Week 4: More on Structure Formation and Characterization of CMB fluctuations

February 8, 2017

1 Mass function (Press-Schechter theory)

In CDM models, the power spectrum determines everything about structure in the Universe. In particular, it leads to what people refer to as "bottom-up" and/or "hierarchical clustering." To begin, note that the power spectrum $P(k)$, which decreases at large k (small wavelength) is psychologically inferior to the scaled power spectrum $\Delta^2(k) = (1/2\pi^2)k^3P(k)$. This latter quantity is monotonically increasing with k, or with smaller wavelength, and thus represents what is going on physically a bit more intuitively. Equivalently, the variance $\sigma^2(M)$ of the mass distribution on scales of mean mass M is a monotonically decreasing function of M; the variance in the mass distribution is largest at the smallest scales and smallest at the largest scales. Either way, the density-perturbation amplitude is larger at smaller scales. It is therefore smaller structures that go nonlinear and undergo gravitational collapse earliest in the Universe. With time, larger and larger structures undergo gravitational collapse. Thus, the first objects to undergo gravitational collapse after recombination are probably globular-cluster–sized halos (although the halos won't produce stars. As we will see later, the first dark-matter halos to undergo gravitational collapse and produce stars are probably $\sim 10^6\,M_\odot$ halos at redshifts $z \sim 20$. A typical L_* galaxy probably forms at redshifts $z \sim 1$, and at the current epoch, it is galaxy groups or poor clusters that are the largest mass scales currently undergoing gravitational collapse.

One of the things we can do with a power spectrum is calculate the dark-matter–halo mass function $n(M)$, the abundance of dark-matter halos as a function of their mass M. This calculation is due to Press and Schechter (PS) when they were Caltech graduate students in the early 1970s, although there is a similar calculation in the then-Soviet literature by Doroshkevich a few years earlier. The basic idea of the PS theory is to identify the high-density peaks in the primordial density distribution that will then rise to gravitationally bound objects, and then count them. The argument is as follows.

We begin by noting that inflation predicts that the distribution of primordial perturbations is *gaussian*. One of the consequences of this prediction is that if we measure the density δ_R of the primordial Universe, smoothed over a radius R, in a large number of different regions of the Universe, the distribution of densities will be a Gaussian with a variance $\sigma^2(M)$ that can be obtained from

the power spectrum as we calculated last week (and where M is the mass enclosed on average by spheres of radius R). I.e., the distribution we will obtain for δ_R is

$$P(\delta_R) = \frac{1}{\sqrt{2\pi\sigma(M)}} \exp\left[-\frac{1}{2}\frac{\delta_R^2}{\sigma^2(M)}\right]. \tag{1}$$

Note also that in the early Universe, before the density-perturbation amplitude has had a chance to grow, $\sigma(M) \ll 1$.

With time, the density-perturbation amplitude will grow, so the mass variance is in fact a function $\sigma(M,t)$ of time t as well as mass, and in linear theory, $\sigma(M,t)$ will evolve with time as calculated last week. From spherical-top-hat collapse, we have a linear-theory critical overdensity ($\delta_c = 1.69$ in EdS universe) at which point a spherical overdensity undergoes gravitational collapse. If we then go to some random position in the Universe, the probability that the smoothed density δ_R at that point is greater than the critical overdensity for collapse is

$$p(\delta_R > \delta_c) = \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{\delta_c}{\sqrt{2}\sigma(M)}\right)\right], \tag{2}$$

where $\mathrm{erf}(x)$ is the error function. The fraction of the total mass in the Universe condensed into objects of mass $> M$ is therefore,

$$F(>M) = 1 - \mathrm{erf}\left(\nu/\sqrt{2}\right), \tag{3}$$

where $\nu = \delta_c/\sigma(M,t)$. Note here that $\sigma(M,t)$ is the variance *as calculated in linear theory*, since it is being compared with δ_c, the linear-theory overdensity. Note also that to arrive at this equation, we have followed PS and multiplied, somewhat arbitrarily by a factor of 2. The reason is that in this formulation, it is only the overdense regions that undergo gravitational collapse. But if so, does that mean that the matter in the underdense half of the Universe never becomes bound into some object? Clearly not, according to PS—everything must wind up somewhere (in the words of Myron Cohen, "Everybody gotta be someplace")—and so they introduced a factor of two that has become perhaps the most famous and successful fudge factor in astrophysics.

The rest is easy. The comoving number density $n(M)\,dM$ of collapsed objects with mass in the range $M \to M + dM$ is determined from

$$M\frac{n(M)}{\bar{\rho}(M)} = \left|\frac{dF}{dM}\right|, \tag{4}$$

where $\bar{\rho}$ is the mean density, or

$$n(M) = \frac{\bar{\rho}}{M^2}\left|\frac{d\ln\sigma}{d\ln M}\right|\sqrt{\frac{2}{\pi}}\nu\exp\left(-\nu^2/2\right). \tag{5}$$

This is the mass function for collapsed objects, according to PS. Note that the mass function is normalized so that $\int_0^\infty M\,n(M)\,dM = \bar{\rho}$. What this implies is that every element of mass in the Universe is in a halo of some given mass. Also note that in an Einstein-de Sitter Universe, $\delta_c = 1.69$, but it differs if $\Omega_m \neq 1$. Note also that if $\Omega_m \neq 1$ today, then Ω_m varies with time, and so δ_c is, for a non-EdS universe, a function of time, although a fairly weakly varying function of time.

The PS mass function defines a characteristic mass $M_*(t)$ through $\sigma(M_*, t) = \delta_c$. Today, this mass is presumably close to the dark-matter–halo mass for an L_* galaxy (under the simplifying assumption that the galaxy luminosity is proportional to the dark-matter–halo mass, or perhaps a bit larger. For smaller masses, $n(M)$ varies roughly as a power law with M. The mass function decreases exponentially with masses $M > M_*$. The PS formalism also allows us to see that how hierarchical clustering occurs. The dependence of $\sigma(M, t)$ on mass M remains constant at all times; the linear-theory amplification of $\sigma(M, t)$ with time t is equal for all masses. Therefore, $M_*(t)$ is a monotonically increasing function of mass, and the characteristic mass of collapsed objects increases with time.

With the PS mass function, we can appreciate one of the current problems facing galaxy-formation theory: For a power spectrum $P(k) \propto k^n$, $\sigma(M) \propto M^{-(n+3)/6}$, implying that $n(M) \propto M^{(n-9)/6}$. At galaxy scales ($\simeq$Mpc), the CDM power spectrum has a spectral index $n \simeq -2$ implying $n(M) \propto M^{-2}$. However, the observed galaxy luminosity function is roughly $\phi(L) \propto L^{-1}$. Therefore, CDM models for galaxy formation predicts too many low-mass galactic halos. The most likely resolution is that some of the gas-dynamical or star-formation processes required to create a luminous galaxy in a dark-matter halo become inefficient in low-mass halos. This is a very active area of research.

