

Optical response of grating-coupler-induced intersubband resonances: The role of Wood's anomalies

L. Wendler

Anna-Siemsen-Straße 66, D-07745 Jena, Germany

T. Kraft, M. Hartung, A. Berger, and A. Wixforth

Sektion Physik, Ludwigs-Maximilians-Universität München, Geschwister-Scholl-Platz 1, D-80539 München, Germany

M. Sundaram, J. H. English, and A. C. Gossard

Materials Department, University of California, Santa Barbara, California 93106

(Received 18 April 1996; revised manuscript received 12 August 1996)

Grating-coupler-induced collective intersubband transitions in a quasi-two-dimensional electron system are investigated both experimentally and theoretically. Far-infrared transmission experiments are performed on samples containing a quasi-two-dimensional electron gas quantum confined in a parabolic quantum well. For rectangular-shaped grating couplers of different periods we observe a strong dependence of the transmission line shape and peak height on the period of the grating, i.e., on the wave-vector transfer from the diffracted beams to the collective intersubband resonance. It is shown that the line shape transforms with increasing grating period from a Lorentzian into a strongly asymmetric line shape. Theoretically, we treat the problem by using the transfer-matrix method of local optics and apply the modal-expansion method to calculate the influence of the grating. The optically uniaxial quasi-two-dimensional electron gas is described in the long-wavelength limit of the random-phase approximation by a local dielectric tensor, which includes size quantization effects. Our theory reproduces excellently the experimental line shapes. The deformation of the transmission line shapes we explain by the occurrence of both types of Wood's anomalies. [S0163-1829(97)05403-9]

I. INTRODUCTION

The optical and electronic properties of electron systems with reduced dimensionality, such as quasi-two-dimensional electron gases (Q2DEG's) formed at, e.g., a modulation-doped semiconductor heterojunction have been widely studied in the recent past. Especially the collective excitation spectrum of such a system has attracted a lot of attention as it represents one of the Q2DEG's most fundamental properties,¹ but also because of potential device applications.²⁻⁴ Here, both types of collective charge-density excitations, *intrasubband* as well as *intersubband plasmons* in the far-infrared (FIR) regime, have been intensively studied for the last twenty years (see, e.g., Ref. 5).

It is known (see, e.g., Ref. 6) that the intrasubband plasmons always exist for wave vectors \mathbf{q}_{\parallel} larger than the wave vector of the freely propagating light, $q_{\parallel} > \omega/c$, where $\mathbf{q}_{\parallel} = (q_x, q_y)$ is the in-plane wave vector ($q_{\parallel} = |\mathbf{q}_{\parallel}|$) of the collective excitation [c denotes the vacuum speed of light and ω is the (angular) frequency], assuming that interfaces of the sample are parallel to the x - y plane. But differently, each intersubband excitation is accompanied by two branches of dispersion curves due to the polariton effect.⁷ One branch is located to the left of the light line ($q_{\parallel} < \omega/c$) and thus is the dispersion relation of a radiative virtual mode, and the second, appearing to the right of the light line $q_{\parallel} > \omega/c$, describes a nonradiative normal mode. Commonly, one calls the Q2D radiative virtual intersubband modes the collective intersubband resonances (ISR's), collective intersubband

transitions, or dimensional resonances and the nonradiative normal modes the intersubband plasmons.

Unfortunately, the investigation of the mode dispersion of the Q2D plasmons $\omega_p^{\text{Q2D}}(q_{\parallel})$ is not directly accessible in FIR spectroscopy. In general, the Q2D plasmons are accompanied by different collective intrasubband and intersubband transitions due to the intersubband coupling (ISC). In this case, the resulting spectrum is of hybrid type. Only for weak ISC nearly pure intrasubband and intersubband plasmons occur. In such a situation, intrasubband plasmons can only be excited with fields polarized parallel to the heterointerfaces and that have wave vectors $q_{\parallel} > \omega/c$, whereas ISR's and intersubband plasmons can only be excited with electromagnetic fields having components polarized perpendicular to the interfaces of the sample^{8,9} and that have wave vectors $q_{\parallel} < \omega/c$ and $q_{\parallel} > \omega/c$, respectively. To solve this problem, usually the FIR radiation is coupled to the Q2DEG through a metallic grating of period d above the electron system.⁹⁻¹¹ This way, different discrete values of the probe wave vector $\mathbf{k}_{\parallel n} = (k_{xn}, k_y)$ of the diffracted electromagnetic field parallel to the 2D plane, $k_{xn} = (\omega/c) \sin \Theta_0 + (2\pi/d)n$; $n = 0, \pm 1, \pm 2, \dots$ (Θ_0 is the ray angle of the incident light measured from the z axis, which is assumed to be perpendicular to the interfaces of the sample), can be excited. Assuming the incident light propagation in the x - z plane and the stripes of the grating along the y axis, the incident light couples at the discrete wave vectors $q_{\parallel} = k_{xn}$ to the collective excitations $\omega_p^{\text{Q2D}}(q_{\parallel})$ in the Q2DEG, provided the grating is sufficiently close to the electron gas. Here, the grating cou-

pler serves as a special optical coupling arrangement: (i) to convert the in-plane electromagnetic field component $E_x(x, z|\omega)$ of the incident wave into the perpendicular component $E_z(x, z|\omega)$ of the scattered waves and (ii) to induce a wave-vector transfer $k_x \rightarrow k_{xn}$ parallel to the interfaces of the system. Moreover, grating couplers provide unique advantages over other techniques such as the use of prisms,¹² Brewster angle orientation,¹³ or multiple internal reflection (waveguide) geometry:¹⁴ normal incidence of the FIR radiation and a relatively large component of the electric field along the direction of confinement can be easily achieved.

The aim of this paper is to investigate both experimentally and theoretically the influence of the grating coupler on the coupling efficiency and the resulting transmission line shape of the Q2D ISR as a function of the grating coupler period. Both intrasubband plasmons (see, e.g., Refs. 9–11 for experiments and Refs. 15–19 for theory) and ISR's (see, e.g., Refs. 20–23 for earlier experiments and Refs. 24 and 25 for theory) are intensively investigated. Here, we wish to focus on the excitation of the ISR. Unfortunately, the theories developed up to now on the grating-coupler-induced excitation of collective Q2D plasmon modes are restricted to strong approximations, e.g., optically isotropic media, 2D gratings, perfectly conducting gratings, simple half-space geometry, etc., which are unsuitable to describe correctly the optical response of the Q2DEG synthesized in multilayer systems with grating. Moreover, as we will point out below, the influence of the grating coupler on the line shape is much more pronounced for the ISR than for the intrasubband plasmon and thus it needs to be considered carefully. Because the perpendicular field component is zero for the zeroth-order beam, the grating-coupler-induced ISR's are nonvertical transitions in \mathbf{k} space, i.e., involve wave-vector transfers of $k_{xn}, n = \pm 1, \pm 2, \dots$. Experiments^{26,27} indicate that the power absorbed at the ISR frequency is a function of the ratio of the grating period d to the intersubband resonance wavelength λ_{sub} . Our theoretical investigations are completely universal in the framework of local optics and can be applied to any given optically uniaxial multilayered structure with gratings of finite height.

II. EXPERIMENT

The experiments are performed on parabolic quantum well's (PQW's).²⁸ Here, the conduction-band edge of a GaAs-Ga_{1-x}Al_xAs quantum well is graded such that it turns out to have a parabolic shape in the growth direction. This is achieved by a proper variation of the percentage of one of the ingredients (namely the Al content x). An electron that is experiencing this profile “sees” a potential that could also result from a homogenous positive background such as a doping layer. From Poisson's equation and a back-of-the-envelope calculation the potential of this (in the case of a PQW fictitious) positive background with density n_+ can be related to the growth parameters of the PQW by

$$n_+ = \frac{\epsilon_0 \epsilon_s}{e^2} \frac{d^2 E_C}{dz^2} = \frac{\epsilon_0 \epsilon_s}{e^2} \frac{8\Delta}{W^2}. \quad (1)$$

Here, ϵ_s denotes the mean static dielectric constant of Ga_{1-x}Al_xAs forming the PQW, ϵ_0 is the dielectric permit-

tivity in vacuum, Δ is the energy height of the parabola from its bottom to the edges, the electron charge is $-e$ and W is the width of the grown PQW. Once such a structure is remotely doped, the donors release electrons into the well, which in turn will screen the man-made parabolic potential and form a wide and nearly homogeneous electron layer.

The collective response of such systems has been the subject of many different experiments and theoretical investigations over the last few years (see, e.g., Ref. 29). The most surprising effect that has been observed in the early days of the spectroscopy of PQW's is that the electron system absorbs radiation independently from the electron density in the PQW only at a single well-defined frequency, which can be related to the shape of the bare confining potential alone. Self-consistent effects such as the renormalization of the subband energies in the structures due to electron-electron interactions seem to turn out to be of no importance for the optical absorption. In a celebrated paper Brey, Johnson, and Halperin³⁰ and Yip³¹ generalized Kohn's theorem,³² originally derived for the cyclotron resonance in 3D bulk semiconductor structures. They showed that for the case of a bare perfectly parabolic confinement potential long-wavelength radiation only couples to the center-of-mass (CM) part of the Hamiltonian, leaving the relative coordinates completely untouched. The reason for this unique phenomenon is that for a bare harmonic confinement and only for this the Hamiltonian can be separated into one part containing only CM coordinates and one part containing only relative coordinates. The interaction Hamiltonian for the incident light can be shown to couple only to the CM part. In other words, long-wavelength radiation excites a mode in the electron system, which is only connected with the CM motion. This collective intersubband mode is sometimes referred to as Kohn's mode or more figurative “sloshing mode.” Per construction, it corresponds to the classical plasma oscillations of a bulk 3D electron gas of density n_+ : $\omega_p = [n_+ e^2 / (\epsilon_0 \epsilon_s m_e)]^{1/2}$, where m_e is the effective conduction-band-edge mass. If we use Eq. (1) relating n_+ to the growth parameters it reads

$$\omega_p = \Omega = \left(\frac{8\Delta}{W^2 m_e} \right)^{1/2}, \quad (2)$$

where Ω is the confining frequency of the bare parabolic potential: $V_0(z) = \frac{1}{2} m_e \Omega^2 z^2$.

