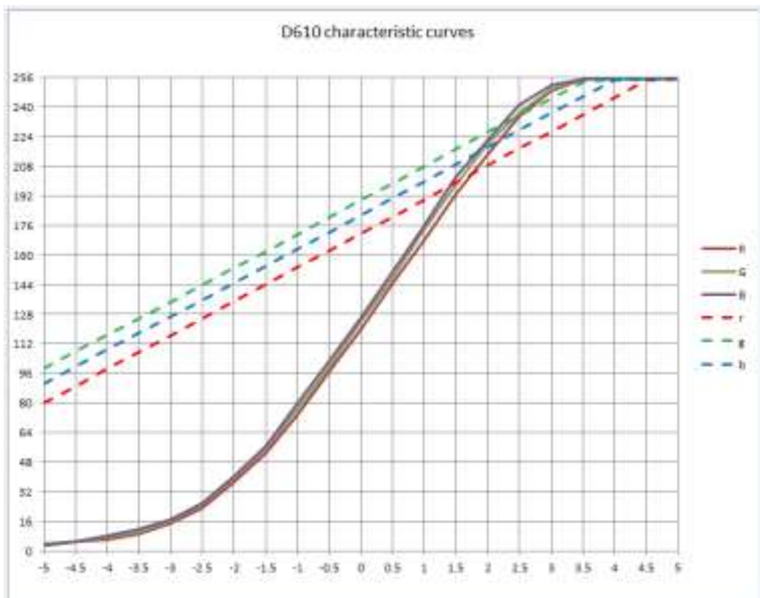
Digital Image Noise

A digital camera records an image by capturing the light in a scene in a sensor as a charge and then converting that to digital information (raw data). The raw data might be saved in the camera in a raw file or used to create an image to save in the camera (usually a JPEG) or both.

This plot shows the default relationship between the raw values (dashed lines) and JPEG values for a D610.



1

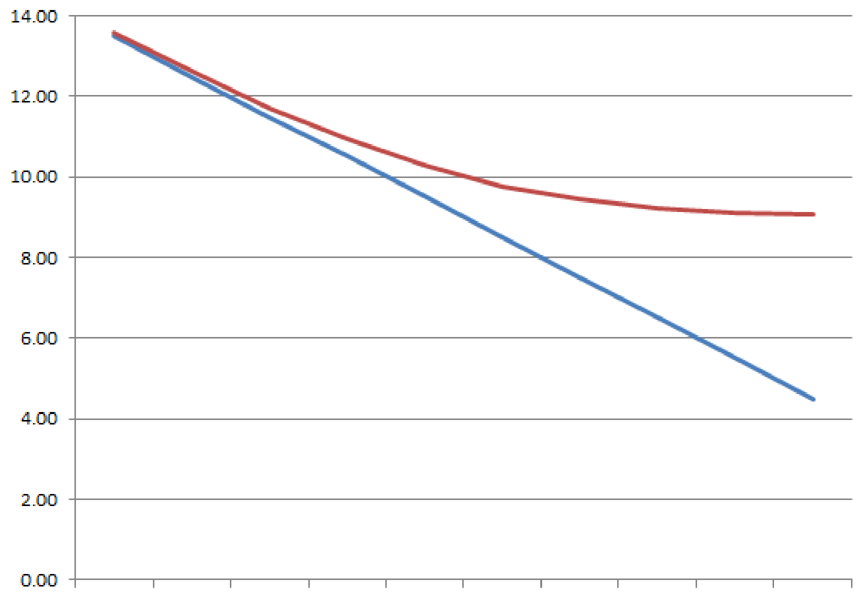
The vertical scale represents the 8-bit JPEG range from 0 to 255. The horizontal scale is the number of stops below and above middle gray (EC). The 14-bit raw values (dashed lines) are linear, scaled to fit the JPEG range by multiplying the 14-bit raw values by 256/16384.

The EC value of 0 for middle gray is at the center of the middle zone that ranges from ½ stop below to ½ stop above middle gray. The upper limit for most raw files is close to 3½ stops above middle gray which ranges from 2½ to 3½, centered on 3. Most cameras will show highlight warnings around that level.

The raw file is composed of straight lines and the JPEG as S-shaped curves. Other color cameras have the same characteristics as the brightness goes from black to white. The only difference is in the location of the JPEG middle gray.

