Hyperthreading and “Almost Amdahl”

What is Hyperthreading and what can it Do?

Hyperthreading is when a CPU chip has more states than cores (“Execution Units” here).

In this case, if one thread of execution blocks (waiting for a memory fetch, for instance), then the next thread can resume execution.

If we let \( H \) be the fraction of a CPU’s capacity that one hyperthread can keep busy, then the remaining unused capacity is \( (1-H) \). If another hyperthread can keep \( H \% \) of that capacity busy, then that leaves \( (1-H)(1-H) \) remaining unused capacity and so on.

If we have \( n \) hyperthreads, then the final remaining unused capacity is \( (1-H)^n \). The capacity actually in use would then be \( 1-(1-H)^n \).

If one thread can only keep the CPU \( H\% \) busy, then the speed-up is potentially:

\[
SU = \frac{1}{1-(1-H)^n}
\]

A Lidar Application: Four Cores with Two Hyperthreads per Core

Note that this is upside-down from our usual convention. Sorry. I got this from someone else.