MASSIVE MIMO PROCESSING AT THE SEMICONDUCTOR EDGE: EXPLOITING THE SYSTEM AND CIRCUIT MARGINS FOR POWER SAVINGS

Yanxiang Huang[*†], Claude Desset[*], André Bourdoux[*], Wim Dehaene[†*], Liesbet Van der Perre[†]

[*]IMEC, Kapeldreef 75, Leuven B-3001, Belgium
[†]KU Leuven, Kasteelpark Arenberg 10, Leuven B-3001, Belgium

Abstract—Massive MIMO has the potential to bring great spectral and energy efficiency improvements, making it a very promising technology for future wireless systems. Essential to achieve the gains in practice, is the ability to realize the many antenna paths at low complexity.

In this paper, we consider the potential of processing at the semiconductor edge by allowing voltage over-scaling and complete antenna signal failures, focusing on the per-antenna digital functionality that dominant the DSP complexity. The impact of the resulting hardware errors on the performance of Massive MIMO transmission is analyzed. It shows that the inherent redundancy in the system brings a solid tolerance to sporadic hardware errors. Potential control tactics are introduced, that could further optimize the operation of the error-prone circuitry. We anticipate that by exploiting the system and circuit margins, up to 40% power reduction could be achieved on the considered DSP functions without sacrificing performance in many traffic scenarios.

Index Terms—Digital Signal Processing, Massive MIMO, CMOS, Low-power computing, Reliability

I. INTRODUCTION

Massive MIMO (MaMi) opens up a new dimension of wireless communications by using an excess of base-station antennas, relative to the number of active terminals. This technique allows for very efficient spatial multiplexing, attainable using linear processing in a time-division duplex mode [1]. Conceptually, a 10x or more increase in system capacity can be achieved with MaMi. Perhaps even more important is the significant gain in reliability due to flattening out of deep fades, hardening of the channel, and array gain. This especially benefits cell edge users, and could be essential for low power terminals as in Machine Type Communications (MTC).

An obvious concern is how the large number of antennas (and associated transceivers and signal processing) will affect the complexity and energy consumption of the base station. An in-depth analysis confirms the spectacular promise: it is anticipated that the overall complexity and energy consumption in terms of J/bit can be lowered by a factor of 10 to 100 with respect to current base stations [2].

This stunning improvement results from the fact that, on the one hand, much less transmitted power is needed thanks to the array gain, and, on the other hand, relatively low complexity hardware can suffice, as the individual antenna signals do not need to be high precision [3]–[5]. The resiliency to non-ideal hardware (e.g. analog non-linearity) is demonstrated in [5]. We expect further considerable energy reduction potential if we can extend the low accuracy need to the CMOS implementation of signal processing – to the point of sporadically processing distortion or even full failure of one or a few individual antenna signals. Indeed it open the door to operating semiconductor circuits with much lower design margins (comparing with the traditional specification set at design-time), and at lower supply voltage (and hence power).

From the CMOS technology perspective, integrated circuits (ICs) are facing ever increasing variability challenges in recent technology nodes (65nm and smaller). The process, voltage and temperature (PVT) variability are considered the three main contributors to circuit variability. Conventionally, to cope with this variability challenge, ICs are designed at the worst PVT corners, to ensure that they always

operate correctly (Fig. 1). This approach introduces considerable margins, leading to reduced peak performance and wasted power consumption. For instance, [6] shows for 28-nm technology, the performance difference (in terms of speed) is as large as 2.2X between typical case and worst-case. Dynamical scaling techniques manage power dissipation and temperature using variable V_{dd}. The third approach, i.e. error resilient, scales the V_{dd} more aggressively while accepting that errors might occur on certain chips. This approach has been proven to enable significant energy savings while maintaining excellent performance for wireless communication [6], [7].
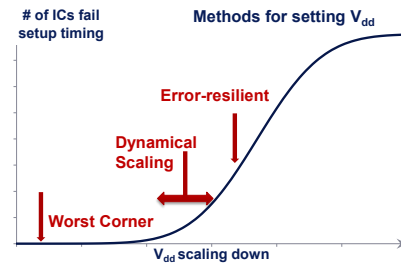

Fig. 1. Various methods on scaling down the V_{dd} for guardband margin reduction.

