Optical and Atomic Force Microscopy Investigation of Two-dimensional Crystals

Bálint Fülöp, J. Gergő Fehérvári

November 4, 2022

1 Introduction

The object of this student measurement is to introduce the physics master students, who specialize in Nanotechnology or Theoretical Physics to the Atomic Force Microscopy (AFM) measurement method. Experiments will be conducted via the Semilab AFM-101 device in the BME Nanoelectronics Lab. We examine the surface morphology and thickness of two kinds of two-dimensional (2D) crystals: graphene and hexagonal boron nitride (hBN). First, we produce microscopic 2D crystal flakes using the mechanical exfoliation method and characterise them under optical microscope. Then, we explore the Semilab AFM-101 device and discuss its working principle. In the following, we apply the obtained knowledge to the AFM characterisation of the exfoliated 2D crystal samples. Using post-processing techniques, the measurements reveal information about the surface contamination and the thickness of the flakes. Finally, the measured thickness can be compared to the optical micrographs and a colour-to-thickness conversion can be established for hBN.

2 AFM Fundamentals

2.1 Working principle

The AFM is a branch of scanning probe microscopy (SPM), where the imaging is performed by moving a scanning tip over a rectangular area on the sample surface line-by-line, which is called a *raster scan*. Its working princpile is shown in Fig. 1a). The scanning tip is on the down-facing side of an elastic cantilever, which is moved with respect to the sample by using piezoelectric actuators. A laser beam is directed on the upper face of the cantilever, where it is reflected to a photodetector. During measurement, the tip is moved close enough to the sample surface that the cantilever bend due to tip-sample interactions. The bending of the cantilever is measured by the displacement of the reflected laser spot on the photodetector. Using a control loop, the actuators are commanded to perform the raster scan in the xy direction and to follow the sample height in the z direction.

2.2 Four-quadrant detection

To detect the beam deflection, a four-quadrant photodetector is used. It consists of four individual silicon photodetectors, i.e. photodiodes, that provide a voltage response proportional to the light intensity. These photodetectors are aligned in a plane with a small gap between them, as shown in Fig. 1b. The detector is initialised by centering the laser spot, when the difference of the signal of each quadrant is zero while the sum serves as the total amount of incident light, called *light level*.

As the cantilever bends, the laser beam is deflected and the spot moves off the center position, causing the difference signal to increase. Due to the measurement of the difference, this setup is very sensitive to the displacement of the spot, while the light level is used to correct a change in the laser intensity over time. In
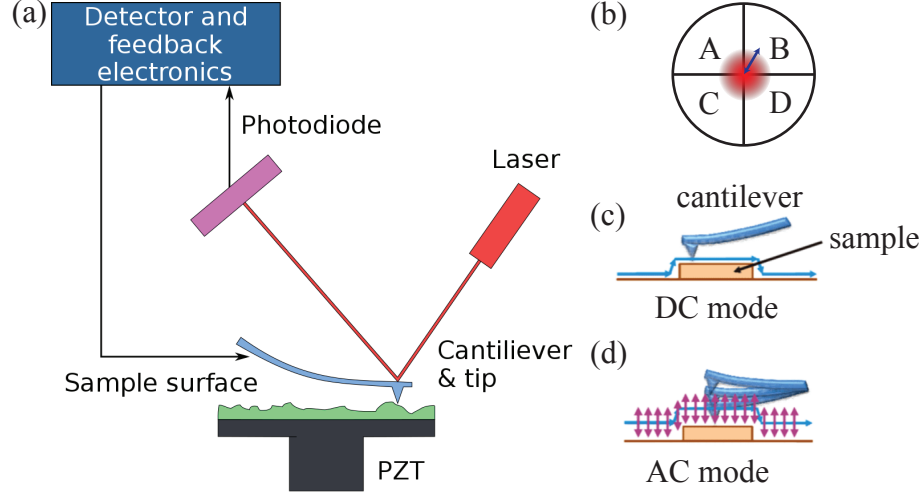
Figure 1: Basic AFM set-up and Modes of Operation.[?] (a) Schematics of an Atomic Force Microscope system: a scanning tip on an elastic cantilever with a reflective coating on the bakcside, a laser beam, a four-quadrant photodetector, the peripheral electrical circuitry, and piezo actuators (PZT) enabling horizontal and vertical movement of the cantilever. (b) The four-quadrant detector with the centered laser beam. This type of detector can detec very small beam deflections, as represended by the blue arrow. (c) Schematic representation of the DC or contact mode. (d) AC or tapping mode.

