Femtosecond pump-probe near-field optical microscopy

B. A. Nechay, a) U. Siegner, M. Achermann, H. Bielefeldt, b) and U. Keller
Swiss Federal Institute of Technology Zurich, Institute of Quantum Electronics, ETH Hönggerberg-HPT, CH-8093 Zürich, Switzerland

(Received 8 December 1998; accepted for publication 2 March 1999)

We have developed an instrument for optically measuring carrier dynamics in thin-film materials with $\sim 150 \text{ nm}$ lateral resolution, $\sim 250 \text{ fs}$ temporal resolution, and high sensitivity. This is achieved by combining ultrafast pump-probe laser spectroscopic techniques, which measure carrier dynamics with femtosecond-scale temporal resolution, with the nanometer-scale lateral resolution of near-field scanning optical microscopes (NSOMs). We employ a configuration in which carriers are excited by a far-field pump laser pulse and locally measured by a probe pulse sent through a NSOM tip and transmitted through the sample in the near field. A novel detection system allows for either two-color or degenerate pump and probe photon energies, permitting greater measurement flexibility over earlier published work. The capabilities of this instrument are proven through near-field degenerate pump-probe studies of carrier dynamics in GaAs/AlGaAs single quantum well samples locally patterned by focused-ion-beam (FIB) implantation. We find that lateral carrier diffusion across the nanometer-scale FIB pattern plays a significant role in the decay time of the excited carriers within $\sim 1 \mu m$ of the implanted stripes, an effect which could not have been resolved with a far-field system. © 1999 American Institute of Physics. [S0034-6748(99)03106-8]

I. INTRODUCTION

Many dynamical processes in condensed matter and molecular or atomic systems take place on picosecond or sub-picosecond time scales. Examples include energy relaxation in semiconductor systems and charge transfer processes in molecular nanostructures. Though a variety of measurement techniques can probe such dynamical processes, only all-optical techniques have reliably achieved subpicosecond temporal resolution. Ultrafast optical spectroscopic techniques have yielded a wealth of information about dynamics of electronic excitations in a variety of materials. However, the diffraction limited lateral resolution of conventional far-field techniques limits the physical information which can be obtained about the properties of nanometer-scale lateral structures, since their inherent inhomogeneity cannot be spatially resolved. Furthermore, far-field measurements can tell us little about the transport mechanisms on the submicron lateral scale. The investigation of ultrafast dynamical processes on the nanometer scale has gained increasing interest in recent years with improvements in nanofabrication technologies leading to nanostructures with potential applications in fields such as lasers and electronic devices. Therefore, combining ultrafast all-optical spectroscopic techniques with the advantages of high spatial resolution in an instrument that measures optical signals in the near field would prove invaluable in exploring such processes. Although alternate means of measuring near-field optical signals exist, near-field scanning optical microscopy (NSOM), in which a nanometer-scale aperture at the end of a fiber is scanned in the near field of a sample, is the most established and easiest to interpret. NSOM has proven useful in exploring nanometer-scale physics and chemistry in a variety of materials. Recently, a number of groups have successfully combined NSOM with ultrafast optical spectroscopic techniques to measure carrier dynamics with femtosecond-scale temporal resolution and nanometer-scale lateral resolution. This article presents the development of a new instrument of this class, which combines a femtosecond pump-probe technique with NSOM, and discusses the issues involved. A novel detection system design adds measurement flexibility by allowing for both degenerate and two-color measurements. Ultrafast measurements of carrier dynamics in nanometer-scale ion implanted single quantum well samples demonstrate the capabilities of the system—$\sim 150 \text{ nm}$ spatial resolution, $\sim 250 \text{ fs}$ time resolution, and high sensitivity—and show that information about physical processes can be obtained that is not accessible with far-field systems.

II. BACKGROUND

In pump-probe experiments, a pump pulse excites carriers in the sample. These excited carriers modify the absorption and the refractive index probed by a subsequent probe pulse. The measurement of the differential probe transmission signal, i.e., $\Delta T = T_{\text{probe on}} - T_{\text{probe off}}$, or reflection signal, i.e., $\Delta R = R_{\text{probe on}} - R_{\text{probe off}}$, versus pump-probe time delay gives a measure of the dynamics of the excited carrier population. Here, $\Delta T = T_{\text{pump on}} - T$ and $\Delta R = R_{\text{pump on}} - R$ (i.e., the pump-induced changes in the probe), $T$ and $R$ are the linear transmission and reflection of the probe, respectively. $P_{\text{probe}}$ is the probe power incident on the sample, and $\eta$ is the collection

a)Electronic mail: nechay@iqe.phys.ethz.ch
b)Present address: Experimentalphysik VI, Universitaet Augsburg, D-86135 Augsburg, Germany.
efficiency of the detection system. In the simplest case, $\Delta T$ and $\Delta R$ are proportional to the excited carrier concentration.\textsuperscript{1}

The time resolution of such a measurement is limited only by the pump and probe pulsewidths. In all but a small handful of experiments, pump and probe excitations are done in the far field—limiting the achievable lateral resolution to the diffraction limit of $\lambda/2$. However, this limitation can be circumvented by incorporating the lateral resolution of NSOM.

