CMB-S4 Data Visualization

Astrophysics Capstone

Jeremy Webb, Advisor: Colin Bischoff

Spring 2023

1 Overview

The project referred to as Cosmic Microwave Background-Stage 4 (CMB-S4) has many primary science goals and countless more science can be done with its data besides. Through unprecedented precision, requiring exponentially more data to be collected than ever previously done, CMB-S4 will be able to probe earlier in the history of the universe than any before.[1] A combination of both large and small aperture, ground-based telescopes will allow many different types of observations to be conducted. Small aperture telescope arrays (SATs) will allow a deep observation of a narrow region of the sky. LATs will allow shallow observations of a wide region. Telescopes will primarily use detector pairs observing at two different frequencies each, for a total of eleven different observing bands.

The scope of the observations to be done means that a tremendous amount of data will be collected during each observation, compounding over the life of the project. Eleven observing bands across tens of thousands of detectors, each taking data during an observation at a rate of between 20 and 440 Hz results in billions of data points over a single observation, and countless more over the array's lifetime. This data must be cleaned and processed before being analyzed. [3] The process of analyzing big data sets such as this has historically been done piecemeal through scripts run by individuals as the data is collected. A more efficient approach is warranted as data sets get exponentially larger, as is required for the precision sought by this project. One of the options for storing, cleaning, and processing this data is utilization of the Structured Query Language, or SQL (sometimes pronounced 'sequel'). This language allows for logical structuring of data, a minimization of duplicate information, and relatively efficient, complex queries over multiple tables.

Before construction of the project can begin, simulations of time-ordered observing data, as well as simple statistics thereof, are generated. [4] These simulations allow the construction of the relevant data processing pipelines to be formatted prior to the collection of real observing data. An ambitious goal for this project was to design a website meant for public access to the observing statistics. With this tool, any user could use the gathered data to generate plots of any segment of the data. This would allow more free access to the project's data and more efficient cleaning of the raw data, which would be especially beneficial considering the amount of related science which can be done with the given observations [2].

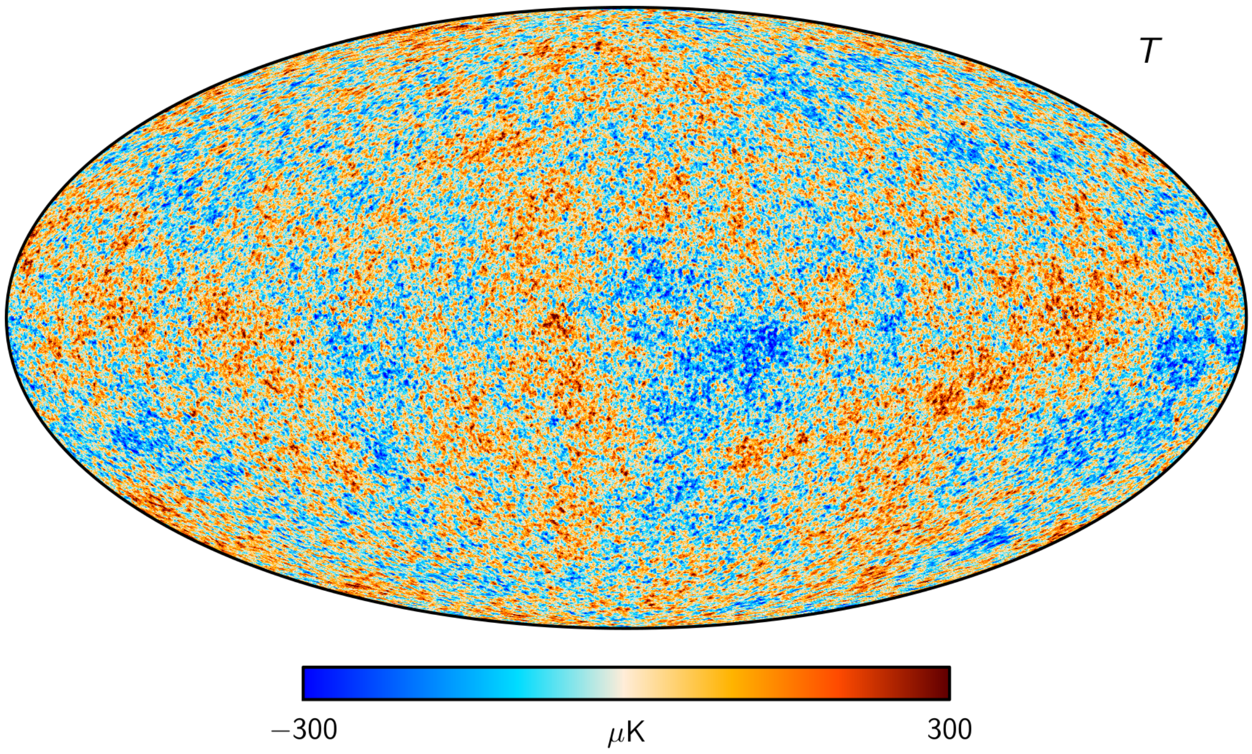

Figure 1: CMB temperature map. Results of Planck-2020-A4[2]

2 CMB Science

Of all use cases for CMB-S4's data, there are a few primary and many secondary goals. These goals stretch across many disciplines and, through its mm-wavelength observations, the project will provide data unprecedented in scope. Herein will be discussed one of the primary goals for the project. This single goal is presented to establish a scale for just how much science can be done with the project. Other project goals will be touched on only briefly, however a reader should bear in mind that even the primary goal description has been condensed[3] and each other goal is comparatively deep. The primary goal of CMB-S4, as mentioned in the introduction, is establishing stricter bounds on the tensor to scalar ratio of primordial perturbations in the early universe. To understand this, an introduction to CMB science is warranted and follows.

