Summarizer:
Trading Communication with Computing Near Storage

Gunjae Koo*, Kiran Kumar Matam*, Te I†, H.V. Krishna Giri Nara*, Jing Li‡,
Hung-Wei Tseng†, Steven Swanson‡, Murali Annavaram*

*University of Southern California
†North Carolina State University
‡University of California, San Diego
Motivation – *High Data Movement Cost*

- Limited data bandwidth
- High access latency

- Data computation @ host
- Data transfer from storage
Near Data Processing (NDP)

- **Host**
  - CPU
  - Storage interface

- **Storage Processor (SP)**

**Diagram Details:**
- Data computation @ host
- Data transfer from storage
- External (host – storage)
- Internal
Near Data Processing (NDP)
Near Data Processing (NDP) on SSDs

Data computation @ host

Data computation @ storage

Data transfer from storage

Garbage collection

Wear-leveling

W/O NDP

With NDP
Obstacles to in-SSD processing

- **Less powerful embedded processor**
- **Dynamic computation resource availability**

**Summarizer: Dynamic NDP framework for SSD**
Summarizer – Basic Concept
Summarizer – Basic Concept

- Host
- CPU
- Storage interface
- Monitoring resources
- AP
- Database
Normal Page Read Request

Host CPU
- User Applications / Operating Systems
- NVMe Host Driver
  - RD (LBA)

Host Memory
- SQ
- CQ

Storage Interface (PCIe / NVMe)

SSD Firmware
- I/O Controller
  - (NVMe command decoder)
- Flash Translation Layer (FTL)
  - (RD) PPA

SSD SoC Interconnection
- Flash Controller
- DRAM Controller
- NAND Flash
- SSD DRAM

Summarizer
- Task Controller
- User Functions

Request Queue
Response Queue
Normal Page Read Request

- **Host CPU**
  - User Applications / Operating Systems
  - NVMe Host Driver

- **Host Memory**
  - SQ
  - CQ

- **Storage Interface (PCIe/NVMe)**

- **SSD Firmware**
  - I/O Controller (NVMe command decoder)
  - Flash Translation Layer (FTL)

- **SSD SoC Interconnection**
  - Flash Controller
  - DRAM Controller
  - SSD DRAM
  - Page data

- **NVMe Host Driver**

- **SSD DRAM**

- **User Functions**
  - User Applications / Operating Systems
  - SSD Firmware
  - I/O Controller (NVMe command decoder)
  - Flash Translation Layer (FTL)
  - SSD SoC Interconnection
  - Flash Controller
  - DRAM Controller
  - SSD DRAM
  - Page data

- **Summarizer**
  - Task Controller
  - User Functions
Normal Page Read Request

Host CPU

User Applications / Operating Systems

NVMe Host Driver

Host Memory

Storage Interface (PCIe / NVMe)

SSD Firmware

I/O Controller (NVMe command decoder)

Flash Translation Layer (FTL)

Request queue

Response queue

SSD SoC Interconnection

Flash Controller

DRAM Controller

NAND Flash

Page data

Task Controller

Summarizer

User Functions
Summarizer – Initialization (Function Offloading)

Host CPU

User Applications / Operating Systems

Storage Interface (PCIe / NVMe)

SSD Firmware

I/O Controller (NVMe command decoder)

Flash Translation Layer (FTL)

SSD SoC Interconnection

Flash Controller

NAND Flash

SSD DRAM

Summarizer

Task Controller

User Functions

f#1  foo()

Function registration

New NVMe command

NVMe Host Driver

INIT (foo)

Request queue

Response queue

foo()
Summarizer – Computation (Dynamic mode)

Host CPU
User Applications / Operating Systems

Host Memory

Storage Interface (PCIe / NVMe)

New NVMe command
NVMe Host Driver
RD&PROC( LBA,foo)

I/O Controller
(NVMe command decoder)

Flash Translation Layer (FTL)
RD&PROC(PPA,foo)

Request queue
Response queue

I/O Controller

SSD SoC Interconnection

Flash Controller
DRAM Controller

SSD DRAM
NAND Flash

Task Controller

User Functions
f#1 foo()
f#2 goo()
Summarizer – Computation (Dynamic mode)

Host CPU
- User Applications / Operating Systems
- NVMe Host Driver

Storage Interface (PCIe / NVMe)

Host Memory
- SQ
- CQ

SSD Firmware
- I/O Controller (NVMe command decoder)
- Flash Translation Layer (FTL)
  - RD&PROC(PPA,foo)
  - RD&P(PPA,foo)
  - RD&P(PPA1,foo)

SSD SoC Interconnection

SSD DRAM

DRAM Controller

Page data

Task Controller
- TQ

User Functions
- f#1
- foo()
- f#2
- goo()
Summarizer – Computation (Dynamic mode)

Host Memory

- User Applications / Operating Systems
- NVMe Host Driver

Storage Interface (PCIe / NVMe)

- SSD Firmware
  - I/O Controller (NVMe command decoder)
    - Flash Translation Layer (FTL)
      - RD&PROC(PPA,foo)
- Task Controller
  - TQ
    - TQ is full
  - User Functions
    - f#1
    - foo()
    - f#2
    - goo()

SSD SoC Interconnection

- Flash Controller
- DRAM Controller
- NAND Flash
  - Page data

Host CPU

- User Applications / Operating Systems
- NVMe Host Driver
Summarizer – Finalization

