

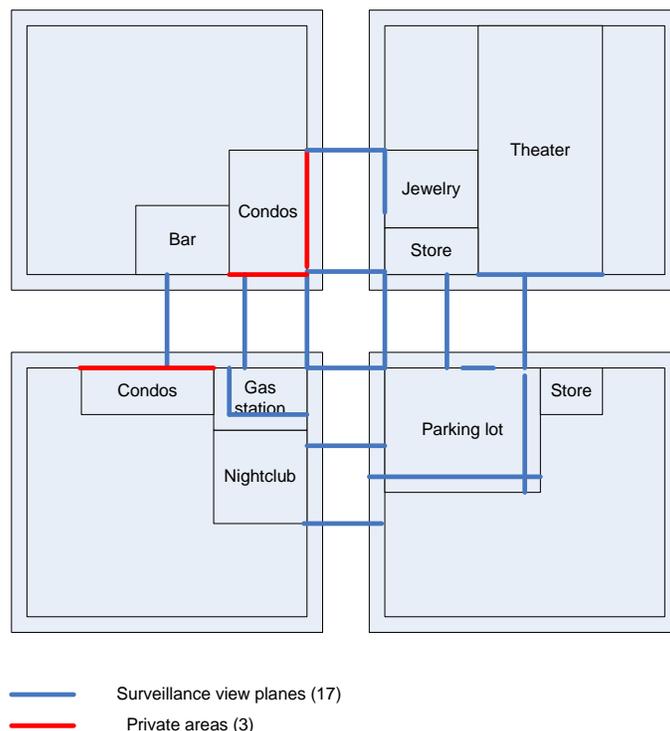


## Wi4Net White Paper:

### *Outdoor Camera Surveillance Design*

Outdoor surveillance requires great flexibility to be efficient. Surveillance has to be done over distances from few feet to over 300 feet at all hours of day and night, under monitored and non-monitored conditions. There is a need to see the whole scene and at the same be able to recognize details, such as car license plates and faces. Images have to be seen under full sun and in nights without moon, with situations where part of the image is flooded with light and other parts are under very low illumination. Furthermore, the amount of information generated has to be minimized, so it is possible to network the information (often wirelessly) back to a server location and maintain storage within reasonable limits. Still, image quality has to be good enough to be used as evidence, as mentioned, under challenging illumination conditions.

Initially, the surveillance target locations should be defined. It is optically impossible to see everything from the selected surveillance locations or to see everywhere all the time. Instead, surveillance view planes have to be defined, targeting important, viable views from the various surveillance locations. The diagram below depicts a 4-city block area, with sidewalks and local commerce, for which were defined 17 surveillance planes and 3 private ones.





Surveillance is not permitted in the private areas and these areas would need to be masked. As each surveillance location characteristics are different, view planes are defined per locations, along with viewing specifications per view plane, such as continuous, periodic, time-driven or event-driven viewing, periodicity, duration, frame rate, etc.

Based on the specific requirements for each location, the number and type of observation points (i.e. cameras) can be defined.

Generally, citywide style surveillance systems are composed of the following components:

- Cameras
- Local “hubs”, with the following typical options:
  - Multi-camera support
  - Power back-up (typical durations of 5 to 30 minutes)
  - Local video storage (typically 10 to 30 days, which is equivalent to 0.5 to 1 TB of storage capacity)
  - One or more radios
- Repeaters and backhaul radios
- Video management solution, with servers, storage systems and monitor stations

This document focuses on the camera design portion of the network.

## Cameras

The main components of a digital camera are the digital light sensor and the optical system. IP cameras send the images in digital form, so a processor is used to compress the information and make it available using the Internet Protocol (IP).

A camera’s purpose is to capture objects, and a object’s definition correlates with the number of pixels available (camera resolution) in the sensor. Consequently, a larges number of pixels results in better details in the image. The table below presents the most common reference resolutions in the industry.

