Astronomy 203 Problem Set #7: Solutions

2 December 1999

1. Gain and gain dispersion in photoconductors. *In photoconductors, the photoconductive gain g is in general different for each photo-generated carrier, and is given by* $g = t/t_0$, *where* t *is the length of time the carrier lives before recombining, and* t_0 *is the* transit time, *the time it takes an unimpeded carrier to travel all the way through the detector. Like most decay times (e.g. radioactive decay),* t *is distributed exponentially, in the sense that the probability that a carrier lifetime lies between* t *and* t + dt, p(t)dt, *is given by*

$$p(t)\,dt \propto e^{-t/t_B}\,dt \quad , \tag{1}$$

where t_B *is called the* mean lifetime *of the carriers.*

a. *Normalize* p(t); *that is, find the proportionality constant in Equation* (1) *that makes the integral of* p(t)dt *equal to unity.*

Let

$$p(t)\,dt = Ae^{-t/t_B}\,dt \quad , \tag{2}$$

and set the integral over all possible *t* equal to 1:

$$\underset{1\,=}{\int_0^\infty p(t)dt} = A\int_0^\infty e^{-t/t_B}dt = t_B A\int_0^\infty e^{-x}dx = \tau_B A\left(-e^{-x}\right)\Big|_0^\infty = \tau_B A \quad . \tag{3}$$

So $A = 1/t_B$, and

$$p(t)\,dt = \frac{1}{t_B}e^{-t/t_B}\,dt \quad . \tag{4}$$

b. *Show that the average value of* g *is* $G = \overline{g} = t_B/t_0$; *that is, that* t_B *really is the mean carrier lifetime.*

Using the formula just obtained for $p(t)$,

$$\begin{aligned} G &= \overline{\left(\frac{t}{t_0}\right)} = \frac{1}{t_0}\int_0^\infty tp(t)dt = \frac{1}{t_0 t_B}\int_0^\infty te^{-t/t_B}dt \\ &= \frac{t_B}{t_0}\int_0^\infty xe^{-x}dx = \frac{t_B}{t_0}\left[-xe^{-x}\Big|_0^\infty + \int_0^\infty e^{-x}dx\right] \\ &= \frac{t_B}{t_0} \quad , \end{aligned} \tag{5}$$

where a simple substitution and integration by parts is done in the second line.

c. *Show that the gain dispersion,* $\beta = \overline{g^2}/\overline{g}^2$, *is exactly equal to 2.*

The numerator of this expression is

$$\overline{g^2} = \overline{\left(\frac{t^2}{t_0^2}\right)} = \frac{1}{t_0^2 t_B}\int_0^\infty t^2 e^{-t/t_B} dt$$
$$= \left(\frac{t_B}{t_0}\right)^2 \int_0^\infty x^2 e^{-x} dx = \left(\frac{t_B}{t_0}\right)^2 \left[-x^2 e^{-x}\Big|_0^\infty + 2\int_0^\infty x e^{-x} dx \right] \tag{6}$$
$$= 2\left(\frac{t_B}{t_0}\right)^2 \quad ,$$

where again a simple substitution and integration by parts appears in the second line, and the results of part b were used. With this,

$$\beta = \frac{\overline{g^2}}{G^2} = 2 \quad . \tag{7}$$

2. Coherent detection vs. incoherent detection. *Suppose you have five detectors, working at wavelengths 1 cm, 1 mm, 100 μm, and 10 μm, that are capable of quantum-limited heterodyne detection or background-limited direct detection in a diffraction-limited beam and in relative bandwidth $\Delta\nu/\nu = 10^{-4}$. (Never mind whether or not this is possible!) Suppose that the product of cold optics transmission and quantum efficiency is 0.2, and that $G = \beta = 1$ in each case. You use each mode of each detector to observe a 1 arcsec diameter, 1000 K object, with a 1 m diameter room temperature (300 K) Cassegrain telescope that has 20% of the primary's aperture blocked by the secondary (that is, its emissivity is 0.2). You observe in each case until a signal-to-noise ratio of 10 is achieved. How long does each of the eight measurements take? Plot the elapsed exposure time, as a function of wavelength. (You'll find it most convenient to display these results on a log-log plot.) Over what range of wavelengths is direct detection significantly more sensitive than heterodyne detection?*

