PhD Thesis

Resistive switching in ultrasmall nanogap devices

László Pósa

Supervisor: Dr. András Halbritter
Department of Physics
Budapest University of Technology and Economics

BUTE
2019
1 Introduction 4

2 Overview of the research field 7
   2.1 Resistive Switches ................................. 7
      2.1.1 Mechanism of the resistive switching ............ 9
      2.1.2 Application of resistive switching devices ...... 13
      2.1.3 Resistive switching in SiO$_2$ .................... 19
      2.1.4 Resistive switching in Ag$_2$S .................... 24
   2.2 Fabrication of nanometer-sized gaps .................. 26
      2.2.1 Electromigration of metal nanowires ............. 27
      2.2.2 Electrical breakdown of graphene nanoribbons .... 29
      2.2.3 Charge transport through dielectrics ............ 31
   2.3 Charge and heat transport in graphene ............... 35
      2.3.1 Electrical properties of graphene ............... 35
      2.3.2 Thermal properties of graphene, energy dissipation 38

3 Experimental techniques 41
   3.1 Mechanical setup .................................. 42
   3.2 Electrical setup .................................. 44
   3.3 Measurement control program ....................... 45
   3.4 Conclusion ....................................... 50
4 Nanofabricated Ag$_2$S memristors 51
4.1 Research history 51
4.2 Sample preparation 55
4.3 Resistive switching in asymmetrically shaped Ag-Ag$_2$S-Ag on-chip memristors 59
4.4 Conclusions 63

5 Nanogap formation in graphene 65
5.1 Sample preparation and experimental techniques 66
5.1.1 CVD growth and transfer 66
5.1.2 Sample design and patterning 70
5.1.3 Electrical characterization and cleaning 72
5.2 High yield formation of nanogaps 78
5.2.1 The electrical breakdown process 78
5.2.2 Characterization of the nanogaps 78
5.3 Mechanism of the breakdown 86
5.4 Controlled electrobreakdown 97
5.5 Conclusions 101

6 Nanometer-sized SiO$_x$ resistive switches 103
6.1 The formation process of SiO$_x$ resistive switches 103
6.2 Real time response of SiO$_x$ resistive switches 107
6.2.1 Dead time rule 108
6.2.2 Multiple timescales in SiO$_x$ switches 114
6.3 Further characteristics of the switching 115
6.4 Conclusions 117

7 Summary 119

8 Acknowledgments 122

Publications related to the thesis statements 124

References 125
List of symbols and abbreviations

\begin{itemize}
    \item $A$: Surface
    \item $A_g$: Cross-section of graphene
    \item $d$: Gap size
    \item $g$: Thermal conductance
    \item $i, I$: Electric current
    \item $j_{ox}$: Current density of oxygen molecules
    \item $k_B$: Boltzmann constant
    \item $L$: Length of graphene stripe
    \item $L_H$: Thermal healing length
    \item $M$: Memristance
    \item $n_{ox}$: Molar concentration of oxygen molecules
    \item $p$: Pressure
    \item $p_{ambient}$: Pressure
    \item $p_x$: Joule-heating power per unit length
    \item $q$: Electric charge
    \item $R_{high}$: High voltage resistance
    \item $R_j$: Junction Resistance
    \item $R_{low}$: Low voltage resistance
    \item $R_s$: Series Resistance
    \item $t$: Nearest-neighbour hopping energy
    \item $t'$: Next nearest neighbour hopping energy
    \item $t_g$: Thickness of graphene
    \item $t_{ni}$: Thickness of Si$_3$N$_4$
\end{itemize}
\( t_{ox} \)  
Thickness of SiO\(_2\)

\( T \)  
Temperature

\( T_0 \)  
Room temperature

\( v \)  
Voltage

\( \bar{v} \)  
Averaged velocity of oxygen molecules

\( V_{\text{high}} \)  
High voltage level

\( V_{\text{low}} \)  
Low voltage level

\( V_{\text{reset}} \)  
Reset voltage

\( V_{\text{set}} \)  
Set voltage

\( V_{\text{step}} \)  
Bias step size

\( V^+, V^- \)  
Positive and negative threshold voltage

\( W \)  
Width

\( \kappa_g \)  
Thermal conductivity of graphene

\( \kappa_{ni} \)  
Thermal conductivity of Si\(_3\)N\(_4\)

\( \kappa_{ox} \)  
Thermal conductivity of SiO\(_2\)

\( \mu \)  
Mass of an oxygen molecule

\( \rho_{gox} \)  
G-SiO\(_2\) thermal boundary resistivity

\( \phi \)  
Magnetic flux

\( \Phi \)  
Potential barrier height

ANN  
Artificial neural networks

BUTE  
Budapest University of Technology and Economics

CMOS  
Complementary Metal-Oxide-Semiconductor

CNN  
Cellular Neural Network

CNP  
Charge Neutrality Point

CNT  
Carbon Nanotube

CVD  
Chemical Vapor Deposition

DAQ  
Data Acquisition Card

EB  
Electrobreakdown

EBL  
Electron Beam Lithography

ECM  
Electrochemical Metallization

ESD  
Electrostatic Discharge

FPGA  
Field-Programmable Gate Array

HRS  
High Resistance State

IPA  
Isopropyl Alcohol

LRS  
Low Resistance State

LTP  
Long Term Potentiation

MCBJ  
Mechanically Controllable Break Junction

PCB  
Printed Circuit Board
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMMA</td>
<td>Poly(Methyl Methacrylate)</td>
</tr>
<tr>
<td>RRAM</td>
<td>Resistive Random Access Memory</td>
</tr>
<tr>
<td>STDP</td>
<td>Spike-Timing Dependent Plasticity</td>
</tr>
<tr>
<td>STM</td>
<td>Scanning Tunneling Microscope</td>
</tr>
<tr>
<td>STP</td>
<td>Short Term Plasticity</td>
</tr>
</tbody>
</table>
In the last few decades there is increasing demand for data storage capacity. Nowadays more than $10^{18}$ (quintillion) bytes of data is generated daily. Data intensive applications are spreading to more and more fields and sectors such as economics, business, social media or health care, extracting useful, hidden information from the large datasets. These applications require intensive data transfer back and forth between the processors and memories, indeed most of the processing time consists of data manipulation and I/O operations. However, today’s complementary metal-oxide-semiconductor (CMOS) based architectures are less and less capable to process this amount of data. Due to the memory-access bottleneck, the processors spend a lot of time waiting for the data. In addition the memory access and communication cause large energy consumption. The need to store, process, analyze or link this large amount of data brings great challenges. New architectures are needed for data intensive applications which integrate the data storage and processing to the same platform. Moreover the CMOS memory devices are also facing reliability problems upon further downscaling. In the last few decades many different concepts have been proposed to overcome the limitations of flash memories or dynamic random access memories (DRAM).

The non-volatile, two-terminal resistive switches offer low power consumption and highly scalable alternatives for the semiconductor memory devices. One of their main advantages lies in the integration of the processor and the memory, since they can be used both for data storage and processing. The crossbar architecture of resistive switching devices would support massive computation parallelism. Moreover, the resistive switches are ideal building blocks for neuromorphic computing by mimicking the characteristics of synapses. The latter application would dramatically decrease the hardware cost of neural networks.
The goal of my research is to realize truly nanometer-scale resistive switching systems using electron-beam lithography technique. The on-chip implementation provides enhanced mechanical stability, better integration into more complex systems and in addition it is a prerequisite for commercial applications. During my work I focused on the fabrication of resistive switches whose size is smaller than the resolution of the standard lithography process, that is 10 nm. For this purpose I developed a measurement setup and control program which allows us to break electrically an initially continuous metal or graphene wire such that the spacing of the two broken electrodes is less than 5 nm. In this thesis the few nanometer-sized gap between the conductive electrodes is referred as nanogap. Using this technique the lower size limit of the resistive switches can be tested. Graphene has a central role in my PhD work, since it has atomic structural stability even at room temperature. Furthermore, graphene is chemically inert, does not contaminate the resistive switches by metal atoms, flexible and transparent. These properties make graphene a perfect electrode material.

In Chapter 2 at first I present an overview about the resistive switches by discussing their terminology, the classification of the switching mechanisms and their possible applications. A special emphasis is put on the discussion of Ag$_2$S and SiO$_2$ systems. Afterwards, I discuss the fabrication protocol of nanometer-sized gaps using a controlled electrobreakdown (EB) technique, devoting a special attention to the nanogap formation in graphene and its research history. The electrical characterization of the formed nanogaps is also presented. Finally, I give a short introduction to the electrical and thermal properties of graphene.

In Chapter 3 I introduce a novel measurement and fabrication setup which allows us to reduce the size scale of the device below the 10 nm regime using controlled electrical breakdown technique.

In Chapter 4 I introduce my research contributions to the geometrical asymmetry induced filament formation in Ag$_2$S resistive switching system. I describe the operation and the optimization steps of the first nanofabricated resistive switching devices of our group.

Chapter 5 is devoted to the formation of nanometer-sized gaps in graphene constrictions. After the introduction of sample fabrication process I demonstrate a high yield fabrication technique of sub-5 nm sized nanogaps and the characterization techniques to confirm their size and purity. Afterwards, the mechanisms of the graphene breakdown are studied under different environmental conditions. Finally, the optimization steps are presented to achieve feedback controllable breakdown in graphene devices.

Chapter 6 introduces my study about sub-10 nm sized SiO$_2$ based resistive switches.
I investigate the physical time scales of the switching phenomena, especially the technologically significant dead time which enables to program the unipolar resistive switches to both resistance states at zero bias. Next, I study whether the extremely small size of the switching region affects the switching parameters. Finally, in Chapter 7 I summarize my work in 4 thesis points.
In the first section of this chapter I summarize the terminology and the research field of resistive switches with special focus on SiO$_2$ and Ag$_2$S systems, also introducing their possible applications. In the next part I give an overview about how nanogap devices are established with the electrical breakdown of an initially continuous nanowire. The formed nanogap between the conductive electrodes could serve as the place of the resistive switching. In the last section I introduce the electrical and thermal properties of graphene, which could be a promising electrode material for nanometer-sized resistive switches.

2.1 Resistive Switches

A resistive switch is a passive, two terminal electric component, whose resistance can be varied by applying proper voltage to the terminals. They have at least two different resistance states that can be altered reversibly. Therefore, they show nonlinear resistance behavior, but in contrast to conventional nonlinear resistors the resistive switches preserve their new resistance state even if the voltage is released. It means that resistive switches exhibit non-volatile memory behavior, and since they are passive elements, no external power supply is needed to store the memory state.

In many cases the resistive switches are also called memristors or memristive systems. The phrase of "memristor" is the concatenation of memory and resistor. It was theoretically predicted by Chua in 1971 as the fourth fundamental circuit element besides the resistor, capacitor and inductor [5]. These elements provide connection between a pair of four basic variables: electric charge ($q$), electric current ($i$), voltage ($v$) and magnetic flux ($\phi$), as it is illustrated in Figure 2.1. Two further equations
come from the definition of current and from the Faraday’s law. The memristor completes the symmetry consideration by introducing the memristance \( M \) which connects the charge \( q \) and the magnetic flux \( \phi \), in the following way:

\[
M(q) = \frac{d\phi(q)}{dq},
\]

where \( M \) is the memristance, its unit is Ohm (\( \Omega \)). Using the \( v(t) = d\phi(q)/dt \) and \( i(t) = dq/dt \) definitions, the previous equation can be rewritten to the following form:

\[
M(q(t)) = \frac{v(t)}{i(t)}.
\]

If the memristance \( M \) is constant, the memristor is a linear resistor. The interesting behavior starts with the nonlinear \( \phi - q \) relationship. If we write the charge as the integration of the current we get a more expressive equation:

\[
M \left( \int_{-\infty}^{t} i(\tau)d\tau \right) = \frac{v(t)}{i(t)}.
\]

Now, it is clearly seen that the actual value of the resistance depends on the current flow through the device in the past. The memristor remembers its past and keeps its resistance even after the external voltage (or current) drive is released. The \( \phi - q \) relationship is unique, from the same initial state if the same amount of charge flows through the device, we get the same final state.

Years later Chua generalized his theory by introducing the memristive systems, which are independent of the magnetic flux [7]:

\[
i(t) = G(w, v)v(t),
\]

\[
\dot{w} = f(w, v)
\]

The resistance of a memristive system is governed by the external inputs \( v \) and a set of internal state variables \( w \) (Equation 2.4). The actual values of the internal variables depend on their value in the past and on the inputs (Equation 2.5). Chua later showed that all two-terminal, non-volatile resistive switching memories are memristors regardless of the device materials and operating mechanism [8]. Many groups use simultaneously the term of memristor and memristive system for the resistive switches. The fabricated resistance change memory devices actually differ from the ideal memristor and they belong to the more general group of memristive
systems [9]. In this thesis the phrase of memristor, memristive system and resistive switches are used as synonyms and refer to the devices whose resistance can be varied reversibly depending on the applied voltage.

The first device which was presented as a memristor was manufactured by HP in 2008, which was based on a titanium-dioxide thin film layer [10]. Soon the research field of the memristors became very attractive due to their promising future. The resistance change ability was discovered in many other materials and nowadays the memristive systems have very wide literature.

It has to be noted that the resistive switching effect has been actually investigated since the 1960s. The first studies were performed in Al$_2$O$_3$ [11], SiO [12], NiO [13], Nb$_2$O$_5$ [14], Ta$_2$O$_5$ [15], however the formalism of the memristive systems was not applied. The robustness of these devices were not sufficient for any practical applications. In the early 2000s the interest on resistive switching rose again, among others the research group of Sharp [16] and Samsung [17] were also working on memristors, but these devices were referred as Resistive Random Access Memories (RRAM).

### 2.1.1 Mechanism of the resistive switching

Over the last decade a large variety of resistive switches were manufactured using several material compositions and designs. These various memristive systems can not be described with the same or at least similar models because their working
mechanisms can be fundamentally different. Their operation not only depends on the material of the active region but also on the electrode materials and their combination [18].

The resistance change memories have typically a capacitor-like structure, where a thin insulator film is sandwiched by two electrically conductive electrodes (see Figure 2.1.b). The electrodes may be made of metals, semiconductors or other conductive material such as carbon nanotube (CNT), graphene, amorphous carbon or indium tin oxide (ITO). Regarding the role of the electrodes intrinsic and extrinsic resistive switching can be distinguished. In the intrinsic case the electrode material has no function expect of serving good electrical contact. The resistance transition is the result of atomic rearrangements in the insulator layer. In contrast, during the extrinsic switching at least one of the electrodes plays an active role. The electrode material fuels the insulator layer by metal ions, forming filament inside the insulator [19].

Figure 2.1.b shows the schematics of the different states of the resistive switches. The freshly made device (1) has a high resistance due to the insulator layer, at first a formation process (2) is needed to establish a reversibly switchable device. For most memristive systems it is achieved by applying a high enough voltage for a longer time and therefore it is called electroforming. As a result of the high electric field a conductive channel forms (red) through the dielectric (blue) connecting the conductive electrodes (orange) [20]. It is important to make a difference between the electroforming and the dielectric breakdown. In the latter case a permanent reduction in the resistance occurs, switching back to the original state is not possible. According to the mechanism of the forming process and the structures of the conductive channels we distinguish several types of resistive switching memories.

The transition from high resistance state (HRS) to low resistance state (LRS) is generally called set process (4) while the reversed transition is called reset process (3). The switching modes can be classified into two types: unipolar or bipolar. In unipolar case the switching happens if the voltage reached a threshold value regardless of the polarity. In most cases the reset process happens at lower voltage level (reset voltage, \( V_{reset} \)) and the device switches to the LRS at higher voltage (set voltage, \( V_{set} \)). This kind of switching characteristic is shown in Figure 2.2.a. The initial state of the device is the HRS which is also called the OFF state. If we sweep up the bias voltage the current is almost zero until the bias reaches \( V_{set} \). Then the current abruptly increases and the memory switches to the LRS or the ON state. In order to prevent the device from a permanent damage, current compliance has to be applied during the set process. Increasing the bias again from zero voltage, the current also increases showing low resistance. If the current compliance is removed,
the high current density destroys the conductive filament at $V_{\text{reset}}$ and the device switches to the HRS again.

Less often, the set voltage is lower than the reset voltage (see in Figure 2.2.b) and current limitation is not needed, which is a great advantage. However, in this case only the ON state is available at low signal level. Unipolar switching does not show voltage polarity dependency, the same characteristic behavior can be seen at both voltage polarities. Typically the HRS of an unipolar device has very large resistance compared to the LRS, exhibiting large OFF/ON resistance ratio, even $\approx 10^4$-$10^5$.

In case of bipolar switching the set and the reset processes are induced at different polarity. In this case a certain intrinsic asymmetry is required. This asymmetry can stem from the different electrode materials [21], different geometry of the electrodes [2] or inducing inhomogeneity in the dielectric layer during the deposition [22] or during the electroforming process [23]. This switching mode is illustrated in Figure 2.2.c. The set event occurs at positive voltage polarity at the threshold voltage of $V_{\text{th}}^+$ and the reset event occurs at $V_{\text{th}}^-$ at the opposite polarity. The two threshold voltages typically are not equal in magnitude. In case of bipolar switching usually the resistance ratio is not so high as for unipolar switches.

![Figure 2.2: Illustrations of I-V characteristics for the a-b) unipolar and c) bipolar switching modes. In unipolar case according to the relative value of the switching voltages two cases can be distinguished: a) $V_{\text{set}} > V_{\text{reset}}$ and b) $V_{\text{set}} < V_{\text{reset}}$.](image)

Concerning the conductive path two different structures can be distinguished: filamentary and interface type switching. In case of filamentary type, the electroforming process causes soft breakdown and a conduction path evolves across the insulator layer. During the resistance transition the applied electric field tunes the structure of the filament resulting in resistance change. Using different electric signals multiple resistance states can be achieved, providing a multi-bit storage media. However the precise control of the resistance is very hard because the resistance switching is a
highly non-linear process, which means a slightly higher bias voltage induces more intensive resistance change. The filamentary switching can produce both unipolar and bipolar switching modes. Since the switching happens in a very small, localized region in the dielectric, the parameters of the switching is independent of the encapsulating device size.

In case of interface type switching, the resistance change occurs at the interface of the electrode and the insulator, typically metal oxide. The difference of the work functions between the metal and the oxide layers induces Schottky barrier. The applied external bias changes the distribution and the density of oxygen vacancies. The resistance change is the result of the variation of the height and width of the Schottky barrier, thus the parameters of the resistive switching have clear size dependence. Interface type memristors mostly show bipolar switching mode [6].

Currently the exact physical mechanisms that occur inside resistive switches are not fully understood, it is still the topic of active research. Figure 2.3 shows a possible classification of the driving forces which may take place in the resistive switching. Moreover, multiple different phenomena may exist simultaneously which makes it even harder to understand the exact process [24].

Figure 2.3: Classification of resistive switches based on the switching mechanisms. The corresponding memory technologies are also presented. Adapted from [24].

RRAM research covers a wide range of topics like materials science, investigation of the switching mechanisms, manufacturing, integration etc. There is still no material combination which would overcome all the others. One of the most promis-
ing opportunity to improve the performance and reliability of RRAMs is the better control of the manufacturing processes.

During my PhD work I studied the switching effect of SiO$_2$ which is based on phase change mechanism. In Section 2.1.3 I will give a detailed description about its switching properties and the literature background. Furthermore as an initial step to realize different kinds of memristive systems using nanofabrication techniques I made measurements on Ag$_2$S memristors as well. In this case the filament formation and destruction is attributed to electrochemical metallization. In Section 2.1.4 I will also give a brief introduction to Ag$_2$S memristors.

### 2.1.2 Application of resistive switching devices

After the pioneering study of HP in 2008 [10], the research of resistive switches significantly expanded. The obvious possibilities of their application supplemented with the simple device structure took the interest of both scientists and industry. In the last decade several application fields were proposed, which can be broadened in the future. A classification of the possible applications is sketched in Figure 2.4, suggested by Mazumder et al. [25].

As the electronic devices can be divided into two categories: analog and digital, this classification can be also used concerning the application of memristors. Under digital operation the resistive switch has two or multiple well defined states between the LRS and the HRS, while for analog usage the resistance can be tuned between the two limits (LRS and HRS) continuously.

Since the memristors are passive components, building integrated circuit using only resistive switches is not realistic, for instance they can not provide amplification. However, most of the memristors are compatible with the conventional CMOS technology, they can be combined into CMOS/memristor system. These hybrid architectures offer many new perspectives or performance enhancements [26].

#### Digital and analog systems

Both in analog and digital circuits the memristors can be used as electrically controllable resistances. In case of amplifiers, substituting the resistor in the feedback path for a memristor the gain can be programmed. The same idea can be used in voltage comparators as well, where the comparator voltage could be varied [27, 28]. The application of memristors as oscillator [29], adaptive filters [30] or ultrawideband receivers [31] is also proposed. Another promising application is building chaotic systems based on Chua’s circuit, which could be realized in smaller size using memristors [32, 33]. A device with chaotic behavior could be applied for secure
communication or cryptography purposes, such as image encryption [34]. Memristors also offer new functionality and area benefit compared to the transistor based digital configurable circuits.

The resistive switches can serve as configurable interconnects between two circuit elements. A connection can be established or broken if the bridging memristor is programmed to low or high resistance state, respectively [35]. Furthermore, the memristors are also capable of logical operations. It was demonstrated that using 3 memristors in parallel NAND operation can be executed, where 2 memristors take part in the operation and the solution is stored in the third one [36]. It is known that any Boolean logic operation can be realized in the proper network of NAND gates.

Beside the use of single memristors, there are lot of applications concerning crossbar arrays of memristors. The design of the crossbar array structure (see Figure 2.5) was proposed to create memory arrays in a very efficient way from resistive switching elements [37]. The bit and word lines are perpendicular to each other and metal-insulator-metal based memristors are located at the crossing points of the lines.

Crossbar arrays can be also used as reconfigurable logic. The field-programmable-gate-array (FPGA) contains logical blocks, which can be wired together by reconfigurable interconnects. Significant part of the FPGA is reserved for the memory that store the actual configuration and small part is used for the computation. The routing bits between the logic blocks could be stored by resistive switches located above the CMOS layer providing a huge area benefit [38, 39].
Resistive random access memories (RRAMs)

The most evident and straightforward application of the resistive switches is the memory cell. This field of application is in the most mature state because it has been in the focus of research since the early 2000s [40, 41]. There has been several branches of devices and denominations according to the exact switching mechanism, such as phase change memories (PCM) [41], resistance RAM [42], conductive bridge random access memory (CBRAM) [43, 44] etc.. Nevertheless, all of them are based on resistance change induced by electric field and can be regarded as the subsets of memristors. In this thesis RRAM is used irrespectively of the structure of the memory cell. The RRAMs are proposed to replace the flash memories, which suffer from charge loss problem as their feature size shrinks [45]. Memristors approached or exceeded the NAND flash memory cells in many parameters, there are several reports which list the switching parameters of RRAMs [46–49]. Beside the replacement of the conventional memories, they also open the door to new kinds of architectures. The flash memories were introduced into the information storage hierarchy between the DRAM and HDD. The memristors are ideal to further reduce the gap between the DRAM and NAND flash, so called storage class memory (SCM) [50] due to their high density, high speed, high endurance and non-volatility properties.

A more progressive idea is to break the von Neumann bottleneck. In the architecture of the modern computer the processing unit and the memory are separated...
and the data moves between the two units. In case of data intensive computation the performance is often degraded due to the data transfer between the two units. The CPU has to wait for the data. Moreover, the data transfer needs significant energy which enhances the power consumption as well. So far several concepts were introduced to merge the storage and the computation blocks such as Computation-In-Memory (CIM) [51, 52], iMemComp [53] or 3D integration [54, 55]. These new architectures enable parallel data intensive computations with high speed and low power consumption.

In the crossbar array of RRAMs a certain memory cell can be switched between the LRS and HRS by applying the proper voltage between the corresponding word line and bit line (see Figure 2.5.b, red arrows). However, if the selected device is in the HRS, while the adjacent ones are in the LRS then the current and the power during the operation could be much higher than expected due to the sneak current paths (see Figure 2.5.b, black arrows). To solve this problem selection devices have to be connected in series to the memory cells. A possible solution is to combine the resistive switch with a transistor, which would serve as selector and current limiter device. This structure is called 1T1R. However this structure also has a limitation because a transistor must be built under the RRAM and thus the 1T1R structure cannot be scaled down to an ideal size and keep the cost low. The device selection problem can be solved by an external diode as well (1D1R structure), since in the sneak current path the current flows in the opposite direction through one of the devices. Moreover, the diodes have simple structure which can be stacked up to the memory cell directly, supporting the high-density integration. Nevertheless, due to the unidirectional feature, diodes are only compatible with unipolar RRAM. The sneak current can be also suppressed by a device with bidirectional highly nonlinear I-V characteristics which is called selector devices and the corresponding structure is called 1S1R. Typically the resistance of these selector devices decreases exponentially by the bias voltage. Since the voltage drops on the unselected devices are lower than on the selected one, the selector devices in the sneak current path exhibit much larger resistance than the one in the selected path. This phenomena significantly reduces the sneak current and the power consumption of a switching operation. Recently different types of selector devices were investigated such as insulator-metal transition, ovonic threshold switching, tunnel barrier, etc. [46]. These elements have the same goal, increasing the non-linearity of the current voltage characteristic. However, several memristive materials also show intrinsic highly nonlinear current-voltage behavior, which allows us to save the additional selector devices [56, 57].
Neuromorphic systems

The artificial neural networks (ANN) attempt to process information like the human brain. The biological neural networks consist of neurons, which work as low power consumption computing elements. The communication between the neurons happens through electrical or chemical signals at their connection called synapse. The synapses function as memory as well, since the strength of their connectivity can vary. Learning is considered as long-term changes in the synaptic strengths. Due to the highly parallel architecture of the brain, it is much more efficient in the data-centric processing, such as real-time image recognition, speech recognition or language processing than the traditional von Neumann architecture. However, the imitation of the human brain is a very complex task, there are about hundreds of billions of neurons each with thousands of synaptic connections. Realizing ANNs is an intensively investigated field, but so far the real breakthrough were achieved with software level ANNs, like the recent success of deep neural networks in beating the best Go player in the world [58]. However, the massively parallel computation of software ANNs requires large computer clusters. The development of hardware based ANNs would strongly reduce the hardware demand of neuromorphic computing systems.

Figure 2.6: Diagram of a) single and b) multi layer feedforward neural networks with three inputs and two outputs. The outputs are given by the weighted sum of inputs. c) Single layer ANN realized by a crossbar array structure of resistive switches.