our case, the displacement is only interesting in one direction, therefore, the following *bending signal* is fed to the control loop:

$$b(t) = \frac{(V_A + V_B) - (V_C + V_D)}{V_A + V_B + V_C + V_D}.$$

2.3 DC mode

The simplest mode of operation is often referred to as *DC mode*, shown in Fig. 1c. The tip is pushed into the sample and dragged over the scanning area in the xy plane while the z actuator maintain constant the cantilever bending, i.e. the tip–sample force and therefore their distance: once the tip gets close to a hump during its movement, the bending signal gets larger (the interaction force increases) and the tip must be elevated. The opposite case is when it runs above a trench, then the bending signal decreases (since the interaction force decreases). The tip must be pushed closer to the sample to keep the interaction strength constant.

2.4 PI Controller

This positional fine-tuning is realised via a *proportional-integral (PI) controller*. Let b_0 be the desired amount of bending, called *set-point* signal, and $b(t)$ the actual bending signal called *reference signal* as the scanning in the xy-plane is in progress. The *error signal* $e(t)$ is calculated as

$$e(t) = b(t) - b_0.$$

The goal of the PI controller in DC mode is to ensure the constant bending, thus, it adjusts the z position according to the equation

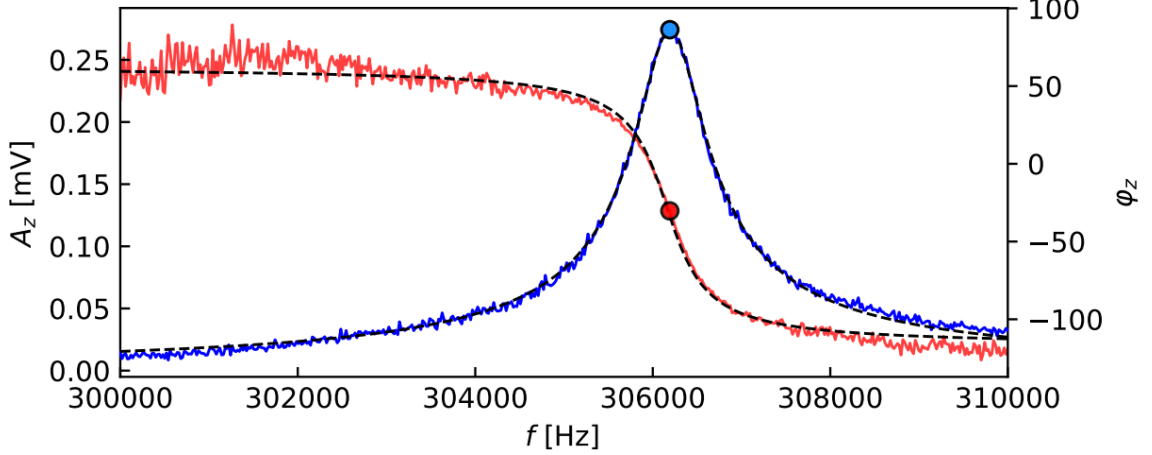$$z(t) = P \cdot e(t) + I \cdot \int_0^t e(t') \, dt'.$$

2

Figure 2: Initial frequency sweeps to determine the resonance frequency f_0. If the machine is properly set, the regular Lorentzian resonance curve is measured.

The accuracy and the speed of the control is set by the parameters P and I. By performing a raster scan with constant height adjustment, this leads to a raised relief map $z(x, y)$, which corresponds to the surface topography.

