Filesystem Design Description Document

Overview

This document provides a comprehensive design and implementation guide for the NAND flash-based Filesystem, ensuring reliability, robustness, and scalability. The system leverages a journaling mechanism, metadata management, ECC for error correction, and non-contiguous block allocation to handle the challenges of flash memory, such as power failures and bad blocks.

---

Design Goals

1. Reliability: Data integrity and recovery mechanisms against power failures.

2. Efficiency: Optimal storage utilization and seamless handling of bad blocks.

3. Scalability: Support for large files and numerous files.

4. Error Correction: BCH-based ECC for correcting up to 8-bit errors per page.

---

Key Features

1. Non-Contiguous Block Allocation

- Files span non-adjacent pages to bypass bad blocks.

- Each page’s spare area stores metadata:

- File Chunk ID (4 bytes)

- Next Chunk Pointer (4 bytes)

- ECC (16 bytes for BCH-based 8-bit correction)

2. Journaling System

- Logs metadata operations to ensure atomicity and consistency.

- Enables power-failure recovery by replaying or rolling back operations.

- Stored in a reserved block to avoid interference with file data.

3. Metadata Management

- File Allocation Table (FAT): Tracks logical-to-physical mapping and file metadata.

- Page Metadata: Stored in the spare area for in-page linkage and ECC.

4. Power-Failure Recovery

- Scans journal for incomplete operations and applies recovery logic.

- Ensures metadata and file content consistency.

5. BCH-Based ECC

- Implements BCH (Bose-Chaudhuri-Hocquenghem) coding to correct up to 8-bit errors per 512-byte block.

- Ensures data reliability under harsh conditions.

---

Data Structures

File Allocation Table (FAT)

```c

#define MAX\_FILES 64

typedef struct {

uint32\_t file\_id; // Unique ID for the file

uint32\_t first\_chunk\_address; // Address of the first page of the file

uint32\_t file\_size; // Total size of the file in bytes

uint32\_t create\_time; // Timestamp for file creation

uint32\_t close\_time; // Timestamp for file closure

} FileAllocationTable[MAX\_FILES];

```

Page Metadata

```c

typedef struct {

uint32\_t file\_chunk\_id; // ID of the file this page belongs to

uint32\_t next\_chunk; // Physical address of the next page in the file

uint8\_t ecc[16]; // ECC data for error correction

} PageMetadata;

```

Journal Entry

```c

typedef struct {

uint32\_t operation\_id; // Operation type (e.g., create, write, close)

uint32\_t logical\_addr; // File ID or logical address

uint32\_t physical\_addr; // Page address

uint32\_t status; // PENDING or COMMITTED

} journal\_entry\_t;

```

---

Filesystem Operations

1. File Creation

```c

bool create\_file(uint32\_t file\_id, uint32\_t \*start\_address) {

for (int i = 0; i < MAX\_FILES; i++) {

if (FileAllocationTable[i].file\_id == 0) {

FileAllocationTable[i].file\_id = file\_id;

FileAllocationTable[i].first\_chunk\_address = find\_free\_page();

FileAllocationTable[i].file\_size = 0;

FileAllocationTable[i].create\_time = get\_current\_time();

journal\_entry\_t journal = {

.operation\_id = OPERATION\_FILE\_CREATE,

.logical\_addr = file\_id,

.physical\_addr = FileAllocationTable[i].first\_chunk\_address,

.status = PENDING

};

set\_journal\_entry(&journal);

\*start\_address = FileAllocationTable[i].first\_chunk\_address;

return true;

}

}

return false;

}

