# Synthesis Methods and Moiré Structures Formed by Graphene on Non-metal Substrates: A Review

Erika Tomsič

*University of Nova Gorica, Vipavska 11c, SI-5270 Ajdovščina*

**Abstract**

We review the synthesis methods of heterobilayers formed by graphene and different non-metal monolayer substrates (e.g. $MoS_2$, $hBN$, $WSe_2$). Recent developments in the field consent the heterostructure assembly layer-by-layer via deterministic wet and dry transfer methods, which would be difficult to fabricate by traditional bottom-up approach. Here we describe and compare the alternative methods used in literature for various substrate materials. We show the varied moiré patterns formed by graphene and the mentioned van der Waals materials, where present in literature, and offer, through scanning tunneling microscopy (STM) images, indication of the strength of interaction between the two layers.

*Keywords:* graphene, heterostructures, moiré patterns

## 1. Introduction

Since a breakthrough work in 2004 [1], research on graphene has expanded greatly. Graphene ($Gr$) is interesting because of its high electrical conductivity and elastic strength. Similarly, other related two-dimensional materials, such as hexagonal boron nitride ($hBN$), molybdenum disulfide ($MoS_2$), zinc selenide ($ZnSe$), received great interest. When two (or more) such atomically thin graphene-like materials are placed on top of each other, their properties change and a material with new hybrid properties is created. This paves the way for the design of new materials and nano-devices.

*Email address:* erika.tomsic@ung.si (Erika Tomsič)

Nevertheless, the preparation of such so-called van der Waals materials remains a major challenge. Heterostructure fabrication methods vary considerably, as do their results. One of the consequences of the fabrication of van der Waals artificial structures is the observation of moiré patterns (MP). They arise in twisted artificial bilayers and in bilayers of materials that have a slightly different periodicity. MPs can drastically change the properties of the material and thus provide foundation for the potential engineering of specific electronic devices through the modulation of their properties. The aim of this review is to describe the state-of-the-art synthesis methods that provide a good starting point for the fabrication of various van der Waals heterostructures. This article begins with a brief overview of the properties of graphene (in section 2.1) and then, in each of the following sections, reviews the synthesis technologies of heterostructures composed of graphene and an elected 2-dimensional material: hBN (in section 2.2), molybdenite (in section 2.3), tungsten disulfide (in section 2.4), molybdenum diselenide (in section 2.5) and phosphorene and silicene (in section 2.6). In the respective sections, there are brief discussions of the feasibility of such heterostructures and comments on the produced moiré pattern.

## 2. Methods and Results

### *2.1. Graphene (Gr)*

Before introducing the features pertaining to the heterostructures, it is crucial to understand the structure and the properties of graphene. Graphene is a 2-dimensional crystal, as can be seen in the Figure 1, a single layer of carbon atoms stacked in honeycomb network. Carbon atoms are bonded

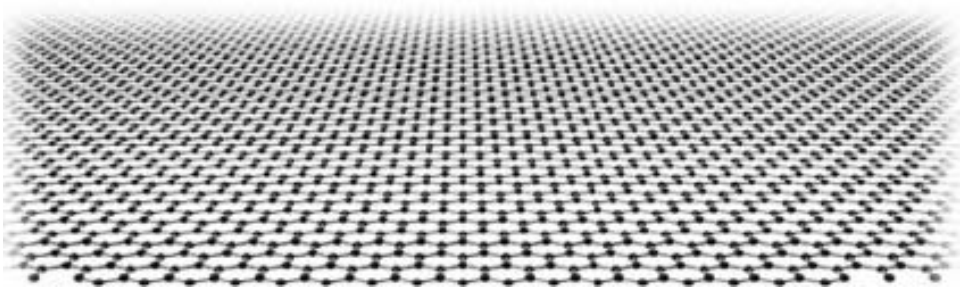

Figure 1: Schematic of graphene hexagonal atomic structure. Carbon atoms are shown by black bullets. Image adopted from [2].

to three nearest neighbours with one shared electron, which partially gives

some remarkable properties to graphene, including high thermal conductivity, mechanical stiffness and extraordinary electron transport properties. The study of graphene earned Geim and Novoselov a Nobel prize in 2010. Now is considered an archetype for other 2-dimensional materials [3]. The original technique to produce graphene is mechanical exfoliation, first published by Geim and Novoselov themselves, in 2004 [1]. Graphene is obtained by micromechanical cleavage of bulk graphite. This is feasible, because the in-plane bond strength is calculated to be ca. 3.6 eV; while the inter-plane bond strength, this is to say the bond between atoms of different layers, is around 40 meV. This relatively weak van der Waals bond is crucial for the success of the mechanical exfoliation method [3]. Graphene devices on a typical substrate such as $SiO_2$ are far from ideal. They present properties that are inferior to the expected intrinsic properties of graphene and display irregularities and disorder. It has been demonstrated that the overall quality of graphene-based devices is dramatically increased when graphene is placed on atomically flat crystal surfaces. Furthermore, despite its electrical efficiency, the use of graphene in the fabrication of nanoscale devices is limited by the presence of a zero band gap. To overcome this limitation, it is proposed to fabricate graphene-based heterostructures in the quest of tailoring its band structure, such as creating additional Dirac points or opening band gaps, through the interaction in the bilayer.

### *2.2. Gr/hBN*

Of particular interest, as a graphene substrate, is hexagonal boron nitride. hBN is a layered material whose planes have the same atomic structure as graphene (see Figure 2). The hBN lattice shows a honeycomb lattice of N

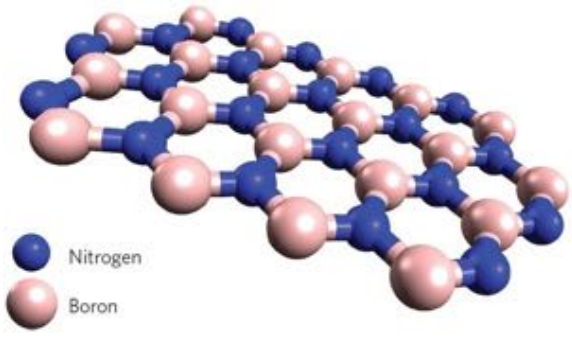


Figure 2: *hBN* atomic structure. Despite retaining a hexagonal lattice akin to graphene structure, they have dissimilar physical properties: while graphene is a zero band gap semiconductor, hBN is a wide bad gap insulator. Image adopted from [4].

and B atoms with a 1.8% longer lattice constant than graphene. It is an insulator that couples only weakly to graphene but its potential generates a graphene superlattice and leads to changes in the graphene's electronic spectrum [5].

### 2.2.1. *Fabrication process of Gr/hBN*

In the literature, the fabrication processes of devices composed of graphene-hBN heterobilayer are quite diverse. In particular, mechanical transfer processes are often used, as in [6]. hBN flakes are micromechanically cleaved from hBN single crystals. hBN is grown in-home by the method described by [7]. The so-obtained flakes are exfoliated onto 285 nm thick $SiO_2$ thermally grown over silicon wafer. Thin flakes, up to a monolayer, are sufficiently transparent to add to the optical path and visible using an optical microscope, as shown in Figure 3a. Similarly, graphene is here exfoliated

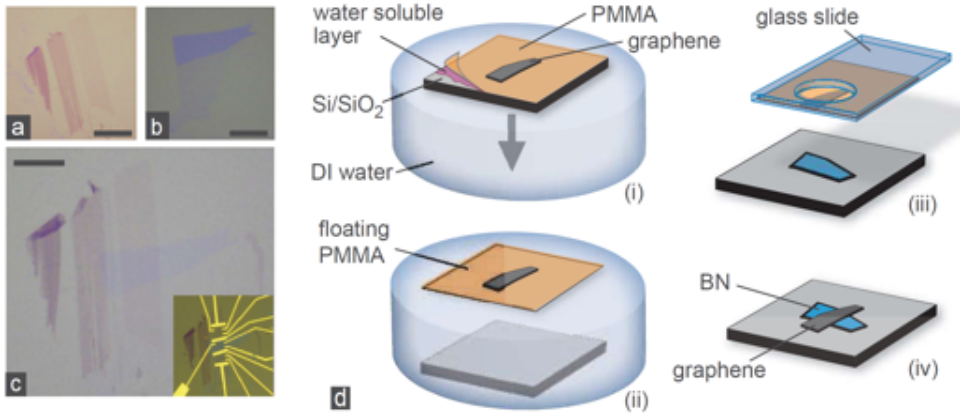


Figure 3: Microscope contrast of h-BN (a) and the graphene flake (b), and $Gr/hBN$ after the transfer process (c). Scale bar is 10 µm. Schematic illustration of the PMMA carrying layer transfer method (d): the water soluble layer dissolves (i) and the graphene flake adhered to PMMA stays afloat (ii). The graphene/PMMA is attached to a glass slide, aligned to hBN flake (iii) and brought into contact with hBN flake (iv). Image adopted from [6].

over a $SiO_2/Si$ wafer coated by a polymer consisting of a PMMA (polymethyl methacrylate) layer and a water-soluble polyester resin (Mitsubishi Rayon aquaSAVE) (see Figure 3 b). The water-soluble layer dissolves in deionized water bath and leaves the PMMA with graphene flake floating in the bath, as in Figure 3d. The so achieved PMMA membrane is placed on a glass slide combined with a micromanipulator with the graphene flake facing down. Using the optical microscope graphene is exactly aligned with the substrate, in this case the hBN flake with sub-micrometer resolution,

and put in contact with it (see Figure 3c). The system is annealed to 110 °C, to encourage the adhesion between PMMA and the substrate. Afterwards, the PMMA is dispersed in acetone. A deterministic all-dry transfer method has been also developed [8]. The setup is very similar to the one described above. In this case, a thin film of PDMS (polydimethylsiloxane) is used for the transport of graphene or the desired 2D material. A small stamp of PDMS (Gelfilm produced by Gelpak) is adhered over a microscope glass slide, as shown in Figure 4. Thin flakes of graphite, up to single layers, are obtained by subsequent cleavages of single crystal graphite with the use of Nitto tape and then placed over the prepared PDMS. The glass slide is then linked to a micromanipulator and aligned and put in contact with the desired substrate, as described before. However, the difference is that, the PDMS can be now slowly raised from the surface with the manipulator, detaching from the surface and leaving behind the transferred flake. The advantage of using this technique is not only the speed in the trans-

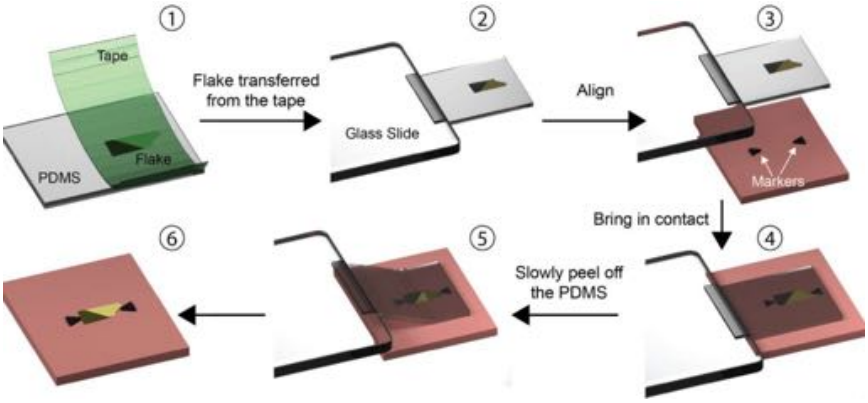


Figure 4: Schematic of all-dry PDMS-assisted transfer method. Graphene is exfoliated onto a PDMS stamp (1), which is put on a glass slide applicable to a micromanipulator (2). Graphene flake is aligned with the substrate (3) and brought in contact with it (4). PDMS is slowly peeled off (5), so that graphene remains attached to hBN on the substrate (6). Image adopted from [9].

fer process, but, more importantly, the absence of any solution or liquid which could impair the cleanness of the employed material. The resulting device is virtually contamination-free. The great drawback of the method is the difficulty of the discrimination between single-layer and few-layer flakes due to the transparent substrate, which does not provide an optical contrast as good as the one created by $SiO_2$. A mixed approach for obtaining Gr/hBN heterostructure is also reported [10]. Commercially accessible

hBN (Momentive AC6004) was exfoliated onto $SiO_2$. Here graphene was grown by means of low-pressure chemical vapor deposition (CVD). No recipe of CVD growth of graphene directly over hBN is described and, to the author's knowledge, available in literature (whereas, the growth of Gr-BN laterally hybrid monolayers is generally known, see [11]). In 2013, a different method for the transfer of 2-dimensional materials such as graphene and hBN was introduced [12]. In case of this pick-up technique, the van der Waals force between the 2D materials is exploited, without the need to expose the inner interface to polymers. A hBH flake is exfoliated onto poly-propylene carbonate (PPC), which is subsequently placed on a PDMS stamp on a microscope slide. By the use of a manipulator the hBN flake is then aligned to the selected graphene flake and brought in contact with it. Due to the van der Waals interaction, graphene prefers adhesion to hBN rater than the substrate (usually $SiO_2$). The pick-up process is favored by a stage temperature of 40 °C, yielding a nearly 100% success rate. The process, as shown schematically in Figure 5, can be repeated to achieve a discretionary number of layers. A device is then produced by transferring the stacked flakes to the desired substrate. Heating up to 90 °C the glass slide can be removed along with the PDMS stamp. Finally, PPC is dissolved in chloroform. As said, with the van der Waals pick-up technique very

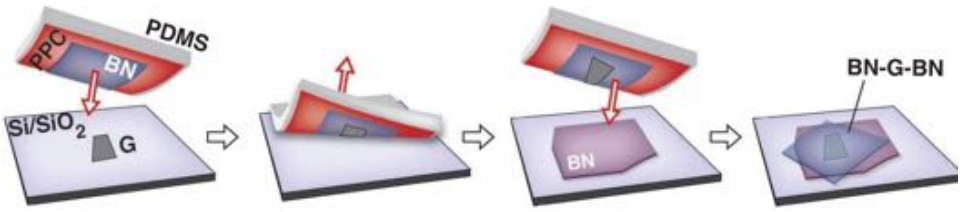


Figure 5: Schematic of the process where an hBN flake on PCC on PDMS (1) is used as a stamp to pick up monolayer graphene (2). The so-formed heterostructure is then transferred over another hBN flake (3) forming this way a hBN/Gr/hBN heterostructure. Image adopted from [12].

clean multilayer heterostructures can be achieved, since only the topmost layer is in direct contact with the transfer polymer. The technique is time-consuming and only concurs to the transfer of heterostructures.

### *2.2.2. Moiré pattern of Gr/hBN*

A lattice mismatch, such as is found between Gr and hBN, and the relative rotation due to random stacking angle, lead to a superlattice formation called moiré lattice. On STM images periodic modulations of the sur-

face are observed, which have various periodicities depending on the substrate and the rotational angle. Not only the topography of graphene is changed but also its electronic structure. Moiré superlattice was observed

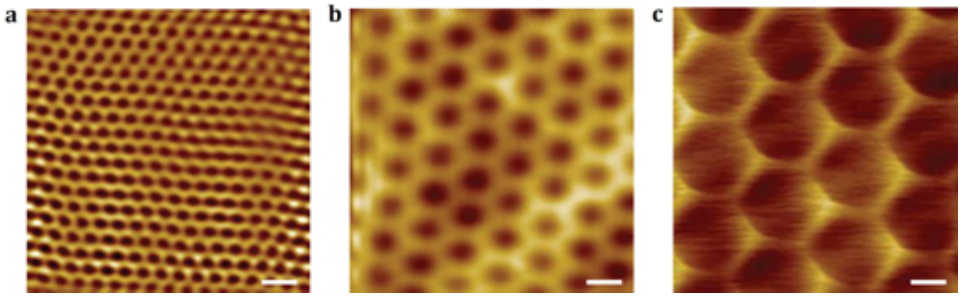


Figure 6: STM images of MP (Gr/hBN) with periodicities of 2.4 nm (a), 6.0 nm (b) and 11.5 nm (c). The scale bars in all images are 5 nm. Image adopted from [10].

on a $Gr/hBN$ fabricated device [10]. The moiré pattern is seen as a small modulation of the apparent surface height. The images in Figure 6 show MP with different periodicities. The periodicity wavelength of the pattern increases as the twist angle decreases. The van der Waals interaction between hBN and graphene, however weak, is not negligible, as the apparent corrugation of the moiré ranges from 0.2 Å to 0.9 Å [13].

### 2.3. $Gr/MoS_2$

Molybdenum disulfide is one of the most studied transition metal dichalcogenides (TMDCs) [9, 14, 15, 16]. Like hBN, molybdenite is considered a graphene-like layered material.

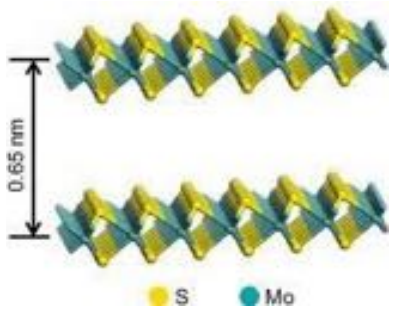


Figure 7: Crystal structure of bilayer $MoS_2$; the yellow bullets represent sulfur atoms and turquoise bullets are molybdenum. Image adopted from [14].

Its bulk material consists of covalently bonded S-Mo-S sheets interacting trough the van der Waals force (see Figure 7), with B-N bond length of about 2.4 Å. It is a semiconductor (unlike hBN, being an insulator) of indirect band gap of 1.3 eV. When reaching the monolayer size it gains a direct band gap of 1.8 eV [14]. Among all the TMDCs (with formula $MX_2$, where $M = Mo, W$ and $X = S, Te, Se$), $MoS_2$ is known to have strong light-matter interactions and excitonic effects. As such, it is the object of studies to combine the advantages of both materials, the high carrier mobility of graphene, and the bandgap of molybdenite.

*2.3.1. Fabrication proccess of Gr/$MoS_2$ and the resulting moiré pattern*

In the following subsection, we describe a graphene transfer experiment onto $MoS_2$ by [16]. A solution of chlorobenzene and PMMA is spin-coated on graphene, grown by CVD on a copper foil. After drying, it results in a layer of 50 nm thickness. The copper substrate is then etched in ammonium persulfate till it is completely removed [17]. After rinsing in isopropanol alcohol, the PMMA with the graphene flake is drawn onto the desired substrate, in this case $MoS_2$ bulk single crystal. The PMMA is removed using acetone. A trapped layer of water at the graphene-$MoS_2$ interface is observed, probably due to the transfer process, which remaizs present even in ultra high vacuum (UHV) conditions; it can be removed by annealing the system to 300 °C. In a similar experiment, it is shown how a large-area $MoS_2$ monolayer can be formed in-house using CVD, and graphene can be transferred onto it [18]. A $SiO_2/Si$ substrate is submitted to $MoO_3$ and sulfur vapor flux up to the temperature of 650 °C. When naturally cooled down, a continuous layer of randomly ordered $MoS_2$ crystal is formed. To stack graphene over molybdenite the previously-described method of PMMA support is used. In 2017, there has been reported a

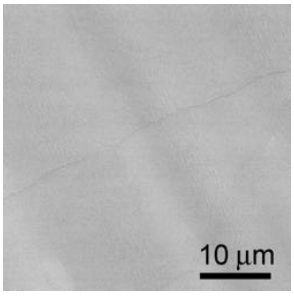


Figure 8: SEM image of the "old"-method transferred graphene flake with a fracture. Image from [19].

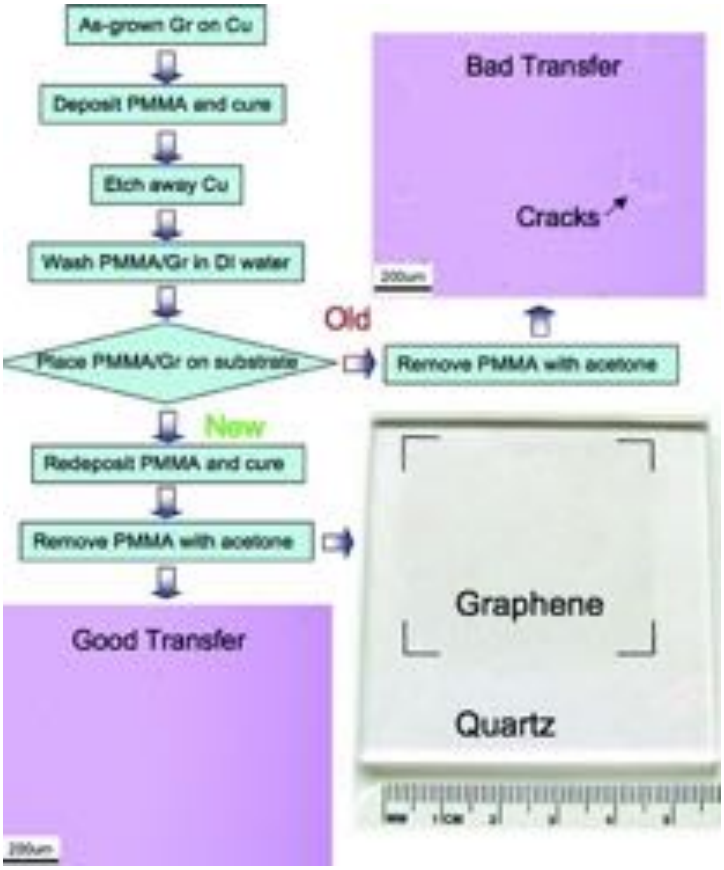


Figure 9: Schematic of "old" and "new" PMMA-assisted method for graphene transfer. Fracture are present on the surface when "old" method is used (upper right image), whereas there aren't any fractures when "new" method is used. Image adopted from [19].

modified PMMA-assisted method for the transfer of graphene directly onto $MoS_2$ supported by SiC(0001) substrate [20]. The technique's advantages are extensively described in a reference paper [19]: the metal surface (commonly Cu foil), on which graphene layer is grown, is usually much rougher than the destination substrate, $SiO_2$ or other van der Waals material such as $MoS_2$. Graphene may not make full contact on the substrate, and strain is produced. This results in fractures and formation cracks (see Figure 8 and upper right image of Figure 9 ) in the transferred graphene after PMMA dissolution when "old" transfer technique is used. Improvement is found when, after transferring the graphene flake over the target substrate, an additional amount of liquid PMMA is placed over it. This way, the earlier-

deposited PMMA is partly dissolved allowing graphene to mechanically relax underneath. Graphene contacts better with substrate, not allowing further strain. Lesser or even no fractures are formed when PMMA is finally dissolved.

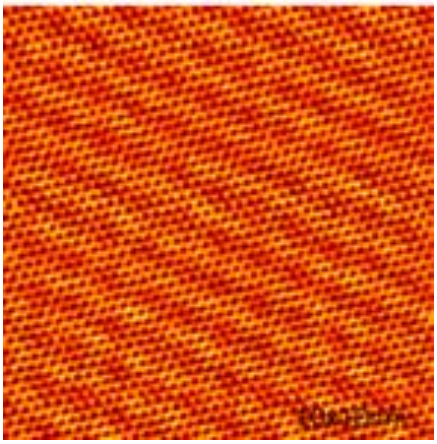

Figure 10: The high resolution STM image demonstrates the cleanliness and flat surface of graphene supported on $MoS_2$. The graphene honeycomb lattice is noticeable, along with a faint moiré periodicity apparent only at very high tunneling current ($I_t = 33\,\text{nA}$ and $V_{bias} = -50\,\text{mV}$). Image adopted from [21].

Graphene on $MoS_2$ presents only a weak moiré structure, as can be seen in Figure 10, visible at only high tunnelling currents of STM [21]. The moiré-generated corrugation is calculated to be only 0.3 Å. This implies that only a weak interaction is destablished between the two layered materials, differently to what happens for the hBN substrate. The weak interaction is ascribed to the big lattice mismatch (the bond length between two carbon atoms is of about 1.42 Å, while the B-N bond length in hBN is 2.41 Å) [16].

### *2.4. Gr/$WS_2$*

The success of Gr/$MoS_2$ heterostructures has enticed the study of other members of the TMDCs family as a substrate for graphene. Being an atomically thin semiconductor of analogue atomic structure, tungsten disulfide ($WS_2$) represents a good replacement for molybdenite. Like the closely related $MoS_2$, the bulk $WS_2$ present a band gap of 1.3 eV; whereas monolayer $WS_2$ has a direct band gap augmented to 1.8 eV.

### 2.4.1. *Fabrication proccess of Gr/$WS_2$*

In literature, the fabrication of Gr/$WS_2$ heterostructure is conducted either by a mixed approach, CVD of tungsten disulfide over the supporting material and transfer of graphene from a different substrate [22], or by a double transfer of both materials. Commonly, the double-transfer approach is used, as in [23]. Here $WS_2$ flakes were exfoliated directly onto $SiO_2$ by micromechanical cleavage of bulk material. Subsequently, graphene is transferred by the wet PMMA-assisted technique formerly described. On

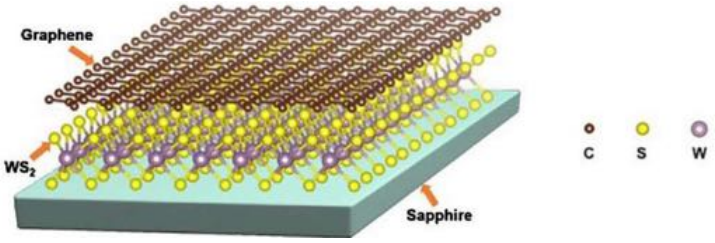


Figure 11: Schematic of Gr/$WS_2$ heterostructure over sapphire. Image adopted from [22].

the contrary, Li et al. realized Gr/$WS_2$ heterostructure by synthesizing the $WS_2$ under-layer directly over the supporting sapphire by CVD [22] (see Figure 11). The graphene, also fabricated by CVD on another sapphire support, was transferred over it. The authors observe a retained good quality of graphene after the transfer process by Raman spectroscopy. Recently (in 2020), Gr/$WS_2$/hBN heterostructure over $TiO_2$ substrate has been fabricated by a subsequent transfer of the three materials [24]. hBN flakes have been obtained by micromechanical cleavage and exfoliated on the substrate. $WS_2$ was placed over it using a polycarbonate film with the help of a micromanipulator. Finally, graphene was transferred on top by a PMMA-assisted method.

### 2.4.2. *Moiré pattern of Gr/$WS_2$*

To the knowledge of the author no STM study of graphene moiré over $WS_2$ was performed. A local band gap modulation in the ARPES (angle-resolved photoemission spectroscopy) mapping due to the superlattice potential is, however, an indirect evidence of a moiré periodicity [24].

## 2.5. *Graphene on other TMDCs*

If few papers are available about the preparation of Gr/$WS_2$, even fewer exist on Gr/$MoSe_2$. Fewer still, or none, on other TMDCs such as $WSe_2$,

$MoTe_2$ and $WTe_2$. Relatively broader is the literature of DFT (density functional theory) calculations of such systems [25, 26, 27, 28].

### 2.5.1. *Fabrication proccess of* $Gr/MoSe_2$

Nevetheless, Froelicher et al. [29] succeeded in producing $Gr/MoSe_2$ heterostructure on $SiO_2$ with the deterministic dry-transfer process proposed by Castellanos-Gomez et al. [8] using PDMS stamps (see Figure 4). Control photoluminescence spectra of such samples do not present any defect-induced background, showing good quality of the graphene. Yet, contamination "pockets" on such structures and decoupling regions (see Figure 12), observed by optical microscope suggest that the transfer technique should be somewhat refined.

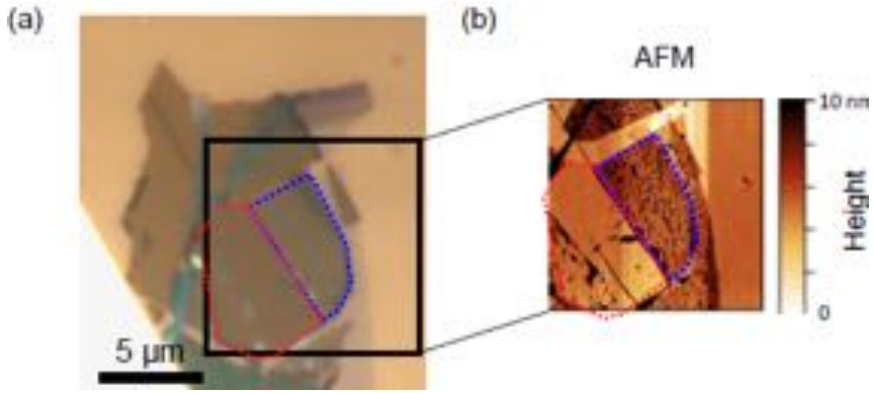


Figure 12: Optical image of the fabricated $Gr/MoSe_2$ device (a) and the corresponding AFM (atomic force microscopy) image (b). In red, the atomically flat perfectly coupled region and, in blue, the uncoupled region with contamination pockets; image adopted from [29].

## 2.6. *Graphene on phosphorene and silicene*

Phosphorene and silicene, considered to be the two-dimensional analogues of graphene for phosphorus and silicium, are not much studied yet; at least compared to the relatively extensive knowledge on graphene. Phosphorene or two-dimensional black phosphorus (BP) is a layered material and the most stable phosphorus allotrope. What differentiates it from graphene is not only its orthorombic atomic structure, see Figure 13, but also a non-zero band gap (0.3 eV bulk BP, 2.0 eV monolayer BP). Its properties are such that it fills the gap provided by graphene and TMDCs and make it attractive in ultrafast photonics applications [30]. On the other hand, silicene possesses a hexagonal honeycomb structure similar to that of graphene,

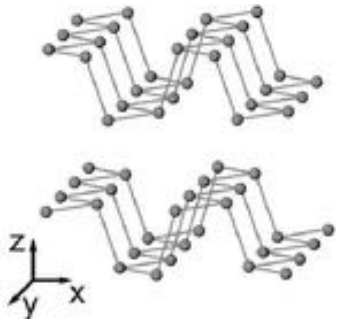


Figure 13: Schematic diagram of the black phosphorus atomic structure. Its layered structure is made of sheets of atoms arranged in a puckered honeycomb lattice. Image adopted from [31].

but has a periodically buckled topology. As graphene, it has a zero band gap, which can be tuned by an external electric field. Phosphorene- and silicene-based heterostructures are predicted to have remarkable electronic and thermal properties; unfortunately their assembly is technically demanding as both suffer from oxidation when exposed to free air.

*2.6.1. Fabrication of Gr / BP and the resulting moiré pattern*

A *Gr / BP* device was successfully fabricated in 2018 [32]. Experiments were performed in a glove box under argon gas to avoid the degradation of BP. A BP flake of appropriate size was mechanically exfoliated by Nitto tape on an oxidized silicon wafer. The graphene was positioned on top by the dry-transfer method already mentioned above. The sample was annealed at 200 °C in UHV to obtain conformal contact at the interface. The so-obtained heterostructure was then surveyed on STM. The superposition of hexagonal graphene and orthorhombic BP lattices leads to different moiré patterns, depending on both, the lattice mismatch and rotation angle [32]. In Figure 14 a variety of such MP is shown. On each image is shown the coexistence of two sets of moiré superlattices, denoted by $\lambda_1$ and $\lambda_2$ (wavelengths of each superlattice), which arise due to the two different BP crystallographic vectors in its orthorhombic lattice.

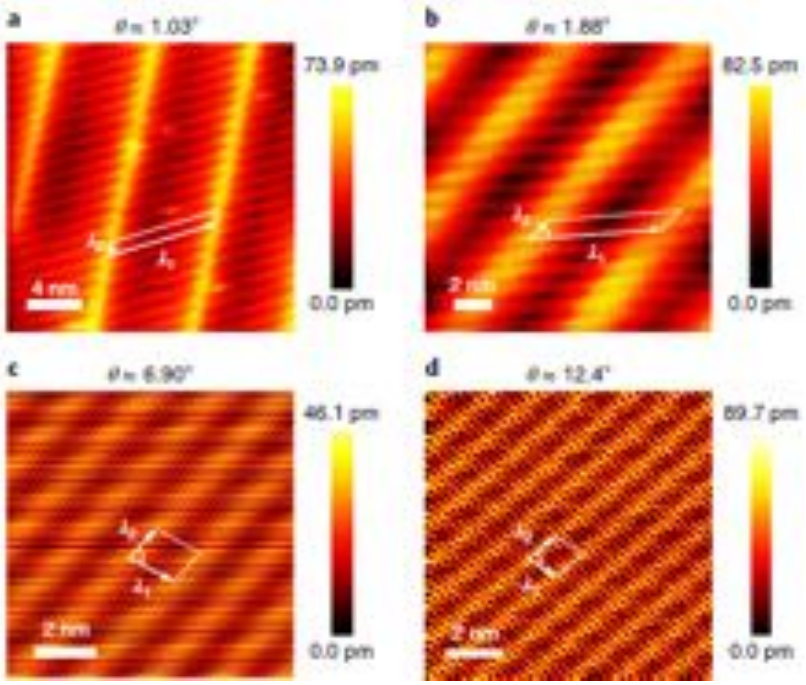


Figure 14: High resolution STM images of $Gr/BP$ for different rotation angles $\theta$ between the BP layer and the graphene layer of the heterostructure, adopted from [32]. The topography shows the coexistence of two superlattices denoted by the wavelengths $\lambda_1$ and $\lambda_2$.

