Novel Address Mappings for Shingled Write Disks

Weiping He, David Du
Outline

• Backgrounds
• Motivations/Objectives
• Approaches
• Evaluations
Background

- Traditional HDDs (perpendicular magnetic recording) are reaching areal data density
- Shingled magnetic recording is a new promising technology
SWD Characteristics

- Sequential write is preferred
- Write/update a block *in place* may destroy the valid data on the subsequent tracks if any
- General approaches to updates:
  - **In-place update:**
    - Extra reads/writes 1 write = 2 reads + 3 writes
  - **Out-of-place update:**
    - Copy-on-write
    - Mapping table and Garbage collection
- **Write amplification**
  - Update may incur extra read/write operations

\[
\text{Shingling direction}
\]

\[
\begin{align*}
\text{a} & \\
\text{b} & \\
\text{c} & \\
\end{align*}
\]
Tradeoff Between Space and Performance

- SWD Layout
  - Tracks are organized into bands
  - There are safety gaps between bands

A good candidate: \( W=2, \ N=4 \)
\[ SG = W \frac{N}{N + W - 1} \]  
\[ WAR = \frac{1}{N} \sum_{i=0}^{N-1} (1 + 2i) = N \]
Objective

• Good balance between capacity and performance
  – Reduce the write amplification overhead
Motivation for New Schemes

- General rule
  - Delay the use of track(s) in the middle of the bands, e.g., 3rd tracks

<table>
<thead>
<tr>
<th>Affected Tracks</th>
<th>Single Band comparison</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Conventional</th>
<th>Novel</th>
</tr>
</thead>
<tbody>
<tr>
<td>25% no</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>
Motivation for New Schemes

- General rule
  - Delay the use of track(s) in the middle of the bands, e.g., 3rd tracks

<table>
<thead>
<tr>
<th></th>
<th>Conventional</th>
<th>Novel</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>50%</td>
<td>Track1</td>
<td>no</td>
</tr>
</tbody>
</table>

Affected Tracks

Single Band comparison
Motivation for New Schemes

• General rule
  – Delay the use of track(s) in the middle of the bands, e.g., 3rd tracks

<table>
<thead>
<tr>
<th></th>
<th>Conventional</th>
<th>Novel</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>50%</td>
<td>Track1</td>
<td>no</td>
</tr>
<tr>
<td>75%</td>
<td>Track 1,2</td>
<td>Track1</td>
</tr>
</tbody>
</table>

Affected Tracks

Single Band comparison
Motivation for New Schemes

• General rule
  – Delay the use of track(s) in the middle of the bands, e.g., 3rd tracks

<table>
<thead>
<tr>
<th>Affected Tracks</th>
<th>Single Band comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional</td>
<td>Novel</td>
</tr>
<tr>
<td>25% no</td>
<td>no</td>
</tr>
<tr>
<td>50% Track1</td>
<td>no</td>
</tr>
<tr>
<td>75% Track 1,2</td>
<td>Track1</td>
</tr>
<tr>
<td>100% Track 1,2,3</td>
<td>Track 1,2,3</td>
</tr>
</tbody>
</table>
Motivation for New Schemes

- General rule
  - Delay the use of track(s) in the middle of the bands, e.g., 3rd tracks

<table>
<thead>
<tr>
<th></th>
<th>Conventional</th>
<th>Novel</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>Track 1</td>
<td>no</td>
</tr>
</tbody>
</table>

Affected Tracks

Two Bands comparison
Motivation for New Schemes

• General rule
  – Delay the use of track(s) in the middle of the bands, e.g., 3rd tracks

<table>
<thead>
<tr>
<th></th>
<th>Conventional</th>
<th>Novel</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>Track 1</td>
<td>no</td>
</tr>
<tr>
<td>50%</td>
<td>Track 1,2,3</td>
<td>no</td>
</tr>
</tbody>
</table>

Affected Tracks

Two Bands comparison
Motivation for New Schemes

• General rule
  – Delay the use of track(s) in the middle of the bands, e.g., 3rd tracks

<table>
<thead>
<tr>
<th></th>
<th>Conventional</th>
<th>Novel</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>Track 1</td>
<td>no</td>
</tr>
<tr>
<td>50%</td>
<td>Track 1,2,3</td>
<td>no</td>
</tr>
<tr>
<td>75%</td>
<td>Track 1,2,3,5</td>
<td>Track1,5</td>
</tr>
</tbody>
</table>

Affected Tracks

Two Bands comparison
Motivation for New Schemes

- General rule
  - Delay the use of track(s) in the middle of the bands, e.g., 3rd tracks

