CMB ANALYSIS

J. RICHARD BOND AND ROBERT G. CRITTENDEN
Canadian Institute for Theoretical Astrophysics
Toronto, ON M5S 3H8, CANADA

1. Abstract

We describe the subject of Cosmic Microwave Background (CMB) analysis, its past, present and future. The theory of primary Gaussian anisotropies, those arising from linear physics operating in the early Universe, is in pretty good shape so the focus has shifted to the statistical pipeline which interfaces the data with the theory: mapping, comparing, cleaning, compressing, forecasting, estimating. There have been many algorithmic advances in the analysis pipeline in recent years, but more are needed for the forecasts of high precision cosmic parameter estimation to be realized. For secondary anisotropies, those arising once nonlinearity develops, the computational state of the art currently needs effort in all the areas: the Sunyaev-Zeldovich effect, inhomogeneous reionization, gravitational lensing, the Rees-Sciama effect, dusty galaxies. We use the Sunyaev-Zeldovich example to illustrate the issues. The direct interface with observations for these non-Gaussian signals is much more complex than for Gaussian primary anisotropies, and even moreso for the statistically inhomogeneous Galactic foregrounds. Because all signals are blended, the separation of components inevitably complicates primary CMB analyses as well. [1]

2. Introduction

What is CMB Analysis? The subject we call "CMB analysis" is a blend of basic theory, simulation and statistical data analysis. The goal of CMB analyzers is to extract from the data the physics of the various signals that make it up. CMB analyzers may therefore be theorists or experimentalists. Depending upon how advanced the state-of-the-art is, the relevant analysis subject of the day may lean more towards the theory/simulation side or

[1] To appear in Proceedings of NATO ASI: Structure formation in the Universe, eds. N. Turok, R. Crittenden.

more towards the statistical analysis (and Monte Carlo simulation) side. In the CMB field, we adopt a theorists's distinction between *primary*, *secondary* and *foreground* anisotropies. The primary ones are those we can calculate using linear perturbation theory (or, in the case of cosmic defects, with linear response theory). This covers the all important epoch of photon decoupling near redshift 1100, and even until the current time on very large scales, and to redshifts of a few on intermediate scales. Secondary anisotropies are those associated with nonlinear phenomena, either calculable via (weakly) nonlinear perturbation theory, semi-analytic methods or by more direct simulation of nonlinear patterns. Of course what we observe does not know of this distinction. Gravitational lensing effects on the CMB, quadratic nonlinearities and the kinetic Sunyaev-Zeldovich effect associated with Thomson scattering from flowing matter, reionization inhomogeneities, the thermal Sunyaev-Zeldovich effect associated with Compton upscattering from hot gas, the Rees-Sciama effect associated with nonlinear potential wells, all come under this heading. We also traditionally call emission by dust in high redshift galaxies a secondary process, and even emission from extragalactic radio sources. On top of this, there are various emissions from dust and gas in our Milky Way galaxy, the foregrounds — signals which are nuisances to the primary CMBologist but of passionate interest to interstellar medium astronomers. Fortunately, most of the secondary and foreground signals have very different dependencies on frequency (Fig. 1), and rather statistically distinct sky patterns (Fig. 2). Collateral information from non-CMB observations can also be brought to bear to unravel the various components.

Primary Theory and Current Data: We think we know how to calculate in exquisite detail the primary signal. The fluctuations are so small at the epoch of photon decoupling that linear perturbation theory is a superb approximation to the exact non-linear evolution equations. Intense theoretical work over three decades has put accurate calculations of this linear cosmological radiative transfer problem on a firm footing, and there is a speedy, publicly available and widely used code for evaluation of anisotropies in a wide variety of cosmological scenarios, "CMBfast" [1], further modified by many different groups to attack even more structure formation models.

The simplest (least baroque) versions of inflation theory predict that the fluctuations from the quantum noise that give rise to the structure we observe form a Gaussian random field. Linearity implies that this translates into temperature anisotropy patterns that are drawn from a Gaussian random process and which can be characterized solely by their power spectrum. The emphasis is therefore on confronting the theory with the data in the power spectrum space, as in Fig. 3. The panels show how $\mathcal{C}_\ell$ responds as various cosmological parameters are individually changed.

Until a few years ago, the emphasis was on restricted few-parameter sequences, but now we consider parameter spaces of much larger dimension, for forecasts of how well proposed experiments can do, and for the first round of data from a staged sequence of high-precision experiments covering large areas of sky: ground-based single dishes and interferometers; balloons of long duration (LDBs, $\sim$ 10 days) and eventually ultra-long duration (ULDBs, $\sim$ 100 days), as well as those that fly in a single night; NASA's MAP satellite in 2001 and ESA's Planck Surveyor in 2007. The lower right panel of Fig. 3 gives a forecast of how well we think that the two satellite experiments can do in determining C_ℓ if everything goes right. This rosy prognosis should be compared with the compressed bandpowers shown in the other panels estimated from all of the current data: DMR and some 19 other pre-conference "prior" experiments, TOCO [2], unveiled about the time of the conference, and three that appeared after the conference, the Boomerang-LDB [3], its North America test flight, Boomerang-NA [4], the Maxima-I [5] flight, and the first results from CBI, the Cosmic Background Imager [6]. Although new methods have been applied to these data sets, much was in place and the algorithms were well tested on previous data, in particular DMR, SK95, MSAM, QMAP, etc.

Although we sketch the current state of the art in CMB analysis in this paper, we note that this is a very fast moving subject now, with a handful of groups and a large number of researchers working on the broad range of issues, often as parts of the experimental teams associated with the increasingly complex experiments. A snapshot of the state two years ago and references to the relevant papers was given in [29], but much has occurred to improve things since then. The MAP plan for analysis is described in [34]. One situation that does have to be rectified in the future and should be borne in mind as one reads this article is the strong simplifying role played by assumptions of signal and noise Gaussianity in CMB analysis that cannot be correct in detail, and could lead to misleading results in practice.

Primary Parameter Sets for Gaussian Theories: A "minimal" inflation-motivated parameter set, e.g., $\{\omega_b, \omega_{cdm}, \omega_{hdm}, \omega_{er}, \Omega_k, \Omega_\Lambda, \tau_C, \sigma_8\, n_s, \tilde{r}_t, n_t\}$, involves about a dozen parameters. For many species we use $\omega_j \equiv \Omega_j \mathrm{h}^2$ rather than Ω_j since it directly gives the physical density of the particles rather than as a ratio to the critical density: thus, ω_b for the baryons, ω_{cdm} for the cold dark matter, ω_{hdm} for the hot dark matter (massive but light neutrinos), and ω_{er} for the relativistic particles present at decoupling (photons, very light neutrinos, and possibly weakly interacting products of late time particle decays). The non-relativistic matter density is $\omega_m \equiv \omega_{hdm} + \omega_{cdm} + \omega_b$.

The curvature energy relative to the closure density is $\Omega_k \equiv 1 - \Omega_{tot}$ and Ω_Λ is the energy relative to closure in the cosmological constant. Ω_Λ can also be interpreted as a vacuum energy, a sector which may be more complex and require more parameters. For example, a scalar field which dominates at late times, "quintessence", has potentials and other interactions to set. A simple phenomenology of quintessence has added to Ω_Q an average equation of state pressure-to-density ratio $w_Q \equiv \bar{p}_Q/\bar{\rho}_Q$.

τ_C is the Compton depth to Thompson scattering from now back through the period when the universe was reionized (by early starlight).

To characterize the initial fluctuations, amplitudes and tilts for the various fluctuation modes present are required. For adiabatic scalar perturbations, the amplitude is most appropriately the overall power in curvature fluctuations, but is often replaced, e.g., by σ_8^2, a density bandpower on the scale of clusters of galaxies, $8\,\mathrm{h}^{-1}$ Mpc. The scalar tilt is n_s. For tensor modes in the CMB, induced by gravity waves, the amplitude parameter is some measure of the ratio of gravity wave to scalar curvature power, $\tilde{r}_t$; usually a ratio of $\mathcal{C}_\ell$'s is used, e.g., at ℓ=2 or 10. Inflation theories often give relationships between the tensor tilt n_t and $\tilde{r}_t$ which can be used to simplify the space e.g., [7]. Two parameters for the amplitude and tilt of scalar isocurvature modes could be added to the mix. The characterization of the modes present in the early universe by 2 parameters each could be expanded to include parameters describing the change of the tilts with scale ($dn_s/d\ln k$, *etc.*), the change of that change, and so on: relatively full functional freedom in $n_s(k)$ is possible in inflation models, and substantial freedom also exists for $n_t(k)$.

Well and Poorly Determined Parameter Combinations: Comparison of the compressed bandpowers in Fig. 3 with the $\mathcal{C}_\ell$ sequences as $\omega_b, \Omega_{tot}, \Omega_\Lambda, \omega_{hdm}$ and n_s, are individually varied show the discriminatory power of the current data. Fixing the age at 13 Gyr as we do here in these sequences defines a specific relation between ω_k, ω_Λ and ω_m. From these figures, we may expect that the four can be determined reasonably well ($\omega_b, \Omega_{tot}, n_s$ and the overall amplitude). This is borne out in the full study with the real data, modulo parameter correlations which bring in some dependence on the other parameters into the best determined combinations (parameter eigenmodes [9, 10, 7]). For example, it is actually a combination of Ω_{tot} and Ω_Λ which is well determined, it just happens to be Ω_{tot}–dominated. There is another orthogonal combination which is very poorly determined, an example of a near-degeneracy among cosmological parameters no matter how good the CMB data gets e.g., [9]. Other near-degeneracy examples are less severe: e.g., $\Omega_{hdm}/(\Omega_{cdm} + \Omega_{hdm})$ is not well determined by current data, but could be by Planck.

Accompanying the idealized bandpower forecasts are forecasts for how well the cosmological parameters that define the functional dependence of C_ℓ in inflation models can be determined, after integrating (marginalizing) the likelihood functions over all other parameters. In one exercise that allowed a mix of nine cosmological parameters to characterize the space of inflation-based theories (the 11 above with ω_{er} fixed and n_t slaved to $\tilde{r}_t$), COBE was shown to determine one combination of them to better than 10% accuracy, LDBs and MAP could determine six, and Planck seven e.g., [10]. All currently planned LDBs could also get two combinations to 1% accuracy, MAP could get three, and Planck five. With the current data, one can forecast 4 to ± 0.1 accuracy, which is actually borne out in detailed computations with the data.

Two panels in Fig. 3 show what happens when two classes of isocurvature models are included, one a Gaussian isocurvature CDM model with the spectral index allowed to tilt arbitrarily, another a set of non-Gaussian cosmic defect models. Both apparently do not fare well with the current data. Primary anisotropies in defect theories are more complicated to calculate, because non-Gaussian patterns are created in the phase transitions which evolve in complex ways and for which large scale simulations are required. The non-Gaussianity means that the comparison with the data should be done more carefully, without the Gaussian assumption that goes into the bandpower estimations.

The Non-Gaussian Secondary and Foreground Signals: The secondary fluctuations involve nonlinear processes, and the full panoply of N-body and gas-dynamical cosmological simulation techniques are being brought to bear on the calculations.

The Santa Barbara test [14] compared simulations of an individual (final state) cluster of $10^{15}\, M_\odot$ done with a variety of (adiabatic) hydrodynamical and N-body techniques, in particular grid-based Eulerian methods and both grid-based and smooth particle (SPH) Lagrangian methods. (Eulerian grids are fixed in comoving space, Lagrangian grids adapt to the density of the flow.) Fig 4 shows this cluster, computed with the treePM-SPH code of Wadsley and Bond [16, 17], seen at redshift 0 in weak lensing, SZ and X-rays, at contour thresholds designed to represent potential observability. For a review of the great strides that have been made in SZ experimental methods and results on individual clusters up to 1998 see [15].

Another hydrodynamical example, the simulation of a supercluster [17, 18] using the treePM-SPH code, is used to illustrate the computational challenges once we extend beyond individual cluster simulations. Fig. 5 shows thermal SZ maps of the supercluster at redshift 0.5. The computation evolved, from redshift 30 to 0, a high resolution 104 Mpc diameter patch with $100^3/2$ gas and $100^3/2$ dark matter particles, surrounded by gas and dark particles with 8 times the mass to 166 Mpc, in turn surrounded by "tidal" particles with 64 times the mass to 266 Mpc. There were a total of 1.6 million particles including the medium and low resolution regions, but much larger simulations are possible in this era of massive parallelization.

The treePM technique used is a fast, flexible method for solving gravity and can accurately treat free boundary conditions. The SPH part of the code includes photoionization as well as shock heating and cooling with abundances in chemical (but not thermal) equilibrium, incorporating all radiative and collisional processes. The species considered were H, H^+, He, He^+, He^{++} and e^-. The extra cooling associate with heavy elements injected into the intracluster media during galactic evolution was ignored, but so was the feedback on the medium from energy injected from galactic supernovae: the computation does not have high enough spatial resolution to calculate galaxy collapse.

We now describe some aspects of simulation design for this supercluster simulation [17, 18]: One first decides on the mass resolution to achieve. This is set by the lattice spacing of particles on an initial high resolution grid, a_L, chosen to be ~ 1 Mpc comoving to ensure that the waves needed to treat the target objects forming in the simulation will be adequate, in this case clusters.[2] Next one needs to determine the spatial resolution, of the gas and of the gravitational forces, preferably highly linked. In Lagrangian codes like treePM-SPH, this varies considerably, being very high in cluster cores, moderate in filaments, and not that good in voids. Good calculations of the cluster cores were a target, so resolution better than ~ 40 kpc was desired. The best resolution obtained was about 15 kpc. Given that, the size of the high resolution part of the simulation volume is determined by CPU limitations on the number of particles that can be run in the desired time. This may mean the high resolution volume distorts considerably during the simulation. To combat this, layers were added of progressively lower resolution volumes, to ensure accurate large scale tides/shearing fields operate on the high resolution patch. For the calculations shown, the High Resolution region of interest (grid spacing a_L, 100^3 sphere) sits within a

[2]The mass resolution limits the high k power of the waves that can be laid down in the initial conditions (Nyquist frequency, π/a_L), but for aperiodic patches there is no constraint at the low k end: the FFT was used for high k, but a power law sampling for medium k, and a log k sampling for low k, the latter two done using SlowFT, i.e., a direct sum over optimally-sampled k values, with the shift from one type of sampling to another determined by which gives the minimum volume per mode in k-space. By contrast, more standard periodic big box calculations are limited by the fundamental mode. This simulated supercluster could not be done in a periodic code with this number of particles because of gross mishandling of the significant large scale tidal forces.