Starting at the brightest raw value, with each full stop drop in brightness the raw value drops by a half in an arithmetic progression – 16384, 8192, 4096, 2048, 1024, etc. This sequence can be expressed as a logarithmic progression in powers of 2 as 14, 13, 12, 11, 10, etc. The sequence is directly proportional to the number of photons reaching the sensor. Each time the shutter speed is doubled (exposure time is cut in half) half as many photons reach the sensor. The same thing happens with the aperture where each smaller full stop lets half as many photons reach the sensor in a fixed period of time.

Although a 14-bit raw file can theoretically represent a full 14 stops of raw values, the dark end of the raw sequence (1, 2, 4, 8, 16, 32, etc.) is not very useful because they are integers, not floating point values. The numbers themselves cannot represent the actual brightness values as precisely as the charge that is recorded on the sensor, a floating point number. To overcome this, a fixed value (the black level) is added to the floating point number before it is converted to an integer. The black level integer might be about 512 for Sony, 1024 for Nikon or Fuji, 2048 for Canon, etc. The sequence of values stored in the raw file will no longer be a straight line but its linearity will be restored when the black level is added back.



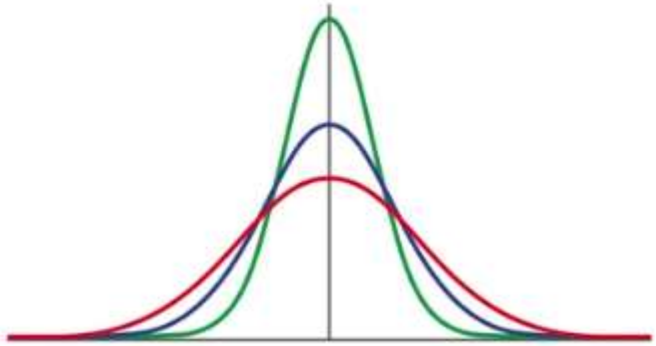
The red line represents the sequence of values (not linear) actually recorded in the raw file that represents the blue linear values that were the actual charge on the sensor. This mathematical trick allows us to record very small values using three or four decimal digits of precision rather than possibly one or two. When subtracting the black level the maximum possible raw value is also reduced but this only affects the brightest full stop of tonality.

The raw values cannot record a charge above the sensor's full well capacity at base ISO. If the ISO is increased above base ISO, the raw file cannot save the results even if the sensor itself has not reached its full well capacity. Both of these limits are reached at about 3½ stops above middle gray.

If not enough light reaches the sensor the value that is saved might actually be a little less than the black level but the net raw values will be 0.

Where does noise come from?

The biggest cause of noise in digital images is the light (signal) itself. Photons arrive at the sensor in a random fashion. When very few of them arrive during the exposure they are not evenly distributed over a portion of the sensor. Rather than recording exactly the same value for each location their values are distributed over a small range of possible values as a normal distribution.

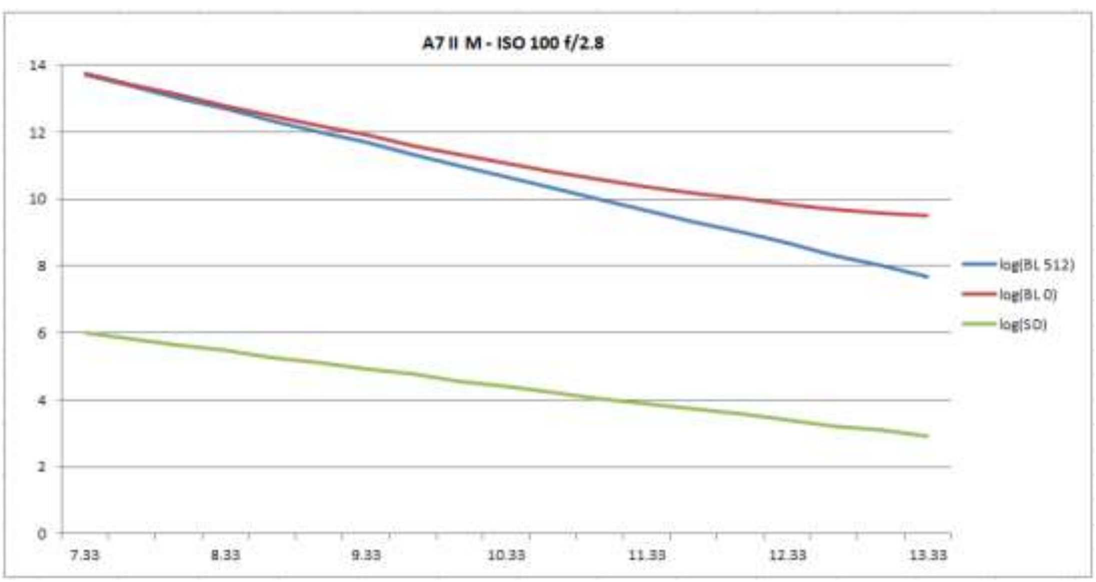


3

When the distribution is very close to the average (green curve) the standard distribution (σ) will be relatively small and the noise will not be easy to see. When it is spread out (red curve) it will be easier to see the noise. At either limit of the exposure range some of the distribution can be truncated and the distribution is no longer normal.