Another important application of the PS halo abundance is the *cluster-abundance normalization* of the power spectrum, which is still very much in vogue. Although the relation between a galaxy luminosity (which we measure) and the mass of its host halo (the abundance of which PS theory predicts) is extremely uncertain, the relation between the observed properties of galaxy clusters and their host halos is not quite so bad. In particular, the hot gas in a relaxed cluster can exist in hydrostatic equilibrium in the cluster's gravitational potential well, and the temperature of this gas can be measured quite well through the x-ray radiation it emits (and which we see), and the cluster mass thus determined. Such mass determinations agree fairly well with those based on dynamics of the resident galaxies, and also with the increasingly precise mass determinations from strong and weak gravitational lensing. Therefore, measurements of cluster abundance can be compared with the predictions of PS theory. The typical rich cluster with which such an analysis is carried out is $M \sim 10^{14.5}\, h^{-1}\, M_\odot$, and such clusters come from a comoving radius $R \sim 6.5\, \Omega_m^{-1/3}\, h^{-1}$ Mpc. By comparing the PS prediction for the cluster abundance with the observed abundance, one finds that the power-spectrum normalization (given here by σ_8 for the mass distribution) is $\sigma_8 \Omega_m^{0.56} \simeq 0.5 - 0.6$. Note that the Ω_m dependence here is due to the shape of $P(k)$ at $k^{-1} \sim 8\, h^{-1}$ Mpc; it has nothing to do with the linear-theory growth factor. The result is consistent with $\sigma_8 \simeq 1$ for the mass distribution and $\Omega_m \simeq 0.3$.

The PS formalism was developed in the early 1970s. Throughout the 1990s, N-body simulations of structure formation became increasingly precise, and were therefore able to provide increasingly precise tests of the very simple and naive PS theory. The basic conclusion is that the PS theory does remarkably well at predicting the mass function over a very wide variety of distance scales and for a wide variety of power spectra. As expected, however, the theory does become inaccurate at some point, and that point seems to have been reached now by numerical simulations. There has been a more complicated successor to PS developed by Sheth and Tormen that considers things like collapse of nonspherical overdensities, but ultimately calibrates several undetermined parameters to the simulations. Since the theory is calibrated to simulations, it provides better agreement with the simulations. Thus, for more precise numerical work, many authors now use the Sheth-Tormen mass function over the PS mass function.

Finally, keep in mind that the PS mass function is derived assuming that the distribution of primordial perturbations is Gaussian. If the primordial distribution is non-Gaussian, then the mass function will be different. It is straightforward to work this out.

2 Biasing

It has been observed empirically that the correlation functions for different types of galaxies are different, with more massive galaxies generally being more strongly clustered. The autocorrelation function for clusters is even stronger, and so is the autocorrelation function for radio sources and AGN. What this implies is that different tracers of the mass distribution are distributed differently with respect to each other, and therefore, that at most one type of galaxy can be distributed in the same way as the mass. In other words, the fractional density distribution of galaxies $\delta_g(\vec{x}) \neq \delta(\vec{x})$. This is the idea of biasing.

The differences in the clustering strengths of different types of galaxies can be understood theoretically—qualitatively, at least, if not quantitatively. To do so, we make a *peak-background split* as follows. Recall that $\delta(\vec{x})$ can be written as a sum over Fourier modes. We can therefore write $\delta = \delta_+ + \delta_s$, where δ_+ is the sum over the long-wavelength Fourier modes for δ and δ_s is the sum over the short-wavelength modes. We assume that $\delta_+ \ll 1$; i.e., the large-scale density contrast is small. We then remember that all peaks in the density field with $\delta > \delta_c \equiv \nu\sigma(M)$ will form bound objects. In an overdense region of the Universe, $\delta_+ > 0$. The requirement for $\delta > \delta_c$ therefore translates to the requirement $\delta_s > \nu\sigma - \delta_+$, where σ is here the variance of the mass distribution on scale of the mass of the tracer population (e.g., a galaxy scale). The probability for a given peak to be higher than $\nu\sigma$ (for $\nu \gg 1$), is $P(>\nu) \propto \nu^2 e^{-\nu^2/2}$, which is controlled by the exponential. Therefore, the probability to have a high-density peak in this overdense region is greater by a fractional amount,

$$\begin{aligned}
\frac{\delta P}{P} &= \frac{P(>\nu,\delta_+) - P(>\nu,\delta_+=0)}{P} \\
&\simeq \frac{\exp\left[-(\nu\sigma-\delta_+)^2/2\sigma^2\right] - \exp\left[-(\nu\sigma)^2/2\sigma^2\right]}{\exp(-\nu^2/2)} \\
&\simeq \nu\frac{\delta_+}{\sigma}. \qquad (6)
\end{aligned}$$

Therefore, the density perturbation for high-density peaks in the overdense region is $\delta_{\rm peak} \simeq (\nu/\sigma)\delta_+$ from which it follows that the autocorrelation function for high-density peaks is $\xi_{\rm peak}(r) \simeq (\nu/\sigma)^2\xi_{\rm mass}(r)$ or $\xi_{\rm peak}(r) \simeq \nu^2\xi(r)/\xi(0)$, where $\xi(0)$ is the zero-lag autocorrelation function evaluated on the scale of the tracer population. One must take care to not overinterpret this result. In particular, the approximations that are made are for realistic situations near the border of applicability, and the mapping between realistic tracers (e.g., galaxies or clusters) and high-density peaks is not precise. The moral of the story—that rarer objects should be more highly clustered—still holds. The other conclusion is that in the limit of weak clustering, $\xi \ll 1$, we expect the bias to be linear; i.e., for $\delta_{\rm tracer} \propto \delta$, independent of r.

It is instructive to get some rough numbers for galaxy clusters, which are very highly clustered. If we look at Abell richness class $R \geq 1$ clusters, their number density is $n \simeq 6\times10^{-6}\,h^3\,{\rm Mpc}^{-3}$. This corresponds for typical CDM models to $\nu \simeq 3$. The cluster autocorrelation function is observed

to be $\xi_{cc}(r) \simeq (r/r_0)^{-1.8}$ with $r_0 = 15 - 20\,h^{-1}$ Mpc, or $\xi_{cc} \simeq 10\,\xi_g g(r)$, roughly ten times the galaxy-galaxy autocorrelation function, which is consistent with $\nu \simeq 3$. Likewise, Lyman-break galaxies, are thought to be formed at high-density peaks at the redshifts $z \sim 3$ at which they are observed, and so their clustering is expected to also be very highly biased. Observations are consistent with this, although it is difficult to make precise statements given the unclear mapping between luminous objects and the dark-matter halos they occupy.

3 Limber approximation

In principle, the power spectrum $P(k)$ can be measured by constructing a three-dimensional distribution of the mass in the Universe by assuming that galaxies trace the mass (or are linearly biased with respect to it) and then doing a redshift survey to find the three-dimensional distribution of galaxies over a large volume of the Universe. One then either measures the correlation function, or one carries out the appropriate Fourier transforms to find the power spectrum.

If, however, one has a galaxy survey without redshifts, then one has only a distribution of galaxies on the two-dimensional surface of the sky, rather than the distribution in three dimensions. The *Limber approximation* (which might also be attributed to Vera Rubin) relates the two-dimensional angular correlation function to the three-dimensional correlation function, or equivalently, the two-dimensional angular power spectrum to $P(k)$. Suppose there is a three-dimensional density field $\delta(\vec{x})$ with power spectrum $P(k)$ and an observer at the origin sees a projection,

$$p(\boldsymbol{\theta}) = \int_0^\infty dr\, q(r) \delta(r\theta_1, r\theta_2, r), \tag{7}$$

where r is a distance along a line of sight near the $\hat{\mathbf{z}}$ direction, $\boldsymbol{\theta} = (\theta_1, \theta_2)$ is a position on the sky (near the north pole), and $q(r)$ is a weight function.

