Supplementary information about The PIONEER database

Antoine Salden[1,2], Maik Budde[2,3], Carolina A. Garcia-Soto[4,5], Omar Biondo[6,7], Jairo Barauna[5,8], Marzia Faedda[1,9], Beatrice Musig[8,10], and Olivier Guaitella[4]

[1]Department of Physics, University of Trento, 38100 Povo, Italy
[2]Department of Applied Physics and Science Education, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands
[3]Instituto de Plasmas e Fusão Nuclear, Instituto Superior Técnico, Universidade de Lisboa 1049-001, Lisboa, Portugal
[4]LPP, CNRS, École Polytechnique, Sorbonne Université, Université Paris-Saclay, IP-Paris 91128, Palaiseau, France
[5]Department of Organic Chemistry, Biochemistry and Catalysis, University of Bucharest, Bucharest, Romania
[6]Research Group PLASMANT, Department of Chemistry, University of Antwerp, Universiteitsplein 1, Wilrijk B-2610, Belgium
[7]DIFFER, 5612AJ Eindhoven, The Netherlands
[8]Instituto de Carboquímica ICB-CSIC, Zaragoza, 50018, Spain
[9]Faculty of Energy and Fuels, AGH University of Krakow, Krakow, Poland
[10]Institut Jean le Rond d'Alembert, CNRS UMR 7190 - Sorbonne Université, Saint-Cyr l'Ecole, 78210, France

August 15, 2023

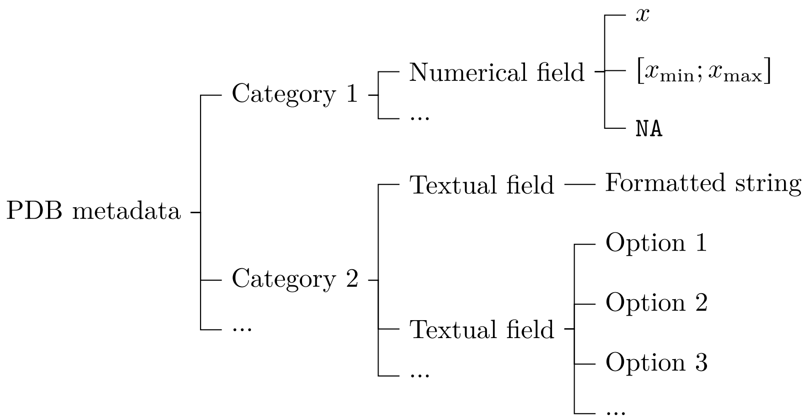
Contents

1 General

The Pioneer database (PDB) is divided into two parts: performance data and metadata. The performance data originate from measurements reported in literature, where they are typically provided in form of plots or tables. The PDB and its dedicated online tool allow to compare large amounts of performance data to derive trends leading eventually to process optimisation. Performance data is provided in form of plain text files with only two comma-separated columns of numbers with a point (.) as decimal separator, without a header (as it is inferred from the metadata). The first column contains what is henceforth called *process parameters*. These are the independent variables of the experiments, i.e. the x-values like power, pressure etc., see section 7. The second column contains the so-called *performance parameters*. These are the dependent variables of the experiment, i.e. the y-values like conversion, selectivity etc., see section 7. The metadata contain additional information about the measurements that are crucial for their interpretation. Metadata are provided in table format following the template discussed subsequently.
Before elaborating on the actual data input, the structure of the PDB metadata is discussed. The metadata is grouped thematically in *categories*. Within each category, information is entered into *fields*, i.e. the columns of the table. *Essential* and *conditional* fields are distinguished. Essential fields contain crucial information for the assessment of the plasma-catalytic process. In the best case scenario, all of them are given in the respective publication. When fields are listed in the description of a category from section 2 onward, essential fields are indicated by a regular bullet point (•). Conditional fields are by no means less important than essential ones, but can rather be left empty depending on other fields. For instance, most fields in the `catalyst` category are left empty, when no catalyst is used. Thus, conditional fields are meant to save time. Listed in the following, they are indicated by a plus (+). In conclusion, all fields are strongly recommended as data useful for valuable comparison with other work from the community. A subgroup of essential as well as conditional fields are those fields that contain the *process parameters* defined in the first paragraph of this section. Generally speaking, process parameters are the experimental settings in the pursuit of highest performance. The user of the PDB thus encounters process parameters on two occasions: on the one hand as typical x-values in the performance data and on the other hand as input to fields of the metadata. Hereafter, *fields that contain process parameters* and *parameters* are used synonymously. The total of fields belonging to the same measurement make up what is hereafter called a *data set*, corresponding to a row of the table. Note that here the input of data is addressed. In the back end, a data set is broken up into (x,y)-pairs for more flexibility in data handling. Data is exported also in that format.

To ensure comparability, a template is used for inputting information into the PDB metadata. With respect to information entered, fields can be divided in *numerical* and *textual* fields. They are filled with numbers or text, respectively. For example, parameters are usually numerical fields. A numerical field contains either (i) one number x if the numerical value is known and does not change during the experiment; (ii) a range of values between x_{min} and x_{max} –given as array-like notation $[x_{\text{min}}; x_{\text{max}}]$– if the numerical value is not exactly known, or when it changes in the course of the measurement; or (iii) `NA` if the numerical value is not known. For further use, an aggregate function is applied to array-like data to obtain a single number, by convention the mean.

A textual field contains a string of text. There are a few instances where text can be entered freely as long as some format is followed. However, usually the field is filled by selecting from a pre-defined list of options in the template. Most of these lists are fixed but some might be extended in the future depending on the experimental data provided. This paragraph just gives a general overview. In the in-depth discussion of the fields of each category, it gets more clear what exactly is supposed to be filled in each field. The following tree illustrates in a general manner the structure of the PDB metadata.



The metadata of the PDB are divided into six categories

- `data identification`

- `gas mixture`
- `plasma source`
- `catalyst`
- `separation unit`
- `output data`

In the following, all the fields included in a given category are described with the relevant information of units, format, limitations/choices made.

2 Data Identification

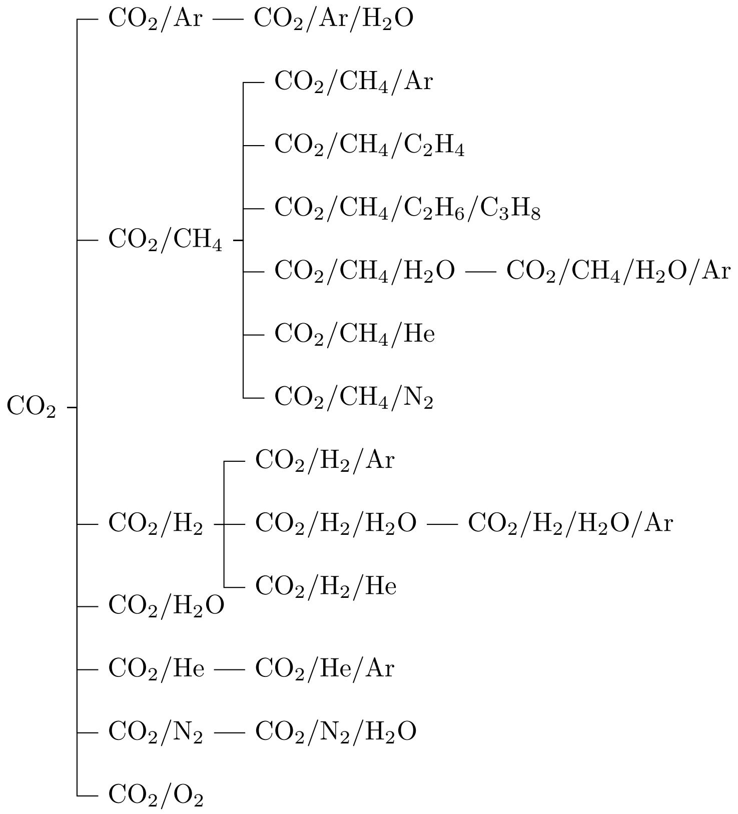
The `data identification` category contains six fields that are all essential.

