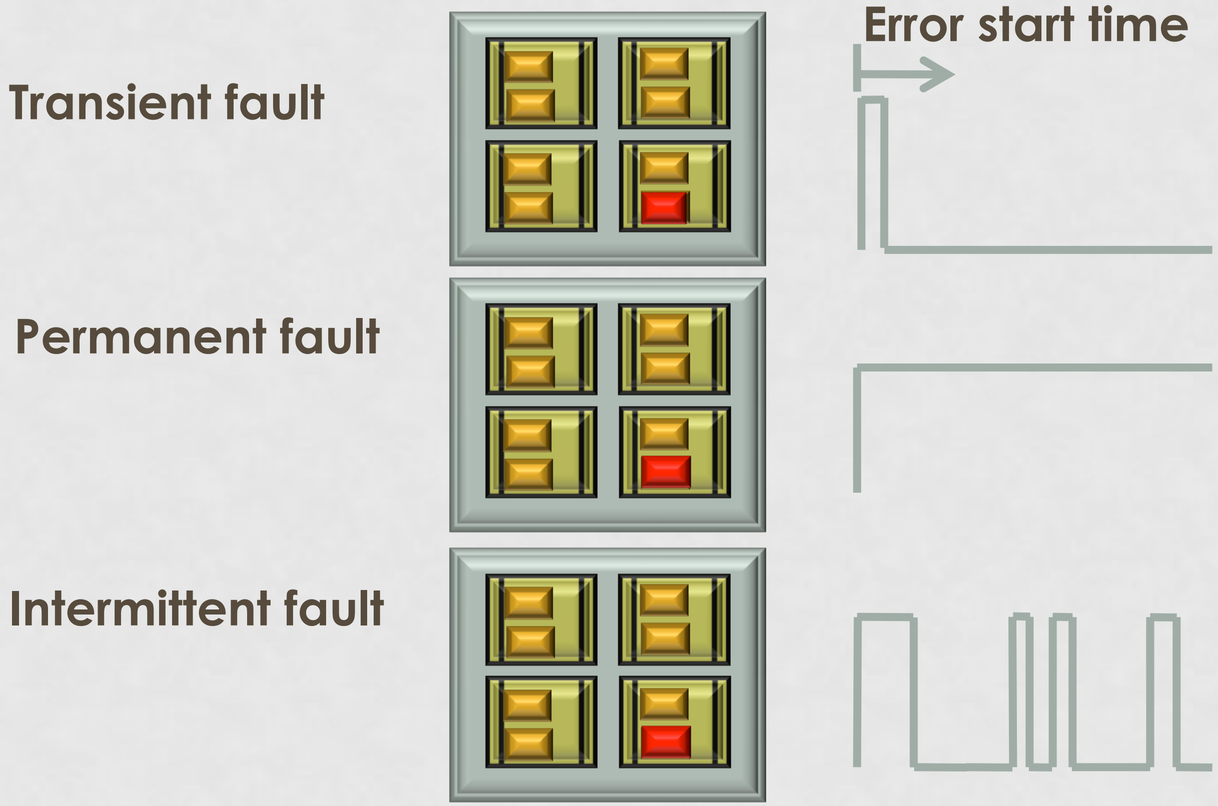
INTERMITTENT HARDWARE ERRORS RECOVERY: MODELING AND EVALUATION

LAYALI RASHID, KARTHIK PATTABIRAMAN AND SATHISH GOPALAKRISHNAN

INTERMITTENT FAULTS-DEFINITION

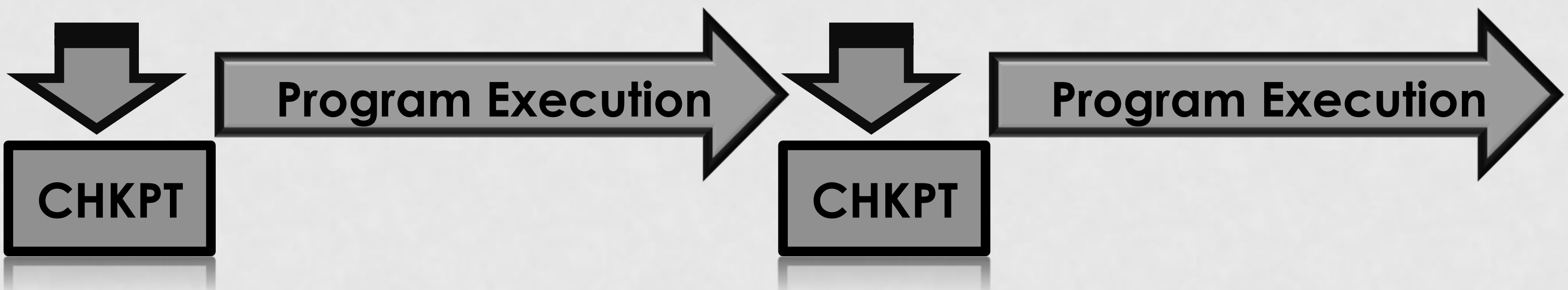
- Hardware errors that appear non-deterministically at the same microarchitectural location.
- 40% of the real-world failures in processors are caused by intermittent faults [1].



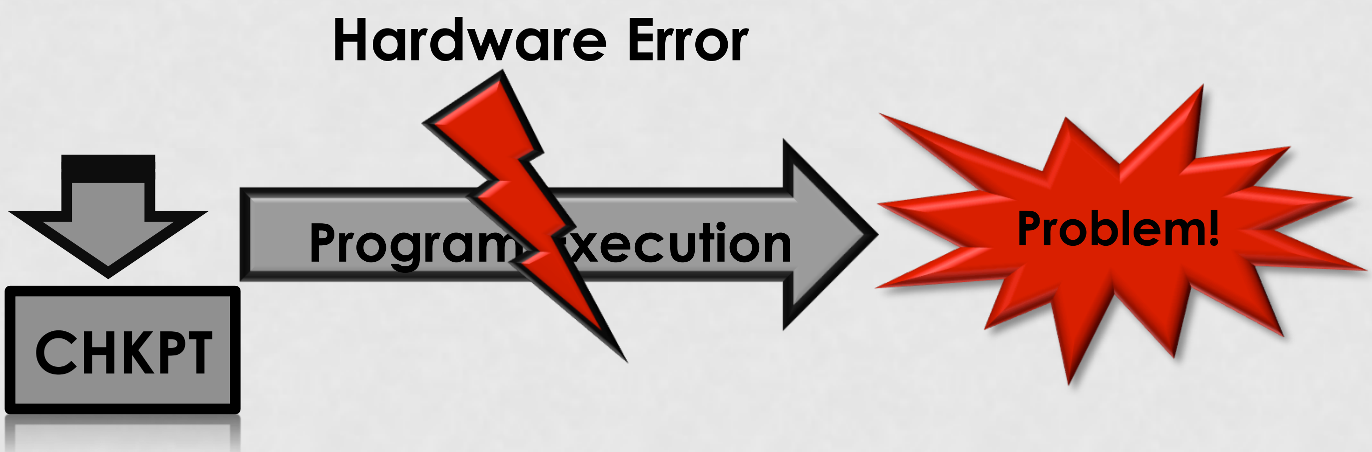
CONTRIBUTIONS

- Build a model of chip multiprocessor running a parallel application using Stochastic Activity Networks.
- Propose intermittent fault models that abstract real intermittent faults at the system level.
- Evaluate the performance of a processor after applying different recovery options.

