Atomic and Molecular Data for Plasma Models

Shuaiyu Fan

Student number: 19139508

MSc Scientific Computing

Dept of Physics and Astronomy

University College London

London

ucapfan@ucl.ac.uk

ABSTRACT

The QDB database has been created to provide a repository for collecting atomic and molecular data about plasma. However, the reactions now included in the database is not sufficient for plasma modelling. And the process of add new data is continuous. For this purpose, we do data mining for a given list of species to extract data from the NIFS database and the NFRI database.

For both databases, we retrieve the basic information as well as sets of numerical data for collision processes. And we extract a total number of 9466 reaction records from the two databases, including 4105 reactions in the NIFS database and 5361 reactions in the NFRI database. Since that there are no data related to photon impact in the NIFS database, all the data retrieved from this database belong to heavy particle impact processes. As for the NFRI database, we obtained the information on data about 2586 cross sections and 2775 rate coefficients for both heavy particle impact and photon impact. Meanwhile, we further classify the retrieved data by experimental data, theoretical data, and evaluated or recommended data for each database, and do statistics for this classification. For the NIFS database, the data obtained from experiments account for the largest proportion; while the number of data calculated by theoretical methods is the largest in the NFRI database.

KEYWORDS

Plasma chemistry; Atomic and Molecular data; Database; Data mining; Python

ACKNOWLEDGE

Thanks to Professor Jonathon Tennyson for providing me with the opportunity to participate in this project, and the helpful guidance and support during my work. Doctor Bridgette Cooper and Doctor Stephen Fossey also do me a favor in project section and work arrangement.

In this project, we retrieve needed data from the NIFS database and the NFRI database. Thank the National Institute of Nuclear Fusion Science (NIFS) of Japan and the National Fusion Research Institute (NFRI) of Korea for supplying the accessible databases to the public online.

In addition, due to the outbreak of COVID-19, both supervisors and students have experienced unprecedented challenges this year. And after returning to China, the poor physical and mental conditions have greatly hindered and delayed my project progress. I am very thankful to my supervisor Professor Jonathon, Doctor Bridgette and Doctor Stephen for their kind help and understanding in the process.

Finally, I would like to express my gratitude to all those who helped and accompanied me in my life and study at University College London during this year. Learning here as a student of MSc Scientific Computing is a precious and unforgettable memory in my life.

LIST of TABLES

LIST of FIGURES

CONTENT

1. INTRODUCTION

With the development of plasmas in industrial applications, more and more research groups pay attention to the study of plasmas. Based on atomic and molecular data, scientists are able to create plasma models to replicate the realistic process of driving plasmas at the sub-microscopic level [1-2]. These models can provide guidance for the generation and application of plasmas, so designing a reasonable model is of critical importance. In order to improve the reliability and practicality of plasma models, it is necessary to improve the accessibility and accuracy of atomic and molecular data.

There are an increasing number of databases related to plasmas. These databases aim to collect as much plasma-related reactions as possible, and then use the existing reactions and parameters extracted from these reactions for plasma modelling.

The Quantemol Database (QDB) [3] is created to collect both cross sections and rate coefficients for key collision processes. Particularly, for low temperature plasma [4-9], in order to model plasma accurately, it is important to assemble the appropriate data set. At present, the QDB database has a total number of 5000 reactions. However, the data now included in the database are far from enough and the process of collecting is continuous. For this reason, we do this project to do data mining for some existing databases, retrieve more atomic and molecular data for plasma modelling.

The following section gives the basic information about plasmas. Section 3 introduces sources of data used for this project while section 4 gives an overview of the QDB database and also classifies the types of reaction processes. Section 5 explains the aims of this project. Section 6 describes the methods and program in detail. Section 7 summarizes the retrieved data and shows the results, and the last section provides conclusions and discusses the possible work in the future research.

2. BASIC and FUNDAMENTALS

2.1 The definition of plasmas

Plasmas are a form of ionized gas-like substance being composed of both n ions, electrons and neutrals in both ground and excited states. At the macro level, these ionizing gases are electrically neutral. Plasmas are widely found in the universe and are often regarded as the fourth state of matter in addition to solid, liquid and gas. Plasmas are not only highly conductive, but also highly coupled with electromagnetic fields.

2.2 The generation of plasmas

Plasmas can be created by applying energy to the gas [**10**]. In this way, the electronic structure of atoms or molecules can be reorganized, and therefore species and ions are excited. Specifically, we are able to use thermal energy, electric energy or energy carried by electromagnetic radiations [**11**].

There are many means to produce low temperature plasma, including ultraviolet radiation, X-ray, electrical discharge field, heating and so on. For most of the scientific research and industrial applications, electromagnetic fields such as DC glow discharges [**12**], radio frequency discharge [**13**] (low frequency AC discharges and high frequency discharges), microwave discharge [**14**], and dielectric-barrier discharge (DBD) [**15**], are used to excite plasmas.

2.3 The classification of plasmas

(a) The impact of temperature on plasma

Plasmas can be divided into two types: one is the low-temperature plasma and the other is the high-temperature plasma [**16-18**]. The electron temperature and the ion temperature can be combined to represent the plasma temperature. The ionization rate of the high temperature plasma is close to 100%. In general, low temperature plasmas have different electron and ion temperatures. And the ionization rate of low temperature plasma is usually low, below 10%.

(b) The impact of pressure on plasma

In addition to temperature, plasmas are particularly sensitive to pressure [**16-19**].

When the pressure is very low, the likelihood of collision between the neutral particle and the neutral particle decreases, under which the charge exchange and neutral reaction are the main process for most electron and heavy particle collisions. In this way, different reaction processes will dominate in different pressure ranges, and therefore the resulting plasma will be applied in different fields. Table 1 lists the corresponding plasmas at different pressure regimes and their application areas.

Table 1 The impact of pressure on application

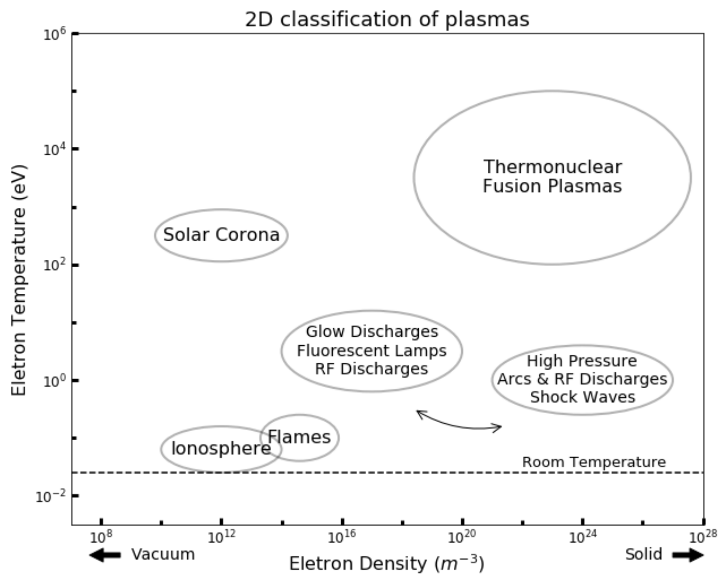
Pressure Regimes	Example Applications
Very low pressure: 1–30 mTorr	Ion bombardment etching
Low pressure: 30–100 mTorr	Ion etching with neutral assistance
Medium pressure: 100–500 mTorr	Ion etching with neutral coverage
High pressure: 500–1000 mTorr	neutral deposition

(c) The impact of power on plasma

Power [20] is another parameter to consider. Although electron temperature is not sensitive to changes in power, changes in power can have a significant impact on the charged matter density, so does the plasma dissociation.

Variance in electronic temperature and electronic density can lead to the changes in the properties of the plasma [21], so under a different range of physical conditions, plasma can be classified as various categories which are listed in Fig 1.

Fig. 1 2D classification of plasmas



2.4 Local or Non-local thermodynamic equilibrium plasmas

According to the concept of Local Thermodynamic Equilibrium [22] which plays an important role on the study of plasmas, plasmas can be classified into two types: LTE plasmas or non-LTE plasmas.

(a) Local thermodynamic equilibrium plasmas

For LTE plasma, chemical reactions and transitions are controlled by the collision process, rather than through the radiation process, and it is necessary that the collision phenomenon is micro-reversible. When the local gradient of these properties is kept low enough, the equilibrium status can be reached [21].

Considering the Griem criterion [23], most plasmas including laboratorial low-density plasmas are not in the scope of LTE plasma.

(b) Non-local thermodynamic equilibrium plasmas

The phenomenon that plasmas usually deviate from local thermodynamic equilibrium can be explained by Boltzmann distribution. Besides, the mass weight of heavy particles is significantly different from that of electrons, which is another reason for the deviation from LTE [24]. In addition, strong gradients in plasmas and corresponding diffusion phenomena are also the causes of this deviation from LTE.

The temperature of heavy particles and that of electrons are two separate temperature model used to describe non-LTE plasmas. Considering that electrons have the higher temperature than heavy particles, the temperature of plasmas, i.e. the gas temperature, are usually determined by the temperature of the heavy particles. The main characteristics of LTE plasmas and non-LTE plasma are summarized in Table 2, where T_e means electronic temperature; T_h means heavy particle temperature [21].