The many-particle picture of this phenomenon^{33,34} is that the renormalization of the (1-0) subband separation frequency $\Delta\Omega_{10} = \Omega_{10}(0) - \Omega_{10}(n_{2\text{DEG}})$, where $\Omega_{10}(n_{2\text{DEG}}) = (\mathcal{E}_1 - \mathcal{E}_0)/\hbar$ is the subband separation frequency of the PQW with the 2D electron density $n_{2\text{DEG}}$ (sheet carrier concentration), cancels for $q_{\parallel} = 0$ with the collective frequency shift $\Delta_p^{10} = \omega_p^{10}(q_{\parallel} = 0) - \Omega_{10}(n_{2\text{DEG}})$ of the (1-0) intersubband plasmon $\omega_p^{10}(q_{\parallel})$. It is shown^{33,34} that in such a situation this mode is pinned at $\omega_p^{10}(q_{\parallel} = 0) = \Omega_{10}(0) = \Omega$ independently from the electron density. Whereas Kohn's theorem states that the absorption spectrum has only one peak at $\omega = \Omega$, the mode spectrum of the freely oscillating Q2DEG consists of all types of Q2D plasmons. It should be noted that for $q_{\parallel} \neq 0$ Kohn's theorem is not valid.

For our experiments we use different samples that have been cut from one single wafer. On top of these samples we fabricate metal gratings of different period d and then com-

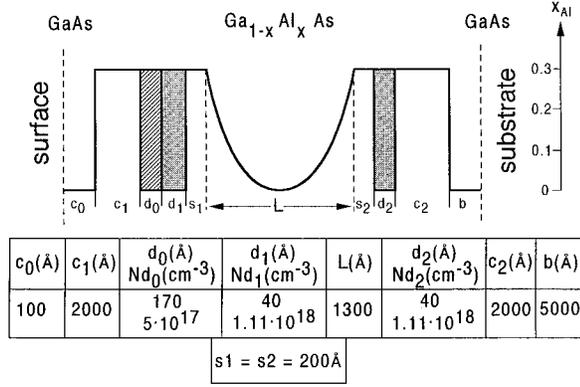


FIG. 1. Schematic arrangement of the layer structure of the sample containing the PQW used in the experiments.

pare their transmission spectra. Any differences in the spectra obtained for the different samples thus can be related to the effect of the different grating couplers. The sample used in our experiments is a standard PQW (sample PB48 of Ref. 29), schematically drawn in Fig. 1, in which the Al content x was varied during growth between $x=0$ in the center of the well and $x=0.3$ at its edges. To grade the structure, we used the digital alloy technique as described elsewhere.²⁸ This structure is grown on a semi-insulating GaAs substrate and consists of the following layers: layer 1 is the 10-nm-thick undoped GaAs cap layer (c_0), layer 2 is a 200-nm-thick undoped $\text{Ga}_{0.7}\text{Al}_{0.3}\text{As}$ layer (c_1), layer 3 is a 17-nm-thick Si-doped ($N_d = 5 \times 10^{17} \text{ cm}^{-3}$) $\text{Ga}_{0.7}\text{Al}_{0.3}\text{As}$ layer (d_0), layer 4 is a 4-nm-thick Si-doped ($N_d = 1.11 \times 10^{18} \text{ cm}^{-3}$) $\text{Ga}_{0.7}\text{Al}_{0.3}\text{As}$ layer (d_1), layer 5 is the 20-nm-thick undoped $\text{Ga}_{0.7}\text{Al}_{0.3}\text{As}$ spacer layer (s_1), layer 6 is the 130-nm-thick $\text{Ga}_{1-x}\text{Al}_x\text{As}$ PQW (L), layer 7 is the 20-nm-thick undoped $\text{Ga}_{0.7}\text{Al}_{0.3}\text{As}$ spacer layer (s_2), layer 8 is a 4-nm-thick Si-doped ($N_d = 1.11 \times 10^{18} \text{ cm}^{-3}$) $\text{Ga}_{0.7}\text{Al}_{0.3}\text{As}$ layer (d_2), layer 9 is a 200-nm-thick undoped $\text{Ga}_{0.7}\text{Al}_{0.3}\text{As}$ layer (c_2), and layer 10 is the 500-nm-thick undoped GaAs buffer layer (b). The substrate is a 500- μm -thick GaAs wafer. From the growth parameters we calculate the density of the fictitious charge of $n_+ = 7.4 \times 10^{16} \text{ cm}^{-3}$, which according to Eq. (2) corresponds to an expected resonance energy of about $\hbar\omega_p(q_{\parallel}=0) \approx 11 \text{ meV}$. A semitransparent NiCr electrode on top of the sample serves as a gate to vary the carrier density in the well and alloyed In pellets at its corners provide Ohmic contacts to the electron system. On top of the NiCr gate a 50-nm-thick rectangular Ag grating with stripes along the y axis of width a , spacing b between the stripes and period $d = a + b$ is deposited. For the different samples investigated here, we used grating periods of $d = 4, 6, 10, 20, 27, 40$, and $80 \mu\text{m}$, respectively. The metallization for all gratings was chosen so that the mark-to-space ratio (aspect ratio) is close to $t = a/b = 1$ (one also introduces the mark fraction $f = a/d$ and the open space fraction $r = b/d$). For comparison, also one sample without any grating coupler has been fabricated. To avoid Fabry-Pérot-type interference effects, the sample substrate is wedged by a small angle of approximately 3° .

Magnetotransport measurements revealed a typical sheet carrier concentration of $n_{2\text{DEG}} = 2 \times 10^{11} \text{ cm}^{-2}$ for all samples. To be able to directly compare the experimental

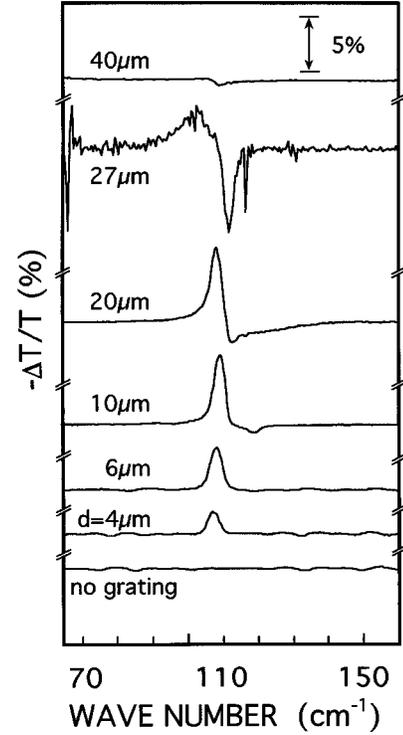


FIG. 2. Experimentally obtained relative transmission spectra for the ISR in a PQW using grating couplers of different grating periods. With increasing grating period first the height of the occurring peak becomes larger, then the lines become strongly asymmetric, and for large grating periods the relative transmission nearly vanishes.

results for different samples we made sure that during the actual measurement of the FIR transmission for all samples $n_{2\text{DEG}}$ was exactly the same by eventually applying a small correction bias to the gate electrode.³⁵

Our transmission experiments are performed at low temperatures ($T = 2 \text{ K}$) with the sample mounted in the center of a superconducting solenoid. Experimentally, we determine the relative change in transmission $-\Delta T/T = [T(0) - T(n_{2\text{DEG}})]/T(0)$. $T(0)$ is the transmission of the sample with the well being completely depleted, $T(n_{2\text{DEG}})$ the one at finite carrier densities. The spectra have been taken with a Fourier transform spectrometer and a Si composite bolometer was used to detect the transmitted radiation, which is guided by a 10-in-long oversized brass waveguide between the sample and the bolometer.

In Fig. 2 we depict the results of our experiments, where we plot the relative transmission $-\Delta T/T$ for seven samples, investigated as a function of the frequency for zero magnetic field. As expected from the dipole selection rules, there is no detectable absorption at the frequency of the ISR for the sample without grating. With increasing grating period a peak develops at the expected resonance frequency of the ISR. Its line shape and peak height, however, strongly depend on the grating used in the experiment. For small grating periods $d < \lambda_{\text{sub}}$, where $\lambda_{\text{sub}} = 2\pi c / (\sqrt{\epsilon_s} \omega)$ is the wavelength of the FIR radiation at the ISR frequency in the GaAs substrate with static dielectric constant ϵ_s , the peak has a Lorentzian shape and its height increases with increasing period. For $d = \lambda_{\text{sub}}$, however, the line shape becomes totally

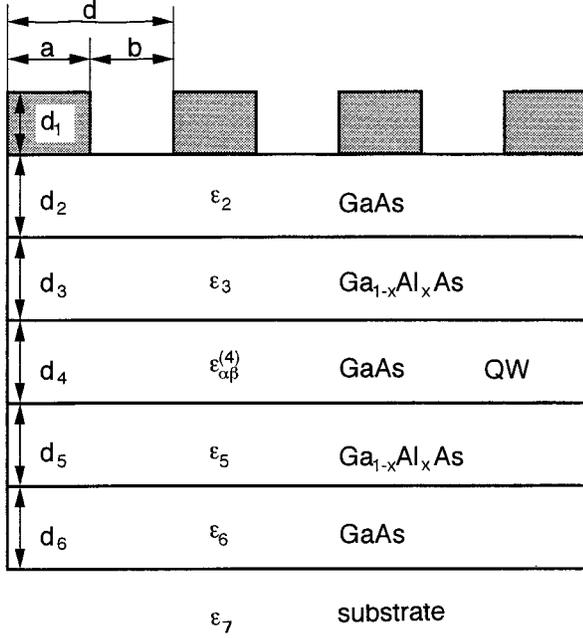


FIG. 3. Schematic arrangement of the geometry of the multilayer system with grating used in the theoretical calculations.

distorted, indicating some interference effect. For even larger grating periods $d > \lambda_{\text{sub}}$ the relative transmission rapidly decreases. For $d = 40 \mu\text{m}$, the resonance is barely detectable and for a grating with period $d = 80 \mu\text{m}$, no detectable resonance is left. It is also interesting to note that the sign of the resonant structure changes when the grating period crosses the condition $d = \lambda_{\text{sub}}$. Whereas for $d < \lambda_{\text{sub}}$ the maximum of $-\Delta T/T$ is positive, for $d > \lambda_{\text{sub}}$ the maximum change in transmission becomes negative, i.e., it becomes a minimum.

