Sun™ StorEdge™ A7000 Remote Dual Copy System Administrator's Guide

Campus-Wide

THE NETWORK IS THE COMPUTER™

Sun Microsystems, Inc.
901 San Antonio Road
Palo Alto, CA 94303-4900 USA
650 960-1300 Fax 650 969-9131

Part No. 805-4600-11
December 1998, Revision A

Send comments about this document to: `docfeedback@sun.com`

Contents

Figures

Tables

Preface

The *Sun StorEdge A7000 Remote Dual Copy System Administrator's Guide (Campus-Wide)* describes the Remote Dual Copy (RDC) software and its use.

What is Remote Dual Copy?

Campus-Wide RDC lets you replicate disks between different, physically-separate StorEdge A7000 servers in real time. RDC is conceptually similar to the local disk mirroring scheme of RAID 1, but it performs its mirroring operations over longer distances using peer storage processors. RDC is intended to be used as part of Disaster Recovery and Business Continuance plans, which often rely on redundant storage of critical information across physically separate sites.

Campus-Wide Remote Dual Copy lets you place each StorEdge A7000 server up to 3 kilometers (approximately 2 miles) apart using the Fiber-Optic REFLECTIVE MEMORY™ Systems (FORMS™).

Note – The actual distances supported depend on optical losses in the fiber plant between the two sites; you may also increase the distance with fiber optic extenders. Refer to the associated hardware specifications for more detailed physical considerations.

FIGURE P-1 shows a typical Campus-Wide RDC configuration connecting primary and secondary data centers.

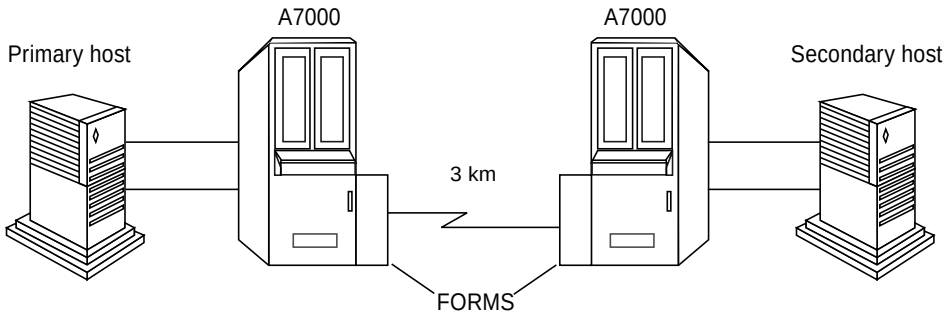

FIGURE P-1 Campus-Wide RDC Configuration Diagram

How This Book Is Organized

Chapter 1 describes general information about RDC and describes RDC terminology.

Chapter 2 describes RDC configuration files and commands.

Chapter 3 describes RDC operations including synchronization, administrative utilities, rehearsals, and mirrored partitions.

The *Glossary* contains definitions of terms used in this document.

Using UNIX Commands

This document may not contain information on basic UNIX® commands and procedures such as shutting down the system, booting the system, and configuring devices.

Refer to the software documentation that you received with your system.

Documentation Conventions

TABLE P-1 Documentation Conventions

Typeface or Symbol	Meaning	Examples
`AaBbCc123`	The names of commands, files, and directories; on-screen computer output.	Edit your `.login` file. Use `ls -a` to list all files. `% You have mail.`
`AaBbCc123`	What you type, when contrasted with on-screen computer output.	`% su` `Password:`
AaBbCc123	Book titles, new words or terms, words to be emphasized. Command-line variable; replace with a real name or value.	Read Chapter 6 in the *User's Guide.* These are called *class* options. You *must* be `root` to do this. To delete a file, type `rm` *filename.*
[ ]	In syntax, brackets indicate that an argument is optional.	**`sd_stats`** [**`-d`** *sec*] [**`-r`** *n*[:*n*][,*n*]...] [**`-z`**]

Shell Prompts

TABLE P-2 Shell Prompts

Shell	Prompt
C shell	*machine_name*%
C shell superuser	*machine_name*#
Bourne shell and Korn shell	$
Bourne shell and Korn shell superuser	#

Related Documentation

TABLE P-3 Related Documentation

Application	Title	Part Number
Release	*Sun StorEdge A7000 Remote Dual Copy Release Notes (Campus-Wide)*	online
Reference	*Direct Access Storage Device (DASD) Manager User's Guide*	805-4884-xx
Reference	*SCSI Target Emulation Release Note*	online
Reference	*Simulation of Count-Key-Data (SIMCKD) Release Notes*	online
Reference	*Sun StorEdge A7000 DataShare Facility Systems Administrator's and User's Guide*	805-4366-xx
Reference	`mpadmin(1M)` man page `rdc_copy_primary_to_secondary(1RDC)` man page `rdc_restore_primary_from_secondary(1RDC)` man page `rdc_refresh_primary_from_secondary(1RDC)` man page `rdc_suspend_mirroring(1RDC)` man page `rdc_resume_mirroring(1RDC)` man page `rdc_disable_mirroring(1RDC)` man page `sd_stats(1SCM)` man page	online
Reference	*Sun StorEdge A7000 Service Manual*	805-6489-xx

Sun Documentation on the Web

The `docs.sun.com`SM web site enables you to access Sun technical documentation on the Web. You can browse the `docs.sun.com` archive or search for a specific book title or subject at:

```
http://docs.sun.com
```

Sun Welcomes Your Comments

We are interested in improving our documentation and welcome your comments and suggestions. You can email your comments to us at:

`docfeedback@sun.com`

Please include the part number of your document in the subject line of your email.

CHAPTER 1

General Information About RDC

This chapter describes Remote Dual Copy and its capabilities:

Overview

Campus-Wide Remote Dual Copy (RDC) features include:

- Real-time disk mirroring between physically separate primary and secondary sites
- 100 Mbyte per second intersite links using Fiber-Optic REFLECTIVE MEMORY Systems (FORMS™)
- Mainframe and SCSI-attached host support
- Minimal impact on primary subsystem performance
- No host software dependency
- Granularly configurable per logical volume (noncritical volumes may be excluded)
- Mutual backup (also known as a bilateral relationship)
- Continued remote data access despite local disk failure (depending on how RDC is configured)
- Suspend and resume control of remote mirroring on a logical volume basis
- Active logging when RDC is suspended or interrupted
- Optimized resynchronization following disk, link, subsystem, and storage platform outages
- Disaster rehearsal rollback provisions
- Locally mirrored (RAID 1), RAID 5 protected, striped (RAID 0), and linear source and target volumes

RDC Terminology

This section describes important RDC terms that are used throughout this manual. The Glossary contains more general terminology.

Term	Description
Asynchronous mirroring	In contrast to synchronous mirroring, asynchronous mirroring confirms to the originating host that the primary I/O transaction is complete before updating the remote image. Deferring the secondary copy removes the long distance propagation delays from the I/O response time (there is uncertainty, as the mirroring operation is not confirmed to the originating application in real time). Asynchronous mirroring is used when the distance between primary and secondary sites may introduce prohibitive latency times to the synchronous operations.
Full synchronization	Complete disk-to-remote-disk copy, which is the most time-consuming of the synchronization operations. In most cases, a secondary disk is synchronized from its source primary disk. However, restoration of a failed primary disk may require reverse synchronization, using the surviving remote mirror as the source.
Logging	Also known as RDC logging, this is a method for tracking disk updates that have not been remotely copied while the remote service is interrupted or impaired. The blocks that no longer match their remote pairs are identified for each source disk. RDC uses this log to re-establish a remote mirror through an optimized update synchronization rather than a complete disk-to-disk copy.
Mutual backup	Each StorEdge A7000 can concurrently transmit and receive remote dual copies to and from its remote counterpart. Each subsystem contains primary disks in an RDC pair that are accessible by local hosts, as well as remote mirrors secondary to remote hosts. Mutual backup may be used where critical applications and storage are split across sites, and both sites require remote redundant copies.
Primary	Subsystem or volume on which the host application is principally dependent.
Rollback synchronization	Operation used during recovery rehearsals. RDC logging keeps track of test updates applied to the secondary subsystem during the rehearsal. When the primary is restored, the test updates are overwritten with the blocks from the primary image, restoring matching remote pairs.

Term	Description
Rolling disaster	A collection of damaging events spread over a period of hours that may potentially impair multiple components in an RDC configuration.
Scoreboard	A special log that tracks hits to a disk, rather than a running log of each I/O event.
Secondary	Remote counterpart of the primary where data copies are written to and read from. Remote copies are transmitted without host intervention between peer A7000 servers. An A7000 server may act as primary storage for some volumes and secondary (remote) storage for others.
Synchronization	The process of establishing an identical copy of a source disk onto a target disk as a precondition to RDC mirroring.
Synchronous mirroring	Synchronous mirroring is limited to short distances (tens of kilometers) because of the detrimental effect of propagation delay on I/O response times.
Takeover rehearsal	Simulation of a switchover to the secondary server after a staged failure of the primary server.
Update synchronization	Copies only those disk blocks identified by RDC logging, reducing the time to restore remotely mirrored pairs.

Remote Dual Copy Basics

Remote Dual Copy is a natural extension of multicast stable storage for the A7000. These sections describe the differences between local and RDC storage behavior.

FIGURE 1-1 to FIGURE 1-5 show simplified data flows for each operation using one channel path on one of the A7000 subsystems. In normal operations, adjacent subsystems actively service host requests concurrently over multiple channels. These interactions have been omitted from the figures for clarity.

I/O Operations in a Local Configuration Without RDC

See FIGURE 1-1; TABLE 1-1 is a legend for the figures in this manual.

In non-RDC A7000 configurations, host disk update write data (D) is placed in the data cache of the directly-connected subsystem (dsp1). The write data is copied to the adjacent subsystem (dsp2) write-cache *safe* area (SAFE) before the host is notified of I/O completion. The StorEdge A7000 power conditioning unit keeps both cache memories and the A7000 in a nonvolatile state.

The safe copy in subsystem dsp2 protects against data loss in a primary subsystem failure. Under normal circumstances, the safe copy is discarded when the write has been permanently destaged to disk by dsp1. The safe copy is used only when the primary subsystem is unable to destage its data cache. Then the adjacent subsystem takes responsibility for updating the disk by destaging the safe copy over an alternate device path. FIGURE 1-2 shows that process.

TABLE 1-1 Legend for Figures in This Manual

Legend Item	Definition
— — →	Control/data flow
1, 2,...N	Process sequence
CA	Channel adapter like block multiplexer (BMC), SCSI target (STE) or ESCON
Cache	cache memory implemented by the StorEdge A7000 cache memory cards
D	Write or read data
DA	SCSI disk controller/adapter
dspN	Subsystem identifier: dsp1, dsp2...dspN
FORMS	Fiber-optic REFLECTIVE MEMORY SYSTEM
MEMORY CHANNEL	High-speed reflected cache interconnect among subsystems (dsp1 to dsp2, or dsp3 to dsp4)
N*	Conditional step
NR	Remote portion of a step in the process sequence
RDC	Special RDC space in cache card

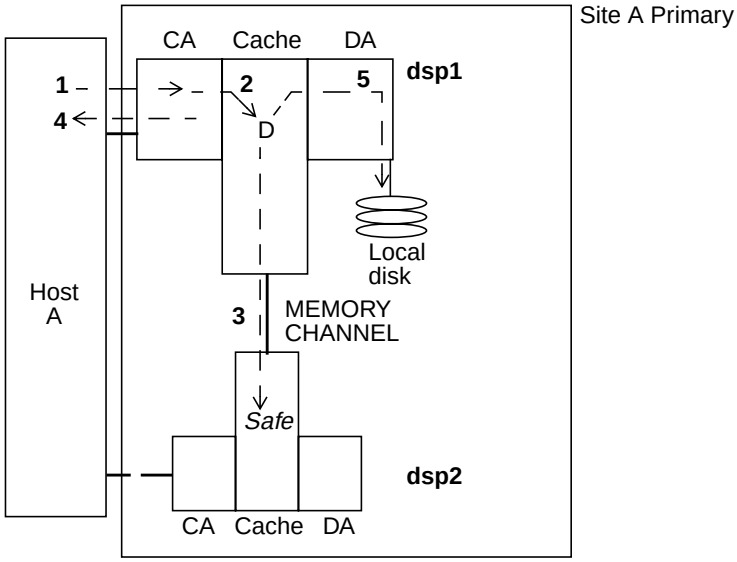

FIGURE 1-1 Local Write

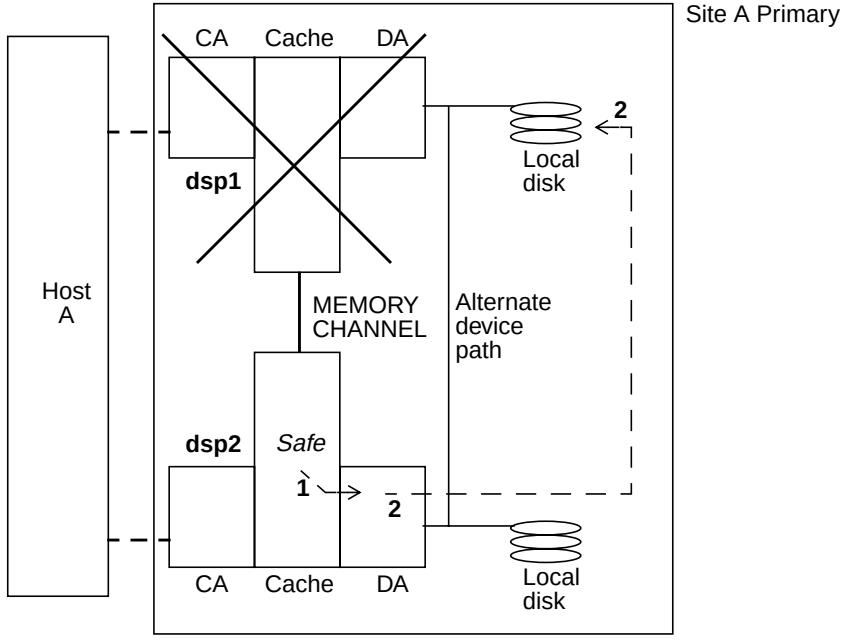

FIGURE 1-2 Destaging Safe Copy

RDC extends the pairing between the local primary and adjacent subsystems by nonintrusively maintaining a second safe copy at a third subsystem some distance away (see FIGURE 1-3). The multicast properties of A7000 MEMORY CHANNEL hardware and FORMS make this extension possible.

This second safe copy is stored in a special RDC cache area on the remote subsystem. When the primary subsystem destages updates to disk, RDC requests that the remote subsystem do the same. When both sites have confirmed destaging, the data cache block can be deallocated and the safe copies discarded.

Remote Volume Access Considerations

As with the local case, the second RDC safe copy is not intended for direct use by hosts but is a standby duplicate if the primary subsystem fails. If this failure occurs, the remote subsystem destages its latest remote dual copies to disk to record the last known state of the primary. The secondary hosts can access the remote mirrors to begin processing. In practice, use additional application-level recovery procedures, including checkpoint validation, to ensure that the surviving data is consistent. Any anomalies encountered must be reconciled before restoration is complete.

To guarantee a consistent view of the data from a secondary host, suspend RDC updates from the primary subsystem and completely destage the RDC cache to secondary disks. See Chapter 2 for RDC commands and Chapter 3 for detailed procedures.

While remote mirroring is underway, declare RDC-managed volumes at the secondary subsystem offline to the secondary hosts. You can declare these secondary volumes online for local access from secondary hosts only after the remote updates stop and the RDC cache has been destaged to secondary disks.

MVS Considerations

The IBM MVS operating system enforces unique volume serial numbers (VOLSERs) under a given system image. RDC mirroring results in identical logical copies; the primary and secondary volumes have the same VOLSER. Therefore, in configurations where the same MVS image is connected to the primary and secondary subsystems, declare only one of these volumes online. Under normal operations, only the primary volumes are declared online.

If the primary volumes become unavailable, the secondary volumes can be declared online using the guidelines for accessing remote volumes.