We want to calculate the two-dimensional power spectrum $P_p(\ell)$ for the projection $p(\theta)$. To do so, we note that the contribution to a shell of width Δr about r_0 is

$$\Delta p(\boldsymbol{\theta}) \simeq q(r_0) \int_{r_0 - \Delta r/2}^{r_0 + \Delta r/2} dr\, \delta(r_0\theta_1, r_0\theta_2, r). \tag{8}$$

The Fourier transform of $\Delta p(\boldsymbol{\theta})$ is the plane of the sky is then

$$\widetilde{\Delta p}(\boldsymbol{l}) \equiv \int d^2\theta\, \Delta p(\boldsymbol{\theta}) e^{-i\boldsymbol{l}\cdot\boldsymbol{\theta}} = q(r_0) \int \frac{d^3k}{(2\pi)^3} \tilde{\delta}(\vec{k}) \int d^2\theta\, e^{i(r_0\mathbf{k} - \boldsymbol{l})\cdot\boldsymbol{\theta}} \int_{r_0 - \Delta r/2}^{r_0 + \Delta r/2} dr\, e^{ik_3 r}, \tag{9}$$

where $\mathbf{k}$ are the components of $\vec{k}$ in the plane of the sky. The $d^2\theta$ integral is $(2\pi)^2 \delta_D(r_0\mathbf{k} - \boldsymbol{l})$, so

$$\widetilde{\Delta p}(\boldsymbol{l}) = \frac{\Delta r q(r_0)}{r_0^2} \int \frac{dk_3}{2\pi} \tilde{\delta}\left(\frac{l_1}{r_0}, \frac{l_2}{r_0}, k_3\right) j_0\left(\frac{k_3 \Delta r}{2}\right). \tag{10}$$

Multiplying two Fourier modes and taking the expectation value, we find

$$\left\langle \widetilde{\Delta p}(\boldsymbol{l}) \widetilde{\Delta p}(\boldsymbol{l}') \right\rangle = (2\pi)^2 \delta_D(\boldsymbol{l} - \boldsymbol{l}') \frac{(\Delta r)^2 q^2(r_0)}{r_0^2} \int dk_3\, P\left(\sqrt{\frac{\ell^2}{r_0^2} + k_3^2}\right) j_0^2\left(\frac{k_3 \Delta r}{2}\right). \tag{11}$$

The Dirac delta function states that $\Delta p(\boldsymbol{\theta})$ is statistically isotropic in the plane of the sky, a consequence of the statistical homogeneity of $\delta(\vec{x})$ in space, and the thing multiplying $(2\pi)^2\delta_D(\boldsymbol{l}-\boldsymbol{l}')$ is the contribution to the angular power spectrum C_ℓ from this Δr interval.

The next step is to carry out the dk_3 integral. To do so, we note that the spherical Bessel function here has width $\delta k_3 \sim 1/\Delta r$. We are also dealing with a small region of the sky near the north pole, so we are considering only planar Fourier modes with $\ell \gg 1$, and in particular, we take Δr to be large enough so that $\ell \gg r_0/\Delta r$. Therefore, $k_3 \ll \ell/r_0$ wherever j_0^2 is non-negligible. Physically, in the small-angle (large-ℓ) approximation in which we are working, only Fourier modes very nearly perpendicular to the line of sight contribute to the projected power (which makes sense), and so we can replace $P(\sqrt{\ell^2/r_0^2 + k_3^2})$ by $P(\ell/r_0)$. We then have

$$\left\langle \widetilde{\Delta p}(\boldsymbol{l})\widetilde{\Delta p}(\boldsymbol{l}')\right\rangle = (2\pi)^2\delta_D(\boldsymbol{l}-\boldsymbol{l}')\Delta r\frac{q^2(r_0)}{r_0^2}P\left(\frac{\ell}{r_0}\right). \tag{12}$$

The final step is to simply sum the contributions from each Δr over the line of sight:

$$C_\ell = \int_0^\infty dr\,\frac{q^2(r)}{r^2}P\left(\frac{\ell}{w}\right). \tag{13}$$

This is the desired result, the Limber approximation (in Fourier space) relating the two-dimensional angular power spectrum C_ℓ for a two-dimensional projection of $\delta(\vec{x})$ with weight function $q(r)$ to the three-dimensional power spectrum $P(k)$ for $\delta(\vec{x})$.

The two-dimensional angular correlation function $w(\theta)$ is then obtained from

$$w(\theta) = \int d^2\ell C_\ell e^{i\boldsymbol{l}\cdot\boldsymbol{\theta}}. \tag{14}$$

In most of the books, the derivation of the Limber equation is done in configuration space, and you are encouraged to check that out.

4 Redshift-space distortions

Galaxy surveys determine the distance to a galaxy by the redshift and then applying the Hubble law. However, density perturbations induce peculiar velocities, so the distance inference can get screwed up. The redshift $1+z$ will then have to be replaced by $(1+z)(1+\delta v/c)$, where δv is the component of the peculiar velocity along the line of sight. Remember, before moving on, that $\vec{x}(t) = a(t)\vec{r}(t)$ and the peculiar velocity $\delta\vec{v} = a(t)\vec{u}(t)$. We assume here that we are observing galaxies at small redshift, so $a \simeq 1$.

The apparent distance to a galaxy is $s_z = cz/H = v_H/H + \hat{r}\cdot\vec{\delta v}/H = r + \hat{r}\cdot\delta\vec{v}/H)$, where v_H is the Hubble-flow velocity, and r is the true distance. The apparent position is thus

$$\vec{s} = \vec{r}\left[1 + \frac{\hat{r}\cdot\delta\vec{v}(\vec{r})}{rH}\right], \tag{15}$$

from which it follows that the apparent and real differential volume elements are

$$d^3s = d^3r \left[1 + \frac{\hat{r}\cdot\delta\vec{v}(\vec{r})}{rH}\right]\left[1 + \frac{d}{dr}\frac{\hat{r}\cdot\delta\vec{v}(\vec{r})}{H}\right]. \tag{16}$$

The density field an always be decomposed into Fourier modes, so let's consider a single Fourier mode $\delta(\vec{r}) = \delta_0\cos(\vec{k}\cdot\vec{r}+\alpha)$, where α is a phase, and let us suppose that the angle between $\vec{k}$ and $\vec{r}$ is θ and let $\mu = \cos\theta$. We know that

$$\delta\vec{v}_{\vec{k}} = -\frac{iHf(\Omega_m)a}{k}\delta_{\vec{k}}\hat{k}, \tag{17}$$

so

$$\frac{\hat{r}\cdot\delta v(\vec{r})}{H} = -\frac{\mu\Omega_m^{0.6}\delta_0}{k}\sin(\vec{k}\cdot\vec{r}+\alpha). \tag{18}$$

The apparent density $\rho_s(\vec{r})$ will be related to the true density $\rho_r(\vec{r})$ by $\rho_s d^3s = \rho_r d^3r$, so for small-amplitude perturbations,

$$\delta_s(\vec{r}) = \delta_r(\vec{r}) - 2\left(\frac{\hat{r}\cdot\delta\vec{v}(\vec{r})}{rH}\right) - \frac{d}{dr}\frac{\hat{r}\cdot\delta\vec{c}(\vec{r})}{H}. \tag{19}$$

