Mobility in gated GaN_xAs_{1-x} heterostructures as a probe of nitrogen-related electronic states

J. Buckeridge[1] and S. Fahy[2,3]

[1]*University College London, Kathleen Lonsdale Materials Chemistry, Department of Chemistry, 20 Gordon Street, London WC1H 0AJ, United Kingdom*

[2]*Tyndall National Institute, Lee Maltings, Cork, Ireland*

[3]*Department of Physics, University College Cork, Ireland*

(Received 19 August 2011; published 31 October 2011)

Calculations are presented of *n*-type carrier mobility in a gated, double-quantum-well InGaAs/GaN_xAs_{1-x} heterostructure (with $x \approx 0.01$), where transport occurs in the lowest two subbands (one primarily in the InGaAs layer and the other primarily in the dilute nitride layer) and carrier confinement is provided by surrounding AlGaAs layers and gating fields from front and back gates. Such a device allows a controlled interaction between carriers and localized nitrogen cluster states near the conduction band edge in the dilute nitride. Varying the In composition in the GaInAs layer and the gating fields is predicted to alter the Fermi level, relative to the localized nitrogen cluster states, and cause a strong modulation of the *n*-type mobility. For appropriately designed heterostructures, the mobility is predicted to vary by up to a factor of three when the Fermi level is resonant with nitrogen cluster electronic states. Thus, the mobility measurements in such a heterostructure can be used as a spectroscopic probe of the localized states in the dilute nitride layer.

DOI: 10.1103/PhysRevB.84.144120 PACS number(s): 71.55.Eq, 72.80.Ng, 72.20.Dp, 72.80.Ey

I. INTRODUCTION

Dilute nitride alloys have attracted considerable attention in recent years, both because of their unusual physical properties and their potential device applications. When a small fraction of the As atoms in GaAs or InGaAs are replaced by N, the energy gap of the material decreases rapidly, for example, by 150 meV when the N concentration is 1%.[1] This is in marked contrast to the usual alloying effect, whereby the bandgap of the alloy $A_{1-x}B_xC$ changes almost linearly in x from that of AC to BC as x increases from 0 to 1; in this conventional alloy behavior, one would therefore expect the addition of N to GaAs to *increase* the band gap from the GaAs value (1.5 eV) to the GaN value (3.4 eV). The reduction in band gap makes InGaNAs, in which the addition of In and N in a 3:1 ratio allows lattice matching to GaAs, one of the more promising materials for long-wavelength (1.3 and 1.5 μm) telecommunications lasers based on a GaAs substrate[2,3] and for extending the wavelength range of GaAs-based solar cells further into the infrared.[4] However, the addition of these small concentrations of nitrogen to GaAs has also been found to cause a drastic reduction in *n*-type carrier mobility.[5–8] For the ternary GaN_xAs_{1-x} alloys, *n*-type carrier mobilities of about 200 $cm^2\ (Vs)^{-1}$ are usually found, and mobilities for the quaternary alloy InGaNAs—typically less than 200 $cm^2\ (Vs)^{-1}$—are unacceptably low for most applications at present.

The dramatic effect of substitutional nitrogen on the band gap of GaAs can be understood using the band anticrossing (BAC) model developed by Shan *et al.*[9] They used hydrostatic pressure techniques to demonstrate that the reduction in band gap can be understood as due to an interaction between the conduction band edge (CBE) and a higher-lying set of localized nitrogen resonant states. In the ideal case, N substitutes onto As sites in GaAs or InGaAs. Although N is isoelectronic with As, it is much smaller and more strongly electronegative. This leads to the formation of quasilocalized electron states associated with the substitutional N (Refs. 10 and 11), which, through strong interaction with the GaAs CBE, lead to a drastic reduction in band gap for a N concentration of a few percent,[1,9,12–14] a marked increase in CBE effective mass with a nonmonotonic composition dependence,[15–17] and a dramatic increase in carrier scattering with an associated reduction in *n*-type carrier mobility in the materials.[5–8,18–21]

In a previous work,[18] we described how the composition dependence of the CBE (as determined using the BAC model, including interactions with isolated substitutional N levels only) is closely related to the *n*-type carrier scattering cross section in the ultradilute alloy and sets a limit on the carrier mobility in dilute nitride materials. This limit was $\sim$1000 $cm^2\ (Vs)^{-1}$, roughly an order of magnitude larger than that seen in experiment. In a more complete study,[22] the interaction of the CBE with both single N substitutional states and cluster states associated with complexes of up to four N atoms, was used to determine an energy-dependent scattering rate. The formation of N clusters at random in the dilute alloy occurs because GaN_xAs_{1-x} samples are grown under conditions far from thermodynamic equilibrium, giving an approximately random distribution of N on the group V sites. The mobility in Ref. 22 was calculated through solution of the Boltzmann transport equation in the relaxation time approximation and the resulting room-temperature mobilities of less than 300 $cm^2\ (Vs)^{-1}$ for N concentrations of 1%–2% were in good agreement with experiment. In those transport calculations, N cluster states dominated the scattering near the CBE and played a crucial role in limiting the carrier mobility. These cluster states have also been used to account for the nonmonotonic composition dependence of the electron effective mass in dilute nitrides.[15,16] The values of the mobility and effective mass, which have been calculated on the basis of the localized cluster state picture, provide strong indirect support for the existence of localized and quasilocalized states near the CBE, as do the magnetotunneling spectroscopy measurements of Endicott *et al.*[23] However, a more direct verification of their existence has been elusive.

 ©2011 American Physical Society

In the present work we describe a method of probing N cluster states near the CBE using mobility measurements in a proposed gated double-well heterostructure. The structure would consist of AlGaAs/Ga(In)As/GaNAs/AlGaAs layers, as discussed in detail below, with both a front and a back gate. From the model of dilute nitride localized state scattering used in previous mobility calculations[22] for bulk materials, we predict the carrier mobilities in such structures and find that the effects of specific localized states near the CBE can be resolved. Due to the drastically lowered electron mobility in the dilute nitrides, the cluster states are extremely difficult to probe in conventional low-temperature transport measurements for bulk or single-well heterostructures. However, we propose a heterostructure, in which the carriers are largely contained in the Ga(In)As layer, so that their electronic wave function has only a small overlap with N states in the dilute nitride layer. By varying the back-gate voltage, the Fermi level can be swept through a portion of the spectrum of localized N levels in the dilute nitride layer, with large reductions in carrier mobility predicted when the Fermi level is resonant with a localized level. Due to the practical limitations of electrical breakdown in the devices,[24] we assume variations in the back-gate field of the order of 10^4 V cm^{-1}. We show how this limit, together with device parameters such as layer thicknesses and alloy compositions, can be used to estimate which portions of the spectrum of localized levels can be probed.

Our calculations assume heterostructures in which the dilute nitride layer thickness is approximately 3 nm, for which we assume a bulklike distribution of localized state energy levels. We use the spectra of localized levels for various N concentrations, calculated by Lindsay *et al.*[15,16,22,25,26] using parameterized tight-binding methods in the linear combination of isolated nitrogen states (LCINS) approach, as the spectra to be probed by our device. We use a modified Poisson-Schrödinger solver to determine the transverse electronic wave functions in the heterostructure. We then use the transverse wave functions to determine the alloy scattering rates, generalizing the quantum-well scattering approach described in Ref. 22 to calculate the N alloy scattering rate. The mobility is determined by solution of the Boltzmann equation in the relaxation time approximation.

We predict sharp reductions (∼65%) in the low-temperature (4 K) mobility as the back-gate field is varied for samples with N concentration $x = 0.36\%$ in the dilute nitride layer. For higher N concentrations ($x = 1.2\%$) we also find variations in carrier mobility. However, at higher N concentrations either higher back-gate fields or higher In contents are required to confine the carriers in the InGaAs layer. This means that the overlap of the wave function of the carriers in the InGaAs layer with the dilute nitride layer is lower, leading to lower variations in mobility when the Fermi level (E_F) is resonant with a localized state.

The rest of the paper is organized as follows: In Sec. II we describe in detail our proposed heterostructure device; in Sec. III we describe the method of calculating the transverse wave functions; in Sec. IV we describe how the mobility is determined, incorporating different scattering mechanisms for different layers; in Sec. V we present our mobility calculation results; and in Sec. VI we summarize the main points of our study. Details of the method used to find the N alloy scattering rate in the double-well heterostructure, how the sub-band occupations vary with gate field, which design parameters remain constant in our calculations and why, and the effect of varying those parameters other than the N content that do not remain constant [the In content and Ga(In)As layer thickness] are presented in Appendixes A–D.

II. DESCRIPTION OF HETEROSTRUCTURE

A schematic of our proposed heterostructure device is shown in Fig. 1. The AlGaAs layers have Al content of 40%, and the Si doping density in the doped layer is 10^{18} cm^{-3}. (The 40% Al content is chosen to confine the carriers in the channel. Reducing this value to 30% will have little effect on our results. This point is discussed in Appendix C.) For simplicity all donors are assumed to be fully ionized. The electric field due to the ionized dopant layer creates a triangular potential well in the InGaAs layer and gives rise to a carrier subband concentrated there. The band-gap reduction due to N in the dilute nitride layer creates a second quantum-well subband, concentrated in the dilute nitride. The relative energy of the two subbands can be varied by altering the gating fields. In and N content and the layer thicknesses are taken to be

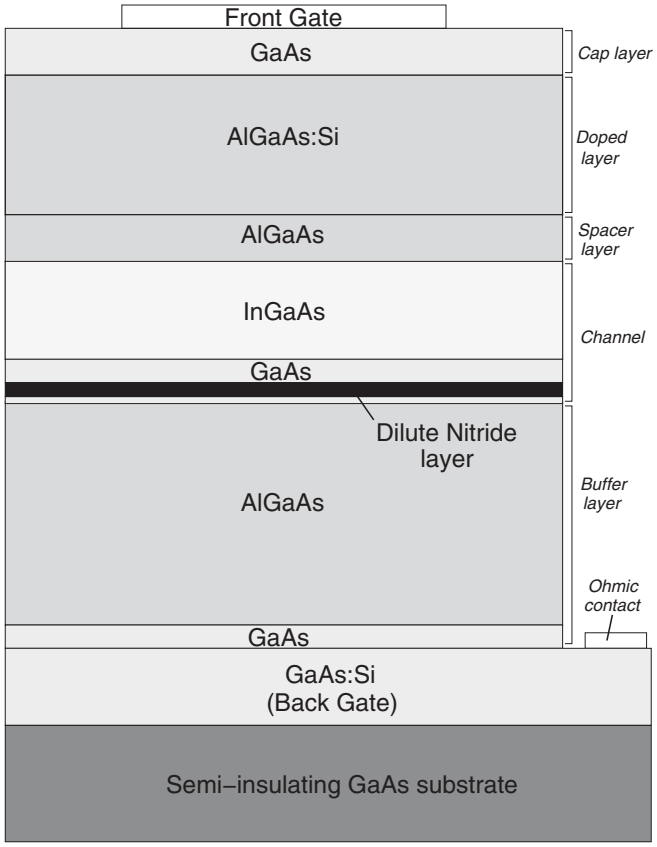

FIG. 1. Schematic of heterostructure device, including front and back gates. The back-gate geometry shown here is an example of how back-gating may be acheived in real devices; the details of the actual structure of the back gate do not affect our calculations. The AlGaAs layers have Al content 40% and the doped layer has a Si concentration of 10^{18} cm^{-3}. Layer thicknesses, In content in the InGaAs layer and N content in the dilute nitride layer, and front- and back-gate fields are variable in our calculations (see text for details).

variable aspects of the device design, their values determining the occupation of the subbands and how they respond to gating fields, essentially determining which sections of the spectrum of localized N cluster states can be probed by varying the gating fields.

The heterostructure is grown on a semi-insulating GaAs substrate (taken to be a pure insulator), with a buried back gate. In our schematic (Fig. 1) we show a back-gate geometry consisting of a heavily Si-doped GaAs layer grown on the GaAs substrate. However, the details of the design of the back gate do not affect our calculations; we assume the presence of a surface charge density at the top of the back gate in our calculations; that is, we do not directly model the various layers composing the back gate. We show them in our schematic for illustrative purposes as an example of how the back gate may be structured in real devices. Grown on this is the buffer region, consisting of a relatively thick (approximately 50 nm) layer of undoped $Al_{0.4}Ga_{0.6}As$ grown on a thin GaAs spacer layer (typically 5 nm in thickness). The buffer layer provides a barrier between the channel and the back gate. Above the buffer layer is the channel region. This consists of a thin ($\sim$1 nm) GaAs layer at the bottom, with the dilute nitride layer grown on top of it. The dilute nitride layer consists of GaN_xAs_{1-x} with $x \leqslant 2\%$ and is approximately 3 nm thick. A layer of GaAs is grown on top of it (about 5 nm thick), and a layer of InGaAs (10–25 nm thick, with a variable In content not larger than 10%, which ensures that the layer thickness is less than the critical thickness[27]) is then grown on top of the GaAs layer, completing the channel. Grown on top of the channel is an $Al_{0.4}Ga_{0.6}As$ spacer layer (about 10 nm thick), with a Si-doped $Al_{0.4}Ga_{0.6}As$ layer grown on top of this (typical thicknesses of this are 25–30 nm). Grown on top of this is an undoped GaAs cap (about 10 nm thick) with the metal front gate completing the proposed heterostructure.

We summarize the material parameters used in our heterostructure simulation in Table I. Energies are given with reference to the GaAs valence band maximum (VBM). The band gap and dielectric constant variation in x for $Al_xGa_{1-x}As$ are taken from Ref. 28, with the conduction band discontinuity between AlGaAs and GaAs, $\Delta E_{C,Al}$, given by[29] $\Delta E_{C,Al} = 0.62 \times [E_{AlGaAs} - E_{GaAs}]$, where E_{GaAs} is the temperature-dependent GaAs band gap [see Eq. (1)] and E_{AlGaAs} is the AlGaAs band gap. Theoretical studies[30] have predicted an increase in dielectric constant with x in GaN_xAs_{1-x}. However, we take a value of 10.6 as the dielectric constant, which is a recent experimental result[31] for $x = 0.01$. Because the dilute nitride layer in our calculations is very thin, we neglect the variation of this dielectric constant with x. The band gap and dielectric constant variation in x for $In_xGa_{1-x}As$ are taken from Ref. 32, with the conduction band discontinuity between InGaAs and GaAs, $\Delta E_{C,In}$, given by[33] $\Delta E_{C,In} = 0.9 \times [E_{GaAs} - E_{InGaAs}]$, where E_{InGaAs} is the InGaAs band gap. The temperature-dependent GaAs band gap E_{GaAs} is given by[34]

$$E_{GaAs}(T) = E_{GaAs}(0) - \frac{\alpha_T T^2}{T + \beta_T}, \tag{1}$$

with $E_{GaAs}(0) = 1.519$ eV, $\alpha_T = 5.405 \times 10^{-4}$ eV K^{-1}, and $\beta_T = 204$ K. The dielectric constant of GaAs is taken to be $\epsilon_{GaAs} = 12.9$.[35]

The thickness d of the doped layer, in combination with the doping density N_D, the charge per unit area on the front gate $\epsilon_{GaAs}\epsilon_0\mathcal{E}_{FG}$ (where ϵ_0 is the vacuum permittivity and $\mathcal{E}_{FG}$ is the front-gate electric field), the charge per unit area on the back gate $\epsilon_{GaAs}\epsilon_0\mathcal{E}_{BG}$ (where $\mathcal{E}_{BG}$ is the back-gate electric field), and the constraint that the device be overall charge neutral, determines n_{ch}, the total charge per unit area in the channel, as

$$-en_{ch} = \epsilon_{GaAs}\epsilon_0 (\mathcal{E}_{FG} + \mathcal{E}_{BG}) - eN_D d. \tag{2}$$

Here e is the electronic charge.

Our heterostructure is designed so that the charge density in the two lowest-energy subbands is confined, either in the InGaAs layer or in the dilute nitride layer; that is, one band is concentrated in the dilute nitride layer, the other in the InGaAs layer near the AlGaAs interface. (For very low N content, $x \leqslant 0.1\%$, in the dilute nitride, this does not strictly hold because the quantum confinement of the subband in the dilute nitride layer is weak and a significant portion of its charge lies in the InGaAs layer.) Which subband is occupied, or in what proportion the total occupation is spread over the two subbands, is controlled by the front- and back-gate fields.

