A paper published by Northlight Images, written by David Goldstein.

All technical queries should be directed to the author, who may be reached at:
savingenergygrowingjobs@yahoo.com
The web version of this article can be found at:
http://www.northlight-images.co.uk/article_pages/guest/physical_limits.html

Physical Limits in Digital Photography

By David B. Goldstein

Abstract

Digital photography has made immense progress over the past 15 years, particularly in *increasing resolution* and *improving low-light performance*. But continuing progress in these aspects of performance is not possible, because cameras are now approaching the limits imposed by the basic nature of light.

This article looks at the physics that explains these limits. It shows how the fact that light has the characteristics of both particles and waves puts caps on resolution and "film speed" that are not far beyond the performance of current high-end dSLRs, and are already impacting point-and shoot cameras.

Resolution is *limited by the wave phenomenon* of diffraction: the current generation of full-frame sensors are diffraction-limited at about f/10 and above, and the best APS sensors at about f/7. More pixels can only improve resolution at wider apertures and if new lenses are up to the task. Pocket cameras are diffraction limited at f/2.4 and above, which is faster than any available lens.

High ISO performance is *limited by the number of photons available per pixel*. ISOs above about 3200 will always result in noise, no matter how much sensors are improved. Actually, the dynamic range of current dSLRs implies that noise in the deepest shadows is limited by photon count even at low ISOs.

These limits change the nature tradeoffs that photographers have been making concerning film or sensor size, depth of field, and sensitivity.

I. Introduction

Quantum mechanics requires that the wave and particle natures of light or matter cannot be seen at the same time. But this only applies for a single observation; if we make large numbers of observations at once, we can see both wave-like and particle-like properties at the same time.

In the last few years, technologies have emerged in consumer markets that allow anyone to see evidence of both the wave and particle nature of light at the same time. The technologies are those of digital photography.

This paper explores how the limitations imposed by quantum mechanics can be seen by examining digital photos taken with state-of-the-art current equipment. It will also show the converse—that the potential for improving the resolution of digital photos and the ability to take them in low light is low—that technology is now pushing the physical limits.

At the heart of a digital camera is a bundling of several million photon detectors arranged in a pattern that allows the user to see:

- The particle nature of light. By looking at the variation in color and light level from pixel to pixel when the sensor is illuminated by a uniform source, the user is seeing the differing number of photon counts; and
- The wave nature of light. This is visible in the form of diffraction by a fuzzy, low contrast image.

Indeed, both of these phenomena are so pronounced that they limit the "film speed" and resolution of digital cameras to levels within a factor of two or so of the performance of current high-quality product offerings. This paper computes, in a simple approximate form, the physical limits to film speed and resolution for popular sizes of cameras, and shows how these limits are binding—or nearly so--at current levels of engineering. The approximate form is appropriate because, as will be demonstrated, the arbitrariness of useful criteria for sharpness and noise make it useless to try for more accuracy of calculation.

II. Film Speed

What is the limit on signal to noise based on fundamental physics?

This calculation looks at the number of photons incident on a pixel and shows that current technology is close to physical limits. Film speeds of ISO 3200 are so close to physical limits that large increases are impossible without compromising other parameters.

The calculation begins by determining how many photons of light are available to be measured by a digital sensor (or film). The energy of a photon of light at λ=500 nm—the mean wavelength of visible light-- is given by $E=h\nu$;
since $h = 6*10^{-34}$, and $\nu=c/\lambda$ or $6*10^{14}$ sec^{-1},

$E=3.6*10^{-19}$ joules

How many photons are incident on a pixel of sensor?

To start with, sunlight is 1.37 kW/m^2 at the top of the atmosphere, or a little less than 1 kW/m^2 at ground level with normal incidence of sunlight. (Normal [90°] incidence provides the brightest possible light.) Of this light, about 45% is visible, the rest being mainly infrared and a little ultraviolet. So the visible light on a square meter of illuminated object is about 400 W. A very bright object will reflect almost all of this light.

But the meaningful criterion for visible noise is how the camera renders the darker parts of the image.

The ratio of the lightest light level that a sensor (or film) will capture to the darkest is called "dynamic range" by photographers, and is expressed as an exponent of 2, called an "f-stop". The dynamic range of the most advanced digital cameras at a high sensitivity—ISO 3200—is about 8 f-stops[1], or a ratio of 2^8. The issue of dynamic range is complex, but this estimate will prove surprisingly robust in calculating the number of photons incident ona pixel in the darkest renderable shadows regardless of the sensitivity of the sensor or film to light, at least with current technology. So the range of light we need to analyze is 1/256 of the 400 W/m^2; or 1.5 W/m^2.

With each photon at energy of $3.6*10^{-19}$ joules, this is a photon density about $4 *10^{18}$ photons per second per square meter.

