

chapter eight | new moving image technologies

“Digital” has become a trendy buzzword, redolent of the computer – the icon of our age. Calling anything digital implies praise and precision, even though the meaning of the term is rarely understood.¹

‘Everything in this place is either too old or it doesn’t work ... and you’re both!’ – Will Hay in *Oh, Mr. Porter!* (1937, dir. Marcel Varne)

The ‘D word’ has now become an inescapable element of moving image technology. Since the early 1990s, digital videotape has been replacing its analogue predecessors in broadcast studios. The last mainstream professional analogue format, Betacam SP, was declared obsolete in 2002 when Sony announced that it would no longer manufacture VTRs in this format and would phase out spare parts availability over a seven-year timespan. In the domestic arena the Digital Versatile Disc (DVD)² has all but replaced VHS for the sale and rental of prerecorded media in the developed world. The digital stills camera has become one of the fastest selling consumer electronics products in the first decade of the twenty-first century, leading to a rapid decline in film-based photography by amateurs. The use of digital imaging in film production began primarily with the integration of digitally generated special effects – a technique known as computer generated imagery (CGI). This was marked spectacularly in 1992 by the release of *Terminator 2: Judgement Day* (dir. James Cameron) and *Jurassic Park* (dir. Steven Spielberg).

By the end of the decade low-budget feature films were being originated and edited entirely on digital video before being transferred to 35mm film for release. By the time of writing computer-based technologies for cinema projection were becoming available, though specific economic factors exist which will prevent their widespread use as an alternative to 35mm for some years to come. Moving image archivists have been using digital processes for restoring audio recordings since the early 1990s. From around the turn of the century the image manipulation technology which was previously only available in the CGI departments of lavishly funded Hollywood studios came down in price to the point at which archivists could

start using it to scan films and the correct defects such as scratches and colour dye fading. Long-term storage of media content in digital form is proving to be a major problem, of which more below.

So, what does the ‘D word’ actually mean, why has this technology come – seemingly from nowhere – to establish itself so widely and what are its implications for the long-term future of recorded moving images?

Computers and moving images

With the exception of digital audio, which has been covered in chapters four and five to the extent that it is necessary to contextualise the role of related technologies, all the other technologies discussed in this book record and reproduce moving images by using what is termed *analogue* (from the Greek *analogos*, meaning equivalent or proportionate) methods. The terms ‘analogue’ and ‘digital’ can to all intents and purposes be treated as antonyms. An easy way to imagine how analogue recording works is to think of it as an analogy: images or sounds are represented, or analogised, as continuously variable physical quantities which can be read and written by a machine which converts them to and from visible images and audible sounds. These variable physical quantities can take a number of different forms. For example, the pattern of silver salts on photographic film is an analogue of the pattern of light which existed at the time and in the place a photograph was taken; or the pattern of magnetised iron oxide on an audio tape is an analogue of the changes in air pressure in a given space at a given time, as they were detected by a microphone. In the case of still photographic images, the only machine needed to read this recording is the human eye. In others, notably videotape, some very sophisticated mechanical and electronic engineering is needed to produce the hardware which can reproduce the recording as meaningful images and sounds.

Digital media does not record a direct representation of a continuous process of change. Instead, it represents that process as information, or *data* (from the Latin ‘datum’, meaning ‘something given’). That information takes the form of numbers, hence the word ‘digital’. Computers are used to encode images and sounds into digital data, and then reproduce them by converting the data back into an analogue signal which can be displayed on a screen or played through a loudspeaker. Doing this may at first sight seem illogical because it would appear to introduce an extra, unnecessary process. But representing recorded images and sounds digitally has one crucial advantage from which several others flow: digital data can be copied with 100 per cent accuracy, whereas an analogue recording cannot.

Copying an analogue recording of any description involves losing some of its accuracy, or quality. Even if you duplicate a film by contact printing it, the individual grains of dye in the source element could never be placed in precise alignment with their counterparts on the unexposed duplicate stock. Therefore, a small amount of distortion will be introduced into the copy. If a film is printed optically, even more distortion is introduced, because the light which is shone through the source element will be refracted by a lens before reaching the destination stock. To copy an

analogue magnetic audiotape it has to be played back using one machine, and the electrical signal which results fed to the recording head in another. The amplification and signal processing electronics between the two tapes will inevitably degrade the signal by introducing unwanted modulation (noise) during the copying process.

In the digital domain none of this applies. Because the image or sound recording is represented as a series of numbers, it is only necessary to copy those numbers (the data) accurately – something which can be checked by comparing the source and destination elements after the copying is complete – in order to produce a ‘clone’ or perfect reproduction of the original, totally free of any imperfections. For an industry which relies almost entirely on the underpinning technology of being able to mass-produce copies of an original recording, this is obviously a significant advantage. The introduction of visual effects (not just obvious and spectacular ones such as dinosaurs, but routine effects such as fades to black, which are used repeatedly in virtually every feature film) previously worked by applying the effect optically during the copying process. So, for example, a scene could be filmed in which an actor performs against a plain background. This could be superimposed with footage taken at a given location in duplication to a new element which combines the two images. But as the combined image is a generation removed from either original, it is of lower quality. Working digitally, these images could be manipulated in a computer and the result output to film again without any loss of resolution, contrast or colour depth. This is just one example of the actual and potential advantages of using digital media rather than analogue.

A computer is essentially a very large, very powerful calculator. It stores and performs calculations on data electronically. This is achieved by exploiting the fact that an electrical circuit has two states – open or closed. For example, the light switch in your living room is either on or off. We can then give these states a number – 0 for off, 1 for on. Each one of these states is the basic, smallest unit of digital data possible, known as a *bit*. It is then possible to combine sequences of bits to form larger numbers. So for example, sequence of two bits would offer four possible states – 00 (1), 01 (2), 10 (3) and 11 (4). The grouping of bits used by almost all modern computers as the working unit of data is a sequence of 8 (256 possible states), known as a *byte*. Possibly the commonest data units represented as bytes in a computer are letters of the alphabet – for example, a capital ‘A’ is 01000001 (which is 65 if expressed as a decimal integer). A computer stores and performs calculations on bytes by containing literally billions of tiny electronic circuits which its processor can open or close as required, thereby representing the data. Given that the byte is so small as to be meaningless as an expression of ‘real life’ data storage, the memory (data capacity) of a computer is usually expressed in multiples: a *kilobyte* consists of 1,024 bytes (2^{10}) and is referred to by the abbreviation ‘k’ or ‘kb’; a *megabyte* is 1,048,576 bytes, or 1,024k (2^{20}), with the abbreviation of ‘mb’; a *gigabyte* is 1,024mb (2^{30}) with the abbreviation of ‘gb’; and a *terabyte* is 1,024gb (2^{40}) and is so big that it is not used frequently enough for a common abbreviation to have become established. Moving image data usually falls into the megabyte or gigabyte category – for example, a feature film on DVD will typically occupy between 4 and 8gb.

The first machines which could store data and carry out mathematical calculations on it were mechanical. Arguably the earliest design of any significance was that of the ‘Difference Engine’, produced by the British mathematician Charles Babbage in 1822. Although it was too mechanically complex to actually be built at the time, one was built by the Science Museum in London in 1991 from Babbage’s original plans, and the three-ton computer was found to work exactly as he had predicted. The Difference Engine stored data and programs (a set of instructions determining what calculations a computer carries out and in what order)³ on punched cards and performed calculations using clockwork mechanisms. Though its processing power was infinitesimal compared to anything electronic, it is estimated that Babbage’s computer could execute simple calculations three to five times faster than the human brain.⁴

Mechanical devices for simple and repetitive data processing functions became established on an industrial scale in the latter half of the nineteenth century. A notable pioneer in this field was the American engineer Herman Hollerith (1860–1929), who, from 1879 onwards, developed machines for counting and sorting punched cards electromechanically, and counting each instance of a data item encoded on them. Because the only data processing carried out by these machines was to accumulate a record of each time a given pattern of holes passed through a mechanism – they did not perform any actual calculations – they could not be termed computers in the strict sense of the word. But they established the precedent for automated data processing on an industrial scale, one which would eventually provide the groundwork for the computers which would be capable of dealing with the volumes of data needed to represent images and sounds. The company which would eventually become International Business Machines (IBM) was founded on the back of Hollerith’s inventions in 1911.

The next significant step was the emergence of machines that could perform full-scale data processing operations (i.e. ones based on carrying out mathematical functions) electronically, i.e. without the need for mechanical moving parts. The technology used to achieve this was the same one with which Lee de Forest was able to electronically amplify audio signals: the thermionic valve. They acted as the basic unit of storage for each ‘byte’, and because they stored data in the form of electrical energy rather than through a mechanical, solid-state process, enabled calculations to be performed far more quickly than was the case previously. The first generation of valve-driven computers, which were developed in the 1940s for code breaking and other military applications during World War Two, could hardly be described as ‘micro’, though: one such example, the American ENIAC (Electronic Numerical Integrator and Calculator), built by the US Army to calculate the trajectory of artillery shells, weighed 25 tons and consumed over 200 kilowatts of power (enough to light a small town). Its processing power was roughly equivalent to that of the first hand-held calculators, sold in the 1970s.

