Quantum dot infrared photodetectors: Comparison of experiment and theory

H. Lim, W. Zhang, S. Tsao, T. Sills, J. Szafraniec, K. Mi, B. Movaghar, and M. Razeghi*
Center for Quantum Devices, Department of Electrical and Computer Engineering, Northwestern University, Evanston, Illinois 60208, USA

(Received 24 August 2004; revised manuscript received 13 January 2005; published 17 August 2005)

We present data and calculations and examine the factors that determine the detectivities in self-assembled InAs and InGaAs based quantum dot infrared photodetectors (QDIPs). We investigate a class of devices that combine good wavelength selectivity with "high detectivity." We study the factors that limit the temperature performance of quantum dot detectors. For this we develop a formalism to evaluate the optical absorption and the electron transport properties. We examine the performance limiting factors and compare theory with experimental data. We find that the notion of a phonon bottleneck does not apply to large-diameter lenslike quantum dots, which have many closely spaced energy levels. The observed strong decrease of responsivity with temperature is ultimately due to a rapid thermal cascade back into the ground states. High temperature performance is improved by engineering the excited state to be near the continuum. The good low temperature (77 K) performance in strongly bound QDIPs is shown to be due to the high gain and the low noise achievable in these micron size devices.

DOI: 10.1103/PhysRevB.72.085332 PACS number(s): 73.21.La, 73.22.−f

I. INTRODUCTION

Quantum dot infrared photodetectors (QDIPs) based on InAs and InGaAs technology and the Stranski-Krastanov self-assembly technique are becoming serious competitors to quantum well infrared photodetector (QWIP) devices.[1–5] The field is becoming very exciting because of the science generated by quantum dots (QDs) and because the wavelength sensitive detectivities[6–12] D^* are now in the regime 10^{10} to 10^{11} cm $\mathrm{Hz}^{1/2}$/W. These devices have operating temperatures in the range $T\sim 77$ K to 200 K. Though QDIPs are still a long way from room temperature operation, they can begin to compete with QWIP technologies with the advantage that normal incidence radiation can be absorbed. It is therefore of some interest to examine in more detail, some of the fundamental issues raised by this technology. One would like to decide whether a particular class of material and a particular growth method can ever achieve good room temperature detectors. Many authors have posed this question in the past, but most of the time the debate has concentrated on devices in which the photoexcited state is quasifree.[13–18] It has long been thought that this is the best overall strategy, so most of the modeling work has also been done for bound-to-continuum transitions. Recent data from the Center for Quantum Devices (CQD) and the University of Southern California (USC) group[10,11] seem to contradict this idea. For a while it seemed that the highest detectivity (3×10^{11} cm $\mathrm{Hz}^{1/2}$/W) QDIPs are with those where the excited state is still far from the continuum.[13–18] These "bound-to-bound" transition devices[6,10] are a unique class of QDIPs with particular advantages of their own. They constitute photodetectors, which are highly voltage sensitive, and they could find their way into sensing applications.

The recent work of Chakrabarti *et al.*[13] has demonstrated that weakly bound photoexcited states also "work well" and give the best high temperature performance to date. The authors have demonstrated a $D^*\sim 10^{11}$ cm $\mathrm{Hz}^{1/2}$/W at 77 K and a remarkable 6×10^9 cm $\mathrm{Hz}^{1/2}$/W at 200 K. The photoexcited state is weakly bound compared to Ref. 10, and the dark current is forced to wait behind an $Al_xGa_{1-x}As$ barrier placed near the drain. The QDIP of Ref. 13, however, exhibits a very broadband response, acting more like a heat sensor than a wavelength selective detector. So it seems that if we want wavelength selectivity, then we do need the stronger binding energy to get the narrow line, but presently, at the cost of poor high temperature performance.

The bound-to-bound QDIPs can have reasonable wavelength selectivity from 3–25 μm. The QDIPs can be made with wavelength modules and operate in multicolor. One can envisage focal plane arrays, cameras, and sensors that can detect and identify objects that emit light in well-defined spectral ranges.[6,11] The emitter can be an object or molecule that has been specially tagged and excited to emit in a selected frequency range. The breakthrough made with QWIPs (Refs. 1 and 2) has had exciting biological, medical, and environmental applications. Being able to pick up many wavelengths simultaneously enormously increases the scope of these applications. It gives the detector far greater potential in medical and molecular sensor applications.[19] Heat imaging has been shown to help early diagnosis of skin cancer[2] and IR imaging, as a whole, is now a well-established and invaluable medical diagnostic tool.

With these types of applications in mind, it is very important to have high detectivities and be able to operate at higher than cryogenic temperatures. The main objective of this work is therefore to identify the physical mechanisms and factors that control the performance and suggest how this performance could be improved. In order to do this, we must first understand how the present range of observed detectivities are obtained and what factors limit present performance. We consider two relatively high performance prototypes made in our own center at Northwestern University. The device architectures are shown in Fig. 1 and are referred to as "D_1" and "D_2," respectively. The discussion will also reference the detector made by the USC group, which has a de-

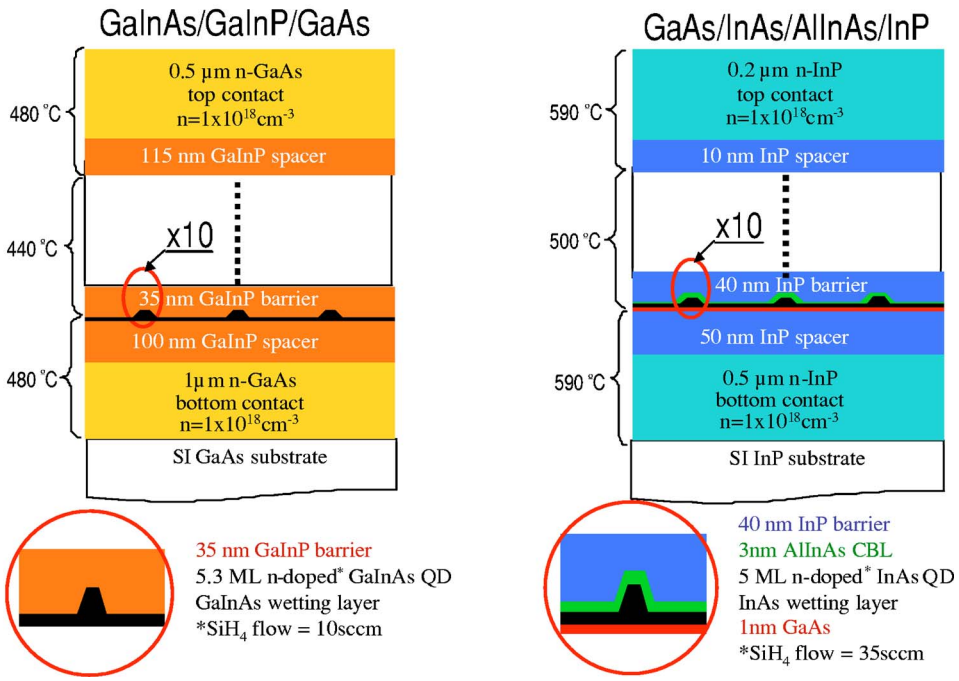

FIG. 1. (Color online) The device structures for QDIPs (left: D_1 and right: D_2).

tectivity of 3×10^{11} cm Hz$^{1/2}$/W, and that we shall call "D_3."[11] The University of Michigan group device, with the best high temperature performance, will be called "D_4" and will be considered the best example of a shallow excited state system.[13]

The objective of this paper is to give the reader a unified, self-consistent picture of the physics of self-assembled QDIPs. This involves relating the photocurrent to the quantum dot (QD) geometry and to the energy level structures; then doing the same for the dark current and showing that the same physical model also explains the observed dark current characteristics. We also analyze the important differences between QDIPs with shallow photoexcited states and those with strongly bound excited states. In the past, this was not done systematically, and the analysis turns out to be crucial if one wants to achieve good high temperature performance. We will mainly focus on D_1 as it is the simplest archetypal device structure. Once we understand D_1, it is relatively simple to understand the strongly bound D_2 and D_3 structures, and the shallow bound state D_4 structures.

The band diagrams for D_1 and D_2 are shown in Fig. 2. The corresponding experimental Fourier transform infrared (FTIR) photocurrent spectra are shown in Fig. 3. Our experimental methods are summarized in Ref. 6. The measured responsivity versus bias is shown in Fig. 4. Most of the discussion will be dedicated to D_1 (InGaAs QD/InGaP/GaAs) because it is the simpler system. The special features of D_2 are that the energy levels are "boxed in." This feature is useful to keep the dark current noise low, the gain high, and the optical excitation bound-to-bound. In device D_3, the QDs are not doped, so that charge has to flow in from the electrodes and fill the dots first. This helps to keep the dark current noise very low and reduces impurity trapping and scattering.

II. ENERGY LEVELS AND OSCILLATOR STRENGTHS

Atomic force microscopy (AFM) images[20] taken in our group show that the QDs in D_1 and D_2 have a lenslike shape. The base lengths and heights for D_1 and D_2 are approximately 40 and 4 nm, 50 and 5 nm, respectively. The barrier from the QD into the InGaP is shown in Fig. 1. The thickness of the wetting layer is not known but we have assumed 1.5 monolayers as in similar InAs/GaAs QDIP structures. These are the parameters fed into the wave function calculation.[21] We have used the effective mass embedding method of Gershoni *et al.*[22] to calculate QD wave functions and matrix elements. This method assumes that one already knows the local effective masses m^* (InGaAs: $0.05m_e$, InGaP: $0.11m_e$, InAs: $0.042m_e$, InP: $0.077m_e$). The error in the choice of an average strain corrected effective mass is $\sim$30%.[23–26] One forces the exact eigenstates to be linear combinations of some selected basis state. Given the near cylindrical symmetry of the lens, a wise choice for the basis is to select 700 eigenstates of another cylinder that is larger than the QD, and then embed the QD inside this cylinder. The QD eigenstates Ψ are then linear combinations of the large cylinder eigenstates as

$$\Psi_m(E_s) = \sum_{n,l} a_{n,l}(E_s) J_m(K_{mn}\rho) e^{im\phi} \sin\left(\frac{l\pi z}{L}\right), \qquad (1)$$

where E_s are the new eigenvalues, J_m are the Bessel functions with ρ denoting the base radial coordinate, m is the eigenvalue of angular momentum in the plane, and l is the index of the confined eigenstates in the z direction with cylinder height L. The energy levels of the basis set are defined by the eigenvalues $J_m(RK_{mn})=0$, where R is the big cylinder radius.