2.5 AC mode

A bit more complicated is the *AC mode*, shown in Fig. 1d, which has multiple subvariants. If the cantilever is subject to a harmonic excitation $V_{\mathrm{drive}}(t) = A_{\mathrm{drive}} \cdot \sin(2\pi f_{\mathrm{drive}} \cdot t)$, then a harmonic voltage signal with negligible phase delay is measured as a bending signal

$$b(t) = A_{\mathrm{osc}}(t) \cdot \sin\left(2\pi f_{\mathrm{osc}}(t) \cdot t + \Delta\varphi\right) \quad \Delta\varphi \approx 0.$$

Time-dependence of the amplitude and frequency of the oscillation represent the fact, that both depend on the tip–sample distance. They are both ideal candidates as feedback variables. One of these parameters can be selected together with a set-point and their difference should be kept constant: in the amplitude-modulated (AM–AFM) case the error signal reads as $e(t) = A_{\mathrm{osc}}(t) - A_0$, whereas in the frequency-modulated case (FM–AFM) it is simply $e(t) = f_{\mathrm{osc}}(t) - f_0$. And yet again the continuous adjustment of the vertical position enables topography mapping, while the sudden changes in the map of the feedback variable, which inevitably occur, may provide extra information.

At this special laboratory event, the AM–AFM measurement is realized, which is often referred to as *tapping mode*, since the tip–sample contact is reduced, it only occurs within a small portion of the period.

2.6 Sweeping Procedure

To achieve maximal oscillation amplitude the cantilever must be excited at its resonant frequency. This is easily determined: far away from the sample in the z direction, a sinusoidal frequency sweep is performed while the amplitude and phase response are simultaneously measured using the lock-in principle. Such example curves can be seen in Fig. 2.

These data sets were fitted via the regular harmonic oscillator functions

$$A(f; A_{\mathrm{fit}}, f_0, Q, A_0) = A_{\mathrm{fit}} \frac{f}{\sqrt{f^2 + \left(\frac{Q}{f_0}\right)^2 \cdot (f^2 - f_0^2)^2}} + A_0$$
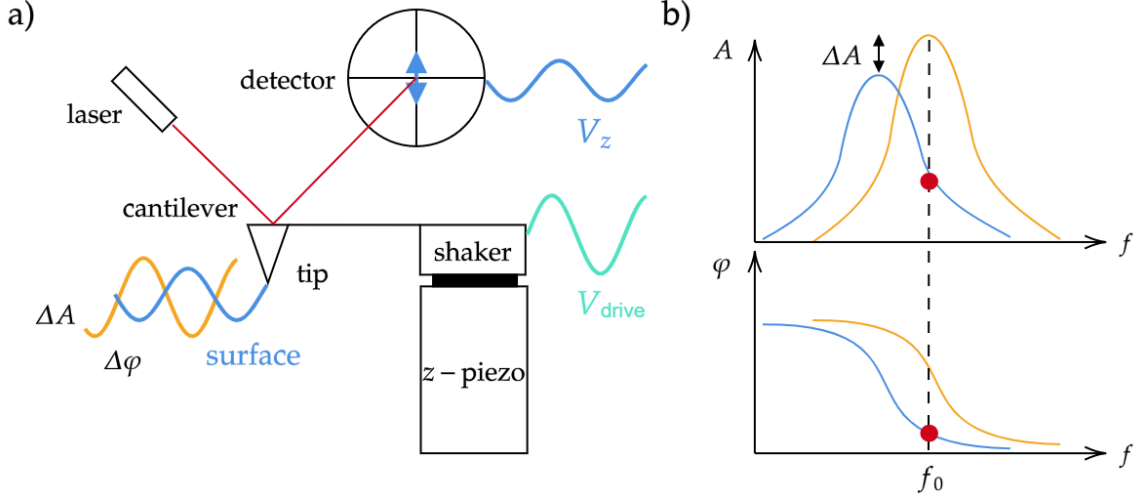
and

3

Figure 3: Effect of decreasing sample–tip distance in a dynamic AFM measurement. (a) The sensing schematics with a harmonically driven piezo shaker (V_{drive}), inducing oscillation far away (orange) and near the sample (blue). The oscillation as a harmonically modulated bending is measured by the photodetector via the laser sensing, the output is a voltage signal (V_z). (b) Measuring in the attractive regime induces a shift to the resonance curve towards the smaller frequencies, which is sensed by the decrease in oscillation amplitude.

$$\varphi(f; f_0, Q, \varphi_0) = \arctan\left(Q\frac{1 - \left(\frac{f}{f_0}\right)^2}{\frac{f}{f_0}}\right) + \varphi_0.$$

This way, the f_0 resonance frequency and A resonance amplitude are both determined, alongside the Q quality factor of the oscillator and the constant shifts A_0 and φ_0. The latter may be used to shift the phase to zero at the resonance.