### *2.6.2. Fabrication of Gr/silicene and the resulting moiré pattern*

Yet, a different, noteworthy technique was used for the construction of a graphene-silicene heterostructure [33]. A graphene monolayer is grown by CVD on a Ru(0001) substrate. Si atoms are then deposited on top, as shown in Figure 15. When annealed to 900 K an intercalation process initiates resulting in silicone nanoflakes, monolayers or multilayes. At high dosages of Si a diffusion mechanism takes place leading to the formation of monolayer silicene. The examination of the so-produced heterostructure by STM reveals the existence of two periodicities, see Figure 16. The first, a honeycomb feature, can be easily assigned to the primary graphene hexagonal lattice, which suggests that the graphene layer remains intact despite the intercalation process. The second, longer range periodicity confirms the formation of a moiré-induced corrugation. DFT calculations show that the interaction between graphene and silicene layers is weak [33].

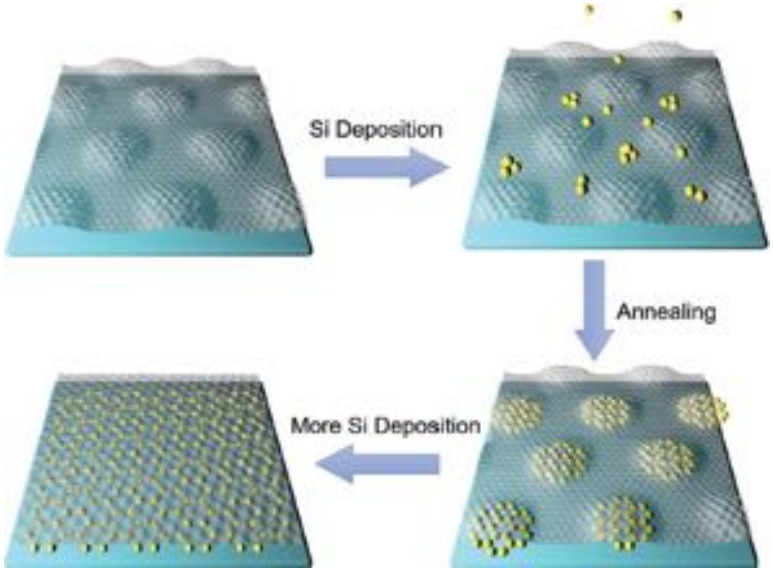


Figure 15: Schematic of the intercalation process of Si under a graphene layer from [33]. Small silicene nanoflakes form when a small Si dosage is applied, whereas with more Si deposition there is produced monolayer silicene encapsulated between the ruthenium substrate and graphene.

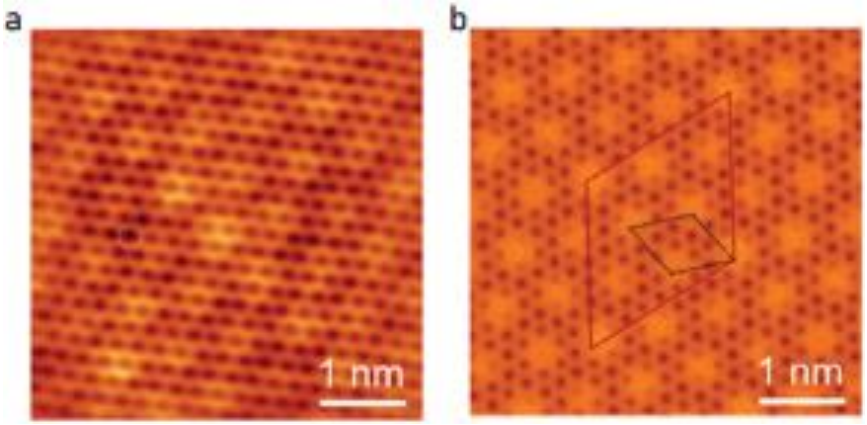


Figure 16: STM image of Gr/silicene/Ru(0001) (a) showing the intact carbon lattice compared with the simulated STM image of such configuration (b) [33]. The primary graphene hexagonal lattice can be seen, along with a longer range moiré periodicity.

## 3. Conclusion

Presently, many deterministic transfer techniques of van der Walls materials, with some variations depending on the specific materials, are at one's disposal. As graphene properties vary extensively with the environment, the cleannes of the developed sample should be the main object of the transfer procedures, even to the disadvantage of expedition. Consequently, allowing for few exceptions, given the nature of the materials, the use of dry-transfer methods is to be preferred. The employment of the new substrates over the standard $SiO_2$ substrate gives not only a renewed insight on the properties of the graphene but also, adding another van der Walls material and tuning the relative rotations to create a variation of moiré structures, will enable the development of a number of novel device characteristics. This tunability of the created Moiré pattern is already showing promise for the creation of novel structures with tailored properties such as tunneling transistors and photovoltaic devices.Additional effort should be put into the consideration of vertical heterostructures where graphene is associated to less studied TMDCs ($WSe_2$, $MoTe_2$ and others). Computational analyses of such heterostructures have already predicted interesting properties for nanoelectronic and spintronic applications which only await experimental validation.


## Acknowledgments

This work was funded in part by the Slovenian Research Agency (research core funding No. P1-0055 and by the project N1-0024) and the Young Researchers program.


## References


[1] K.S. Novoselov, A.K. Geim et al. *Electric field effect in atomically thin carbon films.* Science 306.5696 (2004): 666-669, https://doi.org/10.1126/science.1102896. 1, 2

[2] A.K. Geim, K.S. Novoselov *The rise of graphene.* Nanoscience and technology: a collection of reviews from nature journals. 2010. 11-19, https://doi.org/10.1038/nmat1849. 2

[3] A.K. Geim *Nobel Lecture: Random walk to graphene.* Reviews of Modern Physics 83.3 (2011): 851, https://doi.org/10.1103/RevModPhys.83.851. 3

[4] T.T. Tran et al. *Quantum emission from hexagonal boron nitride monolayers.* Nature nanotechnology 11.1 (2016): 37, https://doi.org/10.1038/nnano.2015.242. 3

[5] K. Watanabe, T. Taniguchi, H. Kanda. *Direct-bandgap properties and evidence for ultraviolet lasing of hexagonal boron nitride single crystal.* Nature materials 3.6 (2004): 404-409, https://doi.org/10.1038/nmat1134. 4

[6] C.R. Dean et al. *Boron nitride substrates for high quality graphene electronics.* Nature nanotechnology 5.10 (2010): 722-726, https://doi.org/10.1038/nnano.2010.172. 4

[7] T. Taniguchi, K. Watanabe. *Synthesis of high-purity boron nitride single crystals under high pressure by using Ba-Bn solvent.* J. Cryst. Growth 303, 525–529 (2007), https://doi.org/10.1016/j.jcrysgro.2006.12.061. 4

[8] A. Castellanos-Gomez et al.*Deterministic transfer of two-dimensional materials by all-dry viscoelastic stamping.* 2D Materials 1.1 (2014): 011002, https://doi.org/10.1088/2053-1583/1/1/011002. 5, 12

[9] R. Frisenda et al. *Recent progress in the assembly of nanodevices and van der Waals heterostructures by deterministic placement of 2D materials.*Chemical Society Reviews 47.1 (2018): 53-68, https://doi.org/10.1039/C7CS00556C. 5, 7

[10] M. Yankowitz et al. *Emergence of superlattice Dirac points in graphene on hexagonal boron nitride.* Nature Physics 8.5 (2012): 382-386, https://doi.org/10.1038/nphys2272. 6, 7

[11] G.H. Han et al. *Continuous Growth of Hexagonal Graphene and Boron Nitride In-Plane Heterostructures by Atmospheric Pressure Chemical Vapor Deposition.* Acs Nano 7.11 (2013): 10129-10138, https://doi.org/10.1021/nn404331f. 6

[12] L. Wang et al. *One-Dimensional Electrical Contact to a Two-Dimensional Material.* Science 342.6158 (2013): 614-617, https://doi.org/10.1126/science.1244358. 6

[13] R. Decker et al. *Local Electronic Properties of Graphene on a BN Substrate via Scanning Tunneling Microscopy.* Nano letters 11.6 (2011): 2291-2295, https://doi.org/10.1021/nl2005115. 7

[14] X. Li, H. Zhu. *Two-dimensional $MoS_2$: Properties, preparation, and applications.*Journal of Materiomics 1.1 (2015): 33-44, https://doi.org/10.1016/j.jmat.2015.03.003. 7, 8

[15] C. Lu et al. *$MoS_2$: Choice Substrate for Accessing and Tuning the Electronic Properties of Graphene.*Physical review letters 113.15 (2014): 156804, https://doi.org/10.1103/PhysRevLett.113.156804. 7

[16] H. Coy Diaz et al. *Interface properties of CVD grown graphene transferred on to $MoS_2$(0001).*Nanoscale 6.2 (2014): 1071-1078, https://doi.org/10.1039/c3nr03692h. 7, 10

[17] A. Pirkle et al. *The effect of chemical residues on the physical and electrical properties of chemical vapor deposited graphene transferred to $SiO_2$.*Applied Physics Letters 99.12 (2011): 122108, https://doi.org/10.1063/1.3643444. 8

[18] W. Zhang et al. *Ultrahigh-Gain Photodetectors Based on Atomically Thin Graphene-$MoS_2$ Heterostructures.*Scientific reports 4 (2014): 3826, https://doi.org/10.1038/srep03826. 8

[19] X. Li et al. *Transfer of Large-Area Graphene Films for High-Performance Transparent Conductive Electrodes.* Nano letters 9.12 (2009): 4359-4363, https://doi.org/10.1021/nl902623y. 8, 9

[20] D. Tomer et al. *Spatial inhomogeneity in Schottky barrier height at graphene/$MoS_2$ Schottky junctions.*Journal of Physics D: Applied Physics 50.16 (2017): 165301, https://doi.org/10.1088/1361-6463/aa61d9. 9

[21] H. Coy Diaz et al. *Direct Observation of Interlayer Hybridization and Dirac Relativistic Carriers in Graphene/$MoS_2$ van der Waals Heterostructures.* Nano letters 15.2 (2015): 1135-1140, https://doi.org/10.1021/nl504167y. 10

[22] Z. Li et al. *Q-switching of waveguide lasers based on graphene/$WS_2$ van der Waals heterostructure.* Photonics Research 5.5 (2017): 406-410, https://doi.org/10.1364/PRJ.5.000406. 11

[23] J.D. Mehew et al. *Fast and Highly Sensitive Ionic-Polymer-Gated $WS_2$–Graphene Photodetectors.* Advanced Materials 29.23 (2017): 1700222, https://doi.org/10.1002/adma.201700222. 11

[24] S. Ulstrup et al. *Direct observation of minibands in a twisted graphene/$WS_2$ bilayer.* Science advances 6.14 (2020): eaay6104, https://doi.org/10.1126/sciadv.aay6104. 11

[25] F. Zhang et al. *Strain effects on the Schottky contacts of graphene and $MoSe_2$ heterobilayers.* Physica E: Low-dimensional Systems and Nanostructures 103 (2018): 284-288, https://doi.org/10.1016/j.physe.2018.06.023. 12

[26] E.S. Souza et al. *Probing the local interface properties at a graphene–$MoSe_2$ in-plane lateral heterostructure: an ab initio study.* Physical Chemistry Chemical Physics 20.26 (2018): 17952-17960, https://doi.org/10.1039/C8CP02343C. 12

[27] M. Sun et al. *Effects of structural imperfection on the electronic properties of graphene/WSe2 heterostructures.*Journal of Materials Chemistry C 5.39 (2017): 10383-10390, https://doi.org/10.1039/C7TC03131A. 12

[28] A.K. Geim, I.K. Grigorieva *Van der Waals heterostructures.* Nature 499.7459 (2013): 419-425, https://doi.org/10.1038/nature12385. 12

[29] G. Froehlicher et al. *Charge Versus Energy Transfer in Atomically Thin Graphene-Transition Metal Dichalcogenide van der Waals Heterostructures.* Physical Review X 8.1 (2018): 011007, https://doi.org/10.1103/PhysRevX.8.011007. 12

[30] A. Carvalho et al. *Phosphorene: from theory to applications.* Nature Reviews Materials 1.11 (2016): 1-16, https://doi.org/10.1038/natrevmats.2016.61. 12

[31] A. Castellanos-Gomez et al. *Isolation and characterization of few-layer black phosphorus.* 2D Materials 1.2 (2014): 025001, https://doi.org/10.1088/2053-1583/1/2/025001. 13

[32] Y. Liu et al. *Tailoring sample-wide pseudo-magnetic fields on a graphene–black phosphorus heterostructure.* Nature nanotechnology 13.9 (2018): 828-834, https://doi.org/10.1038/s41565-018-0178-z. 13, 14

[33] Y. Li et al. *Stable Silicene in Graphene/Silicene Van der Waals Heterostructures.* Advanced Materials 30.49 (2018): 1804650, https://doi.org/10.1002/adma.201804650. 14, 15

# Cross-correlation between extragalactic diffuse gamma-ray background and cosmic shear as a tool in indirect dark matter searches

Veronika Vodeb

*University of Nova Gorica, Vipavska 13, SI-5000 Nova Gorica*

**Abstract**

Gamma-rays represent stable standard model (SM) particles, which could originate from dark matter (DM) annihilation or decay under the Weakly Interacting Massive Particles (WIMP) hypothesis. They are good astrophysical messengers, as they travel through the Universe in almost straight lines and are not easily absorbed. Despite all that, the gamma-ray signal coming from DM annihilation or decay is hard to identify because of large astrophysical backgrounds which dominate the gamma-ray flux measured by telescopes. One method for extracting the DM signal from the measured flux is by cross-correlating the gamma-ray signal with other tracers of DM density distribution, such as the weak-lensing shear. As both gamma-ray emission as well as the weak-lensing shear originate from the same DM distribution, a certain level of correlation between the two signals should be detectable. In this paper, the cross-correlation method and how it can be used to search for correlations between the measured gamma-ray flux and the observed weak-lensing shear is presented.

*Keywords:* cosmology, dark matter, cross-correlation, gamma-rays, weak gravitational lensing

## 1. Introduction

The dark matter (DM) puzzle represents one of the main unanswered questions in modern cosmology. Ever since its existence has been firmly established in the last century, there is a strong endeavour in the physics community to detect the DM signal in direct, indirect or collider experiments and

*Email address:* veronika.vodeb@ung.si (Veronika Vodeb)

pinpoint its physical properties, such as mass, interaction cross-sections, distribution in the Universe, etc.

Although all the DM searches so far have been unsuccessful in detecting the particle DM signal, they have provided us with good constraints on the properties of DM particles and significantly constrained the parameter space of particle DM. With new generations of detectors and experiments being build, the aim is to constrain this parameter space even more and gain further insight into the properties of these particles. Better precision of the new generation detectors and more abundant data obtained by these telescopes, in space as well as on the ground, could significantly increase the sensitivity of our experimental setups to the DM signal. In parallel with the experimental development, new and more sophisticated techniques are being implemented for the analysis of experimental data, granting an even greater power to the already quality and abundant data obtained by the telescopes. Such methods of data analysis are particularly useful in DM searches, as they enable us to look for the DM signal even if it is deeply buried in the emissions coming from the astrophysical sources. Extracting a signal from data which is dominated by background events is a very challenging task. In the recent years many novel approaches to data analysis have been established in the field of indirect DM searches, one of which is the *cross-correlation method* proposed in Ref. [1].

*$\gamma$-rays in the Universe.* $\gamma$-rays that travel to Earth and are detected in experiments can be emitted in a variety of processes happening in the Universe. The WIMP hypothesis assumes that DM particles too can interact with standard model (SM) particles. Different particle physics models predict different cross-sections for interactions between the dark and the baryonic sector. Due to a predicted large mass of DM particles ($m_\chi \gtrsim$ few 100 GeV), DM particles are expected to annihilate (or decay) into heavy SM particles, such as W and Z bosons, heavy quarks, $\tau$ leptons, Higgs, etc. Such heavy SM particles are unstable and further decay into lighter and more stable SM particles. The products of DM annihilation or decay which can be detected on Earth are therefore limited to SM particles that are stable enough to survive the travel over very large distances, do not get easily absorbed and can be efficiently detected in experiments. One type of stable SM particles which fit into this description are $\gamma$-rays. From SM particle physics models we know that $\gamma$-rays are emitted as decay or annihilation products of a variety of SM particles. They can travel large distances without getting significantly absorbed, and they are also relatively easily detected in experiments, via the pair production mechanism. Another important prop-

erty of $\gamma$-rays is that they travel through the Universe in relatively straight lines and can therefore provide us with reliable information about the angular position of their point of origin. All these aspects make $\gamma$-rays very promising and widely used messengers in the indirect DM searches.

*Extragalactic diffuse $\gamma$-ray background.* As already mentioned, $\gamma$-rays also originate from different astrophysical processes happening in the Universe. Astrophysical sources contribute the majority of the $\gamma$-ray flux measured in experiments on Earth and represent an enormous background in the search for the DM signal. In order to minimize the astrophysical background in the search for DM signal, the $\gamma$-ray flux coming from different astrophysical components has to be carefully modelled and subtracted from the overall measured $\gamma$-ray flux as it is done in [2]. To do so, the expected $\gamma$-ray emissions coming from the Galactic Center (GC), the Galactic Plane (GP), active galaxies, and galaxy clusters are individually modelled and subtracted, and a residual emission which remains after subtracting these models from the measured $\gamma$-ray emission is called *the extragalactic diffuse $\gamma$-ray background* (EDGB) [3]. It represents the sum of the $\gamma$-ray flux coming from DM annihilation or decay, and the $\gamma$-rays emitted by the sub-threshold astrophysical sources. The latter represents the emission coming from astrophysical sources which produce a flux of $\gamma$-rays which is too small to be identified as coming from a point source. The flux coming from such sources is therefore interpreted as an astrophysical diffuse $\gamma$-ray background, showing a certain degree of anisotropy which can be analysed to give information about the origin of this emission component. The same goes for the emission coming from DM, which is also expected to produce a signature anisotropic pattern.

*Measuring the cross-correlation.* Measurements of the cross-correlation between different tracers of density distribution in the Universe have been performed [4], recently also by cross-correlating the EDGB and cosmic shear in Ref. [5], which was done using the Fermi-Large Area Telescope (LAT) [6] and Dark Energy Survey (DES) data. They have shown that the expected cross-correlation signal is indeed detected when cross-correlating the Fermi-LAT 9-year data with DES 1-year data. They report a cross-correlation signal peaking at small scales which they attribute to the blazar emission, while also detecting a large-scale component which they investigate in terms of astrophysical sources and particle DM emission. This is an important result showing that the method can actually be used in further exploration in the field.

The cross-correlation method will become especially useful with future $\gamma$-ray telescopes and weak lensing surveys, such as Cherenkov telescope array [7] (CTA) and Euclid [8], which will provide us with even more data and upgraded measurements resolutions. Studying the results of the improved measurements in terms of cross-correlation will give us more detailed features in the measured cross-correlation signal and further enhance our capability to distinguish the DM signal in the EDGB from the astrophysical background.

An importance of this approach is that it will give great power to our capabilities to analyse the $\gamma$-ray data used to indirectly search for DM. In case the cross-correlation between the different matter density distribution tracers, such as cosmic shear or galaxy distributions, and the measured $\gamma$-ray anisotropies can be detected, this is will be an important results in the DM searches and cosmology in large.

The paper is structured as follows. Section 2 describes the general formalism of the cross-correlation technique. Section 3 focuses on the window functions of the different signal and background components, in particular for the annihilating and decaying DM, the different classes of astrophysical sources, and the cosmic shear. In section 4 the three-dimensional power spectra defining the correlation between the two different $\gamma$-ray emissions is described: the DM emission, and the astrophysical emission, with the cosmic shear emission. In section 5 the predicted cross-correlation signal between the $\gamma$-ray emission components and the cosmic shear is calculated, and the meaning of these results is explained. Section 6 contains the conclusions and a brief summary of the results.

## 2. The cross-correlation formalism

A certain level of cross-correlation signal between the anisotropies seen in the EDGB and the weak lensing catalogs should be detected since both maps relate to the DM distribution in the Universe. The EDGB relates to the DM distribution through the component of the $\gamma$-ray emission which comes from the DM decay and annihilation, while the weak lensing relates to the DM distribution as it is a gravitational tracer of matter in the Universe dominated by DM. After detecting the cross-correlation signal, the signal has to be further analysed to evaluate if its features and contributions at different angular scales could be due to the presence of the DM component in the EDGB, or they can be explained solely by the astrophysical component.

The motivation to look for the cross-correlation signal between the extragalactic gamma-ray emission and the cosmic shear is that the detection of such a signal would show two important things: first, it would provide evidence that the measured weak lensing effects are a consequence of DM distribution in the Universe and not a consequence of some alternative theory of gravity, and second, it would help disentangle the DM signal from the astrophysical background coming from unresolved astrophysical sources. To see how the measured cross-correlation signal can be evaluated for the case of different $\gamma$-ray emission components and the cosmic shear, let us take a look at the corresponding formalism used in this approach, as it is described in Refs. [9] (from here after Camera, *et al.* (2015)) and [10].

For some density component with density field $g(\chi, \vec{n})$, we describe the measured intensity of the field $I_g(\vec{n})$ coming from the direction $\vec{n}$ in the sky as

$$I_g(\vec{n}) = \int d\chi \, g(\chi, \vec{n}) \, \widetilde{W}(\chi), \tag{1}$$

where $\widetilde{W}(\chi)$ represents the window function for this density field, and $\chi$ is the comoving distance at redshift $z$. We can define the normalized window function as

$$W(\chi) = \langle g \rangle \, \widetilde{W}(\chi), \tag{2}$$

where the density field $g$ is averaged over the whole sky. Using this definition, we can calculate the average intensity over the whole sky simply as the integral of the normalized window function over the distance:

$$\langle I_g \rangle = \int d\chi \, W(\chi). \tag{3}$$

We define the fluctuations in the intensity field as the difference between the source intensity and the average intensity:

$$\delta I_g(\vec{n}) \equiv I_g(\vec{n}) - \langle I_g \rangle. \tag{4}$$

Any function defined on a sphere can be expanded in terms of spherical harmonics. This means that we can write the intensity fluctuations $\delta I_g$ as a sum of spherical harmonics

$$\delta I_g(\vec{n}) = \langle I_g \rangle \sum_{lm} a_{lm} \, Y_{lm}(\vec{n}) \tag{5}$$

with dimensionless coefficients $a_{lm}$ calculated as

$$a_{lm} = \frac{1}{\langle I_g \rangle} \int d\vec{n} \, \delta I_g(\vec{n}) \, Y^*_{lm}(\vec{n}) = \frac{1}{\langle I_g \rangle} \int d\vec{n} \, d\chi \, f_g(\chi, \boldsymbol{r}) \, W(\chi) \, Y^*_{lm}(\vec{n}), \tag{6}$$

where the function $f_g$ is defined as

$$f_g \equiv \frac{g}{\langle g \rangle} - 1. \tag{7}$$

Using the above mentioned dimensionless coefficient of the expansion $a_{lm}$, we can define the angular power spectrum (APS) as

$$C_l^{(ij)} = \langle \sum_m a_{lm}^{(i)} a_{lm}^{(j)*} \rangle, \tag{8}$$

where the indices $i$ and $j$ label the two components that are being cross-correlated, i.e. $i = \gamma$-rays and $j =$ cosmic shear. Defining the three-dimensional power spectrum (3DPS) $P^{ij}$ through the two-point correlation function (2PCF) of the density functions, $f_i$ and $f_j$,

$$\langle \hat{f}_{g_i}(\chi, \vec{k}) \hat{f}^*_{g_j}(\chi', \vec{k}') \rangle = (2\pi)^3 \delta^3(\vec{k} - \vec{k}') P^{ij}(k, \chi, \chi'), \tag{9}$$

and assuming that the Limber approximation holds for such power spectrum, we obtain the following expression for the APS,

$$C_l^{(ij)} = \frac{1}{\langle I_i \rangle \langle I_j \rangle} \int \frac{d\chi}{\chi^2} W_i(\chi) W_j(\chi) P_{ij}(k = l/\chi, \chi), \tag{10}$$

where the 3DPS can be written in the form of the decomposition into the 1-halo and 2-halo terms:

$$P^{ij}(k) = P^{ij}_{1h}(k) + P^{ij}_{2h}(k) \tag{11}$$

where:

$$P^{ij}_{1h}(k) = \int dm \, \frac{dn}{dm} \hat{f}^*_i(k|m), \tag{12}$$

$$P^{ij}_{2h}(k) = \left[ \int dm_1 \, \frac{dn}{dm_1} b_i(m_1) \hat{f}^*_i(k|m_1) \right] \times \left[ \int dm_2 \, \frac{dn}{dm_2} b_i(m_2) \hat{f}^*_i(k|m_2) \right] P^{\rm lin}(k), \tag{13}$$

and $b_i(m)$ ($b_j(m)$) represent the linear bias between the component $i$ ($j$) and matter, and $P^{\rm lin}(k)$ is the linear power spectrum.

The measured APS between a $\gamma$-ray map and a cosmic shear map will be a superposition of APS of individual components in the $\gamma$-ray emission

($\gamma$) with the cosmic shear ($\kappa$). The $\gamma$-ray emission components, which we will include in our further discussion are DM, blazars (BLA), misaligned active galactic nuclei (mAGN), and star-forming galaxies (SFG):

$$C_l^{\gamma\,\kappa} = C_l^{\gamma_{\rm DM}\,\kappa} + C_l^{\gamma_{\rm BLA}\,\kappa} + C_l^{\gamma_{\rm mAGN}\,\kappa} + C_l^{\gamma_{\rm SFG}\,\kappa}. \tag{14}$$

The 1- and 2-halo terms of the 3DPS are different for various combinations of cross-correlated components, and are discussed in detail in section 4. In the next chapter we will take a look at the forms that the window functions take in case of the different $\gamma$-ray emission components and the cosmic shear.

## 3. Window functions of the $\gamma$-ray emission components and cosmic shear

The window functions, also called weight or selection functions, contain information about the geometry of the survey and the used analysis technique. In the case of the decaying particle DM, the window function is

$$W^{\gamma_{DMd}}(E,z) = \frac{1}{4\pi}\frac{\Omega_{DM}\rho_c}{m_\chi \tau_d}\frac{dN_d[E(1+z)]}{dE}\, e^{-\tau[E(1+z),z]}, \tag{15}$$

while the annihilating DM has a window function

$$W^{\gamma_{DMa}}(E,z) = \frac{(\Omega_{DM}\rho_c)^2}{4\pi}\frac{\langle\sigma_a v\rangle}{2m_\chi^2}(1+z)^3\Delta^2(z)\frac{dN_a[E(1+z)]}{dE}\, e^{-\tau[E(1+z),z]}, \tag{16}$$

where the $\Delta^2$ function is called the clumping factor and it gives information about the clustering of DM in the Universe,

$$\Delta^2(z) = \frac{\langle\rho^2\rangle}{\bar{\rho}^2} = \int_{m_{\rm min}}^{m_{\rm max}} dm \frac{dn}{dm}\int d^3\vec{x}\,\frac{\rho^2(\vec{x}|m)}{\bar{\rho}^2}. \tag{17}$$

In the above equations, $\Omega_{DM}$ represents the DM density parameter, $\rho_c$ is the critical density, $m_\chi$ and $\tau_d$ are the mass and the decay lifetime of DM particles, respectively. $dN_d/dE$ ($dN_a/dE$) is the number of photons emitted per single decay (annihilation) in the energy range ($E$, $E+dE$), $\tau$ is the optical depth for absorption, $\langle\sigma_a v\rangle$ is the annihilation cross-section averaged over velocity, and $\rho$ is the density distribution of DM in the Universe.

A significant uncertainty in the calculated window function for the annihilating DM is the amount of subhaloes hosted by the main haloes. This

can be accounted for with the inclusion of the so-called subhalo boost factor $B(M,\chi)$ in the last term of eq. 17, so that the last integral becomes

$$\int d^3\vec{x}\,\frac{\rho^2(\vec{x}|m)\,[1+B(M,\chi)]}{\bar{\rho}^2}.$$

Different contributions of subhaloes to the clumping factor $\Delta^2$ will give different overall window functions for the annihilating DM. In this paper we will use three different values for the boost factor, giving us a higher (HIGH), lower (LOW) or no-subhaloes (NS) window functions as defined in Camera, *et al.* (2015) [9].

In case of astrophysical sources, the window function is characterised in terms of the source luminosity $\mathcal{L}$, and it takes the form

$$W^{\gamma_i}(E,z) = \frac{A_i(\chi)\langle g_{\gamma_i}\rangle}{4\pi E_0^2}\left(\frac{E}{E_0}\right)^{-\alpha_i} e^{-\tau[E(1+z),z]}, \tag{18}$$

where $i$ stands for different astrophysical components, such as blazars, mAGNs or SFGs, the factor $A_i$ represents a normalisation constant, which depends on the definition of the luminosity, and $\langle g_{\gamma_i}(\chi)\rangle$ is the mean luminosity produced by an astrophysical component at distance $\chi$, such that

$$\langle g_{\gamma_i}(\chi)\rangle = \int_{\mathcal{L}_{\rm min}}^{\mathcal{L}_{\rm max}} d\mathcal{L}\,\mathcal{L}\,\rho_{\gamma_i}(\mathcal{L},z), \tag{19}$$

while $\rho_{\gamma_i}$ is the $\gamma$-ray luminosity function of the $i$ component.

As for the cosmic shear, the window function is of the form

$$W^{\kappa}(\chi) = \frac{3}{2}\,H_0^2\,\Omega_m\,[1+z(\chi)]\,\chi\int_{\chi}^{\infty} d\chi'\,\frac{\chi'-\chi}{\chi'}\,\frac{dN}{d\chi'}(\chi'), \tag{20}$$

where $\Omega_m$ is the matter density parameter, and $dN/d\chi$ represents the redshift distribution of background sources. The cosmic shear window function is normalised to unity, so that $\int d\chi\, W^{\kappa}(\chi) = 1$.

The left panel of Fig. 1 shows the redshift dependence of the window functions for the different boost models of the annihilating DM, the decaying DM, cosmic shear, while the right panel shows the redshift dependence of the window function for different classes of the astrophysical sources. In the case of the astrophysical sources two different models are taken into account, as described in Camera, *et al.* (2015) [9]. The window functions are integrated above $E > 1\,\mathrm{GeV}$ and normalised to the total EDGB intensity measured by the Fermi-LAT telescope above 1 GeV, so that the comparison

of different window functions is easier. As it is seen from the figure, the window functions that are best to use for cross-correlation with annihilating DM (left panel of Fig. 1 dashed blue line) are the two shear window functions, as the have the large values as small redshift $z$. The bigger the overlap of the two window functions used in the analysis, the higher the cross-correlation signal, which can give us information on the composition of the EDGB.

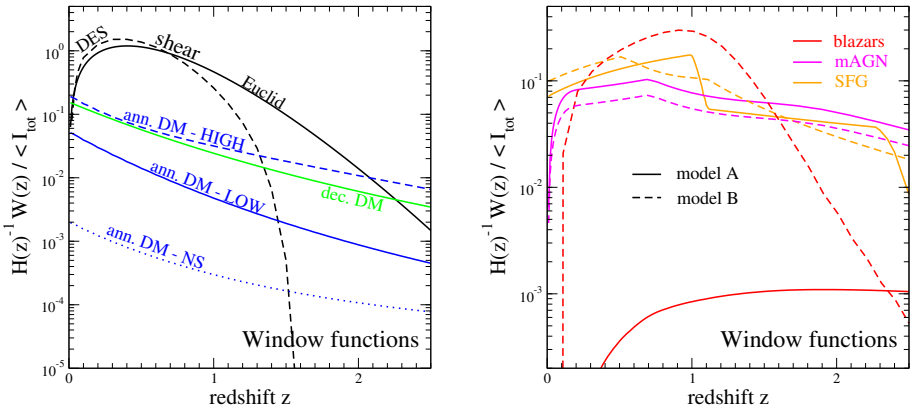


Figure 1: Left: The redshift dependence of the window functions for the different boost models of the annihilating DM, the decaying DM, and cosmic shear coming from two different weak-lensing surveys (DES and Euclid). Right: The redshift dependence of the window function for different classes of the astrophysical sources, for two different models describing the blazar emission. All the window functions are integrated above $E > 1\,\mathrm{GeV}$ and normalised to the total EDGB intensity measured by the Fermi-LAT telescope above 1 GeV, so that the comparison of different window functions is easier. Figure adopted from Camera, *et al.* (2015) [9].

