Digital Television

Edit. Note: In response to requests from many film workers and others not yet inducted into the intricacies of digital television, the following introduction has been prepared by David A. Howell of the Society's Editorial Staff, with substantial parts and suggestions contributed by many members of the Board of Editors. This primer provides background information, the philosophy and fundamental science of digital television, but little of the engineering. The next paper, by E. S. Busby, is also a digital television background paper, but here the emphasis is on showing the principles from an engineering point of view, i.e. it is shown how television can be digitized by using analog-to-digital and digital-to-analog converters and other equipment.

It is recommended that the reader examine both papers and select the approach most suitable for him. Digital television is topical, challenging, abstruse and highly technical. We hope that, having the advantage of two background papers, many more *Journal* readers will benefit from the digital television papers that are sure to come in the future.

A Primer on Digital Television

By DAVID A. HOWELL

Introduction

When the average person hears the phrase "digital television," he is inclined to respond, "I know what 'digital' means in terms of a digital clock or watch, but how can television be digital? Where are the numbers?" The answer is that the principle of digital television involves the use of numbers in the generation, manipulation, recording and transmission of television images, but — unlike the digital clock — the actual numbers are not displayed.

It is the aim of this paper to discuss the fundamentals of digital television in a simplified way, taking into account the widely varying backgrounds of *Journal* readers. (Readers well versed in electronics or video systems or computers may wish to obtain their introduction to digital television via the somewhat more substantive and challenging paper by E. Stanley Busby, Jr., which immediately follows this one.) We shall in this paper compare digital and conventional (we could say "analog") television and examine some of the encouraging possibilities and great difficulties facing developers of digital television systems. It is hoped that this paper and the one following will provide sufficient information for readers to derive the maximum benefit from the more technical and specific papers that follow.

Concepts Needed to Understand Digital Television

In order to grasp the principles of digital television, one should be fairly familiar with conventional television. One should also know the significance of certain terms which may either be wholly new or have special meanings in the field of electronics. Such terms include:

- analog and digital
- continuous vs discrete data
- resolution
- nonlinearities and phase distortion
- sampling and quantizing
- modulation techniques
- encoding
- bandwidth

All of these concepts will be covered at least briefly in this paper.

The most direct way to approach these mysteries is to clarify by examples the distinction between *analog* and *digital*. There are several analog and digital devices which by now are familiar to almost everyone. A simple mercury thermometer and an ordinary slide rule are good examples of analog devices; in the thermometer, the height of the mercury as measured on an appropriate scale is proportional — or analogous — to the temperature of the surroundings; the slide rule is a kind of analog computer in that the distances between the scribed lines are made proportional to the logarithms of numbers so that by adding distances one can in effect multiply numbers.

The ubiquitous electronic pocket calculator and of course the digital clock are obvious examples of digital devices. Comparing the slide rule and the pocket calculator, we can illustrate two of the essential differences between analog and digital systems. Analog devices operate with *continuous* data, which means that over their operating range any desired number can be set in or read out. Digital devices, on the other hand, deal with *discrete* or stepped data: whatever the least significant digit is, it can only change by at least one whole unit. And this distinction relates directly to the second essential difference — that of resolution. Resolution, in the sense used here, provides a measure of the smallest possible increment of change in the variable output of a device. The resolution of any analog device depends on the exactness of the analogy that is used, on a factor which relates the magnitude of a number to the precision of its representation by the device (the scale factor) and to some extent on the estimating skill of the operator. The resolution of the readout of a digital device depends exclusively on the number of significant figures or "places" that one is willing to pay for. (Four and six-digit calculators are quite cheap, while twelve-digit ones are more expensive — and more accurate than most people need.) The accuracy of the readout obtained with a digital device *in no way* depends on the operator's estimate.

We shall see later in this paper that when it comes to digital television, extra digits are very important but also very expensive in terms of money, complexity and something called bandwidth.

Conventional Television in Brief

Conventional television is analog television. When a scene is scanned, the current from the cathode of a television camera tube increases with the amount of light — from each spot in the scene — falling on it. This variable current is used to obtain a signal which is processed in the studio to yield a broadcast signal. The signal that is radiated from the broadcasting antenna passes through the atmosphere to a receiving antenna where it is sensed as a weak "field strength" — a few millivolts per meter of antenna or less. In the receiver the process is reversed: an increasing positive signal voltage applied to the control grid of the picture tube causes an increase in luminance at a given point on the screen. Thus, in a television camera, an electronic signal is used as an "analogy" to represent a pattern of light, while in the receiver, a pattern of light is generated as an "analogy" to represent an electronic signal. For color transmissions, red, green and blue spot brightnesses are converted to currents in the camera and back to tiny colored light spots in the receiver. The eye of the viewer blends the three colors of the spots additively to produce a full-color picture, and since the pictures are sent in "frames," persistence of vision permits the viewer to perceive smooth and continuous motion — just as with motion pictures.

