The PAPI libpfm4 Transition

and unrelated

Software Prefetch Research

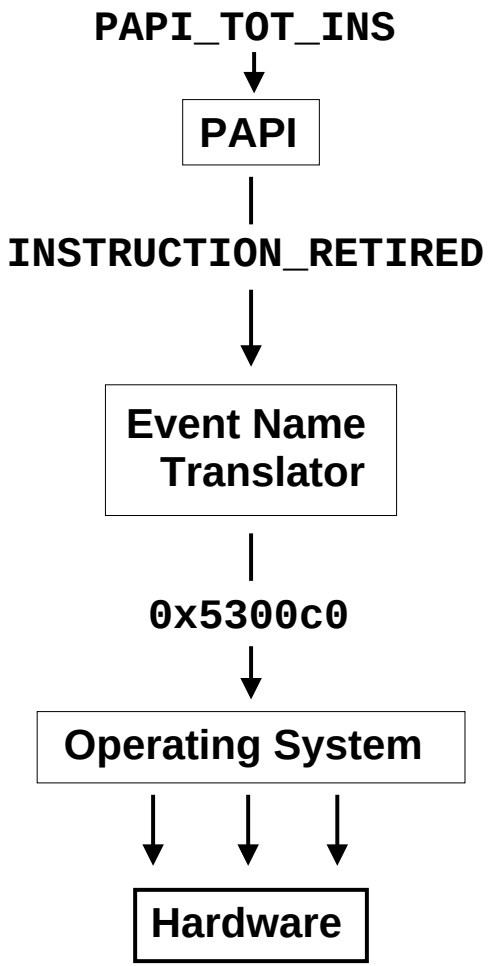
Vince Weaver

ICL Lunch Talk

11 February 2011

Part I: The PAPI libpfm4 Transition

Layers of Abstraction



libpfm3

- Used by PAPI since version 3.0 for Linux: perfctr, perfmon2, perf_events
- No longer supported
- No support for newer chips
- Not really designed for perf_events

libpfm4

- Still under development
- Supports newest processors
- Designed for perf_events
- Just incompatible enough with libpfm3 to be annoying

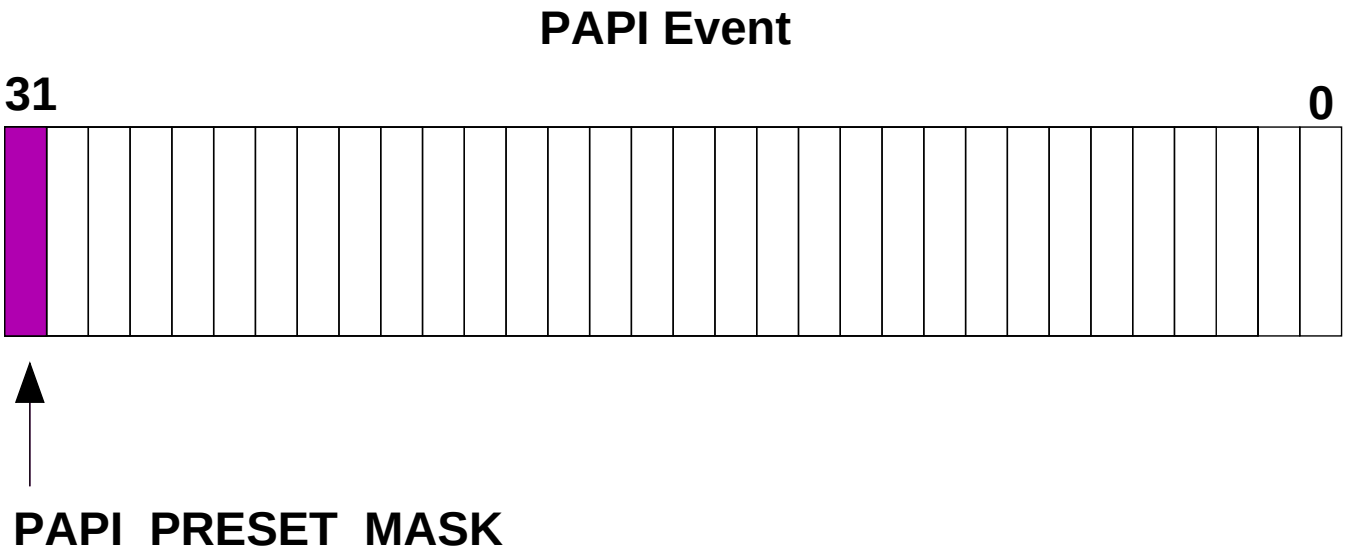
Features Not Supported by Linux 2.6.38

but PAPI/libfpm4 will support once there is support

- AMD Lightweight Profiling (LWP)
- Intel HW Cycle-count Register
- Uncore Events (Intel, AMD 15h, Power)
- Nehalem Offcore Response
- Sampling Interfaces (IBS / PEBS)
- Newer Processors (Sandy Bridge, Bulldozer)

