Camera Training

ARRIFLEX D-21 – Digital Camera Basics

Last changes: 2009-10-21

Digital Camera Basics

Digital video cameras generally consist of the same main components: Camera lens, imager, image processing and video output. How well a camera performs depends on construction, quality and combination of these components. The following explanations shall provide a few basics that allow estimating the eligibility of a particular camera for cinematic productions.

1 Imager

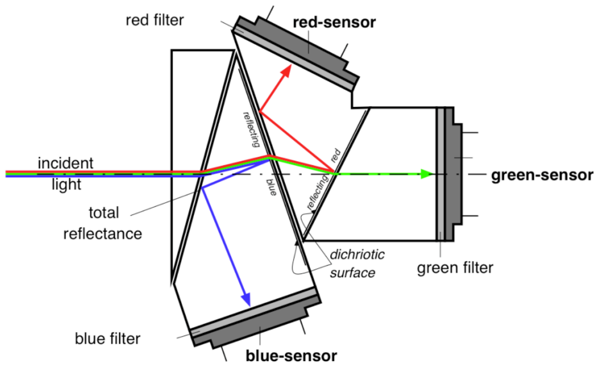
The imager in a digital video camera consists of a single or three semiconductor chips with an array of light-sensitive cells (photocells). It transfers incident light into electric charge, while it is only able to differentiate a change in luminance and therefore only delivers a black and white image. In order to create a colour image, the incident light has to be split into its color components (e.g. red, green, blue) during capture.

1.1 Photocells

When exposed to light, photocells create an electric charge that is proportional to the amount of incident light. After an exposure, these electric charges are read out and passed on as brightness values in an electronic raster image. A photocell is often referred to as a "pixel" (short form for picture element), which easily causes confusion as a pixel on the sensor is not necessarily the same as a pixel in the output image. Depending on the imager technology, an output pixel may originate from photocells of different sensors (3-chip imager) or from a number of photocells on the same sensor.

1.2 Three-Chip Imager

In a three-chip imager, the incident light is sent through a beam splitter to divide the image into three colour components. A sensor, on one of the three surfaces where the light exits the beam splitter then captures the luminance values for one colour component. At each end of the beam splitter, a sensor then captures of the colour component images. All three sensors have to be exactly aligned, as even a small misalignment will produce colour fringes in the image. 3-chip imagers are sensitive to temperature changes. A difference in the surrounding temperature between factory and shooting location already leads to slight misalignment. Another problem occurs when HD high-speed lenses are used. Light entering these lenses at a great off-angle is absorbed by some beam splitters, which causes vignetting (Image gets darker in the corners).

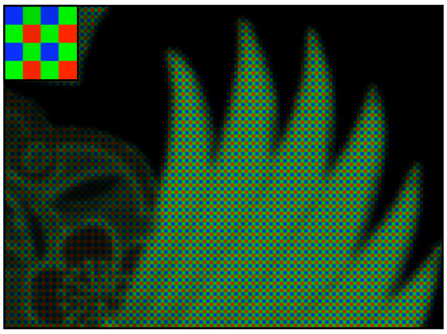


1.3 Single-Chip Imager

To enable the separation of color components with single chip imagers, manufacturers of digital cameras mostly rely on the color filter array (CFA). A CFA is a small mosaic pattern of colour filter dyes, applied directly onto the sensor. One form of a CFA is the Bayer pattern, made from green, red, and blue filter dyes, which is used for the D-21. Each photocell is covered with only one dye, so only the information for one color component is captured. The luminance information for the two missing color components is then interpolated from surrounding pixels in a so-called de-bayering or color reconstruction process.

All CFA imagers have in common that the effective output resolution is lower than the photocell count on the sensor. This needs to be compensated by oversampling. Thus, the D-21 sensor has a 2880 x 2160 photocell raster, which results in alias-free 2K resolution after de-bayering.

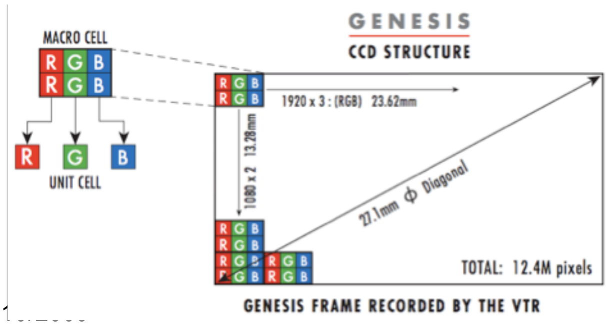
Before processing, a raster image is referred to as raw data (image) or ARRIRAW file. After color reconstruction, it is referred to as RGB data/RGB image.



Enlarged image section from a raw data image with Bayer structure

Another type of single-chip imager can be found in Sony F35 and Panavision Genesis. On this sensor, the photocells are covered with filter dyes in a stripe pattern. To retrieve a full color image, each RGB output pixel is calculated from a group of 3 x 2 photocells (see image, macro cell).

Similar to a CFA sensor, stripe pattern sensors have a loss of effective resolution in consequence of the macro cell structure. In case of F35 and Genesis, the macro cell structure delivers a 1920 x 1080 image from the 5670 x 2160 photo cells on the sensor.



Source: Panavision

1.4 Sensor Types

1.4.1 CCD (Charge-Coupled Device)

A CCD sensor consists of an array of photocells and shift registers, which transport the charges away from the photocells. As there are different principles used for this transport, several types of CCDs can be found. All types of CCDs have in common that the charge read out sequentially, line by line, which makes this technology less flexible. As CCDs have been around since the 70’s, this technology is extensively mature.

1.4.2 CMOS (Complementary Metal Oxide Semiconductor)

A CMOS sensor also consists of an array of photocells. Unlike CCDs a CMOS works without shift registers and allows an individual readout of the charge from any pixel. Aside from applications, such as readout of a smaller window in favor of higher available frame rates, CMOS offers a number of advantages, which will be explained in the next section. CMOS sensors did not get much attention until the 90’s. In the beginning, these sensors exhibited a raised noise level compared to CCDs. Technical advances, however, soon turned CMOS sensors into a strong competitor for CCDs, which why many camera manufacturers (especially in the digital still picture industry) switched to this technology.

