UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Quantifying nanoscale carrier diffusion with ultrafast optical and photocurrent microscopy

by

Alexander Block

Quantifying Nanoscale Carrier Diffusion with Ultrafast Optical and Photocurrent Microscopy

Doctoral Thesis

by

Alexander Block

Thesis Advisor:
Prof. Dr. Niek F. van Hulst

ICFO – Institut de Ciències Fotòniques
BIST – Barcelona Institute of Science and Technology
UPC – Universitat Politècnica de Catalunya

September 2019

Thesis advisor:

Prof. Dr. Niek F. van Hulst

Thesis committee:

Prof. Dr. Markus Lippitz (University of Bayreuth, Bayreuth, Germany)
Prof. Dr. Mischa Bonn (Max Planck Institute for Polymer Research, Mainz, Germany)
Prof. Dr. Simon Wall (ICFO – Institut de Ciències Fotòniques, Castelldefels, Spain)

Abstract

Heat transport in solids is one of the oldest problems in physics, dating back to the earliest formulations of thermodynamics. The classical laws of heat conduction are valid as long as the observed time and length scales are larger than the relaxation time and mean free path of the underlying microscopic heat carriers, such as electrons and phonons. With the advent of ultrafast lasers and nanoscale systems these regimes can now be surpassed and new refined models of heat transport are needed. In particular, the interaction of ultrashort light pulses with matter can excite electrons to high temperatures, leading to a local non-equilibrium of electrons and phonons. Under these conditions, also the transport properties of the carriers are altered.
So far, these effects have typically been studied in the time domain. The cooling of photo-excited hot electrons has been studied both in metals as well as novel 2D materials, such as graphene. However, due to a lack of spatio-temporal resolution, it has not been possible to distinguish the effects of hot-electron diffusion from other cooling mechanisms, such as electron-phonon coupling.
In this thesis, I directly track such ultrafast heat and carrier diffusion in space and time with ultrafast microscopy. By using the recently developed technique of probe-beam-scanning transient-absorption microscopy on thin gold films I directly resolve, for the first time, a transition from hot-electron diffusion to phonon-limited diffusion on the picosecond timescale. I support the understanding of these complex dynamics by theoretical modeling of the thermo-optical response based on a two-temperature model.
I apply the same technique to study hot carrier diffusion in atomically thin monolayer graphene. By comparing differently prepared samples, I study the strong influence of external parameters, such as production type, substrate, and environment on carrier diffusion.
Finally, I study hot carrier diffusion in exfoliated and encapsulated graphene devices with a novel technique of ultrafast spatio-temporal photocurrent microscopy based on the photothermoelectric effect. I extract diffusion dynamics for electrically characterized samples with the help of theoretical spatio-temporal modeling, thereby testing the fundamental relationship between electrical and thermal carrier transport.
The precise quantification of ultrafast and nanoscale carrier transport with these state-of-the-art techniques leads to a broader understanding of non-equilibrium dynamics and could ultimately help the design, optimization, and heat management of the next generation of ultra-compact (opto-) electronic devices, such as solar cells, photodetectors, or integrated circuits.

Resumen

El transporte de calor en sólidos es uno de los problemas más antiguos de la física, que se remonta a las primeras formulaciones de la termodinámica. Las leyes clásicas de la conducción de calor son válidas cuando las escalas de tiempo y longitud observadas son mayores que el tiempo de relajación y la trayectoria libre media de los portadores de calor microscópicos subyacentes, como los electrones y los fonones. Con la llegada de los láseres ultrarrápidos y los sistemas a nanoescala, estos regímenes ahora pueden superarse por lo que se necesitan nuevos modelos refinados de transporte de calor. En particular, la interacción de pulsos de luz ultracortos con la materia puede excitar electrones a altas temperaturas, lo que lleva a un desequilibrio local de electrones y fonones. En estas condiciones, también se modifican las propiedades de transporte de los portadores de calor.

Hasta ahora, estos efectos han sido típicamente estudiados en el dominio del tiempo. El enfriamiento de electrones calientes fotoexcitados se ha estudiado tanto en metales como en nuevos materiales bidimensionales, como el grafeno. Sin embargo, debido a la falta de resolución espacio-temporal, no ha sido posible distinguir los efectos de la difusión de electrones calientes de otros mecanismos de enfriamiento, como el acoplamiento de electrones y fonones.

En esta tesis, hago un seguimiento directo de la difusión del calor y sus portadores en el espacio y el tiempo con microscopía ultrarrápida. Al utilizar la técnica recientemente desarrollada de microscopía de absorción transitoria con escaneo de sonda en películas de oro delgadas, resuelvo directamente, por primera vez, una transición de la difusión de electrones calientes a la difusión limitada por fonones en la escala de tiempo de picosegundos. Apoyo la comprensión de estas dinámicas complejas mediante el modelado teórico de la respuesta termo-óptica basada en un modelo de dos temperaturas. Aplico la misma técnica para estudiar la difusión de portadores calientes en una capa de grafeno atómicamente delgado. Al comparar muestras preparadas de manera diferente, estudio la fuerte influencia de los parámetros externos, como el tipo de producción, el sustrato y el entorno sobre la difusión del portador.

Finalmente, estudio la difusión de portadores en dispositivos de grafeno exfoliados y encapsulados con una técnica novedosa de microscopía de fotocorriente espacio-temporal ultrarrápida basada en el efecto fototermoeléctrico. Extraigo dinámicas de difusión para muestras caracterizadas eléctricamente con la ayuda del modelado espacio-temporal teórico, probando así la relación fundamental entre el transporte eléctrico y térmico.

La cuantificación precisa del transporte de los portadores ultrarrápido y a nanoescala con estas técnicas de vanguardia lleva a una comprensión más amplia de la dinámica del no equilibrio y podría, en última instancia, ayudar al diseño, la optimización y la

gestión del calor de la próxima generación de dispositivos (opto-) electrónicos ultra-compactos, como células solares, fotodetectores o circuitos integrados.

Contents

List of abbreviations

2DEG	Two-dimensional electron gas. 10
AOM	Acousto-optical modulator. 35
APD	Avalanche photodiode. 34
BBO	β-barium borate. 33
BFP	Back focal plane. 39
CVD	Chemical vapor deposition. 44
CW	Continuous wave. 33
EOM	Electro-optical modulator. 35
ESA	Excited state absorption. 17
FDTR	Frequency-domain thermoreflectance. 18
FET	Field effect transistor. 10
FoV	Field of view. 41
FWHM	Full-width at half-maximum. 13
galvo	Galvanometric. 33
GDD	Group delay dispersion. 35
GSB	Ground state bleach. 17
HB	Hall bar (sample name). 74
hBN	Hexagonal boron nitride. 44
IRF	Instrument response function. 35
lock-in	Lock-in amplifier. 43
MSD	Mean square displacement. 13
NA	Numerical aperture. 18
NIR	Near infrared. 9
PBS-TAM	Probe-beam-scanning transient-absorption microscopy. 21
PC	Photocurrent. 76
PSF	Point spread function. 18
PTE	Photothermoelectric. 75
SE	Stimulated emission. 16
SEM	Scanning electron microscope. 41
SHG	Second harmonic generation. 33
SS-TAM	Sample-scanning transient-absorption microscopy. 20
SUTM	Scanning ultrafast thermo-modulation microscopy. 51
TA	Transient absorption. 16
TAM	Transient-absorption microscopy. 18
TBP	Time-bandwidth product. 36
TC	Two contact (sample name). 74

TCSPC	Time-correlated single photon counting. 18
TDTR	Time-domain thermoreflectance. 18
TEM	Transmission electron microscope. 64
Ti:sapph	Titanium sapphire. 33
TMD	Transition metal dichalcogenide. 10

1 Introduction

1.1 Heat transport: Past and present

The transport of heat is one of the oldest problems in physics, with a line of research traceable from Isaac Newton to Albert Einstein and beyond. This research has helped to connect the macroscopic understanding of thermodynamics to the microscopic world of tiny invisible particles carrying the thermal energy. Amazingly, the topic is still very relevant today and promises to stay so as technology advances. The emergence of nanoscale systems and ultrafast switching rates continues to push the boundaries of our understanding of the flow of electrons, phonons, and more exotic quasi-particles through solid-state systems. As the length scales of such devices becomes smaller than the mean free path and the modulation time scales become shorter than the relaxation times of the carriers, deviation from the classical laws creates the need for new theory to explain the dynamics. Ongoing theoretical and experimental advances promise to push humankind towards the creation and optimization of efficient future electronic and photonic nanoscale and ultrafast technologies.

Newton's law of cooling, interpreted from a modern perspective from his paper of 1701, states that the rate of cooling of a hot object is proportional to the difference in temperature of the object and the surrounding environment[1]. This work laid the foundations for what would become the first rigorous mathematical description of heat conduction by Joseph Fourier in 1822, a law which has held its ground for nearly two centuries[2]. More than 30 years later, in 1855, Adolf Fick found that the macroscopic transfer of salt concentration in aqueous solutions due to the random motion of microscopic particles (already called molecules at the time) is governed by the same differential equation as Fourier's law describing transfer of heat by conduction[3]. Therefore, both processes are essentially described by the same mathematical theory, today known as Fick's laws of diffusion.

This macroscopic description of the transport of heat and its analogy to particle diffusion paved the way for the idea that heat in a solid, too, could be carried by a random motion of microscopic particles, which was thought of as impossible during Fick's lifetime*.
The same decade showed major advances in the kinetic theory of gases by Krönig, Clausius, and Maxwell, connecting the concepts of molecular motion and the temperature in a gas[5–7].
Finally, in 1905, Albert Einstein famously demonstrated that a statistical description of the random molecular (Brownian) motion leads to the same diffusion law as derived from macroscopic thermodynamic concepts, such as heat and entropy[8]. This important link between macroscopic thermodynamic observables and underlying statistical mechanics of microscopic particles lead to the modern view in which thermal energy is carried by the random motion of different microscopic particles.
Since then, enormous progress has been made in the understanding of heat conduction in solids. Both classical and quantum descriptions of the main carriers of heat, the phonon and the electron, have been made by describing their dispersion relation, density of states, carrier statistics, scattering interactions, as well as mean free path and carrier speed[9]. The basic idea is the following: In all types of solids, heat can be transported by phonons, i.e., quantized lattice vibrations, which follow Bose-Einstein statistics and can be divided spectrally into acoustic and optical modes. This phononic thermal conduction is the dominant heat transport mechanism for electrically insulating materials. For electrical conductors, which are in general also good thermal conductors, heat is more efficiently transported by free electrons, following Fermi-Dirac statistics. In a semiclassical picture both types of heat carriers can be thought of as moving freely (on straight lines) for an average distance, called the mean free path, before scattering with other electrons, phonons, as well as lattice defects, impurities, and grain boundaries[10].
Although many decades of research have contributed to the microscopic description of heat, greatly advancing our understanding of the (quantum) nature of heat carriers, the classical macroscopic laws of heat diffusion are still used to describe heat conduction in solids under the basic assumption of a local thermal equilibrium. Although temperature might vary in space and time, the dynamics are still well described by Fourier's law, as long as the length scales are large compared to the mean free path of the heat carriers and variations or perturbations are slow compared to the internal relaxation times of the carriers, i.e., slow enough to establish a local thermodynamic equilibrium in the vicinity around every point.
The advent of nanoscale systems, ultrafast switching, and novel materials, have all lead to the breakdown of Fourier's law and the need for new types of descriptions, which can be summarized under the keyword of *non-equilibrium thermodynamics*[11].
Recent years have shown exciting new effects in thermal transport, such as heat waves, ballistic motion, as well as hydrodynamic carrier motion, which can only be understood in a framework beyond the classical laws of heat conduction[12–14]. These effects typically occur due to nanoscale confinement, e.g. to lower dimensions, such as nanowires, ultra-thin films, or novel 2D materials such as graphene[15]. In particular, the advent of ultrafast lasers, as well as electronic nanoscale systems, has made evident the in-

*This was four decades before experimental evidence for the existence of the electron by J.J. Thompson[4]

triguing phenomenon of a situation in which a material's electrons and lattice phonons are not in a thermal equilibrium. When the electronic system is partially decoupled from the phonons and high temperature "hot electrons" need to be accounted for in two-temperature models for novel transport dynamics and thermoelectric effects[16].
Therefore, there is an ongoing effort in both fundamental and applied research to understand and describe situations where heat is transported differently by electrons and phonons out of equilibrium.
The aim to disentangle the electronic and phononic contributions to thermal transport has important implications for technological applications, as well. Modern optoelectronic devices, such as solar cells and photodetectors, rely on the direct energy transfer from light to the electronic or excitonic system. The subsequent efficient, long range transport of these charges is crucial as they contribute to a voltage and ultimately to the storage of the excess energy in form of electrical (potential) energy, while any transfer of this energy to lattice vibrations (phonons) acts as a loss channel[17]. Similarly, as integrated electronic and photonic circuits are constantly miniaturized towards the nanoscale, their thermal management becomes critical[18]. Recently, it was proposed that by harvesting hot electrons, the efficiency of solar cells could be theoretically boosted past the Shockley-Queisser limit, from 33% to up to 66%[19,20].

1.2 This thesis: Electrons versus phonons

Inspired by these exciting new effects, the work presented in this thesis aims at disentangling and following both electronic and and phononic heat and their coupling directly in space and time in such non-equilibrium situations. There are two main experimental difficulties to disentangle electron and phonon thermal conductivity and to study non-equilibrium dynamics. The first is that, once created, temperature differences always tend to equilibrate according to the second law of thermodynamics. In case of electron-phonon coupling this is typically very fast, on the order of femtoseconds to picoseconds[21]. Secondly, to be able to tell apart the contributions from electronic and phononic heat, one needs an experimental observable which can distinguish the two.
In this thesis, I attack these challenges by two conceptual routes, namely the use of ultrafast lasers, and by choice of the right experimental observable, from transient absorption to photocurrent.
To this end, I describe an experimental scheme which aims at directly tracking, in space and time, the electronic heat after pulsed laser illumination. This is done in different solid state systems such as metals, semiconductors, as well as single layer graphene. The basic idea here is to separate electronic and phononic heat transport by studying the dynamics on femtosecond time scales, i.e., faster than the typical electron-phonon thermalization time of a few picoseconds, via transient-absorption microscopy. Another route to disentangle electronic from the phononic transport is to combine ultrafast laser illumination with a photocurrent read out. Compared to transient absorption, which can be modulated by both electronic and lattice heat, the *current* signal is inherently electronic in nature, as it can only be generated by conduction electrons as their charge flow is measured. The photothermoelectric current is an ideal observable, as it has been shown to directly probe the electron heat[22]. Hence, I describe a spatio-temporal transient photocurrent experiment, based on the photother-

moelectric effect on high quality dual-gated graphene devices, aimed at extracting the electronic part of hot carrier diffusion in electrically characterized devices.

1.3 Thesis outline

- In chapter 2 I describe a selection of basic concepts of heat diffusion in solids that are important in the context of this thesis. These include the different types of thermal transport for varying classes of materials, as well as the concepts of electron and phonon heat. I give a general mathematical introduction to the heat equation and its solutions as a framework to quantify diffusion experimentally.

- In chapter 3 I give a brief introduction to the technique of transient absorption and its applicability to study transport phenomena in solids. I consider the combination of transient-absorption spectroscopy with microscopy to resolve such dynamics. I describe the expected point-spread functions and their evolution mathematically. By presenting exemplary simulations of different diffusion dynamics I aim to show the limitations for quantifying diffusion with this technique.

- In chapter 4 I describe the experimental building blocks for spatio-temporally resolved transient-absorption microscopy. I go into detail about a selection of experimental topics, such as calibrating the spatial and temporal resolution and quantifying the signal-to-noise. I present spatio-temporal carrier diffusion dynamics on a bulk silicon sample, as well as atomically thin molybdenum diselenide.

- In chapter 5 I present the experimental results of ultrafast heat diffusion in thin gold films. I resolve, for the first time, a transition from hot-electron to phonon-limited diffusion dynamics with transient-absorption microscopy. I accompany the experiment with a full 3D thermo-optical response calculation based on a two-temperature model to guide the interpretation of the observations.

- In chapter 6 I use transient-absorption microscopy to track the hot carrier dynamics in monolayer graphene. The effect of production type, substrate and environment is probed by comparing the dynamics in the differently prepared samples, including encapsulated and suspended graphene.

- In chapter 7 I continue the study of hot carrier diffusion in graphene with a different technique. I present, for the first time, a spatio-temporally resolved measurement of the ultrafast light induced hot carriers by probing the photothermoelectric current produced at the p-n junction of dual-gated, encapsulated graphene device. The measurements are interpreted with the help of a model, simulating the observed photocurrent signals and their dependence on carrier diffusion.

- Finally, in chapter 8 I give a brief summary of the work in this thesis as well as an outlook of possible future developments based on this work.

2

Heat diffusion in solids: Theoretical considerations

In this chapter I give an introduction to the general topic of heat transport in different materials.
I discuss, in general terms, the physics of heat and carrier transport via diffusion in different classes of solids, such as metals, semiconductors, and graphene. This diffusion occurs due to different underlying carriers, such as electrons, phonons and excitons and happens on different time scales. Thus, it is linked to different physical properties, such as the density of states and dispersion relations of the carriers, as well as different macroscopic material properties, such as thermal conductivity and heat capacity. Therefore, a vast range of underlying physics can be investigated. On the other hand, the signature of diffusion, which is a growing average spatial displacement from a region of high concentration remains the same in all cases.
Therefore, in the first part, I give a brief overview of the expected connection between diffusivity and other material properties for each of the three material types under study and introduce relevant relations to other quantities, such as mean free path, electrical conductivity (via the Wiedemann-Franz law), and electrical mobility (via the Einstein relation). I also introduce the two-temperature model which describes the dynamics of two thermodynamic subsystems, such as electrons and phonons, as they exchange energy and diffuse. This leads me, in the second part, to consider a mathematical description of the general case of diffusion. I derive the spatio-temporal equations and present their solutions relevant for an experimental way of tracking and quantifying the diffusivity, which is expected to hold, in first approximation, independent of the underlying diffusion mechanism.

2.1 Heat transport in solids

Heat transport is a classical problem of physics, dating back to the beginning of the nineteenth century and the foundations of thermodynamics. The classical laws describing heat transport are the Fourier law of 1822, and Fick's diffusion laws of 1855.
Fourier's law states that the heat flux $\mathbf{q}$ is proportional to the temperature gradient ∇T, and that the heat always flows from the warmer to the colder region*. The proportionality constant is known as the thermal conductivity κ, which is a scalar for isotropic materials. This leads to the following relation, known as Fourier's law of heat conduction[2],

$$\mathbf{q} = -\kappa \nabla T. \tag{2.1}$$

With the advent of thermodynamics it was realized that due to conservation of energy, the internal energy per unit volume U of a system changes in time according to a continuity equation as the sum of the divergence of the heat flux vector and any energy exchanges s_i (per unit volume and time), acting as sources or drains of the internal energy[10],

$$\frac{dU}{dt} = -\nabla \cdot \mathbf{q} + \sum_i s_i. \tag{2.2}$$

Furthermore, the definition of the volumetric heat capacity C describes the change of the internal energy per unit volume to the system's temperature T,

$$C = \frac{\partial U}{\partial T}. \tag{2.3}$$

By combining equations 2.1, 2.2, and 2.3 one arrives at the following equation, describing the spatio-temporal evolution of a temperature profile $T(\mathbf{x}, t)$,

$$C\frac{\partial T}{\partial t} = \nabla(\kappa \nabla T) + \sum_i s_i. \tag{2.4}$$

In the absence of any energy exchange or external heating, i.e., $\sum_i s_i = 0$, this reduces to the following partial differential equation for the evolution of the temperature T,

$$C\frac{\partial T}{\partial t} = \nabla(\kappa \nabla T), \tag{2.5}$$

which has the form of a diffusion equation.

2.1.1 Thermal diffusion

A spatio-temporal quantity u is said to diffuse, when it follows the following differential equation, known as the diffusion equation[23], $\partial u/\partial t = \nabla(D\nabla u)$, where D is called the diffusion coefficient or diffusivity. Comparison with equation 2.5 shows that for heat conduction in a solid the temperature profile $T(\mathbf{x}, t)$ follows such a diffusion law,

*Here, I focus on the internal heat transport within a solid via thermal conductivity. Therefore, other modes of heat transport, i.e., convection, which only occurs in liquids, or thermal radiation, are not considered.

where the diffusion coefficient, called thermal diffusivity D_{th}, is given by the ratio of the thermal conductivity κ, measured in W/mK, and volumetric heat capacity C, in J/m^{3}K, as[24]

$$D_{\mathrm{th}} = \frac{\kappa}{C}. \tag{2.6}$$

This macroscopic description of thermal conduction does not rely on any microscopic assumption of the heat carriers and can be seen as a phenomenological material property. The thermal diffusivities of many materials have been measured and tabulated over the years. For example, the measured value for gold, $D_{\mathrm{th}} \approx 1.3\,\mathrm{cm}^2/\mathrm{s}$, represents a typical thermal diffusivity for noble metals[25].

2.1.2 Mean free path

A first approach to establish a connection between the macroscopic quantity of diffusivity and the microscopic random motion of scattering particles comes from the kinetic theory of gases and the related description of Brownian motion. Under the assumption of a single species of particles that move with a mean velocity v and scatter, on average, after a mean free time τ_{scat}, and therefore after a mean free path $\ell = \tau_{\mathrm{scat}} \cdot v$, the diffusion coefficient can then be estimated as

$$D \approx \frac{1}{2}\frac{\ell^2}{\tau_{\mathrm{scat}}} = \frac{1}{2}\ell v. \tag{2.7}$$

This simple description is very useful to explain many manifestations of thermal transport, as one type of heat carrier usually dominates the thermal conductivity. However, the situation becomes more complex when one considers both electrons and phonons, which are not necessarily in thermal equilibrium.

2.1.3 Two-temperature model

Today, it is well known that heat in solids is mainly carried by two types of (quasi-) particles, namely phonons and electrons[9]. Over many decades, physicists have uncovered the ways in which phonons, i.e., collective lattice vibrations, as well as electrons, e.g. free electrons in electrical conductors, contribute to thermal conduction in different materials[†].

It has been proposed, that to describe both types of heat transport, one should consider the total thermal conductivity as being composed of an electronic part κ_e and a lattice (or phononic) part κ_l[11],

$$\kappa = \kappa_e + \kappa_l. \tag{2.8}$$

If one imagines these two subsystems as a mixture of two gases flowing through the crystal, one may assign both thermodynamic subsystems their own internal energy U_α and temperature T_α, where $\alpha =$ "e" or "l" stands for electron or lattice, respectively. Then, the corresponding volumetric heat capacities C_α can be defined via

$$C_e = \frac{\partial U_e}{\partial T_e} \quad \text{and} \quad C_l = \frac{\partial U_l}{\partial T_l}. \tag{2.9}$$

[†]Additionally, heat may also be transported by other quasi-particles, such excitons in semiconductors[26].

Finally, each subsystem's internal energy per unit volume should evolve according to a continuity equation analogous to equation 2.2. Considering no external heat sources or drains, the only energy exchange terms (s_i) will come from the coupling between the electron and phonon systems. This energy exchange is assumed to scale with the difference of the two temperatures[27] and its strength is quantified by an electron-phonon coupling parameter G (measured in $\mathrm{W/m^3K}$). For now, thermal exchange with the environment is excluded. This leads to the following set of semiclassical coupled differential equations, known as the two-temperature model[28],

$$\begin{aligned} C_e \frac{\partial T_e}{\partial t} &= \nabla(\kappa_e \nabla T_e) - G \cdot (T_e - T_l), \\ C_l \frac{\partial T_l}{\partial t} &= \nabla(\kappa_l \nabla T_l) + G \cdot (T_e - T_l). \end{aligned} \tag{2.10}$$

From here, a few interesting limiting cases may be studied. First of all, in case of a thermal equilibrium between electrons and lattice, one should recover the classical heat equation. By adding the two equations and then setting $T \equiv T_e = T_l$, the following equation is recovered,

$$(C_e + C_l)\frac{\partial T}{\partial t} = \nabla((\kappa_e + \kappa_l)\nabla T). \tag{2.11}$$

This shows that at local equilibrium, the diffusivity is actually given by $D = (\kappa_e + \kappa_l)/(C_e + C_l)$. Calculations of the electronic heat capacity show that it is typically a factor 100 smaller than the lattice heat capacity, and can therefore be neglected at equilibrium[9]. Which part of the thermal conductivity dominates heat transport depends on the type of material, as discussed before. For metals κ_e dominates, while for insulators the thermal conductivity is typically dominated by κ_l[29].
Note that additional coupling of the solid to the substrate may be substantial, especially for thin film geometries, leading to an effective three-temperature model. In case the heating comes from an ultrafast laser pulse, a source term is typically added to the first equation of 2.10, as the photon energy at optical frequencies is typically first transfered only to the electron system when the laser pulse duration is shorter than the typical electron-phonon coupling time[30]. It should also be noted here that both the electron and phonon populations are approximated here as a simple thermal populations, not considering more complex density of states models for these carriers, such as subdivisions into longitudinal and transverse optical and acoustic branches for phonons[9,31], or non-thermal distributions for electrons at short times after femtosecond excitation[30,32,33]. Therefore, the coefficients such as the electron-phonon coupling parameter are to be seen as an empirical parameter rather than a coupling parameter that can be derived directly from microscopic theory[31]. However, the success of the two-temperature model in describing ultrafast relaxation processes and its relative simplicity make it a useful model to describe hot-electron dynamics without the need for additional fitting parameters and is therefore considered here as starting point to describe lateral heat flow in gold films on the sub-picosecond and nanometer scale.

2.1.4 Hot-electron diffusion in metals

For metals, the classical thermal diffusion (Eq. 2.6) governs heat transport whenever the electrons and lattice are in thermal equilibrium, which can be observed on

timescales from tens of picoseconds to seconds. Yet, the study of ultrafast light-matter interaction has shown that there is a competing process within the first picoseconds after absorption by ultrafast laser pulses[30]. Here, the thermalization is typically assumed to happen in three steps. First, as photons are absorbed, they transfer their energy, on the order of 1 eV for visible or near-infrared (NIR) light, to conduction electrons, bringing them into a non-thermal state. These excited electrons scatter with each other and establish a high-temperature Fermi-Dirac distribution among the electron system within tens to hundreds of femtoseconds, while the lattice system stays at ambient temperature[‡]. The term "hot electrons" has been used ambiguously in literature to describe either the strictly non-thermal electrons before this elevated electron temperature has been reached, or the quasi-thermalized electrons with temperatures still much higher than the lattice. In this thesis, I use the latter definition.
The reason that the electrons are heated up to much higher temperatures than the lattice is that the electron heat capacity is about a factor 100 smaller than the total heat capacity. It has been predicted, for example by Schoenlein *et al.*[21], that these "hot" electrons can diffuse at a faster rate, governed by a diffusion coefficient which is not limited by the total heat capacity C, but rather only by the electron subsystem's heat capacity C_e, i.e.,

$$D_{\mathrm{hot-e}} \approx \frac{\kappa_e}{C_e}. \tag{2.12}$$

This can be seen as the limiting case of the two-temperature model (Eqs. 2.10) for either the uncoupled case ($G = 0$) or, more realistically, the case of $T_e \gg T_l$. Considering the factor 100 smaller heat capacity this leads to the prediction, $D_{\mathrm{hot\text{-}e}} \approx 100 \cdot D_{\mathrm{th}}$, although for $T_e = 1000 - 1500\,\mathrm{K}$ and $T_l \approx 300\,\mathrm{K}$ the ratio T_e/T_l is only $3-5$.
Chapter 5 describes, for the first time, a measurement of the transition from such fast hot-electron diffusion to thermal diffusion via spatio-temporal transient-absorption microscopy and discusses in more detail the underlying dynamics and the validity of these limits based on a two-temperature model similar to the one presented above (Eqs. 2.10).

2.1.5 Wiedemann-Franz law

For electrically conducting materials, such as metals, there is an direct relationship between the electronic part of the thermal conductivity κ_e and the electrical conductivity σ, as both are transport phenomena carried by the free (conduction band) electrons. This is described by the Wiedemann-Franz law[9],

$$\frac{\kappa_e}{\sigma} = \mathcal{L}T, \tag{2.13}$$

where T is the temperature and the proportionality constant $\mathcal{L}$ is the Lorentz number, which can be calculated to be given by $\mathcal{L} = \pi^2/3 \cdot (k_B/e)^2$, where k_B is the Boltzmann constant and e the elementary charge.

[‡]The electron-electron thermalization time may become comparable to the electron-phonon relaxation time, causing deviations from the two-temperature model, especially for low excitation energy and cryogenic lattice temperatures[30,32,34].

2.1.6 Carrier diffusion in a semiconductor

Semiconductors are materials with an optical band gap and a conductivity between that of a metal and an insulator, where charges can be carried by electrons and holes. In well-known semiconductors such as silicon, the conduction properties can be controlled via impurity doping. By implanting elements with more electrons (called n-doping) an excess of free electrons leads to an enhanced electrical conductivity. If an element with fewer electrons is implanted, the resulting p-doped silicon also conducts better due to an excess of holes. In the latter case, the holes would be called the majority carriers, while the remaining free electrons are the minority carriers (and vice versa for n-doping). Optical excitation with photon energies above the band gap creates an electron-hole pair, i.e., a neutral exciton. Depending on the doping level, the neutral exciton can diffuse before recombination and can even be separated and controlled with applied voltages and geometries, as in p-n junctions.
The conductivity for the general case of electron and hole conduction reads

$$\sigma = e(n\mu_e + p\mu_p), \tag{2.14}$$

where e is the elementary charge, n and p are the electron and hole carrier density, respectively, and μ_e and μ_p are the electron and hole mobility, respectively. For a strongly doped semiconductor, the dominant contribution to the conductivity typically comes from the majority carriers. Yet, the generation, diffusion and recombination of minority carriers is also important in many applications, such as solar cells, where charges need to be separated.
For both types of carriers, the diffusivity D_α ($\alpha =$ "n" or "p") is related to its corresponding electrical mobility μ_α via the so-called Einstein relation[35],

$$D_\alpha = \mu_\alpha k_B T/e. \tag{2.15}$$

For carriers which decay with a lifetime τ, such as minority carriers in a semiconductor due to recombination, an average diffusion length L can be defined as

$$L = \sqrt{D\tau}. \tag{2.16}$$

Well-known semiconductors such as silicon have been very well characterized in terms of electrical properties, including drift-diffusion models, due to their importance in transistor technology[36]. For example, the motion of electrons in heterostructure field effect transistors (FETs) are modeled with the two-dimensional electron gas (2DEG) approximation[37]. Yet, the direct spatio-temporal imaging of carrier diffusion after photo-excitation for the first semiconductor has only been reported recently for a GaAs bulk sample[38].
Another interesting case is that of the novel class of two-dimensional transition metal dichalcogenides (TMDs). These materials, such as molybdenum disulfide ($MoSe_2$), have been described using basically the same concepts known from 3D semiconductors, and first attempts have recently been made to study their carrier diffusion with transient absorption microscopy[39–41].

2.1.7 (Hot) Carrier diffusion in graphene

Graphene is a novel two-dimensional material with a unique dispersion relation and high reported carrier mobilities. In contrast to both metals and semiconductors,

graphene shows a linear dispersion relation, such that the electrons and holes can be described as massless Dirac fermions[42]. The dispersion relation for the energy E as a function of momentum k reads

$$E_{\pm}(k) = \pm\hbar v_F k, \tag{2.17}$$

where $v_F \approx 10^6$ m/s is the Fermi velocity.
It is also possible to change the carrier density of graphene by electrostatic gating[42]. This directly effects the Fermi level E_F, as it scales with the carrier density n as

$$n = \frac{E_F^2}{\pi\hbar^2 v_F^2}. \tag{2.18}$$

It should be noted that in graphene, quantities like carrier density are 2D analogues of 3D concepts. For example, n is a surface carrier density, measured in cm^{-2}, compared to volumetric density (in cm^{-3}) for a conventional solid.
An important link between diffusivity D and carrier mobility μ is the Einstein relation, as introduced above for semiconductors in equation 2.15. There, the scaling with temperature was a specific result valid only for semiconductors. The more general formulation of the Einstein relation reads[36]

$$D = \frac{\mu n}{e\frac{dn}{d\phi}}, \tag{2.19}$$

which depends on the carrier density n and its derivative with respect to the chemical potential ϕ. For doped graphene, the derivative can be taken from equation 2.18 with respect to the Fermi-level,

$$\frac{dn}{d\phi} \approx \frac{dn}{dE_F} = \frac{2n}{E_F}, \tag{2.20}$$

where the Fermi velocity is assumed to be independent of E_F, a generally well-accepted assumption for monolayer graphene[42].
This leads to the following relation between diffusivity and Fermi energy, or alternatively carrier density,

$$D = \frac{E_F\mu}{2e} = \frac{\hbar v_F\sqrt{\pi n}\mu}{2e}. \tag{2.21}$$

This simple relation has, to my knowledge, been discussed only very few times in the literature[43,44]. The same result can be obtained by calculating D from the ratio of thermal conductivity κ and the electronic heat capacity C_e (as in Eq. 2.12), by using the Wiedemann-Franz law (Eq. 2.13),

$$\kappa = \frac{\pi^2 k_B^2 T_e \sigma}{3e}, \tag{2.22}$$

with the conductivity $\sigma = ne\mu$ and the following heat capacity for graphene,

$$C_e = \frac{2\pi E_F k_B^2 T_e}{3\hbar^2 v_F^2}. \tag{2.23}$$

Note that the diffusion coefficient has the same units as in 3D (cm^2/s), while κ, C_e, σ, and n are each reduced by one length dimension, as they are normalized to area

instead of volume for graphene.
The above relations are assumed to hold in the Sommerfeld expansion regime, when $T_e < T_F = E_F/k_B$[35]. Therefore, it remains an open question whether the diffusion of hot carriers (with temperatures on the order of $T_e = 10^3\,\mathrm{K}$) should be modified as $k_B T_e$ approaches E_F.

2.2 Diffusion equation: Mathematical treatment

As introduced at the beginning of section 2.1, the most general equation describing diffusion is known as the diffusion equation. Consider a general spatio-temporal profile $u(\mathbf{x}, t)$, with $\mathbf{x} \in \mathbb{R}^n$ ($n = 1, 2$, or 3), which could be a concentration or temperature profile, evolving in an isotropic medium according to the diffusion equation[23],

$$\frac{\partial u}{\partial t} = \nabla(D\nabla u). \tag{2.24}$$

In general, D can depend on space $\mathbf{x}$, time t, and concentration u.
In the following sections, I give a mathematical summary of a selection of the most useful solutions to slightly different diffusion problems with relevance to this thesis, starting from equation 2.24. The dynamics are always assumed to start at time $t = 0$, the initial conditions will be defined as $u(\mathbf{x}, 0)$ and solutions will describe the spatio-temporal evolution for $t \geq 0$.

2.2.1 Solution to the heat equation

The simplest case of a constant D leads to the so-called heat equation,

$$\frac{\partial u}{\partial t} = D\nabla^2 u. \tag{2.25}$$

The fundamental solution to equation 2.25, defined as the solution to an initial Dirac delta function $u(\mathbf{x}, 0) = \delta(\mathbf{x})$, is given by[23]

$$\Phi(\mathbf{x}, t) = \frac{1}{\sqrt{(4\pi D t)^n}} \exp\left(-\frac{|\mathbf{x}|^2}{4Dt}\right), \tag{2.26}$$

where n is the number of spatial dimensions. This happens to be an n-dimensional normalized Gaussian function with standard width $\sigma_D = \sqrt{2Dt}$.
In the scope of this thesis I am particularly interested in solutions to problems with a Gaussian profile, with an initial width σ_0, as an initial condition,

$$g(\mathbf{x}) \coloneqq u(\mathbf{x}, 0) = A^n \exp\left(\frac{-|\mathbf{x}|^2}{2\sigma_0^2}\right). \tag{2.27}$$

Experimentally, this comes from the shape of a laser (Gaussian beam), being focused by an objective to a spot with a Gaussian-shaped spatial profile.
The boundary condition far away is then set to $u(\pm\infty, t) = 0$, as the Gaussian function drops to zero exponentially and the information from the initial perturbation can only reach out as fast as the diffusion propagates. Experimentally, this corresponds to

a homogeneous sample over the interaction region. If, e.g., u describes temperature, the ambient temperature outside the initial Gaussian acts as an offset and does not change the dynamics. Therefore, it can safely be subtracted.
The solution to this problem can now be calculated by convolution with the fundamental solution (Eq. 2.26),

$$\begin{aligned} u(\mathbf{x},t) &= \int_{\mathbb{R}^n} \Phi(\mathbf{x}-\mathbf{y},t)g(\mathbf{y})d\mathbf{y} \\ &= \frac{A^n}{\sqrt{1+\frac{2Dt}{\sigma_0^2}}^n} \exp\left(\frac{-|\mathbf{x}|^2}{2(\sigma_0^2+2Dt)}\right). \end{aligned} \tag{2.28}$$

This solution is again a Gaussian profile with a time dependent amplitude and width. The amplitude decays with a typical $t^{-\frac{n}{2}}$-law, while the time dependence of the width reads

$$\sigma(t) = \sqrt{\sigma_0^2 + 2Dt}. \tag{2.29}$$

From this, it becomes clear that diffusion is characterized by a mean square displacement (MSD) $\langle x^2 \rangle$ ($= \sigma^2$ for the Gaussian distribution) that grows linear in time[§].
If one measures the full-width at half-maximum (FWHM) instead of σ, the relation $\mathrm{FWHM} = 2\sqrt{2\ln 2}\,\sigma$ for Gaussian functions can be used to convert equation 2.29 to

$$\mathrm{FWHM}^2(t) = \mathrm{FWHM}_0^2 + 16(\ln 2)Dt. \tag{2.30}$$

In particular, the derivative of FWHM^2 with respect to time is directly proportional to diffusivity D, as

$$\frac{\partial \mathrm{FWHM}^2(t)}{\partial t} = 16(\ln 2)D. \tag{2.31}$$

This result indicates how for a diffusion problem, the diffusivity D can be directly observed by tracking the temporal width evolution and extracting the slope of the graph of the squared width.

2.2.2 Heat equation with exponential decay

If the diffusing quantity also decays exponentially in time, with time-constant τ, equation 2.25 becomes

$$\frac{\partial u}{\partial t} = D\nabla^2 u - \frac{1}{\tau}u. \tag{2.32}$$

Then, the fundamental solution reads

$$\Phi(\mathbf{x},t) = \frac{1}{\sqrt{(4\pi Dt)^n}} \exp\left(-\frac{|\mathbf{x}|^2}{4Dt} - \frac{t}{\tau}\right), \tag{2.33}$$

[§]Note that for n-dimensional Gaussians the mean square displacement, as it is defined as a sum over the squares for each dimensions grows with a factor n. Yet, if one measures 1D cut-lines or summed profiles across 2- or 3-dimensional Gaussian profiles, the width will always be measured as equation 2.29, independent of n. This has led to incorrect calculations of diffusivities (by a factor 2) in literature[40,41,45,46]. Other publications by the same (and other) authors have corrected this mistake.

and the solution for a Gaussian initial distribution becomes

$$u(\mathbf{x},t) = \frac{A^n}{\sqrt{1+\frac{2Dt}{\sigma_0^2}}^n} \exp\left(\frac{-|\mathbf{x}|^2}{2(\sigma_0^2+2Dt)} - \frac{t}{\tau}\right). \tag{2.34}$$

This is just the solution of before (Eq. 2.28) with an additional exponential decay only in time (on top of the standard $t^{-\frac{n}{2}}$ decay typical for diffusion). Therefore, the evolution of the Gaussian *width* will not be effected by the decay and evolve in time, as before, according to equation 2.30.
If one includes additional nonlinear terms, for example an Auger recombination term $(-\gamma u^2)$ to equation 2.32, relevant for exciton populations at high concentrations[47], the solution will be more complex and a deviation of the linear squared width evolution will be observed.