- `data of input in the PDB: formatted as yymmdd`
  The `data of input in the PDB: formatted as yymmdd` field contains the date at which the data set was added to the PDB following the format yymmdd, e.g. the 20^{th} of March 2023 would be `230320`.
- `contributor`
  The `contributor` field contains the name of the person that added the data set to the PDB, starting with the initials followed by the surname.
- `publication year (yyyy)`
  The `publication year (yyyy)` field contains the written-out year in which the manuscript containing the data to be added was published. Inherently, only peer-reviewed data can be added.
- `first author name`
  The `first author name` field contains the surname of the first author for the publication.
- `DOI`
  The `DOI` field contains the referable digital object identifier (DOI) of the publication. In the back end the DOI is used to generate a `Bibtex` entry for the publication.

3 Gas Mixture

The `gas mixture` category contains eight fields of which two are conditional.

- `gases`
  The `gases` field lists the gases in the treated mixture. In principle, every mixture is allowed as long as it contains CO_2. For multiple gases in the mixture, the individual components are listed, separated by a forward slash (`/`) and starting always with CO_2. Before a new gas mixture is added, the list of already entered mixtures should be consulted. At the time of publication, the following mixtures are already defined



The order of gas components is arbitrary, e.g. the proportion of each component is irrelevant for their order. In the back end, the order gets resorted with the following precedence: [CO_2, CH_4, H_2, Ar, N_2, He, H_2O, C_2H_4, O_2, C_2H_6, C_3H_8].

- `proportion (%/%/%...)`
  The `proportion (%/%/%...)` field lists the fraction of each gas component in the mixture in percent again separated by a forward slash (`/`). Obviously, this list has to follow the same order as the `gases` parameter before and all fraction together should add up to 100 %.

- `pressure (mbar)`
  The `pressure (mbar)` parameter contains the total pressure p in mbar. Atmospheric pressure corresponds to 1013.25 mbar.

- `gas flow (ml/min) or (sccm)`
  The `gas flow (ml/min) or (sccm)` parameter contains the total inlet gas flow rate ϕ used in the experiment in either $\mathrm{mL\,min^{-1}}$ or in sccm.

+ `residence time (s)`
  The `residence time (s)` field contains the residence time τ_{res} in s. In principle, τ_{res} is the time that a molecule on average spends in the plasma zone. However, since often only the reactor volume is known also the time that a molecule on average spends in the reactor is valid. It is calculated according to

$$\tau_{\mathrm{res}} = \frac{V}{\phi \frac{p_{\mathrm{atm}}}{p} \frac{T}{T_{\mathrm{std}}}}, \tag{1}$$

  where V is the volume (of plasma or reactor), ϕ the total flow rate, p the total pressure, p_{atm} atmospheric pressure, T the (gas) temperature and $T_{\mathrm{std}} = 273.15\,\mathrm{K}$ the temperature at standard conditions. Given that all essential parameters are filled into the PDB, τ_{res} is calculated in the back end. In case of missing parameters, τ_{res} is not considered. If τ_{res} is still known someway and filled in here, it overwrites everything done in the back end.

- `temperature (K)`
  The `temperature (K)` parameter contains the temperature T in K. It refers to the temperature reported which differs from publication to publication.

+ `relevant temperature`
  The `relevant temperature` field is meant to give additional information on which temperature is given

in the previous parameter. The most valuable information would be the in situ gas temperature T_g that is unfortunately often not known. Valid options are in order of decreasing value:

  - `rotational/gas (OES)`
  - `rotational (IR absorption)`
  - `rotational (laser spec.)`
    All of these options are in situ non-invasive techniques. Optical emission spectroscopy (OES) measures the light emission from excited atoms and molecules. IR absorption contains for instance Fourier transform infrared (FTIR) as well as quantum cascade laser (QCL) absorption spectroscopy. Laser spectroscopy contains for instance Raman or Rayleigh scattering and laser-induced fluorescence (LIF) spectroscopy. The Doppler broadening of spectral lines allows for the determination of the gas temperature T_g characterising the translational motion of particles. From molecular bands, the rotational temperature T_rot of molecules is determined either through spectral fitting or the Boltzmann plot method. Due to strong translation-rotation energy transfer it is often valid to assume that T_g and T_rot are equal [1]. Particularly, IR absorption and laser spectroscopy have the advantage that they can measure the species densities of interest at the same location as T.
  - `active zone (fiber optical probe)`
  - `active zone (thermocouple)`
  - `active zone (IR camera through IR window)`
    All three techniques provide temperature information from within the plasma reactor. However, the probed location might not be the most relevant for the conversion process, for the adjustment of the flow rate or calculation of residence time. Furthermore, whatever object is inserted into the plasma volume might affect the plasma properties and therefore, interfere with the real performance.
  - `outer surface (thermocouple/probe)`
  - `outer surface (IR Camera)`
    The temperature of the outer surface and inside the reactor are very likely different. A calibration by heating the reactor inside and measuring T on the outside is an improvement but might still not depict the non-thermal plasma temperature distribution properly.
  - `imposed (oven/cooling bath)`
    Actively imposing a temperature from the outside is common practice in conventional catalysis. In combination with a plasma as additional heat source, this must be considered only a temperature indication.
  - `downstream`
    The gas downstream of the reactor has cooled down to an unknown extent since leaving the plasma and therefore does not give information about T at the location of conversion. In very hot plasma like MW this might however be the only option to measure the temperature with a thermocouple.
  - `other`
    Very exotic methods to determine the temperature that do not need a separate option are grouped in `other`.
  - `NA`
    In case a temperature is given without origin, `NA` is the correct choice.

- `main output gas diagnostic`
  The `main output gas diagnostic` field lists the used diagnostic techniques of the exhaust to assess the performance of the processing. Possible options are gas chromatography (`GC`), mass spectrometry (`MS`), FTIR, gas analyser (`GA`), `other` for unconventional methods and `NA` if not known. If multiple techniques are required to track all components, also lists of the above mentioned are valid. Attention is required to take into account changes in total particle number due to (chemical) reactions like conversion [2].

4 Plasma Source

The `plasma source` category contains fifteen fields of which four are conditional.

- `source type`
  The `source type` field details the used plasma source for conversion. At the time of writing, the following types are valid

  - `corona`
    In a corona discharge, the size of the electrodes are typically quite different so that the electric field decreases significantly from the smaller to the larger electrode, e.g. in a pin to plane geometry, to keep the plasma from going thermal. Positive as well as negative corona discharges are reported depending on the applied high voltage.
  - `DBD`
    In a dielectric barrier discharge (DBD), at least one dielectric is placed between the electrodes limiting the current, leading to the formation of streamers instead of sparks.
  - `GDA`
    Gliding arc discharges (GDA) leverage a combination of diverging electrodes and high gas flow rate to extinguish the plasma in space to avoid the transition to a thermal arc. A special case is the gliding arc plasmatron (GAP) or tornado discharge that uses a vortex gas flow.
  - `glow discharge`
    Glow discharges are ignited by applying a voltage across a gap. The current is often limited through an outer resistor.
  - `MW`
    When igniting a discharge by means of microwave (MW) radiation, the plasma generated resembles thermal hot plasma at atmospheric pressure, but the plasma maintains electron and neutrals out of equilibrium.
  - `NRP`
    In nanosecond-repetitively pulsed (NRP) discharges, the exciting voltage pulses have durations below ţs which give the gas not enough time to significantly heat up ensuring a non-thermal state.
  - `RF`
    Comparable to a MW, the radio frequency (RF) discharge is defined through the MHz frequency range (usually at 13.56 MHz) of the exciting voltage.
  - `spark`
    A discharge is ignited when the electric field over a gap surpasses the needed breakdown field strength. The current is limited by the available charge or the impedance of the power supply. Sparks are at the limit to thermal plasma as for high enough current the contribution of thermal ionisation becomes larger than that of electron impact ionisation.
  - `other`
    Unusual plasma sources that match none of the above mentioned types are sorted as `other`. Depending on future developments of the research field, more source types might be added.

