THEORY OF OPERATION SECTION

Contents

1. RAID Architecture Overview

The objectives of the RAID technology are the low cost, high reliability, and high I/O performance of disk storage devices. To achieve these objectives, this subsystem supports levels 1, and 5 of RAID technologies (in this section, part of level 3 RAID technology is explained to make the outline of RAID5 more understandable). The features of the levels of RAID technologies are described below.

1.1 RAID0 and RAID1

RAID0 uses small-scale disk storage devices instead of conventional expensive large-scale disk storage devices for cost reduction. To increase the reliability of RAID0 devices, all disk devices are duplexed in RAID1.
Strictly speaking, RAID0 and RAID1 subsystems are not disk array systems; their performance is the same as conventional disk subsystems. Controlling these subsystems are rather simple and many RAID0 and RAID1 subsystems have been put into market by several vendors.

1.2 RAID3

In RAID3, a stream of data to be transferred is split and distributed into two or more disk devices (1 parity group) on a byte or bit basis (striping). This enables two or more disk devices to run simultaneously and increase the speed of data transfer between the DKC and disk drives. In addition, since parity data for the parity group is created and stored on a separate disk device (parity disk), data can easily be recovered even if a device in a parity group gets inoperative or causes a read error. This enhances the reliability of the disk storage subsystem.
Since RAID3 lacks the parallel I/O capability and it entails a long latency time because of the need to drive two or more drives at a time, it exhibits little performance in applications that process small-volume data repeatedly (transaction processing) though it shows a real advantage when running applications that process large-volume data in a single run (e.g., scientific computations).

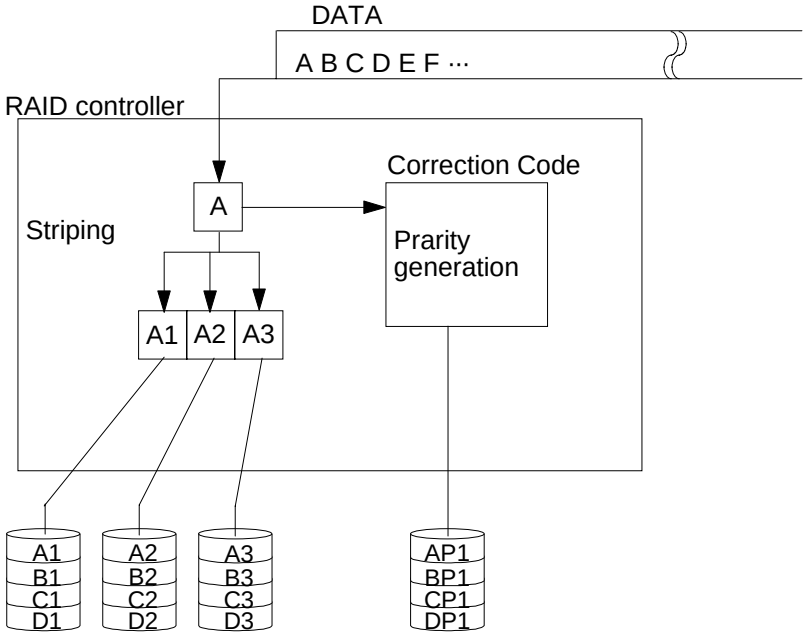


Fig. 1.2 Outline of a RAID3 (3D + 1P) System

1.3 RAID5

Small-sized records are intensively read and written randomly in transaction processing. This type of processing generates many I/O requests for transferring small amounts of data. In such a situation, greater importance is placed on increased I/O performance (parallel I/O processing) than on increase in the rate of transferring large-volume data. RAID5 has been introduced to be suitable for this type of transaction processing.

In RAID5, the striping size is set to that of blocks which are transferred in a small-scale I/O processing mode and which are distributed in two or more disk devices (1 parity group). This entails the RAID controller to access each disk device only for one stripe equivalence of data and allows it to perform I/O operations on other disks in parallel, increasing the I/O performance substantially; though its data transfer rate remains the same as that of conventional subsystems in small-scale I/O applications. In large-scale (sequential) I/O applications, it permits the blocks in the same parity group to be processed in parallel as does RAID3, resulting in an increase in data transfer rate. In addition, like RAID3, RAID5 uses a parity disk for improved reliability. In individual writes to single-blocks (small-scale writes), however, this parity scheme raises various problems for the following reasons:

- The parity disk entails rewrites of new parity data.
- It is necessary to read the old parity data and old (before-update data) to generate new parity data.

Consequently, RAID5 entails extra reads from the data and parity disks in small-size write operations (this is called “write penalty”). If the parity data were fixed at a single disk device, the parity disk would be occupied during a single write that is executed to update the parity data, thus making it impossible to perform parallel I/O processing. To alleviate this problem, RAID5 adopts a system of distributing parity data on several disks in the group. This will not solve the write penalty problem completely.

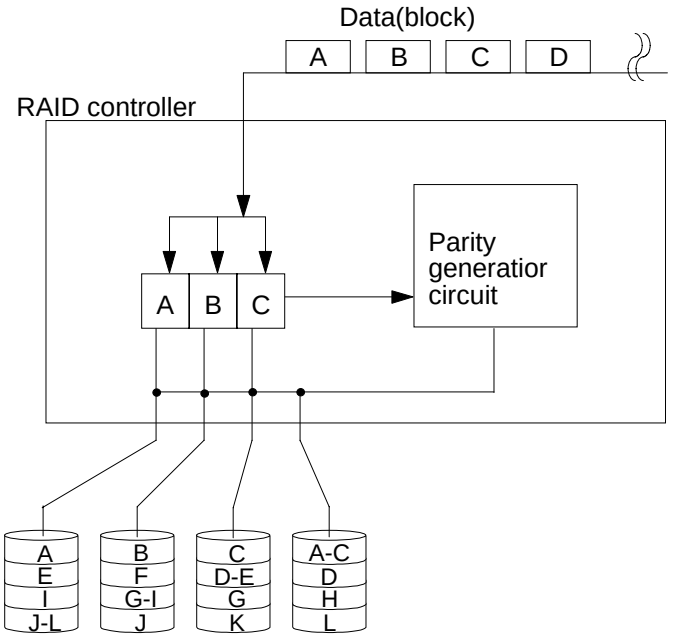


Fig. 1.3 Outline of RAID5 (3D + 1P)

1.4 Application of the RAID technologies DKC410 and DKU405

RAID5 will not show its stuff during transaction processing if a single I/O operation spans over two or more disk devices (when the stripe size is too small). Consequently, adequate consideration should be given to the stripe size.

In this subsystem, the stripe size is 58 KB which is equivalent to one track of the 3390-3 so that the 3390-3 can be emulated. This is because it is anticipated that a single transaction will not span over two or more tracks when the 3390-3 is emulated. This subsystem also uses cache memory to preclude write penalties from occurring wherever possible. The cache memory can pool the data to be written onto the disk drives and prefetch old data.

- On write penalty

A parity group of a this subsystem (level 5) system consists of four disk devices (3D+1P). Since the parity data is made up of parity data for three data disks in the group, once a 1 stripe equivalence of partial write occurs in the group during transaction processing, it becomes necessary to regenerate the corresponding parity data in that group.

Since parity data is calculated using the formula shown below, “data established before update,” “parity established before update,” and “data established after update” are required to generate parity. The extra processing required to read this “data before update” is referred to as a write penalty.

“New parity data”
= (“Data before update” EOR “Data after update”) EOR “Parity before update”

2. Specifications

2.1 Subsystem Specifications

Subsystem specifications are shown in the following table.

Table 2.1 Subsystem specifications

Item			Specifications	
			Separate Model	Single Cabinet Model
Subsystem	Maximum number of disk unit/Subsystem		6	1
	Maximum number of disk drives	Disk unit	96	48
		Subsystem	512	48
	RAID level		RAID5/RAID1	←
	RAID group		RAID5:3D+1P RAID1:2D+2D	←
	Maximum number of RAID group		126 (Spare disk:8) 124 (Spare disk:16)	11 (Spare disk:4)
	Maximum number of spare disk drives		max.16	4
	Maximum number of volumes		4,096	←
	Support emulation type	Mainframe	3390-1/2/3/9/L,3380-J/K/E	←
		Open system	OPEN-3/8/9/K	←
Controller	Cache memory capacity		512MB to 32GB	512MB to 16GB
	Shared memory capacity		512MB to 1,536MB	←
	Internal path architecture		Hierarchical Star Net(HSN)	←
	Internal path transfer rate		max. 3.2GB/s	←
Device I/F	DKC-DKU interface		Fibre(FC-AL) / Dual Port	←
	Data transfer rate (MB/S)		100MB/S	←
	Number of HDD/FC-AL		8 to 32	8
	Number of DKA		2 to 8	2
	Number of array control processor		8 to 32	8
Channel I/F	Support channel option	Mainframe	Serial channel:4S/8S Fibre Short Wavelength: 4MS Fibre Long Wavelength: 4ML	←
		Open system	Fibre Short Wavelength:4GS/8GS Fibre Long Wavelength:4GL/8GL	←
	Data transfer rate (MB/s)	Serial channel	10/17	←
		Fibre channel	100	←
	Number of CHA options		1 to 4	3
	Number of channel ports	4S/4GS/4GL/4MS/4ML	4/8/12/16	4/8/12
		8S/8GS/8GL	8/16/24/32	8/16/24
Non stop maintenance	DKC	Control PCB	✓	←
		CM/SM memory module	✓	←
		Power supply,fan	✓	←
		Battery	✓	←
		Microcode	✓	←
	DKU	Disk drive	✓	←
		Power supply,fan	✓	←
		Control PCB	✓	←

2.2 Disk Drive Specifications

Disk Drive specifications are shown in the following table.

Table 2.2 Disk Drive Specifications

Item		DKR2B-J18FC DKR2C-J18FC *1 DKR2D-J18FC *4 DKR2D-J18FD *5 DKR2E-J18FD *7 DKR2F-J18FD *10	DKR1B-J47FC DKR1C-J47FC *2 DKR2D-J47FD *6 DKR2E-J47FD *8 DKR2F-J47FD *11	DKR1C-J72FC DKR2D-J72FC *3 DKR2E-J72FC *9 DKR2F-J72FC *12	DKR2E-J146FC DKR2F-J146FC *13
Formatted capacity (User area)		18.46GB	47.19GB	72.91GB	143.7GB
Diameter of disk		3 inch	3 inch	3 inch	3 inch
Number of heads		9/6*1/3*4/5*5/5*7/3*10	23/16*2/10*6/5*8/3*11	24/10*3/5*9/3*12	10/5*13
Number of disks		5/3*1/2*4/3*5/3*7/2*10	12/8*2/5*6/3*8/2*11	12/5*3/3*9/2*12	5/3*13
Revolution speed (min^{-1})		10,025	10,025	10,025	10,025
Seek time (ms) (Read/Write)	MIN.	0.5 / 0.7	0.5 / 0.7	0.5 / 0.7	0.5 / 0.7
	AVE.	5.2 / 6.0	5.7 / 6.5	5.7 / 6.5	4.9 / 5.4
	MAX.	12.0 / 13.0	12.0 / 13.0	12.0 / 13.0	10.0 / 11.0
Average latency time (ms)		2.99	2.99	2.99	2.99
Internal data transfer rate (MB/s)		30.2 to 45.6	30.2 to 45.6	33.6 to 56.6	57.3 to 99.9
Interface data transfer rate (MB/s)		Max.100	Max.100	Max.100	Max.100

*1: DKR2C-J18FC *2: DKR1C-J47FC *3: DKR2D-J72FC *13: DKR2F-J146FC
*4: DKR2D-J18FC *6: DKR2D-J47FD *9: DKR2E-J72FC
*5: DKR2D-J18FD *8: DKR2E-J47FD *12: DKR2F-J72FC
*7: DKR2E-J18FD *11: DKR2F-J47FD
*10: DKR2F-J18FD

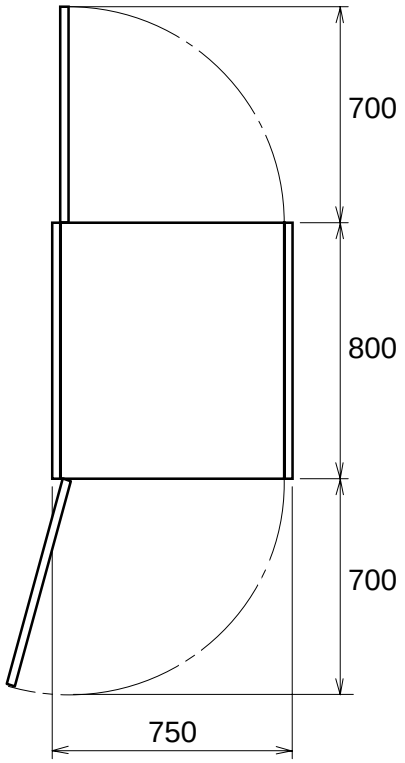
Item		DKS2A-K18FC DKS2B-K18FC *15	DKS1A-H180FC	DKS2B-K36FC DKS2C-K36FC *14	DKS2C-K72FC
Formatted capacity (User area)		17.87GB	172.55GB	35.76GB	71.50GB
Diameter of disk		2.5 inch	3 inch	2.5 inch	2.5 inch
Number of heads		10 / 4 *15	24	8 / 4 *14	8
Number of disks		5 / 2 *15	12	4 / 2 *14	4
Revolution speed (min^{-1})		14,904	7,200	14,904	14,904
Seek time (ms) (Read/Write)	MIN.	0.5 / 0.7	0.8 / 1.1	0.4 / 0.8	0.4 / 0.8
	AVE.	3.9 / 4.5	7.4 / 8.2	3.8 / 4.2	3.8 / 4.2
	MAX.	7.5 / 8.0	16.0 /17.0	7.0 / 8.0	6.7 / 7.1
Average latency time (ms)		2.01	4.17	2.01	2.01
Internal data transfer rate (MB/s)		48.1 to 64.0	35.3 to 63.5	68.5 to 88.3	74.5 to 111.4
Interface data transfer rate (MB/s)		Max.100	Max.100	Max.100	Max.100

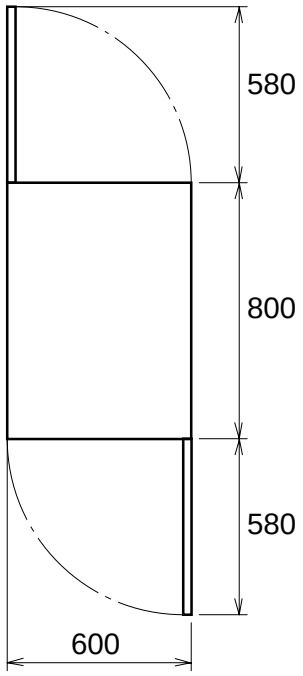
*15: DKS2B-K18FC *14: DKS2C-K36FC

2.3 Physical Specifications

DKC410I/DKU405I physical specifications are shown in the following figures and table.

DKC410I DKU405I (Unit: mm)





Item		DKC410I		DKU405I	
Dimension (mm)	Width	750 *1	750 *1	600	600
	Depth	800	800	800	800
	Height	1,790	1,790	1,790	1,790
Weight (kg)		410 *2	480 *3	440 *4	480 *5
Heat Output (kW)		1.46 *2	2.32 *3	2.85 *4	3.43 *5
Power Consumption (kVA)		1.57 *2	2.41 *3	3.14 *4	3.72 *5
Air Flow (m^3/min.)		18 *2	18 *3	12 *4	12 *5

*1: This includes the thickness of side covers (16 mm × 2).
*2: These values are used when DKC410I has 8 GB Cache Memory, two Fibre 8-port Adapter and a Additional Disk Adapter.
*3: These values are used when DKC410I has full options.
*4: These values are used when DKU405I has full mounted the 18 GB HDD.
*5: These values are used when DKU405I has full mounted the 47 GB or 72 GB HDD.

2.4 Equipment Layout

The overview of DKC410I/DKU405I Disk Subsystem is depicted in Figure 2.4.
DKC410I Disk Controller controls up to six DKU405I units. For details on the individual components,see the Location Sections.

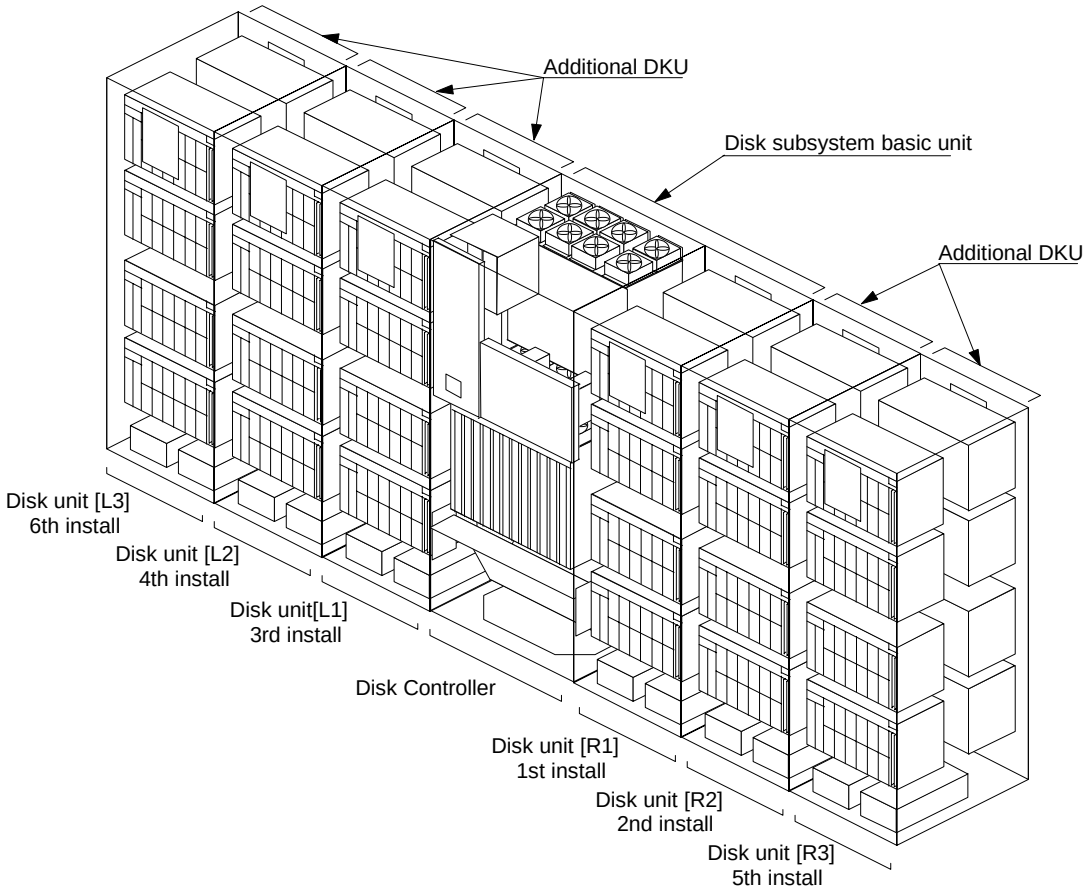


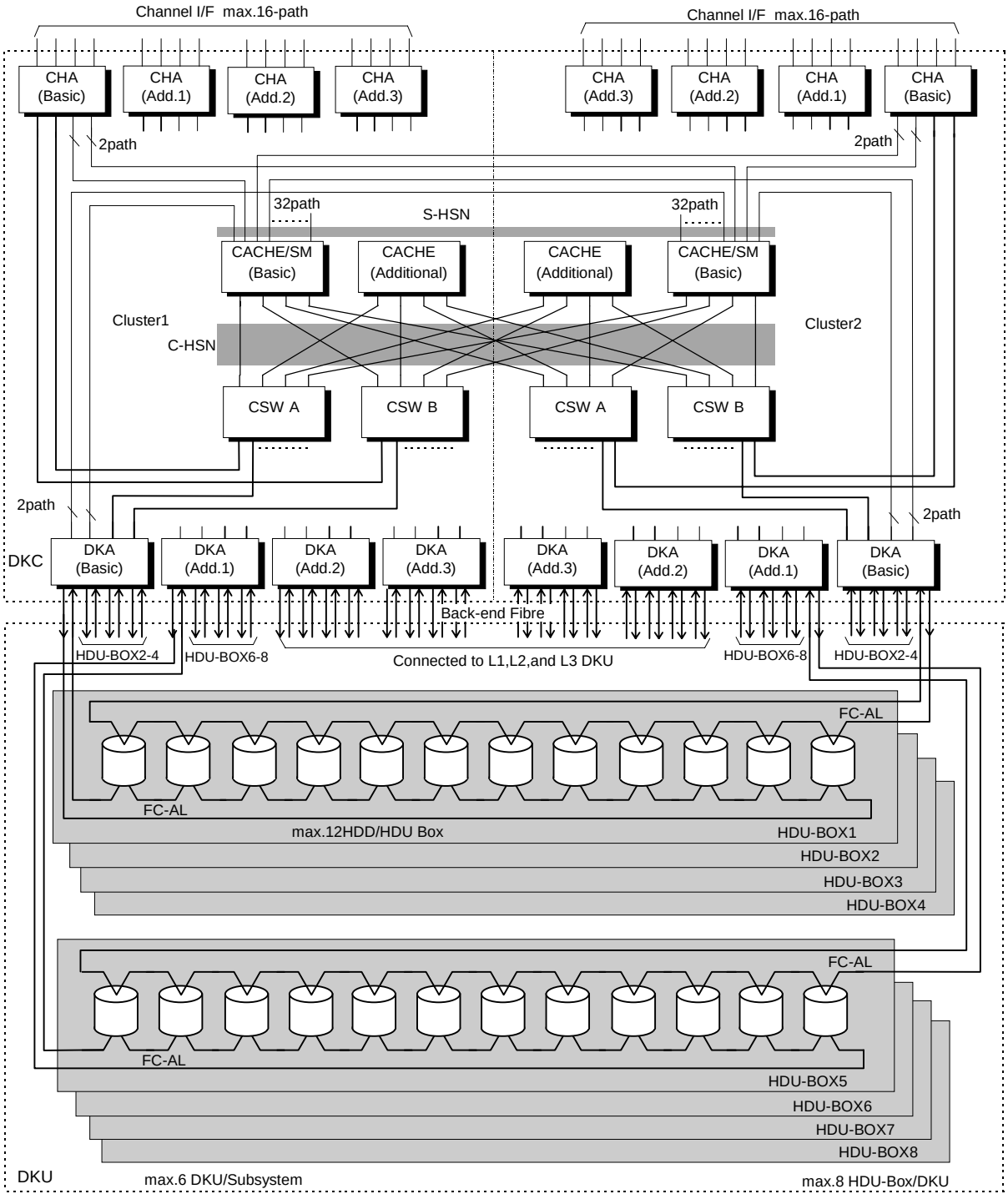
Fig. 2.4 Overview of Disk Subsystem

Addition of the DKU frames to be done in the following order: 1st DKU, 2nd DKU, 3rd DKU, 4th DKU, 5th DKU and 6th DKU. However, when high performance is required or mixture of different types of RAID level is installed, connection of the 3rd and 4th DKUs omitting the 2nd DKU is permitted.

3. Internal Operation

3.1 Hardware Block Diagram

The hardware of the RAID400 disk subsystem is composed of channel adapters, disk adapters, cache memory and disk units.



S-HSN: Shared Memory Hierarchical Star Net
C-HSN: Cache Memory Hierarchical Star Net
CSW: Cache Switch

In the maximum configuration of the subsystem, up to 32 HDDs can be connected through one FC-AL.

3.2 Software Organization

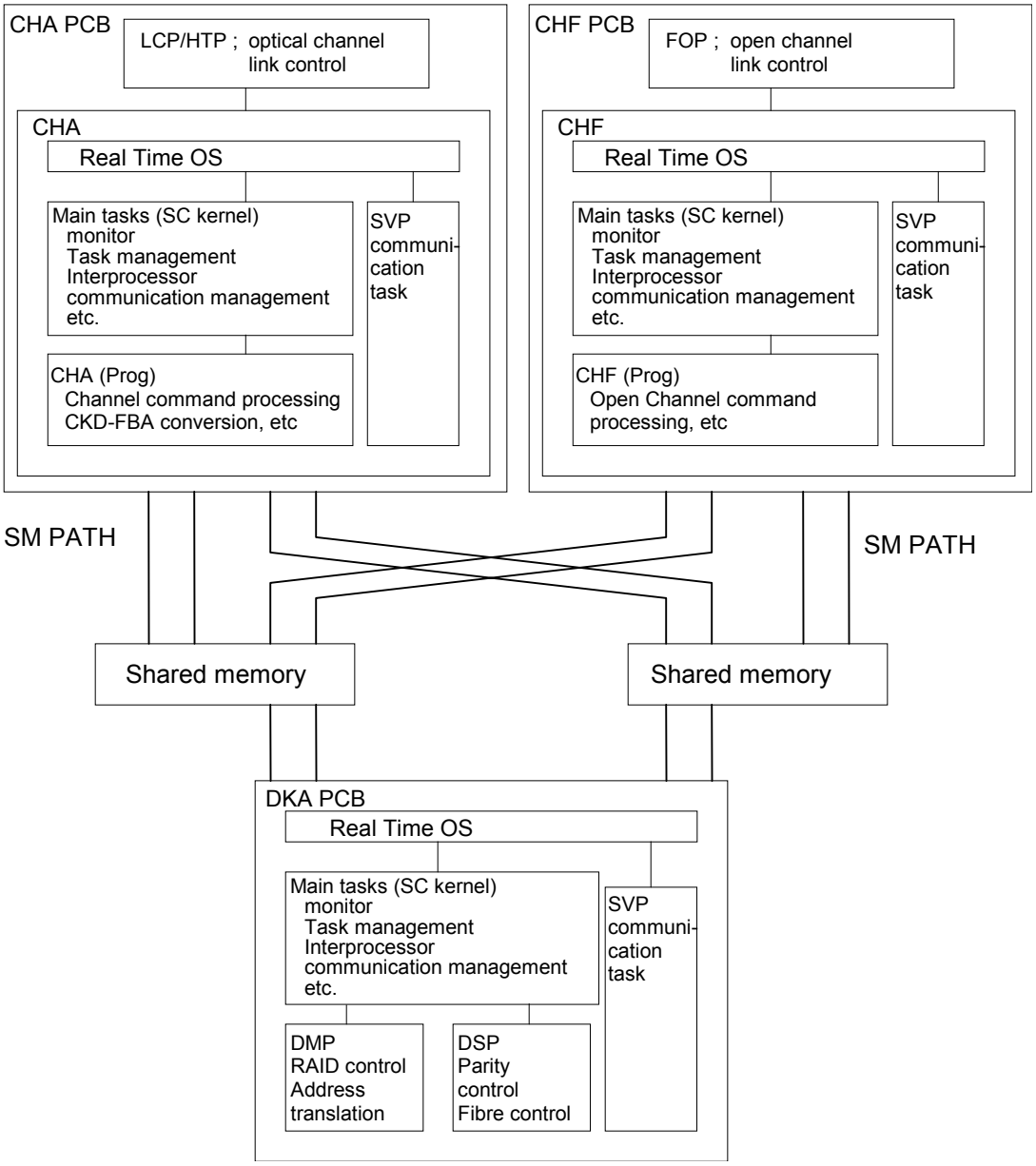


Fig. 3.2 Software Organization

Real Time OS:

A basic OS for controlling the RISC processor. Its primary tasks are to control and switch between the main tasks and SVP communication tasks.

Main tasks:

Made up of DKC control tasks (CHA Prog, CHF Prog, DMP, DSP) and the SC kernel tasks that supervise the DKC control tasks. They switch the control tasks by making use of the SC kernel's task switching facility.

SVP communication task:
Controls the communication with the SVP.

LCP (Link Control Program):
Controls the optical channel links.

HTP (Hyper Transfer Program):
Controls the FICON channel links.

FOP (Fibre Open Program):
Controls the open channel links.

CHA (Prog):
Is a channel command control layer that processes channel commands and controls cache and data transfer operations. It is located in the CHA. CHA Prog is recognized by the logical volume number and logical block number.

CHF (Prog):
Is a open channel command control layer that processes open channel commands and controls cache and data transfer operations. It is located in the CHF. CHF Prog is recognized by the logical volume number and logical block number.

DMP (Disk Master Program):
Is a RAID control layer and provides cache control, logical-to-physical address translation, and RAID control functions. DMP is located in the DKA. DMP is recognized by the logical volume number and logical block number.

DSP (Disk Slave Program):
Is a Fibre drive control layer and provides Fibre control, drive data transfer control, and parity control functions. It is located in the DKA. DSP is recognized by the physical volume number and LBA number.

Shared memory:
Stores the shared information about the subsystem and the cache control information (director names). This type of information is used for the exclusive control of the subsystem. Like CACHE, shared memory is controlled as two areas of memory and fully non-volatile (sustained for approximately 96 hours). The size of shared memory must be 20 MB for 1 GB of cache.

SM PATH (Shared Memory Access Path):
Access Path from the processors of CHA, CHF, DKA, PCB to Shared Memory.

3.3 Data Formats

(1) Data Conversion Overview
Since the disk subsystem uses SCSI drives, data in the CKD format are converted to the FBA format on an interface before being written on the drives. The data format is shown in Fig. 3.3.1.

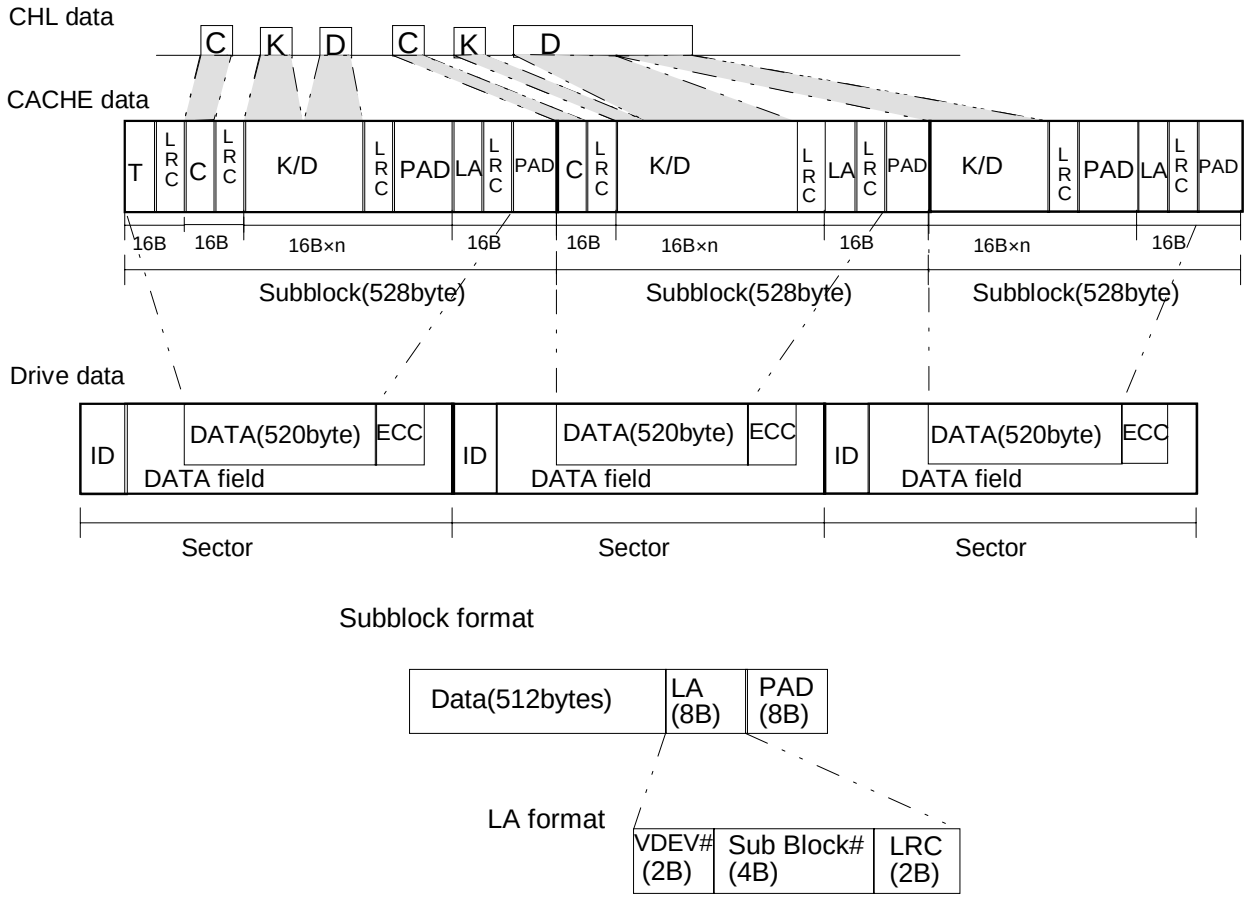
CKD-to-FBA conversion is carried out by the CHA. Data is stored in cache (in the DKC) in the FBA format. Consequently, the drive need not be aware of the data format when transferring to and receiving data from cache.

Each field of the CKD-format record is left-justified and the data is controlled in units of 528-byte subblocks (because data is transferred in 16-byte units). Each field is provided with data integrity code (LRC). An address integrity code (LA: logical address) is appended to the end of each subblock. A count area (C area) is always placed at the beginning of the subblock.

Four subblocks make up a single block. The first subblock of a block is provided with T information (record position information).

If a record proves not to fit in a subblock during CKD-to-FBA conversion, a field is split into the next subblock when it is recorded. If a record does not fill a subblock, the subblock is padded with 00s, from the end of the last field to the LA.

On a physical drive, data is recorded data fields in 520-byte units (physical data format). The format of the LA in the subblock in cache is shown in Fig. 3.3.1. The last 8 bytes of the LA area are padding data which is insignificant (the reason for this is because data is transferred to cache in 16 byte units). When data is transferred to a drive from cache, the last 8 bytes of each LA area are discarded and 520 bytes are transferred.



Subblock format

Fig. 3.3.1 Data Format

(2) Block format

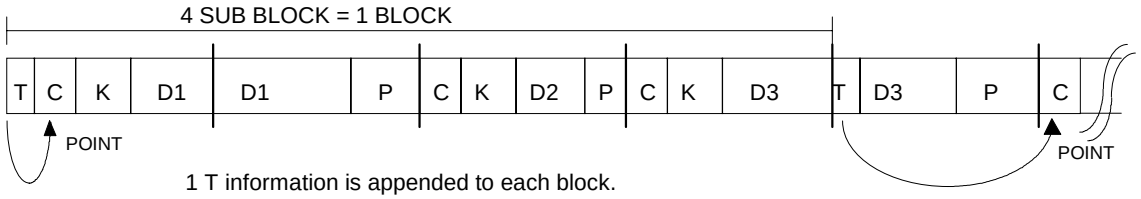


Fig. 3.3.2 Block Format

The RAID system records T information for each block of 4 subblocks as positional information that is used during record search. This unit of data is called a block.

1 block = 4 subblocks = 2 KB

The T information is 16 bytes long. However, only two bytes have meaning and the remaining 14 byte positions are padded with 0s. The reason for this is the same as that for the LA area. Unlike the LA, the insignificant bytes are also stored on the drive as are.

As seen from Fig. 3.3.2, the T information points to the closest count area in its block in the form of an SN (segment number). The drive computes the block number from the sector number with the SET SECT and searches the T information for the target block. From the T information, the drive computes the location of the closest count area and starts processing the block at the count area. This means that the information plays the role of the AM of the conventional disk storage.

(3) Data integrity provided

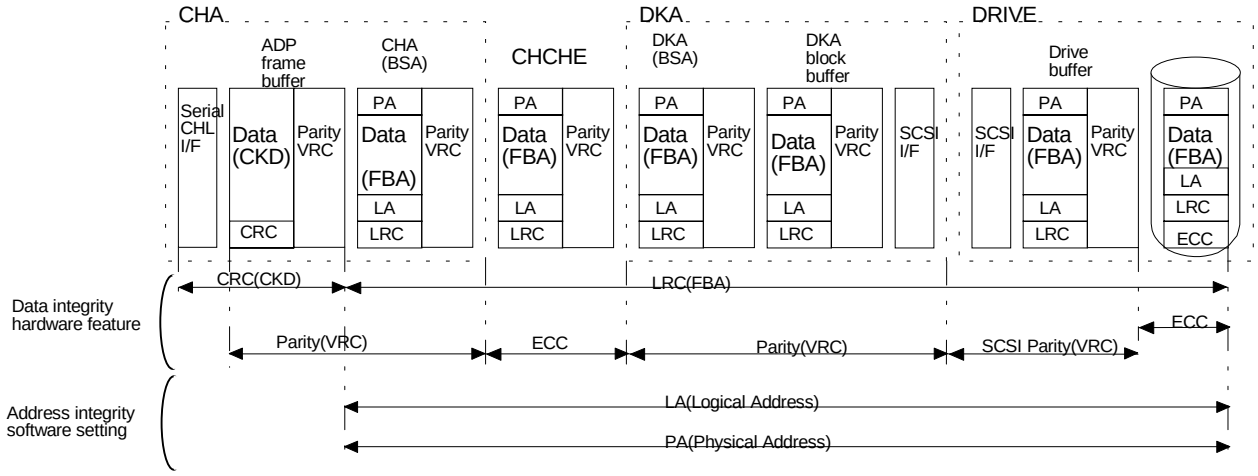


Fig. 3.3.3 Outline of Data Integrity

In the DKC410 and DKU405 system, a data integrity code is appended to the data being transferred at each component as shown in Fig. 3.3.3. Since data is striped onto two or more disk devices and the address integrity code is also appended. The data integrity codes are appended by hardware and the address integrity codes by software.

3.4 Cache Management

Since the DKC requires no through operation, its cache system is implemented by two memory areas called cache A and cache B so that write data can be duplexed. To prevent data loss due to power failures, cache is made non-volatile by being fully battery-backed (48 hours). This dispenses with the need for the conventional NVS.

The minimum unit of cache is the 16 KB segment. Cache is destaged in segment units. Emulation Disk type at three or four segments make up one slot. The read and write slots are always controlled in pair. Cache data is enqueued and dequeued usually in slot units. In real practice, the segments of the same slot are not always stored in a contiguous area in cache, but are stored in discreet areas. These segments are controlled suing CACHE-SLCB and CACHE-SGCB so that the segments belonging to the same slot are seemingly stored in a contiguous area in cache.

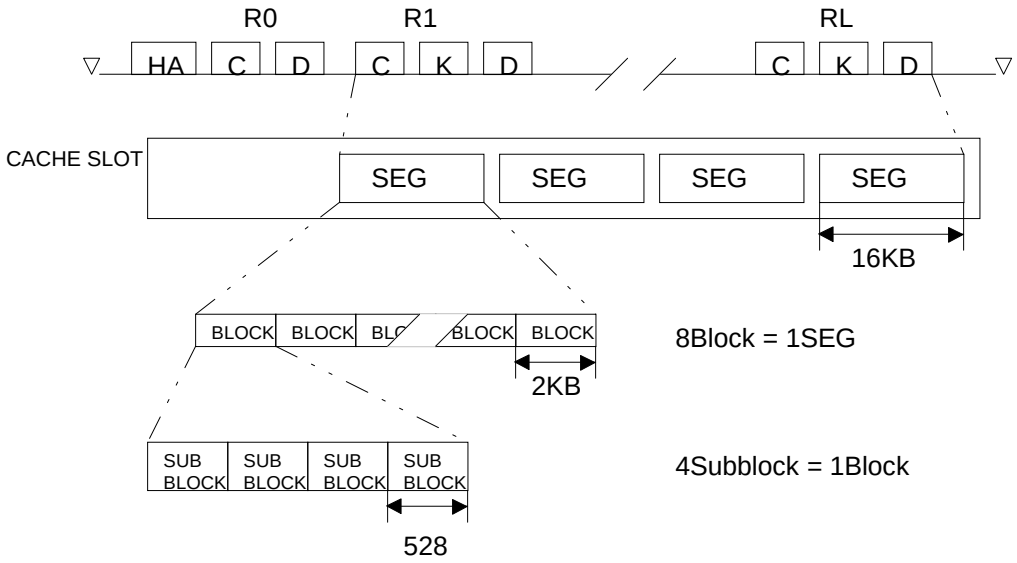


Fig. 3.4 Cache Data Structure

For increased directory search efficiency, a single virtual device (VDEV) is divided into 16-slot groups which are controlled using VDEV-GRPP and CACHE-GRPT.

1 cache segment = 8 blocks = 32 subblocks = 16 KB
1 slot = 1 stripe = 4 segments = 64 KB

The directories VDEV-GRPP, CACHE-GRPT, CACHE-SLCB, and CACHE-SGCB are used to identify the cache hit and miss conditions. These control tables are stored in the shared memory.

In addition to cache hit and miss control, the shared memory is used to classify and control the data in cache according to its attributes. Queues are something like boxes that are used to classify data according to its attributes.

Basically, queues are controlled in slot units (some queues are controlled in segment units). Like SLCB-SGCB, queues are controlled using a queue control table so that queue data of the seemingly same attribute can be controlled as a single data group. These control tables are briefly described below.

(1) Cache control tables (directories)

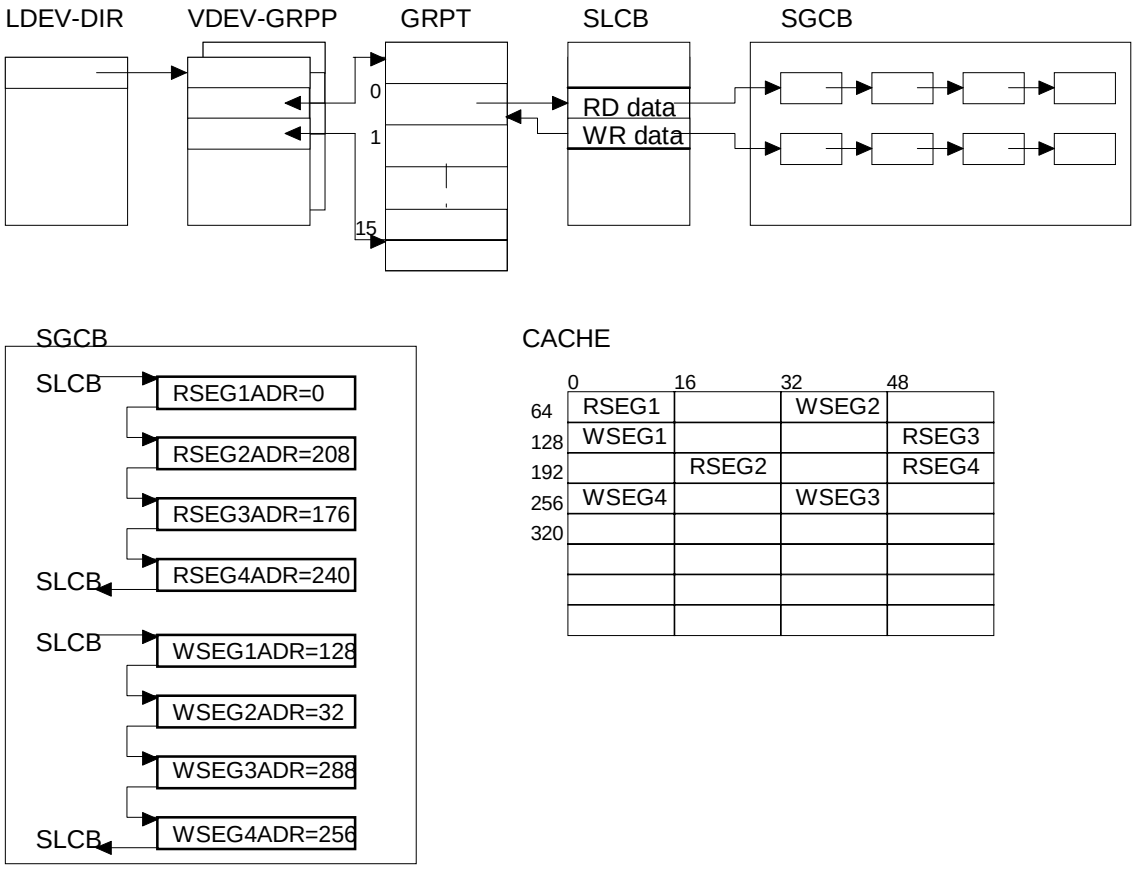


Fig. 3.4.1 Cache Control Tables

LDEV-DIR (Logical DEV-directory):

Contains the shared memory addresses of VDEV-GRPPs for an LDEV. LDEV-DIR is located in the local memory in the CHA.

VDEV-GRPP (Virtual DEV-group Pointer):

Contains the shared memory addresses of the GRPTs associated with the group numbers in the VDEV.

GRPT (Group Table):

A table that contains the shared memory address of the SLCBs for 16 slots in the group. Slots are grouped to facilitate slot search and to reduce the space for the directory area.

SLCB (Slot Control Block):

Contains the shared memory addresses of the starting and ending SGCBs in the slot. One or more SGCBs are chained. The SLCB also stores slot status and points to the queue that is connected to the slot. The state transitions of clean and dirty queues occur in slot units. The processing tasks reserve and release cache areas in this unit.

SGCB (Segment Control Block):

Contains the control information about a cache segment. It contains the cache address of the segment. It is used to control the staged subblock bit map, dirty subblock bitmap, and other information. The state transitions of only free queues occur in segment units.

(2) Cache control table access method (hit/miss identification procedure)

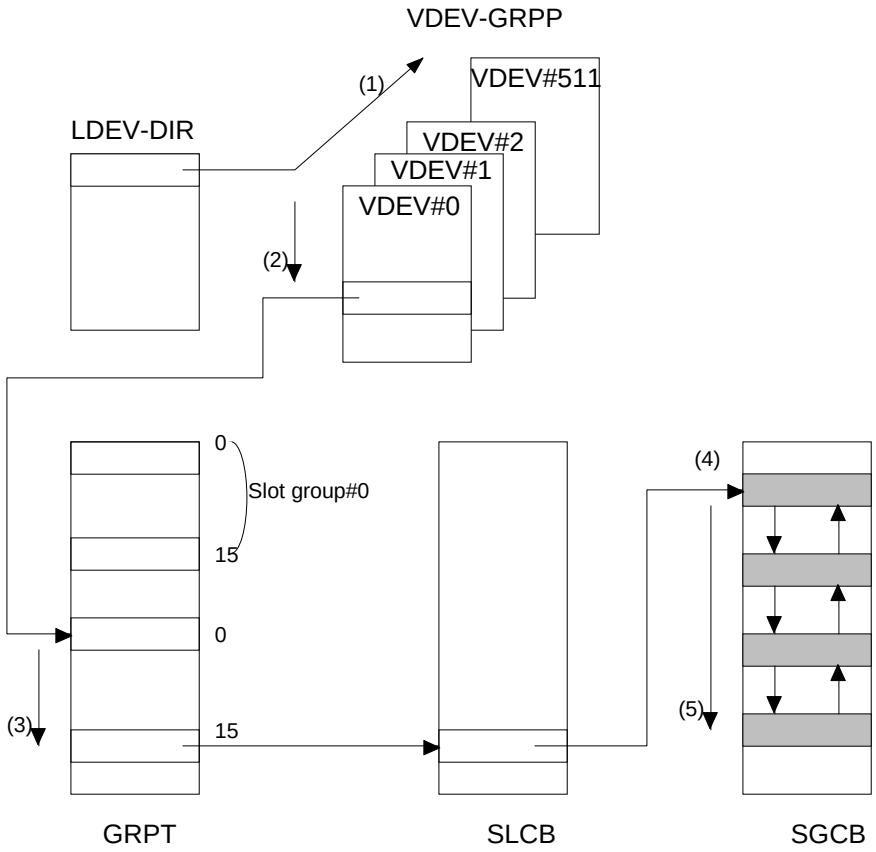


Fig. 3.4.2 Outline of Cache Control Table Access

1. The current VDEV-GRPP is referenced through the LDEV-DIR to determine the hit/miss condition of the VDEV-groups.
2. If a VDEV-group hits, CACHE-GRPT is referenced to determine the hit/miss condition of the slots.
3. If a slot hits, CACHE-SLCB is referenced to determine the hit/miss condition of the segments.
4. If a segment hits, CACHE-SGCB is referenced to access the data in cache.

If a search miss occurs during the searches from 1. through 4., the target data causes a cache miss.

Definition of VDEV number
Since the host processor recognizes addresses only by LDEV, it is unaware of the device address of the parity device. Accordingly, the RAID system is provided with a VDEV address which identifies the parity device associated with an LDEV. Since VDEVs are used to control data devices and parity devices systematically, their address can be computed using the following formulas:
Data VDEV number = LDEV number
Parity VDEV number = 1024 + LDEV number

From the above formulas, the VDEV number ranges from 0 to 2047.

(3) Queue structures

The DKC410 and DKU405 uses 10 types of queues to control data in cache segments according to its attributes. These queues are explained below.

- CACHE-GRPT free queue

This queue is used to control segments that are currently not used by CACHE-GRPT (free segments) on an FIFO (First In First Out) basis. When a new table is added to CACHE-GRPT, the segment that is located by the head pointer of the queue is used.

- CACHE-SLCB free queue

This queue is used to control segments that are currently not used by CACHE-SLCB (free segments) on an FIFO basis. When a new slot is added to CACHE-SLCB, the segment that is located by the head pointer of the queue is used.

- CACHE-SGCB free queue

This queue is used to control segments that are currently not used by CACHE-SGCB (free segments) on an FIFO basis. When a new segment is added to CACHE-SGCB, the segment that is located by the head pointer of the queue is used.

- Clean queue

This queue is used to control the segments that are reflected on the drive on an LRU basis.

- Bind queue

This queue is defined when the bind mode is specified and used to control the segments of the bind attribute on an LRU basis.

- Error queue

This queue controls the segments that are no longer reflected on the drive due to some error (pinned data) on an LRU basis.

- Parity in-creation queue

This queue controls the slots (segments) that are creating parity on an LRU basis.

- DFW queue (host dirty queue)

This queue controls the segments that are not reflected on the drive in the DFW mode on an LRU basis.

- CFW queue (host dirty queue)

This queue controls the segments that are not reflected on the drive in the CFW mode on an LRU basis.

- PDEV queue (physical dirty queue)

This queue controls the data (segments) that are not reflected on the drive and that occur after parity is generated. Data is destaged from this queue onto the physical DEV. There are 32 PDEV queues per physical DEV.

The control table for these queues is located in the shared memory and points to the head and tail segments of the queues.

(4) Queue state transitions

Figure 3.4.4 shows the state transitions of the queues used in. A brief description of the queue state transitions follows.

- State transition from a free queue

When a read miss occurs, the pertinent segment is staged and enqueued to a clean queue. When a write miss occurs, the pertinent segment is temporarily staged and enqueued to a host dirty queue.

- State transition from a clean queue

When a write hit occurs, the segment is enqueued to a host dirty queue. Transition from clean to free queues is performed on an LRU basis.

- State transition from a host dirty queue

The host dirty queue contains data that reflects no parity. When parity generation is started, a state transition occurs to the parity in-creation queue.

- State transition from the parity in-creation queue

The parity in-creation queue contains parity in-creation data. When parity generation is completed, a transition to a physical dirty queue occurs.

- State transition from a physical dirty queue

When a write hit occurs in the data segment that is enqueued in a physical dirty queue, the segment is enqueued into the host dirty queue again. When destaging of the data segment is completed, the segment is enqueued into a queue (destaging of data segments occur asynchronously on an LRU basis).