To go further, we note that $(\hat{r}\cdot\delta\vec{c}/H) \sim \delta_0/k$ while $(d/dr)(\hat{r}\cdot\delta\vec{c}/H) \sim \delta_0$. Moreover, the survey probes wavenumbers $k \gg r_{\rm max}^{-1}$, where $r_{\rm max}$ is the depth of the survey, and so the second term is negligible compared with the third term in the equation above. We then use $(d/dr)(\hat{r}\cdot\delta\vec{c}/H) = -\mu^2\Omega_m^{0.6}\delta(\vec{r})$ to conclude that in redshift space, the density contrast is

$$\delta_s(\vec{r}) = \delta_r(\vec{r})(1+\mu^2\Omega_m^{0.6}). \tag{20}$$

However, it is the matter that gives rise to the peculiar velocity, so this $\delta_s(\vec{r})$ is the *mass* perturbation in redshift space, while what is seen is the distribution of luminous matter, or galaxies. We can thus write $\delta_g = b\delta = \delta + (b-1)\delta$, where the first term gives rise to peculiar velocities, while the second does not. Then, in redshift space, the density perturbation for galaxies will be

$$\delta_{s,g} = \delta_r(1+\mu^2\Omega_m^{0.6}) + (b-1)\delta_r = \delta_{r,g}\left(1 + \frac{\mu^2\Omega_m^{0.6}}{b}\right). \tag{21}$$

Therefore, the anisotropy in the clustering as a function of $r_\parallel$ inferred from the clustering in redshift space and clustering as a function of $r_\perp$ inferred from angular clustering can determine $\beta = \Omega_m^{0.6}/b$. From the data, a value of $\beta \simeq 0.5$ is obtained.

All of the analysis until now has been linear; i.e., applicable to large-scale density perturbations and large-scale peculiar-velocity fields. In linear theory, redshift-space distortions tend to compress structures along the line of sight. Consider a spherical overdensity in the linear regime. The side closest to us will be moving away from us at a slightly larger velocity than the Hubble flow, while the more distance edge moves away from us with a velocity slightly smaller than the Hubble flow. However, if we are looking at a gravitationally collapsed virialized object (e.g., a cluster), structure is spread out in redshift space (these are called "fingers of God") because of the virial velocity spread. The line-of-sight velocity distribution can be modeled as a one-dimensional Gaussian distribution of velocity dispersion σ, and this essentially smooths structure on scales $k \gtrsim \sigma$: i.e., $\delta_k \rightarrow \delta_k e^{-k^2\mu^2\sigma^2/2}$. If we alternatively model the distribution as an exponential (rather than

Gaussian), then $\delta_k \to \delta_k(1 + k^2\mu^2\sigma^2/2)^{-1}$. If we average over the entire Universe (rather than consider just a cluster), then $\sigma = \sigma_p/\sqrt{2}$, where $\sigma_0 \simeq 300 - 400$ km/sec is the *pairwise velocity dispersion* at $1\,h^{-1}$ Mpc. This is the rms line-of-sight velocity difference between pairs of galaxies separated by $1\,h^{-1}$ Mpc. We thus expect the power on measured in redshift space on scales

$$k \gtrsim \sqrt{2}\frac{3\times10^5\,\mathrm{km/sec}}{300\,\mathrm{km/sec}}\frac{h}{6000\,\mathrm{Mpc}} \simeq 0.3\,h\,\mathrm{Mpc}^{-1}, \tag{22}$$

to be suppressed by redshift-space smearing.

5 Zeldovich approximation

Earlier we considered the collapse of a spherical overdensity. Needless to say, overdensities will never be precisely spherical, and so they will not necessarily collapse to a point. The *Zeldovich approximation* allows us to see what happens with a non-spherical perturbation. In this approach, we define the proper (or Eulerian) coordinate $\vec{x}(t) = a(t)\vec{q} + b(t)\vec{f}(\vec{q})$, where $\vec{q}$ is the comoving coordinate (or Lagrangian coordinate) at time $t = 0$, well before the perturbation-amplitude has grown, and $\vec{f}(\vec{q})$ is the "displacement field," which tells us where the matter originally at $\vec{q}$ has moved. The Lagrangian coordinate $\vec{q}$ for each particle remains constant. Therefore, the mass density ρ_0 in $\vec{q}$ space is constant. The density in $\vec{x}$ space is then given by the Jacobian of the transformation between $\vec{q}$ and $\vec{x}$ space. Thus, the physical density (i.e., in $\vec{x}$ space) is

$$\rho = \rho_0\left[\left(1-\frac{b}{a}\alpha\right)\left(1-\frac{b}{a}\beta\right)\left(1-\frac{b}{a}\gamma\right)\right]^{-1}, \tag{23}$$

where $(-\alpha, -\beta, -\gamma)$ are the eigenvalues of $\partial f_i/\partial q_j$, the "strain" or "deformation" tensor. If $\vec{f}(\vec{q})$ originates from a growing mode, then it is irrotational, meaning that $\vec{f}(\vec{q}) = \vec{\nabla}\psi(\vec{q})$, and $(\partial f_i/\partial q_j) = (\partial^2\psi/\partial q_i\partial q_j)$.

Linearizing equation (23), $\delta = -(b/a)(\alpha+\beta+\gamma) = -(b/a)\vec{\nabla}\cdot\vec{f}$. Since the eigenvalues are equal to $\partial f_i/\partial q_j$ in a coordinate frame aligned with the principal axes of $\partial f_i/\partial q_j$, we find that the scaled peculiar velocity is

$$\vec{u} \equiv \frac{1}{a}\left(\dot{\vec{x}} - \frac{\dot{a}}{a}\vec{x}\right) = \left(\frac{\dot{b}}{a} - \frac{\dot{a}b}{a^2}\right)\vec{f}, \tag{24}$$

which, incidentally, satisfies $\dot{\delta} = -\vec{\nabla}\cdot\vec{u}$. Note that the time dependence $\delta(t) \propto b/a$. Thus, for example, in an EdS universe, $\delta \propto t^{2/3}$ implies $b(t) \propto t^{2/3}$. More generally, $b(t)/a(t) \propto D(t)$, where $D(t)$ is the linear-theory growth factor: $D(t) \equiv \delta(t)/\delta(t_{\rm today})$, so the Zeldovich approximation is

$$\vec{x}(t) = a(t)\left[\vec{q} + D(t)\vec{f}(\vec{q})\right]. \tag{25}$$

This is first-order Lagrangian perturbation theory, as opposed to first-order Eulerian perturbation theory, which we did earlier. Lagrangian perturbation theory is found to be more accurate than Eulerian perturbation theory essentially because it works until turnaround.

From the expression for ρ/ρ_0, it can be seen that collapse occurs first along the shortest axis. If we consider some peak in the initial three-dimensional density distribution, then $\vec{f}(\vec{q}) = 0$ at that

point, while $\partial\vec{f}/\partial\vec{q}$ describes isodensity ellipsoids around the peak. The density $\rho \to \infty$ at the time when $b(t)/a(t) = [\max(\alpha, \beta, \gamma)]^{-1}$. The axis corresponding to the largest eigenvalue of $\partial\vec{f}/\partial\vec{q}$ is the shortest axis. We thus see that since primordial overdensities will generically be nonspherical, gravitational collapse results in a *Zeldovich pancake*. For a realistic overdensity, with baryons and dark matter, the baryons will shock when the overdensity collapses to a sheet, and the dark matter will presumably somehow virialize.