The story of computing from that point on was one of decreasing physical size and increasing processing power. The transistor, invented in 1947, did the same job as the valve but was about a hundredth of the size and consumed an insignificant

amount of power. The 'integrated circuit', more commonly known as the 'silicon chip' or 'microchip', was first produced in 1958, and in 1971 the first microprocessor – a device which combines most of a computer's processing functions onto a single chip – went on sale. Within three decades the processing power available from 25 tons of power-guzzling valves had been replaced by a miniature system of circuits little bigger than a postage stamp. Furthermore, the speed and complexity of the calculations performed by microchips has been rapidly increasing ever since. The term 'Moore's Law' is commonly used to describe this phenomenon, which refers to a prediction made by Gordon Moore, the founder of the world's largest microchip manufacturer, Intel. In an essay written in 1965, Moore stated his belief that the number of transistors on an integrated circuit (and therefore its processing power) could be doubled, approximately every eighteen months (an estimate he later revised to two years).⁵ So far, this prediction has proven to be roughly accurate.

The product with which the microprocessor has become primarily associated is the personal computer (PC), which had its origins in the late 1970s and began to be sold on a significant scale from the early 1980s. It was 'personal' in the sense that the processor, memory, offline storage (such as floppy and hard disc drives) and user interfaces (i.e. keyboard, screen and later, mouse) were all combined into a single unit, without the processing power being shared between a number of users. It should be borne in mind, however, that these are not the only devices to contain microprocessors, which are now an integral part of practically every item of hardware used to record and reproduce moving images.

In crude terms, it needed roughly twenty years of Moore's Law taking effect before the processing power of an average microprocessor was sufficient to handle audio encoded as digital data with a quality equivalent to that of an LP record or an FM radio transmission, and then a further fifteen until it was able to manipulate digital video at a resolution which is comparable to a broadcast PAL or NTSC transmission shown on a consumer television set. The following quote is taken from a British national newspaper article which appeared in 1994, describing the use of 'multimedia' computer-based video images in teaching:

The picture is grainy, jerky and usually runs in a tiny window in the middle of your computer screen, to minimise the amount of memory and disc space needed to store and display the moving images.⁶

Within a decade, video manipulation and editing would be possible using home computers of average power and memory capacity. In the rest of this chapter we shall consider some of the ways in which computing has shaped the evolution of moving image technologies at the start of the twenty-first century.

Sound

Audio was the first element of moving image technology to be computerised, with location and studio sound for film and video being recorded digitally by the mid-

1980s in most of the developed world. Put crudely, this is because the volume of data needed to represent an audio waveform and the complexity of sums involved in converting it to and from sound and data are significantly lower than with video or film – so much lower, in fact, that 10–15 years of Moore's Law separates the systematic use of computer technology for audio and moving images.

At some stage, all media content which is encoded as digital data has to undergo what is called *analogue to digital conversion*. This is the process whereby the representations of light and air pressure which are detected by the eye and the ear are turned into the ones and zeros which can be stored and processed by a computer. In the case of audio, the underlying technique for achieving this is known as *sampling*. As will be recalled from chapter one, a 'moving' image consists of multiple still photographs, exposed and projected in rapid succession. To computerise this impression of movement, it is only necessary to create digital versions of each still frame, and then to determine the speed and duration at which they will be displayed. The passage of time, therefore, is neatly divided into a series of discrete segments. The same does not hold true for a recorded soundtrack. In the analogue domain, this consists of a single pattern of modulation which varies infinitely across the surface area of the film, disc or tape on which it is recorded. By placing the medium in contact with the playback device at the same speed as the one which recorded it, the passage of time can be made identical in both recording or playback.

With digital audio it is not as simple as that. Numerical data is not infinitely variable – it is a set volume of ones and zeros. A zero does not gradually become a one over a millimetre or so of magnetic tape: that tape will encode a zero followed by a one, in just the same way that our light switch is either on or off. Therefore a way is needed of determining how that data can be made to represent changes in air pressure (which is detected by the human ear as sound) over time. This is where sampling comes in. The technique has its origins in an influential research paper written by Harold Nyquist, a Swedish physicist who emigrated to the US in 1928 and ended up as a researcher for the Bell Telephone Laboratories.⁷ As with Lee de Forest and electronic amplification, Nyquist was primarily interested in the potential application of digital audio in long-distance telephone links. He determined that the passage of time had to be carved up into a series of discrete segments (comparable to frames in a film or each interlaced scan in a cathode ray tube). The *sampling rate* determines how long the audio information in each 'sample' is reproduced for, or, in other words, how many samples are recorded and reproduced within a set period of time. Within each segment, it is necessary to record data which represents the frequencies of the sounds to be reproduced. Because most audio recordings consist of many different frequencies being audible at once, the sampling rate directly affects the frequency range which can be recorded and reproduced, and consequently the perceived quality of the audio to the human ear. Basically, the higher the sampling rate, the better the sound. However, not only does the sound quality increase with the sampling rate; so does the memory, storage space and computer processing power needed to record and reproduce it digitally.

At the time of Nyquist's research, the computing power did not exist to put any of his ideas into practice. Even the valve-driven behemoths such as the ENIAC were still over a decade away. But, as with Babbage and the Difference Engine, further experiments during the mid-1950s vindicated his theories, and the first commercial applications for digital audio were indeed in telecommunications. But because the memory and processing requirements increased in proportion to the perceived quality of the recorded audio, digital sound for radio, recorded music, film and television did not become feasible until microprocessors were available in the mid-1970s. The key developments in digital audio during the 1970s and 1980s are covered in chapter four, though it is worth noting here that sound capture and editing using personal computers did not become a reality until the mid-1990s.

While the studio recorders which used modified UMatic and Beta videocassette mechanisms and the CD audio players which were sold to consumers since 1982 did contain microprocessors, the speed at which they needed to carry out analogue to digital conversion and vice-versa was relatively low. This was because the first generation of digital audio recording technology used an uncompressed signal. *Compression* is an essential technique for digitally encoding many types of video and audio. In essence, it reduces the volume of data needed to encode a given amount of media content to a given quality, and does so in one of two ways. *Lossy* compression works by discarding (or 'losing') data relating to parts of the audio spectrum which it is believed most human ears cannot detect (extremely high or low frequencies, mainly), leaving only the audible parts left. *Lossless* compression, on the other hand, preserves the entire encoded spectrum, but uses complex mathematical equations, known as algorithms, in the software that carries out the encoding in order to compress it into a smaller data storage space. The main benefit of compression is obvious: you can store more sound using less data. The key drawback is that more processing power is needed in the analogue-digital-analogue conversion process than with an uncompressed data stream. Furthermore, if heavy lossy compression is used, the sound quality will suffer. For example, many of the compression protocols now in use would enable Beethoven's ninth symphony to fit on a 1.44mb floppy disc, if you were prepared to listen to it with the sound quality of a telephone handset. In contrast, an uncompressed recording of it on a compact disc (which were designed for a playing time of 74 minutes, it is rumoured because Beethoven's ninth symphony was the Sony founder Akio Morita's favourite piece of music), in stereo, and with a 16-bit sampling frequency of 44,100 per second, would occupy close to 650mb.

In the first generation of digital audio technology no compression was used, the reason being that on tape-based systems and CDs, storage space was not at a premium and, unlike with consumer digital video a decade later, there was no economic barrier to providing sufficient processing power to decode 'raw' digital audio without the helping hand of asymmetric compression. Furthermore, the electronics industry was anxious to prove that digital audio could match the quality of the most expensive analogue systems then in use. This was one of the reasons why the compact disc was designed for a storage capacity of 650mb – almost 500

times the capacity of any removable data storage medium in widespread use at the time of its launch – despite the formidable engineering challenge that this imposed.

The convergence of digital audio and the personal computer took longer. Moore's Law gradually boosted PC processing power throughout the 1980s, but storage space remained at a premium. The data storage methods used by dedicated digital audio hardware in the early 1980s imposed limitations which made them incompatible with PCs. Either they used magnetic tape, access to which was strictly linear (i.e. it was impossible to access any part of the data stream instantaneously or at random), or the optical compact disc, which until the late 1990s was a 'read only' medium which could only be mass-duplicated in a factory.

The key device which enables a personal computer to store and retrieve large amounts of data quickly and at random is the hard disc. The first ones (sometimes known as Winchester discs after a type of rifle, because their inventor, Ken Haughten, was a keen amateur hunter) were sold by IBM in 1973, and consisted of a sealed unit containing one or more rigid steel platters coated with magnetic oxide, and mounted on a rotating spindle. The reading and writing head was on a pivoting assembly suspended above the rotating platter, which was divided into 'tracks' and 'sectors' on which blocks of data were stored. Being a magnetic medium the data could be erased and rewritten without practical limit, and being a disc, random access was possible. The problem was that until the early 1990s, hard discs were very expensive, limited in capacity and not very reliable. 'Head crashes', in which the mechanism which positioned the head above its required track and sector failed, causing it to make physical contact with the disc surface and thereby destroying it, were common occurrences (and even with today's hard discs, they are not unheard of). By the late 1980s, a top-of-the-range PC costing four figures might contain a hard disc with a capacity of 20mb; in 1995, 500mb was a typical figure – still less than the data capacity of a single CD.

The ability to work with uncompressed audio data on a personal computer did not become firmly established until the late 1990s (in another illustration of the exponential effect of Moore's Law, this was only around 3–5 years before the same could be said of digital video). Three factors had the greatest influence here: the ongoing effect of Moore on processor power, steady but significant increases in typical hard disc capacities and the emergence of the recordable compact disc. Soon analogue audio recording and post-production had almost totally disappeared from the professional film and broadcast industries.

