Reading/Recording Devices

Herman Vermariën
Vrije Universiteit Brussel

Edward McConnell
National Instruments

Yufeng Li
Samsung Information Systems America

96.1 Graphic Recorders

Herman Vermariën

A **graphic recorder** is essentially a measuring apparatus that is able to produce in real time a hard copy of a set of time functions with the purpose of immediate and/or later visual inspection. The curves are mostly drawn on a (long) strip of paper (from a roll or Z-fold); as such, the instrument is indicated as a strip chart recorder. The independent variable time (t) then corresponds to the strip length axis and the physical variables measured are related to the chart width. Tracings are obtained by a writing process at sites on the chart short axis (y) corresponding to the physical variables magnitudes with the strip being moved at constant velocity to generate the time axis. Graphs cannot be interpreted if essential information is absent; scales and reference levels for each physical variable recorded and for time are all a necessity. Additional information concerning the experimental conditions of the recording is also necessary and is preferably printed by the apparatus (data, investigated item, type of experiment, etc.). The capacity of the graphic recorder is thus determined by its measuring accuracy, its ability to report additional information and its graphical quality, including the sharpness of tracings, the discriminability of tracings (e.g., by different colors), and the stability of quality with respect to long-term storage. Simple chart recorders only produce tracings on calibrated paper; more-advanced graphic recorders generate tracings and calibration lines, display additional information in alphanumeric form on charts, store instrument settings and recorded data in memory (which can be reproduced on charts in diverse modes), have a built in waveform monitor screen, and can communicate with a PC via standard serial interfacing. The borderlines between these types of intelligent graphic recorders and, on the one hand, digital storage oscilloscopes equipped with a hard copy unit and, on the other hand, PC-based data acquisition systems

with a laser printer or a plotter, become very unclear. The property of producing the hard copy in real time is probably the most discriminating factor between the graphic recorder and other measuring systems that produce hard copies. Graphic recorders are used for test and measurement applications in laboratory and field conditions and for industrial process monitoring and control. Graphic recorders are intensively used in biomedical measurement applications [1].

Whereas the time axis is generated by moving the chart at constant velocity, the ordinate can be marked in an analog or a digital manner. Analog recorder transducers generate a physical displacement of the writing device, e.g., a pen or a printhead. With digital transducers, moving parts are absent and the writing device is a stationary rectilinear array of equidistant writing points covering the complete width of the chart; the writing act then consists in activating the point situated at the site corresponding to the signal magnitude and putting a dot on the paper. Analog recorders thus can produce continuous lines, whereas digital recorders generate dotted lines. If ordinate and time axis resolutions are sufficient, digital recordings have excellent graphic quality regarding visual impression of continuity. Analog transducers can be used in a discontinuous mode and thus handle a set of slowly varying signals; during the scanning cycle a dot is set on the paper by the moving writing device at the sites corresponding to the magnitudes of the signal. A single digital transducer and a single analog transducer applied in the scanning mode can handle a set of signals; a single analog transducer can process only one signal in the continuous mode. For a digital transducer the number of signals recorded is essentially unlimited; it is thus programmed to draw the necessary calibration lines. With analog transducers calibrated paper is used. In this case, generally ink writing is applied and different colors ensure excellent tracing identifiability. With the digital array, dot printing can be more or less intensified for different channels or more than one adjacent points can be activated, resulting in more or less increased line blackness and thickness. However, tracing identification is usually performed by alphanumeric annotations.

In **analog graphic recorders,** the transducer can be designed in a direct mode or in a servo mode. In the direct mode the signal magnitude is directly transduced to a position of the writing device (e.g., the simple galvanometric type). In the servo mode (also called *feedback* or *compensation*) the position of the writing device is measured by a specific sensor system and the difference between the measured value and the value of the signal to be traced is applied to the transducer motor, resulting in a movement tending to zero the difference value and thus to correct the position [2,3]. In both methods the moving parts set a limit to the system frequency bandwidth. Moreover, in the feedback mode velocity and acceleration limitations may be present; thus linear system theory description of the apparatus behavior with respect to signal frequency may not be applicable. As such, the bandwidths of servo systems can be dependent on the writing width. Movement of the writing device can be generated by a rotation or by a translation. In the latter case the writing part is mechanically guided; primarily, the servo method is applied [3]. A rotation is generally obtained with a galvanometric motortype [3,4]; the galvanometer may rotate a pen, an ink jet, a light beam. The inertia of the moving part is the major parameter determining the bandwidth of the system. Translational devices allow a bandwidth of a few hertz. Higher bandwidths can be obtained with galvanometric pen types (about 100 Hz), ink jets (up to 1 kHz), and optical systems (up to 10 kHz) [1], but, being replaced by dot array recorders or data acquisition systems, these types are disappearing from the market. Major reasons are inherent errors and limitations of these analog types, maintenance needs of moving parts and ink devices, cost of photographic paper, and the lack of the possibilities of digital types.

Whereas moving parts restrict the analog recorder bandwidth, a corresponding capacity of **digital graphic recorders** is determined by the sampling frequency and the writing frequency. According to the sampling criterion the sampling frequency should be twice the highest signal frequency. This implies two samples to display a complete sine wave period, which can hardly be called a good graphic representation. Ten samples may be a minimum. The sampling frequency is a pure electronic matter; the maximal writing frequency of the dot array is the limiting factor in real time. Alternatively, if the signal spectrum exceeds the real-time bandwidth of the recorder, data can be stored at a sufficient sampling rate in memory and reproduced off-line at a slower rate which can be handled by the apparatus. Most digital recorders have

this memory facility; some recorders are specifically referred to as "memory" recorders when their off-line capabilities largely exceed their online performance.

Real-time recording is primarily performed as a function of time (*t–y* recorders). On the other hand, *x–y* recording is another way of representing the data. In this case the relation between two physical variables is traced and the independent variable time is excluded (apart from the fact that, if a dashed line is used, each dash can represent a fixed time interval). In standard analog *x–y* recorders the chart is stationary (e.g., electrostatically fixed to the writing table); two similar analog writing transducers with identical input circuitry are assembled in the recorder. The first transducer (*y*) translates a long arm covering the width of the paper at which the second transducer (*x*) carrying the pen is moved. Evidently, recorders with memory facilities and appropriate software may produce an *x–y* recording in off-line mode by setting dots while the paper progresses. Recording accuracy can be formulated in similar terms as for any measuring instrument [5]. This accuracy is determined, on the one hand, by the input signal conditioning, similar for digital as well as for analog types and, on the other hand, by the accuracy of the recording transducer and its driver electronics. In the digital type, digitization bit accuracy, sampling frequency, dot array resolution, and dot writing frequency are major parameters. In the analog type, typical inconveniences of analog transducer systems can be found (such as static nonlinearity, noise and drift, dead zone and hysteresis, limited dynamic behavior); servo systems are known to be more accurate as compared with direct systems. For example, drift in analog recording can be the result of a small shift of the calibrated chart along the *y*-axis; the latter is excluded if the recorder draws its own calibration lines.

With respect to graphic quality, clarity and sharpness of the tracings are important (within a large range of writing velocities). Tracing quality depends on the writing velocity, i.e., the velocity of the writing device with respect to the moving paper. Evidently, the flow of writing medium (e.g., ink or heat) should be more or less adapted to this writing velocity to prevent poorly visible tracings at high velocities and thick lines at low velocities. Good identifiability of overlapping curves is essential. Sufficient dot resolution (with adequate interpolation techniques) is important in discontinuous types for easy visual inspection. A graphic recorder can be designed as a single-channel instrument or can have a multichannel input. Inputs can be standard or modular, so that the user can choose the specific signal conditioners for the application and the number of channels. A recorder can be called "general purpose" or can be assembled in a specific measuring apparatus (e.g., in biomedical applications such as electrocardiography and electroencephalography). The recorder can be portable for operation in the field or can be mounted in a laboratory rack or in a control panel. The paper can be moved vertically or horizontally on a writing table ("flat bed" recorder) and can be supplied from a roll or in Z-fold. Besides strip chart recorders and *x–y* recorders, circular chart recorders exist. In this case the chart rotates, one rotation corresponds to a complete measurement interval, and the chart is provided with appropriate calibration lines adapted to movement of the pens writing on it.

Apart from the low-bandwidth translational pen devices there is a decreasing interest in analog graphic recorders. They are being replaced by thermal dot array recorders or by data acquisition systems. Nevertheless, as some of them may still be manufactured and a number of apparatus may still be in use, different techniques are mentioned. A description of analog recorder principles and performances can be found in Reference 1. Galvanometric recorders apply rotational transducers. The direct as well as the servo principles are used. The direct type makes use of the d'Arsonval movement as applied in ordinary galvanometers [1, 2, 4]. Dynamically, the galvanometer acts as a mechanical resonant system and the bandwidth is thus determined by its resonant frequency, the latter being dependent on the inertia of the moving parts. Evidently, rotation gives rise to inherent errors in graphic recorders. If pen tips (perpendicular to the pen arm, thermal or ink) are used, the rotation of the pen arm fixed to the galvanometer coil occurs in a plane parallel to the chart plane, so the recording is curvilinear instead of rectilinear, introducing an error with respect to the time axis and to the ordinate axis (the ordinate value being proportional to the tangent of the rotation angle). Calibrated paper with curvilinear coordinate lines may solve this problem; nevertheless, the tracing is deformed and zero offset is critical. Rectilinear recording can be realized with pen systems, ink jets, and light beams. Rectilinear pen recording can be

approximated with pen tips in case of “long-arm” pens and by mechanical rectilinearization; alternatively “knife-edge” recording is a solution [1]. In the case of ink jet and light beam recorders the rotation plane and the chart plane do not have to be parallel; writing then occurs at the intersecting line of both planes and is thus essentially rectilinear. In ink jet recording a miniature nozzle through which ink is pumped is mounted in a direction perpendicular to the axis of the galvanometer. In optical recording a sharp light beam is reflected by a small mirror connected to the galvanometric moving coil toward the photosensitive paper. In these methods miniaturization of the moving parts gives rise to higher resonant frequencies and thus higher bandwidths. Whereas a typical bandwidth for a galvanometric pen system is 100 Hz, the bandwidth for an ink jet system can be 1000 Hz and for an optical system 10 kHz may be reached. In the fiber-optic cathode ray tube (FO-CRT) no mechanical moving parts are present and thus there are no mechanical limits on bandwidth. The FO-CRT is essentially a one-dimensional CRT. A phosphor layer at the inside of the screen converts the cathode ray into ultraviolet (UV) light. This UV light is guided by an outer faceplate composed of glass fibers onto the photosensitive chart. As in ordinary oscilloscopes, the deflection of the spot is directly proportional to the signal applied at the input of the deflection unit. The bandwidth is determined by the driving electronics. The system can be used in scanning mode as the beam intensity is easily controlled. In the following paragraphs further details will be given on translational pen recorders and thermal dot array recorders.

Translational Pen Recorders

In **translational pen recorders** the writing device is usually a fiber-tip pen with an ink cartridge. In discontinuous applications the writing device can be a printhead with different color styli or with a colored ribbon. A manual or automatic pen lift facility is included. During recording, the writing device is translated along the *y*-axis as it is linked to a mechanical guidance and a closed-loop wire system. A motor and wheels system pulls the wire and thus the writing device. In some designs a motor and screw system is applied. Translational recorders are primarily designed as a servo type. The position of the pen is accurately measured, and the difference voltage between the input signal and the position magnitude (following appropriate amplification and conditioning) drives the servomotor. Servo motors can be dc or stepper types; servo electronics can be analog or digital. Position sensing can be potentiometric (“potentiometric” recorders): the pen carriage is equipped with a sliding contact on the resistor (wire wound or thick film) which covers the complete width of the paper. More recently developed methods use optical or ultrasonic principles for position sensing; with these methods contacts are absent resulting in less maintenance and longer lifetime. For example, in the ultrasonic method the pen position is sensed by a detector coil from the propagation time of an ultrasound pulse, which is imparted by a piezoelectric transducer to a magnetostrictive strip covering the chart width. Accordingly, brushless dc-motors are used in some apparatuses. In the servo system accuracy is determined for the larger part by the quality of the sensing system. A poor contact with the resistor can give rise to noise; there may be a mechanical backlash between pen tip and the sliding contact on the potentiometer. The velocity of the pen carriage is limited, about 0.5 to 2 m/s dependent on motor and mechanics design. This results in a bandwidth of the recorder depending on the amplitude of the tracing: the –3 dB frequency fits in the range from 1 to 5 Hz for a full-scale width of 200 to 250 mm. Alternatively, the pen response time to a full-scale step input is given (5 to 95% of full-scale tracing): 0.1 to 0.5 s. Overshoot of the pen step response is extremely small in accurate designs.

In most pen recorders each tracing can cover the complete width. As such, pens must have the possibility to pass each other resulting in a small shift between adjacent pens (a few millimeters) along the time axis. In some apparatus, tracings can be synchronized with a POC-system (“pen offset compensation”); signals are digitized, stored in memory, and reproduced after a time delay correcting for the physical displacement of the pen. If immediate visual inspection is required, applying POC can be inconvenient as a consequence of this time delay. In process monitoring, slowly varying signals such as temperature, pressure, flow, etc. are followed. These signals can be handled by a single transducer in a discontinuous way; all input signals are scanned during the scanning cycle and for each signal a dot is

TABLE 96.1 Pen Recorders

Designation	Description	Manufacturer	Approximate Price (U.S.$)
LR8100	Test, meas.; 4,6,8 c. ch.; POC; printer; display; memory; analysis; alarm; interface	Yokogawa E. C.	11,500 (8 ch.)
LR102	Test, meas.; 1,2 c. ch.	Yokogawa E. C.	1,800 (2 ch.)
LR122	Test, meas.; 1,2 c. ch.; *x–y*; alarm; interface	Yokogawa E. C.	2,300 (2 ch.)
MC1000	Test, meas.; 4,6,8,12 c. ch.; POC; printer; waveform display; analysis; alarm; interface	Graphtec C.	20,200 (12 ch.)
BD112	Test, meas.; 2 c. ch.; POC	Kipp Z.	2,700 (2 ch.)
BD200	Test, meas.; 4,6,8 c. ch.; POC; display; *x–y*; alarm; interface	Kipp Z.	12,200 (8 ch.)
L250	Test, meas.; 1,2 c. ch.; POC	Linseis	2,000 (2 ch.)
L2066	Test, meas.; 1 to 6 c. ch.; POC; *x–y*; interface	Linseis	8,100 (6 ch.)
MCR560	Test, meas.; 2,4,6 c. ch.	W+W	3,800 (2 ch.)
DCR540	Test, meas.; 1 to 4 c. ch.; POC; display; *x–y*; interface	W+W	6,600 (4 ch.)
PCR500SP	Test, meas.; 2,4,6,8 c. ch.; POC; display; *x–y*; analysis; alarm; interface; transient option	W+W	15,900 (8 ch.)
Omega640	Test, meas.; 1,2,3 c. ch.	Omega	4,300 (3 ch.)
Omega600A	Test, meas.; 1 to 6 c. ch.; POC; printer; *x–y*; memory; interface	Omega	23,700 (6 ch.)
µR1000	Process mon.; 1,2,3,4 c. ch., 6 s. ch.; POC; printer; display; analysis; alarm; interface	Yokogawa E. C.	4,000 (4 c. ch.)
µR1800	Process mon.; 1,2,3,4 c. ch., 6,12,18,24 s. ch.; POC; printer; display; analysis; alarm; interface	Yokogawa E. C.	5,600 (4 c. ch.)
DR240	Process mon.; 30 s. ch.; printer; display; analysis; alarm; interface	Yokogawa E. C.	6,500 (30 s. ch.)
RL100	Process mon.; 1,2 c. ch., 6 s. ch.; printer; alarm	Honeywell	1,500 (2 c. ch.)
DPR100C/D	Process mon.; 1,2,3 c. ch., 6 s. ch.; POC; printer; display; alarm; analysis; interface	Honeywell	3,000 (6 s. ch.)
DPR3000	Process mon.; 4 to 32 s. ch.; printer; display; alarm; analysis; interface	Honeywell	7,700 (32 s. ch.)
4101	Process mon.; 1 to 4 c. ch., 6 s. ch.; POC; printer; alarm	Eurotherm	2,400 (4 c. ch.)
4180 G	Process mon.; 8,16,24,32 s. ch.; printer; waveform display; alarm; analysis; interface	Eurotherm	9,800 (32 s. ch.)
Sirec L	Process mon.; 1,2,3 c. ch.	Siemens	1,200 (3 c. ch.)

Note: c. ch. = continuous channel; s. ch. = scanned channel.

printed ("multipoint" recorder). The minimum scanning time is dependent on the moving writing device. For chart progression, dc and stepper motors are used. Calibrated paper is pulled by sprocket wheels seizing in equidistant perforations at both sides of the chart. Translational pen recorders range from simple purely analog design to intelligent microprocessor-controlled types handling a large number of channels with a broad range of control and monitor facilities (e.g., printing of a report after alarm).

Table 96.1 displays a set of translational pen recorders; some of them are equipped with a printhead. Under "Description" the major application is given: test and measurement or process monitoring. Furthermore the following are indicated: the number of continuous (c. ch.) and scanned (s. ch.) channels; the availability of POC, a printer (for additional information or for trace printing), a display for alphanumeric information (such as calibration values for each channel), or even a waveform display, data memory (allowing memory recorder functioning), *x–y* recording facility, alarm generation (after reaching thresholds of recorded variables), and standard serial interface options allowing communication with a PC (introduction of recorder settings, storage, and processing of recorded data, etc.). Table 96.2 gives a summary of pen recorder specifications (multipoint types also included).

