

Copyright 2008 SPIE and IS&T.

This paper was published in *Electronic Imaging, Digital Photography IV* and is made available as an electronic reprint with permission of SPIE and IS&T. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.

# Does resolution really increase image quality?

Christel-Loïc Tisse, Frédéric Guichard, Frédéric Cao  
DxO Labs, 3 Rue Nationale, 92100 Boulogne, France

## ABSTRACT

A general trend in the CMOS image sensor market is for increasing resolution (by having a larger number of pixels) while keeping a small form factor by shrinking photosite size. This article discusses the impact of this trend on some of the main attributes of image quality. The first example is image sharpness. A smaller pitch theoretically allows a larger limiting resolution which is derived from the Modulation Transfer Function (MTF). But recent sensor technologies ( $1.75\mu\text{m}$ , and soon  $1.45\mu\text{m}$ ) with typical aperture  $f/2.8$  are clearly reaching the size of the diffraction blur spot. A second example is the impact on pixel light sensitivity and image sensor noise. For photonic noise, the Signal-to-Noise-Ratio (SNR) is typically a decreasing function of the resolution. To evaluate whether shrinking pixel size could be beneficial to the image quality, the tradeoff between spatial resolution and light sensitivity is examined by comparing the image information capacity of sensors with varying pixel size. A theoretical analysis that takes into consideration measured and predictive models of pixel performance degradation and improvement associated with CMOS imager technology scaling, is presented. This analysis is completed by a benchmarking of recent commercial sensors with different pixel technologies.

**Keywords:** Signal-to-Noise-Ratio (SNR), Modulation Transfer Function (MTF), tonal range, spatial resolution, information transfer capacity, pixel size, CMOS APS image sensor.

## 1. INTRODUCTION

In response to growing consumer demand for higher resolution and more compact digital cameras in mobile phones, the pixels in CMOS image sensors have become smaller. This reduction of pixel size is being made possible by CMOS and micro-optics technologies scaling. Nowadays state-of-the-art imager design rules scale down into the sub-micron regime (*i.e.*  $0.18\mu\text{m}$  -  $0.09\mu\text{m}$ ), and pixel size can be as small as  $1.45\mu\text{m} \times 1.45\mu\text{m}$ . Unfortunately, technology scaling has detrimental effects on pixel performance. Smaller pixels have worse light-gathering ability and more non-idealities. As a result, reducing pixel size and increasing pixel count (*i.e.* the number of pixels in the image) while keeping the size of an imaging sensor array fixed, does not always yield a better image quality.

Spatial resolution and light sensitivity are two fundamental characteristics of image sensor that must be considered for characterizing and optimizing image quality. These characteristics are generally obtained from the Modulation Transfer Function (MTF) and the system Signal-to-Noise-Ratio (SNR). In Section 2 we describe the effects of technology scaling on a variety of pixel properties for conventional active pixel sensors (APS). To better understand how these parameters influence the measures of MTF and SNR, we show some simulations by using an extensive model of the performance of CMOS imager pixels from  $5.2\mu\text{m}$  to  $1.45\mu\text{m}$ . We will see that, even though changing pixel size clearly has opposing effects on MTF and SNR curves, it is difficult to examine the image quality tradeoff between spatial resolution and noise directly from these measurements. In Section 3 we introduce the notion of image information capacity for determining the optimal pixel size. Image information capacity quantifies the maximum visual information that a sensor could optimally convey from object to image, and is an objective measure of image quality. Our theoretical analysis is completed in Section 4 by the comparison of the image information capacity of commercial sensors using  $2.8\mu\text{m}$ ,  $2.2\mu\text{m}$  and  $1.75\mu\text{m}$  pixels.

## 2. SENSOR PERFORMANCE

### 2.1 Trends in Pixel Design

Active Pixel Sensor (APS) is the most popular type of CMOS imager architectures. The APS pixels under consideration in this paper are: (*i*) the 4-T type pinned photodiode with Correlated-Double-Sampling (CDS); the 4-T pixel adds a

transfer gate and a Floating Diffusion (FD) node to the reset, source follower, and row select (or read) transistors of the basic 3-T pixel; (ii) the 2.5-T pixel, where the buffer of the 4-T design is shared between two adjacent pixels; (iii) the 1.75-T pixel architecture,<sup>1,13</sup> in which four neighboring pixels share these same transistors; and (iv) the 1.5-T pixel,<sup>13</sup> in which four pinned photodiodes share only reset and source follower transistors, the read transistor being removed. Sharing transistors improves the fill factor for the APS structure, and is a slight counterbalance to the photodiode process implants increase necessary for preserving the full well capacity ( $E_{FullWell}$  in electrons) of a smaller photodiode area.

## 2.2 Performance Measures and Modeling

The Optical Efficiency (OE), which characterizes the photon-to-photon efficiency from the pixel surface to the photodetector, is affected as CMOS process technology scales to smaller planar feature size. The optical tunnel through which light must travel before reaching the photodetector becomes narrower, but its depth does not scale as much. The pixels' angular response performance to incident light decreases because of longer focal length of the micro-lens that focuses the incoming light onto the photodiode.<sup>4</sup> This phenomenon is also known as *pixel vignetting*. Experimental evidence<sup>2,3</sup> and electromagnetic simulations<sup>5</sup> using new tools based on Finite-Difference Time-Domain<sup>7</sup> (FDTD) show that pixel vignetting becomes extremely severe as technology scales down, which results in significant OE reduction from about 35-40 percent for 3.2 $\mu$ m off-axis pixels to more than 75 percent for 1.45 $\mu$ m off-axis pixels (50 percent fill factor) with light incident at a 20° angle. The pixel aperture width and the structure of the interconnections stack are also critical limiting factors of photon collection inside pixels due to the dominant diffractive effect of light on sub-wavelength scales and the spatial crosstalk arising from light propagation between adjacent pixels, respectively.

The internal Quantum Efficiency (QE), which refers to the conversion efficiency from photons incident on the photodiode surface to photocurrent, is a function mainly of metallurgical process parameters (*e.g.* doping densities) and photodiode geometry. QE varies very little as photodiode dimensions shrink.<sup>2</sup> It is important to note, however, that the pixel photo-response is not flat over the visible spectrum, and the internal QE actually shifts toward shorter wavelengths as junction depth gets shallower.<sup>8</sup>

In addition to lower OE, and lower internal QE, smaller pixels cause higher photon shot noise (inherent to the stochastic nature of the incident photon flux, governed by Poisson statistics), and have higher leakage signals and more non-uniformities. We follow A.E.Gamal<sup>9</sup> *et al.* for describing the different temporal and spatial noise processes associated with these non-idealities and for modeling their impact on sensor Signal-to-Noise-Ratio (SNR) and Dynamic Range (DR). As a function of the photocurrent  $E_S$  in electrons [e<sup>-</sup>], the SNR in decibels (dB) is

$$SNR_{dB}(E_S) = 10 \log_{10} \frac{P_{Signal}}{P_{Noise}} \approx 10 \log_{10} \frac{E_S^2}{\sigma_S^2 + \sigma_{DC}^2 + \sigma_{READ}^2 + \sigma_{DSNU}^2 + \sigma_{PRNU}^2 + \sigma_{Quantization}^2}, \quad (1)$$

