J. Webster (ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*

PRINTERS

Though information technology has made great strides in recent decades, the "paperless office" remains unrealized (1), in part because printed documents have tangible features that their electronic counterparts cannot yet match. Moreover, the integration of computer, information, and printing technologies makes it possible to generate and print documents with unprecedented freedom and speed. The objective of printing is to place pigmented marks on media so that the marks form text, graphics, and pictures in an arrangement specified by their creator as part of a document, repeating the process until the desired number of copies has been reproduced. This article covers the essential technologies of computer printers: marking technologies, digital halftoning, color science, media handling, precision electromechanics, and electronic document descriptions.

Mechanical reproduction of documents, the key to the dispersal of knowledge for the past six centuries, depends on marking technologies that produce consistent results. Modern marking technologies, like their predecessors, divide into two fundamental technological groups—impact and nonimpact. Impact printing transfers pigment to the print medium by mechanically striking the medium. Mechanized impact printing, as opposed to manual impact printing such as handwriting, reaches back to the invention of the printing press in the mid fifteenth century. Early printing presses pressed an ink-coated negative relief image formed of metal or wood onto the media to form the images one page at a time. Nonimpact printing is a collection of technologies that do not use mechanical force to apply pigment to the print medium. Nonimpact printing had its start in Senefelder's development of lithography in the early nineteenth century. The lithographic process takes advantage of the chemical affinities and solubility of inks to transfer the desired image to the media.

Modern impact printing transfers ink, typically held in a ribbon, to the print medium through the forceful impact of a printhead (2). Early computer teletypes such as the ASR-33 functioned much like manual and electric typewriters. The type (characters in negative relief) is arranged on type bars that individually strike through the ink ribbon. In a daisy wheel printer, the type is instead arranged on a circular wheel. The wheel is rotated until the appropriate character or symbol is facing the ribbon, then the wheel strikes the ribbon to mark the appropriate character. Chain printers use an array of impact elements to form one whole line of text at a time. The array is a set of vertical bars, each containing all possible characters. Each bar is moved vertically until the desired character for that location is positioned, and then the array is struck against the medium through the ink ribbon to form a line of text. Dot matrix, the most flexible of the impact printing technologies, uses a printhead consisting of an array of pins. As the printhead scans horizontally across the print medium, the pins are individually actuated to strike the print medium through a ribbon, producing letters, symbols, and even images as bit maps. The mechanical spacing of the moving pins places a limit on the resolution that can be achieved by a dot matrix printer.

Impact printing technologies are no longer in widespread use, giving way to a new class of approaches that are generally faster and quieter and produce higher quality output. Coupled with computer systems, these new approaches provide the user with tremendous flexibility in forming, transmitting, and printing documents. These technologies are the focus of this article. The three most frequently used nonimpact marking technologies are lithography, electrophotography, and ink jet. When choosing the technology to use for a particular document application, the features, capabilities, and costs of each should be weighed. Currently the dominant factor in this decision is the number of copies of the document required. Broadly speaking, lithography is used to print long page runs, that is, more than 500 or so copies of the same document. Electrophotography is typically used for runs of 1 to 500 copies or for longer runs where variation of document content is required. Color ink jet, popular in the small-office-home-office–desktop environment, is typified by very short page runs.

The following section presents a single representative from among the numerous technological variations for each of the three predominant marking technologies, offset lithography, electrophotography, and ink jet. The section after presents topics in printing that are distinct from the choice of marking technology: color science, digital halftoning, and page representations. The last section presents the mechanical challenges of printing.

Marking Technologies

Marking technologies, responsible for physically placing pigment on the print medium, are central to printing technology. Printers are often classified by the marking technology they employ, though important technological distinctions are also present in media handling, on-board computational power, page description language, and so on. Many marking technologies have been developed over the past decades. Two technologies are dominant in computer printers: electrophotography, which is the technology base for laser printing, and ink jet. Each of these categories contains many technological variations. This section will provide an overview of these technologies, examining the most popular variant from each category in detail. Before examining these relatively new marking technologies, it is useful to examine their much older sibling from the world of commercial printing, offset lithography.

Offset Lithography. The majority of mass-audience printed material, such as newspapers, magazines, pamphlets, and telephone books, is produced by offset lithography, the modern descendent of the lithographic process invented in the early nineteenth century. Although generally not considered a computer printer technology, offset lithography still bears mentioning due to the immense volume of material produced this way. In offset lithography, a metal plate, shaped as a cylinder or roller, is etched with the image to be printed. The plate is treated so that the areas that are to receive ink are hydrophobic, while the areas that are not to receive ink are hydrophilic. In a modern print shop, automated equipment etches and prepares plates from a computer specification. In the offset lithographic press, ink is applied to the plate, and since lithographic inks are hydrophobic, the ink only adheres to the appropriate areas. The press transfers the ink by contact from the plate to a rubber roller, called the blanket. The blanket then transfers the ink by contact to the print medium. The rubber blanket's compliance provides more uniform contact with the print medium than would be admitted by the plate, facilitating printing on a wider range of media. In addition, with the blanket as an intermediary, the plate has the same image as the final print, rather than the mirror image.

Offset lithography beautifully reproduces monochrome and color images and text. Although it has a relatively narrow color gamut, custom inks can be used to duplicate a desired color with high precision. Color consistency is very good over many copies. Generally offset lithography is used only for long page runs, since etching, mounting, and aligning a set of plates incurs relatively high prepress and startup costs, which must be amortized over a long printing run to be economically feasible. Offset lithographic presses can require as many as 500 test prints for setup and calibration. Historically a dedicated human operator regulated and maintained the quality of the printing process, but this is changing as real-time controls are incorporated into automated setup devices for registration and color balance. See Ref. (3) for an in-depth look at lithography.

Electrophotography (Laser Printers). Laser printers are based on electrophotography, the same marking technology used in a photocopier. Electrophotograhic reproduction centers on the *photoreceptor*, a belt or drum consisting of at least two layers, a photoconductive layer and a conductive substrate. In darkness, the photoreceptor can hold a static charge, but when exposed to light it discharges. The desired image is "painted" in static electricity and then developed with *toner*, small charged plastic particles. The toner is transferred

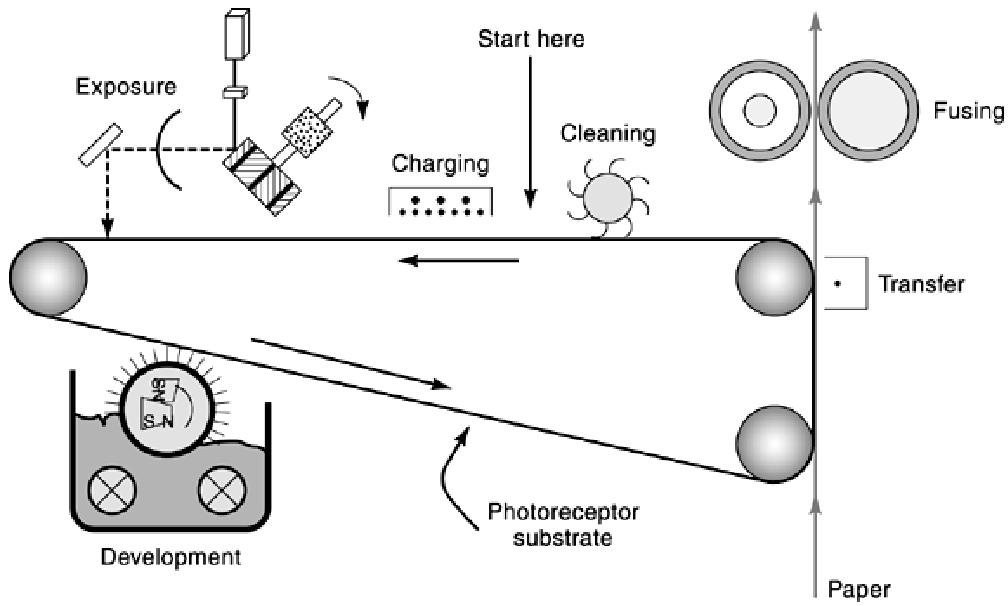

Fig. 1. A schematic depiction of the subsystems of the electrophotographic process as explained in the text. Many other technology choices for the subsystems and physical arrangements are possible, but the sequence of the process steps is relatively uniform between implementations.

to the print medium and then fused. The electrophotographic process consists of six steps: charge, expose, develop, transfer, fuse, and clean. The photoreceptor transports the image, in its various forms, between the subsystems. This subsection will discuss the steps of the electrophotographic process for monochrome printing, illustrated in Fig. 1. Further exposition is provided in Refs. 4,5,6.

