Understanding SSD Over Provisioning

Kent Smith
Sr. Director Product Marketing
Flash Components Division, LSI
What is SSD Over Provisioning (OP)

- It is part of all SSDs that use Flash memory
  - Required due to the inability to overwrite Flash without it first being erased
  - Therefore not part of HDDs
- The portion of the SSD capacity held in reserve (unavailable to the user):
  - Garbage collection (the major use)
  - SSD controller Firmware storage (small %)
  - Spare blocks (small %)
  - Some SSDs include other data protection beyond ECC, like RAISE™ technology (space requirement varies)
Why is More OP a Good Thing?

- OP will consume part of the storage capacity otherwise available to users
- Most users presume that a lower OP is better to provide maximum user storage capacity
- Most users do not understand that higher OP generally provides:
  - Higher write performance
  - Lower “Write Amplification”
  - Longer Flash life (endurance)
  - Space for data protection beyond ECC
How is OP Calculated?

- The ratio of OP vs. total user capacity

\[
\left( \frac{\text{Physical Capacity} - \text{User Capacity}}{\text{User Capacity}} \right) = \text{Over Provisioning}
\]

- 128GB physical flash capacity
- 120GB user capacity
- 7% (6.67%) OP
- However, the “true” physical capacity is usually misunderstood…
What is a Gigabyte?

- Some confusion due to the use of two different base numbers (radix)
- Binary GB is 7.37% more than Decimal GB
- Most OS’s display the “binary” representation for all categories (system memory, storage, networking, etc.)

<table>
<thead>
<tr>
<th></th>
<th>Binary</th>
<th>Decimal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential Notation</td>
<td>$2^{30}$</td>
<td>$10^{9}$</td>
</tr>
<tr>
<td>Actual Number of Bytes</td>
<td>1,073,741,824</td>
<td>1,000,000,000</td>
</tr>
<tr>
<td>Naming Convention</td>
<td>Gibibyte\textsuperscript{IEC}</td>
<td>Gigabyte\textsuperscript{SI}</td>
</tr>
<tr>
<td>Typical Uses for That Radix</td>
<td>System Memory</td>
<td>Storage/Networking</td>
</tr>
</tbody>
</table>

IEC – International Electrotechnical Commission
SI – International System of Units
### True Physical OP on SSDs

<table>
<thead>
<tr>
<th>Over Provisioning Percentages</th>
<th>0%</th>
<th>7%</th>
<th>16%</th>
<th>28%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marketed OP*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True Physical OP*</td>
<td>7%</td>
<td>15%</td>
<td>25%</td>
<td>37%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SSD Physical Cap</th>
<th>Resulting SSD User Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>64, 60, 55, 50</td>
</tr>
<tr>
<td>128</td>
<td>128, 120, 110, 100</td>
</tr>
<tr>
<td>256</td>
<td>256, 240, 220, 200</td>
</tr>
</tbody>
</table>

- An SSD listed with 128GB is marketed as “0% OP”, but in reality the true physical OP is ~7%
Performance Test – Environment

- Test result data points are based on post-garbage collection, steady state operation
- All preconditioning uses the same transfer size and type as the test result
  - E.g., random 4K results are preconditioned with random 4K transfers until it reaches steady state operation
- Test conducted on a single SSD to isolate the OP variable

Hardware:
- Intel Core i5-2500K 3.30 GHz
- 4 GB RAM 1333 MHz
- Intel H67 Express Chipset
- Intel RST 10.1.0.1008 (AHCI Enabled)
- Windows 7 Professional (32-bit)

Software:
- VDBench V5.02 (main test SW)
- IOMeter V1.1.0 (cross check)

SSD:
- MLC 24nm Toshiba NAND Flash
- SF-2281 FSP
Performance Test Results

Sequential Writes (128K sustained)

- Low (0%)
- Real World (25%)
- Medium (50%)
- Very High (75%)
- Typical SSDs (100%)

Low and Real World entropy are the same here

Sequential Writes are not affected by OP

SandForce Driven SSDs take advantage of different entropy levels with DuraWrite™
Performance Test Results