In a B&W sensor the variation will be seen as luminance or shot noise. In a color sensor the noise may also lead to color speckles.

For extremely dark subjects or when there is no light at all, there is still some unavoidable read noise. In this case the noise may overwhelm the signal leaving us with nothing to see. But this is only important for images captured at very high ISO settings.

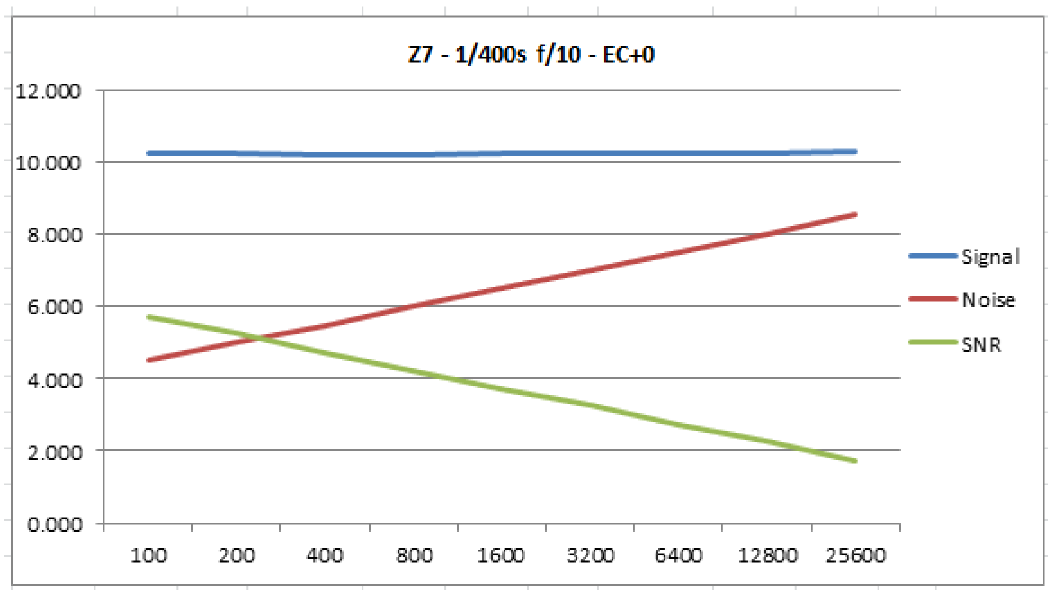


4

This shows the relative importance of the relationship between the signal (red and blue lines) and the noise (green line) for a typical camera at base ISO over a six stop range. At the beginning there is about an 8 stop difference between the signal and the noise. That drops to about a 5 stop difference at the end. Both the signal (blue line) and noise (green) lines are perfectly straight but not parallel. There would be no noise visible in such an image until much later when the separation between signal and noise lines get within about two or three stops apart. By that point the image would be so dark that you could not see the noise even if you could measure it.

At higher ISO settings the lines start out closer together. The noise will become apparent sooner. At extremely high ISO the noise may be visible throughout the image.

ISO is not the source of noise; it's the reduction in exposure or image brightness. If the exposure is changed to keep middle gray at the same raw level:



5

Increasing the ISO results in a lower exposure and makes noise more visible. The signal to noise level (SNR) will drop as the ISO is increased.

There are other things that can make noise more visible.

It will be easier to see the noise if the image is enlarged. If you crop an image you will need to enlarge the result. You will see the noise more easily when viewing it on your monitor at 100%.