- where
- $P_{Signal}$  is the input signal power.
  - $P_{Noise}$  is the average input referred noise power.
  - $\sigma_S^2$  ( $\approx E_S$ ) is the photon shot noise average power, which is signal dependent.
  - $\sigma_{DC}^2$  ( $\approx E_{DC}$ ) is the power of the Dark Current ( $E_{DC}$ ) shot noise arising from the statistical variation (*i.e.* Poisson distribution) over time on the number of dark current generated electrons  $E_{DC}$ .
  - $\sigma_{READ}^2$  ( $\approx \sigma_{Reset}^2 + \sigma_{Readout}^2 + \sigma_{FPN}^2$ ) is the read noise power;  $\sigma_{Read}$  combines (i) pixel reset circuit noise  $\sigma_{Reset}$ , also known as  $kTC$  noise, (ii) readout circuit noise  $\sigma_{Readout}$  due to thermal and flicker noise whose spectrum is inverse proportional to the frequency in MOS transistors, and (iii) offset Fixed Pattern Noise (FPN)  $\sigma_{FPN}$  due to device mismatches; in the 4-T APS architecture, the major part of reset noise and FPN noise is eliminated by CDS, but this requires that the time between the two CDS sampling moments to be short enough to ensure the maximum correlation between the flicker noise components of the samples.
  - $\sigma_{DSNU}^2$  is the Dark Signal Non Uniformity (DSNU) noise power; DSNU noise results from the fact that each pixel generates a slightly different amount of dark current under identical illumination.
  - $\sigma_{PRNU}^2$  is the Photo Response Non Uniformity (PRNU) noise power; PRNU noise  $\sigma_{PRNU}$ , commonly known as gain FPN, describes the pixel-to-pixel gain variation across the image sensor array under uniform illumination;  $\sigma_{PRNU}$  ( $\approx K_{PRNU} \times E_S$ ) is signal dependent and often expressed as a percentage  $K_{PRNU}$  of the average image signal; it mainly affects sensor performance under high illumination.

- $\sigma_{\text{Quantization}}^2 (\approx K^2/12)$  is the quantization noise power that arises from the discrete nature of an  $n$ -bit analog-to-digital conversion; the quantization noise is proportional to the *sensor conversion gain*  $K (\approx E_{\text{FullWell}} / 2^n)$  in  $[e^-]$  per digit number (DN).

All of noise powers in Eq. (1) are measured in  $[e^-]^2$ . A classification between *temporal* and *spatial* noise sources distinguishes (i) photon shot noise, DC shot noise, reset noise, readout circuit noise, and quantization noise from (ii) offset FPN noise, DCNU noise, and PRNU noise, respectively. Temporal noise and spatial noise also determines DR, which quantifies the sensor's ability to detect a wide range of illumination in a scene. DR is expressed as the ratio of the largest non-saturating input signal  $E_{\text{Max}}$  to the smallest detectable input signal  $E_{\text{Min}}$  (*i.e.* noise floor under dark conditions) as follows

$$DR_{dB} = 20 \log_{10} \frac{E_{\text{Max}}}{E_{\text{Min}}} = 20 \log_{10} \frac{E_{\text{FullWell}} - \sigma_{DC}}{\sqrt{\sigma_{DC}^2 + \sigma_{\text{READ}}^2 + \sigma_{\text{DSNU}}^2 + \sigma_{\text{Quantization}}^2}}. \quad (2)$$

DR decreases with full well capacity (and inevitably with pixel size) and as exposure time  $\Delta t$  and/or temperature is/are increased. This is because dark current is a linear function of  $\Delta t$  and is roughly doubling every  $6^\circ\text{C}$  (dark current performance measured for  $3.2\mu\text{m}$  pitch 2.5-T pixel).<sup>2</sup>

Spatial resolution is another critical aspect of image sensor performance. An image sensor performs spatial sampling of the input image projected by the lens onto its (rectangular) pixel array, *i.e.* the focal plane. Assuming an ideal thin lens, the focal plane would result in a perfectly sharp (digital) image. However, photosites are not infinitely small, which implies an intrinsic limit to spatial resolution described by Nyquist (uniform) sampling theory. Spatial resolution below the Nyquist spatial frequency ( $f_N \approx (2 \times \text{Pixel Pitch})^{-1}$  in line pairs or cycles per millimeter) to avoid aliasing and Moiré patterns is measured by the Modulation Transfer Function (MTF). The MTF is mathematically related to the Pixel Response Function (PRF) by calculating the magnitude of its Fourier Transform in a given direction. Several parameters degrade the detector MTF by causing low-pass filtering. The pixel active area geometrical shape (or pixel aperture area) with electronic crosstalk (*i.e.* photocarrier diffusion effect) and optical crosstalk are the main determining factors of the overall detector MTF.<sup>10</sup> For sake of simplicity, a first-order approximation of sensor MTF is obtained by considering only the ideal geometrical PRF (*i.e.* uniform pixel sensitivity within the active area) convolved with an anisotropic (Gaussian or exponential-type) blur filter. The two-dimensional MTF for a traditional *L-shaped* pixel design is then given by

$$MTF_{2D}(w_1, w_2) = \left| G_\sigma(w_1, w_2) \right| \cdot \left| \frac{a(B-b)}{AB-ab} \cdot \frac{\sin(\frac{1}{2} \cdot a \cdot w_1)}{\frac{1}{2} \cdot a \cdot w_1} \cdot \frac{\sin(\frac{1}{2} \cdot (B-b) \cdot w_2)}{\frac{1}{2} \cdot (B-b) \cdot w_2} \cdot e^{+j\frac{1}{2}aw_1} \cdot e^{-j\frac{1}{2}(B-b)w_2} + \frac{B(A-a)}{AB-ab} \cdot \frac{\sin(\frac{1}{2} \cdot (A-a) \cdot w_1)}{\frac{1}{2} \cdot (A-a) \cdot w_1} \cdot \frac{\sin(\frac{1}{2} \cdot B \cdot w_2)}{\frac{1}{2} \cdot B \cdot w_2} \cdot e^{-j\frac{1}{2}(A-a)w_1} \cdot e^{-j(\frac{1}{2}B-b)w_2} \right|, \quad (3)$$

where •  $w_i (=2\pi f_i)$  is the angular frequency in radians per pixel.

- $A, B, a$  and  $b$  are the dimensions of the L-shaped active area, as described in Figure 1(a), with  $a \leq A \leq \text{Pixel Pitch } (P)$  and  $b \leq B \leq P$ .
- $G_\sigma(w_1, w_2)$  is the frequency response of the Gaussian convolution kernel filter with standard deviation  $\sigma$ .

Eq. (3) shows that the modeled MTF of an *L-shaped* pixel is symmetrical about the DC component, but it is not isotropic. This is illustrated in Figures 1(b-c). Note that the Nyquist frequency increases for small pixel size. The result is an improvement of detector MTF and higher spatial resolution.

