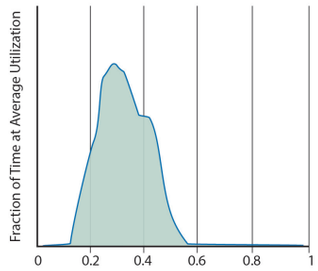
SMiTe: Precise QoS Prediction on Real-System SMT Processors to Improve Utilization in Warehouse Scale Computers

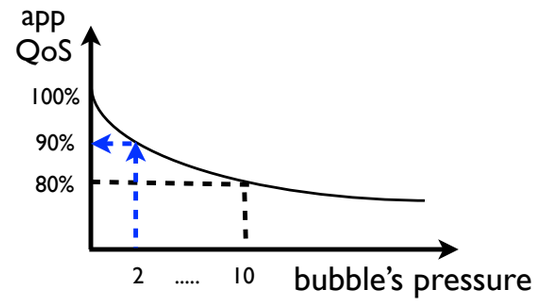
Yunqi Zhang, Michael A. Laurenzano, Jason Mars, Lingjia Tang
Clarity-Lab, Electrical Engineering and Computer Science, University of Michigan, Ann Arbor

Goal: Improve Data Center Utilization

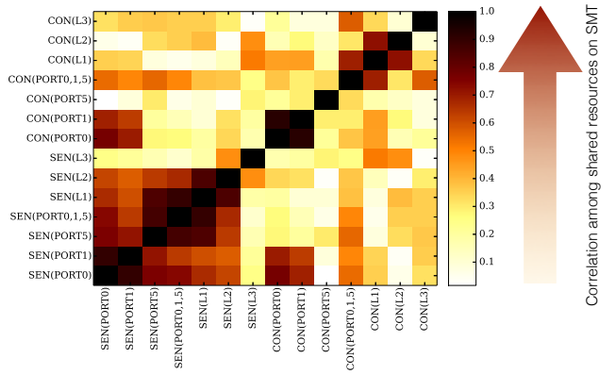


Precise interference prediction identifies "safe" co-locations to improve server utilization

SMT Co-location is Harder than CMP



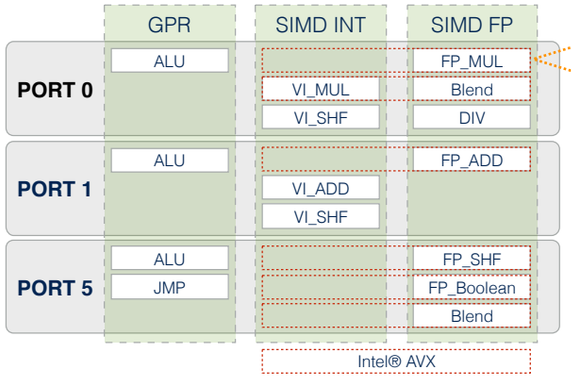
Unified approach for CMP co-location



Unified approach does **not** work for SMT

Solution: Ruler-based Methodology

Max utilization in each resource sharing dimension



Decoupled Quantification

Direct Interference Measurement

```
loop:
    mulps   %xmm0, %xmm0
    ......
    mulps   %xmm7, %xmm7
    ......
    jmp loop
```
(a) FP_MUL (PORT0)

```
loop:
    addps   %xmm0, %xmm0
    ......
    addps   %xmm7, %xmm7
    ......
    jmp loop
```
(b) FP_ADD (PORT1)

