Roadmap for CMOS image sensors: Moore meets Planck and Sommerfeld

Peter B. Catrysse* and Brian A. Wandell
Department of Electrical Engineering
Stanford University, Stanford, CA 94305

The steady increase in CMOS imager pixel count is built on the technology advances summarized as Moore's law. Because imagers must interact with light, Moore's Law impact differs from its impact on other integrated circuit applications. In this paper, we investigate how the trend towards smaller pixels interacts with two fundamental properties of light: photon noise and diffraction. Using simulations, we investigate three consequences of decreasing pixel size on image quality. First, we quantify the likelihood that photon noise will become visible and derive a noise-visibility contour map based on photometric exposure and pixel size. Second, we illustrate the consequence of diffraction and optical imperfections on image quality and analyze the implications of decreasing pixel size for aliasing in monochrome and color sensors. Third, we calculate how decreasing pixel size impacts the effective use of microlens arrays and derive curves for the concentration and redirection of light within the pixel.

Keywords: image sensor, microlens, imaging lens, photon noise, diffraction

1. INTRODUCTION

There has been a steady increase in the quality of images captured using complementary metal oxide semiconductor (CMOS) sensors. Even high-quality imaging applications that traditionally use charge-coupled device (CCD) imagers, such as single lens reflex (SLR) cameras, are increasingly equipped with high spatial resolution CMOS image sensors.[1] In part, the improved image quality is due to increased pixel density that is built on Moore's law: Over the past 25 years, the minimum lithographic feature size has decreased by thirty percent every three years.[2, 3]

CMOS technology scaling can improve image sensor performance in a variety of ways. In the conventional active pixel sensor (APS), scaling can be used to decrease pixel size and improve spatial resolution. Alternatively, scaling can be used to shrink transistor size and increase photosensitive area in the pixel (fill-factor). Different market segments may choose different strategies. For example, SLR cameras typically use scaling to increase photosensitive area. Small mobile devices, on the other hand, have principally used technology scaling to increase spatial resolution. But there has been relatively little analysis of the tradeoffs in image quality as one chooses between spatial resolution and sensitivity improvements.[4] Such analyses are important because the value of scaling for CMOS image sensors differs from the unalloyed value of scaling in other integrated circuit applications. Specifically, the sensor applications differ because (a) the properties of light, the input to the sensor, do not scale, and (b) the human visual system, not another integrated circuit, ultimately receives the sensor output.

Two fundamental properties of light must be considered to understand the effects of scaling on image sensor performance: photon noise (Planck)[5] and diffraction-limited blurring (Sommerfeld).[6] First, consider photon noise. As pixel size shrinks, the mean photon count at the photodetector falls. The Poisson variance of the photon signal at the pixel equals the mean photon count, so that the signal-to-noise ratio (SNR) of the photon signal decreases. Thus, reducing pixel size inevitably increases image noise. Second, consider diffraction. As pixel size shrinks one can increase spatial sampling density (pixel pitch). But image resolution does not improve without bounds because the spatial detail in the image is limited by diffraction, i.e., the spatial spread caused by light passing through a finite aperture.

A further complexity of the image quality analysis is that the output of CMOS imagers is transformed by several computational steps beyond the sensor, and the image is ultimately delivered to the human visual system. Consequently, image system evaluation must include measures of perceptual image quality, and these are beyond the traditional

* Email: pcatryss@stanford.edu, Phone: (650) 736-2309

Digital Photography, edited by Nitin Sampat, Jeffrey M. DiCarlo, Ricardo J. Motta, Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 5678 © 2005 SPIE and IS&T · 0277-786X/05/$15

engineering tools. The evaluation of image appearance requires tools to assess the implications of sensor noise and spatial resolution, such as the color metrics defined by the CIE[7] and perceptual testing to assess noise visibility.[8]

In this paper, we make three observations concerning the effect of decreasing pixel size on image quality under low illumination conditions. First, decreasing pixel size increases the likelihood that photon noise will become visible. Recent perceptual measurements of noise visibility show that photon noise becomes visible when the SNR is around 30 dB,[8] and we assess the consequences of pixel scaling given this constraint. Second, pixel size below the blur circle of the taking lens will not improve system spatial resolution. We quantify the theoretical limits of system spatial resolution for both monochrome and color image sensors, and we illustrate the expected artifacts for realistic lenses. Third, we calculate the implications of decreasing pixel size on the effective use of microlenses. These are small lenses, placed at the surface of the pixel, whose purpose is to focus (concentrate) photons on the photodetector. As pixel size shrinks the aperture shrinks, and the ability of the microlens to concentrate light decreases.

2. METHODS

We use the Image Systems Evaluation Tools (ISET) software to quantify the consequences of decreasing pixel size on image quality.[9] This software is based on a model of the imaging pipeline that includes (1) a radiometric description of the scene, (2) optical transformation of the scene radiance to the irradiance signals at the sensor, (3) sensor capture, (4) digital image processing for display, and (5) several perceptual metrics.

A simulation using ISET is initiated by the creation of a scene. A scene consists of a radiometric dataset in which each pixel represents the spectral radiance $L(x,y;\lambda)$ at a spatial location (x,y) expressed in units of $W/(s\,m^2\,sr\,nm)$. ISET includes several calibrated radiometric datasets, including both real-world and computer-generated scenes.

The imaging optics convert the scene radiance data into a spectral irradiance image $E(x,y;\lambda)$ in $W/(s\,m^2\,nm)$, which is incident on the image sensor. This is described by the camera equation,[10] which produces the ideal geometric image irradiance distribution

$$E(x,y;\lambda)=\frac{\pi T(\lambda)R(x,y;\lambda)}{1+4(1-m)^2(f/\#)^2}L\left(\frac{x}{m},\frac{y}{m};\lambda\right), \tag{1}$$

where $T(\lambda)$ is the spectral transmittance, $R(x,y;\lambda)$ is the relative illumination factor to account for off-axis irradiance, and m (< 0) is the magnification and

$$f/\#=\frac{f}{D} \tag{2}$$

is the f-number of the imaging system, a dimensionless quantity, which is the ratio of its focal length and the diameter of its clear aperture. A circular diffraction-limited lens is completely characterized by its f/#. To include the effects of image blurring that occur due to diffraction and aberrations, the incoherent image formation process is completed by a convolution operation[11]

$$E_{image}(x,y;\lambda)=PSF(x,y;\lambda)\otimes E(x,y;\lambda), \tag{3}$$

where $PSF(x,y;\lambda)$ is the point-spread function of the imaging lens and $\otimes$ denotes the convolution operator. This convolution operation is valid over small image regions (isoplanatic). The full calculation requires dividing the image plane into smaller sections over which there is negligible change in the PSF. Each locally isoplanatic section has its own PSF. ISET includes a variety of imaging optics models ranging from an ideal diffraction-limited lens to realistic lens models that include all the effects of monochromatic and chromatic aberrations, pupil aberrations, vignetting, and diffraction.[12]