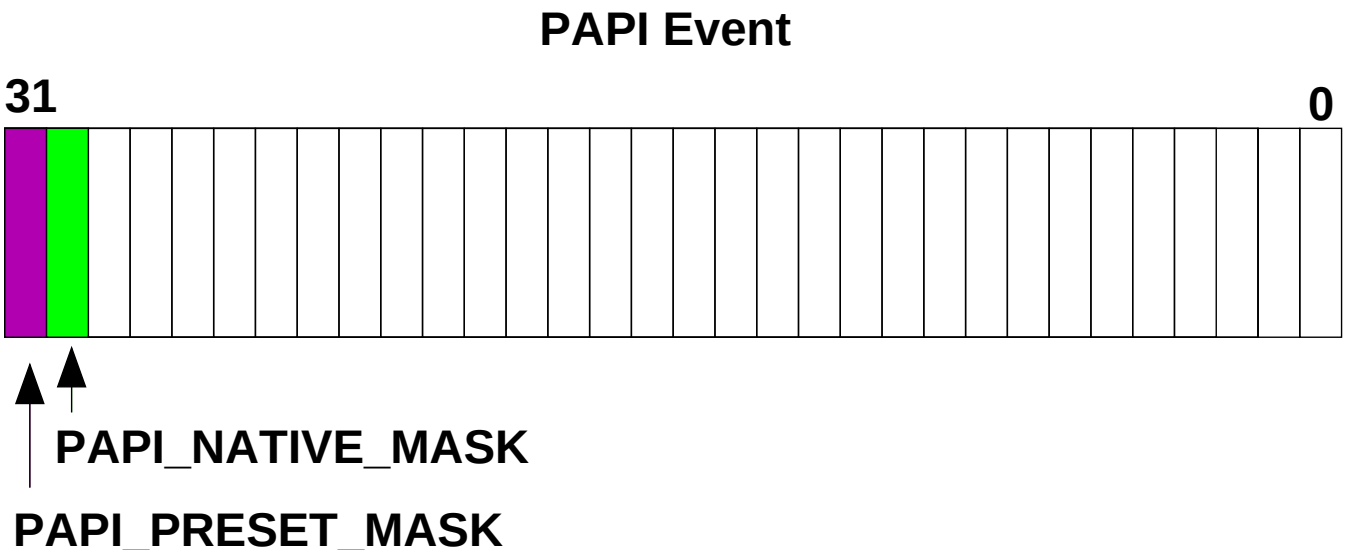
Can Current PAPI Handle All of These New Features?

