Level 3 data product for the Electron Proton Helium INstrument

Janna Martens

E-mail: `martens@physik.uni-kiel.de`

Abbreviations

CIR corotating interaction region

CME coronal mass ejection

COSTEP Comprehensive Suprathermal and Energetic Particle Analyser

EPHIN Electron Proton Helium INstrument

ESA European Space Agency

FMD Failure Mode D

FME Failure Mode E

GLE ground-level enhancements

GCR galactic cosmic ray

LION Low Energy Ion and Electron Instrument

NASA National Aeronautics and Space Administration

PHA pulse-height analysis

PHR pulse-height ratio

SOHO Solar and Heliospheric Observatory

SEP solar energetic particles

1. Introduction

This study gives an insight into the background of the creation of the Level 3 data product for EPHIN. A short introduction for the instrument and the SOHO mission is given, followed by an explanation of the Level 1 data products used for the creation of the Level 3 data. The aim of the Level 3 data product is to

(i) simplify masking the data with settings of the ring, quiet time detection, or avoiding overflows in the detectors,

(ii) increase accuracy by using the correct calibration of the instrument for the energy losses in the detectors and

(iii) enable flux calculation by including parts of science data in the pulse-height analysed data (see Section 3).

The data product itself will be introduced, as well as the algorithm used for its creation. Lastly, an exemplary spectrum is determined.

2. Instrument

2.1. Solar and Heliospheric Observatory

SOHO is a joint mission by the National Aeronautics and Space Administration (NASA) and the European Space Agency (ESA). Its primary scientific objective is to explore the Sun, including its core, corona, and the Solar Wind [2].

For this purpose, the spacecraft orbits around the Lagrangian point L1. This point lies on the line connecting Earth and the Sun, at a distance of 1.5 million kilometers from Earth. At this position, the orbital period around the Sun is synchronised with that of Earth, allowing for an uninterrupted view of the Sun. The spacecraft's orbit is depicted in Figure 1. In 1998, SOHO temporarily lost contact with the mission operations centre, resulting in a data gap of approximately six months. Fortunately, the spacecraft was fully recovered and was able to resume its nominal observations.

The Electron Proton Helium Instrument described in Section 2.2 is part of the Comprehensive Suprathermal and Energetic Particle Analyser (COSTEP) experiment onboard the SOHO spacecraft. In addition to EPHIN, the Low Energy Ion and Electron Instrument (LION) is also a part of this experiment.

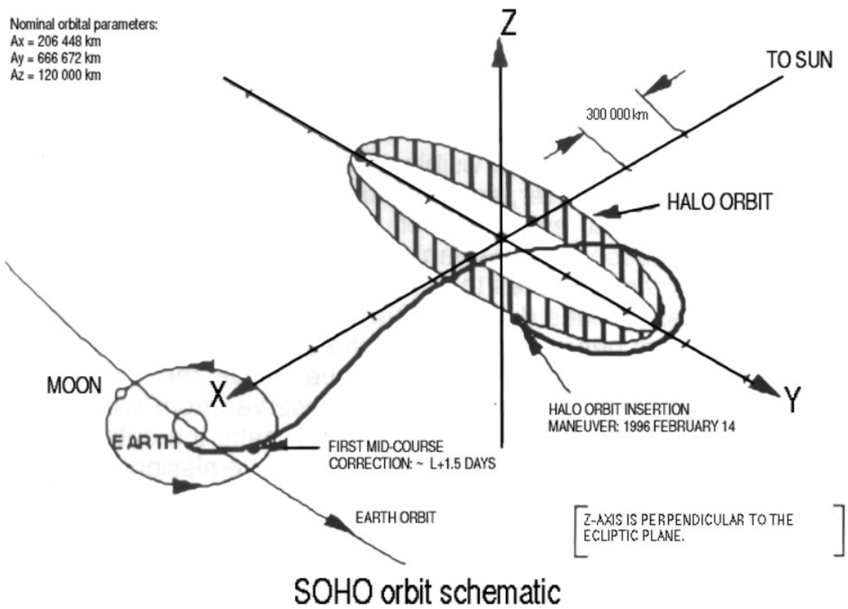

Figure 1. Schema of the SOHO orbit around L1 (`www.nasa.gov/mission_pages/soho/`, last accessed: August 9th, 2023).