1.4.3 Windowing – Readout of a Reduced Image Area

Reading out an area smaller than the full sensor size is called windowing. This reduces the amount of information coming from the sensor and thus allows capturing higher frame rates (e.g. 50 fps full sensor readout and 100 fps 50% window). Due to their design principle CCD sensors offer limited windowing capabilities. CMOS sensors allow readout of individual photocells and thus, unrestrained use of windowing. Even though windowing has no effect on the image angle, the reduced frame size causes the projected image to appear at a different scaling factor (magnified). If a scene needs to appear in the same scale, the different image size caused by windowing needs to be compensated by using a wider lens.

1.4.4 Artifacts

Aliasing

Image elements containing very fine structures or patterns may cause jagged lines, flickering areas or shifting colors in the output image. As this effect depends on the spatial frequency of the structure or pattern in the frame, it often can be eliminated by moving the camera slightly closer to or further away from the object – or, if a zoom lens is used, by a slight change of the focal length setting.

Blooming

Blooming describes an effect, which makes a bright spotlight appear bigger. This effect is only occurs with CCD sensors. CMOS sensors are immune to blooming.

Clipping

Clipping occurs on the upper limit of the dynamic range, i.e. high exposure. Neither CMOS- nor CCD sensors offer good overexposure handling. The photocells collect their charge proportional to the amount of incident light until they reach their maximum capacity and abruptly become saturated (clipping). Differences in brightness beyond this limit will not be reproduced. Negative film stock provides more reserves for overexposure as its characteristic curve has a very soft shoulder at the top. When overexposed, differences in brightness are therefore reproduced with decreased contrast ratio rather than being cut off, as with a sensor.

Dark Spots

Dust particles that have settled on the imager's front surface may become visible as dark spots in the captured image. The degree of this effect depends on the aperture of the lens, but also on the distance between front surface (where dust may have settled) and the image plane (the photo cells). Therefore a 3-chip imager is less likely to have problems with dust than single-chip imagers. Changing the lens only in a clean environment combined with regular dust checks, however, is a good remedy against dark spots.

Horizontal Smear

Bright image elements inside or around the edges of the captured frame may produce horizontal trails in the output image. Depending on the sensor design, the effect may also show a slight difference in luminance between the left and right image half. This effect is called horizontal smear and relates to readout rather than transport of electric charge and thus can occur for both CCD and CMOS.

Noise

Increasing the camera's sensitivity setting (e.g. EI settings for the D-21 or +dB settings for other) or adding electronic gain in postproduction produces more visible noise in the output image and may even cause other image artifacts to become apparent. Some cameras offer countermeasures, such as adaptive noise cancellation, which only works to a certain degree before it negatively affects the image quality (e.g. details, color resolution).

Pixel Failure

Defective photocells deliver improper luminance values causing bright red, green or blue pixels in the output image. This type of problem may occur at any time and may have different causes (e.g. heavy vibration or shocks, electrical discharge, laser rays projected on the sensor though the lens). Removing defect pixels in postproduction is a costly process, but some cameras can detect and remove defect pixels so this will not be necessary. Different manufacturers have different approaches to get rid of defective pixels. Sony users need to regularly perform automated black balancing in order to keep the risk of occurring defect pixels small. The ARRIFLEX D-21 automatically performs a defect pixel correction for each individual exposure.

Smear

Smear is an effect showing a bright stripe across the image. This effect can be seen in dark scenes with solitary glaring spots or lines of light. CCD sensors are quite sensitive to smear due to the way electric charges are transported for readout, but there are different measures that can be taken to minimize this effect (e.g. mechanical shutter). CMOS sensors generally are much more smear resistant than CCDs and usually do not require pre-emptive measures.

1.5 Imager Characteristics

An imager can generally be characterized by its sensitivity to light, its dynamic range, the available bit-depth (even though this is not a part of the exposure, but the readout process) and its resolution.

1.5.1 Sensitivity

The amount of collected charge in a photocell is proportional to that of the incident light whereas smaller photocells collect less charge than bigger ones. Additionally, chips of the same design (same CCD- or CMOS technology) generate a noise level that is mostly independent of the amount of photocells they have. This noise level obscures the charge collected by the photocell and thus reduces the range available to create useable signals. As the result, sensors with larger photocell structures provide better exposure sensitivity performance.

When comparing a 6 megapixel sensor and a 12 megapixel sensor of the same size and design (e.g. 24 x 18 mm), the 12 MP sensor has twice as many photocells, each photocell half the size and the sensor (roughly) half the sensitivity.

However, there are further factors influencing the sensitivity to light:

- Total photocell surface area used to create one output pixel (i.e. the number of photocells combined to one output pixel) (also see: **Photocells** and **Resolution**).
- The size of the light-sensitive photocell surface (fill factor). CCD sensors provide approx. 70 to 100%, CMOS sensors carry more circuits and have a smaller fill factor of approx. 50%.
- A camera’s sensor is located behind a number of filters absorbing e.g. UV and IR light, filtering color components, or keeping away very fine structures (spatial frequencies) to avoid aliasing. How much light gets lost depends on quality and amount of these filters.
- Noise level of different sensor designs.

1.5.2 Dynamic Range

The dynamic range represents the difference between the darkest and brightest tones that can be captured in a single exposure. The noise floor caused by the imager electronics limits the dark end of this range. The bright end is limited by the maximum collectible charge until clipping is reached. Simply speaking, of two sensors with the same size the one with lower resolution (or larger photo cells) is the one with the bigger dynamic range.

1.5.3 Bit-Depth

The bit-depth (or quantization) of an image determines the amount of tonal steps that are rendered across the dynamic range. An 8 bit system can show 2^8=256 tonal steps per color component. If an original image requires more tonal steps to be correctly reproduced than the imaging system can offer, an artifact called "banding" (i.e. visible steps in a tonal range) appears. Putting an A/D converter with more bit-depth into a camera offers better tonal representation, but comes at the cost of increased storage and power-consumption.