Original Event Layout



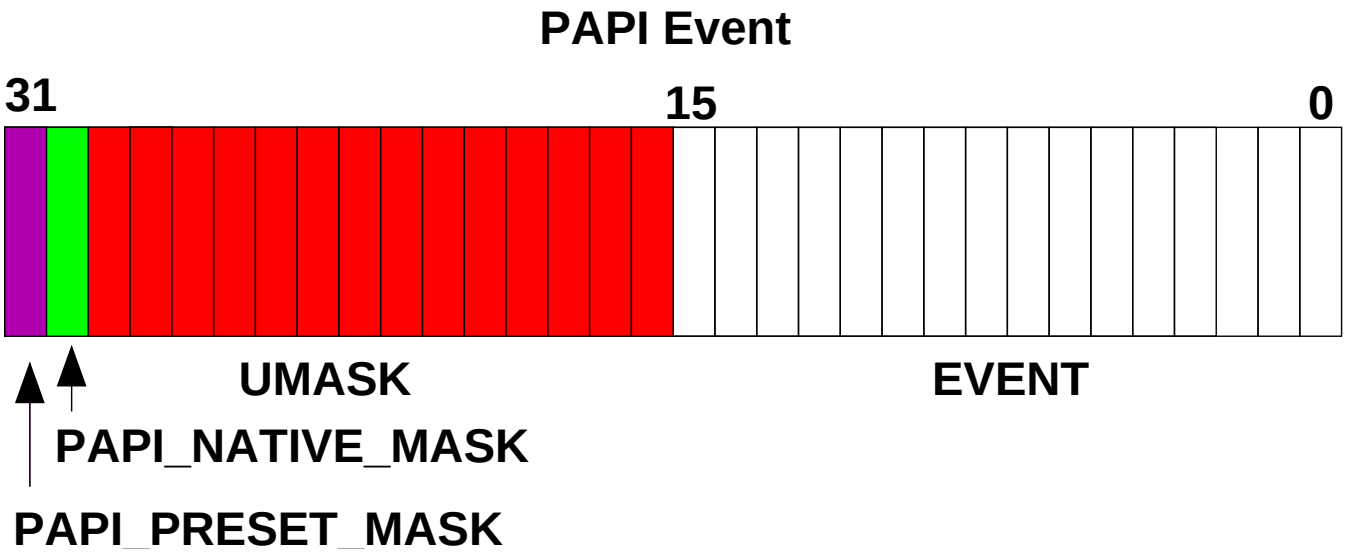
PAPI_L1_TCM

PAPI 3.0 (2004)



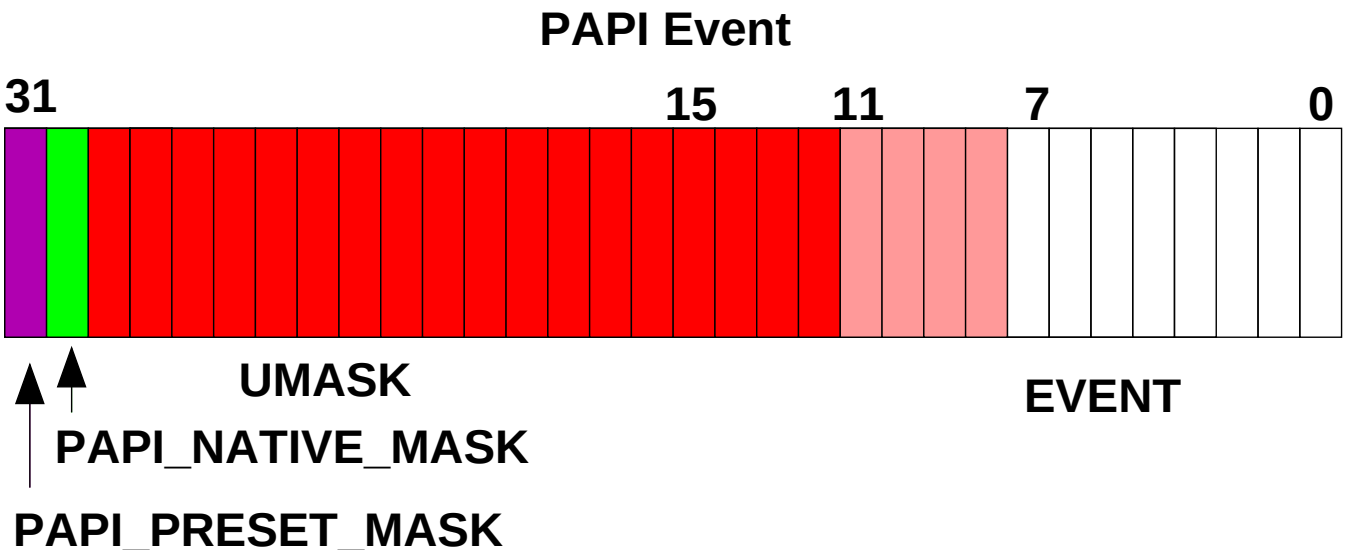
LAST_LEVEL_CACHE_REFERENCES

PAPI 3.5 (2006)



