NEXT-GENERATION SSD VERIFICATION PLATFORM WITH TERA-SCALE NAND CAPACITY AND STORAGE SIGNAL PROCESSING SUPPORT

NVRAMOS 2012-10-16

Seil Lee, Myunghyun Rhee, and Sungroh Yoon

Advanced Computing Laboratory, Seoul National University

Contents

Trends & Issues

- Trends of NVRAM
- Demand for Storage
- NAND Scaling and Challenges

Verification Platform for SSD Development

- SSD Core Technology and Controller
- Existing Verification Platforms
- Our Platforms

Implement of NAND Controller

- Specification
- NAND Status Monitor
- ONFI Interface

Experiments

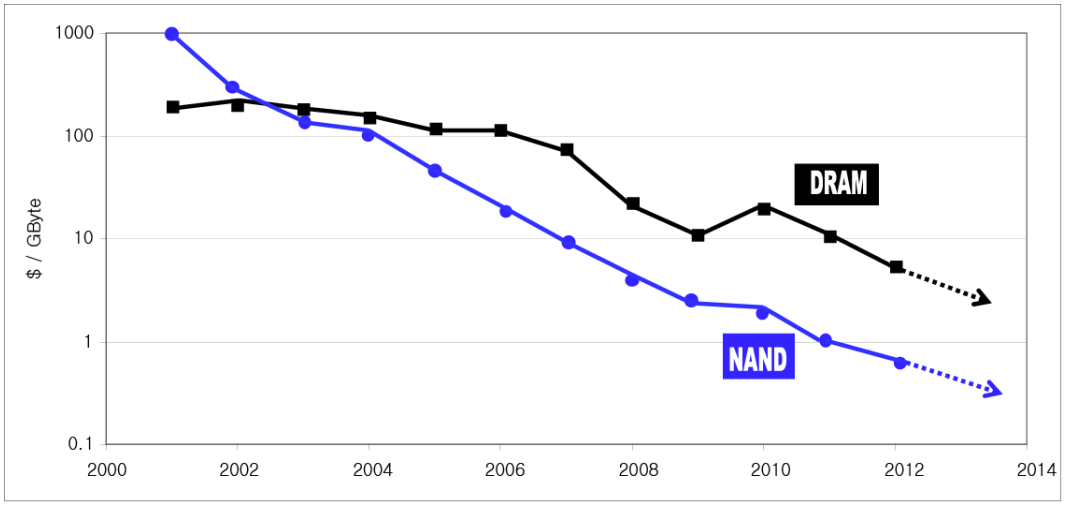
- Maximum performance test
- FTL
- ECC
- NSM TEST

Trends & Issues

- Trends of NVRAM
- Demand for Storage
- NAND Scaling and Challenges

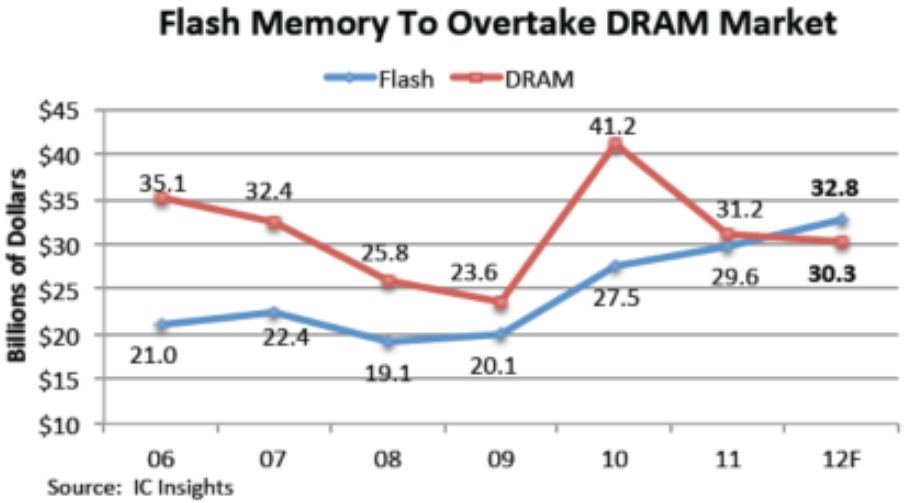
Trends of NVRAM

- Cost per GB Trend in 2012



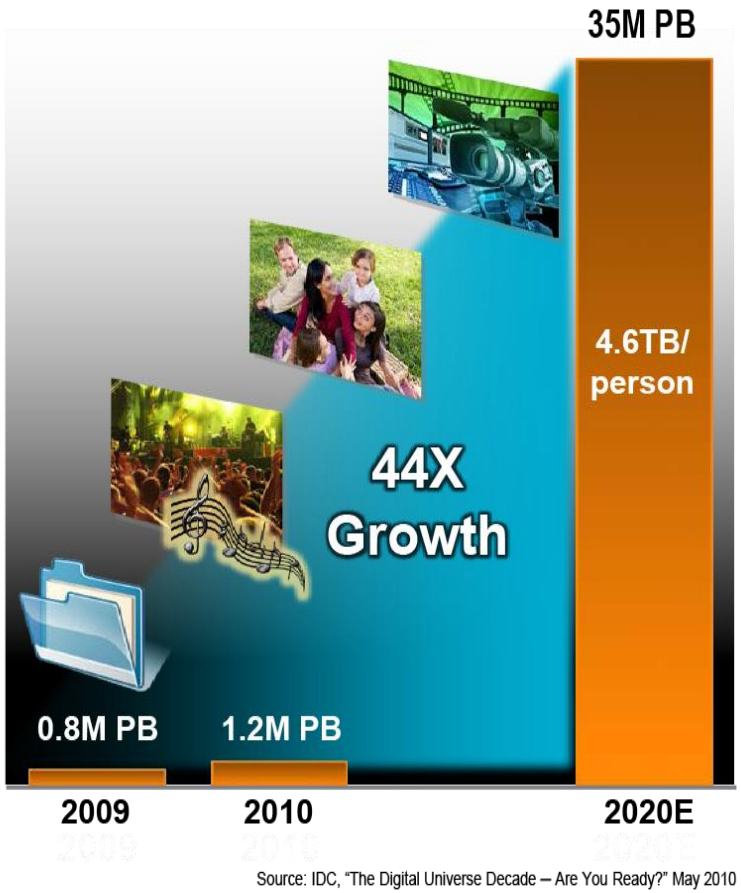
Source = DRAM Exchange/IDC/ASML

- Market Trend in 2011



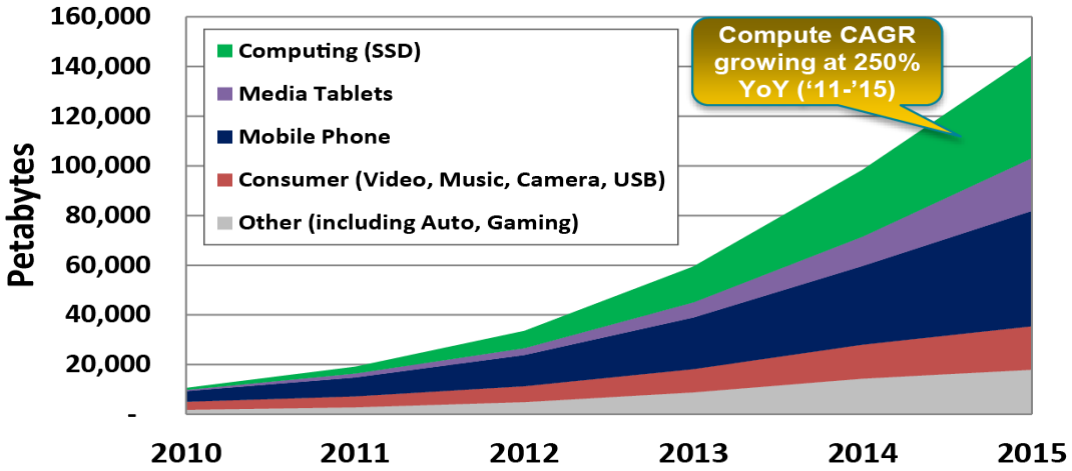
Market of Storage Device

Storage Exploding



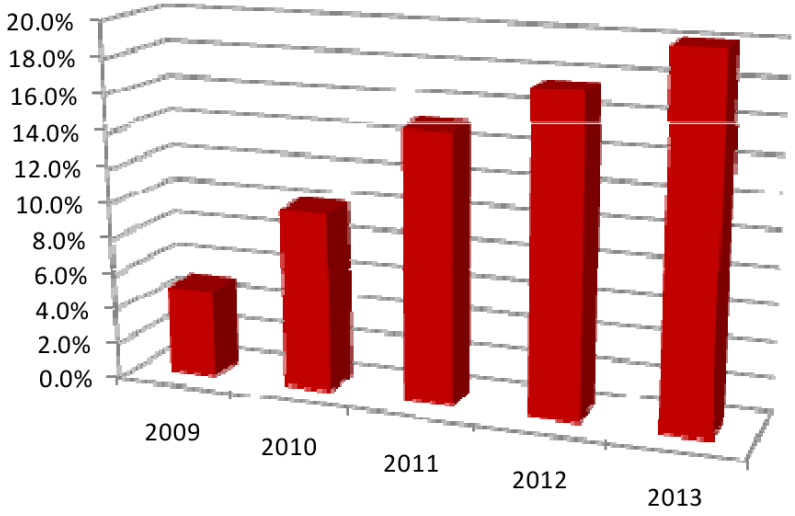
Source: IDC, "The Digital Universe Decade – Are You Ready?" May 2010

NAND Flash Worldwide Usage



Source: Gartner December, 2011
"Forecast: Semiconductor Consumption by Electronic Equipment Type, Worldwide, 4Q11 Update"

