Hard Data on Soft Errors

A Global-scale Survey of GPGPU Memory Soft Error Rates

Imran Haque
Department of Computer Science
Stanford University

http://cs.stanford.edu/people/ihaque
http://folding.stanford.edu
ihaque@cs.stanford.edu

Resilience Workshop @ CCGrid 17 May 2010
Motivation

- GPUs originate in error-insensitive consumer graphics
- Neither ECC nor parity on most* graphics memory

- How suitable is the installed base of consumer GPUs (and consumer GPU-derived professional hardware!) for error-sensitive general purpose computing?

* of which, more later
Why would a comp bio group care?

<table>
<thead>
<tr>
<th>CUDA-Enabled Package</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Folding@home</strong> (molecular dynamics)</td>
</tr>
<tr>
<td><strong>OpenMM</strong> (molecular dynamics)</td>
</tr>
<tr>
<td><strong>PAPER</strong> (3-D chemical similarity)</td>
</tr>
<tr>
<td><strong>SIML</strong> (1-D chemical similarity)</td>
</tr>
</tbody>
</table>

We’ve written a lot of CUDA-enabled software, and we run it on a lot of GPUs.
Methodology – MemtestG80

• Custom software, based on Memtest86 for x86 PCs
• Open source (LGPL), available at https://simtk.org/home/memtest

• Variety of test patterns:
  – Constant (ones, zeros, random)
  – Walking ones and zeros (8-bit, 32-bit)
  – Random words (on-GPU parallel PRNG)
  – Modulo-20 pattern sensitivity
  – Novel iterated-LCG integer logic tests
  – Bit fade
MemtestG80 – Validation

- Negative control – verify that it doesn’t throw spurious errors in “known-good” situations
  - Known-good PSUs, machines located in air-conditioned environments.

- 93,000 iterations on 700 MiB on GeForce 8800GTX
- >180,000 iters on 320MiB on each of 8 x Tesla C870

- No errors ever detected.
MemtestG80 – Validation

- Positive control – verify that it does throw errors in situations that generate errors
- Overclocking generates memory errors (violation of timing constraints; loss of signal integrity)
- Tested GeForce 9500GT (memory clock = 400MHz) at 400, 420, 430, 440, 450, 475, 500, 530 MHz
  - 20 iterations for each frequency (only 10 @ 530MHz)
  - Cooled down and reset to 400MHz between tests
MemtestG80 – Validation

Positive control displays pattern sensitivity of memory tests
Methodology – Folding@home

- Expect a low error rate and environment sensitivity, so must sample *many* cards in diverse environments
- Ran for ~7 months over 50,000+ NVIDIA GPUs on Folding@home (>840 TB-hr of testing)
- >97% of data tested 64 MiB RAM, $k=512$ logic LCG
Methodology – Folding@home

- We achieve good sampling over the NVIDIA consumer product line, and a few pro cards as well.
- Sampled similar numbers of stock and (shader) overclocked boards

![Graph showing number of cards sampled with #iters > cutoff]

<table>
<thead>
<tr>
<th>Card Family</th>
<th># cards ≥ 300,000 iter.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumer graphics cards</td>
<td>17648 total</td>
</tr>
<tr>
<td>GeForce GTX</td>
<td>5520</td>
</tr>
<tr>
<td>GeForce 8800</td>
<td>5478</td>
</tr>
<tr>
<td>GeForce 9800/GTS</td>
<td>4923</td>
</tr>
<tr>
<td>GeForce 9600</td>
<td>1516</td>
</tr>
<tr>
<td>Other Desktop GeForce</td>
<td>181</td>
</tr>
<tr>
<td>Mobile GeForce</td>
<td>30</td>
</tr>
<tr>
<td>Professional graphics cards</td>
<td>89 total</td>
</tr>
<tr>
<td>Quadro FX</td>
<td>83</td>
</tr>
<tr>
<td>Quadroplex 2200</td>
<td>6</td>
</tr>
<tr>
<td>Dedicated GPGPU cards</td>
<td>37 total</td>
</tr>
<tr>
<td>Tesla T10</td>
<td>27</td>
</tr>
<tr>
<td>Tesla C1060</td>
<td>10</td>
</tr>
</tbody>
</table>
Results

• We call a failure if any test in a MemtestG80 iteration failed (ignore exact WER)

• Model: each card has its own probability of error (test failure) = $P_f$. Cards are drawn iid from an underlying distribution $P(P_f)$

• What is the distribution of failure probabilities?
Population of failing cards has a mode around $P_f = 2 \times 10^{-5} \approx 4$ failures/week.
GT200 has typical $P_f = 2.2 \times 10^{-6}$ (one-tenth of G80!)

*Both archs. show monotonic decline in zero-error populations.*
Analysis – GeForce vs Tesla

Tesla traces are rougher from poorer sampling, but appear to represent same error distribution as GeForce data.
Analysis – Test Mutual Information

- Consider mutual information between tests as a nonlinear covariance measure.
- Mod-20 test is unique
- Random blocks test is a good logic workout
- Logic tests measure a failure mode distinct from memory tests
What about “Fermi”?  

- NVIDIA’s new Fermi (GF100) architecture adds SECDED ECC (disabled in consumer GeForce line?), GDDR5 memory bus ECC, and L1/L2 caches.

- Does Fermi redesign affect architectural vulnerability (error rate or error type)?
  - G80/GT200 typically failed on Mod-20 test first.

- FAH test does not run (yet) on Fermi; used standalone MemtestG80 w/reporting capabilities.
  - In-house: 1 GeForce GTX 480, 1 Tesla C2050
  - Public: 44 GeForce GTX 470, 43 GeForce GTX 480
Results – Fermi

- **Tesla**: no app-level errors seen, at least one double-bit error reported by ECC

- **GeForce**: most cards exhibited memory errors – observed in-house $P_f = 1.6 \times 10^{-5}$
  - Non-overclocked cards vulnerable to 8-bit walking zeros
  - RAM-overclocked first failed 8- or 32-bit walking zeros
  - Core/shader-overclocked failed random blocks

- Very different vulnerabilities than G80/GT200 – but problems still exist!
Acknowledgments

• Pande lab, Stanford University

• Simbios (NIH Roadmap GM072970)

• NVIDIA

• Folding@home donors
Summary

• Wrote MemtestG80 to test for GPU memory errors.
• Verified proper operation of MemtestG80 with negative and positive control tests.
• Ran MemtestG80 on over 50,000 GPUs, 840+ TB-hr

• 2/3 of tested GPUs exhibit pattern-sensitive soft errors
• Architecture makes a difference: GT200 is much more reliable than G80; GF100 introduces a new set of vulnerabilities

• GT200 Tesla cards on FAH performed similarly to GeForces (but GF100 ECC seems to make a difference on Tesla C20xx)
Conclusions

• Sufficiently high hard error rate (2%) that explicit testing is warranted.

• Some form of ECC appears to be crucial for reliable GPGPU computation.

https://simtk.org/home/memtest

ihaque@cs.stanford.edu