How many of these get to the pixels? The answer depends on the size of the sensor or film, so I will address different sizes in turn.

Full frame 35mm sensors

Assume a state-of-the-art 35mm camera with 21-24 megapixels[2] (this assumption matters) taking a picture at its highest film speed, ISO 3200 (this assumption would appear superficially to matter, but turns out to be relatively unimportant) with a 50mm lens and the object one meter away, taking a picture at f/16 and 1/3200th of a second, the correct exposure for ISO 3200 (these assumptions don't matter but are made just to make the derivation easier to follow).

The field of view of the lens at 1 m is 720 x 480 mm, or .35 square meter, so the photon density is about $1.5*10^{18}$ photons per second. This photon density represents the darkest shadow capable of being captured with resolution by a professional-grade camera at this "film speed".

Each point in this field of view reflects into a hemispheric solid angle (optimistically; some reflectors will reflect into a full sphere); the amount of light that can be transmitted to the film is the reflected light times the ratio of the solid angle of the lens aperture as seen from a point on the reflector to a hemisphere. A 50mm lens at f/16 has an aperture diameter of 3mm and a radius of 1.5 mm, so the ratio is $\Pi*1.5^2/2\Pi*1000^2$, or 10^{-6}.
So the photon density on the aperture is about $1.5*10^{12}$ photons per second. These photos strike 21-24 megapixels at a rate of $7*10^4$ photons per pixel per second.

---

[1] A typical dynamic range for film is about 7 f-stops (an f-stop is a factor of 2); pocket digital cameras perhaps have a little less. As of spring 2009, the four most advanced digital cameras had dynamic ranges of about 8 stops at ISO 3200 (see dxomark.com). So this paper provides a range based on these limits.

[2] These spec's apply to the Canon 1Ds Mark III and 5D Mark II, both at 21 megapixels, and the Nikon D3X and the Sony Alpha 900, both at 24 megapixels; the pixel size is also about the same for smaller-frame cameras (image size about 15 mm by 22 mm) produced by numerous manufacturers.

For a proper exposure at 1/3200 second, the reception rate is 20 photons per pixel for an exposure. This is small enough for noise (statistical variation in the number of photons registered per pixel) to be evident even with perfect photon detection.

This rate is quite robust to changes in assumption. Had we chosen a different combination of shutter speed and aperture, we still would have gotten the same answer for the proper exposure. While we chose the film speed to see how much a speed of ISO 3200 is pushing hard against the limits of physical laws, it turns out that with all of the professional cameras we examined, the dynamic range is a very strong negative function of the sensitivity.

Thus, as sensitivity increases from ISO 100 or 200 to ISO 3200 (or higher in some cameras), *the dynamic range drops nearly as fast as the sensitivity increases*. Thus as the camera gains 5 f-stops in sensitivity, the dynamic range drops 3 to 4.5 stops[3]. So the increase in sensitivity in the deepest shadows is actually only ½-2 stops. Most of the increase in sensitivity comes from boosting the highlights and midrange: the darkest shadows rendered at low sensitivity are almost as dark (less than factor-of-two difference in one case; about factor-of 4 in the other). The photon reception rate for the darkest renderable shadows is not that different at different "film speed" sensitivities!

This reception rate of about 20 photons per pixel is already a noise risk, since the signal to noise ratio for random events such as photon strikes is √n: this is a S/N ratio of 4.5—not very good. (A S/N of 4.5 assures with high probability that an observed signal is real, but it allows lots of variation from pixel to pixel of the image darkness when the light incident is actually the same. This makes for *grainy, speckled* photos*.)*

And this theoretical number is reduced fourfold by the fact that the calculation above implicitly assumes that we don't have to worry about color. But in fact, we do care about color, and the current technology for digital photography measures only the existence of the photon, not its energy (color). And even color-responsive sensors would only work to reduce noise if the light incident on the pixel was monochromatic, which is not always the case. So the signal to noise calculation above is only valid for black-and-white photography.

Current sensor technology detects color by sending the photons through filters. Each pixel sees light filtered by either a red filter or a blue filter or one of two green filters. Thus the number of photons that make it through the filter channels is at most 1/3 of the number calculated above, or 6 photons.

Even with perfectly transmissive lenses and 100%-effective sensors, the signal to noise ratio for red and blue will be 2.4 (=√6), so color noise will be a big problem in the deep shadows and a significant problem even in the milder shadows. Note that color noise (called "chrominance noise" by photographers) means that we have to be concerned about how nearby pixels rendering a uniform subject will have random variations in color as well as in brightness.