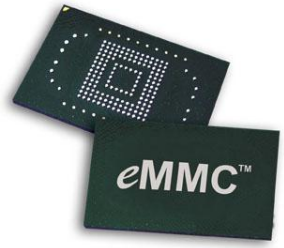
e·MMC Share of Total Flash Market

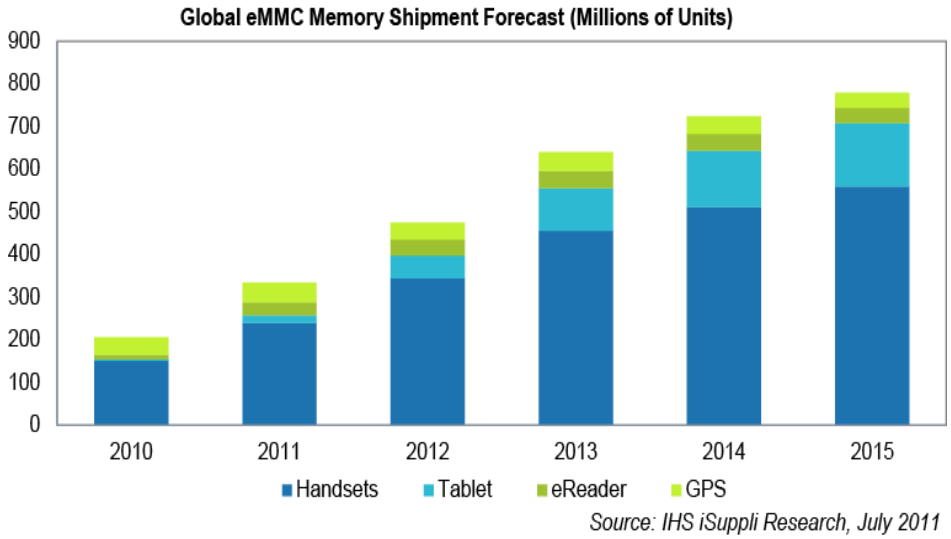


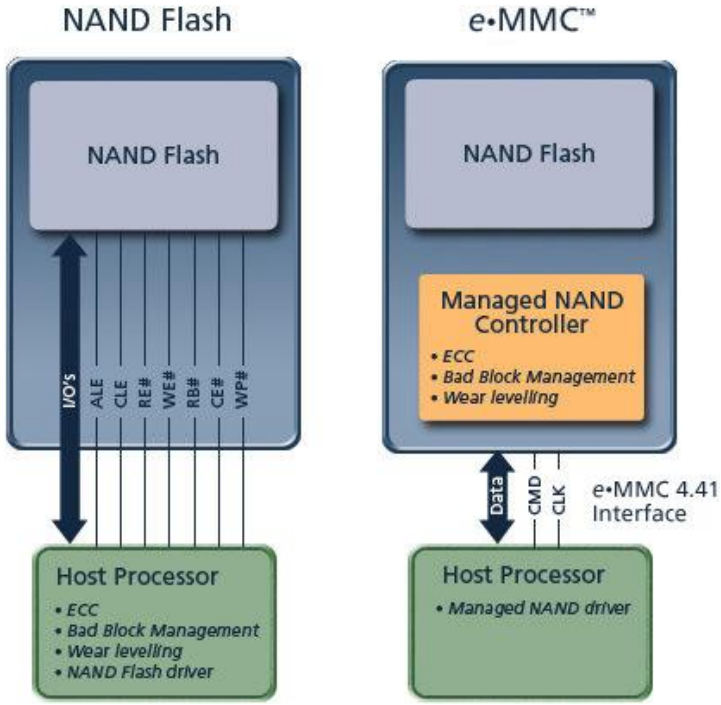
Source: Micron Marketing, 2010

e•MMC

- Embedded MultiMedia Card
  - Not a card, but a chip
  - Embedded storage solution with MMC interface, flash memory and controller
  - Propelled by increased usage in smart mobile devices







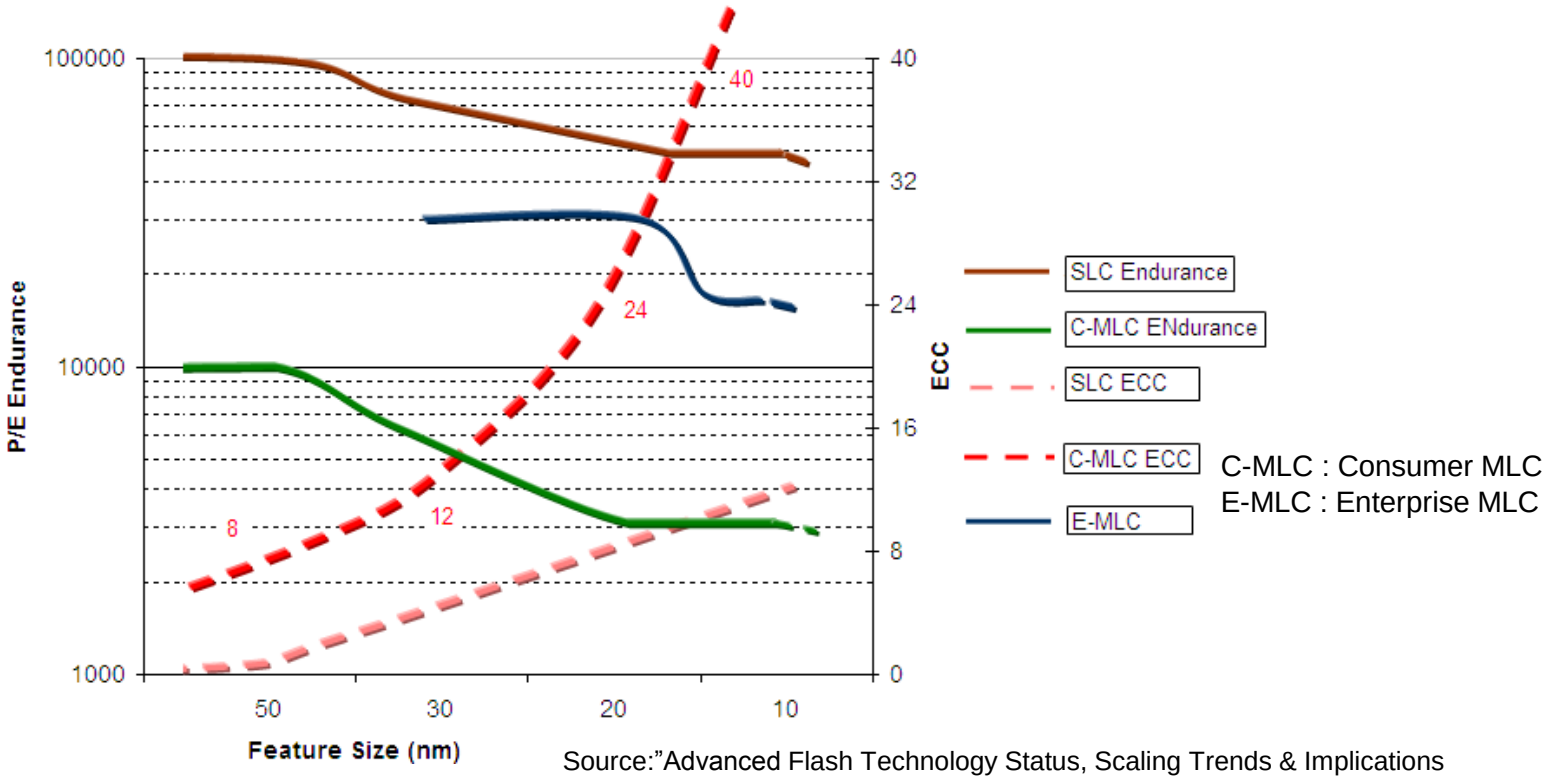
M&A in Flash Market

- **Apple** acquired Israel-based **Anobit** for about $400 million (Dec, 2011)

- **SK Hynix** merged an SSD controller maker **LAMD** for ₩287 billion (Jun, 2012)

NAND Scaling and Challenges

- As NAND Technology node is scaled down
  - Capacity grows up
  - Number of electron per cell / signal Integrity decreases
  - Reliability(endurance) decreases



Source:"Advanced Flash Technology Status, Scaling Trends & Implications to Enterprise SSD Technology Enablement", Flash Memory Summit 2012

Verification Platform for SSD Development

- SSD Core Technology and Controller
- Existing Verification Platforms
- Our Platforms

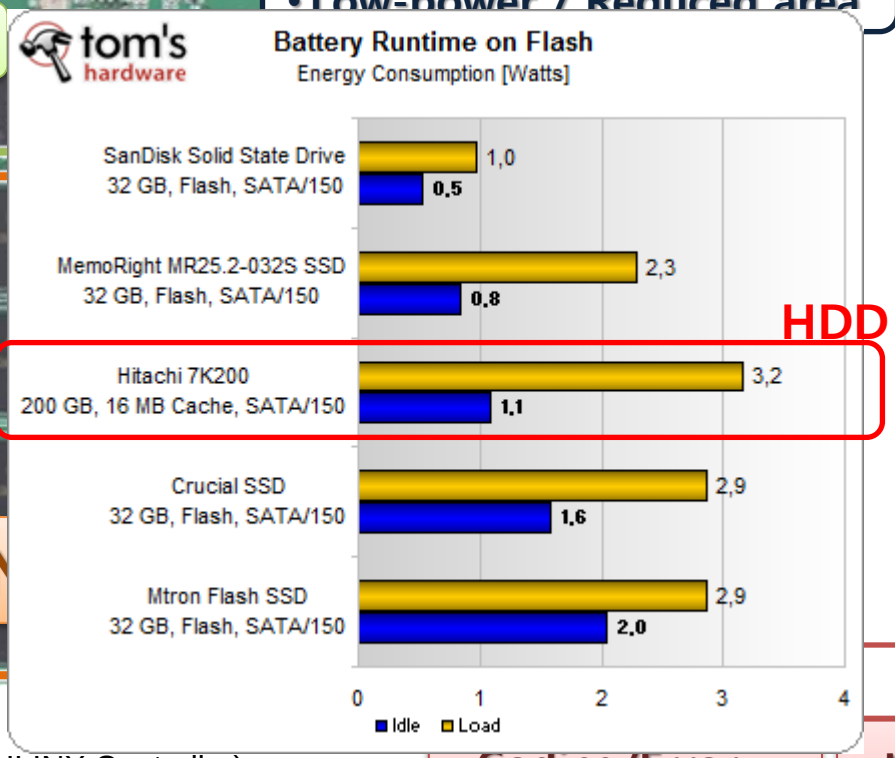
SSD Core Components and Technology

Host Interface

- From Phy to Link/Trans. layer for High-speed interface
- Low-power / Reduced area

DRAM

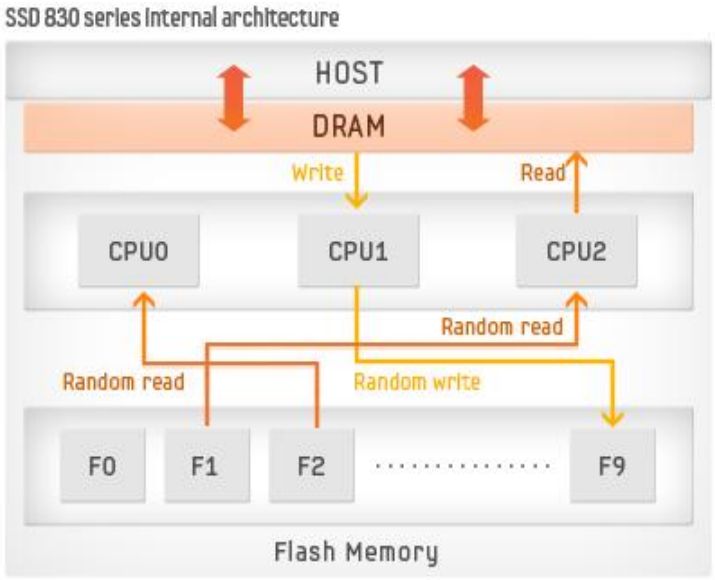
NAN



HDD

SSD
(INDILINX Controller)

Multicore Architecture in Samsung SSD



ETC

- Low-power
- Sudden power loss protection

- Coding/Error Correction
- Better UBER
- Reduced area

Development of SSD Controller

- Tera-scale SSD controller should…
  - Be able to handle multi-channel (more than 10) NAND interface
  - Contain high speed host-interface