The cosmic microwave background (CMB) is light from the moment of last scattering in the early universe. At that point, the plasma that was the universe became transparent to photons. These photons were scattered by electron collisions until this moment, where the universe cooled enough to allow atoms to form. The free electrons became bound and photons were now able to move in a much less impeded path.

In a perfectly homogeneous ideal fluid, there would be no fluctuations in any aspect of the fluid on any scale. Given the complicated nature of fluid dynamics, however, this is not realistically possible; it certainly was not the case in the early universe. Evidence for inhomogeneities in the early universe can be seen by any observation of the CMB. One such observation is apparent in the CMB temperature map shown above (see Figure 1). It should be noted that there are many steps in processing observational data before eventually arriving at such an image. The fluctuations seen in CMB temperature are evidence of primordial perturbations in density and temperature in the early universe. These also lead to the CMB's polarization. This is the focus of the science which aims to detect gravitational wave relics from before the time of last scattering.

To understand how gravity waves can be probed with the CMB, one must first understand its polarization. Light is geometrically polarized in two separate components, called E and B modes. These modes correspond to the light's orientation and amplitude and are separated by a 45 degree rotation about a vector normal to it's plane, i.e. the direction the wave travels. For example, the component of a wave which is moving into the page/screen that is oriented toward the top and bottom is its E-mode, while the portion oriented toward the corners are its B-mode.

E and B modes are created in different ways. E-modes originate with scalar density perturbations, while B-modes were, if correct, sourced from primordial gravitational waves.[4] B-modes, however, may also be present in current observations of the CMB due to more recent gravitational lensing. Distinguishing the original B-modes from those created more presently can be accomplished by, along with removal of galactic foreground noise, observing small angular scales. This allows B-modes which are caused by gravitational lensing to be removed, since this type of polarization should fluctuate significantly with analysis of separate, small-angular-scale observations. The ratio of these perturbations is referred to as r and is a parameter which can be found by taking the power spectrum of CMB maps.[3]

CMB-S4 has a few primary goals: a more precise measurement of the tensor to scalar ratio, r, as mentioned above; measuring light relics from the early universe, which are predicted by some models[5] as a component of the dark energy of the universe; detection of mass density fluctuations in the late universe by gravitational lensing- largely from dark matter; detection of transient astrophysical phenomenon in the mm-wavelength range. A great deal more science can also be done outside of these primary goals.[1]

3 CMB-S4 Structure

Data produced by CMB-S4 will begin as raw temperature data. This raw data will be cleaned and analyzed to determine viability. Examples of non-viable data may include that which is collected during weather phenomenon in the atmosphere, previously unknown faults in the detectors, or any number of others. The raw data will be collected and stored in a table before being cleaned.

Once an observation's data is cleaned, statistics will be generated based on the observation. These statistics will be discussed further in the eponymous section below, 4.3. Information pertaining to the observations themselves will be stored in a separate table and is also discussed in its own sections below, 3.2 & 4.2. Another set of tables will be created, prior to observations being made, which contain data about the detectors to be used. [3.1 & 4.1] This is where the more in-depth discussion of the architecture will begin.

From above, the timeline for the project's data will be, roughly, as follows. First, after construction of the telescopes are complete and calibrations and verification have been done, tables will be constructed which will store information about all detectors in all telescopes (see sections 3.1 & 4.1). A new table will be made which will hold information regarding observations to be made. [3.2 & 4.2] This will be updated as needed throughout the life of the project. Once an observation has been made, the raw data will be stored separately according to established protocols. This will be analyzed and cleaned before being tagged as nonviable or viable. The viable data will then have various statistics generated from it. [4.3] These statistics are the basis of the science to be done.

