Fault Characterization Through FPGAs Undervolting

Behzad Salami, Osman S. Unsal, and Adrian Cristal Kestelman

Presented by Alberto González

Underscaling the supply voltage *below the nominal level*:

- **Power/Energy Efficiency**: Reduces quadratic ally dynamic and linearly static power.
- **Reliability**: Increases the circuit delay and in turn, causes timing faults.

**Aggressive Undervolting is not DVFS!**
Motivation

Contribution of FPGAs in large data centers is growing, expected to be in 30% of datacenter servers by 2020 (Top500 news).

- In comparison to ASICs, energy efficiency of FPGAs is a serious concern.
- Nominal voltage reduction of FPGAs is naturally applied for different generations.

Our Aim:
Undervolting FPGAs below the nominal level to achieve energy efficiency.

Subsequent Study:
How is the reliability affected through FPGAs Undervolting?
Voltage Scaling Capability in Xilinx

Voltage Distribution on Xilinx Platforms

VC707

Voltage Regulator
- Power Management Bus (PMBus).
- Hardwired to the host.

Evaluated Xilinx Platforms

VC707: performance-efficient design
KC705: power-efficient design
Experimental Methodology

Detailed study on FPGA BRAMs, which are a set of bitcells in the row-column format.

Experimental Methodology:
1. **HW**: Transfer content of BRAMs to the host.
2. **SW**: Analyze data, and adjust voltage of BRAMs.

Operating frequency is set to the maximum, i.e., ~500mhz.
1. Voltage $V_{nom} = 1$V.
2. $V_{min}$ & $V_{crash}$ are slightly different.
3. More than 10X energy efficiency.
4. Exponential fault rate increase.
5. VC707 experiences relatively more fault rate.

**Voltage Guardband:**

1. **DRAM** - Multiple Vendors [Sigmetrics2017]: 16%
2. **GPU** - NVidia [Micro2015]: 20%
3. **CPU** - ItaniumII [ISCA2013]: 12%
4. **FPGA** - Xilinx [our work- FPL2018]: 39%
Fault Variability between BRAMs

- BRAMs clustering using K-Mean clustering.
- Majority of BRAMs are low-vulnerable.
- ~36% of BRAMs never experience faults.
- Fully non-uniform fault distribution.

VC707

<table>
<thead>
<tr>
<th>%BRAMs</th>
<th>Average Fault Ratio (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.8%</td>
<td>0.86%</td>
</tr>
<tr>
<td>9.4%</td>
<td>0.24%</td>
</tr>
<tr>
<td>88.6%</td>
<td>0.02%</td>
</tr>
</tbody>
</table>

KC705

<table>
<thead>
<tr>
<th>%BRAMs</th>
<th>Average Fault Ratio (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9%</td>
<td>0.74%</td>
</tr>
<tr>
<td>5.7%</td>
<td>0.17%</td>
</tr>
<tr>
<td>93.4%</td>
<td>0.01%</td>
</tr>
</tbody>
</table>

VCCBRAM = Vcrash

* Different scales in y-axis * *Pattern = 18’h3FFFF *
Thanks!
Contact:
Behzad Salami
behzad.salami@bsc.es
Backup
Outline

• Background
  – What does Undervolting mean?
  – Motivation: FPGAs Undervolting

• First Contribution: Undervolting Xilinx FPGAs
  • Experimental Methodology
  • Overall Power and Reliability Trade-off

• Second Contribution: Fault Characterization
  • Fault Variability
  • Fault Types
  • Impact of the Environmental Temperature

• Related Work

• Summary and Future Works
Permanent ‘1’ to ‘0’ bit-flips

Permanent:

- There is no considerable change on the rate and location of faults over time.
- Validated by repeating experiments for 100 times.

‘1’ to ‘0’ bit flips:

- Experimentally proved that the majority of faults are ‘1’ to ‘0’ bit flips.
- No matter for ‘0’ and ‘1’ permutations.

Conclusion:

Permanent ‘1’ to ‘0’ bit-flips can be translated as stuck-at-0, at a certain voltage, temperature, etc.
Related Works of Undervolting

• Simulation-based: (Lack of precise information of the real hardware.)

- Thundervolt: ASIC-based DNN (DAC2018)
- Minerva: ASIC-based DNN (Micro2016)
- Bravo: CPU (HPCA2017)

• Real Commercial/Customized Devices
  - CPUs: Itanium II (ISCA2013), X86 (IOLTS2017)
  - Multicore CPU: ARM (HPCA2017, ISPASS2018)
  - GPUs: NVidia (Micro2015)
  - DRAMs: Multiple Brands (Sigmetrics2017)
  - SRAMs: Customized (ISQED2017)
  - FPGAs: Xilinx (Our Work - FPL2018)

Focus of Previous Works:

(1) Covered in our work for FPGAs

• Voltage Guardband
• Fault Characterization at Critical Region
• Impact of Environmental Conditions

(2) Not-covered in our work on FPGAs (Future Work)

• Dynamic Vmin Prediction
• Fault Mitigation at Critical Region
• Application Profiling
Future of FPGA Undervolting needs more advanced voltage designs, by vendors:

1. Many FPGA platforms, e.g., Zynq are not equipped with voltage scaling capability.
2. There is no standard about the voltage distribution among platform components.
3. Voltage regulators are hardwired to the host through PMBus interface.
4. In many cases, several components on the FPGA platform share a single voltage rail.
5. Vendors set unnecessarily conservative voltage guardbands that increase the energy.
6. There is no publicly-available circuit-level information of FPGAs.
Fault Characterization at CRITICAL Region

Environmental Temperature

- **Methodology:** Adjusting environmental temperature, monitoring on-board temperature via PMBus.

- **Experimental Observation:**
  - At higher temperatures, fault rate is significantly reduced.
  - The rate of this reduction is highly platform-dependent (VC707 > KC705).

- **Inverse Temperature Dependency (ITD):**
  - For nano-scale technologies, under ultra low-voltage operations, the circuit delay reduces at higher temperatures since supply voltage approaches the threshold voltage.

* y-axis: VCCBRAM (V), y-axis: fault rate (per 1Mbit) *
### Summary

- We **experimentally** showed how Xilinx FPGAs work under aggressive low-voltage operations.
- There is a **conservative voltage guardband** below the nominal level.
- BRAMs **power** is significantly reduced through Undervolting; however, **reliability** degrades below min safe voltage.
- We **characterized** the behavior of Undervolting faults at the critical region.

### Future Works

- **Dynamic Vmin scaling**, adapted by frequency and temperature.
- More advanced designs, where other components such as **I/O, DDR, DSP** are undervolted.
- Efficient Fault Mitigation Techniques.
- **Profiling applications** such as Deep Neural Networks (DNNs), among others.
- Extending Undervolting for other commercial FPGAs such as **Intel/Altera**.