The remainder of the optical path, from sensor surface and silicon substrate, includes a microlens, color filter, planarization and interconnect layers, contacts, vias and dielectric layers that wire active devices into circuits. Different approaches can be taken to characterize this part of the optical path. We analyze the behavior of on and off-axis pixels using a phase space approach. The relationship between the surface irradiance image and the image incident on the silicon substrate can be summarized by an optical efficiency[13]

$$E_{substrate}(x,y;\lambda) = OE(x,y;\lambda)\,E_{image}(x,y;\lambda). \tag{4}$$

For color image sensors, different optical efficiencies represent the respective color channels: OE_R, OE_G, OE_B. Each of these optical efficiencies includes the appropriate color filter transmission properties.

The optical path is followed by an electronic path, which converts of the optical image into a voltage image

$$V(x,y) = gAt_{\exp}\int QE(\lambda)\,E_{substrate}(x,y;\lambda)\frac{hc}{\lambda}d\lambda, \tag{5}$$

where g is the conversion gain in $\mu V/e^-$, A is the photodetector area inside the pixel in m^2, $t_{\exp}$ is the exposure time in s, $QE(\lambda)$ is the spectral quantum efficiency, which describes the conversion efficiency from photons to electrons, c is the speed of light in vacuum, and h is Planck's constant. The voltage image is subsequently digitized to form a digital representation of the captured image. The ISET software includes a quantitative model of the photodetector circuitry that accounts for a wide variety of sensor parameters, such as pixel fill-factor, dark noise, read noise, and many others. The software also includes a variety of image processing algorithms, such as demosaicking, color conversion, and color balancing, and rendering methods, e.g., to allow the digital image to be viewed on a monitor. In addition, it includes metrics to evaluate the perceptual quality of the final rendered image.[9]

3. RESULTS

We describe three consequences of the reduction in pixel size on image quality. First, we calculate the mean number of photons incident per pixel at the sensor under a variety of low-light imaging conditions. This number is of interest because the Poisson variability in photon capture (photon noise) is visible at low photon capture rates (< 1000 photons/pixel/image).[8] We describe the incident photon levels and illustrate the consequences for image quality of working with small pixels and thus low absorption rates. Second, we analyze spatial resolution with respect to the pixel size. We calculate the spatial blur expected from a diffraction-limited lens and a realistic example lens. We then illustrate the interaction between blurring and the typical color sampling implemented by conventional Bayer patterns. Third, we discuss the implications of pixel shrinkage for microlens design. Pixel size also determines the size of the microlens aperture. Pixel size reduction places limits on the f/# of the microlens, and this has significant consequences for the ability to concentrate the light at the sensor surface onto the photodetector buried within the pixel.

1. Photon noise

1.1. Incident photon levels

Figure 1 shows the mean incident photon level per pixel as a function of pixel size and photometric exposure (lux-sec). The photon level is represented by the shading in the image. The dashed curves show three equal-photon levels: 1000, 2000 and 4000 photons/pixel. The pixel size spans the range found in many current image sensors (2-8 μm), and the photometric exposure range covers medium to low light level conditions, from 0.01-0.001 lux-sec. The values were calculated assuming an equal energy spectrum. These photometric exposure values would be obtained by imaging the darker half of a typical computer display with an f/2.8 lens and a 10-ms exposure duration.

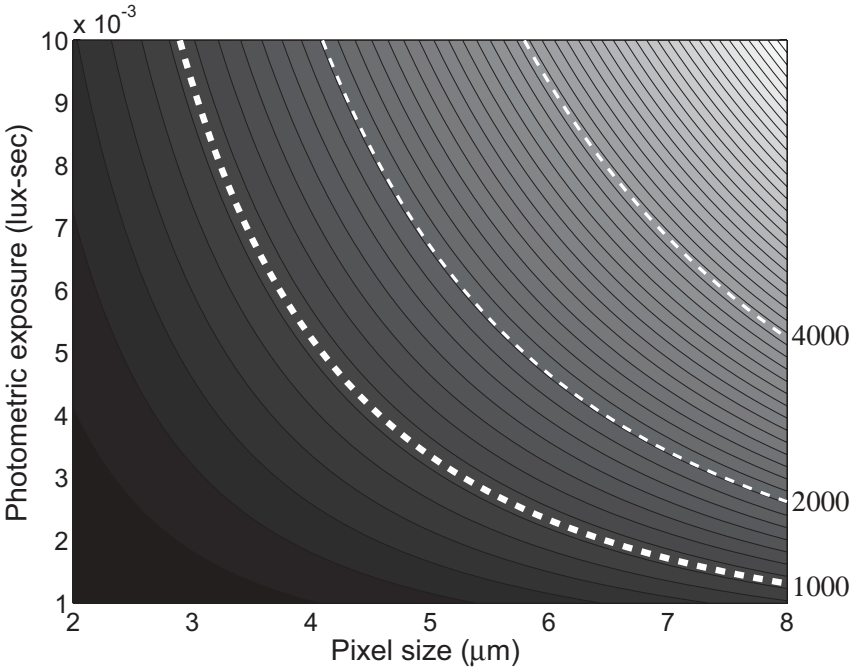

Figure 1: Mean photon count incident on a pixel as a function of pixel size (μm) and photometric exposure (lux-sec). Counts are calculated for an equal-energy spectral source. The lines represent equal-photon counts. The dashed lines represent levels of 1000, 2000 and 4000 photons.

The figure measures the number of incident photons that can be captured by an ideal pixel (100% fill-factor, 100% quantum efficiency; no pixel induced noise); it also defines an upper bound on the pixel SNR. Any loss of light due to scattering, imperfect detector quantum efficiency, the presence of color filters, incomplete fill-factor, will decrease the number of photon absorptions and decrease SNR. Further, any sources of electronic noise in the pixel will also decrease SNR.