In NSOM, a NSOM probe tip confines light to subwavelength dimensions by a nanometer-scale aperture at its apex. By raster scanning the tip across the sample in the near field, i.e., at tip-sample distances much less than the wavelength, nanometer-scale spatial variations of optical properties of the sample can be observed. The NSOM probe tip typically consists of an optical fiber which has been tapered to nanometer-scale dimensions through either pulling\textsuperscript{9} or etching\textsuperscript{11} techniques and then aluminum coated at $90^\circ$ angles, forming a small, typically $\sim 100$ nm size aperture. This aperture size, along with the tip-sample flying height and thickness of the active region of the sample, determines the lateral resolution of NSOM.

One can envision various configurations combining femtosecond pump-probe techniques with NSOM. First, the pump and probe pulses can excite the sample either globally (i.e., far field) or locally (i.e., through the NSOM tip), allowing for three possibilities: global pump/local probe, local pump/global probe, and local pump and probe. Second, the probe signal can be measured either in reflection or transmission. Last, the excitation and detection can either be at two different wavelengths (i.e., two color) or they can be degenerate.

A few of these configurations have been realized by various groups. Stark \textit{et al.}\textsuperscript{7} used a two-color global pump/local transmitted probe configuration, where spectral filtering was used to avoid the detection of background pump light. Smith \textit{et al.}\textsuperscript{8} used a degenerate equal-pulse correlation setup\textsuperscript{12} with both pump and probe sent through the NSOM tip. Vertikov \textit{et al.}\textsuperscript{10} used a global pump with the reflected probe measured in near-field collection mode, and used a two-color configuration to separate the differential reflection signal from the background pump light. Last, though not a pump-probe experiment according to the definition used in this article, Levy \textit{et al.}\textsuperscript{9} measured luminescence intensity autocorrelation using two equal power global pump pulses and picking up the luminescence with the NSOM tip. In this experiment, the higher pump photon energy was spectrally filtered out from the luminescence. In the system described in this article, a global pump/local transmitted probe configuration was chosen, shown in Fig. 1(a), with a novel detection system that allows for both degenerate or two-color excitation. This novel combination allows for a higher degree of measurement flexibility than earlier published work.\textsuperscript{7–10}

The choice between global or local pump excitation is largely determined by the requirements of the chosen experiment. Global pump excitation was chosen mainly due to the greater excitation intensity that can be applied in the far field, as opposed to the near field. Specifically, pump excitation through the NSOM tip is limited in the maximum excitation intensity due to the very low power throughput of NSOM tips and the tip damage threshold of a few mW of input power to the fiber.\textsuperscript{13,14} In global pump excitation, the excitation intensity is only limited by the damage thresholds of the sample so potentially larger nonlinear responses can be measured.

Since global pump excitation was chosen, the differential probe transmission signal had to be measured locally with the NSOM tip. The main consideration in the decision between measuring in illumination versus collection mode was signal-to-noise. The general experience in the NSOM experiment community is that there is considerably more signal to work within illumination mode NSOM, as opposed to collection mode.\textsuperscript{4} Therefore, the higher efficiency of illumination mode NSOM determined our preference for this configuration.

Signal-to-noise was also the main consideration in choosing a transmission geometry over a reflection geometry. Specifically, for samples which have comparable total reflected and transmitted powers, one would expect that a larger power can be measured in transmission than in reflection mode due to a larger solid angle of the collection optics in transmission. As for two-color versus degenerate experiments, the measurements discussed in this article were all done with degenerate pump probe, which suited the physics being explored for these samples.