The *charge* step deposits a uniform static charge on the photoreceptor. Typically, this is performed by a corona discharge, produced by a corotron or scorotron. A corotron is a thin wire to which a high ac and dc voltage is applied. The voltage creates a corona (breakdown of the surrounding air), which transfers charge to the photoreceptor. A scorotron is a corotron with the addition of a control grid between the wire and the photoreceptor. Voltage is applied to the grid to limit and to improve the uniformity of the charge on the photoreceptor. Consistent, uniform charging of the photoreceptor is necessary for accurate image reproduction.

The *expose* step produces a latent image, a pattern of charged and discharged areas, of the desired output on the photoreceptor. In a traditional light lens photocopier, the photoreceptor is discharged in the areas that are not to receive toner by bright light reflected off the original document. In this case a process called *charged-area development* (*CAD*) is used to develop the latent image, covering the remaining charged areas with toner. In a printer or digital photocopier, the latent image is produced by an addressable light source, a laser or light-emitting diode (*LED*) array. For most text images, the total toner area coverage is between 5% and 10%. For this reason, printers and digital copiers use the addressable light source to discharge areas of the image that are to receive toner, reducing the *duty factor* of the light source (the percentage of time the light source is on). In this case, a process called *discharged-area development* (*DAD*) is used to develop the latent image, covering the discharged areas with toner.

When the light source is a laser, the output image is *rasterized* (broken up into lines from top to bottom) similarly to the way a video raster is painted on the screen of a monitor by the electron beam. The light source, typically a diode laser, remains fixed in place, while the laser beam, reflected off a rotating polygonal mirror with constant angular velocity, sweeps across the photoreceptor. Each face of the mirror causes the laser to sweep out one line across the photoreceptor. The laser is modulated on and off by a bit stream, producing regions on the photoreceptor that are uncharged or charged, respectively. The combination of the laser and the rasterizing optics is collectively referred to as a *raster output scanner*, or *ROS*. The resulting pattern of charges on the photoreceptor is called the *latent image*.

Another popular addressable light source is the *LED bar*. LEDs may be constructed in silicon chip arrays and then assembled to produce an exposure system, covering the full width of the print medium, called an *image bar*. Each of the individual LEDs may be modulated directly by addressing logic contained in the carrier for the image bar. The drive electronics may also contain compensating resistors that trim the intensities of the individual LEDs so that the illumination from each is uniform across the bar. The bar is placed in the appropriate exposure location, and the LEDs are turned on and off by a bit stream similarly to the laser imaging case. LED bars avoid the architectural (they are smaller than the laser and the optical system) and control (no rapidly moving parts) constraints that govern the use of laser diodes. However, the loss of a single LED shows up readily as an image quality defect that requires the purchase and installation of an expensive new image bar.

In both cases the imaging system imposes a two-dimensional grid of dots on the photoreceptor. Each of these dots is called a *pixel* (from "picture element"), analogous to the well-known pixel of video display technology with the exception that most electrophotographic imaging technologies are capable of producing only binary (two-level—on–off) pixels. One dimension of the two-dimensional grid is achieved by moving the photoreceptor. This dimension is called the *process direction*, because the medium moves through the system in this direction, or the *slow scan direction*, and corresponds to the vertical dimension in video rasters. The spatial frequency of the lines taken in the process direction is a function of the photoreceptor speed and the scan speed of the laser or the strobing frequency of the LED bar. The direction perpendicular to the slow scan direction is called the *fast scan direction* and corresponds to the horizontal sweep in the video raster. The spatial frequency of the pixels in this direction is governed by the frequency of modulation provided to the laser for ROS systems or by the LED spacing in LED bars. When the two-dimensional grid of pixels is designed, the designer specifies a certain addressability. This quantity indicates how many dots per inch (*dpi*) may be written to the photoreceptor and is, for historical reasons, often specified in multiples of 300. Thus, when a printing system is advertised as being 600×1200, the raster lines are placed 1/600 in. (42.3 μm) apart and the modulation of the imaging system is 1200 dpi in the fast scan direction.

"Addressability" is often confused with "resolution." Addressability is associated with the imaging system's ability to space dots closer or farther from one another. Resolution is the ability of an optical system to discriminate fine detail, referring in this case to the imaging system's ability to reproduce fine structure in an image. The difference between these two terms derives from the size and shape of the dot produced by the imaging system. Smaller dots will preserve image detail better than larger dots at the same addressability. The imaging system does not use exactly rectangular dots, but usually elliptical ones, and thus it is impossible to fill a pixel exactly. The dot size is often made larger than a pixel in order to avoid holes at the corners of the pixels that would receive no exposure. Overfilled dots reduce the resolution of the printer at constant addressability. Marketing statements tends to focus on the addressability, which easier to evaluate. The issues of addressability versus resolution arise in the other printing technologies as well.

The *development* step uses toner to develop the latent image. Toner consists of pigmented, electrostatically charged plastic particles, 5 μm to 25 μm in diameter. In the developer housing, the toner is mixed with larger carrier particles or beads, 80 μm to 700 μm in diameter, which serve two purposes. First, extremely fine powders such as toner are difficult to transport, and can produce dirt inside the machine when they escape the housing, or spots on portions of the document that were supposed to be white. The carrier beads may carry up to 1000 toner particles, preventing powder contamination of other system components or the image. Second, the carrier beads charge the toner particles triboelectrically, that is, by friction. A photomicrograph of a carrier bead and its attached toner is shown in Fig. 2. The magnetic brush development system is the most widespread.

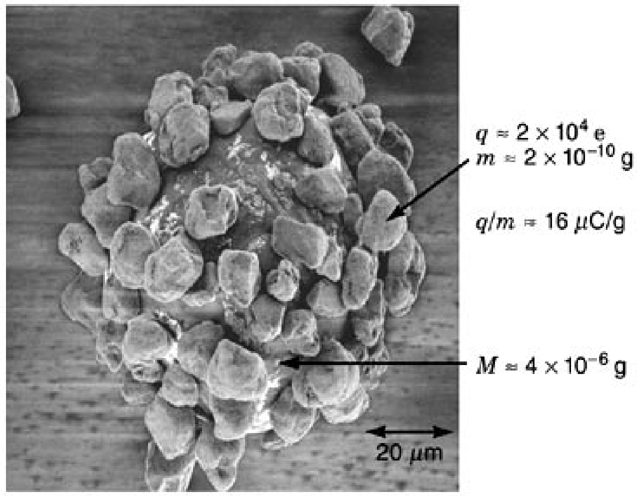

Fig. 2. A photomicrograph of a single developer bead with attached toner particles. The carrier bead is often composed of a coated ferrite core. In the developer housing, the carrier beads and toner particles are agitated to form a tribocharged material called *developer*. The mass of the carrier and the mass and charge of the toner particles are indicated.