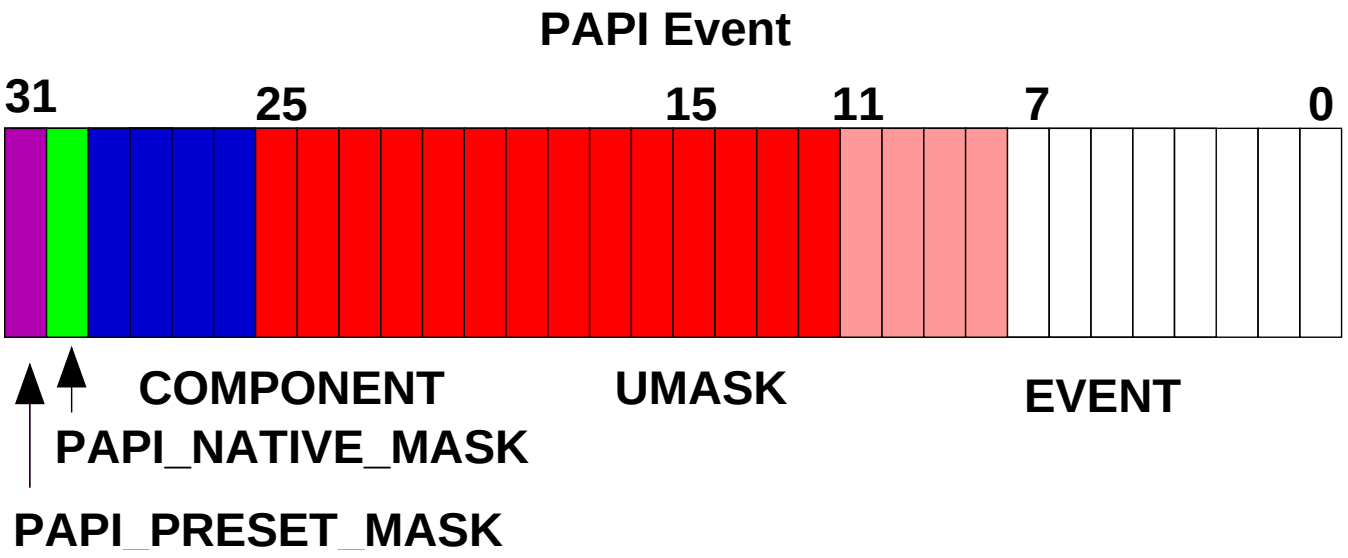
L2_RQSTS:SELF_DEMAND_MESI

PAPI 3.6 (2008)



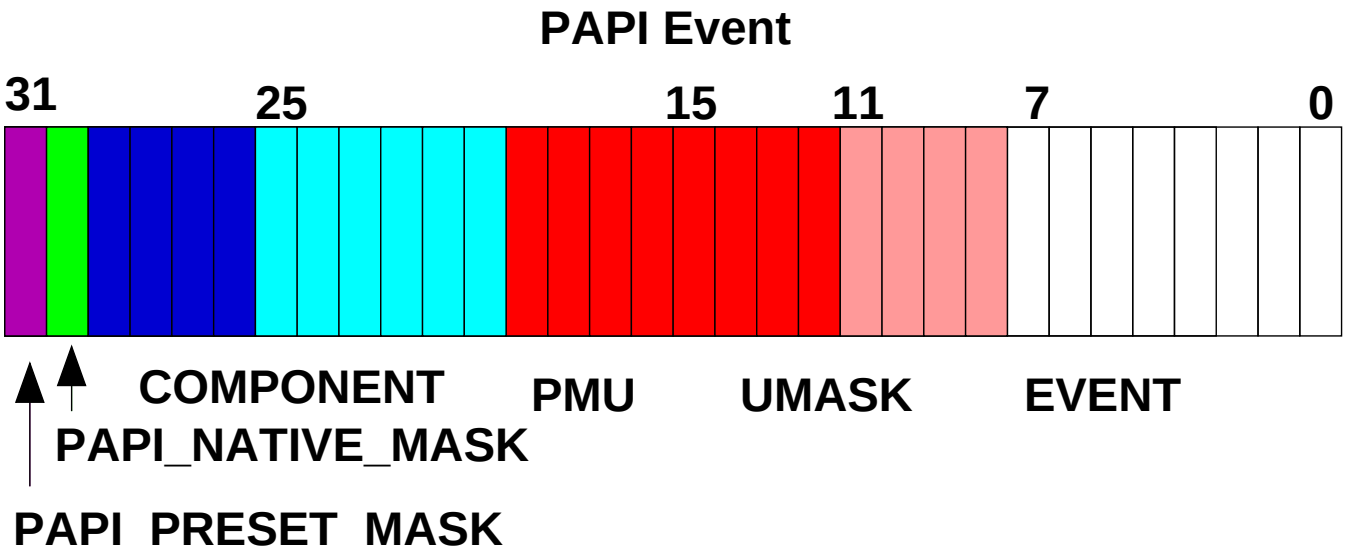
BRANCH_RETIRED:MMNP:MMNM:MMTP:MMTM

PAPI 4.0 (2010)