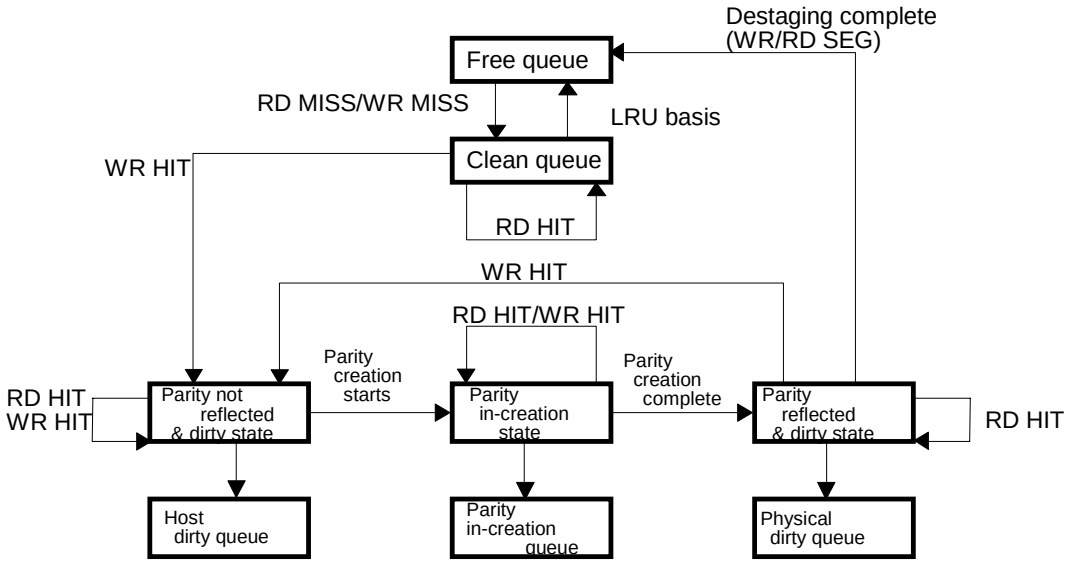
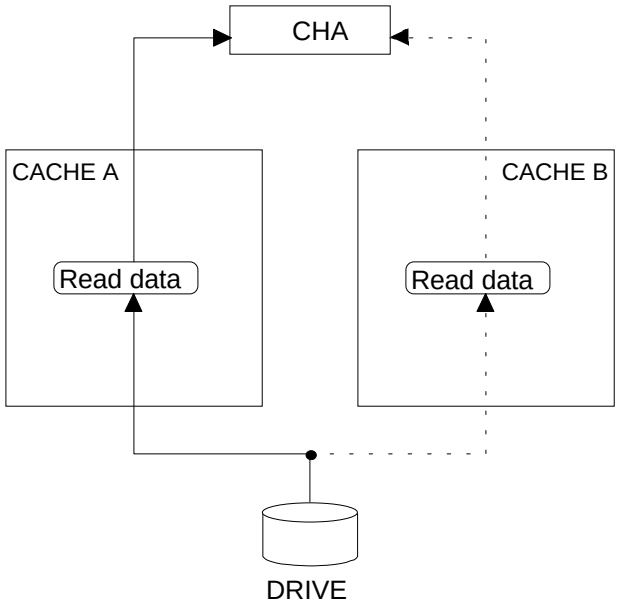


Fig. 3.4.4 Queue Segment State Transition Diagram

(5) Cache usage in the read mode



The cache area to be used for destaging read data is determined depending on whether the result of evaluating the following expression is odd or even:

(CYL# x 15 + HD#) / 16

The read data is destaged into area A if the result is even and into area B if the result is odd.

Fig. 3.4.5 Cache Usage in the Read Mode

Read data is not duplexed and its destaging cache area is determined by the formula shown in Fig. 3.4.5. Staging is performed not only on the segments containing the pertinent block but also on the subsequent segments up to the end of track (for increased hit ratio). Consequently, one track equivalence of data is prefetched starting at the target block. This formula is introduced so that the cache activity ratios for areas A and B are even. The staged cache area is called the cache area and the other area NVS area.

(6) Cache usage in the write mode

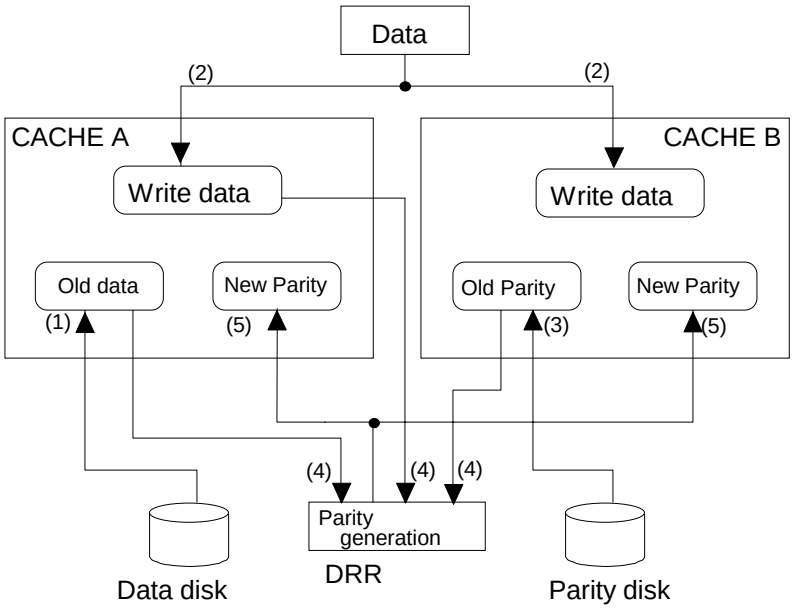


Fig. 3.4.6 Cache Usage in the Write Mode

This system handles write data (new data) and read data (old data) in separate segments as shown in Fig. 3.4.6 (not overwritten as in the conventional systems), whereby compensating for the write penalty.

(1) If the write data in question causes a cache miss, the data from the block containing the target record up to the end of the track is staged into a read data slot.
(2) In parallel with step (1), the write data is transferred when the block in question is established in the read data slot.
(3) The parity data for the block in question is checked for a hit or miss condition and, if a cache miss condition is detected, the old parity is staged into a read parity slot.
(4) When all data necessary for generating new parity is established, it is transferred to the DRR circuit in the DKA.
(5) When the new parity is completed, the DRR transfers it into the write parity slots for cache A and cache B (the new parity is handled in the same manner as the write data).

The reason for writing the write data into both cache areas is that data will be lost if a cache error occurs when it is not yet written on the disk.

Although two cache areas are used as explained above, the read data (including parity) is staged into either cache A or cache B simply by duplexing only the write data (including parity) (in the same manner as in the read mode).

(7) CFW-inhibited write-operation (with Cache single-side error)
The non RAID-type Disk systems write data directly onto disk storage in the form of cache through, without performing a DFW, when a cache error occurs. In this system, cache must always be passed, which fact disables the through operation. Consequently, the write data is duplexed, and a CFW-inhibited write-operation is performed; that is, when one cache subsystem goes down, the end of processing status is not reported until the data write in the other cache subsystem is completed. This process is called CFW-inhibited write-operation.

The control information necessary for controlling cache is stored in the shared memory.

(8) Shared memory

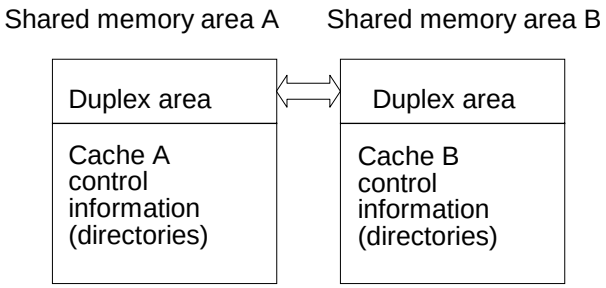


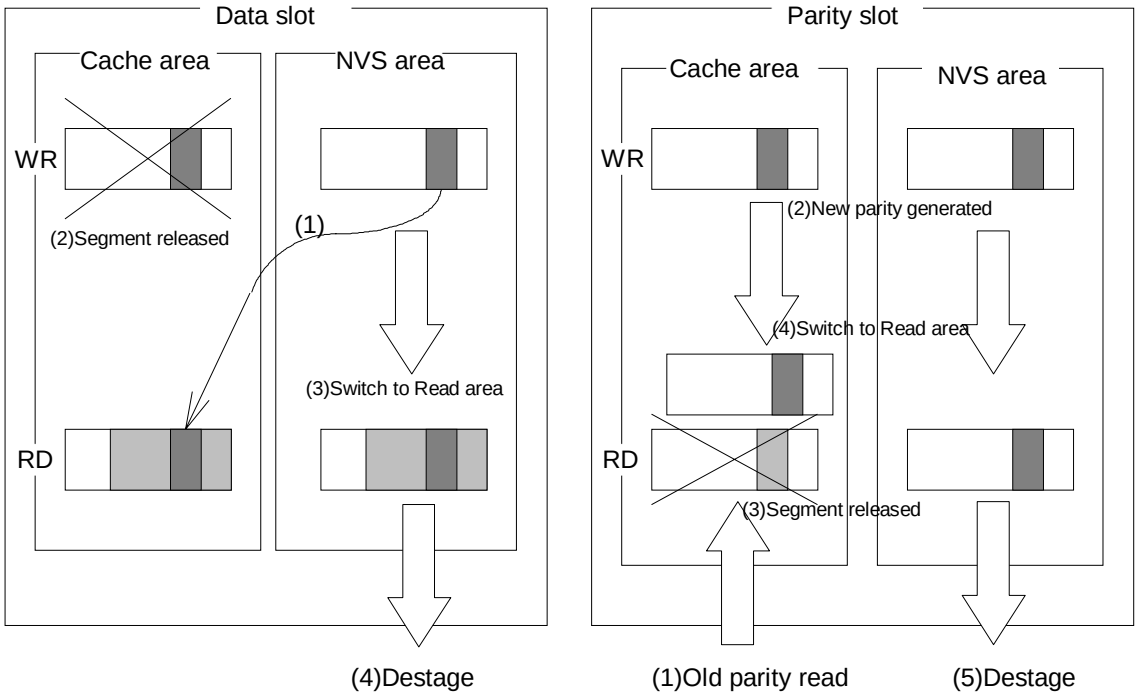
Fig. 3.4.7 Outline of Shared Memory

This system has two areas of cache memory, shown in Fig. 3.4.7, as it has two areas of cache memory. One part of its internal data is fully duplexed (this serves as the role of the conventional ECM). The other part of the shared memory area contains the control information about the corresponding cache area (shared memory area A for cache A and shared memory area B for cache B). If an error occurs on one side of shared memory (A or B), the corresponding cache area becomes inoperative (equivalent to a cache error).

Like cache, shared memory is made non-volatile (approximately 96 hours) to prevent data loss in case of power failures.

3.5 Destaging Operations

(1) Cache management in the destage mode (RAID5)
Destaging onto a drive is deferred until parity generation is completed. Data and parity slot transitions in the destage mode occur as shown in Fig. 3.5.1



① The write data is copied from the NVS area into the read area.
② The write segment in the cache area is released.
③ Simultaneously, the segment in the NVS area is switched from write to read segment.
④ Destaging.
⑤ The read segment in the NVS area is released.

① Parity generation correction read (old parity) occurs.
② New parity is generated.
③ The old parity in the read segment is released.
④ The segments in the cache and NVS areas are switched from write to read segment.
⑤ Destaging.
⑥ The read segment in the NVS area is released.

Fig. 3.5.1 Cache Operation in the Destage Mode

Write data is stored in write segments before parity is generated but stored in read segments after parity is generated. When drive data is stored, therefore, the data from the read segment is transferred.

(2) Cache management in the destage mode (RAID1)
Data slot is destage to primary/secondary drive.

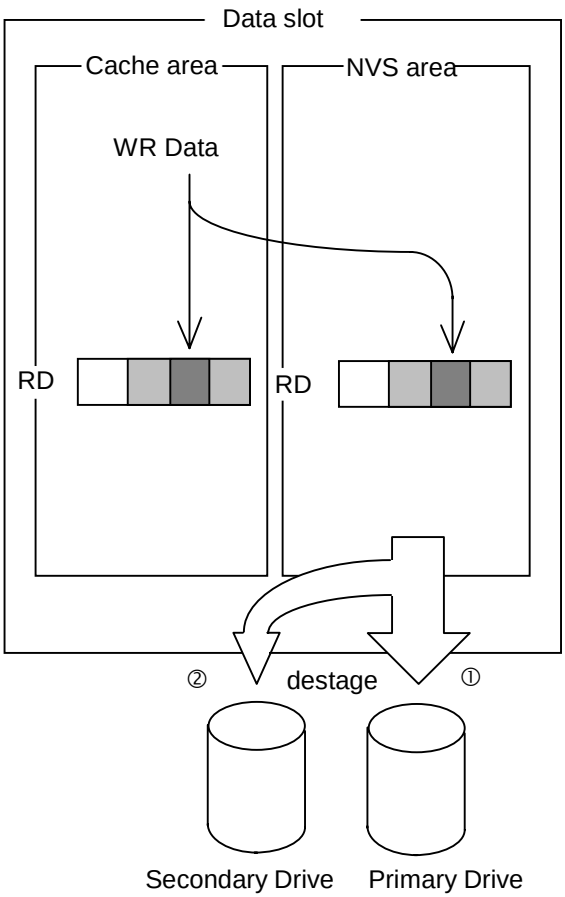


Fig. 3.5.3 RAID1 asynchronous destage

① Destage to primary drive.
② Destage to secondary drive.
③ The data read segment in the NVS area is released.

(3) Blocked data write

The purpose of blocked data write is to reduce the number of accesses to the drive during destaging, whereby increasing the subsystem performance. There are three modes of blocked data write: single-stripe blocking, multiple-stripe blocking, and drive blocking. These modes are briefly explained below.

- Single-stripe blocking

Two or more dirty segments in a stripe are combined into a single dirty data block. Contiguous dirty blocks are placed in a single area. If an unloaded block exists between dirty blocks, the system destages the dirty blocks separately at the unloaded block. If a clean block exists between dirty blocks, the system destages the blocks including the clean block.

- Multiple-stripe blocking

The sequence of stripes in a parity group are blocked to reduce the number of write penalties. This mode is useful for sequential data transfer.

- Drive blocking

In the drive blocking mode, blocks to be destaged are written in a block with a single drive command if they are contiguous when viewed from a physical drive to shorten the drive's latency time.

The single- and multiple-stripe blocking modes are also called in-cache blocking modes. The DMP determines which mode to use. The drive blocking mode is identified by the DSP.

3.6 Operations Performed when Drive Errors Occur

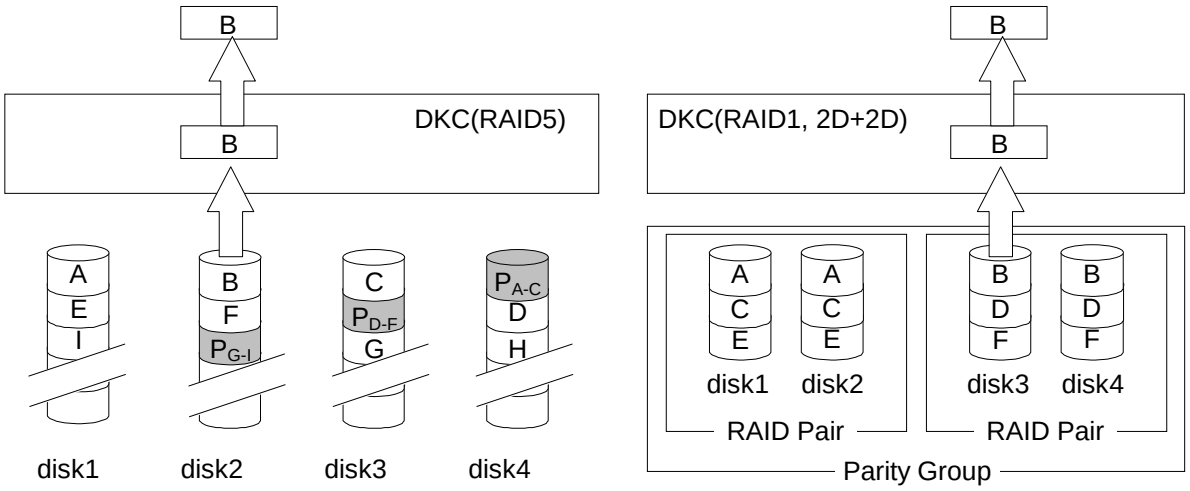
(1) I/O operations performed when drive errors occur

This system can recover target data using parity data and data stored on normal disk storage even when it cannot read data due to errors occurring on physical drives. This feature ensures non-disruptive processing of applications in case of drive errors. This system can also continue processing for the same reason in case errors occur on physical drives while processing write requests.

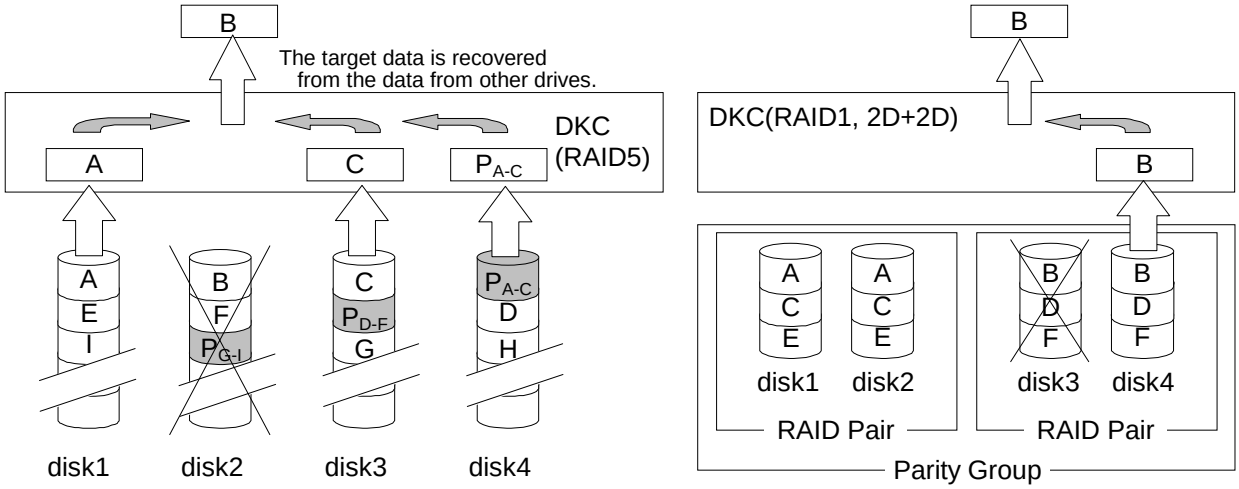
Figure 3.6.1 shows the outline of data read processing in case a drive error occurs.

Request for reading data B

(i) Normal time



(ii) When a disk error occurs



A,B,C • • • ; Data (A=A', B=B', C=C')

P ; Parity data

Fig. 3.6.1 Outline of Data Read Processing

(2) Data integrity feature and drive errors
This system uses spare disk drives and reconfigures any drives that are blocked due to errors or drives whose error count exceeds a specified limit value using spare disks.

Since this processing is executed on the host in the background, this system can continue to accept I/O requests. The data saved on spare disks are copied into the original location after the error drives are replaced with new ones.

1. Dynamic sparing
This system keeps track of the number of errors that occurred, for each drive, when it executes normal read or write processing. If the number of errors occurring on a certain drive exceeds a predetermined value, this system considers that the drive is likely to cause unrecoverable errors and automatically copies data from that drive to a spare disk. This function is called dynamic sparing.In RAID1 method, this system is same as RAID5 dynamic sparing.

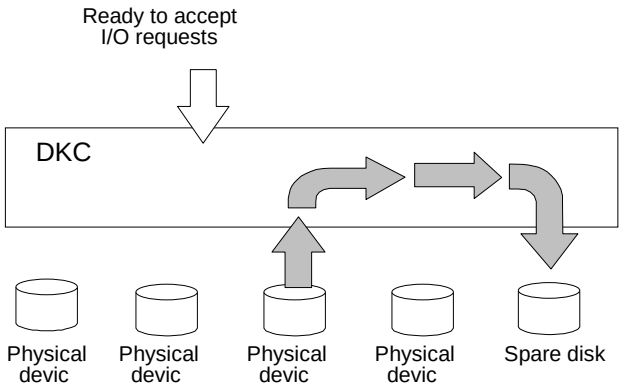


Fig. 3.6.2 Outline of the Dynamic Sparing Function

2. Correction copy
When this system cannot read or write data from or to a drive due to an error occurring on that drive, it regenerates the original data for that drive using data from the other drives and the parity data, and copies it onto a spare disk. In RAID1 method, this system copies data from the another drive to a spare disk.

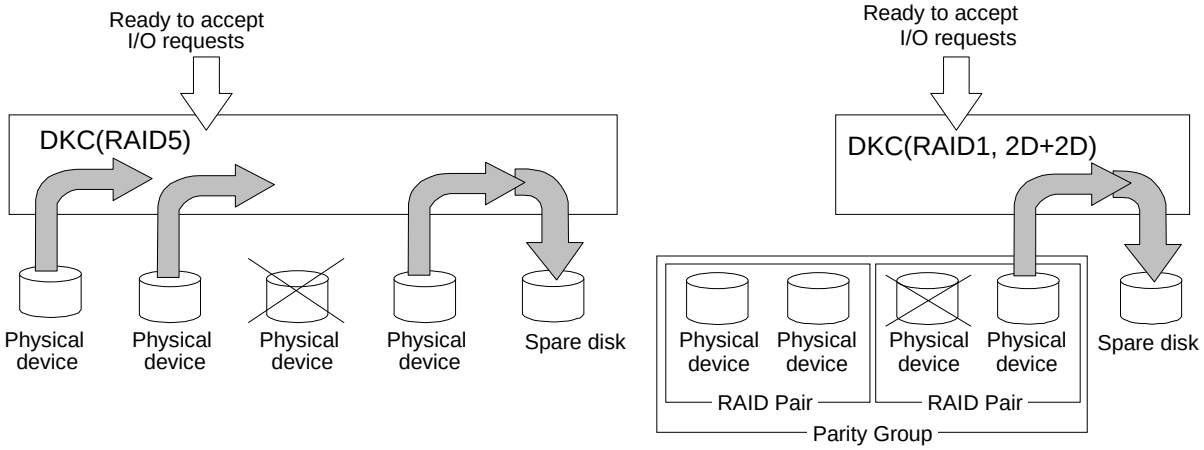


Fig. 3.6.3 Outline of the Correction Copy Function

3.7 Inter Mix of Drives and Emulation types

3.7.1 Drives to be Connected

The models of disk units which are connectable with the RAID400 disk subsystem and the specifications of each disk unit are shown in Table 3.7.1.

Table 3.7.1 Disk Drive Specifications (1/2)

Item		DKR2B-J18FC DKR2C-J18FC *1 DKR2D-J18FC *4 DKR2D-J18FD *5 DKR2E-J18FD *7 DKR2F-J18FD *10	DKR1B-J47FC DKR1C-J47FC *2 DKR2D-J47FD *6 DKR2E-J47FD *8 DKR2F-J47FD *11	DKR1C-J72FC DKR2D-J72FC *3 DKR2E-J72FC *9 DKR2F-J72FC *12	DKR2E-J146FC DKR2F-J146FC *13
Formatted capacity (User area)		18.46GB	47.19GB	72.91GB	143.7GB
Diameter of disk		3 inch	3 inch	3 inch	3 inch
Number of heads		9/6*1/3*4/5*5/5*7/3*10	23/16*2/10*6/5*8/3*11	24/10*3/5*9/3*12	10/5*13
Number of disks		5/3*1/2*4/3*5/3*7/2*10	12/8*2/5*6/3*8/2*11	12/5*3/3*9/2*12	5/3*13
Revolution speed (min^{-1})		10,025	10,025	10,025	10,025
Sector length (byte)		520(512)	520(512)	520(512)	520(512)
Seek time (ms) (Read/Write)	MIN.	0.5 / 0.7	0.5 / 0.7	0.5 / 0.7	0.5 / 0.7
	AVE.	5.2 / 6.0	5.7 / 6.5	5.7 / 6.5	4.9 / 5.4
	MAX.	12.0 / 13.0	12.0 / 13.0	12.0 / 13.0	10.0 / 11.0
Average latency time (ms)		2.99	2.99	2.99	2.99
Internal data transfer rate (MB/s)		30.2 to 45.6	30.2 to 45.6	33.6 to 56.6	57.3 to 99.9
Interface data transfer rate (MB/s)		Max.100	Max.100	Max.100	Max.100

*1: DKR2C-J18FC *2: DKR1C-J47FC *3: DKR2D-J72FC *13: DKR2F-J146FC
*4: DKR2D-J18FC *6: DKR2D-J47FD *9: DKR2E-J72FC
*5: DKR2D-J18FD *8: DKR2E-J47FD *12: DKR2F-J72FC
*7: DKR2E-J18FD *11: DKR2F-J47FD
*10: DKR2F-J18FD

Table 3.7.1 Disk Drive Specifications (2/2)

Item		DKS2A-K18FC DKS2B-K18FC *15	DKS1A-H180FC	DKS2B-K36FC DKS2C-K36FC *14	DKS2C-K72FC
Formatted capacity (User area)		17.87GB	172.55GB	35.76GB	71.50GB
Diameter of disk		2.5 inch	3 inch	2.5 inch	2.5 inch
Number of heads		10 / 4 *15	24	8 / 4 *14	8
Number of disks		5 / 2 *15	12	4 / 2 *14	4
Revolution speed (min^{-1})		14,904	7,200	14,904	14,904
Sector length (byte)		520(512)	520(512)	520(512)	520(512)
Seek time (ms) (Read/Write)	MIN.	0.5 / 0.7	0.8 / 1.1	0.4 / 0.8	0.4 / 0.8
	AVE.	3.9 / 4.5	7.4 / 8.2	3.8 / 4.2	3.8 / 4.2
	MAX.	7.5 / 8.0	16.0 /17.0	7.0 / 8.0	6.7 / 7.1
Average latency time (ms)		2.01	4.17	2.01	2.01
Internal data transfer rate (MB/s)		48.1 to 64.0	35.3 to 63.5	68.5 to 88.3	74.5 to 111.4
Interface data transfer rate (MB/s)		Max.100	Max.100	Max.100	Max.100

*15: DKS2B-K18FC *14: DKS2C-K36FC

The RAID400 disk subsystem can connect up to 512 disk (Single Cabinet Model : 48 disks) drives mentioned above, though the number of connectable disk drives varies with the emulation types and the RAID configuration. These will be explained in detail in Section 3.7.2.

SVP displays each drive model as the following table.

Disk drive model	SVP screen
DKR2B-J18FC	DKR2B-J018FC
DKR2C-J18FC	DKR2C-J018FC
DKR2D-J18FC	DKR2D-J018FC
DKR2D-J18FD	
DKR2E-J18FD	DKR2E-J018FC *1
DKR2F-J18FD	DKR2F-J018FC *1
DKS2A-K18FC	DKS2A-K018FC
DKS2B-K18FC	DKS2B-K018FC
DKS2B-K36FC	DKS2B-K036FC
DKS2C-K36FC	DKS2C-K036FC *2
DKR1B-J47FC	DKR1B-J047FC
DKR1C-J47FC	DKR1C-J047FC
DKR2D-J47FD	DKR2D-J047FC
DKR2E-J47FD	DKR2E-J047FC *1
DKR2F-J47FD	DKR2F-J047FC *1
DKR1C-J72FC	DKR1C-J072FC
DKR2D-J72FC	DKR2D-J072FC
DKR2E-J72FC	DKR2E-J072FC *1
DKR2F-J72FC	DKR2F-J072FC *1
DKS2C-K72FC	DKS2C-K072FC
DKR2E-J146FC	DKR2E-J146FC
DKR2F-J146FC	DKR2F-J146FC
DKS1A-H180FC	DKS1A-H180FC

*1: On some SVP μ version (ex. 01-18-6x/xx), it would be shown “DKR2D-JxxxFC”.
*2: On some SVP μ version (ex. 01-18-6x/xx), it would be shown “DKS2B-K036FC”.

3.7.2 Emulation Device Type

Refer to 3.5 Volume Specification in MULTIPLATFORM SECTION about OPEN Volume Type.

(1) Separate Model emulation list.
The emulation types of disk controller and disk units of the RAID400 Separate Model are shown in Tables 3.7.2.1 to 3.7.2.4.

Table 3.7.2.1 List of RAID400 Separate Model Emulation Types for RAID5 (3D+1P)

Item		Emulation contents						
Emulation Type	DKC	3990-3						
	DKU	3390-9	3390-3/3R	3390-2	3390-1	3380-K	3380-E	3380-J
Storage capacity (GB/volume)		8.51	2.84	1.89	0.95	1.89	1.26	0.63
number of volumes / parity group	DKU-F405I-18J4	6	18	27	55	28	42	84
	DKU-F405I-18K4	5	17	26	53	27	40	81
	DKU-F405I-36K4	11	35	53	107	54	81	163
	DKU-F405I-47J4	15	47	64/71	—	64/72	—	—
	DKU-F405I-72J4	24	73	109	—	111	—	—
	DKU-F405I-72K4	23	71	107	—	109	—	—
	DKU-F405I-146J4	48	144	—	—	219	—	—
	DKU-F405I-180H4	57	173	—	—	250	—	—
Maximum number of parity groups	DKU-F405I-18J4	126	126	126	74	126	97	48
	DKU-F405I-18K4	126	126	126	77	126	102	50
	DKU-F405I-36K4	126	117	77	38	75	50	25
	DKU-F405I-47J4	126	87	64	—	64	—	—
	DKU-F405I-72J4	126	56	37	—	36	—	—
	DKU-F405I-72K4	126	57	38	—	37	—	—
	DKU-F405I-146J4	85	28	—	—	18	—	—
	DKU-F405I-180H4	71	23	—	—	16	—	—
Maximum number of volumes	DKU-F405I-18J4	756	2268	3402	4070	3528	4074	4032
	DKU-F405I-18K4	630	2142	3276	4081	3402	4080	4050
	DKU-F405I-36K4	1386	4095	4081	4066	4050	4050	4075
	DKU-F405I-47J4	1890	4089	4096	—	4096	—	—
	DKU-F405I-72J4	3024	4088	4033	—	3996	—	—
	DKU-F405I-72K4	2898	4047	4066	—	4033	—	—
	DKU-F405I-146J4	4080	4032	—	—	3942	—	—
	DKU-F405I-180H4	4047	3979	—	—	4000	—	—
Subsystem capacity (user area)	DKU-F405I-18J4	51 GB to 6434 GB	51 GB to 6441 GB	51 GB to 6430 GB	52 GB to 3867 GB	53 GB to 6668 GB	53 GB to 5133 GB	53 GB to 2540 GB
	DKU-F405I-18K4	43 GB to 5361 GB	48 GB to 6083 GB	49 GB to 6192 GB	50 GB to 3877 GB	51 GB to 6430 GB	50 GB to 5141 GB	51 GB to 2552 GB
	DKU-F405I-36K4	94 GB to 11795 GB	99 GB to 11630 GB	100 GB to 7713 GB	102 GB to 3863 GB	102 GB to 7654 GB	102 GB to 5103GB	103 GB to 2567 GB
	DKU-F405I-47J4	128 GB to 16084 GB	133 GB to 11613 GB	121 GB to 7741 GB	—	121 GB to 7741 GB	—	—
	DKU-F405I-72J4	204 GB to 25734 GB	207 GB to 11610 GB	206 GB to 7622 GB	—	210 GB to 7552 GB	—	—
	DKU-F405I-72K4	196 GB to 24662 GB	201 GB to 11485 GB	202 GB to 7685 GB	—	206 GB to 7622 GB	—	—
	DKU-F405I-146J4	408 GB to 34720 GB	408 GB to 11450 GB	—	—	413 GB to 7450 GB	—	—
	DKU-F405I-180H4	485 GB to 34439 GB	491 GB to 11300 GB	—	—	472 GB to 7560 GB	—	—

Note : The emulation type 3390-3 and 3390-3R cannot be intermixed in the subsystem.

Table 3.7.2.2 List of RAID400 Separate Model Emulation Types for RAID5 (3D+1P)

Item		Emulation contents				
Emulation	DKC	3990-6/6E/I-2105				I-2105
Type	DKU	3390-9	3390-3/3R	3390-2	3390-1	3390-L *1
Storage capacity (GB/volume)		8.51	2.84	1.89	0.95	27.8
number of volumes / parity group	DKU-F405I-18J4	6	18	27	55	—
	DKU-F405I-18K4	5	17	26	53	—
	DKU-F405I-36K4	11	35	53	107	—
	DKU-F405I-47J4	15	47	64/71	—	4
	DKU-F405I-72J4	24	73	109	—	7
	DKU-F405I-72K4	23	71	107	—	7
	DKU-F405I-146J4	48	144	—	—	14
	DKU-F405I-180H4	57	173	—	—	—
Maximum number of parity groups	DKU-F405I-18J4	126	126	126	74	—
	DKU-F405I-18K4	126	126	126	75	—
	DKU-F405I-36K4	126	117	77	38	—
	DKU-F405I-47J4	126	87	64	—	126
	DKU-F405I-72J4	126	56	37	—	126
	DKU-F405I-72K4	126	57	38	—	126
	DKU-F405I-146J4	85	28	—	—	126
	DKU-F405I-180H4	71	23	—	—	—
Maximum number of volumes	DKU-F405I-18J4	756	2268	3402	4070	—
	DKU-F405I-18K4	630	2142	3276	4081	—
	DKU-F405I-36K4	1386	4095	4081	4066	—
	DKU-F405I-47J4	1890	4089	4096	—	504
	DKU-F405I-72J4	3024	4088	4033	—	882
	DKU-F405I-72K4	2898	4047	4066	—	882
	DKU-F405I-146J4	4080	4032	—	—	1764
	DKU-F405I-180H4	4047	3979	—	—	—
Subsystem capacity (user area)	DKU-F405I-18J4	51 GB to 6434 GB	51 GB to 6441 GB	51 GB to 6430 GB	52 GB to 3867 GB	—
	DKU-F405I-18K4	43 GB to 5361 GB	48 GB to 6083 GB	49 GB to 6192 GB	50 GB to 3877 GB	—
	DKU-F405I-36K4	94 GB to 11795 GB	99 GB to 11630 GB	100 GB to 7713 GB	102 GB to 3863 GB	—
	DKU-F405I-47J4	128 GB to 16084 GB	133 GB to 11613 GB	121 GB to 7741 GB	—	111 GB to 14011 GB
	DKU-F405I-72J4	204 GB to 25734 GB	207 GB to 11610 GB	206 GB to 7622 GB	—	195 GB to 24520 GB
	DKU-F405I-72K4	196 GB to 24662 GB	201 GB to 11485 GB	202 GB to 7685 GB	—	195 GB to 24520 GB
	DKU-F405I-146J4	408 GB to 34720 GB	408 GB to 34720 GB	—	—	389 GB to 49039 GB
	DKU-F405I-180H4	485 GB to 34439 GB	491 GB to 11300 GB	—	—	—

Note : The 3390-6 emulation type is available for the specific customer (RPQ).
Therefore, do not define this emulation type, even though SVP shows this emulation on the volume (LDEV) installation menu.

*1 : If DKU emulation type is 3390-L, DKC emulation type must be I-2105.
DKU emulation type is 3390-L necessitates a HOST Software with PTF as follows.

- OS/390 V2R10 and higher releases.
  PTF: UW99406 (APAR: OW47497, OW50734)
  UQ60515 (APAR: PQ42534)

Table 3.7.2.3 List of RAID400 Separate Model Emulation Types for RAID1 (2D+2D)

Item		Emulation contents						
Emulation	DKC	3990-3						
Type	DKU	3390-9	3390-3/3R	3390-2	3390-1	3380-K	3380-E	3380-J
Storage capacity (GB/volume)		8.51	2.84	1.89	0.95	1.89	1.26	0.63
number of volumes / parity group	DKU-F405I-18J4	4	12	18	37	18	28	56
	DKU-F405I-18K4	3	11	17	35	18	27	54
	DKU-F405I-36K4	7	23	35	71	36	54	109
	DKU-F405I-47J4	10	31	47	—	48	—	—
	DKU-F405I-72J4	16	48	73	—	74	—	—
	DKU-F405I-72K4	15	47	71	—	72	—	—
	DKU-F405I-146J4	32	96	—	—	146	—	—
	DKU-F405I-180H4	38	115	—	—	175	—	—
Maximum number of parity groups	DKU-F405I-18J4	126	126	126	110	126	126	73
	DKU-F405I-18K4	126	126	126	117	126	126	75
	DKU-F405I-36K4	126	126	117	57	113	75	37
	DKU-F405I-47J4	126	126	87	—	85	—	—
	DKU-F405I-72J4	126	85	56	—	55	—	—
	DKU-F405I-72K4	126	87	57	—	56	—	—
	DKU-F405I-146J4	126	42	—	—	28	—	—
	DKU-F405I-180H4	107	35	—	—	23	—	—
Maximum number of volumes	DKU-F405I-18J4	504	1512	2268	4070	2268	3528	4088
	DKU-F405I-18K4	378	1386	2142	4095	2268	3402	4050
	DKU-F405I-36K4	882	2898	4095	4047	4068	4050	4033
	DKU-F405I-47J4	1260	3906	4089	—	4080	—	—
	DKU-F405I-72J4	2016	4080	4088	—	4070	—	—
	DKU-F405I-72K4	1890	4089	4047	—	4032	—	—
	DKU-F405I-146J4	4032	4032	—	—	4088	—	—
	DKU-F405I-180H4	4066	4025	—	—	4025	—	—
Subsystem capacity (user area)	DKU-F405I-18J4	34 GB to 4289 GB	34 GB to 4294 GB	34 GB to 4287 GB	35 GB to 3867 GB	34 GB to 4287 GB	35 GB to 4445 GB	35 GB to 2575 GB
	DKU-F405I-18K4	26 GB to 3217 GB	31 GB to 3936 GB	32 GB to 4048 GB	33 GB to 3890 GB	34 GB to 4287 GB	34 GB to 4287 GB	34 GB to 2552 GB
	DKU-F405I-36K4	60 GB to 7506 GB	65 GB to 8230 GB	66 GB to 7740 GB	67 GB to 3845 GB	68 GB to 7689 GB	68 GB to 5103 GB	69 GB to 2541 GB
	DKU-F405I-47J4	85 GB to 10723 GB	88 GB to 11093 GB	89 GB to 7728 GB	—	91 GB to 7711 GB	—	—
	DKU-F405I-72J4	136 GB to 17156 GB	136 GB to 11587 GB	138 GB to 7726 GB	—	140 GB to 7692 GB	—	—
	DKU-F405I-72K4	128 GB to 16084 GB	133 GB to 11605 GB	134 GB to 7649 GB	—	136 GB to 7620 GB	—	—
	DKU-F405I-146J4	272 GB to 34312 GB	272 GB to 34312 GB	—	—	275 GB to 7726 GB	—	—
	DKU-F405I-180H4	323 GB to 34601 GB	326 GB to 11431 GB	—	—	330 GB to 7607 GB	—	—

Table 3.7.2.4 List of RAID400 Separate Model Emulation Types for RAID1 (2D+2D)

Item		Emulation contents				
Emulation	DKC	3990-6/6E/I-2105				I-2105
Type	DKU	3390-9	3390-3/3R	3390-2	3390-1	3390-L *1
Storage capacity (GB/volume)		8.51	2.84	1.89	0.95	27.8
number of volumes / parity group	DKU-F405I-18J4	4	12	18	37	—
	DKU-F405I-18K4	3	11	17	35	—
	DKU-F405I-36K4	7	23	35	71	—
	DKU-F405I-47J4	10	31	47	—	3
	DKU-F405I-72J4	16	48	73	—	4
	DKU-F405I-72K4	15	47	71	—	4
	DKU-F405I-146J4	32	96	—	—	9
	DKU-F405I-180H4	38	115	—	—	—
Maximum number of parity groups	DKU-F405I-18J4	126	126	126	110	—
	DKU-F405I-18K4	126	126	126	117	—
	DKU-F405I-36K4	126	126	117	57	—
	DKU-F405I-47J4	126	126	87	—	126
	DKU-F405I-72J4	126	85	56	—	126
	DKU-F405I-72K4	126	87	57	—	126
	DKU-F405I-146J4	126	42	—	—	126
	DKU-F405I-180H4	107	35	—	—	—
Maximum number of volumes	DKU-F405I-18J4	504	1512	2268	4070	—
	DKU-F405I-18K4	378	1386	2142	4095	—
	DKU-F405I-36K4	882	2898	4095	4047	—
	DKU-F405I-47J4	1260	3906	4089	—	378
	DKU-F405I-72J4	2016	4080	4088	—	504
	DKU-F405I-72K4	1890	4089	4047	—	504
	DKU-F405I-146J4	4032	4032	—	—	1134
	DKU-F405I-180H4	4066	4025	—	—	—
Subsystem capacity (user area)	DKU-F405I-18J4	34 GB to 4289 GB	34 GB to 4294 GB	34 GB to 4287 GB	35 GB to 3867 GB	—
	DKU-F405I-18K4	26 GB to 3217 GB	31 GB to 3936 GB	32 GB to 4048 GB	33 GB to 3890 GB	—
	DKU-F405I-36K4	60 GB to 7506 GB	65 GB to 8230 GB	66 GB to 7740 GB	67 GB to 3845 GB	—
	DKU-F405I-47J4	85 GB to 10723 GB	88 GB to 11093 GB	89 GB to 7728 GB	—	83 GB to 10508 GB
	DKU-F405I-72J4	136 GB to 17156 GB	136 GB to 11587 GB	138 GB to 7726 GB	—	111 GB to 14011 GB
	DKU-F405I-72K4	128 GB to 16084 GB	133 GB to 11605 GB	134 GB to 7649 GB	—	111 GB to 14011 GB
	DKU-F405I-146J4	272 GB to 34312 GB	272 GB to 34312 GB	—	—	250 GB to 31525 GB
	DKU-F405I-180H4	323 GB to 34601 GB	326 GB to 11431 GB	—	—	—

Note : The 3390-6 emulation type is available for the specific customer (RPQ).
Therefore, do not define this emulation type, even though SVP shows this emulation on the volume (LDEV) installation menu.

*1 : If DKU emulation type is 3390-L, DKC emulation type must be I-2105.
DKU emulation type is 3390-L necessitates a HOST Software with PTF.
■ OS/390 V2R10 and higher releases.
PTF: UW99406 (APAR: OW47497, OW50734)
UQ60515 (APAR: PQ42534)

(2) Single Cabinet Model emulation list.
The emulation types of disk controller and disk units of the RAID400 Single Cabinet Model are shown in Tables 3.7.2.5 to 3.7.2.8.

Table 3.7.2.5 List of RAID400 Single Cabinet Model Emulation Types for RAID5 (3D+1P)

Item		Emulation contents						
Emulation	DKC	3990-3						
Type	DKU	3390-9	3390-3	3390-2	3390-1	3380-K	3380-E	3380-J
Storage capacity (GB/volume)		8.51	2.84	1.89	0.95	1.89	1.26	0.63
number of volumes / parity group	DKU-F405I-18J4	6	18	27	55	28	42	84
	DKU-F405I-18K4	5	17	26	53	27	41	82
	DKU-F405I-36K4	11	35	53	107	54	81	163
	DKU-F405I-47J4	15	47	71	—	72	—	—
	DKU-F405I-72J4	24	73	109	—	111	—	—
	DKU-F405I-72K4	23	71	107	—	109	—	—
	DKU-F405I-146J4	48	144	—	—	219	—	—
	DKU-F405I-180H4	57	173	—	—	250	—	—
Maximum number of parity groups	DKU-F405I-18J4	11	11	11	11	11	11	11
	DKU-F405I-18K4	11	11	11	11	11	11	11
	DKU-F405I-36K4	11	11	11	11	11	11	11
	DKU-F405I-47J4	11	11	11	—	11	—	—
	DKU-F405I-72J4	11	11	11	—	11	—	—
	DKU-F405I-72K4	11	11	11	—	11	—	—
	DKU-F405I-146J4	11	11	—	—	11	—	—
	DKU-F405I-180H4	11	11	—	—	11	—	—
Maximum number of volumes	DKU-F405I-18J4	66	198	297	605	308	462	924
	DKU-F405I-18K4	55	187	286	583	297	440	891
	DKU-F405I-36K4	121	385	583	1177	594	891	1793
	DKU-F405I-47J4	165	517	781	—	792	—	—
	DKU-F405I-72J4	264	803	1199	—	1221	—	—
	DKU-F405I-72K4	253	781	1177	—	1199	—	—
	DKU-F405I-146J4	528	1584	—	—	2409	—	—
	DKU-F405I-180H4	627	1903	—	—	2750	—	—
Subsystem capacity (user area)	DKU-F405I-18J4	51 GB to 562 GB	51 GB to 562 GB	51 GB to 561 GB	52 GB to 575 GB	53 GB to 582 GB	53 GB to 582 GB	53 GB to 582 GB
	DKU-F405I-18K4	43 GB to 468 GB	48 GB to 531 GB	49 GB to 541 GB	50 GB to 554 GB	51 GB to 562 GB	50 GB to 554 GB	51 GB to 561 GB
	DKU-F405I-36K4	94 GB to 1030 GB	99 GB to 1093 GB	100 GB to 1102 GB	102 GB to 1118 GB	102 GB to 1123 GB	102 GB to 1123 GB	103 GB to 1130 GB
	DKU-F405I-47J4	128 GB to 1404 GB	133 GB to 1468 GB	134 GB to 1476 GB	—	136 GB to 1497 GB	—	—
	DKU-F405I-72J4	204 GB to 2247 GB	207 GB to 2281 GB	206 GB to 2266 GB	—	210 GB to 2308 GB	—	—
	DKU-F405I-72K4	196 GB to 2153 GB	201 GB to 2216 GB	202 GB to 2225 GB	—	206 GB to 2266 GB	—	—
	DKU-F405I-146J4	408 GB to 4493 GB	408 GB to 4498 GB	—	—	413 GB to 4553 GB	—	—
	DKU-F405I-180H4	485 GB to 5335 GB	491 GB to 5404 GB	—	—	472 GB to 5197 GB	—	—

Note : The emulation type 3390-3 and 3390-3R cannot be intermixed in the subsystem.

Table 3.7.2.6 List of RAID400 Single Cabinet Model Emulation Types for RAID5 (3D+1P)

Item		Emulation contents				
Emulation	DKC	3990-6/6E/I-2105				I-2105
Type	DKU	3390-9	3390-3	3390-2	3390-1	3390-L *1
Storage capacity (GB/volume)		8.51	2.84	1.89	0.95	27.8
number of volumes / parity group	DKU-F405I-18J4	6	18	27	55	—
	DKU-F405I-18K4	5	17	26	53	—
	DKU-F405I-36K4	11	35	53	107	—
	DKU-F405I-47J4	15	47	71	—	4
	DKU-F405I-72J4	24	73	109	—	7
	DKU-F405I-72K4	23	71	107	—	7
	DKU-F405I-146J4	48	144	—	—	14
	DKU-F405I-180H4	57	173	—	—	—
Maximum number of parity groups	DKU-F405I-18J4	11	11	11	11	—
	DKU-F405I-18K4	11	11	11	11	—
	DKU-F405I-36K4	11	11	11	11	—
	DKU-F405I-47J4	11	11	11	—	11
	DKU-F405I-72J4	11	11	11	—	11
	DKU-F405I-72K4	11	11	11	—	11
	DKU-F405I-146J4	11	11	—	—	11
	DKU-F405I-180H4	11	11	—	—	—
Maximum number of volumes	DKU-F405I-18J4	66	198	297	605	—
	DKU-F405I-18K4	55	187	286	583	—
	DKU-F405I-36K4	121	385	583	1177	—
	DKU-F405I-47J4	165	517	781	—	44
	DKU-F405I-72J4	264	803	1199	—	77
	DKU-F405I-72K4	253	781	1177	—	77
	DKU-F405I-146J4	528	1584	—	—	154
	DKU-F405I-180H4	627	1903	—	—	—
Subsystem capacity (user area)	DKU-F405I-18J4	51 GB to 562 GB	51 GB to 562 GB	51 GB to 561 GB	52 GB to 575 GB	—
	DKU-F405I-18K4	43 GB to 468 GB	48 GB to 531 GB	49 GB to 541 GB	50 GB to 554 GB	—
	DKU-F405I-36K4	94 GB to 1030 GB	99 GB to 1093 GB	100 GB to 1102 GB	102 GB to 1118 GB	—
	DKU-F405I-47J4	128 GB to 1404 GB	133 GB to 1468 GB	134 GB to 1476 GB	—	111 GB to 1223 GB
	DKU-F405I-72J4	204 GB to 2247 GB	207 GB to 2281 GB	206 GB to 2266 GB	—	195 GB to 2141 GB
	DKU-F405I-72K4	196 GB to 2153 GB	201 GB to 2216 GB	202 GB to 2225 GB	—	195 GB to 2141 GB
	DKU-F405I-146J4	408 GB to 4493 GB	408 GB to 4493 GB	—	—	389 GB to 4281 GB
	DKU-F405I-180H4	485 GB to 5335 GB	491 GB to 5404 GB	—	—	—

Note : The 3390-6 emulation type is available for the specific customer (RPQ).
Therefore, do not define this emulation type, even though SVP shows this emulation on the volume (LDEV) installation menu.

*1 : If DKU emulation type is 3390-L, DKC emulation type must be I-2105.
DKU emulation type is 3390-L necessitates a HOST Software with PTF.
■ OS/390 V2R10 and higher releases.
PTF: UW99406 (APAR: OW47497, OW50734)
UQ60515 (APAR: PQ42534)

Table 3.7.2.7 List of RAID400 Single Cabinet Model Emulation Types for RAID1 (2D+2D)

Item		Emulation contents						
Emulation	DKC	3990-3						
Type	DKU	3390-9	3390-3	3390-2	3390-1	3380-K	3380-E	3380-J
Storage capacity (GB/volume)		8.51	2.84	1.89	0.95	1.89	1.26	0.63
number of volumes / parity group	DKU-F405I-18J4	4	12	18	37	18	28	56
	DKU-F405I-18K4	3	11	17	35	18	27	54
	DKU-F405I-36K4	7	23	35	71	36	54	109
	DKU-F405I-47J4	10	31	47	—	48	—	—
	DKU-F405I-72J4	16	48	73	—	74	—	—
	DKU-F405I-72K4	15	47	71	—	72	—	—
	DKU-F405I-146J4	32	96	—	—	146	—	—
	DKU-F405I-180H4	38	115	—	—	175	—	—
Maximum number of parity groups	DKU-F405I-18J4	11	11	11	11	11	11	11
	DKU-F405I-18K4	11	11	11	11	11	11	11
	DKU-F405I-36K4	11	11	11	11	11	11	11
	DKU-F405I-47J4	11	11	11	—	11	—	—
	DKU-F405I-72J4	11	11	11	—	11	—	—
	DKU-F405I-72K4	11	11	11	—	11	—	—
	DKU-F405I-146J4	11	11	—	—	11	—	—
	DKU-F405I-180H4	11	11	—	—	11	—	—
Maximum number of volumes	DKU-F405I-18J4	44	132	198	407	198	308	616
	DKU-F405I-18K4	33	121	187	385	198	297	594
	DKU-F405I-36K4	77	253	385	781	396	594	1199
	DKU-F405I-47J4	110	341	517	—	528	—	—
	DKU-F405I-72J4	176	528	803	—	814	—	—
	DKU-F405I-72K4	165	517	781	—	792	—	—
	DKU-F405I-146J4	352	1056	—	—	1606	—	—
	DKU-F405I-180H4	418	1265	—	—	1925	—	—
Subsystem capacity (user area)	DKU-F405I-18J4	34 GB to 374 GB	34 GB to 375 GB	34 GB to 374 GB	35 GB to 387 GB	34 GB to 374 GB	35 GB to 388 GB	35 GB to 388 GB
	DKU-F405I-18K4	26 GB to 281 GB	31 GB to 344 GB	32 GB to 353 GB	33 GB to 366 GB	34 GB to 374 GB	34 GB to 374 GB	34 GB to 374 GB
	DKU-F405I-36K4	60 GB to 655 GB	65 GB to 719 GB	66 GB to 728 GB	67 GB to 742 GB	68 GB to 748 GB	68 GB to 748 GB	69 GB to 755 GB
	DKU-F405I-47J4	85 GB to 936 GB	88 GB to 968 GB	89 GB to 977 GB	—	91 GB to 998 GB	—	—
	DKU-F405I-72J4	136 GB to 1498 GB	136 GB to 1500 GB	138 GB to 1518 GB	—	140 GB to 1538 GB	—	—
	DKU-F405I-72K4	128 GB to 1404 GB	133 GB to 1467 GB	134 GB to 1476 GB	—	136 GB to 1497 GB	—	—
	DKU-F405I-146J4	272 GB to 2995 GB	272 GB to 2999 GB	—	—	275 GB to 3035 GB	—	—
	DKU-F405I-180H4	323 GB to 3557 GB	326 GB to 3592 GB	—	—	330 GB to 3638 GB	—	—

Table 3.7.2.8 List of RAID400 Single Cabinet Model Emulation Types for RAID1 (2D+2D)

Item		Emulation contents				
Emulation	DKC	3990-6/6E/I-2105				I-2105
Type	DKU	3390-9	3390-3	3390-2	3390-1	3390-L *1
Storage capacity (GB/volume)		8.51	2.84	1.89	0.95	27.8
number of volumes / parity group	DKU-F405I-18J4	4	12	18	37	—
	DKU-F405I-18K4	3	11	17	35	—
	DKU-F405I-36K4	7	23	35	71	—
	DKU-F405I-47J4	10	31	47	—	3
	DKU-F405I-72J4	16	48	73	—	4
	DKU-F405I-72K4	15	47	71	—	4
	DKU-F405I-146J4	32	96	—	—	9
	DKU-F405I-180H4	38	115	—	—	—
Maximum number of parity groups	DKU-F405I-18J4	11	11	11	11	—
	DKU-F405I-18K4	11	11	11	11	—
	DKU-F405I-36K4	11	11	11	11	—
	DKU-F405I-47J4	11	11	11	—	11
	DKU-F405I-72J4	11	11	11	—	11
	DKU-F405I-72K4	11	11	11	—	11
	DKU-F405I-146J4	11	11	—	—	11
	DKU-F405I-180H4	11	11	—	—	—
Maximum number of volumes	DKU-F405I-18J4	44	132	198	407	—
	DKU-F405I-18K4	33	121	187	385	—
	DKU-F405I-36K4	77	253	385	781	—
	DKU-F405I-47J4	110	341	517	—	33
	DKU-F405I-72J4	176	528	803	—	44
	DKU-F405I-72K4	165	517	781	—	44
	DKU-F405I-146J4	352	1056	—	—	99
	DKU-F405I-180H4	418	1265	—	—	—
Subsystem capacity (user area)	DKU-F405I-18J4	34 GB to 374 GB	34 GB to 375 GB	34 GB to 374 GB	35 GB to 387 GB	—
	DKU-F405I-18K4	26 GB to 281 GB	31 GB to 344 GB	32 GB to 353 GB	33 GB to 366GB	—
	DKU-F405I-36K4	60 GB to 655 GB	65 GB to 719 GB	66 GB to 728 GB	67GB to 742 GB	—
	DKU-F405I-47J4	85 GB to 936 GB	88 GB to 968 GB	89 GB to 977 GB	—	83 GB to 917 GB
	DKU-F405I-72J4	136 GB to 1498 GB	136 GB to 1500 GB	138 GB to 1518 GB	—	111 GB to 1223 GB
	DKU-F405I-72K4	128 GB to 1404 GB	133 GB to 1467 GB	134 GB to 1476 GB	—	111 GB to 1223 GB
	DKU-F405I-146J4	272 GB to 2995 GB	272 GB to 2999 GB	—	—	250 GB to 2752 GB
	DKU-F405I-180H4	323 GB to 3557 GB	326 GB to 3592 GB	—	—	—

Note : The 3390-6 emulation type is available for the specific customer (RPQ).
Therefore, do not define this emulation type, even though SVP shows this emulation on the volume (LDEV) installation menu.

