Flash SSDs in Mobile Storage Systems

Tao Xie

Supported by NSF under grant CNS-0834466

Department of Computer Science
San Diego State University

Content

- Introduction
- Motivation
- Research Tasks and Preliminary Results

SAN DIEGO STATE UNIVERSITY

Introduction

- This project develops a hybrid disk array architecture, which integrates small capacity flash SSDs with HDDs (hard disk drives) to form a robust and energy-efficient storage system for mobile data-intensive applications.
- An array of new data management techniques for data-intensive mobile applications will be developed.
- A prototype and a simulation toolkit will be implemented.
- It will also promote teaching, learning, and training.

SAN DIEGO STATE UNIVERSITY

Stationary Data Centers

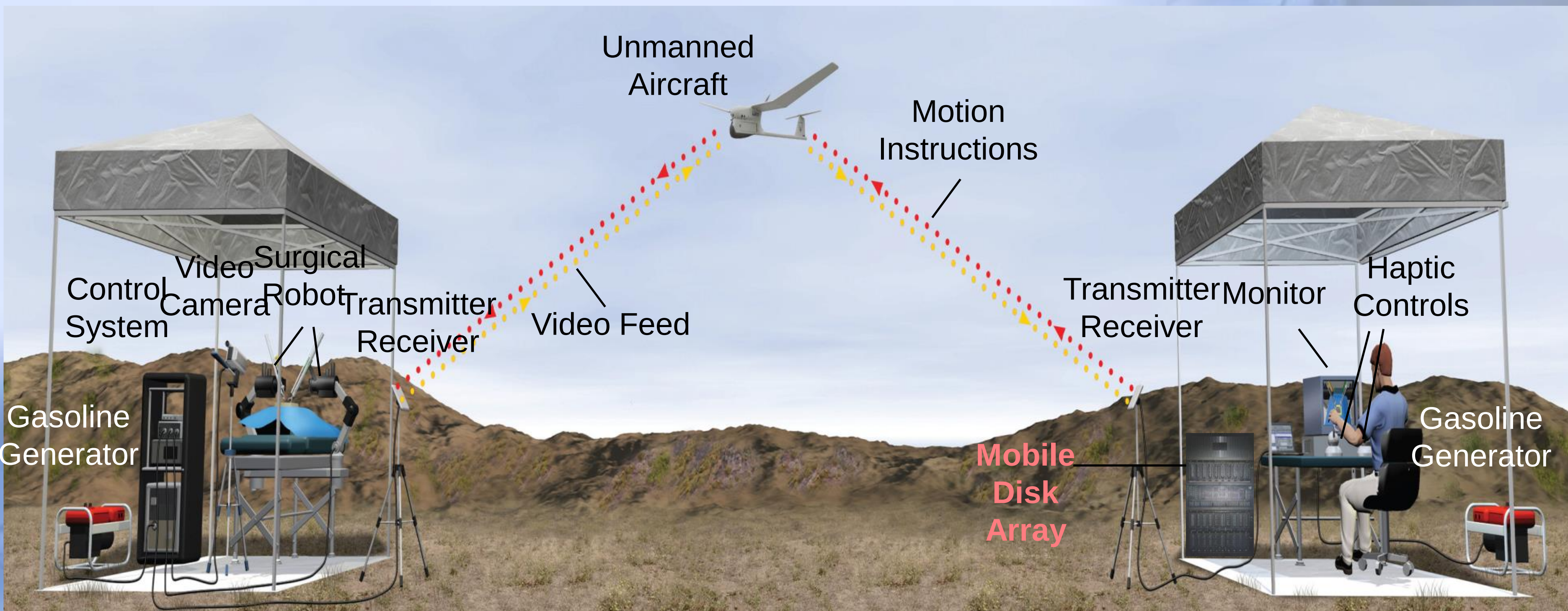
SAN DIEGO STATE UNIVERSITY

Mobile Disk Arrays

- Existing mobile disk array consists of an array of independent small form factor hard disks connected to a host by a storage interface in a mobile computing environment.
- Hard disks have some intrinsic limitations such as long access latencies, high annual disk replacement rates, fragile physical characteristics, and energy-inefficiency.
- Due to their severe application environments, mobile disk arrays must be energy-efficient, extremely reliable, highly fault-tolerant and physically robust.

SAN DIEGO STATE UNIVERSITY

Application One: Remote Surgery



SAN DIEGO STATE UNIVERSITY

Application Two: Mobile Data Center

San Diego State University

New Challenges of Mobile Disk Arrays

- Very limited power supply
- Stringent reliability requirement
- High demands on fault-tolerance
- Robust physical characteristics

File Open Issues That Will Be Addressed

1. The lack of a high-performance, highly reliable, and energy-efficient storage architecture
2. New energy-saving data management schemes for mobile data-intensive applications
3. Understanding of the relationship between disk energy saving techniques and disk reliability
4. The absence of an energy-ware fault-tolerant mechanism
5. A prototype and a simulation toolkit

San Diego State University

Five Research Tasks

1. Develop a hybrid disk storage architecture
2. Develop a reliability model
3. Establish an energy conservation infrastructure
4. Develop an energy-aware fault-tolerant mechanism
5. Design and implement a mobile disk array prototype and a simulation toolkit

SAN DIEGO STATE UNIVERSITY

Task 1: Developing a flash disk assisted storage architecture

1. We are implementing a novel *f*lash d*i*sk s*t*orage architecture (FIT) for high performance, energy conservation and highly reliable mobile disk arrays.