III. THEORY

For the theoretical investigation of the FIR transmission spectra of the PQW structure with grating coupler we use our theory, developed in Ref. 36, which is based on the transfer-matrix method of local optics for anisotropic media and apply the modal-expansion method to calculate the influence of the grating. Any details of the theory may be found in this paper. The combination of both methods results in a generally computationally efficient and stable formalism of the optical response of multilayer systems with grating.³⁶ As shown in Fig. 3 the PQW under consideration is modeled by a six-layer system $\nu = 1, \dots, 6$. Each layer is, in general, characterized by its dielectric tensor $\varepsilon_{\alpha\beta}^{(\nu)}(\mathbf{x}, \omega)$, where $\alpha, \beta = x, y, z$, and by its thickness $d_\nu = |z_\nu - z_{\nu-1}|$. The layer $z_1 < z < z_0$ contains the rectangular-groove grating of height $h \equiv d_1$ and periodicity $d = a + b$. In the grating region we have for the filled stripes $\varepsilon_{\alpha\beta}^{(1)}(\mathbf{x}, \omega) = \varepsilon_\xi(\omega) \delta_{\alpha\beta}$, where $\xi = a$ if $md < x < md + a$ and $\xi = b$ if $md + a < x < (m+1)d$, with $m = 0, \pm 1, \pm 2, \dots$. In the experiments samples with metal gratings are used. In this case we use the local Drude dielectric function $\varepsilon_a(\omega) = 1 - \omega_p^2 / \omega(\omega + i\gamma_a)$ for the metallic stripes, where ω_p is the plasma frequency and γ_a is the phenomenological damping constant, and we assume for the spacing between the stripes $\varepsilon_b = 1$. The semiconductor

layers of the quantum-well structure are described by two different types of dielectric tensors: (i) $\varepsilon_{\alpha\beta}^{(\nu)}(\mathbf{x}, \omega) = \varepsilon_\nu(\omega) \delta_{\alpha\beta}$ in $z_\nu < z < z_{\nu-1}$, if the ν th layer is filled by an isotropic medium and (ii) $\varepsilon_{\alpha\beta}^{(\nu)}(\mathbf{x}, \omega) = \varepsilon_{\alpha\beta}^{(\nu)}(\omega) \delta_{\alpha\beta}$ is a diagonal tensor if this layer forms the QW and contains the Q2DEG, which behaves optically uniaxial. This is true in the absence of an external magnetic field, where we have $\varepsilon_{xx}^{(\nu)} = \varepsilon_{yy}^{(\nu)} \neq 0$, $\varepsilon_{zz}^{(\nu)} \neq 0$ but $\varepsilon_{xy}^{(\nu)} = \varepsilon_{yx}^{(\nu)} = \varepsilon_{xz}^{(\nu)} = \varepsilon_{zx}^{(\nu)} = \varepsilon_{yz}^{(\nu)} = \varepsilon_{zy}^{(\nu)} = 0$. For the dielectric properties of the background of the GaAs and $\text{Ga}_{1-x}\text{Al}_x\text{As}$ layers we use the so-called ε_s approximation: $\varepsilon_\nu(\omega) = \varepsilon_{s\nu}$, i.e., we neglect the dynamical properties of the optical phonons, but include their screening by the static dielectric constant $\varepsilon_{s\nu}$. This is true because the frequencies of the optical phonons are well above the frequency of the ISR in the studied PQW's. That layer that contains the Q2DEG we describe by a macroscopic local dielectric tensor, which includes the size-quantization effects on the electrons in the PQW. The dielectric tensor of the Q2DEG (nonlocal) is derived in the framework of the random-phase approximation of the current-response scheme in Ref. 6, wherefrom the nonvanishing components in the optical limit ($q_{\parallel} \rightarrow 0$) follow in the form

$$\varepsilon_{xx}^{(\nu)}(\omega) = \varepsilon_{yy}^{(\nu)}(\omega) = \varepsilon_{s\nu} \left(1 - \frac{\omega_0^2}{\omega(\omega + i/\tau_{\parallel})} \right), \quad (3)$$

$$\varepsilon_{zz}^{(\nu)}(\omega) = \varepsilon_{s\nu} \left(1 - \frac{\omega_0^2 f_{10}}{\omega^2 - \Omega_{10}^2 + i\omega/\tau_{\perp}} \right), \quad (4)$$

where

$$f_{10} = \frac{2m_e \Omega_{10}}{\hbar} z_{10}^2 \quad (5)$$

with

$$z_{10} = \int_0^{a_{2\text{DEG}}} dz \varphi_1^*(z) z \varphi_0(z). \quad (6)$$

Herein, $\varphi_K(z)$, $K = 0, 1, 2, \dots$ is the envelope wave function of the PQW and we have defined the plasma frequency by $\omega_0 = [n_{2\text{DEG}} e^2 / (m_e \varepsilon_0 \varepsilon_{s\nu} a_{2\text{DEG}})]^{1/2}$, where $a_{2\text{DEG}}$ is the effective layer thickness of the Q2DEG. Further, τ_{\parallel} and τ_{\perp} are the phenomenological longitudinal and transverse relaxation times, respectively, $\Omega_{10} = \Omega_{10}(n_{2\text{DEG}}) = (\mathcal{E}_1 - \mathcal{E}_0)/\hbar$ is the subband separation frequency of the two lowest electric subbands of the effective confining potential (non-parabolic) of the PQW and f_{10} is the oscillator strength of the transition $0 \rightarrow 1$. For the PQW under consideration the subband (bottom) energies \mathcal{E}_K , the envelope wave functions $\varphi_K(z)$, the Fermi energy E_F and the oscillator strengths $f_{KK'}$ are calculated self-consistently in the framework of the Hartree approximation using the method described in our recent paper.³⁴

Assuming monochromatic electric and magnetic fields, $\mathbf{E}(\mathbf{x}, t) = \mathbf{E}(\mathbf{x}, \omega) \exp(-i\omega t)$ and $\mathbf{H}(\mathbf{x}, t) = \mathbf{H}(\mathbf{x}, \omega) \exp(-i\omega t)$, respectively, these fields are given by the wave equations

$$\nabla[\nabla \cdot \mathbf{E}(\mathbf{x}, \omega)] - (\nabla \cdot \nabla) \mathbf{E}(\mathbf{x}, \omega) = \frac{\omega^2}{c^2} \vec{\varepsilon}(\mathbf{x}, \omega) \mathbf{E}(\mathbf{x}, \omega), \quad (7)$$

and

$$\begin{aligned} & \nabla[\nabla \cdot \mathbf{H}(\mathbf{x}, \omega)] - (\nabla \cdot \nabla) \mathbf{H}(\mathbf{x}, \omega) \\ & = -i\omega \varepsilon_0 \nabla \times [\vec{\varepsilon}(\mathbf{x}, \omega) \mathbf{E}(\mathbf{x}, \omega)]. \end{aligned} \quad (8)$$

We assume that the incident plane wave travels in the half-space $z > z_0$ ($\nu=0$), filled by vacuum, within the x - z plane in the negative z direction with the wave vector $\mathbf{k}^{(0)} = (k_x, 0, -k_z^{(0)})$, having the components $k_x = (\omega/c) \sin \Theta_0$ and $k_z^{(0)} = (\omega/c) \cos \Theta_0$. In this case the here considered p polarization (TM waves) with

$$\mathbf{E}(x, z | \omega) = [E_x(x, z | \omega), 0, E_z(x, z | \omega)] \quad (9)$$

and

$$\mathbf{H}(x, z | \omega) = [0, H_y(x, z | \omega), 0] \quad (10)$$

is independent from the s polarization (TE waves).

According to the Floquet-Bloch theorem we have in the grating region $\mathbf{E}^{(1)}(x + md, z | \omega) = \exp(ik_x md) \mathbf{E}^{(1)}(x, z | \omega)$, where the Bloch wave-vector component k_x is defined in the first Brillouin zone: $-\pi/d < k_x \leq \pi/d$. Whereas in a multilayer system of homogeneous layers $\mathbf{k}_{\parallel} = (k_x, k_y)$ is a conserved quantity of the whole system, the periodic structure of the grating coupler produces an infinite number of propagating waves with $\mathbf{k}_{\parallel n} = (k_{xn}, k_y)$, where $k_{xn} = k_x + G_n$ and $G_n = (2\pi/d)n$, $n=0, \pm 1, \pm 2, \dots$ is the reciprocal lat-

tice vector. In the presence of the grating the reflected and the transmitted beams are represented by Fourier series (Rayleigh expansion³⁷):

$$\begin{aligned} \mathbf{E}(x, z | \omega) &= \sum_{n=-\infty}^{\infty} \exp(ik_{xn}x) \mathbf{E}_n(z, \omega), \\ & \text{if } z > z_0 \text{ and } z < z_1. \end{aligned} \quad (11)$$

This ansatz has to fulfill the wave equation in each layer. In the region $z > z_0$ we have

$$\begin{aligned} \mathbf{E}^{(0)}(x, z | \omega) &= \sqrt{\frac{\varepsilon_0}{\mu_0}} \left\{ A_p^{(0)} \exp[i(k_x x - k_z^{(0)} z)] (k_z^{(0)} \mathbf{e}_x + k_x \mathbf{e}_z) \right. \\ & + \sum_{n=-\infty}^{\infty} B_{pn}^{(0)} \exp[i(k_{xn} x + k_{zn}^{(0)} z)] \\ & \left. \times (-k_{zn}^{(0)} \mathbf{e}_x + k_{xn} \mathbf{e}_z) \right\}, \end{aligned} \quad (12)$$

where $A_p^{(0)}$ is the field amplitude of the incident wave, $B_{pn}^{(0)}$ is the amplitude of the n th diffraction-order reflected wave, and \mathbf{e}_α is the unit vector along the α axis. From the dispersion relation in vacuum $k_n^{(0)} = |\mathbf{k}_n^{(0)}| = \omega/c$ it follows that $k_{zn}^{(0)} = [\omega^2/c^2 - k_{xn}^2]^{1/2}$. For a layer filled by an isotropic semiconductor we have

$$\begin{aligned} \mathbf{E}^{(\nu)}(x, z | \omega) &= \sqrt{\frac{\varepsilon_0}{\mu_0 \varepsilon_\nu(\omega)}} \sum_{n=-\infty}^{\infty} \exp(ik_{xn}x) \{ A_{pn}^{(\nu)} \exp[-ik_{zn}^{(\nu)}(z - z_{\nu-1})] (k_{zn}^{(\nu)} \mathbf{e}_x + k_{xn} \mathbf{e}_z) \\ & + B_{pn}^{(\nu)} \exp[ik_{zn}^{(\nu)}(z - z_{\nu-1})] (-k_{zn}^{(\nu)} \mathbf{e}_x + k_{xn} \mathbf{e}_z) \}, \end{aligned} \quad (13)$$

where $k_{zn}^{(\nu)} = [\varepsilon_\nu(\omega) \omega^2/c^2 - k_{xn}^2]^{1/2}$ is valid. $A_{pn}^{(\nu)}$ is the amplitude of the n th-order diffracted wave propagating downwards and $B_{pn}^{(\nu)}$ of that propagating upwards in layer ν . For the anisotropic layer containing the Q2DEG it follows that

$$\begin{aligned} \mathbf{E}^{(\nu)}(x, z | \omega) &= \sqrt{\frac{\varepsilon_0}{\mu_0 \varepsilon_{xx}^{(\nu)}(\omega)}} \sum_{n=-\infty}^{\infty} \exp(ik_{xn}x) \{ A_{pn}^{(\nu)} \exp[-ik_z^{(\nu)}(z - z_{\nu-1})] (k_{zn}^{(\nu)} \mathbf{e}_x + k_{xn} \mathbf{e}_z) \\ & + B_{pn}^{(\nu)} \exp[ik_{zn}^{(\nu)}(z - z_{\nu-1})] (-k_{zn}^{(\nu)} \mathbf{e}_x + k_{xn} \mathbf{e}_z) \} \end{aligned} \quad (14)$$

with $k_{zn}^{(\nu)} = [\varepsilon_{xx}^{(\nu)}(\omega) \omega^2/c^2 - \varepsilon_{xx}^{(\nu)} k_{xn}^2 / \varepsilon_{zz}^{(\nu)}]^{1/2}$.