Medium Resolution region ($2\,a_L$, 80^3), in turn within a Low Resolution region ($2\,a_L$, 64^3). The mean external tide of the entire patch is linearly evolved and applied during the calculation.

It is clear from Fig. 3 that the dominant signals are quite patchy if one is interested in amplitudes above a few μK. This has the disadvantage for analysis that non-Gaussian aspects of the predicted patterns are fundamental, and an infinite hierarchy of connected N-point spectra beyond the 2-point spectrum C_ℓ are required to specify them (as for defect models). However, it also seems reasonable that we could represent the results in terms of extended emission from localized sources. This concentration of power is characteristic of many types of secondary signals. Indeed for some, such as radiation from dusty star-burst regions in galaxies, they can be treated as point sources, much smaller than the observational resolution of most CMB experiments. (However, highly accurate emission from dusty galaxies is too difficult to calculate from first principles, and statistical models of their distribution must be guided by observations.)

Extended and point source models for the temperature fluctuations express the signal as a projection along the line-of-sight of a 3D random field, $\mathcal{G}(\mathbf{r},t)$, which is the convolution of profiles $g(\mathbf{r}|\mathcal{C},t)$ surrounding objects of some class $\mathcal{C}$, with the comoving number density $n_{\mathcal{C}*}(\mathbf{r},t)$ defining a point process for those objects, e.g., [25] and [7] §5.1, 5.3. In [25, 26], the points were referred to as shots, from shot-noise, although of course continuous clustering of the shots as well as their uncorrelated Poissonian self-correlation has to be included. Simple derivations for the form of $\mathcal{G}(\mathbf{r},t)$ and g and the relation of the 2D power spectrum $\mathcal{C}_\ell$ to the 3D power spectrum of the $\mathcal{G}$ and hence of $n_{\mathcal{C}*}$ are given in [7].

For dusty star-burst regions, the shots are galaxies and the profile g involves the dust density and temperature [25, 26], while for the SZ effect, the dominant emission comes from clusters and groups and the profile g is a constant times the pressure [25, 7, 27].

Various attempts have been made to use hydrodynamical codes to do the an entire region from redshift 0 to 2, say. However, it should be clear from the discussion of numerical setup that approximations are always needed to do this. One approach was to calculate the full pressure power spectrum from a single simulation, $\langle|\widetilde{\mathcal{G}}(k,t)|^2\rangle$, and appropriately project it to determine the SZ $\mathcal{C}_\ell$ [19, 20]. A more ambitious approach was to tile the space with boxes out to some redshift [21, 22, 23, 24], but because of computer limitations and resolution requirements, the boxes were relatively small and computationally expensive, so tricks were needed, e.g., rotating the few simulated boxes seen at different redshifts to make the predictions.

The shot model does give a way to go forward semi-analytically, by laying down pressure profiles around a catalogue of halos. The halos and their properties could be computed with N-body-only simulations of much larger boxes than those requiring hydro. Another approach is to use "peak-patches", a method for identifying halos in the initial conditions of calculations which has been shown to be accurate for clusters [27]. For this approach, the entire space can be tiled with contiguous boxes that are smoothly joined, and have all the required long wavelength power to treat the really rare halo concentrations, superduperclusters, that can appear.

What to do for the pressure profiles $p(\mathbf{r}|\mathcal{C}, t)$ (g up to a constant) is of course debatable. One strategy is to use profiles calculated from hydro simulations, but what has invariably been done is even simpler: pressure profiles scaled by bulk halo properties of the observed form derived from low redshift X-ray cluster data (and extrapolated to the less well known higher redshift regime). For example, adopting a spherically symmetric isothermal "beta" profile is common: $p(\mathbf{r}|\mathcal{C}, t) = p_c(1 + r^2/r^2_{core*})^{-3\beta/2}\vartheta(R_{\mathcal{C}*} - r)$, where p_c is the central pressure, r_{core*} is a core radius, and $R_{\mathcal{C}*}$ is an outer truncation radius exterior to which the local pressure contribution is taken to be zero, and interior to which ϑ is unity. The radius r here is comoving and the $*$ subscript denotes comoving radii (with the expansion factor $a(t)$ taken out of the physical radii). $\beta \approx 2/3$ is a reasonable fit to the X-ray data.

This rapid semi-analytic peak-patch method for simulating clusters using isothermal beta-profiles was used in the construction of the two SZ layers (thermal and kinetic) of the sandwich in Fig. 2. The lower panels of Fig 4 show peak-patch-derived SZ maps for 3 cosmologies, and the lower corner shows the distribution of $\Delta T/T$ in pixels, with its non-Gaussian distinct tails. The right two panels contrast the contribution for the ΛCDM example of clusters below redshift 1/2 with those above. The top panel of Fig. 5 shows the ΛCDM example with a filter scale appropriate for the Planck satellite resolution and with an ideal resolution for the cluster system, below two arcminutes, which interferometers and large single dishes can achieve. Interferometers typically filter the low ℓ, losing large scale power, though. The lower panel of Fig. 3 shows the SZ $\mathcal{C}_\ell$ spectrum derived for the flat ΛCDM simulation shown, contrasting the contributions from clusters and groups above redshift 1/2 with all of them. Both Poissonian and clustering contributions are included, since the simulation has them.

Because the power shown is not distributed in a democratic fashion as it is for Gaussian theories, but is concentrated in cold spots below 218 GHz, in hot spots above, the naive visual comparison with bandpowers derived using a Gaussian assumption about the distribution of power is misleading.

The assumption here that the SZ effect is dominated by the high pressure regions associated with clusters and groups of galaxies, with only weak modifications coming from the intercluster medium, is borne out visually in the supercluster simulation: the emission from the gas in filaments outside of the groups and poor clusters that reside in them is weak, below observable levels. What we can also see, however, is that an anisotropic pressure distribution is expected, with elongation along the filament directions, inviting improvements over spherical approximations.

To properly analyze the SZ effect in ambient fields, having fast Monte Carlo simulation methods along the lines developed here is clearly essential. That does not take away from the challenge, as yet unmet, of how best to do the analysis of such distinctly non-Gaussian signals, which consist of extended rather than point sources, and which, with even relatively small beams, will be somewhat confused by overlapping sources. At least with the cluster system, we have hope that we can model the individual elements with hydrodynamical simulations. For other secondary sources such as dusty galaxies, a priori theoretical models are not really feasible, and the properties of the sources will have to be derived from the observations. At least in those cases, for typical beam sizes, the emission can be treated as from point sources.

The *foreground* signals from the interstellar medium are also non-Gaussian, but are not well modelled by even the extended source model, since their emission power spectrum $\mathcal{C}_\ell$ has been shown to rise to lower ℓ; further, no simplifications such as statistical isotropy apply. Exactly what to do for the analysis of foregrounds is under intense study. Using other data sets will clearly play a strong role in the final analysis however, e.g., templates for the distribution of shorter wavelength dust radiation from IRAS+DIRBE data.

3. Acting on the Data

The data comes in as raw timestreams, which must be cleaned of cosmic rays, obvious sectors of bad data, etc. Encoded in it are the sky signals, the noise, unwanted instrument, atmospheric and other residual signals. The pointing matrix from time bits to a chosen pixel space allows the spatial signal to be separated from the noise; a byproduct is a map which completely represents the data if the pixelization is fine enough relative to the experiment's beam. It used to be that the cosmology was drawn from the maps, perhaps using new bases beyond the pixel one, possibly truncated to *compress* the data into more manageable chunks (e.g., using signal-to-noise eigenmodes to get rid of informationless noisy components of the map, e.g., [8] and below). The statistical distribution of operators on the pixelized data can be estimated from the map, in particular the power spectrum $\mathcal{C}_\ell$, which represents a *radical compression* of the data, but of course is no longer a complete representation of the map. However, under the assumption that

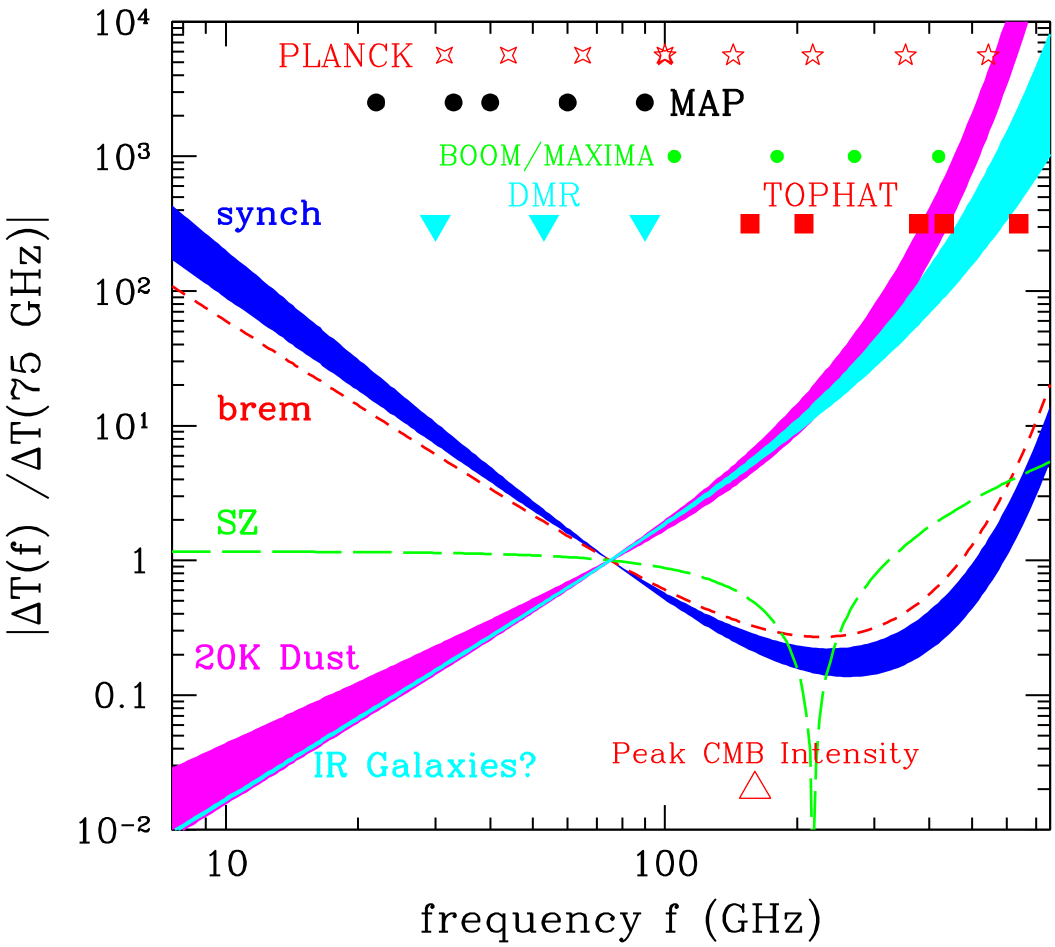

Figure 1. Foregrounds and secondary anisotropies depend differently on photon frequency, the key property for separating the components. Plotted is the frequency dependence of the effective thermodynamic temperature fluctuations, normalized to their values at 75 GHz. Thus the primary CMB fluctuations correspond to a horizontal line in this figure, synchrotron and bremsstrahlung dominate at low frequencies, dust at high. The bands represent a measure of our uncertainty in the appropriate foreground shapes. The highly distinctive shape for the Sunyaev-Zeldovich (SZ) effect is negative at low frequencies, positive at high. The actual level of contamination depends on the angular scale. The foreground emission is minimal around 90 GHz, not far from where the CMB intensity peaks. Detector frequencies for some notable experiments are denoted by the symbols at the top. Currently, detectors below 100 GHz are HEMTs, above are bolometers. Note the wide coverage planned for the Planck satellite which uses both.