2.2. Electron Proton Helium INstrument

The setup of EPHIN is illustrated in Figure 2. It comprises six different semiconductor detectors A-F. Detector G serves as the anticoincidence component and is connected to a photomultiplier. Detectors A and B are ion-implanted detectors, which enables the segmentation as depicted in the bottom right corner of the figure. The production method of the detectors is further explained in [4].

The segmentation enables differentiation between various angles of incidence. Additionally, the *ring* (consisting of the outer segments A1-A5 and B1-B5) can be deactivated during periods of heightened particle fluxes to reduce the geometry factor. The count rates in the central segment of detector A, A0, determine, whether the ring of detectors A and B is enabled or disabled.

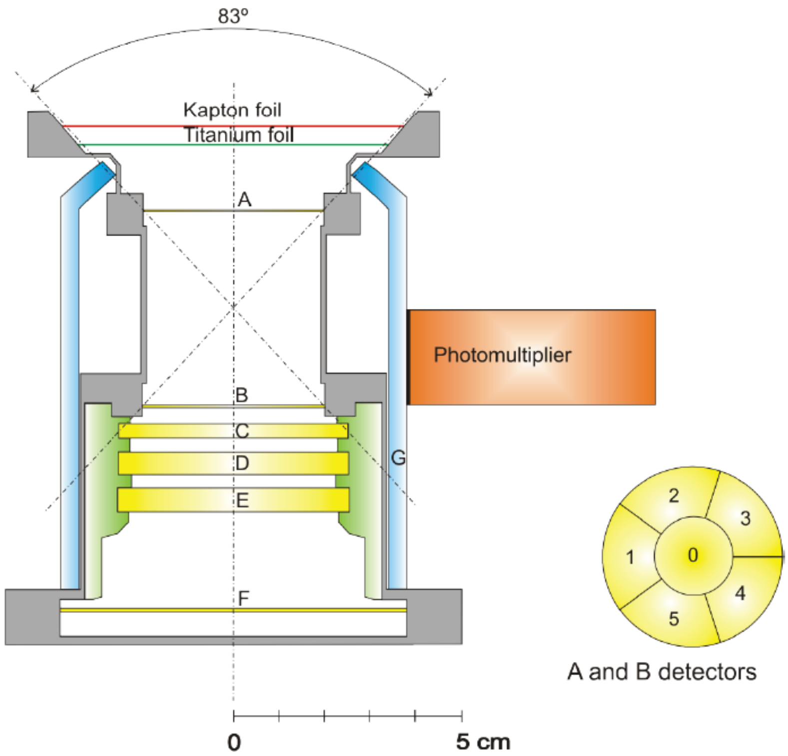

Figure 2. Setup of the Electron Proton Helium Instrument [1].

Detectors A to E, with detectors C to E being thicker than detectors A and B, provide pulse-height analysed output. Consequently, when properly calibrated, it is possible to determine the energy deposited by a particle in these detectors. This capability does not extend to detectors G and F. Since these detectors are incorporated to achieve a precise definition of the geometric boundaries of the instrument, only the knowledge about whether or not the response threshold was exceeded is important for the analysis.
The Kapton foil includes an aluminium layer for the reflection of light to protect the instrument from heating. The titanium foil shields the underlying detectors from light. These foils are very thin, permitting forward incident particles to enter the instrument without significant energy loss.

2.3. Coincidence logic
The $\frac{dE}{dx}$-E-method employed in EPHIN operates on the fundamental principle of determining particle types based on their energy loss in detector A and their penetration depth as a measure of kinetic energy. There are 13 different particle channels, with four channels each dedicated to electrons, protons, and helium, respectively. Additionally, one channel is designed for penetrating particles that pass through detector F. Every detector has a threshold below which the particle will not be considered. Detector A, in particular, has four different thresholds for the distinction between the various particle types. These are listed in Table 1 and the corresponding coincidence conditions are shown in Table 2. The threshold conditions for detectors A and B remain consistent irrespective of the segment penetrated by the particle. An 'A0' indication signifies that the particle has surpassed this threshold, while '$\overline{\text{A1}}$' implies that the energy loss falls below this threshold.

Threshold	A0	A1	A2	A3	A4	B0	C0	D0	E0	F0	G0
Value/keV	30	270	973	2000	5320	60	359	581	582	150	~100

Table 1. Thresholds for the different detectors used for the coincidence logic [4]. The threshold for detector G is calculated from the threshold G0 = 1.60 pC of the scintillation counter.

The priority specifies which particles (with high priority) will be considered for the data with pulse-height analysis (PHA), while those with low priority are overwritten during periods of elevated count rates when it becomes impractical to record all data.
The coincidence logic is used for a preliminary distinction of particle types. Heavier ions with $z > 2$ are sorted into the helium channels.