Camera resolution	Pixels			Image
	Horizontal	Vertical	Total raw	Mbit/image
CIF	352	288	101,376	2.43
4 CIF	704	576	405,504	9.73
1 Megapixel	1,280	720	921,600	22.12
2 Megapixel	1,920	1,080	2,073,600	49.77
5 Megapixel	2,592	1,944	5,038,848	120.93



The following main camera characteristics should be observed when selecting cameras:

- **PTZ Capability:** Capability to pan (horizontal movement), tilt (vertical movement), and zoom (enlarge parts of the scene). This capability is abbreviated as “PTZ”.
- **Auto-focus Capability:** Required to provide sharp images, adjusted for the distance to the object of interest.
- **Minimum Scene Illumination (Light Sensitivity):** This capability is essential, particularly scenes are to be captured at night. This parameter is defined in “lux”. Camera sensitivity depends on many factors, including sensor type, lens F-number (opening), shutter opening duration, and IRE (threshold value between black and white and stands for Institute of Radio Engineers). The sensitivity is different for color and black and white images. The best cameras have a black and white sensitivity of 0.0008 @ F1.4, 1/15s, 30 IRE. Unfortunately no manufacturer fully specifies the conditions for their specifications. It is very difficult to compare specifications from different manufacturers as they do not specify the complete settings under which the sensitivity readings were done. We had to resort to field evaluations to be able to compare performances.
- **Light Dynamic Range:** This capability is essential to be able to view scenes that have both bright and dark objects or areas in the same image. The best cameras have a dynamic range of 75 dB.
- **Preset and Tour Capability:** With this capability, cameras can store several PTZ preset locations, along with camera “tour” (for example using the earlier mentioned preset location), as programmed by the user.
- **Masking Capability:** With this capability, the camera can mask specific areas of the image. This feature is primarily used to address privacy concerns.
- **Event Detection:** This capability is used to detect motion or specific objects.

In citywide outdoor surveillance designs, the following two camera options are best suited: (1) PTZ (Pan, Tilt and Zoom) cameras, and (2) megapixel fixed (no PTZ) cameras.

PTZ cameras can pan horizontally 360°, can tilt vertically about 90° and zoom up to 35 times (at which zoom level the wider angle can be divided in 1225 scenes). Although a single camera can provide 360° coverage, some obstruction may exist, so in some cases two PTZ cameras are recommended. The big advantage of the PTZ camera is that it can capture close and distant scenes with the same resolution and perfect focus. The disadvantage is that there will be breaks between scene captures. This is often acceptable, as scenes do not change over short intervals.

Still, some surveillance view planes may require continuous coverage at a specific distance, for example to provide for car license plate reading. In such cases, fixed cameras can be used as a complement. Typically, fixed cameras focus only on one specific surveillance plane (i.e. fixed focus), and as a result objects distant from the targeted plane will be out of focus. The advantage of fixed cameras is that they allow continuous monitoring, that can still be event triggered, but can be a sole solution in an outdoor city-type environment. One issue is



that criminals learn where fixed cameras are pointed and avoid these areas. This situation does not occur with PTZ cameras.

Megapixel fixed cameras are used to compensate for the lack of optical zoom in fixed cameras, as they offer the possibility of digital zoom (only to an extent), while still keeping the desired resolution. On the other hand, megapixel cameras generate large amounts of information, parts of which are not used. Overall, these cameras provide a considerable burden on (wireless) backhaul networks and storage solutions. The use of these camera need to take into account also the impact on the overall solution, and given practical limitations, should be used with parsimony.

The following two type of megapixel technologies are available: (1) Megapixel fixed camera with advanced CCD (Charged-Couple Device) using H.264 video compression, and (2) megapixel fixed cameras with CMOS (Complementary Metal Oxide Semiconductor) using JPEG2000 picture compression. The advanced CCD cameras provide a better image quality and a better sensitivity to light due to the lenses this technology uses over each pixel. This is better described in the appendix.

At the offer stage, the surveillance design can often not be developed in great detail, since a close collaboration with the stakeholders is required to understand all requirements and issues (such as crime statistics, specific hotspots, etc.). An integrator's experience in surveillance camera design is important to review before selecting a vendor.

### ***Image Compression Algorithms***

A digital image is composed of pixels, aligned in a series of horizontal lines. The number of pixels defines the camera resolution. Cameras are defined according to their resolution, and the main classifications are shown in the table below.

Camera Resolution	Uncompressed (Mbit/s)	Typical Compression Mbit/s @ 15 fps			
		M-JPEG	JPEG 2000	MPEG-4	H.264
CIF	2.4	1.4	1.2	0.4	0.2
4 CIF	9.7	5.5	4.7	1.7	0.8
1 Megapixel	22.1	12.5	10.6	3.9	1.7
2 Megapixel	49.8	28.1	23.9	8.7	3.9
5 Megapixel	120.9	68.3	58.1	21.1	9.5

Regular TV has a resolution of 720x480, while the best HDTV has a resolution of 1920x1080. The most common format used in surveillance is CIF (Common Intermediate Format). CIF does not provide enough resolution for face recognition, and 4CIF is the preferred format today (providing 4 time better resolution than CIF).



