OBJECTIVE: Improving the lifetime of Cache Sub systems.

Emerging Non-Volatile Memory (NVM) such as STT-RAM (Spin Transfer Torque Random Access Memory), PCM (Phase Change Memory), ReRAM (Resistive RAM) has more advantages compared to SRAM. NVM has low leakage power and high density. Despite its advantages, the major challenge of NVM is their limited write endurance and permanent/transient failures due to the thermal fluctuations, read disturbance and process variation. The write endurance of both SRAM and DRAM is above 10^{15}, whereas STT-RAM is below 4×10^{12}. The higher frequency of write operations will reduce the lifetime of the cache. NVM is high inevitable against soft errors.

Recent research is going on to overcome these drawbacks and to improve the lifetime of NVM. Different approaches are Write Minimization technique, Wear Leveling technique and Error Correcting Codes. In Write Minimization techniques, number of write operation on a particular cell is restricted. In Wear Leveling techniques writes are uniformly distributed among the cells. Error correcting methods are used to overcome the permanent/transient errors.

Equal Writes technique works by recording the number of writes on a block and changing the cache-block location of a hot data item to redirect the future writes to a cold block to achieve wear leveling. During the write operation unwanted bit flips and unsuccessful write are the consequence of the faults. Considering the consequences, the following improvements are proposed to the existing solution.

1. The bit error rate of 0 to 1 and 1 to 0 flipping is more in extreme case. The most common error correcting code, a SEC-DED (Single Error Correction, Double Error Detection) is used to correct the single error and detects the double error in hot blocks. With the help of SEC-DED algorithm, reliability can be ensured for the data bits stored in the Cache block.

2. Whenever a write operation is done on the hot block all bit flips at each write. On a write, quick bit-by-bit inspection of the original data word and the new data word write should only be done, writes either the new data word or the “flipped” value of it. This technique reduces the write time by half, more than doubles the write endurance and achieves commensurate savings in write energy under the same instantaneous write power constraint.

3. Malicious software can easily force cache to be flushed continuously, which produces writes to certain cells repeatedly (known as selective attack) and wears out cache. Although existing wear-leveling approaches could evenly distribute writes under selective attack, the overall endurance of cache is still severely impacted, and therefore it is suboptimal. So, security refresh algorithm has to be designed to avoid such attacks.

Base Paper:

Mittal, Sparsh, and Jeffrey S. Vetter. "Equalwrites: Reducing intra-set write variations for enhancing lifetime of non-volatile caches." *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 24.1 (2016): 103-114.