MuMMI : Multiple Metrics Modeling Infrastructure

Valerie Taylor, Xingfu Wu, Charles Lively (TAMU)
Hung-Ching Chang, Kirk Cameron (Virginia Tech)
Shirley Moore (UTEP), Dan Terpstra (UTK)

NSF CSR Large Grant

Petascale Tools Workshops 2013

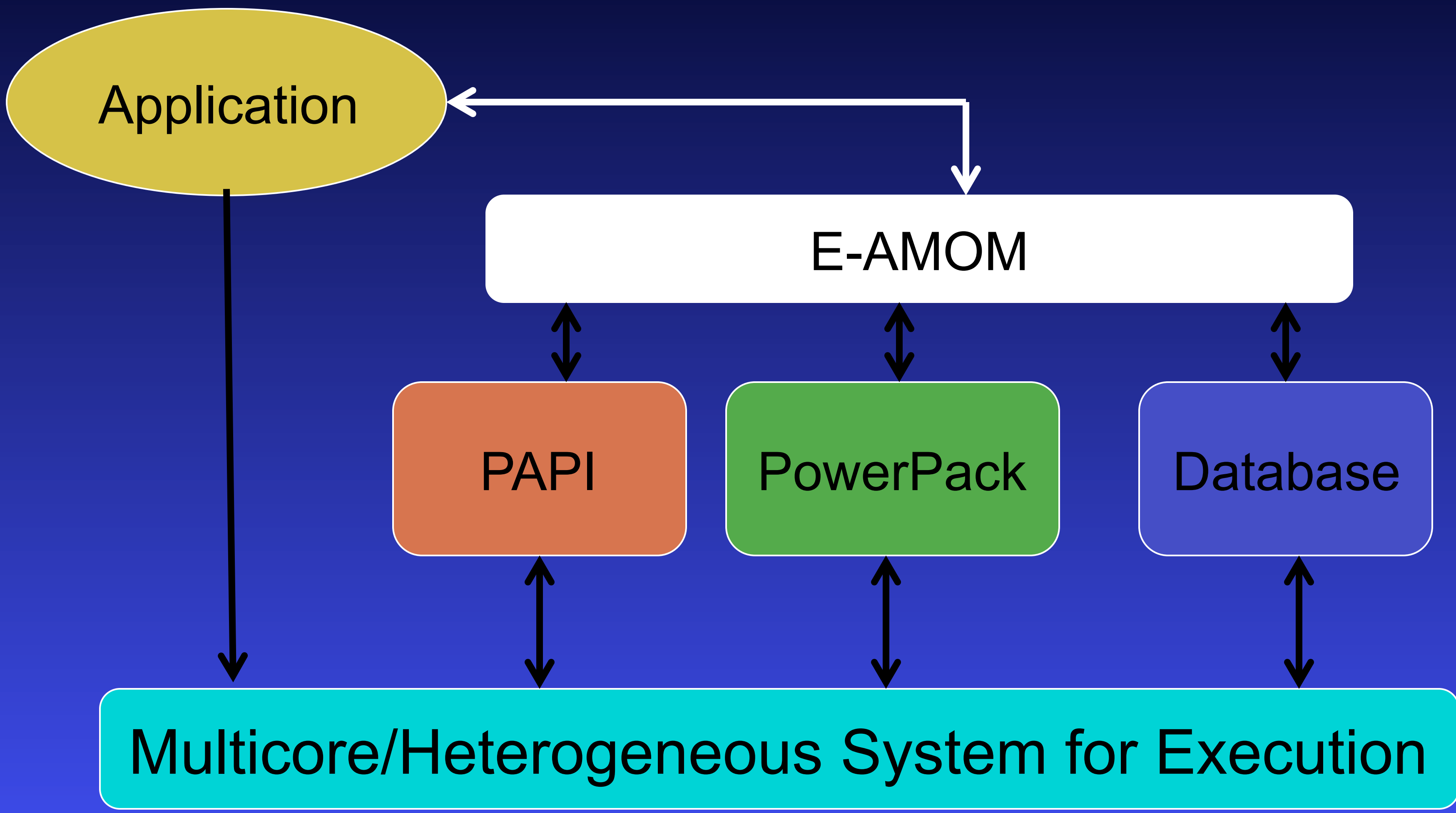
http://www.mummi.org

Motivation

Rank	Name	Vendor	# Cores	R_{MAX} (PFLOPS/S)	Power (MW)
1	Tianhe-2	NUDT	3,120,000	33.9	17.8
2	Titan	Cray	560,640	17.6	8.3
3	Sequoia	IBM	1,572,864	17.2	7.9
4	K computer	Fujitsu	705,024	10.5	12.7
5	Mira	IBM	786,432	8.16	3.95