These amplitude and phase curves represent the situation when there is negligible interaction between the sample and tip, because these are separated at a macroscopic distance. The main experimental observation is that the resonance curves shift to lower frequencies once the forces become significant, as illustrated in 3. This height interval is the *attractive regime*, which is the focal point of AC imaging [1]. If the resonance curve shifts to the left (lower frequencies), then the forced oscillation at the original f_0 resonance frequency will have a smaller amplitude.

Another effect contributing to the decrease is the presence of non-conservative forces, i.e. the damping of the oscillation. In this approach the effect of the shift and the damping can not be separated, however, the net amplitude decrease can be used to define at which height we expect to reach the surface. This is done using an *amplitude set-point*, which is determined as a percentage. Setting 30 % amplitude set-point means, that the tip is set to the height z above the (x_i, y_j) point for which

$$100 \cdot \frac{A_{\text{fit}} - A(z)}{A_{\text{fit}}} = A_{\text{sp}} = 70\%.$$

In summary, the amplitude set-point is the reference, the amount of decrease in the oscillation amplitude which characterizes the surface.

Once the set-point is chosen and the surface is reached, the scanning may be started. The PI regulator circuit will continuously measure the A_{sp} damping while keeping a record of the necessary $\Delta z(x_i, y_j)$ changes.

---

[1] It is important to note that further approaching the tip will lead to a shift towards higher frequencies. This happens in the *repulsive regime*
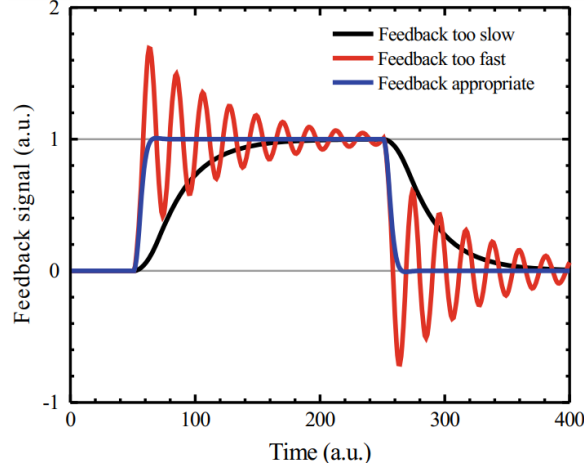
Figure 4: "Schematic (exaggerated) step response of the AFM feedback signal to a modulation of the setpoint with a square wave. For too-slow feedback settings (black line), too fast feedback settings (red line), and appropriate feedback settings (blue line)." Image and caption adapted from Ref. **?**.

The main ideas of the measurement are this simple, however, the quality of the measured topography is highly dependent on the settings of the feedback circuit.

2.7   Parameter Setting

To thoroughly understand the effect of the parameter settings, let us briefly revise what happens in an over (small P, large I)- or under-damped (large P, small I) feedback circuit.

At the beginning of the raster scan, the cantilever is oscillating on the surface, at the amplitude set-point $100 \cdot -\Delta A/A \equiv A_{\mathrm{sp}}$, thus the PI controller's output indicates no change in the z position. Once a high enough surface element is met by the cantilever, the oscillation gets damped. The error signal is said to be proportional to the ΔA changes in amplitude: its proportional (P) component suddenly increases but it is immediately suppressed by the small value of P. In addition, due to the large timescale of integration, the error signal is smoothed over in the integral component (I). Once the integral part accumulates enough large error inputs, the cantilever gets elevated and the amplitude increases back to the set-point value. For trench-like structures the opposite happens: above the trench, there is no damping due to the increased tip-sample distance, thus the oscillation amplitude gets larger. Similarly, the P component is not dominant enough and the integral part delays the response. These slow feedback settings may even make the cantilever get into continuous contact with the surface. As a result, the topography image gets smeared, and the contrast around the edges of sample structures becomes significantly lower.