Thermal Dot Array Recorders

In **thermal dot array recorders**, apart from the chart-pulling system, no moving parts are present; the writing transducer is essentially a rectilinear array of equidistal writing points which covers the total width of the paper. Although some apparatuses apply an electrostatic method [1], the thermal dot array

TABLE 96.2 Pen Recorder Specifications

Type (Test, Measurement)	Recording Width (mm)	Chart Velocity max. (mm/min)	Chart Velocity min. (mm/h)	Pen Velocity, max. (m/s)	Pen Step Response time(s)	Bandwidth (–3 dB) (Hz)	Number of Continuous Channels, max.
LR8100	250	1200	10	1.6		5	8
LR102	200	600	10	0.4	0.5	1.5	2
LR122	200	600	10	0.4	0.5	1.5	2
MC1000	250	1200	7.5	1.6			12
BD112	200	1200	6		0.2		2
BD200	250	1200	5		0.25		8
L250	250	1200	6	1	0.12	3.6	2
L2066	250	3000	1	1	0.3	2	6
MCR560	250	600	10		0.3	1.5	6
DCR540	250	600	10		0.3	1.5	4
PCR500SP	250	1200	10	2	0.15	4.5	8
Omega640	250	600	30	0.5			3
Omega600A	250	600	10	0.1			6

Type (Process Monitor)	Recording Width (mm)	Chart Velocity max. (mm/min)	Chart Velocity min. (mm/h)	Pen Step Response Time (s)	Printhead Scanning Cycle (s)	Number of Channels (max.)[a]
µR1000	100	200	5	1	10	4c, 6s
µR1800	180	200	5	1.5	10 (6s)	4c, 24s
DR240	250	25	1	—	2	30s
RL100	100	8	10	3.2	5	2c, 6s
DPR100C/D	100	100	1	1	0.6	3c, 6s
DPR3000	250	25	1	—	5	32s
4101	100	25	1	2	5	4c, 6s
4180 G	180	25	1	—	3	32s
Sirec L	100	20	1		—	3c

[a] c = continuous; s = scanned.

and thermosensitive paper are generally used. In this array the writing styli consist of miniature electrically heated resistances; thermal properties of the resistances (in close contact with the chart paper) and the electric activating pulse form determine the maximal writing frequency. The latter ranges in real-time recorders from 1 to 6.4 kHz. Heating of the thermosensitive paper results in a black dot with good long-term stability. The heating pulse is controlled in relation to the chart velocity in order to obtain sufficient blackness at high velocities. Tracing blackness or line thickness is seldom used for curve identification; alphanumeric annotation is mostly applied. With the dot array a theoretically unlimited number of waveforms can be processed; the apparatus is thus programmed to draw its own calibration lines. Different types of grid patterns can be selected by the user. Moreover, alphanumeric information can be printed for indicating experimental conditions.

Ordinate axis resolution is determined by the dot array: primarily, 8 dots/mm; exceptionally, 12 dots/mm (as in standard laser printers). The resolution along the abscissa depends on the thermal array limitations and programming. Generally, a higher resolution is used (mostly 32 dots/mm, maximally 64 dots/mm) except for the highest chart velocities (100, 200, 500 mm/s). At these high velocities and consequently short chart contact times, dots become less sharp and less black. Most of the dot array instruments are intended for high-signal-frequency applications: per channel sampling frequencies of 100, 200, and even 500 kHz are used in real time. These sampling frequencies largely exceed the writing frequencies; during the writing cycle, data are stored in memory and for each channel within each writing interval a dotted vertical line is printed between the minimal and the maximal value. For example, a sine wave with a frequency largely exceeding the writing frequency is represented as a black band with a width

TABLE 96.3 Thermal Dot Array Recorders

Designation	Description	Manufacturer	Approximate Price (U.S.$)
WR 5000	8 a. ch.; memory	Graphtec C.	17,100 (8 ch.)
WR 9000	4,8,16 a. ch.; monitor; memory; *x–y*; analysis; FFT	Graphtec C.	11,300 (4 ch.)
Mark 12	4 to 52 a. ch., 4 to 52 d. ch.; monitor; memory	W. Graphtec	30,300 (16 a. ch.)
MA 6000	2 to 16 a. ch.; monitor; memory; *x–y*; analysis; FFT	Graphtec C.	23,500 (8 ch.)
ORP 1200	4,8 a. ch., 16 d. ch.; monitor; memory; *x–y*	Yokogawa E. C.	11,600 (8 a. ch.)
ORP 1300	16 a. ch., 16 d. ch.; monitor; memory; *x–y*	Yokogawa E. C.	18,000 (16 a. ch.)
ORM 1200	4,8 a. ch., 16 d. ch.; monitor; memory; *x–y*	Yokogawa E. C.	14,100 (8 a. ch.)
ORM 1300	16 a. ch., 16 d. ch.; monitor; memory; *x–y*	Yokogawa E. C.	21,900 (16 a. ch.)
OR 1400	8 a. ch., 16 d. ch.; monitor; memory; *x–y*	Yokogawa E. C.	16,100 (8 a. ch.)
TA 240	1 to 4 a. ch.	Gould I. S.	8,500 (4 ch.)
TA 11	4,8,16 a. ch.; monitor; memory	Gould I. S.	18,900 (16 ch.)
TA 6000	8 to 64 a. ch., 8 to 32 d. ch.; monitor; memory	Gould I. S.	33,900 (16 a. ch.)
Windograf	2 to 4 a. ch.; monitor	Gould I. S.	10,200 (4 ch.)
Dash 10	10,20,30 a. ch.; monitor; memory	Astro-Med	22,500 (10 ch.)
MT95K2	8 to 32 a. ch., 32 d. ch.; monitor; memory; *x–y*; analysis	Astro-Med	32,600 (8 a. ch.)
8852	4 a. ch., 24 d. ch.; monitor; memory; *x–y*; analysis; FFT	Hioki E. E. C.	22,300 (4 a. ch.)
8815	4 a. ch., 32 d. ch.; memory; *x–y*	Hioki E. E. C.	4,500 (4 a. ch.)
8825	16 a. ch., 32 d. ch.; monitor; memory; *x–y*; analysis	Hioki E. E. C.	28,800 (8 a. ch.)

Note: a. ch. = analog channel; d. ch. = digital channel.

equal to the sine amplitude. In this way the graphs indicate the presence of a phenomenon with a frequency content exceeding the writing frequency. As data are stored in memory they can be reproduced at a lower rate thus revealing the actual high-frequency waveform captured. Some apparatuses use a much lower sampling rate in real time and only perform off-line: in this case the apparatus is indicated as a "memory" recorder. Digitization accuracy ranges from 8 to 16 bit, whereas the largest number of dots full scale is 4800. In this way the useful signal may be superposed on a large dc-offset: it can be written or reproduced with excellent graphic quality with the offset digitally removed and the scale adapted.

In a high-performance recorder a waveform display is extremely useful to avoid paper spoiling, in real-time and in off-line recording as well. The display is also used for apparatus settings. Signals can be calibrated and real physical values and units can be printed at the calibration lines. Via memory *x–y* plots can be obtained. Some apparatuses allow application of mathematical functions for waveform processing and analysis: original and processed waveforms can be drawn together off-line. A few types are equipped with FFT software. Computer interfacing, a large set of triggering modes (including recording at increased velocity after a specific trigger), event channels, etc. are standard facilities. Table 96.3 shows a set of thermal dot array recorders (under "Description": number of analog channels (a. ch.) and digital channels (d. ch.); waveform monitor; signal data memory, *x–y* facility, mathematical analysis, FFT) and Table 96.4 gives specifications.

Concluding Remarks

Table 96.5 gives addresses and fax and phone numbers of manufacturers of recorders mentioned in Tables 96.1 and 96.3. It should be remarked that prices mentioned in these tables hold for purchasing a complete functioning apparatus (number of channels indicated) from firms in Belgium representing the manufacturers and having provided the data sheets from which specifications were derived. With the expression "a complete functioning" apparatus a standard system is meant, thus including simple input couplers (in case of a modular design), standard RAM and analysis software, no specific options. Obviously, the list of manufacturers is incomplete. It should be mentioned that the number of manufacturers of graphic recorders is decreasing; a significant and increasing amount of applications has been taken over by

TABLE 96.4 Thermal Dot Array Recorder Specifications

Type	Recording Width (mm)	Thermal Array Resolution (dots/mm)	Chart Velocity max. (mm/s)	Chart Velocity min. (mm/h)	Maximal Writing Frequency (dots/s)	Time Axis Resolution max. (dots/mm)	Time Axis Resolution min. (dots/mm)
WR 5000	384	8	200	1	1600	64	8
WR 9000V	200	8	100	1		32	
WR 9000M	200	8	100	1		32	
Mark 12	384	8	200	1	1600	64	8
MA 6000	205	8	100	1		40	
ORP 1200	201	8	100	10	1600	32	16
ORP 1300	201	8	100	10	1600	32	16
ORM 1200	201	8	100	10	1600	32	16
ORM 1300	201	8	100	10	1600	32	16
OR 1400	201	8	250	10	6400	32	25.6
TA 240	104	8	125	36	1000	32	8
TA 11	264	8	200	36	1600	16	8
TA 6000	370	8	200	36	1600	16	8
Windograf	104	8	100	36	800	32	8
Dash 10	256	12	200	60	1200	12	6
MT95K2	400	12	500	1	2000	48	4
8852	100	8	25	10	200	16	8
8815	104	6	8	10	50	12	6
8825	256	8	20	10	200	10	10

Type	Sampling Frequency Real-Time, max. (kHz)	Bit Accuracy, max. (bits)	No., max., Channels[a]	Display Dimensions (mm)	Display Resolution (pixels)	Sampling Frequency Memory, max. (kHz)	Samples Stored/ Channel, max.
WR 5000	64	14	8a	—	—	64	32 k
WR 9000V	250	12	8a	192× 120	640× 400	250	512 k
WR 9000M	50	14	8a	192× 120	640× 400	50	512 k
Mark 12	200	16	52a, 52d	97× 77	256× 320	200	2 M
MA 6000	500	16	16a	192× 120	640× 400	500	512 k
ORP 1200	100	14	8a, 16d	127[b]	320× 240	100	32 k
ORP 1300	100	14	16a, 16d	127[b]	320× 240	100	32 k
ORM 1200	100	14	8a, 16d	127[b]	320× 240	100	128 k
ORM 1300	100	14	16a, 16d	127[b]	320× 240	100	128 k
OR 1400	100	16	8a, 16d	127[b]	320× 240	100	256 k
TA 240	5	12	4a	—	—	—	—
TA 11	250	12	16a	198× 66	640× 200	250	500 k
TA 6000	250	12	64a, 32d	224× 96	640× 200	250	500 k
Windograf	10	12	4a	178[b]	800× 350	—	—
Dash 10	250	12	30a		256× 64	250	512 k
MT95K2	200	12	32a, 32d			200	500 k
8852	1.6	8	4a, 24d	178[b]		100×10^3	1 M
8815	12.5	8	4a, 32d	—	—	500	30 k
8825	8	12	16a, 32d	254[b]	640× 480	200	500 k

[a] a = analog; d = digital.
[b] Diagonal.

"paperless" recorders, i.e., data acquisition systems. Nevertheless, the possibility of generating graphs in real time remains an important feature, e.g., to provide evidence of the presence of a specific phenomenon. In recent years analog recorders have become less used and manufactured (apart from the translational pen types, especially in industrial process monitoring). Thermal array recorders have become more important: the quality and long-term stability of thermal paper have improved and cost levels are comparable with calibrated paper for ink recording. In new designs, recorders provide more capabilities

TABLE 96.5 Companies that Make Graphic Recorders

Astro-Med, Inc. Astro-Med Industrial Park, West Warwick, RI 02893, USA fax (401) 822 - 2430/phone (401) 828-4000	Kipp & Zonen, Delft BV Mercuriusweg 1, P.O. Box 507, NL-2600 AM Delft, The Netherlands fax 015-620351/phone 015-561000
Eurotherm Recorders Ltd. Dominion Way, Worthing, West Sussex BN148QL, Great Britain fax 01903-203767/phone 01903-205222	Linseis GMBH Postfach 1404, Vielitzer Strasse 43, D-8672 Selb, Germany fax 09287/70488/phone 09287/880-0
Gould Instrument Systems, Inc. 8333 Rockside Road, Valley View, OH 44125-6100, USA fax (216) 328-7400/phone (216) 328-7000	Omega Engineering, Inc. P.O. Box 4047, Stamford, CT 06907-0047, USA fax (203) 359-7700/phone (203) 359-1660
Graphtec Corporation 503-10 Shinano-cho, Totsuka-ku, Yokohama 244, Japan fax (045) 825-6396/phone (045) 825-6250	Siemens AG, Bereich Automatisierungstechnik Geschäftsgebiet Processgeräte, AUT 34, D-76181 Karlsruhe, Germany fax 0721/595-6885/phone 0721/595-2058
Hioki E. E. Corporation 81 Koizumi, Ueda, Nagano, 386-11, Japan fax 0268-28-0568/phone 0268-28-0562	Western Graphtec, Inc. 11 Vanderbilt, Irvine, CA 92718-2067, USA fax (714) 770-6010/phone (800) 854-8385
Honeywell Industrial Automation and Control 16404 North Black Canyon Hwy., Phoenix, AZ 85023, USA phone (800) 343-0228	W+W Instruments AG Frankfurt-Strasse 78, CH-4142 Münchenstein, Switzerland fax +41 (0) 6141166685/phone +41 (0) 614116477
	Yokogawa Electric Corporation Shinjuku-Nomura Bldg. 1-26-2 Nishi-Shinjuku, Shinjuku-ku, Tokyo 163-05, Japan fax 81-3-3349-1017/phone 81-3-3349-1015

and appear more intelligent, obviously leading to increased complications with respect to instrument settings and thus increased need for training and experience in the use of the instrument.

Defining Terms

Analog graphic recorder: A graphic recorder that makes use of an analog transducer system (e.g., a moving pen).

Analog recorder bandwidth: The largest frequency that can be processed by the analog recorder (–3 dB limit).

Digital graphic recorder: A graphic recorder that makes use of a digital transducer system (e.g., a fixed dot array).

Graphic recorder: A measuring apparatus that produces in real time a hard copy of a set of time-dependent variables.

Maximal sampling frequency: Maximal number of data points sampled by the digital recorder per time unit (totally or per channel).

Maximal writing frequency: Maximal number of writing (or printing) acts executed by the digital recorder per time unit.

Thermal dot array recorder: A digital recorder applying a fixed thermal dot array perpendicular to the time axis.

Translational pen recorder: An analog recorder with one or several pens being translated perpendicularly to the time axis.

References

1. H. Vermariën, Recorders, graphic, in J.G. Webster, Ed., *Encyclopedia of Medical Devices and Instrumentation*, New York: John Wiley & Sons, 1988.
2. D.A. Bell, *Electronic Instrumentation and Measurements*, 2nd ed., Englewood Cliffs, NJ: Prentice-Hall, 1994.
3. A. Miller, O.S. Talle, and C.D. Mee, Recorders, in B.M. Oliver and J.M. Cage, Eds., *Electronic Measurements and Instrumentation*, New York: McGraw-Hill, 427–479, 1975.
4. R.J. Smith, *Circuits, Devices and Systems: A First Course in Electrical Engineering*, New York: John Wiley & Sons, 1976.
5. W.H. Olson, Basic concepts in instrumentation, in J.G. Webster, Ed., *Medical Instrumentation: Application and Design*, *3rd ed.*, New York: John Wiley & Sons, 1998.

96.2 Data Acquisition Systems

Edward McConnell

The fundamental task of a **data acquisition system** is the measurement or generation of real-world physical signals. Before a physical signal can be measured by a computer-based system, a sensor or transducer is used to convert the physical signal into an electrical signal, such as voltage or current. Often only a plug-in data acquisition (DAQ) board is considered the data acquisition system; however, a board is only one of the components in the system. A complete DAQ system consists of sensors, signal conditioning, interface hardware, and software. Unlike stand-alone instruments, signals often cannot be directly connected to the DAQ board. The signals may need to be conditioned by some signal-conditioning accessory before they are converted to digital information by the plug-in DAQ board. Software controls the data acquisition system — acquiring the raw data, analyzing the data, and presenting the results. The components are shown in Fig. 96.1.

Signals

Signals are physical events whose magnitude or time variation contains information. DAQ systems measure various aspects of a signal in order to monitor and control the physical events. Users of DAQ systems need to know the relation of the signal to the physical event and what information is available in the signal. Generally, information is conveyed by a signal through one or more of the following signal parameters: state, rate, level, shape, or frequency content. The physical characteristics of the measured signals and the related information help determine the design of a DAQ system..

All signals are, fundamentally, analog, time-varying signals. For the purpose of discussing the methods of signal measurement using a plug-in DAQ board, a given signal should be classified as one of five signal types. Because the method of signal measurement is determined by the way the signal conveys the needed information, a classification based on this criterion is useful in understanding the fundamental building blocks of a DAQ system.

As shown in the Fig. 96.2, any signal can generally be classified as analog or digital. A digital, or binary, signal has only two possible discrete levels of interest — a high (on) level and a low (off) level. The two digital signal types are on–off signals and pulse train signals. An analog signal, on the other hand, contains information in the continuous variation of the signal with time. Analog signals are described in the time or frequency domains depending upon the information of interest. A dc type signal is a low-frequency signal, and if the phase information of a signal is presented with the frequency information, then there is no difference between the time or frequency domain representations. The category to which a signal belongs depends on the characteristic of the signal to be measured. The five types of signals can be closely paralleled with the five basic types of signal information — state, rate, level, shape, and frequency content.

