Last products of Disk-File Development at Hursley and Millbrook

October 12, 1990

W J P Case

ID (Information Development)
MP 095 Ext 5524

IBM Internal use only

Contents

Redwing (0681)

First customer shipment (FCS)	April 1990
Size	Standard 5.25-inch footprint
Interface	SCSI
Capacity	857MB
Average access time	12ms
Tracks per inch (TPI)	1677
Rotational speed (RPM)	4986
Bits per inch (BPI)	30 320
Data channel	PRML (Partial response maximum likelihood)
Number of disks	12 thin film
Data rate	4.0MB per second

Redwing was first called Saffron, Osprey and then Kite. Work started in 1983 under John Heath with a very small development team. The first HDA was designed by Mike Hatchet. The first design was in a standard 5.25-inch footprint and included two actuators. The actuators were mounted at right angles to each other, with each actuator only having twelve heads so that each actuator only had heads for half the surfaces in the HDA. The heads initially were the same as on Kestrel. At that time there was concern about the amount of data accessible by a single actuator.

The first HDA was used to check the actuator tilt with varying temperature. The heads and preamplifiers were as on Kestrel. Special circuits were designed to write a correctly wrapped Piccolo-like servo track onto the top or bottom of one of the actuators. A Kestrel servo pattern was recorded onto all the other heads on this actuator. By selecting a particular head, a PES could be obtained to check how far that head was from its on-track position. This information and the selection of a head was done by a WHAT tester that was modified to have an IEEE interface. The position data measured by this system was put onto 5.25-inch disks, which were then read into API.2. Once in API., graphs were drawn of the offset against time as the environmental chamber was cycled through a variety of temperature excursions at various speeds.

The information from this showed how similar the temperatures of either end of the spindle had to be. At this time it was also thought possible that a "wedge" of servo information could be written onto each track position of all the heads. This wedge could have been used by a digital servo to offset the actuator, depending on the selected head. A disclosure was filed for moving the ends of the spindle, to achieve no added delay when changing heads. To keep the ends of the spindle the same was considered to be very difficult as one end of the spindle would be close to the electronic card and the other end might have been in free air.

This section was written by John Drummond as part of the 0681 Technical Reference manual.

Redwing Model 1000 and Model 500 disk drives are designed and manufactured to give quality, reliability, and performance in a standard 5.25-inch footprint. The Model 1000 with 12 disks offers the user over 850 megabytes of formatted storage capacity; the Model 500 with eight disks offers the user over 470 megabytes of formatted storage capacity.

Each drive is manufactured to stringent standards for quality and performance. All sensitive components are assembled in a rigidly controlled clean-room environment. Additionally, extensive quality control and testing procedures are used throughout the manufacturing process to achieve product reliability.

The standard features of Redwing Model 1000 and Model 500 are:

- SCSI industry-standard interface
- Conformity with ANSI X3.131-1986 and draft ANSI/SCSI-2 specifications (X3T9.2/86 - 109 revision 10 at the time of writing)
- SCSI disconnect and reconnect facility
- SCSI data transfer rate up to 4 megabytes per second (synchronous)
- Data transfer rate, to or from the disk surface, of 3 megabytes per second
- Disk speed of 4986 rpm (average latency − 6.02 milliseconds)
- Average seek time 11.2 milliseconds (mix of four reads to one write):
  - 10.8 milliseconds for reads
  - 12.8 milliseconds for writes
- Microprocessor control, integrated controller
- 128 kilobyte data buffer
- Look-ahead prefetches data into read buffer
- Linear voice-coil driven actuator
- Reserved head-landing zone
- Automatic actuator landing and locking in the landing zone on DASD power off
- Dedicated digital servo control
- Small head-change time
- Mini-monolithic low-profile heads
- Inline head suspension with 0° offset
- 20 data heads (Model 1000) or 11 data heads (Model 500)
- Thin-film medium
- Horizontal or vertical mounting
- Disk enclosure shock-mounted to frame

- Dynamic spindle-brake
- High MTBF (Mean time between failures)
- No preventive maintenance required
- Basic assurance tests (BATs)
- No external diagnostics required
- Internal error recovery including rereads, and error checking and correction (ECC) for data errors
- Internal track-defect management.

Product characteristics

DASD characteristics	Model 1000	Model 500
Disk diameter	5.25 inches	
Number of disks	12	8
Number of data heads	20	11
Number of servo heads	1	
Number of actuators	1	
Type of actuator	Linear	
Type of servo system	Dedicated digital	
Type of data channel	Partial response	
Linear density	30,320 BPI (bits per inch)	
Track density	1677 tracks per inch	
Areal density	50.9 megabits per square inch	
Unformatted DASD capacity		
Total number of cylinders	1460	
Sectors per track	58	
Gross track capacity	36 096 bytes	
Gross cylinder capacity	721 920 bytes	397 056 bytes
Gross DASD capacity	1 054 003 200 bytes	579 701 760 bytes
Formatted DASD capacity 512-byte sectors		
Cylinders available for user	1458	
Sectors per track	58	
User capacity per track	29 696 bytes	
User capacity of DASD	857 699 840 bytes	471 758 848 bytes
Formatted DASD capacity 520-byte sectors		
Cylinders available for user	1458	
Sectors per track	58	
User capacity per track	30 160 bytes	
User capacity of DASD	871 101 400 bytes	479 130 080 bytes

DASD performance

Data-transfer rate to or from disk	3.0 megabytes per second
SCSI bus data-transfer rate	4.0 megabytes per second maximum
Rotational speed	4986 rev per minute
Latency	6.02 milliseconds

	Read operations	**Write operations**
Single cylinder seek time	1.0 milliseconds	2.0 milliseconds
Average seek time	10.8 milliseconds	12.8 milliseconds
Full-stroke seek time	26.0 milliseconds	28.0 milliseconds

Motor start (to ready)	30 seconds typical
Motor stop	25 seconds

Physical characteristics

Height	82.55 mm
Width	146.0 mm
Depth	203.2 mm
Weight	3.5 kg

Reliability

Mean time between failures	> 150 000 powered on hours
Unrecoverable data errors	< 10 in 10^{14} bits read
Head positioning error rate	< 10 in 10^{7} seeks

Operating environment

Temperature range (operating ambient)	10°C to 40°C (50°F to 104°F)
Maximum wet bulb	27°C (81°F)
Relative humidity	8% to 80% non-condensing
Altitude	-304.8 to 3048 m (-1000 to 10 000 ft)

Functional description

Redwing Models 1000 and 500 have the following components:

- Read/write, control, servo, and SCSI interface electronics
- A sealed disk enclosure comprising:
  - A dc spindle motor
  - A track-positioning actuator
  - Read/write heads and disks
  - An air filtration system.

These components perform the following functions:

- Generate and interpret control signals
- Analyze a servo signal and position the heads over the desired track
- Read and write data
- Provide a contamination-free environment.

The disk enclosure is attached to the frame by four shock mounts. The electronics are mounted on a single card, which is attached directly to the frame.