Television and video

A video signal is digitised by dividing the frame into individual dots, or 'pixels'. The number of pixels per frame determines the *resolution* (the amount of detail in the image, and consequently the volume of data needed to represent each frame) of the image, and this varies according to the format in use and the television system being encoded. For example, the system used for encoding the video stream on a conventional DVD uses a resolution of 720 (horizontal) x 576 (vertical) pixels for PAL

(625 lines of analogue horizontal resolution), or 720 x 480 for NTSC (525 lines). Data recording the luminance and chrominance of each pixel is stored on a predetermined scale, the detail of which varies in a similar way to the sampling rate in digital audio. For display on a conventional television the digital to analogue converter in a VTR will convert this grid of pixels into conventional, 525- or 625-line signal, at the appropriate scanning rate and interlaced if need be.

Once microprocessors which were fast enough to carry out these functions became available, it made obvious sense for videotape recording to go digital. One of the reasons that videotape had superseded 16mm film for the majority of origination in the television industry during the 1970s and 1980s was the lower cost and greater flexibility in editing. With film it was often necessary to make a separate print (the 'cutting copy'), cut the camera negative to match it only when all the editing decisions were finalised, in order to minimise damage to the original, and only then strike the transmission print. This was a very time consuming and expensive process. In circumstances when it was logistically impossible (such as editing news footage), editing the original camera element directly was necessary and this often caused substantial, visible damage (such as scratches and dirt) in transmission. Archivists have had to expend considerable time and energy in restoring television material from the 1960s, 1970s and 1980s which has been mutilated in this way.⁸ With videotape, the way in which content is edited is to selectively copy footage from several original camera source tapes, assembled onto a single edited master. Sometimes, for example, if special effects or complex titles, visual effects or sound dubbing is required, this can mean going through several generations of tape, thereby lowering the picture quality through generational signal loss – if the video format is analogue. Digital video offered the potential to eliminate this problem, thereby enabling the final edited master of each production to 'clone' the quality of the original material.

This was the next major area of moving image technology to go digital after audio, and did so following a similar pattern: professional studio use first, then digital television and video marketed directly to consumers. In the broadcasting industry of the 1980s, the analogue Betacam SP format superseded U-Matic and retained a dominant market share throughout the decade. The first commercially marketed digital videotape format for television production, the Ampex 'D2' system, was launched in 1986 but failed to achieve significant sales. The D2 VTRs only had analogue inputs and outputs: all the analogue-digital-analogue conversion took place inside the machine. This negated the principal advantage of going digital, as the process of decoding and then re-encoding the signal when copying between two machines for editing introduced the same generational loss as with analogue systems. Although it was uncompressed, the D2 format only encoded a composite video signal in order to minimise the volume of data and processing power needed.

The format which signalled the systematic introduction of digital video into the broadcast industry was Sony's Digital Betacam, launched in 1993. On a pragmatic level, it used the same design of cassette and tape transport mechanism as Betacam SP, thereby meaning that the new VTRs would be 'backwards compatible',

i.e. they would also play the old analogue tapes, of which many broadcasters possessed libraries of tens of thousands. Like its analogue predecessor, 'Digibeta' (as it soon came to be known) encoded a *component* signal, which split the video information up into three signals. This is a further refinement of the NTSC principle of broadcasting luminance and chrominance information separately, in which the luminance is transmitted as before but the colour information is divided into two extra signals. Thanks to being a component system, Betacam SP had acquired a reputation for being comparatively resistant to generational signal loss even in the analogue version. Although the digital version incorporated a 2:1 compression ratio in order to keep the physical tapes compatible, it quickly became established as the gold standard in broadcast video technology.

The demonstrable image quality of Digibeta was, however, reflected in the price of VTRs and media which, throughout the format's life cycle, have remained at levels where it is only really available to major broadcasters and production companies. It was therefore inevitable that a group of other tape-based formats would emerge to fill the gap left by what had become 'semi professional' analogue video formats, notably U-Matic and S-VHS. Perhaps the most widespread were Sony's DVCAM (launched 1995), which used a cassette design similar to the DAT audiotape, and Panasonic's DVC Pro (1996). These have much higher compression ratios than Digibeta – typically between 5:1 and 10:1, depending on the specific format, and therefore do not match its image quality or durability. But the far lower price of hardware and blank media marked the beginning of digital videotape becoming available to a wider user base, with the new generation of formats catering to a range of markets, ranging from the amateur 'home movie' makers who would have used Super 8 a generation earlier to producers of promotional and training videos.

A process of 'going digital' was, therefore, well underway by the mid-1990s. But none of this would have been apparent to viewers in the home at this point, who were still watching their 525-line NTSC or 625-line PAL analogue CRT television sets, which received an analogue composite signal either from the airwaves or a VHS VCR. The image compression technology available thus far had enabled professional and semi-professional digital video to become a reality, but had not brought prices down or reliability up enough for it to be mass-marketed to consumers. The technology which enabled it to make that jump has the somewhat uninspiring title of MPEG-2.

MPEG stands for Motion Picture Experts' Coding Group, a committee of computer scientists founded by Leonardo Chiariglione under the umbrella of the International Standards Organisation in January 1988. Its aim was to establish software standards for encoding and decoding digital video which could be used in a range of consumer and industrial applications, from the small displays in hand-held videogames to digital cinema projection on full-sized screens.⁹ In the following years it produced five main sets of standards (MPEG-1, -2, -3, -4 and -7), of which MPEG-2 has become the method of choice for the two large-scale consumer digital video applications: digital terrestrial television transmission, and the DVD. MPEG-2 proved especially suitable for this role because of two unique aspects of the lossy compression technique it uses. Firstly, it is what is termed an 'asymmetrical' system:

In audio and video compression, where the encoder is more complex than the decoder, the system is said to be asymmetrical. MPEG works in this way. The encoder needs to be algorithmic or adaptive whereas the decoder is 'dumb' and carries out fixed actions. This is advantageous in applications such as broadcasting where the number of expensive and complex encoders is small, but the number of simple and inexpensive decoders is large.¹⁰

From this explanation we can deduce that the key to putting new technology in the hands of the consumer is that it has to be as cheap and simple as possible. The computing power needed to decode the video data on a Digibeta tape, for example, could never be produced at a price which consumers would be able or willing to pay. MPEG-2 solves this problem by not only encoding the moving images themselves, but also small amounts of instructional code which tells the player's decoder how to interpret the image data. To a certain extent, it could be argued that some of the software is built into the media.

The broadcast standard tape formats which preceded it, by contrast, simply contained the image data in a predetermined format, which the VTR had to decode using entirely its own software. Compared to the 'dumb' decompression of MPEG this has a tendency to sap processing power, especially if the compression in use is lossless (because in this case it has to extract the uncompressed data first, before decoding it into an analogue picture). To understand how this technique works let us consider a hypothetical example. A camera photographs a static landscape shot, half of which depicts clear blue sky. A broadcast standard tape format would probably encode all the different blue pixels in the frame individually, even if the difference in luminance and chrominance is so slight as to be invisible to the naked eye. Even if the next shot is a panning close-up on fast-moving action, each frame would generate the same volume of data as the landscape shot which preceded it, and the decoding process would require the same computing power. This is because it would decode and convert to analogue every pixel individually, regardless of the relative similarity or difference between their respective luminance and chrominance properties.

MPEG-2 goes about it a different way. In our landscape shot, instead of the data stream individually representing tens of thousands of blue pixels, it might contain just one, followed by an instruction which says 'repeat this for the next X pixels'. In this way thousands of bits of data storage can be dispensed with instantly, and the microprocessor doing the decoding only has to perform one operation rather than thousands. In the bottom half of the frame, where more detail is needed, more of the pixel characteristics can be defined individually. This is, of course, what defines MPEG-2 as being a 'lossy' compression system – in this example we are 'losing' the detail from the sky, even if the varying shades of blue are so slight that most people would not notice them. MPEG-2 carries this principle even further, by analysing the relationship between whole frames during the encoding process. The video stream is divided into a sequence of 'intra' (I) and 'predicted' (P) frames (the exact composition of the sequence is flexible and can be determined when encoding). If, for example, a sudden flash of lightening were to appear against the sky in our hypothetical landscape shot, then instead of encoding two entire frames, the second

could be substituted for an instruction which says 'repeat frame A as frame B, but in frame B shade these pixels as so where the lightening bolt appears'. Once again, this reduces both the volume of data needed to represent the video stream, and the number-crunching required to decode and display the picture.