Particle	Channel	Energy range	Priority	Coincidence condition
Electron	E150	0.25-0.70 MeV	4	$A0\ \overline{A1}\ B0\ \overline{C0}\ \overline{D0}\ \overline{E0}\ \overline{F0}\ \overline{G0}$
	E300	0.67 - 3.00 MeV	4	$A0\ \overline{A1}\ B0\ C0\ \overline{D0}\ \overline{E0}\ \overline{F0}\ \overline{G0}$
	E1300	2.64 - 6.18 MeV	4	$A0\ \overline{A1}\ B0\ C0\ D0\ \overline{E0}\ \overline{F0}\ \overline{G0}$
	E3000	4.80 - 10.4 MeV	4	$A0\ \overline{A1}\ B0\ C0\ D0\ E0\ \overline{F0}\ \overline{G0}$
Proton	P4	4.3 - 7.8 MeV	4	$A1\ \overline{A4}\ B0\ \overline{C0}\ \overline{D0}\ \overline{E0}\ \overline{F0}\ \overline{G0}$
	P8	7.8 - 25.0 MeV	4	$A1\ \overline{A3}\ B0\ C0\ \overline{D0}\ \overline{E0}\ \overline{F0}\ \overline{G0}$
	P25	25.0 - 40.9 MeV	4	$A1\ \overline{A2}\ B0\ C0\ D0\ \overline{E0}\ \overline{F0}\ \overline{G0}$
	P41	40.9 - 53.0 MeV	4	$A1\ \overline{A2}\ B0\ C0\ E0\ D0\ \overline{F0}\ \overline{G0}$
Helium	H4	4.3 - 7.8 MeV/nuc	40	$A4\ B0\ \overline{C0}\ \overline{D0}\ \overline{E0}\ \overline{F0}\ \overline{G0}$
	H8	7.8 - 25.0 MeV/nuc	16	$A3\ B0\ C0\ \overline{D0}\ \overline{E0}\ \overline{F0}\ \overline{G0}$
	H25	25.0 - 40.9 MeV/nuc	4	$A2\ B0\ C0\ D0\ \overline{E0}\ \overline{F0}\ \overline{G0}$
	H41	40.9 - 53.0 MeV/nuc	4	$A2\ B0\ C0\ D0\ E0\ \overline{F0}\ \overline{G0}$
Integral	INT	E > 8.70 MeV P > 53.0 MeV H > 53.0 MeV/nuc	0	$A0\ B0\ C0\ D0\ E0\ F0\ \overline{G0}$

Table 2. Coincidence channels of EPHIN, with energy ranges, priorities, and conditions [4].

2.4. Malfunctions

SOHO/EPHIN has delivered great quantities of data during the past 27 years. Nevertheless, several difficulties have arisen, since two of its semiconductor detectors are no longer functioning properly. Figure 3 illustrates the single count rates for all detectors on SOHO/EPHIN along the instrument's internal temperature measurements. Detector E is showing a significant increase in the count rates early in the mission. This led to Failure Mode E (FME) being implemented in 1997. In the later stages of the mission, detector D behaved similarly. Failure Mode D (FMD) was implemented in 2017. Due to the malfunctioning detectors D and E, the coincidence logic no longer functions as originally intended. For each particle type, instead of having four coincidence channels, only two channels remain operational. The E150, P4, and H4 channels remain properly defined because particles stopping inside detector B can be accurately distinguished. When the particle further penetrates into the instrument, however, there is no distinction possible between it stopping inside either detector C, D, or E. As a result, the other coincidence