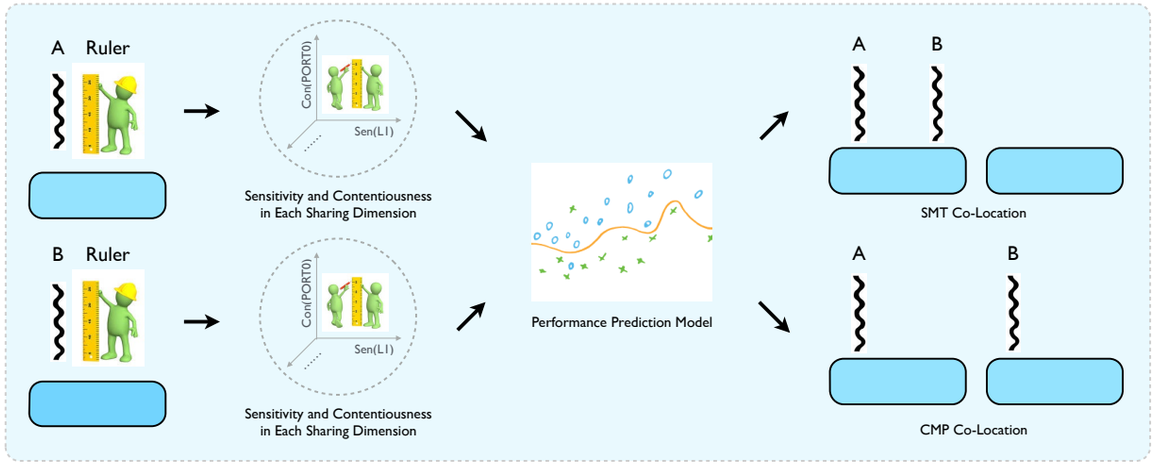
```
loop:
    shufps  %xmm0, %xmm0
    ......
    shufps  %xmm7, %xmm7
    ......
    jmp loop
```
(c) FP_SHF (PORT5)

```
loop:
    addl    %eax, %eax
    ......
    addl    %edx, %edx
    ......
    jmp loop
```
(d) INT_ADD (PORT0,1,5)

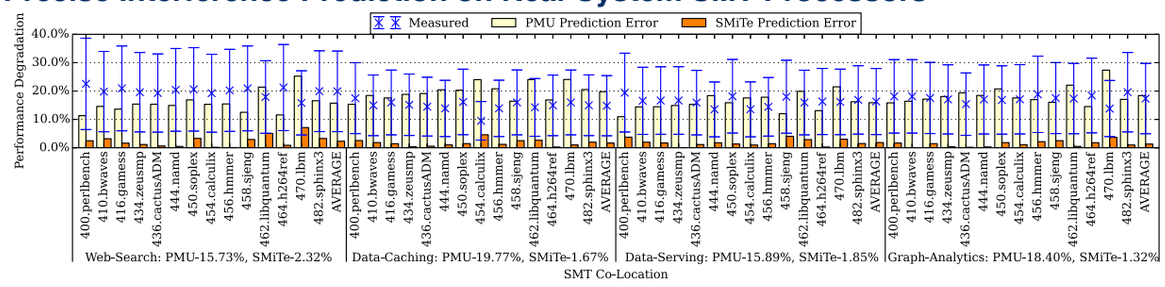
```
#define MASK 0xd0000001u
#define RAND (lfsr = (lfsr >> 1) ^ (unsigned int)(0 - (lfsr & 1u) & MASK))
......
    while (1) {
        data_chunk[RAND % FOOTPRINT]++;
        ......
        data_chunk[RAND % FOOTPRINT]++;
    }
```
(e) MEM (L1, L2 Cache)

```
......
    first_chunk = data_chunk;
    second_chunk = data_chunk + FOOTPRINT / 2;
    while (1) {
        for (i = 0; i < FOOTPRINT / 2; i += 64) {
            first_chunk[i] = second_chunk[i] + 1;
        }
        for (i = 0; i < FOOTPRINT / 2; i += 64) {
            second_chunk[i] = first_chunk[i] + 1;
        }
    }
```
(f) MEM (L3 Cache)

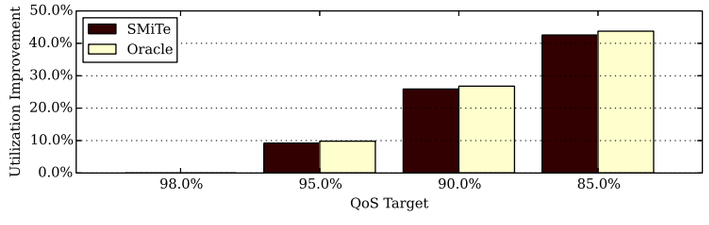
SMiTe Methodology Overview



Precise Interference Prediction on Real-System SMT Processors



Data Center Utilization Improvement



Commodity Processor

< 2% Prediction Error

42% Utilization Improvement