Both of these abilities were essential requirements for the first consumer medium which utilised MPEG-2, the DVD. News of the imminent launch of the format came in 1995:

'Sony of Japan and Philips of the Netherlands plan to mount an assault on Hollywood to persuade US movie studios to make films and other entertainment products on their new digital video disc system. Toshiba and Matsushita have an alternative, non-compatible system. Although both camps have tried rapprochement, they seem on course for a format war in the digital disc market just as over the first generation of video cassettes when Sony struggled to establish its Betamax format over Matsushita's ultimately successful VHS system.'¹¹

Despite its infamous precedent, the anticipated format war did not materialise, and the first pre-recorded Hollywood films on DVD went on sale in Japan in the winter of 1996, and in March 1997 in the United States. Unlike the case of CD audio, processing power was at a premium this time, and so was data capacity. Not using any compression at all simply was not an option – uncompressed NTSC video, for example, would gobble up 26mb *per frame*! By greatly increasing the writing density of a DVD (which, by taking advantage of a decade's development in electromechanical technology, was made possible by using a lower wavelength laser) and sandwiching two data storage layers, each of 4.7gb, onto each disc which could be read by adjusting the frequency of the laser pickup head, it was possible to squeeze 8.5gb onto a carrier with the same physical dimensions as a 1982 CD. To encode a two-hour feature film into this capacity and with a picture quality that would look acceptable on an average domestic television set, a compression ratio in the order of 20:1 was needed. It was to achieve this that the asymmetric design and dynamic, lossy compression features of MPEG-2 were exploited. Due mainly to the economies of scale gained by the fact that the disc transport mechanism in many consumer players was a slightly modified version of those already in mass production for audio and data CD drives, market saturation happened very quickly. Sales in the US almost tripled in the three years to 2003, and at the time of writing this format has effectively superseded VHS as the principal format for sales of pre-recorded media.¹²

Digital terrestrial television has been the other main consumer application for MPEG-2. Since digitisation became a commercial reality in videotape recording during the late 1980s, engineers started looking for ways of transmitting the pictures digitally, replacing the VHF and UHF analogue radio signal used to carry television images since that method was developed by EMI and RCA in the 1930s. The reason is simple. Only a very small part of the total electromagnetic spectrum (all possible wavelengths of electromagnetic energy)¹³ can be used for the wireless transmission

of electromagnetic energy which can be made to represent images and sounds, known as radio waves. Within the radio waves spectrum, the number of wavelengths which can be used to carry individual signals is finite, and each analogue PAL or NTSC television broadcast consumes a disproportionately large amount of this total 'bandwidth'. As John Watkinson explains:

There is only one electromagnetic spectrum, and pressure from other services such as cellular telephones makes efficient use of bandwidth mandatory. Analogue television broadcasting is an old technology and makes very inefficient use of bandwidth. Its replacement by a compressed digital transmission will be inevitable, for the practical reason that the bandwidth is needed elsewhere.¹⁴

As with the scheduled broadcasts in the two years before the outbreak of World War Two, Britain has led the way with digital television broadcasting. The first steps came in the early 1980s when the BBC's engineers developed the NICAM (Near Instantaneously Compacted and Expanded Audio Multiplex) protocol. This was a means of broadcasting two channels of digital audio alongside the existing PAL analogue video transmission. Its first public broadcast was of a classical music concert in July 1986, and it has continued to be used with all terrestrial analogue broadcasts in the UK ever since (at the time of writing practically all televisions and consumer VCRs sold in the UK include a NICAM audio receiver). Although NICAM has also been adopted in some Scandinavian countries, the system was not modified for NTSC use or used on any significant scale outside Europe, where a variety of analogue stereo audio methods for television use emerged during the late 1970s and early 1980s.

At this point MPEG-2 was still some years away, but the issue of digital television was now firmly on the agenda. In September 1993 the Digital Video Broadcasting (DVB) group was formed as a consortium consisting of broadcasters and electronics manufacturers, principally in Europe. Over the following five years it developed and tested three transmission standards, all of them based on MPEG-2: DVB-T (terrestrial), DVB-S (satellite) and DVB-C (cable). In addition to the MPEG-2 video and audio streams, DVB also included data transmission standards for related services, such as the so-called 'interactive' content which can be broadcast alongside each television channel, e.g. subtitles, alternative audio tracks or additional textual information about the programme being broadcast.

After two years of experiments and development of infrastructure, the first regular DVB service began in the UK on 15 November 1998. Sweden followed suit in April 1999, and over the next two years most of Western Europe and Australia did likewise. At the time of writing, the US still uses NTSC analogue for terrestrial broadcasting, though the French owned cable operator Canal+ had offered DVB-C transmissions since 1997. Thanks to the very low bandwidth needed to broadcast the MPEG-2 programme stream it is possible to transmit tens of digital channels alongside the existing analogue ones using the DVB-T system – between thirty and forty in the case of Britain, depending on the transmitter.

This of course leads into the question as to when the original purpose of going digital will be realised, by switching off the analogue signal. History shows us that the greater the market saturation of a consumer technology, the greater the resistance is to the onset of its obsolescence. Videotape formats tend to be superseded comparatively quickly (every decade or so, roughly speaking) in the broadcast and professional sectors because these users derive direct economic benefits from investing in the new technology – for example in the greater flexibility by not having to worry about generational loss when editing video content digitally. The same is not true of media technology sold directly to consumers, who tend to be reluctant to invest unless there is either a demonstrable functionality benefit or they are given no choice. When the BBC announced its competition between the Baird and EMI transmission systems in 1935, the Baird 30-line standard which the winner was intended to supersede could be killed off quickly and uncontroversially, because only a tiny number of 30-line receivers had been sold. The Goldmark/CBS colour system was abandoned after small-scale experimental broadcasts for the same reason.

With consumer technology, however, the situation is very different. The 405-line to PAL conversion in Britain, which began in July 1967, offers a typical example of how technological change tends to happen. At first, the majority of the population were not willing to invest in a new television – the slightly sharper picture simply was not worth what the new sets cost (and the price of colour was stratospheric). Furthermore, 'set-top boxes' were sold to convert the new BBC2 channel, which was only broadcast in PAL, to 405-line so that it could be received on existing sets. As a result the two systems had to be broadcast side-by-side until January 1985, and therefore the conversion took over 17 years. Another example is the commercial lifetime of VHS, which has lasted for almost thirty years and it is still the principal consumer format for off-air recording. Despite JVC's attempt to 'do a Betacam' by introducing a digital version of VHS which was backwards compatible with analogue tapes in 1995, the hardware has hardly sold at all, except to a few enthusiasts.

The set-top 'digibox', which includes a digital to analogue converter and connects by wire to an existing PAL set, has also been the primary method through which DVB has been introduced, and, as with the DVD, the rate of market penetration has certainly been significantly higher than with any comparable analogue technology. The UK regulatory body for public broadcasting, Ofcom, estimated that as of 30 June 2004, 55 per cent of households had access to digital broadcasting, either in terrestrial, satellite or cable form.¹⁵ Given that in most Western countries, television broadcasting also has a public service dimension (it has even been argued that having access to television should be considered a basic human right), any change in the broadcast standard which will render expensive capital items in the homes of low-earners obsolete inevitably becomes as much of a political issue as a technological one, as has the regulation of public broadcasting throughout its history. A report published by the UK government in October 2004 concluded that 'digital television [has to] be near-universally available, generally affordable and have been taken up by the majority of consumers before the analogue signal can be switched off'.¹⁶ Its authors set a target of 95 per cent of the population to be receiving digital television

'on their main television set' before the analogue plug is finally pulled¹⁷ and hinted at the potential political fallout that would result from imposing 'the penalty costs of forced conversion'.¹⁸ It is likely, therefore, that analogue television broadcasting will still be with us for some time to come, and furthermore that another potential benefit of DVB will take a long time to roll out. At the time of writing, only the Faroe Islands and an experimental zone in Berlin consisting of one or two suburbs had switched off analogue television. It should also be remembered that while the 'digibox' remains the receiving technology of choice, converting the MPEG-2 stream to an analogue PAL signal and then passing it by wire to a standard television results in a far lower quality of picture than displaying the digital pixels directly on a CRT or TFT, as is the case on the monitor of a personal computer.

Piracy

As has been noted in chapter six, the illegal copying and/or commercial exploitation of media content has been a major concern pretty much ever since moving image technology has been used on a commercial basis. The industry argued that the advent of consumer VCRs in the late 1970s turned this into an endemic problem, i.e. that the extent of piracy was systematic and that it was resulting in significant loss of revenue. The response of copyright owners has traditionally taken either or both of two forms: legal action and technological barriers. The latter have usually worked on the twin principles of making it sufficiently difficult to copy prerecorded media content as to stop everyone from doing it except organised criminals and a small minority of consumers with above-average technical skills, and a preference for distributing content on media which are inherently resistant to unauthorised use (e.g. the VHS format, which suffers massive generational loss when copied).

The advent of digital media raised the stakes significantly. It will be recalled that the key advantage of digitising video and audio is that it can then be 'cloned' – copied without any loss of signal quality. If legitimate users of the technology can do this, then so can pirates. When the first mass-marketed form of digital media, the audio CD, went on the market, the music industry initially did not perceive the move to digital as being a problem. The discs could only be mastered and duplicated in purpose-built factories, and although purchasers could make analogue copies on magnetic tape, they had been doing that with LP records for decades previously and the industry had still survived. The emergence of the recordable CD (CD-R) in the late 1990s changed all that.