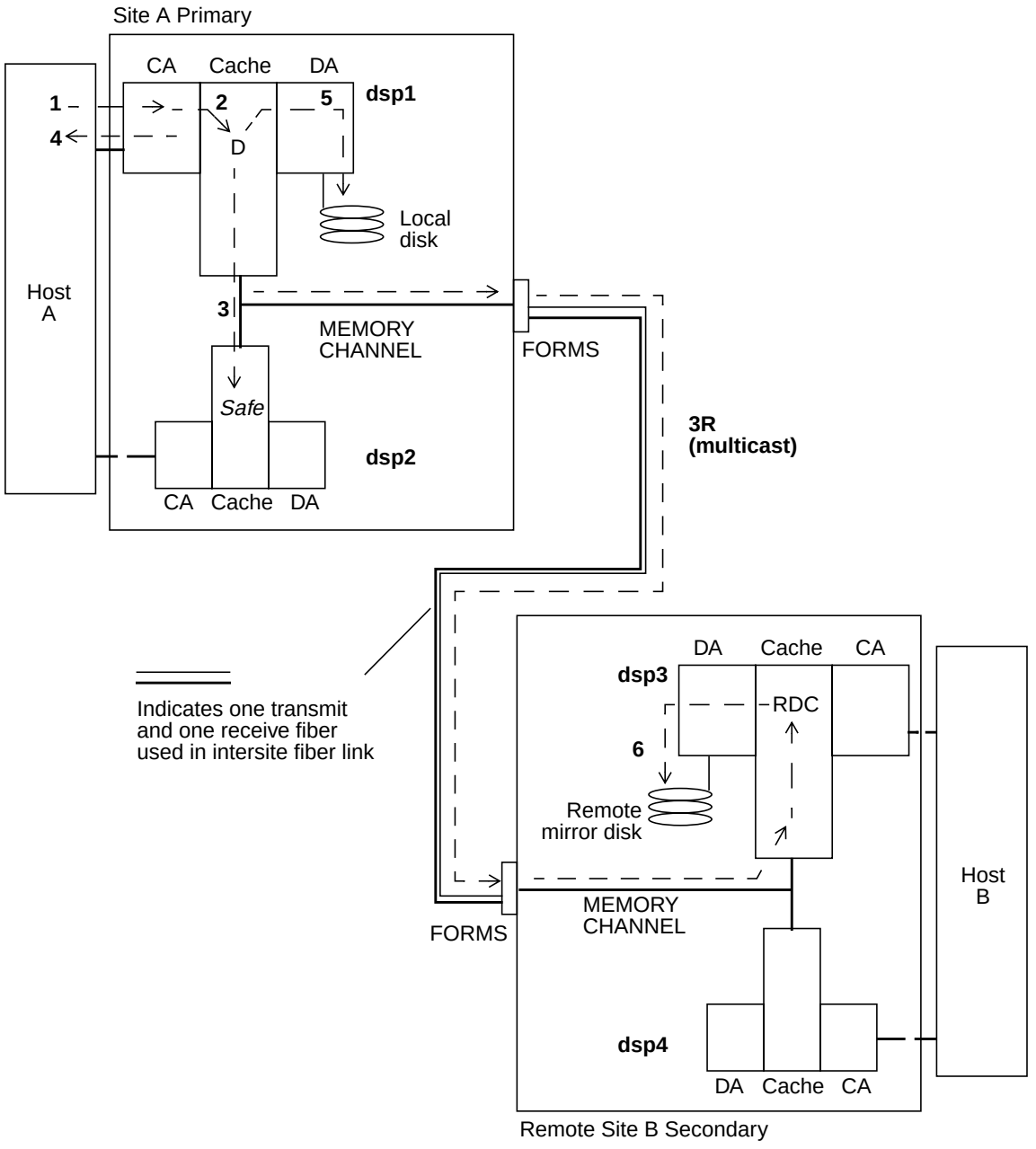

FIGURE 1-3 Real Time Remote Write

What Happens When Pairs Do Not Match?

Physical separation provides highly-available storage configurations, but requires special attention to logistics. Intersite link outages, component failures, and other situations temporarily cause the primary and secondary images to be unequal, or out of sync. RDC uses several logging and resynchronization techniques to re-establish mirror copies.

Full synchronization (that is, a complete volume-to-remote volume copy) is used when the remote volume specified for redundancy is newly formatted or its contents are unknown. Update resynchronization provides a streamlined way to make remote pairs match if their changes have been logged earlier by RDC.

Rollback resynchronization is a form of update synchronization used for disaster recovery rehearsals and other applications where test updates must be reversed. Chapter 3 describes the mechanism to use in a given circumstance. When multiple volumes require synchronization, RDC performs several of these techniques concurrently.

A7000 Failover and Failback Considerations

The A7000 has a fault-resilient two-subsystem architecture optimized to increase the host data availability at the primary site if there is a subsystem failure. During local failover at the primary site, the surviving primary subsystem continues to remote-dual-copy its principal volumes. Volumes principally owned by the failed subsystem are no longer remotely dual copied.

Note – Some A7000 RDC configurations use host-based software to mirror volumes across primary subsystems dsp1 and dsp2. In this case, RDC in the surviving subsystem (for example, dsp2) continues to replicate its half of the mirror to dsp4, ensuring an up-to-date image is always available at the secondary site.

The surviving primary subsystem ensures that any cached writes not destaged by the failed subsystem are written to disk. As shown in FIGURE 1-4, the surviving subsystem (dsp2) destages its safe copy of dsp1 writes over the alternate device path to dsp1 volumes.

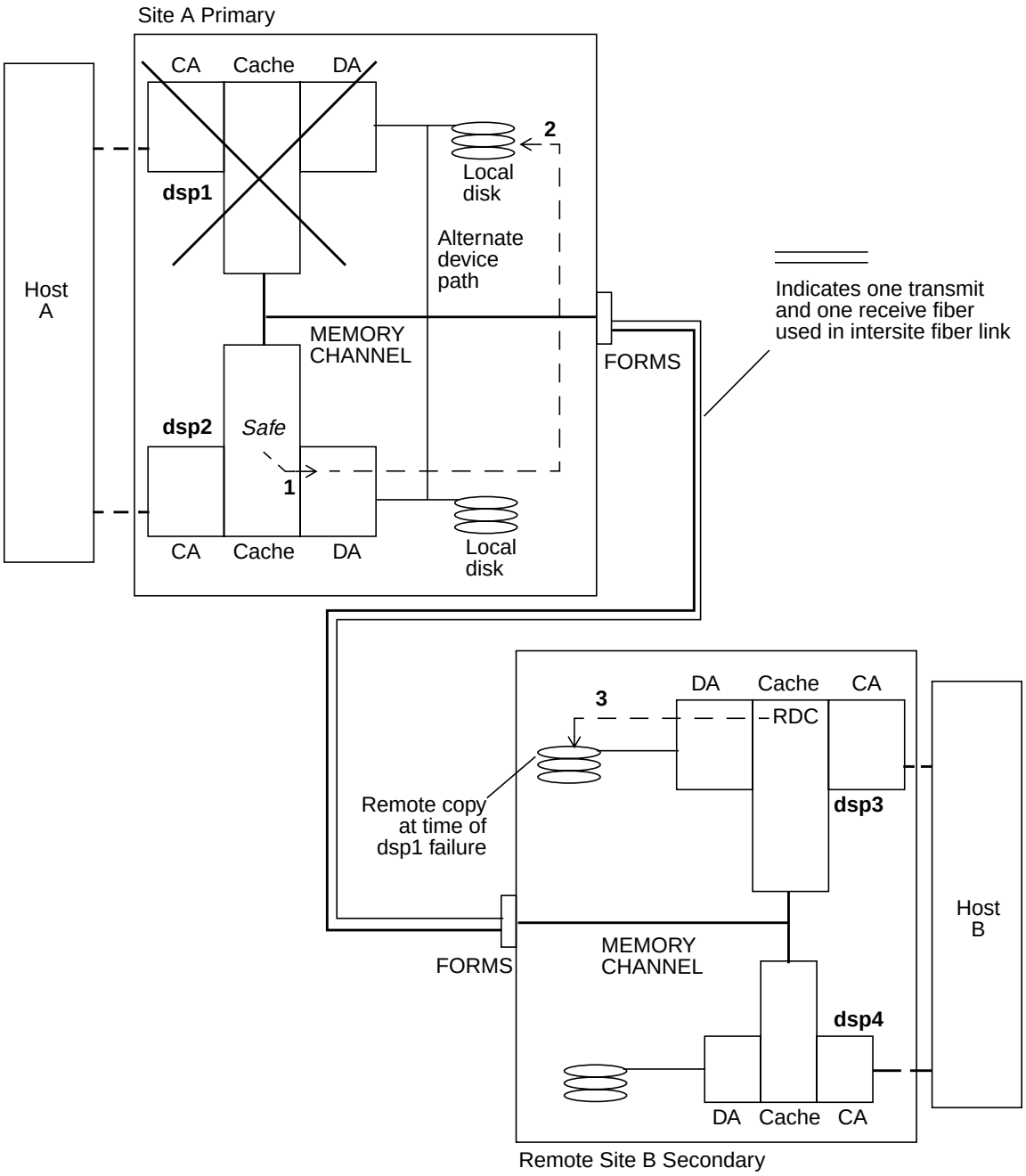

FIGURE 1-4 Destaging Safe Copy During Failover

Similarly, the remote companion subsystem dsp3 detects no response (the loss of *heartbeat*) from dsp1 and destages any previously received cached copies to its secondary volumes. RDC makes no further attempt to preserve a remote synchronization state across a failover/failback sequence. For example, RDC disregards any previous logging on the failed subsystem dsp1.

For mainframe volumes, channel paths may have been switched to the surviving subsystem dsp2. RDC is not aware of writes to the failed subsystem's volumes during failover (see FIGURE 1-5). Upon failback, fully synchronize all RDC volumes from the failed subsystem dsp1 to its remote companion dsp3. The full volume copies ensure that all writes to the failed subsystem volumes during failover are replicated at the remote site, and that the remotely mirrored relationship is restored.

Note – Start synchronization only after the failback configuration is fully operational. Full volume synchronization competes with host workload for read access to the primary devices. Perform synchronization during periods of reduced host activity.

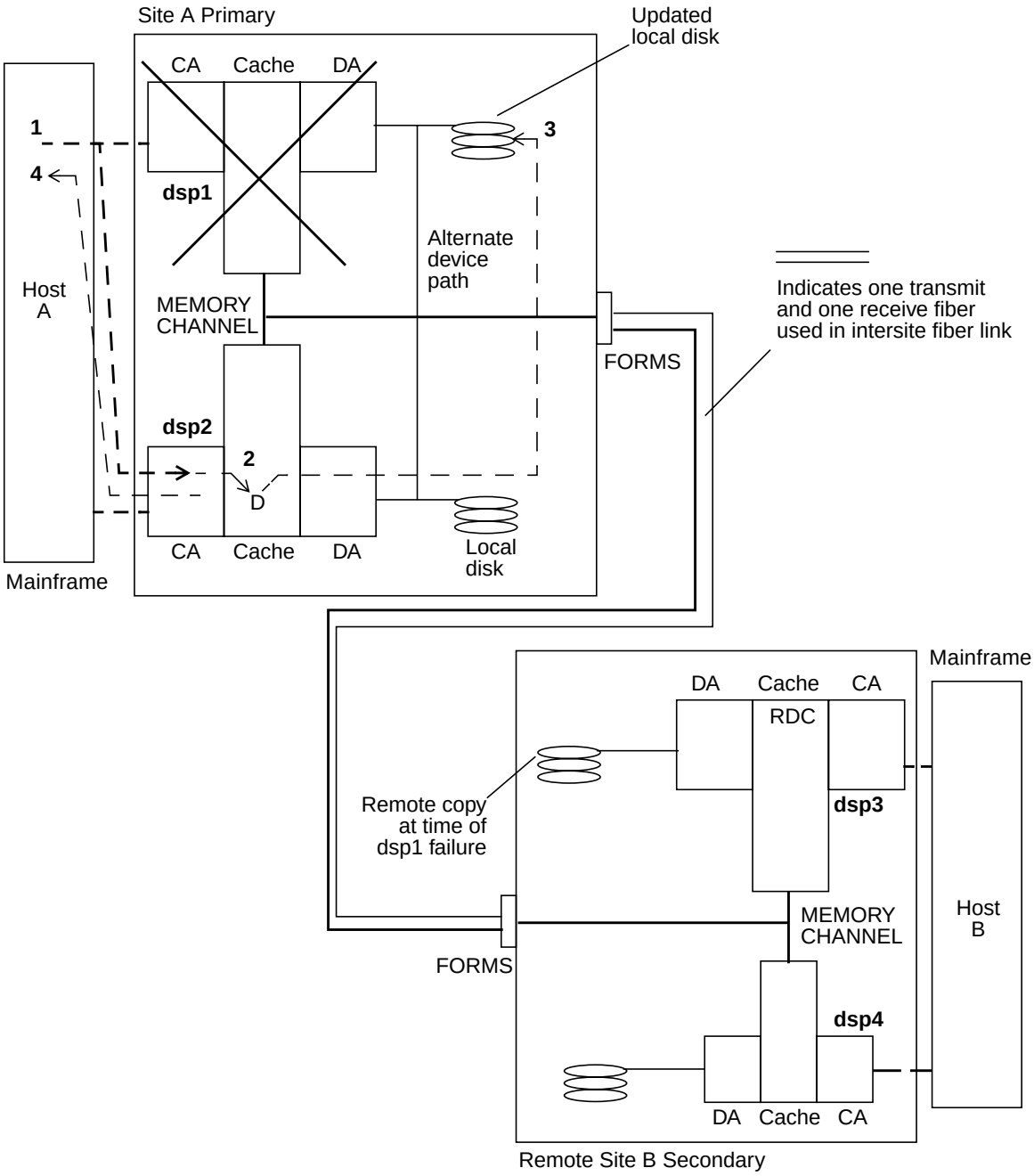

FIGURE 1-5 Mainframe Failover

CHAPTER 2

RDC Files and Commands

This chapter describes the RDC configuration files and commands available from the subsystem console command line. Chapter 3 describes using these commands for RDC operations.

This chapter describes the most-commonly used commands first, followed by the least-used commands.

Note – Issue all `rdc_*` commands from the primary subsystem.

Configuration Files

Configuration files contain information for the dual copy pairs. The RDC commands rely on two default configuration files: `/etc/rdc.cf` and `/etc/sd.cf`. You can create customized configuration files depending on your StorEdge A7000 connection and disaster recovery plans.

`/etc/rdc.cf` File

The default Remote Dual Copy (RDC) configuration file is `/etc/rdc.cf`. This file defines the RDC volume pairs on which RDC operates.

You can specify a configuration file other than the default `/etc/rdc.cf` file with the `-f` option on the RDC command line. For example, use a configuration file to specify multiple dual copy volume pairs with a single RDC command instead of entering a command for each pair.

The StorEdge A7000 supports two RDC configurations; `/etc/rdc.cf` can contain both configurations:

- Mirroring between volumes across separate A7000s; this configuration requires Fiber-Optic Reflective Memory (FORM) hardware and customer-supplied fiber connecting the A7000s
- Mirroring between volumes across adjacent subsystems (nodes) on the same A7000; this configuration uses the intersubsystem Memory Channel bus

To read the manual page for `/etc/rdc.cf` in a subsystem window, type:

```
# man rdc.cf
```

FIGURE 2-1 shows an example of this file. TABLE 2-1 explains each field.

Note – Volumes in one RDC volume pair cannot appear as volumes in another RDC volume pair.

```
#pri_node      pri_vol         sec_node   sec_vol         Reserved
dsp1        /dev/rdsk/0d4     dsp3      /dev/rdsk/0d5        0
dsp1        /dev/rdsk/1d4     dsp3      /dev/rdsk/1d5        0
dsp1        /dev/rdsk/2d4     dsp3      /dev/rdsk/2d5        0
```

FIGURE 2-1 Example `/etc/rdc.cf` File

TABLE 2-1 `/etc/rdc.cf` File Fields

Fields	Description
pri_node	Specifies the primary or source node (subsystem).
pri_vol	Specifies the primary or source volume.
sec_node	Specifies the secondary or target node (subsystem).
sec_vol	Specifies the secondary or target volume.
0	Reserved value. You must specify 0.