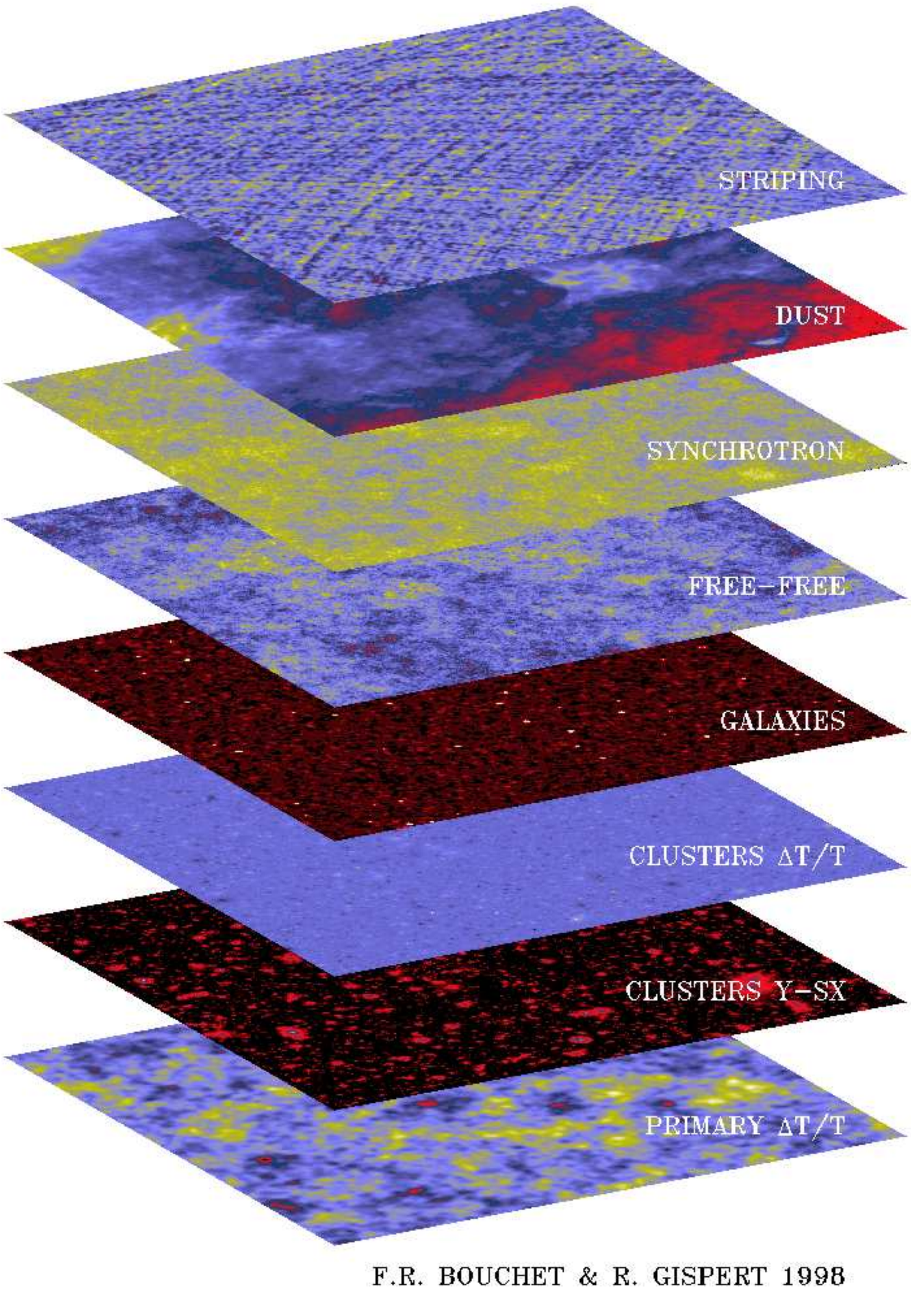

Figure 2. A schematic view by Bouchet and Gispert of many of the possible microwave foregrounds that need to be separated from the primary CMB anisotropy pattern shown in the 10° map at the bottom. These include noise effects (striping), galactic foregrounds (dust, synchrotron and bremsstrahlung or free-free emission) as well as secondary anisotropies (extragalactic radio and infrared galaxies, CMB upscattering by hot gas in clusters, "clusters Y-SX" and kinetic SZ effect (the Doppler effect arising from moving clusters), "clusters $\Delta T/T$"). Each of these has a unique temperature pattern on the sky, and all but the kinetic SZ effect have a different spectral signature (Fig. 1).

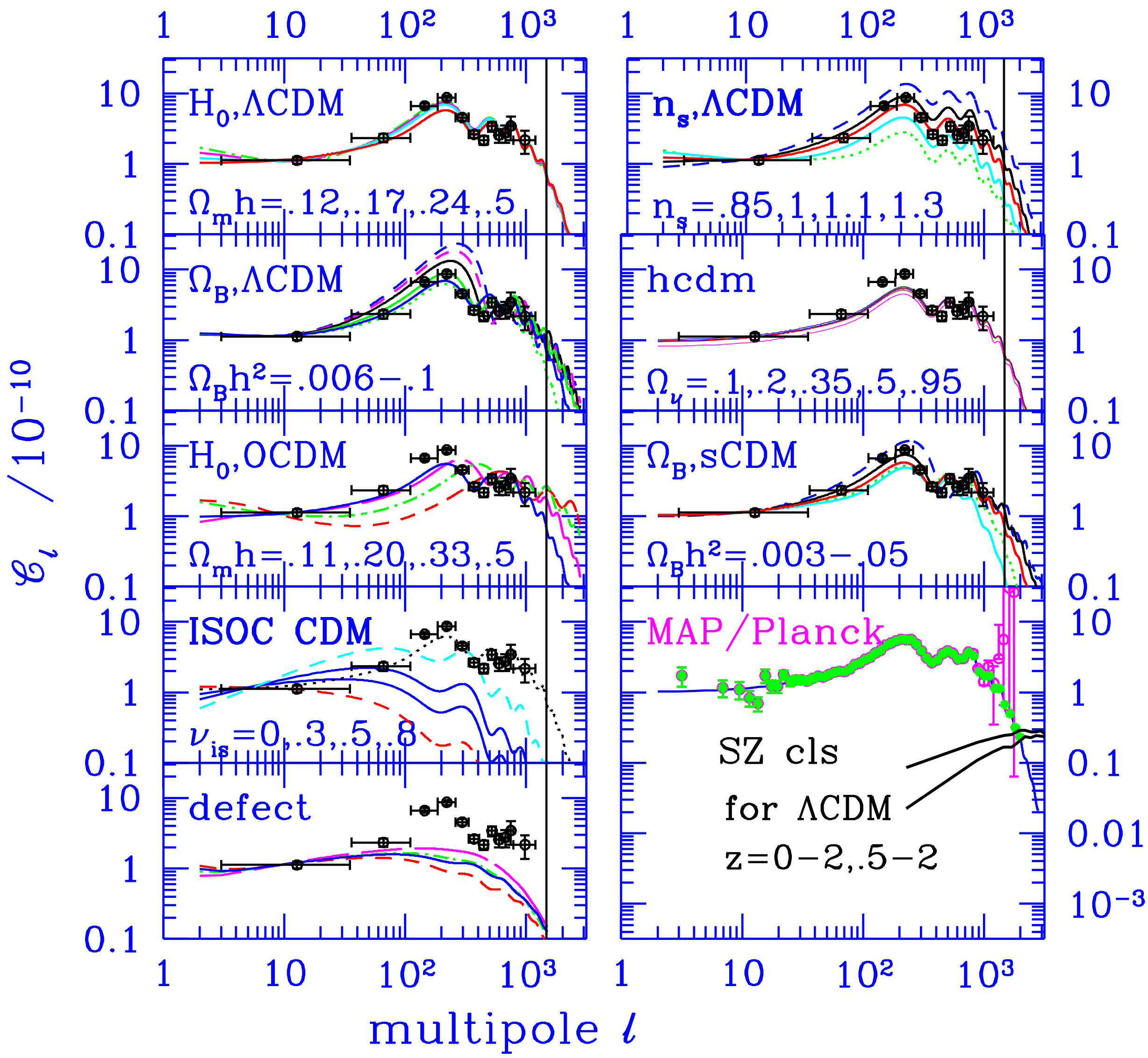

Figure 3. The $\mathcal{C}_\ell$ anisotropy bandpower data, including Boomerang, Maxima-1 and CBI, compressed to 13 bands using the methods of [11] are compared with various $\mathcal{C}_\ell$ model sequences, each for universes with age 13 Gyr (left to right): (1) untilted flat ΛCDM sequence with H_0 varying from 50 to 90, Ω_Λ from 0 to 0.87; (2) n_s varying from 0.85 to 1.25 for an $H_0 = 70$ (Ω_Λ=.66) ΛCDM model – dotted is 0.85 with gravity waves, next without, upper dashed is 1.25, showing visually why n_s is found to be nearly unity; (3) $\Omega_B \mathrm{h}^2$ varying from 0.0063 to 0.10 for the H_0=70 Ω_Λ=.66 ΛCDM model; (4) neutrino fractions Ω_{hdm}/Ω_m varying from 0.1 to 0.95 for an H_0=50 Ω_Λ=0 sequence with Ω_m=1; (5) H_0 from 50 to 65, Ω_k from 0 to 0.84 for the untilted oCDM sequence, showing the strong ℓ-shift of the acoustic peaks with Ω_k; (6) $\Omega_B \mathrm{h}^2$ varying from 0.003 to 0.05 for the $H_0 = 50$ Ω_Λ=0 sCDM model; (7) an isocurvature CDM sequence with positive isocurvature tilts ranging from 0 to 0.8; (9) sample defect $\mathcal{C}_\ell$'s for textures, etc. from [12] – cosmic string $\mathcal{C}_\ell$'s from [13] are similar and also do not fare well compared with the current data. The bottom right panel is extended to low values to show the magnitude of secondary fluctuations from the thermal SZ effect for the ΛCDM model. The kinematic SZ $\mathcal{C}_\ell$ is significantly lower. Dusty emission from early galaxies may lead to high signals, but the power is concentrated at higher ℓ, with a weak tail because galaxies are correlated extending into the $\ell \lesssim 2000$ regime. Forecasts of how accurate $\mathcal{C}_\ell$ will be determined for an sCDM model from MAP (error bars growing above $\ell \sim 700$) and Planck (small errors down the $\mathcal{C}_\ell$ damping tail) are also shown.

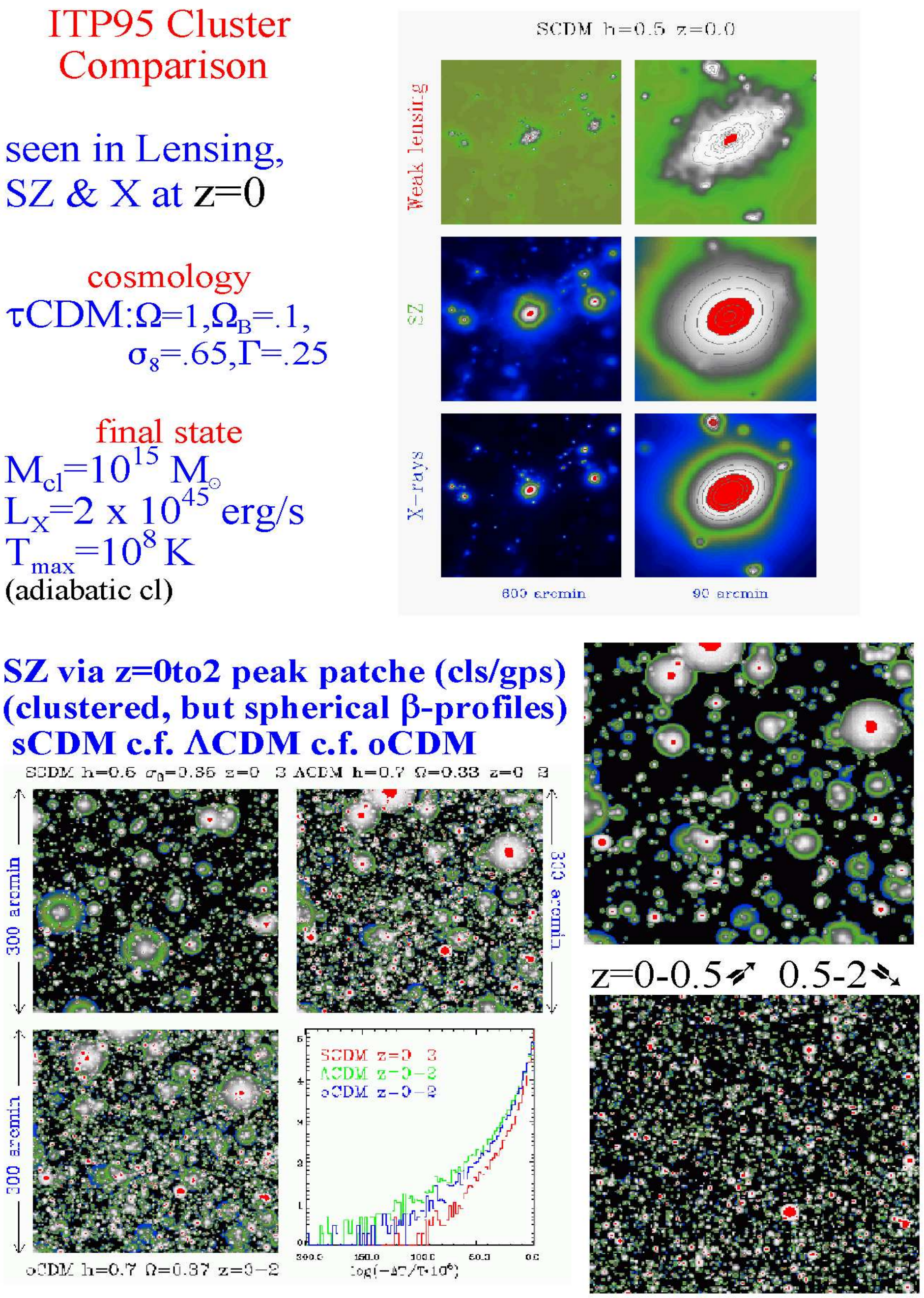

Figure 4. SZ maps for the ITP cluster comparison seen at redshift 0 subtending the angles shown. The observing wavelength was taken to be in the Rayleigh-Jeans region of the spectrum (so $\Delta T/T = -2y$ here). The core dark regions interior to the white areas are above 32×10^{-6}; the dark contours surrounding the white are at 2×10^{-6}, levels now accessible to ground based instrumentation. Blank field SZ surveys using interferometers and bolometer experiments promise to revolutionize our approach to the cluster system, especially at $z \gtrsim 0.5$. The bottom figures shows SZ maps for peak-patches for 3 cosmological models. The observing wavelength was taken to be in the Rayleigh-Jeans region of the spectrum (so $\Delta T/T = -2y$ here). The core dark regions interior to the white areas are above 32×10^{-6}, and the the dark contours surrounding the white are at 2×10^{-6}. These levels are now accessible to ground based instrumentation. Blank field SZ surveys using interferometers and bolometer experiments promise to revolutionize our approach to the cluster system, especially at $z \gtrsim 0.5$. Although the contours do go into the far field of the cluster region because they probe the electron pressure, one needs to go below 10^{-6} to fully detect the filaments. The effect is more pronounced in ΛCDM and oCDM universes than in the sCDM model.