These issues are plainly visible to the photographer. For example, the difference between black-and-white photos and color is evident in lower noise for monochrome. This phenomenon

---

[3] Data on dynamic range of real cameras are taken from dxomark.com.

has been noted in numerous camera reviews in the popular press. The author has also noted this by taking pictures at an effective ISO of 25,000 on a 21 megapixel camera[4] and looking at them on a computer screen. See Figure 1 for an example.

The conclusion then is that it will be nearly impossible to increase film speed without increasing noise dramatically in the deep shadows and significantly in the middle dark shadows. Speed can only come at the expense of reducing the pixel count by making each one bigger. Or at the expense of dynamic range compared to what is expected of quality photography in color.

This latter conclusion appears to be largely a consequence of the underlying physics. If the ability to amplify the signal in the darkest shadows is limited by photon variation, then ***there is a fixed limit to how dark an object can be and still be "visible"/rendered at all.*** Exposure at a given ISO is determined by measuring the light reflected from an 18% reflectance grey card in the bright part of the picture. If we try to make that 18% card brighter and brighter on the photo, but cannot make the darkest shadows any brighter, most or all of the increase in film speed can only be achieved at the bright end of the brightness histogram; the dark end won't change. Thus the ratio of darkest black to whitest white must deteriorate as fast as the bright end is amplified.

What about those other assumptions? First the ones that don't matter. The 50 mm lens doesn't matter because if you doubled the focal length (without changing the sensor size), the aperture at f/16 would be twice as wide and have 4 times the area, but the field of view would be ¼ the area of the reflector, so the photon density on the aperture would be the same. So the lens focal length doesn't matter. Suppose the reflector were 2m away instead of the assumed 1m. Then the field of view of the lens would be twice as big, the area 4 times as big, but the solid angle of the aperture would be ¼ the size. So the number of photons incident on the aperture would be the same.
Next the ones that do. If the camera has ½ the megapixels, then each receives twice the photons and the noise problem is ½ as bad (or more correctly $1/\sqrt{2}$ times as bad). Thus the camera can use twice the ISO speed for the same noise level. The number of photons needed for a good exposure is inversely proportional to film speed, so lowering the film speed to ISO 100 increases the photon count 16 times and improves the signal to noise ratio 4 times. BUT as noted above, with current technology, the dynamic range improves almost as fast as film speed, so the photon count in the darkest shadows is not much higher at low sensitivity. The darkest shadows are darker, but they are just about as noisy.

Smaller sensors for interchangeable lens cameras

For more typical size moderately priced digital interchangeable-lens cameras, the pixel count is typically 10-15 megapixels for a sensor 1.5 to 2 times smaller in linear dimensions. So the typical pixel density is the between the same and 2 times higher, meaning the film speed limit is the same to half as high. Most of these cameras have lower dynamic range than their top-of-the-line competitors, so the signal-to-noise ratios will be slightly better worse than full frame.

[4] This is done by intentionally underexposing 3 stops with the ISO setting at 3200.

Pocket cameras

If the sensor is 1/5 the size of 35mm, typical of point-and-shoot cameras, the lens with the same angle of view has 1/5 the diameter of the assumed 50 mm lens, so the area of the aperture is 1/25 as large. Thus it captures only $1/25^{th}$ the number of photons. Most cameras of that size use 8 to 12 or 15 megapixels, so the number of photons per pixel would be about 1/10 the level calculated above.

This is so troublesome that we can see why pocket cameras may not even offer ISO 1600 and have visible problems with noise at 400 and even lower. (The cameras that offer high ISO (3200 to 6400) do so by combining pixels, effectively meaning that there are about 1/4 as many of them as you would expect, reducing resolution by a factor of about 2.)

In practice, the noise problem isn't quite as much worse than this calculation suggests, but that is because the dynamic range of pocket cameras is much less than for the larger and more expensive alternatives. We can't quantify here how much difference this makes because measured data on dynamic range are not published in photography magazines or manufacturer literature for such products.

Ways Around the Problem

The noise problem is fundamental, in that a pixel of a given size only receives the number of pixels calculated here. But noise can be reduced at the expense of resolution by averaging over adjacent pixels. Most cameras and essentially all computer software allow noise reduction by losing information.

Noise could also be reduced by binning pixels, and some cameras do this already. If four pixels are binned, obviously the combination receives four times the number of photons. If the four are chosen to have each of the four color channels, we don't lose as much information on color as one might think.

Smaller pixels also look less noisy. Imagine a ten by ten array of pixels where one pixel is now. Each pixel will now be exposed to, most likely, one or zero photons in the deep shadows. So on a pixel level, the noise will be extreme. But at the level of a viewer, the tiny pixels will be like a pointillist painting, and the noise will be less bothersome.

One can envision smart software that bins pixels only in the shadows, allowing higher resolution in those parts of the picture that are bright enough to support low noise.

