Electromagnetic radiation from bound electrons

Contents

Introduction

General

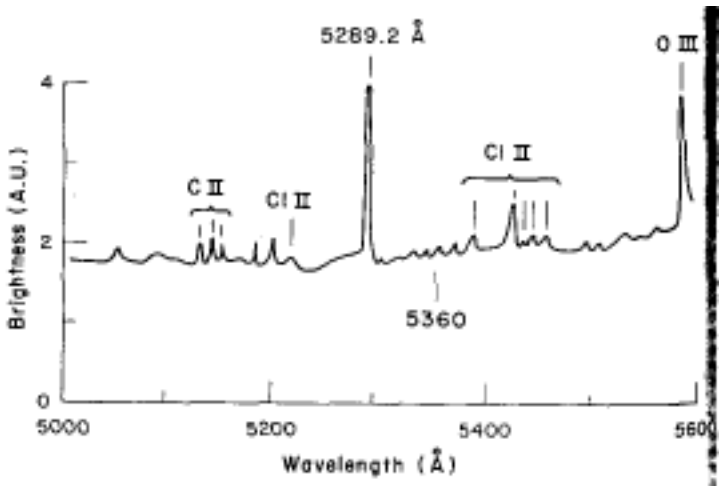
The object is to use light emitted from the plasma to give information on a well defined parameter. The systems fall into three parts:

1) Collection of light.
   - Transmission/absorption of materials
   - Discrete optics
   - Fiber optics
   - Problems with spatial asymmetries and techniques for overcoming them based on multiple sight lines and on active spectroscopy

2) Selection of a portion of the spectrum in such a way that it can be measured.
   - diffraction gratings
   - interference filters (include multilayers)
   - absorption filters

3) Analysis.

Transitions of electrons between states produces narrow lines, not a continuum. Assume energy level structure is known. Not well justified for ionized heavy atoms present in a plasma.



Detectors started as film. Now photo diode or micro channel arrays are used. Fiber optics lead in the light, avoiding lenses. The photo diode arrays are preceded by an image intensifier, and followed by an electro to optical output (for isolation). 300 x 500

pixels are possible. Each pixel (diode) is about 50 μm in size, and center to center about 50 μm. On a 1/2 m instrument (see later) $1A^0$ (10^{-8} m) corresponds to a few pixels, and around 300 A can be covered at once. For hydrogen such a system would allow a temperature resolution of 1 eV.

CCD cameras

In the past few years, there has been a general replacement of tube imaging cameras with the newer solid state technology of CID (Charge Injected Display) and CCD (Charge Coupled Display) technologies. In general vision products and in the consumer and broadcast applications, the choice has been the CCD solid state imaging camera. Instead of an imaging tube with a photosensitive surface at its forward end, the device upon which the lens focuses the scene sample is the surface of the solid state chip, which is

significantly smaller in size. A single CCD chip is roughly the thickness of a quarter and may be as small as 1/3 inch square.

The theory of operation of a CCD imaging chip, simply stated, is that photons strike the surface of the imaging chip, which is divided into individual elements (rather than the continuous photosensitive surface of an imaging tube). Each element samples the light at that point and converts the light to a charge, which is then read out to camera circuitry.

The surface of the imaging chip therefore resembles a matrix of individual elements, which in turn define the spatial resolution capability of the camera.

In color video cameras, it is typical to sample the red, green, and blue components of a scene in order to reproduce color. In the older tube technology, one could either spin a filter in front of a single tube (applicable for slow frame imaging) or use three independent tubes, each with a red, green, and blue filter respectively, to sample the three components of the scene and report back to the circuitry for each point sampled. Optically, the use of a beam splitter, an optical prism that splits the focused scene, allows each tube to see the image with the least optical distortion since all three sensors see the image in the same plane.

In CCD cameras, it is also possible to place a color filter directly on adjacent rows of imaging elements, creating a series of "stripes." With three stripes (R, G, and B), it is possible to sample the scene directly through the same lens for red, green, and blue, at precisely the same time on the same chip. Obviously, the components are not being registered precisely in line since the red element is a row above the green, and the blue a row below the red, resulting in a mild optical distortion. Further, if the manufacturer quotes that there are 768 elements in the x direction of the matrix, and the stripes run vertically, the real spatial resolution of the chip is closer to 768 divided by 3, or 256.

In a three chip camera, three separate chips, each of which, for example, could have 768 elements in the x direction, are used. The focused image goes through a beam splitter, which allows each chip to sample the full scene in each respective color over the entire chip for full resolution. In addition, the optical performance of the system is superior for color sampling compared to the single chip, as each color sample uses the same precise

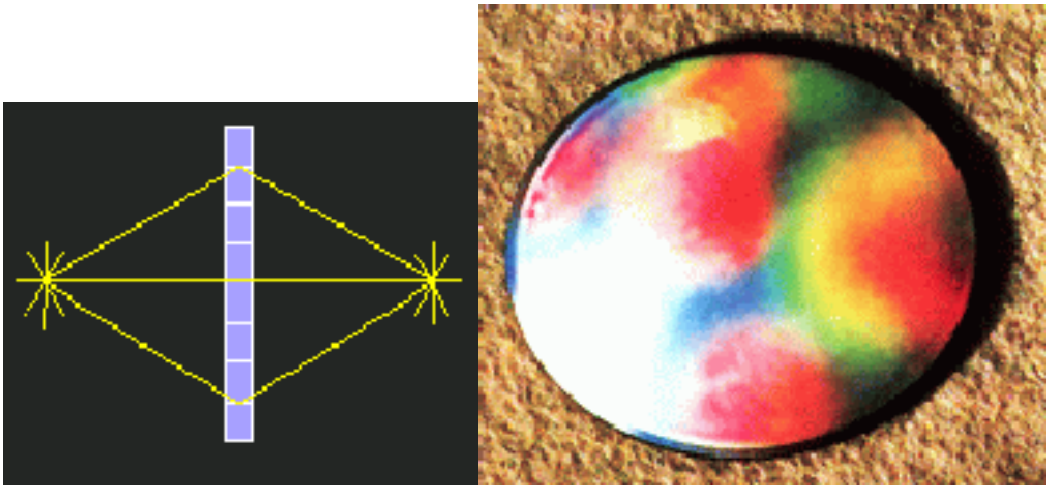
optical alignment. If each chip were the same specification of the single chip camera, then the spatial resolution in color could therefore be three times greater than the single chip example.

Unfortunately, the quotation of spatial resolution for single chip CCD cameras is tricky, due to the less demanding audience for its purchase. Single chip cameras are often used for less discriminating industrial applications as opposed to the broadcast industry, which uses the majority of three chip cameras. It is not uncommon to see a 680 TV line resolution quoted for a three chip camera and a 600 TV line resolution quoted for the

single chip. This is because there are still no standards requirement for specifications for cameras. The single chip color camera is therefore often quoted as operating in a "Black and White" mode or "Green filter mode," as it is sometimes called, wherein all elements are used to sample the same color, not all three. Without this critical differentiation spelled out clearly (and it rarely is), it would seem the difference would normally seem minimal.

Notwithstanding, CCD-based cameras have added a great deal to video technology, and with each generation, provided greater application to a wider range of imaging, bringing with them the advantage of durability, low lag characteristics, minimal power usage, small size, and improved light.

MCP

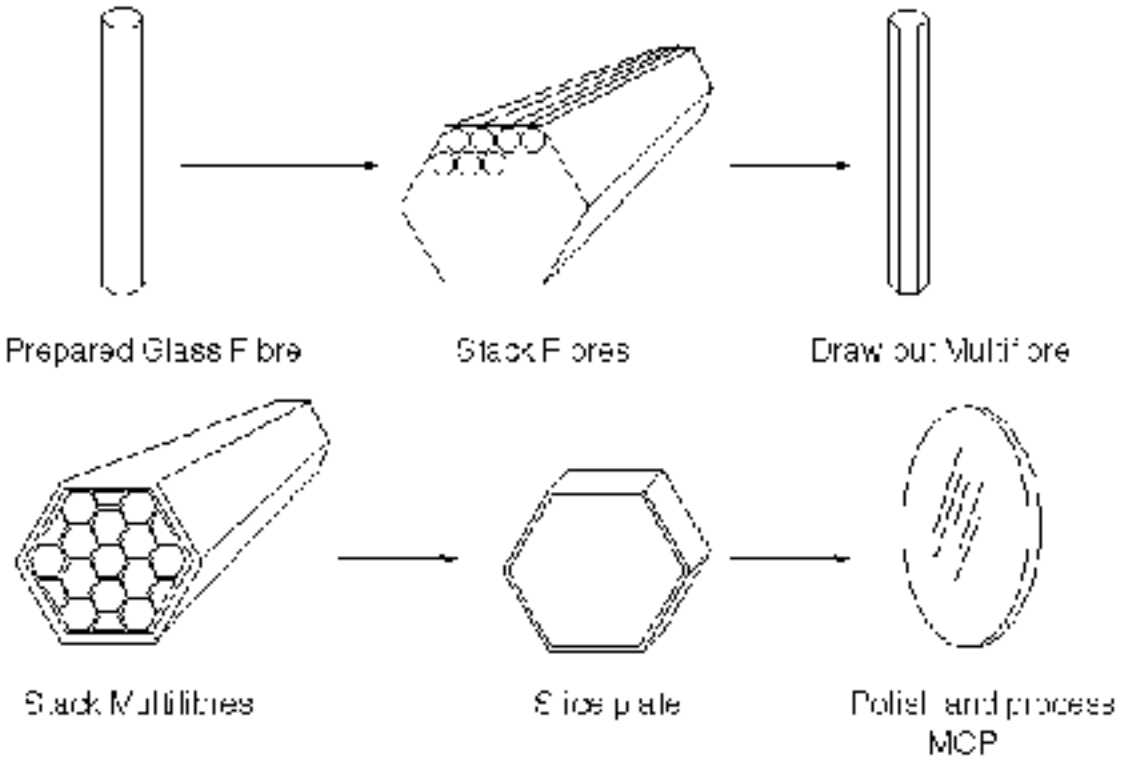


Micro-channel plates (MCPs), flat glass discs containing millions of small diameter pores, have been used for many years as compact electron multipliers: mainly in night vision image intensifier tubes (which provide the now familiar grainy green images we see on our televis screens during night-time news reports from war-torn regions of the world). More recently, similar devices have been used to focus X-rays by reflection from the inside surfaces of the channels. These provide a compact, low mass, relatively inexpensive alternative to conventional X-ray focusing mirrors. The technoogy shows great potential for imaging in X-ray astronomy.

MCPs were developed in the late 1960s as compact imaging electronmultipliers - each channel acting as a continuous dynode multiplier. They have been used in image intensifiers and in detectors of various photons,X-rays, neutrons, and charged particles. In

these applications MCPs usually have circular channels - simply for ease of manufacture. The reflectiongeometry that MCP optics employ requires the channels to be square, as shownin the picture above.

Manufacture of MCPs. MCPs are manufactured by combining individual glass fibres to create the channels and glass matrix which comprise the plate. These fibres are prepared by inserting a bar of etchable `core' glass into a tube of `cladding' glass. These fibres are then stacked into an array which is drawn out to produce a hexagonal `multifibre'. The multifibres are stacked to form a hexagonal `boule' and then sliced to produce individual MCPs which are polished and then have the core glass etched out to open up the channels. At this point MCP destined to become electron multipliers are fired in a hydrogen oven to render the glass surface semi-conducting and allow replacement of photoelectrons via the electrodes which are deposited by evaporation on the faces of the plate. These last two stages are unnecessary in the case of MCP optics.



This diagram shows the conventional type of MCP with circular channels. To manufacture optic-type square channel MCPs, square fibres are packed in square, rather than hexagonal, arrays.