Table 2 Comparison between LTE and Non-LTE

	LTE plasmas	Non-LTE plasmas
Current name	Thermal plasmas	Cold plasmas
Properties	$T_e = T_h$ High electron density	$T_e \gg T_h$ Lower electron density
Examples	Arc plasma (core)	Glow discharges

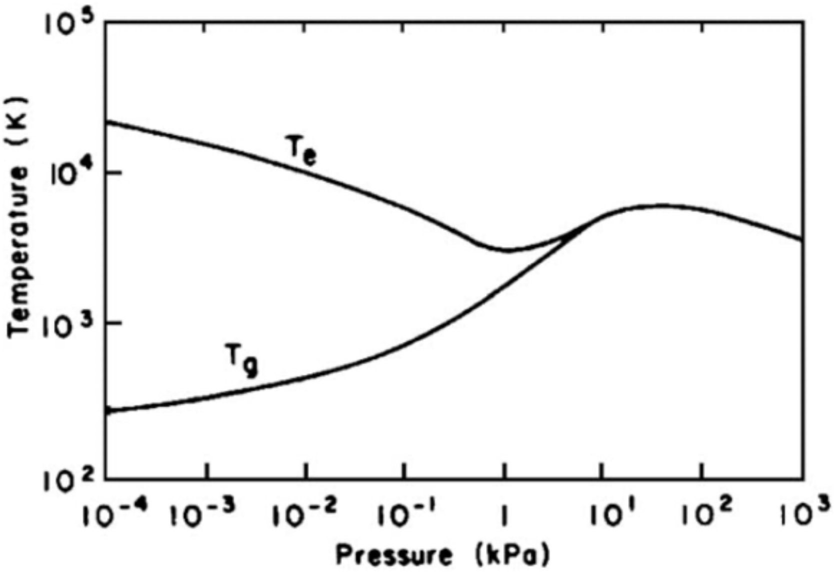
(c) Atmospheric pressure plasmas

Pressure plays an important rule on the transformation between the arc discharge and the glow discharge [21]. Fig 2 displays the evolution of plasmas.

Low pressure plasmas belong to non-LTE plasmas which has a higher electron temperature than the temperature of heavy particles. Whether the plasma is in the LTE or non-LTE state depends to some extent on the size of the feed power density.

In summary, when the power density is high, the plasma is LTE plasma, and when the feed power density is low or the presence of pulsed power supply, non-LTE plasmas are induced, because the continuous short pulse will cause the equilibrium state to not be successfully established.

Fig. 2 Evolution of the plasma



3. DATA SOURCES

There are a wide variety of databases aiming at satisfying the requirements of plasma researchers. In this paper, we focus on the National Institute of Nuclear Fusion Science (NIFS) Atomic and Molecular Database as well as the data center for plasma properties (DCPP) of the National Fusion Research Institute (NFRI, Korea).

3.1 NIFS

(a) The development of the NIFS database

The National Institute of Nuclear Fusion Science (NIFS) [25-27] Atomic and Molecular Database are the numerical database that provides information on data for plasma models. The data information of this database has been available to the general public since 1997.

This database mainly provides three types of information: (1) rate coefficients and cross sections related to dissociation, charge transfer and ionization for heavy particle impact processes; (2) rate coefficients and cross sections related to dissociation, excitation and ionization for electron impact processes; and (3) backscatter coefficients and sputtering rates of solids. At the same time, the bibliographic data can also be obtained from the NIFS database.

From 1975 on, many atomic and plasma physics research groups of universities in Japan has focused on compiling and publishing atomic data [28-29]. In the 1980s, the data stored in atomic databases was accessible for retrieving through a large computer system at Nagoya University's Institute of Plasma Physics [30]. And then, the relational database called Oracle replaced the original database system in 1997. With the help of the new system, the NIFS Atomic and Molecular Database (NIFS DB) became accessible online since then, which allows scientists in relevant research fields to display and extract the data they required online at anytime and anywhere [31-32].

The NIFS database was originally built to collect information on data about cross sections for ionization and excitation by electron impact. However, the database was later expanded and now contains both cross sections and rate coefficients for various collision processes in plasmas, including dissociation, recommendation, charge transfer as well as neutralization for both electronic impact and heavy particle impact. Therefore, the construction of this database not only promotes fusion plasma research, but also plays a positive role in the development of other research fields, particularly in astrophysics.

(b) Overview of the NIFS database

The NIFS database is made up of several sub databases according to collision processes. All the databases are accessible online for the public. Table 3 lists the names of all the sub databases as well as the information about each database for NIFS.

Table 3 NIFS Atomic and Molecular Database

Type	Name	Description	Record
Heavy particle impact	CHART	Cross sections for charge transfer and ionization by heavy particle collision.	7618
	CMOL	Cross sections and rate coefficients for heavy particle–molecule collision processes.	2139
Electron impact	AMDIS	Cross sections and rate coefficients for ionization, excitation, and recombination by electron impact.	747048
	AMOL	Cross sections and rate coefficients for electron–molecule collision processes.	3180
Sputtering rates	SPUTY	Sputtering yields for solids.	2349
Backscattering coefficients	BACKS	Energy and particle back scattering coefficients of light ions injected into surface.	485
Bibliography	ORNL	Bibliography on atomic collision.	78097

As mentioned before, the NIFS database is divided into six different databases, including CHART, CMOL, AMIDS, AMOL, SPUTY and BACKS. Besides, we can also use ORNL to find the bibliography about atomic collisions. Specifically, there are two databases named CHART and CMOL related to heavy particle collision processes, while the AMIDS database and the AMOL database are combined to store the information on data for electronic collision processes. We notice that the AMIDS database has the largest number of data, which is almost 750000. While the number of reaction records in the BACKS database is the smallest, with less than 500 data.

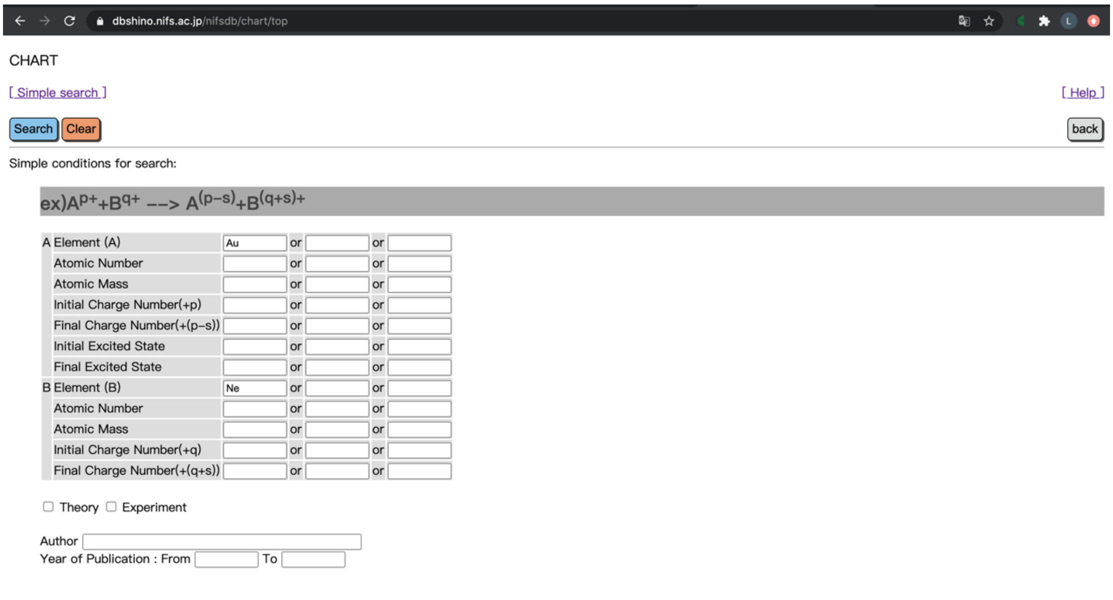
For each record listed in the NIFS database, we are able to retrieve its numerical data tables and graphs. These numerical data are displayed in terms of either electron energy (eV) vs. cross sections (cm^2) or temperature (K) vs. rate coefficients (cm^3s^{-1}). Besides, there are also some other basic information attached to each reaction record, for example, the reference information such as authors and date of publication, as well as additional information on data including methods of theoretical calculations or experiments, and data type of theoretical/experimental/evaluated data. Note that only AMDIS and CHART has the information on methods. The main original data source of records included in the NIFS database is the publication.

(c) User Interface of the NIFS database

All the information on data listed in the NIFS database are accessible online nowadays [30]. As mentioned before, the NIFS database consist of six sub databases. Particularly, the AMDIS database is further subdivided into four smaller databases which are used to store cross sections and rate coefficients for excitation (EXC), ionization (ION), recombination (REC), and dissociation (DIO) by electron impact. The information on data historically stored in DIO are migrated to AMOL since 2000. For NIFS, each sub database has its own search webpage.

Users are allowed to search for target reaction records in two different ways. Firstly, by combining various and multi-condition queries on the search webpage, users are able to search and extract the required data. Here we take the search webpage for the CHART database for example (see Fig 3). After specifying an element name, the initial and/or final states, or ionic stage, a list of reaction processes meeting these conditions will be displayed.