Psychophysical experiments show that spatially uniform photon noise becomes visible at an SNR of 33:1.[8] This level corresponds to the bold dotted equal-photon line at 1,000 photons/pixel. For example, a photometric exposure of 0.01 lux-sec contains just enough photons so that the SNR of an ideal 3-μm pixel will render the photon noise at the threshold of visibility.

1.2. Visibility of photon noise: monochrome images

Figure 2 illustrates the visibility of photon noise. The image shows a simulation of a Macbeth Color Chart[14] captured by a monochrome sensor comprising an array of ideal pixels. The patch luminance ranges from 32 to 1.25 cd/m^2. This range is comparable to the darker half of a typical computer display. To create these images we simulated an ideal 3-μm pixel in an ideal 640x480 monochromatic sensor. By ideal sensor we mean zero dark signal non-uniformity (DSNU) and zero photo-response non-uniformity (PRNU). The images were calculated using 10-ms exposure duration and two different diffraction-limited imaging lenses (f/2.8 and f/8.0). In the image calculated using the f/2.8 lens, the number of incident photons/pixel ranges from 3000 (brightest patch; lower left) to 120 (darkest patch, lower right). For this image, the photon noise is near threshold visibility on a calibrated display. Using the f/8.0 lens, the photon count decreases and the photon noise becomes plainly visible throughout the image.

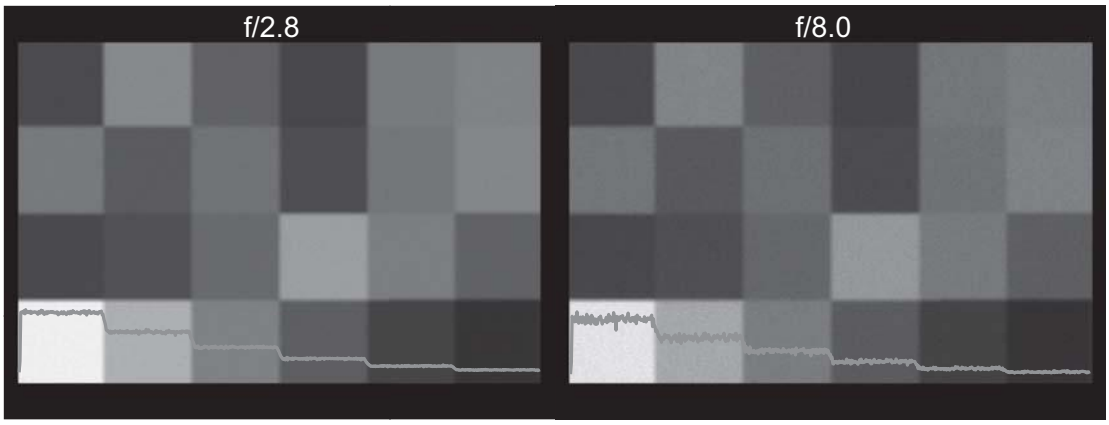

Figure 2: Images of a Macbeth Color Chart simulated for a system comprising a diffraction-limited imaging lens and an ideal monochromatic image sensor with ideal 3-µm pixels. The images were simulated assuming a mean luminance of 10 cd/m^2 and 10-ms exposure duration. Images illustrate the effect of photon noise on image quality for f/2.8 (0.031-0.0012 lux-sec) and f/8.0 imaging lenses (0.0039-0.0002 lux-sec). The traces show the signal level for the bottom six patches. The photon noise becomes easily visible when the lens is changed from f/2.8 to f/8.0, decreasing the photometric exposure.

1.3. Visibility of photon noise: color images

The analysis of monochrome imaging noise is simple because the sensor data (including photon noise) are copied directly to (scaled and linearized) display output values. This transformation preserves the noise contrast. In color imaging, however, several additional image processing transformations take place between data acquisition and display. These transformations alter the noise distribution in the display[7] and can change the noise visibility.[8] We illustrate the perceptual consequences of these transformations for (a) an ideal 3-µm pixel, and (b) an ideal CIF (288x352) format red-green-blue (RGB) image sensor. The simulated pixels were arranged in a 2-by-2 Bayer-pattern (RGGB). The spectral transmittance of the RGB color filters consists of conventional curves with very high transmittance (Figure 3, inset). The sensor data were demosaicked (bilinear interpolation), and transformed into display space with a linear transformation (3x3 matrix) designed to minimize the CIELAB color error (ΔE_{ab}) of a Macbeth Color Checker. Finally, the color channel data were scaled to create an 18% gray image mean.

Figure 3 contains images simulated under three different illumination conditions. The image in panel (a) shows a high quality image under excellent conditions (mean 0.1 lux-sec). The simulation assumes a bright mean luminance (100 cd/m^2), an f/2.8 diffraction-limited lens, and a 10-ms exposure time. The other two images show simulations for lower light levels (10 cd/m^2) using f/2.8 (b) and f/8.0 (c) diffraction-limited lens, respectively. By comparing these images with the one in (a), we can see that color noise is plainly visible in the f/2.8 image (0.01 lux-sec). The f/8.0 image is unusable (0.0125 lux-sec). The results shown in Figure 3 are for an ideal system. It is possible to simulate increasingly realistic examples that include the effects of monochromatic and chromatic aberrations, pupil aberrations, vignetting, and diffraction [12]. Addition of these effects makes the lens model more realistic, but also results in a performance decrease when compared to the ideal diffraction-limited case.

2. Diffraction, sampling and color

In this section, we consider how the imaging lens affects the resolving power of a digital imaging system. We first calculate the relationship between f/# of a diffraction-limited lens and spatial blurring. This analysis defines the upper limit on system spatial resolution. We then consider a realistic case of blurring calculated from the imaging properties of a high-quality multi-component lens.

