

# 3rd PhD Meeting

Konstantinos Iliakis



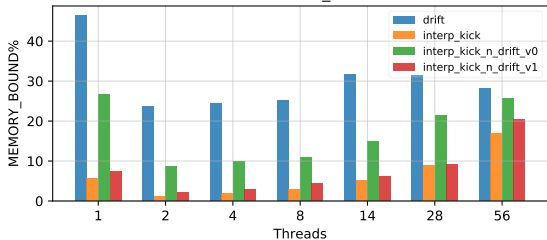
September 29, 2017

# Table of Contents

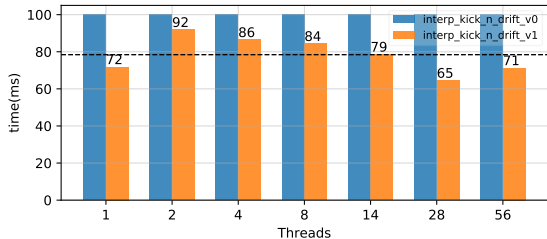
- 1 Code Optimization
  - PYPAPI Use-Case
- 2 Code Profiling
  - One-Turn Feedback Module
- 3 GPU Benchmarking
  - FFT Convolution
- 4 PAPI Deluxe Library
- 5 TechLab Resources

# Interleaved computation of kick and drift

4M-2K-MEMORY\_BOUND%

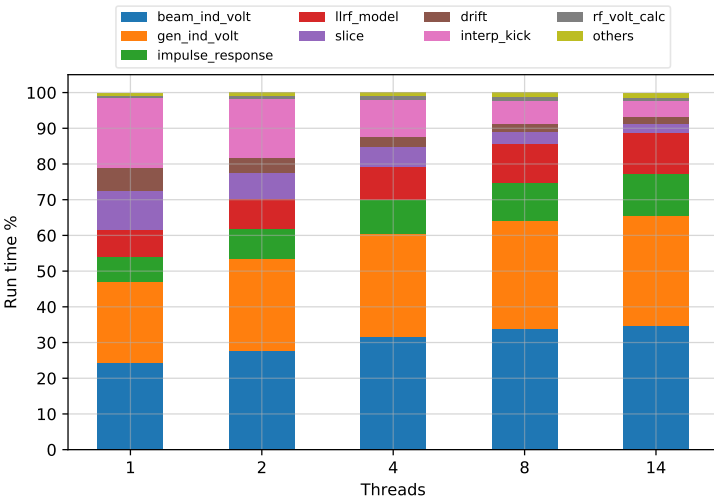


4M-2K-time(ms)



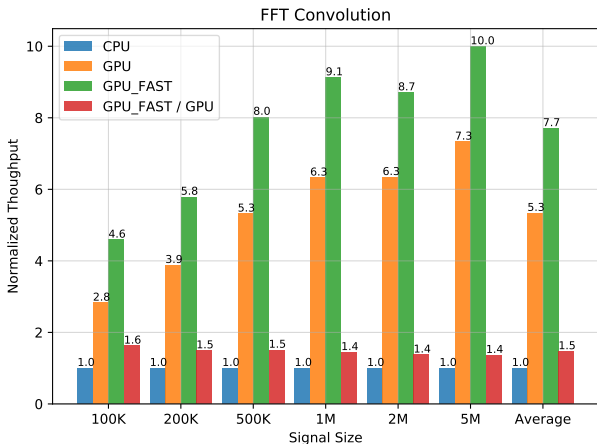
- `drift()`: In:  $dE-dt$ , Out:  $dt$ , memory bound
- `interp_kick()`: In:  $dE-dt$ , Out:  $dE$ , compute intensive
- `kick_n_drift_v0()`: First compute kick for all particles and then drift
- `kick_n_drift_v1()`: Compute kick and drift at the same time
- $MEMORY\_BOUND = \frac{STALLS\_LDM\_PENDING}{CYCLES} \%$
- Metric extracted with PYPAPI

# OTFB Execution Time Breakdown



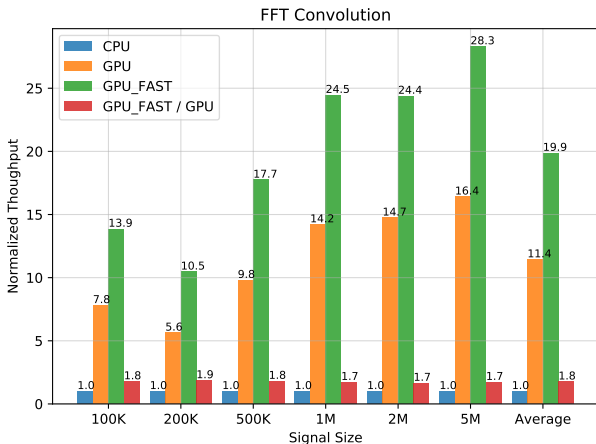
- new signal processing module implemented by Helga
- 50-65% of execution time on Convolutions (with FFTs)