Fibe optics

The cylindrical dielectric waveguide, in the form of an optical fibre, is now the world's first choice medium for long distance communications. This market demand has also prompted the design of a large number of waveguide devices and components for use in fibre optic systems, such as power splitters, combiners, multiplexers etc. Conceptually, the simplest optical waveguides are the step index and graded index planar waveguides, and the most straightforward way to introduce students to the basic principles of waveguiding is to examine the ray model of such waveguides. In this manner students can effectively learn about the principles of modes, mode cut-off, single mode operation and basic waveguide design; concepts which are fundamental to the understanding of light propagation in optical fibres, and to the design of integrated optical components.

Transitions

Two energy levels i (upper) and j (lower). Electrons can make three types of transitions. 1) spontaneous decay from upper to lower, with emission of photon $h\nu_{ij} = E_i - E_j$. Probability A_{ij}. 2) Atom with electron in lower energy state can absorb a photon by transition to upper level. Energy density per unit frequency is $\rho(\nu)$, then probability is $B_{ji}\rho(\nu)$. 3) Induced decay from E_1 to E_j in the presence of radiation: probability is $B_{ij}\rho(\nu)$.

Einstein coefficients A and B related by fundamental thermodynamics. Thermal equilibrium says number of atoms in any quantum state is given by Boltzmann distribution, so N_j/N_i related to $h\nu$ (but allow for degeneracy). Radiation is in thermal equilibrium so given by black body level (gives (ν)). Total rate at which atoms make transitions from i to j is equal to the rate from j to i relates A to B. B calculated by Dirac for electric dipole transitions in terms of the line strength, the square of the atomic dipole moment. Alternatively can express in terms of oscillator strength f, the ratio of the number of classical oscillators to the number of lower state atoms required to give the same line-integrated absorption (it describes the relative strength of the transition).

Equilibria

Thermal

Atoms adopt Boltzmann between all possible states and radiation energy density corresponding to all possible transitions has the b;lack body level of system temperature. Never achieved (maybe in stars).

Local thermal

As above but radiation not necessarily thermal.

Saha Boltzmann populations

Consider complication of allowing state (e.g. j) to have an ionized atom and thus the electron is free. Now must consider possible ionization states (single or multiple). Recognize degeneracy of states i and j, because of statistical weight (degeneracy of quantum level) and when ionized, the number of possible states of the free electron. Result is

$$\frac{n_e n_{i+1}}{n_i} = \frac{g_{i+1}}{g_i}\left[\frac{2m^3}{h^3}\left(\frac{2\pi T}{m}\right)^{3/2}\right] e^{-\chi_i / T}$$

where n refers to number density, n_i atoms in ionization stage i, n_{i+1} in ionization stage i+1, χ_i is ionization energy(energy difference between states i and i+1), g are the number of states in the specified level. For H ground state n_i versus the ion i+1, χ_i = 13.6 eV so that at T = 1 eV $n_{i+1}/n_i = 4\text{x}10^{21}\text{m}^{-3}/n_e$, so that if $n_e < 10^{21}\text{m}^{-3}$ the plasma is almost fully ionized.

Non thermal populations

LTE requires high n so that collisional transitions dominate radiative transitions (otherwise radiation important). For the case where this is not true, rate coefficients must be used to calculate populations. The important electron processes are

Radiative

transitions between bound states,
free-bound transitions (recombination and photo ionization)

Collisional

electron impact ionization/de excitation
impact ionization/three body recombination,
dielectronic recombination/auto ionization

Collisional processes a and b correspond to radiative a and b, except that the transitions are induced by electron collision and there is no spontaneous decay. 2c is a process where an atom capture an electron into an upper level while using the electron's energy loss to excite another electron already in the atom to an upper level. Inverse is auto ionization

Transition probabilities for radiative processes are determined by Einstein coefficients. For collisional processes, (e.g. radiative recombination) use individual cross sections $\sigma_{ij}(v)$ defined by setting the number of such collisional events per unit path length of an electron of velocity v in a density n_i of candidate atoms equal to $\sigma_{ij} n_i$. The rate at which any atom undergoes these collisions with electrons in the velocity interval d^3v is

$$\sigma_{ij}|v| f(v) d^3 v$$

The total rate wrt all types of collisions is written in terms of the rate coefficient

$$\int \sigma_{ij} f(\boldsymbol{v}) v d^3\boldsymbol{v} = n_e \langle \sigma_{ij} v \rangle$$

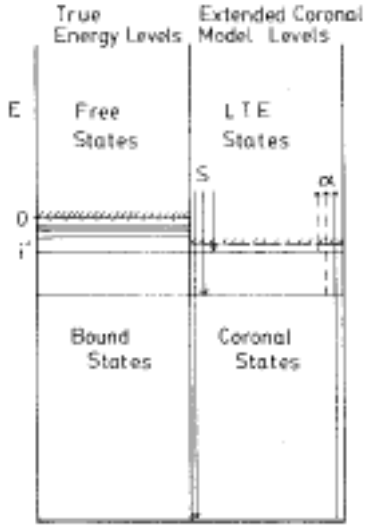
$$\langle \sigma_{ij} v \rangle = \frac{\int \sigma_{ij} f v d^3\boldsymbol{v}}{\int f d^3\boldsymbol{v}}$$

When neither complete not LTE exists one can still calculate populations of atomic levels by writing simultaneous equations for each state and then setting the total rate of transitions to and from that state equal to zero. However the number of equations is too large to handle, so that truncation must be used (assumes populations of high levels are small).

Coronal

At low density assume all upward transitions are collisional; (radiation densities low) and all downward transitions are radiative (since density low. Plasma is optically thin (radiation escapes). This only keep spontaneous emission and collisional excitation. If an excited state can be populated by collisional excitation from the ground state then this

is dominant, so excited state is determined by balancing excitation against radiative de excitation



Time dependent

Keep time derivatives in equations determining populations. Assume excited state populations are in equilibrium with the ground state even though the ground states of different ionization stages are not in equilibrium with each other.

Rate coefficients (collisional processes)

This is a separate sub field of physics. Semi classical is often OK. Consider electron of kinetic energy E losing DE during collision.

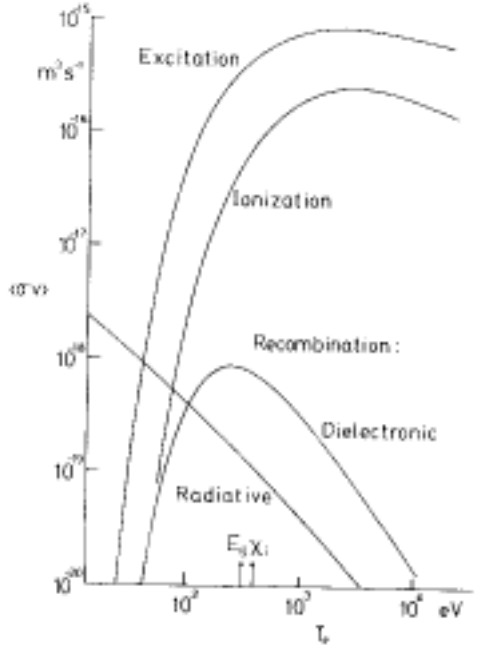
Radiative recombination. $\Delta E > E$, so final sate is negative (electron captured) and the energy change has discrete possible values $E - E_j$ where E_j is the j^{th} energy level of the recombined atom (E_j is negative).

Collisional ionization. Occurs to a continuum of states provided $\Delta E > \chi_i$ (atom's electron can escape) and $E > \Delta E$ (incident electron remains free).

Collisional excitation. Occurs with discrete ΔE ($= E_{ij}$) provided $E > \Delta E$.

Dielectronic recombination. Extension of above to $E < \Delta E$ (incident electron is captured)

Charge exchange recombination. Bound electron is transferred between two atoms in a collision. Very important as a diagnostic (see later). Can usually ignore its influence on excited states because number of neutrals is small. However radiative recombination may not dominate over charge exchange. Atomic states preferentially populated by c/x are those that have maximum overlap integral with eigen functions of the electron in the neutral H atom.



Rate coefficients for collisional processes involving CV (a four times ionized He like atom with two bound electrons). χ_i = 392 eV.

Line Broadening

Natural

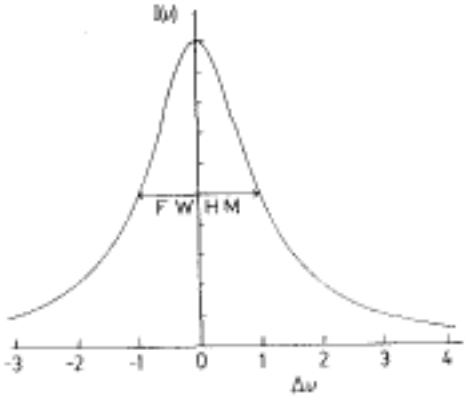
Quantum states have a spread of energy- arises because of interactions with em fields of virtual or real photons. Lifetime of an upper state is finite, hence uncertainty principle give E spread. Consider only spontaneous transitions to lower state:

$$\Delta E = \frac{h}{2\pi\tau};$$

$$\frac{2}{\tau} = \sum_j A_{ij}$$

$$\Delta\nu = \frac{\Delta E}{h} = \frac{1}{2\pi\tau} = \frac{\sum_j A_{ij}}{4\pi}$$

Line shape is determined by shape of energy broadening. It is the Fourier transform of the square root of the exponential decaying probability of the atom being in the i^{th} state, and the line shape is called Lorentzian



Doppler

Thermal motion causes a shift

$$\Delta\nu = \nu - \nu_0 = \frac{\nu_0 v}{c}$$

v is line of sight velocity. For a distribution of electrons we have a line shape

$$I(\nu) \quad f\left(\frac{\nu}{\nu_0} - 1\right)$$

and for a Maxwellian distribution

$$I(\nu) = I(\nu_0)\exp\left(\frac{-(\nu - \nu_0)^2 c^2}{2\nu_0^2 v_t^2}\right)$$

$v_T^2 = T/m$ of emitting atom. Then full width at half maximum (FWHM) is

$$\Delta\nu_{1/2} = \nu_0 \left(\frac{v_T}{c}\right)\left(2\ln(2)\right)^{1/2}$$

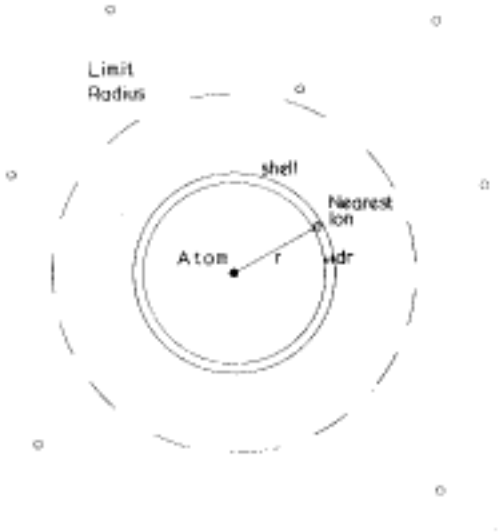
Pressure (Collisional broadening or Stark effect)

Arises from interaction of nearby particles on emitting atom. Two approaches. Collisional Lorentz approach: $\Delta\nu >> 1/\tau_p$ (τ_p is duration of perturbing interaction: occasional collisions interrupt wave train. If time between collisions is τ, then assume duration of collision << τ. The Fourier transform theory tells us that there is a frequency width associated with a sinusoidal wave form of duration $\tau = 1/\tau$. Allowing for statistics of collisions we get coherence of wave as exp(-t/τ) (Poisson statistics). Fourier transforming this give a natural line width as a Lorentzian, with τ the collision time.

Quasi static approach: $\Delta\nu >> 1/\tau_p$: atom radiates in effectively static environment during emission. An individual radiator experiences instantaneous shifts in wavelength - the average over all possible perturbations (hence shifts) gives the line width and shape. Most important perturbing effect is the E field of nearby particles (Stark effect).