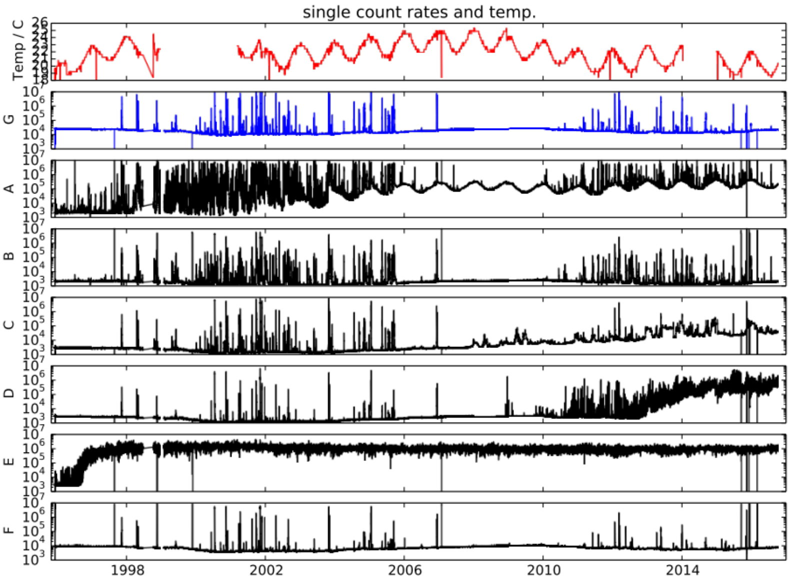

Figure 3. Single count rates for all detectors of SOHO/EPHIN from 1995 to 2018, as well as the temperature of the instrument [3].

channels are merged. Fortunately, the integral channel can still be used, as detector F exhibits no unusual properties.

3. Data products

There are several data products converted from the binary Level 0 data provided by the COSTEP experiment.

3.1. Level 1

Generally, there are 5 different types of Level 1 data for EPHIN.

(i) Science data, including counters and histogram data (.sci)

(ii) Pulse-height analysed data, all coincidence types, nominal observation (.pha)

(iii) Pulse-height analysed data, stopping particles, nominal observation (.phx)

(iv) Pulse-height analysed data, all coincidence types, non-nominal observation (.phr)

(v) Housekeeping data (.hsk)

3.1.1. Science data The Science data shows one row per 59.9 seconds (NOTE: for Chandra/EPHIN this is 65.6 seconds). The data structure is given in Table 3. The single detector counters are a measure of how many particles hit this detector in the specified minute. The specifications GM, GR, S1, S23 and S indicate different incidence angles:

GM particles passing through A0 and B0

GR particles passing through corresponding ring segments (A1/B1, A2/B2, ...)

S1 particles passing through one centre segment and one outer segment

S23 particles passing through different outer segments

Column	Property	Type
TIME TAGS		
0	Year	integer
1	Day of Year	integer
2	Millisecond of Day	integer
3	Minutes since launch + 2 decimals (part of minute)	float
4	Milliseconds since Year 1	integer
SINGLE DETECTOR COUNTER		
5	G0	integer
6-11	A0, A1, A2, A3, A4, A5	integer
12-17	B0, B1, B2, B3, B4, B5	integer
18-21	C, D, E, F	integer
COINCIDENCE COUNTER		
22-24	P4GM, P4GR, P4S	integer
25-27	P4GM, P4GR, P4S	integer
28-31	H4GM, H4GR, H4S1, H4S23	integer
32-35	H8GM, H8GR, H8S1, H8S23	integer
36-39	E150, E300, E1300, E3000	integer
40	INT	integer
41-43	P25GM, P25GR, P25S	integer
44-46	P41GM, P41GR, P41S	integer
47-50	H25GM, H25GR, H25S1, H25S23	integer
51-54	H41GM, H41GR, H41S1, H41S23	integer
55-60	CT0, CT1, CT2, CT3, CT4, CT5	integer
HISTOGRAM CHANNEL		
61-92	Bin1, ..., Bin32	integer
EPHIN STATUS INFO		
93-97	first 5 EPHIN status bytes	integer
98	LION dataset length	integer
99	EPHIN dataset length	integer
100	status flag	integer

Table 3. Data structure of Level 1 .sci data.

S sum of S1 and S23

To obtain the total number of particles being detected in these particle channels, data from these columns must be added. The electron channels have already been summed over all directions. *CT* refers to control channels. The Histogram data provides an estimate of the energy spectrum during times of elevated particle fluxes, where the memory in the scientific data frame is filled rapidly.

The status bytes correspond to different states of the instrument's operation, such as the detectors' CPU status and failure modes. The status bytes in the SCI dataset is composed of 8 bits, the interpretation of which is given in Table 4.

The LION and EPHIN dataset length correspond to the Level 0 data.

The status flag is a shortened value for the determination of the instrument's operational status. The status flag (or *status word*) is created through the combination of status Bits with a specific statement about the condition of the instrument. An overview is given in Table 4. Nominal observation refers to a mode of observation, where High Voltage is on, no failure mode is enabled and the ring segment switching is disabled. In failure mode E (and failure mode D, respectively) no pulse-height analysis is triggered by a detection event in the detector.