Lower bit-depth

Higher bit-deph

Film negative is usually scanned with 16 bit quantization (65536 tonal steps per color channel). Digital motion picture cameras usually provide 10 to 14 bit quantization at the camera head. If the signal is output using HD-SDI and recorded to e.g. HDCAM SR, the quantization is effectively limited to 10 bits. HDCAM, its predecessor only offered 8 bits. Storing to a lower bit depth even though the camera head offers higher bit-depth, however, does make sense as the internal image processing of the camera can use the higher bit-depth images to deliver an optimized output with lower bit-detph.

1.5.4 Resolution

The resolution limit of an imaging system is defined by the finest spatial frequency (black and white bar pattern) that still can be differentiated in the output image. This evaluation is done by gradually plotting the contrast response (modulation) for increasing spatial frequencies, which delivers the modulation transfer function (MTF) for the imaging system.

The resolution limit depends on different factors, such as output pixel count (not necessarily identical to photocell count) and the characteristics of the entire optical system in front of the sensor including the lens and the filter pack (especially the optical low pass filter used to eliminate those spatial frequencies, that would cause aliasing [moiré effect] in the output image). For cameras with a CFA- or Bayer pattern imager, the quality of the color reconstruction algorithms also has great impact on the result.

For more information on resolution evaluation, please download the **4K+ Systems Technology Brochure** by Dr. Hans Kiening, which can be found in the Tutorials section on the ARRI website <http://www.arri.de>.

1.5.5 Depth of Field

When a lens is focused on a certain distance, only objects located at this distance will be in absolute focus. Objects in front of or behind this focus plane are projected with decreasing sharpness depending on the distance from the focus plane. The distance-range at which objects in front or behind the focus plane appear to be in focus (are projected with sufficient focus) is called depth of field (DOF).

A large DOF is favorable for applications such as news gathering as the camera operator usually does not have enough time to set up scenes for shooting. Each shot has to work on the first go; if it is a politician walking by, a group of people marching by in a protest rally or a rally car racing by.

Scenic productions greatly rely on a short DOF, which allows separating actors or objects from a scene. This is also called selective focus. Using the DOF has many applications, such as guiding the viewers' attention, emphasizing a character's emotions, or creating a surprise effect.

Factors influencing the focus range:

- Increasing the lens aperture (= smaller f-number on the lens) decreases the DOF.
- A smaller object-to-image ratio produces a larger DOF. A wide angle lens thus delivers a larger focus range than a telephoto lens.

Another causality resulting from the object-to-image ratio can only be influenced by choice of the camera:

- Using a camera with a smaller image field (active area) provides a larger DOF.

The table to the right shows imager sizes for different cameras in comparison to the Super 35 format, which represents the standard for scenic productions.

Type	Width (mm)	Height (mm)
ANSI Super 35	24,9	14,01
RED One	24,4	13,72
D-20 (16:9)	23,76	13,37
Panavision Genesis	23,62	13,28
ANSI Super 16	11,95	6,72
2/3" HD Cameras (F-23, F-900, Viper,...)	9,6	5,4

Table: Dimensions of different image fields compared to Super 35 film (format 1.78:1)

When using cameras with smaller imagers the DOF can be reduced by shooting wide-open apertures or by using longer lenses. These solutions, however, often introduce other problems:

- Working with a fully opened aperture usually causes the need for ND-filters. This can lead to problems when the number of available filter frames in a matte box are already taken up by e.g. polarizing- or color conversion filters. The use of filters sometimes also causes unwanted reflections in the image.
- Using lenses with longer focal range means that the camera has to be placed at a greater distance. This solution, however, does not work in sets offering only a confined space.

As an alternative to these solutions, video cameras with smaller imagers can also be equipped with a 35 mm image converter. Depending on the camera in use, these converters can be attached either to a camera's fixed lens or directly to the camera's B-4 lens mount. The front end of the converters provides a PL-Mount for the use of 35 format lenses. The attached 35 format lens projects its image onto a frosted glass screen inside the adapter, which is then captured by the video camera's imager. As the projection has the size of a 35 mm frame, the resulting image shows the DOF characteristics of a 35 format camera. The drawbacks of these converters are:

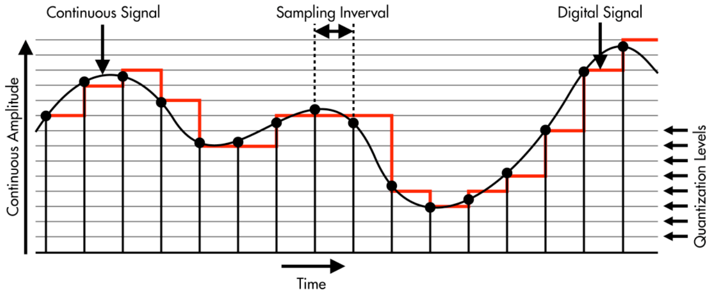
- a brightness loss of approx. one stop.
- a considerably larger and less stable camera.
- an additional possible source of operational errors.

2 Signal Processing

Once the imager converted the incident light into an electrical charge, this charge is turned into a signal for output to the recorder through several processing steps. The following is an explanation of common processing steps.

2.1 Analogue/Digital Conversion

The A/D conversion turns a continuous analogue signal into a discrete set of digital values.



After an exposure, the amount of charge collected by a photocell is recorded as a measured value (sample). Then, a digital value (code value, short form: CV) is assigned to the measured value, depending on its amplitude. The accuracy of the digital representation for the measured analogue value is determined by the set of digital values, the so-called quantisation depth or bit depth.

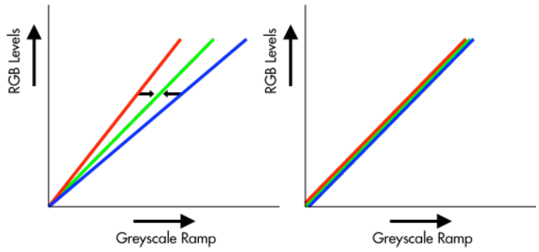
A/D converters with 8bit quantisation depth provide 2^8 = 256 digital levels. 10bits already deliver 1024 levels and 12bits (as used in the D-21) 4096 levels.