Electrons approximated by impact approach, while ions by quasi static approach. For H broadening proportional to E field, and no shift. Must calculate statistical distribution of E field experienced by atoms. For ions as perturbers, provided atom is close to one ion, then E varies as $1/r^2$. Proportion of that space in that area corresponding to r and hence probability P(E) dE of experiencing E is proportional to the volume of the spherical shell $4\pi r^2 dr$, so intensity distribution is

$$I(\nu)d\nu \quad P(E)dE \quad -r^2 dr$$
$$E \quad r^{-2}; \quad -r^2 dr \quad E^{-5/2}dE$$
$$\Delta\nu \quad E;$$
$$I(\nu)d\nu \quad (\Delta\nu)^{-5/2}d\nu$$



Nearest neighbor

Cut off must be applied when other ions affect E, i.e. when $4\pi r^3/3 = 1/n_i$, i.e. at $\Delta\nu$ proportional to E proportional to $n_i^{2/3}$. Full calculations give FWHM as

$$\Delta\lambda_{1/2} = \frac{\Delta\nu_{1/2}}{\nu_0}\lambda_0 = 0.04 n_{20}^{2/3} \quad nm$$

Multiple broadening

Must convolve

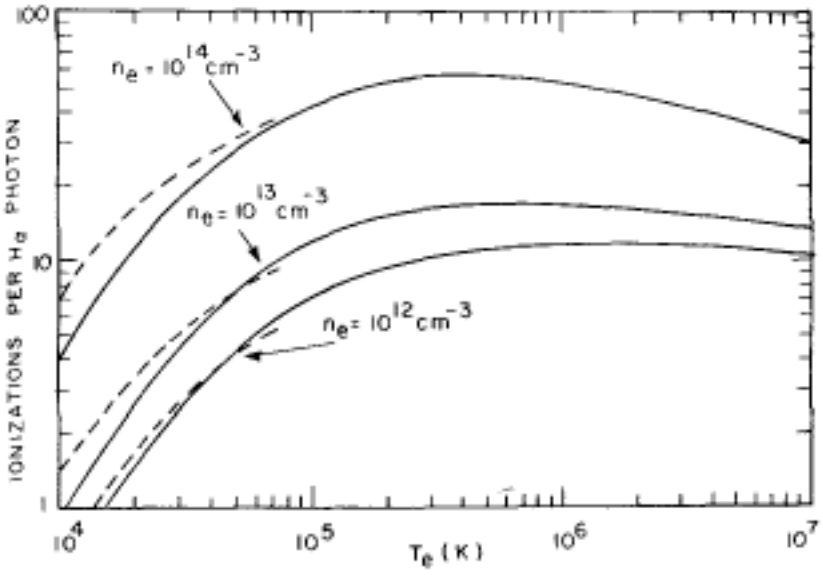
$$f(\Delta\nu) = \int f_1(\Delta\nu - \Delta\nu')\, f_2(\Delta\nu')\, d(\Delta\nu')$$

Two Gaussians: Gaussian with width = square root of sum of squares. Two Lorentzians: Lorentzian with width = sum of widths. Lorentzian plus Gaussian = Voigt.

Line Intensities

Measure absolute line intensity. Know A_{ij} then have direct measurement of excited state density. Use model to obtain ground state density. Example for H is measurement of plasma source rate $S = n_n n_e \langle\sigma v\rangle$. Because excited state density and ionization rate depend on collisions, in coronal equilibrium the ratio of ionization rate to line intensity is independent of n. For $T > R_y$ the ratio is also approximately independent of T because

virtually all collision velocities are capable of ionizing or exciting the atom. Therefore ratio of H_α emission to ionization rate is constant.

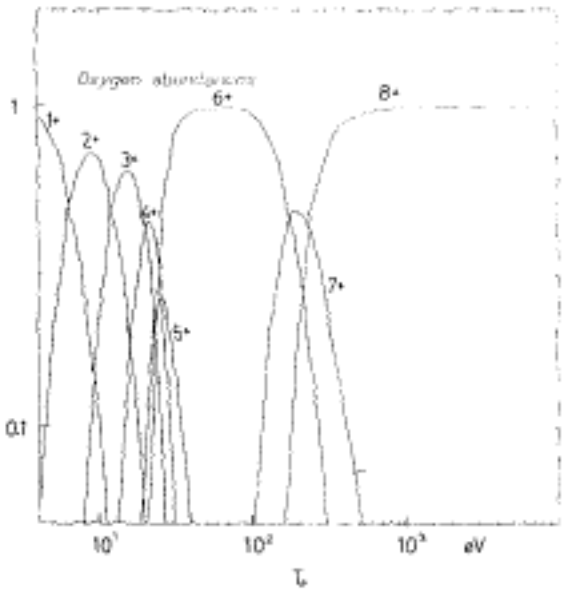


Number of ionization events per H_α photon

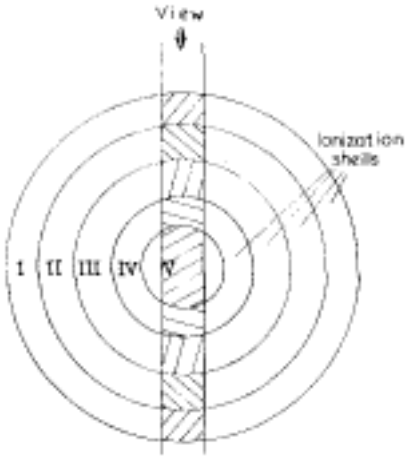
For higher charge atoms, a model for the excited state population and transition rates are required. Usually want total density of a specific ionization stage - then need to know ionization stage populations and measurements of at least one ionization stage. . In ionization equilibrium the dominant ionization stage is a strong function of Te. See figure ionization to next highest level occurs when collisional excitation exceeds radiative recombination, i.e. when

$$T > \frac{\chi_i}{\ln\left[18mc^2/\left(Z^2\chi_i\right)/\left(\chi_i/T\right)\right]}$$

Note need $T_e(r)$ and ionization balance to estimate impurity density.



Fractional abundance's of different ionization stages (oxygen in coronal equilibrium)



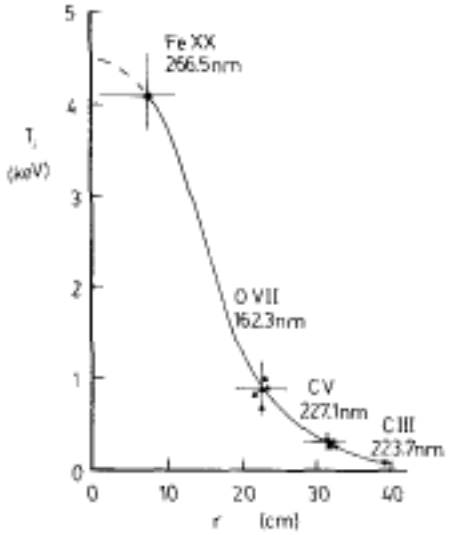
Different ionization stages appear as shells

Taking ratios of intensities of lines from the same species cancels the direct density dependence and leaves a temperature dependent parameter. Thus line ratios give a temperature measurement.

Doppler width

Gives direct measurement of temperature. If equilibration time is fast enough, this can represent the working gas ion temperature. Choosing different impurity species gives a

profile of T_e, because they position of the shells of different ionization stages f different impurities can be measured.



Flow velocity

If mean velocity parallel to line of sight is appreciable, then the emission line is shifted. Because thermal velocity is prop to inverse square of mass, heavier ions will generally have smaller ratio of line width to line shift, and give better results. Measure difference between parallel and anti parallel using e.g. a rotating mirror to view both directions, and a fast vibrating mirror to scan the spectral line during this process.

Stark widths

Dominates only at very high density, and then line width gives density.

Bolometry

Pyro electric detectors: ferro electric crystals - change in spontaneous polarization when a lattice expansion is caused by heating. Current output proportional to temperature change rate.
Thermopiles: The Sebeck effect in antimony and bismuth,

Bolometers: thermistors, semiconductor foils, metal foils, composite and superconducting.

To obtain the total radiated power out, one could perform an elaborate sum over all different states and transitions. Instead we use a radiation bolometer. This requires the spectral response to be independent of frequency as far as is possible. One can use an absorbing element whose temperature rise (e.g. measured as a resistance rise, with a Wheatstone bridge) is equal to the total energy flux divided by the bolometer's thermal capacity. Solid state detectors (e.g. surface barrier detectors) can be used, but uncertainties in electron hole pair production makes for difficult calibration.

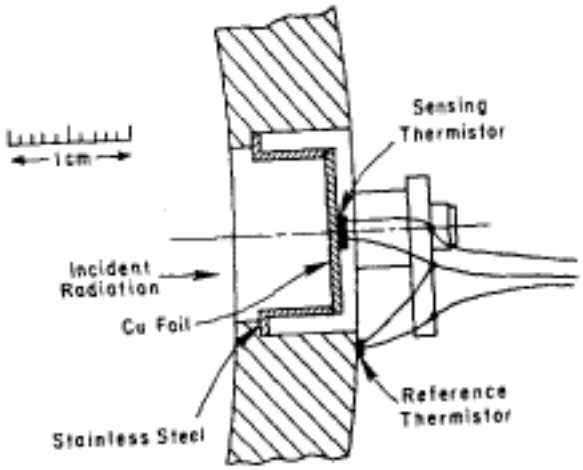
The thermal detectors work by heating the chip detector. For a detector element suspended on two wires connected to a heat sink at some ambient temperature T_a, if a radiant power P is uniformly incident, the heat balance is

$$P = C\frac{d\Delta T}{dt} + \frac{\Delta T}{R_w} + A\sigma\left(T^4 - T_a^4\right)$$

where C is the heat capacitance (JK^{-1}), R_W is the thermal resistance of the wires ($W^{-1}K$), A is the area s is Stefan-Boltzmann constant and ΔT is the change in temperature = $T - T_a$. Terms are 1) heating, 2) conduction down wires, and 3) radiative heat losses.

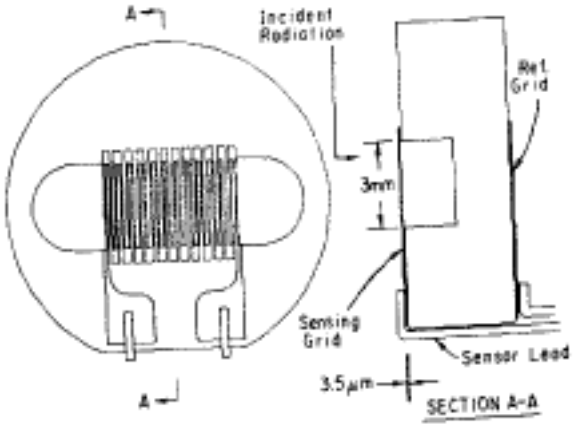
Use imaging and/or collimated system, and use sufficient chords to reconstruct the emissivity profile. Fast neutral particles also transport energy out, and these are also measured with the absorbing element technique. To remove the effect of the neutrals, different neutral gases can be employed (to absorb neutrals but not the radiation).

Mount a thermistor on the back face of a thin (30 mm) Cu foil. Epoxy is used to insulate the Cu and thermistor (wafers of manganese, cobalt and nickel oxides mounted on electrically insulating but thermally conducting sapphire. Increase temperature increase the free carrier population, thus decreasing the resistance). Foil is mounted in a stainless steel tube, which in turn is mounted on a Cu heat sink. A ref. thermistor is mounted on th heat sink. The changing resistance is measured with a Wheatstone bride circuit. The time response of the system is found by flashing a photographers strobe (duration about 1 ms). The resulting curve is used in unfolding the absorbed power from the measured temperature versus time. Typical maximum temperature is around 150 ^{0}C.