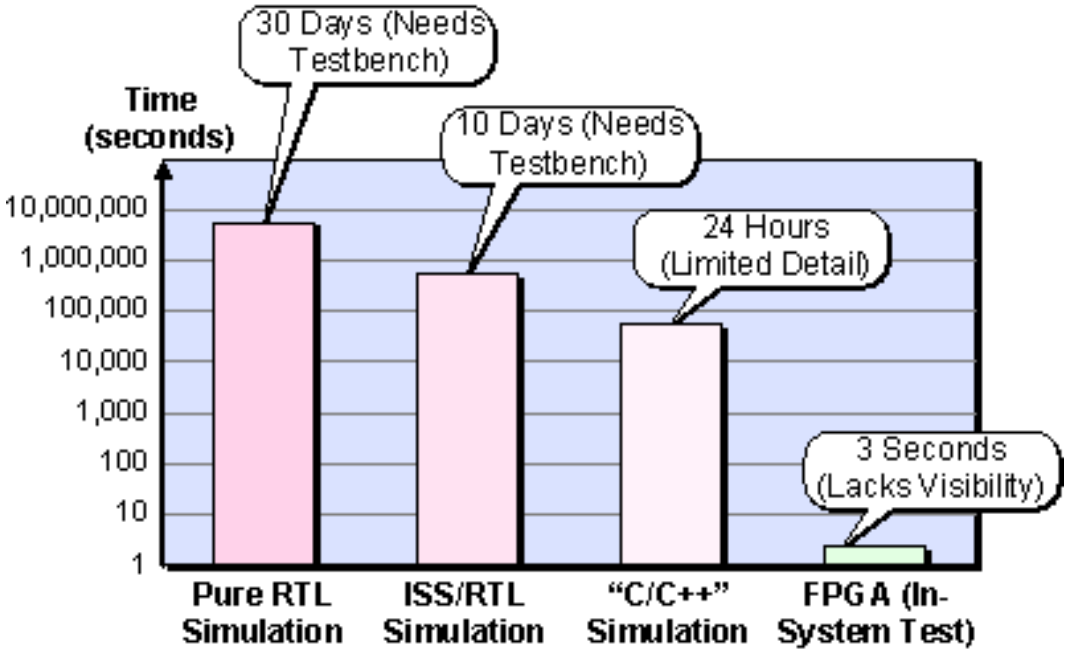
For Speed

  - Carry out garbage collection very efficiently
  - Have high performance ECC module

For Reliability

Limitation of S/W Simulation

- SSD system contains various complicated components
  - CPU, SRAM, DRAM controller, Flash controller, …
- The traditional S/W simulation is very useful but often slow to emulate the whole system
- FGPA-based simulation can help this out

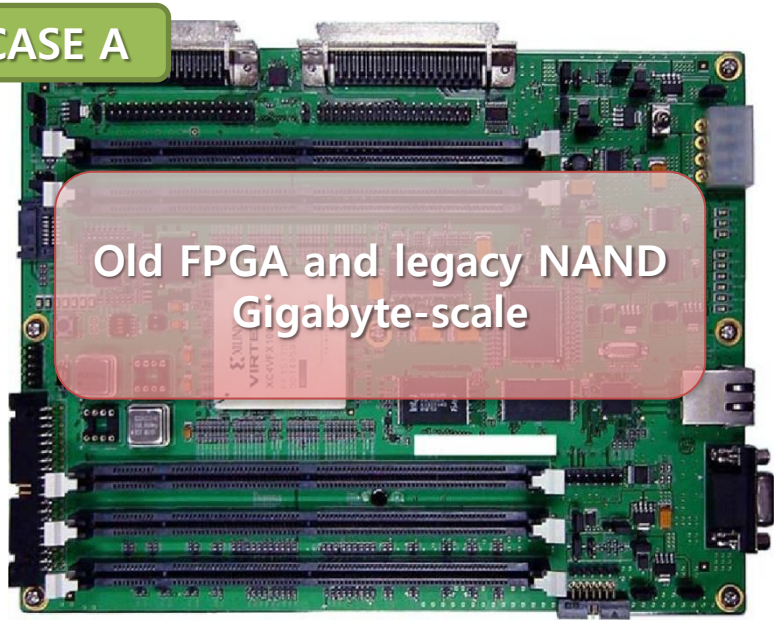


Existing Solutions

	Case A	Case B	Case C
FPGA	Virtex-4	Virtex-5 x 2	Virtex-2
Capacity	4DIMM Bank 2GB/Bank	320GB	32GB
NAND	Legacy	Legacy	Legacy
Host Interface	PATA-to-SATA	N.A.	Ethernet
Source	"Development Platforms for Flash Memory Solid State Disks", Hongseok Kim at al., ISORC 2008	"FPGA-Based Solid-State Drive Prototyping Platform", Yu Cai at al., FCCM 2011,	"BlueSSD: An Open Platform for Cross-layer Experiments for NAND Flash-based SSDs", Sungjin Lee at al., WARP 2010

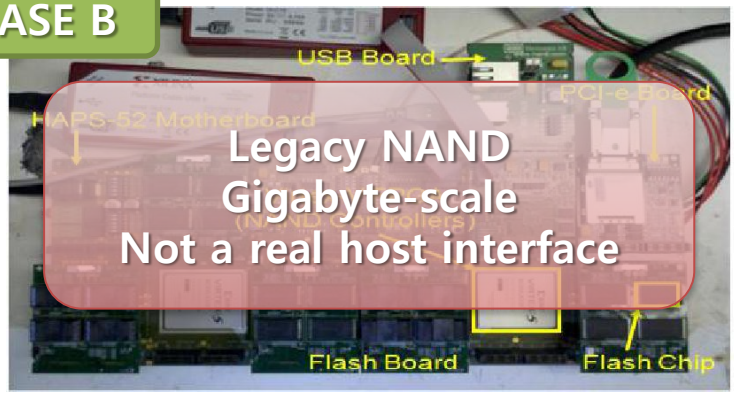
Existing Solutions

CASE A



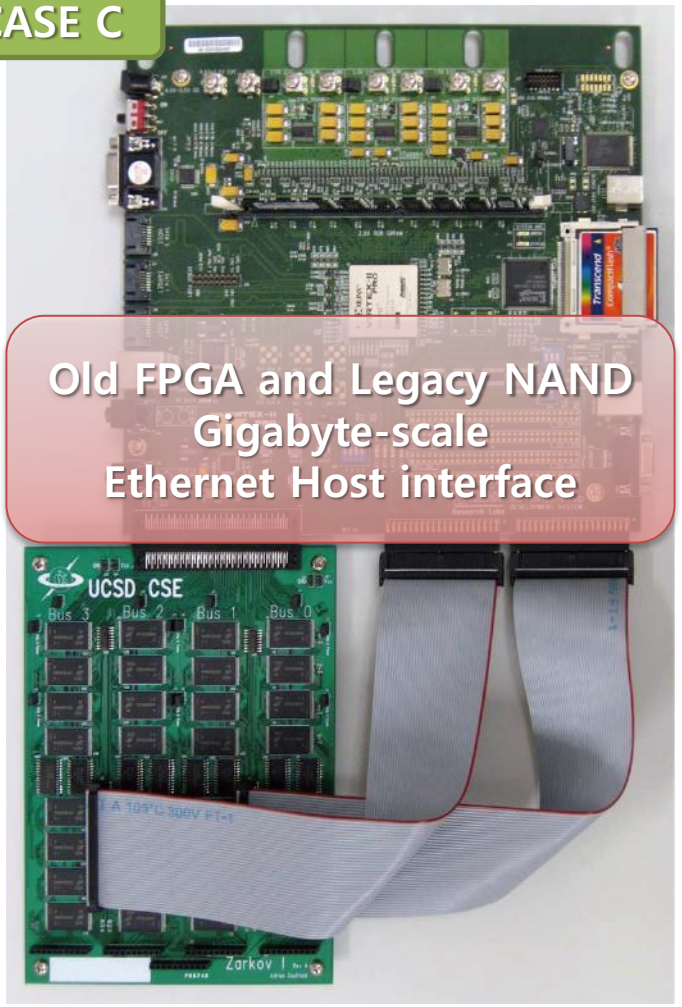
Source:"Development Platforms for Flash Memory Solid State Disks", Hongseok Kim at al., ISORC 2008

CASE B



Source:"FPGA-Based Solid-State Drive Prototyping Platform", Yu Cai at al., FCCM 2011

CASE C

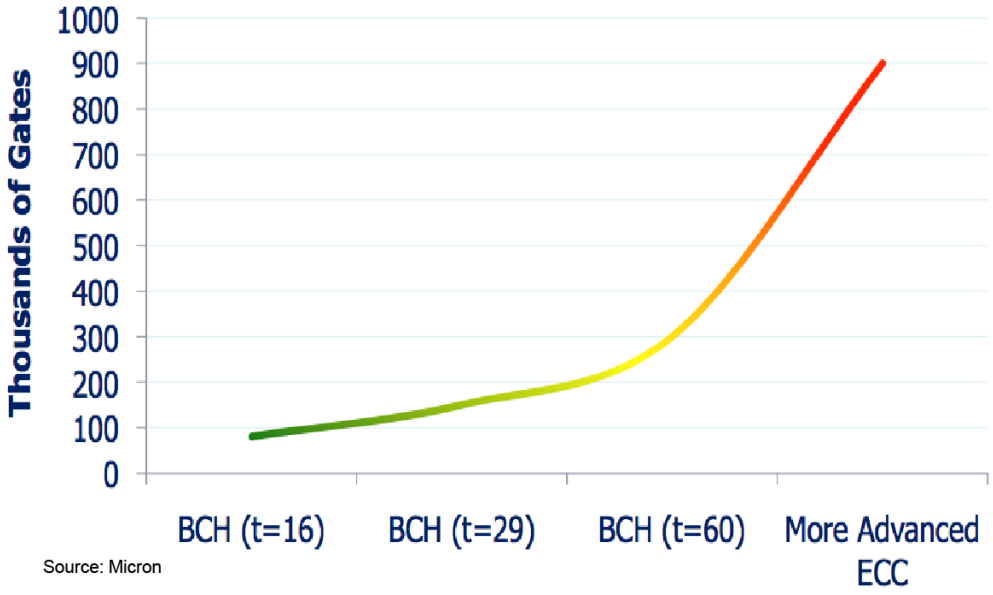


Source:"BlueSSD: An Open Platform for Cross-layer Experiments for NAND Flash-based SSDs", Sungjin Lee at al.