In this system, the carrier beads are also magnetic. The toner-covered carrier beads form brushlike chains on a revolving shell, bringing the beads into contact with the photoreceptor. The resulting physical agitation in the development nip serves to break the adhesive and electrostatic forces binding the toner to the carrier and frees the toner to move under the influence of the photoreceptor latent image.

The developer housing is biased at a voltage between the photoreceptor's charge and discharge voltages. This dc bias produces two polarities of field between the housing and the photoreceptor. In DAD, used in digital printing, the areas that were exposed by the imaging system—the *development* field—points toward the photoreceptor, attracting the tribocharged toner. Meanwhile, in the unexposed photoreceptor regions (the areas intended to be white in the final image), the electric field (*cleaning* field) points toward the development roll, causing toner to remain on the roll. Thus, the charged toner can discriminate between the image and background regions of the image.

The *transfer* step moves the developed image to the print medium, generally paper. The medium is brought in contact with the photoreceptor. A transfer corona, with polarity opposite the toner, pulls the toner from the photoreceptor to the paper. Large particles tend to be transferred more efficiently than small particles, placing a limit on the size reduction of toner particles. In a typical, well-functioning system, between 90% and 100% of the toner is transferred from the photoreceptor to the print medium.

The *fusing* step permanently fixes the toner to the print medium, typically by applying heat and pressure by passing the medium between a pair of heated rollers. The rollers heat the toner sufficiently above the plastic's glass transition temperature to allow it to melt and fuse with the print medium. The pressure forces the melted toner into intimate contact with the paper fibers. When the toner cools, it undergoes thermal contraction. For images that cover a large percentage of the paper, the thermal contraction can cause the paper to curl, necessitating a decurling step to obtain flat sheets.

The *cleaning* step prepares the photoreceptor for the next image by removing any remaining toner left from the transfer step. This is typically performed by a third corona, which discharges the toner left on the photoreceptor, coupled with a bright light that discharges the photoreceptor. A brush or elastomer blade, similar to the one in the development stage, wipes the toner from the photoreceptor. Finally, an erase lamp removes any remaining charge from the photoreceptor.

Laser printers are very quiet and fast. The printers range from desktop models that print 2 to 4 pages per minute at an addressability of 300 × 300 dpi, to commercial printers at up to 2400 dpi. The fastest of these devices can print and bind a 250-page book with covers, inserted tabs, and binding in less than two minutes.

Ink Jet. Ink jets are a family of technologies that propel liquid ink directly onto the print medium. The physics supporting the ink jet process was investigated during the nineteenth century (7,8). Ink jet technologies fall into two categories, continuous and drop-on-demand. Continuous ink jets propel a continuous stream of droplets of electrically charged liquid ink, which is modulated by an electric field, deflecting the stream to a reservoir or to the print medium as desired. Continuous ink jet technology, first successfully produced and marketed in the 1950s, is several decades older than drop-on-demand, but has not developed a significant market presence, probably because the ink reservoir makes it too messy for home or office use (9).

Drop-on-demand technologies shoot ink upon request rather than deflecting a continuous stream. There are two principle drop-on-demand technologies, piezoelectric and thermal. Piezoelectric, developed in the early 1970s, uses a piezoelectric element (see Piezoelectricity) to propel individual droplets of ink to the print medium. Though piezoelectric is the simplest drop-on-demand technology, it is expensive to manufacture compared to thermal ink jets.

Thermal ink jets (*TIJs*), also known as bubble jets, are the dominant drop-on-demand printing technology. Developed independently at Hewlett-Packard (*HP*) and Canon in the late 1970s, TIJs use the explosive evaporation of a bubble of ink to propel a liquid ink droplet to the print medium. Bubble jets have been more successful than their various ink jet cousins for several reasons. First, fabrication techniques for TIJs are very similar to semiconductor fabrication techniques. In fact, TIJs are the most successful application of microelectromechanical systems (MEMSs) to date. These devices can be produced so inexpensively that the print head of an ink jet printer is typically produced as part of the disposable ink cartridge (10). Second, thermal transducers can be packed very tightly together. For example, HP currently markets a 300-nozzle 600 dpi print head, which rivals the print quality of a low-end laser printer for office applications. For further reading on TIJs see, for example, Refs. 11,12,13.

The print head of a thermal ink jet printer is an array of nozzles, each of which draws from a common ink reservoir and can print dots independently. In practice, there are two nozzle configurations, top shooter and side shooter. Functionally these configurations are almost identical. Printing a pixel consists of four stages: bubble nucleation, drop ejection, bubble collapse, and channel refill. The stages for a side-shooter configuration are shown in Fig. 3.

In bubble nucleation, current is passed through a resistive element in the ink-filled channel of the nozzle, heating the neighboring ink very rapidly. When a liquid is heated, typically numerous small bubbles nucleate individually, but with this high rate of heating, on the order of 10^8 K/s, the entire thin layer of ink directly next to the heater nucleates simultaneously. The water-based inks used in TIJ printers have a nucleation temperature of about 280°C. In the drop ejection stage, the expanding bubble, driven by the high vapor pressure of water at the nucleation temperature, pushes a droplet of ink out of the nozzle. The life of the vapor bubble is 10 μs to 20 μs. With such short temperature risetimes and the close spacing of the elements, one issue in TIJ is the crosstalk of temperature to adjacent heating elements. Conditions in neighboring elements will modulate the current necessary to achieve bubble nucleation in a drop-producing resistor, and overheating may occur if the thermal adjacency effects are ignored. Sophisticated algorithms to manage these effects are implemented in many TIJ printheads and/or drive electronics.

When a bubble collapses, the energy it contains is concentrated around the heating element, causing mechanical and thermal degradation of the element. Because of this continual wear, most ink jet printers incorporate the printhead as part of the disposable ink cartridge, which is periodically replaced. As the droplet leaves the nozzle, a number of smaller droplets are formed due to variations in pressure and the eventual separation of a *ligament* of ink connecting the main droplet back to the channel. These smaller droplets will appear on the print medium as a single dot, so long as the medium is close enough to the nozzle, and the print speed is low enough. To ensure that dots of the same size are produced consistently, the ink should refill

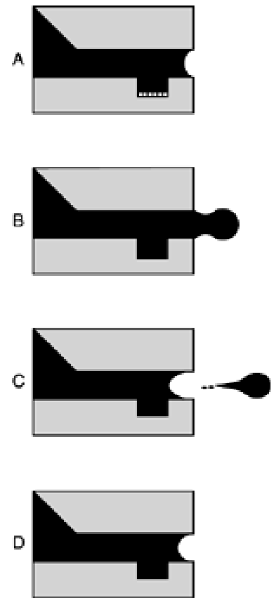

Fig. 3. A schematic representation of the thermal ink jet drop ejection process. In step A the channel is filled with ink and the heating resistor has just begun to nucleate a vapor bubble in the heater well. In step B the heater well is full of vapor and the droplet is beginning to form at the nozzle orifice. In step C the drop has been ejected, leaving the channel empty. In step D the channel is refilling in preparation for the next droplet formation.

the nozzle to near the original state in the nozzle refill stage. The maximum frequency at which dots can be printed is fundamentally limited by the time it takes the bubble to collapse. In practice, the frequency is also limited by the time it takes the channel to refill completely. Making the frequency too high will cause the droplets to become smaller and move faster, a result of the channel being underfilled. In current desktop ink jets, individual nozzles are capable of printing approximately 12,000 dots per second, limited by the channel refill time. Some experimental printheads can eject up to 50,000 dots per second, approaching the fundamental limits imposed by the physics of bubble collapse.