   The only real difficulty of this problem is to make sure that apples are compared to apples: one must remember that the heterodyne receiver we have discussed in class detects only a single mode (one polarization and $A\Omega = \lambda^2$), but detects two IF sidebands. We therefore must restrict the direct detector to a single polarization and a diffraction-limited beam, and restrict the heterodyne receiver to intermediate frequencies of half the resolution bandwidth $\Delta\nu$, so that both systems are exposed to the same power.

   The angle θ = 1 arcsec is small compared to the FWHM widths of all of the diffraction-limited beams we're considering here, so the incident power from the source is

$$P_S = \frac{1}{2} B_\nu(T_S)\Delta\nu A\pi\left(\frac{\theta}{2}\right)^2 \quad , \tag{8}$$

   where $B_\nu(T)$ is the Planck function, T_S = 1000 K, and the factor of ½ represents the polarization restriction. The heterodyne signal-to-noise ratio is

$$\frac{S}{N} = (1-\varepsilon)\frac{\tau\eta}{\beta}\frac{P_S}{h\nu_S}\frac{1}{1+\frac{\tau\eta}{\beta}\varepsilon\overline{N}}\sqrt{\frac{2\Delta t_H}{\Delta\nu_{IF}}} \quad , \tag{9}$$

   where, $\overline{N} = \overline{N}(T_B)$ is the photon mode occupation number at the background temperature (300 K). For the direct detectors, we have

$$\frac{S}{N} = (1-\varepsilon)P_S \sqrt{\frac{\tau\eta}{\beta}\frac{\Delta t_D}{h\nu\overline{P_B}}\frac{1}{1+\frac{\tau\eta}{\beta}\varepsilon\overline{N}}} = (1-\varepsilon)P_S \sqrt{\frac{\tau\eta}{\beta}\frac{\Delta t_D}{\varepsilon h^2\nu^2\overline{N}\Delta\nu}\frac{1}{1+\frac{\tau\eta}{\beta}\varepsilon\overline{N}}} \quad , \tag{10}$$

where we have noted that the detector receives background radiation in a single mode, to match that assumed for the heterodyne detector. Solving for t_D and t_H, with S/N = 10, and accounting for double-sideband response in the heterodyne receiver, we get:

$$\begin{aligned} \Delta t_H &= 100h^2\nu^2\Delta\nu\left(\frac{1+\frac{\tau\eta}{\beta}\varepsilon\overline{N}}{\frac{\tau\eta}{\beta}(1-\varepsilon)P_S}\right)^2 \quad , \\ \Delta t_D &= 100\varepsilon h^2\nu^2\overline{N}\Delta\nu\frac{1+\frac{\tau\eta}{\beta}\varepsilon\overline{N}}{\frac{\tau\eta}{\beta}(1-\varepsilon)^2 P_S^2} \quad . \end{aligned} \tag{11}$$

Note that at very long wavelengths, at which $\overline{N} >> 1$, the times converge to the same value, one that is independent of properties of the detectors and cold optics:

$$\Delta t \xrightarrow[\overline{N}>>1]{} 100h^2\nu^2\Delta\nu\left(\frac{\varepsilon\overline{N}}{(1-\varepsilon)P_S}\right)^2 \quad , \tag{12}$$

as they must; for very long wavelengths the noise is dominated by background fluctuations, rather than photon noise. Putting in the numbers, we get

Wavelength (μm)	t_H (s)	t_D (s)
1	1.9×10^{3}	1.2×10^{-19}
10	6.4×10^{-4}	2.1×10^{-7}
100	1.7×10^{-1}	1.0×10^{-2}
1000	4.3×10^{2}	1.9×10^{2}
10000	1.1×10^{7}	1.0×10^{7}