In summary, for a fixed sensor die size, smaller pixels theoretically allow a higher spatial resolution but have more non-idealities and worse light sensitivity, and consequently lower DR and SNR performances. Some of the advances in image sensor technologies described above have made it possible to partially compensate for such noise performance degradation. In the next subsection we take into consideration both existing and predictive models of APS pixel performance associated with CMOS imager technology scaling to simulate detector MTF and SNR versus pixel size.

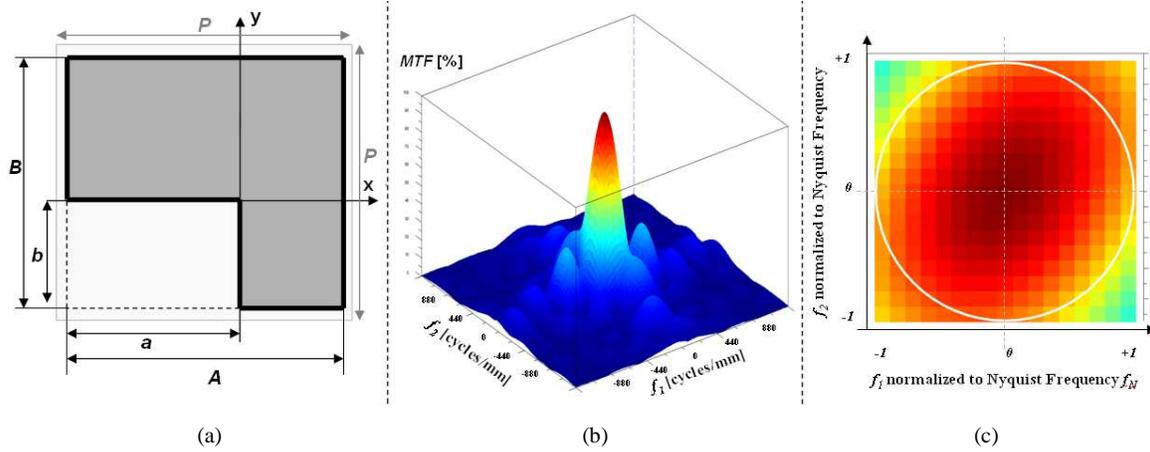


Fig. 1. (a) Layout description of an L-shaped pixel design; (b) 2D MTF simulation for  $P = 3.2\mu\text{m}$  pixel size with the dimensions  $A = B \approx P$ ,  $a \approx 0.55 \times P$  and  $b \approx 0.28 \times P$ ; this simulation assumes no crosstalk between pixels ( $\sigma = 0$ ); (c) Same MTF with spatial frequency normalized to the Nyquist frequency  $f_N = (2P)^{-1}$ ; in the general case where  $a \neq b$ , note the anisotropy of the detector MTF.

### 2.3 Simulations and Predictive Performance

In our SNR simulations we first estimate the mean number of photons  $\eta_{\text{photons}}$  incident on a single pixel (per exposure interval  $\Delta t$  in seconds) as function of pixel size  $P$  and photometric exposure  $H$  (in lux.s), through the following equation:

$$\eta_{\text{photons}}(\lambda, \Delta t) = \frac{H(\lambda, \Delta t)}{K_m(\lambda) E_{\text{photon}}(\lambda)} \cdot \alpha P^2, \quad (4)$$

- where
- $\alpha$  is the pixel fill factor ( $0 < \alpha \leq 1$ ).
  - $P^2$  is the pixel area (in  $\text{m}^2$ ).
  - $K_m$  is the ratio between luminous flux and energetic flux;  $K_m \approx 683 \text{ lm/W}$  for a wavelength  $\lambda = 555 \text{ nm}$ .<sup>11</sup>
  - $E_{\text{photon}} (= h \nu)$  is the energy of a photon (in Joules), equal to the product of Planck's constant  $h$  and the optical frequency  $\nu$ ;  $E_{\text{photon}} \approx 2.58 \cdot 10^{-19} \text{ J}$  for a wavelength  $\lambda = 555 \text{ nm}$ .

Assuming that the surface of object(s) in the scene (depicted by the camera system) is Lambertian, photometric exposure  $H$  can be described in a similar way to Cartrysse *et al.*<sup>12</sup> by

$$H(\lambda, \Delta t) = \frac{T_{\text{lens}}(\lambda)}{1 + 4(1-m)^2 (f/\#)^2} \cdot R(\lambda) E_{\text{scene}} \cdot \Delta t, \quad (5)$$

- where
- $T_{\text{lens}}$  is the spectral transmittance of the lens ( $0 < T_{\text{lens}} \leq 1$ ).
  - $m (< 0)$  is the magnification of the lens.
  - $f/\# (= f/D)$  is the relative aperture (or f-number) of the imaging system, equal to the ratio of the focal length  $f$  to the circular aperture diameter  $D$ .
  - $R$  is the coefficient of reflection ( $0 < R \leq 1$ ).

Figure 2(a) shows the mean number of photons per pixel for typical values of scene illuminance  $E_{\text{scenes}}$ , from  $10$  to  $10^4$  lux, assuming that all photons in the visible range have roughly the same energy (estimation for  $\lambda = 555 \text{ nm}$ ). This estimation is obtained for ideal pixels (*i.e.* fill factor  $\alpha = 1$ ) with exposure time  $\Delta t = 100 \text{ ms}$ , and for an  $f/2.8$  lens with magnification  $m = -10^{-3}$  and transmittance  $T_{\text{lens}} = 0.85$ .

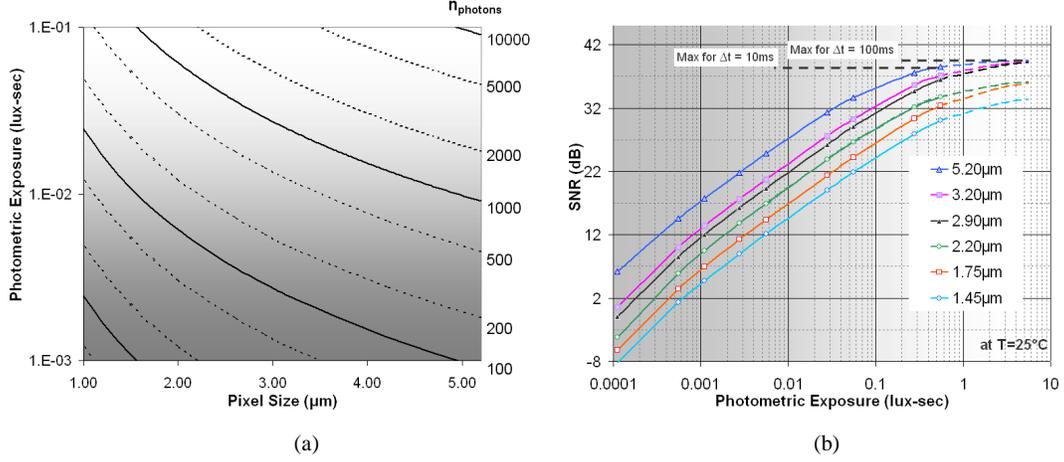


Fig. 2. (a) Mean incident photon level per pixel for different pixel sizes and a photometric exposure range that covers low to high illuminance level conditions; (b) SNR as a function of photometric exposure for 10-bit image sensors with different pixel size.