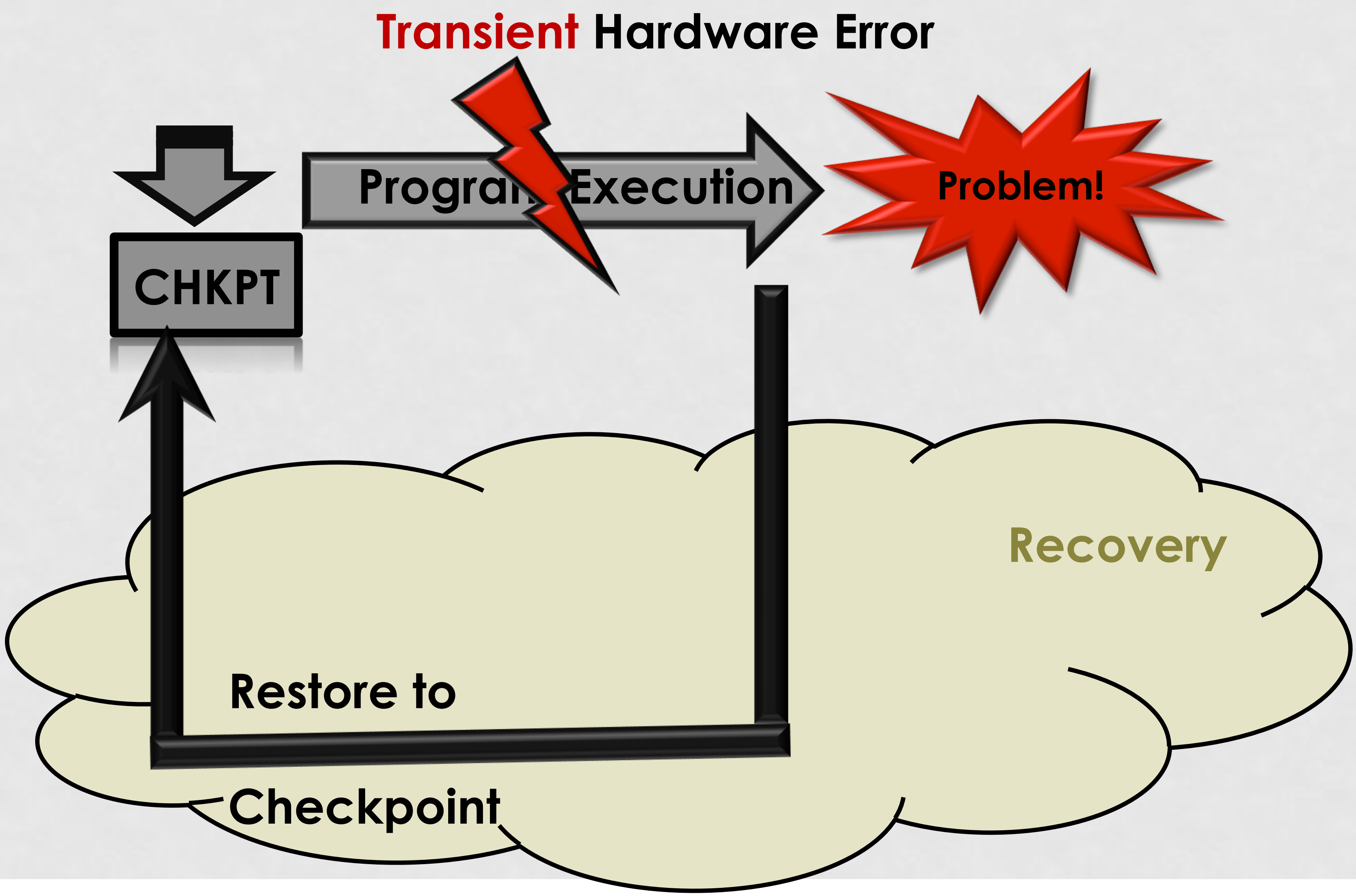
RECOVERY-MOTIVATION



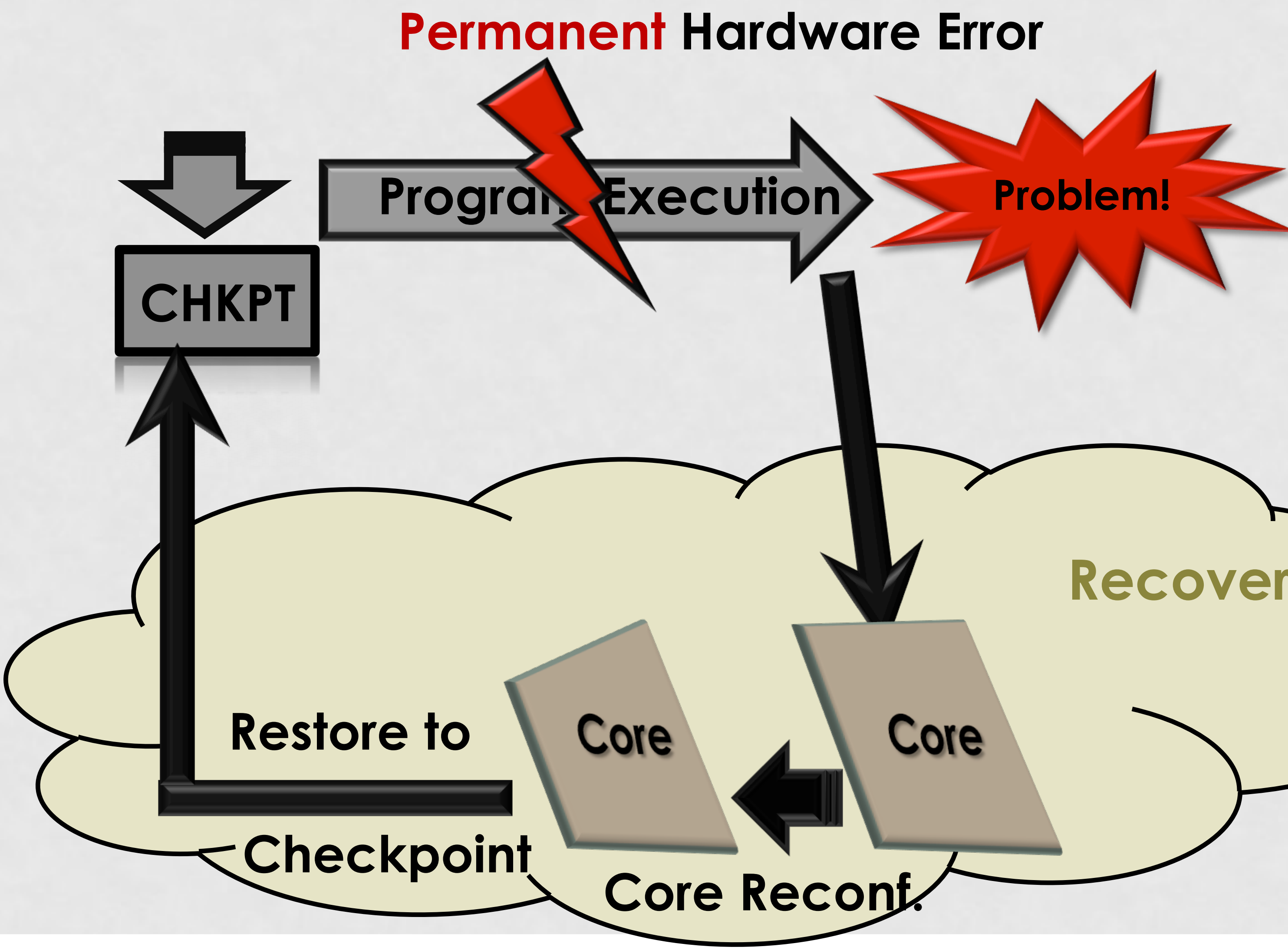
RECOVERY-MOTIVATION



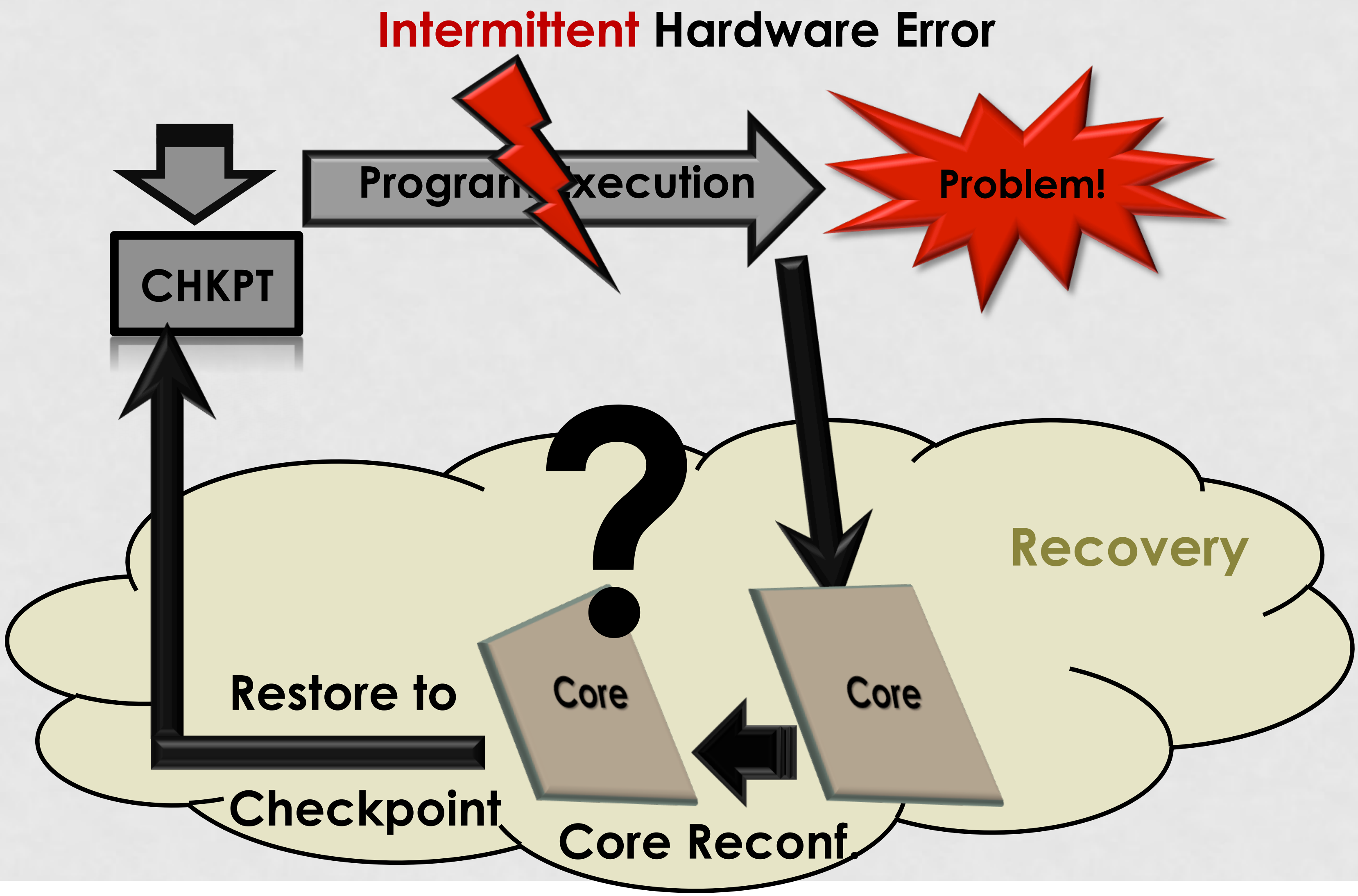
RECOVERY-MOTIVATION



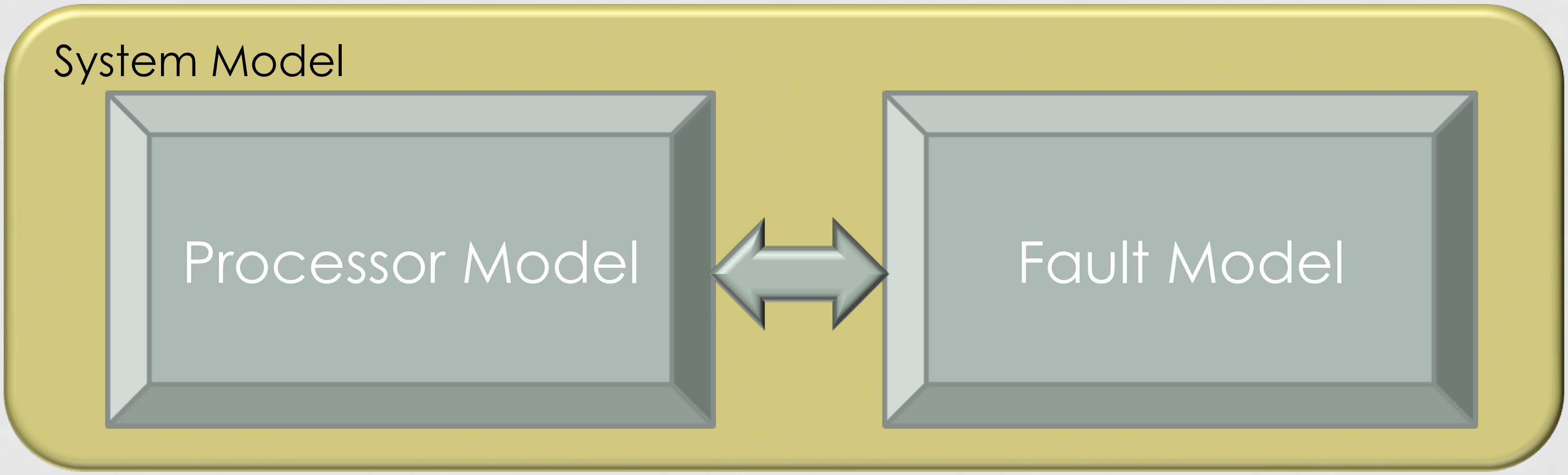
RECOVERY-MOTIVATION



RECOVERY-MOTIVATION