```

2. File Write

```c

bool write\_file(uint32\_t file\_id, uint8\_t \*data, uint32\_t size) {

for (int i = 0; i < MAX\_FILES; i++) {

if (FileAllocationTable[i].file\_id == file\_id) {

uint32\_t current\_page = FileAllocationTable[i].first\_chunk\_address;

while (has\_next\_chunk(current\_page)) {

current\_page = get\_next\_chunk(current\_page);

}

if (page\_is\_full(current\_page)) {

uint32\_t new\_page = find\_free\_page();

set\_next\_chunk(current\_page, new\_page);

current\_page = new\_page;

}

write\_to\_flash(current\_page, data, size);

FileAllocationTable[i].file\_size += size;

journal\_entry\_t journal = {

.operation\_id = OPERATION\_FILE\_WRITE,

.logical\_addr = file\_id,

.physical\_addr = current\_page,

.status = PENDING

};

set\_journal\_entry(&journal);

return true;

}

}

return false;

}

```

3. File Closure

```c

bool close\_file(uint32\_t file\_id) {

for (int i = 0; i < MAX\_FILES; i++) {

if (FileAllocationTable[i].file\_id == file\_id) {

FileAllocationTable[i].close\_time = get\_current\_time();

journal\_entry\_t journal = {

.operation\_id = OPERATION\_FILE\_CLOSE,

.logical\_addr = file\_id,

.status = PENDING

};

set\_journal\_entry(&journal);

commit\_fat\_to\_flash();

return true;

}

}

return false;

}

```

4. Chunk Manipulation Functions

```c

bool page\_is\_full(uint32\_t page\_address) {

return check\_page\_usage(page\_address) >= PAGE\_SIZE;

}

bool has\_next\_chunk(uint32\_t page\_address) {

PageMetadata meta;

read\_page\_metadata(page\_address, &meta);

return meta.next\_chunk != INVALID\_PAGE;

}

uint32\_t get\_next\_chunk(uint32\_t page\_address) {

PageMetadata meta;

read\_page\_metadata(page\_address, &meta);

return meta.next\_chunk;

}

void set\_next\_chunk(uint32\_t current\_page, uint32\_t next\_page) {

PageMetadata meta;

read\_page\_metadata(current\_page, &meta);

meta.next\_chunk = next\_page;

write\_page\_metadata(current\_page, &meta);

}

```

5. File Read

```c

bool read\_file(uint32\_t file\_id, uint8\_t \*buffer, uint32\_t size) {

for (int i = 0; i < MAX\_FILES; i++) {

if (FileAllocationTable[i].file\_id == file\_id) {

uint32\_t current\_page = FileAllocationTable[i].first\_chunk\_address;

uint32\_t read\_size = 0;

while (current\_page != INVALID\_PAGE && read\_size < size) {

uint32\_t chunk\_size = min(PAGE\_SIZE, size - read\_size);

read\_from\_flash(current\_page, buffer + read\_size, chunk\_size);

if (!verify\_ecc(current\_page)) {

if (!correct\_ecc(current\_page, buffer + read\_size, chunk\_size)) {

return false;

}

}

read\_size += chunk\_size;

current\_page = get\_next\_chunk(current\_page);

}

return read\_size == size;

}

}

return false;

}

```

---

Journaling and Power-Failure Recovery

Set Journal Entry

```c

void set\_journal\_entry(journal\_entry\_t \*entry) {

uint32\_t journal\_page = find\_free\_journal\_page();

write\_to\_flash(journal\_page, (uint8\_t \*)entry, sizeof(journal\_entry\_t));

}