## 4. Three-dimensional power spectra

The 3DPS is another factor in the APS defined in eq. (10). Let's take a look at the expressions used to calculate the 1- and 2-halo terms of the 3DPS. The 3DPS quantifies the correlation between two density fluctuations $\delta(\vec{x})$, where different components, such as the decaying and annihilating DM, astrohysical sources and cosmic shear scale differently with the density fields. Both the decaying DM and cosmic shear depend linearly on the DM density, while the annihilating DM scales with density squared $\delta^2$. For these components, the equations describing the 1- and 2-halo terms of the 3DPS

are therefore:

$$P_{1h}^{\delta\delta}(k,z) = \int_{M_{\rm min}}^{M_{\rm max}} dM\, \frac{dn}{dM}\, \tilde{v}^2(k|M), \tag{21}$$

$$P_{2h}^{\delta\delta}(k,z) = \left[\int_{M_{\rm min}}^{M_{\rm max}} dM\, \frac{dn}{dM}\, b_h(M)\, \tilde{v}(k|M)\right]^2 P_{\rm lin}(k,z) \tag{22}$$

for correlation between the cosmic shear and decaying DM, and

$$P_{1h}^{\delta\delta^2}(k,z) = \int_{M_{\rm min}}^{M_{\rm max}} dM\, \frac{dn}{dM}\, \tilde{v}^2(k|M)\, \frac{\tilde{u}^2(k|M)}{\Delta^2}, \tag{23}$$

$$P_{2h}^{\delta\delta^2}(k,z) = \left[\int_{M_{\rm min}}^{M_{\rm max}} dM\, \frac{dn}{dM}\, b_h(M)\, \tilde{v}(k|M)\right] \times \tag{24}$$

$$\times \left[\int_{M_{\rm min}}^{M_{\rm max}} dM\, \frac{dn}{dM}\, b_h(M)\, \frac{\tilde{u}^2(k|M)}{\Delta^2}\right] P_{\rm lin}(k,z) \tag{25}$$

for cross-correlation between cosmic shear and annihilating DM. In the above equations $\tilde{v}(k|M)$ is the Fourier transform of $\rho_{DM}(\vec{x}|M)/\bar{\rho}_{DM}$, and $\tilde{u}(k|M)$ is the Fourier transform of $\rho^2_{DM}(\vec{x}|M)/\bar{\rho}^2_{DM}$.

For the $\gamma$-rays induced by astrophysical sources, the equations are similar to those above, with a couple of differences. First, the quantity characterising the sources is no longer the mass $M$, but instead the luminosity $\mathcal{L}$, second, the halo mass function $dn/dM$ needs to be replaced by the $\gamma$-ray luminosity function, and third, each class of astrophysical sources $i$ is described by its own bias factor $b_{\gamma_i}(\mathcal{L},z)$. The equations defining the 1- and 2-halo terms of the correlation between $\gamma$-rays emitted by a class of astrophysical sources $i$ with another class of astrophysical sources $j$ (labelled $\gamma_i\gamma_j$), the decaying DM or cosmic shear (labelled $\gamma_i\delta$), and annihilating DM (labelled $\gamma_i\delta^2$) are, respectively:

$$P_{1h}^{\gamma_i\gamma_j}(k,z) = 0, \tag{26}$$

$$P_{2h}^{\gamma_i\gamma_j}(k,z) = \left[\int_{\mathcal{L}_{\rm min}(z)}^{\mathcal{L}_{\rm max}(z)} d\mathcal{L}\, \rho_{\gamma_i}(\mathcal{L},z) b_{\gamma_i}(\mathcal{L},z) \frac{\mathcal{L}}{\langle g_{\gamma_i}\rangle}\right] \times$$

$$\times \left[\int_{\tilde{\mathcal{L}}_{\rm min}(z)}^{\mathcal{L}_{\rm max}(z)} d\mathcal{L}\, \rho_{\gamma_j}(\mathcal{L},z) b_{\gamma_j}(\mathcal{L},z) \frac{\mathcal{L}}{\langle g_{\gamma_j}\rangle}\right], \tag{27}$$

$$P_{1h}^{\gamma_i\delta}(k,z) = \int_{\mathcal{L}_{\rm min}(z)}^{\mathcal{L}_{\rm max}(z)} d\mathcal{L}\,\rho_{\gamma_i}(\mathcal{L},z)\frac{\mathcal{L}}{\langle g_{\gamma_i}\rangle}\tilde{v}[k|M(\mathcal{L})], \tag{28}$$

$$P_{2h}^{\gamma_i\delta}(k,z) = \left[\int_{\mathcal{L}_{\rm min}(z)}^{\mathcal{L}_{\rm max}(z)} d\mathcal{L}\,\rho_{\gamma_i}(\mathcal{L},z) b_{\gamma_i}(\mathcal{L},z)\frac{\mathcal{L}}{\langle g_{\gamma_i}\rangle}\right]\times \\ \times\left[\int_{M_{\rm min}}^{M_{\rm max}} dM\,\frac{dn}{dM}\,b_h(M)\tilde{v}(k|M)\right]P_{\rm lin}(k,z), \tag{29}$$

$$P_{1h}^{\gamma_i\delta^2}(k,z) = \int_{\mathcal{L}_{\rm min}(z)}^{\mathcal{L}_{\rm max}(z)} d\mathcal{L}\,\rho_{\gamma_i}(\mathcal{L},z)\frac{\mathcal{L}}{\langle g_{\gamma_i}\rangle}\,\frac{\tilde{u}[k|M(\mathcal{L})]}{\Delta^2}, \tag{30}$$

$$P_{2h}^{\gamma_i\delta^2}(k,z) = \left[\int_{\mathcal{L}_{\rm min}(z)}^{\mathcal{L}_{\rm max}(z)} d\mathcal{L}\,\rho_{\gamma_i}(\mathcal{L},z) b_{\gamma_i}(\mathcal{L},z)\frac{\mathcal{L}}{\langle g_{\gamma_i}\rangle}\right]\times \\ \times\left[\int_{M_{\rm min}}^{M_{\rm max}} dM\,\frac{dn}{dM}\,b_h(M)\,\frac{\tilde{u}(k|M)}{\Delta^2}\right]P_{\rm lin}(k,z). \tag{31}$$

## 5. Cross-correlation of $\gamma$-ray emission with cosmic shear

Using the ingredients defined in the sections above, we can calculate the cross-correlation APS between the cosmic shear and the $\gamma$-ray emitters as defined in Eq. (10), where the total cross-correlation APS is the sum of contributions of different classes of sources, as described in Eq. (14).

The results for cross-correlation of $\gamma$-ray emitters with the weak-lensing signal of Euclid survey are shown in Fig. 2. The curves are arbitrarily normalised so that their shapes are easier to compare. The cross-correlation APS between $\gamma$-rays coming from DM and cosmic shear are shown in the left panel, while the results for the cross-correlation of cosmic shear with different classes of astrophysical sources are shown in the right panel. The results show, that the shape of the cross-correlation signal in case of annihilating DM does not depend much on the chosen subhalo boost model, which is seen in the fact that all blue lines in the left panel of Fig. 2 (denoting the results for the annihilating DM) show approximately the same shape. On the right panel of that same figure, the cross-correlation signal between astrophysical $\gamma$-rays and cosmic shear shows a similar shape for blazars, mAGNs and SFGs, although the uncertainties in this case are most significant for the first two.

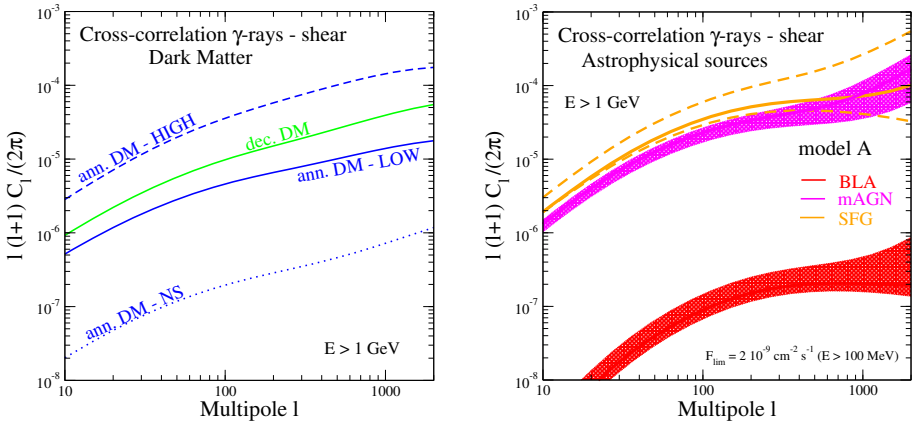


Figure 2: The results for cross-correlation of $\gamma$-ray emitters with the weak-lensing signal of Euclid survey. The curves are arbitrarily normalised so that their shapes are easier to compare. Left: The cross-correlation APS between $\gamma$-rays coming from DM and cosmic shear. The APS for the annihilating DM using three different boost models are shown in blue, while the results for the decaying DM are shown in green. The results show, that the shape of the cross-correlation signal in case of annihilating DM does not depend much on the chosen subhalo boost model. Right: The results for the cross-correlation of cosmic shear with different classes of astrophysical sources. The red line denotes the result of blazars, the purple for mAGNs, and the yellow for SFGs. Figure adopted from Camera, *et al.* (2015) [9].

What we can see from Eq. (10) is that the cross-correlation spectrum depends on the overlap of the window functions of the two signals which are being correlated, so that the different $\gamma$-ray emitting components will provide a more or less significant cross-correlation signal, as well as different shapes of the cross-correlation spectra. The amplitudes of the window functions are normalised to the total average intensity, so they are comparable in this analysis based on their shapes. As we can conclude from the window functions shown in Fig. 1, the weak-lensing window function and the different DM window functions all peaks at intermediate to low redshifts, which means that the cross-correlation between these components could provide a significant contribution to the measured cross-correlation signal, which could be detected in the experiments.

## 6. Conclusions

Cross-correlation with the cosmic shear maps is a tool used to analyse these anisotropy patterns in the EDGB, which helps us disentangle the DM signal from the astrophysical backgrounds. Different components of the as-

trophysical background will produce different shapes and amplitudes of the cross-correlation signal, when cross-correlated with the cosmic shear. The same applies to the decaying and annihilating DM. Different shapes and amplitudes of the cross-correlation signals can give us an insight into the contribution of each component to the overall EDGB. Therefore, the cross-correlation method represents an important analytical tool for extracting the DM signal from the astrophysical background component in the EDGB. As shown in the paper, the cross-correlation between the weak-lensing window function and the different DM window functions give further insight to how the cross-correlation between these components could provide a significant contribution to the measured cross-correlation signal, which could be detected in the experiments.

**Acknowledgements**
This research was supported by the Slovenian Research Agency's Young Researchers program.

**References**

[1] S. Camera, M. Fornasa, N. Fornengo, M. Regis, A novel approach in the weakly interacting massive particle quest: Cross-correlation of gamma-ray anisotropies and cosmic shear, The Astrophysical Journal 771 (1) (2013) L5. doi:10.1088/2041-8205/771/1/l5. 2

[2] M. Ackermann, M. Ajello, A. Albert, W. B. Atwood, L. Baldini, J. Ballet, G. Barbiellini, D. Bastieri, K. Bechtol, R. Bellazzini, et al., The spectrum of isotropic diffuse gamma-ray emission between 100 mev and 820 gev, The Astrophysical Journal 799 (1) (2015) 86. doi:10.1088/0004-637x/799/1/86. 3

[3] M. Fornasa, M. A. Sánchez-Conde, The nature of the diffuse gamma-ray background, Physics Reports 598 (2015) 1–58. doi:10.1016/j.physrep.2015.09.002. 3

[4] F. Calore, A. Cuoco, T. Regimbau, S. Sachdev, P. D. Serpico, Cross-correlating galaxy catalogs and gravitational waves: a tomographic approach (2020). arXiv:2002.02466. 3

[5] S. Ammazzalorso, D. Gruen, M. Regis, S. Camera, S. Ando, N. Fornengo, K. Bechtol, S. Bridle, A. Choi, T. Eifler, et al., Detection of cross-correlation between gravitational lensing and gamma-rays, Physical Review Letters 124 (10). doi:10.1103/physrevlett.124.101102. 3

[6] S. Abdollahi, F. Acero, M. Ackermann, M. Ajello, W. B. Atwood, M. Axelsson, L. Baldini, J. Ballet, G. Barbiellini, D. Bastieri, et al., Fermi large area telescope fourth source catalog, The Astrophysical Journal Supplement Series 247 (1) (2020) 33. doi:10.3847/1538-4365/ab6bcb. 3

[7] D. Mazin, The cherenkov telescope array (2019). arXiv:1907.08530. 4

[8] L. Amendola, S. Appleby, D. Bacon, T. Baker, M. Baldi, N. Bartolo, A. Blanchard, C. Bonvin, S. Borgani, et al., Cosmology and fundamental physics with the euclid satellite, Living Reviews in Relativity 16 (1). doi:10.12942/lrr-2013-6. 4

[9] S. Camera, M. Fornasa, N. Fornengo, M. Regis, Tomographic-spectral approach for dark matter detection in the cross-correlation between cosmic shear and diffuse gamma-ray emission, Journal of Cosmology and Astroparticle Physics 2015 (06) (2015) 029–029. doi:10.1088/1475-7516/2015/06/029. 5, 8, 9, 12

[10] N. Fornengo, M. Regis, Particle dark matter searches in the anisotropic sky, Frontiers in Physics 2. doi:10.3389/fphy.2014.00006. 5

# Bound debris and mass fallback rate in stellar tidal disruption events

Taj Jankovič

*University of Nova Gorica, Vipavska 13, SI-5000 Nova Gorica*

**Abstract**

Highly energetic stellar tidal disruption events (TDEs) provide a way to study black hole characteristics and their environment. A key issue in understanding TDEs, disk formation and their light curves, is the amount of matter, which either falls in the black hole, forms a disk, or escapes in the interstellar space. We simulate TDEs with the `PHANTOM` code in generalized Newtonian potential and research, how the amount of bound debris and mass fallback rate evolve with time for stars initially on orbits with different eccentricities and penetration factors. We find, that in the case of hyperbolic orbits the amount of bound debris is decreasing with penetration factor $\beta$. In the case of elliptic orbits, we derive an increase of bound debris with the value of $\beta$. For parabolic orbits the amount of bound debris stays close to 50%. Mass fallback rate reaches its peak faster for orbits, where the distance to the pericenter in smaller during the first passage. The peak value of the fallback rate can increase or decrease with the value of the penetration factor, depending on the initial eccentricity of the orbit.

*Keywords:* tidal disruption event, black hole, smoothed particle hydrodynamics, generalized Newtonian potential, bound debris, fallback rate

## 1. Introduction

The majority of galactic centers contain a supermassive black hole (mass $10^5$–$10^{9.5}\,\mathrm{M}_\odot$), surrounded by a central stellar cluster. Their high gravitational potential makes it impossible for even photons to escape, which is why different methods for detection are used. There is a probability of $10^{-5}$–$10^{-4}$

*Email address:* `taj.jankovic@ung.si` (Taj Jankovič)

/galaxy/year that a star from such a cluster is scattered and brought in the proximity of the black hole, where it is disrupted by black hole's tidal force [1, 2]. The fate of the stellar debris depends on its total energy. Parts of the debris with positive energy are unbound and escape from the gravitational potential of the black hole. On the other hand, debris with negative total energy is bound and returns in the black hole's vicinity, where it can form an accretion disk, which may emit radiation for months to years.

In a TDE a part of debris is unbound and the other bound (or accreted). Since the paper by Martin Rees in 1988, which made a crude estimate that half of the star's material is on bound and half on unbound orbits, the exact ratio is speculated to be determined by initial orbital parameters [3]. We numerically model tidal disruption events, in order to ascertain this theory. Simulations of tidal disruption events were done with the `PHANTOM` code, which is based on the smoothed particle hydrodynamics method (SPH). In the calculations of the supermassive black hole's gravitational field, we took into account relativistic corrections to the Newtonian potential. We carried out an analysis of the stellar debris after the disruption. We focused on the temporal evolution of the amount of the bound debris and the debris fallback rate. Computations were performed for different types of trajectories (elliptic, parabolic, hyperbolic) and for different values of the penetration factor $\beta$.

## 2. Methods

Simulations of TDEs were performed with the `PHANTOM` code [4]. `PHANTOM` is widely accepted as software for TDE studies. It has been used to study accretion disk formation as well as debris fallback rate for different stellar structures [5, 6, 7, 8]. Here we adopt a similar procedure as in [5]. The star was created as an object with $R_\star = 1\,\mathrm{R}_\odot$ and $M_\star = 1\,\mathrm{M}_\odot$ composed out of $103,806$ particles (determined in `Phantom` to closely match the appropriate equation of state), with the polytropic equation of state ($\gamma = 5/3$). The star was put in a point on orbit with the common center of mass corresponding to values of coordinates $x = 1000\,\mathrm{R}_\odot$ and $z = 0$. Value of $y$ coordinate was calculated from the equation of orbit, which depends on the eccentricity $e$: elliptic ($e = 0.99$), hyperbolic ($e = 1.01$), or parabolic ($e = 1.0$). A thing of note is that we use a value of Courant force parameter $C_{\mathrm{force}} = 10^{-9}$ in order to reach a better convergence of hydrodynamical quantities. Courant force parameter controls the simulation timestep during calculations of external forces on a particle.

We model a supermassive black hole as a non-rotating body with mass $M_{\mathrm{BH}} = 10^6\,\mathrm{M}_\odot$, located in the origin, with the value of accretion radius $R_{\mathrm{acc}} = 4R_{\mathrm{g}}$, where $R_{\mathrm{g}} = GM_{\mathrm{BH}}/c^2$ is the gravitational radius. $G$ and $c$ are the gravitional constant and the speed of light, respectively. The SMBH's external field includes general relativistic corrections — we adopt the generalized Newtonian potential [9]. Therefore, the specific potential energy of particles can be expressed as

$$\epsilon_{\mathrm{pot}} = -\frac{GM_{\mathrm{BH}}}{r} - \left(\frac{2R_{\mathrm{g}}}{r-2R_{\mathrm{g}}}\right)\left[\frac{r-R_{\mathrm{g}}}{r-2R_{\mathrm{g}}}\dot{r}^2 + \frac{r^2\dot{\theta}^2}{2}\right], \tag{1}$$

where $\theta$ is an angle measured in the orbital plane and $r$ is distance between the particle and the origin. This potential reproduces several relativistic properties with higher precision than other pseudo-Newtonian approximations [9]. Bound particles were considered as particles with negative specific total energy $\epsilon_{\mathrm{tot}}$ and particles, that fell into the SMBH

$$\epsilon < 0 \text{ and } r < 4R_{\mathrm{g}}. \tag{2}$$

Total energy was calculated as

$$\epsilon_{\mathrm{tot}} = \frac{1}{2}\left[\frac{r^2\dot{r}^2}{(r-2R_{\mathrm{g}})^2} + \frac{r^3\dot{\theta}^2}{r-2R_{\mathrm{g}}}\right] - \frac{GM_{\mathrm{BH}}}{r}. \tag{3}$$

In order to estimate the debris boundness, we introduce a new quantity

$$\mathcal{R} = \frac{M_{\mathrm{bound}}}{M_{\mathrm{tot}}}, \tag{4}$$

where $M_{\mathrm{bound}}$ is the mass of bound debris and $M_{\mathrm{tot}}$ the total mass. In our case $M_{\mathrm{tot}} = 1\,\mathrm{M}_\odot$. Furthemore, the critical eccentricity $e_{\mathrm{crit}}$ can be derived

$$e_{\mathrm{crit}} = 1 \pm 2\frac{R_\star}{r_{\mathrm{t}}\beta}, \tag{5}$$

where $r_{\mathrm{t}}$ is the tidal radius defined as $r_{\mathrm{t}} \equiv R_\star\left(\frac{M_{\mathrm{BH}}}{M_\star}\right)^{1/3}$ [10]. Parameter $\beta$ is defined as a ratio between the tidal radius and the distance to the pericenter $\beta \equiv r_{\mathrm{t}}/r_{\mathrm{p}}$. For $e_{\mathrm{crit}} > 1$ value of eccentricity, above which total debris is unbound, can be estimated. On the other hand, $e_{\mathrm{crit}} < 1$ gives an estimate of eccentricity below which every particle is bound.

*2.1. Simulations*

TDEs were simulated for orbits with eccentricities $e = 0.99,\ 1.0,\ 1.01$ and $\beta = 1,\ 2,\ 3,\ 4,\ 5$. Disruptions were simulated until 13.3 h. Tables 1, 2 and 3 show an overview of simulations with corresponding values for $e_{\text{crit}}$ (see Equation 5), the time of the first passage $t_1$ as well as quantity $\mathcal{R}$ (see Equation 4).

In the case of **parabolic orbits**, we notice, that higher $\beta$ results in shorter $t_1$ — star is closer the SMBH. The estimated percentage of bound material is 50% for every $\beta$. This follows from the total energy of the common center of mass for debris on parabolic orbit $\epsilon_{\text{tot}} = 0$ and assumption of an equal energy spread during the first encounter towards more negative and more positive values.

Table 1: Values of the first passage time $t_1$ and the amount of bound debris $\mathcal{R}$ in TDEs of a star initially on different parabolic orbits.

| **Orbit** | $t_1[h]$ | $\mathcal{R}[\%]$ |
|---|---|---|
| $e = 1.0,\ \beta = 1$ | 9.7 | 50 |
| $e = 1.0,\ \beta = 2$ | 8.1 | 50 |
| $e = 1.0,\ \beta = 3$ | 7.6 | 50 |
| $e = 1.0,\ \beta = 4$ | 7.4 | 50 |
| $e = 1.0,\ \beta = 5$ | 7.2 | 50 |

Table 2: Values of the first passage time $t_1$ as well as critical eccentricities $e_{\text{crit}}$ and the amount of bound debris $\mathcal{R}$ in TDEs of a star initially on different elliptic orbits.

| **Orbit** | $e_{\text{crit}}$ | $t_1[h]$ | $\mathcal{R}[\%]$ |
|---|---|---|---|
| $e = 0.99,\ \beta = 1$ | 0.980 | 9.8 | 75 |
| $e = 0.99,\ \beta = 2$ | 0.990 | 8.3 | 100 |
| $e = 0.99,\ \beta = 3$ | 0.993 | 7.9 | 100 |
| $e = 0.99,\ \beta = 4$ | 0.995 | 7.8 | 100 |
| $e = 0.99,\ \beta = 5$ | 0.996 | 7.8 | 100 |

A star with the common center of mass on an **elliptic orbit** has negative

total energy and is considered bound. However, tidal forces produce energy spread during the first passage. Under certain conditions, this stretch can be large enough, that in some parts of debris kinetic energy overcomes its potential energy. Table 2 shows, that only parameters of one simulated orbit (with $e = 0.99$ and $\beta = 1$) result in a value above $e_{\text{crit}}$—only in this case a part of debris is expected to become unbound. **Hyperbolic orbits** are characterized by positive total energy. The consequence is a lower amount of bound debris after tidal disruption—only orbit with $\beta = 1$ is below the critical eccentricity.

Table 3: Values of the first passage time $t_1$ as well as critical eccentricities $e_{\text{crit}}$ and the amount of bound debris $\mathcal{R}$ in TDEs of a star initially on different hyperbolic orbits.

| **Orbit** | $e_{\text{crit}}$ | $t_1[h]$ | $\mathcal{R}[\%]$ |
|---|---|---|---|
| $e = 1.01,\ \beta = 1$ | 1.020 | 9.7 | 25% |
| $e = 1.01,\ \beta = 2$ | 1.010 | 8.0 | 0% |
| $e = 1.01,\ \beta = 3$ | 1.007 | 7.4 | 0% |
| $e = 1.01,\ \beta = 4$ | 1.005 | 7.0 | 0% |
| $e = 1.01,\ \beta = 5$ | 1.004 | 6.8 | 0% |

### *2.2. Fallback rate*

Tidal disruptions are affected by initial orbital parameters. During the disruption of a star initially on elliptic more debris is bound in comparison to disruptions on parabolic or hyperbolic orbits. The consequence is a different evolution of debris, which can be seen in Figure 1. This also affects the observed lightcurve. However, there is currently no available method to accurately simulate realistic TDE curves.

A quantity, which can be related to the lightcurve, is the fallback rate of the debris $\dot{M}$ [3]. This was also confirmed observationally — at late times of a tidal disruption, the lightcurves of multiple TDEs decay as $\propto t^{-5/3}$ [11]. There were several studies, which analyzed the effect of the initial orbital parameters on $\dot{M}$ [12, 13]. They simulated TDEs in Newtonian and Kerr spacetime. We calculate $\dot{M}$ from the debris mass distribution over the total energy [12, 10]. We also compare it to the Eddington accretion

$$\dot{M_{\text{Edd}}} = \frac{4\pi G M_{\text{BH}} m_{\text{p}} c}{\eta c^2 \sigma_{\text{t}}}, \tag{6}$$

where $m_\mathrm{p}$ and $\sigma_\mathrm{t}$ are proton mass and Thompson cross-section, respectively. $\eta = 0.1$ is radiative efficiency. Large $\dot{M}$ would result in an increase in radiation pressure, which can delay or even stop accretion, if $\dot{M} > \dot{M}_\mathrm{Edd}$.

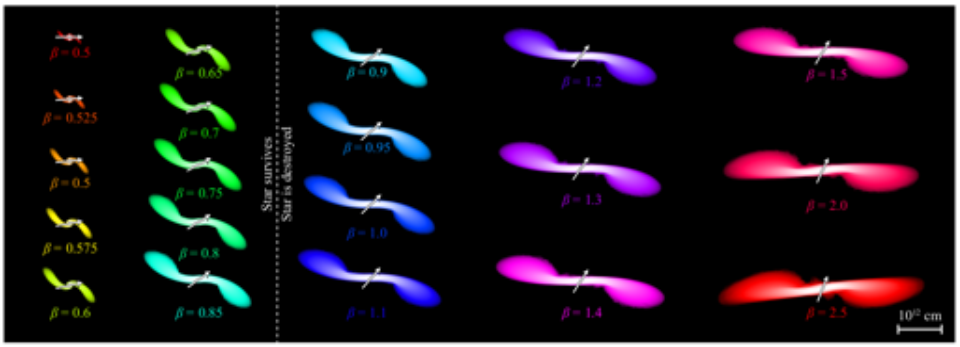


Figure 1: Snapshots of the density plot for debris configurations after a tidal disruption for stars initially on parabolic orbits with different $\beta$. White color corresponds to the maximum density and black corresponding to $10^{-6}\,\mathrm{g/cm^3}$. $\beta$ ranges from 0.5 (upper left corner) to 2.5 (bottom right corner). The white arrows indicate the direction towards the SMBH. The white dashed line separates simulations in which a core survives the encounter. Figure is adopted from [12].

## 3. Results

We calculate the percentage of bound particles $\mathcal{R}$ at different times (for all orbits) to evaluate the temporal dependence of boundness. Particles with negative energy or particles, which were accreted ($r < 4R_\mathrm{g}$), were considered as bound. Calculations of $\dot{M}$ were performed after the star has made the first passage and the particles start to move on ballistic orbits. We note, that the mass distribution of debris over total energy remains constant until the second passage — hence, orbits are ballistic.

### *3.1. Elliptic orbits*

At the beginning of simulations, particles on elliptic orbits are bound. When a star enters the tidal zone, SMBH's tidal forces begin stretching the star in an almost radial direction towards SMBH. This also affects energy distribution of stellar elements — energy of parts, which are closer to the SMBH, shifts to lower values, while the opposite happens to particles, that are farther away. Figure 2 shows debris mass distribution over total energy for different $\beta$ after disruption. The distribution is moved by a value equal to the average energy. In this way an easier comparison between different

orbits is possible. For smaller $\beta$ two prominent peaks are observed, which correspond to particles in the tidal head and tidal tail. For larger $\beta$ there is only one peak in the center of mass distribution. Moreover, the debris spans over a larger range in energy (also in space).

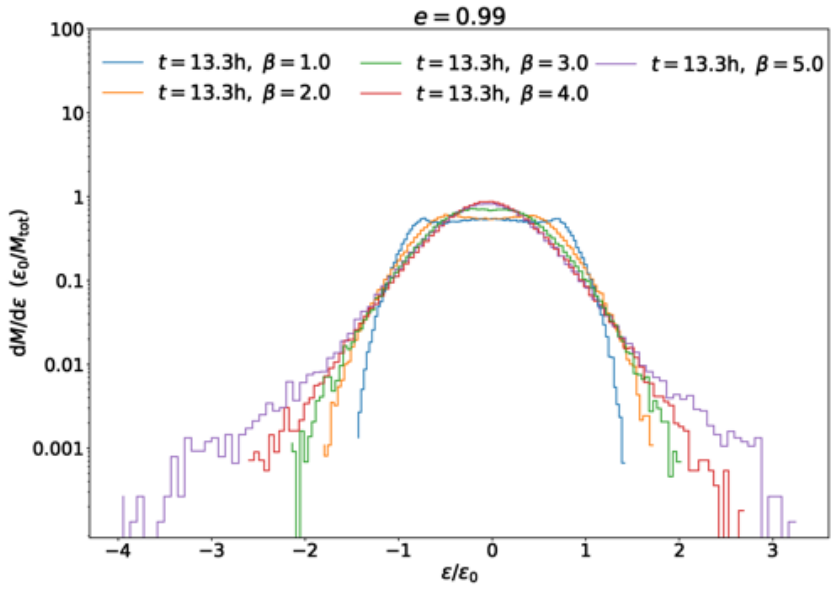


Figure 2: Distribution of debris mass over total energy for a star initially on different elliptic orbits. The distribution is moved, so that the center coincides with $\epsilon = 0$. Different colors correspond to orbits with different $\beta$. For larger $\beta$ the distribution is wider due to the effect of tidal forces.

The common center of mass for stars on elliptic orbits possesses negative energy. The consequence is, that initially all particles are bound (see Figure 3). However, after the first passage, an energy spread induced by tidal forces occurs. This change in more apparent for orbits with smaller $\beta$ — particles on these orbits have larger values of the apocenter distance and are less bound. Value of $\mathcal{R}$ decreases in certain cases. The most obvious change is for $\beta = 1$, where the value of $\mathcal{R}$ after the first passage is closed to the estimated value from Table 2. Debris fallback rate is illustrated in Figure 4. The peak of the mass return rate is higher for more bound orbits with larger $\beta$. On the other hand, smaller $\beta$ results in a slower decay of the $\dot{M}$, which is proportional to the power-law $t^{-5/3}$.

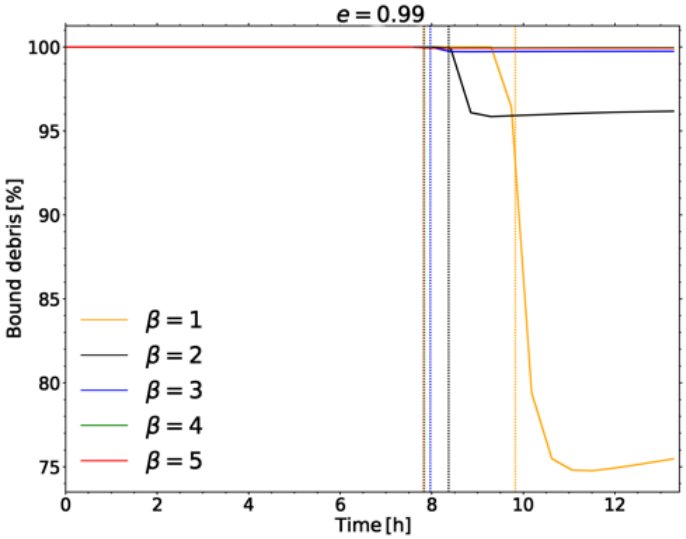


Figure 3: The percentage of the bound debris $\mathcal{R}$ at different times for elliptic orbits. Lines with the same colors correspond to an orbit with the same parameters. Vertical dotted lines represent times of first passages (see Table 2). $\mathcal{R}$ decreases after first passage due to an energy spread caused by tidal forces.

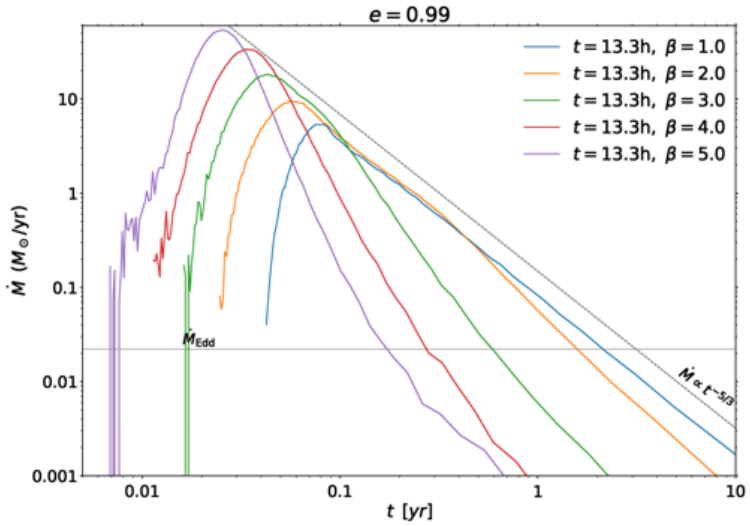


Figure 4: Mass fallback rate for stars initially on elliptic orbits. Values of Eddington accretion limit $M_{\rm Edd}$ and decay proporotional to $t^{-5/3}$ are shown. Different colors correspond to distruptions with different $\beta$. $\dot{M}$ reaches peak faster for orbits with larger $\beta$.