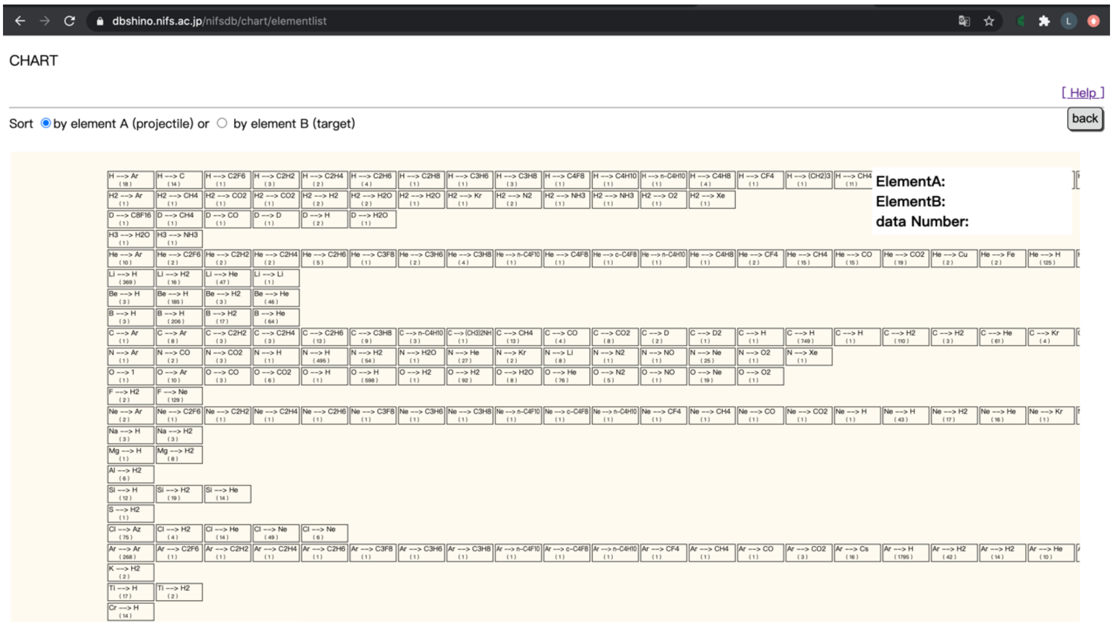
Fig. 3 The basic user interface to search data by combining various queries for NIFS



And then, on the page of process lists, we can use either the standard setting or the custom setting to display the fundamental information on data. To be specific, users are able to customize the data structure of the target records by choosing to display different information, such as the process title, methods used to obtain the data, process type, data type and so on.

Secondly, the NIFS database also provides users with the periodic table interface to simplify the process of searching data. In this way, users are able to retrieve the information on target atomic and molecular data by simply clicking the name of the element. Fig 4 shows the periodic table interface in the NIFS database.

Fig. 4 The periodic table interface to search data for simplicity for NIFS



3.2 NFRI

(a) The development of NFRI database

The data center for plasma properties (DCPP) has been founded by the National Fusion Research Institute (NFRI) of Korea in 2006. Since its establishment, the NFRI database has been collecting relevant information on data about plasma properties continuously. So far, the number of plasma properties data in the NFRI database have approached 100000[34].

Since that the previous system used by the NFRI database fails to include a wide variety of information related to cross sections for electron impact processes, rate coefficients for heavy particle impact processes, sticking coefficients for surface reaction processes, as well as thermodynamic data. To solve this problem, a new system has been developed and already been in use now.

(b) Overview of NFRI database

The NIFS Bulk and Surface Chemistry Database (DB) [35] contains a variety of different types of reaction processes, including thermodynamic reactions, collision reactions, and surface reactions, which are required by researchers and scientists in plasma fields to compute plasma parameters. These data are presented in both digital and picture forms.

Currently, there are a total number of 100000 records included in the NFRI database. Collision processes has the largest number of records, which is composed of nearly 70000 rate coefficients, 35000 cross sections and another 4 differential cross sections. And the data of collision processes are collected by three ways: experiments, theoretical calculation, or recommendation. Most data for this type are obtained from papers published previously. For surface reaction processes, there are three groups of reactions classified by sputtering yields, which is Xe, Kr and Ar. The target count of each groups is 15. All the data about surface reactions are collected from published articles. Lastly, for thermodynamic reactions, this kind of data is calculated by quantum chemistry theory methods. Table 4 shows the information on data stored in the NFRI database.

Table 4 NFRI Bulk and Surface Chemistry Database

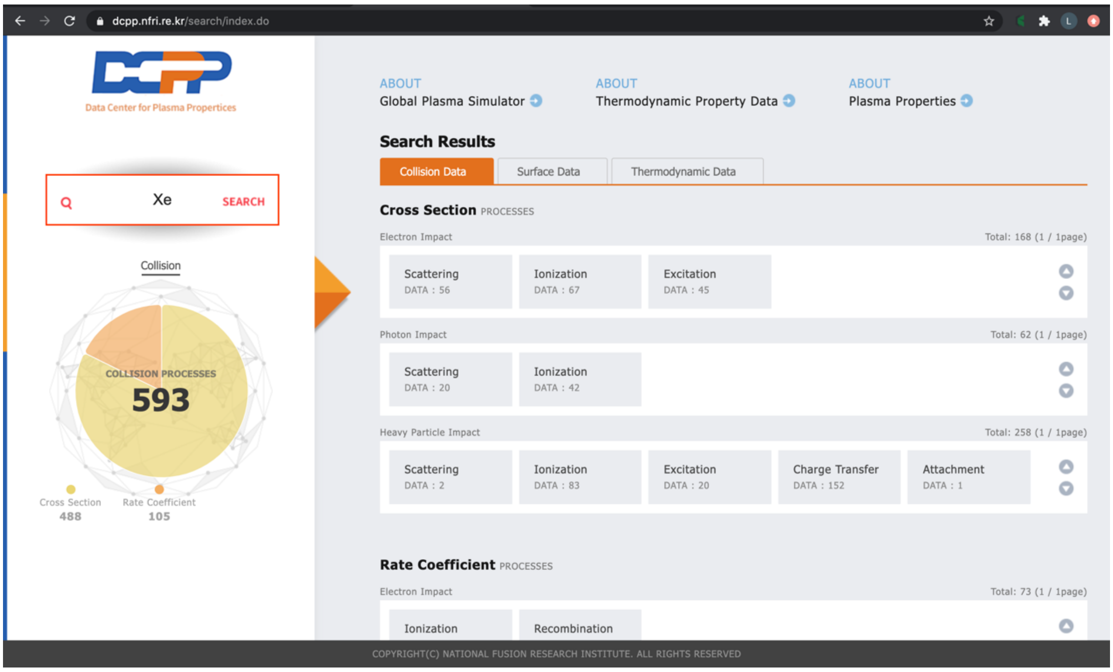
Type	Name	Description	Record
Collision processes	Cross section	Cross section processes for electron impact, heavy particle impact and photon impact	34799
	Differential cross section	Differential cross section processes for ionization and excitation	4
	Rate coefficient	Rate coefficient processes for electron impact, heavy particle impact and photon impact	69210
Surface reactions		Surface reaction processes for sputtering yield	45
Thermodynamic reactions		Thermodynamic data were calculated by quantum chemistry methods	78

Furthermore, in order to improve the accuracy of the data included in the NFRI database, groups of experts and scientists with appropriate experience conduct objective and critical evaluation for the data. At present, 90 chemical species and 700 cross section processes have been evaluated [36].

(c) User Interface of NFRI database

The NFRI database can be categorized into three types as mentioned before. Users are allowed to search data directly on the main search webpage, or search data after entering the collision/surface/thermodynamic search webpage first. Fig 5 displays the user interface in the NFRI database [37].

Fig. 5 The user interface to search data by species for NFRI



Compared to the user interface of the NIFS database, the NFRI database has a rather simple user interface. By entering the given specie on the web browser, users are able to obtain all the reaction processes for this specie form the NFRT database. Take collision processes for example, we are able to specify the type of targe reaction processes. And then we can get a sets of target records in forms of tables and graphs.

4. The QDB DATABASE

4.1 Overview of QDB

The Quantemol Database (QDB) [1] is a relational database which aims to provide coefficient rates and cross sections for a list of main reactions as well as sets of chemistries. The data related to both electron scattering and heavy particle collision processes are included in the QDB database. And for every reaction set, there are records of cross section processes, or rate coefficient processes which are generated from cross sections. QDB mainly collects the information on data about two-body reactions, and therefore the data now presented in QDB are suitable for low-pressure plasmas [38]. Therefore, the construction of QDB facilitates the development of low-temperature plasmas modeling [39-40].

Currently, the number of databases aimed to provide atomic and molecular data for plasma modelling is constantly increasing, such as the LXCat project which supplies the information on data about low temperature plasma models [41-45]. However, the LXCat project only includes data related to electronic collision processes but without records for heavy particle impact processes. In addition, even though multiple datasets are allowed to exist for a single process in either the QDB database or the LXCat project, the two platforms work in different ways in this case. Specifically, according to different application, the QDB database is designed to recommend a suitable dataset of all the datasets available, while no recommendation will be given by the LXCat project. That means the web-based platform of the LXCat project allows its users to choose a reasonable dataset by themselves.

In the QDB database, the species are the basic data items and the user can specify the state of a certain specie according to the relevant information. Electrons, photons, and M are the three generic species currently considered in the QDB database, where M is the third body existed in the three-body reaction processes. Besides, there are another 405 atomic and molecular species included in the QDB database.

