X-692-75-213
PREPRINT

NASA TM X- 70972

SENSITIVITY LIMITS OF AN INFRARED HETERODYNE SPECTROMETER FOR ASTROPHYSICAL APPLICATIONS

M. M. ABBAS
M. J. MUMMA
T. KOSTIUK
D. BUHL

SEPTEMBER 1975

GSFC
GODDARD SPACE FLIGHT CENTER
GREENBELT, MARYLAND

ABSTRACT

The sensitivity of an ideal heterodyne spectrometer approaches the quantum detection limit provided the local oscillator power is sufficiently large and the shot noise dominates all other sources of noise. The post-integration minimum-detectable-number of photons/sec for an ideal heterodyne system is $\sqrt{B/\tau}$ where B is the IF bandwidth and τ is the integration time. For astronomical observations, however, a number of factors (Δ_i) tend to degrade the sensitivity, a fact which becomes significant particularly when the laser power is insufficient. A discussion and an evaluation of the degradation in sensitivity is given for a heterodyne spectrometer employing a HgCdTe photodiode mixer and tunable diode lasers. The minimum detectable source brightness is considered as a function of the mixer parameters, transmission coefficient of the beam splitter and local oscillator emission powers. The degradation in the minimum detectable line source brightness which results from the bandwidth being a fraction of the line width is evaluated and plotted as a function of the wavelength and bandwidth for various temperature to mass ratios. It is shown that the minimum achievable degradation ($\pi_i(\Delta_i)$) in the sensitivity of a practical astronomical heterodyne spectrometer is ~ 30. Estimates of signal-to-noise ratios with which infrared line emission from astronomical sources of interest may be detected are given.

1. Introduction

Infrared heterodyne spectroscopy provides a powerful tool for identification of molecular and atomic species in astronomical sources and for determination of in situ physical conditions through measurement of the line profiles. In principle the sensitivity of a heterodyne spectrometer may approach the so-called quantum detection limit $h\nu\Delta\nu$,[1] but in practice a number of factors tend to degrade the sensitivity significantly. In this paper we evaluate and discuss some degradation factors with a view to optimizing the sensitivity of the system, and we discuss its advantages and limitations for astronomical observations.

The real advantages of heterodyne detection over broad-band techniques are its ultra-high spectral resolution (capable of achieving Doppler limited spectroscopy $\Delta\lambda/\lambda \leq 10^{-6}$), its high spatial resolution ($A\Omega \sim \lambda^2$), and its relatively high detection efficiency. These properties make heterodyne techniques particularly well suited to the detection of weak atomic and molecular line sources in regions as diverse as interstellar clouds, HII regions, comets, the upper atmospheres of planets and the earth's stratosphere. Information about kinetic temperatures, turbulence effects, gross velocity shifts etc., may be obtained through a study of the line profiles. Important discoveries in molecular astronomy have been made using heterodyne techniques at radio wavelengths[2,3] (longer than 1 mm) but it has only recently become possible to do heterodyne spectroscopy in the middle[4-8] and far[9] infrared, opening three new decades (1 μm - 1 mm) to exploration using heterodyne techniques. We examine some limitations of the heterodyne techniques in the spectral range 1 μm - 1 mm for the detection of atomic and molecular lines and of continuum radiation.

We begin by recognizing that for gases characterized by an excitation temperature T, the maximum observable spectral line intensity will approach the specific intensity of the Planck function for that wave length and temperature under conditions of thermal equilibrium. We assume that the intensity of any continuum background is small compared with that of the line spectrum, and that the lines are characterized by a Doppler profile. We then determine the minimum detectable line source brightness for an ideal heterodyne spectrometer having a bandwidth equal to the Doppler width ($\Delta\nu_D$) and we compare this with the black-body source radiance at various temperatures. Various system limitations (quantum efficiency, etc.) increase the minimum detectable source brightness above the ideal case and so are introduced as degradation factors ($\Delta_i \geq 1$), which exercise a cumulative effect as a product function $\pi_i(\Delta_i)$. Some of the Δ_i can be evaluated in simple fashion (chopping, source polarization, quantum efficiency, etc.) but others are treated in more detail. The degradation factor caused by insufficient local oscillator power is calculated as a function of the transmission coefficient of the beam splitter, over a range of parameters characterizing the photodiode mixers presently available near 10 μm, and for various local oscillator powers. The degradation factor due to dividing the available source line power into elements of bandwidth B is evaluated as a function of $B/\Delta\nu_D$. Curves indicating the optimum bandwidth for a given Doppler broadened line are given as a function of the wavelength and of the temperature to mass ratios. Finally, conclusions are drawn regarding the best $\pi_i(\Delta_i)$ one could hope to achieve in a present day heterodyne system near 10 μm, and the usefulness of the technique is demonstrated for several classes of astronomical sources.

2. Sensitivity Limits of an Ideal Heterodyne Receiver

In astronomical heterodyne detection, infrared radiation from the source is mixed with the radiation from a much stronger local oscillator and a signal is detected at the difference frequency called the intermediate frequency (IF) (Fig. 1). The mixing process translates the frequency scale by an amount equal to the local oscillator frequency, preserving the spectral characteristics of the source[10]. When the local oscillator power (P^{LO}) is sufficiently large, the shot noise from it dominates all other sources of noise in the detector and the minimum detectable IF power corresponds to the quantum detection limit[1],

$$P_{min} = h\nu\Delta\nu, \tag{1}$$

where a detector with a quantum efficiency $\eta = 1$ is assumed. This limit corresponds to the detection of a single photon per resolution time $\Delta\nu^{-1}$ (sec) of the system. A simple derivation of this limit can be given on the basis of the particle nature of light.[11] The signal-to-noise ratio (S/N) obtained at the IF for a source power P^S is then

$$\frac{S}{N} = \frac{P^S}{h\nu\Delta\nu} = \frac{P^S}{h\nu B} \tag{2}$$

where $B = \Delta\nu$ is the bandwidth of a single resolving element of the system. If the signal is integrated for some time (τ), the signal-to-noise ratio is improved by a factor $(B_\tau)^{\frac{1}{2}}$ and the post-integration signal-to-noise ratio is

$$\frac{S}{N} = \frac{P^S}{h\nu}\left(\frac{\tau}{B}\right)^{\frac{1}{2}} \tag{3}$$

The corresponding ideal minimum detectable power becomes

$$P^{s}_{min} = h\nu\left(\frac{B}{\tau}\right)^{1/2} \text{ watts}, \tag{4}$$

and the minimum detectable number of signal photons becomes

$$N^{s}_{min} = \left(\frac{B}{\tau}\right)^{1/2} \text{ (photons sec}^{-1}\text{)} \tag{5}$$

in a particular polarization and in a single sideband. The minimum detectable source surface brightness R^{s}_{min} (photons cm^{-2} sec^{-1} str^{-1}) is obtained from

$$P^{s}_{min} = R^{s}_{min}\, h\nu\, A\,\Omega. \tag{6}$$

The etendu, $A\,\Omega$ (with A as the detector area and Ω the field of view), is a constant of the system and is $\sim\lambda^2$ [12]. The minimum detectable source surface brightness from (4) and (6) is

$$R^{s}_{min} = \frac{1}{\lambda^2}\left(\frac{B}{\tau}\right)^{1/2} \text{ (photons cm}^{-2} \text{ sec}^{-1} \text{ str}^{-1}\text{)} \tag{7}$$

Note that Eq. (4) is referred to the detector and that Eq. (7) is referred to the remote source by assuming that the system has been properly matched to the diffraction limit of the collecting optics (telescope) such that $A\,\Omega \sim \lambda^2$ is preserved. In this case, the heterodyne field-of-view is $\theta \sim 1.2\ \lambda/D$ where D is the diameter of the primary mirror. A commonly used convention among infrared astronomers is to refer the brightness to the power incident per unit telescope area per unit frequency interval (watts m^{-2} Hz^{-1}).

One flux unit (or one Jansky) equals 10^{-26} watts m^{-2} Hz^{-1}. The minimum detectable flux over a source bandwidth $\Delta\nu_s$ is

$$F^s_{min} = \frac{P^s_{min}}{A\Delta\nu_s} = \frac{h\nu}{A\Delta\nu_s}\left(\frac{B}{\tau}\right)^{\frac{1}{2}} \tag{8}$$

where A is the telescope area in m^2.

It is not very useful to calculate P^s_{min}, R^s_{min}, and F^s_{min} for fixed bandwidth over three decades of wavelengths since linewidths can be expected to vary drastically over this wavelength range. For our purposes, we set the bandwidth equal to the Doppler line width at each λ and calculate P^s_{min}, R^s_{min}, F^s_{min} as a function of wavelength for various temperature-to-mass ratios (T/M). A telescope collecting area of 1 m^2 is assumed in calculating F^s_{min}. The Doppler bandwidth of an atomic or molecular line is given by

$$B = \Delta\nu_D = 7.16 \times 10^{-7}\,\frac{c}{\lambda}\,(T/M)^{\frac{1}{2}}\ Hz \equiv \frac{\beta}{\lambda} \tag{9}$$

where T is the temperature ($^{\circ}$K) and M is the atomic or molecular mass (amu). The expressions for minimum detectable source power (P^s_{min}), source radiance (R^s_{min}), and flux (F^s_{min}) become

$$P^s_{min} = \frac{hc}{\lambda^{3/2}}\left(\frac{\beta}{\tau}\right)^{\frac{1}{2}}\ \text{watts} \tag{10}$$

$$R^s_{min} = \frac{1}{\lambda^{5/2}}\left(\frac{\beta}{\tau}\right)^{\frac{1}{2}}\ \text{photons cm}^{-2}\text{sec}^{-1}\text{str}^{-1} \tag{11}$$

and

$$F^s_{min} = \frac{hc}{\lambda^{1/2}}\,(\beta\tau)^{-\frac{1}{2}}\ Wm^{-2}\ Hz^{-1} \tag{12}$$

R^s_{min} is compared with the single side-band radiance of a black-body continuum source in Fig. 2. From the Planck function, the power radiated

in bandwidth B is

$$W^s_{BB} = \frac{2hc}{\lambda^3} \quad \frac{B}{(e^{h\nu/kT} - 1)} \quad (W\ cm^{-2} str^{-1}) \tag{13}$$

and

$$R^s_{BB} = \frac{2}{\lambda^3} \frac{B}{(e^{h\nu/kT} - 1)} \tag{14}$$

The curves shown in Fig. 2 have been calculated for a particle with M = 30 amu. R^s_{min} is shown for T = 300°K and 50°K and for integration times of 1, 10^2, and 10^4 seconds. At 10 μm and 300°K, the Doppler linewidth is ~ 68 MHz. The rapid improvement in R^s_{min} with increasing wavelength is a consequence of the decreasing line width (i.e., less LO shot noise in B = $\Delta\nu_D$) and of the increasing field-of-view (as λ^2). However, the improvement due to increasing field-of-view (FOV) is replaced by a degradation factor when the FOV exceeds the source size (see Section 3). The FOV for a diffraction limited 1 m telescope is also shown in Fig. 2.

A similar set of curves for the minimum detectable flux F^s_{min} and for the black-body flux

$$F^s_{BB} = \frac{2hc}{\lambda} \quad \frac{1}{(e^{h\nu/kT} - 1)} \quad W\ m^{-2}\ Hz^{-1} \tag{15}$$

are shown in Fig. 3. In the Rayleigh-Jeans limit ($h\nu \ll kT$), the black-body flux becomes $F^s_{BB} = 2\ kT$, and the received power from both sidebands of width B is 4 kTB watts, the familiar radio result.

It is clear from Figs. 2 and 3 that heterodyne detection is potentially a very powerful tool for high resolution astronomical studies. Section 6 shows that even after the system performance is degraded to take account of practical limitations, heterodyne detection remains a very powerful technique.

