PRiME Power-efficient, Reliable, Many-core Embedded systems

EPSRC Engineering and Physical Sciences Research Council

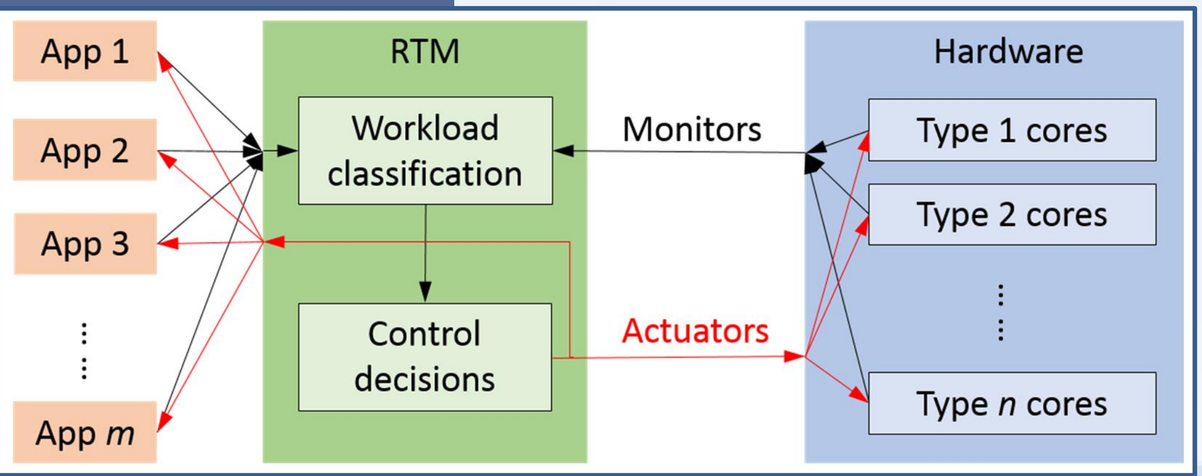
Model-free RTM with workload classification