Both CDs and DVDs encode data as minute indentations in the surface of a reflective disc. In the factory-pressed CD or DVD, the discs are produced from a 'glass master' etched by a laser and then coated with a transparent, reflective coating on a production line. In 1989 the Japanese electronics manufacturer Taiyo Yuden produced a stand-alone CD recorder that 'burnt' the indentations into a dye which was already present on the pre-manufactured disc. The machine cost \$149,000 and even the blank discs cost three figures. Its main application was to burn a single copy of a completed CD in order to check it for errors before the expense of producing

a glass master was incurred. The price of this technology dropped gradually during the early 1990s, but nowhere near the point at which consumers could make cloned copies of commercially sold audio CDs for less than their cover price. All that changed in 1995, when Philips produced the first computer-based CD writer which sold below the \$1,000 mark. Set-top audio CD recorders soon followed, and today the ability to copy CDs and DVDs is a routine feature of virtually every personal computer sold in the Western world. Furthermore, the blank media typically retail for less than \$1 a piece. Recordable DVD technology was marketed to the public from the launch of the Panasonic DVR-A03 in June 2001, and since then its cost and availability has followed a similar trajectory.

The 'Video CD' (VCD) was the first optical disc video format to be systematically exploited by pirates. It was a very low resolution system, based on the MPEG-1 protocol, which could encode up to 70 minutes of video on a conventional CD. It was used extensively as a consumer format in the Middle East, China and Asia during the late 1990s, but failed to gain a market foothold in the West. This was due mainly to the poor picture quality and the fact that Hollywood refused to release media on Video CD because by that stage, the DVD was already in an advanced stage of development. Due to the unprecedentedly low cost of hardware and blank media, it proved to be the ideal medium for pirates in poorer, developing countries: so much so that it was estimated that in China in 1998, pirate VCDs outsold legitimate copies of feature films by a ratio of 35 to 1.¹⁹

With the digital piracy genie well and truly out of its bottle, the industry had to resort, once again, to its two traditional lines of attack, legal and technical. At the time of writing almost all commercially published DVDs for retail sale incorporate at least two copy protection methods: Macrovision (see chapter six) to prevent unauthorised duplication to VHS, and the 'Content Scrambling System' (CSS) to prevent unauthorised duplication of the digital data. CSS works by exploiting a so-called 'hidden area' of the DVD near the inner rim, which can be written onto on a factory-pressed disc but not a home-recordable one. The hidden area includes an encryption key, which enables the software in a DVD player to decode the encrypted video and audio data. Simply cloning the data from the main area of the disc onto a blank will produce an unplayable copy, because the CSS encryption key will not have been copied along with it. Unfortunately, the designers of this system failed to take account of the fact that if the hidden area can be read by a set-top player, it can also be read by illegally used decryption software running on a personal computer. At the time of writing such software is said to be easily available from the Internet, and will enable the pirate to decrypt the video and audio data before burning the 'en clair' version to a recordable DVD.

Although the same rationale applies to CSS as with other technological attempts to restrict piracy (i.e. that only organised criminals and computer geeks will have the motivation and the technical knowledge to circumvent it), the media industry of the early twenty-first century increasingly felt that it was fighting a losing battle. The bottom line is that if human brains can dream up encryption systems, other human brains can crack them: as one film industry technical specialist recently told me, 'as

soon as we build a ten-foot wall, they build an eleven-foot ladder'. But in the digital domain the stakes are higher, so the laws are getting tougher. This has resulted in the emergence of new legislation designed specifically to combat copyright fraud perpetrated using digital means, both in the US and Europe. In 1998 the United States Government passed the Digital Millennium Copyright Act, while the European Union Copyright Directive of 2001 is currently being implemented by EU member states. Both laws specifically make it an offence to circumvent anti-piracy measures incorporated into commercially-sold computer software or digital audiovisual media, and the latter even makes it an offence to use equipment or software designed for this purpose, even if no copyright is actually infringed by doing so. In 2004 following intense lobbying by Hollywood and the music industry, the US government also went one step further. It proposed to introduce legislation that would have the effect of nullifying the Betamax decision of 1981 (see chapter six), in effect outlawing the possession any equipment which *could* be used for piracy, even if its owners did not use it for this purpose. When civil rights campaigners pointed out that this would turn all VCR owners into criminals, the government backed down – but the fact that it was ever seriously considered in the first place gives a powerful demonstration of just how worried copyright owners had become over the issue of digital piracy.

The Internet

The Information Highway will transform itself, even more than it is at present, into the Information Toll Road.²⁰

This issue of piracy leads directly onto the other key area of computing technology that is used to store and deliver audiovisual media in digital form – the Internet. The Internet is, in effect, a global telecommunications network for computers. The function it performs is no more and no less than providing the ability to transport large quantities of digital data over long distances, including across physical (such as oceans) and international boundaries, in a very short time and at relatively low cost.

The Internet's origins lie in the Cold War, specifically in a project which was ostensibly a communications network for the mainframes used for research in university computing departments in the 1970s, the real purpose of which was a military communications network designed to enable the government to continue functioning in the aftermath of a nuclear attack. This was ARPAnet, named after the US Government's Advanced Research Projects Agency, which was established by the Eisenhower administration in 1957. The network began operation in 1969, and over the next two decades grew to encompass a number of other long-distance computer communications systems which had been established independently. This was achieved through a combination of dedicated data communications lines, existing telephone lines, radio and satellite links.

The process which enabled this infrastructure to grow and transform into the Internet we know now encompasses both political and technological factors. The first key political one was a gradual move from military and governmental control

of the interlinked networks which comprised the system as a whole, to civilian and private sector management. Two key milestones on this path were in 1979, when commercial use of the Internet was permitted for the first time, and in 1995 when the US government effectively privatised the network infrastructure within its own borders. Flowing from these political decisions is the economic structure which finances the Internet. Being a largely autonomous collection of interlinked computers, the user (be that an individual or organisation) pays a flat rate to use the network as a whole rather than a price calculated on distance, as is the case with other mass communications networks such as the postal and telephone services. Users who wish to receive data which is made available through the Internet and users who wish to make their data available to others buy access from an Internet Service Provider (ISP), which acts as a gateway to the wider network. In the case of the former, the charge is usually levied by the 'online' time a consumer's computer is connected to the ISP's. In the case of the latter, the data provider's computer, known as a 'server', is permanently connected to the ISP's, who generally charges by bandwidth (i.e. the volume of data flowing in or out). The ISPs in turn purchase access to the telecommunications 'backbones' which make up the global Internet and amortise that cost in what they pass on to their customers. The important thing to note about this economic model is that charges for data transfer are *not* levied according to where that data has come from or is being sent to, as with the telephone and postal services. To the end user, therefore, it is no more expensive to communicate with the other side of the world than the next street. Given that legal attempts to combat piracy depend on legislatures that are, by and large, incapable of working across international boundaries, an infrastructure which enables audiovisual media not only to be copied but digitally 'cloned' across international boundaries quickly and easily is obviously a big concern for rights holders.²¹

The other factors which grew the Internet into a medium of mass communication were technological. To start with there was the effect of three decades of Moore's Law, which in this case did not only apply to the processing power of individual computers, but also to the speed at which data could be reliably modulated into an analogue signal for transmission along a conventional telephone line. This meant that by the mid-1990s, the average personal computer in someone's home or office was powerful enough to carry out a range of consumer applications which would potentially benefit from the use of a long-distance data communications network, and the Internet had become fast and accessible enough to deliver that use effectively. The other important factor was the old chestnut of standardisation. Just as Dickson's 35mm film format became a truly global medium because the core technical variables were standardised, so did the Internet because the same thing happened to the software used to pass data through the Internet. These standards include the Transmission Control Protocol (TCP), which emerged during the late 1970s, the File Transfer Protocol (FTP), Hypertext Transfer Protocol (HTTP) and Hypertext Mark-Up Language (HTML), which form the basis of the World Wide Web (WWW), the system of transferring text and graphics 'pages' through which most end users access the Internet.

There was a significant repositioning of technological goalposts when Internet-based audiovisual media began to emerge, relative to consumer hardware which was designed specifically and only for recording and reproducing it. The device through which the media was delivered was the personal computer, a machine which is designed to have other applications besides recording and playing audio and video. In particular, its principal data storage medium was the hard disc, which until the late 1990s was a lot less reliable and could not offer the same capacity as removable media such as CDs, DVDs and digital videotape. And unlike with offline media, bandwidth was very severely restricted. Until the early 2000s, the highest speed at which a personal computer could connect to the Internet equated to approximately 5kb per second – at that rate, the volume of data equivalent to a feature film on DVD would literally take days to download. With the advent of the Digital Subscriber Line (DSL), colloquially known as ‘broadband’, in the early years of the twenty-first century, this has increased tenfold to around 50kb per second, but even at that speed our feature film would still take 10–12 hours.

The first audiovisual medium to be transmitted by Internet on a significant scale was audio, for the same reasons that it led the way with offline media. The breakthrough came with the emergence of the ‘MP3’ format in the late 1990s, which was a very heavily lossy compression format that just about enabled 2–3 minutes of CD quality audio to be squeezed into under a megabyte. Being an asymmetric system the encoding process used more computing power, but it was still well within the capabilities of an average PC. At around the turn of the century large amounts of content, mainly popular music, had been illegally encoded into this format and made available for public download on websites throughout the world. The industry responded through a combination of legal action against the owners of the servers holding the illegal media content in the country where they were situated, and the establishment of officially-sanctioned sites which legitimately sold MP3 files for download. Even so, music industry representatives were still repeatedly claiming at the time of writing that illegal downloading was hitting their turnover to the order of tens of per cent.