3. Limitations to the Sensitivity of a Real Heterodyne System

The expressions for minimum detectable power and source brightness considered in section 2 are based on the assumption that all source photons radiated into the heterodyne field-of-view are collected by a shot-noise-limited photomixer with unity quantum efficiency. In actual practice, however, there are a number of factors which degrade the sensitivity significantly from the ideal case. An understanding of these factors is important in achieving the highest sensitivities for the spectrometer. For astronomical application, even small gains in sensitivity are of major interest since the integration time required for a fixed minimum detectable power is reduced by the square of that gain. The various factors which must be included in evaluating the sensitivity of the heterodyne spectrometer for astronomical applications are considered here. Any factor which leads to an increase in the ideal minimum detectable source brightness R^s_{min} (Eq. 7) is written as a degradation factor, Δ. The minimum detectable source brightness can thus bewritten as

$$R^s_{min} = R^s_{ideal} \prod_I (\Delta_I) \qquad (16)$$

where the various degradation factors Δ_I to be considered are due to:

(a) Quantum efficiency of the photomixer: $\Delta_Q = 1/\eta$: for HgCdTe photomixers, the best possible quantum efficiency without an anti-reflection coating on the chip is $\eta \sim 0.67$ ($\Delta_Q \sim 1.5$).[13] The best-commercially available detectors have $\eta \sim 0.5$ ($\Delta_Q \sim 2$). It is also important to note that η is the dynamic quantum efficiency (LO on),

not the static quantum efficiency (LO off) and that too high LO powers can cause a lowered dynamic quantum efficiency.

(b) Polarization: $\Delta_{POL} = 2$ for an unpolarized source, but may vary between $1 \rightarrow \infty$ for other sources as the linear polarization changes from parallel to perpendicular with respect to the local oscillator polarization.

(c) Chopping: $\Delta_{chop} = 2$ for a Dicke type chopper, $\sqrt{2}$ for overlapped frequency switching, and 1 for the unchopped case.

(d) Losses in Optics: $\Delta_{optics} = \frac{1}{\alpha}$ where α is the total transmission coefficient of the collecting optics up to but not including the beam splitter.

(e) Phase front misalignment and beam spot size effects: Δ_{phase}. This degradation is caused by mis-match of the phase fronts and spot sizes of the local oscillator beam and the source beam at the photomixer and has been considered by Cohen[14] in detail. It will not be discussed here except to note that proper matching of the phase-fronts and Airy disks allows one to make $\Delta_{phase} \sim 1.2$.

(f) Beam filling factor: Δ_{FF}, the degradation when the source fails to fill the heterodyne field of view (which is assumed equal to the diffraction limited field of view of the telescope).

$$\Delta_{FF} \simeq (A^{dif\ f}/A^{s}) \simeq 1.4\ \lambda^{2} D^{-2}\ \theta_{s}^{-2} \qquad (17)$$

where θ_s is the angular extent of the source and D is the diameter of the primary mirror.

(g) Detector and pre-amplifier noise: Δ_D is the degradation caused by shot-noise due to d.c. bias current in the detector, and the thermal noise contributed by the detector and by the IF amplifier. An evaluation of Δ_D is made in section 4 for a wide range of detector characteristics, transmission coefficients of the beam splitter, and local oscillator powers.

(h) Degradation in measurement of line profiles: Δ_L. This degradation in the sensitivity, compared with that which is obtained when the bandwidth is matched to the Doppler linewidth, is a consequence of dividing the line profile into elements of width B. A derivation and an evaluation of Δ_L as a function of linewidth to bandwidth ratios, and of wavelength is given in section 5 for various T/M ratios.

(i) Various other degradation factors may be introduced in specific experiments under adverse conditions but have not been treated here. Examples would include phase cancellation introduced by air turbulence ("seeing") and false heterodyne signals which are sometimes caused by source or reference continuum shot noise.

4. Degradation Due to Detector Noise and Pre-Amplifier Noise

The shot-noise limited performance of a heterodyne spectrometer is dependent upon the availability of wideband mixers and sufficiently powerful tunable lasers in the desired spectral range. For sufficiently large local oscillator powers, the local oscillator shot noise dominates all other noise sources, leading to the quantum limit of sensitivity given by (1).

The requirements placed on the detector, of low noise and wide bandwidth, are adequately met in the 5-17 μm range by helium cooled Ge:Cu photoconductive detectors and liquid nitrogen cooled HgCdTe photodiodes which have bandwidths of the order of 1 GHz[13,15,16]. Of the two types of detectors, the HgCdTe photodiode is preferred for the tuneable heterodyne spectrometer for astronomical observations because of the higher heterodyne signal-to-noise ratio which may be obtained, (a factor of two better than photo-conductors) and higher operating temperatures. Also, the local oscillator power required to reach the shot-noise-limit with HgCdTe is smaller than for Ge:Cu photoconductors and can be supplied by tuneable semi-conductor lasers.

Sufficiently powerful lasers tunable over many wave numbers in the spectral range of interest, are not readily available at present. Adequate emission powers are available from gas lasers (e.g. CO_2,CO) at specific wavelengths but the tunability is limited to a narrow band around the discrete laser lines. Waveguide CO_2 lasers can be tuned over $\sim$1 GHz for each line. Electronic tuning at the IF

extends the spectral bandwidth to ~ 2 GHz on either side of the discrete local oscillator frequencies. The versatility of heterodyne spectroscopy depends on the availability of tunable local oscillators, and cannot be fully realized with the available gas lasers. Although continuously tunable lasers (over many wave numbers) are not available as yet, semitunable diode lasers have recently become available.

The first use of cryogenically cooled PbSe diode lasers as local oscillators for heterodyne detection was achieved with the infrared heterodyne spectrometer developed at the Goddard Space Flight Center[4]. Astronomical and laboratory sources (both continuum and molecular lines) were detected at 25 MHz resolution at 8.5 μm. Although it has been demonstrated that diode lasers can be used as local oscillators, the emission powers available are still limited to a few hundred μW in a single mode, and are generally insufficient for shot noise limited operation of the heterodyne spectrometer. The additional sources of noise in the detector thus have to be considered. The degradation in the sensitivity of the heterodyne spectrometer due to all additional sources of noise in the photodiode (HgCdTe) is considered in this section.

The equivalent circuit of a photodiode mixer consists of a shunt conductance G_d, a shunt capacitance C_d and a series resistance R_s. The output of the photodiode contains noise current components corresponding to: (i) shot noise due to local oscillator and background-induced DC photocurrents I_o and I_b and dark current I_d (ii) thermal

noise in the detector at its physical temperature T_m and (iii) IF amplifier noise given by an equivalent input noise temperature T_{IF} which includes the effect of the impedance mis-match between the photodiode and the IF amplifier. The signal to noise ratio of the heterodyne receiver (post-integration) is then given by[16]

$$\frac{S}{N} = \frac{(\eta e/h\nu)\ P^{s} I_o\ (B\tau)^{1/2}}{(I_o+I_d')eB+2k(T_m+T_{IF}')BG_{deq}}\ , \tag{18}$$

where

$$I_d' = I_b + I_d$$

and

$$G_{deq} = [G_d(1+R_sG_d) + \omega^2 R_s C_d^{\,2}] \simeq G_d(1+\omega^2/\omega_c^{\,2}) \tag{19}$$

for $R_sG_d \ll 1$; the frequency at which the available output power for the case of a conjugate match to the IF amplifier is down by 3 dB is defined by $\omega_c \simeq 1/C_d(R_s/G_d)^{1/2}$. The d.c. photocurrent is related to the local oscillator power by

$$I_o = \eta e P_{LO}/h\nu \tag{20}$$

It should be noted that the detector quantum efficiency η shown explicitly in (18) has been considered as a degradation factor separately in section 3. However, η also affects the S/N ratio through I_o as shown in (20) and this dependence will be included in Δ_D.

The two terms in the denominator in (18) correspond to shot noise and thermal powers respectively. With sufficient laser power, all other terms in the denominator are negligible compared with the local oscillator shot noise term ($I_o e\mathcal{B}$), and the S/N ratio reduces to the ideal value (3) for $\eta = 1$. When the local oscillator power is insufficient, (18) has to be maximized by proper choice of the photodiode parameters (G_d, C_d, R_s), by impedance matching between the detector and the IF amplifier, and by an optimum division of the signal and local oscillator powers at the beam splitter.

Denoting α for the transmission coefficient of the beam splitter for the incident local oscillator power, the signal and local oscillator powers incident on the photomixer are:

$$P^{LO} = \alpha P_o^{LO} \text{ and } P^s = (1 - \alpha)\, P_o^s \qquad (21)$$

Introducing (21) into (18) and comparing it with the shot-noise limit $S/N = (\eta P_o^s / h\nu\, (\tau/B))^{1/2}$, the degradation in the S/N ratio or in the minimum detectable source brightness is found to be:

$$\Delta_D = \frac{(1+\epsilon/\alpha)\ (\eta e^2 \alpha P_o^{Lo}) + (2kT_{eff} G_{deq}) h\nu}{\alpha(1-\alpha)\ (\eta e^2 P_o^{Lo})} \qquad (22)$$

where $\epsilon = I'_d / I_{LO}$, the ratio of dark current to the local oscillator photocurrent in the mixer is $I'_d / I_o = I'_d / \alpha I_{LO} = \epsilon/\alpha$ and $T_{eff} = (T_m + T'_{IF})$. The second term in the numerator corresponds to the degradation introduced by thermal noise and the factor ϵ/α accounts

for the degradation due to the shot noise generated by the dark current. For sufficiently large local oscillator power for a given value of α, the degradation factor becomes

$$\Delta_D = \frac{(1+\epsilon/\alpha)}{(1-\alpha)} \tag{23}$$

When the photodiode dark current is also small compared with the local oscillator photocurrent so that $\epsilon \rightarrow o$, the degradation factor approahces unity for sufficiently small values of α. For smaller local oscillator powers, the degradation factor and the optimum value of α for minimum degradation may be estimated from the plots of Δ_D vs α for various values of the parameter $T_{eff}\ G_{deq}$, the local oscillator power P_o^{LO} and ϵ.

The degradation factors in Figs. (4-6) have been given as a function of the parameter

$$T_{eff}\ G_{feq} = (T_m + T'_{IF})\ (1 + \frac{\omega^2}{\omega_c^2})\ G_d \tag{24}$$

For the liquid nitrogen cooled HgCdTe photodiode considered here $T_m \sim 80^oK$. The quantity T'_{IF} is the equivalent input noise temperature of the IF pre-amplifier and includes the effect of impedance mismatch between the photodiode and the amplifier. When the two impedances are matched $T'_{IF} = T_{IF} = T_o(F-1)$, where T_o is the room temperature and F is the noise figure of the amplifier. For the unmatched case T'_{IF} may be calculated from

$$T'_{IF} = T_{IF}\ (\frac{R_{IF}}{4R_o} + \frac{R_o}{4R_{IF}} + \frac{1}{2}) \tag{25}$$

where R_{IF} is the input resistance of the IF amplifier and R_o is the resistive component of the output impedance of the photodiode. For $R_o \gg R_{IF}$, T'_{IF} is approximated by $T'_{IF} = (R_o/4R_{IF})\ T_{IF}$. A plot of the variation of the parameter $(T_{eff}\ G_{deq})$ with ω/ω_c is shown in Fig. 7 for various values of the photodiode shunt conductance G_d, and assumed values of $R_{IF} = 50\,\Omega$, $R_s = 20\,\Omega$, $C_D = 2$ pF and $\lambda = 10\,\mu$m. Lower values of $T_{eff}\ G_{deq}$ over the IF bandwidth may be obtained by employing a low noise-temperature impedance matching device.