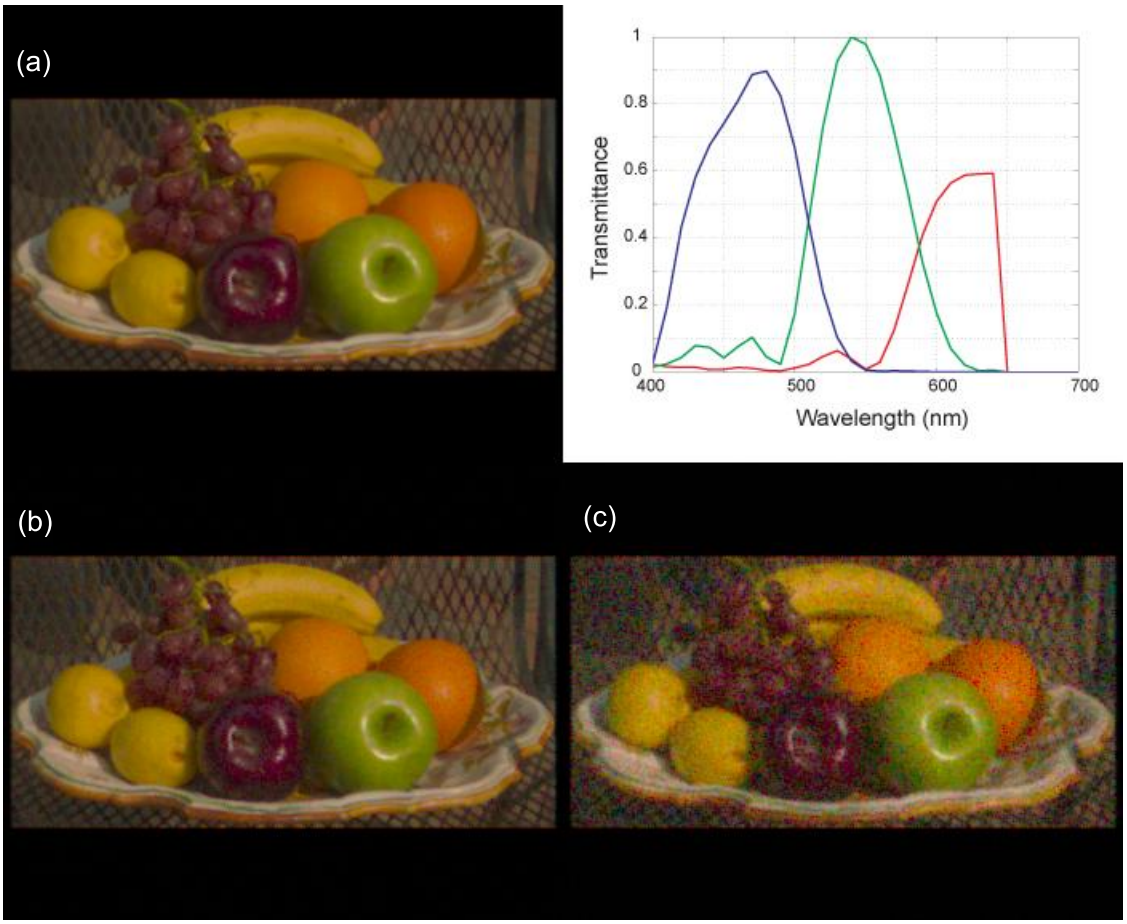

Figure 3: Simulated color images under three different illumination conditions. (a) A simulation assuming a bright mean luminance (100 cd/m^2) and an f/2.8 diffraction-limited lens (0.1 lux-s). (b,c) Simulations under lower mean scene luminance (10 cd/m^2) using an f/2.8 (b) and an f/8.0 (c) diffraction-limited lens, respectively. The simulation assumed an ideal sensor and ideal 3-µm pixels. Inset shows the spectral transmittance of the color filters.

The on-axis blurring of a circular diffraction-limited imaging lens is characterized by a point-spread function in the form of an Airy pattern.[11] Figure 4 shows the diffraction-limited blur circle at 550 nm for an f/2.8 (panel a) and an f/8.0 (panel b) lens. The extent of blurring is usually summarized by the diameter of the main lobe of the Airy pattern:

$$\Delta = 2.44 \times \lambda \times (f/\#) \tag{6}$$

The extent of blurring is also a function of the wavelength of the incident light. The wavelength dependence is shown by lines in Figure 4c, for several f/#s. Each line shows the extent of blurring for light in the visible wavelength range from 400 (blue) to 700 nm (red). For example, an f/2.8 lens produces a 3.76-µm spot size for middle-wavelength light (550 nm).

To avoid spatial aliasing due to sampling by the pixel array, lens blurring must be coordinated with pixel pitch.[15] For a monochrome sensor, there must be two samples per Airy disk diameter. Thus, the pixel pitch must equal the radius of the Airy disk. A finer pixel pitch (higher sampling rate) will not improve the spatial resolution of the imaging system. Under-sampling, on the other hand, produces aliasing artifacts. Hence, to avoid aliasing the pixel pitch of a monochrome sensor should equal or exceed the blur circle radius.

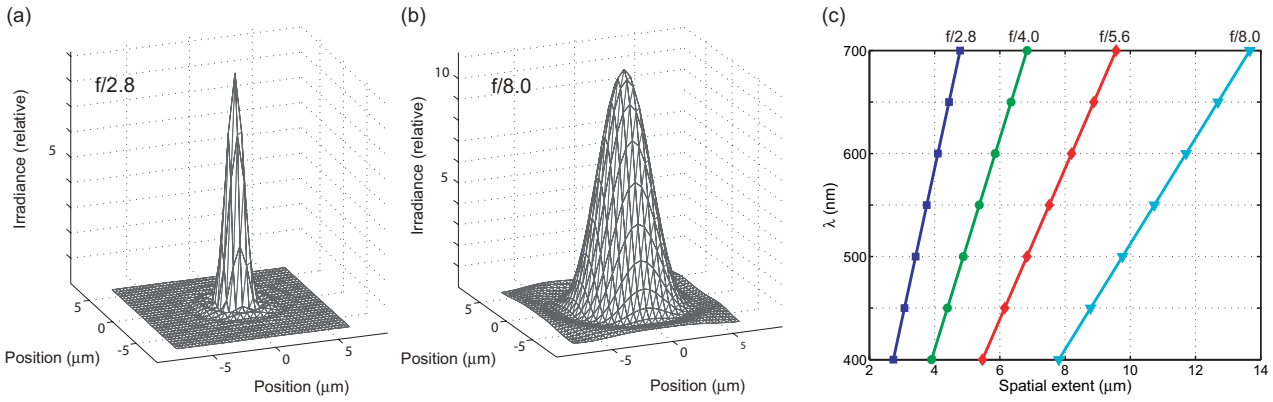

Figure 4: The on-axis point spread function (PSF) of a circular diffraction-limited imaging lens forms an Airy pattern. The mesh plots show the PSF at 550 nm for an f/2.8 (panel a) and an f/8.0 (panel b) lens. The diameter of the Airy disk is measured in panel c for several wavelengths and imaging lens f/#s in panel c.