In a TIJ printer, the nozzle array is typically manufactured as part of the disposable ink cartridge. In a piezoelectric ink jet printer, a single piezoelectric printhead, which has higher manufacturing costs but suffers less wear and tear than a TIJ printhead, is used for the life of the printer. The printhead is mounted on a scanning carriage, which moves across the print medium, corresponding to the fast scan direction in laser printers. The print medium is moved line by line through the printer in the process direction. Print quality depends in part on the accuracy of the electromechanical systems that move the scanning carriage and print medium.

When combined with appropriate media, ink jet printers produce high-quality color output, with typically a very low initial investment. The pixel addressability (up to 1200 $\times$ 600 dpi) and color capabilities render these systems capable of near-photographic image quality even in devices priced for the consumer market. However, the page cost for ink jets is relatively high, due to the cost of the ink cartridges, and ink jets are slower than laser printers, limited by the physics of bubble collapse. Full-width arrays exist that avoid the scanning carriage in consumer TIJ printers. However, paper curl and crinkle will continue to be a problem until the solvent wetting problems are solved.

Imaging Technologies

Printing requires a number of technologies in addition to the marking technologies that place pigments on the two-dimensional grid of image pixels. This section discusses technologies and issues that are distinct from the choice of marking technology: page representation, color science, and digital halftoning.

Representing the Page Image. In a shared network environment such as the modern office, multiple different marking devices may be available for printing a document. Each of these systems has its own device-dependent way of preparing the document for output, based on the device's capabilities, configuration, and current state. The device possesses this information, but in general, the user's workstation does not. Thus, a standardized mechanism for describing the appearance of a page is required to allow interoperability of a wide range of devices. A *page description language* (*PDL*) supplies this interoperability. The PDL provides a computationally expressed generic interface between computers, operating systems, and printers. PDLs, such as Adobe's PostScript (14) and HP's PCL, specify the final appearance of the document, but allow the marking device to decide how to achieve that appearance.

Observe that this idea of a device-independent page representation language permits documents to be sent to arbitrary places and printed in geographically distant locations with consistent output (for the most part). Since it is less expensive to send the electronic representation of a document than to send the hardcopy version, this advance has permitted the distribute-and-print model of document delivery. Thus, meeting preread materials may be emailed to the recipients and printed locally if desired. When the recipients meet together, their copies of the documents are very similar. Adobe has extended this concept to include both printers and display devices with their Acrobat product and its Portable Document Format or (*PDF*). Acrobat permits viewing electronically or in hard copy with similar results.

The PDL provides the device-independent abstraction for specifying *what* an image should look like without specifying *how* it should be printed. Notice that the addressability of the device, the unprintable regions of the page, the color mixing formulas, and the orientation of the paper in the paper trays are not part of the page description. Only items necessary to describe the ultimate appearance of the printed page are specified. The printing device itself determines how to render the image so that it matches the author's intent.

A PDL is a special-purpose programming language with variables, loop constructs, and logical decision elements. A PDL interpreter, embedded in the printer or in the printer driver in the attached computer, executes the page description language code and produces the raw data for the imaging device, whether laser, inkjet, or other. Objects in the language include characters, fonts, color, shapes, fill patterns and colors, images, position, and orientation.

On a workstation, the printer's device driver translates the application's internal representation for the document into PDL. Some operating systems supply an application program interface (*API*) for printing that supplies an easily accessible, uniform document representation across applications.

The Science of Color Printing. All visual technologies were first implemented in black and white, followed by a migration to color. Television, computer monitors, and liquid-crystal displays (*LCDs*) are common examples of this progression. Printing technologies have followed the same evolutionary path. Advances have enabled full color printing of magazines, newspapers, and (recently) documents that were previously restricted to monochrome. Judicious usage of color improves the readability, comprehension, and appearance of printed materials. The objective of printing in color is to reproduce an image from some source to a printed page so that the printed image closely resembles the source image. Grasping the issues and technological solutions in color printing requires a basic understanding of the science of color. In-depth coverage of color science and color printing are in Refs. (15) and 16.

Countless chemicals, both natural and artificial, absorb visible light, and their corresponding spectra are nearly infinite in number. Thus, artificially reproducing an image of the observed world would seem at first to be an insurmountable problem. The problem is considerably reduced in dimensionality by the qualities of the sensors of the human visual system. Four types of cells serve as light receptors in the human retina. These cells are divided into two classes, rods and cones. The rods are responsible for achromatic vision, especially at low light levels. The cones, concentrated at the fovea, or optical focal point, respond to colors at medium to high light levels. The three types of cones differ in the portion of the visual spectrum to which they respond. They are labeled red, green, and blue, although their spectral sensitivities overlap considerably. The normal human visual system is capable of distinguishing 300,000 colors. As these cells receive color stimuli, their responses, in combination with sophisticated processing in the visual cortex, enable our perception of color. Thus, different spectra that produce the same stimulus to the eye's cones will be observed as the same color. The problem of reproducing color may be reduced to reproducing the stimulus rather than reproducing the full spectral behavior of objects.

When colors are combined, the result is another, intermediate color. Color mixing can be either *additive* or *subtractive*, depending on whether light or pigments are being mixed. When mixing light, combining red, blue, and green in equal amounts produces white, because each component adds spectrally to the mixture. This additive color model is used to describe the operation of stage lighting, input scanners, and CRT monitors. The printing process is the opposite, since the pigments reflect the ambient light. Combining the subtractive colors, cyan, magenta, and yellow, produces black on the page, because each pigment subtracts from the white of the substrate.

The Pantone system (17) presents one solution to the problem of reproducing specific visual stimuli. The Pantone matching system consists of 1025 color patches and provides recipes for mixing small subsets of the 17 basic Pantone inks in order to produce the patch colors lithographically. Colors in the image to be printed must be matched against a Pantone patch, and then following the corresponding recipe will reliably reproduce that color. While this technique provides a pragmatic solution to a particular problem, it does not supply a quantitative metric space for describing color.

In 1898 Albert Harry Munsell developed the ordered color space that bears his name. This approach places a set of 1450 colors in a three-dimensional space. The colors and their ordering are chosen to sample the continuous color space with a set of patches that are equally spaced perceptually. This useful system permits placement of an arbitrary color in a defined location within a quantized volume, but does not provide numerical values for measurement or analysis.

The quantification of the human visual system's response to spectral inputs has been a longstanding, problem in visual psychophysics. Much of the historical progress in color science has been focused on discovering a set of coordinate axes that provide a perceptually linear description of color space. Such a representation would enable a system for *colorimetry*, the measurement of color. Successful parametrizations are based on the principles of trichromatic representation.

The problem of reproducing color is considerably simplified by the nature of the detection system, because the normal human visual system detects color using only three different types of sensors, that is, the three types of cones in the retina. Direct measurement of the input–output mapping of the system would require detailed neurological knowledge of the cone response curves and of the postprocessing elements of the visual cortex. Because these sorts of direct analytical techniques are currently unavailable, a phenomenological approach is necessary.

In psychophysical color experiments, the assumption is that an arbitrary monochromatic color may be visually matched by an observer using a linear combination of intensities of three primary colors (considered as basis vectors). This indirect calibration of the visual system consists of the determination of the mixing coefficients for each of a series of monochromatic colors. These three coefficients, called tristimulus values, may be plotted as a function of wavelength to yield candidate tristimulus curves.