2.2.3 Time-dependent diffusivity

For the heat equations considered above, I have assumed a constant diffusion coefficient. Yet, if one includes an explicit time-dependence of the diffusivity, $D(t)$ into the diffusion model, equation 2.24 becomes

$$\frac{\partial u}{\partial t} = D(t)\nabla^2 u. \tag{2.35}$$

Following the treatment of J. Crank[23], one can introduce a transformation to a new time variable s, via $ds = D(t)dt$, and equation 2.35 becomes

$$\frac{\partial u}{\partial s} = \nabla^2 u. \tag{2.36}$$

Now, the heat equation can be solved as above and the time variable can be replaced in the solution. In particular, to measure the diffusivity via the evolution of the slope of FWHM^2, an analogous result to equation 2.31 is established: $\partial\mathrm{FWHM}^2/\partial s = 16\ln 2$. By re-substituting ds with dt one finally arrives at

$$\frac{\partial \mathrm{FWHM}^2(t)}{\partial t} = 16(\ln 2)D(t). \tag{2.37}$$

Therefore, even for a time dependent diffusivity, the slope of the graph of the squared width $\mathrm{FWHM}^2(t)$ gives a direct measure of this diffusivity $D(t)$.
In summary, in the three exemplary cases of a pure heat equation with constant D, a heat equation with an additional exponential decay (and constant D), as well as a heat equation with a time dependent $D(t)$, the slope of the graph of the squared width $\mathrm{FWHM}^2(t)$, divided by the constant factor $16\ln(2) \approx 11$, gives a direct measure of the underlying diffusivity.
The above calculations are valid for $n = 1, 2, 3$ spatial dimensions. In the following sections I focus on 2D dynamics, if not explicitly stated otherwise. This should be thought of as either samples which are 2D in nature, or thin samples, where carrier or heat diffusion is measured in the lateral dimensions (x,y), while the z-dynamics are either negligible, being averaged over, or only indirectly probed due to reflection geometry.

3

Transient absorption microscopy: Theoretical considerations

In this chapter I discuss basic concepts and prerequisites to study carrier and heat diffusion in solids with transient-absorption microscopy.
I briefly introduce the fundamentals of the technique of transient absorption known from time-resolved spectroscopy, as well as its combination with microscopy and applicability to studying the dynamics of solids. Further, I describe details and limitations of an experimental scheme to measure such diffusion directly via transient-absorption microscopy, describing mathematically the expected spatial profile of the transient absorption signal for different scanning modes. In the last part, I show various simple simulations of diffusing populations. I analyze their width evolution depending on complications such as a spatial nonuniform offset and the effect of multiple species, showing the limitations of width monitoring to extract the correct underlying diffusion coefficients.

3.1 Transient absorption microscopy

Here, I introduce transient absorption (TA) as a general technique and describe its use for solid state physics. Then, I introduce the more recent development of combining this technique with microscopy.

3.1.1 Transient absorption: From molecules to solids

Since the advent of ultrafast pulsed lasers, it has become feasible to study processes on the femtosecond to nanosecond time-scales[48]. The technique known as transient-absorption spectroscopy has evolved rapidly over the last 30 years. The basic idea is the following: A short laser pulse, the *pump*-pulse, illuminates the sample under study. The absorbed light leads to excitation and/or heating of the material. After a controlled time delay, a second *probe*-pulse interrogates the sample.
In such experiments, *differential absorption* ΔA is measured in practice by recording the difference in transmission ΔT (or reflection ΔR) with and without pump illumination. Light-induced changes in the sample can be monitored as a function of pump-probe time delay t. After scaling to the unperturbed transmission T (or reflection R) of the probe beam, these read

$$\frac{\Delta T}{T}(t) = \frac{T_{\mathrm{pumpOn}}(t) - T_{\mathrm{pumpOff}}}{T_{\mathrm{pumpOff}}}, \tag{3.1}$$

and

$$\frac{\Delta R}{R}(t) = \frac{R_{\mathrm{pumpOn}}(t) - R_{\mathrm{pumpOff}}}{R_{\mathrm{pumpOff}}}. \tag{3.2}$$

This technique is being used to study excited state dynamics in chemistry, biology, and solid-state physics[48–50]. Typically, the laser beams are weakly focused either through a vial or onto a solid target. By measuring in the time and wavelength domain, useful information about the excited state decay pathways can be deduced. These include many different photo-physical processes from molecules, such as vibrational relaxation, inter-system crossing, or internal conversion, all the way to charge carrier dynamics in semiconductors, or optically induced phase transitions in solids.
In the rest of the chapter, I will usually think of the transient absorption as being measured in transmission (ΔT), although equivalent results hold for reflection (ΔR). A simple picture for the three basic processes which can lead to a transient-absorption signal can be understood in a few-level system, as shown in figure 3.1a.
As this technique probes changes to the absorption (reflection or transmission) with and without the pump beam, it is sensitive to the dynamics of the excited states of the system. In this example, the pump pulse causes an excitation from the ground state $|g\rangle$ to an excited state $|e_2\rangle$, possibly followed by a relaxation to a lower lying excited state $|e_1\rangle$. The populated excited state (and de-populated ground state) can modulate the incoming probe light, in this example at lower photon energy than the pump, via three mechanisms.
The first process, stimulated emission (SE), can force the excited state population back down to the ground state. Therefore, more light at the probe beam wavelength is produced when the pump beam is on and, in a transmission experiment, a positive

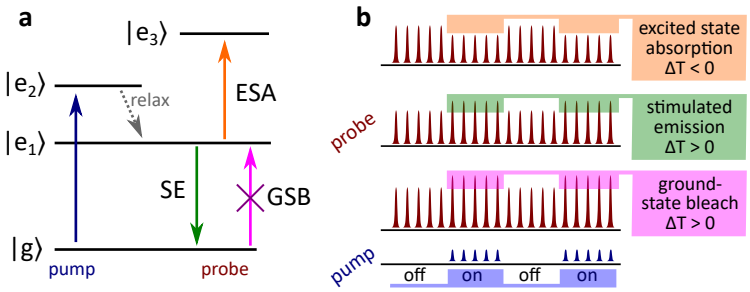

Fig. 3.1: Fundamental processes of transient absorption. (**a**) Four level diagram illustrating the three basic transitions leading to transient absorption, namely excited state absorption (ESA), stimulated emission (SE) and ground state bleach (GSB). (**b**) Temporal representation of the measurement scheme. The probe pulse trains (red) are modulated due to the presence of the pump light (blue). ESA leads to decreased transmission, while the other two processes lead to an increase in transmission. These small modulated differences can be isolated from the background with high sensitivity via lock-in detection.

transient transmission $\Delta T > 0$ will be recorded*.
Secondly, excited state absorption (ESA) of the probe light to a higher lying state $|e_3\rangle$, may cause a reduction of the probe light transmission, i.e. $\Delta T < 0$.
Finally, the pump's effect of depopulating the ground state may also cause a decreased absorption of the probe. This process is called ground state bleach (GSB) and also causes a positive ΔT. Figure 3.1b depicts these changes in probe beam signals for a modulated pump beam illumination. To detect these small signal differences, they can be isolated by de-modulation with a lock-in amplifier.
This technique of transient absorption can be understood more deeply in a general framework of third order nonlinear optics. A rigorous description of nonlinear spectroscopy calculates the differential absorption ΔT as the time-integral of the third order polarization $P^{(3)}(t';t)$ and the electric field of the probe beam $E_{\mathrm{pr}}(t)$[52,53],

$$\frac{\Delta T}{T}(t) \propto \frac{\int_0^\infty dt' E_{\mathrm{pr}}(t') \cdot P^{(3)}(t';t)}{\int_0^\infty dt' |E_{\mathrm{pr}}(t')|^2}. \tag{3.3}$$

This may be thought of as a heterodyne detection of the radiating third order polarization $P^{(3)}$, created by the interaction of the sample with the three electric fields of the pump (twice) and probe laser (E_{pu}, E_{pu}, E_{pr}), with the electric field of the probe beam E_{pr} at the detector. Under the assumptions of the rotating-wave approximation, strict time ordering, and when the pump and probe pulse widths $\tau_{\mathrm{pu/pr}}$ are short compared to the pump-probe time delay t as well as the de-phasing time of the excited state, the problem simplifies considerably. Then one arrives at the picture of the two pump fields causing an excitation event leading to an excited state population $p_1 \propto E_{\mathrm{pu}} \cdot E_{\mathrm{pu}}$. Then, the third (probe) field interaction after the delay t leads to a third order sample response $P^{(3)} \propto E_{\mathrm{pr}} \cdot E_{\mathrm{pu}}^2 \cdot S^{(3)}$, scaling with the involved transition dipoles and the sign depending on the type of interaction (ESA, SE, or GSB). Here,

*In case of a reflection experiment, the sign of the transient reflection ΔR depends on the sample/substrate reflection geometry[51].

effects such as a decaying population has been separated into the sample response $S^{(3)}$. Finally, including spatial dependence (x), the total signal scales as

$$\frac{\Delta T}{T}(x,t) \propto \frac{E_{\mathrm{pu}}^2(x) \cdot E_{\mathrm{pr}}^2(x) \cdot S^{(3)}(x,t)}{|E_{\mathrm{pr}}(x)|^2} = \frac{I_{\mathrm{pu}}(x) \cdot I_{\mathrm{pr}}(x) \cdot S^{(3)}(x,t)}{I_{\mathrm{pr}}(x)}, \tag{3.4}$$

where $I_\alpha = E_\alpha^2$, with $\alpha =$ "pu" or "pr", is the intensity of the pump or probe light, respectively.
The main result here is that the transient absorption signal scales with the intensity of the pump laser, and is (after normalization) independent of the intensity of the probe laser for each point in space. Yet, the spatial shape of both pump and probe beams are important when calculating the point spread function (PSF) for scanning TA microscopy, as described in more detail in section 3.1.4.
When both pulses overlap in time, i.e., when the pump-probe delay is shorter than the pulse lengths, many different time orderings need to be considered for the four-field interaction and strong nonlinear signals may be created, known as "coherent artifacts"[53].
The picture presented here for pure electronic transitions in a four-level system is useful to understand the principles of TA spectroscopy, especially for the study of molecular systems. Yet, the situation becomes more complex when considering solids. In semiconductors, exciton dynamics are typically explained with band diagrams, while off-resonant transient heating can also be studied in other materials, where the changes to reflection and transmission are thought of in terms of thermally induced changes to the permittivity. Recently, a few different pump-probe techniques have emerged to study thermal and carrier diffusion properties. These include time-domain and frequency-domain thermoreflectance (TDTR and FDTR, respectively)[54–57] and transient thermal grating techniques[58–62].

3.1.2 Combining transient absorption with microscopy

The described powerful technique of time-resolved spectroscopy typically studies macroscopic samples, as weakly focused laser beams pass through liquids in vials, or reflect off large areas of solid samples. This allows to extract temporal dynamics, yet with no correlation to microscopic features. On the other hand, traditional microscopy is used to study features of samples on the sub-micrometer length scale. To be able to obtain simultaneous spatial and temporal information from sub-micron sized features on the femtosecond time scale, it is desirable to combine ultrafast techniques with microscopy.
Imaging fluorescence with time-correlated single photon counting (TCSPC) has recently been employed to visualize sub-nanosecond to nanosecond energy transfer dynamics in molecular solids and quantum dot solids[63,64]. Yet, here the time-resolution is limited by the timing electronics, on the order of 50 ps.
To study spatio-temporal processes on the nanometer and femtosecond scales, microscopy must be combined with ultrafast pump-probe techniques. Recently, transient-absorption microscopy (TAM) has emerged as a powerful technique to study carrier diffusion in semiconductors, molecular solids, and 2D materials[20,38,45,65–68].
The complications that arise from this approach are due to the strong focusing with high numerical aperture (NA) objectives, which necessitates precise alignment and

leads to less signal-to-noise at a given fluence. Additionally, the high-NA objective leads to aberrations and dispersion of the ultrafast laser pulses. These trade-offs will be further discussed in chapter 4. For now, I will focus on a description of how to track a nanometer-femtosecond diffusion process in space and time with transient absorption microscopy.

3.1.3 Tracking diffusivity experimentally

The key results from chapter 2.2, in particular equations 2.30 and 2.37, suggest a simple way of experimentally tracking diffusivity, by simply monitoring the (squared) width of the spatial profile over time. I have shown mathematically, that this approach is valid for a single diffusing profile following a heat equation, such as equation 2.25 or 2.32, with or without exponential decay, and even for time-dependent diffusivities. Any *linear* optical response of a homogeneous sample to a pulsed beam excitation with a Gaussian spatial shape will result in an initial Gaussian profile $u(\mathbf{x}, 0)$. It should then be possible to track the evolution of this profile in space and time with an experimental observable. In case this observable is proportional to the profile, the diffusivity over time may be measured in a very direct way, by simply fitting Gaussian functions to the measured spatial profiles. The spatial resolution of this technique can be defined in multiple ways. As in any far field technique, the spot sizes of the individual laser beams are diffraction limited. However, in analogy to super-resolution microscopy[69], the ability to observe small changes in the width is only limited by the precision of the Gaussian fit, therefore depending on the ratio of signal to background and noise. For typical signal levels reported in this thesis at a few seconds integration times, I have obtained a spatial precision on the order of 20 nm[70].

3.1.4 Gaussian point spread function in nonlinear microscopy

To combine the nonlinear optics of transient absorption spectroscopy with advanced microscopy, I model the expected spatial response functions for different scanning modes, using explicit spatial dependencies of the intensities. As seen in equation 3.4, transient absorption scales as the product of the pump intensity $I_{\mathrm{pu}}(x)$ and the sample third order susceptibility $S^{(3)}(x)$. In the following treatment, I consider different scanning modes and calculate the expected spatial responses as a function of the spatial coordinate x, after pixel-by-pixel scanning on a point detector. During the scanning, the total transmitted (or reflected) light is always fully collected at each pixel. Therefore, the x-dependence of the scanning signal will depend on the three spatial profiles ($I_{\mathrm{pu}}(x)$, $I_{\mathrm{pr}}(x)$ and $S^{(3)}(x)$), while the denominator of equation 3.4 will be independent of the scanning coordinate. Also, for now, I will assume the sample response as independent of carrier diffusion. This can be thought of as the initial ($t \approx 0$) situation. Also, for simplicity, all the spatial dependencies will be called x. However, they are valid for the other dimension, y, as well, as there is complete separability between the spatial dimensions.
Both the pump and probe intensities are modeled as Gaussian functions in space,

$$I_i(x) = G(x_i, \sigma_i) \coloneqq A_i \exp\left(-\frac{(x - x_i)^2}{2\sigma_i^2}\right), \tag{3.5}$$

with i = "pu" or "pr" for the pump or probe focal distribution, respectively. The spatial shape of the response $\Delta T(x)$ depends on the scanning mode and the interaction region. In the next two sections, I discuss the two relevant scanning modes for this thesis.

Collinear pump and probe, scanning sample, point detector

Here I describe a type of pump-probe microscopy where both beams are tightly focused to the same volume, while the sample is scanned through this focus region. The image is formed by collecting the probe light on a point detector as a function of the sample position. In this thesis, I call this technique sample-scanning transient-absorption microscopy (SS-TAM).
For collinear pump and probe beams $x_{\mathrm{pu}} = x_{\mathrm{pr}}$ ($= 0$ without loss of generality), and sample scanning, the transient response will reflect the third order susceptibility $S^{(3)}(x)$ of the sample, convoluted with the product of $I_{\mathrm{pu}}(x) = |E_{\mathrm{pu}}(x)|^2$ and $I_{\mathrm{pr}}(x) = |E_{\mathrm{pr}}(x)|^2$, which can be seen as the PSF of this type of microscopy,

$$\Delta T(x) \propto (S^{(3)} * (I_{\mathrm{pu}} \cdot I_{\mathrm{pr}}))(x) = (S^{(3)} * \mathrm{PSF}_{\mathrm{TAM}})(x), \tag{3.6}$$

where "$*$" denotes spatial convolution. Here, the final x-coordinate represents the axis formed after point-by-point scanning and collecting all the transmitted (or reflected) probe light on a point detector. Wide-field TA imaging would result in a different shape. A sketch of this situation is shown in figure 3.2a.
The point spread function $\mathrm{PSF}_{\mathrm{TAM}}$ is formed, according to equation 3.3, by the product of the two Gaussian intensities,

$$\begin{aligned}
\mathrm{PSF}_{\mathrm{TAM}}(x) &= G(0, \sigma_{\mathrm{pu}}) \cdot G(0, \sigma_{\mathrm{pr}}) \\
&= A_{\mathrm{TAM}} \cdot \exp\left(-\frac{x^2}{2\sigma_{\mathrm{pu}}^2}\right) \cdot \exp\left(-\frac{x^2}{2\sigma_{\mathrm{pr}}^2}\right) \\
&= A_{\mathrm{TAM}} \cdot \exp\left[-\frac{x^2}{2} \cdot \left(\frac{1}{\sigma_{\mathrm{pu}}^2} + \frac{1}{\sigma_{\mathrm{pr}}^2}\right)\right] \\
&= G(0, \sigma_{\mathrm{TAM}}),
\end{aligned} \tag{3.7}$$

where the final width is given by the reciprocal square sum of the pump and probe focus widths,

$$\frac{1}{\sigma_{\mathrm{TAM}}^2} = \frac{1}{\sigma_{\mathrm{pu}}^2} + \frac{1}{\sigma_{\mathrm{pr}}^2}. \tag{3.8}$$

This is the expected response when scanning a small particle (size $\ll \sigma$) through the collinear beams. This can be seen as the limit of a delta peak as sample response, $S^{(3)}(x) \propto \delta(x)$ convolved with the PSF,

$$\begin{aligned}
(G(0, \sigma_{\mathrm{TAM}}) * \delta)(x) &= A_{\mathrm{TAM}} \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2\sigma_{\mathrm{TAM}}^2}\right) \cdot \delta(x - y) dy \\
&= A_{\mathrm{TAM}} \exp\left(-\frac{x^2}{2\sigma_{\mathrm{TAM}}^2}\right) \\
&= G(0, \sigma_{\mathrm{TAM}}).
\end{aligned} \tag{3.9}$$

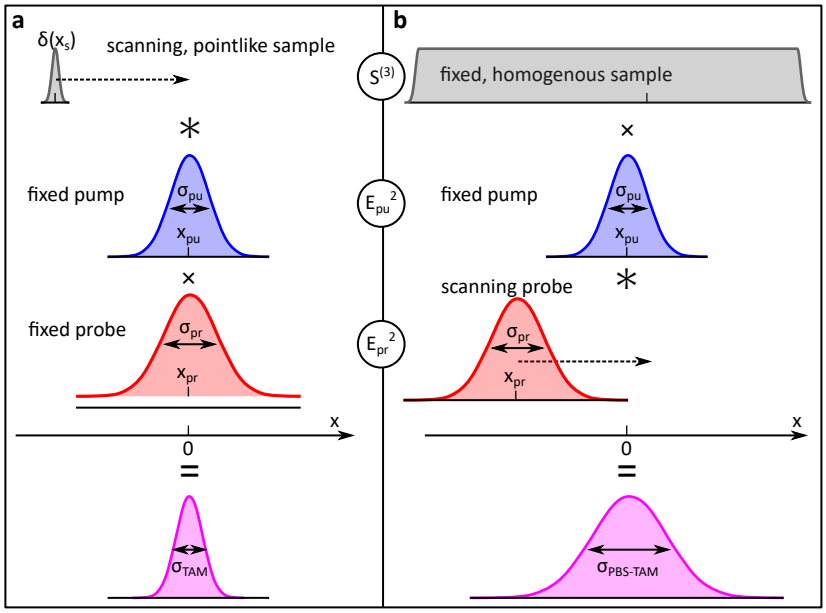

Fig. 3.2: Sketch illustrating the point spread functions for different scanning modes of transient-absorption microscopy. (**a**) The case of collinear and fixed pump and probe beams, while a point-like sample is scanned (as in Eqs. 3.6 to 3.9). After pixel-by-pixel collection on a point detector, the width of the resulting profile σ_{TAM} is narrower than the individual beam size. (**b**) The situation of a fixed pump beam, yet a scanning probe beam over a homogeneous sample (as in Eq. 3.10). Here, the resulting width after point-by-point detection $\sigma_{\mathrm{PBS\text{-}TAM}}$ is wider than the individual beam profiles. The symbol "$*$" denotes spatial convolution, while "$\times$" stands for multiplication with the same x-axis.

In conclusion, the width of the point spread function σ_{TAM} of this collinear ("sample-scanning") type of nonlinear microscopy is always smaller than the PSF of the individual beams, as shown in equation 3.8. For example, in case $\sigma_{\mathrm{pu}} = \sigma_{\mathrm{pr}}$, the transient absorption microscope will have a $\sqrt{2}$ better resolution than linear microscopy with either beam.

Fixed pump on sample, scanning probe, point detector

Consider, on the other hand, a homogeneous sample, extended far beyond the pump and probe spots ($S^{(3)}(x) = \text{const.}$), where the probe beam is scanned over the fixed pump ($x_{\mathrm{pu}} = 0$) and the transmitted (reflected) power is collected on a point detector as a function of the probe beam scanning position. In this thesis, I call this technique probe-beam-scanning transient-absorption microscopy (PBS-TAM).
This situation is depicted in figure 3.2b. Now, the total response is again evaluated according to equation 3.4, with a constant denominator as all the probe light is collected, but where the two PSFs are displaced relative to each other in space, as the

scanning is equivalent to an integral over the probe space coordinate x_{pr},

$$\begin{aligned}\Delta T(x) &\propto \int_{-\infty}^{\infty} G(0,\sigma_{\mathrm{pu}}) \cdot G(x_{\mathrm{pr}},\sigma_{\mathrm{pr}}) dx_{\mathrm{pr}} \\ &= G(0,\sigma_{\mathrm{pu}}) * G(0,\sigma_{\mathrm{pr}}) \\ &= G(0,\sqrt{\sigma_{\mathrm{pu}}^2+\sigma_{\mathrm{pr}}^2}).\end{aligned} \tag{3.10}$$

Here, in the last step, the variance for the convolution of two Gaussian functions is given as the sum of the two original variances[71],

$$\sigma^2_{\text{PBS-TAM}} = \sigma^2_{\mathrm{pu}} + \sigma^2_{\mathrm{pr}}. \tag{3.11}$$

This type of nonlinear "probe-beam-scanning" microscopy does *not* image the sample via raster-scanning with a given PSF. Instead, the experimental observable is given by a spatial convolution of the initially excited, diffusing profile with another, constant Gaussian profile (the probe pulse). By scanning a focused probe beam over the excited area in order to get spatio-temporal diffusion information, the measured profile will be, again, a Gaussian profile, as convolution of two Gaussians results in another Gaussian as shown above (see Eq. 3.10).
To observe diffusion dynamics as described in section 2.2 via PBS-TAM, the finite size of the scanning probe beam can now be included. Under the assumption of linear sample response (discussed in the next section), equation 2.30 will be only modified in terms of a bigger initial width, FWHM_0, as

$$\mathrm{FWHM}_0^2 = \mathrm{FWHM}^2_{\text{excited}} + \mathrm{FWHM}^2_{\text{probe-beam}}. \tag{3.12}$$

Importantly, the temporal evolution, i.e., the slope of $\mathrm{FWHM}^2(t)$, according to equation 2.31 (and 2.37), is still a valid measure of the sample's diffusivity.

3.1.5 Assumptions for measuring diffusion with ultrafast microscopy

Here, I list a set of basic assumptions, or requirements, assumed so far that, if fulfilled, should make PBS-TAM a robust technique to measure and correctly quantify diffusion dynamics.

- Before the arrival of the pump pulse ($t < 0$) the sample is in a steady state, i.e., the pulses are separated sufficiently so that all decay or cooling dynamics caused by one pulse are over when the next pulse arrives.
- At $t = 0$ the sample is illuminated with a beam that can be described by a 2D Gaussian profile $p(x, y) = G(\mathbf{x}, \sigma_{\mathrm{pu}})$ with a spatially homogeneous (if any) offset and relatively small, random noise.
- The sample's response after absorption, and possible redistribution dynamics within the temporal resolution, is described by the function $u_0 = u(x, y, t \approx 0)$, which is *proportional* to $p(x, y)$.
- $u(x, y, t)$ evolves in time according to a diffusion law (Eqs. 2.25, 2.32, or 2.35).

- The experimental observable $\Delta T/T(x, y, t)$ (or $\Delta R/R$) is *proportional* to the underlying profile $u(x, y, t)$, after spatial convolution with the (also Gaussian) probe beam.

- The diffusion dynamics are extracted by Gaussian fitting, or an equivalent measure of the MSD $\langle \mathbf{x}^2 \rangle$, and analyzing the graph of squared width as a function of time, according to equation 2.31.

In this and the next chapters I will explore many cases in which the above assumptions are not, or only partially, fulfilled and discuss what information can be extracted from the data, nonetheless.

3.2 Simulated diffusion: Width dynamics for selected cases

The assumptions listed in the section above are obviously not always going to be fulfilled in reality. Yet, the idea to measure and quantify diffusion dynamics even in the presence of a slight nonlinearity or a small, yet non-uniform background should not be discarded. To get an idea what to expect, I will simulate some of the simplest cases which could lead to "artifacts" in the width evolution analysis, including a Gaussian offset, as well as the case of two diffusing species, with the same or opposite sign.
Since the dynamics in x and y are completely separable, I discuss the following simulations in 1D. I have checked that analogous 2D simulations lead to exactly the same results.

3.2.1 Example 1: One diffusing species

Here, I simulate the simple case of a single diffusing quantity, including an exponential decay in time, i.e., the solution to the diffusion model of equation 2.32. I consider one diffusing species with the following, experimentally realistic parameters:

- The initial profile $p(x)$ is a Gaussian with $\mathrm{FWHM}_{\mathrm{pu}} = 0.6\,\mu\mathrm{m}$.

- The width evolves, beginning at $t = 0$, according to a diffusion law with constant diffusivity D, i.e., $\mathrm{FWHM}^2(t) = \mathrm{FWHM}^2_{\mathrm{pu}} + 16(\ln 2)Dt$.

- The temporal evolution of the amplitude is zero at $t < 0$ then peaks at $t = 0$ and decays as described by the function $A(t) = A_0 \cdot B/(B + t) \cdot \exp(-t/\tau)$, with $B = \mathrm{FWHM}^2_{\mathrm{pu}}/(4(\ln 2)D)$. Here, the first term A_0 is a constant amplitude factor, $B/(B + t)$ describes the typical "1/t" decay from a diffusion law in two dimensions, and the last term is an additional exponential decay with time constant τ (see Eq. 2.34).

- The final spatio-temporal *observable* is the spatial convolution of the simulated profile with the probe beam Gaussian function with $\mathrm{FWHM}_{\mathrm{pr}} = 0.9\,\mu\mathrm{m}$. This makes the absolute widths more similar to the experimental values, but has no effect on the diffusion dynamics (as discussed above).

For the first two examples I choose the parameters $A_0 = 1$ and $\tau = 10\,\mathrm{ps}$. Figures 3.3 and 3.4 summarize the results for diffusion coefficients of $D = 5\,\mathrm{cm}^2/\mathrm{s}$ and $D = 500\,\mathrm{cm}^2/\mathrm{s}$, respectively.

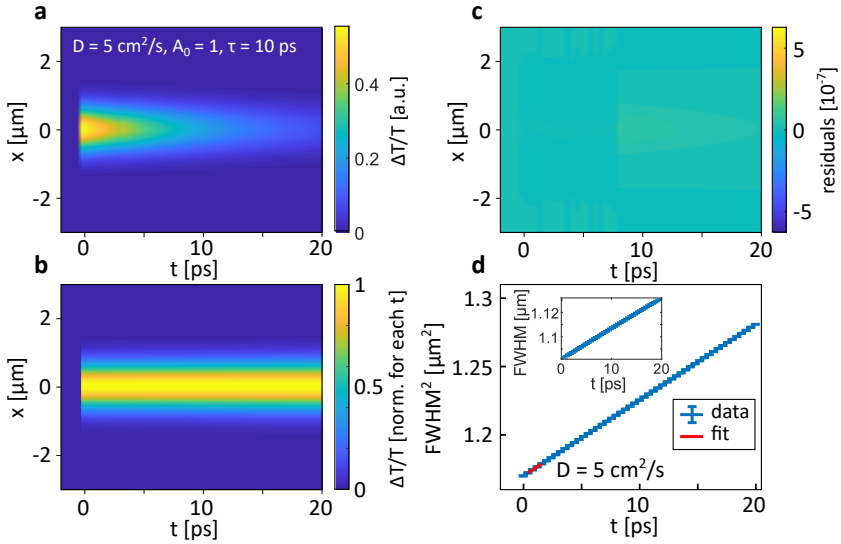

Fig. 3.3: Simulation results of a linear diffusion model with $D = 5\,\mathrm{cm}^2/\mathrm{s}$. (**a**) The spatio-temporal evolution of the simulated "observable" $\Delta T/T(x, t)$, as the initial ($t = 0$) profile diffuses and decays for $t > 0$. (**b**) The same data as in **a**, normalized for each vertical slice, i.e., for each t. The spread is too small to see by eye. (**c**) The Gaussian width analysis fit residuals, in the same absolute units as **a**. The low values indicate no deviation from a Gaussian shape. (**d**) Squared width evolution as a function of t from the Gaussian fits. From the slope of the curve, a diffusivity of $D = 5\,\mathrm{cm}^2/\mathrm{s}$ is recovered (see Eq. 2.31), which matches the simulated value. The inset shows the linear (not squared) width evolution, which also looks like a linear growth in time for this limited temporal observation window.

Figure 3.3a shows the simulated spatio-temporal maps with a diffusion coefficient of $D = 5\,\mathrm{cm}^2/\mathrm{s}$, as it decays in time with a decay constant of $\tau = 10\,\mathrm{ps}$. The maximum value is lower than the original amplitude $A_0 = 1$, as the profiles were convolved with a unit-area Gaussian to simulate the probe-beam scanning, while conserving the total signal strength. To visualize the transport, figure 3.3b shows the same map, normalized to the maximum value for each time slice. For this low diffusivity, it is hard to see a broadening by eye. Therefore, a Gaussian fitting was performed at each point in time t. The residuals of this fitting procedure are displayed in figure 3.3c in the units of $\Delta T/T$. Finally, figure 3.3d shows the squared width evolution over time, $\mathrm{FWHM}^2(t)$, while the inset also shows the (non-squared) width evolution, $\mathrm{FWHM}(t)$.
The same four graphs are displayed in figure 3.4 for the 100-fold higher diffusivity $D = 500\,\mathrm{cm}^2/\mathrm{s}$.
The results of the two simulated cases can be summarized as follows. As simulated, the FWHM^2 of the distribution rises linearly with the predicted slope (see Figs. 3.3d and 3.4d). It can also be seen that if one plots FWHM as a function of time, which

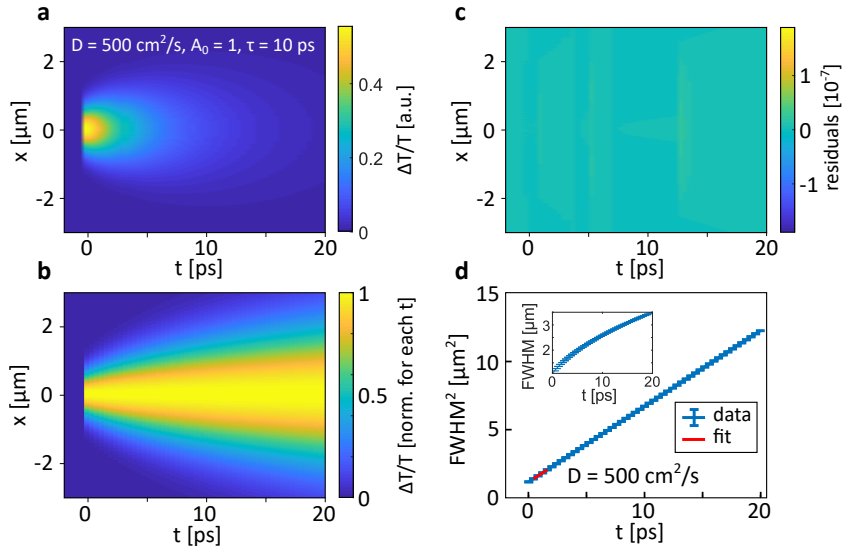

Fig. 3.4: Further simulation results of a linear diffusion model with higher diffusivity of $D = 500\,\mathrm{cm^2/s}$. The four panels show the same quantities as in figure 3.3. As before, the recovered diffusivity matches exactly the simulated value of $D = 500\,\mathrm{cm^2/s}$ and the residuals are negligible. This time, a clear spot widening can be seen in the normalized map (**c**) and the nonlinear (square root) dependence of FWHM(t) can be seen in the inset to **d**.

mathematically should follow a square root behavior, it depends on the level of D (as well as the observation time window and initial FWHM) whether a deviation from a linear slope can be seen. In practice, additional noise will make this distinction even harder. Note that the goodness of fit seems very high, as estimated by the error bars in the graphs of the width evolution, which come from the 68% confidence interval of the Gaussian fitting. In accordance, the residuals are extremely small (below 10^{-7}), as expected for noiseless perfect Gaussian simulated data.

3.2.2 Example 2: One species with offset

As a next step, I explore the possibility of a temporally constant offset to the data. The trivial case of a temporally and *spatially* constant offset will not change the measured dynamics at all, as long as the Gaussian fitting function includes an offset, as well.

Yet, when the offset does have a spatial shape, the situation becomes more complex. To this end, the same diffusing species as before is simulated on top of a small offset. The offset is also a Gaussian function in space, with a larger width compared to the diffusing profile ($\mathrm{FWHM_{offset}} = 3\,\mu\mathrm{m}$), and an amplitude of 1%, relative to the peak of $u(x,t)$. This choice is inspired by observed datasets, where such profiles could come from residual, nano- to microsecond heating and diffusion of the previous pulses of the pulsed laser.

Since this offset is constant in time, while $u(x,t)$ decays exponentially, this offset be-

comes ever more important for longer times. For example, after $t = 2\tau$ ($= 20\,\mathrm{ps}$ in this example), $u(x,t)$ will have decayed to $e^{-2} = 14\%$, which is closer to, yet still one order of magnitude above, the background level of 1%.
The results of this simulation, with $D = 5\,\mathrm{cm}^2/\mathrm{s}$, are summarized in figure 3.5.

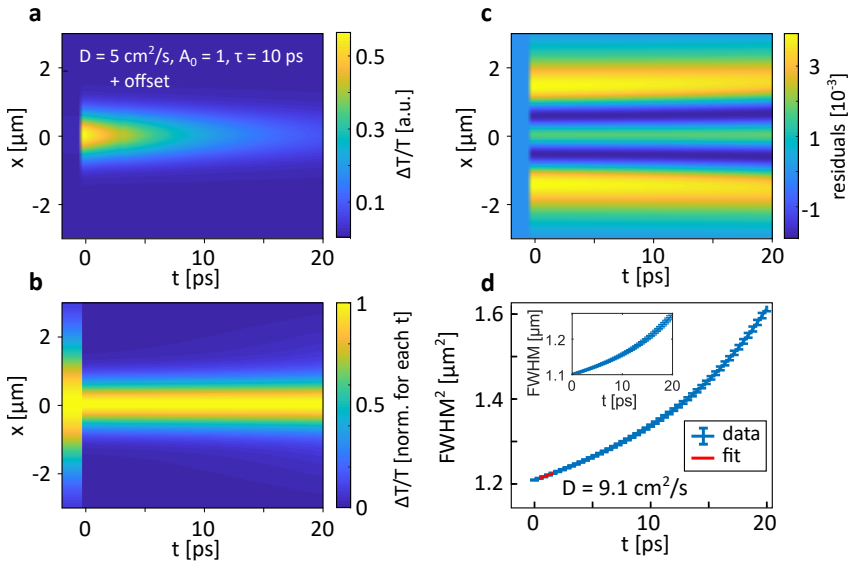

Fig. 3.5: Simulation results of linear diffusion with a small, yet spatially nonuniform background. (**a**) Spatio-temporal dataset. Neither the background level of 1%, nor the broadening is visible on this scale. (**b**) The data of **a**, normalized for each time t. Here, the broad Gaussian background becomes visible before time zero. (**c**, **d**) The residuals and the squared width evolution of the single Gaussian width analysis, respectively. Already at $t = 0$, the slope overestimates the input diffusivity of $D = 5\,\mathrm{cm}^2/\mathrm{s}$, due to decaying signals, yet constant (broad) background. The deviation from the single Gaussian profile, is also reflected in the five-fold residual shape, albeit at small (per mille level) amplitudes.

While the raw data in figure 3.5a looks nearly identical to that of figure 3.3a, the spatially broad offset can be seen in the normalized data of figure 3.3b before time zero. The important differences, though, lie in the width evolution and residuals. First of all, the squared width evolution shown in figure 3.3d does not grow linearly in time. Instead, it seems to grow faster with growing t. This comes from the fact that as the narrow Gaussian decays but the broad Gaussian remains, the single Gaussian fit "sees" an ever more broad distribution of what is essentially a sum of the two Gaussians. This is also reflected by the fact that the residuals, displayed in 3.3c, although still small (on the per mille level), are much bigger than before, and show a typical five-fold pattern, as the single Gaussian over- and underestimates the sum of the two Gaussians in space.
The most surprising result is that even the very initial diffusivity, extracted from the slope of $\mathrm{FWHM}^2(t)$, reads $9\,\mathrm{cm}^2/\mathrm{s}$. This is nearly a factor 2 steeper than the expected input diffusivity of $D = 5\,\mathrm{cm}^2/\mathrm{s}$.
To conclude, on the one hand, for a spatially non-uniform background the width evo-

lution is prone to show an evolution which, if interpreted as carrier diffusion, yields non-physical results. Also, the fact that the fits to single Gaussian curves are reasonably successful, shown by very small errors on the width parameter estimation and per mille level residuals, is not a guarantee for extracting real diffusion dynamics. On the other hand, for this simple case of a temporally constant background, which is already present before time zero, it should be possible to measure the background at $t < 0$ and subtract the spatial profile from the entire dataset. I have followed this approach when analyzing the experimental data in the following chapters.