6 The three-point correlation function, bispectrum, and non-Gaussianity

Inflation predicts that *primordial* perturbations are extremely close to Gaussian. What this means is that the odd-point correlation functions (3-pt, 5-pt,....) all vanish: $\langle\delta_1\delta_2\delta_3\rangle = \langle\delta_1\delta_2\delta_3\delta_4\delta_5\rangle = \cdots = 0$, where $\delta_i = \delta(\vec{x}_i)$. It also means that the even-point correlation functions can all be written in terms of the two-point correlation function. For example, the four-point correlation function is $\langle\delta_1\delta_2\delta_3\delta_4\rangle = \langle\delta_1\delta_2\rangle\langle\delta_3\delta_4\rangle + \langle\delta_1\delta_3\rangle\langle\delta_2\delta_4\rangle + \langle\delta_1\delta_4\rangle\langle delta a_2\delta_3\rangle$. The zero-lag n-point correlation functions are all determined by the probability distribution, $P(\delta) \propto e^{-\delta^2/2\sigma^2}$. Thus, $\left\langle\delta^2\right\rangle = \sigma^2$, $\left\langle\delta^4\right\rangle = 3\sigma^4$, $\left\langle\delta^6\right\rangle = 15\sigma^6$, etc. And, of course, $\langle\delta^n\rangle = 0$ for n odd.

Now it is important to realize that it is only the *primordial* perturbations that are Gaussian. In linear perturbation theory, the perturbations also remain Gaussian. This can be seen by noting that $\delta(\vec{x}, t) \propto D(t)$ in linear theory, so the spatial distribution is unchanged. However, if we go beyond linear theory, then it must be true that gravitational amplification of perturbations will induce non-Gaussianity. The simplest way to see this is to note that the one-point probability distribution $P(\delta)$ (which is initially Gaussian) at late times will become non-Gaussian. A Gaussian distribution is symmetric about $\delta = 0$. However, when structures go nonlinear, there will be regions of the Universe that have very large density contrasts; for example, at virialization, a virialized object will have a density contrast $\delta \simeq 200$. However, $\delta = \delta\rho/\bar{\rho}$ is bounded from below by -1 (since $\delta\rho$ can never be larger in magnitude than $\bar{\rho}$). Thus, nonlinear evolution of the mass distribution must result in a non Gaussian density distribution.

If we go to second order in perturbation theory (i.e., solve the Euler, continuity, and Poisson equation to second order in small perturbations), then a non-zero three-point correlation function ($\langle\delta_1\delta_2\delta_3\rangle \neq 0$) arises. Just as the power spectrum is the two-point correlation function in Fourier space, one can define a *bispectrum*, which is the Fourier-space three-point correlation function, as follows:

$$B(\vec{k}_1, \vec{k}_2, \vec{k}_3) = \left\langle\tilde{\delta}(\vec{k}_1)\tilde{\delta}(\vec{k}_2)\tilde{\delta}(\vec{k}_3)\right\rangle. \tag{26}$$

A nightmarishly tedious calculation shows that in second order in perturbation theory,

$$B(\vec{k}_1, \vec{k}_2, \vec{k}_3) = 2\delta_D(\vec{k}_1 + \vec{k}_2 + \vec{k}_3)\left[P(k_1)P(k_2)P_n^{(}2)(\vec{k}_1, \vec{k}_2) + \text{cyclicpermutations}\right], \tag{27}$$

where

$$P_n^{(}2)(\vec{k}_1, \vec{k}_2) = \frac{5}{7} + \frac{\vec{k}_1 \cdot \vec{k}_2}{2k_1k_2}\left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right) + \frac{2}{7}\frac{(\vec{k}_1 \cdot \vec{k}_2)^2}{k_1^2k_2^2}. \tag{28}$$

Strictly speaking, this result is only valid in an EdS universe, but the more general result isn't too much different. The three-point correlation function can then be obtained from the bispectrum by simply evaluating $\langle\delta(\vec{x}_1)\delta(\vec{x}_2)\delta(\vec{x}_3)\rangle$. Rather than carry this out in full generality, we simply

point out that the zero-lag three-point correlation function, $\left\langle \delta^3(\vec{x}) \right\rangle$ is related to the skewness $S \equiv \left\langle \delta^3 \right\rangle / \left\langle \delta^2 \right\rangle^2$. To second order in perturbation theory, $S = 34/7$ (again in a EdS universe, although the result is pretty much the same with nonzero Λ). To third order in perturbation theory, a nonzero *kurtosis* $K \equiv [\left\langle \delta^4 \right\rangle - 3 \left\langle \delta^2 \right\rangle^2] / \left\langle \delta^2 \right\rangle^3$ arises. Calculation of this quantity is one of the few examples where pencil-and-paper theorists in cosmology actually earn their living.

7 Nonlinear clustering

When the density-perturbation amplitude grows to $\left\langle \delta^2 \right\rangle \sim 1$, linear theory breaks down. One consequence is that the matter distribution becomes non-Gaussian. Another consequence, however, is that the shape (i.e., the k dependence) of the power spectrum $P(k)$, which remains unaltered in linear theory, gets changed by nonlinear evolution. Nonlinear evolution of the power spectrum is extremely complicated. In the *quasi-linear* regime, $r \gtrsim 8\,h^{-1}$ Mpc, where $\sigma^2(M) \lesssim 1$, higher-order perturbative calculations, such as those that give the bispectrum, can be used to determine the change in the power spectrum. However, the galaxy power spectrum is best determined on smaller scales. How do we relate the predictions of the highly nonlinear theory with measurements at these scales? One option is to decide that its too complicated and ignore these measurements. Another option (probably the best idea) is to determine the nonlinear evolution of the power spectrum with *N-body simulations*, where the primordial mass distribution is modeled by a huge number N of point masses and their trajectories then simply determined by Newton's laws in an expanding spacetime.

Then there are analytic techniques. One uses the *stable-clustering hypothesis* which says that once objects have undergone gravitational collapse, they are no longer affected by the expansion of the Universe, and their internal structures remain fixed. If so, then $\xi(r_{\rm proper})$, where $r_{\rm proper}$ is the physical distance is constant in time on small scales. This small-scale behavior is then sewn onto the large-scale linear-theory correlation function through some recipe. Another, more recent but related, technique is the *halo approach* to clustering. In this approach, one notes that numerical simulations seem to indicate that virialized dark-matter halos assume a *Navarro-Frenk-White (NFW)* density profile,

$$\rho(r) \propto \frac{\rho_0}{(r/r_c)(1 + r/r_c)^2}, \tag{29}$$

where ρ_0 and r_0 parameterize the halo density and core radius. The halo approach then notes that these halos undergo biased clustering, with a bias parameter that increases with halo mass. Analytic expressions for the two-point correlation function and power spectrum for the mass can then be obtained.