  Note that the source type could be further divided into a spatial type, i.e. corona, DBD, GDA, glow discharge and spark, as well as a temporal type, i.e. MW, NRP and RF. In consequence, combinations are possible. For instance, Moss et al. report experiments on a NRP corona discharge [3]. In the PDB, priority is given to the temporal type which is why [3] is listed as NRP discharge. This decision is motivated by the fact that the time scale determines whether a certain discharge type can fully develop. Again for the example of [3], for NRP discharges the ns time scale is usually not sufficient to form a conventional corona discharge.

- `geometry`
  The `geometry` field allows to specify the geometry of a certain discharge type. At the time of writing, the following options are valid.

  - `coaxial 1 dielectric`
    Cylindrical reactor geometry with one of the electrodes covered with a dielectric. It is not distinguished whether the inside of the outer electrode or the outside of the inner electrode is covered.
  - `coaxial 2 dielectric`
    Cylindrical reactor geometry with both electrodes covered with a dielectric.

  - `coaxial no dielectric`
    Cylindrical reactor geometry with no dielectric in between.
  - `pin to pin`
    Two pin- or needle-shaped electrodes ensure a reproducible location of plasma ignition.
  - `pin to plane`
    The discharge is ignited around the pin electrode where the electric field is strongest.
  - `pin to ring`
    The ring shaped electrode allows high gas throughput by letting the gas flow along the pin and through the ring. The plasma might rotate around the pin along the ring.
  - `pin to sphere`
    Comparable to the pin to pin geometry with less predictability but also less electrode degradation.
  - `plane to plane`
    Simple geometry with two opposing planes to which a voltage is applied.
  - `ring to ring`
    Like the pin to ring configuration two rings allow for high gas throughput.
  - `surfacewave`
    The plasma is ignited with MW power from a surfatron in a dielectric tube. Gas flows along the tube axis.
  - `three blades`
    The gas flow forces the plasma along the diverging blade-shaped electrodes until it extinguishes. The three electrodes make the plasma three-dimensional.
  - `torch`
    The discharge is allowed to rapidly expand into a volume.
  - `two blades`
    The gas flow forces the plasma along the diverging blade-shaped electrodes until it extinguishes. Due to the two electrodes, the plasma is more or less limited to a plane.
  - `vortex-stabilized`
    Tangential gas inlet creates a vortex flow inside the reactor that stabilises the discharge and protects the wall.
  - `other`
    Very exotic geometries that match none of the above mentioned types are sorted as `other`. Depending on future developments of the research field, more geometries might be added.

  Some geometries are specific for a certain discharge type while others might be shared between different types. See figure 1 for an illustrative but not exclusive overview.

- `volume (cm3)`
  The `volume` field contains either the volume of the reactor or the plasma in cm^3. The volume is of course very important for the residence time of the particles.

+ `relevant volume`
  The `relevant volume` field gives more information about the volume in the preceding parameter. If no volume is known this can be left empty. The options are the following

  - `plasma (estimated or measured)`
    The most valuable information is the volume of the plasma itself as here the gas is converted. Unfortunately, it is very hard to determine the plasma volume. The popular possibilities are (i) a known reactor volume is entirely filled with a homogeneous plasma so that the plasma volume equals the reactor volume. Naturally, this is more of a qualitative assessment. (ii) space-resolved measurements of some quantity, e.g. density of an excited species or temperature, give an idea about the dimension of the plasma. However, the measured profile does not necessarily represent the plasma profile and (iii) plasma imaging of emission. Just like (ii) this suffers from the fact that the limits of the plasma are often spread out and some kind of truncation is needed.
  - `reactor`
    The reactor volume is usually easy to determine. The more localised the plasma, the larger the overestimation of the time the particles spend in the discharge, i.e. a lot of gas just streams through the reactor without ever experiencing plasma exposure. This holds not only for obvious cases like spark discharge in pin-to-pin configuration but also for plasma filaments in DBDs.

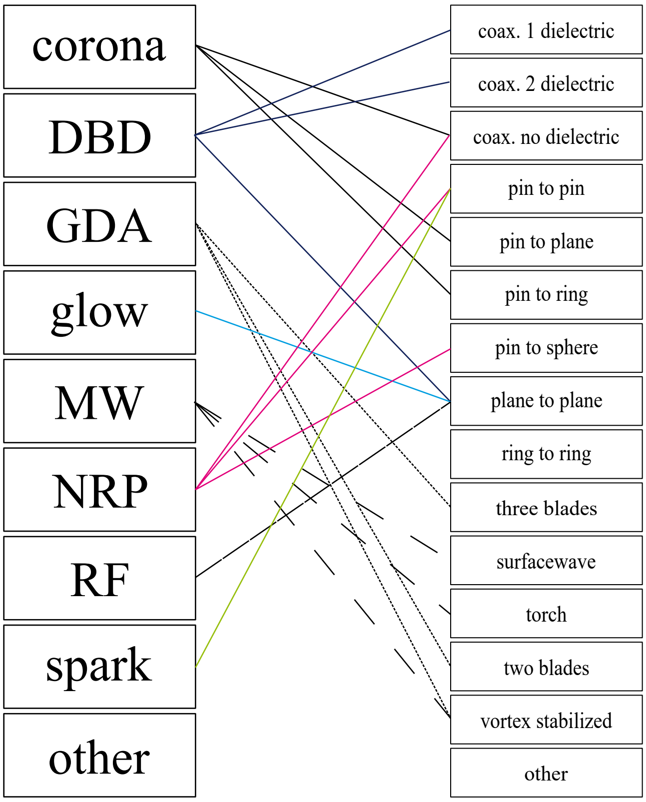


Figure 1. Typical combinations of source type and geometry. For better visibility, the connections from and to `other` are omitted.

  - `reactor - macroscopic catalyst`
    In case of in situ usage of a catalyst, the catalyst material occupies space that otherwise could be filled with plasma. In other words, the catalyst severely influences the plasma from an exclusively physical standpoint already. Knowing the volume of remaining void space allows to an extent to take this effect into account.
  - `NA`
    In case a volume is given without origin, `NA` is the correct choice.

- `reactor material in direct plasma contact 1 (mainly dielectric)`
  The `reactor material in direct plasma contact 1 (mainly dielectric)` field contains a material used in the reactor that is in direct contact with the plasma, e.g. the electrodes or the reactor wall. Even though principally arbitrary, in case of multiple materials including a dielectric material, e.g. in a DBD, the latter should be included as `reactor material in direct plasma contact 1 (mainly dielectric)`. Eventually, every material is valid. At the time of writing the following options are included
  - `alumina`
  - `boron nitride`
  - `ceramic`
  - `glass`
  - `graphite`
  - `mullite`
  - `quartz`
  - `starch`
  - `teflon`
  - `aluminum`
  - `brass`
  - `copper`
  - `gold`
  - `iron`
  - `kovar`
  - `nickel alloy`
  - `palladium`
  - `platinum`
  - `rhodium`
  - `stainless steel`
  - `tungsten`
  - `unknown metal`

  Asking for the difference between the reactor material and a catalyst is a valid question as the boundaries are often blurry. Generally, the metadata in the PDB follows the assessment of the authors of the original publication, i.e. a material reported as catalyst is included in the `catalyst` category. Note also that for the reactor material the material's name is consciously written out instead of using the chemical formula to underline the inherent vagueness of the material composition. Usually, the reactor material is only superficially defined and characterised compared to catalyst materials. For instance, not all stainless steels are the same.

+ `reactor material in direct plasma contact 2 (mainly metal)`
  The `reactor material in direct plasma contact 2 (mainly metal)` field contains a material used in the reactor that is in direct contact with the plasma, e.g. the electrodes or the reactor wall. Even though principally arbitrary, in case of multiple materials including a dielectric material, e.g. in a DBD, the metal should be included as `reactor material in direct plasma contact 2 (mainly metal)`. If there is only one material, it is mentioned in the preceding parameter and this parameter is left empty.

