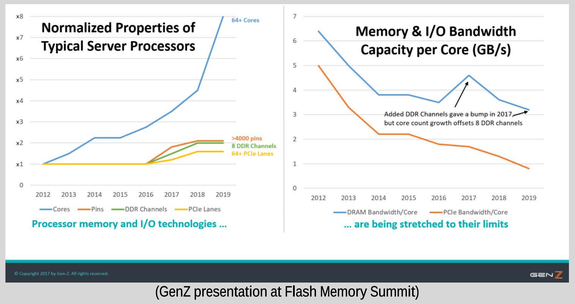
Creating the Optimal Regression Farm Infrastructure That Meets All Your Team's Simulation Requirements

David Lacey, Ed Powell

Hewlett Packard Enterprise

Challenge

Memory bandwidth/core is shrinking



(GenZ presentation at Flash Memory Summit)

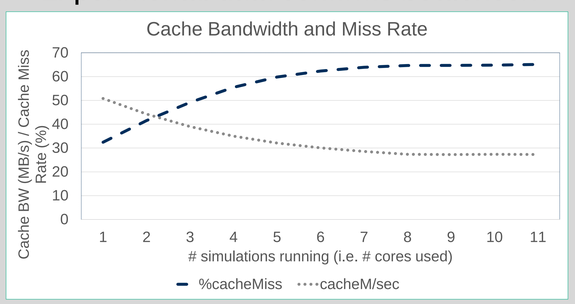
- Adding more cores to the processor does not increase individual EDA workload performance linearly
- More cores running EDA workloads such as simulations are competing for the same memory bandwidth

Impact on Sim Performance

Simulation Performance Degrades

- Each individual simulation takes longer to complete as a single socket is loaded with multiple simulations

Experimental data confirms bottleneck

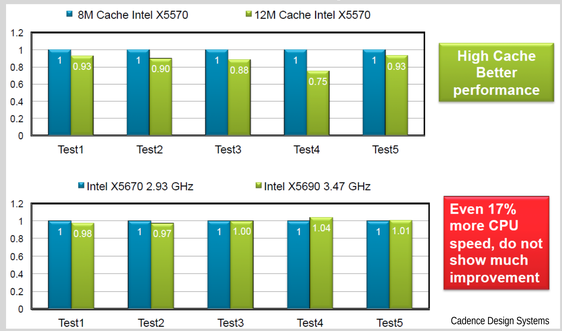


- Cache bandwidth drops and the cache miss rate rises as more simulations use additional cores on a socket
- Cannot keep each simulation fed with the memory data and results in slower simulation speeds

The memory bottleneck decreases the simulation performance as number of jobs running increases

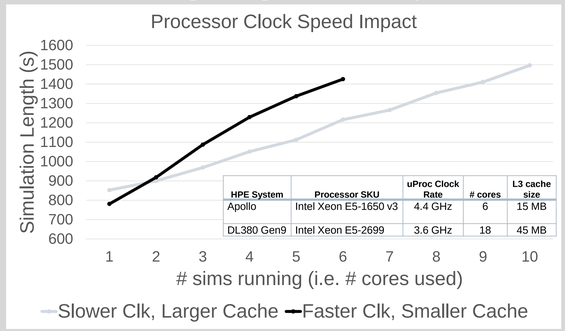
Optimize HW Configuration

Maximize Cache per Core



- It is better to choose processors with larger L3 cache over a processor with less L3 cache but a faster clock rate

Don't target higher clock speeds

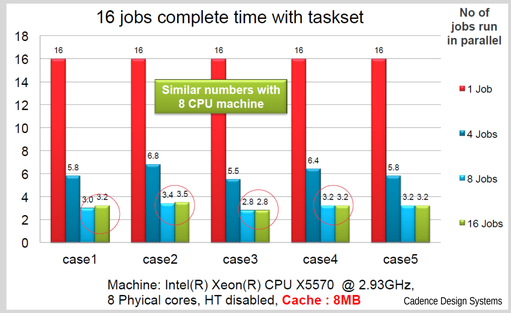


HPE System	Processor SKU	uProc Clock Rate	# cores	L3 cache size
Apollo	Intel Xeon E5-1650 v3	4.4 GHz	6	15 MB
DL380 Gen9	Intel Xeon E5-2699	3.6 GHz	18	45 MB

- Performance boost provided by the faster clock rate is nullified by the limited amount of L3 cache

Optimize Total Cycles

Do use all available licenses
Don't run more jobs than cores



- Experiments run measuring relative time to complete 16 simulations using different number of parallel jobs