From the small, round, often snowy, black-and-white television pictures of thirty years ago to the large, sharp, colorful pictures of today, the principles have scarcely changed. Analog television has been greatly refined but it is still analog. One might compare it further with the slide rule: both have been used widely for a long time in basically their present form; both require a good eye and good judgment to set up and use; both have had to contend with problems of resolution or reading error; and both are now facing competition from digital devices (pocket calculators and digital systems for television studios).

Dealing with the Shortcomings of Conventional Television

The word "electronics" did not appear in any dictionary printed before 1940; yet whole new industries — electronic data processing and high-speed communications (especially television) — have grown up since that time, based on electronics. The need to process greater and greater volumes of data at ever higher speeds assured that engineers would seize quickly on new technology: electromechanical relays were supplanted by vacuum tubes and these in turn gave way to transistors and then to integrated circuits. A number of techniques of data processing and signal processing had been known in principle for a number of years — but their actual application awaited development of electronic systems capable of sufficiently rapid switching. Thus, digital television systems, based on some of the same principles as the digital voice transmission systems used by the military in World War II, are only now becoming really feasible.

When digital television was recognized, just within the last few years, as a potentially marketable commodity, there was immediate interest in the possibility of developing equipment to perform image processing and control, special effects, image storage (as in videotape recording) and perhaps even transmission and reception.

The obvious place to start was with the easily distorted voltage waveform produced by the conventional television camera. The quality of the picture on the viewer's television set is limited ultimately by the purity, accuracy and stability of this waveform. Yet, there are certain impairments — degradations of the signal and image — that tend to be cumulative when the signal is relayed many times from one link to the next as in tandem transmission. Signal nonlinearity, for example, can with each repetition result in greater and greater loss of the intermediate ranges of contrast. In the film industry, such nonlinearities are evident in the way that spurious dye characteristics multiply in duplication stages. Phase distortion is a cumulative problem with any wideband communications system, but especially with high-speed data and image transmission systems. In television, the effect is of fringing — like diffraction rings — at edges where the contrast changes abruptly. It is due to unequal delay (phase shifting) of different frequency components within the signal as they pass through different impedance elements — filters, amplifiers, ionospheric inhomogeneities, etc. The optical equivalent is seen when white light is passed through a prism and dispersed into colors: red is delayed or refracted more than blue.

Television engineers have worked on these and other problems for decades. Sometimes one problem is "solved" by making another even worse. Even with the most up-to-date principles and hardware, impairments due to phase distortion and modulation products continue to cause difficulty. Now, however, digital handling of television signals is opening up possibilities that have been difficult or impossible with conventional analog television systems. These possibilities include (1) retiming signals from videotape recorders or from satellite transmission systems in which Doppler distortions have been introduced, and (2) regenerating signals at intervals along a transmission path with the aim of minimizing the effect of distortions and noise introduced by the transmission medium.

Digital Television — Sampling and Quantizing

In digital television systems, the voltage waveform that is generated by the camera to represent the brightness of a picture element is measured or "sampled" millions of times each second. Each sample is then "quantized": it is assigned the number of the nearest step that the system can resolve. Two kinds of quantizing error must be considered here because (1) a sample that is exactly between two steps can be quantized either way and (2) all digits beyond the resolving limit will be dropped. (The second kind of error is evident on a six-digit calculator where, for example, the number $3\frac{1}{3}$ would be quantized to 3.33333 and all subsequent 3's would be dropped.)

In electronics, such "quantizers" are called analog-to-digital (A/D) converters. Many of them have enough digits — we could say "places" or "significant figures" — to be able to resolve to one part in 256 (0.39%) or better. With this level of resolution, the common 1-V peak-to-peak signal used in video processing would have each sample rounded off to the nearest multiple of 3.9 mV. Experimental quantizers have been built that can resolve at least to one part in 1024.

(The still photographer often uses the principle of digitizing or quantizing when he prepares to print a negative with an enlarger. He may either make a series of test exposures (say 10, 20, 40 and 80 seconds long) or he may make one exposure using a step wedge. Either way, one part of the resulting test print is likely to be very close to the density and contrast that he wants, and he can choose his future exposures with the enlarger accordingly. Motion-picture laboratories have used step wedges to quantize their printing exposures in a similar way. Photographers, furthermore, either minimize nonlinearities by hewing to the straight-line part of the H & D curve or else use them creatively by going to the toe or knee.)