3.2.3 Example 3: Two diffusing species with the same sign

The next example builds on the ideas of the previous one. This time, however, there is no constant background. Instead, I consider two diffusing species, both starting out at $t = 0$ with the same width as before, $\mathrm{FWHM}_{\mathrm{pu}}$. The two species have different diffusion speeds and different temporal decays. In particular, there is one species with high amplitude that diffuses quickly and decays quickly, and a second species with low initial amplitude which diffuses slowly, but also decays slowly. This approach is inspired by the case of having an electron and phonon subsystem, as in the case of noble metals (see chapter 5), yet here no coupling between the two species is considered. I choose the following parameters:

- Two independent amplitudes $A_1 = 10$ and $A_2 = 1$
- Two independent diffusivities $D_1 = 100\,\mathrm{cm}^2/\mathrm{s}$ and $D_2 = 5\,\mathrm{cm}^2/\mathrm{s}$
- Two independent decay constants $\tau_1 = 1\,\mathrm{ps}$ and $\tau_2 = 100\,\mathrm{ps}$

The results of this simulation are summarized in figure 3.6.
The simulated spatio-temporal dataset, shown in figure 3.6a shows a spike at time zero which quickly decays to a weaker signal level which survives over the 20 ps observation window. This reflects the two species decaying with time constants of $\tau_1 = 1\,\mathrm{ps}$ and $\tau_2 = 100\,\mathrm{ps}$. Figure 3.6b shows the normalized dataset. As before, it is not easy to see the width broadening by eye from this representation. Therefore, the single Gaussian fitting analysis is applied once again, resulting in the residual pattern shown in figure 3.6c, showing a clear five-fold pattern on the per-mille level. The width evolution is shown in figure 3.6d. It can be seen that the squared width rises fast within the first picoseconds, then decreases and finally increases again at a slower rate.
Intuitively, it is expected that in the beginning the species with the higher amplitude should dominate the dynamics. Later, the species with the slower decay rate should be left over. In this example, this results in a transition from a high to a low diffusivity. Yet, it is harder to imagine what should happen to the width in the transition region. Again, mathematically, it stands to reason that a single Gaussian analysis is not adequate, as the profile is made up of a sum of two Gaussians. Yet, if one Gaussian is much smaller than the other in amplitude, or if the widths are comparable, the final curve will differ only very little from a Gaussian profile. Again, the residual magnitude and pattern should report on this. A five-fold residual pattern is observed, especially in the transition region, although the absolute residuals are small, as before, on the per mille level. This means in practice, as noise will be present, it will not be straightforward to identify the presence of two species just by the goodness of fit.

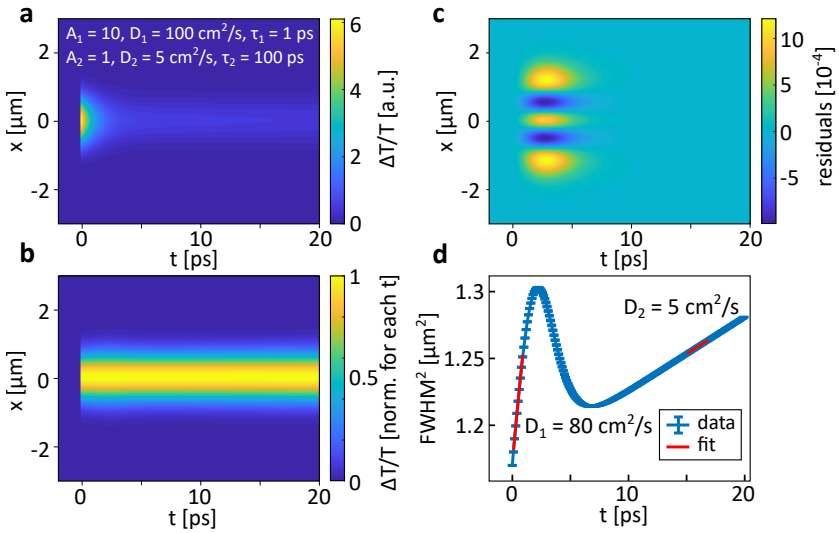

Fig. 3.6: Simulation results of linear diffusion of two uncoupled species with independent amplitudes, decays, and diffusivities. (**a**) The spatio-temporal dataset shows the fast decay within 1 ps, followed by a longer lived signal. (**b**) The data normalized at each t. On this scale, the initial broadening is barely visible. (**c**) The residuals from the Gaussian fitting shows a five-fold pattern at per mille amplitudes between 0 and 6 ps. (**d**) The squared width fit result shows a fast rising width in the first picoseconds, followed by a crossover regime with decreasing FWHM to a slower diffusion regime after 10 ps. The extracted diffusion coefficients are 80 cm^2/s for the first picosecond and 5 cm^2/s for the later times. An analysis of the diffusivity in the transition regime would result in a non-physical negative D.

Importantly, the fact that the width goes down in the transition region should not be interpreted as "shrinking". There has to be an effective decrease of FWHM, to bridge the gap between a high and low diffusivity regime for two species starting at the same initial width. However, this is a result of the interplay of two *positively* diffusing and decaying species.
A plausible solution to this issue would be to fit the resulting data to a sum of two Gaussians. In practice, for realistic values this approach did not prove very helpful. The additional free parameters involved in the fit make the convergence to the actual width evolution very unlikely. These attempts are not shown, as they resulted in very high uncertainties and unreliable fit results.

3.2.4 Example 4: Two diffusing species with opposite signs

As a final example, I simulate two diffusing species with opposite sign. The same simulation as Example 3 is implemented, with the only difference being the amplitude $A_2 = -1$, instead of $+1$ of before. This approach is also inspired by real data, e.g. the case of hBN encapsulated graphene described in chapter 6.
The results are summarized in figure 3.7.
Figure 3.7a shows the spatio-temporal datasets. A transition from the quickly decaying

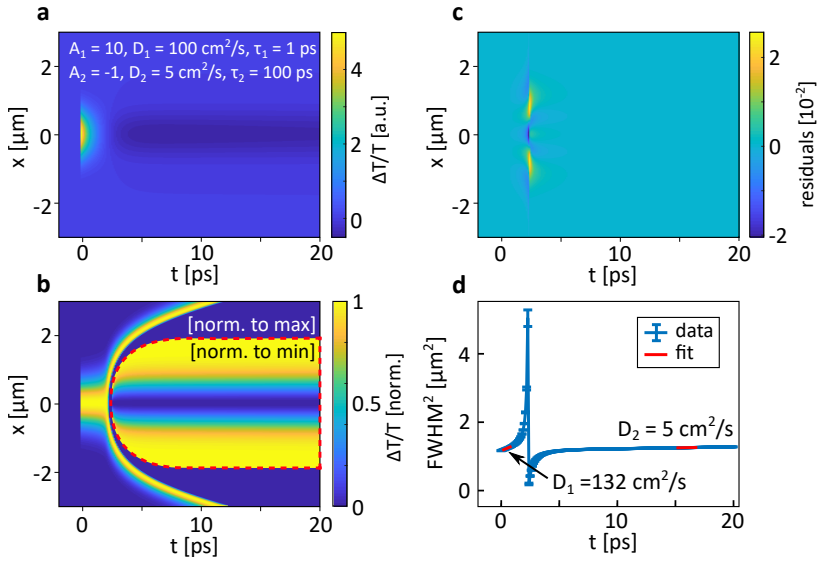

Fig. 3.7: Simulation results of linear diffusion of two uncoupled species with independent decays, diffusivities, as well as amplitudes of opposite sign. (**a**) The spatio-temporal dataset shows the fast decaying positive profile, followed by a longer lived negative signal. (**b**) An overlay of different regions normalized for each t to the maximum and minimum value, respectively. Because of the sign change, the normalization is chosen differently across the regions, separated by a red dashed line. (**c**) Residual pattern of the single Gaussian fitting routine. The highest residuals, on the percent level, come from the cross-over regime. (**d**) Squared width evolution, including a striking "pole" region around $t \approx 2.4\,\mathrm{ps}$, where the width measure seems to diverge. This region is characterized by percent level residuals at the pole, and per mille level residuals just before and after the pole.

large positive amplitude, to the long-lived, lower amplitude, negative signal is visible. In figure 3.7b I have chosen to normalize the spatio-temporal dataset to both the maximum, as well as the minimum for each time delay, and to display both resulting maps in specific regions, separated by a dashed red line. As before, figures 3.7c and 3.7d show the single Gaussian fitting residuals and squared width evolution, respectively.
In principle, one could expect to observe a width evolution as before, dominated by D_1 in the beginning and by D_2 in the end, even if the sign of the data has flipped. However, there is a further complication. As the signal changes sign, there must be a point in time where the signal is zero. Therefore, the width at this point will be ill-defined.
Looking at the simulation results, this is characterized by a "pole" region around 2.4 ps. While this could have been guessed, it is still striking to see the behavior at the crossover region from positive to negative signal. For example, the error bars of the FWHM and residual numbers don't tend to get very big near the pole, but instead stay around 1%. One might conclude the Gaussian fitting worked quite well, gaining confidence that the evolution is real. Yet, again, the five-fold pattern of the residuals indicate the need to be careful when interpreting the meaning of the width evolution.

As the slope of such a pole region can get arbitrarily high, these regions should be excluded when quantifying the diffusion.

3.2.5 Summary

In summary, I have presented different examples of simulated spatio-temporal diffusion, analyzed with Gaussian width fitting. The effect of spatially non-homogeneous backgrounds, multiple species, and sign changes were studied, and the importance of evaluating the residual patterns to judge the applicability of the fitting routine has emerged. The experimental estimation of physically reasonable diffusivities by single Gaussian fitting and squared width slope estimation is shown to lead to deviations from the underlying diffusion speeds. In particular, the interesting case of multiple species diffusing at the same time was considered. Here, the emergence of complex dynamics, such as decreasing FWHM, as well as diverging FWHM have been observed. However, this behavior can be understood, preventing unphysical interpretations, such as negative diffusion. Even for quite complicated scenarios, regimes where the measured diffusivity comes close to the actual values have been found. The learned lessons will be applied when discussing real transient absorption microscopy data in the following chapters.

4

Transient absorption microscopy: Experimental setup

In this chapter I introduce the basic experimental setup that is used in this thesis and discuss general considerations for the measurement of nanometer-femtosecond diffusion processes by probe-beam-scanning transient-absorption microscopy.
I describe the experimental details of how to trade off and push the limits of temporal resolution, spatial resolution, and sensitivity in a transient absorption microscope to study energy transfer processes with 0.25 ps temporal resolution, 20 nm spatial precision, as well as 10^{-6} relative signal sensitivity. Additionally, I provide tutorial descriptions of specific techniques I have found to be especially useful in the development of the experiment, such as how to overlap the beams in time and space. In the last part of the chapter, I present experimental results of tracking carrier diffusion in semiconductors. I study two exemplary types of semiconductors, bulk silicon as well as the novel two-dimensional material $MoSe_2$.

4.1 Setup: Description and characterization

As mentioned in the previous chapters, I am interested in directly imaging nanoscale ultrafast diffusion processes with a novel type of pump-probe microscopy. The basic idea of this class of experiments is depicted in figure 4.1a.

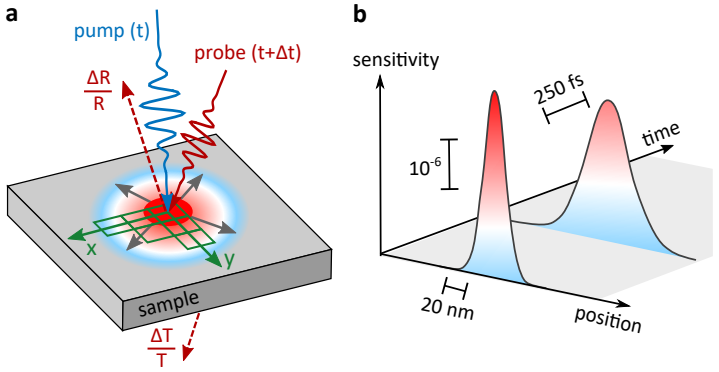

Fig. 4.1: Schematic of a probe-beam-scanning transient-absorption microscopy experiment. (**a**) An optical pump pulse is focused onto the sample under investigation, creating a local distribution of heat or carriers. A probe beam is spatially scanned over the illuminated area, for varying pump-probe delays. Transient reflection ($\Delta R/R$) or transmission ($\Delta T/T$) is recorded as a function of spatial offset and pump probe time delay. (**b**) The resulting profile's width evolution is measured with high temporal and spatial resolution and with high relative signal sensitivity. By monitoring the width of the transient-absorption profile as a function of time, carrier or heat diffusion can be tracked in a direct way.

An ultrafast optical pump pulse is focused onto the sample under investigation, creating a local distribution of either heat, carrier population or other type of energy distribution. To investigate the evolution of this distribution in space, a probe beam is spatially scanned over the pump beam for a given pump-probe time delay. By repeating this type of spatial scan for many different time-delays, a full spatio-temporal dataset is acquired. An equivalent measurement is to perform temporal pump-probe time delay scans for different spatial offsets. Depending on the sample's transparency, either the transient reflection $\Delta R/R$ or transmission $\Delta T/T$ is recorded.
By precisely monitoring the spatial width as a function of pump-probe time delay, as indicated in figure 4.1b, transport dynamics from the sample can be extracted in a very direct way, by probing the dynamics with high temporal resolution, high spatial precision, and by resolving relative signal strengths with high sensitivity. As shown in chapter 2.2, for general diffusion processes with Gaussian initial conditions, the diffusion coefficient $D(t)$ is extracted by analyzing the temporal derivative (slope) of the squared width evolution,

$$D(t) = \frac{1}{16 \ln 2} \frac{\partial}{\partial t} \mathrm{FWHM}^2(t). \tag{4.1}$$

In analogy to super-resolution microscopy, the precision of the width measurement can be far beyond the diffraction-limit and is ultimately only limited by the signal-to-noise

ratio.
A sketch of the setup used for this type of measurement is shown in figure 4.2.

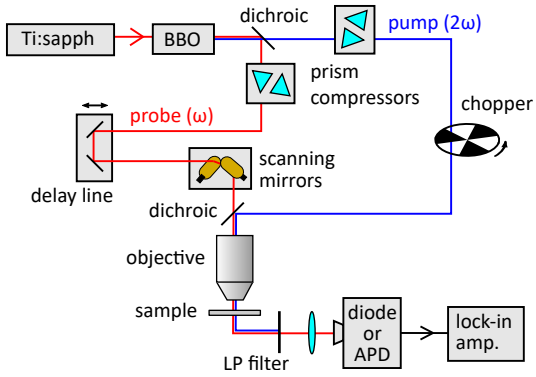

Fig. 4.2: Schematic beam paths for probe-beam-scanning transient-absorption microscopy. The fundamental (probe) beam is sent through a BBO crystal to create the second harmonic (pump) beam. The pump and probe are separated by a dichroic beamsplitter and individually sent through prism pairs for dispersion control. The pump is modulated by an optical chopper, while the probe travels over a mechanical delay line and scanning galvo mirrors. The beams are recombined and focused onto the sample by an objective. A longpass (LP) filter isolates the probe beam, which is detected with a lock-in detection scheme. For transient reflection, the *reflected* probe is detected instead (not shown). Additionally, the role of pump and probe can also be reversed, such that the fundamental (with frequency ω) is modulated as the pump, and the second harmonic (frequency 2ω) is detected with a shortpass filter as the probe.

The laser source used in the work presented in this thesis is a tunable, pulsed titanium sapphire (Ti:sapph) oscillator, Coherent Mira 900, pumped by a 532 nm continuous wave (CW) diode laser, Verdi V18, at 14 W pumping power. The Ti:sapph outputs pulsed light with a pulse duration of about 150 fs at 76 MHz repetition rate and time-averaged power of about 1.5 W. The center wavelength is tunable between 700 and 1000 nm, with a bandwidth of about 6 nm. The laser is focused onto a type 1 phase matched ($\theta = 29.2°$) β-barium borate (BBO) crystal to create frequency doubled pulses via second harmonic generation (SHG). After separation of the two wavelengths with a dichroic beam splitter, both beams propagate through prism pairs for precompensation of all sources of dispersion, e.g. the microscope objective and all lenses (see next section). Next, to achieve spatially non-elliptical, Gaussian laser modes, both beams are individually spatially filtered by transmission through pinholes (not shown in Fig. 4.2). The second harmonic (pump) beam is modulated by an optical chopper, while the fundamental (probe) beam is sent onto a variable delay line and over the scanning galvanometric (galvo) mirrors. Both beams are recombined by another dichroic beam splitter before entering the microscope objective and focused onto the sample*. The

*In this thesis, I quantify the laser fluence F impinging onto the sample, i.e., at the focus of the microscope objective, via the time-averaged power P (in W) at the back entrance of the microscope objective, the laser repetition rate f (76 MHz), and the "$1/e$"-radius of the Gaussian beam at the focus $r_{1/e}$, via: $F = P/(f \cdot \pi r_{1/e}^2)$. Note that for Gaussian beams multiple definitions for the beam size exist. In particular, $r_{1/e} = 1/\sqrt{2} \cdot r_{1/e^2} = \sqrt{2} \cdot \sigma_{\mathrm{RMS}} = 1/(2\sqrt{\ln 2}) \cdot \mathrm{FWHM}$.

reflected or transmitted light (Fig. 4.2 only shows transmission) is wavelength filtered to isolate the probe beam, which is sent onto a detector, such as a balanced photodiode or avalanche photodiode (APD). As the pump beam is being modulated, the probe signal can be demodulated with a lock-in amplifier, as introduced in chapter 3.1, to yield the transient reflection (transmission) signal. This signal is collected as a function of pump-probe time delay as the delay line is translated, and spatial offset as the probe beam is scanned over the pump with the galvo mirrors.

4.1.1 Temporal overlap and resolution

In this section, I describe the crucial steps needed to perform transient-absorption microscopy experiments with high temporal resolution, including an easy way to find time zero, as well as pulse dispersion control.

Finding time zero

Fine tuning the pump-probe delay for ultrafast spectroscopy is typically done with a retroreflector mounted on a micrometer precision stage, where 1 mm movement corresponds to 6.7 ps ($= 2 \cdot 1\,\mathrm{mm}/c$). Yet, to get within tens of picoseconds of equal beam paths can be a challenge. Simply measuring the beam paths is the first, coarse alignment. Other ways to find time zero rely on fast electronic detection, such as monitoring the beams with a fast photodiode or time-correlated single photon counting (TCSPC). Here, TCSPC is used to temporally overlap the pump and probe beams efficiently. Each beam is individually directed onto a single photon counting APD connected to a TCSPC system. Figure 4.3 shows the TCSPC histograms of both beams, after the path length has been adjusted iteratively.

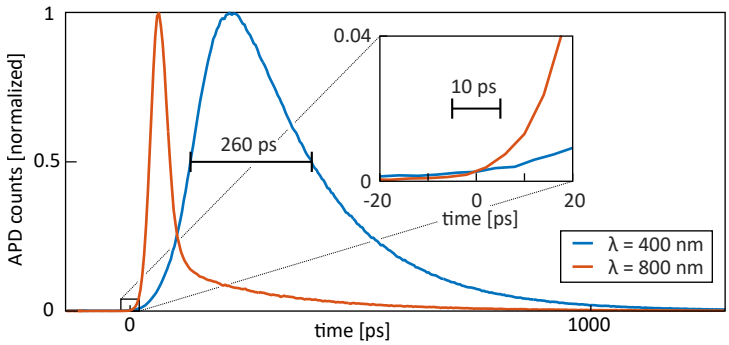

Fig. 4.3: Finding time zero with TCSPC. The onset of the TCSPC histograms of the two pulses have been overlapped by changing the optical path difference of the pump and probe beams. This technique allows for relatively quick alignment to find time zero ($\Delta t = 0$). Although the two colors produce very different instrument response widths, the pulses can be overlapped to within about 10 ps, by making the onsets coincide. The inset shows the zoom into the relevant region.

The absolute time difference between the fundamental ($\lambda = 800\,\mathrm{nm}$) and the second harmonic ($\lambda = 400\,\mathrm{nm}$) pulses is measured and adjusted to within about 10 ps. This

is done after passage through the objective, beam splitters, lenses, etc., to include all differences between the two beams. Due to the start-stop configuration of the TCSPC electronics, the onset of the TCSPC signal comes from the quickest photon-to-electron conversion events, while the average lag-time produces the width of histogram towards later times. Figure 4.3 shows that this temporal response curve, known as the instrument response function (IRF) for lifetime measurements, can be very different for the two colors. Here, the fast electronic readout of an APD optimized for NIR light, shows an IRF with a FWHM of 45 ps for $\lambda = 800\,\text{nm}$. Yet, for $\lambda = 400\,\text{nm}$, the same detector and electronics produce a FWHM width of 260 ps. This difference is not important here, though, as only the *onsets* of both histograms have to be overlapped to find time zero.
After initial coarse alignment of the beam paths and a measurement of Δt with TCSPC, one beam path can be adjusted with respect to the other on the order of millimeters to overlap them to within about $\sim 10\,\text{ps}$, as shown in the inset of figure 4.3. Then, a short delay line of 25 mm is enough to find the exact t_0 via transient absorption on the sub-picosecond timescale, limited only by the precision of the delay stage actuator.

Dispersion compensation

Any dispersive media in either of the two beam paths affects the pulse duration of the femtosecond pulses because of group delay dispersion (GDD), therefore decreasing the overall temporal resolution of the experiment. This includes all lenses, nonlinear crystals in case of SHG, thick, high-index crystals in case of modulation with acousto-optical modulator (AOM) or electro-optical modulator (EOM) technology, as well as the objective, where a higher NA typically implies more dispersion due to longer optical path length through glass elements. The first strategy to avoid this problem is to reduce the number of dielectric elements in the beam path. This can be achieved by replacing glass lenses or objectives with reflective mirrors, or replace AOM/EOM modulation methods with optical choppers. Here, I use glass lenses, as reflective optics have other disadvantages, such as off-axis alignment for spherical mirrors, and limited NA and non-Gaussian point spread functions due to mirror obstruction of the beam path for reflective objectives. I do use an optical chopper instead of AOM or EOM, which limits the modulation frequency to about 10 kHz, without significant loss of sensitivity, as discussed in section 4.1.3.
The effect of dispersion is illustrated by measuring the autocorrelation of the 800 nm pulsed Ti:Sapph oscillator with and without passage through a 5 cm long BK7 glass rod, as an estimation of the kind of GDD that a high NA objective can cause. Figure 4.4a shows the two measured autocorrelations of the 800 nm pulses.
These data are obtained by splitting the laser into two parts and overlapping them onto a BBO crystal at an angle. While each beam can create SHG in the direction of the beam, the SHG created by the both beams together travels into a different direction, due to the phase-matching condition. This signal is isolated and monitored as a function of the delay between the two beams. The only dielectrics the beam passes before this measurement are four standard lenses. An increase in the FWHM of the autocorrelation from $\Delta t_{\text{in}}^{\text{AC}} = 195\,\text{fs}$ to $\Delta t_{\text{out}}^{\text{AC}} = 250\,\text{fs}$ is observed, shown in figure 4.4a. To retrieve the (FWHM) pulse width Δt from the autocorrelation width Δt^{AC}, in principle one needs to know the shape of the pulse[72]. Yet, if one assumes

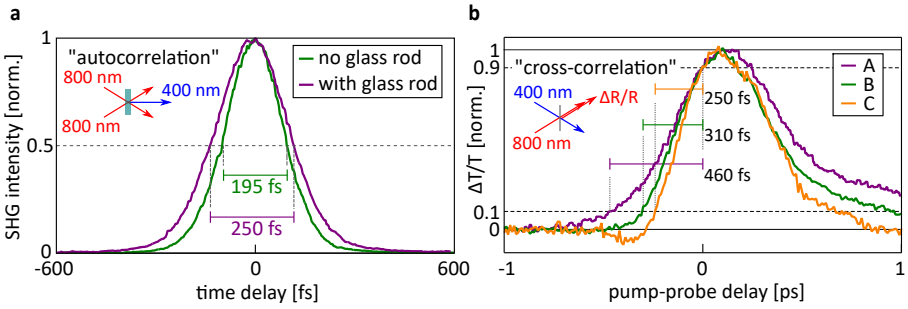

Fig. 4.4: Estimation and compensation of group velocity dispersion. (**a**) Autocorrelation of the $\lambda = 800\,\text{nm}$ beam with and without a 5 cm long BK7 glass rod in the beam. The presence of the glass leads to significant broadening of the autocorrelation, i.e., SHG from a BBO crystal. (**b**) Determination of the $10-90\%$ rise time of the transient absorption signal $\Delta T/T$ from a graphene sample during an iterative prism compression process. Curve A shows the response after coarse installation of the prisms at their calculated separation distance. Curve B is the result of an iterative adjustment of the distance of the prisms in the pump beam path to minimize the rise time. Curve C is the result after subsequent prism distance iteration for the probe beam path.

a Gaussian shape, the following simple relation holds, known from the convolution of two identical Gaussians (c.f. eq. 3.10),

$$\Delta t_\alpha = \frac{1}{\sqrt{2}} \Delta t_\alpha^{\text{AC}}, \tag{4.2}$$

where $\alpha =$ "in" and "out" stand for the pulse before and after the dispersive element, respectively. By measuring the spectrum of the laser (not shown), I find a Gaussian shaped spectrum with a FWHM of $\Delta\lambda = 7.26\,\text{nm}$, centered around $\lambda = 800\,\text{nm}$. Converting to spectral bandwidth, I find $\Delta\nu = c\Delta\nu/\lambda^2 = 3.40\,\text{THz}$. Using equation 4.2, I calculate the pulse width without the glass rod $\Delta t_{\text{in}} \approx 138\,\text{fs}$. From there, the time-bandwidth product (TBP) can be calculated as $\text{TBP} = \Delta t_{\text{in}} \cdot \Delta\nu = 0.47$. This is reasonably close to the value for a transform limited, Gaussian shaped pulse of $\text{TBP} = 0.441$[72].
Then, the additional GDD from the glass rod can be estimated directly from the increase of the autocorrelation width[73],

$$\text{GDD} = \frac{1}{4\ln 2}\sqrt{(\Delta t_{\text{out}} \cdot \Delta t_{\text{in}})^2 - (\Delta t_{\text{in}})^4} \approx 5500\,\text{fs}^2. \tag{4.3}$$

To precompensate for dispersive elements with a GDD on this order of magnitude, I install prism pairs for pulse compression in both pump and probe beam sections[74]. I choose fused silica and SF10 glass for the 400 nm and 800 nm beams, respectively. This choice comes from the calculated prism distances for dispersions of this order of magnitude, fitting easily on the optical table ($30-50\,\text{cm}$). To quantify the temporal resolution of the entire system, I measure the transient response rise time of a graphene sample, as its electron heating response time is expected to be much faster than the laser pulse width used here[75]. After coarse installation of the prism pairs at the estimated distance, I observe a $10-90\%$ rise time of 460 fs, shown as the purple curve

(A) in figure 4.4b. After iteration of the prism pair distance to minimize the rise time for the pump beam (curve B, green) and the probe beam (curve C, orange), I arrive at a final rise time of 250 fs. The achieved pulse compression is satisfactory, as it corresponds to a cross correlation FWHM of 230 fs in a 1.4 NA oil immersion objective from a laser with a pulse width of about $\sim$ 150 fs, and its corresponding second harmonic pulse.

4.1.2 Spatial resolution and scanning

In this section I describe techniques and characterization methodologies to achieve high spatial resolution in transient-absorption microscopy.
All beams used in this work are tightly focused with high NA objectives (between 0.5 and 1.4) to achieve high spatial confinement. Yet, the final spatial resolution can be defined in different ways and depends on the mode of operation, as explained before. Each beam is focused to a spot size limited ultimately by the NA of the objective and the light's wavelength. I have explained in chapter 3.1.4 that while the PSF for SS-TAM is smaller than the sizes of the individual beams, for PBS-TAM it is actually larger ($\sigma_{\text{PBS-TAM}} = (\sigma_{\text{pu}}^2 + \sigma_{\text{pr}}^2)^{1/2}$). Importantly, this PSF is only the starting width for a diffusion experiment and slight changes to the width can still be tracked. To characterize the beams, I estimate the spot sizes at the sample plane with a "knife-edge method" described below.

Overlapping foci in the sample plane

To achieve the best spatial resolution and signal levels, it is important that the pump and probe beams focus to the same volume. In the lateral dimensions (x,y) this is achieved coarsely by adjusting the incoming beams to overlap in position and angle as they enter the microscope objective. This is typically done by imaging the back-reflected beams from a flat sample (e.g. glass coverslip or mirror) and adjusting the beams to make the reflected beams' shape symmetric. Further fine adjustment is achieved by moving one beam with respect to the other with the galvo mirrors, thus optimizing the TA signal.
In the z-direction, the focus overlap depends on the collimation of the two beams and the chromatic aberrations of the optical elements in the illumination pathway, in particular the microscope objective. For example, for 400 and 800 nm light, even semi-apochromat and apochromat objectives both show varying focal lengths for incoming collimated beams on the order of a few micrometers, i.e., often larger than the focal depth, or Rayleigh length, as depicted in figure 4.5a.
A possible solution to this problem is to slightly decollimate one or both of the beams to compensate for this offset, as shown in figure 4.5b. By measuring the beam profile with a knife edge method[76], i.e., scanning a step-like structure through the area around the focus of the beam and measuring transmitted or reflected power as a function of z, the z-offset can be estimated and corrected. Here, a 50 nm evaporated gold film with a scratched edge is scanned through both pump and probe beams individually with a piezo sample scanner. This method has the advantage of being independent of

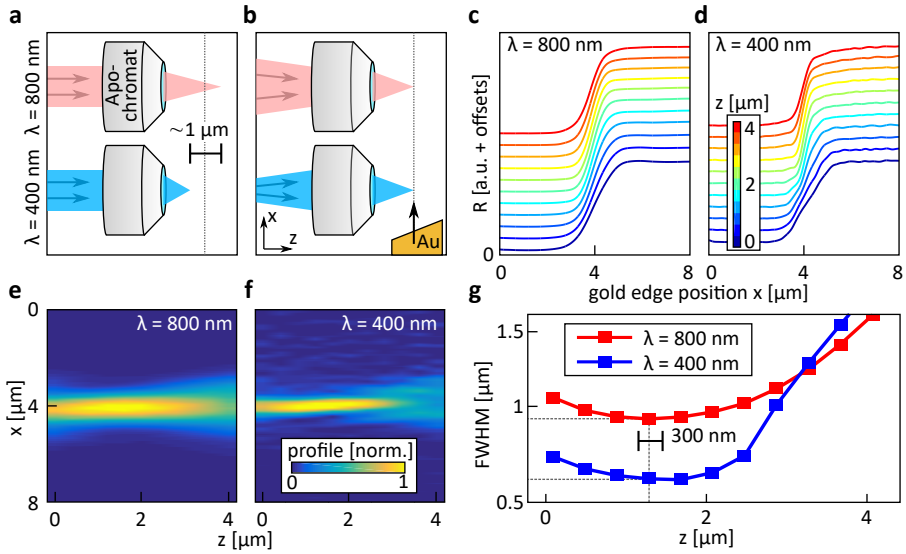

Fig. 4.5: Beam overlap through profiling with the knife-edge method. (**a**) Chromatic aberration leads to micrometer scale differences in focal z-position for different colors, even in apochromat objective lenses. (**b**) By beam profiling and iterative (de-)collimation the focuses of the beams can be brought to better overlap. (**c**,**d**) Reflected power is recorded as a function of position of a gold edge scanned through the probe (**c**) and pump (**d**) beam. (**e**,**f**) The spatial derivatives of the profiles along the x-axis represent the beam profiles as a function of scanning direction, x, and axial offset, z, for (**e**) the probe and (**f**) the pump beam. (**g**) Extracted width (FWHM) for both beams by Gaussian fits to line profiles, i.e., vertical cuts, of **e** and **f**, showing the z-dependence. After optimization, the minimum width, corresponding to the focus, of the two beams overlaps to within about 300 nm in z. The pump and probe focal sizes were measured as $\mathrm{FWHM_{pu}} \approx 0.6\,\mu\mathrm{m}$ and $\mathrm{FWHM_{pr}} \approx 0.9\,\mu\mathrm{m}$. The same characterization was also done for the other spatial dimension, y, yielding similar results.

the imaging pathway[†].

Figure 4.5c shows the reflected intensity of the probe beam ($\lambda = 800\,\mathrm{nm}$) as a function of the gold edge position x for varying z-positions. Figure 4.5d shows the same measurement for the pump ($\lambda = 400\,\mathrm{nm}$). To retrieve the beam shape in the x-z-dimension, I take the spatial derivative of the traces along x. This recreates, in first approximation, a 1D projection of the beam profile, as it reverses the action of integration due to the partial beam reflection and integrated collection of the intensity. Recently demonstrated deviations between measured and actual widths on the order of $50-100\,\mathrm{nm}$ for these type of knife-edge measurements, depending on wavelength and polarization[77], have been ignored here. Figure 4.5e and f show the resulting x-z-beam profiles for the two beams, respectively. The same procedure is repeated for the y-z-profile (not shown here). By fitting Gaussian functions to the z-slices, the

[†]It should be noted that the idea to monitor the sample plane with an imaging system composed of the same, infinity corrected objective, a tube lens, and a camera, and trying to minimize the back reflected spot sizes of both beams does not work for decollimated beams.

minimum z-position, i.e., the focal plane, is extracted, as shown in figure 4.5g. I present the result after multiple iterations of beam profiling and slight adjustments of the collimation lenses of down to sub-millimeters. Finally, the foci are overlapped to within about 300 nm, resulting in an improved spatial resolution and TA signal level. The final result are beams which have been purposely decollimated to overlap in the focal plane. The lateral profiling of the beams results in $\mathrm{FWHM_{pu}} = 0.6\,\mu\mathrm{m}$ and $\mathrm{FWHM_{pr}} = 0.9\,\mu\mathrm{m}$, close to the diffraction limit. From this a PSF for SS-TAM of about 0.5 µm and an initial width for a PBS-TAM experiment of about 1.1 µm is expected, according to equations 3.8 and 3.12, respectively. The advantage is that the foci now overlap satisfyingly in all three spatial dimensions, a prerequisite for high signal-to-noise PBS-TAM on the nanoscale.

Galvo mirror scanning

To scan the probe beam independently of the pump beam, the probe beam has to be manipulated after beam separation and before beam recombination. The most straightforward way to do this is by steering the beam with scanning galvo mirrors. This was displayed before in a simplified way in figure 4.2, without any lenses between the galvo mirrors and the objective. This would lead to the beam translating away from the objective for very small mirror angle movements. Instead, an imaging systems is installed between the mirrors and the objective. To understand how to place the scanning optics, i.e., the relative positioning of the mirrors, lenses, and objective, I use ray transfer matrix analysis, also known as ABCD matrix analysis.
In this formalism, the distance x and the angle θ with respect to the optical axis of a beam traveling close to the optical axis can be traced while traversing optics, such as lenses, by multiplying a 2×2 matrix to the vector (x, θ). Here, the scan lens (f_1), tube lens (f_2), and even the objective lens (f_{obj}) are approximated as thin lenses. As shown schematically in figure 4.6, the distances between the lenses (d_1 to d_4) are set to the conjugated condition, i.e., the scan lens and tube lens form a 4f-system from the scanning mirror pivot point to the back focal plane (BFP) of the objective. The sample plane coincides with the focal plane of the objective.

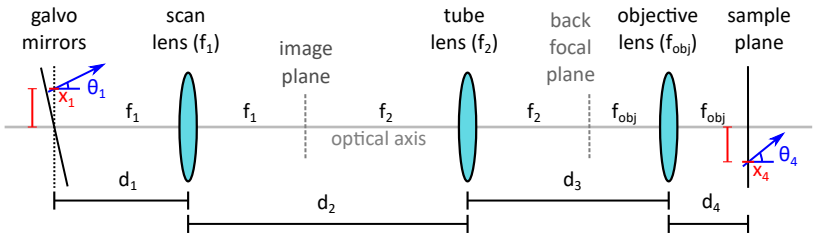

Fig. 4.6: Schematic of the scanning galvo mirror positioning. The positions of the lenses used in the experiment are shown for the corresponding ray transfer matrix calculation. With this configuration, the beam angle θ_1, controlled by the galvo mirrors, is translated into sample plane displacements x_4, without offsetting the beam from the optical axis at the back focal plane of the objective.

The final position and angle (x_4 and θ_4) can be calculated from the input position x_1

and angle θ_1 by matrix multiplication with $\mathbf{M}_{\text{total}}$,

$$\begin{pmatrix} x_4 \\ \theta_4 \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \theta_1 \end{pmatrix} = \mathbf{M}_{\text{total}} \cdot \begin{pmatrix} x_1 \\ \theta_1 \end{pmatrix}, \tag{4.4}$$

where $\mathbf{M}_{\text{total}}$ is given by the multiplication, in reverse order, of the ray transfer matrices of all the free space propagations and traversed thin lenses,

$$\mathbf{M}_{\text{total}} = \mathbf{S}_{d_4} \cdot \mathbf{L}_{f_{\text{obj}}} \cdot \mathbf{S}_{d_3} \cdot \mathbf{L}_{f_2} \cdot \mathbf{S}_{d_2} \cdot \mathbf{L}_{f_1} \cdot \mathbf{S}_{d_1}. \tag{4.5}$$

The free space propagation matrix $\mathbf{S}_d$ for propagation distance d and the thin lens matrix $\mathbf{L}_f$ for a lens with focal distance f are defined as

$$\mathbf{S}_d = \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{L}_f = \begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix}. \tag{4.6}$$

By inserting all the distances and lenses as depicted in figure 4.6, equation 4.4 can be expressed as

$$\begin{pmatrix} x_4 \\ \theta_4 \end{pmatrix} = \begin{pmatrix} 0 & -f_{\text{obj}} \frac{f_1}{f_2} \\ \frac{1}{f_{\text{obj}}} \frac{f_2}{f_1} & 0 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \theta_1 \end{pmatrix}. \tag{4.7}$$

This simple result only takes this form when the lenses are placed in the conjugated positions. In particular, the vanishing first element of $\mathbf{M}_{\text{total}}$ relies on the fact that $d_2 = f_1 + f_2$ and $d_4 = f_{\text{obj}}$ (as drawn in Fig. 4.6). To understand the spatial scanning condition, I evaluate the position x_4 of the beam at the sample plane,

$$x_4 = -f_{\text{obj}} \frac{f_1}{f_2} \theta_1. \tag{4.8}$$

From this it becomes evident that the final position at the sample plane, x_4, only depends on the angle at the scanning mirrors, θ_1, which is the desired function of the scanning mirrors.
Here, it should be noted that objectives are typically not specified by their focal length, but by their magnification M, defined assuming an image distance L. L depends on the objective manufacturer ($L_{\text{Nikon}} = 200\,\text{mm}$, $L_{\text{Olympus}} = 180\,\text{mm}$, $L_{\text{Zeiss}} = 165\,\text{mm}$). Using the relations $M = L/d_4$ and $1/L + 1/d_4 = 1/f_{\text{obj}}$, the objective focal length can be calculated from the magnification via:

$$f_{\text{obj}} = \frac{L}{M+1}, \tag{4.9}$$

and is typically around 3 – 4 mm for the high NA objectives used in this thesis.
Finally, to understand the imaging of the mirrors into the BFP of the objective, it is also instructive to calculate the beam at the position of the back focal plane x_{BFP}. The ray matrix calculation reveals:

$$\begin{pmatrix} x_{\text{BFP}} \\ \theta_{\text{BFP}} \end{pmatrix} = \begin{pmatrix} -\frac{f_2}{f_1} & 0 \\ 0 & -\frac{f_1}{f_2} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \theta_1 \end{pmatrix}. \tag{4.10}$$

This shows that the mirrors are imaged onto the back focal plane of the objective. The beam diameter is expanded by the ratio of focal distances of tube and scan lens f_2/f_1 around the unchanging mirror pivot point x_3, which is aligned into the center of the objective aperture and does not change during scanning. The angles of the mirrors are mapped to angles at the back focal plane of the objective, demagnified (if $f_2 > f_1$) by the inverse ratio f_1/f_2. The maximum (full) scan angle $\phi^{\max}$ of the mirrors can also directly be calculated from equation 4.7, as

$$\phi^{\max} = 2\theta_1^{\max} = -\frac{f_2}{f_1}\frac{2x_4^{\max}}{f_{\text{obj}}} = \frac{\text{FoV}}{f_1}\frac{f_2}{f_{\text{obj}}}, \tag{4.11}$$

where the field of view FoV is defined as the full range of possible sample positions $\text{FoV} = 2x_4^{\max}$. In case the objective manufacturer's choice of tube lens (e.g. $f_2 = 180\,\text{mm}$ for Olympus) is used, the term f_2/f_{obj} simplifies to $M + 1$, i.e.,

$$\phi^{\max} = \frac{\text{FoV}}{f_1}(M+1). \tag{4.12}$$

An equivalent description would be that the Fourier transform function of the scan lens maps the galvo mirror angles to position changes in the intermediate image plane (also drawn in Fig. 4.6), which is then imaged into the sample plane by the 4f-system consisting of the tube and objective lens.
In summary, I have demonstrated a sensible galvo mirror scanning lens configuration and simple calculations that have proven helpful when designing the setup. Most experiments in this thesis were performed with a scan lens $f_1 = 150\,\text{mm}$ and a tube lens $f_2 = 400\,\text{mm}$, chosen with a ratio to widen the beam diameter to match the back aperture of the objective. For an Olympus objective ($L = 180\,\text{mm}$) with a $40\times$ magnification, a full scan angle of 10 mrad (about half a degree) scans a sample area of about 16 µm, calculated via equation 4.11. This is more than the maximum area needed for PBS-TAM with tightly focused beams.

