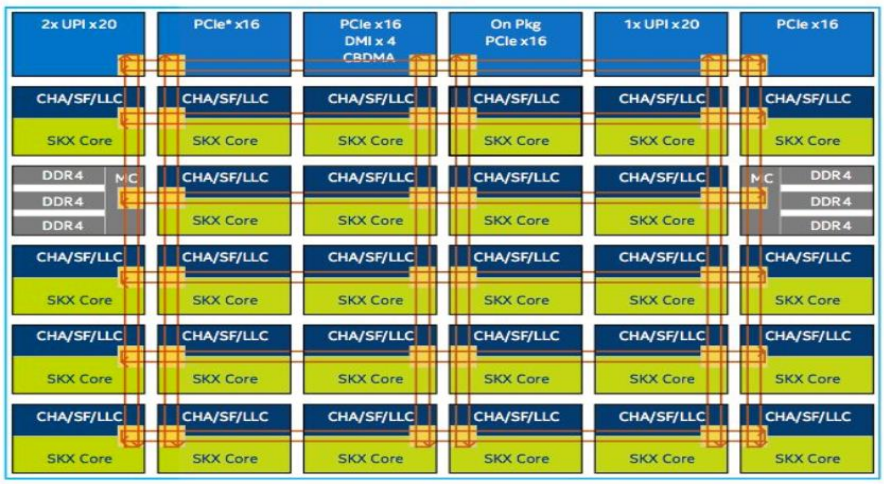
EE155/Comp122: Parallel Computing

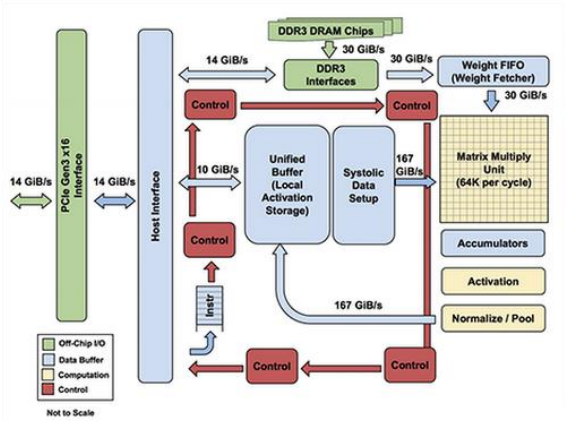
For full details, see https://www.ece.tufts.edu/ee/155 or e-mail joel.grodstein@tufts.edu



Why do CPUs have so many cores nowadays? Don't CPU designers know that using lots of cores is really hard? Why don't they just put one really, really fast core on a chip?

This is an NVidia Tesla GPU. There are 3840 cores. What the hay !!?! Are they for real?

Yes, the cores are those little dots!



This is a Google TPU. It does neural-net inferencing really well – and powers many of Google's servers.

We'll learn about parallel computing at the hardware-software edge; what's inside parallel computers, how to program them, and how to make them run really, really fast once we understand the hardware.

We'll also use our knowledge to try and write programs that run really slow – so slow that they might clobber the memory system and crash the computer (a.k.a., a "memory virus")

We'll learn how to make neural networks run really fast – on CPUs, GPUs, and also in Google's mysterious Tensor Processing Unit – the new parallel computer that powers neural networks on their servers.