- `power (W)`
  The `power (W)` parameter contains the power in W.

+ `relevant power`
  The `relevant power` parameter elaborates on the power determination method. The possible options are

  - `NA`
    If a power is given but it is unclear how it was determined `NA` is the correct choice.
  - `Lissajous`
    The Lissajous or Manley method [4] is the recommended way to determine the deposited power in a DBD. The voltage across the discharge gap is plotted against the transferred charge. From the area of the resulting figure the energy deposited in one period is determined and multiplied with the frequency and the power is obtained.
  - `U*I`
    The straightforward multiplication of voltage U and current I is often the only way to get the power.
  - subtractive method (RF or MW)
    In case of radiofrequency (RF) and microwave (MW) powered plasmas, the electrical signal behaves like a wave rather than as current so that a part of the provided power is actually reflected. The power delivered to the load is the forward power minus the reflected power.
  - `nominal readout generator`
    The lowest order approach is the nominal power value read from the power supply. Since it is unclear what fraction of this power actually ends up in the plasma, the nominal readout is diagnostically less conclusive.
  - `at plug`
    The power taken directly from the grid, i.e. at plug, appear as an even lower order approach than the nominal readout from the generator and suffers from the same problem of course. However, it comes with the additional benefit of aiding in an assessment of the total energy efficiency.
  - `other`
    In case of unfamiliar methods to determine the power that do not fit in one of the before mentioned options, `other` is the correct choice.

- `SEI (J/L)`
  The specific energy input (SEI) is probably the most popular process parameter as it is a measure of the energy deposited per molecule, mol or L. The SEI here is calculated according to

$$\mathrm{SEI}[\mathrm{J\,L^{-1}}] = \frac{P[\mathrm{W}]}{\phi[\mathrm{sccm}]} \times 1000\,\mathrm{mL\,L^{-1}} \times 60\,\mathrm{s\,min^{-1}}, \tag{2}$$

where P is the power and ϕ is the total inlet gas flow rate. Note that ϕ in literature often corresponds to the flow rate provided by the mass flow controllers in the experiment, i.e. no corrections to account for temperature and pressure are implemented. That way, the SEI must be interpreted as a macroscopic process parameter rather than a microscopic one. In the PDB, the same approach is followed. When performance parameters are reported against power and the total inlet flow rate is known, the SEI is calculated in the back end according to equation (2). The upside of this approach is maximum comparability as almost all publications give P and ϕ. Note however that few publications do correctly account for the differences of pressure and/or temperature, therefore the comparison amongst publications should be done with caution. It is also worth mentioning that in consequence the PDB might contain the same data set multiple times, namely when the original publication itself reports on the power, as well as the SEI dependence. The latter is simultaneously recalculated.

+ `SEI (eV/molecule)`
  In units of eV molecule^{-1}, the SEI resembles much more a microscopic parameter than in J L^{-1}. Its value is obtained from equation (2) by multiplying with the molar volume V_m and dividing by the electron charge e and the Avogadro number N_A. The calculation of V_m requires the gas temperature and the pressure in the plasma though. For the majority of publications, at least one of the two is not measured. In literature, there is the shaky practice of using $V_\text{m} = 24.5\,\text{mol L}^{-1}$ even though that value is only valid for standard conditions, i.e. room temperature and atmospheric pressure. It must be emphasised that a so-derived SEI grants not only less insight in the plasma process than the use of equation (2), but might also lead to erroneous conclusions. The complication gets consolidated by the fact that many authors giving the SEI in eV molecule^{-1} do not mention how V_m was calculated.
  In conclusion, in the PDB the following compromise is chosen. The SEI in J L^{-1} is the preferred value. The absence of ambiguity here is considered more important than the slight loss of microscopic insight. The SEI in eV molecule^{-1} is reduced to a conditional parameter that is never calculated but only given if the authors of the original publication use it. If either T or p are not known, the comparison is only possible within the small number of publications giving the SEI in eV molecule^{-1}. Because of the aforementioned reasons, that comparison must be inspected with utter caution.

- `voltage shape`
  The `voltage shape` field contains information about the applied voltage pattern. Valid options are

  - `none`
    In case of DC voltage or an applied MW field there is no shape defined. `none` is the correct choice then.
  - `sinusoidal`
    The most common voltage shape is a sinusoidal voltage at a certain frequency.
  - `ns-pulses`
    Voltage pulses below ţs length are included here.
  - `square long pulses`
    Voltage pulses longer than ţs are included here.
  - `other`
    Other voltage shapes that do no match the other options are included as `other`.

- `voltage (V)`
  The `voltage (V)` parameter contains the amplitude of the applied voltage in V. In case of MW excitation this field is left empty. For sinusoidal excitation half of the peak-to-peak voltage is given. An applied DC bias can be mentioned in the comments.

- `frequency (Hz)`
  The `frequency (Hz)` parameter contains the frequency of the applied electromagnetic field in Hz. In case of a DC voltage, the frequency is 0 Hz while for MW excitation the MW frequency is inserted.

- `modulation frequency (Hz)`
  The `modulation frequency (Hz)` parameter contains the frequency of the the modulating field. 0 Hz corresponds to no modulation.

- `main plasma diagnostic`
  The `main plasma diagnostic` field contains the main diagnostic used to characterise the plasma. Note that the diagnostic does not necessarily provide information to the remaining fields of the PDB. Although very valuable, an in-depth discussion of the ins and outs of each diagnostics is out of the scope of this paper. Hence only a brief summary of the information obtained by each diagnostic is mentioned here with some recommended literature. The valid options are

  - `none`
    In case the plasma was not characterised at all, 'none' is the correct choice.
  - `only electrical`
    In case only the electrical properties (voltage, current), mainly for power determination, were investigated, `only electrical` is the correct choice.
  - `Langmuir probe`
    A Langmuir probe was used to determine electron temperature, electron density and electric potential of the plasma [5].

  - `OES`
    The plasma was characterised by optical emission spectroscopy. OES can provide plenty of information regarding electronically excited species. Several parameters can be measured but it is not limited to the following: atomic density of O, N, H, etc. (by actinometry), estimation of reduced electric field, rotational and vibrational temperature, etc. [6, 7].
  - `Raman`
    A form of Raman laser scattering was used. Raman scattering is capable of quantifying densities of molecules such as CO_2, CO, O_2 and N_2 and rotational temperatures [7, 8].
  - `IR absorption`
    Infrared absorption spectroscopy was used to characterise the plasma for quantification of IR active molecule densities and vibrational temperatures [7].
  - `(TA)LIF`
    For the characterisation of the plasma a form of laser induced fluorescence spectroscopy was used. This technique is used for measurement of molecular and atomic densities but also for radicals (e.g. OH) [7, 9].
  - `CRDS`
    Cavity ring down spectroscopy was used to determine the absolute densities of gas phase species in the plasma [7].
  - `Rayleigh`
    Rayleigh laser scattering was used to characterise the plasma specially the gas temperature [7, 8].

5 Catalyst

The `catalyst` category is the most extensive one. Evidently, only heterogeneous solid catalysts are considered here. For clarity, it is divided in four subcategories.

5.1 Catalyst Coupling

The `catalyst coupling` subcategory contains only two essential fields.