Source: Top500 list (June 2013)

MuMMI (Multiple Metrics Modeling Infrastructure) Project



http://www.mummi.org

E-AMOM

- Start with large set of counters
- Refine set to identify important counters
- Regression analysis to obtain equations
- Focus on:
  - Runtime
  - System power
  - CPU power
  - Memory power

Counters

PAPI_TOT_INS	PAPI_L2_ICM
PAPI_FP_INS	PAPI_CA_SHARE
PAPI_LD_INS	PAPI_HW_INT
PAPI_SR_INS	PAPI_CA_ITV
PAPI_TLB_DM	PAPI_BR_INS
PAPI_TLB_IM	PAPI_RES_STL
PAPI_VEC_INS	Cache_FLD_per_instruction
PAPI_L1_TCA	LD_ST_stall_per_cycle
PAPI_L1_ICA	bytes_out
PAPI_L1_ICM	bytes_in
PAPI_L1_TCM	IPC0
PAPI_L1_DCM	IPC1
PAPI_L1_LDM	IPC2
PAPI_L1_STM	IPC3
PAPI_L2_LDM	IPC4
PAPI_TOT_INS	IPC5

First Reduction: Spearman Correlation

Example: NAS BT-MZ with Class C

Hardware Counter	Correlation Value	Hardware Counter	Correlation Value
PAPI_TOT_INS	0.9187018	PAPI_L1_ICA	0.4876423
PAPI_FP_OPS	0.9105984	PAPI_L1_ICM	0.4449848
PAPI_L1_TCA	0.9017512	PAPI_L2_ICM	0.4017515
PAPI_L1_DCM	0.8718455	PAPI_CA_SHARE	0.3718456
PAPI_L2_TCH	0.8123510	PAPI_HW_INT	0.3813516
PAPI_L2_TCA	0.8021892	PAPI_CA_ITV	0.3421896
Cache_FLD	0.7511682	Cache_FLD	0.3651182
PAPI_TLB_DM	0.6218268	PAPI_TLB_DM	0.3418263
PAPI_L1_ICA	0.5487321	PAPI_L1_ICA	0.2987326
Bytes_out	0.5187535	Bytes_in	0.26187556

Regression Analysis

Counter	Regression Coefficient
PAPI_TOT_INS	1.984986
PAPI_FP_OPS	1.498156
PAPI_L1_DCM	0.9017512
PAPI_L1_TCA	0.465165
PAPI_L2_TCA	0.0989485
PAPI_L2_TCH	0.0324981
Cache_FLD	0.026154
PAPI_TLB_DM	0.0000268
PAPI_L1_ICA	0.0000021
Bytes_out	0.000009