Vision

This discussion is based on the physical properties of light, not on the technology of cameras, so it applies equally to other forms of imaging, including the human eye. And, of course, the eyes of other animals as well.

If noise is a problem at the resolution of a normal lens on a camera, it is equally a problem for vision. So my observation that a digital camera with a f/1.4 lens at $1/30^{th}$ of a second can see more shadow detail than my eye can is not an accident, it is a consequence of the fact that both eye and camera are facing the same photon limits and the eye has a somewhat slower lens. The eye attempts to solve this problem in very low light by using receptors that are not sensitive to color. Dark vision is also less clear, meaning that "pixel binning" must be going on as well.

III. Diffraction

It is well know from elementary physics texts that the minimum distance between images that can be resolved when they are cast through a single round hole aperture is given by the angle $\Theta = 1.2\ \lambda/d$, where λ is the wavelength and d is the diameter of the aperture. (This is true for small angles as described next). This affects different film or sensor sizes in different ways.

Full frame 35mm sensors

For a typical light wavelength of λ=500nm and a 35mm camera with a 50mm lens set at f/10, d=5 mm and Θ=1.2x10-4. This angle corresponds to an image size of 6µm at 50mm, or 1/4000 of the vertical dimension of the sensor (24mm—a "35mm camera" has a sensor size of 24 by 36 mm).

This means that a resolution of 4000 lines[5] or pixels in the vertical dimension is the most that could ever be resolved at f/10. This is significant because a vertical resolution of 4000 is, to first approximation (but see the next paragraph), equal to 24 megapixels, a value that already is reached by state-of-the-art cameras. This implies that to make full use of the sensor capabilities of the best 2009 cameras already in production requires the use of apertures less than about f/8. And since the quality of the optics limit the resolution of most lenses today when apertures are much larger than f/5.6 or f/8, this suggests that about 25 megapixels is the current practical limit for a 35mm camera, a limit that will apply until there are some breakthroughs in optics that can perform better at f/4 and faster than they do at f/8 or 10.

This calculation is oversimplified in a number of ways, many of them offsetting. First, it ignores the fact that pixels are arranged in regular rectangular grids. Suppose the diffraction limit implies a resolution no higher than 4000 lines, but the actual separation of the objects is 10% higher, or 3600 lines. A 4000-line grid won't resolve a 3600 line image well because the image won't line up with the pixel array. The number of pixels needed to resolve precisely an image of 3600 lines is 7200: each line requires a pair of pixels, one exposed and one not.

The more rigorous calculation illustrates the concept of the Nyquist frequency: the lowest frequency needed to extract all of the available information from the image. The Nyquist frequency is equal to twice the most detailed resolution one is seeking.

So to resolve all data up to a frequency corresponding to 4000 lines would require a Nyquist frequency of 8000 vertical lines, corresponding to 100 megapixels.

---

[5] The word "lines" refers here to cycles of a sine wave or line pairs.

But this is a worst case calculation. At frequencies below the Nyquist frequency, we can still see some of the image data but not all.

A *practical* upper limit would be double the number of pixels compared to the diffraction limit as opposed to increasing them four times in order to capture all the information available, according to several bloggers who have looked at this issue in real photos[6]. So the diffraction would limit the needed level of megapixels to 50 at f/10.

This is an important factor in the real world because real images have all sorts of spatial frequencies. The upper limit of resolving power for full information reproduction is not as important with a range of frequencies as it would appear from a resolution test that contains all of the information at the high spatial frequency. This is another reason why resolution limits at the Nyquist limit have minimal practical value, and why the Nyquist limit of 100 megapixels is not the best estimate of where the visual limit caused by diffraction will be.

The second major oversimplification is that the diffraction limit referred to above is not the limit of how small an image can be rendered *sharply*—it is the limit to how small a detail can be distinguished *at all*. It does not consider that the quality of the image will be greatly degraded.

The Airy disk of a circular slit consists of a widening and reshaping of the image of a point source, spreading the central point but also consisting of a series of secondary rings. All of this divergence from a point adds noise to the image; it is seen as reduced contrast within dimensions noticeably larger than Θ.

At an angle of $\Theta = 1.2\ \lambda/d$, the first minimum of the diffraction pattern of one point source overlaps the maximum of the adjacent one's pattern; this is conventionally accepted as the minimum angular separation required to distinguish two separate sources, and is referred to by physicists as the Rayleigh criterion. But the contrast difference between the two adjacent maxima is minimal. To render two adjacent sources with clarity (minimal loss in contrast) would require at a minimum that the *first minima* of the diffraction patterns overlap. This doubles the special frequency of the tightest pattern that can be rendered clearly as a result of diffraction. That is, it reduces the number of vertical lines that a camera needs to resolve at the assumed conditions from 4000 vertical lines to 2000. So the needed number of megapixels for sharp-looking (high contrast) pictures (at the assumed aperture of f/10) is about 24[7].