Both species and reactions are searchable in the QDB database. Currently, this database is composed of 4099 different kinds of reaction processes, with 2888 records of cross section processes which depend on electron energy and 2259 rate coefficient records depending on temperature in the Arrhenius form. Since that the QDB database permits multiple sets of data to be used for the same reaction process, we are able to obtain different records for some reactions when searching in the database. In addition, the information on data about both cross sections and rate coefficients may be available for one reaction process. The QDB database also compiles most of the existing reaction processes into chemistries. The QDB database has compiled 29 sets of chemistries so far, and 8 of which have undergone a certain degree of validation, for example, the sets of chemistries for $SF_6/CF_4/O_2$ [46-48] and $SF_6/CF_4/N_2/H_2$ [49]. Table 5-7 show the types of reaction processes currently considered in the QDB database for heavy particle impact, photon impact and electron impact separately.

Table 5 QDB Plasma Chemistries and Reactions Database for heavy particle impact

Type	Name	Description	Process Title
Dissociation	HDS	collisional dissociation	$AB + C \rightarrow A + B + C$
	HDN	dissociative neutralization	$AB^- + C^+ \rightarrow A + B + C$
	HDC	dissociation and charge transfer	$AB + C^+ \rightarrow A^+ + B + C$
	HDI	dissociation and ionization	$AB + C^* \rightarrow A^+ + B + C + e$
Ionization	HIN	collisional ionization	$A + B \rightarrow A + B^+ + e$
	HIA	association and ionization	$A + B \rightarrow AB^+ + e$
	HDI	dissociation and ionization	$AB + C^* \rightarrow A^+ + B + C + e$
	HPI	Penning ionization	$A + B^* \rightarrow A^+ + B + e$
Electron detachment	HED	electron detachment	$A^- + B \rightarrow A + B + e$
	HGN	associative electron detachment	$A^- + B \rightarrow AB + e$
(de) Excitation	HEX	excitation	$A + B \rightarrow A + B^*$
	HDX	collisional deexcitation	$A + B^* \rightarrow A + B$
Charge transfer	HCX	charge transfer	$A^+ + B \rightarrow A + B^+$
	HDC	dissociation and charge transfer	$AB + C^+ \rightarrow A^+ + B + C$
Neutralization	HNE	neutralization	$e + B^- \rightarrow B + 2e$
	HDN	dissociative neutralization	$AB^- + C^+ \rightarrow A + B + C$
Association	HAS	association	$A + B \rightarrow AB$
	HIA	association and ionization	$A + B \rightarrow AB^+ + e$
	HGN	associative electron detachment	$A^- + B \rightarrow AB + e$
Other	HIR	interchange	$A + BC \rightarrow AB + C$
	HMM	ions recombination	$A^- + B^+ \rightarrow A + B$

Table 6 QDB Plasma Chemistries and Reactions Database for photon impact

Name	Description	Process Title
PDS	photodissociation	$AB + hv \rightarrow A + B$
PEX	photoexcitation	$A + hv \rightarrow A^*$
PRD	radiative decay	$A^* \rightarrow A + hv$

Table 7 QDB Plasma Chemistries and Reactions Database for electron impact

Type	Name	Description	Record
Dissociation	EDS	dissociation	$e + AB \rightarrow e + A + B$
	EDA	dissociative attachment	$e + AB \rightarrow A + B^-$
	EDE	dissociative excitation	$e + AB \rightarrow A^* + B + e$
	EDI	dissociative ionization	$e + AB \rightarrow A + B + 2e$
	EDR	dissociative recombination	$e + AB^+ \rightarrow A + B$
Ionization	EIN	ionization	$\mathrm{e} + A \rightarrow e + A^+ + e$
	EDI	dissociative ionization	$e + AB \rightarrow A + B + 2e$
	ETI	electron total ionization	$e + A \rightarrow e + e + \sum A^+$
Electron total	ETS	electron total scattering	$e + A \rightarrow e + \sum A$
	ETI	electron total ionization	$e + A \rightarrow e + e + \sum A^+$
	ETA	electron total attachment	$e + A \rightarrow \sum A^-$
(de) Excitation	EEX	electronic excitation	$e + A \rightarrow e + A^*$
	EDE	dissociative excitation	$e + AB \rightarrow A^* + B + e$
	ECX	change of excitation	$e + A^* \rightarrow e + A^{**}$
	EVX	vibrational excitation	$e + A \rightarrow e + A[v{=}*]$
	EDX	deexcitation	$e + A^* \rightarrow e + A$
Recombination	ERC	recombination (general)	$e + A^{+z} \rightarrow A^{+(z-1)}$
	EDR	dissociative recombination	$e + AB^+ \rightarrow A + B$
	ERR	radiative recombination	$e + A^+ \rightarrow A + hv$
Attachment	EDT	electron attachment	$e + A + B \rightarrow A + B^-$
	EDA	dissociative attachment	$e + AB \rightarrow A + B^-$
	ETA	electron total attachment	$e + A \rightarrow \sum A^-$
Other	EEL	elastic scattering	$e + A \rightarrow e + A$
	EMT	momentum transfer	

4.2 Process types in QDB

All the reaction records listed in the QDB database can be classified into two types based on whether they belong to cross section processes or rate coefficient processes. Particularly, the rate coefficients can be obtained base on the cross sections. In the QDB databases, the rate coefficients are accessible in forms of three parameters needed in the Arrhenius formula, which are $A, n, and\ E$. And we are able to calculate the rate coefficients at a desired temperature for each reaction process. The Arrhenius formula of electron impact reaction process and heavy particle impact reaction process is slightly different in form [1].

For electronic collision processes:

- The Arrhenius formula: $A\left(\frac{T_e}{1\ eV}\right)^n exp\left(-\frac{E}{T_e}\right)$

T -- electron temperature in eV
E -- activation energy in eV
A -- Arrhenius coefficient (its unit depends on the order of reactions)

For heavy-particle collision processes:

- Arrhenius formula: $A\left(\frac{T_g}{300\ K}\right)^n exp\left(-\frac{E}{T_g}\right)$

T -- gas temperature in K
E -- activation energy in K
A -- Arrhenius coefficient (its unit depends on the order of reactions)

As in both cases, the units of Arrhenius coefficient (A) change according to the order of reactions, we list the corresponding relationships between the units of A and the reaction orders in Table 8 as below.

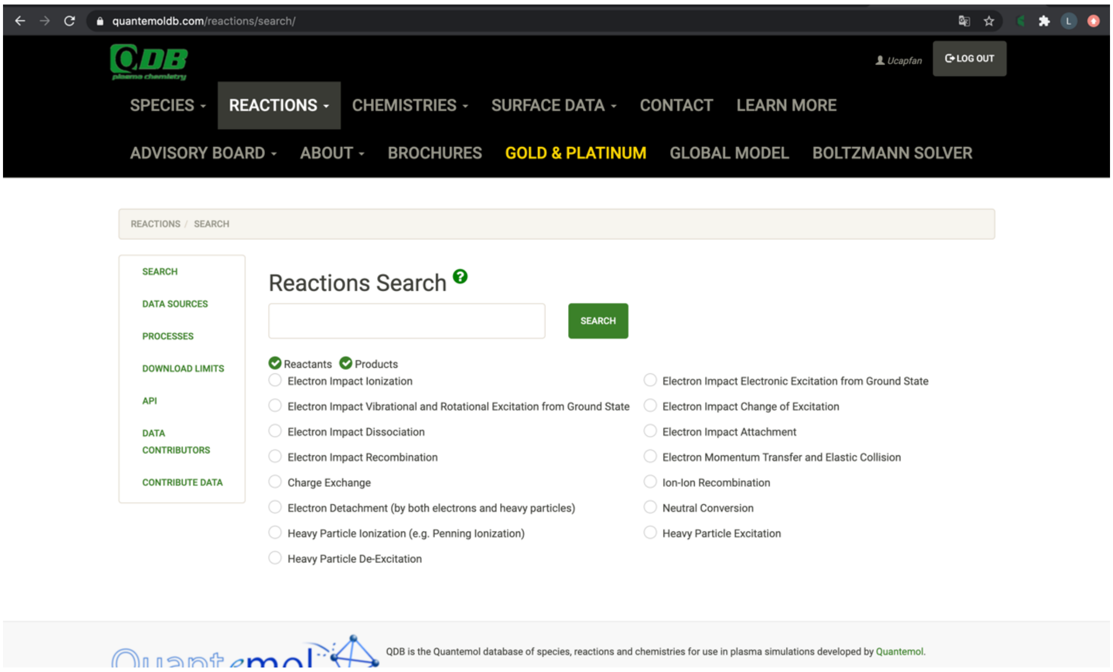
Table 8 The relationships between Arrhenius coefficient (A) and reaction orders

Type	Units of A	Examples
First-order reactions	s^{-1}	photodissociation and photoexcitation
Second-order reactions	cm^3s^{-1}	electron-impact reactions or two-body heavy-particle reactions
Third-order reactions	cm^6s^{-1}	three-body reactions

4.3 User Interface of QDB

In the QDB database, we can search for species as well as reactions [50]. Here we pay attention to reactions. The search interface for reaction processes is rather simple. We are able to search for reactions by simply entering one or more species. Also, we can determine whether these species acting as the reactants or the products. Meanwhile, in order to find the target reaction processes more precisely, we can further specify the process type when searching. In addition, QDB also supply the search page which can be used to search for data sources by DOI or the name of author. Fig 6 displays the user interface to search for data by species in the QDB database.

Fig. 6 The user interface to retrieve data by species for QDB



5. AIMS of the PROJECT

With the development of plasma technology and its successful application in diverse industrial fields, plasma research has come to the center of the stage. Many databases related to plasma have been created, the QDB database is one of them. In fact, the QDB database is a repository used to store both chemistries and key reactions needed for low temperature plasma models. QDB collects the information on data about electron impact processes, heavy particle impact processes, as well as photon impact processes. At present, the number of reaction processes included in the QDB database is nearly 5000, and 8 of 29 complete sets of chemistries have been validated with suitable experimental data. However, it is a continuous process to add new reaction records about cross sections and rate coefficients to the QDB database.