Compression algorithms allow for different compression factors, from lossless to very large compression factors. This is another area where vendors' claims are overstated. For example, a vendor might feature lossless compression, while stating throughput (i.e. data rate requirements) at maximum compression, thus with loss in picture quality present.

The pixel information is composed of three-color information (RGB - red, green and blue or YCbCr - luma and chroma) and results in 24 bits of information per pixel. As can be seen in the previous table, the number of bits in an image grows exponentially with the resolution. Compression techniques were then developed to reduce the amount of information required to code the images. M-JPEG and JPEG2000 encode image by image and provide a conservative compression. MPEG-4 and H.264 code sequences of images, benefiting from the redundancy existing between them and provide a much greater compression. Therefore, the latter two techniques are ultimately suitable for streaming video applications. H.264 is the newest compression method and provides by far the best results. This compression technique provides for the most cost-effective implementation and best expansion capability.

### **Conclusions**

In the surveillance design of citywide surveillance systems, as a first step, a needs-assessment will reveal the surveillance requirements & locations, priority of viewing planes, hotspots, etc. Furthermore, the backhaul network capability needs to be reviewed for each of the locations, along with considerations for future expansion. Also, budget and funding constraints need to be taken into account.

Based on the parameters of each of the locations, the surveillance camera design can be carried out. PTZ cameras are the only options to provide full area coverage, while maintaining reasonable throughput requirements. One-PTZ and two-PTZ camera setups are most common for citywide deployments, providing for a good balance between coverage, cost and throughput & storage requirements. A reason for a two-PTZ setup is to increase coverage (and compensate for potential obstacles).

For specific viewing planes requiring continuous monitoring all the time, fixed cameras can complement the camera design. Megapixel cameras are a logical choice in citywide camera designs, given the large coverage area requirements. When utilizing fixed megapixel cameras, consideration should be given to impact on (wireless) backhaul network and storage capacity. Camera pointing and focus needs to accommodate the available bandwidth for this camera, not the capability of the camera (for instance, a fixed camera might have 5 megapixel capability, but network constraints would only allow configuration to 1 megapixel). H.264 encoding crucial to successful implementation of megapixel cameras.

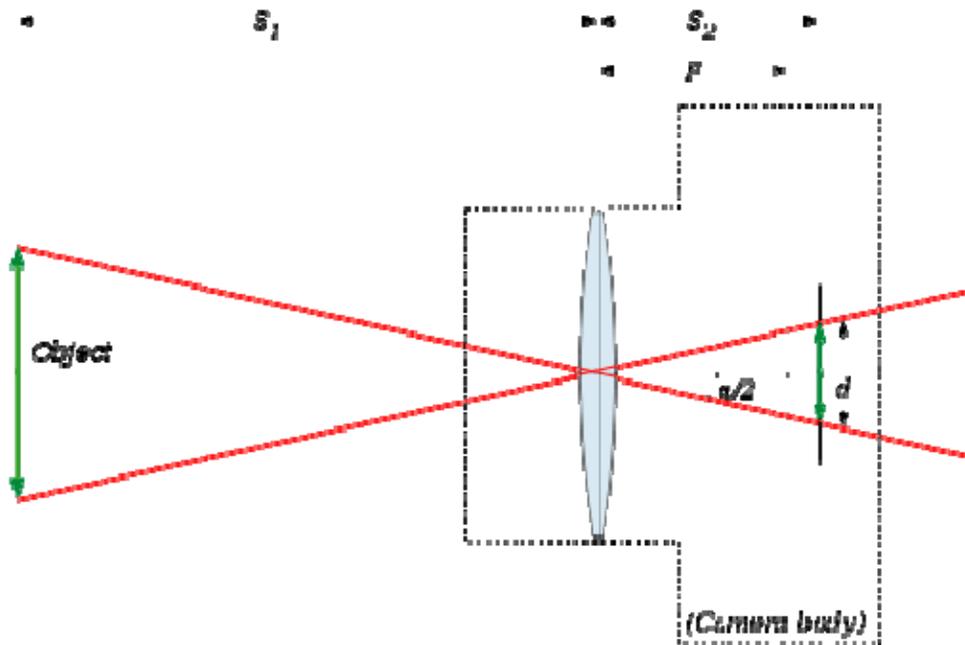


## APPENDIX

### *Image Optics*

The light coming from the scene is received by the camera lenses and projected over a sensor in the camera. The following parameters define the optical geometry:

- The lens or shutter aperture, defined by  $D$
- The sensor size, defined by  $d$
- The focal length defined by distance between the lens and the sensor, represented by  $F$  or  $f$  (typically 3.5 mm to 120 mm)
- The F-stop or the f-number is the ratio between the focal distance and the aperture diameter



To project a sharp image of distant objects,  $S_2$  needs to be equal to the focal length,  $F$  or  $f$ , which is attained by setting the lens for infinity focus. Then the angle of view is given by:

$$\alpha = 2 \arctan \frac{d}{2f}$$

$d$  represents the size of the film (or sensor) in the direction measured. For example, for a sensor that is 6.3 mm wide,  $d = 6.3$  mm would be used to obtain the horizontal angle of view.

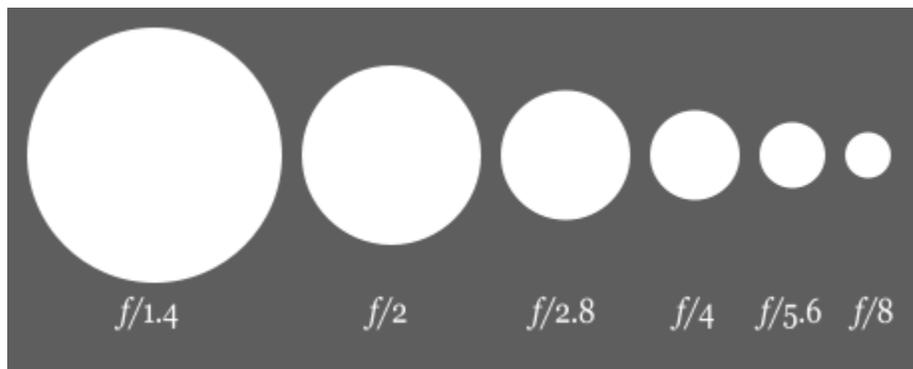


In optics, the f-number (sometimes called focal ratio, f-ratio, or relative aperture) of an optical system expresses the diameter of the entrance pupil in terms of the focal length of the lens; in simpler terms, the f-number is the focal length divided by the "effective" aperture diameter. It is a dimensionless number that is a quantitative measure of lens speed, an important concept in photography.

The f-number  $f/\#$ , often notated as  $N$ , is given by:

$$f/\# = N = \frac{f}{D}$$

$f$  is the focal length and  $D$  is the diameter of the entrance pupil. By convention, " $f/\#$ " is treated as a single symbol, and specific values of  $f/\#$  are written by replacing the number sign with the value. For example, if the focal length is 16 times the pupil diameter, the f-number is  $f/16$ , or  $N = 16$ . The greater the f-number, the less light per unit area reaches the image plane of the system; the amount of light transmitted to the film (or sensor) decreases with the f-number squared. Doubling the f-number increases the necessary exposure time by a factor of four.



*Aperture diagram*

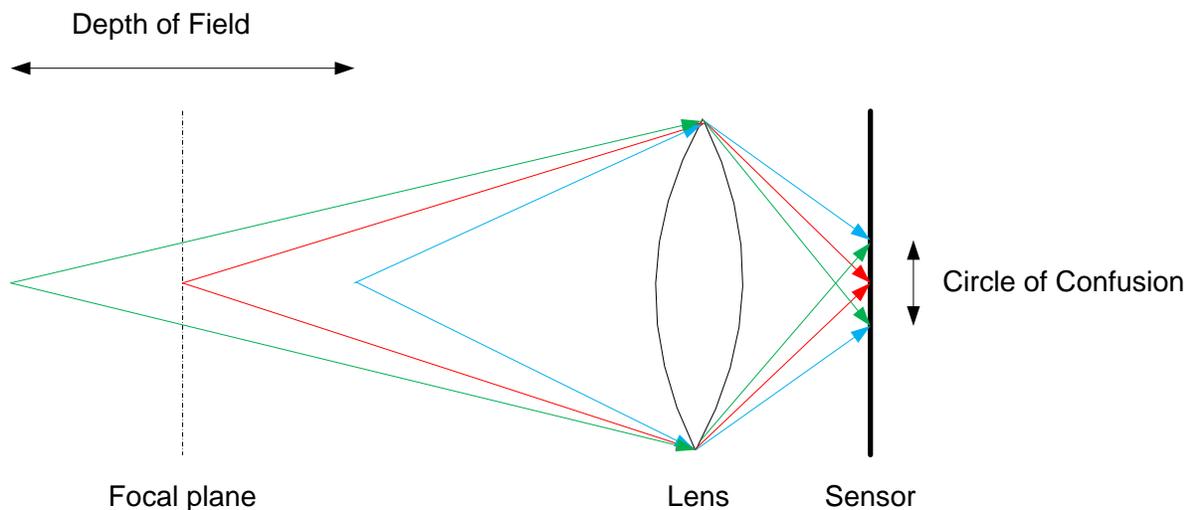
The pupil diameter is proportional to the diameter of the aperture stop of the system. In a camera, this is typically the diaphragm aperture, which can be adjusted to vary the size of the pupil, and hence the amount of light that reaches the film or image sensor. The common assumption in photography that the pupil diameter is equal to the aperture diameter is not correct for many types of camera lens, because of the magnifying effect of lens elements in front of the aperture.