Basic understanding of the signal representing the physical event being measured and controlled assists in the selection of the appropriate DAQ system.

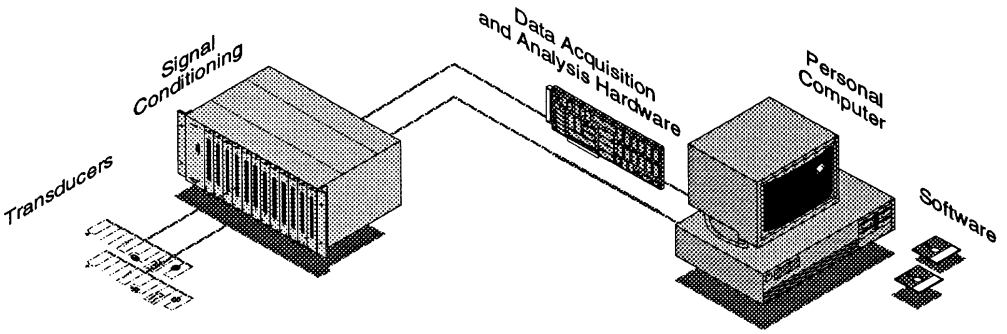


FIGURE 96.1 Components of a DAQ system.

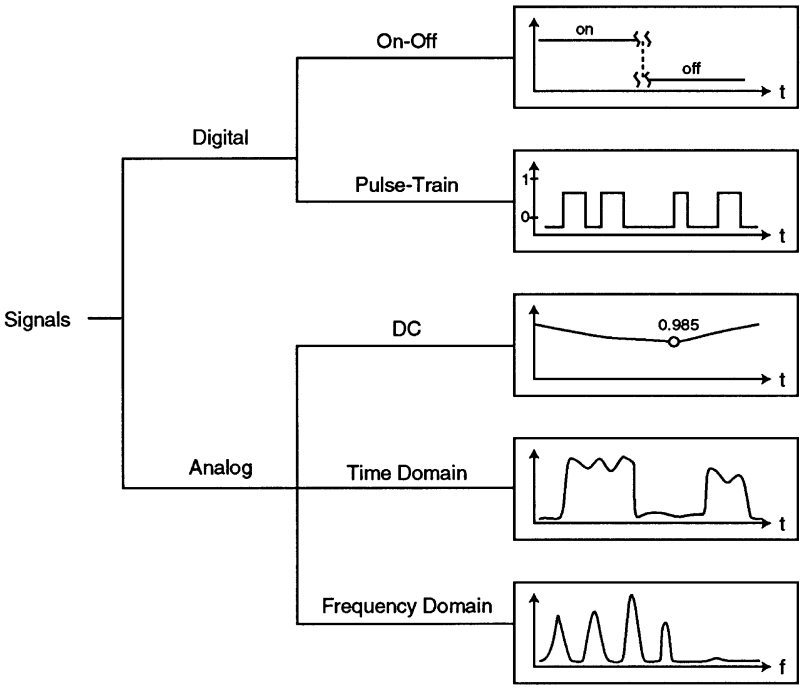


FIGURE 96.2 Classes of signals.

Plug-In DAQ Boards

The fundamental component of a DAQ system is the plug-in DAQ board. These boards plug directly into a slot in a PC and are available with analog, digital, and timing inputs and outputs (I/O). The most versatile of the plug-in DAQ boards is the multifunction I/O board. As the name implies, this board typically contains various combinations of analog-to-digital converters (ADCs), digital-to-analog converters (DACs), digital I/O lines, and counters/timers. ADCs and DACs measure and generate analog voltage signals, respectively. The digital I/O lines sense and control digital signals. Counters/timers measure pulse rates, widths, delays, and generate timing signals. These many features make the multifunction DAQ board useful for a wide range of applications.

Multifunction boards are commonly used to measure analog signals. This is done by the ADC, which converts the analog voltage level into a digital number that the computer can interpret. The analog multiplexer (MUX), the instrumentation amplifier, the sample-and-hold (S/H) circuitry, and the ADC compose the analog input section of a multifunction board (see Fig. 96.3).

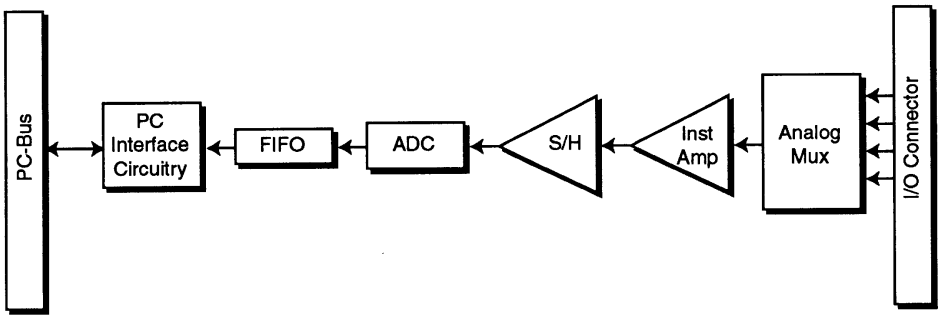


FIGURE 96.3 Analog input section of a plug-in DAQ board. *Note:* FIFO = first-in first-out buffer, S/H = sample-and-hold, Inst. Amp = instrumentation amplifier, and Mux = analog multiplexer.

Typically, multifunction DAQ boards have one ADC. Multiplexing is a common technique for measuring multiple channels (generally 16 single-ended or 8 differential) with a single ADC. The analog MUX switches between channels and passes the signal to the instrumentation amplifier and the S/H circuitry. The MUX architecture is the most common approach taken with plug-in DAQ boards. While plug-in boards typically include up to only 16 single-ended or 8 differential inputs, the number of analog input channels can be further expanded with external MUX accessories.

Instrumentation amplifiers typically provide a differential input and selectable gain by jumpers or software. The differential input rejects small common-mode voltages. The gain is often software programmable. In addition, many DAQ boards also include the capability to change the amplifier gain while scanning channels at high rates. Therefore, one can easily monitor signals with different ranges of amplitudes. The output of the amplifier is sampled, or held at a constant voltage, by the S/H device at measurement time so that voltage does not change during digitization.

The ADC transforms the analog signal into a digital value which is ultimately sent to computer memory. There are several important parameters of A/D conversion. The fundamental parameter of an ADC is the number of bits. The number of bits of an A/D determines the range of values for the binary output of the ADC conversion. For example, many ADCs are 12-bit, so a voltage within the input range of the ADC will produce a binary value that has one of $2^{12} = 4096$ different values. The more bits that an ADC has, the higher the resolution of the measurement. The resolution determines the smallest amount of change that can be detected by the ADC. Resolution is expressed as the number of digits of a voltmeter or dynamic range in decibels, rather than with bits. Table 96.6 shows the relation among bits, number of digits, and dynamic range in decibels.

TABLE 96.6 Relation Among Bits, Number of Digits, and Dynamic Range (dB)

Bits	Digits	dB
20	6.0	120
16	4.5	96
12	3.5	72
8	2.5	48

The resolution of the A/D conversion is also determined by the input range of the ADC and the gain. DAQ boards usually include an instrumentation amplifier that amplifies the analog signal by a gain factor prior to the conversion. This gain amplifies low-level signals so that more accurate measurements can be made.

Together, the input range of the ADC, the gain, and the number of bits of the board determine the minimum resolution of the measurement. For example, suppose a low-level ±30 mV signal is acquired

using a 12-bit ADC that has a ±5 V input range. If the system includes an amplifier with a gain of 100, the resulting resolution of the measurement will be range/(gain * 2^{bits}) = resolution, or 10 V/(100 * 2^{12}) = 0.0244 mV.

Finally, an important parameter of digitization is the rate at which A/D conversions are made, referred to as the sampling rate. The A/D system must be able to sample the input signal fast enough to measure the important waveform attributes accurately. In order to meet this criterion, the ADC must be able to convert the analog signal to digital form quickly enough.

When scanning multiple channels with a multiplexing DAQ system, other factors can affect the throughput of the system. Specifically, the instrumentation amplifier must be able to settle to the needed accuracy before the A/D conversion occurs. With multiplexed signals, multiple signals are being switched into one instrumentation amplifier. Most amplifiers, especially when amplifying the signals with larger gains, will not be able to settle to the full accuracy of the ADC when scanning channels at high rates. To avoid this situation, consult the specified settling times of the DAQ board for the gains and sampling rates required by the application.

Types of ADCs

Different DAQ boards use different types of ADCs to digitize the signal. The most popular type of ADC on plug-in DAQ boards is the successive approximation ADC, because it offers high speed and high resolution at a modest cost.

Subranging (also called half-flash) ADCs offer very high speed conversion with sampling speeds up to several million samples per second.

The state-of-the-art technology in ADCs is sigma–delta modulating ADCs. These ADCs sample at high rates, are able to achieve high resolution, and offer the best linearity of all ADCs.

Integrating and flash ADCs are mature technologies still used on DAQ boards today. Integrating ADCs are able to digitize with high resolution but must sacrifice sampling speed to obtain it. Flash ADCs are able to achieve the highest sampling rate (gigahertz) but are available only with low resolution. The different types of ADCs are summarized in Table 96.7.

TABLE 96.7 Types of ADCs

Type of ADC	Advantages	Features
Successive approximation	High resolution	1.25 MS/s sampling rate
	High speed	12-bit resolution
	Easily multiplexed	200 kS/s sampling rate
		16-bit resolution
Subranging	Higher speed	1 MHz sampling rate
		12-bit resolution
Sigma–delta	High resolution	48 kHz sampling rate
	Excellent linearity	16-bit resolution
	Built-in antialiasing	
	State-of-the-art technology	
Integrated	High resolution	15 kHz sampling rate
	Good noise rejection	
	Mature technology	
Flash	Highest speed	125 MHz sampling rate
	Mature technology	

Analog Input Architecture

With a typical DAQ board, the multiplexer switches among analog input channels. The analog signal on the channel selected by the multiplexer then passes to the programmable gain instrumentation amplifier (PGIA), which amplifies the signal. After the signal is amplified, the sample and hold (S/H) keeps the analog signal constant so that the ADC can determine the digital representation of the analog signal. A

good DAQ board will then place the digital signal in a first-in first-out (FIFO) buffer, so that no data will be lost if the sample cannot transfer immediately over the PC I/O channel to computer memory. Having a FIFO becomes especially important when the board is run under operating systems that have large interrupt latencies, such as Microsoft Windows.

Basic Analog Specifications

Almost every DAQ board data sheet specifies the number of channels, the maximum sampling rate, the resolution, and the input range and gain.

The number of channels, which is determined by the multiplexer, is usually specified in two forms — differential and single ended. Differential inputs are inputs that have different reference points for each channel, none of which is grounded by the board. Differential inputs are the best way to connect signals to the DAQ board because they provide the best noise immunity.

Single-ended inputs are inputs that are referenced to a common ground point. Because single-ended inputs are referenced to a common ground, they are not as good as differential inputs for rejecting noise. They do have a larger number of channels, however. Single-ended inputs are used when the input signals are high level (greater than 1 V), the leads from the signal source to the analog input hardware are short (less than 5 m), and all input signals share a common reference.

Some boards have pseudodifferential inputs which have all inputs referenced to the same common — like single-ended inputs — but the common is not referenced to ground. These boards have the benefit of a large number of input channels, like single-ended inputs, and the ability to remove some common-mode noise, especially if the common-mode noise is consistent across all channels. Differential inputs are still preferable to pseudodifferential, however, because differential is more immune to magnetic noise.

Sampling rate determines how fast the analog signal is converted to a digital signal. When measuring ac signals, sample at least two times faster than the highest frequency of the input signal. Even when measuring dc signals, oversample and average the data to increase the accuracy of the signal by reducing the effects of noise.

If the physical event consists of multiple dc-class signals, a DAQ board with interval scanning should be used. With interval scanning, all channels are scanned at one sample interval (usually the fastest rate of the board), with a second interval (usually slow) determining the time before repeating the scan. Interval scanning gives the effects of simultaneously sampling for slowly varying signals without requiring the additional cost of input circuitry for true simultaneous sampling.

Resolution is the number of bits that are used to represent the analog signal. The higher the resolution, the higher the number of divisions the input range is broken into, and therefore the smaller the possible detectable voltage. Unfortunately, some DAQ specifications are misleading when they specify the resolution associated with the DAQ board. Many DAQ board specifications state the resolution of the ADC without stating the linearities and noise, and therefore do not give the information needed to determine the resolution of the entire board. Resolution of the ADC, combined with the settling time, **integral nonlinearity** (INL), **differential nonlinearity** (DNL), and noise will give an understanding of the accuracy of the board.

Input range and gain determine the level of signal that should be connect to the board. Usually, the range and gain are specified separately, so the two must be combined to determine the actual signal input range as

$$\text{signal input range} = \text{range}/\text{gain}$$

For example, a board using an input range of ±10 V with a gain of 2 will have a signal input range of ±5 V. The closer the signal input range is to the range of the signal, the more accurate the readings from the DAQ board will be. If the signals have different input ranges, use a DAQ board with the feature of different gains per channel.

Data Acquisition Software

The software is often the most critical component of the DAQ system. Users of DAQ systems usually program the hardware in one of two ways — through register programming or through high-level device drivers.

Board Register-Level Programming

The first option is not to use vendor-supplied software and program the DAQ board at the hardware level. DAQ boards are typically register based; that is, they include a number of digital registers that control the operation of the board. The developer may use any standard programming language, such as C, C++, or Visual BASIC, to write series of binary codes to the DAQ board to control its operation. Although this method affords the highest level of flexibility, it is also the most difficult and time-consuming, especially for the inexperienced programmer. The programmer must know the details of programming all hardware, including the board, the PC interrupt controller, the DMA controller, and PC memory.

Driver Software

Driver software typically consists of a library of function calls usable from a standard programming language. These function calls provide a high-level interface to control the standard functions of the plug-in board. For example, a function called SCAN_OP may configure, initiate, and complete a multiple-channel scanning DAQ operation of a predetermined number of points. The function call would include parameters to indicate the channels to be scanned, the amplifier gains to be used, the sampling rate, and the total number of data points to be collected. The driver responds to this one function call by programming the plug-in board, the DMA controller, the interrupt controller, and CPU to scan the channels as requested.

What Is Digital Sampling?

Every DAQ system has the task of gathering information about analog signals. To do this, the system captures a series of instantaneous "snapshots" or samples of the signal at definite time intervals. Each sample contains information about the signal at a specific instant. Knowing the exact time of each conversion and the value of the sample, one can reconstruct, analyze, and display the digitized waveform.

Real-Time Sampling Techniques

In real-time sampling, the DAQ board digitizes consecutive samples along the signal (Fig. 96.4). According to the **Nyquist sampling theorem**, the ADC must sample at least twice the rate of the maximum frequency component in that signal to prevent aliasing. **Aliasing** is a false lower-frequency component that appears in sampled data acquired at too low a sampling rate. The frequency at one half the sampling frequency is referred to as the Nyquist frequency. Theoretically, it is possible to recover information about those signals with frequencies at or below the Nyquist frequency. Frequencies above the Nyquist frequency will alias to appear between dc and the Nyquist frequency.

For example, assume the sampling frequency, f_s, is 100 Hz. Also assume the input signal to be sampled contains the following frequencies — 25, 70, 160, and 510 Hz. Figure 96.5 shows a spectral representation of the input signal.

The mathematics of sampling theory show us that a sampled signal is shifted in the frequency domain by an amount equal to integer multiples of the sampling frequency, f_s. Figure 96.6 shows the spectral content of the input signal after sampling. Frequencies below 50 Hz, the Nyquist frequency ($f_s/2$), appear correctly. However, frequencies above the Nyquist appear as aliases below the Nyquist frequency. For example, F1 appears correctly; however, F2, F3, and F4 have aliases at 30, 40, and 10 Hz, respectively.

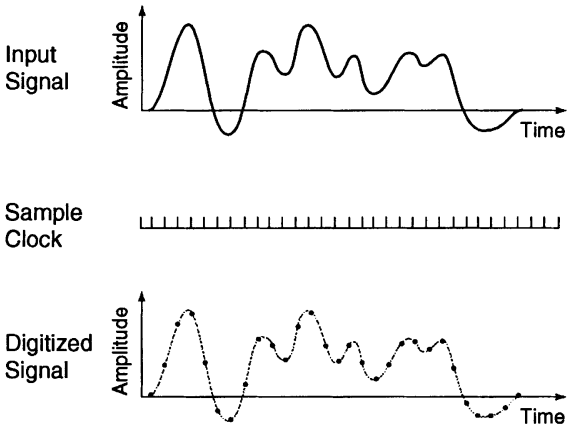


FIGURE 96.4 Consecutive discrete samples recreate the input signal.

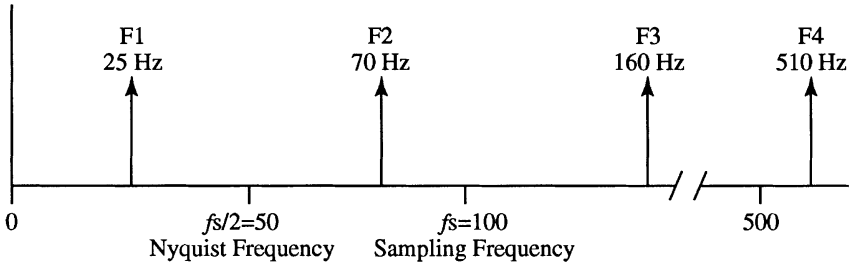


FIGURE 96.5 Spectral of signal with multiple frequencies.

The resulting frequency of aliased signals can be calculated with the following formula:

Apparent (Alias) Freq. = ABS (Closest Integer Multiple of Sampling Freq. – Input Freq.)