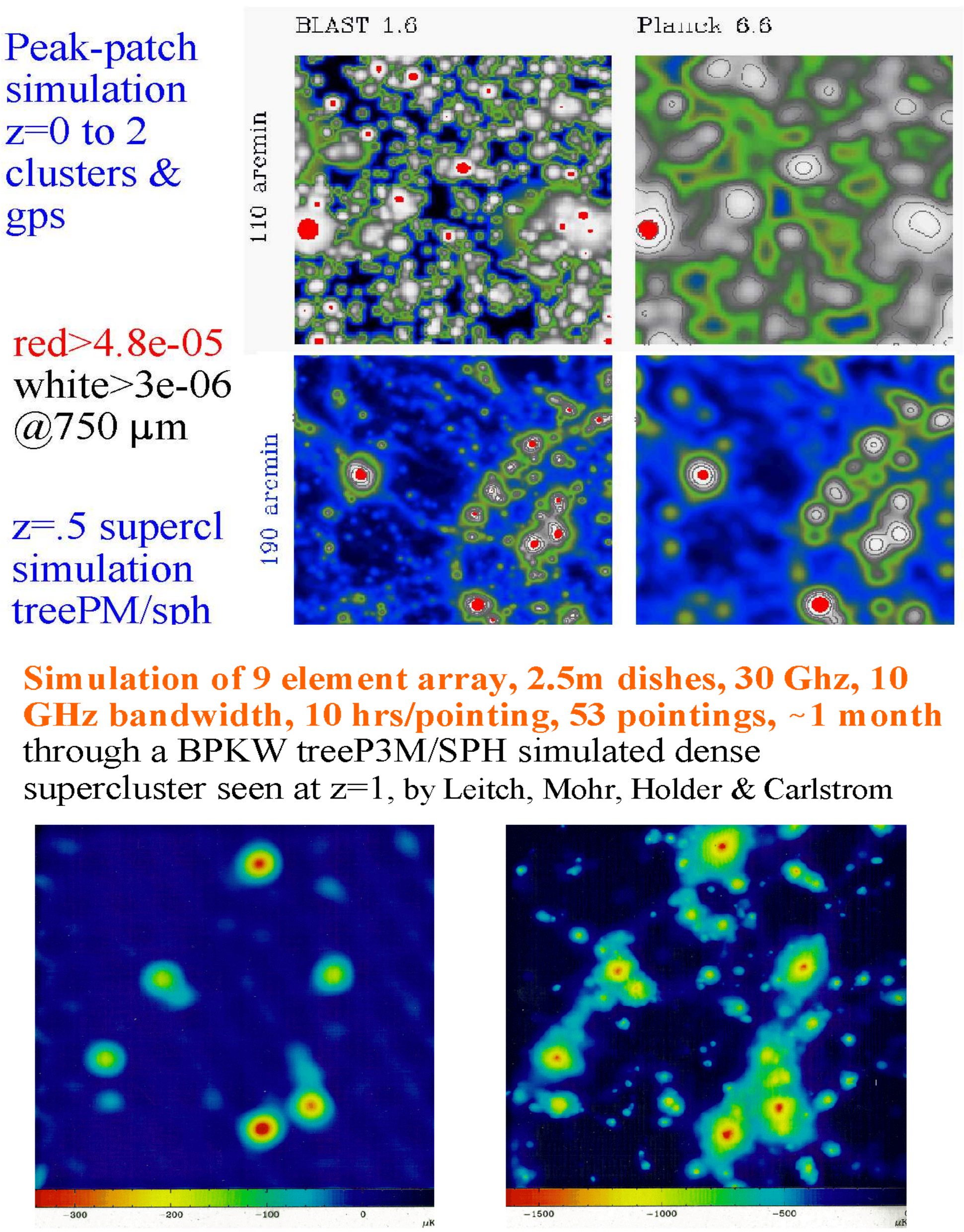

Figure 5. SZ map for peak patches (above) and a supercluster region seen at redshifts 1, 0.5 and 0.2, subtending the angles shown. The core dark regions interior to the white areas are above 48×10^{-6}, and the the dark contours surrounding the white are at 3×10^{-6} at $750\mu m$, levels that can be achieved with blank field SZ bolometer experiments. Filamentary bridges between the clusters are typically below 10^{-6}, but the far field of clusters can be probed. The bottom left panel passes the supercluster region at redshifts 1 in the right panel through a simulated 30 GHz HEMT-based interferometer with the characteristics indicated.

the sky maps only contain Gaussian signals characterized by an isotropic power spectrum, and with fine enough binning of the bandpowers, this is a complete representation of the statistical information in the map. For such Gaussian theories (e.g., inflation-motivated ones), cosmological parameters can be quickly estimated directly from the power spectrum. The radical compression has allowed millions of theoretical models to be confronted with the data in the large cosmological parameter spaces described above, e.g., [28], a situation infeasible if full statistical confrontation of all the models with the maps was required. Basic aspects of the Boomerang and Maxima pipelines are described in [29], [31] and [32]. Pipelines were also developed for DMR and for QMAP. The MAP pipeline as currently envisaged is described in [34]. The Planck pipeline is still very much under discussion and development, complicated by the necessity of delivering high precision component maps and derived parameters from such a huge volume of data.

Bayesian Chains: The data analysis pipeline can be viewed as a Bayesian likelihood chain:

$$\begin{aligned}
&\mathcal{P}(parameters|data,\ theoretical\ framework)\\
&= \mathcal{P}(data\ timestreams|maps,\ noise)\\
&\ \otimes\mathcal{P}(noise|noise - obs)\\
&\ \ \otimes\mathcal{P}(maps|signals)\\
&\ \ \ \otimes\mathcal{P}(signals|parameters)\\
&\ \ \ \ \otimes\mathcal{P}(parameters|prior\ knowledge)\\
&\ \ \ \ \ \otimes[1/\mathcal{P}(data|theoretical\ framework)]
\end{aligned} \tag{1}$$

Each conditional probability in the unraveling of the chain is very complex in full generality. This means the exact statistical problem is probably not solvable. Simplifying approximations of sections of the chain have made it tractable. Of particular widespread use has been the assumption that both noise and signals are Gaussian-distributed, and others are being explored, especially by Monte Carlo means, but much remains to be done for us to attain the high precision levels that have been forecast.

We now describe the various terms. We begin with the last two. The priors $\mathcal{P}(parameters|prior\ knowledge)$ for the cosmological parameters are often taken to be uniform, but can represent information from other experiments, e.g., from LSS observations, from prior CMB experiments, from age or H_0 constraints. Without prior information, the hope is that the updating information from the rest of the probability chain will be so concentrated that the precise nature of the prior choice will not matter very much. As we have seen, these non-CMB experiments are crucial, even given the most ideal CMB experiment, in order to break near-degeneracies among cosmological parameters.

The quantity $\mathcal{P}(data|theoretical\ framework)$ is an overall normalization. It can address whether the overall theory is crazy as an explanation of the data. It has some characteristics in common with goodness-of-fit concepts in "frequentist" probability analysis.

We often refer to entropies $\mathcal{S} \equiv \ln \mathcal{P}$ in the Gibbs sense so the Bayesian chain involves a sum of the individual conditional entropies that make up the whole.

Signal/noise separation and $\mathcal{P}(d|\Delta, \eta, P_{ctp})\mathcal{P}(\eta|w)\mathcal{P}(\Delta)$: The information comes in the form of a timestream d_{ct} for bit of time t in channel c, as well as a pointing vector on the sky, $\hat{q}_c(t)$. The sky signals are functions of $\hat{q}$ and frequency. Once the sky is gridded into pixels, these define signal maps. We denote the sum of these by Δ_{cp}, for pixel p and frequency channel c. The pointing matrix, P_{ctp}, is a operator mapping timebit t to pixel p with some weight. The difference $d - P\Delta$ is the noise, η, plus further residuals $\mathcal{R}$ (which may be on the sky or be systematic experimental signals). The probability is

$$\mathcal{P}(d|\Delta, \eta, P_{ctp}) = \prod_{ct} \delta(d_{ct} - (\sum_p P_{ctp}\Delta_{cp} + \mathcal{R}_{ct} + \eta_{ct})). \tag{2}$$

Map-making for a given channel involves using just the information encoded in d and P to separate what is sky signal (and residuals) from what is noise, i.e., construction of $\mathcal{P}(\Delta|d)$.

The pointing matrix is an onto projection operator, with many more time bits than pixels. The simplest example for P_{ctp} is the characteristic set function for the pixel $\chi_p(\hat{q}_c(t))$ (one inside the pixel, zero outside). For a chopping experiment such as MAP, the two pixels where the two beams are pointing (separated by $141°$) are coupled at a given time bit. There are also choices to be made about scattering the information in the timestream among pixels. For example, one can include the beam in Δ or in P. At the moment the former is preferred, but there are good arguments for the latter if the beam changes in time (e.g., asymmetric beams with rotations). In that case, $\chi_p(\hat{q}_c(t))$ is replaced by $A_p\mathcal{B}(\hat{q}_p - \hat{q}_c(t))$, where $\mathcal{B}$ is the beam function and A_p is the pixel area, and the derived signal Δ_{cp} does not have the beam function convolved with it as it does in the "top hat" χ_p case. Of course, any other function normalized to the pixel area upon integration can be used, e.g., approximations to the beam which might not contain all sidelobe information about it. As long as the signals are appropriately convolved, and the pixelization choices are fine enough, the results should be identical. Pointing uncertainty further complicates the simplicity of P.

In map-making, there is an implicit assumption on $\mathcal{P}(\Delta|prior)$, the prior probability of the map, namely that it is uniform: any value is a priori possible. The map distribution should involve integration over all possible noise distributions; in practice, only the measured noise is used, which delivers the maximum likelihood map $\overline{\Delta}_{cp}$ and a pixel noise correlation $C_{N,cpc'p'}$, allowing for possible channel-channel correlations in the noise.

For many experiments in the past, one could just compute the average $\overline{\Delta}$ and variance C_N of the measured d's contribution to each pixel, relying on the central limit theorem to ensure a Gaussian-distribution. However, this also relied on an implicit assumption, namely that η was uncorrelated from time bit to time bit, which is not correct with the $1/f$ noise in detectors. It was also often the case that the noise was so strong compared with the signal that the latter could be ignored in solving for Δ_{cp}; this is no longer the case.

So far it has been essentially universal to assume that η is Gaussian-distributed stationary noise with noise power w_ω^{-1} a smooth function. That is, the noise is completely specified by its covariance $c_{n,ct,c't'} \equiv \mathcal{S}[\langle \eta_{ct}\eta^\dagger_{c't'}\rangle]$, which is assumed to be a function of $t-t'$, hence diagonal in the time frequencies ω. $\mathcal{S}$ denotes the smoothing operation. Drifts and non-Gaussian elements can be included with timestream filtering and in the catch-all residual $\mathcal{R}$ using time templates (see below). Ignoring $\mathcal{R}$ for the moment, the probability of the time-ordered data given a map and noise-power w, $\mathcal{P}(d|\Delta, w)$, is just a Gaussian:

$$
\begin{aligned}
\mathcal{P}(d|\Delta,w) &= \exp[-(d-P\Delta)^\dagger w(d-P\Delta)/2]/\{(2\pi)^{N_{tbits}/2}\sqrt{\det(w^{-1})}\} \\
&= \exp[-(\overline{\Delta}-\Delta)^\dagger W_N(\overline{\Delta}-\Delta)/2]/\{(2\pi)^{N_{pix}/2}\sqrt{\det(W_N^{-1})}\} \\
&\qquad \otimes \exp[-\bar{\eta}^\dagger \widetilde{w}\bar{\eta}/2]/\{(2\pi)^{(N_{tbits}-N_{pix})/2}\sqrt{\det(\widetilde{w}^{-1})}\}, \qquad (3)\\
W_N\overline{\Delta} &= P^\dagger w d,\ W_N \equiv P^\dagger w P \equiv C_N^{-1} = \langle n^\dagger n\rangle^{-1}, \qquad (4)\\
n &\equiv \overline{\Delta}-\Delta = C_N P^\dagger w\eta, \quad P^\dagger w\bar{\eta}=0,\ \bar{\eta}\equiv\langle\eta|d\rangle = d-P\overline{\Delta}, \\
\widetilde{w} &= w(1-P(P^\dagger wP)^{-1}P^\dagger w)\,. \qquad (5)
\end{aligned}
$$

The maximum likelihood solution for w_ω is $(\mathcal{S}[|\mathrm{FT}(d-P\overline{\Delta})|^2])^{-1}$, where FT denotes the Fourier transform operator. Note that the $\bar{\eta}$ noise term is multiplied by a projector which is perpendicular to the spatial sector; i.e., $C_N P^\dagger w\bar{\eta}$=0. The first Gaussian in eq.(3) is $\mathcal{P}(\overline{\Delta}|\Delta)$. We can interpret it as the integration over the spatial Gaussian random noise field n defined in eq.(4) of the product of $\mathcal{P}(\overline{\Delta}|\Delta,n) = \delta(\overline{\Delta}-(\Delta+n))$ and the distribution of n, $\exp[-n^\dagger C_N^{-1}n/2]/[(2\pi)^{N_{pix}/2}[\det C_N]^{1/2}]$.

Even with stationarity, the convolution of w with d appears to be an $O(N_{tbit}^2)$ operation, but it is really $O(N_{tbit})$ because $w(t-t')$ generally goes nearly to zero for $t - t' \gg 0$. Similarly, the multiplication of the pointing matrix is also $O(N_{tbit})$ because of its sparseness. Thus, we can reduce the timestream data to an estimate of the map and its weight matrix in only $O(N_{pix}^2)$ operations.

What has proved quite effective to solve this is an iterative method described in [31]: the map on iteration $j+1$ is estimated from $W_*\Delta_{(j+1)} = W_*\Delta_{(j)} + P^\dagger w_{(j)}\eta_{(j)}$ where the noise and noise power on iteration j are determined from $\eta_{(j)} = d - P\Delta_{(j)}$, $w_{(j)}^{-1}(\omega) = \mathcal{S}[|FT(\eta_{(j)})|^2(\omega)]$. Here $W_* \equiv P^\dagger w_* P$ is a matrix chosen to be easy to invert: e.g., w_* being constant was used for Boomerang [31]. As the maps converge, the noise orthogonality condition holds, so the solution is the correct one. What $w(\omega)$ looks like is white noise at high temporal ω, crashing to zero because of $1/f$ noise in the instruments, and possibly because of time filtering that has been imposed on the timestream. These low frequencies correspond to large spatial scales, and the iteration converges slowly there, inviting multigrid speedup of the algorithm, which is now being implemented [33]. The final $\overline{\Delta}$ and C_N are computed using d and the converged $w_{(j)}$ in eq.(4). In some instances, we can work directly with $W_N\overline{\Delta}$ and W_N rather than $\overline{\Delta}$ and C_N, avoiding the costly $O(N_{pix}^3)$ inversion. If sections of the map are cut out after W_N determination, through a projector χ_{cut} which is zero outside and one inside the cut, the cut weight matrix is $(\chi_{cut}W_N^{-1}\chi_{cut})^{-1}$. This corresponds to integration over all possible values of the now unobserved $\overline{\Delta}$'s outside of the cut, which increases the errors on the regions inside because of the correlation. Unfortunately, two matrix inversions are required, the first potentially quite large even if the cut is severe.