Mirroring Between Volumes Across Separate StorEdge A7000s

When RDC mirroring between separate A7000s, the primary node is the local node dsp1 and the secondary node is the remote node dsp3:

```
#pri_node      pri_vol        sec_node   sec_vol        Reserved
dsp1        /dev/rdsk/0d4     dsp3      /dev/rdsk/0d5        0
dsp1        /dev/rdsk/1d4     dsp3      /dev/rdsk/1d5        0
dsp1        /dev/rdsk/2d4     dsp3      /dev/rdsk/2d5        0
```

Mirroring Between Subsystems (Nodes) in the Same StorEdge A7000

When RDC mirroring across adjacent nodes in the same A7000, the primary node is the local node dsp1 and the secondary node is the adjacent node dsp2:

```
#pri_node      pri_vol        sec_node   sec_vol        Reserved
dsp1        /dev/rdsk/0d4     dsp2      /dev/rdsk/0d6        0
dsp1        /dev/rdsk/1d4     dsp2      /dev/rdsk/1d6        0
dsp1        /dev/rdsk/2d4     dsp2      /dev/rdsk/2d6        0
```

/etc/sd.cf File

The /etc/sd.cf file is part of the Storage Cache Management (SCM) component of the StorEdge A7000. It must include entries for each remote, or secondary, subsystem.

One key parameter in the file is auto-resync. This parameter controls automatic RDC synchronization. The default value is 0, disabling automatic synchronization:

- Disable auto-resynchronization for normal operations. You can temporarily enable auto-resynchronization to quickly validate link and secondary subsystem recovery following controlled outages (for example, during disaster preparedness drills).
- After a link or secondary site failure, the storage administrator at the primary site determines if the failures are isolated or part of a larger *rolling disaster.*

If auto-resync is enabled (set to 1), the following synchronizations automatically occur for RDC:

- After a transient or momentary link failure, the link automatically recovers and resynchronizes any modifications that have occurred since failure.
- After a secondary side failure is repaired, the primary link automatically starts an update to resynchronize any modifications that have occurred since failure.

Note – If the secondary volume has been replaced, you must perform an explicit full synchronization to restore the contents from the primary volume. See the rdc_copy_primary_to_secondary command described in this chapter and Chapter 3 for more information.

Caution – Do not start resynchronization if it is likely that the failure is an early warning of a rolling disaster that could permanently disrupt the resynchronization attempt. If the resynchronization is permanently interrupted, the secondary site is left in an inconsistent state, further complicating the recovery process.

Summary of RDC Commands

Note – Issue all `rdc_*` commands from the primary subsystem.

TABLE 2-2 RDC Commands Summary

Command	Description
`rdc_copy_primary_to_secondary`[1]	Starts copying the current primary volume contents to the secondary volume. It also enables concurrent RDC mirroring between the primary and secondary volumes
`rdc_suspend_mirroring`	Suspends RDC mirroring and copy operations between primary and secondary volumes and starts independent RDC logging on these volumes
`rdc_resume_mirroring`	Resumes RDC mirroring from primary volumes to secondary volumes currently suspended by RDC. It also updates the corresponding secondary volumes to match the primary volumes, based on RDC logs maintained while mirroring was suspended
`rdc_restore_primary_from_secondary`	Starts copying the current secondary volume contents to the primary volume. It also resumes concurrent RDC mirroring from the primary volumes to the secondary volumes
`rdc_refresh_primary_from_secondary`	Resynchronizes the primary volume from the secondary volume based on RDC logs maintained while RDC mirroring was suspended. It also enables RDC mirroring between the primary and secondary volumes
`rdc_disable_mirroring`	Terminates all RDC mirroring services between the specified primary and secondary volumes and discontinues the RDC relationship between these volume pairs
`sd_stats`	Displays RDC and cache statistics for the subsystem on which it is invoked

1. Issue all `rdc_*` commands from the primary subsystem

RDC Command Parameters

Each `rdc_*` command uses the parameters in TABLE 2-3. If you do not specify parameters, the command uses the default configuration file `/etc/rdc.cf`.

TABLE 2-3 RDC Command Parameters

Parameter	Definition
–f config_file	Specifies the configuration file defining the volume pairs.
pri_node	Specifies the primary or source node (subsystem).
pri_vol	Specifies the primary or source volume.
sec_node	Specifies the secondary or target node (subsystem).
sec_vol	Specifies the secondary or target volume.
0	Reserved value. You must specify 0.

rdc_copy_primary_to_secondary

`rdc_copy_primary_to_secondary`
`rdc_copy_primary_to_secondary` *–f config_file*
`rdc_copy_primary_to_secondary` *pri_node pri_vol sec_node sec_vol 0*

After you type the command, the response is displayed:

```
Overwrite secondary with primary? (Y/N) [N]
```

The default response is no (**N**). If you type **Y** (yes), the `rdc_copy_primary_to_secondary` command starts a full copy of the primary volume to the secondary volume. It also enables concurrent RDC mirroring between the primary and secondary volumes; any new writes to the primary are also mirrored to the secondary volume. See FIGURE 2-2. Use `sd_stats` to monitor the operation.

When To Use This Command

Use this command when all of the following cases exist:

- The contents of the primary and secondary volumes may be different
- The primary volume has the desired contents
- You wish to completely overwrite the contents of the secondary volume
- No logging information exists to incrementally resynchronize the volumes

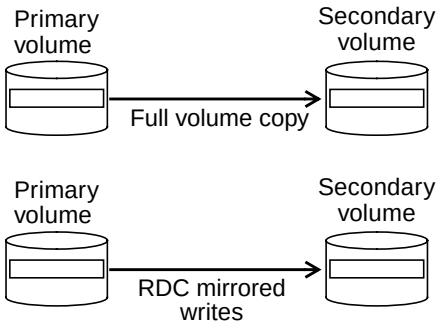

FIGURE 2-2 `rdc_copy_primary_to_secondary` Command Result

rdc_suspend_mirroring

`rdc_suspend_mirroring`
`rdc_suspend_mirroring` *–f config_file*
`rdc_suspend_mirroring` *pri_node pri_vol sec_node sec_vol 0*

After you type the command, the response is displayed:

```
Suspend RDC mirroring and log? (Y/N) [N]
```

The default response is no (**N**). If you type **Y** (yes), the `rdc_suspend_mirroring` command temporarily suspends any RDC mirror, copy, refresh, and restore operations between the primary and secondary volumes. It also starts independent RDC logging on these volumes.

RDC logging identifies the parts of each volume that are updated while RDC is suspended. RDC later uses these logs to resynchronize the volumes, when you issue the `rdc_resume_mirroring` or `rdc_refresh_primary_from_secondary` commands.

When To Use This Command

Use this command to temporarily stop RDC volume updates while maintaining the RDC logs.

Tip – You can use this command to create a point-in-time secondary copy of the primary volumes. Quiesce the workload of the primary volumes to ensure a clear point-of-consistency prior to suspending RDC. Once RDC mirroring is suspended, the workload may be reactivated and RDC logging will keep track of the changed blocks for later resynchronization.

---

rdc_resume_mirroring

Note – Before using this command, quiesce the workload to the affected volumes—stop the host application from writing to the volumes until the updates and copies are complete and the volume pairs are fully synchronized. This ensures that the primary and secondary volumes match before mirroring of new updates resumes. Use `sd_stats` to view the synchronization progress.

`rdc_resume_mirroring`
`rdc_resume_mirroring` *–f config_file*
`rdc_resume_mirroring` *pri_node pri_vol sec_node sec_vol 0*

The `rdc_resume_mirroring` command resumes RDC mirroring from the primary volume to the secondary volume. It also uses the RDC logs to resume any copy, refresh, or restore operations that were underway when RDC was suspended. Any new writes to the primary are also mirrored to the secondary volumes. See FIGURE 2-3. Use `sd_stats` to monitor the operation.

When To Use This Command

Use this command when mirroring on the specified primary volume has been suspended and active RDC logs exist.

Tip – You can use this command after an A7000 subsystem failure or intersite link failure.

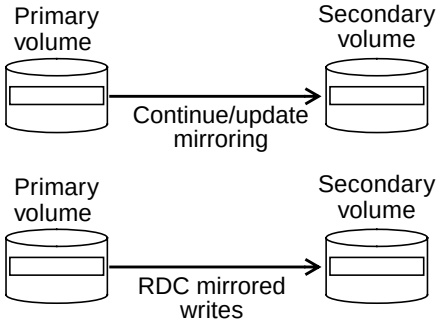

FIGURE 2-3 `rdc_resume_mirroring` Command Result

rdc_restore_primary_from_secondary

Note – Before using this command, quiesce the workload to the affected volumes—stop the host application from writing to the volumes until the updates and copies are complete and the volume pairs are fully synchronized. This ensures that the primary and secondary volumes match before mirroring of new updates resumes. Use `sd_stats` to view the synchronization progress.

`rdc_restore_primary_from_secondary`
`rdc_restore_primary_from_secondary` *–f config_file*
`rdc_restore_primary_from_secondary` *pri_node pri_vol sec_node sec_vol 0*

After you type the command, the response is displayed:

```
Overwrite primary with secondary? (Y/N) [N]
```

The default response is no (**N**). If you type **Y** (yes), the `rdc_restore_primary_from_secondary` command starts a full copy of the secondary volume to the primary volume. It also enables concurrent RDC mirroring between the primary and secondary volumes; any new writes to the primary are also mirrored to the secondary volumes. See FIGURE 2-4. Use `sd_stats` to monitor the operation.

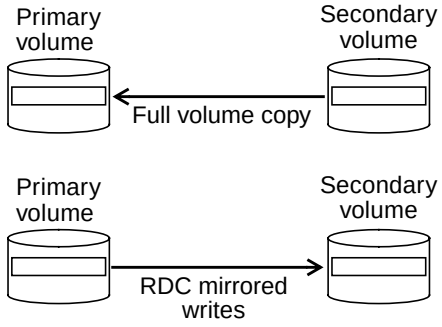

FIGURE 2-4 `rdc_restore_primary_from_secondary` Command Result

When To Use This Command

Use this command when all of the following cases exist:

- The contents of the primary and secondary volumes may be different
- The secondary volume has the desired contents
- You wish to completely overwrite the contents of the primary volume
- No logging information exists to incrementally resynchronize the volumes
- You intend to resume mirroring from the primary to the secondary volume

---

Tip – You can use this command to restore the contents of a replaced primary volume.

---

rdc_refresh_primary_from_secondary

Note – Before using this command, quiesce the workload to the affected volumes—stop the host application from writing to the volumes until the refresh is complete and the volume pairs are fully synchronized. This ensures that the primary and secondary volumes match before mirroring of new updates resumes. Use `sd_stats` to view the synchronization progress.

`rdc_refresh_primary_from_secondary`
`rdc_refresh_primary_from_secondary` *–f config_file*
`rdc_refresh_primary_from_secondary` *pri_node pri_vol sec_node sec_vol 0*

After you type the command, the response is displayed:

```
Refresh primary with secondary? (Y/N) [N]
```

The default response is no (**N**). If you type **Y** (yes), the `rdc_refresh_primary_from_secondary` command resynchronizes the primary volume from the secondary volume, based on RDC logs maintained while mirroring was suspended. It also enables concurrent RDC mirroring between the primary and secondary volumes; any new writes to the primary are also mirrored to the secondary volumes. See FIGURE 2-5. Use `sd_stats` to monitor the operation.

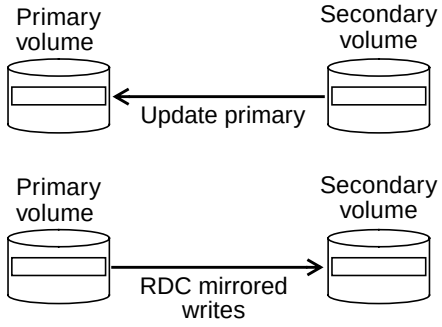

FIGURE 2-5 `rdc_refresh_primary_from_secondary` Command Result

When To Use This Command

Use this command when all of the following cases exist:

- RDC mirroring was suspended and RDC logging was active
- The secondary volume has the desired contents
- You wish to overwrite those segments of the primary volume that are different from the secondary as defined in the RDC logs
- You intend to resume mirroring from the primary to the secondary volume

---

Tip – You can use this command to roll back volume changes on the primary to a point-in-time image captured on the suspended secondary volume.

---

rdc_disable_mirroring

`rdc_disable_mirroring`
`rdc_disable_mirroring` *–f config_file*
`rdc_disable_mirroring` *pri_node pri_vol sec_node sec_vol 0*

Description

After you type the command, the response is displayed:

```
Discard scoreboards and terminate RDC association? (Y/N) [N]
```

The default response is no (**N**). If you type **Y** (yes), the `rdc_disable_mirroring` command terminates all RDC mirroring services between the specified primary and secondary volumes and discontinues the RDC relationship between these volume pairs. It also discards any active RDC logs that track temporary differences between primary and secondary volumes.

When To Use This Command

Use this command when the primary and secondary volumes no longer need to be associated with each other as RDC volume pairs.

As `rdc_disable_mirroring` discards any active RDC logs, use one of the following commands to re-establish identical mirrored pairs and re-enable RDC mirroring:

- `rdc_copy_primary_to_secondary`
- `rdc_restore_primary_from_secondary`

Note – Use `rdc_suspend_mirroring` instead of `rdc_disable_mirroring` to temporarily suspend mirroring and maintain the RDC volume pair association and related RDC logs.

Getting RDC and Storage Cache Statistics: the `sd_stats` Command

`sd_stats` displays storage cache information and can be used to view RDC statistics. Refer to the `sd_stats(1SCM)` man page for more detail about this command. The basic command syntax for RDC is:

```
/usr/install/scm/bin/sd_stats [-d sec] [-r n[:n][,n]...] [-z] [-M]
```

Note – Brackets [ ] contain optional arguments.

TABLE 2-4 describes the `sd_stats` options.

After you type `sd_stats` with the desired options, you can use the following keys:

- `f`—scroll forward through the display
- `b`—scroll backward through the display
- `m`—toggle between regular cache statistics and RDC screens
- `t`—display the cumulative key cache statistics generated since the last time the counters were reset (the `sd_stats -z` command resets the counters)

During synchronization operations, `sd_stats` displays:

- Percentage of the synchronization completed for each volume
- Arrows indicating the direction of the synchronization
- Status of each dual copy volume

TABLE 2-4 `sd_stats` Options

Option	Description
`-d` *sec*	Sets the display update time to sec seconds.
`-M`	Displays RDC statistics.
`-r` *n*[:*n*][,*n*] ..	Specifies the cache descriptor for one volume or a combination of a single volume, an inclusive range of volumes, and multiple volumes. The default is all volumes are displayed. n is the numbers of the specified decimal volumes. A colon (:) is a separator specifying an inclusive range of volumes. A comma (,) is a separator specifying another volume. The following two examples specify the same volumes (3, 6, 7, 8, 9, 10, 11, 12, 14, and 15): **`-r 3,6:7,8,9:12,14:15`** **`-r 3,6:12,14,15`**
`-z`	Clears the statistics first before displaying current statistics.

Example sd_stats RDC Screen Display

The sd_stats RDC statistics display includes the fields in TABLE 2-5. Type sd_stats -M to display RDC-specific statistics; press the m key to toggle between regular cache statistics and RDC screens after typing the sd_stats command.

TABLE 2-5 sd_stats Fields for RDC

Field	Description
partition	Specifies the primary dual copy partition.
link status	Indicates whether mirroring is enabled, disabled, or synchronization direction if a synchronization is in progress.
mirrored partition	Specifies the secondary dual copy partition.
dual copy status	Indicates current status of the dual copy. If a synchronization is in progress, it also displays the estimated time to completion in hours, minutes, and seconds.
recovery needed	Displays the percentage of segments of the partition that require recovery.
recovery completed	Displays the percentage of the partition that has been resynchronized.

See FIGURE 2-6. The ========= in the link status field indicates that the link is up and mirroring is enabled. The asterisks (*) in the recovery fields are graphical representations of the percentage of recovery needed and completed. The arrows (>) indicate the synchronization direction.

```
-----------------------------------------------------------------
SAMPLE 109056   ****** Dual Copy Statistics ******    17:40:10

partition     link status    mirrored partition  dual copy status

/dev/rdsk/0d4 *=========*    dsp3:/dev/rdsk/2d4  enabled
/dev/rdsk/1d4 * >>>>>>> *    dsp3:/dev/rdsk/3d4  resync time
00:02:50
-----------------------------------------------------------------
partition     recovery needed  recovery completed

/dev/rdsk/1d4  [**************** ] 84.23%[***         ] 15.78%
-----------------------------------------------------------------
```

FIGURE 2-6 sd_stats Link Status Display

Example sd_stats Screen Display

The sd_stats statistics include the fields in TABLE 2-6. Use the m key to toggle between these statistics and the RDC statistics. Use the t key to display the cumulative cache statistics generated since the last time the counters were reset. Type sd_stats -z to reset the counters.