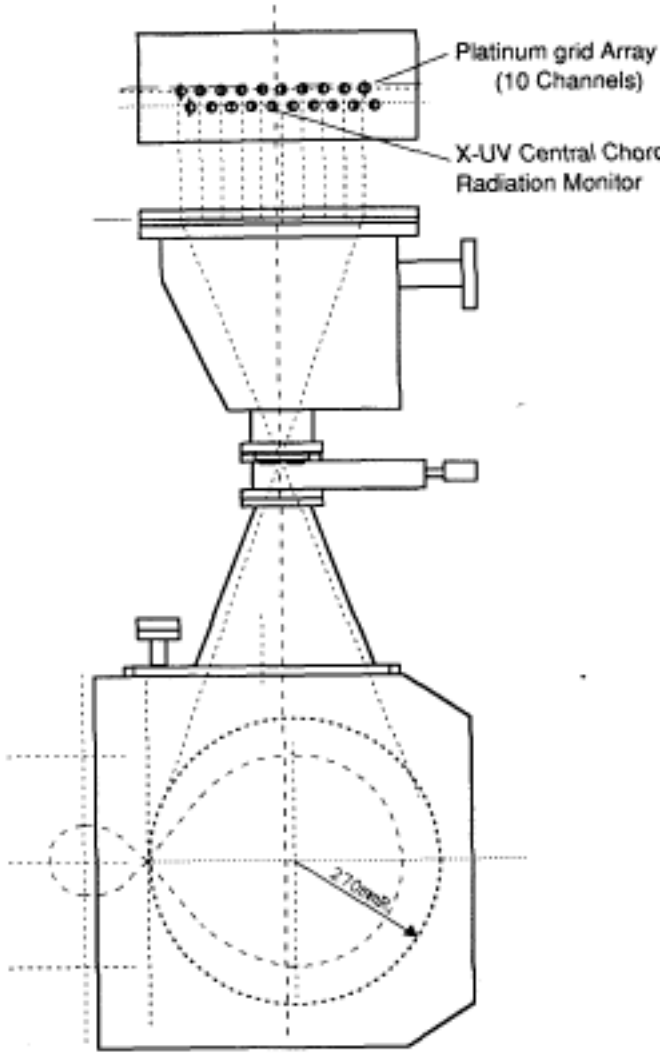
Thermistor bolometer size is 0.5 x 0.5 x 0.04 mm.

Can use a die-cut platinum grid, 3.5 mm thick, available commercially as a strain gauge or temperature sensor. Its thermal capacitance is 1/10 that of the thermistor, which offsets the disadvantage of a 10 times smaller resistance-temperature coefficient of platinum compared to thermistors. The active and reference sensors and two fixed resistors are connected as a Wheatstone bride. A feedback circuit is used to maintain the sensor at constant temperature by modulating the electrical power dissipated in the sensor. This directly measures the incident power; the output is directly proportional to the absorbed power and is independent of sensor thermal capacitance and thermal resistance. Alternatively a conventional circuit can be operated to measure the temperature.



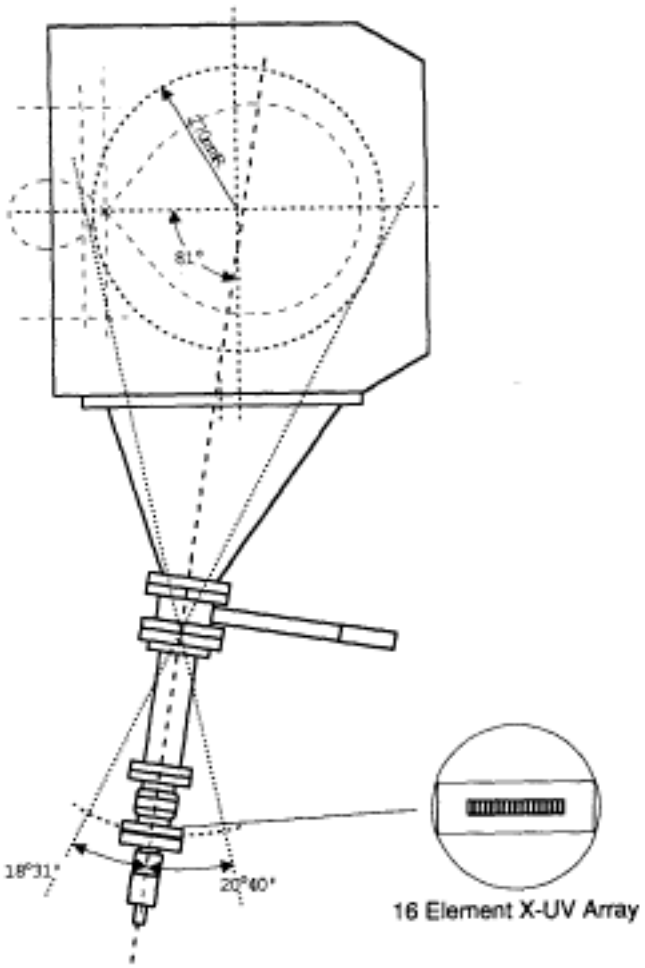
Must beware of alternating JxB forces on the strands of the sensor (noise). With the system operated in temperature feedback mode the thermal characteristics of the

detector must be used in the analysis. It would be useful to blacken the platinum to stop reflection of incident light



TEXT-U Platinum Grid Bolometer Array and X-UV Radiation Monitor (Poloidal View)

Recent advances in fabrication technology have produced a silicon X-UV enhanced photo diode having stable, flat responses to photon energies 25 eV to 6 keV with quantum efficiencies around 100% and linearity over 6 decades. The high sensitivity, fast response (μs) makes for a very good bolometer system



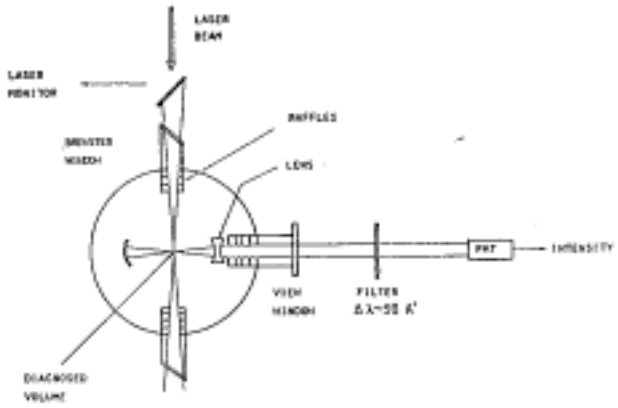
TEXT-U High Resolution Scannable X-UV Photodiode Array (Poloidal View)

Resonance Fluorescence

Use a tunable laser to radiate plasma with em radiation at some resonance line frequency. Typically use a dye laser at e.g. H_α. Radiation induces transitions between atomic levels of chosen transition. If radiation is intense enough, the induced transitions dominate al other processes (spontaneous and collisional) between the levels, and the result is to equalize the occupancy of each quantum state of the two levels. i.e. the effective temperature is infinite.

Generally enhance population of upper state, so that observed radiation increase. When radiation observed is from the same transition as the excited transition, we have fluorescence. Note radiation observed is actually from spontaneous transitions from upper level. Induced photons are emitted in the direction of the laser beam and observed photons are not.

For hydrogen Balmer series (H_α, etc.) the population of the upper level is increased by two or so, as initial populations of the n = 2 and n = 3 levels are about the same. Detecting the fluorescence gives information on neutral population with spatial localization.



Localization using resonance fluorescence.

Zeeman splitting

A B field splits an emission line into three components, an unshifted π component and two symmetrically shifted σ components. The shift of the σ components is approximately proportional to the magnitude of B.

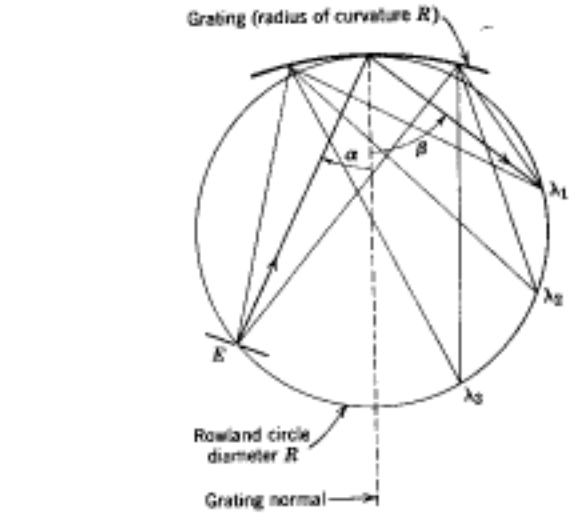
The polarization dependence of the Zeeman effect can be used too obtain the B field direction. When viewed perpendicularly to the B field, the π component is polarized with the E vector parallel to the B field and the σ components perpendicular. By measuring the angle of polarization of one of these components the B field angle can be found.

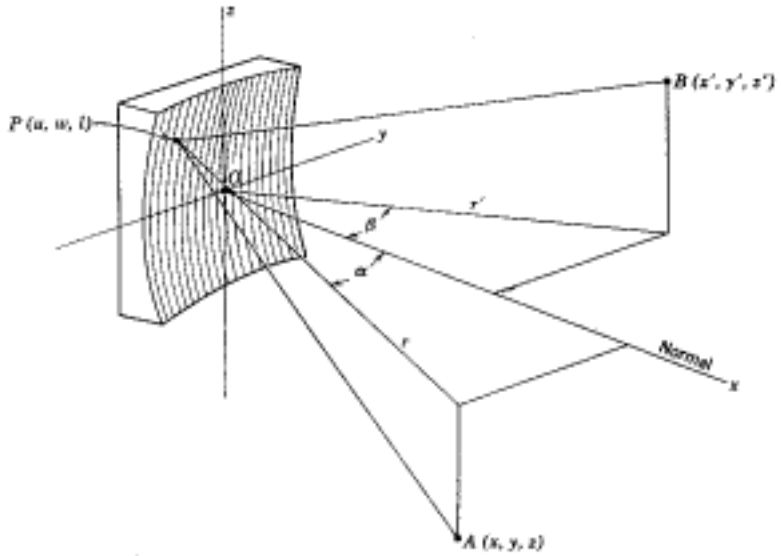
Passive systems suffer from a lack of spatial resolution, and a dominating edge emission region. An injected Li beam enhances the results, and localizes the emission

region. Further improvements follow using resonance fluorescence of the injected Li beam.

Spectrometers and monochrometers

High absorbency of air, need vacuum instruments to view $\lambda < 2000$ A. Dispersion element is usually prism or diffraction grating. Advantages of gratings are wider wavelength coverage, higher resolving power, less scattered light, almost linear dispersion, precise wavelength measurements





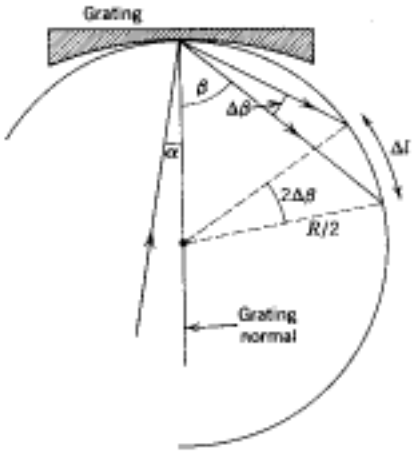
Concave diffraction grating

Rowland 1882. If a concave grating is placed tangential to a circle of a diameter equal to the radius of curvature of the grating such that the grating lies on the circumference, the spectrum of an illuminating point will be focused on the circle. Condition for two rays from adjacent grooves to reinforce at B is that path length must equal an integral number of wavelengths. Path difference for two rays reflected from grooves w apart with a constant groove separation d is mλw/d. Thus reinforcement of all rays at B for arbitrary path APB, where P varies over grating surface, represented by a path function F = AP + BP + mλw/d. Now Fermat says that point B is located such that F will be an extreme for any point P. A and B are fixed, while P may be any point on the grating, so that the

condition for F to be an extreme means $\partial F/\partial l = 0$ and $\partial F/\partial w = 0$. The solution that for B to be the best horizontal focal point means that r must be a circle - the Rowland circle. Diffracted light of all wavelengths will be focused horizontally in the circumference of this circle equal to the radius of curvature of the grating, provided the entrance and grating are on the circle and the grating is normal to the diameter. With vertical entrance slits and vertical rulings it is important that the spectrum be focused onto the horizontal plane.