MODEL OVERVIEW



- Rollback-Only
- Permanent Reconfiguration
- Temporary Reconfiguration

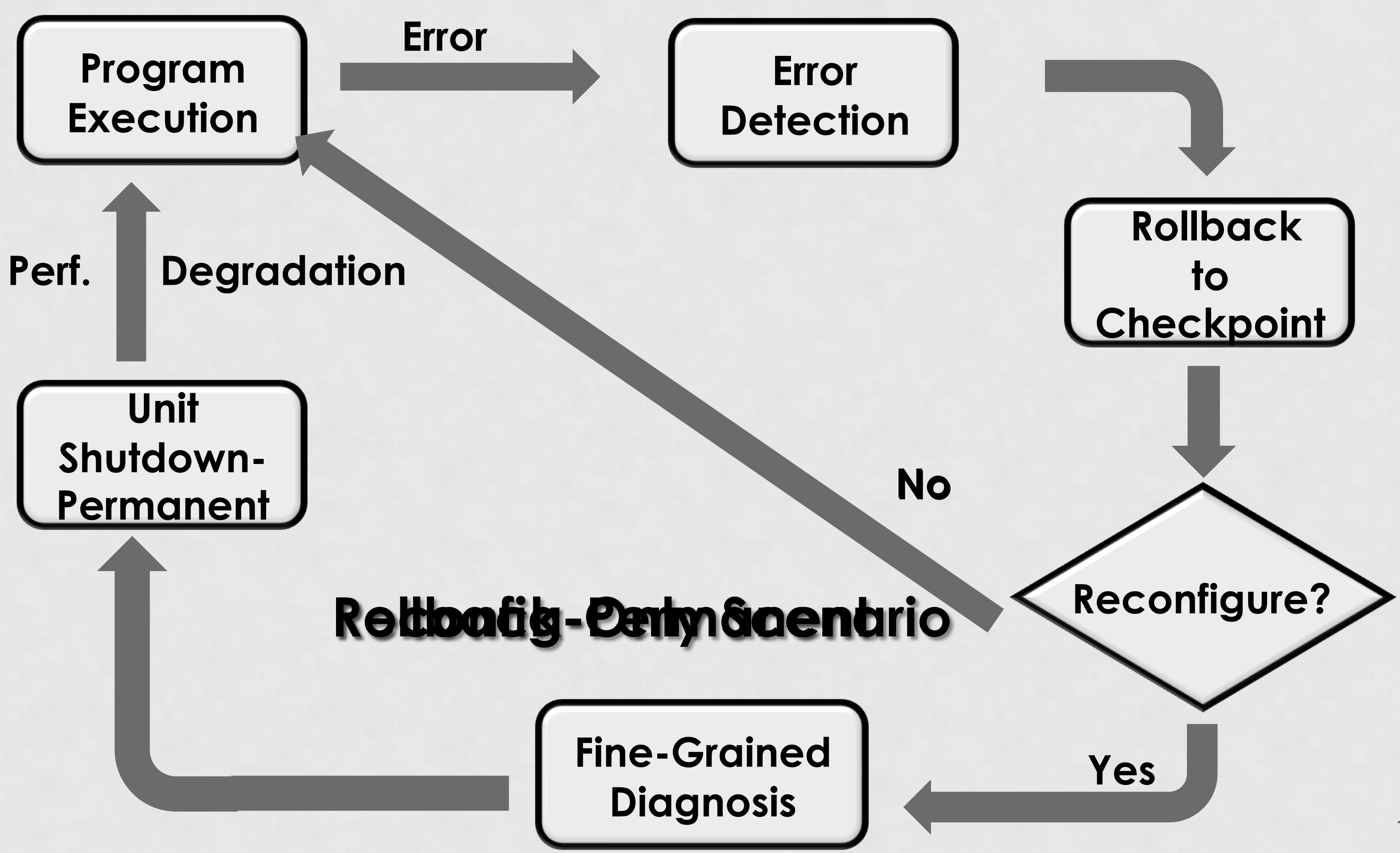
- Base
- Exponential
- Weibull

KEY FINDINGS

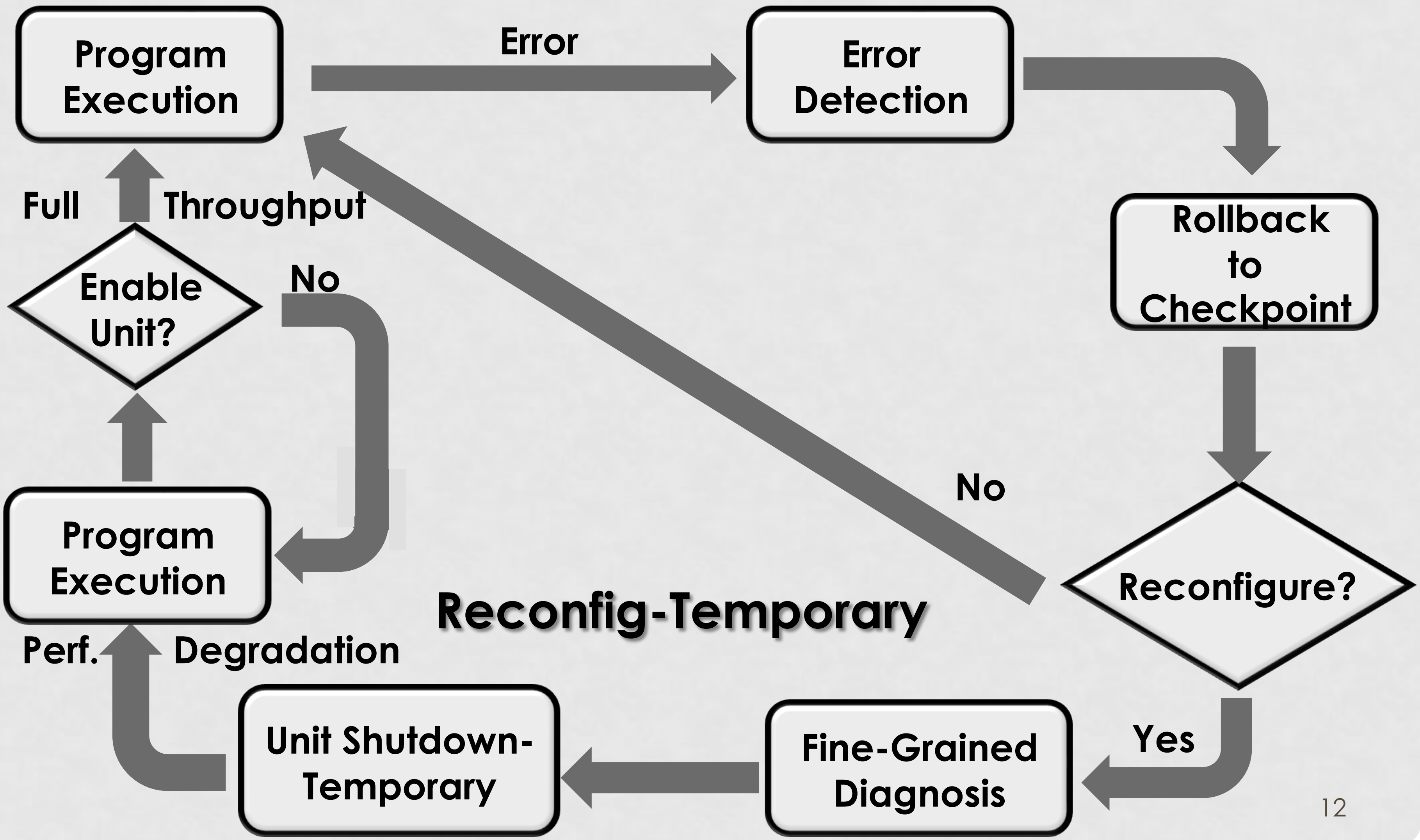
- **Error rate** and the **relative importance** of the error location are the main factors in finding the **best recovery** for high intermittent failure rates.

- **Permanent shutdown** of the defective unit results in a **slight improvement** of the performance **compared** to the **temporary shutdown**.