The solution seems straightforward: let's increase the proportional part. This is reasonable up to a certain extent. Once the feedback becomes too fast, even the simple electrical noise in the error signal is enough to induce unwanted high-frequency z modulation. This means that besides being excited by the shaker piezo, the cantilever is dragged up and down by the z piezo as the feedback circuit struggles with minimizing the error signal and the resulting image becomes noisy. In this case, the P parameter must be decreased up to the point, where no large fluctuation is present. The schematic output for too fast, too slow, and ideally set PI circuits, when the tip encounters a positive and a negative unit step, can be seen in Fig. 4.

Finding all the ideal settings in an AFM device for a given sample and tip system is more like an art, but general guidelines can be identified. Fig. 5 provides a great outline of the most common imaging artifacts simulated for a rectangular reference sample. Studying these will help you to find and eliminate the cause of faulty imaging, moreover, the interested reader is highly encouraged to read through the work of Canale et. al. (Ref. **?**), where all the frequently occurring instances are covered in great detail.
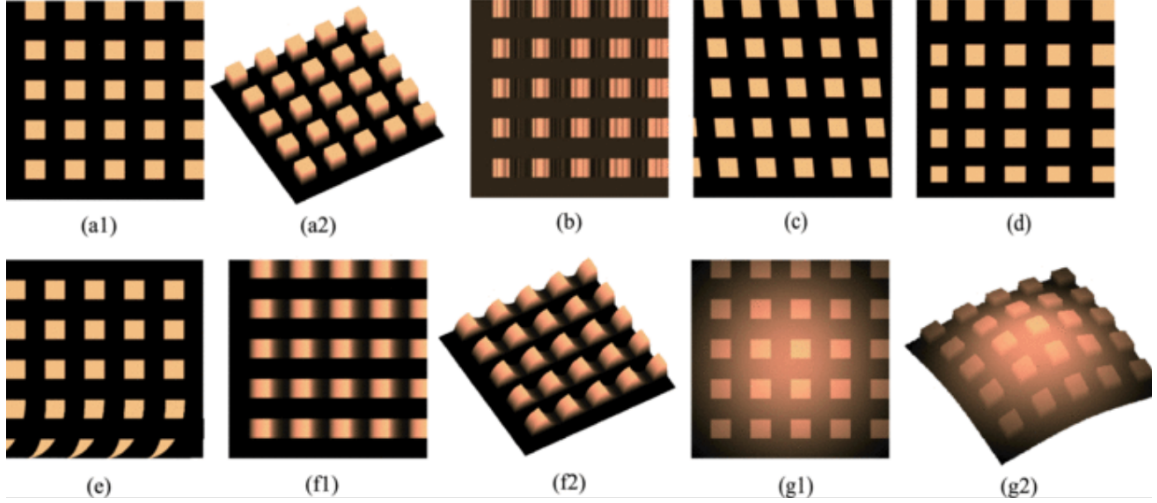
Figure 5: "Simulated AFM image artifacts. (a1 and a2) Ideal grating without artifacts. (b) Artifacts caused by vibrations, (c) x-to-y cross-coupling motions, (d) the presence of hysteresis at the lateral axes, (e) the creep and drift effect, (f1 and f2) insufficient vertical feedback controller bandwidth, and (g1 and g2) xy-to-z cross-coupling motions." Image and caption adapted from Ref. **?**.

3 2D crystals and heterostructures

In recent years, 2D crystals became one of the most researched areas in nanophysics.**?** Originated from the discovery of graphene, many more layered crystals became available in a single or few atomic layer thick form, which are interesting on their own, but also can be stacked on top of each other creating a new, artificial material that may exhibit properties unprecedented in the component crystals alone or simply that unite their advantageous properties. Since the layers are stabilised by the van der Waals interaction, these nanostructures are called *van der Waals heterostructures*. In the following, we briefly discuss two different 2D crystals that are important building blocks of them. We will investigate these materials during this event.

3.1 Graphene

After the discovery of graphene, its remarkable mechanical, chemical and electronical properties quickly made it a promising material in various applications as well as a good tool in the fundamental research.

