|  |  |  |
| --- | --- | --- |
|  | CPU | GPU |
| Design Orientation | Latency-oriented | Throughput-oriented |
| Clock Frequency | High clock frequency | Moderate clock frequency |
| Control | Sophisticated control with branch prediction for reduced branch latency and data forwarding for reduced data latency | Simple control without branch prediction and data forwarding |
| Caches | Large caches to convert long latency memory accesses to short latency cache accesses | Small caches to boost memory throughput |
| ALUs | Powerful ALUs: reduced operation latency | Energy efficient ALUs: many, long latency but heavily pipelined for high throughput |
| Always Used in | Sequential parts where latency matters (10x faster than GPUs for sequential code) | Parallel parts where throughput matters (10x faster than CPUs for parallel code) |