### *3.2. Parabolic orbits*

A star initially on a parabolic orbit has its common center of mass centered at $\epsilon_{\rm tot} = 0$. The effect of tidal forces during the first passage produces an equal energy spread in the directions of positive and negative energy. This effect can be seen in Figure 5. Note that the mass distribution over total energy is unshifted in this case, contrary to Figures 2 and 8.

Temporal evolution of boundness of debris can be seen in Figure 6. After disruption value in all cases remains close to 50% as expected from Table 1. Debris in disruptions with larger $\beta$ is deeper in the potential well of the SMBH. Therefore, it reaches higher velocities and has a more extended energy spread. The result is, that it makes the second passage faster, which also impacts debris fallback rate. Initially $\dot{M}$ is larger for orbits with larger $\beta$, which can be seen in Figure 7. However, $\dot{M}$ reaches larger values at late times for orbits with smaller $\beta$. At late times $\dot{M}$ is well described by $t^{-5/3}$.

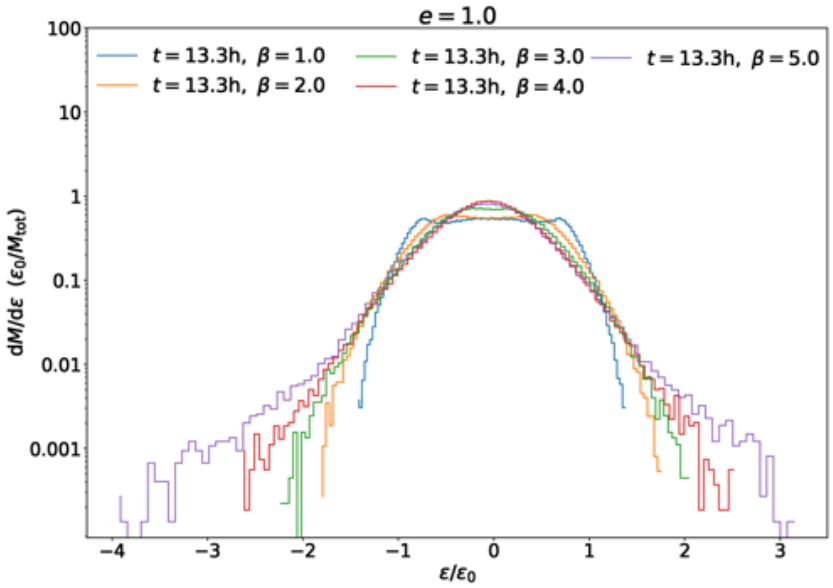


Figure 5: Distribution of debris mass over total energy for a star initially on different parabolic orbits. Different colors correspond to orbits with different $\beta$. For larger $\beta$ the distribution is wider due to the effect of tidal forces.

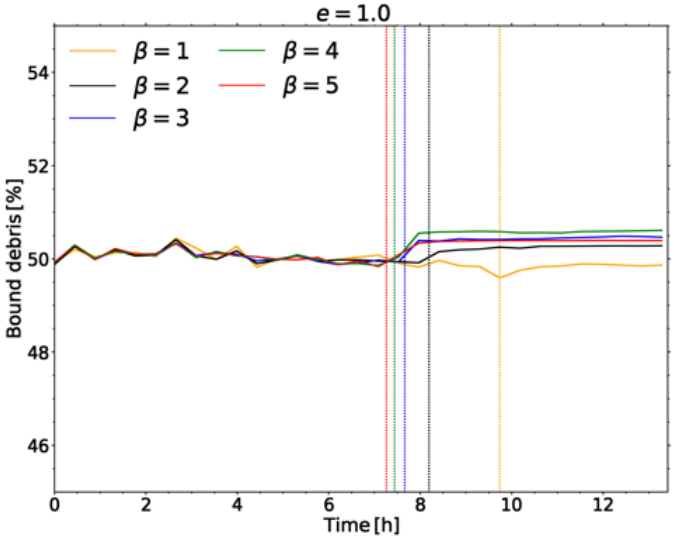


Figure 6: Percentage of bound debris $\mathcal{R}$ at different times for parabolic orbits. Lines with the same colors correspond to an orbit with the same parameters. Vertical dotted lines represent times of first passages (see Table 1). After the first passage $\mathcal{R}$ slightly increases due to numerical error. This increase is less than 1%.

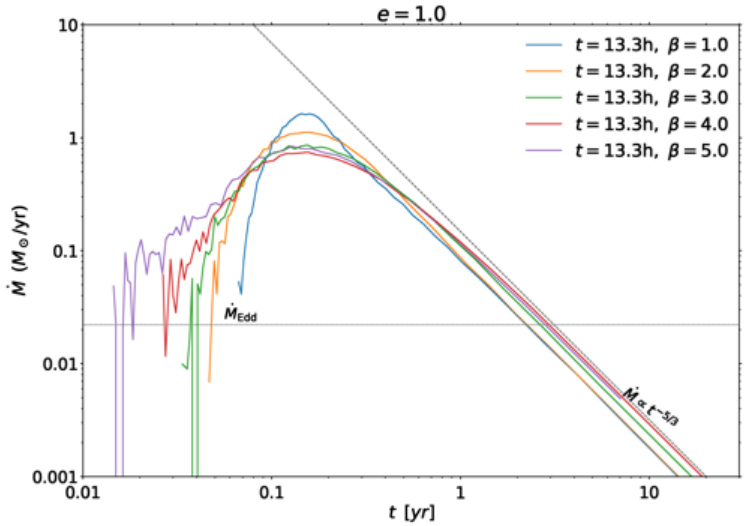


Figure 7: Mass fallback rate for stars initially on parabolic orbits. Values of Eddington accretion limit $M_{\rm Edd}$ and decay proporotional to $t^{-5/3}$ are shown. Different colors correspond to distruptions with different $\beta$. $\dot{M}$ is initially larger for orbits with larger $\beta$.

### 3.3. *Hyperbolic orbits*

Hyperbolic orbits do not have bound debris during the start of simulation (Figure 9). However, when the star starts to approach SMBH and the effect of tidal forces increases, parts of debris, which are closer to the SMBH, lose energy (become more bound) and vice versa. Similar as for parabolic and elliptic orbits, the energy spread is larger for orbits with larger $\beta$. We note, that the shape of distributions of debris mass over total energy is similar for the elliptic, parabolic and hyperbolic orbits (Figures 2, 5, 8). This behavior is a consequence of similar distances from the SMBH during the first passage — the distance between a SMBH and a star either on elliptic, parabolic or hyperbolic orbit with the same value of $\beta$ is similar.

After disruption, when debris starts to move out the tidal zone, the effect of tidal forces diminishes and $\epsilon_{\mathrm{kin}}$ dominates over potential energy. Boundness of orbit with $\beta = 1$ stabilizes at $\sim 22\%$, whereas values for other orbits are between $0 - 1.2\%$. Larger $\beta$ leads to lower $\mathcal{R}$. $\dot{M}$ reaches values above Eddington accretion only for orbit with $\beta = 1$. In other cases, the amount of debris is less than $0.01\,M_{\odot}/\mathrm{year}$.

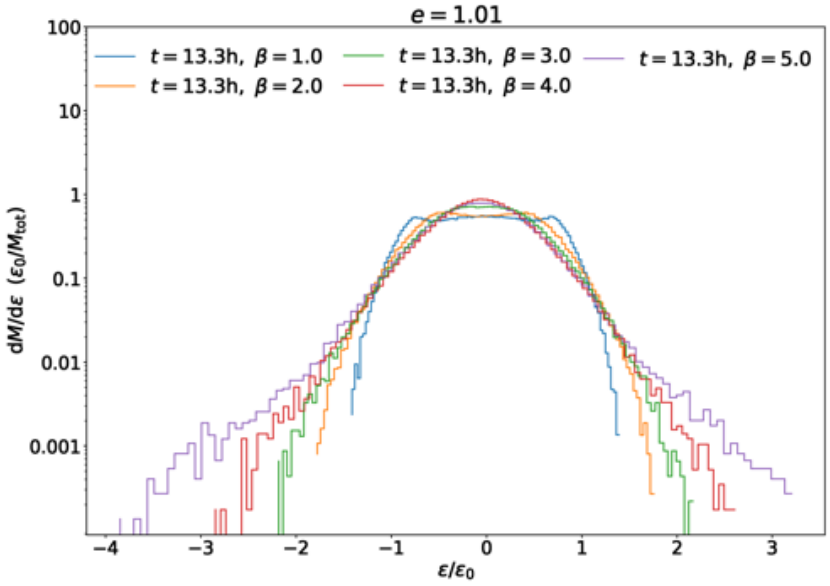


Figure 8: Distribution of debris mass over total energy for a star initially on different hyperbolic orbits. The distribution is moved, so that the center coincides with $\epsilon = 0$. Different colors correspond to orbits with different $\beta$. For larger $\beta$ the distribution is wider due to the effect of tidal forces.

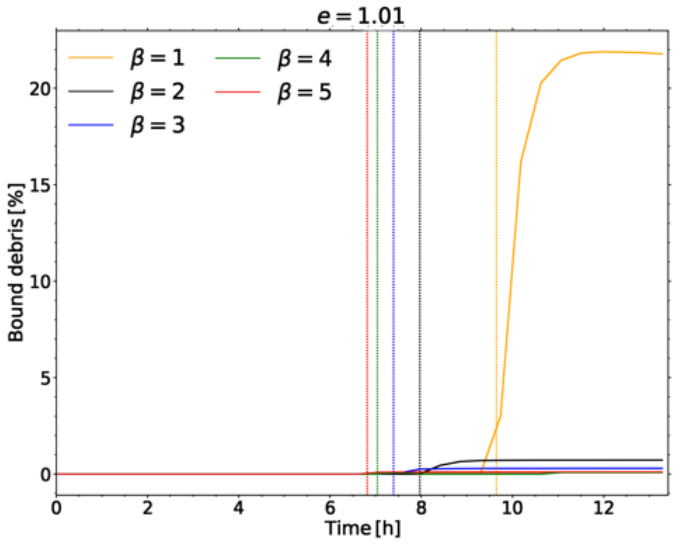


Figure 9: Percentage of bound debris $\mathcal{R}$ at different times for hyperbolic orbits. Lines with the same colors correspond to an orbit with the same parameters. Vertical dotted lines represent times of first passagse (see Table 3). $\mathcal{R}$ increases after first passage due to an energy spread caused by tidal forces.

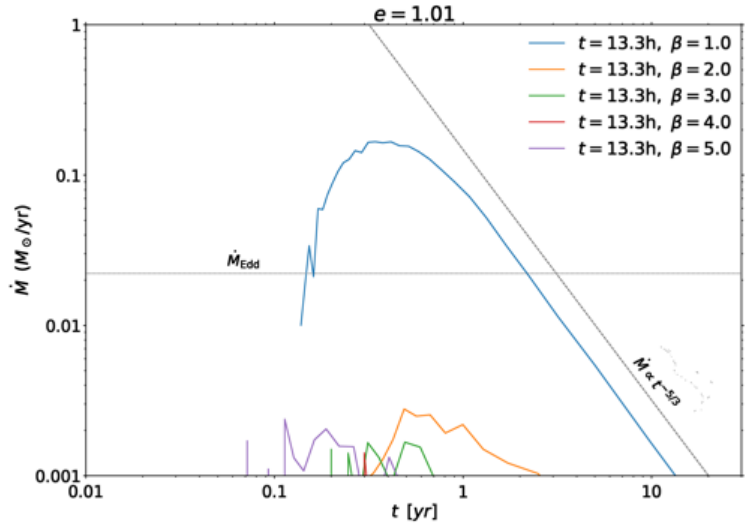


Figure 10: Mass fallback rate for stars initially on hyperbolic orbits. Values of Eddington accretion limit $M_{\rm Edd}$ and decay proporotional to $t^{-5/3}$ are shown. Different colors correspond to distruptions with different $\beta$. $\dot{M}$ reaches super-Eddington accretion rate only for $\beta = 1$.

## 4. Discussion and conclusion

By means of SPH simulations, we investigated, how the amount of bound debris changes after a tidal disruption for stars initially on different orbits. We also calculated the debris fallback rate. Our results confirm some of the previous findings, however, there are also discrepancies. We note, that oscillations for earlier times in $\dot{M}$ evolution are a consequence of oscillations in the tails of $\mathrm{d}M/\mathrm{d}\epsilon$. This problem could be alleviated by smoothing $\mathrm{d}M/\mathrm{d}\epsilon$ with a Gaussian filter or by logarithmically binning the data, and fitting the resulting histogram with a sensible curve [14, 13].

Our conclusion is, that TDEs of stars initially on **elliptic orbits**, lead to highly super-Eddington mass fallback rate $\dot{M}$. The peak of $\dot{M}$ increases for orbits with larger penetration factor $\beta$, which is in accordance with other authors [15]. Furthermore, our simulations suggest, that for weakly bound elliptic orbits (e.g. for $\beta = 1$), the amount of bound debris decreases below 100%. This confirms analytic conclusions from [10]. Disruptions of stars initially on **parabolic orbits** lead to $\dot{M}$, which is super-Eddington, but lower than for elliptic orbits. Contrary to [12] we find, that the peak value of $\dot{M}$ decreases with $\beta$ [12]. However, we emphasize, that this could be a consequence of an application of a different post-Newtonian external potential. Simulations in Kerr space-time by [16] reach a similar conclusion as us [16].

Particles on **hyperbolic orbits** have larger total energy. We confirm, that this leads to a lower percentage of bound matter. However, similar as in the case of elliptic orbits, the initial amount of bound matter can change during the first passage. The difference, in comparison to its elliptic counterparts is, that initially there is no bound matter. During the first passage, the energy spread from SMBH's tidal forces can cause a decrease in energy for debris. We conclude, that disruptions of stars on hyperbolic orbits with small values of $\beta$ can lead to flares, which could be observed [1]. We note, that an SPH resolution test was performed. However, the increase of particles up to $N = 10^7$ has not revealed the desired convergence. Therefore, we decreased the value of the Courant force parameter. This procedure has not been applied before and needs to be investigated into greater detail, which we leave for future work.

The simulations in this paper offer an insight into the TDE process for stars on different orbits — results offer a way to realize the consequences of different initial orbital parameters on the amount of bound debris and mass fallback rate. This work can be further expanded. A few possibilities include taking into consideration exact external relativistic gravitational

potential (either Schwarzschild or Kerr) or accurate initial stellar structure (e.g. generated with MESA code).

**Acknowledgements**

This research was supported by the Slovenian Research Agency grants P1-0031, I0-0033, J1-8136 and the Young Researchers program. We acknowledge the use of PHANTOM for detailed hydrodynamical simulations [4].

**References**

[1] N. C. Stone, The Tidal Disruption of Stars by Supermassive Black Holes, Springer International Publishing, 2015. doi:10.1007/978-3-319-12676-0. 2, 13

[2] T. Alexander, Stellar dynamics and stellar phenomena near a massive black hole, Annual Review of Astronomy and Astrophysics 55 (1) (2017) 17–57. doi:10.1146/annurev-astro-091916-055306. 2

[3] M. J. Rees, Tidal disruption of stars by black holes of $10^6$–$10^8$ Solar masses in nearby galaxies, Nature 333 (6173) (1988) 523–528. doi:10.1038/333523a0. 2, 5

[4] D. J. Price, et. al., Phantom: A smoothed particle hydrodynamics and magnetohydrodynamics code for astrophysics, Publications of the Astronomical Society of Australia 35 (2018). doi:10.1017/pasa.2018.25. 2, 14

[5] C. Bonnerot, E. M. Rossi, G. Lodato, D. J. Price, Disc formation from tidal disruptions of stars on eccentric orbits by Schwarzschild black holes, Monthly Notices of the Royal Astronomical Society 455 (2) (2015) 2253–2266. doi:10.1093/mnras/stv2411. 2

[6] D. Liptai, D. J. Price, I. Mandel, G. Lodato, Disc formation from tidal disruption of stars on eccentric orbits by Kerr black holes using general relativistic smoothed particle hydrodynamics (2019). arXiv:1910.10154. 2

[7] C. Bonnerot, W. Lu, Simulating disc formation in tidal disruption events, Monthly Notices of the Royal Astronomical Society (May 2020). doi:10.1093/mnras/staa1246. 2

[8] E. C. A. Golightly, C. J. Nixon, E. R. Coughlin, On the diversity of fallback rates from tidal disruption events with accurate stellar structure, The Astrophysical Journal 882 (2) (2019) L26. doi:10.3847/2041-8213/ab380d. 2

[9] E. Tejeda, S. Rosswog, An accurate Newtonian description of particle motion around a Schwarzschild black hole, Monthly Notices of the Royal Astronomical Society 433 (3) (2013) 1930–1940. doi:10.1093/mnras/stt853. 3

[10] K. Hayasaki, et. al., Classification of tidal disruption events based on stellar orbital properties, The Astrophysical Journal 855 (2) (2018) 129. doi:10.3847/1538-4357/aab0a5. 3, 5, 13

[11] S. van Velzen, et al., The first tidal disruption flare in ZTF: From photometric selection to multi-wavelength characterization, The Astrophysical Journal 872 (2) (2019) 198. doi:10.3847/1538-4357/aafe0c. 5

[12] J. Guillochon, E. Ramirez-Ruiz, Hydrodynamical simulations to determine the feeding rate of black holes by the tidal disruption of stars: The importance of the impact parameter and stellar structure, The Astrophysical Journal 767 (1) (2013) 25. doi:10.1088/0004-637x/767/1/25. 5, 6, 13

[13] E. Gafton, S. Rosswog, Tidal disruptions by rotating black holes: Effects of spin and impact parameter, Monthly Notices of the Royal Astronomical Society 487 (4) (2019) 4790–4808. doi:10.1093/mnras/stz1530. 5, 13

[14] J. Law-Smith, J. Guillochon, E. Ramirez-Ruiz, The tidal disruption of sun-like stars by massive black holes, The Astrophysical Journal 882 (2) (2019) L25. doi:10.3847/2041-8213/ab379a.
URL http://dx.doi.org/10.3847/2041-8213/ab379a 13

[15] K. Hayasaki, N. Stone, A. Loeb, Finite, intense accretion bursts from tidal disruption of stars on bound orbits, Monthly Notices of the Royal Astronomical Society 434 (2) (2013) 909–924. doi:10.1093/mnras/stt871. 13

[16] E. Tejeda, E. Gafton, S. Rosswog, J. C. Miller, Tidal disruptions by rotating black holes: Relativistic hydrodynamics with Newtonian codes,

Monthly Notices of the Royal Astronomical Society 469 (4) (2017) 4483–4503. doi:10.1093/mnras/stx1089. 13

# Raman lidar as a tool for probing atmospheric quality above the Cherenkov Telescope Array

Miha Živec

*Center for Astrophysics and Cosmology, University of Nova Gorica, 5270 Ajdovščina, Slovenia*

**Abstract**

Ground-based gamma-ray astronomy made a considerable progress with the adaptation of Imaging Atmospheric Cherenkov Telescopes (IACTs). A IACT is sensitive to Cherenkov light created by particles in gamma-ray induced air showers. Compared to other high energy detectors, the experiments that relay on air-shower measurements use atmosphere as a calorimeter. Even though IACTs are usually located at sites, where atmospheric conditions are very stable, local atmosphere should be continuously monitored, in terms of molecular density profiles, aerosol extinction profiles, and clouds. A next generation instrument, which will increase sensitivity of gamma-ray detection in 100 GeV to 10 TeV range for a factor of 5-10 is the Cherenkov Telescope Array (CTA). For measurements of atmospheric properties above the CTA a LIDAR (LIght Detection And Ranging) system will be used. Operating at 355 nm, where the maximum efficiency of CTA is reached, it will be able to detect fine structures in the lowest part of the atmosphere and provide input data for CTA telescope calibration. Performance goals of a CTA Raman LIDAR and its prototypes are presented in this paper.

*Keywords:* CTA, lidar, remote sensing, atmosphere

## 1. Introduction

Since the first detection of high energy photons from the Crab nebula with the Whipple IACT in 1989, the ground based very high-energy (VHE) gamma-ray astronomy has become a rapidly growing scientific discipline [1]. With the detection of more than 150 sources current arrays have demonstrated

*Email address:* miha.zivec@ung.si (Miha Živec)

huge observational potential of IACT astronomy. The CTA will explore in depth our Universe in VHE gamma rays and investigate cosmic processes leading to the emission of ultra-relativistic particles with unprecedented sensitivity. In energy range from 100 GeV to 10 TeV CTA performance will be 5-10 times better compared to existing experiments (Fig. 1). While this is the main CTA operational energy range, the observatory will uniformly cover energies from 20 GeV to 300 TeV. In addition to VHE astronomy, CTA will tackle some of major questions in astrophysics and fundamental physics. The studies will include search of origins of cosmic rays, exploration of the vicinity of neutron stars and black holes, and search for dark matter and other beyond Standard Model phenomena. To cover the whole sky, there will be two sites - one on the northern (La Palma, Spain) and one on the southern hemisphere (Paranal, Chile). The main site will be the southern site (CTA-South) composed of 4 Large Size Telescopes (LST), 25 Medium Size Telescopes (MST) and 70 Small Size Telescopes (SST) and built at an altitude of about 2100 m a.s.l. The northern site (CTA-North) will comprise 4 LSTs and 15 MSTs and will be positioned at the Roque de los Muchachos, of 2400 m a.s.l.

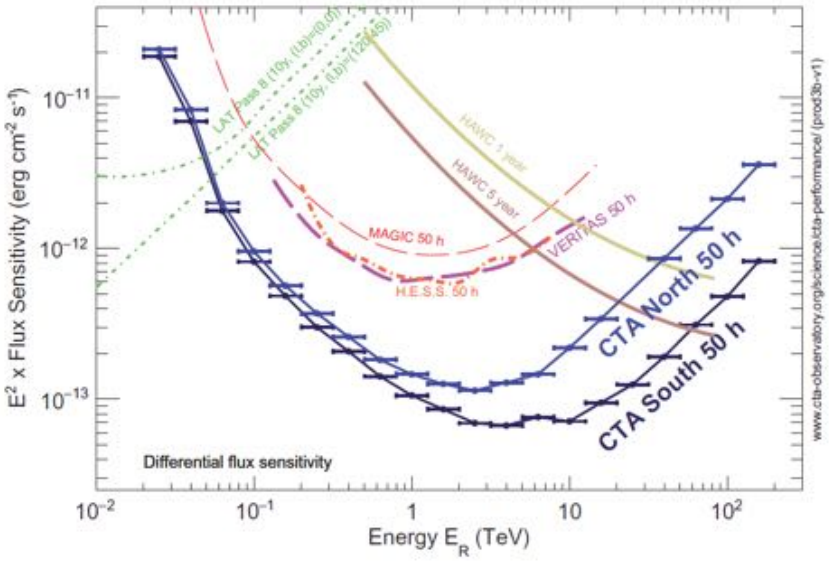


Figure 1: Comparison of the performance of CTA with Fermi-LAT, HAWC, MAGIC and H.E.S.S. The curves shown give only an indicative comparison of the sensitivity of the different instruments, as the method of calculation and the criteria applied are different. Figure adopted from [1].

When a very high energy photon enters the atmosphere, it collides with nuclei of atmospheric gases, forming a shower of secondary charged particles. Propagating with speeds greater than the speed of light in the atmosphere, they emit Cherenkov photons, which can be detected by IACTs. The main systematic uncertainty in the measurement of VHE cosmic gamma ray properties comes from the need to use the atmosphere as a calorimeter [2]. An atmospheric transmissivity depends on the height and wavelength at which Cherenkov photons are emitted. Due to different disturbances and changes in aerosol type and loading, this dependence also varies in time. The changes created by differences in atmospheric aerosol composition affect air shower starting point and its development. Photon loss is determined by absorption in the atmosphere and scattering out of IACTs field of view. Cloud coverage can partially or completely obstruct Cherenkov light emitted by air showers. Even though air-shower can penetrate through a cloud layer, the path through the cloud affects air-shower profile and its development (including shower maximum, penetration depth and energy profile) and reflects or absorbs Cherenkov light reducing yield detected by IACTs. Different atmospheric density profiles lead to variations in Cherenkov light flux of up to 60% [3]. To correctly calibrate IACT measurements, constant monitoring of aerosol density is required. For this the best suited tool is LIDAR working at the same wavelength as IACTs.

## 2. The atmosphere above the CTA

Most events that affect atmospheric optical properties occur in the troposphere and tropopause. These two layers are closest to the Earth surface, the later is reaching altitudes of 9 km above Polar Regions and 17 km above equator [4]. At CTA sites with approximate latitudes of 24° South and 28° North the troposphere thickness is about 15 km.

Atmospheric molecular content varies slowly, usually with time scales on the order of months, while aerosol concentrations change much quicker. Aerosols are most prominent in the lowest part of the troposphere, the planetary boundary layer (PBL), which experiences a regular daily cycle due to solar irradiation (Fig. 2). PBL, which can be several hundred meters thick, is directly influenced by interactions of the atmosphere with the Earth's surface [5, 6, 7]. The PBL is characterized by turbulent transport of moisture and aerosols. Larger aerosol density in the PBL strongly influences light travel due to absorption and scattering. Even though it almost disappears at night, it may leave some residual layer that interfere with

Cherenkov light [2]. In addition extraordinary events such as desert dust intrusions or salt and sea water droplets carried by strong winds, residual layers can change the light extinction in a matter of minutes. This low altitude events usually degrade transmission of all Cherenkov light, while vertical structures of aerosols and clouds only degrade part of the incoming light, making analysis much more complex. The third possible source of aerosol loading at CTA sites are stratospheric aerosols from volcanic eruptions, but these only influence Cherenkov light from exceptionally high shower productions, which are in any case on the energy threshold of IACTs.

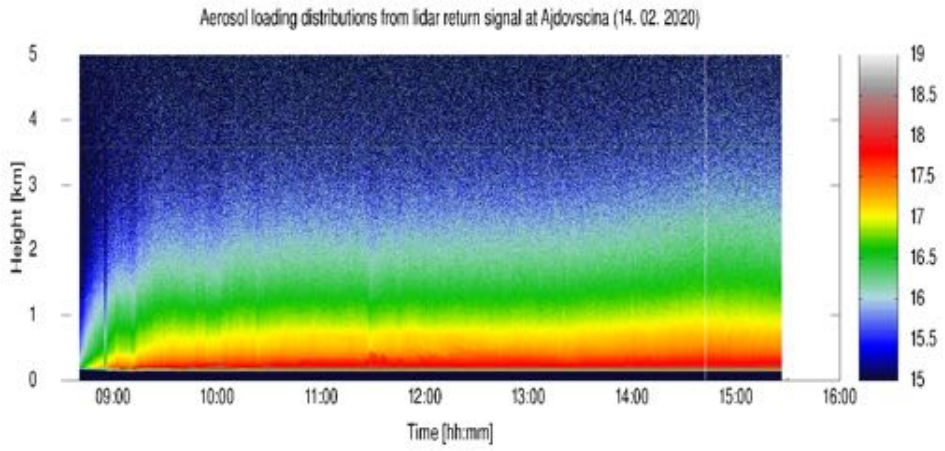


Figure 2: The formation of planetary boundary layer as seen with LIDAR above Ajdovščina, Slovenia on February 14 2020.

### *2.1. Atmospheric effects on light propagation*

The atmosphere is formed by gas and particles in suspension [6, 7]. When a beam of photons passes through this medium light scattering occurs and some portion of the beam is dissipated in all directions. The scattering process can either be elastic, where the scattered photons have the same energy as the incident ones, or inelastic, where energy exchange with the scatterers occur and photons energy changes.

Mie-Rayleigh scattering is elastic scattering on molecules and particles. If the scatterers are much smaller than the wavelength of the incoming light, we are talking about Rayleigh scattering. Mie scattering arises when the

scatterer is of the comparable size or bigger than the wavelength of the scattering light resulting in an interference pattern from waves that wrap themselves around and those that go through the particles. Intensity peaks are usually multiples of the wavelength, depending on the index of refraction of the scattering material. The original Mie scattering theory describes scattering on spherical particles, it can however be extended to non-spherical shapes.

Raman scattering is inelastic scattering with energy transfer between a photon and a molecule of atmospheric gas. During the transfer, the molecule changes its quantum state, usually through roto-vibrational excitation. The energy of the scattered photon can be decreased by Stokes Raman scattering, or can be increased through Anti-Stokes Raman scattering. The wavelength shift between the incident and Raman spectrum wavelength depends on the change in molecular energy level, and corresponds to a specific molecule type. Vibrational scattering has much larger intensity than rotational scattering, but it is still much weaker than Rayleigh scattering.

### 3. General requirements for a CTA Raman LIDAR

LIDAR is a remote sensing device that measures backscattering of emitted laser pulses on aerosols and molecules along its field of view [8]. This detected lidar signal is described with a single–scattering lidar equation as

$$P(r) = P_0 k \frac{c\tau_0}{2} A \frac{\beta(r)}{r^2} T^2(r), \tag{1}$$

where $P(r)$ is a received backscattered power from distance $r$, $P_0$ is a power transmitted by laser, $k$ is system efficiency, $c$ speed of light, $\tau_0$ pulse duration, $A$ effective area of the detector, $T(r)$ describes how much light gets lost while traveling to distance $r$ and back, $\beta(r)$ is a backscattering coefficient and is used to determine the strength of the backscattered signal.

LIDARs have proven to be a powerful tool for environmental studies [9]. Successful characterization of the atmosphere has been made at night using these systems, and in other fields of astronomy, the LIDAR technique has proven to be useful for the determination of the atmospheric extinction of starlight. Elastic LIDARs are the simplest LIDAR class, which only uses elastically backscattered light. The power measurement of this system depend on two unknown physical quantities (the total optical and backscattered coefficient), meaning an assumption is needed. As a result multiple assumptions need to be made, or boundary calibrations introduced,

limiting the uncertainty of the height-dependent atmospheric extinction to about 20%. Introducing a Raman channel to the system allows simultaneous and independent measurement of the extinction and backscattering coefficient. The assumptions are not needed anymore and the uncertainty of retrieved atmospheric extinction can be reduced below 5% if multiple Raman lines are used. An example of a range corrected signal in one elastic scattering channel, one Raman line and extinction coefficient belonging to Raman line are shown in Fig. 3.

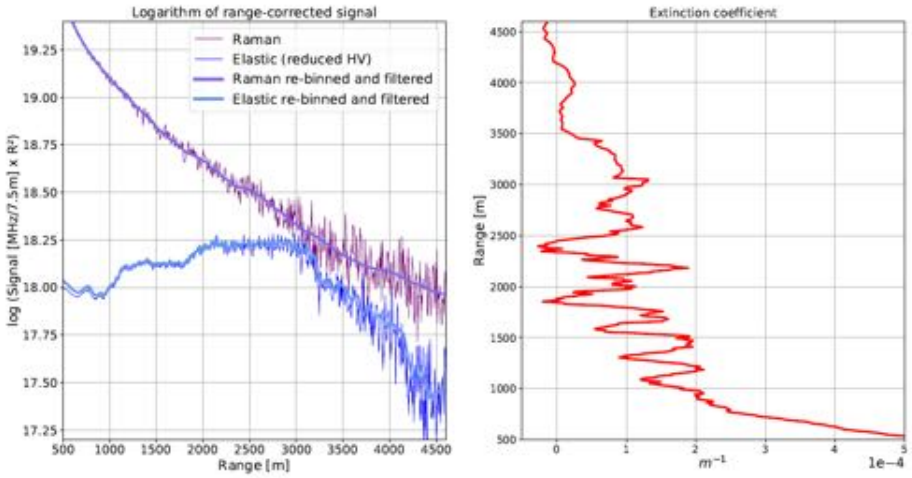


Figure 3: Left: Range-corrected signal from a test run with one Raman and one elastic channel. Right: Inverted Raman signal yields atmospheric extinction profile above Universitat Autònoma de Barcelona, Spain. Figure adopted from [10].

In the case of LIDAR usage at CTA, the device has to meet the requirements imposed by the observatory. To calibrate CTA for any atmospheric effects along its line of sight, the LIDAR must measure extinction profile up to the altitude of 15 km a.s.l. where extended air-showers are typically formed [10]. Since the CTA can operate down to the zenith angle of 60°, this means LIDAR range must be at least 30 km. To cover the typical air-shower range of sever kilometers, the LIDAR range resolution must be at least or better than ∼300 m. Early studies [9] have already shown that the aerosol transmission profiling is best determined with powerful Raman LIDARs. Such LIDARs should be equipped with near-range optics, in order to determine the full ground layer transmission reliably and such meet the aerosol transmission accuracy requirement. Moreover, stratospheric aerosol extinction should be accessible to the LIDAR, at least when pointing towards the

zenith.

The atmosphere must be characterized using at least two wavelengths present in the Cherenkov light spectrum, to be able to measure the transmittance with absolute accuracy of 0.03 [10]. A Nd:YAG laser, operating at second and third harmonic frequency, provides the capability to characterize atmosphere at 532 nm and 355 nm which cover measured Cherenkov light spectrum. Working at the CTA maximum sensitivity, laser light from the LIDAR would blind the CTA telescope, so the LIDAR data collection must be performed a few minutes before and after the CTA observation period, and during Wobble position. To achieve this, LIDAR must perform full characterization of an extinction profile during a minute long time interval. This can be performed with a short burst of about 500 laser pulses at a shooting frequency higher than 10 Hz. The problem of collecting large enough amount of light from this short duration measurements is solved by using bigger mirrors than those usually used in LIDARs for meteorological atmosphere monitoring.