The aim of this project is to add both cross sections and rate coefficients for heavy particle impact processes as well as photon impact processes. In the previous literature research, we have considered a variety of databases including the National Institute of Nuclear Fusion Science (NIFS) Atomic and Molecular Database, the UMIST Database for Astrochemistry (UDfA), the KInetic Astrochemistry Database (KIDA), as well as the National Fusion Research Institute (NFRI) bulk and surface chemistry database. After comparing all these databases, we choose to focus on the NIFS database and the NFRI database. And we mainly rely on data mining methods to extract the information on data in the program.

For the NIFS database, we are expected to retrieve cross sections and rate coefficients for heavy particle impact, since that there are no records for photon impact included in this database. And the information on data related to heavy particle processes are stored in two different sub databases in NIFS. We extract cross sections for charge transfer and ionization by heavy particle collisions in the CHART database, while we can use the CMOL database for heavy particle – molecule collision targets.

For the NFRI database, we need to extract cross sections and rate coefficients of reaction processes for both heavy particle impact and photon impact. All the relevant data are stored in the Collision database.

In both databases, we would like to obtain the fundamental information on data such as the type of processes, the process title, theory or experiment, as well as the reference information including authors, journal name, date of publication and so on. In addition, we also need to retrieve a set of numerical data of each record for the given species. After retrieving all the reaction records needed, we need further connect the data with process types listed in the QDB database.

6. METHODS and PROGRAM

In this project, we mainly rely on data mining methods in Python language to retrieve data from existing plasma databases. Through the preliminary literature research of various plasma databases, we decide to focus on NIFS Atomic and Molecular Database as well as NFRI Bulk and surface chemistry database. However, the webpage and data structure of these two websites are completely different. For this reason, when extracting data, we need to write code for each website separately.

6.1 Methods

Data mining is the method we mainly used in this project. The basic process of it is: (1) Send requests; (2) Get contents of responses; (3) parse the obtained HTML format data; (4) Save data in database or write to TXT file. Here we write the main program based on webdriver. Given that the sites of NIFS and NFRI are encrypted to protect data, and the pages of these websites are dynamically loaded, we need to add anti-crawler code to the program and also consider user agent information.

(a) Data mining

The essence of a web crawler is a program. It can automatically download webpages and obtain the corresponding data of target webpages according to the requirements of users. It can then selectively grab corresponding webpages and links from the Internet to provide users with the data resources.

Since that python has powerful modules, and can conveniently process characters, it is very convenient to write code for data mining, so python is a commonly used language for data mining. The Python language provides a large number of third-party libraries for data mining. The more common ones are requests library, LXML library, and beautiful soup library [51-52]. Users can choose a suitable library for their work.

To do data mining for a website, we generally extract a set of URLs first, and each URL has a corresponding page, and then further extract the URL links in each page repeatedly. The general processes of data mining are (also see in Fig 7):

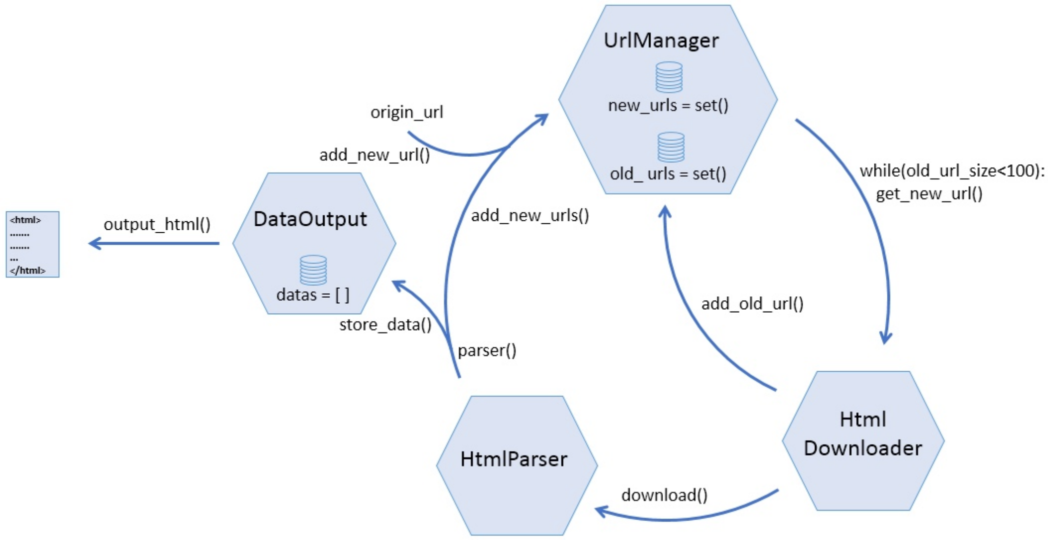
Step 1. Send a request to the target web server: Use the “get” or “post” method of the requests library to send a request to the web server needed to be extracted and then wait for a response.

Step 2. Get contents of the response: When the target site server responds, it will send back a Response object which contains the content of the target webpage.

Step 3. Parse the content of webpages: The data obtained in the previous step may be in HTML format. In this step, a third-party parsing library or regular expression can be used to parse the obtained content.

Step 4. Processing the content of webpages: Perform follow-up processing on the webpage content parsed in the previous step, obtain the required information, and finally output or save the data.

Fig. 7 The basic processes of data mining



(b) User Agent

User Agent is part of the Http protocol, which belongs to the header field [53]. User Agent is also referred to as UA. It is a special string header, which is an identifier that provides information about the type and version of the browser, operating system and its version, kernels of browser, etc. With the help of this identification, the website displayed in different layouts to provide users with a better experience.

In order to distinguish whether it is a real user or a web crawler sending a request to the server, the website usually adds UserAgent information. When doing data mining for the data of webpages, it is necessary to simulate the UserAgent.

(c) Webdriver

Webdriver is an automated testing tool in Python's selenium library [54]. It can fully simulate the operations of the browser without processing complex “requests” and “posts” information, which further improves work efficiency and is very user-friendly.

The focus of data mining is to find the corresponding elements of the required data in the web page. Table 9 lists some basic methods as well as XPath method provided by Webdriver for element positioning.

Table 9 Methods used to find elements in Webdriver

Type	Keyword	Programming statement
Basic	Link text	find_element_by_link_text
	Partial link text	find_element_by_partial_link_text
	Class name	find_element_by_class_name
	Name	find_element_by_name
	Id	find_element_by_id
	CSS selector	find_element_by_css_selector
	Tag name	find_element_by_tag_name
XPath language	XPath	find_element_by_xpath

6.2 NIFS

(a) Program strategy

In this project, we are expected to retrieve the information on data about cross sections and rate coefficients for heavy particle impact and photon impact. However, in the NIFS database, there has no records for photon impact processes, so we can only extract reactions related to heavy particle impact. Meanwhile, this part of data is stored in two sub databases in NIFS. This means we need to retrieve cross sections and rate coefficients in both CHART database and CMOL database. As mentioned before, there are two ways we can use to obtain data in the NIFS database, using either the basic user interface by constructing complex queries or the periodic table interface.

Specifically, the CHART database only stores records of cross sections for ionization and charge transfer by heavy particle collisions. And we are able to retrieve data from this database by entering a list of species in either element A or element B boxes on the search webpage. To extract data for heavy particle - molecule collision processes, we can rely on the CMOL database. Different from the CHART database which has no rate coefficients records, reactions about cross section processes as well as rate coefficient processes are included in the CMOL database. Meanwhile, as there are three data types – cross section, rate coefficient and other – in this database, it is not enough to just enter the given species in the search form as before when searching for data. We need to further specify the type of data after entering a specie.

In addition, note that there may exits more than one records in the NIFS database for one reaction process. We should extract only one record for each reaction in this case. To solve this problem, we have drawn up two rules of selection: 1) for a reaction process with both experimental and theoretical records, we select the first record of the experimental records; 2) and for a reaction process with only experimental data or theoretical data, we choose to extract the first piece of data in either experimental or theoretical group.

(b) Code Construction

For the code part, we define several functions to do data mining for a given list of species in the NIFS database. First of all, we define a function named "do_search" used to choose the sub databases as well as identify the data type in the CMOL database.

For both the CHART database and the CMOL database, we are expected to retrieve not only the numerical data tables but also the fundamental reference data about the reactions. And we retrieve these two types of data separately. For the basic information on data, we define a function named "get_custom_content" to extract the information including the type of reaction processes, theory or experiment, author and other reference data. On the data display page, we could either use the standard setting or the custom setting for this part of information. Here, we choice to use the latter to customize our selection process to obtain all types of data we need. And for this part, we save all the retrieved data in one CSV file so that we can easily do statistics and classification in the follow-up research work.

As for the numerical data related to reaction processes, we also define a function named "get_numerical_content" to obtain the data. The units of cross sections are given in terms of cm^2 as a function of electron energy whose units are given in terms of eV. And for rate coefficients, these reactions are given in units of cm^3s^{-1} as a function of temperature in K. For numerical data tables and graphs, we store piece of record in a separate CSV file using the id/record number of the reaction process as the file name. To store the numerical data sets obtained, we create a folder for each specie, and then save all the reaction records related to this specie in the same folder. In this way, we can summarize these numerical information tables, which is convenient for future search.