These results are plotted in Figure 1. Three important features of coherent and incoherent detection are evident in this figure. First, the observations at the longest wavelengths take a *very* long time (about three months). Next, the integration times for the two techniques approach each other at long wavelengths. Finally, the integration times diverge dramatically at short wavelengths; direct detection gets faster monotonically with higher frequency, and heterodyne detection actually reaches a minimum. Direct detection is significantly better that heterodyne detection at wavelengths less than about 300 μm. Both of the latter effects are due to the dependence of $\overline{N}$ on T: it is very large at 1 cm, but very small at 1 μm. The rms photon noise is $\sqrt{\overline{N}(\overline{N}+1)}$, and thus keeps getting smaller with shorter wavelengths; however, coherent detection involves quantum noise in the amount $(\Delta N)_{\mathrm{rms}} = 1$ (§23.2-3), and this makes a huge difference at short wavelengths.

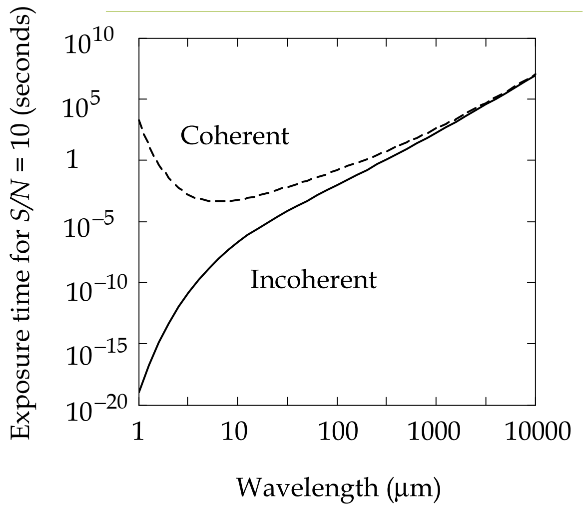

Figure 1: Exposure time as a function of wavelength, as required for S/N = 10 on a 1" diameter, 1000 K blackbody with a 1 m diameter telescope, using coherent and incoherent detection.

3. Background-limited spectrometers. *You are given a collection of spectrometers that work in the 5-40 μm range with a spectral resolution of $\Delta\nu/\nu$ = 1/1000 and (cold) instrumental transmission $\tau = 0.15$, and with detectors that have quantum efficiency 0.3 and unit photoconductive gain. You can use them either on the airborne Stratospheric Observatory for Infrared Astronomy (SOFIA) or the satellite-borne Space Infrared Telescope Facility (SIRTF). SOFIA is an ambient-temperature telescope (270 K), 2.5 meters in diameter, and 10% of its area is obscured. SIRTF is cryogenic (< 5 K), and is in space, so the only source of background is the zodiacal light. The zodiacal light can be thought of as a superposition of two low-emissivity blackbodies, one with $\varepsilon = 3\times10^{-8}$ and temperature 285 K, and the other with $\varepsilon = 2\times10^{-7}$ and temperature 200 K, for this range of wavelengths. SIRTF's primary mirror is 85 cm in diameter. Diffraction-limited beams are used. Compare the NEP of these spectrometers used for observation of point objects on SOFIA and SIRTF, at wavelengths 5, 10, 20 and 40 μm.*

SOFIA has warm optics with transmission 1 - ε = 0.9, which therefore emit a background power given by

$$P_B = \frac{8\varepsilon h\nu\Delta\nu}{e^{h\nu/kT_B}-1} \quad , \tag{13}$$

in a diffraction-limited beam (total $A\Omega=4\lambda^2$), where T_B is the temperature of the telescope. The zodiacal emission is present, but is vastly smaller than this because the emissivities are so small compared to ε = 0.1, and can therefore be neglected. The background-limited *NEP* is

$$NEP = \frac{1}{1-\varepsilon}\sqrt{\frac{2h\nu P_B}{\tau\eta/\beta}\left(1+\frac{\eta}{\beta}\tau\varepsilon\overline{N}\right)} \quad , \tag{14}$$