Control electronics

The main functions of the single control electronics card are:

- Power-on sequence:
  1. Runs a power-on reset of the DASD
  2. Starts running the basic assurance tests (BATs)
  3. Accelerates the spindle to operating speed
  4. Calibrates the servo
  5. Recalibrates the DASD to track 0
  6. Completes the BATs
- Monitors various timers for factors such as head settle and servo failure
- Issues signals to control accessing, track following, and error recovery
- Controls the voice-coil-motor power-amplifier driver to drive the actuator to the desired position
- Monitors error conditions of the servo and acts appropriately if an error occurs
- Contains the velocity profile for the servo while accessing
- Provides digital compensation for the servo loop
- Keeps track of the actuator position and determines, from the logical block number, the target track for a seek operation
- Controls and interprets all interface signals between the using system and the DASD.

Disk enclosure

Spindle

The spindle is driven by an internal brushless dc motor.

Disks

There are twelve thin-film-coated disks attached to the spindle of a Model 1000 and eight in a Model 500. On both models, one disk surface is reserved for servo information. Twenty disk surfaces are used for data storage in a Model 1000; eleven disk surfaces in a Model 500.

Heads

Twenty data heads (Model 1000) or eleven data heads (Model 500), and one servo head are mounted on arms attached to the head carriage.

Actuator

A compact linear actuator drives the head carriage to the desired position. The head carriage is retained against the inner stop by a latch mechanism to prevent damage during shipment or replacement.

If power fails, there is sufficient energy in the spindle to provide, by regeneration, the power to drive the actuator to the inner stop. Crash stops are provided to minimize any damage to the disk enclosure if the servo loses control.

Contamination Control

The internal air circulation is driven by the ventilated disk stack. Air flows out from the spindle, across the disk surfaces, and is returned to the space above the top disk through two parallel filters. Pressure balance is maintained by a breather filter; positive pressure exists across all external seals. A labyrinth breather is fitted.

Disk format

The following section describes the allocation of cylinders, the size of data blocks, and the reassignment of data blocks.

Track format

Each track is divided into 58 physical sectors. One data block is assigned to each sector in sequence.

Cylinder allocation

Redwing Model 1000 has 1460 cylinders of 20 tracks each; the Model 500 has 1460 cylinders of 11 tracks each. The cylinders are allocated into areas, some of which are used for customer data, and others only by the DASD's internal controller. Cylinders are numbered from 0 at the outer diameter through 1459 at the inner diameter for specification only. The using system addresses the data areas by logical block addresses.

The cylinders are allocated as follows:

Customer data area: Cylinders 2 through 1459.

This area of the DASD can be accessed by the using system by use of commands across the SCSI interface. These cylinders can be formatted with a data block size of either 512 or 520 bytes. If a track-surface defect is encountered, the Reassign Block command can be used to reassign blocks within the data area.

Device support cylinders: Cylinders 0 and 1.

These cylinders contain information required by the DASD controller. These cylinders cannot be accessed directly from the using system by Read or Write commands. Some of the information held on these cylinders is read or written indirectly as a result of commands such as Mode Sense, Mode Select, and Inquiry. Because data blocks on these cylinders cannot be reassigned by the system, multiple copies of vital information are held on these cylinders. These cylinders contain:

- Manufacturing-defect list (P list)

  This list contains the locations of all the sectors containing surface defects at the time of manufacture. The list is used by manufacturing when originally formatting the DASD. It is never altered by the system or by the DASD controller.

  This list is empty at the time of manufacture. An entry, containing the location of the defective sector, is added to this list each time a block is reassigned.

  The P and G lists together identify all the currently defined defective sectors. The Format command can be used to format the DASD using either the P list or the P and G lists of defects.

- Microcode load

  A microcode load for the DASD controller is written on cylinder 0 when the DASD is manufactured; it is loaded from cylinder 0 every time the DASD is powered on. The using system should provide a utility to allow the microcode load held in the reserved area to be rewritten in the field with a later level of microcode. This is done using the Write Buffer command with the "download and save" option.

- Vital product data (VPD)

  This data is product information that is written during the manufacturing process. The VPD for Redwing contains the following:

  - Product type
  - Model number
  - Serial number
  - PROM Microcode level
  - RAM Microcode level
  - DE part number
  - DE EC level
  - Card part number

- Controller variables

  This is vital information that must be retained when power is off, including:

  - Data block size (512 or 520 bytes).
  - Mode-select saveable parameters.
  - The operation that was in progress when the DASD was powered off if that operation must be resumed; for example, a Format or Reassign Block command that has not completed.
  - DASD configuration variables.

  This is data that is moved in the course of a block reassignment.

- Read/Write test area

  This area is used during basic assurance tests (BATs) to verify that read and write operations are working correctly on all heads.

Data block sizes

In the customer data area, data block sizes of 512 bytes and 520 bytes are supported. The using system can select the data block size by use of the Mode Select command. During manufacture, this area is formatted with a data block size of 512 bytes. Only interleave factors of 0 or 1, both of which specify no interleave, are supported.

When shipped from manufacturing, the data field of each sector in the customer data area is written with a data pattern that can be read in either 512 or 520 byte format, but the block size is set to 512 bytes.

The reserved area is formatted with a data block size of 512 bytes. The using system cannot select a different data block size for this area.

Data block reassignment

When the DASD is manufactured, and following a Format Unit command, data blocks are assigned to sectors sequentially, skipping over defective sectors.

Eleven sectors are allocated for spares at the end of each cylinder on a Model 1000; six sectors on a Model 500. Some of these sectors may be used for blocks that were reassigned because defective sectors were detected when the DASD was manufactured.

In the customer data area, blocks can be reassigned by a Reassign Block command from the using system.

Reassignment consists of moving the data block from the sector containing the defect to the next sector and pushing down all subsequent blocks on the cylinder as far as the next spare sector. The physical sequential ordering of logical blocks is therefore maintained. When all the spare sectors on a cylinder have been used, another reassignment on that cylinder causes a block to be moved to the adjacent cylinder and the blocks on adjacent cylinders to be pushed down as far as the next spare sector. This could be up to three cylinders away.

If a spare sector is not available within three cylinders, the Reassign Block command terminates with a check condition status and a sense of "no defect spare location available."

Controller functions

The following paragraphs describe some functions of the controller in Redwing.

Read-buffer look-ahead

After a read operation, the data blocks following the last block requested by the Read command are read into the data buffer until the buffer is full. If data requested by a following Read command has been prefetched into the buffer, it is transferred to the initiator without the delay needed for the DASD to rotate to the required data block.

When an SCSI command is received from an initiator during a read-ahead operation, the response varies with the command:

- When the SCSI command is an Inquiry or Request Sense command, the read-ahead operation is continued while the command is processed.
- When the SCSI command is a Read command and the data identified by the starting logical block address (LBA) is in the buffer, Redwing:
  - Processes the Read command by transferring the data from the buffer that had been prefetched by a previous read-ahead operation.

  - Continues with a new read-ahead operation to read data from the disk that has not already been buffered and that is required for the Read command, in addition to data for the next read-ahead operation.
  - Completes the next read-ahead operation by filling the data buffer with the next sequential data.
- When the SCSI command is a Read command, and the data identified by the LBA is not in the buffer, or if the command is any other command than those defined above, the Redwing:
  - Ends the read-ahead operation.
  - Purges the data buffer of the data fetched during a preceding read-ahead operation, but only if the new command requires the use of the buffer.
  - Processes the SCSI command.

