Relieving Self-Healing SSDs of Heal Storms

Li-Pin Chang, Sheng-Min Huang, Kun-Lin Chou
Speaker: Sheng-Min Huang

Embedded Software and Storage Lab
National Chiao-Tung University, Taiwan
Outline

• Introduction
• Heal Storm
• Virtual Wear Leveling
• Experiment Results
• Conclusion
Flash Wear Out Dynamics

• Charge may trap in tunnel oxide after PE cycles.

• The threshold voltage shift will become intolerably large and create erroneous bit values.

![Diagram of Flash Memory Structure](image-url)

- Control gate
- Gate oxide
- Floating gate
- Tunnel oxide
- Source
- Drain
- Silicon substrate
- Trapped electrons

Vr1, Vr2, Vr3, Vr4, Vr5, Vr6, Vr7, Vpass

TLC (111) (011) (001) (101) (100) (000) (010) (110)
Flash Healing

• Trapped charge (stress) dissipate slowly over time.
  • Accelerate this process under high temperature.

• Healing : Heated-Accelerated Self Recovery [2, 3]
  • Word line heaters to create high temp.
  • Block heal operation for system software
    • Heal nearly worn-out blocks
  • Time cost is about one second [3, 4]
  • Relieve about 80% stress [5]
Heal Storms

• **Wear Leveling (WL)**
  • Strive to balance the erase count of all blocks

• **Self-healing flash memory** heals flash blocks when blocks reach their PE cycle limit.

• **Heal storm**, blocks undergo block-healing within a short period of time.

![Graphs](image-url)
Negative Effects of Heal Storms

- Read Response degradation
- Write throughput fluctuation
- Unpredictable reliability
Virtual Wear Leveling

• Leverage the effect of erase count balancing from WL

• Virtual erase count
  • \( vec_i = eci + \delta_i \)
  • Operate conventional WL on vec

Desired effect

Leveraging wear leveling

\[ \begin{align*}
vec_i &= eci + \delta_i \\
\delta &= ec
\end{align*} \]
Virtual Wear Leveling in Action

(a) 
Erasure count

Physical block number

max

(b) 
Erasure count

Physical block number

max

(c) 

Physical block number

max

(d) 

Physical block number

max
Progressive Delta Leveling

• In the rest of the SSD lifetime,
  • the difference among erase counts remains unchanged.
    ⇒ Lots of blocks have unused PE cycles.
  • all blocks have the same $\delta$ (i.e. the difference $= 0$)
    • Increase $\delta_i$ with different rate
    • Update $\delta_i$ only after block healing
Experiment Setup

• Flash memory parameters
  • 16 flash chips
  • 16 KB per page
  • 4MB per block
• Latency
  • 0.5 ms for page read
  • 1.6 ms for page write
  • 2.9 ms for block erase
  • 1024 ms for block heal

<table>
<thead>
<tr>
<th>Workload</th>
<th>Disk vol. size</th>
<th>Total write</th>
<th>Total read</th>
<th>Avg req. size write</th>
<th>Avg req. size read</th>
<th>Scale factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>HM0</td>
<td>39</td>
<td>92</td>
<td>49</td>
<td>37.8</td>
<td>36.3</td>
<td>1/1650</td>
</tr>
<tr>
<td>RSRCH0</td>
<td>277</td>
<td>46</td>
<td>5</td>
<td>37.7</td>
<td>38.9</td>
<td>1/4000</td>
</tr>
<tr>
<td>STG0</td>
<td>113</td>
<td>62</td>
<td>15</td>
<td>38.1</td>
<td>52.8</td>
<td>1/2700</td>
</tr>
<tr>
<td>STG1</td>
<td>113</td>
<td>27</td>
<td>117</td>
<td>36</td>
<td>88.3</td>
<td>1/1900</td>
</tr>
<tr>
<td>WDEV0</td>
<td>136</td>
<td>31</td>
<td>9</td>
<td>35.8</td>
<td>40.7</td>
<td>1/5200</td>
</tr>
<tr>
<td>EX9</td>
<td>400</td>
<td>220</td>
<td>52</td>
<td>47.1</td>
<td>92.2</td>
<td>1/180</td>
</tr>
<tr>
<td>MSN</td>
<td>169</td>
<td>96</td>
<td>321</td>
<td>33.6</td>
<td>35</td>
<td>1/20</td>
</tr>
<tr>
<td></td>
<td>GB</td>
<td>GB</td>
<td>GB</td>
<td>KB</td>
<td>KB</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Experimental workloads and their characteristics.

<table>
<thead>
<tr>
<th>Name</th>
<th>Mapping scheme</th>
<th>Wear Leveling algorithm</th>
<th>Virtual wear leveling</th>
</tr>
</thead>
<tbody>
<tr>
<td>LWL</td>
<td>Page mapping</td>
<td>Lazy Wear Leveling</td>
<td>No</td>
</tr>
<tr>
<td>vLWL</td>
<td>Page mapping</td>
<td>Lazy Wear Leveling</td>
<td>Yes</td>
</tr>
<tr>
<td>HL</td>
<td>Page mapping</td>
<td>Heal Leveling</td>
<td>No</td>
</tr>
<tr>
<td>DH</td>
<td>Page mapping</td>
<td>Dheating</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 2: Flash management methods involved in our experiments.
I/O Performance

- LWL suffered transient variation by heal storms.
- DH had low write throughput because of high garbage collection overhead.
Reliability

• The % of blocks with a high bit error rate (BER) [6] should not fluctuate over time.

• Increasing gradually was good for system software to predict the SSD retirement.
Lifespan

- Our method did not affect the SSD lifespan
Erase count distribution

- Many PE cycles in blocks were wasted in HL.
Experimental Result Summary

• Conventional Wear leveling
  • Suffered heal storms that flash memory were occupied by block-healing operations.

• Dheating
  • Extremely high write amplification because of inaccurate hot/cold identification and local garbage collection in pools.

• Heal Leveling
  • Unexpected short device lifespan because of large variation in erase counts.
Conclusion

• Software-controlled block healing radically extends the SSD lifespan.

• Heal storm damages predictability of performance and reliability.

• Virtual wear leveling leverages conventional wear leveling to disperses block healing over time.

• Possible application of virtual wear leveling
  • Software-controlled bit density [7]
  • Erasing in MLC mode: vec+=2.2
  • Erasing in SLC mode: vec+=1
Thank you

Q & A
Reference