In this paper, we present the results of our investigation to apply the same error resilient techniques in MaMi. Indeed, in MaMi communication systems specifically, we show that the perceived performance will hardly be affected by sparse processing failures, while the power consumption can be considerably reduced as error resilient hardware is utilized. In that view, we assess antenna outage impacts and propose damage control strategies.

II. SYSTEM DESCRIPTION

In a MaMi system, the base station (BS) is equipped with M antennas and serves K single-stream users simultaneously, each equipped with a single antenna. Unless otherwise specified, we select $M = 100$ and $K = 10$ as typical values in this work, and focus on the TDD option of LTE.

Fig. 2 illustrates the BS architecture of a MaMi system. The BS consists of central digital modem functionality, per-antenna processing including (I)FFT operations for OFDM (de)modulation, digital front-end (DFE), analog front-end (AFE) and power amplifier (PA). The signal processing complexity and power consumption of the digital modem scale linearly with K, while the (I)FFT, DFE and AFE complexity scales linearly with M. The MaMi digital processing complexity [2] is summarized by billion complex floating-point arithmetic operations per second in Table I. The data transfer overhead is included. For a typical MaMi system, the digital processing effort is dominated by the per-antenna functionality (mainly FFT and DFE filtering operations). This is due to its linear dependency to the massive BS antenna number M.
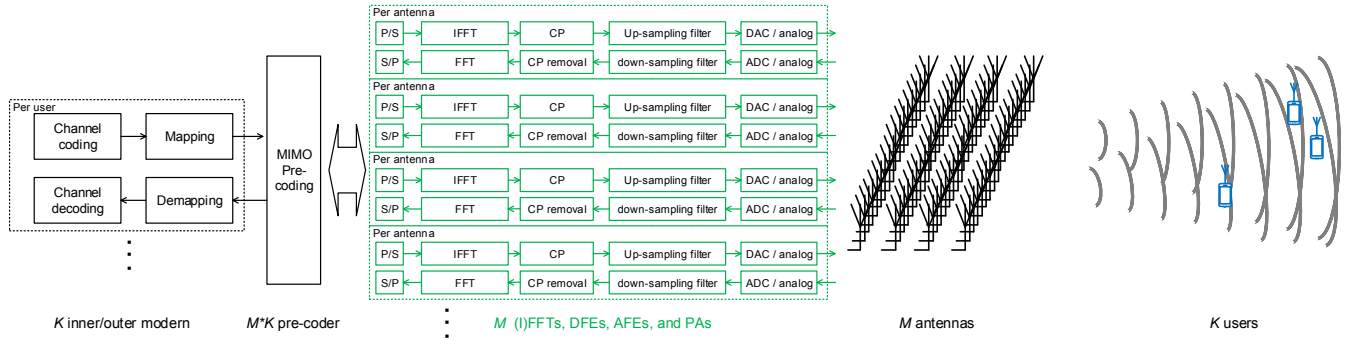
Fig. 2. System model of the MaMi system. Each BS is equipped with M antennas and serves K users. Typically, $M = 100$ and $K = 10$.

For a typical MaMi system, the digital processing effort (see Table I) is dominated by the per-antenna functionality (mainly FFT and DFE filtering operations) [2]. Therefore in order to minimize the area cost and the energy budget of the BS, our work focuses on demonstrating the possibilities of accepting digital hardware errors in the DFE (incl. FFT). In this paper, we specifically focus on the DL, which is typically carrying most of the data and thus power dominant, yet very similar UL DFE essentially could benefit from the same complexity reduction.

TABLE I
COMPLEXITY OF DIGITAL COMPONENTS FOR 100x10 MAMI, WITH 20 MHZ BANDWIDTH, 3 BPS/HZ (16-QAM, CODING RATE 3/4)

Subcomponent	Downlink (DL) [GOPS]	Uplink (UL) [GOPS]	Training [GOPS]
Inner modern	175	520	290
Outer Modern	7	40	0
DFE incl. (I)FFT	920	920	920

III. DIGITAL HARDWARE ERROR EFFECTS

For the MaMi system, the digital hardware errors in (I)FFT & DFE introduced by silicon unreliability and by adventurous design methodologies result in incorrect bit results during signal processing. This can be regarded as digital distortion noise. We characterize the quality of the signal as the signal to digital distortion ratio (SDDR):

$$SDDR = 10 \cdot log \frac{\sigma_s{}^2}{\sigma_d{}^2} \quad (1)$$

where $\sigma_s{}^2$ and $\sigma_d{}^2$ are the powers of error-free DFE output, and noise power of digital distortion due to circuit unreliability, respectively. Voltage-overscaling error (temporary and local error), and antenna outage (hard and full antenna error) are discussed.