MAP is a difference experiment. Differencing is a form of filtering. To a good approximation the MAP noise is white, so $w_{(j)} = w_*$ can drop out, and the critical element is computation of $P^\dagger P$. Hinshaw *et al.* [34] describe the current pipeline for MAP. They choose W_* to be $diag(P^\dagger w_* P)$, but otherwise use this same iterative algorithm.

Since P is all that differentiates signal from noise in the timestream, it is essential that it be sufficiently complex; in practice this means many cross linkages among pixels in the scan strategy. This is because there are random long term drifts in the instruments that make it hard to measure the absolute value of temperature on a pixel, though temperature differences along the path of the beam can be measured quite well because the drifts are small on short time scales. If there are not sufficient cross-links, the maps are *striped* along the directions of the scan pattern, and the weight matrix W_N has to be called upon to demonstrate that these features are not part of the true sky signal we are searching for. The rapid on-board differencing for MAP effectively eliminates this.

Pixelization: Pixels are any discrete basis that gives a complete representation of the data. Square top-hat functions tiling the space in a grid quite a bit smaller than the beam size were the norm in the past. In some cases generalized pixels were used, involving direct projection of the data onto spatially extended mode functions (e.g., SK95). As the size of the data increased, more elaborate hierarchical schemes were needed. For COBE, the Quadrilateralized Spherical Cube was used: the sky was broken into six base pixels corresponding to faces of a cube. Higher resolution pixels were created by dividing each pixel into four smaller pixels of approximately equal area, in a tree with pixels that are physically close having their data stored close to each other. There are (quite) small errors associated with the projections of the sphere onto the faces. At resolution r, there are $6 \times 2^{2(r-1)}$ pixels. The COBE data came with resolution $r = 6$, 6144 pixels of size $2.6°$, and we sometimes use resolution 5, still safely smaller than the $7°$ beam.

For large sky coverage, it is beneficial to have a pixelization which is azimuthal, with many pixels sharing a common latitude, to facilitate fast spherical harmonic transforms [35] between the pixel space, where the data and inverse noise matrix are simply defined, and multipole space, where the theories are often simple to describe. The FSHT operation count is a $\mathcal{O}(N_{pix}^{3/2})$ operations, with $\mathcal{O}(N_{pix} \ln N_{pix}^2)$ potentially possible to achieve. HEALPix[36] is an example which has been adopted for Boomerang, MAP and Planck: it has a rhombic dodecahedron as its fundamental base, which can be divided hierarchically while remaining azimuthal. There is also an extensive software package that accompanies it to allow manipulations. See also [37] for the alternative hierarchical "igloo" choice, with many HEALPix features, and some other advantages.

Templates in Time and Space: Templates are specific temporal patterns that we probe for in the timestream, $\upsilon_{ct,A}$. Templates could have frequency structure (c) as well. They are often associated with specific spatial patterns $\Upsilon_{cp,A}$ that we can probe for in the map or directly in the timestream with $\upsilon_{ct,A}=\sum_p P_{ctp} \Upsilon_{cp,A}$. The unknowns are the amplitudes κ_A (and the errors on them), and possibly other parameters characterizing their position, orientation and scale. The residual signal associated with the templates is therefore $\mathcal{R}_{ct} = \sum_A \upsilon_{ct,A} \kappa_A$, a concept of wide applicability. The templates can model contamination by radiation from the ground, balloon or sun, components of atmospheric fluctuations and cosmic ray hits. Often they are synchronous with a periodic motion of the instrument. In maps, they have been used in the removal of offsets, gradients, dipoles, quadrupoles etc. and looking for point sources and extended sources (e.g., IRAS/DIRBE structures). If we are removing a pattern, we can do this by directly projecting it from the map, and modifying the map statistics, as described in the next subsection. Note that the map itself defines spatial templates, the individual pixels.

Since $\upsilon_{ct,A}$ is of the same mathematical form as P_{ctp} we can solve simultaneously for Δ_{cp} and κ_A, delivering an extended map $\overline{\Delta}, \bar{\kappa}$, with appropriate error covariances for each, and including a cross term. The assumption is that the *a priori* probability for the template amplitudes κ is uniform, as for the map amplitudes. One could also marginalize the unobserved κ_A, i.e., integrate over them, which leads to a projected form of w. A nice way to deal with this is to assume that the prior probability for κ is Gaussian with zero mean and correlation $K_{AA'} = \langle \kappa_A \kappa_{A'} \rangle$. Letting the dispersion become infinite reduces to the projection.

Marginalization yields modified noise weights and a modified average map:

$$\begin{aligned} &\widetilde{w} \equiv w - w\upsilon(\upsilon^\dagger w \upsilon + K^{-1})^{-1}\upsilon^\dagger w = (w^{-1} + \upsilon K \upsilon^\dagger)^{-1} \,, \\ &\widetilde{W}_N \overline{\Delta} = P^\dagger \widetilde{w} d \,, \quad \widetilde{W}_N \equiv P^\dagger \widetilde{w} P \,, \quad \widetilde{C}_N \equiv \widetilde{W}_N^{-1} . \end{aligned} \tag{6}$$

Letting the eigenvalues of $K \to \infty$ gives the uniform prior, and in that case the modified w has the template vectors projected out, i.e., $\upsilon\widetilde{w}\upsilon$=0. The effect of marginalizing out κ is to enhance the noise, in a way that loses time-translation invariance: i.e., $\widetilde{w}$ is not diagonal in temporal frequency ω, and cannot be computed directly by the FFT, followed by smoothing. However, the relationship eq.(6) allows $\widetilde{c}_n \equiv \widetilde{w}^{-1}$ to be computed with one $N_\kappa \times N_\kappa$ matrix inversion in addition to the FFT.

$\mathcal{P}$(**signals**|**parameters**): The signal component of the map,

$$\Delta_{cp} = \sum_j s_{(j)cp} + \sum_A \Upsilon_{cp,A} \kappa_A \,, \tag{7}$$

includes the frequency-independent primary, $s_{(1)cp}$ say, foreground and secondary anisotropy components $s_{(j)cp}$ and additional template-based signals $\Upsilon_{cpA}\kappa_A$. Sample templates Υ_{cp} are: a constant for an average offset, $\hat{q}_{pi}$ for the 3 dipole components, $\hat{q}_{pi}\hat{q}_{pj} - \delta_{ij}/2$ for the 5 quadrupole components, the pattern of dust emission in the IRAS/DIRBE maps, the beam pattern for point sources, $\Upsilon_{cp} = \mathrm{beam}_c(\hat{q}_p - \hat{q}_{source})$.

For primary anisotropies, the signal is often assumed to be Gaussian,

$$\ln \mathcal{P}(s|y) = -\tfrac{1}{2} s^\dagger C_T^{-1} s - \tfrac{1}{2} \mathrm{Tr} \ln C_T - N_{pix} \ln \sqrt{2\pi} \,, \tag{8}$$

where $C_T(y) \equiv W_T^{-1}$ is the theory pixel-pixel correlation matrix and we have denoted the set of cosmological parameters it depends on by y_α. (Tr denotes trace.) A great advantage of assuming all signals and noise to be

Gaussian is that marginalizing all signal amplitudes $\sum s_{(j)cp}$ and template amplitudes κ_A yields another simple Gaussian:

$$\begin{aligned}
&\ln \mathcal{P}(\overline{\Delta}|y) = -\tfrac{1}{2}\overline{\Delta}^{\dagger}\widetilde{W}_t\overline{\Delta} + \tfrac{1}{2}\mathrm{Tr}\ln \widetilde{W}_t - N_{pix}\ln\sqrt{2\pi}\,, \qquad (9)\\
&W_t \equiv (C_N + C_T)^{-1}\,,\ \widetilde{W}_t \equiv (\widetilde{C_N} + C_T)^{-1}\,,\ \widetilde{C_N} \equiv C_N + \Upsilon K \Upsilon^{\dagger}\,,\\
&\widetilde{W}_t \equiv W_t - [\Upsilon^{\dagger}W_t]^{\dagger}(K^{-1} + \Upsilon^{\dagger}W_t\Upsilon)^{-1}[\Upsilon^{\dagger}W_t]\,.
\end{aligned}$$

As for the time templates, the marginalization can be thought of as inducing enhanced noise in the template spatial structures.

The other part of the probability chain, $\ln\mathcal{P}(\Delta|\overline{\Delta})$, gives a Gaussian for the signal given the observations, with mean the Wiener filter of the map:

$$\begin{aligned}
\ln \mathcal{P}(\Delta|\overline{\Delta}) &= -\tfrac{1}{2}(\Delta - \langle\Delta|\overline{\Delta}\rangle)^{\dagger}C_{\Delta\Delta}^{-1}(\Delta - \langle\Delta|\overline{\Delta}\rangle)\\
&\quad -\tfrac{1}{2}\mathrm{Tr}\ln C_{\Delta\Delta} - N_{pix}\ln\sqrt{2\pi},\\
\langle\Delta|\overline{\Delta}\rangle &= C_T\widetilde{W}_t\overline{\Delta}\,, \quad \langle\delta\Delta\delta\Delta^{\dagger}|\overline{\Delta}\rangle = C_T - C_T\widetilde{W}_tC_T\,. \qquad (10)
\end{aligned}$$

The brackets indicate an ensemble average. The lemma

$$B(A+B)^{-1} = (A^{-1} + B^{-1})^{-1}A^{-1} \qquad (11)$$

between two invertible matrices A and B is convenient to remember.

We can also determine the distribution, given the observations, of the various component signals $s_{(j)}$ and of the noise; all of these are Gaussians with means and variances given by:

$$\begin{aligned}
&\langle s_{(j)}|\overline{\Delta}\rangle = C_{T(j)}\widetilde{W}_t\overline{\Delta},\ \ \langle\delta s_{(j)}\delta s^{\dagger}_{(j')}\rangle = C_{T(j)}\delta_{jj'} - C_{T(j)}\widetilde{W}_tC_{T(j')},\\
&\langle n|\overline{\Delta}\rangle = C_N\widetilde{W}_t\overline{\Delta},\ \ \langle\delta n\delta n^{\dagger}\rangle = C_N - C_N\widetilde{W}_tC_N,\\
&\delta s_{(j)} \equiv s_{(j)} - \langle s_{(j)}|\overline{\Delta}\rangle,\ \ \delta n \equiv n - \langle n|\overline{\Delta}\rangle. \qquad (12)
\end{aligned}$$

For foregrounds and secondaries, all higher order correlations are needed since the distributions are non-Gaussian. One sample non-Gaussian prior distribution that has been analyzed for these signals and even for primaries is the so-called maximum entropy distribution, which has been widely used by radio astronomers in image construction from interferometry data. Although "max-ent" is often a catch-all phrase for finding the maximum likelihood solution, the implementation of the method involves a specific assumption for the nature of $\mathcal{P}(s_{(j)cp}|\mathrm{theory})$. For positive signals, $s > 0$, it is derived as a limit of a Poisson distribution, $P_n = \mu^n e^{-\mu}/n!$, in the Stirling formula limit,

$$\begin{aligned}
&\ln n! = \Gamma(1+\mu\zeta) \approx \mu\zeta\ln(\mu\zeta) - \mu\zeta + \ln(2\pi\mu\zeta)/2\,:\\
&\ln\mathcal{P}(s|y) = (-\zeta\ln(\zeta) + \zeta - 1)\mu \ \text{ where }\ \zeta = s/\mu\,. \qquad (13)
\end{aligned}$$

The subdominant $\ln(2\pi\mu\zeta)$ term is dropped and the Poisson origin is largely forgotten, replaced by an interpretation involving the classic Boltzmann entropy, $-\zeta\ln\zeta$, with the constraint on the average $\langle\zeta\rangle = 1$ and μ now interpreted as a measure.

Another example of positive signals is standard emitting sources. These often have power law distributions over some flux range, with leading term $\ln\mathcal{P} = -\gamma\ln\zeta$. These have to be regulated at small and/or large ζ to converge.

For symmetric positive and negative signals, a possible form is

$$\begin{aligned} \mathcal{S}_P(s_{(j)}) &= \ln\mathcal{P}(s_{(j)}|y) = -x\ln(x+\sqrt{1+x^2}) + \sqrt{1+x^2} - 1, \\ x^2 &\equiv s^\dagger_{(j)} C^{-1}_{T(j)} s_{(j)}, \end{aligned} \tag{14}$$

which reduces to $-x^2/2$ in the small fluctuation (small x) limit, but has non-Gaussian wings. This form has been used in CMB forecasting for Planck by [38, 41], but any form which retains the basic $-x^2/2$ limit for small x and a leveling off at high x can have a similar effect. For example, instead of the $-\text{arcsinh}(x)$ for $\partial\ln P/\partial x$, the more drastic deviation from Gaussian, $-\arctan(x)$, has been used in radio astronomy.

Of course, in spite of the analytic form for $\mathcal{S}_P$, the integration over the various $s_{(j)}$ to get $\mathcal{S}(\overline{\Delta}|y)$ does not have an analytic result. Nonetheless, iterative techniques can be used to solve for the maximum entropy solution, and errors can be estimated from the second derivative matrix of the total entropy.

Cleaning and Separating: Separating foregrounds and secondaries from the primary CMB into statistically accurate maps is a severe challenge. It has been common to suppose $s_{(j)cp} = f_{(j)c}s_{(j)p}$ is separable, into a product of a given function of frequency times a spatial function, but this is clearly not the case for foregrounds, and for most secondary signals, although for some, such as the Sunyaev-Zeldovich effect, it is nearly so (Fig. 1). Another approach which is crude but reasonably effective is to separate the signals using the multifrequency data on a pixel-by-pixel basis, but the accuracy is limited by the number of frequency channels.