- `coupling type`
  The essential `coupling type` parameter details how the plasma catalytic system is realised. Possible entries are
  - `none`
    In cases of pure plasmolysis without any deliberately added catalyst, `none` is selected. Note that `none` as coupling type also demands `none` in the following `position` parameter.
  - `packed bed`
    The catalytic material is fixed in the discharge volume fully or partially creating a bed of material.
  - `coating on wall`
    The reactor walls are partially coated with a catalytic material.
  - `fluidized bed`
    The catalytic material in the form of particles become fluidized when an upward-flowing gas imposes a high enough drag force to overcome the downward force of gravity making the particles behave like a fluid.
  - `monolithes`
    Monolithes allow to substructure the plasma-catalytic volume, usually in the form of individual channels.
  - `other`
    Other uncommon coupling types that do not fit any other option are grouped in `other`.
- `position`
  - `none`
    In cases of pure plasmolysis without any deliberately added catalyst, `none` is selected. Then all subsequent (conditional) `catalyst` fields are left empty.
  - `in situ`
    The catalyst is in direct contact with the plasma so that even short-lived species can reach the catalyst surface.

  - `early post-discharge`
    The catalyst is placed not directly in the plasma but shortly after. That way, some short-lived species might still reach the surface. Often this position is used when the plasma is very hot and could damage the catalyst.
  - `downstream`
    Plasma and catalyst are separated in space with the latter being somewhere downstream. Rather than a plasma-catalytic system, a plasma reactor is followed by a conventional catalytic system.

5.2 Catalyst Composition

The `catalyst general` subcategory identifies the catalyst consisting of support, active metal and potentially promoter. It contains three conditional fields.

+ `composition (activeMetal-promoter/support)`
  The `composition (activeMetal-promoter/support)` field allows to specify how the catalyst is built up. It is very important to follow the pattern, where the dash (`-`) separates the active metal and the promoter and the slash (`/`) both of them from the support, even if one or multiple components are missing. Usually, the chemical formula is used. For instance, an alumina support without any active metal and promoter would be `-/Al2O3`. An exception are zeolites, that usually have complex chemical formulas but relatively easy abbreviations. It is then recommended to use the abbreviation but indicate that it is a zeolite, e.g. `-/zeolite HZSM5`. The mixture of supports is indicated with (+).

+ `proportion (% mass ratio)`
  Similar to the `proportion (%/%/%...)` field in the `gases` category, the proportions in % of the used catalyst with respect to mass are given. Pay attention that the proportions follow the order active metal/promoter/support. When a component is missing, it has proportion `0`. If a mixture of support is used, the total of the sum of both components should be used.

+ `support`
  The support of the active phases should be indicated using the chemical formula except for zeolites, SiO2 materials, and others that might have particular composition but belonging to a specific group of materials. The latter is meant to be kept general to increase comparability though while the independent *support* parameter is used to give additional information. For example, `-/SiO2` is a valid composition for `quartz wool`, `SiO2 powder` or `glass` as support.

5.3 Catalyst Pre-treatment Before Reaction

The pre-treatment method is often crucial for the catalyst performance. The scope of pre-treatment methods in literature varies significantly though. As minimal common ground, only the very last pre-treatment step before use in the plasma-catalytic process is detailed. Often this occurs in the same reactor. The `catalyst last pre-treatment before reaction` subcategory contains four conditional fields.

+ `pre-treatment method`
  The `pre-treatment method` field coarsely classifies the pre-treatment method. Possible options are

  - `none`
    Select `none` if no pre-treatment method is used. In that case, all other parameters of the `catalyst last pre-treatment before reaction` category are left empty.
  - `thermal`
    The conventional approach is to heat the catalyst while exposing it to a given (static or flowing) gas. Calcination, annealing or reduction fall into this subcategory.
  - `plasma`
    A plasma is used for the pre-treatment of the catalyst under similar conditions as in thermal pre-treatment. In this case, plasma is the source of rise in temperature, electrons, ions and other reactive species that can induce calcination, annealing or reduction.

+ `pre-treatment gas composition`
  In principle, any gas composition can be used. At the time of writing, the existing options are `NA`, `Air`, `Ar`, `Ar/CH4`, `Ar/H2`, `H2`, `N2`, `N2/H2` and `O2`. `NA` is selected when the gas mixture is unclear. The proportions are no indicated in the database. If there is no pre-treatment the field is left empty.

- `pre-treatment temperature (K)`
  The `pre-treatment temperature (K)` parameter contains the final temperature used in the last pre-treatment step before catalyst use in K. If the `pre-treatment method` field is filled with `plasma` a temperature is often not known. `pre-treatment temperature (K)` parameter is then left empty.

- `pre-treatment time (min)`
  The `pre-treatment time (min)` contains the time the catalyst has been pre-treated in the last step in min.

5.4 Catalyst Conditions

The amount, density and macroscopic shape of the catalyst material that is selected for plasma catalysis plays a role in the results and measured activity. The `catalyst conditions` category requires the operational conditions regarding the catalyst to be filled in.

- `total catalyst mass used (g)`
  The total mass of the catalyst introduced in g.

- `catalyst density (g/cm3)`
  The bulk density of the used catalyst in $\mathrm{g\,cm^{-3}}$ as calculated from the total used catalyst mass divided by the total occupied volume including void space.

- `support shape (or bulk shape)`
  The macroscopic shape of support or bulk includes the following options

  - `NA`
    If the shape of the support or bulk is not specified in the publication `NA` is the correct choice.
  - `aerogel`
    Aerogel like materials includes a large variety of nanostructured porous solids. In the preparation of these materials, supercritical or subcritical conditions are used [10].
  - `beads`
    Beads are small spherical particles typically made of silica, alumina or glass.
  - `coating` The catalytic material is coated on the reactor walls or in another structure placed in situ, early post-discharge or downstream.
  - `foam`
    Foams are materials known for the low density and high porosity with a variety of composition (metal, ceramics, etc.).
  - `monolithes`
    Monolithes are macroscopic structures that usually contain some kind of channel the gas can flow through. A popular variation is the honeycomb structure used in cars exhausts.
  - `particles`
    `particles` is used as an umbrella term for small particles of undefined shape that are usually obtained from grinding larger particles. The only thing those particles have in common is their size that is defined by sieving them. In publications the size is then often given in mesh.
  - `pellets`
    Pellets are small cylindrical particles.
  - `powder`
    Very small particles that leave only very little void volume are entered as `powder`.
  - `wool`
    Very common material to use on plasma catalysis to hold powder materials in place or even as support of active phase.
  - `xerogel`
    Xerogels are similar to aerogels but they are prepared via subcritical drying while keeping a fraction of the original gel pore texture.
  - `other`
    Macroscopic structures that do not fit any other option, e.g. a mixture of different support shapes, are entered as `other`.

  Note that the options given above are not necessarily used consistently in the community. Usually, data is entered *as reported* with the options `NA`, `particle` and `other` being exceptions for grouping. However, the shape is really valuable only if the following field `size (mm)` is filled as well.

- `size (mm)`
  The size is intended as of the macroscopic structures the catalysts which could range from nanometer to milimeter scale.

5.5 Catalyst Characterization

To understand why a catalyst performs well, it must be well characterised. This `catalyst characterization` category considers the results of possible physicochemical characterization that has been carried out on the materials.

- `relative permittivity`
  It is rare to find this material property reported in plasma catalysis papers, despite being a very important feature of the packing material regarding its interaction with plasma and charged species.
- `nanoparticles size (nm)`
  Assuming that the active phase on the nanometer scale is supported on a given material, the size of such nanoparticles (NPs) should be indicated in this category. Electron Microscopy (TEM or SEM) is a common technique used for the measurement of NPs.
- `porosity (nm)`
  This property refers to the pore diameter, normally measured via nitrogen physisorption or Hg porosimetry depending on the porosity of the sample. The smaller a pore is, the harder it is for the plasma to enter. Therefore, we prioritize information about the larger pores (macropores) over the smaller ones.
- `catalyst pore volume (micro/meso/macro) (cm3/g)`
  The volume usually is obtained via nitrogen physisorption and given as a volume instead of diameter. Again, macropores are prioritized over meso and micro pores.
- `specific surface area (m2/g)`
  Also obtained via nitrogen physisorption, usually refers to the area calculated via BET theory.
- `main catalyst characterization diagnostic`
  The characterization techniques used in the publication that define the catalyst properties previously listed should be reported here using the common acronyms (e.g. XRD, TEM, TPR...). The description of such techniques are outside the scope of this publication.