where as usual η = 0.3 is the quantum efficiency, τ =0.15 is the transmission of the cold optics, β is the gain dispersion (assumed to be 1), and $\overline{N}$ is the photon mode occupation number appropriate for the given frequency and the temperature of the telescope. Putting the numbers in from above, we get:

$$\begin{aligned} NEP &= 3.1\times10^{-16}\ \mathrm{W\ Hz^{-1/2}} \qquad (\lambda = 5\ \mu\mathrm{m}) \quad , \\ &= 1.6\times10^{-15}\ \mathrm{W\ Hz^{-1/2}} \qquad (\lambda = 10\ \mu\mathrm{m}) \quad , \\ &= 2.2\times10^{-15}\ \mathrm{W\ Hz^{-1/2}} \qquad (\lambda = 20\ \mu\mathrm{m}) \quad . \end{aligned} \tag{15}$$

For SIRTF, there are no warm optical components to emit background, and therefore no 1/1-ε term out in front of the expression for *NEP*. Also, the values of ε and $\overline{N}$ for both of the zodiacal light components are very small, and therefore $\eta\tau\varepsilon\overline{N}$? 1; we can ignore this term in the NEP. Thus

$$NEP = \sqrt{\frac{2h\nu\left(P_{B1}+P_{B2}\right)}{\eta\tau/\beta}} \quad , \tag{16}$$

where the P_Bs are the power emitted by the two zodiacal light components, obtained from the expression above but with the appropriate emissivities and temperatures, and where now τ = 0.15×0.9 = 0.135, since the transmission loss from the obscured part of the telescope is now in the cryogenic part of the system. Putting in all of these numbers again, one obtains

$$\begin{aligned} NEP &= 2.2\times10^{-19}\ \mathrm{W\ Hz^{-1/2}} \qquad (\lambda = 5\ \mu\mathrm{m}) \quad , \\ &= 1.3\times10^{-18}\ \mathrm{W\ Hz^{-1/2}} \qquad (\lambda = 10\ \mu\mathrm{m}) \quad , \\ &= 2.2\times10^{-18}\ \mathrm{W\ Hz^{-1/2}} \qquad (\lambda = 20\ \mu\mathrm{m}) \quad . \end{aligned} \tag{17}$$

A spectrometer on SIRTF is therefore more than 1000 times more sensitive (*NEP* is a factor of 1000 smaller) than the same spectrometer on SOFIA. This means that any given observation would take more than 1,000,000 times longer on SOFIA than on SIRTF. Now you know why we are so anxious to have a cryogenic infrared telescope in space.

Note that to make this comparison we have tacitly assumed that we were to observe objects for which the same power were contained within the two different diffraction-limited beams; since those beams differ in diameter by a factor of about 3, we really mean *point sources*. The comparison would be slightly different for extended sources, but SIRTF would still be the more sensitive instrument, by orders of magnitude.

By the way, the emissivity and temperature components given above for the zodiacal light were derived from observations by IRAS and COBE, and pertain to the *minimum* of the zodiacal emission, at the ecliptic poles. The emissivity of each component is larger by about a factor of three for directions closer to the plane of the solar system.

4. The characteristic matrix and energy conservation.

   a. *Show that the determinant of the characteristic matrix is unity.*

   For any single dielectric layer,

$$\begin{aligned} |M| &= \begin{vmatrix} \cos\kappa\ell & -i\sin\kappa\ell/Y \\ -iY\sin\kappa\ell & \cos\kappa\ell \end{vmatrix} \\ &= \cos^2\kappa\ell - \left(-iY\sin\kappa\ell\right)\left(-i\sin\kappa\ell/Y\right) = \cos^2\kappa\ell + \sin^2\kappa\ell = 1 \quad . \end{aligned}$$

You have learned in Math 164 (I hope) that the determinant of the product of two nonsingular matrices is equal to the product of their determinants (see, e.g. B. Kolman, *Elementary Linear Algebra* [London: Macmillan, 1970], pp. 136-137, if I'm wrong); thus the characteristic matrix of a stack of dielectric layers is also unity.