Flag Bit Value	Remarks
0	Nominal observation
1	Failure mode E
2	Ring A/B off
4	E patch uploaded
8	Commissioning
16	Standby or maintenance
32	Calibration
64	Ring segment switching enabled
128	Failure mode D

Table 4. The calculation of the status word is performed through the combination of the active Bits.

3.1.2. Pulse-height analysed data The following data structure applies to the PHA, PHR, and PHX data. The PHX data exclusively encompasses stopping particles, while PHR data is generated when the instrument operates in non-nominal mode (e.g. when detector A0 registers counts in saturation and the particles can not be properly detected). In the following, each dataset will be referred to as PHA for simplification.

Each row inside the PHA files contains information about an individual detected particle, resulting in multiple lines per minute. The dataset's columns are shown in Table 5. The coincidence logic is detailed in Table 2. The ADC value can be converted into the energy loss of the particle within the respective detector. This is dependent on the Low/High Flag. For low energy losses, the *high gain* is employed, while high energy losses must be detected using the *low gain* mode to prevent detector overload.

3.1.3. Housekeeping data The .hsk dataset contains information about the instrument's electronics, the spacecraft's properties and attitude data. However, this dataset is not utilised for the creation of the Level 3 data format and will not be further explained.

Column	Property	Type
TIME TAGS		
0	Year	integer
1	Day of Year	integer
2	Millisecond of Day	integer
3	Minutes since launch + 2 decimals (part of minute)	float
PHA DATA		
4	Coincidence	integer
5	Segment in A	integer
6	Segment in B	integer
7	ADC value in A	integer
8	Low/High Flag in A	integer
9	ADC value in B	integer
10	Low/High Flag in B	integer
11	ADC value in C	integer
12	Low/High Flag in C	integer
13	ADC value in D	integer
14	Low/High Flag in D	integer
15	ADC value in E	integer
16	Low/High Flag in E	integer
17	Priority Flag	integer
18	Status Flag	integer

Table 5. Data structure of pulse-height analysed data.

3.2. Level 2

The Level 2 dataset is derived from the Level 1 data. The calculation of particle energy loss relies on conversion factors documented in [4]. Fluxes of the particles can be computed as well, incorporating a fixed geometry factor for each coincidence channel. This geometry factor varies with the setting of the ring due to a variation in the geometric structure due to the change in detector sizes. In the case of stopping particles, with a known total energy, this can be transformed into particle intensities. The reader is referred to the Level 2 data documentation for more information on this data product.

3.3. Level 3

The Level 3 dataset is structured similarly to the Level 1 PHA data. Each row corresponds to the pulse-height analysed output of one individual particle.

There are several differences to the Level 1 data format, such as the time tags. In addition to the year and day of year, the month and day of the month are included. Several columns from the Level 1 data are directly transferred, such as the coincidence, the segments in A and B, and the priority and status flags. The ADC values from the Level 1 data are converted to energy losses. For SOHO/EPHIN, the conversion factors from Table 7 are employed.

These values refer to the full scale of the ADC, meaning that this range of energy is distributed

Column	**Property**	**Type**
TIME TAGS		
0	Year	integer
1	Month	integer
2	Day in Month	integer
3	Day of Year	integer
4	Millisecond of Day	integer
DATA		
5	Coincidence	integer
6	Segment in A	integer
7	Segment in B	integer
8	Priority Flag	integer
9	Energy loss in A	float
10	Full scale in A	float
11	Energy loss in B	float
12	Full scale in B	float
13	Energy loss in C	float
14	Full scale in C	float
15	Energy loss in D	float
16	Full scale in D	float
17	Energy loss in E	float
18	Full scale in E	float
19	Status Flag	integer
20	Ring ON/OFF (0/1)	integer
21	Number of counts for coinc. in L1 SCI-File	integer
22	Number of PHA words for coinc.	integer
23	Overflow Flag	integer
24	B0 counts	integer

Table 6. Data structure of Level 3 data.

across 1024 (0...1023) values. The values from Table 7 corresponding to the detector and High/Low Gain setting are utilised in place of the High/Low Flag from the Level 1 data product. The energy losses ΔE in the detectors are calculated as

$$\Delta E = ADC \cdot FS/1023. \tag{1}$$

ADC refers to the value taken from the Level 1 data and FS is the corresponding value from Table 7.
The Ring ON/OFF Flag in column 20 is set to 0 when the ring is active and 1 when the ring is deactivated during periods of high particle fluxes. This information is extracted from the status flag.