It is a single layer of a hexagonal (or honeycomb) lattice of carbon atoms, which is a trigonal Bravais lattice with a two-atom unit cell, called A and B sites, see Fig. 6a. Its Brillouin zone is also hexagonal, where the most important regions are the corners of the zone. Of the six corner points, there are two non-equivalent ones, called K and K'. Electronically, it is a semimetal with charge carriers that are massless Dirac fermions. Semimetal means that there is no band gap between the conduction band and the valence band. The touching points are the Brillouin zone corners (K and K'). Massless Dirac fermions means that the dispersion relation is linear both for electrons and holes in the vicinity of these points, which are called therefore *Dirac points*, see Fig. 6b.

For practical electronical applications, the most important property of a graphene sample is its charge carrier mobility as defined in $v = \mu E$, where v is the electron velocity in a quasi-classical picture and E is the electric field. This is closely related to the mean free path in the material, i.e. how often the electrons are scattered. Among various conducting media, graphene can have an exceptionally high mobility, but, since graphene is a 2D material, any kind of on-surface contamination acts as a scatterer to the conduction electrons and leads to a reduction of the mobility, therefore, they need to be avoided.

Contamination is inherently present in the environment as the graphene sample is directly exposed to various organic chemicals during the processing of a sample into a nanocircuit. Therefore, a protection
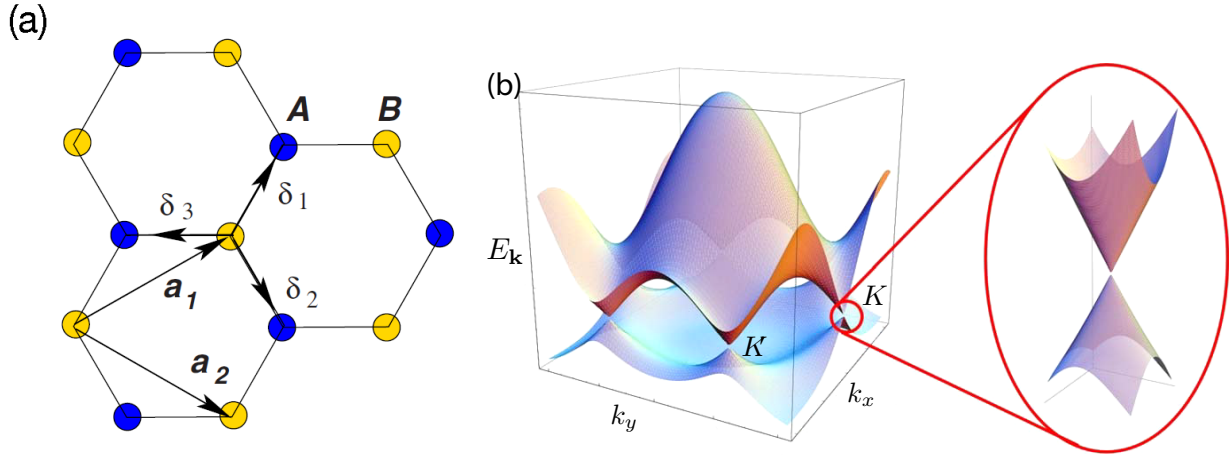
Figure 6: (a) Real space lattice of a graphene sheet. The hexagonal structure corresponds to a diatomic trigonal lattice with lattice vectors $\mathbf{a}_1$ and $\mathbf{a}_2$. (b) The band structure of graphene with the two non-equivalent Dirac points K and K'. Inset: close-up of the quasi-linear touching bands near the K point. Figure adapted from Ref. **?**.

mechanism is needed, which is discussed next.

3.2 Hexagonal boron nitride (hBN)

This material has also a hexagonal crystal lattice similarly to graphene, but with nitrogen and boron atoms on the A and B sublattices, respectively. Since there is no symmetry between the atomic sites anymore, the material exhibits a large band gap at the K and K' points and is therefore an insulator. Chemically and mechanically very stable, this material is able to protect the underlying graphene flake from any harm or contamination from the environment, and preserve its electronic quality. In recent years, surrounding the graphene flake from above and below by hBN, called "hBN encapsulated graphene structures", became a common practice in the research of van der Waals heterostructures.