On the face of it, sampling and quantizing would seem to work to our disadvantage. After all, instead of taking the whole voltage waveform as is done in conventional television, we only take pieces of it. Furthermore, there must inevitably be a quantizing or rounding-off error made each time we substitute the nearest quantizing unit step for an actual measured value. Nevertheless, if we take our samples accurately and frequently (the Nyquist Sampling Theorem says that the interval between successive samples must be equal to or less than one-half of the period of the highest frequency present in the signal) and then quantize in small steps to minimize the rounding-off errors, we *can* use the collected and quantized samples to recover a waveform indistinguishable from the original.

Still, why is it better to have this collection of discrete samples than the whole continuous waveform? The answer is that quantized samples can be encoded to make up a new signal that in principle can be processed, recorded, transmitted and ultimately converted back into an analog signal for playback — all with much less likelihood of errors than the original signal could ever be. Instead of manipulating the waveform itself — or some characteristic of it, such as its instantaneous amplitude — we can manipulate *information about the waveform* and this information can be used eventually to reconstruct the desired waveform.

Modulation and Encoding

To see why this is so, we must examine *modulation techniques* and *encoding.* There are many forms of communication, but virtually all of them impose some kind of intelligence on some kind of "carrier" and this imposing is technically called "modulation." Most people know of amplitude modulation and frequency modulation, but probably not many know the significance of these terms or the fact that they can be considered as examples of analog modulation (see Fig. 1). Various kinds of discontinuous transmission (also called pulse-type or discrete or sampled) are also possible (Fig. 2). Among these are pulse-amplitude modulation (PAM), pulse-duration modulation (PDM) and pulse-position modulation (PPM).

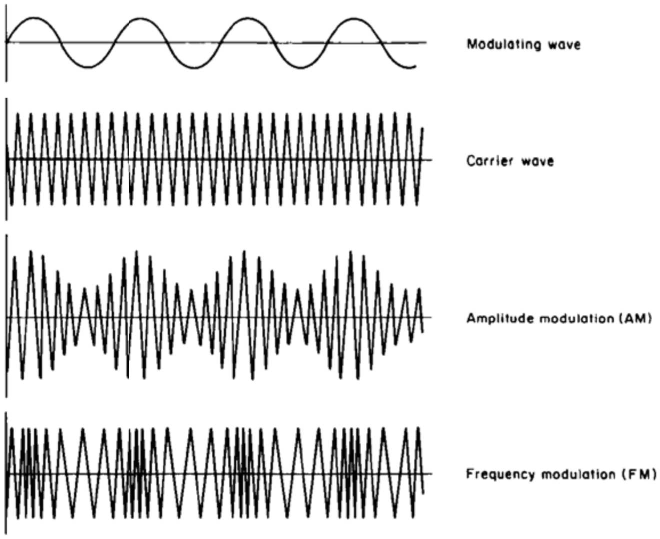

Fig. 1. Modulation for continuous-wave or "analog" transmission. A characteristic of one continuous wave (the carrier) is varied in accordance with another continuous wave (the modulating wave). In AM, the varying amplitude (vertical scale) of the modulating signal - proportional, say, to the signal put out by a microphone - is impressed on the carrier wave to make it vary in amplitude. In FM, the modulating signal causes the carrier to vary in frequency while the carrier amplitude is kept constant.

Under some conditions, it is advantageous to *encode* the information represented by a signal that has already been pulse modulated according to one or another scheme. For example, if a PAM signal is encoded, we can obtain a pulse-*code* modulated (PCM) signal. Britain's Alec H. Reeves invented PCM in 1939 and found it highly noise-resistant; it is therefore well-suited to such communications tasks as digital television.* The encoding in this case would be accomplished by deriving a number proportional to the amplitude of each pulse (see Fig. 2c); *the number describing* the pulse amplitude (rather than the pulse amplitude itself) is then expressed in the form of several discrete pulses and is transmitted in this way. Figure 3 illustrates this for the greatly simplified case of the noise-resistant transmission of a voice signal.

Use of Binary Notation

Because the simplest electronic switches have only two essential positions, "on" and "off", it is often very convenient to use a binary code to represent the sampled amplitude levels. Binary codes use only two digits, zero and one, and these binary digits (called "bits") can be easily represented (as shown in Fig. 3) by negative and positive pulses. In some systems, nonexistence of a pulse is a "zero" and the existence of one is a "one"; there are many ways to implement a binary code.