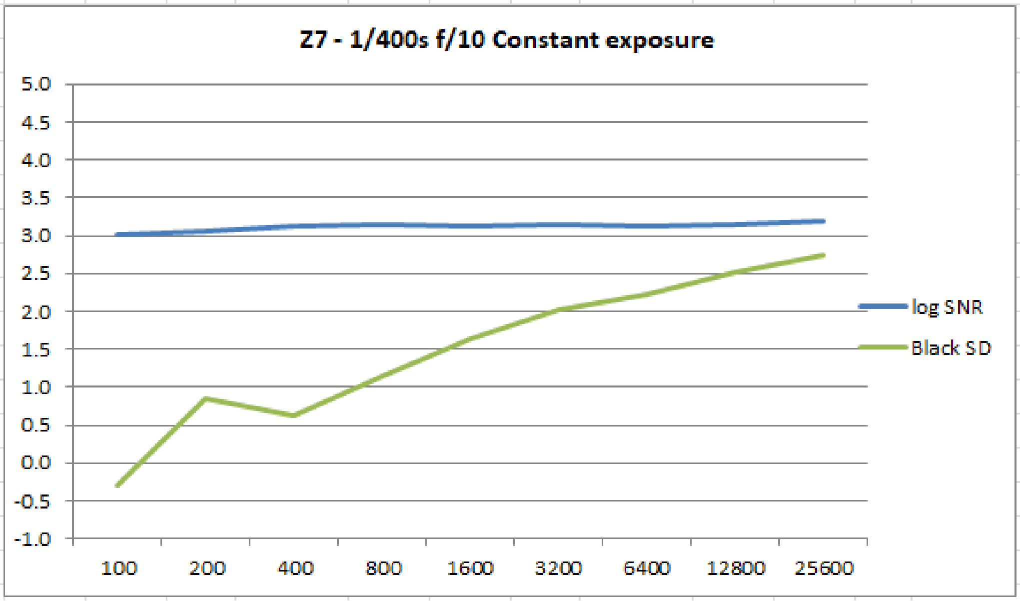
Noise will be easier to see in areas of smooth tonality than where there is detail or texture because you would expect to see a wider σ there.

Noise is harder to observe in images where the contrast is high since the eye is drawn to the lighter parts of the image. The noise might be there in the shadows but the viewer may not notice it or care. Shadow recovery can bring noise with it.

Attempts to remove noise after the fact can be counterproductive because it can reduce sharpness. The photographer needs to proceed with caution keeping in mind how it will ultimately be viewed.

More on Read Noise

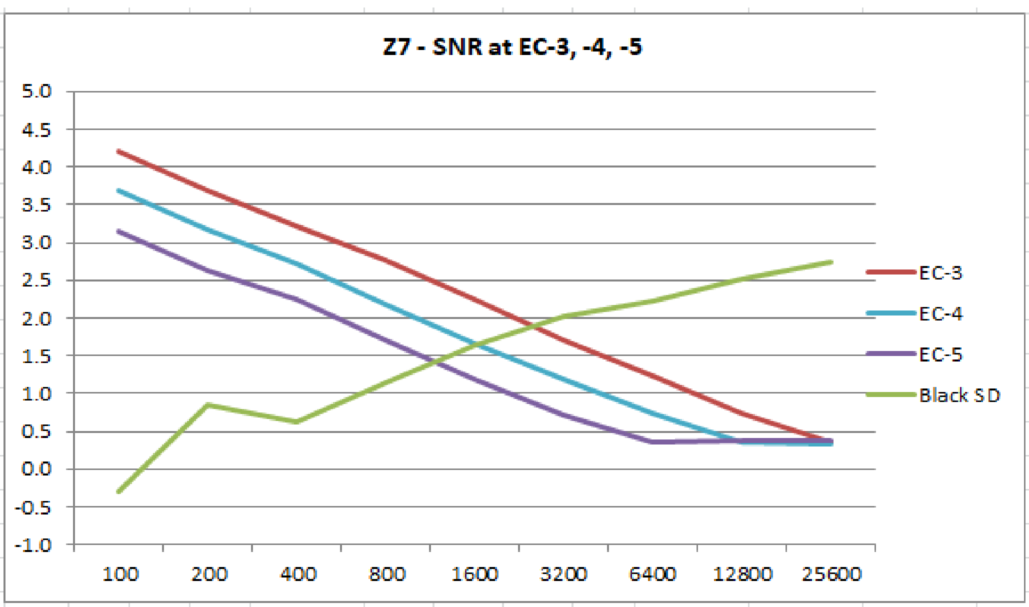
Although read noise plays a part in the calculation of a camera's dynamic range, it actually has no impact on SNR. A series of constant exposure while increasing the ISO results in an almost constant level of SNR:



6

We end up with the same level of visible noise and the SNR is not affected by the black noise. It barely wavers when there is a significant change in read noise in this dual gain sensor.