The minimum values of the degradation factor $(\Delta_D)_{min}$, corresponding to the optimum value of the transmission coefficient of the beam splitter, are shown as a function of the parameters $T_{eff}\ G_{deq}$ for various local oscillator powers in Fig. 8. For a given value of $T_{eff}\ G_{deq}$, the degradation factor is improved significantly when the LO power is increased. This reduction in Δ_D becomes more significant when $T_{eff}\ G_{deq} > 1$. For a HgCdTe diode with $G_d = 4 \times 10^{-4}$ and $T_{eff}\ G_{deq} \lesssim 1.5$ (Fig. 7), the degradation factor Δ_D is of the order of 5 for LO powers $\sim 200\,\mu$W and reduces to 2-3 for LO powers of $500\,\mu$W - 1 mW.

5. Degradation Introduced by Linewidth to Bandwidth Mismatch

The foregoing analysis of the achievable sensitivity level of a real heterodyne system has focussed on detection of spectral lines (single side-band) and has assumed that the bandwidth (B) of the smallest frequency resolving element (e.g. an RF filter on a multi-spectral line receiver) was equal to the Doppler width ($\Delta\nu_D$) of the line. For spectroscopy, one would like to operate in two modes: (1) a low resolution mode which covers a wide range of frequencies at resolutions comparable to the line width and which is used to survey and to find the lines and (2) a high resolution mode which covers the entire line profile simultaneously at a resolution much smaller than the line width and which is used to study the line profiles.

We shall limit our present treatment to the case of unsaturated Doppler broadened lines characterized by (λ, T,M) and calculate the degradation (Δ_L) in signal-to-noise as a function of $B/\Delta\nu_D$. If the total radiance, integrated under the line, were instead to be radiated uniformly in a rectangular line of width $B = \Delta\nu_D$, then Δ_L would equal unity. We will show that for a Doppler line, a minimum for Δ_L occurs when $B = 1.2\ \Delta\nu_D$ for which $\Delta_L = 1.3$. It should be noted that if the line is saturated at line center over the resolution bandwidth (B), then $\Delta_L = 1$ and the measured S/N will correspond to that of a black-body (single side-band) with temperature T(^{o}K). Also, source turbulence (e.g. in interstellar clouds) and source rotation (e.g. in planetary atmospheres) may result in non-Doppler profiles in some cases. With

these facts in mind, we proceed to calculate Δ_L for Doppler-broadened lines.

The intensity profile of a Doppler broadened line is:[17]

$$I(\nu) = I_o \, e^{-[(2(\nu-\nu_o)/\Delta\nu_D)(\ell n\ 2)^{1/2}]^2} \tag{26}$$

where the line width $\Delta\nu_D$ is given in eq. (9). The total line intensity is obtained by integrating (31) over all frequencies

$$I_{Tot} = \int_{-\infty}^{\infty} I(\nu)\ d\nu, \tag{27}$$

which may be evaluated by introducing $\omega = (2(\nu-\nu_o)/\Delta\nu_D\ (\ell n\ 2)^{\frac{1}{2}}$ and $\nu = \nu_o + \omega\Delta\nu_D\ /2(\ell n\ 2)^{\frac{1}{2}}$ so that

$$I_{Tot} = \frac{I_o\Delta\nu_D}{2(\ell n\ 2)^{1/2}} \qquad 2\int_o^{\infty} e^{-\omega^2}\ d\omega = \frac{1}{2}\left(\frac{\pi}{\ell n\ 2}\right)^{1/2} I_o\Delta\nu_D, \tag{28}$$

or

$$I_o = 2\left(\frac{\ell n\ 2}{\pi}\right)^{1/2} \frac{I_{Tot}}{\Delta\nu_D} \tag{29}$$

Considering the portion of the line transmitted through a filter with a bandwidth B centered on the line, the total number of photons received in the filter bandwidth from (26, 29) is

$$I_B = T_{Tot}\left[2\pi^{-1/2} \int_o^{\omega_B} e^{-\omega^2}\ d\omega\right] \tag{30}$$

$$= I_{Tot}\ \mathrm{erf}\ (\omega_B)$$

where $\omega_B = (2(B/2)/\Delta\nu_D)\ (\ell n\ 2)^{\frac{1}{2}} = (\ell n\ 2)^{\frac{1}{2}}\ \frac{B}{\Delta\nu_D} = 0.8325\ \frac{B}{\Delta\nu_D}$

Thus the fraction of photons received in the filter bandwidth B is

$$F_B = \frac{I_B}{I_{Tot}} = \text{erf}\ (\omega_B) \tag{31}$$

When the line profiles are measured with a heterodyne system, the resolution bandwidth is some fraction of the Doppler width

$$B = a_1\ \Delta\nu_D \tag{32}$$

The minimum detectable source brightness of a Doppler broadened line is obtained from (7) and (30)

$$\begin{aligned} R^L_{min} &= \frac{1}{\lambda^2}\ \left(\frac{B}{\tau}\right)^{1/2} \frac{1}{F_B} \\ &= \left[\frac{1}{\lambda^2}\ \left(\frac{\Delta\nu_D}{\tau}\right)^{1/2}\right] \frac{a_1^{1/2}}{\text{erf}(\omega_B)} \\ &= R^L_{Doppler}\ \Delta_L \end{aligned} \tag{33}$$

The factor $\Delta_L = a_1^{\frac{1}{2}}/\text{erf}\ (\omega_B)$, with $a_1 = B/\Delta\nu_D$ and $\omega_B = (\ell n\ 2)^{\frac{1}{2}}\ a_1$, is the degradation (at line center) in the minimum detectable number of photons in the line, which is introduced by the fact that the line is divided into elements of width B. A plot of Δ_L vs $a_1 = B/\Delta\nu_D$ is shown in Fig. 9. The asymptotes for large and small values of a_1 are given by $\Delta_L \rightarrow a_1^{\pm\frac{1}{2}}$. The 3 dB and 6 dB points (with respect to optimum) as well as the optimum point at which there is minimum degradation are shown. This minimum value ($\Delta_L = 1.3$) occurs at the bandwidth

corresponding to $a_1 = 1.2$. It may be noticed that there are two values of a_1 for each degradation factor. The optimum bandwidth, which provides sufficient spectral information with the least degradation in the minimum detectable line brightness is given by the value of a_1 about the 3 dB point on the left hand side of the curve.

The above discussion considers the optimum bandwidth for a given Doppler broadened line, assuming that B can be readily varied with λ so as to keep a_1 constant. In practice a heterodyne system has an IF resolution bandwidth which remains fixed when measuring intensity of different line widths and at different wavelengths. The degradation factor in this case may be considered more conveniently by letting the fraction of the intensity within the bandwidth B be parameterized as a function of $a_2 = \Delta\nu_{D/B}$. The quantity a_2, the normalized Doppler line width, is the inverse of a_1. The minimum detectable line brightness may thus be rewritten in terms of a_2 from (33).

$$R^L_{min} = R^L_{Doppler} \quad \Delta_L \tag{34}$$

when the degradation factor is

$$\Delta_L = a_2^{-1/2}/\mathrm{erf}(\omega_B)$$

with

$$\omega_B = (\ell n\ 2)^{1/2}/a_2$$

and

$$a_2 = \Delta\nu_{D/B} = \frac{2.15 \times 10^4}{\lambda B} \left(\frac{T}{M}\right)^{\frac{1}{2}} \tag{35}$$

a_2 is shown as a function of the wavelength λ in Fig. 10 for various values of the ratio T/M. The (T/m) ratios indicated on the upper line assume B = 25 MHz, and those on the lower line assume B = 125 MHz. The lines representing degradation of 3 dB and 6 dB with respect to optimum are also shown.

The plots of degradation factor Δ_L as a function of wavelength (1-1000 μm) applicable for various values of the parameters $(T/M)^{\frac{1}{2}}/B$ are given in Fig. 11. The values indicated on the plots correspond to a bandwidth of B = 25 MHz with the T/M ratios indicated on the curves. The curves may, however, be read for any other bandwidth with the quantity $(T/M)^{\frac{1}{2}}/B$ kept constant for each curve. The minimum for each curve corresponds to $B = 1.2\ \Delta\nu_D$, the right branch corresponds to $B > 1.2\ \Delta\nu_D$ and the left branch corresponds to $B < 1.2\ \Delta\nu_D$.

6. Astronomical Observations and Comparison With Direct Detection

We may now estimate the actual sensitivity of a typical heterodyne spectrometer for detection of line emission from astronomical sources. An estimate of the numerical values of the degradation factors is summarized in Table 1 for a typical case, where $P_o^{LO} = 500\,\mu W$, $G_d = 4 \times 10^{-4}$, B = 50 MHz and $\Delta\nu_D = 300$ MHz. The total degradation in sensitivity referred to the quantum detection limit is seen to be $\sim$ 90 for the case considered here. If sufficiently large laser power is available so that the shot noise limit is reached and $\Delta_D \sim 1$, the total degradation is $\sim$35. The minimum detectable source fluxes discussed in section 2 (Figs. 2 and 3) are thus to be multiplied by factors $\sim$35 - 90 depending upon the local oscillator power available.

With the estimated degradation factor given here, how does the sensitivity of a heterodyne receiver compare with that direct detection?

The minimum detectable flux values shown in Fig. 3 seem to indicate a poor sensitivity for the heterodyne receiver when compared with the values achieved in the direct detection modes in the infrared and optical frequency ranges. An ideal heterodyne F_{min}^{s} of $\sim 5 \times 10^{-24}$ Wm^{-2} Hz^{-1} at $\lambda = 10$ μm with $\tau = 100S$ (Fig. 3), for example, compares with actually achieved direct detection values of $\sim 10^{-24}$ -10^{-25} Wm^{-2} Hz^{-1} at far infrared (20μ - 1 mm), and $\sim 10^{-27}$ -10^{-28} Wm^{-2} Hz^{-1} at optical wavelengths. Does this mean that the heterodyne detection mode is less sensitive than direct detection? If not, then how are the sensitivities in the two modes of detection to be compared?

The difference in the P^S_{min} values, as shown in the following, has its origin in the entirely different optical bandwidths over which measurements are made in the two cases. To compare the sensitivies, it is more meaningful to compare the S/N ratios obtained when the optical bandwidths are identical. For heterodyne detection we have from (3)

$$\left(\frac{S}{N}\right)_{Het} = \frac{P^S}{\Delta_{Het} h\nu} \left(\frac{T}{B}\right)^{\frac{1}{2}} \tag{36}$$

where Δ_{Het} represents the total degradation factor in the sensitivity. For direct detection, the post integration S/N ratio is:

$$\left(\frac{S}{N}\right)_d = \frac{P^S}{\Delta_d (NEP)_d} \tau^{\frac{1}{2}} \tag{37}$$

where $(NEP)_d$ is the noise equivalent power in $WH_z^{-1/2}$ of the detector and Δ_d is the total degradation factor in the sensitivity of the direct detection system ($\sim$10 consisting of a telescope, dispersing element, and detector). The ratio of the two S/N ratios is a figure of merit with which to compare the sensitivities of the two modes, i.e.

$$F = \frac{(S/N)_{Het}}{(S/N)_d} = NEP' \frac{\lambda}{hc\ B^{1/2}} \tag{38}$$

where $NEP' = (\Delta_d/\Delta_{Het})\ (NEP)_d$. Residual photon noise, after cancellation of sky and telescope background emission, will eventually limit the achievable system NEP in either case.