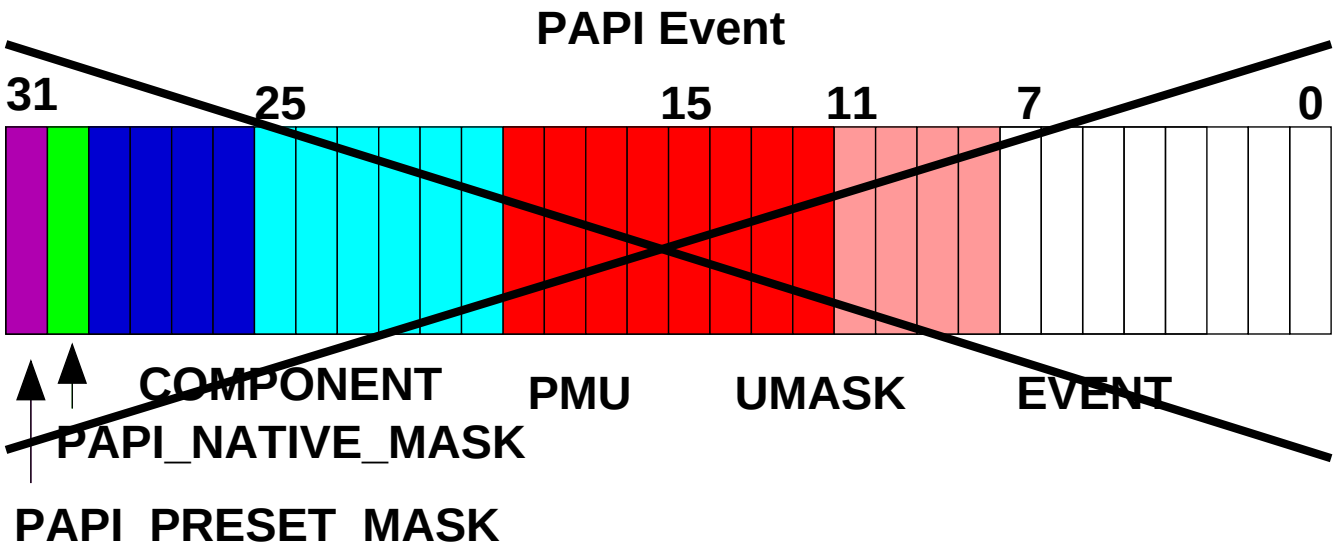
LM_SENSORS.applesmc-isa-0300.temp10.temp10_input

PAPI 5.0???




```
nhm::BR_INST_RETIRED:ALL_BRANCHES
nhm_unc::UNC_DRAM_PAGE_MISS
ix86arch::UNHALTED_CORE_CYCLES
perf::ext4:ext4_discard_blocks
```

Not Enough Bits!




```
PM_DC_PMC_9:lpid_mask=0xff:lpid=0x22:pid_mask=0x3fff:pid=0x1b2d:markin

nhm::OFFCORE_RESPONSE_0:DMND_DATA_RD:DMND_RFO:REMOTE_DRAM:LOCAL_DRAM
```

Move to libpfm4 and String-based Events

- Have a dynamically updated table containing the event names in use as full strings
- A 32-bit PAPI native event is assigned to each string, allowing backward compatibility with current PAPI interface
- Must make sure that event name lookup is not on the critical path to avoid performance regressions

Part II: Investigating Prefetching Using Hardware Performance Counters

Quick Look at Core2 HW Prefetch

- Instruction prefetcher
- L1 Data Cache Unit Prefetcher (streaming).
  Ascending data accesses prefetch next line
- L1 Instruction Pointer Strided Prefetcher.
  Looks for strided access from particular load instructions.
  Forward or Backward up to 2k apart
- L2 Data Prefetch Logic.
  Fetches to L2 based on the L1 DCU

x86 SW Prefetch Instructions (AMD)

- `PREFETCHNTA` – SSE1, non temporal (use once)
- `PREFETCHT0` – SSE1, prefetch to all levels
- `PREFETCHT1` – SSE1, prefetch to L2 + higher
- `PREFETCHT2` – SSE1, prefetch to L3 + higher
- `PREFETCH` – AMD 3DNOW! prefetch to L1
- `PREFETCHW` – AMD 3DNOW! prefetch for write

Investigating adding a PAPI_PRF_SW Pre-defined Event

- Can multiple machines count SW Prefetches?
- Does the behavior of the events match expectations?
- Will people use the preset?

