

HPC platform efficiency and challenges for a system builder

Martin Hilgeman

Technical Director, HPC

martin.hilgeman@dell.com



The landscape is changing

“We are no longer in the general purpose era... the argument of tuning software for hardware is moot. Now, to get the best bang for the buck, you have to tune both.”

- Kushagra Vaid, general manager of server engineering, Microsoft Cloud Solutions

<https://www.nextplatform.com/2017/03/08/arm-amd-x86-server-chips-get-mainstream-lift-microsoft/amp/>



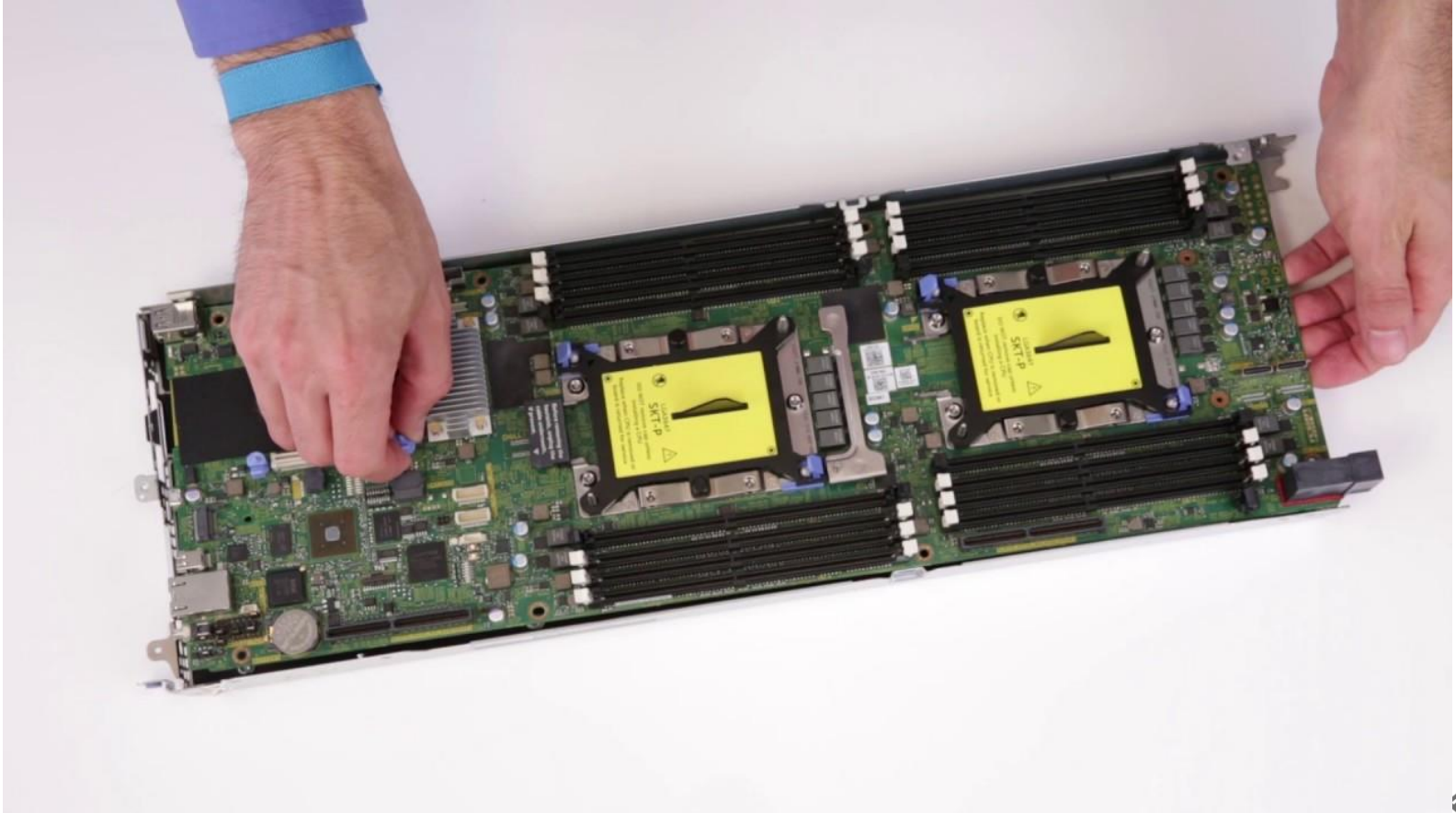
4 nodes in 2U – the HPC standard



The evaluation of the dual socket

Model	6100		6220		6320		6420
Year	2010	2011	2012	2013	2014	2015	2017
CPU	X5570	X5690	E5-2690	E5-2695 v2	E5-2699 v3	E5-2699A v4	8180
Cores	4	6	8	12	18	22	28
GHz	3.33	3.46	2.9	2.7	2.3	2.4	2.5
TDP (W)	95	130	135	135	145	145	205
DIMM	12/(3 channels)		16/(4 channels)		16/(4 channels)		16/(6 ch)
PSU	1400		1400		1600		2000
Price	\$1,500	\$1,700	\$2,050	\$3,549	\$3,805	\$4,938	\$10,009

The de-facto standard node in HPC



Improving performance - what levels do we have?

- Challenge: Sustain performance trajectory without massive increases in cost, power, real estate, and unreliability
- Solutions: No single answer, must **intelligently turn** “Architectural Knobs”

$$\begin{matrix} \textcircled{1} & \textcircled{2} & \textcircled{3} & \textcircled{4} & \textcircled{5} \\ \underbrace{(Freq) \times \left(\frac{cores}{socket}\right) \times (\#sockets)}_{\text{Hardware performance}} & \times & \underbrace{\left(\frac{inst\ or\ ops}{core \times clock}\right)}_{\text{What you really get}} & \times & \underbrace{(Efficiency)}_{\text{Software performance}} \end{matrix}$$

Hardware performance

What you really get

Software performance

Turning the knobs 1 - 4

1

Frequency is unlikely to change much Thermal/Power/Leakage challenges

2

Moore's Law still holds: 130 -> 14 nm. LOTS of transistors

3

Number of sockets per system is the easiest knob.