*1 : If DKU emulation type is 3390-L, DKC emulation type must be I-2105.
DKU emulation type is 3390-L necessitates a HOST Software with PTF.
■ OS/390 V2R10 and higher releases.
PTF: UW99406 (APAR: OW47497, OW50734)
UQ60515 (APAR: PQ42534)

(3) Specifications for coexistence of elements

Table 3.7.2.9 shows permitted coexistence of RAID levels, HDD types, and emulation types respectively.

Table 3.7.2.9 Specifications for Coexistence of Elements in RAID 400

Item	Specification	Remarks
Coexistence of RAID levels	• RAID1/5 can exist in DKC.	
Coexistence of numbers of HDDs composing ECC group (coexistence of configurations 3D+1P and 2D+2D)	• The numbers of HDDs inside an ECC group are 4 HDDs and they are applicable to RAID5 (3D+1P) and RAID1 (2D+2D). • All HDD types support RAID5 (3D+1P) and RAID1 (2D+2D) configuration. • In case of Separate Model configurations 3D+1P and 2D+2D can coexist within units of same DKA pair. • In case of Single Cabinet Model configurations 3D+1P and 2D+2D can coexist within units of same DKA pair.	
Coexistence of HDD types	• HDD types can coexist in each ECC group. • The specification for selecting the spare HDD can be common in the subsystem if there exists a spare HDD having the same capacity as that of the data HDD, though the HDD types coexist.	
Coexistence of emulation types	• Emulation types can coexist in each ECC group. • LDEV ID addressing must be of the same emulation type (3380-K/E/J, 3390-1/2/3/9/L, or 3390-3R) for every 32-address boundary.	An emulation of the same system should be set for every 32 addressings due to generation restriction by the host.

3.8 4096 logical addresses

The host connection interface specification are outlined in Tables 3.8.1 and 3.8.2.

Table 3.8.1 List of Allowable Maximum Values of Host Connection Interface Items on the DKC Side

	ESCON channel	Fibre channel
Maximum number of CUs	16	
Maximum number of SSIDs	64	
Maximum number of LDEVs	4096	

Table 3.8.2 Allowable Range of Host Connection Interface Items on DKC Side

	ESCON channel	Fibre channel
CU address	0 to F	
SSID	0004 to FFFD	
Number of logical volumes	1 to 4096	

The specifications for number of logical paths are summarized in Table 3.8.3.

Table 3.8.3 Specification for Number of Logical Paths

ESCON channel	Per port	256 paths *1
	Per channel processor	512 paths
	Per channel adapter	1024 paths
Number of logical paths per DKC		Number of mounted LCPs × 256

*1: Maximum number of HOST paths per port are 16 paths.
(16 HOST paths × 16 CUs = 256 logical paths)

Table 3.8.4 Specification for Number of Logical Paths (FICON)

FICON channel	Per port	512 paths *1
	Per channel processor	512 paths
	Per channel adapter	1024 paths
Number of logical paths per DKC		Number of mounted HTPs × 512

*1: Maximum number of HOST paths per port are 16 paths.
(32 HOST paths × 16 CUs = 512 logical paths)

3.11 High-Speed LDEV Formatting

3.11.1 Outlines

In DKC, the LDEV formatting time is automatically shortened by providing the HDD with the LDEV formatting function and by making the HDD perform the formatting concurrently (except for Disk drive model = DKSXX).

Item No.	Item	Contents
1	SVP operation	The operation is performed by selecting functions from the Maintenance menu.
2	Display of execution status	Display of the execution progress in the SVP message box (%)
3	Execution result	Normal/abnormal LDEV: Same indications as the conventional ones are displayed. Normal/abnormal PDEV: STATUS is displayed.
4	Recovery action when a failure occurs	Same as the conventional one. However, a retry is to be executed in units of ECC. (Because the LDEV-FMT is terminated abnormally in units of ECC when a failure occurs in the HDD.)
5	Operation of the SVP which is a high-speed LDEV-FMT object	When an LDEV-FMT of more than one ECC is instructed, the high-speed processing is performed.
6	PS/OFF or powering off	The LDEV formatting is suspended. No automatic restart is executed.
7	SVP PC powering off during execution of an LDEV-FMT	After the SVP is rebooted, the indication before the PC powering off is displayed in succession.

3.11.2 Estimation of LDEV Formatting Time

Formatting time of Disk drive model = DKRXX

HDD Capacity	Formatting time
146 G bytes	approx. 90 min.
72 G bytes	approx. 72 min.
47 G bytes	approx. 52 min.
18 G bytes	approx. 20 min.

Formatting time of Disk drive model = DKSXX

HDD Capacity	Formatting time *
18 G bytes	Approx. 20 min/ECC group
36 G bytes	Approx. 40 min/ECC group
72 G bytes	Approx. 80 min/ECC group
180 G bytes	Approx. 200 min/ECC group

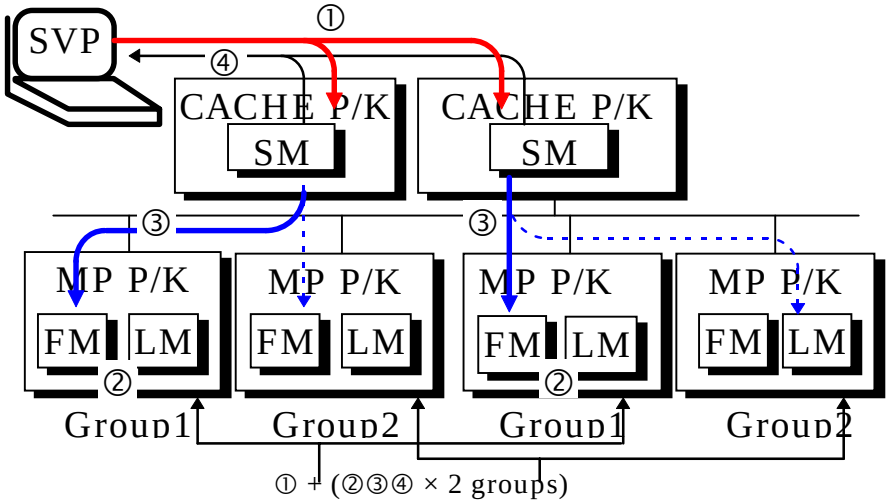
* Formatting time increases by Host commands.

3.12 High-Speed Online Replacement of Microprogram

The microprograms are stored in the shared memory and transferred in a batch. Thus the number of times of the transfer from the SVP to the DKC via the LAN is reduced and the online microprogram replacement is speeded up by recovering two or more processors at the same time.

3.12.1 Outline

A microprogram storage area is reserved in the shared memory. The microprogram by which each processor operates is stored in it, and data is written into the flash memory from the shared memory by rebooting the processor. The reason for executing the writing during reboot processing is that it is intended to unify the microprogram writing processings executed in the PCB replacement and cold replacement (not supported at present). In addition, when the microprograms are stored in the shared memory, processors can execute the microprogram writing processings at the same time and the processing time can be shorted substantially.



Two or more processors execute the processings at the same time.

[Processing sequence]

① The SVP transfers the microprograms (DKCMAIN, MAINDIFF, RAMBOOT, LCP/LCDG, and MCP) to the shared memory.
② The SVP executes a reboot instruction for each processor of the first group.
③ In the reboot processing, the microprograms are transferred from the shared memory to the flash memory and the writing processing is executed.
④ The SVP monitors the status of the processors executing the reboot processing and writing processing to the flash memory and waits until the status changes to normal finally.
⑤ The SVP executes the processings of steps ② to ④ repeatedly for each group.

3.12.2 Processing Time

The estimated time of microprogram replacement is shown below. The time below is required for the process for the programs DKCMAIN, MAINDIFF, RAMBOOT and LCP. The SVP and DKU are processed in the conventional processing sequences and the time required for the sequences are not included in the processing time.

Processing time required for microprogram replacement	
RAID400 method (Estimated time for 4 groups)	RAID300 method (Estimated time for 4 groups)
20 min. / 64 processors	**24 min./ 48 processors**

3.13 HRC

3.13.1 HRC Components

There are two modes of the update copy, a synchronous copy and asynchronous copy. Description of the asynchronous is described in chapter 3.17 HRC Asynchronous.

(1) HRC Components

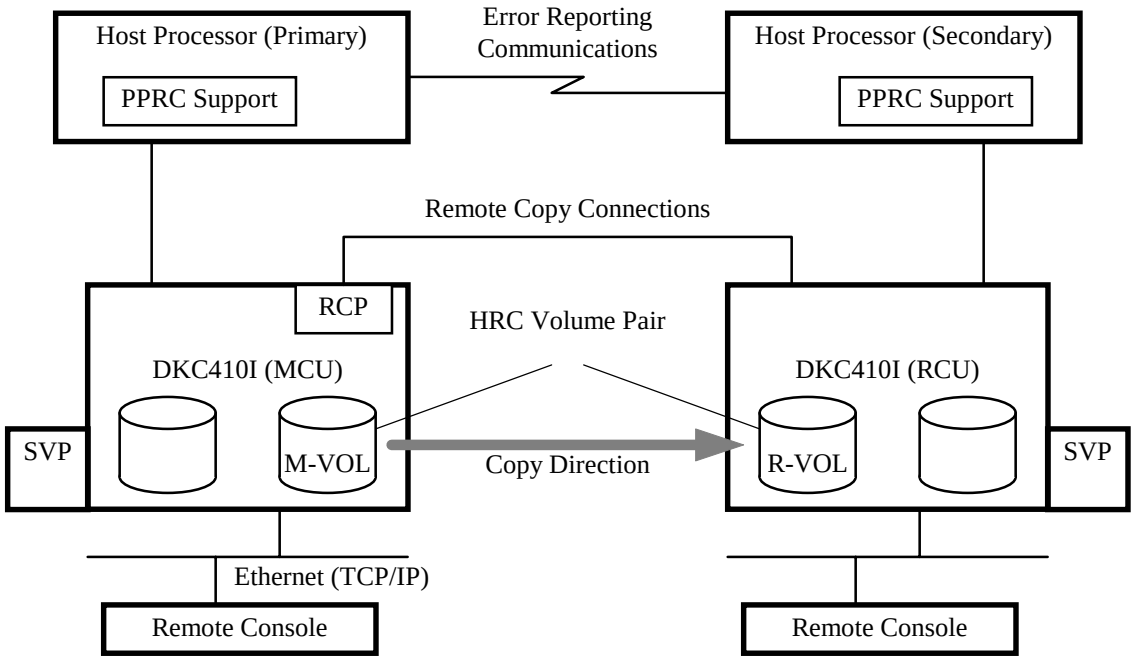


Fig. 3.13.1-1 HRC Components

(a) HRC Volume Pair

An **HRC volume pair** consists of two logical volumes, an M-VOL and an R-VOL, in different DKC410I subsystems.

An **M-VOL** (main volume) is a primary volume. It can be read or written by I/O operations from host processors.

An **R-VOL** (remote volume) is a secondary or a mirrored volume. Under control of the DKC410I subsystems, contents of an M-VOL and updates from host processors are copied to an R-VOL. Read or write I/O operations from host processors to R-VOLs are rejected.

Note ; R-VOL Read Only function

HRC has R-VOL Read Only function to accept read commands to R-VOL of suspended pairs of HRC.

R-VOL Read Only function becomes effective with SVP system option setting for RCU of HRC.

With this function, RCU accepts all RD commands including CTL/SNS commands and WR command to cylinder zero, head zero, record three of R-VOL. (It is necessary to change VOLSER of the volume.)

The RCU rejects some PPRC commands such as ADDPAIR to the R-VOL nevertheless the status of the R-VOL looks ‘Simplex’. They must be controlled by system administration.

With this function, RCU displays the status of the R-VOL as ‘Simplex’ instead of ‘Suspended’. It is necessary to accept I/O to R-VOL.

MCU copies cylinder zero of the pair at RESYNC copy unconditionally, besides the ordinary RESYNC copy.

With this function, if DKC Emulation type is 2105, CSUSPEND command to R-VOL of suspended Pair of HRC is rejected.

The M-VOLs of the HRC volume pairs and the R-VOLs of other HRC volume pairs can be intermixed in one DKC410I subsystem.

Note: Do not use M-VOLs or R-VOLs from hosts that have different CU emulation types (2105 and 3990) at the same time. If you use the M-VOLs or R-VOLs from the 2105 and 3990 hosts simultaneously, an MIH message might be reported to the 3990 host.

Note: When 3380 volume is used as M/R-VOL, CU emulation type of package used for connection between MCU and RCU should be other than 3990-6/6E/2105.

Note: When 3390-L volume is used as M/R-VOL, CU emulation type of package used for connection between MCU and RCU should be other than 3990-3/6/6E.

Note: When 3390-L volume is used as M/R-VOL, the reported FIRST/LAST CYL OF SYNC of the CQUERY command has the possibility that the one for the maximum 4CYL deviates.

(b) MCU and RCU

An **MCU** (main disk control unit) and an **RCU** (remote disk control unit) are disk control units in the DKC410I subsystems to which the M-VOLs and the R-VOLs are connected respectively.

An MCU controls I/O operations from host processors to the M-VOLs and copy activities between the M-VOLs and the R-VOLs. An MCU also provides functions to manage HRC status and configuration.

An RCU executes write operations directed by the MCU. The manner to execute write operations is almost same as that of I/O operations from host processors. An RCU also provides a part of functions to manage HRC status and configuration.

Note that an MCU/RCU is defined on each HRC volume pair basis. One disk control unit can operate as an MCU to control the M-VOLs and an RCU to control the R-VOLs.

Note: When Serial interface is used for Remote Copy connection paths, the controller emulation type of the connection port can be different with MCU and RCU.
But, when MCU and RCU are connected from the same HOST, the controller emulation type of the connection port must be the same (3990-3/6/6E or, 2105) with both MCU and RCU.

Note: Control unit serial number is used for pair configuration management. So one pair configuration can not include two control unit having same serial number.

(c) Remote Copy Connections

There are two kinds of Serial interface (ESCON/ACONARC) and Fibre channel interface of connection form.
Refer to “3.13.7 Fibre Channel connection” for Fibre channel interface connection.

An MCU and an RCU must be connected with serial interface (ESCON) cables. Only multi mode ESCON cables whose length is up to 3km can be connected to the DKC410I subsystems. In order to locate the DKC410I subsystems more than 3km apart, IBM 9032/9033 ESCON directors or 9036 ESCON repeaters are required.

At least two independent remote copy connections should be established between an MCU and an RCU.

(d) RCP

An **RCP** (remote control port) is a serial interface port to which an RCU is connected. Any serial interface port of the DKC410I subsystems can be configured as an RCP.

When an MCU communicates with an RCU through ESCON interface protocol, the RCP plays the role of a host processor channel. The RCP supports ESCON dynamic connection. A serial interface port of the RCU to which the MCU is connected can be connected to host processor channels by using dynamic switching capability provided by ESCON directors.

However an RCP can not communicate with host processor channel. Channel interface paths must be connected to other serial interface ports.

(e) SVP and Remote Console

An **SVP** provides functions to set up , modify and display HRC configuration and status.

A **remote console** is a personal computer compatible with the PC/AT. It should be connected to DKC410I subsystems with an Ethernet network (TCP/IP) . A remote console provides same managing HRC functions as an SVP. Several DKC410I subsystems can be connected with one Ethernet network.

For remote console, Hitachi provides only two software components, an HRC application program and dynamic link library. Both of them require Microsoft Windows operating system. A personal computer, Ethernet materials and other software products are not provided by Hitachi.

(f) Error Reporting Communications

Error reporting communication is a communication means between host processors. An MCU generates the sense information when it fails in keeping synchronization of HRC volume pair. The sense information causes the corresponding message to be displayed on the host processor console. For the reference during disaster recovery at the secondary (recovery) site, this console message should be transferred to the secondary site through the error reporting communication.

The error reporting communications may be configured by using channel-to-channel communications, Netview technology or other interconnect technologies, depending on installation. Hitachi does not provide any product for error reporting communications.

(g) PPRC Support

HRC provides a host processor interface compatible with IBM PPRC. TSO commands, DSF commands and disaster recovery PTFs provided for PPRC can be used for HRC.

(i) DKC emulation type = 2105

The RESETHP option of the CESTPATH command reset host's I/Os, so you have to stop them in advance.

Blank Sheet

Blank Sheet

3.13.2 HRC Theory of Operations

(1) HRC Copy Activities

HRC executes two kind of copy activities, **initial copy** and **update copy**.

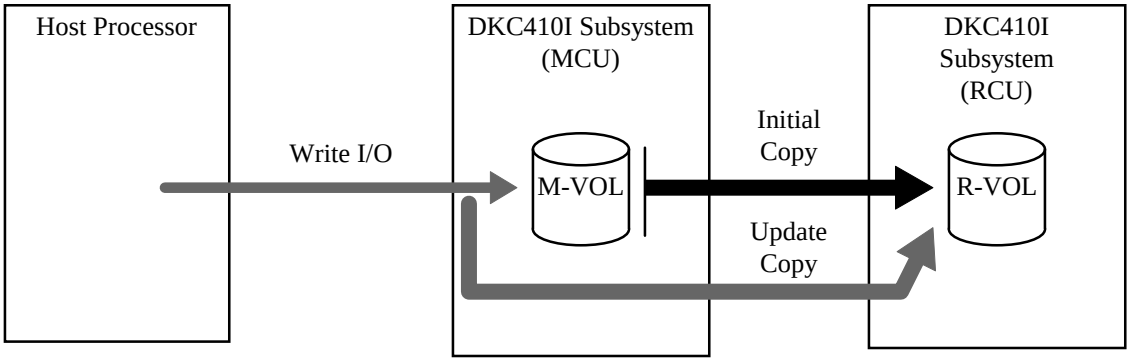


Fig. 3.13.2-1 HRC Copy Activities

(a) Initial Copy

Responding to an Establish HRC Volume Pair operation from an SVP/remote console or an ESTPAIR PPRC command, HRC begins initial copy. Data field of record zero and following records on all tracks, except for alternate and CE tracks, are copied from M-VOL to R-VOL. The initial copy operation is performed in ascending order of cylinder numbers.

"No copy" can be specified as a parameter to the initial copy. When "no copy" is specified, HRC will complete an Establish HRC Volume Pair operation without copying any data. An operator or a system administrator should be responsible for ensuring that data on the M-VOL and the R-VOL is already identical.

"Only out-of-sync cylinders" can also be specified as a parameter to the initial copy. This parameter is used to recover (re-establish) HRC volume pair from suspended condition. After suspending HRC volume pair, the MCU maintains a cylinder basis bit map which indicates the cylinders updated by I/O operations from the host processors. When this parameter is specified, HRC will copy only cylinders indicated by the bit map.

(b) Controlling Initial Copy

Number of tracks copied by one initial copy activity can be specified by an SVP/remote console or an ESTPAIR PPRC command.

Number of volume pairs for which the initial copy are concurrently executed and priority of each volume pair can be specified from an SVP/remote console.

(c) Update Copy

Responding to the write I/O operations from the host processors, HRC copies the records updated by the write I/O operation to the R-VOL.

The update copy is a **synchronous remote copy**. An MCU starts the update copy after responding only channel-end status to the host processor channel, and sends device-end status after completing the update copy. The MCU will start the update copy when it receives:

- the last write command in the current domain specified by preceding locate record command;
- a write command for which track switch to the next track is required;
- each write command without being preceded by locate record command.

If many consecutive records are updated by single CCW chain which does not use locate record command, the third condition above may cause the significant impact on performance.

(d) Update Copy for Cache Fast Write Data

Cache fast write (CFW) data does not always have to be copied because CFW is used for temporary files, such as sort work data sets. These temporary files are not always necessary for disaster recovery.

In order to reduce update copy activities, HRC supports a parameter which specifies whether CFW data should be copied or not.

(e) Special Write Command for Initial Copy and Update Copy

In order to reduce overhead by the copy activities, HRC uses a special write command which is allowed only for copy activities between the DKC410I subsystems. The single write command transfers control parameters and an FBA formatted data which includes consecutive updated records in a track. It reduces interlocks on ESCON interface protocol and overhead required for converting FBA-to-CKD format and CKD-to-FBA format.

(2) HRC Read I/O Operations

Responding to read I/O operations, an MCU transfers the requested records form an M-VOL to a host processor. Even if reading records from the M-VOL is failed, the R-VOL is not automatically read for recovery. The redundancy of the M-VOL itself provided by RAID5 or RAID1 technique would recover the failure.

Blank Sheet

(3) HRC Volume Pair Status

All volumes in a DKC410I subsystem are in one of the states shown in Table 3.13.2-1.

Status of the M-VOLs or the R-VOLs are kept by the MCU and the RCU respectively. The MCU is responsible to keep status of the R-VOLs identical to status of the M-VOLs. However, in the case of communication failure between the MCU and the RCU, they could be different.

From an SVP/remote console or by using an appropriate command for IBM PPRC, status of M-VOLs or status of R-VOLs can be obtained from the MCU or the RCU respectively.

Table 3.13.2-1 HRC Volume Status

Status	Description
Simplex	This volume does not belong to HRC volume pair. When the initial copy is started by an Add Pair operation, the volume is changed to "pending duplex" state.
Pending Duplex	The initial copy is in progress. Data on HRC volume pair is not fully identical. When completing the initial copy, the volume will be changed to "duplex" state.
Duplex	Volumes in HRC volume pair are synchronized. All updates from the host processors to the M-VOL are duplicated to the R-VOL.
Suspended	Volumes in HRC volume pair are not synchronized. When the MCU can not keep synchronization between HRC volume pair due to, for example, failure on the update copy, the MCU will put the M-VOL and the R-VOL in this state. When the MCU or the RCU accepts an Suspend operation from an SVP/remote console, the M-VOL and the R-VOL will be put in this state. When the RCU accepts the Delete Pair operation from the SVP/remote console, the MCU will detect the operation and put the M-VOL in this state.

For "suspended" volumes, cause of suspension is defined as shown in Table 3.13.2-2. The cause of suspension is displayed in the form of "Suspended (cause of suspension)" on an SVP/remote console.

Table 3.13.2-2 HRC Volume Status - Sub-status of Suspended Volume

Cause of Suspension	Description
M-VOL by Operator	The Suspend operation with "M-VOL failure" option was issued to the M-VOL. This cause of suspension is defined only for the M-VOLs.
R-VOL by Operator	The Suspend operation with "R-VOL" option was issued to the M-VOL or the R-VOL. This cause of suspension is defined for both the M-VOLs and the R-VOLs.
by MCU	The RCU received a request to suspend the R-VOL from an MCU. This cause of suspension is defined for only the R-VOLs.
by RCU	The MCU detected an error condition of the RCU which caused HRC volume pair to be suspended. This cause of suspension is defined only for the M-VOLs.
Delete Pair to RCU	The MCU detected that the R-VOL had been changed to "simplex" state by the Delete Pair operation. This cause of suspension is defined only for the M-VOLs.
R-VOL Failure	The MCU detected an error condition on the communication between the RCU or I/O error on the update copy. This cause of suspension is defined only for the M-VOLs. The cause of suspension of the R-VOLs are usually set to "by MCU" in this situation.
MCU IMPL	The MCU could not find valid control information in its non-volatile memory during its IMPL procedure. This situation may occur after the power supply failure.
Initial Copy Failed	The volume pair was suspended before completing the initial copy. Even if no write I/O has been issued after being suspended, the data in the R-VOL is not completely identical to the M-VOL.

Blank Sheet

3.13.3 HRC Control Operations

This section describes HRC control operations from an SVP or a remote console.

(1) Add RCU Operation

An Add RCU operation makes an MCU register the specified disk control unit as an RCU and establish the logical paths to the RCU. This operation also provides a function to modify the Remote Copy options which will be applied to all Remote Copy volume pairs in this subsystem. Refer to "SVP SECTION 4.3 Add RCU... Screen" for GUI image of this operation.

To register the RCU, the following parameters are required:

RCU S#	Serial number of the RCU.
SSID	SSID (subsystem identifier) of the RCU.
Num. of Path	Number of logical paths which should be established to the RCU on the remote copy connections.

Path parameters to specify one logical path are shown below. Up to four sets of the path parameters, as many as the "Num. of Paths" parameter specifies, must be specified. The description of path parameters is similar to the channel path definitions in IOCDS (I/O configuration dataset). In the IOCDS, a logical path is described with a sub-channel number, a link destination address and a logical address of the control unit image. A "Port" parameter, instead of the sub-channel number, is used to specify the serial interface port in the DKC410I subsystem.

Port	The serial interface port of the DKC410I subsystem where the logical path begins from. Before this operation, the serial interface port must be set to the RCP mode.
Link Adr	The link destination address. Similar to the logical path definitions in the IOCDS, this is the destination port address on the ESCD which is set to provide the dynamic connection capability. If the remote copy connection does not include the dynamic connection, x'00' must be specified.
Logical Adr	The logical control unit address of the control unit image. If the RCU is configured as an SPSD (single path storage director), either x'00' or x'01" must be specified. Otherwise, x'00' must be specified.

Fig. 3.13.3-1 shows an example for the RCU and the path parameters.

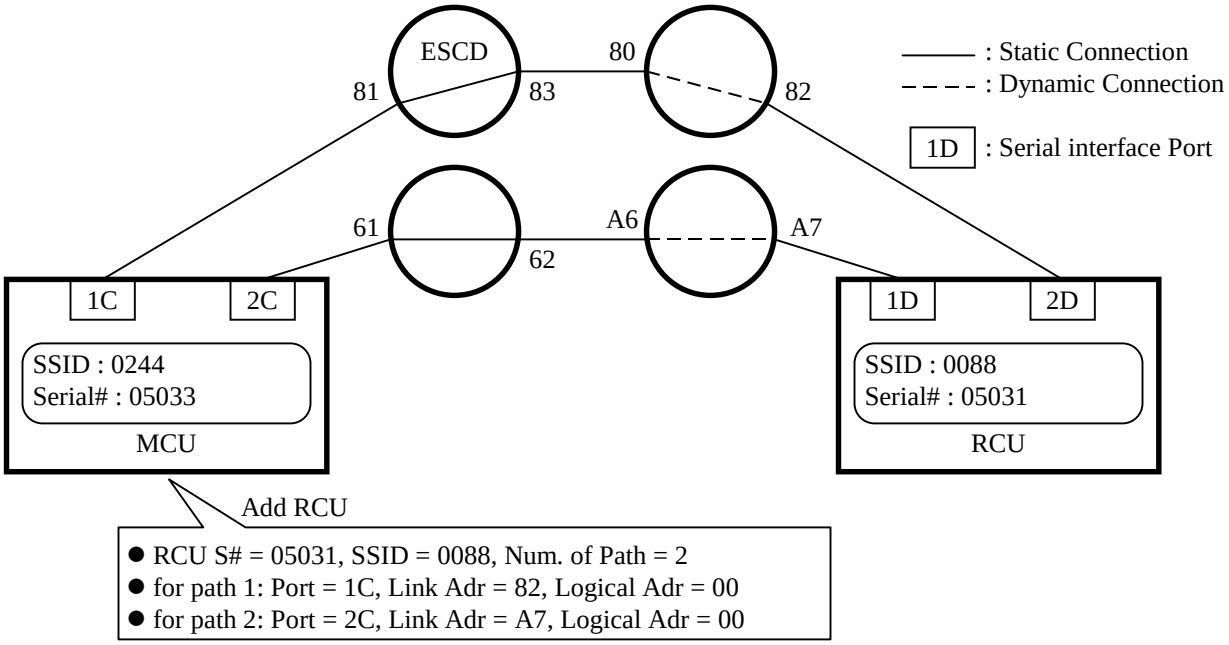


Fig. 3.13.3-1 Add RCU Operation Parameters Example

The following parameters modify the Remote Copy options which will be applied to all Remote Copy volume pairs in this subsystem. Refer to “SVP SECTION 4.5 RCU Option... Screen” for GUI image to set these parameters.

Minimum Paths	When the MCU blocks the logical path due to communication failure, if the number of remaining paths becomes less than the number specified by this parameter, the MCU will suspend all of the Remote Copy volume pairs. The default value is set to “1”. If the installation requirements prefers the subsystem I/O performance to the continuation of Remote Copy, value between “2” and the number of the established logical paths can be specified.
Maximum Initial Copy Activities	It specifies how many HRC initial copy can be simultaneously executed by the MCU. If more Remote Copy volume pairs are specified by an Add Pair operation, the MCU will execute the initial copy for as many volumes as specified by this parameter. The initial copy for other volumes is delayed until one of the initial copy is completed. This parameter can control the performance impact caused by the initial copy activity. Note: Default value of this parameter is “4”.
Incident of RCU	This parameter specifies whether the link incident record generated by the RCU is to be sent or not to the host processor connected to the MCU.
PPRC supported by HOST	If “Yes” is specified, the MCU will generate the sense information which is compatible with IBM PPRC when the HRC volume pair is suspended. If “No” is specified, the MCU will generate only service information messages. Even if the SSB (F/M=FB) is specified by the Suspend Pair Operation, the x’FB’ sense information will not be reported to the HOST.
Service SIM of Remote Copy	If “Report” is specified, the Remote Copy Service SIM will be reported to the HOST. If “Yes” is specified in PPRC supported by HOST option, DEV_SIM of HRC will not be reported. If “Not Report” is specified, the Remote Copy Service SIM reporting will be suppressed. Refer to “SIM Reference Codes Detected by the Processor for Remote Copy” in SIM-RC SECTION.

Note that these parameters will be applied to ALL RCUs registered to the MCU. If different parameters are specified, the last parameter will be applied.

(2) Edit Path Operation

An Edit Path operation makes the MCU add/delete the logical path to/from the registered RCU.

To add a logical path, the same path parameters as an Add RCU operation are required. The added logical path will be automatically used to execute the copy activities.

When deleting a logical path, pay attention to the number of remaining logical paths. If it becomes less than the number specified by "Minimum Paths", Remote Copy volume pair could be suspended.

(3) RCU Option Operation

An RCU Option operation modifies the Remote Copy options described in "3.13.3(1) Add RCU operation".

(4) Delete RCU Operation

A Delete RCU operation makes the MCU delete the specified RCU from RCU registration. All logical paths to the specified RCU will be removed.

If some volumes connected to the specified RCU are active R-VOLs, this operation will be rejected. All R-VOLs must be deleted by a Delete Pair operation before a Delete RCU operation.

(5) RCU Status Operation

An RCU Status operation makes the MCU display the status of RCU registration. It also provides the current status, time of registration and time of changing status for each logical path.

The current status of each logical path is defined as follows:

Normal	This logical path has been successfully established and can be used for the Remote Copy activities.
Initialization Failed	The link initialization procedure between the RCU is failed. It occurred due to Missing physical path connection between MCU and RCU, or connecting MCU with HOST as RCU.
Resource Shortage (RCU)	Establish Logical Path link control function has been rejected by the RCU. All logical path resources in the RCU might be used for other connections.
Serial Number Mismatch	The serial number of the control unit which is connected to this logical path does not match to the serial number specified by "RCU S#" parameter.
Invalid Port	The serial interface port specified by "Port" parameter is not in the RCP mode.

(6) Add Pair Operation

An Add Pair operation makes the MCU establish a new Remote Copy volume pair. It also provides function to modify the Remote Copy options which will be applied to the selected Remote Copy volume pair. Up to 4096 Remote Copy volume pairs can be established in one DKC410I subsystem.

To establish Remote Copy volume pair, following parameters are required:

RCU	The disk control unit which controls the R-VOL of this Remote Copy volume pair. It must be selected from RCUs which have already been registered by Add RCU operations.
R-VOL	Device number of the R-VOL.
Priority	Priority (scheduling order) of the initial copy for this volume pair. When the initial copy for one volume pair has been terminated, the MCU selects and start the initial copy for another volume pair which has the lowest value of this parameter. For the Add Pair operations, the value "1" through "256" can be specified. For establishing HRC volume pair by TSO command or DSF command for PPRC, "0" is implicitly applied to. "0" is the highest priority, "256" is the lowest, and default value for the Add Pair operation is "32". For the volume pairs to which the priority have been specified, the MCU prioritizes the volume pairs in the arrival order of the Add Pair operations or TSO/DSF commands. If the MCU are performing the initial copy for the number of volume pairs, as much as the value of "maximum initial copy activities", and accepts further Add Pair operation, the MCU does not start other initial copy until one of the copy being performed will be completed.
Operation Mode	It specifies what kind of remote copy capability should be applied to this volume pair. "Remote Dual Copy" means HRC respectively.
Initial Copy	It specifies what kind of initial copy activity should be executed for this HRC volume pair. The kind of the initial copy can be selected out of: - "Entire Volume" specifies that all cylinders excluding the alternate cylinder and the CE cylinders should be copied. - "Only Out-of-Sync Cylinders" specifies that only cylinders which have been updated during this HRC volume pair is in "suspended" state - "None" specifies that the initial copy does not need to be executed. The synchronization between volume pair must have been ensured by the operator.

Remote Copy option parameters which will be applied to this Remote Copy volume pair are as follows:

Initial Copy Pace	It specifies how many tracks should be copied at once by the initial copy. “15 Tracks” or “3 Tracks” can be specified. When “15 Tracks” is selected, elapsed time to complete the initial copy becomes shorter, however, the subsystem I/O performance during the initial copy could become worse. This parameter is valid only for HRC volume pair. Note: The default value of this parameter is “15”.
DFW to R-VOL	It specifies whether the DFW capability of the R-VOL is required or not. If “DFW required” is specified, the HRC volume pair will be suspended when the RCU can not execute the DFW due to, for example, cache failure. If the installation requirements prefers the continuation of HRC to the subsystem I/O performance, “DFW not required” is recommended. This parameter is valid only for HRC volume pair.
CFW Data	It specifies whether the records updated by CFW should be copied to the R-VOL or not. “Only M-VOL”, which means that CFW updates are not copied, is recommended because CFW data is not always necessary for disaster recovery.
M-VOL Fence Level	It specifies by what conditions the M-VOL will be fenced (the MCU will reject the write I/O operations to the M-VOL). - “R-VOL Data”: The M-VOL will be fenced when the MCU can not successfully execute the update copy. - “R-VOL Status”: The M-VOL will be fenced when the MCU can not put the R-VOL into “suspended” state. If status of the R-VOL is successfully changed to “suspended”, the subsequent write I/O operations to the M-VOL will be permitted. - “Never”: The M-VOL will never be fenced. The subsequent write I/O operations after the HRC volume pair has been suspended will be permitted. This parameter is valid only for HRC volume pairs.

Blank Sheet

(7) Delete Pair Operation

A Delete Pair operation makes the specified Remote Copy volume pair being terminated. It can be operated on either the MCU or the RCU.

- When operated on the MCU, both the M-VOL and the R-VOL will be put into the "simplex" state.
- When operated on the RCU, only the R-VOL will be put into the "simplex" state. The M-VOL will be suspended when the MCU detects this operation. To complete deleting this volume pair, the MCU requires another Delete Pair operation.

When the MCU accepts this operation and it can not communicate with the RCU, this operation will be rejected. "Delete Pair by Force" option can make the MCU complete this operation, even if it can not communicate with the RCU.

For the purpose of the recovery operation simply, "Delete All Pairs" option is provided in the delete pair operation. This option is need to use "Delete Pair by Force" option together, and specifies that the all volume pairs in the same RCU (CU Image) should be deleted. In the case of the delete operation at the RCU, specifies that the all volume pairs in the same serial number of the MCU and the same CU image of the MCU should be deleted.

(8) Suspend Pair Operation

A Suspend Pair operation makes the MCU or the RCU suspend the specified Remote Copy volume pair.

The option parameter of this operation are as follows:

SSB (F/M=FB)	The MCU and the RCU will generate sense information to notify the suspension of this volume pair to the attached host processors. This option is valid only for HRC volume pairs.
M-VOL Failure	The subsequent write I/O operations to the M-VOL will be rejected regardless of the fence level parameter. This option can be selected only when operating on the MCU. This option is valid for only HRC volume pairs.
R-VOL	For HRC volume pairs. This option can be accepted by the MCU and the RCU.

(9) Pair Option Operation

A Pair Option operation modifies the Remote Copy option parameters which has been applied to the selected Remote Copy volume pair. Refer to "3.13.3(6) Add Pair Operation" for the option parameters.

(10) Pair Status Operation

A Pair Status operation makes the MCU or the RCU display the result of the Add Pair operation or the Pair Status operation to the specified Remote Copy volume pair, along with the following information:

Initial Copy Complete	When this Remote Copy volume pair is in "pending duplex" state, it indicates how many cylinders have been successfully copied by the initial copy. When this Remote Copy volume pair is in "suspended" state, it indicates how many cylinders are currently identical between this Remote Copy volume pair. This information is provided only by the MCU.
Pair Status	It indicates the status of the M-VOL or the R-VOL. Definition of the volume states is described in "3.13.2(4) HRC Volume Pair Status".
Last Update	It indicate the time stamp when the volume pair status has been updated. Note that the time stamp value is obtained from an internal clock in the DKC410I subsystem.
Pair Established	It indicates the time stamp when the volume pair has been established by an Add Pair operation. Note that the time stamp value is obtained from an internal clock in the DKC410I subsystem.

(11) Resume Pair Operation

A Resume Pair operation restart the suspended Remote Copy volume pair. It also provides function to modify the Remote Copy options which will be applied to the selected Remote Copy volume pair.

"Out-of-Sync Cylinders" are recorded in the form of cylinder-bit-map allocated in SM (shared memory) of the DKC410I. If the MCU is powered off and the cylinder-bit-map is not retained due to the battery being discharged, the MCU resumes the initial copy as follows:

(a) For the HRC volume pair in "pending duplex" state, the initial copy is automatically resumed. The all cylinders of this volume will be copied.
(b) For the HRC volume pair in "suspended" state, the all cylinders of this volume will be copied responding to the Resume Pair operation.

(12) Port Operation

All serial interface ports in the DKC410I subsystem are initially set to the LCP mode, to which the host processor channels can be connected. At least two serial interface ports, one port from each storage cluster, must be set to the RCP mode for remote copy connections.

A Port operation makes the DKC410I subsystem change the operating mode of the specified serial interface port(s).

Before changing the operating mode from the LCP mode to the RCP mode, all channel paths to the specified port must be removed using host processor console or ESCD commands.

Before changing the operating mode from the RCP mode to the LCP mode, all RCUs which are connected through the specified port must be deleted by an Delete RCU operation.

Note: The Define Configuration & Installation operation also provide the function to set the operating mode of each serial interface mode.

3.13.4 Managing HRC Environment

(1) Setting Up HRC Volume Pairs

(a) Sequence of Operations

Sequence of operations to establish the HRC volume pairs are shown below.

Table 3.13.4-1 Operations to Set Up HRC Volume Pairs

Step		Operation	
		SVP*	Others
1	Set appropriate serial interface ports to the RCP mode.	Port	
2	Establish logical paths between the DKC410I HRC subsystems	Add RCU	Before this step, remote copy connections must be established between DKC410I subsystems.
3	Ensure that the R-VOLs are offline from host processors		If necessary, perform the following system command. <In case of MVS system> • VARY OFFLINE <In case of VM system> • VARY OFFLINE from guest OS • VARY PATH OFFLINE from VM
4	Establish HRC volume pairs.	Add Pair	

* : Operations form the SVP/remote console attached to the MCU.

Several volume pairs can be specified within one Add Pair Operation. After completing an Add Pair operation, another Add Pair operation can be executed to establish another HRC volume pairs.

Be sure to vary the R-VOLs offline from the attached host processors before executing the Add Pair operation. The RCU will reject the write I/O operations to the R-VOLs once the Add RCU operation has been accepted.

(b) Considering HRC Parameters

Setting of the “fence level” parameter to the Add Pair operation and the “PPRC supported by host” and “Service SIM of Remote Copy” option to the Add RCU operation depends on your disaster recovery planning. Refer to “3.13.5(1) Preparing for Disaster Recovery” for these parameters.

Setting of the “CFW data” and “DFW to R-VOL” parameters to the Add Pair operation and the “minimum paths” parameter to the Add RCU operation depends on your performance requirement to the DKC410I subsystem at the primary site. Refer to “3.13.3(6) Add Pair operation” and “3.13.3(1) Add RCU operation” for these parameters.

Setting of the “maximum initial copy activities” parameter to the Add RCU operation and the “priority” and the “initial copy pace” parameters can control performance effect from the initial copy activities. Refer to “3.13.4(1)(c) Controlling Initial Copy Activities” for more detailed description.

Refer to “3.13.3(1) Add RCU operation and “3.13.3(6) Add Pair operation” for other parameters.

(c) Controlling Initial Copy Activities

To control performance effect from the initial copy activities, the “maximum initial copy activities” parameter and the “priority” and the “copy pace” parameters can be specified:

- The “maximum initial copy activities” parameter controls the number of volumes for which the initial copy are concurrently executed;
- The “priority” parameter specifies the executing order of the migration copy on volume pair basis;
- The “copy pace” parameter specifies how many tracks should be copied by each initial copy activity.

Refer to the following example for the “maximum initial copy activities” and the “priority” parameters.

Example
Conditions:
- The Add Pair operation specifies that devices 00~05 should be M-VOLs.
- “Maximum initial copy activities” is set to “4” (this is the default value).
- “Priority” parameters for devices 00~05 are set to “3”, ”5”, ”5”, “1”, “4”, and “2” respectively.

Under the above conditions, the MCU will performs the initial copy:
- for devices 00, 03, 04 and 05 immediately.
- for device 01 when one of the initial copy has been terminated.
- for device 02 when the initial copy for the second device has been terminated.

(2) Suspending and Resuming the HRC Volume Pairs

This section describes the operations to suspend or resume the HRC volume pair, which are necessary for the following sections in this chapter.

The Suspend Pair operation with the “R-VOL” option parameters can suspend the specified HRC volume pairs while the M-VOLs are still accessed from the attached host processors. The “SSB” option should not be selected to prevent the sense information from being generated.

To resume the suspended HRC volumes pairs, the Resume Pair operation must be executed.

Refer to “3.13.3(8) Suspend Pair Operation” and “3.13.3(6) Add Pair Operation” for more detailed description.

(3) Managing Power On/Off of HRC Components

(a) Cutting Power to HRC component

Cutting power to the RCU or the ESCDs on the remote copy connections, or other equivalent events which make the MCU unable to communicate with the RCU should be controlled in order not to affect the remote copy activities. If the MCU detects these events when it intends to communicate with the RCU, it would suspend all HRC volume pairs.

To avoid this problem, the applications on the primary host processors must be terminated or all HRC volume pairs must be suspended or terminated, before performing these events.

Refer to “3.13.4(2) Suspending and Resuming the HRC Volume Pairs” for the operations to suspend and resume the HRC volume pairs.

(b) Power Control Interface at the Secondary Site

In the secondary site, It is not recommended to use the power control interface which remotely cuts the power to the RCU or the ESCD on the remote copy connections in order to avoid the situation described in “3.13.4(3)(a) Cutting Power to HRC components”.

(c) Power-on-sequence

The RCU and the ESCDs on the remote copy connections must become operable before the MCU accepts to first write I/O operation to the M-VOLs.

After the power-on-reset sequence of the MCU, It communicates with the RCU in order to confirm the status of the R-VOLs. If it is not possible, the MCU retries the confirmation until it is successfully completed or the MCU accepts the first write I/O operations to the M-VOLs.

If the MCU accepts the first write I/O operation before completing the confirmation, the MCU will suspend the HRC volume pair. This situation is critical because the status of the R-VOL can not be changed, that is, remains “duplex” state.

(4)Executing ICKDSF to HRC Volume Pairs

The updates by the channel programs which specify “diagnostic authorization” or “device support authorization” are not reflected to the R-VOL. ICKDSF commands which issue the write I/O operations to the M-VOL must be controlled. The HRC volume pairs must be suspended or terminated before performing ICKDSF commands.

Refer to “3.13.4(2) Suspending and Resuming the HRC Volume Pairs” for the operations to suspend and resume the HRC volume pairs.

3.13.5 HRC Error Recovery

(1) Preparing for Disaster Recovery

(a) Considering Fence Level Parameter

Table 3.13.5-1 shows how the fence level parameter of the Add Pair operation has an effect on the write I/O operations to the M-VOL after the HRC volume pair has been suspended. You should select one of the fence level considering the "degree of the currency" of the R-VOL required by your disaster recovery planning. The SVP or remote console, which is connected to either the MCU or the RCU, can display the fence level parameter which has been set to the HRC volume pairs.

Table 3.13.5-1 Effect of the Fence Level Parameter

Failure		Subsequent write I/O operations to the M-VOL will be ...		
		"Data"	"Status"	"Never"
1)	The update copy has failed,	rejected*	-	-
2)	(1) & however the status of the R-VOL could have been successfully changed to "suspended" state.	rejected*	accepted	accepted
3)	(1) & furthermore the status of the R-VOL could not have been changed to "suspended" state.	rejected*	rejected*	accepted

*) Sense bytes includes "command reject" and x'0F' of format/message.

Note: "Data" and "Status" has an effect when an HRC volume pair of "duplex" state is suspended. For HRC volume pairs which are in "pending duplex" state, subsequent write I/O operations will not be rejected regardless of Fence Level parameter.

1) Fence Level = "Data"

The data of the R-VOL is always identical to the M-VOL if once the HRC volume pair has been successfully synchronized. You can reduce the time to analyze whether the R-VOL is current or not in your disaster recovery procedures.

However, this parameter will make the M-VOL not accessible from your applications whenever the HRC copy activity has failed. Therefore you should specify this parameter to the most critical volumes for your disaster recovery planning.

Most of the database system supports duplexing the critical files, for example log files of DB2, for its file recovering capability. It is recommended to locate the duplexed files on the volumes in the physically separated DKC410I subsystems, and establish HRC volume pairs for each volumes by using physically separated remote copy connections.

Note 1: If the failure has occurred before completing the initial copy, the R-VOL can not be used for disaster recovery because the data of the R-VOL is not fully consistent yet. You can became aware of this situation with referring status of the R-VOL in your disaster recovery procedures. Refer to "3.13.5(2)(b) Analyzing the Currency of R-VOLs" for more detailed description.

Note 2: Only the difference between the HRC volume pair must be the last update from the host processor. HRC is a synchronous remote copy. The MCU reports a "unit check" if it detects the failure on the write I/O operation including the update copy to the R-VOL. Therefore, the operating system and the application program does not regard the last (failed) I/O operation as successfully completed.

This parameter is functionally equivalent to "CRIT=YES" parameter for IBM PPRC

2) Fence Level = Never

The subsequent write I/O operations to the M-VOL will be accepted even if the HRC volume pair has been suspended. Therefore the contents of the R-VOL can become "older" (behind the currency of corresponding M-VOL) if the application program continue updating the M-VOL. Furthermore, the status of the R-VOL which will be obtained from the RCU can not be in a "suspended" state.

To use this parameter, your disaster recovery planning must satisfy the following requirements:

- The currency of the R-VOL should be decided by referring the error message which might have been transferred through the error reporting communications or analyzing the R-VOL itself with other files which are confirmed to be current.
- The data of the R-VOL should be recovered by using other files which are ensured to be current.

This parameter is functionally equivalent to "CRIT=NO" parameter for IBM PPRC

3) Fence Level = Status

The level of this parameter is between "Data" and "Never". Only when the status of the R-VOL can be ensured, the subsequent write I/O operations to the M-VOL will be permitted. Therefore the disaster recovery procedure of deciding the currency of the R-VOL can be reduced.

(b) Transferring the Sense Information through Error Reporting Communications

When the HRC volume pair is suspended, the MCU generates the sense information which notifies the host processor of the failure. This will help in deciding the currency of the R-VOLs in the disaster recovery procedures by transferring the sense information, or the system console message caused by the sense information, with the system time stamp information.

The sense information can be selected out of:

- x'FB' of format/message. The sense information is compatible with IBM PPRC and result on a corresponding system console message, for example IEA491E of MVS, if the operating system supports it.
- service information message whose reference code means that the HRC volume pair has been suspended.

Note: The first version of HRC is not completely certified under the operating system which does not support IBM PPRC. Therefore the x'FB' sense information must be selected.

The error reporting communications are essential if you use the fence level of "Status" or "Never".

(c) File Recovery Procedures Depending on Installations

HRC is a synchronous remote copy. All updates to the M-VOLs are copied to their R-VOLs before completing each channel program of the write I/O operations. When the HRC volume pairs have been suspended or the MCU has become inoperable due to a disaster, therefore, many data "in progress" could remain in the R-VOLs. That is, some data set might be still opened, or some transactions might not be committed yet. All breakdown cases should be previously considered.

Therefore, even if you have selected the fence level of "Data" for all HRC volume pairs, you should establish the file or volume recovery procedures. The situation which should be assumed is similar to that where the volumes have became not accessible due to the disk controller failure in non-remote copied environment.

If you use the fence level of "Status" or "Never", the suspended R-VOLs could become "ancient" compared to other volumes. This situation might cause a data inconsistency problem among several volumes.
You should prepare, in your disaster recovery, for recovering some files or some volumes which have become "ancient" by using:

- files for file recovery, for example DB2 log files, which have been confirmed to be current. To ensure the currency of these files, it is recommended to use the fence level of "Data" for these critical volumes.
- the sense information with the system time stamp which have been transferred through the error reporting communications.
- full consistent file or volume backups, if the sense information and the system time stamp can not be used.

(d) CSUSPEND/QUIESCE for IBM PPRC

PPRC recommends to customers to establish their disaster recovery planning where the CSUSPEND/QUIESCE TSO command is programmed to be issued responding to the IEA491E system console messages. This procedure intentionally suspend the remaining volume pairs when some volume pairs have been suspended due to a disaster.

The CSUSPEND/QUIESCE TSO command will be supported as the enhancement to HRC.