Spatial resolution for sample-scanning TAM

For a basic test of the spatial resolution, the system's PSF can be estimated by imaging samples with feature sizes well below the diffraction limit. Here, two samples of lithographically defined nanorods (size $\ll \lambda$) are imaged with SS-TAM.
Figure 4.7a shows the $\Delta T/T$ image of an array of identical gold nanorods with pump and probe wavelengths $\lambda_{\text{pu}} = 400\,\text{nm}$ and $\lambda_{\text{pr}} = 800\,\text{nm}$, respectively. The corresponding scanning electron microscope (SEM) image is shown on the same scale as an inset. By fitting Gaussian functions to the individual antenna response profiles, the average width is determined as an estimate of the PSF of this transmission SS-TAM at these wavelengths. The FWHM is determined as about 330 nm. Figure 4.7b shown a transient reflection image of a different nanoantenna sample with varying lengths and wavelengths $\lambda_{\text{pu}} = 450\,\text{nm}$ and $\lambda_{\text{pr}} = 900\,\text{nm}$. Here, the PSF of is measured as a FWHM of 380 nm. For comparison, figure 4.7c shows the simultaneously recorded image of the probe beam reflection R, with a PSF of about 610 nm. The SS-TAM PSF is therefore significantly smaller than the reflection microscopy PSF, as well as the individual beam profiles measured before. This shows the nonlinear nature of SS-TAM, as summarized in equation 3.8.

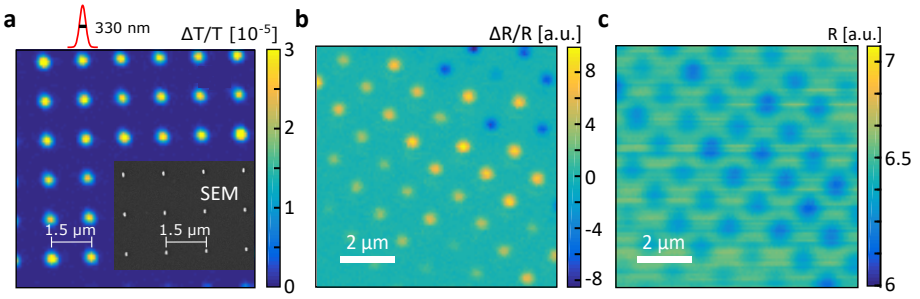

Fig. 4.7: Transient-absorption scans of gold nanoantennas, as a resolution standard for SS-TAM. (**a**) $\Delta T/T$ map of lithographically defined gold nanoantennas of the same length, recorded at $\Delta t \approx 0$. The point spread functions of this nonlinear microscopy have a FWHM of about 330 nm, at $\lambda_{pr} = 800$ nm. The inset shows an SEM image of the nanoantennas. (**b**) $\Delta R/R$ map of lithographically defined gold nanoantenna with varying lengths, recorded at $\Delta t \approx 0$. The point spread functions have a FWHM of about 380 nm, at $\lambda_{pr} = 900$ nm. (**c**) Reflection map, recorded at the same time as **b**, demonstrating the improved spatial resolution of SS-TAM. The FWHM point spread function of this transmission microscopy is about 610 nm.

The different sign of the antenna response seen in figure 4.7b is likely due to a competition between bleaching of the surface plasmon band and an excited state absorption[78]. Although this interpretation is not the focus of this study, it shows how TAM can reveal more information together with a higher spatial resolution than regular microscopy.

Spatial precision for probe-beam-scanning TAM

As mentioned before, sub-diffraction limited information about diffusion can be extracted with PBS-TAM. The resolution here is now directly related to the precision of the spatial profile, i.e., how well two spatial profiles with different widths can be distinguished. One straightforward way to quantify this is to extract confidence intervals for the width parameter of the Gaussian fitting. This measure is complemented by checking the trace-to-trace spread of the extracted width for noisy data. By analyzing the error-bars and the point-to-point variation for the data shown in figure 5.3 of the next chapter, the width precision of this technique is estimated to be about 20 nm for this type of measurement with probe powers on the order of 100 µW, relative signals below 10^{-4} and data averaging over a few seconds, i.e. averaging tens of lines with 75 pixels and 5 ms/pixel integration time[70].
Another source of uncertainty in the PBS-TAM width estimation comes from the conversion from rotation angle of the galvo mirrors to an absolute length scale. The mirror displacement is gauged directly in the sample plane with an antenna sample, like the one shown in figure 4.7a. Here, the center-to-center distance of the nanoantennas of 1.5 µm have been fabricated by electron beam lithography with nanometer precision. By scanning an area on the order of 10×10 µm on such a sample, the applied driving voltage of the galvo movement controller can be converted to a precise displacement in the sample plane. All mirror-scanning measurements presented in this thesis have been calibrated in this way.

4.1.3 Sensitivity: Noise and lock-in detection

The advantage of time resolution in any transient-absorption measurement comes at the cost of inherently low relative signal levels, as only the differential changes to the observed transmitted (or reflected) light are measured on top of a large background. Chong *et al.* calculated and measured the response of a single molecule and a small metal nanoparticle in nonlinear microscopy to be around 10^{-7} and 10^{-4} relative signal, respectively[79]. To be able to measure these low relative intensity levels, on top of a large background signal, one needs to overcome low frequency noise[80]. Therefore, I modulate the pump beam and record only the probe beam, demodulating the signal at the modulation frequency with a lock-in amplifier (lock-in). It is often claimed that $1/f$-noise dominates the noise level between kHz and MHz modulation, without proof in terms of a noise spectral analysis[81,82]. The source of such noise has been attributed both to laser intensity fluctuations, as well as electronic noise[83,84]. The actual noise spectrum depends on many experimental factors, and can be dominated by sources other than just laser noise and shot noise[85]. In particular, mechanical vibrations of individual optomechanical elements on an optical table can contribute significantly to the total noise.
For these reasons, here, the TAM noise scaling is characterized to find out what modulation frequencies and integration times are needed to reach the high signal-to-noise ratio and sensitivity required for PBS-TAM to work as a tool to track diffusion on the nanoscale.

Noise characterization

First, I analyze the spectrum of the noise of the probe beam on the detector, reflected off a blank coverslip. By recording time traces for 5 s and varying the demodulation frequency of the lock-in, the noise level is estimated by taking the standard deviation of the traces. The low pass filter settings of the lock-in are given by the time constant t_c (in ms) and the roll-off ΔL, measured in decibel per octave (dB/oct). Figure 4.8a shows the measured noise level as a function of the demodulation frequency. A decrease in the relative noise for increasing frequency is observed, with a plateau above about 3 kHz.
Next, I choose an experimental modulation frequency of 6.4 kHz, as it shows the lowest noise for the achievable modulation frequencies with the optical chopper, and perform the transient absorption experiment. The data is collected from the output of the lock-in (with its own integration time and filter) with an analog-to-digital converter with a sampling rate of ($4\,\mathrm{ms}^{-1}$). Figure 4.8b shows the standard deviation (noise) of the TAM data collected on graphene, for different settings of the time constant and roll-off, as a function of the number of time bins used for signal averaging. The dashed line shows the calculated behavior of a purely shot noise limited probe beam with 3.5 µW transmitted power, when sampled at the different averaging rates. The calculated relative noise scales as $1/\sqrt{N}$ with the total number of collected photons N, and includes a correction for the detector quantum efficiency. Note that both the time constant as well as the roll-off ΔL of the lock-in influence its effective "total" integration time. It can be seen that, depending on the settings of the lock-in, the setup is capable of detecting nearly shot noise limited signals below 10^{-6} with only about 100 ms integration time and an optical chopper running at $f = 6.4$ kHz, i.e.,

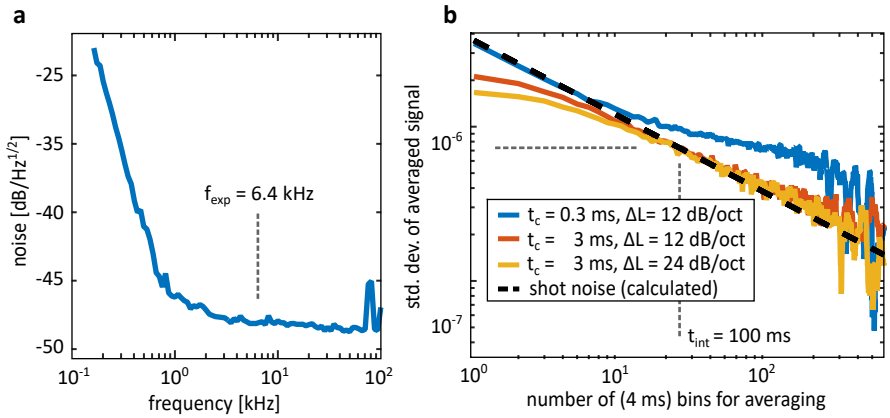

Fig. 4.8: Noise characterization to estimate smallest measurable signal. (**a**) Noise spectrum recorded with only the probe beam on the detector and varying the lock-in demodulation frequency. An experimental modulation frequency of 6.4 kHz is chosen. (**b**) Noise level, estimated from the standard deviation of averaged $\Delta T/T$ time traces on a graphene sample with increasing number of averaging bins. The three solid curves show data collected at different settings of the lock-in low pass filter's time constant t_c and roll-off ΔL. The dashed line shows the calculated relative noise of a shot noise limited signal (probe laser, 3.5 µW at the photodiode).

without the need to use AOM or EOM technology to reach MHz modulation. This is a great advantage both for simplicity of the optical setup, as well for the temporal resolution, as AOM and EOM crystals introduce a large GDD chirp onto the laser pulses, significantly increasing the pulse width. In the rest of the thesis, if not stated otherwise, a roll-off of $\Delta L = 24\,\mathrm{dB/oct}$ and a time constant of $1-3\,\mathrm{ms}$ is used, as these values have shown to lead to the lowest noise levels for this PBS-TAM setup.

4.2 Imaging graphene with transient-absorption microscopy

So far I have shown that TAM can perform well in terms of temporal resolution, spatial resolution, and sensitivity. Furthermore, compared to linear microscopy, additional information can be extracted. As transient absorption probes the sample's photo-excited state properties, contrast can be significantly enhanced. As an example, I show SS-TAM maps of two different graphene samples in figure 4.9.
Figure 4.9a shows a $\Delta T/T$ map recorded on graphene, made by chemical vapor deposition (CVD), which was deposited by wet transfer onto a glass substrate. Line-shaped and pointlike features can be seen on the sample, similar to shapes which have been previously attributed to wrinkles and polymer residue particles, respectively[86]. Figure 4.9b shows a $\Delta T/T$ map on graphene encapsulated between two layers of hexagonal boron nitride (hBN), while figure 4.9c shows the simultaneously recorded reflection. The $\Delta T/T$ image is very sensitive to the edge of the graphene, while the transmission microscopy shows the biggest contrast at the edge of the hBN. Additional features

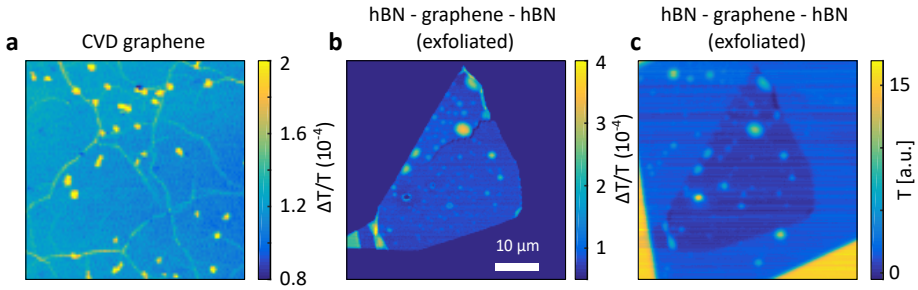

Fig. 4.9: SS-TAM imaging of different graphene samples. Spatial images of transient transmission of (**a**) CVD graphene and (**b**) hBN encapsulated graphene. (**c**) shows the transmission (T) map recorded during the acquisition of **b**, corrected for probe beam fluctuations. The $\Delta T/T$ map shows the highest contrast at the graphene boundaries, while the contrast in reflection is dominated by the hBN edge (yellow corners). All maps were recorded at the pump-probe delay with the highest $\Delta T/T$ signal, i.e., $\Delta t \approx 0$. The scale bar of **b** is valid for all three images.

that look like folds, creases, grain boundaries, and bubbles can be distinguished and imaged with high spatial resolution, demonstrating the power of SS-TAM, even just for a single pump-probe delay[87]. Even more information can be extracted when this type of scan is performed as a function of Δt. Indeed, SS-TAM has been successfully applied for materials characterization, biological imaging, and art imaging[88].

4.3 Tracking exciton diffusion in semiconductors

As a first concrete example of the method of PBS-TAM explained in this chapter, I present spatio-temporal imaging of the exciton diffusion in two different semiconductors, bulk silicon, as well as flakes of atomically thin monolayer $MoSe_2$.

4.3.1 Bulk silicon

As described in chapter 2.1.6, the conductivity properties of silicon have been studied over many decades, as it is the industrial standard for computer chips, diodes and solar cells. Direct relationships between electric conductivity, mobility and carrier diffusivity are known from the Einstein relation.
Here, I present spatio-temporal imaging data on bulk p-type silicon, from an industrial wafer. I choose this sample because it is characterized in terms of doping level and therefore gives a clear prediction of the exciton diffusion.
To investigate the temporal decay dynamics, I spatially overlap the pump and probe beams and vary the time-delay. Figure 4.10a shows the resulting trace.
The transient reflection shows a negative step response, followed by a biexponential decay with fast (6 ps) and slow (210 ps) components.
To visualize exciton diffusion, I proceed to scan the probe beam over the fixed pump beam for many different time delays. Figure 4.10b shows the resulting spatio-temporal dataset. I extract the FWHM at each time-delay by fitting Gaussian profiles to the

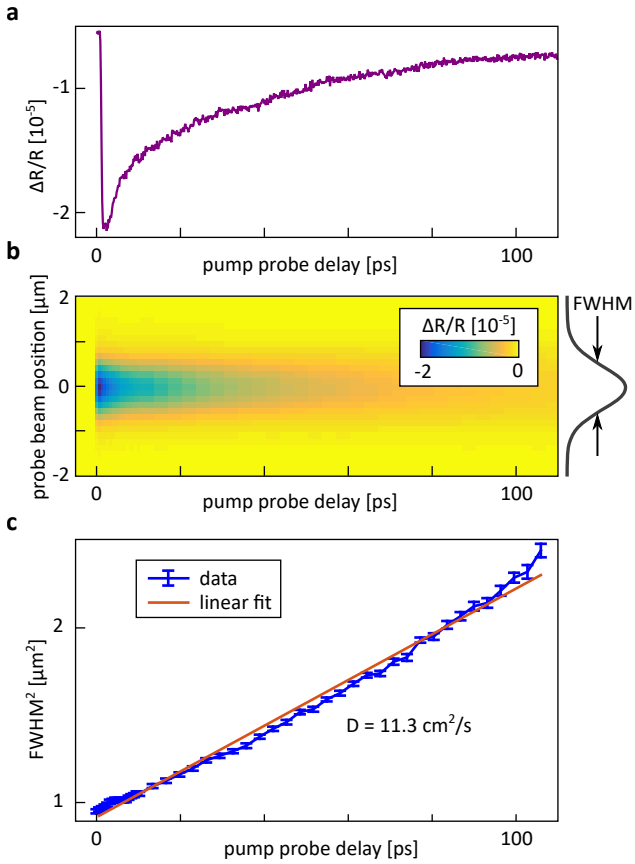

Fig. 4.10: Spatio-temporal PBS-TAM datasets recorded on silicon. (**a**) Temporal $\Delta R/R$ dynamics, recorded at $x = y = 0$, i.e., for overlapping beams. The temporal response fits to a biexponential decay with time constants $\tau_1 = 6\,\text{ps}$ and $\tau_1 = 210\,\text{ps}$, respectively (not shown). (**b**) Spatio-temporal dataset recorded by scanning the probe beam over the fixed pump beam. (**c**) Squared width evolution, extracted from **b** by Gaussian fitting to vertical slices, i.e., for each pump-probe delay (symbols). The error bars show the 68 % confidence intervals of the Gaussian fits. The red curve shows a linear fit, resulting in a diffusion coefficient of $11.3\,\text{cm}^2/\text{s}$, according to equation 2.31.

$\Delta R/R$ spatial profiles, as previously explained, and plot the resulting temporal evolution of the squared width, FWHM^2, in figure 4.10c. The squared width is observed to rise approximately linearly with pump-probe delay. By fitting the slope, together with equation 2.31, a diffusion coefficient of $11.3\,\text{cm}^2/\text{s}$ is extracted.
This value fits well with reported values for the majority carrier diffusivity reported in literature for degenerately doped, p-type silicon[89].

4.3.2 Monolayer molybdenum diselenide

The emergence of two-dimensional materials has opened up an exciting platform for material science and fundamental physics research. Beyond graphene, TMDs have surfaced as atomically thin semiconductors with an optical band gap. While exciton diffusion in bulk silicon, probed in reflection geometry, was shown to give sensible results, the situation becomes more difficult if the sample is only one atom thick, as signal-to-noise is likely to be worse. However, recent studies have shown that it is possible to quantify exciton diffusion in monolayer TMDs with PBS-TAM[39–41].
Here, I investigate samples of commercially available flakes of molybdenum diselenide ($MoSe_2$) on a sapphire coverslip. The results are summarized in figure 4.11.
Figure 4.11a shows the temporal response for collinear pump and probe beams. The transient reflection response shows a negative peak, followed by a biexponential decay with time constants of $\tau_1 = 1\,\mathrm{ps}$ and $\tau_2 = 300\,\mathrm{ps}$, respectively. The inset shows SS-TAM imaging of the differently sized, triangular, monolayer $MoSe_2$ flakes. These scans are used to find homogeneous parts of the samples on which to perform the PBS-TAM scans. Different patterns can be observed, including a darker spot located roughly in the center of the flakes. The flake on the right hand side of figure 4.11 is an example of a flake which was chosen to show enough homogeneity over a size of a few micrometer, so that PBS-TAM can be performed on the sample.
Figure 4.11b shows the full spatio-temporal dataset on such a homogeneous region of a $MoSe_2$ flake. As before, Gaussian fitting is performed for each vertical slice to the data. Figure 4.11c shows the squared width evolution and linear fit to the data in the first 25 ps. A diffusion coefficient of $11.6\,\mathrm{cm}^2/\mathrm{s}$ is obtained. The value obtained here is quite similar to the value measured by Kumar *et al.*[41] While the study of exciton diffusion is not the main focus of this thesis, these results are seen as a proof of principle that nanoscale ultrafast diffusion can be indeed be resolved with PBS-TAM.

4.4 Summary

In summary, I have described the hardware and techniques to set up and characterize a system for spatio-temporally resolved transient-absorption microscopy capable of tracking heat and carrier diffusion in solids. The combination of ultrafast lasers, microscopy and lock-in detection yields high temporal resolution, high spatial precision and the necessary sensitivity to resolve small relative changes.
I have measured the spatio-temporal transient absorption dynamics on bulk silicon, as well as atomically thin monolayer $MoSe_2$ flakes. The observed diffusion dynamics are interpreted as exciton diffusion and match the expected behavior.

Acknowledgments

The experiments reported in this chapter were performed under the supervision of Dr. Matz Liebel and Prof. Dr. Niek F. van Hulst. The results were discussed and interpreted together with Dr. Matz Liebel and Prof. Dr. Niek F. van Hulst. The nanoantenna sample was fabricated by Lisa Saemisch under the supervision of Prof. Dr. Niek F. van Hulst, the CVD graphene sample wet transfer was done by Gabriele

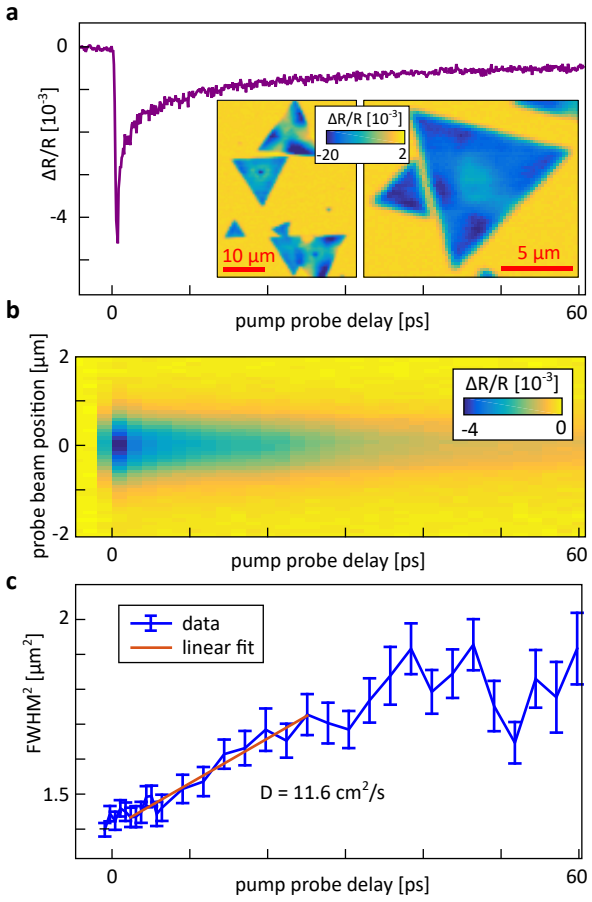

Fig. 4.11: PBS-TAM dataset on monolayer $MoSe_2$ flakes. (**a**) Temporal dynamics, recorded at $x = y = 0$, i.e., for overlapping beams. The temporal response fits to a biexponential decay with time constants of $\tau_1 = 1\,\text{ps}$ and $\tau_2 = 300\,\text{ps}$, respectively (not shown). The insets show SS-TAM imaging of the $MoSe_2$ flake sample. (**b**) Spatio-temporal dataset recorded by probe beam scanning over fixed pump. (**c**) Squared width evolution, extracted from **b** by Gaussian fitting to vertical slices, i.e., for each pump-probe delay (symbols). The error bars show the 68 % confidence intervals of the Gaussian fits. The red curve shows a linear fit, resulting in a diffusion coefficient of $11.6\,\text{cm}^2/\text{s}$. The linear fit was only performed over the time window up to 25 ps, because the data at longer time delays shows a slight sub-linear behavior.

Navickaite, and the hBN encapsulated graphene sample was exfoliated by Mathieu Massicotte, under the supervision of Prof. Dr. Frank Koppens. The monolayer $MoSe_2$ flake sample was purchased from the company 2DLayer.

5

Hot-electron diffusion in thin gold films

In this chapter I study hot-electron diffusion in thin gold films via probe-beam scanning transient-absorption microscopy.
The ultrafast response of metals to light is governed by intriguing non-equilibrium dynamics involving the interplay of excited electrons and phonons. The coupling between them gives rise to nonlinear diffusion behavior on ultrashort timescales. Here, I use scanning ultrafast thermo-modulation microscopy to image the spatio-temporal hot-electron diffusion in thin gold films. By tracking local transient reflectivity, two distinct diffusion regimes are revealed: an initial rapid diffusion during the first few picoseconds, followed by about 100-fold slower diffusion at longer times. Remarkably, a slower initial diffusion is found than the value previously predicted for purely electronic diffusion. A comprehensive three-dimensional model is developed, based on a two-temperature model and evaluation of the thermo-optical response, taking into account the delaying effect of electron-phonon coupling. The simulations describe well the observed diffusion dynamics and let us identify the two diffusion regimes as hot-electron and phonon-limited thermal diffusion, respectively.

This chapter is based on the following publication:
A. Block, M. Liebel, R. Yu, M. Spector, Y. Sivan, F. J. García de Abajo, N. F. van Hulst, Tracking ultrafast hot-electron diffusion in space and time by ultrafast thermo-modulation microscopy *Science Advances* **5**, eaav8965 (2019).

5.1 Introduction

The ultrafast carrier diffusion dynamics in noble metals, such as gold, is of particular importance for heat management in nanoscale devices, as well as femtosecond laser ablation, but has thus far only been studied in the time domain[21,28,30,90–101]. While these time-resolved studies have uncovered numerous aspects of ultrafast carrier dynamics, the nanoscale spatial transport has remained largely unexplored.
Here, I use novel ultrafast microscopy to gain insight into the complex non-equilibrium dynamics of hot electrons in gold.
When a metal interacts with light, its electrons are excited above the Fermi level (see figure 5.1a).

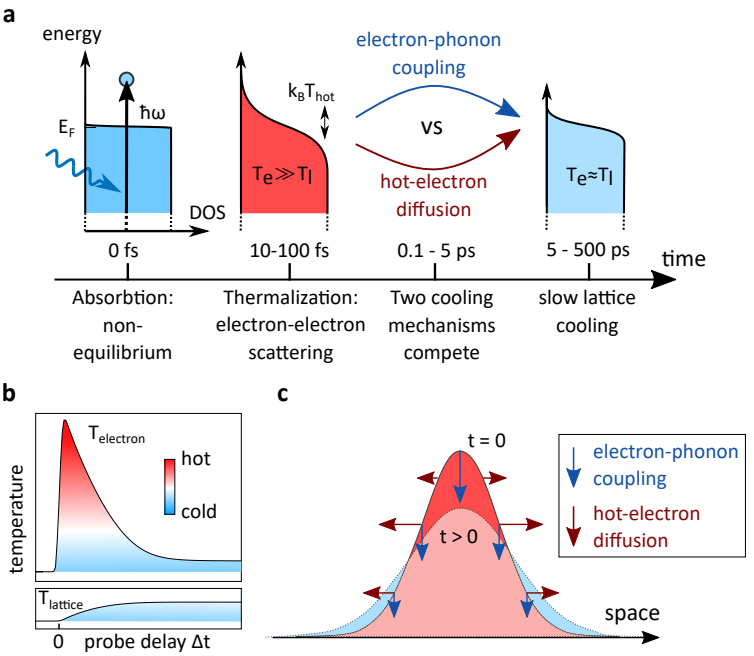

Fig. 5.1: Schematic of non-equilibrium light-metal interaction. (**a**) The energy distribution of the conduction band electrons at ambient temperature is perturbed by optical excitation. It quickly evolves to a quasi-thermalized "hot-electron" Fermi-Dirac distribution with high electron temperature (T_e), while the lattice temperature (T_l) stays close to the ambient level. Subsequent cooling due to electron-phonon coupling and hot-carrier diffusion leads to thermal equilibrium between the electron and lattice subsystems. (**b**) Schematic electron and lattice temperature evolution. Optical excitation to the electron system with its low heat capacity leads to a rapid increase for the electron temperature and subsequent equilibration with the lattice. (**c**) The two competing cooling mechanisms of hot-electron diffusion and electron-phonon coupling should be discernible in the spatial domain, as only hot-electron diffusion leads to a broadening of an initially excited area.

These initially excited electrons trigger a cascade of cooling mechanisms that involve energy transfer between different metal subsystems, of which conduction electrons and lattice vibrations (phonons) are the dominant ones. Indeed, the ultrafast ther-

mal response is commonly described by considering the conduction band electrons and the ionic lattice as separate thermodynamic subsystems with distinct thermal properties, such as heat capacity and thermal conductivity[27]. Since the heat capacity of the electrons is substantially smaller than that of the lattice, the electron subsystem quickly reaches a high temperature Fermi-Dirac distribution, which I refer to as "hot electrons", while the lattice stays close to ambient temperature under the excitation intensities considered here. As depicted in figure 5.1b, the electrons subsequently cool and thermalize with the lattice. This typically happens within a few picoseconds after excitation. The electron relaxation and thermalization of the two subsystems are a direct result of an interplay between electron-phonon coupling and hot-electron diffusion[97]. Purely time-resolved studies typically cannot separate these two contributions and have consistently neglected lateral heat flow[90–92,97], which becomes particularly important when considering nanoscale systems. Directly resolving hot-electron diffusion, a crucial step for understanding the ultrafast heat dynamics, has thus far not been possible due to a lack of spatio-temporal resolution[95,97–100,102]. Optically induced thermal diffusion has been studied on the micro- to nanosecond scale via thermal grating spectroscopy[103–106], yet the distinction of hot-electron diffusion from electron-phonon coupling remains a challenge with this technique[61], as well as the limitation to relatively large samples.

Here, I address this shortcoming by interrogating thin gold films with the recently developed scanning ultrafast thermo-modulation microscopy (SUTM) technique. Figure 5.1c shows how hot-electron diffusion could be distinguished from electron-phonon coupling in the spatial domain, as only the former leads to a broadening of the spatial distribution. I directly measure the spatio-temporal evolution of a locally induced hot-electron distribution on nanometer length scales with femtosecond resolution. Specifically, in this experiment, an optical pump pulse illuminates a thin gold film, thus creating hot carriers in the metal. The subsequent thermal response of the metal is measured by employing a probe pulse that interrogates the sample at a well-defined time-delay, Δt, with respect to the pump pulse. By spatially raster-scanning the probe beam relative to the stationary, tightly focused pump spot at each time-delay, spatio-temporally resolved transient reflection ($\Delta R/R$) maps are obtained. In this experiment, spatial heat diffusion manifests itself as a broadening of the initially excited area, which is quantified with a nanometer accuracy, far beyond the diffraction limit, by accurate determination of the SUTM spatial-response function. Considering the electron redistribution dynamics described above (Fig. 5.1a), I expect to identify distinct diffusion dynamics regimes, each dominated by a different diffusion mechanism, depending on the state of thermalization of the sample.

5.2 Spatio-temporal dynamics

I image the light-induced thermal dynamics of a 50 nm thin gold film using SUTM. The sample is optically excited with a 450 nm (2.76 eV) pump pulse and interrogated with a 900 nm (1.38 eV) probe pulse, as outlined above. The beams are focused to spots of full-width at half-maximum (FWHM) at the sample plane of 0.6 and 0.9 µm, respectively, and $\Delta R/R$ maps are recorded at different times after photo-excitation by varying the pump-probe delay between −5 and 30 ps.

Figure 5.2a shows transient reflection images, recorded at three different pump-probe

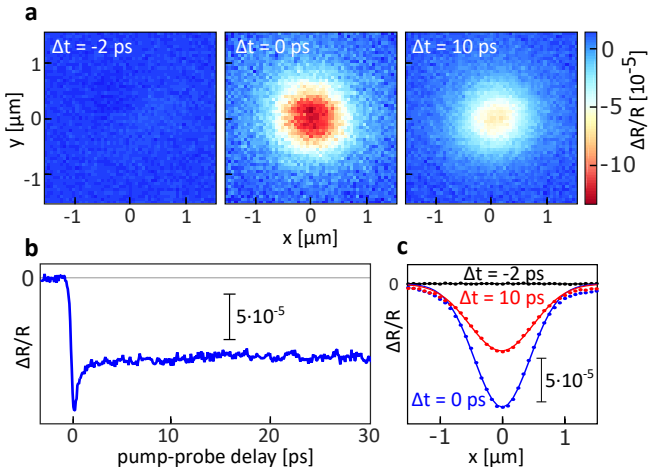

Fig. 5.2: Hot-electron dynamics. (**a**) Scanning ultrafast thermo-modulation images of the gold film recorded by spatially scanning the probe beam ($\lambda_{pr} = 900\,\text{nm}$) relative to a fixed pump beam position ($\lambda_{pu} = 450\,\text{nm}$, centered at $x = y = 0$ in the image) for selected pump-probe delays. (**b**) Transient reflectivity dynamics for collinear pump and probe pulses, exhibiting two distinct exponential decay contributions with time constants of 1 ps and 0.9 ns, respectively. An offset at $\Delta t < 0$ has been subtracted from the data. (**c**) Spatial profiles (dots) and Gaussian fits (curves) for three selected pump-probe delays, extracted from **a** by cutting horizontal lines through the center of the spots ($y = 0$).

time-delays Δt for a fixed pump fluence F of $1.0\,\text{mJ/cm}^2$, which is well below the damage and ablation threshold of the metal[107]. At $\Delta t = -2\,\text{ps}$, i.e., when the probe interrogates the sample before photo-excitation by the pump, the $\Delta R/R$ response is negligible. Then, at $\Delta t = 0\,\text{ps}$, a negative $\Delta R/R$ spot emerges around the pump beam position $x = y = 0$, with up to 10^{-4} contrast. Subsequently, at $\Delta t = 10\,\text{ps}$, a reduced response is observed, as the signal has decayed. The negative sign of $\Delta R/R$ indicates that the heated area exhibits decreased reflection. The data were recorded with an integration time of 5 ms/pixel. To investigate the temporal decay dynamics, I spatially overlap the pump and probe beams ($x = y = 0$) and vary the time-delay. Figure 5.2b shows the resulting trace. The transient reflection shows a negative step response with a 300 fs rise time, close to the instrument temporal resolution, followed by a biexponential decay with fast (1 ps) and slow (0.9 ns) components. While these data contain information about the temporal carrier dynamics, the spatial diffusion information is provided by the $\Delta R/R$ maps of figure 5.2a. Therefore, I fit the central cross sections of the images with Gaussian functions (Fig. 5.2c). The FWHM increases by about 100 nm from 1.05 µm at $\Delta t = 0\,\text{ps}$ to 1.15 µm at $\Delta t = 10\,\text{ps}$, a result that is attributed to spatial heat diffusion. The accuracy of this method is ultimately limited by the signal-to-noise ratio of the response function, which dictates how well the profiles can be fit. As explained in chapter 4.1.2, a FWHM accuracy of about 20 nm is observed.

I further investigate this evident diffusion behavior, as I want to track the rate at which it takes place during the thermalization process. Therefore, I proceed to record transient reflectivity line profiles over many pump-probe delays. Figure 5.3a shows

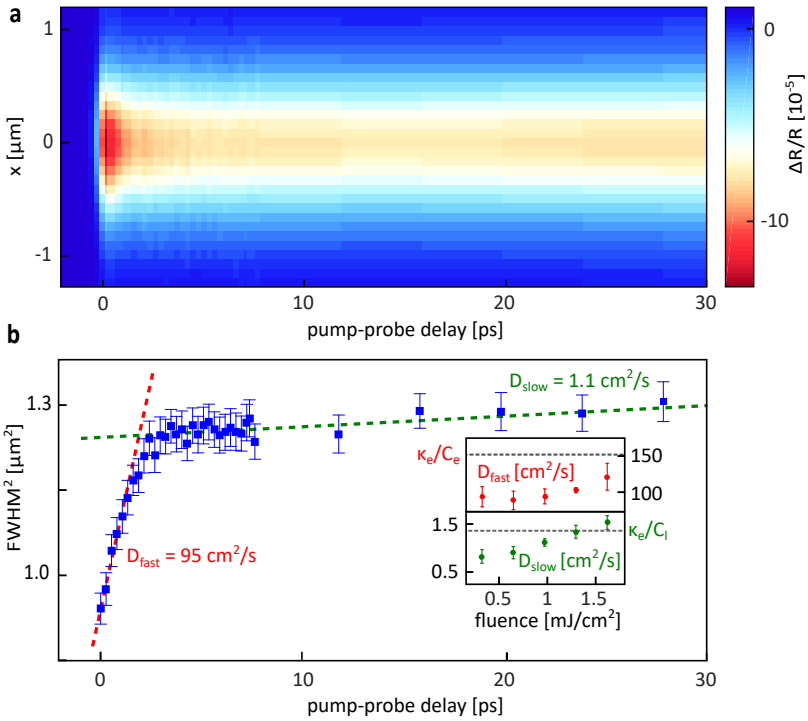

Fig. 5.3: Two-step diffusion dynamics. (**a**) Spatio-temporal dynamics of the transient reflection signal $\Delta R/R$. The probe beam is scanned over the pump beam (1D scan across spot center, vertical axis) as a function of pump-probe delay (horizontal axis) at a pump fluence of 1.0 mJ/cm^2. The offset at $\Delta t < 0$ has been subtracted from the data. (**b**) Squared-width evolution of the $\Delta R/R$ profile, extracted by Gaussian fitting to the spatial profile at each pump-probe delay (symbols). The error bars show the 68 % confidence intervals of the Gaussian fits. The initial and final diffusion coefficients are extracted by fitting slopes in the two regions (dashed lines) and comparing to equation 5.1. A fast diffusion of $D_{\text{fast}} = 95\,\text{cm}^2/\text{s}$ is found within the first few picoseconds, followed by slower diffusion ($D_{\text{slow}} = 1.1\,\text{cm}^2/\text{s}$) after >5 ps. The inset shows the extracted diffusion coefficients for pump fluences in the 0.3 – 1.6 mJ/cm^2 range, along with the ratio of electron thermal conductivity κ_e and electron (lattice) heat capacity C_e (C_l).

a typical spatio-temporal dataset for $F = 1.0\,\text{mJ/cm}^2$. An estimate of the time-dependent diffusion coefficient $D(t)$ is obtained in a first, semi-quantitative manner by assuming the following relationship between the width (FWHM) and D for an initial Gaussian profile, which I adopt from a general treatment of diffusion problems as explained in chapter 2.2,

$$\frac{\partial \text{FWHM}^2(t)}{\partial t} = 16 \ln 2 D(t). \tag{5.1}$$

I extract the FWHM at each time-delay by fitting Gaussian profiles to the $\Delta R/R$ maps, as previously explained, and plot the resulting temporal evolution of the squared width, FWHM^2, in figure 5.3b. An initial fast spreading is observed, revealed by an increase

in FWHM^2, followed by a much slower broadening at longer time-delays. The two diffusion regimes appear to correlate with the fast and slow temporal decay regimes of $\Delta R/R$. I estimate the initial and final diffusion coefficients by fitting lines to the curve in figure 5.3b (Eq. 5.1). Excluding the transition region, I restrict the fit to the $\Delta t = 0-1$ ps and $5-30$ ps intervals, yielding $D_{\text{fast}} = 95\,\text{cm}^2/\text{s}$ and $D_{\text{slow}} = 1.1\,\text{cm}^2/\text{s}$, respectively. I repeat the same procedure for multiple pump fluences and summarize the extracted coefficients, along with their standard errors, in the inset to figure 5.3b. The above analysis provides a simple first measure to quantify diffusion. Yet, it relies on the assumption of a single diffusing profile and its proportionality to transient reflection, while ignoring the underlying electron and phonon subsystems, as well as their individual thermal contributions to the reflection signal. Considering these (non-linear) contributions to the dynamics, the asymptotic linear fitting is too simplistic. To fully understand the nature of the time-dependent diffusion mechanisms at work, in particular the transition between the two regimes of the observed spot broadening dynamics, it is necessary to model the response of the system more rigorously.