A plot of F as a function of the bandwidth B is given in Figure 12 for $\lambda = 10\,\mu m$ for various values of NEP'. The heterodyne receiver thus has an advantage over direct detection in the upper part of the figure above the line F = 1 for the entire bandwidth range shown (neglecting residual photon noise). The advantage is large when NEP' is large and becomes a disadvantage, over the range of bandwidth shown, only when NEP' $\sim 10^{-16}$. A plot of F vs. λ for a fixed bandwidth B = 25 MHz is shown in Fig. 13 for various values of NEP'. A heterodyne receiver with a bandwidth B = 25 MHz thus has an advantage over direct detection (which increases with λ) for all wavelengths above 1 μm when NEP' $\geq 10^{-15}$, and above 10 μm when NEP' $\geq 10^{-16}$. In this comparison, it is assumed that the $A\Omega$ product is the same in both cases. Also additional factors which would favor the direct detection mode and lower the figure of merit have been ignored (e.g. atmospheric turbulence effects may limit the sensitivity of a heterodyne receiver for ground based observations).

With the estimates which have been given here for the practical detection limit of a heterodyne receiver what are the S/N ratios with which infrared line emission from astronomical sources of interest may be detected?

As an example, we consider infrared emission from planetary nebulae (e.g. NGC-7027, NGC-6572, NGC-7009). Fine structure line emission from SIV (10.5 μ), NeII(12.8 μ), ArIII(8.9 μ) and ClIV(11.5 μ), has been detected[18,19] using relatively broadband instruments with spectral resolution $\sim$0.3 - 0.5 cm^{-1}.

Table 2 shows the observed flux from NGC-7027 measured within a 6' beam for the four lines[18]. The flux received by the heterodyne spectrometer F^s - Wm^{-2} Hz^{-1} is based on an assumed Doppler linewidth $\Delta\nu_D$ = 300 MHz and a telescope area of 1 m^2. The minimum detectable flux F^s_{min} is calculated from (13) assuming a total degradation in sensitivity Δ_{Het} = 90, and an integration time of 1/2 hr. The S/N ratios shown are for the 50 MHz interval centered on the respective lines. Using a multi-channel line receiver (filter-bank), it is possible to measure the entire line profile simultaneously at 50 MHz resolution. Not only are these four lines easily detectable, but their line shapes should be measurable using the heterodyne approach.

Acknowledgement

We thank Mr. Gary Burgess for the assistance in carrying out the numerical computations presented in this paper, and Dr. Keith Ogilvie for reviewing the manuscript and useful comments.

REFERENCES

1. Teich, M.C., in Infrared Detectors, Vol. 5 of Semiconductors and Semimetals, 361, Academic Press (1971).

2. Buhl, D., Nature 234, No. 5328, 332 (1971).

3. Gordon, M.A. and L.E. Snyder, Ed., Molecules in the Galactic Environment, Wiley-Interscience (1973).

4. Mumma, M., T. Kostiuk, S. Cohen, D. Buhl, and P.C. von Thuna, Nature, 253, 514 (1975).

5. Peterson, D.W., M.A. Johnson, and A.L. Betz, Nature, 250, 128 (1974).

6. DeBatz, D., P. Granes, J. Gay, and A. Journet, Nature Phys. Sci. 245, 89 (1973).

7. De Graaw, Th., and H. van de Stadt, Nature Phys. Sci. 246, 73 (1973).

8. Nieuwenhuijzen, H., Mon. Not. R. Astro. Soc. 150, 325 (1970).

9. De Graaw, Th., private communication (1975).

10. Cummings, H.Z. and H.L. Swinney, 134, in Progress in Optics, Vol. VIII (1970).

11. Yariv, A., Introduction to Optical Electronics, Holt, Rinehart and Winston, Inc. (1971).

12. Siegman, A.E., Proc. IEEE, 54, 10, 1350 (1966).

13. Spears, D.L., T.C. Harman and I. Melngailis, Proc. IRIS Specialty Group Meeting (1973).

14. Cohen, S.C., Preprint (1975).

15. Melchior, H., M.B. Fisher, and F.R. Arams, Proc. IEEE, 58, 1466 (1970).

16. Peyton, B.J., A.J. DiNardo, G.M. Kanischak, F.R. Arams, R.A. Lange, and E.W. Sard, IEEE J. Quantum Electron, QE. 8, 252 (1972).

17. Mitchell, A.G. and M.W. Zemansky, Resonance Radiation and Excited Atoms, Cambridge Univ. Press (1961).

18. Holtz, J.Z., T.R. Geballe and D.M. Rank, Ap.J. (Letters), 164: L29, 1971.

19. Gillett, F.C., W.J. Forrest and K.M. Merrill, Ap. J., 183, 87 (1973).

Table 1. Typical values of Δ_i's for

$P_o^{LO} \sim 500\ \mu W$ in a single mode, $G_d = 4 \times 10^{-4}$

$B = 50$ MHz and $\Delta\nu_D = 300$ MHz

Δ_i	Typical Value
$\Delta_Q = 1/\eta$	2
Δ_{Pol}	2 for unpolarized source
Δ_{chop}	2 for Dicke type chopping
$\Delta_{optics} = 1/\alpha$	1.2
Δ_{phase}	1.2
Δ_{FF}	≥ 1
Δ_D	3 1^*
Δ_L	2.6
Δ_{TOTAL} ($P_o^{LO} \sim 500\ \mu W$)	90 35^*

*$P_o^{LO} \sim$ a few milliwatts, i.e. the shot-noise limit.

LIST OF FIGURES

Fig. 8. The minimum degradation factor Δ_D (corresponding to optimum values of α) vs $T_{eff}\, G_{deq}$ for the indicated LO powers. The d.c bias current is assumed to be $I_d = 0.2$ ma.

Fig. 9. The line degradation factor Δ_L vs a_1. The 3 dB and 6 dB points refer to the optimum value of $\Delta_L = 1.3$.

Fig. 10. Plots of a_2 vs λ for the indicated T/M ratios for assumed bandwidths of B = 25 MHz and B = 125 MHz.

Fig. 11. The line degradation factor Δ_L vs λ for the indicated T/M ratio for an assumed bandwidth of 25 MHz.

Fig. 12. Plot of F vs B for various values of $NEP' = (\Delta_d/\Delta_{Het})(NEP)_d$.

Fig. 13. Plot of F vs λ for various values of $NEP' = (\Delta_d/\Delta_{Het})(NEP)_d$.