Perhaps the most promising application of resistive switches is the implementa-
tion of ANN hardware using CMOS/memristor hybrid circuits. The CMOS components would act as neurons, while the memristors would ensure the variable strength synaptic connections built on top of the CMOS layer. Figure 2.6.a-b illustrate the simplest class of ANNs, the single and multi layer feedforward neural network. In case of single-layer network (Figure 2.6.a) the two outputs are given from the weighted average of the three inputs and can be expressed by the following equation:

\[ y_j = \sigma \left( \sum_i x_i w_{ij} \right), \]  

(2.6)

where \( y_j \) is the output, \( x_i \) is the input, \( w_{ij} \) is the synaptic weight and \( \sigma \) is some nonlinear function executed by the neurons. Figure 2.6.c shows the schematic of the same 3x2 network implemented in a crossbar structure of resistive switches. The pre-neurons (inputs) and post-neurons (outputs) are represented as CMOS devices and each are connected to the different lines of the crossbar. The analog input voltages are multiplied by the resistance of the memristors. The output currents are the sum of the weighted input voltages. In multi-layer network (Figure 2.6.b) the inputs go through some hidden layers of calculation before the output, which can be also constructed by multiple crossbar structures. where \( y_j \) is the output, \( x_i \) is the input, \( w_{ij} \) is the synaptic weight and \( \sigma \) is some nonlinear function executed by the neurons. Figure 2.6.c shows the schematic of the same 3x2 network implemented in a crossbar structure of resistive switches. The pre-neurons (inputs) and post-neurons (outputs) are represented as CMOS devices and each are connected to the different lines of the crossbar. The analog input voltages are multiplied by the resistance of the memristors. The output currents are the sum of the weighted input voltages. In multi-layer network (Figure 2.6.b) the inputs go through some hidden layers of calculation before the output. It can be also constructed by multiple crossbar structures.

When the ANN works each neuron is programmed to generate periodical spikes independently. These spikes represent the information to be proceeded. The weights of the synapses are continuously modulated by the differential signal drops on its two terminals. The computation is done at the synapse and its result is automatically stored at the same time. Using similar structure of memristor arrays recently several advanced synaptic functions were demonstrated, such as spike-timing dependent plasticity (STDP) [59, 60]. The STDP determines the strength of the synapses, which based on the relative spike timing of the pre-neuron and the post-neuron. The short term plasticity (STP) and long term potentiation (LTP), which is result in the short and long term memory, were also presented on memristors [61].

It should be noted that, beside the ANNs there is another implementation of
neuromorphic architectures, the cellular neural network (CNN), where the communication is allowed between the neighbouring units only [62].

2.1.3 Resistive switching in SiO$_2$

The resistive switching capability of silicon monoxide (SiO) was discovered in the 1960s along with several other oxides in metal-insulator-metal structures by measuring repeatable negative differential resistance characteristics. [11, 12, 15]. The switching in silicon oxide can be realized by both extrinsic and intrinsic ways. In extrinsic case the insulator layer only serve as solid matrix in which the ions of the metal electrodes can diffuse. If the oxide layer is doped by enough metal ions, conductive filaments can be formed connecting the electrodes electrically. During the programming of the resistance states the cross section of this filament is tuned. Extrinsic switches have electrode dependence, it can be realized only with diffusive metals such as silver, gold or copper. Typically the devices have an asymmetric structure, only one electrode is made of these metals, the other one has a passive role. Both bipolar and unipolar switching behavior can be observed in these systems by electrochemical metallization or thermochemical effect respectively [12, 63-65]. In the intrinsic case the electrodes have no role in the switching effect, a silicon rich conductive filament evolves inside the SiO$_x$ which ensures the electrical contact between the two sides. This type of switches shows only unipolar characteristics. In some cases, when metal electrodes are used, it is hard to distinguish the extrinsic and the intrinsic switching. During my PhD I was working on the intrinsic kind of SiO$_x$ switches, therefore I give a detailed description about their main properties.

In the 1960s different models were proposed to explain the switching behavior, such as ion injection [12], conductive filaments formation [66] or electron impact ionization in the insulator [67]. The devices typically have Au-SiO$_x$-Al structure and the diffused gold atoms were attributed to the switching effect. Conductive Si-O-Si chain was also proposed when gold atoms had not been found inside the insulator layer by proton back-scattering, but the exact composition of the conductive channel was not revealed. Nevertheless, most of the properties of intrinsic SiO$_x$ switches were demonstrated such as thickness independent switching voltages and temperature dependence. The first clear intrinsic switching of SiO$_2$ was realized by Yao et al. [68] in 2009 using a metal free arrangement [69]. Later in situ transmission electron microscopy [19] revealed the formation of conductive filament made of local enrichment of silicon. The tested devices had typically vertical geometry (see Figure 2.7.a, top image), where a well defined thick oxide layer was deposited between two conductive electrodes. However, nanogap devices were also used, shown in the
bottom image of Figure 2.7.a, in which the switching region was formed between two electrodes with several 10 nm spacing. These wider nanogaps were created by either electron beam lithography or electrobreakdown process. Several kinds of materials were tested as electrodes: metals, semiconductors, ITO or different carbon allotropes as well. The behavior of the devices was independent of the electrodes. The switching sites were formed by applying high voltage (typically $> 10$ V) to the two electrodes for a longer time (electroforming). During this process the pristine, initially insulator layer converts to a switchable state.

Figure 2.7: a) Schematic of the structure of the vertical (top) [70] and planar SiO$_x$ device (down) using the broken ends of the CNT as electrodes (bottom) [71]. b) Characteristic I-V curve of a SiO$_x$ based resistive switch. The read, set and reset regions, defined by the resistance transitions, are presented at the top. This curve was recorded by myself. c) Top panel: Series of set (6 V), reset (14 V) and read (1 V) voltage pulses. The bottom panel shows the current corresponding to each read pulse [71]. d) Demonstration of resistance transition induced by 50 ns long set and reset voltage pulse. The final resistance can be tuned by varying the length of the pulse [71].

Figure 2.7.b shows the hysteretic I-V curve of an intrinsic SiO$_x$ switch measured at low frequency by me. Starting from the initially low resistance state (LRS, red curve),
when the bias voltage reaches the reset region \((V_{\text{reset}} \approx 7.6 \text{ V})\) a sudden jump occurs in the current and the device switches to the high resistance state (HRS or OFF). During the subsequent reverse voltage (blue curve) the device preserves its OFF states until the voltage enters the set region \((V_{\text{set}} \approx 4.6 \text{ V})\) where the conductance suddenly increases and the ON state is restored. Due to the unipolar characteristic the device shows the same behavior at negative polarity with the similar threshold voltages. Since the reset voltage is always larger than the set voltage \((|V_{\text{reset}}| > |V_{\text{set}}|)\) current compliance is not needed during the switching.

The SiO\(_x\) based resistive switches exhibit large OFF/ON resistance ratio \( (>10^5)\), non-volatile properties, the resistance states could be stable as long as \(10^5\) s and good endurance \( (>10^4\) write-erase cycles) was observed. As we will see in Chapter 6, the resistance states can be also programmed by the proper choice of voltage pulses despite of its unipolar nature. In Figure 2.7.c the top panel shows a series of voltage pulses of 6 V (set pulse) and 15 V (reset pulse), which serve to switch the contact to LRS and HRS respectively. Between these pulses the low voltage resistance was measured by applying 1V high pulses (read pulses). The bottom panel shows the corresponding current for each read pulse. After applying the set and reset pulses the current changes more than 4 order of magnitudes.

The SiO\(_x\) resistive switches show fast switching speed, it can reach 50ns [72]. Furthermore, the resistance of the LRS and the HRS can be tuned by the length and the height of the set and reset pulses, which enables multi-bit memory. Figure 2.7.d shows that the device can be reset by a 50ns length and 8 V height pulse. The ON state can be partially recovered by a 50ns, 4 V pulse and fully set to LRS by a longer \((100\text{ns})\) one. Similar behavior can be observed by tuning the pulse amplitude, the higher the set (reset) pulses the lower (higher) the final resistance of the device [73].

To get insight into the mechanism of the switching several samples with different parameters were investigated. By varying the geometry of the samples the switching voltages \((V_{\text{set}} \text{ and } V_{\text{reset}})\), the electroforming voltages \((V_{\text{forming}})\) and the corresponding currents were found to be independent of the cross section of electrodes [72]. This nonscaling behavior suggests that the switching occurs only in a small localized part of the insulator layer and confirms the filamentary nature. Furthermore the switching voltages and currents show thickness-independence also indicating that the filament builds up and ruptures only at the weakest point [70] and the other parts of the filament do not change. The forming voltage, however, shows linear dependence on the insulator thickness, i.e. a constant electric field is needed to establish the filament for the first time [69]. The switching region can be created only at the surface of the insulator, as the surface is more defective and more readily forms conductive filament [72]. Electroforming and switching can not be induced in oxygen rich environment.
and at low temperature, only the high resistance state can be achieved [19, 72].

![Figure 2.8: High resolution TEM images about the active region of SiO$x$ based resistive switches. a) The pristine amorphous SiO$x$ after the nanogap formation, but before the electroformation. The nanogap region after b) electroformation c) set process and d) reset process. The circled areas mark the nanocrystalline volumes and the insets show the enlarged part of the nanogap region [19].](image)

The exact mechanism of the resistive switching was revealed by in situ imaging of the conductive filament by a transmission electron microscope (TEM) [19]. The structure of the insulator layer was investigated before and after electroforming and switching events. The studied switchable SiO$x$ region was formed in a $\approx 15$ nm sized nanogap between amorphous carbon electrodes. Figure 2.8 shows a series of high-resolution TEM images about the nanogap region, the insets present the enlarged part of the nanogap. Before the electroformation (Figure 2.8.a) only amorphous silica can be observed, however after the formation of the switching site nanocrystalline structures appear (see Figure 2.8.b). The electroforming can be associated with the local enrichment of the silicon in the nanogap region by striping away the oxygen atoms. After that, due to the high electric field this silicon rich region can transform to nanometer-sized crystals. According to the electron diffraction pattern the structure of the Si crystals does not correspond to the conventional semiconductor, diamond cubic Si-I phase. The intersected lattice spacings refer to the semi-metallic Si-III phase [74, 75], which explains the metallic conduction of the ON state. The TEM images taken after the set and reset process are shown in Figure 2.8.c and
d respectively. The circled regions mark the crystalline part and they are enlarged in the insets. However, the filament does not consist of one continuous crystal, instead several nanocrystals along the nanogap. The electron transport happens across these nanocrystals. During the resistance transition the size of the crystals grows or shrinks. The set process can be associated as the electric field induces crystallization, while the reset process is thermally induced amorphization. This picture is consistent with the observation of suppression of switching and electroforming at low temperature and ambient condition. The Si-III phase cannot be formed at the temperature of liquid nitrogen [76], while in oxygen rich environment the Si nanogap region is oxidized again, when it is heated by high current density. However at room temperature the Si-III phase is stable which accounts for the non-volatile property.

Another model is proposed to the mechanism of the intrinsic switching by the research group of J. C. Lee [77]. They assume that the charge transport occurs along the defects in the SiO₂. Since the sidewall is more defective the filament tends to form there. The rest of the defects evolve during the electroformation and after the soft breakdown the concentration of these defects is fixed and localized. During the resistance transitions the defects are converted between low and high conductive states. The conversion, however, takes place only at the gap region, at the minimum cross-sectional area. The other parts of the filament remain intact. The proposed conductive defect is Si-H-Si and the non-conductive defect is (SiH)₂, the transition between them happens through H⁺ release and H desorption. The switching mechanism and charge transport were supported by a detailed energy band model as well.

In Figure 2.7 c-d we can see that the high resistance state can be achieved at low voltage if the applied pulse has fast enough falling edge. This behavior contradicts the unipolar switching mode, the device should always switch back to LRS during the reserved voltage sweep. This effect was observed by both the research group of James M. Tour and J. C. Lee. The former group noted that after the fast falling edge in the reset pulse (10 ns) the device preserved its HRS despite of the unipolar nature. They discussed this phenomena as “during the falling edge in a reset pulse, the device is in a “hot” state since the voltage starts from the reset region, as opposed to a “cool” state in the set operation. This “hot” state may prevent the set process incurred during the falling edge. The detailed study of this aspect has not yet been done.” [71]. The latter research group examined the final state after the reset pulse as the function of falling time of the pulse and the temperature [78]. As the falling edge time was shortened they explored a temperature dependent characteristic time, where the final state changes from ON to OFF. They referred to this effect as backward-scan effect. This effect would have a great technological impact by
integrating the advantages of unipolar and bipolar devices: it makes both states achievable at low voltage level using solely unipolar signals. Despite of the many notable results the dynamics of the switching phenomena has not been investigated in SiO\textsubscript{x} memristors by time-resolved measurements. A detailed study could also reveal how we can achieve the ON and OFF states at low bias while the memristive system shows unipolar behavior. In Section 6.2 I present my research regarding the real time response of SiO\textsubscript{x} resistive switches. Furthermore, I investigate whether the switching capability is still maintained in the sub-10nm regime and whether the switching properties are modified due to the ultrasmall size. These issues also have not been clarified yet.

2.1.4 Resistive switching in Ag\textsubscript{2}S

In Ag\textsubscript{2}S memristive systems the electrochemical metallization (ECM) is considered as the main driving force of the resistive switching. In case of ECM one of the electrodes is electrochemically inert, such as Pt [79] or PtIr [80], while the other one is active, which is typically Ag or Cu [81]. The active electrode serves the atoms to form the conductive filament, while the passive electrode only provides electrical contact. The two electrodes are separated by a thin layer of insulator or ionic conductor. In many cases the dielectric also contains the cations of the active electrode, like Ag\textsubscript{2}S [82], AgI [83], Cu\textsubscript{2}S [84], but it is not required [85, 86].

When we apply positive voltage to the active electrode the high electric field ionize the atoms of silver electrode and the ions diffuse inside the insulator layer towards the opposite passive electrode. If the dielectric layer also contains the same cations they also start to migrate to the passive electrode. At the passive electrode these ions are reduced back to atoms. If the ion migration takes long enough time a conductive filament evolves between the two electrodes. By applying the reversed bias polarity the opposite process takes places, the atoms of the filament migrate back to active electrode and the filament thins or breaks fully. The electrochemical metallization based filament growth is illustrated in Figure 2.9 [87]. In most ECM systems the narrowest cross-section and thus the resistance can be tuned precisely by the parameters of the driving signal [88]. The exact structure of the filament depends on the magnitude of cation mobility and the oxidization, reduction rate. Among other it could be cone-shaped or branched [89].

Due to the mechanism mentioned above the memristors based on ECM show bipolar switching mode. The different sign of the set and reset voltage can be explained by the different material of the active and the passive electrodes. However, the same switching behavior was found in Ag-Ag\textsubscript{2}S-Ag structure solely due to the
The goal of my research was to implement this single material Ag$_2$S memristor into a nanofabricated on-chip device which offers higher mechanical stability and the possibility of integration to more complex circuits. I present my results in this field in Section 4.

In case of Ag$_2$S memristors the movement of the silver ions during the resistance transition is aided by the two crystallographic modifications of Ag$_2$S [90, 91]. The as-grown silver-sulfite has monoclinic acanthite phase, which is a semiconductor with bandgap of 1.3 eV and low conductivity. At elevated temperature (451 K) this phase transits to argentite, whose band gap is 0.3 eV. Argentite is a superionic conductor which results in high ionic mobility [92, 93]. It is reported, that the applied electric field can reduce the temperature of the phase transition [94], but both phases are stable at room temperature. During resistive switching due to the local Joule heating and the large electric field at the active region the acanthite phase transforms to argentite and the enhanced ion conduction facilitates the movement of the Ag ions [93, 95]. It results in rapid filament growth, the set and reset operation can be induced by sub-ns voltage pulses [96].
2.2 Fabrication of nanometer-sized gaps

The size of CMOS devices have been shrinking continuously in accordance with Moore’s law. The gate width of the FinFets has reached the 10 nm size and it is expected to decrease further. However it is assumed that the traditional CMOS devices will face their size limitation soon due to the material properties of the compounds, such as enhanced tunneling current through the gate oxide. New kinds of devices are needed, which exploit the different behavior of the materials at nanoscale. The size of the resistance change memories can be scaled down even below 10 nm due to the filamentary nature. In order to integrate these nanoscale devices into an electrical circuit, electrodes with few nanometer spacing must be fabricated.

The electrical characteristics of the nanoscale devices strongly depend on the local electrode geometry. Highly reproducible and precisely controlled nanogap formation is essential and one of the most challenging issues of manufacturing nanoscale electronics. The sub-10 nm sized structures can not be fabricated by the standard lithography methods yet, additional steps or different approaches are needed. So far several different methods were developed to make extremely small gaps between metal electrodes, however the reproducibility is still not solved.

One of the main driving force of few nanometer sized gap fabrication has been the investigation of single molecular junctions. Most of the papers about the nanogap formation have the goal of making electrical contacts with molecules. Molecular electronics considers single molecules as the building blocks of the future electronic circuits such as single molecule transistors [97] or diodes [98]. Furthermore, nanogap structures were applied as biosensors [99] and gas sensors [100] as well.

For large statistical analysis a simple sample preparation is desired, where the electrodes can be redefined easily. For this purpose the scanning tunneling microscope (STM) [101] and the mechanically controllable break junction technique (MCBJ) [102] are perfect tools. Both methods are based on the mechanical rupture of metal contacts. Using one sample thousands of nanocontacts can be defined. In case of STM setup the tip is approached or touched to the surface of a conductive material to create tunneling or atomic sized contact respectively. The junctions can be established even between two different materials and a fresh contact can be created at another point of the sample surface. However its mechanical stability is rather limited. The MCBJ technique is based on the mechanical rupture of a metal wire ~100 micron diameter. The setup can be easily placed into liquid helium which results in sub-nm stability for several hours, but we can not study the contact of two different materials. Both molecular contacts [103, 104] and memristors [2, 96] have been investigated using these techniques.
However, these methods are not suitable for making highly integrated devices or fabricating three or more terminals. Their room temperature stability is not sufficient as well. An alternative solution is proposed by the electromigration technique, which is based on the electrical breakdown of a continuous nanowire made by lithography. By controlling the breakdown process properly few nanometer sized gaps can be achieved. The on-chip fabrication offers higher stability and allows the integration into a complex circuit.

It must be noted, besides the techniques mentioned above there are several other methods to create nanosized gaps such as focused ion beam lithography [105], electron-beam lithography [106], atomic force lithography[107], electrochemical plating [108] or on-wire lithography [109, 110].

2.2.1 Electromigration of metal nanowires

The electromigration is the motion of metallic ions in a metal wire induced by large current density. If this ion transport exists for a long time the wire gets narrower and finally breaks. This phenomena is known for more than a century but the practical significance increased in the late 1960s when the life time of the integrated circuits (IC) decreased as the size of microelectronic components and interconnects decreased. Since the reliability of the ICs is a critical requirement, nowadays the electromigration has become a highly studied effect [111].

The first empirical model was given by Black, who derived a formula for the mean time of the failure of a metal wire by taking relevant physical parameters into account:

\[
MTTF = \frac{A}{j^n} \cdot e^{-\frac{E_a}{k_B T}},
\]

where A is a constant which comprises the geometry and the material properties, \(j\) is the current density, \(E_a\) is the activation energy, \(T\) is the temperature and \(k_B\) is Boltzmann's constant. The value of the exponent of the current density (n) is typically between 1 and 2 depending on the exact failure mechanism. The expression shows that both the current density and the temperature play an important role in the breakdown.

The charge carriers flowing through a conductor scatter on lattice defects, impurities and grain boundaries. The scattering events exert force on the metal ions, which is called electron wind force and has the same direction as the current density. The local electric field also acts on the metal ions and the direction of the resulting force (direct force) is opposite to the wind force. The movement of the metallic ions
is caused by the competition of these two microscopic forces. The larger current density both increases the wind force and enhances the mobility of ions due to the higher temperature. The migration can easily accelerate. The migration is the most dominant at the surface and at the grain boundaries, where the ions are less bound and the scattering events are more frequent.

If the bias voltage is increased above a critical value the electromigration enhances and the current drops abruptly (see Figure 2.10.a). At the weakest point of the wire the cross section starts to shrink which results in even higher current density and temperature. The temperature can even reach the melting point of the metal and thereby a significantly larger gap will obtained with nanosized metal islands inside. In order to get a reproducible, few-nm sized gap, the control of the electromigration process is essential.

Figure 2.10: a) Current vs bias voltage during the uncontrolled electromigration of a gold nanowire. The wire suddenly breaks when the bias reaches a critical value. The inset shows the last 40 mV part of the bias ramp [112]. b) Schematic illustration of the electrical circuit. The total resistance consists of the series resistance ($R_s$) and junction resistance ($R_j$). During the electromigration only $R_j$ changes. c) The acquired I(V) traces during the controlled electromigration of a gold nanowire. When the current decreased by 1-6% a new bias ramp was started from lower voltage value. By repeating the voltage ramp several times the resistance was increased until the wire broke [113].

The total resistance of the circuit can be divided into two parts (see Figure 2.10.b). The junction resistance ($R_j$) changes during the electromigration, while the series resistance ($R_s$) keeps constant. The series resistance plays an important role during the migration [112]. As the wire becomes narrower the junction resistance starts to increase ($R_j(t)$). Under fixed bias the electrical power on the contact changes as $P_j = \frac{V_b^2}{(R_s+R_j)^2} R_j$. If at the beginning of the migration $R_s \gg R_j(t=0)$, then as the
junction resistance increases more and more power dissipates in the critical volume raising further the temperature. But if the junction resistance is the dominant from the beginning ($R_s < R_j$), then the explosion of the contact be avoided. Samples with low series resistance show much better yield for nanogap formation [112].

In the case when the series resistance can not be reduced below the junction resistance active feedback control is required. If $R_s$ is not significantly larger than $R_j$, a precursor can be seen right before the breakdown. The inset of Figure 2.10.a shows that before the electrical breakdown the current decreases by few percents. The wire starts to thin at this moment. If the current is monitored fast enough it is possible to stop the voltage ramp before the metal wire fully breaks. During the typically applied method the current is measured with high sampling rate ($> 10$kHz) and the voltage is decreased to low value when the current changes with certain percentage. Performing this voltage ramp cycle several times the junction resistance can be increased step by step. The I-V traces of a feedback controlled electromigration are shown in Figure 2.10.c [113]. Before the breakdown the two electrodes are connected by only an atomically thin wire. In this regime the conductance changes step wisely due to the quantized conductance and the discrete atomic jumps inside the junction [112]. The yield of nanogap formation using electromigration technique could be higher than 90% [112, 114].

2.2.2 Electrical breakdown of graphene nanoribbons

The most preferred electrode material to fabricate nanogap devices is gold due to its inert property to many chemical reactants. However, at room temperature there are stability problems since the gold has large surface diffusion. Furthermore, some kind of memristors show intrinsic switching which means that the electrodes play only a passive role. In these systems it is important to avoid the diffusion of metal ions into the switching matrix, because they can also form metallic filaments. However, at nanoscale it is hard to exclude this risk.

Using graphene as electrode material offers a simple solution for the problems mentioned above. At room temperature the graphene has outstanding mechanical stability due to the strong covalent bonds. The metal-free realization ensures that the insulator layer is not fuelled with metal ions by the electrodes. Furthermore, the 2D structure of graphene ensures more defined nanogap geometry, which may lower the device variability. In contrast of the one atom thick graphene for metal electrode the closest points of the two sides could be far from the substrate surface. The one atom thin graphene also provides the possibility to fabricate transparent and flexible electronics. Graphene was demonstrated as a perfect electrode material in case of
several types of devices such as DNA sequencer [115], single molecule devices [116] or memristors [71, 117].

One of the simplest techniques to make nanometer-sized gaps is the electrical breakdown which is very similar to the electromigration. This procedure was also used for carbon nanotubes to establish molecular contacts [118]. The first graphene nanogaps were by fabricated Prins et al. on multilayer, exfoliated graphene using the feedback control of the bias voltage [116]. They patterned a few hundred nanometers wide constriction into the graphene sheet and after high current density was driven through the sample. They found that under ambient conditions the breakdown process was controllable, that is, before the graphene stripe had broken, the resistance increased significantly as the precursor of the breakdown. During the feedback controlled breakdown they swept the bias voltage to zero when the resistance changed by 10% and a new sweep was started. In this way they could gradually narrow down the flakes. They attributed the breakdown mechanism to the oxidation of the carbon atom, therefore this process was called electroburning. The yield of nanogap formation under ambient conditions was found between 26-92% depending on the initial resistance of the multilayer graphene [116, 119]. The speed of the sampling rate and accordingly the feedback control was 200 µs. In vacuum graphene nanogap formation has not been investigated and there were no systematic studies which compare the effects of environmental conditions to the breakdown mechanism. However, in case of carbon nanotubes (CNT) the gap size could be controlled by regulating the oxygen partial pressure. The gap size resulting from uncontrolled electrobreakdown was always larger than 30 nm in oxygen, air and argon atmosphere. In contrast, under high vacuum the gap size was less than 10 nm for 50% of the samples [120].

In Section 5.3 I study in detail the breakdown mechanism under different environmental conditions. I show that clearly distinguishable processes take place at ambient and in vacuum, which can be attributed to burning and sublimation respectively [3]. Similarly in metals, there is also a risk of formation of too large gap and charge islands. In order to exclude the possibility of tunneling through carbon islands, the electric transport has to be measured as the function of gate voltage. The diamond-like structure on the bias voltage - gate voltage map would imply the contamination or partial breakage of graphene [121].

It should be noted that beside the electrical breakdown process, there are several other approaches to make nanometer-sized gaps into a graphene sheet, such as AFM lithography based on mechanical cutting [122] or local oxidation [123], STM lithography [124] or helium ion beam etching [125].
2.2.3 Charge transport through dielectrics

After the electrical breakdown of the constriction the charge transport measurement is one of the most precise methods to characterize the formed nanogap. In order to be able to interpret the electrical measurements on nanogap devices I briefly summarize the most important transport mechanism in dielectrics.

Depending on the role of the bulk insulator two different conduction types can be distinguished. One is when the dielectric does not take part in the charge transport, only the electrodes and the electrode-insulator interface determine transport properties. It is also called as electrode-limited conduction. In the other case the insulator is also involved in the conduction through the charge traps due to the structural defects in the material. This kind of conduction mechanisms are called bulk-limited conduction. By detailed analysis the trap level, trap spacing, trap density, dielectric relaxation time, etc. can be extracted. However, at the same time several conduction mechanisms may contribute to the current through the dielectric layer.