Idle-time functions

If Redwing is inactive for 40 seconds, the actuator arm is moved to the inner cylinder and then to the outer cylinder, and then to a random location in the high flying height area of the disk (outer cylinder area). If Redwing is inactive in the high flying area for nine minutes, it is randomly positioned to a new location in this area. If a command is received during this idle-time movement of the actuator arm, this function is ended and the command is processed.

Servo-code architecture

This subsection was written by Pete Baker.

System model

The model for the servo control of the Redwing file is based on the Lee file. It is a third-order state variable description of the actuator system. The model does not include:

- Friction
- Coil inductance and effect of shorted turn
- Resonances.

The system equations are as follows

$$\dot{x}_1(t) = x_2(t)$$

$$\dot{x}_2(t) = nx_3(t) + nu(t)$$

$$\dot{x}_3(t) = 0$$

x_1 = position error signal (PES)
x_2 = velocity, meters per second
x_3 = bias, amps
u = control, amps
$n = \frac{K_f}{Mass}$

where K_f = force constant, Newtons per amp
Mass is in kilograms

Figure 1. System model for Redwing actuator

The above state equations are converted to discrete time with a sampling period of 64 μs (sector time) and then normalized such that position is in units of tracks, and velocity is in tracks/sector time. The system equations become

$$x(k+1) = \Phi x(k) + \Gamma u(k)$$

where

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

x_4 = integrated PES, and k is discrete time variable.

The matrixes Φ and Γ are

$$\Phi = \begin{bmatrix} 1 & 1 & \Gamma_1 & 0 \\ 0 & 1 & \Gamma_2 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ 0 \\ 0 \end{bmatrix}$$

$$\Gamma_1 = \frac{K_f \, Tpm \, T^2}{2 \, Mass}$$

$$\Gamma_2 = 2\,\Gamma_1$$

Tpm = track pitch, tracks per metre
T = sector time, 64 μs

Definition of algorithms

Track-follow

Controller

: Because bias is an uncontrollable state, it is not fed back. The control law is taken as a weighting of the state variables x_1, x_2, and x_4, that is:

$$u(k) = -\,[\,K_1 x_1(k) + K_2 \hat{x}_2(k) + K_4 x_4(k)\,]$$

where $\hat{x}_2(k)$ is the estimate of $x_2(k)$

The feedback gains K_1, K_2, and K_4 are chosen by frequency response methods whereby the discrete system model is translated to the frequency domain, and the feedback gains are adjusted for maximum phase margin at zero crossover frequency and maximum open loop gain at the spindle frequency of 83.1 Hz.

Estimator

: Redwing uses a reduced-order current estimator to generate the state variables, that is only velocity and bias are estimated. The PES gives x_1 directly. This reduces the number of calculations required, but with increased sensitivity to noise in the PES signal. If time had been available, a full-order estimator would have been implemented, although it was thought that any significant reduction in sensitivity to noise would have an unacceptable phase delay. It is a "current" estimator because it uses the current input to update the estimates used in generating the output.

Because integrated position is not observable, it is not included in the estimator. The Φ and Γ matrixes for the estimator design then reduce to:

$$\Phi_e = \begin{bmatrix} 1 & 1 & \Gamma_1 \\ 0 & 1 & \Gamma_2 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\Gamma_e = \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ 0 \end{bmatrix}$$

The resulting estimator equations, after going through the algebra, are:

$$\hat{x}_2(k+1) = \bar{x}_2(k+1) + L_2 Residual(k+1)$$

$$\hat{x}_3(k+1) = \hat{x}_3(k) + L_3 Residual(k+1)$$

$$\bar{x}_2(k+1) = \hat{x}_2(k) + \Gamma_2 \hat{x}_3(k) + \Gamma_2 u(k)$$

The $Residual(k+1)$ is calculated as follows:

$$Residual(k+1) = x_1(k+1) - \bar{x}_1(k+1)$$

The "hat" indicates that the variable is estimated, and the "bar" indicates that the variable is projected. The state x_1 is PES.

Filters

Two biquadratic filters of the following form are used to reduce sensitivity to mechanical resonances:

$$y(k) = a_1 y(k-1) + a_2 y(k-2) + b_0 u(k) + b_1 u(k-1) + b_2 u(k-2)$$

giving a Z transform of

$$\frac{Y(z)}{U(z)} = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 - a_1 z^{-1} - a_2 z^{-2}}$$

They are both implemented as elliptic filters, one centred at 4500 Hz and the other at 6300 Hz.

Seek

Controller: The seek function has two different implementations, dependent on whether the initial distance-to-go is:

- Greater than 7 tracks, in which case a 2- or 3-phase seek is implemented
- 7 tracks or less, when a position loop is used.

Both implementations hand over to a modified track-follow routine when they fulfil a criterion based on remaining distance to go and velocity.

The first seek algorithm consists of:

1. An accelerate phase – where maximum current is applied until such time as either:
   - The maximum velocity (1.44 metres per second) is reached and the coast phase is entered

   or
   - The remaining distance to go demands that the decelerate current is applied.
2. A coast phase – where the actuator velocity is kept at maximum
3. A decelerate phase – where a current is applied to bring the actuator to rest at the required track. The current required is recalculated every sector.

The second seek algorithm is essentially an extended track-follow algorithm where the position-error signal is replaced by the tracks to go. The integrator is switched out and a fixed value of integrated PES is used.

For both algorithms, the seek part of the access is exited when either:

- The distance to go is less than 0.25 tracks
- The velocity is less than 0.125 tracks per sector, and the distance to go is less than 0.5 tracks.

The seek algorithms include code to:

- Check that maximum velocity is not exceeded by more than 1 track per sector; should this arise the seek is aborted, the actuator retracted, and status posted.
- Check that the PES is valid, that is, bad coherence and error bits are not set. If this is not the case the prediction of the PES is substituted for the PES, and the Bad Sector flag set.
- Check that the difference between the PES and the predicted PES is less than 1 track. If this check fails, the prediction is substituted for the PES, and the Bad Sector flag set.
- Check whether the demodulator is out of synch or the guardband area has been entered. If either of these occur, the seek is aborted, the actuator retracted, and status posted.

Track-follow

The track-follow compensator is designed to give an open loop gain of 0 dB at 600 Hz with enough phase margin to maintain a closed-loop gain rise of less than 5 dB. An integrator on the PES reduces the dc PES error to zero.

The track-follow algorithm includes code to:

- Check that the PES is valid; that is, bad coherence and error bits are not set. If this is not the case, the prediction of the PES is substituted for the PES, and the Bad Sector flag set.
- Check that the difference between the PES and the predicted PES is less than 0.25 track. If this check fails, the prediction is substituted for the PES and the Bad Sector flag set.
- Check whether the demodulator is out of synch or the guardband area has been entered. If either of these occur, track-follow is aborted and the actuator retracted.
- Check that the head is within 10% of track centre. Failure to meet this criterion causes the write-inhibit line to be activated and a condition status posted.
- Check that the head is within 50% of the track centre. Failure to meet this criterion causes the write-inhibit line to be activated and the actuator to retract. A condition status is posted.

Motor speed control

Written and originally designed by Bill Case.