A 100 mm lens with an aperture setting of  $f/4$  will have a pupil diameter of 25 mm. A 135 mm lens with a setting of  $f/4$  will have a pupil diameter of about 33.8 mm. The 135 mm lens'  $f/4$  opening is larger than that of the 100 mm lens but both will transmit the same amount of light to the film or sensor. Other types of optical system, such as telescopes and binoculars may have a fixed aperture, but the same principle holds: the greater the focal ratio, the fainter the images created (measuring brightness per unit area of the image).



In a camera system, the light reflects from the objects in the image and is received by lenses, and directed towards an image sensor. These lenses have to be adjusted to focus the image. Only objects at an exact distance can be precisely focused. All other objects will be out of focus. So, for a fixed lenses camera the distance to the objects we want in focus should be pre-defined.

A single point in the image becomes a circle if it is not at the exact distance and this circle is called (uncertainty circle). The human eye can tolerate some amount of out of focus and this is defined by resolution of the camera. Once the maximum size of the circle of uncertainty is known, it is possible to define the Depth of Field (DoF) of the image, in which the objects are considered in focus.



The diameter of the lens defines the aperture through where light is received, larger the aperture more light is received. This aperture is known as F/stop and can be controlled by a mechanical iris. It is a fraction of the lens diameter and is defined by a number  $F/i$ . So,  $f/2.8$  allows nearly all light coming through the lens to reach the sensor, while  $f/16$  allows a much smaller amount.

The aperture has a major impact on the depth of field; larger the aperture smaller the depth of field. The angle of vision of the camera depends on its focal length. Larger the focal length, narrower is the angle of vision and narrower the Depth of Field.

### ***Image Sensors in Cameras***

An image sensor is a device that converts an optical image to an electric signal. It is used mostly in digital cameras and other imaging devices. An image sensor is typically a charge-coupled device (CCD) or a complementary metal–oxide–semiconductor (CMOS) active-pixel sensor. Today, most digital still cameras use either a CCD image sensor or a CMOS sensor. Both types of sensor accomplish the same task of capturing light and converting it into



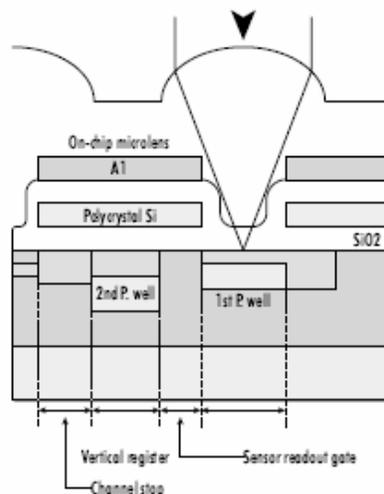
electrical signals. A CCD is an analog device. When light strikes the chip it is held as a small electrical charge in each photo sensor. The charges are converted to voltage one pixel at a time as they are read from the chip. Additional circuitry in the camera converts the voltage into digital information. A CMOS chip is a type of active pixel sensor made using the CMOS semiconductor process. Extra circuitry next to each photo sensor converts the light energy to a voltage. Additional circuitry on the chip may be included to convert the voltage to digital data. CCD sensors offer a better image quality, due to better dynamic range, uniformity and shuttering capability. CMOS offers Windowing capability (capability to extract only part of the image).