Even this may be an overestimate because it assumes monochromatic light. For red light, the diffraction is about 40% worse, and so for multi-colored light points, diffraction will produce color fringing with red on the outside. This makes for a visually fuzzier image than for a single-wavelength approximation.

---

[6] One of the more detailed explanations is Brain P. Lawler, "Resolving the halftone resolution issue: how many dpi does it take to make an lpi?"

[7] To resolve 2000 vertical lines at the Nyquist frequency requires 4000 by 6000 pixels.

This effect of seeing less sharp pictures about a factor of two in resolution less than the Rayleigh diffraction limit has been noted in blogs[8], where sample pictures show that detail close to the diffraction limit can be observed but looks fuzzy. The author has noted it as well, observing such evident degradation in image quality at f/22 that he no longer uses f-stops smaller than f/10. This can be seen in Figures 1-3. This empirical observation aligns well with the rough concept described above where the diffraction limit for clear and crisp pictures from a 21-24 megapixel camera at f/10 is 2000 vertical lines, which is consistent with the calculated Nyquist frequency of 4000 vertical lines.

The other visual evidence that this is going on is that when I look at extreme enlargements of my sharpest pictures, the limit imposed by the optics seems comparable to the limit imposed by pixilation. For wider apertures, it looks like a finer pixel grid would improve the image, but not by a lot.

So the practical "speed limit" for 35mm cameras is close to the current level of about 24 megapixels for a perfect f/10 lens. In order to make use of higher resolution sensors will require both the use of wider apertures than f/10 (at all times) and probably the invention of better optics that can produce more sharpness at wider apertures than is typical of current lenses. And of course, as shown above, each doubling of megapixels cuts the photon count in half. And at the 21-24 megapixels calculated above, we are pushing the issue on film speed and noise, so greater resolution comes at the expense of noise and speed.

A Digression on Digital Camera Sensor Design

This discussion is oversimplified in a number of ways that interact with each other, relating to the form of the image we are trying to record, the way that digital camera sensors record color information and reduce artefacts caused by regular image details that interfere with the regular pattern of pixels, and the color of the images.

To analyze these effects we would have to construct a matrix of calculations that looked at the following issues in all possible combinations:

- Image characteristics: what sort of image detail are we trying to record? The discussion above implicitly assumed that the detail we were trying to distinguish is two point sources located closed together, or equivalently, a pattern of points located on square grid. The calculation would be a little different for:
    - Parallel lines a fixed distance apart. The diffraction formula would no longer have the factor 1.22 in this case.
    - A thin line (imagine a spider web strand) that we want to record as a line rather than have it disappear into the background.
    - A sharp edge. When we look at spatial frequencies, the ability to render sharp edges or fast patterns of gradation from light to dark depends on the ability to see regular patterns at a higher spatial frequency than what we want to record. This is

[8] One of the better examples is http://www.luminous-landscape.com/tutorials/understanding-series/u-diffraction.shtml

analogous to the need to register higher sound frequencies to reproduce a specific wave shape. Thus to distinguish a 2000 Hz note on a flute from the same pitch on a violin depends on the ability of the stereo to reproduce sounds at 4000, 6000, and even 20,000 Hz. This would be mathematically equivalent to the case of looking at two points or two lines except for its interaction with the other issues discussed next.

- Color. The discussion above assumes light is monochromatic with a wavelength of 500 nm. This is greenish light. But the author (and the reader?) was imagining white light when visualizing this calculation and its effects. So we would actually have to consider at least three cases:
  - White light. In this case, we would have three concentric Airy disks: one smaller in diameter for blue light, and one larger for red light. The red disk would be about 40% bigger since red light has a wavelength as long as 700 nm. If the image we are taking is in black and white, diffraction would be 40% worse than calculated. If the image were to be in color, we would see a red ring around a white Airy disk. The camera would likely have difficulty in rendering the color edge of the ring, however. We would need to consider separately the ability to distinguish two adjacent point sources that were:
  - The same frequency as each other
  - Different frequencies
  - White or any combination of all three frequencies
- Digital camera technology. A digital sensor used a Bayer filter over the photon sensor, which is a 2 by 2 array of red, blue and two green filters, one color over each pixel. It also employs an anti-aliasing filter than refracts a small fraction of the light incident on any pixel to several of the adjacent pixels, so a source of a given color that only illuminates one pixel and thus is of unknown color will also illuminate (to a lesser extent) adjacent pixels with different filters. This filtering reduces the modulation transfer function of a sensor at spatial frequencies close to or above the limit of that the number of pixels can render.
- Which diffraction limit are we interested in? Most discussions of diffraction in photography concentrate on the Rayleigh criterion—the bare ability to distinguish between two points being separate. This corresponds to the first minimum of the Airy disk of one point overlapping the central maximum of the adjacent one. It allows astrophotographers to see when the image they are studying corresponds to one star or two. But as I argued previously, this criterion is not as useful for conventional photography because the contrast is so low when it is barely met. I suggested that a more meaningful criterion appears to be for the first minimum of one point to overlap the first minimum of the adjacent one: this would allow almost the full contrast range to be recorded.

