

Domain-specific computing

- ▶ H&P cite performance gains via domain-specific architectures
- ▶ Expansion of historical concepts: math co-processors for early x86 chips, and GPUs
- ▶ Key: architecture maps to a specific application or problem class (TPU example)
- ▶ **Tensor Processing Unit (TPU)** used by Google for DNN & optimized to outperform CPUs and GPUs in cycles and energy efficiency
 - ▶ Removal of features required by general-purpose CPUs
 - ▶ Add a large two dimensional multiply unit
 - ▶ Use eight-bit integers since DNN do not require high precision