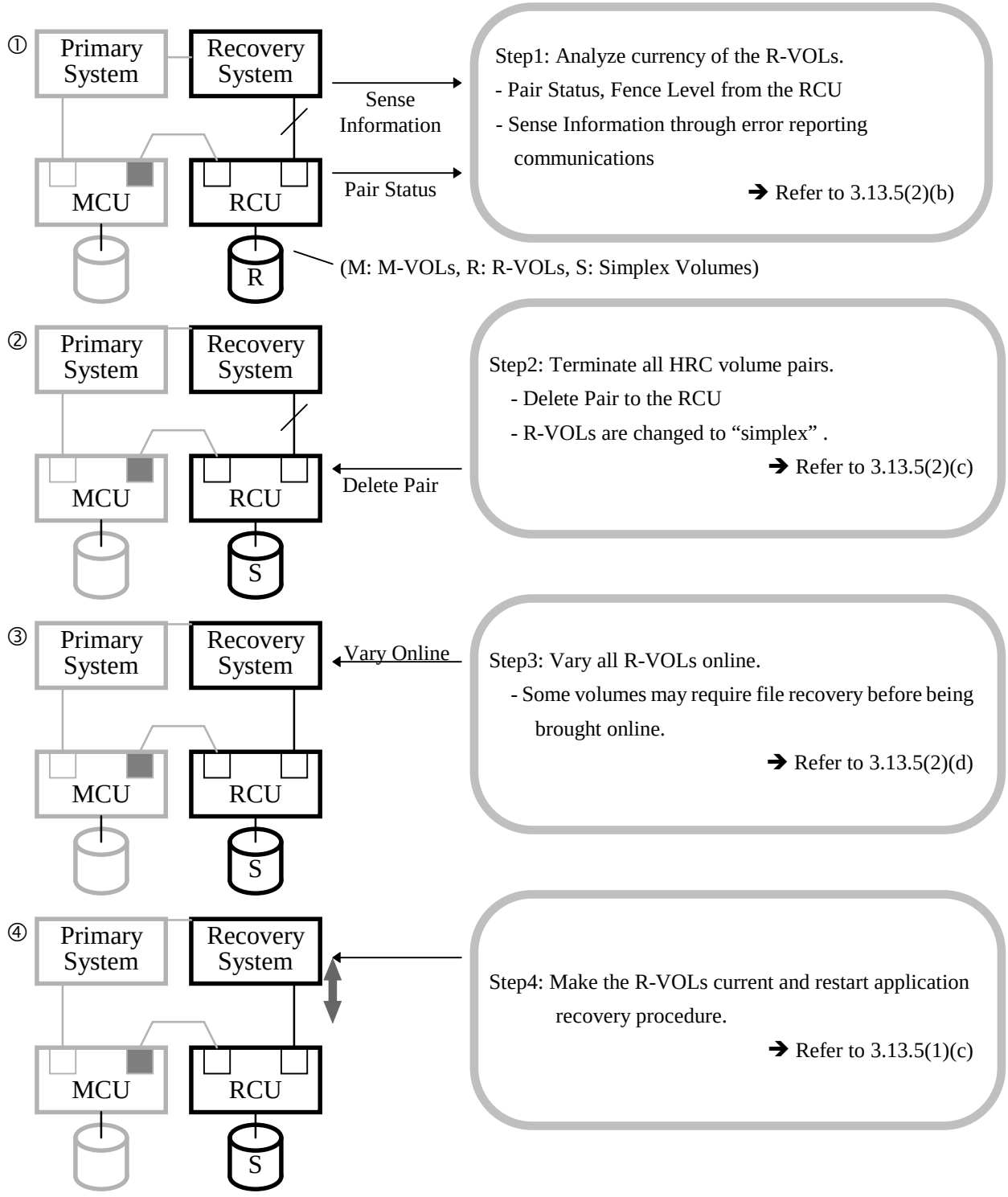
(e) All SIM of the HRC clear option

For the purpose of the restrain to report the HRC SIM which should be generated at the disaster and the recovery operations during the OS IPL, “Clear SIM” button is provided on HRC Main Control Screen at the SVP and the RMC. This function is able to use for the disaster recovery operation, and specifies that the all Remote Copy SIM in the subsystem should be deleted.

(2) Disaster Recovery Procedures - Switching to the Recovery System

(a) Summary

⓪ Primary system and MCU becomes inoperable due to disaster.



(b) Analyzing the Currency of R-VOLs (Step 1)

1) Analyzing Status of the R-VOLs and Fence Level Parameter

Table 3.13.5-2 Currency of the R-VOLs

	Fence Level for this HRC volume pair		
Status of R-VOL	Data	Status	Never
Simplex	To be confirmed	To be confirmed	To be confirmed
Pending Duplex	Inconsistent	Inconsistent	Inconsistent
Duplex	Current	Current	To be analyzed
Suspended (Initial Copy Failed)	Inconsistent	Inconsistent	Inconsistent
Suspended (by other reason)	Current	Suspected	Suspected

Table 3.13.5-2 shows how to analyze the currency of the R-VOLs referring the status of the R-VOLs and the fence level parameter which have been specified when establishing the HRC volume pairs. The status of the R-VOLs must be obtained from the RCU in your disaster recovery procedures. The fence level parameter must be previously field since it cannot be obtained From RCU.

Meaning of the results or further actions shown in each column of table 3.13.5-2 are as follows:

To be confirmed This volume does not belong to any HRC volume pair. If you have certainly established the HRC volume pair for this volume and you have never deleted it, you should regard this volume as inconsistent.

Inconsistent The data on this volume is inconsistent because not all cylinders have successfully been copied to this volume yet. You can not use this volume for the applications unless this volume is initialized (or successfully copied from the M-VOL at later time).

Current The data on this volume is completely synchronized with the corresponding M-VOL.

To be analyzed The currency on this volume can not determined. To determine the currency, further analysis described in (2) of this section should be performed.

Suspected The data on this volume must be “older”, behind the currency of corresponding M-VOL. You should restore the consistency of this volume at least, and the currency of this volume if required. The system time information which might have been transferred through the error reporting communications or time of suspension obtained from the Pair Status operation will help you decide the last time when this volume was current.

2) Further Analysis by Referring to Other Information

The M-VOLs, to which the fence level parameter has been set to “Never”, will accepts the subsequent write I/O operations regardless of the result of communication to change the R-VOL into the “suspended” state. Therefore, the status of the R-VOL should be analyzed by referring to the following information:

- The sense information through the error reporting communications. If the sense information which denote the suspension of this volume is found, you can return to table 3.13.5-2 with assumption of the “suspended” state.
- The status of the M-VOL obtained from the MCU, if possible. You should return to table 3.13.5-2 with assumption of the same status as the M-VOL and fence level of “Status”.
- The other related files, for example DB2 log files, which have been confirmed to be current.

(c) Terminating HRC Volume Pairs (Step 2)

The "Delete Pair" operation to the RCU terminates the specified HRC volume pairs. These R-VOLs will be changed to “simplex” state. Specified “Delete Pair by Force” option and “Delete All Pairs” option the all volume pairs in the same serial number of the MCU and the same CU image of the MCU should be deleted. Refer to “3.13.3(7) Delete Pair Operation”.

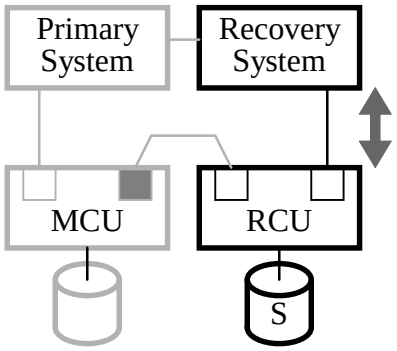
(d) Vary all R-VOLs online (Step 3)

In the case of the OS IPL, execute the “Clear SIM” operation at the SVP or the RMC before OS IPL. Refer to “2.23.19 Clear reported SIM” Operation.

(3) Disaster Recovery Procedures - Returning to the Primary Site

(a) Summary

⓪



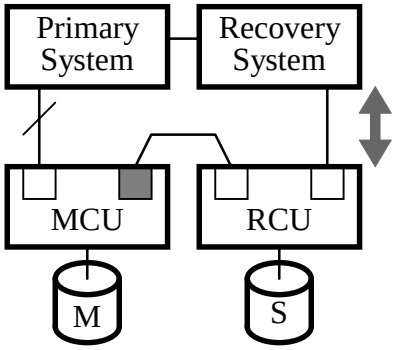
Applications are working at the recovery system.

①

Step1: Make the components in the primary system operable.

Repair Actions

Repair Actions



②

Step2: Terminate the HRC settings remaining in the MCU.

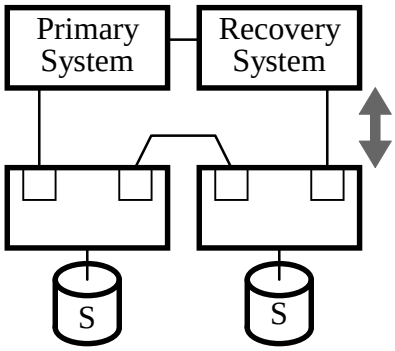
(1) Delete Pair: Delete all HRC volume pairs.

(2) Delete RCU

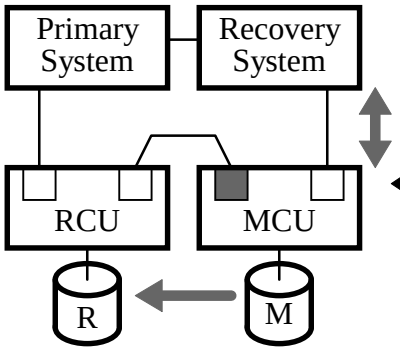
(3) Port: Change from RCP to LCP mode.

➔ Refer to 3.13.5(3)(b)

Delete Pair
Delete RCU
Port



③



Port
Add RCU
Add Pair

Step3: Establish HRC with the reverse direction

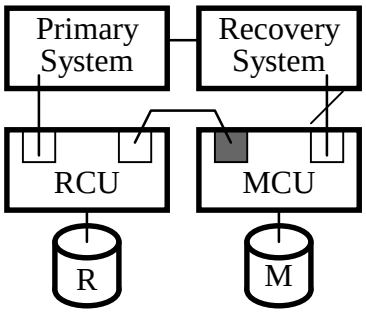
(1) Port: Change from LCP to RCP mode.

(2) Add RCU

(3) Add Pair

➔ Refer to 3.13.5(3)(c)

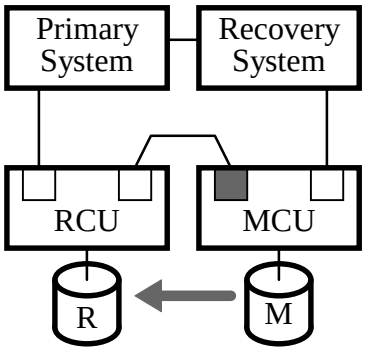
④



Vary Offline

Step4: Halt related applications and vary all M-VOLs offline from the recovery system.

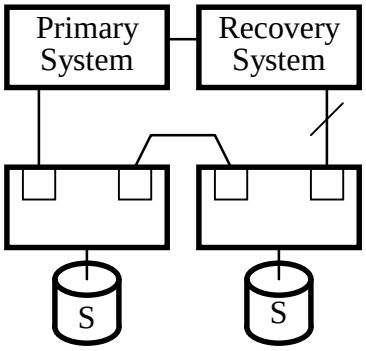
⑤



Pair Status

Step5: Confirm all HRC volume pairs become "duplex" state.

⑥



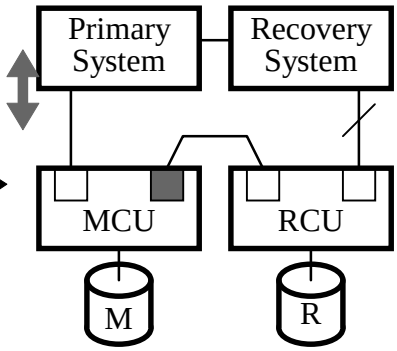
Delete Pair
Delete RCU
Port

Step6: Terminate all HRC settings.
(1) Delete Pair: Delete all HRC volume pairs.
(2) Delete RCU
(3) Port: Change from RCP to LCP mode.
➔ Refer to 3.13.5(3)(d)

⑦

Step7: Establish HRC pair with the original direction and start applications.
(1) Port: Change from LCP to RCP mode.
(2) Add RCU
(3) Add Pair
➔ Refer to 3.13.5(3)(e)

Port
Add RCU
Add Pair



(b) Terminating the HRC Settings Remaining in the MCU (Step2)

After the DKC410I subsystem becomes operable, the remaining registration of the HRC volume pairs and the RCU should be deleted by performing the Delete Pair operation and Delete RCU operation respectively.

To complete the Delete Pair operation, the “delete pair unconditionally” option is required because the original R-VOLs do not belong presently to any HRC volume pairs. The MCU will change the specified M-VOLs into “simplex” state without checking the current status of the corresponding R-VOLs.

Specified “Delete Pair by Force” option and “Delete All Pairs” option the all volume pairs in the same RCU should be deleted.

Note that the status of M-VOLs may be “Suspended (Delete Pair to RCU)” because of Delete Pair operation issued to the RCU in step 2 of “3.13.5(2) Disaster Recovery Procedures - Switching to the Recovery System”. It is normal condition in this situation.

Before performing the Delete RCU operation, all HRC volume pairs must be deleted.

If you want to use same remote copy connections for step 3, the serial interface ports which have been set to the RCP mode should be changed to the LCP mode by the Port operation.

(c) Establish HRC with the Reverse Direction (Step3)

The HRC volume pair should be established with the reverse direction to synchronize the original M-VOLs with the original R-VOLs. The procedures for this step are same as those described in “3.13.4(1) Setting Up HRC Volume Pairs”. Note that the DKC410I subsystems in the original primary site and the recovery site are treated as the RCUs/R-VOLs and the MCUs/M-VOLs respectively.

Do not select “only out-of-sync cylinders” or “none” parameter to the Add Pair operations. The volumes in the original primary site are now behind the volumes in the recovery site. Furthermore the updates to the volumes in the recovery site have not been accumulated in cylinder bit map.

(d) Terminate Applications and HRC Settings at the Recovery Site (Step 4~6)

HRC settings with the reverse direction must be deleted after halting the applications in the recovery site (step 4) and confirming that all HRC volume pairs are in “duplex” state (step 5).

Specified “Delete Pair by Force” option and “Delete All Pairs” option the all volume pairs in the same RCU should be deleted.

If you want to use same remote copy connections for step 7, the serial interface ports which have been set to the RCP mode should be changed to the LCP mode by the Port operation.

(e) Establish HRC Pair with the Original Direction and Start Applications (Step 7)

The HRC volume pair should be established with the original direction to synchronize the original M-VOLs with the original R-VOLs. The procedures for this step are same as those described in “3.13.4(1) Setting Up HRC Volume Pairs”.

Do not select “only out-of-sync cylinders” or “none” parameter to the Add Pair operations. The volumes in the original primary site are now behind the volumes in the recovery site. Furthermore the updates to the volumes in the recovery site have not been accumulated in cylinder bit map.

Blank Sheet

Blank Sheet

Blank Sheet

Blank Sheet

Blank Sheet

Blank Sheet

Blank Sheet

Blank Sheet

3.13.6 Guard Concept for Maintenance Operation in HRC

(1) In this maintenance manual.....
SVP requires Password input for each setting operation. In each paragraph of this maintenance manual, we does not describe the password operation. If SVP requires to input password, please call TSC what the password is.

(2) Basic concept for password.
HRC applications which are installed in the SVP have guard process especially if the operator selects a button to execute HRC setting (Edit path, Add Pair, Delete Pair..). SVP requires the password input. These operations must be concerned with the direct access to customer data. So it is very dangerous operation for the system so that SVP needs guard operation.

(3) summary table
We arranged these concept as the following table.

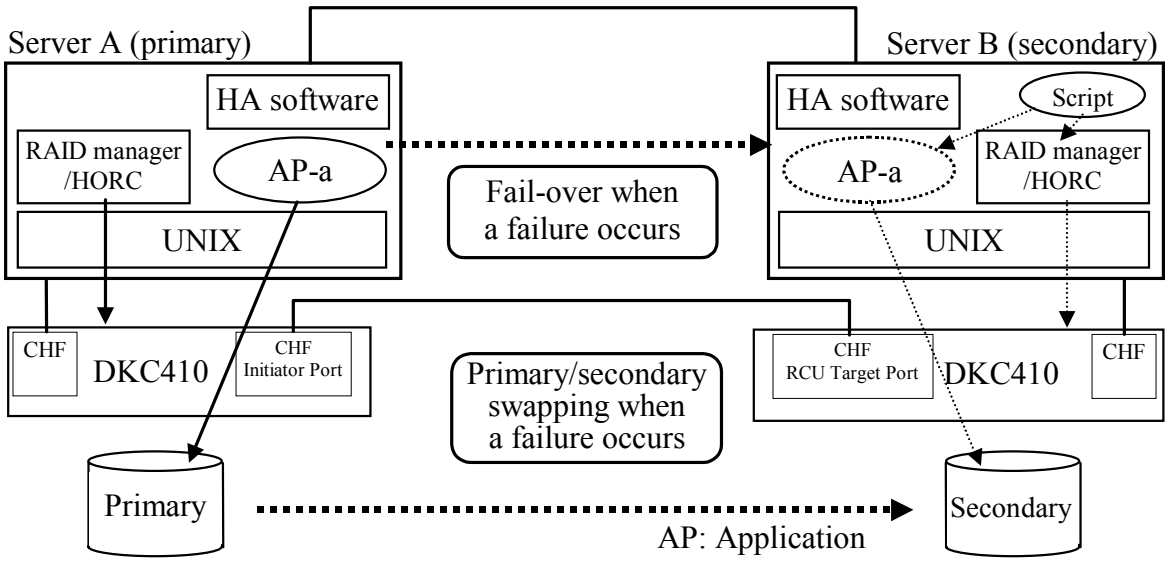
#	Button (Function)	Operation which needs a password input
1	Add RCU	Y
2	Edit Path	Y
3	Delete RCU	Y
4	RCU Option	Y
5	RCU Status	N
6	Add Pair	Y
7	Delete Pair	Y
8	Pair Option	Y
9	Pair Status	N
10	Suspend Pair	Y
11	User	— (*1)
12	Usage	N
13	Connect	— (*1)
14	Port	* (*2)
15	Renew	N
16	Exit	N
	Explanatory notes	Y : It's necessary to input password. The all password are fixed N : Not necessary to input password * : It's necessary to input password when the operator uses some setting function. — : It is not supported(Displayed as gray color),

(*1) "User" and "Connect" functions are supported in the remote console application only. In the SVP, these button is displayed as gray to guard.
(*2) SVP does not require the password input at the first button selection. If the user selects port changing function (LCP-RCP), SVP requires the password input.

3.13.7 Fibre Channel Connection

An explanation is done about the Fibre Channel connection which is connection classification between CUs in this chapter.
Therefore, the part where description overlaps with Serial interface (ESCON/ACONARC) connection is omitted.

3.13.7.1 Component

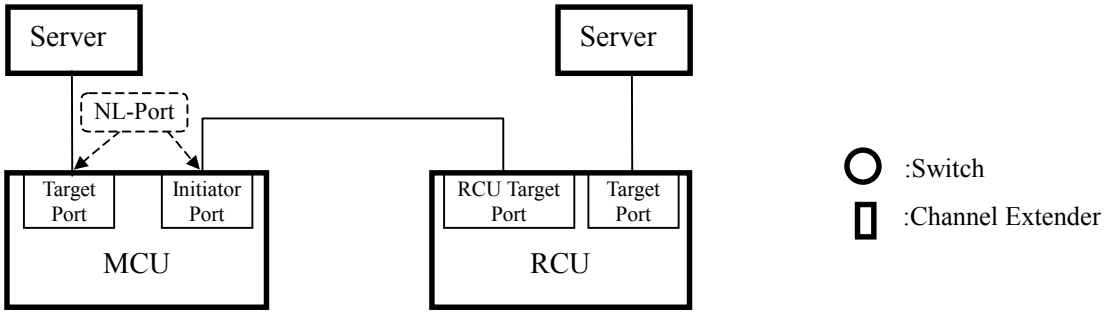


(1) Initiator Port
An **Initiator Port** (remote control port) is a Fibre Channel interface port to which an RCU is connected. Any Fibre Channel interface port of the DKC410I subsystems can be configured as an Initiator Port.
But, as for the channel port of the host computer, it can't communicate. A pass from the host computer must be connected with other Fibre Channel interface ports.

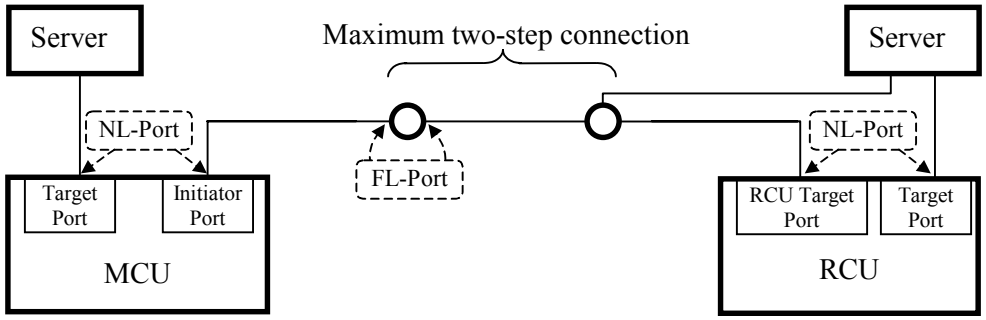
(2) RCU Target Port
An **RCU Target Port** (remote control port) is a Fibre Channel interface port to which an MCU is connected. Any Fibre Channel interface port of the DKC410I subsystems can be configured as an RCU Target Port.
It can be connected with the channel of the host computer by the Fibre Channel switch.

(3) Connection composition.
A connection composition example is shown in the following.

(a) Direct connection



(b) Switch connection



(c) Extender connection

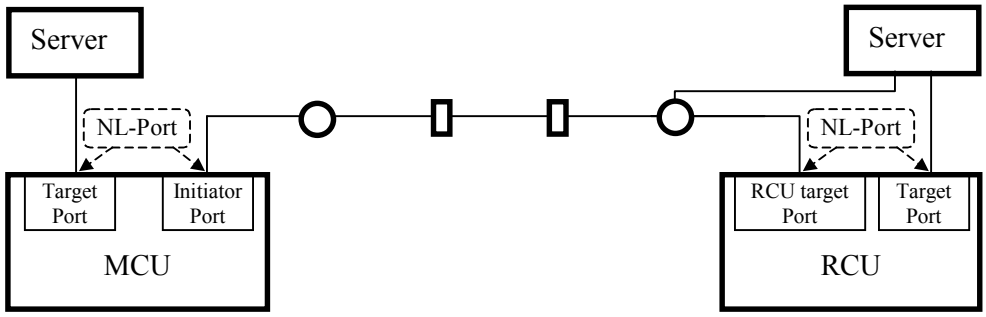


Fig. 3.13.7.1-2 A connection composition example is shown in the following

3.13.7.2 Hardware Requirements

(1) Fibre Channel Port adapter requirement
Fibre Channel Port adapter(CHF) which connects MCU and RCU in Fibre Remote Copy isn't acceptable a thing except for WP411-B is used.

(2) Distance between MCU and RCU
You must connect MCU and RCU with Optical Fibre cable.
With ShortWave(Optical Multi Mode), an extreme head, 500m. An extreme head is 10km with LongWave(Optical Single Mode).
By connecting Switch, with ShortWave, an extreme head, 1.5km. An extreme head is 30km with LongWave.
But, Switch is a maximum two-step connection

A distance between MCU and RCU isn't restricted by connecting Channel Extender.

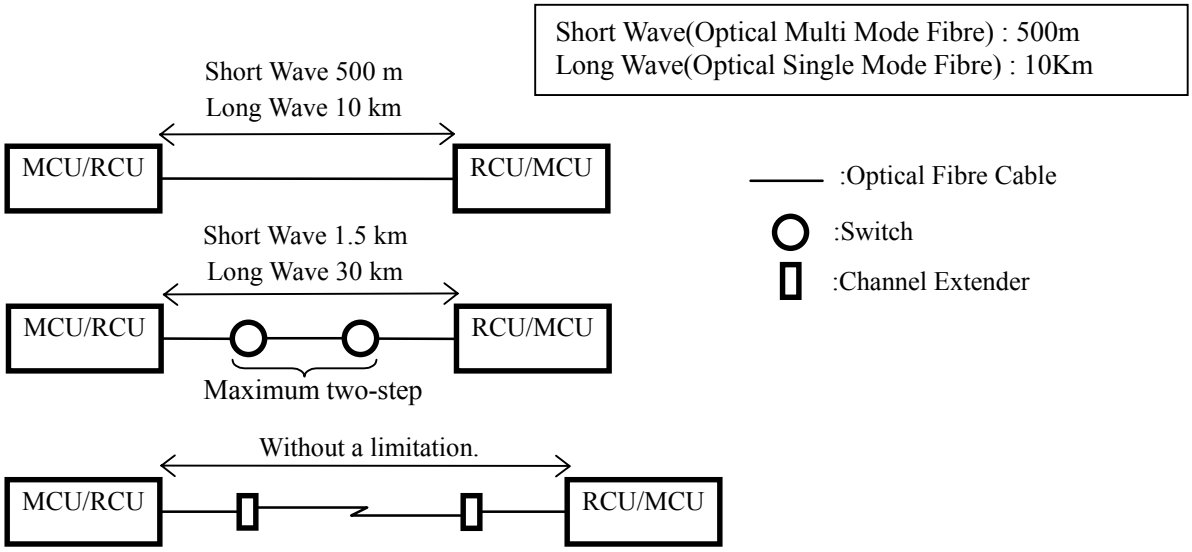


Fig. 3.13.7.2-1 Distance between Disk subsystem

3.13.7.3 Port Setting

You must set up the connection port of MCU and RCU prior to the pass formation in Initiator port or the RCU target port from the usual target port.
Refer to “Install section 6.5 (1) RCP/Initiator/RCU-Target Port Setting” for the establishment.
And, Port topology of Initiator port and the RCU target port must be setup of the following.

- Direct connection : Fabric = OFF,FC-AL
- A connection via Switch : Fabric = ON, FC-AL or Point to point

And, refer to “Install section 6.3.2.6 SCSI I/F xxxx” for the establishment.

(1) About the changing unit of a setup of a port
The standard mode and the high-speed mode exist in CHF P/K. Therefore, the unit which can be set up to the usual target port/Initiator port /RCU target port is different.
And, as for the high-speed mode, port classification mixture inside P/K can’t be set up.
A port establishment switching unit is collected in the following.

P/K Type	4 Port	
Setting Mode	Standard Mode	High-Speed Mode
Changing Unit	2 Port (Port#0/1 or Port#2/3)	2 Port (Port#0/1)
In P/K, port classification mixture.	It is possible.	Impossibility.

3.13.7.4 Add RCU Operations

The following parameter is necessary to register RCU as a Fibre Channel connection.
Refer to “Install section 6.5 HORC Pair Establish” for the GUI image of this operation.

Port Type	Serial: Serial channel interface is used for the connection of MCU and RCU. Fibre: Fiber channel interface is used for the connection of MCU and RCU.
Controller ID	Don’t change this option. Default is 02 fixation.
RIO MIH Time	Data transfer complete waiting time to RCU from MCU. Usual: 15[Sec]. Avail range: 10[Sec] ~ 100[Sec]
Path Blockade Watch	When the connection of MCU and RCU Fibre Channel, time to detect it as a pass difficulty (link going down) on the MCU side is set up. Usual: 40[Sec]. Avail range: 0[Sec] ~ 45[Sec]
MCU Port	Initiator port of the DKC410-I subsystem which a logic pass is set up to. You must set up a Fibre Channel interface port in Initiator port before this operation.
RCU Port	The Fibre Channel interface port of the place of the connection. You must specify a RCU target port.

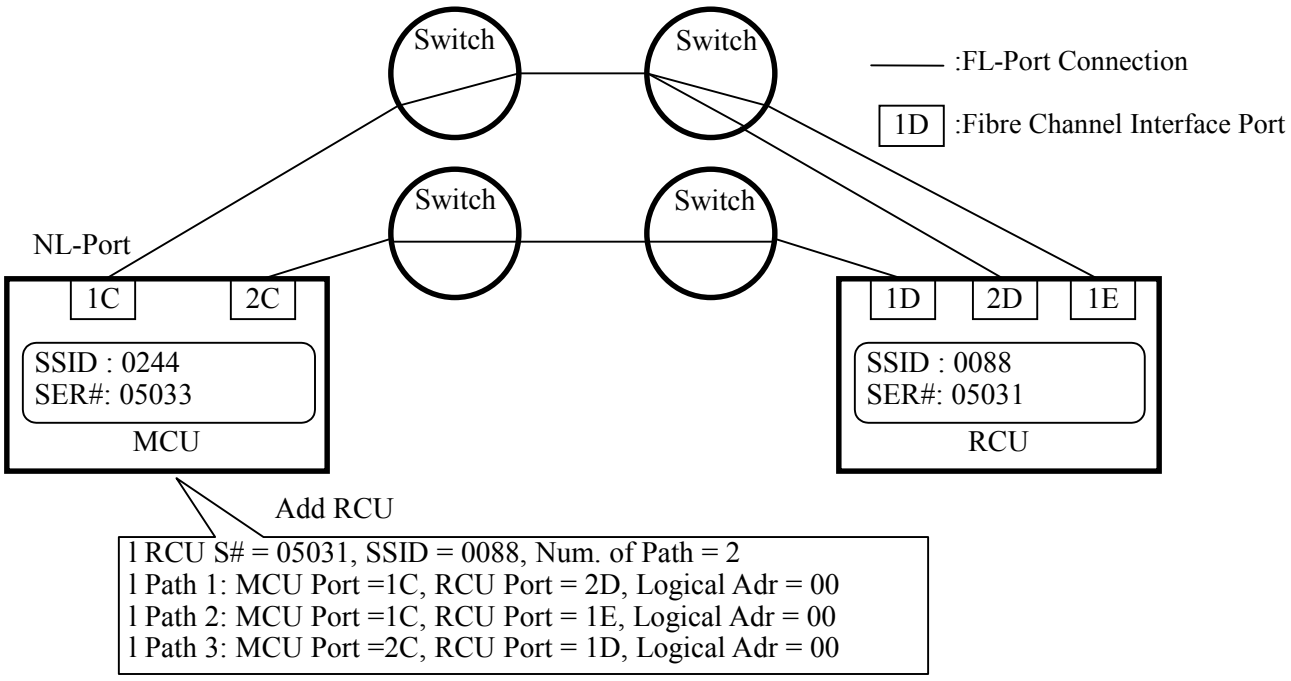


Fig. 3.13.7.4-1 Add RCU Operation

3.13.7.5 Registration the Consistency Group

When HORCA is used, the following parameter is necessary prior to the pair established.
Refer to “Install section 6.5 (3) C/T group assingnation” for the GUI image of this operation.

Port Type
- Serial: When Group concerned is used under the Serial channel interface connection.
- Fibre: When Group concerned is used under the Fibre channel interface connection.

This option is effective only with MCU.

3.13.7.6 Mixing Fibre Connection and ESCON Connection

If the P-VOL that is assigned by TrueCopy with the Fibre connection and the P-VOL or M-VOL that is assigned by TrueCopy or TC390 with the serial interface (ESCON) connection exist in the same subsystem, the Hitachi Lightning 9900™ has the restrictions.

If the status of Fibre TrueCopy pair is COPY or PAIR when there is no serial interface (ESCON) connected pairs, or all serial interface connected pairs are suspended, you cannot create a pair connected with a serial interface (ESCON) using TrueCopy or TC390, or recover a pair connected with a serial interface (ESCON) from the suspended status.

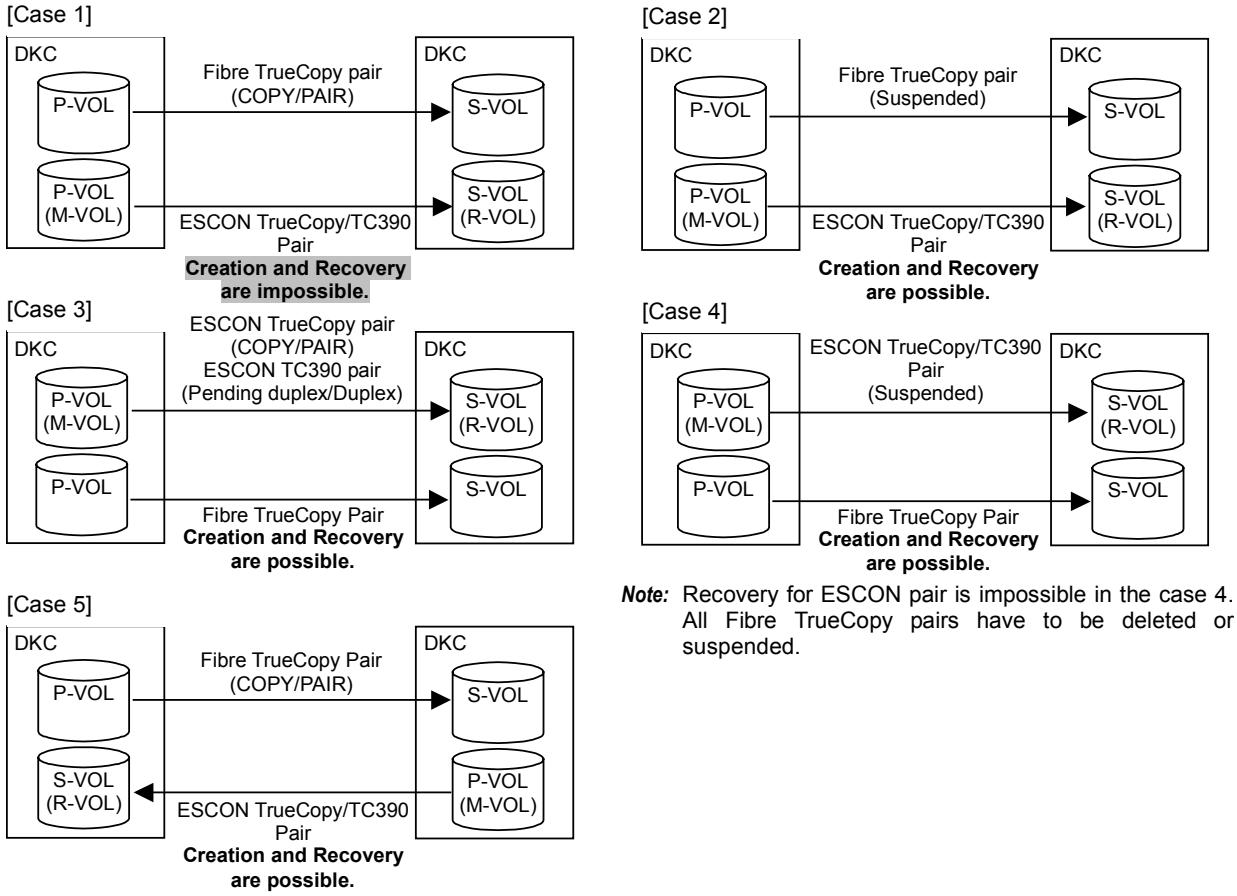
To create or recover a pair connected via ESCON, you have to put all Fibre TrueCopy pairs into the suspended status. After that, you can create and recover a pair that is connected via ESCON using TrueCopy or TC390.
Table 3.13.7.6-1 shows whether you can create and recover a pair or not in these conditions.

Table 3.13.7.6-1 Possible or Impossible to Create/Recover a Pair

TrueCopy Fibre Connection Pair Condition (P-VOL)	Pair Creation/Recovery			
	Fibre TrueCopy pair (P-VOL)	Fibre TC390 pair (M-VOL)	ESCON TrueCopy pair (P-VOL)	ESCON TC390 pair (M-VOL)
No pairs	Possible	Possible	Possible	Possible
All pairs being suspended	Possible	Possible	Possible	Possible
Including Copy or Pair status	Possible	Possible	Impossible	Impossible

Note 1: P-VOL connected via Fibre with TrueCopy and S-VOL connected via ESCON, or S-VOL connected via Fibre with TrueCopy and P-VOL connected via ESCON can exist in the same subsystem.

Note 2: When P-VOL or M-VOL connected via ESCON with TrueCopy or TC390 exists (in any pair conditions), you can create and recover a pair using P-VOL connected via Fibre with TrueCopy. Notice that you cannot recover a pair connected via ESCON if you create a pair via Fibre with TrueCopy when all pairs connected via ESCON has been suspended.



Note: Recovery for ESCON pair is impossible in the case 4. All Fibre TrueCopy pairs have to be deleted or suspended.

Fig. 3.13.7.6-1 Possible or Impossible to Create/Recover a Pair

3.14 HMRCF & HOMRCF

3.14.1 Overview

(1) Main object
1) Reduce Backup time.
2) Easy testing with the data which applications are actually using before system upgrade.

(2) Function Outline
1) Making a duplicated volume.
2) There is no conflict on volume, because the duplicated volume is on another physical storage.
3) Three destination volumes can be with one master volume.
Those three pairs can split independently each other.
4) HMRCF (Hitachi Multiple RAID Coupling Feature) can be controlled by PPRC Command interface.
5) HOMRCF (Hitachi Open Multiple RAID Coupling Feature) can be controlled from RAID Manager.

Table 3.14.1 Outline of HMRCF & HOMRCF

No.	Items	Specification
1	Coupling object	One logical volume (LDEV)
2	Support emulation type of LDEV	HMRCF : 3390-1/2/3/3R/3A/3B/3C/9/L, 3380-J/E/K/KA/KB/KC and CVSs of them. HOMRCF : OPEN-3/8/9/K/E/L/M and CVSs and LUSEs of them.
3	Requirement for create a pair	(1) Pair LDEVs have to be a same track format and same capacity. (2) Pair LDEVs have to exist in a same subsystem. (3) It is not possible to share a destination volume at same time.
4	Support of CVS (Customized Volume Size)	HMRCF : Supported HOMRCF : Supported
5	Combination of RAID level between master volume and destination volume	RAID1(2D+2D)←→RAID1(2D+2D) RAID5(3D+1P)←→RAID5(3D+1P) RAID5(3D+1P)←→RAID1(2D+2D)
6	Data protection	There is a parity protection for both master volume and destination volume.
7	RESYNC pattern	From Master Volume data to destination volume
8	Time for transition from Duplex to Split.	3 min./VOL(3390-3) without IO (Depend on the number of pairs and the load of DKC)
9	When the destination volume can be accessed from HOST.	The destination volume can be accessed at only Split status.

10	Cooperation with HRC/XRC/CC	HMRCF : Supported The master volume of HMRCF can be a volume of HRC/XRC/CC. If the master volume of HMRCF and the master volume of HXRC/CC is same, Reverse Resync and Quick Restore of HMRCF is impossible. The secondary volume of HXRC/CC must not be used for HMRCF volume.
11	Cooperation with HODM	It is not supported.
12	Cooperation with HORC	HOMRCF : Supported The master volume of HOMRCF can be a M-VOL or R-VOL of HORC.
	Cooperation with HORC (only for HOMRCF)	Supported • The M-VOL or R-VOL of HORC can be a primary volume of HOMRCF. • The secondary volume of HMRCF can be a M-VOL of HORC.
	Cooperation with HMBR (only for HOMRCF)	Supported • A primary volume of HOMRCF can access from HOST like a usual volume. • A secondary volume of HOMRCF can not access from HOST if the pair is not SPLIT status. In SPLIT status, a secondary volume of HOMRCF can access from HOST like a usual volume.
13	Cooperation with HIHSM	Supported • The source, destination or RESERVE volume of HIHSM can not be the primary, secondary or RESERVE volume of HMRCF/HOMRCF. • The primary, secondary or RESERVE volume of HMRCF/HOMRCF can be the source volume of HIHSM. But if HMRCF/HOMRCF P-VOL or RootVOL has already 3 pairs, it can not be the source volume of HIHSM. • When HMRCF/HOMRCF pair which is combined with HIHSM is split, the migration of HIHSM is canceled.

14	HCAFE function	The HCAFE function provides the same function as IBM Flash Copy. You can operate the HCAFE function by using the PPRC TSO and DFSMSdss. The HCAFE function forms a pair by virtually or physically copying S-VOL data to the T-VOL. (A pair formed by means of the HCAFE function is especially called relationship.) When the relationship is established, a host can execute a reading/writing from/to T-VOL data that is a virtual or physical copy of S-VOL data. When establishing the relationship, the user can specify an extent for copying (to be referred to as extent). The relationship is canceled at the time when a copying of data in the extent is completed.

In case of HMRCF/HOMRCF volume is selected for HIHSM volume

	P-VOL	S-VOL	RootVOL	NodeVOL	LeafVOL	ReserveVOL
Source VOL	Possible*[1]	Possible	Possible*[1]	Possible*[2]	Command Reject	Possible
Target VOL	Command Reject	Command Reject	Command Reject	Command Reject	Command Reject	Command Reject
ReserveVOL	Command Reject	Command Reject	Command Reject	Command Reject	Command Reject	Command Reject

(*1) It is impossible, if HMRCF/HOMRCF P-VOL or RootVOL has already 3 pairs.
(*2) It is impossible, if HMRCF/HOMRCF NodeVOL is already paired with 2 LeafVOLs.
If you want to execute migration of HIHSM, you need to delete pair and reset ReserveVOL of HMRCF/HOMRCF.

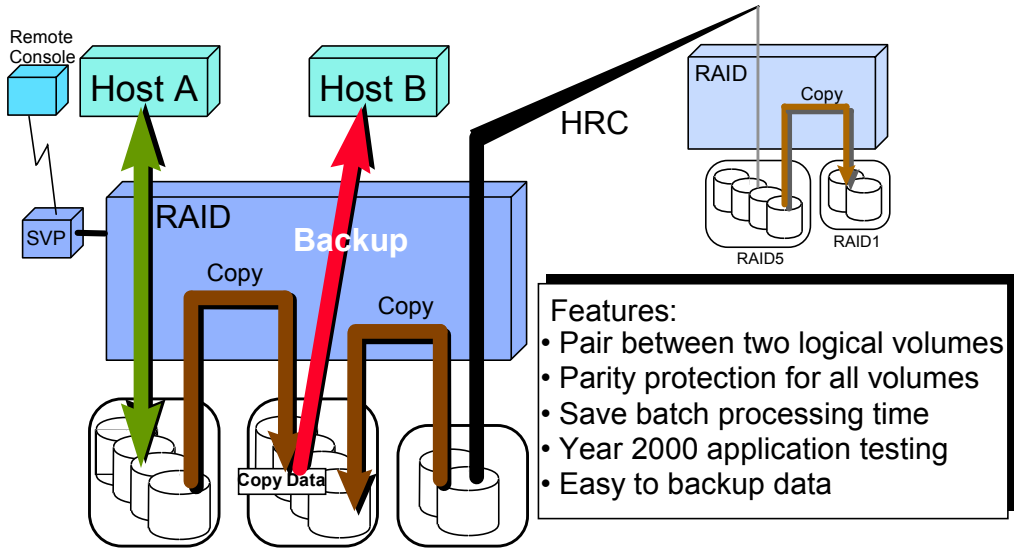
In case of HIHSM volume is selected for HMRCF/HOMRCF volume

	Source VOL	Target VOL	ReserveVOL
P-VOL	Command Reject	Command Reject	Command Reject
S-VOL	Command Reject	Command Reject	Command Reject
RootVOL	Command Reject	Command Reject	Command Reject
NodeVOL	Command Reject	Command Reject	Command Reject
LeafVOL	Command Reject	Command Reject	Command Reject
ReserveVOL	Command Reject	Command Reject	Command Reject

If you want to add pair of HMRCF/HOMRCF, you need to cancel HIHSM migration.

⚠ CAUTION

Copy process is done asynchronously with HOST i/o according to differential bit map.
Differential bit map is recorded on shared memory. So if shared memory is lost by offline micro exchange or volatile PSon etc., DKC lost differential bit map.
In these cases DKC treat as whole volume area has differential data, so copy process will take longer time than usual. And if the pair is SPLIT-PEND status, the pair become SUSPEND status because lost of differential bit map.
Primary volumes and secondary volumes of HMRCF/HOMRCF pairs should be placed on many RAID groups separately. And HMRCF/HOMRCF pairs which are operated at the same time should be placed in other RAID groups. HMRCF/HOMRCF pairs which are concentrated at very few RAID groups may influence HOST I/O performance.
If DKC is busy, increase Cache, DKA and RAID groups. And secondary volumes of HMRCF/HOMRCF pairs should be palaced in the increased RAID groups. HMRCF/HOMRCF pairs in very busy DKC may influence HOST I/O performance.



3.14.2 Construction of HMRCF & HOMRCF

- HMRCF & HOMRCF can be controlled from SVP, Remote Console and HOST.
- DKA with LA exchange have to be exist in the subsystem.

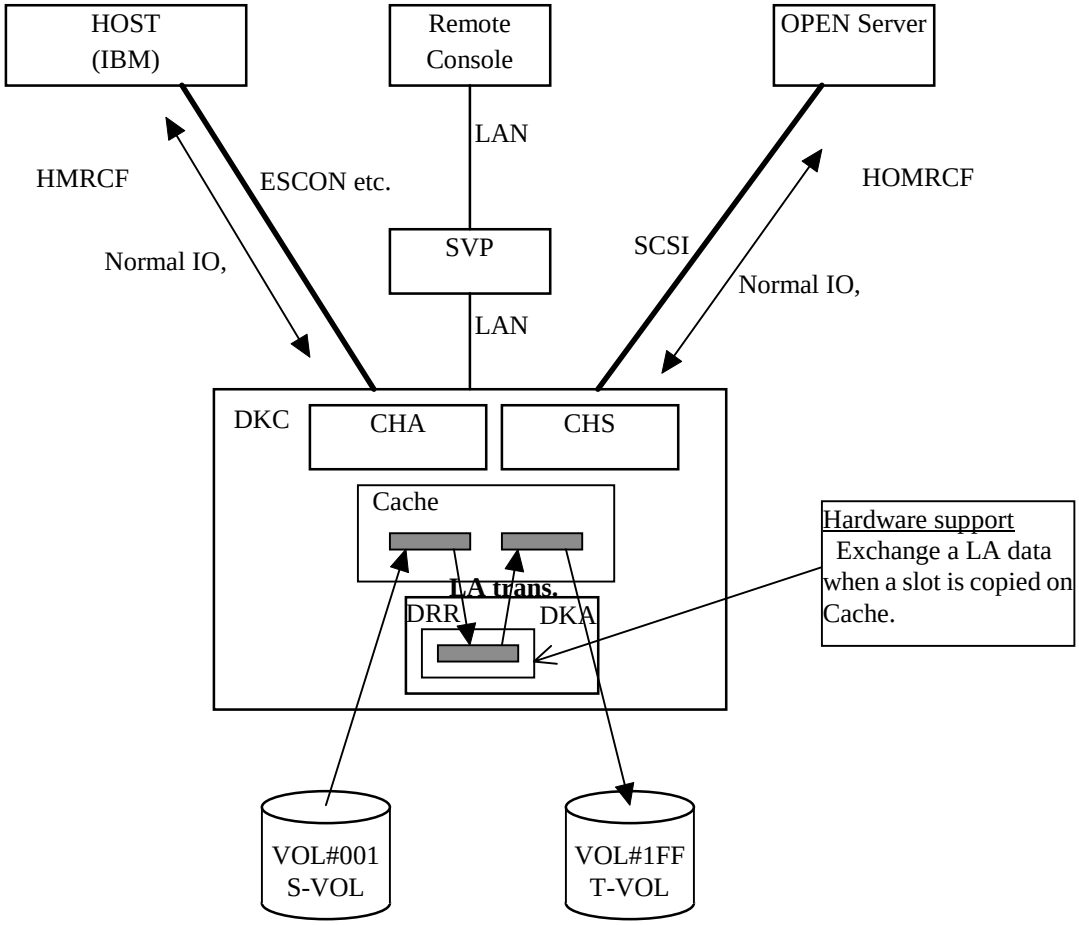


Fig. 3.14.2 Construction of HMRCF & HOMRCF

3.14.3 Status transition

Table 3.14.3 Status of HMRCF & HOMRCF

No.	Status	Definition
1	Simplex	There is no pair with the volume.
2	Pending	At this status, there is a job to copy the data from the master volume to the destination volume.
3	Duplex	The copy from master to destination is finished. The destination volume can not be accessed from HOST.
4	Split Pending	At this status, there is a job to copy the differential data from the master volume.
5	Split	The pair is splited. The destination volume can be accessed from HOST. At this status, the position of write data from the HOST is recorded on a bitmap to reduce the copy time on RESYNC.
6	Resync	There is a job to copy the differential data from master to destination.
7	Suspend	(1) There is an error with the pair. (2) After a running copy job was stopped by operation of SVP, the pair status is “suspend”.
8	Cafe	This is a status which a pair enters when the relationship definition is requested by the host command. In this status, the S-VOL data is being copied to the T-VOL in the background. In the case of the No Copy, the background copy is not performed.

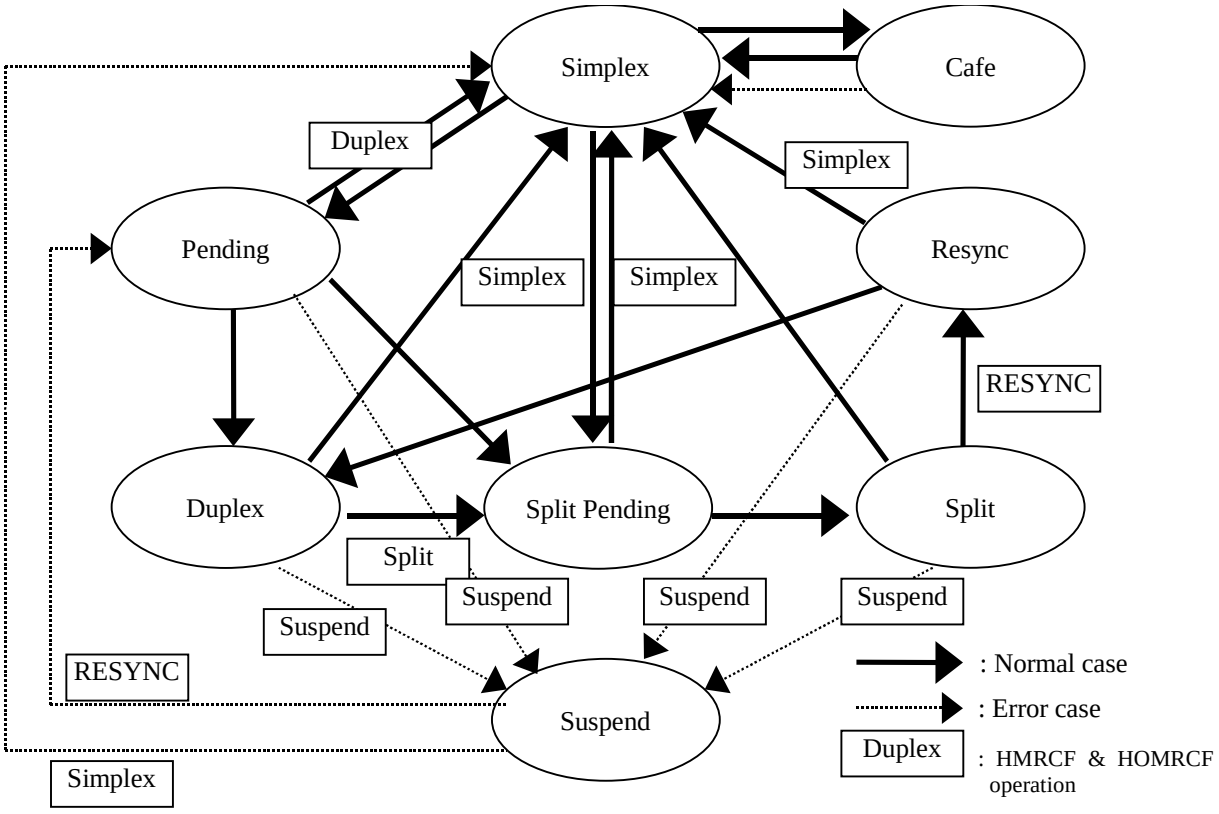


Fig. 3.14.3 Status transition map

3.14.4 Interface

(1) Outline

HMRCF & HOMRCF support a command set to control HMRCF & HOMRCF functions. This command set is a common interface in a subsystem. So the command from different HOST is translated to HMRCF & HOMRCF command at each command process.