3.1 Focalplanes

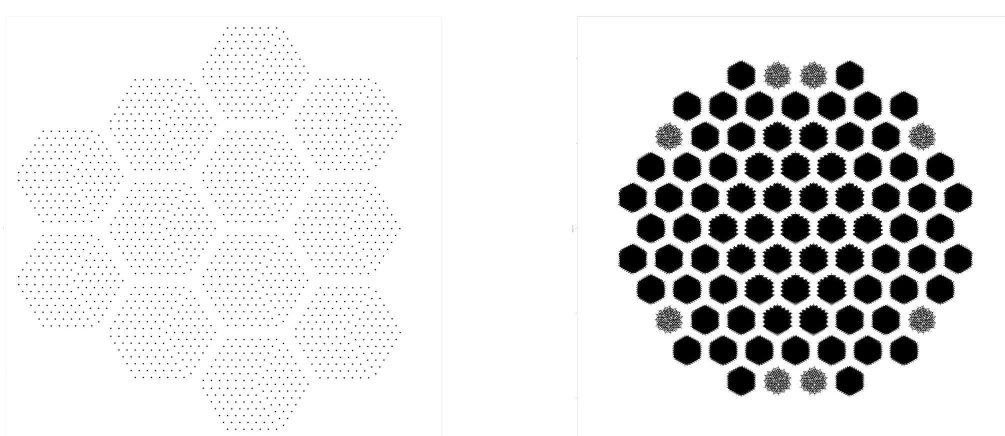

Figure 2: Sample focalplanes for a SAT (left) and LAT (right). The focalplane for the SAT is one of three which will be located in separate, but connected tubes. Each individual dot in the image corresponds to a detector pair. The hexagonal structure of each image are single wafers. In the LAT image, with the exception of the 30 GHz wafers, no individual detectors are discernible due to their much greater density compared to the SAT.

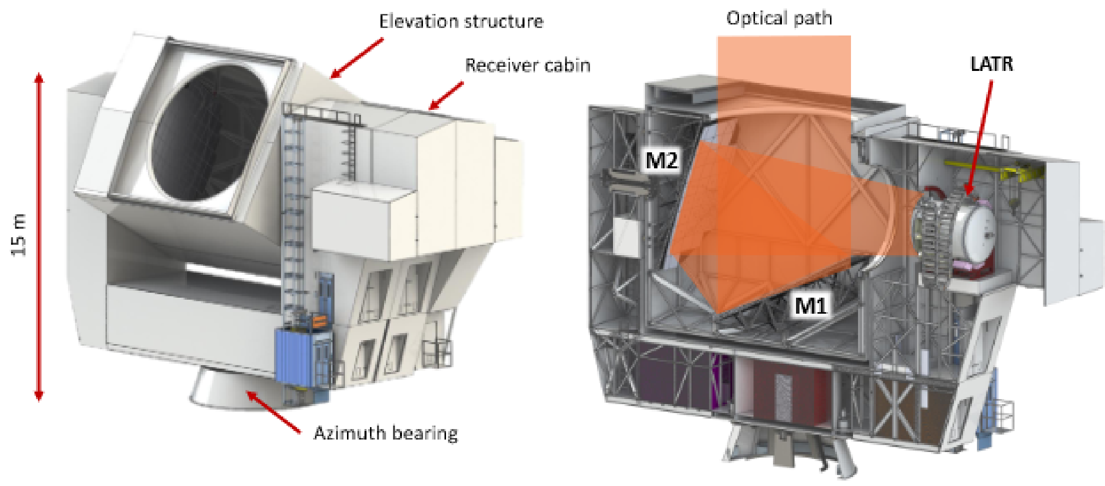

Figure 3: Images of the proposed LAT telescopes. The separation of the elevation structure housing the first two reflection points, M1 and M2, from the Receiver (LATR) is necessary due to the size of the detector array. It would be unfeasible to construct a LAT in which all pieces move co-dependently in all directions.[6]

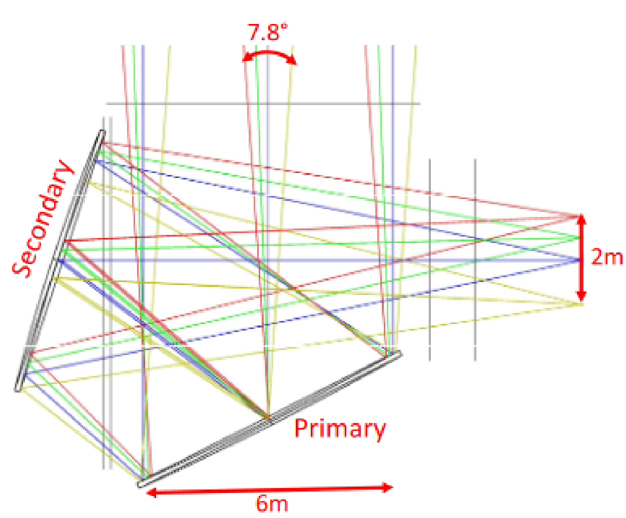

Figure 4: Examples of light paths for a LAT telescope. The telescope would be capable of observing 7.8 degrees of separation at once. The housing for these mirrors would be contained separately from the receiver.[7]