Challenging for power/density/cooling/networking

IPC still grows

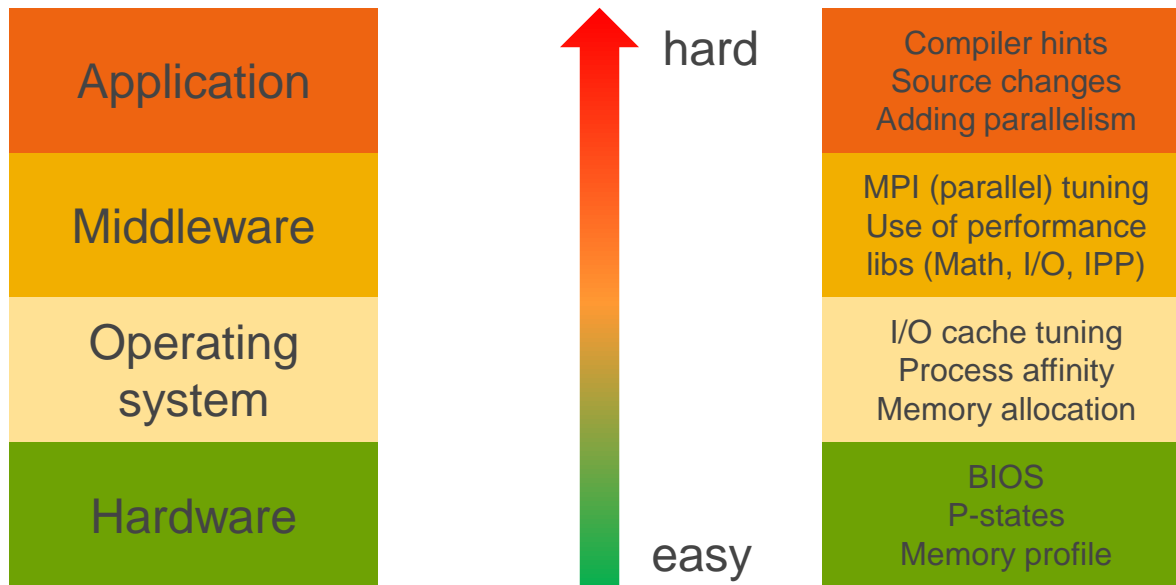
4

FMA, AVX, accelerator implementations for algorithms

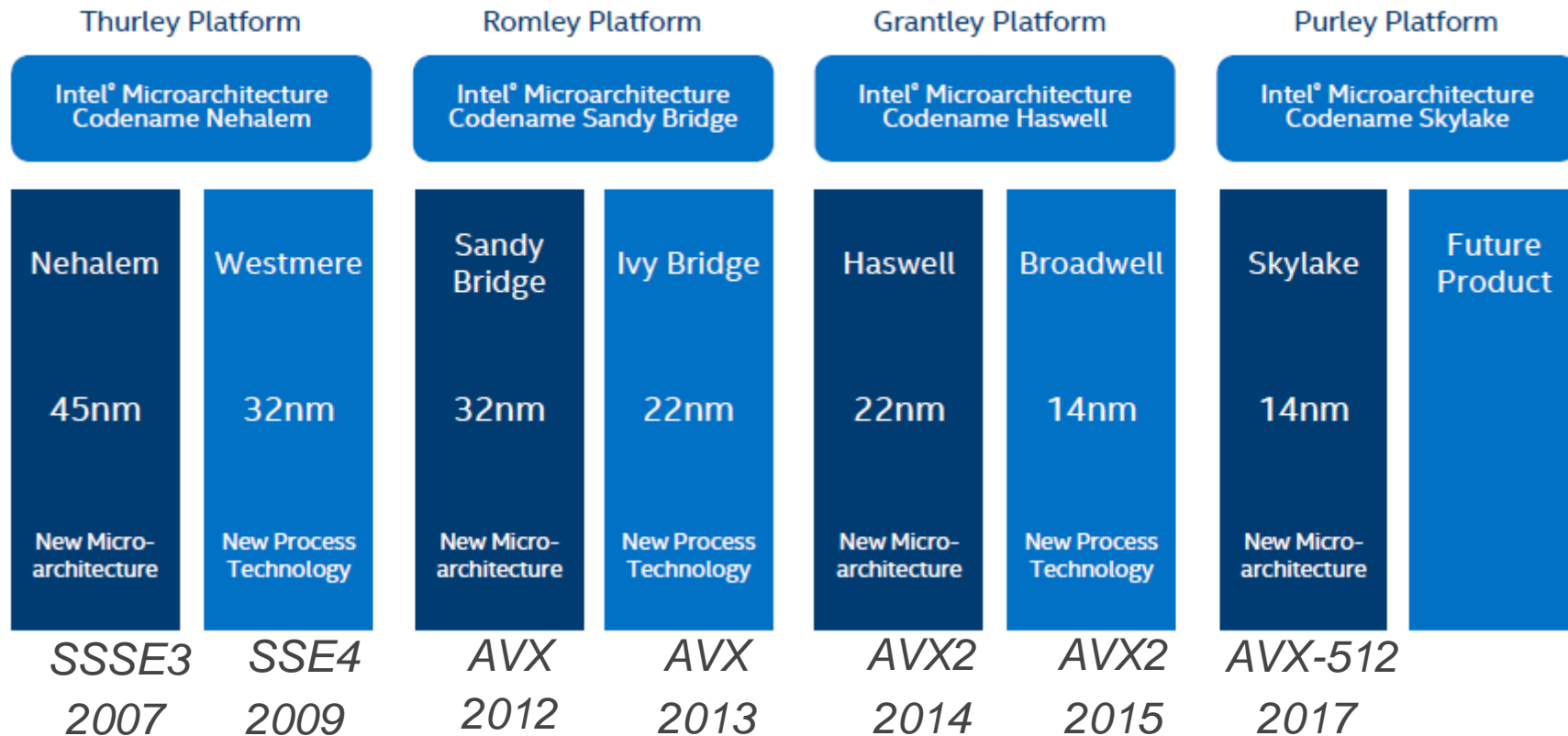
Challenging for the user/developer

Turning knob #5

Hardware tuning knobs are limited, but there's far more possible in the software layer



New capabilities according to Intel



The state of ISV software

Segment	Applications	Vectorization support
CFD	Fluent, LS-DYNA, STAR CCM+	Limited SSE2 support
CSM	CFX, RADIOSS, Abaqus	Limited SSE2 support
Weather	WRF, UM, NEMO, CAM	Yes
Oil and Gas	Seismic processing	Not applicable
	Reservoir Simulation	Yes
Chemistry	Gaussian, GAMESS, Molpro	Not applicable
Molecular dynamics	NAMD, GROMACS, Amber,...	PME kernels support SSE2
Biology	BLAST, Smith-Waterman	Not applicable
Molecular mechanics	CPMD, VASP, CP2k, CASTEP	Yes

Bottom line: ISV support for new instructions is poor. Less of an issue for in-house developed codes, but programming is hard

What does Intel do about these trends?

Problem	Westmere	Sandy Bridge	Ivy Bridge	Haswell	Broadwell	Skylake
Inter CPU bandwidth	No problem	Even better	Two snoop modes	Three snoop modes	Four (!) snoop modes	<ul style="list-style-type: none"> • UPI • COD snoop modes
Memory bandwidth	No problem	Extra memory channel	Larger cache	Extra load/store units	Larger cache	<ul style="list-style-type: none"> • Extra load/store units • +50% memory channels
Core frequency	No problem	<ul style="list-style-type: none"> • More cores • AVX • Better Turbo 	<ul style="list-style-type: none"> • Even more cores • Above TDP Turbo 	<ul style="list-style-type: none"> • Still more cores • AVX2/FMA • Per-core Turbo 	<ul style="list-style-type: none"> • Again even more cores • optimized FMA • Per-core Turbo based on instruction type 	<ul style="list-style-type: none"> • More cores • Larger OOO engine • AVX-512 • 3 different core frequency modes