The change in mass is assumed to be abrupt. But if we know the local variation and strain distribution as shown in

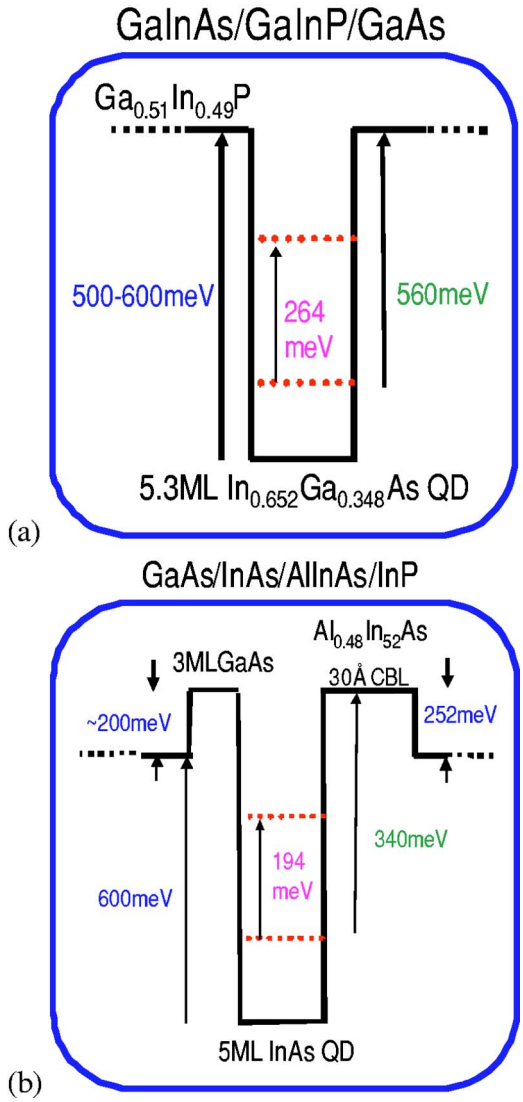

FIG. 2. (Color online) The band diagrams for (a) D_1 and (b) D_2.

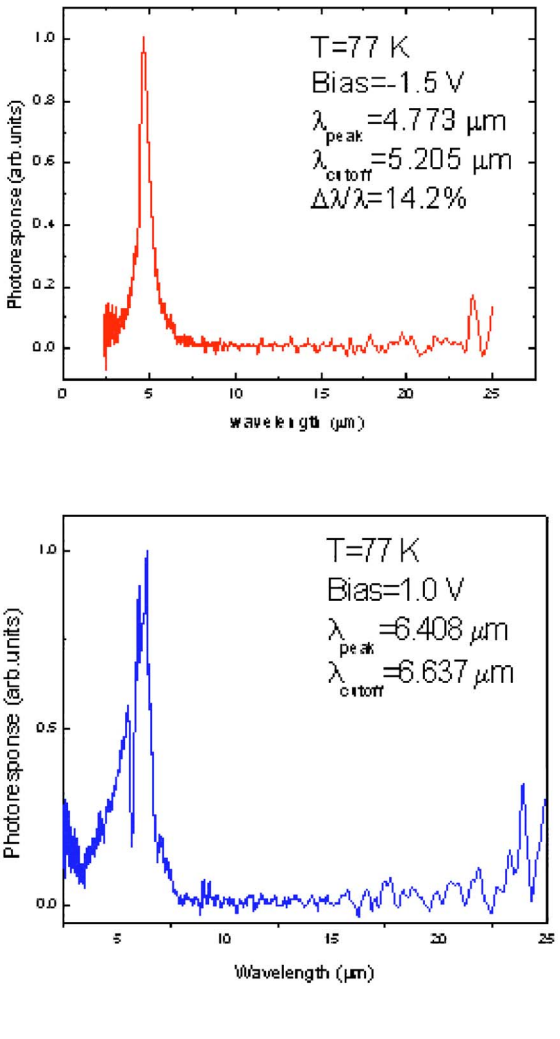

FIG. 3. (Color online) Experimental photocurrent spectra for (a) D_1 and (b) D_2.

Ref. 26 for example, it is in principle a simple matter to upgrade the calculation. Note that due to the rotational symmetry of the "ideal lens," we can assume that the exact eigenstates can be classified using the angular momentum index m (m states greater than 0 are twofold degenerate $=+,-$) and this considerably simplifies the problem. An optical transition with s polarized light is only allowed between states differing in m by one unit.

The real QD arrays are more complex and include size variations, counterion effects, atomic diffusion into the dot, strain, and nonuniformity. We believe that the above simple Ansatz is a good starting point to understand the trends for energies and oscillator strengths. The margin of errors can be quantified as done below.

The bias-induced Stark shift along the growth axis z is a small correction of order 10^{-3} eV at an external field of $F=10^7$ V/m in both devices D_1 and D_2 and has been neglected. The spatial variations of mass are higher order refinements analyzed in Ref. 23.

When taking into account the effect of strain on the magnitude of the confinement potential and effective masses at the interface using the method of Refs. 23–25, one obtains the following sources of error. The conduction band edge could be as low as 533 meV instead of the 700 meV assumed above. Taking the barrier to be 600 meV shifts the photocurrent producing excited state to −0.101 meV instead of the −0.136 shown in Fig. 5. The ground state is shifted up to −0.411 eV instead of the −0.49 eV given. These estimates were obtained using a quantum well approach, which is good enough for flat lenslike QDs of diameter 40 to 50 nm and height ~5 nm as we have in our devices. A full eight-band $\mathbf{k}\cdot\mathbf{p}$ formula for the effective mass with account of dot geometry is given in the recent work of Califano.[23] The effect of the external field on the wave function and energy levels is small in the present range of applied fields.

The embedding technique has been successfully used by a number of authors[22,26,27] and so has the eight-band $\mathbf{k}\cdot\mathbf{p}$ method.[28–30] The best and most reliable technique is still the first-principle pseudopotential method described in Ref. 31. But even in this case, as the authors point out, high accuracy is not attainable unless atomic diffusion and doping contamination are also included.

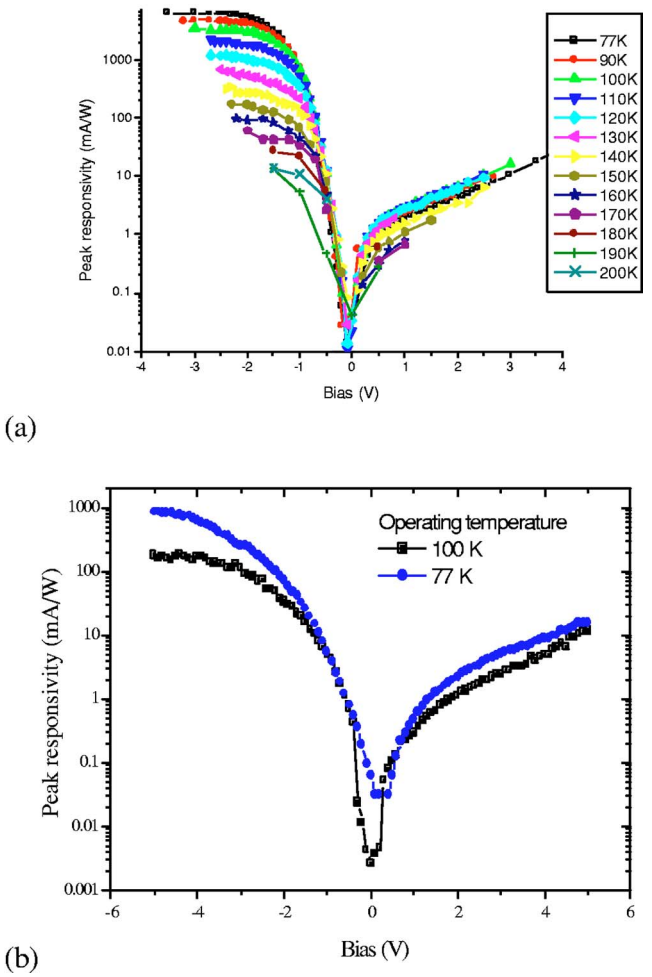

FIG. 4. (Color online) Peak responsivity of (a) a device similar to D_1 but with a higher quantum dot doping (semilogarithm) and (b) device D_2 (semilogarithm).

The embedding method used here is semiempirical. The advantage is that we can easily incorporate the more "precise effective mass variations" derived from first principles[31] or from an experiment into the calculation when needed and exploit the relative mathematical simplicity.

Let us now apply this scheme to understand the data. Looking at Fig. 3 and Fig. 4, we see that according to the FTIR spectral and responsivity behaviors, the measured peak wavelengths (4.7 μm and 6.4 μm) involve excited states that are relatively far ($\sim$136 meV for D_1, and $\sim$149 meV for D_2) from the continuum. The energy level and oscillator strength (f value) calculations for D_1 and D_2 are shown in Fig. 5 and Fig. 6. The numbers show that high oscillator strengths in the range $0.1<f<1$ are almost always associated with transitions from a state (m,n) to a state $(m+1,n)$, i.e., the transition is strong to the level directly above. If the index "n" also differs the transition is weaker and $f<0.1$. The higher "Δn" is, the weaker the transition is. This is true for a simple cylinder, box, or harmonic oscillator as well and is not restricted to the lens.

If we look at possible candidates for transitions matching our measured detection spectrum (these are shown by the

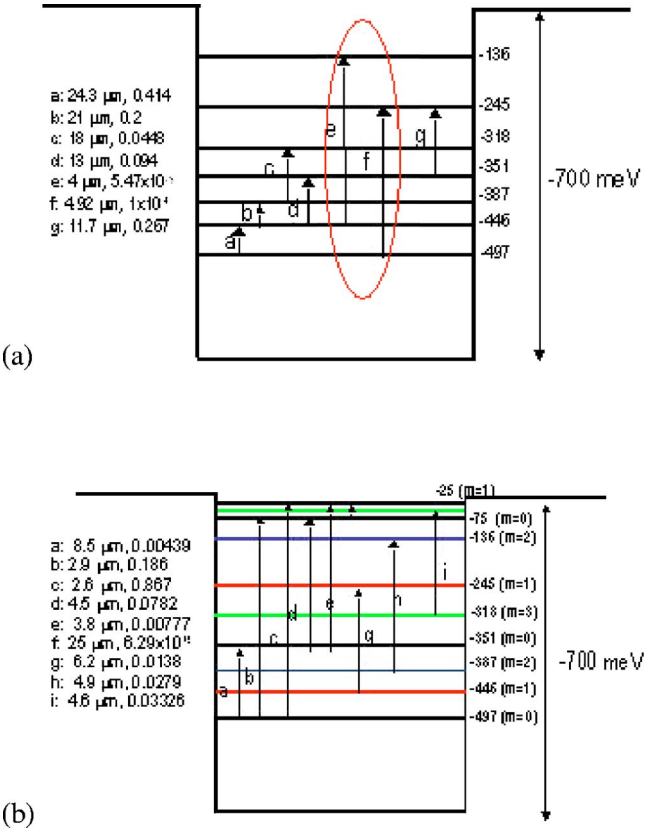

FIG. 5. (Color online) GaInAs QD/GaInP/GaAs energy levels and possible transition: (a) normal incidence, (b) z polarization.

circled arrows in Fig. 5) we note that the f values are $f\sim10^{-4}$ and $f\sim5\times10^{-2}$. These transitions are not the strongest and are not expected to be strongly visible in the absorption at all. The z polarized light, however, has far more respectable oscillator strengths for these energies (see Fig. 5). This is because the neighboring transitions in the z direction are more confined, and the energy difference is large when l changes by only 1 unit. Indeed one can say that the flat lens structure is like a submicron QWIP.