BASIC HETERODYNE SYSTEM

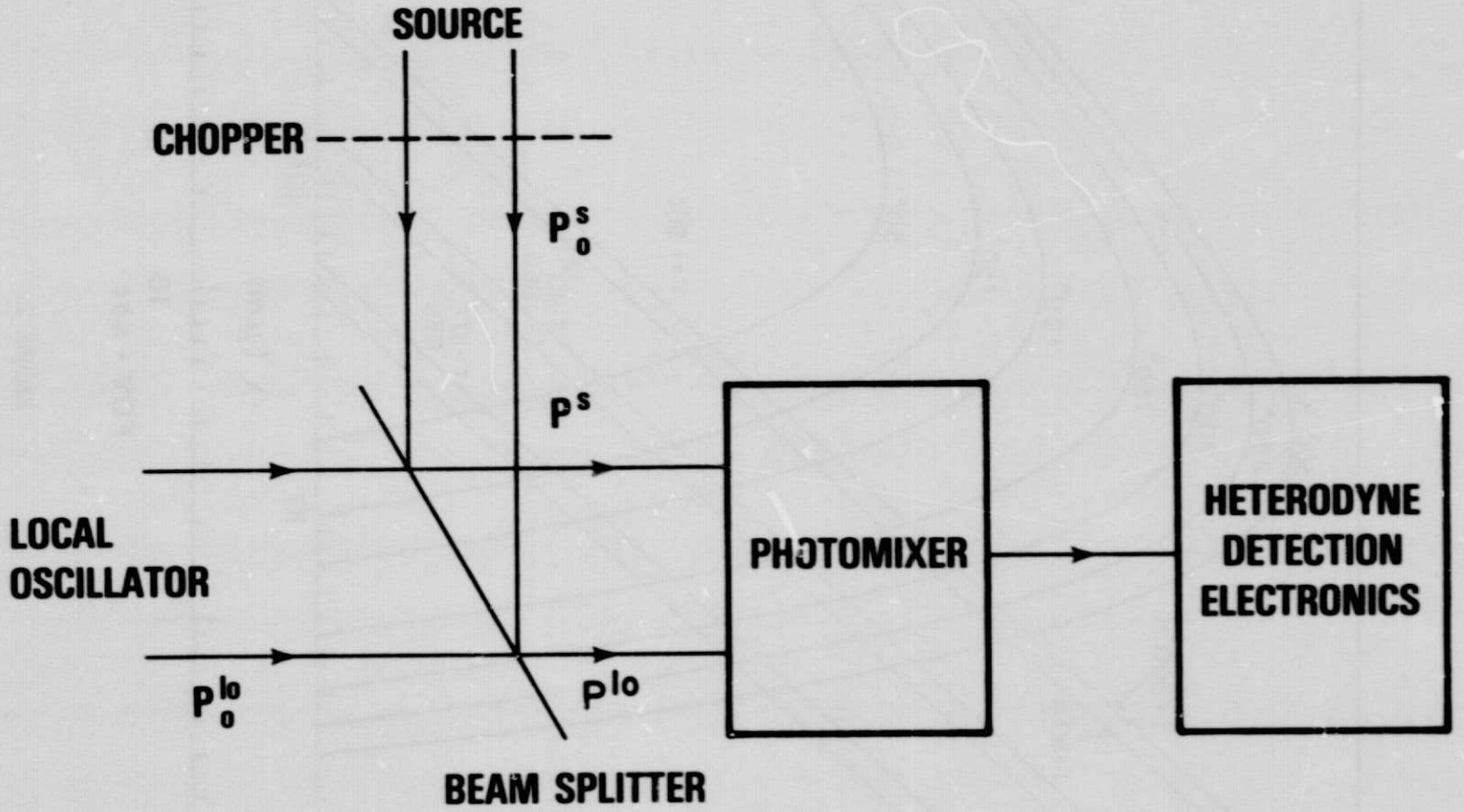

FIGURE 1

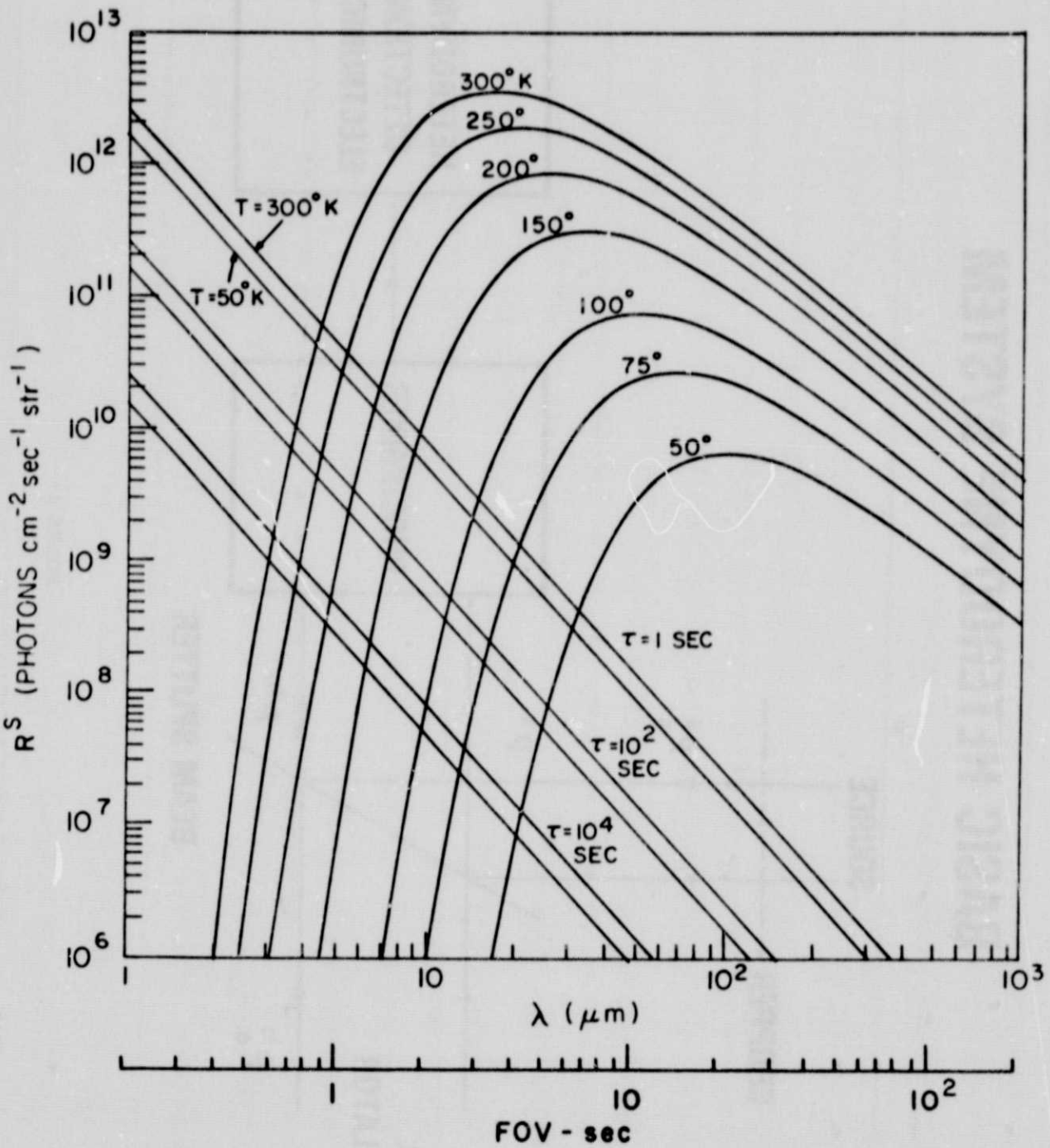
10^{13}
10^{12}
10^{11}
10^{10}
10^{9}
10^{8}
10^{7}
10^{6}
R^S (PHOTONS $cm^{-2} sec^{-1} str^{-1}$)
300°K
250°
200°
150°
100°
75°
50°
T = 300°K
T = 50°K
τ = 1 SEC
$\tau = 10^2$ SEC
$\tau = 10^4$ SEC
1
10
10^2
10^3
λ (μm)
1
10
10^2
FOV - sec

FIGURE 2

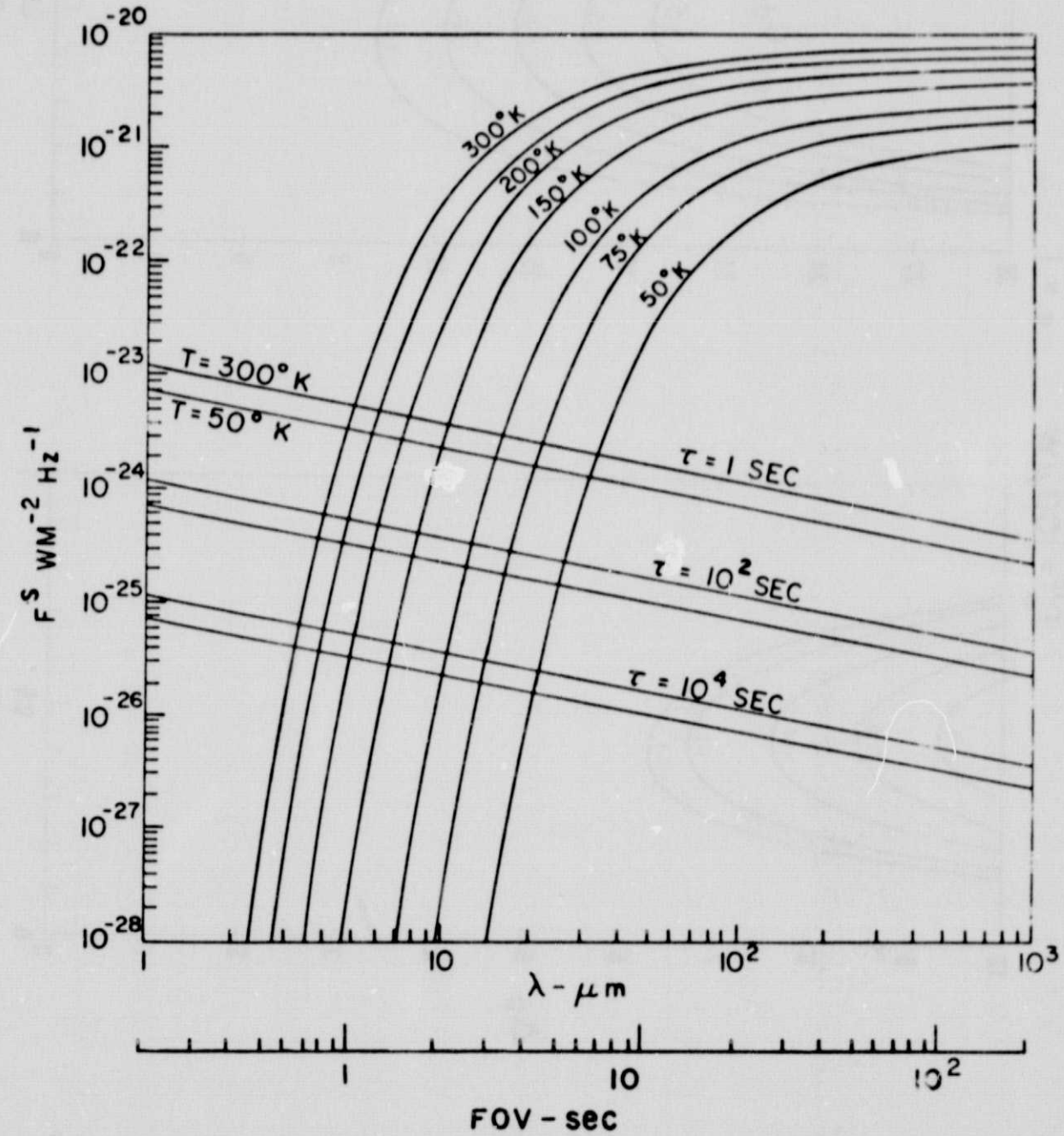
10^{-20}
10^{-21}
10^{-22}
10^{-23}
10^{-24}
10^{-25}
10^{-26}
10^{-27}
10^{-28}
F^S WM^{-2} Hz^{-1}
300°K
200°K
150°K
100°K
75°K
50°K
T = 300° K
T = 50° K
τ = 1 SEC
τ = 10^2 SEC
τ = 10^4 SEC
1
10
10^2
10^3
λ - μm
1
10
10^2
FOV - sec

FIGURE 3

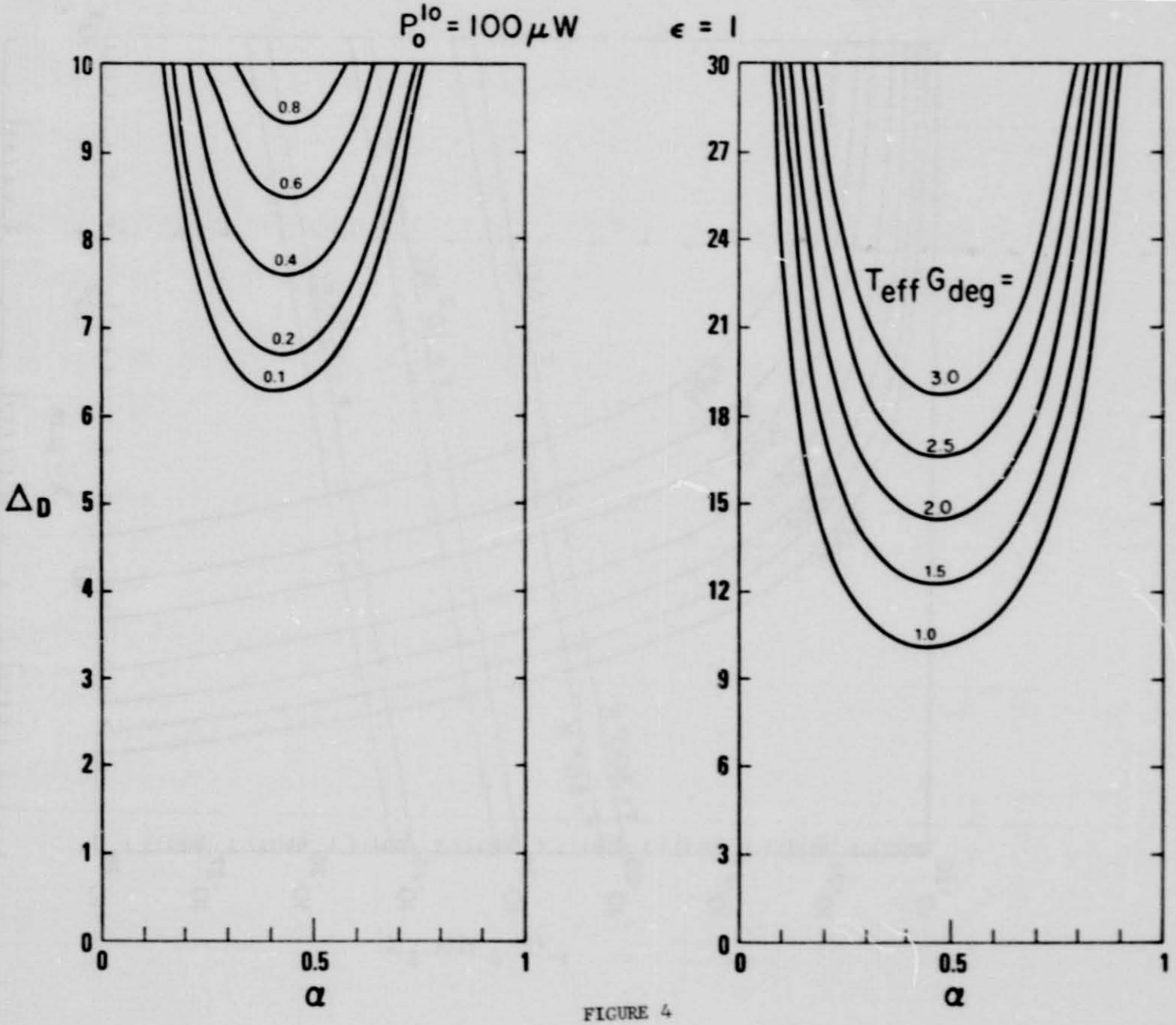
$P_o^{lo} = 100\,\mu W$
$\epsilon = 1$
Δ_D
α
0.8
0.6
0.4
0.2
0.1
$T_{eff}\ G_{deg} =$
3.0
2.5
2.0
1.5
1.0