The sd_stats sample display in FIGURE 2-7 contains the cache statistics and cumulative cache statistics displays.

TABLE 2-6 sd_stats Fields

Field	Description
cd	Specifies the cache descriptor number.
cached_partition	Specifies the disk partition being monitored.
disk_io	Displays the number of Kbytes per second read from or written to physical disks.
cache	Displays the number of Kbytes per second read from or written to data cache.
write_blocks	dirty—the number of dirty cache blocks that have not been queued for destaging. todisk—the number of blocks that have been queued for destaging and are waiting to be written to disks. The blocks displayed here are eventually moved to disk_io(writes).
accesses/s	Displays the number of I/O operations (reads per second + writes per second) serviced by the cache per second. The accesses/s total is the sum of the hits and misses.
read/s	Displays the number of read requests (hits) serviced per second. misses/s—Displays the number of read misses per second.
write/s	Displays the number of write requests serviced per second. misses/s—Displays the number of write misses per second.
%readh	Displays the percent of reads that are cache hits.
%writeh	Displays the percent of writes that are cache hits.
cachesize	Displays the cache size, as specified in the /etc/sd.cf file. Cache size is calculated as: *cachesize = (cache_mem + 2* write_cache_mem)* *1024*1024
blocksize	Displays the blk_size as specified in the /etc/sd.cf file.

TABLE 2-6 `sd_stats` Fields

Field	Description
Write blocks available	Displays the available write cache in blocks. `Net` *n* in the field represents each MEMORY CHANNEL network. A network is added for each MEMORY CHANNEL board that is added.
LRU stat	Displays the Least Recently Used (LRU) algorithm statistics: Blocks—the number of cache blocks in the LRU (always the total number of blocks in the cache). This figure stays constant. Requeued—unused Optimized—cache blocks not requeued to the tail of the LRU.
Total Cache Memory Usage	Displays the size of system memory that is used by the cache when cache starts.
Total Stats Memory Usage	Displays the size of system memory that keeps the cache statistics data.

```
SAMPLE 49    ****** A7000 Storage Cache ******     08:49:59


                          disk_io       cache        write_blocks
 cd cached_partition  reads writes  reads writes  dirty todisk failed
 ------------------------------------------------------------------------
  0 /dev/rdsk/38d5         0      0   1511      0      0      0      0
  1 /dev/rdsk/41d4         0   1151      0   1205     33     45      0
  2 /dev/rdsk/48d4       305      0   1205      0      0      0      0
  3 /dev/rdsk/53d4         0   1159      0   1213     33     45      0
                     ------ ------ ------ ------
      Kbytes/s total:   305   2310   2716   2418

 accesses/s    read/s    write/s   %readh   %writeh
             (misses/s) (misses/s)
 ---------------------------------------------
    680.33     339.68     302.48     89.9     100.0
             (  38.18 ) (   0.00 )


SAMPLE 54  ****** A7000 Storage Cache (Cumulative) ******      08:50:04
                            disk_io                  cache
 cd cached_partition      reads     writes      reads     writes
 ------------------------------------------------------------------
  0  /dev/rdsk/38d5            0          0       59782          0
  1  /dev/rdsk/41d4           33      47448          52      47528
  2  /dev/rdsk/48d4        11805          0       46932          0
  3  /dev/rdsk/53d4           33      46720          52      46800
                       ---------- ----------  ---------- ----------
        Kbytes total:     11871      94168      106818     94328

  accesses        read        write    %readh  %writeh
          (    misses) (    misses)
 ------------------------------------------------------------
    26605       13336        11791      90.0    100.0
          (      1478) (         0)

 cachesize  blocksize
 --------------------
   372736K       4096

 Write blocks available:
 ----------------------
 Net 0:   8036   Net 1:      0    Net 2:      0    Net 3:      0
```

```
LRU stats:  Blocks     Requeued    Optimized
-------------------------------------------
             84992         1485        75174

Total Cache Memory Usage: 11804 Kbytes
Total Stats Memory Usage: 20 Kbytes
```

FIGURE 2-7 Sample `sd_stats` Display

CHAPTER 3

RDC Operations

Remote Dual Copy (RDC) operations range from establishing initial mirrored pairs to recovering from a failed primary storage platform. This chapter describes each operation.

Before You Begin Using RDC

This section describes things to consider before you start using the full capabilities of Remote Dual Copy.

- Primary and secondary sites defined

  In the examples in this chapter, Site A is the local primary site and Site B is the remote secondary site. Depending on the example, either site can be the primary or secondary of the remote dual copy operation. All RDC operations start from the primary subsystem, as you issue the commands from the primary subsystem. See Chapter 2.

- Monitoring RDC operations

  Use `sd_stats` at the A7000 console to monitor the progress of RDC operations. Use a separate `sd_stats` window for each subsystem. Chapter 2 describes this command. You may monitor a subset of RDC statistics on the secondary subsystem from the secondary subsystem console.

- Mutual backup considerations

  Sometimes the distinction between primary and secondary sites is blurred. As applications are geographically distributed, a storage subsystem at Site B may function as a remote disk backup to Site A and as a direct storage resource for applications on Host B. Under these circumstances, keep remote-dual-copies of Host B volumes at Site A. This reciprocal backup arrangement supported by RDC is called *mutual backup*.

  With mutual backup, the RDC volumes considered primary by Site B are administered from the Site B A7000 console. Site B remote-dual-copied devices are considered secondary volumes at Site A.

- Selecting volumes for RDC

  One initial RDC planning step is to determine the volumes to include in the remote mirroring operations. Balance remote accessibility and recoverability against capacity usage and I/O response time.

  Generally, you will include the following critical volumes in the remote mirroring configuration:
  - Databases and database management system (DBMS) logs
  - Access control files

  You can enable these volumes for remote copy individually on the RDC command line or collectively using a configuration file, as described in Chapter 2.

  You can exclude volumes from the RDC configuration if they can be reconstructed at the recovery site or if they seldom change:
  - Temporary volumes (such as those used in sorts)
  - Spool files
  - Paging volumes

- Host disks and A7000 pseudo devices

  The A7000 maps host disks into pseudo devices, also known as logical volumes. These pseudo devices are identified by partition names of the form:

  `/dev/rdsk/`*`xyz`*

  where *xyz* is the partition designator; for example, 10d4. The `sd_stats` and RDC commands use partition names in their configuration files, command line options, and displays.

  Use the A7000 DASD Manager print configuration utility to determine the partition name associated with a a particular host disk. The files `/etc/ckd.cf`, `/etc/ste.cf`, and `/etc/dsf.cf` define the host disk-to-partition mappings for Count-Key Data (CKD), SCSI Target Emulation (STE), and DataShare Facility (DSF), respectively.

  Partition names may also be aliases for logical volumes that are physically housed on two or more Head Disk Assemblies (HDA). For example, a host device can be mapped into a RAID 1 pseudo device, sometimes known as a mirrored partition (mp). The data for that pseudo device is mirrored across two separate HDAs under the control of the same A7000 subsystem. RAID 5 logical volumes are similarly structured, but the data is spread over several drives with a rotating parity structure.

  RDC operates at the pseudo device level and is not aware of the physical data distribution; it can copy primary volumes into physically-dissimilar secondary pseudo devices, as long as their format and capacities match. RDC does not distinguish between CKD volumes and STE pseudo devices.

Establishing Mirrored Images for the First Time

Before using RDC mirrored updates, make sure that the contents of the local and remote volume pairs match. In the examples, Site A is the local primary site and Site B is the remote secondary site. Depending on the example, either site can be the primary or secondary of the remote dual copy operation. All RDC operations start from the primary subsystem, as you issue the commands from the primary subsystem.

Existing Primary and Newly-Formatted Secondary

If the primary volume has already been updated locally (without RDC mirroring) and contains important information, you must first copy its image to the newly formatted secondary volume. Perform this full synchronization for the specified primary and secondary with the `rdc_copy_primary_to_secondary` command. RDC can perform this initial volume-to-secondary volume copy while concurrently forwarding new online updates to the secondary.

To fully synchronize specific remote volumes while applying new updates, type:

```
dsp1# rdc_copy_primary_to_secondary -f config_file
```

FIGURE 3-1 shows the full resynchronization process. It also enables concurrent write updates from the local to remote volumes. These steps repeat until the entire volume is copied. Use `sd_stats` to monitor the progress. RDC will enable up to 8 concurrent resynchronizations. As these complete, others begin until all the requested volume pairs have been synchronized.

The data flow is:

1. RDC on dsp1 requests disk blocks from the active primary volume. The data may already be resident in the dsp1 data cache, or may require a local disk access.
2. RDC transmits the cache blocks, with destaging instructions, over the intersite fiber link (**2R**) to an RDC cache region on remote dsp3.
3. RDC on dsp3 updates its remote disk and acknowledges the update to dsp1.

These steps repeat until the entire volume is copied. Use `sd_stats` to monitor the progress. RDC will enable up to 8 concurrent resynchronizations. As these complete, others begin until all the requested volume pairs have been synchronized.

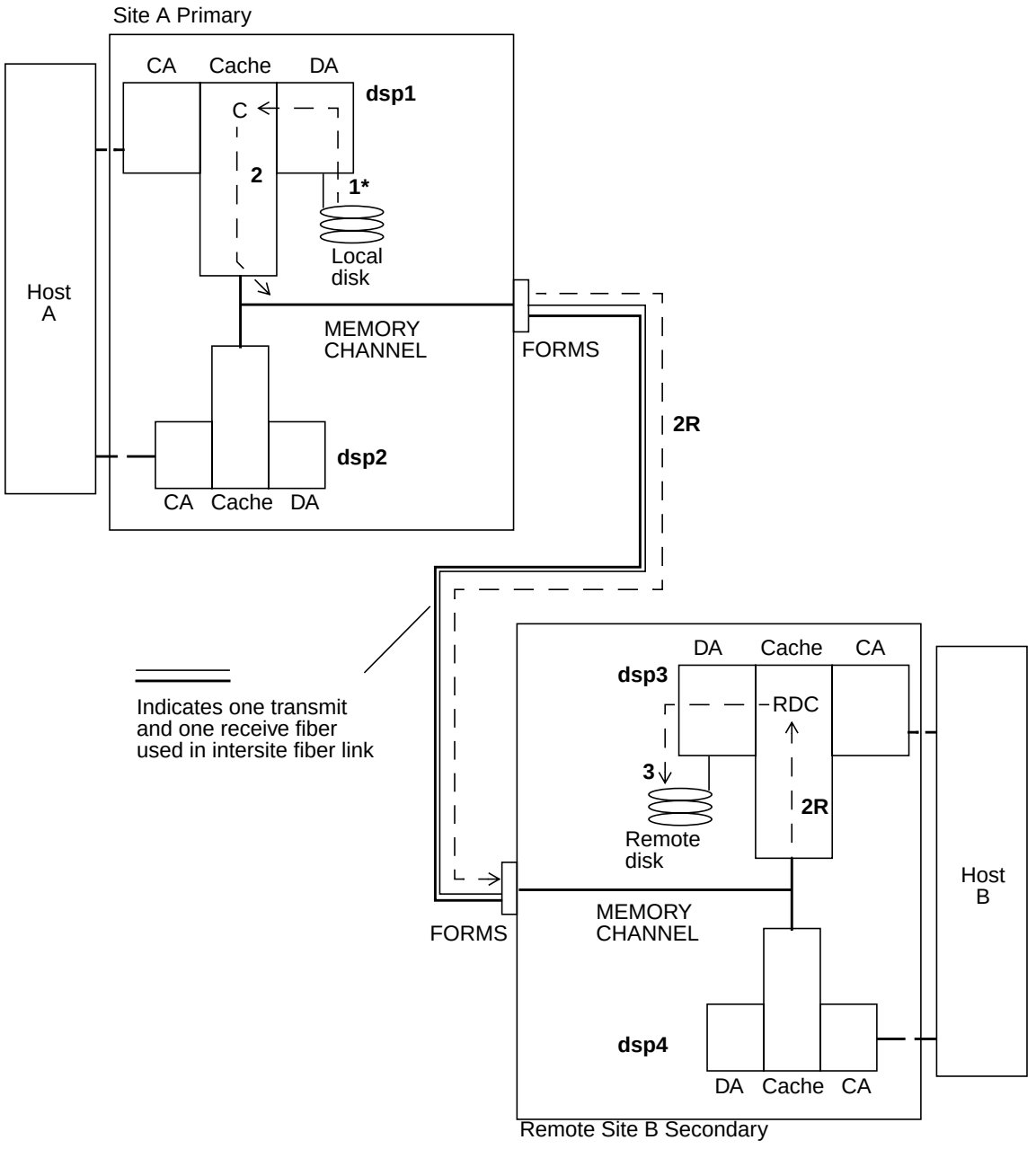

FIGURE 3-1 Full Synchronization (Volume-to-Volume Copy)

RDC Mirroring

You realize the full benefits of real-time RDC mirroring once the local and remote pairs are synchronized. RDC mirroring is enabled when the full volume copy was requested using the `rdc_copy_primary_to_secondary` command. Mirroring continues concurrently with the volume copy process as the StorEdge A7000 server receives new host updates.

FIGURE 3-2 shows the real-time remote mirroring from an A7000 at Site A to a remote A7000 at Site B. The concurrent copy has been omitted from the drawing.

1. Host A issues a write request to its directly-attached A7000.
2. The Channel Adapter (CA) deposits the write data (D) into dsp1 data cache.
3. The write block is reflected to a *safe* region in dsp2's cache using the interconnecting MEMORY CHANNEL (REFLECTIVE MEMORY System). This write-cache area is reserved to keep a second local copy of the active cache updates until they are destaged to disk.

   The A7000 multicasts transmissions on the MEMORY CHANNEL in real-time to the remote A7000 over the interconnecting fiber (**3R**). The Fiber-Optic REFLECTIVE MEMORY (FORMS) hardware transparently deposits this third copy of the write data in an RDC-specific region of dsp3's cache board.
4. The conditions for multicast stable storage are met and the host is notified of I/O completion.
5. Software on dsp1 periodically destages the write data from cache to the appropriate local disk using the Disk Adapter (DA).
6. RDC on dsp1 requests that dsp3 destage its copy of the write data to the appropriate remote mirror disk. This request is sent over the intersite fiber link.

Note – Steps 1 through 5 are identical to a simple local (non-RDC) write, making the remote dual copy operation invisible to the host.