2. The basic idea of the FIT architecture is to construct mobile disk arrays by using both non-volatile solid-state flash SSDs and small-factor hard disk drives.

SAN DIEGO STATE UNIVERSITY

Flash SSD

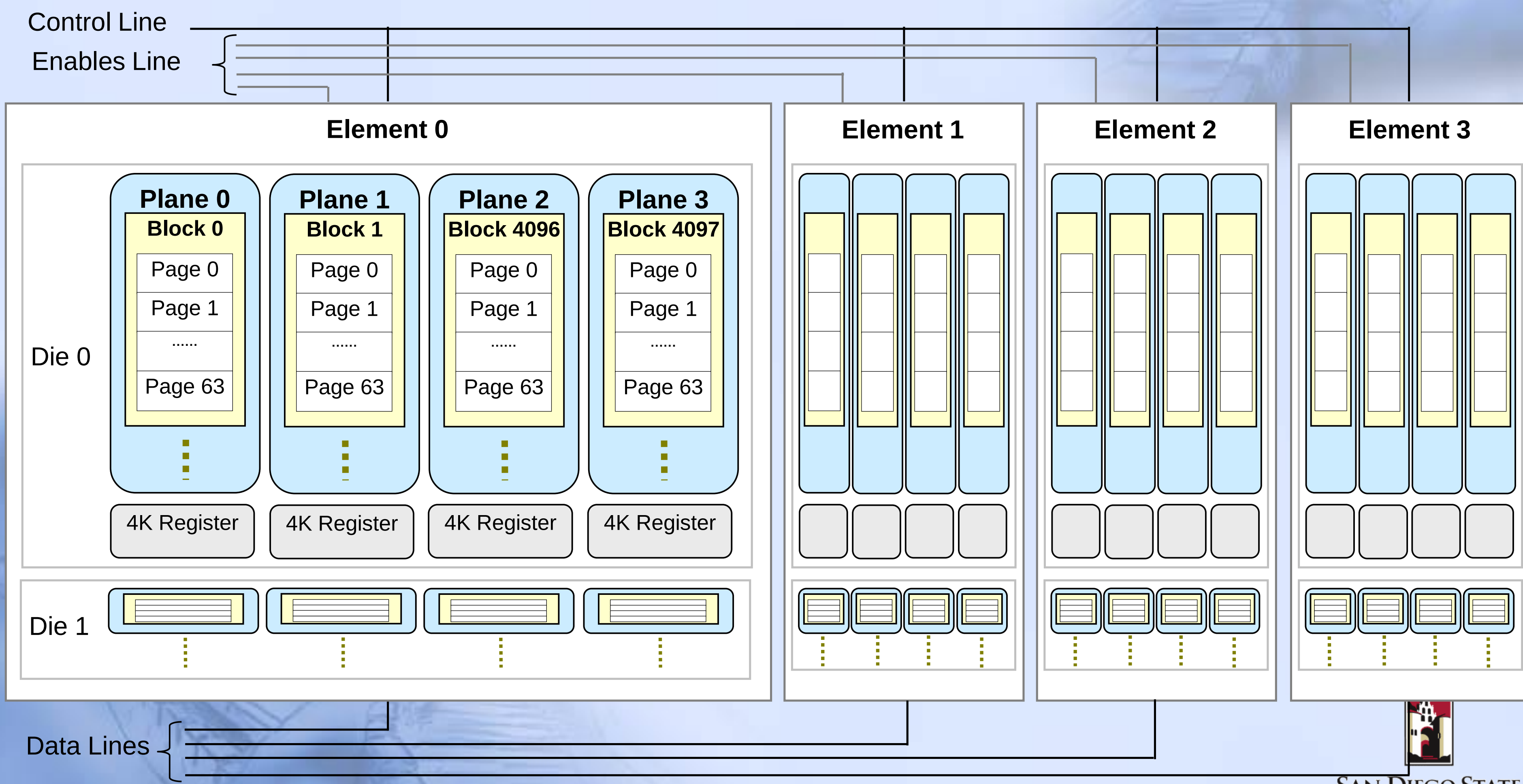
1. They are physically robust with high vibration-tolerance and shock-resistance.
2. They inherently consume much less energy than mechanical mechanism based hard disks.
3. They offer much fast read access times.
4. Very recent breakthrough largely relaxes the three constraints on existing flash disks: small capacity, low throughput, and limited erasure cycles.