### III. EXPERIMENTAL SETUP

The experimental setup is shown in Fig. 1(b). The home-built NSOM instrument consists of either a pulled or etched NSOM tip whose aperture is held within $\sim 10$ nm of the sample surface using shear-force feedback.\textsuperscript{15} Piezoelectric tuning fork shear force detection\textsuperscript{16} is used to avoid the effects of scattered light which can be present in optical feedback techniques. In our case, we use a 32768 Hz quartz tuning fork from Grieder Bauteile AG. The sample stage design and coarse and fine tip approach systems are similar to the designs described in Ref. 4, Secs. 5.3.4 and 5.5, respectively, while the coarse sample positioning is done with a manual translation stage. Both the feedback-controlled tip-sample flying height and the sample scanning are controlled by commercial Park Scientific Instruments Scanning Force Microscope electronics, which was adapted for the NSOM.

A mode-locked laser source (either a home-built Ti:sapphire or commercial Cr:LiSAF laser) generates a train of
ometry. These background signals dwarf the differential power, which cannot be easily avoided with the NSOM geometry unaffected by the pump and the backreflected pump. Specifically, as mentioned in the previous section, the input pump and probe pulse trains are modulated at 1 and 1.05 MHz, respectively. The effect of pump excitation on probe transmission corresponds to a nonlinear response which mixes the pump and probe signals, introducing a differential probe transmission signal at the 50 kHz difference frequency, which can be measured with lock-in detection. The high frequencies of the pump and probe modulation reduce thermal effects of the tip-sample interaction and, together with the high difference frequency, reduce the effect of laser noise.

Despite these measures, the unaffected probe and the back-reflected pump can still affect the measurement by increasing the measured shot and laser noise or even saturating the detector. Unfortunately, detection of the unaffected probe signal cannot be avoided, since it cannot be purely optically separated from the desired differential probe signal. Fortunately, this unaffected probe signal is small enough that it does not introduce significant problems, only slightly increasing the noise floor above detector dark noise. However, the back-reflected pump signal is typically many orders of magnitude larger, so that measures need to be taken to reduce this signal as much as possible. In order to avoid the noise and saturation problems associated with detected pump signal, other groups have used two-color setups with spectral filtering or symmetric arrangements of pump and probe through equal-pulse correlation, sending both pump and probe through the fiber so that both the transmitted pump and unaffected transmitted probe background signals are small. In the experiment described in this article, two measures are taken to reduce the back-reflected pump while avoiding the wavelength limitations of the two-color requirement and the limitations on the measurement flexibility in the equal-pulse setup. These are shown in Fig. 2. First, the pump and probe beams are orthogonally polarized so that a Glan–Thomson polarizer filters out the pump by a factor of typically more than 100, and even as high as 500, while passing most of the probe. Though the polarization of the probe light output from the NSOM tip is not perfectly linear, measurements reveal that the polarization stays quite constant over the pump-probe measurement period. Therefore, no artifacts are introduced into the pump-probe scan by the polarization sensitive

\[ \Delta T^* P_{probe} \eta \]

\[ \Delta T/T, \sim 0.5\% \]

This pW range signal is accompanied by large background signals which include the transmitted probe power which is unaffected by the pump and the backreflected pump power, which cannot be easily avoided with the NSOM geometry. These background signals dwarf the differential probe transmission signal. Therefore, modulation and lock-in detection is used to differentiate between the differential probe transmission and the unaffected probe and back-reflected pump. Specifically, as mentioned in the previous section, the input pump and probe pulse trains are modulated at 1 and 1.05 MHz, respectively. The effect of pump excitation on probe transmission corresponds to a nonlinear response which mixes the pump and probe signals, introducing a differential probe transmission signal at the 50 kHz difference frequency, which can be measured with lock-in detection. The high frequencies of the pump and probe modulation reduce thermal effects of the tip-sample interaction and, together with the high difference frequency, reduce the effect of laser noise.

Despite these measures, the unaffected probe and the back-reflected pump can still affect the measurement by increasing the measured shot and laser noise or even saturating the detector. Unfortunately, detection of the unaffected probe signal cannot be avoided, since it cannot be purely optically separated from the desired differential probe signal. Fortunately, this unaffected probe signal is small enough that it does not introduce significant problems, only slightly increasing the noise floor above detector dark noise. However, the back-reflected pump signal is typically many orders of magnitude larger, so that measures need to be taken to reduce this signal as much as possible. In order to avoid the noise and saturation problems associated with detected pump signal, other groups have used two-color setups with spectral filtering or symmetric arrangements of pump and probe through equal-pulse correlation, sending both pump and probe through the fiber so that both the transmitted pump and unaffected transmitted probe background signals are small. In the experiment described in this article, two measures are taken to reduce the back-reflected pump while avoiding the wavelength limitations of the two-color requirement and the limitations on the measurement flexibility in the equal-pulse setup. These are shown in Fig. 2. First, the pump and probe beams are orthogonally polarized so that a Glan–Thomson polarizer filters out the pump by a factor of typically more than 100, and even as high as 500, while passing most of the probe. Though the polarization of the probe light output from the NSOM tip is not perfectly linear, measurements reveal that the polarization stays quite constant over the pump-probe measurement period. Therefore, no artifacts are introduced into the pump-probe scan by the polarization sensitive
A second method of reducing the back-reflected pump light involves a confocal arrangement consisting of a 50 μm pinhole that spatially filters out the light outside of a 1.25-μm-diam sample area, thereby reducing the back-reflected pump by as much as a factor of 100.