PROCESSOR MODEL

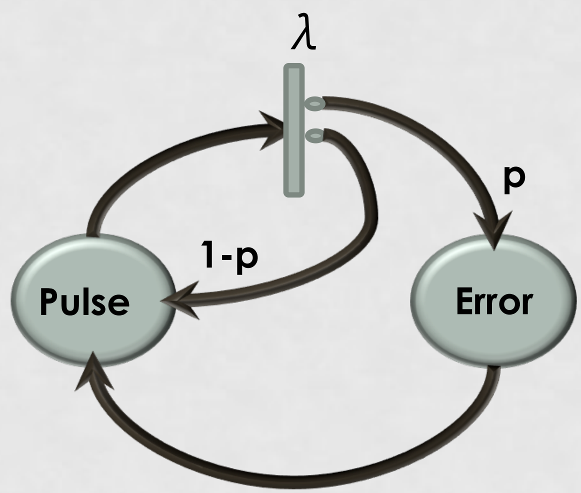


PROCESSOR MODEL



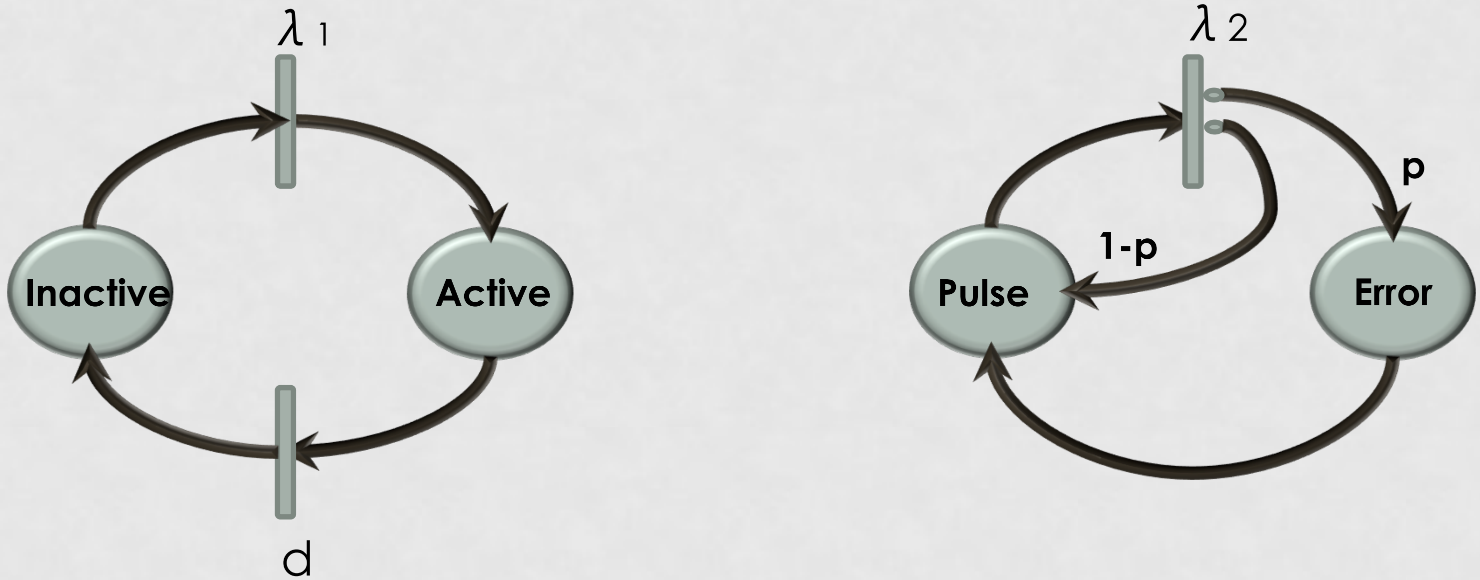
FAULT MODEL-BASE FAULT MODEL

- Abstract physical fault models.
- Prune down the space of system configurations.



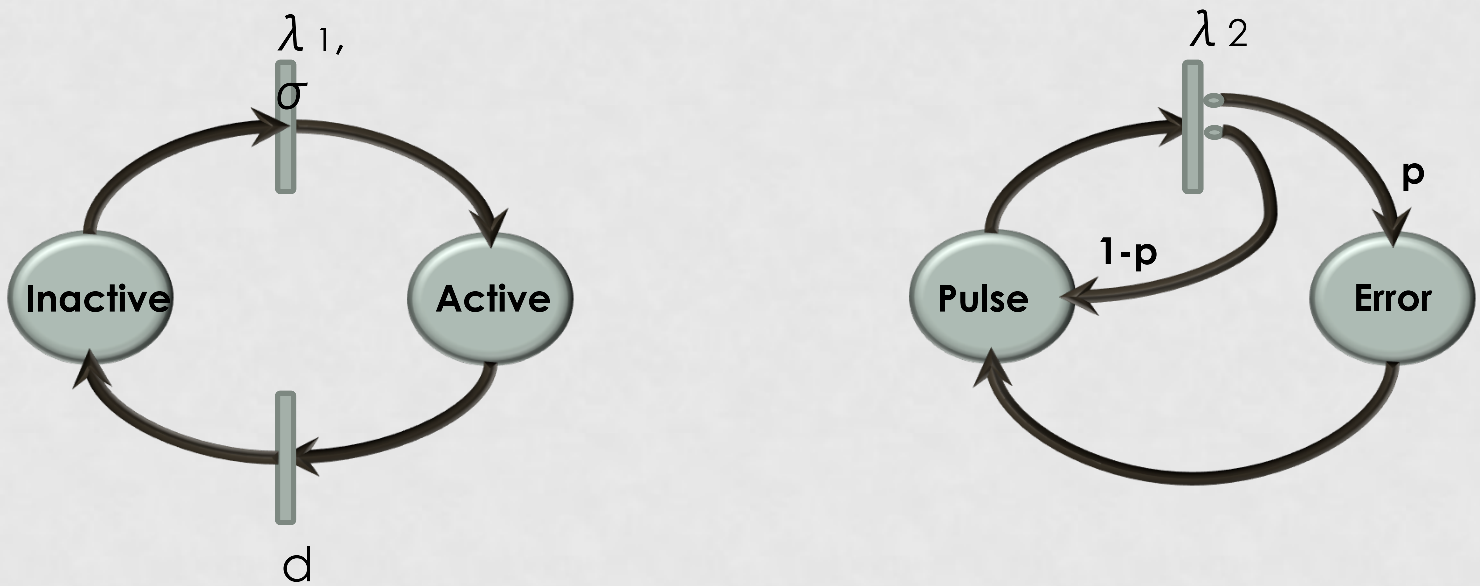