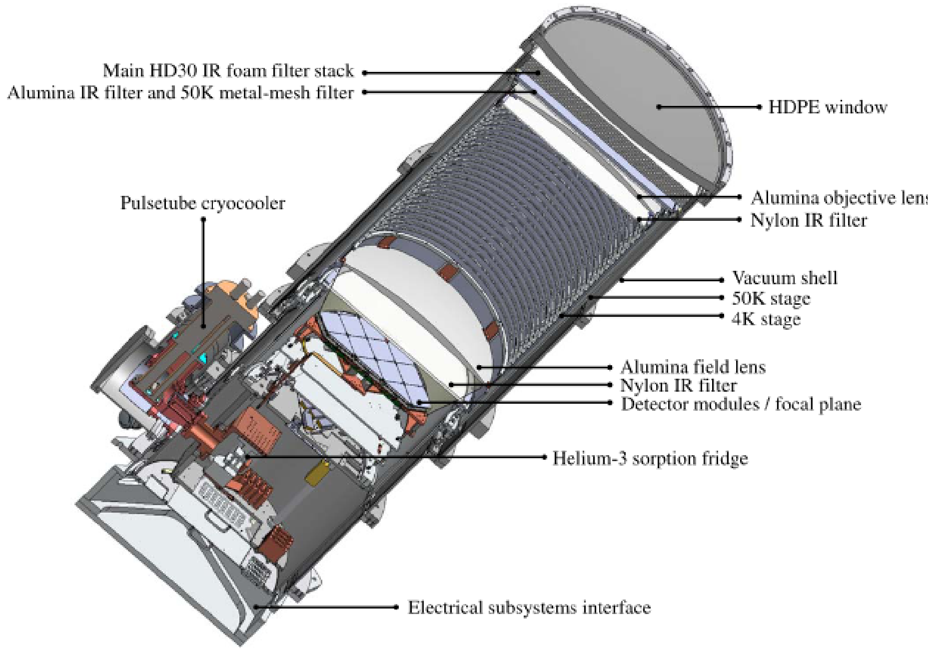

Figure 5: A figure of a BICEP array SAT. This single tube design will be utilized thrice over in each SAT for CMB-S4. These three apparatuses will house a focalplane each with two observing bands per tube.[8]

An example of a focalplane for a LAT and a SAT can be seen in Figure 3. Note that the project outlined herein utilizes data only from LATs, though the methods are generally applicable. Both of these types of telescope arrays have a distinctive hexagonal layout to the arrangement of detectors. Each of these hexagons is referred to as a wafer. These wafers lie within tubes which contain lenses to manipulate the path of the light being observed. Each of these wafers contains detectors, the number of which is mostly determined by the wavelength of light to be observed. Shorter wavelength detectors are, themselves, smaller. Since wavelength and frequency are inversely proportional, there is a greater density of higher frequency detectors per wafer. This can be seen in the LAT figure (2) where the wafers containing only low frequency detectors have a more open structure than those containing higher frequency detectors, which have no distinguishable internal structure due to the density of detectors within.

The difference in focalplanes between the LATs and SATs stems from the telescope structure itself. As can be seen in Figures 4 & 6, the light ray paths for each telescope are quite different. By definition, a LAT must have a substantially larger collecting area than that of a SAT. This necessitates a housing which can move largely independent of collecting area; this is opposed to a SAT, whose collecting tube(s) moves as a whole. Also relevant to construction of focalplane housings is the use case for each telescope. More information about observations to be done can be found in 3.2 & 4.2.

Figure 5 shows a single SAT tube. Contrary to the case of BICEP, for example, CMB-S4 will have not one, but three of these tubes working in concert for each SAT. This allows for scans of a small region deep enough to achieve the accuracy necessary for constraint of r. These tubes will each house a focalplane, an example of one of which can be found in Figure 2. This

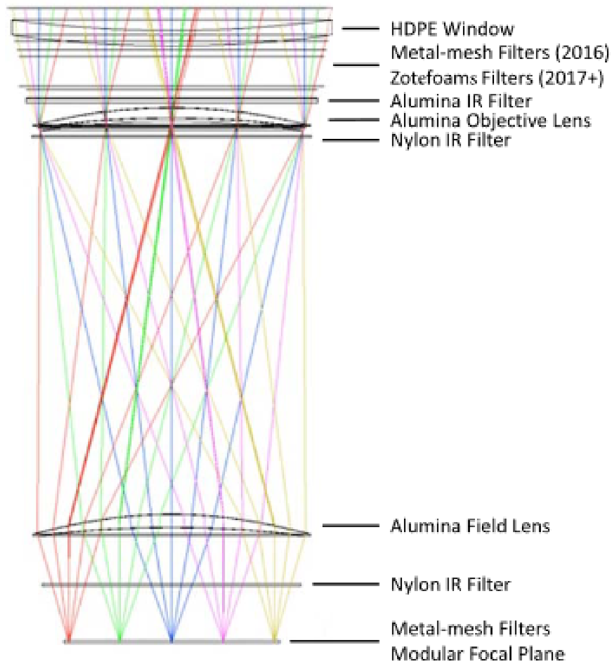

Figure 6: An example of the light rays for the BICEP SAT. mirrors in each tube focus the light on different detectors. Each cavity in the tube will be maintained at a different temperature, with the focalplane housing resting at an order of magnitude of hundreds of milliKelvin. This assures no detector heat is mistaken for signal from the 2.7 Kelvin CMB observations.[8]

focalplane layout, consisting of a number of wafers, each made of individual detector pairs, allows for greater strength in the area of the focalplane which is weakest[9] and is largely designed with engineering concerns in mind. Each detector in a pair will observe at separate frequencies. For example, the focalplane shown in the figure above (LHS, 2) has detector pairs observing at 85 and 145 GHz. The layout of the proposed higher and lower frequency tubes' focalplanes are slightly different to account for the increased and decreased number of detectors, respectively. Despite the differences, it maintains the overall structure of wafers of detector pairs.