These precautions significantly improve the attainable signal-to-noise ratio of our system, allowing us to measure noise-equivalent powers (NEP) of differential probe transmission signals typically below 1.4 pW (rms), for a 1 s lock-in time constant. For a 10 nW (rms) transmitted probe, this NEP is 1.1 pW (rms), corresponding to a sensitivity in the differential probe transmission ratio, ΔT/T, of ~1 × 10^{-4}.

The dominant source of the measured noise in the differential probe transmission signal depends on the total power of the background signals, including both back-reflected pump and unaffected probe. Specifically, for the lowest background signal levels, the measured noise is determined by the dark noise of the detector which, for the case of the avalanche photodiode (APD) module—i.e., APD plus amplifier—used as the photodetector in this system, gives a dark NEP of typically 350 fW/√Hz. At moderate background signal levels, the measured noise would be dominated by shot noise, which is proportional to the square root of the total detected background signal power. Laser noise grows linearly with the background signal power and only would become significant once one exceeds ~1 μW of detected background signal. However, since the background signals in this system range from ~5 to ~100 nW, the noise in the differential probe transmission signal is largely dominated by shot noise or detector dark noise.

This has effects in the measurable signal to noise ratio. Specifically, if one assumes a small backreflected pump power, then increasing the probe power (for example, by increasing tip aperture diameter) much beyond 1 μW would not improve the signal-to-noise ratio since the laser noise is proportional to the probe power. However, in regions where the differential probe transmission noise is dominated by shot noise or dark noise, i.e., at detected background signal powers below ~1 μW, one can improve the signal-to-noise by increasing the probe power.

One artifact of the commercial APD module detection is an electrical mixing of the pump and probe pulses in the module. This introduces a time-delay dependent background signal at the 50 kHz difference frequency at which the differential probe transmission signal is measured. For large back-reflected pump signals, this can be significant enough to warp the pump-probe trace. In all the measurements shown in this article, this mixing effect was sufficiently suppressed by the reduction of the background signals. Nevertheless, new APD designs are being explored which avoid this mixing signal and which reduce the dark noise.

Besides signal-to-noise considerations, time resolution and the limits imposed on it by group velocity dispersion (GVD) is also an important issue. Of highest concern is the GVD of the NSOM fiber, a meter of which can stretch a 100 fs pulse at 800 nm to a chirped 1 ps pulse. Therefore, a GVD precompensation setup was introduced before the fiber coupler in order to add a negative GVD to the pulse, which balances the positive GVD of the ~50 cm length of fiber. The GVD precompensator consisted of a standard two-prism setup, in which different spectral components of the pulse travel different distances, introducing a net negative GVD which depends on prism spacing and insertion. Figure 3(a) shows the pulselength of the output of a 46 cm open-ended fiber after precompensation, deconvoluted from measured autocorrelation traces, as a function of prism separation, for a 90 fs input pulse. As can be seen, the pulselength can be largely recovered with proper GVD optimization. This precompensation should not be affected by the NSOM tip itself, which does not show much effect on GVD.

One issue of consideration in sending ultrastable pulses through NSOM fibers is the possibility that nonlinear effects in the fiber, such as self phase modulation and stimulated Raman scattering, can warp the spectrum of the measured output pulse. To explore this effect, spectral measurements were made on the outputs of open-ended NSOM fibers, where the ~100 fs input pulses were first precompensated for the NSOM fiber length with the prism setup described in the previous paragraph. These measurements are shown in FIG. 4. Spectra of precompensated fiber output pulses for various fiber output powers and fiber input spectrum as a reference.
Fig. 4. One can see that, for 25 mW output power, the output pulse spectrum is significantly warped with respect to the input pulse spectrum—revealing significant nonlinear effects. However, for output powers below ~4 mW, no significant nonlinear effects were observed. Though these results were measured with open ended fibers, it has been shown that the subwavelength aperture does not have an effect on the spectrum. Therefore, since optical powers within NSOM fibers cannot be much higher than 4 mW due to tip damage, the nonlinear effects are not significant.