Result: more variation



A case study on power



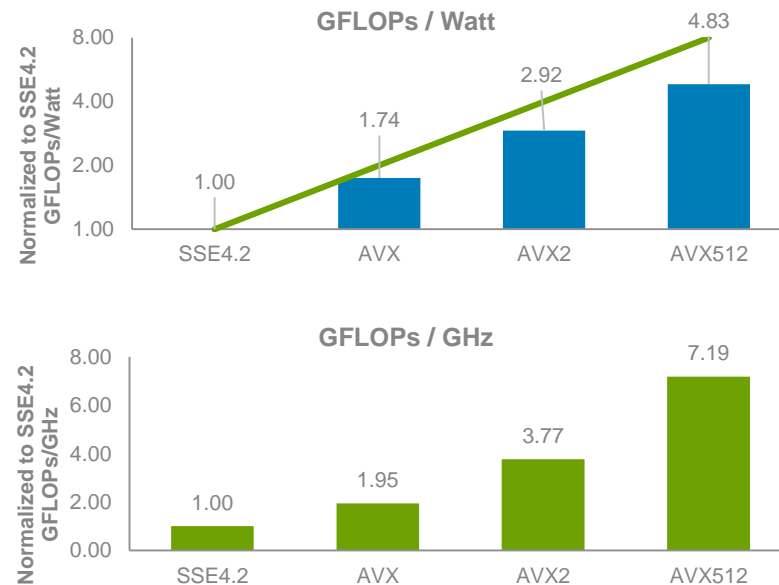
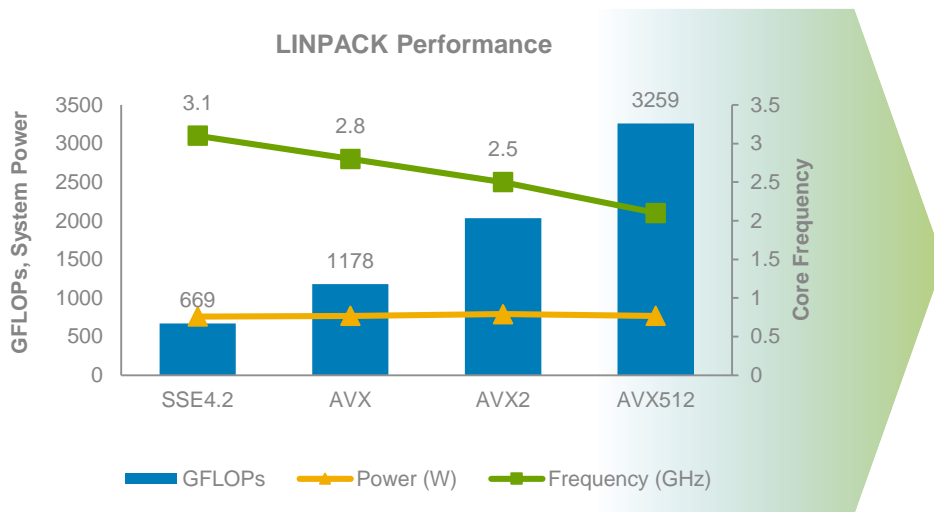
Key Aspects of Acceleration

We have lots of transistors... Moore's law is holding; this isn't necessarily the problem

We don't really need lower power per transistor, we need lower power per **operation**

How to do this?

Performance and Efficiency with Intel® AVX-512