The absorption coefficient $\alpha(\omega)$ is related to the corresponding oscillator strength f_{ge} of a given transition by the formula

$$f_{eg}=\frac{2m^*}{\hbar^2}(E_e-E_g)|x_{eg}|^2, \tag{2}$$

$$\alpha=\frac{\hbar\omega n_{op}N_d}{\varepsilon\varepsilon_0 c}\frac{2\pi}{\hbar}(N_g-n_e)|\langle g|ex|e\rangle|^2\frac{\Gamma}{(\hbar\omega-\hbar\omega_{eg})^2+\Gamma^2}, \tag{3}$$

$$\alpha=\frac{\pi\hbar N_d n_{op}e^2}{m^*\varepsilon\varepsilon_0 c}\left\{\frac{\Gamma}{(\hbar\omega-\hbar\omega_{eg})^2+\Gamma^2}\right\}(N_g-n_e)f_{ge}, \tag{4}$$

where f_{ge} is the oscillator strength for the transition and x_{eg} the in-plane x-transition matrix element from a ground state g to an excited state e with energy difference $\Delta E=(E_e-E_g)$

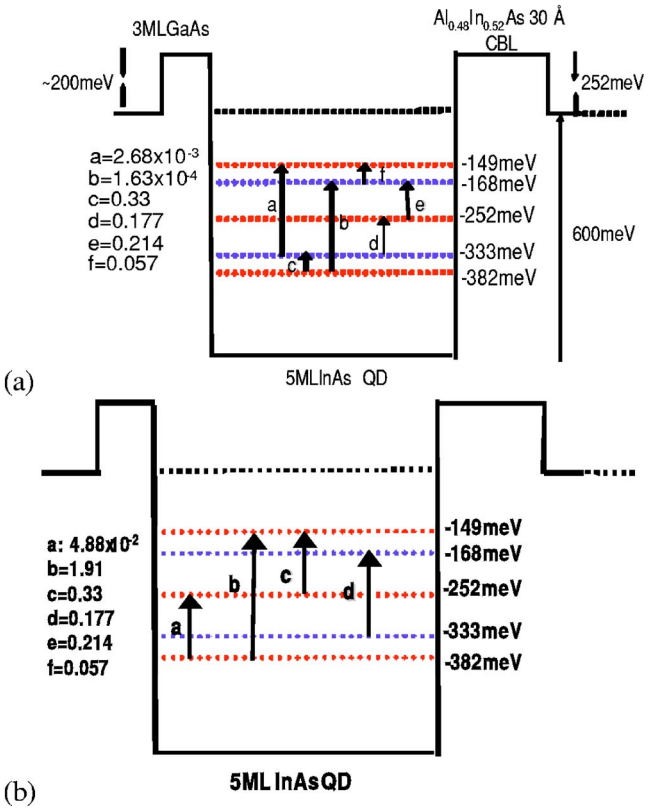

FIG. 6. (Color online) InAs QD/GaAs/AlInAs/InP energy levels and possible transition: (a) normal incidence, (b) z polarization.

$=\hbar\omega_{eg}$. The lifetime broadening is Γ, N_d is the number of dots per unit volume, c the velocity of light, n_{op} its refractive index, and ε_0 and ε are the permeabilities of free space and the medium, respectively. N_g and n_e are occupation probabilities of the ground and excited states, respectively. To compute N_g, we shall, for simplicity, use the Boltzmann distribution in the quantum dot levels but evaluate the band occupation with the Fermi function so that (low bias)

$$N_g = \frac{e^{-E_g/kT}}{\sum_s d_s e^{-E_s/kT} + \sum_t e^{-E_t/kT} + \int_{\varepsilon_c} d\varepsilon \rho(\varepsilon) f(\varepsilon)/N_d} \quad (5)$$

with $N_d = 10^{22}/\mathrm{m}^3$, $\Gamma = 0.02$ eV, $m^* = 0.05 m_e$, the device length $L_d = 10^{-6}$ m, $f(\varepsilon)$ is the Fermi function, and $\rho(\varepsilon)$ is the three-dimensional density of states near the band edge. The sum over s in Eq. (5) runs over the bound QD energy levels, $d_s = 1$ and $m = 0$ or 2 for $m > 0$. The sum over t runs over possible trap levels.

The doping has been selected to give roughly two electrons per dot. The energy level of the Si counterion trap will depend on the position of the ion in the wetting layer and QD and can be anywhere, from the ground state to the continuum. We can therefore assume a constant density of states of dopant states of total strength two levels per dot. In addition there may also be traps due to the InGaP barriers[32] and long-lived polaronic states and resonances[33] (to be discussed later in conjunction with experimental data). The temperature dependence of N_g, allowing all the levels shown in Fig. 5 and Fig. 6 for D_1 and a polaronic trap level at 28 meV above the ground state and $g_c = 1$, is plotted in Fig. 7.

Putting in the numbers in Eqs. (3) and (4), the absorbance at resonance "αL" can be written

$$\alpha L \sim 0.1(N_g - n_e) f_{ge}, \quad (6)$$

and from Eq. (5) and Fig. 7 with $f_{ge} \sim 10^{-2} - 10^{-3}$ at a wavelength of ~ 5 μm, we have an absorbance $\alpha L_d \sim 10^{-3} - 10^{-4}$ in a device of $L = 1$ μm or $\alpha \sim 10^2 - 10^3$ m^{-1}. This is three orders of magnitude less than the estimates given by Phillips for QD transitions to weakly bound states.[15] Our estimate is made at low temperatures, when the occupation of the ground state $N_g \sim 1$ and $n_e \sim 0$. At high temperatures it follows from Eq. (5) that the thermal redistribution of carriers in the dot reduces the strength of a particular transition. The reduction caused by N_g depends on the number of states in the QD and on temperature. Taking the numbers for D_1 with $g_c = 1$ from Figs. 5 and 6 we have the quasi-ideal result for N_g shown in Fig. 7. The reduction in N_g is a factor of 2 up to

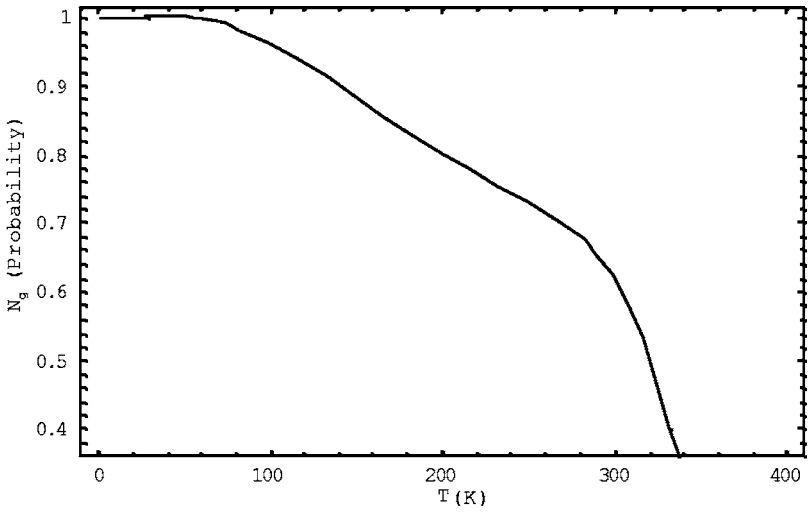

FIG. 7. Probability of staying in the ground state with polaron correction.

300 K. In fact the change may be even less than that since Fig. 5 suggests that the relevant transition denoted by e is not from the ground state but from the first excited state.

From Fig. 5 it would seem that the relevant photoexcitation e has low values of oscillator strength and is unlikely to produce such a high value of peak responsivity R_{peak} as shown in Fig. 4. However, one of the remarkable features of the III-V based QDIP is the high gain, which compensates for the low absorption coefficient in thin devices. Let us therefore examine the quantum efficiency and gain in detail. We will, in the end, show that there is good evidence that the transitions indicated in Figs. 5 and 6, and the f values shown, are indeed the true values (within a factor of 2 or 3).

III. ESCAPE RATES FROM EXCITED STATES

A. The field and temperature dependent escape rates

Consider the escape processes from a QD eigenstate. In the presence of the external field the escape is into Airy functions.[34,35] The Airy functions are the solutions of the effective mass Schrödinge r equation in an electric field. The optical absorption coefficient from quantum well Stark-Wannier states to Airy bands is computed in Ref. 34. Note that as soon as the bias is switched on, the QD bound states are no longer, strictly speaking, bound. They become quasilocalized. The Airy waves penetrate the energy spectrum of the dot and this allows the charge to escape by a variety of pathways. The charge has two extreme paths to come out: (i) the pure horizontal adiabatic tunnel path and (ii) the pure vertical activation pathways. In addition, we also include all the transitions in between those two extrema. The single phonon-assisted matrix element from the optically excited state e to a continuum Airy state A_μ with energy E_μ can be formally written in terms of the electron-phonon interaction coupling[34,35] V_{ep} as (absorption)

$$\nu_{e\mu} = \frac{2\pi}{\hbar}\sum_q \left| \int_0^\infty d\vec{r} A_\mu(r) V_{ep}(q) \Phi_e(r) \right|^2 n_q \delta(E_\mu - \hbar\omega_q - E_e), \tag{7}$$

where for example $A_{\mu=0}(x') = \int_0^\infty ds \cos(s^3/3 - sx')$ when $x' > 0$ and $x' = (2e\mathrm{Fm}^*/\hbar^2)^{1/3} x$. The form of $V_{ep}(q)$ for acoustic and optic modes in InGaAs is given in Appendix A.

For a one-optic phonon transition or one-acoustic phonon transition, the attempt frequency is typically $\nu_{ec} \sim 10^{12}-10^{13}$ Hz, reduced by the tunnel overlap from the localized QD level into the free part of the wave. We assume that the particle can tunnel and be excited to any intermediate Airy wave. The Airy state starts at each lattice site in the field (z) direction, the intermediate transitions span the two extreme paths. One can take care of the spatial tunnel factor and approximately sum over all paths by assuming a decay that is exponential with tunneling distance and obtain the elegant analytical expression

$$W_{ec} = \nu_{ec} e^{-E_{ec}(F)/kT} = \nu_{ec} g_c \left\{ \frac{e^{-E_{ec}/kT} - e^{-\varsigma E_{ec}^{3/2}/eFa} e^{-\varsigma E_{ec}^{1/2}} e^{eFa/kT}}{1 - e^{-\varsigma E_{ec}^{1/2}} e^{eFa/kT}} \right\}, \tag{8}$$

$$\varsigma = a\left(\frac{2m_e^*}{\hbar^2}\right)^{1/2}, \tag{9}$$

where E_{ec} is the energy difference between the photoexcited state and the continuum; a the lattice constant, and ν_{ec} is a phonon-assisted prefactor, which is assumed constant for simplicity, though it should in principle vary somewhat with temperature and energy.[36] It is known for the pure tunnel path to be $\sim E_e/\hbar$, and g_c a density of final states[1] factor, which is $g_c \sim 10^4 (kT/e)^{3/2} \sim 10$ (all QD-band edge states reachable within kT) at low bias. The product $\nu_{ec} g_c$ is in the range 10^{13} Hz (an optic phonon frequency) to 10^{14} Hz (includes a density of states enhancement prefactor). In Eq. (8), the paths sum changes from tunnel-like to activated as we raise the temperature, the turning point is at $\varsigma(E_{ec})^{1/2} = eFa/kT$. The escape efficiency out of the photoexcited state multiplied by the electric field $[\nu_{ec} e^{-E_{ec}(F)/kT}/(\nu_0 + \nu_{ec} e^{-E_{ec}(F)/kT})]F$, is plotted in Figs. 10(a) and 10(b) for the calculated escape energy of 136 meV.