Last, a major issue of concern in most NSOM measurements are topographical artifacts in which the topography of the measured surface can alter the optical image, independent of the variation in optical properties. This is of lesser concern in this instrument since the main physics which are explored involves the temporal dependence of the carrier dynamics—i.e., the temporal shape of the pump-probe scan—and how this temporal dependence varies across the sample, issues in which the topographical artifacts are not significant. However, for samples with larger topographical variations, topographical artifacts must be taken into account when interpreting spatial variations in pump-probe amplitudes and in determining spatial resolution.

V. INSTRUMENT PERFORMANCE—FIB SAMPLE RESULTS

In order to prove the capabilities of the system in terms of its performance and its usefulness for measuring interesting physics, samples needed to be measured which had a reasonably predictable lateral variation in carrier dynamics. The samples chosen were undoped 80 Å single quantum well GaAs/Al0.3Ga0.7As samples with a 12 nm top barrier, which were mounted on glass disks and selectively etched to remove the opaque substrates, thus allowing for transmission experiments. Lateral patterning of damage on the nanometer scale was introduced using FIB implantation of Ga ions at 50 keV energies. Such ion implantation is known to reduce the measured nonlinearity and to introduce trap states, leading to fast carrier trapping times. These samples then had the advantage of having a flat topographical profile, transition energies suitable to the available 840 nm laser sources, and a designed lateral profile of ion damage which was known to have strong effects on carrier dynamics. Furthermore, the thin top barrier avoided surface trapping effects while the optically active well layer stayed close to the sample surface, thus maintaining good achievable lateral resolution. Two samples were measured, both using the same GaAs/Al0.3Ga0.7As quantum well structures, which varied only in their FIB implantation pattern—200 nm implanted stripes at a ~3 × 10^12 ions/cm^2 dose with 400 nm spaces between stripes for sample 1, and 100 nm stripes at a ~8 × 10^11 ions/cm^2 dose with 2 μm spaces for sample 2. All measurements were performed at 300 K, exciting at the band edge with a pump fluence of ~4 μJ/cm^2.
The gray region corresponds to the implanted stripe. Therefore, since the fast decay times cannot be explained by any local characteristic of the material, it must be due to a nonlocal effect of transport. Since drift effects are expected to be insignificant, we conclude that, besides the slow recombination of electron hole pairs, the carrier dynamics in the unimplanted regions of the sample are strongly affected by diffusion, which has also been shown to affect the decay of carriers in other types of semiconductor nanostructures.\(^8,8,26\) Specifically, the initial excited carrier concentration is reduced at the implanted stripes due to fast trapping. The resulting concentration gradient causes carriers in the unimplanted regions to diffuse toward the implanted stripes and get trapped. Therefore, the excited carrier concentrations in the unimplanted regions are quickly depleted by this diffusion process, at a rate which is much faster than the recombination rate and which is relatively independent of the distance across the implantation pattern.

The effect of diffusion on carrier dynamics can also be seen in measurements of sample 2, whose 2 \(\mu\)m spacing allowed us to study carrier dynamics farther away from the FIB stripe. A higher-throughput, \(~300\) nm aperture size NSOM tip was used here. Figure 8 shows the pump-probe amplitude and fast decay time from a double-exponential fit as a function of distance across the FIB pattern. Indeed, the decay time is roughly constant over a distance of \(~800\) nm around the implanted stripe—consistent with the results seen on sample 1—and increases to a value equal to the measured recombination time of 250 ps half way between the implanted stripes. The spatial profile of the diffusive component of the decay time roughly follows a quadratic distance dependence.\(^2^7\) This is plausible since carriers excited close to the edges of the implanted stripe have a shorter distance to diffuse, resulting in enhanced diffusion, whereas diffusion of carriers excited \(~1\) \(\mu\)m from the stripes is suppressed so that the decay in the carrier population is dominated by recombination. It is clear that the strong variation of the diffusion dynamics on a sub-\(\mu\)m scale could not have been resolved with a far-field system. These results demonstrate that the development of a femtosecond-scale time-resolved NSOM opens up new possibilities for measuring the physics of carrier dynamics in nanometer-scale structures.

ACKNOWLEDGMENTS

The authors would like to acknowledge Dieter Pohl and Renato Zenobi for access to their NSOM tip fabrication facilities; Gunter Steinmeyer for helpful advice on detector considerations; and Andreas Schertel of Micron, GmbH for FIB implantation. This work was supported by the Swiss National Science Foundation, program NFP 36.

---