Core2

- `SSE_PRE_EXEC:NTA` – counts NTA
- `SSE_PRE_EXEC:L1` – counts T0 (`fxsave`+2, `fxrstor`+5)
- `SSE_PRE_EXEC:L2` – counts T1/T2
- Problem: Only 2 counters available on Core2

AMD (Istanbul and Later)

- PREFETCH_INSTRUCTIONS_DISPATCHED:NTA
- PREFETCH_INSTRUCTIONS_DISPATCHED:LOAD
- PREFETCH_INSTRUCTIONS_DISPATCHED:STORE
- These events appear to be speculative, and won't count SW prefetches that conflict with HW prefetches

Atom

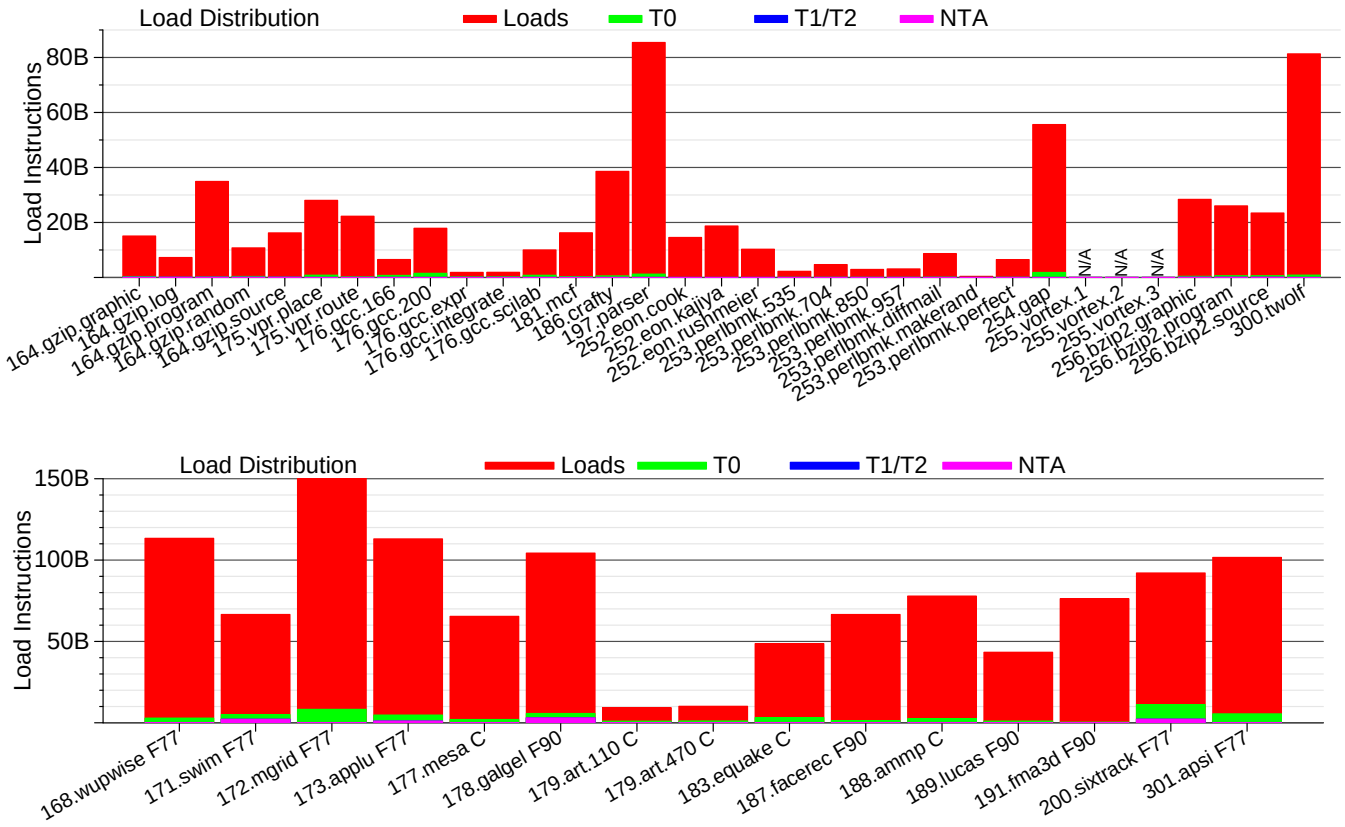
- PREFETCH:PREFETCHNTA
- PREFETCH:PREFETCHT0
- PREFETCH:SW_L2
- These events will count SW prefetches, but numbers counted vary in complex ways

Does anyone use SW Prefetch?