5.3 Modeling

I model the spatio-temporal evolution of the pump-induced changes to the reflectivity of the sample in three basic steps (see Fig. 5.4).

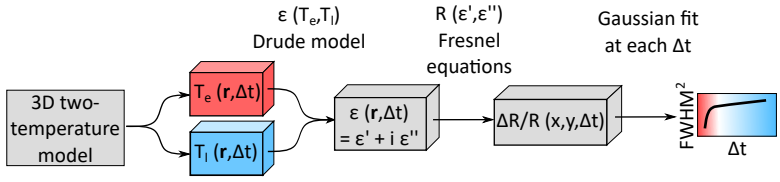

Fig. 5.4: Schematic of the theoretical modeling. The spatio-temporal evolution of the optically excited gold film is simulated with a full 3D-space two-temperature model. The temperature-dependent complex permittivity is obtained from the calculated spatio-temporal electron and lattice temperature maps, including the thermal dependence of electron-electron and electron-phonon scattering, as well as thermal expansion of the lattice. Then, $\Delta R/R(x, y, \Delta t)$ is calculated using the thin-film Fresnel equations and extracting its spatial dynamics using the same Gaussian fitting as in the experimental data analysis.

First, the electron and lattice temperature distributions in the gold film, $T_e(\mathbf{r}, \Delta t)$ and $T_l(\mathbf{r}, \Delta t)$, resolved in space ($\mathbf{r} = x, y, z$) and time ($\Delta t$), are calculated by means of a 3D two-temperature model[27,28]. Then, the spatio-temporal dynamics of the gold film permittivity $\epsilon(\mathbf{r}, \Delta t)$ is calculated, considering the effect of both temperatures on the Drude response. Finally, the permittivity is converted to the observable, transient reflection $\Delta R/R(x, y, \Delta t)$, using the Fresnel equations. This procedure allows for a direct comparison of the measured data to the predictions of the model for the spatio-temporal diffusion by Gaussian fitting of the resulting $\Delta R/R$ maps to obtain FWHM^2 as a function of Δt. I describe the three steps involved in more detail in the following sections.

5.3.1 Two-temperature model

In this model, the electron and lattice temperature T_e and T_l are assumed to be time- and position-dependent, with their spatio-temporal behavior being described by the following coupled equations[27],

$$\begin{aligned} C_e(T_e)\frac{\partial T_e}{\partial t} &= \nabla(\kappa_e(T_e)\nabla T_e) - G\cdot(T_e - T_l) + S - S_{es}, \\ C_l\frac{\partial T_l}{\partial t} &= \nabla(\kappa_l(\nabla T_l) + G\cdot(T_e - T_l) - S_{ls}, \\ C_s\frac{\partial T_s}{\partial t} &= \nabla(\kappa_s(\nabla T_s) + S_{ls} + S_{es}, \end{aligned} \tag{5.2}$$

where C_e/C_l (C_s) and κ_e/κ_l (κ_s) are the volumetric heat capacity and thermal conductivity of the gold electron/lattice (glass substrate), respectively; G is the electron-phonon coupling coefficient; and S_{es} (S_{ls}) is an energy exchange parameter at the gold-glass interface for the gold electrons (lattice). The temperature scalings, parameter values, and further description can be found in Ref. [70]. This model is an extension of the two-temperature model introduced in chapter 2.1.3. The source term $S(\mathbf{r}, t)$ for the energy absorbed from the pump pulse is defined via its Gaussian spatial and temporal profiles. The calculation of the 3D spatio-temporal electron and lattice temperatures $T_e(\mathbf{r}, \Delta t)$ and $T_l(\mathbf{r}, \Delta t)$ for the thin film geometry was first implemented numerically (MATLAB) on a finite mesh. While this approach works, specially designed software to solve differential equations on finite geometries are better suited. Therefore, the final simulations for this work have been performed by our collaborators R. Yu and F.J.G. de Abajo, using COMSOL Multiphysics. The final results of these two-temperature model simulations are the spatio-temporal electron and lattice temperature profiles. For a comparison with the experiment, these have to be converted into the observable $\Delta R/R$. To do this, the change of the optical properties of the gold film with respect to these two temperatures needs to be estimated.

5.3.2 Thermo-optical response

I calculate the spatio-temporal dependence of the gold film permittivity ϵ as a function of both $T_e(\mathbf{r}, \Delta t)$ and $T_l(\mathbf{r}, \Delta t)$, from above, using a Drude model for the near-infrared probe wavelength,

$$\epsilon_{intra} = \epsilon_\infty - \frac{\omega_p^2(T_l)}{\omega(\omega + i\gamma_{re}(T_e, T_l)}. \tag{5.3}$$

Here, ϵ_∞ is the high-frequency permittivity. The plasma frequency ω_p depends on lattice temperature T_l due to volume expansion of the lattice, affecting the free conduction band electron density n_e. Namely, $\omega_p(T_l) = \sqrt{(e^2/\epsilon_0 m_{\text{eff}})\cdot n_e(T_0)/(1+\beta\Delta T_l)}$, where β is the thermal expansion coefficient, $n_e(T_0)$ is the unperturbed density, m_{eff} is the effective electron mass, ϵ_0 is the vacuum permittivity, and e is the elementary charge. γ_{re} represents the total rate of relaxation collisions that conserve momentum and energy of the electron subsystem, given by $\gamma_{re}(T_e, T_l) = \gamma_{e-ph}(T_l) + \gamma_{e-e}^{Um}(T_e)$, where γ_{e-ph} is the electron-phonon collision rate, depending on the lattice temperature as $\gamma_{e-ph}(T_l) = BT_l$, and γ_{e-e}^{Um} is the Umklapp electron-electron collision rate, depending

on T_e as $\gamma_{e-e}^{Um}(T_e) = \Delta^{Um} A T_e^2$. All parameter values and further discussion can be found in Ref. [70]. This analytical model for the explicit dependence of the permittivity on both electron and lattice temperature was provided by our collaborators M. Spector and Y. Sivan.

5.3.3 Differential reflection and transmission

From the temperature dependent permittivity of the gold film, $\epsilon(T_e, T_l)$, I can estimate the spatio-temporal reflectivity maps, using the well known Fresnel formulas for thin films[108]. Figure 5.5 shows the slab geometry used for this calculation.

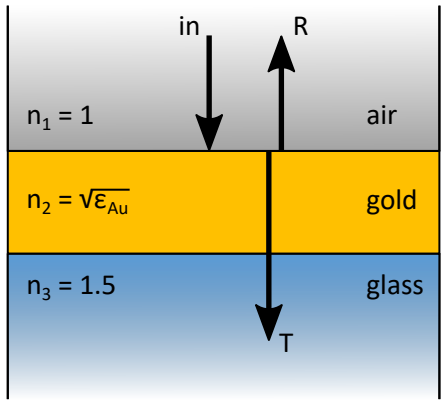

Fig. 5.5: Schematic for the calculation of reflectivity R and transmissivity T in a thin film geometry. The refractive index of air and glass were set to 1 and 1.5, respectively, while the complex refractive index of gold comes from the thermo-optical calculation of ϵ_{Au}.

The air above and the glass substrate below the gold film are modeled as semi-infinite slabs with refractive indices $n_1 = 1$ and $n_3 = 1.5$, respectively, while the complex refractive index of the metal film ($n_2 = n_2' + in_2''$) is taken from the square root of the permittivity calculated from the thermo-optical response above.
The complex reflection and transmission coefficients r and t are calculated via

$$\begin{aligned} r &= \frac{r_{12} + r_{23} e^{2i\beta}}{1 + r_{12} r_{23} e^{2i\beta}}, \\ t &= \frac{t_{12} + t_{23} e^{i\beta}}{1 + r_{12} r_{23} e^{2i\beta}}, \end{aligned} \tag{5.4}$$

with $\beta = \frac{2\pi}{\lambda_0} n_2 h$, where h is the thickness of the film. Then, the reflectivity R and transmissivity T are given by

$$\begin{aligned} R &= |r|^2, \\ T &= \frac{n_3}{n_1} |t|^2. \end{aligned} \tag{5.5}$$

Finally, the differential reflectivity can be calculated by inserting the refractive index $n_2 = \epsilon_{\mathrm{Au}}$ from equation 5.3 for the elevated temperatures,

$$\frac{\Delta R}{R}(T_e, T_l) = \frac{R(T_e, T_l) - R(293\,\mathrm{K}, 293\,\mathrm{K})}{R(293\,\mathrm{K}, 293\,\mathrm{K})}. \tag{5.6}$$

5.3.4 Simulation results

The simulated evolution of the electron and lattice temperatures $T_{e/l}$ at $x = y = z = 0$ is shown in figure 5.6a. A rapid increase in electron temperature to more than 1000 K

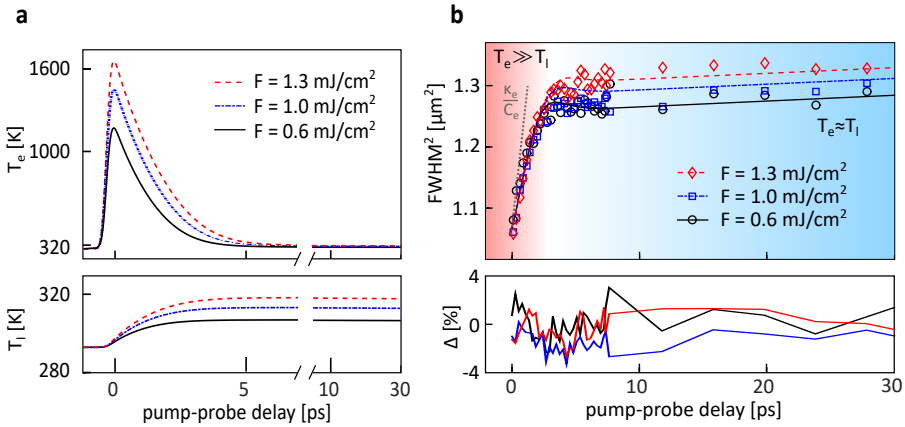

Fig. 5.6: Results of the full modeling. (**a**) Predicted temporal evolution of the electron (top) and lattice (bottom) temperatures at the beam center for the three pump fluences used in the experiment, extracted from finite element method simulations using the two-temperature model. (**b**) Theoretical (curves) and experimental (symbols) evolution of the squared width of $\Delta R/R$ for different pump fluences F (top). For the simulated data, these were obtained by fitting Gaussian curves to the spatio-temporal simulation data, after the full 3D thermo-optical response calculation. In accordance with figure 5.3b, a fast diffusion regime is identified at high electron temperatures, within the first few ps, followed by a thermalized regime (>5 ps) dominated by phonon-limited transport, with orders-of-magnitude lower diffusivity. The initial slope for a diffusivity of $D = \kappa_e/C_e$ is shown for comparison. The bottom panel shows the difference Δ between the data and the full model calculation.

is observed, as well as a subsequent cooling and thermalization with the lattice to a temperature of a few tens of degrees above the ambient level of 293 K.
Figure 5.6b shows the calculated evolution of the squared FWHM together with the experimental results for the three pump fluences under consideration. The comparison shows that the calculated evolution describes the experimental data well for both thermalization regimes, as well as the fluence dependence. The relative difference Δ of the squared width between the full model and the experimental data lies below $\pm 4\,\%$ over the entire range of studied time-delays.

5.4 Discussion

Using scanning ultrafast thermo-modulation microscopy, I have imaged the thermo-optical dynamics of gold films and identified two regimes of thermal diffusion dominated by hot-electrons and lattice modes (phonons), respectively. Intuitively, the two observed regimes of heat diffusion may be understood from limiting cases of the two-temperature model. At early times, when $T_e \gg T_l$, the electron-lattice thermalization time can be estimated as $\tau_{\text{e-ph}} = C_e(T_e)/G$, and the hot-electron-dominated diffusion can be estimated as $D_{\text{fast}} = \kappa_e/C_e$[21]. In the long term, after thermalization of the electrons and the lattice ($T_e \approx T_l$), the two-temperature model simplifies to a diffusion equation with $D_{\text{slow}} = (\kappa_e + \kappa_l)/(C_e + C_l) \approx \kappa_e/C_l$. This is the well-known thermal diffusivity of gold[25]. Inserting the values of electron-phonon coupling constant, heat capacity, and thermal conductivity leads to $\tau_{\text{e-ph}} = 1-3\,\text{ps}$ (for $T_e = 300-1000\,\text{K}$), $D_{\text{fast}} = 152\,\text{cm}^2/\text{s}$ and $D_{\text{slow}} = 1.36\,\text{cm}^2/\text{s}$. However, $D_{\text{fast}} = \kappa_e/C_e$, also shown as a gray dashed line in figure 5.6c, only serves as a preliminary estimation, which overestimates the values observed by our spatio-temporal imaging (see also inset to Fig. 5.3b). Obviously, the delaying effect of electron-phonon coupling on the diffusion dynamics needs to be taken into account. Modeling the full dynamics from absorption through electron-lattice thermalization to spatio-temporal permittivity dynamics allows us to predict the evolution of the spatial width of $\Delta R/R$, resulting in good quantitative agreement with the experimentally determined time-dependent diffusion. In particular, it illustrates how the elevated temperature of the conduction-band electrons correlates with a fast carrier diffusion and how both electron-phonon coupling and hot-electron diffusion contribute to the transition between the two mentioned regimes. I note that the model relies purely on reported material constants and has no free adjustable parameters. In addition, the dependence of the calculated width evolution on the laser fluence agrees well with the experimental data.
In the transition regime, around 5 ps time-delay, a subtle decrease of FWHM is predicted, which lies within the uncertainty of the experimental data. In fact, the two-temperature-model predicts a substantial decrease of FWHM for the electron temperature distribution (i.e., apparent negative diffusion) in this transition regime. The effect is less visible after conversion to the observable, $\Delta R/R$, both experimentally and in simulation, presumably due to the influence of the lattice temperature in the crossover regime.
Recent experiments have studied the effects of non-thermal electron distributions, as well as a transition from ballistic to diffusive electron transport[90,91]. It will be interesting to study these effects in real space with SUTM at higher temporal resolution.

5.5 Conclusion

In summary, I have tracked thermally induced diffusion in a thin gold film from absorption to thermalization in time and space with femtosecond and nanometer resolution. A transition from hot-electron to phonon-limited diffusion is resolved. The electronic and phononic cooling regimes are interpreted with the help of full two-temperature and thermo-optical modeling. The predicted dynamics agree well with the experimental observations. The insight gained on hot-carrier dynamics from direct spatio-

temporal imaging is crucial to understand the interplay of electrons and phonons in ultrafast nanoscale photonics and thus to design nanoscale thermal management in nano-optoelectronic devices. In particular, the control of heat exchange between the electron and lattice systems is important in device functionality. More generally, the excess energy of hot electrons finds applications in an increasingly wide range of systems, such as thermoelectric devices, broadband photo-detectors, efficient solar cells, and even plasmon-enhanced photochemistry.
Here, I have applied my method to gold, the "gold" standard for many such applications of electron heat. Certainly, ultrafast thermo-modulation microscopy is equally suitable to study a vast range of other materials and systems in the future.

Acknowledgments

The experiments reported in this chapter were performed under the supervision of Dr. Matz Liebel and Prof. Dr. Niek F. van Hulst. The experimental results were discussed and interpreted together with Dr. Matz Liebel, Prof. Dr. Niek F. van Hulst, Renwen Yu, Prof. Dr. F. Javier García de Abajo. The 3D finite element calculations were performed by Renwen Yu under the supervision of Prof. Dr. F. Javier García de Abajo. The thermo-optical response theory was provided by Marat Spector and Dr. Yonatan Sivan. The theoretical models were discussed and interpreted together with Dr. Yonatan Sivan, Renwen Yu, Prof. Dr. F. Javier García de Abajo, Dr. Matz Liebel, and Prof. Dr. Niek F. van Hulst.

6

Graphene hot-carrier dynamics tracked with ultrafast all-optical microscopy

In this chapter I study the hot-carrier diffusion dynamics in monolayer graphene. Electronic transport in graphene occurs with an exceptionally high charge carrier mobility, up to 200.000 cm^2/Vs at room temperature. The mobility of a graphene device, however, depends strongly on external parameters, such as temperature, production method, and environment (e.g. substrate). Here, we study the effect of these external parameters on the diffusion of heat in graphene, in particular hot-carrier diffusion. Whereas the charge carrier mobility is typically measured through electrical measurements, we employ an all-optical technique, probe-beam-scanning transient-absorption microscopy, to track, in space and time, electronic heat diffusion in graphene.

6.1 Graphene: Introduction

Graphene is a two-dimensional material, consisting of single layers of carbon atoms in a honeycomb lattice. Since its experimental discovery in 2004, graphene has been characterized both electrically and optically[109–112]. One of the most important observations is that graphene has high electrical conductivity and charge carrier mobility, which can be tuned by electrostatic gating. It is expected that high charge carrier mobility also leads to fast diffusion of heat, i.e., hot-carrier diffusion[42,113]. Importantly, the transport properties of graphene have been shown to be strongly influenced by external parameters. These include the different production types, such as CVD or exfoliation, the environment, such as the substrate, as well as the Fermi level[42].
To test this dependence I study three different samples with different production methods and environments. I present temporally, as well as spatio-temporally resolved transient-absorption dynamics, including their dependence on pump-laser fluence, to study the effect of these external variables on hot-carrier diffusion in monolayer graphene.

6.1.1 Thermal diffusivity expected from carrier mobility

The field of electronic transport in graphene has developed over a decade by now. By determining the electric conductivity at a known carrier density, the electron mobility can be measured (see Eq. 2.14). The reported values for graphene's electron mobility vary over orders of magnitude from 10^3 to $10^5\,\mathrm{cm}^2/\mathrm{Vs}$[114,115]. Higher mobilities have generally been found for suspended, as well as hBN-encapsulated graphene samples[116–118]. In contrast, carrier or heat *diffusivity* is reported far less, and the assumed relation between mobility and diffusivity has not been consistent across the literature.
One of the most cited experimental carrier diffusivities for monolayer graphene comes from one study using PBS-TAM. Ruzicka *et al.* report measurements of graphene hot-electron diffusion coefficients of $D = 11\,000\,\mathrm{cm}^2/\mathrm{s}$ in epitaxial graphene on a Si-terminated 6H-SiC crystalline wafer surface[65] and $D = 5500\,\mathrm{cm}^2/\mathrm{s}$ in CVD graphene on quartz substrates[119]. The authors relate these values to mobility using the Einstein relation for semiconductors (see eq. 2.15), with an extra factor 2, presumably due to electron and hole degeneracy. As this equation includes a temperature, they insert an elevated electron temperature estimated from the photon energy and Fermi-Dirac statistics alone, disregarding the dependence on fluence, as >2000 K. Their estimated mobilities of 70 000 and $60\,000\,\mathrm{cm}^2/\mathrm{Vs}$ are much higher than other reports based on electrical measurements for epitaxial or CVD graphene at room temperature, respectively[118,120], and their results have not been reproduced since.
Further, Chen *et al.* report diffusivities of about $10\,000\,\mathrm{cm}^2/\mathrm{s}$ using a transient grating technique[121]. They calculate mobilities of $120\,000\,\mathrm{cm}^2/\mathrm{Vs}$ with the Einstein relation for semiconductors, yet without the extra factor 2, inserting mean electron temperatures of $650-750\,\mathrm{K}$. In this technique, spatial diffusion information is inferred in an indirect way from time-resolved data.
Rengel and Martín have performed Monte Carlo simulations on carrier diffusion as a function of carrier density[44], substrate[122], and impurity density[123]. Contrary to the previous mentions, they consider the Einstein relation introduced in chapter 2.1.7 for

graphene (Eq. 2.21), which seems to be supported by their Monte Carlo method. However, they calculate a low field diffusivity of 40 000 cm^2/s for suspended graphene, based on their *calculated* mobility of about 500 000 cm^2/Vs, a much higher value than any measured mobilities for room temperature graphene[118]. These reported diffusivities on the order of 10 000 cm^2/s suggest charge carrier mobilities on the order of 100 000 cm^2/Vs. These values are much higher than realistic values for the types of samples that were used, where mobilities below 10 000 cm^2/Vs are expected[118].
Thus, there seems to be a controversy in the literature and this lack of agreement calls for decisive experiments to study the hot carrier diffusion in monolayer graphene, and its dependence on production type, substrate, and environment.

6.2 Experimental Results

6.2.1 Three different samples: Description and imaging

In this study I investigate and compare the diffusion dynamics of three different graphene samples. To get a first look at the samples, I image them with transmission laser confocal microscopy, using $\lambda = 900\,nm$ light. The top row of figure 6.1 shows transmission images of all three samples under study here.

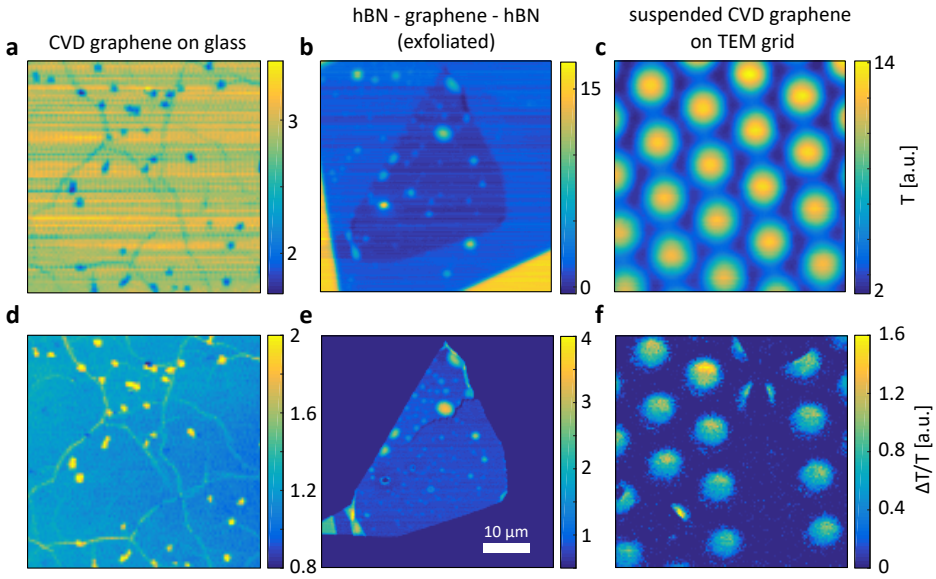

Fig. 6.1: Imaging different graphene samples with transmission microscopy and SS-TAM. The top row shows the transmission microscopy images of (**a**) CVD graphene on glass, (**b**) exfoliated and hBN encapsulated graphene, and (**c**) CVD graphene suspended on a TEM grid. (**d**-**f**) The bottom row shows the corresponding transient-absorption microscopy images of the three samples. The $\Delta T/T$ images shows enhanced contrast of small imperfections (**d** and **e**) and allows for a better distinction of the location of the graphene (**e** and **f**). The 10 µm scale bar shown in **e** is valid for all images.

Figure 6.1a shows a 50 × 50 µm region of the first sample, large-scale commercially

grown CVD graphene (full size: $5 \times 5\,\mathrm{mm}$). The graphene has been transferred from its copper foil substrate to a standard glass cover slip by wet transfer at ICFO.
Figure 6.1b shows the second sample, an exfoliated flake of graphene, which has been encapsulated by thin flakes of exfoliated hBN from both sides, and dry transfered onto a glass cover slip. This sample was fully fabricated at ICFO. The highest transmission contrast comes from the hBN edge, while the region of the graphene can be seen as a slightly darker area.
Figure 6.1c shows the image of the third sample, a commercially available sample with CVD graphene suspended on a transmission electron microscope (TEM) grid, consisting of an array of holes with diameters of $6\,\mu\mathrm{m}$. The transmission image resolves the hole array, while the location of the graphene is not visible on this scale, due to its transparency.
To improve the contrast, I proceed to image the transient absorption $\Delta T/T$ for all three samples via SS-TAM, with $\lambda_{\mathrm{pu}} = 450\,\mathrm{nm}$, $\lambda_{\mathrm{pr}} = 900\,\mathrm{nm}$. The resulting images are shown in the bottom row of figure 6.1.
Figure 6.1d shows the TA signal of the CVD sample supported on glass. Compared to the transmission image, the $\Delta T/T$ shows lower noise levels and a better contrast of the visible round and line-shaped features. As discussed in chapter 4.2, shapes like these have been attributed to wrinkles and polymer residue particles.
For the hBN encapsulated sample, seen in figure 6.1e, the difference in contrast compared with the transmission image is even stronger. Now, a very high contrast between the graphene and the hBN flakes is observed. Again, the features within the graphene have been discussed in chapter 4.2.
Finally, figure 6.1f shows the transient absorption image of the suspended sample. Here, bright and dark areas within the TEM grid holes are visible, which are interpreted as regions covered with graphene and regions with no graphene, respectively.
These transient absorption images were all acquired at a time delay close to time zero ($\Delta t \approx 200\,\mathrm{fs}$), which gives the maximum signal strength. However, the relative signal strengths are not directly comparable due to different pump laser powers. The next section provides a comparison of the temporal evolution at equal powers.

6.2.2 Temporal response

The first step towards the spatio-temporal imaging is to observe the temporal transient-absorption dynamics. Figure 6.2 shows the $\Delta T/T$ signal, recorded at a pump fluence of $0.9\,\mathrm{mJ/cm^2}$, for the three samples under study here.
Both CVD samples (blue and red curves) show an initial rise, limited by the instrument response, followed by a decay within a few picoseconds. The response of the suspended CVD sample is scaled by a factor 0.5 for better comparison to the supported CVD graphene. I observe that the suspended CVD graphene shows a significantly slower decay than the supported sample. This is likely the result of the enhanced coupling to the substrate, providing an additional decay channel for hot carriers. The complex cooling dynamics via various mechanisms involving optical and acoustic phonons in graphene, and substrate phonons is an active field of research[124–131].
The exfoliated and hBN encapsulated sample shows a different temporal response (black curve in figure 6.2). The signal also first rises, but then crosses over to negative $\Delta T/T$ within the first picosecond. Subsequently, the signal decays towards zero on

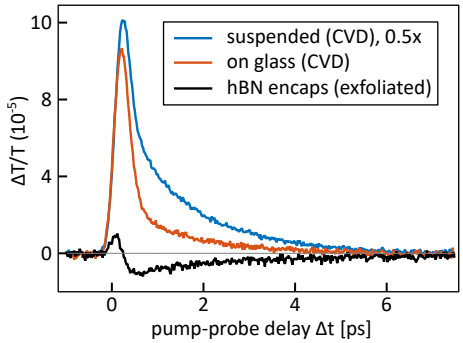

Fig. 6.2: Transient-absorption time traces on the three different graphene samples. The suspended CVD graphene (blue curve) shows the strongest response. Its signal has been multiplied by 0.5 for better visibility. The glass-supported CVD sample (red curve) shows a faster decay than the suspended sample. The hBN encapsulated graphene (black curve) shows more complex dynamics, i.e., a transition from positive to negative signal within the first picosecond.

similar timescales as the suspended sample. The amplitude of the total signal never reaches the levels of the other two samples, possibly due to a partial cancellation of the positive and the negative contribution. The first hypothesis of what is happening here is that the scaling of the observable, $\Delta T/T$, is highly nonlinear with the underlying hot-electron temperature profile that the pump beam creates.
To test this hypothesis, I proceed to measure the TA time traces as a function of pump laser fluence for the hBN encapsulated sample. The results are shown in figure 6.3.

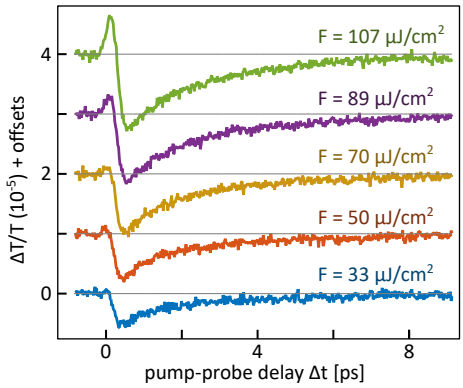

Fig. 6.3: Pump fluence dependence of $\Delta T/T$ time traces for hBN encapsulated graphene. For the lowest pump fluence $F = 33\,\mu J/cm^2$, only a negative peak and subsequent decay is visible. With increased fluence, the emergence of a positive spike within the first picosecond is observed, showing a nonlinear behavior in the scaling of the positive-to-negative ratio. The curves have been offset vertically for visibility.

I observe that for the lowest pump fluence the $\Delta T/T$ response only shows a negative signal amplitude decaying on a few picosecond timescale. Then, with increased

fluence, a positive spike emerges within the first picoseconds, before crossing over to negative $\Delta T/T$ and decaying. These intriguing crossover dynamics show a nonlinear scaling with fluence. For example, the positive signal roughly doubles for a fluence increase from $89-107\,\mu J/cm^2$.
The results here seem to reproduce the trends observed by Malard *et al.* and Chen *et al.*, which have been explained as a combined response of inter- and intraband contributions to the optical conductivity, which is expected to scale linearly with $\Delta T/T$[132,133].
This complex influence of the carrier temperature on the transmission can be understood in terms of the density of states of the graphene electrons. The carrier heating leads to a broadened Fermi-Dirac distribution, similar to the previous description for metal conduction band electrons in chapter 5.1. Therefore, more free carriers are present in both bands of graphene, i.e., more electrons in the conduction band and more holes in the valence band.
Intraband transitions are optical transitions within a band coupled to optical phonons to overcome a momentum mismatch. As more carriers in each band can be promoted by intraband transitions, this leads to enhanced absorption, i.e., negative $\Delta T/T$ response[132,134].
On the other hand, interband transitions are direct transitions between the two bands. Here, the population of states at the probe photon transition energy $\pm\hbar\omega_{pr}/2$ suppresses these transitions*. This phenomenon, known as Pauli blocking, reduces the absorption and leads to a positive $\Delta T/T$ response[132,134]. Malard *et al.* estimate that for electron temperatures between $300-800\,K$, the intraband contribution dominates and the response is linear in T_e. For higher T_e, on the other hand, the interband contribution starts to cause a significant nonlinear behavior and eventually a sign change in the response[132].
Note that the Fermi level is also expected to significantly influence these dynamics. Differently prepared samples are known to have varying Fermi levels, due, e.g., to charge accumulation at the substrate and impurity interfaces.
A possible additional interplay of these transient-absorption dynamics with a so-called "coherent artifact", known from TA-spectroscopy, has not been discussed in the literature, yet might contribute to the dynamics as well. As explained in chapter 3.1, the idea is that TA data within the optical pulse width may be cause by other nonlinear processes, such as four-wave-mixing, as the strict time ordering of pump and probe pulses is not given in this regime.

6.2.3 Spatio-temporal dynamics

To gain more insight into the spatial carrier movement during this hot-electron cooling process, I perform spatio-temporally resolved PBS-TAM to extract diffusion dynamics, as explained before. Figure 6.4 shows a summary of the resulting spatio-temporal datasets for the three different samples.
Figure 6.4a-c shows the spatio-temporally resolved PBS-TAM data for the three samples under study. In all three cases the signal decays temporally within the first few picoseconds. To be able to judge the spatial evolution of the light-induced response

*Note that to cause small absorption changes on the order of 10^{-5}, only the high energy tail of the Fermi-Dirac distribution needs to reach $\hbar\omega_{pr}/2$.

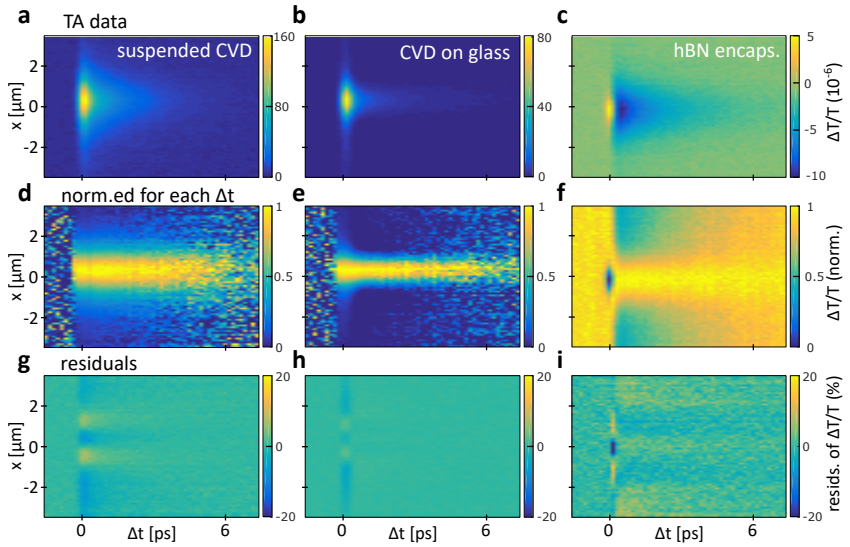

Fig. 6.4: Spatio-temporal datasets collected with PBS-TAM. The top row shows the spatio-temporal $\Delta T/T$ datasets for (**a**) the suspended CVD graphene (**b**) CVD graphene supported on glass, and (**c**) exfoliated and hBN encapsulated graphene. For all cases, the background at $\Delta t < 0$ has been subtracted. In the middle row the spatial profiles have been normalized to the maximum (**d**,**e**) or minimum (**f**) for each pump probe delay Δt. The bottom row shows the residuals of the Gaussian fits to every spatial profile relative to the maximum signal. Both CVD samples (**g** and **h**) show a five-fold residual pattern on the percent level. (**i**) The hBN encapsulated sample shows the strongest residuals of up to 20% in the sign change region.

independent of the decay, the middle row shows the data normalized to the maximum (Fig. 6.4d-e) or the minimum (Fig. 6.4f) value for each time-delay.
To quantify the evolution of the width, I proceed to fit Gaussian functions to each of the spatial profiles as a function of Δt for all three samples. Figure 6.4g-i shows the residual maps of this fitting procedure. While all three samples show a five-fold residual pattern around time zero, the hBN encapsulated sample has the strongest deviation from a Gaussian shape, with residuals up to $\pm 20\,\%$.
Figure 6.5a shows the graph of the squared widths, FWHM^2, as a function of Δt for the three samples under investigation.
All three samples show distinct diffusion dynamics. The suspended CVD graphene sample starts with an initial width (at $\Delta t \approx 0$) of about $2\,\mu\mathrm{m}$ ($= \sqrt{4\,\mu\mathrm{m}^2}$) and shows a roughly linear increase of squared width. The CVD graphene supported on glass starts from a narrower width, closer to $1.4\,\mu\mathrm{m}$, and can be seen to shrink within the first picosecond, followed by barely detectable broadening over 5 ps, while the signal decays. The hBN-encapsulated sample shows even more peculiar behavior. In fact, the width can be seen to decrease, then increase rapidly, then decrease, all within 1 ps, and finally increase with a slope slightly below that of the suspended sample.
The desired measure of the diffusivity should be given by the slope of the curves (see Eq. 2.37). For the suspended sample, the purple dashed line is fit to the data within

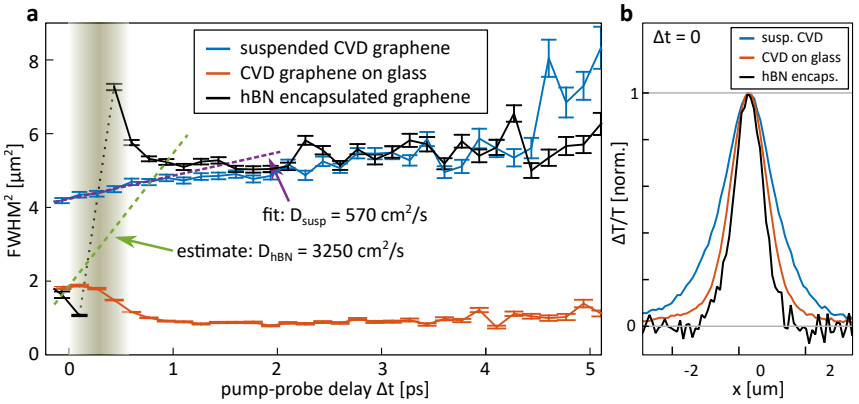

Fig. 6.5: Spatio-temporal width dynamics for the three different samples. (**a**) shows the $\mathrm{FWHM}^2(t)$ evolution, extracted from the Gaussian fits to the data of figure 6.4. The errorbars show the 68 % confidence intervals. The dashed lines are estimated slopes for the suspended and hBN encapsulated graphene. The shaded area highlights the region of high uncertainty for the extraction of diffusion dynamics of the hBN encapsulated sample, due to a "pole" as the $\Delta R/R$ signal changes sign. (**b**) shows the initial ($\Delta t = 0$) spatial profiles for the three different samples.

the first picosecond, showing an estimated diffusivity of about $570\,\mathrm{cm}^2/\mathrm{s}$. For the supported CVD sample, no discernible broadening can be observed. In fact, as the width decreases in the first picosecond, a negative diffusion coefficient would be obtained, which is not physical. Finally, for the hBN encapsulated sample, as the TA signal changes sign around the delay of $\Delta t \approx 0.3\,\mathrm{ps}$, the width dynamics should not be regarded as reporting on the diffusivity in the vicinity of this "pole" region, shown as a shaded region in figure 6.5a. Similar to the simulations of two species with opposite sign shown in chapter 3.2.4, the Gaussian fit close to the pole region shows high confidence, i.e., small errorbars. There, I concluded that this measure does not reflect the true uncertainty for estimating the underlying diffusion dynamics at such a pole region. Excluding the shaded area, I conservatively estimate the diffusivity with the green dashed line to about $3250\,\mathrm{cm}^2/\mathrm{s}$, based on the slope connecting the measured widths at $\Delta t \approx 0$ and $\Delta t \approx 1\,\mathrm{ps}$.

To illustrate the difference in initial width, the spatial profiles at $\Delta t \approx 0$ are shown in figure 6.5b. It is observed that the suspended sample has a much larger width than the other two samples. Note that the samples are measured under the same conditions, such as objective lens, incoming beam collimation, etc. Possible reasons for this discrepancy will be discussed in the discussion (section 6.3).