B. Many-body assisted detrapping

A carrier trapped in the excited state e in Fig. 5(a) of a QD can be detrapped by the normal inelastic electron-electron scattering process, or by collisions with drifting charges, because there is a current flowing in the device. The first (intrinsic zero current) process is not effective at the low carrier densities present in D_1 and D_2. Here the average distance between charges is of order 50 nm, and given a dielectric constant $\varepsilon \sim 13$, we obtain a Coulomb interaction energy $V_{ee} \sim 5$ meV between carriers, which is less than kT. The second process is the current-induced detrapping process and has been studied by Kochman and coworkers using Monte Carlo simulations.[16] The authors have shown that current-assisted detrapping out of the QD is significant at high currents with shallow bound states. But the current-assisted detrapping process is always a small contribution to the photoconductive gain. An analytical estimate of this rate as a function of bias is given in Appendix B.

The intradot "Hubbard" repulsive energy U is important in small QDs and has been discussed in the work of Williamson *et al.*[31] and Fricke *et al.*[37] In our devices, the bare value of U is around 0.05 eV and is not negligible. Using a mean field argument, one can argue that the effect of the Hubbard U correlation is to introduce an extra energy shift. The value of the electron energy shift differs according to how many charges occupy the QD. These occupation number-dependent energy shifts broaden the photocurrent signal. The correlation broadening estimated here is within the broadening of the observed signal, which is $\sim$30 meV, and is important for facilitating optic phonon cascades included in ν_0. So we conclude that apart from structural inhomogeneities, correlations can also broaden the photocurrent signal. Electron correlations will also prevent an electron from entering the dot. The uncompensated repulsive potential is $V(r) \sim \langle n_d \rangle e^2/4\pi\varepsilon\varepsilon_0 R$ (R=dot radius and $\langle n_d \rangle$ average extra charge occupation), which is approximately an optic phonon energy (30 meV). The repulsive barrier helps to lower the capture rate in an occupied QD.

IV. RESPONSIVITY

A. The responsivity formula

Now we can use Eqs. (3), (5), and (8) and write for the responsivity (photocurrent/optical power=Amp/W) associated with the transition from level g to level e, as shown in Fig. 4,

$$R=\left\{\frac{1}{\nu_0}\frac{2\pi\Gamma}{[(\hbar\omega-\hbar\omega_{ge})^2+\Gamma^2]\hbar}|\langle g|ex|e\rangle|^2\right\}\left\{\frac{I_0(T,V)}{A_L\varepsilon\varepsilon_0 c}\right\}\times[e^{-E_{eff}/kT}](N_g-n_e) \tag{10}$$

in terms of absorption coefficient α, gain g, and quantum efficiency η we have

$$R=g\frac{e\alpha(\omega)L}{\hbar\omega}\left\{\frac{\nu_{ec}e^{-E_{ec}(F)/kT}}{\nu_0+\nu_{ec}e^{-E_{ec}(F)/kT}}\right\}=\frac{e}{\hbar\omega}g\eta, \tag{11}$$

$$\eta=\alpha(\omega)L\left\{\frac{\nu_{ec}e^{-E_{ec}(F)/kT}}{\nu_0+\nu_{ec}e^{-E_{ec}(F)/kT}}\right\}, \tag{12}$$

where $E_{ec}(F)$ is an effective escape energy defined by (8)

$$g=\frac{\mu F}{LC_{be}}, \tag{13}$$

$$I_0=\frac{Ae\mu(F,T)FN_d}{(1-n_e)C_{be}(F)}\nu_{ec}\rightarrow n_e<1. \tag{14}$$

The quantity g is the gain and is defined as the ratio of the recombination time to the transit time: η is the quantum efficiency; $\alpha(\omega)L$ is the absorbance, here assumed to be <1; A is the current area; and μ is the band mobility not including capture from quantum dots but it does include scattering from dots and wetting layers and also, for example, the effect of the extra barrier in D_4; ν_0 is the relaxation rate from the photoexcited state to all other states. This relaxation rate would be slow and a phonon bottleneck if the QD had, for example, only one other eigenstate at an energy separation larger than several optic phonon energies. This is not so in our system, and $\nu_0(T)$ is the reason for the strong temperature decrease of the responsivity, which is discussed later. The quantity C_{be} is the quantum mechanical capture rate into the QD excited state, and $(1-n_e)$ is the probability that this state is empty. Here $n_e<1$, the case $n_e\sim 1$ is treated later. In this paper, the capture rate will be deduced from the experimental gain g. We will deduce a number of $\sim 10^{10}$ Hz, which is exactly in the right range for a deep level.[15–17] Estimates in the literature are in the range $C_{be}\sim 10^{11}-10^{12}$ Hz for shallow excited states, which are reachable by acoustic phonon emission.

B. Responsivity summary

There are three main sources of temperature dependence: (a) from the escape rate W_{ec}, (b) from the occupation probability of the (optical) ground state N_g, and (c) from the lifetime $\nu_0(T)$ to which the escape rate also has to be added as shown in Eq. (11). The band mobility $\mu(T)$ is expected to be only weakly dependent on temperature in comparison.

For high accuracy, the line shape need to be convoluted with a Gaussian distribution over the escape energy E_{ec} (dot size variations) and intradot electron-electron interaction energy shifts.

The peak responsivity $R_{\text{peak}}(T,V)$ data for D_1 and D_2 are shown in Figs. 4(a) and 4(b) as a function of bias for different temperatures. Focusing on one temperature in Fig. 4, we see that when the field is strong enough to pull down the QD escape barrier, then $R(T,V)$ saturates with V. At this point, the temperature dependence of $R(T,V)$ is almost exclusively due to (i) the thermal redistribution of electrons in the QD, that is, the factor N_g and (ii) the lifetime processes contained in the relaxation rate $\nu_0(T)$. So at peak value in voltage, the temperature behavior of R is dictated by the thermally induced changes in the absorption and the excited state lifetime. In Ref. 33 it is shown that in InAs/GaAs QDs, N_g is the main cause of absorption decrease at high temperatures. The authors however cannot explain the too rapid fall in absorption that starts at low temperatures with the simple "bare QD energy level model." They need to invoke polaron resonances to explain their data. The bare (not including the electron-phonon coupling) energy level distributions of D_1 and D_2 are shown in Fig. 5 and Fig. 6. The redistribution of charge into the bare QD eigenstates causes N_g to vary only slowly with temperature. In order to explain the strong decrease with temperature, the authors of Ref. 33 refer to the work of Hameau *et al.*[38] They have computed the exact new QD eigenstates including the electron-phonon coupling up to one phonon occupation. They find a new twofold degenerate polaron eigenstate with energy intermediate between "*s* and *p*." The idea is that the new level is an efficient trap for the carrier excited out of the ground state.

We have incorporated this polaron trap in Eq. (6) and evaluated N_g in Fig. 7, but the discrepancy between the behaviors of N_g shown in Fig. 7 and the responsivity data is here still too big to be simply due to thermal redistribution. The change seen is not just a factor of 2 or 3 when we go up to 170 K, but more than one order of magnitude. To identify the source of the strong temperature decrease of responsivity, we need to consider one more piece of vital information. This is contained in Ref. 13, where we have an InAs/GaAs system with a weakly bound excited state. The weakly bound excited state gives a more stable performance in $R(T,V)$ as we raise the temperature. This gives us the supporting information we need to come to a firm conclusion regarding the strong temperature dependence of responsivity.

C. The temperature dependence of the responsivity and the absence of the phonon bottleneck

What is happening is that in QDs of such large size, the energy levels are closely packed as shown in Fig. 5. The phonon bottleneck, which is in principle one of the most important features of QDs,[39,40] is no longer effective in such a situation. Apart from the fact that the excited state has six lower lying states to relax into ($m>0$ levels are doubly degenerate), there will also be the intermediate polaron resonances that Hameau *et al.*[38] and Sakaki *et al.*[41] have dis-

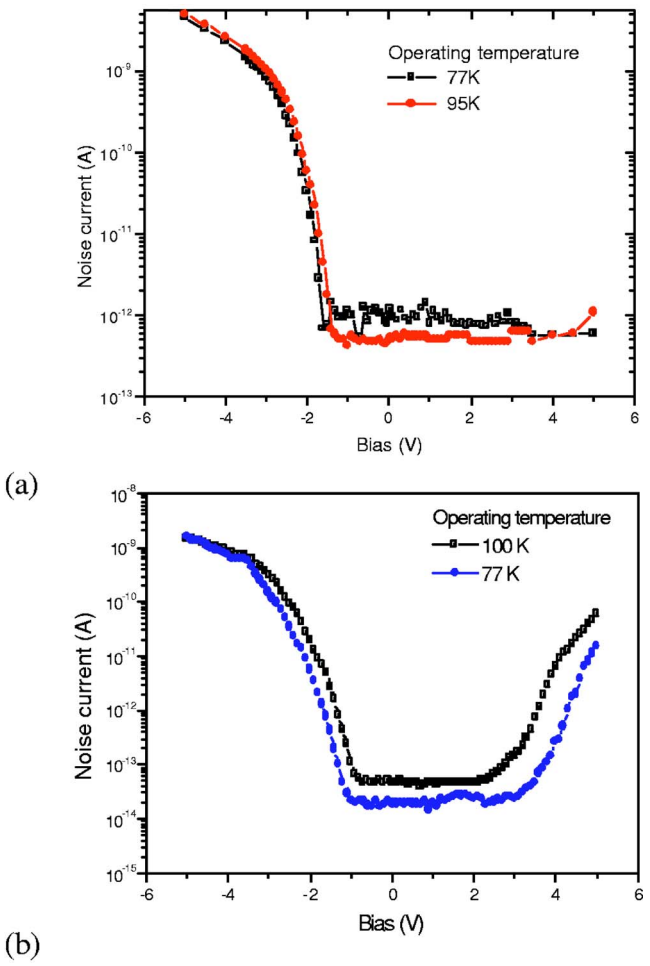

FIG. 8. (Color online) Noise current for (a) D_1 (semilogarithm) and (b) D_2 (semilogarithm).

cussed and mentioned above. These authors have shown how effective temperature is in increasing the two-phonon relaxation rates. If you add to this the further complication of atomic disorder inside the dot, and the existence of the occasional silicon counterion trap as well as Auger broadening and wetting layer traps, one arrives at a picture of a dense ladder of eigenstates with rapid resonant two to three phonon thermalization of charge. Furthermore as we go up in temperature, it is possible for the carrier to first absorb an acoustic phonon in order to complete a resonant multioptic process to a lower energy. So if the rate ν_0 in true bottleneck situation is typically 10^9-10^{10} Hz, this rate will quickly go up to a value of $\sim 10^{12}$ Hz[38,42] as we go up in temperature. However, when the excited state is very close to the continuum as in the device of Ref. 13, the electric field and temperature can make the ionization step compete favorably with the relaxation step, and so the responsivity remains relatively high as we go up in temperature. This view is also supported by the data of Kim *et al.*[43] who have measured at $T=300$ K a responsivity as high as 5 A/W in a device where the photo excited state is "immediately" trapped by a laterally addressed high electron mobility transistors (HEMT) channel.