Running Short of FPGA Resources

- The More channel capacities, the more ECC modules

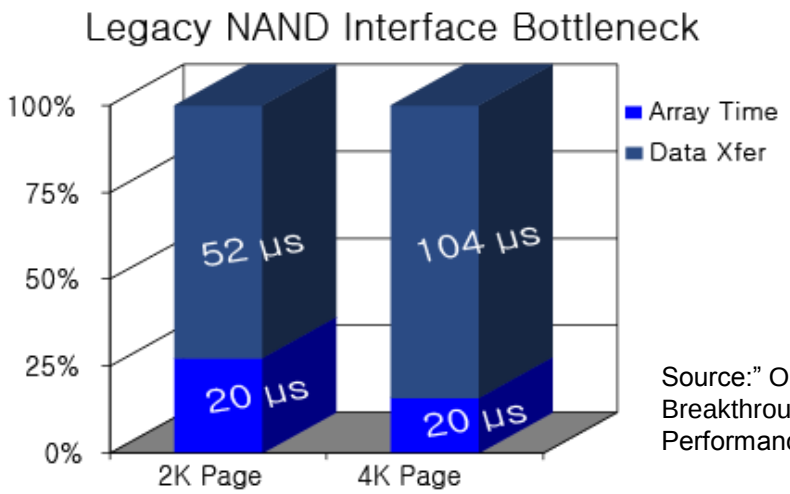
- The more reliable operation by ECC module requires the more of the FPGA's resources



Proposed Platform

- Brand new FPGA : Xilinx Virtex-6

- Support Various NAND chips, not only legacy NAND but also ONFI NAND chips



Source:" ONFi: Achieving Breakthrough NAND Performance", MEMCON08 2012

- Extend NAND channel interface up to 14-Channel
- Various helpful functions for the development
  - Power measurement per channel interface

Target Specification

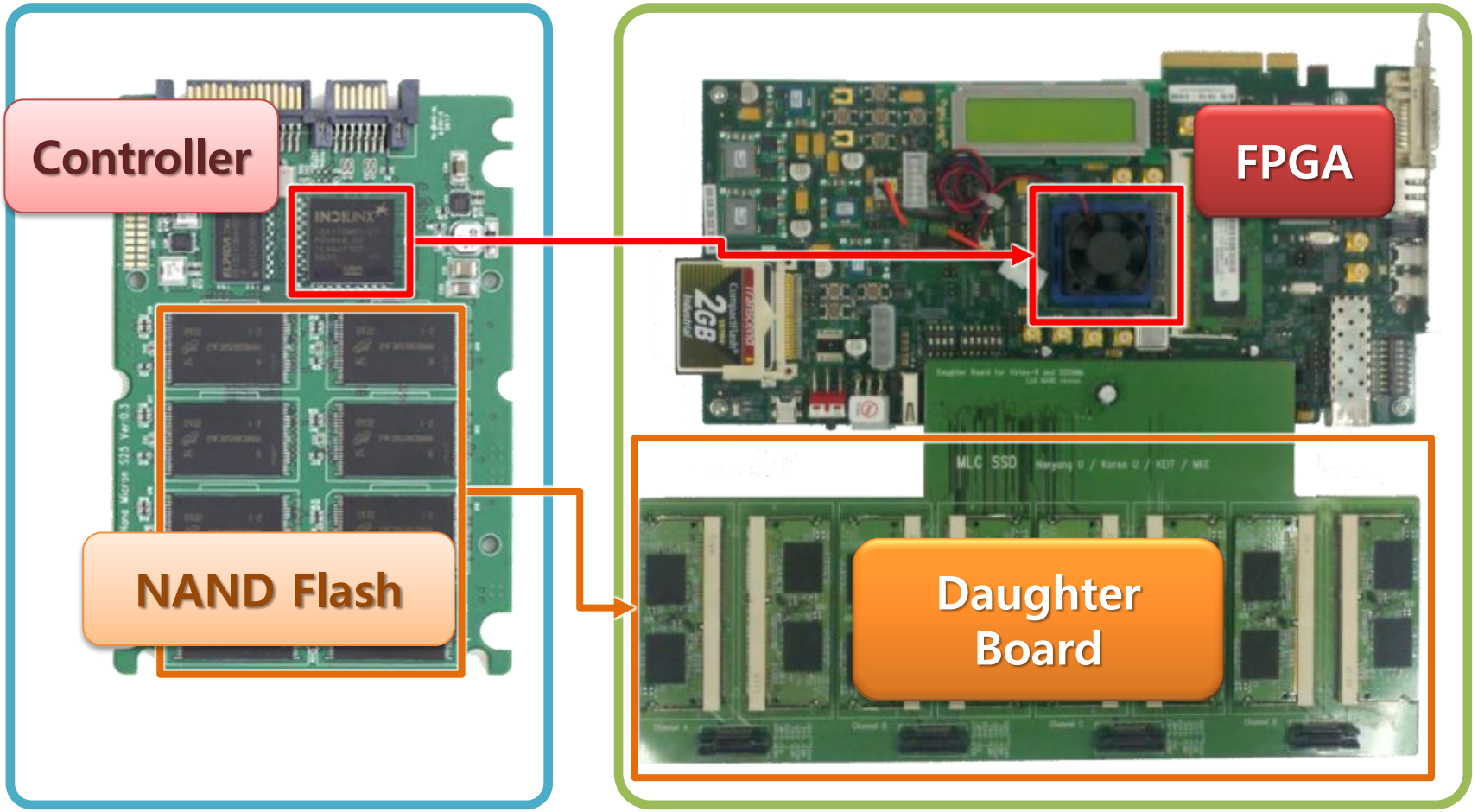
	Goal	Unit
Capacity	4,096	Gigabyte
Read Speed	1,000	MB/s
Write Speed	800	MB/s
IOPS	80	KIOPS
H/I Speed	6	Gbit/s
H/I Power Consumption	100	mW
SSP Throughput	1,200	Mbit/s

Samsung SSD 840 Pro vs 830		
	Samsung SSD 830 (256,512GB)	Samsung SSD 840 Pro (256,512GB)
Sequential Read	520MB/s	540MB/s
Sequential Write	400MB/s	450MB/s
Random Read	80K IOPS	100K IOPS
Random Write	36K IOPS	78K IOPS
Active Power Use	0.24W	0.068W
Idle Power Use	0.14W	0.042W

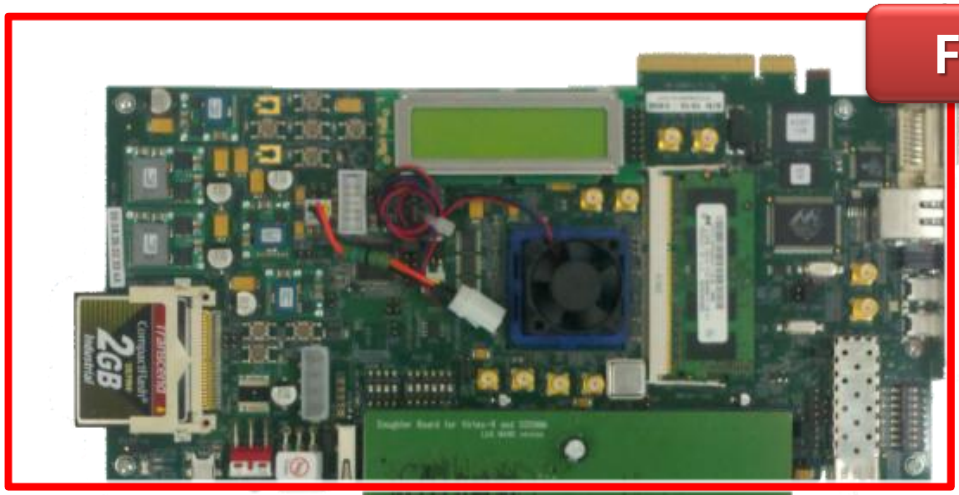
Platform v1

Real SSD

Verification Platform

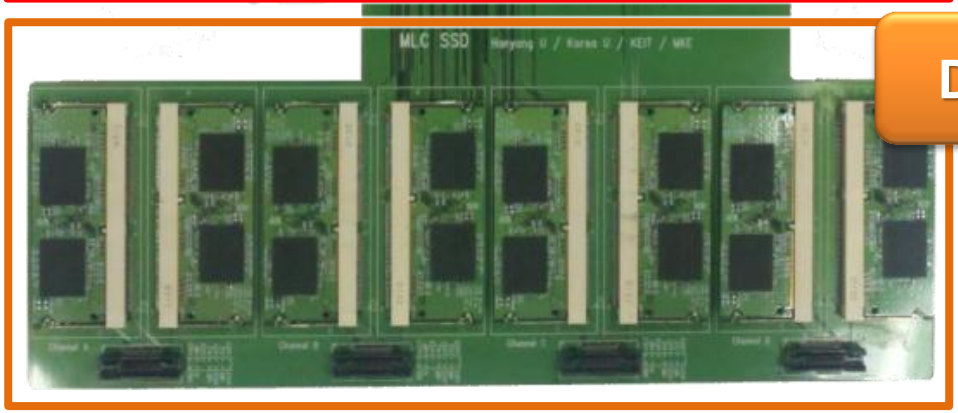


Platform v1 Details



FPGA Board

- Xilinx Virtex6 240LXT
- DDR3 SODIMM 512MB
- USB JTAG, UART
- PCIe x8

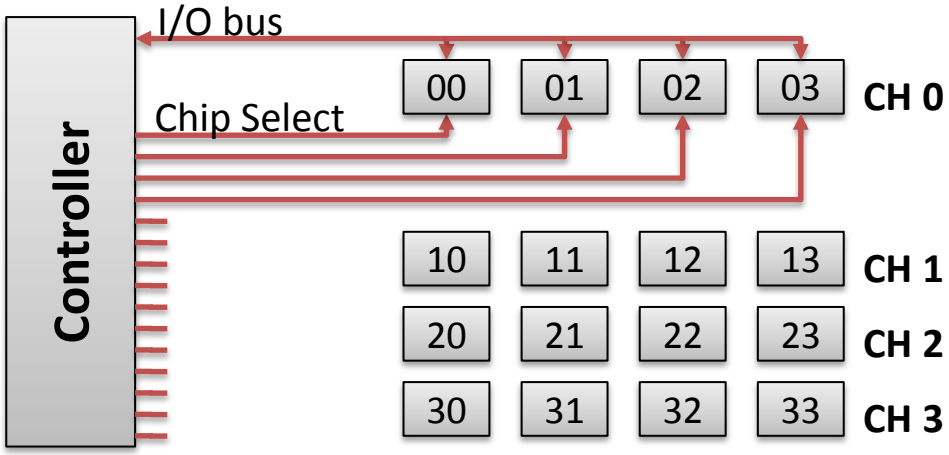


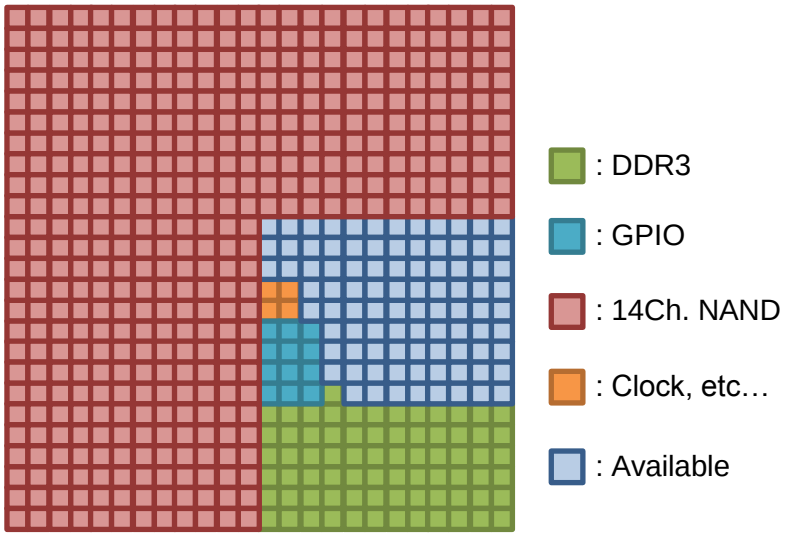
Daughter Board

- MAX up-to 4CH/ 6WAY
- Piggy-back (SODIMM)
- MICTOR connector for debugging
- Both legacy and ONFI NAND

Limitations of Platform v1

- Too few I/O pins, which limits capacity (30 signals per one channel)
- Use of Ready/Busy signals not possible (suboptimal I/O management)