In the following, we discuss the activities during this lab event. We will produce, locate, and investigate graphene and hBN crystal flakes as described below.

3.3 Mechanical exfoliation

2D crystal flakes can be produced via various methods including mechanical exfoliation, chemical vapour deposition, or intercalation among others; each having its own pros and cons. We focus on the *mechanical exfoliation* method, which yields the highest quality and single crystal flakes, but each exfoliated flake has to be identified individually.

The usual process of mechanical exfoliation is quite simple: put a piece of graphite (or other bulk crystal to cleave) on a strip of adhesive tape (see Fig. 7), fold a few times, then push the tape on a standard Si/SiO_2 chip substrate, and voilà, the graphene is on the substrate ready to be located with an optical microscope. However, many small details matter that we will discuss during the event.

3.4 Optical investigation

Identification and location of the flakes is conventionally done by using optical microscopy. Optical visibility of the flakes is not straightforward, and a good optical contrast depends on the thickness of the oxide layer underneath.
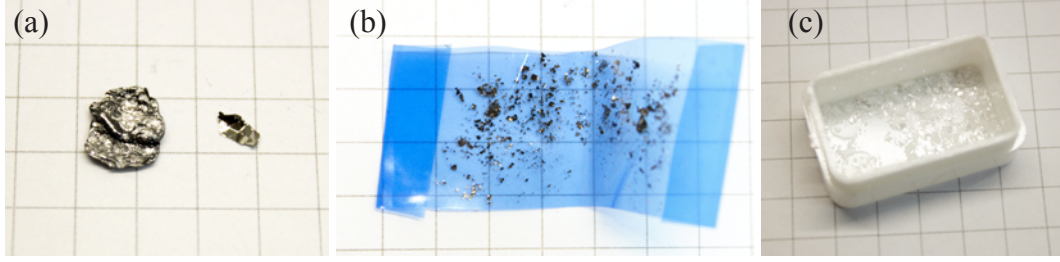
Figure 7: (a) Graphite crystals ready for exfoliation. Try to choose a shiny one! The grid in the background corresponds to 5 mm in each picture. (b) A "blue" adhesive tape with graphite crystals on it after a few foldings. (c) Bulk hBN crystals in a plastic carrier. One piece of this sugar-like crystal is enough for one (or multiple) exfoliation yielding hundreds of potential flakes.

The visibility of a, e.g., graphene flake is a result of an interference effect of the light beams reflected from different layer interfaces in the graphene/SiO$_2$/Si structure. To model this effect, we consider the case of normal light incidence from air (refractive index, $n_0 = 1$) onto the structure.[?] The lowermost Si layer is considered semi-infinite with a complex refractive index $n_3(\lambda)$ that, importantly, is dependent on λ, (for example, $n_3(\lambda = 400\text{nm}) \approx 5.6 - 0.4i$). The SiO$_2$ layer is described by thickness d_2 and another λ-dependent but real refractive index $n_2(\lambda)$ ($n_2(400\text{nm}) \approx 1.47$). These are considered known. The graphene sheet has a thickness $d_1 = 0.34$ nm (which is equal to the extension of the π orbitals out of plane) and its refractive index $n_1(\lambda)$ is described by that of bulk graphite $n_1(\lambda) \approx 2.6 - 1.3i$.

Therefore, the reflected light intensity can be written as:

$$
\begin{aligned}
I(n_1) =\ & \Bigg| \left(r_1 e^{i(\Phi_1+\Phi_2)} + r_2 e^{-i(\Phi_1-\Phi_2)} \right. \\
& + r_3 e^{-i(\Phi_1+\Phi_2)} + r_1 r_2 r_3 e^{i(\Phi_1-\Phi_2)} \Big) \\
& \times \left(e^{i(\Phi_1+\Phi_2)} + r_1 r_2 e^{-i(\Phi_1-\Phi_2)} \right. \\
& + r_1 r_3 e^{-i(\Phi_1+\Phi_2)} + r_2 r_3 e^{i(\Phi_1-\Phi_2)} \Big)^{-1} \Bigg|^2 ,
\end{aligned}
\tag{1}
$$