The accumulated values for all photocells are the digital representation of the scene that is projected on the sensor. As explained above, the bit depth determines, how accurate the digital image reproduces the brightness differences of the scene and the accuracy of the following processing steps.

2.1.1 White Balance

As human colour vision quickly adapts to changing lighting situations, a white piece of paper, illuminated with daylight or tungsten light, is always perceived as white. For the imager of a camera, however, pure white is a fixed combination of certain spectral colour components. Therefore, a camera's output image of a white paper ideally would be white, but mostly is slightly reddish or bluish.

When no white balance is applied to adjust a camera to the prevalent colour temperature, the output image of a neutral grey ramp from black to white (grey wedge) shows different gradients for the three colour channels (RGB). To avoid the deviation of the colour channels, a correction value has to be applied to their gradients. The result is a neutrally coloured reproduction of the grey ramp.



When shooting on film, the camera can be matched to the colour temperature by using colour negative stock for tungsten or daylight.

Using gel filters on the lights or in front of the camera is another option for white balancing, but is mostly used to convert daylight to tungsten, tungsten to daylight, or to reduce green colour purity errors of e.g. neon lamps.

2.1.2 Adjusting Sensitivity

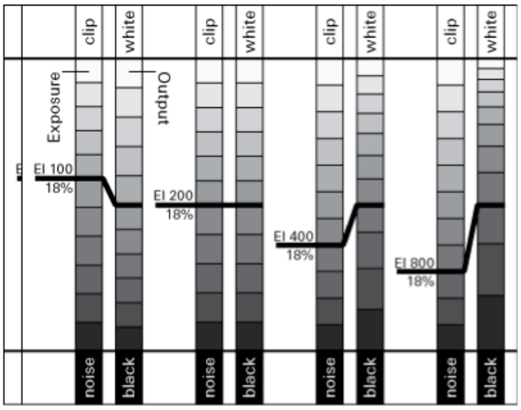
There are a number of reasons that can restrict the possibilities of setting the lighting for a scene or that completely prohibit the use of artificial lighting. Examples would be confined spaces, budget limitations and not least the artistic or aesthetic claim.

Different tools can be used to match a camera to the prevalent lighting conditions. If a scene is too bright or too dark, the lens aperture may be altered, ND-filters may be added or removed, or film stock of a different sensitivity may be used. The latter option, of course, is only available when shooting film. The imager of an electronic camera cannot be exchanged. In principle, these cameras only offer a single sensitivity, given by the properties of the sensor. In response to this, electronic cameras allow adjustment through variable signal amplification. This, however, cannot be equated with the use of negative film stock with different sensitivity, which shall be explained in the following example.

A scene shall be shot using available light at 24fps and F-stop 2.8. The exposure meter indicates that at this lens aperture, we will need a sensitivity of 400 ISO/ASA. The imager of our digital camera has a base sensitivity of 200 ISO. In order get an exposure according to the camera's sensitivity, we would have to open the iris by one stop. However, the DP demands the iris to be left at 2.8 as only then, he will get the depth of field he wants.

Result:

- The middle grey (18%) in the scene is underexposed by one stop.
- The sensor's range below middle grey is reduced by one stop, giving us less definition in the blacks (see image).
- The range above middle grey is extended by one stop, which delivers more detail in the highlights.
- The sensor's output has to be amplified in the camera so it provides a visually correct reproduction of the underexposed image.
- The middle grey, rendered too dark by the sensor, is raised to the correct level for the camera output.
- The blacks are spread, whereas no additional information can be obtained. This only brings forward more noise.
- The highlights need to be compressed to prevent the topmost stop to be pushed into the signal limit (clipping), which would cut off the increased detail. The compression causes a reduction of the contrast in the upper stops/signal levels.



Exposure of the camera sensor and resulting dynamic range distribution by example of different D-21 characteristic curves (EI 100 – 800).

The dynamic range distribution does not shift for negative film stock of different sensitivity. It always offers the same amount of stops above and below middle grey. Adjusting the sensitivity of an electronic camera therefore rather can be compared to raising the levels of underexposed film stock during telecine/colour-timing/grading.

As amplifying the sensor's signal affects visible noise in the image, the following limits apply:

- The minimum sensitivity, or base sensitivity, is based on a signal with no amplification. Some cameras also offer signal dampening, causing a dynamic range shift in the opposite direction.
- The maximum sensitivity produces an image with maximum acceptable noise levels.

Depending on what is defined to be the maximum acceptable noise level, one and the same camera could achieve quite different maximum sensitivity values. For motion picture work, we stick to the international standard ISO 12232, which defines the following limits:

- An image with a signal to noise ratio (SNR) of at least 40:1 is considered to be excellent (first excellent image).
- An image with an SNR of 10:1 is considered to be just about acceptable (first acceptable image).
  An amplification causing this amount of noise is recommended as maximum value for the sensitivity rating.

Note: The mentioned ISO standard is intended for measurement of digital still picture cameras. As noise is perceived differently in moving and still pictures, depending on its appearance (chaotic or in a fixed pattern), this standard can only serve as a guideline. In the end, the user will decide how much noise is considered to be acceptable for the application at hand.

ARRI generally recommends selecting the EI characteristic curve or amplification depending on the intended dynamic range distribution, while proper exposure should be ensured by sufficient lighting.

3 Output

In order to describe the output of a digital camera, we refer to characteristics, such as aspect ratio, resolution, image frequency (fps), colour sampling scheme, output range and contrast characteristic or characteristic curve.

3.1 Aspect Ratio and Image Resolution

HD video has a native aspect ratio of 16:9 (1.78:1) and a resolution of either 1920x1080 or 1280x720 square pixels. The resolution is denoted by the number of lines i.e. 720 or 1080. When an HD image is cropped in height, e.g. to an 1.85:1 format, more complicated descriptive terms have to be used (e.g. 1080 1.85:1 crop), Sony and Panasonic use non-square pixels for the formats HDCAM, HDV or DVCPRO HD, whereas the use of 1440 or 1280 “rectangular” pixels in width conduces the reduction of required bandwidth, while the image is then scaled back to correct width for output.