Flash SSD vs. HDD

2.5" SATA 3.0Gbps SSD		2.5" SATA 3.0Gbps HDD
Solid NAND flash based	**Mechanism type**	Magnetic rotating platters
64GB	**Density**	80GB
73g	**Weight**	365g
Read: 100MB/s, Write :80MB/s	**Performance**	Read: 59MB/s, Write: 60MB/s
1W	**Active Power consumption**	3.86W
20G (10~2000Hz)	**Operating Vibration**	0.5G (22~350Hz)
1,500G for 0.5ms	**Shock resistance**	170G for 0.5ms
0°C to 70°C	**Operating temperature**	5°C to 55°C
None	**Acoustic Noise**	0.3 dB
MTBF >2M hours	**Endurance**	MTBF < 0.7M hours

SAN DIEGO STATE UNIVERSITY

Internal Structure of a SSD with Four Elements



SAN DIEGO STATE UNIVERSITY

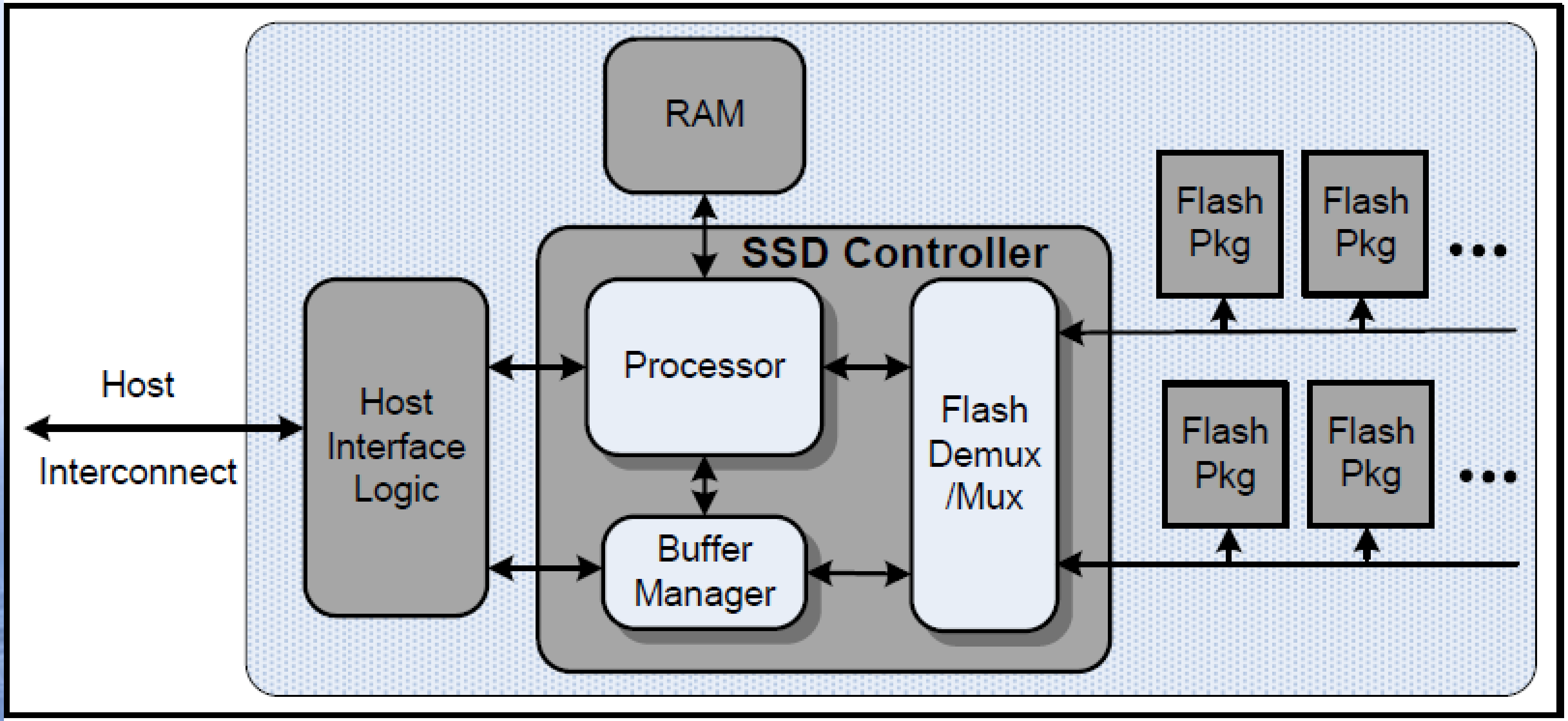
Flash SSD Datasheet

Page Read to Register	25μs
Page Program (Write) from Register	200μs
Block Erase	1.5ms
Serial Access to Register (Data bus)	100μs
Die Size	2 GB
Block Size	256 KB
Page Size	4 KB
Data Register	4 KB
Planes per die	4
Dies per package (2GB/4GB/8GB)	1,2 or 4
Program/Erase Cycles	100 K

Table 1: Operational flash parameters

San Diego State University

Logic Diagram



San Diego State University

FTL (Flash Translation Layer)