Detector	High gain/MeV	Low gain/MeV
A	3.153	31.53
B	3.067	46.01
C	15.321	158.93
D	20.762	233.57
E	22.387	251.85

Table 7. Conversion factors for EPHIN determined by M. Hörlöck.

Occasionally, not every particle can undergo proper pulse-height analysis, resulting in a discrepancy between the number of PHA words for a given coincidence channel and the corresponding counts from the Level 1 .sci file. Consequently, both of these values are included in the Level 3 data product.
In cases where one or more detectors register an overflow (ADC value equals 1023), rendering it impossible to accurately resolve the energy loss, the corresponding particle is flagged using the overflow flag in column 23.This flag is presented in a format that directly indicates which detector experienced the overflow. Specifically, '1' corresponds to detector A, '2' to detector B, and so on. An overflow in all detectors would be denoted by a value of '12345'. The numbers for detectors without overflow are set to '0'. Therefore, an overflow in detectors B, C, and E would be represented by a value of '2305'. By restricting the overflow flag to '0', only properly resolvable energy losses are considered in the analysis.
Additionally, column 24 provides the B0 counter from the Level 1 .sci data, which can be used for the determination of quiet time periods without large solar events.

4. Algorithm for the Level 3 data product creation

The subsequent section provides an explanation of the algorithm employed in generating the Level 3 data product. In the second to last cell (referring to the Jupyter notebook file), the data range can be specified. Currently, this is done by entering a start year and day of year and a range for both properties. An improvement involving entering start and end dates will be included when possible.
There is an option to set the variable `test` to `True`, which triggers the conversion using values from [4] to ensure data comparability with the Level 2 dataset. This approach was used during the validation process.
The cell below is used to run the necessary functions.

```
#start measuring computing time
startTime = time.time()

pha_paths = []
sci_paths = []
pha_paths, sci_paths =
    get_entries(startyear,rangeyear,startdoy,rangedoy)

for i in range(len(pha_paths)):
    startloopTime = time.time()
    year = pha_paths[i].parent.parts[7]
    name = pha_paths[i].stem[3:]
    doy = name[2:5]
    try:
        data_pha, data_sci = load_data_local()
```

```
    except UnboundLocalError:
        continue

    complete_conversion(name, year, doy, data_pha, data_sci)
    executionTime = (time.time() - startloopTime)
    print('Execution time: ' + str(round(executionTime,2)) + ' s')

#check how long that took
executionTime = (time.time() - startTime)
print('Complete execution time: ' +
    str(datetime.timedelta(seconds=executionTime)))
```

The time required for each data file conversion as well as the cumulative computation time is printed. The function `get_entries` returns a list of the paths for the Level 1 PHA and SCI data files within the specified time period. For each item in the list, corresponding to one data file per day, the conversion process is executed. The Level 1 data files are then read using the `load_data_local()` function.

Within this function, the `pandas read_csv()` method is used to read the data files. However, certain days, specifically October 31, 1996, and October 4, 2017, are excluded as Failure Modes E and D were activated on those days. Conversion for these dates will be integrated at a later stage. Additionally, an error message is generated if a file cannot be located. In the event of such occurrences, the variables `data_pha` and `data_sci` are set to 'None,' resulting in an error, and the next data file is processed.

If the reading of the data was successful, the function `complete_conversion` is called.

```
def complete_conversion(name, year, doy, data_pha, data_sci):
    if len(data_pha) != 0:
        #SET OVERFLOW FLAG
        data_pha = set_overflow_flag(data_pha)

        #REMOVE CONSECUTIVE DUPLICATES (MORE THAN 3 IN A ROW)
        data_pha = remove_duplicates(data_pha)

        #DELETE UNREALISTIC SEGMENTS IN DETECTORS A AND B (>5)
        data_pha = delete_segments(data_pha)

        #CALCULATE ENERGIES FROM PHAs
        data_pha = calc_energies(data_pha)

        #DAY, MONTH
        data_pha = data_pha.reset_index(drop=True)
        day, month = get_date(data_pha)
        data_pha['month'] = month
        data_pha['day'] = day

        #RING ON/OFF FLAG
        data_pha = ring_set(data_pha)

        #NUMBER COINCIDENCES IN TIMEFRAME + NUMBER PARTICLES WITH SAID
           COINCIDENCE IN SCI FILE
        if int(year) < 1996 or (int(year) == 1996 and int(doy) < 305):
           #NO FAILURE MODES!
```

```
            channels = ['E150', 'E300', 'E1300', 'E3000', 'P4', 'P8', 'P25', 'P41', 'H4', 'H8', 'H25', 'H41', 'INT']
        elif (int(year) == 1996 and int(doy) > 305) or 1996 < int(year) < 2017 or (int(year) == 2017 and int(doy) < 277): #FME
            channels = ['E150', 'E300', 'E1300', 'E1300', 'P4', 'P8', 'P25', 'P25', 'H4', 'H8', 'H25', 'H25', 'INT']
        else: #FMD and FME
            channels = ['E150', 'E1300', 'E1300', 'E1300', 'P4', 'P25', 'P25', 'P25', 'H4', 'H25', 'H25', 'H25', 'INT']
            nopha = [1,2,5,6,9,10] #coincidences where no pha words should occur
            data_pha = data_pha.drop(data_pha[data_pha['coinc'].isin(nopha)].index)

        data_pha = coinc_counts(data_pha, data_sci, channels)

        #ADD B0 COUNTS FROM SCI FILE
        data_pha = add_b0(data_pha, data_sci)

        #WRITE IN DATA FILES
        data_pha = write_data(name, year, doy, data_pha) #data_hsk
        write_overflow_file(name, year, data_pha)
    else:
        print("EMPTY FOR " + name)
```

