Fig. 1(a).

Baseband transmitter: On the transmit side, a sequence of data bits, referred to as data packet, is encoded using a high-performance error-correction code and then passed through an interleaver which generates a pseudo-random permutation of the input bit stream. This serial bit stream is then converted into parallel streams and each of these streams is individually modulated (with either 16QAM or 64QAM) before they are spreaded and multiplexed to a single stream in the spreading unit. Finally, this stream of multiplexed data symbols modulates a root-raised cosine (RRC) pulse-train which is then transmitted over the mobile channel.

Baseband Receiver: The main task of the receiver is to extract the originally transmitted bit stream from the distorted received signal using sophisticated equalization and channel decoding algorithms, which are the most challenging blocks in terms of implementation complexity. While the equalizer attempts to undo the destructive effects of the mobile channel, the decoder corrects errors in the equalized data packet, exploiting the redundancy and structure in the transmitted bit stream imposed by the channel encoder. Rather than deciding on hard bits, a soft-decision equalizer produces reliability-indicators, referred to as log-likelihood ratios (LLRs), representing the probability for each bit being logic-0 or logic-1. The magnitude of an LLR reflects the confidence, and the sign shows whether a decision would be in favor of logic-0 or logic-1, respectively. A soft-decision channel decoder works on LLRs instead of simple bits. Clearly, a soft receiver (based on LLRs) implies higher implementation complexity in terms of silicon area and power consumption compared to a hard receiver but the considerable gain in performance, required to fulfill the demanding 3GPP specifications, justifies the overhead. An important performance metric in such a system is the block-error rate (BLER) which is the probability that the channel decoder fails to decode a data package. This metric is usually measured as a function of the signal-to-noise ratio (SNR) at the input of the receiver, representing the ratio of the user signal power over the noise and interference power.

Hybrid automatic repeat request (HARQ): A key feature on the terminal side of an HSPA+ downlink is the HARQ operation, which allows for rapid retransmission of erroneously received data packets. HARQ is a crucial mechanism to enable *high average throughput*, the ultimate performance metric of such systems, over a wide range of rapidly varying mobile channel conditions. The main principle of HARQ is depicted in Fig. 1(b). The received data packets are buffered in the LLR storage prior to decoding. In case the channel decoder fails to decode a data packet, a retransmission is requested by the receiver. In contrast to traditional ARQ-based communication systems where simply the retransmitted data packet is decoded, the HARQ operation combines the retransmitted data packet with the (stored) information (i.e., LLRs) of previous transmissions, increasing the probability of correct decoding. The higher the quality of the combined LLRs used by the soft-decision channel decoder, the lower the average number of retransmissions required to successfully deliver a data packet even under channel errors.

2.2. Error Resilience to Channel Noise

The above functionality reveals the main characteristic of such systems; their ability to operate reliably under channel noise. This is clearly indicated in Fig. 2 that depicts the decoding-failure probability of a data packet (i.e., BLER) evolving over the incremental HARQ retransmissions for three different SNR regimes. In the high SNR (29dB) regime, the channel decoder is able to decode roughly 95% of all data packets already after the initial transmission. For the medium SNR (11dB) regime, the channel decoder is still able to deliver a considerable fraction of

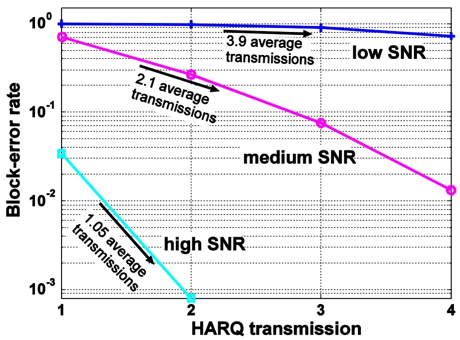

Figure 2: Decoding Error Probability.

the data packets in the initial transmission, revealing the inherent resilience of the system to noisy input data. However, in the low SNR (3dB) regime, the channel-induced noise corrupts the data too severely, and virtually all data packets are scheduled for retransmission. While in a traditional ARQ-based system this would drive the throughput performance to zero at this low SNR, the LLR combination in the HARQ unit increases the decoding probability after each retransmission due to more reliable LLRs as shown in Fig. 2. It is apparent that more retransmissions reduce the throughput.