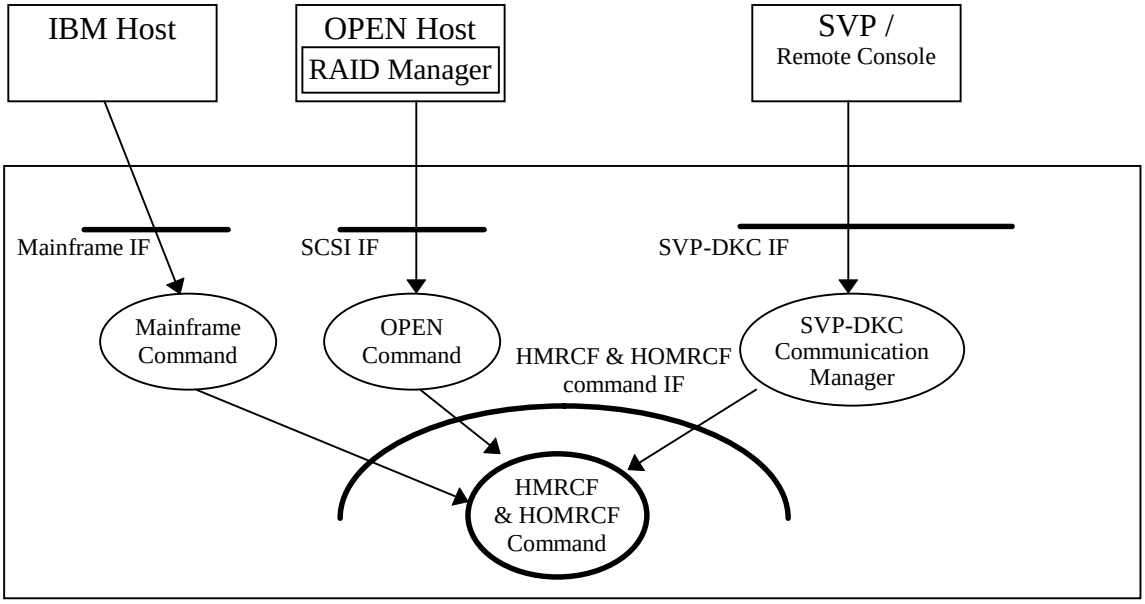


Fig. 3.14.4 Outline of HMRCF & HOMRCF IF

Notice
It is necessary to define Command Device before using RAID Manager on OPEN HOST.
(See INST05-880)
Please do not define Command Device on heavy loaded path.

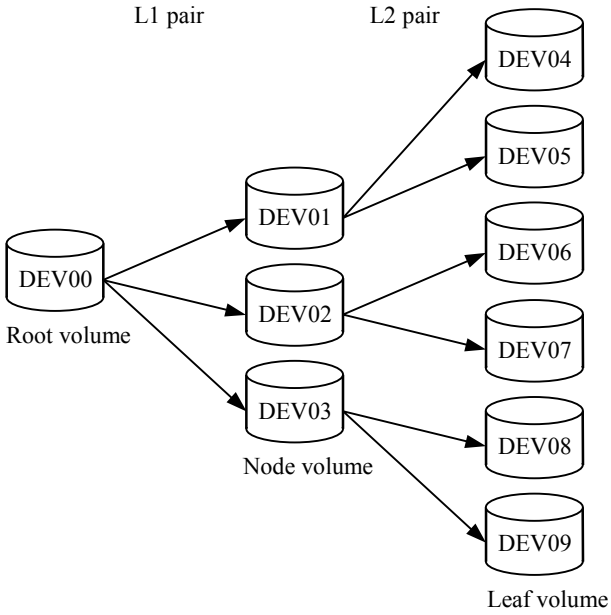
(2) HMRCF & HOMRCF operation

Table 3.14.4 HMRCF & HOMRCF operation

No.	Command	Operation
1	Duplex	Create a pair and start initial copy
2	Split	Split the pair
3	RESYNC	Resume the pair and start differential copy
4	Simple	Delete the pair
5	Suspend	Suspend the pair action
6	Status Check	Requirement for status information
7	Reserve	Marking and Unmarking the volume for a candidate of destination volume

3.14.5 Cascade function

It is possible to make a pair with a existed Target volume as a new Source volume by using Cascade function. See below figure.
This function is available for HOMRCF only.



No.	Content	Specification	Note
1	Pair structure	A Target volume of L1 pair (=Node volume) can be a Source volume of L2 pair.	
2	Number of copies	Root : Node = 1 : 3 Node : Leaf = 1 : 2 Root : (Node + Leaf) = 1 : 9	
3	Add pair condition	L2 pair have to be created after L1 pair.	The reverse is not true.
4	Split pair condition	L2 pair is able to execute split pair request only when the L1 pair is already split status.	
5	Resync pair condition	• No conditions.	
6	Delete pair condition	• No conditions. • When L1 pair is deleted, then L2 pair become L1 pair.	
7	Combination with HORC	Possible But Node volume and Leaf volume are treated as a target volume.	
8	Combination with HIHSM	Possible But Leaf volume is not able to move.	

- Name of volume type
  The source volume of first pair : Root Volume
  The target volume of first pair : Node Volume
  The source volume of 2nd pair : Node Volume
  The target volume of 2nd pair : Leaf Volume
- Name of pair
  - The first pair (A pair of root volume is source volume) : L1 pair
  - The second pair (A pair of node volume is source volume) : L2 pair
- Name of pair chain
  - a chain of L1 pair and L2 pair with a node volume: stream

3.14.6 Reverse-RESYNC

(1) Reverse-RESYNC Function/Quick Restore Function

The Reverse-RESYNC function is an extension of the RESYNC function of the MRCF.
The Quick Restore function is a similar function with Reverse-RESYNC, but it is speeds up the operation.

When a pair in the Split status is requested to perform the Reverse-RESYNC, the differential data between the target volume and the source volume is copied to the source volume from the target volume.
When a pair in the Split status is requested to perform the Quick Restore, a volume map in DKC is changed to swap contents of Source volume and Target volume without copying the Source volume data to the Target volume. The Source volume and the Target volume are resynchronized when update copy operations are performed for pairs in the Duplex status.

Note on RAID Level and DCR swap:
Quick Restore operation changes places of the data which primary volumes and secondary volumes and location of DCR of HMRCF/HOMRCF pairs have. So the operation may change RAID levels, HDD types of the volumes. For example, if the primary volumes is RAID1 and the secondary volumes is RAID5, Quick Restore operation changes the primary volume to RAID5 and the secondary to RAID1.
If you want previous state, do following actions.
step1 : stop HOST I/O to the pair
step2 : split the pair
step3 : quick restore the pair
step4 : restart HOST I/O to the pair

And because of replacing locations of DCR setting, you must operate 1 or 2 shown below.
1. Set same locations of DCR of Source volume and Target volume.
2. Reset DCR of Source volume and Target volume before Quick Restore, and set DCR of Source volume and Target volume after the pair transits to Duplex status by Quick Restore.

Unless you operate this, I/O performance to the same data may be down for the change of the locations of cache residence area after Quick Restore.

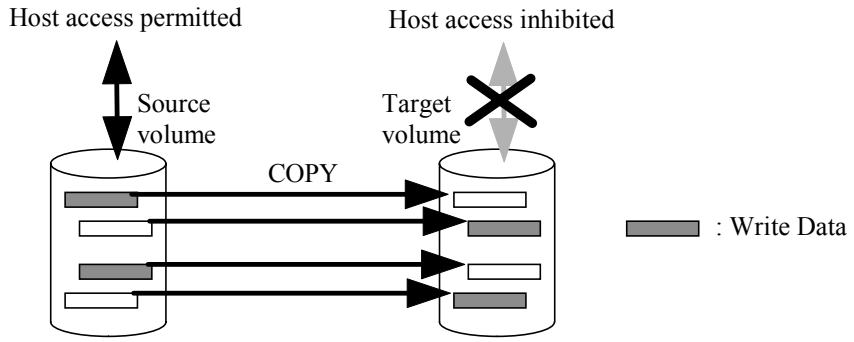


Fig. 3.14.6-1 Normal RESYNC Process

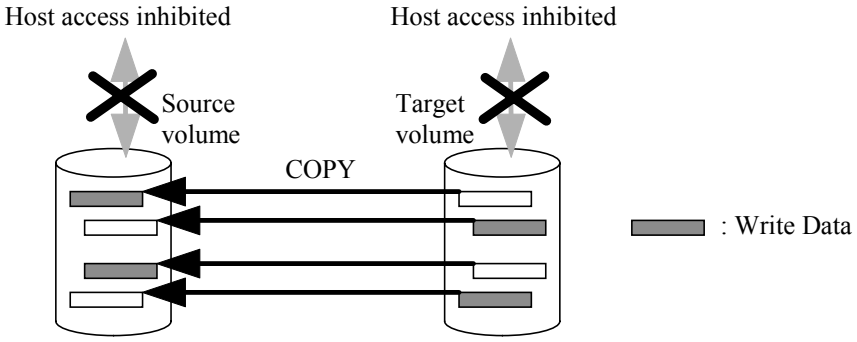


Fig. 3.14.6-2 Reverse RESYNC Process

(2) Specifications

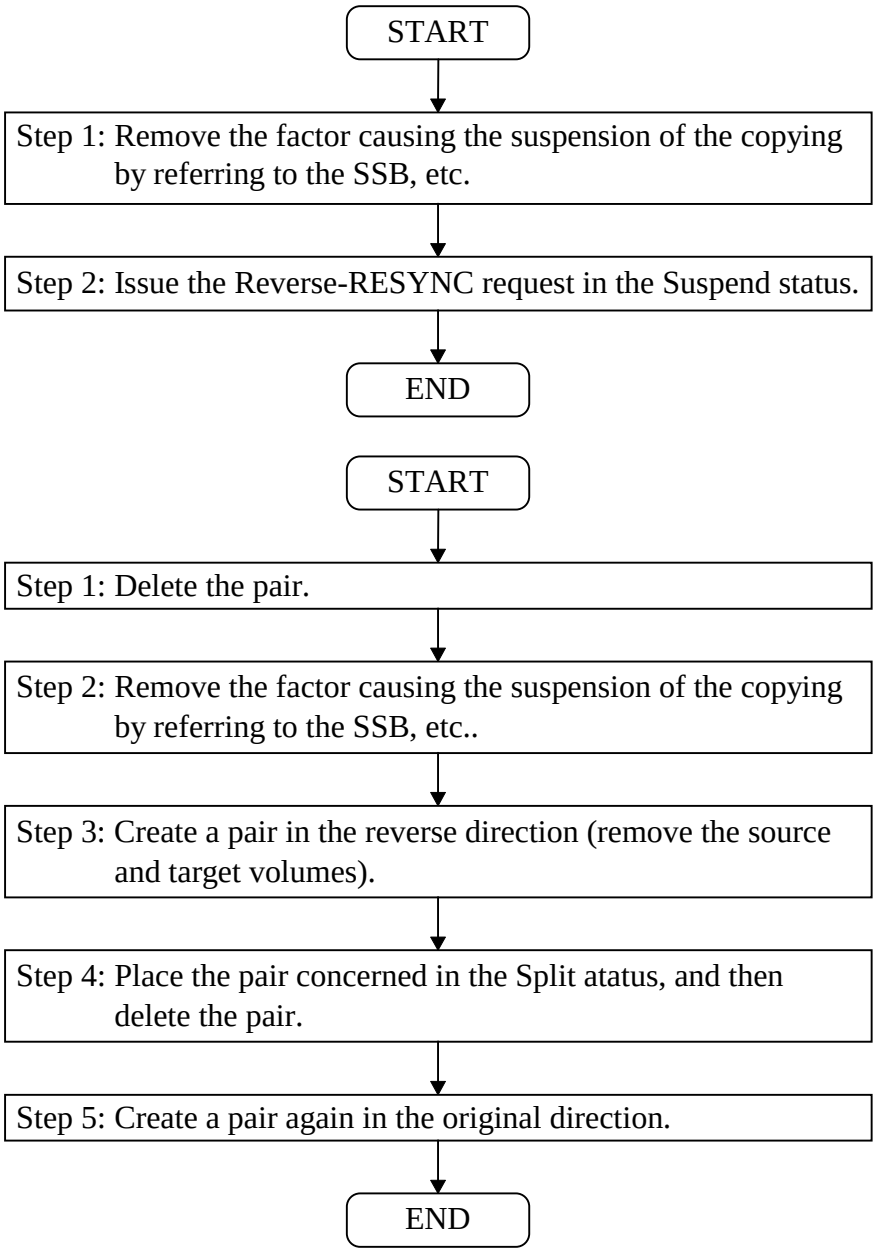
No.	Item	Description
1	RESYNC copy pattern	• The data of the target volume is copied to the source volume. • The copy pattern can be selected by specifying a unit of operation. Specified operation unit : SVP/RMC: In units of pair operation at a time RAID manager: In units of command
2	Copy range	• In the case of the Reverse-Copy and Quick Restore in the Split status, range for merging writing into the regular and target volumes
3	Copy format	• Same format as that of a copy in the Duplex status
4	Applicable LDEV type	• HMRCF : 3390-1/2/3/3R/3A/3B/3C/9/L, 3380-J/E/K/KA/KB/KC and CVSs of them. • HOMRCF : OPEN-3/8/9/K/E/L/M and CVSs and LUSEs of them.
5	Host access during copying	(1) In the case of the main frame volume • Source volume: Reading and writing disabled Target volume: Reading and writing disabled (2) In the case of the open volume • Source volume: Writing disabled Target volume: Reading and writing disabled Note: The reason why the source volume is not disabled to read is to make the volume recognizable by the host and it does not mean that the data is assured.
6	Specification method	• SVP/RMC: Add a specification for the RESYNC pattern onto the Pair Resync screen.
7	Conditions of command reception	• The pair concerned is in the Split status. • In the case of Quick Restore, the pair must not be combined CVS Volume and Normal Volume. • Another pair sharing the source volume is in the Suspend or Split status. → If this condition is not satisfied, the CMD RJT takes place. • When the Reverse-Resync or Quick Restore is being executed by another pair which is sharing the source volume, it is impossible to change the pair status of the pair concerned. (However, the pair deletion and pair suspension requests are excluded.) • The source volume or the target volume of the pair concerned has no pair of the HRC/HORC or in the Suspend status. (See Item No.14 in this table.) • The source volume or the target volume of the pair concerned has no pair of the XRC/CC.
8	Status display during copying	• SVP/remote console HMRCF : RESYNC -R HOMRCF : COPY(RS-R) The display of the attribute, source or target, is not changed. • RAID manager Pair status display: RCPY The display of the attribute, source or target, is not changed.

9	Condition after the copying terminates normally	• The pair concerned enters the Duplex status. • The conditions of the host access after the status transition are shown below. (1) Main frame volume Source volume: Reading and writing enabled Target volume: Reading and writing disabled (2) Open volume Source volume: Reading and writing enabled Target volume: Writing disabled
10	Effects on another pair	In another pair sharing the source volume, the part actually copied becomes the difference after execution of this function. Example: Pair of the other target volumes in the 1:3 configuration
11	Operation when the copying terminates abnormally	(1) The pair concerned enters the Suspend status. (2) The source volume of the pair concerned is enabled to read and write. → Data is not assured. The target volume of the pair concerned is disabled to read nor write in the case of the main frame volume and disabled to write in the case of the open volume. (3) The status of a pair sharing the source volume is not changed.
12	Operation when a suspension request is received during copying	Same as above
13	Relation to the cascade function	• The Reverse-RESYNC and Quick Restore cannot be executed for the L2 pair.
14	Relation to the HRC/HORC	• In the case where "M-volume of the HRC/HORC" = "Source volume of the MRCF" → If the HRC/HORC pair is not Suspend status, the Reserve-Resync and Quick Restore cannot be executed. (Command Reject) • In the case where "R-volume of the HRC/HORC" = "Source volume of the MRCF" → If the HRC/HORC pair is not Suspend status, the Reserve-Resync and Quick Restore cannot be executed. (Command Reject) • In the case where "Target volume of the MRCF" = "M-VOL of the HRC/HORC" → If the HRC/HORC pair is not Suspend status, the Reserve-Resync and Quick Restore cannot be executed. (Command Reject) • A pair of the HRC/HORC cannot be created with the volume of the MRCF executing the Reserve-Resync or Quick Restore. (Command Reject)

Notice:
After the Reverse-Resync or Quick Restore is executed, make sure that the pair transits to DUPLEX status before the Resume of the HRC/HORC is executed. If Resme is executed before the pair transits to DUPLEX status, the HMRCF/HOMRCF pair can transit to SUSPEND status.

(3) Action to be taken when the pair is suspended during the Reverse-RESYNC
The recovery procedure to be used when the pair executing the Reverse-RESYNC is suspended owing to some problem or it explicitly placed the Suspend status by a command from the SVP/remote console/RAID manager is explained below.

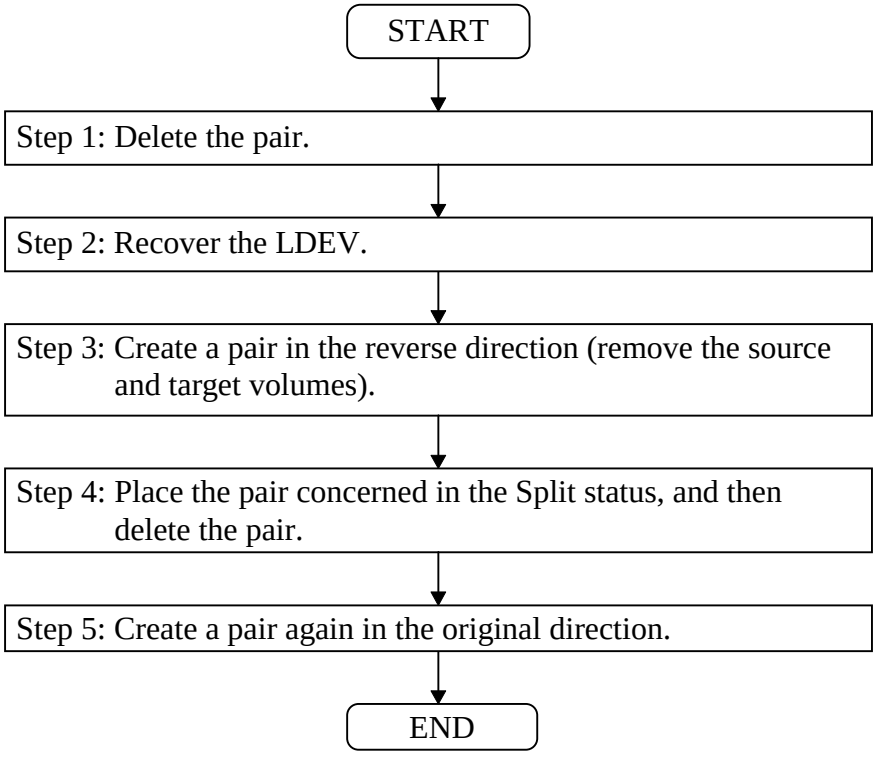
(a) Case 1: A case where the Suspend status can be recovered without recovering the LDEV concerned.
This is equivalent to a case where the pair encounter an event that copying cannot be continued owing to a detection of pinned data or a staging time- out.
Or, it is equivalent to a case where the pair explicitly placed in the Suspend status by a command.
<<Recovery procedure>>



(b) Case 2: A case where the Suspend status cannot be recovered unless the LDEV concerned is recovered.
This is equivalent to a case that the LDEV is blocked.
To recover the blockade of the LDEV, an LDEV formatting or LDEV recovery is required. Both of them cannot be executed in the state that the MRCF pair is created. (A guard works against it.) Therefore, delete the pair once, recover the LDEV, and then create the pair once again.
However, care should be taken because, in the pending state, the data of the source volume is copied to the target volume if simply create the pair again. Recover the blockade following the procedure below.

The following procedure is applicable just to a restoration of the source volume using the target volume. A procedure for directly restoring the source volume when the target volume is blocked is not included.

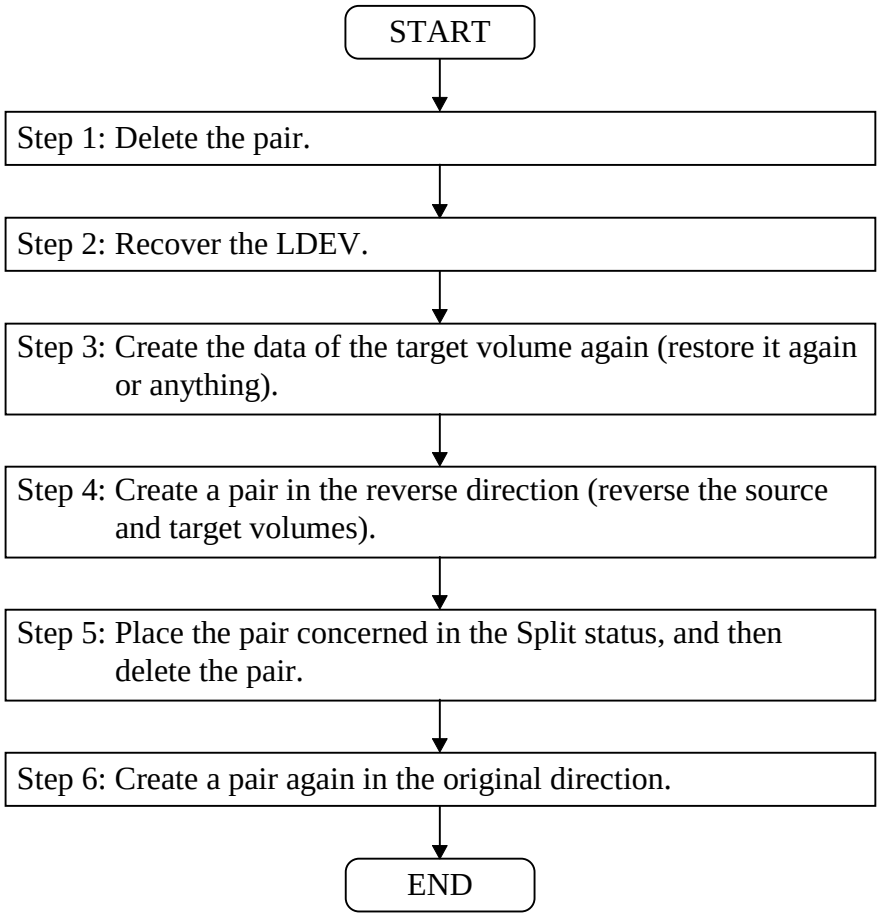
• Case 2-1: A case where the source volume is blocked
<<Recovery procedure>>



• Case 2-2: A case where the target volume is blocked

A recovery procedure for restoring data of the target volume is added because the copy source of the Reverse-RESYNC cannot be accessed.

<<Recovery procedure>>



3.14.7 HCAFE function

Operations that can be done for the HCAFE pair are shown below.

Operation	Pair status		
	Simplex	Cafe	The others
Delete Pair	See Table 2.4.	Yes	See Table 2.4.
Relationship definition	Yes	No	No
The others	See Table 2.4.	No	See Table 2.4.

The S-VOL in the Cafe status and the T-VOL in the status other than Cafe can be shared. The S-VOL in the status other than Cafe and the T-VOL in the Cafe status can be shared also. Therefore, the configuration shown in Figure 3.14.7-1 can be formed.
The pair (L2) in the second layer (L2=Layer2) can be formed under the pair (L1) in the first layer (L1=Layer1). The L2 pair can be formed only when either of the L1 and L2 pairs is the Cafe pair. The L3 pair (L3=Layer3) cannot be formed when both of the L1 and L2 pairs are the Cafe pairs or both of the L1 and L2 pairs are other than the Cafe pairs.

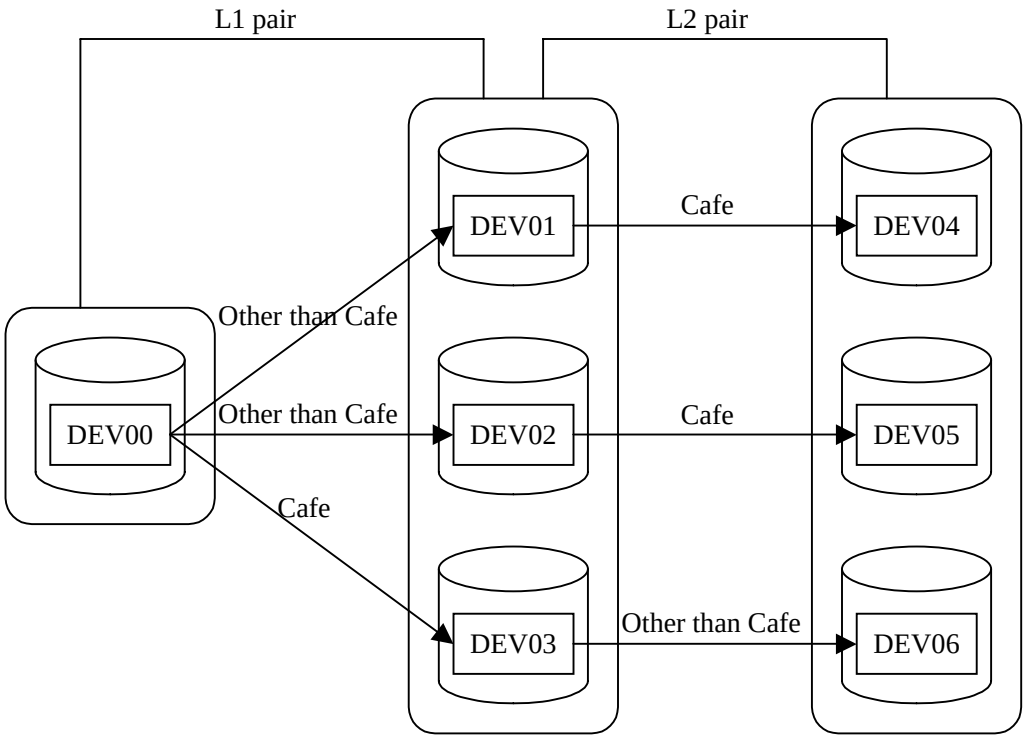


Figure 3.14.7-1 HMRCF Extended Configuration Formed by Means of HCAFE

Relation between the L1 pair statuses and operability of the L2 pair is shown below.

Relation between the L1 pair statuses and operations of the L2 pair

	Operation of L2 pair						
L1 pair status	Add Pair	Split Pair	Resync Pair	Reverse Resync/ Quick Restore	Suspend	Delete	Relationship definition
Pending	–	–	–	–	–	OK	NG
Duplex	–	–	–	–	–	OK	NG
SP-Pend	–	–	–	–	–	OK	NG
V-Split	–	–	–	–	–	OK	NG
Split	–	–	–	–	–	OK	OK
Resync	–	–	–	–	–	OK	NG
Reverse Resync / Quick restore	–	–	–	–	–	OK	NG
Suspend	–	–	–	–	–	OK	NG
Cafe	OK	NG	OK	NG	OK	OK	NG

"-" (dash) means that both L1 and L2 pairs are in the status other than Cafe.

Relation between the L2 pair statuses and operability of the L1 pair is shown below.

Relation between the L2 pair statuses and operations of the L1 pair

	Operation of L1 pair						
L2 pair status	Add Pair	Split Pair	Resync Pair	Reverse Resync/ Quick Restore	Suspend	Delete	Relationship definition
Pending	–	–	–	–	–	OK	OK
Duplex	–	–	–	–	–	OK	OK
SP-Pend	–	–	–	–	–	OK	NG
V-Split	–	–	–	–	–	OK	NG
Split	–	–	–	–	–	OK	OK
Resync	–	–	–	–	–	OK	NG
Reverse Resync / Quick restore	–	–	–	–	–	OK	NG
Suspend	–	–	–	–	–	OK	OK
Cafe	OK	NG	OK	NG	OK	OK	NG

"-" (dash) means that both L1 and L2 pairs are in the status other than Cafe.

3.14.8 Notes on powering off

When performing a powering off, take notice of the following.

Item	Note	Reason
1	(MRCF) Take care that the time required for the copying becomes longer. Make a schedule taking the copying time into consideration.	If data in the shared memory has volatilized when the next powering on is performed, the following phenomena occur. • When the pair is in the Pending or Resync status, the data, from which a copying has been completed before the powering off, is also treated as data to be copied again. · Even if no I/O has been issued, the rate of data identity does not reach 100% when the pair status is changed to Duplex. · The data that has become the one to be copied again is copied to the secondary volume after the pair status is changed to Duplex. • When the pair is in the Duplex status, the data, from which a copying has been completed before the powering off, is also treated as data to be copied again. · The rate of data identity will be 0%. · The copying of the data, which has become the one to be copied again, is performed in the state in which the pair is in the Duplex status. • When the pair is in the Split status, the whole volume will be a differential between the two volumes. · The rate of data identity will be 0%. · Data of the whole volume is copied when a resynchronization is performed.
2	(MRCF) As to a pair in the SP-Pend status, complete the copying of it and put it in the Split status.	If data in the shared memory has volatilized when the next powering on is performed, the following phenomenon occurs. • When the pair status is SP-Pend, it is changed to Suspend.
3	(HCAFE) Perform a powering off of the subsystem after the copying is completed.	If data in the shared memory has volatilized when the next powering on is performed, the following phenomena occur. • The relationship is dissolved. • The secondary volume is detached.

3.15 TPF

3.15.1 An outline of TPF

TPF stands for Transaction Processing Facility.
TPF is one of operating system (OS) for mainframes is mainly used for airline customer reservation systems.
To correspond to TPF, DKC must support logical exclusive lock facility and extended cache facility. The former is a function which called MPLF (Multi-Path Lock Facility) and the later is a function which called RC (Record Cache).

DKC has implemented a special version of microprogram which supports the MPLF and RC functions of TPF feature(RPQ#8B0178), described in IBM public manuals;
(a) IBM3990 Transaction Processing Facility support RPQs (GA32-0134-03)
(b) IBM3990 Storage Control Reference for Model 6 (GA32-0274-03)

(1) An outline of MPLF

This facility provides a means, using a DKC, to control concurrent usage of resources in host systems via use of logical locks. A logical lock may be defined for the control of a shared resource, where the sharing of that resource must be controlled. Each shared resource has it own name called Lock Name. Every Lock Name controls multiple lock states (2 to 16).
The following figure shows the outline of I/O sequence which uses MPLF.
DKC recognizes up to 16 MPLF users. In this figure, user A and user B are used. These users may belong to same HOST or different HOSTs. Each user must indicate MPLP (Multi-Path Lock Partition) when uses MPLF. MPLP is a means of logically subdividing the MPLs (Multi-Path Locks) for a user set. The maximum number of MPLP is four. Each MPLP has numbered from 1 to 4. The process to get permission for use MPLF is called CONNECT.
The connected user execute SET LOCK STATE process using Lock Name. The MPL corresponding to specified Lock Name is assigned to the user. This assignment is canceled by UNLOCK process. HOSTs can share the DASD without contradiction by using this MPLF.

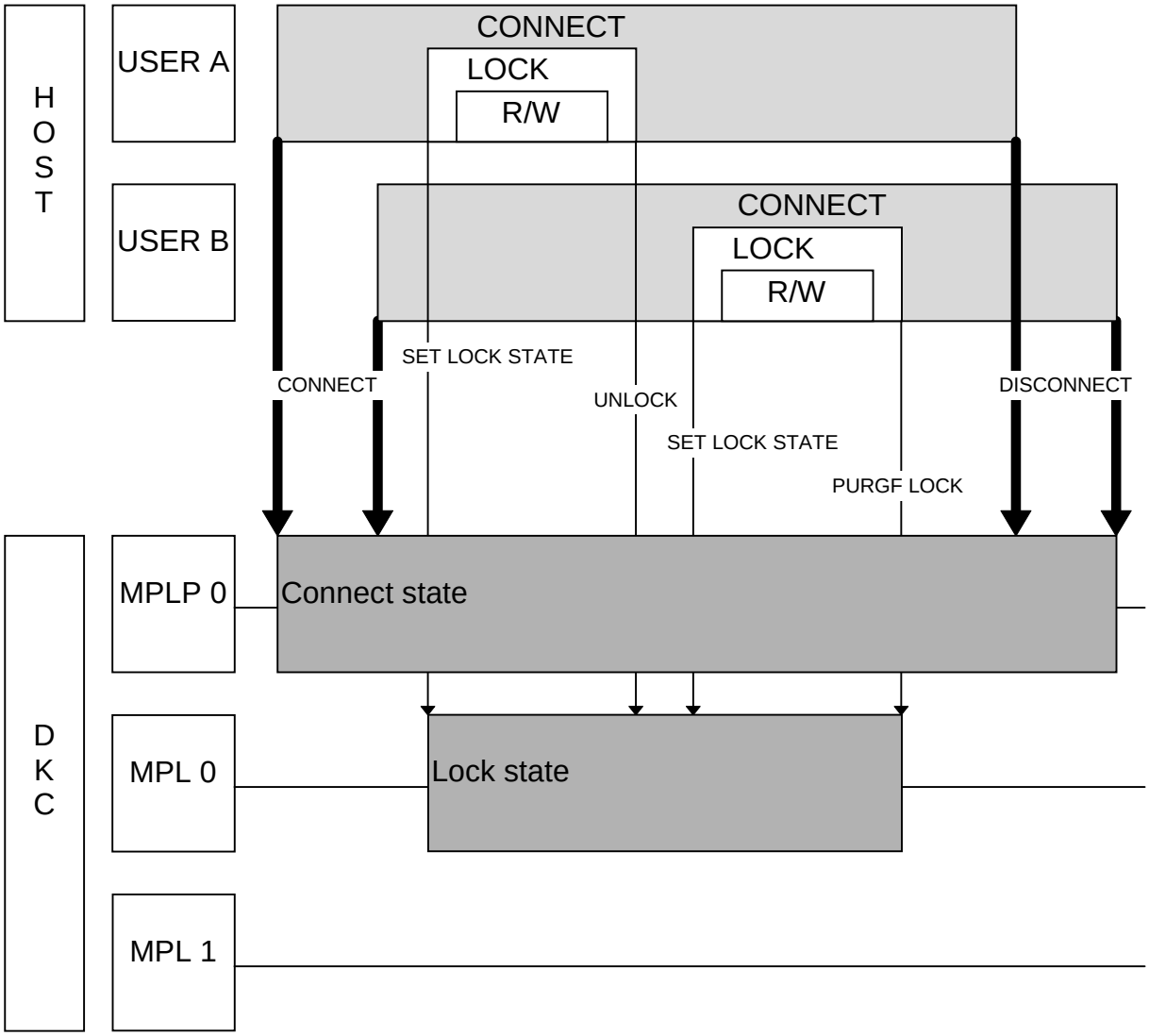


Fig. 3.15.1 An outline of MPLF

(2) An outline of RC

RC has following two feature.
(a) Record Mode Chain
(b) Record Caching

The followings explain these features.
(a) Record Mode Chain

Record Mode Chain consist of following 4 command chains.

1) Mainline Processing (Read)
2) Mainline Processing (Write)
3) Capture
4) Restore

To execute Record Mode Chain, subsystem must be initialized for Record Caching, and Record Mode must be allowed for the addressed device. Under these conditions, Record Mode Chain works when Record Mode Chain is indicated in the Extended Global Attributes of Define Extent command. Otherwise, the chain is processed in standard mode.

A Mainline Processing chain consists of a Define Extent command, a Locate Record command, and a single Read Data or Write Update Data command.

A Capture chain consists of a Define Extent command followed by a Seek command and multiple Read Count, Key, and Data commands.

A Restore chain consists of a Define Extent command, a Locate Record command, and multiple Write Update Data commands.

(b) Record Caching

Record Caching is the naming contract with Track Caching used in standard model. At the point first initialization completes, all of cache is allocated to Track Slot as standard model. Record Cache will be allocated if Set Cache Allocation Parameters Order is issued.

3.15.2 TPF Support Requirement

(1) OS

TPF Ver.4.1.

(2) Hardware

The following table shows subsystem hardware specification for TPF support.

Table 3.15.2 TPF Support Hardware Specification

Item	Description
Number of MODs	Max. 4096/box
Number of LCUs/Box	Max. 16
Number of SSIDs/LCU	1
Cache/SM capacity	(Refer to INST01-40)
RAID level	5 or 1
Emulation type (1) LCU (2) Device	 3990-3/6 or 2105-E20 *1 3390-1/2/3/9 or 3380-K
Physical Disk drive	(Refer to THEORY02-20)
Number of Host paths	Max. 32
Number or Multi-Path Locks	16k/LCU (when up to 8 LCUs) 8k/LCU (when 9 ~ 16 LCUs)
Option features; (1) CVS = Virtual LVI (2) DCR = FlashAccess (3) HRC (4) HMRCF = Shadow Image (5) HIHSM (6) HODM (7) HMBR	 Available Available Available Available Available *2 (Not Available) (Not Available)

*1: 2105-E20 is supported 01-14-6X and 01-16-XX or higher of DKC MAIN version.

*2: HIHSM supports only a Monitor function.

3.15.3 TPF trouble shooting method

Basically there is no difference between TPF environment and MVS (as a standard operation system) for trouble shooting.

A example order is below;
(a) To gather system error information by Syslog ...etc.
(b) To gather DKC error information by SVP dump operation.
- Normal dump which contains TPF dump data also. (Refer SVP02-530)

(c) To send above data to T.S.C.

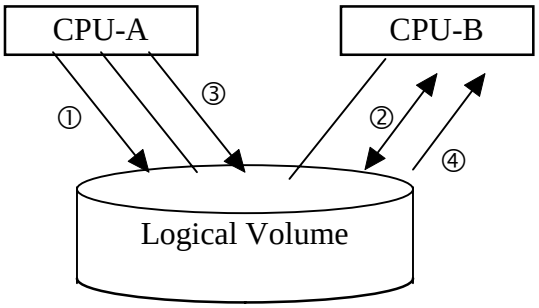
3.15.4 The differences of DASD-TPF(MPLF) vs DASD-MVS

(1) Data-exclusive method by MPLF function

MVS environments
(a) Logical volume(Device) is the unit of data-exclusive between several CPUs.
(b) “Device” is owned by one CPU during its processing(accessing), and “Device-busy” status is reported to another CPU’s accesses.
(c) “Device-end” status is used to notify the waiting CPUs
(d) when the device becomes free.

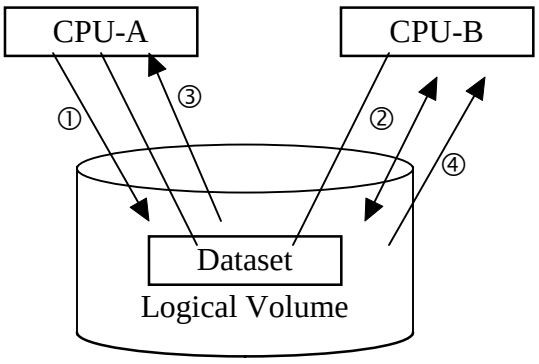
TPF environments
(a) Logical “Lock” is used for this purpose, instead of logical volume(device) of MVS.
(b) Most Read/Write CCWs have a unique: Prefix-CCW(Set Lock) to own the target lock.
And only when the request-lock is granted to, its CCW continues the following Read/Write processes.
DSB=”4C” is for granted / DSB=”0C” is for NOT-granted(wait).
(c) “Attention” status is used to notify the waiting CPUs when the lock becomes free.
(d) The relationship between Lock and Dataset is quite free.
Usually TPF users(customers) have their own definitions.

MVS environments



① Reserve/Read&Write Access by CPU-A (Successful).
② CPU-B’s trial is rejected by Device-busy (Failed).
③ Terminate its process and release the volume.
④ Free(Device-end) will be sent. CPU-B can use this volume.

TPF environments



① Set Lock/Read&Write process(*1) by CPU-A (Successful).
② CPU-B’s trial is rejected by Not-granted (failed).
③ Terminate with Unlock, by CPU-A.
④ Free(Attention) will be sent.(*2). CPU-B can use this Dataset.
(*1)Typical CCW chain:
- Set lock State(x27/order(x30));
- Read Subsystem Data(x3E);
- Tic (to be continued if granted)
- (ordinary CCW chain)
(*2)This report’s path/Address is usually different from above②.

Fig. 3.15.4-1 Environments of DASD-TPF and DASD-MVS

(2) No path-group

MVS environments
(a) Each CPU establishes the Path-group on every DASD Online-devices, using the all connected paths.
(b) Channel and DASD (Control Unit) rotates the I/O service path to meet each occasion within this group.
(c) “Device-end” status can be reported through any-path of this group.

TPF environments
(a) TPF OS/CPU does not establish this Path-Group, even if its configuration has multiple-paths for DASD.
(b) But the Channel rotates the I/O request-path, within the connected paths. (Like old MVS way.)
(c) “Attention” report is restricted to one “Connect-Path” which has been defined during IPL (Vary-online) procedure.

(3) Connect Path/Device

(a) TPF system issues “Connect order” to define ;
- Lock tables on each DASD control-unit,
- Report path & Device for Attention interrupt.

(b) This order is code(x33) of Perform Subsystem Function (x27) command.
(c) This order is issued during the IPL process of each CPU.
(d) CPU (channel) only has the capability to change this path and device definition.

Table 3.15.4-1 Order-list of Perform Subsystem Function (x27) command

Order	Meaning	Function
x10	Commit	RC
x11	Discard	
x18	Prepare for Read Subsystem Data	
x19	Destage Modified Tracks	
x1B	Set Special Intercept Condition	
x20	Set Cache Allocation Parameters	
x21	Suspend/Resume Function	
x22	Prepare to Read Lock Data	MPLF
x30	Set Lock State	
x31	Purge Lock	
x32	Unlock	
x33	Connect	
x34	Disconnect	

For detailed, please see the IBM RPQ manual;
“IBM 3990 Transaction Processing Facility Support RPQs” (GA32-0134-03)

(4) Channel Re-drive function

MVS environments
(a) In general, the Channel selects the most proper path (in the Path-group) for each I/O request.
(b) In MVS environments, there is no this kind of function.

TPF environments
(a) In TPF environments,
to keep a fast IO response & IO request-order(fast-in-fast-out),
this kind of special function has been introduced. (Thus is our conjecture.)
(b) By the channel-monitor data,
Once IO request is rejected with Control-unit busy by DASD,
Sub-channel repeats a reconnect-trial to DASD (with some short interval),
until (1) to get-into DASD or (2) to reach the trial-count threshold
(in this case IO request is registered to some waiting Queue in the channel.)
(c) And once IO request is accepted by DASD control-unit,
next Control-unit judges this would be accepted or not using “Lock” status.

(5) Fixed Record-size

Dataset structure in DASD

In general, TPF software makes the following logical structure in DASDs.

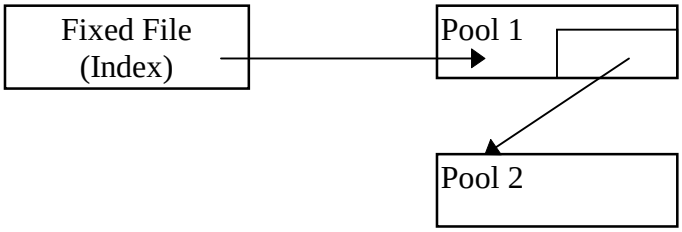


Fig. 3.15.4-2 Logical structure in DASDs

Table 3.15.4-2 Pool records Classification

Logical (usable) size	Physical size
381 Bytes	384 Bytes
1055 Bytes	1056 Bytes
4095 Bytes	4096 Bytes

There exist only three lengths for pool records.

Table 3.15.4-3 More detailed classification

381 Record	1055 Record	4095 Record
SLT (Small, Long Term)	LLT (Large, Long Term)	4LT (4KB, Long Term)
SST (Small, Short Term)	LST (Large, Short Term)	4ST (4KB, Short Term)
SDP (Small, Long Term, Duplicated)	LDP (Large, Long Term, Duplicated)	4DP (4KB, Long Term, Duplicated)

(6) Prime/Dupe MODs pairs

(a) To improve Data-integrity of DASD,
TPF system often makes the Data-duplications
on different two DASD subsystems.
(b) The following figure shows one example of this pairs.
Prime MOD(module)s and Dupe MODs are never located on
one-side of subsystem(spread to all subsystems).

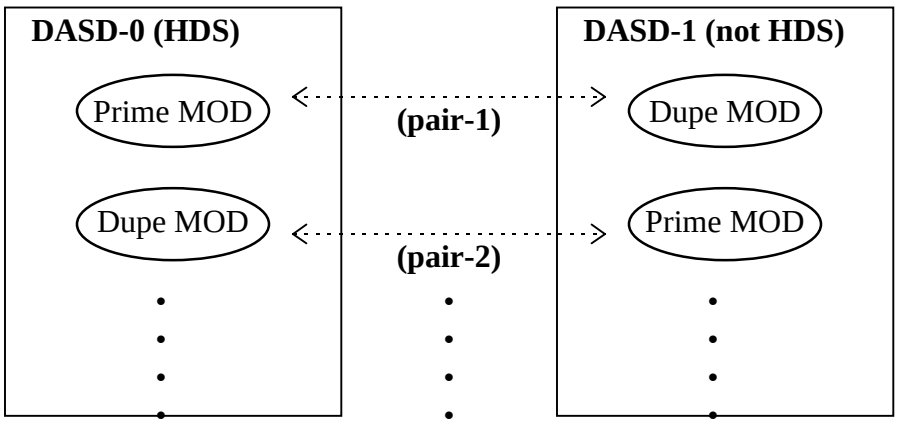


Fig. 3.15.4-3 Prime/Dupe MODs pairs

(7) Data Copy procedures

The Copy procedures are taken for the following purposes;
(a) To make a pair (To copy data from Prime MOD to Dupe),
(b) To recover the failed data (To copy the remaining data to the re-init MOD.)

There exist two ways to make a pair.
(a) AFC (All File Copy), and
(b) AMOD copy.

(a): In this copy process, the destination-drive of the copy keeps “Offline” status, and just after the completion of this copy, the source-drive is made “Offline” and the destination-drive is changed to “Online”. From the view-point of TPF software, there exists only one MOD, independent of copy process.

(b): In this copy, both source-drive and destination-drive are made “Online”.
TPF software can distinguish both drives, even if in the copy process.

3.15.5 Notices for HRC-option setting

<SVP operation>

(1) RCU Option (see SVP04-170)

- We strongly recommend to choose “No”-side in “PPRC support by host” column of “RCU Option” window.
- We strongly recommend to choose “Not Report”-side in “Service SIM of Remote Copy” column of “RCU Option” window.

(2) Pair Option (see SVP04-230)

- We strongly recommend to choose “Copy to R-VOL”-side in “CFW Data” column of “Pair Option” window.

(3) Suspend Pair (see SVP04-290)

- We strongly recommend not to activate (to keep blank) “SSB(F/M=FB)”-lone of “Suspend Pair” window.

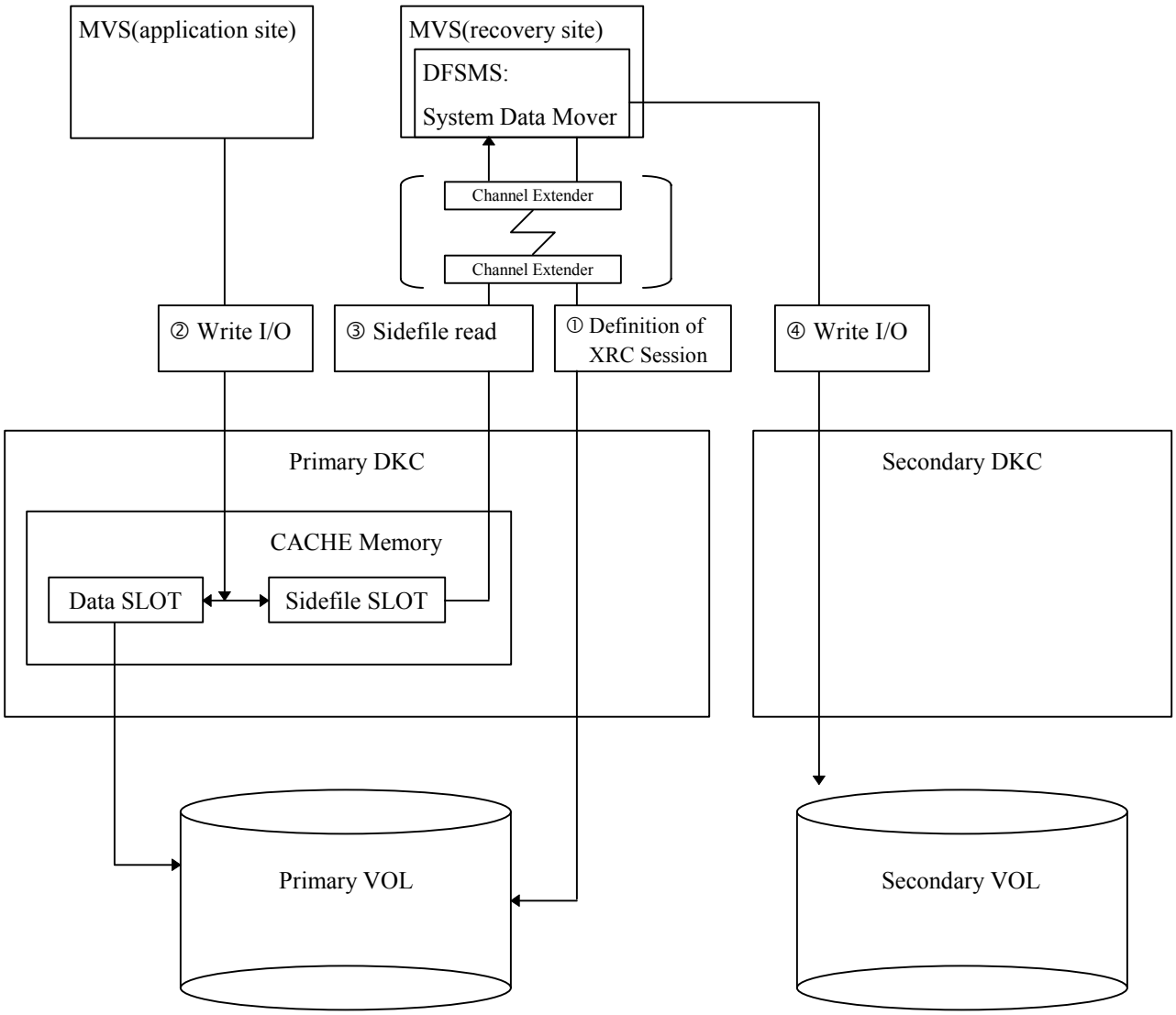
<Host (TPF-OS) consideration>

In MVS-OS world, DKC with HRC expects (requires) Customers to extend I/O Patrol Timer to prevent many MIHs from reporting.
In TPF-OS world also, same consideration is required, so please discuss with the Customer to find the opportunity to extend “Stalled Module Queue” timer over 5 seconds.

3.16 HXRC

3.16.1 An outline of HXRC

HXRC(Hitachi eXtended Remote Copy) function provides for data replication at distance in order to recover at a disaster.



- System Data Mover defines a XRC pair session.(①)
- When a write command is issued to primary volume from application site, primary DKC makes a Sidefile data (replication) on cache memory.(②)
- System Data Mover reads a Sidefile data non-synchronously at distance, and writes it to secondary volume.(③, ④)

Fig.3.16.1-1 An outline of HXRC

3.16.2 HXRC Support Requirements

3.16.2.1 OS

(1) OS level
(a) MVS/ESA 4.3.0 or upper.
(b) DFSMS/MVS 1.1.0 or upper.

<Restriction of SMS 1.4 environment>
The Maximum number of HXRC pair per CU image is up to 128 under SMS 1.4 environment. CCA (Channel Connection Address) can be specified to 128 logical devices per CU images such '00' - '7F'. CCA address '80' - 'FF' for HXRC may be rejected by System Data Mover.

(2) Conditions of HXRC using
(a) The following conditions must be satisfied by OS before starting HXRC function,
• CACHE ON
• NVS ON
and must be CACHE ON status on DKC.
When CACHE OFF/NVS OFF commands are issued by OS or Cache malfunctions (includes 'Ref code=FFEE: Area temporary blocking) occur, HXRC function is stopped.

(b) I/O Patrol Value
(I) Without CHL Extender
• Current patrol time(more than 30sec).
(II) With CHL Extender
• More than 70sec.

(c) Session ID
• Up to 64Session ID's can be utilized per 1CU for Concurrent Copy and HXRC.
• Up to 4Session ID's can be utilized per 1CU for HXRC.
• Only 1Session ID can be utilized per 1VOL for HXRC.

Table 3.16.2.1-1 Number of session of volume

SESSION	DEV TYPE			
	3390-1/2/3/3R/9		3390-L	
	CC SESSION	XRC SESSION	CC SESSION	XRC SESSION
CC only	16	–	6	–
CC + XRC2	15	1	5	1
CC + XRC3	13	1	3	1

(d) HRC/HODM
• HXRC cannot be used for the same volumes using HODM pair volumes.
• HODM cannot be used for the same volumes using HXRC pair volumes.
• HXRC must be for that volumes after deleting HODM pair volumes.

3.16.2.2 Hardware

(1) HXRC Support Hardware Specification.

Table 3.16.2.2-1 HXRC Support Hardware Specification

CU Type	3990-6/6E/2105 (*1)
DEV Type	3390-1/2/3/3R/9/L
DKC model	Primary:RAID400 Secondary:RAID400/RAID300/RAID200HA/DKC80/DKC90
RAID level	RAID5/RAID1
Channel	ESCON/FICON

(*1) Do not intermix of DKC emulation type '2105' and 3990-6/6E in the same DKC.
If you change DKC emulation type '2105', the following operation.
- Delete All CC/XRC pairs
- Change DKC emulation type 2105 of All CHE PK
- RESUME CC/XRC pairs

(2) CACHE SIZE

Cache capacity should be doubled from current cache size.
(The amount of Sidefile data may occupy up to 60% of total cache capacity.)

