Title
Multiscale multichroic focal planes for measurements of the cosmic microwave background

Author
Cukierman, Ariel Jozef

Publication Date
2018

Peer reviewed|Thesis/dissertation

Multiscale multichroic focal planes for measurements of the cosmic microwave background

by

Ariel Jozef Cukierman

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge:

Professor Adrian T. Lee
Professor William Holzapfel
Professor Aaron Parsons

Spring 2018

Multiscale multichroic focal planes for measurements of the cosmic microwave background

Abstract

Multiscale multichroic focal planes for measurements of the cosmic microwave background

by

Ariel Jozef Cukierman

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Adrian T. Lee, Chair

We report on the development of multiscale multichroic focal planes for measurements of the cosmic microwave background (CMB). A multichroic focal plane, i.e., one that consists of pixels that are simultaneously sensitive in multiple frequency bands, is an efficient architecture for increasing the sensitivity of an experiment as well as for disentangling the contamination due to galactic foregrounds, which is increasingly becoming the limiting factor in extracting cosmological information from CMB measurements. To achieve these goals, it is necessary to observe across a broad frequency range spanning roughly 30-350 GHz. Depending on the foreground complexity, it may become necessary to observe across an even larger range. For this purpose, the Berkeley CMB group has been developing multichroic pixels consisting of planar superconducting sinuous antennas coupled to extended hemispherical lenslets, which operate at sub-Kelvin temperatures. The sinuous antennas, microwave circuitry and the transition-edge-sensor (TES) bolometers to which they are coupled are integrated in a single lithographed wafer.

We describe the design, fabrication, testing and performance of multichroic pixels with bandwidths of 3:1 and 4:1 across the entire frequency range of interest. Additionally, we report on a demonstration of multiscale pixels, i.e., pixels whose effective size changes as a function of frequency. This property keeps the beam width approximately constant across all frequencies, which in turn allows the sensitivity of the experiment to be optimal in every frequency band. We achieve this by creating phased arrays from neighboring lenslet-coupled sinuous antennas, where the size of each phased array is chosen independently for each frequency band. We describe the microwave circuitry in detail as well as the additional benefits of a multiscale architecture, e.g., mitigation of beam non-idealities, reduced readout requirements and polarization-wobble cancellation. Finally, we discuss the design and fabrication of the detector modules and focal-plane structures including cryogenic readout components, which enable the integration of our devices in current and future CMB experiments.

To my very cool and totally normal parents, who told me I was allowed to do whatever I want, so long as I earn a doctorate.
– Dr. R

Contents

List of Figures

Preface

I have no illusions about the readership of this thesis. Realistically, a small number of graduate students will read a small number of sections, where the word "read" is being used here generously. I've tried to keep this in mind while writing, and I hope the result is useful at least to those few, those happy few...

The main goal, of course, has been to provide a written record of the main thrust of my research (and to fulfill the university's graduation requirements). Beyond that, however, I suppose I settled on two secondary goals. The first is simply to put in writing all the little details and insights that are neither generally known nor publication-worthy but still probably worth recording *somewhere* if only to save some poor future graduate student a month of wasted effort. I settled on this goal as much for the benefit of my colleagues as for myself. It's amazing what the human brain can forget. Many sections have been included in this thesis mainly as a service to my future forgetful self. Perhaps most usefully, I've advertised some software I've written over the years for various purposes. Feel free to email me at `ajcukierman@berkeley.edu` for a copy. (And then feel free to email again when it doesn't work properly.) The second goal I settled on is essentially pedagogical in nature: I've attempted to demystify several topics that typically confuse or simply overwhelm incoming graduate students. I do not claim to have *succeeded* in demystifying these topics, but I hope it makes the reader feel better that that was at least the aim. I certainly felt overwhelmed when I first started. I remember many group meetings early on when all I really understood was that certain words seemed to be important and that I should probably look them up. In many cases, however, it was surprisingly difficult to find a clear, introductory explanation, so I've attempted on several topics to fill that niche. A more minor goal has been to outline ideas for future research trajectories. Here and there, I've mentioned what was suboptimal or incomplete in my own work and how to plausibly improve on it.

In our line of work, which is extremely collaborative, it is sometimes difficult to assign credit fairly. I've tried but almost certainly failed to acknowledge all the efforts of others that made this thesis possible. I've relied heavily on the designs, calculations and measurements of others. In some cases, I give credit explicitly, but it was simply impractical to do so in every instance. For that I apologize to the relevant parties. I only managed to be consistent, I think, when it came to figures. Any figure that was not generated on my computer or by my camera is credited in the caption to an individual, a collaboration or a publication.

If any errors or gross omissions are noticed, please email me at `ajcukierman@berkeley.edu`. I won't be able to edit this manuscript once it's been submitted, but I'd still like to correct my own misconceptions and fill in the gaps in my knowledge.

Well, the past six years have certainly been enjoyable, and I've learned a tremendous

amount about a great diversity of topics. Our research occupies a strange no-man's land that spans the gap between the loftiest of philosophical speculations and the most mundane do-it-yourself engineering. I hope this thesis conveys some of the excitement and joy to be found in our work. I hope it is illuminating and perhaps even useful. But mostly I hope that some clever graduate student comes along and does a much better job than I have. It just seems more exciting that way.

Ari J. Cukierman
Berkeley, CA
April 2018

Acknowledgments

I may have had some help along the way.

I must first thank my advisor, Adrian Lee, who took a chance on a thoroughly unqualified first-year graduate student. Thank you for everything from signing this dissertation to chatting with me at the student-faculty mixer in August of 2012. Through Adrian, I became involved in the POLARBEAR, South Pole Telescope, LiteBIRD and Simons Observatory collaborations, which have provided a great wealth of experiences and introduced me to a huge number of talented and wonderful people. At Berkeley, I was allowed a remarkable degree of academic freedom. Thank you for your confidence. I promise I spent the money wisely.

It was my good fortune to have a desk for five years next to Toki Suzuki, who taught me most of what I know about CMB detectors. Thank you for answering all my annoying questions the first few years. And thank you for answering them a second time after I'd forgotten the answers.

The older graduate students were very kind and welcoming to me. I learned a tremendous amount from them. Thanks to Liz George, Nick Harrington, Bryan Steinbach and Ben Westbrook. (I already thanked you, Toki. Don't be greedy.) I can only hope I've been even half as helpful to the graduate students who came after me.

We've had a number of wonderful postdocs over the years, to whom I am also indebted. Thanks to Mike Myers, who seemed to know everything, Yuji Chinone, Nathan Whitehorn, Darcy Barron and Tijmen de Haan, who have shared their time and wisdom with me.

Thanks to Chris Raum, who taught me the art of *proper* nanofabrication. Thanks to Pete Thuesen, Warner Carlisle, Gordon Long, Tommy Gutierrez and Abel Gonzalez of the physics machine shop for teaching me the subtleties of machining and for working closely with me on my designs.

I'd like to thank Bill Holzapfel for serving on my dissertation committee and for at times acting almost as a second advisor. Thanks to Paul Richards for your sage wisdom and for all the stories of what it was like way back when. Thanks to Dick Plambeck for showing me that it's possible to be a *coherent* mm-wave astronomer (pun intended) and for a truly magnificent sense of humor. Thanks to Akito Kusaka for your advice and for many stimulating conversations.

Thanks to Aaron Parsons for serving on my dissertation committee, for teaching one of the more interesting courses I ever took and for doing very cool research that I'm always eager to learn more about. Thanks to Martin White for serving on my qualifying committee and for asking challenging but educational questions. The other committee members were suspiciously quiet during that portion of the exam.

It has also been a pleasure working with all of the younger graduate students and even the undergraduates. (Yes, undergraduates are people, too.) I've learned quite a lot from them and even relied on them for much of the work in this thesis. I've tried to be a benevolent senior graduate student. I hope you remember me fondly.

Chapter 1

Cosmic microwave background

The afterglow of the Big Bang, which we call the *cosmic microwave background* (CMB), has been an invaluable tool in understanding the origin and evolution of our universe. In this chapter, we describe the Big Bang model and the crucial role that the CMB has played in its establishment.

1.1 Big Bang model

Imposing isotropy and homogeneity on the universe, which is sometimes called a *Copernican principle*, yields the Friedmann-Robertson-Walker (FRW) metric[1]

$$ds^2 = -dt^2 + a^2(t)\left[d\chi^2 + S_k^2(\chi)d\Omega^2\right], \tag{1.1}$$

where $a(t)$ is called the *scale factor*, t is the time coordinate, Ω is a solid angle, χ is a radial spatial coordinate and

$$S_k(\chi) \equiv \frac{\sin\left(\sqrt{k}\chi\right)}{\sqrt{k}} \tag{1.2}$$

for any $k \in \mathbb{R}$.[2] The present time is denoted t_0. By convention, we often take $a_0 \equiv a(t_0) = 1$. In general, we will use a subscript 0 to denote cosmological parameters evaluated at the present time. The entire spatial part of the metric is proportional to a. The coordinate χ is called a *co-moving distance* and is useful for labeling points in a universe which is expanding or contracting. If two points are separated by a distance χ today, then they are separated by a distance $a(t)\chi$ at time t. At the present time, the co-moving distance is equal to the *physical* distance. The parameter k determines the spatial *curvature* of the universe. When $|\sqrt{k}\chi| \ll 1$, we have $S_k(\chi) \approx \chi$. In this case, we say that the geometry of the universe is *flat*, since the spatial part of the metric is just that of three-dimensional Euclidean space modulo the scale factor which stretches distances as a function of time.

With the FRW metric, we can compute the Ricci tensor $R_{\mu\nu}$ and solve the tt component

[1] We set $c = k_B = 1$ in this chapter.

[2] When $k < 0$, we have $S_k(\chi) = \sinh(\sqrt{-k}\chi)/\sqrt{-k}$.

of the Einstein field equations to obtain the *Friedmann equations*

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}, \qquad \frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p), \tag{1.3}$$

where ρ is the density and p is the pressure of a perfect, isotropic fluid that fills the universe. Both quantities are, in general, functions of the scale factor a. We can make some preliminary observations about the Friedmann equations. If $k > 0$, there is potentially a value of a for which $\dot{a} = 0$, which we call a turning point. If the universe had been expanding, i.e., $\dot{a} > 0$, it will now begin to contract, i.e., $\dot{a} < 0$. We call this possibility a *closed universe*. If $k < 0$, then there can be no turning points, in which case the universe expands forever. We call this possibility an *open universe*.

The *Hubble parameter* is defined as

$$H \equiv \dot{a}/a. \tag{1.4}$$

For most of the last century, the present value of the Hubble parameter was only known at the order-of-magnitude level, so it was often parameterized as

$$H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}, \tag{1.5}$$

where h was understood to be $\mathcal{O}(1)$. Today, we know $h \approx 0.7$ with a precision at the level of a few percent [72]. The Hubble parameter gives the recession velocity v as a function of distance d via Hubble's law: $v = Hd$. Having set $c = 1$, the Hubble parameter actually has units of inverse length or inverse time, so we can define the *Hubble length* H^{-1} and the *Hubble time* H^{-1}. The present values of these parameters are $H_0^{-1} \approx 3h^{-1}$ Gpc ≈ 4 Gpc and $H_0^{-1} \approx 10h^{-1}$ Gyr ≈ 14 Gyr. These quantities set length and time scales for cosmological processes. In fact, the present value of the Hubble time immediately yields a first estimate for the age of the universe. If we assume the expansion is linear in time, i.e., $a = \dot{a}t$, where $\dot{a} = \text{const.}$, then $t_0 = a_0/\dot{a}_0 = H_0^{-1}$.

The *critical density* ρ_c is defined as the density required *today* for the universe to be flat, i.e., $k = 0$. This is just a definition and *a priori* has no bearing on the actual density of the universe, although it happens to be true that the geometry of the universe has been measured to be very close to flat. The condition is, from the Friedmann equations (Eq. 1.3),

$$\rho_c = \frac{3}{8\pi G}H_0^2 \approx 5 \times 10^{-6} \text{ GeV/cm}^3, \tag{1.6}$$

which is the energy equivalent of 5 protons per cubic meter.[3] The ratio of the actual density to the critical density is defined as

$$\Omega \equiv \rho_0/\rho_c. \tag{1.7}$$

The Friedmann equation for the present time is, then,

$$\Omega = 1 + \frac{k}{H_0^2}. \tag{1.8}$$

[3]Contrast this with the number density of air molecules at standard temperature and pressure, which is approximately 3×10^{25} m^{-3}.

Here we see that Ω is in one-to-one correspondence with k. By measuring the density of the universe, we can estimate the curvature of the universe or vice versa.

The density of the universe is a sum of various contributions, which are typically divided into three categories: matter, radiation and dark energy. Each contribution has a certain density ρ_i and an associated normalized density $\Omega_i = \rho_{i,0}/\rho_c$. We choose these three categories, because each has a density that depends on the scale factor in a particular way. We take *matter* to consist of non-relativistic particles, so the main contribution to the energy density comes from the masses of the particles. Therefore, the density of matter is simply diluted as the universe expands, and we have $\rho_m = \rho_{m,0}/a^3 = \Omega_m \rho_c/a^3$. Current measurements indicate $\Omega_m = 0.308 \pm 0.012$ [72]. The two main components of matter in the universe are *cold dark matter* (CDM), which interacts only gravitationally, and ordinary or *baryonic matter*, which consists mainly of protons, neutrons and electrons.[4] The parameter combinations that are best constrained are $\Omega_b h^2 = 0.02226 \pm 0.00023$ and $\Omega_{\text{CDM}} h^2 = 0.1186 \pm 0.0020$ [72], so there is more than five times as much dark matter as ordinary matter. We take *radiation* to consist of relativistic particles, which contribute to the total energy density mainly through their kinetic energies and not through their masses. The number density of these particles dilutes in the same way as the non-relativistic particles, but the wavelengths associated with their momenta also stretch with the expansion of space. The density scales, then, as $\rho_{\text{rel.}} = \rho_{\text{rel.},0}/a^4 = \Omega_{\text{rel.}} \rho_c/a^4$. The two main components of radiation in the universe are photons and neutrinos, although the neutrinos have transitioned from being relativistic to non-relativistic. Current measurements indicate $\Omega_\gamma h^2 = 2.473 \times 10^{-5} (T_0/2.7255)^4$, where the present-day temperature of cosmological background photons is $T_0 = 2.7255 \pm 0.0006$ [72]. The neutrino contribution is less constrained: $0.0012 \leq \Omega_\nu < 0.016$, where the lower limit comes from measurements of neutrino mixing and the upper limit from measurements of the cosmic microwave background [72]. A better upper limit can be achieved by combining CMB data with other cosmological probes. Relative to matter, the relativistic contribution is quite small today, but, since $\rho_{\text{rel.}} \propto a^{-4}$, the photons and neutrinos dominated the energy density of the universe at $a \lesssim 1/4000$. Finally, we have dark energy, which is a poorly understood component that is responsible for the *accelerating* expansion of the universe. Most commonly, dark energy is assumed to consist of a constant energy density filling the entire universe, i.e., $\rho_\Lambda = \rho_{\Lambda,0} = \Omega_\Lambda \rho_c$. A constant dark energy is mathematically equivalent to Einstein's cosmological constant, and the terms are often used interchangeably. Current measurements indicate $\Omega_\Lambda = 0.692 \pm 0.012$ [72], making dark energy the primary component of the present-day energy density of the universe. Formally, we can define an energy density associated with the curvature of the universe, which is

$$\Omega_k = -\frac{k}{H_0^2} = 1 - \Omega. \tag{1.9}$$

This is not a physical energy density and is constrained by the actual density Ω. Current measurements indicate $\Omega_k = -0.005^{+0.016}_{-0.017}$, where the uncertainties indicate 95% confidence limits [72]. This measurement is consistent with a flat universe, i.e., $\Omega_k = 0$. Then the

[4] Although electrons are not baryons in a particle-physics context, it is common in cosmology to include them under this label. The overwhelming majority of mass in ordinary matter comes from atomic nuclei, which are baryons proper.

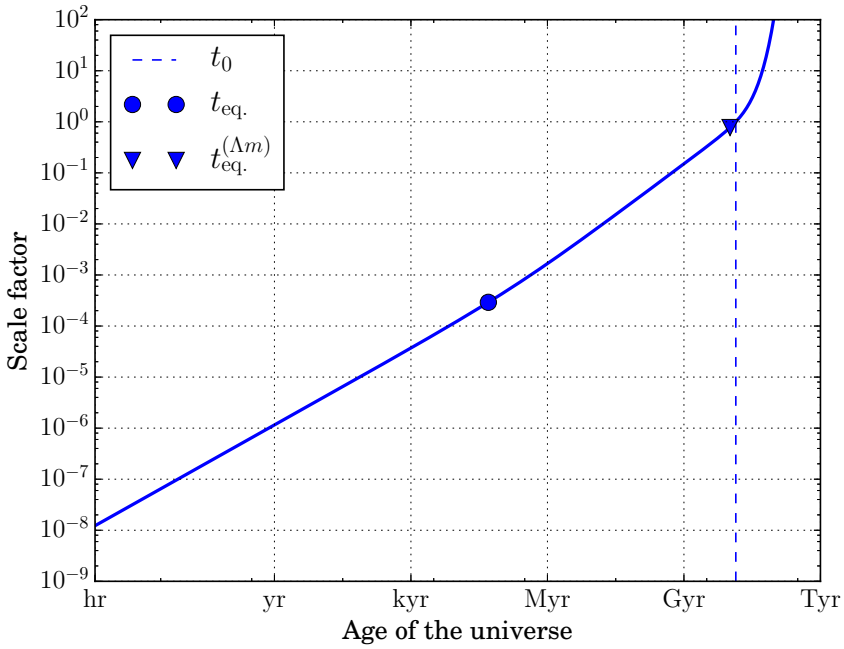


Figure 1.1: Expansion history of the universe using parameters recommended by the Particle Data Group [70]. The time of matter-radiation equality $t_{\text{eq.}}$ is indicated as well as the time of matter-dark-energy equality $t_{\text{eq.}}^{\Lambda m}$. The present time t_0 is also marked.

Friedmann equation can be rewritten as

$$H^2 = H_0^2 \left(\frac{\Omega_m}{a^3} + \frac{\Omega_{\text{rel.}}}{a^4} + \Omega_\Lambda + \frac{\Omega_k}{a^2} \right). \tag{1.10}$$

Substituting $H = \dot{a}/a$, this is nothing more than a differential equation for the function $a(t)$, which can be solved by integration, in some cases analytically but often numerically. The expansion history of our universe is shown in Fig. 1.1. For most of the history of the universe, the expansion is a power law in time, which appears as a straight line on a logarithmic plot. Currently, the expansion is beginning to accelerate. The extrapolation to future times is plotted, where we see a deviation from linearity on the logarithmic plot. The point at which the scale factor vanishes is chosen to be at $t = 0$ and is called the *Big Bang* or the *initial singularity*. For the first $50,000$ years, the energy density of the universe was dominated by radiation. It then transitioned to being matter dominated. The transition to being dominated by dark energy occurred relatively recently at $t \approx 10$ Gyr. That this transition occurred so recently in a universe in which transitions are rare is sometimes called the "Why now?" problem.

In Fig. 1.1, we plotted the scale factor as a function of time, but these are not the most convenient quantities to relate to observations. Since the cosmic expansion stretches wavelengths, the scale factor is directly related to redshift z by

$$a = \frac{1}{1+z}. \tag{1.11}$$

With spectroscopic data, it is possible to estimate redshifts and, therefore, the value of the scale factor at the time of emission or absorption. The time coordinate can be related to a co-moving distance by $d\chi = dt/a$. This substitution can be made before integrating the Friedmann equations. By substituting z for a, we have a relationship between χ and z. The observational challenges are to estimate redshift and distance. A plot of this relationship is called a *Hubble diagram*, which is named after Edwin Hubble, who in 1929 used the period-luminosity relation for Cepheid variable stars to estimate distances to "extra-galactic nebulae", which we would now simply call galaxies, and looked for a correlation with redshift [49]. His distance measurements were systematically biased, which inflated his estimate of the Hubble constant to $h \approx 5$. Using current values for the density parameters, this would imply that the age of the universe is only 2 Gyr, i.e., younger than the Earth. The highest redshift in his data set was only $z \approx 0.003$. Over the following century, more precise and extensive measurements have been made out to $z \sim 1$. These measurements have established the presence of dark energy [74, 86] and have constrained the Hubble constant to percent-level precision [7, 85].

The above discussion assumed that the universe consists of a perfect, isotropic fluid, which is, of course, an approximation. On large scales, e.g., > 10 Mpc, the approximation becomes more and more legitimate. The average motion of galaxies will conform to the Hubble law, but each individual galaxy may have a *peculiar velocity* relative to the *Hubble flow*. Peculiar velocities tend to be on the order of 100-1000 km/s, so the Hubble flow tends to dominate the observed motion at distances of more than ~ 10 Mpc.

1.1.1 ΛCDM

The standard model of cosmology describes a hot, dense state followed by cosmic expansion in a universe consisting of Standard-Model particles, cold dark matter (CDM) and dark energy in the form of a cosmological constant (Λ). This is often referred to as the ΛCDM model. The model is consistent with all current measurements [72].

Moving anti-chronologically closer and closer to the initial singularity at $t = 0$, the universe becomes arbitrarily hotter and denser. At early times, the universe is dominated by relativistic particles, since $\rho_{\text{rel.}} \propto a^{-4}$ and $\rho_{\text{non-rel.}} \propto a^{-3}$. The temperature of relativistic particles decreases as $T \propto 1/a$. As the universe cools from this hot, dense state, quarks condense into nucleons. When the temperatures are above $T \sim 1$ MeV $\sim 10^{10}$ K, which corresponds to a cosmic age of $t \sim 1$ s, the neutrons and protons are in chemical equilibrium through beta and inverse beta decays, and the electrons and positrons are in chemical equilibrium with photons through annihilation and pair production. After this time, the neutrinos decouple from matter, which produces the cosmic neutrino background (CνB) and freezes the neutron-to-proton ratio at $n/p \approx 1/7$. At the slightly lower temperature of $T \sim m_e = 511$ keV, photons no longer have enough energy to efficiently pair produce, so all of the positrons and the overwhelming majority of electrons annihilate to produce additional photons. This injection of photons ultimately increases the photon temperature relative to the neutrino temperature. Since the neutrinos are decoupled, however, this temperature difference persists after annihilation. The relationship between the two temperatures can be calculated to be $T_\nu = (4/11)^{1/3} T_\gamma \approx 0.7 T_\gamma$ assuming three neutrino species. As the neutrino temperature falls below 1 eV $\approx 10^4$ K, which occurs at a redshift $z \approx 5000$, the

neutrinos become non-relativistic and the dependence on the scale factor changes. The exact dependence is unknown due to the large uncertainty in the neutrino masses. A method for detecting the neutrino background is to search for electron energies beyond the tritium-decay endpoint [102]. This filters out electrons from the standard beta decay $^3\mathrm{H} \rightarrow {}^3\mathrm{He}\ e^-\ \overline{\nu}_e$ and retains only the products of the neutrino capture process $^3\mathrm{H}\ \nu_e \rightarrow {}^3\mathrm{He}\ e^-$. Since neutrino flux from other sources is expected to be negligibly small in this energy range, any events will be interpreted as cosmic neutrinos. The PTOLEMY (Princeton Tritium Observatory for Light, Early-Universe, Massive-Neutrino Yield) experiment will use 100 g of tritium to search for cosmic neutrinos [16]. The expected event rate, assuming the local neutrino density is close to the cosmic average, is on the order of 10 events per year.

During the first three minutes, the temperatures are high enough to allow nucleons to combine and low enough to prevent immediate disintegration. These nuclear reactions produce a spectrum of atomic nuclei. We call this event *Big Bang nucleosynthesis* (BBN). A robust prediction is that the primordial abundance of ^{4}He is $Y_p \approx 25\%$ with the rest of the matter overwhelmingly in ^{1}H. Big Bang nucleosynthesis also produces trace amounts of ^{2}H, ^{3}He and ^{7}Li. Since the energy density of neutrinos affects the expansion rate and, therefore, the freeze-out temperature of neutron-proton conversions, precision measurements of the primordial abundances can be used to constrain the total number of neutrino species N_ν and vice versa. See Cyburt et al. [24] for a comprehensive review of BBN.

After Big Bang nucleosynthesis, the universe is radiation dominated. The baryons are completely ionized, which allows the photons to exchange energy with each other through Thomson scattering and achieve an approximate thermal equilibrium with a temperature $T \propto 1/a$. We sometimes refer to this system as a baryon-photon fluid. Small inhomogeneities seed the gravitational collapse of dark matter into regions of over- and under-densities and produce sound waves or *acoustic oscillations* that propagate in the baryon-photon fluid. As the baryon-to-photon ratio is $\eta \approx 6 \times 10^{-10}$, the thermodynamic properties of the photons dominate the fluid. The sound speed is then given by $c_s^2 \sim \partial P_\gamma / \partial \rho_\gamma = 1/3$, since $P_\gamma = \rho_\gamma / 3$ for a photon gas in thermal equilibrium. This turns out to be a good approximation even through the point of matter-radiation equality. A more precise treatment accounts for the baryons and argues for adiabaticity in the fluid perturbations. This yields

$$c_s^2 = \left(\frac{\partial P}{\partial \rho} \right)_S = \frac{1}{3(1+R)}, \tag{1.12}$$

where $R = 3\rho_b/(4\rho_\gamma) \approx 670a$. Matter-radiation equality occurs at $z \approx 3000$, where $R < 1$. The approximation $c_s = 1/\sqrt{3}$ is only off by 10% in this case. During this period in the universe's history, there is a maximum distance an acoustic wave can have traveled, which we call the *acoustic horizon* or the *sound horizon*. As we will discuss below, the ionized baryons combine over a relatively short period of time to form neutral atoms, which dramatically suppresses the Thomson scattering rate and effectively decouples the photons from matter. We call this event *recombination*, since this is the conventional astrophysical term to describe ions combining with electrons to form neutral atoms; in this particular instance, however, it is a misleading term, since the atoms had never been combined before. The acoustic horizon at the time of recombination is, therefore, frozen into the background of free-streaming photons. Today these photons have been redshifted into the microwave portion of the electromagnetic spectrum and are, therefore, referred to as the *cosmic microwave*

background (CMB). Photon decoupling occurs at $z \approx 1100$, from which we can calculate a co-moving acoustic horizon of $\chi_H \approx 150$ Mpc. The estimate is almost entirely determined by the parameter combinations $\Omega_m h^2$ and $\Omega_{\text{rel.}} h^2$. The sound horizon is a characteristic length scale of perturbations in the universe.

Before recombination, the baryon-photon fluid supported sound waves but was not vulnerable to gravitational collapse. During this time, however, dark-matter over-densities were growing. After recombination, when the universe was electrically neutral and the baryons were decoupled from the photons, the baryons were able to fall into the dark-matter potential wells. This era is sometimes called the *cosmic dark ages*, since there were few charged particles for light to scatter from. It is an important period in the history of the universe, however, since this is when large-scale structure forms. It may be possible to probe this epoch by observing the hyperfine transition of neutral hydrogen, which has a rest wavelength of 21 cm (1.4 GHz) and is expected to create an absorption trough in the spectrum of the background photons at redshifts between approximately 40 and 200 [83]. As the baryons fall into the dark-matter potential wells, the first galaxies and stars form, which fills the universe with ultraviolet photons around $z \approx 30$. The Lyman-α photons excite hydrogen atoms, which then relax to one of the two hyperfine states. The hyperfine population levels are then set by the kinetic temperature of the gas rather than by the background photons. This is called the Wouthuysen-Field effect and creates a 21-cm absorption profile between redshifts of approximately 30 and 20. This profile has recently been claimed to have been detected by the EDGES (Experiment to Detect the Global Epoch of Reionization Signature) collaboration [21]. The measurement is extremely difficult, however, mainly due to the galactic synchrotron emission, which is approximately $10,000$ times brighter than the cosmological signal, so it will be necessary to wait for corroboration from an independent experiment before drawing any conclusions from the EDGES result. The ultraviolet photons heat and ionize the gas. This is called the epoch of *reionization*. By $z \approx 7$, the universe is essentially completely ionized. Approximately 5% of the background photons scatter from these reionized baryons.

After reionization, galaxies continue to form, and star formation reaches its peak around $z \sim 2$. At $z \approx 0.4$, which corresponds to a lookback time of roughly 4 Gyr, the energy density of the universe becomes dominated by dark energy, which accelerates the cosmic expansion. We are living in the transition period between matter domination and dark-energy domination.

1.2 Background radiation

At temperatures greater than 13.6 eV, we will have substantial ionization of hydrogen. Photons will interact strongly with the free electrons,[5] and we, therefore, expect the photons and baryonic matter to be in thermal equilibrium. As the universe expands and cools,

[5]There is also some scattering of photons from protons and, in principle, from ^{4}He ions. Since, however, the Thomson cross section goes as $1/m^2$, where m is the mass of the non-photon particle, the rate of proton scattering is suppressed by roughly 4×10^6 relative to electron scattering.

electrons and protons will recombine[6] to form neutral hydrogen, which has a much lower cross section for scattering photons. When the ionization fraction and the density of baryonic matter are sufficiently low, the interactions between light and matter will become so rare as to make the universe effectively transparent on large scales. Light and matter will no longer be in equilibrium with each other, and photons will stream freely through the universe. The light that comes to us from this equilibrium era is called the *cosmic microwave background* (CMB), because it is most intense in the microwave portion of the electromagnetic spectrum with a peak frequency of 160 GHz.

1.2.1 Recombination

Before recombination, when the baryons and photons are in thermal equilibrium, the thermodynamics are dominated by the photons on account of the very large photon-to-baryon ratio. With more than a billion photons per baryon, we can treat the baryons approximately as a system embedded in a thermal bath of photons. The cross section for an individual baryon to interact with a photon is much higher than the cross section for an individual photon to interact with a baryon, so the temperature of the total system is dominated by the photons. We mentioned above that a photon gas in thermal equilibrium in an expanding universe has a temperature that evolves as $T = T_0/a$. In the absence of photon scattering, the baryons would relax to a Maxwell distribution with a temperature that falls as $T_b \propto 1/a^2$. Before recombination, however, the baryon temperature is tied to the photons, i.e., $T_b = T_\gamma = T_0/a$.

If the photons and baryons are assumed to be always in thermal equilibrium, which is not true but is useful for a first approximation, then we can use the Saha equation to trace the ionization history. The ratio of free electrons to neutral hydrogen atoms is, then,

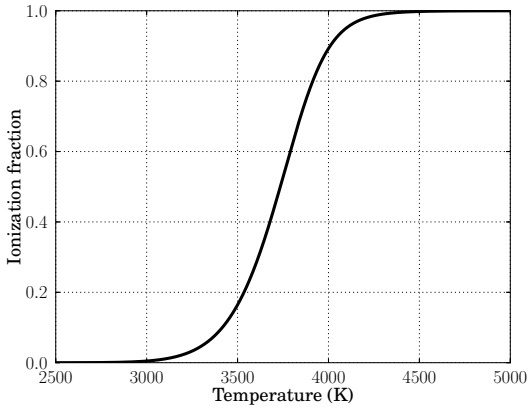
$$\frac{n_e}{n_{\mathrm{H}^{(s)}}} = \sqrt{\frac{g_e}{g_{e^{(s)}}} \frac{n_{Q,e}}{n_{\mathrm{H}^{(s)}}}} \exp\left(-\frac{B_s}{2\tau}\right). \tag{1.13}$$

where n_e is the number density of free electrons, $n_{\mathrm{H}^{(s)}}$ is the number density of hydrogen atoms in state s, $g_e = 2$ is the electron spin degeneracy, $g_{e^{(s)}}$ is the electron degeneracy of the s state of hydrogen, $n_{Q,e}$ is the quantum concentration of electrons and B_s is the ionization energy of the s state of hydrogen. Naively, this seems to imply that the ionization will fall off at $T \sim \mathrm{Ry}/2 = 8 \times 10^4$ K. The problem is that the coefficient in Eq. 1.13 is rather large: setting $n_{\mathrm{H}^{(s)}} \leq n_{b,0}$, the coefficient when $T = \mathrm{Ry}/2$ is at least 10^8. So the universe is still almost completely ionized even at $T = \mathrm{Ry}/2$. The ionization does not fall off until substantially lower temperatures. This is a convenient limit, since the low temperature creates a strong thermodynamic preference for neutral hydrogen atoms to be in the ground state.