The parameters in these approaches are then determined by fitting to simulations, and so they should be seen as theory-inspired semi-analytic fits to simulations, rather than "first-principles" theory. Still, they can be quite handy, especially is someone else has done the simulations with which they are calibrated. Qualitatively, nonlinear evolution *increases* the power on small scales. It is said that nonlinear evolution moves power from large scales to small scales. Although this statement is difficult to quantify precisely, it is easy to see how this occurs. Consider a single sinusoidal primordial density perturbation $\delta(\vec{x}, t_i) = \delta_i \sin(\vec{k} \cdot \vec{x})$, with an initial amplitude $\delta_i \ll 1$. In linear theory (as long as $\delta \ll 1$), the spatial dependence remains sinusoidal. When $\delta \sim 1$,

however, the spatial distribution will still be periodic, although not necessarily sinusoidal. The true density distribution will include higher harmonics: $\delta(\vec{x}) \propto \sum_n A_n \sin(n\vec{k}\cdot\vec{x})$. In this way, power is introduced on smaller scales.

8 Characterization of CMB fluctuations

We measure the temperature $T(\hat{n})$ of the CMB as a function of position $\hat{n} = (\theta, \phi)$ on the sky. If we restrict our attention to a patch of sky sufficiently small to be considered flat, then we can approximate the coordinates on this small patch by Cartesian angular coordinates θ_x and θ_y. The temperature map $T(\theta_x, \theta_y)$ can then be represented equivalently in terms of its Fourier components,

$$T(\boldsymbol{\ell}) = \int \frac{d^2\boldsymbol{\theta}}{(2\pi)^2} e^{i\boldsymbol{\ell}\cdot\boldsymbol{\theta}} T(\boldsymbol{\theta}), \tag{30}$$

where $\boldsymbol{\theta} \equiv (\theta_x, \theta_y)$.

The full-sky temperature map is analogously expanded in terms of spherical harmonics $Y_{\ell m}(\hat{n})$ through,

$$T(\hat{n}) = \sum_{\ell m} a^T_{\ell m} Y_{\ell m}(\hat{n}) \tag{31}$$

and

$$a^T_{\ell m} = \int d\hat{n}\, T(\hat{n}) Y^*_{\ell m}(\hat{n}), \tag{32}$$

since spherical harmonics constitute a complete orthonormal basis for scalar functions on the two-sphere.

We also measure the CMB linear polarization as a function of position on the sky. The linear polarization is specified in terms of Stokes parameters Q and U. These, however, are coordinate-dependent quantities. Suppose that they are measured with respect to some x-y axes and that we then consider some other x'-y' axes rotated by an angle α with respect to the x-y axes. Under this rotation, the Stokes parameters (Q, U) transform as components of a symmetric trace-free (STF) 2×2 tensor,

$$\begin{pmatrix} Q & U \\ U & -Q \end{pmatrix} \Rightarrow \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} Q & U \\ U & -Q \end{pmatrix} \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}. \tag{33}$$

Alternatively and equivalently, if we represent the polarization by a complex number $P = Q + iU$, then $P \to Pe^{2i\alpha}$ under a rotation of the coordinate axes by an angle α; i.e., the polarization is a spin-2 field.

Anything we say about Stokes parameters Q and U are thus tied to the coordinate system we choose. We will therefore want to find a coordinate-system–independent representation of this tensor field if we are to make statements about physics that are independent of coordinate system. Later, we will do this on the full sky, but we first do the simpler case of a flat sky (which also serves as a good approximation to a small region of the sky).

8.1 Harmonic analysis on a flat sky

Once the polarization, $Q(\boldsymbol{\theta})$ and $U(\boldsymbol{\theta})$, has been measured as a function of position $\boldsymbol{\theta} = (\theta_x, \theta_y)$ on a flat region of sky, we have measured the polarization tensor field,

$$\mathcal{P}_{ab} = \frac{1}{\sqrt{2}} \begin{pmatrix} Q(\boldsymbol{\theta}) & U(\boldsymbol{\theta}) \\ U(\boldsymbol{\theta}) & -Q(\boldsymbol{\theta}) \end{pmatrix}, \tag{34}$$

where the normalization is chosen so that $\mathcal{P}^{ab}\mathcal{P}_{ab} = Q^2 + U^2$.

We now define gradient ("E modes") and curl ("B modes") components of the tensor field that are independent of the orientation of the x-y axes by

$$\nabla^2 E = \partial_a \partial_b \mathcal{P}_{ab}, \qquad \nabla^2 B = \epsilon_{ac} \partial_b \partial_c \mathcal{P}_{ab}, \tag{35}$$

where ϵ_{ab} is the antisymmetric tensor.

Writing,

$$\mathcal{P}_{ab}(\boldsymbol{\theta}) = \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} \tilde{\mathcal{P}}_{ab}(\boldsymbol{\ell}) e^{-i\boldsymbol{\ell}\cdot\boldsymbol{\theta}}, \qquad \tilde{\mathcal{P}}_{ab}(\boldsymbol{\ell}) = \int d^2\boldsymbol{\theta} \mathcal{P}_{ab}(\boldsymbol{\theta}) e^{i\boldsymbol{\ell}\cdot\boldsymbol{\theta}}, \tag{36}$$

the Fourier components of $E(\boldsymbol{\theta})$ and $B(\boldsymbol{\theta})$ are

$$\begin{pmatrix} \tilde{E}(\boldsymbol{\ell}) \\ \tilde{B}(\boldsymbol{\ell}) \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \cos 2\varphi_{\boldsymbol{\ell}} & \sin 2\varphi_{\boldsymbol{\ell}} \\ -\sin 2\varphi_{\boldsymbol{\ell}} & \cos 2\varphi_{\boldsymbol{\ell}} \end{pmatrix} \begin{pmatrix} \tilde{Q}(\boldsymbol{\ell}) \\ \tilde{U}(\boldsymbol{\ell}) \end{pmatrix}, \tag{37}$$

where $\varphi_{\boldsymbol{\ell}}$ is the angle $\boldsymbol{\ell}$ makes with the $\hat{x}$ axis. This relation can be inverted,

$$\begin{pmatrix} \tilde{Q}(\boldsymbol{\ell}) \\ \tilde{U}(\boldsymbol{\ell}) \end{pmatrix} = \sqrt{2} \begin{pmatrix} \cos 2\varphi_{\boldsymbol{\ell}} & -\sin 2\varphi_{\boldsymbol{\ell}} \\ \sin 2\varphi_{\boldsymbol{\ell}} & \cos 2\varphi_{\boldsymbol{\ell}} \end{pmatrix} \begin{pmatrix} E(\boldsymbol{\ell}) \\ B(\boldsymbol{\ell}) \end{pmatrix}. \tag{38}$$

Thus, for a pure B mode in the $\hat{x}$ direction ($\varphi_{\boldsymbol{\ell}} = 0$), we have (as shown in the right panel of **Figure 1**) $\tilde{Q}(\boldsymbol{\ell}) = 0$ and $\tilde{U}(\boldsymbol{\ell}) = \tilde{B}(\boldsymbol{\ell})$. For a pure E mode in the $\hat{x}$ direction, we have (as shown in the left panel of **Figure 1**) $\tilde{Q}(\boldsymbol{\ell}) = \tilde{E}(\boldsymbol{\ell})$ and $\tilde{U}(\boldsymbol{\ell}) = 0$. Thus, *in an E mode, the polarization varies parallel/perpendicular to the direction of the Fourier mode, while for a B mode the polarization varies along directions* 45° *with respect to the direction of the Fourier mode.*

Since the combined temperature/polarization map is described by three sets, $\tilde{T}(\boldsymbol{\ell})$, $\tilde{E}(\boldsymbol{\ell})$, and $\tilde{B}(\boldsymbol{\ell})$, of Fourier components, the two-point statistics of the temperature/polarization field are determined by a total of six power spectra, $C_\ell^{\mathrm{X}_1,\mathrm{X}_2}$, defined by

$$\left\langle \tilde{X}_1(\boldsymbol{\ell}) \tilde{X}_2(\boldsymbol{\ell'}) \right\rangle = (2\pi)^2 \delta(\boldsymbol{\ell} + \boldsymbol{\ell'}) C_\ell^{X_1 X_2}, \tag{39}$$

where $\mathrm{X}_1, \mathrm{X}_2 = \{\mathrm{T}, \mathrm{E}, \mathrm{B}\}$. Here the angle brackets denote an average over all realizations of the temperature map.