3.16.2.3 Micro-program

(1) HXRC Support from 1'st version of Main Frame Micro-program.
(2) CNT extender version 4.9 or upper level code is recommended.
(3) HXRC support Modes

Table 3.16.2.3-1 HXRC support Modes

Mode	Description
Mode45	Sleep wait suppression (see modes 61, 85, 86, 97) When mode45 = ON and 61 = ON, write I/Os for LDEV are blocked by threshold specified by SDM. Mode45 OFF : Sidefile threshold activates sleep wait timer at sleep wait threshold. Mode45 ON : Sidefile threshold does not activate sleep wait timer at sleep wait threshold.
Mode61	Enables 'DONOT BLOCK' option of XADDPAIR command. (see mode 45) Should be holding mode61 = OFF, if the SDM does not support the function. Mode61 OFF : 'DONOT BLOCK' option is ignored. Mode61 ON : 'DONOT BLOCK' option is activated.
Mode 85, 86	Variable Sidefile threshold. (see modes 45, 97, 98) Mode85 86 ON, OFF : Threshold sleep wait/SCP/Puncture = 30/40/50% Mode85 86 OFF, OFF : Threshold sleep wait/SCP/Puncture = 40/50/60% Mode85 86 OFF, ON : Threshold sleep wait/SCP/Puncture = 50/60/70% Mode85 86 ON, ON : Threshold sleep wait/SCP/Puncture = 60/70/80%
Mode97	Variable sleep wait timer duration. (see modes 45, 85, 86) Mode97 OFF : Sleep wait timer duration = 100ms at sleep wait threshold. Mode97 ON : Sleep wait timer duration = 10ms at sleep wait threshold.
Mode98	Selectable SCP or session cancel at SCP threshold. (see modes 45, 85, 86) Mode98 OFF : SCP activates at SCP threshold. Mode98 ON : Forced session cancel at SCP threshold.
Mode118	Warning SIM reporting at sleep wait threshold. (see modes 45, 85, 86, 97) Mode118 OFF : Not report SIM. Mode118 ON : Reports SIM. (30sec interval over sleep wait threshold)

(4) Device Blocking Function and Load Balancing Control
DKC does not block write I/Os for the logical device which is specified DONOTBLOCK option not to affect performance impact for application programs.
Device blocking function and new load balancing control are supported from DKCMAIN micro rev. 01-12-00 or upper revision.

<Requirements>
It is necessary the following conditions to activate DONOTBLCOK option.
For Operating system
- The operating system should support DONOTBLOCK option.

For RAID system
- Set system option MODE61 = ON for DONOTBLOCK option.
  DKC performs current load balancing control, if MODE61 = OFF (default).
- Should be holding MODE61 = OFF (default), if the operating system does not support the function.

3.16.2.4 HXRC recommendations

(1) Recovery site CPU is the most ideal location for Data Mover.

(2) Data Mover's path should be utilized only to read Sidefile.

(3) Subsystem configurations
- Cache capacity : Should be doubled from current cache size.
- Confirmation for Number of channel paths for system data mover (SDM)
- Confirmation for Work loads for the subsystem

(4) Utility device for primary volume
- Should be prepare for each XRC session.
- A low activity device should be selected as a Utility Device
- Utility Device should be specified at 1'st time before establishing pair volumes.

(5) System Data Mover (SDM)
- Confirmation for PTF levels
- No record found problem. : APAR # OW30183, OW33680
- Necessary tuning for SDM data set : Capacity, Geometry of the data set

(6) DB2
- Broken VSAM index file problem : APAR # II08859

(7) Others
- CPU MIPS : Enough for HXRC environment
- LINE CAPACITY : Enough for HXRC environment with channel extender.

(8) HXRC with FICON

Table 3.16.2.4-1 HXRC and FICON configuration

		Record set transfer path (System Data Mover - DKC)	
		FICON	ESCON
Application site (System and DKC)	ESCON	Supported	Supported
	FICON	Supported	Not recommended (*1)

*1: If the path of Application site is FICON, System Data Mover (SDM) path should be also FICON in order to balance the performance of Application path and SDM path.

(9) HXRC

If the ANTA5107E (RC=9014 REAS=604 OR REAS=608) console message is displayed the XADDPAIR operation for HXRC pairs, the operation might be unsuccessful.
In this case, you may check the HXRC Option Installed. If HXRC Option not install please install HXRC option.

Before DKC Main VER.01-16-XX, Hitachi-Extended Remote Copy Option effects 3990 and 2105 dkctype.
After DKC Main VER.01-17-XX, Hitachi-Extended Remote Copy Option effects only 2105 dkctype.
After DKC Main VER.01-17-XX, You can use Hitachi-Extended Remote Copy for 3990 dkctype without this option.

Caution:
In this case, the following restriction applies if using XRC for 3990 dkctype without this option and DKCMAIN MICRO VER DOWN TO before 01-16-XX.

To apply XRC volumes, take the following steps:
1. Delete all the XRC pairs.
2. VER DOWN DKC MAIN MICRO PROGRAM.
3. Install the Hitachi-Extended Remote Copy Option.
4. Re-establish XRC pairs.

3.16.3 Online Maintenance while Concurrent Copy(CC)/HXRC using

(1) Availability of Installation and DE-installation.

Component	Maintenance Type	During initial copy		Established		Suspend	
		Primary	Secondary	Primary	Secondary	Primary	Secondary
HDD canister	Installation	*	x	*	x	*	x
	De-installation	*	x	*	x	*	x
Cache PCB	Installation	*	x	*	x	*	x
	De-installation	*	x	*	x	*	x
CHA	Installation	x	x	x	x	x	x
	De-installation	x	x	x	x	x	x
DKA	Installation	x	x	x	x	x	x
	De-installation	x	x	x	x	x	x

x: Maintenance is available.

*: Maintenance is available but it should take place when workload is low. The following are recommendations.
When a maintenance operation is needed while CC/HXRC is using, I/O's for CC/HXRC pair volumes or CC/HXRC itself should be stopped before the maintenance operation.
If the maintenance operation must be done while CC/HXRC is using, you must confirm that the usage of Sidefile monitor less than 20% of total Cache capacity before you start the maintenance operation. Only when the usage of Sidefile monitor is less than 20% of total Cache capacity, you can proceed the maintenance operation.
Refer to "Monitoring" in the SVP SECTION about Sidefile monitor.
Select the [Information] icon in the 'SVP' window.
Next select the [Monitor] menu in the 'Information' window and select [start...].
Next select the 'Sidefile' box in the 'Item' menu in the 'Monitoring' window and select [OK].

(2) Availability of the System tuning .
When the following System tuning operation is needed while CC/HXRC is using, CC/HXRC should be stopped before the System tuning operation.
- It is impossible to change the DKC No, SSID, or DKC Emulation type by System tuning operation while CC/HXRC is using.
- When the DRV emulation type of CC/HXRC pair volumes are 3390-3 or 3390-3R, it is impossible to change the emulation type between 3390-3 and 3390-3R by CHANGE EMULATION operation while CC/HXRC is using.

(3) Availability of the Replacement.

Component	Maintenance Type	During initial copy		Established		Suspend	
		Primary	Secondary	Primary	Secondary	Primary	Secondary
Logical Device	Blockade	**	**	**	**	**	**
	Recovery	**	**	**	**	**	**
	Format	**	**	**	**	**	**
	Verify	x	x	x	x	x	x
HDD canister	Replace	x	x	x	x	x	x
Cache PCB	Replace	*	x	*	x	*	x
CHA	Replace	x	x	x	x	x	x
DKA	Replace	x	x	x	x	x	x
LTM PCB	Replace	x	x	x	x	x	x

x: Maintenance is available

*: Maintenance is available but it should take place when workload is low. The following are recommendations.
When a maintenance operation is needed while CC/HXRC is using, I/O's for CC/HXRC pair volumes or CC/HXRC itself should be stopped before the maintenance operation.
If the maintenance operation must be done while CC/HXRC is using, you must confirm that the usage of Sidefile monitor less than 20% of total Cache capacity before you the start maintenance operation. Only when the usage of Sidefile monitor is less than 20% of total Cache capacity, you can proceed the maintenance operation.
Refer to "Monitoring" in the SVP SECTION about Sidefile monitor.
Select the [Information] icon in the 'SVP' window.
Next select the [Monitor] menu in the 'Information' window and select [start...].
Next select the 'Sidefile' box in the 'Item' menu in the 'Monitoring' window and select [OK].

**: When a maintenance operation is needed while CC/HXRC is using, CC/HXRC should be stopped before the maintenance operation.

3.17 HRC Asynchronous

3.17.1 Components

Asynchronous mode is one of the update copy modes of the HRC volume pairs. The HRC asynchronous subsystem consists of the same components as HRC synchronous with the following exceptions:

- A set of the HRC volume pairs named consistency group is introduced.
- Only 1-to-1 and n-to-1 (n≤4) configuration is supported
- For n-to-1 configuration, the XRC time-stamping capability is required.
- A communicating facility to transfer error information from the primary system to the secondary system is not required.

Note) In this document, the term n-to-m means that n-MCUs and m-RCUs are connected to each other to establish the HRC volume pairs. N and m is number of control units of physical unit bases, not of control unit image bases.

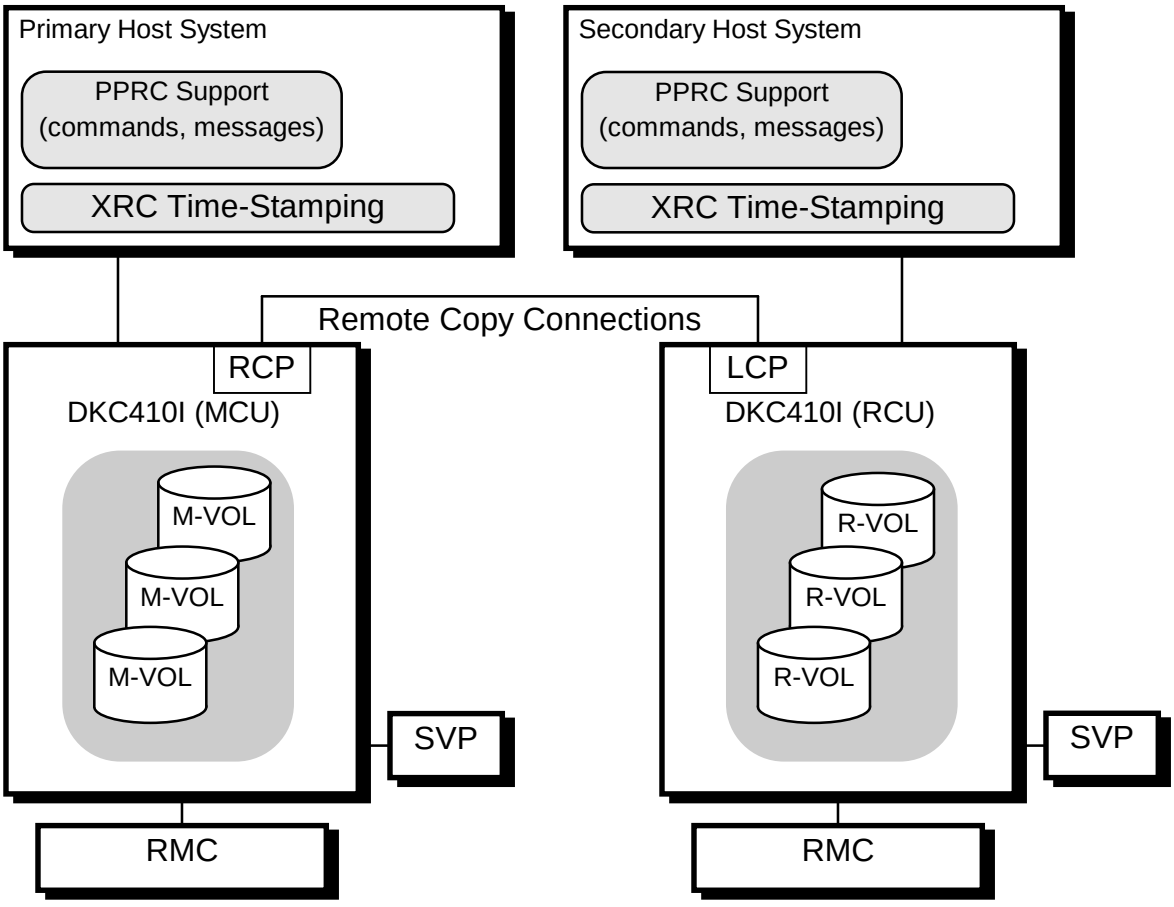


Fig. 3.17.1-1 HRC Asynchronous Subsystem Components

(1) MCU and RCU

- Both MCU and RCU must be RAID300 or upper model.
- Maximum configuration is 4-to-1. For n-to-1 (n >1) configuration, XRC time-stamping capability is required.
- To use XRC time-stamping capability, control unit emulation must be 3990-6 basic or enhanced mode.
- Maximum number of control unit image pairings (established by Add RCU or ESTPATH) is 16.

(2) Consistency Group

- HRC asynchronous ensures update-sequence-consistency across several volume pairs. It also provides some group-based operations. A set of volume pairs treated by such group-based functions is called a consistency group.
- HRC asynchronous supports 16 consistency groups at maximum. Every HRC asynchronous volume pair belongs to one consistency group.

(3) PPRC Support

- Although HRC asynchronous is not fully compatible to PPRC, it can be controlled and monitored with PPRC host facilities, PPRC TSO commands, ICKDSF PPRC commands and some console messages. For this purpose, MVS/DFP 3.2.0 or higher level and ICKDSF release 16 or upper are available.
- Only fundamental facilities are available. Neither P/DAS SWAP nor CGROUP is supported.
- If the primary system (and the secondary system) consists of several CPU complexes, SYSPLEX timer must be installed for the common time reference.

(4) XRC Time-Stamping Capability

- In case of N-to-1 configuration, the XRC time-stamping capability requires to be installed in the primary host system. MVS/DFP 3.2.0 or higher level is required.
- In order to get benefit of time-stamping capability during copy-back process (pair establishment from the secondary to the primary subsystem), the XRC time-stamping capability recommends to be installed in the secondary system.
- If the primary system (and the secondary system) consists of several CPU complexes, SYSPLEX timer must be installed for common time reference.

3.17.2 Consistency Group

(1) HRC Asynchronous Volume Pairs and Consistency Group

- Every HRC asynchronous volume pair belongs to one consistency group.
- When establishing the HRC asynchronous volume pair, an operator specifies the consistency group number that the volume pair will belong to with a new parameter of Add Pair and ESTPATH.
- The consistency group must be registered prior to pair establishment.

(2) Functions of Consistency Group Basis

(a) Ensuring Update Sequence Consistency

- The updated records are copied to the corresponding R-VOLs in the same order as the M-VOLs have been updated by the primary host systems.
- The update sequence consistency is ensured within a consistency group. The updated records of the different consistency groups may be copied in the different order from the original.

(b) Suspending Volumes Pairs (Error Level)

- When one R-VOLs is not updated correctly due to the failure, all the HRC asynchronous volume pairs will be suspended with keeping update sequence consistency.
- When establishing the HRC asynchronous volume pair, an operator specifies whether other volume pairs will be suspended together or not against the failure of the volume pair. A new parameter called **Error Level** is defined for this purpose.

Error Level = Group	When this volume pair is suspended due to the failure, all volume pairs in the same consistency group will be suspended together.
Error Level = Volume	Even if this volume pair is suspended due to the failure, other volume pairs in the same consistency group will not be suspended, as long as the failure prevent.

(c) Providing the Consistency Time

The latest time stamp value of the update that has been successfully copied to the R-VOL is called a **consistency time**. The consistency time is a group basis indication. It means that all the updates performed before or at the consistency time have been successfully copied to the R-VOLs in the consistency group.

- The consistency time can be displayed with the following operations issued to the R-VOL of duplex or suspended state. If 'LOCAL' is specified for timer type, Consistency time is not displayed.
  - RMC (or SVP) Pair Status panel
  - PPRC CQUERY command (only at suspended state.)
- At the R-VOLs in duplex state, the consistency time is a ticking value. Any R-VOL displays the consistency time in that instance. It can be used for feeling how long the R-VOLs are behind the M-VOLs.
- Whenever the volume pair is suspended, the consistency time of the R-VOL is frozen.
  - ① If the update sequence consistency between the R-VOL and other R-VOLs in the consistency group is ensured, the R-VOL indicates the latest consistency time of the consistency group.
  - ② Otherwise, the R-VOL indicates the latest time stamp value of the update that has been successfully copied to the R-VOL. It may be older than other R-VOLs because the consistency time of the consistency group is still ticking.
- If the R-VOL is in suspended state, the supplementary status that indicates whether the consistency time is of the consistency group (case (a) above) or the R-VOL (case (b) above) is also displayed.

(d) Consistency Group Basis Operations

In order to make the disaster/failure recovery procedure simple, the following consistency group basis operations are provided.

• Operations at the RCU

① Deleting all suspended volume pairs except for inconsistent volume pairs
② Deleting all volume pairs regardless of the consistency among them
③ Suspending all volume pairs

• Operations at the MCU

① Suspending all volume pairs
② Deleting all volume pairs behind this unit (except for M-VOLs behind other MCU)
③ Resuming all suspended volume pairs behind this unit (except for M-VOLs behind other MCU)

(3) Configuration of the Consistency Group

(a) Disposition of the Volume Pairs

• All R-VOLs that belong to the same consistency group must be located behind one RCU.
• The M-VOLs that belong to in the same consistency group can be located behind up to 4 different MCUs.
• Up to 16 consistency groups can be established within one pair of MCU and RCU. The RCU supports up to 16 consistency groups.
• The R-VOLs of the different consistency groups can be located behind the different RCU.
• Up to 1,024 volume pairs can belong to one consistency group.

(b) Primary Host Systems and Consistency Group

1) Primary host systems and timer type

- Every update I/O to the M-VOL of the same consistency group must be time-stamped by using common timer facility. Table 3.17.2-1 shows the relationship between the primary host system and available timer resource.
- When registering the consistency group, an operator must specify which timer resource should be used for the consistency group based functions.

Table 3.17.2-1 Primary host systems and timer resource

Primary Host System	Timer Resource	Notes on Configuration
MVS with the XRC time-stamping capability	System timer (CPU TOD clock)	• N-to-1 (N≤4) configuration is possible. • SYSPLEX timer must be installed if the primary system consists of the several CPU complexes.
Other main frame host systems	Local timer (MCU internal clock)	• Only 1-to-1 configuration allowed. • The consistency time is not displayed.
Open host systems		

2) Restrictions and notes on the primary host systems

- The primary host systems can not access the volume pairs of the same consistency group unless they have the common timer reference.
- The M-VOLs updated by the same primary host system can belong to the same or different consistency group if the M-VOLs have no requirement on update sequence consistency (i.e. they are updated independently of each other.) However it is recommended for them to belong to the different consistency groups because, for example, they might be suspended together against the failure
- Because of the same reason, the independent M-VOLs accessed by the independent primary host systems is recommended to belong to the different consistency groups, even if they can use common timer reference.

3.17.3 HRC Asynchronous Theory of Operations

(1) Update Copy

The updates from the primary host systems and additional control information are queued in the cache storage of the MCU, and sent to the RCU independent of host I/O processes. The RCU stores the data and control information into the provisional spaces allocated in the cache storage. According to the time-stamp and the sequence information, the RCU promotes the updates in the provisional spaces the formal data of the R-VOLs in the same order as they have been performed at the MCU.

(a) Receiving Time-stamp Information

In case of the Timer Type of System specified, the MCU receives the time-stamp information as follows:

- When directed to establish the HRC asynchronous volume pair, the MCU reports the state-change-interrupt (SCI) to all the attached host systems. The host system issues a series of sense group commands to recognize what status of the device has changed. The MCU generates the response as if the device became a member of an XRC session. This response activates the XRC time-stamping capability if installed in the host systems.
- Once activated, MVS IOS routine attaches the time-stamp information (contents of time-of-day clock) to each I/O operation to read and write the device. The time-stamp information indicates when the corresponding update has been issued at the primary host system. It is transferred to the MCU at the beginning of each I/O operation.

(b) Creating Recordset

- When accepting the updates from the primary host systems, the MCU creates a set of information called a **recordset**. A recordset includes:

  ① updated record
  ② time-stamp information received
  ③ sequence number
  ④ record locations (device, cylinder, track and record number) and record length

- The **sequence number** is the number of recordsets the MCU has created for the consistency group. That is, all recordsets in each MCU and each consistency group are independently numbered.
- The recordset information other than the updated records is stored and queued into the exclusive spaces allocated in the cache storage.
- The updated records are stored as the host-dirty data and do not occupy the exclusive space until the following events happen before the recordset is sent to the RCU
  — The same record is updated again, or
  — The host-dirty status is removed by the de-staging process.
  When the above mentioned event happens, the MCU moves the updated records into an exclusive space, called **Sidefile,** in the cache storage.

(c) Sending Recordset to the RCU

- The MCU sends the recordset in a similar manner as HRC synchronous. That is, the MCU and RCP port act as the host processor channel and issue I/O operation, called Remote I/O(RIO), to the RCU.
- The RIO transfers the time-stamp information, the sequence number, the record locations and length, and the updated records in the FBA format (not in the CKD format) by one channel command, like HRC synchronous. However the parameter length and detailed specification of this channel command is different from HRC synchronous. Therefore the micro code of the store-and-forward type channel extender (i.e. Channelink and UltraNet of CNT corp.) should be upgraded to support the command.
- Unlike the recordset offloading of the XRC Data Mover, the MCU sends the recordset with directly specifying the device address of the R-VOL. The RIO independently activates each R-VOL. Furthermore, the MCU may send several recordsets by one RIO even if their sequence numbers are not contiguous to each other. Therefore the recordsets are usually sent in different order from their arrivals to the MCU.

(d) Storing Recordset into Sidefile Space

- The RCU stores the received recordsets into the spaces exclusively allocated in the cache storage. The exclusive space is called a Sidefile. The updated records in the Sidefile are not treated as the formal data. That is, the host I/O processes and the de-staging processes do not access the records in the Sidefile at this time.
- The records in the Sidefile will be promoted the formal data later. A term **settle** means to promote the records in the Sidefile.
- The RCU also allocates exclusive spaces in the cache storage so that the Sidefiles form a queue.
  - The RCU makes a queue per MCU and per consistency group.
  - This queue is not of the FIFO fashion. Each entry of the queue is previously assigned to each sequence number. The arrived recordsets are queued into the corresponding entries with indexed by their sequence number. The entries for the recordsets that do not arrive yet are left empty. As a result, the RCU lines up the recordsets in the order of their sequence number.

(e) Selecting Recordset and Deciding Consistency Time

- The RCU selects the recordset to be settled with the following algorithm.

  ① Checks if there are the valid entries at the top of all the queues in the consistency group. If one of them is empty, the RCU waits for the entry.
  ② When all the top entries are filled with the valid Sidefiles, the RCU selects one entry that has the smallest time-stamp. It can be settled.
  ③ Repeats step 1 and 2.

- Figure 3.17.3-1 shows an example. All the top entries are filled with the recordsets of S11/T3, S21/T2, S31/T1 and S41/T5. The RCU selects the recordset of S31/T1 to be settled because the T1 is the smallest time-stamp. Then the top entry S31/T1 is removed from the MCU3's queue. The S32 becomes the top but it is empty. The next recordset S11/T2 will be selected when recordset S32 arrives and its time-stamp is smaller than T2.

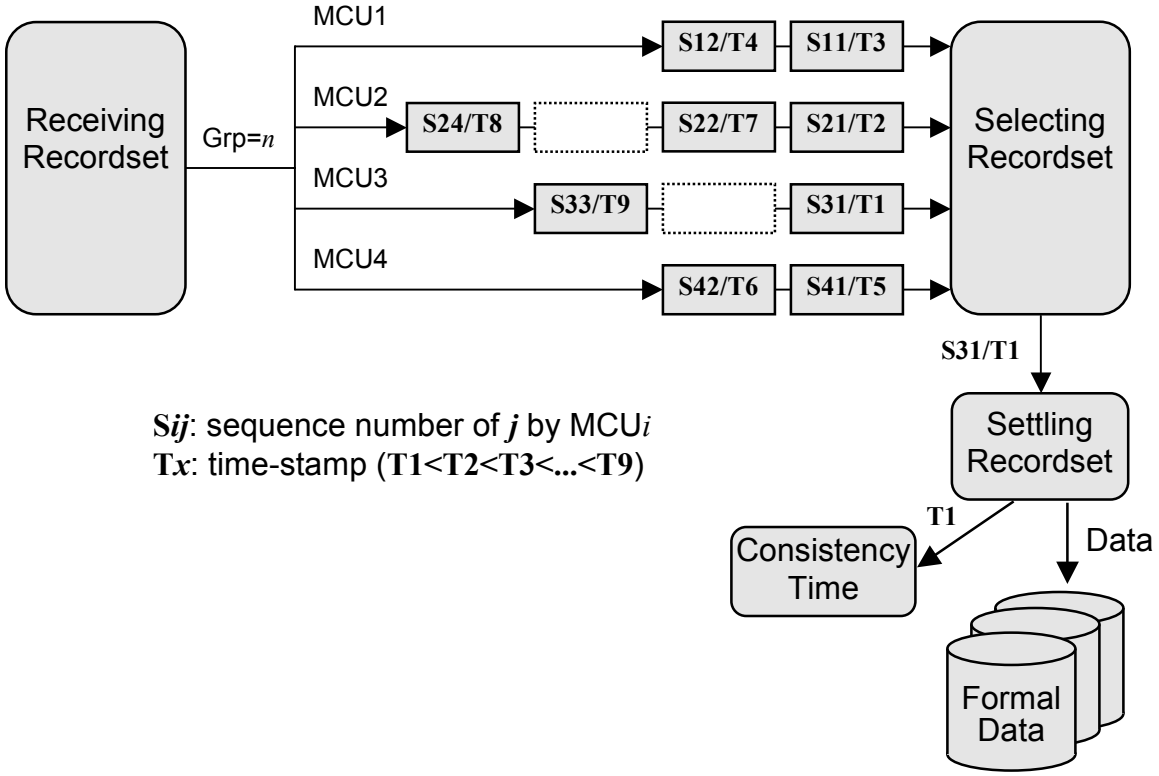


Figure 3.17.3-1 Example of Selecting Recordset at the RCU

(f) Settling Recordset

- The recordset selected by the algorithm described in section(e) will be marked as host-dirty and treated as the formal data after that time. The time-stamp value of the recordset is promoted to the consistency time.
- The RCU settles the updated records in the recordset as follows:
  — If the corresponding track is not in cache storage (track-miss), the cache directory of the Sidefile is changed to be the formal data. No data is moved.
  — If the corresponding track is in cache (track-hit), the updated records in the recordset are copied to the existing cached-track and the cache space for the Sidefile is released.

3.17.4 MCU-RCU Communications to Maintain Asynchronous Copies

(1) Dummy Recordset

The RCU needs to receive the recordset continuously from all the MCUs even if the MCU does not have to create the new recordset.

- The MCU creates and sends a dummy recordset when it has received no update I/O in a second. The dummy recordset contains only the sequence number and the time-stamp information. Contents of the time-stamp information is generated by incrementing the largest time-stamp that the RCU has (the MCU reads it from the RCU before creating the dummy recordset) by one.
- The RCU receives the dummy recordset and puts it into the queue. It can help other recordsets being selected.

Another purpose of the dummy recordset is have the RCU be aware of the disaster. If the RCU can not receive any recordset in the predetermined duration (this time can be specified by Maximum Copy Delay Time parameter), the RCU regards such situation as the disaster and suspends all the HRC asynchronous volume pairs.

Due to these reasons, the MCU and RCU always need to communicate with each other once the HRC asynchronous volume pair established and not all the pairs have been suspended.

(2) Change-Status Recordset

When the volume pairs are suspended or deleted due to operations or failure, the update sequence consistency must be ensured. In order to meet this requirement, the negotiation on changing volume pair status is made by means of recordset. The recordset used for this purpose also has only the sequence number and the time-stamp information.

(3) PS OFF Notification Recordset

At the power-off sequence, the MCU creates and sends the recordset to notify of the power-off event. The recordset used for this purpose also has only the sequence number and the time-stamp information.

3.17.5 Failure Detected by the RCU

(1) Pair Suspend by the RCU

Table 3.17.5-1 shows the failures detected by the RCU and the volume pairs to be suspended due to the failure.

Table 3.17.5-1 Volume Pairs to be Suspended by the RCU-detected Failure

Failures	Volume pairs to be suspended
The RCU could not settle the pending recordset or could not communicates with the MCU before the maximum copy delay time expired.	All pairs in the consistency group
The RCU could not receive the recordset successfully due to the hardware failure.	All pairs in the consistency group, or only the affected pair (depending on the failure)
The RCU detected the logical error while selecting the recordset to be settled.	All pairs in the consistency group
The RCU could not settle the recordset due to the hardware failure, the track condition, or the logical error.	All pairs in the consistency group, or only the affected pair (depending on the failure)

(2) Pair Suspended and Re-synchronization

For the HRC asynchronous volume pairs, both the MCU and RCU maintain the bit map for pair re-synchronization. When/after the volume pair(s) suspended, the cylinders that contain the following records are marked in the bit map as modified(to-be-copied later):

- The recordsets that have been created by the MCU but not sent the RCU yet. After marking the cylinders as modified, the recordsets are discarded.
- The recordset that have reached at the RCU but not settled yet. After marking the cylinders as modified, the recordsets are discarded.
- The records updated by the primary system after the volume pair(s) suspended

At the beginning of the pair re-synchronization, the contents of the RCU's bit map are sent to the MCU and merged into the MCU's bit map. The MCU performs the initial copy for the pair re-synchronization according to its bit map. That is, the cylinders that contain the lost recordset are re-synchronized at this time.

3.17.6 Inflow Control for Sidefiles

- As described in section 3.17.4, both the MCU and RCU create the Sidefiles for storing the recordsets. Since the Sidefile is an exclusive space in the cache storage, both the MCU and RCU perform the inflow control to prevent the subsystem overload.
- Both the MCU and RCU use the threshold value specified with the RMC/SVP panel.

(1) Inflow Control by MCU

- When the amount of Sidefiles reaches at the threshold, the MCU responds to the update I/Os from the primary system with the state-change-pending (SCP) or channel-command-retry request.
- If no recordset has been sent to the RCU after the specified time duration, the MCU will suspend all the volume pairs and reset the SCP condition in order to avoid the system being hung up.

(2) Inflow Control by RCU

- When the amount of Sidefiles reached at the threshold, the RCU responds to the command that transfers the recordset from the MCU with the channel-command-retry request. Only the recordset of the sequence number necessary to continue settling the pending recordsets is accepted.
- If the recordset has not been settled after the specified time duration, the RCU will suspend all the volume pairs and reset the channel-command-retry condition in order to avoid the MCU being hung up.

3.17.7 HRC Asynchronous Control Operations

This chapter describes the RMC (or SVP) operations for HRC asynchronous.

3.17.7.1 DKC Options

(1) *Async Option* - Modifying HRC Options on Physical Unit Basis

- Async Option panel provides the function to modify asynchronous options.
- These options are effective to entire physical control unit (i.e. all M-VOLs and R-VOLs behind the control unit.)
- These options can be modified when no asynchronous volume pair is established.
- These options may be modified before/after performing Port, Add RCU, and Add Group operations.

Table 3.17.7-1 Async Options on Physical Control Unit Basis

Option Name	Description
Pending Update Data Rate	• It specifies the amount of cache storage in percent that allows to be used for storing recordset (Sidefile). • When the amount of cache storage for the recordset reaches the specified threshold, the MCU and RCU invokes its own inflow control as follows: — The MCU responds to the update I/Os from the primary system with the state-change-pending (SCP) or channel-command-retry request. — The RCU responds to the command that transfers recordset from the MCU with the channel-command-retry request. However the specific recordset that will help the RCU settle the pending recordset is still accepted. • Any percent between 30% and 70% can be specified with a unit of 10%. The default is 50%.
Offloading Timer	• It specifies how long the MCU can continue the inflow control described above. • The MCU stops the inflow control and suspends all asynchronous volume pairs after the specified time expires unless no recordset has been offloaded to the RCU. • Every minute between 1 to 20 and “None” can be specified. If “None” is specified, the MCU will immediately become suspend, when the Sidefile threshold is exceeded.

Notice : If Pending Update Data Rate is modified when asynchronous volume pair is established, the host I/O timeout may occur.

3.17.7.2 Registering/Monitoring/Deleting the Consistency Group

(1) *Add Group* - Registering Consistency Groups

- The consistency group can be registered with RMC (or SVP) attached to the MCU.
- The consistency group must be registered prior to the volume pair establishment.
- The consistency group has its own attributes and parameters, consistency group number, timer type, and others. They are specified when registered.
- The consistency group is registered in the RCU too. However it is not necessary to be specified. When the volume pair is established, the MCU directs the RCU to register the consistency group.

(a) Consistency Group Number

- The consistency group number is described with one digit of hexadecimal character.
- The volume pair control operations require the consistency group number as the parameter. The pair status displayed by Pair Status operation of RMC and CQUERY command also includes the consistency group number.

(b) Timer Type

- The timer type must be specified out of System, Local and None when the consistency group is registered.

Table 3.17.7-2 Timer Type Attributes

Timer Type	Meaning
System	The system timer (CPU TOD clock) provided by the XRC time-stamping capability is used for controlling this consistency group.
Local	The local timer (*internal* TOD clock of this MCU) is used to for controlling this consistency group.
None	The system timer (CPU TOD clock) provided by the XRC time-stamping capability is used for controlling this consistency group. The R-VOLs in this consistency group can be located behind the different RCUs. *However the update sequence consistency across the RCUs is not ensured*. This timer type should be selected only when volume pairs are established from the original secondary to the original primary volumes (*copy back*).

- Table 3.17.7-3 shows the related configuration and timer type to be specified.

Table 3.17.7-3 Timer Types to Be Specified

Configuration			Timer type to be specified	
Host system	XRC time-stamping capability	MCU-to-RCU	For *P-to-S* copy (original direction)	For *S-to-P* copy (copy back)
Main frame	Installed	N-to-1 (n >1)	System	None
		1-to-1	System	System
	Not installed	1-to-1	Local	Local
Open systems	(unavailable)	1-to-1	Local	Local

(c) Timeout Parameters

• The following parameters can be specified to modify the expiration time for the timeout event of the consistency group basis.

① Maximum copy delay time

Table 3.17.7-4 Maximum copy delay time to Be Specified

Name in the RMC panel	Time Out: Write Pending [min]
Available range	3 min. to 15 min. or "None"(no time out event occurs)
Default	5 min.
Description	It specifies the maximum delay allowed for asynchronous copy. Based on this parameter, the RCU will suspend all R-VOLs if following time out event occurs: • The RCU has received the updated data but it can not be settled in the specified time. • The RCU has had no communication from the MCU until the specified time expires

Note.) The RCU stores the updated data received into a *provisional* space of cache storage, and will make it available for use later. The term *settle* means to making it available for use.

When the User of SE use ASYNC HRC function in N-to-1 configuration, they must reset the maximum copy delay allowed with the notice of as follows.
If not take, it may cause suspension of HRC pairs.

(1) Execute ASYNC HRC with the maximum delay allowed = 'NONE".
(2) You can recognize the current copy delay with the difference the Time Stamp of HOST I/O and the Consistency Time of "Group Status" in RMC (SVP).
(3) Execute "Suspend Pair" for all pairs on the CT Group by RMC (SVP), and reset the maximum delay allowed with over the current copy delay. If it is longer than maximum time (15 min.), reduce the HOST I/O rate, or you should leave it "NONE".
(4) Restart ASYNC HRC with "Resume Pair" for all pairs on the CT Group by RMC (SVP).

② Maximum RCU-ready-wait time

Table 3.17.7-5 Maximum RCU-ready-wait time to Be Specified

Name in the RMC panel	Time Out: RCU Ready [min]
Available range	1 min. to 10 min or "None"(no wait)
Default	5 min.
Description	During the power-on-reset procedure, the MCU intends to communicate with the RCU and will suspend all the volume pairs if it can not communicate until the specified time expires.

(2) *Delete Group* - Deleting Consistency Group Registration

- If no volume pair belongs to the consistency group, the group can be deleted.
- This operation is available only at the MCU. The registration to the RCU is automatically deleted, when the last volume pair in this group is deleted.
- In N-to-1 (N >1) configuration, deleting the consistency group does not affect the consistency group that has been registered in another MCU.

(3) *Group Option* - Modifying HRC Options on Group Basis

- This operation allows MCU to delete the consistency group currently registered.
- The Group Option can be operated only when no volume pair belongs to this group.

(4) *Group Status* - Displaying Consistency Group Status

- The options and the working status can be displayed on this panel.

Table 3.17.7-6 Consistency Group Status

Item	Contents	Displayed by:	
		MCU	RCU
Consistency group number	Consistency group number in one digit of hexadecimal character.	Yes	Yes
RCU serial number/SSID	Serial number and SSID of the RCU that belongs to this group.	Yes	No
Volume list	List of volumes that belong to this group and are behind this control unit.	Yes (*1)	Yes
Consistency time	Current consistency time of this group.	Yes (*2)	Yes
Timer type	Specified timer type of this group.	Yes	Yes
SEQCHK (*3)	At least one volume pair of this group has the SEQCHK status.	Yes (*2)	Yes
Maximum copy delay time	Specified maximum copy delay time.	Yes	Yes
Maximum RCU-ready-wait time	Specified maximum RCU-ready-wait time.	Yes	No

(*1) In N-to-1 (N >1) configuration, it does not include the M-VOLs behind other MCU.

(*2) The consistency time and SEQCHK status are decided by the RCU. The MCU displays these items after reading them from the RCU. If the MCU can not communicate due to communication failure, the latest contents are not displayed. Therefore these items should not be used for disaster recovery.

(*3) SEQCHK is one of the R-VOL statuses.

3.17.7.3 Pair Status

- Suspending and Deleting state are newly defined to indicate the status in transition.
- Suspended by MCU powered-off is added as the caused of suspension.
- To indicate whether the update sequence consistency is kept or not, the *subsidiary pair status* (Group or Volume) and SEQCHK indicator are newly defined.

(1) Status in Transition - *Suspending* and *Deleting*

When suspending or deleting the HRC asynchronous volume pairs, the MCU and RCU intend to process all pending recordset before changing pair status. It takes longer time for the asynchronous volume pairs to change to suspended or simplex state than the synchronous volume pairs. Therefore adding to the conventional pair statuses (simplex, pending, duplex, suspended), two new statuses are introduced.

(a) Definitions and conditions of transition

Suspending This volume pair is *in transition from duplex or pending to suspended state*. When cause of suspension (failure or operation) is detected, all affected volume pairs change to suspending state. After completing suspension, they will automatically change to suspended state.

Deleting This volume pair is *in transition from duplex, pending or suspended to simplex state.* When accepting delete pair operation, all affected volume pairs change to deleting state. After completing delete pair operation, they will automatically change to simplex state.

(b) Indication of pair status

- Suspending and deleting statuses can be indicated *only* on RMC (or SVP) main control and pair status panels.
- For main frame host systems, these states are not indicated. Status in transition is treated as follows:
  - — In case of operations (suspend pair or delete pair), status is not changed until transition completes. After completing status transition, affected volume pairs are changed to suspended or simplex state.
  - — In case of failure, affected volume pairs are changed to suspended state when cause of suspension is detected.
  - — In any cases, the MCU or RCU report the state change interrupt (SCI) after completing status transition. IEA491E or IEA494I console messages appear on the system console at this time.

(2) Suspended by MCU Powered-Off

When the MCU is being powered off, the MCU suspends all volume pairs behind it. The suspended volume pairs will automatically return to their original state (duplex or pending) when the MCU is powered-on again. During this suspension, the suspended R-VOLs indicate by MCU powered-off as the cause of suspension.

(a) Conditions of transition

- During power-off sequence, the MCU notifies the RCU of the power-off event. The RCU changes all related R-VOLs to suspended states and sets by MCU powered-off as the cause of suspension.
  - Only R-VOLs which are in duplex or pending state are affected. The cause of suspension of the R-VOL that is already suspended is not changed.
  - Only R-VOLs of which corresponding M-VOLs are behind the MCU are affected. The volume pairs between other MCU are not affected.

- During power-on sequence, the MCU notifies the RCU of the power-on event. The RCU changes all R-VOLs in suspended by MCU powered-off state to original (duplex or pending) state.

(b) Indication of pair status

- Only the RCU can indicate this cause of suspension.
- This cause of suspension is displayed as follows:
  - RMC (or SVP) Main Control panel: OFF in the Sub column
  - RMC (or SVP) Pair Status panel: Suspended (MCU PS OFF)
  - PPRC CQUERY command: SUSPENDED(5) with the specific indicator in the serial number field

(3) Consistency Status - *Group* or *Volume*

(a) Definition of the status

Group Update sequence consistency between this R-VOL and other R-VOLs in this consistency group is ensured. This R-VOL can be used for disaster recovery at the secondary system after deleting the HRC volume pair.
This status is indicated when:
— All volume pairs in this consistency group have been suspended due to the failure that affected the consistency group (not the specific volume pair.)
— The volume pair having Error Level = Group has been suspended due to the failure.
— This volume pair has been suspended by suspend pair operation with *Group* parameter.

Volume Probably this volume pair has been suspended alone. Update sequence consistency between this R-VOL and other R-VOLs in this consistency group is not ensured. This R-VOL can not be used for disaster recovery at the secondary system. This status is indicated when:
— This volume pair has Error Level = Volume and has been suspended due to the failure that did not affect entire group.
— This volume pair has been suspended by suspend pair operation with *Volume* parameter.

(b) Indication of the status

- Only the RCU can display this cause of suspension.
- This status has the meaning described above only when the R-VOL is in suspended state.
- This status is displayed as follows:
  — RMC (or SVP) Main Control panel: GRP or VOL in the Sub column
  — RMC (or SVP) Pair Status panel: GRP or VOL as “Suspended by:” item
  — PPRC CQUERY command: SUSPENDED(x) with the specific indicator in the serial number field

(4) Alert for Non-Time-stamped Updates - *SEQCHK*

SEQCHK indicates that the volume pair has accepted some updates without time-stamp attached while the consistency group is specified to use system timer.

(a) Set/reset conditions

When the RCU is settling the updated data, the RCU turns this indicator on if the updated data is not time-stamped, otherwise turns this indicator off.

(b) Indication of this indicator

- Only the RCU can display this indicator.
- This status can be indicated unless the volume is in simplex status.
- This status is displayed as follows:
  - RMC (or SVP) Main Control panel: SEQ (on) or blank (off) in the SEQ column
  - RMC (or SVP) Pair Status panel: SEQCHK (on) or blank (off)
  - PPRC CQUERY command: The specific indicator in the serial number field

3.17.7.4 Controlling Volume Pairs

- Basic operations to control/monitor the HRC asynchronous volume pairs are same as the HRC synchronous pairs. That is, at first the volume pairs to be operated must be selected on the main control panel, and then the operation and its parameters/options should be specified.
- For the HRC asynchronous volume pairs, some group basis operations are supported. Several volume pairs within the consistency group to which the selected volume belongs can be deleted, suspended, resumed by the group basis operations. For these operations, any volume pair within the consistency group may be selected.

(1) *Add Pair* - Establishing HRC Volume Pairs

- This operation has the MCU establish the HRC volume pair(s). Selected volumes will be the M-VOLs and the R-VOLs should be specified by the parameters.
- For asynchronous volume pairs, the MCU-RCU logical connection must be established and the consistency group must be registered prior to this operation.

(a) Basic parameters and pair options

The basic parameters and the pair options of Add Pair are described in and respectively. "Changed" column of these tables indicates the difference between HRC asynchronous and conventional HRC.

Table 3.17.7-7 Add Pair - Basic Parameters

Item	Parameters	Changed	Description
Pair configuration	- R-VOL - RCU	No	Specifies the serial number and SSID of the RCU and the logical device number of the R-VOL
Initial copy control	- Initial Copy - Priority	No	Specifies whether the initial copy for this volume pair is necessary ("Entire Volume") or not ("None") and its priority
Update copy mode	- Copy Mode	Yes	Specifies the update copy mote for this volume pair out of Synchronous and Asynchronous. For the asynchronous volume pair, specifies the consistency group number to which this volume pair will belong.

Table 3.17.7-8 Add Pair/Resume Pair - Option Parameters

Option Name	Changed	For:	Description
Initial copy pace	No	Any	Specifies the initial copy pace.
Fence Level	No	Synch.	Specifies the fence level out of "Data", "Status" and "Never" for synchronous volume pairs. This option is valid for the HRC synchronous volume pairs.
CFW data	No	Any	Specifies whether the CFW updates should be copied to the R-VOL or not.
DFW to R-VOL	No	HODM	Specifies the volume pair should be suspended when DFW is disabled/failed at the RCU. This option is valid only for HODM volume pairs.
Error Level	Yes (new)	Asynch.	Specifies whether all the volume pairs in the same consistency group should be suspended together or not when this volume pair is suspended. Group All the volume pairs group should be suspended together. Volume Only this volume pair may be suspended. This option is valid only for the HRC asynchronous volume pairs and the default is Group.
Pair Resume	Yes (new)	Asynch.	Specifies whether all the suspended volume pairs, which belong to the same consistency group and whose M-VOLs are behind this MCU, should be resumed together. Group All the volume pairs should be resumed together. Volume Only this volume pair may be resumed. This option is valid only for Resume Pair operation of the HRC asynchronous volume pairs and the default is Group (currently Volume).

(b) End conditions

The Resume Pair operation with the Group option gets the normal end condition before the initial copy for each volume pair begins. If some unusual conditions (pinned tracks, correction-access status of the parity group, etc.) prevents the volume pair from being re-established, the volume pair would be still in suspended status. Therefor the result of this operation should be confirmed by referring to the actual pair status or the console messages that indicate pair status change.

(2) *Delete Pair* - Deleting HRC Volume Pairs

- In order to make the disaster recovery operations at the secondary subsystem simple, a new option that specifies the volume pairs to be deleted is supported (***Delete*** option).
- The pair status of the asynchronous volume pairs will be Deleting when the operation completes, and then Simplex after the internal process completes.

(a) Volume pairs to be deleted

Table 3.17.7-9 Delete Pair - Delete Option

Option Name	Operable at:		Description
	MCU	RCU	
Consistent Volumes	No	Yes	Specifies that the volume pairs that meet the following conditions should be deleted. — Belongs to the same consistency group as the selected volume pair and; — Is in Suspended status and; — Has the consistency status of Group. Other volume pairs than described above are not deleted. This is the default when operated at the RCU.
Group	Yes	Yes	Specifies that the all volume pairs (but volume pairs behind other MCU when operated at the MCU) in the same consistency group should be deleted.
Volume	Yes	Yes	Specifies that only selected volume pairs should be deleted. This is the default when operated at the MCU.

- This option is valid only when the asynchronous volume pair is selected.
- "Consistent Volumes" and "Group" can be specified when only one volume pair is selected.

(b) Force option for asynchronous volume pairs

The Force option specifies that the volume pair should be deleted even if the MCU and RCU can not communicate with each other. For the asynchronous volume pairs, this option is still effective but can be specified with the Delete option of Group.

(c) End conditions

The Delete Pair operation for the asynchronous volume pairs gets the normal end condition when the operation is accepted by the MCU/RCU. The pair status of the volume pairs to be deleted are Deleting at this time. And then the pair status will be Simplex after the internal process (negotiation between the MCU and RCU) completes. Therefor the result of this operation should be confirmed by referring to the actual pair status or the console messages that indicate pair status change.

(d) Consistency time/status after deleted

Once the volume pair deleted, the consistency time/status of the volume pair will be reset. Therefor the consistency time/status should be memorized prior to this operation

(3) *Suspend Pair* - Suspending HRC Volume Pairs

- A new option that specifies the volume pairs to be suspended is supported (Suspend option).
- The pair status of asynchronous volume pairs will be Suspending when the operation completes, and then Suspended after the internal process completes.

(a) Parameters and options

Table 3.17.7-10 Suspend Pair - Parameters and Options

Item Name		Operable at:		Description
		MCU	RCU	
SSB (F/M=FB)		Yes	Yes	Specifies that IEA494E console message should be generated.
Suspend Kind		-	-	Only "R-VOL" can be selected for the asynchronous volume pairs.
Suspend				Specifies the volume pairs to be suspended.
	Group	Yes	Yes	Specifies that all volume pairs in the same consistency group as the selected volume pair should be suspended together.
	Volume	Yes	Yes	Specifies that only the selected volume pair should be suspended.
Pending Update				Specifies how the pending recordset should be treated.
	Drain	Yes	Yes	Specifies that the volume pair(s) should be suspended after all pending recordset are settled. Refer to (3) in this section for notes on this option.
	Purge	Yes	Yes	Specified that that volume pair(s) should be suspended even if pending recordset remain. The pending recordset may be purged. If the MCU/RCU discard the pending recordset, the MCU/RCU marks the cylinders that contain discarded recordset as modified in its shared memory. The marked cylinders will be copied when the Resume Pair operated.

(b) Update sequence consistency

Regardless of the Pending Update parameter, the update sequence consistency across the volume pairs to be suspended is ensured. However, if some volume pairs have been suspended with the Volume option and other related volume pairs are updated after that, the update sequence consistency is not ensured.

(c) Notes on Drain option

- If the Purge option specified, the volume pairs are suspend when the RCU accepts this operation. On the other hand, if the Drain option is specified, the volume pairs will be suspended when the RCU completes the steps described bellow.
  ① Accepts this operation and;
  ② Has finished settling all the pending recordsets and;
  ③ Completes the negotiation with all MCUs (reports ready-for-suspension to all the MCUs and receives their acknowledgements) without further recordset generated.

- Therefore, the procedures to get the R-VOLs whose contents are frozen at specific point in time relative to the application, are as follows
  ① Quiesce the application (Quiesce all update activity to the volume pairs) and;
  ② Perform the Suspend Pair operation with Drain option and;
  ③ Confirm that all volume pairs complete to be suspended and;
  ④ Restart the application.

- Time to disrupt the application (between 1 and 4 above) mainly depends on time for the RCU to follow steps (a) to (c) described above. That is, it depends on how many recordsets are pending and how much I/O workload the MCUs/RCU have. Since the RCU will break off above steps and forcibly suspend the volume pairs if time specified by the maximum copy delay time parameter expires, this parameter can help decide the expected disruption time.

(d) End conditions

The Suspend Pair operation for the asynchronous volume pairs gets the normal end condition when the operation is accepted by the MCU/RCU. The pair status of the volume pairs to be suspended are Suspending at this time. And then the pair status will be Suspended after the internal process (negotiation between the MCU and RCU) completes. Therefore the result of this operation should be confirmed by referring to the actual pair status or the console messages that indicate pair status change.

(4) *Resume Pair* - Resuming HRC Volume Pairs

The Pair Resume option is newly supported. It specifies whether all suspended volume pairs, which belong to the same consistency group and whose M-VOLs are behind this MCU, should be resumed together or not. Refer to Table 3.17.7-8 for this option.

(5) *Pair Option* - Modifying HRC Options on Volume Pair Basis

Refer to Table 3.17.7-8 for the options on volume pair basis. The Error Level option can be changed regardless of the pair status (even if the volume pair is in Suspended state), but become effective at the next time of suspension.

(a) Modified items on Pair Status panel

Table 3.17.7-11 Pair Status - Modified Items on Pair Status Panel

Items	Indicated at:		Description
	MCU	RCU	
M-VOL and R-VOL	Yes	Yes	Indicates control unit image number and device number in form of "c:dd".
Pair Synchronized	Yes	Yes	Indicates the amount (%) of cylinders that are marked as modified in the bitmap for pair re-synchronization. Refer to (2) below for more detailed information.
Pair Status			Adding to the contents for the synchronous HRC volume pairs, the following are added for the asynchronous HRC volume pairs.
Suspending	Yes	Yes	Indicates that the volume pair is in transition state to Suspended.
Deleting	Yes	Yes	Indicates that the volume pair is in transition state to Simplex.
Suspended[MCU PS OFF]	No	Yes	Indicates that the volume pair has been suspended due to the power-off event of the MCU.

(b) The Pair Synchronized indicator

- For the asynchronous volume pairs, both MCU and RCU maintain the bitmap for pair re-synchronization. The indicated percentage is calculated as follows.

— The M-VOL of Pending state indicates the remaining cylinders to be copied for pair re-synchronization.
— The R-VOL of Suspended state indicates the cylinders that contain the recordsets lost at the RCU (reached at the RCU but can not be settled before suspension).
— The M-VOL of Suspended state indicates the cylinders that contain;

① The tracks that have not copied yet by the initial copy and;
② The records updated by the primary system after suspension and;
③ The recordsets lost at the MCU (created in the MCU but can not be sent to the RCU before suspension) and;
④ The recordsets lost at the RCU.