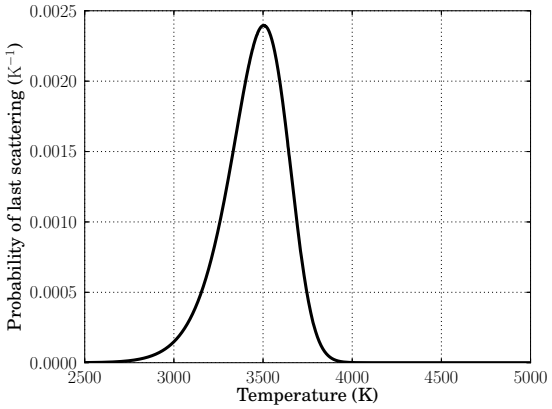
We can solve for the fractional ionization $X = n_e/(n_e + n_{\mathrm{H}})$. The result is plotted as a function of temperature in Fig. 1.2(a). We see that the drop in ionization begins around 4000 K and plateaus around 3000 K, which is substantially lower than our naive expectation based on the ionization energy of hydrogen. The postponement of recombination is mainly due to the very low baryon density: at $T = \mathrm{Ry}/2$, we have $n_b = 6 \times 10^6$ cm^{-3}.[7]

[6] They were never combined before this point, but the process $e + p \to \mathrm{H}$ is conventionally called *recombination* in astrophysics.

[7] See footnote 3.



(a) The ionization fraction of hydrogen as a function of photon temperature.



(b) The probability that a photon scatters for the last time as a function of temperature.

Figure 1.2: Recombination. These curves assume the photons and baryons are in thermal equilibrium at all times, so that the Saha equation is a valid approximation. A more precise treatment would see recombination delayed slightly by the large cross section for recombination photons at low ionization levels.

A more precise treatment would account for the fact that the photons and baryons actually begin to go out of equilibrium. Recombination is also delayed slightly by several effects such as the large Lyman-α cross section at low ionization levels. We have, however, a first approximation of the recombination process.

Since the scale factor and redshift are directly related to temperature, we can read off of Fig. 1.2(a) that recombination occured at a redshift of $z_{\text{rec.}} \approx 10^3$. The age of the universe at this time is $t_{\text{rec.}} \approx 4 \times 10^5$ yr.

With the ionization fraction, we can calculate the Thomson scattering rate. An interesting quantity we can derive is the probability that a photon scatters for the *last* time, which is shown as a function of temperature in Fig. 1.2(b). Just like Fig. 1.2(a), this should be thought of as a first approximation, but it contains the main qualitative features of interest. The probability of last scattering is narrowly distributed about ~ 3500 K. This means that the information contained in the CMB comes to us from a fairly well-defined moment in the history of the universe. Rather than seeing a superposition of different-temperature blackbodies in the CMB, we see a near perfect blackbody with a single well-defined temperature. The CMB photons come to us from approximately the same time and, therefore, approximately the same distance. We often refer to the imaginary sphere from which the CMB photons originate as the *last scattering surface*. Precision measurements, of course, resolve the non-zero width of the last scattering surface.

1.2.2 Blackbody form

During the era in which photons and baryons are in equilibrium, the photons energies relax to a Planck distribution with a number density per unit frequency given by

$$n_\nu = 8\pi \frac{\nu^2}{e^{h\nu/T} - 1}. \tag{1.14}$$

After the photons decouple from the baryons, there is no way for photons to exchange energy with each other. The photons are no longer in thermal equilibrium even with each other. If the spectral distribution were perturbed in some way, it would *not* relax back to a Planck distribution, since there are no thermalizing interactions to redistribute the photon energies. The evolution of the spectral distribution is, therefore, *not* determined by thermodynamics but only by the volumetric dilution and by the wavelength stretching associated with Hubble expansion. Interestingly, however, a spectral distribution that takes a Planck function as its initial condition will maintain a Planck distribution with a temperature that evolves as

$$T = T_0/a, \tag{1.15}$$

where T_0 is the present-day photon temperature. The evolution is identical to that of the isentropic expansion of a photon gas, for which thermal equilibrium persists at all times.[8] We have the peculiar situation, then, that the CMB *looks* like a photon gas in thermal equilibrium, although there has been no mechanism for thermalization since recombination nearly 14 Gyr ago.

Current measurements indicate $T_0 = 2.725$ K $= 0.235$ meV [33, 72], from which we can use Eq. 1.14 to find that present-day number density of CMB photons is $n_0 = 411$ cm^{-3}. The peak frequency of the blackbody distribution is given by the Wien displacement law: $\nu_0^{\text{peak}} = 2.82 k_B T/h = 160$ GHz. The specific intensity, i.e., intensity per unit frequency per unit solid angle, of the CMB is given by

$$B_\nu = \frac{2h}{c^2} \frac{\nu^3}{e^{h\nu/(k_B T)} - 1} \tag{1.16}$$

and is plotted in Fig. 1.3. The photon energy density is $\rho_\gamma = 4\sigma T^4/c = 4\sigma (T_0/a)^4/c$. At the present time, this evaluates to $\rho_{\gamma,0} = 0.26$ eV cm^{-3}. The average energy per photon is $\rho_{\gamma,0}/n_0 = 0.63$ meV.

Discovery and characterization

An early hint of the existence of the background radiation came in 1941 with a measurement of cyanogen (CN) in molecular clouds in the Milky Way [3, 65]. From the relative strengths of the transition lines, it was possible to estimate the relative populations of the energy states and, therefore, an effective temperature, which turned out to be 2.3 K. This result remained a bit of a mystery for more than two decades, until it was realized that a thermal bath of photons could be invoked to set the CN temperature.

[8] The entropy of a photon gas in thermal equilibrium in a volume $V = L^3$ is $S \propto VT^3$, so the constant-entropy or isentropic condition is $T \propto 1/L$. The physical side length of a co-moving volume is proportional to the scale factor a, so the cosmological equivalent of isentropic expansion requires $T \propto 1/a$.

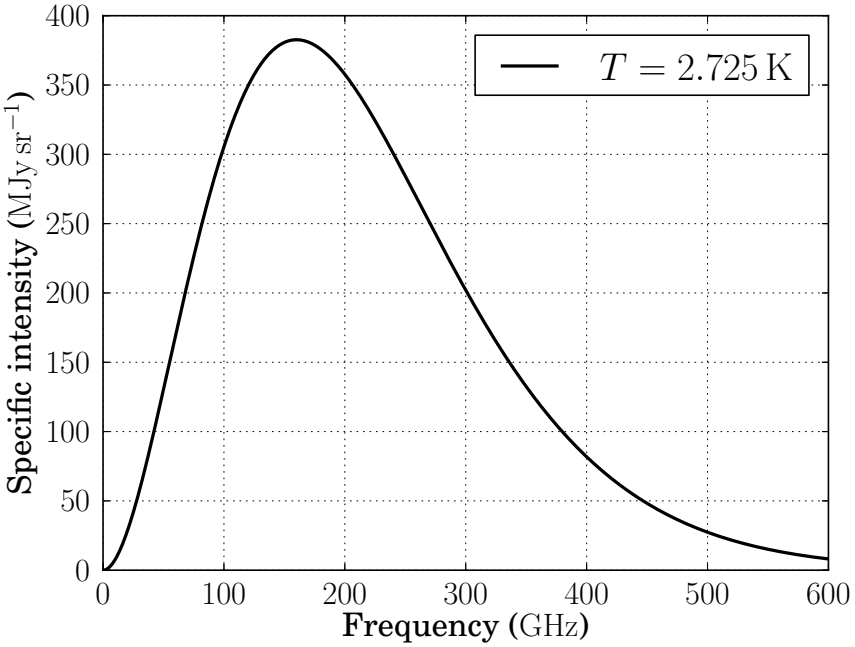


Figure 1.3: CMB blackbody spectrum.

The discovery of the cosmic microwave background is generally attributed to Arno Penzias and Robert Wilson. In 1965, they published the results of about 9 months of data alongside a companion article written by the four Princeton cosmologists Robert Dicke, Jim Peebles, Peter Roll and David Wilkinson [27, 73]. The Penzias/Wilson article is remarkably unassuming with a title that is perhaps the greatest understatement in the history of science: "A measurement of excess antenna temperature at 4080 Mc/s".[9] Essentially, they measured an unaccounted-for 4-GHz isotropic signal that was equivalent to what would be expected from a blackbody of temperature $3.5 \pm 1.0\,\mathrm{K}$. The Princeton article is much bolder with the title "Cosmic black-body radiation".

One problem in these early days was that the measurements were all done in the Rayleigh-Jeans portion of the blackbody distribution. We calculated above from the Wien displacement law that the peak frequency of a 2.725-K thermal radiator is 160 GHz. That puts the 4 GHz of Penzias and Wilson decidedly in the Rayleigh-Jeans limit. In this regime, the specific intensity is given by

$$B_\nu = \frac{2k_B T}{c^2}\nu^2. \tag{1.17}$$

But now suppose we have some emissivity ϵ. Then the specific intensity becomes

$$B_\nu = \frac{2k_B \nu^2}{c^2}\epsilon T, \tag{1.18}$$

[9]Before Hz became the SI unit of frequency in 1960, people simply used "cycles per second" or cps. Of course, it took a few years for Hz to become the *de facto* standard, and that is why Penzias and Wilson were using Mc/s (megacycles per second) in 1965.

where we have grouped the factors into known and unknown quantities. We see that measurements of the Rayleigh-Jeans distribution can only ever constrain the product ϵT so that temperature and emissivity are entirely degenerate. To determine the temperature of the CMB, it is necessary to take data at higher frequencies where the Planck and Rayleigh-Jeans distributions diverge. By the late 1970s, a peak in the spectrum had been observed indicating a temperature of approximately 3 K but with poorly understood deviations from a blackbody curve [104]. Ultimately, these deviations turned out to be due to calibration and other systematic errors.

The blackbody form of the CMB was established unambiguously in 1990 by the Far InfraRed Absolute Spectrophotometer (FIRAS) on board the Cosmic Background Explorer (COBE) satellite [62]. The FIRAS is an interferometer that compares the sky signal with that of an internal calibrated blackbody. In the preliminary 1990 analysis based on only 9 minutes of data, they found an excellent fit to a blackbody spectrum with temperature $2.735 \pm 0.06\,\mathrm{K}$. In 1994, an analysis based on 6 weeks of data found a temperature of $2.726 \pm 0.010\,\mathrm{K}$ [63]. A refined analysis based on the entire 10-month data set was published in 1996 and reported a temperature of $2.728 \pm 0.004\,\mathrm{K}$ [34]. This paper contains the famous plot of the CMB spectrum with error bars that are smaller than the thickness of the line used to plot the theory curve. The peak of the CMB spectrum has an intensity of about $400\,\mathrm{MJy/sr}$, while the 1σ uncertainties in the data points between $2\,\mathrm{cm}^{-1}$ and $22\,\mathrm{cm}^{-1}$ are less than $300\,\mathrm{kJy/sr}$. Below $18\,\mathrm{cm}^{-1}$, the 1σ uncertainties are less than $30\,\mathrm{kJy/sr}$.

The current best estimate (PDG 2014) for the CMB temperature is $2.7255 \pm 0.0006\,\mathrm{K}$, which comes from a paper by Dale Fixsen in which he uses WMAP data to recalibrate the FIRAS data set and then combines this result with the results of 15 other experiments to arrive at the value $2.72548 \pm 0.00057\,\mathrm{K}$ [33].

1.3 Inflation

There are a few problems with the standard Big Bang model.

The *horizon problem* refers to the apparent fine tuning of the statistical isotropy observed in the CMB. The co-moving distance to the CMB is $\chi_{\mathrm{CMB}} \approx 3H_0^{-1} \approx 45$ Gyr. At the time of recombination, the causal horizon, i.e., the farthest co-moving distance that light could have traveled, was $\chi_{\mathrm{rec.}} \approx 0.06H_0^{-1}$. Then, in our measurements of the CMB, the angle subtended by this horizon is only $\theta_H \approx 1^\circ$. This means that a point on the last scattering surface has only been in causal contact with points within about 1°. This is a major problem, because the CMB is uniform to about 1 part in 10^5; moreover, its angular power spectrum, which we discuss in Sec. 1.4.2, appears to be statistically uniform across the sky. How can points that have never been in causal contact appear to be in thermal equilibrium with perturbations that are drawn from the same distribution and coherent on scales larger than horizon? This paradox is sometimes called the *horizon problem*. It can be solved by introducing a period of exponential expansion in the early universe called the *inflationary epoch*. Thinking in a reverse-time sense, the exponential compression brings points much closer together than they would have in a universe that was radiation dominated at the earliest times. This increases the size of the horizon and brings the entire observable universe into causal contact at early times. This is how thermal equilibrium is established.

The *flatness problem* refers to the apparent fine tuning of the spatial curvature of the universe. The curvature Ω_k of the universe is consistent with zero to approximately 1% [7]. We will see, however, that the curvature has been growing for most of the expansion history of the universe, which suggests a fine tuning of the initial conditions to ensure near-perfect flatness today. We can define a time-*dependent* curvature parameter $\Omega_k = -k/(a^2H^2) = -k/\dot{a}^2$. Then from time t_1 to time t_2, the curvature grows by the factor

$$\frac{\Omega_{k,2}}{\Omega_{k,1}} = \left(\frac{\dot{a}_1}{\dot{a}_2}\right)^2. \tag{1.19}$$

In a decelerating universe, as was the case for most of cosmic history, i.e., during the radiation-dominated era with $a \propto t^{1/2}$ and during the matter-dominated era with $a \propto t^{2/3}$, the curvature grows (in magnitude). Only since dark energy came to dominate the universe around $z \approx 0.4$ have we had an accelerating expansion ($a \propto \exp\left(H_0\sqrt{\Omega_\Lambda}t\right)$), which will bring the geometry back toward flatness. But this has been occurring for less than a doubling in the expansion history of the universe, whereas the radiation- and matter-dominated eras span many orders of magnitude in the growth of the scale factor (Fig. 1.1). Since neutrino decoupling at $t \sim 1$ s or $a \sim 10^{-10}$, the curvature parameter Ω_k has grown by approximately 16 orders of magnitude. To measure $\Omega_k < 1\%$ today, then, requires $\Omega_{k,\nu\text{ dec.}} \lesssim 10^{-18}$. This suggests a fine tuning that could be solved with a period of accelerated expansion in the early universe that drives the curvature to zero to an arbitrary number of decimal places. It would require, however, an expansion factor of many orders of magnitude. This expansion factor is usually parameterized by the number of e-foldings, i.e., the natural logarithm of the expansion factor. The number of required e-foldings depends on when the accelerated expansion ends and the decelerated radiation-dominated expansion that we are familiar with begins.

These problems can be solved by postulating a period of accelerated expansion in the early universe that we call *inflation* or the *inflationary epoch* [40]. The accelerated expansion increases the causal horizon arbitrarily and drives the curvature to zero. A benefit that is sometimes mentioned is the dilution of magnetic monopoles, which explains why there are none for us to detect today.

The main idea is that there is a period of roughly constant energy density, which creates an accelerated expansion similar that of a universe dominated by dark energy, that undergoes a phase transition in a process called *reheating* to the hot Big Bang that we are familiar with. Many models belong to the category of *slow-roll* inflation, for which a scalar field falls towards a potential minimum slowly enough for the energy density to be approximately constant on cosmological timescales. The equation of motion for a scalar field in an expanding universe is

$$\ddot{\phi} = -3H\dot{\phi} - V'(\phi). \tag{1.20}$$

The first term on the right-hand side is a damping or frictional term, which will tend to slow down evolution of ϕ. The kinetic energy is suppressed, so the energy density is dominated by the potential energy of the scalar field. The Friedmann equation is, then, $H^2 \sim V(\phi)/M_P^2$, where M_P is the Planck mass. It is common to define *slow-roll parameters* $\epsilon \propto (V'/V)^2$ and $\eta \propto V''/V$. When the slow-roll parameters are small, the potential is approximately flat. Then the slow-roll parameters are approximately constant during inflation.

Since the energy scales at which inflation must have taken place are approaching the Planck scale, quantum fluctuations in the scalar field couple to the spacetime metric and create the seeds of the anisotropies we observe in the CMB and, ultimately, the large-scale structure we observe in the universe today. This is a nice additional benefit of the inflationary hypothesis, since the initial conditions, i.e., near uniformity with small perturbations drawn from an arbitrary probability distribution, are otherwise something of a mystery. The perturbations to the FRW metric are typically classified according to their symmetry properties; we have *scalar*, *vector* and *tensor* perturbations. The scalar perturbations translate to *density* fluctuations, and the tensor modes translate to gravitational waves. In the simplest models, the perturbations are drawn from a Gaussian distribution with a power spectrum $P(k)$, where k is the magnitude of a co-moving wavevector. The power spectrum depends only on the magnitude of $\mathbf{k}$, so long as the universe is statistically isotropic and homogeneous. The 2-point correlation is related to the power spectrum by

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}')\rangle \sim \frac{\delta^3(\mathbf{k}+\mathbf{k}')}{k^3}P(k), \tag{1.21}$$

where $\delta(\mathbf{k})$ is the Fourier transform of the overdensity field $\delta(\mathbf{r}) \equiv (\rho(\mathbf{r}) - \langle\rho\rangle)/\langle\rho\rangle$.

The Fourier modes evolve so long as the Hubble length is larger than the wavelength. During reheating, the Hubble length decreases, and the modes are frozen when $k = a_* H_*$, where a_* is the scale factor when mode k is frozen and H_* is the associated Hubble parameter at this time. The power spectrum of scalar perturbations becomes

$$P_\zeta(k) \sim \frac{H_*^2}{\epsilon_* M_P^2}. \tag{1.22}$$

This power spectrum is expected to be close to constant as a function of k. Since the modes exit the Hubble horizon at different times, however, there is a slight deviation from flatness, which is parameterized at first order by the *scalar index* or *primordial tilt* n_s. We define the scalar index by the relation

$$n_s(k) - 1 \equiv \frac{d\log P_\zeta}{d\log k}, \tag{1.23}$$

so $n_s = 1$ gives a perfectly flat power spectrum. The vector perturbations dilute away during the period of accelerated expansion and are, therefore, often ignored. The power spectrum of tensor perturbations becomes

$$P_t(k) \sim \left(\frac{H_*}{M_p}\right)^2, \tag{1.24}$$

where the tilt is defined by

$$n_t(k) \equiv \frac{d\log P_t}{d\log k}. \tag{1.25}$$

Since scale invariance is broken by the different exit times from the Hubble horizon, we expect n_s to be close but not equal to 1 and n_t to be close but not equal to 0.

The *tensor-to-scalar* ratio is defined to be

$$r_* \equiv \frac{P_t(k_*)}{P_\zeta(k_*)} = 16\epsilon_*. \tag{1.26}$$

When values of the tensor-to-scalar ratio are reported, they must be given at a fiducial scale k_*. For instance, $r_{0.002}$ is understood to denote the tensor-to-scalar ratio evaluated at $k_* = 0.002$ Mpc^{-1}. We can now derive a relation between the tensor-to-scalar ratio and the energy scale at horizon exit. We have

$$r_* \sim 16 \frac{V_*}{A_s M_P^4}, \tag{1.27}$$

where $A_s = P_\zeta(k_*)$ is constrained by measurements of the temperature anisotropies in the CMB to be $A_s = 2 \times 10^{-9}$ at $k_* = 0.05$ Mpc^{-1} [72]. The quantity V is actually a potential energy *density*, so the energy scale at horizon exit is

$$V_*^{1/4} \sim 10^{16} \text{ GeV} \left(\frac{r_*}{0.01}\right)^{1/4}. \tag{1.28}$$

Here we see that the inflation occurs naturally at the energy scale of grand unified theories (GUT). The dependence on the tensor-to-scalar ratio is rather weak. A modest measurement of r will, therefore, give a relatively precise indication of the inflationary energy scale. On the other hand, a modest decrease in the energy scale of inflation can suppress r to undetectably small values.

We can derive some consistency relations. We have $n_s - 1 \approx -6\epsilon_* + 2\eta_*$ and $n_t \approx -2\epsilon_a st$, so $n_t = -r/8$. Given an inflationary model, we can compute allowed values for the slow-roll parameters ϵ_* and η_*, which immediately constrain the allowed values for the observables n_s and r. A measurement of r would immediately constrain the allowed values of n_t, which is also, in principle, an observable. From measurements of n_s and upper limits on r, the *Planck* collaboration has been able to constrain several common inflationary models as shown in Fig. 1.4. The current best limit on the tensor-to-scalar ratio is [18]

$$r_{0.05} < 0.07, \tag{1.29}$$

and the best measurement of the spectral index is [7]

$$n_s = 0.968 \pm 0.006, \tag{1.30}$$

which shows that the scalar index differs from 1 with a 5σ statistical significance.

The number of e-foldings N_* is the natural logarithm of the expansion factor during inflation. We can set a lower bound on N_* by requiring that the horizon and flatness problems are solved. If we additionally require that inflation ends above the TeV scale, which must be true since these energy scales have been probed in particle colliders, then we have $N_* \gtrsim 40$. For GUT-scale inflation, we have $N_* \gtrsim 60$. During the inflationary epoch, the scalar field varies by an amount $\Delta\phi$. This is sometimes called the field *excursion* and is related to the number of e-foldings by $N_* \sim \Delta\phi / \sqrt{\epsilon_*} M_P$. Setting $N_* \gtrsim 40$ implies the *Lyth bound* [60]

$$\frac{\Delta\phi}{M_P} \gtrsim \sqrt{\frac{r}{0.01}}. \tag{1.31}$$

A model with a field excursion exceeding the Planck scale belongs to the category of *large-field inflation*.

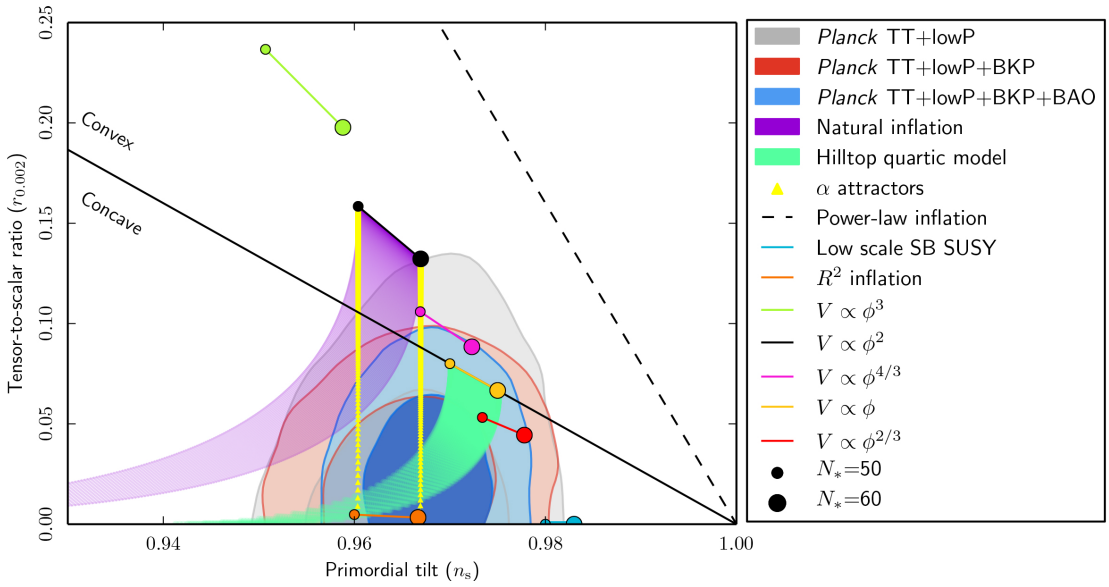


Figure 1.4: Exclusion plot for models of inflation from Planck Collaboration et al. [80].

1.4 Anisotropies

The perturbations set during the inflationary epoch act as the initial conditions for the radiation-dominated era and allow the universe to depart from perfect isotropy and homogeneity. During this era, dark matter begins to gravitationally collapse and form potential wells. In the meantime, sound waves propagate in the baryon-photon fluid and are themselves perturbed by the growing dark-matter potential wells. The anisotropies are frozen into the photons at the time of recombination. During the epoch of reionization, approximately 6% of the background photons scatter one additional time. This partially obscures the information from the last scattering surface and is often parameterized by an optical depth $\tau = 0.066 \pm 0.016$ [72]. For this reason, a parameter combination that is frequently reported is $A_s e^{-2\tau}$. The ripples in the baryon-photon fluid are small enough to allow for the validity of low-order perturbation theory. Given a set of cosmological parameters, the anisotropies in the CMB can be computed with a program such as the Code for Anisotropies in the Microwave Background (CAMB) [58]. This makes the CMB an excellent tool for constraining cosmological parameters.

The anisotropies in the CMB were first detected by the Differential Microwave Radiometer (DMR) on board the Cosmic Background Explorer (COBE) [96]. The best measurements to date were made by the *Planck* satellite [78] and are shown in Fig. 1.5. The RMS fluctuations are at the level of 100 μK on the 2.7-K average background. The anisotropies are conventionally decomposed in spherical harmonics as

$$\Delta T(\hat{\mathbf{n}}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_\ell^m(\hat{\mathbf{n}}) \tag{1.32}$$

for some complex coefficients $a_{\ell m}$. The assumption of statistical isotropy implies that the

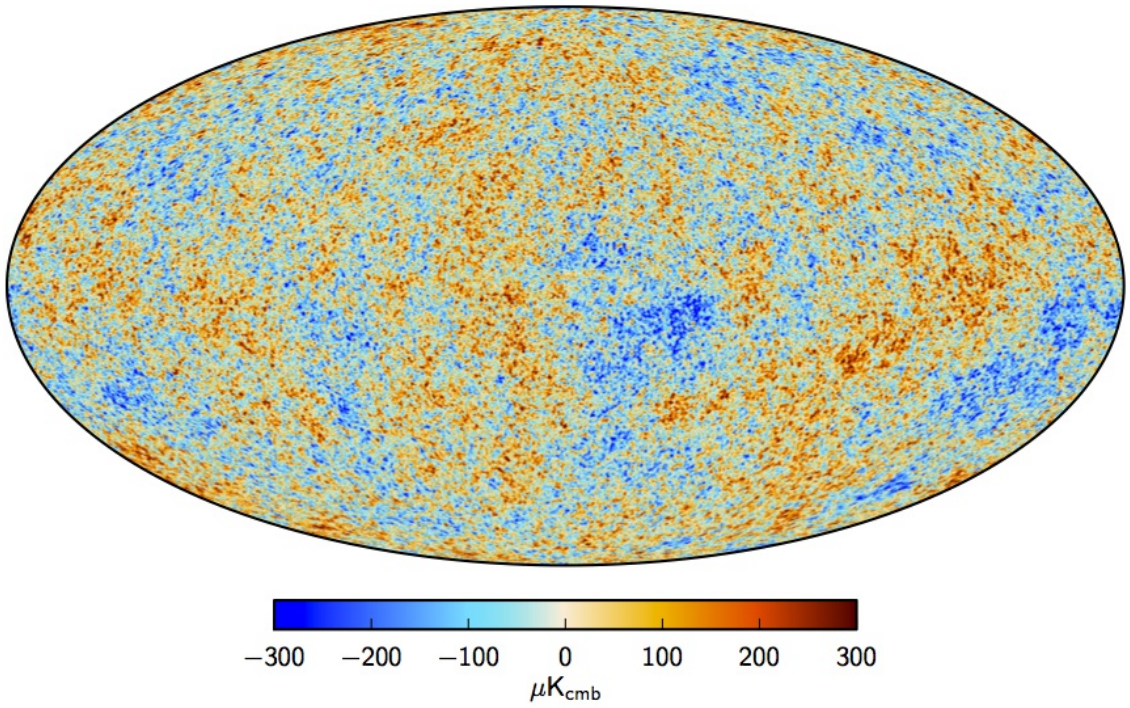


Figure 1.5: The full-sky *Planck* CMB map [78]. The map uses a Mollweide projection, which is a type of equal-area map projection.

2-point correlation depends only on the angular distance and not on the angular coordinates. We have

$$\langle \Delta T(\hat{\mathbf{n}}) \Delta T(\hat{\mathbf{n}}') \rangle = \frac{1}{4\pi} \sum_{\ell=0}^{\infty} (2\ell+1) C_\ell P_\ell(\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}'), \tag{1.33}$$

where $P_\ell(x)$ is the Legendre polynomial of order ℓ and C_ℓ is defined by

$$\langle a_{\ell m} a^*_{\ell m} \rangle = \delta_{\ell\ell'} \delta_{mm'} C_\ell. \tag{1.34}$$

In estimating the C_ℓ, we are limited by the fact that there are a finite number of modes at each multiple moment ℓ. At most, then, we have $2\ell + 1$ estimates of each C_ℓ. If we reduce the fraction of the sky f_{sky} that we measure, then we are effectively reducing the number of modes we can access. The effective number of estimates is, then, $(2\ell+1) f_{\text{sky}}$. For Gaussian perturbations, the sample variance on our estimate of C_ℓ is

$$\Delta C_\ell^2 = \frac{2}{(2\ell+1) f_{\text{sky}}} C_\ell^2. \tag{1.35}$$

This sort of sample variance is called *cosmic variance* and is due to the fact that there are only a finite number of modes available at each multipole moment ℓ. This can only be improved upon by moving a cosmological distance away from the Earth so as to make a statistically independent measurement of the background radiation. Any noise in the experiment will increase the variance further.

The $\ell = 0$ component of the spherical-harmonic decomposition is called the *monopole* and is nothing other than the average sky temperature T_0, which we discussed in Sec. 1.2.2.

The $\ell = 1$ component is called the *dipole anisotropy* and is dominated by the motion of the Earth relative to the Hubble flow. The universe is isotropic and homogeneous only when we average over very large volumes. The Earth, the Solar System, the Milky Way Galaxy and even the Local Group move with respect to the frame in which the background radiation appears isotropic. This peculiar motion, which is sometimes called *aether drift* in reference to the idea that the universe has a preferred reference frame, blueshifts the CMB in the direction of motion and, therefore, generates an angular dependence. The apparent temperature is given by

$$T(\theta) = \frac{1}{\gamma(1 - v\cos\theta)} T_0, \tag{1.36}$$

where v is the peculiar velocity of the observer, θ is the angle from the direction of maximum peculiar velocity and $\gamma = 1/\sqrt{1 - v^2}$. For small v, the anisotropy is

$$\Delta T = T_0 v \cos\theta + \mathcal{O}(v^2), \tag{1.37}$$

which is proportional to $Y_1^0(\theta, \phi)$, i.e., a dipole. There were several hints of a dipole anisotropy starting in the late 60s. The first serious measurement is often considered to be the Berkeley result from 1977 [95]. In this experiment a twin-antenna Dicke radiometer sensitive to 33 GHz was flown on a U2 aircraft at an altitude of 20 km. There were 8 flights between December 1976 and May 1977, and each yielded about 3.5 hours of data. They found a dipole amplitude of 3.5 ± 0.6 mK in the direction of the galactic coordinates $(\ell, b) = (248°, 56°)$. The dipole amplitude translates to a Solar-System velocity of 390 ± 60 km/s. In the PDG 2014, the amplitude of the dipole anisotropy due to the motion of the Solar System is given as 3.355 ± 0.008 mK, where the maximum occurs at galactic coordinates $(\ell, b) = (263.99° \pm 0.14°, 48.26° \pm 0.03°)$. These numbers ultimately come from the WMAP 5-year results [47]. The anisotropy indicates a velocity of 369 km/s. Now we see that the assumption $v \ll c$ was a good one. The corrections to higher-order modes are at the level of 0.1% and have been detected by the *Planck* collaboration [77].

The higher-order modes will be discussed in Sec. 1.4.2.

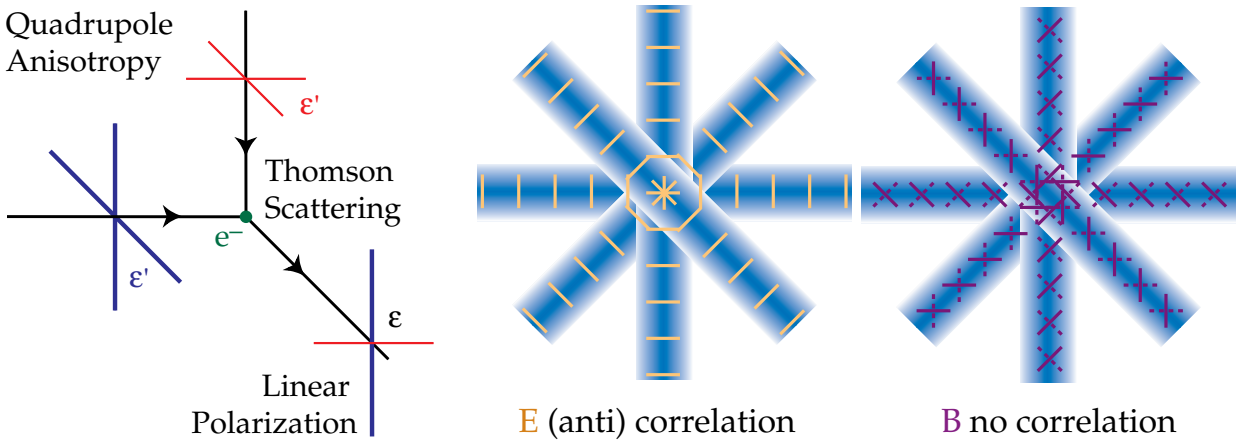
1.4.1 Polarization

The CMB is expected to contain a polarized component due to Thomson scattering from quadrupole anisotropies at the time of last scattering. This is illustrated in Fig. 1.6(a). A convenient decomposition of the CMB polarization field is in terms of so-called E- and B-modes. The E-modes are even parity, and the B-modes are odd-parity. For observations on relatively small patches of sky, it is convenient to work in the *flat-sky approximation*. In this case, the spherical-harmonic decomposition can be replaced by a two-dimensional Fourier transform. The spherical-harmonic indices ℓ and m are replaced by a 2D vector $\boldsymbol{\ell}$. In this approximation, the Fourier-space representations of the modes are defined by

$$E(\boldsymbol{\ell}) = Q(\boldsymbol{\ell})\cos(2\phi_{\boldsymbol{\ell}}) + U(\boldsymbol{\ell})\sin(2\phi_{\boldsymbol{\ell}}) \tag{1.38}$$

and

$$B(\boldsymbol{\ell}) = -Q(\boldsymbol{\ell})\sin(2\phi_{\boldsymbol{\ell}}) + U(\boldsymbol{\ell})\cos(2\phi_{\boldsymbol{\ell}}), \tag{1.39}$$



(a) Thomson scattering from a quadrupole anisotropy.