Now suppose we have a given temperature/polarization map and then consider a parity inversion; e.g., a reflection about the x-axis. Then

$$\theta_y \to -\theta_y, \quad Q \to Q, \quad U \to -U, \quad \ell_x \to \ell_x, \quad \ell_y \to -\ell_y. \tag{40}$$

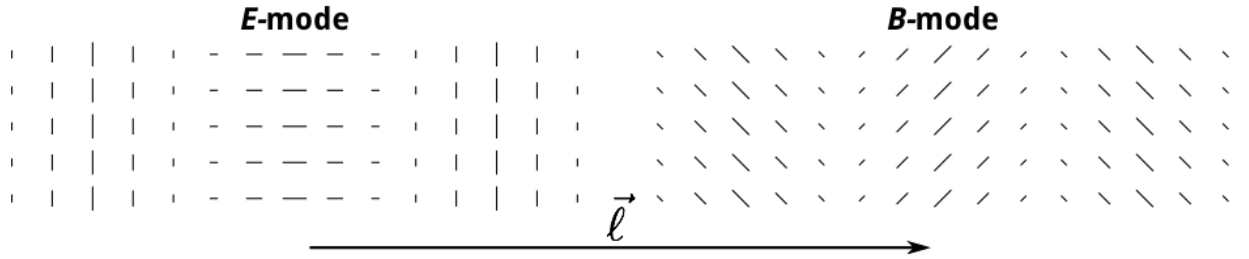

Figure 1: Shown are the polarization pattern associated with a single E mode (left) and a single B mode (right) with a horizontal wavevector $\boldsymbol{\ell}$. The E mode features a variation of the polarization along directions parallel/perpendicular to the direction of $\boldsymbol{\ell}$ (Stokes parameter Q in a coordinate system aligned with $\boldsymbol{\ell}$), while in the B mode the variation in the polarization is along directions 45° with respect to $\boldsymbol{\ell}$ (Stokes parameter U in coordinates aligned with $\boldsymbol{\ell}$). (From C. Bischoff.)

Also,

$$\tilde{T}(\boldsymbol{\ell}) \to \tilde{T}(\boldsymbol{\ell}), \quad \tilde{E}(\boldsymbol{\ell}) \to \tilde{E}(\boldsymbol{\ell}), \quad \tilde{B}(\boldsymbol{\ell}) \to -\tilde{B}(\boldsymbol{\ell}). \tag{41}$$

Thus, T and E have the same parity, while B has the opposite parity. If the physics that gives rise to temperature/polarization fluctuations is parity conserving, we then expect $C_\ell^{\rm TB} = C_\ell^{\rm EB} = 0$. In this case, the statistics of the temperature/polarization map are determined entirely by the four power spectra, $C_\ell^{\rm TT}$, $C_\ell^{\rm TE}$, $C_\ell^{\rm EE}$, and $C_\ell^{\rm BB}$.

8.2 Harmonic analysis on the full sky

If our maps extend beyond a small region of the sky, we will have to deal with the curvature of the sky. We thus generalize the tensor Fourier analysis that we carried out above for STF 2×2 tensors to tensors that live on the 2-sphere. In the usual spherical polar coordinates θ, ϕ, the sphere has a metric, $g_{ab} = {\rm diag}(1, \sin^2\theta)$. The polarization tensor $\mathcal{P}_{ab}$ must be symmetric $\mathcal{P}_{ab} = \mathcal{P}_{ba}$ and trace-free $g^{ab}\mathcal{P}_{ab} = 0$, from which it follows that,

$$\mathcal{P}_{ab}(\hat{n}) = \frac{1}{\sqrt{2}} \begin{pmatrix} Q(\hat{n}) & U(\hat{n})\sin\theta \\ U(\hat{n})\sin\theta & -Q(\hat{n})\sin^2\theta \end{pmatrix}, \tag{42}$$

where the factors of $\sin\theta$ follow from the fact that the coordinate basis (θ, ϕ) is orthogonal but not orthonormal.

We use a colon (:) to denote a covariant derivative on the surface of the sphere (e.g., $S^a{}_{:a}$ denotes the divergence of S^a) and a comma to denote a partial derivative [e.g., $S_{,a} = (\partial S/\partial x^\alpha)$]. Appendix A of arXix:astro-ph/9611125 reviews the rules of differential geometry on the sphere in the notation we use here.

Any STF 2×2 tensor field on the sphere can be written as the 'gradient', $E_{:ab} - \frac{1}{2} g_{ab} E^{:c}_{c}$, of some scalar field $E(\theta, \phi)$, plus the 'curl,' $(1/2)\,(B_{:ac}\epsilon^c_b + B_{:bc}\epsilon^c_a)$, of some other scalar field $B(\theta, \phi)$, For comparison, a vector field is analogously decomposed as $V_a = \nabla_a E + \epsilon_{ab}\nabla_b B$. Since any scalar field

on the sphere can be expanded in spherical harmonics (e.g. for the temperature),

$$\frac{T(\hat{n})}{T_0} = 1 + \sum_{\ell=1}^{\infty} \sum_{m=-\ell}^{\ell} a^{\rm T}_{\ell m}\, Y_{\ell m}(\hat{n}), \qquad \text{where} \qquad a^{\rm T}_{\ell m} = \frac{1}{T_0} \int d\hat{n}\, T(\hat{n}) Y^*_{\ell m}(\hat{n}), \tag{43}$$

it follows that the polarization tensor can be expanded in terms of basis functions that are gradients and curls of spherical harmonics,

$$\mathcal{P}_{ab}(\hat{n}) = T_0 \sum_{\ell=2}^{\infty} \sum_{m=-\ell}^{\ell} \left[a^{\rm E}_{\ell m} Y^{\rm E}_{(\ell m)ab}(\hat{n}) + a^{\rm B}_{\ell m} Y^{\rm B}_{(\ell m)ab}(\hat{n}) \right]. \tag{44}$$