# Tesla K20X



- CPU: fftconvolve from scipy
- GPU: cuFFT library, allocation and data moving CPU-GPU per turn
- GPU FAST: Data persist on the GPU, allocation and data moving only 1 per run
- All the GPU FFTs are out-of-place
- GPU Model: Tesla K20X, Peak Double Precision performance: 1.31TFlops

# Pascal P100



- CPU: fftconvolve from scipy
- GPU: cuFFT library, allocation and data moving CPU-GPU per turn
- GPU FAST: Data persist on the GPU, allocation and data moving only 1 per run
- All the GPU FFTs are out-of-place
- GPU Model: Pascal P100, Peak Double Precision performance: 4.7TFlops

# Higher-level C/C++ PAPI Library

## Supports:

- 1 Collection of Native and Preset events.
- 2 Combination of events to form metrics.
- 3 Multiplexing to allow the collection of more events.
- 4 Collection of events for single or multi-threaded modules.
- 5 Metrics found in [Intel64 and IA-32 Architecture Optimization Manual](#).

# Resources Available on TechLab(CERN)

## Current TechLab systems

Hardware type	Specs summary	HEP-SPEC06	OS & Kernel
<a href="#">x86_64 Quad Socket Xeon E5-4650</a>	4 nodes <a href="#">SandyBridge</a> and 4 nodes Westmere-EX	499.38	SLC 6.5
<a href="#">x86_64 Intel Xeon Phi 7120</a>	4 nodes, each with dual socket 8 cores SandyBridge + Xeon Phi 7120P		SLC 6.5 + Intel MPSS 3.1.2
<a href="#">GPU Nvidia Tesla K20X GPU</a>	4 nodes, each with dual socket 8 cores SandyBridge		SLC 6.5 + CUDA 5.5 Prod Release
<a href="#">GPU Nvidia GTX1080 GPU (Pascal architecture)</a>	Dual core host system with one GPU inside the box.		<a href="#">CentOS</a> 7.3 + CUDA 8.0 Prod Release
<a href="#">GPU Nvidia Pascal P100 GPU (Pascal architecture)</a>	Dual core host system with one GPU inside the box.		<a href="#">CentOS</a> 7.2 + CUDA 8.0 Prod Release
<a href="#">GPU AMD FirePro W8100</a>	1 node, dual socket 8 cores SandyBridge + AMD GPU		<a href="#">CentOS</a> 7.1
<a href="#">x86_64 Intel Atom C2750 Moonshot cartridge</a>	10 cartridges (up to 45)	53.40	SLC6 2.6.32-358.23.2.el6.x86_64 gcc 4.4.7 20120313
<a href="#">ARM64 X-Gene Moonshot cartridge</a>	X-Gene 1, 8 cores @ 2.4 GHz, 64 GB of RAM	56.52	Ubuntu 14.04, kernel 3.13, gcc 4.9.2
<a href="#">Maxeler Data Flow Engine</a>	1 node, dual socket 8 cores SandyBridge + Galava PCI-e DFE card		Maxeler OS
<a href="#">PPC64le Palmetto</a>	IBM Turismo, 4 physical cores (32 logical) @ 3 GHz, 64 GB of RAM	112.29	<a href="#">CentOS</a> 7.2.1511, kernel 3.10.0-327.el7.ppc64, gcc 4.9.2
<a href="#">PPC64le Wistron</a>	Wistron Polaris, dual socket 128 cores @ 3.325 GHz, 267 GB of RAM	602.46	Fedora 24, kernel 4.5.5-300.fc24.ppc64le, gcc 6.1.1
<a href="#">ARM64 ThunderX</a>	Dual socket 96 cores, 264 GB of RAM	342	<a href="#">CentOS</a> 7.2, kernel 4.2.0, 4.8.5
<a href="#">FPGA Altera Arria10</a>	dual socket, 40 cores (Hypert-threading), 65 GB of RAM		<a href="#">CentOS</a> Linux release 7.3.1611, kernel 3.10.0-514.21

Figure 3: For more info: [TechLab Available Systems](#)



# Thank you for your attention