It is clearly better to incorporate the knowledge we do have on the $s_{(j)cp}$ spatial patterns. This has been done so far by either assuming the prior probability for $\mathcal{P}(s_{(j)cp}|\text{theory})$ is Gaussian or of the max-ent variety. Even the obviously incorrect Gaussian approximation for the foreground prior probabilities has been shown to be relatively effective at recovering the signals, in simulations for Planck performed by Bouchet and Gispert [39]; see also [40].

The Poisson aspect of the maximum entropy prior makes it well-suited to find and reconstruct point sources, and more generally concentrated ones: for Planck-motivated simulations, it has been shown to be better at recovery than the Gaussian approximation [41]. Errors on the maximum entropy maps are estimated from the second derivative matrix of the likelihood function. This means that non-Gaussian aspects of the errors are ignored. Further, foregrounds and secondary anisotropies have non-Gaussian distributions which are certainly not of the max-ent form, and can only be determined by Monte Carlo methods, simulating many maps — and that only if we know the theory well enough to construct such simulations, a tall task for the foregrounds. Fortunately, with current data these issues have not been critical to solve before useful cosmological conclusions could be drawn, but much more exploration is needed to see what should be done with separation for the very high quality Planck and MAP datasets.

Templates are very useful for foreground removals. For example, maps at higher or lower frequencies than the target one, where the foregrounds clearly dominate, can be the templates. The marginalized formulas given above get rid of the templates but of course the amplitude and distribution may be of great interest. The other part of the likelihood (before marginalization) is $\ln \mathcal{P}(\kappa|\overline{\Delta})$; with the Gaussian prior assumption, it is a Gaussian distribution with the mean a weighted template-filtering of the data, akin to a Wiener-filtering:

$$\langle\kappa|\overline{\Delta}\rangle = [K^{-1} + \Upsilon^\dagger W_t \Upsilon]^{-1}\Upsilon^\dagger W_t \overline{\Delta} \rightarrow [\Upsilon^\dagger W_t \Upsilon]^{-1}\Upsilon^\dagger W_t \overline{\Delta}\,,$$
$$C_{\kappa\kappa} = \langle\delta\kappa\delta\kappa^\dagger|\overline{\Delta}\rangle = [\Upsilon^\dagger W_t \Upsilon]^{-1}, \quad \delta\kappa = \kappa - \langle\kappa|\overline{\Delta}\rangle. \tag{15}$$

An example [43] of the use of templates was comparing Wiener maps and various statistical quantities for the SK95 data under the hypothesis that it was pure CMB or CMB plus a "dust template" – a cleaned IRAS $100\mu m$ map with strong input from the DIRBE 100 and 240 μm maps [42]. A correlated component in the 30–40 GHz SK95 data was found [43], but at a low level compared with the CMB signal, i.e., not much dust contamination, in agreement with what other methods showed [44].

Templates for Point Sources: Another useful method in finding sources and cleaning data is to look for known or parameterized non-Gaussian patterns with the position, and possibly orientation and scale of the templates being allowed to float. An example is (optimal) point source removal, in which the map is modeled as a profile times an amplitude κ plus noise plus the other signals present [45]. As the source position varies, eq.(15) defines a $\overline{\kappa}$–map. The dimensionless map $\bar{\nu} \equiv C_{\kappa\kappa}^{-1/2}\overline{\kappa}$ gives the number of σ a signal with the beam pattern shape has. If $|\bar{\nu}|$ is small, e.g., below 2.5-sigma, it will be consistent with the Gaussian signals in the map; if large, e.g., above 5-sigma, it will stick out as a non-Gaussian spike ripe for removal. The question then is how many σ one cleans to, and what impact setting such a threshold has on the statistics of the cleaned map, $\Delta' = \Delta - \Upsilon\langle\kappa|\overline{\Delta}\rangle$ with its corrected weight matrix $\widetilde{W}_t' = W_t - W_t\Upsilon C_\kappa \Upsilon^\dagger W_t$. (Actually $\widetilde{W}_t'$ always acts on Δ, so removing $\Upsilon\langle\kappa|\overline{\Delta}\rangle$ is not necessary in the limit of a wide prior dispersion for κ.)

Fig. 6 shows an application of this method to DMR. A comparison is also made between the mean noise field and the $\overline{\kappa}$ field: they are quite similar.

A similar optimal filter was used in [46] for ideal forecasting and in [47] for SZ source finding (based on the beta-profile). Another approach is to not pay such attention to the detailed optimal filter, but use a Gaussian-based filter with variable scale to try to identify point sources. [48] used a Mexican hat filter, i.e., the second derivative of a Gaussian. This continuum wavelet method is optimal only in the case of a scale invariant signal, a Gaussian beam and no noise. In the pure noise limit with no signal, the straight Gaussian is optimal. The best choice lies in between.

Comparing: The simplest way to compare data sets A and B is to interpolate on overlap regions dataset A onto the pixels of dataset B. This of course will not always be possible, especially if the pixels are generalized ones, e.g., modal projections such as in the SK95 dataset. We would also like to compare nearby regions even if overlap is only partial. This requires an extrapolation as well as interpolation. In [49], a simple mechanism was described based upon Bayesian methods and a Gaussian interpolating theory for comparing CMB data sets. The statistics of the two experiments is defined by a joint probability for the two datasets, $P(A, B|C_T)$, where the theory matrix C_T has AB as well as AA and BB parts connecting it. A probability enhancement factor $\beta = \ln P(A|B, C_T)/P(A|C_T)$ was introduced, which is invariant under A, B interchange. What was visually instructive was to compare the A-data Wiener-filtered onto B-pixels with the Wiener-filter of B-data. They should bear a striking resemblance except for the errors. The example used in [49] was SK95 data for A and MSAM92 for B, which showed remarkable similarities: indeed that the two experiments were seeing the same sky signal, albeit with some deviations near the endpoint of the scan. Although these extrapolations and interpolations are somewhat sensitive to the interpolating theory, if it is chosen to be nearly the best fit the results are robust. The same techniques were used for an entirely different purpose in [50], testing whether two power spectra with appropriate errors, in this case from Boomerang and Maxima, could have been drawn from the same underlying distribution. The conclusion was yes.

Compressing: We would like to represent in a lossless way as much in-

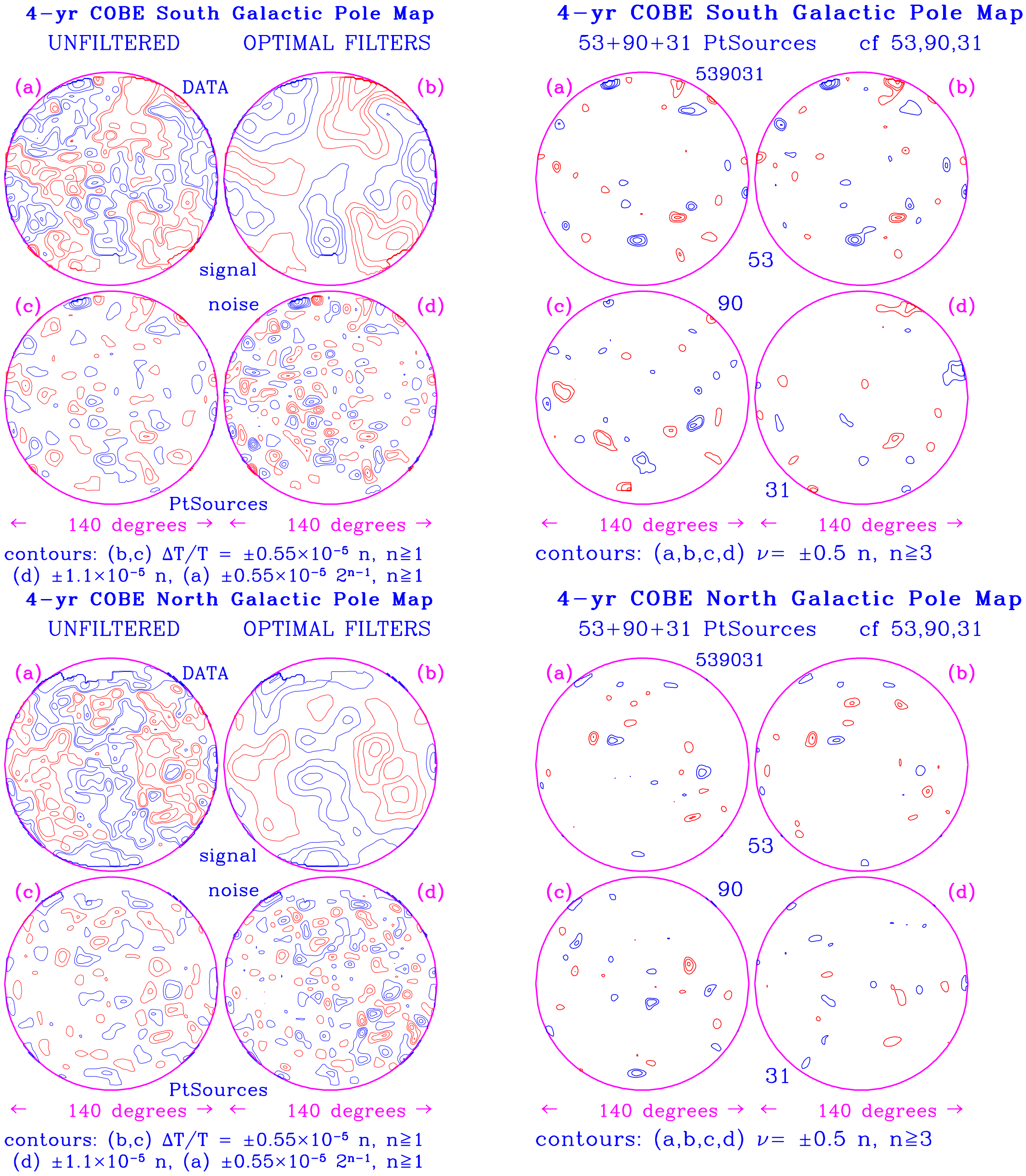

Figure 6. The upper left panel shows the unfilterd DMR map (combined 53+90+31 GHz data) of a 140° diameter region centered on the South Galactic Pole, with the contours indicated, (b) and (d) show the Wiener-filtered signal and noise maps, while (c) shows the fluctuations of height $\nu\sigma$, with $|\nu| > 1.5$. The lower left panel shows the same for the North Galactic Pole. Note the similarity between (c) and (d), that is that the Point Sources found at these low contours are consistent with random concentrations of noise, the logical conclusion within the statistics: there are no rare events that are good source candidates. We chose such a low cut so that one could see the effects of making 2 or 3σ cuts just by counting in, each positive contour being 0.5 higher than the previous. The right panels show the contours of ν rather than $\Delta T/T$ for the individual (A+B) frequency maps. A true source would be expected to persist in the 3 maps, as well as being in the combined one, the strength variation across channels depending upon the emission mechanism. Again there are no obvious hits. Apart from being a useful way to detect point sources, this is a nice mechanism for looking for anomalous local non-Gaussian patterns, to be investigated to determine if we are dealing with funny noise or true signals by appropriate follow-up investigations.

formation as we can in much smaller datasets. Timestreams to maps (and map-orthogonal noise) are a form of compression, from N_{tbits} to N_{pix}. Map manipulations, even for Gaussian theories, are made awkward because the total weight matrix involves C_N, often simplest in structure in the pixel basis, and C_T, most often naturally represented in a spherical harmonic basis. Signal-to-noise eigenmodes ξ_k are a basis in which $\widetilde{C}_N^{-1/2} C_T \widetilde{C}_N^{-1/2}$ is diagonal, hence are statistically independent of each other if they are Gaussian. They are a complete representation of the map. Another S/N basis of interest is one in which $C_T^{1/2} \widetilde{C}_N^{-1} C_T^{1/2}$ is diagonal, since $\widetilde{C}_N^{-1}$ comes out directly from the signal/noise extraction step. Finding the S/N modes is another $O(N_{pix}^3)$ problem. Typically, $\langle \xi_k^2 \rangle$ falls off dramatically at some k, and higher modes of smaller eigenvalues can be cut out without loss of information. These truncated bases can be used to test the space of theoretical models by Bayesian methods e.g., [57], and determine cosmic parameters. Further compression to bandpowers is even better for more rapid determination of cosmic parameters. The related concept of parameter eigenmodes is discussed later.

Quadratic Estimation of Power Spectra and Parameters: Estimating power spectra and correlation functions of the data has traditionally been done by minimizing a χ^2 expression,

$$\chi^2 = Tr[W(\overline{\Delta\Delta^\dagger} - C(y))W(\overline{\Delta\Delta^\dagger} - C(y))] \, / \, 4, \qquad (16)$$

where W is some as yet unspecified weight matrix. The 1/4 is from the pair sum. The critical element is to make a model $C(y)$ for the pixel pair correlation $\overline{\Delta\Delta^\dagger}$ which is as complete a representation as possible. For example, if we adopt a linear dependance about some fiducial $C_* = C(y_*)$,

$$i.e., \; C(y) = C_* + \sum_\beta C_\beta \delta y^\beta \, , \text{ then}$$

$$\sum_\beta \mathcal{F}_{\alpha\beta} \delta y^\beta = \tfrac{1}{2} \mathrm{Tr}[W C_\alpha W (\overline{\Delta\Delta^\dagger} - C_*)], \quad C_\alpha \equiv \partial C / \partial y^\alpha \, , \qquad (17)$$

$$\mathcal{F}_{\alpha\beta} \equiv \tfrac{1}{2} \mathrm{Tr}[W C_\alpha W C_\beta] - \tfrac{1}{2} \mathrm{Tr}[W C_{\alpha\beta} W (\overline{\Delta\Delta^\dagger} - C_*)], \quad C_{\alpha\beta} \equiv \frac{\partial^2 C}{\partial y^\alpha \partial y^\beta} ,$$

$$\text{Fisher information matrix :} \quad F_{\alpha\beta} \equiv \langle \mathcal{F}_{\alpha\beta} \rangle = \tfrac{1}{2} \mathrm{Tr}[W C_\alpha W C_\beta] \, . \qquad (18)$$

Operators linear in pixel pairs are called *quadratic estimators*. Since $\mathcal{F}_{\alpha\beta}$ has Δ dependence, this is not strictly a quadratic estimator. However, if the Fisher matrix $F_{\alpha\beta}$, the ensemble average of $\mathcal{F}_{\alpha\beta}$, is used, it is. Iterations with F rather than $\mathcal{F}$ converge to the same y_* if y is continually updated by the δy. $C_* = C_N + C_T$ has been assumed.