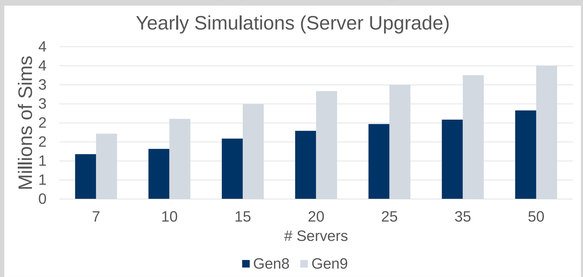
More Servers Brings More Value From Simulation Licenses

Licenses	# two socket servers	cores/ socket used	Speed of reference simulation (ns/sec)	Total simulated ms achievable in a year	# of average length (130k ns) sims/year
100	4	13	78.7	248,188	1,909,141
100	5	10	88.6	279,409	2,149,300
100	6	9	93.9	296,123	2,277,870
100	7	8	97.6	307,898	2,368,450
100	8	7	104.3	328,993	2,530,712
100	9	6	109.0	343,691	2,643,773
100	10	5	118.9	374,821	2,883,236
100	13	4	126.2	398,126	3,062,507
100	17	3	137.2	432,746	3,328,816
100	25	2	146.6	462,288	3,556,065
100	50	1	154.7	488,016	3,753,968

- To gain more overall cycles, run all licenses on whatever cores you have, ignoring negative simulation performance impact of running multiple cores

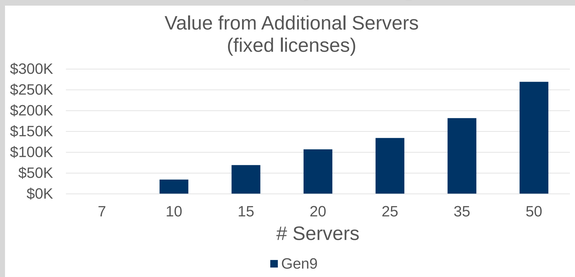
Optimize Total Costs

Server Upgrades Bring Value



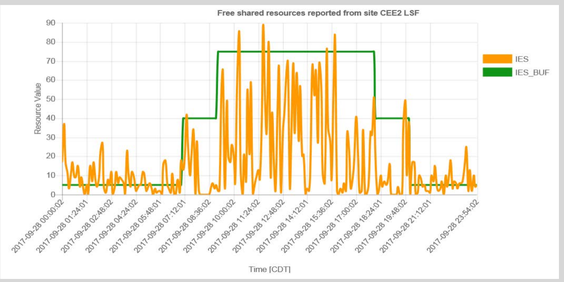
- Investing in regular server refreshes brings immediate results
- Experimental results between Gen8 and Gen9 HPE servers showed a 40% increase in simulation capacity

HW Costs Offset by Adding Servers



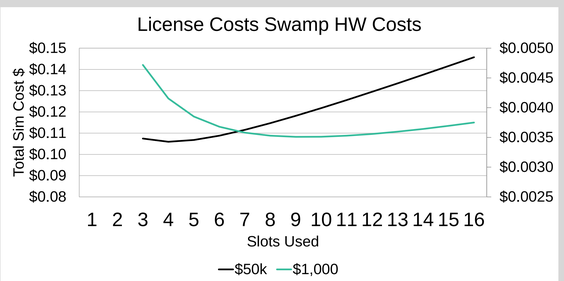
- Investing in additional hardware to enable fewer jobs per core brings direct value back through better license utilization
- This license utilization creates increase license cost value

Optimize Interactive Jobs



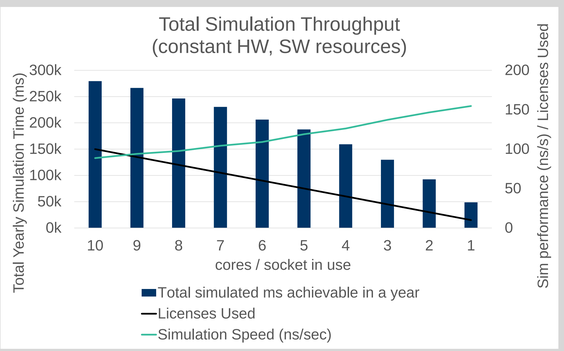
- Engineers are an organization's most expense resource
- Efforts should be made to maximum their productivity
- Allocate more licenses to interactive use during work hours

License Costs Dominate



- When considering total costs of simulation jobs (HW costs, license costs, infrastructure costs), license costs quickly dominate the overall costs
- Investing in more hardware and faster hardware directly optimizes your overall job cost

Don't Leave Licenses Idle



- With a fixed server count, leaving licenses idle just to force less cores/socket to be in use does not produce enough individual simulation performance increase to yield more overall cycles