In case of electrode-limited conduction the simplest model assumes rectangular potential barrier with an average height of $\Phi$ and width of $d$. If the barrier is thin enough ($< 10$ nm) the wave function of the electrons penetrates through the insulator layer and tunneling current can flow.

**Direct tunneling**

If the applied bias is low ($eV < \Phi$) the distortion of the potential barrier is not significant, the electrons see the full width of barrier, as illustrated in Figure 2.11.a. This conduction regime is called direct tunneling and the Simmons model gives a generalized description. In case of similar electrodes with the same work function the current ($I$) can be expressed by the following form [126, 127]:

$$I = \frac{A e}{4\pi^2 \hbar d^2} \left[ \left( \Phi - \frac{eV}{2} \right) \exp \left( \frac{-2d}{\hbar} \sqrt{2m_e \left( \Phi - \frac{eV}{2} \right)} \right) 
- \left( \Phi + \frac{eV}{2} \right) \exp \left( \frac{-2d}{\hbar} \sqrt{2m_e \left( \Phi + \frac{eV}{2} \right)} \right) \right],$$

\hspace{1cm} (2.8)

where $A$ is the surface of the tunnel junction, $m_e$ is the electron mass, $e$ is the elementary charge and $\hbar$ is the reduced Planck’s constant. By fitting this expression to the measured current data the main parameters of the tunnel junction can be extracted (see Figure 2.12.a).
Figure 2.11: Schematic energy band diagram of a) Direct tunneling, b) Fowler-Nordheim tunneling, c) Schottky emission and d) Thermionic-field emission in metal-insulator-metal structure.

At small bias $eV \ll \Phi$ Equation 2.8 simplifies to a linear current-voltage relationship, which allows us to assign a resistance value to the tunnel junction (see inset of Figure 2.12.a).

$$I \sim V \exp \left( -\frac{2d\sqrt{2m_e\Phi}}{\hbar} \right).$$  \hspace{1cm} (2.9)

**Fowler-Nordheim tunneling**

At higher bias ($eV > \Phi$) the trapezoidal shape of the barrier distorts to triangular (see Figure 2.11.b) and the Simmons model is not valid any more. This regime is called field electron emission or Fowler-Nordheim tunneling. The current can be written as [128]

$$I \sim V^2 \exp \left( -\frac{4d\sqrt{2m_e\Phi^3}}{3\hbar e} \cdot \frac{1}{V} \right).$$  \hspace{1cm} (2.10)

In case of F-N tunneling the plot of $\ln(I/V^2)$ versus $1/V$ should be linear with negative slope. On the other hand, the current in direct tunneling regime shows logarithmic growth on the same axis (see Figure 2.12.b). The transition between the two tendencies refers to the validity of the Simmons model and can be used to determine the fitting limits. This technique is called as transition voltage spectroscopy [128, 129].
Schottky emission

In case of Schottky conduction the electrons can gain enough energy by thermal fluctuations to overcome the energy barrier, Figure 2.11.c shows energy band diagram of this process. This conduction mechanism is also called as thermionic emission and it is a very often observed conduction mechanism, especially at higher temperature. The expression of current density is [130]

\[ J \sim T^2 e^{\frac{-\Phi - \sqrt{e^3 V/(4d\pi e_r \epsilon_0)}}{k_B T}}, \quad (2.11) \]

where T is the temperature, \( k_B \) is the Boltzmann’s-constant, \( \epsilon_0 \) is the permittivity in vacuum and \( e_r \) is the dielectric constant. For Schottky conduction the plot of \( \ln(I/T^2) \) vs \( V^{1/2} \) is linear and the barrier height can be obtained from the intercept.

Thermionic-field emission

This conduction mechanism lies in the intermediate regime of Schottky emission and the field electron emission. In this case the electrons do not have enough energy
to overcome the energy barrier, but have larger energy than the Fermi-level of the source metal electrode, due to the thermal excitations (see in Figure 2.11.d).

**Poole-Frenkel emission**

In real dielectrics there are several structural defects which can serve as charge traps and induce additional energy levels inside the band gap of the insulator. The transport through these charge traps typically shows strong temperature dependence, due to the thermally activated nature of the process.

![Figure 2.13: Schematic energy band diagram of a) Poole-Frenkel emission and b) Hopping conduction in metal-insulator-metal structure.](image)

The conduction mechanism of Poole-Frenkel emission is similar to the Schottky emission. The electrons from the source electrode reach the drain electrode through several charge traps. The electrons are stuck in the localized states until the random thermal excitation provides enough energy to move further. The applied electric field across the dielectric film can reduce the potential barrier of the trap ($\Phi_T$). The corresponding energy band diagram is shown in Figure 2.13.a. The current density can be written as [131]

$$J \sim \frac{V}{d} \exp \left(-\frac{\Phi_T - \sqrt{e^3V/(d\pi\varepsilon_r\varepsilon_0)}}{k_B T}\right). \quad (2.12)$$

Since both the temperature and the electric field are the driving force of the charge transport, this conduction mechanism is dominant at high temperature and high electric field. The corresponding plot is the $\ln(I/V)$ vs $V^{1/2}$, where the measured points scale to a line and the intercept gives the trap energy level ($\Phi_T$).
Hopping conduction

During the hopping conduction the electrons hop from one trap to another due to the quantum tunneling effect, as illustrated in Figure 2.13.b. The current density is [132]

\[ J = ean\nu \exp \left( \frac{eaV}{d\kB T} - \frac{E_a}{\kB T} \right), \]  

where \( a \) is the mean hopping distance, \( n \) is the electron concentration in the conduction band of the dielectric, \( \nu \) is the frequency of thermal vibration of electrons at trap sites and \( E_a \) is the activation energy. The corresponding plot is the log(I) vs V.

Besides these two most relevant bulk-limited conduction mechanism there are several other ones such as space-charge-limited conduction, ionic conduction or grain boundary conduction, however they are not discussed here.

2.3 Charge and heat transport in graphene

During my PhD work most of my measurements were performed on graphene to fabricate nanometer-sized gaps and later to utilize them as the electrodes of nanometer-sized devices. Graphene has remarkable properties, such as outstanding mechanical strength, ultra-high electron mobility, high thermal conductivity, transparency and flexibility, which makes graphene a very promising material for the future applications. Since its first isolation in 2004 by Novoselov and Geim [133] the production techniques of graphene have been developed significantly. Now it is solved to grow wafer-scale graphene with good quality, which is essential for industrial applications. Although these techniques do not provide ultra-high mobility, it is not required if we used graphene as electrode material. The advantages of graphene were listed in Section 2.2.2, while in this section I introduce the electrical and thermal properties of graphene, which are relevant to my research.

2.3.1 Electrical properties of graphene

The hexagonal structure of graphene can be described as a triangular lattice with two atoms per unit cell. The bonding distance between the carbon atoms is 1.42 Å (see Figure 2.14.a). The 2s, 2p\(_x\) and 2p\(_y\) orbitals are hybridized, which overlap with the neighbouring atoms and form \( \sigma \) bonds in the plane of the graphene. According to the Pauli principle, these bands have a filled shell and form a deep valence band. They are responsible for the robustness and flexibility of the graphene [134]. The
remained 2p\textsubscript{z} orbitals are perpendicular to the graphene sheet and can be treated independently from the other three ones. The 2p\textsubscript{z} orbitals form π bonds, which are delocalized to the whole graphene sheet. These bonds are involved in all observed electron transport properties at the relevant energy scales. The basis vectors of the reciprocal space also form triangular lattice and the Brillouin zone is also hexagonal, as illustrated in Figure 2.14.b. The tight binding approach by considering the nearest- and next-nearest neighbor hopping gives a good approximation to the band structure. The dispersion spectrum can be expressed in the following form [135]:

\[ E_{\pm}(k) = \pm t\sqrt{3 + f(k)} - t'f(k), \quad (2.14) \]

\[ f(k) = 2\cos(\sqrt{3}k_ya) + 4\cos\left(\frac{\sqrt{3}}{2}k_ya\right)\cos\left(\frac{3}{2}k_xa\right), \quad (2.15) \]

where \( t \) is the nearest-neighbour hopping energy (≈ 2.8 eV), \( t' \) is the next nearest-neighbour hopping energy (0.02t < t' < 0.2t). The positive and the negative sign correspond to the conduction (π\textsuperscript{*}) and the valence (π) band respectively. In most cases \( t' \) can be neglected because \( t' \ll t \). This approximation results in a symmetric band structure around zero energy, which located at the corners of the Brillouin-zone (K and K' points). At these points the two bands cross each other with zero band gap and zero density of states (see Figure 2.14.c). These points are called charge neutrality points (CNP) or Dirac-points due to the local linear dispersion (if \( t'=0 \)). Since there is only one p\textsubscript{z} electron per atomic site, the valence band is fully occupied, while the conduction band is empty, the graphene can be considered as a zero band gap semiconductor. The local linear energy dispersion around the CNP can be written as

\[ E_{\pm}(q) \approx \pm v_F |q| + O\left([q/K]^2\right), \quad (2.16) \]

where \( q \) is the momentum vector measured from the CNP (\( k = K + q, \ |q| \ll |K| \)) and \( v_F \) is the Fermi velocity, equal to 3\( ta/2 \approx 1 \cdot 10^6 \text{m/s} \) [136]. Due to the linear dispersion, the charge carriers around the Dirac cones behave as a massless fermions. The two nonidentical CNPs (K and K') lead to chiral charge carriers in graphene, which can be described as a pseudospin. In neutral graphene the Fermi energy lies exactly at Dirac points, however due to the dopants, defects and surface contamination the Fermi level can be shifted.

The charge carrier density can be tuned electrostatically if we use the graphene as one plate of a capacitor, while the other plate is the doped silicon under the graphene.
Figure 2.14: a) The honeycomb lattice structure of graphene in real space. The primitive unit cell and the basis vectors are also shown. b) The Brillouin-zone of the reciprocal lattice. c) Dispersion relation of the conduction and the valence band of the graphene. The Dirac cones are located at the corners of the hexagonal lattice (K and K’) [134].

The two plates are separated by a dielectric such as SiO$_2$ or Si$_3$N$_4$. The charge carrier density by applying $V_g$ gate voltage on the doped silicon can be written as,

$$n = C_g(V_g - V_0)/e,$$  \hspace{1cm} (2.17)

where $C_g$ is the capacitance per unit area, $V_0$ is the offset of the CNP caused by the doping and $e$ is the elementary charge.

If the mean free path is smaller than the device geometry, the charge transport is diffusive and the conductivity can be derived from the relaxation time approximation of the Boltzmann equation. It can be given as

$$\sigma = |n| e \mu,$$  \hspace{1cm} (2.18)

where $\sigma$ is the longitudinal conductivity and $\mu$ is the mobility. In case of graphene supported on SiO$_2$ substrate the mobility rarely exceeds $2 \cdot 10^4$ cm$^2$/Vs [137, 138] even for cleaned, exfoliated graphene. The corresponding mean free path is $l_e = 230$ nm ($l_e = \mu \hbar k_F/e$, $k_F = \sqrt{|n| \pi}$) at $n=10^{12}$ cm$^{-2}$ [134]. According to Equation 2.18, the conductivity should vanish at the CNP, however due to the short range scattering it has a theoretical minimum value of $\sigma_{min} \approx 4e^2/\pi h$ [139, 140]. In the transport measurements the minimum of the conductivity is in the range of $\sigma \approx 2 - 5e^2/h$ due to the charge inhomogeneity induced by defects or doping [138, 141, 142].

The electrical quality of the graphene sheet is typically characterized by charge carrier mobility measurement. In accordance with Equation 2.17 and 2.18, the mobility can be derived from the gate voltage dependent transport measurement by
\[
\mu = \frac{1}{C_g} \frac{\partial \sigma}{\partial V_g}.
\]  

(2.19)

The mobility of clean, defect free, exfoliated suspended graphene can reach \(10^6 - 10^7\,\text{cm}^2/\text{Vs}\) [138, 143]. This ultra-high mobility of graphene allows us to study quantum Hall effect [144, 145] or ballistic transport [146, 147] even at room temperature. However, on SiO\(_2\) substrate the electronic properties are highly influenced by the defects, roughness and impurities of SiO\(_2\) making charge puddles. Furthermore the electrons of graphene interact with the surface phonons of the substrate. Therefore the typical mobility on SiO\(_2\) substrate is in the range of \(1000 - 10000\,\text{cm}^2/\text{Vs}\) [137, 138].

The resistance of a rectangular graphene to width \(W\) and length \(L\) can be given by the Ohm’s law:

\[
R = \frac{1}{\sigma} \frac{L}{W} = \rho \frac{L}{W}
\]  

(2.20)

In case of arbitrary shape sample, the resistance is still proportional to with the inverse of conductivity, but the proportionality factor is a geometry dependent constant.

### 2.3.2 Thermal properties of graphene, energy dissipation

In order to understand the thermal properties of the graphene at first the vibration modes must be inspected. The unit cell has 2 carbon atoms which results in 3 acoustic (A) and 3 optical (O) phonon modes. Both of them have longitudinal (L), transverse (T) and out-of-plane (Z) modes. The longitudinal acoustic (LA) and transverse acoustic (TA) modes starts linearly at the center of the Brillouin-zone (\(\Gamma\)), while ZA has quadratic dispersion, see in Figure 2.15. Due to the strong in-plane covalent sp\(^2\) bonds and the low mass of carbon atoms, the speed of sound is very high in graphene.

The specific heat (C) has two components. One stems from the energy stored by the phonons (\(C_p\)) and the other comes from the free conduction electrons. Above 1K the phonons dominate the specific heat [149], which increases with the temperature until the Debye-temperature of the in-plane phonons, \(\Theta_D \approx 2100\,\text{K}\). Above this temperature the specific heat is nearly constant at \(C \approx 25\,\text{J/molK}\approx 2100\,\text{J/kgK}\) according to the Dulong-Petit law. The specific heat of graphene is proposed to be the same as that of graphite above 100\,\text{K} [150, 151]. At room temperature the specific heat is about 700\,\text{J/kgK} [150, 152].
Figure 2.15: Calculated phonon dispersion for monolayer graphene [148]

The specific heat does not only determine the amount of energy stored within the material, but also how quickly the graphene heats or cools. Its characteristic time is the thermal time constant, which can be given by $\tau = RCV$, where $R$ is the thermal resistance of heat dissipation and $V$ is the volume. The thermal time constant is very small for nanoscale graphene objects, it is in the range of 10 ns [153] to 0.1 ns [154].

Like the specific heat, the thermal conductivity ($\kappa$) also consists of two parts, one from the phonons and other one from the electrons. In case of metals, where the free electron density is high, the electronic contribution dominates and the Wiedemann-Franz law gives the relation between the electrical conductivity and the thermal conductivity. However in graphene due to the strong covalent bonds, the heat transfer is dominated by the phonons [155]. The in-plane heat conductivity of graphene ($\kappa_g$) is one of the largest among the known materials, it could be 2000 – 4000 W/mK for suspended graphene [155, 156]. However, when the graphene is in contact with a substrate the thermal conductivity decreases because of the enhanced phonon scattering. At room temperature the heat conductivity of graphene supported on SiO$_2$ substrate is in the order of 100 – 1100 W/mK [157, 158]. The suppressed value is the result of the coupling of the phonons of graphene with the surface phonons of the substrate. In case of graphene nanoribbons (GNR) if the sample dimension reaches the phonon mean free path the heat conduction decreases further [159, 160]. At room temperature in graphene supported on SiO$_2$ the mean free path is $l_{ph} \approx 100$ nm [161].

The self-heating of graphene is considered as the scattering of the electrons with the optical phonons [162, 163]. The high electric field accelerates the electrons and if they gain enough energy ($\approx 160 meV$) optical phonons are emitted and the electrons scatter back. Later the optical phonons decay into acoustic phonons [150, 162].
electron backscattering causes current saturation effect in graphene [162, 164, 165] similarly than in CNT [166, 167], see in Figure 5.5.a. If the graphene is supported on SiO$_2$ the electrons can interact directly with the surface phonons of the substrate, which results in the heating of the SiO$_2$.

There are two relevant pathways for the dissipated heat to the environment. One is the intrinsic heat conduction of the graphene to the metal contacts. In the diffusive regime ($L,W \gg l_{ph}$) the heat conduction can be written by the Fourier's law,

\[ q = -\kappa \nabla T, \]  

(2.21)

where $q$ is the heat flux density and $\nabla T$ is the temperature gradient. In wider temperature range we need to take into account the temperature dependence of the thermal conductivity.

The other pathway is the heat conduction to the substrate. The heat flow perpendicular to the graphene sheet is dominated by the weak van der Waals coupling. The relevant physical quantity for the heat transfer across the interface is the thermal boundary resistance per unit area ($\rho$), whose value is on the order of $5 \cdot 10^{-9} - 4 \cdot 10^{-8}$ m$^2$K/W [168-170] in case of SiO$_2$ substrate. This limited heat flow at the interconnects could be a thermal bottleneck in case of highly scaled graphene devices [169].

Although the heat conduction of nanoscale supported graphene is the fraction of the bulk suspended graphene, but in nanoscale the heat conduction of most materials are also suppressed. Few tens of nanometer sized copper interconnect has about 100 W/mK thermal conductivity based on the Wiedemann–Franz law [150, 171].
The main goal of my PhD work is to create smaller resistive switching devices than the resolution of present lithography techniques. In order to reach the sub-10 nm regime the dimension of the initially nanofabricated structure is further reduced by controlled electrical breakdown process. This Chapter is devoted to the detailed introduction of the novel measurement setup which is suitable to establish nanometer-sized gaps in conductive wires. By confining the active region of the resistive switches into the nanogap region the switching characteristics of different memristive materials can be studied at ultrasmall scale.

During my MSc studies I was working on the fabrication of atomic-scale nanostructures using electromigration technique at the Solid State Physics Laboratory of the Department of Physics, BUTE. I developed a low temperature experimental setup and a measurement control program to perform controlled electromigration on nanometer sized metal wires at 4 K. The first measurements were performed on nickel and platinum MCBJ samples and on nanofabricated platinum devices made by the Nanotechnology Group of Tyndall National Institute. After the breakdown process the formed gaps were characterized electrically by measuring tunneling current, which revealed few nanometer large gap sizes [172].

After starting my PhD work I improved the nanogap fabrication process and I extended this method on single-layer chemical vapor deposited graphene nanostripes. The experimental setup was optimized for more simple device handling and the measurement control program was modified for faster feedback control. This setup was used for all the measurement presented in my thesis. Furthermore, the collaboration with the Department of Physics, University of Basel and the Institute of Technical Physics and Materials Science, Hungarian Academy of Sciences Centre for Energy
Research (MTA EK MFA) made it possible to fabricate specifically designed devices using their nanofabrication facilities.

### 3.1 Mechanical setup

In order to characterize electrically the fabricated on-chip nanodevices we have to connect them to our instruments via electrical wires. For this purpose we need to bridge the several order of magnitudes differences in size. As the first step, during the device preparation hundreds of micrometers large metal pads are defined on the silicon wafer (see Section 4.2 and Section 5.1.2). In the next stage we need to establish electrical contact between these metal pads and our instruments. To achieve this at first the silicon chip is fixed to a printed circuit board (PCB) (see Figure 3.1.a) which leads the contacts of the samples to the pins of a 25-pins D-sub connector. The doped Si wafer is fixed to the square metal layer at the middle of the PCB using silver paste. The conductive paste ensures the electrical contact to the backgate. In order to get better ohmic contact to the backgate the native oxide layer has to be removed on the back side of the Si wafer by scratching its surface. The tuning of local electric potential has an important role in case of graphene devices to determine the local quality of graphene sheet prior establishing the resistive switch. In case of characterization of Ag$_2$S memristors the backgate is always grounded. Around the large metal plate there are 25 smaller pins, connected to a 25-pin D-sub connector. The electrical contact is established to the nanojunctions by bonding these pins to the large metal pads of the samples.

The bonding is made by F&K Devoltec manual wedge bonder (see Figure 3.1.b). In the bonder head a 25 µm diameter aluminium wire, kept by a wedge, is guided through a needle hole. The vertical position of the bond head is controlled by motors while the stage can be moved horizontally manually. When the aluminium wire is touched to the metal surface the bonder starts vibrating the needle with ultrasonic frequency. Due to the surface friction the aluminium wire melts and sticks to the metal surface. This procedure is performed both at the pins of the PCB and at the metal pads of the devices. Finally the bonder breaks the aluminium wire.

After the electrical connections are established the samples have to be treated very carefully due to the electrostatic discharge (ESD). When two differently charged objects touch each other very large currents can be induced suddenly. In micro- and nanoelectronic devices the ESD can cause failure by breaking the circuit elements. Therefore, we have to avoid heavily charging materials and ground all objects and workers who get in touch with the samples. If the samples are already charged then the discharging has to be a slow, controlled process. For this purpose all ESD-safe
tools are made of high resistance and dissipative materials.

Attending to the precautions considerations we put the PCB into a vacuum tight sample holder, shown in Figure 3.1.c. On both sides of the cap there is a 25 pin connector, which ensure the electrical feedthrough. The pressure inside the sample holder can be decreased as low as $10^{-7}$ mbar using turbomolecular pump. The sample holder also serves as an electrical shield. The outer D-sub plug is connected to a switch box, which leads each pin to a BNC connector (see Figure 3.1.c). The outer part of the BNC is always on the ground, while the inner part can be switched between ground (down position of the adjacent switch) or float (up position) states. The sample is always connected to the inner part. The variable switch position serves to protect the samples against the ESD. When the switch is in the ground position the samples can not be charged and all electrical peaks from the environment are shorted.
3.2 Electrical setup

Most of the electrical measurements are carried out using National Instruments X-6363 data acquisition card (DAQ) with 2MS/s sample rate and 16-bit ADC resolution. An analog output is used for biasing the sample, while the current is measured by a current amplifier. Two kinds of current amplifiers are used, the first is a commercial one (Femto DLPCA-200) whose gain can be set between $10^3$ and $10^{11}$. In case of the electromigration process or the resistive switching the current can change several orders of magnitude. However, during the electrical measurements the gain can not be modified because a transient voltage peak appears at the input of the amplifier which would damage the device. Therefore, before the gain set the electrical measurement has to be stopped and the sample has to be grounded. This procedure increases the risk of the device failure. To avoid these problems a multi-channel current amplifier is also used. It was made by Dr. Gábor Mészáros [173]. It allows us to acquire the current data with five different, fixed gains simultaneously. The amplifications vary between $10^3$ and $10^8$ and one of them is always in the appropriate voltage level. The outputs of the current amplifier are acquired by the analog inputs of the DAQ. The schematic of the electrical circuit is shown in Figure 3.2.

![Figure 3.2: Schematic of the electrical circuit used for most of the measurements. b) The sample holder used for high frequency measurements.]

To study the timescale of resistive switching sub-$\mu$s long pulses are required, however such high frequency signals can not be supplied by the aforementioned measurement setup. The DAQ can not generate nanosecond long pulses and the BNC and D-sub cables can not transmit these signals without distortion. For high frequency measurements another setup was assembled. Arbitrary waveform generator (Agilent 33220A, 20MHz) is used as voltage source, while both the output voltage and the voltage drop on a series resistor are measured simultaneously by an oscil-
loscope (Picoscope 6402B, 250 MHz analog bandwidth). Thereby the resistance of
the device can be evaluated. Another sample holder is used for these measurements
to pass the applied pulses to the sample with low distortion (see Figure 3.2.b). It
is also a vacuum tight sample holder whose SMA connectors are directly connected
to the instruments via SMA cables. To the inner SMA plugs, a PCB is fixed which
serves the same role as the other PCB. Using this setup the time resolution is in the
order of 10 ns.

3.3 Measurement control program

The feedback controlled electrobreakdown processes are performed by a measure-
ment control program developed by myself. It was written in C# object oriented
programming language. The breakdown timescale during the electromigration in
metals is few tens of milliseconds [174], it would not require fast operation. How-
ever this program is also used for the electrobreakdown of graphene stripes whose
breakdown timescale is several orders of magnitude shorter. It is in the order of few
100 µs [116, 175], thereby fast feedback control is needed. During the bias voltage
ramp the acquisition rate of resistance must happen sufficiently frequently to detect
and react to the sudden change. When the critical resistance jump is observed the
bias voltage has to be dropped to low voltage region. However the specifications of
the DAQ and the PC control do not allow fast feedback process, the communication
between the two devices could take even several 10 ms. If a high voltage was applied
on the sample during this period of time, the breakdown could proceed further. In
order to eliminate this kind of delay time of the feedback control, the breakdown
process is performed by voltage pulses. The pulses are applied periodically, whose
periodicity has a minimal value of ≈ 20 ms which corresponds to the aforementioned
limit of the communication.

Figure 3.3.a shows the electromigration process of a 800 nm long and 100 nm wide
silver nanowire performed by my measurement control program. The breakdown was
performed in cycles and after each cycle the resistance was a bit higher (see Figure
3.3.b). In this specific case after 80 cycles I increased the resistance from 62 Ω to
900Ω. The cycles mean voltage ramps until the precursor of the breakdown, which
is a sudden fall in the current. However, in contrast to the similar measurement
protocols, in case of my program the voltage ramp is not the continuous increase of
bias voltage, rather discrete points as it can be seen in the inset of Figure 3.3.a, which
shows the last part of the first cycle. Each point corresponds to an applied voltage
pulse. The height of the pulse (V_{high}) is presented on the x-axis, while the current
measured during the pulse (I_{high}) is indicated on the y-axis. These pulses are applied
Figure 3.3: Feedback controlled electromigration of a silver nanowire performed by voltage pulses. The amplitude of the voltage pulse ($V_{\text{high}}$) was increased until the precursor of the breakdown, afterwards a new voltage ramp cycle was started. Meanwhile the current was measured simultaneously ($I_{\text{high}}$). The inset shows the last part of the first cycle. b) Evolution of the low bias resistance during the breakdown cycles. The initial resistance was 62 Ω and after 80 cycles it was increased to 900 Ω. Schematics on the top illustrates the structure of the silver nanowire at the different stages of the breakdown process.

periodically and these periods are the basic units of my program. In each period the resistance of the junction is calculated and the pulse amplitude is raised by a fixed value ($V_{\text{step}} = 1 \text{mV}$) if there was no feedback event. In the other case, if feedback event is detected the next pulse is applied with low amplitude ($V_{\text{low}}$) and thereby a new cycle starts. The schematics at the top of Figure 3.3.a illustrate the structure of the silver nanowire at the different stages of the electromigration. If the resistance of the metal junction is close to the conductance quantum, $1G_0=\left(12.9\text{kΩ}\right)^{-1}$, the two sides are connected by only few atoms [176]. Applying this electrobreakdown technique I can break a nanofabricated 100 nm wide and 400 – 1000 nm long silver nanowire in controlled way with 100% yield. In case of asymmetric silver structure, presented in Chapter 4, the yield of the electromigration is lower due to the better
heat conduction of the electrodes and the mechanical stress buildup along the wire [177], but it is still about 50%.