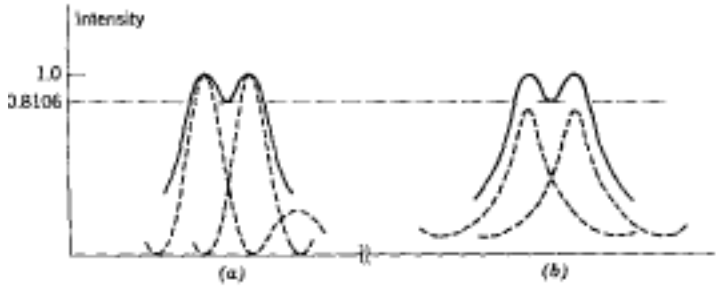
Dispersion

Describes how wavelengths are distributed along the Rowland circle. Want angular dispersion, or more usually the number of angstroms per mm along the circle.. For a 1 m normal incidence spectrometer with 600 line per mm grating, the plate factor is 16 A per mm in first order.



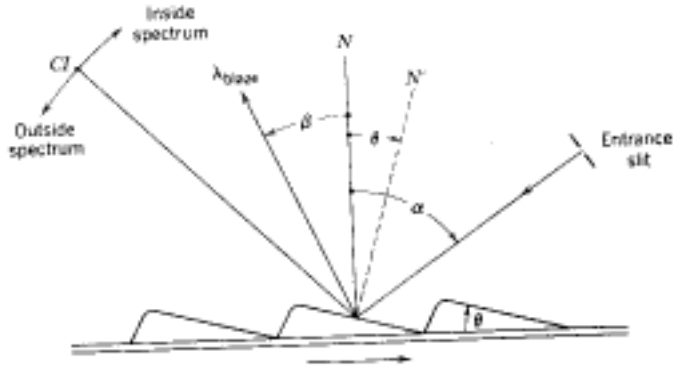
Resolving Power

Determines whether the above separation can be distinguished. Each monochromatic beam itself forms a diffraction pattern, the principle maxima of which are represented by order m. Between such maxima secondary maxima whose intensity decreases the number of ruled lines increases. The angular half width of the principle maxima is the angular distance between the principle maxima and its first minimum, and provides a theoretical limit to what can be resolved



Grating efficiency

Percentage of incident flux returned by grating into a given spectral order. Groove separation determines angular dispersion. Groove shape controls amount of radiation concentrated into a given order. Object is that a specific wavelength of interest must be diffracted in a direction which coincides with direction of the specularly reflected beam from the surface of the facet, i.e. α is incident angle, β is diffraction angle, then want $\alpha - \theta = \beta + \theta$.



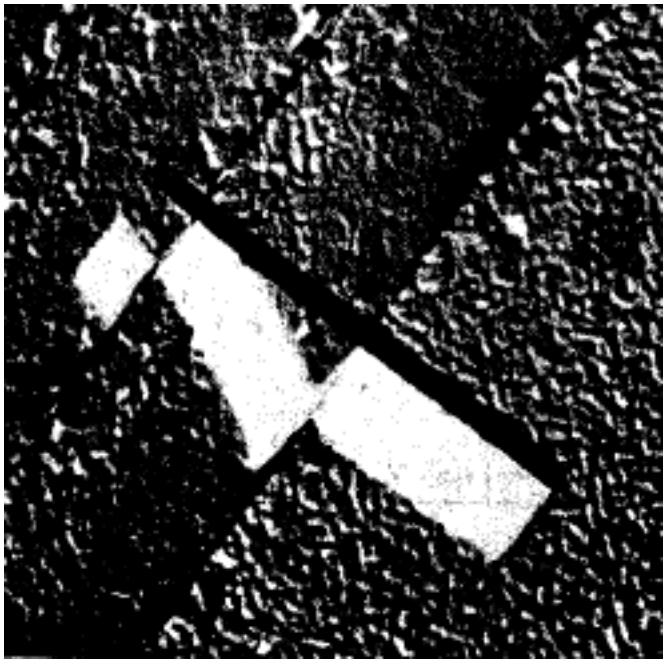
Cross section of blazed grating

$$\theta = \frac{\alpha - \beta}{2}$$

However $\quad m\lambda = d(\sin\alpha - \sin\beta)$

Therefore $\quad m\lambda_{blaze} = 2d\sin\theta\cos(\alpha - \theta)$,

$\alpha = 0$ for normal incidence.

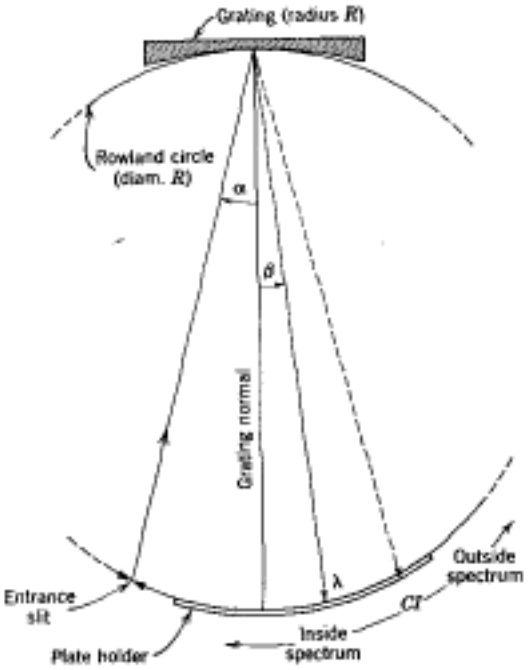


Electron micro graph.

Reflective coatings

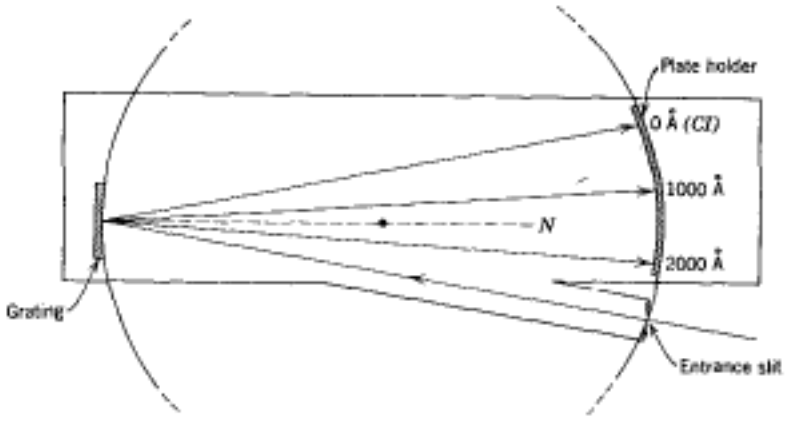
For a given wavelength, total reflection occurs at all grazing angles up to a certain maximum, beyond which reflectance drops rapidly. Conversely, for a given grazing angle all wavelengths longer than some minimum are reflected. As density of surface increases, shorter wavelengths can be reflected. Gold is good.

Basic spectrometers



Basic equation is $\pm m\lambda = d(\sin\alpha + \sin\beta)$ Two types, normal incidence (300 to 2000 A) and grazing incidence (< 300A). Referred to by size of grating and mount, e.g. 1 m normal incidence means 1 m radius of curvature and normal incidence mount.

normal incidence $\alpha < 10^0$.



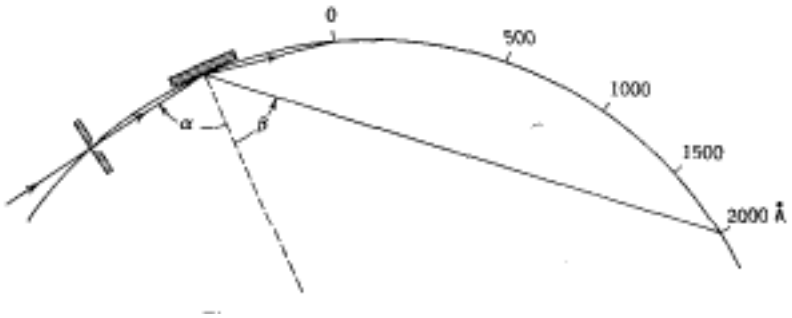
Positions of first order for a 1200 line per mm grating and α = 10 degrees.

Grazing incidence. This is required for decreasing wavelengths because of the decrease in reflectance of all grating materials for wavelengths of around 200 A and below. At an

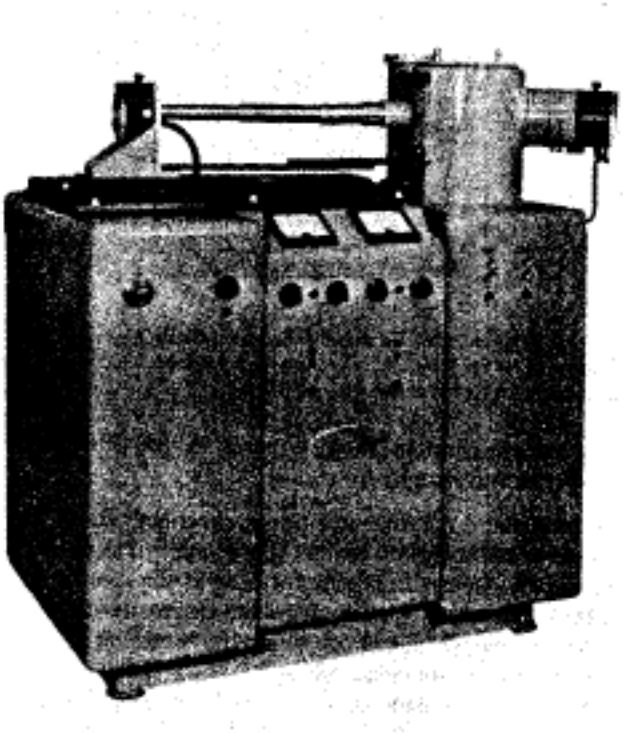
angle of about 2^0 the wavelength cutoff is at best around 5 A· i.e. it is impossible;to look below 5 A with a grating. To get below this (X-rays) one must go to crystals. The regular lattice spacing takes the place of the ruled grating, and according to Bragg's law the grating equation becomes

$$m\lambda = 2d\sin\theta$$

with θ the grazing angle and d the crystal plane separation. The maximum wavelength which can be diffracted is λ = 2d, so maximizing d is useful. Crystals with 2d = 100A have been made.

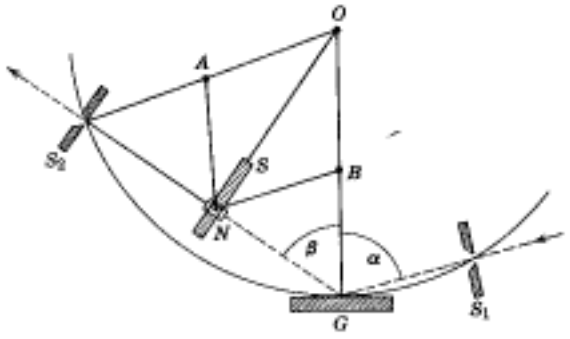


Monochrometers



A McPherson 2m grazing incidence scanning monochromator

To maintain focus and still achieve monochromatic radiation from the exit slit, either one element must move along the Rowland circle or the Rowland circle must rotate about one of the elements. The two elements must move appropriately so that they stay on the circle. In the example below of a grazing incidence system, both slits are fixed and the grating moves on the circle. The angle of incidence and diffraction change, but the angle subtended by the two slits at the center of the grating is constant (i.e. $\alpha - \beta =$ constant).



There are other ways to solve this, e.g. cutting several exit slits in an endless metal band constrained to move on a Rowland circle.

Appendix

p-n junctions

Intrinsic semi conductor - The outer band is full, but separated from the next band by a relatively low energy (1 eV). Thermally excited electrons (from valence to conduction band) leave holes in the valence band. Thus conductivity increases with temperature.