(b) Polarization fields for pure E- and B-modes.

Figure 1.6: Illustrations of CMB polarization [48].

where Q and U are the Stokes polarization parameters and $\phi_{\boldsymbol{\ell}}$ is the angle that $\boldsymbol{\ell}$ makes with the ℓ_x-axis. Stokes Q is the difference in power between the components polarized along the x- and y-axes. Stokes U is the same but for axes oriented at 45° to the x- and y-axes. Some cartoons of pure E- and B-modes are shown in Fig. 1.6(b), the characteristic feature of the B-modes being the spiral patterns. In Q maps, pure E-modes will produce features oriented along the x- and y-axes, whereas U maps will show features oriented at 45° and 135°. This is just the opposite for the B-modes. An example is shown in Fig. 1.7.

The great utility of the E/B decomposition is that the B-modes are generated *only* by tensor perturbations [90], i.e., from gravitational waves propagating in the early universe, whereas the E-modes and the temperature (intensity) anisotropies are sourced by *both* scalar and tensor modes with the scalar modes dominating. The B-modes are, therefore, a kind of clean channel for primordial gravitational waves. The production of gravitational waves is essential to the inflationary model. The search for B-modes in the polarized fluctuations of the CMB is, therefore, an attempt to test and constrain the inflationary hypothesis. A detection would open a window onto physics at energy scales potentially 12 orders of magnitude beyond what can be probed at the Large Hadron Collider (LHC).

Just as we did for the temperature anisotropies, the E- and B-modes can be decomposed into orthogonal functions, although they are, in this case, spin-2 spherical harmonics. But we can still take the variance of the coefficients and define an angular power spectrum. The polarization analogs of the C_ℓ are often denoted C_ℓ^E and C_ℓ^B.

One wrinkle is that the approximation of free-streaming, unperturbed photons is not valid at the level of precision required to measure B-modes sourced by primordial gravitational waves. The main obscuring effect comes from the gravitational lensing of CMB photons by large-scale structure that forms between recombination and the present. The E-mode symmetry is broken, and spurious B-modes are created. In fact, the modes are mixed

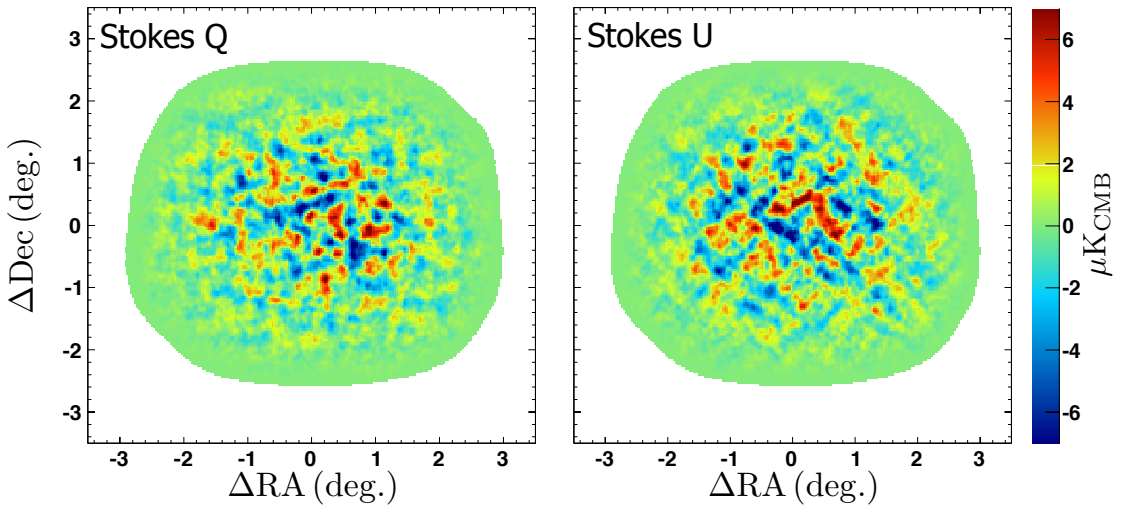


Figure 1.7: Stokes Q and U from one of three observation patches (RA23) presented in Polarbear Collaboration et al. [81]. The polarization measurement is dominated by E-modes, which tend to create features oriented along the x- and y-axes in Q and along the 45° and 135° diagonals in U.

according to the transformations [105]

$$C_\ell^E \to C_\ell^E + \frac{1}{2}\left[W_{1\ell}^{\ell'} + W_{2\ell}^{\ell'}\right] C_{\ell'}^E + \frac{1}{2}\left[W_{1\ell}^{\ell'} - W_{2\ell}^{\ell'}\right] C_{\ell'}^B \tag{1.40}$$

$$C_\ell^B \to C_\ell^B + \frac{1}{2}\left[W_{1\ell}^{\ell'} - W_{2\ell}^{\ell'}\right] C_{\ell'}^E + \frac{1}{2}\left[W_{1\ell}^{\ell'} + W_{2\ell}^{\ell'}\right] C_{\ell'}^B, \tag{1.41}$$

where the $W_{i\ell}^{\ell'}$ are window functions that depend on the power spectrum of gravitational lensing. This mixing implies that B-modes will appear even in the absence of primordial gravitational waves. Notice also that ℓ-modes are mixed.

As polarization measurements become more sensitive, the search for primordial B-modes will become limited by gravitational lensing. For this purpose, several methods of *delensing* have been proposed. For low-noise maps, it is possible to measured E- and B-modes to invert the effects of lensing [89]. For higher-noise maps, it is better to look for correlations with large-scale structure like the cosmic infrared background (CIB) [61, 91].

The E-modes were first detected in 2002 by the Degree Angular Scale Interferometer (DASI) collaboration. The lensing B-modes were first detected indirectly in 2013 by cross-correlating data from the South Pole Telescope (SPTpol) and the *Herschel* satellite [42]. Since then, the POLARBEAR, BICEP (Background Imaging of Cosmic Extragalactic Polarization), SPT and ACT (Atacama Cosmology Telescope) collaborations have made increasingly precise measurements of the lensing B-modes [5, 6, 8, 18, 19, 20, 51, 59, 81], which we will present in Sec. 1.4.2.

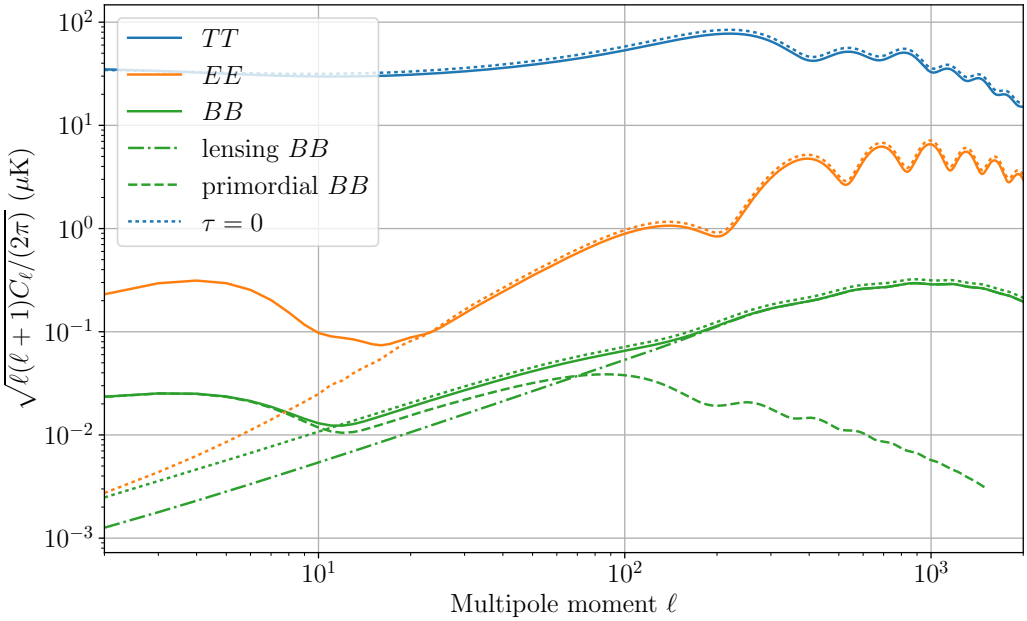
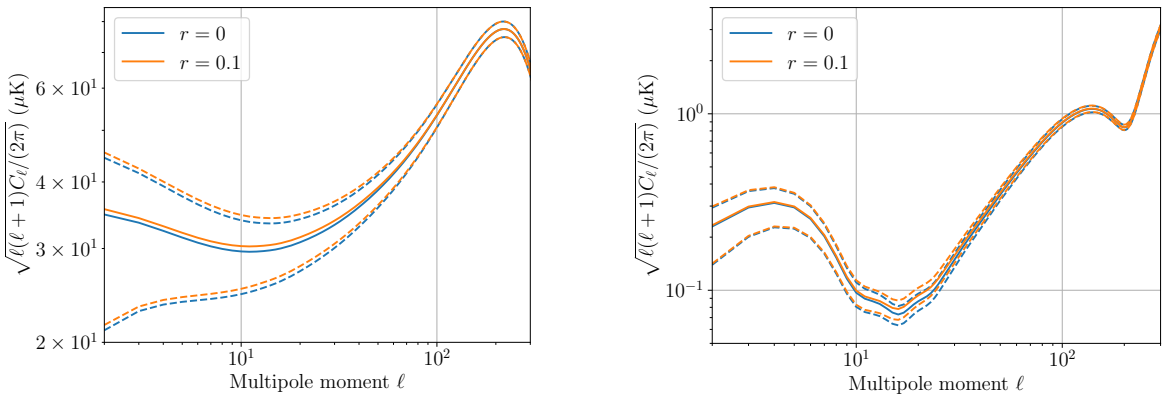


Figure 1.8: Angular power spectra for the temperature, E-mode and B-mode anisotropies simulated with CAMB. We use $r = 0.02$ as a fiducial tensor-to-scalar ratio. The TT and EE spectra are negligibly altered from the $r = 0$ case. The lensing BB spectrum is the same as the total BB with $r = 0$. The dotted lines for all colors show the spectra in the absence of reionization ($\tau = 0$).

1.4.2 Angular power spectra

An *angular power spectrum* is nothing more than a plot of C_ℓ as a function of multipole moment ℓ. Very roughly, features of angular size θ contribute mainly to the spherical harmonics of multiple moment $\ell \sim 180°/\theta$. For example, the characteristic angular scale of a dipole anisotropy is 180°, so it shows up primarily at $\ell = 1$. It is often convenient to use the quantity $D_\ell = \ell(\ell + 1)C_\ell/(2\pi)$, since this keeps the spectrum roughly flat. Some simulated power spectra are shown in Fig. 1.8.

We begin by discussing the temperature power spectrum. At low-ℓ, the spectrum is roughly flat. This is called the *Sachs-Wolfe plateau* and is the portion of the spectrum detected by the DMR in 1992 [96]. The *Sachs-Wolfe effect* is the gravitational redshift of CMB photons that must climb out of potential wells from the surface of last scattering to our telescopes today. Since the gravitational potential is not the same everywhere, this effect alone creates temperature anisotropies. The low-ℓ modes correspond to large angular scales, which roughly correspond to large spatial scales. These large modes remained outside of the acoustic horizon for the duration of the radiation-dominated era and are, therefore, largely unevolved. They give an indication of the primordial power spectrum for which $\ell(\ell+1)C_\ell \approx$ const. The *integrated Sachs-Wolfe effect* (ISW effect) is the gravitational redshift of CMB



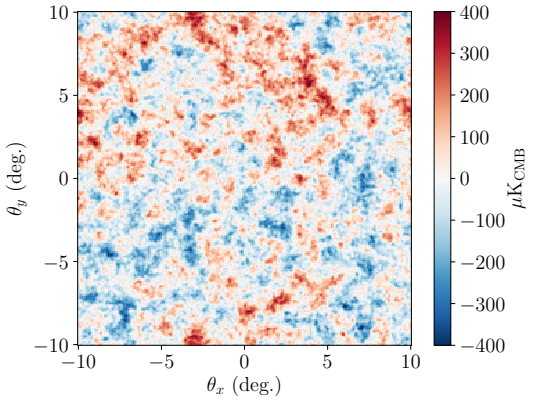
(a) *TT* (b) *EE*

Figure 1.9: Illustration of the cosmic-variance limit on r without B-modes. The dashed lines indicate the 1σ cosmic variance.

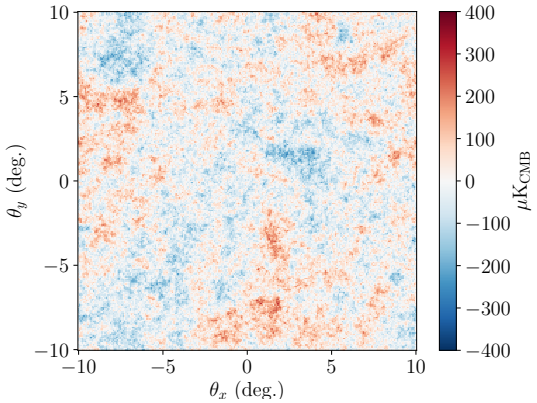
photons due to time-varying gravitational potentials. As the photons are free streaming after recombination, they fall into potential wells and climb back out. The potential may evolve, however, over the time it takes the photon to fall in and climb back out, so the photon energy is not necessarily the same when it emerges. This is an additional source of anisotropy and is the reason for the slight rise, which is called the *ISW rise*, in the power spectrum at the lowest multipole moments.

Because the low-ℓ modes have not been affected by baryon acoustic oscillations, they contain the most direct information about the initial conditions including the possible presence of tensor perturbations. Because there are a finite number of these pristine modes, there is a cosmic-variance limit on the ability of the temperature anisotropies to constrain the amplitude of the tensor power spectrum [55]. In Fig. 1.9(a), we show the Sachs-Wolfe plateau with and without tensor perturbations. We see that, for $r \lesssim 0.1$, the difference between the two cases is not statistically distinguishable from the scatter expected from cosmic variance. Notice also that the curves are essentially identical for $\ell \gtrsim 80$, where the rise from the first acoustic peak begins. For these reasons, the cosmic-variance limit on r from measurement of the temperature anisotropies is $r \lesssim 0.1$. Combined with some polarization sensitivity, the *Planck* collaboration achieved $r < 0.11$ at the 95% confidence level [80].

The sub-horizon modes are perturbed by sound waves or *acoustic oscillations* in the primordial plasma or baryon-photon fluid. The sound waves are themselves perturbed by the growing dark-matter potential wells. At recombination, the pressure of the baryon-photon fluid suddenly vanishes, which freezes in the acoustic oscillations to the photon background. The modes that have undergone a half-integer number of periods imprint regions of maximal compression and rarefaction, which set characteristic length scales in the plasma at the time of recombination. These length scales appear as characteristic *angular* scales in the CMB as illustrated in Fig. 1.10(a). As a comparison we show in Fig. 1.10(b) what the CMB would look like if the Sachs-Wolfe plateau, i.e., $D_\ell \approx$ const., extended indefinitely. A characteristic angular scale manifests itself in the angular power spectrum as an enhancement



(a) Using the temperature power spectrum from Fig. 1.8. The characteristic size of the anisotropies is $\sim 1^\circ$ corresponding to a peak in the power spectrum at $\ell \sim 180^\circ/1^\circ \approx 200$.



(b) Using $D_\ell = 1000\ \mu\mathrm{K}^2$, which is approximately the level of the Sachs-Wolfe plateau, for all ℓ. A characteristic angular scale is more difficult to identify in this case.

Figure 1.10: Comparison of CMB simulations with and without acoustic peaks. For these small flat-sky maps, we draw two-dimensional Fourier amplitudes for each wavevector $\boldsymbol{\ell}$ from a Gaussian distribution whose variance is proportional to C_ℓ. The power spectrum is interpolated for non-integer ℓ.

in a particular range of multipole moments. We call these features *acoustic peaks*. We see in Fig. 1.8 that these acoustic features begin around $\ell \approx 200$ and are spaced by $\Delta\ell \approx 300$. The odd-numbered peaks correspond to standing waves that have undergone an odd number of half periods. In the first half period, the plasma is pulled into the gravitational wells; the fluid is, then, compressed in the potential wells and rarefied at the potential peaks. Undergoing any integer number of periods from this point will leave the qualitative picture unchanged, i.e., compression in the potential well and rarefaction at the potential peaks. For the even-numbered peaks, the story is just the opposite. These correspond to standing waves that have undergone an *even* number of half periods or simply an integer number of periods. In the first half-period, the fluid is compressed in the potential wells just as we described above. In the second half-period, however, the fluid bounces back out of the potential well and is, therefore, rarefied in the wells and compressed at the peaks, which is just the opposite of the odd-numbered peaks. The enhancement, suppression and location of each peak depends on the baryon and dark-matter content of the universe. The location of each peak also depends on the geometry of the universe, which controls how length scales are translated to observed angular scales. Some variations on the temperature power spectrum are shown in Fig. 1.11. The curvature mainly controls the locations of the peaks. The ratio of the baryon density Ω_b to the dark-matter density Ω_c controls the relative heights of the peaks. The first acoustic peak was detected in 2000 by the BOOMERanG (Balloon Observations Of Millimetric Extragalactic Radiation ANd Geophysics) and MAXIMA (Millimeter Anisotropy eXperiment IMaging Array) experiments [26, 41]. Since then, many experiments have made increasingly precise measurements of the temperature power spectrum, most notably the

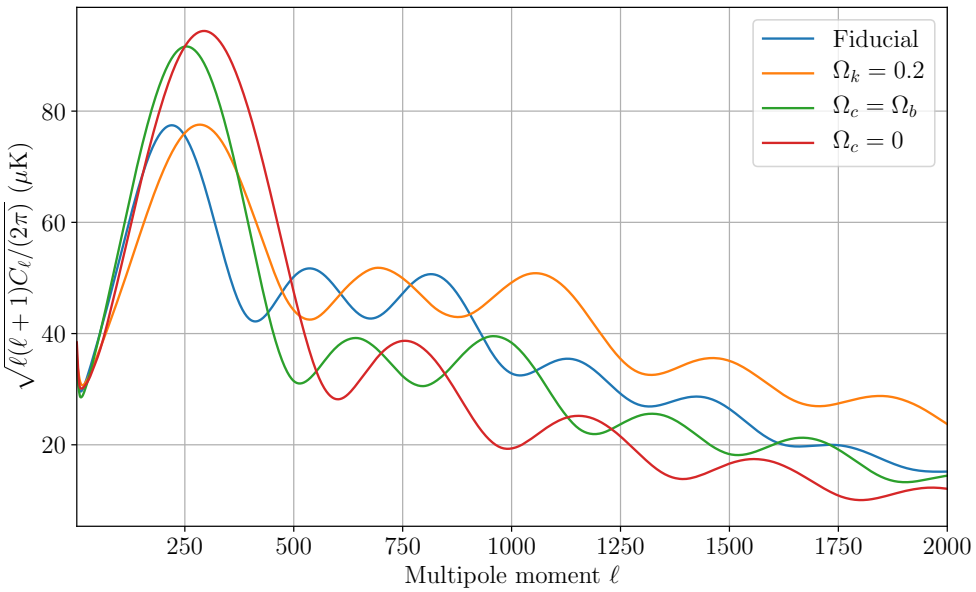
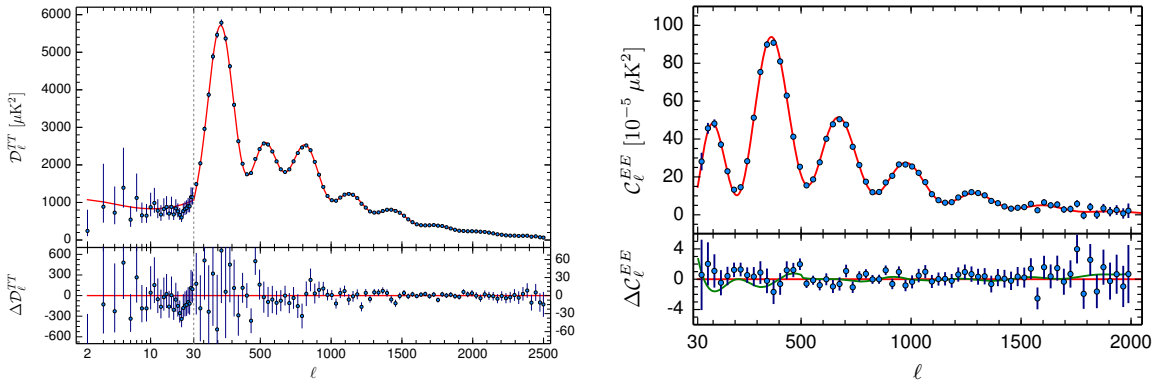


Figure 1.11: Temperature power spectra for a variety of input parameters. The fiducial model has $\Omega_b h^2 = 0.0226$, $\Omega_c h^2 = 0.112$ and $\Omega_k = 0$. We hold Ω_m fixed for these variations.

WMAP [14] and *Planck* satellites [75]. The most recent *Planck* measurements, which are shown in Fig. 1.12(a), are the most precise up to $\ell \sim 2000$. The SPT and ACT collaborations have made measurements to higher ℓ [25, 97].

The power spectrum decays steeply for $\ell \gtrsim 2000$. This is called the *damping tail*. This is due to *Silk damping*, which is the diffusion of photons on scales similar to the mean free path. On these scales, the structure imprinted by acoustic oscillations is washed out. Another effect comes from the finite thickness of the last scattering surface (Fig. 1.2(b)), which suppresses anisotropies on comparable scales. It becomes increasingly difficult to extract information about the primordial plasma for $\ell \gtrsim 2000$.

The quadrupole anisotropies that source polarization (Sec. 1.4.1) are created by diverging or converging fluid velocities that redshift or blueshift, respectively, photons along the wavevector. The velocities are largest when the acoustic oscillations have undergone an odd number of *quarter* periods. Recall that the acoustic peaks in the temperature power spectrum corresponded to integer numbers of *half* periods. So we expect the acoustic peaks in polarization, which is dominated by the E-modes, to be exactly out of phase with temperature, which can be seen in a comparison of the temperature and E-mode spectra in Fig. 1.8. For $\ell \lesssim 1700$, the most precise measurements of the E-mode power spectrum come from the *Planck* collaboration and are shown in Fig. 1.12(b). At higher ℓ, measurements have been made by the SPT and ACT collaborations [45, 59]. Just like the low-ℓ temperature modes, the low-ℓ E-modes can be used to constrain the tensor-to-scalar ratio. In this case, however, the modes with $\ell \lesssim 10$ are dominated by the reionization peak, which we discuss below, so there is an even smaller ℓ-range over which the tensor modes could potentially manifest



(a) Temperature power spectrum.

(b) E-mode power spectrum.

Figure 1.12: Measured power spectra from the *Planck* satellite [78]. The lower panels shows residuals with respect to the best-fit ΛCDM model using both temperature and polarization data.

themselves. We illustrate the dependence in Fig. 1.9(b), where it can be seen that $r = 0.1$ is statistically indistinguishable from $r = 0$.

The B-modes are typically broken into two components, which are shown in Fig. 1.8. The component that is predicted by the standard Big Bang model and that has been detected is that due to the gravitational lensing of E-modes, which peaks near $\ell \approx 1000$. The component due to primordial gravitational waves, which is *not* required by the standard Big Bang model and has *not* been detected, has a peak near $\ell \approx 100$. The amplitude depends on the tensor-to-scalar ratio r. In Fig. 1.8, we have plotted the primordial contribution for $r = 0.02$, where we see that the lensing component dominates above $\ell \sim 80$. The level of the polarized fluctuations in the primordial B-modes is 10-100 nK, and recall that this measurement must be made on a background of 2.7 K. This requires exquisite control of noise and systematic uncertainties and motivates the detector development discussed in the following chapters.

In Fig. 1.8, we have also shown the effects of reionization on the angular power spectra. For $\ell \gtrsim 20$, reionization mainly has the effect of attenuating power. The additional scattering of CMB photons from reionized hydrogen washes out the primordial structure. Reducing the amplitude of the initial power spectrum of perturbations would have the same effect on the angular power spectrum. For this reason, A_s and τ tend to be degenerate model parameters. The best constrained combination is $A_s e^{-2\tau}$. For $\ell \lesssim 20$, there is a dramatic increase in the polarization power as a result of reionization. The additional peaks at $\ell \sim 5$ in the E- and B-modes are called *reionization peaks* or *reionization bumps*. Measuring the E-mode reionization peak will help us to break the degeneracy between A_s and τ. For the B-modes, the reionization peak only appears when there are primordial gravitational waves. The lensing signal is due to all of the deflections of CMB photons along the line of sight between the observer and the last scattering surface, but the bulk of the lensing occurs at relatively low redshifts *after* reionization. For this reason, there is no reionization peak in the lensing B-modes but only in the primordial B-modes. This extra peak is an additional opportunity to detect the presence of tensor modes in the primordial plasma and appears

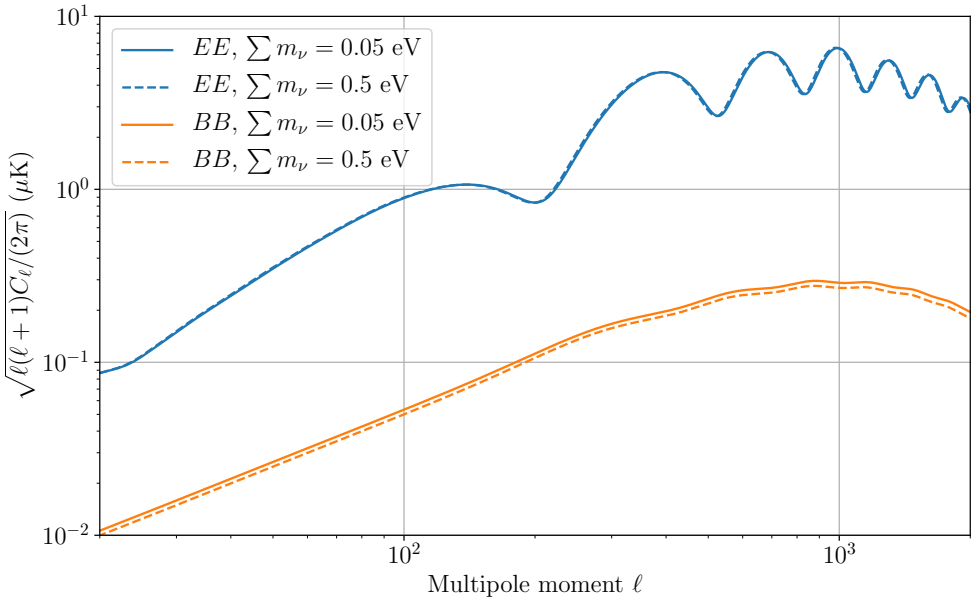


Figure 1.13: The dependence of CMB polarization power spectra on the sum of the neutrino masses.

fortunately in an ℓ-range where the lensing signal is relatively weak. Reducing the lowest multipole moment that an experiment can access can significantly improve the sensitivity to r [94].

The lensing signal can be used to constrain parameters that only weakly affect the temperature spectrum, e.g., the sum of neutrino masses. The neutrinos are free streaming for most of the expansion history of the universe. Unlike the massless photons, the neutrinos interact gravitationally with the other matter components. The heavier the neutrinos are, the more they will tend to suppress the growth of structure in the universe. When there is less structure, there is less gravitational lensing. The neutrino density is related to the sum of the neutrino masses by [72]

$$\Omega_\nu h^2 = \frac{1}{93 \text{ eV}} \sum_j m_{\nu_j}, \tag{1.42}$$

which enters as a parameter in CMB simulators such as CAMB. The current limits are

$$0.05 \text{ eV} \leq \sum_j m_{\nu_j} \leq 0.23 \text{ eV}, \tag{1.43}$$

where the lower limit comes from measurements of neutrino mixing and upper limit from *Planck* combined with measurements of baryon acoustic oscillations and the Hubble parameter [7, 72]. The dependence of the E- and B-mode spectra on the sum of the neutrino masses is illustrated in Fig. 1.13.

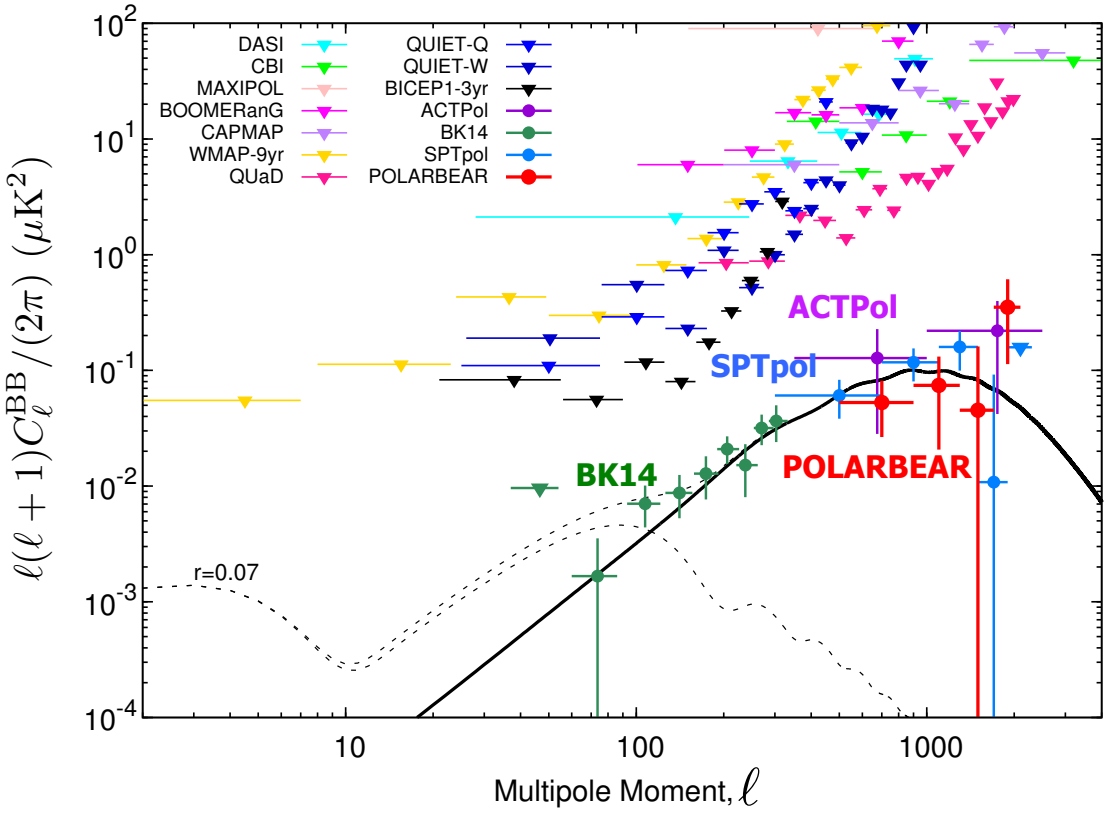
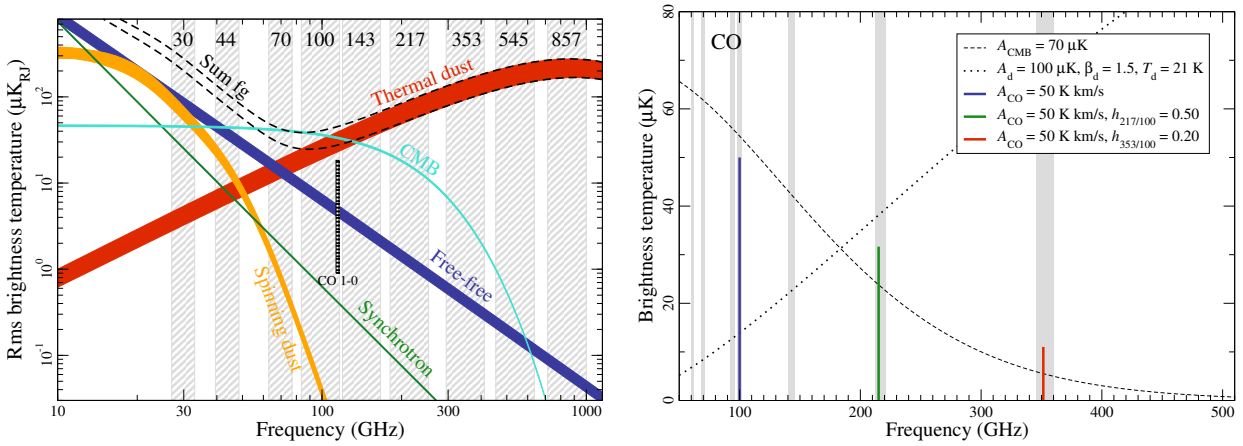


Figure 1.14: The B-mode power spectrum today. Triangles indicate upper limits. The solid black line is the ΛCDM B-mode power spectrum due to gravitational lensing of primordial E-modes. The lower dashed line shows the gravitational-wave contribution to the B-modes assuming a tensor-to-scalar ratio $r = 0.07$. The upper dashed line shows the sum of the B-modes from lensing and from gravitational waves. All current measurements are consistent with lensing B-modes *only*. (Image credit: Yuji Chinone.)