A. Voltage over-scaling (VOS)

For digital circuits, the dynamic power consumption scales with $V_{dd}{}^2$, where V_{dd} is the supply voltage. Therefore, digital circuit designs usually reduce V_{dd} (voltage over-scaling, or VOS) for power saving. This is regarded as error-free power saving as long as the signal setup timing constraint is met [8]. However, the critical (minimum) V_{dd} that guarantees setup-timing closure cannot be determined at the design-time due to PVT variabilities and aging effects. Consequently, digital circuit designers face the risk of introducing hardware errors with VOS: for logic components, the signal from the longest propagation paths are miscaptured [9]; for memory components, this leads to incorrect write/read data/address or data loss [10].

The arithmetic noise tolerance [7], [11]–[13] reduces the power consumption of digital signal processors by gracefully sacrificing

the signal to noise ratio (SNR), admitting that a certain amount of errors occur. Numerous designs [6], [7], [11]–[14] are proposed to reduce the hardware errors at a given power budget. For instance, Fig. 3 provides an example of energy saving in 65nm COMS FIR filter brought by VOS, at the cost of SNR degradation. In addition, [11] utilizes reduced precision redundancy to reduce the power consumption by 40% on a digital FIR filter at the cost of slightly degrading the 23 dB SNR signal into 22 dB, and by 35% for a 64-point FFT when lowering the SNR from 55.5 dB to 55 dB.
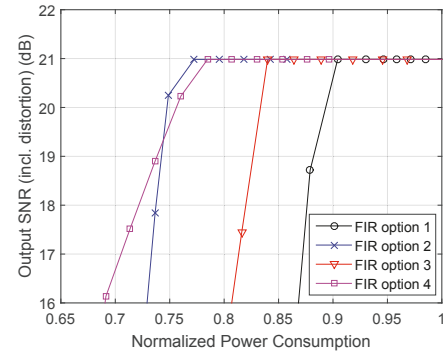

Fig. 3. For various FIR filters, the power consumption is reduced due to VOS, at the cost of signal quality degradation (SDDR increase). [15]

In conclusion, up to 40% power can be saved if the aforementioned error resilient techniques are applied, as the cost of potential sparse antenna processing distortion. The SDDR depends on the operating V_{dd}, the process variability and the environment temperature. This means that even with the same design, different (I)FFT & DFEs of the MaMi might exhibit vastly different SDDR behavior. Section IV discusses the MaMi system-level effects of the VOS error and assures that, circuit degradation on a small portion of antennas can be absorbed in the MaMi system.

B. Antenna outage

Another hardware failure scenario for the DFE is antenna outage (antenna is completely non-operational). This happens when the power supply systems are broken, or when a circuit controlling signal is corrupted, e.g. failure to wake-up the digital circuit.

In an antenna outage scenario, the DFE output is stuck at a fixed value, which is assumed to be the maximum value (DFE output Y = maximum). The SDDR of the outage antenna is $-\infty$, as the signals from the victim antennas are completely lost. This model is regarded as one of the most pessimistic hardware failure. Note that the $-\infty$ SDDR does not imply infinite noise to the whole system, as only

the victim antennas are affected and their PA power are normalized among all antennas. Therefore, single antenna outage should not fail the system entirely.

IV. RANDOM ANTENNA ERROR IMPACT ASSESSMENT

Consider a TDD MaMi system in DL with $M = 100$ and $K = 10$, where the channel estimation and the minimum mean square error (MMSE) MIMO pre-coding are free from digital hardware errors. We simulate the performance over a Rayleigh 20-tap i.i.d. channels. The system is OFDM-based according to LTE parameters, i.e., 1200 loaded subcarriers in a 20 MHz band. The channel is estimated through uplink pilots associated to the different user equipment (UE) in a round robin fashion, i.e., one pilot every 10 subcarriers for a given UE. SNR is defined based on a total transmit power normalized to 0 dB per user. The emitted power is normalized for each antenna. The simulations do not apply error correction coding (ECC), except for Fig. 5(c) where we study the effect of coding on digital hardware errors.