When talking about the digital formats 2K and 4K, we refer to the image container’s width (pixel) instead of its height. 2K generally stands for 2048 pixels and 4K for 4096 pixels in width. As with film, the digital formats also have different sub-formats. According to the current standard published by the Digital Cinema Initiative (DCI), a 2K image has to take up either width or height of a container with 2048x1080 pixels. For 4K images the container size is 4096x2160 pixels. The image can have any aspect ratio inside of this container. A DC 4K image with an aspect ratio of 2.39:1 thus would have 4096x1716 pixels. Joining the Digital Cinema formats are the Full Aperture formats and the Academy formats (see Table).

Format	Resolution	Aspect ratio
Full Aperture 4K	4096x3112	1.33:1
Academy 4K	3656x2664	1.37:1
Digital cinema 4K	4096x2160	1.9:1
Full Aperture 2K	2048x1556	1.37:1
Academy 2K	1828x1332	1.37:1
Digital cinema 2K	2048x1080	1.9:1

Source: Wikipedia

3.2 Scanning Methods (p, i, PsF)

3.2.1 Progressive Scan (p)

After each exposure, the full frame is captured/output from the sensor. This scanning method provides the highest resolution per frame and is comparable to the way images are captured on film. It delivers good material for keying or masking in postproduction. The only problem is fast motion or fast panning, which easily results in jittering/juddering images. Shooting with higher frame rates eliminates this effect.

Progressive material is denoted with the letter “p”, attached to the image format:

- 720p meaning progressive HD material with 1280 x 720 resolution
- 24p for progressive Material at 24 frames per second
- 1080/24p for progressive HD material with 1920x1080 resolution at 24 fps.

3.2.2 Interlaced scan (i)

After each exposure, odd (1,3,5…) or even (2,4,6…) lines of the full frame are alternately captured/output from the sensor, as so called fields. Two sequentially output fields make up one frame. Fields are output at twice the frequency of frames from progressive capture. This principle bases on the afterglow of CRT phosphors and the characteristics of human vision, which results in two fields being perceived as a continuous image once the frequency is high enough. Due to the higher frequency of the fields, this scanning method provides a smooth motion representation. Fast horizontal movement or quick pans, however, cause jagged vertical edges in the image, as the two fields composing a frame are captured one after another. This reduces the quality of keying or masking in postproduction. Interlaced material usually has to be de-interlaced (combining fields to frames), before it can be worked on using today’s post processes.

Interlaced material is denoted with the letter “i”, attached to the image format:

- 1080i meaning interlaced HD material with 1920x1080 resolution
- 25i for interlaced material at 25 frames, i.e. 50 fields per second
- 1080/25i as a combination of the above.

Some people prefer relating to fields- rather than frames per second, e.g. 50i or 60i. As current cameras also offer progressive capture of 50 or 60 fps, the declaration of fields per second can be rather confusing.

3.2.3 Progressive segmented frame (PsF)

This is not a scanning method used for capturing, but a transmission method. Basically, a progressively captured frame is segmented into two fields, which are then transmitted similar to an interlaced signal. This enables devices that usually work with interlaced material to also work with progressive material. The notable difference to an interlaced scan signal is that both fields originate from the same frame and have not been captured sequentially. Therefore, transmitting progressive scan material using PsF means no loss in image quality. Motion representation and resolution is exactly the same as in progressive scan material.

PsF material is denoted with the letters “psf” (or PsF), attached to the image format:

- 1080/25psf for progressive HD material with 1920x1080 resolution at 25 fps, transmitted as PsF.

3.3 Frames per Second (fps)

The HD standards define signal transmission at fixed image frequencies between 23.976 and 60 fps. There are different techniques that allow frame rates in between the fixed frequencies and higher frequencies.

To allow variable speeds (ramps) over standard HD-SDI interfaces, a camera is set to output a continuous signal at a supported frame rate, such as 50 fps while the camera head is ramping from e.g. 20 to 50 fps. Since, at the beginning of the ramp, the camera head delivers less than half of the required frames, the 20 fps are simply duplicated as required. To get rid of duplicate frames they are ‘flagged’ (Variframe flag) in the HD signal. Some recorders only write unflagged frames to the recording medium, others record all frames including the flags and skip the duplicates during playback or they leave duplicate removal for postproduction.

Higher frame rates are mostly realized using a temporary high speed buffer connected via data interface, which is then played back at a standard speed to the attached recorder.

HD standard frame rate restrictions, of course, do not apply if a data interface is used between camera and recorder. Images are simply stored at whatever frequency the camera delivers them. The limiting factors are the maximum data rate of the interface or the image processing speed in the camera (also see **Windowing**).

3.4 Colour Signals and Chroma (Sub-)Sampling

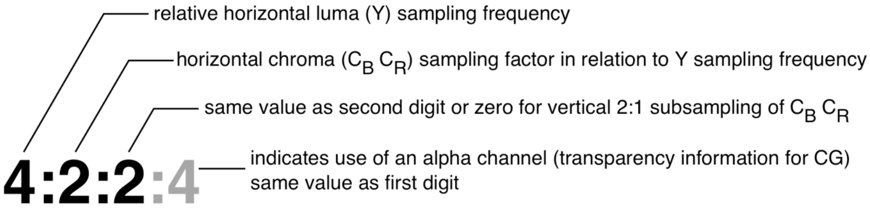
RGB, YUV, 4:4:4 and 4:2:2 are terms often used for digital video. While RGB and YCbCr (correct denotation for the digital component signal often called YUV) represent different forms of colour signals, 4:4:4 and 4:2:2 are a denotation for chroma sampling.

3.4.1 Colour Signals

- **RGB** means the analogue or digital form of signals for the three colour components red, green and blue. RGB signals deliver very high quality images providing more possibilities for postproduction. At the same time RGB signals require high bandwidth and storage space.
- **YCbCr** is called a component signal and consists of a luma channel (luminance) and two chroma channels (colour difference channels), which originate from the RGB signal. Colour difference channels are calculated from a weighted difference of blue- or red signal and the luma signal. YCbCr signals require less bandwidth than RGB signals and thus are more common for economical reasons.