6 Separation Unit

The separation of products from the processed gas stream can positively influence the performance. For example, the separation of oxygen prevents back reactions with CO to CO_2. Not many publications report on this possibility yet. Note that a cool trap after the reactor should not be listed here as such a trap separates products for analysis purposes and not to influence the product distribution. The `separation unit` category contains two essential fields.

- `separation type`
  The `separation type` field indicates what method is used for separation. Valid options are
  - `none`
    If no action is taken to separate products `none` is the correct choice.
  - `carbon bed`
    A bed of carbon catches oxygen and avoids back reactions with the freshly produced CO [11].
  - `MIEC membrane`
    A mixed ionic-electronic conducting (MIEC) membrane can be used to separate a product from the other. A popular option are oxygen-permeable membranes where the oxygen ions and electrons are transported in opposite directions. An example of MIEC membranes are solid oxide electrolyser cell (SOEC) membranes. The transport is driven by a partial pressure gradient between feed side and sweep side and potentially an applied voltage. Usually a high temperature is required for noticable fluxes but the membrane performance might actually be enhanced in contact with plasma [12].
  - `water`
    Generated products are dissolved in water and thereby protected from getting destroyed again.
  - `other liquid solvent`
    Instead of water also other liquids can be used the prevent the destruction of products.

  - `other membrane`
    Any membrane that is not MIEC is listed as `other membrane`.

- `separation position`
  See the `position` field in the `catalyst coupling` category 5.1.

Only these two categories are added but future developments will indicate how to extend this section and many more options can be added.

7 Output Data

The `output data` category connects the metadata with the performance data. It contains five field of which one is conditional.

- `main goal`
  For a coarse classification, the `main goal` field groups a study in one of the following classes
  - `application`
    The `application` option is somehow used as a wildcard for processes that do not fit the other options, e.g. when they are not aiming for CO_2 conversion.
  - `CO2 splitting`
    If the reported experiment aims to promote the reaction $2CO_2 \rightarrow 2CO + O_2$, `CO2 splitting` is the right choice. Note that this is not limited to pure CO_2 plasma but additional gases are not supposed to react with the other gas molecules (e.g. noble gases).
  - `dry reforming of methane`
    If the reported experiment aims to promote the reaction $CO_2 + CH_4 \rightarrow 2CO + 2H_2$, `dry reforming of methane` is the right choice. Note that this is not limited to CO_2 - CH_4 plasma but additional gases are not supposed to react with the other gas molecules. Water is explicitly excluded.
  - `hydrogenation`
    Hydrogenation is a general term for processes aiming to produce hydrogen-containing products from CO_2. Usually, the reaction is indicated as hydrogenation despite of also forming methane in CO_2-H_2 plasma.
  - `methanation`
    CO_2 methanation is a popular form of CO_2 hydrogenation promoting the reaction $CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O$.
  - `RWGS`
    The reverse water-gas shift (RWGS) reaction $CO_2 + H_2 \rightarrow CO + H_2O$ is promoted.

- `x axis`
  The `x axis` field contains mainly information about which process parameter that was used to vary the performance, i.e. the 'setscrew' that the experimenter can adjust from the outside to influence the process. Note that there is a hierarchy of process parameters that becomes important when performance data in the original publication is not given as a function of the the options listed below. For instance, a plot against the catalyst composition must be provided with a new x-value that might actually be the same for all points. Such occasions cause vertical lines in plots of the PDB data. By default the SEI in $\mathrm{J\,L^{-1}}$ should be used. If information is missing, in order of decreasing priority the power, the total gas flow rate or the initial CO_2 fraction should be used.
  - `SEI (J/L)`
    See information in the `plasma source` category.
  - `SEI (eV/molecule)`
    See information in the `plasma source` category.
  - `power (W)`
    See information in the `plasma source` category. When the total gas flow rate, i.e. the `gas flow (ml/min) or (sccm)` field in the `gas mixture` category, see section 3, is given, the SEI is calculated from the power according to equation (2).
  - `% CO2 (init. mix)`
    See information in the `gas mixture` category.
  - `gas flow (mL/min or sccm)`
    See information in the `gas mixture` category.

  - `temperature (K)`
    See information in the `gas mixture` category.
  - `pressure (mbar)`
    See information in the `gas mixture` category.
  - `conversion_CO2 (%)`
    The only performance parameter in the `x axis` field for the popular plot option energy efficiency vs. conversion.
  - `residence time (time in the plasma zone) (s)`
    See information in the `gas mixture` category.
  - `plasma pulse duration (t_on in s)`
    The time s that the plasma is on in every period.
  - `post discharge duration (t_off in s)`
    The time in s that the plasma is off in every period.
  - `process treatment time (min)`
    The total time in min that the plasma is operating in continuous flow condition. Often also mentioned as processing time, this parameter is used to illustrate the operation stability of the plasma(-catalytic) reactor.
  - `GHSV (h-1)`
    The gas hourly space velocity in h^{-1} is the ratio between volumetric flow of reactants (volume per time) and volume of the catalyst in flow reactors.

- `y axis`
  The `y axis` parameter contains so-called performance parameters, i.e. the quantitative measures used to assess the plasma(-catalytic) gas conversion process.
  The most important performance parameter is the conversion χ_i of reactant i (in %) that is assumed to be defined as

  $$\chi_i = \frac{n_i^{\text{conv}}}{n_i^{\text{prov}}}, \tag{3}$$

  where n_i is the number of reactant molecules (often in mol) and the superscripts indicate whether it is the number of molecules converted (conv) or initially provided (prov). Two kinds of conversions are included in the PDB:

  - `conversion_CO2 (%)`
    The reactant i in equation (3) is CO_2.
  - `conversion_CH4 (%)`
    The reactant i in equation (3) is CH_4.

  The second most important performance parameter is the energy efficiency η. Various definitions can be found in literature that can be differentiated by the unit of η. Often, η is given in units of molecule J^{-1}, assuming

  $$\eta = \sum_i \eta_i = \sum_i \frac{\dot{n}_i}{P}, \tag{4}$$

  where $\eta_i = \frac{\dot{n}_i}{P}$ is the energy efficiency of reactant i with $\dot{n}_i$ being the conversion rate of reactant i and P is the power. Following this definition, the PDB includes:

  - `en. eff. CO2 conv. (N_molecule/J)`
    The only reactant i in equation (4) is CO_2. Note that this does not mean that pure CO_2 is treated.
  - `en. eff. CH4 conv. (N_molecule/J)`
    The only reactant i in equation (4) is CH_4. Note that this does not mean that pure CH_2 is treated.
  - `en. eff. syngas prod. (N_molecule/J)`
    This is a production not a conversion energy efficiency. In equation (4), i corresponds to the components of syngas ($CO + H_2$) and $\dot{n}_i$ are their production rates [13].
  - `energy efficiency (N_molecules/J)`
    No explicit mentioning of the converted species indicates that multiple species have been converted. The typical example is dry-reforming of methane (DRM) where CO_2 as well as CH_4 are converted. In equation (4), the conversion rates of all molecules converted are simply summed up.

Another often encountered version of η is given in %. Note that this form of η is less consistently defined/used in literature. A common practice relates the energy efficiency of reactant i to the conversion defined in equation (3) and the standard reaction enthalpy ΔH_i

$$\eta_i = \frac{f_i \chi_i \Delta H_i}{\mathrm{SEI}}, \tag{5}$$

where f_i is the initial fraction of reactant i in the gas mixture.

