NSF/IUCRC CAC PROJECT

MONITORING, VISUALIZING, AND PREDICTING HEALTH STATUS OF HPC CENTERS

Jie Li
Doctoral Student, TTU
06/12/2020

Advisors:
Mr. Jon Hass, SW Architect, Dell Inc.
Dr. Alan Sill, Managing Director, HPCC, TTU
Dr. Yong Chen, Associate Professor, CS Dept, TTU
Dr. Tommy Dang, Assistant Professor, CS Dept, TTU

ON-CHIP PM

- On-chip performance monitoring (PM) exists on almost every microprocessor.
- PM hardware consists of a small number of registers with connections to various other parts of the chip.
- Two methods of using the PM hardware:
  - Aggregate (direct)
    - Involves reading the counters before and after the execution of a region of code and recording the difference.
    - Permits explicit, highly accurate, fine-grained measurements.
  - Statistical (indirect)
    - The PM hardware is set to generate an interrupt when a performance counter reaches a preset value. This interrupt includes the program counter (PC), the text address at which the interrupt occurred.
    - Facilitates a good high-level understanding of where and why the bottlenecks are occurring.

Ref: https://www.drdobbs.com/tools/performance-monitoring-with-papi/184406109

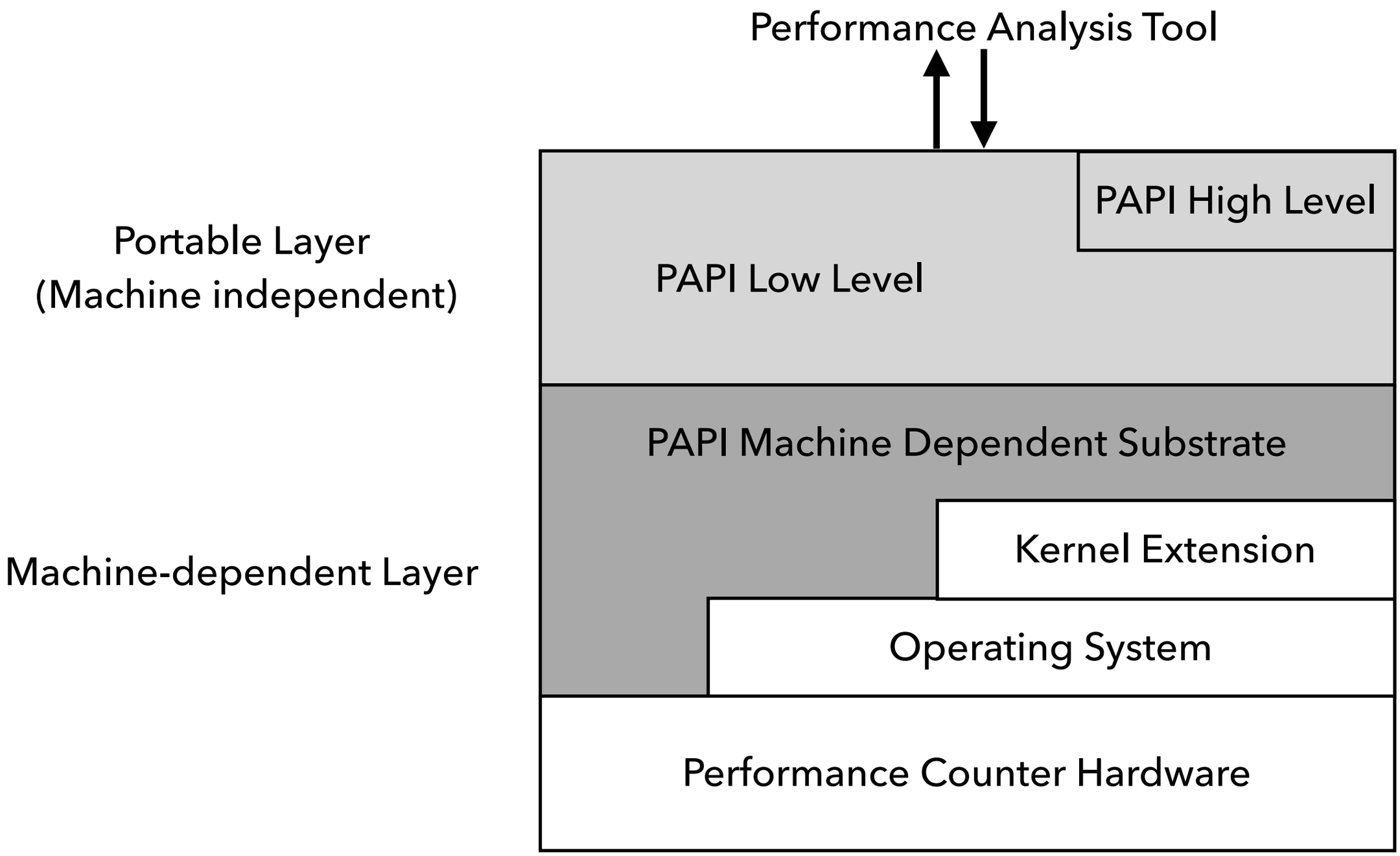
PAPI

- Performance Application Programming Interface (PAPI) library.
- Provides a consistent interface and methodology for using low-level performance counters in CPUs, GPUs, on/off-chip memory, interconnects, I/O system, and energy/power management.
- Enables monitoring of both types of performance events— hardware- and software-related events— in a uniform way.
- Implemented on a wide variety of architectures and operating systems.

AMD CPU: up to Fam17 Zeppelin Zen GPU: ROCm, ROCm-smi	arm Cortex, Cavium ThunderX, ARM64	CRAY THE SUPERCOMPUTER COMPANY Gemini and Aries interconnect, power	IBM Blue Gene Series, Q: 5-D Torus, I/O System, EMON power, energy	IBM Power 5,6,7,8,9 Power monitoring support
IBM Power9 NEST event support via Performance Co-Pilot (PCP) PAPI component	intel Westmore, Sandy/Ivy Bridge, Haswell, Broadwell, Skylake(-X), Kaby Lake, Cascadelake	intel KNC, KNL, Knights Mill including power/energy	intel RAPL (power/energy), powercap	INFINIBAND TRADE ASSOCIATION