3. HARDWARE ERRORS AND IMPACT ON YIELD AND MEMORY OPERATION

As explained in the previous section, the receiver's ability to decode the received stream correctly heavily depends on the operation of the HARQ unit. The main component of this block as shown on Fig. 1(b) is the LLR memory that stores the received data packets and combines them with the corresponding retransmissions. Striving for a fully integrated baseband solution, this storage is typically implemented with SRAM memory cells, which account for a considerable fraction of both silicon area and power consumption of the overall system besides the equalizer and channel decoder. The latency in such a complex wireless system combined with the high data rates involved thereby inflate the required HARQ storage size, which can range up to 253 Kb times the number of bits used to quantize an LLR. Unfortunately, while the continuous scaling of devices allows for the realization of such high density memories in a single chip, the small sizes beyond the sub-65nm node make devices more prone to variation-induced defects. To better understand the nature of such defects we briefly explain the basic sources of hardware errors and their impact on yield and SRAM operation. In general, memory failures can be persistent (i.e., failures due to difference in transistor characteristics causing yield loss) or non-persistent (e.g., soft errors due to radiation) and the probability of both of them increases as supply voltage decreases [2, 6, 7].

Parametric Variations: The primary source of device mismatch, which is the dominant failure mechanism in memory cells, is the intrinsic fluctuation of the threshold voltage (V_{th}) of different transistors due to random dopant fluctuations (RDF) [2-8]. Any mismatch in V_{th} of neighboring transistors in an SRAM cell can result in a failure of the corresponding bit cell. For instance, a cell failure can occur due to, i) unstable read (flipping of the cell while reading) and/or write (inability to successfully write to the cell), ii) increase in the cell access time (access time failure), and iii) failure in the data hold capability of the cell in the standby mode. Since these failures are caused by the variations in the device parameters, they are known as parametric failures [11]. The degree of such failures depends on the size/type of the memory bit-cell, but also on the array organization and strongly on the supply voltage (V_{dd}). Specifically, as the on-chip memory density increases, lowering the supply voltage is the most effective approach for low power operation in order to meet the tight power budgets in wireless communication systems. However, a supply voltage below its nominal value increases the sensitivity of circuits to RDF and thus leads to higher number of failures. Fig. 3 depicts the failure probability of a memory array implemented by medium-sized 6T bit-cells, 15% upsized 6T cells and 8T bit-cells under various voltages in case of slow-fast corner, which was found to be the worst corner for RDF induced memory failures in the 65nm technology node [5, 9]. Such failure rates are directly related to the yield of a memory block. It is apparent that as the effect of intra-die variations increases with technology scaling and lower voltages the memory failures increase and thus yield decreases accordingly. Conventionally, the addition of redundant rows/columns could help to recover from such defects, but as the size of memory and the number of defects increases they are insufficient to avoid yield loss. Moreover, the number and the location of failures due to process variations changes depending on operating condition (e.g., applied V_{dd} and frequency) which cannot be handled efficiently by redundant rows/columns.

Soft Errors: The small size of transistors have made it also easier to upset the stored charge in a node giving rise to soft errors with a rate that is almost constant across technology generations [5]. Such errors do not damage the cell permanently, and studies have shown that they do not depend so much on voltage since they only increase by a factor of 3x for every 500mV decrease in supply voltage as opposed to RDF induced errors that increase by billion times for such a voltage decrease (Fig. 3).

In general purpose systems, techniques such as transistor up-sizing, novel bit-cell configurations (8T) and error correcting codes (ECC) can decrease the failure probability of a memory array to improve yield or enable operation at lower supply voltage. Unfortunately, all these techniques come at an increased cost in terms of silicon area and power consumption overhead. Such additional costs may be prohibitive for wireless communication systems that need to deliver large amounts of data rates at very low energy as part of battery-operated consumer-electronics devices.