- Address Mapping
- Wear Leveling
- Garbage Collection

Mapping Schemes

- Page-level mapping
- Block-level mapping
- Hybrid

Wear Leveling

- Dynamic Wear Leveling

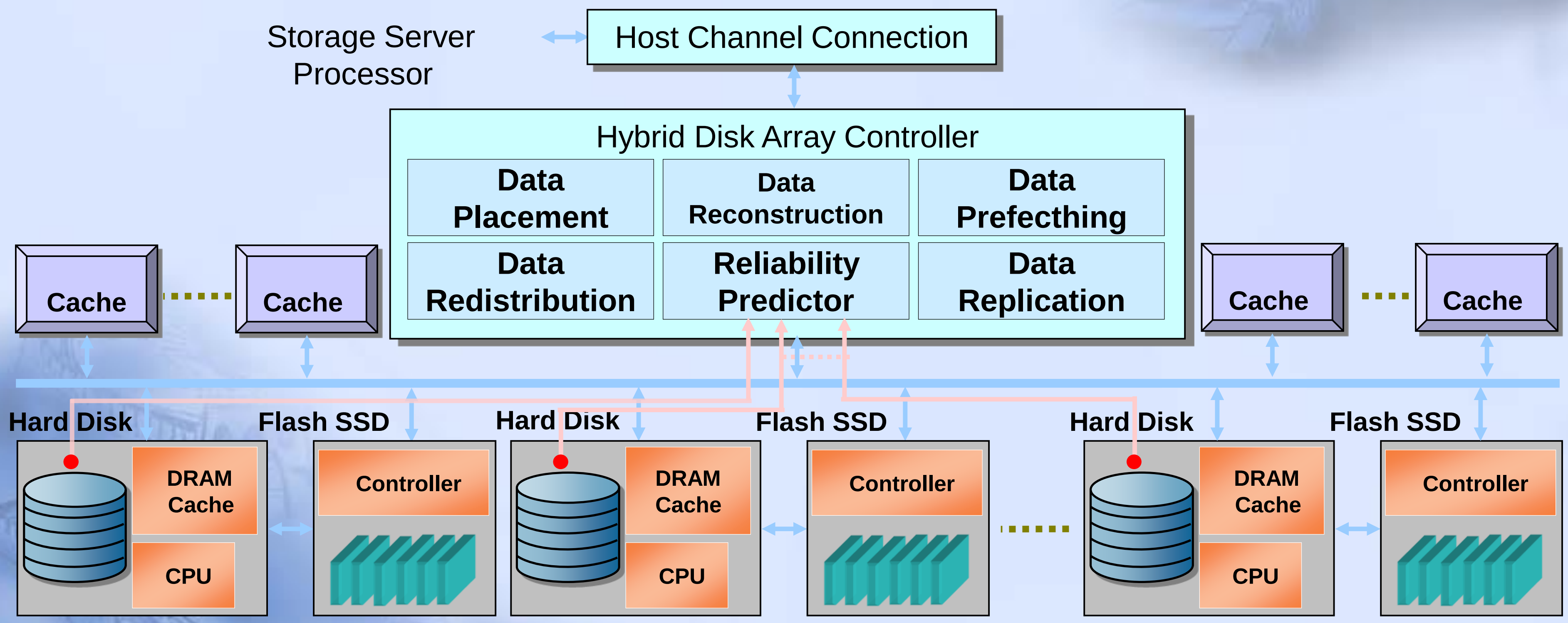
  when a write request arrives it dynamically selects a new free data block based on the number of erasure cycles that the block already has