The block diagram of my measurement control program is shown in Figure 3.4.a. After the first initialization of variables and input/output channels (Panel 1-5) the program enters into the periodic pulse generation routine which is realized by a software timer (Timer tick panel). A period starts with applying the output voltage signal, which is illustrated on the Signal generation panel. The breakdown is induced by the voltage pulse in the middle of the signal. Before and after this voltage pulse the resistance is measured at low voltage level \( (R_{\text{low}}) \) by recording I-V characteristics with \( V_{\text{low}}=10 - 100 \text{mV} \) amplitude. In the first cycle the height of the pulse \( (V_{\text{high}}) \) is the same as the amplitude of the triangular signals \( (V_{\text{low}}) \). The length of the pulse \( (\tau_{\text{pulse}}) \) can be tuned from few \( \mu \text{s} \) to several seconds, while the low level I-V measurements have fixed 1 ms length. Therefore the total length of the signal is in the order of few ms if the pulse length is not longer than \( 1 - 2 \text{ms} \). While the output signal is applied the current is measured simultaneously. After applying the output signal the program reads the acquired current data and calculates the relevant parameters such as \( I_{\text{high}}, \, R_{\text{high}}, \, R_{\text{low}} \). After plotting and saving the data the program check whether the resistance reached the threshold limit. The threshold value means the desired final resistance value, where we would like to stop the breakdown procedure. If it is false the program stays in the loop and checks the feedback condition. Finally according to the feedback the new output signal is generated and new period begins. If the junction resistance exceeds the threshold value the timer is stopped and no other pulse is applied. The timer interval must be longer than the sum of the length of output signal and the calculation time, but it has the minimum value of 20 ms due to the aforementioned communications between the instruments.

The feedback condition corresponds to the relative change of one of these three quantities: \( \Delta R_{\text{high}}/R_{\text{high}}^{\text{ref}}, \, \Delta R_{\text{low}}/R_{\text{low}}^{\text{ref}}, \, \text{or} \, \Delta I_{\text{high}}/I_{\text{high}}^{\text{ref}} \). Before starting the electromigration we have to choose which parameter is monitored. If the relative change of the chosen parameter exceeds a critical value, feedback event occurs. This critical value can be set on the graphical user interface. In case of the metal junctions there are no substantial differences between these three quantities, for graphene this issue is studied in Section 5.4. Typically a smaller value (1-5%) must be given at the beginning of the process and as the resistance grows, it can be increased. Furthermore, in case of graphene large resistance changes can be caused by the Joule-heating induced cleaning, which is easily confused with the breakdown precursor. In order to avoid the false feedback events the resistance changes can be calculated by different reference values (initial value, global extremum or extremum of the last few hundreds of mV). If we would like to suppress the slow resistance changes due to other effects, like
Joule-heating induced cleaning, the reference value should be the maximum current or minimum resistance of the last $100 - 200$ mV voltage interval.

Since the applied waveform has complex shape it is essential to synchronize the timing of the output and input channels. It is realized by a common hardware start trigger. Both the input and the output tasks start to run if a rising edge appears on a programmable function interface (PFI) digital port. When the sampling is over, both tasks wait for another start trigger, meanwhile the output is set to zero bias. The trigger signal is generated by the program at the end of the timer period when the new output signal is generated. Figure 3.4.b shows the graphical user interface of the program. The parameters, such as $V_{low}$, $V_{step}$, $\tau_{pulse}$, threshold resistance, feedback conditions etc. can be set here. During the breakdown procedure the actual resistance and current is plotted on a graph and the breakdown can be stopped manually as well.

Besides the feedback controlled electrical breakdown process, I also implemented the further parts of sample characterization into the measurement control program (see Figure 3.4.a). The measurement automation enables us to perform the electrical characterization quicker and more uniformly which results in higher statistics. In case of graphene devices, discussed in Chapter 5 and 6, before the electrical breakdown it is necessary to measure the back gate dependent resistance of the graphene to determine its quality. During the gate measurement the back gate voltage, supplied by a DAQ controlled voltage amplifier, is varied, while I-V curves are measured at each gate voltage. The frequency and the amplitude of the gate and bias voltages can be set up on the graphical interface. The detailed analysis of the gate dependent resistance is discussed in Section 5.1.3. After the breakdown of the metal or graphene constriction the electrical characterization of the tunnel contact follows by measuring I-V traces. During these measurements very low current signals (few tens of pA) have to be acquired, thus the proper shielding and low noise setup are needed. These measurements can be performed by either multi-channel or Femto DLPICA-200 current amplifiers. The detailed study of the tunneling measurements is presented in Section 5.2.2. The same panel was used for the electroforming process or recording hysteretic memristor I(V) traces if the bandwidth of this setup ($\leq 500$ kHz) was suitable. The characterization of resistive switches requires various triangular signals, for this reason on the graphical interface we can set several parameters, such as the frequency, initial voltage sweep direction or the positive and negative bias limits. After stable resistive switching was established by the I-V measurements, the voltage dependence and the timescales of switching mechanism were investigated by pulsed measurements. During this study a series of voltage pulses are applied while the current is recorded real time. On the interface of the program we can set
Figure 3.4: a) Block diagram of the measurement control program used for the electrical characterization and electrobreakdown process. b) The graphical user interface of the breakdown panel.
among others the length and the amplitude of the pulses or the duty cycle. In case of SiO$_2$ resistive switches due to the high ON/OFF current ratio, the memristor I-V or pulsed measurements can be performed by only the multi-channel current amplifier. The measurement control program can partially process the measured data as well.

3.4 Conclusion

The development of a novel measurement setup for the controlled electrical breakdown of nanofabricated devices was a fundamental part of my PhD work. This included the development of a new high vacuum sample holder, the assembly of an optimized measurement setup and scheme and the development of a versatile measurement program. This setup was used for all the measurements in my PhD work. Furthermore, this setup is also used for ongoing projects, like in recent 1/f-type noise measurements on Ag-nanowire junctions [178]. The importance of this setup lies in the possibility, that we can prepare on-chip nanofabricated devices such that finally the size of the active device region is reduced well below the resolution of e-beam lithography. Compared to other similar measurement protocols [175, 179, 180] the specialty of my setup lies in the pulsed breakdown technique. This allows the exposure of the sample to much shorter voltage intervals than in real-time feedback systems. In my measurements I could go down to pulses as short as 5 \( \mu \text{s} \), while usual real time feedback systems have a sampling rate of 200 \( \mu \text{s} \).
In this chapter I describe my contribution to the study of geometrical asymmetry induced filament formation in Ag$_2$S based memristive system. In Section 4.1 I present the motivation of my work by giving a short introduction about the research history in this field performed at the Department of Physics, BUTE. Afterwards the sample fabrication and the device optimization steps are presented. Finally, in Section 4.3 I discuss the electrical measurements on the nanofabricated Ag$_2$S memristors, which are the first nanofabricated resistive switching devices of our group.

### 4.1 Research history

The Ag$_2$S based memristors are a well studied memristive system in the Solid State Physics Laboratory of Department of Physics, BUTE. The measurements were established in an STM arrangement, where the nanometer sized contact was formed by touching an inert STM tip (PtIr) to the surface of Ag substrate covered by Ag$_2$S thin layer. The sketch of the obtained Ag-Ag$_2$S-PtIr structure is illustrated in the inset of Figure 4.2.a. Using this setup several notable results were presented such as inducing resistance transition with sub-ns pulses [96], investigation of the switching dynamics [181] or the role of the local heat dissipation [95].

The memristive behavior of Ag$_2$S is described by electrochemical metallization (ECM), where the bipolar switching mode is explained by the different electrode materials (see Section 2.1.4). During the electroformation a silver filament is formed inside the insulating material. Afterwards the conductive filament can be ruptured by applying negative voltage to the tip, while it can be recovered by applying positive voltage. This polarity rule never changes during the device operation. In the hysteric
I-V characteristic this rule appears such that the direction of the hysteresis loop is fixed. Figure 4.1 shows the typical I-V traces of Ag$_2$S junctions for various conductances. The left panels show the contacts whose on-state conductance lies between $50 - 300 \ G_0$, where $1 \ G_0 = 2e^2/h = (12.9 \ k\Omega)^{-1}$ is the conductance quantum. In this regime the junctions exhibit regular bipolar switching I-V characteristics described in Section 2.1.1. The current strongly increases as the bias voltage exceeds the positive threshold voltage and the conductance increases as long as the applied voltage is ramped up. The direction of the current–voltage loop, indicated by the arrows, are the same for all the curves. The junction diameter was estimated to 10nm by the Sharvin expression [182]. The switching mechanism corresponds to formation and destruction of Ag filament in the Ag$_2$S matrix. It has to be noted that the switching characteristics significantly changes if the conductance of the on-state decreases below $20 \ G_0$ (panels on the right in Figure 4.1). The junctions still exhibit bipolar switching but stepwise resistance transition occurs between well defined states. After the current jump at the positive voltage polarity the resistance is fixed until the back-switching at the negative voltage polarity [183]. The switching voltage scatters in a broad range and the direction of the loops (see arrows) is random, both switching directions were observed in 50-50% of the cases. Since the conductance of the junctions are close to the conductance quantum, the size of the contact is at atomic scale [176]. The switching of atomic contacts is based on displacement of single or few atoms at the vicinity of the narrowest cross-section.

The presented polarity robustness of the memristive switches (left panels) could be explained by the different electrode materials (silver and PtIr), however after the first rupture of the silver filament both broken ends are made of silver. Therefore the fixed switching loop can not be caused by the material asymmetry, another phenomena must be present in this system. The geometrical asymmetry of the electrodes could be an explanation. This asymmetry effect may be presented in the different behavior of ECM type and atomic switches. In case of Ag$_2$S based memristor the junction has few nanometers size, it is large enough to notice the local geometrical asymmetry of the sharp STM tip and the flat bottom electrode (see insets of Figure 4.2.a). In contrast the atomic switch (right panels) has much smaller size and the local atomic asymmetry determines the switching direction which varies randomly.

To study this issue the switching direction was tested in different arrangements (STM, MCBJ and on-chip) using symmetric electrode material structure, i.e. instead of the PtIr top electrode in this research both electrodes were made of Ag. In this research topic the STM and MCBJ measurements were performed by Á. Gubicza, whereas I performed the electrical measurements on lithographically designed on-chip
Figure 4.1: Typical I-V traces of Ag$_2$S contacts with various conductance. The left panels correspond to the junctions with low on-state conductance (50 – 300 $G_0$) and exhibit regular bipolar switching. On the right side the conductance of the on-state decreases below 20 $G_0$ and the switching mechanism changes to atomic switching. The circuits were not purely voltage driven, which accounts for the back-turning of the I–V curve \cite{183}.

devices. The sample fabrication was first done by Dr. Miklós Csontos, and later, after the first published results \cite{2} I also participated in the device optimization process and I also prepared on-chip devices. In the following I describe briefly the context of these studies about the geometrical asymmetry, these results are based on the work of Dr. Ágnes Gubicza and this topic is discussed in more detailed in her PhD thesis \cite{184}.

In case of an STM setup the geometrical asymmetry is obviously given by the sharp tip and the flat thin layer. Figure 4.2.a shows two characteristic hysteretic I(V) traces, with the same switching direction marked by arrows. By convention, the positive bias corresponds to the case when the positive voltage is applied to the thin and flat layer, while the tip is grounded. At positive bias the filament gets
thicker by the migration of the silver atoms towards the tip. The growth of the conductive channel results in lower resistance. By applying reversed bias the inverse process takes place, the filament thins and becomes less conductive. All curves showed fixed switching direction regardless of the polarity of applied voltage during the electroformation or the initial direction of triangular voltage signal, i.e. the set transition always occurs at positive bias [2].

![Figure 4.2](image-url)  
**Figure 4.2:** Representative hysteric I-V traces measured in a) STM and b) MCBJ arrangement on Ag-Ag$_2$S-Ag junctions. The straight arrows indicate the initial configurations, while the curved arrows illustrate the switching direction. The insets show the sketch of the setups and the biasing convention for the STM. In case of the STM the switching directions were the same for all traces, while for MCBJ random distribution was observed [2].

During the MCBJ measurements a pure silver wire with diameter of 125 µm was mechanically ruptured in a controlled way. After the rupture the as broken electrode surfaces were sulfurized in-situ in vacuum. Finally the junction was closed, and resistive switching traces were recorded. The sketch of the arrangement is illustrated in the inset of Figure 4.2.b. Using this method very high mechanical stability can be achieved. In contrast to STM setup, after the rupture of the wire the local asymmetry is indefinite at the narrowest cross-section. If the geometry plays the crucial role, the switching direction should vary randomly after each rupture. Figure 4.2.b shows an I(V) curves of two different junctions and as it is marked they have opposite switching direction. Overall, both switching directions were represented in 50-50% of the cases, which is in agreement with the expectation.

In order to understand the kinetics of the electric field driven filament evolution
on a microscopic level, molecular dynamical simulations were also performed by Dr.
Dávid Zsolt Manrique at the Physics Department of Lancaster University [2]. The
simulated system reproduced the experimental findings and supported the role of the
geometrical asymmetry.

As the last step to confirm this phenomenon, the switching behavior was tested
on such a nanofabricated device, which mimics the geometry of the STM setup. The
nanofabricated devices have further advantages compared to the STM arrangement,
such as higher mechanical stability and they can be integrated into larger circuits.
This part of the research was my task.

4.2 Sample preparation

The design of the nanofabricated devices were defined by standard electron beam
lithography (EBL) using lift-off process on Si substrate covered by 80\text{nm} thick SiO$_2$
and 140\text{nm} thick Si$_3$N$_4$ on the top. The silicon nitride serves to exclude the resistive
switching behavior of SiO$_2$ as it is discussed in Section 2.1.3 and Chapter 6. Since
both electrodes are realized from silver it enables us to simplify the sample prepara-
tion process into one lithography step. Figure 4.3 shows the design of the first set
of samples. The geometry of the electrodes mimic the asymmetry of the STM setup
by defining a rectangular and a triangular silver structure, which are connected by
a narrow, only few tens of nanometer wide silver channel. The active region of the
resistive switches are formed at this part. Far from the constriction the silver struc-
ture broadens and the electrodes end in 150\times150\text{$\mu$m} large metal blocks. They serv

easily achievable electrical contacts between the resistive switch and the measure-
ment setup. Note that the large patches on the metal pads are the residual part of
the aluminum wire used for the electrical contacting of the devices. For the electron
beam lithography process double-layered poly(methyl methacrylate) (PMMA) resist
were used (see Figure 4.4.a). In the lower, more sensitive resist layer (PMMA 50k)
the electron-beam exposes larger area than in the top layer (PMMA 600k) resulting
in an undercut profile. This resist structure helps to remove the resist easier during
the development and enables higher spatial resolution [185].

The silver was evaporated by electron-beam physical vapor deposition technique
with uniform thickness of 40\text{nm}. To ensure better adhesion between the substrate
and the silver, the surface of Si$_3$N$_4$ was treated by Ar plasma. After the samples were
taken out from the vacuum chamber of the evaporator system they were placed into
another vacuum chamber or acetone as soon as it was possible to avoid the oxygen
and carbon contamination of the surface [186]. For this reason it is important to put
the silver device into the measurement setup as soon as possible.

55
Figure 4.3: Optical (I and II) and SEM (III) images about the asymmetric Ag-Ag$_2$S-Ag on-chip samples after the electrical measurements. The samples are made of two large metal pads (I) and a constriction (III) at the middle with asymmetric geometry. On the SEM image the white scale bar in the lower right corner indicates 200 nm [2].

During the lift-off significantly portion of the devices flaked off the substrate due to the poor adhesion to the Si$_3$N$_4$. To get better device fabrication yield titanium adhesion layer was deposited before the silver. In order to avoid the parallel conduction at the constriction, during the evaporation of the titanium the sample holder was tilted. Since at the constriction the height of the PMMA mask is similar to the width of the structure, the titanium is not present at the constriction under the silver, as illustrated in Figure 4.4.a. Due to the undercut profile of the double layer PMMA a narrow stripe appeared parallel to the edge of the structure, as it can be seen in the SEM image (Figure 4.4.b). The adhesion could be improved by depositing a 1 – 2nm thick titanium layer under the silver. Such a thin Ti layer does not form continuous layer and thus do not provide parallel electron conduction. These deposition processes substantially increased the yield of the sample fabrication.

As it is discussed in Section 4.3 the sulfurization is performed by placing the pure silver devices into sulfur rich atmosphere. The aforementioned device are made of a single, uniformly thick silver layer. However, it is not feasible to expose only the constriction part to the sulfur. Therefore, if we performed the sulfurization process on the freshly made samples, all parts would transform to Ag$_2$S and the devices
would not exhibit resistive switching behavior. We must localize the active region. We attempted to achieve it in two ways. The one is to break or thin the silver wire at the narrowest cross-section using electromigration technique. This thinned region transforms to Ag$_2$S sooner than the other parts of the device. By controlling the sulfurization time it is possible to form few tens of nanometer sized switching region. The second option is to fabricate a shadows mask for the sulfurization process which is open at the constriction, but protects the rest parts of the device. During my work we tested three different mask structures.

As the simplest case we covered the whole chip by few 100nm thick PMMA layer and the polymer was removed at the constrictions using another lithography step. The geometry of the silver junctions was the same as presented above. In order to reduce the device to device deviation due to the sulfurization process, the devices were packed closer to each other in groups of ten. This design aids to expose constrictions to the same amount of sulfur. The device arrangement is shown in Figure 4.4.c. However, the sulfurization was not effective using this PMMA shadow mask, the samples showed hysteric I-V curves only after 30 minutes sulfurization time. The passivization against the sulfur may arise from the long-term mechanical instability of the PMMA slit, the surface of the constriction is covered by a thin polymer layer.

To overcome this problem the shadow mask was prepared by evaporating titanium layer to the top of the silver. It was carried out by tilting the sample holder at two
angles with the same magnitude but opposite direction. Figure 4.5 shows a SEM image about the surrounding of the constriction. Due to the two different tilted evaporation processes the parallel lines are present at both sides of the electrodes. The dashed white lines around the constriction indicate the contour of the titanium, inside this region the silver is not covered for the same reason as discussed. Despite of the more sophisticated mask fabrication these devices did not show better sensibility to the sulfur. After exposing the devices to 30 minutes long sulfurization only a weak resistive switching effect could be observed. Despite of the tilted sample holder a thin titanium layer may have covered the surface of the silver at the constriction as well.

![Figure 4.5: SEM image of a device using titanium shadows mask for sulfurization. The dashed white lines around the constriction highlight the contour of the top titanium mask layer. The white scale bar at the lower left corner indicates 200 nm](image)

The usage of shadow mask can be avoided if the silver layer is not evenly thick. This structure can be realized in the same way as the titanium mask. After the perpendicularly evaporated 40nm thick silver layer, additional silver was deposited in tilted sample holder at two different angles. Thereby the silver is thinner at the constriction in the same way as it is shown in Figure 4.5. The full deep sulfurization takes place sooner at this region than the other parts of the device.

Taking SEM images about the constriction before the sulfurization also puts the cleanness of the devices at risk. During scanning of the surface by the electron beam some organic contamination could burn onto the silver which forms a passivation layer. Thereby after the metal deposition it is not worth putting back the devices into the electron microscope to look at their structure.

The first set of devices and the samples with masks were fabricated by Miklós Csontos. Later I also fabricated devices without masks. All samples were made at Microtechnology Department of MTA EK MFA.

58
4.3 Resistive switching in asymmetrically shaped Ag-Ag\(_2\)S-Ag on-chip memristors

As it is discussed in the previous section the first set of devices were made of a single, uniformly thick silver layer. In order to localize the active region of the resistive switch, at first we need to thin the silver layer at the constriction using electromigration. Figure 4.6.a shows the electromigration process of a sample performed by 500\(\mu\)s long voltage pulses on a current-voltage graph. The initial resistance of the device was 55\(\Omega\). During the controlled migration the current was continuously monitored and a new bias voltage ramp was started from 10mV if the current decreased by 5% in the last 100mV. Left panel of Figure 4.6.b shows the evolution of the device resistance as the function of the bias ramp cycles. The electromigration process was stopped when the resistance reached 250\(\Omega\) because the contact showed instabilities. Afterwards the device resistance was continuously acquired by measuring I(V) characteristics with amplitude of 10mV (right panel of Figure 4.6.b). At room temperature the silver exhibits large surface diffusion, thereby the resistance of the partly electromigrated junction constantly increased and finally the metal connection broke. When the conductance of the junction goes below 10\(G_0\) the cross-section of the contact consist of only few atoms. In this regime the resistance changes in discrete jumps, as the size of the contact decreases by atoms. The plateau at around 1\(G_0\) right before the breaking corresponds to the single atom contact. After the self-breaking of the junction S-shaped tunneling I(V) characteristics were measured (see Figure 4.6.c, black dots) referring to nanometer-sized spacing between the broken silver electrodes. By fitting Simmons-model 1.02nm was obtained to the gap size (red line).

Along the sulfurization process of the junction I followed the work of Ágnes Gubicza, which was developed for STM samples. The sulfurization of the on-chip devices were performed under ambient conditions. The sulfur powder was placed onto the surface of a heater and the devices were placed about 1cm above the sulfur. At this time the chip was already fixed to the PCB presented in Section 3.1. During the optimization of the sulfurization the heater was set to fixed 60°C and the time was varied. In each step the sample was exposed to sulfur vapor for 1min and after the switching behavior was tested in vacuum. If no resistive switching was observed the sulfurization process was repeated again. Finally after 3 minutes the sample showed reproducible resistive switching.

After the sulfurization electroformation was not needed, the contact was in low resistance state \((R \approx 85\Omega)\). By increasing the amplitude of the I-V measurement to 200mV the hysteric I-V traces appeared. Figure 4.6.d shows a representative
Figure 4.6: a) Current-voltage traces of the voltage ramp cycles during the electromigration of asymmetrically shaped silver nanocontacts. The measurement was performed in vacuum at room temperature. b) Evolution of the device conductance during the electromigration process (left panel) and during the self-breaking (right panel). c) I-V characteristic of the tunneling contact after the self-breaking (black dots). By fitting the Simmons model (red line) the main parameters could be extracted (annotation). d) A representative hysteric I-V trace of the nanofabricated Ag-Ag_{2}S-Ag nanojunction. The inset shows the SEM image of the device. The white scale bar in the lower right corner indicates 200 nm and the convention of the bias voltage polarity is also shown. Positive voltage means positive biasing of the bottom flat part. e) 20 consecutive I-V traces acquired on stable nanojunction configurations. $R_{OFF} = 279 \pm 15 \, \Omega$, $R_{ON} = 155 \pm 5 \, \Omega$ \textit{[2].}

I(V) curve of the formed Ag_{2}S memristor. It is highly similar to the ones obtained in the STM and MCBJ setups. The switching direction, indicated by the arrows, agrees with the expectations of asymmetry induced filament evolution, i.e. the set transition occurs at positive voltage. The formation and destruction of the conductive
channel may take place at a few 10 nm long region of Ag$_2$S. The device showed about 150 switching cycles with 1.5 OFF/ON resistance ratio. Afterwards it turned into unresolvable high resistance state. Figure 4.6c shows 20 consecutive I-V traces, which indicates high reproducibility both in ON and OFF resistances. The switching occurs between $R_{ON} = 155 \pm 5 \Omega$ and $R_{OFF} = 279 \pm 15 \Omega$.

The measurements in Figure 4.6 are considered as the first proof of principle results on asymmetrically shaped on-chip Ag-Ag$_2$S-Ag resistive switches. At the same time these are the first results in our lab, where resistive switching was achieved in an on-chip environment instead of the STM-based memristor junctions. These devices had great mechanical stability, the tunnel contact could be sustained for hours, while in the former STM arrangement the contacts were stable only for few minutes [184]. Therefore the devices showed good resistance reproducibility in the consecutive switching curves, however, the poor endurance, vanishing of switching ability after few 100 of switching cycles, refers to the intrinsic instability of the active region.

After the first set of devices our attention turned into the samples fabricated with shadow mask. Applying mask would greatly simplify and make more reliable the sulfurization process. However, most of the samples exhibited only weak effect or nothing when they were exposed to the sulfur regardless of the mask structure. In order to get feedback about the structural changes of the samples during the sulfurization, the resistance was measured in-situ. Figure 4.7 shows the two typical in-situ resistance measurements performed on two different devices on the same chip. It was a general observation that these samples reacted with the sulfur much less sensitively than the first set of devices without mask. The first type of samples (Figure 4.7.a) did not show any reasonable resistance change even after 30 minutes and they did not show resistive switching behavior. On the other hand the device in Figure 4.7.b exhibited a slow resistance increase and finally a sudden jump was observable. Removing from the sulfur rich environment at this point, the resistance remained in the range of few k$\Omega$. However these devices mostly showed unstable resistance jumps with large ON/OFF ratio, presented in Figure 4.7.c-d. The switching direction was not well determined. These observations suggest that the cleanness of the silver plays crucial role in the reliable device fabrication. A thin passivation layer on the surface can hinder the formation of Ag$_2$S region at the constriction. The slight difference in the surface contamination can cause the large variety of device behavior during and after the sulfurization.

Devices without any mask were fabricated again by myself to test the reproducibility. Using subsequent electromigration and sulfurization processes resistive switching could be induced, but it has very low yield presumably due to the surface
Figure 4.7: In-situ resistance measurements during the sulfurization of as-prepared silver devices using PMMA mask. In most cases the resistance of the samples either a) did not show any changes or b) a very slow one in the sulfur rich environment. c-d) Representative hysteretic I(V) curves with high ON/OFF ratio. The arrows indicate the switching direction, which is opposite as the geometrical asymmetry induced case.

contamination. However, neither in stability nor in reproducibility of resistances they did not exceed the prototype. Figure 4.8 shows 10-10 switching cycles of the two most stable devices. The switching direction, indicated by the arrows, agrees with the previous results. However, after 40-60 cycles the switching behavior disappeared, the hysteresis of the I(V) traces closed. We had to increase the amplitude of the I-V measurement to sustain the switching effect, but the resistance of the contacts increased continuously and finally their switching capability failed.