Based on this predicted number of photons per pixel and using Eq. (1) we simulate the sensor SNR for different pixel sizes. The simulation results are plotted in Figure 2(b) for a set of typical pixel parameters that are listed below in Table 1. The pixel performance parameters are derived from Rhodes<sup>2</sup> *et al.*, Cohen<sup>13</sup> *et al.* and Pain<sup>14</sup>. The contributions of read noise and DSNU noise are assumed negligible (*e.g.*  $\sigma_{\text{DSNU}} \leq 0.5\% \times E_{\text{DC}}$ ). The comparative examination of these SNR plots confirms that SNR decreases with pixel size. For photometric exposure from  $10^{-3}$  to  $10^{-1}$  lux.s, photon shot noise is dominant and SNRs increase with photometric exposure at 10dB/decade. Within this photon-noise-limited region, the smallest pixel results in an SNR approximately 10dB lower than that of the largest pixel. At low signal levels, the slope difference between the SNR curves indicates that small pixels are also more sensitive to dark current than large ones. At high level signals, SNR curves flatten out when PRNU dominates. The dashed curves illustrate that peak SNR increases with integration time (until capacity-well saturation). In practice an upper limit on the integration time is dictated by how much loss of contrast information (*cf.* DR) and motion blur artifact can be tolerated in the captured image.

Table 1. Set of typical pixel parameters used in our simulations; these data which are derived from Rhodes<sup>2</sup> *et al.*<sup>\*</sup>, Cohen<sup>13</sup> *et al.*<sup>\*\*</sup> and Pain<sup>14</sup>, include measured and predictive properties of (4T, 2.5T, 1.75T and 1.5T) APS pixels; sensitivity for the 1.45 $\mu\text{m}$  pixel (shown in *italic*) was obtained by creating an empirical model that takes into account OE reduction as discussed in Section 2.2.

Pixel Pitch ( $\mu\text{m}$ )	Full Well ( $\text{ke}^-$ )	Sensitivity ( $\text{ke}^-/\text{lux.s}$ )	Dark Current ( $\text{e}^-/\text{s}$ ) at 25°C / 60°C	PRNU (%)	
5.2 (4T)	28 – 38 <sup>*</sup>	57 <sup>*</sup>	250 – 22 <sup>*</sup>	2000 <sup>*</sup>	<1.05 <sup>*</sup>
3.2 (2.5T)	22 – 33 <sup>*</sup>	22 <sup>*</sup>	65 – 12 <sup>*</sup>	360 <sup>*</sup>	<1.05 <sup>*</sup>
2.90 (2.5T)	19 – 25 <sup>*</sup>	16 <sup>*</sup>	30 – 10 <sup>*</sup>	300 <sup>*</sup>	<1.05 <sup>*</sup>
2.20 (2.5T)	12 – 21 <sup>*</sup>	9.3 <sup>*</sup>	25 – 9 <sup>*</sup>	270 <sup>*</sup>	<1.5
1.75 (1.75T)	9 – 8 <sup>**</sup>	5 <sup>**</sup>	20	25 <sup>**</sup>	<1.5 <sup>**</sup>
1.45 (1.5T)	7 – 4 <sup>**</sup>	3	18	15 <sup>**</sup>	<2.0

We now compare sensor spatial resolution for different pixel sizes. Because, independently of the photosite geometrical shape, the amount of frequency response degradation due to pixel size increase is anisotropic, it can be plotted for one arbitrary direction without loss of generality. In Figure 3 the curved lines define the detector MTFs as a function of *vertical* input spatial frequency. The simulation results are again for ideal square pixels. As expected, for a fixed die size and a fixed imaging optics, sensors with (more) smaller pixels are capable of capturing higher spatial frequencies and are better at preserving thin details. In Figure 3 we also compare the influence of the detector on the overall MTF of the imaging system with that of a diffraction-limited lens operating at  $f/2.8$ . This comparison shows that the effect of diffraction of light becomes a limiting factor of the spatial resolution in image sensors with pixels smaller than 2.2 $\mu\text{m}$ .

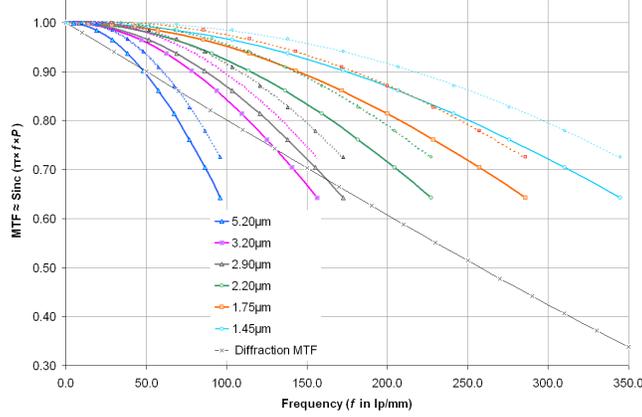


Fig. 3. Slice along the (y) vertical array direction of the imaging detector geometrical MTF with different pixel size (fill factor  $\alpha = 1$ ); the extra dashed curves show comparative MTFs along the same (y) direction for sensors with L-shaped pixels described in Figure 1 (fill factor  $\alpha = \text{constant} \approx 0.45 \times 0.72$ ); *Diffraction MTF* represents the frequency response expected from a perfect, diffraction-limited lens operating at  $f/2.8$  ( $\text{Diff. MTF} \approx 2/\pi \times [\arcsos(f/f_0) - (f/f_0) \times (1 - (f/f_0)^2)^{-1/2}]$  with spatial cut-off frequency  $f_0 \approx (\lambda \times f/\#)^{-1}$ ).

Our theoretical performance analysis of image sensors with varying pixel size shows an inherent difficulty in comparing the SNR and MTF curves to determine the optimal pixel size. The proposed metrics so far do not summarise to a scalar output which makes the tradeoffs between light sensitivity and spatial resolution still depend on many factors. Farrell<sup>15</sup> *et al.* suggested comparing the pixel performance by (i) applying a psychological threshold for the SNR, referred to as MPE30, and (ii) selecting the commonly used value MTF50 for the MTF. The MPE30 metric corresponds to the minimum required photometric exposure to render (uncorrelated) photon shot noise invisible in an image of uniform field, in other words such that  $\text{SNR}(H = \text{MPE30}) \geq 30\text{dB}$ . The MTF50 metric is used to quantify the amount of perceived image blur. A tradeoff function is obtained by plotting MTF50 against  $1/\text{MPE30}$  for each of the simulated sensors. It turns out that this monotone decreasing tradeoff function is not sufficient to identify an obvious optimal pixel size. In the next section we use image information capacity as figure of merit of image sensors with varying pixel size.