### **Sensor Technology**

An interline-transfer CCD has a parallel register that is subdivided into alternate columns of sensor and storage areas. The image accumulates in the exposed area of the parallel register and during CCD readout the entire image is shifted under the interline mask into a hidden shift register. Readout then proceeds in normal CCD fashion. Since the signal is transferred in microseconds, smearing is undetectable for typical exposures. However, a drawback to interline-transfer CCDs has been their relatively poor sensitivity to photons since a large portion of each pixel is covered by the opaque mask. As a way to increase a detector's fill factor, high-quality interline-transfer devices have microlenses that direct the light from a larger area down to the photodiode.

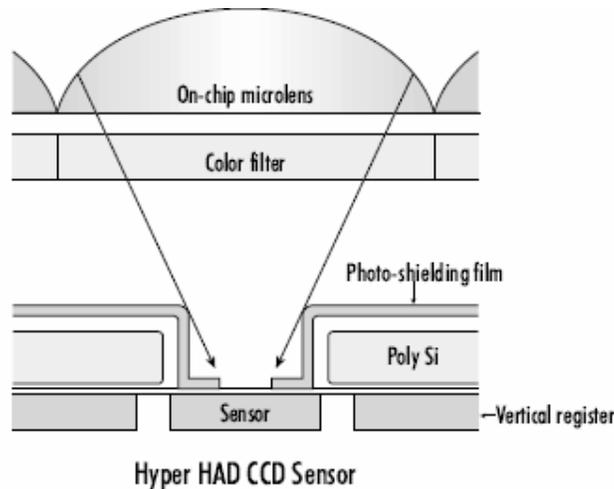
### **Hyper HAD**

Hyper HAD was the first technology to implement the interline-transfer CCD and on-chip microlenses (see figure below). By collecting some of the light falling on the masked area, which is otherwise lost, microlens technology improved the quantum efficiency (QE). Furthermore, microlenses increase the effective fill factor of the CCD from approximately 40% to greater than 75%.



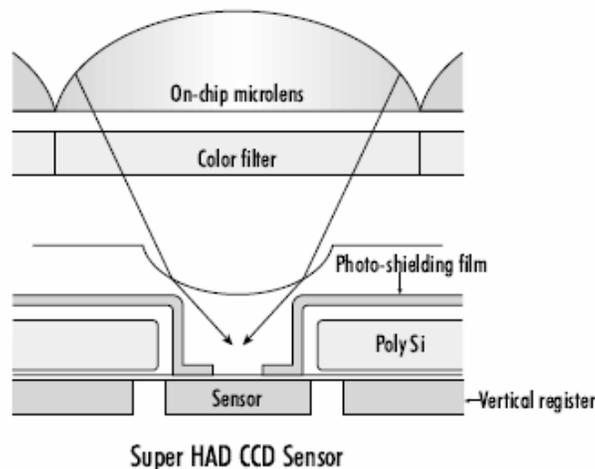


Sony subsequently improved the microlens technology and manufacturing process with the introduction of the Hyper Hole-Accumulation-Diode (HAD) CCD. Hyper HAD CCDs have much closer spacing between microlenses, thus further increasing the light-collection efficiency, even with reduced pixel sizes (see figure below).



### Super HAD

Recently, the Super HAD interline CCD was developed with an additional layer of on-chip microlenses very close to the pixel area (see figure below). When used with wider f-number lenses, a single array of microlenses cannot focus the higher-angle light rays onto the sensing area of the pixel and sensitivity is reduced. A second layer of microlenses helps alleviate this problem by further condensing the beam path, thereby increasing QE. Another improvement in this technology is thinning of the insulating layer between the silicon substrate and polysilicon gate structures, which reduce the light leaking under the mask (smear factor).





## ***Exwave HAD***

In monitoring and surveillance applications, camera sensitivity is one of the most important factors in obtaining an adequate picture in low light conditions. The sensitivity of cameras using the Exwave HAD technology is well over twice that of the cameras using the other HAD technologies.

The original Hyper HAD sensor structure has an OCL (on chip lens) located over each pixel. The result is that light is concentrated on the photo-sensor areas and the sensitivity of the camera is improved. The Exwave HAD takes the Hyper HAD technology a giant step further. The OCL of the Exwave HAD is a nearly gap-less structure, eliminating the ineffective areas between the microlenses. This enables the hole accumulated layer to receive the maximum amount of light. Moreover, the smear level of the Exwave HAD technology is reduced to 1/50th that of the Hyper HAD technology. This leakage is dramatically reduced because the improvement of the unit cell structure minimizes the unnecessary reflection of the light onto the CCD surface.