- `en. eff. CO2 conv. (%)`
  The reactant i in equation (5) is CO_2 with $\Delta H_{CO_2} = 283\,\mathrm{kJ\,mol^{-1}}$ at standard conditions. From equation (5), we can also see that from a measurement χ_{CO_2} vs SEI, η_{CO_2} can be calculated. Many authors do so by themselves in which case data is simply extracted from the publication. Simultaneously, η_{CO_2} is automatically calculated in the back end of the PDB app. To that end, the SEI is converted to $\mathrm{kJ\,mol^{-1}}$ taking into account the real conditions (temperature and pressure) in the calculations of the molar volume.
- `energy efficiency (%)`
  No explicit mentioning of the converted species indicates that multiple species have been converted. The typical example is DRM where CO_2 as well as CH_4 are converted. Various definitions are found in literature, reaching from a simple sum like in equation (4) to more extensive calculations aiming for a global energy efficiency assessment by taking into account the electric energy consumption and the chemical energy stored in the molecules described by the lower heating value. Although we deem these considerations very valuable, caution is appropriate when comparing this performance parameter between different references.

The selectivity S_j of product j is assumed to be defined as

$$S_j = \frac{n_j}{\sum_i n_i \zeta_i}, \tag{6}$$

where $n_{j,i}$ are numbers of product j and reactant i, respectively, and ζ_i is the stoichiometric ratio of the limiting reactants. For instance, for the production of H_2 from CH_4 $\zeta_{CH_4} = 2$ as in principle methane potentially forms two hydrogen molecules. The concrete calculation of the selectivity hence depends on the used gas mixture which is why it is not further detailed here.

- `selectivity_CO (%)`
  Selectivity of carbon monoxide.
- `selectivity_H2 (%)`
  Selectivity of molecular hydrogen.
- `selectivity_CH4 (%)`
  Selectivity of methane.
- `selectivity_CH3OH (%)`
  Selectivity of methanol.
- `selectivity_CH3COOH (%)`
  Selectivity of acetic acid.
- `selectivity_C2H2 (%)`
  Selectivity of acetylene.
- `selectivity_C2H4 (%)`
  Selectivity of ethylene.
- `selectivity_C2H6 (%)`
  Selectivity of ethane.
- `selectivity_C2H5OH (%)`
  Selectivity of ethanol.
- `selectivity_C3H6 (%)`
  Selectivity of propylene.
- `selectivity_C3H8 (%)`
  Selectivity of propane.
- `selectivity_C4H10 (%)`
  Selectivity of butane.

The yield Y_j is the fraction of initial limiting reactant i that is transformed into a given product j, assumed to be defined as

$$Y_j = \frac{n_j}{n_i}. \tag{7}$$

The concrete calculation of the yield hence depends on the used gas mixture which is why it is not further detailed here. See the selectivity for the identification of all the products.

  - `yield_CO (%)`
  - `yield_H2 (%)`
  - `yield_CH4 (%)`
  - `yield_CH3OH (%)`
  - `yield_CH3COOH (%)`
  - `yield_C2H2 (%)`
  - `yield_C2H4 (%)`
  - `yield_C2H6 (%)`
  - `yield_C2H5OH (%)`
  - `yield_C3H6 (%)`
  - `yield_C3H8 (%)`
  - `yield_C4H10 (%)`

  - `carbon_balance (%)`
    The carbon balance is the ratio of moles of carbon atoms entering the reactor over the moles of carbon atoms leaving the reactor. A carbon balance below 100 % can be an indication of carbon deposition or polymerisation.
  - `ratio_H2/CO`
    The ratio of moles of H_2 produced over the moles of CO produced determine the potential for further processing of the syngas.

- `file ID`
  The `file ID` field contains name of the text file containing the performance data set corresponding to the metadata entry.

+ `comment`
  Information that is considered important for the interpretation of data but does not fit any of the preceding metadata fields can be stored as short as possible in the `comment` field.

8 Data processing

When data is reported according to the specified scheme above, the combined data and metadata can be read from disk and processed with some scripting. Most notably this performs data type coercion and extraction from a more flexible 'human-readable' format to a consistent, 'machine-usable' scheme. An example of this would be taking a gas composition string (e.g. 'CO_2/Ar/H_2O') and setting the proportions in the appropriate columns. Some of the computed columns are redundant to some extent –the `authyear` column for instance is just a concatenation of the `first author name` and `publication year (yyyy)` columns– but these are provided for ease of filtering or grouping data, avoiding frequent (re)computation.

8.1 Computed columns

By parsing the input data it is possible to increase the amount of columns in the database without changing the source format. The following columns are computed (or updated) in the back-end when the database is loaded from disk to memory.

`residence time` Calculated (using equation (1)) from available metadata if possible, aggregating array-like data by their mean (e.g. temperate ranges).

`catalyst active_phase` The catalyst metal active phase, as extracted from the `catalyst composition` column (section 5.2). Maintains an underscore in the name due how the regular expression is implemented.

`catalyst support` The catalyst support material, as extracted from the `catalyst composition` column (section 5.2). Note: this is different from the `support` column, which might contain a more descriptive name of the support.

`catalyst promotor` The catalyst promotor material, as extracted from the `catalyst composition` column (section 5.2).

`<species>` The respective percentage of total input gas flow for the species, extracted from the columns `gases` and `proportion (%/%/%...)` (section 3).
Possibilities: $CO_2, CH_4, H_2, Ar, N_2, He, H_2O, C_2H_4, O_2, C_2H_6, C_3H_8$.

`gas components` A string containing the gas composition, equivalent to the `gases` column (section 3), but enforcing a fixed order to the components for consistent sorting/grouping.

`authyear` The concatenation of the columns `first author name` and `publication year (yyyy)` for ease of filtering and grouping data, with an easily understood label. However, does not guarantee uniqueness and will result in grouping multiple publication by the same first author in a year together.

`Calculated` A column to indicate whether the (x,y) data values are those that were read from disk, or if some calculation has been applied to them (e.g. power in W converted to SEI in $\mathrm{J\,L^{-1}}$).

`Catalyst` Boolean column indicating the experiment used a catalyst (i.e. if there is form of catalyst coupling, implying a catalyst is being used) or not; for ease of filtering and grouping data.

`Support material class` Classification of support materials into some material classes (for now using a regular expression lookup table for zeolites, perovskites and ceria-zirconia); for ease of filtering and grouping data.

`Frequency range` Classification of plasma power driving frequencies (into either ‘MW’, ‘RF’, ‘up_to_300kHz’, ‘up_to_30kHz’, ‘up_to_3kHz’, thresholds chosen based on wavelength for the latter three) to complement the `source type` category.

`BibTeX code` The BibTeX citation for a publication, as obtained via the Digital Object Identifier (DOI) system, if available (see section 2).

8.2 Assumptions for calculations

A number of parameters are (re)calculated based on provided metadata for publications: the residence time (1), the SEI in $\mathrm{J\,L^{-1}}$ (2), energy efficiency of CO_2 conv. % (5) but also the `frequency range` column (section 8.1). Each of these calculations rely on data that potentially is given as a range of values, see section 1. By convention, the mean of the range is taken for each of the parameters. All entries for which a (re)calculation has been performed on either the x or y data, have the `Calculated` field set to ‘Calculated’ (rather than ‘Original’). Filtering the database on only `Original` data, this calculated data can be replaced with calculations by the user based on other aggregation, if desired.

As mentioned for equation (2), the calculation that converts power (in W) to SEI ($\mathrm{J\,L^{-1}}$) simply relies on the gas flow rate ϕ (in sccm or $\mathrm{mL\,min^{-1}}$) and therefore does not take into account if ϕ is defined w.r.t. standard conditions or the actual conditions in the discharge regarding pressure and temperature. Since generally in this case both the power and gas flux are know, the SEI can be calculated (except for batch reactors where there is no flow), which allows for a comparison between a larger body of experiments.

Contrary to this, the calculations for the residence time (1) and energy efficiency of CO_2 conversion both take into account the discharge conditions. This discrepancy is a conscious one: the SEI (in $\mathrm{J\,L^{-1}}$) is often reported as a macroscopic process parameter that is calculated from the power and fluxes put into the reactor. For τ_{res} taking into account the conditions inside the plasma is paramount, since it can provide the information on how long a particle is exposed to plasma conditions. Do note that in lieu of the plasma –or active zone– volume, for some experiments an approximation such as the reactor volume is used, see column `relevant volume` in section 4.