The proven inherent resilience of the considered system to channel noise, as discussed in Section 2.2, suggests that such systems may also be able to cope with additional distortions introduced by unreliable hardware. We therefore propose to depart from the 100% error-free computation paradigm and accept hardware-induced errors up to a certain defect rate. This paradigm-change would not only enable more aggressive voltage scaling in wireless communication systems, but it would also facilitate achieving the demanding manufacturing yield targets that are critical for today's cost sensitive applications. In the next sections we investigate the limits of the inherent error resilience of wireless systems by considering the effect of hardware failure

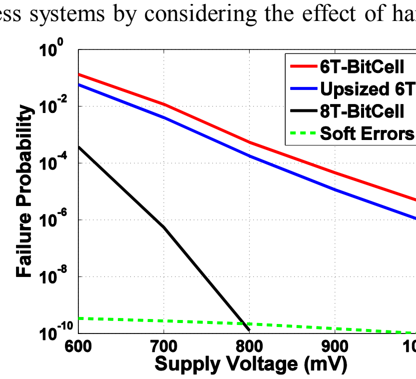

Figure 3: Memory Failure Probability (65nm).

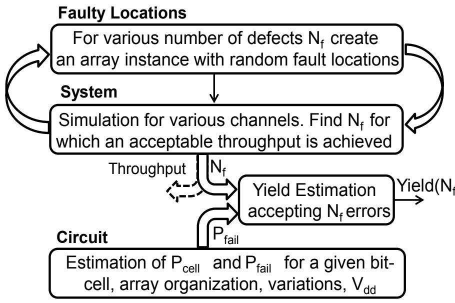

Figure 4: System Level Fault-Simulation Approach.

mechanisms in system-level simulations. Based on these results, we further identify techniques to maintain acceptable system operation beyond this point with minimum additional cost.

4. SYSTEM LEVEL FAULT-SIMULATOR

In the following, we describe our approach for jointly considering circuit misbehavior and the consequences on yield together with the impact on the overall system performance. This analysis will later also be instrumental in identifying the few most critical parts for the operation of the system that may need protection under high hardware failure rates. The primary challenge in estimating the system-performance impact of errors in not-100% operational dies is that meaningful throughput evaluation requires a vast amount of Monte-Carlo simulations averaging over various wireless channel conditions. Unfortunately, as we have departed from the 100% error-free processing paradigm, individual devices may be different since they may be affected by a different number of errors distributed across the storage array according to one out of billions of possible error patterns. To nevertheless capture the effects of the number and the location of defects on the system performance in a meaningful way and to relate the results to yield, we developed the system level fault-simulation methodology depicted in Fig. 4.

Estimation of Cell-Failure probability (P_{cell}): Initially, the failure probability (P_{cell}) of the desired bit-cell type under various degrees of variations and voltage scaling are obtained through Monte-Carlo circuit simulations.

Yield Estimation - (Y): In a conventional, 100% defect-free design the cell-failure probability immediately leads to the failure probability of the overall memory array (P_{fail}) using simple methods that can also consider some advanced robustness techniques such as redundant columns, error correction codes (ECC) and array organizations [6, 7, 11-12]. For example, by assuming that all cell failures are independent, one can easily estimate the array failure probability for an array size of M cells and thus the yield (Y) for this part of the circuit according to:

$$Y(100\%\ reliable) = (1 - P_{cell})^M \quad (1)$$

However, as discussed above, the inherent error resilience of wireless systems may allow tolerating a limited number of failing cells. This relaxation makes the acceptance of faulty dies possible, which otherwise would be discarded, thus improving the yield. Alternatively, the relaxed selection criterion enables meeting the yield target at a reduced nominal supply-voltage which leads to the desired power reduction. Keeping in mind the impact on the throughput we can investigate the yield that can be achieved by tolerating a number or percentage of faulty cells for a given memory array with size M and cell failure propability P_{cell}. To this end, we redefine yield (Y) for the case where chips with at most N_f faulty cells pass the inspection:

$$Y(N_f) = \sum_{i=0}^{N_f} \binom{M}{i} \times (P_{cell})^i (1 - P_{cell})^{M-i} \quad (2)$$