A. Aalsaud, A. Rafiev, F. Xia, R. Shafik, A. Yakovlev

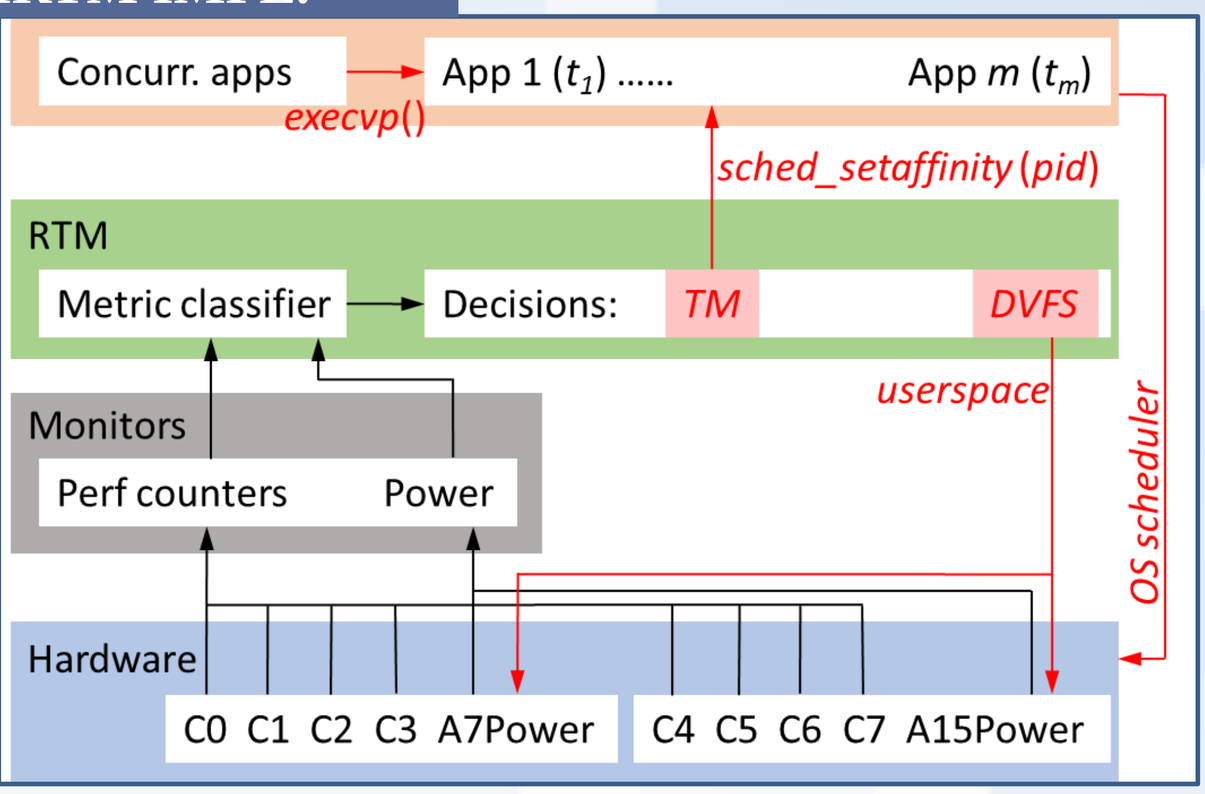
INTRODUCTION

- Runtime optimization to maximize power-normalized performance
- Trading off inter-application concurrency with performance/power
- Workload classification allows the minimization of models to reduce overhead and complexity
- RTMx runtime facility for runtime algorithm plug-ins
- Performance counters as monitors
- Decision space reduced from NP (exponential) to linear
- Negligible time overheads
- Robustness enhancements
- Runtime, per time-interval classification detects different phases of each app
- Odroid XU3 experimental validation with up to 120% performance/power improvements

RTM ARCH.



IRTM IMPL.



WL CLASSES

0: Low activity; 1: CPU-intensive;
2: CPU- & memory-intensive;
3: Memory-intensive.

METRICS

cmr	(*InstRet-Mem*)/*InstRet*
uur	*Cycles*/*ClockRef*

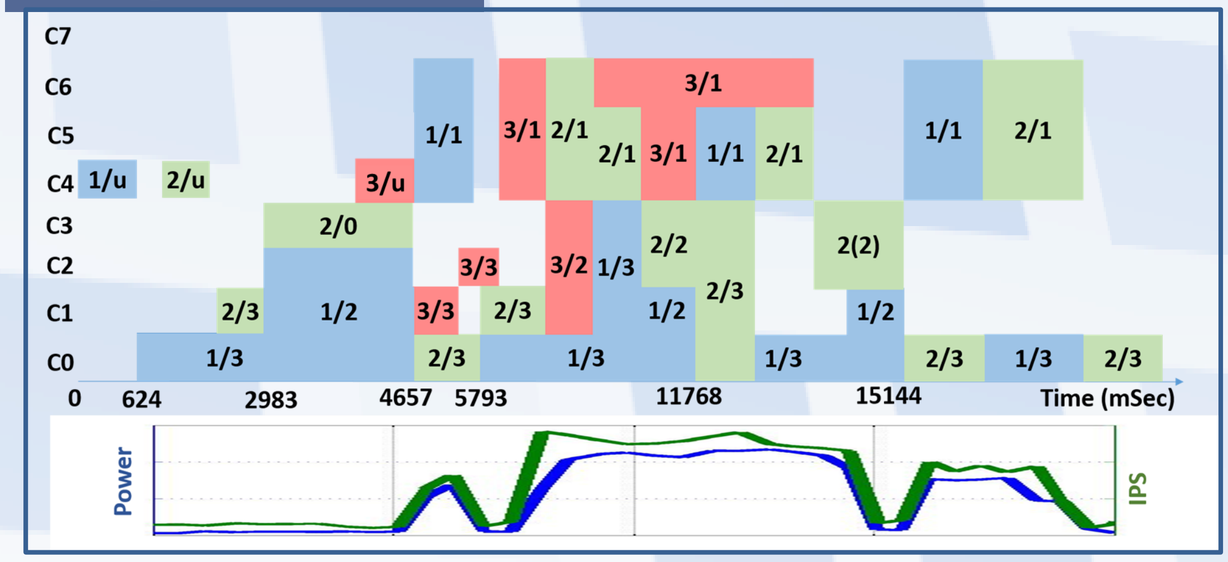
CLSSIFICATION

uur of all cores [0, 0.11]	0: Low activity
cmr per-core [0.3, 1]	1: CPU-intensive
cmr per-core [0.25, 0.3)	2: CPU+memory
cmr per-core [0, 0.25)	3: memory-intensive
out of range	4: unclassified
special class	5: low-parallelizability

DECISIONS

0	*F* = min	A7 = 1	A15 = 0
1	*F* = max	A7 = 0	A15 = max
2	*F* = min	A7 = max	A15 = 0
3	*F* = max	A7 = 1	A15 = 0
4	*F* = min	A7 = 1	A15 = 0
5	*F* = max	A7 = 0	A15 = 1

TRACE



STATE SPACE

model-based	$O((N_{A7DVFS} \cdot N_{A15DVFS}) \cdot (N_{A7} \cdot N_{A15})^{N_{app}})$
WLC-based	$O(N_{app} \cdot N_{class} + N_{core})$

IMPROVEMENT

apps	WLC (inc. OH)	LR (no OH)
fluidanimate	127%	127%
2 diff class apps	68.6%	N/A
3 diff class apps	46.6%	29.3%
2 class 3 apps	24.5%	N/A
3 class 3 apps	44.4%	36.4%
2 class 1 apps	31.0%	N/A

Imperial College London · Newcastle University · The University of Manchester · University of Southampton · intel · arm · NXP · Imagination · Microsoft Research · nmi Semiconductors to Systems · Knowledge Transfer Network ICT