Use of the Internet for video broadcasting (and for that matter, piracy) is today still not as big an issue as it is with audio, because the infrastructure of the Internet itself is simply not yet capable of moving the volume of data needed fast enough to deliver an equivalent quality picture to a PAL or NTSC television broadcast in anything like real time, even with the heavy compression of MPEG-2 and its competing protocols which are now starting to emerge. Assuming that Moore’s Law applies to the growth of Internet capacity as it always has done, this will start to be an issue around the late 2000s, when we can expect to see the same profile of legitimate and illegitimate Internet use with video as we have with audio. While digitised pirate films have been transferred over the Internet on a small scale, these are so heavily compressed and with such a small image that even the ‘grainy, jerky picture which runs in a tiny window of your computer screen’ reported by the *Guardian* journalist in 1994 would seem like a complimentary description. Another issue is that, while MPEG-2 is bound to be with us for a long time due to its use in dedicated DVB and

DVD hardware, it is by no means the only video encoding method available for PC use. There are several in widespread use, with two key competitors to MPEG being the proprietary systems developed by the two major producers of PC operating system software, Microsoft’s Windows Media format and Apple’s Quicktime. Both offer demonstrable advantages over MPEG for certain specific applications, and before Internet video broadcasting becomes a commercial reality, some serious standardisation will have to take place. At the moment, therefore, consumers without a conscience are far more likely to buy pirate DVDs from their local market stall than to download movies from illegal websites.

The one significant use of digital video delivered specifically via the Internet to date has been to distribute political propaganda which for a number of reasons would not be broadcast by terrestrial television, because the latter is politically regulated on a national basis. Probably the best known example is the use of the Internet for this purpose by Islamic terrorists. It has included the distribution of lectures by the terrorist leader Osama bin Laden following the destruction by his organisation Al-Qaeda of the World Trade Center in New York on 11 September 2001, and videotaped footage of the murders by beheading of a number of Western hostages in the aftermath of the Iraq War. The VCD is also believed to have been used on a significant scale for distributing terrorist and extremist political propaganda before the Internet became capable of that function, and for similar reasons – the discs are very cheap and easy to produce, and are small enough to be illicitly shipped without detection.²² The fact that the videos themselves are of such low resolution that they do not even approach the definition of VHS is not the issue. The point is that these Islamic terrorists are specifically exploiting a unique technological attribute of the Internet: its ability to pass digital media across international boundaries without national governments being able to do much about it. The nature and extent to which the Internet will come under legal regulation is an interesting topic for speculation, but so far the only countries where this has happened have been those in which all ISPs are either government-owned or heavily regulated, China and Iran being two key examples. In any Western democracy, barring access to any given content would require every ISP (i.e. public point of access to the Internet) in a given country to cooperate by blocking access to the servers hosting that content by its customers, and even if they did so an individual would still have the option of establishing a telephone link to a foreign ISP and accessing the banned content that way (albeit very slowly and at a significant cost).

One striking and almost farcical example can be found in the form of a story published in 2003 by an Italian newspaper, which alleged that the Prince of Wales is bisexual and had had a gay affair with his butler. A court injunction was swiftly obtained and the grantees in Buckingham Palace doubtless thought that that was that. Far from quelling any interest, their action probably ensured that the allegation gained more credence than if they had just ignored it. The British print and broadcast media referred extensively to the allegation and its original source without explicitly stating what the allegation actually was, thereby inviting anyone who spoke Italian to head for the nearest PC and download the whole story in a matter of seconds.

Prince Charles consequently discovered that a court order obtained in London had no jurisdiction over an Internet server situated in Rome.

In the early years of the twentieth century, therefore, the significance of the Internet in moving image technology is principally political; though its eventual use as a mass communications medium for digital video in similar ways to that of broadcasting and offline media is almost certainly imminent.

Film

Film comes next down the list of moving image technologies to be affected by the digital revolution in chronological terms, despite having been the first one to be invented. Once again, the reason is Moore's Law. Because the first and primary application film was designed for was to enable the projection of a moving image on a surface area of up to hundreds of square feet, the amount of detail (termed *definition or resolution*) in the emulsion of a 35mm film element is many times that of videotape or broadcast television. To put this in perspective, a digital representation of a PAL television frame using the MPEG-2 system is 720 x 576 pixels (about 25mb per frame uncompressed), but a digital representation of a 35mm film frame with the emulsion densities being produced at the turn of the twenty-first century is around 5,000 pixels square (40gb per frame). An uncompressed feature film stored to the resolution of a 35mm camera negative would require tens of terabytes of hard disc space. At the time of writing, the total replacement of film by computer technology is even further in the distance than full resolution Internet video.

It will be recalled from chapters one and two that the production of film-based moving images has three stages: origination in the camera, the post-production stages (such as editing, the introduction of special effects and duplication) and projection in the cinema (and/or telecine transfer for television and video use). The first of these stages to introduce computer technology was post-production, specifically in the emergence of CGI, hence the significance of *Jurassic Park* and *Terminator 2*.

As with the role of telecine technology in providing the interface between film and video-based media, the use of digital intermediate stages in the production of feature films for cinema exhibition requires a way of converting the images on exposed and processed film into digital data, and then, after the footage has been edited and manipulated digitally, outputting the result back onto film so that it can then be duplicated and distributed as any other cut camera negative would be. The first of these processes involves the use of a scanner, sometimes termed a *datacine* in order to distinguish it from a telecine, the first of which were produced by Eastman Kodak using the trade name 'Cineon' in the early 1990s. The difference is that a telecine outputs an analogue signal which conforms to a specific broadcast television standard (usually PAL or NTSC), dividing the picture up into horizontal lines. A datacine contains a CCD array which scans the picture as a grid of individual pixels and simply outputs the data in that form, without dividing it up into lines or interlacing it. The advantage of this method is that with sufficient computing power to process the resulting data, the output of a single scan can be used to produce

every sort of element from a VHS tape to a high resolution data stream for digital cinema projection:

This concept started to become feasible in 1996 when Philips introduced the Spirit DataCine film scanner. It was the first telecine that, in addition to standard and high-definition video output, provided data output with direct film density representation. Its application to long-form film transfer work, e.g. full-length feature films, was made possible by an increase in transfer rate of one to two orders of magnitude over previous data output-type devices.²³

The volume of data produced for each frame, or the resolution, of a given scanner is commonly expressed in multiples of thousands of pixels squared, e.g. '2K' usually describes an image of for example, 2,048 x 1,744 pixels, assuming the Academy ratio.²⁴ At the time of writing, datacine scanners for professional film industry use are being produced by four principal manufacturers: Kodak/Cineon, Oxberry, Philips/Spirit and Quantel, with resolutions ranging from 2K to 6K. It is estimated that the emulsions of most 35mm camera negative stocks have a grain density roughly equivalent to 4-5k.

Once the source footage has been scanned it can be edited and its visual properties manipulated using a wide range of software tools which have been developed for the purpose. When the use of digital post-production was in its infancy it was usually the case that the majority of editing was done conventionally (i.e. using a 16mm or 35mm cutting copy), with only those sections of camera negative which combined CGI and photographically-originated images being processed digitally, before being output back to film and then spliced into the assembled negative. Throughout the 1990s and early 2000s the role of computer technology in post-production has been steadily expanding. The use of analogue video copies made from a telecine transfer for editing dates back to the mid-1980s, when this method started to become substantially cheaper than producing a cutting copy on film. By the late 1990s editors were increasingly making use of PC-based editing software, which at the end of the process would output an *edit decision list* (EDL). An EDL contains precise instructions that enable the as yet untouched camera negative to be *conformed*, or cut in order to exactly match the edited sequence in the computer, before duplication using the conventional route. One of the many advantages this method offers, especially in low-budget production, is to minimise handling of the camera negative, thereby reducing the chances of contamination or damage.

Outputting the processed image data back to film is a key requirement of digital intermediate work, and will remain so unless and until film totally disappears from all stages of production, post-production, distribution and exhibition. Film recorders expose each individual pixel onto a 35mm fine-grain internegative or interpositive element in one of two ways. Until around the turn of the century, CRTs which expose red, green and blue (or yellow, cyan or magenta, depending on whether the destination film stock is positive or negative) colour records sequentially and through filters (it will be recalled from chapter six that the luminance generated by a CRT does not in itself have any colour) were the norm. During the late 1990s laser-

operated film recorders started to be developed, though they remained prohibitively expensive for most operations until the early 2000s. As this description of Kodak's Lightning recorder suggests, lasers are a far more accurate method of burning image data onto film:

The Lightning Recorder's technology is unique in that it uses red, green and blue lasers to expose negative film. The three lasers write directly to each colour layer of the intermediate stock. This combination produces images of unparalleled sharpness and colour saturation. Cinesite [the post-production company from whose website this text is quoted] outputs to 5244 intermediate stock, which is virtually grainless (ASA 3) and has the ability to record the grain pattern from faster stocks so that the digital transformation is unnoticeable in the final edit.²⁵

At the time of writing, studio cameras which originate images as digital data are not being used on any significant scale for major feature-film production, although they are starting to replace 16mm and Super 16 for low-budget feature-film and high-budget television drama and documentary production. The 'Digital High Definition' (known as 'HD') format was developed by Sony in the late 1990s, the key breakthrough being the launch of the HDC-F900 camera in 2000. This incorporated the '24-p' format meaning that, as in a datacine, the camera's CCD exposed progressive, discrete frames at the rate of 24 per second, in order to maximise compatibility with film. HD's most vocal advocate in the film world to date has probably been the science fiction director George Lucas, who shot one of his *Star Wars* series of films almost entirely using the following generation of HD camera, the HDC-F950, in 2003. In the main, though, feature studios have preferred to stick with film: new generations of stock remain compatible with their existing cameras, the perceived image quality of HD is not believed by most people to match the best that film can offer and as yet there is no demonstrable economic advantage in abandoning film for studio origination in cases where extensive special effects are not needed.