The last item (d) is included as long as the MCU can get the information from the RCU. Otherwise, only items (a)-(c) are included. In this case, the item (d) will be included at the beginning of the pair resynchronization.

- This percentage is always calculated based on the total cylinders of the M-VOL (even if the R-VOL is larger than M-VOL).

(c) Added items on Pair Status panel

Table 3.17.7-12 Pair Status - Added Items on Pair Status Panel

Items	Indicated at:		Description
	MCU	RCU	
C/T Group	Yes	Yes	Indicates the consistency group number to which this volume pair belongs.
C/T Type	Yes	Yes	Indicates the timer type that has been specified to control the consistency group. The displayed content is out of System, Local and None.
C/T	Yes (Note)	Yes	Indicates the consistency time of this volume pair in form of "mm/dd/yyyy hh:mm:ss.uuuuuu". (uuuuu: micro seconds)
SEQCHK	Yes (Note)	Yes	Indicates SEQCHK if the volume pair is in SEQCHK status. Otherwise it blanks.
Suspended by	No	Yes	Indicates the consistency status of the R-VOL when this volume pair is in Suspended status. Group The consistency between this R-VOL and other R-VOLs in the consistency group is ensured. Volume The consistency between this R-VOL and other R-VOLs in the consistency group is not ensured. The contents of this R-VOL may be behind other R-VOLs.

(Note) Only the RCU maintains the latest result and can display it. The result at the MCU may be behind the actual result because the MCU get the result from the RCU before displaying. Therefore the result at the MCU can be used *for monitoring the normal activities*, but can not for the disaster recovery at the RCU.

(d) Added items

Table 3.17.7-13 Pair Status - Added Items on Pair Status Panel

Items	Indicated at:		Description
	MCU	RCU	
Grp (Lv)	Yes	Yes	Grp Indicates the consistency group number to which this volume pair belongs. (Lv) Indicates the Error Level of this volume pair, "Grp" for group or "Vol" for volume.
Sub	No	Yes	Indicates the consistency status or the others when the volume pair is in "suspended" state. GRP This R-VOL has the consistency status of Group. VOL This R-VOL has the consistency status of Volume. OFF This volume pair has been suspended due to the power-off event of the MCU.
Seq	No	Yes	Indicates SEQCHK if the volume pair is in SEQCHK status. Otherwise it blanks.

3.17.7.5 Monitoring subsystem Statistics

(1) *Usage* - Displaying Remote I/O Statistics Information

Table 3.17.7-14 Usage - RIO Statistics in Async Copy Category

Item name	Unit	Description
Async IO count	—	Indicates the total number of RIO (Remote I/O) activities completed in a specified interval.
Total number of recordset	—	Indicates the total number of recordset sent to the RCU in a specified interval.
RCU command retries	—	Indicates the total number of RIO command retries requested by the RCU in a specified interval.
MCU command retries	—	Indicates the total number of channel command retries performed by the MCU in order to avoid cache slot conflict between the host I/O and RIO processes.
Average transfer rate	KB/sec.	Indicates the average data transfer rate of the recordset sent to the RCU in a specified interval.
Average RIO response	msec.	Indicates the average RIO response time in a specified interval.

(2) *Information/Monitor* - Displaying Subsystem Resource Usage

- The item named "Async Write Pending Data" is newly supported. It is displayed on the Information/Monitor panel of the SVP and recorded into monitor.dat file.
- "Async Write Pending Data" indicates the total amount of cache space in percent that store the pending recordset (Sidefile).
- The conventional item "Sidefile" indicates the total usage of the Sidefile for HXRC, Concurrent Copy and HRC asynchronous.

3.17.7.6 On-line Micro-program exchange procedure

(1) Version up procedure

- Usually, exchange the Micro-program of MCU first.
- If the ‘Copy Back function’ is running, exchange the Micro-program of RCU first.

(2) Version down procedure

- If in the ‘N-to-1’ configuration, set back to ‘1-to-1’ configuration.
- Execute ‘Delete- Group’ for all CT groups on the subsystem.

3.17.8 Management/Recovery Procedures

3.17.8.1 Managing HRC Asynchronous Subsystems

(1) Checking on SEQCHK Status

- SEQCHK status would be indicated when the asynchronous volume pair in the consistency group with the timer type of System specified accepts the non-time-stamped updates from the primary system.
- SEQCHK status does not affect copy activities of HRC asynchronous and will be removed when the next time-stamped update is successfully copied to the R-VOL. However if the disaster happens before the next time-stamped update, the update sequence consistency between the R-VOL and other R-VOLs in the consistency group is not ensured. Therefor to make the disaster recovery more certain, the source of SEQCHK status should be detected and removed.
- Source of SEQCHK status to be suspected is as follows:
  — An application may issue the update I/Os with bypassing MVS standard I/O procedure.
  — An XRC time-stamping capability may not active at the primary system.
  — An Operating system of the primary system may not support the time-stamping capability.

(2) Checking on the Consistency Time

The consistency time is indicated as a part of pair status of the HRC asynchronous volume pairs. While the primary system continues to update the M-VOLs, the difference between the current time and the consistency time indicates how long the R-VOLs are behind the M-VOLs. The updates to the M-VOLs during this duration may be lost when the disaster happens

- If the disaster recovery design can not accept this delay, performance improvement by adding the remote copy resources (paths, cache amount, etc.) and/or reducing unnecessary I/O workload should be considered to get shorter delay.
- If this delay is close to the time specified by the Maximum Copy Delay Time parameter, the timeout failure may occur and the affected volume pair may be suspended due to the I/O workload fluctuations. The performance improvement described above and/or increasing this parameter setting should be considered.

(3) Planned Outage of HRC Asynchronous Components

The MCU requires communication with the RCU even if it receives no update I/Os from the primary system. Suspending all the duplex pairs is necessary prior to the planned outage of the RCU.

(a) Planned outage of the MCU

- No special procedure is required for the asynchronous volume pairs. The MCU automatically suspends all volume pairs in Duplex/Pending state during the power-off sequence, and will remove the Suspended state at the power-on-reset sequence.

(Note) Perform planned outage operation of the MCU after the Sidefile data becomes a zero. The Sidefile ratio of the subsystem is shown on the Monitor dialog box.

- Note that the volume pairs whose M-VOLs are behind the MCU will not be suspended. In N-to-1 configuration, the volume pairs behind other MCU(s) continue to be copied. If the full-consistent volume set requires to be kept during the planned outage, the procedures are as follows:

① Quiesce the applications.
② Perform Suspend Pair operation with Group and Purge (or Drain) option at the RCU.
③ Perform planned outage operation of the MCU(s).
④ After getting all MCUs ready to resume, perform Resume Pair operation at all MCUs.

(b) Planned outage of the RCU

- Suspending all volume pairs is required prior to the planned outage according to the followings:

① Suspend all volume pairs;
  1) Perform Suspended Pair operation at each MCU and;
  2) Confirm volume status with Pair Status operation.
② Perform planned outage operation of the RCU.
③ Get the RCU ready to resume.
④ Remove Suspended status from all volume pairs;
  1) Perform Resume Pair operation at each MCU and;
  2) Confirm volume status with Pair Status operation.

(Note) If step 1 above is not performed, the MCU detects the communication failure and suspend all affected volume pairs with generating SIM(s) and console message(s) that indicates the failure.

(c) Planned outage of the components on the remote copy connection

The same restriction should be considered and the same procedures should be performed as the RCU for the components (channel extender, ESCON director, etc.) on the remote copy connection.

- Suspending all volume pairs is required prior to the planned outage according to the followings:

① Suspend all volume pairs;
   1) Perform Suspended Pair operation at each MCU and;
   2) Confirm volume pair status with Pair Status operation.
② Perform planned outage operation of the component.
③ Get the component ready to resume.
④ Remove Suspended status from all volume pairs;
   1) Perform Resume Pair operation at each MCU and;
   2) Confirm volume status with Pair Status operation.

(Note) If step 1 above is not performed, the MCU detects the communication failure and suspend all affected volume pairs with generating SIM(s) and console message(s) that indicates the failure.

(d) Planned outage of both MCU and RCU

- The MCU(s) must become not-ready first, and the RCU must be back before the MCU(s).

① Perform planned outage operation of the MCU(s).
② Perform planned outage of operation of the RCU.
③ Get the RCU ready to resume.
④ Get the MCU(s) ready to resume.

- Note that the MCU make all volume pairs suspended if it can not communicate with the RCU during the power-on-reset sequence. Therefor step 4 above should be started after completing step 3. If it is difficult to control due to some installation requirements (ex. Power-Control-Interface setting), consider using the Maximum RCU-ready-wait Time parameter.

(4) ICKDSF on the Asynchronous Volume Pairs

- ICKDSF activities involve write I/Os with device support authorization or diagnostic authorization instead of normal authorization. Since the MCU does not duplicate write I/Os with device support or diagnostic authorization, the HRC volume pairs must be suspended before running ICKDSF. The procedures to do are as follows:

① Perform Suspend Pair operation with Volume option at the MCU to suspend volume pair(s) for which ICKDSF will be performed.
② Perform ICKDSF.
③ Perform Resume Pair operation at the MCU.

(5) Micro-program Exchange

No special procedure is required for the asynchronous volume pairs for the on-line micro-program exchange and the MCU's off-line micro-program exchange.
In the case of HRC Asynchronous components, Suspending all volume pairs is required prior to the RCU's off-line micro-program exchange.

① Suspend all volume pairs;
   1) Perform Suspended Pair operation at each MCU and;
   2) Confirm volume status with Pair Status operation.
② Perform micro-program exchange operation of the RCU.
③ Remove Suspended status from all volume pairs;
   1) Perform Resume Pair operation at each MCU and;
   2) Confirm volume status with Pair Status operation.

(Note) If step 1 above is not performed, the MCU detects the communication failure and suspend all affected volume pairs with generating SIM(s) and console message(s) that indicates the failure.
In step 3, if both micro-program exchange operation is performed, perform Resume Pair operation after waiting 10 minutes after RCU's micro-program exchange is completed.

3.17.8.2 Recovering from Pair Suspended

(1) Recovering from Pair Suspended - Suspended by the MCU

The cause of suspension and the recovery procedures are basically the same as HRC synchronous with exceptions that:

① The cache storage/shared memory of the one side (not both sides) may cause the HRC asynchronous volume pairs to be suspended.
② The MCU requires communicating with the RCU in power-on-reset sequence and suspends all the HRC asynchronous volume pairs if it can not do so.
③ The MCU suspends all the HRC asynchronous volume pairs if no recordset can be sent in the specified time period.

(2) Recovering from Pair Suspended - Suspended by the RCU

Adding to directed by the MCU, the RCU of HRC synchronous may detect the cause of suspension and suspended the failed volume pair(s) on its own initiative.

3.17.8.3 Disaster Recovery - Switching to the Secondary Subsystem

(1) Switching Procedures to the Secondary Subsystem

Basic procedures to switch to the secondary subsystem are as follows:

① Check pair status and memorize the consistency time of the R-VOLs.
② Make the R-VOLs simplex by performing Delete Pair operation.
③ Confirm that the R-VOLs have successfully been changed to simplex.
④ Perform IPL of the secondary system.
⑤ Perform the application restart procedure at the secondary system

(2) Suspending Volume Pairs due to the Disaster

If the RCU receives no communication with the MCU after the time specified by Maximum Copy Delay Time expires, the RCU regards this timeout event as a disaster and suspends all the HRC asynchronous volume pairs.

- Pair status of the R-VOLs normally become Suspended - Group.
- The latest time-stamp of the recordset that has been successfully settled is frozen and indicated as the consistency time of the R-VOLs.

(3) Checking Volume Pair Status

- The R-VOLs of Suspended - Volume should not be used for the disaster recovery. Therefore Error Level of Group should be selected if the volume is essential to the disaster recovery.
- Delete Pair operation with Group option can make only the R-VOLs of Suspended - Group status. Therefore an operator seems to be able to skip this checking. However the volumes of Simplex status can not be distinguished form others after the R-VOLs in Suspended - Group status become Simplex by Delete Pair operation. Therefore at least the volumes in *Simplex* must be checked at this time.

Table 3.17.8-1 Checking R-VOL Status When Switching to the Secondary System

Pair status		Usable for recovery?	Description
Suspended	Group	Yes (*1)	Since the update sequence consistency across these R-VOLs is ensured at point in time indicated by the consistency time, these R-VOLs can be used for the disaster recovery at the secondary system. Note that some updates performed after the consistency time at the primary system may be lost.
Suspended	Volume	No	Since contents of this R-VOL may be behind other R-VOLs in the consistency group, this R-VOL should not be used for the disaster recovery if this volume requires the update sequence consistency with other volumes. Suspected reason for this status are as follows: — This R-VOL has been suspended due to the failure or Suspend Pair operation prior to the disaster. — This R-VOL was in pending status when the disaster happened. — The Error Level of Volume has been specified for this R-VOL and this R-VOL was suspended by the first symptom of the disaster.
Duplex Pending Simplex		No No No	This status does not usually take place in the disaster recovery procedure. This R-VOLs should not be used for the disaster recovery. *It should be especially noted that the volumes in simplex status can not be distinguished from others after deleted by Delete Pair operation.*

(*1) If SEQCHK status should be decided by the RCU using the XRC time-stamping capability, this R-VOL shall not be used for disaster recovery.

(4) Memorizing the Consistency Time

The consistency time of the R-VOLs should be *memorized*. It could help the disaster recovery retrieve the lost updates.

- All the R-VOLs of Suspended - Group status indicate the same consistency time. Therefor any R-VOL of this status can represent the consistency time of the consistency group.
- Once Delete Pair operation performed, the R-VOL *never* indicates the consistency time again. Therefore the consistency time should be memorized before Delete Pair operation.

(5) Deleting Volume Pairs

By Delete Pair operation, the R-VOLs that are used for the disaster recovery should be changed to Simplex status.

- Group option of Delete Pair operation can make all the R-VOLs of Suspended - Group status Simplex. It would be helpful in reducing number of operations.
- Group option of Delete Pair operation does not change the pair status of the R-VOLs if they are other than Suspended - Group. It can prevent the inconsistent volumes from being used for the disaster recovery.

(6) Recovering the Lost Updates

HRC asynchronous provides *no* factory standard procedure to retrieve the lost updates.

- In order to detect and recreate lost updates, it is necessary for customers to check other current information, for example, data base journal log file that had been active at the primary system when disaster occurred. Note that the journal log file entries of most DBMS may be related to time-of-day clock information and the source of the consistency time is time-of-day clock of the primary system (when the timer type of System specified.).
- However such a detection/retrieval would take long time to do. The customers' disaster recovery scenario recommends to be designed to enable such a detection/retrieval *after* the application has been started at the secondary system.
- Maximum Copy Delay Time parameter of Add Group operation can control the maximum time duration during which the updates may be lost.

3.17.8.4 Disaster Recovery - Switching Back to the Primary Subsystem

This section describes the procedures to switch back from the secondary system to the primary system. The basic concept and procedures are the same as HRC synchronous. After volume pairs establishment in the opposite direction, planned switching-back and volume pair establishment in the original direction are performed.

That is, the original R-VOLs are working as Simplex and the application is running on the secondary system by using the original R-VOLs. The original HRC asynchronous configuration/status should remain in the original MCU. In this section, the original MCU/M-VOLs and RCU/R-VOLs are called the primary subsystem and secondary subsystem respectively.

(1) Switching Back Procedures to the Primary Subsystem

(a) Make the primary subsystem and the communication facilities on the remote copy connection operable. Note that all the M-VOLs in the primary subsystem may be suspended since the original R-VOLs are now in Simplex status.

(b) Remove entire HRC asynchronous configuration remained in the *primary* subsystem:

① Make all the volume pairs Simplex by Delete Pair operation.
② Remove registration of the consistency group by Delete Group operation.
③ Remove MCU-RCU logical paths by Delete RCU operation.
④ Change the serial interface port to LCP mode.

(c) Establish HRC asynchronous in the *opposite* direction:

① Change the operating mode of the communication facilities in the opposite direction.
② Change the serial interface port of the secondary subsystem to RCP mode.
③ Register the consistency group at the secondary subsystem by Add Group operation.
④ Establish MCU-RCU path in the opposite direction at the secondary subsystem by Add RCU operation. The primary subsystem is now defined as the RCU.
⑤ Establish the HRC asynchronous volume pairs in the opposite direction. The volumes in primary and secondary subsystem are now defined as the R-VOLs and M-VOLs respectively. Note that Initial Copy of Entire Volume option must be specified.

(d) Quiesce the application at the secondary system.

(e) Confirm all the HRC asynchronous volume pairs are in Duplex status.

(f) Remove entire HRC asynchronous configuration at the *secondary* subsystem:

① Make all the volume pairs Simplex by Delete Pair operation.
② Remove registration of the consistency group by Delete Group operation.
③ Remove MCU-RCU logical paths by Delete RCU operation.
④ Change the serial interface port to LCP mode.

(g) Establish HRC asynchronous in the original direction

① Change the operating mode of the communication facilities in opposite direction.
② Change the serial interface port of the primary subsystem to RCP mode.
③ Register the consistency group at the primary subsystem by Add Group operation.
④ Establish MCU-RCU path in opposite direction at the secondary subsystem by Add RCU operation. The secondary subsystem is now defined as the RCU.
⑤ Establish the HRC asynchronous volume pairs in the original direction. The volumes in primary and secondary subsystem are now defined as the M-VOLs and R-VOLs respectively. Note that Initial Copy of No Copy option may be specified.

(h) Make the primary subsystem online from the primary system and restart the application at the primary system.

(Note) Since the CNT channel extenders have the operating mode, they must be re-configured to change copy direction. The boxes (or nodes) to which the *current* MCU and RCU are connected must be set as channel-mode and device-mode respectively.

(2) Setting the Consistency Group for Switching Back

- For N-to-1 (N≥2) configuration, timer type of "None" muse be specified.
  - — None enables the consistency group across up to 4 RCUs to be established.
  - — However the update sequence consistency across the RCUs is not ensured.
- In 1-to-1 configuration, the update sequence consistency across all R-VOLs is ensured regardless of the timer type. However System is recommended if the secondary system can use the XRC time-stamping capability in order for the consistency time to be indicated.

Table 3.17.8-2 Timer Type for Switching Back

Original Configuration		Timer Time for Switching Back	
Number of MCU-RCU	Timer Type	XRC time-stamping capability at the secondary system?	
		Yes	No
N-to-1 (N≥2)	System	**None**	**None**
1-to-1	System	**System**	**None**
	Local	**Local**	**Local**

3.17.9 Specification Table

Table 3.17.9-1 shown below describes the specifications and support stage of these items.

Table 3.17.9-1 Specification Table

No.	Items	Specification
Operating systems		
1-1	MVS of XRC support level	• Yes
1-2	Other M/F systems	• Yes
1-3	Open systems	• Yes (HORC Asynch)
Support devices		
2-1	Disk subsystems (MCU/RCU)	• RAID300 or higher level
2-2	DKC emulation type	• 3990-6 or -6E
		• 3990-3
2-3	Kind of physical drives	• No limitation
2-4	RAID levels	• RAID5
		• RAID1
2-5	Logical volume emulation	• 3390-1, -2, -3, -3R, -9
		• 3380 (for non-MVS systems)
2-6	CVS'ed logical volumes	•Yes

(To be continued)

(Continued from preceding sheet)

Remote copy configuration		
3-1	Remote copy connections	• ESCON up to 43km
		• ESCD, MuxMaster available
		• CNT channel extenders
3-2	ESCON port setting	• MCU requires RCP setting
3-3	MCU-to-RCU connection	• 1-to-1
		• N-to-1 (N≤4)
3-4	Number of logical CU pairs	• Up to 16
3-5	Number of logical paths	• Up to 4 per a logical CU pair
3-6	Number of volume pairs	• Up to 4,096
3-7	Number of consistency groups	• 1
		• Up to 64 groups
3-8	Remote Copy volume pairs in a same unit stc.	• HRC(Synch.)
		• XRC
		• Multiplatform
		• HORC
Concurrent use of other functions for the HRC asynchronous volume pairs		
4-1	Concurrent copy	• Yes
4-2	Archival function	• Yes
4-3	DCR	• Yes
4-4	HMRCF	• R-VOL = HMRCF S-VOL
		• M-VOL = HMRCF S-VOL
		• M-VOL = HMRCF T-VOL (One side pair status is suspend)
4-5	HXRC	• XRC(R-VOL) = M-VOL
		• another pattern is No
4-6	HRC Synch.	• No
4-7	HODM	• No

(To be continued)

(Continued from preceding sheet)

Detailed functions		
5-1	Timestamp information ***(Timer Type)***	• system TOD clock
		• Internal TOD clock
5-2	Volumes to be suspended against copy failure ***(Error Level)***	• All volumes
		• Only affected volume
5-3	Message when pair suspended (for M/F systems)	• SIM with unique REFCODE
		• PPRC console messages (IEA494I, IEA491E)
5-4	Information for monitoring HRC status	• Volume pair status
		• consistency time
		• consistency status (***Volume*** or ***Group)***
		• Alert to non-time-stamped writes ***(SEQCHK)***
5-5	RESYNC copy	• in addition to different cylinder maintained by an MCU and an RCU
		• in acceding order of cylinder number
		• including “in flight” data
		• under-condition of correction copy
5-6	Copy back to original volumes	• 1-to-1
		• 1-to-N
5-7	R-VOL R/W	• Yes

(To be continued)

(Continued from preceding sheet)

Controlling/managing means		
6-1	Basic means	• Remote console P/P (RMC)
		• SVP (only for CE use)
		• PPRC TSO commands
		• PPRC console messages - IEA494I,IEA491E
		- DEVSERV PATH
6-2	RMC scripting means	• Yes
6-3	RMC operation log-out	• Yes
Control operations		
7-1	Volume basis operations, operable at MCU	• Add Pair
		• Delete Pair
		• Suspend Pair
		• Resume Pair
		• Pair Status
7-2	Group basis operations, operable at MCU	• Add Group*
		• Delete Group*
		• **Delete Pair with *Group***
		• **Suspend Pair with *Group***
		• Group Status
7-3	Volume basis operations, operable at RCU	• Delete Pair
		• Suspend Pair
		• Pair Status
7-4	Group basis operations, operable at RCU	• Delete pairs
		• Suspend Pair
7-5	Tuning parameters, operable at MCU	• Cache usage for recordset*
		• Timeout for pending writes*
		• Timeout for RCU ready*
7-6	Statistics monitor, accumulated by MCU	• Amount of cache usage for recordset*
		• Number of write per volume*
		• Average data length*

* by RMC only

3.18 HIHSM(Hitachi internal Hierarchical Storage Management)

3.18.1 HIHSM Overview

This document describes the function of HIHSM (Hitachi internal Hierarchical Storage Management) that is one of program products.

RAID system can be constructed by several types of physical drive and two types of RAID level (RAID1 and RAID5).

This combination of the type of physical drive and the type of RAID level provides the system that cost and performance are optimized to user environment. However, it is difficult to get information about actual utilization of physical drives in the RAID system unlike other disk subsystems.

(1) HIHSM provides solution of the problem and supports decision of users to determine system construction as described below.
  (a) Load balancing of system resources.
  Unbalance of utilization of system resources makes performance worse. HIHSM supports decision of optimized allocation of logical volumes to physical drives.
  (b) Migration of logical volumes optimized to access patterns to physical drives.
  For instance, RAID5 is suitable to sequential access, and RAID1 of high performance drive is suitable to random access that is required small response time. HIHSM shows types of access pattern to physical drives clearly, and supports migration of logical volumes to suit the access pattern.

(2) HIHSM consists of following subfunction to achieve above purposes. Users can refer utilization of system resources monitored by monitor function, decide reallocation plan by using estimate function, and reallocate the logical volumes by volume moving (migration) function.
  (a) Monitor function monitors and shows utilization of system resources.
  (b) Estimate function estimates utilization of parity groups after migration of logical volumes.
  (c) Volume moving (migration) function moves logical volumes to specified parity groups.
  (d) Preset function makes a migration plan from information that users preset, and moves the logical volumes by the migration plan automatically.

3.18.2 Hardware requirements

none.

3.18.3 Monitor function

(1) Start and end of monitoring
Monitoring starts in DKC by direction from HIHSM utility window. The monitoring continues until the direction of ending the monitoring from HIHSM utility window.

(2) Monitoring information
Monitor function monitors utilization of system resources described below.
(a) CHP utilization and DKP utilization
(b) Starnet utilization
(c) DRR utilization
(d) Parity group utilization: Disk utilization of parity groups. The used time of physical drives in a parity groups
(e) Parity group utilization of each logical volume: The used time of physical drives of synchronous and asynchronous access on each logical volume, averaged by the number of physical drives in the parity group. Parity group utilization means sum of utilization of each logical volume in the parity group.
Above information is collected from DKC to SVP twice a day at request time(AM/PM) automatically, and before the making time of migration plan and before the executing time of migration plan, or by requiring from users. The collected information of 3 months is stored in the hard disk of SVP. When the amount of the information is over 3 months, the oldest information is overwritten by new information.

(3) Viewing monitored information
Users can specify term and kind of the information to see in HIHSM utility window. In HIHSM utility window, average value and maximum value of the utilization of system resources are shown. The CHP, DKP, Starnet, and DRR utilization are shown in one window, and parity group utilization is shown in another window.

Users can see parity group utilization and parity group utilization of each logical volume by following. (1) View parity group utilization in HIHSM main window. (2) Select one parity group in the window to show the utilization of each logical volume in the parity group.

Monitored information can be outputted to FD. User can see the outputted information by HIHSM viewer tool.

3.18.4 Estimate function

Estimate function estimates changes of parity group utilization and parity group utilization of each logical volume after migration of the logical volume to specified parity group. Estimate function estimates the changes from the monitored information.

3.18.5 Volume moving (migration) function

Volume moving (migration) function moves data in logical volume (source volume) to physical location of another logical volume (destination volume). Users specify the volumes in HIHSM utility window.

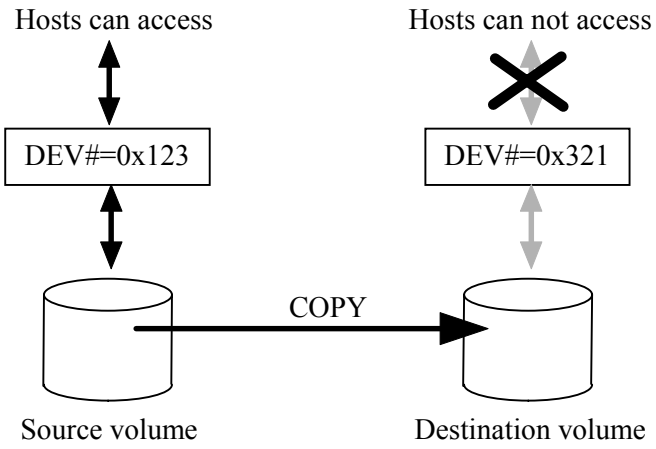


Fig. 3.18.5-1 Before moving

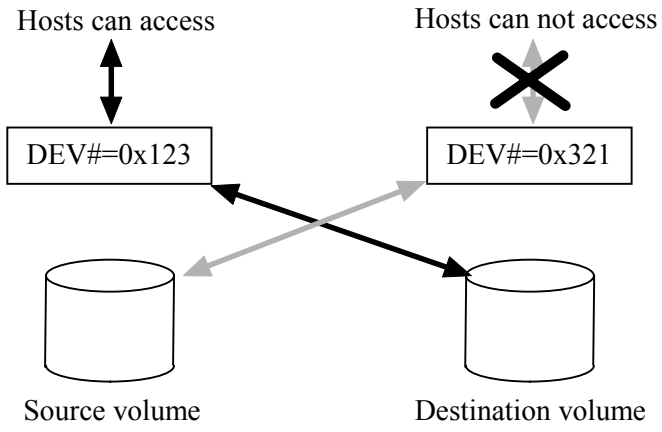


Fig. 3.18.5-2 After moving

(1) Volume moving function Overview

Logical volumes. The Dynamic Optimizer source and target volumes must be specified by LDEV ID, not VOLSER or port/TID/LUN. The Dynamic Optimizer source and target volumes must be within the same 9900 subsystem and have the same device emulation type and same size.

(a) Source Volumes:

- Volumes which are set as command devices(devices reserved for use by the host)
- Volumes which are use by XRC
- Volumes which are used by CC(Concurrent Copy)
- Volumes which have Flash Access(also called DCR) data stored in cache
- Volumes which are in an abnormal or inaccessible condition(e.g., pinned track, fenced)

If the status of volumes that form HRC pairs is *suspended*, the volumes can be used as source volumes. If the status of the volumes is not s*uspended*, the volumes cannot be used as source volumes. If you delete an HRC pair from an MCU, the status of the M-VOL and the R-VOL changes to *simplex*, so that the volumes can be used as source volumes. If you delete and HRC pair from an RCU, the status of the M-VOL changes to *suspended* and the status of the R-VOL changes to *simplex*, so that the volumes can be used as source volumes.

If the status of volumes that form HORC pairs is *PSUS* or *PSUE*, the volumes can be used as source volumes. If not, the volumes cannot be used as source volumes. If you delete and HORC pair from an MCU, the status of the P-VOL and the S-VOL changes to *SMPL*, so that the volumes can be used as source volumes. If you delete and HORC pair from an RCU, the status of the P-VOL changes to *PSUS* and the status of the S-VOL changes to SMPL, so that the volumes can be used as source volumes.

For volumes that form an HMRCF pair or an HOMRCF pair, it depends on the status or configuration of the pair whether the volumes can be used as source volumes, as explained below:

- If the status of the pair is not *Split pending*, the volumes can be used as source volumes. If the status of the pair is *Split pending*, the volumes cannot be used as source volumes.

- The table below explains whether volumes that do not form a cascade pair can be used as source volumes:

Table3.18.5-1 Whether volumes that do not form a cascade pair can be used as source volumes.

If the pair is configured as follows	Can P-VOLs be used as source volumes?	Can S-VOLs be used as source volumes?
If the ratio of P-VOLs to S-VOL is 1:1	Yes	Yes
If the ratio of P-VOLs to S-VOL is 1:2	Yes	Yes
If the ratio of P-VOLs to S-VOL is 1:3	No	Yes

– The Table below explains whether volumes that form a cascade pair can be used as source volumes:

Table 3.18.5-2 Whether volumes that form a cascade pair can be used as source volumes

If the pair is configured as follows	Can P-VOLs be used as source volumes?	Can S-VOLs be used as source volumes?
If the pair is an L1 pair and the ratio of P-VOLs to S-VOL is 1:1	Yes	Yes
If the pair is an L1 pair and the ratio of P-VOLs to S-VOL is 1:2	Yes	Yes
If the pair is an L1 pair and the ratio of P-VOLs to S-VOL is 1:3	No	Yes
If the pair is an L2 pair and the ratio of P-VOLs to S-VOL is 1:1	Yes	No
If the pair is an L2 pair and the ratio of P-VOLs to S-VOL is 1:2	No	No

Caution: If any of the following operations is performed on a source volume, the volume migration process stops:

– XRC operation
– CC operation
– HRC/HORC operation that changes the volume status to something other than *suspended*
– ShadowImage (HMRCF/HOMRCF) operation that changes the volume status to *Split Pending.*

(b) Target volumes:

Target volumes must be reserved prior to migration. The HIHSM remote console software allows you to reserve volumes as HIHSM target volumes.

Hosts cannot access HIHSM-reserved volumes.

The following volumes cannot be reserved as target volumes:

– Logical Unit Size Expansion (LUSE) volumes
– Volumes which are set as command devices (devices reserved for use by the host)
– Volumes which are assigned to Hitachi ShadowImage(HMRCF/HOMRCF)or Hitachi Remote Copy (HRC/HORC) pairs
– Volumes which are used by XRC
– Volumes which are used by CC (Concurrent Copy)
– Volumes which are reserved for ShadowImage operations
– Volumes which have FlashAccess (also called DCR) data stored in cache
– Volumes which are in an abnormal or inaccessible condition (e.g., pinned track, fenced)

(c) Specifying Volumes:
Source volume and destination volume are specified by LDEV number.
An open volume also specified by one LDEV number. If the volumes are set as LUSE, you can specify by one LDEV number. HIHSM will migrate the LDEV which is one of a LUSE and its access is higher.

(d) Unit to moving:
One moving can be instructed by specifying one pair of volumes (LDEV).

(e) Moving of multi volumes:
Moving of volumes can be performed by repeating instruction about each volume.
Maximum 36 volumes can be accepted in the same time.
However, the maximum value of the total number of HMRCF/HOMRCF pairs plus the number of moved volumes of HIHSM is restricted as follows.
① When the number of CU is 4 or less, 1024.
② When the number of CU is 5 or more, 2048.

(f) Abort moving:
Users can direct to abort the instructed moving before completion.
With aborting, the data in the destination volume is not guaranteed.
Users can direct to abort about each LDEV.

(g) Notice when the DKC is maintenance:
HIHSM sometimes fails the volume migration if the Cache or drive replacement installation or deinstallation is executed at the same time.

(2) Conditions for moving
Data moving is performed when all conditions about source volume and destination volume described bellow are satisfied.

(a) Both of the volumes have same emulation type.
(b) Both of the volumes have same size.
(c) There is no PIN data in the source volume.
(d) Both of the volumes are not blockade.
(e) Both of the volumes in same DKC.
(f) The volumes are not instructed to move already and not waiting to move.
(g) The Volumes are not combination of CVS Volume and Normal Volume.

(3) Viewing History
Users can see the history of volume moving (migration) by [History] button in HIHSM utility window.

3.18.6 Decision of volume moving (migration)

HIHSM supports decision of users about disk system performance tuning by logical volume moving (migration). This section describes usage and points to notice about monitor function.

(1) Inspecting utilization of system resources
First of all, using monitoring function, a user investigates whether there exists overloaded resources, or imbalance of resource utilization. Then the user tunes resource utilization in the manner described in the following clause.

Note: due to average system resource utilization, there will be such a case as a portion of system performance will be negatively effected although total performance of a system will be improved. For example, if there exists RAID groups A and B of utilization 20% and 90% respectively, and if the utilization will become 55% and 55% if a logical volume residing in parity group B moves parity group A. Then response time of I/Os to parity group A will be increased while response time of I/Os and throughput to parity group B will be improved.

(2) Tuning Starnet utilization
Since Starnet are common resources in RAID400, migration of logical volumes does not improve system performance.

(3) Tuning CHP utilization
Migration of logical volumes does not improve CHP performance. Therefore if CHPs are overloaded on an average, the user should consider installation of new CHAs. And if the utilization of CHPs are imbalance, the user should consider that channel paths connecting to a CHA, which includes overloaded CHPs, is reconfigured into the connection to another CHA which includes CHPs of lower utilization.

(4) Tuning DKP/DRR utilization

If utilization of DRRs or DKPs is in high average, the user should consider to install new DKAs and HDDs. After installation of DKAs and HDDs, logical volumes which had high traffic of write access (especially of sequential write access) should be migrated to a parity group in newly installed HDDs.

If utilization of each DKP are imbalance, the user should consider migration of logical volumes from current parity group deployed under a DKA pair of high utilization to one under a DKA pair of lower utilization. The estimate function cannot simulate the DKP utilization. Therefore this method should be applied under the prospect of large improvement. For example, there would be least improvement in case of slight difference of utilization of each DKP, or if DRRs are comparatively in high utilization.

(5) Tuning RAID group utilization

If parity groups are in high utilization, the user should consider to install new HDDs. After installation of HDDs, logical volume which had high traffic of I/Os should be migrated to a parity group in newly installed HDDs.

If utilization of each parity group are imbalance, the user should consider migration of logical volumes from current parity group of high utilization to the one of lower utilization.

These method should be applied under the prospect of large improvement. There would be least improvement in case of slight difference of utilization of each parity group, or if DRRs or DKPs are comparatively in high utilization.

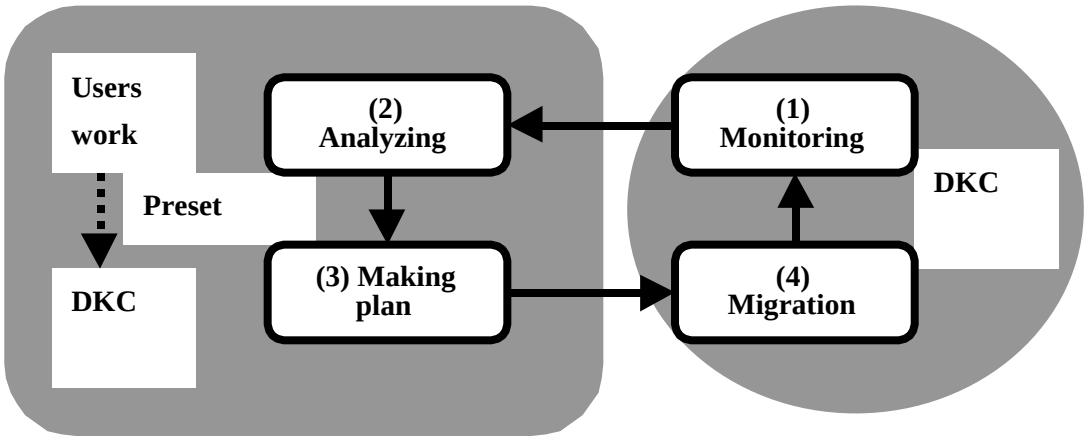
When some kind of errors in the system, utilization of system resource can increase or be unbalanced.

3.18.7 Preset function

(1) Preset function Overview
Preset function provides typical tuning method of HIHSM based on parameters given by users, and performs tuning plan automatically.

(2) Overview of tuning
Tuning by HIHSM can be done by following steps repeatedly.
(a) Monitoring information
(b) Analyzing information
(c) Making volume migration plan (decision of volume migration)
(d) Moving volume (migration)
(a') Monitoring information again to confirm condition and effect of the performed tuning.

Preset function eases users' work by providing typical tuning method of HIHSM based on parameters given by users, and performs tuning plan automatically. When (b) Analyzing and (c) Making plan are done by users, it can make fine tuning.



(3) Process of tuning by preset function

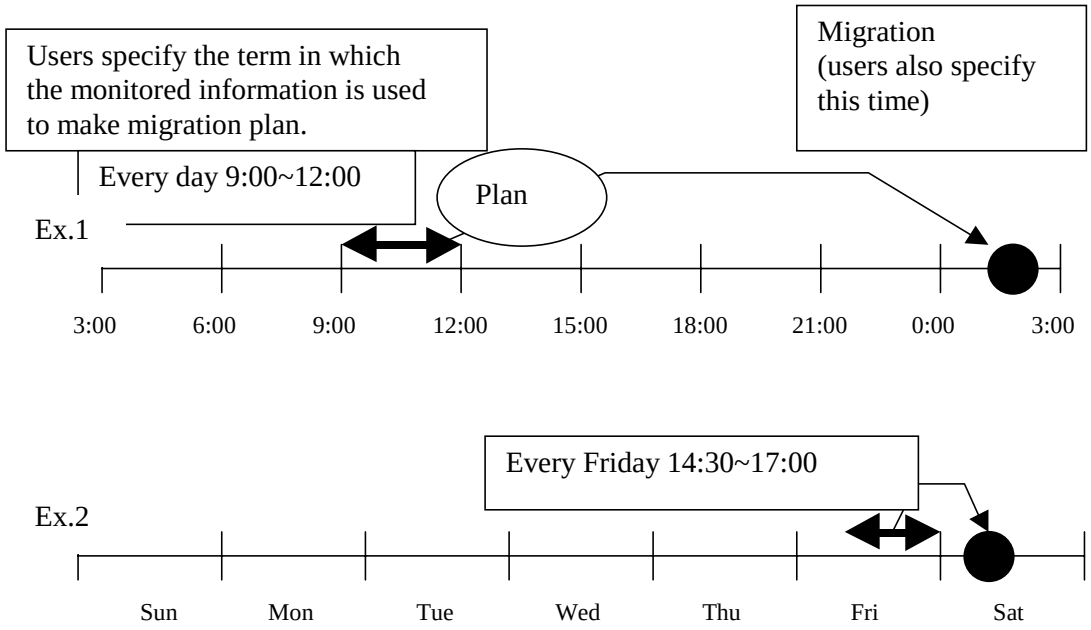
(a) Process of tuning by preset function

Preset function performs following two process.

- Making Plan: Detecting volumes that have problems by monitored information (disk utilization) and making plan of volume Migration
- Migration: Moving volume by the plan.

These two process are repeated by the cycle described above, and users give the cycle. In preset function, this cycle is supposed from one day to one week or several weeks.

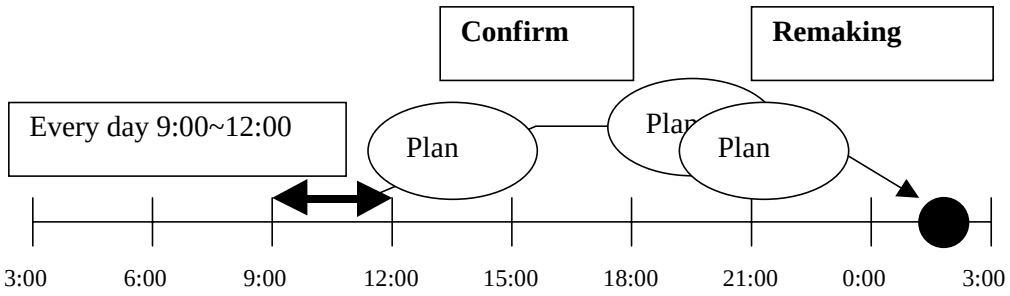
Users can specify the term in which monitored information is used to make the plan from concern of users. For example, users can specify the highest load term in a day or in a week as the referred term to making plan.



(b) Confirm and remaking of plan by users

From making plan by preset function until migration, users can see the plan and delete it if needed.

From making plan by preset function until migration, users can remake the plan manually with changing parameters.



(4) Making plan

(a) Tuning based on disk utilization

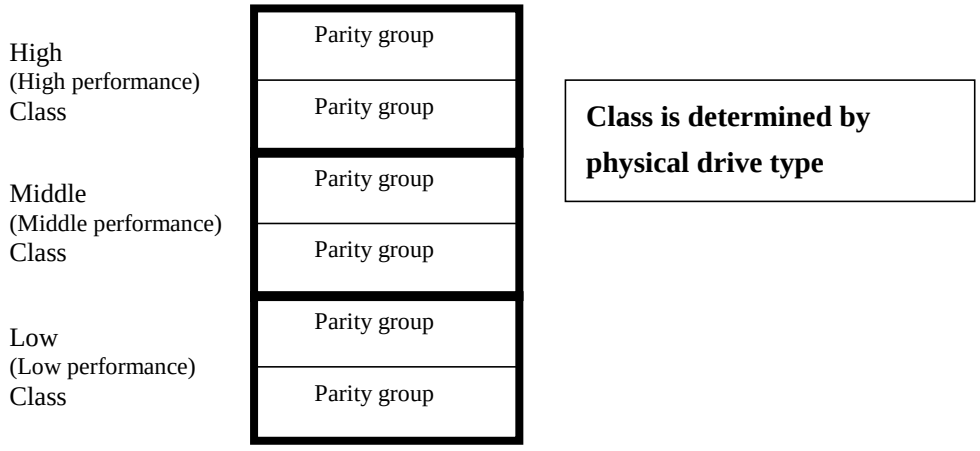
HIHSM monitors utilization of various system resources, but preset function uses disk utilization to make migration plan to solve disk neck as typical tuning method. Users should refer other information if needed.

Users also specified parameters for tuning based on disk utilization. Users should these parameters for their system (preset function provides default value for these parameters roughly).

(b) Hierarchy of parity groups and management by class

Parity groups in DKC have hierarchy by drive type and RAID type. HIHSM optimizes usage of this hierarchy. Preset function manages this hierarchy as class (parity group set).

Parity groups are divided into classes. The classes are ordered from high level (high performance) to low level (low performance). This classification is decided by performance of physical drive type of each parity group.



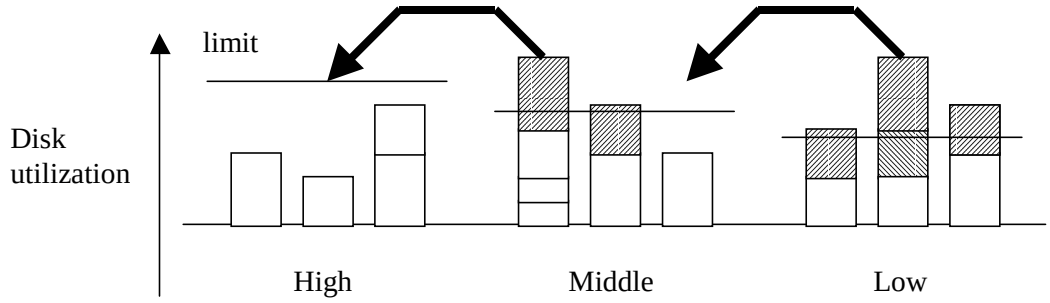
Preset function makes migration plan by the method described bellow. Preset function uses monitored information in the term that users specify to be referred to make the plan.

(b-1) Management by maximum limit of disk utilization

Maximum limit of disk utilization (parity group utilization) is specified to each parity group. Users should specify this limit for their system. (HIHSM uses default value but you should set them.)

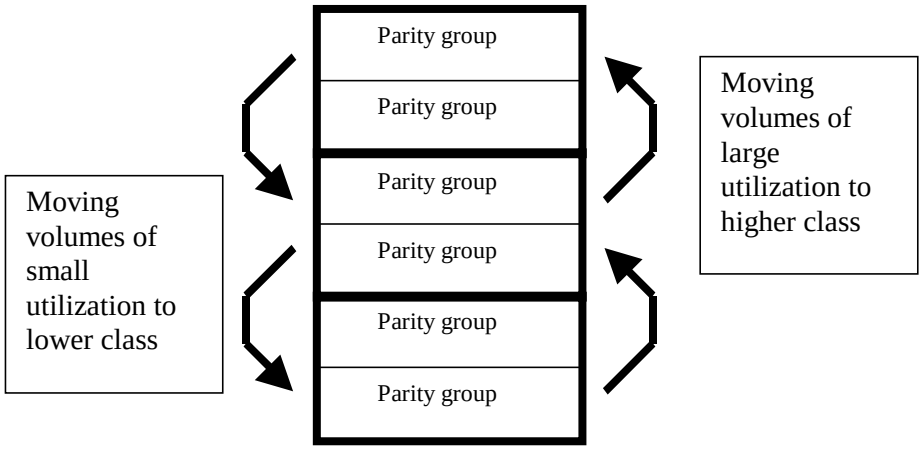
In parity groups that exceed this limit, preset function makes plan of moving volumes in this parity group from this parity groups to another parity group in higher class.

This avoids physical disk neck and provides load balancing of disk utilization.



(b-2) Selection of volumes to migration

In the migration plan described above, volumes of lager disk utilization are selected to be moved from the parity group that exceeds the limit to higher class. Moving lager utilization volumes to higher performance class is expected to makes large tuning effect. And lager utilization means lager amount of access from host, and this also makes large tuning effect. When the reserved (empty) volumes run short in the high class, volumes of smaller disk utilization are selected to be moved from higher class to lower class to make reserve volumes.



Preset function provides some criteria for you to select them, average of disk utilization, average of highest Nth value of disk utilization in the referred term, and a value considering sequential/random access pattern.

(b-3) Specifying maximum limit of disk utilization by users

When same maximum limit of disk utilization described above is given to each class (parity group) classified by drive types, performance of each physical drive type makes performance of each class directly.

When the users specify the limit to each class with bias, users can make the difference of performance of classes lager or smaller.

Users can specify parity groups to fixed parity groups in which volumes do not be moved automatically. Preset function does not make a migration plan about fixed parity groups and volumes in the fixed parity groups.

(b-4) Notice for making plan

HIHSM can make these plan only on the following conditions.

- HIHSM can estimate the disk utilization against all migrated parity groups.
- The disk utilization rate of all migrated parity groups are not over the maximum rate. If the rate of one parity group is over. HIHSM could not make a plan.

(b-5) Notice for reference term

HIHSM could not use old information before the last volume migration in order to reduce the influence of performance by volume migration.

Therefore, HIHSM sometimes fails making a plan by lack of information.

(5) Moving (migration) by preset function

Preset function performs moving (migration) process once a day at time specified by users.

If there is a migration plan made by preset function at the time, volumes are moved by the plan.

Users can specify limit to moving to avoid overload by moving (data copy) process.

If the disk utilization of parity groups in which moving is started exceed the moving process limit, the moving in the parity group will be aborted.

Users can specify time limit to moving. If plan is not completed in the time limit, remain of plan will be executed in next day. If the new plan will be made until the next migration time, those remain plans will be deleted before making the plan.

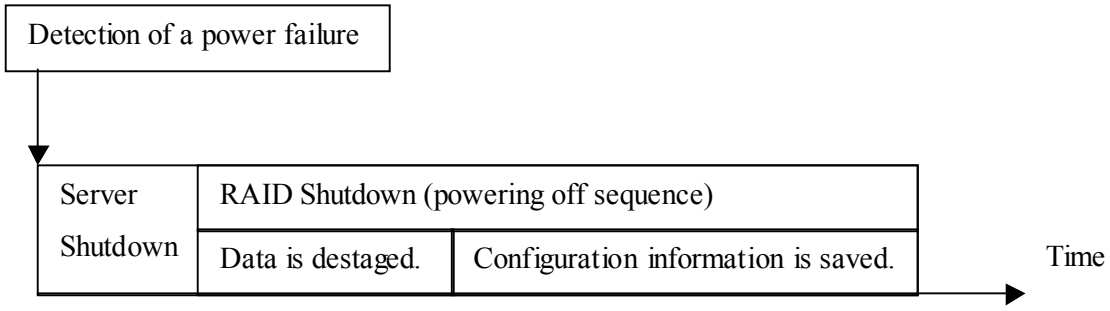
3.19 UPS

3.19.1 Outline

When a power failure occurs, data on the cache is sustained for 48 hours by means of the backup of the memory by the battery. In addition to the above, when the UPS is connected, the UPS functions to enable the subsystem to continue execution of host I/O's when an instantaneous power failure occurs and to save data when a long power failure occurs by backing up the whole subsystem with a battery.

3.19.2 Operation sequence

When a power failure continues, the connected UPS backs up the whole subsystem with a battery, and while the backup is done, the subsystem destages data stored on the cache and saves configuration information by making the powering off sequence operate.



3.19.3 Limiting amount of data which flows into the cache

It is required to control amount of data which is on the cache and has not been destaged to the drive beforehand in order to control the time taken for the powering off. To satisfy the requirement, amount of non-destaged data which flows in the cache is limited for each parity group. However, the amount of the data may exceed the allowable limit owing to a failure of the cache or drive.

3.19.4 Setting a limit on the amount of data which flows in the cache

When connecting the UPS, set the PS Off Timer on the SVP screen. (Refer to the item of system option in the SVP Section.)
For the setting value of the PS Off Timer, refer to the UPS Connection Manual (a separate volume).

3.20 HPAV

3.20.1 Overview

3.20.1.1 Overview of HPAV

HPAV (Hitachi Parallel Access Volume) enables a host computer to issue multiple I/O requests in parallel to each device in the disk subsystem. Usually, host computers are able to issue only a single I/O request to a single device. When a host computer issues one I/O request to a device, the host computer is unable to issue another I/O request to that device. However, HPAV enables you to assign one or more aliases to a single device so that the host computer is able to issue multiple I/O requests. In this way, HPAV provides the host computer with substantially faster access to data in the disk subsystem.

When you assign aliases to a device, you specify the addresses of unused LDEVs (*logical devices* or *logical volumes*) in the disk subsystem. The specified addresses are used as alias addresses.

Throughout this manual, the term *base device* refers to a device to which aliases will be assigned. Also, the term *alias device* refers to an alias.

HPAV operates in either of the following ways: *static PAV* and *dynamic PAV*. These are described next:

- Static PAV
  When static PAV is used, the number of aliases for each base device remains unchanged even if the number of I/O requests to each device changes. As explained later, when dynamic PAV is used, the number of aliases for a base device is likely to increase as the number of I/O requests to the device increases; this means the number of aliases for other base devices may decrease. However, when static PAV is used, the number of aliases remains as specified by the Remote Console user or SVP operation user.
  Before you assign aliases to base devices, you should consider whether I/O requests will converge on some of the base devices. We recommend that you assign more aliases on base devices on which I/O requests are expectedly converge. Otherwise, HPAV might not be able to provide much faster access to data in the disk subsystem.
  The following figure gives an example of static PAV. In this figure, each of the three base devices (numbered 10, 11, and 12, respectively) has two aliases assigned. I/O requests converge on the base device #10, but the number of aliases for each base device remains unchanged.