In an experiment, a single pure monochromatic color is presented to a "normal" subject and compared to a mixture of three primary colors (700 nm red, 546.1 nm green, and 435.8 nm blue) of light. The subject adjusts the intensities of the three primaries, trying to match the mixture to the monochromatic color. In some cases, a match is impossible because theoretically a negative amount of red light is required. That is, the given monochromatic color cannot be reproduced by any mixture of primary colors. In this case, the subject is permitted to add red light to the monochromatic sample to achieve a match. When this is done for all visible wavelengths, a set of color matching functions is obtained. From this set of curves another set (called *XYZ* or tristimulus values) were generated mathematically that were all positive but contained components that are physically unrealizable—a regrettable consequence of the fact that there exists no set of realizable primaries that produce color matching functions that are positive everywhere:

$$\begin{aligned} X &= K\sum R(\lambda_i)S(\lambda_i)\bar{X}(\lambda_i) \\ Y &= K\sum R(\lambda_i)S(\lambda_i)\bar{Y}(\lambda_i) \\ Z &= K\sum R(\lambda_i)S(\lambda_i)\bar{Z}(\lambda_i) \end{aligned}$$

These equations quantify the visual sensation of a "standard" observer to spectral stimuli.

Observe that the tristimulus values are a function of the reflectivity of the object $[R(\lambda_i)]$, the spectral content of the illuminant $[S(\lambda_i)]$, and the response of the detector $(\overline{X}, \overline{Y}, \overline{Z})$, in this case represented by the color matching functions for human eyes. When two objects have the same three tristimulus values, they appear to have the same color when viewed under the same lighting conditions by an "average" observer.

This projection from full spectral space to three values results in certain degeneracies. There are different spectral reflectance curves that can yield the same tristimulus values under certain lighting conditions. This phenomenon, called metamerism, occurs when colors appear identical under certain lighting conditions but may be seen to be different when the illuminant changes.

The tristimulus description of color leads to a three-dimensional coordinate system with axes labeled *X, Y, Z*. However, this coordinate system does not constitute a good metric color space, because it is not perceptually linear. Distances in different parts of this color space represent different degrees of perceived color difference. A color space in which distances have a uniform perceptual interpretation would be more useful for understanding and analyzing color reproduction. Several systems have been proposed, such as the widely used $(L*, a*, b*)$ coordinate system, called *CIE* 1976 or *CIELAB* (after the *Commission Internationale de l'Eclairage*, the official standards body for this subject). The transformation between tristimulus and $L*\ a*\ b*$ is

$$\begin{aligned} L^* &= \begin{cases} 116(Y/Y_\mathrm{N})^{1/3} - 16 & \text{for } Y/Y_\mathrm{N} > 0.008856 \\ 903.3(Y/Y_\mathrm{N})^{1/3} & \text{for } Y/Y_\mathrm{N} < 0.008856 \end{cases} \\ a* &= 500\left[\left(\frac{X}{X_\mathrm{N}}\right)^{1/3} - \left(\frac{Y}{Y_\mathrm{N}}\right)^{1/3}\right] \\ b* &= 200\left[\left(\frac{Y}{Y_\mathrm{N}}\right)^{1/3} - \left(\frac{Z}{Z_\mathrm{N}}\right)^{1/3}\right] \end{aligned}$$

Here $L*$ is the lightness–darkness axis, $a*$ is the red–green axis, and $b*$ is the blue–yellow axis. The triplet $(L*, a*, b*)$ is a point in Cartesian 3-space. The CIELAB color space may also be represented in cylindrical coordinates using hue and chroma. A color's hue and chroma can be calculated from it $L*\ a*\ b*$ coordinates by

$$\begin{aligned} H(a,b) &= \tan^{-1}\left(\frac{b^*}{a^*}\right) \\ C^*(a,b) &= \sqrt{a^{*2} + b^{*2}} \end{aligned}$$

giving the cylindrical coordinates ($L*$, $C*$, h). In this coordinate system, the $L*$ axis describes the neutral gray colors from black ($L* = 0$) to the reference white level (X_N, Y_N, Z_N; $L* = 100$). The reference white level is the "white" that is associated with the image. It is usually obtained from a color measurement of a specular highlight in the image or, in the absence of one in the image, a specular highlight in another image obtained under similar conditions. Hue, the angle around the $L*$ axis, describes the color (red, yellow, green, blue); and chroma, the radial distance from the $L*$ axis, describes the saturation or intensity of the color.

CIELAB in both its Cartesian and cylindrical representations presents a useful (but not perfect) device-independent metric space for describing color. Euclidian distances in this space are a measure of the differences in the perceived colors. These distances are given in units called ΔE. The human visual system can detect color differences of approximately one ΔE unit in this space. If the CIELAB coordinate system were a linear metric space representation of the human visual system, spheres of one just-noticeable-difference (*JND*) radius would have the same size, independent of their location in color space. This is not quite the case for CIELAB, because the human visual system can distinguish colors near the neutral axis ($L*$) better than saturated colors. Other linear color spaces have been proposed but not widely adopted.

The accuracy and reproducibility of a color printing process is commonly described in terms of ΔE differences in the CIELAB space. If the desired and reproduced colors are represented using CIELAB, the distance between them is a measure of the color difference imposed by the printing process. Stabilization of the printing process to produce the desired colors is a complex subject of ongoing research encompassing process control, image processing, sensing, and actuation and is outside the scope of this article.

Each marking device uses a different set of pigments that are mixed together in various proportions to produce the desired colors. A printed color can be represented as the amounts of the various pigments used to print the color. This is called a device-dependent coordinate system, since the coordinate description of a given color depends on the specific pigments used by the device. The colors in an image are typically specified in a device-independent color coordinates, such as CIELAB. The task of the printing system is to transform the device-independent coordinates into the device-dependent coordinates specific to the marking device on which the image is to be printed. This coordinate transformation is multidimensional, nonlinear, and, just to complicate matters, often slowly time-varying. Moreover, the transformation is one-to-many rather than one-to-one, since a single CIELAB color can be generated by multiple pigment combinations in most printers. This transformation, typically between CIELAB and either *RGB* (red–green–blue for monitor phosphors) or *CMYK* (cyan–magenta–yellow–black for printing pigments), is called a *color space transformation*. In practice, this transformation is performed using a lookup table (*LUT*) called the color rendition dictionary (*CRD*). The CRD is similar in function to the recipes used in the Pantone matching process referred to above. Because of storage considerations, the CRD is of limited size and therefore only sparsely populates color space. For this reason it must be interpolated to yield specific values for color conversion (18).

Each color technology is capable of printing a limited range of colors. The color experiments described above, in which some monochromatic color sources required negative amounts of red light, indicate that even sources such as computer monitors cannot reproduce all visible colors. The color range is determined by the pigments used and how they may be mixed. For example, printing a solid composed only of yellow pigment produces a certain $L*a*b*$ value. This value cannot be exceeded by the printing system, since it is not possible to make an image more yellow than the pure yellow pigment. This is not only true of the other primaries, but also of their mixtures. Thus, for each set of pigments a bounded volume in color space is accessible for printing. All colors inside this volume may be constructed from mixtures of the basis set. Colors outside this volume cannot be realized. This volume is called the color *gamut*. Different marking technologies use different pigments and thus have different color gamuts. The differences in color gamuts become a problem when trying to match images printed using different technologies. This is especially true for matching printed images to displayed images. CRT phosphors produce additive colors viewed in transmission, while printed images contain pigments for subtractive colors viewed in reflection. CRTs have a larger gamut than printing inks or toners in most areas except for yellow, where the inks are brighter.

Another consideration in color printing is the quantity of data required to represent a color image. An 8.5 $\times$ 11 in (22 $\times$ 28 cm) sheet printed at 1200 addressable dots per inch contains 16 Mbyte of pixel information. Color printing takes four of these pixel maps, one for each primary color, and stacks them on top of one another for a total of 64 Mbyte of data required to describe a full-page image. A laser printer producing 30 impressions per minute requires a data bandwidth greater than 32 Mbyte/s. Such large data bandwidth requirements are more often associated with video applications, but arise in printing as well.