FIGURE 4

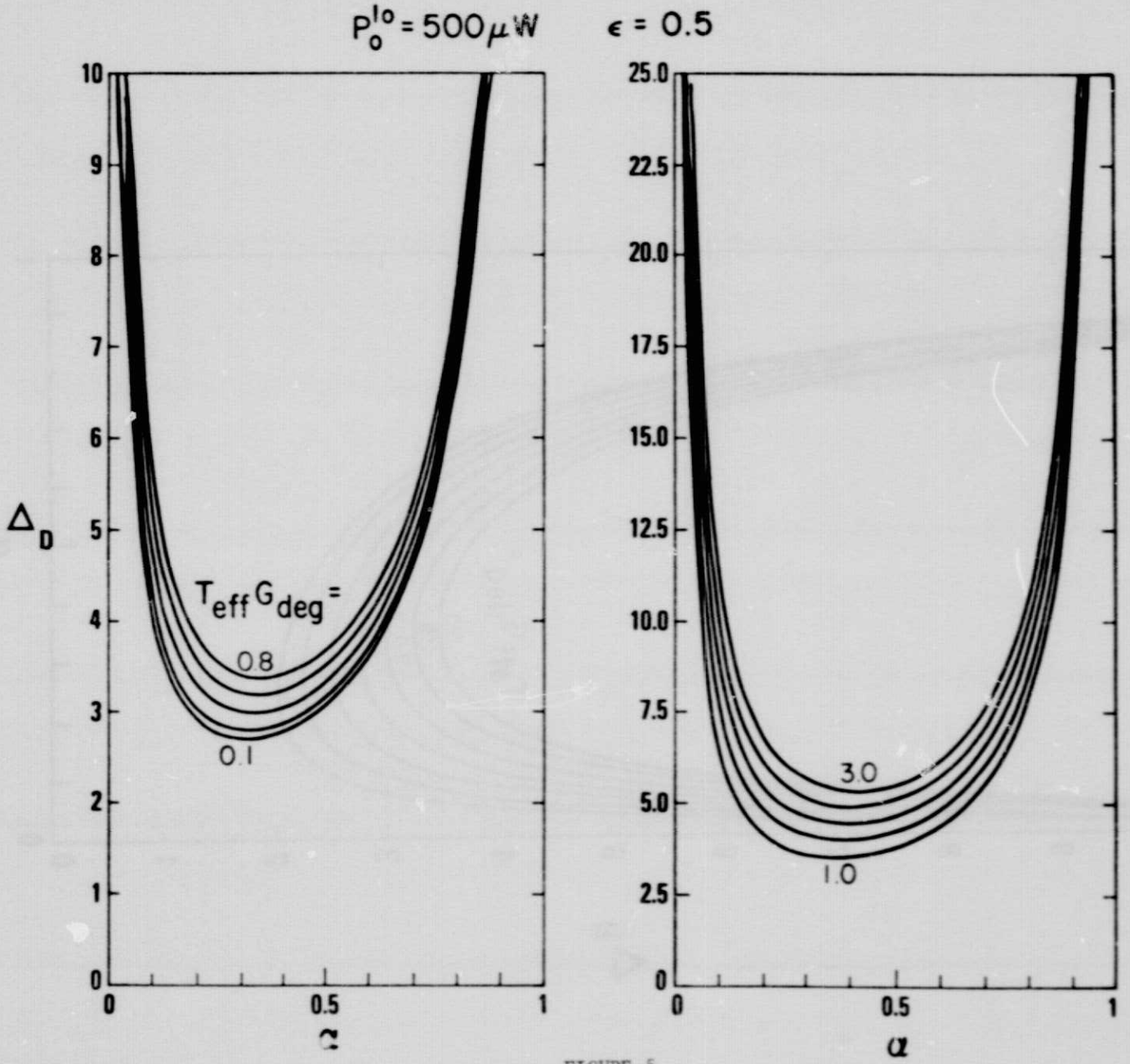
$P_o^{lo} = 500\mu W$
$\epsilon = 0.5$
Δ_D
$T_{eff}\,G_{deg} =$
0.8
0.1
3.0
1.0
α
α

FIGURE 5

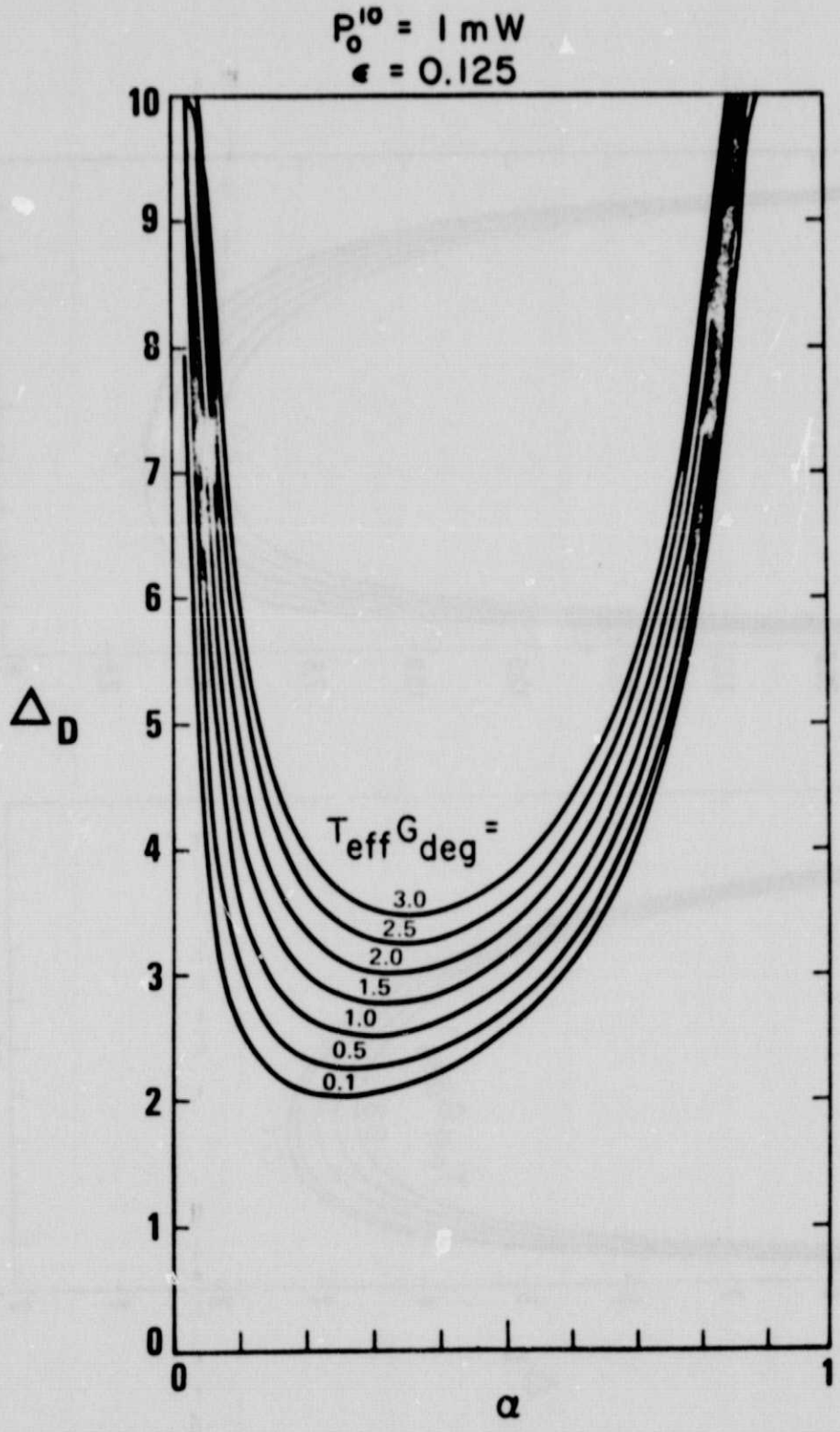
P_0^{10} = 1 mW
ϵ = 0.125
Δ_D
T_{eff} G_{deg} =
3.0
2.5
2.0
1.5
1.0
0.5
0.1
0
1
2
3
4
5
6
7
8
9
10
0
1
α

FIGURE 6

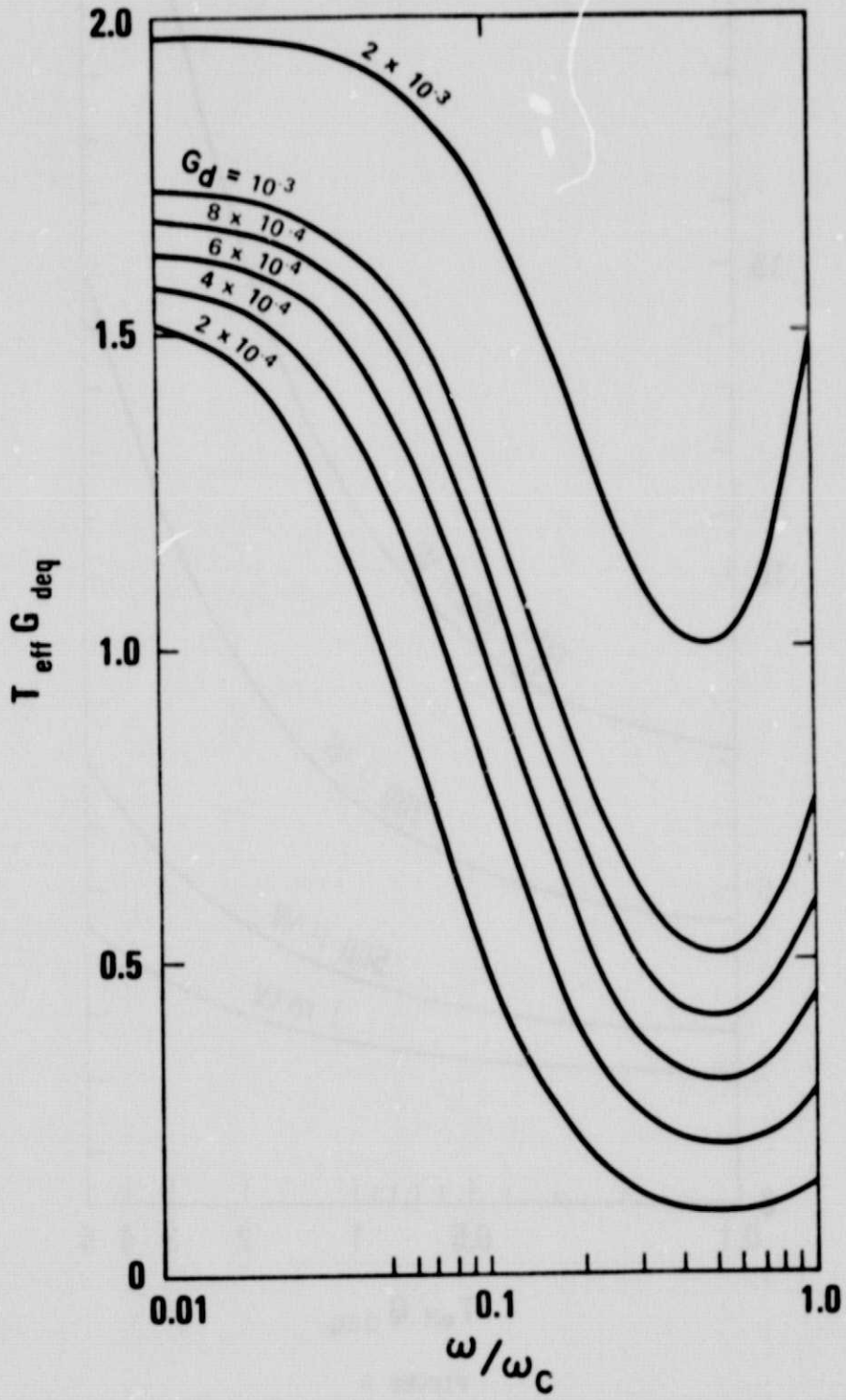
2.0
1.5
1.0
0.5
0
0.01
0.1
1.0
$T_{eff} G_{deq}$
ω/ω_c
$G_d = 10^{-3}$
2×10^{-3}
8×10^{-4}
6×10^{-4}
4×10^{-4}
2×10^{-4}