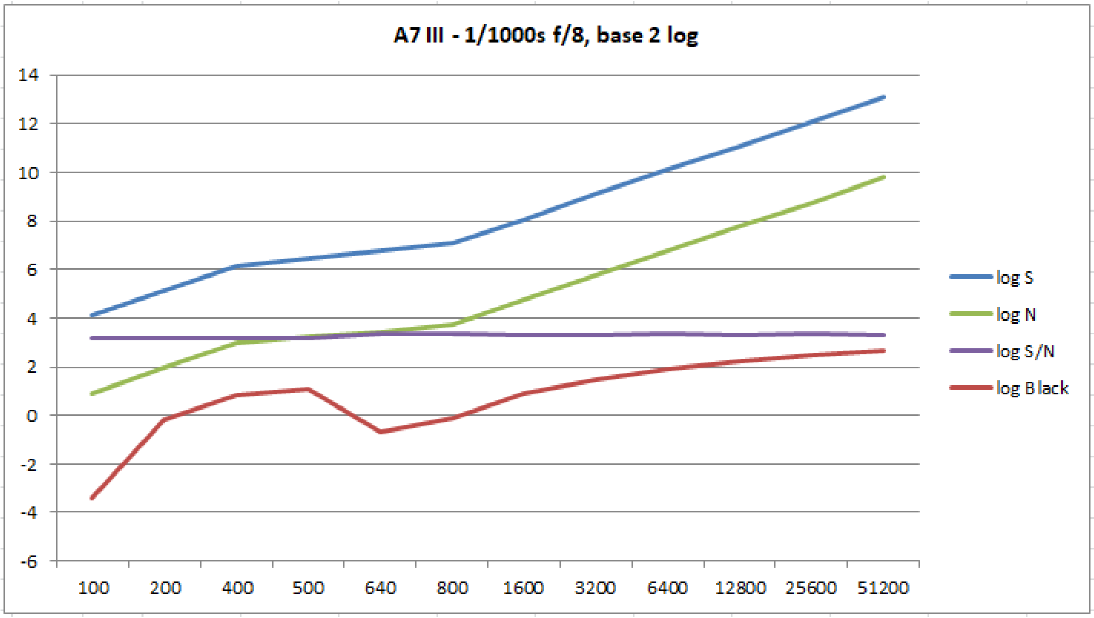
A series of exposures with a constant exposure compensation setting (EC) might be truncated by the camera to keep the result above a certain level.



7

The SNR at EC-4 and -5 have an arbitrary lower limit for SNR in the Z7.

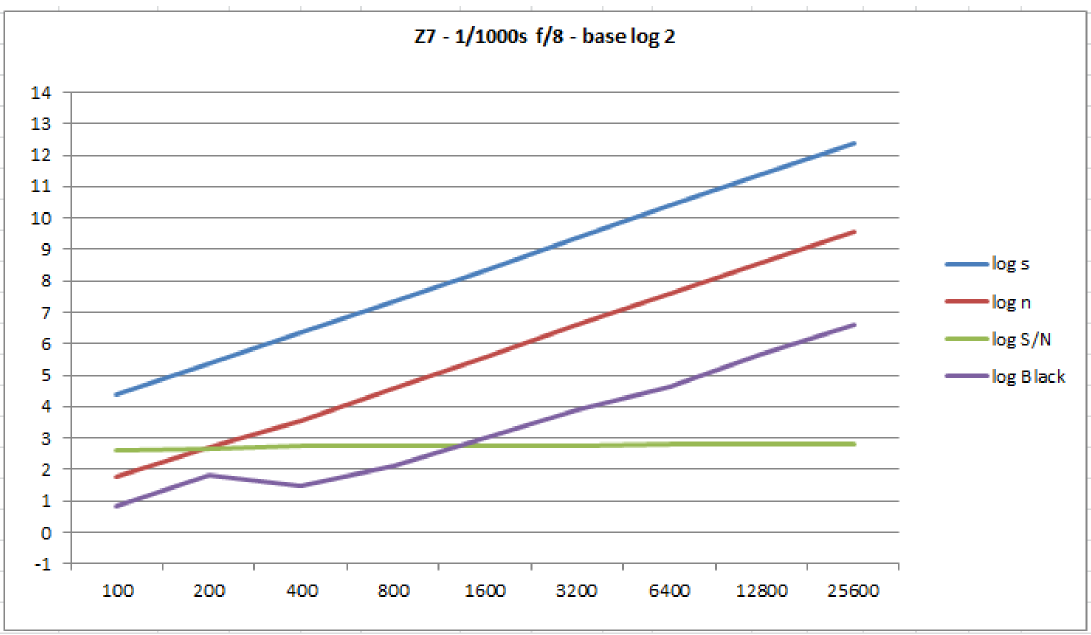
Finally, a plot from an A7 III which has a more pronounced jump in black noise as the camera switches from the low ISO gain to the high ISO gain.