Halftoning. Display devices such as CRTs and LCDs, as well as marking technologies such as photography and dye diffusion thermal transfer, can reproduce continuous tone (contone) images. That is, these devices can produce pixels at many different intensity levels. The most predominant marking technologies, including offset lithography, electrophotography, and ink jets, can produce only a small number of discrete tone levels, often only two. The challenge is to reproduce a continuous tone image to within some perceptual tolerance using a device that can produce only a few discrete tones. See Refs. 19,20,21,22 for more on halftoning.

Halftoning trades spatial resolution for perceived tone levels in order to reproduce an apparent continuous tone image with a binary marking device. Perceptual studies show that humans can resolve about 8 cycles/mm to 10 cycles/mm at a normal viewing distance. At sufficiently high spatial frequencies, the eye integrates the individual binary pixels in an area, perceiving the result as gray. Digital halftoning uses patterns of binary pixels to trick the eye into "seeing" intermediate tones.

Traditional (i.e., nondigital) halftoning uses a photolithographic method to reproduce a contone image using a regular grid of dots of varying radii. Traditional halftoning, developed over a century ago, made it possible to easily reproduce photographs in print. Similarly, digital halftoning allows the reproduction of images in print, but digital technology allows more flexibility than traditional methods.

In the *ordered dithering* algorithm for digital halftoning, marking device pixels are aggregated into *halftone cells* that form a regular or semiregular grid of the image surface. The cells can have any configuration so long as the pixels in a cell are a contiguous group and the cells tile the image without leaving holes. Typically halftone cells are approximately square and rotated by some angle. The colored pixels in a cell are called the *halftone dot*. Ordered dithering algorithms have four parametric components: screen angle, screen frequency, dot pattern, and level assignment. Screen angle and screen frequency specify how marking-device pixels are aggregated into halftone cells. The screen frequency is the number of halftone cells in a given length, chosen as some fraction of the device addressability. For example, a 600 dpi printer might use a halftone cell spacing of 50 lines per inch (*lpi*). The halftone cells form a grid that is not necessarily aligned with the vertical direction of the media. The screen angle is the angle from vertical of the grid of halftone cells, typically 0°, 15°, or 45° in monochrome printing. The dot pattern and level assignment specify how a halftone cell should be filled for a given contone value. The level assignment is the correspondence between the contone value and the number of pixels in the halftone cell that should be marked, while the dot pattern specifies the order in which pixels are marked. In practice, the dot pattern and level assignment are often combined into a threshold array. Figure 4 provides two examples of dot patterns, Fig. 5 provides an example level assignment, and Fig. 6 shows halftone dots corresponding to the dot patterns for a specific contone value.

Dot patterns are divided into two groups, clustered-dot and dispersed-dot. Clustered-dot ordered dithering, also known as *amplitude modulation* (*AM*) halftoning, is the older method and is analogous to traditional photolithographic halftoning. In clustered-dot patterns, the dot is nucleated at the center of the halftone cell, and pixels are successively added at the edge of the dot. A variety of dot patterns can be used for halftoning, generating variously shaped dots, such as round, spiral, square, and line. Clustered dot patterns are mainly used on marking devices that have difficulty producing single, isolated pixels, such as laser printers. Grouping pixels together in clusters makes the low-frequency components of the two-dimensions Fourier transform of the dot relatively high, causing the dots to be more apparent to a human observer and decreasing the effective resolution.

In dispersed-dot halftoning, pixels are turned on throughout the cell with increasing tone value. Dispersed-dot algorithms reduce the visibility of halftone patterns by reducing the low-spatial-frequency components of

4	8	10	1
11	15	14	5
7	16	13	9
3	12	6	2

Round dot

1	9	3	11
13	5	15	7
4	12	2	10
16	8	14	6

Bayer dot

Fig. 4. Dot patterns of a 4 × 4 square halftone cell for a clustered-dot dither (round dot) and a dispersed-dot dither (Bayer dot), providing 17 levels of gray from 0 (darkest) to 16 (lightest). For a specified gray level, the pixels that are numbered strictly higher than the gray level are marked, while the others are unmarked.

Threshold	8	24	40	56	72	88	104	120	136	152	168	184	200	216	232	248	
Gray level	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

Fig. 5. A "linear" level assignment used to map an 8-bit contone value into a gray level for use with a dot pattern as in Fig. 4. The resulting tone reproduction curve will be similar to Figure 7.

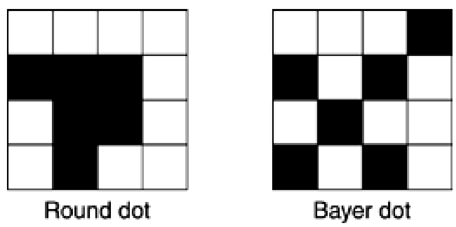

Fig. 6. Gray level 10 for the dot patterns from FIg. 4. Under the threshold array of Fig. 5, contone values 152 to 167 correspond to this level.

the fill pattern. Bayer found a set of necessary and sufficient conditions on the dot pattern that minimize the low-frequency components of the two-dimentional Fourier transform for regions of uniform tone level [20]. Figure 4 shows the Bayer dot pattern, one of several dot patterns that satisfy the conditions. While a dispersed-dot pattern can provide better homogeneity in regions of uniform tone than a clustered-dot pattern, regular structure is still observable.

The *tone reproduction curve* (*TRC*) is a graph of the input contone value versus the percentage darkness of a halftone cell, often measured empirically. An example is shown in Fig. 7. The TRC is monotonically increasing, but not, in general, linear. The number of contone levels (256 for an 8-bit representation) is typically greater than the number of pixels in the halftone cell, in which case the TRC is a piecewise constant function. The TRC is directly affected by changing the level assignment in the ordered dither algorithm. In this way, TRC is used as a control variable in printing to obtain consistency of document appearance (23).

Ordered dithering balances the tradeoff between tone level and screen frequency. Using small halftone cells reduces the number of perceived tones that can be produced within the cell. When a contone image with large areas of slowly varying tone is reproduced using insufficient tone quantization, the reproduced image will have visible contour lines between regions of adjacent contone levels. Larger halftone cells allow more perceived tones to be produced, reducing contouring, at the cost of making the halftone screen more visible.

Blue-noise dithering (19), so called because this technique yields a spatial-frequency spectrum for a uniform region of gray that is zero across low frequencies and flat across high frequencies, does away with halftone cells and screens in order to break the regular patterns that occur in ordered dithering. Ordered

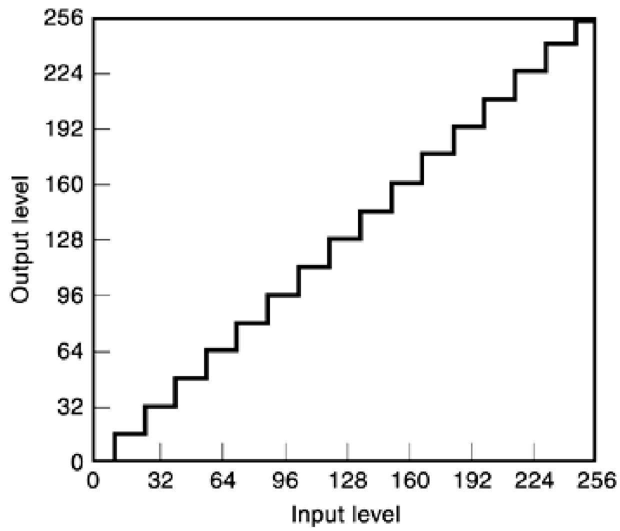

Fig. 7. The input–output relationship for the 16-pixel halftone dot of Fig. 4, using a filling algorithm that yields a linear TRC. The input level is often specified as an 8-bit quantity, and the output level is some measure of the darkness of the resulting halftone dot. Note that the TRC is both piecewise constant and monotonically increasing.

dithering is a point process, that is, only the contone value and threshold are needed to determine whether a pixel is on or off. In contrast, blue-noise techniques use information from neighboring pixels to decide whether to turn a pixel on or off. *Error diffusion*, originally developed by Floyd and Steinberg, is the best-known blue-noise technique. For a given marking-device pixel, error diffusion attempts to distribute the error between the desired contone value, taken to be a real number between 0 and 1, and the printed tone, either 0 or 1, across the neighboring pixels. Note that concepts of screen angle and screen frequency no longer apply in blue-noise dithering. This succeeds in breaking the regular patterns that occur in dispersed-dot ordered dithering, but blue-noise methods are much more computationally intensive than ordered dithering. One compromise is *blue-noise masks*, which are essentially very large ordered dither arrays (256×256 as compared to 16×16) that have blue-noise characteristics (22).