(a) Intensity foregrounds.

(b) CO-line intensities.

Figure 1.15: Galactic foregrounds [79].

The state-of-the-art measurements of the B-mode spectrum are shown in Fig. 1.14. Detections have been made in the range $500 \lesssim \ell \lesssim 2500$ by the ACTPol [59], POLARBEAR [8] and SPTpol [51] collaborations and in the range $70 \lesssim \ell \lesssim 300$ by the BICEP and Keck Array collaborations [18].

1.5 Foregrounds

Precision measurements of the CMB are impeded by foreground emission from a variety of sources. For a ground-based experiment, the brightest foreground is simply the atmosphere. Oxygen and water both have several resonances at millimeter wavelengths. For this reason, ground-based experiments are typically located in high-altitude deserts like the Atacama Desert in Chile and the South Pole. To avoid the atmosphere, many CMB experiments have been flown on balloons that achieve an altitude of ~ 40 km. The disadvantages of a balloon-based experiment include limited weight and time. Typically, the balloons fly for no more than two weeks. To date, there have been three satellite-based CMB experiments: COBE, WMAP and *Planck*. The observing conditions in space are excellent, since there is no atmosphere. The budgets of these projects allow for several years of observations, which is competitive with ground-based experiments. A disadvantage is the inability to perform maintenance and the strict requirements on technological readiness.

Even in space, however, CMB experiments must contend with galactic foregrounds such as thermal emission from dust, synchrotron radiation, rotational modes of carbon monoxide (CO) and bremsstrahlung or free-free emission. A summary of galactic foregrounds from the *Planck* collaboration is shown in Fig. 1.15. The foreground minimum is near 80 GHz, but the foreground contribution is still substantial at this frequency. To characterize the foregrounds, broad frequency coverage is required. With multiple frequency bands, the foreground components can be fit and subtracted. At frequencies significantly lower or higher than 80 GHz, the signal is dominated by foregrounds. The plots in Fig. 1.15 show the

intensity of the foregrounds. *Planck* does not detect polarization in spinning dust, free-free emission or CO [79]. As measurements of CMB polarization become more precise, it will be necessary to characterize the *polarization* properties of the galactic foregrounds. This field is in its infancy but is crucial to the success of future CMB polarimetry experiments.

Chapter 2

Overview of CMB experiments

In this chapter, we describe several experiments in which the Berkeley CMB group is involved.

2.1 POLARBEAR

The POLARBEAR[1] project is to map the CMB on angular scales that cover both the primordial and the lensing B-modes [53]. Such an experiment would have the ability to delens its own data to uncover a low-amplitude inflationary signal. The POLARBEAR telescopes are located in the Atacama Desert at an altitude of approximately 5000 m.

2.1.1 POLARBEAR-1

The first generation of the POLARBEAR project was deployed to the Atacama Desert in 2012. In its first season, POLARBEAR observed three $3^\circ \times 3^\circ$ patches and achieved a map depth sufficient to make one of the first detections of the lensing B-modes [5, 6, 81]. In 2014, a warm (room-temperature) half-wave plate was installed at the primary focus to modulate polarized signals. At the same time, the observation patch was widened to be similar to BICEP's observations [20] and allow access to lower-ℓ modes.

A half-wave plate is a birefringent material whose thickness is controlled so that the electrical length for one linear polarization is one half-wavelength longer than for the other linear polarization. A plane wave that passes through the plate will have its polarization flipped about a particular axis. Signals which are on average unpolarized will pass through unchanged. If the half-wave plate is set spinning, then polarized signals will be modulated at a harmonic of the rotation frequency. Unpolarized signals will not be altered. In this way, a rotating half-wave plate can distinguish between polarized and unpolarized signals. Since the atmosphere is unpolarized, a half-wave plate can be used to filter out the atmosphere. By modulating the sky signal up to higher frequencies, the half-wave plate mitigates so-called $1/f$ noise, which is the noise due to drifts in the detector responsivity or due to atmospheric fluctuations. At the same time, a single polarized detector is now able to

[1] It is officially *not* an acronym, although it is written in all capital letters. Originally, it was meant to be an approximate acronym for POLARization of the Background Radiation.

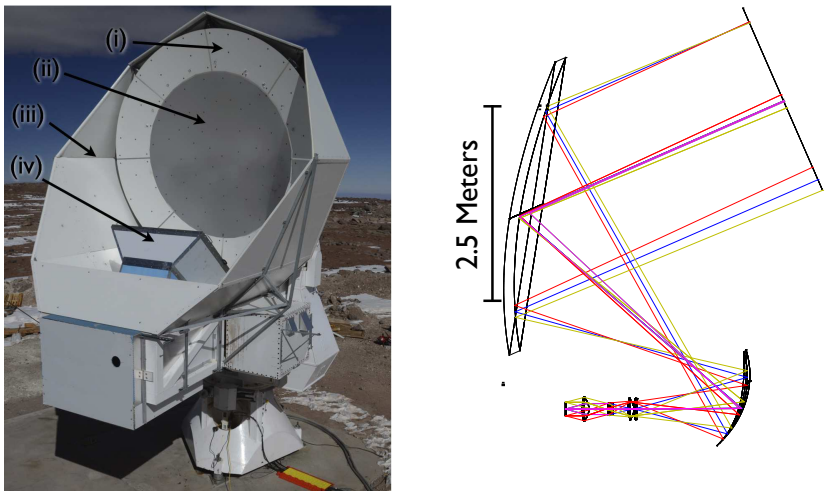


Figure 2.1: POLARBEAR telescope. (*Left*) A photograph of the telescope showing (*i*) the guard ring for the primary mirror, (*ii*) the primary mirror, (*iii*) the ground shield and (*iv*) the prime-focus baffle. (*Right*) A cross-sectional ray-trace diagram. (Image taken from Kermish et al. [53].)

measure *both* linear polarization on the sky, since the half-wave plate will rotate its on-sky polarization vector as a function of time. This removes the need for pair-differencing detectors of orthogonal linear polarizations, which in turn removes all of the systematics associated with asymmetries between the paired detectors. A disadvantage of a half-wave plate is that it contributes an effective non-zero temperature in the optical path. To reduce the extra optical loading from the half-wave plate, it is necessary either to reduce the loss tangent of the birefringent material and the anti-reflection coating or to cool the half-wave plate to cryogenic temperatures. Since there are additional engineering challenges associated with a *cryogenic* half-wave plate, POLARBEAR opted for a *warm*, i.e., ambient-temperature half-wave plate. The data taken with the half-wave plate is still being analyzed at the time of writing. A preliminary noise analysis, however, has been completed, which demonstrates the basic operation and benefits of the half-wave plate [101].

The POLARBEAR telescope, which is also called the Huan Tran Telescope, is an off-axis Gregorian-Dragone telescope and is shown in Fig. 2.1 along with a cross-sectional ray-trace diagram. The optically active area of the primary mirror has a diameter of 2.5 m, which allows for an angular resolution of approximately $3'$ at 150 GHz. The primary mirror creates an image on a secondary mirror. The focus of the primary mirror, which is where the rays converge, is called the *primary focus* and is where the warm half-wave plate was installed. The secondary mirror focuses the light onto a set of three lenses, which constitute the *optics tube*. The optics tube is part of the cryogenic receiver, which includes the detectors. The lenses are cooled to reduce the thermal radiative load on the detectors. Interspersed with the lenses are several low-pass filters, which reflect out-of-band radiation back to the sky. The cryostat window is made of foam, which is both microwave transparent and mechanically robust enough to hold vacuum. The three lenses focus the light onto the *focal plane*, which

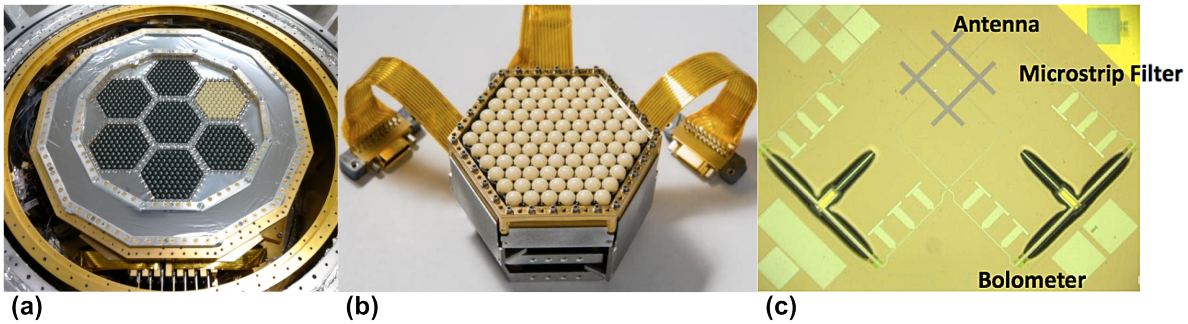


Figure 2.2: The POLARBEAR-1 focal plane. (*a*) The full focal plane showing seven wafers. The one white one uses alumina lenslets instead of silicon. We will discuss lenslets further in Ch. 4. (*b*) A wafer *module*, which integrates the detector wafer, the lenslet array and the *LC* resonators used for frequency multiplexing. When installed in the focal plane, the cables are connected to the cryogenic amplifiers. (*c*) A microscope photograph of the detectors and integrated planar antenna and microstrip bandpass filters. We will discuss this design in detail in Ch. 4. (Image credit: POLARBEAR collaboration.)

is where the detectors are located.

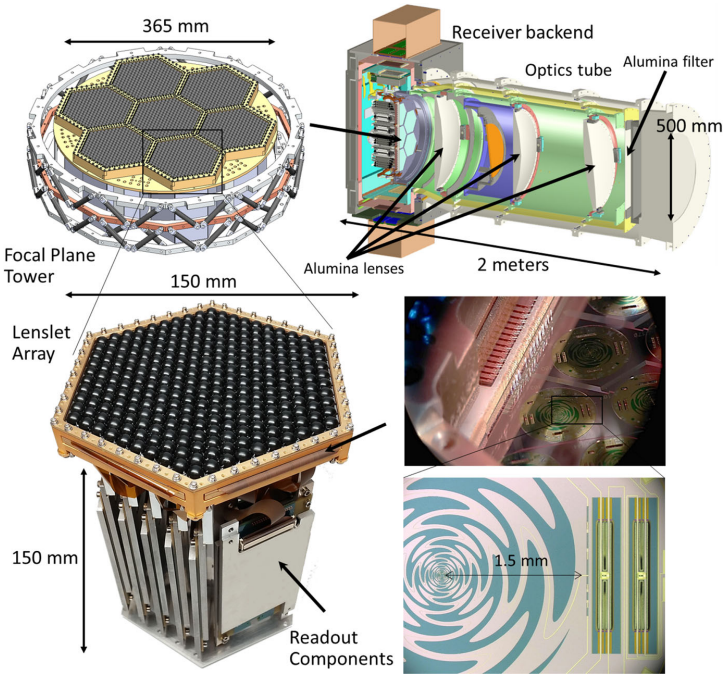
The focal plane of POLARBEAR-1 (PB-1) contains approximately 1000 bolometers and is shown in Fig. 2.2. The detectors are divided among seven wafers, each with a diameter of approximately $4'$. The detector design will be discussed in detail in Chs. 3 and 4. The PB-1 detectors are all sensitive to a bandwidth of approximately 30% centered on 150 GHz, which lies in the middle of an atmospheric window.

A Cryomech helium pulse-tube cooler (PTC) is used to cool the inner shells of the cryostat to 50 and 4 K. Inside the 4-K shell a ^{3}He sorption refrigerator made by Chase Research Cryogenics is used to cool the focal plane to 250 mK.

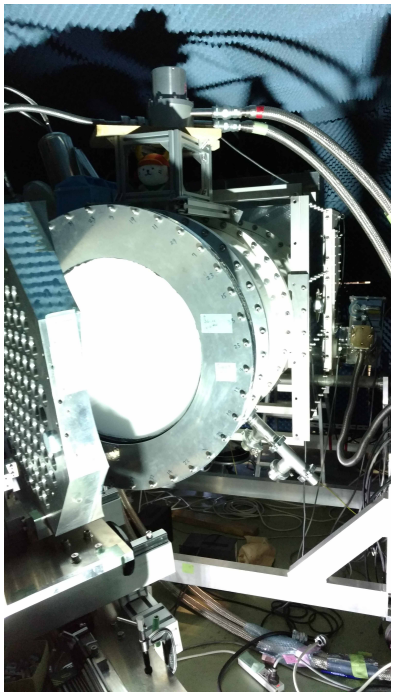
2.1.2 POLARBEAR-2

The next generation of the POLARBEAR project is called POLARBEAR-2 (PB-2) [100]. It will be located at the same site in the Atacama Desert with a similar telescope except for the receiver, which will contain approximately seven times more detectors with sensitivity to bands centered on both 90 and 150 GHz. Extending the frequency coverage to two bands introduces several challenges in the optics as well as in the detectors. Dielectric surfaces must be coated with additional layers of anti-reflection (AR) coating, the polarization modulator must contain multiple birefringent plates [46], and the detector antennas must achieve more than an octave in bandwidth. We call such a receiver *dichroic* to indicate that it is sensitive two colors or frequency bands. Receivers that are sensitive to more than one frequency band are called *multichroic*. We will discuss such detectors in great detail in Ch. 4. The increase in detector count allows for a greater mapping speed, i.e., greater experimental sensitivity, and the addition of a second frequency band aids in the characterization of galactic foregrounds.

PB-2 is being developed in collaboration with the CMB group at KEK (High Energy Accelerator Research Organization) in Japan, which is where the PB-2 receiver is being



(a) Some hardware developed for PB-2, much of which will be discussed in later chapters. (Montage taken from Suzuki et al. [100]) (*Clockwise from upper left*) CAD image of the focal plane, cross section of the optics tube, wire bonds to the detector wafer, sinuous antenna with TES bolometers and a wafer module.



(b) The PB-2 receiver being tested at KEK. This view shows the window of the optics tube. The receiver has been placed in a large microwave-absorbing chamber. At left is a mirror, which is focusing light into the receiver for optical tests.

Figure 2.3: POLARBEAR-2 hardware development.

integrated. Some of the hardware being developed for PB-2 is shown in Fig. 2.3, and a cross section of the optics tube is shown in Fig. 2.4. The focal plane consists of seven wafers just like PB-1, but these wafers are 6′ in diameter, which is an increase of approximately 50%. Each pixel contains four detectors: one for each linear polarization in each of two frequency bands. The optics tube is similar to that of PB-1, except that the lenses must be coated to reduce reflections across a significantly larger frequency range. The lenses and the window are also larger than those of PB-1. Due to the increased mass, PB-2 uses *two* pulse-tube coolers, one for the backend containing the focal plane and one for the optics tube. The focal plane is cooled with the same kind of ^{3}He sorption refrigerator and is also designed for a base temperature of 250 mK.

Simons Array

The Simons Array (SA) is an extension of PB-2 and will consist of three telescopes, which are referred to as PB-2a, PB-2b and PB-2c [100]. The first telescope, PB-2a, is just PB-2 itself. The second telescope will be a copy of PB-2, except that the warm half-wave plate

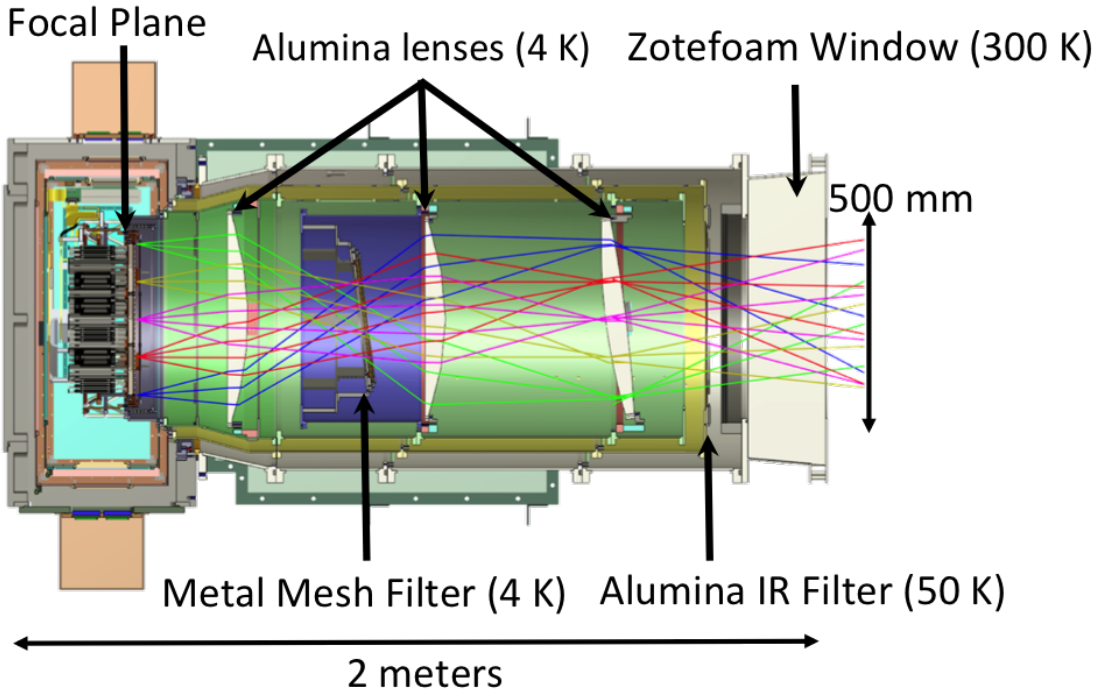


Figure 2.4: Cross section of the PB-2 receiver. (Image credit: POLARBEAR-2 collaboration.) The backend at left contains the detectors and readout electronics, which will be discussed in Chs. 3 and 4. The beige boxes above and below the backend house room-temperature controller boards for the cryogenic SQUID (superconducting quantum interference device) amplifiers. To the right is the optics tube, which contains three cryogenic lenses as well as several low-pass filters, i.e., metal-mesh and alumina filters, to reduce radiative thermal loading on the focal plane. The metal-mesh filters reflect high frequencies, while the alumina filter simply absorbs infrared radiation. For this reason, the alumina filter must be very well heat sunk to the 50-K shell. The foam window must be both microwave-transparent and mechanically strong enough to hold vacuum. Ray traces are also shown, where the outermost rays are meant to correspond to the -10-dB point of the detector beams. The inner walls of the optics tube are covered with absorptive material to prevent stray reflections.

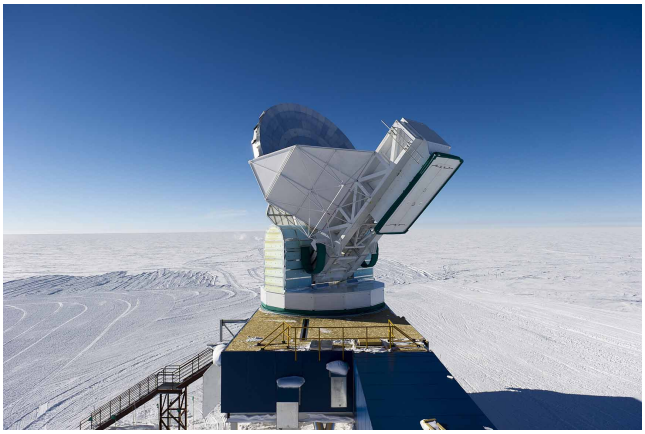
Figure 2.5: Simons Array under construction at the POLARBEAR site in the Atacama Desert. The middle telescope is the Huan Tran Telescope. The large structure in the upper right is the Atacama Cosmology Telescope (ACT). (Image credit: POLARBEAR collaboration.)

will be removed and a cryogenic half-wave plate will be installed. Cooling the half-wave plate lowers the radiative load on the focal plane and improves the experiment's sensitivity. Constructing a continuously rotating cryogenic half-wave plate, however, presents a large number of engineering challenges. The third telescope, PB-2c, will be similar to the first two but sensitive to frequency bands centered on 220 and 270 GHz. These channels will be used to characterize foreground contamination due to galactic dust. While two new telescopes have been built for PB-2a and PB-2b, the Huan Tran Telescope will be used for PB-2c. At that point, of course, PB-1 will be decommissioned. A photograph of the SA telescopes under construction in the Atacama Desert is shown in Fig. 2.5. In total, the SA will have more than 2×10^4 detectors with band centers at 90, 150, 220 and 270 GHz.

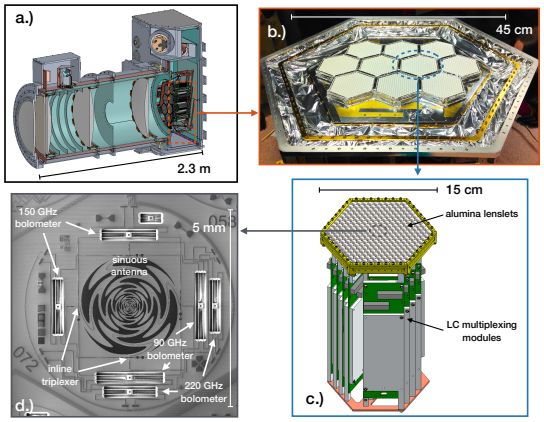
2.2 South Pole Telescope 3rd Generation (SPT-3G)

The most recent generation of the South Pole Telescope (SPT) is called SPT-3G [15]. The SPT, which is located at the geographic south pole, has a 10-m primary mirror and allows for an angular resolution of approximately $1'$ at 150 GHz. With such high angular resolution, SPT is able to resolve high-redshift galaxy clusters and access multipole moments as high as $\ell \sim 10^4$. SPT-3G is an upgrade from the previous generation, which was called SPTpol. For SPT-3G, the telescope mirrors remain the same, but a new receiver has been built, which contains more than 1.5×10^4 detectors sensitive to bands centered on 90, 150 and 220 GHz. The focal plane makes use of trichroic lenslet-coupled sinuous-antenna pixels, the design of which was originally developed for PB-2.[2] The wafer modules, which integrate the detectors

[2]At the focal-plane level, the extension from two bands to three is relatively trivial. These sinuous-antenna multichroic pixels will be discussed in great detail in Ch. 4.



(a) The South Pole Telescope.



(b) Some hardware developed for the SPT-3G receiver. (*a*) Cross section of the optics tube. (*b*) Focal plane. (*c*) CAD image of a wafer module. (*d*) Trichroic sinuous-antenna pixel with TES bolometers (see Ch. 4).

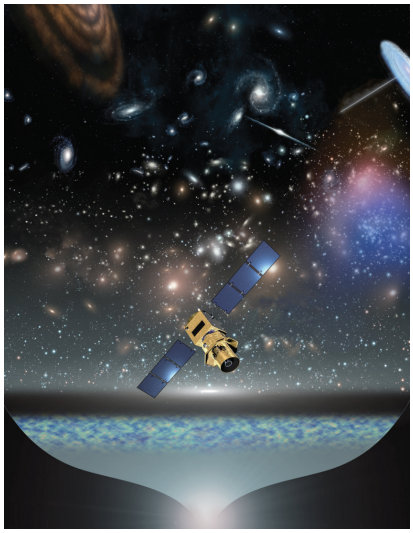
Figure 2.6: South Pole Telescope 3rd Generation (SPT3G). (Image credits: SPT collaboration.)

and some of the cold electronics, are also based on designs developed for PB-2. The PB-2 designs were shared with the SPT collaboration in the early stages of the SPT-3G project. The focal plane is cooled to 250 mK with the same type of ^{3}He sorption refrigerator used in POLARBEAR. Like PB-2, some of the major challenges revolve around the anti-reflection coating of the large-diameter lenses. In the case of SPT-3G, this is even more difficult, since the total bandwidth is larger. Unlike POLARBEAR, SPT does not make use of a half-wave plate. Some hardware developed for SPT-3G is shown in Fig. 2.6. The SPT-3G receiver was deployed to the South Pole in December 2016. The first season was plagued by spurious noise in the readout electronics. Replacement hardware was installed beginning in December 2017, which appears to have reduced the noise to close to expected levels.

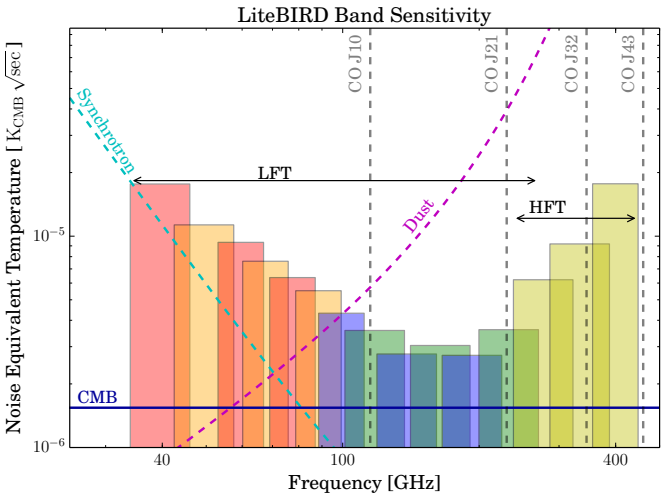
2.3 LiteBIRD

The Lite satellite for the studies of B-mode polarization and Inflation cosmic background Radiation Detection (LiteBIRD) is a proposed satellite telescope that will be launched by the Japan Aerospace Exploration Agency (JAXA) with significant contributions from NASA [64]. The US team submitted a Concept Study Report (CSR) to NASA in July 2016 and has been funded for three years of technology development. LiteBIRD is designed to target the low-ℓ B-modes, especially the recombination peak at $\ell \sim 100$. Ideally, LiteBIRD will also be able to access the reionization peak at $\ell \sim 5$. The threshold mission will produce an upper limit on the tensor-to-scalar ratio of $r \lesssim 10^{-3}$.

A measurement of primordial B-modes at the level of $r \sim 10^{-3}$ will require precise foreground removal. For this purpose, LiteBIRD will be sensitive to 15 spectral bands spanning



(a) Cover image for the CSR.



(b) Frequency bands with foreground contributions indicated. The height of each band indicates the expected noise-equivalent temperature (NET).

Figure 2.7: LiteBIRD. These images were taken from the LiteBIRD Concept Study Report (CSR), which was submitted in July 2016 to NASA.

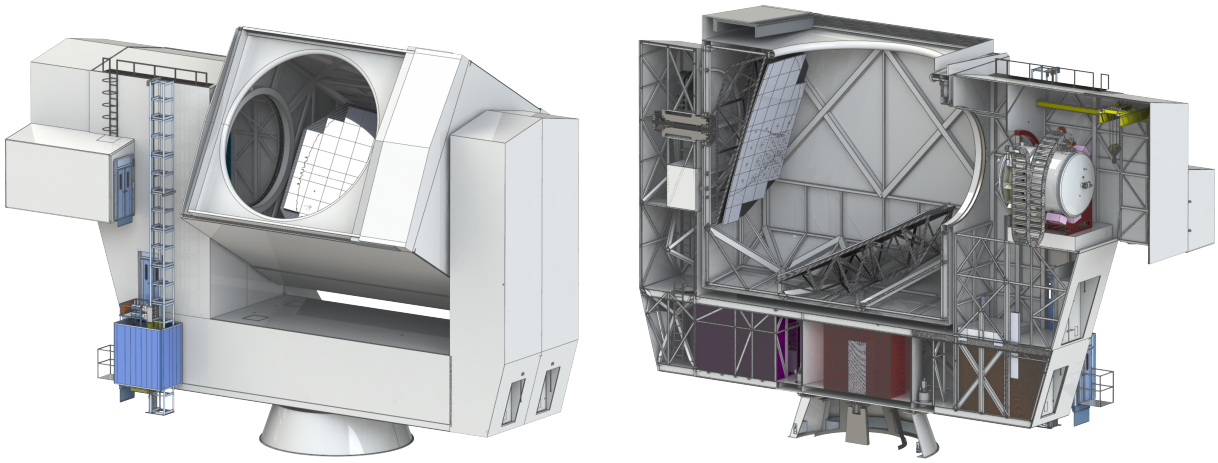
more than a decade in frequency. The bands are shown in Fig. 2.7. The low frequencies will help mainly to constrain galactic synchrotron emission, and the high frequencies will help mainly to constrain galactic dust emission. Altogether, the LiteBIRD experiment will contain approximately 2000 bolometers cooled to 100 mK.

The launch date is proposed for the mid-2020s with a baseline mission duration of three years.

2.4 Simons Observatory

The Simons Observatory (SO) is a project funded mainly by the Simons Foundation to build next-generation CMB telescopes in the Atacama Desert. The collaboration consists mainly of members of the POLARBEAR and Atacama Cosmology Telescope (ACT) collaborations. Large-aperture telescopes with primary-mirror diameters of 6 m will be built to provide sensitivity to high-ℓ modes as well as high-redshift galaxy clusters. The design of the large-aperture telescope (LAT) is shown in Fig. 2.8. Small-aperture telescopes will be built to provide sensitivity to low-ℓ modes. The project is currently transitioning from a planning phase to a fabrication and testing phase. First light is expected in 2021.

Most likely, two detector technologies will be employed: the dichroic feedhorns developed at NIST, Boulder and the lenslet-coupled sinuous antennas developed at UC Berkeley. The focal planes will be cooled to 100 mK with dilution refrigerators.



(a) Exterior. (b) Cross section.

Figure 2.8: Drawings of the Simons Observatory large-aperture camera (LAT) as of March, 2018. (Image credit: SO collaboration.)

2.5 CMB Stage 4

The CMB community is currently planning to propose to the Department of Energy (DOE) a project that will bring together the majority of US-based CMB cosmologists. Telescopes will most likely be built in multiple locations. The collaboration has produced a Science Book [1] and a Technology Book [2]. The Concept Definition Task Force (CDT) has submitted a report to the Astronomy and Astrophysics Advisory Committee (AAAC).

Chapter 3

Detectors for CMB cosmology

Oh, Langley devised the bolometer.
It's really a kind of thermometer
That can measure the heat
From a polar bear's feet
At a distance of half a kilometer.
– From the infrared oral tradition

In this chapter, we will discuss the design and fabrication of detectors, for which the heat from a polar bear's feet is a booming signal at a distance of half a kilometer. We call these detectors *bolometers*, and they are, as the limerick suggests, really just very sensitive thermometers. If they are allowed to absorb infrared or microwave radiation, then they become very sensitive optical sensors.

