LINEARLY COMPRESSED PAGES: A MAIN MEMORY COMPRESSION FRAMEWORK WITH LOW COMPLEXITY AND LOW LATENCY
Gennady Pekhimenko, Vivek Seshadri, Yoongu Kim, Hongyi Xin, Onur Mutlu, Todd C. Mowry (CMU), Phillip B. Gibbons, Michael A. Kozuch (Intel)

CHALLENGES IN MAIN MEMORY COMPRESSION

Uncompressed Page

<table>
<thead>
<tr>
<th>Address Offset</th>
<th>(L_0)</th>
<th>(L_1)</th>
<th>(L_2)</th>
<th>(\ldots)</th>
<th>(L_{N-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>64</td>
<td>128</td>
<td>(\ldots)</td>
<td>(N-1)*64</td>
<td></td>
</tr>
</tbody>
</table>

Compressed Page

<table>
<thead>
<tr>
<th>Address Offset</th>
<th>(L_0)</th>
<th>(L_1)</th>
<th>(L_2)</th>
<th>(\ldots)</th>
<th>(L_{N-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>(\ldots)</td>
<td>?</td>
</tr>
</tbody>
</table>

Virtual Page (4kB)

Virtual Address

Physical Address

Physical Page

Challenge 1: Address Computation

Compressed Data

Metadata (64B):

? (compressible) and ? (zero cache line)

Challenge 2: Mapping and Fragmentation

Challenge 3: Physically Tagged Caches

LINEARLY COMPRESSED PAGES (LCP)

Uncompressed Page (4kB: 64*64B)

4:1 Compression

Solves all 3 challenges

Compressed Data

Metadata (64B):

? (compressible) and ? (zero cache line)

LCP Overview and Optimizations

- Page Table entry extension: compressed type, size
- Operating System management support: 4 memory pools
- Changes to cache tagging logic
- Handling page overflows
- Compression algorithms: BDI and FPC
- Metadata cache: Avoids additional requests to metadata
- Memory bandwidth reduction
- Zero pages and zero cache lines

Key Results: Compression Ratio, Bandwidth, Performance

Compression Ratio

GeoMean

SPEC2006, databases, web workloads, L2 2MB cache

Zero Page  FPC  LCP (BDI)  LCP (BDI+FPC-fixed)  MXT  LZ

GeoMean

Normalized BPKI

FPC-memory (None, LCP-BDI)

(FPC, FPC) (BDI, LCP-BDI)

(BDI, LCP-BDI+FPC-fixed)

Cores  LCP-BDI  (BDI, LCP-BDI)

1  6.1%  9.5%

2  13.9%  23.7%

4  10.7%  22.6%

Average performance improvement