- gcc by default disables SW prefetch unless you specify `-fprefetch-loop-arrays`
- icc disables unless you specify `-xsse4.2 -op-prefetch=4`
- glibc has hand-coded SW prefetch in `memcpy()`
- Prefetch can hurt behavior:
  - Can throw out good cache lines,
  - Can bring lines in too soon,
  - Can interfere with the HW prefetcher

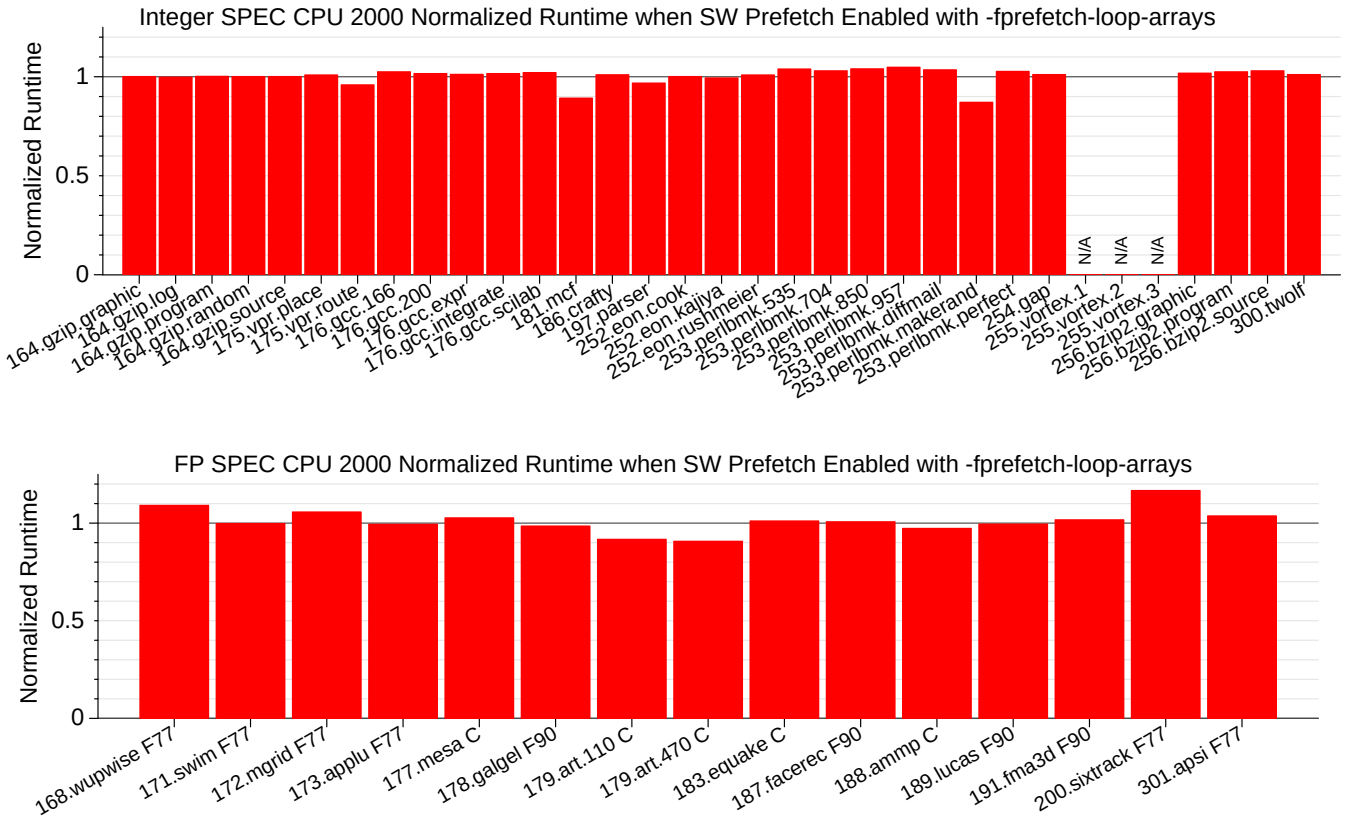
SW Prefetch Distribution

SPEC CPU 2000, Core2, gcc `-fprefetch-loop-arrays`



Normalized SW Prefetch Runtime

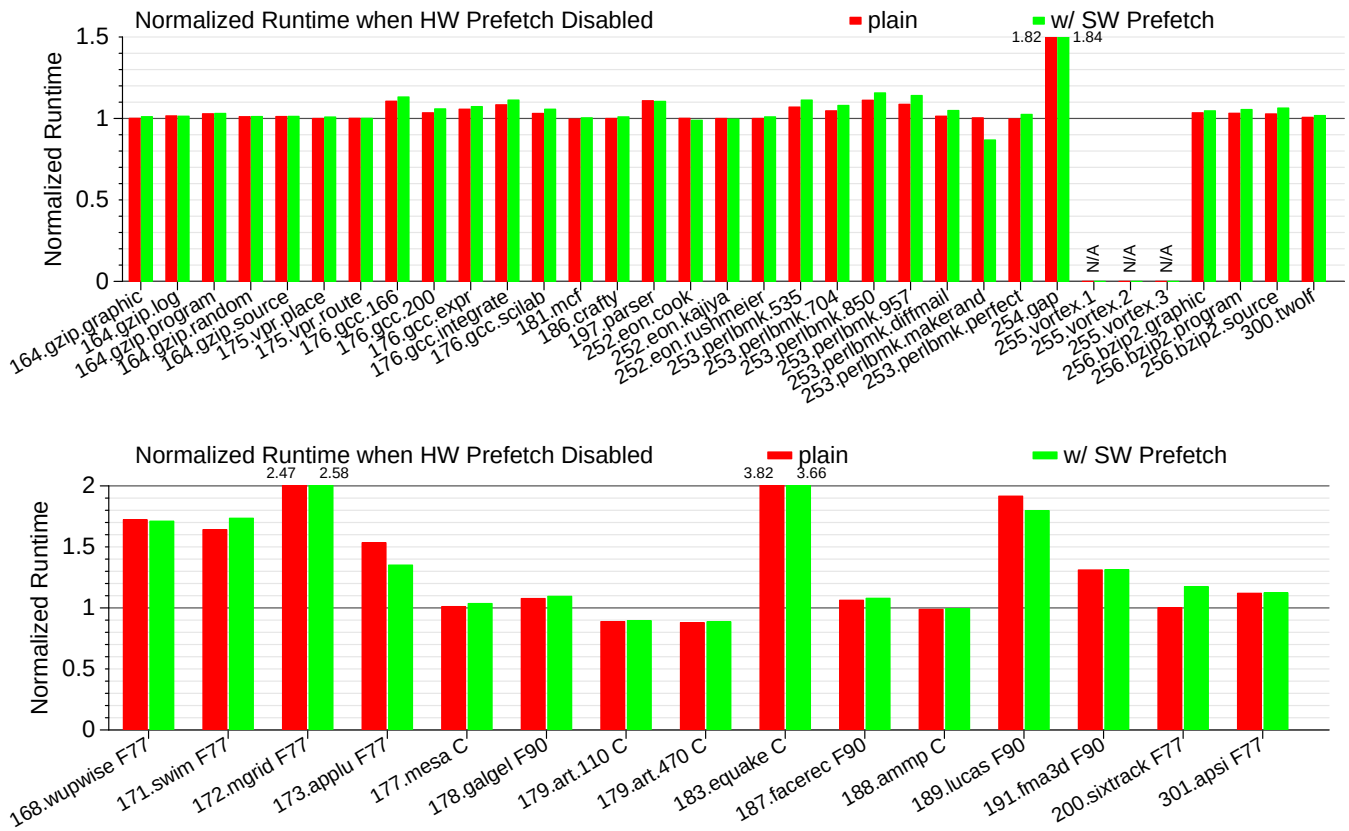