The bolometers that we are regularly fabricating have a per-detector noise-equivalent power (NEP) on the order of 100 aW/$\sqrt{\text{Hz}}$. If this is white noise, then a power difference of 100 aW can be seen with a signal-to-noise ratio of unity in a half-second of integration. To gain some intuition for how sensitive these detectors are, let's compare this noise level to the heat from a polar bear's feet at a distance of half a kilometer. To detect the polar bear's feet, we need to be able to detect excess power above the background radiation, which in this context is simply the thermal radiation from the landscape in which the polar bear is roaming. We will take the temperature of the landscape to be the freezing point of water, i.e., 0° C, although this is perhaps warmer than the environments in which polar bears are typically found. We take the temperature of the polar bear's feet to be similar to human body temperature, i.e., 37° C. So we are looking for a temperature difference on the order of 37 K. One caveat in this little calculation is that a major source of noise in our measurements comes from the optical signal itself. This is called *photon noise* and is due to the variance in the photon arrival rate. The noise levels that we achieve, i.e., $\mathcal{O}(100)$ aW/$\sqrt{\text{Hz}}$, are only possible with a *total* incident optical power, i.e., integrated over all lines of sight, at the level of $\mathcal{O}(10)$ pW. The polar bear will occupy only a tiny solid angle, but the detector will receive radiation from other directions, all of which contribute to the photon noise. To reduce the optical loading to the level of 10 pW, it is necessary for the background temperature to be at the level of $\sim$ 10 K and for the detector to be sensitive to only one electromagnetic mode per

frequency, i.e., to be *single moded*, over a roughly 30-GHz bandwidth in the Rayleigh-Jeans portion of the spectrum. But none of that has to do with the intrinsic detector noise. In the case of the polar bear's feet, where the background temperature is ten times larger than we typically design for, the photon noise will completely dominate every other noise source and become the limiting factor in this measurement. To demonstrate how sensitive our detectors are, then, it is misleading to include in the noise budget the contribution from the optical signal. We will instead compare the size of the signal to the intrinsic detector noise, which we take here to be on the order of 100 aW/$\sqrt{\text{Hz}}$. Denote the detector's absorbing area as $A_{\text{det.}}$. Denote the projected area of the polar bear's feet as $A_{\text{p.b.}}$. Denote the distance to the polar bear as r. Then the solid angle occupied by the polar bear's feet is $\Omega = A_{\text{p.b.}}/r^2$. Take the absorber to be sensitive from a frequency ν_{low} to ν_{hi}, and assume that these frequencies are in the Rayleigh-Jeans portion of the spectrum, which is true for CMB detectors observing a room-temperature source. Assume that the optical power couples to the detector through an absorber, so this is a *multi-moded detector*, i.e., the only restrictions on the number of electromagnetic modes come from the frequency range and the optical throughput $A_{\text{det.}}\Omega$. Then the power reaching the detector from the polar bear's feet is

$$P = \frac{2k_B T A_{\text{det.}}\Omega}{c^2} \int_{\nu_{\text{low}}}^{\nu_{\text{hi}}} d\nu \; \nu^2. \tag{3.1}$$

The difference in power between the background with and without the polar bear's feet is

$$\Delta P = \frac{2k_B A_{\text{det.}} A_{\text{p.b.}}}{3c^2 r^2} \left(\nu_{\text{hi}}^3 - \nu_{\text{low}}^3\right) \Delta T, \tag{3.2}$$

where we have taken the integral and written the solid angle in terms of more interesting parameters. So now we can plug in some numbers. Take the polar bear's feet to be something like 10 cm on a side, so $A_{\text{p.b.}} = 100$ cm^2. We take $r = 0.5$ km as the limerick instructs. We have some freedom in the choice of the detector area and frequency range. It is not difficult to build bolometers which absorb radiation between, e.g., 100 and 1000 GHz. The signal will increase approximately as the cube of ν_{hi}, so there is a large premium on increasing the highest frequency. Very conservatively, we take the frequency range to be 100-200 GHz. We take the absorber area to be 1 cm^2; for reference, the POLARBEAR pixels are 1.5 cm^2 in area, although they are not simple absorbers. In this case, we find $\Delta P = 1.1 \times 10^2$ aW, which is approximately the same as the RMS fluctuations in a half-second of integration due to the detector noise only. If we increase the frequency range to, e.g., 500-1000 GHz (so the *fractional* bandwidth is the same), then the signal strength increases dramatically: $\Delta P = 1.3 \times 10^4$ aW. In this case, if the detector were the only noise source, we could detect the heat from a polar bear's feet at a distance of half a kilometer with a signal-to-noise ratio on the order of 100 in only a half-second of integration.

3.1 Bolometer

A bolometer is a detector that converts an incident signal to heat and measures the resulting temperature change. Most commonly, a temperature-dependent electrical resistance, i.e., a thermistor, is used as a proxy for measuring temperature [56]. At cryogenic temperatures,

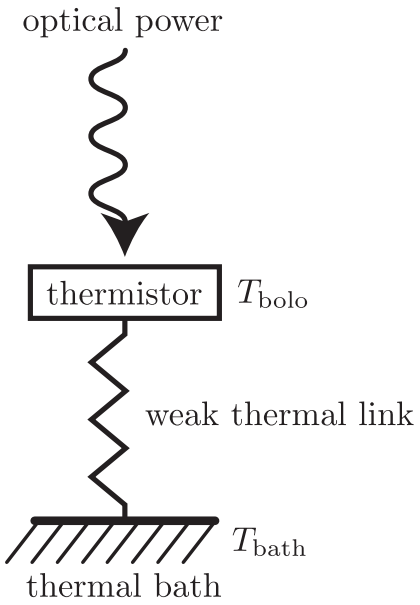


Figure 3.1: A sketch of a bolometer. Optical power is absorbed on a thermistor, whose resistance is monitored electrically as a proxy for temperature T_{bolo}. The electrical current used to monitor the thermistor causes Joule heating, which injects additional power onto the thermistor. We identify the thermistor as the bolometer. Power is allowed to dissipate through a weak thermal link to a thermal bath at temperature T_{bath}.

a semiconductor has an electrical resistance that increases exponentially as the temperature decreases. The steep temperature dependence can be exploited to build a very sensitive thermometer. The thermometer is read out simply by measuring the electrical resistance of the semiconductor. Another common choice is the transition-edge sensor, which we will discuss below and operates at the superconducting transition temperature of a metal film, where the resistance is decreasing dramatically as temperature decreases. To prevent large temperature excursions, the bolometer is allowed to dissipate power to a thermal reservoir as shown in Fig. 3.1. The incident power is absorbed onto the thermistor, which we are identifying with the bolometer. The bolometer temperature T_{bolo} will tend to increase, but some power dissipates to the thermal bath at temperature $T_{\mathrm{bath}} \leq T_{\mathrm{bolo}}$. The amount of power conducting to the bath is controlled by a weak thermal link with thermal conductance G, which is in general a function of temperature. The rate of change of T_{bolo} can be suppressed by increasing the bolometer's heat capacity C, which is also a function of T_{bolo}. At the same time, there is electrical power being dissipated on the bolometer through Joule heating from the readout electronics that are monitoring the thermistor resistance. The electrical power depends on T_{bolo}, since the resistance is temperature-dependent. Then the bolometer temperature is described by

$$C(T_{\mathrm{bolo}})\frac{dT_{\mathrm{bolo}}}{dt} = P_{\mathrm{opt.}} + P_{\mathrm{elec.}}(T_{\mathrm{bolo}}) - \int_{T_{\mathrm{bath}}}^{T_{\mathrm{bolo}}} dT\ G(T), \tag{3.3}$$

where the left-hand side is simply the rate of change of the energy stored in the bolometer and the last term on the right is the power conducting out of the bolometer through the weak thermal link. By a clever choice of the form of $P_{\text{elec.}}(T_{\text{bolo}})$, we can ensure that there is a stable equilibrium bolometer temperature. We call this *electrothermal feedback*.

3.1.1 Transition-edge sensor (TES)

The transition-edge sensor (TES) is a type of bolometer that exploits the steep temperature dependence of the resistance of a metal film near its superconducting transition. If the TES can be operated near its critical temperature T_c, then small changes in temperature are translated into relatively large changes in resistance. To operate stably in its superconducting transition, a TES may be voltage biased. In this case, the power entering the bolometer island is a combination of the optical signal we wish to detect and the electrical power supplied by the readout electronics. A steady state arises when the thermal power flowing from the island to the bath is exactly balanced by the power entering the island, i.e., when

$$P_{\text{cond.}} = P_{\text{opt.}} + P_{\text{elec.}}, \tag{3.4}$$

where $P_{\text{cond.}}$ is the power conducting through the weak thermal link to the bath, $P_{\text{opt.}}$ is the incident optical power and $P_{\text{elec.}} = V^2/R$ is the power supplied by the voltage bias, where V is the voltage bias applied across the TES of operating resistance R, i.e., in the transition. To stay at the superconducting critical temperature, we want the sum on the right-hand side to remain constant. The condition is

$$P_{\text{opt.}} + \frac{V^2}{R} = \text{const.} \tag{3.5}$$

But the resistance varies with temperature. If the optical power increases, the TES warms up, which increases R and decreases $P_{\text{elec.}}$. If the optical power decreases, the TES cools down, which decreases R and increases $P_{\text{elec.}}$. This see-saw behavior keeps the power flowing into the bolometer island constant. We refer to this mechanism as *electrothermal feedback*.[1]

T_c curve

The shape of the superconducting transition is of interest to TES engineers, since the steepness controls the responsivity, the linearity and the stability of the detector. We can map out the superconducting by finely controlling the temperature of a metal film. An example is shown in Fig. 3.2, where the temperature has been controlled at the sub-mK level. We see that the transition in this case occurs over just a few mK, which translates to less than a percent in absolute temperature. If the temperature were controlled coarsely, the R-T relationship would look like a step function. Notice that there is a steep and roughly linear regime about halfway through the transition. Notice also that the resistance is not constant even above the critical temperature. This temperature dependence above the transition is

[1] For a semiconductor bolometer, the resistance *decreases* with increasing temperature, i.e., dR/dT has an opposite-sign slope compared with a TES. Electrothermal feedback can be achieved in this case by *current* biasing the bolometer, so $P_{\text{elec.}} = I^2R$.

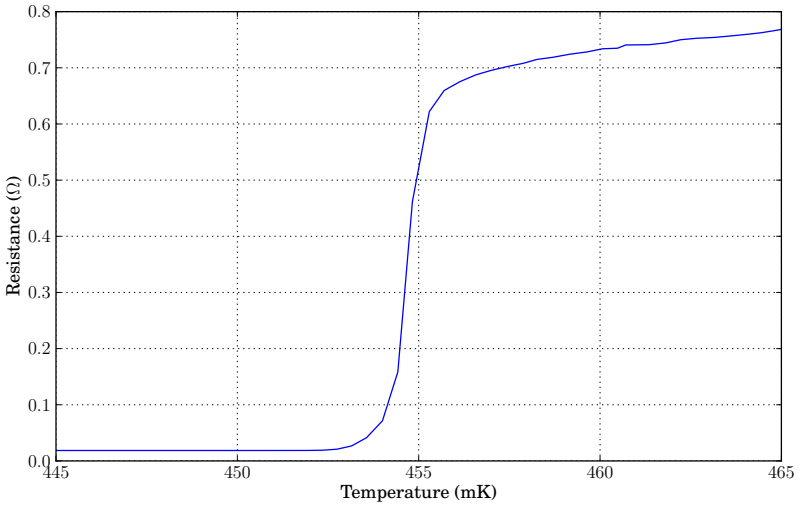


Figure 3.2: A superconducting transition from an AlMn film. The transition is steep but can be resolved with fine temperature control. Notice the slight temperature dependence even above the critical temperature. Also notice the residual resistance below the transition, which is likely an indication of some parasitic series resistance in the readout circuit.

useful for measuring optical signals that are too strong to allow the TES to cool to its superconducting transition. We say that such TESs are *saturated*. Of course, we design the TESs so that they do not saturate in operation on the telescope. For laboratory tests, however, it is common to saturate the detectors, since room-temperature thermal radiation is much stronger than optical loading from the sky.

When measuring a T_c curve, it is usually recommended to scan in both directions in temperature, i.e., make one measurement with temperature slowly increasing and one with temperature slowly decreasing. This helps to check whether there is a poor thermal coupling between the thermometer and the metal film, in which case the metal will be delayed in achieving the same temperature as the thermometer. When two such scans are made, it is typical to see one shifted slightly to the right and one slightly to the left, which is an indication of a small thermal delay. To the extent that the two curves are consistent, we can be reasonably confident in our measurement.

A TES that has cooled below its superconducting transition is said to *latched*. In Fig. 3.2, we see that there is some residual resistance below the transition. The residual resistance is usually a parasitic resistance from other components in the readout circuit. If some of it is on the bolometer island, then it can be used to bring the TES back up above its superconducting transition by applying a sufficiently large voltage; the Joule heating will warm up the TES, the metal will become normal, and the voltage bias can be decreased to lower the TES into the transition. If there is no residual resistance on the island, then the only way to get the TES back into its superconducting transition is to heat the thermal bath above the critical temperature. While the bath is at this elevated temperature, a voltage bias is applied to the now normal-metal TES. Then the bath temperature is lowered, but the TES stays at an

elevated temperature due to the electrical power supplied by the voltage bias.

I-V characteristic

The TES has a distinctive current-voltage relationship, which we often call an I-V curve. An example is shown in Fig. 3.3(a), which shows the characteristic "check mark" of a TES. Above the superconducting transition, the TES behaves like a normal metal obeying Ohm's law, i.e., $I \propto V$. This is the linear portion of the I-V curve of Fig. 3.3(a) at relatively large voltage bias. As the voltage bias is decreased, the TES is lowered into its transition. In the transition, as we discussed above, the electrical power $P_{\text{elec.}}$ is actually constant, so long as the optical power $P_{\text{opt.}}$ is constant. In this regime, we have $I \propto 1/V$, so that $P_{\text{elec.}} = IV = \text{const.}$ We see the beginning of this regime at the lowest bias voltage in Fig. 3.3(a).

Calibrating measurements of the I-V curve is slightly non-trivial, because the SQUID output has an arbitrary offset, i.e., the SQUID output really only provides information about *changes* in the current but not about the absolute amplitude. Furthermore, the SQUID has an *a priori* unknown transimpedance that converts the actual current running through the input coil to an output voltage. The transimpedance can be calibrated by using a known temperature-independent resistance in place of a TES. The absolute amplitude of the current can be extracted by requiring that the linear portion of the I-V curve is, in fact, an instance of Ohm's law, so the line must intercept the y-axis at zero current.

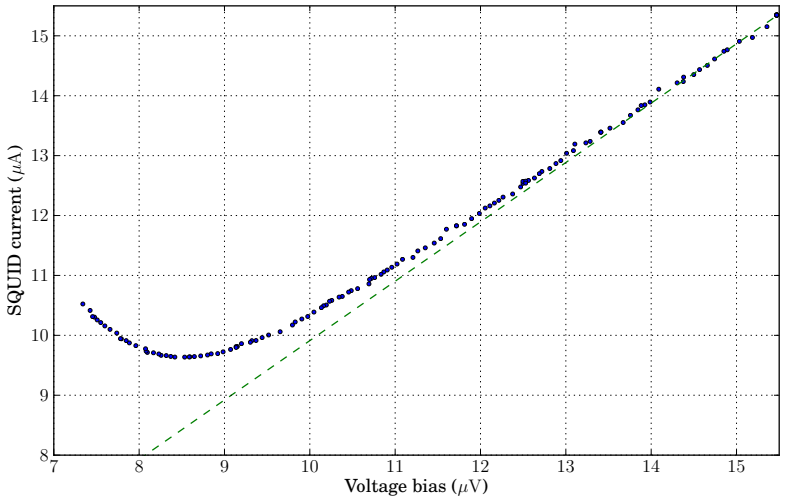
We can convert the I-V curve to a P-R (power-resistance) curve with the relations $P = IV$ and $R = V/I$. The P-R curve derived from the I-V curve of Fig. 3.3(a) is shown in Fig. 3.3(b). Here we see even more clearly the constant-resistance regime, where the TES acts as a normal metal, and the constant-power regime, where the TES is stabilized by electrothermal feedback at its superconducting critical temperature.

Saturation power

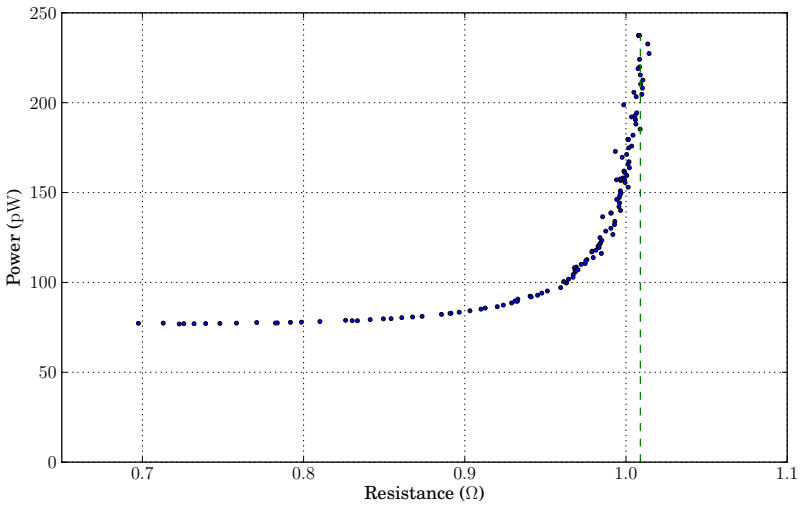
The *saturation power*, often denoted $P_{\text{sat.}}$ is the power that must be supplied to the TES island to keep it out of the superconducting transition. The flat line that the P-R curve asymptotes to at low R is a good definition of the saturation power. The saturation power is the thermal power conducted from the TES to the thermal bath when the TES is at its critical temperature. We can control the saturation power by modifying the thermal conductance. Fine adjustments are usually made with geometric parameters like the length and width of the weak thermal links. We design our TESs so that approximately half of the power entering the TES island in operation is from an optical signal and half from the voltage bias. For most current experiments, this translates to saturation powers in the range of 5-25 pW for frequency bands centered on 90-220 GHz.

3.1.2 Frequency multiplexing

For experiments containing large numbers of bolometers, it is useful to introduce a *multiplexing* scheme. By this, we mean an architecture for sharing wires and amplifiers among many detectors. This is beneficial for a number of reasons. It reduces the total number of amplifiers, which reduces the cost of the experiment. It reduces the thermal loads on the



(a) I vs. V.



(b) P vs. R. This plot contains no additional information beyond the I-V curve, since $P = IV$ and $R = V/I$.

Figure 3.3: TES electrical characteristics. With a relatively large voltage bias, the TES acts like a normal Ohmic metal. As the voltage bias is decreased, the TES enters the superconducting transition, where electrothermal feedback ensures in the absence of an optical signal that the applied electrical power is constant.

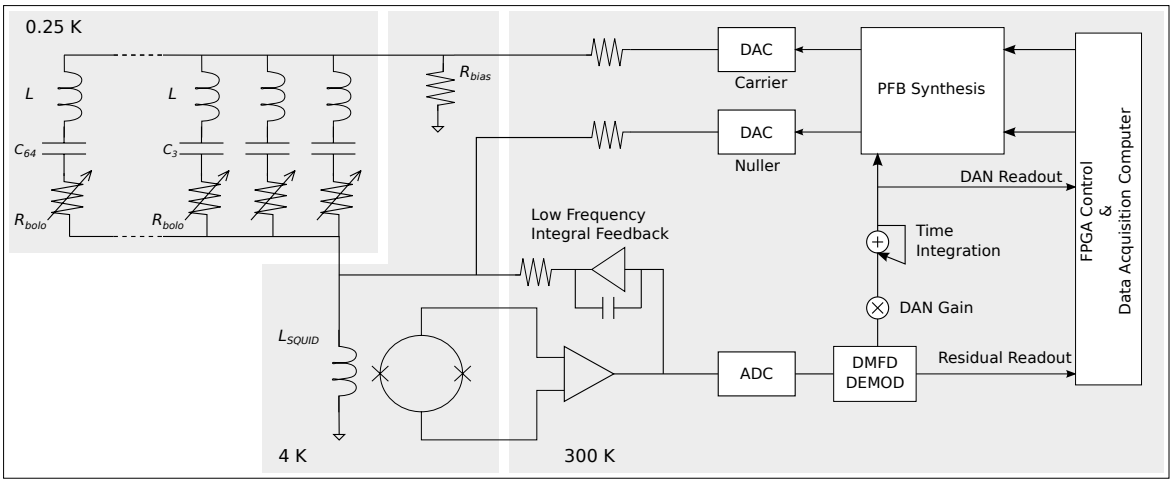


Figure 3.4: Circuit diagram for digital frequency multiplexing (DfMUX). The LC filters are in the upper left. The SQUID amplifier is in the bottom center. The temperature of each part of the circuit is indicated. (Image credit: Bender et al. [13].)

detector temperature stage, since fewer wires are required. It also simplifies the experiment by reducing the number of physical parts.

The particular flavor of multiplexing that is used in the POLARBEAR and SPT collaborations is called *frequency multiplexing* (fMUX). Each bolometer is associated with a resonant frequency. We construct an electrical resonant circuit consisting of an inductor L and a capacitor C in series with a TES of resistance R. So we have an LCR circuit. The width of the resonance in frequency space is proportional to R. We connect multiple such resonances in parallel but tune the resonant frequencies so the passbands do not overlap significantly. The same voltage supply can be used to bias the bolometers, but the bias voltages are now AC carrier signals tuned to the resonant frequencies. The signals exiting the amplifier are demodulated at these resonant frequencies. In this way, we separate our detectors in frequency space but are able to use the same wires and amplifiers for multiple simultaneously. A circuit diagram is shown in Fig. 3.4. Since the carrier signals are now synthesized digitally, we call the current version *digital frequency multiplexing* (DfMUX) [28]. The number of resonances per amplifier is called the *multiplexing factor*. For PB-1, a multiplexing factor of 8 was used. For PB-2, we use a multiplexing factor of 40, and SPT-3G uses 68.

Some of the hardware associated with DfMUX is shown in Fig. 3.5. The mother boards containing the field-programmable gate arrays (FPGAs) are designed by the CMB group at McGill University. Each board is responsible for a certain number of amplifiers. The LC resonators are fabricated commercially based on a design from the Berkeley CMB group [87]. The SQUID amplifiers are fabricated at NIST, Boulder and integrated on a PCB.

The frequency response of the circuit can be characterized simply by sweeping a carrier tone in frequency and reading the output. We often call this a *network analysis*. An example network analysis is shown in Fig. 3.6. We see a *comb* of resonances, which overlap only at the $\sim 10\%$ level. The resonances would overlap less and could be more tightly spaced if the

(a) DfMUX boards from the POLARBEAR-1 and SPTpol generation.

(b) *LC* chip (Image credit: Aritoki Suzuki.)

(c) SQUID card from the PB-1 and SPTpol generation. (Image credit: Quealy [84].)

Figure 3.5: DfMUX hardware.

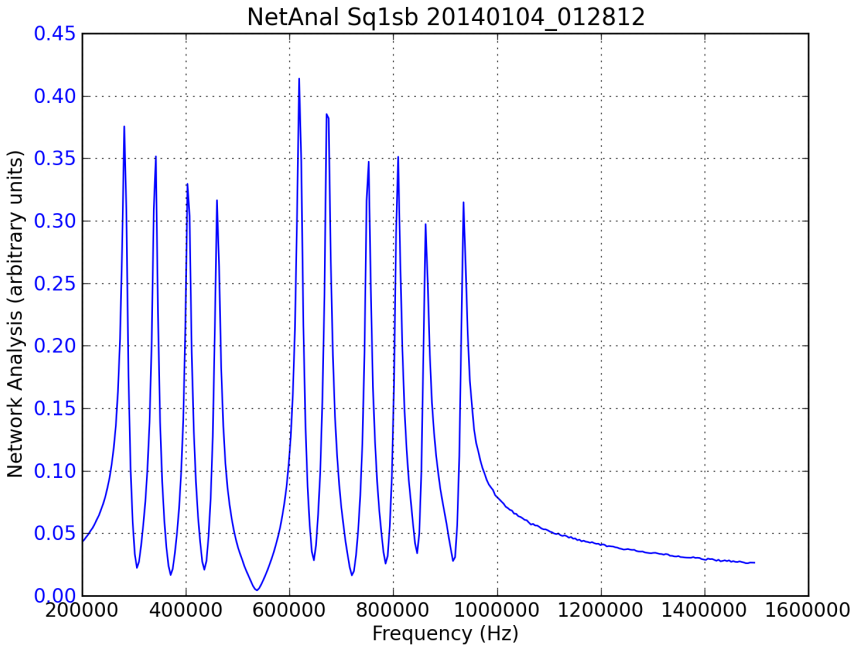


Figure 3.6: Network analysis.

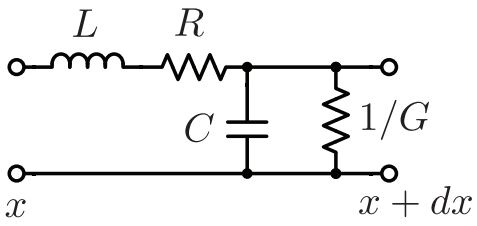


Figure 3.7: Transmission-line circuit model.

series resistance R were smaller. Since R is set by the TES operating resistance, one might ask why this resistance is not simply designed to be smaller. One reason is that there are parasitic series resistances in the circuit at a level that is approximately 10% of the current design TES resistance. We typically design for $R \approx 1\ \Omega$. If the parasitic resistances were reduced, we would be confronted with a problem regarding the electrical time constant of the circuit. As the resonance is made narrower, the electrical time constant increases, i.e., the circuit becomes slower. But we need the circuit to be at least as fast as the bolometer. So there is a limit to how narrow the resonance can be. The bolometer cannot be made arbitrarily slow, since it would eventually begin to affect the scan strategy of the telescope, i.e., the telescope would have to scan more slowly or the half-wave plate would have to rotate more slowly.

3.2 Integrated microwave components

Our bolometers are fabricated on silicon wafers using thin-film photolithographic processes. Several components are integrated on this wafer including a planar antenna and band-defining filters. The integration of these microwave components with the bolometers allows for a simplicity in the design of the focal plane. In this section, we discuss some of the microwave components that are integrated on the wafer.

3.2.1 Microwave engineering

Here we provide a crash course in microwave engineering with special emphasis on transmission lines.

We model a lossless transmission line as an infinite sequence of series inductors and shunt capacitors. We include loss in two places: in series with the inductor and in parallel with the capacitor. The circuit model is shown in Fig. 3.7. This is meant to be an infinitesimal segment of the transmission line. Locally, the values of the circuit components do not change with x. We have a *series inductance* L, a *series resistance* R, a *shunt capacitance* C and a *shunt admittance* G, all of which we take to be per-length quantities, e.g., the actual inductance is $L\,dx$. The transmission line consists of an infinite cascade of such segments. The input impedance Z_0 of this transmission line can be computed by insisting that one additional segment does not change the input impedance of the infinite chain. Dropping

$\mathcal{O}(dx^2)$ terms, it is not too hard to derive

$$Z_0 = \sqrt{\frac{j\omega L + R}{j\omega C + G}}. \tag{3.6}$$

This quantity is often referred to as the *characteristic impedance* of the transmission line. An important limiting case is that of a *lossless* transmission line, for which $R = 0$ and $G = 0$. Then we have

$$Z_0 = \sqrt{\frac{L}{C}}. \tag{3.7}$$

What is remarkable here is that the impedance is pure real, although the circuit consists only of reactive components. This means that we can transfer power to the transmission line as though it were a resistor, the interpretation being that the power then propagates down the line.

We can determine how the voltage and current evolve with x by applying Kirchhoff's laws to the transmission-line segment shown in Fig. 3.7. This produces a pair of coupled differential equations called the *telegrapher's equations*:

$$\frac{dV}{dx} = -I(j\omega L + R), \qquad \frac{dI}{dx} = -V(j\omega C + G). \tag{3.8}$$

This system can be solved to find the following wave solutions for voltage and current:

$$V(x) = V_0^+ e^{-\gamma x} + V_0^- e^{\gamma x}, \qquad I(x) = \frac{V_0^+}{Z_0} e^{-\gamma x} - \frac{V_0^-}{Z_0} e^{\gamma x}, \tag{3.9}$$

where the propagation constant

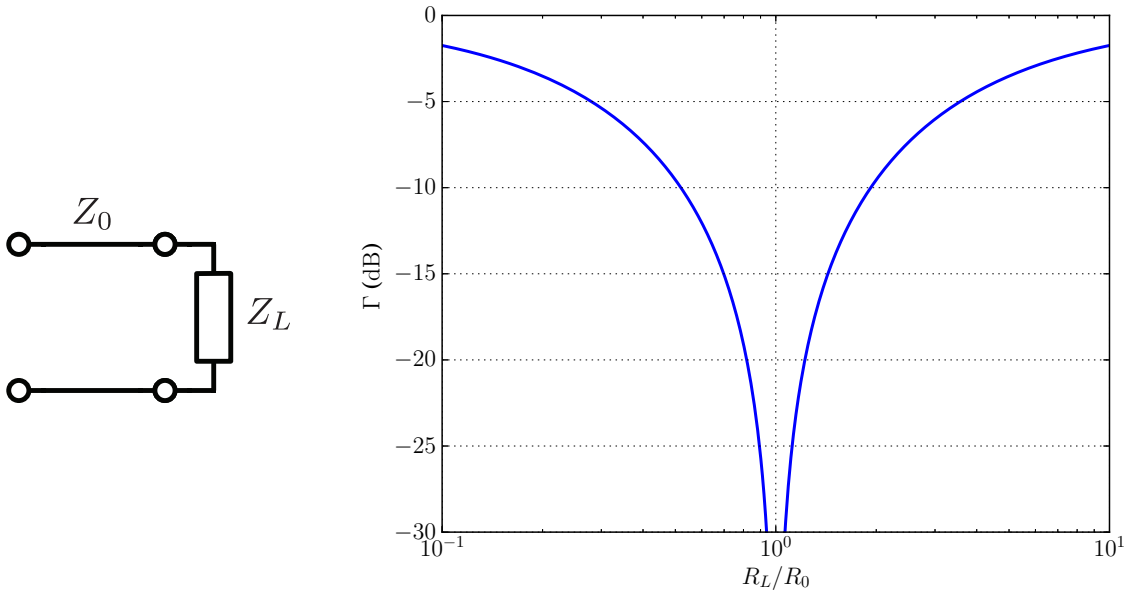
$$\gamma \equiv \sqrt{(j\omega C + G)(j\omega L + R)}. \tag{3.10}$$

The terms labeled with a plus sign are right-going waves, and the terms labeled with a minus sign are left-going waves. These are phasor quantities, for which the time dependence is implicit. When there is no loss, i.e., when $R = 0$ and $G = 0$, the propagation constant is pure imaginary, which produces unattenuated oscillatory solutions. When the circuit is lossy, however, the propagation constant picks up a real part, which contributes an attenuation that increases exponentially with distance.

Until now, we've imagined our transmission lines are infinite. It is time to see what happens when we terminate the lines. A transmission-line termination is shown in Fig. 3.8(a). The termination could be a lumped element such as a resistor. It could also be another transmission line with a different characteristic impedance. The load might also be an antenna, in which case the power is coupled out of the circuit and into free space. In this latter case, the antenna presents an effective load impedance, which we call the *antenna impedance*.

We imagine that the transmission line approaches the termination from $x = -\infty$ and define the *reflection coefficient*

$$\Gamma \equiv \frac{V_0^-}{V_0^+}. \tag{3.11}$$



(a) A terminated transmission line.

(b) The reflection coefficient for a real-impedance transmission line dumping power to a real-impedance load. Notice that the amplitude of the reflection is relatively insensitive to the impedance mismatch. Even a factor-of-2 mismatch only produces $\sim 10\%$ reflection.

Figure 3.8: Terminated transmission line.

The picture is that an incident wave comes from $x = -\infty$; some of the power is transferred to the load, while some is reflected back towards $x = -\infty$. The reflection coefficient Γ is a measure of the *amplitude* of the voltage of the reflected wave. For a lossless transmission line, the *power* will go as $|\Gamma|^2$.

We often measure the magnitude of the reflection coefficient in decibels. Since the power goes as $|\Gamma|^2$, the conversion to decibels is conventionally

$$|\Gamma| \to 20 \log_{10} |\Gamma| \text{ dB}. \tag{3.12}$$

The *return loss* is defined as

$$\text{RL} \equiv -20 \log_{10} |\Gamma| \text{ dB}. \tag{3.13}$$

This is a measure of the power that does not return to the input. If there is no reflection at all, the return loss is infinite; if the reflected wave has an amplitude equal to the incident wave, the return loss is 0 dB.

Set the load at $x = 0$. The load is just like a new medium with characteristic impedance Z_L that the signal is transmitted into. The boundary condition is $V(0)/I(0) = Z_L$, where the voltage and current are continuous across the boundary. It is not too difficult to apply these conditions to derive

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0}, \tag{3.14}$$

which is a Fresnel-like relation for the amplitude of the reflected wave. Notice that we get no reflection when the impedances are matched, i.e., $Z_L = Z_0$. For an open circuit, i.e., $Z_L \to \infty$, we have 100% reflection. For a short circuit, i.e., $Z_L = 0$, we also have 100% reflection but this time with a 180° phase shift.

Perhaps the most common case is to have a low-loss and, therefore, real-impedance transmission line dumping power to another real-impedance element such as another low-loss transmission line, a resistor or an on-resonance antenna. In this case, $Z_L = R_L$ and $Z_0 = R_0$. The reflection coefficient is plotted in Fig. 3.8(b). The reflection amplitude grows relatively slowly with the impedance mismatch. We still have only $\sim 10\%$ reflection even with a factor-of-2 impedance mismatch. So, except in high-precision applications, the impedance matching is quite forgiving.