The interpretation of this binary notation is straightforward. Where, in decimal notation, each "place" that a digit is moved to the left multiplies the value of that digit by ten, in binary notation, each place moved multiplies the value of the digit by two. Thus "10" in binary is "2" in decimal; "101" in binary is "5" in decimal; and "1110111" in binary (Sample 6 in Fig. 3) is 119 in decimal.

The "places" available in a given binary number code limit both the largest binary number that can be expressed and the resolution that can be obtained in that binary code. The largest binary number that can be expressed with *n* bits is $2^n - 1$, and the resolution is limited to one part in 2^n. Thus, four bits give a resolution of one part in 16, five bits one in 32, six bits one in 64, seven bits one in 128 and eight bits one in 256. (For comparison, the highest three-place decimal number is 999 and this permits a resolution of one in 1,000 — since zero is also a possibility.) It appears that eight-bit resolution (where eight bits "describe" each sample) may be the minimum that is acceptable for broadcast television.

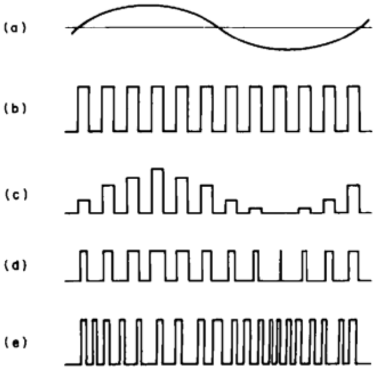

Fig. 2. Pulse modulation techniques available to the communications engineer. (a) The modulating signal; (b) the carrier; (c) pulse amplitude modulation (PAM); (d) pulse duration modulation (PDM) - also called pulse width and pulse length (PWM and PLM); (e) pulse position modulation (PPM).

Pulse Code Modulation and Correlation Techniques

The effectiveness of pulse code modulation is well illustrated when it is used with computer correlation techniques — although no one has suggested that such techniques be used with television systems. If pulse code modulation of the kind shown in Fig. 3 is used, we know *a priori* that the signal is either a positive pulse or a negative one. Computer correlation techniques can then be used at the point of reception to determine if the signal correlates with positive or with negative. In correlation detection, a signal is compared point to point with an internally generated reference. The output of such a detector is a measure of how closely the input signal resembles the reference. The reference signal is at all times a "best guess" by the computer of what the input signal should be at that time. The noise (which can be much larger than the signal), having a random nature, will be uncorrelated and therefore eliminated when correlation techniques are employed.

Pulse-code modulation and computer correlation detection have been used very effectively in aerospace applications such as constructing radar maps of the planet Venus. Although the signal sent out has many kilowatts of power, the return signal is only on the order of a quadrillionth of a watt and deeply embedded in noise. Nevertheless, computer correlation detection of a PCM signal can not only show that the signal is there but also derive useful information from it.

Limitations and Prospects of Digital Television

Everything we have said about digital television so far has been positive. The alert reader, however, should be just about now asking, "What is the 'catch'?" The big problem with digital

---

* Phase-modulated PCM is also popular and might be considered to be derived from pulse-position modulation, but in the present discussion we shall deal only with the form of PCM that in essence is encoded PAM.

