

A Demo of FPGA Aggressive Voltage Downscaling: Power and Reliability Tradeoffs

Behzad Salami^{*†}, Osman Unsal^{*}, and Adrian Cristal^{*†}

^{*}Barcelona Supercomputing Center (BSC), Barcelona, Spain. Emails: {behzad.salami, osman.unsal, and adrian.cristal}@bsc.es

[†]Universitat Politcnica de Catalunya (UPC), Barcelona, Spain.

The power consumption of digital circuits, e.g., Field Programmable Gate Arrays (FPGAs), is directly related to their operating supply voltages. On the other hand, usually, chip vendors introduce a conservative voltage guardband below the standard nominal level to ensure the correct functionality of the design in worst-case process and environmental scenarios. For instance, this voltage guardband is empirically measured to be 12%, 20%, and 16% of the nominal level in commercial CPUs [1], Graphics Processing Units (GPUs) [2], and Dynamic RAMs (DRAMs) [3], respectively. However, in many real-world applications, this guardband is extremely conservative and eliminating it can result in significant power savings without any overhead. Motivated by these studies, we aim to extend the undervolting technique to commercial FPGAs. Toward this goal, we will practically demonstrate the voltage guardband for a representative Xilinx FPGA¹, with a preliminary concentration on on-chip memories, or Block RAMs (BRAMs).

Our experimental results show the voltage guardband to be 39% of the nominal level ($V_{nom} = 1V$, $V_{min} = 0.61V$), which in turn, directly delivers an order of magnitude BRAM power savings. Further undervolting below $V_{min} = 0.61V$ delivers more power savings, up to 40% in our case; however, causes fault generation in some locations of some of BRAMs. These faults are the consequence of timing violations, since the cir-

¹We have undervolted several Xilinx platforms and observed very similar results. This paper demonstrates it on a representative platform, VC707.

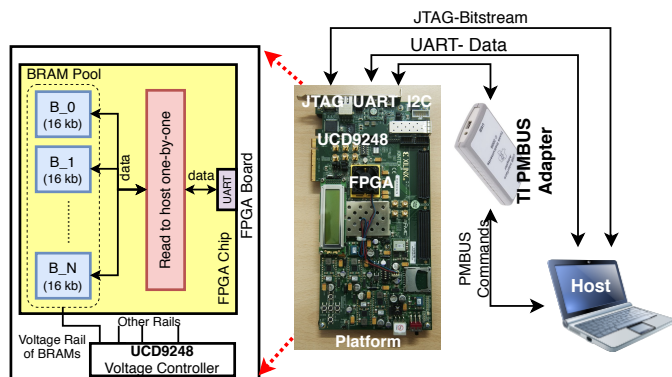
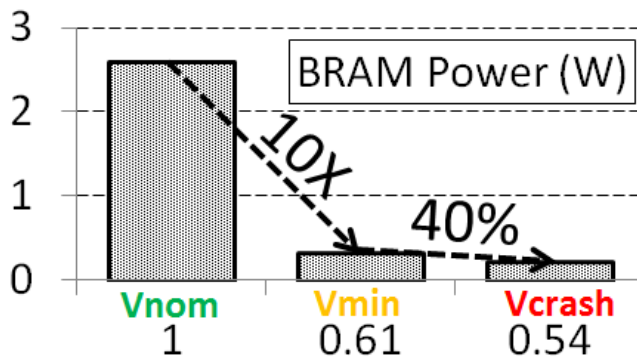
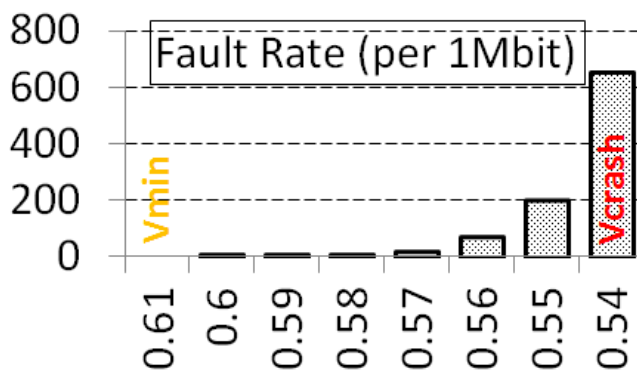


Fig. 1: Experimental Demo Setup to Study FPGA BRAMs Undervolting.



(a) BRAM Power.



(b) BRAM Fault Rate.

Fig. 2: Power and Reliability Trade-off over Voltage Downscaling in FPGA BRAMs, undervolting from $V_{min} = 0.61V$ to $V_{crash} = 0.54V$. (y-axis is the V_{CCBRAM} .)

cuit delay increases by further undervolting. Note that simultaneously downscaling the frequency is a promising approach to prevent the generation of these faults [4]; however, it can limit the energy reduction achievement. Alternatively, our aim is to understand the behavior of these faults, through which customized and low-overhead fault mitigation techniques can be deployed to achieve power saving gains.

The setup of the demo is shown in Fig. 1, where we demonstrate the voltage guardband and also more aggressively-reduced voltage regions where BRAMs experience faults. Our FPGA design includes raw Read/Write accesses to BRAMs,

TABLE I: Summary of Fault Characterization in FPGA-Based BRAMs and comparing with a recent voltage downscaling study on modern DRAMs, i.e., DDR-3.