For example, Figure 4 shows that to avoid aliasing at all wavelengths of visible light using an f/2.8 diffraction limited lens, a monochrome sensor must have a pixel pitch of 1.36 μm or smaller. For the f/8.0 diffraction-limited lens, the required pixel pitch is roughly 3.9 μm.

The sampling rate requirements for color image sensors differ from those of monochrome image sensors. In a monochrome image sensor every pixel in the sensor corresponds to a pixel in the final digital image. Color image sensors, on the other hand, typically measure the image irradiance through a color filter array (CFA) superimposed on the pixels. The most widely used format for the CFA is the Bayer pattern, a periodic array built from a 2-by-2 pixel kernel. Often, the 2-by-2 pixel pattern is comprised of a RGBG arrangement. The data from these four measurements are usually combined to form a pixel in the final display. Hence, such a method implicitly assumes that the sensors in the kernel measure a common spectral radiance in the scene, and the effective sampling rate is that of the Bayer pattern. It follows that to avoid aliasing artifacts the imaging lens must spread light across the extent of the kernel, not just a single pixel. So, for a color image sensor the spatial sampling rate of the image sensor is that of the color filter array pattern rather than the pixel pitch. To avoid aliasing in the color array the pixel pitch must be smaller than that in the monochrome array.

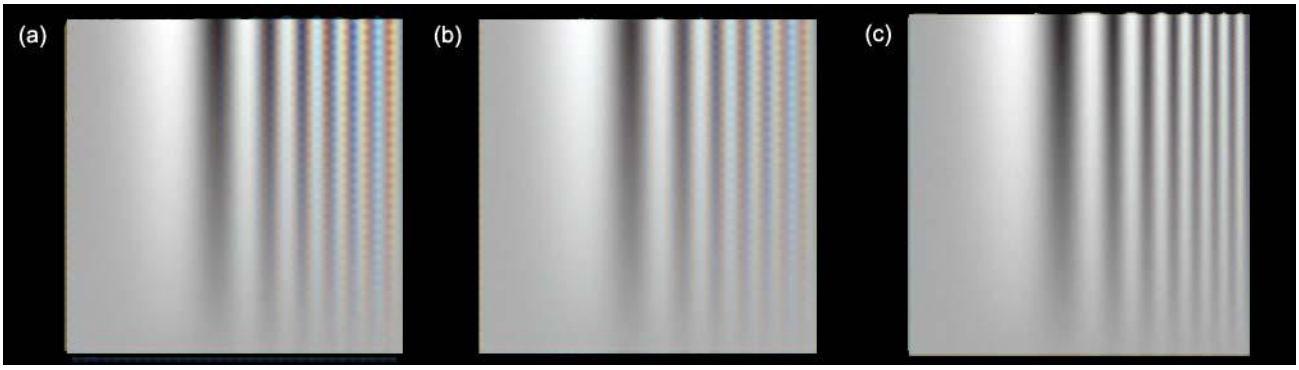

Figure 5: Sampling artifacts introduced by color image sensors. The simulated input image is a monochrome sweep frequency (horizontal dimension) with varying contrast (vertical dimension). In panels (a,b) the scene was simulated assuming diffraction-limited f/2.8 and f/8.0 lenses, respectively. The simulation further assumed an ideal sensor with ideal 3-μm pixels. Color artifacts are visible in panels (a,b). A simulation using an ideal 1.5-μm pixel is shown in panel (c). No color artifacts are visible.

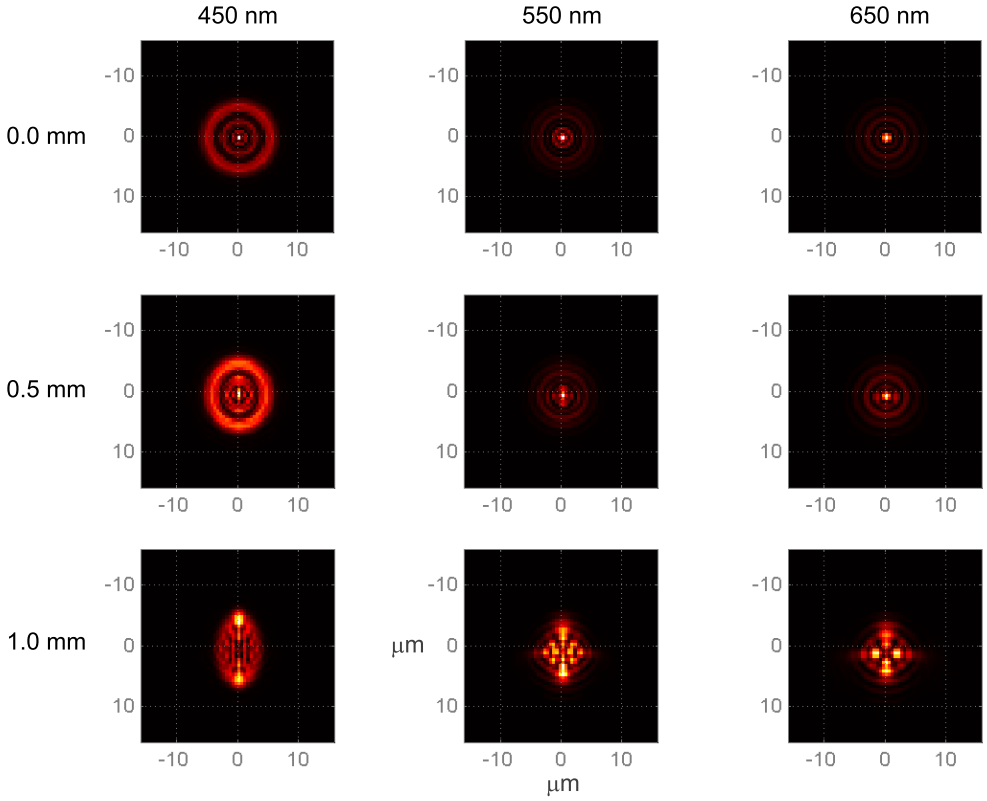

Figure 6: PSFs for a high-quality Edmund Optics Finite Conjugate Micro Video Imaging Lens. The columns contain PSFs at three wavelengths (450, 550 and 650 nm). The rows show three field heights (0, 0.5 and 1.0 mm).