For the example of Figs. 96.5 and 96.6:

$$\text{Alias F2} = |100 - 70| = 30 \text{ Hz}$$

$$\text{Alias F3} = |(2)100 - 160| = 40 \text{ Hz}$$

$$\text{Alias F4} = |(5)100 - 510| = 10 \text{ Hz}$$

Preventing Aliasing

Aliasing can be prevented by using filters on the front end of the DAQ system. These antialiasing filters are set to cut off any frequencies above the Nyquist frequency (half the sampling rate). The perfect filter would reject all frequencies above the Nyquist; however, because perfect filters exist only in textbooks, one must compromise between sampling rate and selecting filters. In many applications, one- or two-pole passive filters are satisfactory. The rule of thumb is to oversample (5 to 10 times) and use these antialiasing filters when frequency information is crucial.

Alternatively, active antialiasing filters with programmable cutoff frequencies and very sharp attenuation of frequencies above the cutoff can be used. Because these filters exhibit a very steep roll-off, the DAQ system can sample at two to three times the filter cutoff frequency. Figure 96.7 shows a transfer function of a high-quality antialiasing filter.

Alias F2 = |100 - 70| = 30 Hz
Alias F3 = |(2)100 - 160| = 40 Hz
Alias F4 = |(5)100 - 510| = 10 Hz

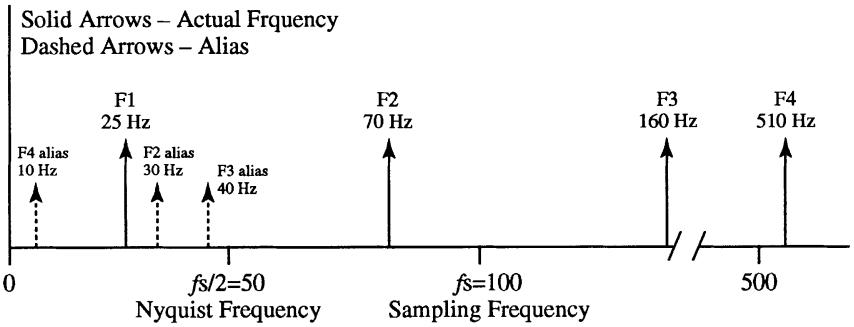


FIGURE 96.6 Spectral of signal with multiple frequencies after sampling at f_s = 100 Hz.

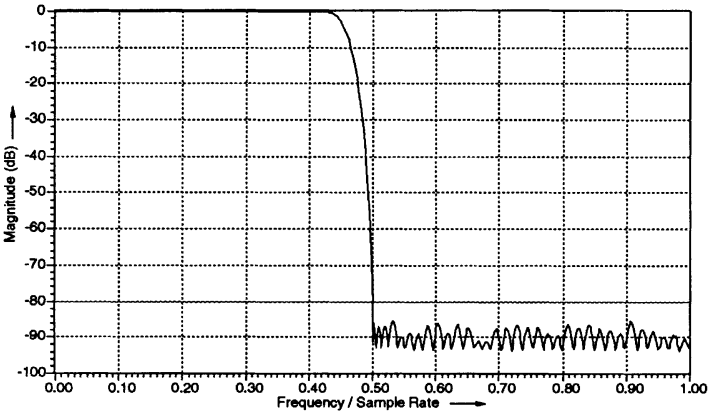


FIGURE 96.7 Magnitude portion of transfer function of an antialiasing filter.

The computer uses digital values to recreate or to analyze the waveform. Because the signal could be anything between each sample, the DAQ board may be unaware of any changes in the signal between samples. There are several sampling methods optimized for the different classes of data; they include software polling, external sampling, continuous scanning, multirate scanning, simultaneous sampling, interval scanning, and seamless changing of the sample rate.

Software Polling

A software loop polls a timing signal and starts the A/D conversion via a software command when the edge of the timing signal is detected. The timing signal may originate from the internal clock of the computer or from a clock on the DAQ board. Software polling is useful in simple, low-speed applications, such as temperature measurements.

The software loop must be fast enough to detect the timing signal and trigger a conversion. Otherwise, a window of uncertainty, also known as jitter, will exist between two successive samples. Within the window of uncertainty, the input waveform could change enough to reduce the accuracy of the ADC drastically.

Suppose a 100-Hz, 10-V full-scale sine wave is digitized (Fig. 96.8). If the polling loop takes 5 ms to detect the timing signal and to trigger a conversion, then the voltage of the input sine wave will change

as much as 31 mV, [$\Delta V = 10 \sin (2\pi \times 100 \times 5 \times 10^{-6})$]. For a 12-bit ADC operating over an input range of 10 V and a gain of 1, one least significant bit (LSB) of error represents 2.44 mV:

$$\left(\frac{\text{Input range}}{\text{gain} \times 2^n}\right) = \left(\frac{10\text{ V}}{1 \times 2^{12}}\right) = 2.44\text{ mV}$$

But because the voltage error due to jitter is 31 mV, the accuracy error is 13 LSB.

$$\left(\frac{31\text{ mV}}{2.44\text{ mV}}\right)$$

This represents uncertainty in the last 4 bits of a 12-bit ADC. Thus, the effective accuracy of the system is no longer 12 bits but rather 8 bits.

External Sampling

Some DAQ applications must perform a conversion based on another physical event that triggers the data conversion. The event could be a pulse from an optical encoder measuring the rotation of a cylinder. A sample would be taken every time the encoder generates a pulse corresponding to *n* degrees of rotation. External triggering is advantageous when trying to measure signals whose occurrence is relative to another physical phenomenon.

Continuous Scanning

When a DAQ board acquires data, several components on the board convert the analog signal to a digital value. These components include the analog MUX, the instrumentation amplifier, the S/H circuitry, and the ADC. When acquiring data from several input channels, the analog MUX connects each signal to the ADC at a constant rate. This method, known as continuous scanning, is significantly less expensive than having a separate amplifier and ADC for each input channel.

Continuous scanning is advantageous because it eliminates jitter and is easy to implement. However, it is not possible to sample multiple channels simultaneously. Because the MUX switches between channels, a time skew occurs between any two successive channel samples. Continuous scanning is appropriate for applications where the time relationship between each sampled point is unimportant or where the skew is relatively negligible compared with the speed of the channel scan.

If samples from two signals are used to generate a third value, then continuous scanning can lead to significant errors if the time skew is large. In Fig. 96.9, two channels are continuously sampled and added together to produce a third value. Because the two sine waves are 90° out-of-phase, the sum of the signals should always be zero. But because of the skew time between the samples, an erroneous sawtooth signal results.

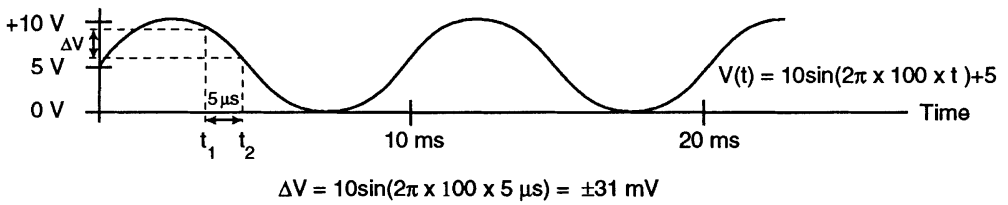


FIGURE 96.8 Jitter reduces the effective accuracy of the DAQ board.

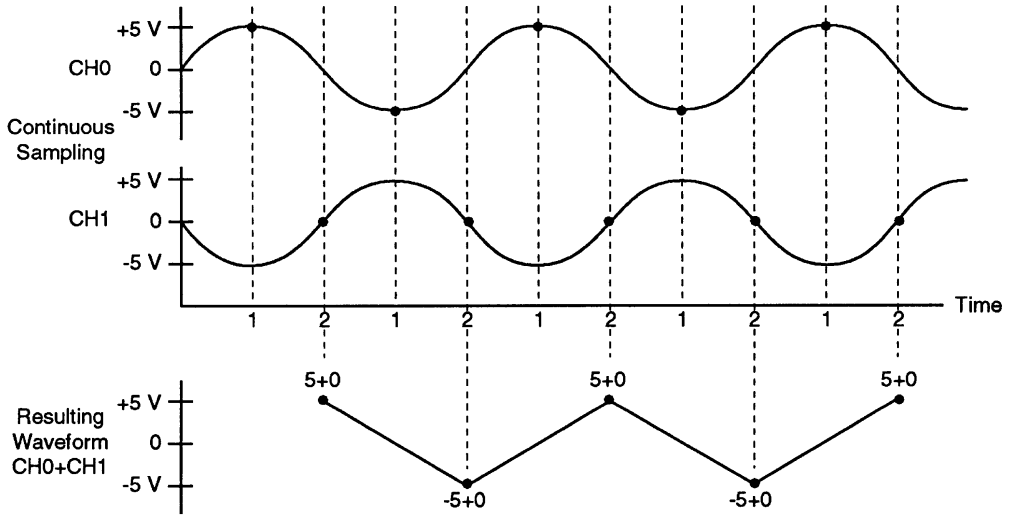


FIGURE 96.9 If the channel skew is large compared with the signal, then erroneous conclusions may result.

Multirate Scanning

Multirate scanning, a method that scans multiple channels at different scan rates, is a special case of continuous scanning. Applications that digitize multiple signals with a variety of frequencies use multirate scanning to minimize the amount of buffer space needed to store the sampled signals. Channel-independent ADCs are used to implement hardware multirate scanning; however, this method is extremely expensive. Instead of multiple ADCs, only one ADC is used. A channel/gain configuration register stores the scan rate per channel and software divides down the scan clock based on the per-channel scan rate. Software-controlled multirate scanning works by sampling each input channel at a rate that is a fraction of the specified scan rate.

Suppose the system scans channels 0 through 3 at 10 kS/s, channel 4 at 5 kS/s, and channels 5 through 7 at 1 kS/s. A base scan rate of 10 kS/s should be used. Channels 0 through 3 are acquired at the base scan rate. Software and hardware divide the base scan rate by 2 to sample channel 4 at 5 kS/s, and by 10 to sample channels 5 through 7 at 1 kS/s.

Simultaneous Sampling

For applications where the time relationship between the input signals is important, such as phase analysis of ac signals, simultaneous sampling must be used. DAQ boards capable of simultaneous sampling typically use independent instrumentation amplifiers and S/H circuitry for each input channel, along with an analog MUX, which routes the input signals to the ADC for conversion (as shown in Fig. 96.10).

To demonstrate the need for a simultaneous-sampling DAQ board, consider a system consisting of four 50-kHz input signals sampled at 200 kS/s. If the DAQ board uses continuous scanning, the skew between each channel is 5 μs (1S/200 kS/s) which represents a 270° [(15 μs/20 μs) × 360°] shift in phase between the first channel and fourth channel. Alternatively, with a simultaneous-sampling board with a maximum 5 ns interchannel time offset, the phase shift is only 0.09° [(5 μs/20 μs) × 360°]. This phenomenon is illustrated in Fig. 96.11.

Interval Scanning

For low-frequency signals, interval scanning creates the effect of simultaneous sampling, yet maintains the cost benefits of a continuous-scanning system. This method scans the input channels at one rate and

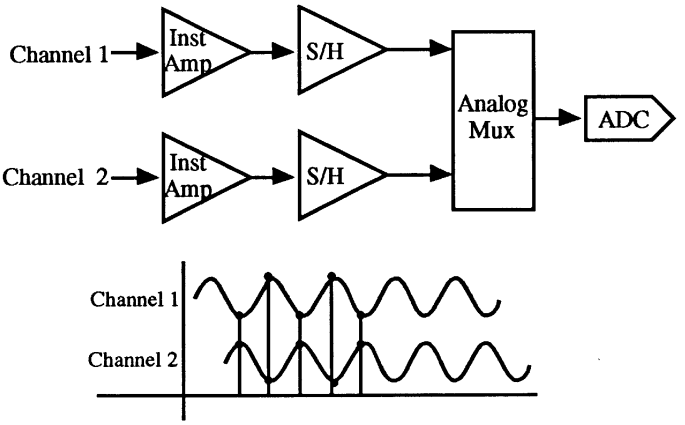


FIGURE 96.10 Block diagram of DAQ components used to sample multiple channels simultaneously.

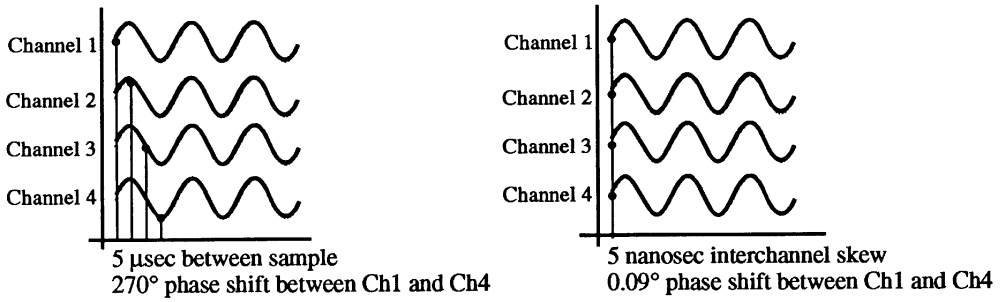


FIGURE 96.11 Comparison of continuous scanning and simultaneous sampling.

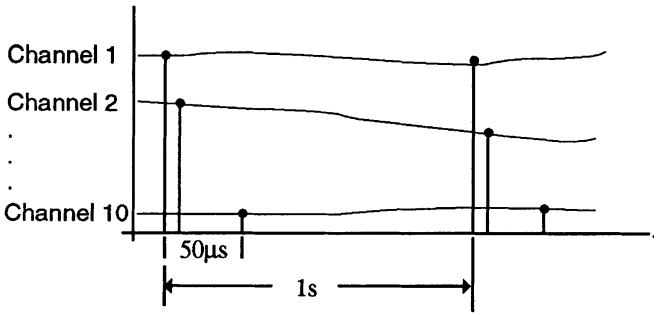


FIGURE 96.12 Interval scanning — all ten channels are scanned within 45 µs; this is insignificant relative to the overall acquisition rate of 1 S/s.

uses a second rate to control when the next scan begins. If the input channels are scanned at the fastest rate of the ADC, the effect of simultaneously sampling the channels is created. Interval scanning is appropriate for slow-moving signals, such as temperature and pressure. Interval scanning results in a jitter-free sample rate and minimal skew time between channel samples. For example, consider a DAQ system with ten temperature signals. By using interval scanning, a DAQ board can be set up to scan all channels with an interchannel delay of 5 µs, then repeat the scan every second. This method creates the effect of simultaneously sampling ten channels at 1 S/s, as shown in Fig. 96.12.

To illustrate the difference between continuous and interval scanning, consider an application that monitors the torque and RPMs of an automobile engine and computes the engine horsepower. Two signals, proportional to torque and RPM, are easily sampled by a DAQ board at a rate of 1000 S/s. The values are multiplied together to determine the horsepower as a function of time.

A continuously scanning DAQ board must sample at an aggregate rate of 2000 S/s. The time between which the torque signal is sampled and the RPM signal is sampled will always be 0.5 ms (1/2000). If either signal changes within 0.5 ms, then the calculated horsepower is incorrect. But using interval scanning at a rate of 1000 S/s, the DAQ board samples the torque signal every 1 ms, and the RPM signal is sampled as quickly as possible after the torque is sampled. If a 5-μs interchannel delay exists between the torque and RPM samples, then the time skew is reduced by 99% [(0.5 ms – 5 μs)/0.5 ms], and the chance of an incorrect calculation is reduced.

Factors Influencing the Accuracy of Measurements

How does one determine if a plug-in DAQ will deliver the required measurement results? With a sophisticated measuring device like a plug-in DAQ board, significantly different accuracies can be obtained depending on the type of board used. For example, one can purchase DAQ products on the market today with 16-bit ADCs and get less than 12 bits of useful data, or one can purchase a product with a 16-bit ADC and actually get 16 bits of useful data. This difference in accuracies causes confusion in the PC industry where everyone is used to switching out PCs, video cards, printers, and so on, and experiencing similar results between equipment.

The most important thing to do is to scrutinize more specifications than the resolution of the ADC that is used on the DAQ board. For dc-class measurements, one should at least consider the settling time of the instrumentation amplifier, DNL, **relative accuracy**, INL, and noise. If the manufacturer of the board under consideration does not supply these specifications in the data sheets, ask the vendor to provide them or run tests to determine these specifications.

Defining Terms

Alias: A false lower frequency component that appears in sampled data acquired at too low a sampling rate.

Asynchronous: (1) Hardware — A property of an event that occurs at an arbitrary time, without synchronization to a reference clock. (2) Software — A property of a function that begins an operation and returns prior to the completion or termination of the operation.

Conversion time: The time required, in an analog input or output system, from the moment a channel is interrogated (such as with a read instruction) to the moment that accurate data are available.

DAQ (data acquisition): (1) Collecting and measuring electric signals from sensors, transducers, and test probes or fixtures and inputting them to a computer for processing: (2) Collecting and measuring the same kinds of electric signals with ADC and/or DIO boards plugged into a PC, and possibly generating control signals with DAC and/or DIO boards in the same PC.

DNL (differential nonlinearity): A measure in LSB of the worst-case deviation of code widths from their ideal value of 1 LSB.

INL (integral nonlinearity): A measure in LSB of the worst-case deviation from the ideal A/D or D/A transfer characteristic of the analog I/O circuitry.

Nyquist sampling theorem: A law of sampling theory stating that if a continuous bandwidth-limited signal contains no frequency components higher than half the frequency at which it is sampled, then the original signal can be recovered without distortion.