FAULT MODEL-EXPONENTIAL FAULT MODEL

- Abstract physical fault models.
- Prune down the space of system configurations.



FAULT MODEL- WEIBULL FAULT MODEL

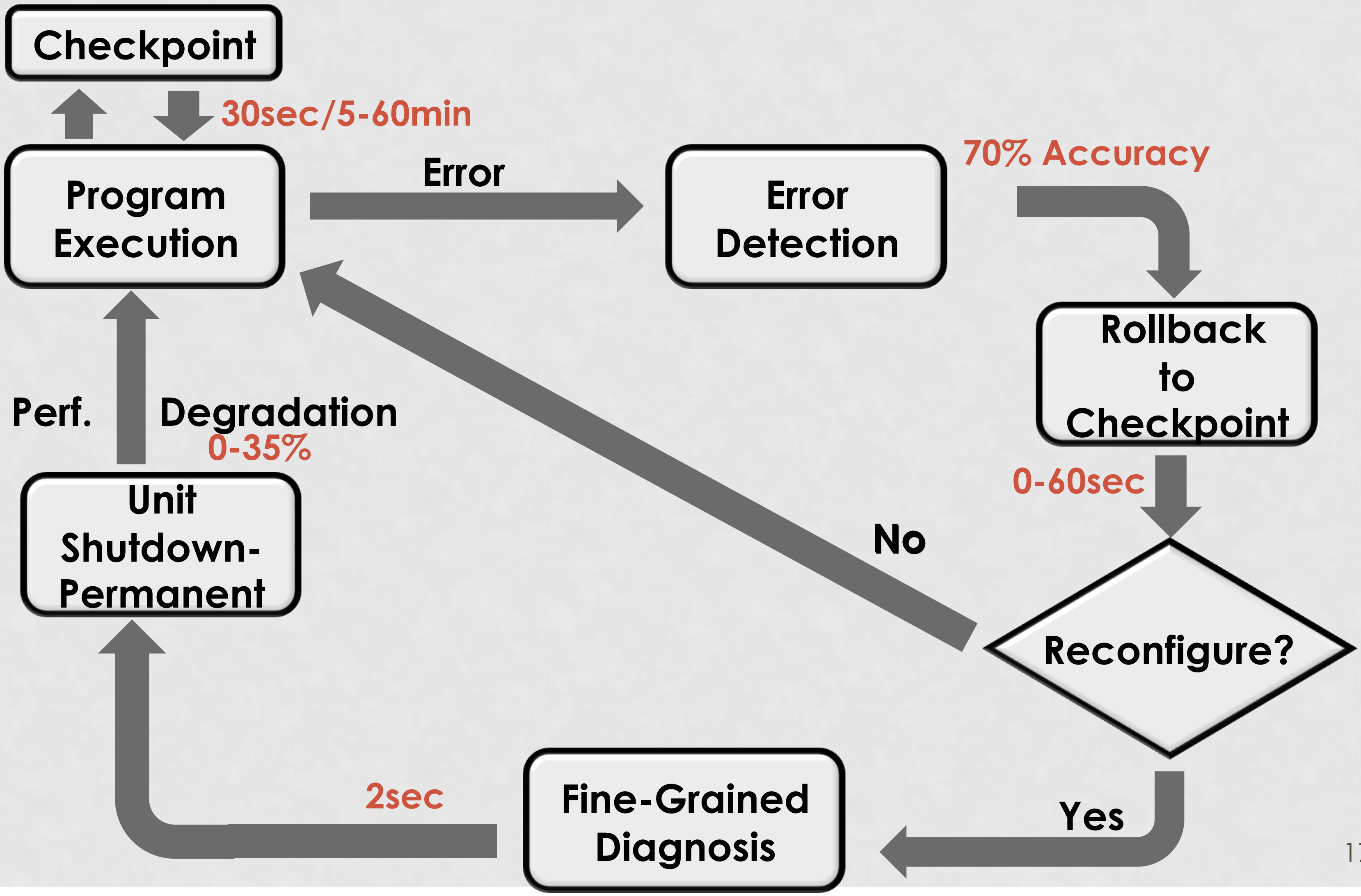
- Abstract physical fault models.
- Prune down the space of system configurations.