A thorough analysis of these issues would require looking at each possible combination of these factors; it would far more than double the length of this paper, and so is not attempted here.

But we can very briefly consider two cases and show that these effects are not very large and to some extent cancel each other. The following calculation is what physicists call a Fermi problem because Enrico Fermi used to write them for his students. It involves a problem in which the goal is to get the underlying structure of the solution correct while essentially making up the input data. A good solution will illustrate that even if ones guesses on the data are wrong (within limits of course) the answer remains the same.

Consider first a two dimensional grid of points that are barely resolved. Since I do not have enough data on how anti-aliasing filters work quantitatively, I will make a number of arbitrary assumptions. They illustrate a general trend that should be true even if the assumptions are mistaken.

Assume that the first point is centered on pixel (1,1) and the second at pixel (1,2.7) the third point is at (1,4.4). The points are 1.7 pixels apart, so we are trying to resolve one line pair using 1.4 pixels, or the equivalent of 4000/1.7 or 2400 line pairs on a 24 megapixel sensor, almost at the limit measured in camera tests.

Without diffraction or anti-aliasing, pixel (1,1) would record 1 unit of light, pixel (1,2) would record .3 units, and pixel (1,3) would record .7 units. Pixel (1,4) would record .6 units and pixel (1,5) would record .7. Thus we would see the second point as distinct from the first, with *very* low contrast. (Distinguishing the second point from the third is problematic in this example, and becomes more so in the proceeding, so we will ignore it here.)

With diffraction at the Rayleigh limit, the Airy disk would have a radius of 1.7 pixels. The first point would record that fraction of the energy of the disk that is within a radius of .5 pixels (plus the corners) from the center. This is a very difficult calculation to make, so I will assume that 80% of the light makes it into pixel (1,0) and the other 20% is spread among pixels (1,2) and the other three neighbors. Pixel (1,1) receives some light from point 2, but less, so it is now at, say, .82 units of light.

Pixel (1.2) is now up to .35 units--.3 from the point source and .05 from diffraction from point (1.1). This pixel also receives diffracted light from the second point source—more of it proportionally. So perhaps it is at .43 units. Pixel (1.3) loses a lot less of its light from diffraction to the right, and also picks up some light from the second maximum of point 1's Airy disk. I assume here that it drops only about .02 to .68 units.

Suppose the anti-aliasing filter sends 5% of each pixel's light to each of the four adjacent pixels. The light detection of pixel (1,1) is now down to .8*.82 or .66; and pixel (1,2) picks up .04 to a level of .47. And pixel (1,1) picks up .03 of light from (1,2) due to the filter But the filter also reduces pixel (1,2)'s own light level by 20% of .42 or .1, so it is at .46. On the other hand, it picks of about .02 from the right. Pixel (1.3) loses 20% and is at .8*.68 or .54 plus whatever it picks up from the right.

So without considering the anti-aliasing filter, the first three pixels record .82, .42 and .68 units of light respectively. After the filter, the results are about .69, .48, .54. In both cases the first two points are weakly distinguishable, but the contrast is lower with the filter's effects included.

The point of this calculation (which is for monochromatic light) is that the anti-aliasing filter reduces contrast and may even take barely distinguishable points and make them indistinguishable, but it will not change the effect of diffraction on digital resolution by factors like 1.4, but rather by something more in the neighborhood of 10%. Another interesting point is the importance of the implicit assumption that the source points of light are equally intense. If point 2 were dimmer than point 1 in this calculation, we might no longer be able to distinguish them. If point 2 were brighter, the difference would be disproportionately larger.

Adding color into the equation affects the results by as much as 40% in both directions—a much bigger effect.

And looking at the other choices of images we might want to study, or of how this calculation would differ if the point sources were at twice the Rayleigh limit, it is evident that the effects of sensor design are to reduce the visibility of diffraction limits by a quite small amount, smaller than the error of other approximations in this derivation. For example, if we repeat the calculation above with 2-pixel separation, the anti-aliasing filter does not reduce resolution at all, but only contrast. Pixels (1,1) and (1,3) would be the centers of their Airy disks and the effects of the filter; so pixel (1,2) would gain brightness from both effects but remain much dimmer that 100%. This is also what is predicted by the Nyquist-Shannon sampling theory: with resolution of two pixels per line pair, we are able to extract all of the available information and do not need to worry about artefacts.