b. *Show that* $|r|^2 + Y_{p+1}|t|^2 / Y_0 = 1$ *, where the amplitude reflection and transmission coefficients r and t are given by Equations 25.28 and 25.29.*

Let's first define

$$M = \begin{pmatrix} a & ib \\ ic & d \end{pmatrix}$$

for convenience; then, $|M| = ad + bc$. Let's also consider the Y factors to be real numbers; this amounts to a claim that the index of refraction is real for each layer, which in turn implies that there is no absorption. In these terms, the amplitude transmission and reflection coefficients are

$$t = \frac{2Y_0}{aY_0 + ibY_0Y_{p+1} + ic + dY_{p+1}} \quad \text{, and}$$

$$r = \frac{aY_0 + ibY_0Y_{p+1} - ic - dY_{p+1}}{aY_0 + ibY_0Y_{p+1} + ic + dY_{p+1}} \quad .$$

Thus

$$\begin{aligned} |r|^2 + \frac{Y_{p+1}}{Y_0}|t|^2 &= \frac{\left(aY_0 - ibY_0Y_{p+1} - ic + dY_{p+1}\right)\left(aY_0 - ibY_0Y_{p+1} - ic + dY_{p+1}\right) + 4Y_0Y_{p+1}}{\left(aY_0 + ibY_0Y_{p+1} + ic + dY_{p+1}\right)\left(aY_0 - ibY_0Y_{p+1} - ic + dY_{p+1}\right)} \\ &= \frac{a^2Y_0^2 + b^2Y_0^2Y_{p+1}^2 + c^2 + d^2Y_{p+1}^2 - 2Y_0Y_{p+1}(ad+bc) + 4Y_0Y_{p+1}}{a^2Y_0^2 + b^2Y_0^2Y_{p+1}^2 + c^2 + d^2Y_{p+1}^2 + 2Y_0Y_{p+1}(ad+bc)} \quad . \end{aligned}$$

But $|M| = ad + bc = 1$, so

$$\begin{aligned} |r|^2 + \frac{Y_{p+1}}{Y_0}|t|^2 &= \frac{a^2Y_0^2 + b^2Y_0^2Y_{p+1}^2 + c^2 + d^2Y_{p+1}^2 - 2Y_0Y_{p+1} + 4Y_0Y_{p+1}}{a^2Y_0^2 + b^2Y_0^2Y_{p+1}^2 + c^2 + d^2Y_{p+1}^2 + 2Y_0Y_{p+1}} \\ &= \frac{a^2Y_0^2 + b^2Y_0^2Y_{p+1}^2 + c^2 + d^2Y_{p+1}^2 + 2Y_0Y_{p+1}}{a^2Y_0^2 + b^2Y_0^2Y_{p+1}^2 + c^2 + d^2Y_{p+1}^2 + 2Y_0Y_{p+1}} = 1 \quad ; \end{aligned}$$

in fact, since $|r|^2 + Y_{p+1}|t|^2 / Y_0 = 1$, multiplied by $cE_I^2 / 8\pi$, embodies energy conservation, the result of part a can be said to be *demanded* by energy conservation.

5. Antireflection coatings.

a. *Calculate and plot the transmission of a 5 μm thick wafer of diamond (n = 2.4) over the visible wavelength range,* $\lambda = 0.35 - 0.7\ \mu\text{m}$ *.*

b. *Repeat the calculation and plot with a layer of MgF_2 (n = 1.38) on each face, with thickness equal to a quarter of a wavelength for incident light at* $\lambda = 0.5\ \mu\mathrm{m}$.