### 3. METHODOLOGY FOR QUANTIFYING IMAGE INFORMATION TRADEOFFS

Following Farrell<sup>15</sup> *et al.*'s approach, we can distinguish two different types of image *distortions* associated with the process of pixel size reduction. The first one is an increase of the amount of visible noise in the image. The other one is a decrease of the amount of image blur. These two phenomena of noise addition and image blurring are usually considered in terms of undesirable (spatial and temporal) variation in pixel intensity values and linear low-pass filtering in the spatial domain, respectively. However, an information theoretical viewpoint can be taken instead where the pertinent criterion for pixel size scaling optimization is the maximum image information capacity  $C$  (in bits) that the sensor could optimally convey. For instance, the limit of information capacity in a perfectly sharp, noise-free image (captured with an ideal imaging sensor) is simply the number of pixels of the sensor  $s$  multiplied by its quantization resolution  $b$  (number of bits per pixel). Note the analogy here with the Shannon formula<sup>16</sup> for the transmission capacity of a discrete noiseless communication channel.

$$C \leq s \times b. \quad (6)$$

Let us first consider the noisy case in which a very thin grey level quantization may become irrelevant if it is much smaller than noise. The effects of the noise can be considered by substituting  $b$  into Eq. (6) by the number of bits  $b'$  ( $\leq b$ ) necessary to encode all the distinguishable grey levels. The information quantity  $b'$  is also known as Tonal Range ( $\text{TR} = \log_2^{-1}(b')$ ) which characterizes the effective number of grey levels of the imaging system. Tonal range is computed through the Riemann integral

$$\text{TR} = \int_{H_{\min}}^{H_{\max}} \frac{1}{\max(\sigma_{\text{noise}}(H), 1)} dH, \quad (7)$$

where  $\sigma_{noise} (= (\sigma_S^2 + \sigma_{DC}^2 + \dots)^{1/2})$  is the standard deviation of the overall noise of the image sensor. The interval of integration  $\Delta H (= H_{max} - H_{min})$  over photometric exposure  $H$  corresponds to the dynamic range of the sensor.

Let now address the effects of image blurring on maximum image information capacity. Image blur can be interpreted as another (channel) constraint which increases statistical correlation among neighboring image points. We expect this constraint to specifically affect the available information transfer rate from object to image, *i.e.* the effective imaging spatial resolution of the sensor  $s'$ . By the reasoning<sup>17,18</sup> which led Shannon<sup>16</sup> to the theorem of entropy change in linear filters, we derive that the two-dimensional spatial resolution loss  $\Delta s$ , for low-pass filter with characteristic  $OTF(w_1, w_2) = MTF(w_1, w_2) \times e^{j\phi(w_1, w_2)}$ , obeys

$$\Delta s = (4f_{N1}f_{N2})^{-1} \int_{-f_{N1}}^{f_{N1}} \int_{-f_{N2}}^{f_{N2}} \log_2(MTF^2(w_1, w_2)) \cdot dw_1 \cdot dw_2, \quad (8)$$

where  $f_{N1}$  and  $f_{N2}$  are Nyquist sampling frequencies in the horizontal and vertical directions (usually  $f_{N1} = f_{N2}$ ),  $MTF(w_1, w_2)$  has nonzero values over the image spectrum, and  $\Delta s (\leq 0)$  is expressed in bits. A traditional method for characterizing the MTF of an image sensor is to measure its spatial frequency response (SFR) to both slanted vertical and horizontal black and white edges (*cf.* ISO standard 12233). These measures are performed in the centre of the FOV, and vertical and horizontal SFRs are averaged to estimate the overall sensor MTF. Although inaccurate - we have demonstrated above that detector MTF can be anisotropic - this one-dimensional MTF gives often a good approximation of the sensor spatial resolution capability in all directions. Assuming that the two-dimensional MTF is now circularly symmetric, the domain of integration in Eq. (8) is also circular since there is no preferred direction of modulation. From the one-dimensional MTF measurement and simplification of Eq. (8), the two-dimensional spatial resolution loss (in bits) is approximated by

$$\Delta s = \pi f_N^{-2} \int_0^{f_N} w \log_2(MTF(w)) \cdot dw. \quad (9)$$

Finally, an upper limit estimate of the image information capacity is obtained by substituting  $s$  into Eq. (6) by the number of effective pixels  $s' (= s \times 2^{\Delta s})$ .

As an example of the information capacity limit for image sensors with different pixel sizes, consider the pixel parameters of Table-1 and MTFs plotted in Figure 3 (see previous section). Just as SNR in image sensors tends to increase as a function of their pixel size and exposure time, the same is true of TR. However, for pixels with same fill factor and active area of the same shape, (geometrical) MTFs with spatial frequency normalized to Nyquist frequency are similar. This means that the relative spatial resolution loss factor  $2^{\Delta s}$  is theoretically independent of pixel size. The image information capacity is essentially limited in this case by sensor resolution, TR and diffraction. In fact, given a constant optical format (e.g. 1/4-inch, corresponding to the diagonal dimension of the imaging area), the number of pixels is inversely proportional to the square of the pixel pitch, whereas TR is typically only about 1.3 bit higher when the pixel pitch is more than tripled from  $1.45\mu\text{m}$  to  $5.2\mu\text{m}$ . Consequently, the image information capacity of the sensor increases (for a fixed die size) as the pixel pitch decreases down to  $1.45\mu\text{m}$  despite the effects of diffraction. Figure 3 illustrates this trend for pixels down to  $1.45\mu\text{m}$  and a perfect (diffraction-limited)  $f/2.8$  lens, and then extends the predictive model of image information capacity down to  $1\mu\text{m}$  pixels by assuming that TR continues to decrease more or less linearly with pixel pitch at approximately  $0.35 \text{ bit}/\mu\text{m}$  (*cf.* dashed trend-line). Note that this is a reasonable assumption as long as the micro-lens has the ability to efficiently focus light onto the photodiode area. According to our prediction results, an optimal pixel size that maximizes the information capacity of the sensor is found for  $P \approx 1.45\mu\text{m}$ . In other words, even under the assumption of *ideal* pixels with a higher OE than predicted by FDTD analysis, shrinking pixel size beyond this  $1.45\mu\text{m}$  limit will lead to reduced performance.

In our theoretical analysis for quantifying image information tradeoffs between blur and noise, we have relied on a number of hypotheses and simplifications regarding the technological properties of pixels. In the next section, off-the-shelf commercial image sensors with different pixel sizes are compared to validate our simulations and to determine whether existing pixels as small as  $1.75\mu\text{m}$  pitch (or possibly smaller) can indeed lead to a higher image information capacity than larger pixels, or if the optimal pixel size has already been reached.