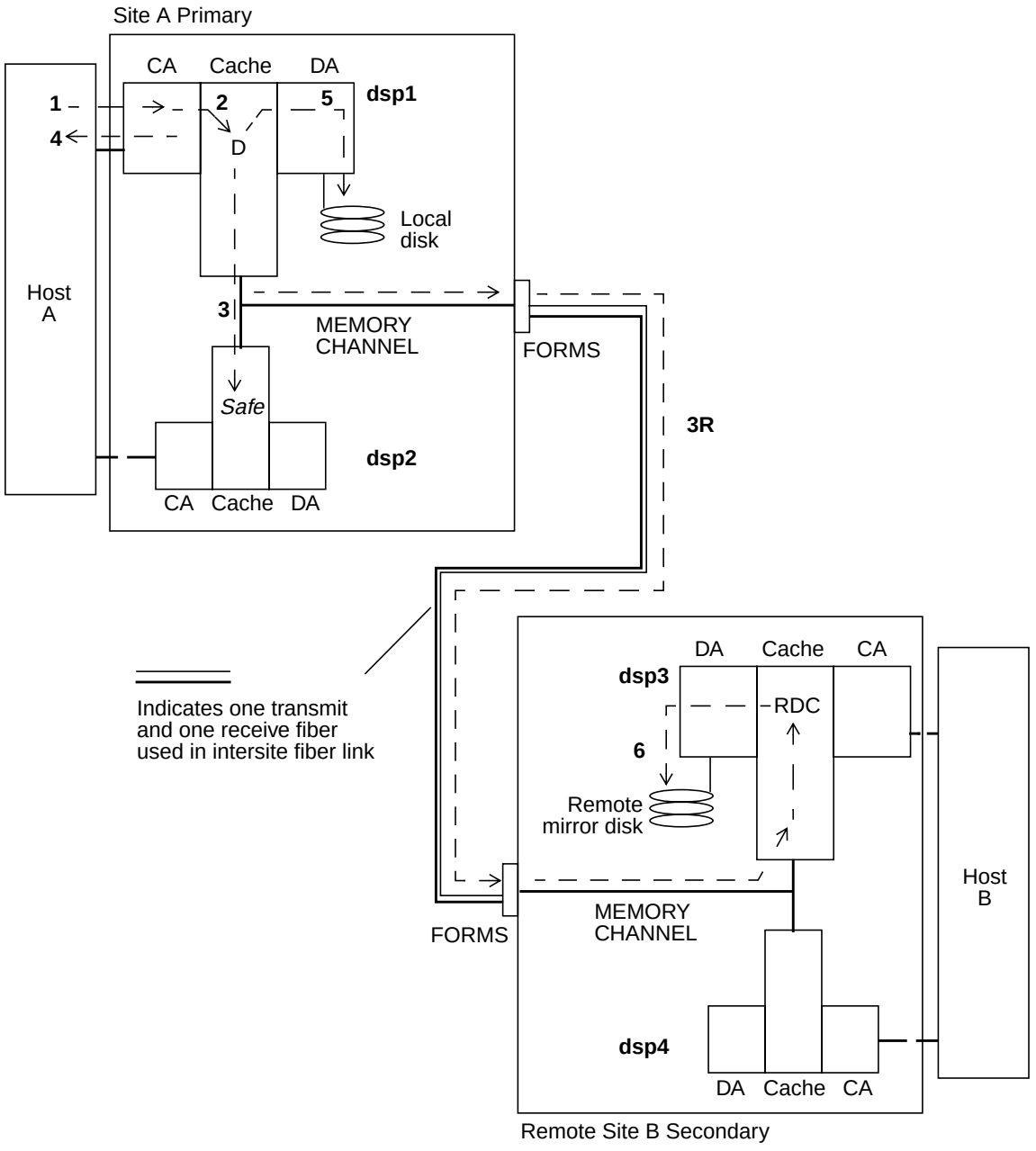

FIGURE 3-2 Real-Time Remote Mirroring

Handling RDC Interruptions

RDC uses a periodic intersite heartbeat to monitor the health of primary and secondary subsystems. The absence of a heartbeat indicates an interruption in the remote dual copy service. Interruptions can arise from a failure or impairment in the intersite link or an outage at the remote site. In some cases, interruptions are introduced intentionally to exercise failure contingencies, as during the disaster recovery rehearsals described later in this chapter.

Primary Logging During RDC Interruptions

During interruptions, RDC software tracks primary volume areas that are being updated locally but have not yet been copied to the secondary site. Under several circumstances, when the RDC service is restored, update resynchronization may be requested using the logging information to update the remote site.

FIGURE 3-3 shows the case where the intersite fiber has been cut.

1. Host A issues a write request to its directly-attached A7000.
2. The Channel Adapter (CA) deposits the write data (D) into dsp1 data cache.
3. The write block is reflected to a *safe* region in dsp2 cache using the interconnecting MEMORY CHANNEL (REFLECTIVE MEMORY System). This write-cache area is reserved to keep a second local copy of the active cache updates until they are destaged to disk. Because the remote service is impaired, RDC keeps a scoreboard log for each affected volume to indicate updates not yet remotely copied (**3S**).
4. The conditions for multicast stable storage are met and the host is notified of I/O completion.
5. Software on dsp1 destages the write data to the appropriate local disk using the Disk Adapter (DA).

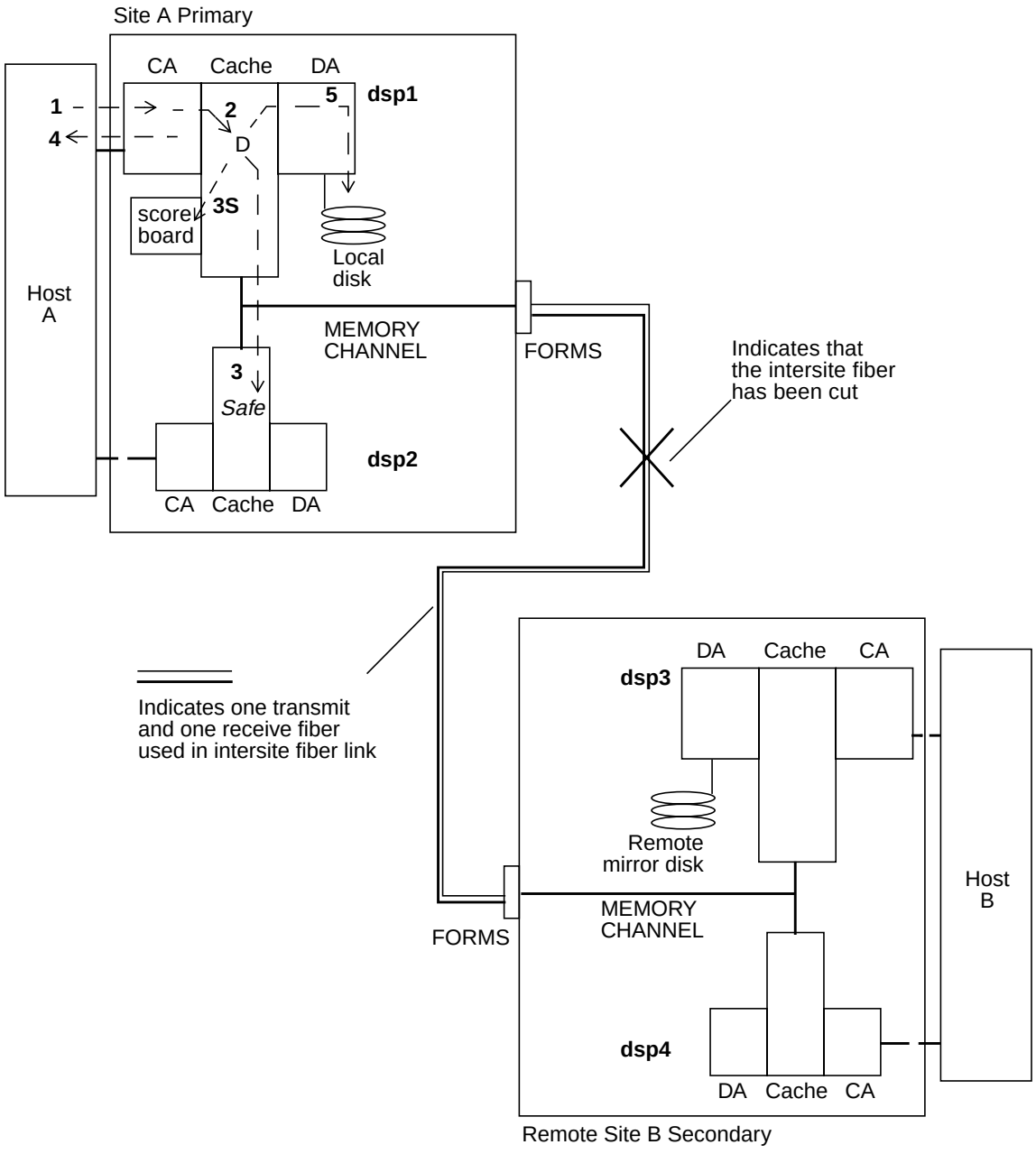

FIGURE 3-3 Primary Write Logging During RDC Interruption

Resynchronizing Volumes After an RDC Interruption

Typically, interruptions in the Remote Dual Copy services are infrequent. Whether the event is an isolated incident or part of a larger disaster determines the action you should take.

Caution – Ensure that you understand the resynchronization process (updating both sites) before invoking it. While resynchronization is underway, the secondary volumes are temporarily inconsistent and cannot be relied on for recovery. Consistency is restored when the resynchronization completes.

If the RDC interruption is the warning of a larger rolling disaster, avoid resynchronization of the sites. Maintain the secondary site in a dated but consistent state, rather than risk a disastrous interruption that leaves the secondary inconsistent and difficult to recover from. This is the reason that the auto-resynchronization option in the `/etc/sd.cf` file is disabled by default.

When you determine that the RDC interruption is an isolated incident and the condition has been repaired, enable update synchronization using the following RDC command:

```
dsp1# rdc_resume_mirroring
```

Note – If the secondary volume state is unknown because of subsystem or disk failure, you may need to make full volume copies to re-establish matching RDC pairs. In this case, use the `rdc_copy_primary_to_secondary` command.

FIGURE 3-4 shows an update resynchronization from the primary subsystem to its secondary subsystem, when the secondary volumes have been left stale by the interruption.

1. RDC on dsp1 examines a scoreboard for one of the RDC-managed volumes affected by the interruption.
2. RDC on dsp1 requests the blocks updated during the interruption from the up-to-date volume. The data may already be resident in dsp1 data cache, or it may require a local disk access.
3. RDC on dsp1 transmits the update blocks to dsp3 RDC cache region using the intersite fiber (**3R**).
4. RDC on dsp3 refreshes its stale mirror image with the updated blocks and acknowledges the action to dsp1.
5. RDC revises the scoreboard to track the remote update.

All steps repeat until the remote mirror is up-to-date. Use `sd_stats` to monitor the resynchronization progress.

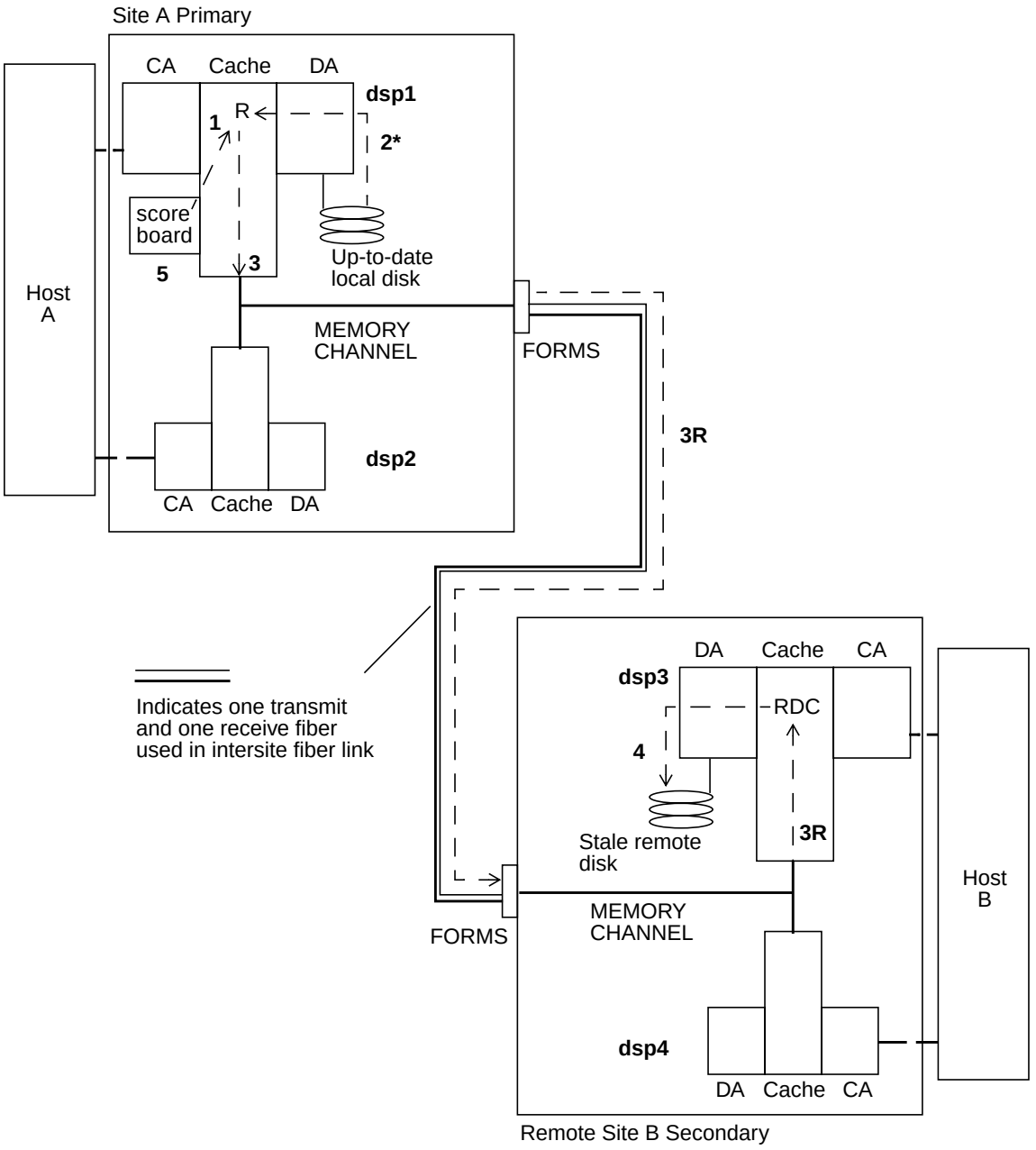

FIGURE 3-4 Update Synchronization of a Secondary

Restoring Volumes After Secondary Site Failures

The effort required to restore a secondary site to its remote dual copy state depends on the hardware and software that is replaced or repaired. This section describes restoring disk, subsystem, and intersite links.

Restoring a Failed Secondary Disk

FIGURE 3-5 shows restoration of a failed secondary disk. It follows the same steps used to establish an initial RDC pair. The concurrent write updates have been omitted from the figure. Type:

```
dsp1# rdc_copy_primary_to_secondary dsp1 /dev/rdsk/abc dsp3
/dev/rdsk/xyz 0
```

where `/dev/rdsk/`*`abc`* is the active primary volume on dsp1 and `/dev/rdsk/`*`xyz`* is the newly formatted secondary volume on dsp3.

1. RDC on dsp1 requests disk blocks from the active primary volume. The data may already be resident in dsp1 data cache, or may require a local disk access.
2. RDC transmits the data blocks, with destaging instructions, over the intersite fiber link to an RDC region on remote dsp3 (**2R**).
3. RDC on dsp3 updates its remote disk and acknowledges the action to dsp1.

All steps repeat until the entire volume is copied. Use `sd_stats` to monitor the restoration process.

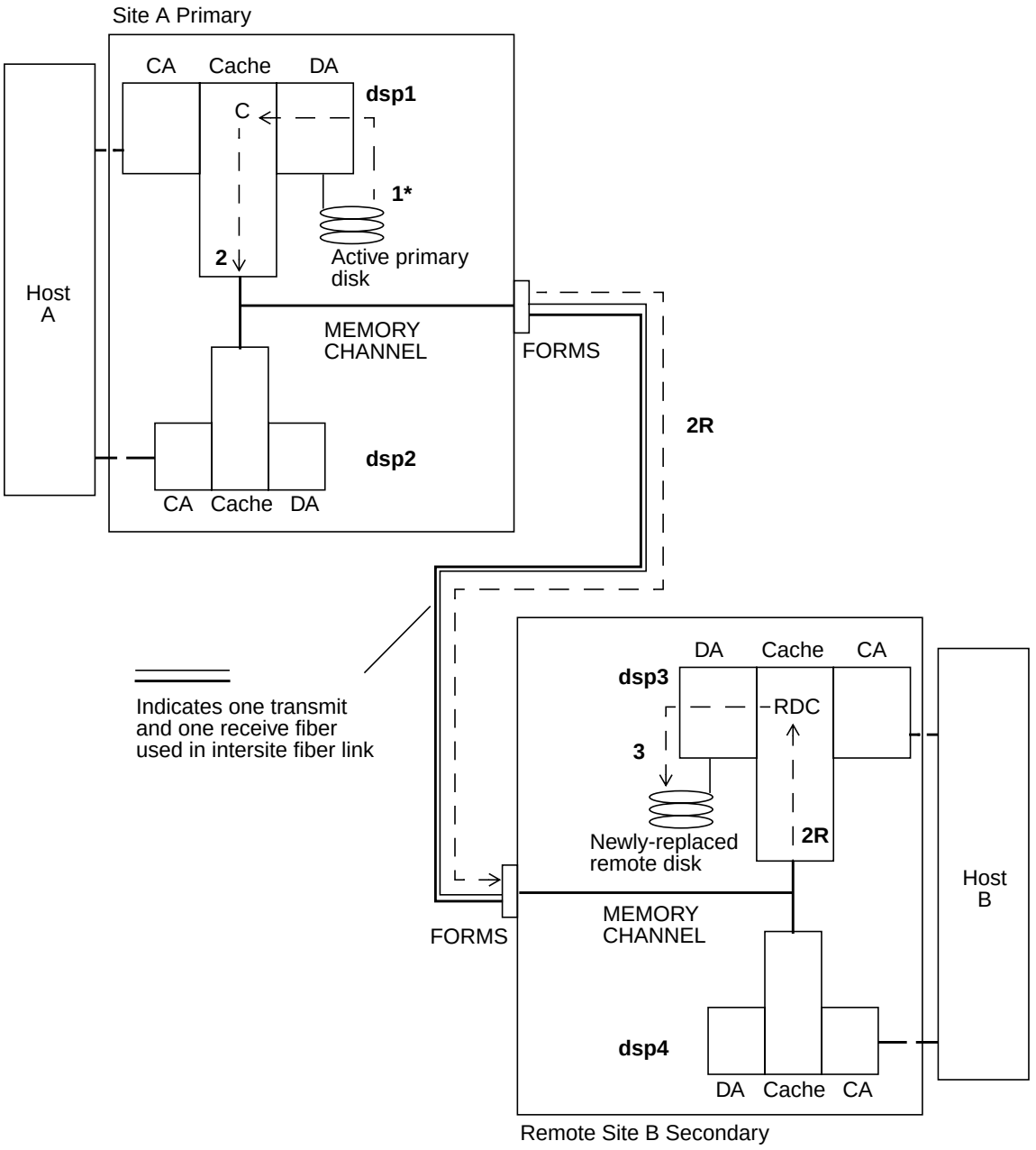

FIGURE 3-5 Restoring a Failed Secondary Disk

Restoring a Failed Secondary A7000

Restoring a failed secondary A7000 may involve both updating and fully resynchronizing the secondary disks, depending on the severity and duration of the failure.