In the grating region we represent the fields by the modal-expansion method,^{38,39} i.e., represent the electromagnetic fields as a sum of the eigensolutions of the wave equation in the grating layer. One solves the wave equation in each stripe of the grating and requires the boundary conditions $\llbracket E_z \rrbracket = 0$ and $\llbracket H_y \rrbracket = 0$, where $\llbracket A \rrbracket$ denotes the change of A evaluated at the interface. This results in the dispersion relation of the modes

$$\frac{(\varepsilon_a \beta_b)^2 + (\varepsilon_b \beta_a)^2}{2 \varepsilon_a \varepsilon_b \beta_a \beta_b} \sin(\beta_a a) \sin(\beta_b b) - \cos(\beta_a a) \cos(\beta_b b) + \cos(k_x d) = 0. \quad (15)$$

Equation (15) determines for a given pair (ω, k_x) a set of eigenvalues $\{k_{zl}^{(1)}\}$. The electromagnetic field in the grating layer is the sum over all eigenfunctions:

$$\mathbf{E}^{(1)}(x, z | \omega) = \frac{c}{\omega} \sum_l \varepsilon_\xi \{ X_{+l}(x) k_{zl}^{(1)} [A_l^{(1)} e^{-ik_{zl}^{(1)} z} - B_l^{(1)} e^{ik_{zl}^{(1)} z}] \mathbf{e}_x + X_{-l}(x) \beta_{\xi l} [A_l^{(1)} e^{-ik_{zl}^{(1)} z} + B_l^{(1)} e^{ik_{zl}^{(1)} z}] \mathbf{e}_z \}, \quad (16)$$

where

$$X_{\pm}(x) = \begin{cases} D_{al} e^{i\beta_{al}(x-md)} \pm F_{al} e^{-i\beta_{al}(x-md)}, & md < x < md+a \\ D_{bl} e^{i\beta_{bl}(x-md-a)} \pm F_{bl} e^{-i\beta_{bl}(x-md-a)}, & md+a < x < (m+1)d \end{cases} \quad (17)$$

and $\beta_{\xi l} = [\varepsilon_{\xi}(\omega)\omega^2/c^2 - (k_{z l}^{(1)})^2]^{1/2}$.

The electromagnetic fields, given by Eqs. (12)–(14) and (16) have to fulfill the electromagnetic boundary conditions $[[E_x]] = 0$ and $[[H_y]] = 0$ at the interfaces between the different layers of the multilayer system. The most profitable method to represent the results is the transfer-matrix method. The transfer matrix relates the field amplitudes in one layer to that of a different layer. Without the grating coupler the electromagnetic field is characterized by two amplitudes in each layer and hence, the transfer matrix is a 2×2 matrix. But in the presence of the grating coupler the electromagnetic field is represented in each layer by the Rayleigh expansion with an infinite number of field amplitudes. It is therefore necessary for practical calculations to restrict on a finite number of scattered modes: $-n_{\max} \leq n \leq n_{\max}$. We arrange the field amplitudes of the different diffraction orders of the transmitted and reflected waves in layer ν in the form of column matrices $[(2n_{\max} + 1)$ -dimensional vectors]:

$$\mathcal{A}_{pn_{\max}}^{(\nu)} = (A_{p0}^{(\nu)}, A_{p-1}^{(\nu)}, A_{p1}^{(\nu)}, \dots, A_{p-n_{\max}}^{(\nu)}, A_{pn_{\max}}^{(\nu)}), \quad (18)$$

$$\mathcal{T}_{pn_{\max}}^{ij} = \begin{pmatrix} [\mathcal{T}_{pn_{\max}}^{ij}]_{11} & \dots & [\mathcal{T}_{pn_{\max}}^{ij}]_{1(2n_{\max}+1)} \\ \vdots & & \vdots \\ [\mathcal{T}_{pn_{\max}}^{ij}]_{(2n_{\max}+1)1} & \dots & [\mathcal{T}_{pn_{\max}}^{ij}]_{(2n_{\max}+1)(2n_{\max}+1)} \end{pmatrix}, \quad (22)$$

where $i, j = 1, 2$. These submatrices are $(2n_{\max} + 1) \times (2n_{\max} + 1)$ -dimensional matrices, where we have arranged the elements $(n, n' = 0, -1, 1, \dots, n_{\max})$ in the following manner: $[\mathcal{T}_{pn_{\max}}^{ij}]_{11} = [\mathcal{T}_{pn_{\max}}^{ij}]_{n=0n'=0}$, $[\mathcal{T}_{pn_{\max}}^{ij}]_{12} = [\mathcal{T}_{pn_{\max}}^{ij}]_{n=0n'=-1}$, $[\mathcal{T}_{pn_{\max}}^{ij}]_{21} = [\mathcal{T}_{pn_{\max}}^{ij}]_{n=-1n'=0}$, \dots $[\mathcal{T}_{pn_{\max}}^{ij}]_{(2n_{\max}+1)(2n_{\max}+1)} = [\mathcal{T}_{pn_{\max}}^{ij}]_{n=n_{\max}n'=n_{\max}}$. The transfer matrix can be written as

$$\mathbf{T}_{pn_{\max}}(\nu, \nu+1) = [\mathbf{P}_{pn_{\max}}(\nu)]^{-1} [\mathbf{D}_{pn_{\max}}(\nu)]^{-1} \mathbf{D}_{pn_{\max}}(\nu+1),$$

where $\mathbf{P}_{pn_{\max}}(\nu)$ is the propagation matrix, which describes the propagation of the diffracted waves in layer ν , and $\mathbf{D}_{pn_{\max}}(\nu)$ is the dynamical matrix, which depends on the polarization of the waves. The different matrices are derived explicitly in Ref. 36, where the reader can find any details. As shown in Ref. 36 the propagation matrix of a homogeneous layer is given by

$$\mathbf{P}_{pn_{\max}}(\nu) = \begin{pmatrix} \mathcal{P}_{pn_{\max}}^1(\nu) & 0 \\ 0 & \mathcal{P}_{pn_{\max}}^2(\nu) \end{pmatrix}, \quad (23)$$

where the elements of the submatrices read

$$[\mathcal{P}_{pn_{\max}}^1(\nu)]_{nn'} = \exp(ik_{zn}^{(\nu)} d_{\nu}) \delta_{nn'}, \quad (24)$$

and

$$\mathcal{P}_{pn_{\max}}^2(\nu) = [\mathcal{P}_{pn_{\max}}^1(\nu)]^{-1}. \quad (25)$$

$$\mathcal{B}_{pn_{\max}}^{(\nu)} = (B_{p0}^{(\nu)}, B_{p-1}^{(\nu)}, B_{p1}^{(\nu)}, \dots, B_{p-n_{\max}}^{(\nu)}, B_{pn_{\max}}^{(\nu)}). \quad (19)$$

Note that the incident wave contains only the zeroth-order $\mathcal{A}_{pn_{\max}}^{(0)} = (A_{p0}^{(0)}, 0, \dots, 0)$. The resulting matrix equation, which relates the field amplitudes in layer ν with that in layer $\nu+1$, is given by

$$\begin{pmatrix} \mathcal{A}_{pn_{\max}}^{(\nu)} \\ \mathcal{B}_{pn_{\max}}^{(\nu)} \end{pmatrix} = \begin{pmatrix} \mathcal{T}_{pn_{\max}}^{11} & \mathcal{T}_{pn_{\max}}^{12} \\ \mathcal{T}_{pn_{\max}}^{21} & \mathcal{T}_{pn_{\max}}^{22} \end{pmatrix} \begin{pmatrix} \mathcal{A}_{pn_{\max}}^{(\nu+1)} \\ \mathcal{B}_{pn_{\max}}^{(\nu+1)} \end{pmatrix}, \quad (20)$$

where the $2(2n_{\max} + 1) \times 2(2n_{\max} + 1)$ -dimensional transfer matrix is

$$\mathbf{T}_{pn_{\max}}(\nu, \nu+1) = \begin{pmatrix} \mathcal{T}_{pn_{\max}}^{11} & \mathcal{T}_{pn_{\max}}^{12} \\ \mathcal{T}_{pn_{\max}}^{21} & \mathcal{T}_{pn_{\max}}^{22} \end{pmatrix}. \quad (21)$$

The submatrices are given by

Herein, we have arranged the elements of the submatrices in the following form: $\mathcal{P}_{11}^1 = \mathcal{P}_{n=0, n'=0}^1$, $\mathcal{P}_{12}^1 = \mathcal{P}_{n=0, n'=-1}^1$, $\mathcal{P}_{21}^1 = \mathcal{P}_{n=-1, n'=0}^1 \dots$, $\mathcal{P}_{1, 2n_{\max}+1}^1 = \mathcal{P}_{n=0, n'=n_{\max}}^1, \dots$, $\mathcal{P}_{2n_{\max}+1, 2n_{\max}+1}^1 = \mathcal{P}_{n=n_{\max}, n'=n_{\max}}^1$. The dynamical matrix of a homogeneous layer is calculated to be

$$\mathbf{D}_{pn_{\max}}(\nu) = \begin{pmatrix} \mathcal{D}_{pn_{\max}}^{11}(\nu) & \mathcal{D}_{pn_{\max}}^{12}(\nu) \\ \mathcal{D}_{pn_{\max}}^{21}(\nu) & \mathcal{D}_{pn_{\max}}^{22}(\nu) \end{pmatrix}, \quad (26)$$

where the elements of the submatrices are given by

$$[\mathcal{D}_{pn_{\max}}^{11}(\nu)]_{nn'} = [\mathcal{D}_{pn_{\max}}^{12}(\nu)]_{nn'} = \delta_{nn'} \quad (27)$$

and

$$[\mathcal{D}_{pn_{\max}}^{21}(\nu)]_{nn'} = -[\mathcal{D}_{pn_{\max}}^{22}(\nu)]_{nn'} = \frac{k_{zn}^{(\nu)}}{\varepsilon_{xx}^{(\nu)}} \delta_{nn'}. \quad (28)$$

Equations (23) – (28) are valid for any optically uniaxial medium. The corresponding expressions for isotropic media follow if one replaces $\varepsilon_{xx}^{(\nu)}(\omega) = \varepsilon_{zz}^{(\nu)}(\omega) = \varepsilon_{\nu}(\omega)$.