When development first started of Saffron, four disk drives were mounted in a Kite drawer. At that time it was anticipated that all of the spindles would be synchronized very precisely to enable Count Key Data (CKD) to be used over the four spindles. To achieve that, a loop was designed by Bill Case, which closed the spindle speed control loop at 30 hz. The loop had to include an integrator to keep the rotational position of each drive accurately positioned to a common index. A measurement of each spindle's rotary position was invented by Eric Newman and the output from this was used as input to the speed control loop. The motor which was first used was a standard brusshless dc motor, with Hall effect devices used to commutate the windings. The need for Hall effect devices was ended when Lawrence Wright invented a way of controlling the commutation without the need for the Hall effect devices.

Procedure used for Motor Speed Control (original Saffron)

At Start Up

Start the Motor turning by using the Hall Effect devices alone. The SID will start to appear when the speed is about 2% below the final correct speed. The time between SIDs is measured (the change in SID times is less than one reset of the SID error logic per SID), when the speed is measured as being within 1% of the final value, the controller can be activated. The control will thereafter be via the PWM into the motor driver.

Normal Running

The time between SIDs is measured for each SID time, however when the control algoritm is started it works on the result after every 45 degrees (ie. 8 per revolution), which is produced by accumulating the individual SID time error over the 45 degrees.

```
The controller algorithms are as follows:

      CMPOP2 = QTIM2 - 1.908*PQTIM1 + .9094*PQTIM2
                     + 1.338*PCOP12 - .3381*PCOP22

      CMPOP1 = QTIM2 - .9344*PQTIM1 + .3381*PCOP11


     Where :-     QTIM2 is the error in the output of L. Wright's
                   SID timer logic or the processor, after 45 degrees.

                  PQTIM1 is the value of QTIM2 at the previous sample.

                  PQTIM2 is the value of QTIM2 2 samples earlier.

                  PCOP12 is the value of CMPOP2 at the previous sample.

                  PCOP22 is the value of CMPOP2 2 samples earlier.

                  PCOP11 is the value of CMPOP1 at the previous sample.

           The coefficients can be expressed in 15 bits plus a sign bit.
```

Before the algorithms are activated the variables PQTIM1,PQTIM2, PCOP12, PCOP22 and PCOP11 are made zero.

During the start up CMPOP1 is used for the first 2.3 seconds and CMPOP2 is made equal to CMPOP1. During this time the processor is used to accumulate the error to be applied to the controller. After this time L.Wright's logic is used to do the accumulation over 45 degrees.

While the algorithms are active the selected output is multiplied by .0115 (.0155 for 20 Mhz clock) and then has a value of .854 added to it (this is the nominal value required by the PWM cicuit.). The output from this summation is constrained to stay between .75 and 1.0 before it is applied to the PWM circuit.

This value '.854' could be adaptively modified by the code to try and get the average value of CMPOP2 to be zero.

This would allow any bearings in the spindle to be tolerated.

```
During start up the maximum value after the multiplication and addition is 43
                the minimum value after the multiplication and addition is -25

   These values can be represented by 13 bits and 1 sign bit.

   The result of multiplying these values by the coefficients will
    results in 28 bits and 1 sign bit.

   The actual maximum in calculation is 1.908 x MAX. or 1.908 x MIN. that is +82 and -48
   This value can be represented by 11 bits and 1 sign bit.

   The 28 (+1) bit answer may thus be represented by shifting the 32 bit
    answer by 3 bits to the left and only thereafter using the most
    significant 15 (+1) bits of the answer.

The maximun output of the algorithm is between .75 and 1 (an 8 bit
 number) which controls the PWM generator.
   This can be taken by first checking for a positive value, shifting
    left 3 bit and then only taking the top 8 bits as input to the
    PWM circuit.

After the calculations the variables must be rotated as follows:

       PCOP11=CMPOP1
       PCOP22=PCOP12
       PCOP12=CMPOP2
       PQTIM2=PQTIM1
       PQTIM1=QTIM2
```

With this high bandwidth a need was for very high current pulses.

The size of these pulses was unacceptable to the SCSI interface which was by then decided to be the interface to be used by Redwing (the name that was current by that time). The bandwidth was therefore reduced to reduce the size of the current pulses. As there was no need for spindle synchronisation on the SCSI interface, the integrator was removed from the servo loop.

Wear on Osprey

This subsection was written by Bill Case and Edward Prager.

- Investigation of bearing runout and its effects on the Osprey actuator.

  Some initial investigation was made into the relative effects of runout, radial play, and outer ring out-of-round on the tilt of the actuator carriage. The runout was measured using a talysurf and a rig for holding a single bearing and turning it slowly using a plastic band from a small electric motor. Radial play was measured by measuring the axial play of the inner bearing that measured the change between the position of the inner bearing when it was first upwardly and then downwardly preloaded, with the outer ring supported horizontally on the measuring surface.

- Design of a simple rig to measure bearing runout, (this proved to be quite inaccurate, but it gave a good estimate of the bearing quality).

- Measurement of bearing out-of-round and radial play (from the axial play).

- Construction of a rig to run the actuator in isolation, which required programmable logic arrays. PLA blowing.

  The rig relied on a crude feedback signal given by a light-sensitive switch. A mica strip attached to the carriage broke the light beam when the carriage passed the centre of the stroke. This caused the direction of current in the coil to change, slowing the actuator and accelerating it in the opposite direction. The delay between break of the beam and current direction change could be altered to produce different access lengths.

  Bearing slippage was reviewed as follows:
  One of the bearings had a white mark painted onto a side. This bearing was photographed with a high-speed video camera. After a series of accesses the position of the painted line was noted and compared with the starting position of this line.

- Preparation of an "idiots" guide to PLA design and blowing.

- The most extensive life testing of a linear actuator that has ever been run.
  17 actuators underwent 1 000 000 000 accesses each.

  After a short test time it became apparent that the ceramic rails being used in the actuator had a considerably greater resistance to wear than steel rails which had previously been used. As a consequence, a full life test of the ceramic rails was started. These tests were run, and by measuring the wear to the bearings and rails using a talysurf, the distributions of actuator tilts caused by bearing and rail wear which would occur in the file were calculated. Thus an estimation of the wear contribution to TMR could be made long before fully operational files were built.

- Continued development of the ceramic guiderails.

  Tested several ceramic rails from different manufacturers, and by talking to the manufacturers and with information from the materials laboratories in Hursley and Havant, a materials spec was produced which controls the attributes of the material that are believed to be critical for its very high resistance to wear.

- Production of a theory of bearing skewing as the major contributor to unrepeatability of the actuator.

The Redwing actuator design is very sensitive to any change in the bearing profiles, because it has a high head stack in proportion to the length of its wheelbase. Three different ways were derived in which the outer ring of the bearing could move to cause tilt. The team then analyzed under what conditions each movement would occur, and changed the carriage and bearing designs to minimize this effect.

Data channel

This uses the new Partial Response Maximum Likelihood (PRML) channel designed in Rochester, based on the work done in the Zurich research laboratory. This channel is fully documented and a copy may be obtained from the Data channel staff in Havant.

SAT

This subsection was written by Chris Ward.

The surface-analysis testing for the Redwing file takes 20 minutes for 850 megabytes. It needs no electronics apart from an ordinary Redwing file card. The only involvement with a controller is for download of microcode at the start of the test and retrieval of summary data on completion. The number of defects missed is so small that there is no need for a process to detect and reassign defective sectors at Extended Run.