The resume command updates any secondary volumes whose contents were unchanged by the failure, and whose primary members have seen little activity during the outage.

```
dsp1# rdc_resume_mirroring -f  config_file
```

where the *config_file* lists volumes to be updated.

Fully synchronize volumes residing on secondary disks that were replaced or whose state is unknown:

```
dsp1# rdc_copy_primary_to_secondary -f  config_fileZ
```

where the *config_fileZ* lists volumes requiring full volume-to-volume copies.

Switching to an Alternate Fiber

Most intersite link impairments are temporary, but a fiber cable can be accidentally cut. Fiber plant suppliers support separately-routed links between the two sites for added protection. They may use an automatic switch or may need to manually switch fiber cable pairs at a patch panel to redirect the signal to the alternate route. In each case, the RDC FORMS receivers sense the temporary loss of optical signal, triggering RDC suspension and logging.

FIGURE 3-6 shows the use of the alternate fiber links to circumvent cable cuts and other long-term intersite fiber impairments.

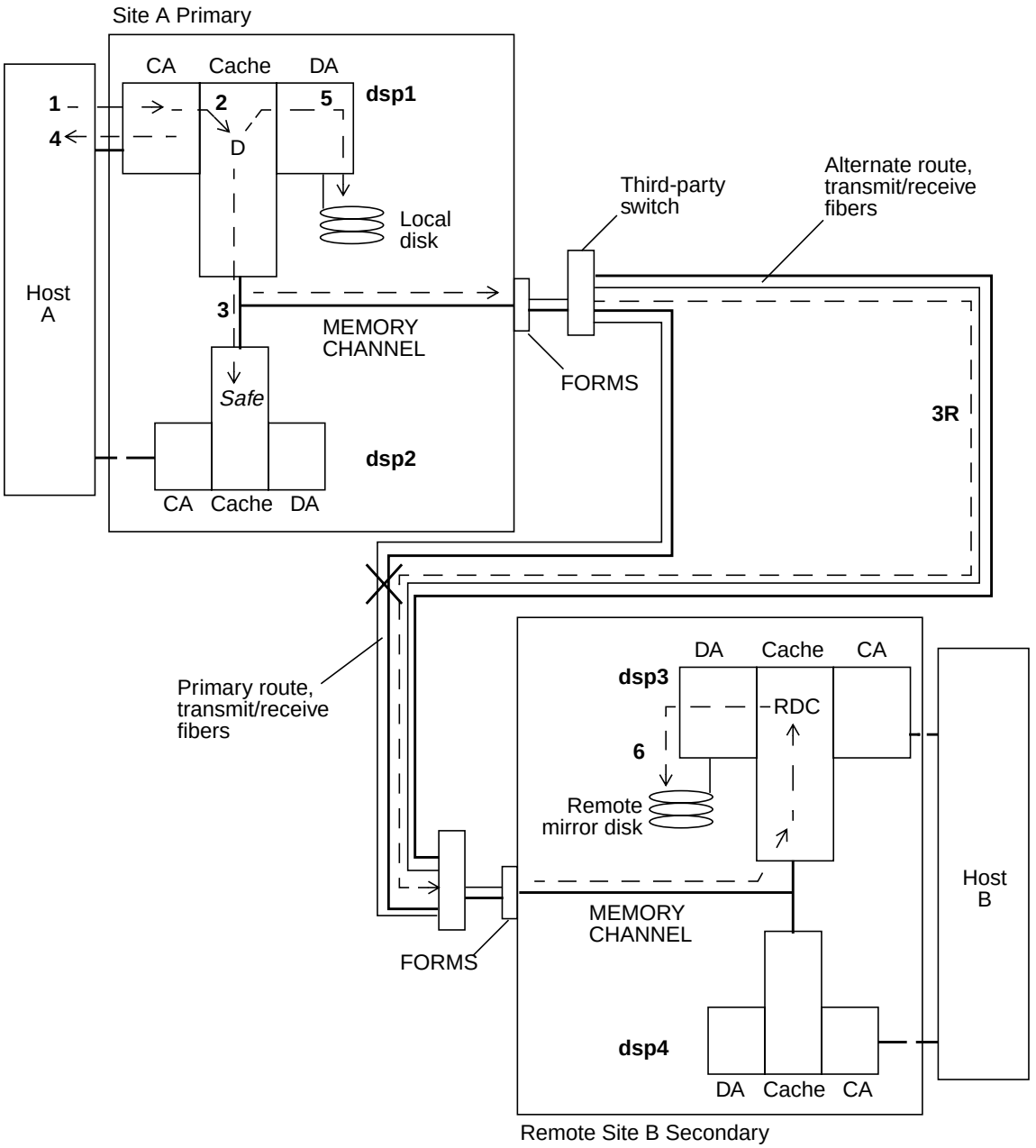

FIGURE 3-6 Alternate Fiber Connections

Disaster Recovery Rehearsals

Validating contingency procedures through rehearsals is an important aspect of any Disaster Recovery Plan. Perform rehearsals regularly and refine them whenever a significant change is made to the primary or secondary processing environments.

RDC lets you perform disaster rehearsals easily, encouraging frequent verification of disaster preparedness.

Rehearsing Recovery From Secondary Volumes

Some organizations simulate a disaster by temporarily interrupting the RDC link and practicing takeover on the secondary site. With RDC, disable the optical transmission on the FORMS Fiber Transmission Module card at the secondary A7000 server. See the *StorEdge A7000 Service Manual* for FORMS switch locations.

Complete these steps before proceeding with recovery:

1. Use `sd_stats` to verify that all RDC secondary volumes have been destaged to disk. Destaging may be in progress if RDC mirroring activity was high prior to the rehearsal.
2. Declare the secondary volumes online for access from the recovery hosts.
3. Start application-level recovery procedures to ensure a consistent starting point for future transactions.

FIGURE 3-7 shows a read operation from the secondary recovery copy.

1. Host B issues a read request on dsp3 against the recovery copy.
2. Step 1 being a cache miss, the appropriate blocks are read from the recovery copy disk into dsp3 data cache.
3. The read operation results are returned to Host B.

Subsequent host reads and writes populate the recovery subsystem data cache and normal local access is restored.

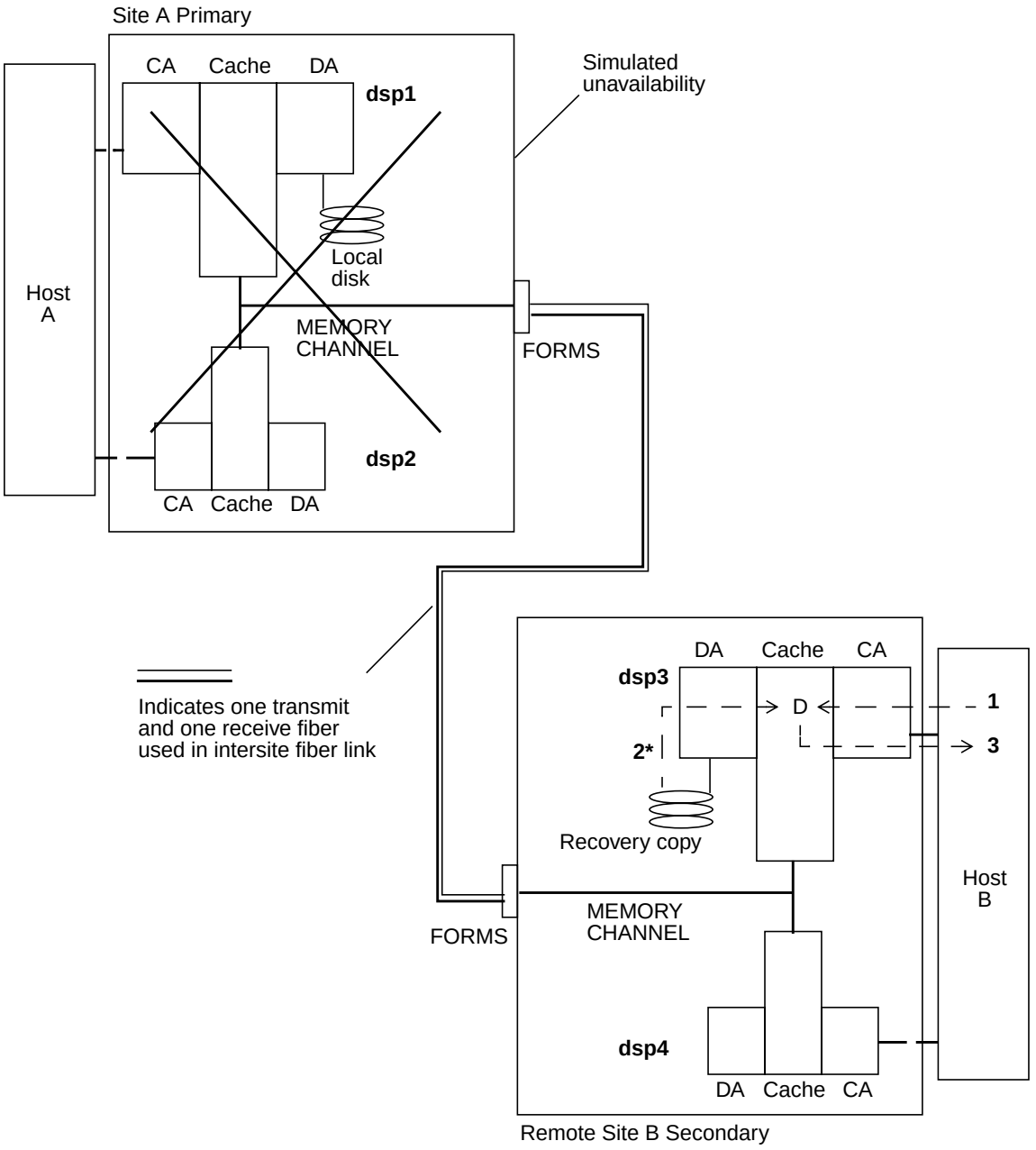

FIGURE 3-7 Reading from Recovery Copy

Secondary Updates During a Takeover Rehearsal

During the rehearsal, you can apply test updates from a secondary host to the secondary volumes to evaluate the recovery mechanisms. If these test updates are not part of the permanent business record, they must be undone when the recovery copy is returned to its secondary RDC role. See “Rolling Back Updates After a Takeover Rehearsal” for the appropriate procedures.

RDC logs track the test updates occurring on the secondary volumes during recovery rehearsals.

FIGURE 3-8 shows secondary write logging during rehearsals.

1. Host B issues a write request to its directly-attached A7000.
2. The CA places the write data D into dsp3 data cache.
3. The write block is reflected to a *safe* region in dsp4’s cache using the interconnecting MEMORY CHANNEL (REFLECTIVE MEMORY System). This write-cache area is reserved, keeping a second local copy of the active cache updates until they are destaged to disk. RDC keeps a scoreboard for each RDC-managed volume to indicate changes from the last known image synchronized with dsp1.

   These logs/scoreboards may be used later to overwrite the changed segments with the most recent data from the primary volumes.
4. The conditions for multicast stable storage are met and the host is notified of I/O completion.
5. Software on dsp3 destages the write data to the local disk using the DA.

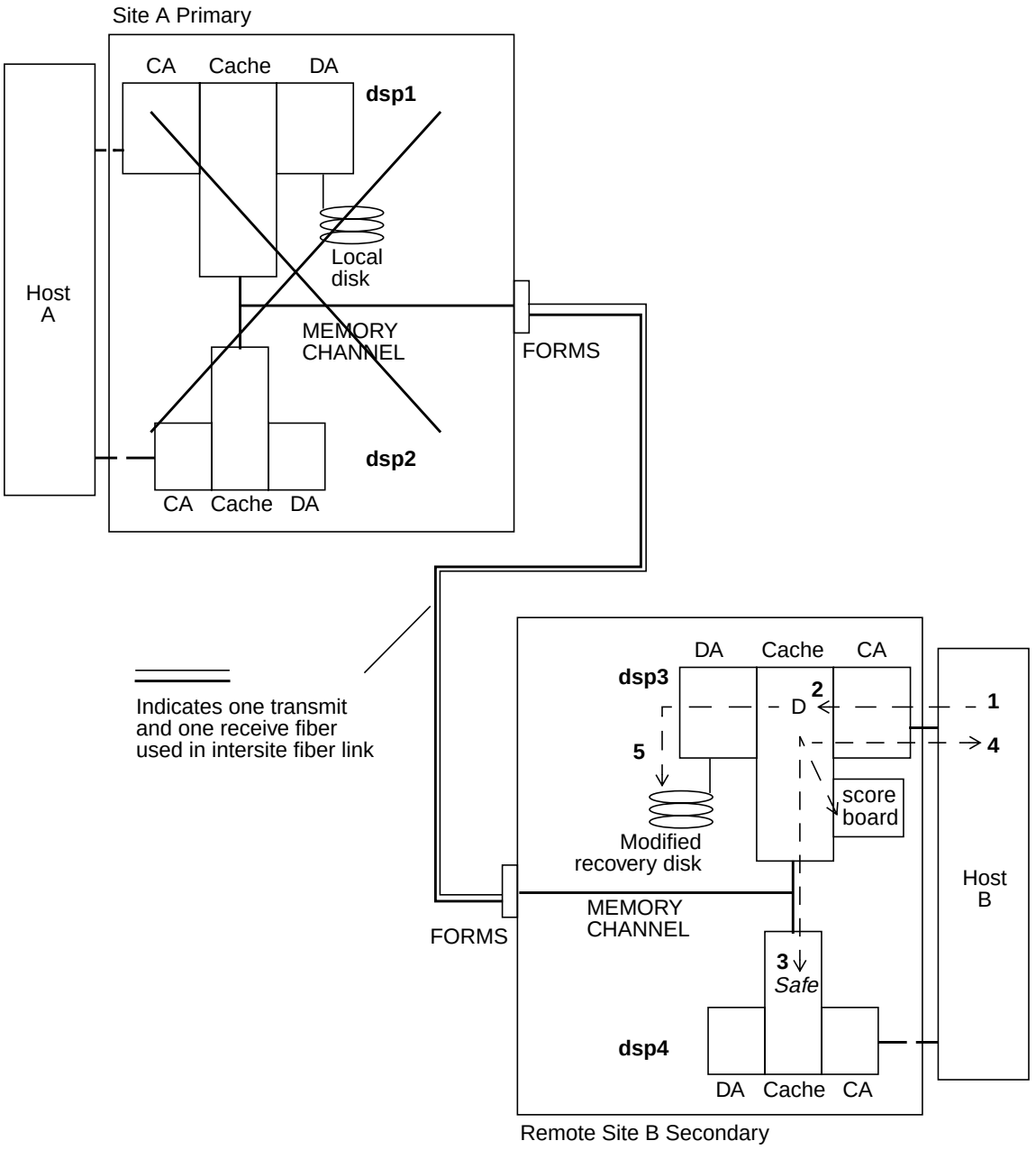

FIGURE 3-8 Secondary Logging

Rolling Back Updates After a Takeover Rehearsal

When the rehearsal completes, RDC uses the secondary logs (scoreboards) to update the appropriate blocks with the current information from the primary devices; the update procedure rolls back the test changes. At the primary console, enable RDC secondary rollback by typing:

```
dsp1# rdc_resume_mirroring
```

A more complete recovery rehearsal requires that the rehearsal updates be captured and later reflected on the primary A7000 before the workload is switched back. See “Updating the Primary Site From the Secondary Site” on page 3-31” for more detail.