8

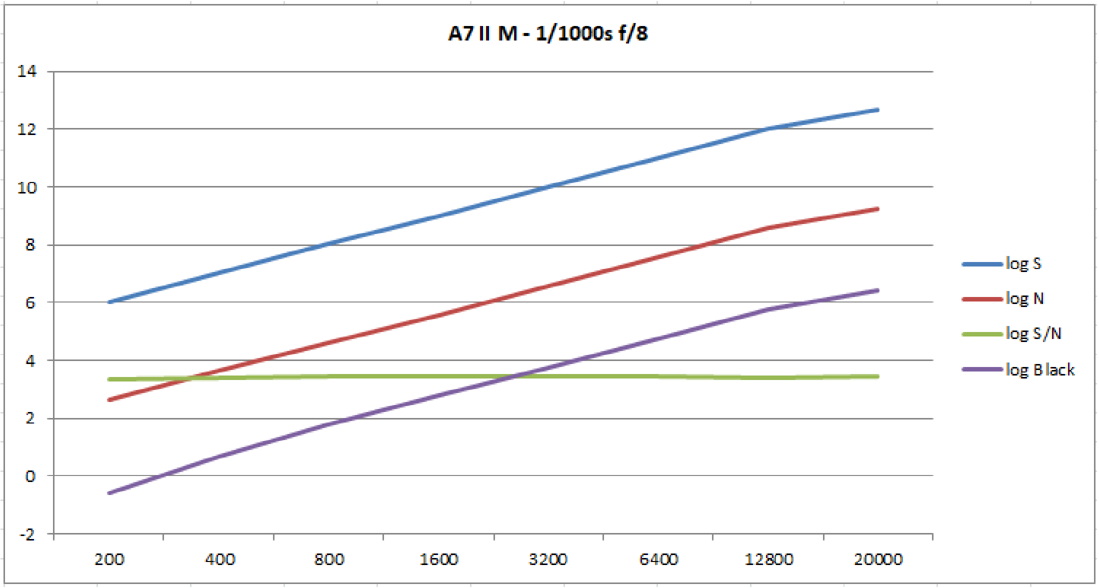
The exposure was the same and the ISO went from 100 to 51200. The dual gain sensor switched between ISO 500 and 640. Although this caused the signal (S) and the noise (N) to change between ISO 400 and 800, the actual ratio of signal to noise (S/N) remained constant. In other words, the noise did not change since the exposure was constant.

The Z7 behaves the same way but the signal and noise do not waver.



