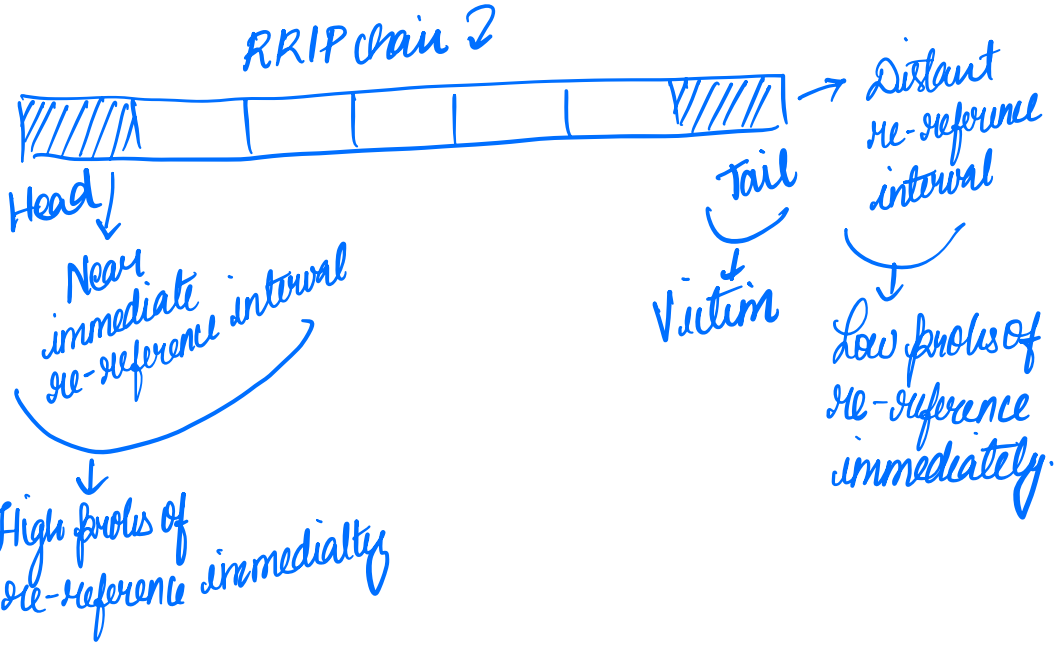
Re-Reference Interval Prediction (RRIP):

Motivation: Design a thrash-resistant and scan-resistant replacement policy that exploits Re-Reference Interval Prediction

1) SRRIP (Static Re-Reference Interval Prediction) ⇒ Scan resistant

2) DRRIP (Dynamic Re-Reference Interval Prediction) ⇒ Scan & Thrash resistant

* Requires only 2-bits per cache block ♡.



Doubt: This paper says that DIP & LRU cant handle Scan. But it can!

RRIP:
→ Prevents cache blocks with distant re-reference interval from evicting blocks that have a near-immediate re-reference interval.

NRU: Not Recently Used ⇒ SRRIP with M = 1

LFU: Least Frequently Used ⇒ Uses a counter to anoint the higher as near-immediate & lower as distant-future

+ mixed, cache friendly ↪ High locality.

Doubt: Other than "Scanning" & "Thrashing" is there something else?

Doubt: Why cant DIP handle scan? → The BIP will retain

(NRU) Not Recently Used:

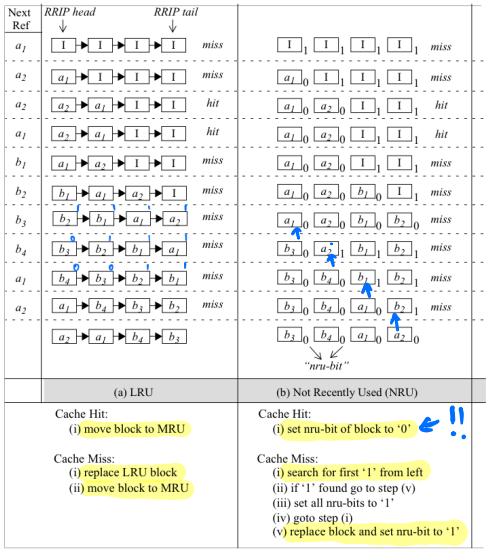
→ Has a single nru-bit → says whether its near-immediate or distant future

→ 0: Near immediate
1: Distant future.

→ On cache fill, the first block from HEAD with nru-bit = 1 is evicted.

→ The newly added block becomes 0.

→ Once all nru-bit = 1 blocks are evicted, all the blocks with nru = 0 are promoted to 1.



Issues with NRU:

→ For scan: Keeps block that will never be used

→ For thrash: Evicts block that may be used later.

↳ Solution → SRRIP & DRRIP

→ RRPV = Re-ref Pred Value.