We investigate the effect of local PVT variation and semiconductor aging. This effect results in VOS hardware errors for a portion of the antenna IFFTs & DFEs, while no digital hardware errors for the remaining antennas. The system bit error rate (BER) degradation is illustrated in Fig. 4(a). The BER performance drops slightly as more antennas are affected. Nevertheless, the degradation is small even with 50% antennas affected. Fig. 4(b) exhibits a larger VOS distortion

noise as designers further exploit the design margin. For a target BER of 10^{-4}, MaMi with 20% antenna failing only requires a channel SNR of -7.4 dB, as opposed to -8.4 dB for error-free case. This shows that, the MaMi system still operates even if a noticeable amount of antennas suffer from digital hardware errors. In Fig. 4(c), the MaMi BER when applying the most pessimistic antenna outage model is shown. For victim antennas, the useful signals are completely lost and a constant value is output from the DFE and emitted by the PA. This corresponds to a infinitely small SDDR for those victim antennas. The resulting BER performance shows larger SNR degradation for the same BER target. Nevertheless, the MaMi system can still cope with the antenna outage error thanks to the redundancy of antennas in the BS, at least for a failure rate up to 10%.

Fig. 5(a) displays that for the 100x10 MaMi, 10% antenna outage leads to slightly more BER degradation for QPSK, comparing with BPSK. This is due to the larger error margin for simpler modulation scheme. For the more sensitive 16-QAM modulation scheme, 10% antenna outage will lead to a huge degradation in DL BER. This implies that for communication systems where channel SNR is worse and simple modulation schemes are used, the reliability requirement of the antennas can be relaxed, to simplify the (I)FFT & DFE design and to reduce the power consumption budget.

The BS antenna redundancy is reduced if the load of the MaMi system increases (the number K of served users or streams is increased). In this scenario, the tolerance for antenna outage is decreased,

(a) slight digital error, i.e. SDDR = 10 dB   (b) extensive digital error, i.e. SDDR = 0 dB   (c) antenna outage (DFE output stuck at a fixed value), i.e. SDDR = -∞
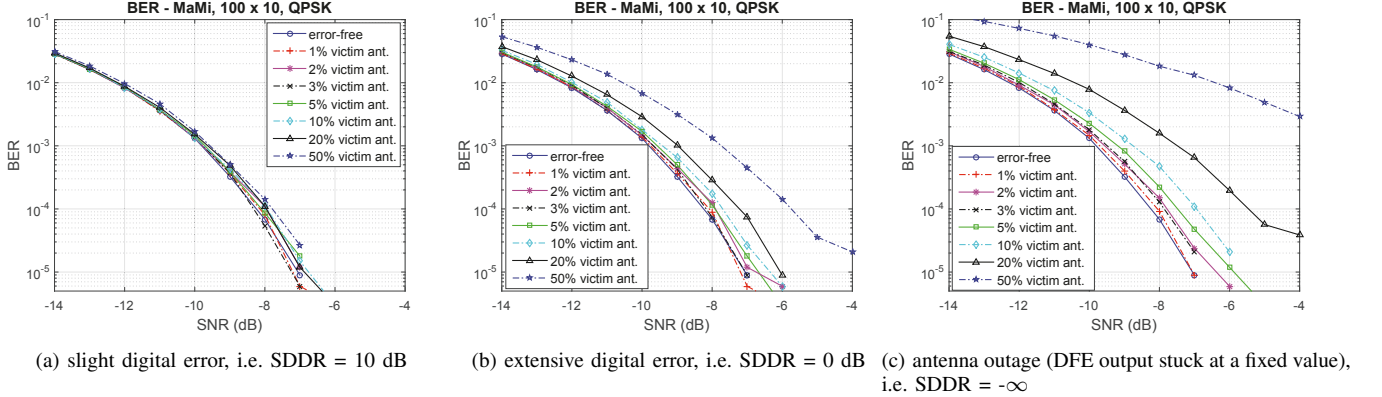
Fig. 4. MaMi system bit BER vs. channel SNR. Randomly chosen victim antennas suffer from VOS digital hardware errors / outage errors. The other antennas are free from this error. (a), (b) and (c) represents different error power (SDDR).