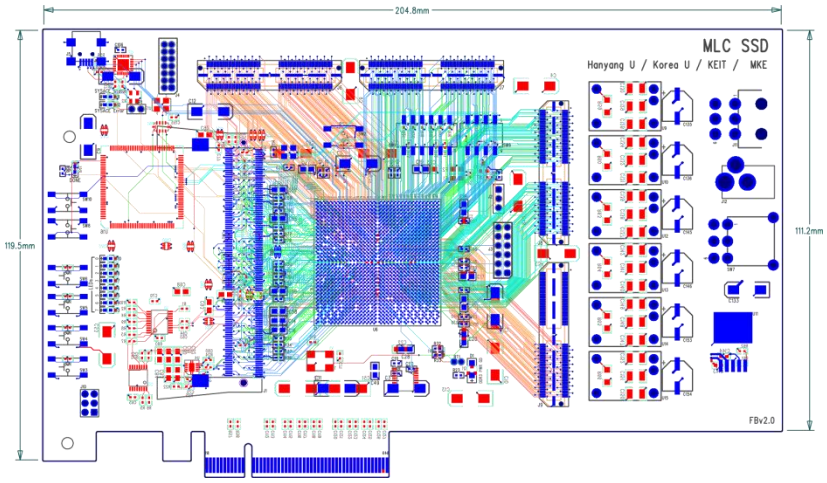


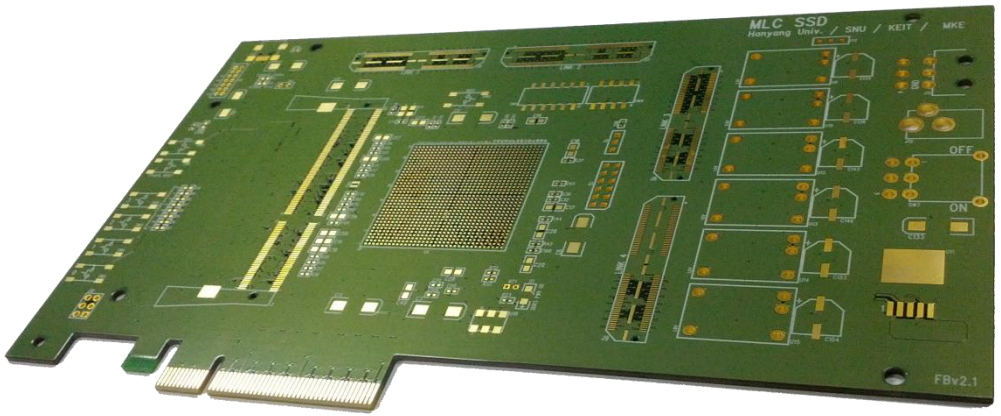


Virtex-6 I/O Pins Mapping Diagram

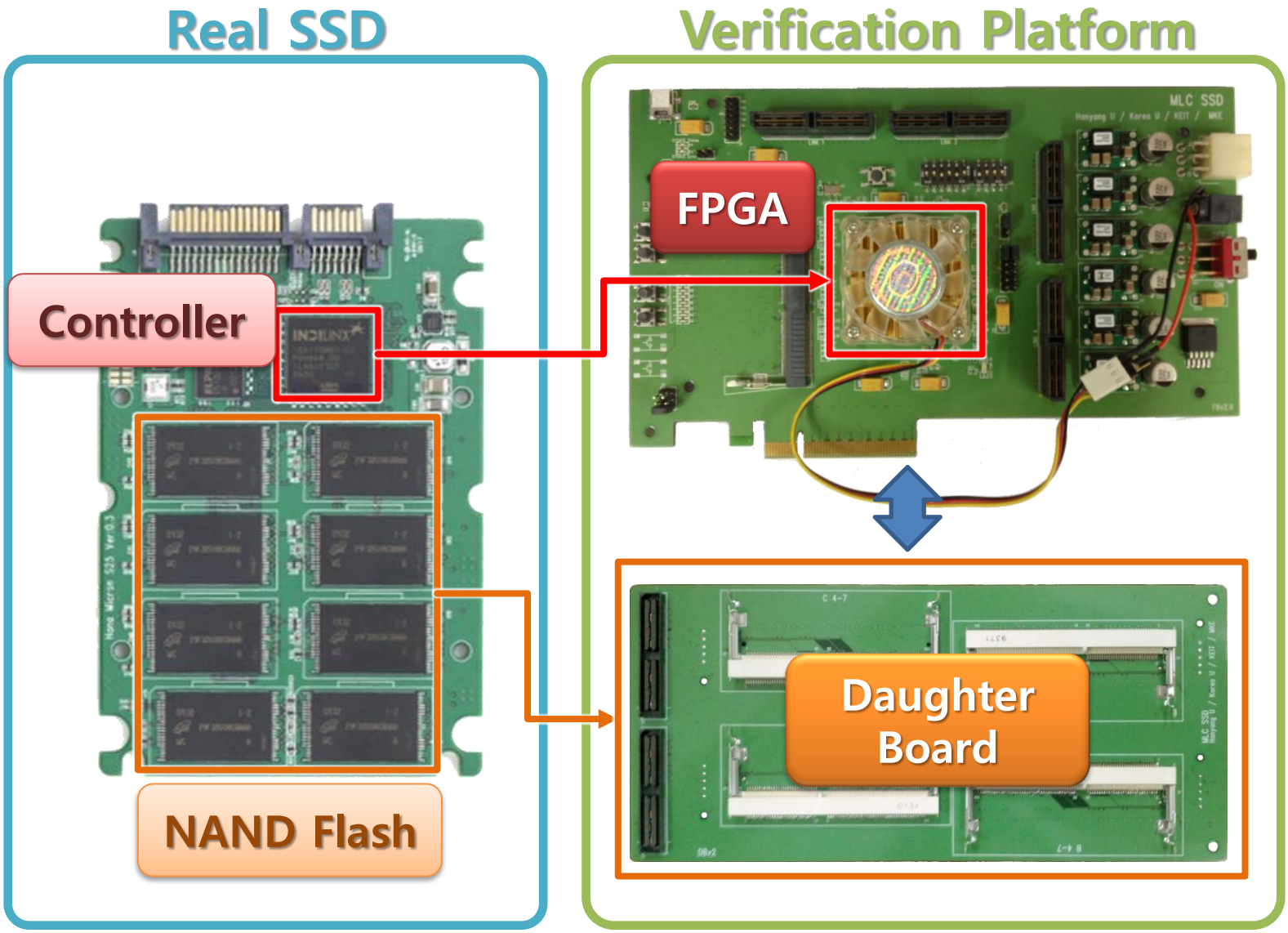
Improvement Ideas

- Design a custom FPGA board
  - Selectively benchmark the reference board
- Maximize utilization of FPGA I/O pins for NAND interface
- Support up to 14-CH/8-Way
- Efficient use of Ready/Busy signals

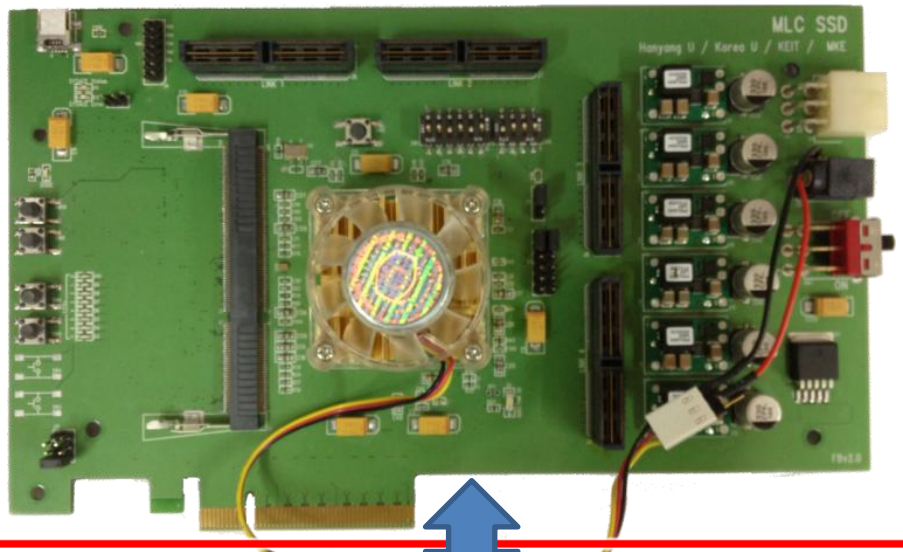




Platform v2


Real SSD
Verification Platform
Controller
FPGA
NAND Flash
Daughter Board

Platform v2 Details



FPGA Board

- Our own design
- Xilinx Virtex6 240LXT
- DDR3 SODIMM 512MB
- MAX up-to 14CH / 8WAY

Daughter Board

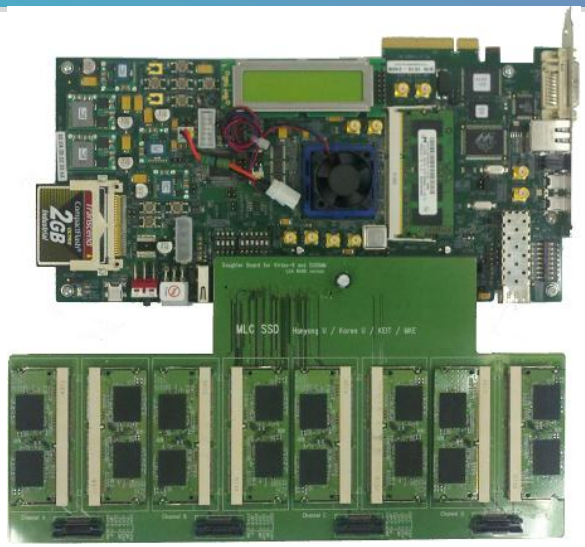
- 2 Daughter board
- Each one for 8CH - One of 8CH, another with 6CH(2CH not used)
- High speed/reliable SAMTEC connector interface
- Piggy-back (SODIMM)
- MICTOR connector for debug

Implement of NAND Controller

- Specification
- NAND Status Monitor
- ONFI Interface

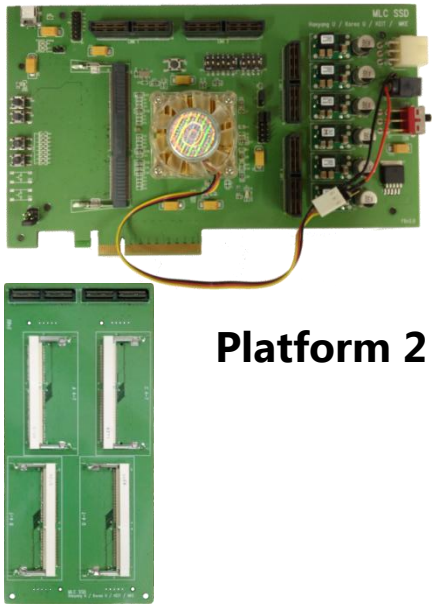
Step of implementing NAND Controller

- **The first stage of NAND Controller**
  - Target: Platform v1
  - 4-channel 4-way
  - Legacy NAND flash
  - No R/B pins