- Static Wear Leveling

  wear leveling all data blocks including those that are not being written to

San Diego State University

The FIT Architecture



SAN DIEGO STATE UNIVERSITY

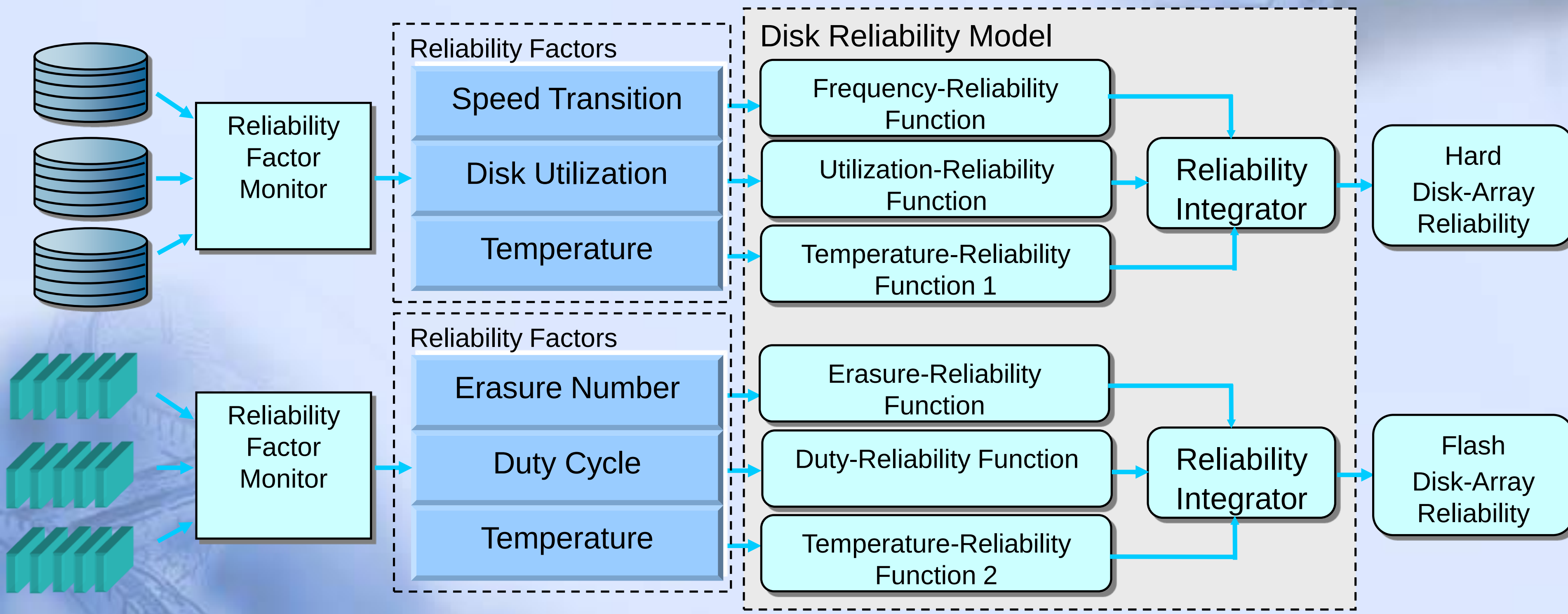
Task 2: Developing a disk reliability predictor

1. Developing a reliability predictor that is capable of estimating failure rate for both hard disks and flash SSDs is challenging.

2. A deep understanding of the relationship between energy saving techniques and disk reliability is an open question.

3. The a *re*liability *p*redictor (REP) will be built.

The Reliability Predictor (REP)



San Diego State University

Preliminary Results of Task 2

- We developed an empirical reliability model, called PRESS (Predictor of Reliability for Energy Saving Schemes) [Xie and Sun, IPDPS'08]

- Fed by operating temperature, disk utilization, disk speed transition frequency, three energy-saving-related reliability affecting factors, PRESS estimates the reliability of entire hard disk array

SAN DIEGO STATE UNIVERSITY

Task 3: Developing data management schemes

- Traditional data management schemes like data placement algorithms only concentrated on improving system performance data reliability, while normally overlooked energy efficiency.

- An array of energy-aware data management software modules: data placement algorithms, data redistribution strategies, data replication policies, and data prefetching schemes will be developed.

SAN DIEGO STATE UNIVERSITY

Modules in Task 3

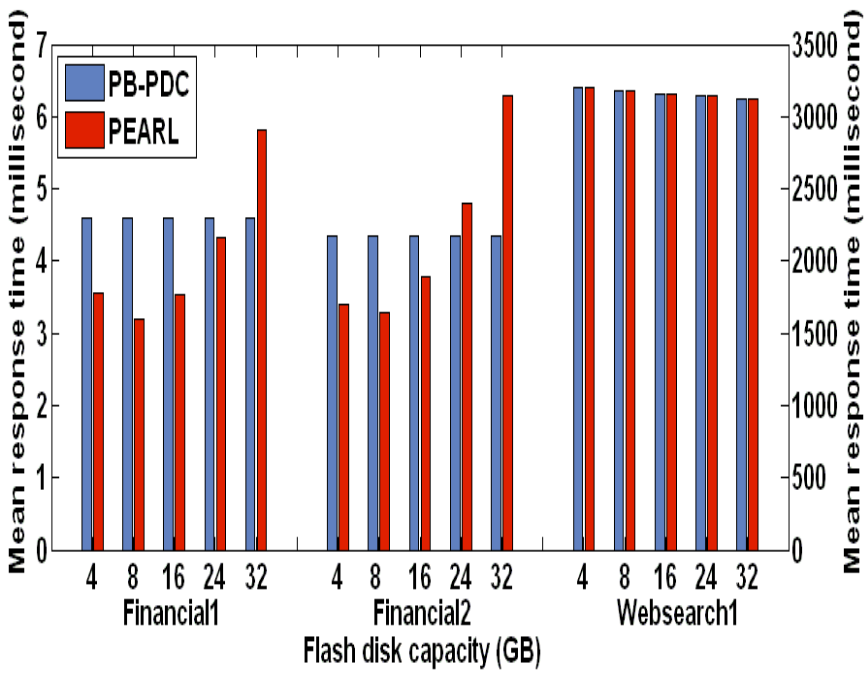
- Energy-Efficient Data Placement Algorithms
- Self-Adaptive and Reliability-Aware Data Redistribution Strategies
- Self-Triggered Data Replication Policies
- Automatic Data Prefetching Schemes

