Approximate Storage in Solid-State Memories

Adrian Sampson
Jacob Nelson
Karin Strauss
Luis Ceze

University of Washington
Microsoft Research & UW
University of Washington

MICRO 2013
<table>
<thead>
<tr>
<th>App</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Music</td>
<td>15.9 GB</td>
</tr>
<tr>
<td>Photos &amp; Camera</td>
<td>11.3 GB</td>
</tr>
<tr>
<td>TripAdvisor</td>
<td>549 MB</td>
</tr>
<tr>
<td>Stay.com</td>
<td>535 MB</td>
</tr>
<tr>
<td>Keynote</td>
<td>389 MB</td>
</tr>
<tr>
<td>Podcasts</td>
<td>375 MB</td>
</tr>
<tr>
<td>Evernote</td>
<td></td>
</tr>
</tbody>
</table>
Figure 6: Output degradation for each benchmark using the...the array has experienced, indicating...

(b) Persistent data sets with approximate MLC.

Impact of bit priority.

Impact of ECP budget.

Moving from ECP...

storage provides an opportunity to reduce the frequency of scrubbing.

Approximate computing is an area of research that seeks to optimize the execution of error-tolerant programs using both hard-ware and software techniques. Programmers control the impact of approximate execution using language features, analyses, or program logics. While much prior work has also considered low-overhead techniques for handling errors in memory, this paper focuses on extending the lifetime of precise storage on approximate MLCs. In the first, the bits whose quality was (statistically significantly) improved in writes...the results above assume a retention time of 9 seconds, or about

While the striped code is imperfect, as discussed in Section 3.2, it fares better than the naive code in practice since it lowers the probability of errors in the high-order bits of words. If we are only interested in one intermediate write speed, using that code reduces the mean error rate between 0.1% and 4.2%. The other benchmarks, which see very little error, could...examine two different...ECP...correction on memory that contains approximate MLCs. In the first, the bits of...extension) while incurring no additional error-correction overhead. Moving from ECP...systems trade off...MCs complement these techniques by allowing even faster writes...motherboard and...For the main-memory applications, in which typical retention time is 10 seconds, or about...and...the striped code leads to better output quality on average. For...the highest bits of the value map to the highest bits of the cells. The results given above use the latter encoding, but we also evaluated three intermediate write speeds, using that code reduces the mean number of complete overwrites the array has experienced, indicating...

Prior work on SRAM and DRAM density...error rates between 0.1% and 4.2%. The other benchmarks, which see very little error,...are always stored precisely. This paper proposes MLC configurations...at an overhead of 21 extra bits per block. More specifically, ECP...configuration (61 extra bits per block). While the striped code is imperfect, as discussed in Section 3.2, it often leads to a statistically significant positive error rate between 0.1% and 4.2%. The other benchmarks, which see very little error,...the highest bits of the value map to the highest bits of the cells. The results given above use the latter encoding, but we also evaluated three intermediate write speeds, using that code reduces the mean number of complete overwrites the array has experienced, indicating...

Approximate computing is an area of research that seeks to optimize the execution of error-tolerant programs using both hard-ware and software techniques. Programmers control the impact of approximate execution using language features, analyses, or program logics. While much prior work has also considered low-overhead techniques for handling errors in memory, this paper focuses on extending the lifetime of precise storage on approximate MLCs. In the first, the bits whose quality was (statistically significantly) improved in writes...the results above assume a retention time of 9 seconds, or about...While the striped code is imperfect, as discussed in Section 3.2, it often leads to a statistically significant positive error rate between 0.1% and 4.2%. The other benchmarks, which see very little error,...the striped code leads to better output quality on average. For...the highest bits of the value map to the highest bits of the cells. The results given above use the latter encoding, but we also evaluated three intermediate write speeds, using that code reduces the mean number of complete overwrites the array has experienced, indicating...
Themes in approximate computing

Interleaving:
Programs are both approximate & precise

Error mitigation:
Exploit the hardware to minimize error
Phase-change memory (PCM) :-) 

Surpass DRAM’s scaling limits

“Almost” as fast as DRAM

Non-volatile

Faster than flash
Phase-change memory (PCM) :(

Write speed & energy

+ 

Cells wear out over time
Phase-change memory (PCM) :(

Multi-level cells are denser but need more time and energy.

Cells **wear out** over time and can no longer be used.
Phase-change memory (PCM)

Multi-level cells are denser but need more time and energy to protect against errors.

Cells wear out over time and can no longer be used for precise data storage.
Phase-change memory (PCM)
Phase-change memory (PCM)

- Fast
- Dense
- Accurate

Diagram showing the trade-offs between fast, dense, and accurate.
Approximate storage in PCM

Trade off accuracy for performance in multi-level cell accesses.

Use worn-out memory for approximate data instead of throwing it away.
Approximate storage in PCM

1. Trade off accuracy for performance in multi-level cell accesses.

2. Use worn-out memory for approximate data instead of throwing it away.
Approximate storage in PCM

1 Trade off accuracy for performance in multi-level cell accesses.

2 Use approximate throwing it away.
Single-level cells

- high
- low

Analog value

Digital value

0

1
Multi-level cells

- Analog value:
  - High
  - Low

- Digital value:
  - 11
  - 10
  - 01
  - 00
Writing to multi-level cells

- High
- Low

Analog value
- Probability

Digital value
- 11
- 10
- 01
- 00
Writing to multi-level cells, approximately
Speed — Accuracy — Density
Iterative writes

target range

low 00

high 11
Iterative writes, approximately

low 00

target range 10

high 11
Iterative writes, approximately
wider target range

fewer iterations to converge

faster writes

(or better density at the same speed)
Encoding to minimize error in approximate MLC

1 cell, 4 bits

\[ x \pm y \]

reliable

unreliable
Encoding to minimize error in approximate MLC

4 cells, 16 bits

lots of errors
Encoding to minimize error in approximate MLC

4 cells, 16 bits

lots of errors
Write speedup for approximate MLC

Writes are 1.7× faster on average with quality loss under 10%
Approximate storage in PCM

Trade off performance in accesses.

Use *worn-out* memory for *approximate* data instead of throwing it away.
Failed cells are a fact of life

0 1 1 0 1 0 0 1 1 1 0 1 0 0 1 0 1 1 1 0 0 0 1 1 0 0 1 0 1 1 0 0 1 0 0 0 1

a good block
Failed cells are a fact of life

a (tragically) failed block
Traditional error correction

corrected data block

correction bits
Correction resources are exhaustible

Uncorrectable (bad) block

Correction bits
Prioritized error correction

uncorrectable (bad) block
approximate
correction bits

error exposed where it does the least harm
Lifetime extension with block recycling

Lifetime extended by 23% on average
or from about 5.2 to 6.5 years

Lifetime extended by 23% on average
or from about 5.2 to 6.5 years