Platform 1

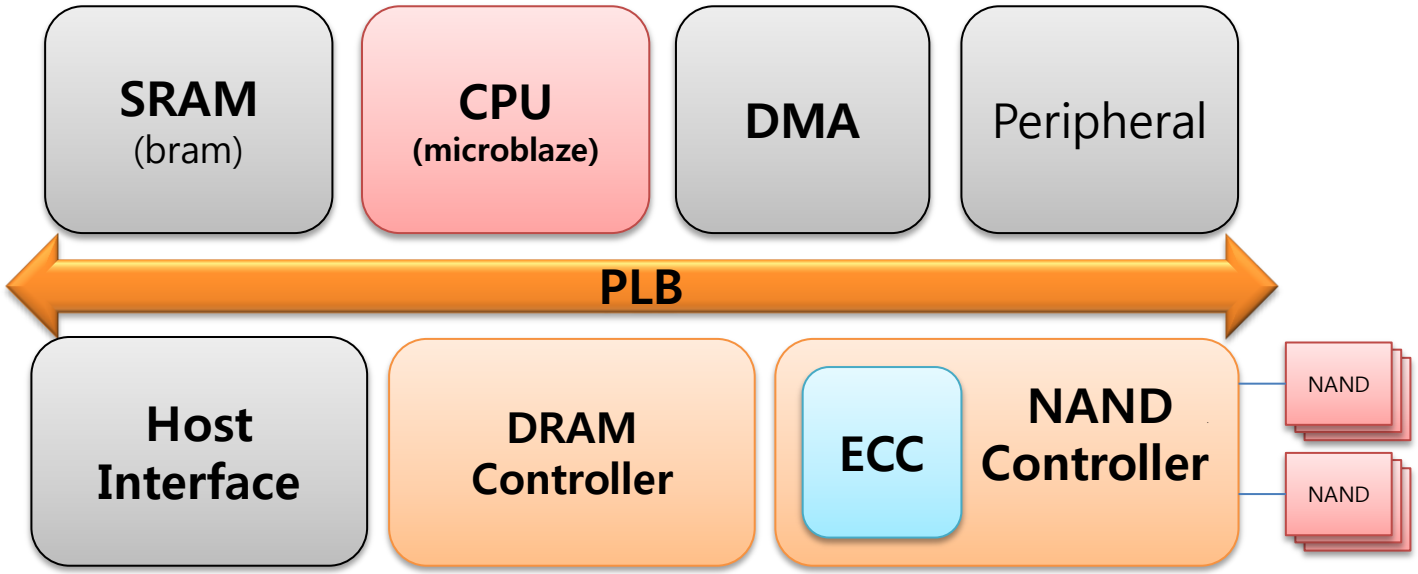
- **Tera-scale NAND Controller**
  - Target: Platform v2
  - 14-channel 8-way
  - ONFI NAND flash
  - 1 R/B pins per channel



Platform 2

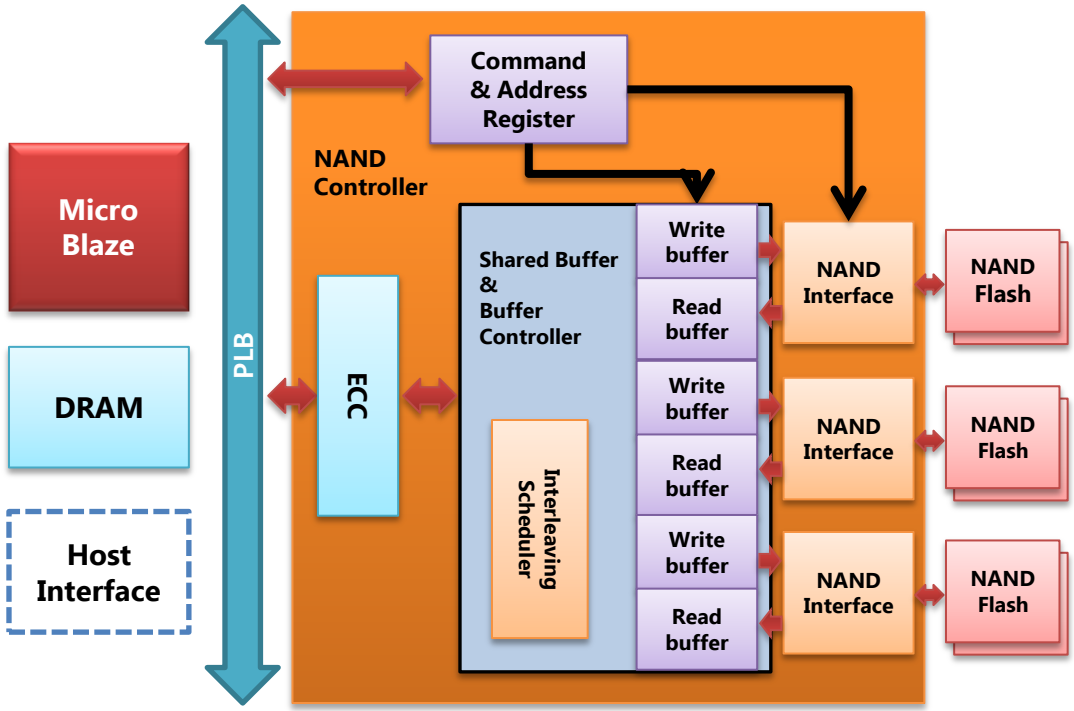
Baseline Architecture

- **Specifications**
  - CPU: 150MHz Microblaze softcore
  - RAM: 512MB SDRAM
  - BUS: PLB



NAND Controller

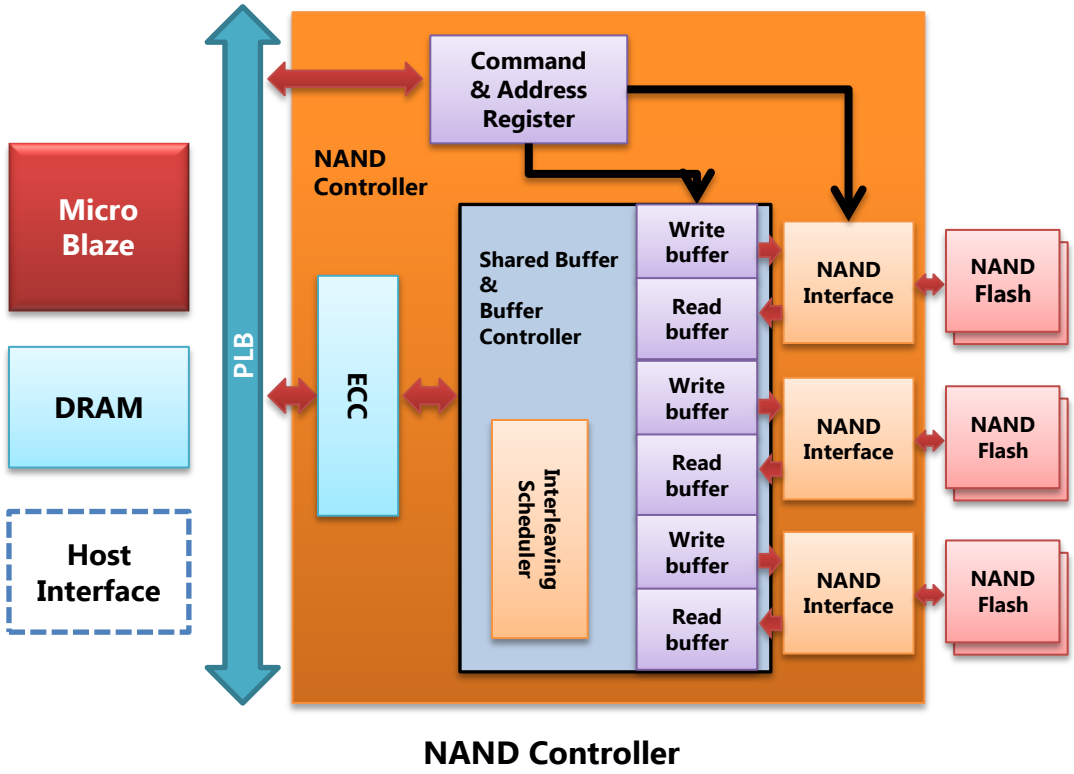
- **NAND Flash**:
  - **8GB Samsung Legacy NAND**
    - 33.3MHz interface clock
  - **16GB Micron ONFI 2.2 NAND**
    - 100MHz interface clock
- **Shared Buffer** (for multi channel)
  - 8 – 32 pages buffer
  - Buffer sharing policy
    : 4-way associativity
- **ECC** : 28-bit BCH
  - Parity 392 bits
- **Capacity**
  - **4 channel 4 way : 128GB (v1)**
  - **14 channel 8 way : 1792GB (v 2)**



NAND Controller

NAND Controller

- **Channel structure**
  - Command / Address / Status Register
  - 16KB Read/Write Buffer (v1)
  - No R/B signal (v)
    1 bit R/B per 1 channel (v2)



NAND Controller

  - **FPGA usage**
    - Microblaze system : 6%
    - NAND Controller : 2% per channel

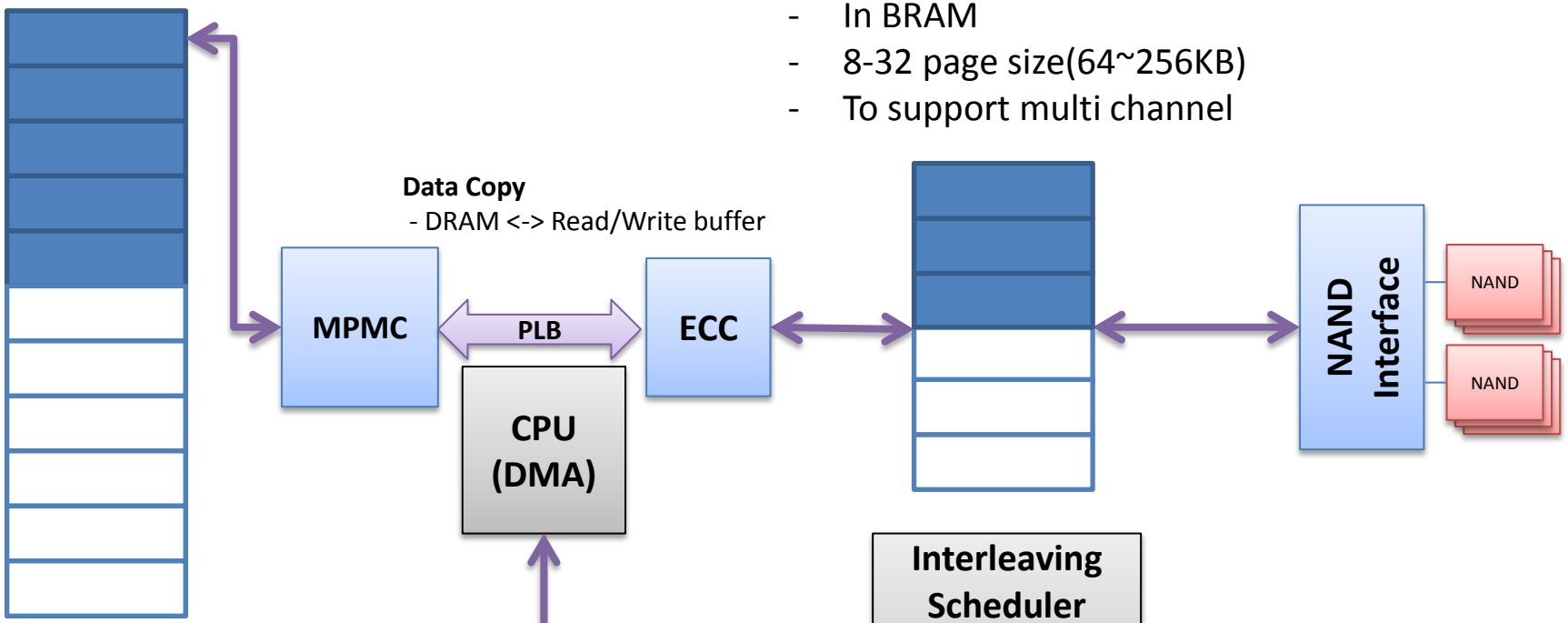
Register Map

- **Register**
  - NAND interface에 System의 커맨드를 전달
- **Address Register**
  - Using Physical Address : FTL에서 직접 입력
  - Table size overhead 없음
  - Physical address를 통해 page 정확한 상태 정보 FTL에 전달