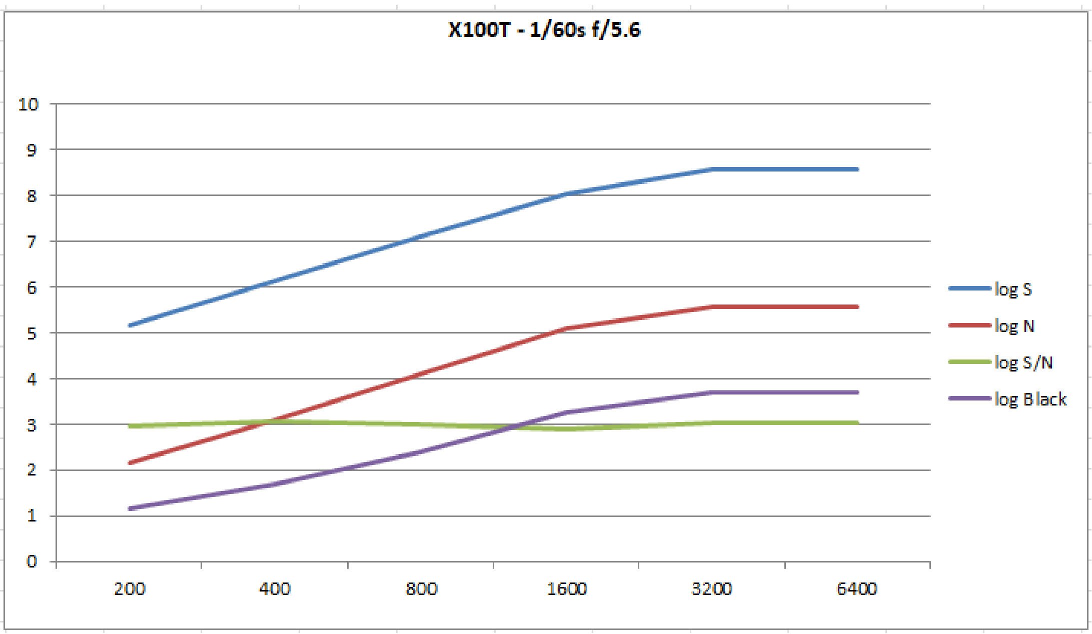
8

The older A7 II with the color filter array gone shows the same S/N behavior but the analog range ends at ISO 20000. Also, the removal of the Bayer array means that base ISO 100 can't be used because the well capacity would overflow.



9

The X100T also behaves the same but since it has an APS-C crop sensor it will end up showing noise sooner because it needs more magnification.

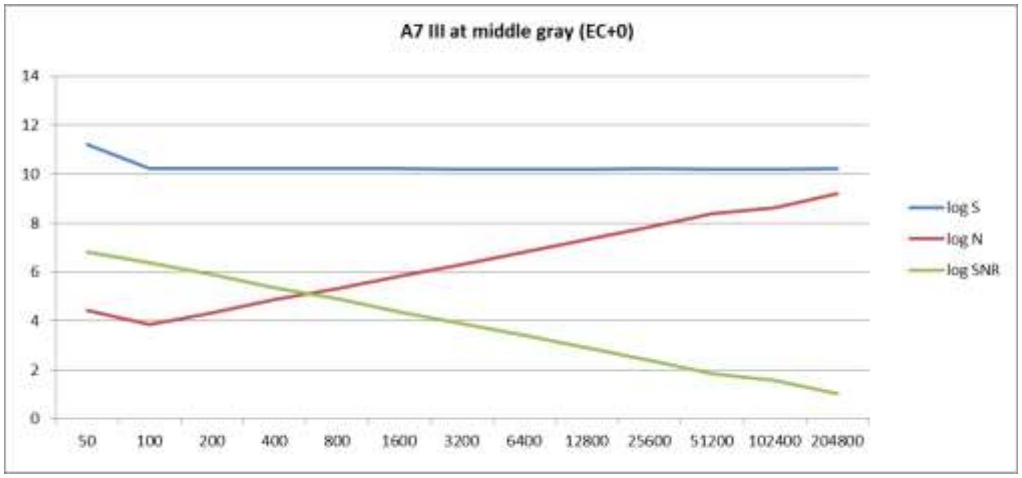


10

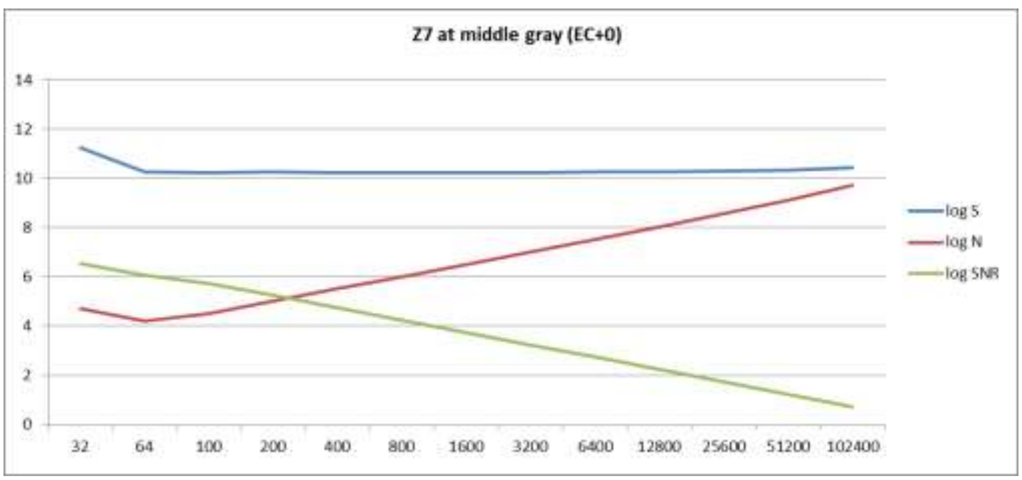
The analog range for the X100T is only from ISO 200 to 1600. Above the analog range the raw data is fudged to provide two stops.