Beside the ECM induced resistance transitions, atomic switching was also observed in sulfurized silver contacts. As it was presented in right panel of Figure 4.1 if the size of the junction is in atomic scale the resistance changes in discrete jumps. Figure 4.9 shows 20 consecutive I-V curves of an atomic switch device. Resistance jumps occur at wide range of voltages and the device preserves its new resistance state until the back-switching at the opposite polarity [183, 187]. The switching direction of a specific atomic switch is fixed, but varies randomly for different junctions
Figure 4.8: 10 consecutive I-V traces of the two stable Ag$_2$S based memristors. The resistance transition occurs between a) $R_{ON}=2.8\,k\Omega$, $R_{OFF}=4.1\,k\Omega$ and b) $R_{ON}=150\,\Omega$, $R_{OFF}=190\,\Omega$. The arrows indicate the switching direction.

according to the local geometry [183]. The atomic switches had outstanding stability, more than thousand switching cycles could be measured without any degradation. This observation is in contrast with the thermal instability of pure silver nanowire during the electromigration, presented in Figure 4.6. The much higher stability of the atomic switch may arise from the stabilization effect of the Ag$_2$S matrix. It confirms that the low endurance of the on-chip Ag$_2$S memristor devices is not the consequence of the poor mechanical stability, rather the intrinsic instability of the switching process.

4.4 Conclusions

Our research group demonstrated stable resistive switching behavior of Ag-Ag$_2$S-Ag memristors in STM and MCBJ arrangement based on local geometrical asymmetry. In contrast to the conventional active-passive electrode structure both electrodes were made of silver. The well defined switching direction was the result of the inhomogeneity of the local electric field. In order to demonstrate this phenomena in on-chip devices I fabricated lateral Ag-Ag$_2$S-Ag devices using standard electron-beam lithography technique. To localize the active region of the resistive switching memories I created nanometer-sized gap in initially continuous silver wires by controlled electromigration and afterwards I sulfurized the created nanogaps in-situ. I have demonstrated that the resistive switching can be established using a simplified, all-Ag sample design, which allows us to fabricate the devices using only one lithography
Figure 4.9: Representative I-V traces of 20 consecutive switching cycles of silver atomic switch. The resistive switching occurs between $R_{ON}=4.5\,k\Omega$ ($G_{ON}=2.9\,G_0$), $R_{OFF}=11\,k\Omega$ ($G_{OFF}=1.2\,G_0$).

step. The fixed direction of the switching confirmed the phenomenon of asymmetry induced filament formation. These are the first results in our lab, where resistive switching was achieved in an on-chip environment instead of the STM-based memristor junctions. The on-chip device showed increased mechanical stability compared to the other two arrangements. In similar devices I have also demonstrated stable room temperature atomic switching phenomenon, indicating that the surrounding Ag$_2$S matrix stabilizes the atomic switching process.
In this chapter I describe my results about the nanogap formation in a continuous graphene nanostripe, which can be used as the electrodes of nanometer sized resistive switches. In the first section I introduce the sample fabrication processes. After demonstration of high yield graphene nanogap fabrication and the characterizations are presented. Next I study the mechanism of the breakdown process in different environment conditions. Finally I describe my optimization steps to achieve controllable breakdown process.

The results presented in this chapter were realized by the collaboration with research group of Michel Calame in Basel. Before our collaboration they had already worked on controlled breakdown of graphene stripe grown by chemical vapor deposition (CVD) technique. They fabricated their samples by photolithography technique, the minimal width of their graphene constriction was about $3\mu$m. However the yield of the nanogap formation was low regardless of the environment conditions. Meanwhile, during the first year of my PhD, I was also working on the same project using mechanically exfoliated graphene. At the beginning of our work I visited their research group at the University of Basel for two months. In order to improve the nanogap formation process I fabricated narrower graphene junction using electron beam lithography with the help of Cornelia Nef and Peter Makk. During my visit I learned to grow graphene on copper foil using CVD technique and the whole sample preparation procedure described in Section 5.1.2. After my visit either the fully fabricated samples or the graphene on copper substrate were sent from Basel to Budapest. These sample preparation processes were performed by Cornelia Nef or Maria El Abbassi. All graphene sheets presented in this thesis were grown in Basel.

The investigation of graphene nanogap formation was performed together however
the aim of its application was different. Research group of Michel Calame utilized
the graphene nanogaps to contact single organic molecules, while our aim was to
establish few nanometer sized resistive switch inside the nanogap. In order to clarify
my contribution to this study at the caption of each figure I note who performed the
measurements and where.

5.1 Sample preparation and experimental techniques

The first trials to make nanometer-sized gaps in graphene were performed on
mechanically exfoliated graphene samples in Budapest. I used adhesive tape to
split graphite into few-layers thick graphene. It is one of the simplest production
 technique, since no special and expensive instrumental or material requirements are
needed [188]. The exfoliated graphene typically has excellent structural quality with
low defect concentration, which provides high electron mobility and it is ideal to
study exotic physical phenomena [145]. The first paper in the field of graphene
nanogap formation by electrical breakdown was also published on multi-layer ex-
foliated graphene flakes [116]. Graphene with various thickness could allow us to
investigate the breakdown process as the function of the number of layers. How-
ever, it is very difficult to control the thickness, the ideal graphene flakes must be
identified individually under optical microscope among the thick ones. Each sample
needs unique design and hence the sample preparation is exhaustive and takes a lot
of time. Due to diversity of the samples it is hard to make identical samples and
study the mechanism of the breakdown with statistical analysis. In the aspect of the
application as an electrode material the low defect density and the high charge car-
rier mobility is useful but not necessary. The high reproducibility and uniformity is
more desired. Therefore after some sample characterizations on exfoliated graphene
our attention was turned to graphene made by chemical vapor deposition (CVD)
technique.

5.1.1 CVD growth and transfer

For industrial applications it is essential to produce large scale homogeneous
graphene sheets. Exfoliation with scotch tape offers only micron scale flakes with
various structures. Since the first isolation of the graphene several production tech-
niques were developed both in a top-down (chemical [189, 190] and mechanical exfo-
liation [133, 191, 192], chemical synthesis [193–195], etc.) and bottom-up (epitaxial
growth [196, 197], unzipping of CNT [198, 199], CVD [200, 201], etc.) approach. One
of the most advantageous technique for large-scale device fabrication is the thermal CVD process [202].

The chemical vapor deposition is a frequently applied chemical process to produce thin films by depositing gaseous reactant onto a substrate. The gas molecules are combined in a reaction chamber typically at elevated temperature. When the precursor gases come into contact with the substrate reaction occurs and thin film forms on the surface. In case of graphene, transition-metals, such as Ni [203], Cu [201, 204], Ir [205] substrates are used to get high quality sheet. The CVD growth of graphitic layer on metal surface is known for half a century [206, 207], but its significance raised after the first characterization of the graphene. The CVD process is a straightforward way to produce large area and reasonably good quality graphene, although some special equipment and strict control of the environment (gas flow rate, pressure, temperature etc) are necessary [208]. The advantage of CVD technique compared to e.g. epitaxial growth on SiC, is the possibility of transferring the graphene onto various substrates.

The copper, besides the role as substrate, has catalytic effect. During the CVD process methane is used as carbon source. It decomposes at the surface of the copper and forms graphene layer. The presence of copper reduces the energy barrier and hence the reaction temperature [201]. The weak bond between the carbon and the copper stabilizes the structure of the graphene. The solubility of carbon atoms into the copper is very small, therefore the growth of the graphene layer is self-limiting. The deposition stops after the surface is covered by one layer thin graphene, hence the ratio of two and three layers graphene is low [209]. After the deposition, the copper can be etched by acid without affecting chemically the graphene and with low contamination of copper atoms.

The optimal parameters of the deposition can be different for each CVD setup and long optimization produce is required to produce good quality graphene reliably. The protocol of the graphene growth and transfer was developed by Dr. Cornelia Nef and Dr. Wangyang Fu. Herein it is described briefly, the schematics of the fabrication steps are illustrated in Figure 5.1. The copper is a polycrystaline 25 µm thick foil with 99.8% purity from Alfa Aesar company. Before the deposition, the copper has to be treated. At first the foil is washed in acetone, in isopropyl alcohol (IPA) and in distilled water to remove the organic contamination. In the next step the native oxide has to be etched chemically by 1:2 mixture of orthophosphoric acid (H₃PO₄) and distilled water. The CVD process is performed in Carbolite HZS horizontal split 3 zone tube furnace at low pressure with base pressure of 0.015 mbar. During the whole process the gas flows are controlled, the hydrogen (H₂) is constantly set to 10sccm flow rate with the pressure of ≈ 0.37 mbar. The growth is performed at
1000 °C, when this base temperature is reached at first the copper is annealed for 10 minutes in the hydrogen atmosphere to remove the residual oxides and increase the grain size. After that we open the methane (CH\(_4\)) source with 25 sccm flow rate for 5 minutes. The methane decomposes at the surface of copper and the graphene starts to grow at several nucleation points in radial direction. The reaction stops when the whole surface is covered by single layer graphene due to lack of the catalytic effect of the copper. The H\(_2\) acts as a control reagent. During the deposition the pressure is \(\approx 0.85\) mbar. At the end of the growth the CH\(_4\) flow and the heating is switched off, while the H\(_2\) flow still remains. When the temperature decreases below 300 °C the H\(_2\) source is also closed and the vacuum tube is flushed with argon gas (100 sccm, 1 mbar). By applying higher temperature or pressure the ratio of multi-layer graphene to the single layer one increases.

![Diagram of graphene growth](image)

**Figure 5.1:** a) Illustration of the graphene growth by CVD process. At first the copper foil has to be cleaned (I) and after the oxide layer must be removed (II). The graphene growth takes place in a furnace at 1000 °C using CH\(_4\) precursor. b) Schematics of the wet graphene transfer from copper to insulating substrate. Adapted from [210].

For electrical measurements the graphene has to be placed from the copper to an insulator substrate. Several methods were proposed to the transfer, we use wet transfer described by Li et al. [211]. In this case the copper foil is etched in diluted acid solution, as illustrated in Figure 5.1.b. During the transfer a lot of attention has to be paid not to damage or contaminate the graphene. The copper is covered
on both sides by the graphene, which protects the metal from the acid. Therefore at first the graphene has to be removed from one of the sides by placing it into oxygen-argon plasma. In order to protect the graphene on the other side we coat it with 300 nm thick poly-methyl methacrylate (PMMA) layer. The polymer layer will also stabilize the graphene after the copper is etched beneath. After removing the graphene the copper is placed on the surface of 0.1 M solution of ammonium persulfate ((NH₄)₂S₂O₈). During the wet etching there is a risk of formation of bubbles which can damage and crack the graphene. It can be avoided by using lower concentration of acid. In order to get rid of the residual of acid and copper the graphene sheet is flushed several times by changing the distilled water bath. Finally we place the insulator substrate below the floating graphene and lift the substrate up, while the graphene sticks on the surface. After the transfer we leave the sample to dry on air for a day and finally the PMMA layer can be removed. Before transfer the wafers were cleaned by acetone, IPA distilled water and Ar/O₂ plasma.

Figure 5.2: Optical image of a transferred and patterned CVD graphene. The multi-layer graphene parts are highlighted in blue circles. During the samples preparation holes can be formed in the graphene sheet (red circle) and the surface get contaminated (green circle) by organic molecules. The sample and the optical image were made by me in Basel.

Doped silicon wafer was used as substrate covered by two different types of insulator layers. The layer thicknesses were chosen to give the maximum optical contrast between the naked substrate and the graphene owing to interference effect. Most of the measurements were performed on 300 nm thick SiO₂ layer. Beside this the structure of 140 nm thick Si₃N₄ layer on 80 nm thick SiO₂ was also used when the silica was not desired beneath the graphene due to its oxygen content or resistive
switching behavior. The CVD graphene can be characterized by several methods. The structural defects like holes or multi-layer islands can be detected by optical microscope. Figure 5.2 shows a shaped CVD graphene on SiO$_2$ substrate. The blue circles indicate the double- or multi-layer graphene parts, while the red circle shows a larger hole in the graphene lattice. As it can be seen, major part of the graphene sheet consist of a single layer. During the fabrication process the contamination by the resist layers can not be removed fully, some residues are highlighted by green circle. More detailed information can be obtained about the quality of the graphene by performing Raman spectroscopy [209, 212].

5.1.2 Sample design and patterning

After transferring the graphene to the Si substrate it is ready to pattern into the desired shape. The design of the graphene devices is defined by electron beam lithography (EBL) process. Two kinds of samples were fabricated whose images are shown in Figure 5.3. During first lithography process two large metal contacts are defined for each device, which give stable electrical contact between the graphene and the measurement setup. The typical layer structure of the metal is 5 nm titanium (Ti) and 50 nm gold (Au) which are made by electron-beam physical vapor deposition technique. In the next step the graphene sheet is tailored using another EBL process. In this case we expose those regions by the electron beam where the graphene is etched by Ar/O$_2$ plasma. For these lithography steps PMMA or ZEP 520A resist was used.

Concerning the chip design some aspects must be taken into account. It is important to localize the nanogap for further investigation and application. For this purpose a nanometer sized constriction is defined (see panels III. in Figure 5.3.a and b), where the current density increases considerably resulting in large local temperature compare to the wider parts of the sample. In the next step we have to consider to minimize the series resistance which serves to get better yield for the nanogap formation as described in Section 2.2.1. For this purpose the metal electrodes have to be positioned as close as possible. During an EBL process we can fabricate hundreds of samples but as the number of samples increases the alignment between the metal and the graphene layer can be shifted even several micrometers. To avoid the intersection between the constriction and the metal pads we have to keep distance between them. Depending on the alignment procedure this spacing was varied between 5 – 20 $\mu$m. The graphene constriction and the metal contacts are connected by larger graphene electrodes, as the dashed white lines indicate in panel II. of Figure 5.3.a. To lower the series resistance of this part high width to length ratio is desired in accordance
with Equation 2.20. Therefore, the graphene broadens with large angle after the nanojunction. Finally, the graphene is etched around the samples to separate them electrically from each other. Owing to the wafer scale graphene sheet hundreds of samples can be fabricated on a single chip, which allows us to make statistics about the breakdown process.

Figure 5.3: Optical (I and II), SEM (a.III) and AFM (b.III) images about the a) rectangular and b) bow tie as-prepared samples. The samples made of two large metal pads (I), two large graphene electrodes between the metal structures (II) and a graphene constriction (III) at the middle. The first sets of samples (a) were made in Basel, the second sets of samples (b) were made in Budapest at MTA EK MFA.

Two different types of samples have been fabricated. In case of the first sets of samples the constriction has rectangular shape by varying their lengths and widths. Table 5.1 contains the parameters of the different geometries. The different aspect ratios vary the ratio of the series resistance ($R_s$) to the junction resistance ($R_j$). The
distance between the metal pads is about 20 µm. The first samples were made by me with the help of Cornelia Nef during my visit in Basel. Later they were prepared by Cornelia Nef and Maria El Abbassi.

<table>
<thead>
<tr>
<th></th>
<th>Width (nm)</th>
<th>Length (nm)</th>
<th>L/W</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>400</td>
<td>800</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>400</td>
<td>400</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>600</td>
<td>300</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Table 5.1: List of the widths and lengths of fabricated devices with rectangular constriction.

The second sets of samples were fabricated with bow-tie shaped constriction (see panel III. in Figure 5.3.b). The width of the narrowest part was varied between 100 nm and 350 nm with 50 nm steps. It allows us to study the effect of the minimal cross section to the breakdown process. By applying better alignment technique the metal contacts were placed 5 µm far from each other, but they could be even closer. These samples were made by me at MTA EK MFA using CVD graphene supported by the University of Basel.

Most of the measurements were performed on the rectangular shaped constrictions, in Section 5.2 and 5.3 all presented results originate from this kind of sample. The bow-tie geometry was only used during the controlled breakdown, described in Section 5.4.

5.1.3 Electrical characterization and cleaning

Following the graphene patterning the sample is ready for the electrical characterization. The same measurement setup was used for graphene devices as introduced in Chapter 3. The yield of the sample preparation is not 100% due to the holes and cracks in the graphene. Since the graphene is not sensitive to the contamination in air, it is worth testing the devices before bonding them and putting into the vacuum chamber. Using needle probe station temporary electrical contacts can be formed and the resistance of the devices can be measured. Under a high working distance microscope metal needles, positioned by micromanipulators, can be touched to the surface of the Ti/Au metal contacts. The resistance measurements are performed by Keithley Source Meter. After testing the samples fixed contact can be created by the bonder. In case of room temperature measurements the same printed circuit board (PCB) is used as shown in Figure 3.1.a. In case of graphene the good electrical con-
tact to the back gate is essential. Figure 5.4 shows the sketch of the electrical circuit used both for characterization and electrobreakdown process of graphene devices. The only difference compared to the measurement setup of the Ag$_2$S memristor is the gating of the devices.

![Schematic of the electrical circuit used for graphene characterization and electrobreakdown.](image)

**Figure 5.4: Schematic of the electrical circuit used for graphene characterization and electrobreakdown.**

The electrical measurement starts by recording current-voltage (I-V) characteristics to determine the resistance of the sample. At low voltage the graphene shows linear, ohmic current-voltage characteristic. However if we measure at high bias, similar to CNTs [166, 167], current saturation effect can be observed. The degree of the saturation depends on the charge carrier density, that is, the distance from the charge neutrality point (CNP). Figure 5.5.a shows a set of I-V characteristics measured at CNP (blue curve) and increasing the gate voltage with the step of 5 V until 40 V (red curve). This effect is attributed to the scattering of the charge carriers with the optical phonons of SiO$_2$ and graphene [164, 165, 213] which finally results in the Joule heating of both the graphene and the substrate.

For each sample gate voltage dependent resistance was measured to determine the averaged quality of the graphene sheet. From the gate response we can conclude on the amounts of the defects, the concentration and type of the surface contamination and the mobility of the charge carriers [141]. Figure 5.5.b-c show the resistance of two freshly prepared samples as the function of the gate voltage under ambient condition (black curves). They differ from the gate response of the ideal clean graphene, which can be described in four parameters [214].
Figure 5.5: a) Set of I-V characteristics of graphene under high bias measured at different gate voltages. The blue curve corresponds to CNP and the red curve to $V_{\text{gate}} - V_{\text{CNP}} = 40$ V. Gate voltage dependent resistance of weakly (b) and highly (c) doped graphene (black curves). Placing them into vacuum the contamination can be removed (blue curves), while the electrical heating causes larger inhomogeneity (red curve). d) Relaxation of the surface contamination after electrical self-heating in vacuum. The measurements were performed by me in Budapest, the samples were fabricated in Basel.

- **Location of the charge neutrality point:** If the minimum of the conductance is at $V_{\text{CNP}} \neq 0$ V the graphene is doped. From the shift of the CNP the doped charge carrier density can be estimated which is proportional to the concentration of the doping. The dopants are n type mainly if the Dirac cone is located at negative gate value and the dopants are p type if the Dirac cone can be found at the positive gate voltage side [141, 215].

- **Hysteric behaviour:** Mostly the dipolar adsorbates such as water are responsible for the hysteresis [216, 217]. These adsorbates serve as a charge trap on the surface of the graphene. Another source of the hysteric behaviour could be
the charge traps in the SiO$_2$ layer or at the Si/SiO$_2$ interface [218]. Hence the screening of the dipole molecules is a dynamic process the size of the hysteresis depends on the speed of gate sweep. The faster the gate sweep the larger hysteresis split [214].

- **Asymmetry of the conductivity:** The different slopes at the two sides of the CNP refer to the different mobility of electrons and holes. Partially it can be ascribed to the doping adsorbates. These adsorbates behave as an energy dependent scattering potential, which gives different scattering cross section for negative and positive charge carriers [215, 219]. It causes reduction in the conductance at one side of the Dirac cone, while there is negligible effect on the other side. The asymmetry can be also induced by the doping of the metal electrode. It can result both n type doping (e.g. Ti) and p type doping (e.g. Au) and the charge carrier density is pinned in the vicinity of the electrodes [220, 221].

- **The conductivity at the Dirac point:** As it is discussed in Section 2.3 the conductance does not vanish fully at the Dirac point, there is some remaining resistance. The minimal value of the conductivity is $\sigma_{\text{min}} = 4e^2/h$. However the conductance does not reach this minimal value because charge inhomogeneities, called as charge puddles, are formed in the graphene due to defects, doping, substrate roughness. At the charge neutrality point even though the graphene is neutral in average the charge puddles locally provide achievable charge carriers [141, 142].

The loosely bound adsorbates can be removed from the surface by putting the samples into vacuum [214, 217]. The cleaning process is observable in the gate response as well, the quality of the graphene improves (see Figure 5.5.b-c, blue curves). The CNP shifts towards to zero gate voltage, the size of the hysteresis and the asymmetry decreases considerably and the higher resistance at the CNP indicates the more homogeneous charge distribution. In case of the sample shown in Figure 5.5.c, the graphene can not be cleaned fully due to the strongly bound contaminations. As the samples are exposed to air again nearly the initial gate dependence returns soon. This effect shows that the contaminations prefer given locations to attach on the surface and edges [217].

The desorption can be accelerated significantly by annealing the graphene in vacuum [214]. The heat treatment can be realized by external source or Joule-heating. Current induced cleaning is a commonly used technique to clean the surface of the graphene [222–224]. For the samples which are ideal for nanogap formation
(see Section 5.1.2), the large current density flows only at the constriction, hence a little part of the sample is heated up, the major part of the graphene remains cold. Figure 5.5.b-c show two different cases about the effect of high current. If the sample is cleaned mostly in the vacuum (Figure 5.5.b) then the position of the CNP does not change, but the maximum resistance decreases significantly. It refers to the enhanced charge inhomogeneity. The contaminants do not adsorb but spread from the hot constriction to the cold part of the graphene. This assumption is also supported by the heat treatment of highly doped graphene (Figure 5.5.c). In that case the current annealing causes substantial changes in the gate dependence. Additional peaks appear with large hysteresis which refers to differently doped regions in the graphene. The peak near the zero gate voltage may refer to the clean graphene at the constriction. However the induced inhomogeneity is not stable in time, as soon as the high current is released the contaminations start to relax back, the peaks are shifting and smearing (see Figure 5.5.d). The current induced heating can clean the vicinity of the constriction, but increases the inhomogeneity of the whole graphene. In respect of the nanogap formation the clearness of the constriction is the critical, the large graphene electrodes are less relevant. The resistance changes during the electrobreakdown process may arise from both the non-linear I-V characteristic (Figure 5.5.a) and the distortion of the gate dependence (Figure 5.5.b-c).

Figure 5.6: Conductivity of the graphene sample shown in Figure 5.5.b as the function of the gate voltage (dots). The charge carrier mobility can be extracted from the slope of the fitted lines (green lines). The measurements were performed by me in Budapest, the sample was fabricated in Basel.
Although the geometry of the samples are not ideal to determine charge carrier mobility due to its complex shape and large size it is possible to make a rough estimation. Assuming homogeneous graphene sheet with constant conductivity the proportionality factor between the resistance and the conductivity can be calculated (see Equation 2.20). In case of the generally used sample with 400x800nm constriction size, this factor is around 5.1. Figure 5.6 shows the conductivity of the same sample as introduced in Figure 5.5.b. For better clarity only the reverse sweeps are plotted. According to Equation 2.19 the slope of the fitted line in vicinity of CNT is proportional to the mobility of electrons or holes. The mobility values are listed in Table 5.2. They give a good quantitative information about the cleaning process and confirm the conception of the cleaning. The high vacuum improves the quality of the graphene, while the heating process rather reduces the mobility. It has to be noted that after the heating process the graphene sheet can not be considered homogeneous concluded from the gate voltage dependent resistance measurements.

This calculation gives lower limit to the mobility values. The proportionality factor between the resistivity and the resistance (Equation 2.20) could be higher than the theoretical one if there are holes or defects in the graphene. The larger proportionality value results in higher conductivity and higher mobility. The mobility values typically fall in the range of 100 – 1500cm²/Vs

<table>
<thead>
<tr>
<th>Mobility (cm²/Vs)</th>
<th>Electron</th>
<th>Hole</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambient</td>
<td>207</td>
<td>579</td>
</tr>
<tr>
<td>Vacuum</td>
<td>479</td>
<td>680</td>
</tr>
<tr>
<td>After heating</td>
<td>453</td>
<td>405</td>
</tr>
</tbody>
</table>

Table 5.2: Mobility of the charge carriers at different stages of the cleaning procedure.

The resistances of the freshly made samples at zero gate voltage vary between wide limits because of the different doping levels. The typical resistance values are 5 – 30kΩ. The resistance of the sample is the sum of the resistance of the constriction, the large graphene electrodes and the contact resistance between the graphene and the gold/titanium metal pads. Contribution of the further parts (metal pads, bonding wires, measurement system) are negligible. The contact resistance between the graphene and the metal pads was measured by fabricating samples with four contact pads. According to the four probe resistance measurements, the contact resistance is at least one order of magnitude smaller than the resistance of the graphene.
5.2 High yield formation of nanogaps

5.2.1 The electrical breakdown process

As it is discussed in Chapter 3 the graphene electrobreakdown is performed by voltage pulses with increasing amplitude. The resistance is measured during the pulse ($R_{\text{high}}$) and after each pulse at low bias ($R_{\text{low}}$) as well. Figure 5.7.a shows the evolution of both resistances as the function of pulse amplitude ($V_{\text{high}}$) during the electrobreakdown of a sample with 400x800nm sized rectangular constriction at zero gate voltage. The length of the applied pulses is 10ms. The gate dependent resistance of the same sample before the EB is shown in Figure 5.7.b. The location of the CNP ($V_{\text{CNP}} \approx -40\text{V}$) indicates that the surface of the sample is highly contaminated. When the pulse amplitude reaches the $\approx 1\text{V}$ magnitude the cleaning of the graphene starts and both resistances change significantly. The sudden and substantial increase of $R_{\text{low}}$ mostly corresponds to the shift of the Dirac cone towards the zero gate voltage due to the cleaning process. The almost monotonic increase of the high voltage level resistance ($R_{\text{high}}$) is not only caused by the cleaning but also by the current saturation effect. An abrupt breakdown event can be observed at $V_{\text{high}} = 9.23\text{V}$, the resistances jumps several orders of magnitude and can not be resolved by the measurement setup any more. Before the breakdown a modest precursor effect can be observed in the resistances. However, the more significant cleaning effect makes the feedback controlled breakdown process difficult, because it is hard to distinguish the cleaning effect from the breakdown. Therefore at first we attempted to make nano-sized gap without precise feedback control of the EB using exfoliated, multi-layer graphene.

The nanogap formation was investigated under both ambient and vacuum ($\approx 1 \cdot 10^{-5}\text{mbar}$) condition. Before our study the graphene EB was performed only in air in the literature.