If the data frame for the Level 1 PHA data is not empty, the conversion process begins. This starts with the `set_overflow_flag` function, which determines the positions where overflow occurs and creates a new column for the overflow flag following the previously explained principle. To eliminate consecutive duplicates in the data frame, the `remove_duplicates` function is employed. Occasionally, the same PHA word is recorded multiple times, likely due to a malfunction rather than proper particle detection. Any minutes associated with such occurrences are removed from the data. Additionally, minutes are filtered from the data frame when non-existent segments are detected. Since detectors A and B are subdivided into segments 0 through 5, higher segment numbers suggest issues in the data generation process. This is accomplished using the `delete_segments` function.
Next, the energies and full scales from Table 7 are calculated using the function `calc_energies`, depending on the value of the High/Low flag in the Level 1 dataset.
The time tags are adjusted by converting the day of year into the month and day using the `get_date` function. Here, the pandas function `to_datetime` is applied. The column for the indication of the ring setting is generated inside the function `ring_set`, which examines the state of the bit corresponding to the setting of the ring within the status word.
The determination of the number of PHA words for one coincidence and minute and the corresponding number of counts in the Level 1 SCI file is complicated by the failure modes. Prior to October 31, 1996, the observation was nominal and all coincidence channels could be used as intended. Between November 1, 1996 and October 3, 2017, only failure mode E was active. During this time, no distinction could be made between the coincidence channels 2 and 3 (corresponding to E1300 and E3000), as well as 6/7 (P25/P41) and 10/11 (H25/H41). Since October 4, 2017, both failure modes D and E have been implemented, leading to a merging between coincidence channels 1, 2, 3 and 5, 6, 7 as well as 9, 10, 11, respectively. This is implemented by adjusting the `channels` list. The following function can then be used for the determination of the number of PHA words (`counts_pha`) and the number of counts in the SCI file (`counts_sci`).

```python
def coinc_counts(data_pha, data_sci, channels):
    #COMBINE DIRECTIONS PROTON AND HELIUM CHANNELS TO ONE CHANNEL
    data_sci['P4'] = data_sci['P4GM'] + data_sci['P4GR'] + data_sci['P4S']
    data_sci['H4'] = data_sci['H4GM'] + data_sci['H4GR'] + data_sci['H4S1'] + data_sci['H4S23']
    data_sci['P8'] = data_sci['P8GM'] + data_sci['P8GR'] + data_sci['P8S']
    data_sci['H8'] = data_sci['H8GM'] + data_sci['H8GR'] + data_sci['H8S1'] + data_sci['H8S23']
    data_sci['P25'] = data_sci['P25GM'] + data_sci['P25GR'] + data_sci['P25S']
    data_sci['H25'] = data_sci['H25GM'] + data_sci['H25GR'] + data_sci['H25S1'] + data_sci['H25S23']
    data_sci['P41'] = data_sci['P41GM'] + data_sci['P41GR'] + data_sci['P41S']
    data_sci['H41'] = data_sci['H41GM'] + data_sci['H41GR'] + data_sci['H41S1'] + data_sci['H41S23']

    #FOR COUNTS OF PHA-WORDS: GROUPBY MSOD AND COINC, CALCULATE SIZE, WRITE IN ORIGINAL DATAFRAME
    data_grouped = data_pha.groupby(['msod','coinc']).size()
    data_grouped = data_grouped.reset_index()
    data_grouped = data_grouped.rename(columns={'0': 'coinc_pha'})
    data_grouped.columns = ['msod','coinc','coinc_pha']
    data_pha = pd.merge(data_grouped, data_pha, on=['msod','coinc'], how = 'right')
    data_pha = data_pha.rename(columns={'size':'coinc_pha'})

    #FOR COUNTS IN SCI-FILE: MERGE SCI FILES WITH PHA FILES, WRITE CORRESPONDING COINCIDENCE COUNT IN 'COINC_SCI' COLUMN IN ORIGINAL DATAFRAME
    data_pha = pd.merge(data_sci[['year','doy','msod','E150','E300','E1300',\\
            'E3000','P4','P8','P25','P41','H4','H8','H25','H41','INT']],\\
            data_pha, on=['year','doy','msod'], how = 'right')
    data_pha['coinc_sci'] = 0
    for i in range(len(channels)):
         data_pha.loc[data_pha.coinc == i,'coinc_sci'] = data_pha[channels[i]]

    #IN CASE THERE ARE MISSING TIMES IN SCI-FILE -> NaN WILL BE PUT INTO PLACE (MUST BE DELETED)
    data_pha = data_pha.dropna(how='any')

    return data_pha
```