The reflected power is also sometimes parameterized by the *standing-wave ratio*

$$\text{SWR} = \frac{1 + |\Gamma|}{1 - |\Gamma|}. \tag{3.15}$$

The wave on the $x < 0$ side of the boundary can be decomposed into a standing-wave component and a traveling-wave component. The more reflection, the greater the amplitude of the standing wave. The standing-wave ratio is a ratio of the amplitude at an antinode of the standing wave to the amplitude at a node of the standing wave. The standing-wave ratio is in one-to-one correspondence with the magnitude of the reflection coefficient and is a popular parameterization in the microwave-engineering literature.

3.2.2 Electromagnetic simulation

We use several software packages to simulate electromagnetic structures.

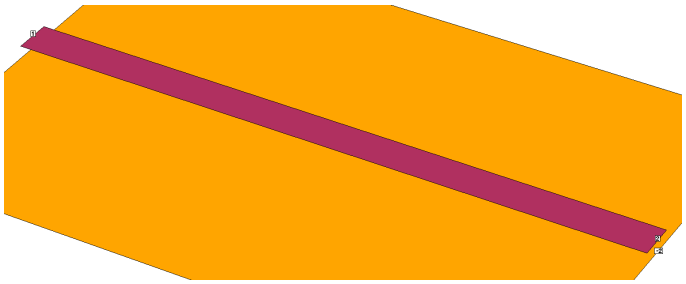
Keysight (formerly Agilent) has an application called the Advanced Design System (ADS), which is capable of both circuit simulations and 3D electromagnetic simulations. The 3D simulations are essentially solving Maxwell's equations numerically for a given geometry and configuration of excitations. In ADS, it is possible to simulate using a finite-element method (FEM) or a method of moments (MoM). The MoM is much faster. The geometry is restricted to be what is sometimes called 2.5D, which means that the structure is created in stacked two-dimensional layers. The layers can be patterned, and the material properties including thickness can be chosen by the user. We have found some bugs in the ADS solver, however, including a scale dependence and have, therefore, used ADS only for its circuit simulator. The scale dependence was noticed by simulating a straight microstrip line mm wavelengths and comparing to the same geometry scaled by a factor of 1000 in each linear dimension and simulated at m wavelengths. The S-parameters did not sum to 100% at the small wavelengths, but they *did* at the large wavelengths. We also noticed unphysically large coupling between microstrip lines and slotlines. This coupling was also *not* reproduced by other software packages. So we decided to stop using ADS for 3D simulations. The circuit simulator, however, has proven to be an extremely valuable tool. It essentially solves circuits analytically but much faster than a human being. The circuit parameters can be tuned in real time. We typically use ADS as a high-level design tool before proceeding to a physical layout.

We use Sonnet Software as a FEM 2.5D simulator. It is especially useful for circuits that are fabricated in thin-film processes, which is exactly how are devices are made (see Sec. 3.3). We use Sonnet to simulate most of the microwave components other than the antenna. This includes bandpass filters, hybrids, power splitters, impedance transformers, etc. By changing the material properties, we can also perform tolerancing analyses, e.g., characterizing the shift in a passband as a function of dielectric constant.

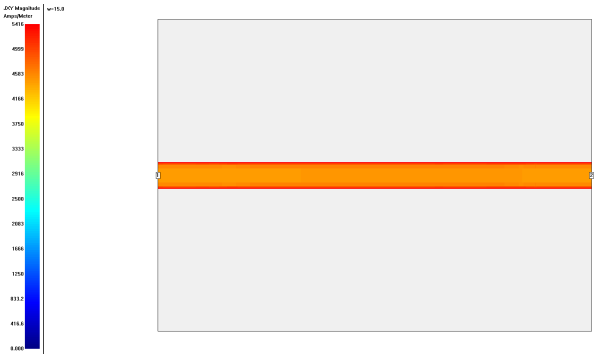
Ansys offers an application called the High Frequency Structure Simulator (HFSS), which is a 3D FEM simulator. Unlike ADS and Sonnet, there are no restrictions on the geometry. The simulator is essentially integrated with a computer-aided design (CAD) program. We use HFSS to simulate three-dimensional structures that cannot be created by stacking two-dimensional layers. An example, which we will discuss in detail in Ch. 4, is the lenslet-coupled planar antenna. The program solves for the electric and magnetic fields at all points within the simulation volume and is able to extrapolate the field structure to the far field. In this way, we can simulate the beam or radiation pattern of an antenna.

3.2.3 Microstrip line

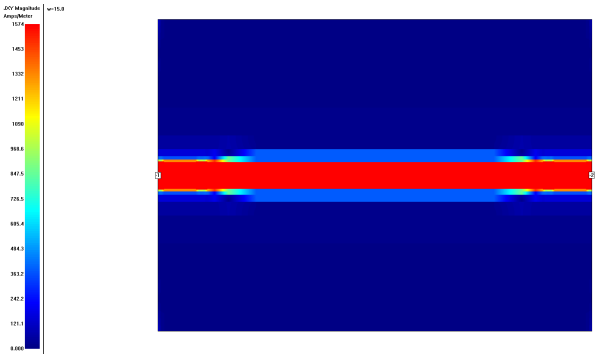
A transmission line that is particularly convenient for our fabrication process and for our electromagnetic requirements is the *microstrip line*, which is nothing more than a strip of metal above a ground plane as shown in Fig. 3.9(a). Qualitatively, the microstrip line can be understood as topologically equivalent to the more generally familiar coaxial cable. The inner conductor has been flattened into the strip, and the outer conductor has been unwrapped and extended infinitely outwards. A useful property of a coaxial cable is its ability to support transverse electric and magnetic (TEM) modes with no cutoff frequency. This is to be contrasted with a waveguide, which has a non-zero cutoff frequency even for its lowest-order modes. For this reason, the waveguide must have cross-sectional dimensions



(a) An isometric view of a microstrip line. Both orange and maroon are meant to indicate niobium. The maroon strip is 15-μm wide and 0.5 μm above the orange ground plane. The z-axis in this image has been exaggerated by a factor of ten. The simulation ports are indicated by the numerical labels at the ends of the strip. With a silicon-oxide dielectric, the characteristic impedance is close to 7 Ω.



(b) Current distribution on the strip layer at 260 GHz. The color scale extends from 0 to 5400 A/m.



(c) Current distribution on the ground-plane layer. Notice that the excitations are confined to the area very close to the strip. The color scale extends from 0 to 1600 A/m.

Figure 3.9: Model of a microstrip line for Sonnet simulations.

at least on the order of a wavelength. A coaxial cable, however, can be made much smaller while supporting even DC, i.e., extremely low-frequency, signals. Qualitatively, this behavior persists in the microstrip line. The lowest-order mode is not exactly but approximately TEM, which we sometimes call quasi-TEM, with no cutoff frequency. So we can build microstrip lines with cross-sectional dimensions much smaller than a wavelength. Typically, we are limited in compactness only by our fabrication capabilities. A comfortable geometry is a 10-μm-wide strip with a height of 0.5 μm from the ground plane. We have used such microstrip lines at frequencies as low as 40 GHz, where the free-space wavelength is 7.5 mm, i.e., hundreds of times larger than the microstrip width and thousands of times larger than the microstrip height. This geometry also tends to constrain the electromagnetic fields similar to a parallel-plate capacitor, so the wave propagation is not significantly influenced by nearby objects. For mechanical reasons, we place a dielectric between the ground plane and the strip layer; otherwise, the strip would be levitating. This dielectric tends to constrain the fields even more, which is again similar to the operation of a parallel-plate capacitor. The current distributions for a Sonnet simulation of a microstrip line are shown in Figs. 3.9(b) and 3.9(c). We see that the current is essentially negligible already just one strip-width away from the microstrip line.

Impedance transformer

Different parts of the microwave circuit require different microstrip impedances. The antenna tends to prefer relatively high impedances, e.g., 30-50 Ω. For reasons discussed in Sec. 4.5.4, the bandpass filters prefer lower impedances such as 5-10 Ω. Microstrip crossovers and crossunders, however, prefer higher impedances such as 20-30 Ω. The microstrip termination on the TES island prefers impedances on the lower side but not too low, e.g., 10 Ω. To accommodate all of these requirements, it is useful to be able to transform impedances at will, which is why we discuss in this section the design of transmission-line *impedance transformers*.

Since the reflection coefficient for a transmission line is controlled by a Fresnel-like equation (Eq. 3.14), we can design impedance transformer by direct analogy to anti-reflection coatings. This is a useful analogy in our field, because quite a lot of effort has already been brought to bear on simulations and optimizations of anti-reflection coatings. Suppose we simulate an anti-reflection coating for transmitting from a material of dielectric constant $\epsilon_{\text{min.}}$ to a material of dielectric constant $\epsilon_{\text{max.}}$. Let the dielectric constant of the ith anti-reflection layer be ϵ_i, where we mean that this is the ith layer from the layer of dielectric constant $\epsilon_{\text{min.}}$. And these layers have corresponding thicknesses d_i. Now we wish to translate this system to a transmission-line impedance transformation. We transmit from a characteristic impedance $Z_{\text{min.}}$ to $Z_{\text{max.}}$. The relationship between the minimum and maximum dielectric constants and impedances is

$$\frac{Z_{\text{max.}}}{Z_{\text{min.}}} = \sqrt{\frac{\epsilon_{\text{max.}}}{\epsilon_{\text{min.}}}}. \tag{3.16}$$

In general, a higher dielectric constant in a real transmission line creates a lower characteristic impedance, since it increases the shunt capacitance. The ith transmission-line segment

from $Z_{\text{max.}}$ must then have impedance

$$Z_i = Z_{\text{max.}} \sqrt{\frac{\epsilon_{\text{max.}}}{\epsilon_i}}. \tag{3.17}$$

So, to be clear, the i label is being used in different directions for impedance and dielectric constant. The electrical length of each transmission-line segment should be the same as each layer of anti-reflection coating. A transmission line of a given impedance Z_i has an effective dielectric constant $\epsilon_{\text{eff.},i}$, which relates a free-space wavelength to a transmission-line wavelength by $\lambda_{\text{eff},i} = \lambda/\sqrt{\epsilon_{\text{eff.},i}}$. This effective dielectric constant can often be extracted quickly in simulation. Demanding constant electrical length then translates to

$$\ell_i = \sqrt{\frac{\epsilon_i}{\epsilon_{\text{eff.},i}}} d_i, \tag{3.18}$$

where ℓ_i is the length of the ith segment from $Z_{\text{max.}}$. Using this prescription, one can repurpose an anti-reflection optimization code for transmission-line impedance transformations.

A fabricated microstrip impedance transformer is shown in Fig. 3.10(a). We see that the transition is from a narrow, high-impedance line to a relatively wide, low-impedance line. In between, there are three segments of intermediate width. The simulated transmission is shown in Fig. 3.10(b), which shows good transmission above 95% over a nearly 4 : 1 bandwidth. The slight ripples are characteristic of a three-segment impedance transformer or three-layer anti-reflection coating.

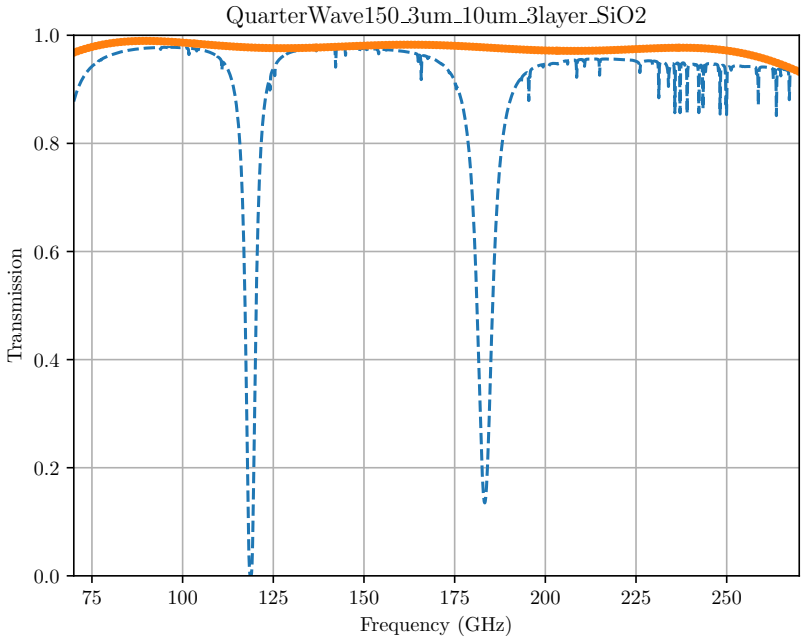
3.2.4 Termination

In most current-generation experiments, the incident radiation couples to an antenna which is fed by a microstrip line.[2] The signal can then pass through bandpass filters and/or mode rejectors on its way to the bolometer island, where the power is dissipated as heat for the bolometer to detect. A resistive element is used to dissipate the heat. This can be accomplished in more than one way. There are two common techniques in the field: a lossy meandering microstrip line and a lumped resistor.

We first discuss the lossy meander. Most experiments use superconducting metals and low-loss dielectrics for their microstrip lines in order to minimize attenuation of the microwave signal. At the bolometer island, however, it is necessary to dissipate the power. This can be achieved by transitioning to a non-superconducting, purposefully resistive metal. The signal will be attenuated along the length of this non-superconducting microstrip line, and the power will be dissipated as heat. The microstrip can be designed to meander so that the path length is large while occupying a relatively small area on the bolometer island. The resistivity of this non-superconducting metal must be relatively low in order to prevent an impedance mismatch between the incoming superconducting microstrip line and the lossy meandering microstrip line. Since the resistivity is low, the attenuation per unit length is relatively small; therefore, the meander must have a large path length in order to dissipate a majority of the power. This tends to make the lossy meanders large, which also increases

[2]Some of the text in this section appeared in Abitbol et al. [2].

(a) Fabricated microstrip impedance transformer.



(b) Simulated performance (solid); atmospheric transmission (dashed).

Figure 3.10: Impedance transformer for microstrip lines (3 μm $\rightarrow$ 10 μm).

the size of the bolometer island. A desirable property of this termination is that it requires only a signal unbalanced microstrip line coming in to the bolometer island. The end of the meander can be left open-circuited, since the reflected power is heavily attenuated by the lossy metal. Since different frequencies pass through a different number of wavelengths in the meander, the absorption efficiency is frequency dependent.

The other main type of termination is a lumped resistor. The incoming microstrip line is terminated by an impedance-matched resistor, and the power is then dissipated as heat on the bolometer island. An advantage of this paradigm is that a lumped resistor tends to be relatively small and represents a minor contribution to the size of the bolometer island. The lumped resistor typically consists of a short section of high-resistivity metal, where the particular geometry is important in determining the lumped resistance. For a single unbalanced microstrip line, the lumped resistor should be shorted to ground; the disadvantage here is that a via is required. For two balanced microstrip lines, the resistor can be differentially fed and, if its resistance is chosen to be twice the microstrip impedance, will dissipate all of the power without a via. Another advantage of the differentially fed termination is that it accepts odd modes but rejects even modes. The lumped-resistor paradigm is relatively insensitive to the termination resistance, because the reflected power goes as

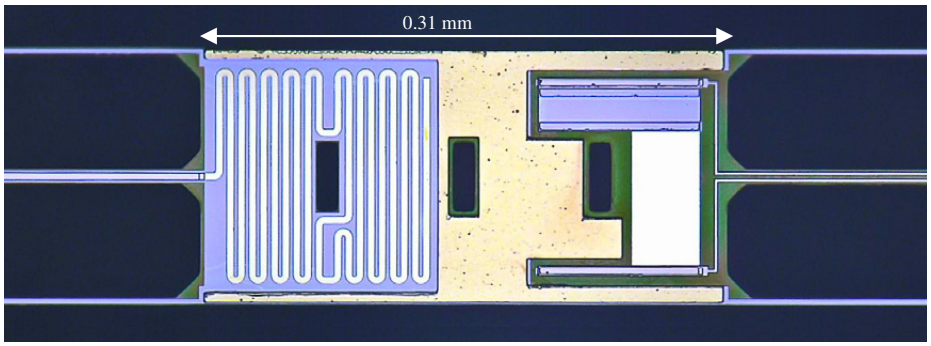
$$|\Gamma|^2 = \left| \frac{R_0 - R_L}{R_0 + R_L} \right|^2 , \tag{3.19}$$

where Γ is the reflection amplitude, R_0 is the characteristic impedance of the microstrip line and R_L is the termination (load) resistance. This relation is plotted in Fig. 3.8(b), where both impedances are assumed to be pure real as is the case in most experiments. We see that the reflection increases relatively slowly as R_L deviates from R_0. Even when the termination resistance differs from the microstrip impedance by a factor of 2, the reflection is only $\sim 10\%$.

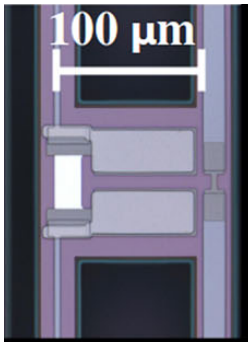
Some experiments that use lossy meanders include BICEP2, ABS, ACTPol and SPTpol [4, 37, 44, 92]. Gold is a popular low-resistivity metal for this purpose. The BICEP2 bolometer island is shown in Fig. 3.11(a), where a lossy gold meander can be seen.

Experiments that use lumped resistors include POLARBEAR, POLARBEAR-2, SPT-3G and CLASS [9, 22, 82, 99]. Titanium is a popular high-resistivity metal for this purpose. The critical temperature of titanium, which is $\sim$ 500 mK, is low enough for frequencies above $\sim$ 40 GHz to break Cooper pairs and see titanium as an effectively normal metal. A POLARBEAR2-style bolometer island is shown in Fig. 3.11(b), where the lumped resistor can be seen. Notice that the lumped resistor is substantially smaller than the lossy gold meander shown in Fig. 3.11(a).

In the realm of lumped terminations, there are two main varieties. The termination can either be differential, i.e., coupling to unbalanced or out-of-phase microstrip lines, in which case there is a virtual ground in the middle of the resistor, or it can be a resistance between the strip layer and the ground plane. The two varieties are shown in Fig. 3.12. These designs were used in all of the prototype devices reported on in Ch. 4 and Ch. 5. The two varieties are useful in different circumstances. If the antenna is differentially fed, then it is natural to simply bring the two lines to the bolometer island and terminate the power on a differentially-fed resistor. This design also removes the need for a via, which may be undesirable in some cases for fabrication reasons. When the microstrip lines from the

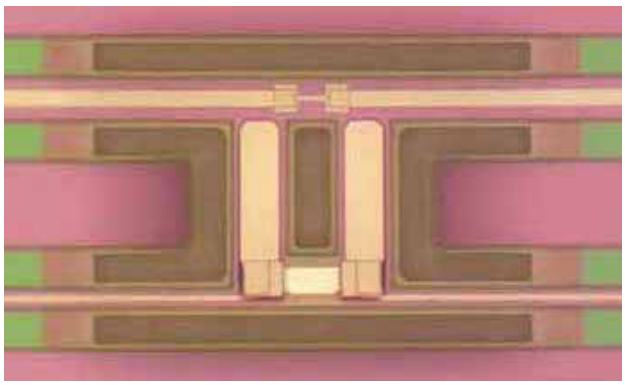


(a) BICEP2 bolometer island with a lossy gold meander (left) and TES (right) [4].

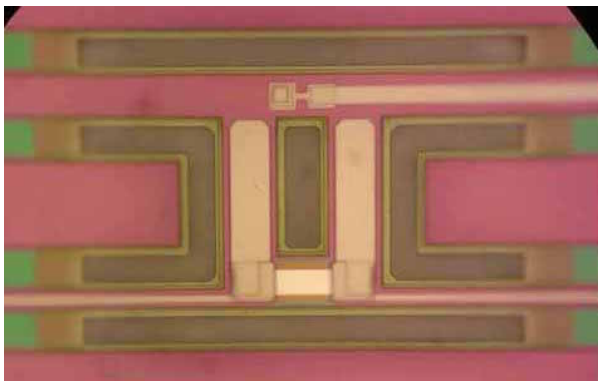


(b) POLARBEAR2-style bolometer island with a lumped titanium resistor (right) and TES (left) [103].

Figure 3.11: Example microstrip terminations. These images are approximately to scale relative to one another. Notice that the lossy gold meander requires a much larger area than the lumped resistor.



(a) Differential termination resistor.



(b) Via termination connected through the insulating SiO to the Nb ground plane.

Figure 3.12: Bolometer islands featuring AlMn TESs and Ti lumped resistors.

antenna have been combined to one, it is more natural to simply terminate the power in a single-ended resistor rather than converting the single microstrip line to two.

3.2.5 Slot antenna

A slot antenna is one that is constructed from a slot in an infinite ground plane. A slot antenna is the dual of a wire antenna of the same geometry, i.e., electric and magnetic fields are approximately interchanged but radiation patterns are similar. Slot antennas have a number of advantages. They are planar, which makes them relatively easy to fabricate compared with three-dimensional antennas. In particular, the design lends itself immediately to standard thin-film microfabrication techniques. The ground plane allows for microstrip technology, which we argued for in Sec. 3.2.3, as well as coplanar waveguides (CPWs) and slotlines. In this paradigm, the microstrip lines run from the bolometer island to excite the slot antenna, and we have the opportunity to place bandpass filters between the two. The ground plane also serves as part of a radiation shield or Faraday cage for all of the electronics that may be installed behind the detector wafer. This is especially important for the SQUID amplifiers, which respond to electromagnetic radiation up to approximately 5 GHz, e.g., cell phones, wi-fi, etc.

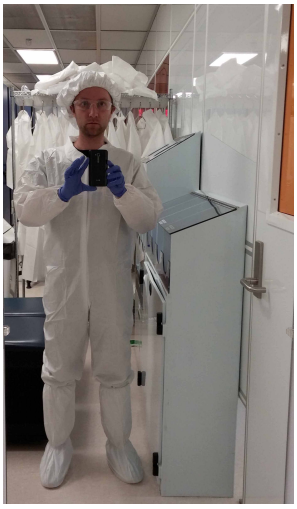
3.3 Fabrication

The fabrication process described here represents a snapshot of the POLARBEAR-2 (PB-2) detector development effort and was used for the prototypes described in Ch. 4 and Ch. 5 except for Sec. 4.8.1. The process used for the pixels reported on in Sec. 4.8.1, which is described in Westbrook et al. [103], was similar but differed in a number of important ways that made it essentially inferior to the more mature process described here. The PB-2 process has evolved beyond what is described here, most notably in the introduction of silicon nitride as a lower-loss microstrip dielectric in the place of silicon oxide, but we do not report on these more recent developments.

Our group fabricates devices in the Marvell Nanolab at UC Berkeley. This is a shared facility, which is available to university-affiliated students and researchers as well as to private businesses. The Nanolab is a class-100 clean room, which means that there are no more than 100 particles of size 0.5 μm per cubic foot. For emphasis, this air quality can be compared to that of a typical indoor room which has more than a million such particles per cubic foot. Some images from the Nanolab are shown in Fig. 3.13. Nanolab users wear full-body "bunny suits" to protect the devices and machines from particle contamination, i.e., mainly to protect the fabrication from the fabricator. For certain activities, additional outerwear is necessary to protect the user from harmful chemicals; this might include chemically resistant aprons and gloves as well as face shields.

Before describing the entire fabrication process, it is perhaps instructive to outline a simple process, which contains most of the key elements. In Fig. 3.14, we show some cartoons for a one-layer process using photolithographic techniques for patterning. We begin with a blank monocrystalline-silicon wafer, which can be simply purchased at the Nanolab or ordered from a company. For our devices, the dopants and crystal orientations don't matter

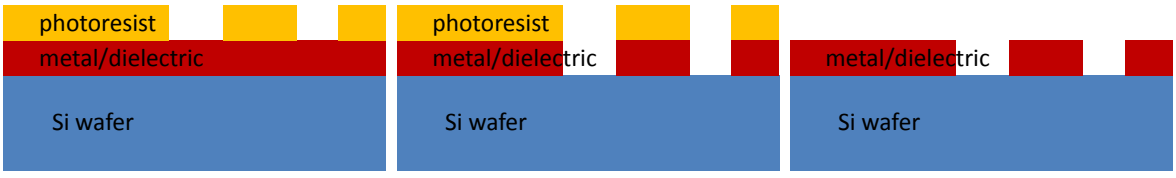
(a) Detail from a completed device showing the artist's signature.



(b) A portrait of the artist as a third-year graduate student.

Figure 3.13: Scenes from the fab.

(a) Crystalline Si wafer (b) Deposit metal/diel. (c) Coat with photoresist



(d) Expose and develop (e) Etch metal/diel. (f) Strip photoresist

Figure 3.14: One-layer example process.

very much. All we need from the silicon is a low-loss isotropic dielectric at low temperatures. The semiconductor properties freeze out well above our operating temperature, and crystalline silicon is negligibly birefringent. In the clean room, we deposit (by methods which will be described below) a thin layer (typically no thicker than 1 μm) of some metal or dielectric (depending on the application of interest). The remaining steps are all aimed at patterning this layer, which we accomplish through a process called *photolithography*. We deposit a layer of a light-sensitive material called *photoresist*. The thickness is typically 1-3 μm. We place a shadow mask above the wafer and shine a laser on it, which exposes some of the photoresist to light. The exposed photoresist can be washed away in a process called *developing*. At this point, we have patterned the photoresist but not the metal/dielectric. To *etch* the metal/dielectric, there are a variety of methods available, which we will discuss below. Ideally, the photoresist protects the underlying metal/dielectric perfectly, but usually the etch process affects the photoresist to some extent as well. It is, therefore, important to know the etch rate of the photoresist as well as that of the metal/dielectric. If the photoresist were to be etched away completely, then the metal/dielectric would be etched in undesirable locations. After the etch is completed, the remaining step is to remove or *strip* the photoresist. There are several methods available, which we describe below. Once the photoresist is removed, we have a patterned thin film of metal/dielectric. For a multilayer process, these steps can be repeated but with several added complications stemming from competing film stresses, overetches into underlying materials, alignment between layers, etc.

Another common process is called *lift-off*. A cartoon of an example lift-off process is shown in Fig. 3.15. In a lift-off process, the photoresist is applied and patterned before the metal/dielectric is deposited. The metal/dielectric is then deposited *on top of* the photoresist. Since the photoresist was patterned, though, some areas of the underlying substrate are exposed, so some of the metal/dielectric is resting on photoresist and some simply on the underlying substrate. Then we dissolve away the photoresist, which carries with it any metal/dielectric that was resting on top. What is left is the metal/dielectric that was resting on the substrate. Lift-off is a useful process for layers that are difficult to etch by other means.

In the Marvell Nanolab, each machine has a name which is typically given in all lowercase. The tool name is usually related to the company that made the tool. For example, we use a plasma etcher called lam7, which was made by Lam Research. We will refer to the Nanolab machines according to this convention.

3.3.1 Thin-film deposition

In this section we describe several methods for depositing thin films of metals and dielectrics. The typical thicknesses for our process are in the range 0.1-1 μm.

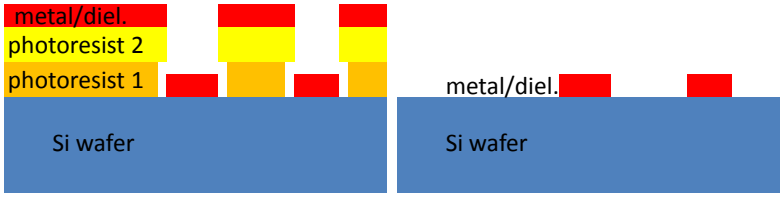
Evaporation

By raising a metal to its boiling point, we can create a metal vapor. If we place our wafer upside-down above the boiling metal, then the vapor will rise until it hits the wafer. A cartoon is shown in Fig. 3.16(a). The wafer is held at a much lower temperature than the

(a) Crystalline Si wafer

(b) Coat with two different photoresist layers.

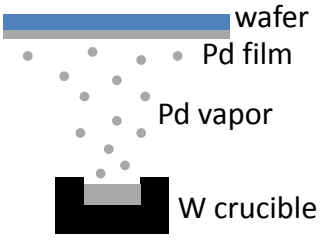
(c) Expose and develop. The bottom layer is more vulnerable to the developer and is, therefore, etched further horizontally creating an overhang.



(d) Metal/dielectric is deposited. The overhang ensures that the layer is broken.

(e) The photoresist is dissolved away and carries some of the metal/dielectric with it.

Figure 3.15: One-layer lift-off example process.



(a) Boiling metal evaporates and condenses on the upside-down wafer.

(b) A tungsten crucible containing palladium at room temperature.

(c) The same crucible heated with an e-beam to $\sim 1500°$ C.

Figure 3.16: Electron-beam evaporation. The crucible diameter is approximately 1 in.

(a) Ions bombard a target, and material rains down and adheres to the wafer.

(b) The mrc944 in the Marvell Nanolab, a multi-target sputter-deposition machine. The tool is about 6 ft. tall.

(c) Deposition chamber of the mrc944 while 1 kW of power is used to strike an argon plasma to sputter titanium. The chamber is about 2 ft. wide.

Figure 3.17: Sputter deposition.

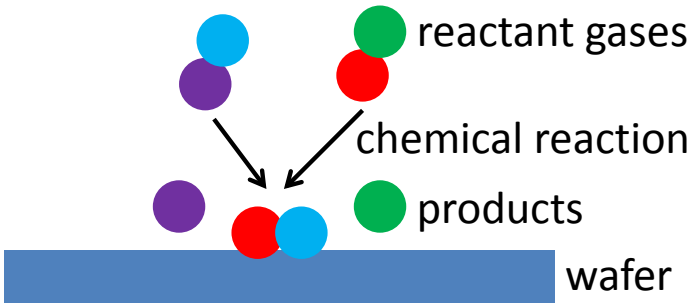
boiling metal, so the vapor particles lose energy and adhere to the wafer. Over time, the condensed metal forms a film on the wafer, which can be controlled at the level of ~ 10 Å.

The metal of interest is held in a *crucible* composed of a material with a significantly higher boiling point. For palladium, we use a tungsten crucible as shown in Fig. 3.16(b). For titanium, we use a graphite crucible. In vacuum, we heat the metal with an electron beam (e-beam) that is focused onto the center of the crucible. This provides intense local heating as can be seen in Fig. 3.16(c). The beam can be steered to induce boiling in other areas of the crucible or simply to recenter.

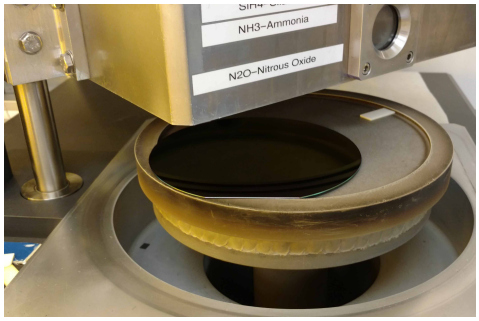
The deposition rate is measured in real time with a *crystal monitor*. This is a mechanical resonator, whose resonant frequency changes as material accumulates. It is placed in the vacuum chamber but away from the wafer. If the acoustic impedance of the metal is given, then the change in resonant frequency can be translated to a change in layer thickness. The output of the crystal monitor can be fed back to the e-beam gun to stabilize the deposition rate. A typical deposition rate for, e.g., palladium is 10 Å/s.

Sputter deposition

Another method of thin-film deposition involves bombarding a metal or dielectric target with ions and then allowing the ejecta to rain down onto the wafer. A cartoon of this process is shown in Fig. 3.17(a). Typically, the vacuum chamber is filled with a low-pressure inert gas like argon. A strong electric field is applied, which ionizes the gas. The ions are accelerated into the target and mechanically etch away material that falls down onto the wafer. When the target is larger in diameter than the wafer and the electric field is applied over a comparably large area, the deposition rate on the wafer can become reasonably uniform as a function of radius. In reality, there is a noticeable radial dependence. Modulo this non-uniformity, the film thickness can be controlled at the level of approximately 50 Å in the sputtering tools we use in the Nanolab. We control the thickness by measuring the deposition rate on dummy wafers and simply timing the sputter deposition on device wafers. A tool called the mrc944, which we use to deposit Ti and AlMn without breaking vacuum, is shown in Figs. 3.17(b) and 3.17(c).



(a) Reacting gases are mixed with extra power provided by an ionizing electric potential. A chemical reaction creates products which adhere to the wafer or are pumped away.



(b) A wafer being removed from the oxford2, a PECVD tool in the Nanolab. The surface on which the wafer is resting is held at 350° C, which helps to expel impurities from the film.

Figure 3.18: Plasma-enhanced chemical vapor deposition (PECVD).