5.2.2 Characterization of the nanogaps

Electrical characterization

Following the EB process we need to characterize the formed nanogap by determining its main parameters, such as width, barrier height or location. The broken graphene can be modelled as a potential barrier with the width of $d$ and height of $\Phi$. If the gap size is only a few nanometers tunneling current can flow through the broken junction (see Section 2.2.3). The current is strongly sensitive to the barrier height and width which allows us to determine these parameters. Figure 5.8.a shows
Figure 5.7: a) Electrobreakdown process of a graphene stripe, $R_{low}$ (black) and $R_{high}$ (red) are measured as the function of pulse amplitude ($V_{high}$) at $V_{gate} = 0$ V. The large resistance changes correspond to the cleaning of the graphene. b) The gate voltage dependence before the EB refers to highly contaminated sample. The measurements were carried out together with Cornelia Nef in Basel.

A typical S-shaped tunneling I-V curve right after the breakdown procedure in vacuum. At low bias, when the applied voltage is much lower than the barrier height there is a linear dependence, the contact shows ohmic behavior and a resistance value can be assigned. The inset shows the linear fit of the tunneling curve between $\pm 0.3$ V, the corresponding resistance is $\approx 190$ GΩ. Figure 5.8.b shows the statistic about the low voltage resistances after the electrobreakdown in vacuum and under ambient condition. Nearly 98% of the samples broken in vacuum show measurable tunneling current which means less than 5 nm gap sizes. Most of the resistances lie in the ideal range of GΩ [121]. The breakdown process is considered as failed if the current is less than 10 pA under $\pm 10$ V bias. In the presence of oxygen the yield drops significantly, only 2 samples out of 11 show tunneling effect. This result suggests that the oxygen concentration plays crucial role in the breakdown process. A detailed investigation of this phenomena is presented in Section 5.3.

Applying higher bias voltage the I-V characteristic deviates from linear behavior and the differential resistance starts to decrease. It is the results of the distortion of the potential barrier from rectangular to trapezoidal (see insets of Figure 5.8.c). This regime is called as direct tunneling and the Simmons model gives a quantitative description (see Section 2.2.3). By fitting the measured current data the main geometry parameters of the gap can be determined. At even higher bias ($V_{bias} > \Phi$) field electron emission takes place and the Simmons model is not valid any more.
Figure 5.8: a) I-V characteristic of a tunneling junction after EB procedure in vacuum. The inset shows the linear fit to the low bias regime, the corresponding resistance is 190 GΩ. b) Histogram of low bias resistances after the EB performed in vacuum (blue) and ambient conditions (orange). c) Fowler-Nordheim plot of a tunneling I-V characteristic. The schematics shows the corresponding energy band diagrams. The transition voltage of this specific trace is 0.32 V. d) Fitting of the Simmons model using the smallest (red line) and largest (blue dashed line) possible junction area to the measured data (black dots). The two extreme fits perfectly cover each other by giving different potential barrier parameters. The fitting limit is ±0.42 V. The measurements were carried out together with Cornelia Nef in Basel [1].

The transition voltage spectroscopy (introduced in Section 2.2.3) was used to get an estimation to the range of the Simmons fit. If we plot the I-V curves on the ln(I/V²) versus 1/V graph (Fowler-Nordheim plot), there is a minimum, called as transition voltage (Vₜ), which corresponds to the transition between the direct tunneling and the electron field emission (see in Figure 5.8.c). During the fittings I chose the validity range of the Simmons model (Equation 2.8) to 1.3 times of the transition voltage.

However the Simmons fitting has to be applied carefully because the fitting is not ambiguous, very similar curves can be obtained using significantly different parameters. The cross-section of the tunneling contact can vary several orders of magnitudes,
the current may flow through a single atom contact \( (A \approx 0.01 \text{nm}^2) \) or through the full width of the broken ribbon \( (A \approx 100 \text{nm}^2) \), that is, 400 nm width with single atom height). However this four orders of magnitudes difference is still much lower than it could be in case of metal contacts due to the 2D structure of the graphene. In order to handle this uncertainty we performed the Simmons fits with these two extreme cross section values and we assumed that these fits provide an upper and a lower limit of the gap size. For the barrier height I didn’t make any constraint. Figure 5.8.d shows the results of two fits, one with continuous red line \( (A=0.01 \text{nm}^2) \) and the other with dashed blue line \( (A=100 \text{nm}^2) \). The corresponding parameters of the tunneling barrier are also presented. As it can be seen, the two curves fully cover each other, while their parameters are quite different. However the large difference in the junction area induces only sub-nm difference in the gap size because the current depends linearly on the cross section but exponentially on the barrier width. The statistical results of the Simmons fit revealed that the gap size ranged from 0.3 nm to 2.2 nm. More detailed analysis about the barrier parameters under different environmental condition is presented in Section 5.3. The yield of the nanogap formation was independent of the geometry parameters of the rectangular constriction, there were no observable differences in respect to the formed nanogaps. However in case of the 600 nm wide constriction the breakdown currents \( (\approx 1.2 \text{mA}) \) were about 1.5 times larger in average than for the 400 nm wide ones \( (\approx 0.8 \text{mA}) \). These values correspond to \( j \approx 5.6 \cdot 10^8 \text{A/cm}^2 \) current density at the moment of the breakdown, which is in good agreement with the literature \cite{175, 225}.

It is possible that the broken graphene can not be modelled as a single potential barrier as it was assumed in the Simmons model. Inside the gap there may be organic residues or carbon island which arise from the sample preparation or the EB procedure. These contaminations can take part in the electrical transport giving false gap size or switching effects \cite{121, 179, 226}. In order to investigate the cleanness of the nanogap current-voltage characteristics were measured on the tunnelling junctions as the function of the gate voltage. The charge islands behave as quantum dots, whose energy level can be tuned by the gate voltage \cite{179}, resulting in gate dependent electrical transport. In Figure 5.9.a the left panel shows the I-V characteristic of a broken sample at zero gate voltage while on the right, the current is shown on color-scale as the function of gate voltage and bias voltage. The tunneling current is independent of the gate voltage which may refer to a clean tunneling junction.

However it is possible that the energy scale of the island is in the order of the thermal fluctuation at room temperature \( (k_B T \approx 26 \text{meV}) \) which smears the effect of the nanoislands. To exclude the false conclusion some samples on Si\(_3\)N\(_4\) substrate were cooled down to 4.2 K to measure gate dependent electrical transport. One of
them is shown in Figure 5.9.b which still did not show any sign for charge island. All these indicate, that the gap did not contain any contaminations.

Figure 5.9: a) Electrical characterization of a nanogap device under ambient conditions. Left panel: single I-V measurement at zero gate voltage. Right panel: Gate voltage dependence of the I-V curves b) Electrical characterization at low temperature (4.4 K) Left panel: single I-V trace at zero gate voltage. Right panel: Gate voltage dependence of the I-V curves. The measurements were performed by me in Budapest, the sample was fabricated in Basel [3].

Raman spectroscopy and scanning techniques

Further information can be gained about the gap and its surroundings using scanning techniques, such as Atomic Force Microscopy (AFM), Scanning Electron Microscopy (SEM) or Raman Spectroscopy. These measurements were performed at
the University of Basel by Dr. Cornelia Nef (SEM and Raman) and Dr. Monica Schönenberger (AFM) during my visit and these results are based on the work of Dr. Cornelia Nef.

Raman spectroscopy is a perfect tool to get an insight to the local quality of the graphene. Among others we can obtain information about the number of layers, concentration of the defects or the structure of the edges while it is fast and non-destructive method [227–229]. Furthermore the Raman spectrum is sensitive to the temperature, which allows us to estimate the local temperature [230]. Due to its versatile usage Raman spectroscopy became one of the most important techniques to characterize graphene structures.

Figure 5.10: a) Temperature dependence of the position of the graphene 2D Raman peak, the graphene D peak (right inset) and the Si peak (left inset) as well as linear fits to the temperature dependence of the peak position. The temperature was varied with a external heater. b) Shift of the graphene 2D peak versus power $P_{el}$ (dots) and corresponding graphene (blue squares) and silicon (purple triangles) temperature due to the self-heating of the graphene nanowire during the EB process. These measurements were performed by Cornelia Nef in Basel [1].

The 2D peak of the graphene, which belongs to the two-phonon scattering, shifts linearly by the temperature. This effect enables to probe the local temperature during the breakdown process. Since the size of the laser spot is $\approx 400\,\text{nm}$ for Raman measurements samples with wider constriction ($2\,\mu\text{m}$) were fabricated. Prior the EB process the setup had to be calibrated by external heating of the graphene. Figure 5.10.a shows the peak position of 2D-band (blue squares) as the function of the external temperature. By fitting a line to the experimental data, the slope
of the shift is $-0.051 \pm 0.002 \text{cm}^{-1}\text{K}^{-1}$, this value is in good agreement with other studies [230, 231]. The shift of the D peak of the graphene (right inset) and the peak of the silicon (left inset) are also presented. These peaks also show a less robust temperature dependence. In the next step the Raman spectra were recorded again, but during the EB process at ambient conditions. The shift of 2D peak was extracted as the function of the power (see Figure 5.10.b colored circles) and using the previous calibration the peak shift could be rescaled to temperature. Finally the temperature-power relationship was obtained as plotted in Figure 5.10.b (blue squares). At the moment of the breakdown, the temperature reaches the 570K. It is lower than reported by other research groups. The observable oxidation of single layer graphene starts to appear at 450°C under the period of 2 hours, but for faster oxidation higher temperature is required [232]. The double or triple layer graphene reacts with oxygen at even higher temperature. However the estimated temperature is an averaged value over the spatial resolution of the laser spot (400nm), the maximum temperature must be higher. During the EB process the heating of the silicon substrate can be also observed, the maximum temperature enhancement is 34K (see Figure 5.10.b purple triangles). This refers considerable heat transport to the substrate.

The Raman spectroscopy is also suitable to investigate the local quality of the graphene [229] after the breakdown process. Figure 5.11.a shows the comparison of Raman spectra measured at the constriction (purple curve and dot) and at the electrodes far from the gap (blue curve and dot). The large 2D peak of the graphene electrode indicates single layer graphene with low disorder density. In contrast at the constriction the presence of amorphous carbon and higher disorder density are indicated by the enhanced D peak and the intensity ratio of G to 2D band. The inset illustrates the intensity map of 2D peak, the contour of the graphene is highlighted by dashed white line.

The broken graphene nanoribbon was also investigated by Scanning Electron Microscope (SEM) and Atomic Force Microscope (AFM) to get higher spatial resolution about its size and location. On the SEM image (see Figure 5.11.b) the location of the gap is visible as darker line along the graphene stripe emphasized by arrows (graphene is light gray, the SiO$_2$ substrate is darker gray). As it was expected, the graphene broke at the rectangular shaped constriction.

More detailed image can be taken about the structure of the gap by AFM. The measurement was performed under ambient condition by Dimension 3100 (Veeco, USA) in combination with PPP-NCHR cantilevers, whose nominal tip curvature is less than 10nm (Nanosensors, Switzerland). The AFM image of a freshly broken sample is shown in Figure 5.11.c. According to the cross-section scanning (turquoise
Figure 5.11: a) Raman spectra recorded on the graphene electrode (blue) and on the burned constriction (purple). The inset shows a map of the integral over the graphene 2D peak, where the white dashed line shows the border of the graphene. b) A SEM image of the graphene constriction after EB (lighter gray graphene, darker gray substrate), the gap in the graphene is visible as a thin line, marked with arrows. c) AFM image and d-e) height profiles across the graphene channel and over the gap. The gap size is estimated to be 4.5 nm wide. These measurements were performed by Cornelia Nef in Basel [1].

lines in Figure 5.11.c and d), the thickness of the graphene is 0.7 nm. It is larger than the theoretical value of 0.35 nm, but it is in good agreement with other AFM measurements [233, 234]. The gap size is measured at multiple places along the slit. At the narrowest position (pink lines in Figure 5.11.c and e) the distance of the two graphene electrodes is 4.5 nm, however this value overestimates the gap size since the tip curvature is in the same order. At another position (blue lines) the width of the gap is larger which indicates that the gap size is not uniform along the slit.
5.3 Mechanism of the breakdown

The exact geometry of the gap and its surrounding essentially influence the operation of the fabricated devices. In respect of application it is crucial to produce the nanogaps with very high yield and uniformity. For this purpose understanding of the breakdown process is essential. The ideal resistance of a nanogap is in the regime of GΩ [121]. In case of higher resistance the tunneling current can not be measured which makes the initial characterization difficult and the gap may exceed 5 nm size. On the other hand tunneling junctions in MΩ regime may be the results of carbon island or residues inside the gap, which give a parallel current flow with the device. Environmental conditions are critical issues in a single-layer material hence every atom is on the surface. In Section 5.2.2 it is discussed that the yield of the nanogap formation in vacuum is very high. In contrast when the breakdown process is performed in ambient condition the yield drops significantly. This finding also suggests that the environment has a strong effect on the EB. The active role of the oxygen was also considered by F. Prins et al., they ascribed the breakdown to burning of the carbon atoms. However the exact mechanism has not been examined in detailed.

In order to study the effect of different environmental conditions I performed a set of measurements with fixed sample geometry, while I monitored the breakdown parameters. For this purpose rectangular shaped constrictions were used with 400 nm width and 800 nm length. The oxygen can not only come from the atmosphere but from the SiO₂ substrate as well. In order to investigate whether the oxygen content of the substrate influences the breakdown process two different insulator layers were used: doped silicon coated by 300 nm thermal oxide (SiO₂) or 80 nm thermal oxide and 140 nm Si₃N₄. The thickness of the dielectric layers were chosen to get the highest optical contrast between the naked substrate and the substrate covered by single layer graphene. In respect of the oxygen source we assume that the surface of graphene at the constriction is clean, the organic residues are removed due to the current induced heating. The melting point of PMMA is 160°C and boiling point is around 200°C, while the graphene starts to burn at 450°C [232].

The breakdown process was examined for both substrates under both ambient and high vacuum (≈ 1 · 10⁻⁷ mbar) condition. In agreement with the results of previous section the yield of the nanogap formation in vacuum was over 90% while in ambient in most cases unresolvable tunneling current was observed. Figure 5.12.a shows the histogram of the tunneling resistance in case of different environments. No substantial differences can be discovered between the two substrates and in case of vacuum most of the samples fall into the ideal resistance regime of few GΩ.
To get more detailed information about the structure of tunneling junctions I fitted the Simmons-model to several I-V characteristics and I plotted the distributions of the gap sizes and barrier heights. I followed the same procedure during the Simmons fittings as it is discussed in Section 5.2.2. The histograms of the fitted gap sizes are shown in Figure 5.12.b. For both substrates the gap sizes are smaller than 4 nm, however the samples on nitride show narrower distribution. The cross section of the tunneling junction has only sub-nm effect on the distributions. The difference between the two substrates is more notable in case of the barrier height (Figure 5.12.c). The fits reveal a wide distribution of barrier heights between 0.4 and 5 eV for SiO$_2$ substrate. The relatively small values compared to the vacuum work function of graphene (4.5 eV) were also observed in previous studies on similar systems [116, 226]. The broad range of barrier heights may point to the role of the SiO$_2$ substrate in the tunneling process. As a sharp contrast, on silicon-nitride narrow distribution is obtained between 3 and 5 eV) which is close to the vacuum work function value. It
suggests that the nitride does not affect the charge transport through the nanogap resulting in more defined tunnel junctions.

The measurements revealed that there is obvious difference between the breakdown under ambient and vacuum conditions regardless of the substrate. The breakdown of graphene and carbon nanotubes were investigated in several studies. In air the electrobreakdown is commonly attributed to oxidation of the carbon atoms both in carbon nanotubes [167, 235] and in graphene [116, 236]. In vacuum much higher breakdown powers were observed both for CNT and graphene, suggesting other failure mechanism. In case of nanotubes among others defect formation, melting, thermally assisted field evaporation of the atoms or failure of the underlying SiO\textsubscript{2} were proposed as the possible breakdown mechanism [237–239]. In graphene mostly only the maximum current density and the temperature right before the breakdown were studied, the exact breakdown mechanism has not been discussed [225, 236, 240–242]. However, they mentioned that the temperature plays crucial role, in vacuum the breakdown temperature of graphene was found above 2000 °C. At reduced oxygen concentration it is also supposed that the oxygen atoms may come from the SiO\textsubscript{2} substrate or from the contaminations [240, 243, 244]. The structural modifications of few-layer graphene under high bias were studied by in situ transmission electron microscope (TEM) measurements [223, 245]. They revealed most contaminations were removed before the breakdown event and when the bias reached a critical value crack formations at the edge of the flakes were observed. These cracks propagated towards the other edge of the samples. In some cases holes were also evolved in the middle of the graphene flakes at the disorders [223]. The mechanism of the graphene breakdown was attributed to sublimation, however they did not carry out systematic measurements to verify the exact mechanism. They also estimated the breakdown temperature to 2000 °C, but it must be the lower limit since the high-energy electron beam can assists the breakdown. The breakdown processes both at ambient condition and in vacuum are supposed to be thermally activated process, in which the temperature plays a significant role. If we examine the reaction rate as the function of the temperature, we could gain important details about breakdown dynamics.

The EB is performed by applying voltage pulses. For all samples the breakdowns happen abruptly, we can not detect any notable precursor right before the resistance jumps regardless of the pulse length. This suggests that most of the carbon atoms leave the contact at the last pulse. By assuming that during all breakdown processes similar number of carbon atoms are involved, the shorter pulse length must imply proportionally faster reaction rate. Since the EB is proposed as thermally activated process the faster reaction rate means higher temperature and hence higher electric power at the moment of the breakdown. In order to study this relationship EB pro-
cedures were performed both in ambient (10³ mbar) and in high vacuum (10⁻⁷ mbar) using varied pulse lengths (5 µs-5 s). This set of measurements were done on both substrates.

Figure 5.13: Breakdown powers (P_{BD}) for all samples as the function of the pulse length in ambient (orange) and in vacuum (blue) conditions. The left panel corresponds to SiO₂, the right panel to Si₃N₄. This pulse time dependent measurements were performed by me in Budapest using the developed pulsing setup introduced in Chapter 3 [3].

Figure 5.13 shows the power values at the moment of the breakdown as the function of the pulse length, for the different environmental conditions. Typically 2-5 points were measured with the same parameters. The most pronounced tendency is the significantly different breakdown power in vacuum (blue dots) and in ambient (orange dots). It suggests two distinct breakdown mechanisms. Furthermore, as it was expected, on average higher power is needed for shorter pulse length independently of the pressure or the substrate. Finally, samples on SiO₂ generally require higher power than samples on Si₃N₄.

At ambient conditions the Joule-heating induced oxidation is assumed. However, in vacuum (10⁻⁷ mbar) the oxygen concentration is too low, there can not be enough molecules to react with carbon atoms. In order to confirm this assumption at first we have to estimated the number of oxygen molecules impact on one atomic sites during the period of one voltage pulse.

According to the kinetic gas theory the current density of oxygen molecules (j_{ox}) can be expressed as
\[ j_{ox} = \frac{1}{4} n_{ox} \bar{v}, \quad (5.1) \]

where \( n_{ox} \) is the molar concentration of the oxygen and \( \bar{v} \) is the average speed of the oxygen molecules. The concentration can be written as \( n_{ox} = \alpha_{ox} \frac{p}{k_BT}, \) where \( \alpha_{ox} = 0.21 \) is the fraction of oxygen in air, \( p \) is the pressure, \( T = 300 \text{K} \) is the room temperature and \( k_B \) is the Boltzmann constant. The average speed of oxygen molecules is \( \bar{v} = \sqrt{\frac{8k_BT}{\pi\mu}}, \) where \( \mu = 5.31 \cdot 10^{-26} \text{kg} \) is the mass of an oxygen molecule.

The area of one atomic site in graphene is \( A \approx 0.0262 \text{nm}^2. \) By substituting all these values into Equation 5.1, the number of oxygen atoms arriving to one atomic site \( (N_{ox}) \) is given by:

\[ N_{ox} \approx 1.5 \cdot 10^7 \cdot \tau \cdot \frac{p}{p_{ambient}}, \quad (5.2) \]

where \( \tau \) is the pulse length in seconds and \( p_{ambient} \) is the atmospheric pressure. Considering the range of the pulse lengths (5s-5\( \mu \)s) and the pressures (10\(^{-7} \text{mbar} \) and 1bar), the number of oxygen molecules hitting an atomic site during a single pulse was varied by 16 orders of magnitude. It should be noted that the thermal time constant of such a large graphene stripe is in the order of few nanoseconds [153, 154], which is much shorter than the applied pulse length. The graphene is heated only when the voltage pulses are applied and the graphene is in steady state during the pulse all along.

To interpret the data in terms of electroburning it is useful to rescale the pulse length to the number of oxygen molecules arriving to an atomic site during a single pulse. This rescaled axis is shown in the top axis of Figure 5.14, while the power values with the same parameters are averaged and their deviations are illustrated by the error bars. At ambient conditions in case of the shortest pulse the number of oxygen molecules is still high enough (\( \approx 10^2 \)) to oxidize the graphene. In contrast under vacuum it can be seen that even at the longest pulse there are not enough oxygen molecules (\( \approx 10^{-2} \)) to react with the carbon atoms. This combination of the pressures and pulse lengths guarantees that the two kind of measurement sets are properly separated in respect of the presence of oxygen, and therefore two fundamentally different types of breakdown procedures are expected. The similar tendency of the breakdown powers in vacuum for both substrates refers to that the same breakdown mechanism takes place regardless of the substrate. This finding contradicts the assumption that the oxygen atoms in the SiO\(_2\) take part in the breakdown process.

To get detailed insight to the distribution of breakdown powers under fixed conditions, larger statistics were collected. These measurements were performed by Maria.
El Abbassi in Basel using 10 ms long pulses for both substrates in vacuum and for SiO$_2$ in ambient. The distributions of the powers for the vacuum measurements are shown in Figure 5.14.b. By fitting Gaussian curves the average values and the standard deviations of the data are calculated and also plotted on the left panel as lighter dots. It can be seen that these points fit to the trend of the pulse length dependent data.

Figure 5.14: a) Average breakdown power as the function of pulse length. On the top axis the number of oxygen molecules hitting to an atomic site during a single pulse is also shown. The error bars represent the standard deviation of the data points. b) Distribution of the breakdown powers for SiO$_2$ and Si$_3$N$_4$ for a pulse length of 10 ms under vacuum. The curves correspond to the Gaussian fit of the distribution. As a reference the mean values and the standard deviations of the fitted Gaussians are represented as lighter symbol on panel (a). Similar measurements were carried out for SiO$_2$ at air. The pulse length dependent measurements were performed by me in Budapest and the larger statistics with 10 ms pulse length were collected by Maria El Abbassi in Basel [3].

We assume that under a certain atmospheric condition the same breakdown mechanism is involved regardless of the substrate. At ambient condition oxidation is supposed, while under vacuum the sublimation is our proposed phenomena. Since we assume that the substrates do not affect the breakdown process, at a given pulse length and pressure the breakdown should happen at the same local temperature of graphene for both SiO$_2$ and Si$_3$N$_4$. To verify our assumption about the dynamics of the breakdown processes we need to rescale the power to the maximal local temperature of the graphene stripe. For this purpose we have to solve the heat equation. For
analytical solution we need to simplify the geometry of the graphene and make some assumptions. The simplified thermal model is illustrated in Figure 5.15.a, where the most relevant dimensions and boundary conditions are also presented. We assume that the large graphene contacts can lead most of the heat to the metal contacts and to the silicon wafer and their temperature do not change a lot. Based on this consideration at the ends of the constriction the temperature is fixed to room temperature \( T_0 = 300 \text{ K} \).

![Figure 5.15](image.png)

**Figure 5.15:** a) The schematic of the device geometry and the boundary condition used in the analytic thermal model. The layer structure of the substrate corresponds to Si\(_3\)N\(_4\). b) The calculated temperature profile along the graphene stripe. The maximum temperature is reached at the middle of the constriction. c) The maximum temperature at the constriction as the function of the electric power for SiO\(_2\) (blue) and Si\(_3\)N\(_4\) (orange).

After the simplification of the geometry the 1D heat equation has the following form:

\[
A_g \kappa_g \frac{d^2T}{dx^2} + p_x - g(T - T_0) = 0, \tag{5.3}
\]

where \( A_g = W \cdot t_g \) is the cross-section of the single layer graphene, \( W \) is the width of the graphene stripe, \( t_g \) is the thickness of single layer graphene, \( \kappa_g \) is the thermal conductivity of the graphene, \( p_x \) is the Joule heating power per unit length and \( g \) is the thermal conductance to the substrate per unit length.

Equation 5.3 has a simple analytic solution, its form along the x-axis is given by

\[
T(x) = T_0 + \frac{p_x}{g} \left( 1 - \frac{\cosh(x/L_H)}{\cosh(L/2L_H)} \right), \tag{5.4}
\]

where \( L_H = \sqrt{\frac{\kappa_g W t_g}{g}} \) is the thermal healing length, \( L \) is the length of the graphene.
The temperature profile along the axis of the graphene stripe is plotted in Figure 5.15.b. We assume that the graphene stripes break at the hottest point, therefore only the maximum temperature value is considered in the further calculations, which can be expressed as

\[ T_{\text{max}} = T_0 + \frac{p_x}{g} \left( 1 - \frac{1}{\cosh(L/2L_H)} \right), \]

(5.5)

To calculate the thermal conductance of the substrates, we have to take into account the layers and their thermal resistance where the heat flows through. The full thermal resistance is the sum of the thermal resistance of each layer. For instance in case of SiO$_2$ coated substrate, 3 layers have to be considered: $\approx 500 \mu$m thick doped Si, 300nm thick SiO$_2$ and the SiO$_2$-graphene interface. However the thermal resistance of doped Si can be neglected.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Value</th>
<th>Unit</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W$</td>
<td>400</td>
<td>nm</td>
<td></td>
</tr>
<tr>
<td>$L$</td>
<td>800</td>
<td>nm</td>
<td></td>
</tr>
<tr>
<td>$t_g$</td>
<td>0.335</td>
<td>nm</td>
<td></td>
</tr>
<tr>
<td>$A_g$</td>
<td>134</td>
<td>nm$^2$</td>
<td></td>
</tr>
<tr>
<td>$\kappa_g$</td>
<td>1000</td>
<td>$\frac{W}{mK}$</td>
<td>[157, 158, 246]</td>
</tr>
<tr>
<td>$\kappa_{oi}$</td>
<td>1.4</td>
<td>$\frac{W}{mK}$</td>
<td></td>
</tr>
<tr>
<td>$\kappa_{ni}$</td>
<td>30</td>
<td>$\frac{W}{mK}$</td>
<td></td>
</tr>
<tr>
<td>$\rho_{goi}$</td>
<td>$1 \cdot 10^{-8}$</td>
<td>$\frac{m^2K}{W}$</td>
<td>[168, 170, 247]</td>
</tr>
<tr>
<td>$t_{oi}$</td>
<td>300 or 80</td>
<td>nm</td>
<td></td>
</tr>
<tr>
<td>$t_{ni}$</td>
<td>140</td>
<td>nm</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.3: List of the parameters used in the thermal model

The thermal conductance through the 300nm thick SiO$_2$ substrate ($g_{oi}$) is

\[ g_{oi} = \frac{1}{\frac{t_{oi}}{\kappa_{oi}W} + \frac{\rho_{goi}}{W}}, \]

(5.6)

where $t_{oi}$ and $\kappa_{oi}$ is the thickness and the thermal conductivity of the silicon-dioxide respectively, $\rho_{goi}$ is the thermal boundary resistivity between the oxide and the graphene.