FIGURE 7

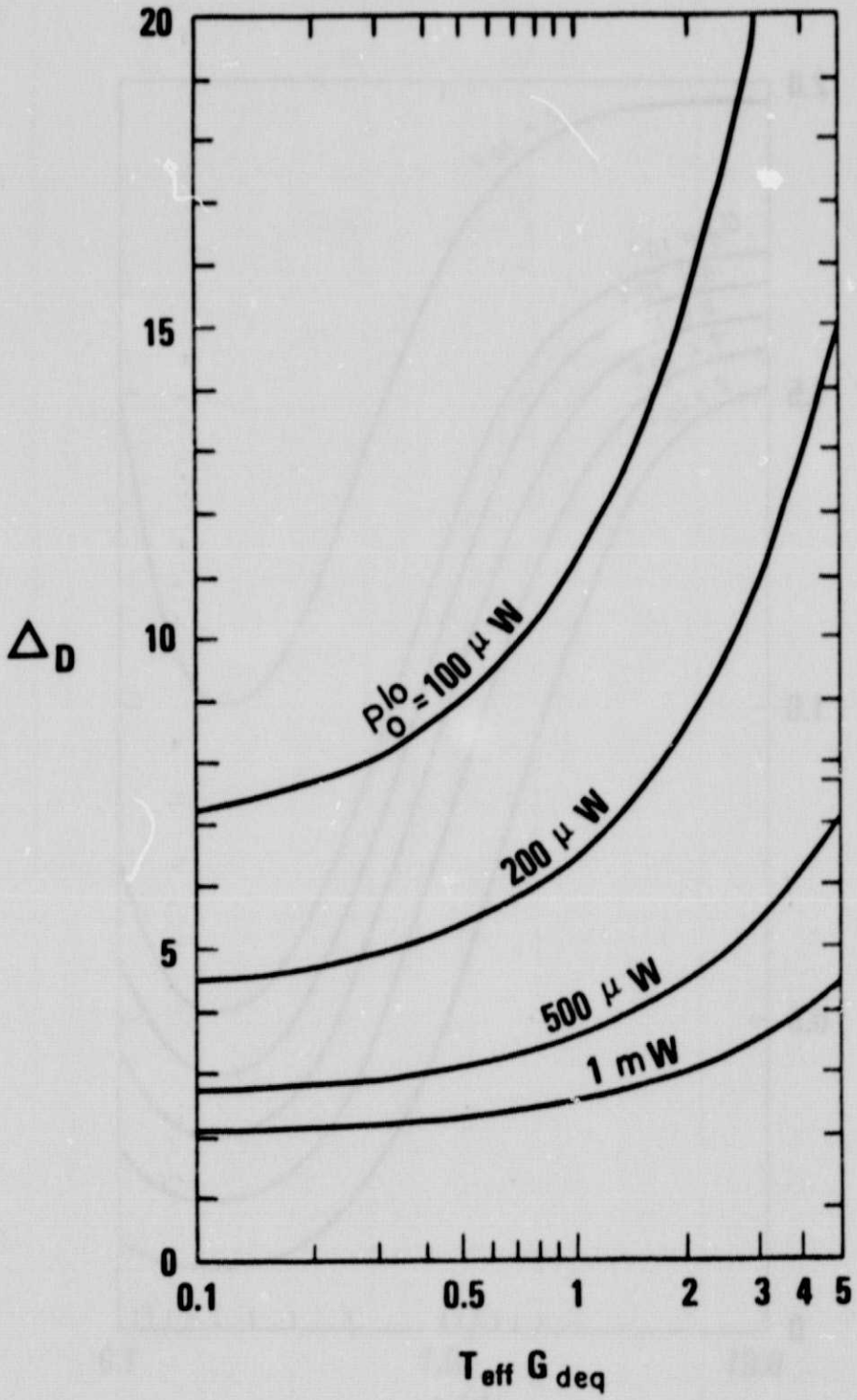
20
15
10
5
0
Δ_D
P_o^{lo} = 100 μW
200 μW
500 μW
1 mW
0.1
0.5
1
2
3
4
5
$T_{eff}\ G_{deq}$

FIGURE 8

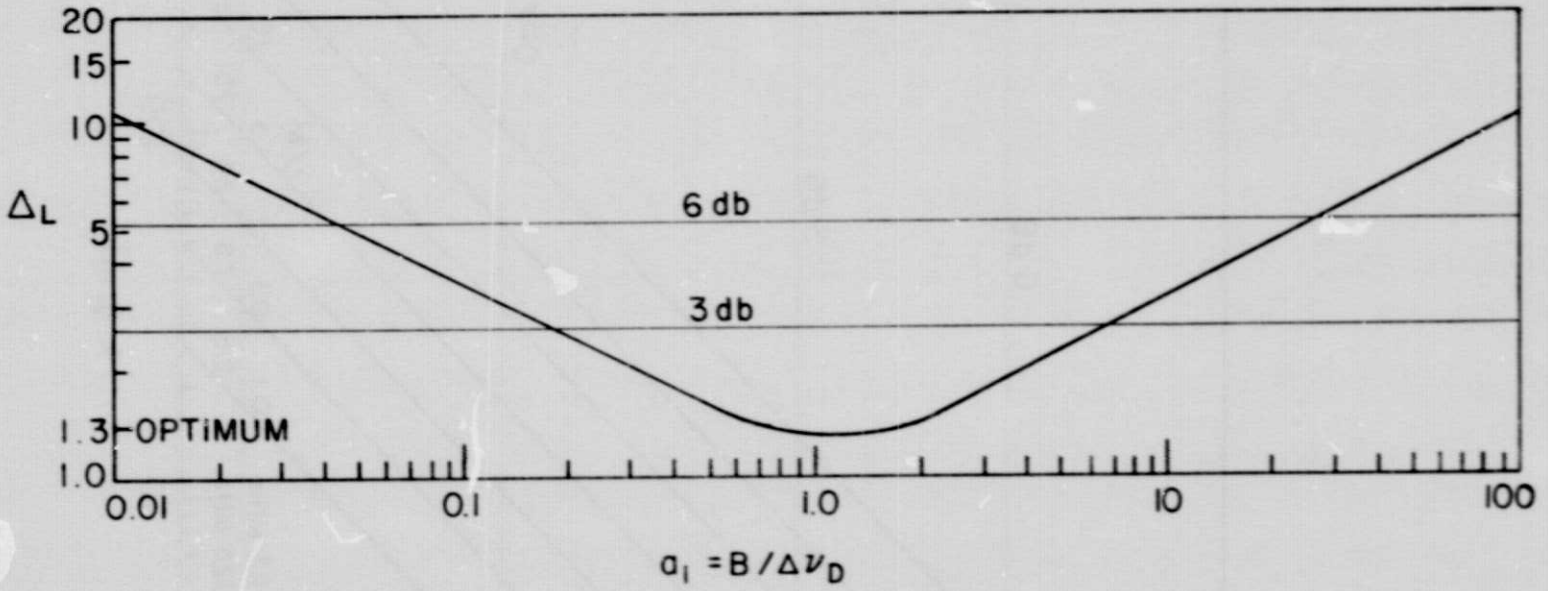
20
15
10
5
1.3
1.0
Δ_L
6 db
3 db
OPTIMUM
0.01
0.1
1.0
10
100
$a_1 = B/\Delta\nu_D$

FIGURE 9

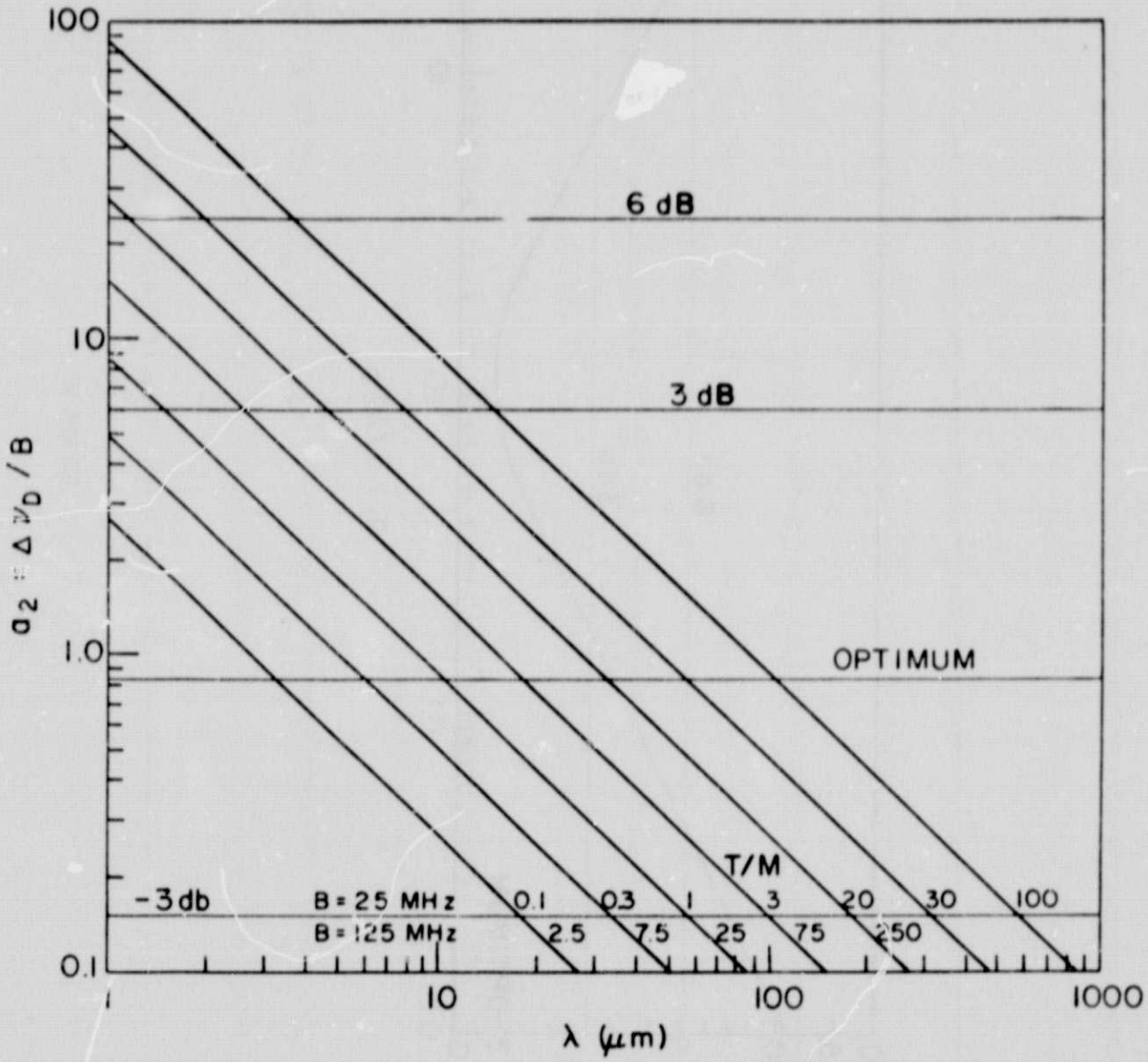
100
10
1.0
0.1
6 dB
3 dB
OPTIMUM
$a_2 = \Delta\nu_D / B$
T/M
-3db
B = 25 MHz
B = 125 MHz
0.1
0.3
1
3
20
30
100
2.5
7.5
25
75
250
1
10
100
1000
λ (μm)