Firstly, the different incidence angles for proton and helium channels must be combined into one channel each, such as P4 and P25. To determine the number of the PHA words for each coincidence, a new data frame is generated in which the data is grouped by the *millisecond of day* and the *coincidence* column. For each combination of the two, the group's size is determined.

Subsequently, this data frame is merged with the original data frame, incorporating the group sizes into a new column.
For the counts of the coincidence channel in the SCI file, the PHA data frame and part of the SCI data frame are merged, based on the time columns. The `channels` list is used to ensure the correct selection of the coincidence channels.
Occasionally, time stamps present in the PHA file do not correspond to entries in the SCI file, resulting in generated NaN values, which are subsequently removed.

Afterward, the B0 counts from the Level 1 SCI file are included in a new column via the function `add_b0`. Finally, the generated data frame is written into a file using the `write_data` function and the overflow file is appended with the rows showing an overflow in any detector, as handled by the `write_overflow_file` function.

5. Changes for Chandra/EPHIN
There is a slight distinction between the Level 3 data product of SOHO/EPHIN and that of Chandra/EPHIN. Due to Chandra's orbit residing within the Earth's radiation belts, an additional column has been introduced as a *position flag*. The value of this flag is derived by comparing the Level 1 SCI data between SOHO and Chandra. When Chandra/EPHIN's B0 counts are 1.7 times or greater than those of SOHO/EPHIN's, it indicates that Chandra is positioned inside the radiation belts. In such cases, the flag is assigned a value of 1; otherwise, it remains at 0 for times outside the belts.
Additionally, unlike SOHO/EPHIN, the Level 1 data for Chandra/EPHIN does not contain a status word. Consequently, it has been manually computed using the following binary bits.
Furthermore, the time resolution is slightly different. While SOHO/EPHIN provides a time

Flag Bit Value	Remarks
1	Automatic Ring switching (1:on, 0:off)
2	Failure Mode A
4	Failure Mode G
8	Ring on/off (0:on, 1:off)

Table 8. Status Bits for Chandra/EPHIN.

stamp every 59.9 seconds, whereas Chandra/EPHIN has an interval of 65.5 seconds.
Additionally, no recalibration has been executed for Chandra/EPHIN. Consequently, the conversion for the energy losses employs the conversion factors introduced in [4] instead of those presented in Table 7.

References

[1] Raúl Gómez Herrero. *Partículas energéticas en la heliosfera interna (1996-2000). Respuesta instrumental y observaciones del sensor EPHIN embarcado en el Observatorio Solar y Heliosferico SOHO (ESA-NASA).* PhD thesis, 2003.

[2] NASA. Solar and Heliospheric Observatory, 2021. `https://www.nasa.gov/mission_pages/soho/overview/index.html`.

[3] Patrick Kühl. Documentation of the SOHO/EPHIN Level3 data product, 2018.

[4] Holger Sierks. *Kosmische Teilchen im Sonnensystem - Messung geladener Teilchen mit dem Kieler Instrument EPHIN an Bord der SOHO-Raumsonde - Ideal und Wirklichkeit.* PhD thesis, 1997.