(a) modulation effects   (b) load effects   (c) coding effects.
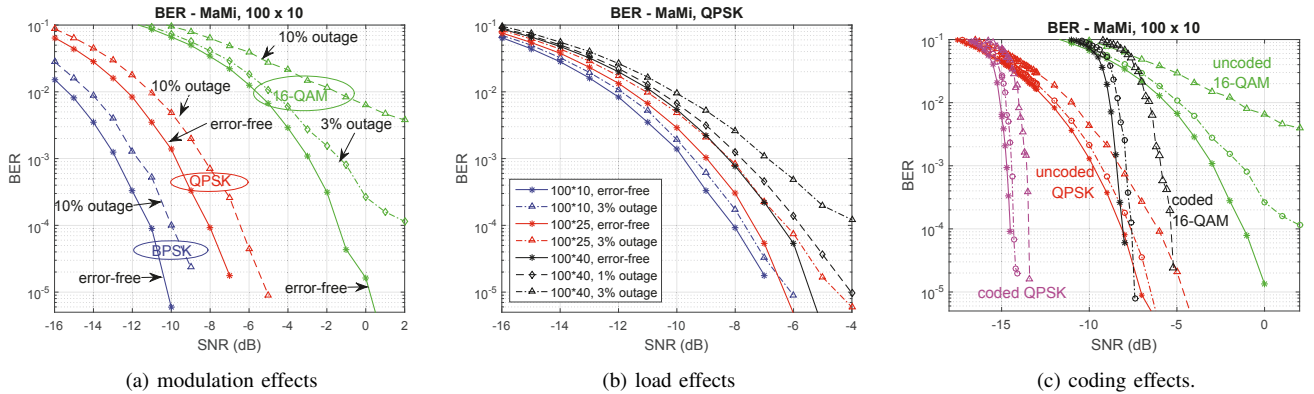
Fig. 5. MaMi system with random antenna outage error for (a) uncoded BPSK, QPSK and 16-QAM; (b) various loads (100x10, 100x25, 100x40); (c) uncoded and coded (3/4 soft LDPC) QPSK, and uncoded and coded (3/4 soft LDPC) 16-QAM. The legend denotes: i) error-free (star shapes), ii) 3% victim antennas (circle shape), and iii) 10% victim antennas (triangle shape).

compared to systems with small K (Fig. 5(b)). Nevertheless, For MaMi systems where $M >> K$, the amount of antenna redundancy is sufficient to provide opportunities for antenna un-reliability.

So far uncoded results were presented. However, errors in MaMi systems can be mostly corrected by error correction codes, e.g. convolutional codes and LDPC codes. Fig. 5(c) shows the BER improvement when 3/4 soft decoded LDPC code is utilized in the MaMi system. At the targeted BER of 10^{-4}, the SNR is 6 dB lower for the coded QPSK, compared with an uncoded case. For this BER, the SNR difference when considering antenna outage is smaller for the coded MaMi system, compared to the uncoded one, although a limited degradation always remains.

V. CONTROLLED ANTENNA OUTAGE

According to Section III A, VOS can bring up to 40% power saving, at the risk of failures for few antennas. The simulation results from Section IV provide the performance when no error detection monitors are equipped and thus the MaMi system operates as an open loop, regardless of the hardware error situation. In this situation, the MaMi manages to sustain system performance even if several antennas are non-operational (outage) due to aggressive VOS or failure.

Moreover, in order to improve the robustness of the system towards hardware errors, we apply techniques to first detect hardware errors, and next either correct, or if this is not possible circumvent the defective hardware. Indeed the distortion originating from digital circuit failures fundamentally differs from the random noise introduced in communication channels. While process variation may feature continuous random distributions, their effect typically translates in discrete error events. Dedicated monitoring circuit can be established to detect these errors and thus these erroneous bits can be labeled unreliable and potentially be corrected. Moreover, eventually some circuit errors get too large or systematic, measures at system-level can be taken to discard this hardware and increase the overall robustness.