Likewise for the efficiency, it is important to account for the actual measurement conditions to establish the number of particles into which energy has been channelled. For *in situ* versus effluent conditions this can be very different. Overestimating the number density favourably improves efficiency, whilst an underestimation similarly negatively impacts it. While the way in which the PDB is structured and reports metadata is not without caveats (section 4.3 of the paper), the distinction between in- versus post-plasma dissociation measurements can be partially addressed by tailored filtering and aggregation of the data contained within.

8.3 Normalisations

Several normalisation functions are available in the interface of the database. As described in the main text, the aim to provide a tool for easy calculation of normalised data within the same interface and compare with different data sets. For flexibility sake, a wide set of normalisations are provided, without restriction on whether they are sensible in a given context. The purpose of each normalised function is described briefly and summarised in table 1 along with the equations utilised.

- `divide abundance`
  As mentioned in the online documentation, abundance A refers to the partial pressure times volume ($p_{\text{partial}} \cdot V$). This term gives an approximate amount or 'abundance' of a species under certain conditions without actually knowing the molecule density, due to the temperature not being known.

- `divide flow`
  The data is divided by the flow as stated in the publication.

- `divide flow stp`
  The data is divided by the flow corrected to standard conditions. This assumes that the flow stated in the publication has not been corrected.

- `divide fraction`
  The data is divided by the initial gas fraction of the given species.

- `divide partial density`
  The data is divided by the partial density when the pressure, volume and temperature are given, see formula in table 1.

- `divide pressure`
  The data is divided by the pressure

- `divide temperature`
  The data is divided by the pressure.

- `divide volume`
  The data is divided by the volume. This normalisation considers the value in the database regardless of the relevant volume.

- `do nothing`
  The data is plotted as it is, clears any previous manipulation

- `multiply abundance`
  The data is multiplied by abundance.

- `multiply flow`
  The data is multiplied by flow.

- `multiply flow stp`
  The data is multiplied by flow corrected to standard conditions.

- `multiply fraction`
  The data is multiplied by the initial gas fraction of the given species.

- `multiply partial density`
  The data is multiplied by the partial density when the pressure, volume and temperature are given.

- `multiply pressure`
  The data is multiplied by the pressure.

- `multiply temperature`
  The data is multiplied by the temperature.

- `multiply volume`
  The data is multiplied by the volume.

Table 1. Overview of the widgets and options for normalising data. The `*` character signifies either the multiplication or division operation.

Widget	**Options**	**Note**
`Normalise function`	`*_abundance`	$A = p_{\text{partial}} \cdot V$
	`*_flow`	
	`*_flow_stp`	$\phi = \phi_{\text{stp}} \cdot \frac{T}{273.15\,\text{K}} \cdot \frac{1013.25\,\text{mbar}}{p_{\text{partial}}}$
	`*_fraction`	
	`*_partial_density`	$n_{\text{partial}} = \frac{p_{\text{partial}}}{k_{\text{B}} \cdot T}$
	`*_pressure`	
	`*_temperature`	
	`*_volume`	
	`do_nothing`	Does not normalise; clears normalisations.
`Normalise on axis`	`x`	
	`y`	
`Normalise species`	`<species>`	Species to consider for the fraction, either: CO_2, CH_4, O_2, Ar, H_2, H_2O, N_2, He
`Normalise aggregate func.`	`min`	
	`mean`	
	`max`	

Either x or y data can be subjected to a normalisation function at a time. In addition, not only CO_2 can be used in the normalisation but also other species such as Ar or He as seen in table 1. Finally, when ranges of parameters are introduced to the database, mean, max and min values could be used to normalise, e.g. max of temperature. This feature allows flexibility to explore with any dataset. It is left to the user of the PDB App themselves to decide whether normalising the CO_2 concentration in the initial mixture by the temperature for example, is useful. Furthermore, any data exported directly from the App is never normalised data, the normalisation operations only update the graph visualisation and will be annotated on exporting the graph.

Acknowledgements

This project has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No. 813393

References

[1] M A Damen et al. “Temperature evolution in a pulsed CO_2-N_2 glow discharge measured using quantum cascade laser absorption spectroscopy”. In: *Plasma Sources Science and Technology* 29.6 (June 2020), p. 065016. DOI: 10.1088/1361-6595/ab8e50 (cit. on p. 5).

[2] N. Pinhão et al. “Influence of gas expansion on process parameters in non-thermal plasma plug-flow reactors: A study applied to dry reforming of methane”. In: *International Journal of Hydrogen Energy* 41.22 (June 2016), pp. 9245–9255. DOI: 10.1016/j.ijhydene.2016.04.148 (cit. on p. 5).

[3] M S Moss et al. “An investigation of CO_2 splitting using nanosecond pulsed corona discharge: effect of argon addition on CO_2 conversion and energy efficiency”. In: *Plasma Sources Science and Technology* 26.3 (Feb. 2017), p. 035009. DOI: 10.1088/1361-6595/aa5b1d (cit. on p. 6).

[4] T. C. Manley. “The Electric Characteristics of the Ozonator Discharge”. In: *Transactions of The Electrochemical Society* 84.1 (1943), p. 83. DOI: 10.1149/1.3071556 (cit. on p. 9).

[5] B. E. Cherrington. “The use of electrostatic probes for plasma diagnostics?A review”. In: *Plasma Chemistry and Plasma Processing* 2.2 (June 1982), pp. 113–140. DOI: 10.1007/bf00633129 (cit. on p. 10).

[6] P. G. Reyes et al. “Optical emission spectroscopy of CO_2 glow discharge at low pressure”. In: *physica status solidi c* 5.4 (Apr. 2008), pp. 907–910. DOI: 10.1002/pssc.200778306 (cit. on p. 11).

[7] Richard Engeln et al. “Foundations of optical diagnostics in low-temperature plasmas”. In: *Plasma Sources Science and Technology* 29.6 (June 2020), p. 063001. DOI: 10.1088/1361-6595/ab6880 (cit. on p. 11).

[8] A F H van Gessel et al. “Laser scattering on an atmospheric pressure plasma jet: disentangling Rayleigh, Raman and Thomson scattering”. In: *Plasma Sources Science and Technology* 21.1 (Jan. 2012), p. 015003. DOI: 10.1088/0963-0252/21/1/015003 (cit. on p. 11).

[9] R. Ono et al. “OH radical measurement in a pulsed arc discharge plasma observed by a LIF method”. In: *IEEE Transactions on Industry Applications* 37.3 (2001), pp. 709–714. DOI: 10.1109/28.924749 (cit. on p. 11).

[10] Atsushi Ishihara et al. “Large Mesopore Generation in an Amorphous Silica-Alumina by Controlling the Pore Size with the Gel Skeletal Reinforcement and Its Application to Catalytic Cracking”. In: *Catalysts* 2.3 (Sept. 2012), pp. 368–385. DOI: 10.3390/catal2030368 (cit. on p. 13).

[11] Fanny Girard-Sahun et al. “Carbon bed post-plasma to enhance the CO_2 conversion and remove O_2 from the product stream”. In: *Chemical Engineering Journal* 442 (Aug. 2022), p. 136268. DOI: 10.1016/j.cej.2022.136268 (cit. on p. 14).

[12] Qiankun Zheng et al. “Coupling of dielectric barrier discharge plasma with oxygen permeable membrane for highly efficient low-temperature permeation”. In: *Journal of Membrane Science* 641 (Jan. 2022), p. 119896. DOI: 10.1016/j.memsci.2021.119896 (cit. on p. 14).

[13] Kuan Lun Pan et al. “Dry Reforming of CH_4 With CO_2 to Generate Syngas by Combined Plasma Catalysis”. In: *IEEE Transactions on Plasma Science* 42.12 (Dec. 2014), pp. 3809–3818. DOI: 10.1109/tps.2014.2360238 (cit. on p. 16).