- **Command Register**
  - Chip selection, Command data, NAND interface 상태 정보 저장
- **Status Register**
  - 모든 chip의 Ready/Busy 및 커맨드 성공/실패 상태 확인

Data Flow

- **NAND Controller Data Flow**



Shared Read/Write buffer
- In BRAM
- 8-32 page size(64~256KB)
- To support multi channel

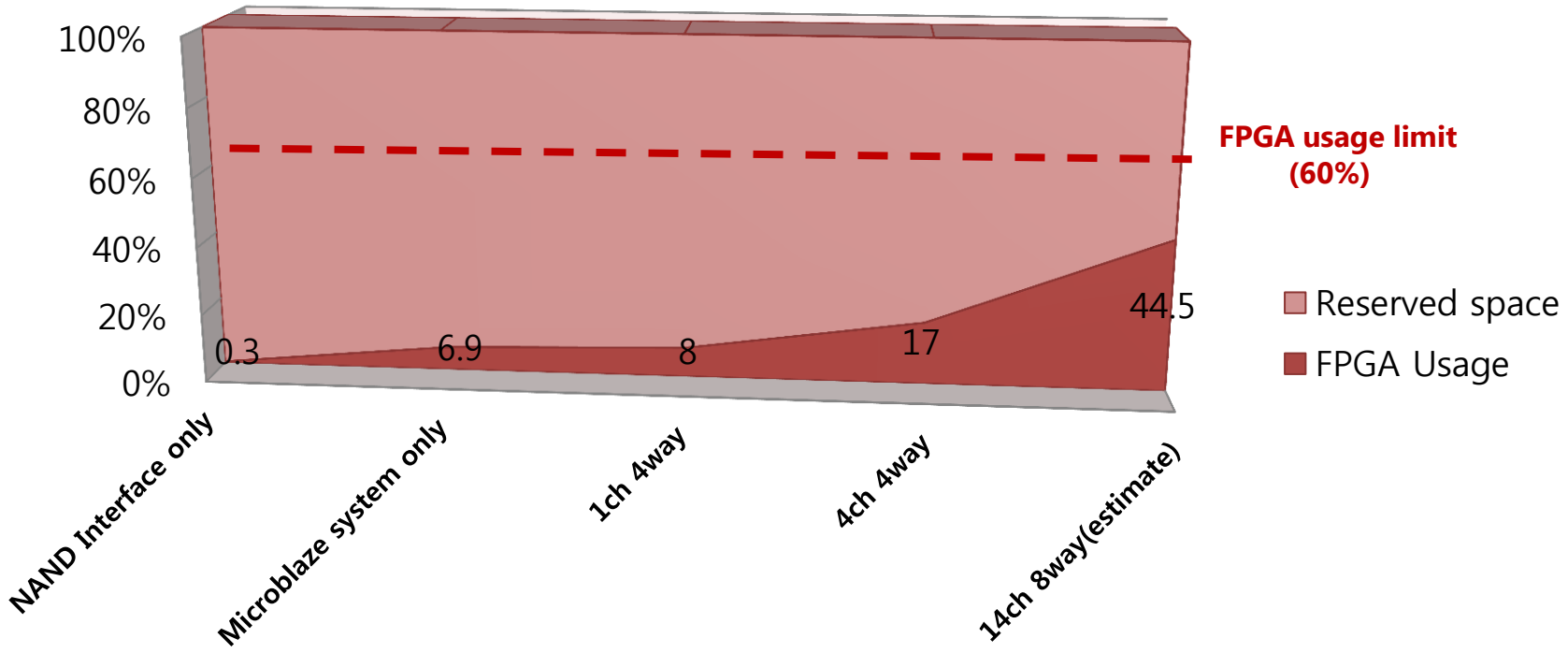
Data Copy
- DRAM <-> Read/Write buffer

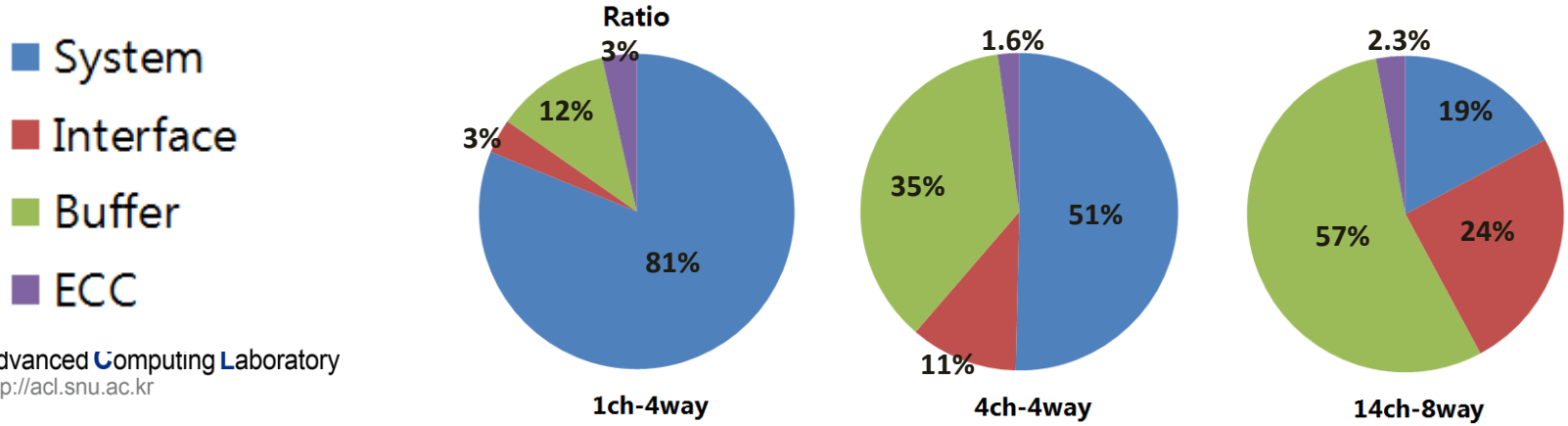
Host Interface Queue
- In DRAM
- priority Queue

Interleaving scheduler
- Register decoder
- Control NAND interface

FPGA usage

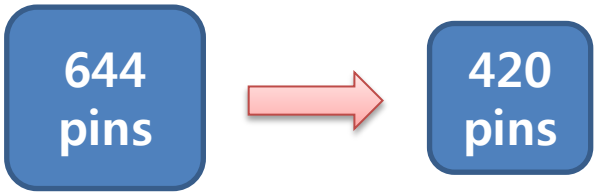
FPGA usage of NAND Controller



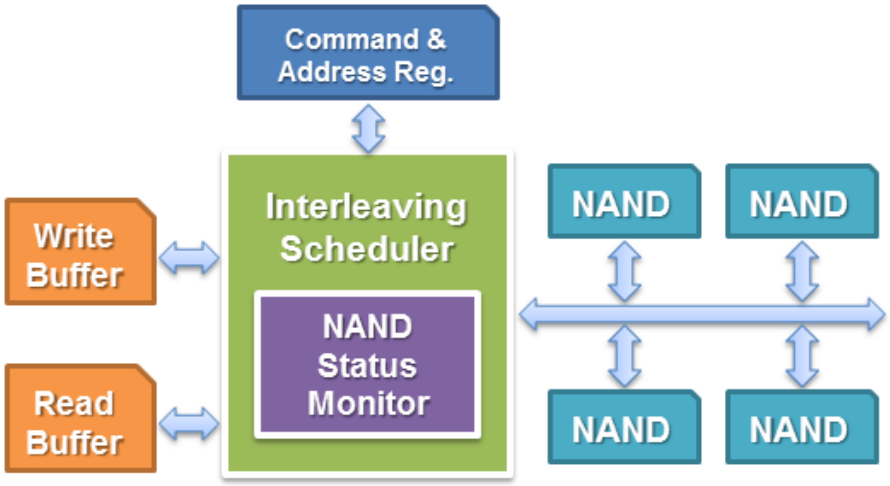


NAND Status Monitor

- **Motivation**
  - Insufficient user I/O pins
  - **Pin counts without R/B**



- **NAND status monitor (NSM)**
  - Use only **1 R/B** or even **without R/B**
  - Update NAND status in near real time



NSM in Interleaving Scheduler

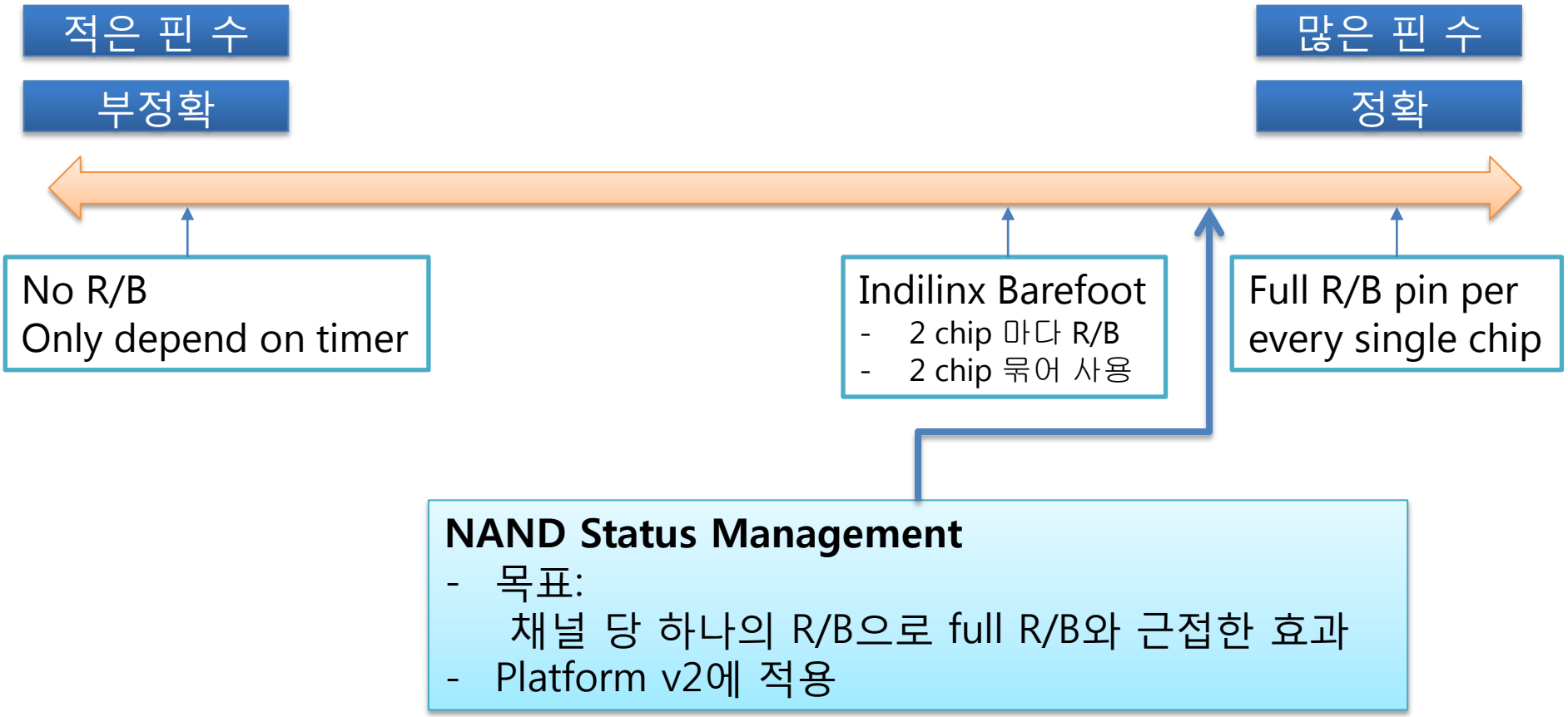
- **Idea**
  - Throw read status command automatically – polling strategy
  - Save past operation time in table