Higher-order subbands $\psi_3(z)$, etc. [where $\psi_i(z)$ is the transverse wave function of the ith subband], may become occupied at high temperatures or at large values of n_{ch}. This is undesirable, so in our calculations we keep temperature T low (4 K) and we limit the ranges of $\mathcal{E}_{FG}$ and $\mathcal{E}_{BG}$ used in our calculations appropriately. We limit $\mathcal{E}_{BG}$ to values of the order of 10^4 V cm^{-1} (except for the 1.2% N concentration case, where we expand the range of $\mathcal{E}_{BG}$ for illustrative purposes), as large fields on back gates are difficult to achieve in practice.[24] We limit $\mathcal{E}_{FG}$ to the range 3×10^5 to 4×10^5 V cm^{-1}. For $\mathcal{E}_{FG} > 4 \times 10^5$ V cm^{-1}, n_{ch} is quite low and can reach zero (meaning the device is off). For $\mathcal{E}_{FG} < 3 \times 10^5$ V cm^{-1}, we have two problems: n_{ch} can become large enough that higher-order subbands become occupied; also at low enough values of $\mathcal{E}_{FG}$ we find that the CBE in the AlGaAs-doped layer can become close enough to the Fermi level in the channel that DX centers, induced by the Si dopants, may become occupied. In our simulation we assume all donors to be fully ionized; therefore, we must avoid values of $\mathcal{E}_{FG}$ that bring the CBE in the AlGaAs-doped layer within 0.13 eV of the Fermi level, where 0.13 eV is taken to be the binding energy of DX^0 centers.[36] Keeping $\mathcal{E}_{FG} > 3 \times 10^5$ V cm^{-1} allows this condition to be easily met, given our device architecture.

TABLE I. Material parameters of $In_xGa_{1-x}As$ and $Al_xGa_{1-x}As$ used in simulating the heterostructure device. E_{GaAs} [see Eq. (1) in the text] is the temperature-dependent GaAs band gap and $\epsilon_{GaAs} = 12.9$ is the GaAs dielectric constant. GaNAs is taken to have the same band gap as GaAs in our calculation because the band gap bowing due to N is represented explicitly using the BAC model (see text for details).

Material	Energy band gap (eV)	Dielectric constant
GaNAs	E_{GaAs}	10.6
$In_xGa_{1-x}As$	$E_{GaAs} - 1.593x + 0.493x^2$	$\epsilon_{GaAs} + 2.15x$
$Al_xGa_{1-x}As$	$E_{GaAs} + 1.2475x$	$\epsilon_{GaAs} - 2.84x$

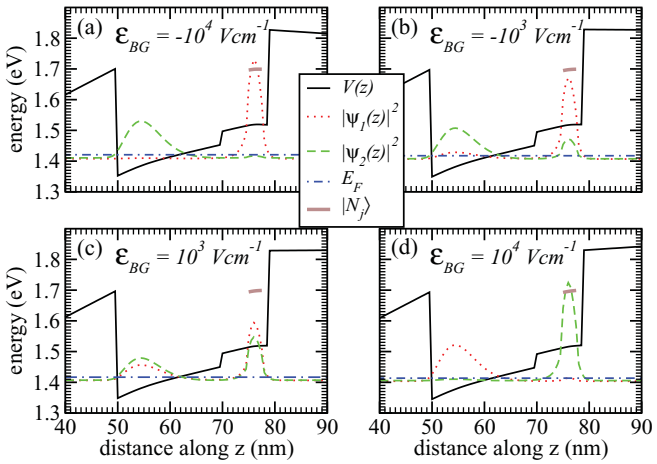

FIG. 2. (Color online) Subband weights $|\psi_i(z)|^2$ for $i = 1$ (red dotted lines), $i = 2$ (green dashed lines), and 1D heterostructure potential $V(z)$ (black solid lines) as a function of depth z in the heterostructure, with $z = 0$ at the top of the device, for various values of the back-gate field $\mathcal{E}_{\rm BG}$. The Fermi level, E_F, is indicated by the (blue) dot-dashed line. The localized N states $|N_j\rangle$ in the dilute nitride layer are indicated by the thick (brown) line. The $|\psi_i(z)|^2$ are plotted at their subband energies E_i. The GaAs VBM is used as the zero of energy. Results are for a 20-nm InGaAs layer (In content 3%), 5-nm GaAs layer, and 3-nm dilute nitride layer (N content 1.2%) in the channel and a constant front-gate field $\mathcal{E}_{\rm FG} = 3.4 \times 10^5$ V cm^{-1}. AlGaAs/InGaAs interface at the top of the channel is at $z = 50$ nm and the GaAs/AlGaAs interface at the bottom of the channel is at $z = 79$ nm. Temperature, $T = 4$ K.

By tuning the gate fields it is possible to bring the energies of the two subbands closer together until they are resonant and hybridize strongly and, by further variation, cause the lowest-energy subband to swap from mainly occupying one of the dilute nitride or InGaAs layers to mainly occupying the other. The swapping of the lowest-energy subband from the dilute nitride layer to the InGaAs layer is illustrated in Fig. 2, with the corresponding subband occupations summarized in Table II. Here $|\psi_i(z)|^2$ for $i = 1,2$ are shown as a function of z for a heterostructure with a 3-nm dilute nitride layer (with a N concentration of 1.2%), a 20-nm InGaAs layer (with In content of 3%), and a 5-nm GaAs layer in the channel, with a constant $\mathcal{E}_{\rm FG} = 3.4 \times 10^5$ V cm^{-1} for different values of $\mathcal{E}_{\rm BG}$ ranging from -10^4 to $+10^4$ V cm^{-1}. The cap, doped, and spacer layers are 10, 30, and 10 nm, respectively, in width and $N_D = 10^{18}$ cm^{-3}. The buffer layer is 50 nm thick. The black line indicates the 1D potential $V(z)$ and the $|\psi_i(z)|^2$ are plotted at their energy levels E_i. The zero of the z axis is at the top of the device. This means that the interface between the AlGaAs spacer layer and the InGaAs layer is at $z =$ 50 nm, and the dilute nitride layer is at $75 \leqslant z \leqslant 78$ (nm). The AlGaAs buffer layer is at $z \geqslant 79$ nm. The temperature of the device is 4 K.

With $\mathcal{E}_{\rm BG} = -10^4$ V cm^{-1} [Fig. 2(a)] the lowest-energy subband is largely located in the dilute nitride layer, with the majority of carriers occupying this subband ($n_{\rm ch} = 6.47$ cm^{-2}, $n_1 = 3.60$ cm^{-2}, $n_2 = 2.87$ cm^{-2}). As we increase the back-gate field through -10^3 V cm^{-1} [Fig. 2(b)], 10^3 V cm^{-1} [Fig. 2(c)] we see a strong mixing of the subbands and a swap over between the positions of the lowest-energy and second-lowest-energy subbands. With $\mathcal{E}_{\rm BG} = 10^4$ V cm^{-1} [Fig. 2(d)] the subbands have completely swapped and the lowest energy subband is now located in the InGaAs layer, with a small tail extending into the dilute nitride layer. The majority of carriers now occupy the InGaAs layer ($n_{\rm ch} = 5.05$ cm^{-2}, $n_1 = 2.80\,{\rm cm}^{-2}, n_2 = 2.25\,{\rm cm}^{-2}$). By increasing $\mathcal{E}_{\rm BG}$ further it is possible to completely empty the dilute nitride layer subband of carriers, so that all the carriers in the channel occupy the InGaAs layer.

TABLE II. Total number of carriers per unit area $n_{\rm ch}$, occupation of lowest subband n_1, and occupation of second-lowest subband n_2 for Figs. 2(a) to 2(d).

Fig. 2	$n_{\rm ch}$ (10^{11} cm^{-2})	n_1 (10^{11} cm^{-2})	n_2 (10^{11} cm^{-2})
(a)	6.47	3.60	2.87
(b)	5.83	3.04	2.79
(c)	5.69	2.94	2.75
(d)	5.05	2.80	2.25

To probe the localized N cluster-related states thought to lie close in energy to the CBE in dilute nitride alloys requires: (i) that we have carriers occupying the subband in the triangular InGaAs quantum well only, with the electron chemical potential of the carriers close in energy to the energies of the localized states to be probed; (ii) that the subband transverse wave function has sufficient overlap with the dilute nitride layer, so that the carriers are scattered by the localized states and we see associated dips in the mobility, but not to such a degree that the mobility of the carriers is drastically reduced; and (iii) that a sufficient range in localized state energy is scanned through as the back-gate field is varied in the range -10^4 to 10^4 V cm^{-1}. When in operation the device will, in general, have one or more subbands occupied. However, by varying the gating fields it is possible to control the occupations of the subbands, so that higher-order subbands can be emptied until only the InGaAs layer subband is occupied. In Sec. V we describe heterostructure architectures and gate fields in which (i) to (iii) are satisfied.

III. DETERMINATION OF TRANSVERSE WAVE FUNCTIONS

The transverse wave functions are determined ultimately through the solution of one-dimensional (1D) equations. We discuss first the representation of the conduction band states in our heterostructure, then discuss the 1D form of the Poisson and Schrödinger equations and their solution.

We represent our Hamiltonian in a plane-wave basis within an effective mass approximation. The effect of nitrogen in the dilute nitride layer is represented by placing a certain concentration x of localized N states $|N_j\rangle$ at an energy $E^0_{{\rm N},j} = 0.18$ eV above the GaAs CBE and incorporating their interaction with the GaAs conduction band using the BAC model.[9] Here the localized states are due to isolated substitutional N atoms and at this stage we ignore the N cluster states, which play such a crucial role in limiting the mobility in dilute nitrides.[22] The isolated substitutional N localized states principally determine the band-edge energy in the dilute

nitrides because the density of isolated N states is much larger than the density of N cluster states in the material.

The area and thickness of the entire heterostructure are A and L, respectively. We begin by considering the entire wave function $\Psi(x,y,z)$. Later we use the translation invariance of the heterostructure along the surface (x and y) directions to write

$$\Psi(x,y,z) = \frac{1}{\sqrt{A}} e^{i(k_x x + k_y y)} \psi(z), \tag{3}$$

where $\psi(z)$ is the transverse wave function. This allows us to perform the calculation through solution of 1D equations only, as $\psi(z)$ depends only on the transverse coordinate z, perpendicular to the surface. The normalization condition on the transverse wave function is

$$\int_0^L |\psi(z)|^2 dz = 1. \tag{4}$$

We specify the structure as being composed of N_s very thin slices, each slice having a chemical composition of either GaAs, $Al_{0.4}Ga_{0.6}As$, $In_yGa_{1-y}As$, or $GaAs_{1-x}N_x$. In physical terms, each slice corresponds to an atomic layer or a few atomic layers. In numerical terms, the slices define the numerical grid on which our transverse wave functions are specified and are typically of the order of 0.5 nm thick; one grid point z_i lies at the center of each slice.

We determine the electron states as linear combinations of effective mass momentum eigenstates $|k\rangle$ (the GaAs, AlGaAs, or InGaAs conduction band states) and localized N states $|N_j\rangle$, located at positions $\mathbf{R}_j$ (the LCINS representation[15,22,25,26]). In layers that have no N present, the state is completely specified in the effective mass representation. To give the complete wave function, conduction band (effective mass) components and N localized state components must be given. The N states are distributed randomly in a layer, usually consisting of several slices. We define the effective mass position eigenstate as

$$|\mathbf{r}\rangle \equiv \frac{1}{\sqrt{\Omega}} \sum_k \exp[-i\mathbf{k}\cdot\mathbf{r}]|k\rangle, \tag{5}$$

where $\Omega = AL$ is the volume of the system. When dealing with 3D wave functions, we sum k over the vectors (k_x,k_y,k_z). When dealing only with the transverse wave function, $k = k_z$ and the coordinate $\mathbf{r}$ is simply z.

The single-particle Hamiltonian in the Poisson-Schrödinger approach can then be written in the form

$$H = H_k + H_N + H_{kN}. \tag{6}$$

Here H_k is the effective mass Hamiltonian,

$$H_k = \sum_k \frac{\hbar^2 k^2}{2m^*} |k\rangle\langle k| + \int V(\mathbf{r})\, |\mathbf{r}\rangle\langle\mathbf{r}|\, d^3\mathbf{r}, \tag{7}$$

where $m^* = 0.067 m_e$ (m_e is the electron mass) is the conduction band effective mass (assumed for simplicity to be independent of composition) and

$$V(\mathbf{r}) = V_c(\mathbf{r}) - e\phi(\mathbf{r}) \tag{8}$$

is the potential at the point $\mathbf{r}$, which depends on the composition-dependent CBE V_c and on the electrostatic potential $\phi(\mathbf{r})$ (see below). The Hamiltonian for the N localized states is

$$H_N = \sum_j E_{N,j} |N_j\rangle\langle N_j|, \tag{9}$$

where $E_{N,j} = E^0_{N,j} - e\phi(\mathbf{R}_j)$ is the energy of the localized state, including the contribution from the electrostatic potential. The interaction between the conduction band states and the localized states is

$$H_{kN} = \sum_j \beta \sqrt{\frac{a_0^3}{4}} [|N_j\rangle\langle\mathbf{R}_j| + |\mathbf{R}_j\rangle\langle N_j|], \tag{10}$$

where β is a parameter which specifies the interaction strength of the localized state with the conduction band states [in our calculations we set $\beta = 2.5$ eV (Ref. 22)], a_0 is the cubic lattice constant, and $|\mathbf{R}_j\rangle$ is the effective mass position eigenstate at the site of the localized state $|N_j\rangle$.

If $|\Psi_i\rangle$ are the eigenstates of H with energy E_i, each contributes $|\langle\mathbf{r}|\Psi\rangle|^2 + \sum_j |\langle N_j|\Psi\rangle|^2 \delta(\mathbf{r} - \mathbf{R}_j)$ to the particle density at $\mathbf{r}$, where δ is the Dirac δ function. Assuming these states have an occupation $f(E_i)$ [which we assume to be given by the Fermi-Dirac distribution, Eq. (30)], the total charge density due to the carriers at $\mathbf{r}$ will be

$$\rho(\mathbf{r}) = -e \sum_i f(E_i) \times \left\{ |\langle\mathbf{r}|\Psi_i\rangle|^2 + \sum_j |\langle N_j|\Psi_i\rangle|^2 \delta(\mathbf{r} - \mathbf{R}_j) \right\}. \tag{11}$$

If we neglect dielectric screening, the electrostatic potential is determined via Coulomb's law by this charge and the external charge ρ_{ext} (the ionized dopants and the gate charges):

$$\phi(\mathbf{r}) = \int \frac{\rho(\mathbf{r}') + \rho_{ext}(\mathbf{r}')}{4\pi\epsilon_0 |\mathbf{r} - \mathbf{r}'|} d\mathbf{r}'. \tag{12}$$

The potential ϕ in Eq. (12) is more conveniently found by solving the equivalent Poisson equation,

$$-\nabla \cdot [\epsilon(\mathbf{r}) \nabla\phi] = \rho(\mathbf{r}) + \rho_{ext}(\mathbf{r}), \tag{13}$$

where dielectric screening is now accounted for by using the position-dependent dielectric function $\epsilon(\mathbf{r})$ of the material.

We now assume that the charge density and electrostatic potential are uniform along the surface of the heterostructure. Note that we are thus assuming that any disorder (in particular, that arising from the random distribution of N in the dilute nitride layer) can be averaged out for the purposes of determining the transverse wave functions and that momentum parallel to the surface is a good quantum number. In subsequently calculating the alloy scattering in the heterostructure, we take account of the small deviations from this. The potentials V_c and ϕ, the charges ρ and ρ_{ext}, and the dielectric constant ϵ depend only on z. Poisson's equation then becomes

$$-\frac{d}{dz}\left[\epsilon(z)\frac{d\phi}{dz}\right] = \rho(z) + \rho_{ext}(z), \tag{14}$$

which is solved, assuming that ϕ and $d\phi/dz = 0$ inside the back gate ($z = L$), by two integrations:

$$\frac{d\phi}{dz} = \frac{1}{\epsilon(z)} \int_z^L [\rho(z) + \rho_{ext}(z)]\, dz, \tag{15}$$

followed by

$$\phi(z) = -\int_z^L \frac{d\phi}{dz} dz. \tag{16}$$

These integrals are performed using Simpson's rule on the grid of points z_i. The external charge density consists of contributions from the doped layer and from the front and back gates:

$$\rho_{\text{ext}}(z) = eN_D\ \chi_D(z) + \epsilon_{\text{GaAs}}\epsilon_0[\mathcal{E}_{\text{FG}}\delta(z) - \mathcal{E}_{\text{BG}}\delta(z-L)], \tag{17}$$

where $\chi_D(z) = 1$ within the dopant layer and zero elsewhere.