```

Power Recovery

```c

void recover\_filesystem() {

journal\_entry\_t entry;

for (uint32\_t i = 0; i < JOURNAL\_ENTRY\_COUNT; i++) {

if (retrieve\_journal\_entry(i, &entry)) {

if (entry.status != JOURNAL\_ENTRY\_COMMITTED) {

switch (entry.operation\_id) {

case OPERATION\_FILE\_CREATE:

rollback\_file\_creation(entry.logical\_addr);

break;

case OPERATION\_FILE\_WRITE:

rollback\_file\_write(entry.logical\_addr, entry.physical\_addr);

break;

case OPERATION\_FILE\_CLOSE:

close

# Filesystem Design Description Document

Overview

This document provides a comprehensive design and implementation guide for the NAND flash-based Filesystem, ensuring reliability, robustness, and scalability. The system leverages a journaling mechanism, metadata management, ECC for error correction, and non-contiguous block allocation to handle the challenges of flash memory, such as power failures and bad blocks.

---

Design Goals

1. Reliability: Data integrity and recovery mechanisms against power failures.

2. Efficiency: Optimal storage utilization and seamless handling of bad blocks.

3. Scalability: Support for large files and numerous files.

4. Error Correction: BCH-based ECC for correcting up to 8-bit errors per page.

---

Key Features

1. Non-Contiguous Block Allocation

- Files span non-adjacent pages to bypass bad blocks.

- Each page’s spare area stores metadata:

- File Chunk ID (4 bytes)

- Next Chunk Pointer (4 bytes)

- ECC (16 bytes for BCH-based 8-bit correction)

2. Journaling System

- Logs metadata operations to ensure atomicity and consistency.

- Enables power-failure recovery by replaying or rolling back operations.

- Stored in a reserved block to avoid interference with file data.

3. Metadata Management

- File Allocation Table (FAT): Tracks logical-to-physical mapping and file metadata.

- Page Metadata: Stored in the spare area for in-page linkage and ECC.

4. Power-Failure Recovery

- Scans journal for incomplete operations and applies recovery logic.

- Ensures metadata and file content consistency.

---

Data Structures

File Allocation Table (FAT)

```c

#define MAX\_FILES 64

typedef struct {

uint32\_t file\_id; // Unique ID for the file

uint32\_t first\_chunk\_address; // Address of the first page of the file

uint32\_t file\_size; // Total size of the file in bytes

uint32\_t create\_time; // Timestamp for file creation

uint32\_t close\_time; // Timestamp for file closure

} FileAllocationTable[MAX\_FILES];

```

Page Metadata

```c

typedef struct {

uint32\_t file\_chunk\_id; // ID of the file this page belongs to

uint32\_t next\_chunk; // Physical address of the next page in the file

uint8\_t ecc[16]; // ECC data for error correction

} PageMetadata;

```

Journal Entry

```c

typedef struct {

uint32\_t operation\_id; // Operation type (e.g., create, write, close)

uint32\_t logical\_addr; // File ID or logical address

uint32\_t physical\_addr; // Page address

uint32\_t status; // PENDING or COMMITTED

} journal\_entry\_t;

```

---

Filesystem Operations

1. File Creation

```c

bool create\_file(uint32\_t file\_id, uint32\_t \*start\_address) {

for (int i = 0; i < MAX\_FILES; i++) {

if (FileAllocationTable[i].file\_id == 0) {

FileAllocationTable[i].file\_id = file\_id;

FileAllocationTable[i].first\_chunk\_address = find\_free\_page();

FileAllocationTable[i].file\_size = 0;

FileAllocationTable[i].create\_time = get\_current\_time();

journal\_entry\_t journal = {

.operation\_id = OPERATION\_FILE\_CREATE,

.logical\_addr = file\_id,

.physical\_addr = FileAllocationTable[i].first\_chunk\_address,

.status = PENDING

};

set\_journal\_entry(&journal);

\*start\_address = FileAllocationTable[i].first\_chunk\_address;

return true;

}

}

return false;

}

```

2. File Write

```c

bool write\_file(uint32\_t file\_id, uint8\_t \*data, uint32\_t size) {

for (int i = 0; i < MAX\_FILES; i++) {

if (FileAllocationTable[i].file\_id == file\_id) {

uint32\_t current\_page = FileAllocationTable[i].first\_chunk\_address;

while (has\_next\_chunk(current\_page)) {

current\_page = get\_next\_chunk(current\_page);