Defining the transfer matrix of the whole sample by

$$\begin{aligned} \mathbf{T}_{pn_{\max}}^G(0, N+1) \\ = \mathbf{T}_{pn_{\max}}^G(0, 1) \mathbf{T}_{pn_{\max}}^G(1, 2) \mathbf{T}_{pn_{\max}}^G(2, 3) \dots \mathbf{T}_{pn_{\max}}^G(N, N+1), \end{aligned} \quad (29)$$

then field amplitudes $(\mathcal{A}_{pn_{\max}}^{(0)}, \mathcal{B}_{pn_{\max}}^{(0)})$ in the half-space $z > z_0$ are related to that in the substrate $z < z_N$ by

$$\begin{pmatrix} \mathcal{A}_{pn_{\max}}^{(0)} \\ \mathcal{B}_{pn_{\max}}^{(0)} \end{pmatrix} = \mathbf{T}_{pn_{\max}}^G(0, N+1) \begin{pmatrix} \mathcal{A}_{pn_{\max}}^{(N+1)} \\ \mathcal{B}_{pn_{\max}}^{(N+1)} \end{pmatrix}. \quad (30)$$

Herein, the transfer matrices $\mathbf{T}_{pn_{\max}}^G(0,1)$ and $\mathbf{T}_{pn_{\max}}^G(1,2)$ result from the two interfaces of the grating with the homogeneous media. Further, the following is valid: The submatrices that result from the interface between two homogeneous media are diagonal, whereas those resulting from the interfaces between the grating and homogeneous media are non-diagonal. Also the last one can be represented in the forms $\mathbf{T}_{pn_{\max}}^G(0,1) = [\mathbf{P}_{pn_{\max}}^G(0)]^{-1} [\mathbf{D}_{pn_{\max}}^G(0)]^{-1} \mathbf{D}_{pn_{\max}}^G(1)$ and $\mathbf{T}_{pn_{\max}}^G(1,2) = [\mathbf{P}_{pn_{\max}}^G(1)]^{-1} [\mathbf{D}_{pn_{\max}}^G(1)]^{-1} \mathbf{D}_{pn_{\max}}^G(2)$. The matrices $[\mathbf{P}_{pn_{\max}}^G(0)]^{-1}$ and $[\mathbf{D}_{pn_{\max}}^G(0)]^{-1}$ are the inverse of the propagation matrix and of the dynamical matrix of the half-space $z > z_0$ ($\nu=0$) given by Eqs. (23) and (26), respectively, and $\mathbf{D}_{pn_{\max}}^G(2)$ is the dynamical matrix of layer $\nu=2$ given by Eq. (26). In the grating region the propagation matrix $\mathbf{P}_{pn_{\max}}^G(0)$ has the same form as given in Eqs. (23) – (25). But the dynamical matrix in the grating region is different from that describing homogeneous layers. The dynamical matrix of the grating is given by

$$\mathbf{D}_{pn_{\max}}^G(\nu) = \begin{pmatrix} \mathcal{D}_{pn_{\max}}^{G,11}(\nu) & \mathcal{D}_{pn_{\max}}^{G,12}(\nu) \\ \mathcal{D}_{pn_{\max}}^{G,21}(\nu) & \mathcal{D}_{pn_{\max}}^{G,22}(\nu) \end{pmatrix}, \quad (31)$$

with

$$[\mathcal{D}_{pn_{\max}}^{G,11}(\nu)]_{ln} = [\mathcal{D}_{pn_{\max}}^{G,12}(\nu)]_{ln} = \Gamma_{ln}^1 \quad (32)$$

and

$$[\mathcal{D}_{pn_{\max}}^{G,21}(\nu)]_{ln} = -[\mathcal{D}_{pn_{\max}}^{G,22}(\nu)]_{ln} = \Gamma_{ln}^2 k_z^{(\nu)}. \quad (33)$$

The different elements of these matrices are

$$\begin{aligned} \Gamma_{ln}^1 &= \frac{1}{d} \int_0^d dx \exp(-ik_{xn}x) X_{+l}(x) \\ &= iD_{al} \frac{\{1 - \exp[ia(\beta_{al} - k_{xn})]\}}{\beta_{al} - k_{xn}} \end{aligned}$$

$$\begin{aligned} &-iF_{al} \frac{\{1 - \exp[-ia(\beta_{al} + k_{xn})]\}}{\beta_{al} + k_{xn}} \\ &+ iD_{bl} \frac{\{1 - \exp[ib(\beta_{bl} - k_{xn})]\}}{\exp(ik_{xn}a)(\beta_{bl} - k_{xn})} \\ &-iF_{bl} \frac{\{1 - \exp[-ib(\beta_{bl} + k_{xn})]\}}{\exp(ik_{xn}a)(\beta_{bl} + k_{xn})}, \quad (34) \end{aligned}$$

$$\begin{aligned} \Gamma_{ln}^2 &= \frac{1}{d} \int_0^d dx \exp(-ik_{xn}x) \frac{X_{+l}(x)}{\varepsilon_\xi} \\ &= iD_{al} \frac{\{1 - \exp[ia(\beta_{al} - k_{xn})]\}}{\varepsilon_a(\beta_{al} - k_{xn})} \\ &-iF_{al} \frac{\{1 - \exp[-ia(\beta_{al} + k_{xn})]\}}{\varepsilon_a(\beta_{al} + k_{xn})} \\ &+ iD_{bl} \frac{\{1 - \exp[ib(\beta_{bl} - k_{xn})]\}}{\exp(ik_{xn}a)\varepsilon_b(\beta_{bl} - k_{xn})} \\ &-iF_{bl} \frac{\{1 - \exp[-ib(\beta_{bl} + k_{xn})]\}}{\exp(ik_{xn}a)\varepsilon_b(\beta_{bl} + k_{xn})}. \quad (35) \end{aligned}$$

Herein, the matrix elements Γ_{ln}^1 ($n=0, -1, 1, \dots, n_{\max}$, $l=1, \dots, 2n_{\max}+1$) are arranged in the form: $\Gamma_{11}^1 = \Gamma_{l=1, n=0}^1, \Gamma_{12}^1 = \Gamma_{l=1, n=-1}^1, \Gamma_{21}^1 = \Gamma_{l=2, n=0}^1, \dots, \Gamma_{12n_{\max}+1}^1 = \Gamma_{l=1, n=n_{\max}}^1, \dots, \Gamma_{2n_{\max}+12n_{\max}+1}^1 = \Gamma_{l=2n_{\max}+1, n=n_{\max}}^1$.

In the absence of the grating the formalism presented here reduces to the well-known 2×2 transfer-matrix method; i.e., each submatrix of the transfer matrix given in Eq. (20) becomes a single complex function or number and the field amplitudes, Eqs. (18) and (19), reduce to only one (that of the zeroth order), respectively:

$$\begin{pmatrix} A_{p0}^{(\nu)} \\ B_{p0}^{(\nu)} \end{pmatrix} = \mathbf{T}_{p0}(\nu, \nu+1) \begin{pmatrix} A_{p0}^{(\nu+1)} \\ B_{p0}^{(\nu+1)} \end{pmatrix}. \quad (36)$$

with

$$\mathbf{T}_{p0}(\nu, \nu+1) = \frac{1}{2} \begin{pmatrix} [1 + \varepsilon_{xx}^{(\nu)} k_z^{(\nu+1)} / (\varepsilon_{xx}^{(\nu+1)} k_z^{(\nu)})] \exp(-ik_z^{(\nu)} d_\nu) & [1 - \varepsilon_{xx}^{(\nu)} k_z^{(\nu+1)} / (\varepsilon_{xx}^{(\nu+1)} k_z^{(\nu)})] \exp(-ik_z^{(\nu)} d_\nu) \\ [1 - \varepsilon_{xx}^{(\nu)} k_z^{(\nu+1)} / (\varepsilon_{xx}^{(\nu+1)} k_z^{(\nu)})] \exp(ik_z^{(\nu)} d_\nu) & [1 + \varepsilon_{xx}^{(\nu)} k_z^{(\nu+1)} / (\varepsilon_{xx}^{(\nu+1)} k_z^{(\nu)})] \exp(ik_z^{(\nu)} d_\nu) \end{pmatrix}. \quad (37)$$

For a multilayer system with a grating coupler, however, transmitted waves of the order $n=0, \pm 1, \pm 2, \dots, \pm n_{\max}$ occur. The quantity measured in the experiments is the time-averaged power transmission coefficient, which is calculated to be

$$\begin{aligned} T_p &= \frac{1}{2\varepsilon_{sN+1} |A_{p0}^{(0)}|^2 k_z^{(0)}} \sum_{n, n' = -n_{\max}}^{n_{\max}} A_{pn}^{(N+1)*} A_{pn'}^{(N+1)} e^{i\phi_{nn'}(x, z)} \\ &\times (k_{zn}^{(N+1)*} + k_{zn'}^{(N+1)}), \quad (38) \end{aligned}$$

where the asterisk symbol * means complex conjugate and

$$\phi_{nn'}(x, z) = x(k_{xn'} - k_{xn}) + (z - z_N)(k_{zn}^{(N+1)*} - k_{zn'}^{(N+1)}). \quad (39)$$