Intel® AVX-512 delivers significant performance and efficiency gains

Source as of June 2017: Intel internal measurements on platform with Xeon Platinum 8180, Turbo enabled, UPI=10.4, SNC1, 6x32GB DDR4-2666 per CPU, 1 DPC. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

Powerful instructions to save power

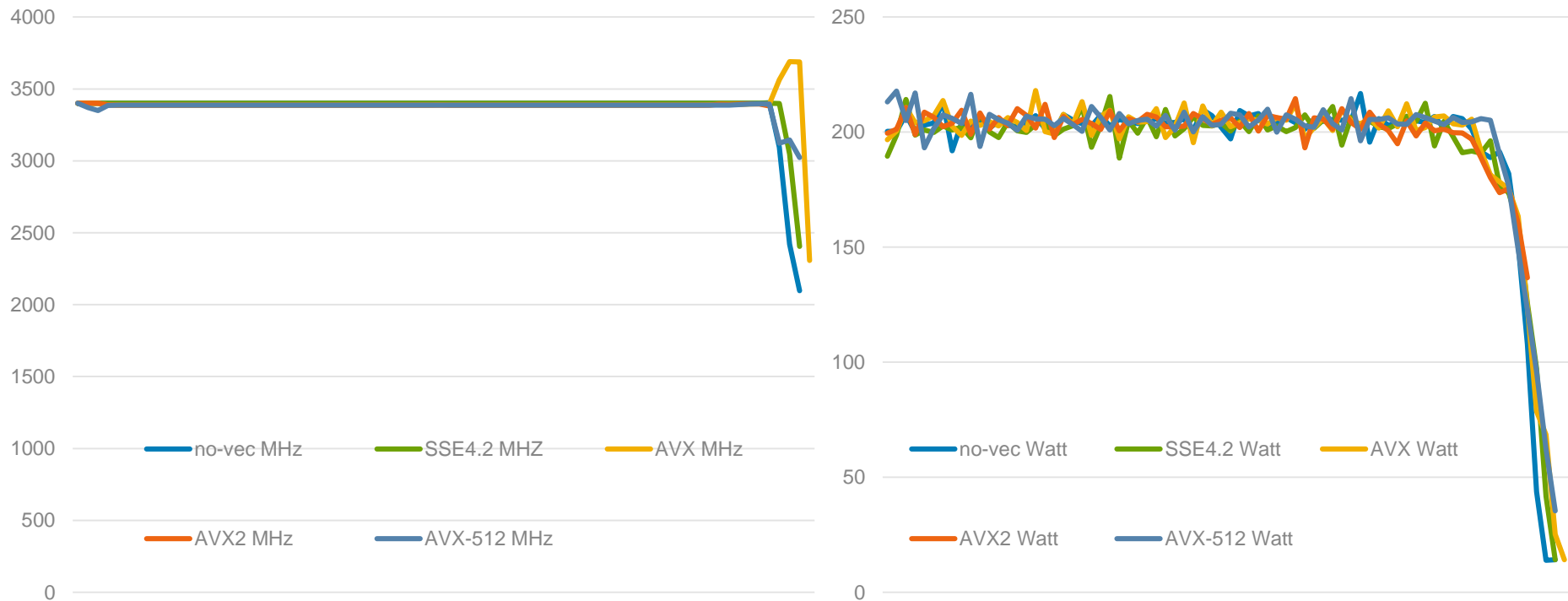
For LINPACK, powerful instructions can bring significant performance gains.
What about real applications?