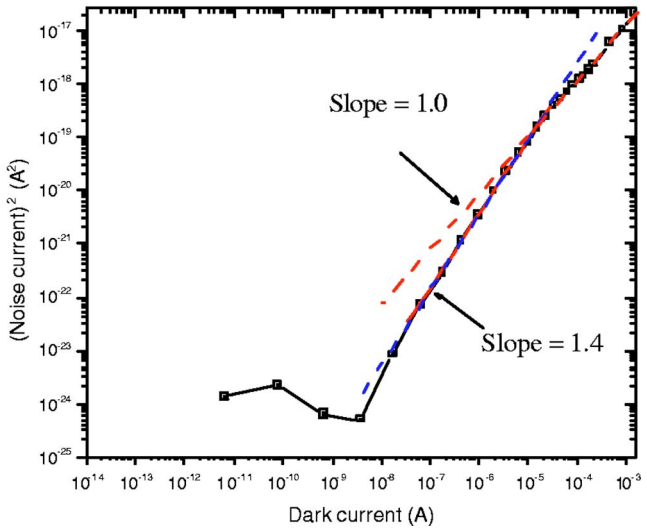

FIG. 9. (Color online) Square of noise current vs dark current (log-log).

V. EXPERIMENTAL AND THEORETICAL RESPONSIVITY

A. Quantum efficiency and responsivity: Comparison with experiment

The quantum efficiency η is defined in Eq. (12). The quantum efficiency can be estimated by measuring the responsivity and the noise, and then using the generation-recombination (GR) noise analysis, which relates the noise I_n and the dark current I_D at a bandwidth $\Delta\nu$ via

$$I_n^2 = 4qg\Delta\nu I_D. \tag{15}$$

For D_1, it was calculated in Ref. 4 that $g\sim 800$ with $\eta\sim 2\times 10^{-4}$. Experimental values of D_2 are only slightly different, but g is strongly bias dependent with maximum $g\sim 1200$ and $\eta\sim 10^{-4}$. The somewhat higher gain is related to the boxed-in dot structure but is not exploited in D_2 device because the high gain is unfortunately also in the region of high noise (see Fig. 8). Going back to D_1, the noise-dark current relationship is shown in Fig. 9. It should be noted that the GR noise formula is in principle valid at any real values of dark current, even when the noise is not exactly proportional to the square root of the dark current. The experimental relationship between noise and dark current is complex, and here it is of the form $I_n^2\sim (I_D)^n$ where $n>1$.[44] This is not surprising as the expected gain is voltage dependent and the present variation of I_D is based on voltage changes. The gain in D_1 was estimated to be ~ 800 for D_1 at $V=-1.5$ V.

Looking at the measured quantum efficiency and the estimated absorbance in Eqs. (6) and (12), we note that in the low temperature and high bias limit, the escape factor saturates to ~ 1, so theoretically $\eta\sim \alpha L\sim 10^{-3}-10^{-4}$ which is the correct number despite the low value of f_{ge}. The corresponding R with $g\sim 1000$ is then $R\sim 3$ A/W, which is also the correct value. Note that the concept of gain implies that the system retains its charge neutrality during the transport process. This is not completely self-evident when we have a high injection barrier and/or a small number of QD layers and must be checked in each device.

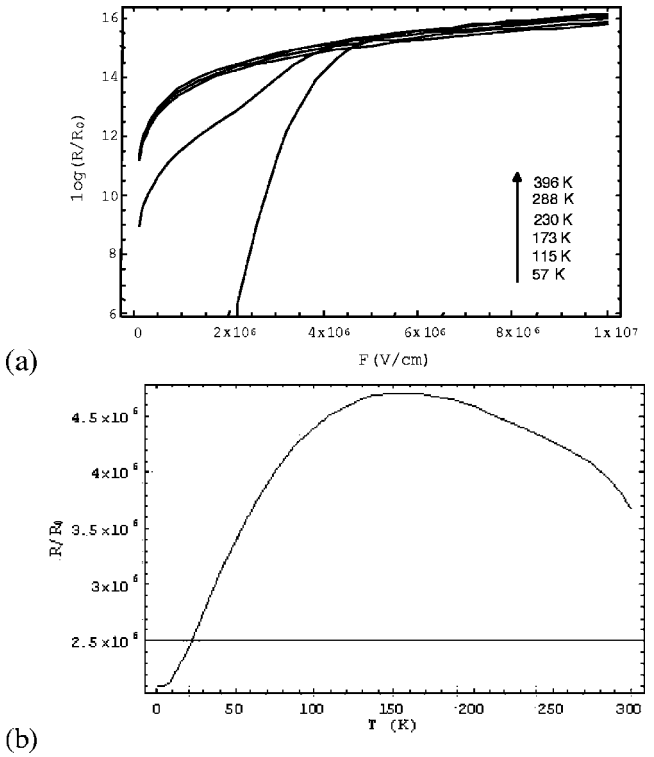

FIG. 10. (a) The "bias-dependent part of the responsivity" (defined as the product $N_g[\nu_{ec}e^{-E_{ec}(F)/kT}/(\nu_0+\nu_{ec}e^{-E_{ec}(F)/kT})]F=R/R_0$) as a function of field (V/cm). (b) The "temperature-dependent part of the responsivity" (defined as $N_g[\nu_{ec}e^{-E_{ec}(F)/kT}/(\nu_0+\nu_{ec}e^{-E_{ec}(F)/kT})]F=R/R_0$ with $n_e\sim 0$ and ν_0 constant) as a function of temperature (K) with polaron correction at $F=6\times10^4$ V/cm for N_g but with constant (T independent) relaxation rate ν_0.

B. Temperature dependence of the responsivity

The experimental temperature dependence of $R(T,V)$ for D_1 is shown in Fig. 4. The theory Eqs. (11)–(13) for $R(T,V)$ predict that increasing the temperature will (i) enhance the escape rate, (ii) decrease N_g, and (iii) reduce the lifetime. The first two processes can be quantified quite accurately, the third one must remain semiphenomenological though we do have the two phonon calculations of Refs. 38 and 42. The theoretical curve for R as a function of bias, for an escape energy of 136 meV is shown in Figs. 10(a) and 10(b). The temperature variation is shown in these figures assuming a perfect bottleneck for illustration. We have calculated the normalized responsivity including the polaron trap in N_g [Fig. 10(a)]. The temperature dependent normalized responsivity is shown in Fig. 10(b). We have not tried to fit the temperature dependence of the true lifetime $1/\nu_0(T)$ and treated it as a constant. Consequently, the experimental drop in temperature is bigger than expected from theory. The difference between theory and experiment can be attributed to $\nu_0(T)$.

VI. THE DARK CURRENT AND NOISE

The experimental dark currents for D_1 and D_2 are shown in Fig. 11. The data for D_1 is plotted on an Arrhenius scale in

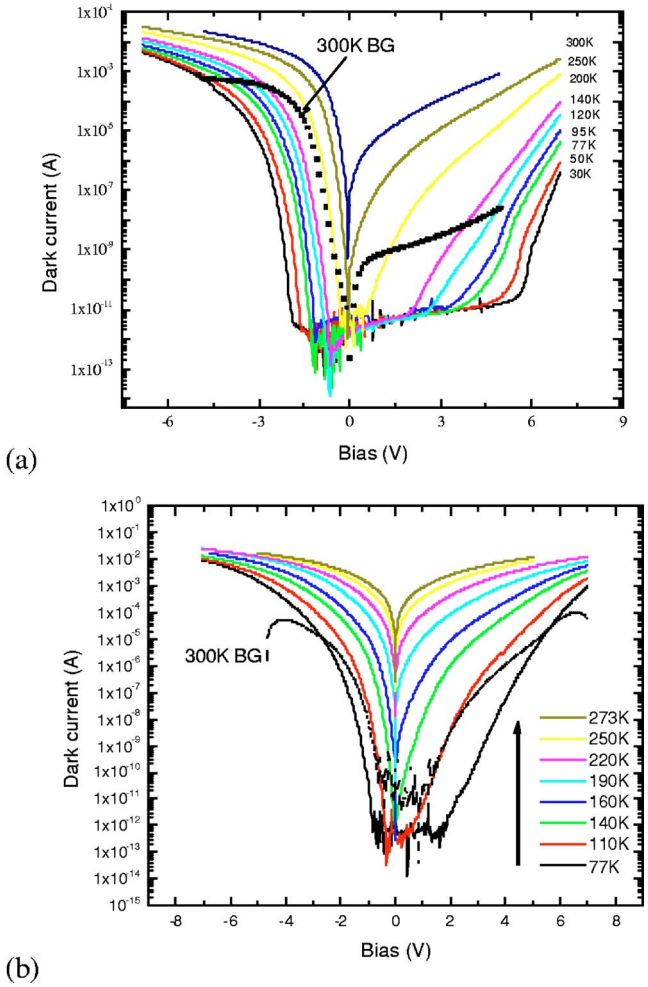

FIG. 11. (Color online) (a) Dark current of D_1 (semilogarithm); (b) dark current of D_2 (semilogarithm).

Fig. 12. QDIP dark currents have been discussed in a number of publications[15,16,18] but for specific geometries and conditions. Although very useful contributions have been made to this field, we need to do our own analysis here.

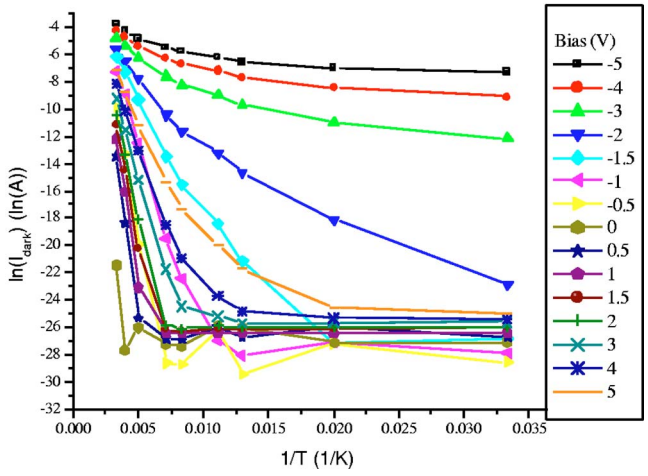

FIG. 12. (Color online) Arrhenius plot of the experimental dark current of D_1.