The structure of the LAT, with a receiver separate from the primary and secondary reflection points, necessitates a single focalplane, as shown also in Figure 2. Toward the edges of the focalplane lie the lower frequency detectors, which can be distinguished by having few enough detectors that white space can be seen as plotted. The middle of the focalplane contains those of highest frequency, and can be distinguished by their more jagged edges. Between and around these lie the middle frequencies' detectors. The middle frequency detectors occupy a greater number of wafers simply because there are more frequencies to be observed in these bands. The proposed layout consists of two LATs, each with eight observing bands ranging from 30 to 270 GHz. The breakdown is proposed as a single low frequency band, two high frequency bands, and five between, near the ideal CMB observing frequencies.

3.2 Observations

Observations can be broken into two main categories based on CMB-S4's science goals. Large angular scale (covering roughly 60% of the sky), shallow, but sensitive observations of the sky are to be conducted by the LATs based in the Atacama Desert in Chile. Deep observations of a small angular scale (roughly 35 degrees) region of sky will be done by SATs located mostly at the South Pole facility. To meet these goals, wildly different approaches must be utilized.

South Pole observing is centered around accounting for the rotating sky. At the geographic pole, where Amundsen–Scott South Pole Station is located, stars never set, instead rotating around the celestial pole. This means that, barring weather and solar/lunar activity, observations of a single patch can be done continuously. The science goal of observing deeply in a single region of sky would be horribly inefficient anywhere accept the South Pole. For example, SAT observing schedules in Chile focus on multiple regions that they prioritize observing, but spend much of their time shifting from one to the next as each rise and set with the rotation of the Earth.

Chile observing is based mostly around the two LATs which will be observing there. These will work in conjunction to observe shallowly over large swaths of the sky. The sheer volume of detectors per LAT means that a significant amount of raw data is collected during each observing period. These periods are broken up by calibrations and maintenance, but otherwise are an almost continuous cycle. One key characteristic of LAT observations is that the telescope has historically been required to slow down near the edges of the focus of the patch in order to change direction. This leads to a feature where the edges of the observing patch contain an increased number of hits.

4 Database

The first challenge in this project was a choice in how the data should be accessible. Since much of the work on previous stages of CMB observation has been implemented through Python and scripts within it, this was a logical starting point. Within Python, there are numerous ways to handle large quantities of data. Chief among them are SQL and Pandas. SQL, or Structured Query Language, allows tables to be constructed whose entries are rows of information about a single object, with the information being segregated into separate columns. This matrix-like organization is convenient to build, query, and visualize. Although Pandas has much of the same framework, it differs in a few main ways. Pandas is a Python module, so has much less new syntax to learn, though SQL has SQLite, their own abbreviated version of SQL for use within Python. Pandas is also much less optimized for querying large volumes of data, whereas SQL was built with