### **EXview HAD**

The introduction of EXview HAD CCDs are the most recent innovation and added another step of sensitivity by improving the QE (Quantum Efficiency) in the near-infrared (NIR) region. Since NIR photons are absorbed at deeper levels in the silicon, using thicker silicon in the chip increases the probability of photon-silicon interaction and thus further increases QE.

## ***Sensor Scan***

Progressive or non-interlaced scanning is a method for displaying, storing or transmitting moving images in which all the lines of each frame are drawn in sequence. This is in contrast to the interlacing used in traditional television systems where only the odd lines, then the even lines of each frame (each image now called a field) are drawn alternately. Interlaced scan present an interline twitter effect associated with interlace. The interlaced images use half the bandwidth of the progressive ones. Actual interlaced video blurs details to prevent twitter and this comes at the cost of image clarity. A line doubler could not restore the interlaced image to the full quality of the progressive image.

## ***Dynamic Range***

Citywide surveillance applications typically require a wider dynamic range, such as 65-75 dB (1:1800 - 1:5600), to deal with challenging (strongly varying) illumination. When imaging such a scene with a 60 dB imager, either detail in the darker areas get lost in the noise ("cut off"), or details in the brighter areas are lost in saturation, or both. This is shown in the following figure.

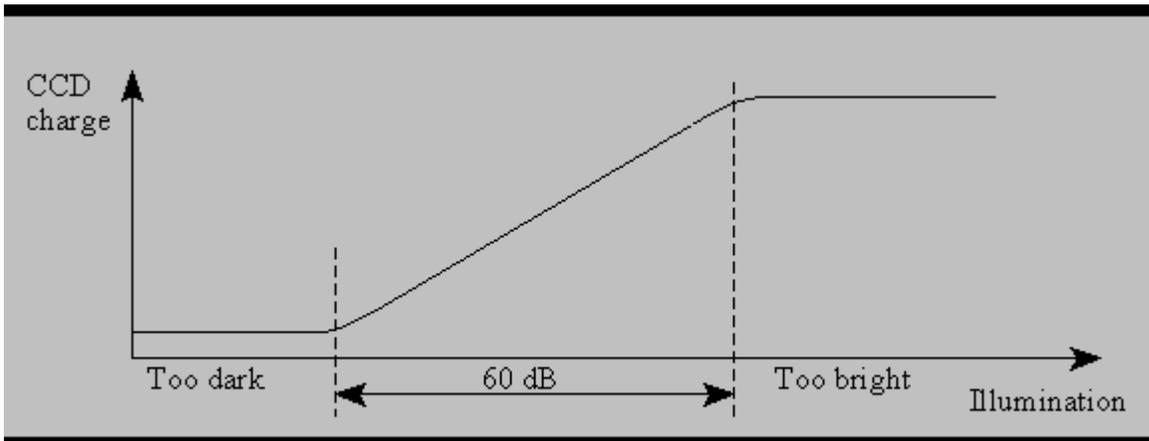
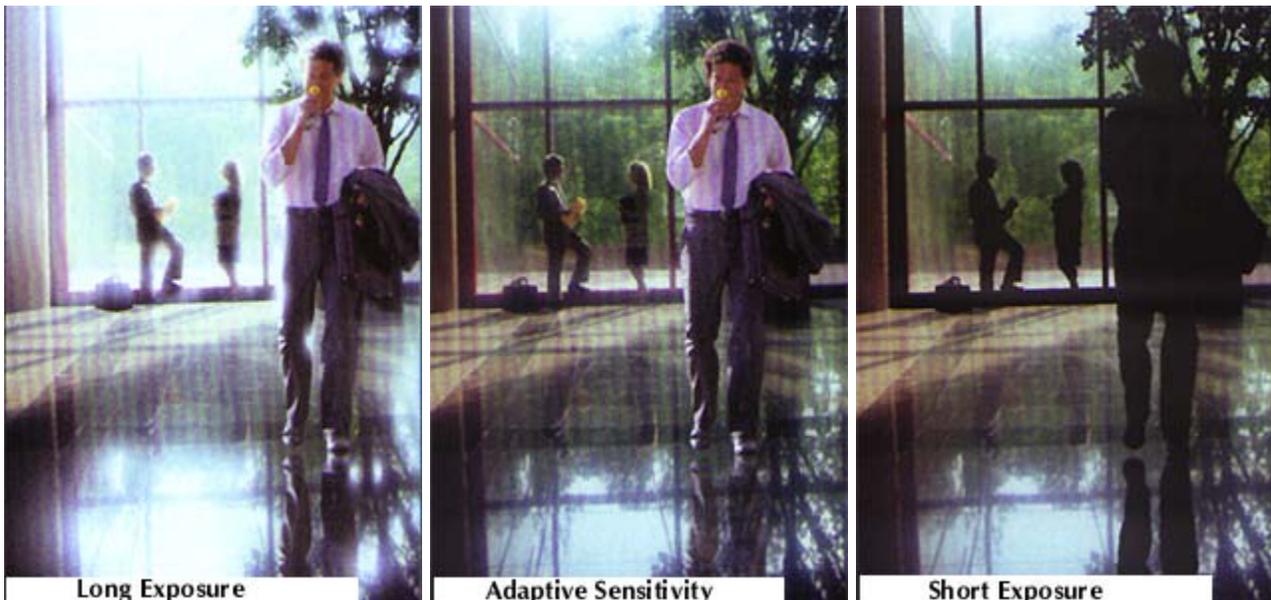