The above equation reveals the yield improvement that a manufacturer can get by not discarding chips with a specific number of faulty cells. Fig. 5 plots $Y(N_f)$ for various P_{cell} and various numbers of N_f. Note that each P_{cell} corresponds to probability of defects due to voltage scaling and parametric variations as we discussed in Section 3. From such a figure we can determine the number of defects that we need to accept for achieving the yield target. For instance in case of $P_{cell} = 10^{-4}$ and $M = 200$Kb, chips with 0.1% defects need to be accepted for meeting the target yield (95%). For determining how many faulty cells N_f can be tolerated we need to evaluate the impact on the throughput of the overall system as we describe next.

Wireless System Simulation - (Throughput): Since we do not require zero defects, we need to assess the worst-case system performance for the dies that pass the selection process (each of which can be affected by different defects within the specified selection criterion). To this end, we consider only the case of N_f defects distributed across the array using random fault-location maps. For a given wireless channel realization, each bit of the received LLRs is mapped to a specific memory cell in the LLR memory array. If the mapped location of the 'bit' indicates a fault in the fault location map, the 'bit' is inverted to indicate a bit-error. These bit flips are considered in the MATLAB Monte-Carlo system simulations and the impact of circuit misbehavior is evaluated using the appropriate system metrics (i.e., average throughput), as also prescribed by the corresponding communication standards.

5. EXPLOITING THE RESILIENCE LIMITS

In this section we evaluate the impact of defects in the LLR storage on the throughput performance of a fully standard-compliant HSPA+ system. We present worst case simulation results for the most noise-sensitive, high throughput 64QAM modulation mode and for a maximum of three retransmissions per data packet over a standard-compliant multipath channel. A minimum mean-square error (MMSE) equalizer is used for the generation of LLRs which are quantized with 10 bits to avoid any throughput-loss due to quantization noise over a wide range of SNR points (according to our simulations of a defect-free system).

Having set the above parameters in our simulation framework we use the approach discussed in Section 4 for injecting errors at random locations of the LLR storage (assuming a medium sized 6T based memory) . In our simulations we cover various choices for N_f. Note that the system-performance results reflect the worst-case behavior of dies with exactly N_f failing cells (i.e., for a given selection criterion) and are thus independent of the failure probability P_{cell}. However, N_f and P_{cell} together define the impact on yield and, due to the dependence of P_{cell} on the supply voltage, also the potential for power savings.

Results: Fig. 6(a) depicts the throughput performance of the considered system for various choices of N_f, specified in % of the size of the LLR storage array. We observe that the throughput is roughly the same as that of the defect-free system (up-to a 0.1% defects), highlighting the inherent resilience of wireless communication systems to unreliable storage. Furthermore, the simulations reveal that the described system is able to meet the required (normalized) throughput (0.53 at 18dB) specified by the standard for this mode of operation (64QAM) withstanding even 10% of defects in LLR storage (corresponding to 2000 defective cells). This indicates that there is no need for protective mechanisms in the LLR storage up to that amount of defects. This resilience allows not only to avoid the cost for protective mechanisms, but also to lower the supply voltage since for example a memory based on conventional 6T cells can function at 0.8V (lower by 200mV compared to 1V in 65nm (Fig. 3, 5)).

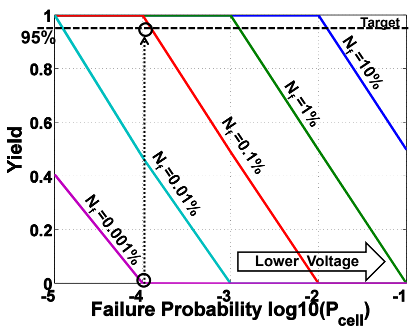

Figure 5: Yield estimation (200Kb array).