Extended Ranges

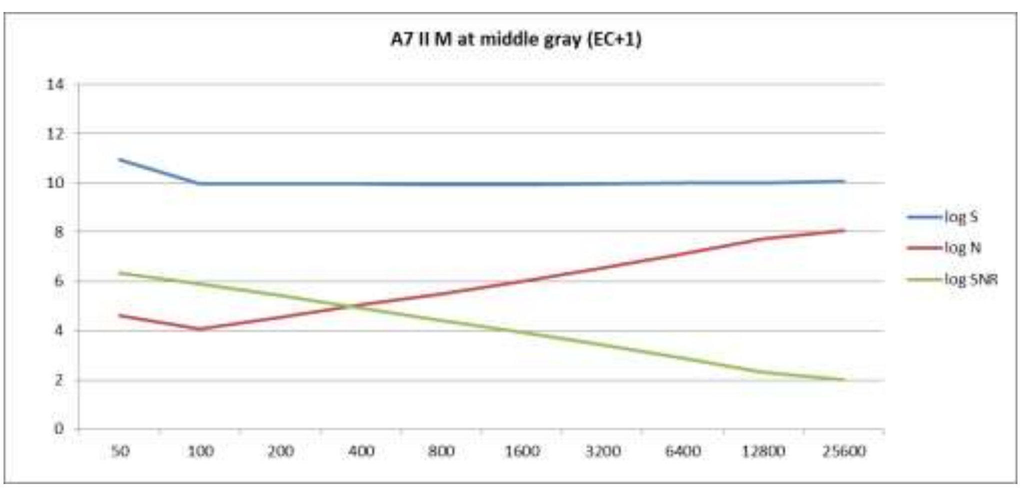
The three cameras that were tested have low and/or high ISO ranges beyond their normal analog ranges. The low ranges cannot actually go below base ISO so they include a warning that the raw highlights might blow out. But the SNR remains linear throughout their entire ranges.



11

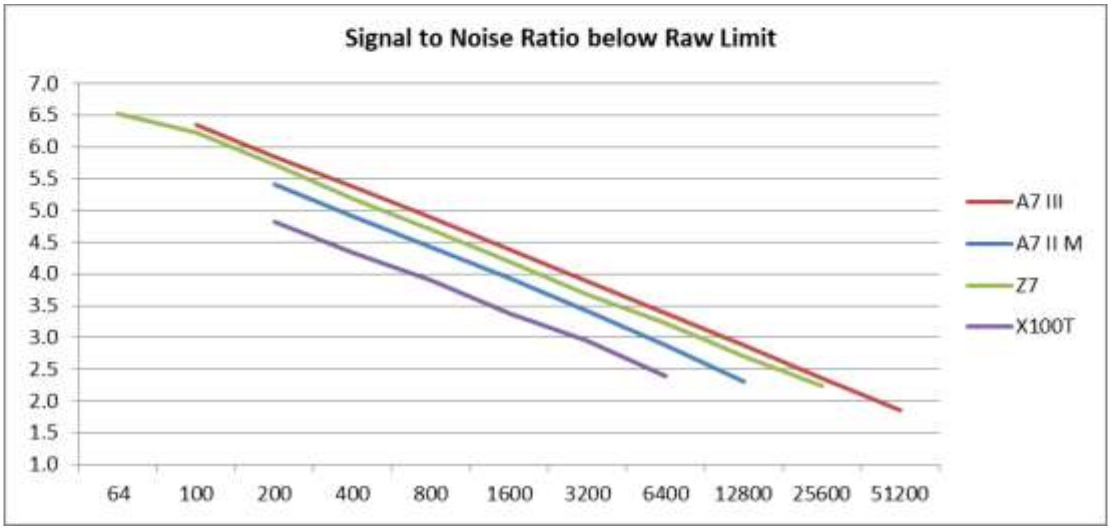


12



13

Here is an objective comparison of the four cameras in this study adjusted for sensor resolution and size.



[14]

Summary

Dynamic range reported by DXOMARK and Photons to Pixels reflects the efforts being made to extend a camera's performance at high ISO settings by reducing the read noise. A second gain range adds about a one stop increase in the upper limit of a camera's ISO.

Shot noise is visible long before the impact of read noise begins to show up. By reducing the read noise with a second gain program there is already so much noise that the difference, although it can be measured, is almost impossible to see.