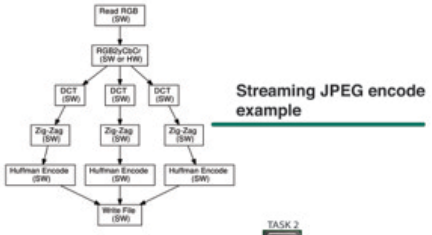
# Use of Simple Analytic Performance Models for Streaming Data Applications on Diverse Architectures


SBS Stream Based Supercomputing Lab
http://sbs.wustl.edu

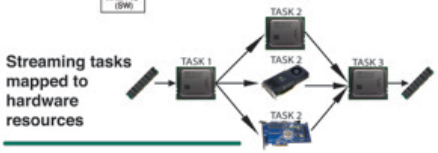
Jonathan Beard, Roger Chamberlain

Washington University in St. Louis


## Background

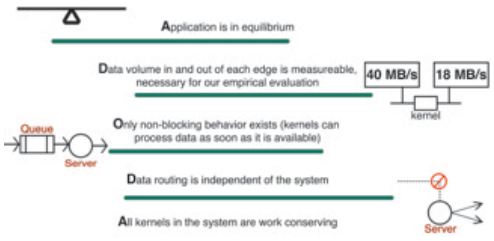



Streaming JPEG encode example




Streaming tasks mapped to hardware resources

## Assumptions




## Overview

**A** model is introduced that uses a generalized gain / loss maximum flow algorithm along with a stochastic M/M/1 queueing network to calculate throughput and estimate buffering capacity at each edge in a stream processing application.

**T**he method presented is applicable to any application that can be formulated as a directed graph.

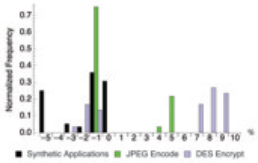
**T**he model can be used where buffering capacities along "virtual-queue" edges may not be known.

**S**imple resource sharing models are used to explore their ability to function in place of computationally expensive models.

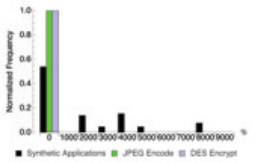
**V**alidation of the flow model approach shows that it works very well ( often within 10% ), despite the simple sharing models.

**R**esults are shown that indicate where the stochastic queueing model fails and where it can succeed.

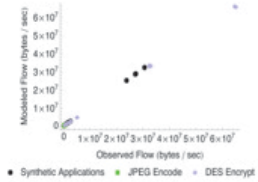
## Results



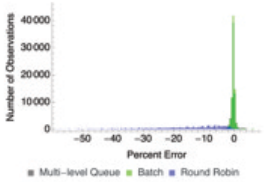

Modeled Flow Percent Error




Modeled Buffer Capacity Percent Error




Modeled vs. Observed Flow




Sharing Model % Error by Scheduler

## Testing Methodology

**S**imple sharing models tested independent of flow and queueing models.

**A**uto-Pipe [CFT+10] is used as the streaming application run-time.

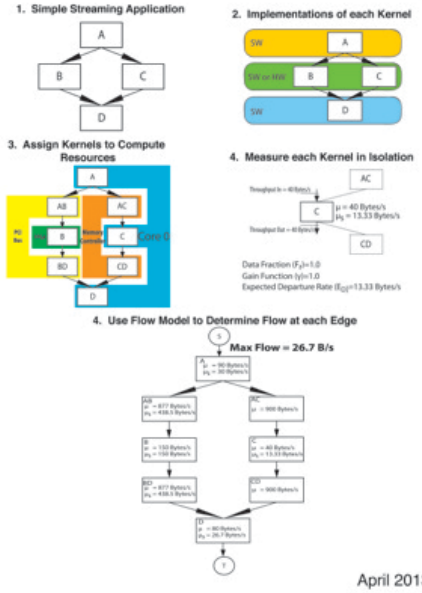
**A** tool called GraphModeler was developed to test models on streaming applications.

**M**ultiple synthetic streaming applications with both deterministic and exponential service rates were used, auto-generated kernels compiled into C and VHDL by GraphModeler then using native tools (GCC and Xilinx ISE / Synplify Premier DP).

**E**mpirical performance data collected using the TimeTrial [LSBC11] performance monitoring system.

**H**ardware:
*Machine 1* - 2 x Xilinx Virtex-4 FPGA, 12 x 2.4 GHz AMD Opteron, 32GB RAM
*Machine 2* - 4 x 3.1GHz Intel Xeon E3, 8GB RAM

## Throughput Modeling




## References and Acknowledgements


[LSBC11] Lancaster, Joseph M., EF Berkley Shands, Jeremy D. Buhler, and Roger D. Chamberlain. "TimeTrial: A low-impact performance profiler for streaming data applications." In *Application-Specific Systems, Architectures and Processors* (ASAP), 2011 IEEE International Conference on, pp. 69-76. IEEE, 2011.

[CFT+10] Chamberlain, Roger D., Mark A. Franklin, Eric J. Tyson, James H. Buckley, Jeremy Buhler, Greg Galloway, Saurabh Gayen, Michael Hall, EFBerkley Shands, and Naveen Singla. "Auto-Pipe: Streaming applications on architecturally diverse systems." *Computer* 43, no. 3 (2010): 42-49.


ISPASS


*This work supported by NSF grants CNS-0905368 and CNS-0931693 as well as by Exegy, Inc.*


April 2013