Doped semi conductor. Add an impurity to an intrinsic semiconductor. n type: impurity has one more outer electron than the intrinsic semiconductor (e.g. germanium doped with arsenic). Arsenic atoms occupy random positions in the crystal lattice. E levels of arsenic are slightly different from those of germanium. The outer electrons from the arsenic reside in energy levels which are slightly below the conduction band. With a small amount of energy (thermal excitation) electrons are boosted into the conduction band without leaving any unfilled states (holes) in the valence band. In a p type, the impurity atoms have one less outer electron than the intrinsic semiconductor (germanium doped with gallium). The gallium atoms occupy random places and leave holes in energy levels which are just above the valence band. Electrons can be thermally excited from the

valence band. This creates holes in the valence band without putting electrons in the conduction band.

p-n junction. A diode. Some electrons near junction diffuse from n-type into p-type, and holes diffuse other way (holes from p to n). This leaves excess of fixed positive charge in the n type region, and an excess of fixed negative charge in the p type region, near the boundary. This produces an electric field (10^7 V/m) which keeps the region free of mobile charges. The potential difference across this 'depleted' region is called the contact potential.

After initial redistribution, net current across region = 0. Fermi energies are the same for both p and n type. The p-type region has mobile holes at the top of the valence band, but they cannot cross into the n type region because of the contact potential. The n-type has mobile electrons at the bottom of the conduction band, but they cannot cross into the p type region.

Reverse bias - (E field same as hat of contact potential). positive voltage to n-type region. effectively increases contact potential. Potential appears across the depletion layer. No charges can cross opposing the potential. But a small current can flow, electrons in p or holes in n can move.

Add forward bias (E field opposite to that due to contact potential) - increases possibility of charges crossing region. Mobile charge carriers, the most energetic, are described by Maxwell Boltzmann distribution. A charge contributes to the current if it has enough energy to cross the barrier. Mobile electrons crossing from n to p, and mobile holes crossing from p to n, thus give a current varying as (exp(eV/kT)).

$$I_1 = \left(I_n + I_p\right)e^{eV/kT}$$

The current due to mobile electrons that have diffused to th p type region, and now cross back to the n type region, and for mobile holes that have diffused into the n region and now cross back into the p region, are given as a straight forward current.

$$I_2 = -\left(I_n + I_p\right)$$

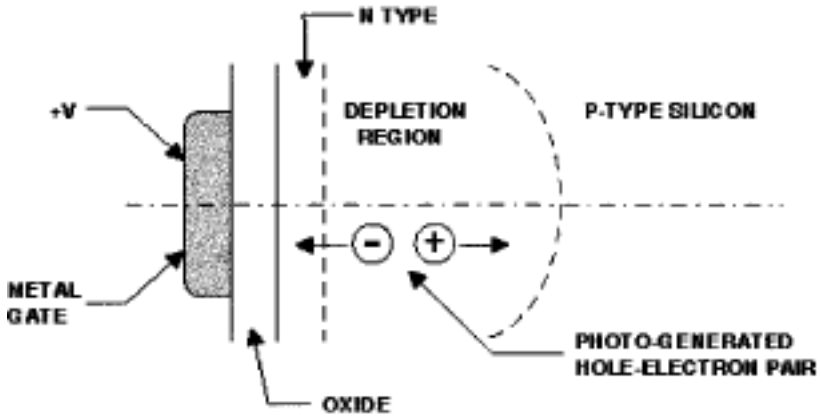
Therefore

$$I = \left(I_1 + I_2\right) = \left(I_n + I_p\right)\left(e^{eV/kT} - 1\right)$$

More on CCD

The MOS Capacitor

The simplest of all MOS (Metal-Oxide-Semiconductor) structures is the MOS capacitor. This device forms the basis for the charge-coupled device; an understanding of the structure is useful in order to comprehend CCD operation. There are two types of MOS capacitors: surface channel and buried channel. These devices differ only slightly in their fabrication; however, the buried channel capacitor offers major advantages which make it the structure of choice for the fabrication of CCDs. Virtually all CCDs manufactured today utilize the buried channel structure, and we will focus on this preferred structure. Typically, the device is built on a p-type substrate. An n-type region (~1μm thick) is formed on the surface. Next, a thin silicon dioxide layer is grown (~0.1μm thick) followed by a metal or heavily doped polycrystalline silicon layer. The latter layer forms the electrode or gate and completes the capacitor.



As usual, the electron seeks a condition of lowest potential energy. However, since the potential energy of the electron is $-|q|\Psi$, with Ψ the electrostatic potential, the electrons will seek a position where the local electrostatic potential is highest.

For the unbiased condition, the number of electrons in the n-type region is characterized by the equilibrium Fermi level. The bands are flat and the electrostatic potential is uniform within the n-type region. However, when the n-type region is fully depleted of electrons, a minimum occurs in the electron potential energy.

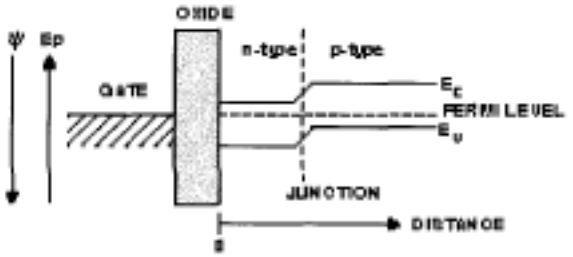


Figure 2A
Equilibrium energy band diagram.

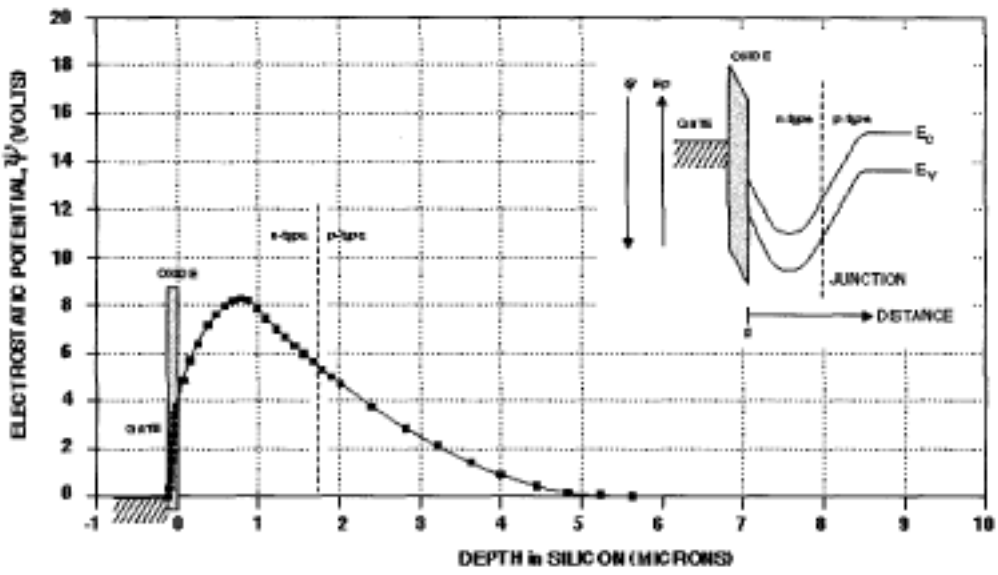


Figure 2B
Energy band diagram and electrostatic potential. Zero bias.

The minimum is located within the n-type region and at some distance removed from the silicon-silicon dioxide (Si-SiO2) interface. This minimum in the energy (or maximum in the potential) is where excess electrons collect. (The difference between the buried and surface channel capacitors is the presence or absence of the n-type layer; in the surface channel structure, the potential maximum is located directly at the Si-SiO2 interface.) The diagram below illustrates the situation where a bias of +10 V has been applied to the electrode with the substrate held at ground potential and some charge (generated optically, for example) has collected in the potential minimum.

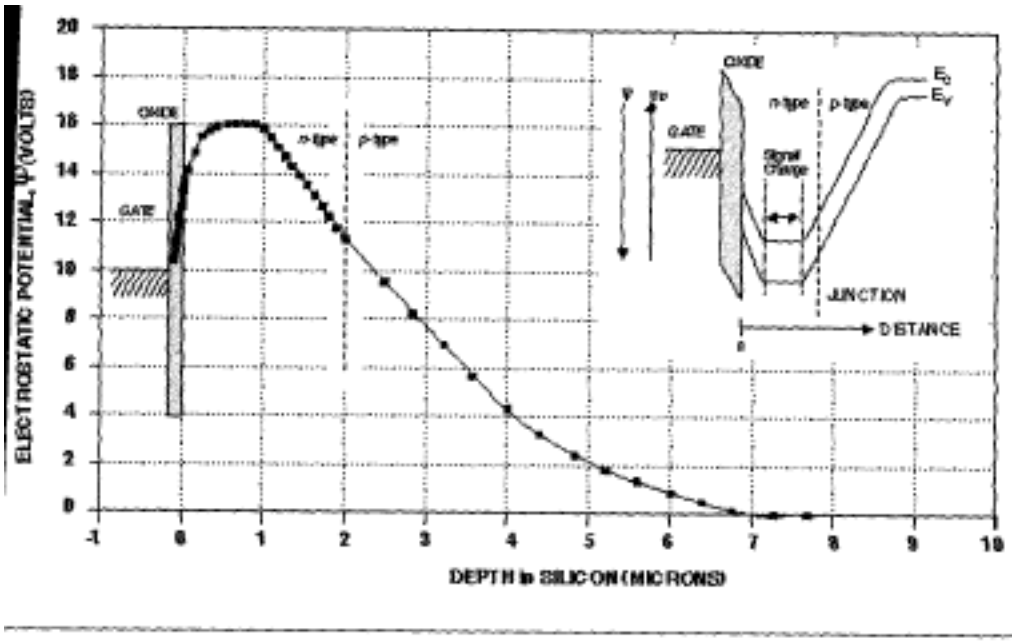


Figure 2C
Energy band diagram and electrostatic potential. 10 volt bias with charge.

Figure 2c illustrates the situation where a bias of +10 V has been applied to the electrode with the substrate held at ground potential and some charge (generated optically, for example) has collected in the potential minimum. The region where the charges collect is referred to as the channel; since this structure is below the silicon surface, the term buried channel is used. Note that the potential in the channel is flat (i.e. the field is zero) in the region where the charge has collected. If it were not, the electrons would move to make the field zero in the channel.

The region where the charges collect is referred to as the channel; since this structure is below the silicon surface, the term buried channel is used. You will note that the potential in the channel is flat (i.e. the field is zero) in the region where the charge has collected. If it were not, the electrons would move to make the field zero in the channel.