Real-time NAND Status Monitor

- **R/B Signal per channel: 1 R/B signal** per 1 channel (Platform v2)
- **Channel interleaving**
  - Initial read status point : 1 Bank 상황에서 측정한 operation time OR Typical Time
  - 2개 이상의 Bank가 Busy인 경우는 R/B signal이 없을 때와 동일하게 동작

Method	Pros	Cons
Fully depend on R/B	Ideal case	많은 수의 IO pin필요
NSM with No R/B (using timer)	IO pin 필요 없음	여러 개의 bank가 동시에 busy일 경우 정확한 완료 타이밍 알기 어려움
NSM with 1 R/B	1개의 R/B로 정확한 Initial Value 설정 가능	Pin count가 ch당 1개 늘어남
Time multiplexing R/B pins	채널당 핀 수 : log(ways) NAND가 느린 특성 이용	MUX를 FPGA 외부에 장착 보드 제작 후 수정 어려움

Real-time NAND Status Monitor

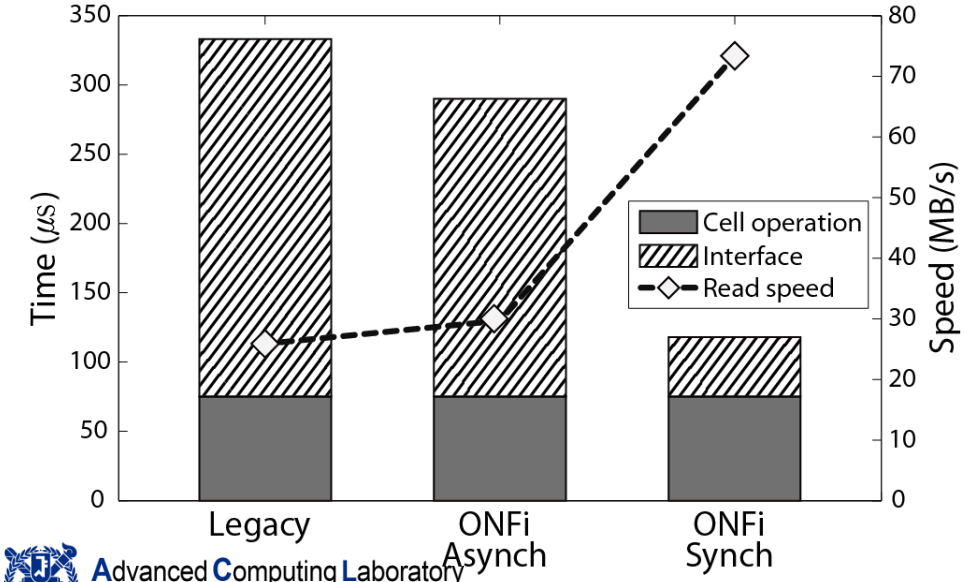


Implementing ONFI Interface

- **Core Features**
  - **ONFI Interface**
    - ONFI 2.2 Interface 구현 - using DQS(Data Strobe)
    - ONFI 2.2 최대 interface 속도 지원(100 MHz, Legacy는 33.3MHz)
    - DDR protocol 적용 (Interface 속도 2배 향상)
  - **Backward compatibility**
    - Legacy NAND 와 ONFI NAND 동시 사용
    - Legacy – Asynchronous mode
    - Synchronous mode (ONFI 2.2)
  - No R/B pins
    - NAND Status Management 모듈 사용
  - **Implementation**
    - Verilog RTL
    - Implemention & debugging in Platform v1

Legacy vs ONFi 2.2

Model	K9LCG08U1M	MT29F128G08CECABH1
Type	Legacy	ONFi2.2 (Synchronous mode)
Test capacity	16GB	32GB
사용 BD	Original Platform 1	Revised Platform 1 for ONFI
Interface Speed	33.3 MB/s	200 MB/s
Write Speed	3.47 MB/s	6.4 MB/s
Read Speed	25.9 MB/s	73.4 MB/s



- Legacy & ONFI
  NAND Interface 속도 최적화

Experiments

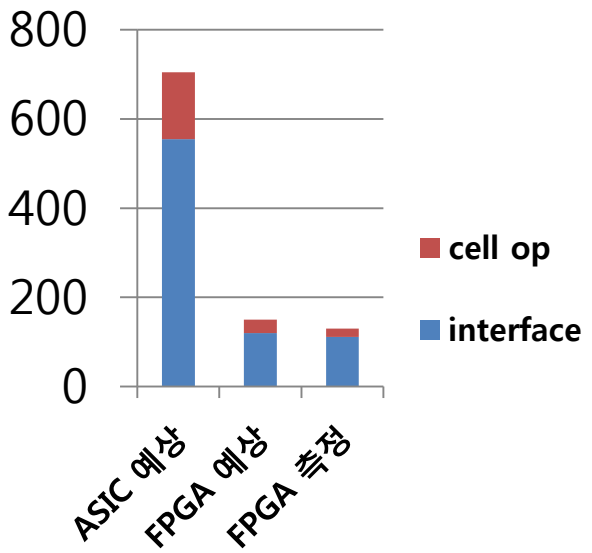
- Maximum performance test
- FTL
- ECC
- NSM TEST

Maximum Performance test

- **Verifying Maximum Capacity**
  - Capacity : 14 channel 8way - > 1.792 GB
  - FPGA elements 사용량 60% 이하 조건
  - 1.792 GB 전체 읽기/쓰기/지우기 정상 동작 확인

- **Maximum throughput test**
  - Throughput : 순차/랜덤 읽기/쓰기 속도
  - Bank(8GB)의 개수 : 14 x 16 = 224
  - 하나의 Bank 최대 쓰기 속도: 6.4MB/s
  - ASIC 컨트롤러 쓰기 속도 예상값: 1433.6MB/s
  - FPGA 예상 최대 쓰기 속도 예상값 : 약 150MB/s

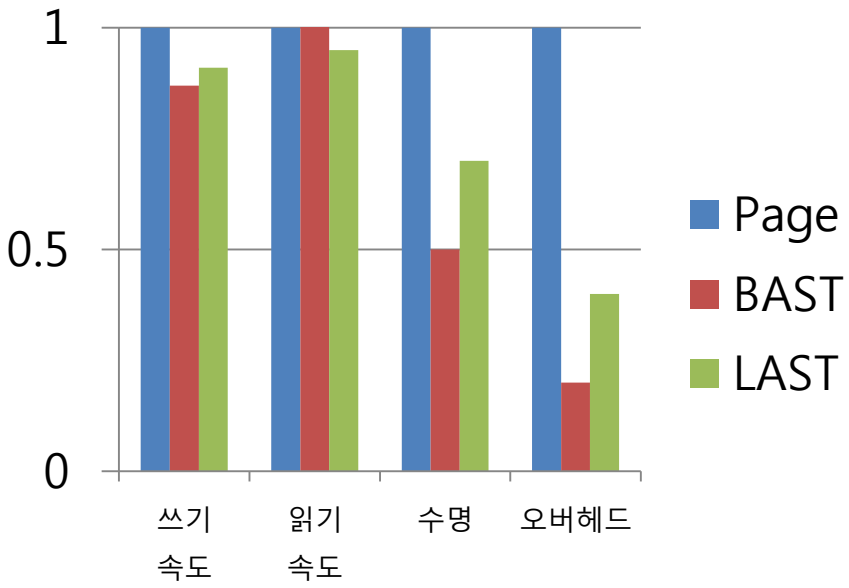
기능성 및 성능 검증



Read speed of NAND Controller
(예상도)

FTL development & verification

- **FTL**
  - Page mapping
  - Block mapping
  - BAST
  - LAST
  - User custom FTL
  - **속도, 수명, 오버헤드 비교**

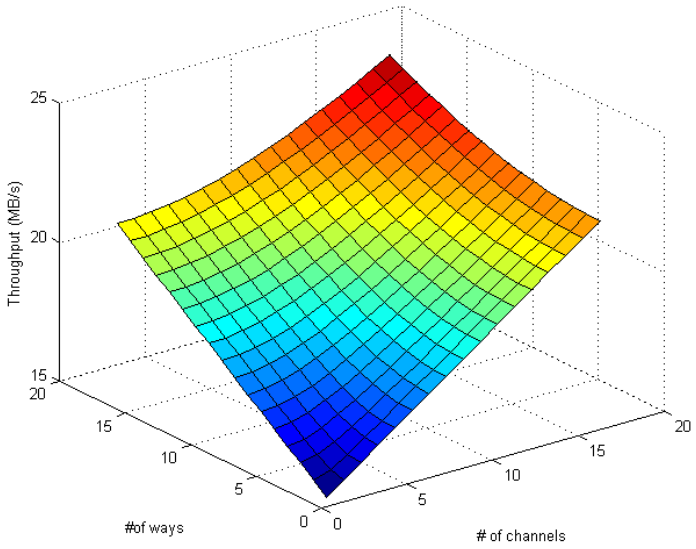


Page mapping 대비 FTL 성능 비교(예상도)

Architecture exploration

- **Architecture exploration**
  - # of channels: 1, 2, 4, 8, 14
  - # of ways: 1, 2, 4, 8
  - Shared buffer
    - Various size & sharing policy

- **Measuring**
  - Channel, way 변화에 따른 throughput
  - Power consumption
  - channel-way trade off
  - ECC 크기 변화에 따른 error 변화



SSD의 대부분의 Factor에 대해 측정 가능

Other experiments

- **ECC**
  - Basic option: 28bit BCH (parity 392 bits)
  - NAND Spare area 제약으로 (440 bits) 더 큰 BCH 적용은 힘듬
  - Other options
    - **Small size BCH**
    - **LDPC**
    - **Architecting ECC module(s)**
  - 결과
    - ECC 사용에 따른 오버헤드 (속도, 게이트 사용량) 비교
    - Error 비율과 오버헤드에 따른 trade off 분석

- **NAND Status Monitor**
  - No R/B vs 1bit R/B per 1 channel 성능 비교