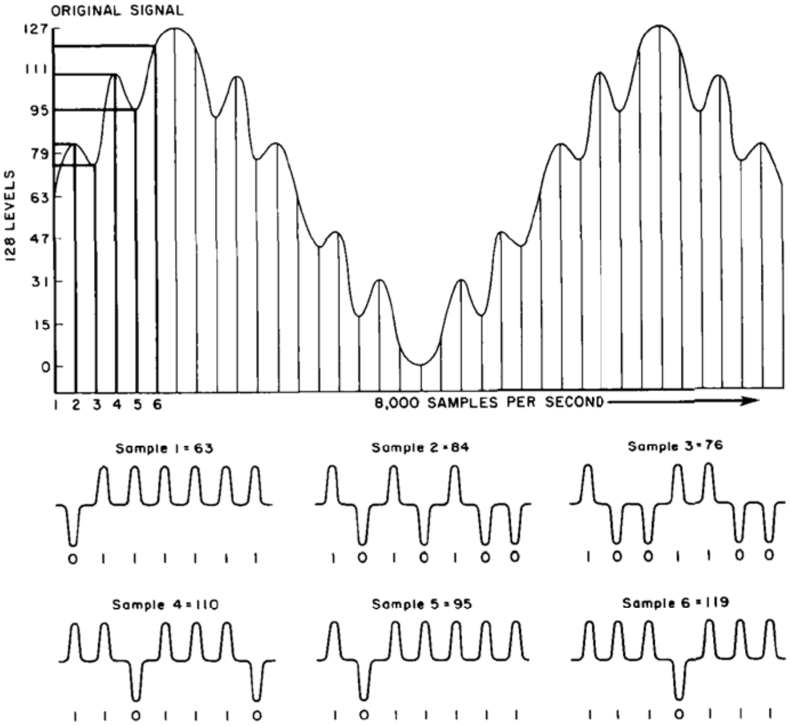

Fig. 3. When PCM is used to transmit a voice signal, the amplitude of the original wave is sampled 8,000 times per second and the sampled values are translated into binary code groups consisting of 1's and 0's (positive and negative pulses). Code groups seven binary digits long make it possible to measure the instantaneous amplitude of the original wave to an accuracy of one part in 128. Often, an extra digit is added to each group of seven for signalling and other functions, making the bit rate 64,000 per second (8,000 × 8). At the receiver, the sequence of pulses is decoded to obtain the original signal.

television, at present, can be summed up in one word: *bandwidth*. Bandwidth is the difference (often measured in megahertz) between the upper and lower limits of a frequency band. The Nyquist Sampling Theorem and the need for at least eight-bit binary numbers to achieve adequate resolution conspire against us to aggravate the wide bandwidth problem that is already the preeminent curse of television. Because an ordinary color television signal has a bandwidth of about 4.5 MHz and every individual cycle can be important, we must sample approximately 11 million times each second (as opposed to the "mere" 8,000 times per second in the voice example considered earlier). Using an eight-bit binary number to describe each sample, it is obvious that we would need to be able to process about 88 million bits per second.

To drive home the sheer magnitude of this number, consider that transmitting this quantity of information is equivalent to sending the Bible (Old and New Testaments) more than ten times in the course of a second. And that is with the most efficient encoding theoretically attainable. To transmit information at such high rates, one needs an information "pipeline" with a greater "diameter" than anything so far used for television. And it is more than just a technical problem: "pipelines" take up "real estate"; the airwaves are not unlimited and other users have their rights too.

This extraordinary bandwidth problem is the primary reason that no one at present is even *considering* broadcasting this data stream, but only using it between the analog camera and the analog transmitter. (Members of the European Broadcasting Union are working with the idea of transmitting digitized television pictures from satellites to ground stations but only *in a tight beam* and only with six-bit resolution.) Still, intense efforts are being made to circumvent — to any degree possible and by any means possible — the requirement for such a large bandwidth.

In the face of such difficulties, how can communications engineers be sanguine about the future of digital television systems? They can because the technology is evolving. More efficient encoding techniques are being developed continually. The use of lasers and light guides can, in principle, permit the transmission of thousands of television programs simultaneously over a single fiber. Integrated circuit (IC) technology is advancing at a fast pace, and the cost of IC units is coming down.

Finally, the motivation to develop digital television certainly exists — and this does not just mean to get a clearer picture on the screen of the viewer at home. The biggest motivating factor may be to get digitized signal processing equipment into studios. With such equipment, special effects, image enhancement and many other functions could be handled with great ease; it is likely that many could even be automated.

It is certainly going to be interesting to watch the development of digital television over the next decade or so. It is hoped that this paper will aid the reader in understanding some of the developments as they occur.

Bibliography

1. J. R. Pierce, *Symbols, Signals and Noise,* Harper & Row, New York, 1961.
2. Henri Busignies, "Communications channels," *Scientific American, 227:* 99–113, Sept. 1972.
3. W. M. Goodall, "Television by pulse code modulation," *Bell Syst. Tech. Jour., 30:* 33–49, Jan. 1951.
4. A. A. Goldberg, "PCM-encoded NTSC color television subjective tests," *Jour. SMPTE, 82:* 649–654, Aug. 1973.
5. Woodrow W. Everett, Jr., *Topics in Intersystem Electromagnetic Compatibility,* Holt, Rinehart and Winston, New York, 1972.
6. Clyde F. Coombs, Jr., Ed., *Basic Electronic Instrument Handbook,* McGraw-Hill Book Co., New York, 1972.
7. Ray Ryan, *Basic Digital Electronics,* Tab Books, Blue Ridge Summit, Pa., 1975.
8. Albert Rose, *Vision — Human and Electronics,* Plenum Press, New York, 1973.
9. Pierre Mertz, "Long-haul television signal transmission," *Jour. SMPTE, 75:* 850–855, Sept. 1966.
10. C. W. B. Reis, "Tutorial — Anyone for digits?" *The BKSTS Jour. 57:* 110–115, April 1975.