The correlation in the parameter errors can be estimated from the average of a quartic combination of pixel values,

$$\begin{aligned}
&\langle \delta y \delta y^\dagger \rangle = \mathcal{F}^{-1}\tfrac{1}{2}\mathrm{Tr}[WC_\alpha WCWC_\beta WC]\mathcal{F}^{-1} \\
&\quad +\mathcal{F}^{-1}\tfrac{1}{4}\mathrm{Tr}[WC_\alpha W(\delta C)]\mathrm{Tr}[WC_\beta W(\delta C)]\mathcal{F}^{-1}\,, \\
&\text{where } (\delta C) \equiv C - C_*; \\
&\text{note } \langle \delta y \delta y^\dagger \rangle = \mathcal{F}^{-1} \text{ if } C = C_* \text{ and } W = C_*^{-1}\,.
\end{aligned}$$

Another approach is to estimate the errors by taking the second derivative of χ^2, $\partial\chi^2/\partial y^\alpha \partial y^\beta = \mathcal{F}_{\alpha\beta}$:

$$\begin{aligned}
&\langle \partial^2\chi^2/\partial y_\alpha \partial y_\beta \rangle = \tfrac{1}{2}\mathrm{Tr}[WC_\alpha WC_\beta] - \tfrac{1}{2}\mathrm{Tr}[WC_{\alpha\beta}W(\delta C)] \\
&\quad \to F_{\alpha\beta} \text{ as } (\delta C) \to 0.
\end{aligned}$$

These two limiting cases hold for the maximum likelihood solution as we now show.

Maximum Likelihood Power Spectra & Iterative Quadratics: If we look for the maximum likelihood solution for a Gaussian signal plus noise, then it is of this form, with $W = \widetilde{W}_t$, the "optimal" weight which gives the minimum error bars, associated with the correlation function $\langle \delta y_\alpha \delta y_\beta \rangle = [\mathcal{F}^{-1}]_{\alpha\beta}$. This is found by differentiating $-2\ln\mathcal{L} = \mathrm{Tr}[\widetilde{W}_t\overline{\Delta\Delta}^\dagger)] - \mathrm{Tr}[\ln\widetilde{W}_t] + N_{pix}\ln(2\pi)$. A solution procedure for this optimal case is the Newton-Raphson method, with the weight matrix updated in each iterative improvement δy_β to $y_{*\beta}$, at a cost of a matrix inversion.

For the MAP satellite, a solution has been proposed which relies upon the statistical isotropy of the signal and a simple form for the noise. Unfortunately, most experiments will produce maps with more complicated noise properties.

Only in the Gaussian or uniform prior cases is the integration over $s_{(j)}$ analytically calculable. The usual procedure for maximum entropy priors is to estimate errors from the second derivative of the likelihood, *i.e.*, effectively use a Gaussian approximation.

Solving for the roots of $\partial\ln\mathcal{L}/\partial y_\alpha$ using the Newton-Raphson method requires that we calculate $\partial^2\ln\mathcal{L}/\partial y_\alpha\partial y_\beta$, the curvature of the likelihood function. Other matrices have been used for expediency, since one still converges on the maximum likelihood solution; in particular the curvature matrix expectation value, i.e., the Fisher matrix $F_{\alpha\beta}$, which is easier to calculate than $\mathcal{F}_{\alpha\beta}$.

The bandpower example: A bandpower in a region b in ℓ-space is defined by a shape $\mathcal{C}_\ell^{(s)}$ and an amplitude q^b. $\overline{\Delta\Delta}^\dagger$ is modeled by

$$C(q) = \widetilde{C}_N + \sum_b C_{Tb}^{(s)} q^b = C_* + \sum_b C_{Tb}^{(s)} \delta q^b \,. \tag{19}$$

The usual technique [11] is to just use this linear form in the q^b, the downside being that the bandpowers could be negative – when there is little signal and the pixel pair estimated signal plus noise is actually less than the noise as estimated from C_N. By choosing $\exp(q^b)$ rather than q^b as the variables, positivity of the signal bandpowers can be assured. (From now on we use the summation convention for repeated indices.)

To solve eq.(17) with the optimal weight by an iterative Newton-Raphson method requires some care in controlling the step size each time q^b is updated. This method has been used for bandpower estimates for COBE and SK95, and for Boomerang and Maxima, among others. The cost is $\mathcal{O}(N_{pix}^3)$ for each parameter, so this limits the number of pixels that can be treated. For $N_{pix} \gtrsim 200K$ or so, the required computing power becomes prohibitive, requiring 640 Gb of memory and of order 3×10^{17} floating-point operations, which translates to a week or more even spread over a Cray T3E with $\sim$ 1024 900-MHz processors, and is clearly impossible for the ten million pixel Planck data set. Numerical roundoff would be prohibitive anyway. A case has been made for highly simplified weights (e.g., diagonal in pixel space) that would still deliver the basic results, with only slightly increased error bars [53]. We note that the compressed bandpowers estimated from all current data and the idealized satellite forecasts shown in Fig. 3 were also computed using this maximum likelihood Newton-Raphson iteration method.

However, to make use of the q^b for cosmological parameter estimation, we need to know the entire likelihood function for q^b, and not just assume a Gaussian form using the curvature matrix. This can only be done by full calculation, though there are two analytic approximations that have have been shown to fit the one-point distributions quite well in all the cases tried [51]. The simplest and most often used is an "offset lognormal" distribution, a Gaussian in $z^b = \ln(q^b + x^b)$, where the offset x^b is related to the noise in the experiment:

$$\begin{aligned}
&\mathcal{P}(q) \propto \exp[-\tfrac{1}{2}(z-\bar{z})^b \mathcal{F}_{bb'}^{(z)} (z-\bar{z})^{b'}], \\
&\mathcal{F}_{bb'}^{(z)} = (\bar{q}^b + x^b)\mathcal{F}_{bb'}^{(q)} (\bar{q}^{b'} + x^{b'}), \\
&\bar{q}^b / x^b = [\mathcal{F}_{bb'}^{(q)0} / \mathcal{F}_{bb'}^{(q)}]^{1/2} - 1 \,,
\end{aligned} \tag{20}$$

where $\bar{q}^b$ is the maximum likelihood value. To evaluate x^b in eq.(20), the curvature matrix evaluated in the absence of signal, $\mathcal{F}_{bb'}^{(q)0}$, is needed. Values of x^b and other data for a variety of experiments are given in [51]. For Boomerang and Maxima analysis, the two Fisher matrices are standard output of the "MADCAP" maximum likelihood power spectrum estimation code [52].

The use of non-optimal weighting schemes has had a long history. For example, a quadratic estimator of the correlation function was used in the very first COBE DMR detection publication, with $diag(W_N)$ for W, a choice we have also used for COBE and the balloon-borne FIRS. Here one is estimating correlation function amplitudes C^α, where α refers to an angular bin: $C_{pp'} = C_{Npp'} + \sum_\alpha C^\alpha \chi_\alpha(\theta_{pp'})$, where $\cos(\theta_{pp'}) = \hat{q}_p \cdot \hat{q}_{p'}$ and $\chi_\alpha(\theta)$ is one inside and zero outside of the bin. As for bandpowers, the distribution is non-Gaussian, so errors were estimated using Monte Carlo simulations of the maps.

For the power spectra, the model for $\overline{\Delta\Delta}^\dagger$ is

$$C_{pp'} = C_{Npp'} + \sum_b q^b \sum_{\ell \in b} \mathcal{C}_\ell^{(s)} P_\ell(\cos\theta_{pp'}) \frac{(\ell + \frac{1}{2})}{\ell(\ell+1)},$$

$$\langle \mathcal{C} \rangle_b = q^b \langle \mathcal{C}^{(s)} \rangle_b, \quad \text{where } \langle \mathcal{C}^{(s)} \rangle_b \equiv \sum_{\ell \in b} \mathcal{C}_\ell^{(s)} \frac{(\ell + \frac{1}{2})}{\ell(\ell+1)},$$

where $\mathcal{C}_\ell^{(s)}$ is the shape within the ℓ-bin b, including the window function in that bin, which takes into account both beam and pixelization effects. In general, the implicit isotropization that makes this only a function of ℓ does not work at high ℓ because of the pixelization, and this can significantly complicate the treatment. (Other weightings are possible in the renormalization from q^b to the conventional bandpower $\langle \mathcal{C} \rangle_b$ [51].)

Any weight matrix W which is diagonal in pixel space is easy and fast to calculate. In [53] it was shown that making W the identity was adequate for Boomerang-like data. The actual approach used a fast highly binned estimate of $C(\theta)$, which can be computed quickly, which was then lightly smoothed, and finally a Gauss-Legendre integration of $C(\theta)$ was done to estimate $\langle \mathcal{C} \rangle_b$.

Another way to proceed is to work with the full $W = \widetilde{W}_t$, but use symmetries and simplifications of specific experiments to speed things up. Since many of the signals are most simply described in multipole space, it is natural to try to exploit this basis when implementing the parameter estimation method described above. We should also try recasting the calculation to take advantage of the simple form the weight matrix C_N^{-1} sometimes has in the pixel basis. Finally, with iterative methods we can exploit approximate symmetries of these matrices which can speed up the algorithms tremendously. In [54], these techniques were applied to MAP simulations to great effect.

More could be said about ways to make power spectrum estimation nearly optimal but also tractable, but much more remains to be explored.

Forecasting: A typical forecast involves making large simplifications to the full statistical problem. For example, the noise is homogeneous and white, and the fluctuations are Gaussian. This was applied to estimates of how well cosmic parameters could be determined for the LDBs and MAP and Planck in the 9 parameter case described earlier [10], and for cases where the power spectrum parameters are allowed to open up beyond just tilts and amplitudes to parameterized shapes [55]. The methods were also applied when CMB polarization was included, and when LSS and supernovae were included [56].

The procedure is exactly like that for maximum likelihood parameter estimation, except one often does not bother with a realization of the power spectrum, rather uses the average (input) value and estimates errors from the ensemble average curvature, i.e., the Fisher matrix: F^{-1} is therefore a measure of $\langle \delta y_\alpha \delta y_\beta \rangle$, where $\delta y_\alpha = y_\alpha - \bar{y}_\alpha$ is the fluctuation of y_α about its most probable value, $\bar{y}_\alpha$. As mentioned above, when the y_α are the bandpowers instead of the cosmic parameters, we get the forecasts of power spectra and their errors, as for MAP and Planck in Fig. 3. Those results ignored foregrounds. However other authors have treated foregrounds with the Gaussian and max-ent prior assumptions in this simplified noise case [39, 41].

Clearly forecasting can become more and more sophisticated as realistic noise and other signals are added, ultimately able to grow into a full pipeline-testing simulation of the experiment.

Current CMB and CMB+LSS Constraints: Although the purpose of this paper is to describe analysis techniques, the comparison of the radically compressed current bandpowers with the models in Fig. 3 invite a brief description of the results that are derived for the 7-parameter "minimal" inflation-motivated parameter set. For a more complete description, see [28, 50, 58].

As the upper right panel of Fig. 3 suggests, the primordial spectral tilt n_s is well determined using CMB data alone: $n_s = 1.05^{+.09}_{-.08}$, rather encouraging for the nearly scale invariant models preferred by inflation theory. The figure also shows that constraints on Ω_Λ will not be very good, but the strong dependence of the position of the acoustic peaks on Ω_k means that it is better restricted: $\Omega_{tot} \approx 1.09 \pm 0.07$. The baryon abundance is also well determined, $\Omega_b h^2 = 0.031 \pm 0.005$, above the 0.019 ± 0.002 estimate from deuterium observations combined with Big Bang Nucleosynthesis theory.

The isocurvature CDM models with tilt $n_{is} > 0$ and the isocurvature defect models (*e.g.* strings and textures) [12, 13] shown in Fig. 3 clearly have more difficulty with the CMB bandpowers.

There is a region of the spectrum that nominally seems to be out of reach of CMB analysis, but is actually partly accessible because ultralong waves contribute gentle gradients to our CMB observables. For example, the size of compact spatial manifolds has been constrained; e.g., [59] find for flat equal-sided 3-tori, the inscribed radius must exceed $1.1(2H_0^{-1}) = 6600\,\mathrm{h}^{-1}$ Mpc from DMR at the 95% confidence limit, with the weaker $> 0.7(2H_0^{-1})$ constraint for asymmetric 1-tori. It is also not as strong if the platonic-solid-like manifolds of compact hyperbolic topologies are considered, though the overall size of the manifold should be at least of order the last scattering surface radius [59].

We have always combined CMB and LSS data in our quest for viable models. For example, DMR normalization determines σ_8 to within 7%, and comparing with the $\sigma_8 \sim 0.6\Omega_m^{-0.56}$ target value derived from cluster abundance observations severely constrains the cosmological parameters defining the models. This is further restricted when the possible variations in the spectral tilt allowed by the COBE data are constrained by the higher ℓ Boomerang data. Similar constrictions arise from galaxy-galaxy and cluster-cluster clustering observations: the shape of the linear density power spectrum must match the shape reconstructed from the data. In [28, 50, 58], a simple shape parameter and σ_8 were used with broad distributions to characterize the LSS data, so that these prior probability choices would not be controversial, but encompass the ranges that almost all cosmologists would believe.

Using all of the current CMB data and the LSS priors, [58] got $H_0 = 70 \pm 9$, $\Omega_\Lambda = 0.63 \pm 0.09$, with n_s, Ω_{tot} and $\Omega_b h^2$ virtually unchanged from the CMB only result. Apart from the sigificant Ω_Λ detection, the dark matter is also strongly constrained, $\Omega_{cdm} h^2 = 0.14 \pm 0.03$.