Relative accuracy: A measure in LSB of the accuracy of an ADC. It includes all nonlinearity and quantization errors. It does not include offset and gain errors of the circuitry feeding the ADC.

Further Information

House, R., "Understanding Important DA Specifications," *Sensors*, 10(10), June 1993.

House, R., "Understanding Inaccuracies Due to Settling Time, Linearity, and Noise," *National Instruments European User Symposium Proceedings*, November 10–11, 1994, pp. 11–12.

McConnell, E., "PC-Based Data Acquisition Users Face Numerous Challenges," *ECN*, August 1994.

McConnell, E., "Choosing a Data-Acquisition Method," *Electronic Design*, 43(6), 147, 1995.
McConnell, E. and Jernigan, Dave, "Data Acquisition," in *The Electronics Handbook*, J.C. Whitaker (ed.), Boca Raton, FL: CRC Press, 1996, 1795–1822.
Potter, D. and A. Razdan, "Fundamentals of PC-Based Data Acquisition," *Sensors*, 11(2), 12–20, February 1994.
Potter, D., "Sensor to PC — Avoiding Some Common Pitfalls," *Sensors Expo Proceedings*, September 20, 1994.
Potter, D., "Signal Conditioners Expand DAQ System Capabilities," *I&CS*, 25–33, August 1995.
Johnson, G. W., *LabVIEW Graphical Programming*, New York: McGraw-Hill, 1994.
McConnell, E., "New Achievements in Counter/Timer Data Acquisition Technology," *MessComp 1994 Proceedings*, September 13–15, 1994, 492–498.
McConnell, E., "Equivalent Time Sampling Extends DA Performance," *Sensors Data Acquisition*, Special Issue, June, 13, 1995.

96.3 Magnetic and Optical Recorders

Yufeng Li

The heart of recording technology is for the process of information storage and retrieval. In addition to its obvious importance in different branches of science and engineering, it has become indispensable to our daily life. When we make a bank transaction, reserve an airplane ticket, use a credit card, watch a movie from a video tape, or listen to music from a CD, we are using the technology of recording. The general requirements for recording are information integrity, fast access, and low cost. Among the different techniques, the most popularly used ones are magnetic and optical recording.

Typical recording equipment consists of a read/write head, a medium, a coding/decoding system, a data access system, and some auxiliary mechanical and electronic components. The head and medium are for data storage and retrieval purposes, and the coding/decoding system is for data error correction. The data access system changes the relative position between the head and the medium, usually with a servo mechanism for data track following and a spinning mechanism for on-track moving. While the data access system and the auxiliary components are important to recording equipment, they are not considered essential in this chapter to the understanding of recording technology, and will not be covered. Interested readers are referred to Reference 1.

Magnetic Recording

At present, magnetic recording technology dominates the recording industry. It is used in the forms of hard disk, floppy disk, removable disk, and tape with either digital or analog mode. In its simplest form, it consists of a magnetic head and a magnetic medium, as shown in Fig. 96.13. The head is made of a piece of magnetic material in a ring shape (core), with a small gap facing the medium and a coil away from the medium. The head records (writes) and reproduces (reads) information, while the medium stores the information. The recording process is based on the phenomenon that an electric current i generates a magnetic flux ϕ as described by Ampere's law. The flux ϕ leaks out of the head core at the gap, and magnetizes the magnetic medium which moves from left to right with a velocity V under the head gap. Depending on the direction of the electric current i, the medium is magnetized with magnetization M pointing either left or right. This pattern of magnetization is retained in the memory of the medium even after the head moves away.

Two types of head may be used for reproducing. One, termed the *inductive head*, senses magnetic flux change rate, and the other, named the *magnetoresistive* (MR) *head*, senses the magnetic flux. When an inductive head is used, the reproducing process is just the reverse of the recording process. The flux coming out of the magnetized medium surface is picked up by the head core. Because the medium magnetization under the head gap changes its magnitude and direction as the medium moves, an electric

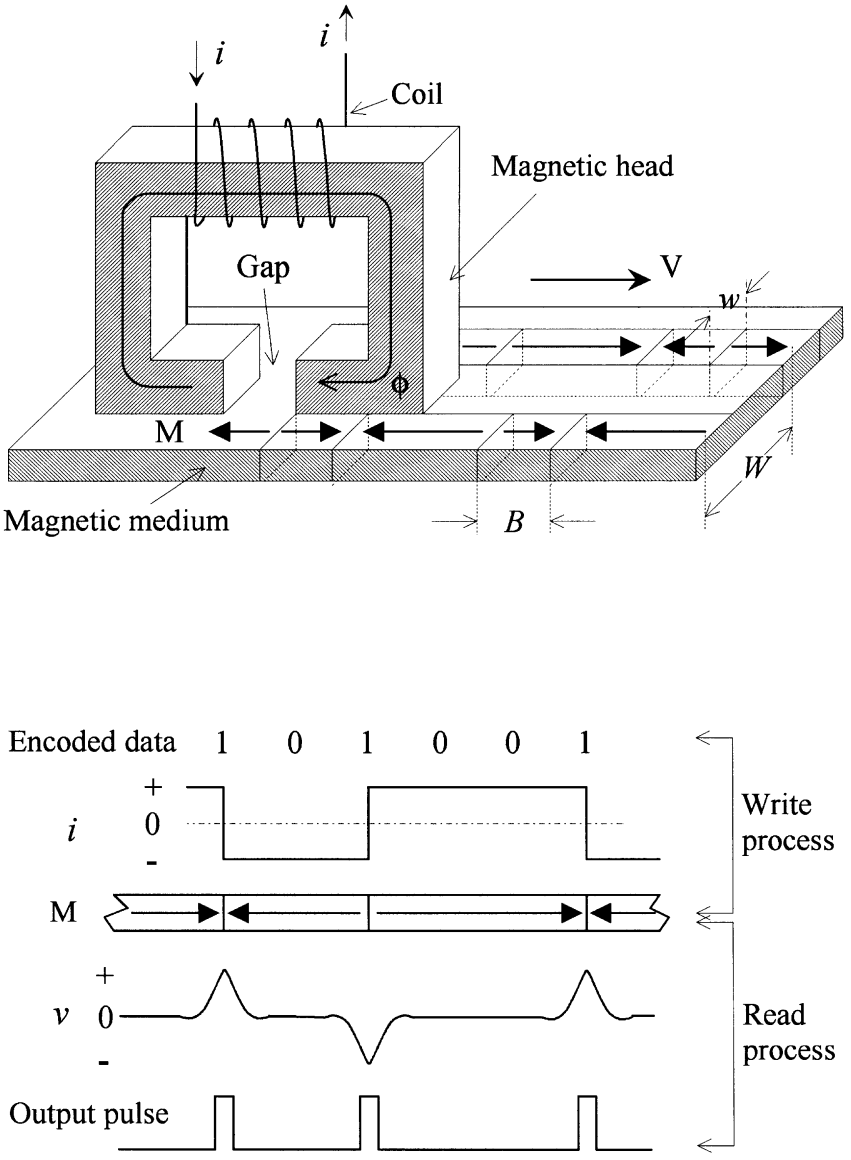


FIGURE 96.13 Conceptual diagrams illustrating the magnetic recording principle (a), and recording/reproducing process (b).

voltage is generated in the coil. This process is governed by Faraday's law. Figure 96.13b schematically shows the digital recording/reproducing process. First, all user data are encoded into a binary format — a serial of 1s and 0s. Then a write current *i* is sent to the coil. This current changes its direction whenever a 1 is being written. Correspondingly, a change of magnetization, termed a *transition*, is recorded in the medium for each 1 in the encoded data. During the reproducing process, the electric voltage induced in the head coil reaches a peak whenever there is a transition in the medium. A pulse detector generates a pulse for each transition. These pulses are decoded to yield the user data.

The minimum distance between two transitions in the medium is the flux change length *B*, and the distance between two adjacent signal tracks is the track pitch *W*, which is wider than the signal track width *w*. The flux change length can be directly converted into bit length with the proper code information. The reciprocal of the bit length is called *linear density*, and the reciprocal of the track pitch is termed

track density. The information storage areal density in the medium is the product of the linear density and the track density. This areal density roughly determines how much information a user can store in a unit surface area of storage medium, and is a figure of merit for a recording technique. Much effort has been expended to increase the areal density. For example, it has been increased 50 times during the last decade in hard disk drives, and is expected to continue increasing 60% per year in the foreseeable future. At present, state-of-the-art hard disk products feature areal densities of more than 7 Mbits/mm^2 ($B < 0.1$ μm and $W < 1.5$ μm). This gives a total storage capacity of up to 6 Gbytes for a disk of 95 mm diameter.

Magnetism and Hysteresis Loop

Magnetism is the result of uncompensated electron spin motions in an atom. Only transition elements exhibit this property, and nearly all practical interest in magnetism centers on the first transition group of elements (Mn, Cr, Fe, Ni, and Co) and their alloys. The strength of magnetism is represented by magnetization M, and is related to magnetic field H and magnetic flux density B by

$$B = \mu_0(H + M) \tag{96.1}$$

where μ_0 is the permeability of vacuum. Since M is a property of a magnetic material, it does not exist outside the magnetic material. H represents the strength acting on a magnetic material from a magnetic field which is generated either by a magnetic material or by an electric current. B is the flux density which determines the induced electric voltage in a coil. The ratio of B with and without a magnetic material is the relative permeability μ of that magnetic material.

When a magnetic field H is applied to a piece of demagnetized magnetic material, the magnetization M starts increasing with H from zero. The rate of increase gradually slows down and M asymptotically approaches a value M_s at high H. If H is reduced to zero, then M is reduced to a lower value M_r. Continuous reduction of H to a very high negative value will magnetize the material to $-M_s$. In order to bring the material to demagnetized state, a positive field H_c is required. Further increase in the H field will bring the trace of M to a closed loop. This loop is the major hysteresis loop, as shown in Fig. 96.14. The hysteresis loop shows that a magnetic material has memory. It is this memory that is used in the medium for storing information. H_c is the coercivity, indicating the strength of magnetic field required to erase the memory of a magnetic material. Magnetic materials with high H_c are "hard" magnets, and are suitable for medium applications if they have high M_r. On the other hand, magnetic materials with low H_c are "soft" magnets, and are candidates for head core materials if they have high M_s and high μ. M_r and M_s are the remanent and saturation magnetization, respectively, and their ratio is the remanent squareness. The flux density corresponding to M_s is B_s.

Magnetic Media

Magnetic media are used to store information in a magnetic recording system. In order to increase the areal density, we need to reduce flux change length B and track width w. Since B is limited by the term $M_r\delta/H_c$, where δ is the magnetic layer thickness, we can reduce B by either decreasing $M_r\delta$ or increasing H_c. However, the amplitude of the magnetic signal available for reproducing head is proportional to the term $M_r\delta w$. If we reduce track width w to increase areal density, we must increase $M_r\delta$ to avoid signal deterioration. In addition, if the magnetic layer is so thin that it causes thickness nonuniformity, more noise will appear in the reproducing process. Therefore, the major requirements for magnetic layer are high H_c, high M_r, and ease of making a uniform thin layer. Additional requirements include good magnetic and mechanical stability.

There are two groups of magnetic media. The first group is called particulate media because the magnetic materials are in the form of particles. This group includes iron oxide (γ-Fe_2O_3), cobalt-modified iron oxide (γ-Fe_2O_3+Co), chromium dioxide (CrO_2), metal particles, and barium ferrite ($BaFe_{12}O_{19}$). Some of these have been used in the magnetic recording for several decades. More recently, another group of media has been developed largely due to the ever-increasing demand for higher storage capacity

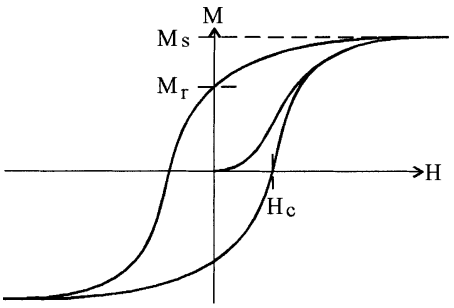


FIGURE 96.14 Hysteresis loop of a magnetic material shows the nonlinear relationship between M and H which results in magnetic memory.

TABLE 96.8 Remanence (M_r) and Coercivity (H_c) Values of Some Commonly Used Magnetic Media (some values are from Reference 5)

Group	Material	M_r (kA/m)	H_c (kA/m)	Application
Particulate	γ-Fe_2O_3	56–140	23–32	Floppy disk, audio, video, and instrumentation tapes
	γ-Fe_2O_3+Co	60–140	44–74	Floppy disk, audio, video, and instrumentation tapes
	CrO_2	110–140	38–58	Floppy disk, audio, video, and instrumentation tapes
	$BaFe_{12}O_{19}$	56	58	Floppy disk
Thin film	Co–Ni	600–1100	30–85	Hard disk
	Co–Fe	1100–1500	60–150	Hard disk
	Co–P	600–1000	36–120	Hard disk
	Co–Ni–Pt	600–1100	60–175	Hard disk
	Co–Cr–Ta	350–900	55–190	Hard disk
	Co–Cr–Pt	300–750	56–200	Hard disk

in the computer industry. This group of media is the thin-film media, where the magnetic layer can be made as a continuous thin film. Most materials in this group are cobalt-based metal alloys. Compared with particulate media, the thin-film media usually have a higher coercivity H_c, a higher remanence M_r, and can be deposited in a very thin continuous film. Table 96.8 lists H_c and M_r for some of the most popularly used particulate and thin-film media. Note that magnetic properties are affected by the fabrication process and film structure. Therefore, their values can be out of the ranges of Table 96.8 if different processes are used.

Magnetic media can be classified into three general forms of applications. Tape is the oldest form and remains an important medium today. It is central to most audio, video, and instrumentation recording, although it is also used in the computer industry for archival storage. Tape is economical and can hold a large capacity, but suffers slow access time. Hard disk is primarily used as the storage inside a computer, providing fast data access for the user, but having poor transportability. Flexible disk is designed for easy data transportation, but is limited in capacity. Besides these three general forms of applications, a hybrid of flexible and hard disk is being gradually accepted. It is a removable rigid disk capable of holding up to several gigabytes of digital data. In addition, magnetic stripes are getting wide use in different forms of cards.

The magnetic layer alone cannot be used as a medium. It needs additional components to improve its chemical and mechanical durability. Typical cross sections of a particulate magnetic tape and a thin-film hard disk are shown in Fig. 96.15. In the case of tape application, iron particles with typical size of 0.5 μm long and 0.1 μm wide are dispersed in a polymeric binder, together with solvents, lubricants, and other fillers to improve magnetic and mechanical stability. This dispersed material is then coated on an abiaxially oriented polyethylene terephthalate substrate. An optional back coat may also be applied to the other side of the substrate. The cross section of a hard disk is more complex. A high-purity aluminum–magnesium (5 wt%) substrate is diamond turned to a fine surface finish, and then electrolessly plated with a nonmagnetic nickel–phosphorus (10 at%) undercoat. This layer is used to increase the

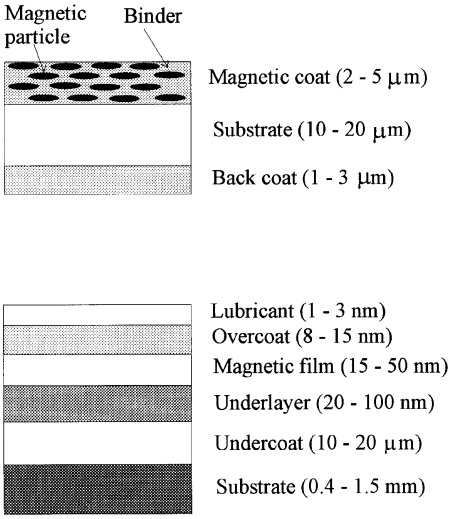


FIGURE 96.15 Cross-sectional views of a particulate magnetic tape (top) and a thin film hard disk (bottom).

TABLE 96.9 Relative Permeability (μ), Saturation Flux Density (B_s), Coercivity (H_c) and Resistivity (ρ) Values of Some Commonly Used Magnetic Head Materials at Low Frequency (some values are from Reference 5)

Material	μ	B_s (T)	H_c (A/m)	ρ (μΩcm)	Application
Ni–Fe–Mo	11000	0.8	2.0	100	Audio tape
Ni–Zn	300–1500	0.4–0.46	11.8–27.6	10^{11}	Floppy and hard disk drives, video and instrumentation tapes
Mn–Zn	3000–10000	0.4–0.6	11.8–15.8	10^{6}	Floppy and hard disk drives, video and instrumentation tapes
Fe–Si–Al	8000	1.0	2.0	85	Floppy and hard disk drives, video and instrumentation tapes
Ni–Fe	2000–4000	1.0	<10	20	Hard disk drives

hardness, reduce the defects, and improve the finish of the Al–Mg alloy, and is polished to a super surface finish. Next, an underlayer of chromium is sputtered to control the properties of the magnetic film, followed by sputtering the magnetic film. Finally, a layer of hydrogenated or nitrogenated carbon is overcoated on the magnetic film, and an ultrathin layer of perfluorinated hydrocarbon liquid lubricant is applied on top. The carbon and lubricant layers are used to improve the corrosion and mechanical resistance of the disk. For a 95-mm disk the finished product should have a surface flatness better than 10 μm and a tightly control surface roughness. In some applications, an arithmetic average roughness (R_a) of less than 0.5 nm is required.

Magnetic Heads

Magnetic heads have three functions: recording, reproducing, and erasing. Usually for stationary head applications such as tape drives, multiple heads are used to perform these functions. For moving head applications such as disk drives, a single head is employed because of the requirements of simple connections and small head mass for fast data access. Most of these heads are the inductive type, where the fundamental design is an inductive coil and a magnetic core. The general requirements for the core materials are high relative permeability μ, high saturation flux density B_s, low coercivity H_c, high electric resistivity ρ, and low magnetostriction coefficient λ. Some of the properties for the commonly used core materials are listed in Table 96.9.