Random Writes (4K sustained)

With DuraWrite and real data, reduce user capacity by 6% and increase performance by 2% (already 6x higher)

Typical SSDs, reduce user capacity by 6% and increase performance by 29%

SandForce Driven SSDs take advantage of different entropy levels with DuraWrite™

Marketed Over Provisioning
Performance Test Results

Random Writes (4K sustained)

- **SandForce Driven SSDs** take advantage of different entropy levels with DuraWrite™

Log scale makes it easier to see typical SSDs or SandForce Driven SSDs with high entropy data benefit more from higher OP.
Write Amplification Test Results

Random Writes (4K sustained)

Write Amplification* vs. Marketed Over Provisioning

Typical SSDs or SandForce Driven SSDs with high entropy data benefit more from higher OP.

Aggregate Entropy (Level)
- Low (0%)
- Real World (25%)
- Medium (50%)
- Very High (75%)
- Typical SSDs (100%)

*(GB Written to Flash / GB Written from Host)

With DuraWrite and real data, write amplification is nearly 10 times lower so the benefit is less significant.

SandForce Driven SSDs take advantage of different entropy levels with DuraWrite™
**TRIM and Over Provisioning**

- **Presumed Valid Data** – Data deleted by the OS or user, but the SSD is not aware of it because TRIM was not present

- **Dynamic OP** – TRIM expands the OP when the OS or user erases data

\[ \text{Visible to the OS and user} \]

\[ \text{GB vs. Billion Bytes} \]

\[ \text{“Marketed OP”} \]

\[ \text{No TRIM 28%* OP} \]

\[ \text{With TRIM 50%* OP} \]

\[ \text{“7.37%”} \]

\[ \text{“0 to 28+%”} \]

\[ \text{Free Space} \]

\[ \text{User Data} \]

\[ \text{Real Over Provisioning} \]

\[ \text{True Valid Data} \]

\[ \text{Real Over Provisioning} \]

\[ \text{Dynamic OP} \]

\[ \text{True Valid Data} \]

*Illustrative numbers only*
### DuraWrite, TRIM and Over Provisioning

DuraWrite provides additional dynamic OP like TRIM

<table>
<thead>
<tr>
<th>SSDs without DuraWrite</th>
<th>SSDs with DuraWrite</th>
</tr>
</thead>
</table>
| **1. OS w/o TRIM**  
  *or*  
  **2. RAID Environment*** | **1. OS with TRIM**  
  *and*  
  **2. No RAID*** |
| **Free Space (OP)**  
  Presumed Valid Data | **Free Space (OP)**  
  True Valid Data |
| **True Valid Data** | **True Valid Data** |

**Free Space (OP)** Larger due to DuraWrite

*Some RAID 0 starting to support TRIM*
User Controlled Higher OP

- Users can increase the OP, but not decrease it.
- During initial setup and formatting, allocate a smaller partition (don’t use the full space)
  - SSD must be either “Fresh Out of Box” (FOB) or secure erased
- Leave the extra space unallocated
- The SSD controller automatically uses this as additional dynamic OP
Summary

- Over provisioning is a key component of any SSD

- Higher OP provides:
  - Higher write performance
  - Lower “Write Amplification”
  - Longer Flash life (endurance)
  - Space for data protection beyond ECC

- TRIM and DuraWrite contribute to OP

- Users can easily increase the OP if desired
LSI is Accelerating Mobile Computing at the FMS 2012!

Visit us at the booth 624-626 to:

• Experience the latest SandForce Driven™ Ultrabook™ systems

• See a live demo of the SandForce Driven Kingston USB SSD with Windows To Go

• Enter to win:
  − SandForce Driven SSDs from SuperSSpeed, Kingston, Corsair, Adata, EDGE, Mushkin, OWC, PNY, DMS, Patriot Memory and Wintec

  Up to 4 winners every 30 minutes!

  − Grand Prize – the latest SandForce Driven ASUS Zenbook Prime!

Don’t miss the drawing on Wednesday 8/22 @ 6:30pm