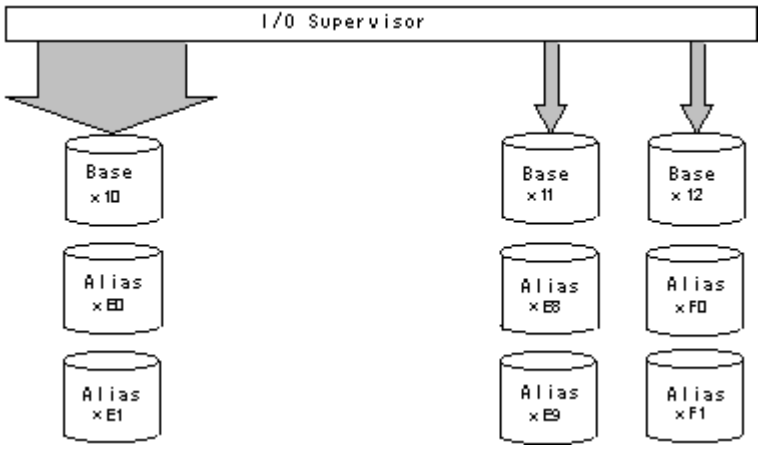


Fig. 3.20.1.1-1 Static PAV

- Dynamic PAV

  When dynamic PAV is used, the number of aliases for a base device may change as the number of I/O requests to the device changes. If I/O requests converge on some of the base devices, the number of aliases may increase for these base devices but may decrease for the other base devices. Dynamic PAV can balance workloads on base devices and optimize the speed for accessing data in the disk subsystem.

  The following figure gives an example of dynamic PAV. In this example, each of the three base devices (#10, #11, and #12) was originally assigned two aliases. As I/O requests converge on #10, the number of aliases for #10 increases to four. For the base devices #11 and #12, the number of aliases decreases to one.

  Dynamic PAV requires the Workload Manager (WLM), a special function provided by the operating system at the host computer. For details, see sections 3.20.2.1 and 3.20.2.2.2.

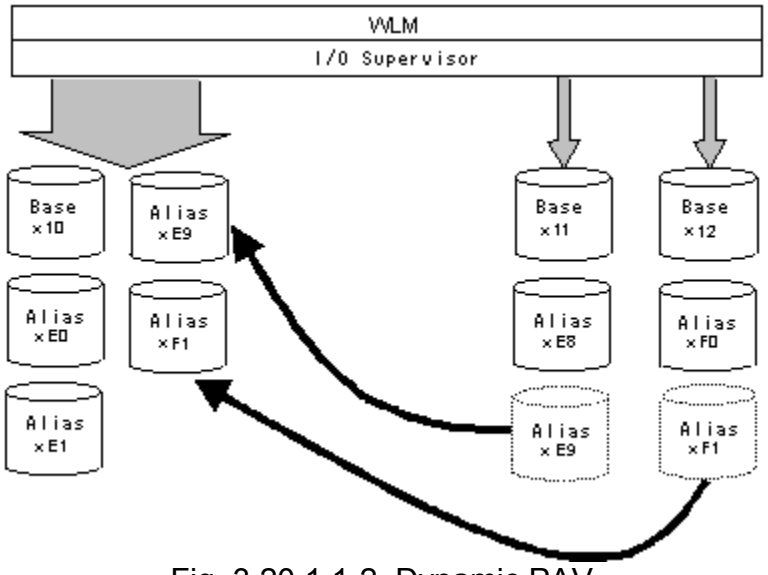


Fig. 3.20.1.1-2 Dynamic PAV

3.20.1.2 How to Obtain Optimum Results from HPAV

To obtain good results from HPAV, you should be aware of the following:

- The best results can be obtained if the number of aliases is "Number of available channel paths minus 1". If the number of aliases is specified this way, I/O operations can use all the channel paths, and thus the best results can be obtained.

- HPAV may not produce good results when many channel paths are used. If all the channel paths are used, no good results can be expected.

- HPAV lets you assign unused devices for use as aliases. If you assign most of the unused devices for use as aliases, only a small amount of free devices are available. It is recommended that you think about adding more disks in the future when you determine the number of aliases to be assigned.
  If we assume that there are 256 devices and we assign the same number of alias devices to each base devices, the number of base devices and alias devices is calculated as explained in Table 3.20.1.2. The recommended ratio of base devices to alias devices is 1:3.
  If you can expect the types of jobs to be passed to base devices, or if you can expect how many accesses should be made to each base device, you should determine the number of aliases for each base device so that it meets the requirements for each base device.

Table 3.20.1.2 The ratio of base devices to aliases

Ratio (base devices : alias devices)	The number of base devices	The number of alias devices
1:3 (recommended)	64	192
1:1	128	128

- Good results cannot be expected on devices that are always shared and used by multiple host computers.

- If dynamic PAV can be used in all the systems, good results can be expected if you assign 8 to 16 aliases to each CU (control unit).

3.20.2 Preparing for HPAV Operations

3.20.2.1 System Requirements

To be able to run, HPAV requires the following operating systems to be installed on the host computer:

- For static PAV
  - OS/390 V1R3 & DFSMS/MVS 1.3 with PTF
  - VM/ESA 2.4.0
- For dynamic PAV
  - OS/390 V2R7 with PTF & DFSMS/MVS 1.5

Note: To perform operations with HPAV, you must have administrator access privileges. Users who do not have administrator access privileges can only view HPAV information. The following restrictions apply when using HPAV.

Table 3.20.1.1 Restrictions that apply when using HPAV

No.	Item	Specifications
1	DKC emulation type	I-2105
2	DKU emulation type	3390-1, 3390-2, 3390-3, 3390-3R, 3390-9, 3390-L
3	SSID Boundary	Should be set 256 LDEV/SSID
4	Number of aliases that can be assigned to a single base device	Up to 15
5	Alias device numbers that can be used	When you set aliases for base devices, you can use the device numbers of unused devices as the alias device numbers. When you set aliases for base devices, you must be aware that the alias devices and the base devices must belong to the same CU.
6	Device functions that can concurrently be used with HPAV	• CVS (Customized Volume Size) • DCR (Dynamic Cache Residence) • LDEV Security
7	Device functions that cannot concurrently be used with HPAV	• HMBR (Hitachi Multiplatform Backup/Restore) • HMDE
8	Caution of Copy function	• Concurrent Copy/XRC - Do not intermix DKC emulation type of 2105 and 3990-3/6/6E in the MCU. • HRC (Hitachi Remote Copy) / HMRCF (Hitachi Multiple RAID Coupling Feature) / HIHSM (Hitachi Internal Hierarchical Storage Manager) - You can be intermixed DKC emulation type of 2105 and 3990-3/6/6E in the same DKC.

3.20.2.2 Preparations at the Host Computer

This section briefly describes arrangements that should be made at the host computer. For detailed information, see the documentation for MVS.

3.20.2.2.1 Generation Definition of Base Devices and Alias Addresses

The address mapping between base devices and the corresponding alias devices must be defined at generation.
The address mapping between base devices and alias devices at the host computer should match the corresponding address mapping at the DKC side. If it does not match, a serious failure might occur during data processing.

The following gives an example of mapping between base devices and aliases devices:

(A)	(B)	(C)	(D)
x 00-x1F:Base x 20-xFF:Alias	x 00-x3F:Base x 40-x7F:Alias x 80-xBF:Base x C0-xFF:Alias	x 00-x7F:Alias x 80-xFF:Base	x 00-x3F:Alias x 40-x7F:Base x 80-xBF:Alias x C0-xFF:Base

Note: The recommended ratio of base devices to aliases is 1:3, if each base device is assumed to be assigned the same number of aliases.

3.20.2.2.2 Setting the WLM Operation Mode

If you want to use dynamic PAV, you must set the WLM (Workload Manager) operation mode to *goal mode*. WLM manages workloads on MVS and can use two operation modes, which include goal mode. In goal mode, WLM manages the system to fulfill the performance goal that was specified before the system began to operate.
You should be aware that static PAV is used instead of dynamic PAV if compatibility mode is used instead of goal mode.
For details on the WLM operation modes, see the documentation for MVS.

3.21 FICON

3.21.1 Introduction

FICON is new mainframe architecture, which is FC-SB-2 protocol based on Fiber channel physical layer protocol (FC-PH) and it is approved by ANSI.

The specification of FICON is below.
- Full duplex data transfer
- Multiple concurrent I/O operations on channel
- High bandwidth data transfer (100MB/s)
- Fewer control unit interfaces
- Pipelined CCW execution

3.21.2 Environment

If you use FICON, below environment is needed.

Table 3.21.2-1

Items	Contents
CPU	z900, G5/G6
OS	OS/390 Rev2.6 and later released
	Z-OS R1.V1 and later released
DKC Emulation type	2105-F20
FICON Director	ED-5000
	ED-6064
	FC9000-64
FICON support DKC Main micro program version	01-17-xx or higher

3.21.3 DKC410I FICON specification

Table 3.21.3-1 shows the specification of DKC410I FICON.

Table 3.21.3-1 DKC410I FICON specification

Items		Contents
Number of ports	Max. per subsystem	16
	Per PCB	2
	Install unit	4
Connectable channel Image per port		32
CU image per port		16
Number of logical paths per CU port		512
Logical paths per subsystem		8192
Support DKC emulation type		2105-F20
Support fiber channel	Bandwidth	1 Gbps
	Cable and connector	SC-Duplex
	Mode	Single Mode Fiber/Multi Mode Fiber

Functions supported by DKC410I FICON are below.
- CVS.DCR
- LDEV Security
- Multiple Allegiance
- HRC w/ OPEN-FC Link (Table 3.21.3-2)
- MRCF
- Concurrent Copy (Table 3.21.3-3)
- HXRC (Table 3.21.3-4)
- HIHSM
- HMDE

Table 3.21.3-2 HRC with FICON

MCU-RCU / CHL-MCU	ESCON	Open Fiber
ESCON	Supported	Supported
FICON	Not supported (*1)	Supported

*1: The link bandwidth of FICON is greater than that of ESCON. Then, in consideration of performance balance of ESCON path and FICON path, if FICON path is used between Channel and MCU, the path between MCU and RCU should be OPEN-FC link, not ESCON link.

Table 3.21.3-3 Concurrent Copy with FICON

		Concurrent Copy data transfer path (SMS – DKC)	
		FICON	ESCON
application site (System and DKC)	ESCON	Supported	Supported
	FICON	Supported	Not recommended (*2)

*2: If the path of Application site is FICON,SMS (Concurrent copy data transfer path) should be also FICON in order to balance the performance of Application path and SMS path.

Table 3.21.3-4 HXRC and FICON configuration

		Record set transfer path (System Data Mover - DKC)	
		FICON	ESCON
application site (System and DKC)	ESCON	Supported	Supported
	FICON	Supported	Not recommended (*3)

*3: If the path of Application site is FICON, System Data Mover (SDM) path should be also FICON in order to balance the performance of Application path and SDM path.

3.21.4 Configuration

(1) Topology

The pattern of connection between DKC410I FICON and channel are such as below.

- Point to point connection
- Switched point to point

FICON director (FICON switch) specifications are below.

Table 3.21.4-1 FICON switch specification

Switch		ED-5000	ED-6064	FC-9000-64
Bandwidth		1Gbps	2Gbps	1Gbps
Connector		SC-Duplex	LC-Duplex	SC-Duplex
Cable	CPU – Switch or Switch -DKC	SC-Duplex to SC-Duplex	LC-Duplex to SC-Duplex	SC-Duplex to SC-Duplex

[Note]

1. Above switches are all that RSD certified connectivity between DKC410I FICON and each switch.
2. Cascading switch configuration is not available.
3. Regarding switch zoning, no mixed mode supports through switch/directors. Then in case of FICON only, zoning is available.

(2) Configuration of FICON and ESCON intermixed

(a) Migration from ESCON to FICON

FICON/ESCON intermix within the same path group is allowed for only migration from ESCON to FICON.

(b) FICON and ESCON intermixed configuration within the same subsystem

If the generation of ESCON and that of FICON are separated, each system can use the same volume, without disturbing each other.

4. Power-on Sequences

4.1 IMPL Sequence

The IMPL sequence, which is executed when power is turned on, is executed by the following four modules:

(1) Boot loader
The boot loader performs the minimum necessary amount of initializations after a ROM boot. Subsequently, the boot loader expands the local memory loader from flash memory into local memory and transfers control to the local memory loader.

(2) Local memory loader
The local memory loader loads the Real Time OS load modules into local memory. Subsequently, the local memory loader transfers control to Real Time OS.

(3) Real Time OS
Real Time OS is a root task that initializes the tables in local memory that are used for intertask communications. Real Time OS also tests the hardware resources.

(4) DKC task
When the DKC task is created, it executes initialization routines. It initializes the most part of the environment it uses. When the environment is established so that the DKC task can start scanning, it notifies the SVP of a power event log. Subsequently, the DKC task turns on the power to the physical drives and, when the logical drives get ready, notifies the host processor of an NRTR.

The control flow of IMPL processing is shown in Fig. 4.1.

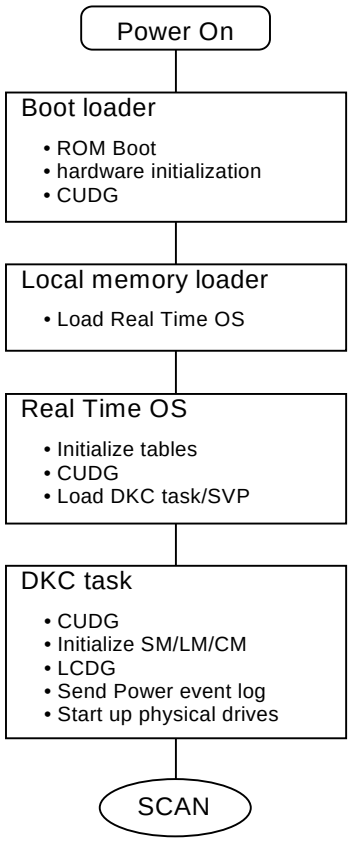


Fig. 4.1 IMPL Sequence

4.2 Drive Power-on Sequence

An overcurrent condition will occur if two or more drives are started at the same time. To preclude the overcurrent condition, DKUs are started at the power supply level, one at a time, at approximately 10 second intervals.

When the logical devices get ready as the result of the startup of the physical drives, the host processor is notified to that effect.

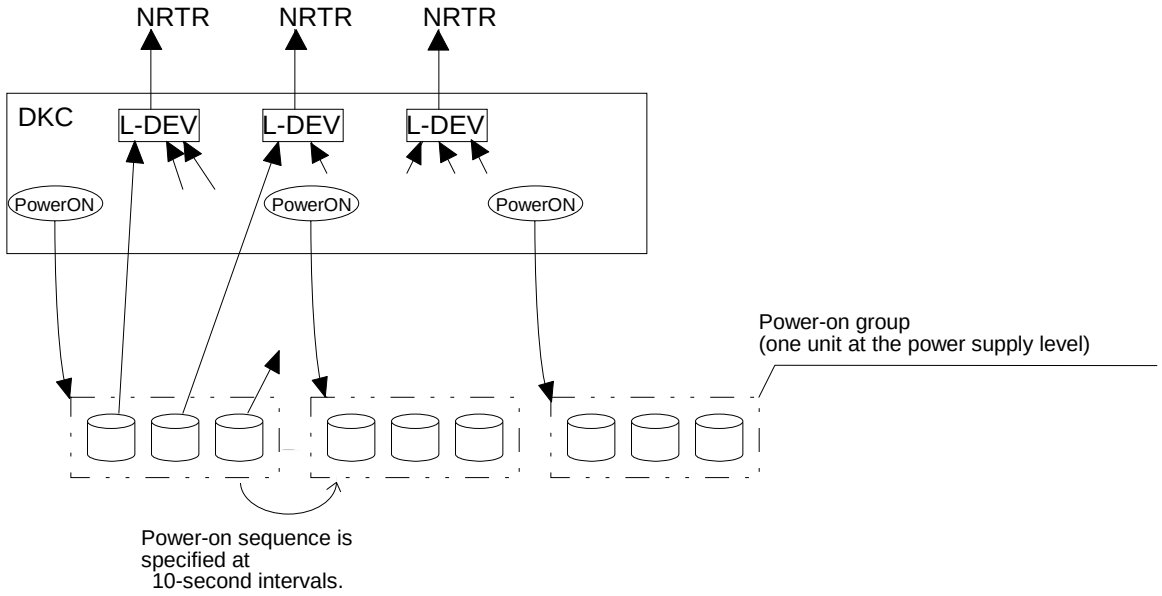


Fig. 4.2 Drive Power-on Sequence

4.3 Planned Stop

When a power-off is specified by a maintenance personnel, this subsystem checks for termination of tasks that are blocked or running on all logical devices. When all the tasks are terminated, this subsystem disables the CHL and executes emergency destaging. If a track for which destaging fails (pinned track) occurs, this subsystem stores the pin information in shared memory. Subsequently, this subsystem saves the pin information, which is used as hand-over information, in flash memory, sends Power Event Log to the SVP, and notifies the hardware of the grant to turn off the power.

The hardware turns off main power when power-off grants for all processors are presented.

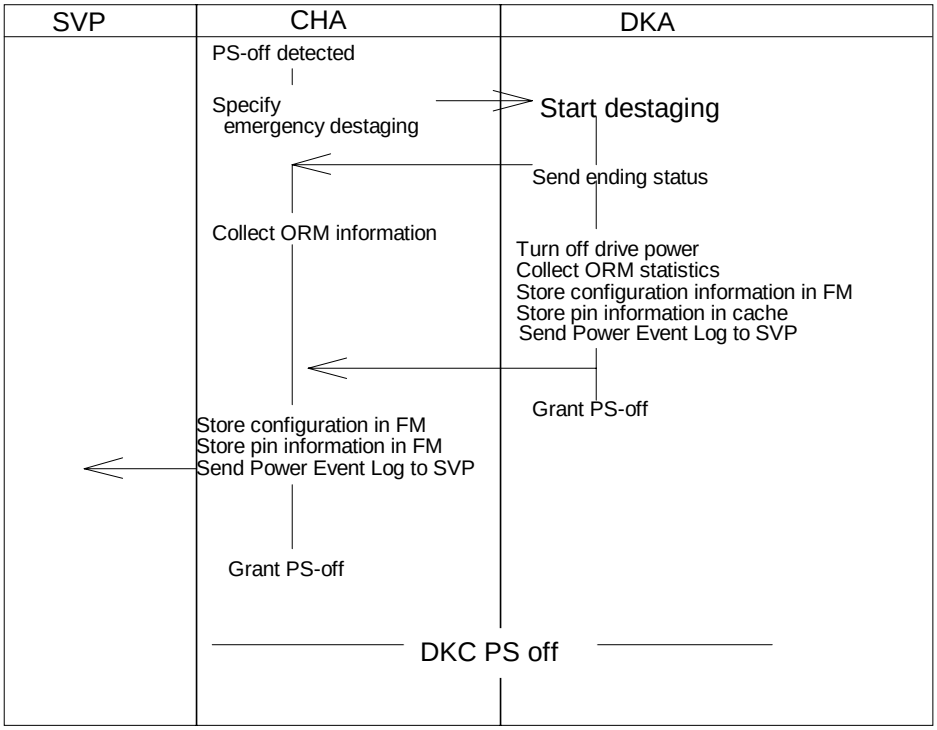


Fig. 4.3 Planned Stop Sequence

5. Appendixes

5.1 Physical - Logical Device Matrixes for Separate Model

Relationship between disk drive number and parity group number (1/12)

HDD BOX number	Disk drive number	C# / R#	Parity group number
HDU-R10	HDU-R100	0/0	01-01
	HDU-R101	0/1	01-02
	HDU-R102	0/2	01-03
	HDU-R103	0/3	01-04
	HDU-R104	0/4	01-05
	HDU-R105	0/5	01-06
	HDU-R106	0/6	01-07
	HDU-R107	0/7	01-08
	HDU-R108	0/8	01-09
	HDU-R109	0/9	01-0A
	HDU-R10A	0/A	01-0B
	HDU-R10B	0/B	01-0C
HDU-R11	HDU-R110	1/0	01-01
	HDU-R111	1/1	01-02
	HDU-R112	1/2	01-03
	HDU-R113	1/3	01-04
	HDU-R114	1/4	01-05
	HDU-R115	1/5	01-06
	HDU-R116	1/6	01-07
	HDU-R117	1/7	01-08
	HDU-R118	1/8	01-09
	HDU-R119	1/9	01-0A
	HDU-R11A	1/A	01-0B
	HDU-R11B	1/B	01-0C
HDU-R12	HDU-R120	2/0	01-01
	HDU-R121	2/1	01-02
	HDU-R122	2/2	01-03
	HDU-R123	2/3	01-04
	HDU-R124	2/4	01-05
	HDU-R125	2/5	01-06
	HDU-R126	2/6	01-07
	HDU-R127	2/7	01-08
	HDU-R128	2/8	01-09
	HDU-R129	2/9	01-0A
	HDU-R12A	2/A	01-0B
	HDU-R12B	2/B	01-0C
HDU-R13	HDU-R130	3/0	01-01
	HDU-R131	3/1	01-02
	HDU-R132	3/2	01-03
	HDU-R133	3/3	01-04
	HDU-R134	3/4	01-05
	HDU-R135	3/5	01-06
	HDU-R136	3/6	01-07
	HDU-R137	3/7	01-08
	HDU-R138	3/8	01-09
	HDU-R139	3/9	01-0A
	HDU-R13A	3/A	01-0B
	HDU-R13B	3/B	01-0C

Relationship between disk drive number and parity group number (2/12)

HDD BOX number	Disk drive number	C# / R#	Parity group number
HDU-R14	HDU-R140	4/0	02-01
	HDU-R141	4/1	02-02
	HDU-R142	4/2	02-03
	HDU-R143	4/3	02-04
	HDU-R144	4/4	02-05
	HDU-R145	4/5	02-06
	HDU-R146	4/6	02-07
	HDU-R147	4/7	02-08
	HDU-R148	4/8	02-09
	HDU-R149	4/9	02-0A
	HDU-R14A	4/A	02-0B
	HDU-R14B	4/B	02-0C
HDU-R15	HDU-R150	5/0	02-01
	HDU-R151	5/1	02-02
	HDU-R152	5/2	02-03
	HDU-R153	5/3	02-04
	HDU-R154	5/4	02-05
	HDU-R155	5/5	02-06
	HDU-R156	5/6	02-07
	HDU-R157	5/7	02-08
	HDU-R158	5/8	02-09
	HDU-R159	5/9	02-0A
	HDU-R15A	5/A	02-0B
	HDU-R15B	5/B	02-0C
HDU-R16	HDU-R160	6/0	02-01
	HDU-R161	6/1	02-02
	HDU-R162	6/2	02-03
	HDU-R163	6/3	02-04
	HDU-R164	6/4	02-05
	HDU-R165	6/5	02-06
	HDU-R166	6/6	02-07
	HDU-R167	6/7	02-08
	HDU-R168	6/8	02-09
	HDU-R169	6/9	02-0A
	HDU-R16A	6/A	02-0B
	HDU-R16B	6/B	02-0C
HDU-R17	HDU-R170	7/0	02-01
	HDU-R171	7/1	02-02
	HDU-R172	7/2	02-03
	HDU-R173	7/3	02-04
	HDU-R174	7/4	02-05
	HDU-R175	7/5	02-06
	HDU-R176	7/6	02-07
	HDU-R177	7/7	02-08
	HDU-R178	7/8	02-09
	HDU-R179	7/9	02-0A
	HDU-R17A	7/A	02-0B
	HDU-R17B	7/B	02-0C

Relationship between disk drive number and parity group number (3/12)

HDD BOX number	Disk drive number	C# / R#	Parity group number
HDU-R20	HDU-R200	0/C	05-01
	HDU-R201	0/D	05-02
	HDU-R202	0/E	05-03
	HDU-R203	0/F	05-04
	HDU-R204	0/10	05-05
	HDU-R205	0/11	05-06
	HDU-R206	0/12	05-07
	HDU-R207	0/13	05-08
	HDU-R208	0/14	05-09
	HDU-R209	0/15	05-0A
	HDU-R20A	0/16	05-0B
	HDU-R20B	0/17	05-0C
HDU-R21	HDU-R210	1/C	05-01
	HDU-R211	1/D	05-02
	HDU-R212	1/E	05-03
	HDU-R213	1/F	05-04
	HDU-R214	1/10	05-05
	HDU-R215	1/11	05-06
	HDU-R216	1/12	05-07
	HDU-R217	1/13	05-08
	HDU-R218	1/14	05-09
	HDU-R219	1/15	05-0A
	HDU-R21A	1/16	05-0B
	HDU-R21B	1/17	05-0C
HDU-R22	HDU-R220	2/C	05-01
	HDU-R221	2/D	05-02
	HDU-R222	2/E	05-03
	HDU-R223	2/F	05-04
	HDU-R224	2/10	05-05
	HDU-R225	2/11	05-06
	HDU-R226	2/12	05-07
	HDU-R227	2/13	05-08
	HDU-R228	2/14	05-09
	HDU-R229	2/15	05-0A
	HDU-R22A	2/16	05-0B
	HDU-R22B	2/17	05-0C
HDU-R23	HDU-R230	3/C	05-01
	HDU-R231	3/D	05-02
	HDU-R232	3/E	05-03
	HDU-R233	3/F	05-04
	HDU-R234	3/10	05-05
	HDU-R235	3/11	05-06
	HDU-R236	3/12	05-07
	HDU-R237	3/13	05-08
	HDU-R238	3/14	05-09
	HDU-R239	3/15	05-0A
	HDU-R23A	3/16	05-0B
	HDU-R23B	3/17	05-0C

Relationship between disk drive number and parity group number (4/12)

HDD BOX number	Disk drive number	C# / R#	Parity group number
HDU-R24	HDU-R240	4/C	06-01
	HDU-R241	4/D	06-02
	HDU-R242	4/E	06-03
	HDU-R243	4/F	06-04
	HDU-R244	4/10	06-05
	HDU-R245	4/11	06-06
	HDU-R246	4/12	06-07
	HDU-R247	4/13	06-08
	HDU-R248	4/14	06-09
	HDU-R249	4/15	06-0A
	HDU-R24A	4/16	06-0B
	HDU-R24B	4/17	06-0C
HDU-R25	HDU-R250	5/C	06-01
	HDU-R251	5/D	06-02
	HDU-R252	5/E	06-03
	HDU-R253	5/F	06-04
	HDU-R254	5/10	06-05
	HDU-R255	5/11	06-06
	HDU-R256	5/12	06-07
	HDU-R257	5/13	06-08
	HDU-R258	5/14	06-09
	HDU-R259	5/15	06-0A
	HDU-R25A	5/16	06-0B
	HDU-R25B	5/17	06-0C
HDU-R26	HDU-R260	6/C	06-01
	HDU-R261	6/D	06-02
	HDU-R262	6/E	06-03
	HDU-R263	6/F	06-04
	HDU-R264	6/10	06-05
	HDU-R265	6/11	06-06
	HDU-R266	6/12	06-07
	HDU-R267	6/13	06-08
	HDU-R268	6/14	06-09
	HDU-R269	6/15	06-0A
	HDU-R26A	6/16	06-0B
	HDU-R26B	6/17	06-0C
HDU-R27	HDU-R270	7/C	06-01
	HDU-R271	7/D	06-02
	HDU-R272	7/E	06-03
	HDU-R273	7/F	06-04
	HDU-R274	7/10	06-05
	HDU-R275	7/11	06-06
	HDU-R276	7/12	06-07
	HDU-R277	7/13	06-08
	HDU-R278	7/14	06-09
	HDU-R279	7/15	06-0A
	HDU-R27A	7/16	06-0B
	HDU-R27B	7/17	06-0C

Relationship between disk drive number and parity group number (5/12)

HDD BOX number	Disk drive number	C# / R#	Parity group number
HDU-R30	HDU-R300	0/18	09-01
	HDU-R301	0/19	09-02
	HDU-R302	0/1A	09-03
	HDU-R303	0/1B	09-04
	HDU-R304	0/1C	09-05
	HDU-R305	0/1D	09-06
	HDU-R306	0/1E	09-07
	HDU-R307	0/1F	09-08
HDU-R31	HDU-R310	1/18	09-01
	HDU-R311	1/19	09-02
	HDU-R312	1/1A	09-03
	HDU-R313	1/1B	09-04
	HDU-R314	1/1C	09-05
	HDU-R315	1/1D	09-06
	HDU-R316	1/1E	09-07
	HDU-R317	1/1F	09-08
HDU-R32	HDU-R320	2/18	09-01
	HDU-R321	2/19	09-02
	HDU-R322	2/1A	09-03
	HDU-R323	2/1B	09-04
	HDU-R324	2/1C	09-05
	HDU-R325	2/1D	09-06
	HDU-R326	2/1E	09-07
	HDU-R327	2/1F	09-08
HDU-R33	HDU-R330	3/18	09-01
	HDU-R331	3/19	09-02
	HDU-R332	3/1A	09-03
	HDU-R333	3/1B	09-04
	HDU-R334	3/1C	09-05
	HDU-R335	3/1D	09-06
	HDU-R336	3/1E	09-07
	HDU-R337	3/1F	09-08

Relationship between disk drive number and parity group number (6/12)

HDD BOX number	Disk drive number	C# / R#	Parity group number
HDU-R34	HDU-R340	4/18	0A-01
	HDU-R341	4/19	0A-02
	HDU-R342	4/1A	0A-03
	HDU-R343	4/1B	0A-04
	HDU-R344	4/1C	0A-05
	HDU-R345	4/1D	0A-06
	HDU-R346	4/1E	0A-07
	HDU-R347	4/1F	0A-08
HDU-R35	HDU-R350	5/18	0A-01
	HDU-R351	5/19	0A-02
	HDU-R352	5/1A	0A-03
	HDU-R353	5/1B	0A-04
	HDU-R354	5/1C	0A-05
	HDU-R355	5/1D	0A-06
	HDU-R356	5/1E	0A-07
	HDU-R357	5/1F	0A-08
HDU-R36	HDU-R360	6/18	0A-01
	HDU-R361	6/19	0A-02
	HDU-R362	6/1A	0A-03
	HDU-R363	6/1B	0A-04
	HDU-R364	6/1C	0A-05
	HDU-R365	6/1D	0A-06
	HDU-R366	6/1E	0A-07
	HDU-R367	6/1F	0A-08
HDU-R37	HDU-R370	7/18	0A-01
	HDU-R371	7/19	0A-02
	HDU-R372	7/1A	0A-03
	HDU-R373	7/1B	0A-04
	HDU-R374	7/1C	0A-05
	HDU-R375	7/1D	0A-06
	HDU-R376	7/1E	0A-07
	HDU-R377	7/1F	0A-08

Relationship between disk drive number and parity group number (7/12)

HDD BOX number	Disk drive number	C# / R#	Parity group number
HDU-L10	HDU-L100	8/0	03-01
	HDU-L101	8/1	03-02
	HDU-L102	8/2	03-03
	HDU-L103	8/3	03-04
	HDU-L104	8/4	03-05
	HDU-L105	8/5	03-06
	HDU-L106	8/6	03-07
	HDU-L107	8/7	03-08
	HDU-L108	8/8	03-09
	HDU-L109	8/9	03-0A
	HDU-L10A	8/A	03-0B
	HDU-L10B	8/B	03-0C
HDU-L11	HDU-L110	9/0	03-01
	HDU-L111	9/1	03-02
	HDU-L112	9/2	03-03
	HDU-L113	9/3	03-04
	HDU-L114	9/4	03-05
	HDU-L115	9/5	03-06
	HDU-L116	9/6	03-07
	HDU-L117	9/7	03-08
	HDU-L118	9/8	03-09
	HDU-L119	9/9	03-0A
	HDU-L11A	9/A	03-0B
	HDU-L11B	9/B	03-0C
HDU-L12	HDU-L120	A/0	03-01
	HDU-L121	A/1	03-02
	HDU-L122	A/2	03-03
	HDU-L123	A/3	03-04
	HDU-L124	A/4	03-05
	HDU-L125	A/5	03-06
	HDU-L126	A/6	03-07
	HDU-L127	A/7	03-08
	HDU-L128	A/8	03-09
	HDU-L129	A/9	03-0A
	HDU-L12A	A/A	03-0B
	HDU-L12B	A/B	03-0C
HDU-L13	HDU-L130	B/0	03-01
	HDU-L131	B/1	03-02
	HDU-L132	B/2	03-03
	HDU-L133	B/3	03-04
	HDU-L134	B/4	03-05
	HDU-L135	B/5	03-06
	HDU-L136	B/6	03-07
	HDU-L137	B/7	03-08
	HDU-L138	B/8	03-09
	HDU-L139	B/9	03-0A
	HDU-L13A	B/A	03-0B
	HDU-L13B	B/B	03-0C

Relationship between disk drive number and parity group number (8/12)

HDD BOX number	Disk drive number	C# / R#	Parity group number
HDU-L14	HDU-L140	C/0	04-01
	HDU-L141	C/1	04-02
	HDU-L142	C/2	04-03
	HDU-L143	C/3	04-04
	HDU-L144	C/4	04-05
	HDU-L145	C/5	04-06
	HDU-L146	C/6	04-07
	HDU-L147	C/7	04-08
	HDU-L148	C/8	04-09
	HDU-L149	C/9	04-0A
	HDU-L14A	C/A	04-0B
	HDU-L14B	C/B	04-0C
HDU-L15	HDU-L150	D/0	04-01
	HDU-L151	D/1	04-02
	HDU-L152	D/2	04-03
	HDU-L153	D/3	04-04
	HDU-L154	D/4	04-05
	HDU-L155	D/5	04-06
	HDU-L156	D/6	04-07
	HDU-L157	D/7	04-08
	HDU-L158	D/8	04-09
	HDU-L159	D/9	04-0A
	HDU-L15A	D/A	04-0B
	HDU-L15B	D/B	04-0C
HDU-L16	HDU-L160	E/0	04-01
	HDU-L161	E/1	04-02
	HDU-L162	E/2	04-03
	HDU-L163	E/3	04-04
	HDU-L164	E/4	04-05
	HDU-L165	E/5	04-06
	HDU-L166	E/6	04-07
	HDU-L167	E/7	04-08
	HDU-L168	E/8	04-09
	HDU-L169	E/9	04-0A
	HDU-L16A	E/A	04-0B
	HDU-L16B	E/B	04-0C
HDU-L17	HDU-L170	F/0	04-01
	HDU-L171	F/1	04-02
	HDU-L172	F/2	04-03
	HDU-L173	F/3	04-04
	HDU-L174	F/4	04-05
	HDU-L175	F/5	04-06
	HDU-L176	F/6	04-07
	HDU-L177	F/7	04-08
	HDU-L178	F/8	04-09
	HDU-L179	F/9	04-0A
	HDU-L17A	F/A	04-0B
	HDU-L17B	F/B	04-0C

Relationship between disk drive number and parity group number (9/12)

HDD BOX number	Disk drive number	C# / R#	Parity group number
HDU-L20	HDU-L200	8/C	07-01
	HDU-L201	8/D	07-02
	HDU-L202	8/E	07-03
	HDU-L203	8/F	07-04
	HDU-L204	8/10	07-05
	HDU-L205	8/11	07-06
	HDU-L206	8/12	07-07
	HDU-L207	8/13	07-08
	HDU-L208	8/14	07-09
	HDU-L209	8/15	07-0A
	HDU-L20A	8/16	07-0B
	HDU-L20B	8/17	07-0C
HDU-L21	HDU-L210	9/C	07-01
	HDU-L211	9/D	07-02
	HDU-L212	9/E	07-03
	HDU-L213	9/F	07-04
	HDU-L214	9/10	07-05
	HDU-L215	9/11	07-06
	HDU-L216	9/12	07-07
	HDU-L217	9/13	07-08
	HDU-L218	9/14	07-09
	HDU-L219	9/15	07-0A
	HDU-L21A	9/16	07-0B
	HDU-L21B	9/17	07-0C
HDU-L22	HDU-L220	A/C	07-01
	HDU-L221	A/D	07-02
	HDU-L222	A/E	07-03
	HDU-L223	A/F	07-04
	HDU-L224	A/10	07-05
	HDU-L225	A/11	07-06
	HDU-L226	A/12	07-07
	HDU-L227	A/13	07-08
	HDU-L228	A/14	07-09
	HDU-L229	A/15	07-0A
	HDU-L22A	A/16	07-0B
	HDU-L22B	A/17	07-0C
HDU-L23	HDU-L230	B/C	07-01
	HDU-L231	B/D	07-02
	HDU-L232	B/E	07-03
	HDU-L233	B/F	07-04
	HDU-L234	B/10	07-05
	HDU-L235	B/11	07-06
	HDU-L236	B/12	07-07
	HDU-L237	B/13	07-08
	HDU-L238	B/14	07-09
	HDU-L239	B/15	07-0A
	HDU-L23A	B/16	07-0B
	HDU-L23B	B/17	07-0C

Relationship between disk drive number and parity group number (10/12)

HDD BOX number	Disk drive number	C# / R#	Parity group number
HDU-L24	HDU-L240	C/C	08-01
	HDU-L241	C/D	08-02
	HDU-L242	C/E	08-03
	HDU-L243	C/F	08-04
	HDU-L244	C/10	08-05
	HDU-L245	C/11	08-06
	HDU-L246	C/12	08-07
	HDU-L247	C/13	08-08
	HDU-L248	C/14	08-09
	HDU-L249	C/15	08-0A
	HDU-L24A	C/16	08-0B
	HDU-L24B	C/17	08-0C
HDU-L25	HDU-L250	D/C	08-01
	HDU-L251	D/D	08-02
	HDU-L252	D/E	08-03
	HDU-L253	D/F	08-04
	HDU-L254	D/10	08-05
	HDU-L255	D/11	08-06
	HDU-L256	D/12	08-07
	HDU-L257	D/13	08-08
	HDU-L258	D/14	08-09
	HDU-L259	D/15	08-0A
	HDU-L25A	D/16	08-0B
	HDU-L25B	D/17	08-0C
HDU-L26	HDU-L260	E/C	08-01
	HDU-L261	E/D	08-02
	HDU-L262	E/E	08-03
	HDU-L263	E/F	08-04
	HDU-L264	E/10	08-05
	HDU-L265	E/11	08-06
	HDU-L266	E/12	08-07
	HDU-L267	E/13	08-08
	HDU-L268	E/14	08-09
	HDU-L269	E/15	08-0A
	HDU-L26A	E/16	08-0B
	HDU-L26B	E/17	08-0C
HDU-L27	HDU-L270	F/C	08-01
	HDU-L271	F/D	08-02
	HDU-L272	F/E	08-03
	HDU-L273	F/F	08-04
	HDU-L274	F/10	08-05
	HDU-L275	F/11	08-06
	HDU-L276	F/12	08-07
	HDU-L277	F/13	08-08
	HDU-L278	F/14	08-09
	HDU-L279	F/15	08-0A
	HDU-L27A	F/16	08-0B
	HDU-L27B	F/17	08-0C

Relationship between disk drive number and parity group number (11/12)

HDD BOX number	Disk drive number	C# / R#	Parity group number
HDU-L30	HDU-L300	8/18	0B-01
	HDU-L301	8/19	0B-02
	HDU-L302	8/1A	0B-03
	HDU-L303	8/1B	0B-04
	HDU-L304	8/1C	0B-05
	HDU-L305	8/1D	0B-06
	HDU-L306	8/1E	0B-07
	HDU-L307	8/1F	0B-08
HDU-L31	HDU-L310	9/18	0B-01
	HDU-L311	9/19	0B-02
	HDU-L312	9/1A	0B-03
	HDU-L313	9/1B	0B-04
	HDU-L314	9/1C	0B-05
	HDU-L315	9/1D	0B-06
	HDU-L316	9/1E	0B-07
	HDU-L317	9/1F	0B-08
HDU-L32	HDU-L320	A/18	0B-01
	HDU-L321	A/19	0B-02
	HDU-L322	A/1A	0B-03
	HDU-L323	A/1B	0B-04
	HDU-L324	A/1C	0B-05
	HDU-L325	A/1D	0B-06
	HDU-L326	A/1E	0B-07
	HDU-L327	A/1F	0B-08
HDU-L33	HDU-L330	B/18	0B-01
	HDU-L331	B/19	0B-02
	HDU-L332	B/1A	0B-03
	HDU-L333	B/1B	0B-04
	HDU-L334	B/1C	0B-05
	HDU-L335	B/1D	0B-06
	HDU-L336	B/1E	0B-07
	HDU-L337	B/1F	0B-08

Relationship between disk drive number and parity group number (12/12)

HDD BOX number	Disk drive number	C# / R#	Parity group number
HDU-L34	HDU-L340	C/18	0C-01
	HDU-L341	C/19	0C-02
	HDU-L342	C/1A	0C-03
	HDU-L343	C/1B	0C-04
	HDU-L344	C/1C	0C-05
	HDU-L345	C/1D	0C-06
	HDU-L346	C/1E	0C-07
	HDU-L347	C/1F	0C-08
HDU-L35	HDU-L350	D/18	0C-01
	HDU-L351	D/19	0C-02
	HDU-L352	D/1A	0C-03
	HDU-L353	D/1B	0C-04
	HDU-L354	D/1C	0C-05
	HDU-L355	D/1D	0C-06
	HDU-L356	D/1E	0C-07
	HDU-L357	D/1F	0C-08
HDU-L36	HDU-L360	E/18	0C-01
	HDU-L361	E/19	0C-02
	HDU-L362	E/1A	0C-03
	HDU-L363	E/1B	0C-04
	HDU-L364	E/1C	0C-05
	HDU-L365	E/1D	0C-06
	HDU-L366	E/1E	0C-07
	HDU-L367	E/1F	0C-08
HDU-L37	HDU-L370	F/18	0C-01
	HDU-L371	F/19	0C-02
	HDU-L372	F/1A	0C-03
	HDU-L373	F/1B	0C-04
	HDU-L374	F/1C	0C-05
	HDU-L375	F/1D	0C-06
	HDU-L376	F/1E	0C-07
	HDU-L377	F/1F	0C-08

5.2 Physical - Logical Device Matrixes for Single Cabinet Model

Relationship between disk drive number and parity group number

HDD BOX number	Disk drive number	C# / R#	Parity group number
HDU-0	HDU-00	0/0	01-01
	HDU-01	0/1	01-02
	HDU-02	0/2	01-03
	HDU-03	0/3	01-04
	HDU-04	0/4	01-05
	HDU-05	0/5	01-06
	HDU-06	0/6	01-07
	HDU-07	0/7	01-08
	HDU-08	0/8	01-09
	HDU-09	0/9	01-0A
	HDU-0A	0/A	01-0B
	HDU-0B	0/B	Spare
HDU-1	HDU-10	1/0	01-01
	HDU-11	1/1	01-02
	HDU-12	1/2	01-03
	HDU-13	1/3	01-04
	HDU-14	1/4	01-05
	HDU-15	1/5	01-06
	HDU-16	1/6	01-07
	HDU-17	1/7	01-08
	HDU-18	1/8	01-09
	HDU-19	1/9	01-0A
	HDU-1A	1/A	01-0B
	HDU-1B	1/B	Spare
HDU-2	HDU-20	2/0	01-01
	HDU-21	2/1	01-02
	HDU-22	2/2	01-03
	HDU-23	2/3	01-04
	HDU-24	2/4	01-05
	HDU-25	2/5	01-06
	HDU-26	2/6	01-07
	HDU-27	2/7	01-08
	HDU-28	2/8	01-09
	HDU-29	2/9	01-0A
	HDU-2A	2/A	01-0B
	HDU-2B	2/B	Spare
HDU-3	HDU-30	3/0	01-01
	HDU-31	3/1	01-02
	HDU-32	3/2	01-03
	HDU-33	3/3	01-04
	HDU-34	3/4	01-05
	HDU-35	3/5	01-06
	HDU-36	3/6	01-07
	HDU-37	3/7	01-08
	HDU-38	3/8	01-09
	HDU-39	3/9	01-0A
	HDU-3A	3/A	01-0B
	HDU-3B	3/B	Spare

5.3 Commands

This subsystem commands are classified into the following eight categories:

(1) Read commands
The read commands transfer data read from devices to channels.

(2) Write commands
The write commands transfer data from channels to devices.

(3) Search commands
The search commands follow a control command and logically search for target data.

(4) Control commands
The control commands include the seek commands that position the head at the specified cylinder and head positions. The SET SECTOR command that executes latency time processing, the LOCATE RECORD command that specifies the operation of the ECKD command, the SET FILE MASK commands that defines the permissible ranges for the write and seek operations, and the DEFINE EXTENT command that defines the permissible ranges for the write and seek operations and that defines the cache access mode.

(5) Sense commands
The sense commands transfer sense bytes and device specifications.

(6) Path control commands
The path control commands enable and disable the exclusive control of devices.

(7) TEST I/O command
The TEST I/O command transfers the specified device and its path state to a given channel in the form of DSBs.

(8) Subsystem commands
The subsystem commands include those which define cache control information in the DKCs and those which transfer cache-related information to channels.

Table 5.3 Command Summary (1/3)

Command Name			Command Code	
			Single Track	Multitrack
Read commands	READ INITIAL PROGRAM LOAD	(RD IPL)	02	
	READ HOME ADDRESS	(RD HA)	1A	9A
	READ RECORD ZERO	(RD R0)	16	96
	READ COUNT,KEY,DATA	(RD CKD)	1E	8E
	READ KEY,DATA	(RD KD)	0E	86
	READ DATA	(RD D)	06	92
	READ COUNT	(RD C)	12	
	READ MULTIPLE COUNT,KEY AND DATA	(RD MCKD)	5E	
	READ TRACK	(RD TRK)	DE	
	READ SPECIAL HOME ADDRESS	(RD SP HA)	0A	
WRITE commands	WRITE HOME ADDRESS	(WR HA)	19	
	WRITE RECORD ZORO	(WR R0)	15	
	WRITE COUNT,KEY,DATA	(WR CKD)	1D	
	WRITE COUNT,KEY,DATA NEXT TRACK	(WR CKD NT)	9D	
	ERASE	(ERS)	11	
	WRITE KEY AND DATA	(WR KD)	0D	
	WRITE UPDATE KEY AND DATA	(WR UP KD)	8D	
	WRITE DATA	(WR D)	05	
	WRITE UPDATE DATA	(WR UP D)	85	
	WRITE SPECIAL HOME ADDRESS	(WR SP HA)	09	
SEARCH commands	SEARCH HOME ADDRESS	(SCH HA EQ)	39	B9
	SEARCH ID EQUAL	(SCH ID EQ)	31	B1
	SEARCH ID HIGH	(SCH ID HI)	51	D1
	SEARCH ID HIGH OR EQUAL	(SCH ID HE)	71	F1
	SEARCH KEY EQUAL	(SCH KEY EQ)	29	A9
	SEARCH KEY HIGH	(SCH KEY HI)	49	C9
	SEARCH KEY HIGH OR EQUAL	(SCH KEYD HE)	69	E9

Table 5.3 Command Summary(2/3)

Command Name			Command Code	
			Single Track	Multitrack
CONTROL commands	DEFINE EXTENT	(DEF EXT)	63	
	LOCATE RECORD	(LOCATE)	47	
	LOCATE RECORD EXTENDED	(LOCATE EXT)	4B	
	SEEK	(SK)	07	
	SEEK CYLINDER	(SK CYL)	0B	
	SEEK HEAD	(SK HD)	1B	
	RECALIBRATE	(RECAL)	13	
	SET SECTOR	(SET SECT)	23	
	SET FILE MASK	(SET FM)	1F	
	READ SECTOR	(RD SECT)	22	
	SPACE COUNT	(SPC)	0F	
	NO OPERATION	(NOP)	03	
	RESTORE	(REST)	17	
	DIAGNOSTIC CONTROL	(DIAG CTL)	F3	
SENSE commands	SENSE	(SNS)	04	
	READ AND RESET BUFFERED LOG	(RRBL)	A4	
	SENSE IDENTIFICATION	(SNS ID)	E4	
	READ DEVICE CHARACTERISTICS	(RD CHR)	64	
	DIAGNOSTIC SENSE/READ	(DIAG SNS/RD)	C4	
PATH CONTROL commands	DEVICE RESERVE	(RSV)	B4	
	DEVICE RELEASE	(RLS)	94	
	UNCONDITIONAL RESERVE	(UNCON RSV)	14	
	SET PATH GROUP ID	(SET PI)	AF	
	SENSE SET PATH GROUP ID	(SNS PI)	34	
	SUSPEND MULTIPATH RECONNECTION	(SUSP MPR)	5B	
	RESET ALLEGIANCE	(RST ALG)	44	
TST I/O	TEST I/O	(TIO)	00	
TIC	TRANSFER IN CHANNEL	(TIC)	X8	

Table.5.3 Command Summary(3/3)

Command Name			Command Code	
			Single Track	Multitrack
SUBSYSTEM commands	SET SUBSYSTEM MODE	(SET SUB MD)	87	
	PERFORM SUBSYSTEM FUNCTION	(PERF SUB FUNC)	27	
	READ SUBSYSTEM DATA	(RD SUB DATA)	3E	
	SENSE SUBSYSTEM STATUS	(SNS SUB STS)	54	
	READ MESSAGE ID	(RD MSG IDL)	4E	

Note 1: Command Reject, format 0, and message 1 are issued for commands that are not listed in this table.

Note 2: TEST I/O is a CPU instruction and cannot be specified directly. However, it appears as a command to the interface.

Note 3: TIC is a type of command but runs only on a channel. It can never be visible to the interface.

5.4 Comparison of pair status on SVP, Remote Console, Raid Manager

Table.5.4 Comparison of pair status on SVP, Remote Console, Raid Manager

NO	Event	Status on Raid Manager	Status on SVP, Remote Console
1	Simplex Volume	P-VOL: SMPL S-VOL: SMPL	P-VOL: SMPL S-VOL: SMPL
2	Copying LUSE Volume Partly completed (SYNC only)	P-VOL: PDUB S-VOL: PDUB	P-VOL: PDUB S-VOL: PDUB
3	Copying Volume	P-VOL: COPY S-VOL: COPY	P-VOL: COPY S-VOL: COPY
4	Pair volume	P-VOL: PAIR S-VOL: PAIR	P-VOL: PAIR S-VOL: PAIR
5	Pairsplit operation to P-VOL	P-VOL: PSUS S-VOL: SSUS	P-VOL: PSUS (S-VOL by operator) S-VOL: PSUS (S-VOL by operator)
6	Pairsplit operation to S-VOL	P-VOL: PSUS S-VOL: PSUS	P-VOL: PSUS (S-VOL by operator) S-VOL: PSUS (S-VOL by operator)
7	Pairsplit -P operation (*1) (P-VOL failure, SYNC only)	P-VOL: PSUS S-VOL: SSUS	P-VOL: PSUS (P-VOL by operator) S-VOL: PSUS (by MCU)
8	Pairsplit -R operation (*1)	P-VOL: PSUS S-VOL: SMPL	P-VOL: PSUS(Delete pair to RCU) S-VOL: SMPL
9	P-VOL Suspend (failure)	P-VOL: PSUE S-VOL: SSUS	P-VOL: PSUE (S-VOL failure) S-VOL: PSUE (S-VOL failure)
10	S-VOL Suspend (failure)	P-VOL: PSUE S-VOL: PSUE	P-VOL: PSUE (S-VOL failure) S-VOL: PSUE (S-VOL failure)
11	PS ON failure	P-VOL: PSUE S-VOL: ------	P-VOL: PSUE (MCU IMPL) S-VOL: ------
12	Copy failure (P-VOL failure)	P-VOL: PSUE S-VOL:SSUS	P-VOL: PSUE (Initial copy failed) S-VOL: PSUE (Initial copy failed)
13	Copy failure (S-VOL failure)	P-VOL: PSUE S-VOL:PSUE	P-VOL: PSUE (Initial copy failed) S-VOL: PSUE (Initial copy failed)
14	Suspending volume (ASYNC only)	P-VOL: COPY or PAIR or PSUE S-VOL: COPY or PAIR or PSUE	P-VOL: Suspending S-VOL: Suspending
15	Deleting volume (ASYNC only)	P-VOL: COPY or PAIR or SUS S-VOL: COPY or PAIR or SUS	P-VOL: Deleting S-VOL: Deleting
16	RCU accepted the notification of MCU's P/S-OFF	P-VOL: ------ S-VOL:SSUS	P-VOL: ------ S-VOL: PSUE (MCU P/S OFF)
17	Sidefile overload (under margin, ASYNC only)	P-VOL: PFUL S-VOL: PAIR	P-VOL: PAIR S-VOL: PAIR
18	Sidefile overload Suspend (over margin, ASYNC only)	P-VOL: PFUS S-VOL: PFUS	P-VOL: PSUS (Sidefile Overflow) S-VOL: PSUS (Sidefile Overflow)

*1 Operation on Raid Manager