There are three prototype designs for CTA LIDAR in progress [9]. One is led by CEILAP (Centro de Investigaciones Laser y sus Aplicaciones) as a part of Argentinian CTA collaboration, one by LUPM (Laboratoire Univers et Particules de Montpellier) in Montpellier France and the third one by IFAE/UAB (Institut de Fìsica d'Altes Energies / Universitat Autònoma de Barcelona) in Barcelona Spain, in which the University of Nova Gorica (UNG) is also involved. All three designs cover at least two elastic lines (355 nm and 532 nm) and two Raman lines (387 nm and 607 nm), with similar pulse strength ($\pm$25 mJ) and duration($\pm$2 ns), but the CEILAP design uses higher pulse frequency (50 Hz), than the European ones (20 Hz). The biggest difference in the design is that the European LIDARs use single 1.8 m mirror, while the one from Argentina uses six 40 cm mirrors. The composite mirror is cheaper and easier to maintain, but the alignment of the LIDAR becomes harder. Rest of the design data are summarized in Table 1.

## 4. Barcelona LIDAR

The Barcelona LIDAR design is the most promising prototype and was awarded the status of a CTAO Pathfinder by the CTAO Council. It is in the final test and adjustment stage before shipping to the Roque de los Muchachos Observatory at La Palma. After its placement near the LAT-1 telescope, the LIDAR will be tested in real conditions of the CTA-North

Table 1: Summary of specifications of all three CTA LIDAR prototypes. Data adopted from [9].

| | **CEILAP** | **IFAE/UAB** | **LUPM** |
|---|---|---|---|
| Housing | Custom-made container | CLUE container | CLUE container |
| Design | Multi-angle | Co-axial | Co-axial |
| Mirror diameter [cm] | 6×40 | 1×180 | 1×180 |
| Mirror f/D | 2.5 | 1 | 1 |
| Elastic 355 nm | yes | yes | yes |
| Raman 387 nm ($N_2$) | yes | yes | yes |
| Raman 408 nm ($H_2O$) | yes | no | no |
| Elastic 532 nm | yes | yes | yes |
| Raman 607 nm | yes | yes | yes |
| Elastic 1064 nm | yes | no | no |
| Laser | Continuum Inlite-II | Quantel Brilliant | Quantel CFR400 |
| Max. Laser power [mJ/p] | 125 @ 1064 nm | 100 @ 355, 532 nm | 90 |
| Pulse duration [ns] | 7-9 | 5 | 7 |
| Beam diameter [mm] | 6 | 6 | 7 |
| Beam divergence [mrad] | <0.75 | 0.5 | 3.5 |
| Pulse frequency [Hz] | 50 | 20 | 20 |
| Light transport | Optical fiber Edmund UV/VIS series | Liquid lightguide Lumatec Series 300 | Liquid lightguide Spectral Labs. |
| Fiber size [mm] | 1 | 8 | 8 |
| Fiber N.A. [mm] | 0.22 | 0.64 | 0.6 |
| Photon detector | PMT Hamamatsu H10721-110 | PMT/HPD Hamamatsu R11920-100 | PMT Hamamatsu R329P |
| Cathode diameter [inch] | 1 | 1.5 | 2 |
| Readout system | LICEL | LICEL | LICEL |

site for a year. During this time the performance of all LIDAR systems will be investigated and first trials of the LST-1 data calibration with LIDAR atmospheric data will be made.

The LIDAR is enclosed in 20 ft standard shipping container that protects the device itself and other instrumentation from environmental conditions. When LIDAR is in operation the container is opened by two linear actuators as shown in Fig. 4. Inside, has a custom made chassis on an alt-azimuth mount, that allows zenith and azimuth angle movements. Its 1.8 m diameter primary mirror is protected by four petals (Fig. 5). The primary mirror has an $f$-number of 1, and thus a focal length of 1.8 m [10]. It was originally produced for the CLUE experiment (Cherenkov Light Ultraviolet Experiment) in Padova [11] with a slumping technique invented at CERN [12, 13]. The mirror was made with float glass with a particularly smooth surface,

Figure 4: Barcelona Raman LIDAR in opened container at its site in Universitat Autònoma de Barcelona, Spain. The blue piston is one of actuators holding container halves in opened position.

specially produced in Venice, Italy[1] and the reflective coating of vacuum-evaporated 80 nm thick aluminum was performed in Asiago, Italy[2]. The roughness of the mirror is 2–3 nm with an initial reflectivity of 95%. After 4 years of operation, the mirror was re-coated at La Palma, Spain[3], because the reflectivity had degraded to 50-60%. In addition to the primary mirror, LIDAR system incorporates two guiding mirrors (Fig. 5e), that are used for guiding laser light in front of primary mirror (co-axial configuration) to reduce the complete overlap distance of the laser beam and the field of view. The dichroic guiding mirrors reflect light at an angle of $61.1 \pm 0.3^{\circ}$ and have very high reflectivity at the wavelengths 355 nm and 532 nm, while other wavelengths are suppressed. To further increase the dynamic range of the system, special near-range optics is used. While main LIDAR optics gives the access to the range greater than $\sim$150 m, the near-range optics covers

[1] Società Italiana Vetri of Porto Marghera
[2] Osservatorio Astrofísico of the University of Padova
[3] Herschel observatory

20 m to 200 m distance in the LIDARs field of view. For now the near-range optics works only in an elastics 532 nm line.

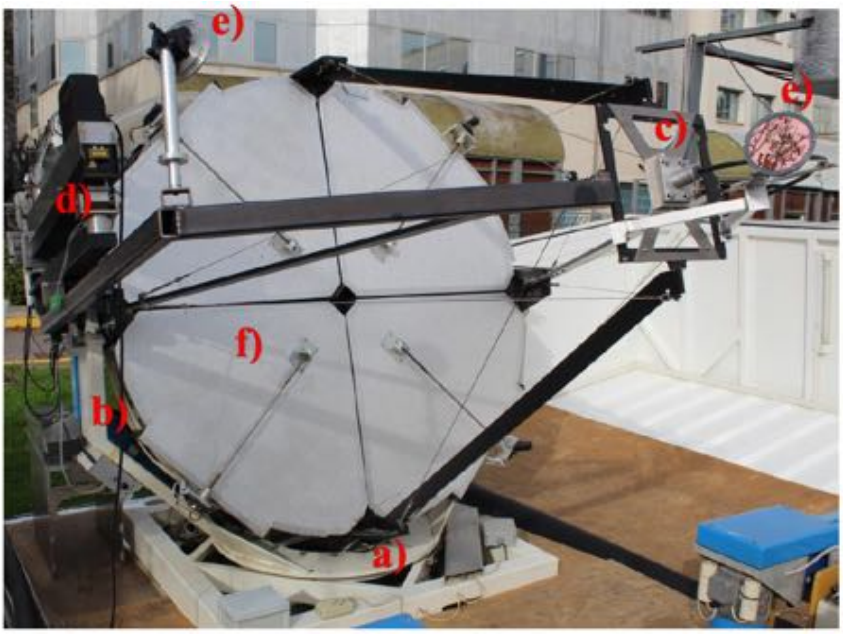


Figure 5: Parts of the chassis: the metal platform that support the azimuth movement (a), the U-form structure that supports the zenith movement (b), the focal plane support (c), the structure for the laser arm (d), the support for the small mirrors to align the laser beam (e) and the petals (f).

As a light source a QUANTEL Brilliant Nd:YAG 1064 nm is used. It is a pulsed laser with a pulse frequency 20 Hz and a ground wavelength of 1064 nm. For the LIDAR purpose second (532 nm) and third (355 nm) harmonics are used, operating at 160 mJ and 70 mJ energy per pulse respectively. The laser with both harmonic generators is placed in a compact module located on the laser arm (Fig. 5d) and moves together with the whole LIDAR system. Laser system incorporates cooling and heating unit, since the beam quality varies with temperature and must be held within optimal temperature range. Both units are attached to the system in such a way, that they move with whole chassis and do not limit LIDAR inclination angles.

To extract Mie and Raman signal, the backscattered light must be split into different wavelengths. This is done using a polychromator specially designed for the purpose of Raman LIDAR. It has four read-out channels,

where two of them are used for elastically backscattered light at wavelengths 355 nm and 532 nm, and two of them for Nitrogen Raman signal at wavelengths 387 nm and 607 nm. The design and a photograph of the equipped polychromator are shown in Fig. 6.

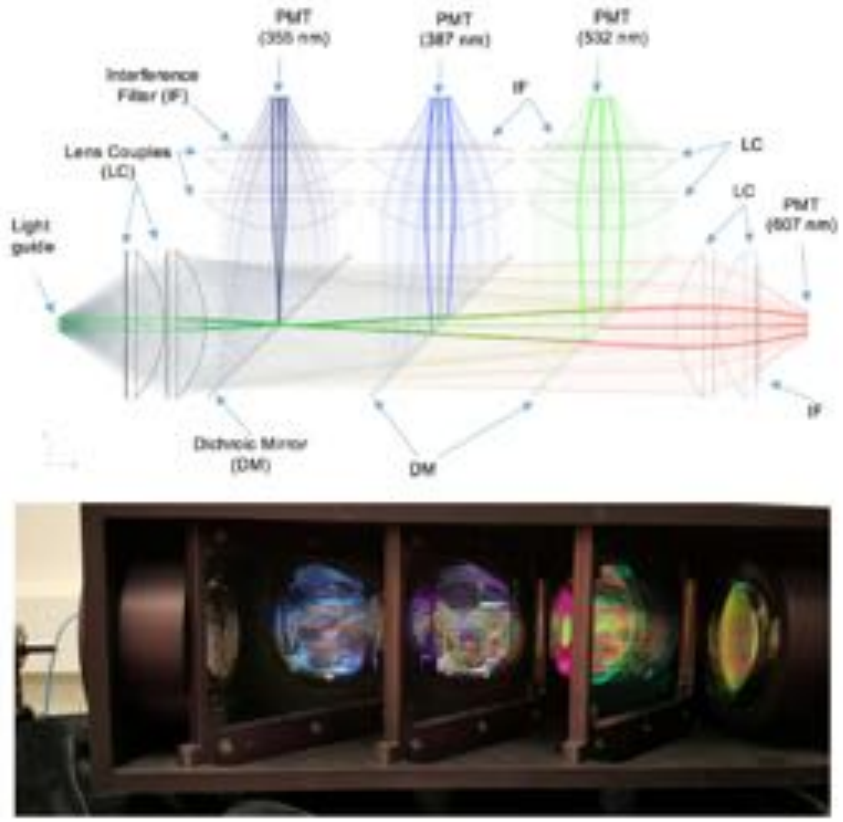


Figure 6: Top: the Zemax design of the Barcelona LIDAR polychromator. Bottom: A picture of the polychromator taken from the bottom side. After collimation by 100 mm Lens Couples, the incoming light is directed towards its respective detector with the help of three Dichroic Mirrors. For each channel, the light is again focused by Lens Couples and the noise is reduced by 10 nm wide Interference Filters. Each of the four wavelengths is then collected by a 1.5 in high-quantum-efficiency Hamamatsu R11920 photomultiplier tubes, which are of the same type same as those used in the Large-Sized-Telescope camera [10].

### *4.1. First results*

The Barcelona LIDAR has seen its first light in July 2018 [14]. Backscattering signals were collected in both elastic channels and a Raman 387 nm channel. When using fully powered laser, the photomultipliers (PMTs), with an

applied voltage of 1500 V, collected signal from a 25 km range, which can clearly be extended even further (see Fig. 7 left). The data taking was performed in less than a minute, with 500 laser shots at 10 Hz pulse rate. In addition to this full range scan, also test with reduced PMT gain, to avoid saturation of elastic lines, was performed (see Fig. 7 right) to show the capabilities of near-range detection, as later implemented into near-range optics.

After first tests, several minor optimizations and modifications were applied to the system. Now it also operates at second Raman (607 nm) channel and has separate near range channel at 532 nm line. Before relocating the LIDAR system to La Palma, the new gating system for elastic PMTs will be connected in order to prevent saturation of the signal in the near-range of the main optics. In addition the primary mirror will be recoated to increase reflectivity, which will even further extend the maximum range of detection.

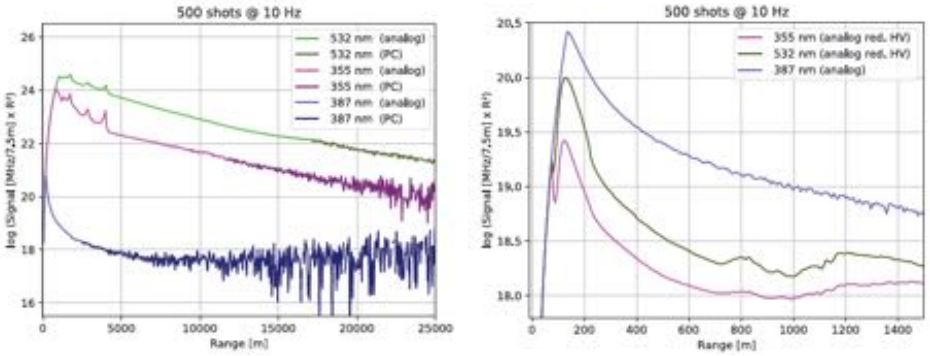


Figure 7: First commissioning signal of the Raman LIDAR at the Universitat Autònoma de Barcelona university campus. On left is the graph of a logarithm of range-corrected signal collected with maximum PMT gain and laser power to demonstrate LIDARs long-range behavior. On the right is a graph of a logarithm of range-corrected signal collected with reduced PMT gain for two elastic lines to demonstrate its short-range behavior [14].

## 5. Conclusions and prospects

Tests performed with Barcelona LIDAR design show, that the system can collect enough data to get complete atmospheric extinction profile for the altitudes of up to 15 km a.s.l. in just 500 laser pulses. The elastic lines can penetrate even further, reaching maximum range of 30 km, which gives

access to the stratosphere, if pointed vertically. This fulfills the criteria set by CTAO.

The LIDAR systems are now in final servicing and pre-testing stage, and will soon be moved to the CTA site for testing stage of the equipment in its working environment. Barcelona LIDAR will participate in a cross-calibration campaign with an EARLINET-calibrated LIDAR for site characterization. After the tests, LIDAR performance will be evaluated. Based on the experience gained from all three designs, the final Raman LIDARs for CTA will be built.

## Acknowledgements

This research was supported by the Slovenian Research Agency grants P1-0031, I0-0033, J1-9146 and its Young Researchers program.

## References

[1] The CTA Consortium, *Science with the Cherenkov Telescope Array*, World Scientific, (2018).

[2] M. Gaug, *CTA Atmospheric Calibration*, EPJ Web of Conferences **144**, 01003, (2017).

[3] K. Bernlöhr, *Impact of atmospheric parameters on the atmospheric Cherenkov technique*, Astrop. Phys. **12** 225, (2000).

[4] J. Rakovec, T. Vrhovec, *Osnove meteorologije za naravoslovce in tehnike (2nd edition)* DMFA, (2000).

[5] L. Wang et al., *Evolution of the planetary layer in the presence of fog and plume*, Geophysical research abstracts, vol. 18., (2016).

[6] L. Wang, *Study of atmospheric aerosol properties in the Vipava valley*, Dissertation, University of Nova Gorica , (2016).

[7] F. Gao, *Study of processes in atmospheric boundary layer over land-sea transition interface using scanning LIDAR*, Dissertation, University of Nova Gorica, (2012).

[8] M. Živec, *Measurements and modeling of air mass motion in the troposphere*, Diploma thesis, University of Nova Gorica , (2016).

[9] M. Doro et al., *Status and motivation of Raman LIDARs development for the CTA Observatory*, ATMOHEAD WORKSHOP, (2013).

[10] O. Ballester et al., *Raman LIDARs and atmospheric calibration along the line-of-sight of the Cherenkov Telescope Array*, 36$^{\text{th}}$ International Cosmic Ray Conference, PoS(ICRC2019)814, (2019).

[11] D. Alexandreas et al., *Status report on CLUE*, NIM A, 360, 385-389, (1995).

[12] P. Baillon et al., *An improved method for manufacturing accurate and cheap glass parabolic mirrors*, NIM A 276(3), 492-495, (1989).

[13] P. Baillon et al., *Production of 300 paraboloidal mirrors with high reflectivity for use in Barrel RICH counter in DELPHI at LEP*, NIM A 277(2), 338-346, (1989).

[14] M. Gaug et al., *The IFAE/UAB Raman LIDAR for the CTA-North*, EPJ Conf. Ser. **197** 02005, (2019).

# Novel technology for grid energy storage in matter - the efficiency of $FeCl_{2(aq)}$ to $Fe^0_{(s)}$ reduction under different electrolysis conditions

Uroš Luin

*University of Nova Gorica, Vipavska 11c, Si-5270 Ajdovščina, Slovenija*

**Abstract**

The grid energy management requires incorporation of energy storage technologies. Intermittent weather dependent renewable sources are exposing grid instability due to unmatched supply and demand. Our patented materially closed cycle energy storage technology is suitable for mid-term, seasonal as well for storage reserves. Energy is stored in the form of matter, as change in red-ox potential of iron ($Fe^{2+}$/ $Fe^0$), without self-discharge. The presented experimental work focus on electrolytes and electrolysis system optimization. Catholyte $FeCl_{2(aq)}$ concentrations 1.5M to 3M (190-380 $gFeCl_2\,dm^{-3}$) are within $>$ 93% range of maximum conductivity, with the peak at 2.25M ($285gFeCl_2dm^{-3}$). $Fe^0$ deposition using constructed electrolysis system operating at $70^{\circ}C$ resulted in $\Delta Gibbs$ efficiency of $94 \pm 4\%$ at current density $100A/m^2$ and $51 \pm 2\%$ at $1500A/m^2$. The HHV efficiency was $81 \pm 4\%$ and $43 \pm 2\%$ respectively. Electrolysis efficiency was about 10% higher after increasing system temperature from 30 to $70^{\circ}C$.

*Keywords:* Energy storage in matter, Power-to-solid, $FeCl_2$ electrolysis

*Email address:* uros.luin@ung.si (Uroš Luin)

## 1. Introduction

Although fossil fuels are consequence of the energy from the Sun, combined with the planetary bio-activity and the historical Geo-chemical events, they are not a renewable source. The rate of natural production of fossil fuels is critically slower and strongly unmatched with the nowadays exploitation rate. Furthermore, notable global climate change is led by the increase of atmospheric greenhouse gases as the byproducts of burning fossil fuels. However, there are also exceptions like burning natural gas coming from Earth's crust to $CO_2$, since $CH_4$ has more than 25 times greater green house gas effect than $CO_2$ over a 100-year period [1]. The shift to weather dependent renewable energy sources is exposing a fundamental weakness in the grid stability due to unmatched electricity production and consumption. Consequently, during off-peak periods the surplus electricity is being curtailed. Therefore, introducing more renewable sources requires ancillary services for grid management to assure peak shaving and load shifting. The magnitude of world's electric energy consumption easily exceeds $20\,000\,TWh\,year^{-1}$. So, environmentally and economically acceptable, high capacity and long term energy storage (ES) technology is necessary to enable global scale application of renewable green energy sources [2, 3].

Currently, a number of different ES technologies are in use; however, their high cost and inefficiency suppress their relevance and functionality. The most mature ES technology is Pumped Storage Hydro-power (PSH) with around 99 % of capacity of the world-wide electricity storage [4]. Despite high initial investment cost, PSH is currently the most cost efficient form of the ES with real-life round-trip efficiency of 72-75 %. The major drawback of PSH is related to its high environmental impact due to high land footprint resulting from a small energy density of only around $1\,Wh\,dm^{-3}$. In addition, a choice of the PSH plant sites is geographically limited [5, 6]. The remaining 1 % storage capacity belongs to batteries, compressed air energy, flywheels and hydrogen storage, mostly used for the secondary regulation [7]. The grid energy can also be stored via conversion into chemical energy contained in substances called energy carriers. Today, by far the most important energy carrier is hydrogen, which can be produced by water electrolysis, stored, transported around (although nowhere near as easily as electricity) and then, when needed, it can be converted back to electricity. This hydrogen can be stored as a compressed gas up to 700 bars, stored as liquid hydrogen or absorbed in hydrides. Under such storage conditions, the hydrogen's energy density is several orders of magnitude higher

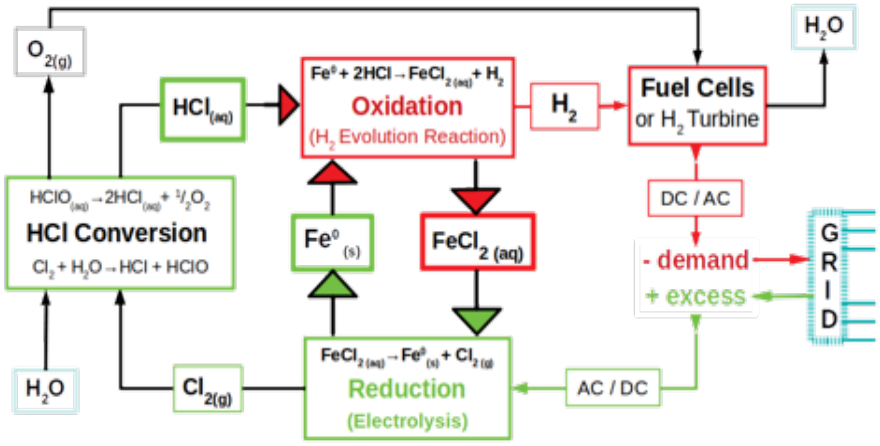


Figure 1: The schematic diagram of our patented technology for energy storage in matter. System parts indicated with green are active during storing when is excess grid energy and are: rectification (alternative to direct current) unit, electrolysis unit with one or several electrolytic cells, $Cl_2/HCl$ conversion unit, HCl tank and elemental iron storage. System parts indicated with red are active during the release of energy from matte and are: hydrogen evolution reactor, hydrogen to electricity conversion unit and direct to alternative current conversion unit. Water enters the materially closed cycle into the HCl conversion unit and it exit the cycle during hydrogen conversion to electricity by fuel cells or turbine. Oxygen released from HCl conversion unit could be stored or released into atmosphere.

than that of PSH, meaning that the land footprint is much smaller due to the significantly smaller storage volumes. The barrier to commercial implementation is in the low overall efficiency and huge capital costs. The efficiency of a commercial water electrolyser is around 70 %, while 10-20 % energy is lost in the hydrogen storage segment and another 50 % in the fuel cells. So, the overall energy efficiency does not exceed 30 % [8, 9]. Another disadvantage of hydrogen storage is the question of safety, mainly related to an uncontrolled release of hydrogen that can cause cryogenic burns or explosions [10, 11].

### *1.1. Description of our process for energy storage in matter*

Our patented [12] material closed-cycle process for converting electricity into matter, allows long-term energy storage without self-discharge, with the possibility to split production and consumption segments, and so delivery of energy on remote locations. Shortly, it is defined as a novel power-to-solid grid energy storage technology, that can store the energy in the oxidative-reductive potential of a metal. The other major advantages of the presented technology are: no geographical constrains and small land footprint for the storage plants, no emission or waste due to a closed working

cycle, no self-discharge during long-term storage and possibility for decoupling of the storage and back-conversion segment that allows many application possibilities. In the case of excess grid energy, the electricity is stored

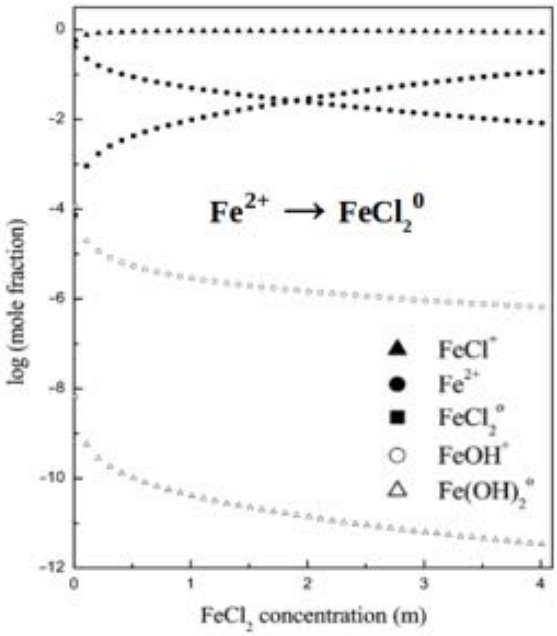


Figure 2: Man-Seung Lee and colleagues reported presence of ionic and non-ionic species in iron chloride solution as a function of concentration [13]. With increasing $FeCl_2$ concentrations, the $Fe^{2+}$ species are decreasing, while $FeCl_2^0$ are increasing.

as a decrease in the oxidation state of the iron species and formation of hydrochloric acid. Catolyte is a $FeCl_{2(aq)}$ solution, where iron chloride is dissociated into cations $FeCl^+$, $Fe^{2+}$, $FeOH^+$ and chloride anions $Cl^-$. Present are also neutral or non-ionic species $FeCl_2$ and $Fe(OH)_2$. Shown in Figure 2, Man-Seung Lee and colleagues [13] found that increasing the concentration decreases the proportion of the ionic species $Fe^{2+}$ and increases the proportion of the neutral species $FeCl_2$. The reduction of ionic iron species from the catholyte solution to $Fe^0$ takes place at the cathode. As anolyte is used 2M $MgCl_2$ aqueous solution. $Cl^-$ anions pass through the anion exchange membrane (AEM) into the anolyte solution and at the anode are oxidized. So, each $Cl^-$ give one $e^-$ to anode according to $2Cl^-_{(aq)} \longrightarrow Cl_{2\,(gas)} + 2e^-$. The evolved chlorine gas leaves the electrolysis system and enters scrubbing chamber for conversion to $HCl_{(aq)}$, through photo-catalytic reaction in aqueous medium. During the latter process, $O_2$ is formed.

Energy storage and release is temporally and spatially separated. Energy stored in matter is not subjected to self-discharge and compared to existing methods is suitable for seasonal energy storage. Storing power depends on the size of electrolysis unit, more precisely on electrode active surface area and current density. Energy capacity depends on the size of suitable storage facilities for: $HCl_{(aq)}$, $Fe^0_{(s)}$ and $FeCl_2$. Energy is released on demand through controlled hydrogen evolution reaction (HER) using $Fe^0$ and $HCl_{(aq)}$ as reactants. Reaction products are $H_2$ as energy carrier and $FeCl_{2(aq)}$. The conversion of hydrogen gas into electricity is done by fuel cells or hydrogen turbines. $FeCl_{2(aq)}$ is reused as the catolyte. Importantly, the described energy storage process runs in a closed material cycle. Research and development work is focused on optimization of electrolysis deposition rate efficiency of the morphological appropriate deposit within applicable current density range. Firstly, to obtain high deposition efficiency is crucial to optimize electrolytes conductivity. Secondly, is crucial to prevent competitive reactions as is water splitting and formation of masking gas bubbles on the electrode surface [14]. Simultaneous evolution of $H_2$ also affects the morphological properties of the deposit and contributes to the formation of porous, non-compact and brittle $Fe^0$ deposits. Electrolysis studies presented here do not involve the conversion of chlorine gas to HCl. The chlorine gas is immediately neutralized in aqueous NaOH solution.

## 2. Methodology

The methodology section is divided into electrolyte optimization and electrolysis studies sections. The purpose of experiments presented here is to optimize preparation and to find electrolyte conductivity dependence on concentration and temperature. Optimization of electrolysis system efficiency and understanding operation under different settings and conditions is in the second section.

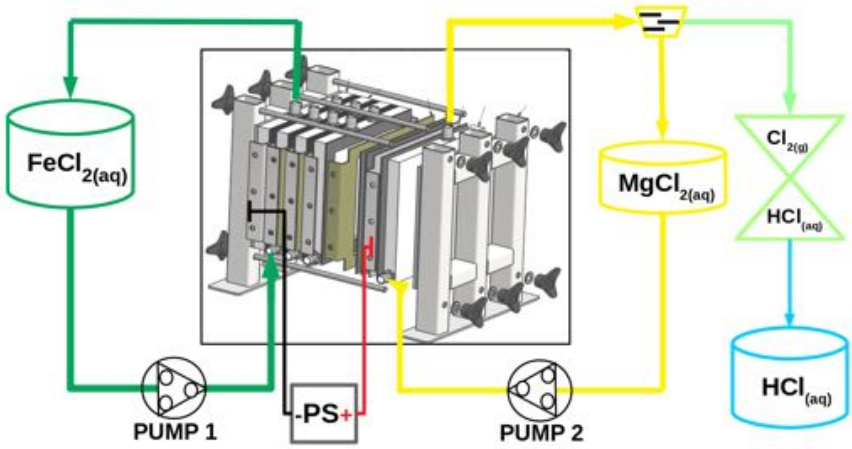


Figure 3: Sketch of the main components of the pilot electrolysis system. Inside rectangle is shown the electrolyser with several electrolytic cells mounted in the dedicated pressing device. With dark green is marked catholyte cycle, while with yellow is marked the anolyte cycle. With light green is marked the $Cl_2/HCl$ conversion and with blue the HCl acid tank. The negative terminal from power supply (PS) is connected to cathode, while positive to anode.

### 2.1. *Electrolytes conductivity dependence on temperature and concentration*

The initial supply of catholyte could be prepared either from commercial $FeCl_2[4H_2O]$ or from reaction between elemental Fe and HCl acid. The cost / benefit ratio suggest initialization from reaction. Higher catholyte pH inhibits the production of hydrogen during electrolysis. $FeCl_2$ solutions of the following concentrations: 1M, 2M, 3M, 4M and 5M, were prepared by reaction of Fe powder ($< 0.2\,mm$ particle size, 99+ % purity) and appropriate concentration of HCl acid respectively. The 2M $MgCl_2$ aqueous solution was prepared using $MgCl_2[6H_2O]$ (Fisher Chem.). De-ionized water (Grade 2) from the Adrona laboratory water purifier was used for all solutions. Oil bath Memmert ONE 10 was used to heat the electrolyte solutions.

Cooling was natural, the measuring chamber was raised above the oil level. Measurements of the pH, conductivity and temperature, were performed using Phoenix EC45-Multi linked to PC for data logging. Due to the purpose of the experiment, automatic temperature correction factor was set to zero. Measuring cell, Sentek VPT80C10, was calibrated using Hamilton ZDL 300 mL standard solution with conductivity 100 mS/cm at 25°C. Through the upper part of the measuring chamber nitrogen flow of 50 L/h. Peristaltic pump Boxer 25K (Boxerpumps, 3RWT1.6, 24VA4pDriver) was used for agitating the solution to reduce temperature gradients.

### 2.2. *Constructed electrolysis system*

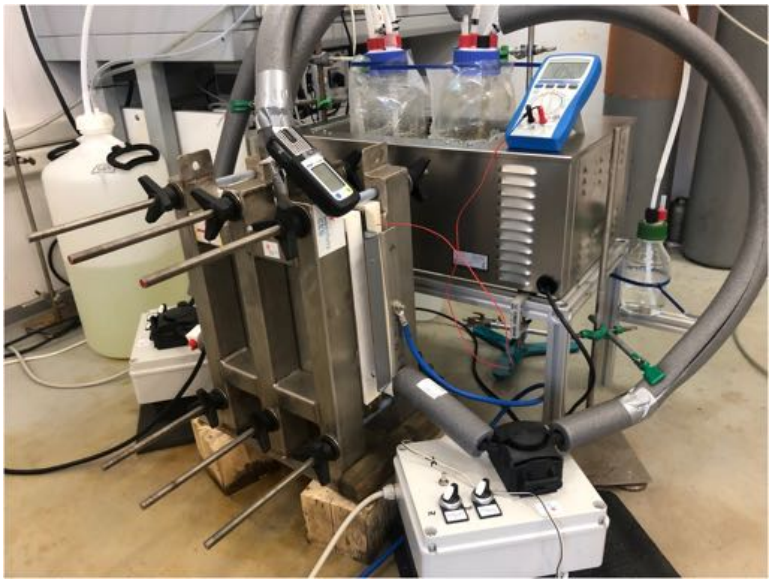

Figure 4: The constructed experimental electrolysis system with visible parts. Inside oil bath are two boro-silicate glass tanks with blue multi-connection caps, single a electrolytic cell mounted into dedicated press and two peristaltic pumps on the sides.

Iron cations are, during electrolysis reduced to $Fe^0_{(s)}$ and deposited on the cathode. Parameters adjusted only by changing the system parts are: resistance of the separator (AEM), active area of the electrodes, gap between the electrodes and mixing quality of the electrolytes. Adjustable parameters without changing parts of the constructed electrolysis system are: overpotential, system temperature, electrolytes concentration and flow rate. Main parameter of interest is the applied over-potential, which is directly proportional to the current density. The relationship between current density and

efficiency is inversely proportional. $Fe^0$ deposition rate $[g/m^2]$ depends on the current density $[A/m^2]$, electrolysis time $[s]$ and current (or Faraday) efficiency $[\%]$.