}

if (page\_is\_full(current\_page)) {

uint32\_t new\_page = find\_free\_page();

set\_next\_chunk(current\_page, new\_page);

current\_page = new\_page;

}

write\_to\_flash(current\_page, data, size);

FileAllocationTable[i].file\_size += size;

journal\_entry\_t journal = {

.operation\_id = OPERATION\_FILE\_WRITE,

.logical\_addr = file\_id,

.physical\_addr = current\_page,

.status = PENDING

};

set\_journal\_entry(&journal);

return true;

}

}

return false;

}

```

3. File Closure

```c

bool close\_file(uint32\_t file\_id) {

for (int i = 0; i < MAX\_FILES; i++) {

if (FileAllocationTable[i].file\_id == file\_id) {

FileAllocationTable[i].close\_time = get\_current\_time();

journal\_entry\_t journal = {

.operation\_id = OPERATION\_FILE\_CLOSE,

.logical\_addr = file\_id,

.status = PENDING

};

set\_journal\_entry(&journal);

commit\_fat\_to\_flash();

return true;

}

}

return false;

}

```

4. Chunk Manipulation Functions

```c

bool page\_is\_full(uint32\_t page\_address) {

return check\_page\_usage(page\_address) >= PAGE\_SIZE;

}

bool has\_next\_chunk(uint32\_t page\_address) {

PageMetadata meta;

read\_page\_metadata(page\_address, &meta);

return meta.next\_chunk != INVALID\_PAGE;

}

uint32\_t get\_next\_chunk(uint32\_t page\_address) {

PageMetadata meta;

read\_page\_metadata(page\_address, &meta);

return meta.next\_chunk;

}

void set\_next\_chunk(uint32\_t current\_page, uint32\_t next\_page) {

PageMetadata meta;

read\_page\_metadata(current\_page, &meta);

meta.next\_chunk = next\_page;

write\_page\_metadata(current\_page, &meta);

}

```

5. File Read

```c

bool read\_file(uint32\_t file\_id, uint8\_t \*buffer, uint32\_t size) {

for (int i = 0; i < MAX\_FILES; i++) {

if (FileAllocationTable[i].file\_id == file\_id) {

uint32\_t current\_page = FileAllocationTable[i].first\_chunk\_address;

uint32\_t read\_size = 0;

while (current\_page != INVALID\_PAGE && read\_size < size) {

uint32\_t chunk\_size = min(PAGE\_SIZE, size - read\_size);

read\_from\_flash(current\_page, buffer + read\_size, chunk\_size);

if (!verify\_ecc(current\_page)) {

if (!correct\_ecc(current\_page, buffer + read\_size, chunk\_size)) {

return false;

}

}

read\_size += chunk\_size;

current\_page = get\_next\_chunk(current\_page);

}

return read\_size == size;

}

}

return false;

}

```

---

Journaling and Power-Failure Recovery

Set Journal Entry

```c

void set\_journal\_entry(journal\_entry\_t \*entry) {

uint32\_t journal\_page = find\_free\_journal\_page();

write\_to\_flash(journal\_page, (uint8\_t \*)entry, sizeof(journal\_entry\_t));

}

```

Power Recovery

```c

void recover\_filesystem() {

journal\_entry\_t entry;

for (uint32\_t i = 0; i < JOURNAL\_ENTRY\_COUNT; i++) {

if (retrieve\_journal\_entry(i, &entry)) {

if (entry.status != JOURNAL\_ENTRY\_COMMITTED) {

switch (entry.operation\_id) {

case OPERATION\_FILE\_CREATE:

rollback\_file\_creation(entry.logical\_addr);

break;

case OPERATION\_FILE\_WRITE:

rollback\_file\_write(entry.logical\_addr, entry.physical\_addr);

break;

case OPERATION\_FILE\_CLOSE:

close\_file(entry.logical\_addr);

break;

default:

break;

}

}

}

}

}

```

---

BCH Error Correction

The BCH-based ECC algorithm