Similarly, the thermal conductance through Si$_3$N$_4$ substrate ($g_{ni}$) is
\[ g_{ni} = \frac{1}{\frac{t_{ox}}{\kappa_{ni} W} + \frac{t_{ni} \kappa_{ox}}{W} + \rho_{gni}} \]  
(5.7)

where \( t_{ox} \) corresponds to the 80 nm thick oxide layer and \( t_{ni} \) to the 140 nm nitride thickness, \( \kappa_{ni} \) is the thermal conductivity of the nitride and \( \rho_{gni} \) is the thermal boundary resistivity between the graphene constriction and the nitride substrate.

Table 5.3 contains the values of parameters used in the thermal model. All the parameters were taken from the literature except the thermal boundary resistivity between Si\(_3\)N\(_4\) and graphene, for which we are not aware of any prior measurement. Based on the assumption that the breakdown of the graphene happens at same temperature regardless of the substrate, we used \( \rho_{gni} \) as a fitting parameter to obtain the least squares deviation between the breakdown temperatures on both substrates at the various pulse lengths and pressures. This method gives the value of \( \rho_{gni} = 4.8 \cdot 10^{-7} \text{Km}^2/\text{W} \) to the thermal boundary resistance between nitride and graphene. This value is one order of magnitude larger than for SiO\(_2\) which indicates a weak van der Waals interaction between the graphene and Si\(_3\)N\(_4\) substrate.

According to Equation 5.5, there is a linear relation between the maximum temperature (\( T_{max} \)) and the power. By substituting the value of parameters 294 K/mW and 409 K/mW were obtained to the slopes for the SiO\(_2\) and Si\(_3\)N\(_4\) substrate respectively, as it is seen in Figure 5.15.c.

As it was mentioned, both the burning and the sublimation, the two preferred phenomena for the breakdown, are thermally activated processes. According to the Arrhenius-law the number of reaction (\( N \)) per unit area and unit time can be expressed as

\[ \frac{N}{A \cdot t} = C \cdot e^{\frac{-E_a}{k_B T}}, \]  
(5.8)

where \( E_a \) is the activation energy and \( C \) is a pre-exponential parameter. We assume that similar number of carbon atoms have to leave the sample to break the constriction for all pulse lengths, so the \( N/A \) is assumed to be the same for any pulse lengths. By rearranging and taking the logarithm of Equation 5.8, we get that the logarithm of the time is proportional to the inverse of the calculated temperature and the slope yields the activation energy,

\[ \log_{10}(\tau) = \log_{10}\left(\frac{N}{C \cdot A}\right) + \frac{E_a}{k_B} \cdot \left(\frac{1}{T}\right). \]  
(5.9)

Figure 5.16 shows the Arrhenius plot of measured data for both substrates on \( \log_{10}(\tau) - 1/T \) graph. The right axis presents the corresponding temperature at
the grid lines. The common linear fits are also plotted. In case of the vacuum measurements the points show clear linear trend, they are close to each other and to the fitted line as well. At ambient conditions the same behavior can be observed, but with larger scattering and significantly different slope. The slopes of the fitted lines revealed $10.4 \pm 2.4\text{eV}$ and $1.38 \pm 0.28\text{eV}$ to the activation energies under vacuum and at ambient respectively. As a comparison, the activation energy of sublimation in graphene lattice in the presence of disorders is $\approx 7\text{eV}$ [223, 248, 249] while for oxidation values between $1$ and $2\text{eV}$ [250] were reported by other research groups.

The top axis shows the number of oxygen molecules arriving on a single atomic site during a single voltage pulse. The activation energy of the breakdown mechanisms can be determined from the slope of the fitted lines. The thermal model was calculated together with Maria El Abbassi [3].

The errors of the activation energies include only the uncertainty of the fit due to the scattering of the measurement points, the error caused by the uncertainty of the thermal parameters are not involved. Most of the parameters is well known, but the heat conductivity of the graphene ($\kappa_g$) and the thermal boundary resistances ($\rho_{gox}$) vary a lot in the different studies. To investigate the sensitivity of the results to these thermal parameters, the model was solved for different sets of parameters. The $\rho_{gmi}$ was always a fitting parameter. According to the calculations, the activation energies are not very sensitive to $\rho_{gox}$. If it is increased or decreased by a factor of 5, than the activation energies change only few percentages. In contrast the variation

---

Figure 5.16: Arrhenius-plot of $1/T$ versus $\log(\tau)$ for SiO$_2$ (blue) and Si$_3$N$_4$ (orange). The top axis shows the number of oxygen molecules arriving on a single atomic site during a single voltage pulse. The activation energy of the breakdown mechanisms can be determined from the slope of the fitted lines. The thermal model was calculated together with Maria El Abbassi [3].
of heat conductivity by the factor of 2 induces 20-30% change in the activation energies. Table 5.4 contains the calculated energy values to the different combinations of thermal parameters and environments.

Although the activation energies have large error due to the sample to sample deviation, uncertainty of the thermal conductivity and the simplification of the thermal model, the two different regimes can be still clearly distinguished by the factor of 7-8 difference between the two activation energies. It justifies that different mechanisms take place in ambient and in vacuum condition. The fitted activation energies are close to the literature values, the sublimation and oxidation are indeed feasible explanation for the involved processes.

$$\kappa_g, \rho_{g ox} \quad 2\kappa_g, \rho_{g ox} \quad 0.5\kappa_g, \rho_{g ox} \quad \kappa_g, 5\rho_{g ox} \quad \kappa_g, 0.2\rho_{g ox}$$

<table>
<thead>
<tr>
<th>$\kappa_g$ (WK$^{-1}$m$^{-1}$)</th>
<th>1000</th>
<th>2000</th>
<th>500</th>
<th>1000</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{g ox}$ (1·10$^{-8}$ m$^2$K/W)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>0.2</td>
</tr>
<tr>
<td>$E_a$ (eV) ambient</td>
<td>1.38</td>
<td>1.15</td>
<td>1.63</td>
<td>1.44</td>
<td>1.37</td>
</tr>
<tr>
<td>$E_a$ (eV) vacuum</td>
<td>10.4</td>
<td>7.52</td>
<td>13.3</td>
<td>11.1</td>
<td>10.3</td>
</tr>
</tbody>
</table>

Table 5.4: List of the calculated activation energies using different thermal conductivity ($\kappa_g$) and boundary resistivity ($\rho_{g ox}$) values for ambient and vacuum conditions.

It has to be mentioned that the temperature distribution was also calculated by finite element simulation using the exact device geometry. The 2D heat equation was solved using the same model for the heat transfer towards the substrate. Furthermore I applied the same procedure to examine the sensitivity of the activation energies to the uncertainty of the graphene heat conductivity. The simulation revealed that the temperature at the constriction is overestimated by the analytic model. The maximum temperature-power relationship is still linear (see Figure 5.15.c), but its slope is lower by 27%.

The main inaccuracy of the simplified analytic model is the assumption that at the end of the graphene stripe is close to room temperature. Actually, substantial temperature increment is still presented at the ends of the rectangular constriction, but the temperature profile decreases strongly moving away from the center. The lower temperature results in lower activation energies. The modified values are listed in Table 5.5 calculated by different graphene heat conductivity values. Under ambient conditions the absolute values of $E_a$ have not changed significantly, they are still between 1 – 2eV, but closer to 1eV. However in vacuum the activation energy decreased by more than 2eV and get closer to 7eV. This value is more reasonable in case of sublimation. Furthermore the activation energy is less sensitive to the heat conductivity.
conductivity, the difference between the lowest and highest value is about 2 eV. The finite element simulation further verified our proposal that in ambient electroburning takes place, while in vacuum the sublimation of the carbon atoms breaks the junction.

<table>
<thead>
<tr>
<th>( \kappa_g ) (WK(^{-1})m(^{-1}))</th>
<th>( \kappa_g ), ( \rho_{gox} )</th>
<th>( 2\kappa_g ), ( \rho_{gox} )</th>
<th>( 0.5\kappa_g ), ( \rho_{gox} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>2000</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td>( \rho_{gox} ) (1·10(^{-8}) m(^2)K/W)</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( E_a ) (eV) ambient</td>
<td>1.19±0.23</td>
<td>1.10±0.20</td>
<td>1.27±0.25</td>
</tr>
<tr>
<td>( E_a ) (eV) vacuum</td>
<td>8.1±1.8</td>
<td>6.9±1.5</td>
<td>9.0±2.0</td>
</tr>
</tbody>
</table>

Table 5.5: The corrected activation energy values considering the exact device geometry in finite element simulation.

5.4 Controlled electrobreakdown

In Section 5.2.1 it is discussed that the cleaning of the graphene surface can cause more dominant resistance increase than the precursor of the breakdown. Since it is hard to distinguish the two effects, it is also hard to introduce a proper feedback control, which would be sufficient to break the junction in small steps, instead of the single step breakdown reported in the previous parts of this chapter. In the following I describe my work on the development of a feedback-controlled breakdown protocol. This development was done recently, after the publication of the results presented in the previous parts of this chapter. There are two possibilities to reduce the effect of cleaning during the EB process. We can perform precleaning treatment under vacuum by applying higher bias on the sample which do not cause breakdown yet. However there is always a risk to damage the device even if the bias is chosen carefully. Furthermore the graphene can not be cleaned fully because right after the heating treatment the surface contaminants start to relax back or at ambient conditions the adsorbates can attach again to the surface. Another possibility is to modify the sample design such a way that cleaning induce smaller resistance changes. During the voltage ramp at first the constriction is cleaned due to the high current density. Although it begins at much lower bias than the breakdown voltage, but as the bias is increased to higher value further part of the graphene is heated and cleaned. The cleaning of the large graphene electrodes can be confused by the breakdown event. By decreasing the resistance of the graphene electrodes compare to the constriction
resistance the temperature profile around the constriction will be narrower and it reduces the cleaning effect.

Figure 5.17.a shows EB process of a pre-cleaned sample by applying 6V bias voltage in vacuum. The slight variation of the resistance curves indicate that most of the contamination is removed, but there are still some kinks. The Joule-heating caused resistance growth at high bias can be clearly seen. The corresponding current ($I_{\text{high}}$) is shown in Figure 5.17. b, the last 150mV is plotted in the inset. It shows nearly monotonically increasing behavior, but as the precursor of the breakdown the current drops off in the last few tens mV. To control the electrobreakdown in more reliable way I examined which parameter is the most sensitive to the breakdown event. To reduce the effect of cleaning, current saturation or any long-term resistance change it is a common method to monitor the change of the resistance or current only in the last few hundreds millivolts. The relative change of $R_{\text{low}}$, $R_{\text{high}}$, and $I_{\text{high}}$ in the last 100mV bias window is presented in Figure 5.17.c as the function of the pulse amplitude. In case of $R_{\text{low}}$ there is no observable precursor before the breakdown. It could be the consequence of the higher noise level of $R_{\text{low}}$. In contrast, $\Delta R_{\text{high}}$ (middle panel) shows less variation, but it has a constant offset due to the Joule-heating. A sudden increment can be detected right before the breakdown, however despite the pre-cleaning procedure, it is still in the range of the fluctuations. Finally the current change is presented in the third panel, which shows the most robust precursor of the breakdown. It is closely related to $\Delta R_{\text{high}}$, but it scales down with the ratio of the examined bias window (100mV) and the actual bias. At low bias, where the cleaning is the dominant effect, larger resistance change is needed to decrease the current by $1 - 2\%$ than in larger voltage regime where the breakdown occurs. In conclusion the high level current is the best parameter to monitor if we would like to actively control the breakdown process.

Besides the feedback method, the sample geometry was also optimized to reduce the effect of cleaning. By using bow-tie geometry other research groups could control the breaking process [175, 251]. It could be related to the more localized heating, the temperature profile is narrower and smaller graphene part is involved in the cleaning before the breakdown. The samples with bow-tie constriction were made with different narrowest cross-section from 100nm to 350nm with 50nm steps (see Figure 5.3).

Figure 5.18.a shows the current during the controlled breakdown of a bow-tie sample with minimal cross-section of 250nm under ambient condition. The first cycle is highlighted by red. The breakdown was performed by 100$\mu$s long voltage pulses. In the first cycle the feedback event occurred when the current dropped by 2% in the last 100mV. As the EB progressed higher current change value had to be
set, otherwise the resistance did not increased from cycle to cycle. As the graphene connection gets narrow both the threshold voltage and the current increase. Figure 5.18.b shows the power of pulses \( P = I_{\text{high}} \cdot V_{\text{high}} \) during the same EB process. The power also decreases in every cycle.

Before the EB, there was no cleaning process. Although the sign of the cleaning can be observed (kinks in the red curves), but it is below the feedback limit. The electrobreakdown processes were controllable independently of the minimal width. The new sample geometry allow us to control the breakdown process without any hazardous cleaning pre-treatment.
Figure 5.18: The recorded a) current and b) power during controlled electrobreakdown of a bow-tie shaped constriction with minimal width of 300 nm under ambient condition. Both the current and the power drop off when the breakdown process starts. c) Evolution of the low bias resistance during the breakdown cycles. d) The critical power of the first bias ramp cycle as the function of the minimum cross section of the bow-tie constriction (dots). The lines correspond to the simulated isotherms on the power-width graph. The measurements were performed by me in Budapest, the graphene was grown in Basel and the sample was fabricated by me in Budapest at MTA EK MFA.

As we saw in Section 5.3, the main parameter of the breakdown is the temperature and thus the applied power. To verify this phenomena for the bow-tie samples as well I collected the critical power of the first cycles during the EBs. All measurements were performed under ambient conditions and 3-5 samples were broken at each constriction widths. The geometry of the graphene is well defined only before the first feedback
event, thus the further cycles are not investigated in this regard. The pulselength was 100 µs for all cases. Figure 5.18.c shows the measured power (black dots) as the function of the constriction width. A clearly visible trend is observed, the wider samples, the higher power is needed to induce the breakdown process. In order to interpret this result I have calculated the temperature by finite element simulation using the exact geometry. The colored lines in the same figure show the isotherms. The experimental data fit well in the range of 950 ± 100 ºC. This finding confirms that the graphene breaks around at fixed temperature if the pulse length is also fixed.

The controlled electrobreakdown allows us to increase the resistance of the graphene junction from step to step. However as the resistance raises large current fluctuation and jumps appear and the process becomes less controllable (see in Figure 5.19.a). This behavior could be caused by the fluctuation of carbon atoms at the nanojunction region or the contamination of the nanogap. The current change ratio has to be set to high value (e.g. > 80 %) or the feedback control has to be switched off to increase the resistance further. Finally the formed gaps typically have lower resistance than we get by the uncontrolled breakdown. Figure 5.19.b shows a tunnel I-V characteristic whose low bias resistance is ≈ 8.9 MΩ and the distribution of tunnel resistance for larger statistic is shown in Figure 5.19.c, they are in the range of 1 MΩ - 1 GΩ. To determine the cleanliness of the nanogaps, gate dependent I(V) characteristic measurements were performed at room temperature. However the contacts were not stable during the measurements, there were some structure in the gate responses, but they were not reproducible. For more accurate investigation low temperature measurement is needed to suppress the thermal instability.

5.5 Conclusions

In conclusion our study was the first to use CVD graphene to establish sub 5 nm nanogaps. The application of CVD graphene has enabled the wafer-scale production of graphene nanogaps. Using electrical transport measurements and scanning techniques we characterized and localized the nanogaps and we obtained 0.3 - 3 nm to the gap sizes. According to the statistical analysis of ≈ 100 devices we could achieve 98 % of nanogap formation. The low temperature measurements verified that the gap did not contain any contaminations or carbon islands.

The systematic study of the breakdown power as the function of pulse length and environmental conditions revealed the details of the breakdown processes under ambient and vacuum conditions. By applying appropriate combination of pulse lengths and pressures two fundamentally different cases were investigated in respect of the presence of oxygen. By setting up a thermal model we rescaled the power to
temperature and determined the activation energy of the breakdown process both in ambient and in vacuum conditions. The significantly different values are consistent with electroburning under ambient conditions and sublimation under high vacuum.

Owing to the optimization of both the feedback control protocol and the device geometry I managed to increase the resistance of the graphene constrictions gradually under ambient. The scaling of the electric power with the constriction width confirmed that the electroburning starts at a well defined temperature regime using fixed pulse length. The resistance of the tunnel junction are below the optimal value, a further optimization is needed in the final stage of the breakdown process.
In this chapter I present my results about the SiO$_x$ based resistive switching. In the first part I introduce the fabrication and electrical characterization steps of sub-10 nm sized resistive switches inside a graphene nanogap region. Afterwards I study the internal timescales of the system, which govern its operation. At first I introduce the phenomenon of dead time, which plays a crucial role in the response to various driving signals. Next, I demonstrate the voltage dependent set and reset times. Finally, I investigate the switching properties of the few nanometer sized region and compare them to larger systems already published in the literature. These measurements were performed by myself in Budapest. Most of the graphene nanowire samples were prepared by Maria El Abbassi in Basel, except for one batch of samples used for the measurement in right panel of Figure 6.6.b that were prepared by Botond Sánta and Miklós Csontos in Budapest using the nanofabrication facilities of MTA EK MFA.

6.1 The formation process of SiO$_x$ resistive switches

After the electrobreakdown of the graphene nanostripe the broken ends can be used as the electrodes of a few nanometer sized device. The SiO$_x$ under the graphene serves as a straightforward possibility to form a resistive switching device in the nanogap region. The confinement of the active region between the nanometer spacing electrodes ensures the small size of the device. Figure 6.1.a-c illustrates the fabrication steps of a specific nanometer sized resistive switch (left panels) and the corresponding electric measurements (right panels). For all the measurements presented in this section rectangular shaped constrictions were used with 800 nm length.
and 400 nm width patterned by electron beam lithography and argon-oxygen plasma etching as shown in the schematics. The right panel of Figure 6.1.a shows the evolution of the low bias resistance ($R_{\text{low}}$) and high bias resistance ($R_{\text{high}}$) during the EB process by applying gradually increasing 500 µs long voltage pulses ($V_{\text{high}}$). The electrical breakdown occurs at 9.8 V pulse height where both resistances increase suddenly (arrow) above the resolution limit of our measurement setup. After the breakdown event no more voltage pulses were applied.

To estimate the gap size of the tunnel junction, I-V measurements were performed and the Simmons-model was fitted to the I(V) curves (see top inset of Figure 6.1.b). The validity regime of the Simmons model were determined by transition voltage spectroscopy (see bottom inset of Figure 6.1.b). During the curve fitting I used the same procedure as presented in Section 5.2.2. The fitting result reveals a gap size of 2.0 ± 0.3 nm and barrier height of 0.6 ± 0.2 eV for this representative I(V) curve. The small value of barrier height compared to the vacuum work function (4.5 eV) refers to the role of the SiO$_x$ substrate in the tunneling process.

After the low bias measurements further I-V measurements were performed with gradually increasing voltage amplitude until 8.75 V (see in the main panel of Figure 6.1.b). The current-voltage characteristics still show S shaped behavior referring to pristine tunneling contact. When the bias amplitude reached 9 V, at first large current fluctuation was observed at high voltage regime and finally at 4.5 V during the backward voltage sweep the device transformed to low resistance state (Figure 6.1.c). The current jump indicates the electroformation of the resistive switch, that is, a conductive channel formed between the graphene electrode in the insulator layer. The switching behavior appears at the negative polarity of the I(V) curve, however, it does not show stable characteristic yet. For all switching traces, presented in this section, the same convention was applied for the color. The blue/red parts refer to the ON/OFF state with the arbitrary threshold of 150 kΩ. Considering that the formation of conducting pathway is assumed to be an electric field driven process [71], the active volume can be confined into a similar size as the nanogap. The small electroformation voltage (9 V) compared to the the common values of 20 – 30 V in other larger SiO$_x$ switches [69, 252] also refers to the small filament size. Formerly it was found that the electroforming voltage increases linearly with the gap width, a constant electric field is needed for the soft breakdown [69].

The breakdown events occur at similar or even higher voltage as the electroforming. It could result in instantaneous electroforming right after the nanogap formation. However, for all samples the SiO$_x$ remained in non-switchable, pristine state after the gap formation. If the switching site had already been formed due to the breakdown voltage, during the subsequent tunneling I-V measurements the
Figure 6.1: Fabrication steps of nanometer sized SiO$_x$ based resistive switches. The schematics on the left illustrate the process, the panels on the right show the corresponding electrical measurements. (a) Resistance during the electrobreakdown process at high bias (red) and low bias (black) as the function of the pulse amplitude. A sudden breakdown occurs at 9.8 V. (b) Electrical characterization of the tunnel junction after gap formation. By fitting the low bias trace to the Simmons model a 2.0±0.3 nm gap size was obtained (top inset). The minimum of the ln(I/V$^2$) vs 1/V plot, $V_T$, defines the voltage interval, where the Simmons fitting is applied (bottom inset). The junction exhibits a S shaped tunneling I(V) curve up to $V_{max} = 8.75$ V (main panel). (c) At a threshold amplitude of 9 V large current fluctuations appear signaling the electroforming process [4].
device would have set to ON state at $3 - 4\, \text{V}$. Instead, S-shape characteristic can be measured until $7 - 9\, \text{V}$. The absence of electroformation can be attributed to its much slower timescale. The characteristic time of the breakdown is less than $500\, \mu\text{s}$ while the duration of the I-V measurements during the electroformation is in the range of few $10\, \text{s}$. In case of other applications of nanogap, such as molecular electronic, the electroformation of SiO$_x$ should be avoided, since the switching effect can mimic the electrical phenomena of the molecules [253]. Owing to the difference in timescales applying short pulses during the breakdown offers a possibility to keep the pristine state of the SiO$_x$ substrate. The low electroforming voltage can be further decreased if the SiO$_x$ layer is exposed to argon plasma treatment [254] or annealing [252].

As a control measurement 50 similar devices were tested fabricated on amorphous silicon nitride (Si$_3$N$_4$) substrate. For these samples larger work function values ($3 - 5\, \text{eV}$) were obtained during the nanogap characterization and switching behavior was not observed. This is in contrast to the devices on SiO$_x$, where all of them show switching phenomena. These findings confirm that in our devices the switching occurs in the SiO$_x$ layer and not due to the intrinsic switching of graphene via atomic movement of carbon atoms [225, 226, 241].

After the electroformation and several I-V measurements the switching characteristic was stabilized and reversible filament formation and disruption could be observed (see in Figure 6.2.a top panel). Starting from the low resistance ON state at zero bias, the conductance of the junction drops abruptly at $V_{\text{reset}} = 5.5\, \text{V}$ switching the device to its high resistance OFF state. During the subsequent reverse voltage sweep, the current increases suddenly at $V_{\text{set}} = 4.4\, \text{V}$ and the device switches back to the ON state. Due to the unipolar nature of the switching the same characteristic behavior can be seen at opposite voltage polarity. Consequently, the device is always set to its ON state at zero bias. The 6.2.a bottom panel shows the distribution of set (blue) and reset (red) voltage based on 60 subsequent I(V) curves of the same device and recorded by the same parameters. Both the set and reset voltages have a sharp Gaussian distribution ($V_{\text{set}} = 3.1 \pm 0.4\, \text{V}$, $V_{\text{reset}} = 5.7 \pm 0.3\, \text{V}$) and no significant difference can be detected between the two voltage polarities. The switching voltages were defined by the value, where the device resistance crosses $150\, \text{k}\Omega$, typically during an abrupt transition.

In accordance with previous studies [72, 255] the electroforming and the resistive switching cannot be induced at ambient conditions. Figure 6.2.b shows a set of I(V) curves on logarithmic current scale at different pressure starting from ON state. As the pressure increases, the reset voltage shifts to lower value and finally under ambient condition (black curve) the sample cannot be set to ON state any more. The switching capability can be recovered after reducing the pressure to the initial
Figure 6.2: a) After a few voltage sweeps reproducible unipolar switching characteristics evolve (top curve). The red/blue colored parts of the traces correspond to resistances higher/lower than the predefined threshold of $V/I = 150 \text{k}\Omega$. The histograms of the set and reset voltage show well-defined, stabilized distribution (bottom histograms). b) Pressure dependence of the reset operation on logarithmic current scale. Under ambient condition (black curve) switching effect can not be induced.

value ($< 10^{-5} \text{ mbar}$).

Summarizing the characterization measurements, all of the basic properties of the observed resistive switching, such as switching voltages and currents, were found to be consistent with previous studies on SiO$_x$ switching devices. The small electroforming voltage is attributed to that the active region is formed in a nanometer scale gap. It was shown that the electroforming voltage scales with the gap size; that is, the electric field is the relevant parameter of the electroforming process. Hence, it can supposed that the active volume of the device is confined to the nanometer scale gap region and has the similar size. My measurements showed that the switching properties of SiO$_x$ based memristors are maintained even at extreme small ($\approx 5 \text{ nm}$) size.

6.2 Real time response of SiO$_x$ resistive switches

As it is discussed in Section 2.1.3, the SiO$_x$ based memristors exhibit reproducible switching between ON and OFF zero bias states. This observation is inconsistent with the unipolar $I(V)$ characteristic, shown in Figure 6.2.a. A reset pulse or triangular signal could only switch OFF the device temporarily. The device should
always switch back to ON state when the bias is swept to zero and the voltage passes the set region. The phenomena was observed by two different groups by applying short pulses, however its dynamics were poorly examined, they solely monitored the final states after applying voltage pulses. Nevertheless, this effect would be a great technological relevance, namely, if we intend to operate a device, whose characteristic is shown in Figure 6.2.a, we should bias it. In this specific case, we should apply a constant 4.5 V, where both states are available. However, at this working point the device shows large current fluctuation and the states are less stable. Furthermore, the constant biasing would result in high energy consumption as well.