The evolution of the magnetic head follows the selection of core materials, as shown in Fig. 96.16. Early heads used laminated molybdebum Permalloy (Ni-Fe-Mo, 79-17-4 wt%). These heads are inex-

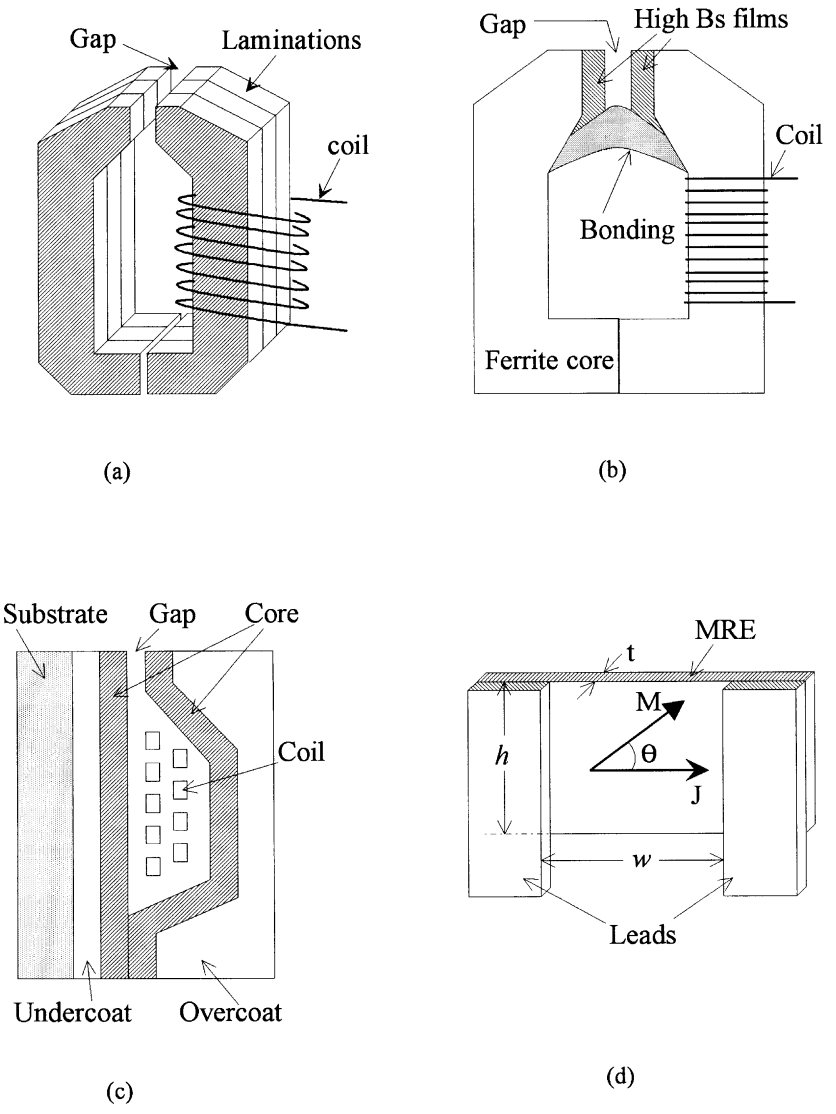


FIGURE 96.16 Schematic illustrations of (a) a laminated head, (b) cross-section of a MIG head, (c) cross-section of a thin film head, and (d) an MR sensor with leads.

pensive to make, and have low H_c and high μ and B_s. The primary drawbacks are frequency limitation, gap dimension inaccuracy, and mechanical softness. Frequency limitation is caused by the difficulty of making the lamination layer thinner than 25 μm. Eddy current loss, which is proportional to layer thickness and square root of frequency, reduces the effective permeability. As a result, laminated heads are seldom used for applications exceeding 10 MHz. Gap dimension inaccuracy is associated with the head fabrication process, and makes it unsuitable for high areal density applications. Lack of mechanical hardness reduces its usable life.

One way to reduce eddy current loss is to increase core material electric resistivity. Two types of ferrite material have high resistivity (four to nine orders higher than Permalloy) and reasonable magnetic properties: Ni–Zn and Mn–Zn. These materials are also very hard, elongating head life during head/medium contacts. The major deficiency of ferrite materials is their low B_s values. In order to record in high H_c media, high flux density B is needed in the head core. When the flux density in the core

material reaches its saturation B_s, it will not increase despite the increase of recording current or coil turns. This saturation starts from the corners of the gap due to its geometry. To remedy this deficiency, a layer of metallic alloy material with much higher B_s is deposited on the gap faces. This type of head is called the metal-in-gap (MIG) head. Sendust (Fe–Si–Al, 85–9.6–5.4 wt%) is one of the materials used for the deposition. MIG heads are capable of recording up to 100 MHz frequency and 180 kA/m medium coercivity.

Thin-film heads capitalize on semiconductor-like processing technology to reduce the customized fabrication steps for individual heads. The core, coil, gap, and insulator layers are all fabricated by electroplating, sputtering, or evaporation. Due to the nature of the semiconductor process, the fabrication is accurate for small dimensions. Small gap dimensions are suitable for high linear and track density, and small core dimensions allow the use of high B_s Permalloy material (Ni–Fe, 80–20 wt%) as core with low inductance for high data rate applications. Thin-film heads are used for high medium H_c, high areal density applications. The high cost of the semiconductor-like process is offset by high throughput: a 150 $\times$ 150 mm wafer can produce 16,000 nanoslider heads. One disadvantage is the limited-band recording capability because the small pole length limits low-frequency response and introduces undershoots. A second disadvantage the Barkhausen noise, which is caused by the relatively small number of magnetic domains in the core. At present, thin-film heads are used up to frequencies of 80 MHz and medium coercivity of 200 kA/m. MIG thin-film heads are also being used for high-coercivity applications.

An inductive head is often used for both recording and reproducing. The optimal performance cannot be achieved because recording and reproducing have contradictory requirements for head design. To solve this problem, the MR head has been developed. The MR head is for reproducing only, and an inductive head is used for recording. As schematically shown in Fig. 96.16, an MR head has a magnetoresistive element (MRE) and two electric leads. The MRE is a Permalloy stripe (Ni–Fe, 80–20 wt%), with thickness t, width w, and height h. An electric current, with density J, passes through the MRE through the leads. The electric resistivity of the MRE is a function of the angle θ between J and MRE magnetization M:

$$\rho_\theta = \rho\left(1 + \frac{\Delta\rho}{\rho}\cos^2\theta\right) \tag{96.2}$$

where ρ_θ is the resistivity at θ, ρ is the resistivity at $\theta = 90°$, and $\Delta\rho$ is the resistivity difference between $\theta = 0°$ and $\theta = 90°$. $\Delta\rho/\rho$ is the anisotropic MR ratio (AMR) of the MRE. Usually a transverse magnetic field is applied to the MRE so that $\theta = \theta_0$ when the head is not reading a magnetic signal. Assume that a magnetic signal from the medium rotates M from θ_0 to θ, then an electric voltage change v will be detected across the MRE as the output signal:

$$v = Jw\rho\frac{\Delta\rho}{\rho}\left(\sin^2\theta_0 - \sin^2\theta\right) \tag{96.3}$$

where θ_0 is the bias angle and is set to 45° for good linearity. In practice, a longitudinal bias is also used along the MRE width direction to stabilize the magnetic domain and reduce large Barkhausen noise. To compare the output between an MR head and an inductive head, we write the inductive head output using Faraday's equation:

$$v = -nV\frac{d\phi}{dx} \tag{96.4}$$

where n is the number of the coil turns, V is the medium velocity, ϕ is the magnetic flux, and x is the coordinate axis fixed on the medium surface. Equations 96.3 and 96.4 tell us that while inductive head

output is proportional to medium velocity and not suitable for low-velocity applications, the MR head can be used for either high- or low-velocity applications.

Recording Process

We can imagine the recording process in two steps. First, the magnetic flux flowing in the head core generates a fringe magnetic field around the gap. Then the magnetic field magnetizes the magnetic medium and leaves a magnetization transition in it. Partly due to the nonlinear nature of the hysteresis loop, the recording process is so complex that there has been no rigorous explanation. However, we can still obtain significant insights into the recording process by using simple models if we keep in mind the limitations.

If we set the origin of a coordinate system at the center of the gap with x axis on the head surface and y axis pointing away from the head, then the longitudinal magnetic field H_x and perpendicular magnetic field H_y of this head can be expressed by the Karlqvist approximation [2]:

$$H_x = \frac{ni}{\pi\left(g + lA_g/\mu A_c\right)} \tan^{-1}\left[\frac{yg}{x^2 + y^2 - \left(g^2/4\right)}\right] \tag{96.5}$$

$$H_y = \frac{ni}{2\pi\left(g + lA_g/\mu A_c\right)} \ln\left[\frac{\left(x - g/2\right)^2 + y^2}{\left(x + g/2\right)^2 + y^2}\right] \tag{96.6}$$

where n is the number of coil turns, i is the current in the coil, g is the gap length, l is the core length, A_g is the core cross-sectional area, μ is the relative permeability of core material, and A_c is the gap cross-sectional area. Both Eqs. 96.5 and 96.6 give accurate results for points $0.25g$ away from gap corners. Since longitudinal recording mode dominates the magnetic recording industry, we will focus on the field H_x. Equation 96.5 shows that the contours of constant H_x field are circles nesting on the two gap corners, as shown in Fig. 96.17. The greater the diameter of the circle, the weaker the magnetic field. Assume a magnetic medium, moving from left to right with a distance d above the head, has a thickness δ and a magnetization M pointing to right. At some instant the recording current turns on and generates the magnetic field H_x above the gap as depicted in Fig. 96.17. On the circumference of $H_x = H_c$, half of medium material has its magnetization reversed and half remains the same, resulting in a zero total magnetization. Since H_x has a gradient, the medium closer to the gap (inside a smaller circle) gets its magnetization reversed more completely than the medium farther away from the gap (outside a bigger circle). Therefore, magnetic transition is gradual in the medium even if the change of recording current follows a step function. Assume the original magnetization is M_r and the completely reversed magnetization is $-M_r$, this gradual change of magnetization for an isolated transition can be modeled by [3]:

$$M = \frac{2}{\pi} M_r \tan^{-1}\frac{x}{a} \tag{96.7}$$

where x is the distance from the center of transition and a is a parameter characterizing the sharpness of the transition as shown in Fig. 96.18. Assuming a thin-film medium and using the Karlqvist approximation for the head field, a is found to be [4–6]:

$$a = \frac{\left(1 - S^*\right)\left(d + \delta/2\right)}{\pi Q} + \sqrt{\left[\frac{\left(1 - S^*\right)\left(d + \delta/2\right)}{\pi Q}\right]^2 + \frac{M_r \delta\left(d + \delta/2\right)}{\pi Q H_c}} \tag{96.8}$$

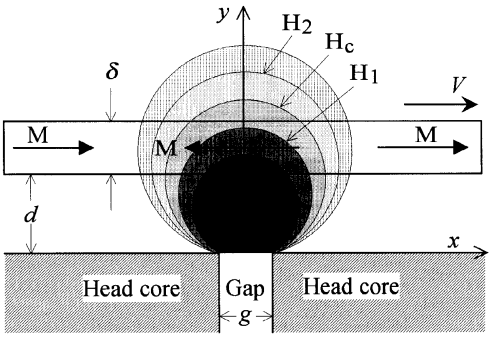


FIGURE 96.17 The constant horizontal fields of Karlqvist approximation are circles resting on the gap corners of a head, and the change of magnetization in the medium is gradual.

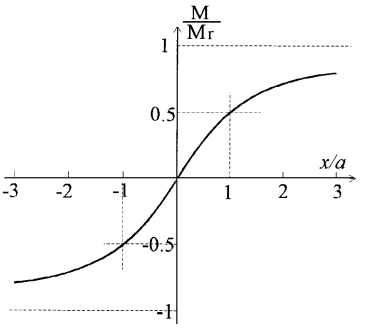


FIGURE 96.18 An isolated arctangent magnetization transition from negative M_r to positive M_r.

where S^* is the medium loop squareness and Q is the head-field gradient factor. For a reasonably well designed head, $Q \approx 0.8$. It is obvious that we want to make parameter a as small as possible so that we can record more transitions for a unit medium length. If the head gap length g and medium thickness δ are small compared with head/medium separation d, and the medium has a squareness of one, then the minimum possible value of a is [7]

$$a_m = \begin{cases} \dfrac{M_r \delta}{2\pi H_c} & \dfrac{M_r \delta}{4\pi H_c d} \geq 1 \\ \sqrt{\dfrac{M_r \delta d}{\pi H_c}} & \dfrac{M_r \delta}{4\pi H_c d} < 1 \end{cases} \tag{96.9}$$

In order to decrease the value of a and therefore increase areal density, we need to reduce medium remanence M_r, thickness δ, head/medium separation d, and to increase coercivity H_c.

Reproducing Process

In contrast to the recording process, the reproducing process is well understood. The flux density induced in the head core is on the order of a few millitesia, yielding a linear process for easier mathematical treatment. The head fringe field is the Karlqvist approximation (Eq. 96.5) and the foundation is the reciprocity theorem. For an isolated transition (Fig. 96.18) with a thin magnetic layer $\delta \ll d$, the induced electric voltage v for an inductive head is [7]:

$$v(x) = \frac{-2\mu_0 V w M_r \delta n}{\pi\left(g + lA_g / \mu A_c\right)}\left[\tan^{-1}\left(\frac{g/2 + x}{a + d}\right) + \tan^{-1}\left(\frac{g/2 - x}{a + d}\right)\right] \tag{96.10}$$

where μ_0 is the permeability of vacuum, μ is the relative permeability of the core, V is the medium velocity, w is the track width, n is the number of coil turns, g is the head gap length, d is the head/medium

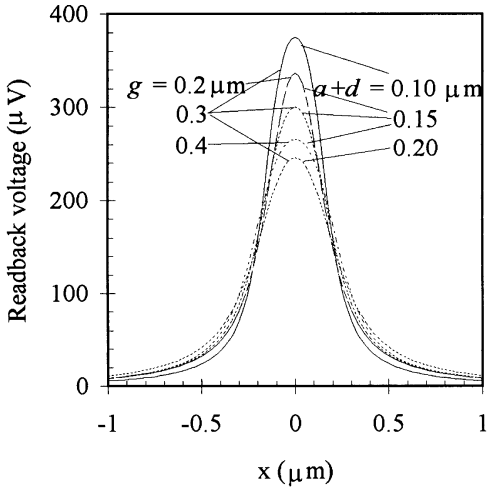


FIGURE 96.19 The reproducing voltage of an inductive head over an isolated arctangent transition shows the effects of gap length *g*, parameter *a*, and head/medium separation *d*.

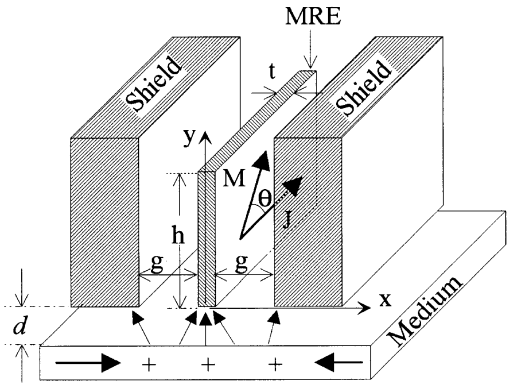


FIGURE 96.20 Schematic diagram of a shielded MR head with a shield to MRE distance *g*.

separation, and *x* is the distance between the center of the medium transition and the center of the head gap. The term $lA_g/\mu A_c$ is closely related to *g* for head efficiency. When a transition passes under the head, its voltage starts with a very low value, reaches a peak, then falls off again, as shown in Fig. 96.19, where the following typical values for a hard disk drive are used: V = 20 m/s, w = 3.5 μm, M_r = 450 kA/m, δ = 50 nm, n = 50, $lA_g/\mu A_c = 0.1g$. The effects of *g* and $a + d$ are shown in Fig. 96.19. Since a greater peak voltage and a narrower spatial response are desired for the reproducing process, smaller *g* and $a + d$ values are helpful.

When an MR head is used for reproducing, the MRE is usually sandwiched between two magnetic shields to increase its spatial resolution to medium signals, as shown in Fig. 96.20. Since the MR head is flux sensitive, the incident flux ϕ_i on the bottom surface of the MRE should be derived as a function of the distance (x) between the center of MRE and the center of the transition [7, 8]:

$$\phi_i(x) = \frac{2\mu_0 w M_r \delta(a+d)}{\pi g}\left\{ f\left[\frac{x+(g+t)/2}{a+d}\right] - f\left[\frac{(x+t)/2}{a+d}\right] + f\left[\frac{x-(g-t)/2}{a+d}\right] - f\left[\frac{(x-t)/2}{a+d}\right]\right\} \tag{96.11}$$

where *g* is the distance between the MRE and the shield, *t* is the MRE thickness, and

$$f(\beta) = \beta \tan^{-1}\beta - \ln\sqrt{1+\beta^2} \tag{96.12}$$

The angle between magnetization and current varies along the MRE height *h*. To find out the variation, we need to calculate the signal flux decay as a function of *y* by

$$\phi_s(y) = \phi_i \frac{\sinh\left[(h-y)/l_c\right]}{\sinh(h/l_c)} \tag{96.13}$$

where

$$l_c = \sqrt{\mu g t/2} \tag{96.14}$$

Then the bias angle θ_0 and signal angle θ can be calculated by

$$\sin\theta_0(y) = \frac{\phi_b(y)}{M_s t} \tag{96.15}$$

and

$$\sin\theta(y) = \frac{\phi_s(y) + \phi_b(y)}{M_s t} \tag{96.16}$$

where ϕ_b is the biasing flux in the MRE and M_s is the saturation magnetization of the MRE. Application of Eqs. 96.15 and 96.16 to Eq. 96.3 and integration over height *h* lead to

$$v = Jw\rho \frac{\Delta\rho}{\rho}\frac{1}{h}\left[\int_0^h \sin^2\theta_0(y)dy - \int_0^h \sin^2\theta(y)dy\right] \tag{96.17}$$

For an MR head with a 45° bias at the center and small height $h << l_c$, the peak voltage is [6]

$$v_p \approx \frac{9\Delta\rho J w M_r \delta(g+t)}{8\sqrt{2}\, tgM_s} \tan^{-1}\left[\frac{g}{2(a+d)}\right] \tag{96.18}$$

The general shape of the reproducing voltage from an MR head is similar to that in Fig. 96.19.