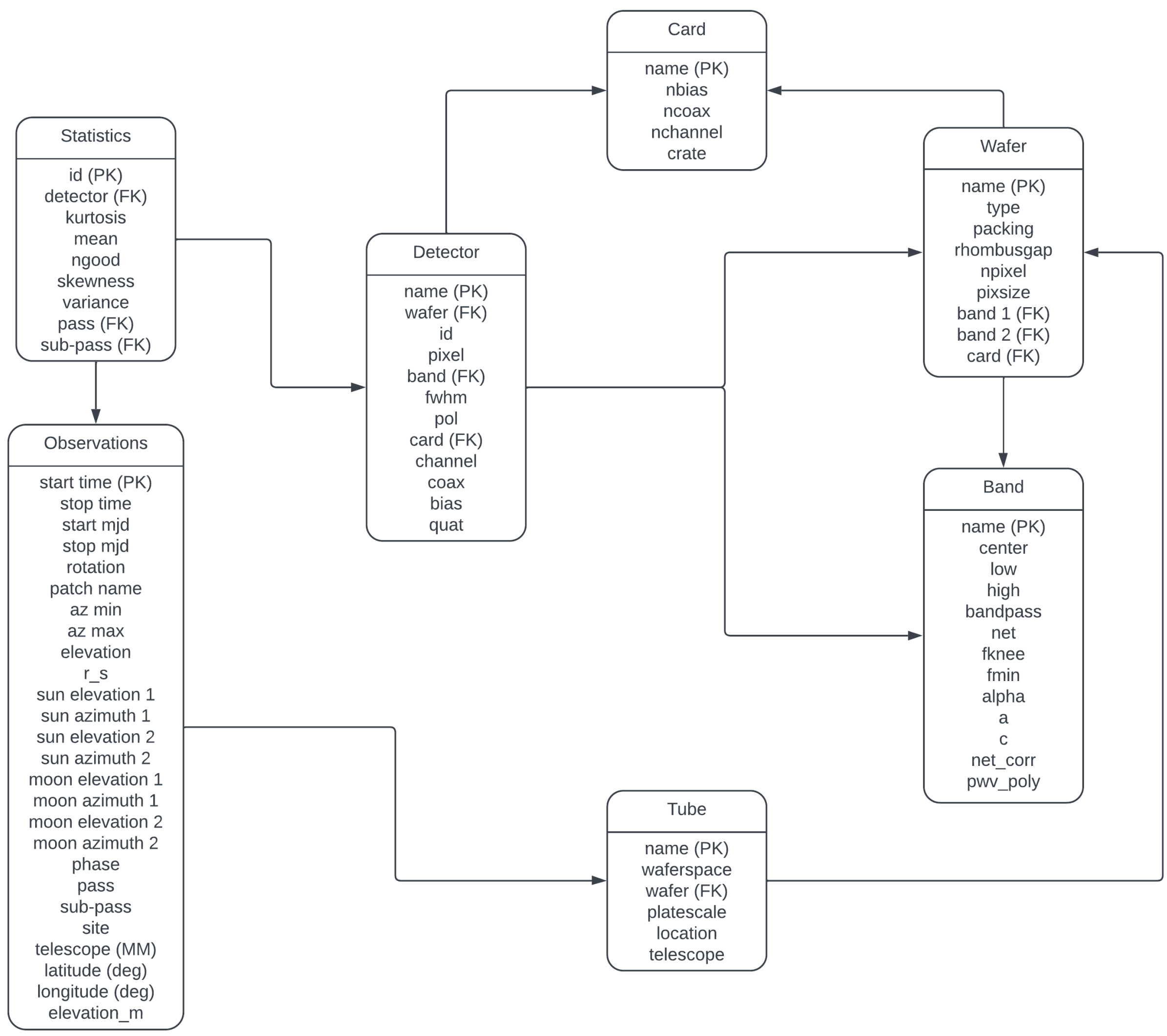

Figure 7: CMB-S4 Data Architecture. Entries labelled PK correspond to the primary key of each table. Those labelled FK are table Foreign Keys referencing the corresponding entry in the table(s) connected by an arrow.

this exact operation in mind. Lastly, Pandas does have many more built-in functions which make complex queries much easier. This, however, can be mitigated by the use of specific built-in data classes in SQL, which will make complex queries much simpler and will reduce the overall volume of data to be stored significantly. The aforementioned built-in data classes in SQL are known as keys. These keys allow simpler complex queries, as well as contributing to the efficiency thereof.

The first of these keys is called a primary key. This differs from the others in that it is present in all SQL tables, even without the user's specification. A primary key is simply a unique identifier for each row (object) in the table. Without users specifying a primary key, an arbitrary integer ID is assigned to each row and is largely inaccessible. Users can, however, specify any unique column of the table to be its primary key instead. An example of this would be if a table contained a list of people, their social security numbers could be a good choice for primary key, since most all other information may be duplicated among entries. These primary keys will be referenced at length, and are specified in Figure 7 as PK.

Foreign keys are the other main type of key used which will be discussed. The existence of foreign keys and their utility in queries was a deciding factor in the choice of SQL over Pandas. A foreign key is a designation given to a non-unique column of a table which points to a unique column of another table. This is also known as a many-to-one relationship between tables. Typically the unique column is the secondary table's primary key, though this is not a requirement. Foreign keys can be found in nearly all tables in Figure 7 and are labelled as FK.

Other types of keys don't have names such as foreign or primary key, but stand for the other types of relationships between database tables. The first is a many-to-many relationship, where many elements of a single table could reference many elements of another. For example, many customers could purchase a single type of product, and a single product could be purchased by many customers. The other type is a one-to-one relationship. This could arise when multiple tables share a common element which is unique in both tables. An example could be a list of students in a school, each having a unique ID, and a list of members of a class, each having a unique ID, which would correspond to the same ID in the school table. These last two types of keys are not currently utilized in the database structure, as seen in Figure 7.

The final reasoning behind the choice of SQL is related to the accessibility of the data. In anticipation of how many different groups will need access to the data, and to facilitate this access, a website was proposed. This website's goal would be to allow complex queries of CMB-S4's large volumes of data by any user regardless of level of knowledge of the database or of SQL queries as a whole. The website would take user input, possibly in the form of drop-down menus, though also with a space for more familiar users to enter their queries as a full string. With all selections made, the website would generate one or more plots for visualization of the returned observing statistics. This would allow fast and simple access tot he data for a much wider audience. Django, the website building platform chosen, also conveniently utilizes SQLite as the framework for its databases. More information on the website can be found in Section 5.

Raw data generated by CMB-S4 will be stored separately from all other data. From the numerous experiments which have come before it, methods are already in place to handle this volume of data. Since the raw data can never be retaken after its collection, care must and will be taken to ensure its longevity and accessibility. The raw data, however, is not all usable. When the raw data is collected, statistics are generated from it. These statistics range from weather to calibration data and are meant to form a picture of how and whether the detectors were working as intended during the observations. These statistics must be easily and quickly accessible. This is the goal of our project. Group members need a way to access and assess the observation quality statistics and through them, generate meaningful conclusions as to the status of detectors during observations. This requires all of that information (statistics themselves, telescope information, and observing schedule) to be readily available and able to be queried efficiently. The rest of this section will detail the methods and decisions implemented to further this goal.

4.1 Telescopes

In order to reference information relating to the focalplanes, multiple tables are used. All but the Band table contain a foreign key referencing at least one other table. The choice to use multiple tables was made in order to reduce the overall amount of duplicate data. If, for example, there were only one table, with a primary key of detector name, each row would need to contain all information relating to the detector itself, its place on the telescope (wafer, tube), its observing band, etc. Since there are only eight bands per LAT, thousands of detectors per wafer, etc., but over a hundred thousand detectors, any of these would be a massive quantity of repeated information taking up unnecessary room in the database. This structure allows the most amount of repeated information to be condensed. Though some of the information is still repeated, for example, each tube in the current focalplane model has the same platescale, which is calculated from the effective focal length and the diameter of the primary mirror, many of these repeated bits of information may not be present in tables containing real telescope data.