This report explains how it has been done, so it can be repeated on future files.

Design features

Several innovative hardware features made this surface analysis testing possible. Some of these features were a natural part of the Redwing file design; others were icluded at an early stage especially for this manufacturing process.

These special features do not increase the hardware cost or part count for the file. In all cases, they use cells in custom logic chips that would otherwise have been spare.

- Recording Subsystem
- Data Encoder/Decoder
- Formatter
- Sector Mark Generator
- Data Integrity Interlock Overrides
- File Control Microprocessor
- Servo Processor.

Theory of test

Signals, noise, and the recording subsystem

: In the Redwing file, the analogue signal from the readback heads is passed through filters and amplifiers into a digitizer. At each bit time, the digitizer creates a 6-bit binary number on a scale from −32 to +31.

Digital gain, timing, and equalization control loops are used so that customer data bits ideally correspond to −16, 0, or +16 on this scale.

Deviations from these ideal values can be caused by a number of effects, such as:

- Surface defects
- Electrical noise
- Adjacent track reading
- Write-over modulation
- Imperfect equalization
- Gain and timing errors during closed-loop control
- Magnetic-domain effects in the head and disk.

These effects are collectively called noise.

If there is no noise in the system, the data decoder decodes every bit correctly. But as the total noise increases, so does the mis-decode rate.

Surface analysis is the manufacturing process that controls the contribution of surface defects to the total noise. It locates sites on the disk surface that give an unacceptably large contribution, and arranges that these sites are never used for recording customer data.

To isolate the effect of surface defects, the other effects must be minimized and the recording subsystem must be made more sensitive to noise.

- Electrical noise

  Electrical noise in the recording subsystem degrades the file soft-error rate in normal operation. The test equipment designer can assume that the product designer minimises electrical noise as part of his normal development.

- Adjacent track reading

  The file is dc-erased during the first phase of the surface test, so that there are no signals on adjacent tracks to be read.

  The servo can place the head half-way between tracks to do this, to ensure that no residual signal remains in this region.

- Write-over modulation

  The maximum possible write current is used for the surface test. It does not matter that this causes the head to write wide, as there is no data on adjacent tracks.

  The dc-erase helps minimize this effect too.

- Imperfect equalization

  The surface test uses a data pattern that reads back as a single-frequency sine wave. Misequalization only affects mixed frequency patterns, so does not affect the surface test.

- Gain and timing errors

  Gain and timing errors are the means by which closed-loop control is maintained. The design of the recording subsystem is such that these errors are small, because these errors contribute to the file soft-error rate. So the test-equipment designer does not have to be concerned with these.

- Magnetic-domain effects.

  Again, these contribute to the file soft-error rate, so other people control these effects.

The recording subsystem in surface analysis mode

: The test involves writing a sequence of 2-bit-long magnets. This reads back as a sine wave; and the gain and timing loops controls things so that the digitized readback signal is ideally (+16,+16,−16,−16) repeated.

The recording subsystem processes the odd bits and even bits through two identical separate decoders. In the ideal case, each of these decoders reads a stream of numbers (+16,−16,+16,−16) and so on. Each decoder subtracts the previous number from the current number, giving (+32,−32,+32,−32); makes the result positive, giving (+32,+32,+32,+32); and subtracts the result from 32, giving (0,0,0,0).

The signal in the ideal case is shown below.

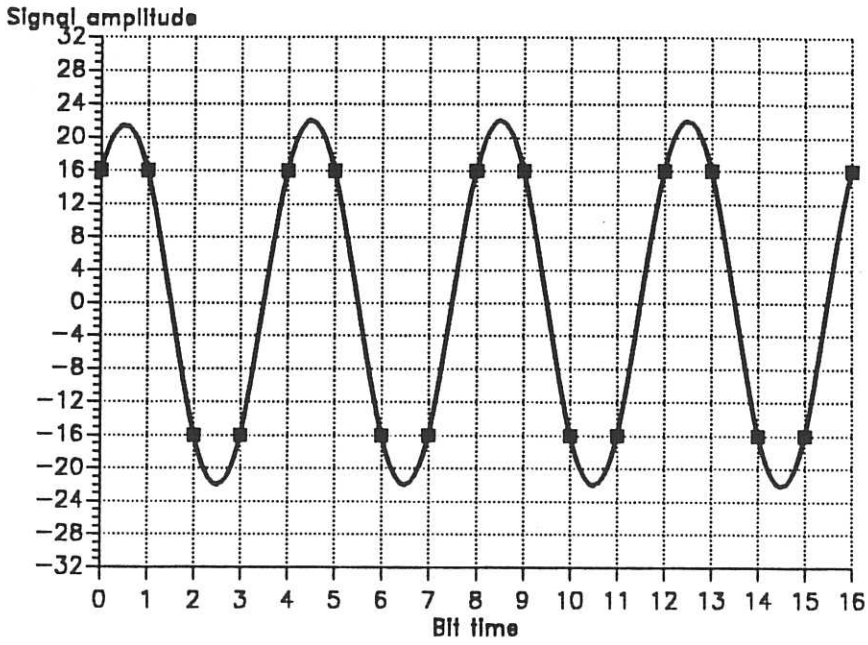

Figure 2. Head-readback waveform of good signal

If there is a disk-surface defect, this reduces the amplitude of the signal at one or more bit times, and when processed through the decoders results in numbers that are accordingly nonzero.

The signal for a typical defect is shown below.

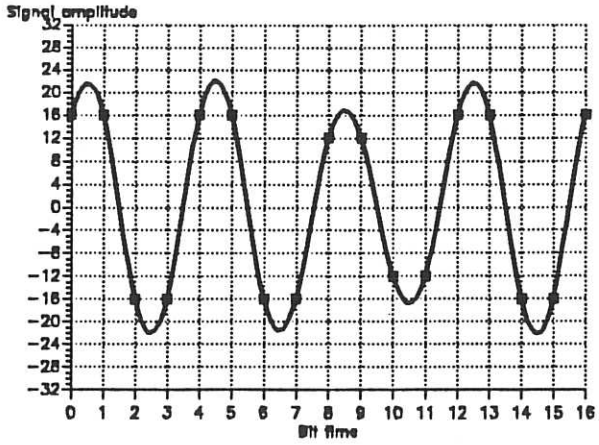

Figure 3. Head-readback waveform of missing bit signal

The output numbers of the decoders are compared at each bit time with a threshold number programmed into the recording subsystem; and the final output is a stream of mostly '1' bits for numbers that pass the comparison, with occasional '0' bits where the numbers fail the comparison, indicating a surface defect.

Both for writing and for reading, the high-speed bit stream passes through the data path that would normally be used for customer data. The only differences noticeable outside the recording subsystem chip are the unusual data patterns of (00110011) for writing and (11111111) for reading.

The data encoder/decoder in surface analysis mode

: In normal operation, the recording subsystem is not capable of storing arbitrary binary data. Some patterns give insufficient signals on the disk for the control loops to operate, and some patterns overflow the Viterbi decoder shift register in the recording subsystem.

To overcome this, a 'Constraint Code' is used. It has been found that of the 512 possible 9-bit patterns, 279 are allowed by the recording subsystem.