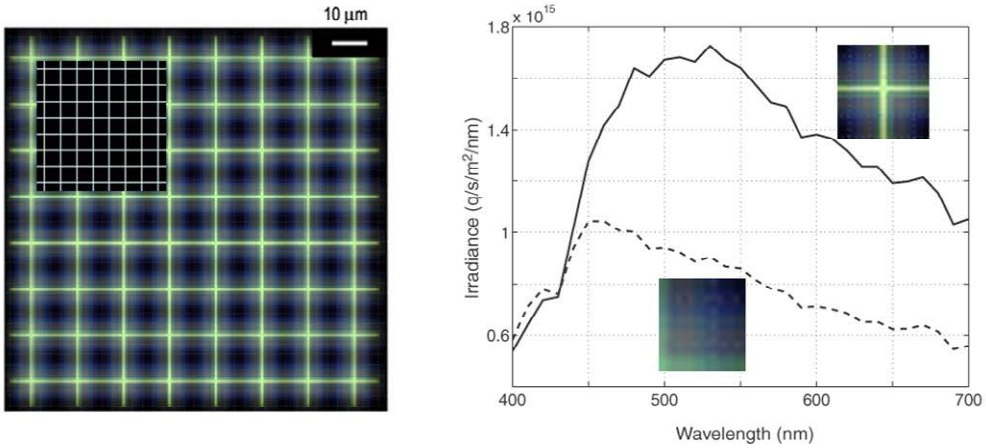

Figure 7: Spectral properties of the irradiance image in a color imaging system simulated with a realistic lens model. (a) The simulated input image is a grid pattern (inset) with a space-invariant (D65) spectral power distribution. The image is the simulated sensor irradiance image produced by the Edmund Optics lens. The uneven color distribution is easily visible. (b) The calculated spectral power distributions in two regions of the irradiance image are shown. Regions near the grid line intersections (solid curve) have a peak irradiance power at 530 nm; regions adjacent to the grid lines peak at 460 nm.

Figure 5 illustrates the appearance of the color artifacts one finds when these assumptions are not met. The simulated input image is a monochrome sweep frequency (horizontal dimension) with varying contrast (vertical dimension). The simulated mean luminance level is 50 cd/m^2. The ideal sensor comprises ideal 3-μm pixels with the color filters show in the inset of Figure 3. The simulation was performed with a 200-ms exposure. The image was rendered using a bilinear interpolation demosaicking algorithm.

In panels (a,b) the scene was imaged through diffraction-limited f/2.8 and f/8.0 lenses, respectively. The rendered images contain colors that are not present in the original. These artifacts are very apparent for the f/2.8 lens (panel a); the reduction in contrast of the f/8.0 lens makes the color artifacts somewhat less visible (panel b). Panel (c) shows that an ideal 1.5-μm pixel, (3-μm sampling rate for the Bayer-kernel) eliminates these color artifacts.

The diffraction-limited point-spread functions in Figure 4ab show that, in principle, spatial resolution can be increased by using pixel pitches smaller than the 5-μm size cited in an earlier roadmap paper by Wong.[16] In practice, however, many lenses fail to reach the diffraction-limit. The images in Figure 6 show the point spread functions of a high-quality Edmund Optics Finite Conjugate Micro Video Imaging Lens[12, 17] at several wavelengths and field heights (distance from the image center). For this lens the spatial blurring at 450 nm is considerably larger than that at 550 or 650 nm (where the lens is nearly diffraction-limited). The point spread function varies from image center to periphery, becoming considerably larger than the diffraction limit for all wavelengths.

Figure 7 illustrates a surprising complexity of color imaging that emerges when simulating the sensor irradiance image. In this case we used a pattern comprising a set of grid lines (inset in panel a) with a D65 (daylight) spectral power distribution. The large image in Figure 7a represents the calculated spectral irradiance pattern when imaging the grid lines with the Edmund Optics Finite Conjugate Micro Video Imaging Lens. Because the spatial spread of light varies with wavelength, the space-invariant input produces a pattern whose wavelength composition varies across space. Figure 7b shows the calculated spectral power distribution in two regions of the irradiance image. At the intersection of the grid lines, the spectral power distribution peaks close to 530 nm. The short wavelength light, however, is scattered more than middle-wavelength light. Thus, the space adjacent to the grid line intersections peaks closer to 460 nm.

The spatial variation of the spectral irradiance image could be a significant factor for a 3-μm pixel. Notice the 10-μm scale bar: the spatial variation introduced by the optics produces an inhomogeneous spectral image across the 2-by-2 Bayer pattern. Additional simulations show that these patterns are very salient in the reproduced image, and that such infelicities can be eliminated by using smaller pixels – at the expense of increased photon noise.

3. Microlens design

The irradiance image at the surface of a digital image sensor must travel down the pixel tunnel, passing through the sensor's back-end technology, to the photodetector in the silicon substrate. Methods for controlling the path of the photons between sensor surface and silicon substrate, a distance usually ranging from a few to more than 10 microns, are called the pixel optics. These methods are increasingly important in image sensor design as pixel size shrinks.[13, 18]

In many digital image sensors an array of microlenses is placed near each pixel aperture on the sensor surface. The position and properties of this microlens array control two important functions of the optical path from the sensor surface to the photodetector. First, the microlenses concentrate photons incident within the pixel aperture onto a compact region at the silicon substrate. In most CMOS imagers, the photodetector only occupies a fraction of the area at within the substrate because the photodetector shares substrate area with transistors that form the pixel circuitry. For example, in the widely used APS pixel three transistors perform tasks including selecting and resetting the pixel. Microlenses are intended to concentrate the photons onto the photodetector rather than allowing them to fall on non-photosensitive positions in the pixel floor. To concentrate photons the microlens properties must be coordinated with those of the imaging lens, as we discuss below.

Second, the microlens array redirects the light path. Consider the pixels at different positions across the sensor array. As pixel position varies from center to edge, the angular distribution of the rays arriving at the pixel changes. The

angular distribution is centered on the pixel aperture for pixels in the center of the sensor. But pixels near the edge receive a bundle of rays that would normally pass through the aperture and travel to substrate positions that are displaced from the pixel floor. These rays would strike the edge of the pixel tunnel or travel to regions within the floor of adjacent pixels. The microlens must redirect these rays so that they travel through the pixel tunnel and reach the photodetector position at the substrate. The loss of light due to the failure to redirect these rays is called pixel vignetting.[19] Redirection of the rays is achieved by appropriate placement of the microlenses with respect to the pixel aperture.