FIGURE 10

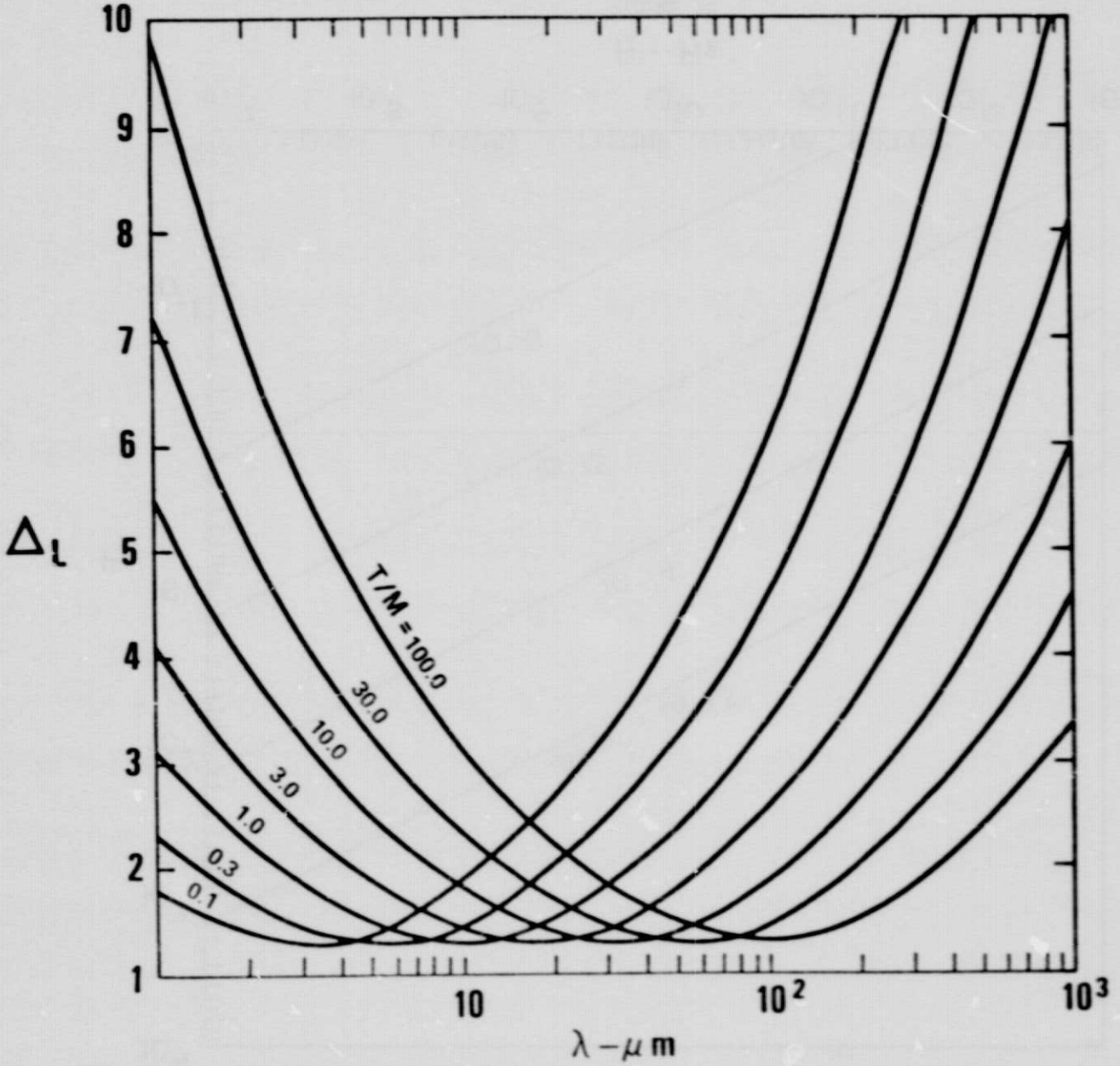
10
9
8
7
6
5
4
3
2
1
Δ_L
T/M = 100.0
30.0
10.0
3.0
1.0
0.3
0.1
10
10^2
10^3
$\lambda - \mu$ m

FIGURE 11

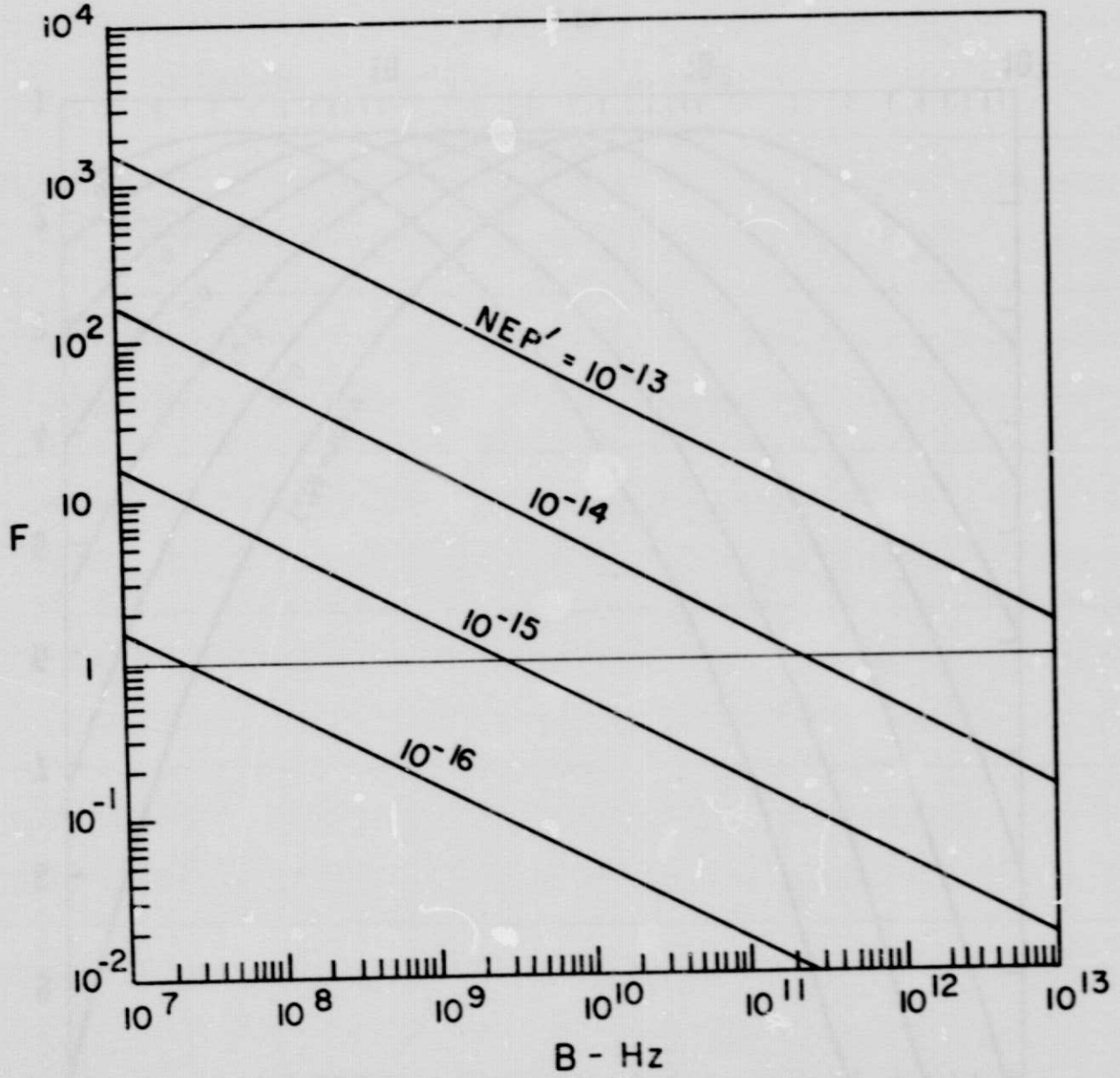
10^4
10^3
10^2
10
1
10^{-1}
10^{-2}
F
NEP' = 10^{-13}
10^{-14}
10^{-15}
10^{-16}
10^7
10^8
10^9
10^{10}
10^{11}
10^{12}
10^{13}
B - Hz

FIGURE 12

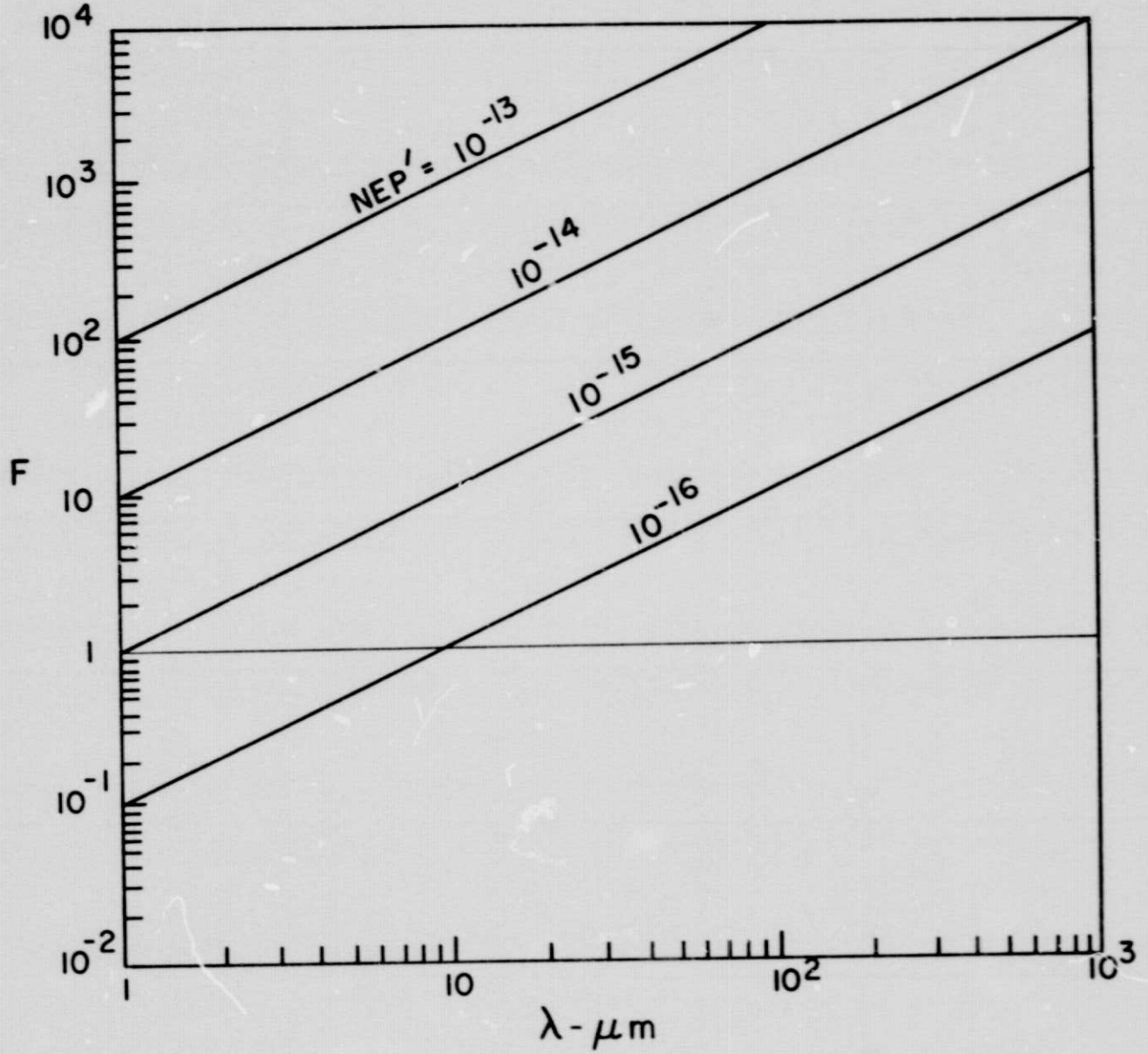
10^4
10^3
10^2
10
1
10^{-1}
10^{-2}
F
NEP′ = 10^{-13}
10^{-14}
10^{-15}
10^{-16}
1
10
10^2
10^3
λ - μm

FIGURE 13