on Core2 (Smaller is Better)



The HW Prefetcher on Core2 can be Disabled

Runtime with HW Prefetcher Disabled

Normalized against Runtime with HW Prefetcher Enabled
on Core2 (Smaller is Better)



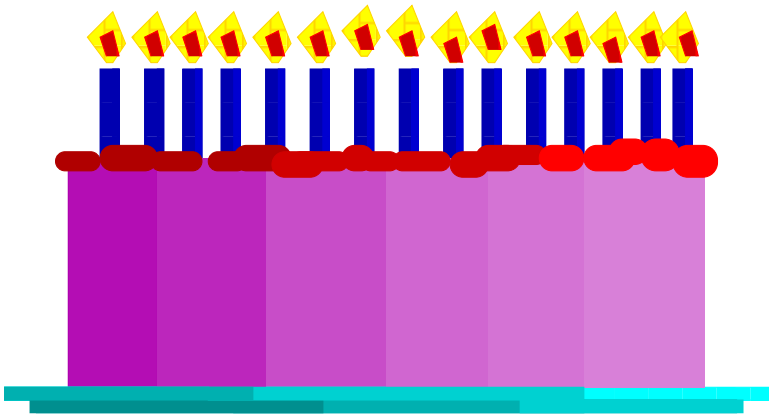
PAPI_PRF_SW Revisited

- Can multiple machines count SW Prefetches?
  Yes.
- Does the behavior of the events match expectations?
  Not always.
- Would people use the preset?
  Maybe.

Questions?

vweaver1@eecs.utk.edu

Questions?



vweaver1@eecs.utk.edu