- Host CPU
- User Applications / Operating Systems
- Storage Interface (PCIe / NVMe)
- SSD Firmware
  - I/O Controller (NVMe command decoder)
  - Flash Translation Layer (FTL)
- SSD SoC Interconnection
- Flash Controller
- DRAM Controller
- NAND Flash
- Results
- User Functions
  - f#1
  - foo()
  - f#2
  - goo()
- Summarizer
- Task Controller
- TQ
- New NVMe command
- NVMe Host Driver
- Finalization (foo)
Summarizer API and NVMe commands

### Initialization
- NVMe command: `INIT_TSKn`
- Transfer a *in-SSD procedure* to SSD memory
- Initialize data structure and temporal variables for in-SSD computation

### Computation
- NVMe command: `READ_PROC_TSKn`
- Page read command is issued with the *flag* indicating the *user procedure* embedded in SSD memory
- Return the special code if the requested page is processed in SSD
- Page data is transferred to the host if the requested page is *NOT* computed in SSD

### Finalization
- NVMe command: `FINAL_TSKn`
- Gather final in-SSD computation results and transfer to the host
Evaluation Platform

- LS2085a intelligent SSD development platform
- ARM cores running FTL and **Summarizer** firmware
- FPGA implementing NAND flash controller
- PCIe Gen. 3 4x lanes for host communication
Evaluation - Performance

TPC-H Query 6

<table>
<thead>
<tr>
<th>Static</th>
<th>Dynamic</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

- **SDD time**
- **Host time**

*Static workload offloading*
Evaluation - Performance

**TPC-H Query 6**

- **SDD time**
- **Host time**

**Static**
- CPU only processing (baseline)
- SSD only processing

**Dynamic**
- SSD only processing
Evaluation - Performance

![Graph showing TPC-H Query6 performance](chart.png)

- **SDD time**
- **Host time**

### TPC-H Query6

**Static**
- 0
- 0.2
- 0.4
- 0.6
- 0.8
- 1

**Dynamic Offloading**

- **Summarizer Dynamic Offloading**

---

**Legend**
- **Dynamic**
- **Static**
Evaluation - Performance

TPC-H Query6

- SDD time
- Host time

SSD processing + transfer time (internal + external + In-SSD processing)

Host CPU processing time

Static

Dynamic

0 0.2 0.4 0.6 0.8 1
Evaluation - Performance

TPC-H Query 6

Execution time normalized to baseline (CPU only)
Evaluation - Performance

TPC-H Query6

Execution time (normalized to baseline)

<table>
<thead>
<tr>
<th>Static</th>
<th>Dynamic</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0.2</td>
<td>0.8</td>
</tr>
<tr>
<td>0.4</td>
<td>0.7</td>
</tr>
<tr>
<td>0.6</td>
<td>0.9</td>
</tr>
<tr>
<td>0.8</td>
<td>1.1</td>
</tr>
<tr>
<td>1</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Legend:
- SDD time
- Host time
Evaluation - Performance

Execution time (normalized to baseline)

- **CPU only**
  - SDD time: 0.70
  - Host time: 0.30

- **Dynamic**
  - SDD time: 0.60
  - Host time: 0.24

Chart Title: TPC-H Query6
Evaluation - Performance

Performance improved by 14%

W/O NDP

Data computation @ host

With NDP

Data computation @ storage

Data transfer from storage

CPU only

Static

Dynamic

SDD time

Host time

Data computation @ host

Dynamic

Dynamic

W/O NDP

With NDP
Evaluation - Performance

TPC-H Query 6

Performance degraded by static NDP
Evaluation - Performance

TPC-H Query 6
- SDD time
- Host time
- Execution time (normalized to baseline)

TPC-H Query 14
- SDD time
- Host time
- Execution time (normalized to baseline)

String Similarity Join
- SDD time
- Host time
- Execution time (normalized to baseline)

16% 10% 20% 7%
Commercial SSD maintains internal bandwidth ≈ external bandwidth
Design Exploration – Higher Internal Bandwidth

Higher internal bandwidth without increasing external bandwidth

Data transfer bottleneck

Storage interface
Design Exploration – Higher Internal Bandwidth

External : Internal bandwidth ratio

TPC-H Query 6
TPC-H Query 1
TPC-H Query 14
String Similarity Join
Average
**Design Exploration – Higher Internal Bandwidth**

*Summarizer is effective if an SSD platform has higher internal bandwidth*

- **TPC-H Query 6**
- **TPC-H Query 1**
- **TPC-H Query 14**
- **String Similarity Join**
- **Average**
Design Exploration – Better SSD Processor

Better embedded processor is cost effective
Design Exploration – Higher Internal Bandwidth

Embedded processor performance
Summarizer is a cost effective NDP solution with powerful storage processors.
**Dynamic NDP framework for SSDs**
- Opportunistically enables in-SSD processing
- Page-level NDP control
- Automatic workload partitioning

**Summarizer programming model**
- Evaluation on the real development platform
- Explored design space for future SSDs
Thank you

Summarizer:
Trading Communication with Computing Near Storage

Gunjae Koo, Kiran Kumar Matam, Te I, H. V. Krishna Giri Nara, Jing Li, Hung-Wei Tseng, Steven Swanson, Murali Annavaram

(We thank to Dell EMC for supporting the SSD development board)