Finally, we need to use functions to connect the data retrieved from the NIFS database with the process types listed in the QDB database.

6.3 NFRI

(a) Program strategy

In the NFRI database, there are records of reaction processes for both heavy particle impact and photon impact. Therefore, we are able to extract cross section processes and rate coefficient processes for the two groups. And for this database, the user interface page is rather simple. In the NIFS database, there are different type of conditions we can use to specify the reactions in data searching, while in the NIFS database, we can only search for reaction processes based on the given species.

To be specific, we can search for all the target plasma data in the Collision database by entering the given species, and then further click each type of sub processes one by one to obtain the numerical data tables and graphs as well as the basic information on data about each reaction.

The challenge we meet in extracting data in the NFRI database is that we cannot retrieve all the information on data for a given specie on one page as we have done in the NIFS database. For the NFRI database, we need to retrieve all the URL links for each record of cross section processes and rate coefficient processes, and then retrieve the information on data for each piece of record by their own URL link.

(b) Code Construction

For the code part, we also define a variety of functions to retrieve data for a given list of species in the NFRI database. Firstly, a function named “do_search” is defined to do the searching process for the given species. In the next step, we define another two functions named “get_id” and “get_search_value” respectively to obtain the needed information in the URL links of each reaction process. Specifically, we use the former one to extract the id number of each record stored in the NFRI database, and the latter one to extract the type of the specie the record belongs to.

Thirdly, we define a function named “write_content” to save all the information on data we needed. All the other functions have been included in this main function as a part. In order to implement the data mining, there are two more functions need to be defined. One is for retrieving the information such as process types and process tiles, which is called “get_content”. We also extract sets of numerical data using this function. And another is defined as “get_plBiDataNum_content” to retrieve reference information data composed of authors, date of publication as well as the reference number. We store all the retrieved data of NFRI in the same way as that of NIFS.

Similarly, we need to link reaction processes between the NIFS and QDB database.

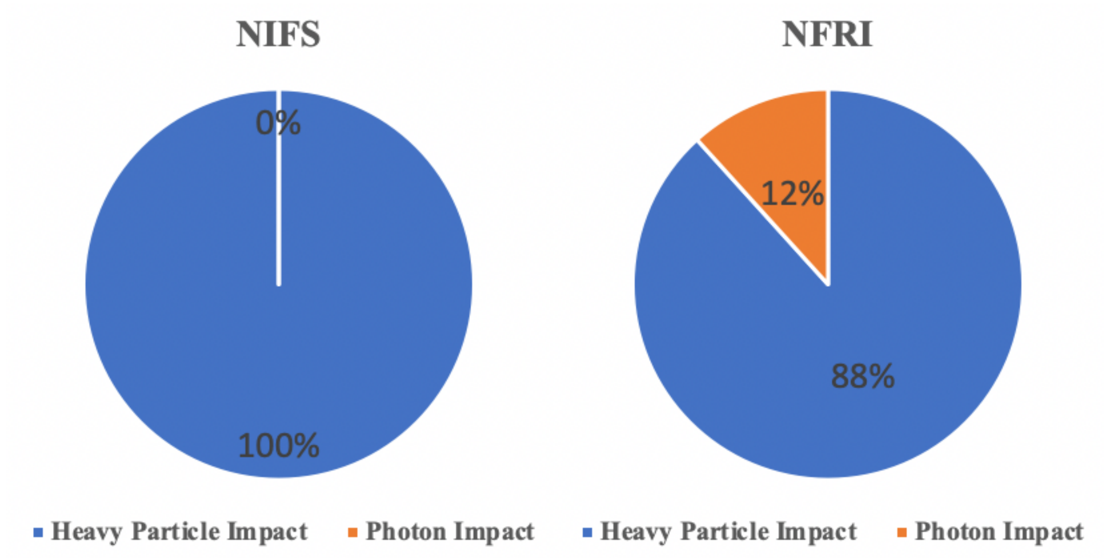
7. RESULTS and DISCUSSION

The aim of this project is to do data mining for two existing databases – the NIFS database and the NFRI database – in order to retrieve the information on atomic and molecular data for the QDB plasma models. As mentioned in the last chapter, we have written code based on the data structure and webpage design of the two databases. As a result, we obtain a total number of 9488 reaction processes which are composed of 8838 records of heavy particle impact and only 628 records of photon impact. Table 10 shows the statistics of retrieved data about heavy particle processes and photon processes in the NIFS database and the NFRI database in total. And Fig 8 shows the proportion of all the data.

Table 10 The statistics of data retrieved from NIFS and NFRI

	NIFS	NFRI	Total
Heavy Particle Impact	4105	4733	8838
Photon Impact	0	628	628
Total	4105	5361	9466

Fig. 8 The proportion of data retrieved from NIFS and NFRI



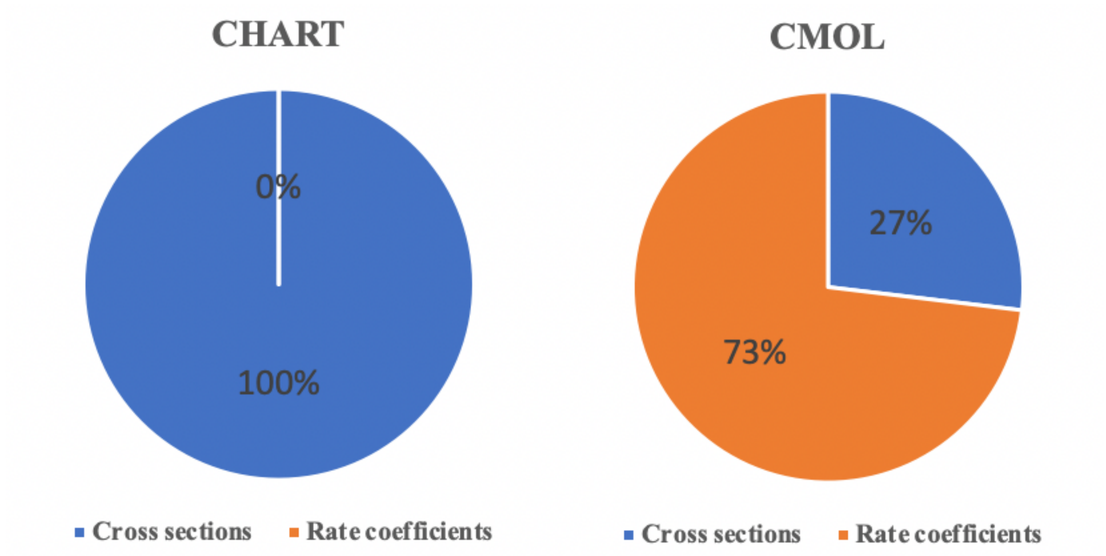
7.1 NIFS

For the NIFS database, we extract 3246 cross section processes and 859 rate coefficient processes for heavy particle impact. And there are no processes related to photon impact in this database. The statistics and proportion of data retrieved from the CHART database and the CMOL database are displayed in Table 11 and Fig 9 as below.

Table 11 The statistics of data retrieved from CHART and CMOL in NIFS

	CHART	CMOL	Total
Cross sections	2930	316	3246
Rate coefficients	0	859	859
Total	2930	1175	4105

Fig. 9 The proportion of data retrieved from CHART and CMOL in NIFS



Specifically, the CHART database only includes the information on data about cross sections, and we extract 2930 data from this database. The number of data related to charge transfer processes (labeled as HCX in the QDB database) secure the first position, with 2092 in total. There are also 810 records of processes belonging to ionization group (labeled as HIN in the QDB database), 8 records for excitation (labeled as HEX in the QDB database), and another 20 reaction records without labels.

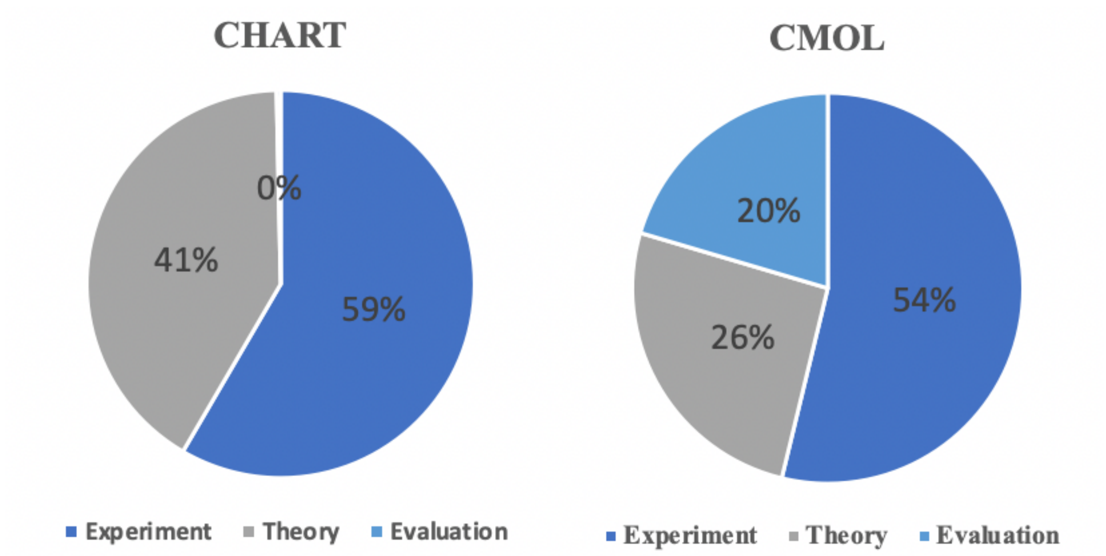
While in the CMOL database, both cross sections and rate coefficients are listed in the database, and the total number of them is 859 and 316 respectively. For cross sections, 625 reactions have been classified into various types of processes, and the remaining 234 records have no process labels. Most data of cross sections in CMOL also belong to charge transfer processes, which is 373. For the rest data, HDS (57), HDI (43), HDC (41), HEX (39) and HIN (37) all have a similar proportion of the retrieved data. And for rate coefficients, all data have process labels, with HEX having the largest data of 186 in total, compared to only 7 records of both HIN and HNE.