Key Finding	FPGA BRAMs [our work]	DRAM (DDR-3) [3]
Below V_{nom} and above V_{min} , no observable fault.	$V_{nom} = 1V, V_{min} = 0.61V$	$V_{nom} = 1.5V, V_{min} = 1.25V$
Type of faults: Permanent	Stuck-at-0	No Information .
Impact of data pattern.	Fault ratio is directly related to number of '1' logic.	Fault ratio is independent of the pattern.
Below V_{min} , fault ratio <i>exponentially</i> increases.	Valid, chip-dependent, on average [0, ~0.06%]	Valid, vendor-dependent, on average [0, ~20%]
Faults tends to cluster in certain locations.	Certain BRAMs	Certain Banks
Fault Inclusion Property (FIP)	Follows this property.	No Information.

while their supply voltage, i.e., V_{CCBRAM} is controlled in the host through the Power Management Bus (PMBus) interface [5]. The on-board voltage regulator with the part number of UCD9248 has the responsibility to handle these PMBus commands, and set appropriate voltage to different components, e.g., BRAMs. Note that other FPGA components, e.g., Look-Up Tables (LUTs), Digital Signal Processors (DSPs), operate at their default nominal voltage levels. Through experiments on this setup, the overall power and reliability trade-off is summarized in Fig. 2, when V_{CCBRAM} is downscaled from the nominal level $V_{nom} = 1V$ to the minimum level that the FPGA practically operates, $V_{crash} = 0.54V$. As can be seen, BRAMs start experiencing faults in regions below $V_{min} = 0.61V$ with an exponentially increasing behavior up to 653 faults per 1Mbit $\sim 0.06\%$. Major observed properties of these faults are summarized as follows:

- There is significant variability of fault rate among different BRAMs, which is the consequence of the inherent process variation. Through our experiments, we observed that more than 38.9% of BRAMs never experience faults. Also, among BRAMs the maximum, minimum, and average fault rate are 2.84%, 0%, and 0.06%.
- Aggressive voltage downscaling causes permanent '1' to '0' bit flips, i.e., stuck-at-0. In other words, first, the faults locations and rate do not considerable change over the time. Second, by experimentally evaluating different data patterns, we observed that the fault rate directly depends on the number of "1" bits since a vast majority of generated faults are '1' to '0' bit-flips. Due to this observation on the behavior of faults, simplified and low-overhead fault mitigation techniques can be potentially deployed.
- More than 90% of these faults are single-bit, and a further 7% are double-bit faults. Due to this observation on the behavior of faults, the built-in Error Correction Code (ECC) of BRAMs can be effective to mitigate these faults. Note that the built-in ECC of BRAMs has the type of Single-Error Correction and Double-Error Detection (SECDED) capability [6], potentially with a good efficiency to mitigate BRAMs faults in low-voltage regions.
- Faulty bitcells in a certain voltage stay faulty in lower voltages, as well, and potentially, expand to other bitcells. This property is called Fault Inclusion Property (FIP), and is observed in CPU caches [7], as well. This paper experimentally confirms that FIP exists in FPGAs, under

aggressive low-voltage operations. FIP can be potentially used to build efficient fault mitigation techniques.

SUMMARY

This paper demonstrated the voltage downscaling approach for commercial FPGAs, as an effective solution for improving the energy efficiency. With a concentration on on-chip BRAMs, we evaluated the subsequent power and fault rate trade-off. We experimentally observed that an extremely conservative voltage guardband exists below the standard nominal level. Eliminating this voltage gap delivers more than an order of magnitude power savings, without any performance or reliability overhead. Further undervolting delivers more power savings up to a further 40%; however, with the cost of faults generation. We presented comprehensive experimental fault characterization. Our experimental observations such as significant fault rate variability among BRAMs, can provide an opportunity to optimize power-reliability trade-off in aggressively low-voltage regimes, for applications implemented onto FPGAs. We summarize our observations and findings in Table. I. Also, we compare our observations with a recent characterization work on DDR-3 [3], mostly in the behavioral-level. Although, there is a technological difference between them, i.e., BRAMs are SRAM-based while DDR-3 are DRAM-based, the comparison highlights their significant similar fault behavior under low-voltage operations.

ACKNOWLEDGMENT

We thank Pradip Bose, Alper Buyuktosunoglu, and Augusto Vega from IBM Watson for their contribution to this work. The research leading to these results has received funding from the European Union's Horizon 2020 Programme under the LEGaTO Project (www.legato-project.eu), grant agreement n° 780681.

REFERENCES

- [1] A. Bacha, et al. "Dynamic reduction of voltage margins by leveraging on-chip ECC in Itanium II processors", in *ISCA*, 2013.
- [2] J. Leng, et al. "Safe limits on voltage reduction efficiency in GPUs: a direct measurement approach", in *MICRO*, 2015.
- [3] K. K. Chang, et al. "Understanding reduced-voltage operation in modern DRAM devices: Experimental characterization, analysis, and mechanisms", in *Measurement and Analysis of Computing Systems*, 2017.
- [4] Nunez-Yanez, et al. "Energy optimization in commercial FPGAs with voltage, frequency and logic scaling", in *IEEE TC*, 2016.
- [5] "Power Management Bus (PMBUS)." <http://pmbus.org>
- [6] Xilinx Co. https://www.xilinx.com/support/documentation/user_guides/ug473_7Series_Memory_Resources.pdf
- [7] M. Gottscho, et al. "Power/capacity scaling: Energy savings with simple fault-tolerant caches", in *DAC*, 2014.