Both dispersed-dot ordered dithering and blue-noise dithering rely on the ability of the marking device to produce single, isolated pixels reliably. These techniques are in use on ink jet printers, but electrophotographic printers largely continue to use AM halftoning. Moreover, the optimality and analysis of these algorithms assume an ideal marking device, able to perfectly fill a pixel. Printing of a larger dot than intended, called dot *gain*, may be considerable in ink jets, depending on the interaction of the liquid ink with the print medium (e.g., card stock versus newsprint). Difficulty printing isolated pixels, called *dot loss*, can be observed in many laser printers. A marking device may variously exhibit dot gain and dot loss under different circumstances. Current research in model-based halftoning uses models, either physical or stochastic, of the marking device to improve halftoning performance (22,24). In this light, adjusting the TRC through the threshold assignment to improve the appearance of ordered dither halftoning may be viewed as one of the earliest model-based techniques.

Another new area in halftoning is the hybrid, or *green-noise*, techniques, which attempt to combine the good homogeneity of blue-noise techniques with allowing local clusters as in AM halftoning, making the technique more suitable for electrophotographic devices and other devices that have difficulty producing isolated pixels (22,24). The term green noise is used because one attempts to make the frequency spectrum 0 for low and high frequencies and flat for intermediate frequencies.

Thus far this section has covered halftoning of a grayscale contone image. Halftoning a color image brings about additional complications (20,21). Color printing, as discussed in the previous section, uses three or more pigments to form the desired colors. Modern printers generally use four colors, cyan, magenta, yellow, and black, in order to achieve a larger color gamut than just three colors. Six or even more colors can be used to extend the color gamut even further. The amount of each pigment is specified as a contone value, but once again, many marking technologies, such as ink jet and electrophotographic printers, only produce binary pixels of each pigment. Each pigment is halftoned and printed in close proximity to approximate the desired color. The halftone dots for the different pigments can be printed either directly on top of each other (dot-on-dot), next to each other (dot-off-dot), or at different screen angles with some overlapping (rotated dot). Dot-off-dot is used in computer monitors, where the three pixels, red, green, and blue, are grouped closely together and appear to yield a single color to the human observer. A pure dot-off-dot is not possible in printing, since for dark colors the area coverage will be so large that the pigments must overlap. On the other hand, dot-on-dot halftoning is sensitive to misregistration. That is, slight error in positioning the different color screens can greatly degrade the color reproduction. Also, it has been found that dot-on-dot yields a smaller color gamut than dot-off-dot or rotated-dot.

Both dot-off-dot and rotated-dot are susceptible to moiré interference patterns generated by overlaying patterns with similar spatial frequencies, which are visually distracting. In four-color printing with traditional screen angles, cyan 75°, magenta 15°, yellow 90°, and black 45°, the typical interference pattern is known as a *rosette*, for its flowerlike shape. For rotated-dot printing, the screen angles of the individual colors can be adjusted to reduce the moiré, but it cannot be overcome altogether.

In Ref. 20, the list of Olaru's design critera for digital screening technologies is presented. Among other criteria, a color screening technology ideally (i) should be free of moiré artifacts, (ii) should have no spatial modulations across final color output, (iii) should have no dot-on-dot overlapping for any printable middletone hues, (iv) should meet the common accepted standards for color appearance over the whole spectrum, and (v) should have no restriction on the number of pigments used. Goals such as these push the development of color extensions of the stochastic halftoning techniques such as blue-noise and green-noise dithering. These techniques, which are free of screen angle and screen frequency, may be able to provide these characteristics at the price of extra computation (20,21,22).

Media Handling. In copying and printing devices, the images ultimately end up on a medium (sheet). The media-handling system is responsible for the transport of media from the input stack, through the marking process, and out the finishing station.

The market for marking devices has placed upward pressure on the speed of the system. Thus, the media paths in the device must present the sheets to the marking elements more quickly and with uncompromised accuracy.

Print jobs may require any one of a variety of media stocks. Media qualities include basis weight, stiffness, surface properties (smooth or rough), and composition (paper or transparencies), among others. There are about 4000 different types of media with various values of these properties. If the marking device is incapable of transporting the stock required for a particular job, the job should be sent to another device that can. *Media latitude* refers to the range of media qualities that the system is able to transport reliably through the device. Wider latitudes indicate that the system can transport a wider range of media and hence is capable of accepting more types of printing jobs.

Media properties are not constant for each media type. For example, paper is hygroscopic—it absorbs water from the air, making it somewhat soggy. While a given medium may perform well when dry, water absorption dramatically changes certain important properties of the sheet. As a result, media in the tray that fed just fine on Friday may jam in the machine Monday morning. In the other extreme, in duplex (two-sided) printing modes, media pass through a fuser in electrophotographic systems. The first pass removes most of the water from the sheet, changing its properties and even its dimensions as it enters the second pass. These changes must be detected and compensating adjustments made by the media handling system.

The critical issues in media handling are extracting individual sheets from the input tray(s), setting and maintaining orientation, and avoiding jams.

The paper path begins at the feeding station. The sheets are typically in a stack in the input side of the device. They must be extracted from the stack one at a time and fed through the various process steps. The removal of a single sheet is complicated by a number of factors. For example, media are usually dimensioned by guillotine cutting. This process tends to "weld" the edges of sheets to one another, making separation difficult. If the stack is compressed, extracting a single sheet every time is problematic. If the sheets are slippery, the friction-based input tray feeder mechanisms may fail to separate them. If more than one sheet is extracted from the input tray, the system will attempt to transport the multiple sheet packet through the system. At some point, the sheets will separate and cause problems and perhaps a jam.

Sheets must have a particular orientation with respect to the feed direction. There are two options for moving rectangular sheets (short-edge or long-edge feed). In either case, small changes in the angle of the sheets will be readily visible because the image edges will not be parallel to the paper edges. This phenomenon is known as *skew*. Media orientation may be altered by differences in the friction of the drive rollers somewhere in the system. This skew must be detected and removed by rotating the sheet into the proper orientation. The paper-path literature (patents and articles) contains many techniques for both sensing and adjustment of skew.

Media jams are a large source of customer dissatisfaction. Jam clearance requires that the user open the machine, gain access to the media path elements, and remove sheets from the printer. There are two classes of jams signaled by the media path. *Hard* jams are those that occur when the media are physically distorted by wrinkling or crumbling. Such media do not transport well and will frequently enter areas of the printer that were not intended for them. At the first sign of such a problem, the machine typically shuts down. *Soft* jams are those that occur when the media were expected to arrive at a sensor location within a certain fixed time window but do not. This condition indicates that the sheet timings are outside their design latitude window. Typically soft jams also prompt a machine shutdown in anticipation of the hard jam that would occur if they were ignored. This is the reason that often the sheets removed during jam clearance are still uncrumpled. The system stopped before something catastrophic occurred.