Cinema exhibition

Any adoption of a new methodology or a new technology must take into account two issues: (i) does the new way decrease the cost of getting the film into distribution, or (ii) is there a different benefit, such as increased creativity? Change for the sake of change in the film world is not common practice.²⁶

Of all the areas of moving image technology in which computers are playing an increasing role, its use for projection in cinemas has arguably generated the most controversy within industry circles, and its future is possibly the most difficult to predict.

It is worth bearing in mind that the electronic projection of television images onto large screens is nothing new. Probably the earliest successfully demonstrated and commercially used technology for this purpose was the Eidophor (from the

Greek *Eidos* – 'image', and *phor* – 'conveyor'), developed by the Swiss scientist Fritz Fischer, first demonstrated in 1943 and remaining on the market in one form or another until 2000. The light source of the Eidophor was the same as in a conventional cinema projector – carbon arc at first, with later models being lit by xenon arc bulbs. The imaging device consisted of a thin layer of oil, known as the 'Eidophor liquid', coated on a reflective surface, the opacity of which could be modulated by an electron beam similar to the ones used in CRT television receivers. In effect, the coated surface acted as the equivalent of the gate in a film projector: by projecting the arc light through it and focusing it with a lens, it was possible to project a television image onto a large screen.²⁷

The Eidophor was used on a small scale for cinema exhibition in Europe and North America, usually for live feeds of special events such as sports and election coverage. Its core market, though, was in non-theatrical venues such as university lecture theatres. Other television projection technologies emerged during the 1970s and 1980s which gradually superseded the Eidophor. The use of high-power CRTs for large-screen projection was pioneered by, among others, the Belgian-American Radio Corporation (Barco): these designs featured three separate tubes, one for each of the primary colours, each of which had to be aligned with the others to produce a properly focused image in projection. In the early 1990s LCD arrays that could withstand the heat of a projection lamp were developed, resulting in the production of small, portable projectors that were suitable for displaying video or the display of a PC monitor in settings such as classrooms and business meetings.

Until this point no one had seriously considered the use of electronic projection as a replacement for 35mm and 70mm film in cinemas. The only moving images that Eidophors, CRT and LCD-based projectors had been able to display thus far were those supplied either by a videotape or a live broadcast – in other words, electronic projection was limited to the definition of PAL or NTSC did not even come close to the image quality offered by film. Two factors changed that: the effect of Moore's Law, which, by the early 2000s, had given us the computers and data storage space capable of handling the volume of data needed for digital moving images to approach the resolution of film, and the invention of the Digital Light Processing (DLP) imaging device.

The DLP technique was initially developed by Larry Hornbeck of the American electronics giant Texas Instruments in 1987. The guts of the system is the 'digital micromirror device', a microscopic mirror mounted on an adjustable hinge. When a light source is placed at right angles to the device, the angle of the micromirror determines whether or not it will refract the light through a lens in front of it. Each DLP chip contains an array of micromirrors of a similar density to the imaging CCDs used in HD cameras and datacines. A DLP cinema projector will contain three chips, one for each of the primary colours. Today, 2K DLP cinema projectors are being marketed, with the launch of 4K models said to be imminent.

DLP has been aggressively promoted by the electronics industry as a replacement for release prints on film. Theoretically it provides the final piece in the digital imaging jigsaw where feature films are concerned: the combination of HD cameras,



Fig. 8.1: A 1.3k DLP projector installed alongside a conventional 35mm film projector in a modern cinema.

computer-based post-production and DLP cinema projection is being promoted as finally enabling moving images to be produced and shown with a quality that was previously only available using film, only without any film. The detractors of digital imaging point out that resolution is not the only relevant criterion in determining the quality of a projected image. Colour depth – the number of separate shades it is possible to colour each pixel – is also an issue, with film's supporters pointing out that with a photochemical emulsion this is infinite, whereas in the digital domain it is

fixed and carries a data storage penalty the higher you go. The inevitable response to this is that Moore's Law will solve that problem sooner or later.

There is another significant issue involved, though; one which will not necessarily be solved by Moore's Law. This is the economic structure through which distribution and exhibition on film has traditionally been supported. A research study on the economic potential of digital cinema projection carried out by Wall Street financial researchers in 2002 noted that 'the hype surrounding the technology was overwhelming' when they first examined the sector in 2000, but concluded that 'structural issues in the industry were likely to serve as a major impediment to adoption'.²⁸ Those structural issues still exist, and if anything represent an even greater impediment than they did in 2002. A key characteristic of 35mm is that the imaging device is contained within the film itself, in the form of the emulsion. The research and development which goes into producing each successive generation of camera negative, intermediate and release print stocks results in a product which delivers a demonstrably 'better' image than its predecessor, but which remains compatible with existing cinema projectors. The projectors themselves are essentially very simple, mechanical devices. A 35mm projector built in the 1940s will, if properly maintained, be able to display the benefit of a new generation of release print stock sold in the twenty-first century. At present a typical new 35mm projector mechanism costs around \$30,000. A state-of-the-art DLP projector costs over \$150,000.

With DLP the technology is in the hardware. Not only is the projector a far bigger investment, but the resolution and colour-depth possible from a given imaging device is fixed and cannot be upgraded. In this respect Moore's Law may even be a negative factor rather than a positive one – whereas with film, a 35mm projector can reasonably be assumed to have a service life of several decades, the rapid growth in computing power may well mean that a year or two after a cinema invests in DLP, its rival venue acquires a newer model of projector, producing a better image and rendering theirs obsolete and economically uncompetitive.

Furthermore, history shows that new technologies which have required a substantial investment at the exhibition end usually fail. The industry wanted exhibitors to install widescreen in the early 1930s, but exhibitors decided that it was not worth it. The same happened when Twentieth Century Fox tried to package magnetic sound with CinemaScope in the 1950s and when Eastman Kodak tried to launch a digital sound system that was not backwards compatible, CDS, in the early 1990s. It is hardly surprising that the electronics industry is lobbying Hollywood hard over DLP, stressing the potential savings to be made by now avoiding the cost of striking large inventories of 35mm prints, which in the case of a big blockbuster can run into millions of dollars. But cinema exhibitors are not about to spend the huge amounts of money needed even for the cinemas in North America and Europe to convert on a meaningful scale, let alone the rest of the world, without a substantial proportion of the studios' savings being passed on to them. So far, no sign of that happening is apparent.

Other obstacles remain, chiefly that of standardisation. Hollywood representatives have for some time taken the line that 4k should be considered a bare minimum for any systematic rollout, in order to show-off high-budget production values to the best advantage. Independent filmmakers and the 'art house' sector are calling for standardisation at 2k, as this would substantially lower the cost of encoding the image data from whatever source it was originated on to the format used by the cinema DLP systems. As things stand, 35mm film is a global, universally agreed standard, and no-one in the industry wants the nightmare of it being replaced by a myriad of incompatible formats and systems, which is precisely what has happened with videotape. Furthermore, film industry representatives are increasingly voicing fears over the potential risk of piracy, especially if the method of delivering the image data to cinemas is some form of online data transmission, which could theoretically be intercepted. Were this to happen the stakes would be even higher still than with DVD piracy, because criminals could potentially get their hands on clonable, cinema-quality copies of films weeks or months before they were scheduled to be broadcast or released on consumer media. Unlike most electronic media, 35mm film is inherently very safe as far as piracy prevention goes: a release print of a feature film is bulky, very heavy, kept in buildings which can be made reasonably secure, requires specialist technical skills to handle, requires expensive, specialist telecine equipment to duplicate and being of much higher contrast than intermediate elements, does not telecine very well.

The result of all these factors is that cinema exhibition is about the one area of moving image technology in which the industry actually seems to be withdrawing from the digital route. A projection engineer I spoke to recently told me that 'the industry doesn't want it [digital projection] and isn't willing to pay for it'. As things stand, the only people who are persisting with digital projection in its current form are public sector arts organisations, of which several in European countries are advocating the use of DLP for enabling low-budget, specialist and archival re-release films to be shown in 'art house' venues when the cost of striking a 35mm print would be a significant barrier. They are operating in the usually mistaken belief that DLP will, for

this market, bring the overall cost of distribution and exhibition down while pushing the range and technical quality of material shown up. One particular example is the British Government's film agency, the UK Film Council (UKFC). In 2003 it announced its 'Digital Screen Network' initiative, a £15 million fund intended to equip up to 250 independent cinemas with DLP projectors. The statements which appeared on UKFC's website, characterised by emotive rhetoric such as 'freeing cinema from the tyranny of 35mm', during the following year demonstrated quite obviously that whoever had thought up the idea had very little technical knowledge; had not addressed fundamental questions of compatibility, standardisation, the cost of encoding and distributing digital media to venues, the cost of maintenance and rapid equipment obsolescence and much else besides; had not sought the advice of any real technical experts (or if they did, had ignored it); and had failed to do the simple sums which would have revealed that UKFC's distribution and exhibition objectives could have been met far more cheaply and efficiently using 35mm. That such a scheme ever got past the drawing board is, in the opinion of this author, almost a criminally negligent waste of taxpayers' money, and a powerful demonstration of the spin and propaganda which surrounds the 'D word' where moving image technology is concerned.