Primary Site Failures

This section contains information needed for handling primary site failures.

Failed Primary Disk

RDC provides continuous data access during primary disk failures. The RDC high-availability features are a superset of RAID 1 and RAID 5 storage protection that can be optionally configured for the primary volumes. The RDC remote disk access features start only after the disk protection schemes on the primary subsystem are unable to provide data access to the local devices.

In the linear and striped (RAID 0) cases, failure of a single disk storing the primary volume triggers RDC to transparently redirect disk reads and writes to the remote storage subsystem.

If the primary logical volume is locally mirrored (RAID 1) across two physical disks on the same subsystem, a single disk failure results in its local mirror disk handling all requests for cache staging on a read miss and cache destaging. RDC relies on the remote site secondary devices only if both local mirrors fail.

If the primary volume is RAID 5 protected, its contents are striped across several physical disks. The local subsystem considers the primary volume inaccessible and yields to RDC remote disk access only when two or more of the disks in the RAID 5 stripe fail.

FIGURE 3-9 shows the RDC transparent pass-through of read requests to the remote subsystem bypassing the local disk failure. Writes continue to be handled as with a working local disk, though no destaging occurs on the primary subsystem.

1. Host A issues a read request for failed local disk on dsp1.
2. If the disk block is resident in dsp1 cache, it is immediately returned. However, if the read results in a cache miss, the attempt to access the local disk fails and RDC forwards the read request to dsp3.
3. RDC on dsp3 acts on the remote request by reading its remote mirror disk.
4. RDC on dsp3 responds to dsp1 with the requested disk blocks. The read data is used to refresh dsp1 cache.
5. The read data is returned to Host A from dsp1 cache.

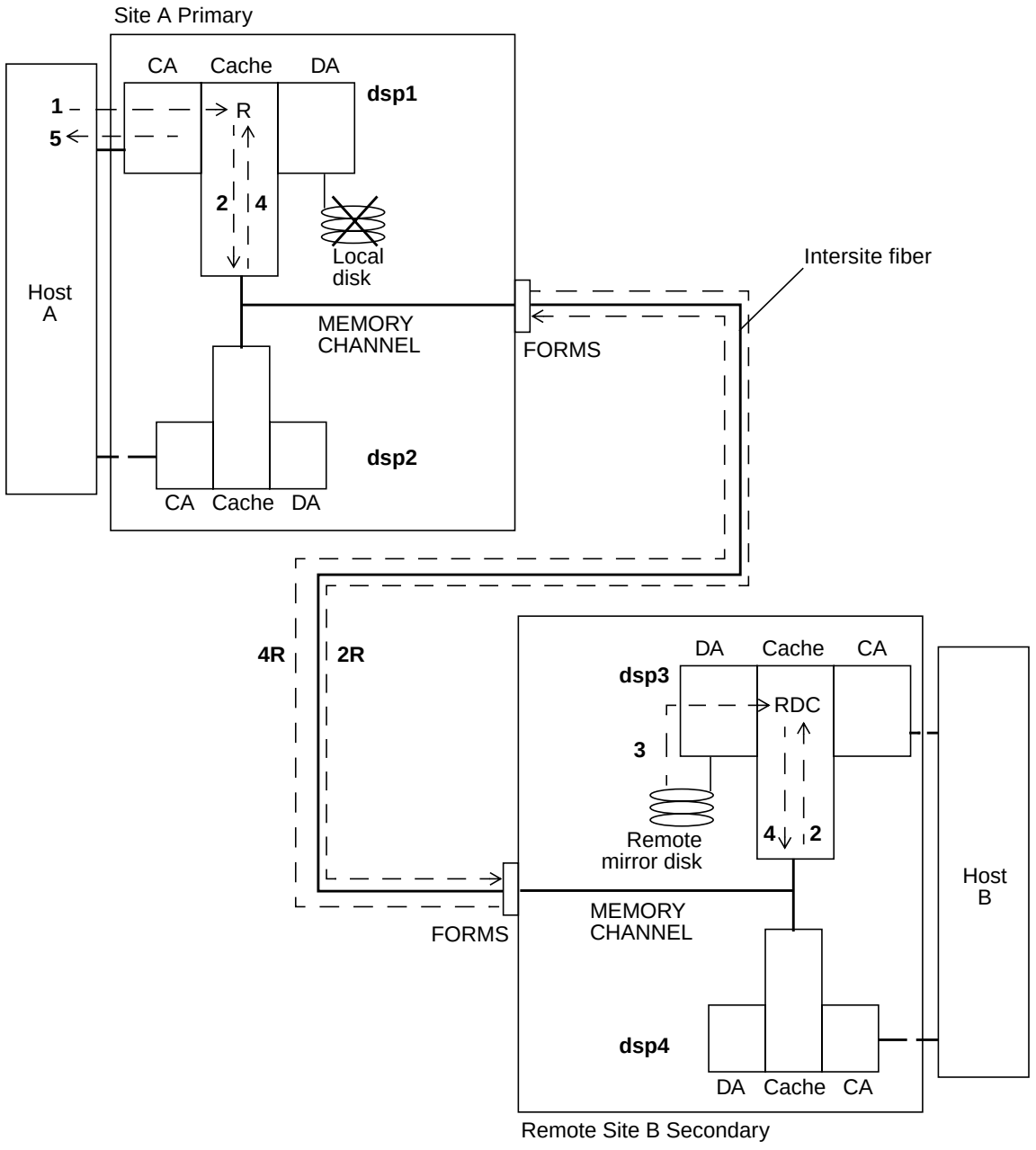

FIGURE 3-9 Pass-Through Reads to Bypass Local Disk Failure

▼ To Restore a Failed Primary Volume (Example)

1. **Keep the failed volume enabled under RDC.**

   RDC marks the device as failed when it is unable to read or write from it. RDC continues to provide read and write services to the host application using the secondary volume at the remote site.

Caution – Disabling RDC on the failed device causes application access to the device to fail and disassociates the failed primary volume from its secondary copy.

2. **Contact Sun Enterprise Services to arrange for replacement of the failed device and restoration of the associated logical volumes.**

3. **After the disk failure is corrected and logical volume formatting restored, restore the primary volume from its secondary image. For example:**

```
dsp1# rdc_restore_primary_from_secondary dsp1 /dev/rdsk/abc dsp3
/dev/rdsk/xyz  0
```

   This is a reverse full synchronization, as the secondary volume (*xyz*) on dsp3 is resynchronizing the new primary (*abc*) volume on dsp1. FIGURE 3-10 shows the full reverse synchronization process.

   1. The data may already be resident in dsp1 data cache, or it may require a secondary disk access. If so, RDC on dsp1 requests blocks from the up-to-date secondary volume on dsp3.
   2. RDC on dsp3 transmits the cache blocks over the intersite fiber link to an RDC region on dsp1 with destaging instructions.
   3. RDC on dsp1 updates its disk.

   All steps repeat until the entire volume is copied. Subsequent reads from Host A are serviced locally from the newly updated disk. Reverse synchronization occurs concurrently with pass-through reads.

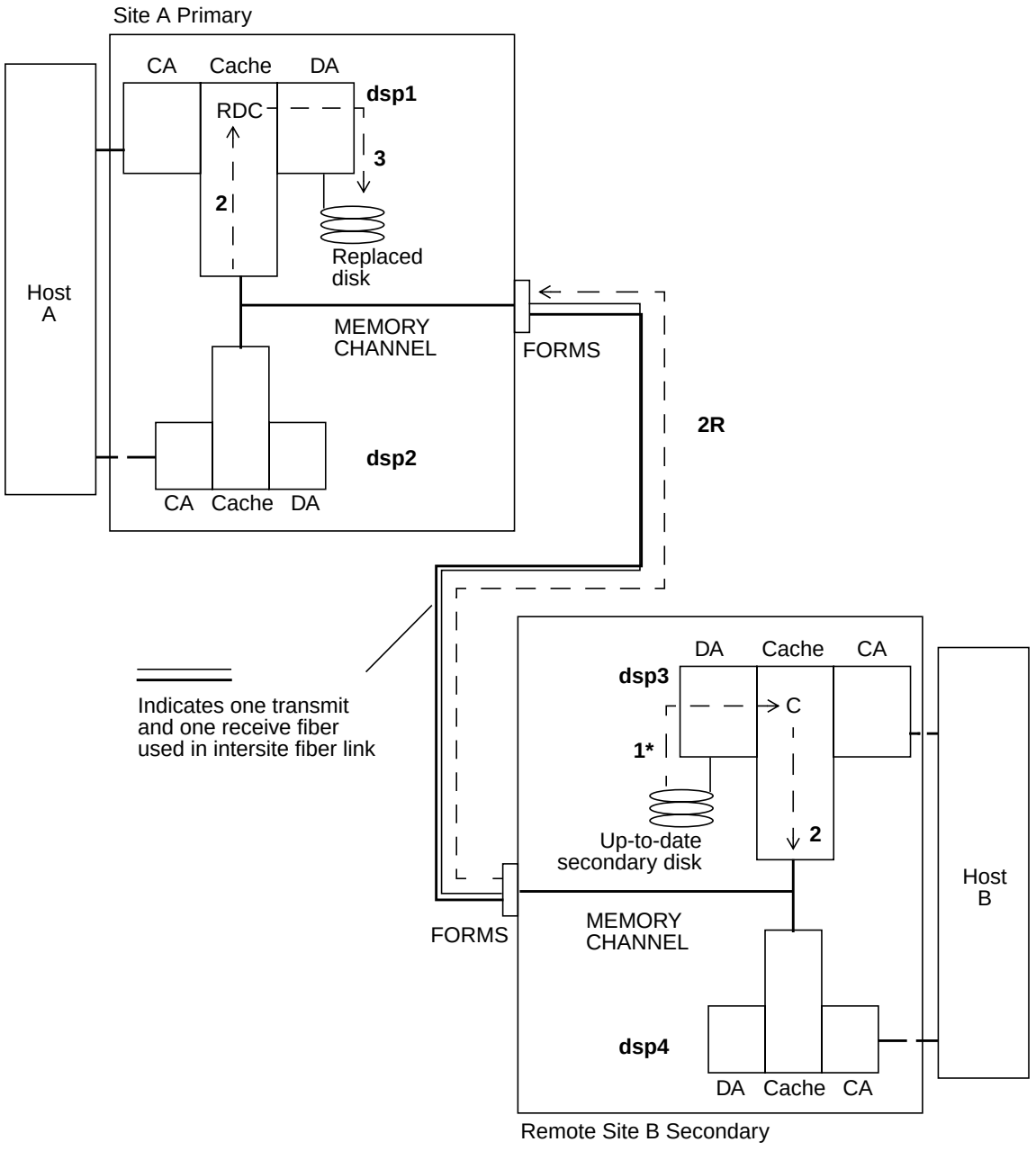

FIGURE 3-10 Reverse Full Synchronization

Recovering From a Primary Site Disaster

RDC minimizes the effects of a disaster at the primary site by keeping the secondary storage images up-to-date. Although the secondary RDC cache contains the latest writes issued on the primary before the disaster, that data may not have been destaged to the secondary disks yet. After detecting an interruption in the RDC service, RDC automatically destages the secondary RDC cache to its corresponding secondary volumes.

After all the secondary volumes have been updated with the latest RDC cache images, the secondary volumes may be varied online to the secondary hosts. The `sd_stats` screen displays confirm that destaging is complete. Run application-level recovery procedures to ensure a well-known state at the secondary site. The workload can then be switched to the secondary hosts for continued business operation.

Until the extent of the primary failure is understood, keep RDC enabled at the secondary site to track disk areas that are being modified. Under some confined disasters, the secondary update logs speed primary restoration, as described in “Completely Restoring a Primary A7000 from the Secondary” on page 3-27.

Completely Restoring a Primary A7000 from the Secondary

If the primary A7000 becomes inoperative and primary data on the primary disks is lost, update logs at the secondary subsystems have little value. You must perform a reverse full synchronization on the repaired or replaced primary A7000. In other words, volume-to-volume copies from the secondary to the primary are required for all RDC-managed volumes. This reverse synchronization process ensures that only the latest data is deposited on the primary disks. At the primary A7000 console, type:

```
dsp1# rdc_restore_primary_from_secondary
```

RDC allows active updates and queries on the primary volumes while the reverse synchronization is taking place. Primary read misses continue to be rerouted to the secondary device while updates are applied directly.

FIGURE 3-11 shows this restoration process for the primary storage platform. The procedure is similar to that of restoring a single primary disk, but a larger number of devices is included in the reverse resynchronization request.

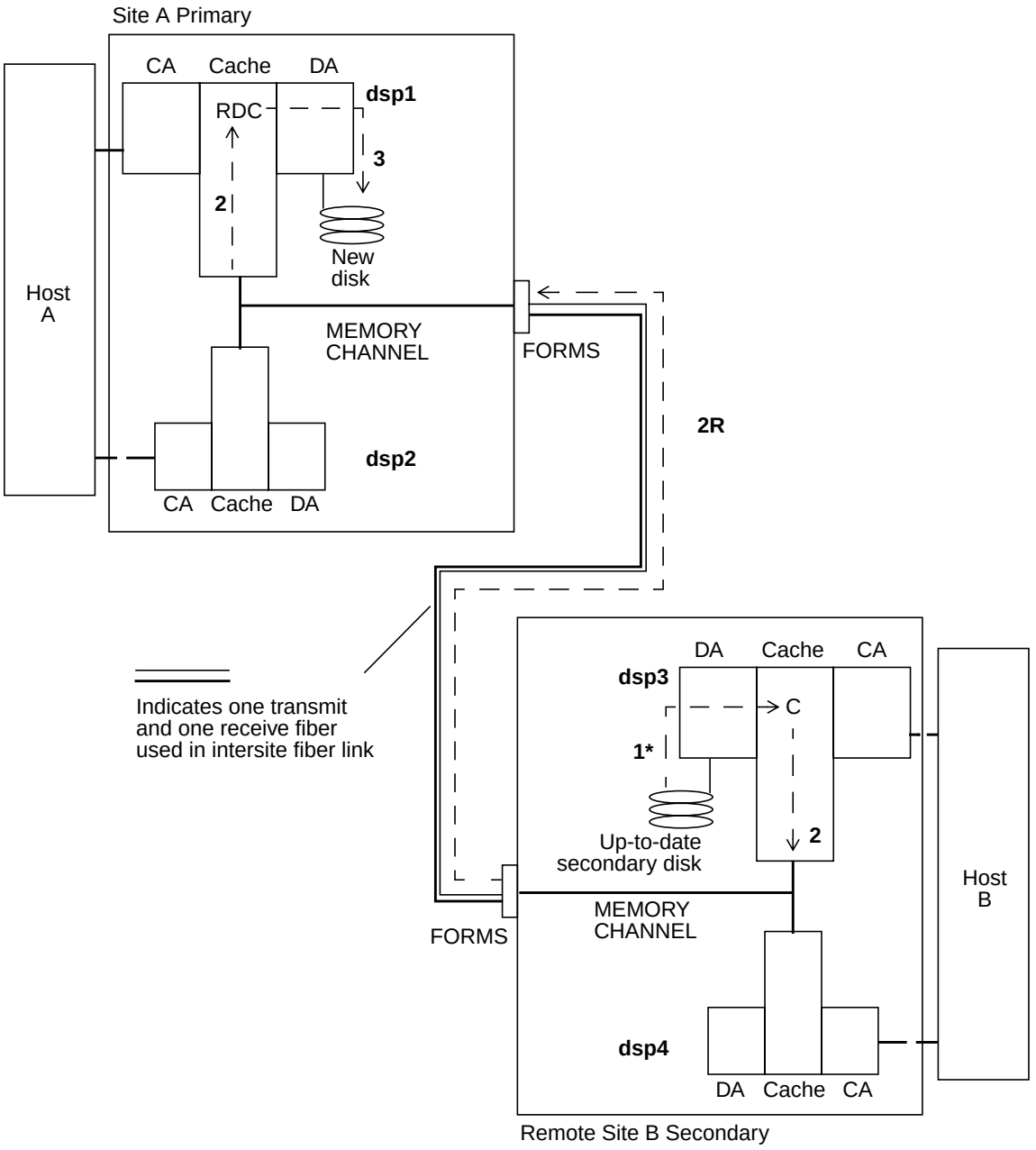

FIGURE 3-11 Restoring Primary from Secondary

Temporarily Suspending Remote Mirroring

Occasionally it is convenient to temporarily suspend remote mirroring between primary and secondary volumes and still keep track of any changes that occur during the suspension. Establishing point-in-time copies for backups or report generation are likely cases. The `rdc_suspend_mirroring` command provides such control at the logical volume level.