It should be noted here that one of the basic assumptions to directly probe diffusion with PBS-TAM, elaborated in chapter 3.1.5, seems to be violated here. That is, the relation between the observable $\Delta T/T$ and the underlying carrier profile, in particular the hot-carrier temperature profile ΔT_e, is nonlinear. While this is effect is most visible in the case of the hBN encapsulated sample, where inter- and intraband effects have been used to explain the sign changes, the other two sample types might also suffer from a nonlinear scaling of transient absorption with electron temperature.

Fluence dependent spatio-temporal dynamics of hBN encapsulated graphene

To further investigate the intriguing results suggesting a highly nonlinear response by the hBN-encapsulated sample, I now perform the spatio-temporal imaging as a function of pump fluence. The results are summarized in figure 6.6.
Figure 6.6a shows four spatio-temporal datasets for varying fluences of the pump laser between 58 and 412 $\mu J/cm^2$. To emphasize the influence of the two contributions on the spatial profiles, figure 6.6b shows the same data scaled to show only the negative valued data. Once again, Gaussian fits are performed for all fluences, and their width evolution is shown in figure 6.6c. The spatial profiles for the specific pump-probe delay of ≈ 0.6 ps are displayed in figure 6.6d.
The first observation is that the width evolution shows a pole around $\Delta t \approx 0.3$ ps, as seen before, for all investigated pump fluences. The shapes of the profiles in figure 6.6d show that the dynamics are not described well by a Gaussian function, as the positive and negative contributions cause the shapes to deviate significantly from a Gaussian curve. By looking at the low fluence data, it seems that such a deviation is also present when there is no sign flip. Therefore, the presence of two contributions can cause deviations to the FWHM, even if the signal might look roughly Gaussian and the width evolution should not be taken as a guaranteed measure of diffusion in such a crossover regime. This is similar to what was described in chapter 3.2.4. This is the reason why the widths determined by Gaussian fitting for the hBN encapsulated sample within the shaded region in figure 6.5a should be excluded from the slope fitting method to extract diffusivities. From the spatio-temporal data it is also observed that the two components with opposite sign influence the dynamics at times Δt well above the pulse lengths $\tau_{\text{cross-corr}}$. Therefore, the simple explanation of a coherent artifact, mentioned in section 6.2.2, is excluded as the sole explanation of the observed transition behavior.

6.3 Discussion

As shown in the previous chapters, the technique of PBS-TAM has proven capable of tracking carriers and estimate their diffusion coefficients in silicon and gold samples. Since graphene is known to be one of the materials with the highest carrier mobility, one might have expected to measure very high diffusivities, as previous reports on unencapsulated, supported CVD graphene observed diffusivities up to $10^4\,cm^2/s$[65]. However, none of the studied samples were observed to diffuse faster than about $600\,cm^2/s$ for pump-probe delays above 1.5 ps. The three differently fabricated samples showed very distinct dynamics, showing a large dependence of hot-carrier transport on external parameters, such as production method and environment. In particular, the CVD graphene supported on glass showed nearly no discernible carrier diffusion, while the suspended CVD sample showed moderate carrier diffusion on the order of $600\,cm^2/s$. This follows the general expectation that CVD graphene shows inferior transport properties, typically ascribed to polymer residue. Additionally, supported samples typically show higher scattering from substrate phonons and defects than suspended ones. The hBN encapsulated, exfoliated graphene, known for high mobility, showed a highly nonlinear transient optical response with a sign change around 0.3 ps, causing a “pole” in the graph of the squared width evolution. Although this makes the direct tracking of

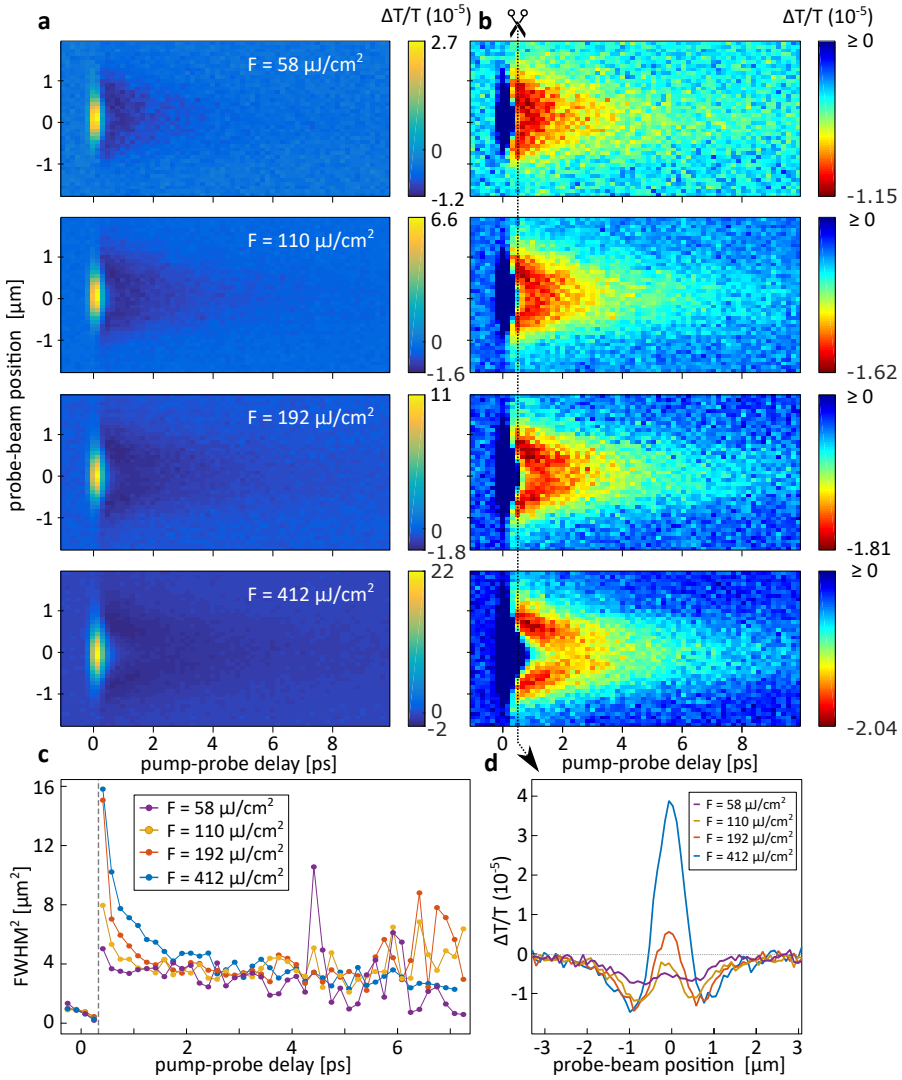

Fig. 6.6: Fluence dependent spatio-temporal transient-absorption microscopy on hBN-encapsulated graphene. (**a**) Spatio-temporal PBS-TAM data for the four studied pump fluences between $58-412\,\mu J/cm^2$. All datasets show a transition form positive to negative signal. (**b**) The same data as in **a**, scaled to show only the negative part. Here, the deviation from a Gaussian shape becomes especially visible. (**c**) The squared width evolution, extracted from **a** via Gaussian fitting. The widths for all fluences show a pole around $\Delta t \approx 0.3\,ps$. (**d**) Cutlines of the spatio-temporal datasets at $\Delta t = 0.6\,ps$. The influence of the positive and negative component on the spatial shape becomes evident, causing a deviation from a Gaussian shape.

the diffusivity more difficult, the data show signs of a significantly higher diffusivity, estimated conservatively to lie around $3000\,cm^2/s$.

A further observation from the data is that the suspended sample showed a larger initial response already at time zero. As the experimental conditions were identical, the possible explanations for such a broader response should reduce. One explanation could be a strong nonlinear relationship from the observable to the underlying spatial profile, scaling differently for the different samples may also change the initial width, although the effect observed here seems too large to be explained by this fact alone. A second explanation could be worse focusing, as the z-position is re-iterated when changing sample. A third interpretation is that there has been substantial broadening of the response within the temporal window of the pulse lengths, $\tau_{\text{cross-corr}} = 250\,\text{fs}$. The focus of the beams onto the sample was achieved by maximizing the $\Delta T/T$ signal while moving the sample in z. This should lead to the minimum spot size on the sample if one considers the final signal as the overlap of the fluence, i.e. power divided by spot size area, of the two beams. Therefore, it seems plausible that there has been a large diffusion occurring within 250 fs, with an increase in FWHM^2 from 2 to 4 μm^2, in accordance with the only other similar work[65]. Calculating $\Delta\text{FWHM}^2/(16(\ln 2)\tau_{\text{cross-corr}})$, a diffusion coefficient can be estimated to be on the order of $7200\,\text{cm}^2/\text{s}$. However, further experiments have been performed to exclude this possibility. Investigating the initial width as a function of the sample height z along the laser focal position, an initial width of around 1.4 µm has been observed with the suspended sample, as well. This suggests the exact focusing conditions to be the most likely cause of the observed effect. Indeed, as I will show in the next chapter (section 7.6.1), the strongest of transient photocurrent signal can come from a z-position away from focus. Although this behavior did not seem to manifest itself in TA, it should be examined more thoroughly in future studies.

Despite the discussed complications, intriguing dynamics were uncovered. At pump-probe delays of more than 1 ps, the upper bounds on hot-carrier diffusion in graphene observed here are in accordance with reports of relative low mobilities in supported, room-temperature graphene. For example, using equation 2.21, a typical mobility of CVD graphene[135] of $\mu = 3000\,\text{cm}^2/\text{Vs}$, and a Fermi energy between $50-400\,\text{meV}$ (i.e. carrier densities $n = 1.8\times10^{11} - 1.2\times10^{13}\,\text{cm}^{-2}$), predicts a diffusivity of $75-600\,\text{cm}^2/\text{s}$.

6.4 Summary

In summary, I have measured the ultrafast optical response, both in the time domain and in space and time, of three differently supported and produced graphene samples to gain insight into hot-carrier diffusion and its dependence on external parameters. Intriguing dynamics were observed, including a highly nonlinear pump laser fluence dependence of the transient absorption signal for hBN encapsulated, exfoliated graphene. A transition from positive to negative transient transmission was observed, which has been hypothesized to come from a competition of inter- and intraband contributions. I have presented upper limits to the hot-carrier diffusion at room temperature for the samples under study, helping to put into context previous reports with very high diffusivity values.

The estimation of diffusivity from electronic mobility and vice versa in graphene has not been been consistent and might have resulted in unrealistically high diffusivity predictions, as mentioned in the introduction to this chapter. In particular, different

versions of the Einstein relation have been employed throughout the literature.
To my knowledge, there has never been a combined, independent measurement of mobility and diffusivity in which the validity of the Einstein relation has been tested. In the following chapter, 7, I present a transient spatio-temporal photocurrent technique on a previously characterized sample in which, for the first time, the diffusivities are estimated in a sample with known mobility and the validity of the Einstein relation is discussed.

Acknowledgments

The experiments reported in this chapter were performed under the supervision of Dr. Matz Liebel and Prof. Dr. Niek F. van Hulst. The results were discussed and interpreted together with Dr. Matz Liebel, Dr. Klaas-Jan Tielrooij, and Prof. Dr. Niek F. van Hulst. The CVD graphene sample wet transfer was done by Gabriele Navickaite, and the hBN encapsulated graphene sample was exfoliated by Mathieu Massicotte, both under the supervision of Prof. Frank Koppens. The CVD graphene on the TEM grid sample was purchased from Ted Pella, Inc.

7

Graphene hot-carrier dynamics tracked with ultrafast photocurrent microscopy

In the previous chapter, I tracked hot carriers in graphene with all-optical transient-absorption microscopy. In particular, I studied the effect of external parameters on the ultrafast transport properties, thereby establishing a link between electron mobility and diffusivity. Here, I take this one step further, by investigating devices with precisely known mobility and Fermi energy and measuring the electronic part of the thermal transport.

To this end, I introduce a novel technique, ultrafast spatio-temporal beam-scanning photocurrent microscopy. This technique measures the photocurrent, generated due to the photothermoelectric effect, in gate-tunable graphene p-n junction devices. Here, the photothermoelectric current directly probes the induced elevated electron temperature above the ambient level. I extract diffusion dynamics, by independently controlling the time delay and spatial offset between two ultrafast pulses and measuring the nonlinear contribution to the photocurrent caused by the interaction of the electronic heat. A simulation based on hot-electron diffusion reproduces many of the observed dynamics for reasonable values of hot-electron diffusivity in a device with known carrier mobilities. Therefore, a direct quantitative comparison between diffusivity and mobility can be established.

7.1 Introduction

The speed at which charge carriers can move through a solid-state device with an applied voltage is typically quantified as the carrier mobility by conductivity measurements. On the other hand, the same carriers are also involved in heat transport. In chapter 6.1.1, I gave an overview of the current status of reported values for carrier diffusivity and electric mobility in the literature. In particular, the relation between the two, i.e., the Einstein relation, has been not used consistently. In chapter 2.1.7, I introduced a simple form of the Einstein relation which I hypothesize should be the one applicable to graphene carriers.
My goal here is to test the validity of this relation for the highly mobile carriers in exfoliated and hBN encapsulated graphene. To this aim, I directly track hot-carrier diffusion with an experimental observable which is sensitive to electron heat in a sample with known carrier mobility.
Here, I present results from two different devices, which have both been characterized in independent electrical measurements. Both devices consist of an exfoliated flake of single layer graphene, encapsulated on both sides by thin layers of hBN and have been contacted for electrical measurements. In the first sample, the graphene is connected to two gold contacts. I call this the TC ("two-contact") sample, which has been described and characterized by Woessner *et al.*[136]. In the second sample, multiple gold contacts connect to the graphene in a hall-bar geometry. I name this the HB ("Hall bar") sample, which has been described by Tielrooij *et al.*[131]. In particular, the HB sample was characterized by four-point probe Hall measurements as a function of electrostatic gate voltage[131]. The resulting graph, showing the carrier mobility μ as a function of carrier density n, is reproduced here in figure 7.1a.

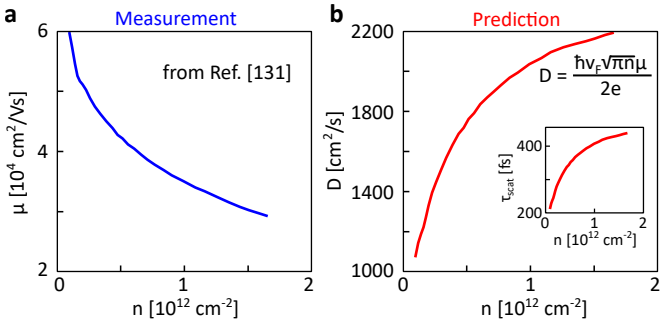

Fig. 7.1: Calculated carrier diffusion dynamics in an electrically characterized device. (**a**) The carrier mobility μ, extracted from a four-probe conductivity measurement as a function of carrier density, from Tielrooij *et al.*[131] (**b**) Estimated diffusivity D from the Einstein relation for graphene (Eq. 2.21, displayed). The inset shows the estimated momentum scattering time τ_{scat}, calculated via equation 2.7 with $v_F \approx 10^6$ m/s.

The sample shows a typical decrease of mobility for higher n, i.e., away from the charge neutrality point. Note that the conductivity σ increases with n, yet sublinearly. Therefore, the mobility decreases, as $\mu = \sigma/(en)$. Now that the mobility of the sample is known, the carrier diffusivity can be predicted. The Einstein relation I have introduced for graphene in equation 2.21, directly relates the carrier mobility to

the diffusivity D and the carrier density n. Figure 7.1b shows the calculated diffusivity for the studied values of the carrier density, assuming a constant Fermi velocity of $v_F = 10^6\,\mathrm{m/s}$. From these values, the mean free path ℓ and the momentum scattering time τ_{scat} can also be calculated using the equations introduced in chapter 2.1.2. The inset to figure 7.1b shows the resulting momentum scattering time, with values of about $250-430\,\mathrm{fs}$ (corresponding to a mean free path of $250-430\,\mathrm{nm}$) for n between 0.2×10^{12} and $1.5\times10^{12}\,\mathrm{cm}^{-2}$.
Now that I have established a prediction for the values of the diffusivity, I aim to track light-induced hot carriers as they diffuse in space and time with an observable that measures electron heat.
In the following sections of this chapter, I describe a number of prerequisite measurements leading up to the final spatio-temporal tracking of ultrafast carrier diffusion.
First of all, I introduce the concept of measuring a photocurrent due to the photothermoelectric effect. Here, a voltage can be generated by an elevated temperature across a junction of two materials with different Seebeck coefficients. Next, I show optical images to characterize the studied devices, which realize such a Seebeck step by differently electrostatically gated areas of graphene. The effect of tuning the gate voltages on the photocurrent is presented, as well as spatial images of the photocurrent from the devices in the p-n configuration.
However, this type of single-laser photocurrent is not temporally resolved. To achieve high temporal resolution the combined effect of two ultrafast laser pulses is studied, thus heating the sample with a variable time delay. A prerequisite to study the difference of two heating sources interacting at the sample is a nonlinear power dependence. Therefore, I present the power dependence of the generated photocurrent, showing a sub-linear behavior. Then, I proceed by introducing the effect of having two ultrafast laser pulses impinging on the devices. I show their temporal dynamics, their combined spatial imaging properties, as well as their intriguing focusing properties. Prepared with all of these details, I tackle the final experiment, studying the combined effect of two heating pulses on the created photocurrent and the dependence on spatial offsets from the junction region, as well as time delay between the two pulses.
Finally, a theoretical model is developed to explain and help to interpret the observed trends, finally allowing a quantification of hot-carrier diffusion and comparison to the predicted dynamics.

7.2 Measuring electron temperature with the photothermoelectric effect

As mentioned before, the carrier density, and therefore the Fermi level of graphene can be tuned by electrostatic gating. In analogy to semiconductors, graphene with an increased electron or hole carrier density is called n- or p-doped, respectively.
Here, I use the photothermoelectric (PTE) effect in graphene p-n junction devices to track hot-carrier diffusion.
The basic idea is to generate a photovoltage across a p-n junction, i.e., a flake of graphene which is electrostatically gated differently on opposite sides of a narrow gap region. A sketch of this is shown in figure 7.2.
Figure 7.2a shows a schematic graphene device with two distinctly electrostatically

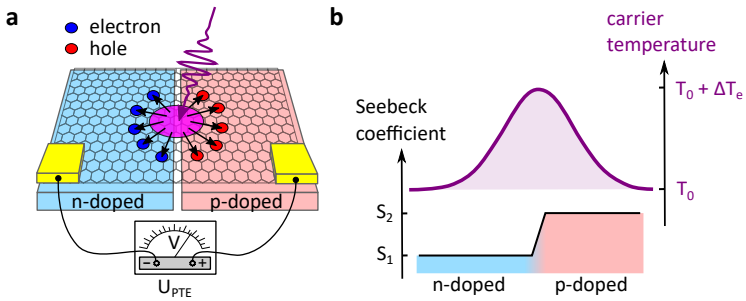

Fig. 7.2: Generation of a photothermoelectric voltage in a graphene p-n junction. (**a**) Schematic device, with an n-doped region shaded blue on the left and a p-doped region shaded red on the right. After illumination with a light pulse, hot electrons and holes diffuse away from the heat source, creating a local photovoltage, which can be measured between two contacts. (**b**) Sketch depicting the spatial profile of the Seebeck coefficient step across the junction, together with the elevated carrier temperature profile. This geometry causes the photothermoelectric voltage in a macroscopic picture of the thermoelectric effect.

gated regions. The two regions are characterized by different Seebeck coefficients, S_1 and S_2. A pulsed laser impinging on the junction region causes an electron temperature rise with peak amplitude ΔT_e with respect to the non-excited region. Figure 7.2b schematically depicts the spatial variation of the Seebeck coefficient and the carrier temperature after pulsed laser illumination. The electron temperature can rise to values on the order of 1000 K, while the rest of the sample stays around ambient temperature $T_0 = 293\,\mathrm{K}$. As measured in the previous chapter, electrons and phonons thermalize on a timescale of a few picoseconds. A voltage U_{PTE} is created due to the photothermoelectric effect. The voltage reads[22,75,137]

$$U_{\mathrm{PTE}} = (S_2 - S_1)\Delta T_e. \tag{7.1}$$

To achieve such a difference in Seebeck coefficients across a small gap region, independent gate voltages can be applied in a split-gate-capacitor type device, as electrostatic gating directly influences the Seebeck coefficient in graphene according to the Mott-Schottky equation[22].
This concept has recently been used to image photocurrent (PC) on hybrid organic-inorganic perovskite[138], as well as graphene devices[75,139]. While these powerful techniques can give spatial information about the generated photocurrent, they lack the simultaneous *spatio-temporal* resolution to track carrier dynamics.
Here, I define the photocurrent $\mathrm{PC} = U_{\mathrm{PTE}}/R$ as the current flowing due to the local photothermoelectric voltage U_{PTE} through a device with total resistivity R, which includes the intrinsic (channel) resistance of the graphene, as well as the device's contact resistances.

7.3 Optical sample characterization

To study the spatial variations of the PTE response of these microscopic devices, I first image them optically. Figure 7.3a shows a white light microscopy image of the TC sample, adapted from Woessner *et al.*[136]

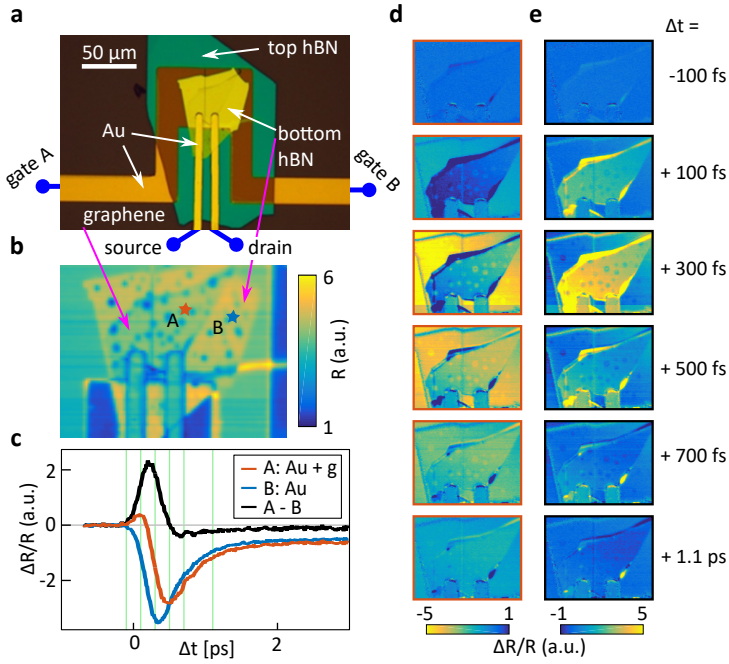

Fig. 7.3: All-optical characterization of a split gate device. (**a**) Optical image of the TC split-gate device sample (from Woessner *et al.*[136]), showing the relevant materials and contact nomenclature. (**b**) Confocal laser scanning reflection map, showing the border between the graphene and bottom hBN. (**c**) The red and blue curves show temporal scans on and off the graphene, at points A and B from **b**, respectively. Both show a mainly negative transient reflection signal. After subtraction of the two curves, the graphene response is revealed, with a large positive spike followed by a slight negative recovery. (**d**) $\Delta R/R$ images recorded at different pump probe delays, indicated by the green lines in **c**. The total signal is shown. (**e**) shows the same signal as **d** with the average signal of the area "B" subtracted. Here, the positive response of the graphene peaks around $\Delta t = 300\,\mathrm{fs}$.

The graphene is barely visible between the top and bottom hBN flakes. The latter can be seen as green and yellow regions, respectively. The two gold back gate contacts, labeled gate A and B, are separated by a small gap underneath the graphene. They are separated vertically from the graphene by the bottom hBN, and therefore act as capacitive electrostatic gates. The other two gold contacts act as the drain and source contacts and directly touch the encapsulated graphene.
Figure 7.3b shows a close-up of the graphene flake imaged with confocal laser scanning reflection microscopy. Here, the region with and without the graphene flake can be distinguished. To enhance this contrast, I take transient-absorption time traces at two

regions on and off the graphene flake, marked A and B, respectively. Figure 7.3c shows the $\Delta R/R$ traces taken at points A and B indicated in figure 7.3b. The TA response differs from the data presented in the previous chapter, as both the gold gate contact and the silicon substrate contribute to the total signal. To isolate the response of the graphene, I subtract the two signals, resulting in the black curve. This response can be seen as a positive peak, at early times and then transitioning to a negative signal at later times, while the signal off the graphene shows mainly a negative peak in intensity and decaying on a slower time scale.
Similarly, SS-TAM can be used to image the devices and get very clear contrast between the involved materials, as seen in figure 7.3d. Here, the transient reflection maps on the TC sample is shown for different time delays, $\Delta t = -0.1$ to 1.1 ps, indicated by the green vertical lines in figure 7.3c. The signal is dominated by a negative response of the region away from the graphene, decaying over a picosecond timescale.
To focus on the response of the graphene, as before, I subtract the signal away from the graphene for all time delays, resulting in the maps shown in figure 7.3e. Here, the graphene flake shows a peak response at $\Delta t \approx 300$ fs and decays quickly. Thus, I have established a way to image the graphene area with high contrast by temporal selection of transient-absorption microscopy in preparation of the photocurrent measurements.

7.4 Single laser photocurrent

The PTE photocurrent is measured via lock-in amplification. The pulsed laser is modulated with an optical chopper at hundreds of hertz. The weak photocurrent signal on the order of nanoamperes is then measured by demodulation of the amplified current (at 10^6 V/A gain) across the source and drain contacts through the graphene sheet (see Fig. 7.4a).

7.4.1 Gate dependence

The dependence of photocurrent across a graphene p-n junction on applied gate voltage typically shows a characteristic six-fold pattern, which indicates the thermoelectric effect as the source of the photocurrent[140]. This response has been observed in both devices studied here[131,136]. Importantly, this directly links an experimental observable, the PTE photocurrent, to the underlying photo-induced electronic heat ΔT_e, the quantity of interest.
Before starting the photocurrent measurement, I find the charge neutrality point of the samples under study by an all-electrical resistivity measurement. Figure 7.4a shows the wiring scheme for the TC device.
The split-gate electrostatic contacts, which are isolated from the graphene flake by the hBN layer, forming a capacitor, are labeled gate A and B. The other two contacts, which actually touch the graphene at opposite sides of the junction, are called source and drain.
Figure 7.4b shows the all-electrical characterization of the TC device for this work for one example gate voltage scan. The source-drain current is monitored for a source voltage of 2 mV as a function of the voltage applied to gate A. The minimum current is found at a voltage of about 0.45 V for gate A, indicating the charge neutrality point, as it is known that graphene's resistance is maximum here[109]. The same measurement

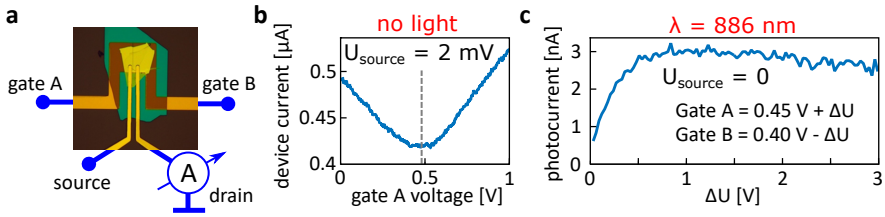

Fig. 7.4: Electrical and photocurrent device characterization. (**a**) Optical image of the TC device, with contact name labels. For the conductivity measurement of the graphene, a source voltage of 2 mV is applied. The current is measured between the source and drain, which directly touch the graphene, as a function of the electrostatic gate voltages, insulated from the graphene by the hBN. (**b**) Device conductivity as a function of voltage on gate A. The charge neutral point, i.e., the highest resistance, is found around 0.45 V. The same measurement is done for gate B (not shown). (**c**) Photocurrent is measured with NIR light impinging on the junction region. This time, the source is grounded and the current is measured as a function of p-n junction voltage ΔU, i.e., symmetrically sweeping away from the charge neutrality point with both gate voltages.

is then repeated for the gate B, as well, with a minimum at about 0.4 V (not shown). Once the charge neutrality point is known, both gates can be tuned symmetrically away from charge neutrality (in opposite directions) to establish the p-n junction.
I proceed to measure the photocurrent response, illuminating the split-gate area with the NIR laser ($\lambda = 886\,\mathrm{nm}$) as a function of the extra voltage ΔU, applied symmetrically to both gates. Figure 7.4c shows the resulting photocurrent, as it rises to its maximum value of about 3 nA for $\Delta U \approx 1\,\mathrm{V}$. With a measured total device resistance of $R = 3.4\,\mathrm{k\Omega}$, this corresponds to a local PTE photovoltage of $U_{\mathrm{PTE}} \approx 10.2\,\mu\mathrm{V}$.
The behavior is in accordance with the previous characterization, showing a maximum photocurrent at about 1 V away from charge neutrality[136].

7.4.2 Photocurrent imaging

To find out at what regions of the device the photovoltage is created, I collect the photocurrent as a function of sample position with respect to the laser focus, thereby creating, point-by-point, a spatially resolved photocurrent map. Figure 7.5a shows a zoom of the part of the device where the graphene is located.
A black contour has been overlaid with the PC image, showing the location of the graphene flake, as well as the metal contacts. This is done by comparison to images of the sample fabrication. Figure 7.5b shows the reflectivity of the sample which is acquired at the same time as the photocurrent, by collecting the reflected light on a photodiode.
Now it becomes evident that the photocurrent of about 5 nA is created along the part of the p-n junction region that has the graphene flake directly above. Additional photocurrent with the opposite polarity is also created at the interface between the right gold contact and the graphene flake, while much less current is created at the left contact. This type of photocurrent at graphene-metal interfaces has also been explained as arising due to the PTE effect, where the difference in Seebeck coefficients between the metal-contacted graphene are considered[141]. The asymmetry between

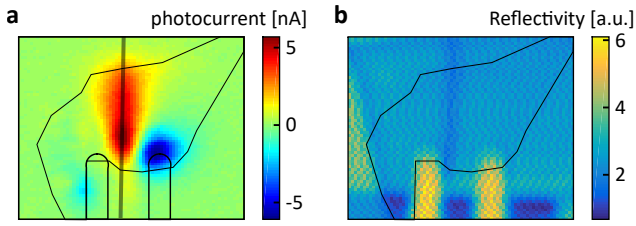

Fig. 7.5: Single laser photocurrent imaging. (**a**) Photocurrent image of the split gate region of the TC device, with $\lambda = 886\,\text{nm}$. The black contour has been overlaid to show the graphene flake and source and drain contact locations. The photocurrent is most efficiently created along the split-gate region, as well as at the graphene-metal contact interface, yet with opposite polarity. (**b**) Scanning optical reflectivity map, acquired simultaneously by recording the reflected light on a photodiode. Here, the main contrast comes from the gold contact regions, yet the gap defining the split-gate can also be seen as a vertical darker line. The two images together show that the *PC* is created exactly at the region, where the graphene overlaps with the p-n junction.

the two contacts is expected to come from a different Seebeck coefficient step from the gated regions with respect to the metal contacted graphene due to Fermi level pinning[141].

7.4.3 Power dependence

As described by Tielrooij *et al.*, the photocurrent response of graphene rises sublinearly, i.e. nonlinearly, with input laser power[131]. As I explain in more detail in the next section, this fact allows the observation of decreased photocurrent in a two-pulse experiment for temporally overlapping beams, and the extraction of temporal, and spatio-temporal, dynamics. Therefore, I study the power dependence of the photocurrent for the TC device imaged in figure 7.5. The results for illumination with the $\lambda = 886\,\text{nm}$ laser are summarized in figure 7.6.

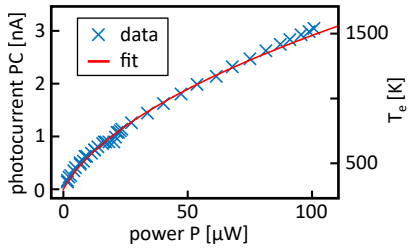

Fig. 7.6: Power dependence of the measured photocurrent. The photocurrent PC, as a function of time-averaged power P of the NIR pulsed laser with $\lambda = 886\,\text{nm}$ (left y-axis, symbols). A sublinear scaling is observed. The red line is a fit to the data with a thermal model (Eq. 7.2), allowing for an estimation of the corresponding electron temperatures, shown on the right y-axis.

Following the treatment of Tielrooij *et al.*, the PTE photovoltage is assumed to be

proportional to the time-averaged increase of the electronic temperature T_e above the ambient temperature T_0[131]. The generated photocurrent is collected over an integrated time window of many milliseconds. With the repetition rate of the laser used here, pulses arrive at the sample every 13 ns, creating high electron temperatures T_e, which decay within a few ps. This extra electronic heat causes the PTE signal, albeit with a very low "duty cycle". By assuming a linear relationship between average and peak increase of T_e, the photocurrent will scale proportional to $T_e - T_0$ as well. The peak T_e as a function of the laser power P scales as[131]

$$T_e = \sqrt[2]{T_0^2 + bP}. \tag{7.2}$$

I fit the data presented in figure 7.6 to the curve $\mathrm{PC}(P) = a(\sqrt[2]{T_0^2 + bP} - T_0)$, where T_0 is set to 293 K. a and b are the only fit parameters. The resulting curve (with $b = 2.1 \times 10^4\,\mathrm{K}^2/\mu\mathrm{W}$) allows us to convert the observed photocurrent into peak electronic temperature T_e (right axis of Fig. 7.6). Note that the exponent $n = 2$ of equation 7.2 comes from an assumed linear temperature scaling law of the electronic heat capacity for highly doped graphene*. For very weakly doped graphene, an exponent of $n = 3$ has been predicted[131]. This model can also be fit satisfyingly to the data, but previous work on these samples have shown that gate voltages on the order of ± 1 V correspond to the high doping regime[131]. The extracted nonlinearity will also be used for the modeling in this chapter.
The relationship between peak carrier temperature T_e and heating power, given in equation 7.2, was derived from a linear scaling of the electronic heat capacity. This is assumed to hold as long as $k_B T_e < E_F$ (due to the Sommerfeld approximation mentioned in chapter 2.1.7). In the device used here an applied voltage of 1 V corresponds to about $E_F = 70$ meV[136], while the highest estimated temperatures here, $T_e = 1500$ K, correspond to 130 meV. Therefore, slight variations to the scaling law could be expected for the highest fluences. Indeed, the data points for high fluence shown in figure 7.6 deviate most from the fit function. More work is needed to study the hot-carrier dynamics in the region $k_B T_e > E_F$, which is accessible with ultrafast laser pulses and gate-tunable devices.

7.5 Dual laser photocurrent

Next, I study the effect of having two femtosecond laser pulses interacting at the p-n junction, as described before[75]. Due to the nonlinearity of the induced peak electron temperature with laser power, the simultaneous heating of two pulses arriving at the sample at the same time should cause less time-averaged photocurrent than the situation where both pulses are separated temporally by more than the cooling time. This situation is depicted in figure 7.7.
Simply said, two pulses causing heat energy increases ΔQ_1 and ΔQ_2 separated by, say, more than 10 ps ($\Delta t \gg 0$) should contribute independently to the total measured

*In detail, the relation assumes that the power P is proportional to the absorbed heat energy ΔQ, and that the heat capacity scales as $C_e(T) = \gamma T$. Then, by integrating $dQ = C_e dT$, we get $\int_{Q_0}^{Q_0+\Delta Q} dQ = \int_{T_0}^{T_e} \gamma T dT$, from which equation 7.2 is obtained by setting $bP = 2\Delta Q/\gamma$.

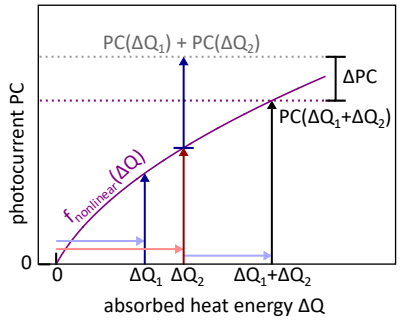

Fig. 7.7: Origin of the decreased photocurrent ΔPC. Due to the sublinear scaling of the photocurrent PC with the absorbed heat energy ΔQ, two independent pulses create more photocurrent than two pulses which overlap in space and time, i.e., act as a single pulse with higher fluence. Therefore, the difference, ΔPC, is a measure for the spatio-temporal overlap of electronic heat.

photocurrent, as

$$\mathrm{PC}^{\mathrm{tot}}_{\Delta t \gg 0} = \mathrm{PC}(\Delta Q_1) + \mathrm{PC}(\Delta Q_2). \tag{7.3}$$

Yet, when both beams overlap in time and space, the sublinear relationship of $\mathrm{PC}(P)$ will cause less photocurrent, therefore

$$\begin{aligned} \mathrm{PC}^{\mathrm{tot}}_{\Delta t \approx 0} &= \mathrm{PC}(P_1 + P_2) \\ &< \mathrm{PC}^{\mathrm{tot}}_{\Delta t \gg 0}. \end{aligned} \tag{7.4}$$

To isolate this decrease in photocurrent $\Delta\mathrm{PC} = \mathrm{PC}^{\mathrm{tot}}_{\Delta t \gg 0} - \mathrm{PC}^{\mathrm{tot}}_{\Delta t}$, a double chopper technique is introduced, as explained in the next section.

7.5.1 Temporal dynamics

The effect of the interaction of two ultrashort heating pulses on the photocurrent is tested by measuring the photocurrent as a function of the time delay between two pulses. To distinguish the two, I name them "pump" and "probe" in analogy to transient absorption. More accurately, this class of experiments could be called "pump-pump-current-probe".

Until now, the single beam photocurrent was measured by modulating the laser with an optical chopper and demodulating the current with a lock-in amplifier. Now, to study the effect of two pulses, both beams are modulated individually with optical choppers, as depicted in figure 7.8a. The photocurrent can be measured by demodulating the current signal at one of the two chopper frequencies as a function of pulse separation. The top black curve of figure 7.8c shows the result of such a measurement.