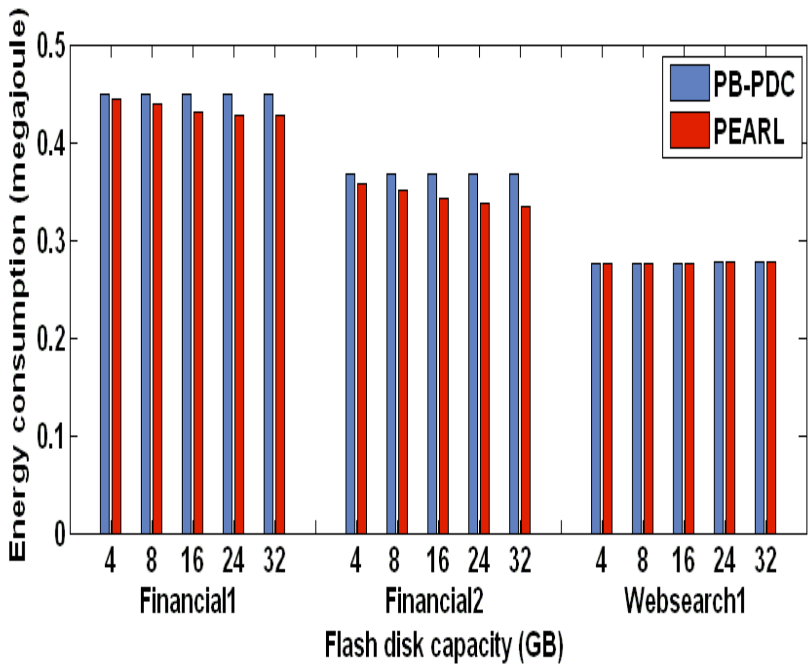
San Diego State University

Preliminary Results of Task 3

- “Dynamic Data Reallocation in Hybrid Disk Arrays” [Xie and Sun, IEEE TPDS]
- “PEARL: Performance, Energy, and Reliability Balanced Dynamic Data Redistribution for Next Generation Disk Arrays” [Xie and Sun, MASCOTS’08]
- “SAIL: Self-Adaptive File Reallocation on Hybrid Disk Arrays” [Xie and Madathil, HiPC’08]

PEARL





SAN DIEGO STATE UNIVERSITY

Task 4: Developing an energy-aware fault-tolerant mechanism for FIT

- Data reconstruction algorithms, which are executed in the presence of disk failure, for mobile storage systems must be reliability-aware, performance-driven and energy-efficient.

- We developed two novel reconstruction strategies that can be applied to mobile storage systems to noticeably save energy while providing shorter reconstruction times and user response times.

SAN DIEGO STATE UNIVERSITY

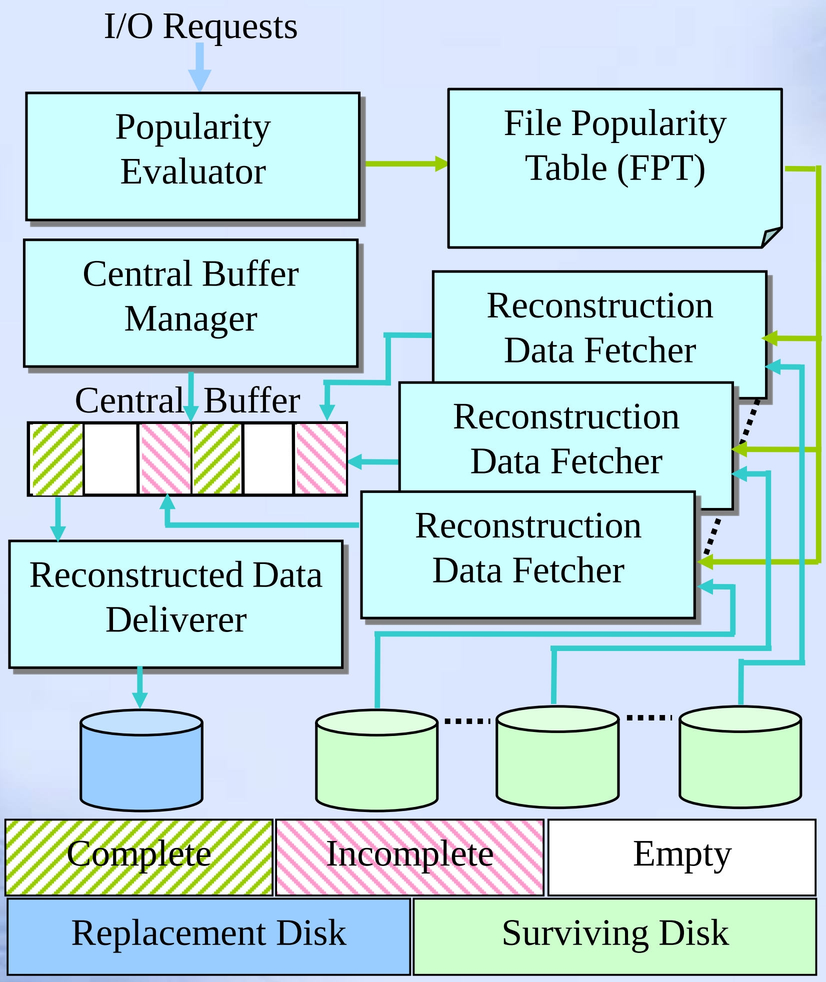
Preliminary Results of Task 4

- "MICRO: A Multi-level Caching-based Reconstruction Optimization for Mobile Storage Systems" [Xie and Wang, IEEE Transactions on Computers, October 2008]