Due to evolution of gaseous products bottom up flow through the half cell compartments is being used. Mixture of liquid and gas flows through the upper cell outlet into the tank where the phases are separated. In the pilot system, electrolyte tanks function as decent degassing unit, although separate unit is shown on the sketch. Gases leave the system, while the liquid phase re-enters through the lower inlet of the cell. Gap or distance between the electrodes is $4.4 \pm 0.2\, mm$. Active cross-section area of the electrolytic cell is $0.0552\, m^2$. During electrolysis, catolyte concentration is decreasing with respect to iron deposition rate and electrolysis time. The stock catholyte volume present in the electrolysis system was over sized to ensure concentration within the optimal conductivity range.

Figure 5: Expanded view showing the components of the electrolytic cell. The two white half-cell parts on the sides are made from PTFE. The black parts are four EPDM rubber gaskets of which two are placed in-between each half cell and electrode, the remaining two are placed in-between each electrode and membrane with visible additional brownish masking tape towards membrane. Two electrodes of which cathode is the left light gray anode is the right dark gray. In the middle the light yellow is the anion exchange membrane AMI7001s (Membrane International, USA).

*2.3.* ***Electrolytic cell***

Used was modified commercial laboratory size electrolyzer (Denora / PFI-Milano) for $NaCl_{(aq)}$ electrolysis. Each cell consist of two half-cell compartments separated by membrane. The cathode is a titanium mesh welded to a titanium frame. The anode is dimensionally stable electrode for the evolution of chlorine gas in chlor-alkali industry. Technically, it is same as cathode, but with addition of few microns thick $IrO_2/RuO_2$ coating over titanium mesh. The housing of each half-cell is made of PTFE (Poly Tetra Fluoro Ethylene). Seals between the components are made from $3\,mm$ thick EPDM (Ethylene Propylene Diene Monomer) rubber. One or more electrolytic cells could be installed. Experiments presented here were performed using single electrolytic cell. Cell components are shown in Figure 5. The components are compressed in a dedicated six-point stainless steel (AISI304) clamping press. The outer dimensions of the electrolysis cell are $300 \times 300\,mm$. Non-conductive masking tape is placed on the inner edge of the EPDM seal toward direction of the membrane. It extends $\approx 7\,mm$ into the cell compartment to shield welding joints between mesh and frame. Consequently, the active cross-sectional area of the cell is $0.0552\,m^2$ ($235 \times 235\,mm$).

The starting time of electrolysis is the moment when the voltage across electrodes is higher than the standard potential of the electrolysis cell ($E^0_{cell} = 1.8\,V\,at\,25\,^\circ C$). The end of electrolysis is the moment when the voltage is lowered below the standard potential of the electrolysis cell and the current is zero. After each experiment, the membrane was immersed in cleaning $3\,\%\,wt$ HCl solution for 12-24 hours, then rinsed with de-mineralized water and stored in $5\,\%\,wt$ NaCl solution until the next experiment. System preheating includes electrolytes, massive metal frame and PTFE parts of the cell. Electrolyte circulation started when the appropriate temperature was reached. The thermocouple was inserted in the catholyte tank close to inlet from the electrolytic cell.

In the next section we firstly present results from initially performed electrolytes optimization studies. Secondly, we focus on the reduction and deposition of iron cations from aqueous iron chloride solution to solid elemental iron on the cathode. The presented experimental results were obtained using constructed scaled-up electrolysis system and covers: the electrolysis efficiency as function of current density, electrolysis efficiency as function of system temperature and the impact of non-optimal initial catholyte concentration on the electrolysis efficiency.

## 3. Results and Discussion

In preliminary electrolysis experiments the anolyte was 10 % $wt\ HCl$ which have negative $pH = -0.62$ and conductivity $\approx 700\ mS/cm$ at 25 °$C$. It's conductivity is relatively high compared to other water based electrolyte solutions. We observed anolyte pH increasing and decreasing of the catholyte pH during electrolysis. A plausible explanation is the migration of $H^+$ from the anolyte through the membrane into the catholyte. During the electrolysis, simultaneous water splitting as hydrogen evolution was observed, since it is favored in acidic catholyte. Consequently, iron deposition current efficiency was lowered on the account of $H_2$ that was simultaneously evolved. In order to inhibit the migration of $H^+$ protons across the membrane, the chemical composition of the anolyte was substituted. In continuation, for all the presented electrolysis studies anolyte solution of 20 % $wt\ MgCl_{2(aq)}$ prepared from $MgCl_2[6H_2O]$ (Fisher Chem.) with initial $pH = 3.7$ has been used. During the first use it changed from transparent to clear yellow solution and the $pH$ was decreased around 1. During further use, anolyte remained stable with $pH = 1 \pm 0.2$. Explanation is that the oxidation of the $OH^-$ species at the anode leads to the formation of $O_2$ and $H^+$ until a self-inhibition $pH$ value is reached. However, anolyte was not subjected to optimization and kept in the system without changing for all studies.

### *3.1. Catholyte preparation and regeneration process*

The catholyte solutions were prepared by reaction between appropriate HCl acid concentration and surplus ($\approx 10$ %) of iron powder, that resulted in a $pH > 4$ over time period of 12-36 hours. Neutralization time depends on the iron surplus, Fe granulation or specific surface, temperature and mixing. The time is shorter using smaller granulation or higher specific surface, higher surplus, at elevated temperature and with applied mixing. Oxidized catholyte was succesfully regenerated through a relatively small addition of HCl acid and Fe powder. Typical amount of HCl is 1-5 vol. % of catholyte solution volume. HCl enables the solubility of the precipitate, while hydrogen cause reduction of oxidized species. Mixing significantly improves the regeneration process because it increases the contact surface and prolongs the contact time between the hydrogen bubbles and the oxidized species in solution. During regeneration one can observe color and clarity change from opaque yellow-brown to clear intense green solution.

The excess iron must be removed from the neutralized solution to improve stability. Solutions were stable when stored in glass bottles without excess iron and flushed with argon or nitrogen gas.

### 3.2. *Comparison of catholyte solution prepared from commercial iron tetrachloride and catholyte solution from HER*

Comparison of pH and conductivity between $FeCl_{2(aq)}$ solutions prepared from commercial $FeCl_2[4H_2O]$ and from HER ($Fe^0 + 2HCl_{(aq)}$), over concentration range $0.75M - 4.6M$ (95-570 $g FeCl_2\, dm^{-3}$), is given in Figure 6. The conductivity deviates slightly due to the measurement error, however the trend of both solutions is consistent. The peak conductivity is at $2.25M$ ($285\, g FeCl_2\, dm^{-3}$). Concentrations in the range $1.5M - 3M$ (190-380 $g FeCl_2\, dm^{-3}$) represent the range $A > 93\,\%$ of maximum conductivity, while $2M - 2.5M$ represent range $B > 98\,\%$ of maximum conductivity.

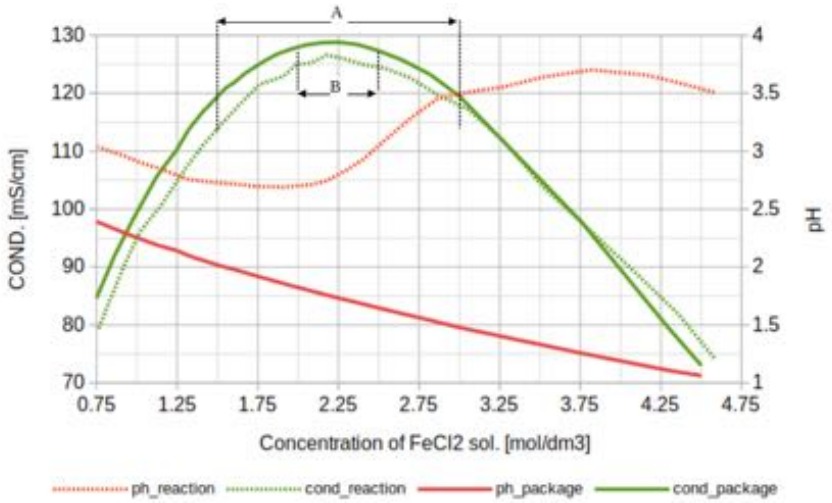


Figure 6: Conductivity and $pH$ comparison between catholyte solution prepared from commercial $FeCl_2[4H_2O]$ marked as continuous curves and catholyte solution from HER marked with dashed curves. Both measurements were done at room temperature ($25\,^{\circ}C$). The peak conductivity was found at $2.25M$ ($285\, g FeCl_2\, dm^{-3}$).

### 3.3. *Results from conductivity dependence on electrolyte temperature and concentration*

Results of temperature and concentration-dependent conductivity measurements are presented in the Figure 7. Despite intensive mixing, the measuring cell is additionally manually agitated, due to the release of dissolved gases from the solution into active area of the measuring cell. Bubbles cause a false decrease in conductivity and consequent deviations from the linearity of the curve are observed. Red values are measured during heating, while blue during cooling. In the case of a 5M solution, the conductivity divergence between heating and cooling is significant. A noticeably higher amount of yellow-brown precipitate formed in the 5M solution compared to other concentrations. The 2M $MgCl_2$ solution ($\approx 19\,\%\,wt$) has the highest conductivity and a steeper curve compared to $FeCl_2$ solutions. Among $FeCl_2$ solutions, 2M has the highest conductivity, followed by 3M with a relatively small difference. In the middle is 1M $FeCl_2$, followed by 4M solution. The 5M solution has the lowest conductivity. The results are consistent with the results of previous concentration-dependent conductivity measurements in Figure 6 and the findings shown in Figure 2.

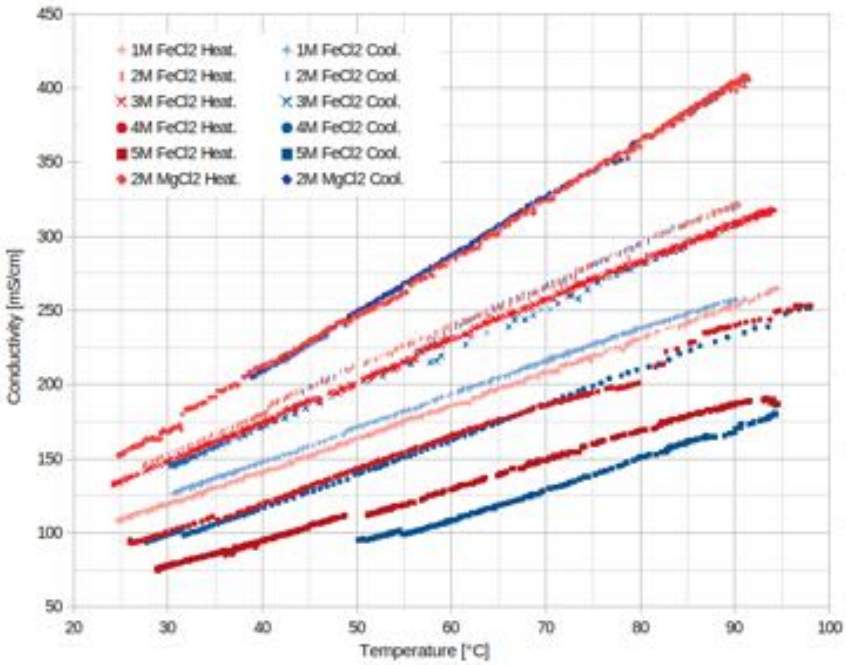


Figure 7: Linear dependence between electrolytes conductivity and temperature.

| E No. | Temp. [°C] | $R_{cell}$ [Ω] | $Fe^0_{d.r.}$ [g/h] | η[V] | U [V] | I [A] | P [W] |
|---|---|---|---|---|---|---|---|
| E17 | 71.1 | 0.020 | 6.1 | 0.17 | 1.91 | 5.6 | 11 |
| E11 | 69.0 | 0.022 | 19.7 | 0.47 | 2.21 | 18.3 | 40 |
| E14 | 71.1 | 0.020 | 34.8 | 0.72 | 2.46 | 33.2 | 82 |
| E15 | 72.4 | 0.020 | 59.2 | 1.20 | 2.94 | 56.8 | 167 |
| E16 | 72.4 | 0.018 | 90.0 | 1.65 | 3.38 | 87.5 | 296 |

| E No. | CD [A/m²] | $Fe^0_{N.d.r.}$ [g/hm²] | $E_{IN}$/g$Fe^0$ [Wh/g] | ΔG Eff. | HHV Eff. |
|---|---|---|---|---|---|
| E17 | 101 | 110 | 1.76 | 94.6% | 80.6% |
| E11 | 332 | 357 | 2.05 | 81.5% | 69.4% |
| E14 | 602 | 630 | 2.35 | 71.0% | 60.5% |
| E15 | 1028 | 1072 | 2.82 | 59.1% | 50.4% |
| E16 | 1584 | 1630 | 3.29 | 50.7% | 43.3% |

Table 1: Average parameter values from electrolysis experiments at different current density and relatively fixed system temperature.

### *3.4. **Results from experimental electrolysis studies***

The $pH$ of the catholyte at the end of the electrolysis was typically higher than 2. To compensate for depleted concentration, the catholyte was regenerated and neutralized after each use, with the addition of surplus iron powder and appropriate concentration of HCl acid. In the following experiments, evolution of hydrogen during electrolysis was not observed and consequently the current deposition efficiency of iron was close to 100 %.

The following results of experimental electrolysis studies provide a relationship between current density, system temperature, catholyte concentration, and system efficiency. Outline of performed electrolysis experiments:
- at different current densities,
- at different system temperatures,
- at non-optimal initial catholyte concentration,

Used are parameters and labels of which description follows:
E No. is the serial number of the performed electrolysis experiment. Temp. $[^\circ C]$ is the average temperature of the catholyte during active electrolysis time. $R_{cell}[\Omega]$ is the average resistance of a cell. $Fe^0_{d.r.}[g/h]$ is the time-normalized elemental iron deposition rate expressed as mass per unit time. $CD[A/m^2]$ is the average current density. $\eta\ [V]$ is the average over-potential during electrolysis. $U[V]$ is the average voltage between the electrodes. $I[A]$ is the average current drained from power supply during electrolysis. $P[W]$ is the average electrical power. $Fe^0_{N.d.r.}[g/hm^2]$ is the time-normalized iron deposition rate expressed as the mass per unit time per

square meter of active cross section area. $E_{IN}/\,gFe^0[Wh/g]$ is the energy input into electrolytic cell per gram of deposited elemental iron. $\Delta G_{Eff}[\,\%]$ is the electrolysis efficiency expressed as the ratio between the Gibbs free energy change for deposited moles of iron and the energy input for deposited moles of iron $(\Delta G_{Fe} \times mol\,Fe^0) \div E_{IN}$. $HHV_{Eff}[\,\%]$ is the electrolysis efficiency expressed as the ratio between higher heating value energy of hydrogen formed from deposited iron mass and the energy input for deposited iron mass $(HHV_{H2} \times mol\,H_2) \div E_{IN}$.

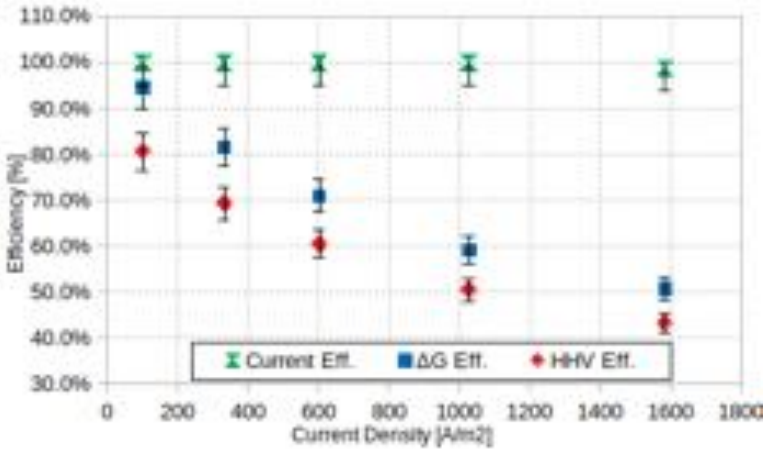


Figure 8: Performed electrolysis experiments at constant system temperature of 70 °C and given current density range, resulted in decreasing $\Delta G_{Eff}$ and $HHV_{Eff}$ efficiency, while current efficiency was close to 100 %.

### 3.5. *Electrolysis at different current densities*

In the Table 1 are average values for each electrolysis. The current density is directly proportional to the connection voltage and inversely proportional to the resistance of the cell. The relationship between energy efficiency and current density from performed electrolysis is shown in Figure 8. The resistance of the electrolysis cell is inversely proportional to the system temperature. In addition to results it was observed that the current density in E17 was decreasing from the initial 127 $A/m^2$ to the final 90 $A/m^2$ during electrolysis. The reverse trend is observed in E16, where the current density increased from the initial 1558 $A/m^2$ to the final 1811 $A/m^2$. E17 and E11 were performed without initial polarization, since they were carried at low power. E15 and E14 include two-minute polarization at $\eta = 0.1\,V$, while E16 at $\eta = 0.2\,V$. Linear dependence of applied potential across electrodes on the average current drained through the electrolysis cell was observed,

| E No. | Temp. [°C] | $R_{cell}$ [Ω] | $Fe^0_{d.r.}$ [g/h] | η [V] | U [V] | I [A] | P [W] |
|---|---|---|---|---|---|---|---|
| E20 | 81 | 0.023 | 33.5 | 0.75 | 2.48 | 32.8 | 81 |
| E14 | 71 | 0.020 | 34.4 | 0.72 | 2.46 | 33.2 | 82 |
| E21 | 49 | 0.029 | 34.4 | 0.96 | 2.73 | 33.1 | 91 |
| E19 | 31 | 0.038 | 34.8 | 1.24 | 3.03 | 32.5 | 98 |

| E No. | CD [A/m²] | $Fe^0_{N.d.r.}$ [g/hm²] | $E_{IN}/gFe^0$ [Wh/g] | ΔG Eff. | HHV Eff. |
|---|---|---|---|---|---|
| E20 | 594 | 607 | 2.42 | 68.4% | 58.7% |
| E14 | 602 | 630 | 2.35 | 71.0% | 60.5% |
| E21 | 600 | 623 | 2.63 | 64.6% | 54.1% |
| E19 | 588 | 629 | 2.83 | 60.7% | 50.2% |

Table 2: Resulting average parameter values from electrolysis at different temperatures and relatively fixed current density.

indicating operation within the resistance regime. Thus, we can conclude that in a given range of current density, electrolysis is not subject to a decrease in conductivity due to the inhibitory mass transport of ionic species from the electrolyte to the electrode surface.

### 3.6. *Electrolysis at different system temperature*

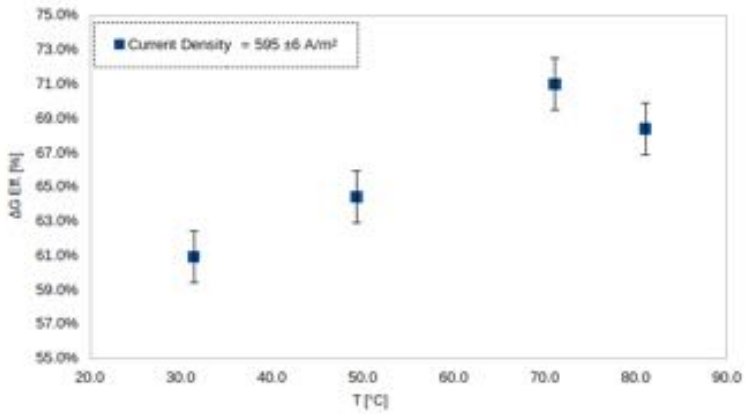


Figure 9: Shown are electrolysis study results of efficiency dependence on system temperature. Electrolysis were done at constant current density ($595 \pm 6\, A/m^2$) and different average system temperature: 31, 49, 71 and 81 $°C$. Electrolysis efficiency is defined as Gibbs free energy change for the amount of deposited elemental iron.

The results from electrolysis performed at different system temperatures and the comparable current density are shown in Table 2. They provide basic understanding of relationship between: efficiency and temperature, temperature and cell resistance, and over-potential variations at different

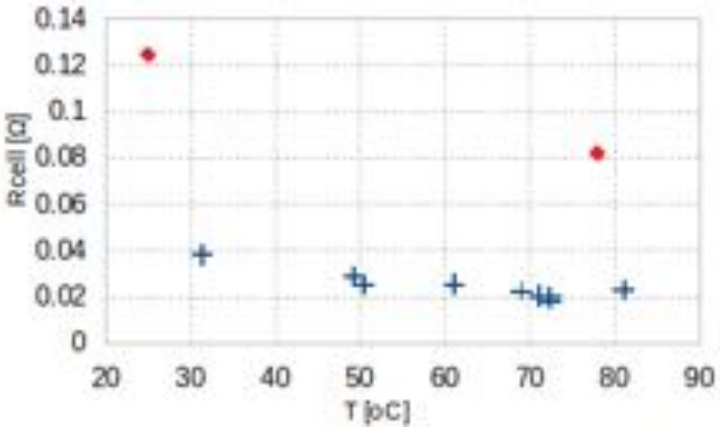


Figure 10: Cell resistance as a function of temperature. E10 and E12 (red rhombus) were performed at non optimal initial 4M $FeCl_2$ catholyte . Others (blue crosses) at 2.5M $FeCl_2$.

system temperatures for defined current density. The following electrolysis: E19, E21, E14 and E20 were performed at a fixed current density of 595 $\pm 6\, A/m^2$ and different system temperatures: 31.4, 49.3, 71.1 and 81.1 $^\circ C$, respectively. The temperature has several impacts that directly or indirectly affect conductivity like: electrolyte viscosity, ion mobility, electrolytes solubility, concentration of dissolved gases and standard potential of the cell. Calculation using temperature correction factors for half cell reactions from literature [15] shows a drop of $E^0_{cell}$ from 1.804 $V$ at 25 $^\circ C$ to 1.734 $V$ at 75 $^\circ C$. Figure 9 shows the relationship between the energy efficiency and the temperature of the system at the comparable current density.

### *3.7. Electrolysis with non-optimal initial catholyte concentration*

Most experiments presented in this paper until now were performed at an initial optimal catholyte concentration of 2.5M. In order to find the impact of initial non-optimal catholyte concentration on the efficiency, were performed two electrolysis (E10, E12) with initial 4M catholyte. E10 was performed at room temperature (25 $^\circ C$) and a current density of 152 $A/m^2$, while E12 at a system temperature of 78 $^\circ C$ and an average current density of 142 $A/m^2$. In Figure 10 is shown the resulting cell resistance dependence on temperature between optimal 2.5M (blue crosses) and non-optimal 4M (red rhombuses) initial catholyte concentration. The resulting over-potential dependence on current density is presented in Figure 11. Electrolysis results using 4M catholyte are in agreement with expectations based on electrolytes measurements of conductivity dependence on concentration and temperature. At lower electrolytic cell conductivity, higher

over-potential for certain current density is required. Consequently, lower electrolysis efficiency was obtained.

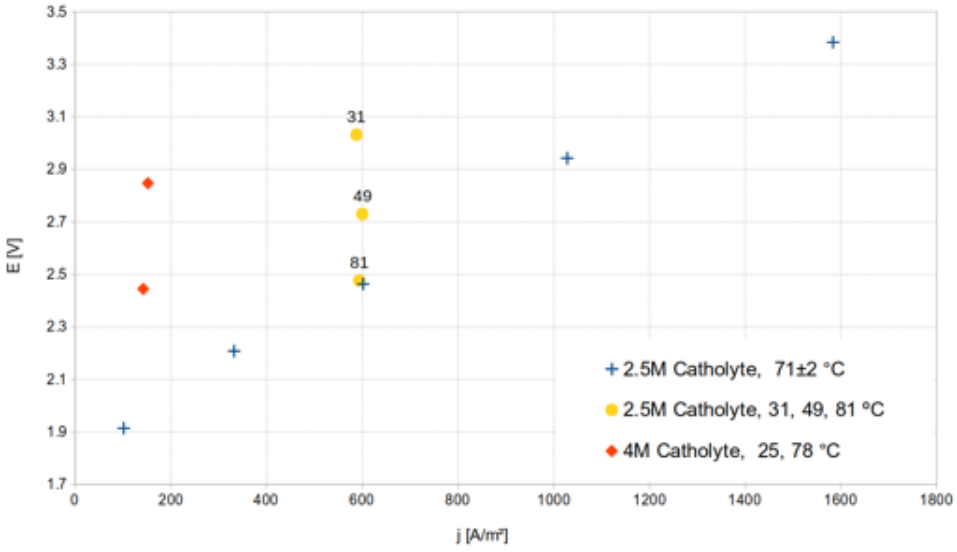


Figure 11: The blue crosses show the relationship between current density and voltage at system temperature of 70 °C and initial catholyte concentration of 2.5M. Electrolysis with 4M $FeCl_2$ are indicated by red rhombuses. The yellow circles represent the electrolyses performed with a variable system temperature of 31.4, 49.3, 71.1 and 81.1 °C and an initial catholyte concentration of 2.5M.

### *3.8. Electrolysis studies graphic summary*

In Figure 11 are shown results from three individual electrolysis studies. Electrolysis with initial 4M $FeCl_2$ as catholyte clearly show that higher cell resistance requires higher voltage at constant current density. Consequently, by using non-optimal electrolytes the electrolysis efficiency is lower as expected. Based on the obtained results, the future electrolysis optimization will include experiments using different AEM, preferably with lower $H^+$ transfer rate and higher ionic conductivity compared to AEM (AMI7001) used in the presented work.

## 4. Conclusions

The experimental results using the constructed electrolysis system under different operation conditions, provided understanding of individual parameter contribution to the electrolysis efficiency. We confirmed the concentration and temperature dependence on electrolyte conductivity and found the maximum conductivity at 2.25M $FeCl_{2(aq)}$ ($\approx 285\, g FeCl_2\, dm^{-3}$). Concentration dependent proportion of ionic $Fe^{2+}$ and neutral species $FeCl_2$ reported in literature [13] is in agreement with our results. Non ionic $FeCl_2$ species do not contribute to conductivity, in contrast to diminishing $Fe^{2+}$ concentration which contribute and so explaining the catholyte peak conductivity. Concentrations $1.5M-3M$ ($190-380\, g FeCl_2\, dm^{-3}$) are within $>93\,\%$ range of maximum conductivity, while $2M-2.5M$ are within $>98\,\%$ of maximum conductivity.

System temperature is directly proportional to electrolytic cell conductivity. Consequently, the increase of system temperature lowers the over-potential at constant current density. Increased electrolysis efficiency of about 10 % over $40\,^{\circ}C$ system temperature difference was observed. Using optimized initial catholyte solution at system temperature $70\,^{\circ}C$ resulted in $Fe^0$ deposition $\Delta Gibbs$ efficiency of $94\pm4\,\%$ at current density 100 $A/m^2$ and $51\pm2\,\%$ at 1500 $A/m^2$. The HHV efficiency was $81\pm4\,\%$ and $43\pm2\,\%$ respectively.

Linear dependence between applied potential and the current density in the range of $100-1000\ A/m^2$ suggest, that mass transport of ionic species from the electrolyte to the electrode surface was not limited. Finally, we recognized AEM as important element for further electrolysis efficiency optimization, especially required at high current densities. Future work will include electrolysis experiments using different AEMs targeting on lower $H^+$ transfer rate and higher ionic conductivity compared to AMI7001.

## 5. Acknowledgment

For the financial support, the authors acknowledge the Institute CES d.o.o.

## 6. Conflicts of interest

There are no conflicts to declare.

[1] IPCC, *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.*, Cambridge University, 996 pp.(2007). https://www.ipcc.ch/report/ar4/wg1/, accessed in June, 2020. 1

[2] R.L. Fares and M.E. Webber, *The impacts of storing solar energy in the home to reduce reliance on the utility*, Nat. Energy 2, 17001 (2017). http://www.webberenergygroup.com/wpnew/wp-content/uploads/Nature-Energy-Storing-Solar-Energy-01-17.pdf, accessed in June, 2020. 2

[3] R.L. Revesz and B. Unel, *Managing the Future of the Electricity Grid: Energy Storage and Greenhouse Gas Emissions*, Harvard Environmental Law Review, 42, (2018). https://harvardelr.com/wp-content/uploads/sites/12/2018/03/revesz_unel.pdf, accessed in June, 2020. 3

[4] International Energy Agency, *Technology Roadmap, Energy Storage*, IEA Publications Paris, (2014). https://www.nachhaltigwirtschaften.at/resources/iea_pdf/iea_technology_roadmap_energystorage.pdf, accessed in June, 2020. 4

[5] M. Aneke and M. Wang, *Energy storage technologies and real life applications – A state of the art review*, App. Energy 179 350–377 (2016). http://eprints.whiterose.ac.uk/154479/1/2016_05_05_MA_Modified_Manuscript_NotMarked.pdf, accessed in June, 2020. 5

[6] S. Rehman, L. M. Al-Hadhrami, M. M. Alam *Pumped hydro energy storage system: A technological review*, Renewable and Sustainable Energy Reviews 44 586–598 (2015). https://ideas.repec.org/a/eee/rensus/v44y2015icp586-598.html, accessed in June, 2020. 6

[7] F. Ausfelder *et al. Energy storage as part of a secure snergy supply* ChemBioEng Reviews, Vol 4, Issue 3, (2017). https://doi.org/10.1002/cben.201700004, accessed in June, 2020. 7

[8] B. Zakerin and S. Syri, *Electrical energy storage systems: A comparative life cycle cost analysis*, Renewable and Sustainable Energy Reviews 42 569–596 (2015). https://econpapers.repec.org/article/

eeerensus/v_3a42_3ay_3a2015_3ai_3ac_3ap_3a569-596.htm, accessed in June, 2020. 8

[9] M. A. Pellow, C. J. M. Emmott, C. J. Barnhartd and S. M. Benson, *Hydrogen or batteries for grid storage? A net energy analysis*, Energy Environ. Sci. 8, 1938-1952 (2015). https://pubs.rsc.org/en/content/articlelanding/2015/ee/c4ee04041d#!divAbstract, accessed in June, 2020. 9

[10] A. Rusin and Stolecka K., *Hazards associated with hydrogen infrastructure*, J. Power Technologies 97 153–157 (2017). http://yadda.icm.edu.pl/baztech/element/bwmeta1.element.baztech-b90297df-d8e1-484e-bf02-c39ca8e57234, accessed in June, 2020. 10

[11] F. Rigas and S. Sklavounos, *Evaluation of hazards associated with hydrogen storage facilities*, J. Hydrogen Energy 30 1501–1510 (2005). https://doi.org/10.1016/j.ijhydene.2005.06.004, accessed in June, 2020. 11

[12] M. Valant, *Procedure for electric energy storage in solid matter*, Patent No. SI 25573 A, WO 2019117818 (A1), (2017). https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2019117818, accessed in June, 2020. 12

[13] Man-Seung Lee *et al. Chemical equilibria in ferrous chloride acid solution* Metals and Material. International, Vol. 10, No. 4, pp. 387-392, (2004). https://link.springer.com/article/10.1007%2FBF03185990, accessed in June, 2020. 13

[14] A. Angulo *et al.*, *Influence of bubbles on the energy conversion efficiency of electrochemical reactors*, Joule 4, 1-25 (2020). https://doi.org/10.1016/j.joule.2020.01.005, accessed in June, 2020. 14

[15] S. G. Bratsch, *Standard electrode potentials and temperature coefficients in water at 298.15K*, J. Phys. Chem., Vol. 18, No. 1, (1989) https://www.nist.gov/system/files/documents/2019/04/02/jpcrd355.pdf, accessed in May, 2020. 15

# Indirect Search for dark Matter in Andromeda galaxy with Cherenkov Telescope Array

Barbara Marčun

*University of Nova Gorica, Vipavska 13, SI-5000 Nova Gorica*

**Abstract**

After more than 40 years since the first evidence of the dark matter existence, its nature is still unknown, and the dark matter particles have not been detected yet. Cherenkov telescope array (CTA) will provide new possibilities to measure $\gamma$-ray emission and search for annihilation and decay products of dark matter particles. In the article some properties of dark matter, Andromeda galaxy (M31) and CTA, and the possibilities for dark matter indirect detection are described. Latest discoveries show that direct interaction between the dark component and Standard Model particles exists, which strongly challenge the Lambda Cold Dark Matter ($\Lambda CDM$) scenario with collisionless dark particles, and leads to the idea that the paradigm should be changed and that dark matter particles are something completely different. Maybe even our understanding of gravity should be modified.

*Keywords:* dark matter, Andromeda galaxy, M31, CTA, rotation curve, universal rotation curve, compactness

## 1. Introduction

In the 1960s at the Carnegie Institution of Washington, Vera C. Rubin and W. Kent Ford, Jr., observed stars in the Andromeda galaxy with the newly developed telescope instrument that amplified starlight. They expected that stars, dust, and gas in this spiral galaxy would spin faster at its center than at its edges. But they found that the orbiting speeds of gas clouds in the outer parts of galaxies remain constant even beyond the visible starlight, which seemed impossible. With this speed, the gas and stars should have been torn apart, and the galaxy, according to the laws of gravity, shouldn't exist. But the intact spiral was there, so something else had to contribute to the mass of the galaxy. But nothing was there, at least nothing they could see. After they published their work in February 1970 [1], they measured velocities in more galaxies, and they found the same pattern.