<table>
<thead>
<tr>
<th></th>
<th>Conventional</th>
<th>Novel</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>Track 1</td>
<td>no</td>
</tr>
<tr>
<td>50%</td>
<td>Track 1,2,3</td>
<td>no</td>
</tr>
<tr>
<td>75%</td>
<td>Track 1,2,3,5</td>
<td>Track 1,5</td>
</tr>
<tr>
<td>100%</td>
<td>Track 1,2,3,5,6,7</td>
<td>Track 1,2,3,5,6,7</td>
</tr>
</tbody>
</table>

Affected Tracks

Two Bands comparison
Scheme 1: R(4123)

1st 25%: 4th tracks
Scheme 1: $R(4123)$

1\textsuperscript{st} 25%: 4\textsuperscript{th} tracks
2\textsuperscript{nd} 25%: 1\textsuperscript{st} tracks
Scheme 1: R(4123)

1\textsuperscript{st} 25\%: 4\textsuperscript{th} tracks
2\textsuperscript{nd} 25\%: 1\textsuperscript{st} tracks
3\textsuperscript{rd} 25\%: 2\textsuperscript{nd} tracks
Scheme 1: R(4123)

1\text{st} 25\%: 4\text{th} tracks
2\text{nd} 25\%: 1\text{st} tracks
3\text{rd} 25\%: 2\text{nd} tracks
4\text{th} 25\%: 3\text{rd} tracks

Similarly, R(1423) can be adopted.
Scheme2: 14R(23)

1\textsuperscript{st} 50\%: 1\textsuperscript{st} and 4\textsuperscript{th} tracks
Scheme2: 14R(23)

1st 50%: 1st and 4th tracks
The following 25%: 2nd tracks
Scheme2: 14R(23)

1<sup>st</sup> 50%: 1<sup>st</sup> and 4<sup>th</sup> tracks
The following 25%: 2<sup>nd</sup> tracks
The last 25%: 3<sup>rd</sup> tracks
Scheme 3: 124R(3)

1st 75%: 1st, 2nd and 4th tracks
Scheme 3: 124R(3)

1\textsuperscript{st} 75%: 1\textsuperscript{st}, 2\textsuperscript{nd} and 4\textsuperscript{th} tracks

The last 25%: 3\textsuperscript{rd} tracks
## Scheme Comparisons

- **SWD setup:**
  - $W = 2$
  - $N = 4$

### Update Performance Prediction

<table>
<thead>
<tr>
<th>scheme</th>
<th>Spatial Locality</th>
<th>Write Amplification</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDD</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>R(4123)</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>14R(23)</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>124R(3)</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>1234</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

5 is best, 1 is worst
Evaluation Design (1/2)

- SWD simulation
  - Disksim, with hp_c3323a disk model
  - **Address mapper**: translate LBAs to PBAs
  - **Write amplifier**: convert an update into several reads/writes accordingly

- SWD setup:
  - 3000 cylinders
  - 1000 blocks per cylinder
  - Band size = 4
  - Write width = 2

<table>
<thead>
<tr>
<th>Trace</th>
<th>Write %</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web_0</td>
<td>0.70123</td>
<td>Update intensive</td>
</tr>
<tr>
<td>Financial_0</td>
<td>0.096978</td>
<td>Update light</td>
</tr>
<tr>
<td>hp_c2247</td>
<td>0.488449</td>
<td>Update moderate</td>
</tr>
<tr>
<td>SYN</td>
<td>1</td>
<td>Sequential write, average size 8 blocks, IAT=(mean 50ms, std. dev 10ms). No update.</td>
</tr>
</tbody>
</table>
Evaluation Design (2/2)

- Performance test points:
  - 25%, 50%, 75%, 100%
- For each test point, “pre-fill” the SWD space to the corresponding percentage to logically convert writes to updates
  - This is done to pass the percentage to the write amplifier
Results

- SWD Space Usage (a) Financial2
- SWD Space Usage (b) hp_c2247
- SWD Space Usage (c) Web_0
- SWD Space Usage (d) SYN
- SWD Space Usage (e) Financial2
- SWD Space Usage (f) hp_c2247
- SWD Space Usage (g) Web_0
- SWD Space Usage (h) SYN
Summary

• Shingled magnetic recording drives
• Write amplification problem
• Achieve good space gain and performance balance with new static address mappings
  – R(4123) or R(1423)
  – 14R(23)
  – 124R(3)
Future Work

• Comparison to out-of-place update SWD
• SWD file system designs
• Construct storage system with SWDs, e.g., RAID and erasure codes
• Hybrid SWDs
Questions?