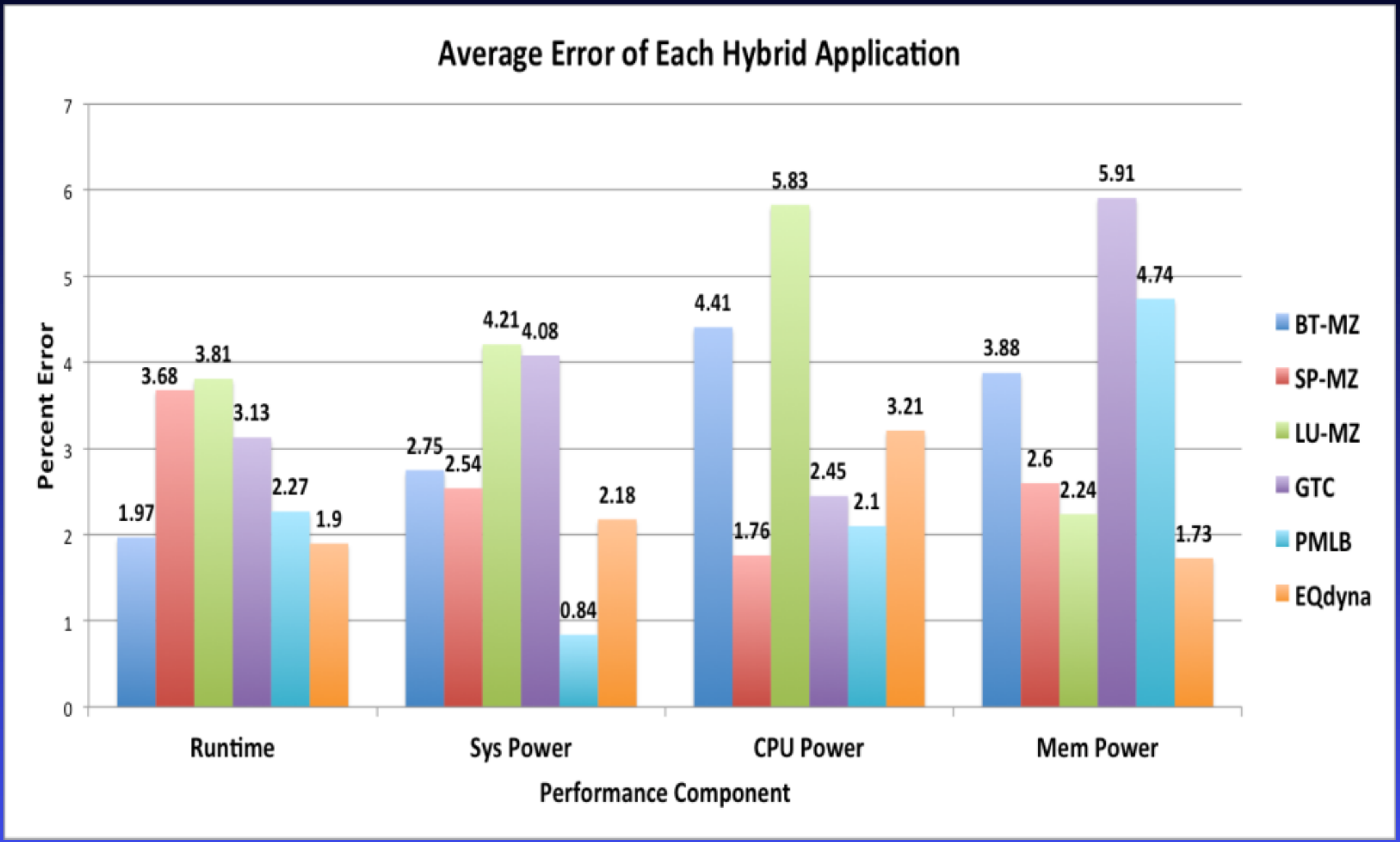
Training Set

- 12 training set points
  - Intra-node: 1x1, 1x2, 1x3 at **2.8 GHz** and 1x4, 1x6, 1x8 at **2.4 Ghz**
  - Inter-node: 1x8, 3x8, 5x8 at **2.8 Ghz** and 7x8, 9x8,10x8 at **2.4 Ghz**
- Predicted 30 points beyond of training set and validated experimentally :
  - 1x4, 1x6, 1x8, 2x8, 4x8, 6x8, 7x8, 8x8, 9x8, 10x8, 11x8, 12x8, 13x8, 14x8, 16x8 at **2.8Ghz**
  - 1x1, 1x2, 1x3, 1x5, 2x8, 3x7, 4x8, 5x8, 6x8, 8x8, 11x8, 12x8, 14x8 16x8 at **2.4 Ghz**