When writing, the data encoder takes each 8-bit unit of customer data, codes it to one of 256 of these patterns, and sends it to the recording subsystem in 9-bit times (≡ 1.125 bytes). The recording subsystem uses a 27 MHz clock to support the 3MB data rate from the formatter. The remaining 23 of the 279 allowable patterns are not used.

Following the encoder, there is a Write Precoder. This is a 2-bit shift register with exclusive-OR feedback, which compensates for the head − disk transfer function.

When writing in surface-analysis mode, the encoder is made to produce either all zeros or all ones at its output; neither of these patterns correspond to customer data input.

This, and the ability to preset the precoder shift register, gives the ability to write dc in either polarity, 1-bit-length magnets (which also erase), or 2-bit-length magnets (which are the surface test pattern) in any of the four possible phases.

When reading in surface-analysis mode, the decoder decodes a (9-bit) byte of all 1's to an (8-bit) byte of zero, and decodes all other patterns to nonzero values.

Bit and byte synchronization
: At the start of each ID or data field, patterns are recorded that enable the control loops to lock in, determine byte boundaries, and determine which is the first data byte.

When surface-analysis mode is selected, the recording subsystem and the encoder/decoder process these synchronization fields in the normal way, and switch over to the surface-test function at completion of the synchronization fields.

The formatter
: The formatter has byte-by-byte control of operations while writing each sector. It determines when write gate and read gate will be asserted, it verifies ID fields, it routes data between the disk and RAM buffers, and it generates and checks CRC and ECC.

In normal operation, the file has identification (ID) fields and data fields. The ID field is a 6-byte field containing the logical block number of the following data field, and miscellaneous other flags. After the data field is a relatively long gap; if the file is instructed to do a multiblock write, it needs this time to turn round the recording subsystem to verify the ID fields between blocks.

The formatter in Redwing is a 10C00, which is controlled by a 32-word microprogram. For surface-test purposes, this allows flexibility to use a single long block per sector, covering the ID and data fields and their synchronization regions. The synchronization for this long block can be written in the gap after a normal data field; there is no need for a long gap because there is no need to read an ID field between blocks.

When reading a surface-test block, the formatter is arranged to compare incoming data with binary zeros. If it detects an error, it stops and interrupts the file microprocessor at the end of the block and indicates 'data-verify error'.

If the file control microprocessor can restart the formatter in a predictable time, testing can continue.

For Redwing, the file-control microprocessor restarts the formatter after 0.5 sectors, with a head switch if required; Redwing has 20 heads, so the basic unit of test is 1 cylinder, which can be tested in 20 revs + 20 sectors. The sector immediately following each defective sector is not tested in the first pass.

After most of the cylinder is tested, a different formatter microsequence is used to test the remaining sectors.

Sector-mark generator
: The Redwing sector-mark generator is consistent with the servo-mark demodulator. It generates a programmably recurring pulse, starting a programmable time after Index.

This is used to start surface test sectors earlier than the start of the ID field for normal sectors.

Data integrity interlocks
: The Redwing file assures data integrity by preventing writing when certain abnormal conditions exist, for instance:

- Servo off-track
- No transitions in the head
- Attempted write through index.

These interlocks can all be overridden by software; this is necessary for the surface-test operations that have been described above.

File control microprocessor
: The FCM has an EPROM for bootstrap, and a RAM for operating code. It can load the RAM from the file (normal operation), or from its SCSI interface (manufacturing and diagnostic use).

It has access to a 256KB data buffer (which is intended for data in transit between the file and a using system).

Using the RAM as program memory and for small quantities of critical data, and the buffer for large quantities of data such as the defect map and various sorting trees, it is possible on Redwing to download and run the whole surface test algorithm on the file card. This includes:

- Pre-erase
- Write and test
- Scratch algorithm
- Reserved area formatting
- Defect map writing
- Other reserved area setup
- Summary data generation
- Tracing and diagnostics.

The Redwing FCM wa programmed mostly in 'C', with several time-critical routines in 8051 Assembler. All programming tools run on a PC or PS/2, and code can be put on diskette or uploaded to VM for sharing or safekeeping. This allows fast and relatively informal code development on a changing base; the surface-test microcode was built in parallel with the product microcode provided by the product programming group.

Servo processor
: The Redwing servo is fully digital. It can be instructed to place the head at any position with respect to track centre, in 1/256 track steps. This is used to do the half-track erase.

Options taken, options available

In designing a surface-analysis test, there are many options. The objective of the test is to provide adequate protection against hard and soft errors; the rate of test escapes can be estimated statistically from extended run data during development evaluation and manufacturing process. Twelve hours at the full manufacturing going rate normally provides more data than does months or years of development scale work; but that is always too late for tester design.

Evaluation and process experience are used to choose the options that maximize yield and minimize process time, consistent with the objective above.

The right choice is strongly dependent on disk surface quality; the disk used in Redwing has typically 20 'random' defects per surface, and may occasionally have a scratch from pack merge that can wipe out one sector for any distance up to the whole radius.

The bare minimum test involves one revolution to write the test pattern and one revolution to read the test pattern once; this takes 12 minutes to operate over all the tracks on a Redwing file.

Some file products dispense with surface analysis altogether; and rely on a very high and stable raw disk-surface quality, which tend to lead to long Extended Run tests to give an acceptable defect-escape rate in the field.

The test implemented for Redwing uses one revolution to erase, one revolution to write, and one revolution to test. This takes 18 minutes; extra revolutions would cost 6 minutes each.

Tester overhead, microcode download, motor start/stop, and reserved area setup account for the remaining 2 minutes.

Threshold selection
: The threshold, or clip level, can be programmed in a range from 0 to 15. The lower the number, the more sensitive the test.

If the threshold is too low, many sectors are flagged because of electrical noise and insignificant defects that do not affect data handling; on Redwing, approximately 1% of sectors are spare and if this allowance is used up the file is rejected. So too low a threshold impacts SAT yield.

If the threshold is too high, significant defects are not flagged. This gives an extended run yield problem caused by hard and soft errors, and a possible exposure in the field.

For Redwing, a value of 8 was chosen.

Multiple test revolutions and voting
: There is noise of many forms in all disk files. This means that a test of a track does not necessarily give the same result as another test of the same track, even on the immediately subsequent revolution.

One way of reducing the effect of this is to perform (say) three test revolutions and decide to flag a sector if it is reported defective in any two of the tests.

Microcode for this is available for Redwing, but it is not necessary with the files and cards currently seen.

Multiple writes
: Sometimes, the effect of a small defect on the signal varies depending on how the magnetic transition is written across the defect. If multiple test revolutions are used, this effect can be reduced by interspersing the test revolutions with write revolutions.

Microcode for this is available for Redwing, but it is not necessary with the files and cards currently seen.

Pre-erase
: When files arrive for surface test, the surface condition of the disks is not known. A pre-erase pass guarantees the surface condition, and eliminates possible side-reading. Pre-erase can be a good investment, as it reduces the need for multiple writes and voting.

Pre-erase is performed on Redwing, with the head at half-track positions.

Post-erase
: After a track has been written and tested, it can be erased to minimize adjacent track interference with the next track.