Chemical vapor deposition

In chemical vapor deposition, two (or more) reactant gases are mixed to create products, at least one of which adheres to the wafer. The chemical reaction can be encouraged by applying an ionizing potential inside the vacuum chamber; this is called *plasma-enhanced chemical vapor deposition* (PECVD). A cartoon of chemical vapor deposition is shown in Fig. 3.18(a). To deposit silicon oxide, we flow silane (SiH_4) and nitrous oxide (N_2O), i.e., laughing gas, into the chamber. The silane is actually diluted with argon (10% silane and 90% argon). We strike a plasma, which encourages a chemical reaction producing SiO_2 as well as some molecular hydrogen and nitrogen. The reaction can also produce some free silicon, which can enrich the deposited silicon dioxide. For this reason, we sometimes refer to our silicon oxide as SiO_x, where stoichiometrically $x < 2$. Enriching the film with silicon tends to increase its dielectric constant.

In the Marvell Nanolab, we use a PECVD tool called oxford2. The deposition rate can be monitored with a film-thickness measurement system called nanospec, which calculates the local thickness based on interference patterns produced with visible light. A typical deposition rate is something like 13 Å/s. For silicon oxide, we can reliably control the thickness and uniformity at the percent level.

3.3.2 Photolithography

As described above, we use a technique called *photolithography* to pattern our wafers. There are multiple steps involved, but the basic idea is to use a light-sensitive material called *photoresist* to create a kind of shield over the parts of the thin film that we want to retain and then to etch away the parts of the thin film that are unshielded. After the etch, we dissolve or burn away the photoresist. An overview of a photolithographic process is shown in Fig. 3.14. These steps are repeated (with small variations) many times throughout the fabrication process, since every layer requires some patterning.

The very first step is to *prime* the wafers by depositing a thin layer of adhesion promoter called hexamethyldisilazane (HMDS). We use a tool called primeoven for this purpose, which dehydrates the wafers before applying the HMDS. Without this adhesion promoter, the photoresist is likely to delaminate during the develop, especially near small features.

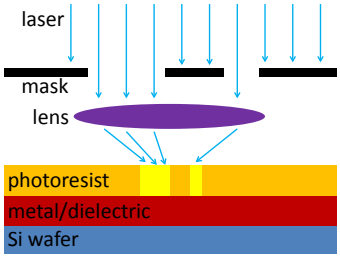
From this point forward, the wafer must be shielded from ultraviolet light except during the exposure. This is because the photoresist is light sensitive. For this reason, the areas of a nanofabrication facility designated for photolithography tend to use yellow filters on all lights including those on microscopes. This is why the photographs in, e.g., Fig. 3.19 all appear to be tinted yellow.

Next we coat the wafer with photoresist. Until recently, our group used a machine called svgcoat1 for this purpose, but this machine has since been decommissioned. The svgcoat1 looked similar to the machines shown in Fig. 3.19(a). This tool is a spin coater, which means that it spins the wafer on a vacuum chuck while dispensing photoresist. The photoresist spreads out across the wafer to create a uniform coating. The thickness can be controlled and is typically 1-3 μm. We choose the thickness based on the overall process including the required resolution on the device features and the etch process for the underlying metal/dielectric layer. We use two types of photoresist, which we refer to as i-line and g-line photoresist. These photoresists are optimized for slightly different wavelengths: i-line is optimized for 365 nm and g-line for 436 nm. But we have found that they are practically interchangeable as far as the exposure is concerned. For most layers, we apply 1.3 μm of g-line resist.

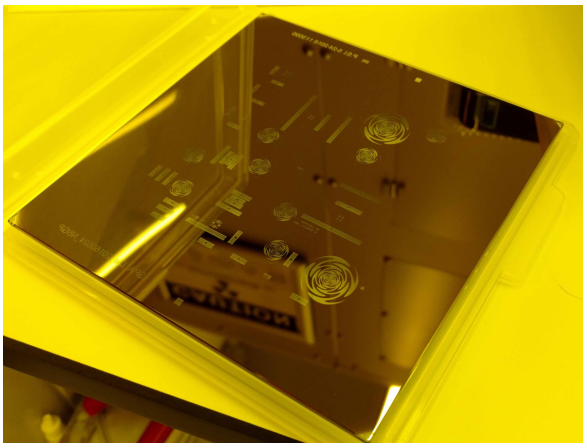
The photoresist is now exposed to ultraviolet (UV) light. We place a *mask* between the wafer and the UV light source to cast shadows corresponding to our design features. A typical mask is shown in Fig. 3.19(c). It is usually made of some sort of glass with a thin film of chrome that is responsible for the shadows. We design our masks in a program called L-Edit made by Tanner Tools. An example of a screenshot from L-Edit is shown in Fig. 3.20. The geometries are exported as gds files and sent to a mask maker for fabrication. For critical masks, we purchase from Photo Sciences, Inc. For less critical masks, we can purchase from the Marvell Nanolab, where the staff are able to fabricate masks subject to some restrictions on feature shapes and sizes; non-rectangular geometries tend to be difficult at the Nanolab. Since our tolerances are no smaller than 0.5 μm, we use near-UV light for the exposure. For finer features, one can use deep-UV or, for *very* fine features, an electron beam, both of which are available in the Marvell Nanolab. We use two different machines for exposures. The one that is in some sense simpler is called ksaligner (Karl S$'$uss aligner), which is a tool for *contact lithography*. This means that the mask features are made at the same scale as the intended on-wafer features. The mask is literally pressed on top of the wafer[3] so that the projected shadows are the same size as the mask features. The wafer is exposed to a UV light source for a few seconds, where the time depends on the thickness of the photoresist. The other exposure tool we use is called gcaws6 (GCA wafer stepper). It is a tool for *stepper lithography*. This means that the projection of the mask does not cover the entire wafer, usually just a small area like a square inch or less. But the tool has the

[3]It is possible to do so-called *proximity lithography*, in which case the mask comes close but does not touch. The fidelity of the pattern transfer is compromised. For non-critical features, though, this is a viable option.

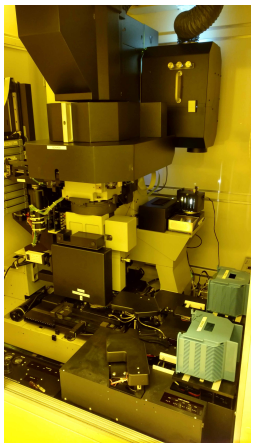
(a) A photoresist spin coater (svgcoat6) on the left right up against a photoresist developer (svgdev6) on the right. These machines are track systems that shuttle the wafer from one end to the other.



(b) A mask controls where ultraviolet light can strike the photoresist. A lens can be used to focus the light to a smaller area than on the mask itself.



(c) A mask for stepper lithography that reduces feature sizes by a factor of five. The mask is 5″ on a side. The substrate is soda-lime glass; the patterned thin film is chrome.



(d) The gcaws6 stepper lithography machine. The tool is about 7′ tall.

Figure 3.19: Photolithography. The area of the Nanolab devoted to photolithography uses yellow filters on all lights (except in the exposure machines) to protect photoresist from ambient ultraviolet light. This is why all of the photographs appear to be tinted yellow.

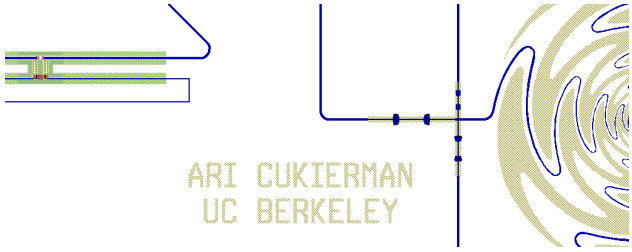
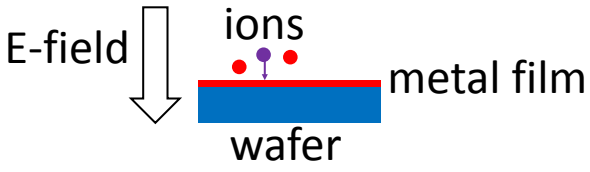


Figure 3.20: Mask design in L-Edit showing the same region of the wafer as Fig. 3.13(a). Each color represents a different mask. The program L-Edit allows the user to see all of the masks stacked on top of each other and to edit them. Some layers are drawn as negatives, e.g., the niobium ground plane (beige), since the features tend to be *slots*; the actual mask will be printed as the negative of what is drawn.

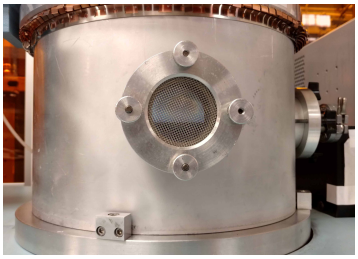
capability to flash this pattern in many places across the wafer. This is useful when features are repeated many times across a wafer. In the gcaws6, a lens focuses the shadows to be five times smaller in each linear dimension on the wafer as compared with the mask. This makes the gcaws6 useful for fine features, since the tolerance on the mask is five times larger than on the wafer. A cartoon of the optics of the gcaws6 is shown in Fig. 3.19(b), and a picture of the gcaws6 is shown in Fig. 3.19(d). The stepping pattern can be programmed and does not have to be regular. Additionally, the gcaws6 has the capability to block parts of the mask with so-called shutters. There are four shutters corresponding to the $\pm x$- and $\pm y$-directions. Sometimes we put multiple independent features on the same mask that we do not intend to appear near each other on the wafer. With the shutters, though, we can control what is transferred to the photoresist at each exposure. This strategy of sharing the same mask helps to save money and time. After the first layer has been patterned, subsequent layers must be aligned to the underlying ones. We place special alignment marks on our masks to aid in this alignment. Both the gcaws6 and the ksaligner allow for manual alignment. The gcaws6 also has an automatic alignment, which uses a program called Micro DFAS (Micro Dark Field Alignment System), that works well once a course alignment has been achieved.

After the photoresist has been exposed, we must wash away the exposed areas. This is called *developing* the photoresist. We used a machine called svgdev1 for this purpose, but it has since been decommissioned. A similar machine is shown in Fig. 3.19(a). It is similar to the coating machine in that the wafer is shuttled on a track and spins on a vacuum chuck. But the chemical that is dispensed is a developer. After the develop, the exposed photoresist has been removed, but the unexposed photoresist remains.

When we apply photoresist for a lift-off process, we actually use *two* layers of photoresist as shown in Fig. 3.15. We perform a longer exposure to make sure we get through both layers. The photoresist used for the bottom layer, though, is chosen to be more vulnerable to the developer. This tends to create overhangs in the photoresist. If there were no such overhangs, it is possible that the metal that we deposit after lithography would conform to the photoresist profile and create a continuous sheet. Then the metal would not be easily lifted off in the dissolving process. The overhang helps to ensure that there is a break between the metal on the substrate and the metal on the photoresist, which produces a clean lift-off.



(a) A strong electric field to ionize a gas and accelerate ions into the substrate, where they remove material chemically and/or mechanically.



(b) A photograph of the ptherm, a versatile "barrel" etcher.

Figure 3.21: Plasma etching.

Sometimes at this point we perform what is called a *descum*. This is a light etch with an oxygen plasma, which we typically perform in a tool called technics-c. The idea is to burn away any residual photoresist in places it shouldn't be. We do this at the cost of burning away some photoresist in the places it should be. This step is not strictly necessary but is easy and fast.

For most layers that will see a plasma etch, we harden the photoresist right before the etch by baking it in an oven at, e.g., 120° C for approximately 30 min. This is useful for particularly aggressive etches that attack the photoresist to a non-negligible extent.

3.3.3 Etching

There are multiple methods available for etching materials. Here we describe the three that are used to fabricate our devices.

Plasma etch

In a plasma etch, one or more gases are ionized in a strong electric field and forced to collide with the wafer. An illustrative cartoon is shown in Fig. 3.21(a). The etch can be accomplished through a combination of mechanical and chemical effects. When the etch is purely mechanical as with an inert gas such as argon, we sometimes call it a *sputter etch*. We often use a sputter etch to remove relatively thin oxides that formed on existing layers before depositing a new layer. This helps with adhesion. When we are etching a layer to pattern it, we usually use an etch that is partially mechanical and partially chemical. We also sometimes use a plasma etch to strip photoresist. The photoresist reacts with oxygen, so an oxygen plasma can burn away the photoresist in a process we sometimes call *ashing*. A photograph of the ptherm, a tool we use for ashing, is shown in Fig. 3.21(b).

Wet etch

A wet etch is a type of chemical etch, for which the wafer is submerged in a liquid that reacts with at least one of the on-wafer materials. It is common to use a liquid etchant for aluminum, since many are commercially available. The etch rate is partially determined by

the temperature of the liquid, which we monitor for the duration of the etch. Typically, we keep the liquid slightly elevated above room temperature, e.g., 50° C. The etch is isotropic in the sense that the liquid etches everything it comes into contact with, so that vertical and horizontal directions are equivalent. Compare this with a plasma etch in which the ions are being accelerated perpendicular to the wafer and not parallel.

Gas etch

A gas etch is similar to a wet etch but uses a gas instead of a liquid. In this case, the etch rate is controlled by the flow rate of the etchant into and out of the process chamber. We use a xenon-difluoride (XeF_2) gas to etch silicon. The etch is isotropic just like the wet etch.

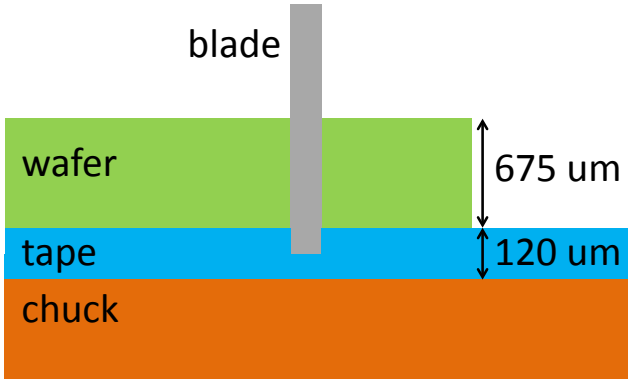
3.3.4 Dicing

The device wafer begins as a circle, but it is possible to cut it into a different shape through a process called *dicing*. Some images related to dicing are shown in Fig. 3.22. This is a non-trivial step, since the wafer is typically made of very thin (0.5-1 mm) monocrystalline silicon. In fact, it is possible to cleave the wafer along the crystal axes. To dice silicon, we use a diamond blade that is approximately 50-μm wide. The wafer is placed an adhesive tape, so that the disconnected pieces do not move around during the dicing process. The blade height is controlled at the level of approximately 10 μm to ensure that the blade cuts all the way through the silicon but not into the vacuum chuck below the tape.

3.3.5 Process

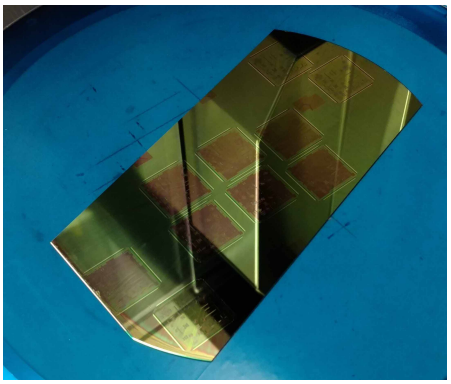
In this section, we describe the process flow to fabricate the prototype devices described in Ch. 4 and Ch. 5 but not Sec. 4.8.1. The process calls for eight masks, half of which were fabricated at Photo Sciences, Inc. and half at the Marvell Nanolab. This fabrication process required only stepper lithography, i.e., all of the features for each layer fit on a single mask. We repeat these features many times across the wafer to produce many instances of the prototype pixels. In order to mate the pixels with an existing POLARBEAR-1 (PB-1) lenslet array, the locations of the dies on the wafer were chosen to allow the pixels to line up with the lenslets. Additionally, alignment marks were placed on the wafer to correspond to those on the lenslet array. At the end of the process, the wafer was diced so that a central hexagon was extracted that mated with the PB-1 lenslet array. Additional dies were placed outside of the hexagon and diced or cleaved off separately for testing with single lenslets but *not* with a lenslet array.

We purchase prime-grade 675-μm-thick 6″-diameter P-type silicon wafers from Silicon Valley Microelectronics. We order them with dielectric coatings: 350 Å of silicon dioxide under 6000 Å of silicon nitride. This is shown topologically in Fig. 3.23(a). Our group used to coat wafers with these layers in the Nanolab using low-pressure chemical vapor deposition (LPCVD), but we experienced too much particle contamination from the relevant machines. We often monitor how much the wafer bends over the course of the process, since this is an indication of the film stresses. We use a tool called flexus to measure the wafer bow. At this

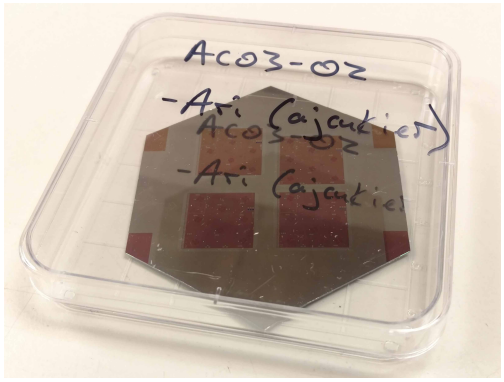


(a) Dicing cartoon.

(b) A diamond blade.



(c) A partial diced wafer.



(d) A wafer that has been diced into a hexagon and is ready for the next step in the process or to leave the clean room.

Figure 3.22: Wafer dicing. This is not necessarily the last step in the process, although it often falls near the end. For our devices, this is the penultimate step. We perform this dicing before releasing our bolometers, since the dicing process would destroy a completed bolometer.

point in the process, it is typical to have a maximum bow of up to 10 μm, i.e., the maximum deviation from flatness across the entire 6″ wafer.

Niobium ground plane

We deposit the niobium ground-plane layer in a sputter-deposition tool called gartek. To clean off any contamination on the SiN surface, we perform a light sputter etch in the load-lock chamber for 4 min. with 6 mTorr of Ar plasma at a power of 100 W. Before opening the gate between the load lock and the deposition chamber, we perform a *presputter* of the Nb target, which means that we bombard the target with Ar ions to clean off any contaminants. For this we apply a power of 3 kW for 3 min. with 5.5 mTorr of Ar plasma. Typically, we ramp up the power over the course of about a minute and then let it sit at the maximum for the remainder of the time. Now we open the gate between the load lock and the deposition chamber. Keep in mind that vacuum has not been broken all this time. We shuttle the wafer into the deposition chamber. A typical deposition uses 2.1 kW of power for 2.75 min. with 5.5 mTorr of Ar plasma. We ramp up the power over the course of approximately 6 s at the outset. This will produce a Nb film with a thickness of about 3500 Å in the center and about 3000 Å around the edges.

Using flexus, we can measure the film stress by comparing the wafer bow to what it was before the deposition. We usually aim for a 200-500 MPa of compressive stress (the wafer arches upwards, where up is the side with niobium). We found that it is important for a niobium film to have compressive stress as opposed to tensile stress in order to ensure that the metal is low loss at microwave frequencies.

We perform a standard photolithographic process to prepare the wafer for a niobium etch. We use two masks: one for the dies and one for the alignment marks intended for mating with a PB-1 lenslet array.

We etch the niobium ground plane in a plasma etcher called lam7. We use a chlorine based etch with 400 W of power for about 30 s. The tool monitors the abundance of certain elements in real time, which allows the user to see when the niobium has been fully etched, at which point we manually terminate the etch. Immediately after the wafer exits the vacuum chamber, we douse it with water to dilute any residual chlorine.

At this point in the process, there are multiple methods available to strip photoresist. We can dissolve the photoresist either by soaking it for several hours in acetone or much more quickly (5 min. or less) on a spinner with an NMP (*N*-Methyl-2-pyrrolidone) spray using a tool called asap-liftoff. We can also burn the photoresist away in an oxygen plasma in less than 10 min. with a tool called matrix. The state of the wafer at the point is shown topologically in Fig. 3.23(b).

Silicon-oxide microstrip dielectric

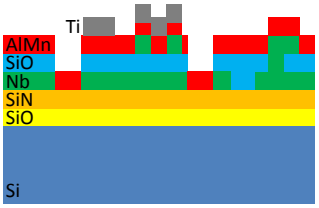
We deposit silicon dioxide in a PECVD tool called oxford2. The temperature of the chuck is held at 350° C. We flow silane (diluted with argon) and nitrous oxide and supply a power of 30 W. The deposition rate can be measured with a tool called nanospec, which was described above. Usually we use a dummy wafer to calibrate the deposition rate before

(a) The wafer comes to the clean room with uniform layers of SiO and SiN.

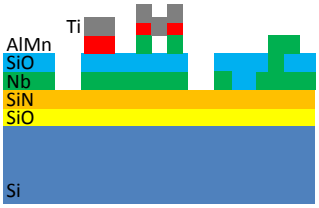
(b) Deposit and pattern the Nb ground plane.

(c) Deposit and pattern the SiO microstrip dielectric.

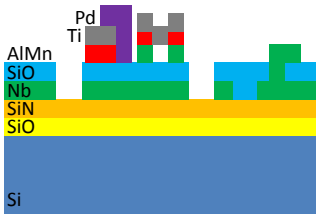
(d) Deposit and pattern the Nb strip layer.



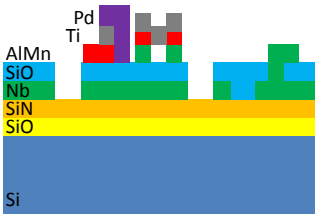
(e) Deposit AlMn immediately followed by Ti and pattern the Ti.



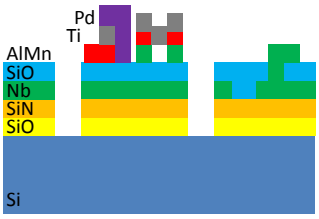
(f) Wet etch the AlMn, which also releases the Ti load resistors (*center*).



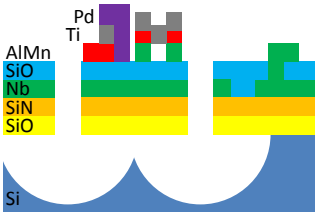
(g) Deposit the Pd bling.



(h) Remove Ti to reveal the AlMn TES (*left of center*).



(i) Etch through the SiN and underlying SiO to expose the silicon wafer.



(j) Etch the exposed silicon isotropically to release the bolometer island (*center*).

Figure 3.23: Cartoon cross sections of the fabrication process. The layer thicknesses are not to scale. The silicon wafer is approximately 1000 times thicker than any other layer. The profiles of the features are drawn much more sharply than they are actually fabricated to be, i.e., the side walls have a gentler slope on an actual wafer. These cartoons are meant to show topologically how the layers combine. Some information is, of course, lost in cross section.

inserting our device wafers into the machine. The rate usually comes to approximately 13 Å/s. For a thickness of 5000 Å, this translates to a deposition time of approximately 6 min.

We perform a standard photolithographic process. The features are mainly holes in the oxide to allow for vias and access to the underlying silicon nitride for the bolometer release that occurs at the end of the process.

The oxide etch is similar in execution to the niobium etch. We use a tool called lam6, which is nearly identical to the lam7 except for the available gases. Using a dummy wafer (or even a device wafer sometimes) and the nanospec, we can calibrate the etch rate. The etch lasts approximately 2 min. If there are areas over which the silicon oxide should be etched away to reveal the niobium ground plane, it is possible to perform an electrical inspection at this point to check whether the oxide has been completely etched.

Just as for the previous layer, we have multiple options available for stripping photoresist. Later on in the process, this will not be true. The state of the wafer at this point is shown topologically in Fig. 3.23(c).

Niobium strip layer

The process for depositing and patterning the niobium strip layer is nearly identical to that for the niobium ground plane. The only difference is that we use a different mask. We also commonly monitor the film stress to make sure it is still compressive.

At this point, it is useful to perform a few electrical checks. We want to make sure that the strip layer is not accidentally shorting to the ground plane in places we don't want it to. We want to make sure that our vias are actually making electrical contact between the strip layer and the ground-plane layer. Niobium is a high-resistivity metal at room temperature, so the measured resistance can actually be quite large depending on the geometry. The state of the wafer at this point is shown topologically in Fig. 3.23(d).

TES (AlMn) and load-resistor (Ti) layers

We deposit the TES layer and load-resistor layer without breaking vacuum in a sputter-deposition tool called mrc944. This machine has four metal targets and a conveyor belt that shuttles the wafer to the target of interest. We perform a presputter on the targets to clean them before the deposition. We also perform a sputter etch on the wafer in the vacuum chamber to clean off any oxides or other superficial contaminants on the wafer surface. We have tuned the deposition times to achieve the desired sheet resistances. For our TES material, we use aluminum manganese (AlMn). This is just pure aluminum that has been doped at a level ~ 4000 ppm with manganese in order to lower its superconducting critical temperature from about 1.2 K to about 450 mK. Concerning the AlMn sheet resistance, it is only important that the TES resistance comes to close to 1 Ω at cryogenic temperatures above the superconducting transition. Even this value is somewhat forgiving for small-scale prototype testing in which the TESs are voltage biased with a DC power supply. For the Ti, we aim for a cryogenic sheet resistance of 4 Ω/□. The residual resistance ratio (RRR), i.e., the ratio of the resistance at room temperature to that at cryogenic temperatures, is usually around 3. We can check this value relatively quickly by making one- or two-layer dummy wafers that we dunk in liquid helium. Empirically, we find that the resistance at 4 K is

essentially the same as the resistance at our operating temperature of 250 mK. One subtlety is that titanium is a superconductor with a critical temperature near 500 mK, but photons above approximately 40 GHz have enough energy to break Cooper pairs, which makes the titanium act like a normal metal at these frequencies.

Since we deposited two layers without any patterning in between, we now have a somewhat complicated lithography and etch procedure. We begin by removing some but not all of the titanium that we will eventually remove. We sometimes refer to this as the AMTI (Aluminum Manganese TItanium) mask. We use a slightly thicker photoresist (2 μm of g-line), since it must survive two etches instead of only one. We etch the titanium in the lam7 for about 75 s. The state of the wafer at this point is shown topologically in Fig. 3.23(e). After this etch and *without* stripping the photoresist, we bring the wafer to a wet bench for a wet aluminum etch. The aluminum manganese behaves just like regular aluminum for this purpose. We use an aluminum etchant made by KMG Electronic Chemicals, Inc. We hold the liquid at 50° C and leave the wafer submerged for approximately 4 min. After the etch, it is important to rinse thoroughly with deionized water. At this point, it is a good idea to perform an electrical check to make sure the aluminum has not been over- or under-etched. The titanium load resistors are particularly fragile, since the aluminum has been etched out from underneath them. They are, therefore, released from the substrate along most of their length. We have test structures on the wafer for electrical inspections of these load resistors.

Since the titanium plasma etch is particularly aggressive, the photoresist is tougher than usual at this point. We can take off the upper toughened layer of photoresist with a quick ash (oxygen plasma etch) in the technics-c. Since the titanium load resistors are now released from the substrate, we cannot use the asap-liftoff, since the NMP spray will destroy the resistors. We also do not want to use the matrix for an oxygen plasma, because it tends to heat the wafer, which can change the critical temperature of the aluminum manganese. So we soak the wafer in acetone overnight. This is longer than necessary but convenient; it is probably only necessary to soak the wafer for an hour or two. After the soak, we rinse with acetone, rinse with isopropanol (IPA), rinse with deionized water and finally blow dry. For good measure, we run a quick ash in the technics-c, e.g., 75 W for 5 min. with 180 mTorr of oxygen gas.

The titanium is not yet completely patterned, but we defer the removal of the remaining material until after the completion of the palladium layer. The state of the wafer at this point is shown topologically in Fig. 3.23(f).

Palladium bling

The bling layer is produced in a lift-off process. We, therefore, begin with the lithography. We coat the wafer with 2 μm of g-line resist followed by 1.3 μm of i-line. These two photoresist layers correspond to those shown in Fig. 3.15. The reason we use two layers is that we want to create an overhang. The developer tends to attack g-line faster than i-line, so this double stack will naturally produce an overhang during the develop. We do not perform a hard bake, since we want the photoresist to be easily dissolved after the metal deposition.

We deposit palladium in an *e*-beam evaporation tool called cha. To aid with adhesion, we deposit a 100-Å layer of titanium before we deposit palladium, but both materials are deposited without breaking vacuum. We use a deposition rate of 1 Å/s for titanium and

10 Å/s for palladium. Typically, we deposit approximately 1 μm of palladium. The palladium mass controls the heat capacity of the bolometer island, which controls the bolometer time constant. For some prototyping purposes, though, e.g., testing microwave devices, there is a large tolerance on the time constant.

The palladium is lifted off by soaking the wafer in acetone overnight. Sometimes we set the wafer upside down for the soak, so the metal falls away from the wafer as the photoresist is dissolved. Other than that, this lift-off step is essentially identical to the photoresist strip described in the previous section. The state of the wafer at this point is shown topologically in Fig. 3.23(g).

Titanium removal

Only after depositing palladium do we return to the titanium layer to remove some excess material. The reason for this order of operations is that the TES is protected underneath the titanium, which is a situation we would like to maintain as long as possible, since aluminum is actually soluble in the photoresist developer. Usually, the TES can survive one develop but not more, which would leave no room for error in the bling lithography step if the TES were unprotected. The TES is not located near a developed part of the photoresist in the remaining step, so it is OK to remove the titanium at this point in the process.

We run a standard lithography process but use 2-μm g-line resist to withstand the aggressive titanium plasma etch. The etch is identical to that described above for titanium. Mainly, this step removes the titanium above the TESs. Since the titanium resistors are still released, we must run an overnight soak to remove the photoresist as described above. The state of the wafer at this point is shown topologically in Fig. 3.23(h).

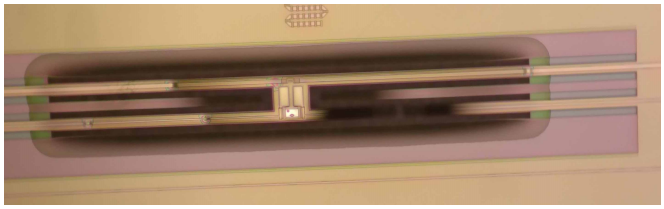
Dice and release

We now perform the last lithography step in the entire process. Since the following plasma etch is somewhat harsh and there are no more opportunities to reapply photoresist, we coat the wafers with 4 μm of g-line resist. The features we are now patterning are for holes in the underlying silicon nitride to give us access to the silicon wafer, which we will etch in order to release the bolometer from the substrate.

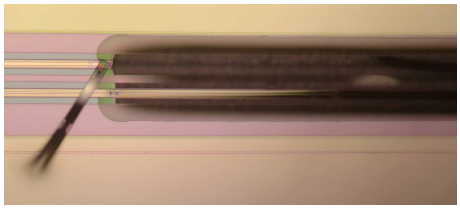
We etch the silicon nitride in a plasma etcher called ptherm. We use carbon tetrafluoride (CF_4) at a flow rate of 50 sccm and strike a plasma with 100 W of power. We sometimes use a dummy wafer to estimate the etch rate by measuring the silicon-nitride thickness with the nanospec. A typical total etch time is 11 min. Ignoring the photoresist, the state of the wafer at this point is shown topologically in Fig. 3.23(i).

We leave the photoresist on the wafer during the dicing procedure, since we want it to protect the wafer from silicon dust churned up during dicing and we need it for the following step.

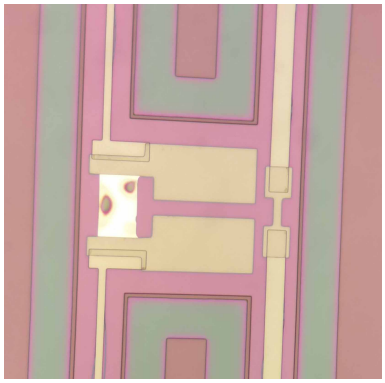
We release the bolometers with a xenon-difluoride XeF_2 gas etch in a machine called xetch. The machines operates by letting into the chamber a certain amount of gas, leaving it there for a chosen amount of time and then pumping it out along with any etch products. The machine can repeat this cycle any number of times. Typically, we let in 3.2 T of XeF_2 for 32 s at a time. We can monitor the etch rate of the silicon with test structures. The number



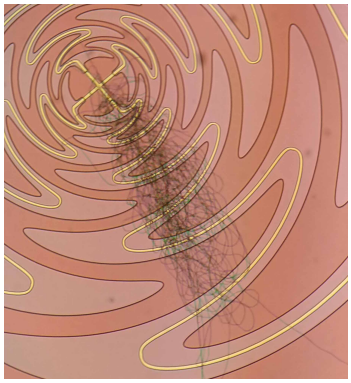
(a) Film stresses cracked the bolometer legs in a number of places. The bottom right leg has broken, and the two pieces have fallen downwards.



(b) Another example of film stresses breaking bolometer legs. In this case, the leg snapped upwards.



(c) After the aluminum etch, the photoresist was not stripped until the following day. Some residual etchant charred the TESs. Normally, the acetone used in the photoresist strip dilutes any remaining etchant.



(d) The Ti etch was too directional and left some thin strips of AlMn and Ti on the corners of the Nb microstrip lines that only revealed themselves during the aluminum wet etch.