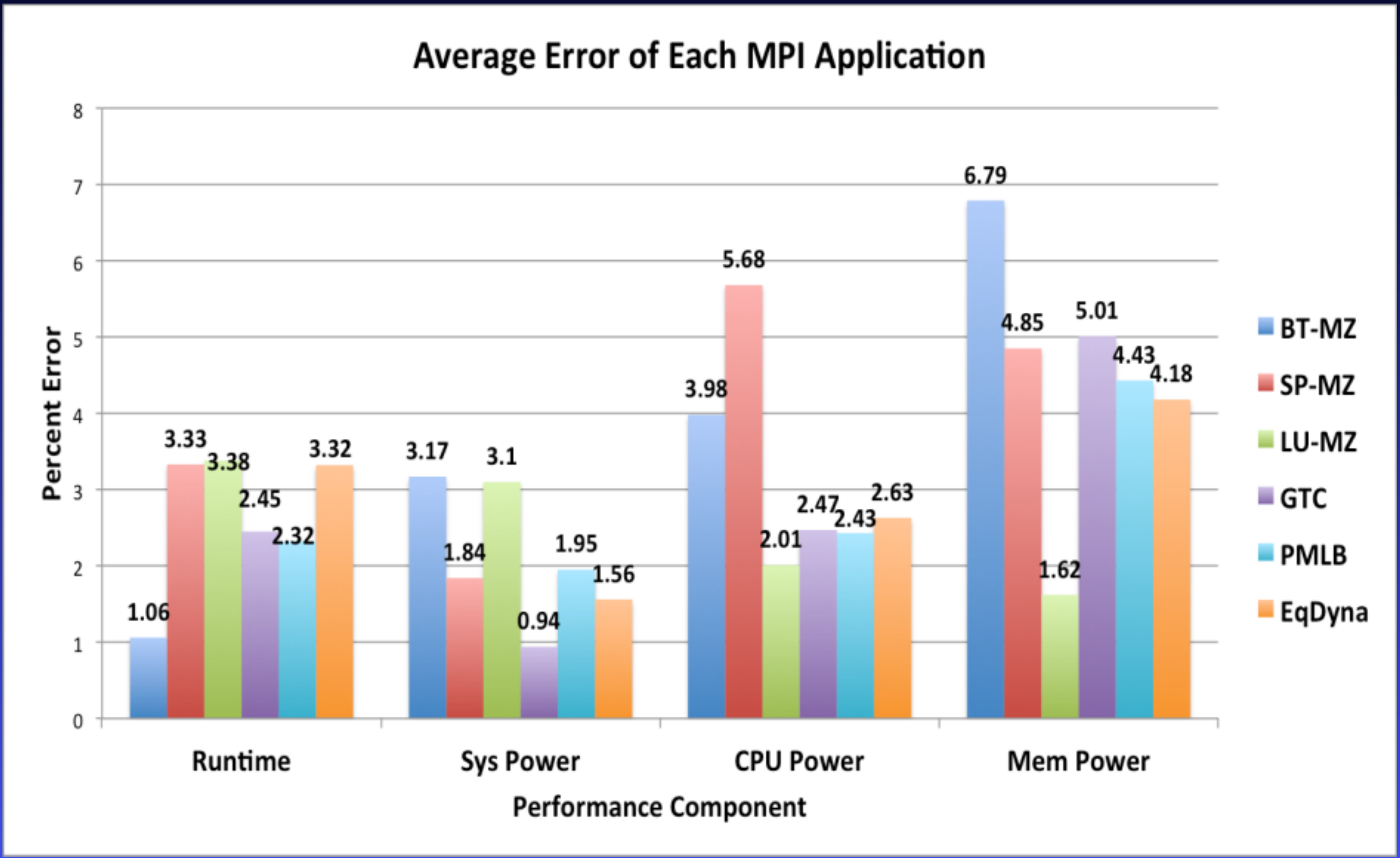
SystemG (Virginia Tech)

Configuration of SystemG	
Total Cores	2,592
Total Nodes	324
Cores/Socket	4
Cores/Node	8
CPU Type	Intel Xeon 2.8Ghz Quad-Core
Memory/Node	8GB
L1 Inst/D-Cache per core	32-kB/32-kB
L2 Cache/Chip	12MB
Interconnect	QDR Infiniband 40Gb/s

