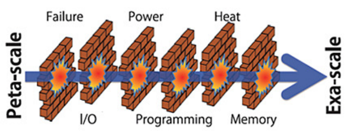
# Toward Smart HPC through Active Learning and Intelligent Scheduling

A Team led by Zhiling Lan at Illinois Institute of Technology

## Problem Statement

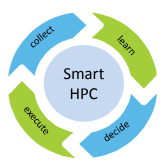
As high performance computing (HPC) continues to grow in scale, energy and resilience become first-class concerns, in addition to the pursuit of performance. These concerns demand significant changes in many aspects of the system stack including **resource management and job scheduling** (aka batch scheduling). Major issues with existing batch scheduling:

- Static
- CPU focused
- Increasing concern of data movement
- Growing demand for energy efficiency
- Dramatic demand for resilience
- Emerging demand for dynamic time-power-resilience tradeoff



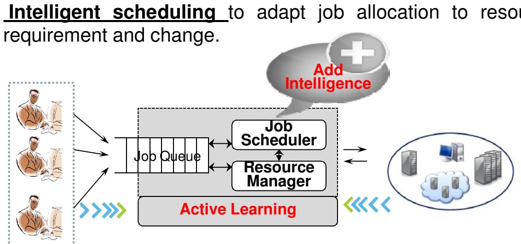

Various walls to go through

## Project Goal




We envision **smart HPC** in which information about resources and applications will be automatically gathered, analyzed, and acted on for improving performance, resilience, and energy efficiency. Specifically, we aim to design and develop a framework named ***SPEaR***:

- **Active learning** to automatically discover and predict job behavior and resource availability.
- **Intelligent scheduling** to adapt job allocation to resource requirement and change.




SPEaR : Scheduling for Performance, Energy, and Resilience efficiency

## Team Members

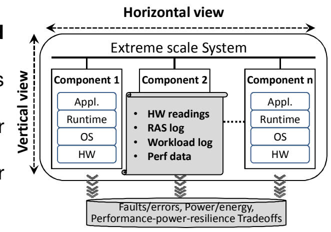
- **Principal investigator:**
  Zhiling Lan
- **Ph.D. students:**
  X. Yang, S. Wallace, E. Berrocal, Y. Fan
- **Collaborators:**
  M. Papka, S. Coghlan, W. Allcock, V. Vishwanath, P. Rich, V. Morozov, J. Jenkins, M. Mubarak, R. Ross, F. Cappello, S. Di, L. Bautista-Gomez (Argonne)
  W. Tang, N. Desai, and L. Yu (Google)

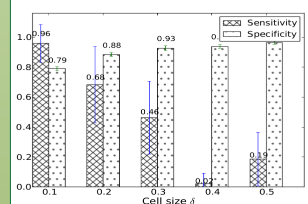
## Reference


1. E. Berrocal, L. Yu, S. Wallace, M. Papka, and Z. Lan, "Exploring Void Search for Fault Detection on Extreme Scale Systems", ***IEEE Cluster'14 [Best Paper]***, 2014.
2. E. Berrocal, L. Bautista-Gomez, S. Di, Z. Lan, and F. Cappello, "Exploring Partial Replication to Improve Lightweight Silent Data Corruption Detection for HPC Applications", ***Euro-Par'16***, 2016.
3. S. Wallace, X. Yang, V. Vishwanath, W. Allcock, S. Coghlan, M. Papka, and Z. Lan, "A Data Driven Scheduling Approach for Power Management on HPC Systems", ***IEEE/ACM SC'16***, 2016.
4. L. Yu, Z. Zhou, S. Wallace, M.E, Papka, and Z. Lan, "Quantitative Modeling of Power Performance Tradeoffs on Extreme Scale Systems", ***Journal of Parallel and Distributed Computing (JPDC)***, 2015.
5. Z. Zhou, X. Yang D. Zhao, P. Rich, W. Tang, J. Wang, and Z. Lan, "I/O-Aware Batch Scheduling for Petascale Computing Systems", ***Cluster'15***, 2015.
6. Z. Zhou, X. Yang, Z. Lan, P. Rich, W. Tang, V. Morozov, and N. Desai, "Improving Batch Scheduling on Blue Gene/Q by Relaxing 5D Torus Network Allocation Constraints", ***IPDPS'15***, 2015. [an extended version appears on ***IEEE TPDS'16***].
7. X. Yang, J. Jenkins, M. Mubarak, R. Ross, and Z. Lan, "Watch Out for the Bully! Job Interference Study on Dragonfly Network", ***IEEE/ACM SC'16***, 2016.

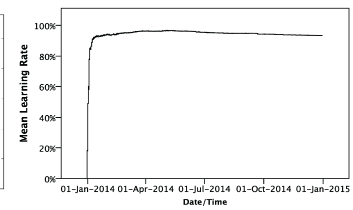

## Active Learning

- Exploring void search for **hard error detection** [1]
- Developing lightweight detectors for **SDCs** [2]
- Exploring statistic methods for **fast power profile learning** [3]
- Developing colored Petri nets for **tradeoff modeling** [4]



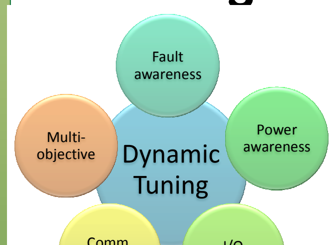




Fault detection accuracy by using void search on an environmental data collected from Mira [1]

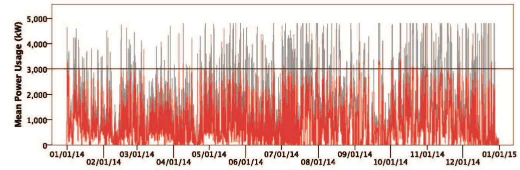



Learning of job power profiles on Mira, which is over 94% after 26 days of the operation [3]

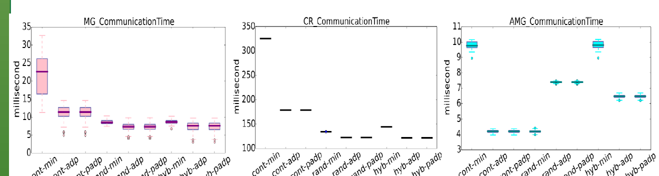
## Intelligent Scheduling




- **Power awareness**: dynamic learn job power profiles and control system-wide power consumption under a user-define cap with a minimal impact [3]
- **I/O awareness**: coordinate ongoing I/O requests from user jobs for available network bandwidth [5]
- **Comm awareness**: to allocate system resources according to coarse-grained application communication patterns [6,7]




*Power consumption using the original Mira scheduler (light gray) w/o power capping versus **our power aware scheduler** (red). The relative degradation caused by our design to job wait time and system utilization is less than 3% [3].*




***Hybrid job placement on Dragonfly** aids in reducing the worst-case performance degradation for less communication-intensive applications while retaining the performance of communication-intensive ones [7].*

## Acknowledgement

This work is supported in part by US National Science Foundation grant **CCF-1422009**. It is also funded in part and used resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under contract **DE-AC02-06CH11357**.