While suspended, RDC tracks new writes to both the primary and secondary volumes in their respective scoreboards. You can later perform an update resynchronization to restore remotely mirrored images and resume active dual copy operations.

The direction of resynchronization determines which updates are applied and which updates are overwritten. Typically, the latest contents are taken from the primary site volume, used to incrementally update the secondary volume, and replace any secondary writes that occurred during the suspended period. A reverse (refresh) resynchronization treats the secondary volume as having the desired contents and updates the primary volume accordingly.

- To suspend remote mirroring and activate logging, type:

```
dsp1# rdc_suspend_mirroring
```

- To resume remote mirroring and resynchronize the secondary volumes on dsp3 from the primary volumes on dsp1, type:

```
dsp1# rdc_resume_mirroring
```

- To reverse the resynchronization, using the secondary volume on dsp3 as the source to overwrite the primary volume, type:

```
dsp1# rdc_refresh_primary_from_secondary
```

- You can also resume remote mirroring by requesting a full volume copy against the suspended devices. To fully resynchronize the secondary volumes on dsp3 using dsp1 volumes as the primary, ignoring any incremental scoreboards, type:

```
dsp1# rdc_copy_primary_to_secondary -f config_file
```

- To resynchronize the primary volumes on dsp1 using dsp3 volumes as the source, ignoring any incremental scoreboards, type:

```
dsp1# rdc_restore_primary_from_secondary
```

Disabling Remote Mirroring

Unlike suspending RDC mirroring with the `rdc_suspend_mirroring` command, disabling RDC breaks the connection between primary and secondary volumes and discards any scoreboards.

Caution – Disable remote mirroring *only* when it is not necessary that the primary and secondary volumes be associated with each other any longer.

Following the RDC disable, a full synchronization (full volume copy) will be necessary to re-establish the RDC relationship and ensure that the volumes' contents match. See "Establishing Mirrored Images for the First Time" on page 3-3.

To permanently terminate RDC remote mirroring and RDC resynchronization services, type:

```
dsp1# rdc_disable_mirroring
```

Note – Use the `rdc_suspend_mirroring` command instead of the `rdc_disable_mirroring` command when you want to only temporarily suspend RDC services.

Updating the Primary Site From the Secondary Site

The primary volume state may be frozen as host workloads are switched to the secondary site. This is true in more realistic disaster rehearsals and may be true in confined disasters. Under these conditions, the secondary site activates its RDC logs (scoreboards) to track changes not yet seen at the primary. When it is time to bring the primary back into service, the scoreboards can be used to refresh the primary contents from the current secondary images. The following RDC command, typed at the primary A7000 subsystem console, starts the reverse update resynchronization from the secondary to the primary:

```
dsp1# rdc_refresh_primary_from_secondary
```

FIGURE 3-12 shows a reverse update resynchronization from the secondary subsystem to its primary subsystem.

1. RDC on dsp1 retrieves the secondary scoreboard from dsp3 for one of the RDC-managed volumes affected by the interruption.
2. RDC on dsp1 requests the blocks updated during the interruption from the up-to-date secondary volume of dsp3. The data may already be resident in dsp1 data cache, or it may require secondary disk access.
3. RDC on dsp3 transmits the updated blocks to dsp1 RDC-region of cache using the intersite fiber link.
4. RDC on dsp1 refreshes its stale image with the updated blocks.
5. RDC on dsp1 revises the scoreboard to track the remote update.

All steps repeat until the primary volume is up-to-date.

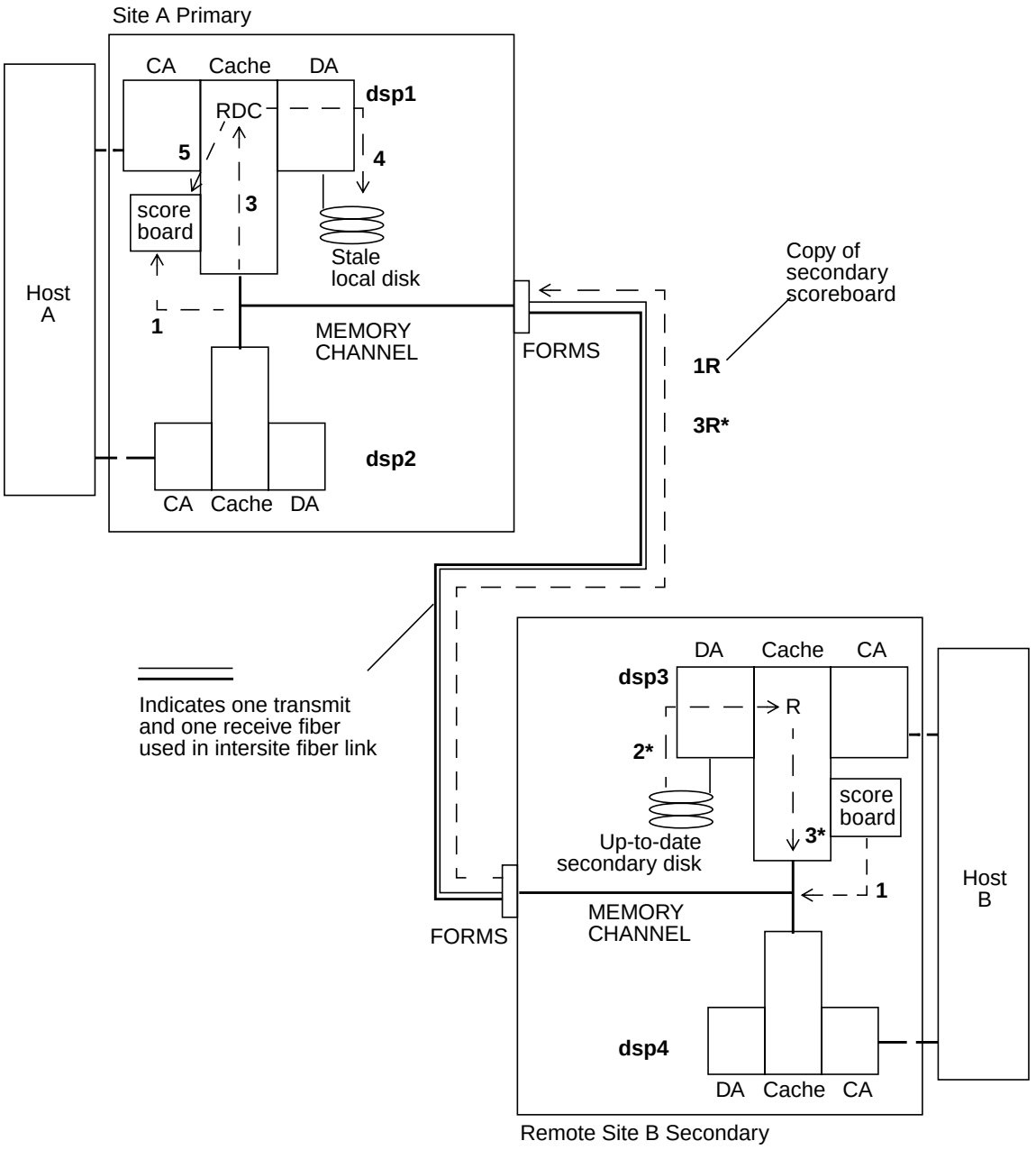

FIGURE 3-12 Reverse Update Synchronization

Mirrored Partitions for Quad Copies

You can combine Remote Dual Copy (RDC) with RAID 1 volumes to create quad copies of each device. Because RAID 1 mirrored partitions have separate administration utilities similar to those of RDC, take care to avoid redundant or unnecessary operations. FIGURE 3-13 shows RAID 1 protected RDC.

Mirrored Partitions

Mirrored partitions (RAID 1) are configured using the `mpadmin(1M)` utility and are completely separate from RDC. Refer to the `mpadmin(1M)` man page for a description of the utility.

- RDC is aware of disk failures only when all members of a mirrored partition fail on the primary or secondary. Single disk failures are transparent to RDC.
- Avoid mirrored partition synchronization operations on dual copy partitions except at initialization.

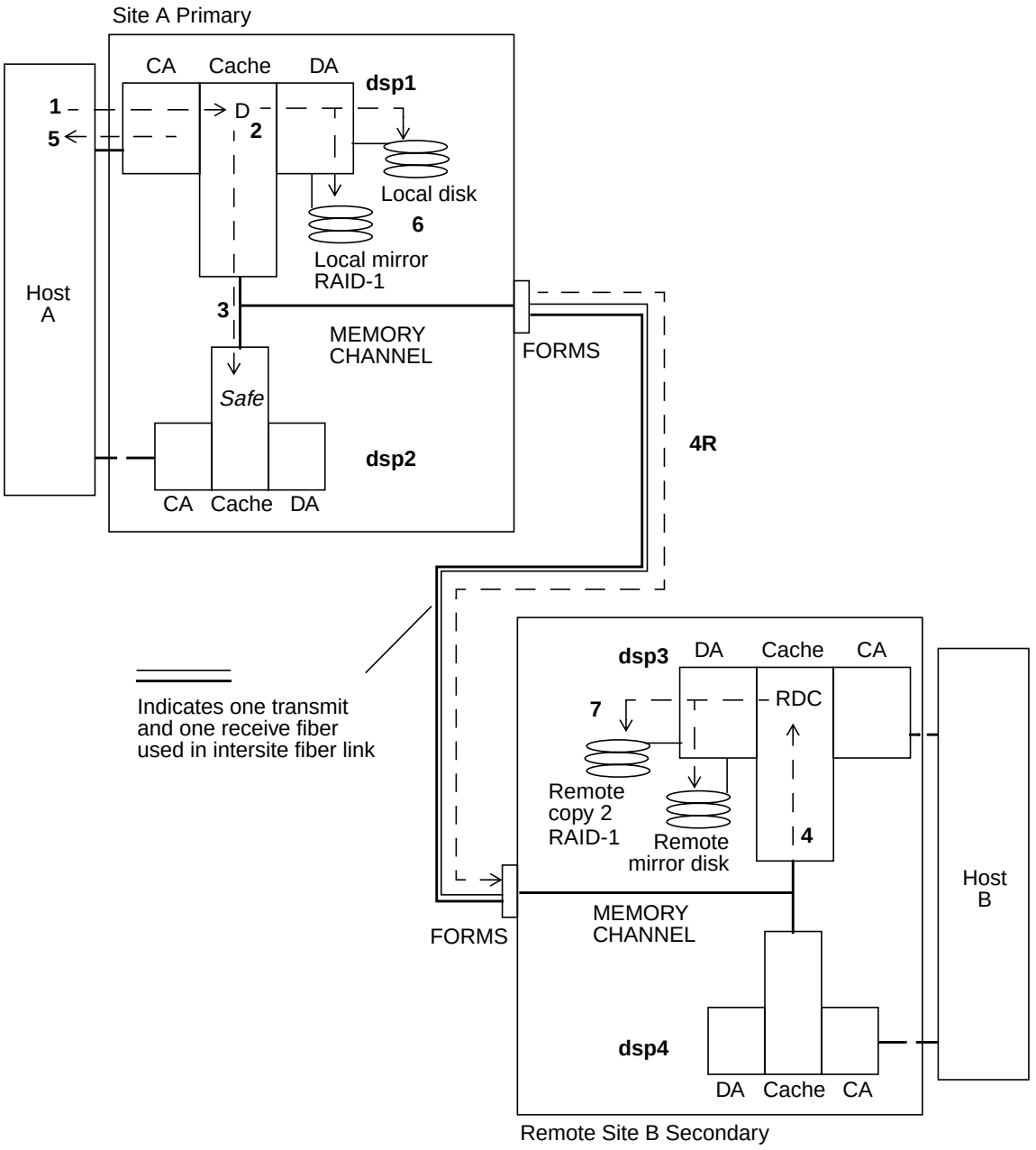

FIGURE 3-13 RAID 1-Protected RDC

Glossary

CKD Count-Key-Data.

disk Host view of the logical storage device. Disks are also referred to as volumes. In an A7000, several host volumes can be physically stored in a single Head-Disk Assembly (HDA). Alternatively, a single host volume can be spread (striped) across multiple HDAs. RDC options are performed on discrete logical volumes rather than on physical HDAs.

failback Process of restoring a failed A7000 subsystem to a fully operational state following failover.

failover Actions taken by the surviving adjacent subsystem in an A7000 configuration to ensure that all cached writes on the failed subsystem are properly destaged to disk. For mainframes, failover may also include switching of selected host channels for continued access to the failed subsystem's volumes using an alternate device path.

FORMS Fiber-Optic Reflective Memory System. FORMS is hardware that provides electro-optical conversion of the MEMORY CHANNEL bus for intersite transmission over fiber. FORMS includes a Transition Module Interface (TMI), a Fiber Transition Module (FTM), Arbitration/Termination Board (ATB), and a power supply.

FTM Fiber Transition Module. See *FORMS*.

heartbeat Health monitoring signal sent by each subsystem to its remote pair confirming that it is operational. The absence of a heartbeat signal is the first indication that the intersite link or remote subsystem is impaired.

HDA Head-Disk Assembly.

local mirrors Also known as mirrored partitions, RAID 1, and mp, which are established in a single A7000 subsystem and do not rely on RDC software or hardware to maintain local synchronization. See *mirrored partitions*.

logging Method for tracking disk updates that have not been remotely copied while the remote service is interrupted or impaired. The blocks that no longer match their remote pairs are identified for each volume. RDC uses these logs (scoreboards) to re-establish matching pairs through an optimized update synchronization rather than a complete volume-to-volume copy.

MEMORY CHANNEL Patented distributed shared memory interconnect between adjacent subsystems in an A7000. Also known as REFLECTIVE MEMORY SYSTEM (RMS).

mirrored partitions Multiple copies of disk partitions that provide RAID 1 protection against possible data loss from disk failure. Data can be read back from any of the members in the set and, in the event of a failure, a read is attempted on each set member until a successful read is accomplished.

mirroring Process of maintaining two or more identical images of a designated disk volume.

MVS Multiple Virtual Storage (mainframe operating system).

node Refer to *subsystem*.

partitions UNIX device names of the form `/dev/rdsk/`*`abc`* associated with the host disks on an A7000.

primary Subsystem or volume being accessed directly by the host application.

RAID Redundant Arrays of Independent Disks.

RDC Remote Dual Copy.

RDC log See *scoreboard*.

rollback synchronization Resynchronization operation that discards any blocks modified during recovery rehearsals.

scoreboard Special log that tracks hits to a disk area rather than a running log of each I/O event.

secondary Remote counterpart of a primary subsystem or volume where copies are destined.

source device Device used as the originating device in the remote copy.

subsystem Combination of storage hardware (channel adapters, device adapters, CPUs, cache memory) and software that is independently powered and configured in an A7000. Typical A7000 storage platforms comprise two adjacent subsystems working as a pair to service a large array of hosts and disks. Subsystems are sometimes referred to as nodes.

takeover rehearsal Simulation of a switchover to the secondary subsystem after a staged failure of the primary subsystem.

target device Device receiving the remote copy.

TMI Transition Module Interface. See *FORMS*.

update resynchronization Resynchronization operation that copies only the blocks modified since an RDC interruption.

virtual partition A logical device that functions identically to a normal physical disk partition, but whose relation to the physical disk partitions is determined from a software-defined mapping.

Virtual partitions contain one or more component disk partitions. The component disk partitions can be actual physical disk partitions or other virtual partitions. Virtual partitioning improves capacity and access speed efficiency.

VOLSER Volume serial number used by the MVS operating system.

volume See *disk*.

Index

S

T

U

V

W