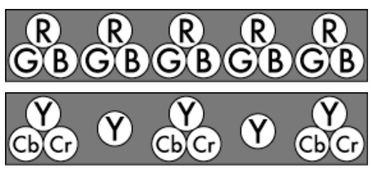
3.4.2 Chroma (Sub-)Sampling

The bandwitdh required by component signals can be reduced by means of so-called chroma subsampling. The idea behind this technique utilizes human vision. A reduction of colour information has lesser influence on the impression of an image, as long as full luminance information is available. For this reason, the chroma channels of a component signal may be sampled with a lower frequency than the luma channel – i.e. subsampling.



The notation of three or sometimes four digits separated by colons represents the relationship of subsampling.

- The first digit represents the relative horizontal sampling frequency (simply speaking: the amount of information per line) of the luma channel. “4” is just a historical reference to standard definition television and provides a base value.
- The second digit represents the factor for horizontal subsampling (i.e. the relation of the amount of information per line) of the chroma channels.
- The third digit represents the factor for vertical subsampling (i.e. the relation of the amount of information for two adjacent lines) of the chroma channels. In professional video, this digit is identical with the second. Only consumer products (e.g. HDV) make use of this kind of subsampling.
- If a fourth digit is given, it indicates the presence of an alpha channel. An alpha channel contains transparency/opacity information used in postproduction (e.g. keying) and is provided for each pixel without subsampling.



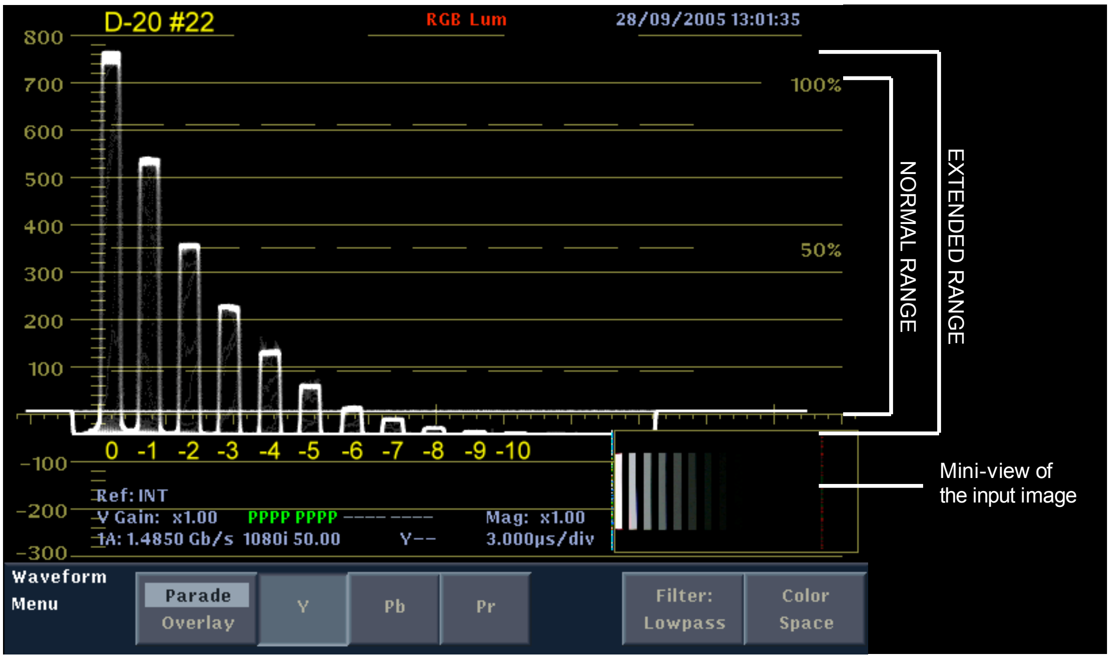
In a 4:4:4 signal, each pixel has full luminance and chrominance information. As the luma component of an RGB signal is a part of all colour channels, an RGB signal is always 4:4:4. In a 4:2:2 signal each pixel contains full luminance information, but only every other pixel per line also contains chrominance information. An RGB signal has to be transformed into a colour difference signal, such as YCbCr, before it can be subsampled. YCbCr signals can be delivered as a signal using 4:2:2 subsampling, or as a 4:4:4 signal without subsampling.

3.5 Signal Range

The luminance levels in a scene are represented as amplitude levels in a recorded video signal. More exposure outputs higher signal levels. Overexposure eventually results in white clipping – image areas with a signal level at the upper limit with no definition left. In analogue video, the signal amplitude is expressed as voltage or IRE units. In digital video, a set of discrete code values is used to represent the continuous luminance range (see **Analogue/Digital Conversion**).

The signal range defines the upper and lower limit of the video signal, which represents the luminance values of black and pure white in the scene:

- Normal range (also called legal or safe range) signals use code values from 64 to 940 (0 – 100%).
  This signal range is HD broadcast compliant, thus is ideal for HDTV production.
- Extended range (also called full range) signals use code values from 4 to 1019 (-10 – 110%).
  This signal range is not HD broadcast compliant, but provides finer tonal gradation for editing in postproduction. Using it for broadcast requires the signal range to be converted to normal range prior to delivery.



Both normal and extended range describe the same overall luminance-range in a scene. Therefore, the selection of normal or extended range has no influence on the dynamic range of a camera. Instead, extended range offers more signal range, i.e. more code values between black and white (for each color channel). This delivers a finer tonal gradation of the overall luminance-range and thus provides better color separation for color keying or color grading in post production.

The decision for extended or normal range should base on the type of production, its intended distribution and the workflow that shall be employed. Examples:

- For an HDTV drama series, shooting 1080 4:2:2 YCbCr normal range will allow fast and efficient finishing.
- A film shoot which uses an HD camera for green screen shots may opt for 1080 4:4:4 RGB extended range to get the best color resolution and the finest tonal separation for more freedom in postproduction.