The data structure of the focalplanes logically follows the structure of the apparatuses, as well as the structure of the file which contains the information that contains the information from when the telescope was modelled. Telescopes are comprised of tubes. Each tube contains some number of wafers. Each wafer is comprised of detectors. These detectors observe at some frequency (Band). Information from some number of wafers gets passed along the same channel (Card). When mechanical failures occur, there are frequently multiple detectors affected. An easy way to track this is to have the information about shared qualities of detectors stored and easily referenced. For example, if many detectors suddenly are out of operational bounds, one could check if these have some feature in common like lying on the same wafer or sharing a card. The established data structure allows these comparisons to be done neatly and efficiently.

Generated telescope information began as a TOML file. This file type was used because it is both easy to import into Python dictionaries and it is able to be viewed and understood as is without being imported by the user first. This avoids the common pitfall of accessing an unfamiliar dictionary, where many queries must be made to understand the structure. This also allows the information to be separated easily into its constituent parts. During creation of the database, functions were created which could parse these focalplane files and write the contents to separate tables. This method is necessary for observation and statistic tables which are frequently altered, but not strictly necessary for the focalplane file. Since the current data is simply a model, a choice was made to create a foundation for the same sequence for the focalplane file as well.

4.2 Observations

The majority of the columns in the Observations table are necessary simply to categorize each observation; the rest of the columns are required to describe the state of the sky during each observing window. Details of starting and ending sun and moon locations in conjunction with starting and ending observing times describe regions of sky which must be avoided, lest damage to the telescope be done. Pass and sub-pass are two important rows which allow the statistics table to reference which observation they came from. During an observation in Chile, a telescope will scan a line across the sky, then move to a new line to start the next row. This, however, is made more difficult due to the Earth's rotation changing the starting coordinates for the next row during the observation of the previous one. The telescope will repeat this process either until the patch is no longer fully visible, or it has covered the entire region. This is one pass. The telescope will then either begin a new pass starting with the first row again, or else move on if the patch is no longer fully visible. Each time the patch is observed is counted as a pass, and each row of that patch observation is a sub-pass.

The table contains a many-to-many key within each row that references the telescope which has done the described observation. This is chosen instead of a foreign key (a many-to-one relationship) since there are many tubes contained in a single telescope and, thus, many rows in the tube table which share a value for the telescope column. This portion of the data structure will likely change once real observations are done. In current simulations, observations are broken up by individual tubes. While it would not be logical to perform an observation with only one tube of many in a telescope at a time, given they cannot be pointed in different directions at once, it becomes apparent there must be another reason for the discrepancy. That reason is simply that mock observation statistic generation is best done piecemeal, otherwise the virtual memory of the modeling computer will fill and the operation will be aborted prematurely. This is due to the large number of detectors in a single LAT. Without this constraint, real observations will be separated by telescope, not by tube, and the key will be changed to a foreign key to match.

South Pole observations differ in one main way from those done in Chile. At the South Pole, the sky turns while the telescope remains nearly stationary during each sub-pass, while in Chile, the telescope is constantly moving. This, however, largely does not affect the data framework. Each observation regardless of the sight can be stored in the same manner. Differences include, but are not limited to, lack of sun during any period of many observation at the South Pole. This will simply mean that some fields are null, where those same fields in a Chile observation will be much less of the time.

The Observations table will need to be updated regularly with each new observation made. This must be done during or prior to the observation, since the statistics generated from the observation must be made quickly after its conclusion, or possibly during the observation itself in order for transients to be detected and alerted in a timely enough manner. The framework that has been created will allow the table to be altered at will. This process may need further refinement once data has begun collection, but is efficient enough for current modelling.

4.3 Statistics

Simulations of observing statistics have been generated and are the basis of the current data structure. These statistics are vitally important for determining data quality. They will be an invaluable troubleshooting tool used to identify viability of components of observations to be used in the map-making step of the data pipeline. The statistics generated in current simulations, however, are not what will be generated during and following observations. As discussed in 3.2, calibrations will be done with the telescope routinely- some as often as at the start of or during each observation. These calibrations can provide information on weather conditions and be used to determine if any detectors are non-functional, among other things. One such example is called an elevation nod, or el-nod. This involves pointing the telescope at a section of sky, then tipping it in elevation. This should produce a change in unpolarized foreground noise, since differing quantities of atmosphere are observed through at different elevations. If foreground noise does not change as predicted, this is an indication of a problem with the telescope itself, which would require intervention to rectify, if possible.

Other statistics could and eventually will be fed into the table, such as weather information (wind speed, temperature, dew point, etc.). These are not directly related to the telescope itself, but affect the observations in profound ways. Lower observing frequencies are more susceptible to changes in atmospheric conditions and, thus, could be used in conjunction with these outside sources to determine validity of observing data from specific time periods.

As mentioned above, simulated observations have been done per telescope tube. This is reflected in the corresponding keys in the statistics table. These three entries, pass, sub-pass, and tube, allow a composite primary key consisting of all three, as well as a foreign key to connect these three fields to the observations table, and the tube and detector alone to reference their respective tables. These connections are crucial, since the statistics table itself contains no data about timing of observations or any more than the detector and tube names. Cross references between these three sets of tables will allow users to search for, for example, all detectors used in a specific observation and which are out of acceptable operating range.