In order to determine $\rho(z)$ in Eq. (14), we must determine the transverse electronic wave functions from our Hamiltonian [Eq. (6)]. To do so, we consider first the transverse wave function for zero momentum parallel to the surface. We then assume that all the summations in the Hamiltonian are over $k = k_z$ only. We want to calculate the effective mass part of the transverse wave function in a plane-wave (i.e., Fourier series) representation. Taking the matrix elements of the effective mass Hamiltonian,

$$\langle k|H_k|k'\rangle = \frac{\hbar^2 k^2}{2m^*}\delta_{k,k'} + \frac{1}{L}\int_0^L V(z)\mathrm{e}^{-i(k-k')z} dz. \tag{18}$$

Since the N localized states are not strictly uniform in the x-y plane, we need to incorporate them in an average way that preserves the approximate conservation of momentum in the x and y directions. If we consider only effective mass states with zero momentum in the x-y plane, then the interaction between the effective mass states and the localized states can be written as

$$\langle k|H_{k\text{N}} = \beta\sqrt{\frac{a_0^3}{4}}\sum_j \langle k|\mathbf{R}_j\rangle\langle N_j| = \frac{\beta}{\sqrt{N_c}}\sum_j \mathrm{e}^{-ikZ_j}\langle N_j|, \tag{19}$$

where N_c is the number of primitive zincblende unit cells (or group-V sites) in the heterostructure and Z_j is the z coordinate of $\mathbf{R}_j$. If we now group together the N states that lie in the mth z slice at z_m, we can define a normalized combination of the localized states in each slice:

$$|N(m)\rangle \equiv \sqrt{\frac{N_s}{x_m N_c}} \sum_{j \atop |Z_j - z_m| < \Delta z/2} |N_j\rangle, \tag{20}$$

where x_m is the nitrogen composition on the group-V sites in z slice m, N_s is the total number of z slices in the heterostructure, and $\Delta z = L/N_s$. We then find that

$$\langle k|H_{k\text{N}} = \beta\sqrt{\frac{x_m}{N_s}}\sum_{m=1}^{N_s} \mathrm{e}^{-ikz_m}\langle N(m)|. \tag{21}$$

We now consider the transverse wave function in subband i to be a linear combination of the effective mass states (with momentum in the z direction) and the slice-averaged localized N states:

$$|\psi_i\rangle = \sum_k \psi_{em,i}(k)|k\rangle + \sum_{m=1}^{N_s} \psi_{\text{N},i}(m)|N(m)\rangle. \tag{22}$$

We obtain the coefficients $\psi_{em,i}(k)$ and $\psi_{\text{N},i}(m)$ as the eigenvectors of the Hamiltonian matrix,

$$\begin{aligned} \langle k|H|k'\rangle &= \langle k|H_k|k'\rangle, \\ \langle k|H|N(m)\rangle &= \beta\sqrt{\frac{x_m}{N_s}}\mathrm{e}^{-ikz_m}, \\ \langle N(m')|H|N(m)\rangle &= \left[E_{\text{N}}^0(m) - e\phi(z_m)\right]\delta_{m,m'}, \end{aligned} \tag{23}$$

where $E_{\text{N}}^0(m)$ is the average N state energy (with respect to the local VBM) in z slice m. The allowed values of the momentum in the z direction are $k = 2\pi n/L$, where $n = 1, N_s$ is an integer. It is unnecessary to include a slice-averaged localized state for slices in which the N composition x_m is zero. A Fourier transform of the effective mass part of the wave function gives the transverse wave function in z:

$$\psi_{em,i}(z) = \frac{1}{\sqrt{L}}\sum_k \psi_{em,i}(k)\mathrm{e}^{ikz}, \tag{24}$$

and for the localized N part

$$\psi_{\text{N},i}(z) = \sum_{m=1}^{N_s} \psi_{\text{N},i}(m)\frac{T_m(z)}{\sqrt{\Delta z}}, \tag{25}$$

where $T_m(z) = 1$ for $|z - z_m| < \Delta z/2$ and zero otherwise.

For nonzero momentum parallel to the heterostructure layers, we note that the only aspect of the transverse Hamiltonian that is altered is the addition of a term $\hbar^2 k_{xy}^2/2m^*$ to the diagonal components of H_k. The state $|N(m)\rangle$ can be replaced with a state that has the appropriate phase in the x-y plane for each localized state,

$$|N(m;k_{xy})\rangle \equiv \sqrt{\frac{N_s}{x_m N_c}} \sum_{j \atop |Z_j - z_m| < \Delta z/2} \mathrm{e}^{i(k_x X_j + k_y Y_j)}|N_j\rangle,$$

and all the matrix elements are the same as for the corresponding case with $k_{xy} = 0$. We could solve this equation separately for each value of $\hbar^2 k_{xy}^2/2m^*$ but is it simpler to take this change into account in perturbation theory. We then find that the effective mass in the x-y plane in subband i, m_i^*, is modified to give

$$\frac{1}{m_i^*} = \frac{1}{L}\int_0^L \frac{|\psi_{em,i}(z)|^2}{m^*(z)} dz, \tag{26}$$

where $m^*(z)$ is the effective mass for a homogeneous system which has the same composition as that of the heterostructure at position z. We define the $m^*(z)$ function as: $m^*(z) = 0.067m_e$ for z in a pure GaAs layer (m_e is the free space electronic mass), $m^*(z) = (0.067 + 0.083x)m_e$ for z in $\text{Al}_x\text{Ga}_{1-x}\text{As}$ layers,[28] $m^*(z) = (0.067 - 0.0435y)m_e$ for z in $\text{In}_y\text{Ga}_{1-y}\text{As}$ layers,[32] and for z in the dilute nitride layer $m^*(z)$ equals the corrected effective mass where hybridization between GaAs conduction band states and localized N states are taken into account [see Eq. (5) of Ref. 22].

Having determined the transverse wave functions as given by Eq. (22), we now determine the charge density $\rho(z)$ used to solve Poisson's equation, Eq. (14). If subband j is occupied with n_j carriers per unit area, then the charge density $\rho_j(z)$ associated with this subband is

$$\rho_j(z) = n_j[|\psi_{em,j}(z)|^2 + |\psi_{\text{N},j}(z)|^2]. \tag{27}$$

Note that the effective mass part and the localized state part of the wave function contribute to the charge density at z. The charge density used to solve Poisson's equation is the sum over all occupied bands of $\rho_j(z)$.

We determine the occupations n_j using the Fermi level E_F and 2D density of states (DOS) function,

$$D(E) = \begin{cases} \frac{jm^*}{\pi\hbar^2} & E_j \leqslant E < E_{j+1} (j = 1, \ldots, \infty), \\ 0 & E < E_1, \end{cases} \quad (28)$$

where E_j is the energy of the jth subband edge, so that

$$n_j = \int_{E_j}^{\infty} f(E) D(E_j) dE, \quad (29)$$

where $f(E)$ is the Fermi-Dirac distribution function

$$f(E) = \frac{1}{e^{(E-E_F)/k_BT} + 1} \quad (30)$$

(where k_B is Boltmann's constant). The Fermi level is determined from the constraint

$$\int_{-\infty}^{\infty} f(E) D(E) dE = n_{\text{ch}}.$$

We have $n_{\text{ch}} = \sum_j n_j$, where the summation is over all occupied subbands.

Using $\rho(z)$, we solve Poisson's equation [Eq. (14)] for the electrostatic potential $\phi(z)$. We then diagonalize our Hamiltonian H, given by Eq. (6) and matrix elements Eq. (23), to recalculate the transverse wave functions. The procedure is repeated self-consistently until the transverse wave functions are converged to within an accuracy of 10^{-8} nm^{-1}. When determining $V(z)$ from Eq. (8), we set the electrostatic potential in the current iteration

$$-e\phi(z) \equiv -e[\mathcal{D}\phi_1(z) + (1 - \mathcal{D})\phi_2(z)], \quad (31)$$

where $\phi_1(z)$ is the total electrostatic potential determined in the present iteration, $\phi_2(z)$ is the electrostatic potential determined in the previous iteration, and $\mathcal{D}$ is a damping factor, set equal to 0.2. If, after a number of iterations, the solution begins to diverge, $\mathcal{D}$ is reduced by a factor of 0.75. In this way, the procedure is allowed to smoothly reach convergence.

IV. DETERMINATION OF MOBILITY

Once we have calculated the transverse wave functions and their energies, we next calculate the mobility μ in the device. To calculate μ we first determine the scattering rate of carriers in subband i, $R_i(E)$, where E is the carrier energy. We have n occupied subbands. Using $R_i(E)$, we then calculate μ through solution of the Boltzmann transport equation in the relaxation time approximation.[22] There are three contributions to the scattering rate in subband i: (i) the N alloy scattering $R_{\text{N},i}$; (ii) the In alloy scattering rate $R_{\text{In},i}$; and (iii) the GaAs phonon scattering rate R_{ph}.

The N alloy scattering $R_{\text{N},i}$ is given by (see Appendix A for derivation)

$$R_{\text{N},i}(E) = \frac{a_0^3}{4N_c\hbar^3} \sum_{j=1}^{n} \sum_{\text{N}} \times \int_{-\infty}^{\infty} \frac{\beta_N^4 m_j^* |\psi_i(z)|^2 |\psi_j(z)|^2 \chi_{\text{N}}(z)\, dz}{[E - E'_{\text{N}}(z)]^2 + [\Gamma_{\text{N}}(z)/2]^2}, \quad (32)$$

where $\psi_i(z)$ is the effective mass part of the transverse wave function, as given by Eq. (24) (we drop the *em* subscript to simplify the notation), and $\chi_{\text{N}}(z) = 1$ for z in the dilute nitride layer and zero elsewhere. By summing over all occupied subbands we account for inter- and intra-subband scattering. Here the N summation is over all localized states determined previously using the LCINS method[15,22,25,26] (not just the isolated N states used in determining the transverse wave functions), each N state having interaction strength parameter with CB states β_{N}, and z dependent energy E'_{N} (shifted due to interaction with the CB), and linewidth Γ_{N}.

We determine the rates of scattering of carriers in subband i from In atoms, $R_{\text{In},i}$, using Fermi's "golden rule."[37] For n occupied subbands, this gives

$$R_{\text{In},i} = \sum_{j=1}^{n} \frac{2\pi}{\hbar} |U_{i,j}|^2 \frac{m_j^*}{2\pi\hbar^2}, \quad (33)$$

where[32]

$$|U_{i,j}|^2 = |V_{\text{alloy}}|^2 y(1 - y) a_{\text{In}}^3 \int_{-\infty}^{\infty} |\psi_i(z)|^2 |\psi_j(z)|^2 \chi_{\text{In}}(z)\, dz$$

is the average square In alloy scattering matrix element. The function $\chi_{\text{In}}(z) = 1$ for z in the $In_yGa_{1-y}As$ layer and zero elsewhere. We take $V_{\text{alloy}} = 0.55$ eV and determine the $In_yGa_{1-y}As$ lattice constant a_{In} as $a_{\text{In}} = 5.653\,25 + 0.405\,15y$ Å(Ref. 32). Because the 2D DOS is constant, $R_{\text{In},i}$ does not depend on carrier energy E. The effective mass in subband i is determined using Eq. (26).

In our calculations the GaAs phonon scattering rate, R_{ph}, is a temperature-dependent phenomenological parameter, independent of carrier energy and determined as that value which gives us mobilities in excellent agreement with the temperature-dependent experimental results of Ref. 38, for a channel with zero In and N content, that is, a pure GaAs channel.

Our total scattering rate of carriers in subband i is

$$R_i(E) = R_{\text{ph}} + R_{\text{In},i} + R_{\text{N},i}(E). \quad (34)$$

Due to occupation by carriers, not all states are available as scattering channels. We therefore must introduce a blocking factor, by replacing $R_i(E)$ with $[1 - f(E)]R_i(E)$, where $f(E)$ is the Fermi-Dirac distribution [Eq. (30)].

Once the scattering rates $R_i(E)$ are determined we calculate the mobility through solution of the Boltzmann transport equation in the relaxation time approximation, following the method used in Ref. 22. Transforming the integral over momentum p in Eq. (30) of Ref. 22 to an integral over energy E (using the dispersion relation $E = E_i + \frac{p^2}{2m^*}$, where E_i is

the CBE in subband i), and evaluating the derivative of the Fermi-Dirac distribution, gives

$$j = \frac{e^2\mathcal{E}}{k_B T \pi\hbar^2} \sum_{i=1}^{n} \int_{E_i}^{\infty} \frac{(E - E_i)e^{(E-E_F)/k_BT}}{R_i(E)(e^{(E-E_F)/k_BT} + 1)^2} dE \quad (35)$$

(here $\mathcal{E}$ is the electric field). Using this expression, combined with the total carrier concentration in the channel $n_{\rm ch}$ determined using the Poisson-Schrödinger solver, the mobility μ is determined from $\mu = j/en_{\rm ch}\mathcal{E}$ to be

$$\mu = \frac{e}{k_B T n_{\rm ch} \pi\hbar^2} \times \sum_{i=1}^{n} \int_{E_i}^{\infty} \frac{(E - E_i)e^{(E-E_F)/k_BT}}{R_i(E)(e^{(E-E_F)/k_BT} + 1)^2} dE. \quad (36)$$

V. RESULTS

We calculate the effect of varying the front- and back-gate electric fields on the carrier mobility in the device, for various nitrogen concentrations in the dilute nitride layer, demonstrating that the carrier mobility may be used to probe localized N states in the dilute nitride layer.

We begin with a N concentration $x = 1.2\%$ and vary the back-gate field over the range $\mathcal{E}_{\rm BG} = -5 \times 10^4$ to 5×10^4 V cm^{-1}. We choose this range of $\mathcal{E}_{\rm BG}$ because, though it may be impractical, it allows us to demonstrate three different modes of operation for the device: The carriers occupy the dilute nitride layer subband alone; the carriers occupy both the dilute nitride layer subband and the InGaAs layer subband, with hybridization occurring between the two subbands; and the carriers occupy the InGaAs layer subband alone. Varying y will allow these different modes of operation to be achieved at more feasible back-gate fields (the effect of varying y on our results is discussed in Appendix D).

We then present results for $x = 0.1\%$ and 0.36%, for which the mode of operation, where only the InGaAs layer subband is occupied by carriers, is achieved at more feasible values of $\mathcal{E}_{\rm BG}$. As noted in Sec. II, having only the InGaAs layer subband occupied by carriers is a necessary condition in order to use the device as a probe of the localized N states, associated with clusters of N atoms, near the band edge in the dilute nitride layer. These cases may, therefore, be used to demonstrate that the device can probe the localized N states, as dips in mobility will be seen when the Fermi level is resonant with these states.

We have kept the cap layer, doped layer, spacer layer, GaAs layer, dilute nitride layer, and buffer layer at constant values of 10, 30, 10, 5, 3, and 50 nm, respectively. We have also kept the Al content, at 40%, and $N_D = 10^{18}$ cm^{-3} constant. A discussion of the reasons for these choices of heterostructure parameters can be found in Appendix C.