One other trend is interesting from this type of calculation. We are trying to see small differences in light level between adjacent pixels in order to resolve closeby points. If the signal to noise ratio is small, these differences will be lost in the noise, So this implies that in the deep shadows, we will lose resolution due to both diffraction and anti-aliasing regardless of how the sensors or the RAW algorithms work.

Other sizes of film or sensor

For other sizes of "film" the diffraction limit varies linearly with sensor dimension, as seen next. As the linear dimension of the sensor increases, the focal length of the normal lens increases proportionally. This means that the aperture diameter increases linearly, and the angle limiting resolution is inversely proportional. Thus if we double the size of the sensor, the number of lines that can be resolved in one dimension doubles also.

Smaller sensors for interchangeable lens cameras

The pixel size for a smaller sensor (roughly 15 by 22 mm, with a little variation bigger or smaller by manufacturer) is about 20% smaller, so diffraction effects begin to be visible at about f/8 instead of f/10. And because of the smaller sensor size, the number of pixels at the same number of lines of resolution is about 40%; this suggests that the ~25 megapixel level of diminishing returns is equivalent to about 10 megapixels, a level that is already exceeded by several models.

As we compare camera lenses, whether they were designed for film cameras or digital, we find that the optics designed for smaller cameras offer higher resolution (in lines per mm) than those for larger cameras, so a film area that is 4 times bigger is less than 4 times sharper. But this is not true for small-sensor camera. In fact, most of the lenses available for these products are designed for and interchange with full frame sensors. So the higher pixel density on smaller-sensor cameras takes us into new territory for exploring the limitations of the lens relative to the sensor. Reviews of the first 15-megapixel cameras have begun to point this out, showing higher levels of resolution in some cases[9] but failing to see much advantage to the finer screen in most cases.

Pocket cameras

Typically the sensor size is 1/5.5 that of 35mm film, thus the diffraction limit for 4000 lines of clear, sharp resolution or 25 megapixels is f/1.8 instead of f/10. Since no current small-sensor camera has a lens that fast, it means that small high-MP cameras are always diffraction limited and that megapixel counts much above about 12—which is currently offered on a few top of the line cameras—are almost pointless. This observation explains why typical small cameras do not even allow f-stops smaller than f/8: at f/8 the diffraction limit is 900 lines, corresponding to about 1.5 megapixels.

The author has also observed this effect: pictures taken at f/8 are visibly, disappointingly less sharp than those taken at wider apertures. I have started using a pocket camera with manual override to assure that I use apertures wider than about f/5, and preferable much wider, whenever possible. This corresponds to about 3.5 megapixels, which is less than the sensor resolution of 7 megapixels, and looks that way, since enlarged photos start to look blurry before they look pixilated.

Large and Medium Format Cameras

A 4x5 (inch) film camera has about 4 times the resolving power (in terms of diffraction limit) of a 35mm camera at the same f-stop. But most large format cameras have slow lenses and are used stopped down for depth-of-field reasons, so this increase usually is not fully realized. So if we wanted our 4x5 camera to perform well at f/32, we would only get 25% greater resolution than a 35mm camera at f/10.We would need only about 50% more pixels than for 35mm to perform at the same level.

No one yet produces a commercial 4x5 digital sensor. The largest size commercially available is 36mm x 48 mm. This is about 1.4 times the size of a 35mm sensor. So it could make use of about twice the megapixels as 35mm. The densest sensor at this size has 39 megapixels. Note that this again limits the f-stop to f/10 to make use of this pixel density.

Vision

As mentioned in the discussion on photons, most of this discussion is about wavelengths of light and lens and sensor dimensions. So it also applies to human and animal vision. Diffraction would appear to impose a hard limit on how eagle-eyed an eagle can be.

III. Conclusions

[9] See http://www.luminous-landscape.com/reviews/cameras/50d.shtml

A. Physics conclusions

Technology today is approaching or even at the limits of resolution, film speed, and image size imposed by the laws of physics for cameras in the 35mm format or smaller. Further significant improvements in speed or resolution will require the use of larger sensors and cameras.

Thus today's equipment is sufficiently powerful to allow a physicist to see directly—without the need for instruments and calculation/interpretations beyond the camera itself and a computer monitor-- the effects of both the photon nature of light and the wave nature of light. A single picture taken at f/22 and ISO 3200 will display uniform unsharpness unrelated to motion blur or focus error and the noise patterns of each primary color as well as overall brightness that reflect the statistical nature of photon detection.