Generation of observing statistics plots is a vital part of determining viability of data. A simple histogram showing the difference between the two el-nod observations could be enough to detect a faulty detector. Plotting weather information could, by itself, be cause for dismissal of data during a period of collection. Being able to quickly access and observe the data in this way is crucial for data quality management and will allow project members to clean the raw data of that which is not automatically filtered out from anticipated sources of error.

5 Future Outlook

CMB-S4's far reaching science cases mean that the data collected from the project must be widely accessible from a central location. Copies of the raw data will be stored separately. The other information will be backed up, but mainly accessible through, at present, NERSC. This computing cluster will host the database containing the described tables, and likely others as needed. The centralization of the database to the project's user base would allow it to be shared among the group easily. For outside persons who need access to the data, a different approach may be utilized.

The nature of any privately owned computing cluster comes with exclusivity, if not difficulty of access. To circumvent this problem, a website has started being developed. This website would allow any person to access the databases. Users would be unable to alter data, but would be allowed to query the data and generate plots natively. An avenue for exporting queried data and generated plots would be a core functionality of the site. This wide-spread access to the observing statistics and schedules could lead to the data making an even bigger impact in fields such as astronomy. A deep, wide observation such as will be conducted in the millimeter-wavelength range has never before been conducted.

A quick data-cleaning turnaround time is crucial for observation of transient phenomenon. These events frequently last less time than data cleaning itself has historically taken. In order to facilitate this goal, the processing of data- from its raw form through to (sub-optimally due to the time constraint) produced maps- must be efficient and largely scripted. The website which is proposed could aid in the process. By having the maps hosted on the website, an alert could be sent to specific sources when a phenomenon is detected in a produced map. An ambitious goal could be to automate even the detection of transients over time through AI training, since image recognition has been a central theme in AI software since its inception.

After consideration of numerous website-building methods, Django was chosen. Django is a Python module which has much of the fundamental coding involved in building a website from scratch premade. It also, conveniently, already utilizes SQLite as a basis for its databases. The main drawback to using Django is that it functions best when tables are created by/within it to be utilized by the website. The framework for utilizing legacy (preexisting, imported) databases exists, but is far from ideal. This has been the major point of difficulty in getting the website functional.

The current state of the website is that it is able to generate basic histograms only, but does not have a working statistics table to draw from. Since the original focalplane file used to create the framework for the database contained information on detectors which were not used in the statistics models, foreign keys from the statistics table point to nothing. The database is coded such that there must be a reference for foreign keys. This is to preempt any null pointers errors in the database. Logically, a statistic should never be added to the database if it did not come from a detector which is already accounted for in its own table. This means that the focalplane tables must be recreated. This issue led to a restructuring of the existing focalplane structure.

The original iteration of the focalplane tables had much of the data missing and nearly all detector information hard-coded in. This is terribly inefficient and would lead to problems as soon as any information needed to be updated, new focalplanes to be added, etc. This is exactly the issue which arose when new data was generated. Currently, the table structure shown in figure 7 is being programmed. Once this structure is finalized, the foreign keys from the statistics table will no longer be referencing nothing and the website development can continue past its current state.

6 Summary

CMB-S4's unprecedented scope allows for exciting new science wrapped in a neat package. However, issues arise from the handling of its massive volume of data and nearly constant scanning. Efficient management of data structures involved with the project will address many problems before their inception. Raw data will be evaluated immediately for viability using observation statistics generated during or immediately following observations. Usable data will then quickly be processed from its raw form into maps for transient phenomenon. The website currently under development will facilitate these points by making the statistics readily accessible and by handling general transient phenomenon alerts. Additional maps will be created more slowly for science cases which require more precision. In future iterations, we hope to bring the site up to full functionality, and to finish developing the pipelines which will be the backbone of real data collection once the project is operational.

References

[1] K. A. *et al.*, "Snowmass 2021 cmb-s4 white paper," 2022.

[2] Y. A. *et al.*, "Planck 2018 results," *Astronomy & Astrophysics*, vol. 641, p. A4, sep 2020.

[3] M. Kamionkowski and E. D. Kovetz, "The quest for b modes from inflationary gravitational waves," *Annual Review of Astronomy and Astrophysics*, vol. 54, pp. 227–269, sep 2016.

[4] W. Hu and M. White, "A CMB polarization primer," *New Astronomy*, vol. 2, pp. 323–344, oct 1997.

[5] C. D. *et al.*, "The physics of light relics," 2022.

[6] N. G. *et al.*, "The simons observatory: instrument overview," in *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy IX* (J. Zmuidzinas and J.-R. Gao, eds.), SPIE, jul 2018.

[7] S. R. D. *et al.*, "Cold optical design for the large aperture simons observatory telescope," 2019.

[8] P. A. R. A. *et al.*, "Bicep/keck XV: The bicep3 cosmic microwave background polarimeter and the first three-year data set," *The Astrophysical Journal*, vol. 927, p. 77, mar 2022.

[9] K. A. *et al.*, "Cmb-s4 science case, reference design, and project plan," 2019.