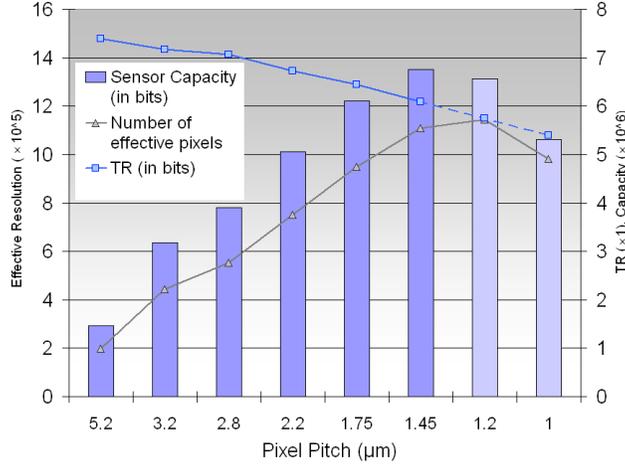


Fig. 4. Image information capacity of the sensor as a function of pixel size for a fixed 1/4-inch imaging area and a perfect (diffraction-limited)  $f/2.8$  lens; TR is given for a 8-bit equivalent grayscale.

#### 4. BENCHMARKING OF COMMERCIAL IMAGE SENSORS

We present here the benchmarking of five commercial CMOS (color) image sensors produced by two of the world’s leading suppliers. We refer to these two suppliers as M1 and M2, respectively. Below in Table 2 is a brief description of the characteristics of the sensors. The pixel size varies between sensors from  $1.75\mu\text{m}$  to  $2.80\mu\text{m}$ . All of the noise and MTF measurements were conducted in RAW format using *DxO Analyzer*.<sup>19</sup> Only the measure values for the green channel and for pixels at the center of the sensor array (*i.e.* on-optical-axis) are reported. To obtain accurate comparable measures of SNR and detector MTF, the sensors under test were mounted with identical lenses with known aperture and optical MTF. The performance of each lens was provided by *TRIOPTICS* measurements.<sup>20</sup> The detector MTF was found by dividing the overall MTF of the resultant imaging system by the lens MTF. The effective exposure times of the sensors were determined by imaging an external LED-panel-based device where LEDs are successively illuminated for a defined time and can be counted in the picture taken.

Table 2. Main characteristics of the sensors used for the benchmarking.

Designation	Manufacturer	Pixel pitch ( $\mu\text{m}$ )	Resolution (pixels)	Optical format (inch)
M2_2.80 $\mu\text{m}$	M2	2.80	1600 × 1200	1/3
M2_2.20 $\mu\text{m}$	M2	2.20	2056 × 1544	1/3.2
M1_2.20 $\mu\text{m}$	M1	2.20	2048 × 1536	1/3.2
M2_1.75 $\mu\text{m}$	M2	1.75	2048 × 1536	1/4
M1_1.75 $\mu\text{m}$	M1	1.75	2048 × 1536	1/4

A SNR performance comparison of image sensors by manufacturer is displayed in Figure 5. For both manufacturers M1 and M2, the SNR of the sensor with the largest pixel is the best as expected. The sensors of Manufacturer M2 perform differently depending on exposure time duration; this is in part due to the presence of a dark current compensation circuit that operates when the *analog* gain (to adjust the sensor sensitivity and conversion factor) is increased in low light conditions.

It is important to note however the disparity in noise performance between image sensors with same pixel size but from different manufacturers, as shown in Figure 6. In this comparison, we included additional sensors from a third manufacturer M3. We also included older sensor versions from manufacturers M1 and M2 using  $1.75\mu\text{m}$  and  $2.2\mu\text{m}$  pixels (referred to as “*bis*”). The gap in SNR performance at mid-dynamic range between image sensors of the same generation can be as high as 5dB.

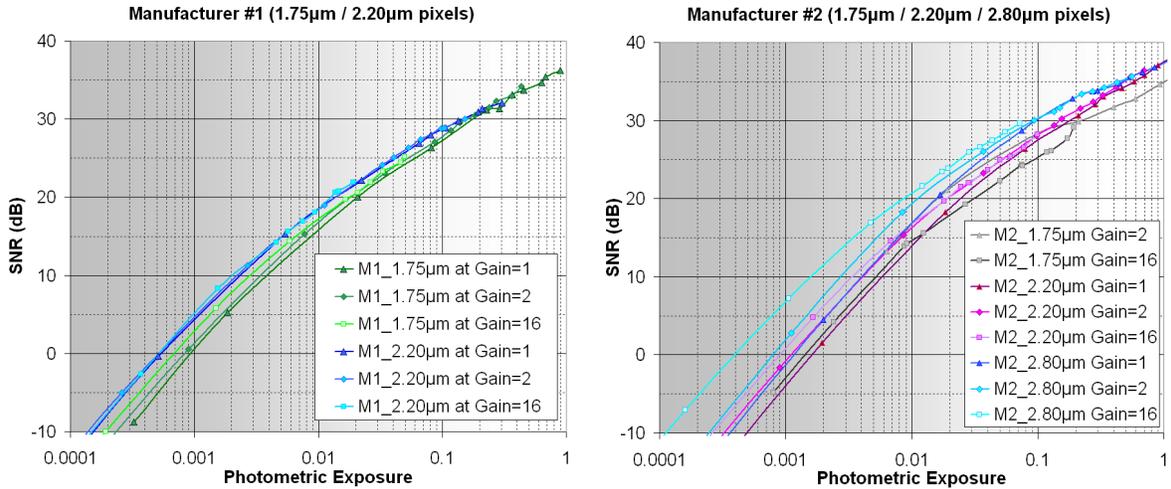


Fig. 5. SNR comparison between image sensors of the same manufacturer but with different pixel sizes; SNR curves are plotted as a function of photometric exposure (in lux.s) and for different *analog* gains (*i.e.* varying exposure times); (Left) Manufacturer M1 and (Right) Manufacturer M2.

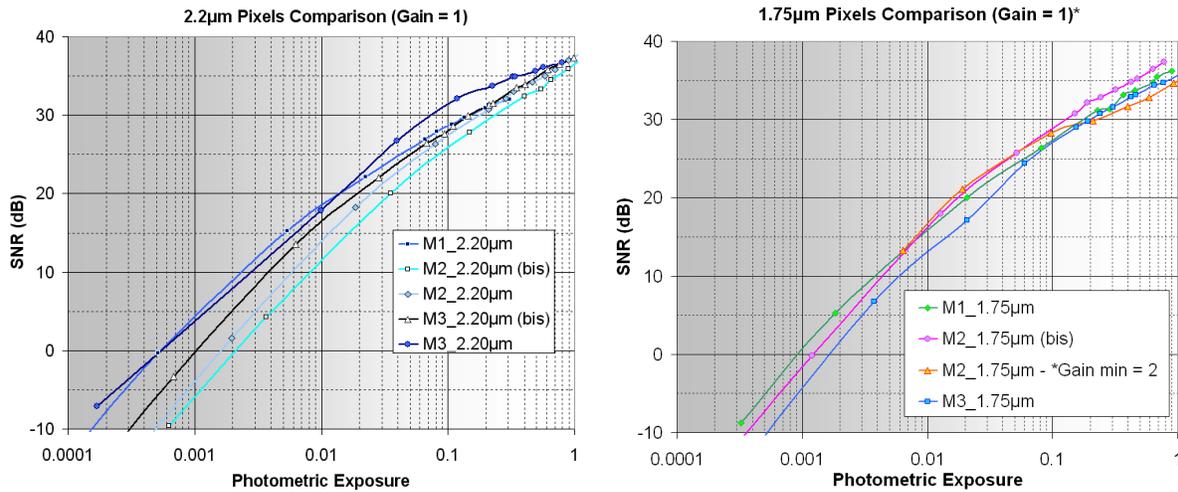


Fig. 6. SNR comparison between image sensors with same pixel size but from different manufacturers; SNR curves are plotted as a function of photometric exposure (in lux.s) and for similar *analog* gains; (Left) 2.2µm pixel pitch and (Right) 1.75µm pixel pitch.