The transmission coefficient depends on both x and z because the grating produces a diffraction pattern along the x direction. It is important to note that not all the orders of diffracted waves can propagate through the sample and from

the sample surface to the detector. Only those waves for which $k_{zn}^{(0)}$ is real and positive, i.e., if $\omega^2/c^2 > [(\omega/c)\sin\Theta_0 + (2\pi/d)n]^2$ are propagating waves above and below the sample. In the case that $k_{zn}^{(0)}$ is pure imaginary, i.e., if $\omega^2/c^2 < [(\omega/c)\sin\Theta_0 + (2\pi/d)n]^2$ the corresponding wave is an evanescent wave with the decay length $L_n = (2|\text{Im}k_{zn}^{(0)}|)$ and thus exists only in the near field of the grating. Because in the experiment the distance between the sample and the detector is usually much larger than the decay lengths of the evanescent waves, which cannot transport energy in the negative z direction if the medium below the sample is nonabsorbing, only the propagating waves are detected. But it is important that the evanescent waves couple with the collective excitations of the Q2DEG inside the PQW, which usually is in the near vicinity of the grating, leading to a resonant decrease of the reflected and transmitted radiation. Hence, the measured transmission coefficient contains, via the boundary conditions at the interfaces of the grating with the homogeneous layers, which couple the different diffracted orders, all information about the evanescent higher-order diffracted waves. In the typical situation only the zero-order diffracted wave is propagating in the surrounding. If in this case the influence of the evanescent waves below the sample on the transmission coefficient of Eq. (39) can be neglected then, the transmission coefficient reads

$$T_p = \frac{|A_{p0}^{(N+1)}|^2 k_{z0}^{(N+1)}}{\varepsilon_{sN+1} |A_{p0}^{(0)}|^2 k_{z0}^{(0)}}. \quad (40)$$

With increasing period d of the grating it becomes possible that the first-order diffracted waves become propagating. In the vacuum above and below the sample this takes place at $d = \lambda_0$, where $\lambda_0 = 2\pi/k^{(0)} = 2\pi c/\omega$ is the vacuum wavelength at the frequency of the ISR. If this becomes true the transmission spectrum as well as the reflection spectrum show the so-called Rayleigh anomaly.^{37,40-43} It consists of a rapid variation in the amplitudes of the diffracted orders corresponding to the onset (evanescent \rightarrow propagating) or disappearance (propagating \rightarrow evanescent) of a particular diffracted order. This is true because with the appearance of a new order of diffracted wave a rearrangement of the field amplitudes of the other propagating diffracted waves is caused. The wavelength where this takes place is called Rayleigh wavelength λ_R . But in general, two types of anomalous effects, called Wood's anomalies,⁴⁴ occur: (i) the above described Rayleigh wavelength type and (ii) the resonance-type anomaly. The second type of Wood's anomaly is connected with the excitation of a leaky surface wave propagating along the metallic grating.^{41,43,45}

If $2n_1 + 1$ waves are propagating below the sample (zero order plus forward and backward diffracted waves of higher order), valid if $d \geq n_1 \lambda_0$, it depends on the detector if the diffraction pattern or the spatial average is measured. Usually the last one is realized and then the transmission coefficient, which only includes the action of the propagating waves below the sample, reads

$$T_p = \frac{1}{\varepsilon_{sN+1} |A_{p0}^{(0)}|^2 k_{z0}^{(0)}} \sum_{n=-n_1}^{n=n_1} |A_{pn}^{(N+1)}|^2 k_{zn}^{(N+1)}. \quad (41)$$

The experimental and theoretical investigations performed in this work are focused on the question about the coupling efficiency of the grating to excite the ISR of the Q2DEG synthesized in the PQW with one occupied subband for a given semiconductor heterostructure. It is known (for details see Ref. 36) that the coupling efficiency depends on the one hand on the grating parameters (period, height, mark-to-space ratio, material) but on the other hand on the properties of the multilayer system (distance between grating and Q2DEG and between the Q2DEG and the lower boundary of the sample, material, etc.). Here, we investigate in detail the coupling efficiency of FIR transmission in dependence on the grating period d and all the other parameters of the sample remain fixed.

IV. NUMERICAL RESULTS AND DISCUSSION

The numerical calculations are done for the six-layer system depicted in Fig. 3 with the following material parameters: layer $\nu=1$ is a Ag grating layer with $\omega_{pa} = 5.69 \times 10^{15} \text{ s}^{-1}$ and $\gamma_a = 7.596 \times 10^{13} \text{ s}^{-1}$ of height $d_1 = h = 50 \text{ nm}$; layer $\nu=2$ is a GaAs layer of thickness $d_2 = 10 \text{ nm}$ and $\varepsilon_{s2} = 12.87$ and layer $\nu=3$ is a $\text{Ga}_{1-x}\text{Al}_x\text{As}$ ($x=0.25$) layer with $\varepsilon_{s3} = 12.21$ and of thickness $d_3 = 241 \text{ nm}$; the fourth layer contains the PQW with the bare confining energy $\hbar\Omega = 11 \text{ meV}$, in which the Q2DEG is synthesized. For simplicity we assume for the background dielectric constant of this layer a homogeneous dielectric constant with the parameters of GaAs: $\varepsilon_{s4} = 12.87$ and $n_{\text{2DEG}} = 2 \times 10^{11} \text{ cm}^{-2}$, $\tau_{\parallel} = 1 \times 10^{-11} \text{ s}$, $\tau_{\perp} = 1 \times 10^{-12} \text{ s}$, and $d_4 = a_{\text{2DEG}} = 18.7 \text{ nm}$. Layer $\nu=5$ is 224 nm thick and consists of $\text{Ga}_{1-x}\text{Al}_x\text{As}$ and the layer $\nu=6$ is a 500 nm-thick GaAs layer. In the numerical calculations the substrate (region $\nu=7$) is filled with GaAs and it is assumed to be of infinite thickness, which is different from the experimental situation. In the experimental sample the substrate is wedged with a mean thickness of approximately 500 μm . In both theory and experiment this is done to avoid the Fabry-Pérot resonances in the substrate.

The numerically calculated relative transmission coefficient $-\Delta T/T$ is plotted in Fig. 4 for different periods d of the grating and fixed mark-to-space ratio $t=1$. The wave vector transfer from the incident wave to the ISR is as follows, for $d=4 \mu\text{m}$: $k_{x1} = 1.57 \times 10^4 \text{ cm}^{-1}$, $k_{x2} = 3.14 \times 10^4 \text{ cm}^{-1}$, etc.; $d=6 \mu\text{m}$: $k_{x1} = 1.047 \times 10^4 \text{ cm}^{-1}$, $k_{x2} = 2.094 \times 10^4 \text{ cm}^{-1}$, etc.; $d=10 \mu\text{m}$: $k_{x1} = 6.238 \times 10^3 \text{ cm}^{-1}$, $k_{x2} = 1.256 \times 10^4 \text{ cm}^{-1}$, etc.; $d=20 \mu\text{m}$: $k_{x1} = 3.14 \times 10^3 \text{ cm}^{-1}$, $k_{x2} = 6.28 \times 10^3 \text{ cm}^{-1}$, etc.; $d=30 \mu\text{m}$: $k_{x1} = 2.094 \times 10^3 \text{ cm}^{-1}$, $k_{x2} = 4.188 \times 10^3 \text{ cm}^{-1}$, etc.; and for $d=40 \mu\text{m}$: $k_{x1} = 1.57 \times 10^3 \text{ cm}^{-1}$, $k_{x2} = 3.14 \times 10^3 \text{ cm}^{-1}$, etc. It becomes obvious from both experiment (Fig. 2) and theory (Fig. 4) that the efficiency to excite the ISR of a Q2DEG inside a given multilayered-quantum-well system increases with increasing period d of the grating coupler. This is true for the given parameters up to $d=20 \mu\text{m}$ with excellent agreement between theory and experiment. It is seen that analogous to the experimental result, the theoretically calculated line shapes of the ISR peak show at smaller periods a Lorentzian form, which becomes more and more asymmetric

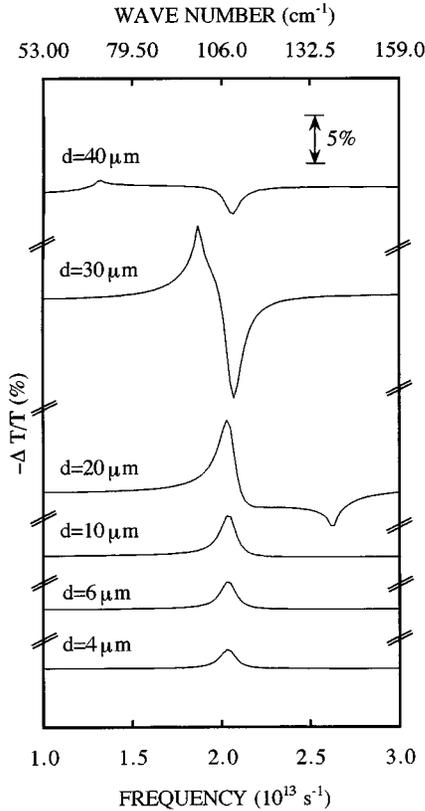


FIG. 4. Relative transmission coefficient $-\Delta T/T$ of the multilayer system in the near vicinity of the ISR of the Q2DEG for different periods d of the grating assuming the region $\nu=7$ to be filled by GaAs.

with increasing grating period. For $d=20 \mu\text{m}$ we have the largest magnitude of the maximum in $-\Delta T/T$, but with an asymmetric shape. Increasing further the period transforms the maximum to a minimum that vanishes for $d \rightarrow \infty$. Please note that in the case of the asymmetric line shape $T(n_{2\text{DEG}}) > T(0)$ is valid; i.e., more light is transmitted through the sample with the Q2DEG.

Because it is difficult to define something like the coupling efficiency for a strongly disturbed line shape, we choose the integrated relative transmission of the detected lines as a measure for it. The result is shown in Fig. 5, where we plot the integrated relative transmission of the resonant structures for the measured and calculated spectra as a function of the grating period. For $d < \lambda_{\text{sub}}$ we observe a steady increase of the integrated signal with increasing period d . Beyond the vertical dashed line, which indicates the condition $d = \lambda_{\text{sub}}$, the integrated relative transmission decreases very rapidly. It should be noted, however, that our experimental setup only integrates over a finite solid angle given by the distance between the sample and the waveguide and by the waveguides aperture. This fact might be important especially for propagating higher-order waves as we point out below.