The resulting trace shows a dip in the photocurrent when the two pulses overlap temporally, i.e., around $\Delta t = 0$. In other words, when the time-delay of the pulses is on the order of a few picoseconds, the nonlinear contribution of the second pulse leads to less total photocurrent, as described in figure 7.7 and equation 7.4. Yet, this signal still suffers from low frequency noise, as one of the two beams intensity fluctuations still directly influence the data. Additionally, the dynamics of interest, i.e., the dip

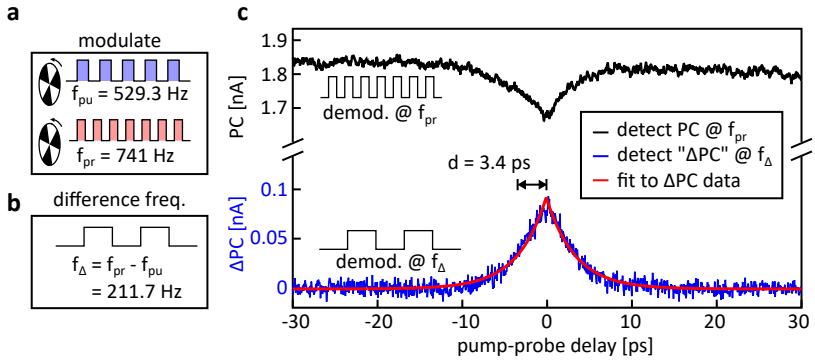

Fig. 7.8: Temporal photocurrent dynamics in a double-pulse experiment. (**a**) Both beams are individually modulated with optical choppers at frequencies f_{pu} and f_{pr}, respectively. (**b**) The difference frequency signal, a square wave at $f_\Delta = f_{pr} - f_{pu}$, is also created electronically as a possible demodulation frequency. (**c**) The photocurrent PC, demodulated at the chopper frequency of the probe, f_{pr}, is shown as the black trace. The blue curve shows the differential photocurrent ΔPC, i.e., the nonlinear effect of both heating responses interacting, isolated by demodulating at the difference frequency f_Δ. The red line is a fit to the data according to equation 7.5.

in the photocurrent signal, is not being isolated from the "background" PC, making it harder to maximize this contribution experimentally. Both of these problems can be solved by demodulating the current signal at the difference frequency of the two modulation frequencies (see Fig. 7.8b), essentially Fourier filtering the signal caused by both heating sources[142]. To avoid interference of harmonics, I choose the prime integer ratio of 7 to 5. By choosing $f_{pr} = 741\,\text{Hz}$ and $f_{pu} = 529.3\,\text{Hz}$, a difference frequency of $f_\Delta = 211.7\,\text{Hz}$ is obtained. I choose a Newport chopper driver with a setting for a 5/7 chopping wheel which also has a trigger-signal output at this difference frequency. Now the lock-in amplifier demodulates at this frequency and therefore isolates the effect of both beams, i.e., the signal of interest, ΔPC, and avoids low frequency noise. The blue, lower curve in figure 7.8c shows this signal. The sign of the signal is set to be positive by the choice of lock-in phase. This "background-free" signal is very useful for the further analysis as well as experimental optimization, as a better spatio-temporal overlap now leads to a signal rise above zero offset.
The data are fit to an empirical model of the form

$$\Delta\text{PC}(\Delta t) = a \exp\left(-\frac{|\Delta t - t_0|}{d}\right), \tag{7.5}$$

where the two key parameters are the amplitude a and an exponential decay time, which I call diversity d. The name comes from statistics, where this curve is known as the probability density function of the Laplace distribution. The result of the fit is shown as the red curve in figure 7.8. It gives a diversity of $d = 3.4\,\text{ps}$, in accordance with previously measured hot-carrier cooling times[131]. In principle, a convolution with the laser pulse width ($\tau_{\text{cross-corr}} = 0.2\,\text{ps}$) should be added as an instrument response function to equation 7.5, leading to a slight blurring of the sharp peak at $\Delta t = 0$. However, as this time scale is an order of magnitude faster than the decay, it does not

significantly change the observed trends and was therefore omitted from the analysis.

7.5.2 Imaging the dual-pulse decreased photocurrent

As I am interested in the spatio-temporal dynamics of the hot carriers, which are assumed to diffuse as they cool, I proceed to image the device with the double-chopping technique, as it improves the signal-to-noise ratio for the differential signal and removes the background offset.

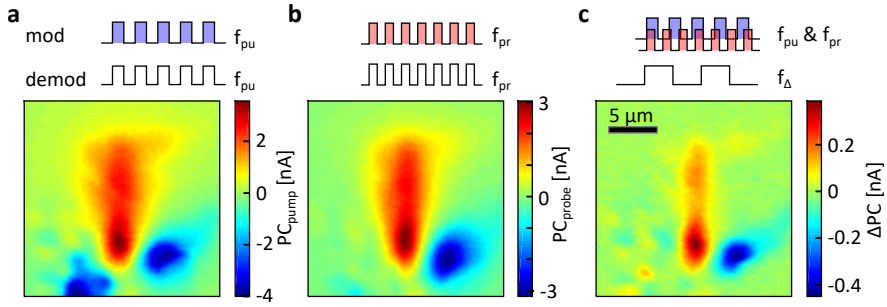

Fig. 7.9: Photocurrent maps in a double pulse experiment. (**a**,**b**) Photocurrent maps of the TC device (as shown in Fig. 7.5) when illuminated with the (**a**) pump and (**b**) probe beam, modulated and demodulated at f_{pu} and f_{pr}, respectively. The spatial overlap of both beams, at $\lambda_{pu} = 443\,\text{nm}$ and $\lambda_{pr} = 886\,\text{nm}$, has been optimized. This can be seen in the similar positions of the positive PC signal along the junction. The negative lobe at the left gold contact shows up more clearly in the blue laser PC imaging channel. (**c**) Nonlinear decreased two-pulse photocurrent ΔPC along the junction at $\Delta t \approx 0$. The signal is now demodulated at f_{Δ}. The sign of ΔPC at the junction is positive, due to choice of lock-in phase, as explained in section 7.5.1.

Figure 7.9a shows the photocurrent measured on the TC sample with the pump beam only. As before, the PC signal along the junction is resolved, as well as some additional PC at the gold contact areas with the opposite polarity.
Figure 7.9b shows the same area imaged with the probe beam only, for the same sample stage movement as before. The PC response can be seen to come from the same region, as the sample is scanned. This is the result of precise alignment of the two beams with respect to each other with the scanning galvo mirrors.
After both beams have been overlapped spatially at the sample plane, I proceed to image the differential photocurrent ΔPC, by demodulating the signal at f_{Δ}. The resulting map is shown in figure 7.9c.
Some information about the time-averaged cooling length can already be extracted from the single laser photocurrent images, such as figure 7.9a and b. I observe a "steady state" width of the PC response of about 4 µm at the 100 nm narrow junction. This is significantly larger than the convolution of the beam width for with the active area. This observation reflects the total cooling length, i.e., the amount of diffusion that can occur before the carriers have cooled to ambient temperature[131]. The cooling length can be estimated from the carrier diffusivity D and the carrier cooling time τ_{cool} as

$$L_{\text{cool}} = \sqrt{D\tau_{\text{cool}}}, \tag{7.6}$$

as introduced in chapter 2.1.6. In accordance with this picture, the ΔPC map (Fig. 7.9c) shows a narrower response, as this signal is created only within the picosecond decay. However, the extraction of a diffusion coefficient from this method alone relies on the knowledge of the exact beam size at the focal plane. As I discuss in the next section, the focusing of the beams onto the sample, in particular by maximizing the PC signal is not straightforward and needs careful consideration. Furthermore, the effects of carrier mobility and cooling dynamics cannot be disentangled this way. Therefore, an ultrafast beam scanning spatio-temporal technique, similar to PBS-TAM is desired, to track hot-carrier diffusion directly.

7.6 Nonlinear spatio-temporal dynamics

To gain direct spatio-temporal information about hot-carrier dynamics on the femtosecond and nanometer scales, I proceed to measure ΔPC as a function of pump-probe time delay as well as spatial offset, in the spirit of the all-optical probe-beam scanning TAM described in the previous chapters. As the PTE signal comes from the electron temperature across the junction, i.e., across the Seebeck coefficient step, it should be possible to trace spatial diffusion of electron heat from beams which are spatially offset from this junction region.

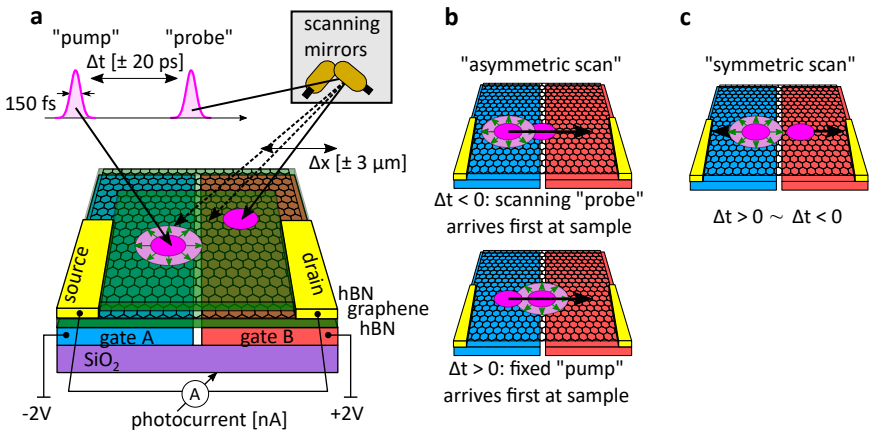

Fig. 7.10: Schematic of the ultrafast spatio-temporal scanning photocurrent microscopy experiment. (**a**) Sketch of the two-pulse spatio-temporal photocurrent mapping in a split-gate junction sample. Both beams' positions, their relative pulse delay, as well as the gate voltages can be individually controlled. The observable is the source-drain photocurrent. (**b**) Spatially asymmetric scanning mode, split into two parts, to show the difference between $\Delta t < 0$ and $\Delta t > 0$ when interpreting the spatio-temporal data. For $\Delta t < 0$ (top), the energy distribution created by the spatially offset "probe" heating pulse has had time to diffuse before the "pump" arrives at the junction. For $\Delta t > 0$ (bottom) the interaction of the probe and pump heating sources should intuitively be less effective at creating ΔPC, as the heat created by the offset probe arrives at the junction after much of the pump heat has decayed. (**c**) Spatially symmetric scanning away from the junction. Here, positive and negative delays Δt are equivalent.

Figure 7.10a shows the schematic of such a measurement. As before, the split-gate device is independently doped by two gates, to create a p-n junction across a narrow gap and the photocurrent is monitored between the source and drain contacts. The position of the "pump" beam is controlled by sample nano-positioning with a piezo stage, while the "probe" can be independently scanned with the galvo mirrors. The nonlinear decrease in photocurrent ΔPC can now be measured as a function of pump-probe time delay, both gate voltages, as well as spatial offset of the pump and probe beam with respect to each other and the p-n junction region.
Figure 7.10b depicts one of the possible spatial scanning modes, i.e., asymmetric scanning with one beam centered on the junction while the other beam is displaced. Here, the difference between negative and positive time delay is explained by showing the two cases of either beam arriving first at the sample. This is shown as a spatial broadening of one of the two beams which has "had time" to diffuse before the other beam arrives at the sample. Figure 7.10c shows a symmetric scanning mode with both beams being displaced simultaneously away from the junction region. For this case, if the two beams are identical, there should be no difference between negative and positive time delay.

7.6.1 Focusing properties

The first observation made by this technique is an anomalous scaling of the signal level when focusing the two beams on the sample.
Intuitively, it could be expected that the two beams, which have been profiled at the sample plane (see section 4.1.2), should lead to the strongest ΔPC signal when the fluence is highest, i.e., when the spots are focused to the smallest size, as long as the spot sizes are big enough to impinge simultaneously on the p- and n-doped sides of the device.
To test this, I perform scans with the pump beam fixed at the center of the junction and the probe beam scanning a 2D area over the pump location at $\Delta t = 0$ as a function of the z-position of the sample, i.e., scanning the sample through the focal plane of the two beams. The results are summarized in figure 7.11.
Figure 7.11a shows ΔPC datasets for varying z-positions of the device, recorded with a fixed pump beam at the center of the p-n junction and the probe beam scanning in two dimensions. The u axis is parallel to the junction axis, while the v-axis lies perpendicular. The resulting spatial signal maps look similar to the all-optical images collected with PBS-TAM, yet somewhat elongated along the junction (u-) axis.
To quantify the spatial extent and its dependence on the sample position relative to the beam focus, I analyze the average of the signal of the central few lines in both dimensions, resulting in the u-z and v-z datasets shown in figure 7.11b. The first striking observation is that the signal seems to be minimum at the waist of the beams. I proceed to quantify this effect by extracting the width and signal strength from these data.
Figure 7.11d shows the FWHM of the ΔPC profiles, extracted from 7.11b and c by Gaussian fitting. For both dimensions, the FWHM goes through a minimum around the same z position, labeled as $z = 0$. The ΔPC profiles along the u-axis show a larger extent compared to those along the v-axis as the photocurrent can be created more efficiently along the p-n junction than when moving away from it.

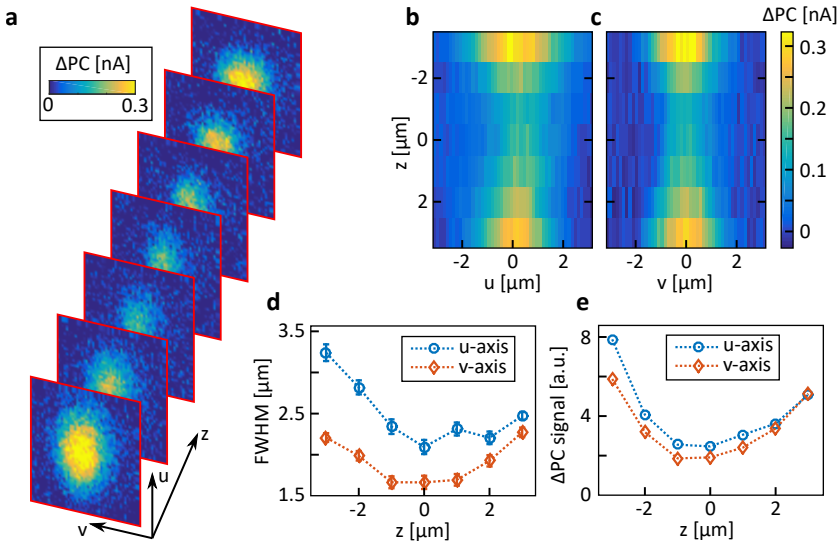

Fig. 7.11: Focusing properties of spatio-temporal photocurrent imaging. (**a**) ΔPC maps of the TC device for varying z-positions. The data is collected with one beam scanning over the other beam, which is centered at the split-gate junction. The u- and the the v-axis are parallel and perpendicular to the junction, respectively. The pump-probe delay is $\Delta t \approx 0$. The resulting maps are elongated in the u-direction. (**b**,**c**) Averaged signal from **a** projected onto the u- and v-axis, respectively. (**d**) FWHM of the ΔPC profile as a function of z, estimated by Gaussian fitting to the profiles shown in b and c. The errorbars show the 68 % confidence intervals of the Gaussian fits. The focus, i.e, minimum beam waist, is at $z \approx 0$. (**e**) Extracted signal strength as the sample is scanned through the focus. Surprisingly, the best focusing position corresponds to a minimum in signal strength.

One might have expected the signal of ΔPC to be maximum at this z-position, as the laser heating power is concentrated to the smallest volume, creating the maximum electron heat and therefore also the maximum nonlinear PC decrease. Figure 7.11e shows the ΔPC signal level as a function of z for both axes. It can be seen that the signal level is higher away from $z = 0$. This means, that there is an optimum photocurrent created when the beams have a size larger than the best focusing conditions (when $\mathrm{FWHM}_{\mathrm{pu/pr}} \lesssim 1\,\mu\mathrm{m}$). For even larger displacements ($z > 3\,\mu\mathrm{m}$) the signal decreases again (not shown). A possible explanation for this behavior is that there is an optimum beam size, when the hot carrier cooling length roughly matches the beam size, such that most of the carriers still interact with each other and don't leave the area of overlap "too efficiently". By estimating $\tau_{\mathrm{cool}} = 2\,\mathrm{ps}$ and D between 2000 and $4000\,\mathrm{cm}^2/\mathrm{s}$, from equation 7.6 one obtains a cooling length between 0.6 and 0.9 µm, comparable to the beam sizes.

In practice this result also means that special care has to be taken when optimizing the system for spatio-temporal measurements. Evidently, maximizing ΔPC does not guarantee minimum spot sizes. Therefore, depending on the type of measurement, a choice must be made how to balance the trade-off between spatial confinement and

signal-to-noise level. On the other hand, as explained in chapter 3.1.4, the initial width is not crucial to resolve small changes of FWHM as a function of Δt, which is rather limited by the signal-to-noise and associated goodness of Gaussian fitting.
Additionally, the knowledge of this effect could be used to study the cooling length in a quite simple way by simply varying the illuminating beam size.
In the following experiments, the beams are focused onto the samples to spots between $1-3\,\mu m$, to balance the benefits of signal-to-noise with spatial resolution.

7.6.2 Asymmetric scan: Experimental results

To investigate the spatio-temporal diffusion dynamics, I now offset the beams with respect to the junction region. In the first approach, I fix the pump beam ($\lambda_{pu} = 443\,nm$) at the junction and scan the probe beam ($\lambda_{pr} = 886\,nm$) perpendicular to the junction axis, i.e. away from the split gate region (see Fig. 7.10b) on the TC device. The two sides of the device are set to p- and n-doping, by setting the voltages of gate A and B to about 1 V above and below the charge neutrality point, respectively.
Figure 7.12a shows the temporal scans of ΔPC as a function of the probe beam offset Δx. The data are offset vertically for better visibility. The black curves are fits to the data with the Laplace curve of equation 7.5. As indicated by the black arrows, the data show a shift in the peak position to higher Δt as the spatial offset is increased. Figure 7.12b shows the same dataset displayed as an image where the y-axis is now the spatial offset and the color scale indicates the signal height. The white lines show the extracted peak positions. In figure 7.12c the data has been normalized to the maximum for each Δx, to be able to highlight this peak shift evolution even for decaying signal strengths.
To gain information about carrier diffusion from these two-dimensional datasets, I quantify the effects of spatial and temporal offset with different measures, involving both horizontal as well as vertical slices through the data.
From the fits to the temporal scans as a function of Δx, the amplitude and diversity are extracted. Figure 7.12d shows the amplitudes, which show maximum values at $\Delta x = 0$. This behavior is expected, as the maximum ΔPC is created when the beams overlap spatially. The diversity, shown in figure 7.12e, seems roughly unchanged over most of the spatial offsets. Note that the trends towards the very high displacements $\Delta x \approx \pm 4\,\mu m$ come from regions with extremely low signal-to-noise and should be regarded as less accurate.
While these measures look at the temporal shape as a function of spatial offset, one might also wonder about the spatial shape as a function of pulse delay. Hence, another way to analyze the data is to interpret the dataset as spatial sections for each Δt and to fit Gaussian profiles to the data, similar to the treatment of the PBS-TAM data in the previous chapters. The result of this Gaussian fitting and extraction of the FWHM is shown in figure 7.12f. The evolution of the FWHM shows an interesting behavior which is not symmetric around $\Delta t = 0$. The minimum width does not seem to coincide with the exact temporal overlap of the beams.
The asymmetry between the times $\Delta t < 0$ and $\Delta t > 0$ in these data is expected. As depicted in figure 7.10, $\Delta t < 0$ corresponds to the probe arriving at the sample before the pump and vice versa. Intuitively, if the signal is interpreted for a given probe offset as coming from the probe arriving at the sample first, and creating hot carriers that

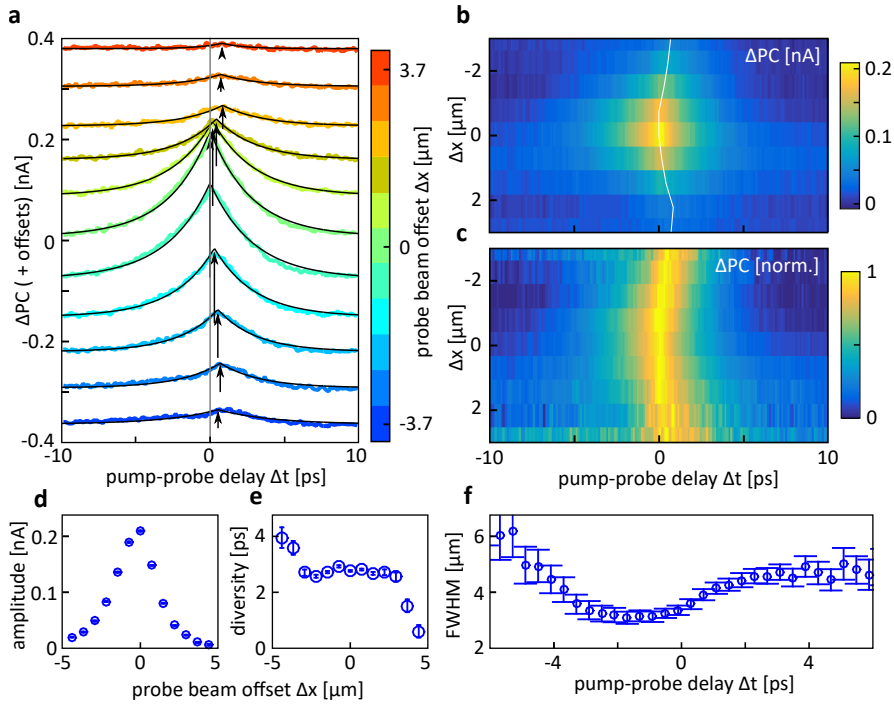

Fig. 7.12: Spatio-temporal transient photocurrent dynamics for an asymmetric scan on the TC device with $\lambda_{pu} = 443\,nm$, $\lambda_{pr} = 886\,nm$. (**a**) ΔPC time traces as a function of the probe beam spatial offset Δx away from the fixed pump beam on the p-n junction. The traces have been offset vertically for visibility. The black curves are fits to the two-sided exponential curve and the vertical arrows indicate the peak of these curves. (**b**) the same dataset as **a** with peak times shown as a white lines. (**c**) and normalized to the maximum value for each Δx. (**d**,**e**) Amplitude and diversity, respectively, as extracted from the fits to the temporal traces, i.e. horizontal slices of the data in **b**. (**f**) Extracted width (FWHM) from Gaussian fits to vertical slices of the data in **b**, i.e., for each Δt. All errorbars show the 68 % confidence intervals of the respective fit parameters.

have had time to diffuse towards the junction area, the strongest ΔPC signal should come at $\Delta t < 0$ for $|\Delta x| > 0$. Yet, the opposite is observed. The peak times actually "bend" towards $\Delta t > 0$, i.e., when the offset probe arrives after the centered pump. A possible explanation of this peculiar behavior will be discussed together with the simulations in section 7.7.2.

7.6.3 Symmetric scan: Experimental results

While the asymmetric scan does show some interesting dynamics, the direct tracking of carrier diffusion is complicated by the asymmetry between the delay times before and after time zero. Another asymmetry comes from the previous use of a combination of a blue and NIR beam for the two pulses. In fact, the usage of the blue laser brings further disadvantages, which are a different power dependence compared to the NIR

beam (not shown), as well as the possibility to create photo-induced doping in the devices[143].

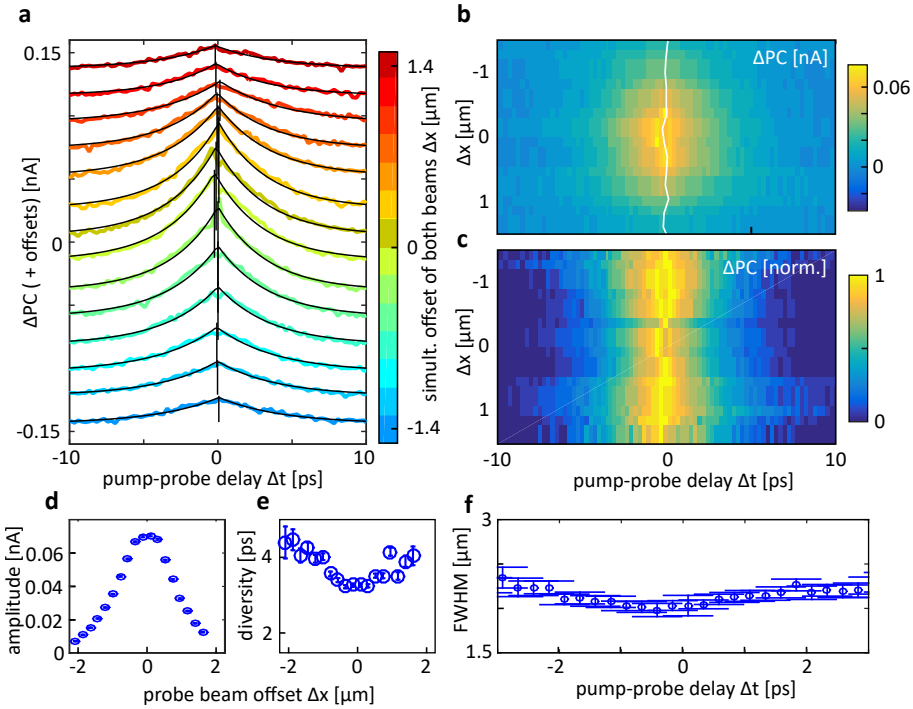

Fig. 7.13: Spatio-temporal transient photocurrent dynamics for a symmetric scan on the HB device with $\lambda_{pu} = \lambda_{pr} = 886\,nm$. (**a**) ΔPC time traces as a function of the simultaneous spatial beam offset Δx of both beams away from the p-n junction. The curves have been offset vertically for visibility. The black curves are fits to the two-sided exponential curve and the vertical lines indicate the peak of these curves. (**b**) the same dataset as **a** with peak times shown as a white line and (**c**) normalized to the maximum value for each Δx. (**d**,**e**) Amplitude and diversity, respectively, as extracted from the fits to the temporal traces. (**f**) Extracted width (FWHM) from Gaussian fits to vertical slices of the data for each Δt. All errorbars show the 68 % confidence intervals of the respective fit parameters.

For these reasons, I proceed to perform a "symmetric scan", i.e., moving two NIR beams ($\lambda_{pu} = \lambda_{pr} = 886\,nm$) simultaneously away from the split-gate area (see Fig. 7.10c). Additionally, the pulses are focused more tightly in this measurement. Before, the beams were focused to about 3 µm to maximize the signal strength while sacrificing some spatial confinement. Now, the beams are focused to about 1 µm to access the best possible spatio-temporal resolution while the signal levels are still acceptable. Furthermore, as I am interested in comparing the dynamics to a device with known mobility, I now also switch to the HB device, as its mobility has been directly measured by a four-point electric characterization, which was not possible for the TC device. The two sides of the device are set to p- and n-doping, by setting the voltages of gate A and B to about 1 V above and below the charge neutrality point, respectively.

The results of this type of "symmetric scan" are summarized in figure 7.13.
As before, the same quantities are shown in the same way as for figure 7.12. This time, figures 7.13a and b show no shift in the peak position, shown as black lines in figure 7.13a and the white curve in figure 7.13b. This is expected, as there is a symmetry between the times $\Delta t < 0$ and $\Delta t > 0$, as illustrated in figure 7.10c.
As the pump and probe beam are identical now, interference fringes are created within the coherence time of the laser when scanning close to $\Delta t = 0$. This can be seen as shifts of the brightest region in figure 7.13c. This minor effect does not significantly influence the further analysis.
The amplitudes (Fig. 7.13d) show, as before, a maximum at $\Delta x = 0$, as expected. The diversity, shown in figure 7.13e shows an increasing trend towards higher displacements this time, even within the trustworthy signal-to-noise regions ($\pm 1\,\mu m$).
The temporal evolution of FWHM, shown in figure 7.13f, now does show a minimum width at $\Delta t = 0$, and a symmetric rise for increasing pump-probe delay.
Contrary to the PBS-TAM data of the previous chapters, to my knowledge there is no simple analytical way to extract diffusivity directly from these data, such as looking at the slope of figure 7.13f. This comes from the added complexity of the nonlinear nature of ΔPC. Therefore, to extract carrier diffusivities from these data, a comparison to a simulation of the full spatio-temporal nonlinear photocurrent dynamics, is desired.

7.7 Simulation

In order to understand the experimental observations shown above, here I describe and present a simulation of the spatio-temporal transient photocurrent measurement. First, I describe the underlying model used to implement the simulation and then show the results and compare them to the experimental data.

7.7.1 Model

This model calculates the photocurrent created due to the elevated electron temperature at the p-n junction after the absorption of both laser pulses as a function of their separation in space and time. The basic assumption is that the absorbed heat energy ΔQ is converted to elevated electron temperature ΔT_e in a nonlinear fashion[75]. Here, in contrast to previous work[75,127,144], I explicitly include both the spatial and temporal dependence of this process,

$$\Delta T_e(x,t) = f_{\text{nonlinear}}(\Delta Q(x,t)). \tag{7.7}$$

The nonlinear function $f_{\text{nonlinear}}$ is the model introduced in section 7.4.3 to describe the power dependence of the PTE photocurrent,

$$f_{\text{nonlinear}}(\Delta Q(x,t)) = \sqrt[2]{T_0^2 + b \cdot \Delta Q} - T_0, \tag{7.8}$$

where $T_0 = 293\,\text{K}$ is the ambient temperature. The factor b is the experimentally determined parameter from section 7.4.3.
For a two-pulse experiment, with absorbed heat ΔQ_α (α ="pu" or "pr") due to the

absorption of the pump or probe pulse, respectively, the total temperature rise $\Delta T_e^{\mathrm{tot}}$ should take into account the absorbed heat of the two and be calculated as

$$\Delta T_e^{\mathrm{tot}}(x,t) = f_{\mathrm{nonlinear}}(\Delta Q_{\mathrm{pu}}(x,t) + \Delta Q_{\mathrm{pr}}(x,t)). \tag{7.9}$$

Finally, the photocurrent PC is assumed to come from the time-averaged extra electron temperature ΔT_e at the location of the split-gate, as this is the location of the stepwise difference in Seebeck coefficients where the photovoltage is created (see Eq. 7.1). The following sum is therefore taken over the junction region in space, and over the full simulation time,

$$\mathrm{PC} = A \cdot \sum_t \sum_{x_{\mathrm{junction}}} \Delta T_e(x,t), \tag{7.10}$$

where the photocurrent $\mathrm{PC} = U_{\mathrm{PTE}}/R$ is given by the local photovoltage and the total device resistance R. The proportionality constant A converts temperature into current, i.e., includes the Seebeck coefficient step $(S_2 - S_1)$ and R. In this simulation I set $A = 1$ and calculate the photocurrent in arbitrary units.
I isolate the decrease in photocurrent due to the (nonlinear) contribution of the two modulated heating sources, ΔQ_{pu} and ΔQ_{pr}. This is analogous to the experimental demodulation at the difference frequency with the double-chopper technique, as introduced in section 7.5. Here, I define the differential photocurrent $\Delta\mathrm{PC}$ as

$$\Delta\mathrm{PC} = \mathrm{PC}(\Delta Q_{\mathrm{pu}}) + \mathrm{PC}(\Delta Q_{\mathrm{pr}}) - \mathrm{PC}(\Delta Q_{\mathrm{pu}} + \Delta Q_{\mathrm{pr}}), \tag{7.11}$$

where the individually calculated currents are obtained by inserting equation 7.7 into equation 7.10.
The input to this simulation is therefore the spatio-temporal pump (and probe) pulse heat energy $\Delta Q_\alpha(x,t)$, with $\alpha =$ "pu" or "pr", respectively, which I model as

$$\Delta Q_\alpha(x,t) = P_\alpha \cdot \exp\left(-\frac{\beta(x+\Delta x_\alpha)^2}{\mathrm{FWHM}_\alpha^2(t)}\right) \cdot \frac{1}{2}\mathrm{erfc}\left(-\frac{t-\Delta t_\alpha}{\tau_\alpha^{\mathrm{rise}}}\right) \exp\left(-\frac{t-\Delta t_\alpha}{\tau_\alpha^{\mathrm{cool}}}\right), \tag{7.12}$$

where P_α is the amplitude; $\beta = 4\ln 2$; Δx_α and Δt_α are the spatial and temporal offset from the coordinates t and x, respectively; $\tau_\alpha^{\mathrm{rise}}$ and $\tau_\alpha^{\mathrm{cool}}$ are the characteristic heating and cooling times, respectively.
Figure 7.14 shows the input $\Delta Q = \Delta Q_{\mathrm{pu}} + \Delta Q_{\mathrm{pr}}$ to the simulation.
This heat model has a Gaussian shape in space, with a time-dependent width $\mathrm{FWHM}_\alpha(t)$, according to a diffusion model with initial width (at $t = \Delta t_\alpha$) of $\mathrm{FWHM}_{0,\alpha}$ and diffusivity D_α,

$$\mathrm{FWHM}_\alpha^2(t) = \mathrm{FWHM}_{0,\alpha}^2 + 4\beta \cdot D_\alpha \cdot (t - \Delta t_\alpha). \tag{7.13}$$

It is hard to see the broadening of the Gaussian width directly in figure 7.14, as the amplitude decays. Therefore, purple lines have been added, showing the contour of the $\mathrm{FWHM}(t)$ for both pulses[†].
The temporal shape of the carrier heating is assumed to be limited by the pump laser pulse length, as the intrinsic electron heating time is much shorter, typically estimated at tens of femtoseconds. Therefore, the pulse length limited rise time is set

[†]Note that as the *squared* width grows linear in time, the width grows according to a squared root law

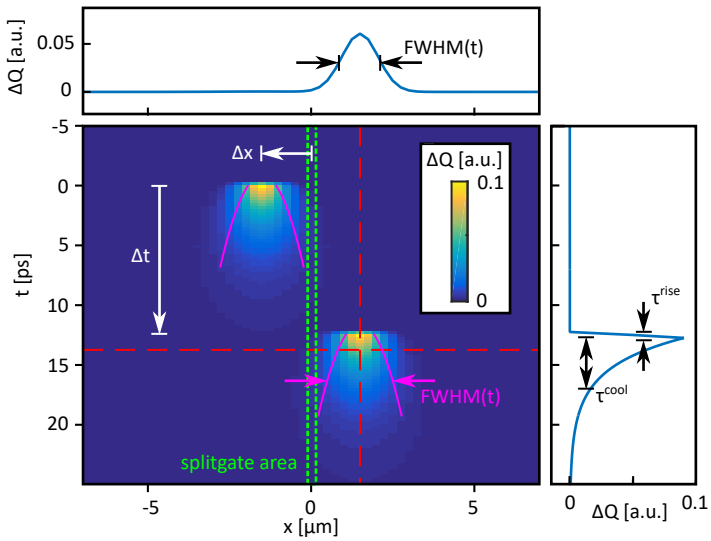

Fig. 7.14: Spatio-temporal heat energy ΔQ as input for the simulation of the graphene p-n junction photocurrent. The input heat energy $\Delta Q = \Delta Q_{\mathrm{pu}} + \Delta Q_{\mathrm{pr}}$ is shown as a function of the spatial and temporal lab-coordinates x and t. Here, a particular set of the variable spatial offset Δx (for the symmetrical scanning) and pump-probe delay Δt are shown. Exemplary spatial and temporal profiles are shown as 1D cutlines along the the red, dashed lines. The top panel shows the Gaussian spatial profile, while the right panel shows the temporal profile, where the variables τ^{rise} and τ^{cool} are labeled. The purple lines show the evolution of FWHM(t) of $\Delta Q_{\mathrm{pu/pr}}$. The position of the split-gate (p-n junction) is shown as a green dashed rectangle.

to $\tau_{\mathrm{pu}}^{\mathrm{rise}} = \tau_{\mathrm{pr}}^{\mathrm{rise}} = 0.15\,\mathrm{ps}$. The subsequent dynamics are modeled with an exponential decay with cooling time $\tau_{\alpha}^{\mathrm{cool}}$.
Although ΔQ_{α} should be proportional to the pulse heating energy, here the heating amplitude P_{α} is inserted as the time averaged laser power at the back entrance to the microscope, i.e., in units of µW, without estimation of the absorbed power fraction. This has the advantage that the conversion to temperature via equation 7.8 is done with the experimentally established conversion factor b, measured in $\mathrm{K}^2/\mu\mathrm{W}$.
Note that t and x represent the "lab"-coordinates, which will be summed over in the calculation of the photocurrent (Eq. 7.10), while the variables Δt and Δx will be used, respectively, as the pump-probe delay and spatial offsets of the beams as they are scanned. In particular, I set $\Delta t_{\mathrm{pu}} \equiv 0$, $\Delta t \equiv \Delta t_{\mathrm{pr}}$; and Δx depending on the scanning mode as $\Delta x \equiv \Delta x_{\mathrm{pu}} = -\Delta x_{\mathrm{pr}}$ for the symmetric scan, or $\Delta x \equiv \Delta x_{\mathrm{pr}}$, $\Delta x_{\mathrm{pu}} \equiv 0$ for the asymmetric scan (c.f. Fig. 7.10b).
The spatial and temporal offsets Δx and Δt depicted in figure 7.14 are quite large, such that there will be nearly no overlap between the deposited heat energy of the pump and the probe. Therefore, $\mathrm{PC}(\Delta Q_{\mathrm{pu}} + \Delta Q_{\mathrm{pr}}) \approx \mathrm{PC}(\Delta Q_{\mathrm{pu}}) + \mathrm{PC}(\Delta Q_{\mathrm{pr}})$ and, according to equation 7.11, $\Delta\mathrm{PC} = 0$. Yet, as Δx and Δt approach zero, there will be more nonlinear contribution to the photocurrent, which will create a finite differential photocurrent $\Delta\mathrm{PC} > 0$.

7.7.2 Simulation results

A typical simulated dataset is presented in figure 7.15.

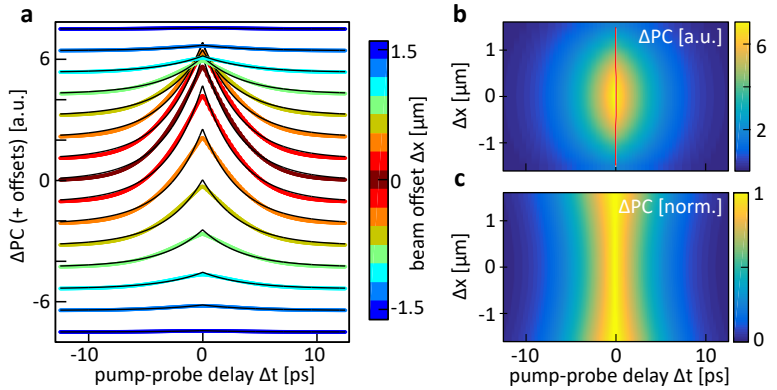

Fig. 7.15: Example of simulation results for a symmetric scan. (**a**) Calculated differential photocurrent ΔPC as a function of Δt and Δx. The black lines are fits to a two-sided exponential function (see Eq. 7.5). The curves are offset vertically for better visibility. (**b**) ΔPC dataset of **a** displayed together with the peak positions (red line). (**c**) The data of **b**, normalized for each Δx. The simulation parameters are $\mathrm{FWHM}_{0,\mathrm{pu}} = \mathrm{FWHM}_{0,\mathrm{pr}} = 1.0\,\mu\mathrm{m}$, $\tau_{\mathrm{pu}} = \tau_{\mathrm{pr}} = 2.0\,\mathrm{ps}$, $D_{\mathrm{pu}} = D_{\mathrm{pr}} = 300\,\mathrm{cm}^2/\mathrm{s}$.

Figure 7.15a shows simulated differential photocurrent ΔPC (see eq. 7.11) as a function of time delay Δt (x-axis) and spatial offset Δx (colors). The black curves are two-sided exponential fits to the data, according to Eq. 7.5. Figure 7.15b shows the same dataset as 7.15a, where the peak positions are shown as a red line. Finally, figure 7.15c shows the ΔPC data normalized to the maximum for horizontal slice, i.e., each Δx.
The observed simulated data looks quite similar to the experimental dataset (c.f. Fig. 7.13). To compare more quantitatively, I proceed to simulate the different scanning modes for the experimental datasets shown here, with the parameters, such as illumination intensity and beam size adjusted to match the experimental conditions, and analyze the results in terms of the extracted parameter evolutions, such as amplitude, diversity, and FWHM.