Figure 7 displays the results of the MTF analysis for the five commercial image sensors under test. Both graphs in Figure 7 present the same data. The detector MTFs plotted as a function of input spatial frequency (in lp/mm) on the left graph confirm that, for a given imaging area (e.g. 1/4-inch optical format) and imaging optics, MTF generally improves for image sensors with smaller pixels. The plots on the right are the same detector MTF curves than on the left but with spatial frequency normalized to the image domain. This time, for a fixed pixel count and field-of-view (*i.e.* variable focal length optics), the detector MTFs plotted as a function of image domain frequency (in cycles/image or cpi) indicate that a large pixel size results in a better MTF. For the sensors using 1.75µm and 2.2µm pixels and having nearly identical (vertical and horizontal) resolution, the detector MTFs on the right graph can also be interpreted as Nyquist normalized MTFs with  $f_N$  located at  $f \approx 1024$  cpi.

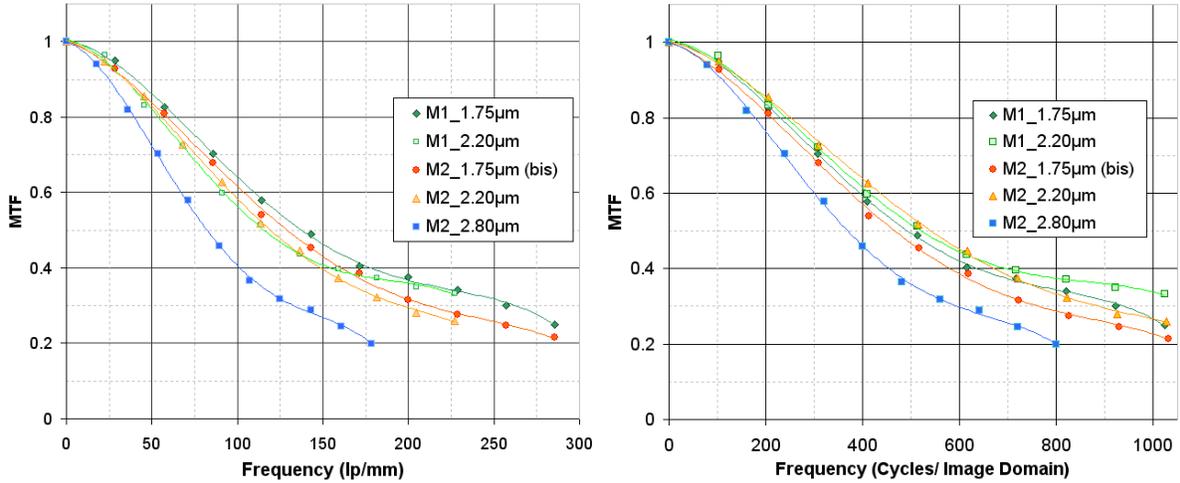


Fig. 7. Detector MTFs as a function of: (Left) input spatial frequency in line pairs per mm; and (Right) spatial frequency normalized to the image domain in cycles per image.

We now calculate TR and the number of effective pixels  $s'$  as discussed in the previous section. To allow sensor comparison, we must first make sure that TR values are computed at identical average photometric exposure  $H_0$ . This is illustrated in Figure 8(a) for targeted  $H_0$  level of 0.4 lux.s. Figure 8(a) also shows that imagers with larger pixel sizes produce (across a wide range of targeted illuminations) images with a higher TR than image sensors with smaller pixels. Finally, the image information capacity results obtained for the five commercial image sensors - with varying pixel size and resolution - are compared in Figure 8(b). This graph shows that for a fixed imaging area, *i.e.* 1/4-inch optical format, the 2.20 $\mu\text{m}$  pixel sensor of each manufacturer is capable of capturing almost the same amount of visual information than its counterpart(s) using smaller pixels. It is interesting to note once again the difference in performance between sensors (with same pixel size) of different manufacturers. For instance, the relative difference in information capacity between M1\_1.75 $\mu\text{m}$  and M2\_1.75 $\mu\text{m}$  is found to be about 20%. Furthermore, when comparing sensors at full resolution, we see the clear advantage in information capacity of sensors using 2.20 $\mu\text{m}$  pixels. All of these observations indicate that using image sensors with pixel size smaller than 2.2 $\mu\text{m}$  (for increasing resolution) does not always yield a higher image information capacity and better image quality.

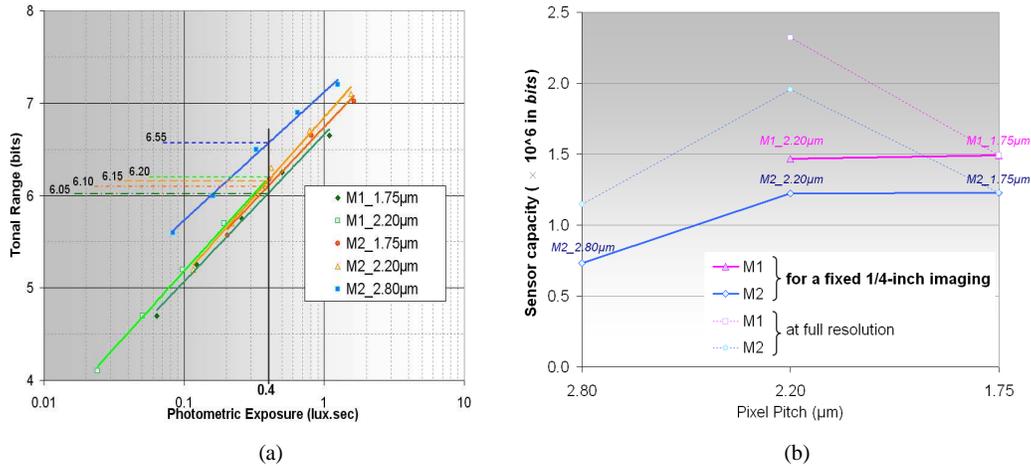


Fig. 8. (a) TR plotted as a function of average photometric exposure; the measurement points correspond to different *analog* gain settings of each sensor; TR is computed for a 8-bit equivalent grayscale; (b) Image information capacity of the five commercial imagers described in Table 2; these imager capacities are calculated using the same diffraction-limited ( $f/2.8$ ) lens model than in Figure 3.