This is not such a good deal as pre-erase; the surface test pattern does not interfere with the next track to produce spurious defects, and on Redwing the head width is generally only about 85% of track pitch.

Microcode for this is available for Redwing, but it is not necessary with the files and cards currently seen.

Off-track testing
: On some files, it is worth performing surface tests offset left and right of the track centre, to locate defects that are not central.

Microcode for this is available for Redwing, but it is not necessary with the files and cards currently seen.

Scratch algorithm
: The scratch algorithm analyzes the raw defect map and tries to find radial scratches, that is defects on the same head and sector in nearly-consecutive tracks.

It flags sites in the middle of such groups, which should fairly clearly have been picked up by the test outright; and it flags sites at either end of such groups, which have sometimes been found to cause data handling problems.

Microcode is available to find scratches; it can be applied multiple times to find different sized scratches, without causing sites to extend continuously.

The scratch algorithm is used with a minimum of 4, a bounce of 5, and an extension of 10.

Summary

It is possible to have a surface test that combines the speed and effectiveness of analogue surface analysis with the low tester cost of a self-test approach. This gives a very low overall process cost, a fast manufacturing process, and the ability to run a surface test in development and field support of the same quality as the manufacturing test.

To do it, there is a need for:

- An advocate in each of the sites concerned with the development of the product.
- A recording subsystem that uses a digitizer.
- Early support from management. Manufacturing management strongly supports this. Development management are not too unhappy when you explain that the component cost will not increase, and can be persuaded that it is worth the reduced manufacturing cost to invest the development time and accept the extra risk of causing a chip design to fail.
- Early involvement with the file design, to ensure that necessary features are designed in.
- Early access to programmers and programming tools, so that the surface test can be brought up at least as soon as anyone wants to use the file for data handling.
- Access to the manufacturing line, so data can be gethered to evaluate the options during EVT and DVT.
- A fallback plan in case something goes irretrievably wrong.

Acknowledgements from Chris Ward

- Jon Coker, for being my advocate in Rochester
- Dave James, for being my advocate in Havant
- Rick Galbraith, for the recording subsystem
- Gordon Dixon, for having faith
- Bill Laishley, for allowing the rest of the chips
- Wallace To, for the data encoder/decoder and the sector mark generator
- John Fairless, for the interlock overrides
- Ian Hamilton, for early microcoding

- Ron Venturi, for access to source code
- Zahur Amin, for later microcoding.

Harrier

The initial part of the section written by Bill Case.

Work on Harrier started in 1988.

This subsection was written by Stan Cutts.

The Harrier product is a DADS subsystem that attaches via SCSI, DFCI, or micro-channel to various IBM systems.

One version attaches to a SCSI (small computer system interface) bus. It has the facility to load any industry-standard SCSI device. The current version of Harrier loads Redwing, and Zumbro 5.25-inch DASD at the front with a pluggable power supply at the rear. Cooling is by an exceptionally quiet fan in a central FRU: the unit with four Redwings achieves 5.4 Bels.

Toboggan 0.25-inch tape drives can be loaded for save and restore operation. Later units of Harrier are to be able to load SCSI CD-ROMs.

The Harrier frame fits a standard EIA rack and is four EIA units high (1 EIA unit = 1.75 inches). The front panel is snap on, with an operator panel that protrudes. Power good and an on/off switch is incorporated. Device power on/off displays are shown under the front panel. The devices are secured by screws: unscewing allows a device to be pulled forward and out of the frame. The dc power cable and SCSI IDC cable are unplugged by hand for removal. The basic Harrier drawer has four partitions each able to contain a Redwing drive.

There are two versions of the Harrier cabinette:

- A four-drawer version that can contain 16 Redwing drives (16GB)
- An eight-drawer version that can contain 32 drives (32GB).

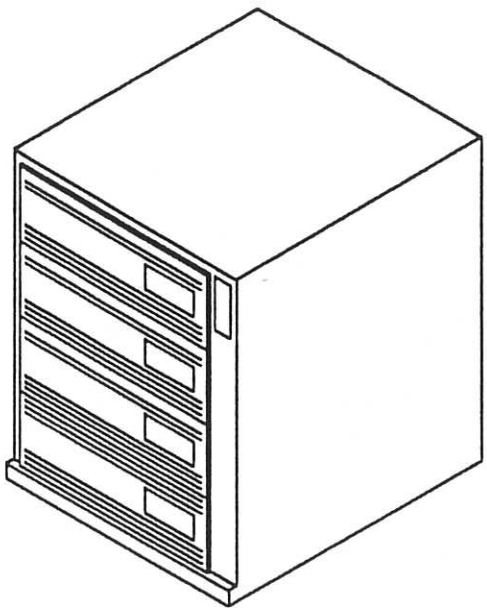

Figure 4. Four Harrier drawers in a tower

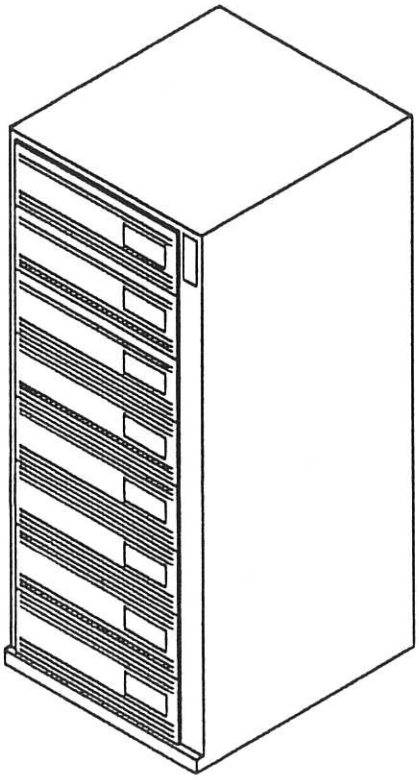

Figure 5. Eight Harrier drawers in a tower

Description of the SCSI drawers

This subsection was written by John Drummond and extracted from the Harrier Technical Reference manual.

Overview

The IBM RISC System/6000 has a storage subsystem product designed for installation in an IBM 7015 POWERserver 930.

Note: In this section the term drawer applies to both device drawers and disk drawers. These differ only in the configuration of the I/O devices installed in them.

The frame can house up to four SCSI I/O devices such as disk drives or tape drives that attach to an ANSI SCSI adapter.

The frame comprises:

- A power supply unit
- A fan assembly
- Four compartments for the SCSI I/O devices

- A detachable front panel
- A controller
- An internal SCSI cable
- Connectors, switches, and lights.