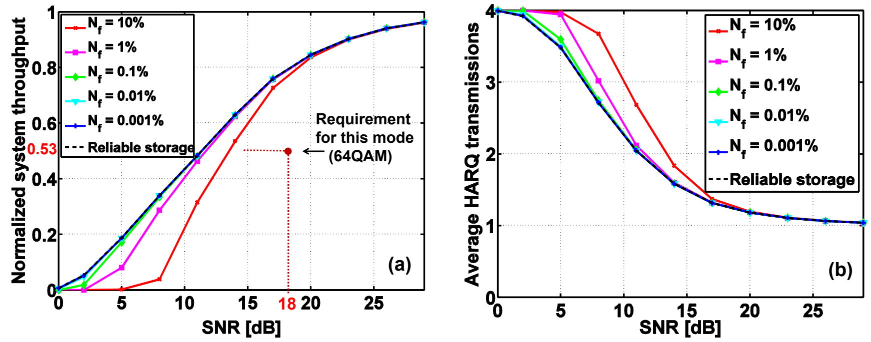

Figure 6: (a) Throughput, (b) Average number of transmissions under various defect rates.

As the number of tolerated defects increases beyond 0.1%, the quality of the LLRs deteriorates to a point that becomes dominant over the effects of the signal-distortions due to the wireless channel, increasing the average number of transmissions required to successfully decode a data packet as shown in Fig. 6(b). As outlined in Section 2, this in turn reduces the throughput and increases the overall energy required to deliver a data packet since the entire transmitter/receiver chain is forced to handle the incurred overhead. Hence, a further yield increase under severe process variations or further voltage scaling (increasing P_{cell}) without degrading yield, requires more sophisticated measures to maintain good system performance (throughput) while tolerating more faults.

6. IMPROVING RESILIENCE AND YIELD

In this section we discuss how selective application of more robust circuit design techniques to the LLR storage is sufficient for allowing the wireless system to operate reliably at a low cost in case of high defect rates.

6.1. Proposed Storage Approach

As discussed above, conventionally, designers would apply expensive methods in terms of area and power to the complete memory array such as ECC or larger transistor/different type of cells (i.e., 8T) in order to enhance the robustness of the memory. However, not all bits are of equal weight (e.g., the sign information is of higher importance than the rest bits for the channel decoder). Hence, such expensive techniques for the protection against failures may not be required for all bit-cells. In order to determine the number of LLR bits that need to be protected in order to obtain an acceptable throughput even with a large number of faulty cells in the remaining bits (corresponding to a better yield), we performed a sensitivity analysis by utilizing the approach discussed in Section 4. Specifically, we consider zero or a very low number of tolerated defects ($N_f^{MSB} \leq 0.01\%$)) in the well protected bit locations starting from the most significant bit (MSB), while in turn considering a high number of tolerated defects in the less well physically protected rest of the bits ($N_f \geq 0.1\%$). In other words, for the sensitive parts of the data we propose to meet the yield target by reducing the cell-failure probability using for example 8T SRAM cells. In exchange, we continue to use area- and energy- efficient, but unreliable 6T cells for the less sensitive parts (bits). We speculate, that we can now tolerate an even higher number of faulty cells for these less significant bits to achieve the yield target even under more severe process variations or at lower voltages. The corresponding analysis reveals that protection of few MSBs is sufficient and allows for a high number of accepted defective cells in the remaining bits without jeopardizing throughput. This is evident by comparing Fig. 7 to Fig. 6, where it is shown that by protecting only the 3-4 MSB bits (rather than all bits) is suffcient for limiting the throughtput loss even under 10% defect rates.

6.2. Efficiency of Protection

Of course by protecting more bits, a higher number of defects can be accepted improving the throughput and yield. However,a higher number of protected bits increases the associated area penalty proportionally. In Fig. 8 we plot the throughput gain (Throughput(N_f)/ defect-free Throughput) achieved by protecting various number of bits divided by the area overhead needed by using more robust cells in the case of $N_f = 10\%$. We assume the use of 8T cells for the protection of bits and we plot the overhead of a hybrid array (8T and 6T cells) over the area of a 6T-based array. By focusing on the point where the system with unprotected storage cells experiences the worst-case throughput penalty compared to the error-free case (here at an SNR=8dB), we observe that protecting 4 bits is optimum. The protection of more bits causes further increase in silicon area without any significant throughput improvement. This observation also proves that the conventional approach of using equal protection for all bits cells is not as efficient as protecting few MSB bits only.