- "Collaboration-Oriented Data Recovery for Mobile Disk Arrays" [Xie and Sharma, ICDCS'09]

SAN DIEGO STATE UNIVERSITY

Architecture of MICRO

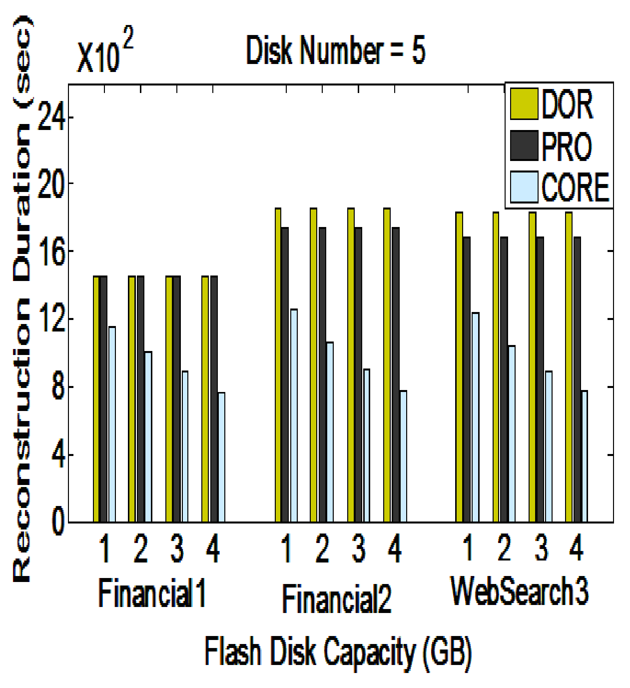


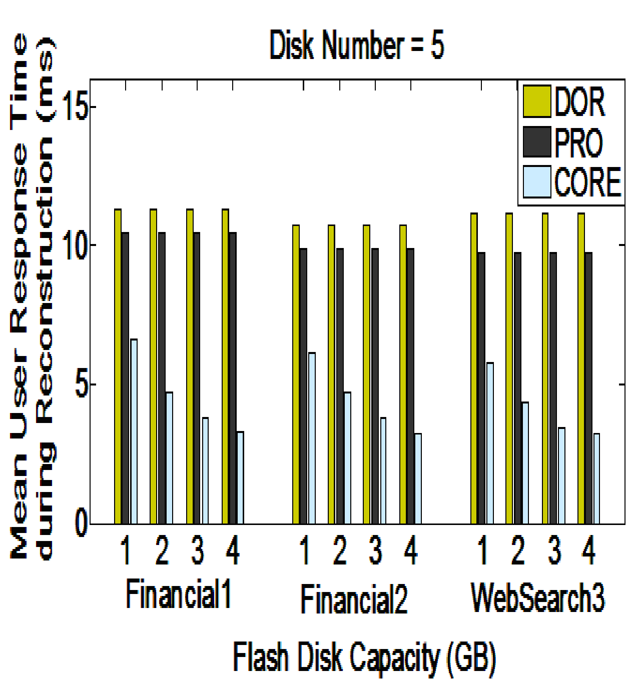
SAN DIEGO STATE UNIVERSITY

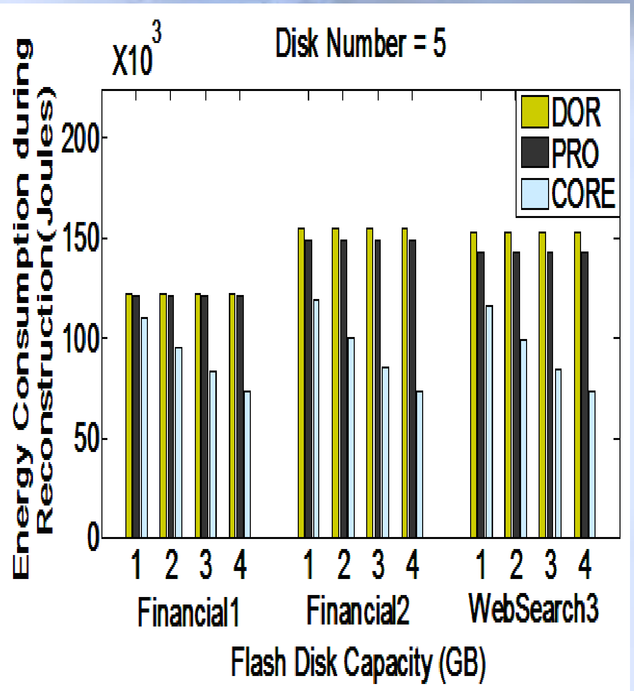
CORE

- We developed a flash assisted data reconstruction strategy called CORE (c̲ollaboration-o̲riented r̲e̲construction) on top of a hybrid disk array architecture.