A. N concentration of 1.2%

The calculated mobility μ as a function of front- and back-gate electric fields is shown in Fig. 3, for a heterostructure with an InGaAs layer thickness $d_{\rm In} = 24$ nm, a GaAs layer thickness $d_{\rm GaAs} = 5$ nm, a dilute nitride layer thickness $d_{\rm N} = 3$ nm, $y = 3\%$, $x = 1.2\%$, and at $T = 4$ K. For illustrative purposes we have varied $\mathcal{E}_{\rm BG}$ from -5×10^4 to 5×10^4 V cm^{-1}, which

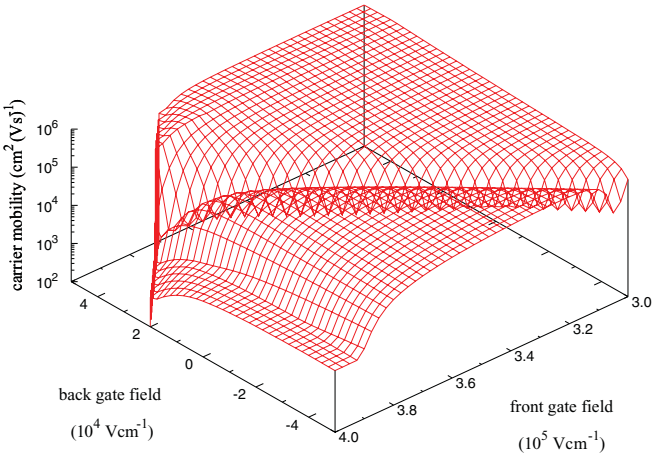

FIG. 3. (Color online) Variation in carrier mobility as a function of applied gate field for a heterostructure with a 24-nm InGaAs layer, a 5-nm GaAs layer, 3-nm dilute nitride layer, In content of 3%, N content of 1.2%, and temperature $T = 4$ K.

may be physically unfeasible in real devices, but allows us to demonstrate three modes of operation of our device: the mode of operation where carriers only occupy the subband in the dilute nitride layer (at, e.g., $\mathcal{E}_{\rm FG} = 4 \times 10^5$ V cm^{-1} and $\mathcal{E}_{\rm BG} = -5 \times 10^4$ V cm^{-1}); the mode of operation where carriers only occupy the subband in the InGaAs layer (at, e.g., $\mathcal{E}_{\rm FG} = 3 \times 10^5$ V cm^{-1} and $\mathcal{E}_{\rm BG} = 5 \times 10^4$ V cm^{-1}); and the mode of operation where both subbands are occupied (at intermediate values of $\mathcal{E}_{\rm FG}$ and $\mathcal{E}_{\rm BG}$). We discuss the effect of varying the gate fields on the subband occupations in Appendix B.

When both $\mathcal{E}_{\rm FG} \geqslant 3.7 \times 10^5$ V cm^{-1} and $\mathcal{E}_{\rm BG} \geqslant 2 \times 10^4$ V cm^{-1}, we have no carriers in the channel and the device is off (a plot of the subband occupations versus gate field can be found in Appendix B, Fig. 12). In this region we set $\mu = 0$. Where the subbands become degenerate and strongly

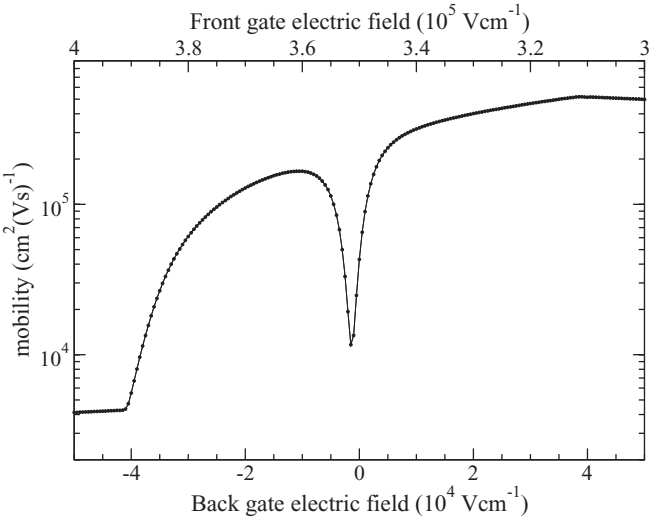

FIG. 4. Variation in mobility along a line taken from Fig. 3, where $\mathcal{E}_{\rm FG}$ and $\mathcal{E}_{\rm BG}$ are simultaneously varied so that $\mathcal{E}_{\rm FG} + \mathcal{E}_{\rm BG}$ is constant. The large dip in mobility is due to strong hybridization of the InGaAs layer and dilute nitride layer subbands as their energies become degenerate.

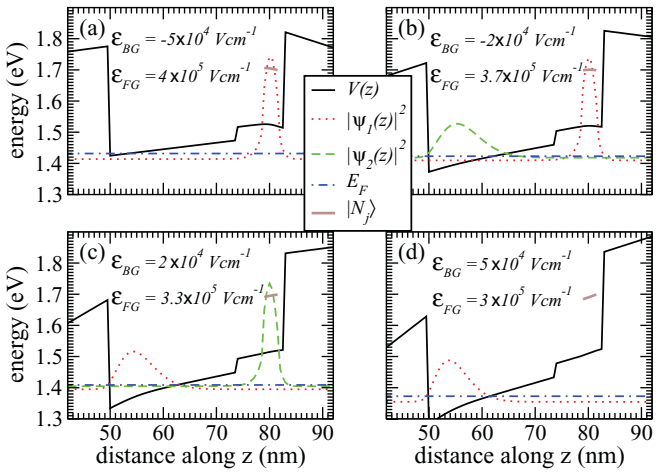

FIG. 5. (Color online) Lowest energy subband weight $|\psi_1(z)|^2$ (red dotted line), second-lowest energy subband weight $|\psi_2(z)|^2$ (green dashed line), and 1D heterostructure potential $V(z)$ (black solid line) as a function of depth z in heterostructure for different points along the line plot shown in Fig. 4. The $|\psi_i(z)|^2$ shown are normalized and plotted at their subband energies E_i. The Fermi level is indicated by the blue dot-dashed line, and the position and energies of the localized N states $|N_j\rangle$ are indicated by the thick brown line. Only the $|\psi_i(z)|^2$ corresponding to occupied subbands are plotted. The subband occupations are summarized in Table III.

hybridize there is a dramatic reduction in μ. This accounts for the large dip visible in Fig. 3, which occurs approximately along the line $\mathcal{E}_{\rm FG} = \mathcal{E}_{\rm BG} + 3.55 \times 10^5$ V cm^{-1} [see Appendix B, Eq. (B3)].

To better illustrate the various features of the plot in Fig. 3, we show in Fig. 4 the mobility variation along the line corresponding to a simultaneous decrease in $\mathcal{E}_{\rm FG}$ from 4×10^5 to 3×10^5 V cm^{-1} and increase in $\mathcal{E}_{\rm BG}$ from -5×10^4 to 5×10^4 V cm^{-1}. This variation in front- and back-gate fields means we keep $n_{\rm ch}$ at a constant value of 5.05×10^{11} cm^{-2} [see Eq. (2)]. In Fig. 5 we plot the subband weights at different values of $\mathcal{E}_{\rm FG}$ and $\mathcal{E}_{\rm BG}$ along this line, similar to the plots demonstrating strong hybridization between the subbands in Fig. 2. The occupations of the subbands for each of Figs. 5(a) to 5(d) are summarized in Table III.

With $\mathcal{E}_{\rm FG} = 4 \times 10^5$ V cm^{-1} and $\mathcal{E}_{\rm BG} = -5 \times 10^4$ V cm^{-1}, we have only the subband in the dilute nitride layer occupied, as shown in Fig. 5(a), and a correspondingly low carrier mobility of $\mu = 4100$ cm^2 (V s)$^{-1}$. As we increase $\mathcal{E}_{\rm BG}$ and decrease $\mathcal{E}_{\rm FG}$, the energy of the InGaAs layer subband decreases relative to that of the dilute nitride layer subband, until it is low enough with respect to the Fermi level that it becomes occupied by carriers. In Fig. 4 this occurs for $\mathcal{E}_{\rm FG} = 3.91 \times 10^5$ V cm^{-1} and $\mathcal{E}_{\rm BG} = -4.1 \times 10^4$ V cm^{-1}, but from Fig. 3 it is evident that the point at which this occurs depends only very weakly on $\mathcal{E}_{\rm BG}$ (the field on the gate furthest from the InGaAs layer), due to the screening effect the carriers in the dilute nitride layer subband have on the field on the back gate. Because a channel in the InGaAs layer with little overlap with the $|N_j\rangle$ states in the dilute nitride layer is now occupied by carriers, we have a dramatic increase in μ.

TABLE III. Occupations of lowest subband n_1 and occupation of second-lowest subband n_2 for Figs. 5(a) to 5(d). The variation in $\mathcal{E}_{\rm FG}$ and $\mathcal{E}_{\rm BG}$ ensures that $n_{\rm ch} = n_1 + n_2$ is kept constant at $n_{\rm ch} = 5.05 \times 10^{11}$ cm^{-2}. The lowest-energy subband switches from the dilute nitride layer in (a) and (b) to the InGaAs layer in (c) and (d).

Fig. 5	n_1 (10^{11} cm^{-2})	n_2 (10^{11} cm^{-2})
(a)	5.05	0
(b)	3.72	1.33
(c)	3.89	1.16
(d)	5.05	0

As we continue to increase $\mathcal{E}_{\rm BG}$ and decrease $\mathcal{E}_{\rm FG}$, the subbands grow closer in energy to each other. This situation is illustrated in Fig. 5(b), for $\mathcal{E}_{\rm FG} = 3.7 \times 10^5$ V cm^{-1} and $\mathcal{E}_{\rm BG} = -2 \times 10^4$ V cm^{-1}. They become degenerate, then swap so that the InGaAs layer subband is now the lowest-energy subband, as shown in Fig. 5(c), for $\mathcal{E}_{\rm FG} = 3.3 \times 10^5$ V cm^{-1} and $\mathcal{E}_{\rm BG} = 2 \times 10^4$ V cm^{-1}. When the subbands are degenerate they both have a large overlap with the dilute nitride layer, and there is therefore a dramatic decrease in μ. This accounts for the large dip in μ seen in Fig. 4 at $\mathcal{E}_{\rm FG} \approx 3.5 \times 10^5$ V cm^{-1} and $\mathcal{E}_{\rm BG} \approx 0$.

As we increase $\mathcal{E}_{\rm BG}$ and decrease $\mathcal{E}_{\rm FG}$ further, we eventually push up the energy of the dilute nitride layer subband relative to the InGaAs layer subband and the Fermi level so that it is no longer occupied by carriers, and all carriers now occupy the InGaAs layer subband alone. This situation is illustrated in Fig. 5(d), for $\mathcal{E}_{\rm FG} = 3 \times 10^5$ V cm^{-1} and $\mathcal{E}_{\rm BG} = -5 \times 10^4$ V cm^{-1}, and is the situation we desire in order to probe the localized N states using mobility measurements in our heterostructure.

To illustrate the effect of varying the gate fields on μ when the InGaAs layer subband is the only one occupied by carriers, we show in Fig. 6 a magnification of the upper right corner of the line plot in Fig. 4, where the mobility is now plotted on a linear, rather than log, scale. The dilute nitride subband empties of carriers when $\mathcal{E}_{\rm FG} = 3.099 \times 10^5$ V cm^{-1} and $\mathcal{E}_{\rm BG} = 4.01 \times 10^4$ V cm^{-1}. When this occurs there is a sudden decrease in the number of carriers being scattered by

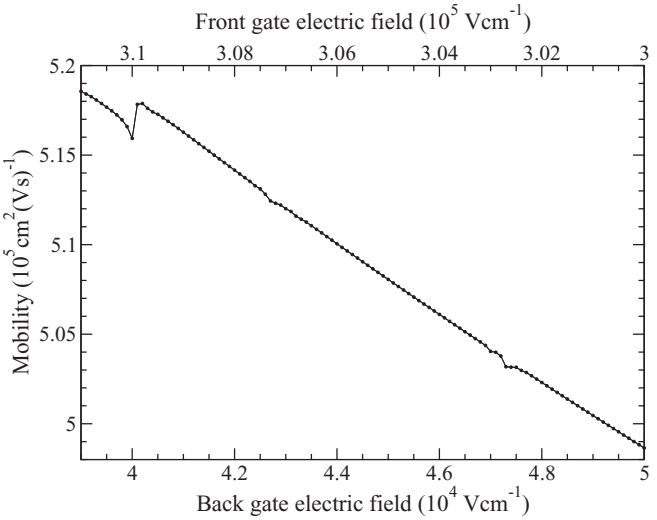

FIG. 6. Magnification of the upper right corner of the line plot shown in Fig. 4. The mobility is shown on a linear rather than log scale here.

N atoms in the dilute nitride layer, and there is therefore a corresponding jump in μ, by 1900 $\text{cm}^2\ (\text{Vs})^{-1}$ (the point at which this occurs depends only very weakly on $\mathcal{E}_{\text{FG}}$, the field on the gate furthest from the dilute nitride layer, due to the screening effect of the carriers in the InGaAs layer subband).

As we increase $\mathcal{E}_{\text{BG}}$ and decrease $\mathcal{E}_{\text{FG}}$ further, there is a general downward trend in μ. This is due to an increase in confinement of the carriers in the InGaAs layer as $\mathcal{E}_{\text{BG}}$ is increased, which leads to an increase in the In alloy scattering rate [Eq. (33)] and hence a decrease in μ. The point to note, however, is that before the dilute nitride layer subband empties, μ is a smoothly varying function of gate field, but once the InGaAs layer subband only is occupied the variation in μ is no longer smooth; instead, we have several small dips.

This succession from smoothly varying μ to a variation including dips once the dilute nitride layer subband empties, can be explained as follows: The linewidths of the localized states depend on the weights of the subband transverse wave functions in the dilute nitride layer [see Eq. (A3)]. When both subbands are occupied, the Fermi level scans through N states that can decay into both subbands. The dilute nitride layer subband has a large weight in the dilute nitride layer, meaning that the states scanned in this case have large linewidths, which, in turn, means they have short lifetimes and are hence relatively delocalized in energy. When the Fermi level decreases below the energy of the dilute nitride layer subband, localized states with the decay channel in the dilute nitride layer cut off from them are probed. The weight of the dilute nitride layer subband in the dilute nitride layer makes no contribution to the linewidths of these states. They therefore have very small linewidths, and hence have very long lifetimes. They make a large contribution to the scattering rate [Eq. (32)], and because they are long-lived, there are sharp resonance points as the Fermi level in the heterostructure sweeps through their energies.

There is a small section of mobility variation with gate field, where the Fermi level is below the energy of the dilute nitride layer subband, but that subband is still occupied. We therefore see a sharp decrease in mobility in this section, as the Fermi level becomes resonant with a localized state, before we see the sudden increase in mobility as the dilute nitride layer subband empties of carriers. However, due to the low temperature in the device (4 K), this section is very small. It can be seen in Fig. 6 at $\mathcal{E}_{\text{FG}} = 3.10 \times 10^5\ \text{V cm}^{-1}$, $\mathcal{E}_{\text{BG}} = 4 \times 10^4\ \text{V cm}^{-1}$. This implies that the statement (i) made in the penultimate paragraph of Sec. II (to probe the localized states, only the InGaAs layer subband should be occupied by carriers) is not strictly true. However, because the range of gate field for which it is not true is very small (due to the low temperature in the device), the statement is a very good approximation.

We show the previously calculated[22] N localized state spectra in Fig. 7, in the vicinity of the GaAs CBE ($\approx$1.519 eV). For the variation in $\mathcal{E}_{\text{BG}}$ from 4.01×10^4 to $5 \times 10^4\ \text{V cm}^{-1}$ shown in Fig. 6, E_F varies from 1.398 to 1.373 eV, and it is this 25-meV section of the spectrum in Fig. 7 that is probed by the heterostructure device. There are three localized states in this range, including one very close to 1.398 eV, the value of E_F when the dilute nitride layer subband is emptied of carriers. We therefore see three corresponding dips in the mobility as E_F is swept through their energies, the first of which occurs directly after the dilute nitride layer subband is emptied of carriers. Because $\mathcal{E}_{\text{BG}}$ is so high in this example, the subband in the InGaAs layer is strongly confined in the triangular well formed there, and its overlap with the dilute nitride layer is low. The dips in mobility as E_F is resonant with the localized states are therefore quite small.

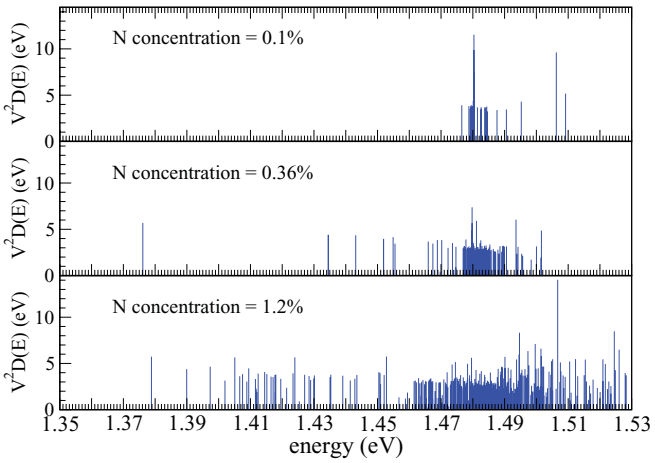

FIG. 7. (Color online) The calculated spectra of the localized N states, plotted as the density of localized LCINS states, weighted by the square of their interaction matrix element with the GaAs CBE state $V^2 = \beta^2/N_c$, for each state vs energy (referred to the GaAs VBM at $T = 4$ K) in the region of the GaAs CBE. The spectra for N concentrations of 0.1%, 0.36%, and 1.2% are shown.