On regular monitors, signals at/below 0% are shown as black and signals at/above 100% are shown as full white. Unless a monitor has been adapted to extended range signals, some of the detail in the blacks and in the highlights are simply not displayed. This does not affect the recorded signal.

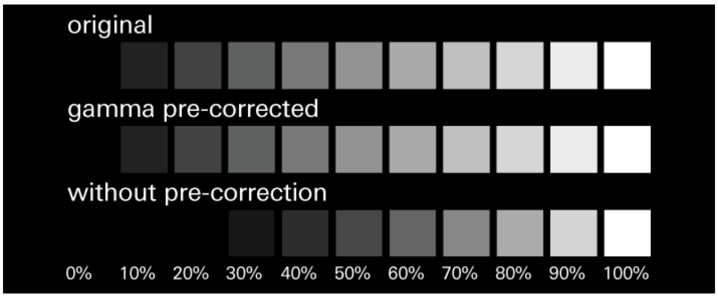
3.6 Transfer Function

The relation of an input to the output of a system can be described as transfer characteristic or transfer function.

As explained in the beginning of this document, the photocells on an imager turn incident light into electric charge that is proportional to the amount of incident light. Therefore, the transfer function of this system describes the relation of scene contrast to output signal, which would be linear.
For a display device, such as a CRT monitor, the transfer function describes the relation between input signal and output luminance (on the screen), which usually is nonlinear and better known as 'gamma'.

3.6.1 Contrast Characteristic

In order to achieve a visually "pleasing"/correct image on a monitor, the input signal has to be adapted to its nonlinearity, which is also called gamma correction. Without adaption, a linear image would appear non-uniform



(illustration below).

This the original transfer function of a camera system, combined with any kind of adaption or gamma pre-correction can be described as the contrast characteristic of a camera.

Providing images with a contrast characteristic that includes gamma pre-correction is essential for cameras that serve the purpose of delivering images that are directly suitable for viewing on monitors or for broadcasting. For digital motion picture cameras, however, the main goal is to transmit the maximum extent of image information. These cameras often provide contrast characteristics for direct viewing on monitors, but in addition they offer contrast characteristics that are optimized to deliver an ideal output of the captured dynamic range, rather than a pleasantly looking image. The decision, which curve is the best, needs to be decided based on the application.

On the ARRIFLEX D-21, the user can select from different characteristic curves for different purposes:

EI 100 – 800

These characteristic curves can be considered "what you see is what you get" settings. Their purpose is to directly deliver a visually correct representation of the camera image on a monitor while retaining as much image information as possible.

The EI (exposure index) value represents the corresponding ASA/ISO sensitivity of negative stock. If the operator selects e.g. EI 400, the exposure meter needs to be set to 400 ASA/ISO (and set fps and shutter angle) to get the correct lens aperture. Using the label ASA or ISO would be wrong, as a digital camera has a fixed sensitivity that can only be adjusted by adding gain (also see: **Adjusting Sensitivity**).

LOG C

This characteristic curve provides a signal output according to Cineon format specifications at 200 ASA/ISO sensitivity. It provides the same tonal step representation that is known from scanned film negative. Each stop of exposure is mapped in equal intervals of digital code values. Therefore, Log C is ideal for processing in a DI/log workflow, just like scanned footage.

Logarithmic material from a film scan or from the D-21 looks very grey and flat, when directly viewed on a monitor. To get a visually correct output on HD monitors the log images need to be adjusted by means of a so-called preview or 1D LUT (look up table). A preview LUT provides gamma pre-correction and an optional sensitivity adjustment, as the camera does not offer direct adjustment of the sensitivity setting in Log C (also see: **Look Up Table**).

LOG F

This characteristic curve provides a signal according to FilmStream specifications at 200 ASA/ISO sensitivity. It was implemented for customers using HD workflows adapted to FilmStream (Viper camera) signals. LOG F delivers an ISO equivalent/EI of 200. As LOG F also does not allow to directly adjust the sensitivity, the same procedure applies as for the LOG C curve.

FilmStream, Panalog and S-log

Manufacturers such as Thomson Grass Valley, Sony and Panavision all provide log-derivate curves to optimize their camera's output. Thomson's curve is called FilmStream (as mentioned above), Panavision's is called Panalog and Sony's is called S-log. With the F35, Sony also offers the possibility of adjusting the camera's sensitivity when shooting with S-log. While this is basically in accordance with the adjustment described under **EI 100 – 800**, the necessity of employing preview LUTs for visually correct monitoring output pretty much defies the purpose.

Custom Contrast Characteristics

Next to the S-log curve, Sony offers the users an individual adjustment of the characteristic curve. This can be done on a laptop and then transferred to the camera, or adjusted directly on the camera. While this possibility is often praised, it bears the risk of changing the output to a degree where it cannot be corrected in postproduction anymore, which means having to go back and shoot again.

Look Management

Probably the most innovative solution is look-management via metadata. The captured material is always captured with the optimized characteristic curve. The look, which is created by the DP directly on the camera or e.g. on his laptop using reference frames, is only displayed in the camera preview output by means of a LUT. The look-metadata is recorded together with the images (or embedded in the images) and can be read out in postproduction. This metadata delivers a first and very detailed information about the Look that is intended by the DP and can be used to reproduce the look using postproduction tools. The D-21, unfortunately, does not allow transmission of metadata along with the images or the use of custom looks (LUTs) on a preview output. However, recorders e.g. from S.two and Codex provide preview LUTs for the monitoring output of the recorded signal. These LUTs and a record of which one is used for which scene are stored along with the recorded material.

3.7 Look Up Table (1D & 3D LUT)

A 1D LUT works like a RGB contrast/brightness adjustment with varying intensity across the contrast range. A 3D LUT, in addition, allows making saturation adjustments. LUTs can be employed to adjust e.g. log material for viewing on a monitor or to create a preview of the intended look of a scene. Depending on the equipment in use, a 1D or 3D LUT can be applied:

- at the monitoring output of a recorder (e.g. Sony SRW-1/SRPC-1 when in 4:4:4 mode, Codex, S.two).
- in a monitor itself (e.g. from Cinetal or Eizo).
- in a box sitting in between recorder and monitor (e.g. from Cinetal, Thomson, Kodak)

ARRI provides a set of standard preview LUTs for the D-21 serving the purpose of:

- applying tonal adjustments to allow visually correct HD monitoring of the camera's Log C output.
- applying a sensitivity adjustment for Log C exposure at EI 100 – 800.