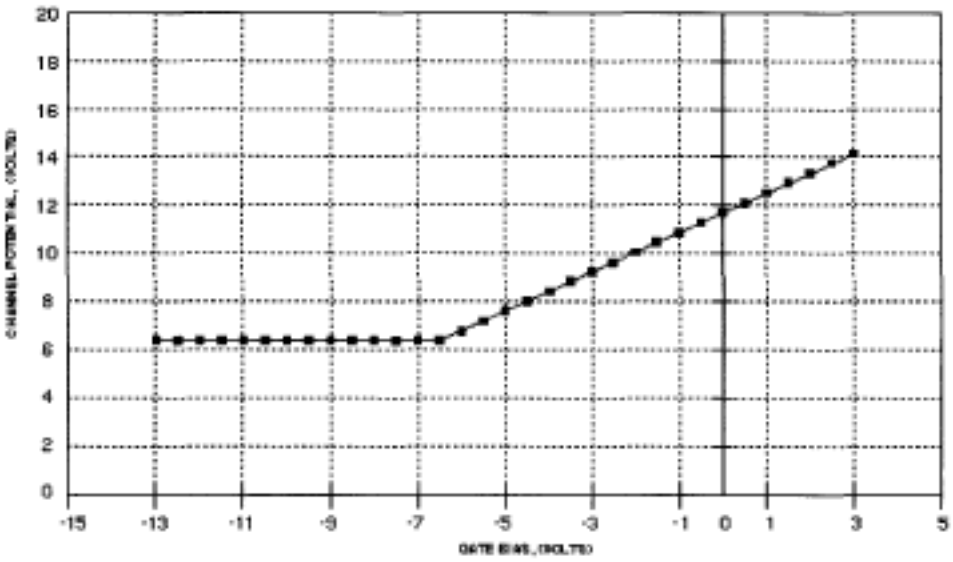
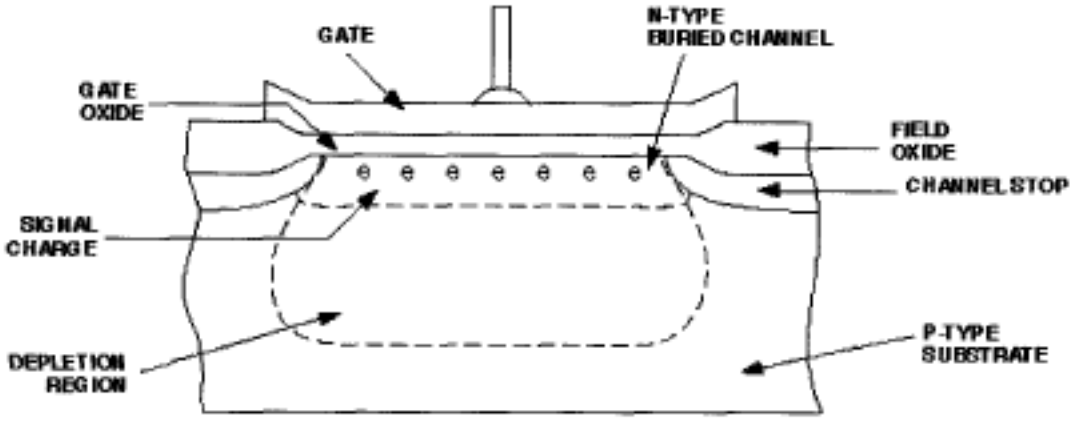


Figure 3
Typical channel potential as a function of gate bias.

The measured variation in channel potential, with the bias applied to the electrode, is illustrated for a typical buried channel capacitor. Over much of the range of the gate voltage, the channel potential is essentially linearly related to the gate bias. However, for values of the electrode bias less than about -6 V the channel potential is pinned at a fixed value. This state is termed inversion. Under these conditions, the bias on the electrode is sufficiently negative to attract holes to the Si-SiO2 interface. Any further decrease in the gate voltage merely attracts more holes to the surface and does not affect the underlying channel potential. This pinned feature of the channel potential curve is utilized in MPP operation which is discussed later. The exact form of the channel potential curve depends principally on doping concentrations, and the oxide thickness.



A practical buried channel structure is illustrated. The additional feature presented in this figure is the presence of field oxide regions on either side of the buried channel structure. These regions usually have heavily doped p-type diffusions beneath the oxide layer and because of the heavy doping combined with the thickness of the field oxide (~0.5-1.5µm), the electrostatic potential beneath the gate in this region is relatively insensitive to the voltage or changes in the voltage on the gate. In this way, charge which is collected in the buried channel device can be confined to a region beneath an electrode on top of thin gate oxide. These confining regions are called channel stops. It is not necessary that the electrode be surrounded on all sides by field oxide. Note that the potential energy of an electron is lower under an electrode that is biased at +10 V than one that is biased at 0V. Consequently, a neighboring gate can also be used to confine the charge. In summary, the features of the buried channel capacitor that make it attractive in scientific imaging applications are:

* the ability to create a potential well in a local region beneath a single electrode;

* the ability to modulate or control the potential under the gate;

* the storage location (channel minimum) is positioned away from the Si-Si02 interface and the states located there;

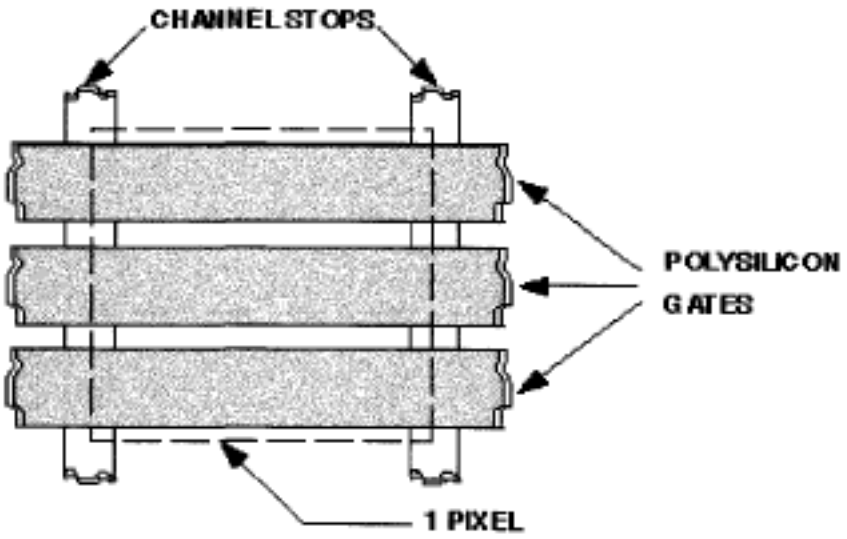
* low dark current makes it possible to store signal charge forlong periods of time (tens of seconds to hours depending on operating conditions);

* the charge that collects can be generated optically, injected electrically, or created by charged particles such as cosmic rays, protons, or high energy photons (x-rays, gamma-rays);

* the ability to move charge from a position beneath one electrode to a second, neighboring electrode rapidly and with very low loss.

CCD Pixel

A single CCD pixel (picture element) is illustrated.

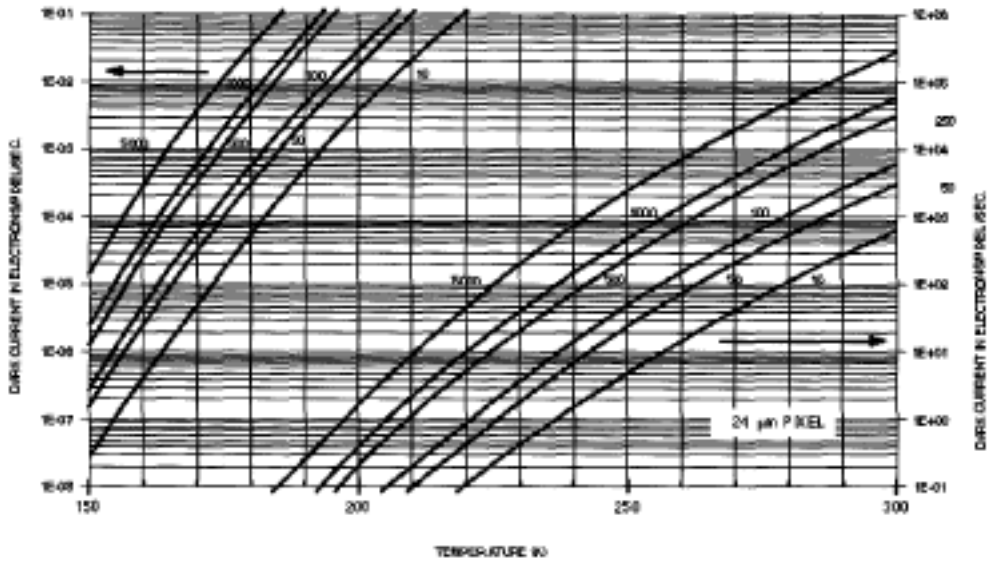


This figure shows three polysilicon gates oriented perpendicular to two-channel stop regions. Between the channel stop regions lies the buried channel. If the potential on the middle electrode is more positive than that applied to either of the other two gates, a local potential energy minimum will be formed under the middle gate. When photons strike the pixel, electron-hole pairs are created via the photoelectric effect. Electrons created within the potential minimum will be collected there. Electrons that are created in the channel stop region or in the substrate beneath the pixel may diffuse to the minimum and be collected. In either case, the holes diffuse to and are collected in the p-type substrate. The quantity of charge that is collected in the well is linearly related to the intensity of the photon flux and to the time over which the light is allowed to fall on the pixel (integration time). Other formats for the CCD pixel exist. There are structures that utilize two polysilicon gates to define a pixel and some that use four. There is even a technology

that uses a single gate in combination with multiple implants to define the pixel region. However, the technology most often used in the fabrication of scientific-grade CCDs is the three phase structure. The prevalence of the three phase technology is due principally to higher yield and process tolerance of this technology.

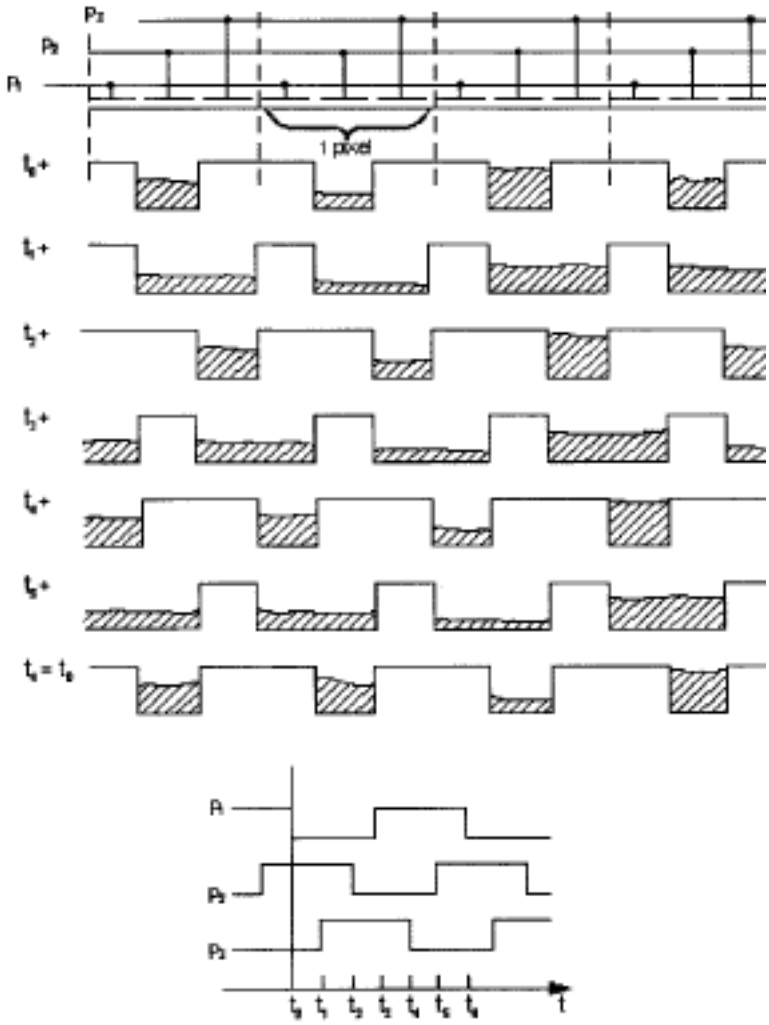
Dark Current

Dark current may be defined as the unwanted charge that accumulates in CCD pixels due to natural thermal processes that occur while the device operates at any temperature above absolute zero. At any temperature, electron-hole pairs are randomly generated and recombine within the silicon and at the silicon-silicon dioxide interface. Depending on where they are generated, some of these electrons will be collected in the CCD wells and masquerade as signal charges at the output. Large quantities of dark current limit the useful full well of the device, because the generation process is random, the dark current will contribute noise. The principle sources for dark current in order of importance are generation at the silicon-silicon dioxide interface, electrons generated in the CCD depletion region, and electrons that diffuse to the CCD wells from the neutral bulk. The first two sources usually dominate the dark current. In addition, the generation rate can vary spatially over the array leading to a fixed pattern. Fortunately, dark current sources are strongly temperature dependent. By sufficiently cooling the device, dark current can be effectively eliminated. The extent of cooling required depends largely on the longest integration time desired and the minimum acceptable signal-to-noise ratio. CCDs are most commonly cooled by using a liquid nitrogen dewar or by coupling the device to a Peltier cooler. The choice depends on how cold one must operate the device to achieve the desired performance. Calculations of the dark generation rate in e/pix/sec as a function of operating temperature for a 24μm2 pixel are illustrated.