In the case considered here for $d < 17.5 \mu\text{m}$ only the zeroth-order diffracted wave is a propagating wave in vacuum and in the GaAs substrate. But for a grating period of $d = 17.5 \mu\text{m}$ at $\omega = 3 \times 10^{13} \text{s}^{-1}$ ($\bar{\nu} \equiv \omega/(2\pi c) = 159.3 \text{cm}^{-1}$) the first-order diffracted wave additionally becomes propagating in the GaAs substrate and for

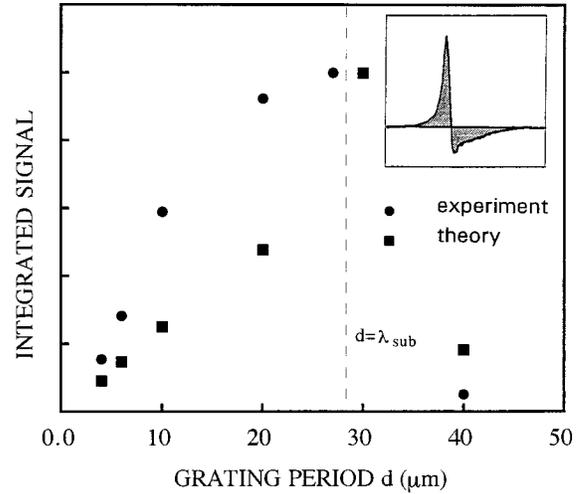


FIG. 5. Integrated relative transmission of the spectra shown in Figs. 2 and 4. As long as the grating period d is smaller than the wavelength λ_{sub} of the light corresponding to the intersubband transition, the integrated relative transmission increases. For $d > \lambda_{\text{sub}}$ the integrated relative transmission rapidly decreases.

$d = 62.83 \mu\text{m}$ this becomes also true in the region filled by vacuum. In the framework of the ϵ_s approximation the onset of the propagation of the first-order diffracted wave in the GaAs substrate, which is effectively formed by the regions $\nu=6$ and 7 , is just at $d = \lambda_{\text{sub}}$, where $\lambda_{\text{sub}} = 2\pi c/(\sqrt{\epsilon_s} \omega)$ is the wavelength of the FIR light in the substrate at the frequency of the ISR. In this case it follows from $k_{z1}^{(6)} = 0$ that $E_{x1}^{(6)} = 0$ and thus the electric field of this wave is polarized pure perpendicular to the interfaces in the substrate and in all the other layers filled with GaAs. The physical situation inside the PQW is very similar but becomes complicated due to the resonance structure of the complex dielectric tensor, which may vary strongly in the near vicinity of the frequency of the ISR. The onset of the propagation of the first-order diffracted wave in the substrate is seen in the calculated relative transmission coefficient. The calculated curves show for $d = 10 \mu\text{m}$, $d = 20 \mu\text{m}$ and $d = 30 \mu\text{m}$ the Rayleigh anomaly at $\omega = 5.3 \times 10^{13} \text{s}^{-1}$ ($\bar{\nu} = 281.5 \text{cm}^{-1}$), $\omega = 2.6 \times 10^{13} \text{s}^{-1}$ ($\bar{\nu} = 138.1 \text{cm}^{-1}$) and $\omega = 1.7 \times 10^{13} \text{s}^{-1}$ ($\bar{\nu} = 90.3 \text{cm}^{-1}$), respectively. It becomes obvious from the theoretical and experimental $-\Delta T/T$ curves that just under these conditions the transmission peaks begin to deform. For conditions below the onset of the Rayleigh anomaly one measures only the zeroth-order beam. But the peak in the relative transmission spectrum results from the evanescent higher-order beams which couple with the nonradiative intersubband plasmons. This situation is changed above the onset of the Rayleigh anomaly, where the first-order diffracted beam couples with the radiative ISR. Because the collective intersubband excitations are nearly dispersionless in the long-wavelength limit, the resulting ISR peak is a sum of the action of all diffracted beams with $k_{x1}, k_{x2}, k_{x3}, \dots$. Our detailed numerical analysis shows that for the infinitely thick GaAs substrate only the Rayleigh-wavelength-type anomaly is responsible for the deformation of the peak. The asymmetric line shape is thus caused by the superimposition of the ISR peak with the Rayleigh anomaly. For the chosen system the resonance-type anomaly is absent in the plotted

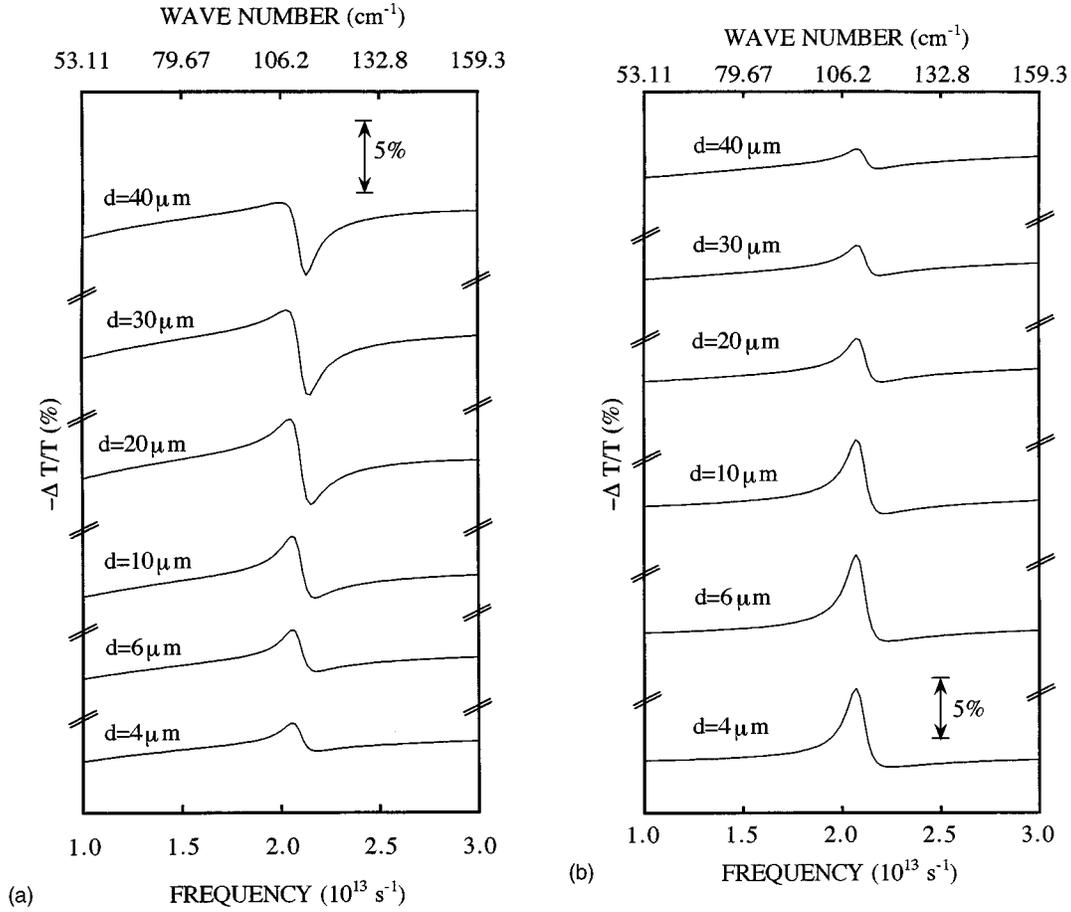


FIG. 6. Relative transmission coefficient $-\Delta T/T$ of the multilayer system in the near vicinity of the ISR of the Q2DEG for different periods d of the grating of height $h = 10$ nm, assuming the region $\nu = 7$ to be filled by vacuum: (a) $d_6 = 4$ μm and (b) $d_6 = 500$ nm.

frequency range; i.e., the comparison of the transmission $T(0)$ and $T(n_{2\text{DEG}})$ gives no indication for the excitation of a leaky surface wave carried by the grating. It is noticeable that the diffracted peaks resulting from the excitation of the intrasubband plasmon $\omega_p^{00}(q_{\parallel})$ at $q_{\parallel} = k_{x1}, k_{x2}, k_{x3}, \dots$ are always symmetrically shaped.

If one would calculate $-\Delta T/T$ for a sample with GaAs substrate of *finite* thickness, at the above-mentioned frequencies the first-order diffracted wave becomes propagating in the GaAs layers of the sample and slightly below these frequencies in the Ga_{1-x}Al_xAs layers. Possible anomalies in this case occurring in the optical spectra due the Rayleigh wavelength effect we call internal Rayleigh anomalies. However, in this frequency range the first-order diffracted wave is an evanescent wave in the vacuum below and above the sample. To answer the question whether these internal Rayleigh anomalies are the cause of the asymmetry in the experimentally detected line shape or not, we investigated the multilayer system of Fig. 3, but assumed the region $\nu = 7$ to be filled by vacuum. The relative transmission coefficient $-\Delta T/T$ is plotted in Fig. 6 for two different thicknesses d_6 , assuming here a height of the grating of $h = 10$ nm. It is seen in Fig. 6 (a) that for $d_6 = 4$ μm the Lorentzian shaped maximum, which appears for $d = 4$ nm, deforms with increasing grating period; quite similar to that shown in Fig. 4, it becomes asymmetrically shaped and for $d = 40$ μm the maximum is transformed into a minimum. The spectra in

Fig. 6(a), calculated for $d = 20$ μm and $d = 30$ μm, show that the internal Rayleigh anomalies are absent and thus they have no significant influence on $-\Delta T/T$. Here, the question arises about the mechanism that is responsible for the pronounced asymmetric profiles. In Fig. 6(b) we plotted the corresponding relative transmission spectra for $d_6 = 500$ nm. It becomes obvious that in this case the best coupling efficiency is obtained for $d = 6$ μm. For larger grating periods the peak height decreases but with a nearly stable line shape. It should be noted that whereas the exponential function of the first-order diffracted wave, appearing in the propagation matrix of layer $\nu = 6$, Eq. (24), is nearly one for small d_6 in the plotted frequency range, it varies rapidly with the frequency for larger d_6 . Such a rapid variation may cause Fabry-Pérot-like resonances of the first-order diffracted wave in the multilayer system. Further, we have calculated the total power absorption $A_p(n_{2\text{DEG}}) = 1 - T_p(n_{2\text{DEG}}) - R_p(n_{2\text{DEG}})$ for the two configurations of Fig. 6 assuming a grating period of $d = 30$ μm. From the comparison of $A_p(n_{2\text{DEG}})$ and $A_p(0)$ (see Fig. 7) it becomes obvious that the $A_p(0)$ spectrum shows for $d_6 = 4$ μm a maximum, which we attribute to the excitation of a leaky surface wave in the Ag grating. This resonance superimposes the ISR resonance to give the asymmetric line shape of the peak. For the smaller layer thickness, $d_6 = 500$ nm, the resonance associated with the surface wave is absent in the plotted frequency range. Thus, we attribute the asymmetric line shape for the