In addition, we pay attention to the way these data obtained as well. In the NIFS database, most records are classified into experimental data, theoretical data and evaluated data, leaving 6 reaction process with missing information. And the total number of the three types are 2340 of E, 1509 of T and only 250 of V. Table 12 shows the statistics of the data classified by E/T/V and Fig 10 displays the proportion.

Table 12 The statistics of data classified by E/T/V in NIFS

	CHART	CMOL		**Total**
	Cross section	Cross section	Rate coefficient	
Experiment	1708	21	611	2340
Theory	1206	227	76	1509
Recommend	10	68	172	250
Total	2924	316	859	4099

Fig. 10 The proportion of data classified by E/T/V in NIFS



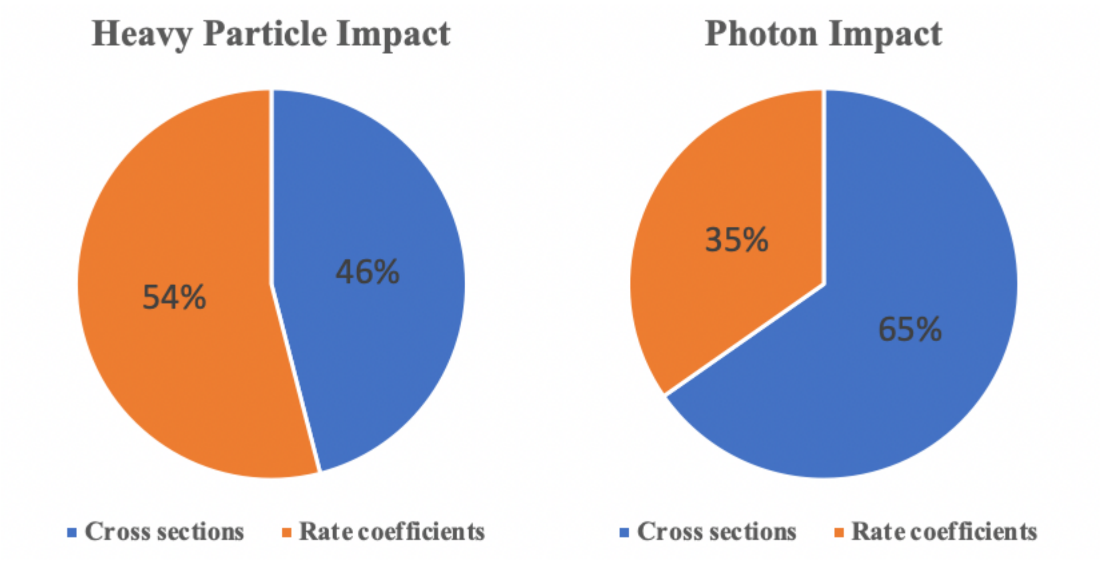
7.2 NFRI

For the NFRI database, the total number of heavy particle impact processes is 4733 and that of photon impact processes is 628. We extract 2176 cross sections and 2557 rate coefficients for heavy particle impact, and 410 cross sections and 218 rate coefficients for photon impact. Here, we notice that the only data source of photon impact in our project is the NFRI database, and the number of data in this category is relatively smaller compared to that in the heavy particle impact category. Table 13 and Fig 11 show the statistics and proportion of data retrieved for heavy particle and photon impact in NFRI separately.

Table 13 The statistics of data retrieved for heavy particle and photon impact in NFRI

	Heavy particle impact	Photon impact	**Total**
Cross sections	2176	410	2586
Rate coefficients	2557	218	2775
Total	4733	628	5361

Fig. 11 The proportion of data retrieved for heavy particle and photon impact in NFRI



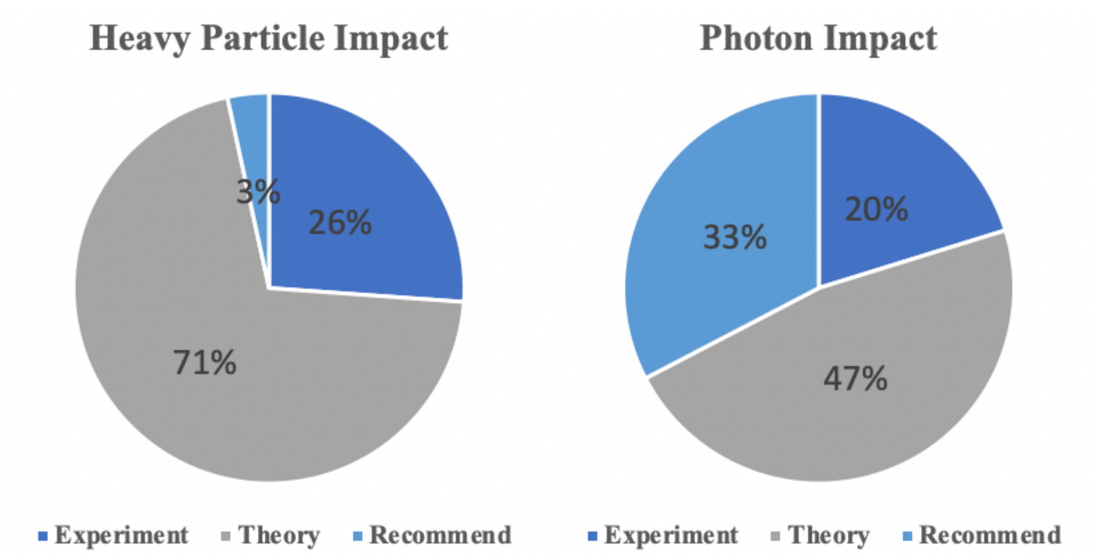
To be specific, for heavy particle impact processes, rate coefficients have a higher percentage which is 54% (2557 records), closely followed by cross sections with 45% (2176 records). These reaction processes are mainly made up of HEX, HCX and HIR. Most of reaction records can be connected to the QDB processes. As for photon impact processes, there are twice as many records of cross sections as that of rate coefficients. The number of reaction records with the corresponding QDB process labels is only 124, including 96 records of PDS and 28 records of PEX.

Meanwhile, we also classify all the retrieved data from the NFRI database into three types: experimental data, theoretical data and recommended data. For heavy particle impact, we extract 1237 data from experiments, 3334 data from theoretical calculations, and 126 data from recommendation. For photon impact, the theoretical data accounts for the highest proportion and the proportion of the recommended data approach 33%. The experimental data takes up the rest 20%. The details about the number and proportion of reaction records retrieved from NFRI are shown in Table 14 and Fig 12.

Table 14 The statistics of data classified by E/T/R in NFRI

	Heavy particle impact		Photon impact		**Total**
	Cross section	Rate coefficient	Cross section	Rate coefficient	
Experiment	449	788	127	0	1364
Theory	1565	1769	186	110	3630
Recommend	162	0	97	108	367
Total	2176	2557	410	218	5361

Fig. 12 The proportion of data classified by E/T/R in NFRI



8. CONCLUSSION and FUTURE WORK

With the development and application of plasma, there are an increasing need for accessible and reliable plasma properties data. For the reason, the QDB database has been created to collect the information on data about both species and reactions for plasma models. However, the existing data included in the QDB database is far from sufficient for modelling plasmas, and therefore the process of accumulating cross sections and rate coefficients is continuous.

In this project, we aim to do data mining for some existing databases to extract the information on data for heavy particle impact processes as well as photon impact processes. At first, we considered a wide variety of databases, such as the NIFS database, the UMIST database [55-56], the KIDA database [57-58], and also the NFRI database. And then, through the basic literature research, we decide to mainly fucus on the NIFS database as well as the NFRI database, finally extracting nearly 10000 records of reactions.

For the NIFS database, we only extract reaction records for heavy particle impact processes from CHART and CMOL, the number of which is 4105 in total. For this database, we retrieve a total number of 3246 cross sections and 859 rate coefficients.

For the NFRI database, we extract 2586 cross sections and 2775 rate coefficients in total. The retrieved data in this database is composed of heavy particle collisions as well as photon impact processes.

In the future work, we would like to continuously add additional new data about cross sections and rate coefficients for heavy particle impact, photon impact, and electron impact, to complete the QDB database. And also, we aim to expand this database to include the information on data about surface reaction processes as well as supplement more reactions involving the third body.

REFERENCES

[1] Bartschat K and Kushner M J 2016 Proc. Natl Acad. Sci. 113 7026–34

[2] Munro J J and Tennyson J 2008 J. Vac. Sci. Technol. A 26 865–9

[3] Tennyson J, Rahimi S, Hill C, et al. QDB: a new database of plasma chemistries and reactions[J]. Plasma Sources Science and Technology, 2017, 26(5): 055014.