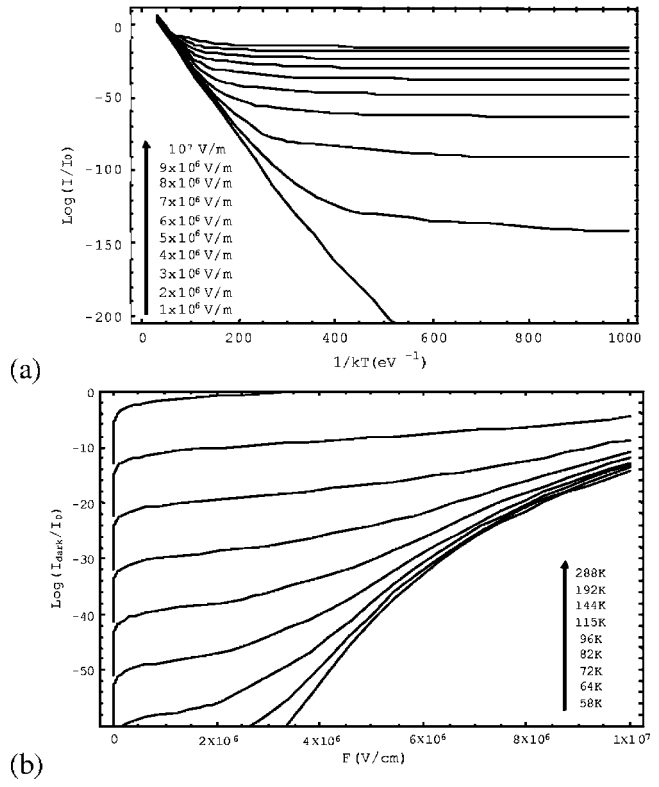

FIG. 13. Theoretical dark current (a) Arrhenius plot and (b) field (V/m) dependence at different temperatures.

Looking at Fig. 12 for D_1 at negative bias, we should note that "below" a negative critical bias (around −1 V) the measurement is instrumentation limited. The first activation energy we could extract with confidence was ~400 meV at low bias, which is indeed roughly the escape energy from the ground state in Fig. 5. To proceed with further insight, we need to know the height of the injection barrier from the contact layer. In D_1, the barrier from the *n*-doped GaAs contact Fermi level to the band edge in InGaP has been measured[45] to be $E_b \sim 0.3$ eV. A value of 0.22 eV was measured for GaAs/InGaP ($10^{16}/\mathrm{cm}^3$) in Ref. 45. This is a higher doping level for the barrier than here.[45] This energy is slightly lower than the energy difference between the ground state of the QD to the edge of the band, see Fig. 5 and Fig. 6. The Fermi level in the dot aligns with the contact reservoir. Indeed the contact resistance has been measured using a dot free device and is much smaller than the bulk dot layer resistance. So we are indeed in the range of parameters of interest, measuring the QD layer resistance.

The dark current as evaluated from dot layer to dot layer consists of two contributions: from carriers that escape the dot by field and thermally assisted escape to the band edge called I_d and carriers that are thermally excited high up in the barrier band called I_b. The Fermi level at low temperatures is assumed to be roughly between the ground and first excited level at $\sim(E_g+E_{e1})/2$ but will move up with bias as charge piles up in the QDs. The above formula assumes symmetric escape barriers. The theoretical dot-to-dot dark current is plotted in Fig. 13. For simplicity we write the dark current in terms of an exponential of an effective field and temperature dependent activation energy $E_D(F,T)$, namely as $(n_e<1)$

$$I_{dot} = \frac{Ae\mu(F)FN_d}{(1-n_e)C_{be}}\nu_{ec}\exp[-E_D(F)/kT], \tag{16}$$

$$I_b = eA\int_{E_c} dE\rho(E)f(E)\mu F,$$

$$I_{dot} = \frac{Ae\mu(F)FN_d}{(1-n_e)C_{be}}\nu_{ec}\left\{\sum_s \frac{g_s^0 f_s}{\left(1+\frac{W_{sc}f_s}{(1-n_e)C_{be}}\right)} \times\left(\frac{e^{-E_{sc}/kT}-e^{-\varsigma E_{sc}^{3/2}/eFa}e^{-\varsigma E_{sc}^{1/2}}e^{eFa/kT}}{1-e^{-\varsigma E_{sc}^{1/2}}e^{eFa/kT}}\right)\right\}, \tag{17}$$

where $f(E_s)$ is the "Fermi function" (with nonequilibrium renormalization of Fermi level and temperature) and

$$I_0 = Ae\mu N_d g_0 \frac{\nu_{ec}}{C_{be}}F = AeN_d g_0 \nu_{ec} gL, \tag{18}$$

where g is the gain and ~500 at 2 V. We can estimate

$$I_0 = [(400\times10^{-6})^2\times(1.6\times10^{-19})\times10^{22}\times100\times10^{13} \times500\times10^{-6}]\sim10^8\mathrm{A},$$

where each term in the product appears in the same order as the right hand side of Eq. (18) (in mks units).

A. Summary

We summarize as follows:

(i) The dark current is only truly activated at high temperatures (see Fig. 12).

(ii) This "activation energy" decreases strongly with bias as was already shown in Ref. 6.

(iii) In the limit of very weak capture $C_{be}\to0$, it follows from Eq. (17) that there is a strong dark current gain. It follows from Eq. (17) that carriers escape the dot and flow around the circuit can stay in the band for a long time before they recombine.

Equation (17) is plotted in Fig. 13 as an Arrhenius plot and as a function of the field with different temperatures. The prefactor I_0 is defined by Eq. (18). This number is based on an experimental gain of $g=500$, together with $L=10^{-6}$ m and $\mu=1$ m^2/V s. The capture rate is deduced from the measured gain with the above numbers, we have, for D_1

$$(1-n_e)C_{be}\sim10^{10}\ \mathrm{Hz}. \tag{19}$$

This is exactly in the right range for a strongly bound level.[15,16] These numbers are empirical estimates. A complete theory should, in principle, produce numbers for mobility, recombination rate, and gain. This is difficult to do at this stage because there are uncontrollable factors such as trap distribution and atomic interdiffusion in the QDs, which change the electronic structure and effective masses. Also, there are noise processes that are not generation-recombination limited.

B. The strong bias asymmetry in the dark current of D_1

The experimental current voltage asymmetry shown in Fig. 11 for D_1 is related to the asymmetry of the contact barriers and to the QD growth conditions.[6] The asymmetry of the QD barrier can be related to the nonuniformity of atomic distributions inside the dot as was recently demonstrated for InAs on GaAs by Hanke *et al.*[46] We have grown different types of (InGaAs QD/InGaP/GaAs) devices, changing the number of QD layers and varying the dot growth conditions. Even the dot free device turns out to be asymmetric. The trend we can report is that the I-V asymmetry grows as the dark current decreases. The observations suggest that the way these dots are allowed to self-assemble and the dislocation densities are the key to the symmetry question. Atomic diffusion[47] can also be a serious problem. It should be noted that the responsivity is also asymmetric see Fig. 4(a). This supports the view that both the contact region and the dot environment contribute to asymmetry. The asymmetry in the responsivity $R(V)$ is shown in Fig. 4(a) and is most interesting. Whereas the negative bias dependence of R is very strong and tunnel-like, the positive bias is for a large part exponential in voltage and "semi-classical" (space charge related) in structure.

VII. DETECTIVITY D^*

A. The detectivity equations

The detectivity D is defined experimentally as the ratio

$$D = \frac{R}{(eI_n^2)^{1/2}}. \tag{20}$$

Theoretically using Eq. (16) we have per unit bandwidth

$$D^* = \frac{R\sqrt{A}}{(egI_D)^{1/2}}. \tag{21}$$

Substituting from Eqs. (11) and (15) we obtain

$$D^* = \left\{ Q\frac{\nu_{ec}}{[W_{ec}+\nu_0(T)]^2} \right\}^{1/2} \frac{\alpha(\omega,T)}{A^{1/2}N_d\hbar\omega} e^{(E_D/2kT-E_{ec}/kT)}, \tag{22}$$

$$D^* = D_0^* e^{(E_D/2kT-E_{ec}/kT)}, \tag{23}$$

where Q is the total number of QDs in the device. In the low temperature regime where α and ν_{ec} are temperature independent, we have with $\nu_{ec}\sim 10^{13}$ Hz and $\nu_0\sim 10^{10}$ Hz $\gg W_{ec}$, and $L=10^{-6}$ m, a $D_0^*\sim 10^7$ cm Hz$^{1/2}$/W for device D_1 (see the more detailed analysis below). The experimental data are shown in Fig. 14 for D_1 and D_2.

B. D^* the geometry dependence

Here $Q=ALN_d$ is the total number of quantum dots in the device. There is in reality only a weak geometry effect via the gain which is $D^*\sim L^{1/2}$ and which would disappear in thick devices where $g<1$.

C. D^* the bias dependence

From Fig. 14, one can observe that D^* reaches a peak value, and then sharply turns down again with increasing voltage. In Eq. (23), we see that the voltage dependent effective activation "energy" $E_D(V)/2$ can still be important and still decrease with V, while the escape energy $E_{ec}(V)$ is already negligibly small and comparable to kT. The responsivity at low temperatures saturates when

$$\frac{4}{3}\left(\frac{2m^*\times 1.6\times 10^{-19}}{\hbar^2}\right)^{1/2}(E_{ec})^{3/2}\sim F \tag{24}$$

with $E_{ec}\sim 0.136$ eV, whereas the escape out of the dot saturates for $E_{gc}\sim 0.45$ eV.

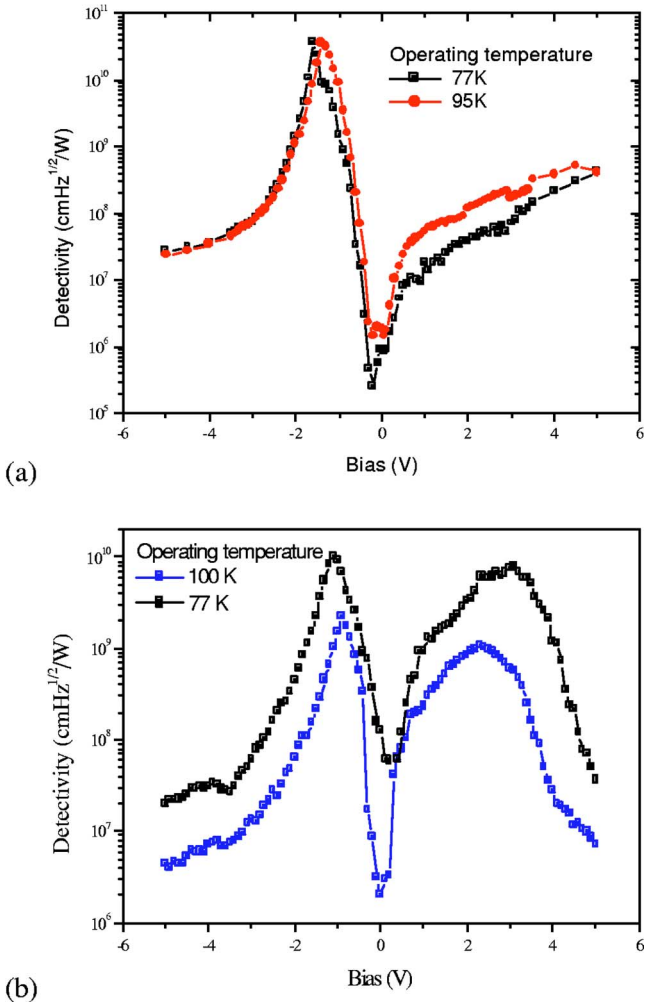

FIG. 14. (Color online) Detectivities of (a) D_1 (semilogarithm) and (b) D_2 (semilogarithm).