These preview LUTs are therefore selected on a shot-by-shotbasis, depending on the ASA rating a shot was exposed for. If desired, the intended look can also be incorporated in these LUTs (e.g a LUT for exteriors, interiors, day or night scenes). To do this, the ARRI preview LUTs need to be loaded into a grading tool (e.g. IRIDAS Speed Grade) and adjusted accordingly. The use of too many LUTs, however, is not recommended, as it means a tremendous administrative effort and sometimes is less satisfying than it sounds. A slight change of lighting during shooting or even a change of viewing conditions during review (in a tent during the day or the hotel suite at nighttime) can cause the LUT to "not work" for a shot. LUTs are no full-fledged substitute for a simple best-light correction.

3.8 HD Video and ARRIRAW

The decision for a certain production/shooting format should be based on the type of production, distribution, and of course, the budget.

3.8.1 4:2:2/4:4:4 HD Video

HD video signals are subject to a number of restrictions, such as a limited bit depth, 1920 x 1080 pixel resolution and specified signal transmission interfaces. This also guarantees for an output that is fully HD compliant without particularities.

- Using HD 4:2:2 lin as production format enables quick and efficient production and therefore represents the ideal option for productions with distribution on (HD) television that do not require complex postproduction.
- HD 4:4:4 lin is recommended for productions that involve a lot of post production/special effects work with primary distribution on (HD) television. It provides images with full color resolution that are ideal for extensive postproduction work, such as delicate color/chroma keying for visual effects work and/or dramatic color timing.
- HD 4:4:4 log, in addition to the above allows the application of DI based workflows and log (film) grading, which is especially advantageous when mixing digitally shot material with scanned film footage.

3.8.2 4:2:2 Mscope

Mscope is a unique feature of the ARRIFLEX D-21. Because of its Super 35-sized 4:3 sensor, the D-21 can be used with anamorphic camera lenses that project an optically, 2x horizontally squeezed widescreen image onto the 4:3 (1.33:1) imaging area, just as on 35 mm motion picture cameras. The result is an oversized HD image with approximately 80% more scanning lines than an equivalent widescreen images derived (cropped) from 16:9 HD material shot with a spherical lens.

To output this oversized image it is split into two HD 4:2:2 streams (dual stream). Each HD stream contains a 1920 x 720 letterbox image in 2.66:1 aspect ratio (180 blank lines on top and bottom). Stream A contains all odd lines (line count starts from 1), stream B all even lines of the original image. Recording in Mscope requires a system that is compatible with HD 4:2:2 dual stream/dual camera input. This type of signal is commonly used for stereoscopic capture with two camera heads.

Image, as recorded by the D-21 with an anamorphic lens

A simple 2:1 stretch and a slight crop from the sides in post results in a Cinemascope format image.

Currently only devices from Codex Digital offer automated recombining of both Mscope streams. Without an automated recombining, both streams are transferred individually. As both streams contain a de-squeezed image, either one can directly be used for editorial/offline without the need for recombining. After the offline-edit, only the material required for further post production and finishing needs to be re-combined to save on processing time.

Re-combining both Mscope streams basically requires two image operations:

1. Crop top and bottom 180 blank lines from both streams.
2. Interleave (de-interlace) both cropped streams beginning with stream A line 1, stream B line 1, A2, B2, A3, and so on, to create the original 1920 x 1440 image.

As the full frame image size (1920 x 1440) is larger than a standard HD image (1920 x 1080), 2K or resolution independent editing tools are required. There are different tools that allow recombining Mscope material as part of their feature set. Which of them is most suitable will depend on the existing infrastructure of the post facility. Examples are the Quantel DI workflow system, IRIDAS FrameCycler and SpeedGrade, Apple Shake (discontinued) or the Open Source Software ImageMagick (this list is not exhaustive).

Similar to the HD 4:2:2 lin format, Mscope HD 4:2:2 lin represents the best option for productions that want the real anamorphic widescreen look for distribution on (HD) television, but which do not require complex postproduction.

3.8.3 ARRIRAW

ARRIRAW is another unique feature of the ARRIFLEX D-21. ARRIRAW provides RGB color space output of the native sensor CFA pixel grid, in 2880 x 2160 (1.33:1 AR) up to 25 fps or 2880 x 1620 (1.78:1 AR) up to 30 fps, providing alias-free 2K resolution after de-bayering. This data contains the maximum quantization depth of 12 bit, maximum color information, maximum resolution, and allows framing for all image aspect ratios up to 1.33:1, including Cinemascope, of course. With an extended set of image processing options, ARRIRAW offers high quality 2K output for mastering in a DI workflow.

ARRIRAW material can only be recorded with ARRIRAW T-Link certified recording systems (currently T-Link is supported by S.two, Codex, Keisoku Giken, Microm) or recorders offering uncompressed recording and output of RGBA HD-SDI signals. The certified recorders generate a live HD preview from the original 4:3 (1.33:1) or 16:9 (1.78:1) raw data frame. If anamorphic lenses are used, the preview can also be de-squeezed for the live output.

Since ARRIRAW is raw Bayer data, the material needs to be processed into RGB images. As editorial mainly uses lower resolution material, it can be captured from the recorder's or playback station's live preview output or it can be rendered by processing tools, such as Pomfort Silverstack, tools from IRIDAS or MetaFuze. To deliver files for effects, conforming and grading, ARRIRAW material is processed using the ARRIRAW Converter (Image Booster), which provides EDL based processing to e.g. dpx image files. All parameters that can be set in for e.g. HD output in the camera can also be set in this processing tool. Next to ARRI's free processing software, manufacturers of postproduction tools are adding features to enable native ARRIRAW editing (Gluetoos for Final Cut Pro), so the processing can be shifted towards the end of the postproduction chain.