A discussion on the effect varying the In content has on E_F and the value of $\mathcal{E}_{\text{BG}}$ at which the dilute nitride subband empties can be found in Appendix D. By varying the In content it is possible to tune the device design to probe different sections of the spectrum (near the band edge), at smaller back-gate fields than the $\sim 5 \times 10^4\ \text{V cm}^{-1}$ we have used here for illustrative purposes, meaning that larger dips in mobility are observable. In Appendix D we also discuss a qualitative method of determining where E_F lies, relative to the localized N states, and in what range it can be swept through, given the thicknesses of the layers, the In content, and the value of $\mathcal{E}_{\text{BG}}$ at which the dilute nitride layer subband empties.

B. N concentration of 0.1%

The calculations shown so far have been for $x = 1.2\%$, where relatively large back-gate fields were required to achieve the mode of operation where only the InGaAs layer subband is occupied. We next examine the variation in μ for a similar heterostructure, but with $x = 0.1\%$. The concentration of N cluster states near the band edge is much smaller in this case. However, we can probe the states at lower values of $\mathcal{E}_{\text{BG}}$, resulting in larger reductions in μ as E_F becomes resonant with them. With $x = 0.1\%$ in the dilute nitride layer, the quantum well formed at that layer is much shallower, and hence the subband we associate with the layer is higher in energy and less tightly bound to the layer. This means that the mode of operation, where all carriers occupy the subband in the InGaAs layer alone, is achieved at lower values of $\mathcal{E}_{\text{BG}}$ than in the previous higher x case.

In Fig. 8 we plot the calculated mobility of a heterostructure with $x = 0.1\%$, $y = 3\%$, $d_{\text{In}} = 18$ nm, $d_{\text{GaAs}} = 5$ nm, and $d_{\text{N}} = 3$ nm at $T = 4$ K, for variation in $\mathcal{E}_{\text{FG}}$ from 3.5×10^5

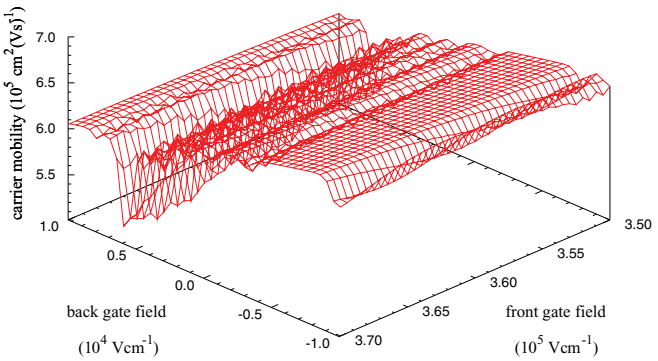

FIG. 8. (Color online) Variation in carrier mobility as a function of applied gate field, for a heterostructure with an 18-nm InGaAs layer, a 5-nm GaAs layer, a 3-nm dilute nitride layer, In content of 3%, N content of 0.1%, and $T = 4$ K.

to 3.7×10^5 V cm^{-1}, and variation in $\mathcal{E}_{BG}$ from -10^4 to 10^4 V cm^{-1}. Decreasing d_{In} from 24 to 18 nm means there is an increase in overlap of the InGaAs layer subband with the dilute nitride layer. This decrease, however, reduces the energy range swept through by E_F for a given variation in $\mathcal{E}_{BG}$. For this range of front- and back-gate field, only the InGaAs layer subband is occupied by carriers. At the $\mathcal{E}_{FG} = 3.5 \times 10^5$ V cm^{-1}, $\mathcal{E}_{BG} = -10^4$ V cm^{-1} corner of the plot, n_{ch} and E_F are at maximum values of 5.76×10^{11} cm^{-2} and 1.510 eV, respectively, decreasing to minimum values of 2.91×10^{11} cm^{-2} and 1.473 eV at the $\mathcal{E}_{FG} = 3.7 \times 10^5$ V cm^{-1}, $\mathcal{E}_{BG} = 10^4$ V cm^{-1} corner of the plot. E_F therefore sweeps through the energies of all the localized N states shown in Fig. 7, with associated dips in the mobilities at resonance.

It is apparent from Fig. 8 that variation in μ has a much greater dependence on $\mathcal{E}_{BG}$ than $\mathcal{E}_{FG}$. The carriers in the InGaAs layer subband screen the field $\mathcal{E}_{FG}$ of the gate nearest to them, so that varying $\mathcal{E}_{FG}$ has little effect on the value of E_F relative to the localized state energies.

To give a clearer understanding of the mobility reduction when E_F is resonant with the localized state energies, we show in Fig. 9 a line plot taken from the surface plot shown in Fig. 8, where $\mathcal{E}_{BG}$ is varied and $\mathcal{E}_{FG}$ is kept constant at

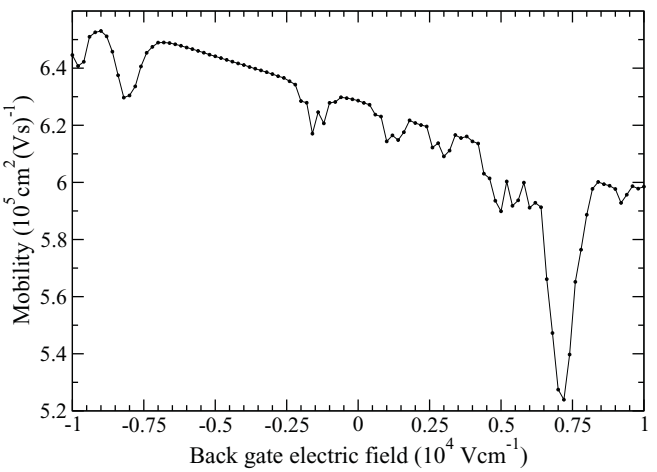

FIG. 9. Mobility variation along the line, taken from Fig. 8 corresponding to a constant $\mathcal{E}_{FG} = 3.5 \times 10^5$ V cm^{-1}.

$\mathcal{E}_{FG} = 3.5 \times 10^5$ V cm^{-1}. As $\mathcal{E}_{BG}$ is increased, there is an overall downward trend in μ. As stated in the $x = 1.2\%$ case, this is due to increasing confinement of the carriers in the InGaAs layer, leading to an increase in In alloy scattering and a smooth decrease in μ. E_F varies from $E_F = 1.510$ eV at $\mathcal{E}_{BG} = -10^4$ V cm^{-1} to $E_F = 1.476$ eV at $\mathcal{E}_{BG} = 10^4$ V cm^{-1}. The dips in mobility occur as E_F sweeps through the energies of the localized N states shown in Fig. 7. The first two dips as $\mathcal{E}_{BG}$ goes from -10^4 to -7500 V cm^{-1} occur as E_F sweeps through the states at energies between 1.510 and 1.506 eV, the next three as $\mathcal{E}_{BG}$ goes from -2500 to 3500 V cm^{-1} occur as E_F sweeps through the states at energies between 1.496 and 1.487 eV, and the rest as $\mathcal{E}_{BG}$ goes from 4500 to 10^4 V cm^{-1} occur as E_F sweeps through the more densely packed cluster of states at energies between 1.485 and 1.478 eV, with the mobility dipping by up to 15% here.

C. N concentration of 0.36%

In Fig. 10, we show the variation in μ as a function of gate field for an identical heterostructure design as that used in the $x = 0.1\%$ case, apart from the N concentration in the dilute nitride layer, which has been increased to 0.36%. In this case we have two distinct regions of mobility variation. First, there is the region where the mobility is relatively low [between approximately 3.3×10^5 and 4.2×10^5 cm^2 (Vs)$^{-1}$] and does not vary greatly with gate field. This occurs for values of $\mathcal{E}_{BG}$ below the line joining the points ($\mathcal{E}_{BG} = -6000$ V cm^{-1}, $\mathcal{E}_{FG} = 3.7 \times 10^5$ V cm^{-1}) and ($\mathcal{E}_{BG} = -4000$ V cm^{-1}, $\mathcal{E}_{FG} = 3.5 \times 10^5$ V cm^{-1}). In this region both subbands are occupied, and E_F is at or above the energy of the dilute nitride layer subband. This means that the N alloy scattering is high, due to carriers occupying the subband in the dilute nitride layer, and the N states have relatively short lifetimes, meaning that sharp resonances do not occur as the Fermi level sweeps through the energy of the N states. Hence the low variation in mobility in this region.

Second, there is the region where the mobility variation is extremely high, as $\mathcal{E}_{BG}$ and $\mathcal{E}_{FG}$ are varied. This occurs for values of $\mathcal{E}_{BG}$ above the line joining the points

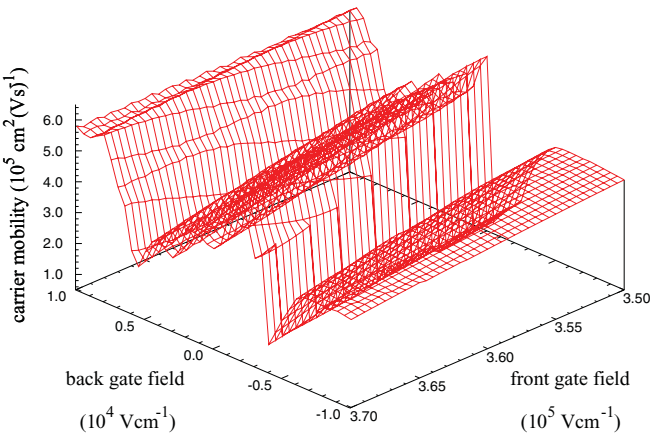

FIG. 10. (Color online) Variation in carrier mobility as a function of applied gate field, for a heterostructure with an 18-nm InGaAs layer, a 5-nm GaAs layer, a 3-nm dilute nitride layer, In content of 3%, N content of 0.36%, and temperature $T = 4$ K.

($\mathcal{E}_{BG} = -6000\,\mathrm{V\,cm^{-1}}, \mathcal{E}_{FG} = 3.7 \times 10^5\,\mathrm{V\,cm^{-1}}$) and ($\mathcal{E}_{BG} = -4000\,\mathrm{V\,cm^{-1}}, \mathcal{E}_{FG} = 3.5 \times 10^5\,\mathrm{V\,cm^{-1}}$). In this region E_F has decreased below the energy of the dilute nitride layer subband. The N states are therefore much more strongly localized in energy (as the decay principal channel in the dilute nitride layer subband has been cut off), and large reductions in μ occur as E_F is resonant with these states. There is a small part of this region where both subbands are occupied, occurring for values of $\mathcal{E}_{BG}$ below the line joining the points ($\mathcal{E}_{BG} = -4000\,\mathrm{V\,cm^{-1}}, \mathcal{E}_{FG} = 3.7 \times 10^5\,\mathrm{V\,cm^{-1}}$) and ($\mathcal{E}_{BG} = -1000\,\mathrm{V\,cm^{-1}}, \mathcal{E}_{FG} = 3.5 \times 10^5\,\mathrm{V\,cm^{-1}}$). Along this line the dilute nitride layer subband empties of carriers, and there is a corresponding large increase in μ.

In the region where both subbands are occupied, there is an overall reduction in μ as $\mathcal{E}_{FG}$ is increased. This can be explained as follows. When $\mathcal{E}_{FG}$ is increased (assuming $\mathcal{E}_{BG}$ remains constant), there is a decrease in n_{ch}. However, the decrease in $\mathcal{E}_{FG}$ is largely screened by the carriers in the InGaAs layer subband. Therefore, the carriers in the dilute nitride layer subband are largely unaffected by the change in $\mathcal{E}_{FG}$, in particular, the subband occupation is unaffected. This means that the decrease in n_{ch} is largely facilitated by a decrease in the occupation of the InGaAs layer subband only, which, in turn, means that a greater proportion of n_{ch} consists of carriers in the dilute nitride layer subband, leading to a greater proportion of carriers being scattered by N states and an overall reduction in μ.

To illustrate the dips in mobility for the 0.36% N concentration case more clearly, in Fig. 11 we show a line plot taken from the surface plot shown in Fig. 10, where $\mathcal{E}_{BG}$ is varied and $\mathcal{E}_{FG}$ is kept constant at $\mathcal{E}_{FG} = 3.7 \times 10^5\,\mathrm{V\,cm^{-1}}$. The dilute nitride layer sub-band is occupied for $\mathcal{E}_{BG} \leqslant -3800\,\mathrm{V\,cm^{-1}}$; once it empties there is a corresponding sudden increase in mobility. The variation in mobility when E_F is above the dilute nitride layer subband energy (for $\mathcal{E}_{BG} \leqslant -6000\,\mathrm{V\,cm^{-1}}$) is much smoother than the variation when it decreases below the dilute nitride layer subband energy, although there is a small feature at $\mathcal{E}_{BG} \approx -8200\,\mathrm{V\,cm^{-1}}$. As $\mathcal{E}_{BG}$ is increased, E_F varies from 1.502 to 1.498 eV, before decreasing below the dilute nitride layer subband energy. The feature at $\mathcal{E}_{BG} \approx -8200\,\mathrm{V\,cm^{-1}}$ is due to resonance between E_F and the group of localized states above 1.498 eV (see Fig. 7). In the region where $-3800 \leqslant \mathcal{E}_{BG} \leqslant 10^4\,\mathrm{V\,cm^{-1}}$, where the dilute nitride layer subband is unoccupied, E_F scans approximately from 1.495 to 1.473 eV, and the dips (of up to ~65%) occur as E_F becomes resonant with the localized N states (shown in Fig. 7) in this range.

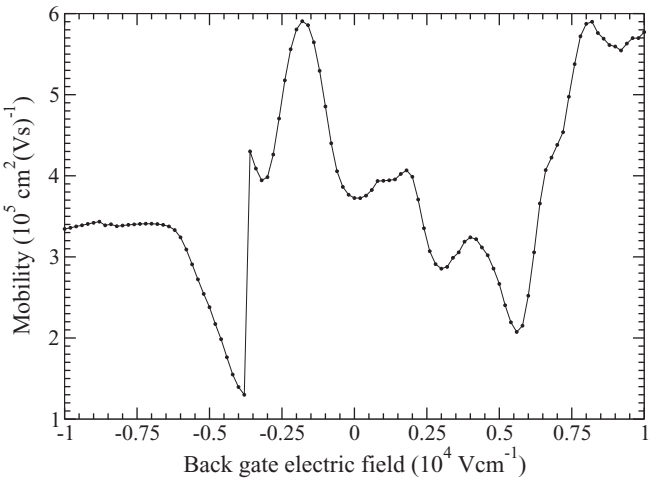

FIG. 11. Mobility variation along the line, taken from Fig. 10, corresponding to a constant $\mathcal{E}_{FG} = 3.7 \times 10^5\,\mathrm{V\,cm^{-1}}$.

From these results and the results for $x = 0.1\%$, we see that, by varying the back-gate field for an appropriate heterostructure geometry, we can probe the localized N cluster states near the dilute nitride CBE. We see substantial reductions in μ in the heterostructure when E_F is resonant with the energies of the localized states. However, because E_F must be below the dilute nitride layer subband energy in order to probe the localized states, the range of states that may be probed is always restricted by the heterostructure design. Localized N states in the dilute nitride layer that lie above the CBE in that layer are, in general, inaccessible to our probing device.

In our calculations we have assumed that all carriers contribute to the conductivity. However, in bulk GaN_xAs_{1-x} one would expect the conductivity to be degraded as electrons can be trapped by localized states below the CBE. In the gated heterostructure device presented in this paper the situation is rather different than in the bulk: Unlike in bulk GaN_xAs_{1-x}, these localized states below the CBE can decay relatively rapidly by coupling to the conduction states in the $GaIn_yAs_{1-y}$ well and, therefore, will not trap charge in the way they might in the bulk. When the coupling to the conduction states in the $GaIn_yAs_{1-y}$ well is weak, as is the case when only the dilute nitride layer subband is occupied, the localized states will have a much smaller decay rate and so may contribute to conductivity degradation. However, we calculate that the density of localized N cluster states below the CBE is one to three orders of magnitude smaller than the density of conduction band states, depending on the N concentration (varying from approximately an order of magnitude smaller for $x = 1.2\%$, to three orders of magnitude smaller for $x = 0.1\%$). Because the density of localized states is lower, we can neglect charge trapping by these states to a good approximation.