Light flux interrupters are frequently chosen as the sensors used in the media path. These U-shaped devices have an LED emitter on one side of the U and a photodetector facing it on the other. As a sheet passes between the LED and the detector, the illumination is occluded and the sensor detects the sheet. For this reason, many systems require transparencies with an opaque stripe down the side. Media-handling system architectures often call for media presence sensors spaced roughly at the process direction size of the media being transported. This is so that each sheet is seen by at least one sensor all the time.

The media path geometry is frequently simplified in order to reduce the opportunities for these disturbances to lead to failures. Bends in the media paths are avoided or managed in order to accommodate a wider latitude of paper thickness and stiffness without failure. The active control of the drive elements in conjunction with the multiplicity of sensors in the system is an ongoing subject of research and engineering technology efforts.

Registration. In monochrome printing systems, where there is only a single color separation, the registration problem is restricted to the alignment of the image to the sheet. The acceptable tolerance levels are in the range of a few tenths of millimeters. In color printing systems, regardless of technology, all the color separations must be aligned with one another within a few tens of micrometers in order not to be visually detectable.

A variety of technologies is available for combining color separations into a page. All of these require that a set of primary colors be mixed or layered on top of each other in order to give the illusion of a continuum of visual stimuli. Whether the technology requires multiple passes over a substrate, the passage of a partially formed image between print stations, or the passage of a four-color printhead over the media, every one of these requires that the timing between the events that lay down the component colors be accurately synchronized. If they are not, then the individual separations become visible and image quality suffers.

In many of these technologies, the imaging stations that supply the component colorants are widely separated, often by more than one page length. In order to achieve the required registration accuracy, the system is calibrated using a set of registration targets. These targets consist of crosshairs printed in each of the colorants and positioned in multiple places across the page. Prior to calibration, these crosshairs do not line up on top of one another. The calibration process brings the various test targets into alignment and involves shifting, scaling, and rotation of the image or hardware in order to achieve exact overlap.

Early printing technologies relied on the stability and robustness of the machining of the hardware components to maintain registration accurately. Recently, automated techniques for both setup and run-time control are used for this important function.

Motion Quality. Halftoning provides a convenient way to simulate a continuum of colors within the printing gamut. However, the halftone dot frequency provides a periodic structure to the image that is subject to visible disturbances. Motion-induced variations in this periodicity produce image quality artifacts that are readily detectable. The severity of these defects is a strong function of the frequencies at which they occur. Throughout the system there are elements that carry the image or portions thereof in its various forms from place to place within the printer. Often these elements receive additional image content as they flow through the system. There are many mechanical opportunities to introduce motion variations within the image.

The detection and isolation of these velocity disturbances can be a tedious exercise. A Fourier analysis of the image is made, and the disturbance frequencies are analyzed. Having discovered the responsible frequencies, the sources must be identified and eliminated or reduced below a visible threshold level. Drive elements must be carefully chosen not to excite resonances in the hardware that affect the images. Gear tooth ratios and roll runout can cause periodic disturbances that must be eliminated when they appear in the image. Frame vibrations can also contribute to this problem.

These problems do not yet yield to control techniques, because of the difficulties in automatic detection and remediation of the problems. Detection requires examination of printed test patterns in two dimensions followed by Fourier analysis. Remediation would require active velocity compensation of the drive elements in phase with the disturbances. These problems have not yet been solved in commercial products, and thus the analysis and treatment of the problem are usually addressed in product engineering.

Conclusion

The subject of printing is much more multifaceted than it appears to the casual user. The complexities of physics, chemistry, electromechanics, and computer science have been almost completely hidden by the companies responsible for the design and manufacturing of the hardware, embedded software, and printing system. The objective of this article has been to introduce the reader to some of what lies under the covers. The authors' desire is to indicate that the scientific and engineering disciplines that undergird the printing process provide a feast of fascinating and unsolved problems.

The paperless office has been a gleam in the eyes of many futurists, but the requirement for hardcopy documents does not seem to be abating. In fact, paper consumption continues to rise as people print their email prior to reading it. North America alone consumes about 15 million tons of uncoated paper annually. In the absence of a portable, compact, robust electronic alternative to replace paper, this trend will persist. Printing remains a topic of academic and industrial interest, and innovative solutions will continue to supply profitable competitive advantage for printing equipment companies for years to come.

BIBLIOGRAPHY

1. A. J. Sellen R. H. R. Harper *The Myth of the Paperless Office*, Cambridge, MA: MIT Press, 2002.
2. T. S. Jewitt Traditional impact printing, in P. Gregory (ed.), *Chemistry and Technology of Printing and Imaging Systems*, New York: Blackie Academic & Professional, 1996.
3. C. Shapiro (ed.) *The Lithographers Manual*, Pittsburgh, PA: Graphic Arts Technical Foundation, 1983.
4. C. B. Duke J. Noolandi T. Thieret The surface science of xerography, *Surface Sci.*, **500**: 1005–1023, 2002.
5. D. M. Pai B. E. Springett Physics of electrophotography, *Rev. Mod. Phys.*, **65** (1): 163–211, 1993.
6. L. B. Schein *Electrophotography and Development Physics*, New York: Springer Verlag, 1988.
7. J. A. F. Plateau On the recent theories of the constitution of jets of liquid issuing from circular orifices, *Phil. Mag.*, **12**: 286 (1856).
8. F. R. S. Raleigh On the stability of jets, *Proc. London Math. Soc.* **104**: 4–13, 1878.
9. R. W. Kenyon Ink jet printing, in P. Gregory, (ed.), *Chemistry and Technology of Printing and Imaging Systems*, New York: Blackie Academic & Professional, 1996.
10. D. J. Drake *et al.* Thermal ink jet printhead fabricating process, US Patent No. 4,789,425, 1988.
11. S. F. Pond *Inkjet Technology and Product Development Strategies*, Carlsbad, CA: Torrey Pines Research, 2000.
12. A. I. Pan Advances in thermal ink jet printing. *Input/Output Imaging Technologies*, pp. 38–44, 1998.
13. I. Rezanka Thermal ink jet—a review. *Color Hard Copy and Graphic Arts*, **1670**: 192–200, 1992.
14. Adobe Systems Incorporated, *Postscript Language Reference Manual*, Reading, MA: Addison-Wesley, 1990.
15. J. A. C. Yule G. G. Field *Principles of Color*, 2nd ed., GAFT Press, 2001.
16. R. W. G. Hunt *The Reproduction of Color*, 6th ed., London: Fountain Press, 2002.
17. Pantone, Inc., *Pantone Color Formula Guide*, Carlstadt, NJ: Pantone, Inc., 2000.
18. W. F. Schreiber Color reproduction system, US Patent No. 4,500,919, 1985.
19. R. Ulichney *Digital Halftoning*, Cambridge, MA: MIT Press, 1987.
20. H. R. Kang *Color Technology for Electronic Imaging Devices*. Bellingham, WA: SPIE Optical Engineering Press, 1997.
21. H. R. Kang *Digital Color Halftoning*, New York: IEEE Press, 2002.
22. D. L. Lau G. R. Arce *Modern Digital Halftoning*, New York: Marcel Dekker, 2001.
23. T. E. Thieret T. A. Henderson M. A. Butler Method and control system architecture for controlling tone reproduction in a printing device, US Patent No. 5,471,313, 1995.
24. D. Kacker T. Camis J. P. Allebach Electrophotographic process embedded in direct binary search, *IEEE Trans. Image. Process.*, **11**: 243–257, 2002.

READING LIST

W. R. Wehl Ink-jet printing: The present state of the art, *Proc. VLSI and Computer Peripherals*, 1989, pp. 46–52.

E. M. Williams *The Physics and Technology of Xerographic Processes*, New York: Wiley, 1984.

RICHARD E. GROFF
University of Michigan
TRACY E. THIERET
Xerox Corporation