Asymmetric Scan: Simulation results

Figure 7.16 summarizes the simulated results for an asymmetric scan, with the powers and beam sizes closely matched to those of figure 7.12.
As for this type of asymmetric scan there is an asymmetry between $\Delta t > 0$ and $\Delta t < 0$, the peak position is added as a parameter to the evaluation, as displayed in figure 7.16b. A striking feature is that for high D and only at high displacements $|\Delta x| > 4\,\mu\mathrm{m}$ the intuitively expected effect of diffusion to shift the peak position towards negative Δt is visible. Before, at lower $|\Delta x|$, which corresponds experimentally to the region of trustworthy signal-to-noise, there is actually a shift to positive Δt, even for negligible diffusivities. This behavior was not anticipated and is ascribed to

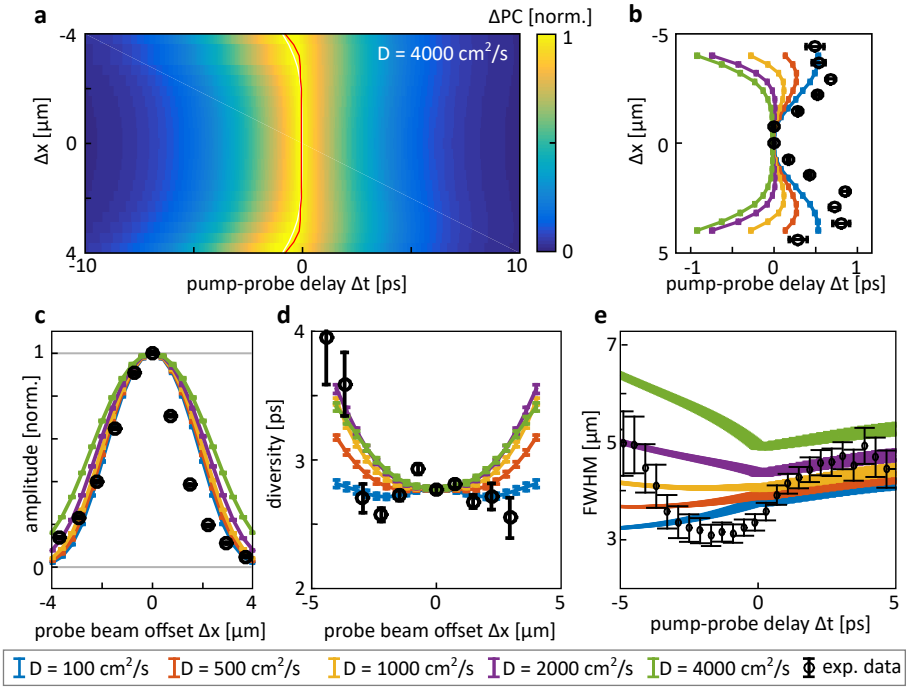

Fig. 7.16: Simulation results for an asymmetric scan for different values of D and comparison to the experimental data. (**a**) ΔPC dataset displayed together with the peak positions (red line). (**b**) Simulated peak positions and comparison to the experimental shifts. The colors represent different diffusion constants, as shown in the legend at the bottom of the figure. The black points are the values extracted from the experimental data (see Fig. 7.13) (**c**, **d**) Amplitude and diversity, respectively, as a function of beam offset, extracted from fits to the Laplace function (see Eq. 7.5). (**e**) Extracted values for FWHM from Gaussian fits to vertical slices of the simulated dataset at each Δt. Simulation parameters: $\mathrm{FWHM}_{0,\mathrm{pu}} = \mathrm{FWHM}_{0,\mathrm{pr}} = 3.0\,\mu\mathrm{m}$, $\tau_{\mathrm{pu}} = \tau_{\mathrm{pr}} = 1.9\,\mathrm{ps}$, $b_{n=2}^{\mathrm{nonlin}} = 2.1 \times 10^7\,\mathrm{K}^2/\mathrm{mW}$, $P_{\mathrm{pr}} = 50\,\mu\mathrm{W}$. Errorbars shows the 68% confidence intervals for all the fitted parameters.

the effect of the instantaneous contribution of the tails of the Gaussian profiles to the photocurrent, although the final reason for this behavior has not been verified. The other measures such as amplitude, diversity, and FWHM (Fig. 7.16c, d, and e, respectively) reproduces the main trends of the experimental data.

Symmetric Scan: Simulation results

I proceed to simulate the other case of symmetric scanning with parameters close to those of the experiment shown in figure 7.13. These simulation results are summarized in figure 7.17.

The simulated trends show quite striking and unintuitive behavior, ultimately due to the inherent nonlinearity of the measurement and an interplay of photocurrent produced due to diffusion and simply due to the tails of the Gaussian spatial distributions

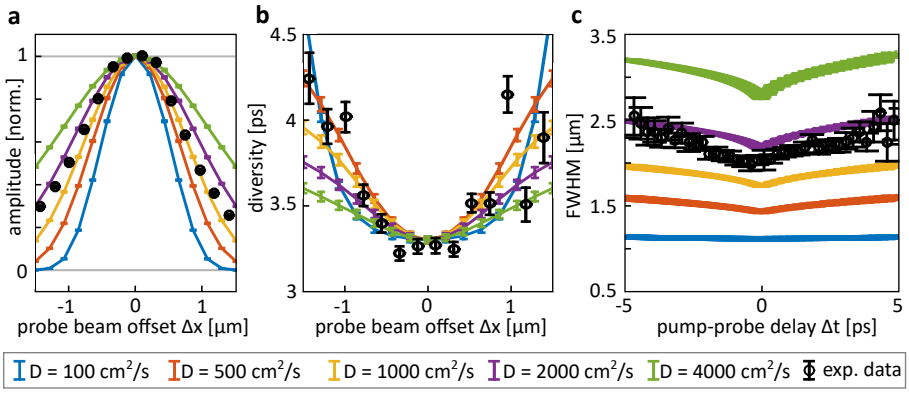

Fig. 7.17: Simulation results for a symmetric scan for different values of D and comparison to the experimental data. (**a** ,**b**) show the extracted parameters from fits to the Laplace function (Eq. 7.5). (**c**) Extracted values for FWHM from Gaussian fits to vertical slices of the simulated dataset at each Δt. Simulation parameters: $\mathrm{FWHM}_{0,\mathrm{pu}} = \mathrm{FWHM}_{0,\mathrm{pr}} = 1.0\,\mu\mathrm{m}$, $\tau_{\mathrm{pu}} = \tau_{\mathrm{pr}} = 2.1\,\mathrm{ps}$, $b_{n=2}^{\mathrm{nonlin}} = 2.1 \times 10^7\,\mathrm{K}^2/\mathrm{mW}$, $P_{\mathrm{pr}} = 100\,\mu\mathrm{W}$. Errorbars shows the 68% confidence intervals for all the fitted parameters.

reaching the split-gate area even for large offsets. For example, the diversities shown in figure 7.17b show very non-monotonous scaling with the diffusivity. While the bending angle of the curves changes, the intermediate diffusivity of $500\,\mathrm{cm}^2/\mathrm{s}$ shows higher values for the diversity than either the lower or the higher D curves, which makes it difficult to exclude any of the simulated diffusivities within the experimental errors from this last graph alone.
Another peculiar fact is that that the FWHM (see Fig. 7.17c), even at $\Delta t = 0$, is not the same for all simulated diffusivities. Instead, it grows for increasing D, starting close to the simulated beam size of $1.0\,\mu\mathrm{m}$, but quickly growing to above $2\,\mu\mathrm{m}$.
Comparison to the experimental data for the amplitudes (Fig. 7.17a), as well as the FWHM (Fig. 7.17c), shows that a simulated diffusivity of about $1800\,\mathrm{cm}^2/\mathrm{s}$ reproduces the trend of the data quite accurately.

7.8 Discussion

After eliminating many complications, the final experiment on the HB sample, using the NIR-NIR laser combination ($\lambda_{\mathrm{pu}} = \lambda_{\mathrm{pr}} = 886\,\mathrm{nm}$), and symmetric scanning type has been performed.
The results of this measurement, summarized in figure 7.13, together with the corresponding simulation results (Fig. 7.17), give a consistent picture of carrier diffusion in this previously electrically characterized device, resulting in an estimated diffusion between 1000 and $2000\,\mathrm{cm}^2/\mathrm{s}$, confirming the predictions made by the Einstein relation for graphene, as shown in figure 7.1b.
An open question remains how the diffusion is influenced by the carrier temperature. On the one hand, the version of the Einstein relation used here is independent of the carrier temperature (contrary to other literature), suggesting a constant diffusivity

over a wide electron temperature range, even for "hot" carriers with temperatures on the order of 1000 K. On the other hand, deviations from this diffusion law might be expected for deviations to the Sommerfeld approximation, i.e., when the electron temperature becomes comparable to the Fermi level, or in truly non-equilibrium situations, e.g. in the first femtoseconds, or for devices with more restricted geometries, such as narrow graphene nanoribbons.

7.9 Summary

I have presented, for the first time, a novel technique of ultrafast photocurrent microscopy with spatially and temporally offset laser pulses, to study the hot-carrier diffusion in graphene p-n junction devices. With the help of simulations I am able to interpret spatially and temporally resolved data to track hot-carrier diffusion in graphene with photothermoelectric photocurrent read-out. The direct connection between carrier mobility and hot-carrier diffusion as predicted by the Einstein relation has been successfully tested by studying electrically characterized devices with known mobilities.

Acknowledgments

The experiments reported in this chapter were performed under the supervision of Dr. Klaas-Jan Tielrooij, Dr. Matz Liebel, and Prof. Dr. Niek F. van Hulst. The results were discussed and interpreted together with Dr. Klaas-Jan Tielrooij, Dr. Matz Liebel, and Prof. Dr. Niek F. van Hulst. The graphene split-gate devices used in this chapter were fabricated by Diana Davydovskaya and Niels Hesp under the supervision of Dr. Klaas-Jan Tielrooij and Prof. Dr. Frank Koppens.

8 Conclusion

8.1 Summary

In this thesis I have explored the possibility to directly resolve, in space and time, the ultrafast heat transport in different materials and with different techniques. I focused on the idea to distinguish electronic and phononic contributions to heat transport dynamics on the nanometer-femtosecond scale. By using a combination of an ultrafast laser pump-probe scheme and advanced scanning microscopy with high-NA, I was able to track, in a very direct way, the diffusion of photo-excited electrons in thin gold films. In particular, a transition from a fast hot-electron diffusion to thermal, phonon limited diffusion was observed, for the first time, on a timescale of a few picoseconds. The experimental observations were understood and interpreted by a full 3D thermo-optical model, based on a two-temperature model and, importantly, including the electronic and phononic temperature contributions to the optical response.
In a next step, the exciting optoelectronic properties of the highly conductive, novel, and truly two-dimensional material, single-layer graphene, were put to the test. Specifically, I studied the electronic heat transport, thereby directly investigating the relationship between carrier mobility thermal diffusivity, as predicted by the Einstein relation. Again, by using state-of-the-art probe-beam-scanning transient-absorption microscopy, I was able to track light induced electronic heat in graphene. I studied the influence of external parameters on hot-carrier transport by comparing the dynamics of samples with different substrates, encapsulations, as well as production types.
As a complementary approach to disentangle electronic from phononic contributions to carrier diffusion, I switched gears to use a purely electronic read-out, the photothermoelectric voltage created at a p-n junction of a electrostatically gated device of exfoliated, encapsulated graphene in a double-pulse, spatio-temporal scanning scheme. This novel technique, which is the first of its kind, allowed me to detect and map out

the photocurrent on a nanometer-femtosecond scale. By modeling the expected carrier dynamics with a spatio-temporal diffusion model, I was able to compare the observed and simulated dynamics to estimate the hot-carrier diffusion. Here, an electrically pre-characterized device with known mobility was studied to test the predictions of the Einstein relation. Within the experiential uncertainties, I was able to estimate a hot-carrier diffusion coefficient consistent with the measured electron mobilities.
With these achievements I believe to have contributed to the understanding of complex non-equilibrium thermodynamics, in particular the decoupling of electronic and phononic heat transport via two novel ultrafast microscopy techniques. By precise quantification of all-optical transient absorption, as well as photothermoelectric photocurrent microscopy, I have been able to put our current understanding of carrier and heat transport to the test.

8.2 Outlook

There is a vast possibility of extending this research both by expanding the fundamental understanding, as well as for technological applications.
First of all, the limits of diffusive heat transport, even just for the electron subsystem have been predicted in terms of ballistic motion, when the length scales approach the mean free path, as well as and non-thermal motion, when the time resolution reaches the scales of electron-electron scattering. It should be possible to probe both of these regimes with the types of ultrafast spatio-temporally resolved microscopy techniques presented in this thesis.
In principle, this type of fundamental understanding should be able to help to improve the efficiency of solar cells, where the light energy needs to arrive at the charge extraction regions by carrier transport. In particular, the generally unused extra energy of hot carriers could boost solar cell efficiency if designed to harvest these carriers.
The field of ultrafast detectors relies on similar fundamental processes of creating currents from optical absorption. In particular, graphene is a promising material to create ultrafast, compact and broadband detectors, for example in the terahertz wavelength regime.

Acknowledgments

First of all, I would like to thank Niek for giving me the great opportunity to conduct my research in his group at ICFO, a fruitful environment with cutting edge facilities, as well as a great team. I am thankful for his guidance, as well as leaving me all the freedoms to figure out and follow my own path, while always being available for scientific (or any other type of) discussions.
To Matz I am tremendously grateful for supporting me from the beginning, for his patience, guidance, source of detailed knowledge mixed with the right intuitions and work ethic, and continued scientific support on any topic.
Thanks to Klaas for fruitful collaborations and discussions over the last years and his continued faith in me for the future.
Further thanks go out to Renwen, for many scientific discussions and collaborations.
Thanks also to my other collaborators, Javier, Yonatan, as well as other contributors to my work, such as Matthieu, Gabi, the mechanical and electronics workshop, as well as facilities departments at ICFO.
I want to thank the entire molecular nanophotonics group, from the ones who have seen me arrive, James, Nicolò, Pablo, Łukasz, Gaëtan, Richard, Anshuman, Ion, to the ones who see me leave, Vikas, Lisa, Esther, Nicola, Alex D, Jana, Unai, Monse, Luca, Paweł and Manuel.
A great thanks to helping me balance out my work life at ICFO goes to the kickers of the tiny futbolín balls of the first half of my PhD, to the kickers of the regular sized soccer balls on the field in the second half, as well as the wakeboard, coffee machine and couches in between.
Finally, a great shout out goes to the unconditional love and support from my family, friends, and partner Lina.

Bibliography

1. Besson, U. The history of the cooling law: When the search for simplicity can be an obstacle. *Science & Education* **21,** 1085–1110 (2012).
2. Fourier, J.-B. J. *The analytical theory of heat* (The University Press, Cambridge, 1878).
3. Fick, A. Ueber Diffusion. *Annalen der Physik* **170,** 59–86 (1855).
4. Thomson, J. J. Cathode rays. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **44,** 293–316 (1897).
5. Krönig, A. Grundzüge einer Theorie der Gase. *Annalen der Physik* **175,** 315–322 (1856).
6. Clausius, R. Ueber die Art der Bewegung, welche wir Wärme nennen. *Annalen der Physik* **176,** 353–380 (1857).
7. Maxwell, J. C. Illustrations of the dynamical theory of gases. —Part I. On the motions and collisions of perfectly elastic spheres. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **19,** 19–32 (1860).
8. Einstein, A. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik* **322,** 549–560 (1905).
9. Kittel, C. *Introduction to solid state physics* (Wiley, New York, 1996).
10. Kaviany, M. *Essentials of heat transfer: principles, materials, and applications* (Cambridge University Press, 2011).
11. Sellitto, A., Cimmelli, V. A. & Jou, D. *Mesoscopic Theories of Heat Transport in Nanosystems* (Springer, Switzerland, 2016).
12. Jou, D., Cimmelli, V. A. & Sellitto, A. Nonequilibrium temperatures and second-sound propagation along nanowires and thin layers. *Physics Letters A* **373,** 4386–4392 (2009).
13. Dong, Y., Cao, B. Y. & Guo, Z. Y. Ballistic-diffusive phonon transport and size induced anisotropy of thermal conductivity of silicon nanofilms. *Physica E* **66,** 1–6 (2015).
14. Cimmelli, V. A., Sellitto, A. & Jou, D. Nonlinear evolution and stability of the heat flow in nanosystems: Beyond linear phonon hydrodynamics. *Physical Review B* **82,** 184302 (2010).
15. Lepri, S. *Thermal transport in low dimensions: from statistical physics to nanoscale heat transfer* (Springer, Switzerland, 2016).

16. Lebon, G., Jou, D. & Casas-Vázquez, J. *Understanding non-equilibrium thermodynamics* (Springer, Berlin, 2008).

17. Bartelt, J. A., Lam, D., Burke, T. M., Sweetnam, S. M. & McGehee, M. D. Charge-Carrier Mobility Requirements for Bulk Heterojunction Solar Cells with High Fill Factor and External Quantum Efficiency >90%. *Advanced Energy Materials* **5,** 1500577 (2015).

18. Han, N. *et al.* Improved heat dissipation in gallium nitride light-emitting diodes with embedded graphene oxide pattern. *Nature Communications* **4,** 1452–1458 (2013).

19. Ross, R. T. & Nozik, A. J. Efficiency of hot-carrier solar energy converters. *Journal of Applied Physics* **53,** 3813–3818 (1982).

20. Guo, Z. *et al.* Long-range hot-carrier transport in hybrid perovskites visualized by ultrafast microscopy. *Science* **356,** 59–62 (2017).

21. Schoenlein, R. W., Lin, W. Z., Fujimoto, J. G. & Eesley, G. L. Femtosecond studies of nonequilibrium electronic processes in metals. *Physical Review Letters* **58,** 1680–1683 (1987).

22. Gabor, N. M. *et al.* Hot carrier–assisted intrinsic photoresponse in graphene. *Science* **334,** 648–652 (2011).

23. Crank, J. *et al. The mathematics of diffusion* (Oxford university press, 1979).

24. Bird, R. B., Stewart, W. E. & Lightfoot, E. N. *Transport phenomena* (John Wiley & Sons, New York, 2002).

25. King, J. A. *Materials Handbook for Hybrid Microelectronics* (Artech House, Boston, 1988).

26. Pillai, C. G. & George, A. M. High-temperature thermal conductivity of NdCoO3 and GdCoO3. *International Journal of Thermophysics* **12,** 207–221 (1991).

27. Anisimov, S., Kapeliovich, B. & Perel-man, T. Electron emission from metal surfaces exposed to ultrashort laser pulses. *Journal of Experimental and Theoretical Physics* **66,** 375–377 (1974).

28. Rethfeld, B., Ivanov, D. S., Garcia, M. E. & Anisimov, S. I. Modelling ultrafast laser ablation. *Journal of Physics D: Applied Physics* **50,** 193001 (2017).

29. Ibach, H. & Lüth, H. *Solid-State physics: An Introduction to Principles of Material Science* **2,** 87 (Springer, Berlin, 2009).

30. Fann, W. S., Storz, R., Tom, H. W. & Bokor, J. Direct measurement of nonequilibrium electron-energy distributions in sub-picosecond laser-heated gold films. *Physical Review Letters* **68,** 2834–2837 (1992).

31. Waldecker, L., Bertoni, R., Ernstorfer, R. & Vorberger, J. Electron-phonon coupling and energy flow in a simple metal beyond the two-temperature approximation. *Physical Review X* **6,** 021003 (2016).

32. Groeneveld, R. H. M., Sprik, R. & Lagendijk, A. Femtosecond spectroscopy of electron-electron and alectron-phonon energy relaxation in Ag aund Au. *Physical Review B* **51,** 11433 (1995).

33. Lin, Z., Zhigilei, L. V. & Celli, V. Electron-phonon coupling and electron heat capacity of metals under conditions of strong electron-phonon nonequilibrium. *Physical Review B* **77,** 075133 (2008).

34. Fann, W. S., Storz, R., Tom, H. W. & Bokor, J. Electron thermalization in gold. *Physical Review B* **46,** 13592–13595 (1992).

35. Ashcroft, N. W. & Mermin, N. D. *Solid State Physics.* (Harcourt Inc., New York, 1976).

36. Nag, B. R. *Electron Transport in Compound Semiconductors* (Springer, Heidelberg, 1980).

37. Weisbuch, C. & Vinter, B. *Quantum Semiconductor Structures: Fundamentals and Applications* (Academic Press, Inc., Boston, 1991).

38. Ruzicka, B. A., Werake, L. K., Samassekou, H. & Zhao, H. Ambipolar diffusion of photoexcited carriers in bulk GaAs. *Applied Physics Letters* **97,** 262119 (2010).

39. Wang, R. *et al.* Ultrafast and spatially resolved studies of charge carriers in atomically thin molybdenum disulfide. *Physical Review B* **86,** 045406 (2012).

40. Cui, Q., Ceballos, F., Kumar, N. & Zhao, H. Transient absorption microscopy of monolayer and bulk WSe2. *ACS Nano* **8,** 2970–2976 (2014).

41. Kumar, N. *et al.* Exciton diffusion in monolayer and bulk MoSe2. *Nanoscale* **6,** 4915–4919 (2014).

42. Das Sarma, S., Adam, S., Hwang, E. H. & Rossi, E. Electronic transport in two-dimensional graphene. *Reviews of Modern Physics* **83,** 407–470 (2011).

43. Zebrev, G. *Graphene Field Effect Transistors: Diffusion-Drift Theory* (IntechOpen, 2011).

44. Rengel, R. & Martín, M. J. Diffusion coefficient, correlation function, and power spectral density of velocity fluctuations in monolayer graphene. *Journal of Applied Physics* **114** (2013).

45. Guo, Z., Manser, J. S., Wan, Y., Kamat, P. V. & Huang, L. Spatial and temporal imaging of long-range charge transport in perovskite thin films by ultrafast microscopy. *Nature Communications* **6,** 7471 (2015).

46. Cui, Q. *et al.* Transient Absorption Measurements on Anisotropic Monolayer ReS2. *Small* **11,** 5565–5571 (2015).

47. Kulig, M. *et al.* Exciton diffusion and halo effects in atomically thin semiconductors. *Physical Review Letters* **120,** 207401 (2018).

48. Berera, R., van Grondelle, R. & Kennis, J. T. M. Ultrafast transient absorption spectroscopy: Principles and application to photosynthetic systems. *Photosynthesis Research* **101,** 105–118 (2009).

49. Dantus, M. & Gross, P. Ultrafast spectroscopy. *Encyclopedia of Applied Physics* **22,** 431–456 (1998).

50. Hartland, G. Measurements of the material properties of metal nanoparticles by time-resolved spectroscopy. *Physical Chemistry Chemical Physics* **6,** 5263–5274 (2004).

51. Gao, C. *et al.* Sign of differential reflection and transmission in pump-probe spectroscopy of graphene on dielectric substrate. *Photonics Research* **3,** A1 (2015).
52. Mukamel, S. *Principles of Nonlinear Optical Spectroscopy* (Oxford University Press, 1999).
53. Hamm, P. Principles of Nonlinear Optical Spectroscopy: A Practical Approach or: Mukamel for Dummies. *University of Zurich* **41,** 77 (2005).
54. Cahill, D. G. *et al.* Nanoscale thermal transport. *Journal of Applied Physics* **93,** 793–818 (2003).
55. Cahill, D. G. Analysis of heat flow in layered structures for time-domain thermoreflectance. *Review of Scientific Instruments* **75,** 5119–5122 (2004).
56. Feser, J. P., Liu, J. & Cahill, D. G. Pump-probe measurements of the thermal conductivity tensor for materials lacking in-plane symmetry. *Review of Scientific Instruments* **85,** 104903 (2014).
57. Jiang, P., Qian, X. & Yang, R. Time-domain thermoreflectance (TDTR) measurements of anisotropic thermal conductivity using a variable spot size approach. *Review of Scientific Instruments* **88,** 074901 (2017).
58. Vega-Flick, A. *et al.* Thermal transport in suspended silicon membranes measured by laser-induced transient gratings. *AIP Advances* **6,** 121903 (2016).
59. Johnson, J. A. *et al.* Direct measurement of room-temperature nondiffusive thermal transport over micron distances in a silicon membrane. *Physical Review Letters* **110,** 025901 (2013).
60. Huberman, S. *et al.* Observation of second sound in graphite at temperatures above 100 k. *Science* **379,** 375–379 (2019).
61. Maznev, A. A., Johnson, J. A. & Nelson, K. A. Non-equilibrium transient thermal grating relaxation in metal. *Journal of Applied Physics* **109,** 073517 (2011).
62. Chen, K. *et al.* Measurement of Ambipolar Diffusion Coefficient of Photoexcited Carriers with Ultrafast Reflective Grating-Imaging Technique. *ACS Photonics* **4,** 1440–1446 (2017).
63. Akselrod, G. M. *et al.* Visualization of exciton transport in ordered and disordered molecular solids. *Nature Communications* **5,** 3646 (2014).
64. Akselrod, G. M. *et al.* Subdiffusive Exciton Transport in Quantum Dot Solids. *Nano Letters* **26,** 3556–3562 (2014).
65. Ruzicka, B. A. *et al.* Hot carrier diffusion in graphene. *Physical Review B* **82,** 195414 (2010).
66. Wan, Y. *et al.* Cooperative singlet and triplet exciton transport in tetracene crystals visualized by ultrafast microscopy. *Nature Chemistry* **7,** 785–792 (2015).
67. Ceballos, F. & Zhao, H. Ultrafast laser spectroscopy of two-dimensional materials beyond graphene. *Advanced Functional Materials* **27,** 1604509 (2017).
68. Massaro, E. S., Hill, A. H. & Grumstrup, E. M. Super-Resolution Structured Pump-Probe Microscopy. *ACS Photonics* **3,** 501–506 (2016).
69. Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313,** 1642–1645 (2006).

70. Block, A. *et al.* Tracking ultrafast hot-electron diffusion in space and time by ultrafast thermomodulation microscopy. *Science Advances* **5,** eaav8965 (2019).

71. Bracewell, R. N. *The Fourier Transform and its Applications* (McGraw-Hill, New York, 1986).

72. Trebino, R. *Frequency-Resolved Optical Gating: The Measurement of Ultrashort Laser Pulses* (Springer Science & Business Media, New York, 2000).

73. Walmsley, L., Waxer, L. & Dorrer, C. The role of dispersion in ultrafast optics. *Review of Scientific Instruments* **72,** 1–29 (2001).

74. Bor, Z. & Rácz, B. Group velocity dispersion in prisms and its application to pulse compression and travelling-wave excitation. *Optics Communications* **54,** 165–170 (1985).

75. Tielrooij, K. J. *et al.* Generation of photovoltage in graphene on a femtosecond timescale through efficient carrier heating. *Nature Nanotechnology* **10,** 437–443 (2015).

76. Suzaki, Y. & Tachibana, A. Measurement of the μm sized radius of Gaussian laser beam using the scanning knife-edge. *Applied Optics* **14,** 2809–2810 (1975).

77. Huber, C., Orlov, S., Banzer, P. & Leuchs, G. Corrections to the knife-edge based reconstruction scheme of tightly focused light beams. *Optics Express* **21,** 25069 (2013).

78. Ahmadi, T. S., Logunov, S. L. & El-Sayed, M. A. Picosecond dynamics of colloidal gold nanoparticles. *Journal of Physical Chemistry* **100,** 8053–8056 (1996).

79. Chong, S., Min, W. & Xie, X. S. Ground-State Depletion Microscopy: Detection Sensitivity of Single-Molecule Optical Absorption at Room Temperature. *The Journal of Physical Chemistry Letters* **1,** 3316–3322 (2010).

80. Fu, D. *et al.* Two-color excited-state absorption imaging of melanins. *Proceedings of SPIE* **6424,** 642402 (2007).

81. Wei, L. & Min, W. Pump-probe optical microscopy for imaging nonfluorescent chromophores. *Analytical and Bioanalytical Chemistry* **403,** 2197–2202 (2012).

82. Davydova, D., de la Cadena, A., Akimov, D. & Dietzek, B. Transient absorption microscopy: Advances in chemical imaging of photoinduced dynamics. *Laser and Photonics Reviews* **10,** 62–81 (2016).

83. Huang, L. & Cheng, J.-X. Nonlinear optical microscopy of single nanostructures. *Annual Review of Materials Research* **43,** 213–236 (2013).

84. Gao, B., Hartland, G. V. & Huang, L. Transient absorption spectroscopy of excitons in an individual suspended metallic carbon nanotube. *Journal of Physical Chemistry Letters* **4,** 3050–3055 (2013).

85. Feng, Y., Vinogradov, I. & Ge, N.-H. General noise suppression scheme with reference detection in heterodyne nonlinear spectroscopy. *Optics Express* **25,** 26262 (2017).

86. Zhao, S., Surwade, S. P., Li, Z. & Liu, H. Photochemical oxidation of CVD-grown single layer graphene. *Nanotechnology* **23,** 355703 (2012).

87. Haigh, S. J. *et al.* Cross-sectional imaging of individual layers and buried interfaces of graphene-based heterostructures and superlattices. *Nature Materials* **11,** 764–767 (2012).

88. Fischer, M. C., Wilson, J. W., Robles, F. E. & Warren, W. S. Invited Review Article: Pump-probe microscopy. *Review of Scientific Instruments* **87,** 031101 (2016).

89. Canali, C., Nava, F. & Reggiani, L. in *Hot-electron transport in semiconductors* 87–112 (Springer, Berlin, 1985).

90. Brorson, S. D., Fujimoto, J. G. & Ippen, E. P. Femtosecond electronic heat-transport dynamics in thin gold films. *Physical Review Letters* **59,** 1962–1965 (1987).

91. Juhasz, T., Elsayed-Ali, H. E., Smith, G. O., Suárez, C. & Bron, W. E. Direct measurements of the transport of nonequilibrium electrons in gold films with different crystal structures. *Physical Review B* **48,** 15488–15491 (1993).

92. Qiu, T. Q. & Tien, C. L. Heat Transfer Mechanisms During Short-Pulse Laser Heating of Metals. *Journal of Heat Transfer* **115,** 835 (1993).

93. Sun, C. K., Vallée, F., Acioli, L. H., Ippen, E. P. & Fujimoto, J. G. Femtosecond-tunable measurement of electron thermalization in gold. *Physical Review B* **50,** 15337–15348 (1994).

94. Suárez, C., Bron, W. E. & Juhasz, T. Dynamics and transport of electronic carriers in thin gold films. *Physical Review Letters* **75,** 4536–4539 (1995).

95. Hohlfeld, J., Müller, J. G., Wellershoff, S. S. & Matthias, E. Time-resolved thermoreflectivity of thin gold films and its dependence on film thickness. *Applied Physics B* **64,** 387–390 (1997).

96. Wellershoff, S. S., Hohlfeld, J., Güdde, J. & Matthias, E. The role of electron-phonon coupling in femtosecond laser damage of metals. *Applied Physics A* **69,** 99–107 (1999).

97. Bonn, M. *et al.* Ultrafast electron dynamics at metal surfaces: Competition between electron-phonon coupling and hot-electron transport. *Physical Review B* **61,** 1101–1105 (2000).

98. Del Fatti, N. *et al.* Nonequilibrium electron dynamics in noble metals. *Physical Review B* **61,** 16956–16966 (2000).

99. Hohlfeld, J. *et al.* Electron and lattice dynamics following optical excitation of metals. *Chemical Physics* **251,** 237–258 (2000).

100. Poletkin, K. V., Gurzadyan, G. G., Shang, J. & Kulish, V. Ultrafast heat transfer on nanoscale in thin gold films. *Applied Physics B* **107,** 137–143 (2012).

101. Heilpern, T. *et al.* Determination of hot carrier energy distributions from inversion of ultrafast pump-probe reflectivity measurements. *Nature Communications* **9,** 1853 (2018).

102. Rethfeld, B., Kaiser, A., Vicanek, M. & Simon, G. Ultrafast dynamics of nonequilibrium electrons in metals under femtosecond laser irradiation. *Physical Review B* **65,** 214303 (2002).

103. Matthias, E. *et al.* The influence of thermal diffusion on laser ablation of metal films. *Applied Physics A* **58,** 129–136 (1994).

104. Käding, O. W., Skurk, H., Maznev, A. A. & Matthias, E. Transient thermal gratings at surfaces for thermal characterization of bulk materials and thin films. *Applied Physics A* **61,** 253–261 (1995).

105. Katayama, K., Sawada, T. & Shen, Q. Detection of photoinduced electronic, thermal, and acoustic dynamics of gold film using a transient reflecting grating method under three types of surface plasmon resonance conditions. *Physical Review B* **58,** 8428–8436 (1998).

106. Takata, Y., Haneda, H., Mitsuhashi, T. & Wada, Y. Evaluation of thermal diffusivity for thin gold films using femtosecond laser excitation technique. *Applied Surface Science* **189,** 227–233 (2002).

107. Demaske, B. J., Zhakhovsky, V. V., Inogamov, N. A. & Oleynik, I. I. Ablation and spallation of gold films irradiated by ultrashort laser pulses. *Physical Review B* **82,** 064113 (2010).

108. Born, M. & Wolf, E. *Principles of Optics* (Cambridge University Press, 2009).

109. Castro Neto, A. H., Guinea, F., Peres, N. M., Novoselov, K. S. & Geim, A. K. The electronic properties of graphene. *Reviews of Modern Physics* **81,** 109–162 (2009).

110. Bonaccorso, F., Sun, Z., Hasan, T. & Ferrari, A. C. Graphene photonics and optoelectronics. *Nature Photonics* **4,** 611–622 (2010).

111. Avouris, P. Graphene: Electronic and photonic properties and devices. *Nano Letters* **10,** 4285–4294 (2010).

112. Peres, N. M. Colloquium: The transport properties of graphene: An introduction. *Reviews of Modern Physics* **82,** 2673–2700 (2010).

113. Balandin, A. A. *et al.* Superior Thermal Conductivity of Single-Layer Graphene. *Nano Letters* **8,** 902–907 (2008).

114. Buron, J. D. *et al.* Graphene mobility mapping. *Scientific Reports* **5,** 12305 (2015).

115. Wang, L. *et al.* One-dimensional electrical contact to a two-dimensional material. *Science* **342,** 614–617 (2013).

116. Bolotin, K. I. *et al.* Ultrahigh electron mobility in suspended graphene. *Solid State Communications* **146,** 351–355 (2008).

117. Dean, C. R. *et al.* Boron nitride substrates for high-quality graphene electronics. *Nature Nanotechnology* **5,** 722–726 (2010).

118. Gannett, W. *et al.* Boron nitride substrates for high mobility chemical vapor deposited graphene. *Applied Physics Letters* **98,** 242105 (2011).

119. Ruzicka, B. A. *et al.* Spatially resolved pump-probe study of single-layer graphene produced by chemical vapor deposition. *Optical Materials Express* **2,** 708–716 (2012).

120. Pallecchi, E. *et al.* High electron mobility in epitaxial graphene on 4H-SiC(0001) via post-growth annealing under hydrogen. *Scientific Reports* **4,** 4558 (2014).

121. Chen, K. *et al.* Non-destructive measurement of photoexcited carrier transport in graphene with ultrafast grating imaging technique. *Carbon* **107,** 233–239 (2016).

122. Rengel, R., Pascual, E. & Martín, M. J. Influence of the substrate on the diffusion coefficient and the momentum relaxation in graphene: The role of surface polar phonons. *Applied Physics Letters* **104** (2014).

123. Rengel, R., Iglesias, J. M., Pascual, E. & Martín, M. J. Effect of charged impurity scattering on the electron diffusivity and mobility in graphene. *Journal of Physics: Conference Series* **647,** 012046 (2015).

124. Stange, A. *et al.* Hot electron cooling in graphite: Supercollision versus hot phonon decay. *Physical Review B* **92,** 184303 (2015).

125. Gao, B. *et al.* Studies of intrinsic hot phonon dynamics in suspended graphene by transient absorption microscopy. *Nano Letters* **11,** 3184–3189 (2011).

126. Wu, S. *et al.* Hot phonon dynamics in graphene. *Nano Letters* **12,** 5495–5499 (2012).

127. Graham, M. W., Shi, S. F., Ralph, D. C., Park, J. & McEuen, P. L. Photocurrent measurements of supercollision cooling in graphene. *Nature Physics* **9,** 103–108 (2013).

128. Graham, M. W. *et al.* Transient absorption and photocurrent microscopy show that hot electron supercollisions describe the rate-limiting relaxation step in graphene. *Nano letters* **13,** 5497–5502 (2013).

129. Song, J. C. W., Reizer, M. Y. & Levitov, L. S. Disorder-assisted electron-phonon scattering and cooling pathways in graphene. *Physical Review Letters* **109,** 106602 (2012).

130. Mihnev, M. T. *et al.* Microscopic origins of the terahertz carrier relaxation and cooling dynamics in graphene. *Nature Communications* **7,** 11617 (2016).

131. Tielrooij, K. J. *et al.* Out-of-plane heat transfer in van der Waals stacks through electron-hyperbolic phonon coupling. *Nature Nanotechnology* **13,** 41–46 (2018).

132. Malard, L. M., Fai Mak, K., Castro Neto, A. H., Peres, N. M. R. & Heinz, T. F. Observation of intra- and inter-band transitions in the transient optical response of graphene. *New Journal of Physics* **15,** 015009 (2013).

133. Chen, K. *et al.* Diversity of ultrafast hot-carrier-induced dynamics and striking sub-femtosecond hot-carrier scattering times in graphene. *Carbon* **72,** 402–409 (2014).

134. Mak, K. F., Ju, L., Wang, F. & Heinz, T. F. Optical spectroscopy of graphene: From the far infrared to the ultraviolet. *Solid State Communications* **152,** 1341–1349 (2012).

135. Lyon, T. J. *et al.* Probing Electron Spin Resonance in Monolayer Graphene. *Physical Review Letters* **119,** 066802 (2017).

136. Woessner, A. *et al.* Electrical detection of hyperbolic phonon-polaritons in heterostructures of graphene and boron nitride. *npj 2D Materials and Applications* **1,** 25 (2017).

137. Song, J. C., Rudner, M. S., Marcus, C. M. & Levitov, L. S. Hot carrier transport and photocurrent response in graphene. *Nano Letters* **11,** 4688–4692 (2011).

138. Liu, S. *et al.* Imaging the Long Transport Lengths of Photo-generated Carriers in Oriented Perovskite Films. *Nano Letters* **16,** 7925–7929 (2016).

139. Ma, Q. *et al.* Competing Channels for Hot-Electron Cooling in Graphene. *Physical Review Letters* **112,** 247401 (2014).

140. Woessner, A. *Exploring flatland nano-optics with graphene plasmons* PhD thesis (Universitat Politècnica de Catalunya, 2017).

141. Tielrooij, K. J. *et al.* Hot-carrier photocurrent effects at graphene-metal interfaces. *Journal of Physics Condensed Matter* **27,** 164207 (2015).

142. Urich, A., Unterrainer, K. & Mueller, T. Intrinsic response time of graphene photodetectors. *Nano Letters* **11,** 2804–2808 (2011).

143. Ju, L. *et al.* Photoinduced doping in heterostructures of graphene and boron nitride. *Nature Nanotechnology* **9,** 348–352 (2014).

144. Sun, D. *et al.* Ultrafast hot-carrier-dominated photocurrent in graphene. *Nature Nanotechnology* **7,** 114–118 (2012).