NAS parallel benchmarks, which are “mini applications” containing kernels for major HPC workload types

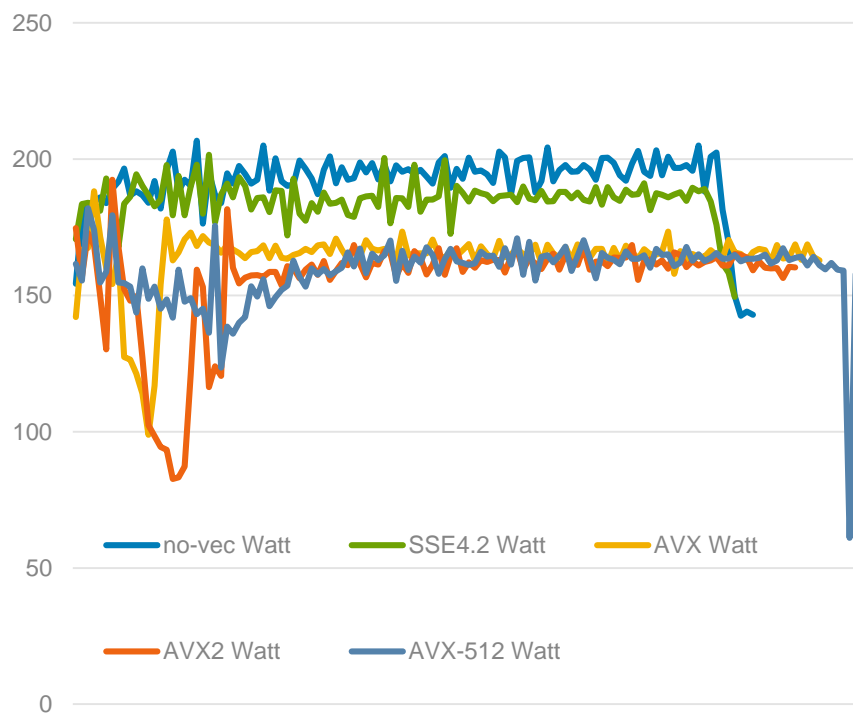
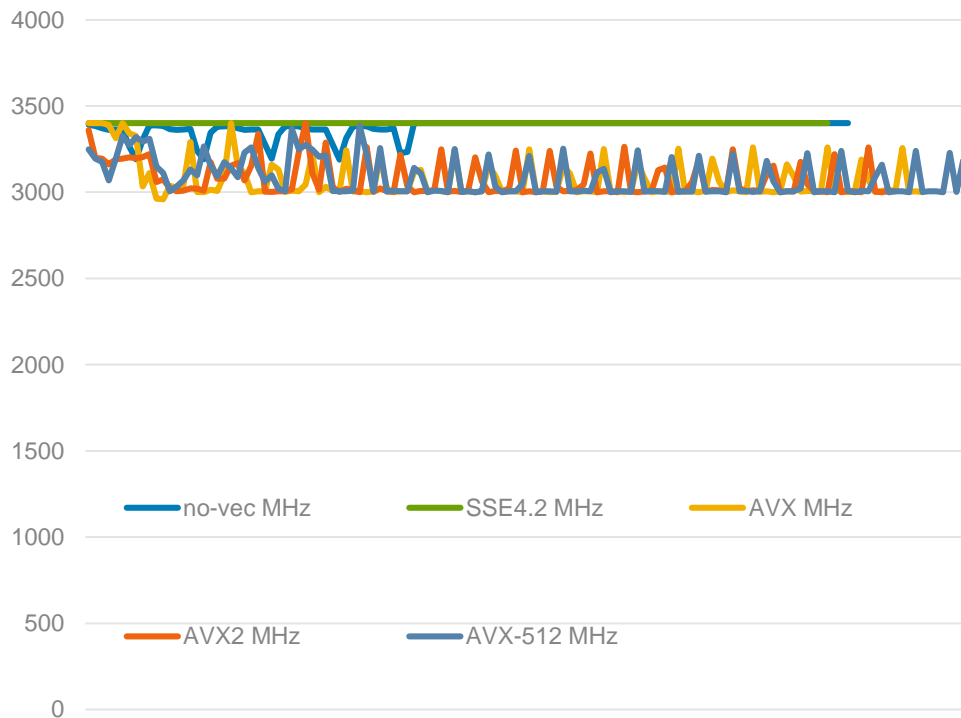
NPB kernels