D. D^* the temperature dependence

From Eq. (22) it follows that the temperature dependence of D^* is determined by the following:

(i) The exponential of the difference between half of the effective activation energy of the dark current $E_D(F)$ and the effective activation energy $E_{ec}(F)$ of the responsivity $[E_D(F)/2-E_{ec}(F)]$.

(ii) The temperature dependence of the absorption, which in turn is mainly due to the carrier redistribution N_g in the dot and band.

(iii) Phonon-assisted resonant cascades down in energy ν_0. The phonon assisted escape to continuum prefactor ν_{ec} is expected to be only weakly temperature dependent in comparison.[36]

VIII. COMPARISON OF D^* THEORY AND EXPERIMENT

A. Comparison D^* with D_1 and D_2

Looking at Eq. (23) for D^* and using the numbers quoted above for the responsivity R, absorbance, and gain g of D_1 we note that at peak value in Fig. 4 ($V=V_{R\max}$) and Fig. 9, R has saturated with V, but the noise I_n has not. So we can rewrite D^* ($V_{R\max}$) for D_1 as

$$D^* = \left[Q \frac{\nu_{ec}}{(W_{ec}+\nu_0)^2} \right]^{1/2} \frac{\alpha(\omega,T)}{A^{1/2} N_d \hbar\omega} \exp\{E_D(V_R)/2kT\}, \tag{25}$$

$\alpha L=0.1$; $\nu_{ec}=10^{13}$ Hz; $N_d=10^{22}/\mathrm{m}^3$; $\hbar\omega=0.25$ eV; $\nu_0=10^{10}$ Hz noting that $D^*(T=95\ \mathrm{K}, V=2)=3\times10^{10}\ \mathrm{cm\,Hz^{1/2}/W}$, neglecting W_{ec}, we extract the formula

$$D^* = 10^8 \exp\{E_D(V_R)/2kT\} \mathrm{cm\,Hz^{1/2}/W} \tag{26}$$

with $E_D(V=2)=0.16$ eV, which in turn implies a drop of $D^*(V=2,T)$ to a maximum possible value of $\sim 10^8 \exp(8/3) \sim 10^9\ \mathrm{cm\,Hz^{1/2}/W}$ at room temperature $kT \sim 0.03$ eV. This is under the assumption that, $W_{ec} < \nu_0$ and the phonon bottleneck exists, that is, $[(\nu_{ec}/\nu_0^2)]^{1/2}$ can be kept constant with temperature. Clearly this is not so in the large lenslike QDIPs, so here the $T=300$ K estimate for D^* will be two to three orders of magnitude lower than predicted by Eq. (24) as ν_0 increases up to 10^{13} Hz. Indeed, the room temperature D^* of D_1 and D_2 was too small to be measured because we did not have enough signal from $R(T=300\ \mathrm{K})$.

When the dark current and noise saturate with bias, as we still increase the voltage, the exponential factor in Eq. (25), which increases D^*, is now reduced ~ 1, and we have $D^*(V=4, T=95\ \mathrm{K}) \sim N_g \times 10^8\ \mathrm{cm\,Hz^{1/2}/W}$. This is roughly the right answer as can be seen from Fig. 4 and Fig. 14. The above result also implies that in devices with weakly bound excited states where $[\nu_{ec}/(\nu_{ec}+\nu_0)^2]^{1/2}$ is indeed roughly constant with temperature, the temperature dependence of the peak detectivity is actually dominated by the exponential (noise related) factor shown in Eq. (25), and not by the temperature dependence of the absorbance (mainly N_g and $\nu_0(T)$). Of course, both effects will pull down D^* with temperature. From this result we can also conclude that the maximum high temperature D^* achievable with these devices is

$$D^* = \left\{ \frac{\nu_{ec}}{[W_{ec}+\nu_0(T)]^2} \right\}^{1/2} \frac{L\alpha(\omega,T)}{(LN_d)^{1/2}\hbar\omega}. \tag{27}$$

This is actually the prefactor in Eq. (25) but is rewritten to exhibit the temperature dependence of the total inverse lifetime at high temperature.

B. Summary of discussion on D^*

From Eq. (27), the high temperature performance of D^* is limited by the magnitude of the absorbance per dot and the lifetime of the excited state. The escape efficiency or ratio of escape rate to total relaxation rate (inverse lifetime) can at best be 1. The detectivity D^* increases ultimately only with the square root of the dot density N_d. It is unlikely that one can improve N_d by more than an order of magnitude. But the oscillator strength could be an order of magnitude better. Altogether a potential increase of a factor 30 over what we have now.

C. Comparison with D_4: 70 layers; shallow bound state

The device with the more weakly bound excited state in Ref. 13 has the best high-T performance. The temperature drop in D^* from 150 K to 200 K is almost exclusively due to the noise current [i.e., the $\exp(E_D/2kT)$ factor]. This proves that $R(T)$ and $\alpha(T)$ can indeed remain stable. We are also dealing with a thick device where $g=1$ in the GR noise formula Eq. (15), so we should be using the form (substituting for W_{ec} and noting that it is greater or approximately equal to ν_0)

$$D^* = \left[Q \frac{\nu_{ec}}{(\nu_0+\nu_{ec})\exp(-E_{ec}/kT)^2} \right]^{1/2} \times \frac{\alpha(\omega,T)}{A^{1/2} N_d \hbar\omega} \left\{ \frac{\mu F}{L C_{be}} \right\}^{1/2} e^{(E_D/2kT - E_{ec}/kT)}, \tag{28}$$

$$D^* = \left[\frac{\mu F}{C_{be}\nu_{ec}} \right]^{1/2} \frac{\alpha(\omega,T)}{N_d^{1/2}\hbar\omega} e^{(E_D/2kT)} \tag{29}$$

with $E_{ec} \sim kT$. The D^* at 140 K is $\sim 10^{11}$ and at 200 K it is $D^* \sim 6\times10^9$, so from Eq. (29), we expect a maximum possible $D^*(T=300\ \mathrm{K})=5\times10^8\ \mathrm{cm\,Hz^{1/2}/W}$.

IX. CONCLUSION AND FUTURE WORK

We have presented data and given a theoretical explanation for the bias and temperature dependence in the observed dark current, responsivity, and detectivity in two categories of recently made high detectivity devices. The cases studied are sufficiently general so that our model covers a large range of possible device types. The present theory covers the ideal case, but is verifiably, already quite close to the final answer. Detailed point-by-point curve fitting of data requires fine details, such as Gaussian convolution due to inhomogeneities, trap distributions, electron correlations, and hot carrier corrections. This is not appropriate in this paper and needs short focused presentations.

Relatively high D^* values have been achieved with relatively low oscillator strengths in D_1, D_2, and D_3. This is because of high gain ($g \sim 1000$) and low noise. The USC group device has a lower noise current than D_1. The results from all four devices investigated suggest that there is hope for band-structure engineering to further improve these values.

A strong bound state gives well-defined, narrow lines. The well-defined excited state is a strongly bound state and has a long "phonon bottleneck lifetime" but only at low temperatures. In the "high voltage range" of interest, at temperatures

of ~77 K, we found that the escape rate was larger than the phonon bottleneck relaxation rate to the ground state. The escape is therefore rate determining at high voltages, but only at low temperatures when the excited state has enough time to escape.

The most important simple observation is that in large diameter, lenslike structures as investigated in our group (D_1 and D_2), the density of QD levels is "high" and the oscillator strengths "QWIP-like" (see Fig. 5 and Fig. 6). Consequently there is no true phonon bottleneck. Furthermore, the photoexcited state is strongly bound and needs bias and temperature to assist the escape out into the continuum. This then implies that the phonon resonance cascades down in energy, which now only need two to three phonons per step, and rapidly becomes faster than the escape rate as we go up in temperature.[38,42] At high doping, the cascade is facilitated by electron-electron lifetime broadening. This considerably reduces the responsivity as we go above ~120 K. In the range 77–100 K, the escape rate enhancement with temperature compensates for the lifetime broadening, giving the illusion of temperature independent responsivity behavior.

When the excited state is closer to the continuum as in device D_4, we have a situation more similar to some QWIPs.[48] The escape is easier, and the lifetime can be escape controlled. However, the oscillator strength is not much higher. The advantage is that, as in QWIPS,[48] one can work at lower biases and avoid field-assisted and/or temperature-assisted trapping losses out and into the ground states. The spectral band in D_4 is broader because the carrier is free to go up in energy,[34] and because of larger (i) size variations in thick layers and (ii) electron correlation broadening with space charge fields.

In conclusion, it seems that for this particular range of wavelengths, we have quite a good scenario with the InGaAs/InGaP, D_1 devices, but unfortunately only up to ~140 K. The oscillator strength and the absorbance for s polarized waves could, we believe, still be improved but the high temperature performance needs a somewhat different strategy. Looking at Fig. 4 and Fig. 5 for D_1, one notes that as in QWIPs, bringing in z polarization admixture while keeping the high gain, would seem like a good option.

We have seen that the high gain is achieved because the capture rate C_{be} is slow (around 10^{10} Hz), the mobility high, and the transit time short. These devices brilliantly exploit the high mobility III-V technology in very thin ($\sim 10^{-6}$ m) systems.

Device D_2 is boxed in by barriers. The design was expected to reach a higher gain and a low dark current. The gain did go up to a peak ~1200, but the overall responsivity in the "voltage of low noise range" of interest did not improve. There were, however, other consequences as well: (i) the escape energy increased to 149 meV necessitating a higher bias for escape and (ii) the ground-to-continuum energy decreased to 382 meV, which implies a higher (dot-to-dot) dark current and thus noise.

Finally we note that for this class of photodetectors, the experimental dark currents and noise levels are reasonably low in the range 77–200 K, both in D_1 and D_2, and this is true also for the USC group devices[10,11] and also true for Ref. 13. The measured responsivity and absorption[33] in the strongly bound wavelength selective devices D_1 and D_2, however, decrease too rapidly with temperatures above 100 K compared to what is needed for high temperature operation. This is a serious problem. The situation is better in the more weakly bound broad band devices of Chakrabarty *et al.*[13] The reason has been given already and is because whereas in the strongly bound excited states the relaxation rate increases faster with temperature than the escape rate, in the weakly bound devices, the escape rate remains at least as fast, and is lifetime determining. This has also been demonstrated in QWIPS[48] where the responsivity stays high up to room temperature. As in the work of Kim *et al.*,[43] a possible way forward for QDIPS would be, therefore, to introduce an adjacent capture layer for the carrier in the photoexcited state. This would, with the help of external or internal fields, block the return down via the phonon cascade.

We have identified a number of interesting and important problems related to QDIP performance, which have to be solved before serious progress in device performance can be made.

ACKNOWLEDGMENTS

The authors are grateful to Dr. R. Grober from Yale University and J. Jiang for useful discussions.