Figure 1: Response of a CCD.

*Response of a CCD*

There are two ways to compensate for a wide dynamic range: (1) Use of non linear sensors or (2) use different exposures for sub-sequent images and averaging the results.

The following illustrations explain the improvements that are possible with increased dynamic range.





The following table provides summary information about various reference camera systems.

Manufacturer	Type	Pan (°)	Tilt (°)	Zoom	Auto focus	Sensor Technology	Scan	F number	Focal Length	Minimum Scene Illumination	Sensitivity	Resolution	FPS	Compression
<b>PTZ</b>														
Axis	233D	360	180	35X Optical; 12X Digital	Yes	1/4 ExView HAD CCD	Progressive	1.4 to 4.2	3.4 to 119 mm	Color: 0.5 lux@30IRE, B&W: 0.008 lux@30IRE	Good	704x480	30	MJPEG, MPEG-4
Pelco	Spectra IV IP	360	90	35X Optical; 12X Digital	Yes	1/4 ExView HAD CCD		1.4 to 4.2	3.4 to 119 mm	Color: 0.55 lux@35IRE, B&W: 0.0002lux@35IRE	Good	768 x494	30	MJPEG, MPEG-4
Sony	SNCRX570N	360	360	35X Optical; 12X Digital	Yes	1/4 ExWave HAD CCD	Interlaced	1.6 to 4.5	3.4 to 122.4 mm	Color: 1.4 lux @ 50IRE, B&W: 0.15lux@ 50 IRE	Average	640x480	30	MJPEG, MPEG-4, limited H.264
<b>FIXED</b>														
Avigilon	IP Camera 1Mp	None	None	Digital only	No	1/3 CMOS 2 Mp	Progressive	Fixed only	Fixed only	Color: 0.2 @ F1.4, B&W: 0.02 @F 1.4	Average	1280x720i	39	JPEG2000
Avigilon	IP Camera 2Mp	None	None	Digital only	No	1/2.5 CMOS 2 Mp	Progressive	Fixed only	Fixed only	Color: 0.2 @ F1.4, B&W: 0.02 @F 1.4	Average	1920x1080i	18	JPEG2000
Avigilon	IP Camera 3Mp	None	None	Digital only	No	1/2 CMOS 2 Mp	Progressive	Fixed only	Fixed only	Color: 0.2 @ F1.4, B&W: 0.02 @F 1.4	Average	2048x1536i	12	JPEG2000
Avigilon	IP Camera 5Mp	None	None	Digital only	No	1/2.5 CMOS 2 Mp	Progressive	Fixed only	Fixed only	Color: 0.3@F1.4, B&W: 0.03 @F 1.4	Average	2592x1944i	12	JPEG2000
Axis	Q1755	None	None	10X Optical; 12X Digital	Yes	1/3 CMOS 2 Mp	Progressive	1.8 to 2.1	5.1 to 51 mm	Color: 2 lux@30IRE, B&W: 0.2 lux@30IRE	Average	1920x1080i	30	MJPEG, H.264