VI. CONCLUSION

We have demonstrated a method, by calculating the mobility in gated dilute nitride heterostructures, of probing the localized states related to N complexes in dilute nitride materials. These states, by acting as strong scatterers of carriers, are thought to be the primary mobility-limiting factor in these materials. In our heterostructure design (shown in Fig. 1) we have two available channels, one in the InGaAs layer and one in the dilute nitride layer. Varying the gate fields allows us to vary the relative energies of the subbands in these layers, until we have only the subband in the InGaAs layer occupied by carriers (so that the Fermi level lies below the energy of the dilute nitride layer subband). By varying the gate fields, it is then possible to scan the Fermi level through the energies of the localized states, associated with N complexes in the dilute nitride layer that lie below the dilute nitride layer subband energy, observing reductions in carrier mobility when the Fermi level is resonant with a localized state.

The mobility was determined in three steps. First, the transverse wave functions were determined using a modified

Poisson-Schrödinger approach, with the interaction between conduction band states and isolated N states determined using the BAC model developed by Shan *et al.*[9] The transverse wave functions were then used to determine the In and N alloy scattering rates, with the full range of LCINS states, including N-N interactions, used in determining the N scattering. The scattering rates were then used to solve the Boltzmann transport equation in the relaxation time approximation for the mobility. We have given outlines of design parameters for the heterostructure, giving estimates of how varying different parameters will affect the mobility versus gate field results. We predict reductions of up to 65% in the carrier mobility for a heterostructure with a N content of 0.36%, varying the back-gate field from -10^4 to 10^4 V cm^{-1}, a feasible range for most back-gate geometries used in practice.

ACKNOWLEDGMENTS

We thank C. Kurdak for useful discussions. This work is supported by Science Foundation Ireland.

APPENDIX A: DERIVATION OF N ALLOY SCATTERING RATE

We solve for the N alloy scattering rate within the independent scattering approximation. We generalize the method to determine the carrier scattering in quantum wells used in Ref. 22, to account for our more complicated heterostructure design, and to account for scattering between subbands as well as scattering within subbands.

As in Ref. 22, we assume the quantum well has area A and thickness L, but now we allow occupation of n subbands instead of only the lowest subband in the well. The localized states are statistically of the same character as those of the bulk at the same N concentration, as the quantum well boundary conditions will have no effect on these states, assuming that the quantum well is not too thin (here we ignore the effects of epitaxial strain, which may cause a relative shift in the energies of the localized states with respect to the CBE). In the effective mass approximation, the extended host GaAs conduction band states of the bulk are replaced with states of the following form:

$$\langle k_i|\mathbf{r}\rangle = \frac{1}{\sqrt{A}} e^{i(k_x x + k_y y)} \psi_i(z), \tag{A1}$$

where $|\mathbf{r}\rangle$ is given by Eq. (5), and $\psi_i(z)$ is the effective mass part of the transverse wave function, as given by Eq. (24) (we drop the *em* subscript to simplify the notation), for subband i within the well. The $\psi_i(z)$'s are normalized as in Eq. (4).

To treat the scattering of carriers in a quantum well by a single defect state $|N\rangle$ located at position $\mathbf{r} = (x,y,z)$, we write the scattering Hamiltonian as

$$\Delta H = \beta_N \sqrt{\frac{a_0^3}{4}} [|\mathbf{r}\rangle\langle N| + |N\rangle\langle \mathbf{r}|], \tag{A2}$$

where $a_0^3/4$ is the volume of a single zincblende primitive unit cell and β_N is the interaction matrix element between the defect state and conduction band states. The state $|N\rangle$ is taken from the full spectrum of localized states derived from LCINS calculations,[15,22,25,26] which includes states associated with pairs and clusters of N atoms, in addition to states associated with isolated N atoms. The β_N parameters are determined using the LCINS method.[22] They depend only on the type of localized state (e.g., single N, N-N pairs), assuming that they are independent of the momentum of the conduction band state. This assumption is justified because the state $|N\rangle$ is highly localized and we only consider states $|k_i\rangle$ near the CBE of subband i.

Due to interaction with the conduction subbands, the localized state has a finite lifetime and broadens into a resonance (quasilocalized state). From Fermi's golden rule the decay rate R_N of the state $|N\rangle$ is given by

$$\begin{aligned} R_N &= \sum_{i,k} \frac{2\pi}{\hbar} |\langle k_i|\Delta H|N\rangle|^2 D_i(E_N), \\ &= \frac{2\pi}{\hbar} \beta_N^2 \frac{a_0^3}{4A} \sum_{i=1}^{n} |\psi_i(z)|^2 D_i(E_N) = \frac{\Gamma_N(z)}{\hbar}, \end{aligned} \tag{A3}$$

where $\Gamma_N(z)$ is the (position-dependent) decay width of the quasilocalized state $|N\rangle$, and $D_i(E)$ is the DOS in subband i. We may posit a time-dependent wave function $|\psi\rangle$ with $\langle N|\psi(t)\rangle \approx \exp[-i(E'_N - i\Gamma_N/2)t/\hbar]$, where E'_N is the localized state energy, shifted due to interaction with the conduction band. The Fourier transform of this gives a Lorentzian line shape $G(E)$ for the weight of the localized state:

$$G(E) = \frac{C}{(E - E'_N)^2 + (\Gamma_N/2)^2}. \tag{A4}$$

The normalization constant C is chosen so that $\sum_{i=1}^{n} \int |\langle N|k'_i(E)\rangle|^2 D_i(E) dE = 1$. When the local DOS is approximately constant over the width of the Lorentzian, so that $D_i(E) \approx D_i(E_N)$, and we have $|\langle N|k'_i(E)\rangle|^2 = |\beta_N \psi_i(z)|^2 G(E)$ (see below), it can be shown that

$$C = \frac{a_0^3}{4} \frac{1}{A}.$$

The scattering rate $R_{N,i}(E)$ of carriers in each subband i is given by

$$R_{N,i}(E) = \sum_{j=1}^{n} \frac{2\pi}{\hbar} |\langle k_i|\Delta H|k'_j\rangle|^2 D_j(E). \tag{A5}$$

The scattering matrix element $|\langle k_i|\Delta H|k'_j\rangle|^2 = \frac{a_0^3}{4A} |\beta_N \psi_i(z)|^2 |\langle N|k'_j\rangle|^2$ is symmetric in i and j, which implies that

$$\frac{|\langle N|k'_i(E)\rangle|^2}{|\beta_N \psi_i(z)|^2} = \frac{|\langle N|k'_j(E)\rangle|^2}{|\beta_N \psi_j(z)|^2} = G(E). \tag{A6}$$

Putting this all together gives the following expression for the N alloy scattering rate of carriers in each subband:

$$R_{N,i}(E) = \sum_{j=1}^{n} \frac{1}{A} \sum_{N} \frac{|\beta_N|^4 |\psi_i(z_N)|^2 |\psi_j(z_N)|^2}{[E - E'_N(z_N)]^2 + [\Gamma_N(z_N)/2]^2} \frac{a_0^6 m_j^*}{16\hbar^3}, \tag{A7}$$

where z_N is the z coordinate of localized state $|N\rangle$. We sum over all subbands to account for inter- and intra-subband scattering, and sum over all localized states $|N\rangle$ to account for scattering (within the independent scattering approximation) by single-N states, N-N pairs, etc. We have put in the explicit form of the 2D DOS in subband j (one spin direction only)

$D_j(E) = Am_j^*/2\pi\hbar^2$. The linewidth for each state $|N\rangle$, Γ_N, depends on z, as seen from Eq. (A3). The effective mass in each subband is determined using Eq. (26). The argument z_N implies that we only include in the summations the values of the transverse wave functions for which z is in the dilute nitride layer.

In determining the spectrum of localized states using the LCINS approach,[15] calculations were performed using a GaAs supercell of volume V, with N atoms distributed in the supercell in a homogeneous statistical distribution. Incorporating this into our present case of a GaNAs quantum well of area A, we assume that the quantum well can be represented by a "slice" of the volume V of the supercell used in the LCINS approach. This implies that we can replace

$$\frac{1}{A}\sum_N \rightarrow \frac{1}{V}\sum_N \int_{-\infty}^{\infty} \chi_N(z)dz$$

in Eq. (A7), where V is the volume of the GaAs supercell ($V = N_c a_0^3/4$, where N_c is the number of group V sites in the supercell) used in the LCINS approach, and $\chi_N(z)$ is a weighting function where $\chi_N(z) = 1$ in the GaNAs layer and $\chi_N(z) = 0$ elsewhere. This gives us the following expression for $R_{N,i}(E)$:

$$R_{N,i}(E) = \frac{a_0^3}{4N_c\hbar^3}\sum_{j=1}^{n}\sum_N \int_{-\infty}^{\infty} \frac{\beta_N^4 m_j^* |\psi_i(z)|^2 |\psi_j(z)|^2 \chi_N(z)dz}{[E - E_N'(z)]^2 + [\Gamma_N(z)/2]^2} \tag{A8}$$

[Eq. (32) of the text]. For our low-temperature calculations $R_{N,i}(E)$ is tabulated for energies up to 225 k_BT above the CBE in subband i, E_i, which is a range high enough above the Fermi level to include all the energies at which the Fermi-Dirac distribution function is substantial.

The shifted energy $E_N'(z)$ and linewidth $\Gamma_N(z)$ of the localized states are determined using the generalized normalization sum rule

$$\sum_i \int W_i(E)D_i(E)dE = 1, \tag{A9}$$

where $W_i(E) = |\langle k_i'|N\rangle|^2$, and the generalized energy sum rule

$$\sum_i \int W_i(E)E D_i(E)dE = E_N \tag{A10}$$

(see Appendix A of Ref. 22). The shifted energy $E_N'(z)$ is determined using the default linewidth [see Eq. (A3)]

$$\Gamma_{N_0}(z) = \frac{a_0^3}{4\hbar^2}\beta_N^2 \sum_i m_i^* |\psi_i(z)|^2, \tag{A11}$$

combined with Eq. (A10), which gives the expression

$$E_N - E_N'(z) = \frac{a_0^3\beta_N^2}{8\pi\hbar^2}\sum_{i=1}^{n}\int_{E_i}^{E_{max}} \frac{[E - E_N]m_i^*|\psi_i(z)|^2}{[E - E_N]^2 + [\Gamma_{N_0}(z)/2]^2}dE, \tag{A12}$$

where we set $E_{max} = E_i + 0.23$ eV. If $E_N \leqslant E_i$ we set $\Gamma_{N_0} = 0$. Having determined $E_N'(z)$ we then use Eq. (A9) to determine the adjusted linewidth $\Gamma_N(z)$ from

$$1 = \frac{a_0^3\beta_N^2}{8\pi\hbar^2}\sum_{i=1}^{n}\int_{E_i}^{E_{max}} \frac{m_i^*|\psi_i(z)|^2}{[E - E_N'(z)]^2 + [\Gamma_N(z)/2]^2}dE, \tag{A13}$$

which holds true for each value of z, thus giving us our function $\Gamma_N(z)$.

APPENDIX B: OCCUPATION OF SUBBANDS AS A FUNCTION OF GATE FIELD

We demonstrate the effects of varying the front- and back-gate fields on the subband occupations in Fig. 12, for the same heterostructure as that used for the mobility results in Fig. 3 (i.e., a heterostructure with $d_{In} = 24$ nm, $d_{GaAs} = 5$ nm, $d_N = 3$ nm, $y = 3\%$, and $x = 1.2\%$ at $T = 4$ K). The range of $\mathcal{E}_{FG}$ and $\mathcal{E}_{BG}$ used allows us to demonstrate three modes of operation of the device: the mode of operation where only the dilute nitride subband is occupied by carriers (at, e.g., $\mathcal{E}_{FG} = 4 \times 10^5$ V cm^{-1} and $\mathcal{E}_{BG} = -5 \times 10^4$ V cm^{-1}), the mode of operation where only the InGaAs layer subband is occupied by carriers (at, e.g., $\mathcal{E}_{FG} = 3 \times 10^5$ V cm^{-1} and $\mathcal{E}_{BG} = 5 \times 10^4$ V cm^{-1}), and the mode of operation where both subbands are occupied by carriers (at intermediate values of $\mathcal{E}_{FG}$ and $\mathcal{E}_{BG}$). The subbands are degenerate along the line where their occupations are equal.

For this large variation in $\mathcal{E}_{BG}$, reasonable values of $\mathcal{E}_{FG}$, and typical heterostructure design we do not have any occupation of higher-order subbands; the only occupied subbands are the InGaAs layer subband and the dilute nitride layer subband. If we were to decrease $\mathcal{E}_{FG}$ below 3×10^5 V cm^{-1}, and also decrease $\mathcal{E}_{BG}$ to below -5×10^4 V cm^{-1}, we would increase n_{ch} to such a degree that unwanted third-order subbands would begin to be populated.

The sum of the two subband occupations equals n_{ch}, and in the 2D space defined by $\mathcal{E}_{FG}$ and $\mathcal{E}_{BG}$, n_{ch} describes a planar surface given by Eq. (2). As can be seen on the right-hand side of Eq. (2), the value of n_{ch} is offset by the doping contribution, $N_D d$, to carrier density from the doped layer. As n_{ch} physically

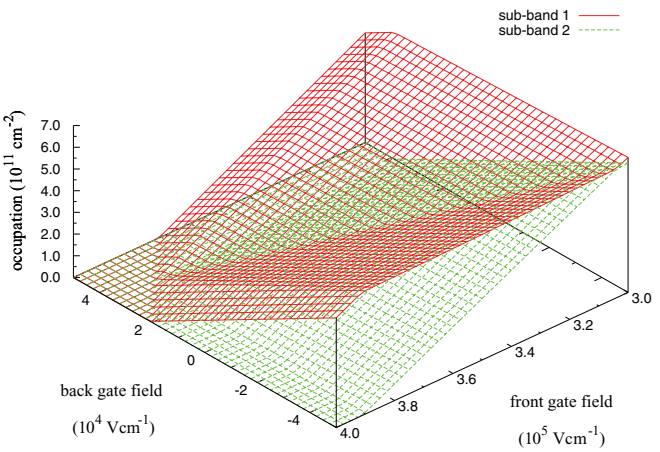

FIG. 12. (Color online) Variation in occupation of the subbands as a function of applied gate field, for a heterostructure with a 24-nm InGaAs layer, a 5-nm GaAs layer, a 3-nm dilute nitride layer, In content of 3%, N content of 1.2%, and temperature $T = 4$ K. Subband 1 is the lower-energy subband.

cannot be negative, the plane does not cross the $\mathcal{E}_{FG}$-$\mathcal{E}_{BG}$ plane and instead goes to zero for negative solutions to Eq. (2). A portion of the n_{ch} plane can be seen in Fig. 12, in the regions where only one subband is occupied (this occurs in the range $\mathcal{E}_{FG} = 3.925$–4×10^5 V cm^{-1}, $\mathcal{E}_{BG} = -5$–2.08×10^4 V cm^{-1}; and also in the range $\mathcal{E}_{FG} = 3$–3.708×10^5 V cm^{-1}, $\mathcal{E}_{BG} = 3.75$–5×10^4 V cm^{-1}). When $n_{ch} = 0$ the device is off. This occurs when $\mathcal{E}_{FG} + \mathcal{E}_{BG} \geqslant 4.208 \times 10^5$ V cm^{-1}.

Decreasing either $\mathcal{E}_{FG}$ or $\mathcal{E}_{BG}$ will increase n_{ch}, as is evident from Eq. (2). A decrease of field on a gate near a subband will largely only affect that subband, with little effect on the further away subband, due to screening. This is evident from Fig. 12, if we consider a decrease in field on one gate while keeping the field on the other gate constant (e.g., as $\mathcal{E}_{FG}$ is decreased for a constant $\mathcal{E}_{BG} - 5 \times 10^4$ V cm^{-1}, or as $\mathcal{E}_{BG}$ is decreased for a constant $\mathcal{E}_{FG} = 3 \times 10^5$ V cm^{-1}). Initially strong variation in the occupation of the subband furthest from the gate whose field is being varied is observed, as this is the only occupied subband. However, when the subband nearest the gate being varied becomes occupied, the variation in field only affects that subband, while the occupation of the further-away subband has only a weak dependence on the gate field being varied.