In the 1970s the optical and 21-cm rotation curves were measured for several galaxies also by other astronomers [2]. The curves were again incompatible with curves that were predicted from the distribution of luminous matter, as shown in Figure 1. The dark component can explain the measured rotation curves, and the idea of the presence of invisible matter was born. It became obvious that there is much more

*Email address:* bbarbara483@gmail.com (Barbara Marčun)

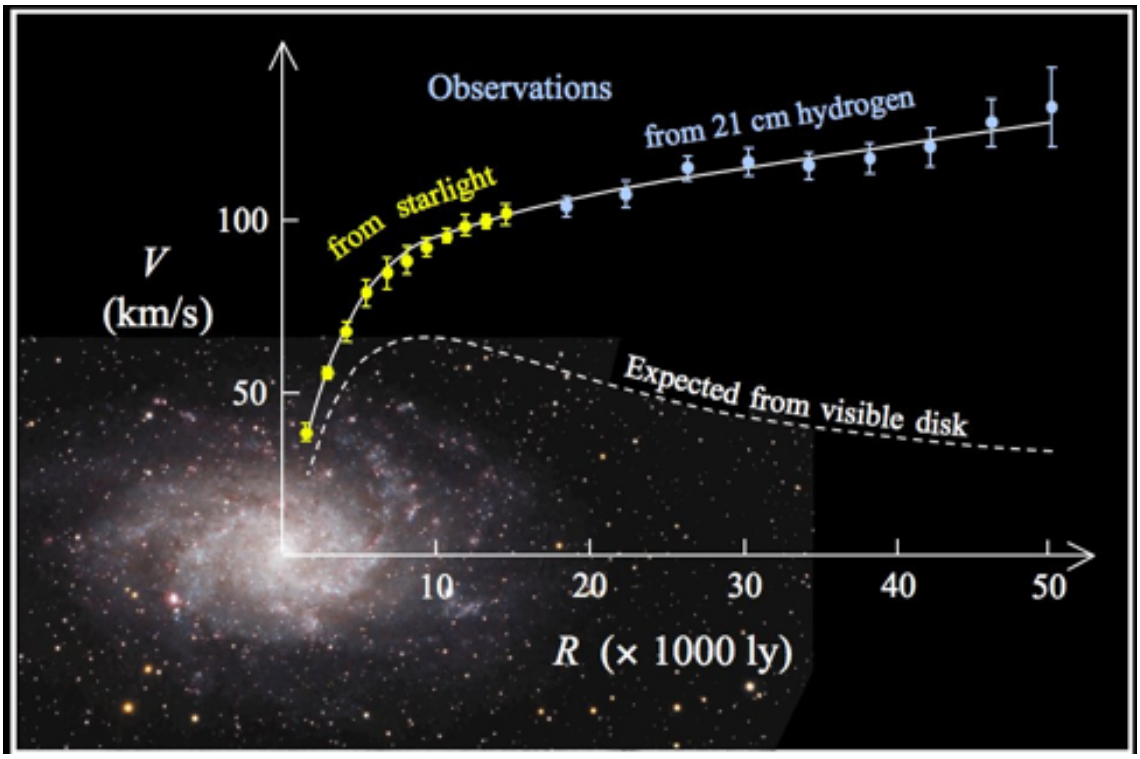


Figure 1: The image of M33 and the corresponding rotation curve. The dashed curve was predicted from the distribution of luminous matter, the solid curve is a fit to the measured data. The dark component explains the discrepancy between the curves. Figure adopted from [3].

invisible matter around us than the type that can be detected. Since then, the dark component has been an important part of astrophysics, elementary particles physics, and cosmology. It is essential to explain the mass distribution in galaxies and clusters of galaxies, the large scale distribution of galaxies, the bottom-up growth of the cosmological structures, the redshift dependence of the expansion of its scale factor, and the relative heights of the peaks in the cosmic microwave background (CMB) fluctuations [3]. This observational evidence and theoretical issues support the idea of a massive dark particle from beyond the Standard Model of the elementary particles. Despite all theoretical and experimental efforts, there were no positive results in creating such particle in colliders or detecting its messengers, so its nature is still unknown.

## 2. Dark Matter

About 27% of the mass-energy density of the Universe is made up of non-relativistic gravitating matter, the rest is dark energy. Roughly 20% of the matter is in the form of known particles, mostly baryons, as the leptons, neutrinos, and photons represent much less mass than the baryons. The other 80% of the matter is made up of material that cannot be directly observed, the dark matter.

DM does not emit, block, or reflect electromagnetic waves. It interacts gravitationally and its gravity is felt in the dynamics of galaxies and galaxy clusters, and in the dynamics of the expansion of the Universe. Its presence could be directly mapped via gravitational lensing which distorts the paths of the light rays from distant objects through foreground clusters of galaxies, dominated by DM, like Abell 1689 (Figure 2),

Figure 2: Galaxy Cluster Abell S1063. A Giant Gravitational Lens, 4 billion light-years away, magnifies and distorts images of galaxies that lie far beyond the cluster. Images are seen as blue arcs that would be otherwise undetected. Figure adopted from [4].

observed by the Hubble Space Telescope. DM is structured over a wide range of mass scales. In the standard cosmology small, dense DM clumps were formed first, and with merging larger structures, like DM halo in Milky Way, were formed. These larger structures retain less massive subhalos. Since DM was first discovered there were many candidates proposed to explain what this mysterious component could be made of. Some of them are listed below.

### *2.1. Massive Compact Halo Objects*

Potential candidates include small black holes, low mass dim dwarfs or planets, so-called massive compact halo objects (MACHOs). Supermassive black holes could also be part of the missing mass. But these hard-to-spot objects would have to play a more dominant role than the one observed in an extensive search for those objects, so this hypothesis has been shelved. Or perhaps there are primordial black holes left over from the Big Bang that have not been found yet.

### 2.2. Neutrinos

Neutrinos could be other candidates. The three known types of neutrinos, electron, muon, and tau neutrinos, have mass small enough to be relativistic. If they made up 22% of the total mass, their motions would have prevented the baryons from assembling into galaxies. So the total mass in neutrinos must be much smaller, and can not exceed 1% or 2% [5].

The fourth type, the sterile neutrino, is also proposed as a DM candidate. If sterile neutrinos exist, they only interact with baryonic matter gravitationally. Sterile neutrino's mass should be around 1 $eV/c^2$. Their existence would answer questions such as why neutrinos have mass. If sterile neutrinos are important contributors to dark matter, they are expected to produce a strong lack of atmospheric muon neutrinos crossing the Earth (Figure 3), which results in a depletion at energies of a few TeV due to matter effects in neutrino oscillations.

IceCube has not found any anomalous lack of muon neutrinos with two independent searches, each consisting of a year's amount of data, or about 100,000 neutrino events. In August 2016 the conclusion was that with 99 percent certainty the eV-mass sterile neutrino does not exist [4]. The IceCube Collaboration has continued the hunt for sterile neutrinos with the Deep Core using lower energy atmospheric neutrinos. The observed muon neutrino flux in IceCube, measured for energies from 10 to 60 GeV, was consistent with the standard three-neutrino hypothesis. This new search looked into three years of IceCube data and again has not found any hint of a sterile neutrino [7].

### 2.3. Weakly Interacting Massive Particles

From the observational evidence, it is clear that DM particles must interact gravitationally, be non-relativistic (cold), and must have very small scattering cross-sections either with themselves or with Standard Model particles. Heavy DM candidates, with masses in the GeV to TeV range, are Weakly Interacting Massive Particles (WIMPs), which are a non-relativistic relic that froze out of equilibrium with the particle bath in the early Universe (the $\Lambda CDM$ scenario).

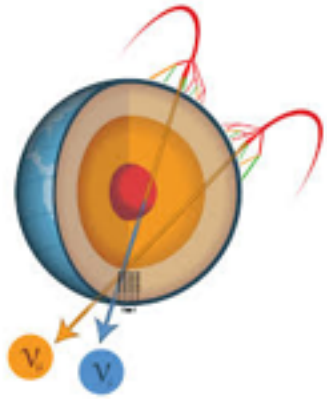


Figure 3: If sterile neutrinos exist, IceCube would measure a lack of atmospheric neutrinos that would otherwise reach the detector on a trajectory through the Earth's core. Figure adopted from [6].

If we assume that WIMPs make up 22% of the energy density of the Universe, that the strength of interaction (coupling constant) is comparable to weak interactions, and that the DM particles self annihilate producing the usual particles from Standard Model, a rough estimate can be done: $m_w c^2 \sim 0,1 - 1\,\mathrm{TeV}$, and cross-section for interacting with normal matter $\sigma_w \sim 10^{-36}\,\mathrm{cm}^2$ [5]. This cross-section is very small but is comparable to that of TeV neutrinos, measured by IceCube, and it is also within the range of accelerators like Large Hadron Collider (LHC) in Geneva. It is also comparable to the range currently probed by the $\gamma$-ray telescopes. The most widely used family of BSM (beyond the Standard Model) theories involve super-symmetry, in which the favored WIMP particle is neutralino with mass $m_\chi \gtrsim$ GeV [8]. In these models neutralino and its anti-particle can annihilate, resulting in Standard Model particles, like neutrino and anti-neutrino, photons,

electron and positron, etc.:

$$\chi + \bar{\chi} \to \nu + \bar{\nu}$$

$$\chi + \bar{\chi} \to \gamma + \gamma$$

$$\chi + \bar{\chi} \to e^- + e^+.$$

In $\Lambda CDM$ scenario WIMPs are expected to populate not only the galactic halos and clusters of galaxies, but also the Galactic center, and the center of the Earth and the Sun. The annihilation rate depends on the mean density of WIMPs and also on a clumpiness of the DM. In the case of the Earth and the Sun, we can only search for annihilation neutrinos since photons or electrons and positrons would be absorbed.

For annihilation of WIMPs in the Galactic halo or the Galactic center, we would expect to detect $\gamma$-rays (with VERITAS, Fermi, ...) and charged particles (with ATIC, PAMELA, ...). PAMELA satellite detected an excess of $\sim$10% in the positron flux, compared to the expected one from cosmic ray interactions in the interstellar medium [9], and even larger excess was found by the ATIC experiment. The excess signal could be interpreted due to DM annihilation, but such a conclusion is very tentative. Other explanations could be astrophysical, such as positrons from nearby supernovae, pulsars (like Geminga), or cosmic ray propagation effects. Fermi found only mild excess from the prediction in the region of the Galactic center, which could be another tentative signal of DM that needs to be further evaluated.

### *2.4. Axions and Axion-Like Particles*

Axion is a hypothetical light boson (pseudo-Nambu-Goldstone boson), which was introduced theoretically to explain why strong interactions do not violate a charge-parity symmetry, that would have been expected at some level. The problem was solved by Peccei and Quinn with a theory in which they predicted the existence of a new light particle, axion [10]. Its mass and coupling to other particles depends only on so-called Peccei-Quinn energy scale,

$$m_a = \frac{6 \cdot 10^{15} eV}{f_a} \sim 10^{-6} - 10^{-2}\,\mathrm{eV},$$

where $f_a \sim 10^9 - 10^{13}\,\mathrm{GeV}$ [5]. Masses of Axion-Like Particles (ALPs) are very poorly constrained and could range from $10^{-10}\,\mathrm{eV}$ to $10^9\,\mathrm{eV}$ [11]. Axions, as possible candidates forDM, are weakly interacting, non-relativistic, and neutral, so they can decay into two photons, each with the energy of $m_a c^2/2$. The lifetime for this process is extremely long. If the mass were $m_a \sim 0,1\,\mathrm{eV} - \mathrm{few\,eV}$, the process would have shown up as a narrow line in optical telescopes, that was not found. Masses below a few eV would lead to a drastic reduction in the number of He-burning red giants. They would lose energy by axion emission, compensated by increased nuclear burning and shortening the lifetime of the stars. This was not detected either.

In a strong magnetic field, axions would convert into an X-ray photon. Experiments in microwave caves are ongoing, so does the experiment CAST (CERN Axion Solar Telescope), using a 9.26 m long LHC magnet with $B \sim 9,5\,\mathrm{T}$. The magnetic field acts as a catalyst to transform axions into X-rays, making them relatively easy to detect. The first hints of the existence of axions and ALPs have already been seen in the anomalous cooling of white dwarfs and the anomalous transparency of the Universe for very high energy (VHE) $\gamma$-rays [11]. These hints are also tentative and do not confirm axions beyond doubt. Large-scale astrophysical $B$ fields can also induce axion-photon oscillations. $B$ is much smaller than in the laboratory, but the conversion region $L$ is much larger, and the product $BL$ can be

large. Sensitivities are restricted to very low-mass particles, far away from the axion band. One example is SN 1987A, which would have emitted a burst of axion-like particles that would have partially converted into $\gamma$-rays in the galactic $B$ field. The lack of a $\gamma$-ray signal in the Gamma-Ray Spectrometer (GRS) instrument of the Solar Maximum Mission (SMM) satellite in coincidence with the observation of the neutrinos emitted from SN1987A provides a strong bound on their coupling to photons. The underlying physics, modeling of the supernovae and the Milky-Way magnetic field, has been improved, resulting in the limit $m_{ALP} \lesssim 4,4 \cdot 10^{-10}\,\mathrm{eV}$ [12, 13].

### *2.5. Collisional dark Matter*

It is also possible that the dark particles are something else, something from beyond our current ideas, and they likely interact with Standard Model particles.

## 3. Indirect Search for dark Matter

### *3.1. Current Results of Indirect DM Searches*

Indirect searches of WIMPs consist of looking for anomalous $\gamma$-ray signals from astrophysical structures that host a large amount of DM. The table 1 summarises the WIMP targets, including the J-factor, which is proportional to the expected magnitude of the signal for a given DM interaction rate. The deepest limits come from searches targeting the dwarf spheroidal Milky Way satellite galaxies.

### *3.2. Gamma-Ray Signal from dark Matter*

The annihilation channels of DM particles are unknown. They could annihilate into pairs of Standard Model particles, like neutrinos, quarks, charged leptons, or gauge bosons. Except for the stable particles, like electrons and positrons, annihilation products must hadronize. That leads to a cascade of electrons, protons, their anti-particles, and $\gamma$-rays, which can also be emitted by processes like inverse-Compton scattering from $e^+$ and $e^-$ interactions with radiation fields, and pion production. The differential flux of $\gamma$-rays from DM annihilation within a solid angle $\Delta\Omega$ is:

$$\frac{d\phi\left(\Delta\Omega\right)}{dE_\gamma} = \left(\frac{x\left\langle\sigma v\right\rangle}{8\pi}\frac{dN_\gamma}{dE_\gamma}\frac{1}{m_\chi^2}\right)\left(\int_{\Delta\Omega} d\Omega \frac{dJ_{ann}}{d\Omega}\right),$$

where $x = 1$ if DM is its own antiparticle and $x = 1/2$ if it is not, $\langle\sigma v\rangle$ is the velocity-averaged annihilation cross-section, $m_\chi$ is the WIMP mass, $dN_\gamma/dE_\gamma$ is the differential spectrum of $\gamma$-rays from the annihilation of a pair of DM particles. $J_{ann}$ is the J-factor, which depends on the spatial distribution of the DM, i.e. the density of the DM particles $\rho(\vec{l})$, and the distance to the DM target: $dJ_{ann}/d\Omega = \int_{l.o.s.} dl\rho^2(\vec{l})$. The integration is performed along the line of sight. The total J-factor integrated over the solid angle of a region of interest $\Delta\Omega$ is

$$J\left(\Delta\Omega\right) = \int_{\Delta\Omega} d\Omega \int_{l.o.s.} dl\rho_\chi^2(\vec{l})$$

and is typically quoted for a given target [8].

## 4. Andromeda galaxy

One of the targets for DM searches with $\gamma$-rays is Andromeda galaxy (M31), the nearest major galaxy to the Milky Way and the biggest galaxy in the Local Group. It is a spiral galaxy at the distance 780 kpc from the Earth. It contains $\sim 10^{12}$ stars and has a diameter of approx. 67 kpc. The Andromeda Galaxy is among the brightest of the Messier objects. Its apparent magnitude is 3.4. Apparent dimensions of the Andromeda galaxy (3.2° by 1°) allow us to study its internal structure in detail. High energy (HE) $\gamma$-ray detection data from CTA can be used to find any hint to the DM annihilation or decay.

Table 1: WIMP targets. J-factors are proportional to the expected magnitude of the signal for a given DM interaction rate. Table adopted from [11].

| Target | Distance (kpc) | J-factor ($GeV^2cm^{-5}$) | Angular Extent (°) |
|---|---|---|---|
| Galactic center/halo | 8.5 | $3 \cdot 10^{22}$ *to* $5 \cdot 10^{23}$ | $> 10$ |
| Known Milky Way satellites | 25 to 300 | $3 \cdot 10^{17}$ *to* $3 \cdot 10^{19}$ | $< 0.5$ |
| Dark satellites | up to 300 | up to $3 \cdot 10^{19}$ | $< 0.5$ |
| Galaxy clusters | $> 5 * 10^4$ | up to $1 \cdot 10^{18}$ | up to $\sim 3$ |
| Cosmological DM | $> 10^6$ | - | isotropic |

M31 is well studied in all wavelengths, so it is an ideal target to study diffuse $\gamma$-ray emission. Two main contributors to the very high energy (VHE) emission from M31 are the diffuse emission from the interaction of cosmic rays with the interstellar medium and unresolved point-source emission. $\gamma$-ray luminosity of normal galaxies (without active galactic nuclei) depends on the number of cosmic ray accelerators, which depends on the star formation rate (SFR), on the escape time of cosmic rays, and the amount of target material. HE $\gamma$-ray flux from M31 was studied by VERITAS [14], HAWK [15], Fermi Large Area Telescope (Fermi-LAT) [16, 17], etc.. Relationship between existing flux measurements for 8 different galaxies and SFR is shown in the Figure 4. Approximately 55 hours of VERITAS observations of M31 are presented in Figure 5, giving upper limits on the total VHE emission. In 25 months of HAWK observations[15] no significant emission in the point sources and extended sources were found, so the upper limits have been computed for the disc and Fermi Bubble-like structures, shown in the Figure 6. The upper limits on Fermi Bubble structure in M31 are greater than the extrapolated flux of the Milky Way for 2 orders of magnitude, and 4 orders of magnitude smaller than the integral flux in the 0.3-100 GeV energy range observed by Fermi-LAT.

More than 7 year Fermi-LAT Pass data of M31 $\gamma$-ray emission was recently analyzed [16]. The results show that $\gamma$-ray emission from M31 is extended, but not through the whole galaxy. The flux is confined to the central parts of the galaxy, it does not fill the disc, and does not extend far from it. The signal is consistent with a uniform-brightness disc with an angular radius of 0.4°, 5 kpc from the center. Excessive $\gamma$-ray emission might originate from annihilation or decay of DM particles. This possibility was studied in [17]. In this study, using a more realistic assumption, that part of the flux comes from the galaxy (not only the bulge), no excess for the presence of DM was found, so the limits for annihilation cross-section were constrained to the thermal cross-section. Results for M31 and also for M 33 are shown in the Figure 7. $\gamma$-ray emission from M31, measured by Fermi, could also originate in the emission of many pulsars, that are expected to be present in the M31 bulge.

## 5. Cherenkov Telescope Array

The Cherenkov Telescope Array (CTA) is the next generation ground-based observatory for VHE $\gamma$-ray astronomy. It will be the world's largest and most sensitive HE $\gamma$-ray observatory, with more than 100 telescopes located in the southern and northern hemispheres (Figure 8). With its large collection area, the detection rate will be 10 times that of current instruments. CTA will have a large field of view and superior sensitivity (Figure 9). Sky survey will be several hundred times faster than with current telescopes.

Extreme cosmic environments like regions around neutron stars and black holes, or extreme cosmic voids between galaxies will be searched. CTA data will also help to explore the nature of DM and its distribution, to find out whether axion-like particles exist or not, and if there are quantum gravitational effects on photon propagation.

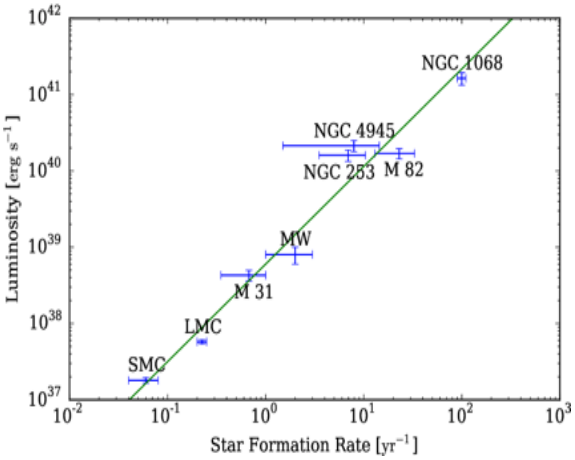


Figure 4: $\gamma$-ray luminosity (0.1 to 100 GeV) vs. SFR for "normal" galaxies overlaid with the best fit, luminosity $\propto$ (SFR)$^{1.28}$. Figure adopted from [14].

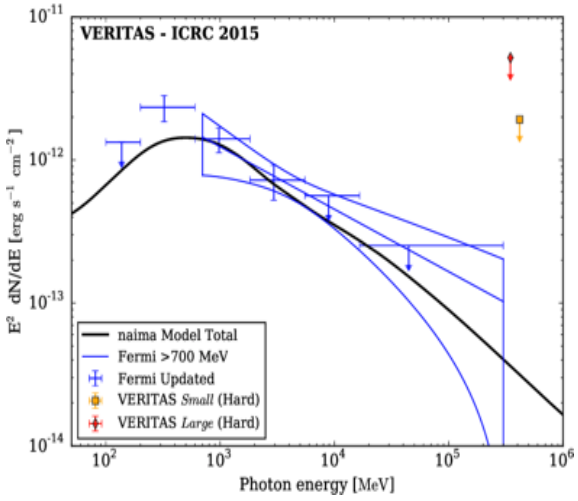


Figure 5: The VERITAS upper limits compared with the Fermi-LAT spectra and the GALPROP model. All upper limits are at 95% CL. Figure adopted from [14].

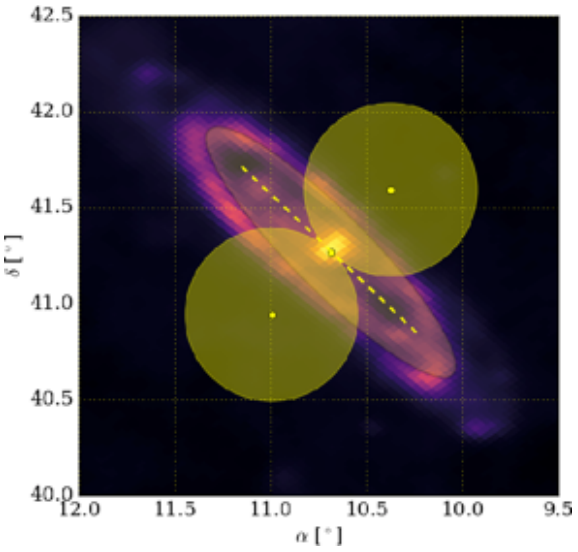


Figure 6: Regions used in the model for the Fermi Bubbles around M31. The M31 galactic plane is modeled as an ellipse with the two Fermi Bubbles as circular regions with radius 0.45°. The background image is the IRIS 100 µm map of the 2.5°× 2.5° region surrounding M31. The dashed line represents the M31 galactic plane. Figure adopted from [15].

### *5.1. Dark Matter Detection with CTA*

CTA will have a much greater potential for DM indirect detection than the current generation of VHE telescopes for several reasons:

1. extended energy range will allow searches for WIMPs with lower mass,
2. probability of detection of DM will be much higher because of the improved sensitivity in the entire energy range (Figure 10),
3. increased field of view and angular resolution will allow more efficient searches for extended sources and spatial anisotropies,
4. improved energy resolution will increase the probability of detecting a DM induced photon spectrum [18].

If direct search experiments or LHC will find any signatures of DM, CTA observations will be able to provide a complementary approach to identify DM

particles. But if direct experiments do not detect it (in case of heavy DM candidates), CTA may be the only way to search for such particles. Among others, star-forming systems like M31 will be searched by CTA, that can provide much more detailed $\gamma$-ray spectra and high-resolution images than Fermi-LAT. Fermi-LAT is sensitive to large-scale interstellar emission in the Milky Way and nearby galaxies, but CTA will probe particles close to their acceleration sites.

The Andromeda galaxy (M31) has a similar SFR as the Milky Way. Studying the similarities and differences in $\gamma$-rays will provide important insights on the cosmic ray properties of both galaxies. The ratio of $\gamma$-ray luminosity over SFR for M31 to that from the Milky Way is 3:1, and the GeV spectrum is noticeably harder. Compared to the Milky Way, the advantages of the Andromeda galaxy observations are the unbiased view and the fact that the SFR is much better known for M31. That is why observations of the Andromeda galaxy with CTA will provide invaluable insights into cosmic ray properties in spiral galaxies.

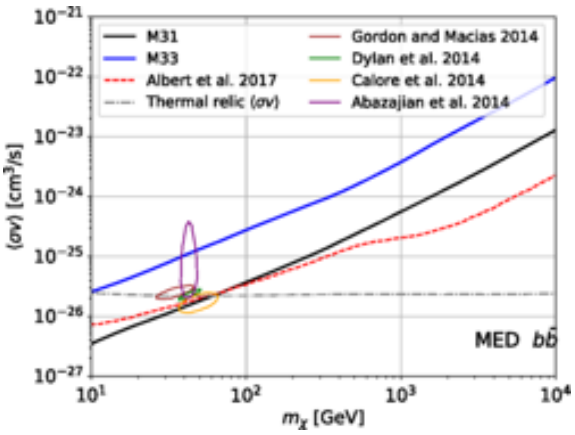


Figure 7: Comparison between the 95% CL upper limits from M31 (black solid line) and M33 (blue solid line) with the limits found from dwarf spheroidal galaxies (red dashed line). The confidence regions for cross-section and mass determined by analysis of the Galactic center excess are shown. The horizontal dashed line shows the thermal relic cross-section. Figure adopted from [17].

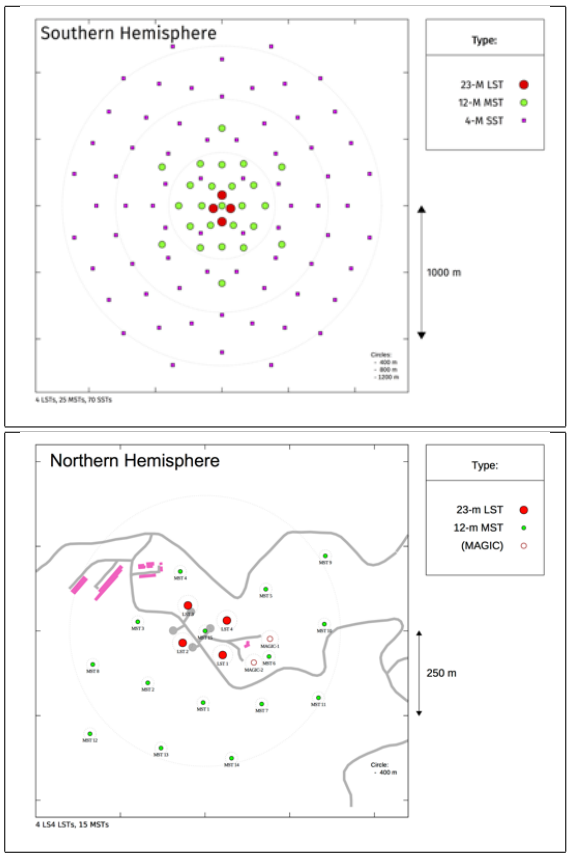


Figure 8: Top panel: The Southern Hemisphere Array in Paranal, Chile, will span the entire energy range of CTA, covering $\gamma$-ray energies from 20 GeV to more than 300 TeV. Bottom panel: The Northern Hemisphere Array in La Palma, Spain, will be more limited in size and will focus on the low and mid-energy ranges from 20 GeV to 20 TeV. Figures adopted from [18].

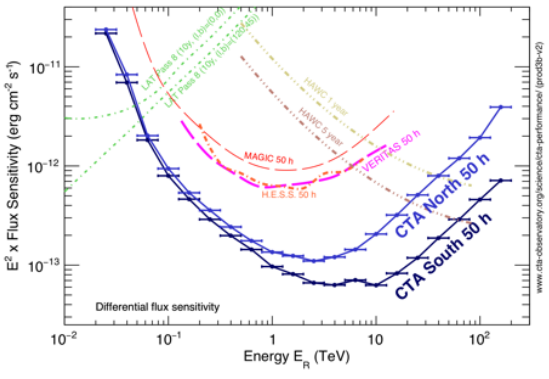


Figure 9: The sensitivity and angular resolution of the CTA arrays compared to a selection of existing $\gamma$-ray detectors. Figure adopted from [18].

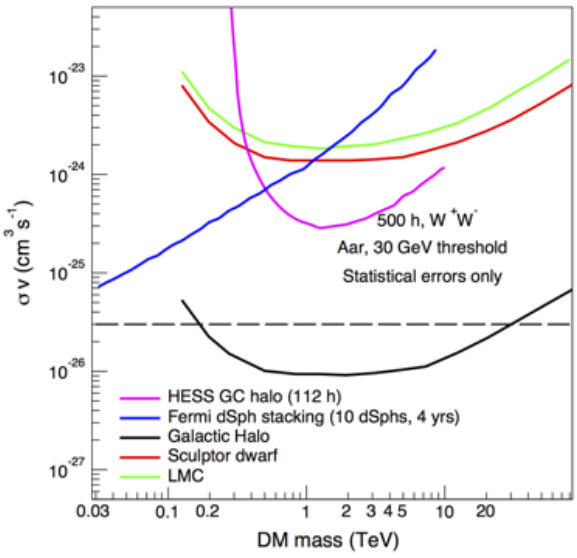


Figure 10: CTA sensitivity to a WIMP annihilation signature as a function of WIMP mass. The dashed horizontal line indicates the likely cross-section for a WIMP which is a thermal relic of the Big Bang. Figure adopted from [18].

### *5.2. CTOOLs*

*CTOOLs* is a publicly available software package developed for the analysis of Cherenkov Telescope Array (CTA) data, and also data from existing Imaging Air Cherenkov Telescopes (H.E.S.S., MAGIC, VERITAS). *CTOOLs* are based on GammaLib, a versatile toolbox for the analysis of astronomical $\gamma$-ray data. Besides CTA, GammaLib supports also the analysis of Fermi-LAT and COMPTEL data [19]. With *CTOOLs* one can simulate the possible astrophysical signal from M31 with CTA, based on the models that can explain the extended emission measured by the Fermi-LAT. For different DM models, simulations can predict whether CTA will be able to detect an extended signal from M31, and what would be the sensitivity of the CTA to constrain the DM models in the direction of the Andromeda galaxy.

## 6. Problems and recent Discoveries

$\Lambda CDM$ scenario successfully describes large structures of the Universe. N-body simulations show that the DM density profile of the virialized structures like galactic halos is universal and the Navarro-Frenk-White (NWF) profile describes it well. But there are problems at the galactic scales:

1. "missing satellite problem": number of the detected satellites around the Milky Way and the predictions of the corresponding simulations do not match,
2. "too big to fail problem": when masses of the detected satellites are compared to the predicted subhalos' masses, discrepancy widens up,
3. the "core-cusp" controversy: the measured inner DM density profiles of galaxies are cored, and not cuspy as predicted in the simplest $\Lambda CDM$ scenario.

Discrepancies between observations and predictions are present in all galaxies, but they are larger in low luminosity systems [20].

### *6.1. Rotation Curves (RCs)*

The RCs of more than 5000 local spiral galaxies were measured within $z \sim 0.1$. They do not have the same profile, but they can be described by the same universal rotation curve (URC) which is a function of the radius with free parameters: luminosity $L$ (or magnitude, or mass) and characteristic radius of the luminous matter $R_{opt}$: $v(R) = v(R, L, R_{opt})$. URC provides the mass distribution of normal disc galaxies in the magnitude range $-23.5 < M_I < -17.2$. The main structural properties of luminous matter and DM of spiral galaxies are encoded in URC.