The parameter is the dark current in pA/cm2 at a reference temperature of 293K. For example, if the system requirements dictate a dark generation rate of less than 0.0008 e/pix/sec. and one has a CCD that generates 10 pA/cm2 at 293K, the device must be cooled to -82o C. Conversely, if the requirements are such that cooling to -50oC is all that is practical and integration times are <1 second, then in order to ensure that, on average, the dark current contributes less than 10 electrons/pix/sec, any device with less than 500 pA/cm2 dark current at 293K will suffice. The curves in our diagram can be scaled for other pixel sizes simply by multiplying the dark current values by the ratio pixel area/(24µm)2.

Charge Transfer Process



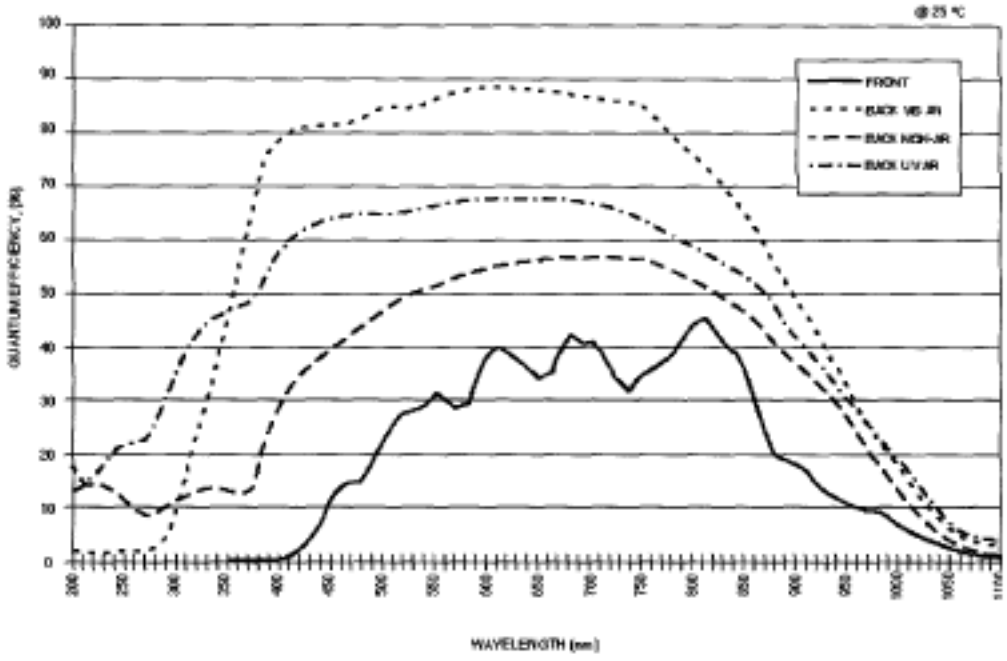
A charge-coupled device consists of a regular array of individual pixels. We illustrate a linear array of pixels, each with three separate gates, referred to as Phase 1, Phase 2, and Phase 3 (P1, P2, and P3). As shown within the linear register, all P1 gates are connected to the same bias or clock line; the same is true for P2 and P3. A single CCD pixel (as defined above) has no means to read out the quantity of charge accumulated beneath the integrating electrode. The process of reading out this signal charge involves moving the packet from the site of collection to a charge detecting amplifier located at the end of the linear array. This is the charge transfer process, which is also illustrated. Assume that charge is collected beneath the P2 electrodes. At time, t1, we apply a positive bias to the P3 electrode equal to the P2 bias, turning the P3 electrodes "on." The charge located entirely beneath the P2 electrodes spills to the region beneath P3 due to self-induced drift

and diffusion. At t2, the P2 electrodes are turned "off" (i.e., the voltage applied to the P2 gates now equals that on the P1 gates). The charge under P2 now moves rapidly to a position beneath the P3 electrodes. Note that the charge moves from a position under the P2 gates to one under P3 gates. Indeed, the process of moving or coupling the charge from one gate to its immediate neighbor gives rise to the name charge-coupled device. Repeating the process using the P3 and P1 gates results in the charge moving to a position under the P1 gates. One more cycle involving the P1 and P2 gates leaves the charge packet residing under the P2 gates again. Note that all charge packets move simultaneously one pixel to the right. In addition, the first packet moves to the charge detection amplifier where the number of signal electrons is measured. N such clock cycles are required to readout an entire N-pixel linear register. At all times within a single pixel, a well and a barrier to the next pixel co-exist. These are required in order to store the charge and to maintain the uniqueness and identity of each charge packet. Note that by interchanging the roles of any two of the gates the charge will move to the left. This flexibility is clearly another advantage of the three phase process. The effectiveness with which the transfer process occurs is measured by the Charge Transfer Efficiency (CTE). Typically, charge may be transferred with an efficiency greater than 99.999% per pixel.

Quantum Efficiency

Quantum efficiency (QE) is the measure of the efficiency with which incident photons are detected. Some incident photons may not be absorbed due to reflection or may be absorbed where the electrons cannot be collected. The quantum efficiency is the ratio of the number of detected electrons divided by the product of the number of incident photons times the number of electrons each photon can be expected to generate. (Visible wavelength photons generate one electron-hole pair. More energetic photons generate one electron-hole pair per each 3.65 eV of energy.) The energy, E, of the incident photon in eV is $E = 1.24/\lambda$, where λ is the photon wavelength in mm. For wavelengths longer than about 300 μm, the number of collected electrons per pixel per second, Ne, is related to the optical input power density, P, by the expression $Ne = 5.03 \times 10^{10} PA\lambda QE$, where P is the optical power density in μW/cm2. A is the pixel area in cm^2, λ is the wavelength in μm, and QE is the quantum efficiency in percent.

Front -Illuminated Devices



The illustration is included which depicts the typical QE versus wavelength of various CCDs. In the front illuminated mode of operation, incident photons must pass through a passivation layer as well as the gate structure in order to generate signal electrons. Photons will be absorbed in these layers and not contribute to the signal. Because of the high absorption coefficient for short wavelength photons in silicon, the quantum efficiency of front-illuminated CCDs is poor in the blue and UV regions. Interference effects in the thin gate structure cause undulations in the QE curve at the longer wavelengths.

Back-Illuminated Devices

In order to increase short wavelength quantum efficiency, there has been developed a technique for thinning the silicon wafer to approximately 15µm and mounting the CCD with the gate structure against a rigid substrate. Light is incident on the exposed, thinned surface. The incident photons do not pass through the front surface electrodes and passivation layers. An enhancement layer is then added to the back surface to create an electric field that forces photo-generated electrons toward the potential wells under the gates. An anti-reflective (AR) coating may be added to increase optical quantum efficiency. See the figue for QE versus wavelength for typical non-AR coated and AR coated SITe CCDs.

Enhancements

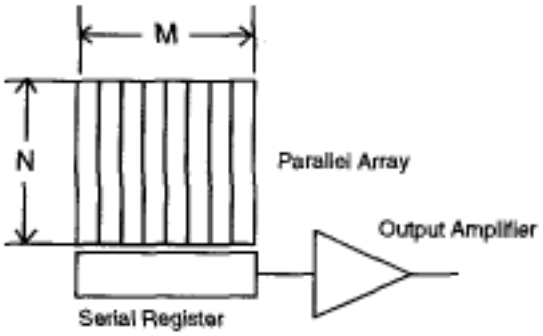
Multi-Phase Pinned Operation Multi-phase pinned (MPP) devices have been developed to reduce or eliminate the Si-SiO2 interface state contribuiton to dark current. As the gate electrode is driven to more and more negative voltages, a point is reached when holes are attracted to the interface. Any further negative voltage merely attracts more holes. The presence of the high density holes at the silicon-silicon dioxide interface fills the interface states and reduces or eliminates this contribution to dark current. During integration, it is desirable to bias all the gates into inversion. However, in the normal CCD with all gates inverted, there is no barrier to separate pixel changes. In the MPP device, an extra implant is placed under one of the electrodes (usually P3). The effect of this implant is to change the channel potential in such a way that when all the gates are biased into inversion, the channel potential beneath P3 still forms a barrier relative to the neighboring two gates. Charge collects under P1 and P2 while remaining separated from neighboring pixels. In MPP operation, it is possible to obtain dark current in the range of 10 to 50 pA/cm2 and in some cases less than 10 pA/cm^2.

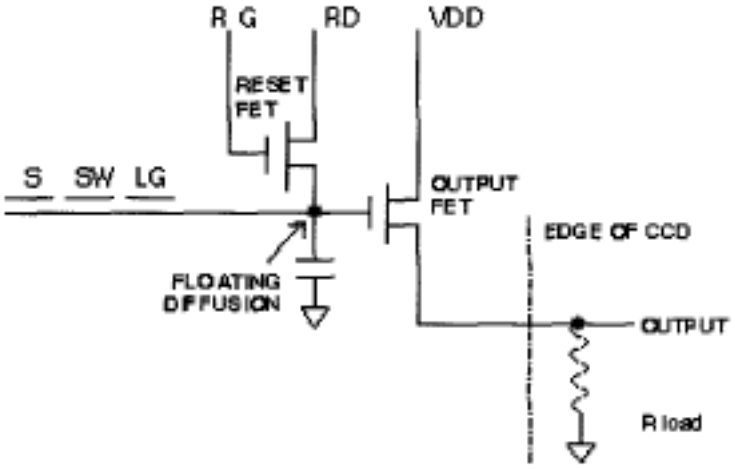
Mini-channel The mini-channel is formed by placing an extra, low dose, n-type implant in the center of the buried channel of a CCD register. Its function is to increase the channel potential locally so that small signal packets (typically 4000 electrons) will be confined to the center of the channel. The advantage of the mini- channel is that it reduces low level signal interaction with defects in the channel. These defects would otherwise reduce the charge transfer efficiency. The addition of the mini-channel is particularly important for devices which operate in a radiation environment where protons and/or neutrons produce lattice damage in the CCD channel.

The CCD Imaging Array

Imaging Array Basics Imaging arrays usually consist of square or rectangular arrays of pixels. An M x N array may be thought of as a collection of M linear registers of N pixels each. The M linear registers are aligned vertically side by side and separated by channel stop regions. An additional independent linear register is placed next to the array with its charge transfer direction orthogonal to that in the array. This serial register is arranged so that there is a single pixel adjacent to each of the M columns, and is terminated in a charge detection output amplifier. Following an integration period, array

readout involves simultaneously clocking all rows of charge packets one pixel toward the serial register.





This transfer process causes the bottom row to transfer into the serial register. The charge packets are then transferred along the serial register toward the output amplifier where they are detected. The resulting data stream is a pixel-by-pixel, row-by-row representation of the image falling on the CCD.

Output Structure At the end of the serial register is an output amplifier that converts electronic charge to a voltage. The output structure consists of two buried channel MOSFETs and a last gate. The function of the last gate is to isolate the floating diffusion from the serial clocks. The last gate has a separate lead and is generally provided a DC bias somewhat more positive than the low rail of the serial gates. The source of the

reset FET lies adjacent to the last gate. This n+-p diode can be modeled as a capacitor between the channel and the substrate. This diffusion is also connected to the gate of the output FET. When the reset FET is off, this diffusion is electrically isolated and is referred to as the floating diffusion or the output node. The total capacitance of the output node, C_T, is the capacitance of the floating diffusion plus the parasitic capacitances associated with metal leads and the gate of the output FET. Reading out a pixel begins by turning on the reset FET and setting the floating diffusion to the reset drain voltage. The reset FET is then shut off, isolating the floating diffusion. When the final serial gate voltage is dropped, any electrons stored under the gate pass under the last gate and onto the floating diffusion. The change in voltage ΔV at the output, is directly related to the quantity of charge ΔQ , that is transferred to the floating diffusion and the gain G at the output FET: $\Delta V = \Delta Q\ G\ / C_T$