Whether we vary $\mathcal{E}_{FG}$ and keep $\mathcal{E}_{BG}$ constant or *vice versa*, the screening of the field interacting with carriers in one subband by the carriers in the other is not perfect. The change in subband occupation as the gate field is reduced (while the other gate field is held constant), once both subbands can be populated, is ~20 times greater for the subband in closest proximity to the gate field being varied than for the further-away subband.

If perfect screening is assumed, one can derive simple analytical expressions for the surfaces defined by the occupations of both subbands, in the 2D space defined by $\mathcal{E}_{FG}$ and $\mathcal{E}_{FG}$. We denote the occupation of the InGaAs layer subband by n_{In}, and denote the occupation of the dilute nitride layer subband by n_N. We have

$$n_{ch} = n_{In} + n_N, \tag{B1}$$

provided that, at most, two subbands are occupied (our desired case). The n_{In} and n_N surfaces are visible in Fig. 12 as the two interweaving planar surfaces, each composed of parts of the subband 1 occupation (n_1) and the subband 2 occupation (n_2) surfaces, which cross each other at the line where $n_1 = n_2$.

One can relate n_1 and n_2 in Fig. 12 [as given in Eq. (29)] to the components in Eq. (B1) as follows: When only one subband is occupied (the range of field where this occurs was given above), $n_1 = n_{ch}$, $n_2 = 0$; when $\mathcal{E}_{FG}$ is increased from 3×10^5 to 3.925×10^5 V cm^{-1} and $\mathcal{E}_{BG}$ is simultaneously decreased from 3.75×10^4 to -5×10^4 V cm^{-1}, $n_1 = n_{In}$ and $n_2 = n_N$; when $\mathcal{E}_{FG}$ is decreased from 3.925×10^5 to 3×10^5 V cm^{-1} and $\mathcal{E}_{BG}$ is simultaneously increased from -5×10^4 to 3.75×10^4 V cm^{-1}, $n_1 = n_N$ and $n_2 = n_{In}$.

The assumption of perfect screening is equivalent to the statements

$$\frac{\partial n_{In}}{\partial \mathcal{E}_{FG}} = \frac{\partial n_{ch}}{\partial \mathcal{E}_{FG}}, \quad \frac{\partial n_{In}}{\partial \mathcal{E}_{BG}} = 0;$$
$$\frac{\partial n_N}{\partial \mathcal{E}_{FG}} = 0, \quad \frac{\partial n_N}{\partial \mathcal{E}_{BG}} = \frac{\partial n_{ch}}{\partial \mathcal{E}_{BG}}.$$

This, in combination with Eq. (2), gives the following expressions for the n_{In} and n_N surfaces

$$-en_{In} = \epsilon_{GaAs}\epsilon_0(\mathcal{E}_{FG} - \mathcal{E}_{FG,0}),$$
$$-en_N = \epsilon_{GaAs}\epsilon_0(\mathcal{E}_{BG} - \mathcal{E}_{BG,0}), \tag{B2}$$

where $\mathcal{E}_{FG,0}$ is the front-gate field at which n_{In} goes to zero ($\mathcal{E}_{FG,0} = 3.925 \times 10^5$ V cm^{-1} in Fig. 12) and $\mathcal{E}_{BG,0}$ is the back-gate field at which n_N goes to zero ($\mathcal{E}_{BG,0} = 3.75 \times 10^4$ V cm^{-1} in Fig. 12). Setting $n_{In} = n_N$ gives us the line in the $\mathcal{E}_{FG}$-$\mathcal{E}_{BG}$ plane along which the subbands are degenerate,

$$\mathcal{E}_{FG} + \mathcal{E}_{BG,0} = \mathcal{E}_{BG} + \mathcal{E}_{FG,0}, \tag{B3}$$

assuming the device is on and no higher order subbands are occupied. Combining Eq. (B2) with Eq. (B1) leads to the relation

$$\epsilon_{GaAs}\epsilon_0(\mathcal{E}_{FG,0} + \mathcal{E}_{BG,0}) = eN_Dd, \tag{B4}$$

which, given our values of $\mathcal{E}_{FG,0}$, $\mathcal{E}_{BG,0}$, $N_D = 10^{18}$ cm^{-3}, and $d = 30$ nm, is accurate to within 2%. Therefore, if either $\mathcal{E}_{FG,0}$ or $\mathcal{E}_{BG,0}$ is determined, and assuming we know N_D and d, the relation between subband occupations and gate fields can be constructed. We discuss in Appendix D 3 how the In content will affect this relation.

Our calculations predict a sudden increase in carrier mobility when the subband in the dilute nitride is emptied of carriers (see Sec. V). This increase will be a detectable feature in mobility versus gate field measurements, and from this feature $\mathcal{E}_{BG,0}$ can be determined. The actual values of $\mathcal{E}_{FG,0}$ and $\mathcal{E}_{BG,0}$ will depend on the channel layer widths and chemical compositions.

APPENDIX C: PARAMETERS KEPT CONSTANT IN THE CALCULATIONS

In the calculations we have presented above, we have kept the cap layer, buffer layer, doped layer, spacer layer, GaAs layer, and dilute nitride layer thicknesses constant. The doping density in the doped layer has also been kept constant. We discuss the reasons for this here.

Varying the thicknesses of the undoped GaAs cap layer and the undoped buffer layer only affects our results by altering the relationships between the gate fields and gate voltages. As the electric field is voltage per unit length, making these layers thicker (thinner) means a larger (smaller) voltage for the corresponding electric field. We have chosen a thickness of 10 nm for the cap layer, which is typical of GaAs/AlGaAs heterostructures (see, for example, Ref. 39). For the buffer layer we have chosen a thickness of 50 nm, consisting of 45 nm of undoped $Al_{0.4}Ga_{0.6}As$ grown on 5 nm of undoped GaAs. In practice, thicker buffer layers may be necessary. However, thickening the buffer layer will have no qualitative effect on our results, besides increasing the required voltage to achieve the same back-gate field (in our calculations the values of front- and back-gate fields we use correspond to voltages of magnitudes no greater than 1 V).

The thickness and doping density of the doped layer only affects our results through the determination of n_{ch} [see Eq.(2)]. We can keep the contribution to the charge per unit area in the channel from the doped layer constant, if, when

varying d, we also vary N_D so that $N_D d$ is kept constant. In our calculations we have chosen $N_D = 10^{18}$ cm^{-3} and $d = 30$ nm. If we were to increase d to 60 nm, we would get the exact same results as before if we decreased N_D to 5×10^{17} cm^{-3}.

Changing the thickness of the AlGaAs spacer layer between the doped layer and the channel has little effect on our calculations. Changing the thickness of this layer from 0 to 20 nm, while keeping all other parameters constant, has no significant effect on the subband energies. While keeping all other parameters constant, changing the thickness from 0 to 2 nm changes the individual subband occupations by ∼0.01%, while changing thickness from 2 to 20 nm changes the subband occupations by ∼0.0001%. However, increasing the thickness of the spacer layer raises the Fermi level with respect to the CBE in the doped layer. This may lead to carriers occupying DX centers in the doped layer at higher values of $\mathcal{E}_{\mathrm{FG}}$ than our minimum value of 3×10^5 V cm^{-1}. We have therefore chosen the reasonable value of 10 nm for the spacer layer thickness in our calculations.

We have chosen an Al content of 40% in the AlGaAs-doped, spacer, and buffer layers. This is a high enough Al content to achieve sufficient confinement in the channel, while ensuring the AlGaAs alloy is a Γ-point direct band gap semiconductor alloy,[28] so that our band-offset model of the potential in the heterostructure is valid. If we decrease the Al content to 30% (a commonly used Al content in AlGaAs/GaAs MODFETs), the only significant effect on our results is that the front-gate field at which DX^0 centers in the doped layer may become occupied is increased. We still find, however, that in our ranges of front- and back-gate field, with an Al content of 30%, the occupation of DX^0 centers can be avoided.

The GaAs layer in the channel helps confine the InGaAs layer subband in the InGaAs layer, while also limiting the overlap of the wave function of that subband with the dilute nitride layer. Large overlaps are undesirable as the mobility will then be drastically reduced. Very low overlaps are also undesirable as then the variation in mobility as E_F becomes degenerate with localized states in the dilute nitride layer would be too low to be detectable. The combined thickness of the GaAs and InGaAs layers affects the degree of overlap, as well as affecting the range of energy probed by the device. When studying these effects, we keep the GaAs layer thickness constant at 5 nm and vary the InGaAs layer thickness only. We do this because the effect of increasing the GaAs layer thickness above 5 nm is equivalent to the effect of increasing the InGaAs layer thickness by the same amount. If the GaAs layer thickness is reduced below 5 nm, a larger proportion of the InGaAs layer subband wave function penetrates through to the dilute nitride layer, meaning there is an increase in overlap and an associated decrease in mobility. This decrease becomes substantial once the GaAs layer thickness is reduced below 3 nm.

To probe the localized states associated with clusters of N in the dilute nitride layer requires that we have carriers occupying the subband in the triangular InGaAs quantum well only. It is therefore helpful to have the energy of the subband in the dilute nitride layer as high as possible with respect to the InGaAs layer subband, as this means it is easier to empty that subband of carriers with an applied field on the back gate, if the subband in the InGaAs layer is at a relatively lower energy. By making the dilute nitride layer as narrow as possible, we maximize the effect quantum confinement has in increasing the energy of the subband in that layer. In our calculations we have chosen a dilute nitride layer thickness of 3 nm, which is sufficiently narrow for quantum confinement to significantly increase the subband energy, but still wide enough for the distribution of nitrogen cluster states to be similar to bulk materials.

APPENDIX D: VARYING THE In CONTENT AND InGaAs LAYER THICKNESS

We discuss here how varying the In content and width of the InGaAs layers affects the results presented in Sec. V. We first present an analytical model that can be used to estimate the Fermi level relative to the localized N states in the dilute nitride layer, when only the InGaAs layer subband is occupied.

1. Simple model to determine the Fermi level, relative to the localized N states in the dilute nitride layer

When only the InGaAs layer is occupied, the effect of the front-gate field on the subband energy can be neglected to a good approximation. One can then use a simple model, derived from parametrizations of our full numerical results, to determine the position of the Fermi level relative to the localized N states in the dilute nitride layer. The parameters in the model depend on the thicknesses of the layers in the channel, the In content, the channel carrier density n_{ch}, and the back-gate field. The front-gate field only affects the model in determining n_{ch}, given by Eq. (2). We denote the InGaAs, GaAs, and dilute nitride layer thicknesses by d_{In}, d_{GaAs}, and d_{N}, respectively (given in nm), and the In content by y.

Consider the case where a known carrier density, n_{ch}, occupies the InGaAs layer sub-band alone, and $\mathcal{E}_{\mathrm{BG}} = 0$. The InGaAs layer subband is located in a quantum well that is approximately triangular near the AlGaAs/InGaAs interface, and has a barrier of height $\Delta E_{\mathrm{C,In}}$ (see Sec. II) at the InGaAs/GaAs interface. As the back-gate field is zero and all the charge in the channel is confined in the InGaAs layer, the potential $V(z)$ will be flat for z in the GaAs and dilute nitride layers in the channel and equal to the GaAs CBE (referenced to the VBM), E_{GaAs}, given by Eq. (1). The energy E_1 of the InGaAs layer subband edge, relative to E_{GaAs}, will then depend on y (because y affects the magnitude of $\Delta E_{\mathrm{C,In}}$) and on d_{In} (through a quantum confinement effect). If we denote the difference between E_{GaAs} and E_1 as ΔE_1, we find that the fit (in eV)

$$\Delta E_1 = F_1(y)\frac{1}{d_{\mathrm{In}}^2} + F_2(y), \tag{D1}$$

where

$$F_1(y) = 0.3967\frac{1}{\sqrt{y}} - 4.8181, \tag{D2}$$

and

$$F_2(y) = 1.4017y + 0.0086 \tag{D3}$$

[where $F_1(y)$ is in eV nm^2 and $F_2(y)$ is in eV] gives values of E_1 in excellent agreement with our full numerical solutions, with errors typically less than 1 meV. As n_{ch} is confined to the

InGaAs subband alone, E_F can then be taken to be $n_{ch}/D(E)$ above E_1, where $D(E)$ is the 2D DOS, $D(E) = m^*/\pi\hbar^2$.

When $\mathcal{E}_{BG}$ is not equal to zero, the energies of the localized states in the dilute nitride layer will vary relative to the energy of the subband edge in the InGaAs layer (but remain fixed relative to the band edge in the dilute nitride layer). As long as the dilute nitride layer subband remains empty of carriers, this effect can be incorporated in our model by analogy to a lever arm, rotating about a pivot point located in the InGaAs layer. We define z_{field} as the distance from the center of the dilute nitride layer to this pivot point in the InGaAs layer, so that the shift in the relative energy between the localized states and the InGaAs layer subband edge is approximately $\mathcal{E}_{BG} z_{field}$. From our full numerical solution, we find that z_{field} is approximately independent of y and n_{ch} and only depends on the layer thicknesses. It can be determined (in nm) as

$$z_{field} = 0.7144 d_{In} - 3.1120 + d_{GaAs} + \tfrac{1}{2} d_N. \qquad \text{(D4)}$$

This fit gives excellent agreement with our full numerical solutions for $\mathcal{E}_{BG} \leqslant 10^4$ V cm^{-1} and $d_{In} \leqslant 24$ nm, with errors typically less than 3 meV. For higher fields, however, the error can increase up to 10–15 meV.

Putting all this together, a very good approximation of E_F, given that only the InGaAs layer subband is occupied, is given by

$$E_{F,approx} = E_{GaAs} - \Delta E_1 + n_{ch}\frac{\pi\hbar^2}{m^*} - \mathcal{E}_{BG} z_{field}. \qquad \text{(D5)}$$

This approximation is only valid if the dilute nitride layer subband is not populated with carriers. However, as shown is Sec. V, we predict a sudden increase in mobility when the dilute nitride layer subband empties of carriers, which should be possible to observe experimentally. Then, knowing the value of $\mathcal{E}_{BG}$ at which this occurs (denoted $\mathcal{E}_{BG,0}$ in Appendix B), and knowing the layer thicknesses, y, and n_{ch}, the portion of the spectra shown in Fig. 7 probed by the device can be estimated using this simple model. For example, in the 0.36% N concentration case presented in Sec. V C, E_F was found to be 1.495 eV when the dilute nitride layer subband was emptied of carriers (at $\mathcal{E}_{BG,0} = -3800$ V cm^{-1}) and decreased to 1.473 eV as $\mathcal{E}_{BG}$ was increased to 10^4 V cm^{-1} (see Fig. 11). Our model, given by Eq. (D5), gives $E_{F,approx} = 1.496$ eV where the dilute nitride layer subband empties, decreasing to 1.471 eV at $\mathcal{E}_{BG} = 10^4$ V cm^{-1}. Similar good agreement is found for the 0.1% N concentration case, given is Sec. V B.

For the 1.2% N concentration case given in Sec. V A we have a discrepancy of 10–15 meV, as here the dilute nitride layer subband does not empty until $\mathcal{E}_{BG} = 4.01 \times 10^4$ V cm^{-1}, a value beyond the accuracy of our model (and beyond the practical limits of experiment). However, if we increase the In content to 10% and keep all the other heterostructure parameters constant, we find that at $\mathcal{E}_{BG} = 10^4$ V cm^{-1} the dilute nitride layer subband is empty of carriers (in general, increasing y will decrease the value of $\mathcal{E}_{BG}$ at which the dilute nitride subband empties of carriers). In this case, from our full numerical solution we find that $E_F = 1.373$ eV, and from our model we find that $E_{F,approx} = 1.374$ eV.

The model presented in this section can be used to account for the effects of varying y and d_{In} on our results. In the next sections we discuss the how the position of E_F varies for different values of y; the change in the value of gate field at which the dilute nitride layer subband empties ($\mathcal{E}_{BG,0}$), and the value at which the subbands are degenerate, for different values of y; and the variation in the range of energy scanned by the probe as d_{In} is varied.