Restricting Ω_{tot} to be unity as in the usual inflation models, with CMB+LSS, $\Omega_b h^2 = 0.030 \pm 0.004$, $\Omega_{cdm} h^2 = 0.17 \pm 0.02$, $\Omega_\Lambda = 0.66 \pm 0.06$ and $n_s = 1.03^{+.08}_{-.07}$. If the equation of state of the dark energy is allowed to vary, $w_Q < -0.3$ is obtained, becoming substantially more restrictive when supernova information is folded in, $w_Q < -0.7$.

Fig. 7 gives a visual perspective on how the parameter estimations for the adiabatic models evolved as more CMB data were added (but with always the LSS prior included for this figure). With just the COBE-DMR+LSS data, the 2-sigma contours were already localized in $\Omega_{cdm} h^2$. Without LSS, it took the addition of Maxima-1 before it began to localize. Ω_k localized near zero when TOCO was added to the April 99 data, more so when Boomerang-NA was added, and much more so when Boomerang-LDB and Maxima-1 were added. Some n_s localization occurred with just "prior-CMB" data. $\Omega_b h^2$ really focussed in with Boomerang-LDB and Maxima-1, as did Ω_Λ.

However, the lesson of CMB analysis done to date is that almost all of the time and debate is spent on the path to a believable power spectrum, with the cosmological parameters quickly dropping out once the spectrum has been agreed upon. And the analysis challenge, requiring clever new algorithms, gets stronger as we move to more LDBs, to MAP and to Planck.

Many people have worked with us on aspects of the CMB analysis touched on here, Carlo Contaldi, George Efstathiou, Andrew Jaffe, Lloyd Knox, Steve Myers, Barth Netterfield, Ue-Li Pen, Dmitry Pogosyan, Simon Prunet, Kris Sigurdsen, Tarun Souradeep, Istvan Szapudi, Neil Turok, James Wadsley; Julian Borrill, Eric Hivon and other members of the Boomerang team.

References

1. Seljak, U. and Zaldarriaga, M. 1996, ApJ 469, 437. CMBfast
2. Miller, A.D. *et al.* 1999, ApJ 524, L1. TOCO
3. de Bernardis, P. *et al.* 2000, Nature 404, 995, astro-ph/00050087
4. Mauskopf, P. *et al.* 2000, ApJ Lett 536, L59. BOOM-NA
5. Hanany, S. *et al.* 2000, ApJ Lett. 545, 5, astro-ph/0005123
6. Padin, S. *et al.* 2000, ApJ Lett., submitted, astro-ph/0012211. CBI
7. Bond, J.R. 1996, in *Cosmology and Large Scale Structure*, Les Houches Session LX, eds. R. Schaeffer J. Silk, M. Spiro and J. Zinn-Justin (Elsevier), pp. 469-674
8. Bond, J.R. 1994, Phys. Rev. Lett. 74, 4369
9. e.g., Efstathiou, G. & Bond, J.R. 1999, M.N.R.A.S. 304, 75, where many other near-degeneracies between cosmological parameters are also discussed
10. Bond, J.R., Efstathiou, G. & Tegmark, M. 1997, M.N.R.A.S. 291, L33
11. Bond, J.R., Jaffe, A.H. & Knox, L. 1998, Phys. Rev. D 57, 2117, astro-ph/9708203
12. Pen Ue-Li, Seljak U. & Turok N. 1997, Phys. Rev. Lett. 79, 1611
13. Allen, B., Caldwell, R.R., Dodelson, S., Knox, L., Shellard, E.P.S. & Stebbins, A. 1997, astro-ph/9704160
14. Frenk, C.S. *et al.* 1999, ApJ 525, 554
15. Birkinshaw, M. 1999, Phys. Rep. 310, 97
16. Wadsley, J.W. & Bond, J.R. 2001, preprint
17. Bond, J.R., Pogosyan, D., Kofman, L. & Wadsley, J.W. 1998, in '"Wide Field Surveys in Cosmology", Proc. XIV IAP Colloquium Conf., ed. S. Colombi & Y. Mellier, (Paris: Editions Frontieres), p. 17, astro-ph/9810093
18. Pogosyan, D., Bond, J.R., Kofman, L. & Wadsley, J.W. 1998, in '"Wide Field Surveys in Cosmology", Proc. XIV IAP Colloquium Conf., ed. S. Colombi & Y. Mellier, (Paris: Editions Frontieres), p. 61, astro-ph/9810072
19. Persi, F.M., Spergel, D.N., Cen, R. & Ostriker, J.P. 1995, ApJ 442, 1
20. Refregier, A., Komatsu, E., Spergel, D.N. & Pen, Ue-Li 2000, astro-ph/9912180
21. Scaramella, R., Cen, R. & Ostriker, J.P. 1993, ApJ 416, 399
22. Seljak, U., Burwell, J. & Pen, Ue-Li 2000, astro-ph/0001120
23. da Silva, A.C., Barbosa, D., Liddle, A.R. & Thomas, P.A. 2000, M.N.R.A.S. 317, 37
24. Springel, V., White, M. & Hernquist, L. 2000, astro-ph/0008133
25. Bond, J.R. 1998, in *The Early Universe*, Proc. NATO Summer School, Victoria, B.C., Canada, August 1986, p. 283-334, ed. W. Unruh and G. Semenoff, Dordrecht: Reidel
26. Bond, J.R., Carr, B.J, & Hogan, C.J. ApJ 1991, 367, 420

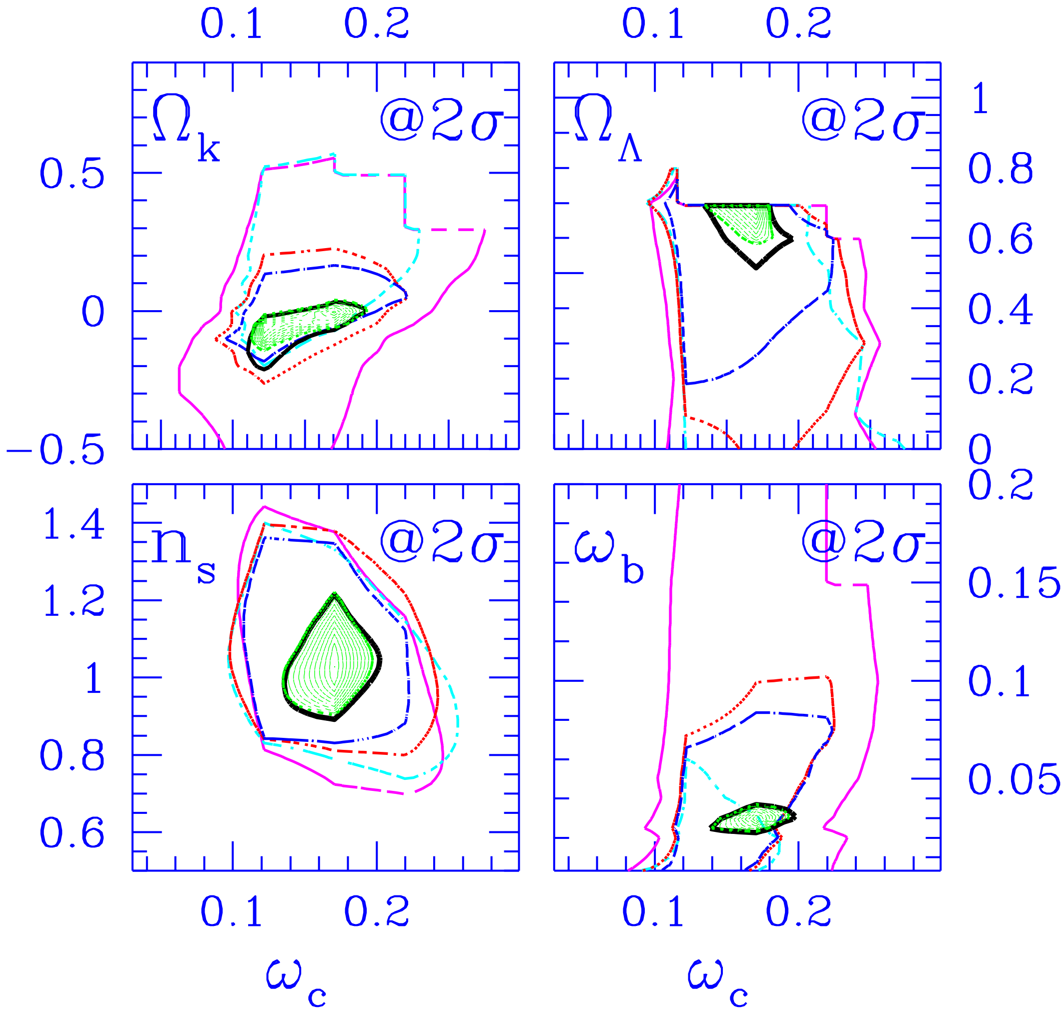

Figure 7. This figure from [58] shows 2-sigma likelihood contours for the dark matter density $\omega_c = \Omega_{cdm}\mathrm{h}^2$ and $\{\Omega_k, \Omega_\Lambda, n_s, \omega_b\}$ for the LSS prior combined with weak limits on H_0 (45-90) and cosmological age (¿ 10 Gyr), and the following CMB experimental combinations: DMR (short-dash); the "April 99" data (short-dash long-dash); TOCO+4.99 data (dot short-dash); Boomerang-NA+TOCO+4.99 data (dot long-dash, termed "prior-CMB"); Boomerang-LDB + Maxima-1 + Boomerang-NA + TOCO + 4.99 data (heavy solid, all-CMB). These 2σ lines tend to go from outside to inside as more CMB experiments are added. The smallest 2-sigma region (dotted and interior) shows SN1+LSS+all-CMB, when SNI data is added. For the Ω_Λ, n_s and ω_b plots , Ω_{tot}=1 has been assumed, but the values do not change that much if Ω_{tot} floats, as Table 1 shows for all-CMB. (The discreteness of the $\mathcal{C}_\ell$–database is responsible for the sharpness of the contour edges, especially evident in the ω_c–Ω_Λ CMB+LSS case. Though different interpolation schemes can round these off somewhat, it is the price of the finite grid.)

27. Bond, J.R. & Myers, S. 1996, ApJ Supp 103, 1; 41; 63
28. Lange, A. *et al.* 2001, Phys. Rev D, in press, astro-ph/0005004
29. Bond, J.R., Crittenden, R., Jaffe, A.H. & Knox, L. 1999, Computing in Science and Engineering 1, 21, astro-ph/9903166, and references therein
30. Ferreira, P.G. & Jaffe, A.H. 1999, M.N.R.A.S. astro-ph/9909250
31. Prunet, S. *et al.* 2001,, Proc. Garching Conference on "Analysis of Large Astronomical Data Sets", astro-ph/0101073
32. Stompor, R. *et al.* 2001, Proc. Garching Conference on "Analysis of Large Astronomical Data Sets", astro-ph/0012418
33. Dore, O., Teyssier, R., Bouchet, F.R., Vibert, D. & Prunet, S. 2001, AstronAp, astro-ph/0101112
34. Hinshaw, G. *et al.* 2001, Proc. Garching Conference on "Analysis of Large Astronomical Data Sets", astro-ph/0011555; see also E.L. Wright 1996, astro-ph/9612006
35. P.F. Muciaccia, P. Natoli & N. Vittorio 1998, ApJ 488, L63
36. K.M. Gorski, E. Hivon & B.D. Wandelt 1998, astro-ph/9812350
37. Crittenden, R. & Turok, N. 1999, astro-ph/9806374
38. Hobson, M.P. & Lasenby, A.N. 1998, M.N.R.A.S. 298, 905, astro-ph/9810240
39. Bouchet, F.R. & Gispert, R. 1999, New Astron. 4, 443, astro-ph/9903176
40. Prunet, S. *et al.* 2000b, AstronAp Lett, astro-ph/0012497
41. Hobson, M.P., Jones, A.W., Lasenby, A.N. & Bouchet, F.R. 1999, M.N.R.A.S. astro-ph/9806387
42. Schlegel, D.J., Finkbeiner, D.P. & Davis, M. 1998, ApJ 500, 525
43. Jaffe, A.H., Finkbeiner, D. & Bond, J.R. 1999, preprint
44. de Oliveira-Costa, A., et al. 1997, ApJ Lett 482, 17
45. Crittenden, R. & Bond, J.R. 2001, preprint
46. Tegmark, M. and de Oliveira-Costa, A. 1998, ApJ 500, 83
47. Haehnelt, M.G. & Tegmark. M. 1996, M.N.R.A.S. 279, 545
48. Cayon, L. *et al.* 2000, M.N.R.A.S. astro-ph/9912471
49. Knox, L., Bond,J.R., Jaffe, A.H., Segal, M. & Charbonneau, D. 1998, Phys. Rev. D, 58, 1443
50. Jaffe, A.H. *et al.* 2001, Phys. Rev. Lett., in press, astro-ph/0007333.
51. Bond, J.R., Jaffe, A.H. & Knox, L. 2000, ApJ 533, 19, astro-ph/9808264
52. Borrill, J. 2000, these proceedings; see also 1999, Phys. Rev. D 59, 7302, astro-ph/9712121.
53. Szapudi, I., Prunet, S., Pogosyan, D., Szalay, A. & Bond, J.R. 2001, ApJ Lett, in press, astro-ph/0010256
54. Oh, S.P., Spergel, D.N. & Hinshaw, G. 1999, ApJ 510, 551, astro-ph/9805339
55. Souradeep, T., Bond, J.R., Knox, L., Efstathiou, G. & Turner, M.S. 1998, in "Proceedings of COSMO-97", ed. L. Roszkowski (World Scientific), astro-ph/9802262
56. Eisenstein, D., Hu, W. & Tegmark, M. 1998, ApJ, astro-ph/9807130
57. Bond, J.R. & Jaffe, A. 1998, Phil. Trans. R. Soc. London 357, 57, astro-ph/9809043
58. Bond, J.R. *et al.* 2000, Proc. IAU Symposium 201, astro-ph/0011378; Proc. CAPP 2000, astro-ph/0011379; Proc. Neutrino 2000, astro-ph/0011377
59. Bond, J.R., Pogosyan, D. and Souradeep, T. 2000, Phys. Rev. D **62**, astro-ph/9912124, astro-ph/9912144