Modeling Results: Hybrid Applications



Modeling Results: MPI Applications

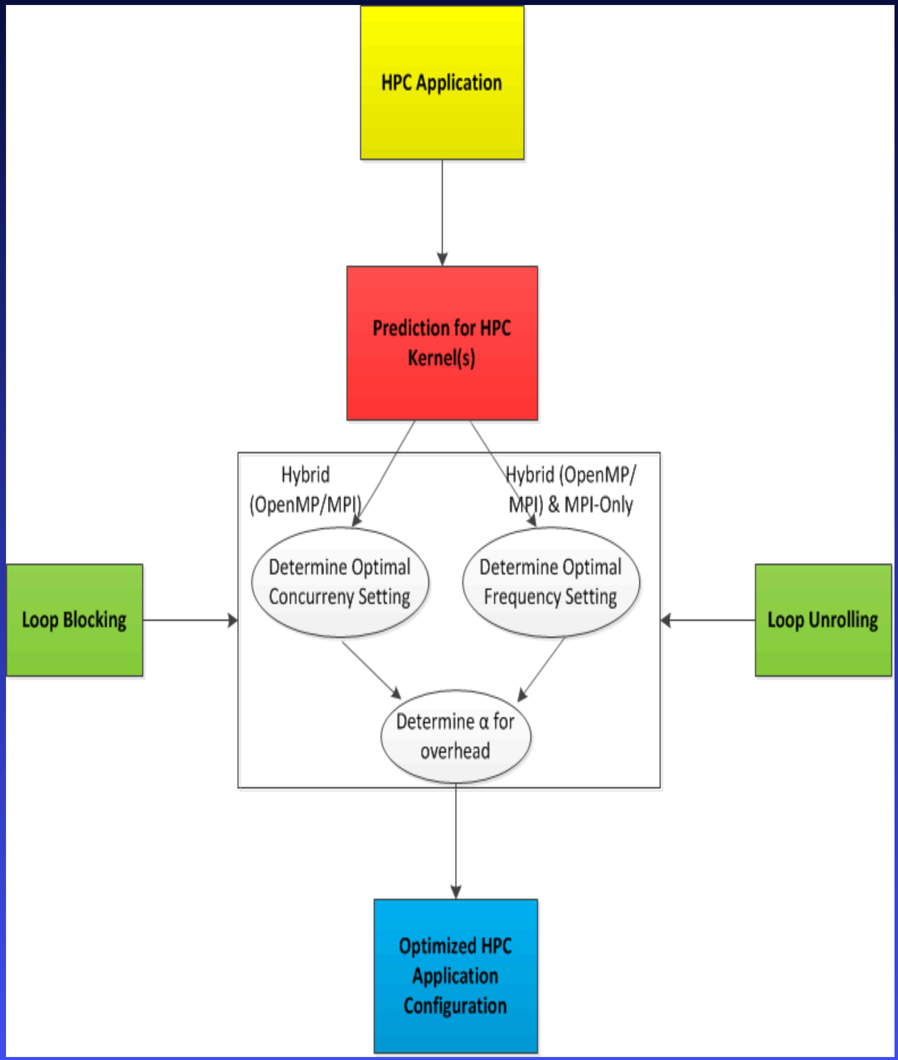


Performance-Power Optimization Techniques

- Reducing power consumption
  - Dynamic Voltage and Frequency Scaling (DVFS)
  - Dynamic Concurrency Throttling (DCT)
- Shortening application execution time
  - loop optimization: blocking and unrolling

Optimization Strategy

1. *Input: given HPC application*
2. *Determine performance of each application kernel*
3. *Determine configuration settings*
   – *setting for DVFS, DCT, or DVFS +DCT*
4. *Estimate performance*
5. *Apply loop optimizations*
6. *Use new configuration settings*



Optimization Strategy: Parallel EQdyna

- Apply DVFS
  - initialization
  - hourglass kernel
  - final kernels
- Apply DCT
  - improved configuration using 2 threads for hourglass and qdct3 kernels
- Additional loop optimizations
  - block size = 8x8
  - loop unrolling to respective kernels

Optimization Results: EQDyna

#Cores	EqDyna Type	Runtime(s)	Total Energy (KJ)	Total Power (W)
16x8	**Hybrid**	458	132.36	289.03
	Optimized-Hybrid	422 (-8.5%)	111.83 (-18.35%)	265 (-9.1%)
32x8	**Hybrid**	261	75.37	288.79
	Optimized-Hybrid	246 (-6.1%)	64.23 (-17.34%)	261.11 (-10.6%)
64x8	**Hybrid**	151	42.08	278.67
	Optimized-Hybrid	145 (-4.14%)	36.23 (-16.15%)	249.89 (-11.52%)

Optimization Strategy: GTC

- Apply DVFS
  - initialization,
  - first 25 time steps of application
  - final kernels
- Apply DCT
  - optimal configuration using 6 threads for pusher kernels after 30 time steps
- Additional loop optimizations
  - block size = 4x4 (100ppc)

Optimization Results: Hybrid GTC

#Cores	GTC Type	Runtime(s)	Total Energy (KJ)	Total Power (W)
16x8	**Hybrid**	453	132.82	293.19
	Optimized-Hybrid	421 (-7.6%)	116.34 (-14.16%)	276.35 (-6.1%)
32x8	**Hybrid**	455	134.03	294.58
	Optimized-Hybrid	424 (-7.31%)	118.44 (-13.16%)	279.35 (-5.45%)
64x8	**Hybrid**	436	128.53	294.79
	Optimized-Hybrid	423 (-3.1%)	114.72 (-12.03%)	271.12 (-8.73%)

Future Work

- Energy-Aware Modeling
  - Performance models of CPU+GPGPU systems
  - Support additional power measures: IBM EMON API for BG/Q, Intel RAPL, NVIDIA Power Management
  - Collaborations with Score-P

- Additional Energy-Aware Optimizations
  - Exploration the use of correlations among counters to provide optimization insights
  - Exploring different classes of applications