If digital hardware designs provide monitors [16]–[19] for each (I)FFT & DFE, we can deploy the MaMi system as a closed loop. One possible counter-measure is to disable the victim antennas, and re-compute the channel estimation (and hence the precoding) for error-free antennas only. This method is equivalent to operating with a reduced number of error-free BS antennas M. Fig. 6 shows the BER performance when the victim antennas are taken out. On the one hand, the BER is worse than Fig. 4(a) and Fig. 4(b), where antennas are affected with moderate digital hardware errors and the signal from those victim antennas are still exploited for communication. On the other hand, the BER in Fig. 6 is superior comparing with Fig. 4(c), on which the antennas are heavily affected by digital hardware errors. This shows that, by detecting the degree of antennas failing (noise power), designers have the option to whether to exploit the victim antennas or to discard them, in order to maximize system performance.

Another error detection strategy is to periodically check the antenna functionality by putting one antenna in testing-mode at a time (Fig. 7). During testing mode, per antenna DSP are supplied with testing inputs. The outputs are compared with pre-computed data. If the results are vastly different, the antenna is detected erroneous, and thus the V_{dd} is increased to reduce error and hence ensure performance. Since suppressing one (1% for MaMi with 100 antenna) antenna into testing-mode would not introduce huge degradation (see Fig. 6), the testing can even be performed on-the-fly during data transmission. This enables timely fine-grained V_{dd} adjustment, which maximize power saving. If however the antenna are permanently damaged and
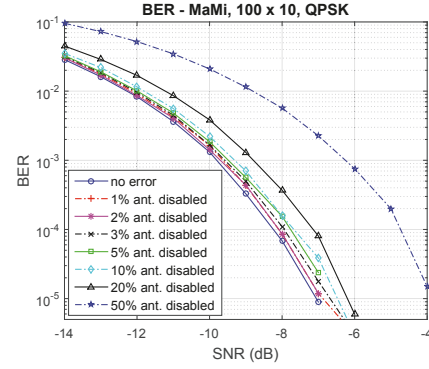

Fig. 6. MaMi system performance when the victim antennas are discarded and the channel estimation and DL is carried out by the remaining error-free antennas.
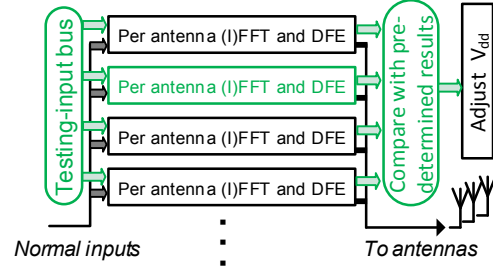

Fig. 7. Periodically checking DFE functionality, and adjusting V_{dd} for power saving. The second antenna is in testing-mode.

thus cannot recover from increasing V_{dd}, the antenna will then be labeled as defected.

VI. CONCLUSION

This paper examines the opportunity of using error-prone digital signal processing components in MaMi systems, and proposes a strategy to maximize power saving while still offering robust operation. The (I)FFTs & DFEs in MaMi are the most critical digital components in terms of area and power consumption as they scale linearly with the massive BS antenna count M. Hardware errors in a number of antenna (I)FFTs & DFEs can be absorbed by the MaMi system thanks to the redundancy coming from the large antenna number. The MaMi system exhibits error resilience even for the worst-case antenna outage scenario.

When the hardware error distortion power is low, e.g. lower than 0 dB , the MaMi system should continue using the erroneous signal as the errors can be corrected at the system-level by other redundant antennas. We equip the system with on-chip monitors, that can detect hardware errors on-the-fly, and propose to discard the error-prone antennas when their SDDR is large, e.g. antenna outage or components suffering from severe aging.

This provides opportunities for the digital hardware designers to embrace cost-efficient and reduced-power digital components at the expense of sacrificing individual antenna reliability, yet maintaining overall systems performance. Up to 40% power can be reduced for the considered digital processing components.

REFERENCES

[1] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 186–195, 2014.

[2] C. Desset, B. Debaillie, and F. Louagie, "Modeling the hardware power consumption of large scale antenna systems," in *Green Communications (OnlineGreencomm), 2014 IEEE Online Conference on*, nov 2014, pp. 1–6.

[3] C. Desset and L. V. der Perre, "Validation of low-accuracy quantization in massive MIMO and constellation EVM analysis," in *Networks and Communications (EuCNC), 2015 European Conference on*, jun 2015, pp. 21–25.