It is impossible to predict if and when cinema projection will eventually go digital, because the barriers to conversion lie more in the economic and political domain than the technological. The speed at which the quality of digital imaging improves is going to have to start outpacing that of film. Up to now it has not; indeed it could even be argued that Moore's Law applies to the evolution of film emulsions as much as with computer-based imaging. There will also need to be a shift in the economics of film distribution so that the investment in digital projectors can be amortised more evenly across the different sectors of the industry than would be the case if cinemas just went out and bought them. Once again, there is no sign of any meaningful progress on this front yet.

Archival restoration and preservation

The 'D word' essentially means two things for archivists, one of them a godsend, the other a nightmare. The godsend is in the use of digital intermediate technology to carry out film restoration functions which could previously only be dealt with by photochemical methods. Until the late 1990s, for example, scratches and dirt could only be removed by invasive chemical treatments applied to original elements and/or expensive photochemical duplication using techniques such as wet-gate printing. Elements with faded colour dyes could also only be restored by photochemical duplication, specifically by manipulating the light source used to expose the destination stock.

These and other defects can now be corrected using what is essentially the same technology that gave us the special effects in *Jurassic Park* and *Terminator 2*. An original element can be scanned, the image manipulated using a computer and the result output to whatever format is needed for access purposes – potentially

anything from a low resolution MPEG file to a new, laser-burnt 35mm intermediate. The importance of retaining original elements (or as near a generation to them as can be acquired) for long-term preservation in appropriate atmospheric conditions is thus heightened, as this enables them to be rescanned as necessary as and when Moore's Law gives us higher-quality scanning and image-manipulation technology. At the time of writing the use of digital technology for restoration work is just coming within the financial reach of well-financed private libraries and large, national public-sector archives, but is not yet cheap enough to be a routine tool used by smaller and more specialist organisations. But in this case it should just be a case of sitting back and waiting for Moore to deliver the goods.

In fact, the issue with which archivists will increasingly have to grapple in relation to digital technology is probably more ethical than technical. In 1990, the joke in *Gremkins 2* in which a cable TV station broadcasts '*Casablanca* – now in full colour and with a happy ending!' was effectively science fiction. In 2004, colourising *Casablanca* to broadcast standard is now well within the capabilities of an average home PC, given the requisite software. This opens up the possibility – indeed, some would say the *probability* – that films will find themselves being 'restored' by individuals and organisations which either do not have the contextual and historical knowledge needed to evolve a coherent model of the 'original' state they are aiming to restore (a hypothetical example would be digitally removing the Dufaycolor réseau pattern in the belief that it is a defect, even though it is in fact a characteristic of the original film) or are working more to a commercial agenda than a cultural one (e.g. digitally removing the Dufaycolor réseau pattern in footage being broadcast in a documentary, in the full knowledge that doing so destroys the technical integrity of the footage but deliberately, on the grounds that uneducated viewers would believe it to be a technical fault). Some PC video editing software in widespread use today (such as Final Cut Pro or Adobe Premiere) actually has the facility to *add* digitally simulated image characteristics such as scratches, flat contrast and a pink, faded hue – presumably in order to suggest to viewers that footage is 'old'.²⁹ Once this digital genie fully emerges from its bottle, therefore, archivists will have to develop effective and considered strategies for protecting the technical integrity of the footage in their care, once representations of it enter the digital domain.

If digital restoration is the archivists' godsend, attempting to preserve 'born digital' moving images is their nightmare. The term 'born digital' refers to media content which has been originated digitally and has never existed in an analogue form considered of equivalent quality. Examples would include a feature film shot on HD, edited digitally and only ever projected using DLP without any film transfer ever taking place, television news footage originated on Digibeta or DVCAM, or home movies shot on Digital 8 or Mini DV. It will be recalled from chapter seven that the only known reliable method of preserving footage originated on videotape in the long term is 'continuous format migration', i.e. copying from one tape format to another as the former approaches obsolescence. When starting with an analogue medium this is bad enough, but with digital two new problems also have to be addressed. In an analogue videotape, partial signal loss will result in a visible defect (a common

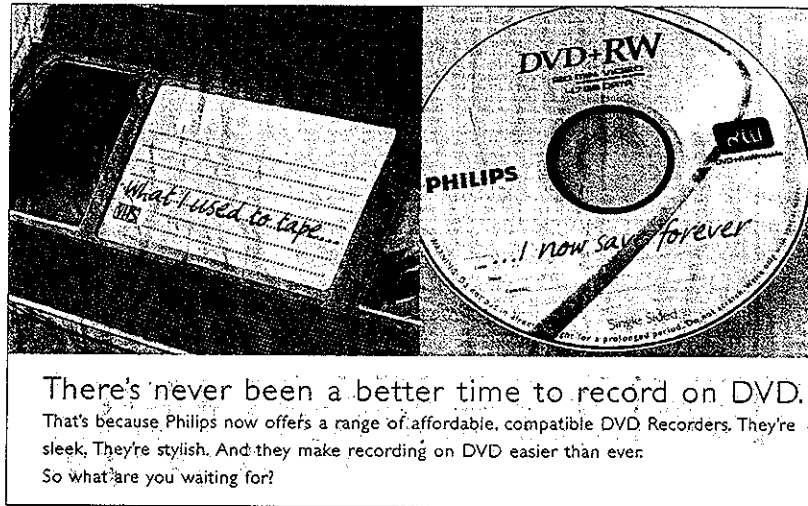


Fig. 8.2 The challenge of preserving 'born digital' in the form of a British newspaper advertisement for consumer DVD recorders, published in the summer of 2003. Many archivists would regard the claim that the data recorded on this medium will be preserved 'forever' as naïve at best, and a cynical manipulation of their customers at worst – especially as the disc shown in the picture is of a format (RW) which is specifically designed to be erased and rerecorded! The implicit and probably unintentional acknowledgement by a major manufacturer of VCRs that their tape format of choice is effectively useless for long-term preservation ('What I used to tape...') is nevertheless remarkable.

example is the momentary loss of picture sync known as a 'dropout'), but the rest of the image will still be displayed. But if, for example, a significant amount of digital data is lost due to chemical decomposition of the magnetic oxide on a tape, one has potentially lost a much larger part of the moving image sequence. So-called error correction protocols – software which uses surviving data to 'guess' the lost part – will work up to a set volume of data loss, but beyond that the whole picture will be lost. In crude terms, an analogue videotape suffering chemical decomposition will play with an imperfect picture on the screen, but a digital one with an equivalent level of damage will play nothing at all. Another issue is that data compression reduces the tolerance of error correction – so, for example, if you lose ten seconds' worth of video on an uncompressed tape, the same volume of data on a compressed tape might represent 20 or even 30 seconds. In practice, archivists have no choice but to try and pre-empt format obsolescence, and to continually examine collections of digital tapes in their care in an attempt to pre-empt physical damage or decomposition.

Compression also imposes another problem, one which actually negates the 'clonability' advantage of digital media. Virtually all the digital video formats in widespread use implement compression of one sort or another. In most cases the only way to decompress them is through digital to analogue conversion. For example, no Digibeta VTR currently on the market will output the datastream bit for bit as it is encoded on the tape. So format migration can only be accomplished by convert-

ing from digital to analogue and then back to digital using a different compression system, or, if one is lucky, 'transcoding' the data from one system to another in the digital domain. Either way will introduce generational loss (the latter marginally less so), just as analogue copying methods will.

The volume of 'born digital' video content being generated in a range of market sectors is now simply so large that it will prove impossible to preserve even a significant proportion of it in the long term. This situation is bought about by the fact that no offline data storage medium currently available has been shown to preserve data integrity in long-term storage for anything approaching that of analogue film, and furthermore no new solutions have emerged (as with the temperature and humidity controlled storage of film, for example) to alleviate this situation. Magnetic tapes shed oxide and suffer from dropouts leading to data loss, and even if they were to refrain from doing so the format obsolescence problem remains. Optical discs – CDs and DVDs – are to all intents and purposes an unknown quantity, but overwhelming anecdotal evidence suggests that recordable DVDs burnt shortly after the format's launch in 2001, handled carefully and stored in cool, dry conditions, have already undergone chemical changes which have rendered them unreadable. In a generation's time we may well find ourselves in the strange position whereby film records from the twentieth century have survived intact, but our digital moving image record of the early twenty-first has largely disappeared.

Conclusion

The 'D word' in essence signifies the use of computers to originate, manipulate, distribute and display moving images and sounds. The theoretical possibilities this offered were realised and developed in the early twentieth century, but the data processing power needed to make them a reality did not become widespread until the 1990s. At the time of writing – an indicative though unintentional catchphrase of this chapter – they had made a profound impact on some forms of moving image technology (for example, videotape recording had almost entirely 'gone digital' save for the continuing use of VHS as a consumer medium) but less so in others (principally cinema exhibition). The only reliable tool we have for predicting what happens next is the ubiquitous Moore's Law – and even then, as demonstrated by the commercial failure of digital cinema projection thus far, it is often not the only factor in play.