Experimental Results of CORE







SAN DIEGO STATE UNIVERSITY

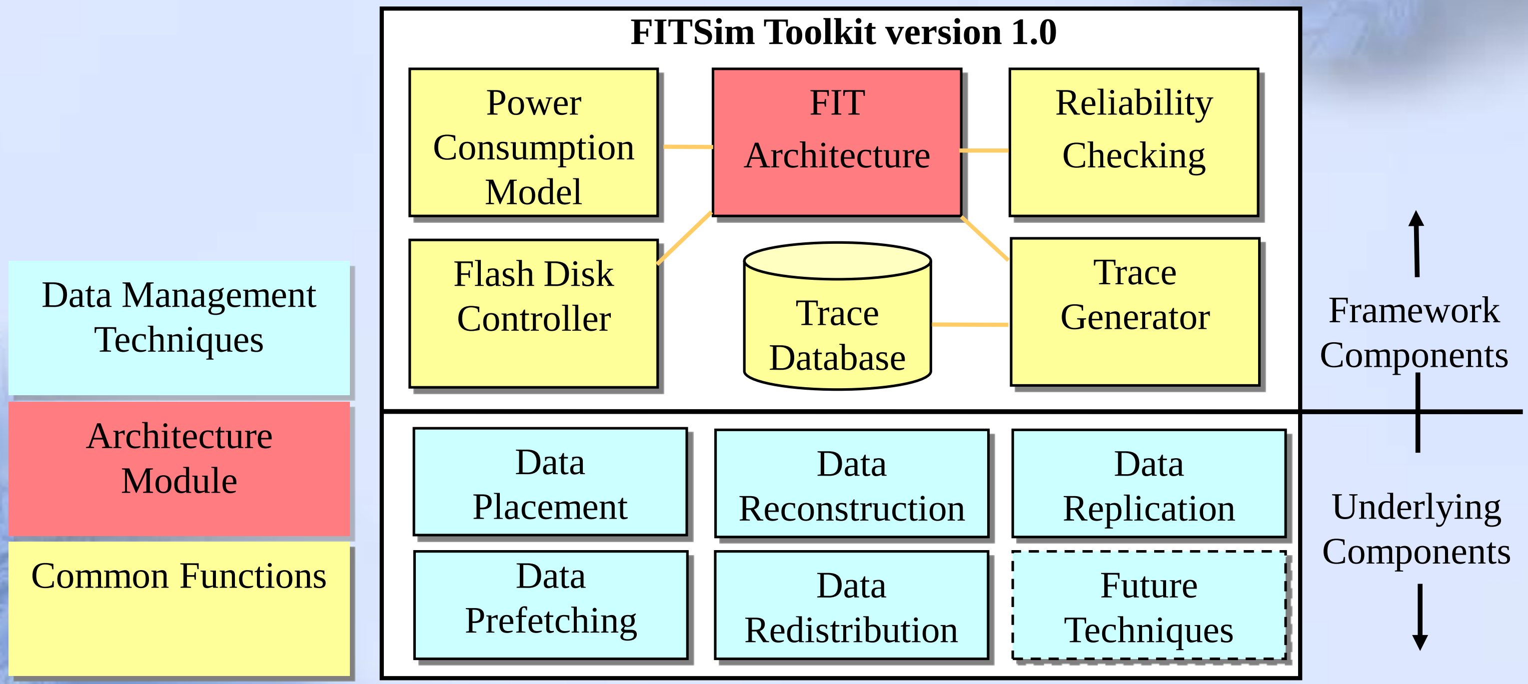
Research Task 5: a prototype and a simulation toolkit

- We are implement a FIT prototype, which will be the first of its kind.
- We are implementing a simulation software toolkit called FITSim Toolkit.

The FIT Prototype

- Implement a FIT prototype to work on Linux
- Evaluate the FIT architecture and the new data management schemes' performance and tradeoffs
- Address the issues that will arise during the realization of the FIT architecture

FITSim



SAN DIEGO STATE UNIVERSITY

Other Active Projects

- A Device-Array Based Flash Storage System for Emerging Data-Intensive and Mission-Critical Mobile Applications: from Architecture Redesign to New File System (NSF under grant CNS-1320738, $440,727, 10/2013 ~ 09/2016)
- CAREER: Architectural Support for Integrating NAND Flash Solid State Disks into Enterprise-Class Storage Systems (NSF under grant CNS-0845105, $436,000, 09/2009 ~ 08/2014)

San Diego State University

Questions?