→ In NRU 2 possible values are there since we use 1 bit. $M = 1 \Rightarrow 2^M = 2^1 = 2$.

↳ 0 → Immediate
1 → Distant.

→ This work introduces the use of 2^M, $M > 1$ values. So we can also track intermediate cases.

SRRIP/DRRIP = NRU, when $M = 1$.

→ SRRIP/DRRIP: Same as NRU. Select the victim with nru-bit > 0. Select victim from HEAD side. But in NRU, you update everything to 1 when there is no 1. Here it differs here.

RRIP hit promotion policy:

→ Hit Priority (HP)

→ Frequency Priority (FP)

SRRIP (Static RRIP): (The heuristic is based on hit, miss statistics).

RRIP-HP:

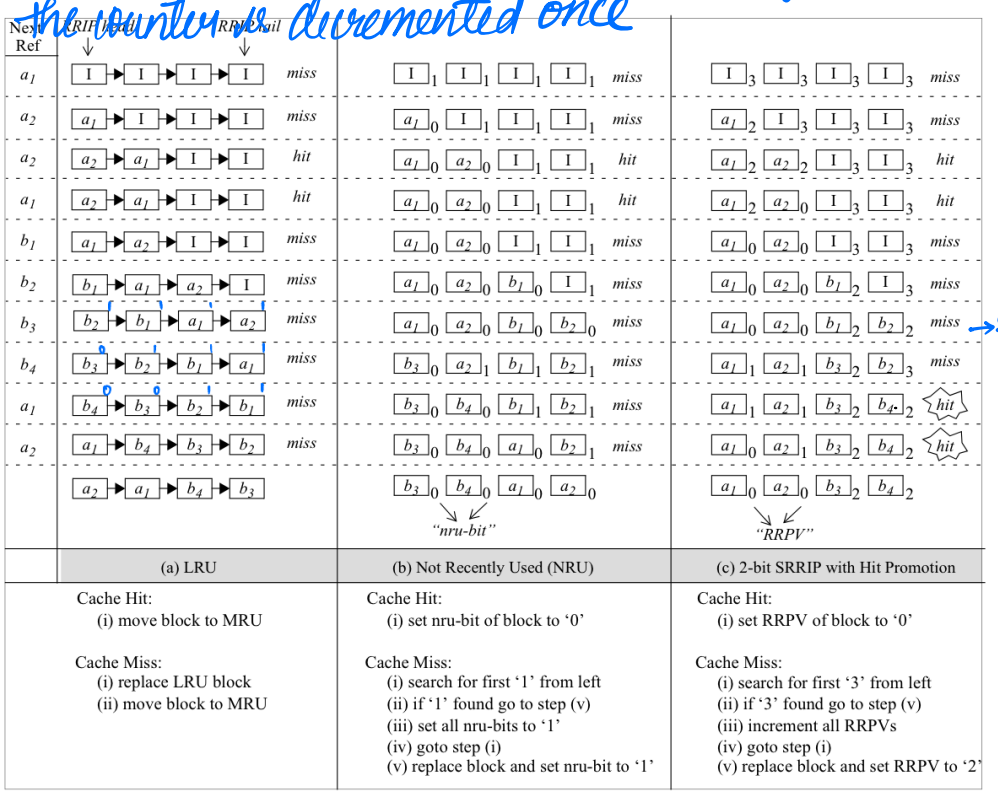
→ Block with hit is converted immediately to near immediate. Make RRPV = 0

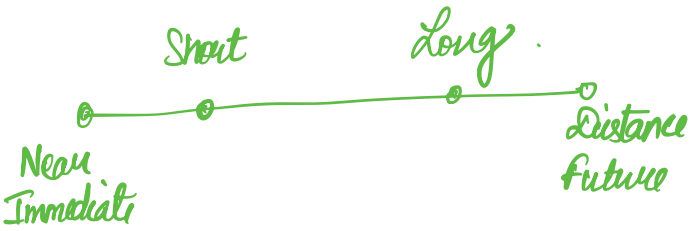
RRIP-FP:

→ Replace blocks with no hits over blocks with atleast 1 hit.

↳ Problematic when the block gets hit only once & never again. ⇒ Occupies space

→ To fix this, at each hit, instead of making RRPV = 0 (like RRIP-HP) the counter is decremented once





RRIP-HP:

Insertion: Long
Eviction: Distance Future
Promotion: Near-intermediate

RRIP-FP:

Insertion: Long
Eviction: Distance future
Promotion: Short

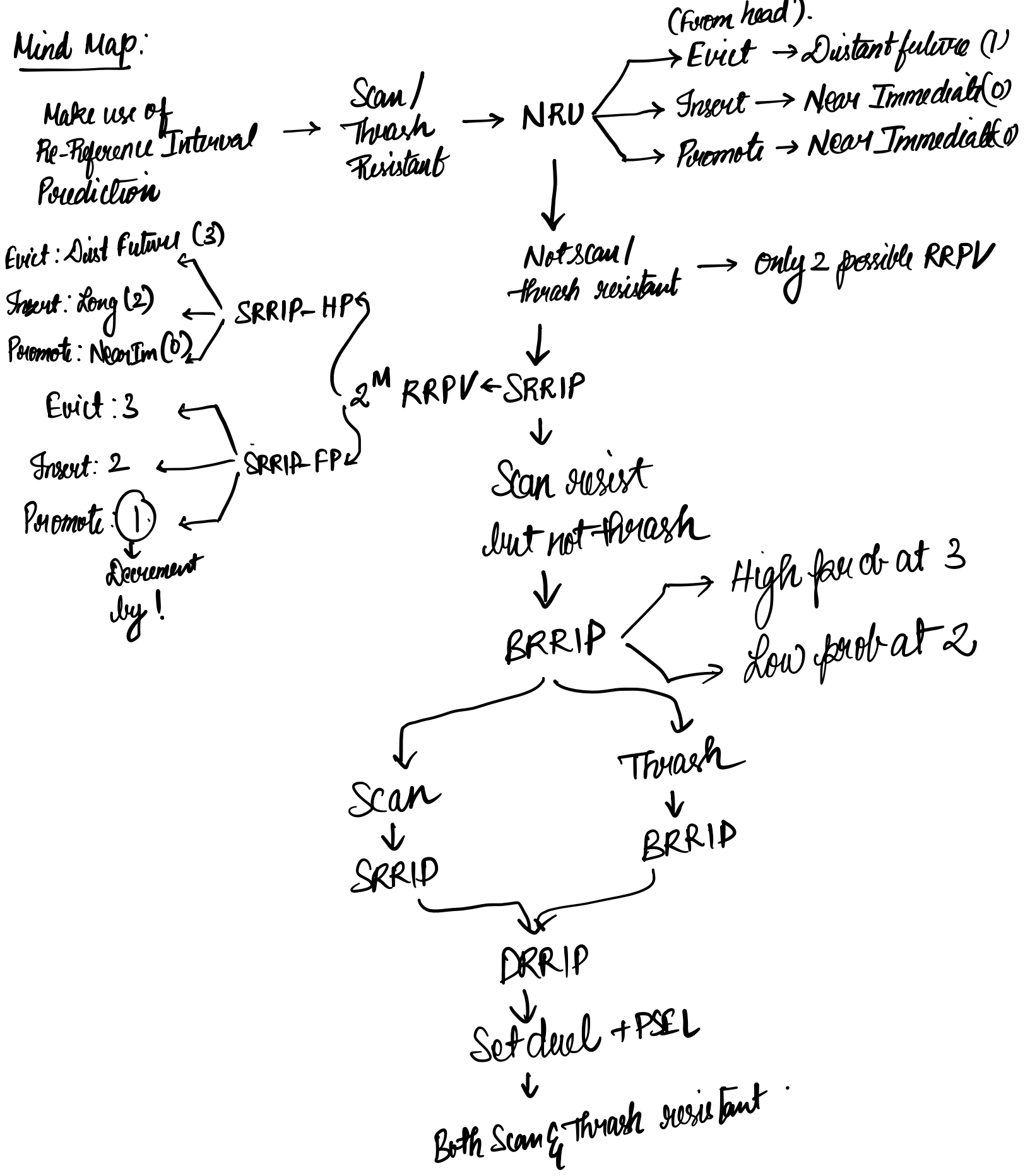
* SRRIP cant store active blocks when scan length > $(2^M - 1) * (A - W)$

A = associativity
W = working set.

* SRRIP cant handle thrashing
  ↳ Solution: BRRIP ⇒ Bimodal RRIP (Similar to BIP)
    ↳ Majority @ distant
    ↳ Some @ long
  But not good for scanning.

Final solution: DRRIP ⇒ SRRIP(scan resist) + BRRIP(Thrash resist) + Set dueling

Doubt: Why sec 4.4 exists?


Mind Map:
Make use of Re-Reference Interval Prediction
Scan / Thrash Resistant
NRU
(From head).
Evict → Distant future (3)
Insert → Near Immediate (0)
Promote → Near Immediate (0)
Not scan / thrash resistant
Only 2 possible RRPV
SRRIP
2^M RRPV
SRRIP-HP
Evict: Dist Future (3)
Insert: Long (2)
Promote: NearIm (0)
SRRIP-FP
Evict: 3
Insert: 2
Promote: (1)
Decrement by 1
Scan resist but not thrash
BRRIP
High prob at 3
Low prob at 2
Scan
SRRIP
Thrash
BRRIP
DRRIP
Set duel + PSEL
Both Scan & Thrash resistant