KERNEL	Description	Workload
CG	Conjugate gradient	Memory latency bound
MG	Multigrid	Memory intensive
FT	Fourier transform	Compute and transpose
BT	Block tridiagonal solver	
SP	Scalar pentadiagonal solver	
LU	Lower-upper Gauss-Seidel solver	

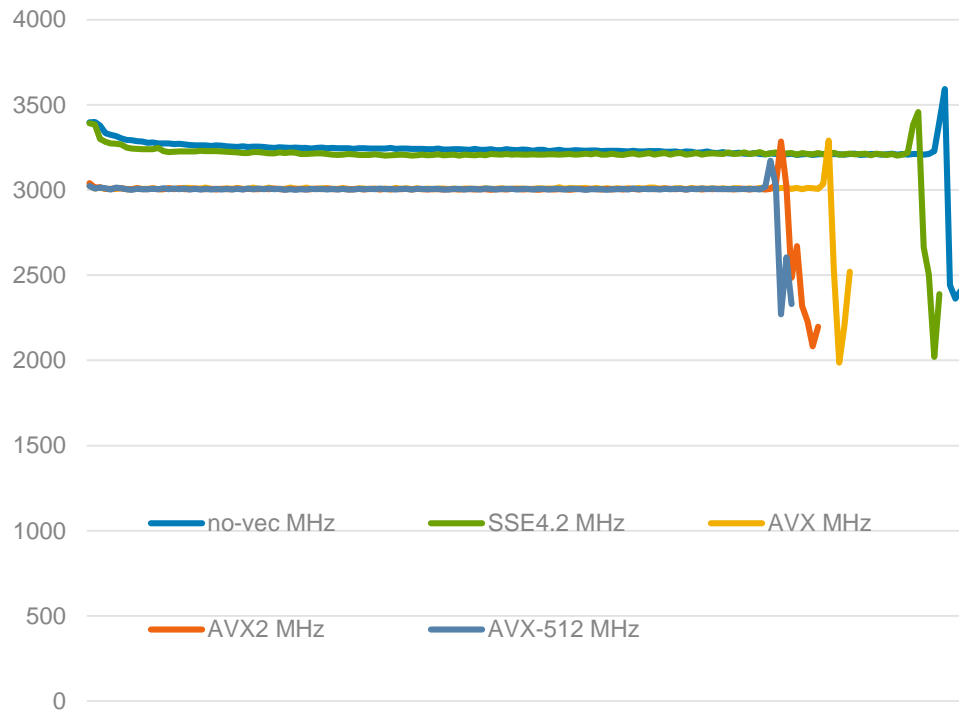
Conjugate Gradient



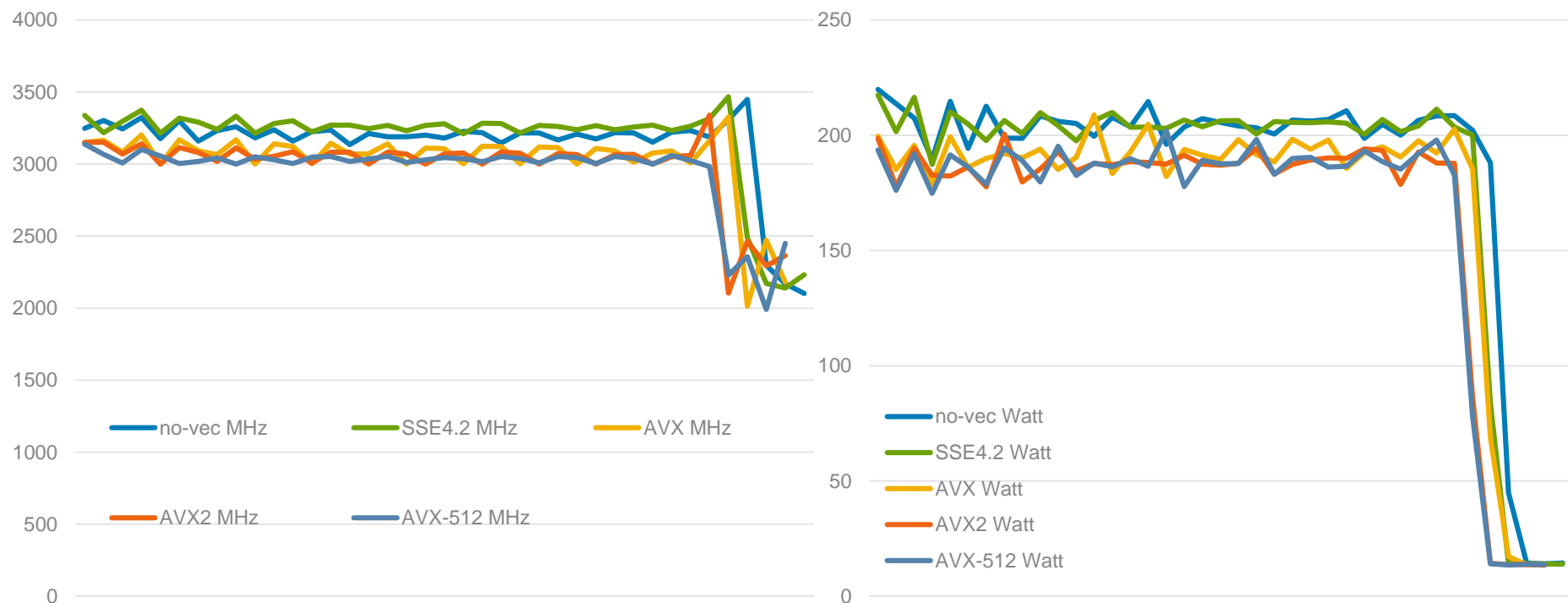
Multigrid



Block tridiagonal solver



Fourier Transformation



D  **LEMC**