3.1. Light concentration

The light gathering efficiency of any optical system is summarized by a quantity called the geometrical extent (also known as etendue or light throughput).[20] It is the optical equivalent of the space-bandwidth product. The geometrical extent (G) of an optical system, say one that converts the scene radiance to image irradiance at the pixel aperture, is the product of the pixel aperture area with the solid angle subtended by the exit pupil of the imaging lens as seen by the pixel. For systems with radial symmetric or systems that are xy-separable, the geometrical extent can be expressed in terms of more familiar properties, i.e., as the product of the imaging lens numerical aperture (NA) with the size of the pixel aperture. In the paraxial approximation, NA and f/# are related by $NA = 1/(2f/\#)$ and G becomes,

$$G_{imaging} = w_{pixel} NA_{imaging} \cong \frac{w_{pixel}}{2(f/\#)_{imaging}}. \tag{7}$$

The geometrical extent of the pixel optics, that is the path from the sensor surface to the photodetector, is the product of the photodetector size and the NA of the microlens,

$$G_{pixel} = w_{detector} NA_{micro} \cong \frac{w_{detector}}{2(f/\#)_{micro}}. \tag{8}$$

In a system comprising several stages, such as the imaging lens and the pixel optics, light throughput is limited by the system component with the smallest geometrical extent. Generally for CMOS imagers, pixel optics is the limiting factor, $G_{pixel} < G_{imaging}$. The microlens is intended to increase the geometrical extent of the pixel optics to that of the imaging optics, $G_{pixel} = G_{imaging}$. For a pixel with a linear fill-factor of 50% ($w_{detector} = w_{pixel}/2$), the numerical aperture of the microlens should satisfy $NA_{micro} = 2NA_{imaging}$ or $(f/\#)_{micro} = (f/\#)_{imaging}/2$. In summary, the microlens should have an f/# that is smaller than that of the imaging lens by a factor that depends on the desired concentration.

The dilemma in pixel scaling for CMOS imagers is the following. Small pixel sizes require imaging lenses with small f/#s to obtain a sufficient photon supply at the sensor surface (see section Photon noise). Concentration of light onto the photosensitive area in small pixels requires a microlens with even smaller f/#. Given the small pixel aperture, the smaller f/# means shorter focal length. The focal length is determined by the sensor thickness (stack height). When CMOS technology scales, the lateral dimensions (e.g., transistor size and pixel size) shrink but the stack height usually increases. When does the stack height become too high to permit the microlens to concentrate the light from the imaging lens onto the photodetector? Figure 8a shows the upper limit on the concentration that can be achieved by a microlens as a function of stack height (microlens focal length) and pixel size (microlens diameter). The surfaces represent the best concentration that can be achieved assuming four different imaging lenses. For a 3-μm pixel whose detector size is half that of the pixel, we want a four-fold area concentration. Figure 8b shows the 4X equal-concentration curves. For this pixel the focal length of the microlens should be smaller than 6, 4.1, 2.75, and 1.75 μm for f/4.0, f/2.8, f/2.0, and f/1.4 imaging lenses, respectively. For image sensor design, it is advisable to avoid a trend towards more metal layers, since metal layers add to the stack height and consequently decrease the microlens f/#.

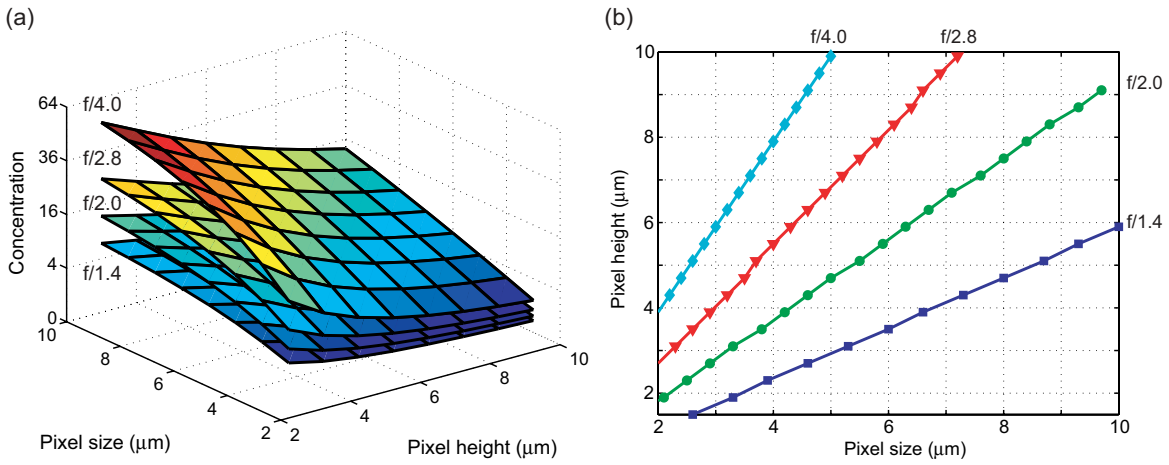

Figure 8: Light concentration. Simulations of the light concentration by microlenses measured as a function of focal length (pixel height) and diameter (pixel size). (a) Surfaces measuring the area concentration for four different imaging lenses with four different f/#s. (b) Equal concentration curves (4X area) as a function of pixel size and pixel height.

3.2. Light redirection

Given the difficulties associated with microlens design, the main purpose of the microlens array is no longer concentration. Rather, the most important function is the redirection of light onto the photodetector for off-axis pixels. To achieve proper redirection of light, the microlenses must be offset from the center of the pixel aperture.