The study of an isolated transition reveals many intrinsic features of the reproducing process. However, transitions are usually recorded closely in a magnetic medium to achieve high linear density. In this case, the magnetization variation in the medium approaches a sinusoidal wave. That is,

$$M(x) = M_r \sin\frac{2\pi}{\lambda}x \tag{96.19}$$

where λ is the wavelength. The reproducing voltage in an inductive head becomes [9, 10]

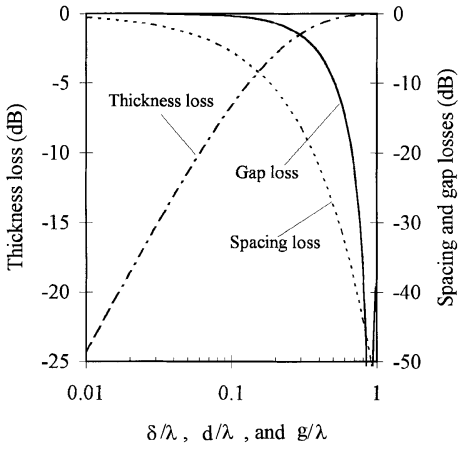


FIGURE 96.21 Spacing, thickness, and gap losses of an inductive head vs. frequency for the reproducing of a sinusoidal medium magnetization.

$$v(x) = \frac{-\mu_0 V w M_r n g}{g + l A_g / \mu A_c}\left(e^{-2\pi d/\lambda}\right)\left(1 - e^{-2\pi\delta/\lambda}\right)\left(\frac{\sin\frac{\pi g}{\lambda}}{\pi g/\lambda}\right)\cos\frac{2\pi}{\lambda}x \qquad (96.20)$$

This equation presents all the important features of the high-linear-density reproducing process. The term $\exp(-2\pi d/\lambda)$ is the spacing loss. It shows that the reproducing voltage falls exponentially with the ratio of head/medium spacing to wavelength. The second term $1 - \exp(-2\pi\delta/\lambda)$ is the thickness loss. The name of this term is misleading because its value increases with a greater medium thickness. However, the rate of increase diminishes for thicker medium. In fact, 80% of the maximum possible value is achieved by a medium thickness of 0.25λ. The last term $\sin(\pi g/\lambda)/(\pi g/\lambda)$ is the gap loss. This term is based on the Karlqvist approximation. If a more accurate head fringe field is used, this term is modified to $\sin(\pi g/\lambda)/(\pi g/\lambda)\cdot(1.25g^2 - \lambda^2)/(g^2 - \lambda^2)$ [11]. It shows a gap null at $\lambda = 1.12g$, and limits the shortest wavelength producible. These three terms are plotted in Fig. 96.21. The most significant loss comes from the spacing loss term, which is 54.6 dB for $d = \lambda$. Therefore, one of the biggest efforts spent on magnetic recording is to reduce the head/medium spacing as much as possible without causing mechanical reliability issues. For an MR head, the reproducing voltage is [11]

$$v \propto \frac{4 M_r i \Delta\rho w \lambda}{ht}\left(e^{-2\pi d/\lambda}\right)\left(1 - e^{-2\pi\delta/\lambda}\right)\left(\frac{\sin\frac{\pi g}{\lambda}}{\pi g/\lambda}\right)\sin\frac{\pi(g+t)}{\lambda}\cos\frac{2\pi}{\lambda}x \qquad (96.21)$$

Digital vs. Analog Recording

Due to the nonlinearity of the hysteresis loop, magnetic recording is intrinsically suitable for digital recording, where only two states (1 and 0) are to be recognized. Many physical quantities, however, are received in analog form before they can be recorded, such as in consumer audio and instrumentation recording. In order to perform such recording, we need to either digitize the information or use the analog recording technique. In the case of digitization, we use an analog-to-digital converter to change a continuous signal into binary numbers. The process can be explained by using the example shown in Fig. 96.22. An electric signal V, normalized to the range between 0 and 1, is to be digitized into three bits. The signal is sampled at time $t = 1, 2, \ldots, 6$. At each sampling point, the first bit is assigned a 1 if the value of the continuous signal is in the top half (>0.5), otherwise assigned a 0. The second bit is assigned a 1 if the value of the continuous signal is in the top half of each half ($0.25 \le V < 0.5$, or >0.75),

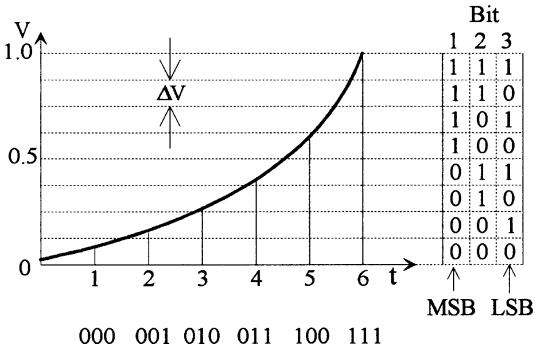


FIGURE 96.22 Schematic illustration of the quantization of a continuous signal to three bits.

otherwise assigned a 0. The third bit is assigned similarly. The first bit is the most significant bit (MSB), and the last bit is the least significant bit (LSB). The converted binary numbers are listed below each sampling point in Fig. 96.22. This process of digitization is termed *quantization*. In general, the final quantization interval is

$$\Delta V = \frac{V}{2^N} \tag{96.22}$$

Where V is the total voltage range and N is the number of bits. Because we use a finite number of bits, statistically there is a difference between the continuous signal and the quantized signal at the sampling points. This is the quantization error. It leads to a signal-to-quantization-noise ratio (SNR) [12]:

$$\text{SNR} = \frac{12P_s}{\Delta V^2} \tag{96.23}$$

where P_s is the mean square average signal power. For a signal with uniform distribution over its full range, this yields

$$\text{SNR} = 2^{2N} \tag{96.24}$$

For a sinusoidal signal, it changes to

$$\text{SNR} = 1.5 \times 2^{2N} \tag{96.25}$$

The SNR can be improved by using more bits. This improvement, however, is limited by the SNR of the incoming continuous signal. The quantized signal is then pulse code modulated (PCM) for recording.

For analog recording, a linear relationship between the medium magnetization and the recording signal is required. This is achieved through the anhysteretic magnetization process. If we apply an alternating magnetic field and a unidirectional magnetic field to a previously demagnetized medium, and then reduce the amplitude of the alternating field to zero before we remove the unidirectional field, the remanent magnetization shows a pseudolinear relationship with the unidirectional field strength H_u up to some level. Figure 96.23 shows such an anhysteretic curve. The linearity deteriorates as H_u gets greater. In applications, the recording signal current is biased with an ac current of greater amplitude and higher frequency. Therefore, it is also termed ac-biased recording. Analog recording is easy to implement, at the price of a lowered SNR because remanent magnetization is limited to about 30% of the maximum possible M_r to achieve good linearity.

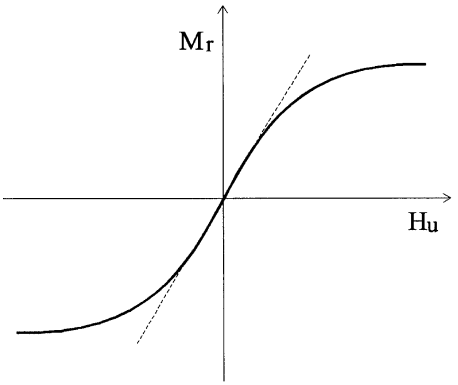


FIGURE 96.23 An anhysteretic remanent magnetization shows a pseudo-linear relationship with the applied unidirectional magnetic field to some H_u level.

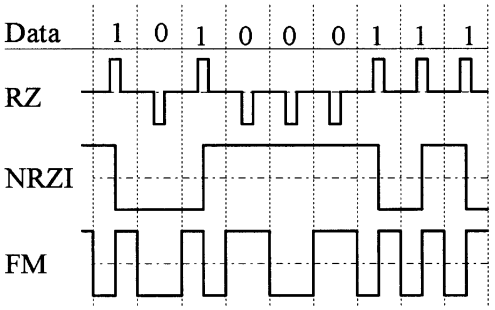


FIGURE 96.24 Comparison of some early developed codes.

Recording Codes

PCM is a scheme of modifying input binary data to make them more suitable for a recording and reproducing channel. These schemes are intended to achieve some of the following goals: (1) reducing the dc component, (2) increasing linear density, (3) providing self-clocking, (4) limiting error propagation, and (5) achieving error-free detection. There are numerous code schemes, only three of the ones developed early are shown in Fig. 96.24. The earliest and most straightforward one is the return-to-zero (RZ) code. In this scheme a positive and negative pulse is used to represent each 1 and 0, respectively, of the data. The main drawback is that direct recording over old data is not possible due to the existence of zero recording current between two data. It also generates two transitions for each bit, therefore reducing the linear density. In addition, it only uses half of the available recording current range for a transition. The non-return-to-zero-invert (NRZI) method was developed to alleviate some of these problems. It changes the recording current from one direction to the other for each 1 of the data, while making no changes for all 0s. However, it has a strong dc component and may lose reproducing synchronization if there is a long string of 0s in the input data. In addition, reproducing circuits are usually not designed for dc signal processing. In frequency modulation (FM) code there is always a transition at the bit–cell boundary which acts as a clock. There is also an additional transition at the bit–cell center for each 1 and no transition for 0s. It reduces the dc component significantly. The primary deficiency is the reduction of linear density since there are two transitions for each 1 in the data.

The most popularly used codes for magnetic recording are the run-length-limited (RLL) codes. They have the general form of $m/n(d, k)$. In these schemes, data are encoded in groups. Each input group has m bits. After encoding, each group contains n bits. In some schemes multiple groups are coded together. d and k are the minimum and maximum 0s, respectively, between two consecutive 1s in the encoded sequence. While d is used to limit the highest transition density and intersymbol interference, k is employed to ensure adequate transition frequency for reproducing clock synchronization. The encoding is carried out by using a lookup table, such as Table 96.10 for a 1/2(2,7) code [13].

TABLE 96.10 1/2(2,7) Code

Before Coding	After Coding
10	0100
11	1000
000	000100
010	100100
011	001000
0010	00100100
0011	00001000

Head/Medium Interface Tribology

As expressed in Eq. 96.20, the most effective way to increase signal amplitude, therefore areal density, is to reduce head/medium spacing *d*. However, wear occurs when a moving surface is in close proximity to another surface. The amount of wear is described by Archard's equation:

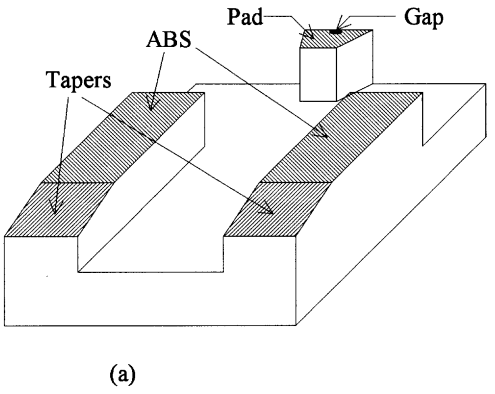
$$V = k\frac{Ws}{H} \tag{96.26}$$

where V is the volume worn away, W is the normal load, s is the sliding distance, H is the hardness of the surface being worn away, and k is a wear coefficient. In order to increase medium hardness H, hard Al_2O_3 particles are dispersed in particulate media and a thin layer of hard carbon ($\approx$ 10 nm) with either hydrogenation or natrogenation is coated on thin-film media of hard disks. A layer of liquid lubricant, typically perfluoropolyethers with various end groups and additives, is applied on top of the carbon film to reduce the wear coefficient k. Load is minimized to reduce wear while keeping adequate head/medium dynamic stability. For applications where the sliding distance s is modest over the lifetime of the products such as floppy disk drives and consumer tapes drives, the head physically contacts the medium during operations. In the case of hard disk application, heads are separated nominally from the media by a layer of air cushion. The head is carried on a slider, and the slider uses air-bearing surfaces (ABS) to create the air film based on hydrodynamic lubrication theory. Figure 96.25 shows two commonly used ABS. Tapers are used to help the slider take off and maintain flying stability. ABS generates higher-than-ambient pressure to lift the slider above the medium surface during operations. The tripad slider is for pseudo-contact applications while the subambient-pressure (SAP) slider is for flying (such as MR head) applications. Because the relative linear velocity between the slider and the medium changes when the head moves to different radii, a cavity region is used in the SAP slider to generate suction force to reduce flying height variation. The ABS is designed based on the modified Reynolds equation:

$$\frac{\partial}{\partial X}\left(PH^3Q\frac{\partial P}{\partial X}\right)+\frac{\partial}{\partial Y}\left(PH^3Q\frac{\partial P}{\partial Y}\right)=\Lambda_x\frac{\partial(PH)}{\partial X}+\Lambda_y\frac{\partial(PH)}{\partial Y}+\sigma\frac{\partial(PH)}{\partial T} \tag{96.27}$$

where X and Y are coordinates in the slider longitudinal and transverse directions normalized by the slider length and width, respectively, P is the hydrodynamic pressure normalized by the ambient pressure, H is the distance between the ABS and medium surface normalized by the minimum flying height, Q is the molecular slip factor, T is time normalized by the characteristic squeeze period, Λ_x and Λ_y are the bearing numbers in the x and y directions, respectively, and σ is the squeeze number. A full derivation and explanation of the Reynolds equation can be found in Reference 14. At present, high end hard disk drives feature a flying height on the order of 20 to 50 nm.

When power is turned off, the slider in the popularly used Winchester-type drives rests on the medium surface. Although the ABS and medium surface look flat and smooth, they really consist of microscopic peaks and valleys. If we model an ABS/medium contact by a flat surface and a sphere tip, the liquid lubricant on the medium surface causes a meniscus force F_m as depicted in Fig. 96.26 [15]:



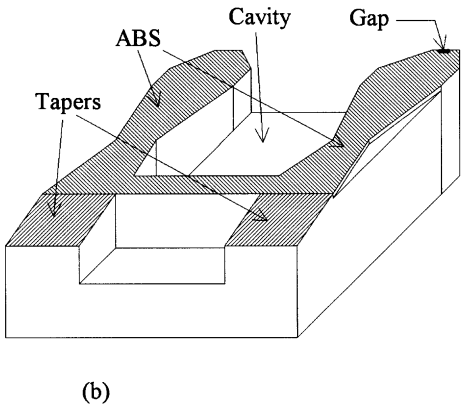


FIGURE 96.25 The ABS of (a) a tri-pad slider for pseudo-contact recording and (b) a SAP slider for conventional flying recording.

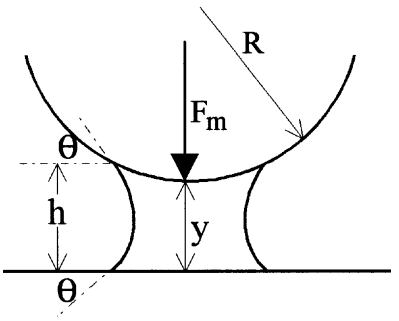


FIGURE 96.26 Formation of meniscus between a sphere tip and a flat surface.

$$F_m = \frac{4\pi R\gamma\cos\theta}{1+y/(h-y)} \tag{96.28}$$

where R is the radius of the sphere, γ is the surface tension of the lubricant, θ is the contact angle between the lubricant and the surfaces, y is the sphere to flat surface distance, and h is the lubricant thickness. Detailed analysis [16] shows that the static friction F at a head/medium interface is a function of several parameters:

$$F = f\left(h, R, A, \eta, \gamma, \theta, E', \phi, \sigma, s\right) \tag{96.29}$$

where A is the ABS area, η is the peak density, E' is the effective modulus of elasticity, ϕ is the peak height distribution, σ is the rms peak roughness, and s is the solid-to-solid shear strength. If friction F is too large, either the drive cannot be started or the head/medium interface is damaged. While friction can be reduced practically by reducing A, γ, and increasing θ, the most effective ways are to control h, σ, η, and ϕ. Too thin a lubricant layer will cause wear, and too thick will induce high friction. This limits h to the range of 1 to 3 nm. σ is controlled by surface texture. Historically, texture is created by mechanical techniques using either free or fixed abrasives. This leaves a surface with a random feature and is unsuitable for controlling η and ϕ. Recently, people started to use lasers [17]. This technique generates a surface texture with well-defined η and ϕ to improve wear and friction performance. Figure 96.27 shows AFM images of a mechanical and a laser texture.