where

$$
r_1 = \frac{n_0 - n_1}{n_0 + n_1}, \quad r_2 = \frac{n_1 - n_2}{n_1 + n_2}, \quad r_3 = \frac{n_2 - n_3}{n_2 + n_3}
\tag{2}
$$

are the relative indices of refraction. $\Phi_1 = 2\pi n_1 d_1/\lambda$ and $\Phi_2 = 2\pi n_2 d_2/\lambda$ are the phase shifts due to changes in the optical path. The contrast C is defined as the relative intensity of reflected light in the presence ($n_1 \neq 1$) and absence ($n_1 = n_0 = 1$) of graphene:

$$
C = \frac{I(n_1 = 1) - I(n_1)}{I(n_1 = 1)}.
\tag{3}
$$

Evaluating this calculation for the visible light wavelength range, one can obtain a contrast map as a function of oxide layer thickness, as shown in Fig. 8. Although multiple oxide layer thicknesses can produce high optical contrast for graphene, due to a few practical considerations, substrates with an oxide layer of 290 nm are commonly used. Here, not only the single layer graphene flakes are visible, but they are also distinguishable from a bilayer and other, thicker flakes as well. Graphite flakes thicker than a few layer quickly become opaque and provide no information about the thickness due to the strong optical absorption of the material. Part of the current laboratory event is to familiarise with the optical appearance of these flakes.
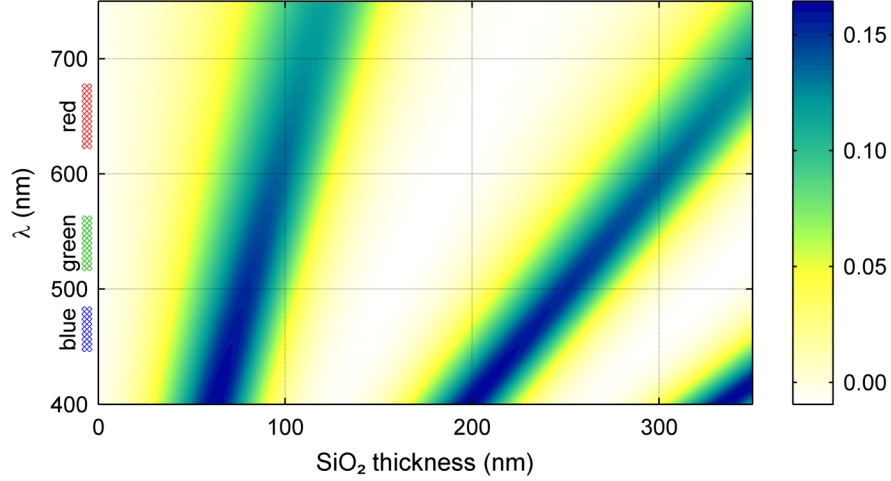
Figure 8: Optical contrast as a function of SiO_2 layer thickness and beam wavelength calculated using Eq. 3

In case of hBN, similar calculations can be made, except that due to the large band gap, which is in the deep ultraviolet range, these flakes are transparent even in their thicker form. Due to the interference condition, different thicknesses lead to a difference in the apparent colour of the flakes in the optical microscope. This can be used to identify the flake thickness via optical microscopy, but the colour also depends on the SiO_2 thickness, therefore, each type of substrate needs to be recalibrated, if one needs precise thickness information (which we do).

Therefore, one of our task will be to investigate the colour-to-thickness relation for hBN on the actual substrate used in the laboratory event.

4   Measurement tasks

- Exfoliate graphene and hBN flakes on a Si/SiO_2 substrate.

- Locate thin and potentially interesting flakes on the chip surface using an optical microscope.

- Take pictures of the selected flakes and their vicinity. Pay attention to the color settings and the potential caveats.

- Load the chips into the Semilab AFM-101 instrument and find the flakes.

- Record the surface topography of the exfoliated flakes and its vicinity. Look for surface contamination. Measure the thickness of the flakes. Do this for multiple flakes.

- Investigate the flake thickness at different amplitude and PID settings. It is known that due to the difference in the material quality, a graphene layer on the SiO_2 surface can be measured with difficulty.[?]

- Try to establish a color-to-thickness relation for the hBN flakes.