The above measurements suggest that it is very unlikely that shrinking the pixels down to  $1.45\mu\text{m}$  will increase the image information capacity of the next generation of sensors. It seems indeed that the optimal compromise (in the sense of image information capacity) for a camera module with an *ideal* 1/4-inch lens operating at  $f/2.8$  has already been achieved by sensors with a pixel size of  $1.75\mu\text{m}$ . The discrepancy between the predicted value of  $1.45\mu\text{m}$  for the optimal pixel size and the measurements is mainly explained by the fact that, for the commercial sensors under test, large pixels produce a better Nyquist normalized MTF response than small ones (our simulations assumed no increasing cross-talk between pixels as their size decreases). For that same reason and because of the rapidly decreasing TR (cf. OE loss problem) for pixel pitch below  $1.75\mu\text{m}$ , halving pixel size and combining photodiode charges or digital values from four adjacent pixels, *i.e.*  $2\times 2$  pixel binning, will not allow an increase of image information capacity.

## 5. SUMMARY AND CONCLUSION

We reviewed the trends in pixel design for CMOS APS imagers. Despite the use of optimized semiconductor process, more advanced design rules and novel pixel architecture based on transistor sharing, the light sensitivity of pixels below  $3.2\mu\text{m}$  pitch decreases drastically with further pixel size reduction due not only to lower pixel aperture but also more severe pixel vignetting and increasing spatial cross-talk. Therefore, when shrinking pixels beyond this limit, it becomes necessary to examine the importance of tradeoffs between spatial resolution and noise. MTF and SNR can be used as indicators of image quality. A simplified model of the effect of pixel size on sensor MTF and SNR was described to simulate and discuss the theoretical performance of pixels from  $5.2\mu\text{m}$  down to  $1.45\mu\text{m}$ . For selecting the optimal pixel size, we designed a metric that characterizes the visual information transfer capacity (from object to digital image) of the sensor. This metric which is defined as the product of the effective spatial resolution of the image detector by its tonal range, takes both MTF and SNR measurements into account. A theoretical maximum of image information capacity was found for a pixel pitch of  $1.45\mu\text{m}$ , in the approximation that the pixel optics has the ability to efficiently focus the incoming light onto the photodiode area (with negligible cross-talk). Finally, this metric was used as a figure of merit to benchmark five low-end commercial image sensors typically designed for camera-phone applications (to be used in combination with an  $f/2.8$  lens). Our experimental results showed a significant disparity in performance between sensors coming from different manufacturers. In general, for a fixed die size, the advantage of commercial  $1.75\mu\text{m}$  pixel sensors over  $2.20\mu\text{m}$  pixel sensors can be very small. With regards to information capacity, this implies that an optimum has already been reached by sensors using  $1.75\mu\text{m}$  pixels, *e.g.* a 1/4-inch camera phone sensor with 3.2 megapixel resolution. In spite of the advances in CMOS pixel technology and design promised by the manufacturers of image sensors, it will become difficult to scale pixel size down to  $1.45\mu\text{m}$  without significant degradation in image quality. In future work, we will perform subjective experiments to quantify the relationship between image information capacity and the preferences of a human observer between image sharpness and image noise visibility to maintain perceptual image quality. Our comparative analysis of image information capacity needs also be extended to colour image quality. This extension requires to determine the number of colours that a sensor can distinguish, up to noise, which can be performed by evaluating *colour sensitivity*<sup>21</sup> instead of tonal range.

## REFERENCES

- <sup>1</sup> Micron MT9P001 & MT9T012 CMOS Image Sensors Process Comparison Report (CWR-0612-801), Chipworks, Ottawa - Canada, September 2007.
- <sup>2</sup> H.Rhodes, G.Agranov *et al.*, "CMOS imager technology shrinks and image performance", Proc. of IEEE Workshop on *Microelectronics and Electron Devices*, 7-18 (2004).
- <sup>3</sup> P.B.Cartryse, X.Liu and A.E.Gamas, "QE reduction due to pixel vignetting in CMOS image sensors", Proc. Of SPIE Conference in *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications* **3965**, 420-430 (2000).
- <sup>4</sup> P.B.Cartryse and B.A.Wandell, "Optical efficiency of image sensors", J. Opt. Soc. Am. (OSA) A **19(8)**, 1610-1620 (2002).
- <sup>5</sup> J.Vaillant, A.Crocherie *et al.*, "Uniform illumination and rigorous electromagnetic simulations applied to CMOS image sensors", Optics Express **15(9)**, 178-184 (2007).
- <sup>6</sup> C.Fesenmaier, B.Sheahan and P.B.Cartryse, "Optical crosstalk in CMOS image sensors", EE362/PSYCH221 Project Report, Stanford University (2007).
- <sup>7</sup> <http://www.optiwave.com/>

- <sup>8</sup> A.Moini, "Vision chips or seeing silicon", Technical Report, Centre for High Performance Integrated Technologies and Systems, The University of Adelaide (1997).
- <sup>9</sup> A.E.Gamal and H.Eltoukhy, "CMOS image sensors", *IEEE Circuits and Devices Magazine* **5(3)**, 6-20 (2005).
- <sup>10</sup> I.Shcherback and O.Yadid-Pecht, "CMOS APS MTF modeling", *IEEE Trans. on Electron Devices* **48(12)**, 2710-2715 (2001).
- <sup>11</sup> J.LMeyzonettes and T. Lépine, *Bases de Radiométrie Optique*, Cépaduès Editions, Chap.2, 70-71 (2001).
- <sup>12</sup> P.B.Cartrysse and B.A.Wandell, "Roadmap for CMOS image sensors: Moore meets Planck and Sommerfield", *Proc. Of SPIE Conference in Digital Photography* **5678**, 1-13 (2005).
- <sup>13</sup> M.Cohen, F.Roy, D.Herault *et al.*, *Proc. of IEEE International Electron Devices Meeting*, 1-4 (2006).
- <sup>14</sup> B.Pain, "CMOS imagers: how did we get here? What are we doing? Where are we going?", *IntertechPira Conference on Image Sensors*, San Diego - USA (2007).
- <sup>15</sup> J.Farrell, F.Xiao and S.Kavusi, "Resolution and light sensitivity tradeoff with pixel size", *Proc. Of SPIE Conference in Digital Photography II* **6069**, 211-218 (2006).
- <sup>16</sup> C.E.Shannon, "A mathematical theory of communication", *The Bell System Technical Journal* **27**, 623-656 (1948).
- <sup>17</sup> D.Middleton, *Statistical communication theory*, McGraw-Hill Book Co., 311-315 (1960).
- <sup>18</sup> B.R.Frieden, "How well can a lens system transmit entropy", *J. Opt. Soc. Am. (OSA)* **58(8)**, 1105-1112 (1968).
- <sup>19</sup> <http://www.dxo.com/>
- <sup>20</sup> <http://www.trioptics.com/>
- <sup>21</sup> J.Buzzi, F.Guichard and H.Hornung, "From spectral sensitivities to noise characteristics", *Proc. Of SPIE Conference in Electronic Imaging* (2007).