APPENDIX A

In these high mobility devices, the lifetime is capture limited. The matrix element giving the capture rate from a state k in a single phonon emission to any quantum dot in the device can be written

$$C_{be}(k) = \frac{2\pi}{\hbar\Omega} N_d \sum_q \left\| \int dr e^{ikr} V_{ep}(q) \Psi_e(r) \right\|^2 \delta(E_k - E_e - \hbar\omega_q) \times (1 + n_q). \tag{A1}$$

Airy functions allow proper counting and evaluation of the escape sum, Airy oscillations have been neglected in matrix elements. In Eq. (A1), n_q is the phonon distribution, V_{ep} the coupling, Ψ_e the normalized wave function of the excited state, and Ω is the volume. The acoustic phonon coupling is of the form (absorption term)

$$V_{ep}(\vec{q}) = e^{iq\cdot r} \frac{D^2}{2\rho_d V c_s^2} \hbar\omega(\vec{q}), \tag{A2}$$

where D is the deformation potential (~6.2 eV), c_s is the velocity of sound, ρ_d is the density, and V the volume. The unscreened LO-optic coupling is of the form (cgs) units

$$|V_{ep}|^2 = \frac{2\pi\hbar\omega_{LO}e^2}{Vq^2}\left(\frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_s}\right), \tag{A3}$$

where the $\varepsilon_s\varepsilon_\infty$ are static and high frequency dielectric permittivities.

If the energy difference is outside the range of a single optic or acoustic mode, then a many phonon process must be considered.[36] An analytical calculation using a simple hydrogenic wave fuction gives an intuitively very attractive result,

which says that the *LO* resonant rate of capture from a state k is essentially just $C_{be} \sim f(k)\omega_0(a^3/\Omega_d)$ where Ω_d is the dot volume and $f(k)$ is the Fermi function. In other words Eq. (A1) says that the rate is just the phonon frequency times the probability of finding the carrier on any site in the dot, i.e., $\sim a^3 \Psi^*_e \Psi_e$, this is $\sim 10^9$ Hz.

APPENDIX B: MANY-BODY ASSISTED DETRAPPING

The simplest estimate is obtained by assuming that the charge in the QD experiences a time fluctuating field caused by carriers passing by. The detrapping rate from e is

$$\nu_{e\text{-}e} = \left(\frac{\Delta V_{ee}}{\hbar}\right)^2 \frac{\nu_I}{\nu_I^2 + E_{ec}^2/\hbar^2}, \tag{B1}$$

where $\Delta V_{ee} \sim 5$ meV, $E_{ec} \sim 136$ meV, and $\nu_I = \mu F/d_e$ is the frequency of the current with d_e denoting the average carrier distance $d_e \sim 50$ nm. When $F = 2.10^6$ V/m and $\mu = 1$ m^2/V s, we have ν_{ee} in the range 10^{10}–10^{11} Hz, which agrees with the simulation estimates.[16]

*Corresponding author. Email address: razeghi@ece.northwestern.edu

[1] B. F. Levine, A. Zusman, S. D. Gunapala, M. T. Asom, J. M. Kuo, and W. S. Hobson, J. Appl. Phys. **72**, 4429 (1992).
[2] C. Gmachl, F. Capasso, D. L. Sivco, and A. Y. Cho, Rep. Prog. Phys. **64**, 1533 (2001).
[3] J. Urayama, T. B. Norris, B. Kochman, J. Singh, and P. Bhattacharya, Appl. Phys. Lett. **76**, 2394 (2000).
[4] A. Rogalski, J. Appl. Phys. **93**, 4356 (2003).
[5] A. Kosterev and F. K. Tittel, IEEE J. Quantum Electron. **38**, 582 (2002).
[6] J. Jiang, S. Tsao, T. O'Sullivan, W. Zhang, H. Lim, T. Sills, K. Mi, M. Razeghi, G. J. Brown, and M. Z. Tidrow, Appl. Phys. Lett. **84**, 2166 (2004).
[7] J. Jiang, K. Mi, S. Tsao, W. Zhang, H. Lim, T. O'Sullivan, T. Sills, M. Razeghi, G. J. Brown, and M. Z. Tidrow, Appl. Phys. Lett. **84**, 2232 (2004).
[8] H. Mohseni, M. Razeghi, and G. J. Brown, Appl. Phys. Lett. **78**, 2107 (2001).
[9] H. Mohseni, Y. Wei, and M. Razeghi, Proc. SPIE **4288**, 191 (2001); Y. Wei, A. Gin, M. Razeghi, and G. J. Brown, Appl. Phys. Lett. **80**, 3262 (2002); **81**, 3675 (2002).
[10] E. Tae Kim, A. Madhukar, Z. Ye, and J. C. Campbell, Appl. Phys. Lett. **84**, 3277 (2004).
[11] Z. Ye, J. C. Campbell, Z. Chen, E. Tae Kim, and A. Madhukar, Appl. Phys. Lett. **83**, 1234 (2003).
[12] S. Chakrabarti, A. D. Stiff-Roberts, P. Bhattacharya, S. Gunapala, S. Bandara, S. B. Rafol, and S. W. Kennedy, IEEE Photonics Technol. Lett. **16**, 1361 (2004).
[13] S. Chakrabarti, A. D. Stiff-Roberts, and P. Bhattacharya, IEEE Photonics Technol. Lett. **16**, 867 (2004).
[14] L. Jiang, S. Li, Nien-Tze Yeh, Jen-Inn Chyi, C. E. Ross, and K. S. Jones, Appl. Phys. Lett. **82**, 1986 (2003).
[15] J. Phillips, J. Appl. Phys. **91**, 4590 (2002).
[16] B. Kochman, A. Stiff-Roberts, S. Chakrabarti, and J. D. Phillips, IEEE J. Quantum Electron. **39**, 459 (2003).
[17] J. Z. Zhang and I. Galbraith, Appl. Phys. Lett. **84**, 1934 (2004).
[18] V. Ryzhii, Jpn. J. Appl. Phys., Part 2 **40**, L148 (2001); V. Ryzhii, I. Khmyrova, and V Mitrin, Semicond. Sci. Technol. **19**, 8 (2004).
[19] B. Jean and T. Bende, in *Topics in Applied Physics*, edited by I. T. Sorokina and K. L. Vodopyanov (Springer, Berlin, 2003), Vol. 89, 511.
[20] S. Tsao, A. V. Gin, K. Mi, J. Szafraniec, W. Zhang, H. Lim, T. O'Sullivan, J. Jiang, M. Rzaeghi, G. J. Brown, and M. Z. Tidrow, in *Proceedings of SPIE, Denver, 2004*, edited by E. L. Dereniak, R. E. Sampson, and C. B.1 Johnson (SPIE, Bellingham, 2004), p. 74.
[21] M. Razeghi, H. Lim, S. Tsao, J. Szafraniec, W. Zhang, K. Mi, and B. Movaghar, Nanotechnology **16**, 219 (2005).
[22] D. Gershoni, H. Temkin, G. J. Dolan, J. Duinsmuir, S. N. G. Chu, and M. B. Panish, Appl. Phys. Lett. **53**, 995 (1988).
[23] M. Califano and P. Harrison, J. Appl. Phys. **91**, 389 (2002).
[24] S. L. Chuang, *Physics of Optoelectronic Devices* (Wiley-Interscience, Hoboken, 1995).
[25] L. R. C. Fonseca, J. L. Jimenez, J. P. Leburton, R. M. Martin, Phys. Rev. B **57**, 4017 (1998).
[26] M. Roy and P. A. Maksym, Phys. Rev. B **68**, 235308 (2003).
[27] M. Califano and P. Harrison, Phys. Rev. B **61**, 10959 (2000).
[28] David M.-T. Kuo and Y. C. Chang, Phys. Rev. B **67**, 035313 (2003).
[29] E. P. Pokatilov, V. A. Fonoberov, V. M. Fomin, and J. T. Devreese, Phys. Rev. B **64**, 245328 (2001).
[30] H. Jiang and J. Singh, Appl. Phys. Lett. **71**, 3239 (1997).
[31] A. J. Williamson, L. W. Wang, and A. Zunger, Phys. Rev. B **62**, 12963 (2000).
[32] M. Razeghi, *The MOCVD Challenge* (Institute of Physics Publishing, Philadelphia, 1995), Vol. 2.
[33] F. Bras, P. Boucaud, S. Sauvage, G. Fishman, and J.-M. Gérard, Appl. Phys. Lett. **80**, 4620 (2002); S. Sauvage, P. Boucaud, F. H. Julien, J.-M. Gérard, and V. Thierry-Mieg, Appl. Phys. Lett. **71**, 2785 (1997).
[34] B. Movaghar and J. Leo, J. Appl. Phys. **65**, 5019 (1989).
[35] B. Movaghar, Semicond. Sci. Technol. **2**, 185 (1987); **4**, 91 (1989); J. Mol. Electron. **3**, 183 (1986); **4**, 79 (1987).
[36] B. K. Ridley, *Quantum Processes in Semiconductors* (Oxford University Press, Oxford 1993).
[37] M. Fricke, A. Lorke, J. P. Kotthaus, G. Medeiros-Ribeiro, and P. M. Petroff, Europhys. Lett. **36**, 197 (1996).
[38] S. Hameau, Y. Guldner, O. Verzelen, R. Ferreira, G. Bastard, J. Zeman, A. Lemaitre, and J. M. Gérard, Phys. Rev. Lett. **83**, 4152 (1999).
[39] David M.-T. Kuo and Y. C. Chang, Phys. Rev. B **66**, 085311 (2002).
[40] J. Urayama, T. B. Norris, J. Singh, and P. Bhattacharya, Phys. Rev. Lett. **86**, 4930 (2001).
[41] T. Inoshita and H. Sakaki, Phys. Rev. B **56**, R4355 (1997).
[42] T. Inoshita and H. Sakaki, Phys. Rev. B **46**, 7260 (1992).

[43] Jong-Wook Kim, Jae-Eung Oh, Seong-Chul Hong, Chung-Hoon Park, and Tae-Kyung Yoo, IEEE Electron Device Lett. **21**, 329 (2000).
[44] R. Grober (private communication).
[45] S. L. Feng, J. Jrynicki, V. Donchev, and J. C. Bourgoin, Semicond. Sci. Technol. **8**, 2092 (1993).
[46] M. Hanke, D. Grigoriev, M. Schmidbauer, P. Schafer, R. Kohler, R. L. Sellin, U. W. Pohl, and D. Bimberg, Appl. Phys. Lett. **85**, 3062 (2004).
[47] P. W. Fry, I. E. Itskevich, D. J. Mowbray, M. S. Skolnick, J. J. Finley, J. A. Barker, E. P. O'Reilly, L. R. Wilson, I. A. Larkin, P. A. Maksym, M. Hopkinson, M. Al-Khafaji, J. P. R. David, A. G. Cullis, G. Hill, and J. C. Clark, Phys. Rev. Lett. **84**, 733 (2000).
[48] H. C. Liu, R. Dudek, A. Shen, E. Dupont, C.-Y. Song, Z. R. Wasilewski, and M. Buchanan, Int. J. High Speed Electron. Syst. **12**, 803 (2002).