[4] Dutuit O et al 2013 Astrophys. J. Suppl. 204 20

[5] Capitelli M, Colonna G, D'Ammando G, Laporta V and Laricchiuta A 2014 Chem. Phys. 438 31–6

[6] Kadochnikov I N, Loukhovitski B I and Starik A M 2013 Plasma Sources Sci. Technol. 22 035013

[7] Capitelli M, Celiberto R, Colonna G, Esposito F, Gorse C, Hassouni K, Laricchiuta A and Longo S 2016 Fundamental Aspects of Plasma Chemical Physics: Kinetics (Springer Series on Atomic Optical and Plasma Physics vol 85) (Berlin: Springer) pp 223–45

[8] Bultel A and Annaloro J 2013 Plasma Sources Sci. Technol. 22 025008

[9] Bartschat K, Tennyson J and Zatsarinny O 2017 Plasma Proc. Polym. 14 1600093

[10] H. Conrads, M. Schmidt, Plasma generation and plasma sources, Plasma Sources Sci. Technol. 9 (2000) 441 – 454.

[11] N.S.J. Braithwaite, Introduction to gas discharges, Plasma Sources Sci. Technol. 9 (2000) 517 – 527.

[12] Machala Z, Marode E, Laux C O, et al. DC glow discharges in atmospheric pressure air[J]. Journal of Advanced Oxidation Technologies, 2004, 7(2): 133-137.

[13] Abramski K M, Colley A D, Baker H J, et al. Power scaling of large-area transverse radio frequency discharge CO2 lasers[J]. Applied physics letters, 1989, 54(19): 1833-1835.

[14] Moslehi M M, Fu C Y, Saraswat K. Low-temperature direct nitridation of silicon in nitrogen plasma generated by microwave discharge: U.S. Patent 4,715,937[P]. 1987-12-29.

[15] Corke T C, Enloe C L, Wilkinson S P. Dielectric barrier discharge plasma actuators for flow control[J]. Annual review of fluid mechanics, 2010, 42: 505-529.

[16] J.F. Behnke, H. Lange, P. Michel, T. Opalinska, H. Steffen, H.-E. Wagner, The cleaning process of metal surfaces in barrier discharge plasmas, International Symposium of High Pressure Low Temperature Plasma Chemistry, September 2 – 4 1996, Milovy, Czech Republic, 1998, pp. 138 – 142.

[17] J.F. Behnke, H. Steffen, H. Lange, Ellipsometric investigations during plasma cleaning: comparison between low-pressure RF-plasma and barruer dscharge at atmospheric pressure, International Symposium of High Pressure Low Temperature Plasma Chemistry, September 2

– 4 1996, Milovy, Czech Republic, 1998, pp. 133 – 137.

[18] M. Cada, O. Churpita, Z. Hubicka, H. Sichova, L. Jastrabik, Investiga- tion of the low temperature atmospheric deposition of TCO thin films on polymer substrates, Surf. Coat. Technol. 177 – 178 (2004) 699 – 704.

[19] T. Yokoyama, M. Kogoma, T. Moriwaki, S. Okazaki, The mechanism of the stabilisation of glow plasma at atmospheric pressure, J. Phys. D: Appl. Phys. 23 (1990) 1125 – 1128.

[20] Ayilaran A, Hanicinec M, Sebastian M, et al. Reduced chemistries with the Quantemol database (QDB)[J]. Plasma Science and Technology, 2019, 21(6): 064006.

[21] M.I. Boulos, P. Fauchais, E. Pfender, Thermal Plasmas: Fundamental And Applications. Volume I, Plenum Press, New York, ISBN: 0-306- 44607-3, 1994, 452 pp.

[22] H.R. Griem, High-density corrections in plasma spectroscopy, Phys. Rev. 128 (3) (1962) 997 – 1003.

[23] H.R. Griem, Validity of local thermal equilibrium in plasma spectroscopy, Phys. Rev. 131 (3) (1963) 1170 – 1176.

[24] Herziger, G., H. Weber, and R. Poprawe, eds. Laser Physics and Applications. Springer Berlin Heidelberg, 2007.

[25] Emoto M, Murakami I, Kato D, et al. Improvement of the NIFS Atom and Molecular Database[J]. Atoms, 2019, 7(3): 91.

[26] Murakami I, Kato D, Kato M, et al. NIFS Atomic and Molecular Database for Collision Processes in Plasma[J]. Fusion science and technology, 2007, 51(2T): 138-140.

[27] Murakami I, Kato D, Kato M, et al. Atomic and Molecular Databases and Data Evaluation Activities at the National Institute for Fusion Science[J]. Fusion Science and Technology, 2013, 63(3): 400-405.

[28] Takayanagi, K.; Suzuki, H. Cross-sections for Atomic Processes Vol.1 (H) IPPJ-DT-48; Institute of Plasma Physics, Nagoya University: Nagoya, Japan, 1975.

[29] Takayanagi, K.; Suzuki, H. Cross-sections for Atomic Processes Vol.2 (He) IPPJ-DT-50; Institute of Plasma Physics, Nagoya University: Nagoya, Japan, 1976.

[30] Kato, T.; Itikawa, Y.; Kanada, K.; Watanabe, R. Database and Retrieval Display System of Atomic Data for Fusion. Phys. Sci. 1981, 23, 198.

[31] Kato, T.; Tawara, H.; Matsunami, N.; Watanabe, R. Database and Retrieval System for Compiled Sputtering Data. J. Nucl. Master. 1984, 128, 1006–1009.

[32] Murakami, I.; Kato, D.; Kato, M.; Sakaue, H. Atomic and Molecular Databases and Data Evaluation Activities at the National Institute for Fusion Science. Fusion Sci. Technol. 2013, 63, 400–405.

[33] The NIFS database. Available online: https://dbshino.nifs.ac.jp/nifsdb

[34] Data Center for Plasma Properties http://dcpp.nfri.re.kr (accessed Jan. 31, 2020).

[35] Park J H, Choi H, Chang W S, et al. A New Version of the Plasma Database for Plasma Physics in the Data Center for Plasma Properties[J].

[36] National Center for Standard Reference Data http://www.srd.re.kr (accessed Jan. 31, 2020).

[37] The NFRI database. Available online: https://dcpp.nfri.re.kr/index.do

[38] A Grill Cold Plasma in Materials Fabrication from Fundamentals to Applications (John Willy & Sons, Inc, 1994), pp. 1–27.

[39] IAEA AtomicandMolecularDataUnithttps://www-amdis.iaea.org (accessed Jan. 31, 2020).

[40] J. Tennyson, S. Rahimi, C. Hill, L. Tse, A. Vibhakar, D. Akello-Egwel, D. B. Brown, A. Dzarasova, J. R. Hamilton, D. Jaksch, and S. Mohr, Plasma Sources Sci. Technol. 26, 055014 (2017).

[41] VAMDC http://www.vamdc.org/ (accessed Jan. 31, 2020).

[42] G. Rixon, M. Dubernet, N. Piskunov, N. Walton, N. Mason, P. Sidaner, S. Schlemmer, J. Tennyson, A. Akram, K. Benson, and J. Bureau, AIP Conf. Proc. 1344, 107 (2011).

[43] M. L. Dubernet, B. K. Antony, Y. A. Ba, Y. L. Babikov, K. Bartschat, V. Boudon, B. J. Braams, H. K. Chung, F. Daniel, F. Delahaye, and G. Del Zanna, J. Phys. B: At., Mol. Opt. Phys. 49, 1 (2016).

[44] L. C. Pitchford, L. L. Alves, K. Bartschat, S. F. Biagi, M. C. Bordage, I. Bray, C. E. Brion, M. J. Brunger, L. Campbell, A. Chachereau, and B. Chaudhury, Plasma Processes and Polymers 14, 1600098 (2017).

[45] Y. Ralchenko, R. E. H. Clark, M. L. Dubernet, S. Gagarin, D. Humbert, P. A. Loboda, N. Moreau, E. Roueff, and D. R. Schultz, AIP Conf. Proc. 1125, 207 (2009).

[46] Babcock L M and Streit G E 1981 J. Chem. Phys. 74 5700–6

[47] Kokkoris G, Panagiotopoulos A, Goodyear A, Cooke M and Gogolides E 2009 J. Phys. D: Appl. Phys. 42 055209

[48] Nelson C T, Overzet L J and Goeckner M J 2012 J. Vac Sci. Technol. A 30 051308

[49] Moseley J T, Olson R E and Peterson J R 1975 Case Stud. At. Phys. 5 1–45

[50] The QDB database. Available online: https://www.quantemoldb.com/

[51] The LXML library in Python. Available online: https://lxml.de

[52] The bs4 in Python. Available online: https://www.crummy.com/software/BeautifulSoup/

[53] The Fake-user Agent library in Python. Available online: https://pypi.org/project/fake-useragent/

[54] The Selenium library in Python. Available online: https://selenium-python.readthedocs.io

[55] Walsh C, Millar T J. An introduction to astrochemical modelling[J]. 2014.

[56] McElroy D, Walsh C, Markwick A J, et al. The UMIST database for astrochemistry 2012[J].

Astronomy & Astrophysics, 2013, 550: A36.

[57] Wakelam V, Herbst E, Loison J C, et al. A kinetic database for astrochemistry (KIDA)[J]. The Astrophysical Journal Supplement Series, 2012, 199(1): 21.

[58] Bartschat K, Kushner M J. Electron collisions with atoms, ions, molecules, and surfaces: Fundamental science empowering advances in technology[J]. Proceedings of the National Academy of Sciences, 2016, 113(26): 7026-7034.