[4] U. Gustavsson, C. Sanchéz-Perez, T. Eriksson, F. Athley, G. Durisi, P. Landin, K. Hausmair, C. Fager, and L. Svensson, "On the impact of hardware impairments on massive MIMO," in *Globecom Workshops (GC Wkshps), 2014*. IEEE, 2014, pp. 294–300.

[5] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 7112–7139, 2014.

[6] Y. Huang, M. Li, C. Li, P. Debacker, and L. Van der Perre, "Computation-skip Error Mitigation Scheme for Power Supply Voltage Scaling in Recursive Applications," *Journal of Signal Processing Systems*, vol. 84, no. 3, pp. 413–424, sep 2016.

[7] R. Hegde and N. R. Shanbhag, "A voltage overscaled low-power digital filter IC," *Solid-State Circuits, IEEE Journal of*, vol. 39, no. 2, pp. 388–391, 2004.

[8] J. P. Kulkarni, C. Tokunaga, P. Aseron, T. Nguyen, C. Augustine, J. Tschanz, and V. De, "4.7 A 409GOPS/W adaptive and resilient domino register file in 22nm tri-gate CMOS featuring in-situ timing margin and error detection for tolerance to within-die variation, voltage droop, temperature and aging," in *ISSCC, 2015 IEEE*, feb 2015, pp. 1–3.

[9] J. Han and M. Orshansky, "Approximate computing: An emerging paradigm for energy-efficient design," in *Test Symposium (ETS), 2013 18th IEEE European*, may 2013, pp. 1–6.

[10] E. Karl, D. Sylvester, and D. Blaauw, "Timing error correction techniques for voltage-scalable on-chip memories," in *Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on*. IEEE, 2005, pp. 3563–3566.

[11] B. Shim, S. R. Sridhara, and N. R. Shanbhag, "Reliable low-power digital signal processing via reduced precision redundancy," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 12, no. 5, pp. 497–510, may 2004.

[12] S. Narayanan, J. Sartori, R. Kumar, and D. L. Jones, "Scalable stochastic processors," in *Proceedings of the Conference on Design, Automation and Test in Europe*. European Design and Automation Association, 2010, pp. 335–338.

[13] G. Karakonstantis, D. Mohapatra, and K. Roy, "System level DSP synthesis using voltage overscaling, unequal error protection & adaptive quality tuning." in *2009 IEEE Workshop on Signal Processing Systems (SiPS)*, 2009, pp. 133–138.

[14] Y. Huang, M. Li, C. Li, P. Debacker, and L. Van der Perre, "Computation-skip error resilient scheme for recursive CORDIC," in *2014 IEEE Workshop on Signal Processing Systems (SiPS)*. IEEE, oct 2014, pp. 1–6.

[15] Y. Liu, T. Zhang, and K. K. Parhi, "Computation error analysis in digital signal processing systems with overscaled supply voltage," *IEEE Transactions on Very Large Scale Integration(VLSI) Systems*, vol. 18, no. 1-4, pp. 517–526, 2010.

[16] D. Ernst, N. S. Kim, S. Das, S. Pant, R. Rao, T. Pham, C. Ziesler, D. Blaauw, T. Austin, K. Flautner, and T. Mudge, "Razor: a low-power pipeline based on circuit-level timing speculation," in *Microarchitecture, 2003. MICRO-36. IEEE/ACM International Symposium on*, dec 2003, pp. 7–18.

[17] K. A. Bowman, J. W. Tschanz, S. L. L. Lu, P. A. Aseron, M. M. Khellah, A. Raychowdhury, B. M. Geuskens, C. Tokunaga, C. B. Wilkerson, T. Karnik, and V. K. De, "A 45 nm Resilient Microprocessor Core for Dynamic Variation Tolerance," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 1, pp. 194–208, jan 2011.

[18] D. Bull, S. Das, K. Shivashankar, G. S. Dasika, K. Flautner, and D. Blaauw, "A power-efficient 32 bit ARM processor using timing-error detection and correction for transient-error tolerance and adaptation to PVT variation," *Solid-State Circuits, IEEE Journal of*, vol. 46, no. 1, pp. 18–31, 2011.

[19] M. Fojtik, D. Fick, Y. Kim, N. Pinckney, D. Harris, D. Blaauw, and D. Sylvester, "Bubble Razor: An architecture-independent approach to timing-error detection and correction," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2012 IEEE International*. IEEE, 2012, pp. 488–490.