EXPERIMENT SETUP

- Used Mobius[2] to simulate the system for 48 hours with a confidence interval of 95%.
- Used useful work[3] measure to model processor throughput in a certain a mount of time.
- Analyzed a model of multiprocessor running coordinated checkpoint.

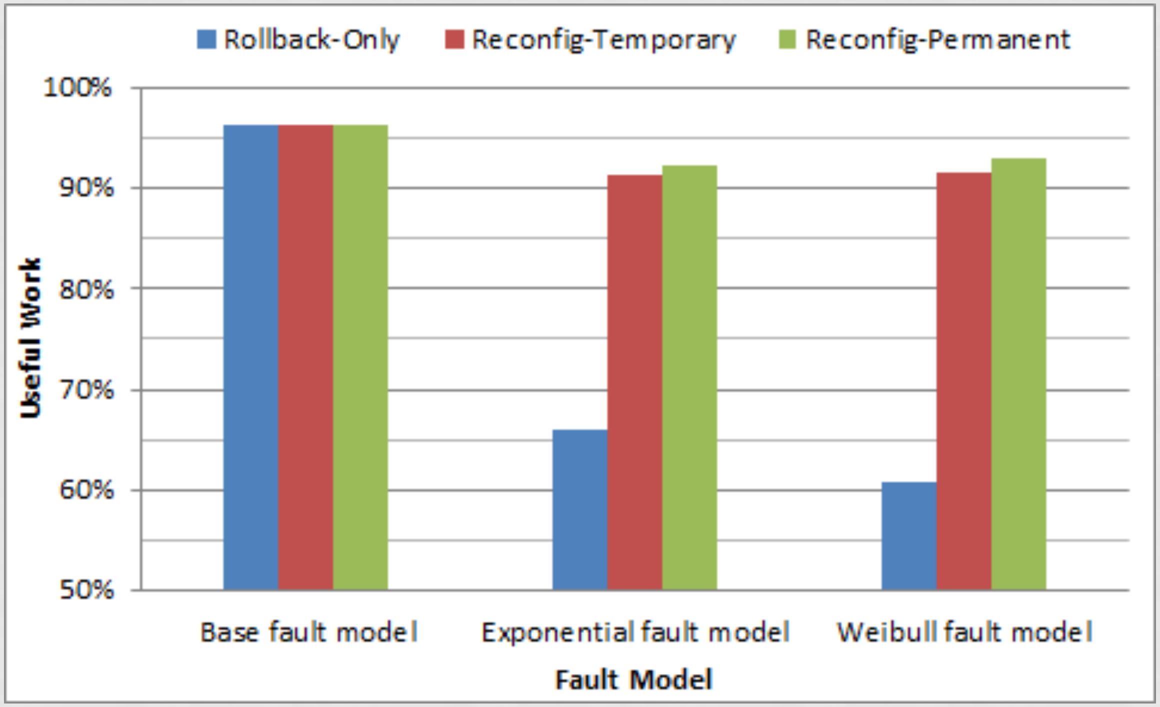
SYSTEM PARAMETERS



RESEARCH QUESTIONS

- **When should we recover from an intermittent fault by shutting down the defective component?**
- **For errors that are tolerated by shutting down the defective component, should the shutdown be permanent or temporary?**

RESULTS-DIFFERENT FAULT MODELS



- Permanent/temporary reconfiguration leads to 27% more useful work than rollback-only for exponential and Weibull fault models.

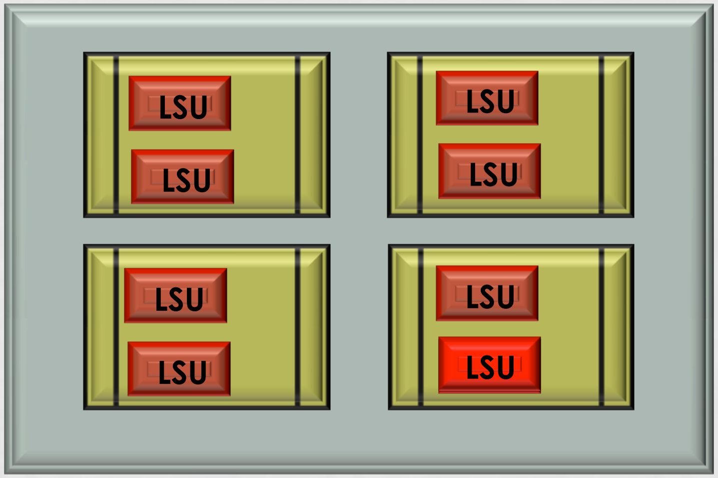
RESEARCH QUESTION

What is the granularity of the disabled component that maximizes the processor's performance?