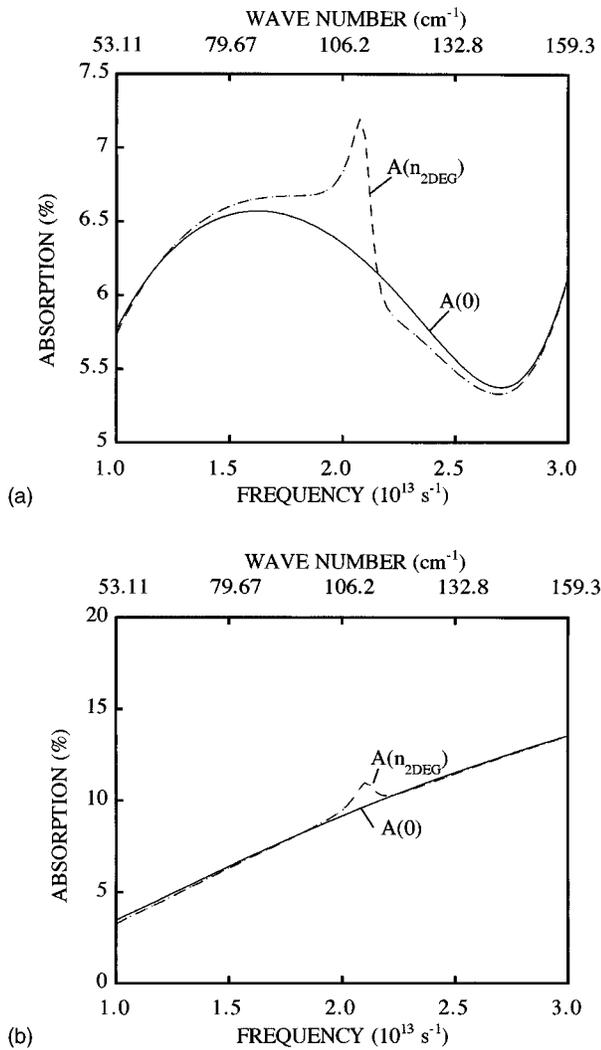


FIG. 7. Total absorption coefficient $A(n_{2\text{DEG}})$ and $A(0)$ for the multilayer system of Fig. 6: (a) $d_6 = 4 \mu\text{m}$ and (b) $d_6 = 500 \text{ nm}$ calculated for a grating period of $d = 30 \mu\text{m}$.

finite width substrate due to the resonance-type Wood's anomaly. It should be noted that similar line shapes of the ISR resonance were observed in Ref. 20 for metal-oxide-semiconductor structures of silicon. In this paper the asym-

metric line shape is assumed to be caused by the superimposition of the ISR with a non-depolarization-shifted ISR. Our rigorous grating-coupler theory, however, shows that the ISR is always accompanied by a depolarization shift that cannot be "screened" by the grating. Thus, it seems that also in this early experiment Wood's anomalies could be responsible for the asymmetric line shape.

V. SUMMARY

In summary, we have measured and calculated the optical response of the ISR of a quasi-two-dimensional electron system in a PQW with different grating couplers on top of this structure. Using the transfer-matrix method of local optics and the modal-expansions method to calculate the influence of the lamellar grating on the electromagnetic fields, the calculated relative transmission describes very well the measured spectrum. It is shown that the coupling efficiency of the grating coupler to excite the ISR increases with increasing period of the grating up to a certain value where the absorption peak starts to deform from a Lorentzian shape to an asymmetric shape. This asymmetric line shape is caused by both types of Wood's anomalies, i.e., due to the propagating higher-order diffracted waves in the sample beyond the threshold of the Rayleigh anomaly and due to the resonance-type anomaly (excitation of a leaky surface wave in the grating region). In this case one excites both radiative and nonradiative collective intersubband excitations.

ACKNOWLEDGMENTS

We gratefully acknowledge many fruitful discussions with J. P. Kotthaus. This work has been financially supported in part by the Deutsche Forschungsgemeinschaft (Project No. We 1532/3-2) and the PHANTOMS network, in part by the Air Force Office of Scientific Research under Grant No. AFOSR-91-0214, by the NSF Science and Technology Center for Quantized Electronic Structures (QUEST) DMR-91-20007, by the NSF DMR-90-02291, and the ONR N000-14-92-J1452. The München-Santa Barbara cooperation has been supported by a joint NSF-European grant (EC-US 015:9826).

¹For an excellent review on the electronic properties of quasi-two-dimensional electron systems, see, e.g., T. Ando, A. B. Fowler, and F. Stern, *Rev. Mod. Phys.* **54**, 437 (1982).
²K.W. Goosen and S.A. Lyon, *Appl. Phys. Lett.* **47**, 1257 (1985).
³M. Helm, E. Colas, P. England, F. DeRosa, and S.J. Allen, Jr., *Appl. Phys. Lett.* **53**, 1714 (1988).
⁴B. F. Levine, C.G. Bethea, G. Hasnain, U. Walker, and R.J. Malik, *Appl. Phys. Lett.* **53**, 296 (1988).
⁵D. Heitmann in *Festkörperprobleme: Advances in Solid State Physics* (Vieweg, Braunschweig, 1985), Vol. 25, p. 429.
⁶L. Wendler and E. Kändler, *Phys. Status Solidi B* **177**, 9 (1993).
⁷L. Wendler and T. Kraft (unpublished).
⁸F. Stern, *Phys. Rev. Lett.* **33**, 960 (1974).
⁹S.J. Allen, Jr., D.C. Tsui, and R.A. Logan, *Phys. Rev. Lett.* **38**, 980 (1977).

¹⁰T.N. Theis, J.P. Kotthaus, and P.J. Stiles, *Solid State Commun.* **24**, 273 (1977).
¹¹E. Batke, D. Heitmann, and C.W. Tu, *Phys. Rev. B* **34**, 6951 (1982).
¹²B.D. McCombe, R.T. Holm, and D.E. Schafer, *Solid State Commun.* **32**, 603 (1979).
¹³M. Nakayama, H. Duwahara, H. Kato, and K. Kubota, *Appl. Phys. Lett.* **51**, 1741 (1987).
¹⁴J. Y. Andersson and G. Landren, *J. Appl. Phys.* **64**, 4123 (1988).
¹⁵Lian Zheng, W.L. Schaich, and A.H. MacDonald, *Phys. Rev. B* **41**, 8493 (1990).
¹⁶Lian Zheng and W.L. Schaich, *Phys. Rev. B* **43**, 4515 (1991).
¹⁷Dongzi Liu and S. Das Sarma, *Phys. Rev. B* **44**, 9122 (1991).
¹⁸C.D. Ager and H.P. Hughes, *Phys. Rev. B* **44**, 13 452 (1991).

- ¹⁹C.D. Ager, R.J. Wilkinson, and H.P. Hughes, *J. Appl. Phys.* **71**, 1322 (1992).
- ²⁰D. Heitmann and U. Mackens, *Phys. Rev. B* **33**, 8269 (1986).
- ²¹E. Batke, G. Weimann, and W. Schlapp, *Phys. Rev. B* **39**, 11 171 (1989).
- ²²J. Pillath, E. Batke, G. Weimann, and W. Schlapp, *Phys. Rev. B* **40**, 5879 (1989).
- ²³E. Batke, in *Festkörperprobleme: Advances in Solid State Physics* (Vieweg, Braunschweig, 1991), Vol. 31, p. 297.
- ²⁴W.J. Li, B.D. McCombe, F.A. Chambers, G.P. Devane, J. Ralston, and G. Wicks, *Phys. Rev. B* **42**, 11 953 (1990).
- ²⁵W.J. Li and B.D. McCombe, *J. Appl. Phys.* **71**, 1038 (1992).
- ²⁶M. Helm, E. Colas, P. England, F. DeRosa, and S.J. Allen, Jr., *Appl. Phys. Lett.* **53**, 1714 (1988).
- ²⁷W.J. Li, B.D. McCombe, F.A. Chambers, G.P. Devane, F. Ralston, and G. Wicks, *Surface Sci.* **228**, 164 (1990).
- ²⁸M. Sundaram, A.C. Gossard, J.H. English, and R.M. Westervelt *Superlattices Microstruct.* **4**, 683 (1988).
- ²⁹A. Wixforth, M. Kaloudis, C. Roche, K. Ensslin, M. Sundaram, J.H. English, and A.C. Gossard, *Semicond. Sci. Technol.* **9**, 215 (1994).
- ³⁰L. Brey, N.F. Johnson, and B.I. Halperin, *Phys. Rev. B* **40**, 10 647 (1989).
- ³¹S. K. Yip, *Phys. Rev. B* **43**, 1707 (1991).
- ³²W. Kohn, *Phys. Rev.* **123**, 1242 (1961).
- ³³L. Wendler and R. Haupt, *Phys. Rev. B* **52**, 9031 (1995).
- ³⁴L. Wendler and T. Kraft, *Phys. Rev. B* **54**, 11 436 (1997).
- ³⁵A. Wixforth, M. Sundaram, K. Ensslin, J.H. English, and A.C. Gossard, *Appl. Phys. Lett.* **56**, 454 (1990).
- ³⁶L. Wendler and T. Kraft (unpublished).
- ³⁷J.W.S. Rayleigh, *Proc. R. Soc. London Ser. A* **79**, 399 (1907).
- ³⁸R. Petit, in *Electromagnetic Theory of Gratings*, edited by R. Petit, Springer Topics in Modern Physics Vol. 22 (Springer-Verlag, Heidelberg, 1980), p. 1.
- ³⁹P. Sheng, R. S. Stepleman, and P. N. Sanda, *Phys. Rev. B* **26**, 2907 (1982).
- ⁴⁰J.W.S. Rayleigh, *Philos. Mag.* **14**, 60 (1907).
- ⁴¹A. Hessel and A. A. Onliner, *Appl. Opt.* **4**, 1275 (1965).
- ⁴²R. Petit and G. Tayeb, *J. Opt. Soc. Am. A* **A7**, 1686 (1990).
- ⁴³M. Nevière, in *Electromagnetic Theory of Gratings* (Ref. 38), p. 123.
- ⁴⁴R.W. Wood, *Philos. Mag.* **4**, 396 (1902).
- ⁴⁵U. Fano, *J. Opt. Soc. Am.* **31**, 213 (1941).