The URC of dwarf disc galaxies needs three free parameters: luminosity, optical radius, and also compactness $C$: $v(R) = v(R, L, R_{opt}, C)$ [20]. The additional

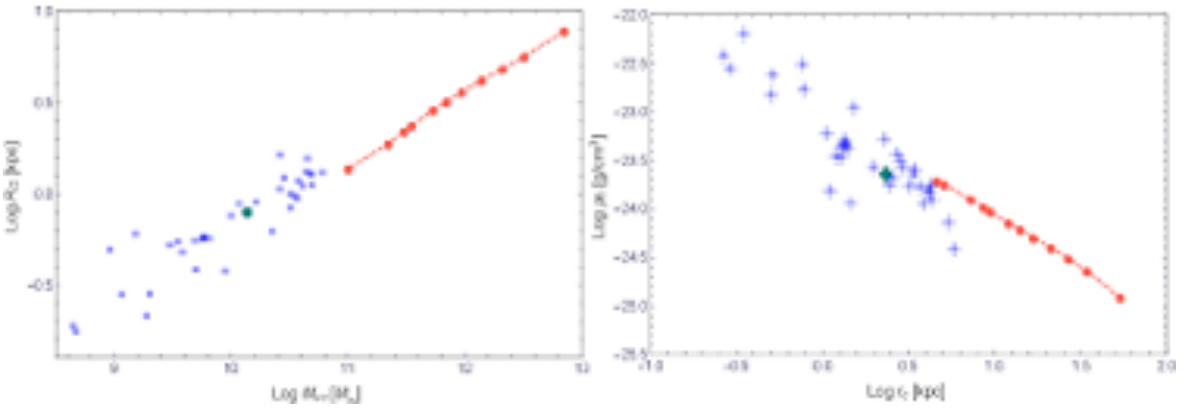

Figure 11: Left panel: the disc scale lenght versus virial mass. Right panel: the central density versus core radius. Red circles: normal spirals, blue squares: dwarf discs, green circle: the average point. Figure adopted from [20].

parameter $C$ was introduced because of the large scatter due to different sizes of stellar discs in dwarf disc galaxies with approximately the same stellar mass and luminosity (Figure 11). Compactness $C$ is the ratio between stellar disc radius $R_D$ predicted from the linear regression $\log R_D = -3.64 + 0.46 \log M_D$, and $R_D$ measured from photometry,

$$C = \frac{10^{(-3.64+0.46 \log M_D)}}{R_D}.$$

With $C$ the universality is restored (Figure 12):

$$\log R_D = -3.99 + 0.38 \log M_{vir} - 0.94 \log C.$$

The compactnesses of the stellar disc $C$ and the DM halo

$$C_{DM} = \frac{10^{(-5.08+0.53 \log M_{vir})}}{R_{DM}}$$

are closely related (Figure 13).

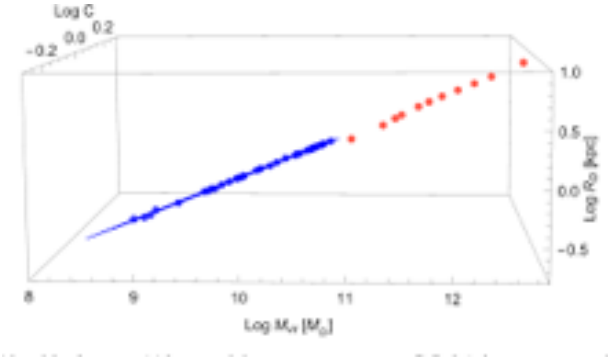

Figure 12: The disc scale lenght versus virial mass and the compactness parameter C. Universal rotation curve exists for a variety of galaxies, from dwarfs to giant ones. Red circles: normal spirals, blue squares: dwarf discs, blue line: result of the fit. Figure adopted from [20].

### *6.2. Results*

The results show:

1. gas fraction in low-luminosity and disc galaxies is $f = M_{HI}/M_{bar} \sim 0.8$, where $M_{HI}$ is mass of gas (HI) and $M_{bar}$ is baryonic mass,
2. surface density of DM halo is constant for objects from dwarf to giant galaxies over 18 blue magnitudes: $\rho_c r_c \simeq 75\, M_{\odot} pc^{-2}$,
3. structural parameters of the dark and luminous matter (LM) do not explicitly depend on luminosity,

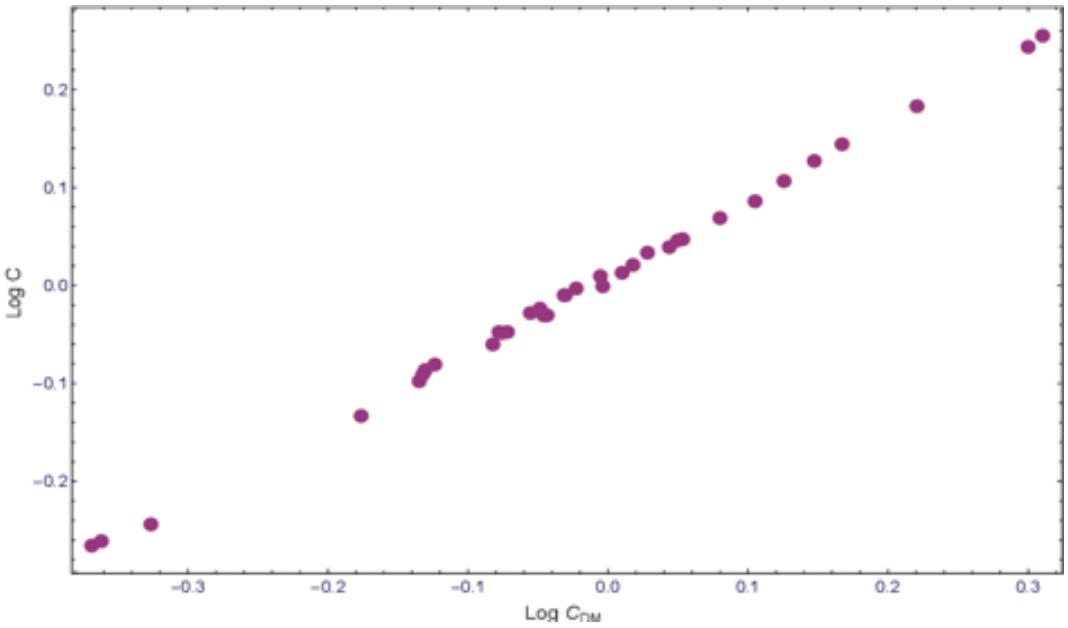


Figure 13: The compactness of the stellar disc versus the compactness of the DM halo. The compactnesses are strongly correlated. The distribution of DM and the stellar distribution follow each other very closely. Figure adopted from [20].

4. DM is cored and the density core is a bit bigger than the stellar disc: $r_c \sim 1.5\,R_D$ to $3\,R_D$,

5. the compactnesses of the stellar disc and of the DM halo are proportional in log-log scale,

6. DM and the stellar distribution are strongly correlated [20].

Recent findings on the rotating Low Surface Brightness (LSB) galaxies gave similar results and also confirm strong entanglement between the luminous and the dark matter [21]. Such strong correlations between LM and DM in all disc systems strongly challenge the $\Lambda CDM$ scenario with collisionless dark particles. A new paradigm for the DM phenomenon is proposed, in which the dark matter scenario should be searched for from reverse-engineering observations at galactic scales [22].

## 7. Conclusions

Such a strong correlation between the dark halo and stellar disc compactness implies that interactions between DM and luminous matter are not weak at all, and it is very unlikely that WIMPs make the DM in galaxies. That is why W-W annihilation and decay products are probably not what we should look for. Perhaps we should change the paradigm and search for something completely different. Even if WIMPs do not exist at all, $\gamma$-ray emission from M31 will provide information about properties of both, dark and luminous matter in the galaxy. Data obtained by CTA will help to test dark matter and cosmological theories and will be very important to resolve the mystery.

## Acknowledgments

I would like to thank the reviewers for useful comments and suggestions that helped me to improve the article.

## References


[1] V. Rubin, W. K. Ford, Jr., "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions". The Astrophysical Journal. 159: 379ff, (1970).

[2] W. K. Huchtmeier, Rotation-curves of galaxies from 21 cm-line observations, Astronomy and Astrophysics, vol. 45, no. 2, p. 259-268, (1975).

[3] P. Salucci, The distribution of dark matter in galaxies, Astron Astrophys Rev 27, 2, (2019).

[4] Galaxy Cluster Abell S1063 and Beyond, Image Credit: NASA, ESA, Jennifer Lotz (STScI).

[5] P. Mészáros, P. The High Energy Universe: Ultra-High Energy Events in Astrophysics and Cosmology, Cambridge University Press, New York, (2010).

[6] M. G. Aartsen et al. (IceCube Collaboration), Searches for Sterile Neutrinos with the IceCube Detector, Phys. Rev. Lett. 117, 071801, (2016).

[7] M. G. Aartsen et al. (IceCube Collaboration), Search for sterile neutrino mixing using three years of IceCube DeepCore data, Phys. Rev. D 95, 112002, (2017).

[8] E. Charles, Sensitivty Projections for Dark Matter Searches with the Fermi Large Area Telescope, (2016), doi:10.1016/j.physrep.2016.05.001.

[9] O. Adriani et al. (PAMELA Collaboration), Observation of an anomalous positron abundance in the cosmic radiation, (2008), doi:10.1038/nature07942.

[10] F. Wilczek, The Lightness of Being: Mass, Ether, and the Unification of Forces, Basic Books, New York, (2008).

[11] E. Charles, Indirect Dark Matter Searches: Status and Prospects, PoS (EDSU2018) 012, (2018).

[12] M. Tanabashi et al. (Particle Data Group), Phys. Rev. D 98, 030001 (2018), The Review of Particle Physics, Axions and Other Similar Particles, 111.

[13] A. Payez et al., Revisiting the SN1987A gamma-ray limit on ultralight axion-like particles, Journal of Cosmology and Astroparticle Physics (JCAP) 1502, 006, (2015).

[14] R. Bird (VERITAS Collaboration), VERITAS Observations of M31 (The Andromeda Galaxy), (2015), doi:10.22323/1.236.0851.

[15] R. Rubenzahl, S. BenZvi, J. Wood (HAWC Collaboration), Limits on the Emission of Gamma Rays from M31 (The Andromeda Galaxy) with HAWC, (2017), arXiv:1708.03012v1.

[16] M. Ackermann et al., Observations of M31 and M33 with the Fermi Large Area Telescope: A Galactic Center Excess in Andromeda?, The Astrophysical Journal, American Astronomical Society, (2017).

[17] M. Di Mauro et al., Search for gamma-ray emission from dark matter particle interactions from Andromeda and Triangulum Galaxies with the Fermi Large Area Telescope, (2019), doi:10.1103/PhysRevD.99.123027.

[18] B.S. Acharya et al. (The Cherenkov Telescope Array Consortium), Science with the Cherenkov Telescope Array, (2018).

[19] J. Knoedlseder et al., CTA Science Tools, A&A, 593, A1, (2016), doi:10.1051/0004-6361/201628822.

[20] E. Karukes, P. Salucci, The universal rotation curve of dwarf disc galaxies, (2016), doi:10.1093/mnras/stw3055.

[21] C. Paolo, P. Salucci, Fundamental properties of the dark and the luminous matter from Low Surface Brightness discs, (2020), arXiv:2005.03520.

[22] P. Salucci, N. Turini, Paradigms and Scenarios for the Dark Matter Phenomenon (2020), arXiv:2003.01424.

# Element specific magnetization dynamics using X-ray magnetic scattering

Arun Ravindran

*Laboratory of Quantum Optics, University of Nova Gorica, Vipavska 11c, SI-5270 Ajdovščina*

**Abstract**

Femtosecond magnetism is the study of magnetization dynamics using ultrashort laser pulses. Magnetization studies in the ultrashort timescales are of much importance due to the exciting new fundamental physics it can uncover and also because of the practical applications such materials offer in areas like data storage and transfer. The use of X-ray magnetic scattering technique for the study of ultrafast demagnetization dynamics has gained popularity due to development of technologies that can produce ultrashort tunable XUV radiations with high efficiency. In this paper basic principles of the technique will be reviewed and relevant works compared.

*Keywords:* magnetic materials, X-ray scattering, pump-probe, High Harmonic Generation

## 1. Introduction

The ultrafast study of magnetization dynamics has been acquiring interest for the past few decades. This can be said to have started by the pioneering work of Beaurepaire *et al.* [1], who found that ultrashort light pulses can induce changes in the magnetization dynamics of a material at femtosecond time scales. Since then, many works have been carried out to understand the dynamics of magnetic materials at ultrashort timescales from a purely scientific interest point of view and also due to their varied technological applications [2, 3, 4, 5, 6].The studies are usually carried out in the pump-probe setup where the sample is initially excited to a non-equilibrium state

*Email address:* arun.ravindran@ung.si (Arun Ravindran)

using a pump laser beam and subsequently the dynamics are studied using a probe beam.

The interaction of an ultrashort laser pulse on a magnetic sample can give rise to many processes [6]. These can be spin-photon interactions in the lowest timescales of around 10 fs, to scattering processes and diffusive currents in the larger femtosecond scales, thermal equilibrium by lattice absorptions in the picosecond regions and, finally, thermal diffusion leading to cooling of the material in the nanosecond regions. All these processes collectively contribute to the ultimate magnetization dynamics in the material. Identifying the contributions of each process to the dynamics in the material has been a key topic of interest for researchers. This can give answers to fundamental questions like how fast can the magnetic reordering process, inside the material, take place after excitation from an ultrafast laser pulse. To do such distinct studies one would require to generate very short coherent pulses, possibly in the femtosecond ranges.

Another interesting topic to look at, is how the magnetic materials will behave, when they are put together in a coupled system, such as an alloy. Such element-specific studies require the light source to be tuned at specific absorption edges pertaining to the element under observation.

In short, a good experimental setup for this kind of research would require a source that has tunable energies close to the absorption edges of the specific elements and also it should have the shortest possible pulse durations. In this paper we will review the principles of X-Ray magnetic scattering technique and how they are useful for the element specific demagnetization dynamic studies. The basic principles of Magneto Optic Kerr Effect and High Harmonic Generation process are explained. A description of the experimental setup is given. Finally, a comparison of works done in this area of research is presented.

## 2. Magneto Optic Kerr Effect (MOKE)

The Magneto Optic Kerr effect arises from the interaction between a polarized light source and a magnetic surface. This interaction can induce changes in the polarization and\or the intensity of reflected light. This is mainly due to the magnetic anisotropy in the material. There are three general geometries of MOKE, namely, polar, longitudinal and transverse based on the orientation of magnetization in the material, the plane of polarization of the incident beam and the reflecting surface. In the transverse

geometry, the magnetization is parallel to the reflecting surface and is perpendicular to the plane of incidence (Figure 4). Here, a change in the magnetization of the material induces a change in the intensity of the reflected light. The intensities of light reflected from the sample surface are given by the equations

$$I_{\pm}^{P} = I_0 \left| \frac{ncos\theta_i - cos\theta_t}{ncos\theta_i + cos\theta_t} \pm \frac{2sin\theta_i cos\theta_i}{n^2 \left(ncos\theta_i + cos\theta_t\right)^2} \epsilon_{xy} \right|^2 \tag{1}$$

$$I^{S} = I_0 \left| \frac{cos\theta_i - ncos\theta_t}{cos\theta_i + ncos\theta_t} \right|^2 \tag{2}$$

where, $I^P, I^S$ are the intensities of reflected beam for P-polarized and S-polarized incident beams respectively. The $\pm$ sign represents the signal for both directions of the applied external magnetic field.$\theta_i$ and $\theta_t$ are angle of incidence and refraction respectively, $n$ is the refractive index of the magnetic material and $\epsilon_{xy}$ is the off-diagonal term of the dielectric tensor [7]. The equations 1 and 2 are obtained by applying the electromagnetic boundary conditions at the interface between vacuum and magnetic material. A phenomenological derivation of these equations are given in [8]. From eq.( 1), the first term is the non-magnetic fresnel coefficient and the second term has a dielectric dependance. This dielectric tensor term, also called the magneto-optic constant, is responsible for the variation in reflected intensities depending on the magnetization in the material. Also, there are no magnetic contribution terms in eq.( 2), and hence we will not see any change in the reflected intensities for the S-polarized light sources.

The Kerr effect is characterized by measuring the reflected intensities for both directions of magnetization and calculating the asymmetry ratio given by

$$\begin{aligned} A =& \frac{I_{+}^{P} - I_{-}^{P}}{I_{+}^{P} + I_{-}^{P}} \\ =& 2Re\left[\frac{sin(2\theta_i)\epsilon_{xy}}{n^4 cos^2(\theta_i) - n^2 + sin^2(\theta_i)}\right] \end{aligned} \tag{3}$$

Measuring this asymmetry parameter gives information about the magnetization dynamics in the material.

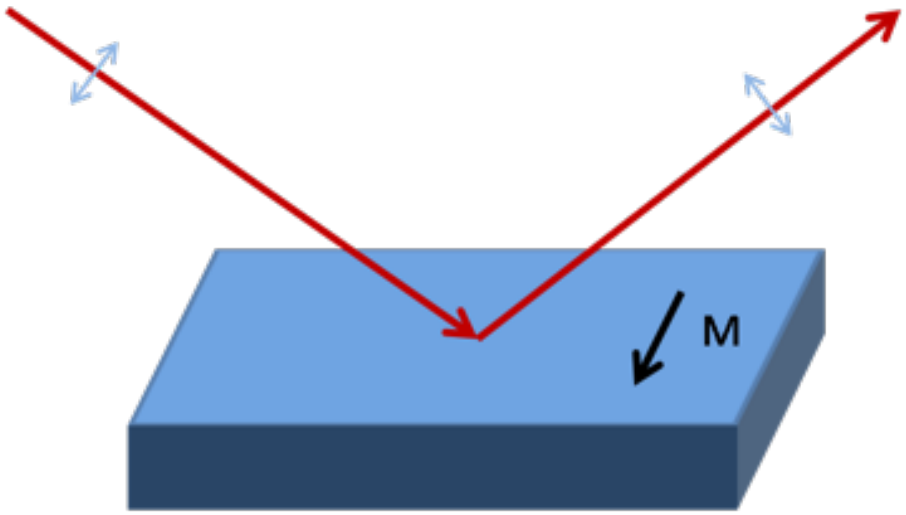


Figure 1: Illustration of the transverse MOKE geometry. The laser light is incident on the sample which is placed in an external magnetic field to contorl the magnetization inside it. The polarization of light is denoted by the arrows on the incident and reflected rays. The magnetization direction on the sample surface is also denoted by M.

## 3. Experimental techniques and setup

### *3.1. The source criteria*

MOKE measurements can be done in visible as well as X-ray ranges of the electromagnetic spectrum. However, visible MOKE has the drawback of not providing element specific measurements. By tuning the source at resonance energies of the absorption edges of the elements in the material, we can differentiate the asymmetry signal of specific elements acquired in the measurement. This enables us to identify the specific contributions of different elements and their respective magnetization dynamics in the material. For the case of M-absorption edges of Fe and Ni, this falls into the Extreme ultra violet(XUV) range of the spectrum. Several processes contribute towards the magnetic signal from a material at different timescales ranging from femtoseconds to nanoseconds. Ultrashort light pulses are required to identify the effects of each process distinctly.

Taking these requirements into consideration, a good source would have very short pulse widths, of the order of a few femtoseconds, and be preferably in the XUV spectral range. A good control over the polarization is also necessary and the energy of the pulse should be tunable.

Commercially available Ti:Sapphire mode-locked laser amplifier systems can produce ultrashort laser pulses in the timescales of 10-100 fs. These are very good systems for ultrafast optical studies. However the pulses produced are in the visible-IR range, where element specific studies are not possible. Free electron lasers, femtoslicing synchrotron radiation and High harmonic generation by an ultrashort laser pulse are some of the source techniques that are suitable for such kind of studies [9]. Even though X-ray Free electron Lasers can produce high powered ultrashort pulses with good intensities, they are not easily available everywhere in the world and getting access to a beamline in a large facility can be a quite tedious process. Femtoslicing synchrotron beamlines to produce ultrashort XUV pulses has the disadvantage of very low photon flux at the material surface. High order harmonic generation(HHG) using an ultrashort laser pulse on the other hand, has been employed in recent studies and is capable of producing ultrashort coherent pulses in the soft X-ray ranges and at very high repetition rates. It has also the added advantage of being a simple table-top setup which is technologically available to laboratories everywhere.

### *3.2. Principles of High Harmonic Generation*

HHG is a non-linear process where a medium is excited using a linearly polarized high power laser source. After the interaction between the laser field and the medium, the medium re-emits light in the higher harmonics of the generating beam. This process is explained in a simple three-step model (refer Figure 2). In the first step, the incoming laser field excites the atoms in the medium and pulls out the electron from the outer shell by tunneling ionization. Next, the tunnelled electron starts oscillating in the laser field of the generating beam. In the final step, after half a cycle of the incoming laser pulse, the oscillating electron recombines with the parent atom, emitting radiation in the higher harmonics of the generating laser pulse. By this process one can generate higher order harmonics in the soft X-ray regime with inert gases (Ar, Ne) as the target material in the generation chamber. Harmonics produced by this process are synchronized with the generating laser pulse and has the same repetition rate. The saturation intensity of the harmonics generated depend on the target material used. For higher harmonics, lighter elements have to be chosen, which in turn may reduce the efficiency of the generation. So, the choice of target material depends on the energies required for the experiment. The generated

beams will have similar spatial and temporal coherence properties to that of the driving laser. They are emitted collinearly with the driving laser and can have very small divergence. These properties make the use of an HHG source highly efficient for resonant magnetic scattering studies.

The laser system at Quantum Optics Laboratory (LKO), University of Nova Gorica makes use of a micra oscillator producing 800 nm light beam with a spectral bandwidth of about 100 nm and power of 380 mW [10]. It has two stage amplification process and produces mode-locked ultra-short pulses with around 35 fs pulse duration and 5 kHz repetition rate. A typical HHG spectrum produced in LKO lab is given in Figure 4 of [10].

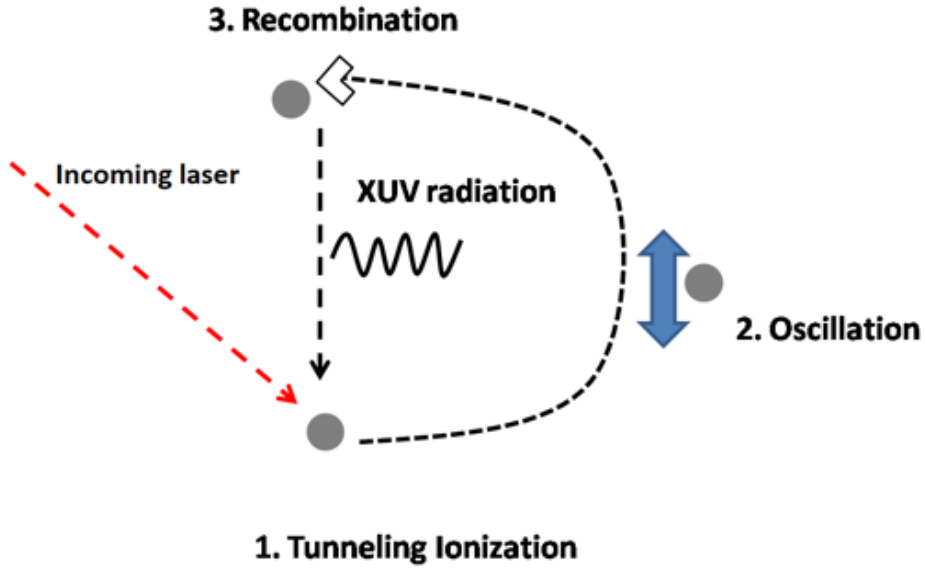


Figure 2: A simple illustration of the three step model for High Harmonic Generation. The incoming laser takes out electron from the parent atom. The ionized atom oscillates in the laser field and finally recombines with the parent atom to give out radiation at higher energies.

### *3.3. Beamline system and measurement procedures*

The time-resolved X-ray magnetic scattering(tr-XRMS) studies are generally carried out in a pump-probe setup making use of the Magneto Optic Kerr effect. Pump-probe technique is a widely used tool for analyzing ultrafast dynamics in very short timescales. Here, the sample is initially

excited to a non-equilibrium state by a pump laser and then the dynamics at ultrashort timescales are recorded using a probe beam. By changing the delay time of the pump pulse w.r.t the probe pulse, the dynamics of the relaxations and ultrafast processes inside the material can be analyzed. A common way of recording the probe beam without any effects from the pump laser is by sending the pump and probe beams in different wavelengths. The XRMS studies use an IR pump beam and an XUV probe beam. They are usually beams split from the same source. The pump beam goes through a delay line system while the probe beam goes into the harmonic generation chamber. A schematic of the XRMS setup currently under commissioning at LKO is provided in Figure 3.

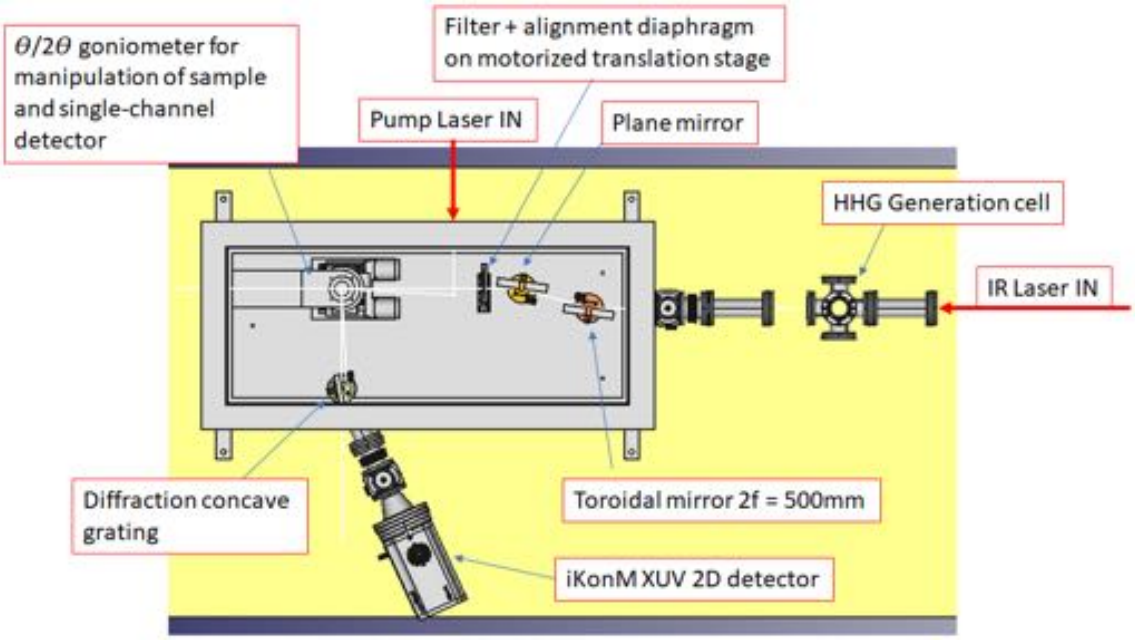


Figure 3: Schematic diagram of the tr-XRMS setup in LKO.

A portion of IR laser split after the amplification stage is focused onto the harmonic generation cell. The other portion travels through the delay line system and gets refocused onto the sample surface. The sample is placed in an external magnetic field generated using an electromagnet. After reflection from the sample surface, the white probe beam is spectrally dispersed onto a 2-D detector using a diffraction grating. This enables us to analyze the magnetic signal of the entire HHG spectrum at the same

time and hence only the delay time needs to be varied for successive measurements. Reflected intensities are measured for both directions of magnetization and at different delay times. The asymmetry signal obtained is plotted against the photon energy to get a clear picture of the absorption edge of each element present in the system. An example of such a plot is shown in Figure 4. Here we can see clearly that the magnetic signals from Fe and Ni are separated according to their absorption-edge energies. Hence, time-resolved XRMS studies at the resonance conditions for the absorption edges can give insight into the element specific contributions to the magnetic signal of a material containing more than one magnetic element.

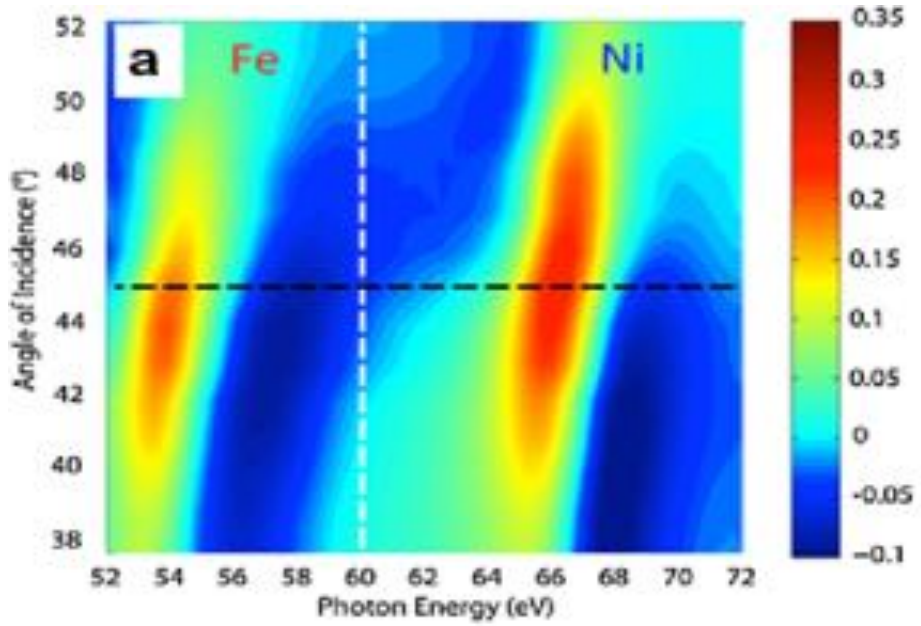


Figure 4: The magnitude of asymmetry signal plotted in a color-coded graph against photon energy and angle of incidence (from [6]). Here we can see the seperation of the absorption edges for Fe and Ni clearly.

## 4. Research and Outlook

One of the simplest and common system, which can be used for element specific magnetization studies, is an alloy of Ni and Fe, called permalloy.

It contains 80% of Ni and 20% of Fe. An interesting phenomena, that has been studied for this material, is how the demagnetization time changes for Fe and Ni when they are in an alloy system, compared to when they are in their elemental form. Several studies have been carried out to analyze this effect producing varied results. [2, 3, 4, 6] The works of Radu *et al.* [4] and Mathias *et al.* [6] can be taken as the general categories of the two different approaches to this question and their contradictory results.

Mathias *et al.* measured the magnetic asymmetry signal of the permalloy sample in the reflection geometry. They used a transverse MOKE setup with the probe pulse tuned at the M-absorption edges of Fe and Ni. In elemental forms , they found that Ni was taking more time to demagnetize (157 fs) than Fe (98 fs). The quenching was much higher in Ni samples than in Fe samples for the same pump fluence. However, in the permalloy system, demagnetization curves observed for Fe and Ni were similar. There was only a small difference in the demagnetization time were Fe demagnetized faster by around 10-20 fs. This suggests a strong coupled exchange interaction between the two elements in the coupled alloy system. Radu *et al.* carried out magnetic scattering measurements in the transmission geometry. They used circularly polarized light source tuned at the L-edges of Fe and Ni. A quicker demagnetization for Ni compared to Fe was reported. They also tried to explain the probable causes for the difference in the measured dynamics. The first argument is based on the geometry of the measurement. The signal acquired after reflection from the sample may contain absorptive and dispersive elements, which prevents a direct quantitative measurement of the elemental magnetic moments. Secondly, the energy ranges for the M-absorption edges falls into the VUV spectral range, where the energies for the spin-orbit and exchange interactions are of comparable magnitudes. This may produce distortions in the 3p electronic states and hence the M3 and M2 edges maybe overlapped. In such a scenario, one cannot use the magneto-optical sum rules to acquire the magnitude of the magnetic moment. Also, at these energy ranges the absorption edges of Fe and Ni are close to each other, increasing the possibility for their respective magnetic signals to overlap. Hence, one loses the ability to distinguish the specific contributions of different elements to the obtained signal.

On this note, there exists a need to address the issue of factors affecting the measured asymmetry signal of a multi-elemental system. A study, which can carry out such measurements in both reflection and transmis-

sion geometry in the same lab, can give a comprehensive and definite understanding of the parameters affecting the measured signals. As an initial step, at LKO, we are planning to carry out the magnetic scattering measurements of permalloy samples with the signal tuned at M-absorption edges of the element. A setup for measuring in both transmission and reflection geometries, so as to compare the results, is also being thought of.

## 5. Acknowledgement

This work was supported by the Young Researchers Program of Slovenian Research Agency.

## References

[1] E. Beaurepaire, J.-C. Merle, A. Daunois, J.-Y. Bigot, Phys. Rev. Lett. 76 (1996) 4250–4253. doi:10.1103/PhysRevLett.76.4250.

[2] C. La-O-Vorakiat, Phys. Rev. Lett. 103 (2009) 257402. doi:10.1103/PhysRevLett.103.257402.

[3] S. Jana, Review of Scientific Instruments 88 (3) (2017) 033113. doi:10.1063/1.4978907.

[4] I. Radu, SPIN 05 (03) (2015) 1550004. doi:10.1142/S2010324715500046.

[5] S. Günther, Phys. Rev. B 90 (2014) 180407. doi:10.1103/PhysRevB.90.180407.

[6] S. Mathias, Journal of Electron Spectroscopy and Related Phenomena 189 (2013) 164 – 170. doi:10.1016/j.elspec.2012.11.013.

[7] H. Höchst, D. Rioux, D. Zhao, D. L. Huber, Journal of Applied Physics 81 (11) (1997) 7584–7588. doi:10.1063/1.365303.

[8] M.Freiser, IEEE Transactions on Magnetics 4 (2) (1968) 152–161. doi:10.1109/TMAG.1968.1066210.

[9] T. Pfeifer, C. Spielmann, G. Gerber, Reports on Progress in Physics 69 (2) (2006) 443–505. doi:10.1088/0034-4885/69/2/R04.

[10] C. Grazioli, Review of Scientific Instruments 85 (2) (2014) 023104. doi:10.1063/1.4864298.