COMPONENT RANK

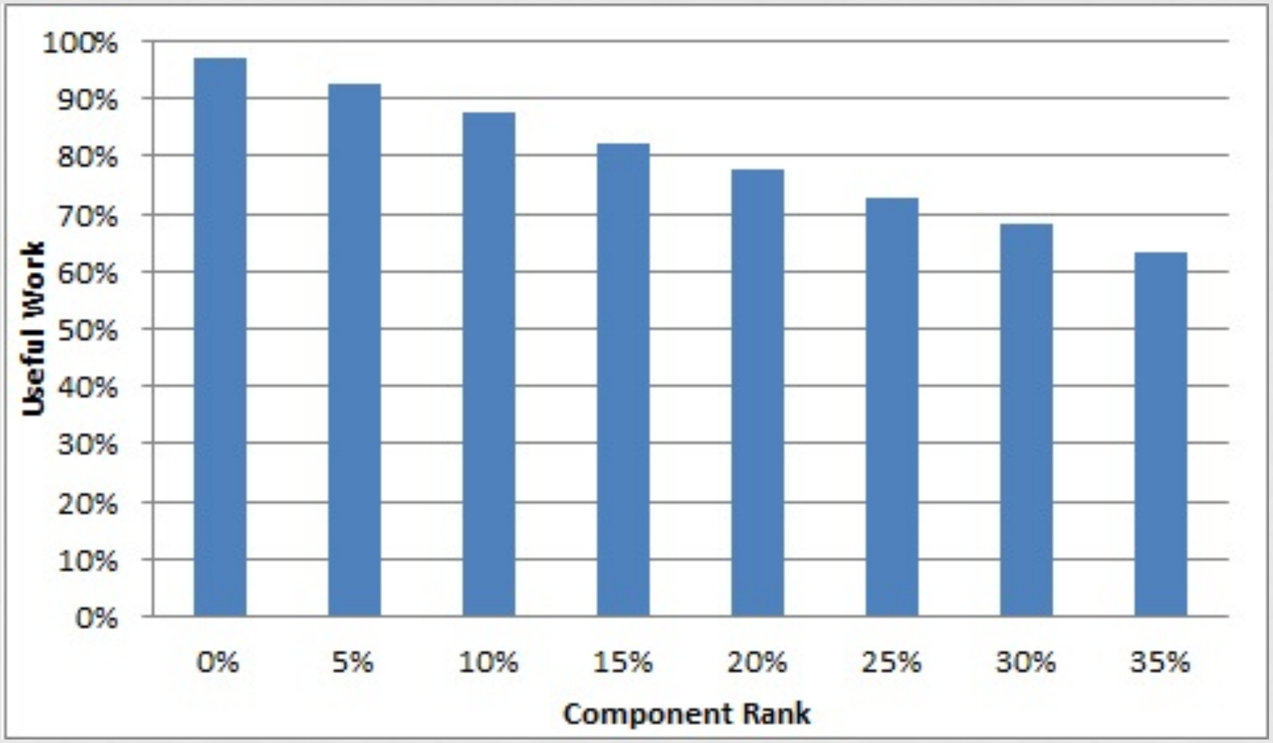
- The maximum percentage of useful work that is lost when the component is disabled.

- 4-core processor, each core has two LSUs and is running a program that is using all the 8 LSUs for 60% of the time.



- Using Amdahl's law, LSU rank is **19%** or **1/(0.4 + (0.6/0.125))**

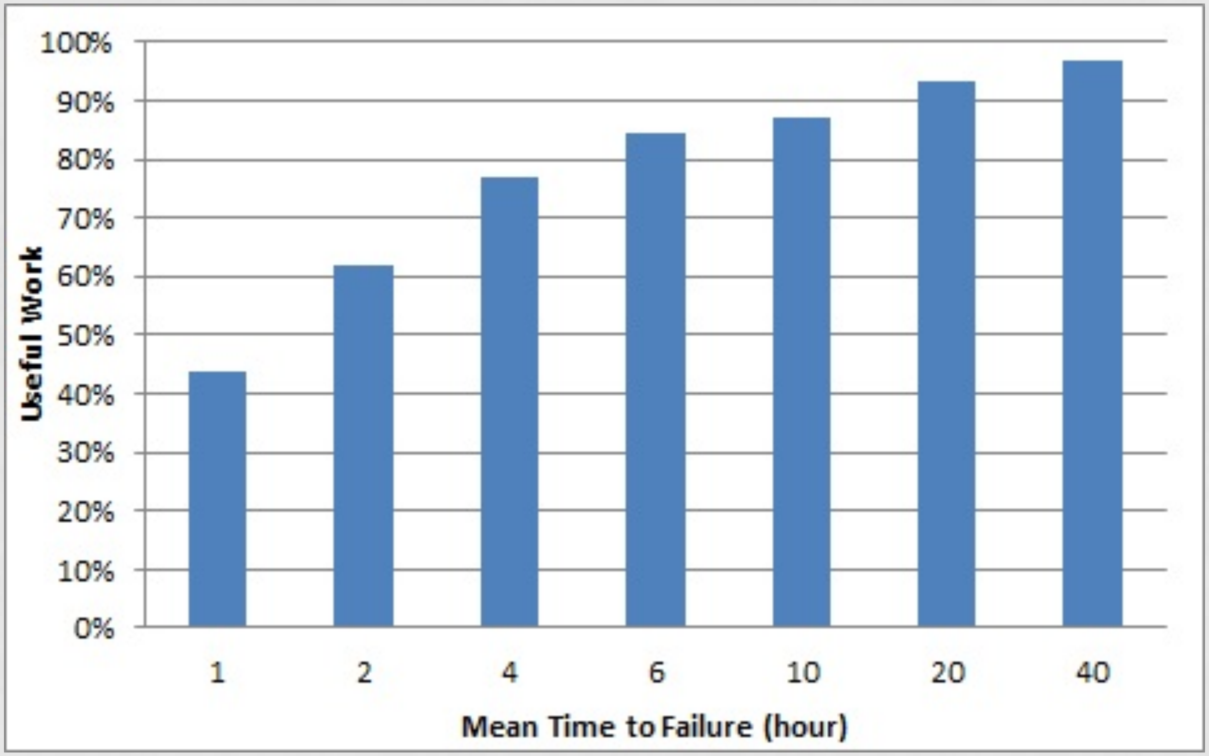
RESULT-EFFECT OF COMPONENT RANK



- For this experiment, components with a rank of 35% or more should be disabled if diagnosed with intermittent errors.

Sensitivity to Fault Rate

RESULTS- SENSITIVITY TO FAULT RATE



- If lost useful work outweighs the rank of the defective component, then the defective component should be disabled.

KEY FINDINGS

- **Error rate** and the **relative importance** of the error location are the main factors in finding the **best recovery** for high intermittent failure rates.

- **Permanent shutdown** of the defective unit results in a **slight improvement** of the performance **compared** to the **temporary shutdown**.

[1] Eurosys, 2011
[2] Tools of MME, 2003.
[3] DSN, 2005.