Similarly we can argue that the use of ECC protection of all

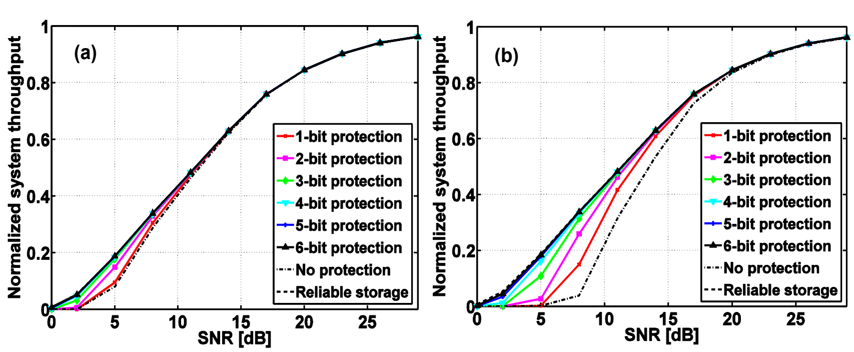

Figure 7: (a) Throughput after protecting various numbers of bits under various defect rates (a) N_f=1% in 6T cells, (b) N_f=10% in 6T cells.

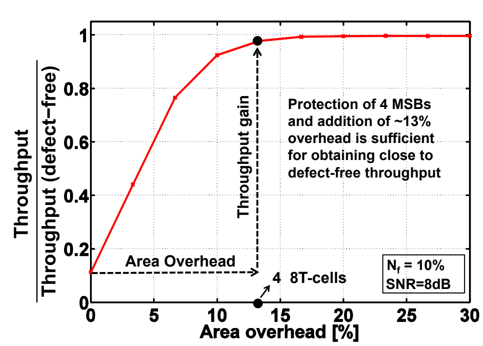

Figure 8: Protection efficiency.

bits is not efficient either. Specifically, a single detection and correction ECC method could be used for the protection of all 10 bits. However, this would result in 35% area overhead compared to the 6T-based array, since 4 redundant bits are required for a single error correction according to Hamming based ECC [5, 6, 8]. Furthermore, the use of higher order ECC for the protection of more bits increases the area and power by more than 50% [6].

Note that for the implementation of such a hybrid memory we could also utilize upsized 6T cells for the protection of the required bits. However, it was shown recently that 8T cells lead to lower area and power overhead for the same stability improvement [9]. In any case, the same techniques applied in the design of such a hybrid memory for multimedia applications [9] can be utilized also here. Nonetheless, a selective protection of LLR bits is very efficient since it protects only what is necessary for obtaining good throughput limiting any unnecessary overhead of conventional techniques (i.e. use of 8T cells in whole memory).

6.3. Potential for Power Reduction

The improvement in throughput achieved by selective protection of significant data bits translates into improvement of yield and offers the potential for power savings through aggressive voltage over-scaling beyond i) the limit of reliable operation of 6T memory cells and ii) the limits imposed by the inherent error resilience of the fully unprotected system described in Section 5. Specifically, as discussed in Section 6.1 the proposed storage scheme can limit the throughput loss, providing acceptable performance even under high number of defects (1%-10%) which could be induced by aggressive voltage scaling down to 0.6V (Fig. 3, 5, 7(b)). In other words, our proposed storage scheme can allow the operation of wireless system with acceptable throughput even at a low supply voltage for the HARQ memory block that can translate to 30% power savings for that block under an iso-area comparison with a conventional 6T SRAM array [9].

Furthermore, the proposed prefential storage does not only reduce the power locally in the HARQ block but also in the overall system. This is evident if we consider Fig. 6(b) and the number of retransmissions. For instance in case of SNR=9dB we can observe that with the utilized partial protection we need 2.4 retransmissions as opposed to 3.5 retransmissions in case of no memory protection under N_f=10%. Therefore, the preferential storage scheme increases the ability of decoder for correct decoding thus reducing the retransmission rate which has an immediate impact on the whole system energy efficiency.