B. A Photographer's Conclusions

Currently available products are pushing the limits of what is physically possible in terms of resolution and low-light performance. Therefore a photographer must relearn how to make choices about what size of camera to use for what assignment, and what tradeoffs to make when shooting concerning aperture and depth of field and concerning film speed versus image quality. Past experience and rules of thumb developed in the film age will give results that are qualitatively and quantitatively different than what our intuition led us to understand.

1. Equipment size/sensor size means better pictures

As equipment pushes closer to the diffraction limit, image quality is directly proportional to image size. Equipment size scales with image size: by simple dimensional analysis, the weight of equipment should scale as the third power of image size. Better pictures require heavier equipment to an even greater extent than was true in the past.

2. Depth of field

Depth of field at a given f-stop is inversely proportional to image size. In the past, photographers got around that by using smaller apertures for large cameras: a setting of f/64 on an 8 by 10 inch camera produced the same depth of field as f/8 on a 35 mm camera, but allowed greater resolution. But with higher sensor resolution for 35mm digital cameras, both cameras are at the diffraction limit measured in angular terms. Therefore there is no real advantage to the large format. Taking advantage of the larger film size will now require larger apertures, limiting creative freedom for 8 x10 users who want more resolution in return for the weight and inconvenience.

3. What does speed trade off against?

With film, sensitivity traded off against many parameters: resolution, color accuracy, and grain being the most significant. With digital, the only tradeoff of importance is with noise. And even this tradeoff refers mainly to noise in the highlights and middle tones, where it is not a big aesthetic problem even at ISOs of 800 or 3200 or even higher (depending on the aesthetic sense of the viewer). Looking at the performance of current best-in-class cameras, higher film speed comes mostly (or in some cases almost entirely) at the expense of dynamic range. While this tradeoff may be due in part to engineering choices made by the manufacturer, much of it is fundamental: at the highest dynamic range currently available, even at ISO 100 the noise in the shadows is pushing the limits of what is acceptable artistically.

4. Film

Digital equipment performs *much better* than film of the same sensor size, which is why a discussion of the limits imposed by the physics of light was not interesting in the past. This better performance, which to the author's eyes is a factor of 3 to 5 compared to the best 35mm slide film, means that past rules of thumb and tradeoff rules must be reconsidered. For example, a rule of thumb for hand-holding the camera used to be that the shutter speed must be at least the reciprocal of the lens focal length. Thus a 50mm lens could only be hand-held reliably at 1/50 second. But if the image *could* be 3-5 times sharper than film, than a three-times-faster shutter speed is needed. Similarly, since depth of field is based on the smallest size of blur that could be considered equivalent to a point, it now is 3-5 times shallower. And not only that, but the small apertures that photographers used to use for large depth of field are no longer available due to diffraction. This is even more the case for smaller sensors, where for pocket cameras sharpness demands the widest aperture available. Depth of field is no longer the creative choice that it used to be.

ILLUSTRATION OF DIFFRACTION AND NOISE IN THE SAME PICTURE

Figure 1. Enlargement of Figure 3, taken at f/22. Note noise especially in shadows on the right hand side of the buildings. Diffraction limits both sharpness and contrast.

Figure 2. Identical photo except taken at f/8, shows how much sharper the picture looks where diffraction isn't an issue. Note noise is the same because exposure is the same.

Figure 3. Full size picture from which Figures 1 and 2 are cropped. (Actually only Figure 1 is enlarged from this precise image since this was taken at f/22).

1 A typical dynamic range for film is about 7 f-stops (an f-stop is a factor of 2); pocket digital cameras perhaps have a little less. As of spring 2009, the four most advanced digital cameras had dynamic ranges of about 8 stops at ISO 3200 (see dxomark.com). So this paper provides a range based on these limits.
2 These spec's apply to the Canon 1Ds Mark III and 5D Mark II, both at 21 megapixels, and the Nikon D3X and the Sony Alpha 900, both at 24 megapixels; the pixel size is also about the same for smaller-frame cameras (image size about 15 mm by 22 mm) produced by numerous manufacturers.
3 Data on dynamic range of real cameras are taken from dxomark.com.
4 This is done by intentionally underexposing 3 stops with the ISO setting at 3200.
5 The word "lines" refers here to cycles of a sine wave or line pairs.
6 One of the more detailed explanations is Brain P. Lawler, "Resolving the halftone resolution issue: how many dpi does it take to make an lpi?"
7 To resolve 2000 vertical lines at the Nyquist frequency requires 4000 by 6000 pixels.
8 One of the better examples is http://www.luminous-landscape.com/tutorials/understanding-series/u-diffraction.shtml
9 See http://www.luminous-landscape.com/reviews/cameras/50d.shtml