To understand the exact dynamics of the resistive switching, we need to know the real-time response of the SiO$_x$ memristive system to different biasing conditions. However this kind of study has not been reported yet. In the followings I present my time-resolved measurements.

### 6.2.1 Dead time rule

It was found that the SiO$_x$ memristor show clear unipolar behavior only if the voltage sweep rate is slow (0.5 Hz, 16 V/s), demonstrated in top curve of Figure 6.3.a. If we increase the speed of the I(V) measurement to 2 Hz (64 V/s) a striking phenomena can be observed, the characteristic resembles to bipolar operation (middle curve in Figure 6.3.a). Starting from ON state the device switches off at $\approx +7$ V but during the reverse voltage sweep it does not switch back to ON state in the set voltage region. The device stays in OFF state at zero bias. At the negative polarity during the forward voltage sweep it still remains in OFF state and finally during the reverse voltage sweep it switches back to the ON state. At an even higher frequency (50 Hz, 1600 V/s bottom curve in Figure 2a) the device switches off in the first quarter of the triangular signal and it does not switch back to the ON state along the rest of the driving cycle.

In order to gain a deeper insight into the nature of this effect, multiple period of triangular voltage signal was applied with the frequency of 10 Hz. The upper panel of Figure 6.3.b shows the driving signal (black curve) and the simultaneously measured current using the same convention for the color. The bottom panel shows the corresponding conductance. Initially the device is in ON state and switches OFF when the voltage reaches the reset region. Afterwards the device does not switch back to the ON state in the next four periods, only in the fifth period, about 500 ms after the reset event. Similar behavior is observed during the following driving periods. It has to be noted, that the periodic small red current peaks at each maximum of the voltage signal correspond to the highly nonlinear behavior of I(V) in the OFF state,
Figure 6.3: (a) Representative $I(V)$ traces measured at different driving frequencies. (b) Demonstration of the dead time by applying multiple periods of a triangular voltage signal with frequency of 10 Hz (black curve, voltage scale not shown). The current is measured simultaneously (red/blue curve). The bottom panel shows the corresponding conductance. (c) Illustration of the dead time rule and the time scales involved in the operation cycle $[4]$. 

The observed phenomena can be described by a simple operation rule: once the device is switched OFF, it is blocked in the OFF state for the period of the dead time, even if the driving signal level would be sufficient for initiating a set transition. Once the dead time has passed, the device can be switched ON again at the first appropriate set voltage level. A similar effect does not appear in the opposite switching direction: after the set process the device can be switched OFF without any dead time as illustrated by the flowchart of Figure 6.3.c. This sketch also shows the other two relevant time scales in SiO$_x$ switches, the set and reset time, which is discussed later. In order to study the process behind the dead time it is essential to measure its length under different conditions. However applying triangular driving signal is not suitable for measuring precisely the length of the dead time. Furthermore the constantly alternating voltage may have an effect on the dead time. For this reason in the following I am investigating the device operation by pulsed measurements.
Figure 6.4.a shows a sequence of set-reset-set voltage pulses with the amplitude of 4.5 V, 9 V and 4.5 V respectively (black curves) and the measured current (red/blue curves). The pulse length was 10 ms with rise/fall time of 1 ms at the edges, which are much shorter than the dead time. Between the pulses the bias voltage is zero. The first two pulses initialize the device to prepare identical OFF states for statistical analysis. The first set pulse switches ON the contact and the subsequent high amplitude pulse resets the device to the OFF state. The beginning of the dead time is considered from this event (see arrows). The third pulse is applied with a varying delay time $\tau_{\text{delay}}$ with respect to the reset pulse. If the delay time is shorter than the dead time (top curve), the device stays in the OFF state, because the set pulse can not induce set transition yet. However, if the delay time is longer than the dead time (bottom curve) the set pulse can set the device to ON state.

Figure 6.4.c shows the probability of the set process at different delay times. Each point was obtained by the average of 20 subsequent pulse sequences at fixed delay time. To get a distribution, the delay time was varied in 25–50 ms steps. If the delay time is much shorter than the dead time ($\tau_{\text{delay}} \ll \tau_{\text{dead}}$) the device never switches ON, the probability is zero, whereas if $\tau_{\text{delay}} \gg \tau_{\text{dead}}$ the device turns to low resistance state with a probability of one. Between the two extreme cases the probability changes between 0 and 1 according to the statistical variation of the dead time. Altogether, this function represents the cumulative probability distribution function of the dead time ($F(\tau_{\text{dead}})$). The fitted Gaussian distribution function is plotted by green line. The corresponding $\rho(\tau_{\text{dead}})$ Gaussian probability density function (the derivative of $F(\tau_{\text{dead}})$) is shown by the green line in Figure 6.4.d. It reveals $
abla_{\text{dead}} = 120 \pm 31$ ms for the dead time, where the error corresponds to the standard deviation of the Gaussian. Note that all probability density functions, shown in Figure 6.4, are normalized to unit amplitude for clarity.

There is an alternative method to determine the length of the dead time, as illustrated in Figure 6.4.b. Using the same initialization pulses the device is erased to OFF state, but afterwards the the bias voltage is kept in the set region at the constant value of 4.5 V (black curve). The simultaneously measured current (red/blue curve) shows that after a certain period of time the current jumps abruptly and the contact switches back to ON state. This pulsing scheme enables us to deduce the value of the dead time directly and clarify whether the applied voltage has an effect to $\tau_{\text{dead}}$. The orange histogram in Figure 6.4.c shows the distribution of the measured dead times based on 75 traces. The corresponding fitted Gaussian probability density function is plotted by the orange line indicating $\tau_{\text{dead}} = 216 \pm 77$ ms.

The two different pulsing schemes revealed similar value for the dead time. Unlike the set and reset time (see later), the dead time does not depend on the driving
Figure 6.4: Determining the length of the dead time by two different biasing conditions. In both cases the device is initialized by a set and subsequent reset transition. Afterwards, a) until the next set voltage pulse the bias is set to zero voltage (unbiased) or b) set voltage is constantly applied (biased). c) The probability of the set transition as the function of the delay time (black dots) using the unbiased pulsing scheme. The fitted Gaussian probability distribution is plotted by the green line. d) Distribution of the dead time measured by different biasing conditions. The green line is the density function of the unbiased dead time deduced from panel c. The orange histogram is the distribution of 65 independently measured dead times using the biased pulsing scheme, the orange line is the corresponding fitted Gaussian [4].

condition, the set process is blocked for the same period of time both in the unbiased case (panel a) and also when the device was continuously driven by the set voltage (panel b). Its length can not be controlled by biasing. The slight difference between the two dead times can be explained by an aging effect. As we write and erase the device several times, the elongation of the dead time can be observed. According to my experience it is more significant if the set voltage is applied over longer periods of time. Figure 6.5.a shows the evolution of the dead time measured by the biased pulsing scheme, shown in Figure 6.4.b. The linear fit (red line) shows a clear tendency, in average the dead time increases by 1.2 ms in every cycle. It may be a permanent effect of the long set voltage signal in surrounding of the filament. When the aging effect was minimized by shortening the length of the set pulses, the difference between the biased and unbiased dead time was even smaller. Based on the room temperature investigation of 29 independent devices on 6 chips, the device to
device variation of the dead time spans an order of magnitude ranging from a few hundreds of milliseconds to a few seconds.

![Graph](image)

**Figure 6.5:** *(a) Demonstration of the aging effect by the sequence of directly measured dead times under biased condition. Fitted line reveals on average 1.2 ms/cycle increasing. (b) The distribution of the dead time during a thermal cycle [4].*

I also studied the temperature dependence of the dead time. Applying the same sequence of pulses as shown in Figure 6.4.a and making the same dead time distribution as shown in Figure 6.4.c and d, I determined the length of the dead time of a specific device first at 300 K, after at 350 K and finally at 300 K again. The corresponding fitted Gaussian probability density functions are plotted in Figure 6.5.b, which shows very robust temperature dependence. As the temperature was increased by 50 K the dead time decreased almost 2 orders of magnitude. When the temperature was reduced to the initial value, the dead time returned almost the same value as before the heat treatment, only a modest decrease could be detected. The orders of magnitude change in the dead time shows that thermally activated process is involved. The modest permanent decrease may be attributed to that the heat treatment may reverse the aging effect. In contrast to the strong temperature dependence, the dead time was always presented with the same order of magnitude in the entire pressure range, where the SiO$_x$ resistive switching was achievable ($\approx 4 \cdot 10^{-6}$ to $5 \cdot 10^{-4}$ mbar).

These findings indicate that instead of a thermally activated process is responsible for the observed dead time rule. The relatively long dead times in my measurements imply that slow microscopic processes are involved. Since the dead time does not depend on the driving condition, the background process can not be voltage driven, it happens spontaneously, even at zero bias. According to in situ HRTEM imaging [19] the reset transition is interpreted as a self-heating induced amorphization which destroys the conducting crystalline nanowire. Supposing this mechanism my results
demonstrate that right after the reset pulse induced amorphization the silicon can not be formed crystalline again by applying electric field. A thermally driven reorganization, presumably slow diffusion process, may take place before the OFF state turn into switchable state.

In recent years there were some suggestions to explain the phenomena of dead time. The research group of M. Tour wrote the followings about this effect: "during the falling edge in a reset pulse, the device is in a "hot" state since the voltage starts from the reset region, as opposed to a "cool" state in the set operation. This "hot" state may prevent the set process incurred during the falling edge. The detailed study of this aspect has not yet been done." [71]. Although they noticed the contradiction between the slow I(V) characteristics and the fast pulsed measurements and they suggested further investigations, they did not publish any other paper about this effect later. Their proposed "hot" and "cool" states could be the same as our "OFF*" and "OFF" states in Figure 6.3.c, respectively. Nevertheless the thermal time constant of such a small systems must be several orders of magnitude shorter ($\tau_{thermal} < 1 \mu s$), than the dead time. Therefore, the "hot" and "cool" modifiers can not refer to the temperature of the region, this could motivate the use of the quotation marks in their interpretation.

Furthermore the dead time must be closely related to backward-scan effect, studied by the group of J. C. Lee. They also observed that the final state of the device depends on the speed of the reverse voltage sweep during the reset operation [78]. Although they give a detailed microscopic model to the switching mechanism, the backward sweep effect was not explained. According to our measurements, however, the dead time is a more general phenomenon, it happens spontaneously, even at zero voltage and not only activated by the backward sweep. Nevertheless, the characteristic time of the reverse voltage sweep is 4-6 orders of magnitude faster ($10 \mu s$-1 ms) than I observed for dead time. This significant difference may be attributed to the difference in size of the two systems. In my case the switching site is confined into a well-defined region of a nanogap with the narrowest cross section of few nanometers. There could be only one or few nucleation points, where the recrystallization process can begin, which may take a considerable amount of time. In contrast for a larger vertical device recrystallization can occur at any segment of a much larger cross section. Therefore, the characteristic time scale is expected to downscale with increasing device cross section, which could explain the significant difference. Furthermore the dead time may be also sensitive to the microscopic details of the SiO$_x$ layer, thus it is expected to be sensitive to the growing method or the pretreatments.

In the 1960s the dead time effect was also observed by J. G. Simmons and G. Dearnaley in Au-SiO-Al devices [12]. Simmons suggested extrinsic switching mech-
anism, he assumed that the silicon monoxide layer is injected by the gold ions. The dead time was attributed to the slow diffusion of trapped charge from localized states to the conduction band and finally to the anode. Few years later Dearneley et al. offered another model to the dead time [66]. They assumed many filaments conduction model consists of either conductive Si-O-Si or metal ions. The reset process was attributed to the breakdown of the filament. During the breakdown the electron scattering is enhanced and the surrounding insulator become polarized. At this state the filament can not be reconnected by applying high electric field, at first the trapped charges have to be relaxed by thermally activated Poole-Frenkel emission. However, the bias independent dead time contradicts to their model, since the characteristic time of the trapped charges and thus the dead time should depend on the driving voltage.

6.2.2 Multiple timescales in SiO\(_x\) switches

The time-resolved measurements revealed that there are another timescales besides the dead time, which governs the switching dynamics of SiO\(_x\) memristive system. Figure 6.6.a shows the real time response (blue/red curve) of the resistive switch to a set (left panel) and a reset (right panel) pulse (black curve). The linear voltage ramp between 0V and 1V before and after the square pulse are used to determine the initial and final resistance at low bias. As it is seen, the transitions between the two resistance states significantly differ from the one in Ag\(_2\)S memristors. The set and reset processes are not a gradual transitions, but the device stays in its initial state for a certain time delay (\(\tau_{\text{set}}\) and \(\tau_{\text{reset}}\)), which is mostly followed by an abrupt switching to the final state. The resolution of the latter switching time (\(\tau_{\text{switch}}\)) is always limited by our instrumental bandwidth. This delayed and abrupt action implies that during both the electric-field driven crystallization (set process) and the Joule-heating induced amorphization (reset process) "hidden" microscopic processes take place, which do not directly affect the resistance of the device. For instance, in case of set transition the nucleation can be a possible process.

In memristive systems the highly nonlinear behavior is a commonly observed phenomena: as the bias voltage is increased the speed of the resistance transitions exponentially accelerate. It enables us to program the device fast at high voltage and read the actual state noninvasively at low bias. However, the length of set and reset times have not been measured yet in SiO\(_x\). For this purpose I applied voltages pulses, shown in Figure 6.6.a, with various amplitudes and I monitored the set and reset times. Figure 6.6.b shows for both the set (left panel) and the reset (right panel) transition, that a modest change in the driving voltage induces
several orders of magnitude changes in the set/reset time. The black dots show the switching time for 20 individual measurements, while colored ones correspond to their averaged values. This result is in accordance with ref [73], in which the resistance change \((R_{OFF}/R_{ON})\) was examined as the function of the pulse length and amplitude. However this resistance ratio does not describe the set/reset transition precisely in this system it is more useful in case of gradual transitions.

For both cases nearly linear tendencies can be observed on the logarithmic set/reset time axis. In case of the reset time the slope of the trend line is larger and thus higher acquisition frequency was needed. For these pulsed measurements the high frequency measurement setup was used, introduced in Chapter 3.

![Figure 6.6: a) Illustration of the time-resolved set and reset transitions. b) Statistical analysis of the set/reset times as a function of the set/reset voltages. The mean values are highlighted in blue/red [4].](image)

### 6.3 Further characteristics of the switching

Finally I turn to further electrical characterizations to clarify whether the shrinkage of the switching region below 10 nm affects the operation or reliability of SiO\(_x\) memristors.

Figure 6.7 shows the fastest switching for both reset (left panel) and set (right panel) transitions. Applying proper driving voltage 50 ns fast resistance transition can be achieved, limited by the instrumental setup. These switching times are in the same range as other research group have published [71, 73]. According to measurements on similarly sized nanodevices [256, 257] and thermal models based on the classical theory of heat conduction [258–260] the thermal time constant of the active junction region is in the range of few nanoseconds. This value is smaller than the length of the shortest voltage pulse (50 ns), the temperature reaches the steady state almost immediately.
state right after the pulse edges. This implies that even faster operation could be performed.

Considering the power values we can give an estimation to the temperature during the resistance transition. The typical electric power during the reset transition is few mW, which is similar as the electrical breakdown of graphene. In the latter case the temperature can exceed the 1000 K, where the heat is dissipated at much larger volume. Although in case of SiO\textsubscript{x} embedded switch the heat is supposed to conducted better towards the substrate, the similarly high temperature is expected at the moment of the reset event. During the set transition the electric power is more than 4 orders of magnitude smaller, therefore, no significant self-heating is expected.

![Figure 6.7](image)

**Figure 6.7:** a) Demonstration the fastest set/reset transitions achieved at the time resolution of our pulsing setup. b) I(V) characteristic starting from HRS and subsequent reverse voltage sweep on logarithmic current scale. The arrows show the direction of voltage sweeps, the vertical dashed line at 1V indicates the ON/OFF ratio of \(\approx 10^5\). c) Endurance test up to \(10^3\) switching cycles by voltage pulses of 3.5 V (set) and 9 V (reset). d) Evolution of ON resistance under constant electrical stress. As the amplitude and the length of the voltage signal increase the device become more conducting (dots). The repeated measurement show good reproducibility (lines) [4].

Figure 6.7.b shows a slowly recorded I(V) curve starting from OFF state and ended in ON state on logarithmic current scale. The vertical dashed line at 1V marks the OFF/ON resistance ratio of \(>10^4\). This value is as high as generally observed in this system. Figure 6.7.c illustrates device’s response to a repeated set/reset pulse train. After each pulse the current was measured at 1 V, as the dashed line shows in
Figure 6.7.b. The samples show excellent endurance, they can be written and erased more than $10^3$ times without any degradation. The high OFF/ON ratio ($>10^4$) is still present.

Multilevel programming is a commonly observed phenomenon in SiO$_x$ based resistive switches [70, 261, 262], but it is not evident, whether it is maintained in such a small size. The conducting filament can be partially recovered and ruptured by the variation of the amplitude and length of the driving signal. In case of SiO$_x$ this phenomenon can arise from the variation of size of the nanocrystals and their relative position to each other. However, as the device’s size shrinks, this degree of freedom may scale down as well.

To study this effect I applied 200 ms long set pulses with different amplitudes (3 V, 3.5 V and 4 V), while the current was measured simultaneously. The initially erased device switched to ON state when the set voltage dependent set time passed and after the evolution of the ON resistance was monitored under the constant voltage stress. For each set voltage the pulsing scheme was repeated by 20 times. Figure 6.7.d shows the evolution of the averaged resistance versus time (dots) after set events. The error bars correspond to the standard deviation of the 20 curves. As the set voltage increases the device jumps to less resistive ON state right after the set process ($t \approx 10^{-4}$ s) and the electrical stress further decrease the resistance in time. It must be noted that in the individual time resolved curves discrete resistance jumps are more typical than the smooth reduction. This result verifies that different ON resistances can be set by varying the parameters of the set pulse enable us for multilevel programming. In order to test its reproducibility I applied the same series of voltage pulses on the same device. The obtained resistance values (lines in Figure 6.7.d) were the same within the error limit. The corresponding microscopic process could be that the constant voltage stress heats the silicon crystals inside the active region which can promote their growth.

### 6.4 Conclusions

In conclusion, I have studied SiO$_x$ based resistive switching memories, whose active regions were confined into few nanometer wide graphene gaps. Since the formation of the switching region is electric field driven, it can be anticipated that the intrinsic resistive switching in the SiO$_x$ layer also takes place within a similarly short length scale. The small dimensions of the switching region was confirmed by the low electroforming voltages. The devices still show fast switching, high endurance, high OFF/ON ratio and multilevel programming. The resistive switching capability of the SiO$_x$ memristive system has not been demonstrated yet in the sub-10 nm size.
The time resolved measurement revealed that the device operation is governed by multiple physical time scales. The resistance transitions do not happen right after the bias voltage reaches the set or reset region. The device stays in the initial state for a certain period of time \((\tau_{\text{set}}, \tau_{\text{reset}})\) and after that the resistance changes abruptly. The timescale of the transitions between the two resistance states are below our experimental resolution (50 ns). The modest variation in the set/reset voltage induces exponential change in the set/reset time, referring to voltage driven mechanisms.

Besides the above mentioned switching times, another fundamental time scale, the dead time was also identified. After switching OFF the device, it can not be set to ON state again as long as the dead time has not passed. The detailed study of dead time revealed that unlike the set and reset time, its length does not depend on the driving conditions. However, increasing the temperature by 50 K the dead time decreased almost two order of magnitudes. These results suggest that instead of voltage driven mechanism, thermally assisted structural rearrangements may take place during the dead time. This phenomenon has fundamental technological impact since it combines the positive properties of unipolar and bipolar switches, using unipolar voltage pulses both resistance states can be achieved at zero bias.
The major conclusions of this Ph.D. work are summarized in the following thesis statements.

1. I developed a novel measurement setup for the controlled electrical thinning and breakdown of nanofabricated junctions. This included the development of a new high vacuum sample holder, the assembly of an optimized measurement setup and the development of a versatile measurement program. This measurement system enables us to reduce the active region of nanofabricated devices well below the resolution of present lithographic techniques. The specialty of this system is the pulsed breakdown technique, which allows us to expose the device to much shorter voltage intervals than in real-time feedback-controlled systems. Applying this setup I could establish a few nanometers wide gaps in nanofabricated Ag wires. I have extended this method on single-layer chemical vapor deposited graphene nanostripes achieving nanogaps with measurable tunnel current with a yield over ($\approx 98\%$). The wafer scale growth of CVD graphene enabled the simultaneous fabrication of a large ensemble of devices on a single chip. The statistical analysis of hundreds of devices reveals typical gap sizes between $0.3\ \text{nm}$ and $2.2\ \text{nm}$ [1–3].

2. I studied the resistive switching phenomena of nanofabricated Ag-$\text{Ag}_2\text{S}$-Ag memristors [2]. I have demonstrated that the resistive switching can be established using a simplified sample design lacking the conventionally employed inert electrode. In this design a simple lithographic step is sufficient to fabricate the base structure, which is an asymmetrically shaped Ag nanowire. I have established the ultrasmall resistive switching region by the controlled
electromigration of the Ag nanowires, and the in-situ sulfurization of the such created nanogaps. I have demonstrated that these devices exhibit the conventional switching characteristics of nanometer-scale Ag$_2$S memristors, such that the direction of the switching is governed by the inhomogeneity of the local electric field due to the geometrical asymmetry of the device. In similar devices I have also demonstrated stable room temperature atomic switching phenomenon, indicating that the surrounding Ag$_2$S matrix stabilizes the atomic switching process.

3. I analyzed the influence of the environmental conditions on the electrical breakdown of graphene nanostripes [3]. The systematic study of the breakdown power as the function of pulse length and pressure revealed two fundamentally different breakdown processes. I have found, that in high vacuum a significantly higher power was needed to achieve the breakdown than in atmospheric pressure air. Using a thermal model I rescaled the breakdown power to the maximal local temperature of the graphene stripe. Assuming thermally activated processes I estimated the activation energies of the physical mechanisms involved in the breakdown. The significantly different activation energies are consistent with oxidation in air and sublimation in high vacuum. Using two different substrates (SiO$_2$ and Si$_3$N$_4$), I found that the oxygen content of the SiO$_2$ substrate does not play role in the breakdown process.

4. I investigated the resistive switching phenomena of graphene-SiO$_x$-graphene devices [4]. The intrinsic resistive switching in the SiO$_x$ layer was confined under a few nanometers wide graphene gap, resulting in a yet unexplored, sub-10 nm size-scale switching region of SiO$_x$. My detailed electrical characterizations revealed that these ultrasmall devices exhibit a significantly smaller electroforming voltage than conventional SiO$_x$ switches, such that the further beneficial properties of larger devices, like fast switching speed, excellent endurance and data retention are maintained. I have performed detailed time resolved measurements to identify the physical timescales governing the device operation. I have demonstrated, that the switching is not a gradual transition: the device keeps its initial state for a certain period of time after the voltage is applied, and finally an abrupt resistance change is observed, which is faster than the $\approx 50$ ns temporal resolution of the measurements. I demonstrated that a modest, linear decrease of the set/reset voltage induces an exponential slowdown of the set/reset operation. I have also identified another fundamental time-scale, the dead time. I found, that after switching OFF the device, it cannot be set to the ON state again as long as the dead time has not passed,
even if the driving signal is sufficient for a set transition. The detailed study of the dead time revealed that its length does not depend on the driving conditions, however it could be decreased significantly by a modest increase of temperature, indicating a thermally activated rearrangement of the switching region.
Throughout the years of my PhD work I have received support from many people to whom I would like to express my gratitude. First of all, I am very grateful to my supervisor András Halbritter for giving me the opportunity to do my PhD in his group and for coordinating my work. The discussions and his ideas were always inspiring. A special thanks goes to Péter Makk for helping me with various problems and for the many discussions and advice since my BSc studies.

I also express my gratitude to Michel Calame for inviting me to his research group and for the fruitful collaboration and discussions. I acknowledge Cornelia Nef for her help during my stay in Basel and our cowork for developing the high yield graphene nanogap fabrication technique. I thank Maria El Abbassi for sharing our research experiences and for our work in respect of graphene breakdown mechanisms. I also thank them for fabricating and sending me the graphene samples.

I am also grateful to Miklós Csontos who gave a lot of advice and helped in the interpretation and discussion of the results. I thanks Mészáros Gábor for building me the multi-channel current amplifier. Further I want to express my gratitude to the people with whom I worked together in our group, Ágnes Gubicza, Botond Sánta, András Magyarkuti and special thanks for Zoltán Balogh with whom I started my work together and we helped each other a lot.

I am very grateful to János Volk for providing me the possibility to use the nanofabrication facilities of MFA for the careful reading of my thesis. I also thank István Lukács for his invaluable help and advice in device fabrication.

I am also pleased to thank my friends and colleagues: Szabolcs Csonka, Zoltán Scherübl, Ádám Butykai, Dávid Szaller, Gergő Fülöp and Bálint Fülöp. They all contributed to lively working days and many discussions about physics and other
topics.

I am also grateful to the support from the mechanical workshop. Béla Horváth, Sándor Bacsa and Krisztián Németh for manufacturing the sample holders and for the lots of ideas. Not to forget Edit Honti, Tímea Varga and Edina Beck from the secretary, who make all administrative matters run smoothly.

Finally, I thank my family for all the support they provided me over all the years of my studies.
Publications related to the thesis statements


$^1$These authors contributed equally to this work.


[34] Z. hui Lin and H. xia Wang. “Image encryption based on chaos with PWL memristor in Chua’s circuit”. In: 2009 International Conference on Communications, Circuits and Systems, 2009, 964–968


127


Y. He, H. Dong, T. Li, C. Wang, W. Shao, Y. Zhang, L. Jiang, and W. Hu, Graphene and graphene oxide nanogap electrodes fabricated by atomic force microscopy nanolithography; Applied Physics Letters, 97, 133301 (2010)


J. G. Simmons, Electric Tunnel Effect between Dissimilar Electrodes Separated by a Thin Insulating Film; Journal of Applied Physics, 34, 2581–2590 (1963)

J. G. Simmons, Generalized Formula for the Electric Tunnel Effect between Similar Electrodes Separated by a Thin Insulating Film; Journal of Applied Physics, 34, 1793–1803 (1963)


[258] D. Cai, Z. Song, H. Chen, and X. Chen, Temperature model for Ge$_2$Sb$_2$Te$_5$ phase change memory in electrical memory device; Solid-State Electronics, 56, 13-17 (2011)


