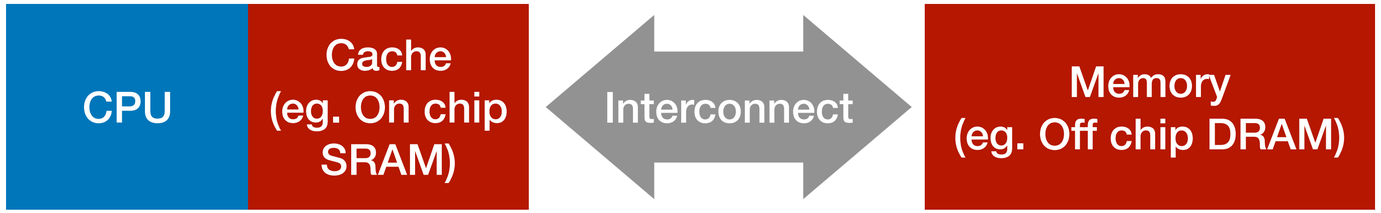
This paper's focus

The breaking point - is there any WASTE of bandwidth?

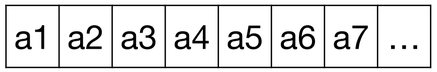
- "Useless or redundant fetches which increase memory traffic"

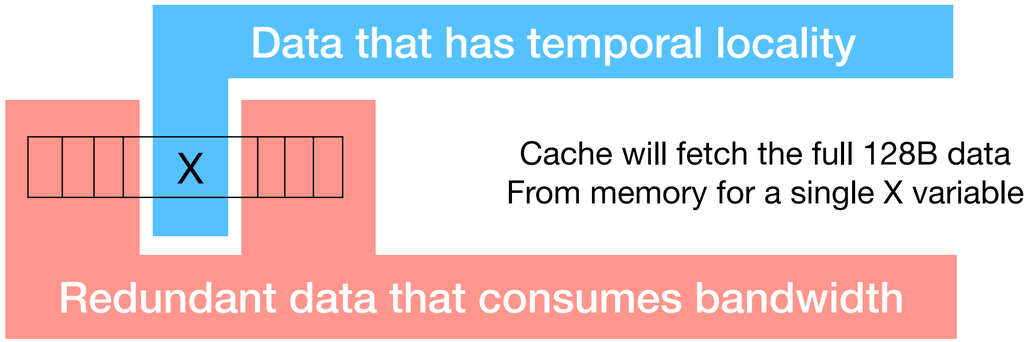


4 bytes **128 bytes** cache line

Then why we set it so big when we design this cache?

ANS: large array with strong spatial locality
can benefit from large cache line size





Cache will fetch the full 128B data
From memory for a single X variable

Conclusion: traditional single cache design is a **compromise**
between different access needs of different types of data.
It is just this compromise that creates the margin to shrink bandwidth.

Author’s proposal

The key idea: FINE TUNE

- What is fine tune? only consume bandwidth that we really need