These calculations can be done in one fell swoop. With vacuum on both sides of the diamond, $Y = 1$ outside it. The characteristic matrix of the diamond wafer is

$$M = \begin{bmatrix} \cos\kappa\ell_W & -i\sin\kappa\ell_W / Y_W \\ -iY_W \sin\kappa\ell_W & \cos\kappa\ell_W \end{bmatrix} \quad ,$$

where, for normal incidence, we have

$$\begin{aligned} d_W &= 0.5\ \mathrm{mm} \quad , \\ \kappa\ell_W &= \kappa n_W d_W = \frac{2\pi n_W d_W}{\lambda} \quad , \\ Y_W &= n_W \quad . \end{aligned}$$

If on the other hand the wafer is coated, the characteristic matrix of the combination is the product of the characteristic matrices of the coating layers and the wafer:

$$\begin{aligned} M = &\begin{bmatrix} \cos\kappa\ell_C & -i\sin\kappa\ell_C / Y_C \\ -iY_C \sin\kappa\ell_C & \cos\kappa\ell_C \end{bmatrix} \begin{bmatrix} \cos\kappa\ell_W & -i\sin\kappa\ell_W / Y_W \\ -iY_W \sin\kappa\ell_W & \cos\kappa\ell_W \end{bmatrix} \\ &\times \begin{bmatrix} \cos\kappa\ell_C & -i\sin\kappa\ell_C / Y_C \\ -iY_C \sin\kappa\ell_C & \cos\kappa\ell_C \end{bmatrix} \quad , \end{aligned}$$

where

$$\begin{aligned} \lambda_0 &= 0.5\ \mu\mathrm{m} \quad , \\ \kappa\ell_C &= \kappa n_C d_C = \frac{2\pi n_C}{\lambda}\frac{\lambda_0}{4n_C} = \frac{\pi\lambda_0}{2\lambda} \quad , \\ Y_C &= n_C = \sqrt{n_W} \quad . \end{aligned}$$

In either case, the amplitude transmission coefficient is given by Equation 25.28,

$$t = \frac{2Y_0}{m_{11}Y_0 + m_{12}Y_0 Y_{p+1} + m_{21} + m_{22}Y_{p+1}} = \frac{2}{m_{11} + m_{12} + m_{21} + m_{22}} \quad ,$$

and the power transmission coefficient by

$$\tau = \frac{Y_2}{Y_0}|t|^2 = |t|^2 \quad .$$

The matrix multiplication, calculation and plotting are easily done in Mathcad. You can use a spreadsheet like Excel to do it, too, but then you must multiply the matrices yourself. Note that the transmission peaks are evenly spaced in wavenumber but not wavelength. Note the significant improvement in the average transmission of the wafer, owing to the antireflection coatings.

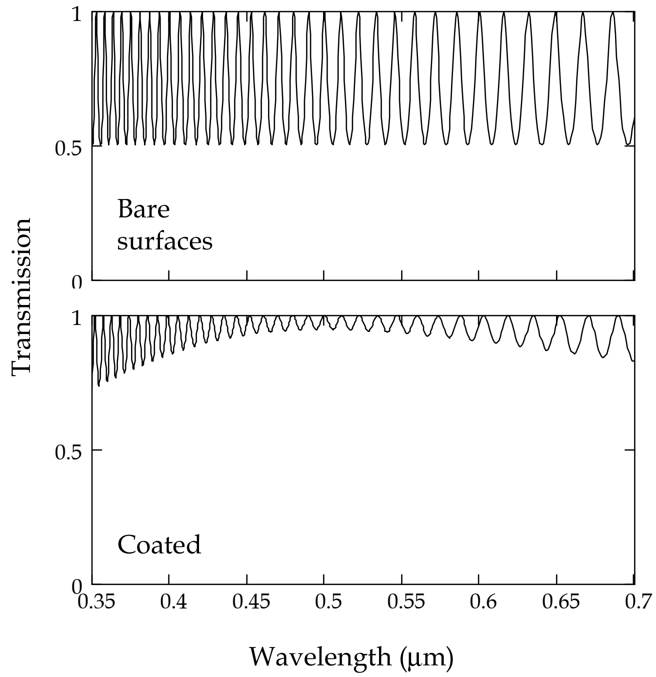

Figure 2: transmission of diamond wafer with bare surfaces (top) and with quarter-wave MgF_2 coatings (bottom).