The expansion coefficients are given by

$$a^{\rm E}_{\ell m} = \frac{1}{T_0} \int d\hat{n} \mathcal{P}_{ab}(\hat{n}) Y^{{\rm E}\,ab\,*}_{(\ell m)}(\hat{n}), \qquad a^{\rm B}_{\ell m} = \frac{1}{T_0} \int d\hat{n}\, \mathcal{P}_{ab}(\hat{n}) Y^{{\rm B}\,ab\,*}_{(\ell m)}(\hat{n}), \tag{45}$$

and

$$Y^{\rm E}_{(\ell m)ab} = N_\ell \left(Y_{(\ell m):ab} - \frac{1}{2} g_{ab} Y_{(\ell m):c}{}^{c} \right), \quad Y^{\rm B}_{(\ell m)ab} = \frac{N_\ell}{2} \left(Y_{(\ell m):ac} \epsilon^c{}_b + Y_{(\ell m):bc} \epsilon^c{}_a \right), \tag{46}$$

constitute a complete orthonormal set of basis functions for the E and B components of the polarization. The quantity, $N_\ell \equiv \sqrt{2(l-2)!/(l+2)!}$, is a normalization factor chosen so that

$$\int d\hat{n}\, Y^{{\rm X}\,*}_{(\ell m)ab}(\hat{n})\, Y^{{\rm X}'\,ab}_{(l'm')}(\hat{n}) = \delta_{\ell\ell'}\delta_{mm'}, \tag{47}$$

for XX′ =EE, EB, and BB. Also, we can integrate by parts to write alternatively,

$$a^{\rm E}_{\ell m} = \frac{N_\ell}{T_0} \int d\hat{n}\, Y^*_{\ell m}(\hat{n})\, \mathcal{P}_{ab}{}^{:ab}(\hat{n}), \qquad a^{\rm B}_{\ell m} = \frac{N_\ell}{T_0} \int d\hat{n}\, Y^*_{(\ell m}(\hat{n})\, \mathcal{P}_{ab}{}^{:ac}(\hat{n}) \epsilon_c{}^b. \tag{48}$$

Finally, since T, Q, and U are real, we get $a^{{\rm X}\,*}_{\ell m} = (-1)^m a^{\rm X}_{\ell,-m}$, where ${\rm X} = \{{\rm T}, {\rm E}, {\rm B}\}$. The temperature/polarization power spectra are now

$$\left\langle a^{{\rm X}\,*}_{\ell m} a^{\rm X'}_{\ell' m'} \right\rangle = C^{\rm XX'}_\ell \delta_{\ell\ell'}\delta_{mm'}, \tag{49}$$

for XX′ =TT, EE, BB, TE, TB, and EB. The C_ℓ here reduce in the small-angle (large-ℓ) limit with those in Section 8.1 as long as the angles in the flat-sky limit are given in radians.

The $Y^{\rm E}_{(\ell m)ab}$ and $Y^{\rm B}_{(\ell m)ab}$ are explicitly given by

$$Y^{\rm E}_{(\ell m)ab} = \frac{N_\ell}{2} \begin{pmatrix} W_{\ell m} & X_{\ell m} \sin\theta \\ X_{\ell m} \sin\theta & -W_{\ell m} \sin^2\theta \end{pmatrix}, \quad Y^{\rm B}_{(\ell m)ab} = \frac{N_\ell}{2} \begin{pmatrix} -X_{\ell m} & W_{\ell m} \sin\theta \\ W_{\ell m} \sin\theta & X_{\ell m} \sin^2\theta \end{pmatrix}, \tag{50}$$

where

$$W_{\ell m}(\hat{n}) \pm i X_{\ell m}(\hat{n}) = \sqrt{\frac{(l+2)!}{(l-2)!}}\, {}_{\pm 2}Y_{\ell m}(\hat{n}), \tag{51}$$

in terms of the spin-2 harmonics ${}_{\pm 2}Y_{\ell m}$. If we replace (Q, U) by $(U, -Q)$, then E $\to -$B and B $\to$ E. This tells us therefore, that a pure-E polarization pattern becomes a pure-B pattern if we rotate each polarization vector by 45°, and *vice versa*, as can be also inferred from the flat-sky treatment. Examples of E and B type polarization patterns are shown in **Figure 2**. The parity properties of T, E, and B found in the flat-sky treatment remain valid on the full sky.

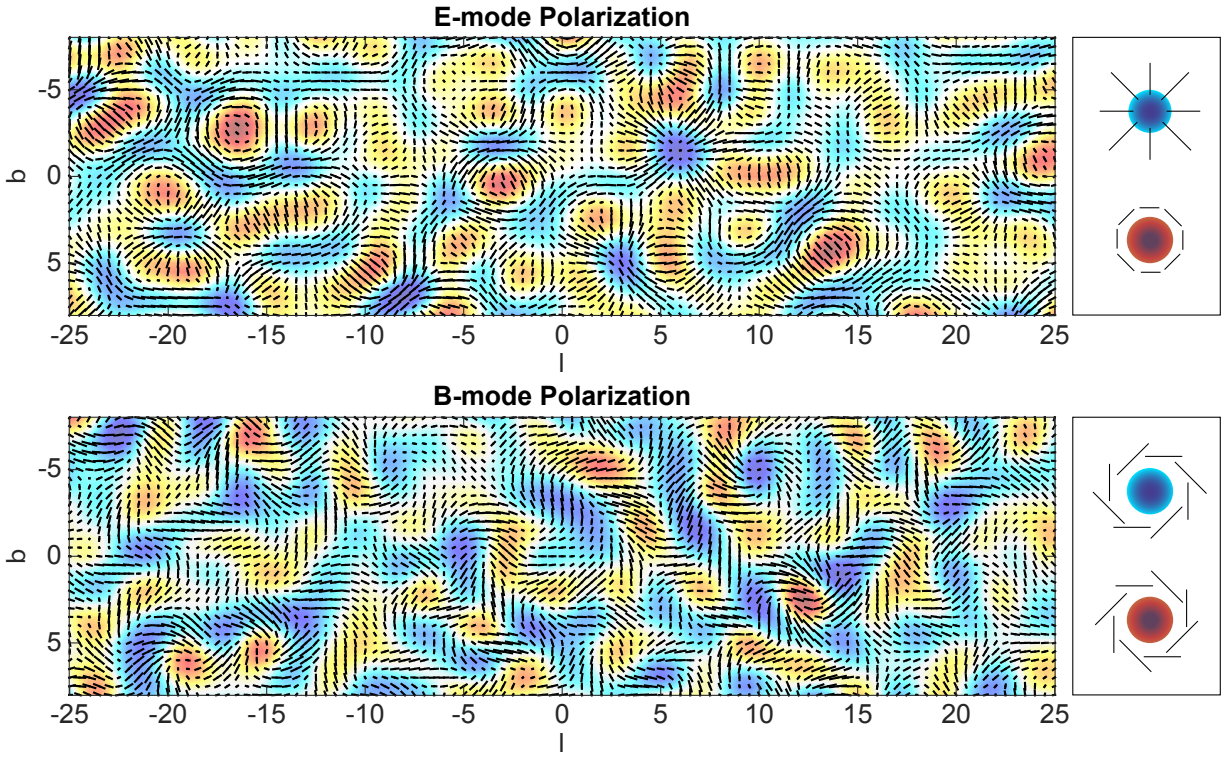

Figure 2: In the top figure we show a polarization pattern composed only of E modes and in the bottom one composed only of B modes. As indicated on the right, it is seen that around hot spots (red) the polarization pattern of the E mode is tangential and radial around cold spots (blue). The polarization pattern surrounding hot and cold spots of the B mode show a characteristic swirling pattern (with different orientation around hot and cold spots).