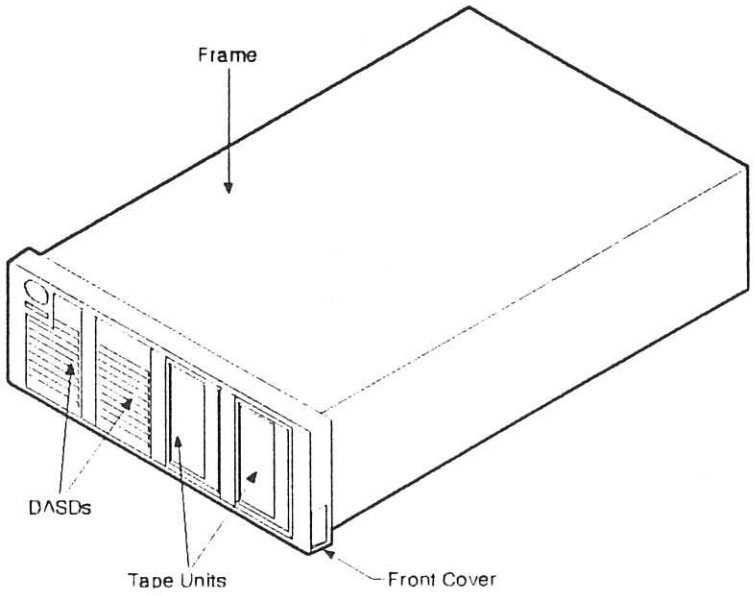

Figure 6. Typical configuration of an SCSI device drawer

The configuration of devices in a SCSI drawer is described in *RISC System/6000 – Hardware Offerings Overview*, GC23-2188.

The following diagram shows the command and data flow in a SCSI drawer:

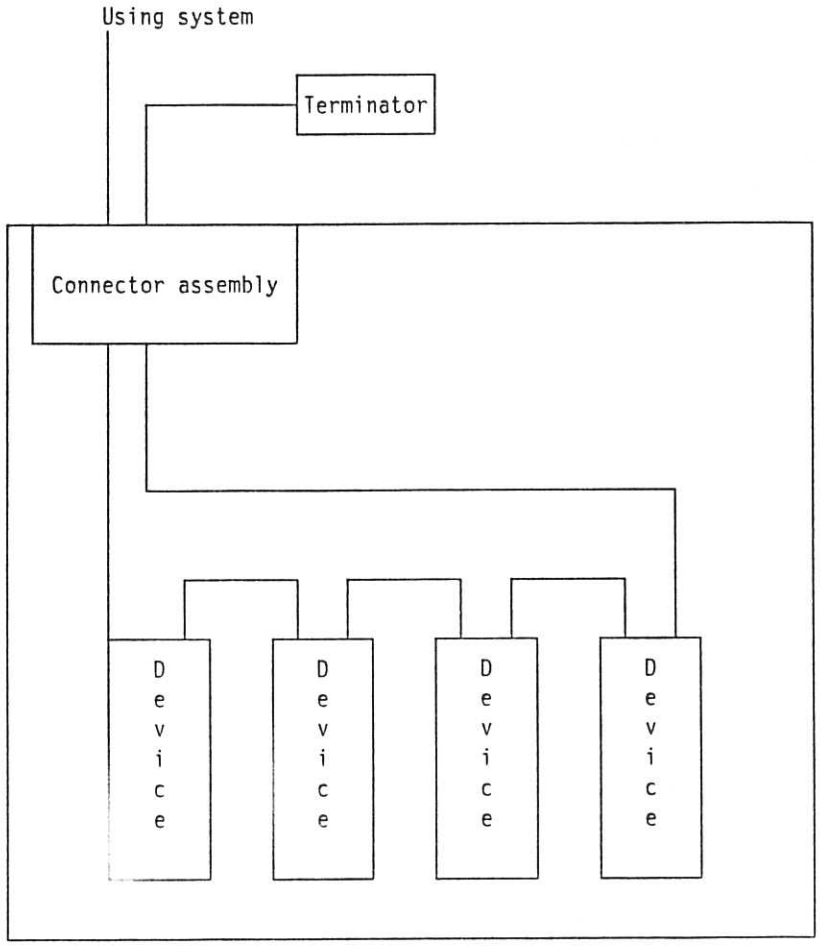

Figure 7. Schematic command and data flow in an SCSI drawer

Commands and data to or from the using system are received at the connector assembly for transmission to or from the individual devices; these are connected in a daisy-chain configuration.

The power supply unit provides power for the devices, cooling fans, and the controller. Electronic interlocks isolate power from the devices should the fans fail or the power supply unit overheat.

Pushbutton switches at each device compartment isolate individual devices. Power connectors are installed on the back panel of the frame.

Some devices are fitted with "ready" lights that illuminate when the device is ready. These lights are visible when the front cover is removed.

This subsection was written by John Drummond and taken from the Harrier Techicnical Reference manual.

Product characteristics

Physical dimensions

Height	171 mm (6.7 in.) nominal
Width	445 mm (17.5 in.) nominal
Depth	686 mm (27 in.) maximum

Weight

Without devices:	25 kg (55 lb)
With four disk drives	43 kg (95 lb)

Power requirements

ac power requirements	
Nominal line voltage	180 V to 260 V
Frequency	48 Hz to 64 Hz
Current:	
Input current at nominal power output (200 W)	2.3 A at 180 V
Input current (continuous)	3.9 A at 180 V ac (225 V dc)
Input current (20 s duration)	4.8 A at 180 V ac (225 V dc)
(1 s duration)	6.0 A at 180 V ac (225 V dc)
(0.1 s duration)	12.0 A at 180 V ac (225 V dc)
(20 ms duration)	39.0 A at 180 V ac (225 V dc)
dc power requirements	
dc line voltage range	240 V to 375 V (equivalent to a nominal ac input of 170 to 260 V)
Maximum starting input current	2.7 A
Maximum inrush current	20 A (for 10 milliseconds)

Environment

Operating environment

Temperature range:	10°C to 40°C (50°F to 102°F)
Relative humidity:	8% to 80%
Maximum wet bulb:	27°C (81°F) non-condensing

Storage environment

Temperature range:	1°C to 60°C (34°F to 140°F)
Relative humidity:	5% to 80%
Maximum wet bulb:	29°C (84°F)

Shipping environment

Temperature range:	-40°C to 60°C (-40°F to 140°F)
Relative humidity:	5% to 100%
Maximum wet bulb:	29°C (84°F)

Altitude

Operating:	0 to 2133 m (0 to 7000 feet)
Non-operating:	-305 m to 12 192 m (-1000 ft to 40 000 ft)

Power line disturbance (PLD)

When the main power line voltage drops below the PLD threshold shown below, the early power-off (EPOW) line from the power supply unit goes to ground. This signal is ORed with the SCSI 'RST' line at the connector assembly, and the devices in the frame are reset. If a reset occurs while a disk is writing, the current block is completed before the reset operation begins.

The 'RST' line does not become inactive until the input line voltage is within the range 180 V to 260 V ac or 225 V to 375 V dc.

PLD operating threshold

Voltage level	Duration	Frequency
<180 V to 160 V ac	2.0 seconds	48 to 64 Hz
<160 V to 130 V ac	0.5 seconds	48 to 64 Hz
<130 V to 0 V ac	30 milliseconds	48 to 64 Hz
<225 V	0	dc

Electromagnetic compatibility

A fully configured Harrier in an IBM rack meets the requirements of:

- FCC[1] class A
- VDE[2] 0871/6.78 class A.

Safety Approvals

The Harrier complies with the following safety standards:

- IEC[3] 950 and 384-14
- VDE 0805/0806
- CSA[4] 22-2 no.220
- UL[5] 478 (5th edition), 796, and 1012

---

1 United States Federal Communications Commission

2 Verband Deutscher Electrotechniker

3 International Electrotechnical Commission

4 Canadian Standards Association

5 Underwriters Laboratories Incorporated