2. The Fermi level as a function of In content

Varying the In content in the InGaAs layer effectively varies the energy of the subband in that layer, relative to the subband in the dilute nitride layer. This is because varying the In content varies the conduction band offset between the InGaAs and GaAs layers, and hence deepens the quantum well in the InGaAs layer. When only the InGaAs layer subband is occupied by carriers, E_F will follow the variation in the energy of that subband as the In content is varied, as E_F will be a constant value of $n_{ch}/D(E)$ above the subband energy. It is therefore possible to vary the portion of the spectrum of localized states probed (which is determined by E_F), by growing heterostructures with different In contents.

If both the InGaAs layer subband and the dilute nitride layer subbands are occupied, the effect of varying In content on E_F is much reduced. This occurs because, with the dilute nitride layer occupied, E_F is pinned close to the energy of the subband in that layer. Varying the In content only affects the energy of the subband in the InGaAs layer, so when both subbands are occupied, an increase in In content (leading to a reduction in energy of the InGaAs layer subband relative to the dilute nitride layer subband) will lead to an increase in the occupation of the InGaAs layer subband (as it is now lower in energy) and a decrease in the occupation of the dilute nitride layer subband. Therefore, in this case varying the In content only affects E_F through the reduction in occupation of the dilute nitride layer subband, an effect that is much smaller than the case where only the InGaAs layer subband is occupied.

For a heterostructure with $d_{In} = 20$ nm, $d_{GaAs} = 5$ nm, $x = 0.1\%$, $T = 4$ K, and at constant $\mathcal{E}_{FG} = 3.5 \times 10^5$ V cm^{-1} and $\mathcal{E}_{BG} = 0$, we have only the InGaAs layer subband occupied for a range of y from 2% to 10%. In this case we find that the relationship between y and E_F follows closely the relationship between InGaAs/GaAs conduction band offset and y. Our model given in Eq. (D5) reproduces the variation determined from our full numerical calculation almost exactly, with errors of only less than 0.5 meV.

It is therefore possible, using our model given in Eq. (D5), to estimate how the portion of the spectrum of localized states probed by a specific heterostructure device will change, if an identical device is grown with a different In content.

3. Effect of In content on the variation of the subband energies with gate field

As varying the In content effectively varies the energy of the subband in the InGaAs layer, relative to the subband in the dilute nitride layer, one can also vary how the subband energies change with gating fields by varying the In content. In particular, the values of $\mathcal{E}_{FG}$ and $\mathcal{E}_{BG}$ at which the subband energies are degenerate, and the value $\mathcal{E}_{BG,0}$ where the dilute nitride subband empties of carriers, may be varied in this way.

For example, in Sec. V A we found that, for our heterostructure design with nitrogen concentration, $x = 1.2\%$, the dilute nitride subband emptied at $\mathcal{E}_{\mathrm{BG},0} = 4.01 \times 10^4$ V cm^{-1}. If we increase the In content y from 3% to 4%, we find that the dilute nitride subband now empties at $\mathcal{E}_{\mathrm{BG},0} = 3.39 \times 10^4$ V cm^{-1}, a decrease of 6200 V cm^{-1}. If we take the change in energy of the subband in the InGaAs layer predicted by our simple model given by Eq. (D5), $\Delta E = \Delta E_1(y = 0.04) - \Delta E_1(y = 0.03)$, divided by z_{field} as an approximation of the change in $\mathcal{E}_{\mathrm{BG},0}$, we get that change to be 6391 V cm^{-1}. This implies that our simple model reproduces quite well the change in field at which the dilute nitride subband empties as y is varied. From our model we can predict that, if we have an identical device, but with y increased to 8%, the dilute nitride layer subband will empty at $\mathcal{E}_{\mathrm{BG},0} = 6719$ V cm^{-1}, which is a within the more practical -10^4 to 10^4 V cm^{-1} range. Our model then predicts that $E_F = 1.408$ eV. From our full numerical solution we find that the subband empties at $\mathcal{E}_{\mathrm{BG},0} = 8675$ V cm^{-1} and that $E_F = 1.403$ eV. Though there is a small discrepancy, it is evident that we can use our simple model to estimate how $\mathcal{E}_{\mathrm{BG},0}$ will vary for similar devices with different values of y.

A similar analysis applies to the field at which the subbands are degenerate, except that in this case, $\mathcal{E}_{\mathrm{FG}}$ and $\mathcal{E}_{\mathrm{BG}}$ must be varied oppositely. This is due to the fact that the point at which the subbands are degenerate depends equally strongly on both gate fields (as both subbands are occupied in this case), which is evident from Fig. 3. Conceptually, one can picture the effect of an increase in In content as causing a rigid shift of the plot in Fig. 3 along the diagonal from $\mathcal{E}_{\mathrm{FG}} = 3 \times 10^5$ V cm^{-1} and $\mathcal{E}_{\mathrm{BG}} = 5 \times 10^4$ V cm^{-1} to $\mathcal{E}_{\mathrm{FG}} = 4 \times 10^5$ V cm^{-1} and $\mathcal{E}_{\mathrm{BG}} = -5 \times 10^4$ V cm^{-1} by an amount $\Delta E_1 / z_{\mathrm{field}}$.

Even though our simple model is not applicable when both subbands are occupied, we can use it to estimate how the fields at which the subbands are degenerate varies with y, because the difference in $\mathcal{E}_{\mathrm{BG}}$ between where the dilute nitride layer subband empties and where the subbands are degenerate is independent of y. One can therefore use the variation in $\mathcal{E}_{\mathrm{BG},0}$ with y to determine how the fields at which the subbands are degenerate vary with y, keeping in mind that in this case $\mathcal{E}_{\mathrm{FG}} + \mathcal{E}_{\mathrm{BG}}$ must remain constant. For example, in our heterostructure presented in Sec. V A, degeneracy occurred when $\mathcal{E}_{\mathrm{FG}} = -1250$ V cm^{-1} and $\mathcal{E}_{\mathrm{FG}} = 3.5125 \times 10^5$ V cm^{-1}. Increasing y from 3% to 4%, we find that degeneracy now occurs at $\mathcal{E}_{\mathrm{FG}} = -7500$ V cm^{-1} and $\mathcal{E}_{\mathrm{FG}} = 3.575 \times 10^5$ V cm^{-1}, a change of 6250 V cm^{-1}. As stated previously, our simple model gives a change in field of 6391 V cm^{-1} for this change in y. We therefore see that our simple model can be used to estimate how the fields at which the subbands are degenerate varies for similar devices with different In content y.

For a given variation in gate field, the relative energies of the subband in the InGaAs layer and the subband in the dilute nitride layer will vary. This variation in relative energy depends on the distance between the two subbands along z, with larger distances giving larger variations, analogous to a lever arm. The distance between the subbands depends on the combined thickness of the InGaAs and GaAs layers. Therefore, if we decrease the InGaAs layer thickness, we decrease the range of relative energy variation for a given variation of gate field. This, in turn, means that a greater range of variation in gate field is required in order to bring the subbands close in energy so that they are degenerate or to empty the dilute nitride subband of carriers. Changing the energy of the subband in the InGaAs layer by varying the In content will then change the value of gate field at which degeneracy of the subbands, or the emptying of the dilute nitride layer subband, occurs by a larger amount. We find that the decrease (increase) in $\mathcal{E}_{\mathrm{BG}}$ ($\mathcal{E}_{\mathrm{FG}}$) per percentage increase in In content changes by $\sim$250 V cm^{-1} per nm decrease in InGaAs layer thickness.

This effect can also be determined from our simple model, by incorporating the change in InGaAs layer thickness in our determination of z_{field}. We then find that the change in field of $\sim$250 V cm^{-1} is reproduced quite well by our model, allowing us to predict how varying the InGaAs (or GaAs) layer thickness will change the effect varying y has on the fields at which degeneracy, or emptying of the dilute nitride layer subband, occurs.

When the gate fields are set to the value at which degeneracy between the subbands, or the emptying of the dilute nitride layer subband occurs, we predict major features in mobility measurements, as demonstrated in Sec. V. When the subbands are degenerate, we predict a large decrease in mobility, and when the dilute nitride layer subband is emptied of carriers, we predict a sudden increase in mobility. Our results on how the gate fields at which these features occur vary for similar devices with different values of y may be helpful in calibrating real devices.

4. Energy range of probe as a function of InGaAs layer thickness

The combined thicknesses of the InGaAs and GaAs channel layers determines the extent to which the variation in gate fields vary the relative energies between the InGaAs layer subband and the dilute nitride layer subband, as stated previously. If the dilute nitride layer subband is unoccupied, then the thicknesses of the layers determine the energy range swept through by E_F for a given variation in $\mathcal{E}_{\mathrm{BG}}$.

In the case of a heterostructure with $y = 5\%$, $x = 0.1\%$, $T = 4$ K, a constant $\mathcal{E}_{\mathrm{FG}} = 3.5 \times 10^5$ V cm^{-1}, and $d_{\mathrm{GaAs}} = 5$ nm, we have only the InGaAs layer subband occupied for a variation in $\mathcal{E}_{\mathrm{BG}}$ from -10^4 to 10^4 V cm^{-1}, and a variation in d_{In} from 10 to 24 nm. We find that the range of energy swept through by E_F for the given variation in $\mathcal{E}_{\mathrm{BG}}$ and d_{In} varies approximately linearly, with the range increasing from 26.7 meV at $d_{\mathrm{In}} = 10$ nm, to 42.0 meV at $d_{\mathrm{In}} = 24$ nm. Using our simple model given by Eq. (D5), we find that the range increases from 21.1 to 41.1 meV for the same variation in d_{In} and $\mathcal{E}_{\mathrm{BG}}$. Our model therefore gives reasonable predictions of the range scanned by the heterostructure, assuming that only the InGaAs layer subband is occupied.

It would appear that increasing d_{In} indefinitely will continue to increase our energy range variation for a given variation in gate field. However, the thickness of the InGaAs layer and GaAs layer also determines the extent to which the wave function of the InGaAs layer subband overlaps with the dilute nitride layer, with larger thicknesses giving lower overlap, for a given N content. If the overlap becomes too small, the reductions in mobility as E_F is resonant with a localized state become too small to be detected. There is therefore a trade-off

in setting the combined thicknesses of the InGaAs and GaAs layers, so that we have as large an extent in variation in relative energies of the subband with the N states as the back-gate field is varied, for an optimum overlap of the subband with the dilute nitride layer. We have therefore chosen the range 15 to 25 nm for the combined thickness of the layers, with the GaAs layer kept at 5 nm thick, as explained previously.

From these results it is evident that our simple model can be used to estimate how the localized state energy range scanned by a device will vary for similar devices with different values of d_{In}. The model will be applicable as only the InGaAs layer subband will be occupied (to a good approximation) if the device can probe the localized states.

5. Summary

We have presented a method of determining, once a device is grown of certain dimensions and $\mathcal{E}_{\mathrm{BG},0}$ is measured, how $\mathcal{E}_{\mathrm{BG},0}$ and the range and position of the portion of the spectrum of localized states probed by the device will change, for similar devices grown with different In contents and InGaAs layer widths. We stress that the range of localized states that may be probed is always restricted by the heterostructure design, and that localized N states in the dilute nitride layer that lie above the CBE are, in general, inaccessible to our probing device. This method will be useful for the design of experiments to probe these localized states in the dilute nitrides using our proposed heterostructure device.

[1] M. Weyers, M. Sato, and H. Ando, Jpn. J. Appl. Phys. **31**, L853 (1992).
[2] H. Riechert, A. Ramakrishnan, and G. Steinle, Semicond. Sci. Technol. **17**, 892 (2002).
[3] J. S. Harris, Semicond. Sci. Technol. **17**, 880 (2002).
[4] J. F. Geisz and D. J. Friedman, Semicond. Sci. Technol. **17**, 769 (2002).
[5] D. L. Young, J. F. Geisz, and T. J. Coutts, Appl. Phys. Lett. **82**, 1236 (2003).
[6] R. Mouillet *et al.*, Solid State Commun. **126**, 333 (2003).
[7] D. Fowler, O. Makarovsky, A. Patane, L. Eaves, L. Geelhaar, and H. Riechert, Phys. Rev. B **69**, 153305 (2004).
[8] A. J. Ptak, S. W. Johnston, S. Kurtz, D. J. Friedman, and W. K. Metzger, J. Cryst. Growth **251**, 392 (2003).
[9] W. Shan, W. Walukiewicz, J. W. Ager, E. E. Haller, J. F. Geisz, D. J. Friedman, J. M. Olson, and S. R. Kurtz, Phys. Rev. Lett. **82**, 1221 (1999).
[10] H. P. Hjalmarson, P. Vogl, D. J. Wolford, and J. D. Dow, Phys. Rev. Lett. **44**, 810 (1980).
[11] X. Liu, M. Pistol, L. Samuelson, S. Schwetlick, and W. Seifert, Appl. Phys. Lett. **56**, 1451 (1990).
[12] P. R. C. Kent, L. Bellaiche, and A. Zunger, Semicond. Sci. Technol. **17**, 851 (2002).
[13] L. Bellaiche, S.-H. Wei, and A. Zunger, Phys. Rev. B **54**, 17568 (1996).
[14] E. P. O'Reilly, A. Lindsay, S. Tomić, and M. Kamal-Saadi, Semicond. Sci. Technol. **17**, 870 (2002).
[15] A. Lindsay and E. P. O'Reilly, Phys. Rev. Lett. **93**, 196402 (2004).
[16] F. Masia *et al.*, Phys. Rev. B **73**, 073201 (2006).
[17] T. Dannecker, Y. Jin, H. Cheng, C. F. Gorman, J. Buckeridge, C. Uher, S. Fahy, C. Kurdak, and R. S. Goldman, Phys. Rev. B **82**, 125203 (2010).
[18] S. Fahy and E. P. O'Reilly, Appl. Phys. Lett. **83**, 3731 (2003).
[19] O. F. Sankey, J. D. Dow, and K. Hess, Appl. Phys. Lett. **41**, 664 (1982).
[20] J. Wu, K. Yu, and W. Walukiewicz, IEEE Proc.: Optoelectron. **151**, 460 (2004).
[21] S. Fahy, A. Lindsay, and E. P. O'Reilly, IEEE Proc.: Optoelectron. **151**, 352 (2004).
[22] S. Fahy, A. Lindsay, H. Ouerdane, and E. P. O'Reilly, Phys. Rev. B **74**, 035203 (2006).
[23] J. Endicott, A. Patane, J. Ibanez, L. Eaves, M. Bissiri, M. Hopkinson, R. Airey, and G. Hill, Phys. Rev. Lett. **91**, 126802 (2003).
[24] M. J. Kelly, *Low-Dimensional Semiconductors* (Clarendon Press, Oxford, 1995), Chap. 5.
[25] A. Lindsay and E. P. O'Reilly, Solid State Commun. **118**, 313 (2001).
[26] A. Lindsay and E. P. O'Reilly, Physica E **21**, 901 (2004).
[27] G. Sęk *et al.*, Vacuum **50**, 219 (1998).
[28] S. Adachi, J. Appl. Phys. **58**, R1 (1985).
[29] J. H. Davies, *The Physics of Low-Dimensional Semiconductors* (Cambridge University Press, Cambridge, 1998), Chap. 3.
[30] J. Wu, W. Walukiewicz, and E. E. Haller, J. Appl. Phys. **89**, 789 (2001).
[31] M. Kondow *et al.*, Appl. Phys. Express **2**, 041003 (2009).
[32] D. Vasileska, C. Prasad, H. H. Wieder, and D. K. Ferry, J. Appl. Phys. **93**, 3359 (2003).
[33] M. Oloumi and C. C. Matthai, J. Phys. Condens. Matter **3**, 9981 (1991).
[34] D. E. Aspnes, Phys. Rev. B **14**, 5331 (1976).
[35] M. Levinshtein, S. Rumyantsev, and M. Shur, *Handbook Series on Semiconductor Parameters* (World Scientific, London, 1996), Vol. 1, pp. 77–103.
[36] D. J. Chadi and K. J. Chang, Phys. Rev. Lett. **61**, 873 (1988).
[37] J. J. Sakurai, *Modern Quantum Mechanics* (Addison-Wesley, Reading, MA, 1994), Chap. 5.
[38] T. Saku, Y. Horikoshi, and Y. Tokura, Jpn. J. Appl. Phys. **35**, 34 (1996).
[39] G. A. C. Jones, P. D. Rose, E. H. Linfield, and D. A. Ritchie, Physica B **227**, 264 (1996).