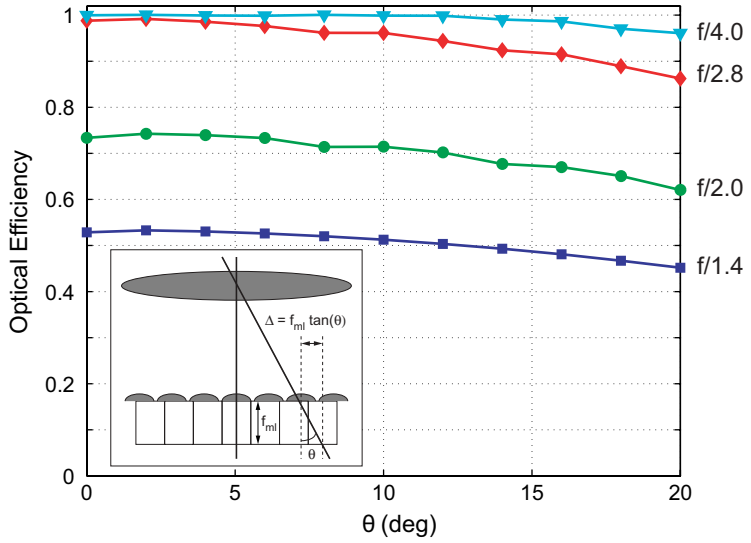

Figure 9. Light redirection. Simulations of light redirection for a 3-µm pixel with an 8-µm pixel height simulated for microlenses. The optical efficiency is measured as a function of the pixel location which is represented by angle of the chief ray in the material (θ). The simulations offset the microlens from the pixel center by an amount Δ computed as from the geometry shown in the inset. The curves measure efficiency for a series of diffraction-limited imaging lenses.

This offset, which is towards the optical axis, can be calculated to a very good approximation using first-order geometrical optics (see Figure 9 inset) and yields

$$\Delta = f_{ML} \tan(\theta), \tag{9}$$

where f_{ML} is the focal length of the microlens and θ is the chief ray angle inside the dielectric. Figure 9 shows the optical efficiency that is achieved for an ideal 3-µm pixel with an 8-µm stack height after this offset is applied for a series of diffraction-limited imaging lenses ranging from f/2.8 to f/8.0. Efficiency remains constant across the field of view and the level is almost equal to the optimal on-axis efficiency. Notice that the low optical efficiency for low f/#s of the imaging lens is caused by the mismatch in geometrical extent between imaging lens and microlens (f/2.7) as explained in the previous section Light concentration.

4. CONCLUSION

Sensors must interact with light: consequently the impact of technology scaling on imagers differs from its impact on other integrated circuit technologies. Using simulations, we quantified three consequences of decreasing pixel size on image quality. Decreasing pixel size increases the likelihood that photon noise will become visible, and we derive a noise-visibility contour on a map that combines photometric exposure and pixel size. Decreasing pixel size also changes the pixel pitch and thus influences spatial resolution. We analyze the implications of decreasing pixel size for aliasing in monochrome and color sensors. Finally, we show how decreasing pixel size restricts the effective use of microlens arrays by limiting the possible f/# of the microlens. We derive curves for the concentration and redirection of light within the pixel.

ACKNOWLEDGMENTS

The authors thank P. Maeda as well as the team at ImagEval, and particularly J. E. Farrell, for useful discussions. P.C. acknowledges partial support from a CIS FMA fellowship from Agilent Technologies, Inc.

REFERENCES

1. Meynants, G., et al., *A 14-megapixel 36 x 24-mm^2 image sensor*, in *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications*, M.M. Blouke, N. Sampat, and R.J. Motta, Editors. 2004, SPIE: San Jose. p. 168-174.
2. Moore, G.E., *Cramming more components onto integrated circuits.* Electronics, 1965. **38**(8).
3. ITRS, *International Roadmap for Semiconductors, 2004 Edition, Executive Summary*. 2004.
4. Chen, T., et al., *How Small Should Pixel Size be?* in *Sensors and Camera Systems for Scientific, Industrial and Digital Photography Applications*, M.M. Blouke, et al., Editors. 2000, SPIE: San Jose. p. 451-459.
5. Planck, M., *Über das Gesetz der Energieverteilung in Normalspektrum.* Ann. d. Phys., 1901. **4**: p. 553.
6. Sommerfeld, A., *Mathematische theorie der diffraction.* Math. Ann., 1896. **47**: p. 317-374.
7. Barnhoefer, U., et al., *Color estimation error trade-offs*, in *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications IV*, M.M. Blouke, N. Sampat, and R.J. Motta, Editors. 2003, SPIE: San Jose. p. 263-273.
8. Xiao, F., J. Farrell, and B. Wandell, *Psychophysical thresholds and digital camera sensitivity: the thousand photon limit*, in *Digital Photography*. 2005, SPIE: San Jose.
9. Farrell, J.E., et al., *A simulation tool for evaluating digital camera image quality*, in *Image Quality and System Performance*, Y. Miyake and D.R. Rasmussen, Editors. 2003, SPIE: San Jose. p. 124-131.
10. Klein, M.V. and T.E. Furtak, *Optics*. 2nd ed. 1986, New York: Wiley.
11. Goodman, J.W., *Introduction to Fourier optics*. 2nd ed. 1996, New York: McGraw-Hill.
12. Maeda, P., P.B. Catrysse, and B.A. Wandell, *Integrating Lens Design with Digital Camera Simulation*, in *Digital Photography*. 2005, SPIE: San Jose.
13. Catrysse, P.B. and B.A. Wandell, *The optical efficiency of image sensor pixels.* Journal of the Optical Society of America A, 2002. **19**(8): p. 1610-1620.

14. McCamy, C.S., H. Marcus, and J.G. Davidson, *A color-rendition chart.* J. Applied Photographic, 1976. **48**: p. 777-784.
15. Rhodes, H., et al., *CMOS imager technology shrinks and image performance*, in *2004 IEEE Workshop on Microelectronics and Electron Devices*. 2004, IEEE. p. 7-18.
16. Wong, H.-S., *Technology and device scaling considerations for CMOS imagers.* IEEE Transactions on Electron Devices, 1996. **43**(12): p. 2131-2142.
17. Edmund Industrial Optics, *Optics and Optical Instruments Catalog*. 2004, New Jersey: Edmund Optics Inc.
18. Catrysse, P.B. and B.A. Wandell, *Integrated color pixels in 0.18-µm complementary metal oxide semiconductor technology.* Journal of the Optical Society of America A, 2003. **20**(12): p. 2293-2306.
19. Catrysse, P.B., X. Liu, and A. El Gamal, *Quantum efficiency reduction due to pixel vignetting in CMOS image sensors*, in *Sensors and Camera Systems for Scientific, Industrial and Digital Photography Applications*, M.M. Blouke, et al., Editors. 2000, SPIE: San Jose. p. 420-430.
20. Steel, W.H., *Luminosity, throughput, or etendue.* Applied Optics, 1974. **13**: p. 704-705.