Optical Recording

The major obstacle to achieving higher areal density in magnetic recording is the spacing loss term. It is a great engineering challenge to keep heads and media in close proximity while maintaining the head/medium interface reliable and durable. Care must also be taken in handling magnetic media since even minute contamination or scratches can destroy the recorded information. In addition, the servo technique of using magnetic patterns limits the track density to about one order lower than the linear density. Optical recording, on the other hand, promises to address all these concerns.

Optical recording can be categorized into three groups. In the first group, information is stored in the media during manufacturing. Users can reproduce the information, but cannot change or record new information. CD-ROM (compact disk–read only memory) belongs to this group. The second group is WORM (write once read many times). Instead of recording information during manufacturing, it leaves this step to the user. This is usually achieved by creating physical holes or blisters in the media during the recording process. Once it is recorded, however, the medium behaves like the first group: no further recording is possible. The third group is similar to magnetic recording. Recording can be performed infinitely on the media by changing phase or magnetization of the media. The most noticeable example in this group is the magneto-optic (MO) technique [18]. Only CD-ROM and the magneto-optic recording are described in the following.

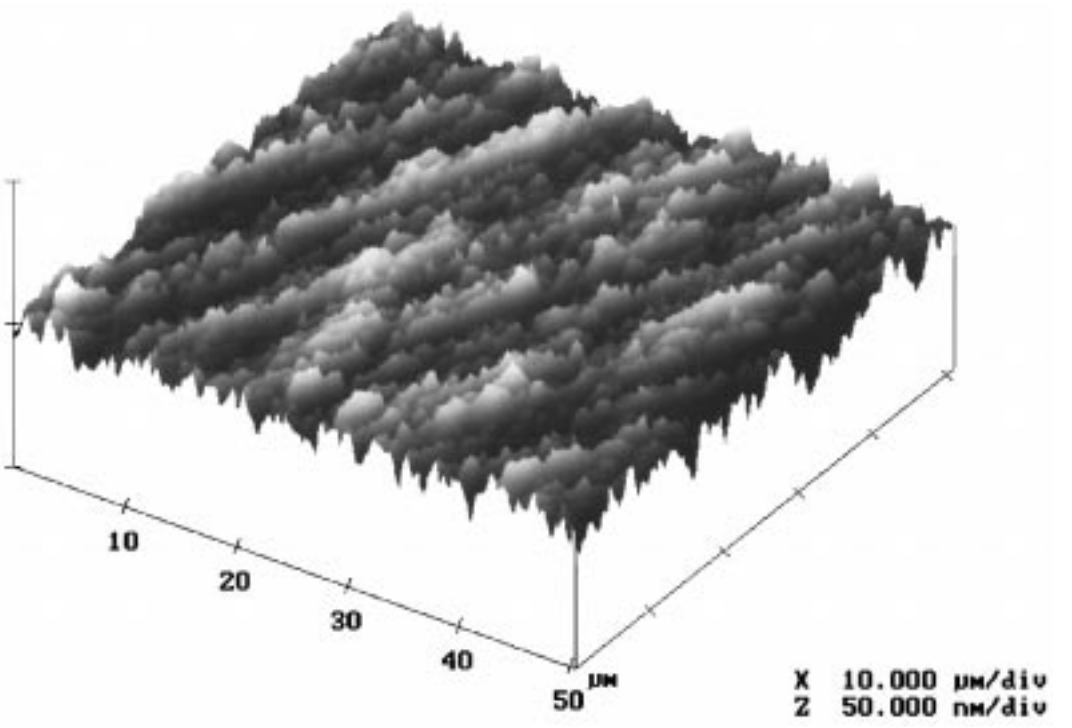
CD-ROM

Figure 96.28 shows the CD-ROM reproducing principle. Data are pressed as physical pits and lands on one surface of a plastic substrate, usually polycarbonate. A layer of aluminum is deposited on this surface to yield it reflective. Lacquer is then coated to protect the aluminum layer. During the reproducing process, an optical lens is used to focus a laser beam on the reflective pit and land surface. The diameter of the lens is D, the distance between the lens and the substrate is h_3, and the substrate thickness is h_2. The diameter of the laser beam is d_2 when entering the substrate, and becomes d_1 when focused on the reflective surface. The width of the pits are designed smaller than d_1. The reflected light consists of two portions: I_1 from the land and I_2 from the pit. According to the theory of interference, the intensity of the reflected light is

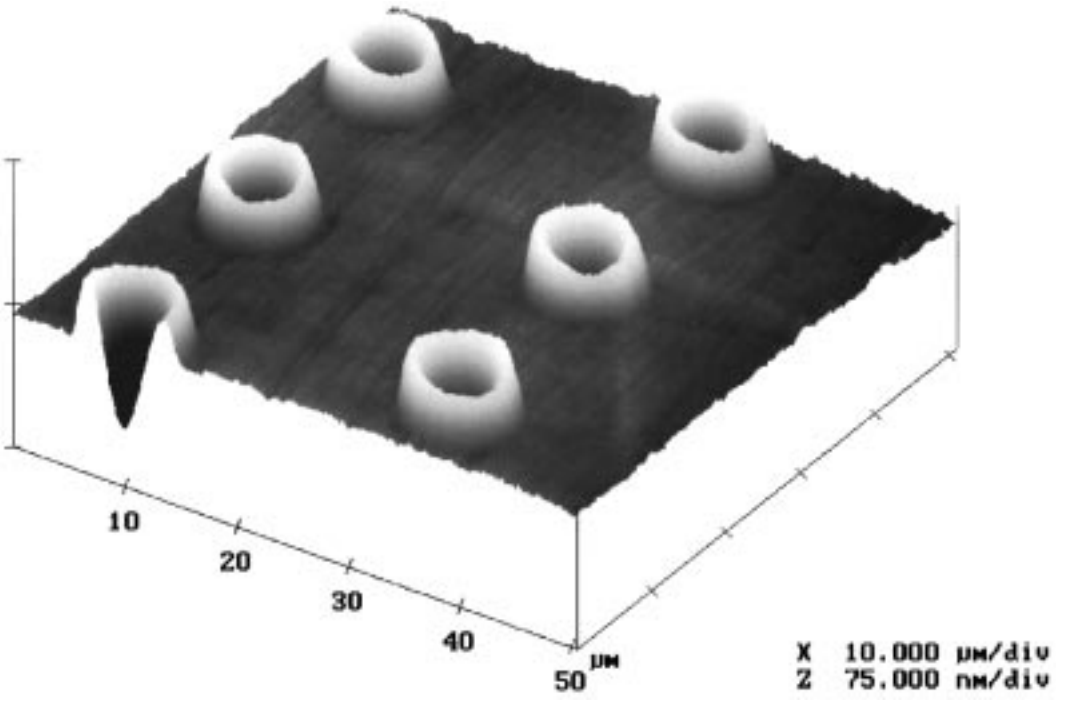
$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\frac{4\pi h_1}{\lambda} \tag{96.30}$$

where λ is the wavelength of the laser and $h_1 \approx \lambda/4$ is the pit height. This leads to

$$I = \begin{cases} I_1 + I_2 - 2\sqrt{I_1 I_2} & \text{if there is a pit } (h_1 = \lambda/4) \\ I_1 + I_2 + 2\sqrt{I_1 I_2} & \text{if there is no pit } (h_1 = 0) \end{cases} \tag{96.31}$$



(a)



(b)

FIGURE 96.27 AFM images of (a) a mechanical texture and (b) a laser texture (Courtesy of J. Xuan).

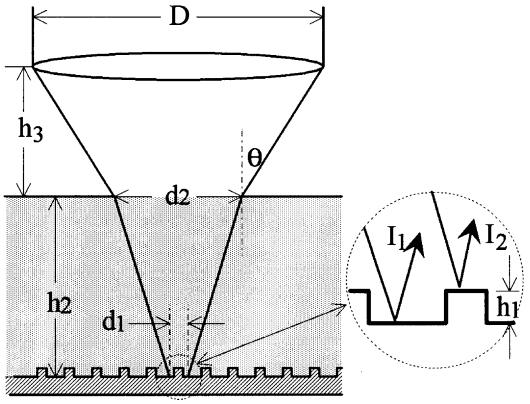


FIGURE 96.28 Schematic representation of the CD-ROM reproducing principle.

This change of light intensity is detected and decoded to yield the recorded data. The reflected light is also used for focusing and track following.

The fundamental limit on optical recording density is the focused beam diameter d_1. For a Gaussian (TEM_{00}) laser, this is the diffraction-limited diameter at which the light intensity is reduced to $1/e^2$ of the peak intensity:

$$d_1 \approx \frac{2\lambda}{\pi\theta} \tag{96.32}$$

where θ is the aperture angle. The following values are typical for a CD-ROM system: λ (gallium arsenide laser) = 780 nm, θ = 27°, h_2 = 1.2 mm, D = 5 mm, h_3 = 4.2 mm. This yields $d_1 \approx 1.0$ μm and $d_2 \approx 0.7$ mm, and sets the areal density limit of optical (including magneto-optic) recording to about 1 Mbit/mm^2. For most CD-ROM applications, the areal density is smaller than this limit, and a disk with 120 mm diameter holds about 600 Mbyte information. In order to increase areal density, we can either reduce light wavelength or increase numerical aperture. Much of the effort has been to adopt a new light source with short wavelength such as a blue laser. Increasing numerical aperture is more difficult because increasing lens diameter is cost prohibitive and reducing h_2 or h_3 is reliability limited. Note that although the beam size on the focus plane is on the order of 1 μm (d_1), it is two to three orders greater at the air/substrate interface (d_2). This means that optical recording can tolerate disk surface contamination and scratches much better than magnetic recording. However, the performance of optical recording does not match magnetic recording in general. The data transfer rate of CD-ROM drives is expressed as multiple (×) of 150 kB/s. Even for a 12× CD-ROM drive, the data access time and data transfer rate are still on the order of 100 ms and 1.8 MB/s, respectively, while for a high-performance rigid disk drive these values are less than 10 ms and greater than 30 MB/s, respectively.

Magnetooptic Recording

The primary drawback of a CD-ROM to an end user is its inability to record. This deficiency is remedied by MO recording technology, as depicted in Fig. 96.29. A linearly polarized laser beam is focused on a layer of magnetic material, and a coil provides a dc magnetic field on the other side of the medium. This dc magnetic field is too weak to affect the medium magnetization at normal temperature. The recording process utilizes the thermomagnetic property of the medium, and the reproducing process is achieved by using the Kerr effect. During recording, the medium is initially magnetized vertically in one direction, and the dc magnetic field is in the opposite direction. The laser heats up the medium to its Curie temperature, at which the coercivity becomes zero. During the cooling process, the dc magnetic field aligns the medium magnetization of the heated region to the magnetic field direction. In the process of reproducing, the same laser is used with a smaller intensity. The medium is heated up to its compensation temperature, at which the coercivity becomes extremely high. Depending on the direction of the mag-

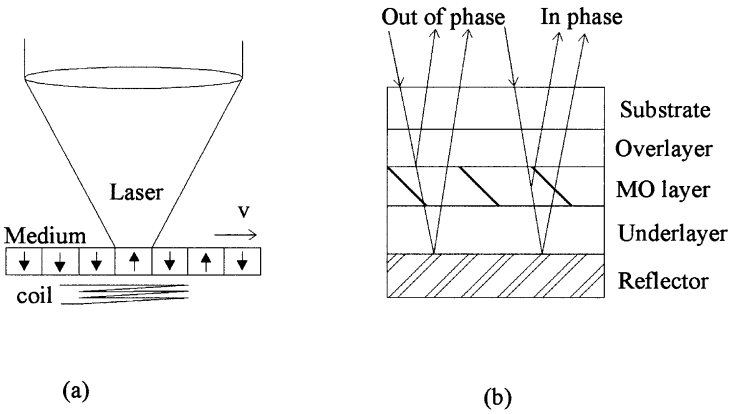


FIGURE 96.29 Schematic illustrations of (a) MO recording/reproducing and (b) quadrilayer medium cross section.

TABLE 96.11 Digital Magnetic and Optical Storage Devices

Description	Manufacturers	Approximate Price, $
Thin-film head for hard disk drive	AMC, Read-Rite, SAE	6.00–9.00
MR head for hard disk drive	AMC, Read-Rite, SAE, Seagate	8.00–12.00
Thin-film hard disk	Akashic, HMT, Komag, MCC, Stormedia	7.00–10.00
Hard disk drive	IBM, Maxtor, Quantum, Samsung, Seagate, WD	0.02–0.20/Mbytes
Floppy drive	Panasonis, Sony	20.00–40.00
Floppy disk	3M, Fuji, Memorex, Sony	0.15–0.50
Removable rigid disk drive	Iomega, Syquest	100.00–400.00
Removable rigid disk	Iomega, Maxell, Sony	5.00–20.00/100 Mbytes
Tape drive	Exabyte, HP, Seagate	100.00–400.00
Backup tape	3M, Sony, Verbatim	4.00–25.00/Gbytes
8× CD-ROM drive	Goldstar, Panasonic	100.00–200.00
Recordable optical drive	JVC, Philips	300.00–500.00
Recordable optical disk	3M, Maxell, Memorex	3.00–15.00/650 Mbytes

netization, the polarization of the reflected light is rotated either clockwise or counterclockwise (Kerr rotation). This rotation of polarization is detected and decoded to get the data. The main disadvantage of MO recording is that a separate erasing process is needed to magnetize the medium in one direction before recording. Recently some technologies have been developed to eliminate this separate erasing process at the cost of complexity.

The medium used in MO recording must have a reasonable low Curie temperature (<300°C). The materials having this property are rare earth transition metal alloys, such as $Tb_{23}Fe_{77}$ and $Tb_{21}Co_{79}$. Unfortunately, the properties of these materials deteriorate in an oxygen and moisture environment. To protect them from air and humidity, they are sandwiched between an overlayer and a underlayer, such as SiO, AlN, SiN, and TiO_2. Another issue with the rare earth transition metal alloys is their small Kerr rotation, about 0.3°. To increase this Kerr rotation, multiple layers are used. In the so-called quadrilayer structure (Fig. 96.29b), the overlayer is about a half-wavelength thick and the underlayer is about a quarter-wavelength thick [18]. The MO layer is very thin (≈3 nm). Light reflected from the reflector is out-of-phase with the light reflected from the surface of the MO layer, and is in-phase with the light reflected from the inside of the MO layer. As a result, the effective Kerr rotation is increased several times.

Compared with magnetic recording, optical recording has the intrinsic advantages of superior reliability and portability. However, its performance is inferior due to slower data access time and transfer rate. Another advantage of optical recording, higher areal density, has been disappearing or even reversing to magnetic recording. Both magnetic and optical recording will be continuously improved in the near future, probably toward different applications. Currently there are some emerging techniques that try to

combine the magnetic and optical recording techniques. Table 96.11 is a short list of representative magnetic and optical devices for digital recording.

References

1. C.D. Mee and E.D. Daniel, Eds., *Magnetic Storage Handbook*, 2nd ed., New York: McGraw-Hill, 1996.
2. O. Karlqvist, Calculation of the magnetic field in the ferromagnetic layer of a magnetic drum, *Trans. R. Inst. Technol. (Stockholm)*, 86, 3–28, 1954.
3. J.J. Miyata and R.R. Tartel, The recording and reproduction of signals on magnetic medium using saturation-type recording, *IRE Trans. Elec. Comp.*, EC-8, 159–169, 1959.
4. M.L. Williams and R.L. Comstock, An analytical model of the write process in digital magnetic recording, *A.I.P. Conf. Proc.*, 5, 738–742, 1972.
5. V.A.J. Maller and B.K. Middleton, A simplified model of the writing process in saturation magnetic recording, *IERE Conf. Proc.*, 26, 137–147, 1973.
6. H.N. Bertram, *Theory of Magnetic Recording*, Cambridge, UK: Cambridge University Press, 1994.
7. C.D. Mee and E.D. Daniel, Eds., *Magnetic Recording Technology*, 2nd ed., New York: McGraw-Hill, 1996.
8. R. Potter, Digital magnetic recording theory, *IEEE Trans. Magn.*, MAG-10, 502–508, 1974.
9. W.K. Westmijze, Studies on magnetic recording, *Philips Res. Rep.*, 8, 161–183, 1953.
10. G.J. Fan, A study of the playback process of a magnetic ring head, *IBM J. Res. Dev.*, 5, 321–325, 1961.
11. J.C. Mallinson, *The Foundations of Magnetic Recording*, 2nd ed., San Diego, CA: Academic Press, 1993.
12. M.S. Roden, *Analog and Digital Communication Systems*, 4th ed., Upper Saddle River, NJ: Prentice-Hall, 1996.
13. P.A. Franaszek, *Run-length-limited variable length coding with error propagation limitation*, U.S. Patent No. 3, 689, 899, 1972.
14. W.A. Gross, L. Matsch, V. Castelli, A. Eshel, T. Vohr, and M. Wilamann, *Fluid Film Lubrication*, New York: Wiley, 1980.
15. J.N. Israelachvili, *Intermolecular and Surface Forces*, London: Academic Press, 1985, 224.
16. Y. Li and F.E. Talke, A model for the effect of humidity on stiction of the head/disk interface, *Tribology Mech. Magn. Storage Syst.*, SP-27, 79–84, 1990.
17. R. Ranjan, D.N. Lambeth, M. Tromel, P. Goglia, and Y. Li, Laser texturing for low-flying-height media, *J. Appl. Phys.*, 68, 5745–5747, 1991.
18. J. Watkinson, *The Art of Data Recording*, Oxford U.K.: Focal Press, 1994.
19. G.A.N. Connell and R. Allen, Magnetization reversal in amorphous rare-earth transition-metal alloys: TbFe, *Proc. 4th Int. Conf. on Rapidly Quenched Metals*, Sendai, Japan, 1981.