lustre	NVIDIA Tesla, Kepler, Maxwell, Pascal, Volta	NVIDIA Power monitoring and capping support (NVML), NVLink	KVM Virtual Environment	vmware Virtual Environment

Ref: https://bitbucket.org/icl/papi/src/master/

PAPI-ARCHITECTURE



Ref: S. Browne, J. Dongarra, N. Garner, G. Ho, and P. Mucci, “A portable programming interface for performance evaluation on modern processors,” International Journal of High Performance Computing Applications, vol. 14, no. 3, pp. 189–204, 2000.

PAPI-COMPONENTS

	Component Name	Description
CPU	perf_event	Linux perf_event CPU counters (default)
	perf_event_uncore	Linux perf_event CPU uncore and northbridge (default)
	...	
GPU	cuda	CUDA events and metrics via NVIDIA CuPTI interfaces
	nvml	NVIDIA hardware counters (usage, power, temperature, fan speed, etc)
	...	
Power	powercap	Linux powercap energy measurements
	rapl	Linux RAPL energy measurements
	...	

Ref: https://bitbucket.org/icl/papi/wiki/PAPI-Overview.md

	Component Name	Description
Network	infiniband	Linux Infiniband statistics using the sysfs interface
	net	Linux network driver statistics
I/O	appio	Linux I/O system calls
	io	Linux I/O statistics from /proc/self/io
	lustre	Lustre filesystem statistics
	...	
Other	coretemp	Linux hwmon temperature and other info
	Vmware	Support for VMware (vmguest and pseudo counters)
	...	

Ref: https://bitbucket.org/icl/papi/wiki/PAPI-Overview.md

Sample events in the perf_event component:

- L1 and L2 cache hit rates, indicate how how cache-friendly a program is.
  - L1 cache hit rate: *1.0 - (PAPI_L1_DCM/(PAPI_LD_INS + PAPI_SR_INS))*
  - L2 cache hit rate: *1.0 - (PAPI_L2_DCM/PAPI_L1_DCM)*
- TLB misses: *PAPI_TLB_DM*
- Floating-point operations: *PAPI_FP_OPS*
- Completed operations per cycle (IPC): *(PAPI_TOT_INS/PAPI_TOT_CYC)*

Ref: S. Browne, J. Dongarra, N. Garner, G. Ho, and P. Mucci, “A portable programming interface for performance evaluation on modern processors,” International Journal of High Performance Computing Applications, vol. 14, no. 3, pp. 189–204, 2000.

DEMO

PLAN

- Explore other events and understand the meaning.
- Investigate the overhead of PAPI.
- Build metrics collector utilizing PAPI interface.

QUESTIONS?/COMMENTS?