6.4 Joint Consideration of Bit-Width and Defects

One of the main system level decisions that need to be taken into account, while designing a wireless system is the degree of quantization. Traditionally designers tend to use more bits for the quantization for ensuring minimum quantization noise and thus minimum impact on throughput. However, a high number of bits increase the size of the required storage, making memories not only larger, but also more prone to hardware errors. This reveals that when deviating from the paradigm of 100% correct operation, circuit level limitations should also be considered when making decisions on quantization. The necessity for such considerations is suggested also by Fig. 9. Although the 10-bit quantization introduce more noise than using more bits at the high SNR points, it actually results in a better throughput compared to 11/12 bits (which would be the selection of designers in case that only channel noise is considered) with cell failures. This can be attributed to the fact that the system becomes more sensitive to failures in the memory which due to its larger size (in case of 11 and 12 bits), becomes more prone to hardware errors.

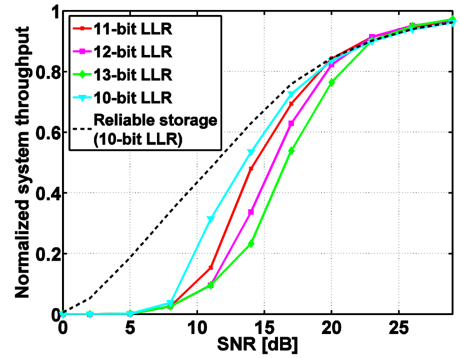

Figure 9: Throughput under various bit-widths (no protection with 10% defects).

7. CONCLUSION

The paper proposes the departure from the paradigm of 100% reliable circuit operation for the design of wireless communication systems. Our study reveals that wireless systems are able to tolerate a considerable number of defects allowing for the acceptance of defective dies. Focusing on the large storage array in the Hybrid ARQ subsystem of the 3GPP HSPA+ standard we show that this not only translates directly to a yield improvement, but also to power savings since circuits can be operated at lower supply voltage. We further show that only partially protecting the memory content ensures reliable operation even under high number of defects at low cost. This preferential storage scheme enables further power savings (through voltage scaling) and reduces the circuit-level overhead required to provide robust operation in sub-65nm process nodes. Overall, our study suggests that taking hardware errors into account already in the system-level design of future wireless systems can be beneficial for achieving robust low power solutions.

8. ACKNOWLEDGMENTS

This research was supported by Swiss National Science Foundation under the project number PP002-119052.

9. REFERENCES

[1] J. M. Rabaey, "Low Power Design Essentials", Springer, 2009.

[2] S. Bhunia, et al, "Low-Power Variation-Tolerant Design in Nanometer Silicon," Springer 2011.

[3] S. Borkar, et al., "Design and reliability challenges in nanometer technologies," *IEEE DAC*, pp.75, 2004.

[4] A. Shrivastava, et al., "*Statistical Analysis and Optimization for VLSI: Timing and Power*", Springer, 2005.

[5] Z. Chishti, et al., "Improving Cache Lifetime Reliability at Ultra-low Voltages," *IEEE MICRO*, 2009.

[6] C. Wilkerson, et al., "Trading off Cache Capacity for Reliability to Enable Low Voltage Operation," *IEEE ISCA*, 2008.

[7] Shi-Ting Zhou, et al. "Minimizing Total Area of Low-Voltage SRAM Arrays through Joint Optimization of Cell Size, Redundancy, and ECC," *IEEE ICCD*, 2010.

[8] Y. Emre, et al., "Memory Error Compensation Techniques for JPEG2000," *IEEE SiPS*, 2010.

[9] I. J. Chang, et al., "A Priority-Based 6T/8T Hybrid SRAM Architecture for Aggressive Voltage Scaling in Video Applications," *IEEE Trans. on CSVT*, 2011.

[10] High speed downlink packet access (HSDPA), Third Generation Partnership Project TS 25.308, Rev. 10.5.0, Jun. 2011.

[11] S. Mukhopadhyay, "Statistical design and optimization of SRAM cell for yield enhancement", *IEEE ICCAD*, 2004.

[12] P. Zuber, et al., "Statistical SRAM analysis for yield enhancement," *IEEE DATE*, 2011.