Figure 3.24: Fabrication bloopers.

of cycles required to fully release all bolometers depends on the distribution of bolometers and the quantity. Usually, we need more than 15 cycles.

Once the bolometers are released, we remove the photoresist by ashing the wafer in the ptherm. Some typical parameters are 100 sccm of O_2 at 100 W for 30 min. total. Often, we pause the etch for 5 min. in the middle to let the wafer cool down. The ptherm is a good tool for this final ash, because it can accept any wafer shape unlike other tools that require a 6″ wafer with a well-defined wafer flat. The state of the wafer is shown topologically in Fig. 3.23(j). At this point, the wafer is ready to be removed from the clean room for testing.

3.3.6 Lessons

In the course of fabricating prototype wafers, many mistakes are inevitably made. It is perhaps amusing to remember some of the more spectacular fabrication failures. Some relevant images are shown in Fig. 3.24. We learned that film stresses are especially important for any released structures, since there is no substrate to keep everything together. We learned that it is important to dilute away etchants to prevent them from slowly over time

chewing away at the on-wafer features. We also learned to control our plasma etches to be less directional to make sure all of the targeted material is removed.

3.4 Testing

After all the trouble of fabricating a device, one must then submit to all the trouble of testing the device. What makes this highly non-trivial is that the devices are designed to work at sub-Kelvin temperatures, so the problem of testing a device is in large part the problem of cooling the device while still allowing in optical signals. The devices are also designed to be extremely sensitive, so great care must be taken to shield the test setup from electromagnetic interference and vibrations. Finally, it turns out to be not at all straightforward to just put everything together without breaking something.

In characterizing our devices, one major obstacle is that it is difficult to isolate the components. A photon that enters the cryostat passes through a window, several filters (to block infrared radiation), a lenslet, a planar antenna, a microstrip line, a bandpass filter and a termination resistor. The signal appears as a temperature fluctuation in the TES that is thermally coupled to the termination resistor. The TES is connected to microstrip lines, wire bonds, a printed-circuit board, to SQUID amplifiers and possibly to flexible cables and LC resonators. While it is possible to characterize the TES portion of this system separately from the optical components, it is very difficult to achieve the opposite, i.e., characterizing the optical components without the TESs. And it is even more difficult – or, at least tedious – to characterize each optical component individually. Typically, we measure the end-to-end performance and try our best to infer which component is responsible for any non-idealities.

3.4.1 Test dewars

The dewars we use for testing are designed similarly to the observation receivers we deploy on our telescopes. These cryostats are non-trivial to design and build, since we simultaneously need to cool our detectors to sub-Kelvin temperatures while letting in microwave radiation from the much warmer outside world. Just achieving a sub-Kelvin temperature is not easy but more or less a solved problem. Since we need to let in microwave signals from the outside, a large amount of effort must be devoted to designing microwave-transmissive but vacuum-sealing windows, which are usually made of judiciously chosen foam, as well as infrared-blocking filters, which are essential for lowering the thermal-radiation load on the cold stages. For reference, the radiation load from a room-temperature (293-K) blackbody on a perfectly-absorbing $6''$ wafer assuming an opening half angle of 45° is at the level of $\mathcal{O}(1)$ W, which can be compared with a typical wafer-stage thermal budget of $\mathcal{O}(5)$ μW. So our filters are achieving something like a million-to-one rejection of thermal radiation. The remaining radiation is usually confined to below 300 GHz if not lower depending on the particular frequency range of interest for science observations. In the field, the radiation load comes mainly from the sky, which has an effective temperature of $\mathcal{O}(10)$ K. In testing, however, we must deal with the roughly 300-K laboratory. For the example scenario outlined above, this produces $\mathcal{O}(10)$ μW of loading. In reality, there are optical inefficiencies at microwave frequencies, so this number can easily come down to $\mathcal{O}(1)$ μW or lower. Since we are now

in the Rayleigh-Jeans portion of the spectrum, this radiative loading scales approximately with the cube of the cutoff frequency, so it is worthwhile to filter aggressively.

The room-temperature thermal radiation also cause problems for our detectors even after the cryogenic issues have been settled. This is because the thermal radiation provides excess optical loading beyond what the detectors have been designed for. Our detectors are single-moded devices, which means that they are loaded only by a single mode of the thermal radiation. The power in a small bandwidth $d\nu$ is $k_B T\ d\nu$. For a 25-GHz bandwidth, which is typical for a passband centered on approximately 100 GHz, the optical loading from the room-temperature thermal radiation is at the level of 100 pW, which is to be compared with typical saturation powers of $\mathcal{O}(10)$ pW. This means that our detectors are virtually guaranteed to saturate in laboratory tests if no additional steps are taken. A common strategy is to include an absorptive filter in the dewar to simply attenuate the microwave signal. This will prevent the bolometers from saturating but may add additional optical systematic errors to the characterization measurements. Another method is to use higher-T_c TESs for laboratory optical testing. This can be accomplished either by having a higher-T_c TES in series with the science-observation TES or by fabricating an entire wafer with a higher T_c. The latter is perhaps a bit wasteful but is fine for prototypes. It is also possible to bite the proverbial bullet on bolometer saturation and operate the bolometer above its superconducting transition where there is still some responsivity due to a non-zero dR/dT.

Typically, the cryogenic are organized so that there is a system with relatively high cooling power that provides a temperature of 4 K or lower and a system that uses this temperature to bring the detectors to sub-Kelvin temperatures. The more traditional method for the super-Kelvin temperatures is to use liquid cryogens. The dewar is structured as a kind of onion with shells at different temperatures. We can use liquid nitrogen to bring the first inner shell to 77 K, and we can use liquid helium to bring the second inner shell to 4 K. The helium can be cooled further to approximately 1.5 K by pumping on the helium vapor to achieve a vapor pressure of 1-2 Torr. The more modern method is to use a helium pulse-tube cooler. This requires electrical power but can run continuously providing approximately 0.5 mW of cooling power at 3-4 K. We usually use pulse tubes on larger dewars and liquid cryogens on smaller dewars.

3.4.2 Sorption refrigeration

Inside the coldest shell, we install a sub-Kelvin cooler. There are many varieties of sub-Kelvin coolers that exploit a diversity of thermodynamic phenomena, but our group has recently been using exclusively ^{3}He sorption refrigerators. The idea is similar to the method used to cool ^{4}He to its lowest temperature: we lower the vapor pressure as much as possible. Since ^{4}He is a boson, it forms a superfluid around 2 K and creeps along the walls of the container forming thermal shorts. This makes it impractical to use ^{4}He to achieve sub-Kelvin temperatures. Since ^{3}He is a fermion, though, we can achieve temperatures as low as approximately 250 mK by lowering the vapor pressure. At this point, the ^{3}He atoms start to combine into effective bosons which create a superfluid.

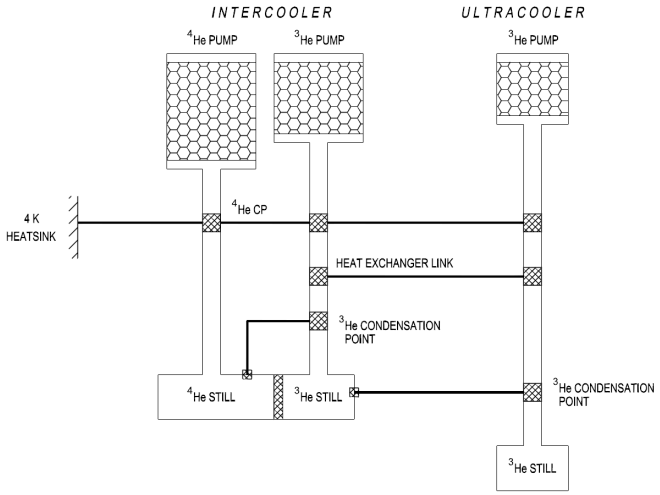
But ^{3}He is rare and, therefore expensive. Unlike many of our ^{4}He systems, in which the pumped gas is simply exhausted to the atmosphere, it is vital to create a closed system that holds on to the pumped ^{3}He. This is the design goal of a ^{3}He sorption refrigerator.

A schematic of a three-stage helium sorption cooler is shown in Fig. 3.25(a), and a fabricated fridge is shown in Fig. 3.25(b). The "intercooler" stages act as thermal buffers for the "ultracooler" stage, which contains the ^{3}He that cools the detectors. Each stage is structured in a similar manner. Each is a closed chamber containing a fixed amount of helium gas. The honeycomb section near the top is meant to indicate the presence of charcoal, whose salient feature in this context is its very large surface area. This large surface area is used to encourage the adsorption of helium atoms. The thin vertical tubes are made of stainless steel, whose salient feature in this context is its very low thermal conductivity. When there is no gas in the chamber, the charcoal is thermally isolated from the still at the bottom, which is where the liquid helium is meant to condense. The bottom of the still can be connected to a heat strap which puts it in thermal contact with the cryostat components we wish to cool to the associated temperature. Each of the tubes has one or more condensation points that are located along the stainless-steel tubes and are simply points of good thermal contact with a relatively cold thermal reservoir. These condensation points are meant to be locations at which the helium gas loses energy, thereby encouraging condensation in the still.

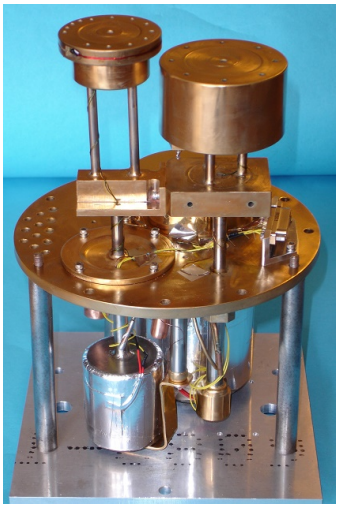
A single stage is operated in the following manner. When no external power is applied, the system asymptotes to the temperature of the so-called 4-K heat sink, which is typically the cold head of a pulse-tube cooler but could also be provided by a large bath of liquid ^{4}He. In any case, we take this heat sink to be relatively stiff with a temperature of something like 4 K. The charcoal is heated electrically to 40 or 50 K. This is hot enough to expel most of the helium that may have adsorbed to the charcoal surface. The chamber is now filled with helium gas. Over time, the gas loses energy to the condensation point, and liquid begins to condense in the still. We reach a steady state, in which there is both liquid and gas helium in the chamber. Now we cut power to the charcoal heater and wait for the charcoal to cool down to approximately 4 K. To accelerate this cooldown, some systems include heat switches which can be activated to put the charcoal in better thermal contact with the 4-K reservoir. As the charcoal cools, more and more helium atoms adsorb to its surface, so the charcoal is essentially pumping on the helium gas and lowering the vapor pressure. The decrease in vapor pressure causes a corresponding decrease in liquid temperature. The system is designed so that the vapor pressure is close to 0 when the charcoal has returned to approximately 4 K. The thermal load on the still causes a steady boil off of the liquid helium. When the helium runs out, the still tends to return back to approximately 4 K.

We can cascade stages to achieve lower and lower temperatures with lower and lower parasitic thermal loads. In Fig. 3.25(a), the warmest stage uses ^{4}He, and the still achieves a base temperature of approximately 1 K. This still is then used as a second condensation on the next coldest stage, which uses ^{3}He. The second-stage still is used as a second condensation point for the coldest stage, which also uses ^{3}He. We cool the ^{4}He still while continuing to heat the ^{3}He charcoal pumps. During this period of time, liquid ^{3}He is condensing in the two coldest stills. Once the liquid ^{4}He has boiled off completely, we cool the ^{3}He charcoal pumps, which lowers the temperatures of the ^{3}He stills. The intermediate stage tends to asymptote to a temperature of approximately 350 mK, while the coldest stage asymptotes to approximately 250 mK.

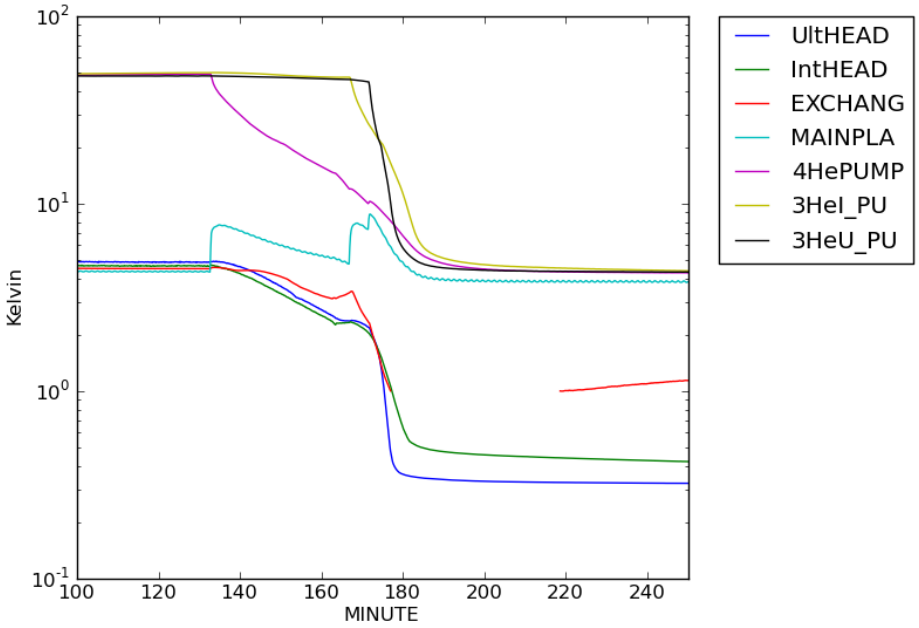
This type of sorption cooler is sometimes called a He10 fridge, since it uses one ^{4}He stage and two ^{3}He stages ($4 + 2 \times 3 = 10$). The heat exchanger shown in Fig. 3.25(a) allows the relatively cold boil-off from the ultracooler to help cool the gas in the intermediate stage.



(a) A schematic of a three-stage helium sorption cooler [17].

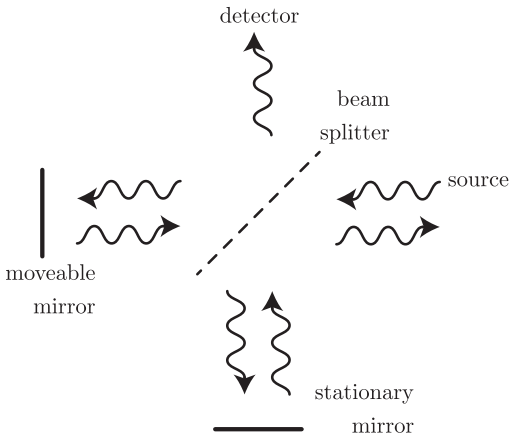


(b) Chase He10 fridge. The orientation here is upside down. (Image credit: Chase Research Cryogenics.)

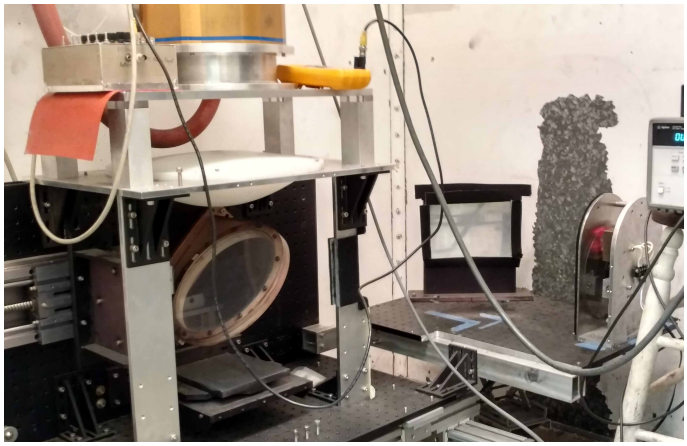


(c) An example fridge cycle. The "pumps" are charcoal. They are held at 50 K during condensation and cooled to 4 K to lower the helium vapor pressure. The ultrahead is the coldest still; the interhead is the intermediate-temperature still. The main plate is the cold head of a pulse-tube cooler; notice that it responds strongly to the dumping of power from the charcoal pumps. The break in the heat-exchanger stream is due to a lack of thermometer-calibration data below 1 K.

Figure 3.25: A three-stage helium sorption refrigerator.



(a) A common design based on a simple Michelson interferometer.



(b) An FTS coupled to an optical dewar for measurements of prototype devices. The red glow from the ceramic heater can be seen at right.

Figure 3.26: Fourier transform spectrometer (FTS).

When the liquid ^{4}He has been exhausted, the heat exchanger responds most obviously and indicates that it is time to cool the ^{3}He charcoal pumps. The heat exchanger also provides some cooling power for an additional temperature stage at approximately 2 K.

The system uses gas-gap heat switches to accelerate the cooling of the charcoal pumps. These heat switches operate on a similar principle to the sorption chambers. Charcoal is used to control how much gas is in a chamber. The only thermal path from one end of the chamber to the other is through the gas. When all of the gas is adsorbed to the charcoal, the ends of the chamber are thermally isolated. When the charcoal is warm and the gas fills the chamber, the two ends are in good thermal contact. So the entire He10 fridge actually contains six charcoal chambers.

The process of bringing the fridge from 4 K to its base temperature of 250 mK is referred to as "cycling" the fridge and involves carefully timed applications of electrical power to heat or cool the six charcoal pumps in the system. The temperature streams from an example fridge cycle are shown in Fig. 3.25(c), where it can be seen that the "ultrahead" is cooled from a little more than 4 K to a little over 300 mK. In this particular system, the readout for the thermometer on the ultrahead was miscalibrated. Using better readout, the temperature was measured to be 270 mK. The other temperatures plotted show the components of the sorption cooler involved in enabling the cooling of the ultrahead.

3.4.3 Fourier-transform spectroscopy (FTS)

We construct a Michelson interferometer as shown in Fig. 3.26(a). We have a source which produces a plane-wave signal. The beam splitter separates the signal into the two arms of the interferometer. The mirrors reflect the signal back to the beam splitter. The signals recombine, and some of the power reaches the detector. The power that does not reach the detector is either dissipated in the optical elements, e.g., a lossy beam splitter, or directed

back towards the source. We control the path difference between the two arms at the sub-wavelength level. As the moveable mirror scans back and forth, the recombined signal passes through interference peaks and troughs. If the source were monochromatic, then a one-way scan of the moveable mirror would produce a sinusoidal response in the detector. For a broadband source, we detect the superposition of many such sinusoidal responses, which is just the inverse Fourier transform of the source spectrum. Taking the Fourier transform gives us the source spectrum. This method is, therefore, called *Fourier transform spectroscopy*. The raw data collected by the detector, i.e., response as a function of mirror position, is called an *interferogram*.

In reality, the beam splitter will not divide the power evenly and may be lossy. At the same time, the detector will have some frequency dependence as well.

In Fig. 3.26(b), we show the FTS that was used for all of the spectra shown in this manuscript. The light source is a ceramic heater, which is placed behind a room-temperature chopper, whose frequency we control and typically set to 10-40 Hz. The detector response is demodulated at this frequency with a lock-in amplifier before being sent to the computer. The heater sits at the focus of a partial parabolic mirror, which parallelizes the rays to couple into the Michelson interferometer. This is the point that corresponds to the source labelled in Fig. 3.26(a). The beam splitter is a thin mylar sheet. The beam-splitter efficiency varies with frequency and has nulls that should be avoided. When measuring in a new frequency range, it is sometimes necessary to use a different thickness for the mylar beam splitter to move the locations of the nulls. The moveable mirror is controlled with a stepper motor. The combined signal travels upwards towards the detector. To better focus the light onto the detector, we use a plastic lens. The window of the test dewar faces downwards, and the detector being measured is centered on the lens.

Define the path difference η to be the extra path length traversed by the signal that reflects off of the moveable mirror. Keep in mind that η is actually *twice* the distance from the beam splitter to the moveable mirror. We define the linear or spectroscopic wavenumber $\sigma = 1/\lambda$, which is really just an inverse wavelength.[4] Let the incident intensity, i.e., from the source, per unit wavenumber be $I_\sigma^{\text{inc.}}$, and let the intensity received by the detector per unit wavenumber be $I_\sigma^{\text{det.}}$. Then we have

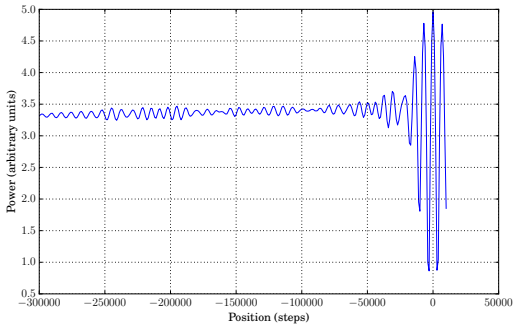
$$I_\sigma^{\text{det.}} = \epsilon_{\text{det.}}(\sigma)\epsilon_{\text{b.s.}}(\sigma) I_\sigma^{\text{inc.}} \frac{1+\cos(2\pi\sigma\eta)}{2}, \tag{3.20}$$

where $\epsilon_{\text{det.}}$ is the detector efficiency and $\epsilon_{\text{b.s.}}$ is the beam-splitter efficiency. We take the definitions of $I_\sigma^{\text{det.}}$ and $I_\sigma^{\text{inc.}}$ to be such that they include the contributions from both positive and negative frequencies. These expressions depend only on the magnitude of the wavenumber, so we integrate only over positive σ to get

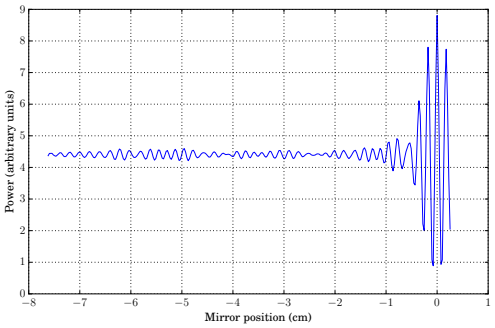
$$I^{\text{det.}} = \int_0^\infty d\sigma\, \epsilon_{\text{det.}}(\sigma)\epsilon_{\text{b.s.}}(\sigma) I_\sigma^{\text{inc.}} \frac{1+\cos(2\pi\sigma\eta)}{2}. \tag{3.21}$$

As the moveable mirror scans, i.e., sweeping in η, the cosine term in the integrand averages

[4] Among mm-wave researchers, it is common to report frequency in units of, e.g., inverse centimeters (cm^{-1}). Strictly speaking, this is a spectroscopic wavenumber. It is proportional to frequency, and it is useful to remember that the conversion is roughly 30 GHz per cm^{-1}.



(a) The raw interferogram.



(b) The same interferogram processed as described in the text.

Figure 3.27: Interferogram from a detector designed to be sensitive to a band centered on 90 GHz. In post-processing, we correct for several effects before computing the Fourier transform.

to 0. The interesting data is then

$$\Delta(\eta) \equiv I^{\text{det.}} - \left\langle I^{\text{det.}} \right\rangle = \int_0^\infty d\sigma\, \epsilon_{\text{det.}}(\sigma)\epsilon_{\text{b.s.}}(\sigma) I_\sigma^{\text{inc.}} \frac{\cos(2\pi\sigma\eta)}{2}, \tag{3.22}$$

where $\left\langle I^{\text{det.}} \right\rangle$ is the detector response averaged over the mirror position. This looks like a Fourier cosine transform, so we can write

$$\epsilon_{\text{det.}}(\sigma)\epsilon_{\text{b.s.}}(\sigma) I_\sigma^{\text{inc.}} = 4 \int d\eta\, \Delta(\eta) e^{-2\pi i \sigma \eta}, \tag{3.23}$$

where the integral is meant to be taken from negative to positive infinity. So we see that the combination of spectra on the left-hand side is proportional to the Fourier transform of the detector signal, i.e., the interferogram. In reality, of course, we will only ever take data over a finite range of mirror positions, so we must estimate the spectra with a well-chosen window function $w(\eta)$, i.e.,

$$\boxed{\epsilon_{\text{det.}}(\sigma)\epsilon_{\text{b.s.}}(\sigma) I_\sigma^{\text{inc.}} \approx 4 \int d\eta\, w(\eta)\Delta(\eta) e^{-2\pi i \sigma \eta}.} \tag{3.24}$$

In practice, the normalization is often very difficult to predict or measure; the proportionality is the most important part of this result.

Interferogram processing

A typical raw interferogram is shown in Fig. 3.27(a). There are several things to notice here. The mirror position is given in motor steps, which must later be translated into physical distances. The zero of the stepper motor's coordinate system is chosen to coincide with the peak of the interferogram. This peak occurs when the path difference between the two arms

of the FTS is 0. At this position, every frequency experiences *constructive* interference, i.e., $\cos(2\pi\sigma\eta) = 1$ in Eq. 3.21. We call this the point of *zero path difference*, which is sometimes abbreviated as ZPD. Away from $\eta = 0$, the frequencies all beat against each. Averaging over mirror positions, the cosine in Eq. 3.21 vanishes. Then the ratio of the peak value to the average should be $2:1$. But it's pretty obvious in Fig. 3.27(a) that the ratio is *not* $2:1$. As we will discuss below, this can be attributed to the non-linear response of the detector and can be corrected for. Another thing to notice in Fig. 3.27(a) is that the data have not been taken symmetrically about the ZPD. A small amount of data have been taken to the right of the ZPD, and a large amount to the left. This is purely a time-saving choice. The interferogram is expected to be symmetric about the ZPD, so we only take a long sweep in one direction. The small amount of data to the right of the peak is used to fine tune the actual point of zero path difference, which does not in general lie exactly on one of the data points. This will be discussed below in the section on phase correction. Finally, it is obvious by eye in Fig. 3.27(a) that the measured interferogram does not fluctuate about a flat line. We sometimes call this a tilt. It is due to imperfect focusing in the optical chain. If the rays are not all traveling perfectly parallel but instead diverge a bit, then some rays will miss the moveable mirror as it backs away from the beam splitter. This is, indeed, what we see in Fig. 3.27(a). As the mirror coordinates becomes more negative, i.e., the mirror moves away from the beam splitter, the power received by the detector decreases. We try our best achieve a good alignment in the optics before taking data, but at some point it is better to just remove the tilt in post-processing. This will be discussed below.

Tilt A tilt in the interferogram is actually almost entirely tolerable. Since it looks like a long-wavelength mode, it shows up in the Fourier transform as a spurious low-frequency signal. It is usually at such low frequencies, that it is obvious even by eye that it is artificial. Typically, the tilt does not significantly affect the main frequency content of the signal. It is pretty easy to remove, however, so there is little cost to adding this functionality to our analysis scripts. We simply fit a line to the interferogram. The one subtlety is that the large oscillations near the ZPD tend to strongly influence the best-fit line; there just aren't enough oscillations for them to average away. So it is better to fit to the portion of the interferogram that does not include such wild oscillations, e.g., the left-most 80% of the signal shown in Fig. 3.27(a). Then we simply subtract the linear component but not the offset, since we are still interested in relating the peak to the average level.

Phase correction During data taking, we try our best to set the origin of the stepper motor's coordinate system at the ZPD position. Since we sample discretely, however, it is usually the case that the true ZPD is somewhere in between two data points. We can correct for this by appealing to the symmetry of the underlying interferogram. Let's refer to $\Delta(\eta)$ as the interferogram, which is true up to an overall offset, and $\tilde{\Delta}(\sigma)$ as its Fourier conjugate. When the origin of the η-axis is at the true ZPD position, the interferogram is a real even function, which immediately implies that $\tilde{\Delta}(\sigma)$ is real and even. We have a small amount of data to the right of the peak in Fig. 3.27(a). We take this data and the same amount from the left-hand side of the peak and compute the Fourier transform of this small subset of the data. If the ZPD position were actually at the zero of the η-axis, then this Fourier

transform would pure real. If, instead, the interferogram is translated a little away from the origin, the Fourier transform will pick up a small phase shift and, therefore, some non-zero imaginary components. The phase shift is

$$\theta = -\arctan \frac{\mathrm{Im}\left[\tilde{\Delta}(\sigma)\right]}{\mathrm{Re}\left[\tilde{\Delta}(\sigma)\right]}, \tag{3.25}$$

which can be translated to the real-space translation

$$\Delta\eta = \frac{\theta}{2\pi\sigma}. \tag{3.26}$$

Sending

$$\tilde{\Delta}(\sigma) \to e^{2\pi i\sigma\Delta\eta}\tilde{\Delta}(\sigma) \tag{3.27}$$

corrects the Fourier transform. Now we can take the inverse Fourier transform to recover an untranslated interferogram. A subtlety here is that the phase shift θ calculated from the data may not be proportional to σ, in which case $\Delta\eta$ is not constant with frequency. In this case, it usually works pretty well to simply take the mean $\Delta\eta$. After taking the inverse Fourier transform, one can check what the maximum imaginary component in the corrected interferogram is as a way of monitoring any unphysical effects of the phase correction. It had also better be true that the estimated $\Delta\eta$ is small compared with the mirror step size, since the raw data already show a clear peak near the zero of the η-axis.

Non-linearity A non-linearity can easily arise in the detector response, especially if the TESs are operating in the upper part of or above the superconducting transition. At first order, we model the non-linearity as a quadratic component in the detector response, e.g.,

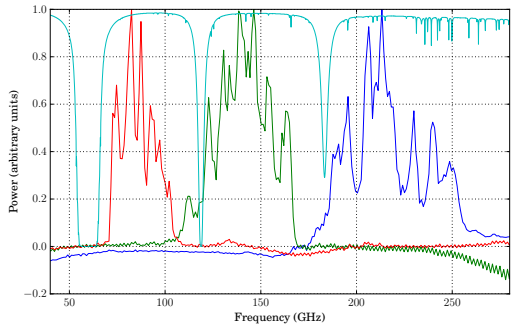
$$f_{\mathrm{meas.}} = f_{\mathrm{true}} + \alpha f_{\mathrm{true}}^2, \tag{3.28}$$

where $f_{\mathrm{meas.}}$ is the measured signal and f_{true} is the true signal. The quadratic term introduces a so-called "2F" signal, i.e., a signal at the twice the frequency of the true signal. This is just because $\cos^2(k\eta) = [1 + \cos(2k\eta)]/2$. This can be seen in the raw Fourier transforms shown in Fig. 3.28(a), where the 90-GHz band has a small bump around 180 GHz and the 150-GHz band near 300 GHz. The other most obvious effect of non-linearity is the suppression or enhancement of peaks in the interferogram. When α is negative, the peaks near the ZPD position tend to be suppressed. This is why the ratio of the peak to the average level is not 2 : 1. By estimating α, we can undo the non-linear effects in post-processing. The true signal can be written in terms of the measured signal as

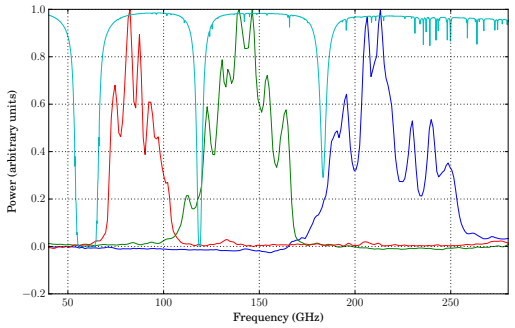
$$f_{\mathrm{true}} = \frac{1 - \sqrt{1 - 4\alpha f_{\mathrm{meas.}}}}{2\alpha}, \tag{3.29}$$

where we have taken the minus sign in the quadratic formula so that, for small α, we have $f_{\mathrm{true}} \approx f_{\mathrm{meas.}}$. Now we can optimize α so that the peak-to-average ratio is 2 : 1, which is best done numerically.

An interferogram that has been processed to correct for all of the effects described above is shown in Fig. 3.27(b). This flatness and peak-to-average ratio are noticeably improved. The phase correction is not obvious, since the raw interferogram was close to centered to begin with.



(a) The spectra without any of the processing described in the text.



(b) The spectra with all of the processing described in the text.

Figure 3.28: FTS spectra (peak-normalized) for detectors sensitive to bands centered on 90, 150 and 220 GHz.

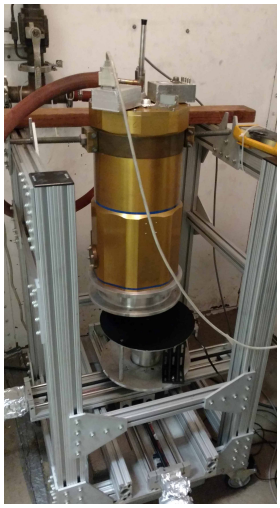
Fourier transform

After subtracting the tilt, phase correcting and removing the non-linear detector response, we are nearly ready to take the Fourier transform of our interferogram. We need to apodize the interferogram to avoid ringing in the spectrum; see, for example, the 150-GHz band in Fig. 3.28(a), which shows very regular ringing out of band. Since we have phase corrected, we feel comfortable filling in what would have been the right half of the interferogram in, e.g., Fig. 3.27(a). With this full, symmetric interferogram, we apply a window function. A triangular apodization is a common choice. Now, we finally take the discrete Fourier transform (DFT) of our processed interferogram. Keep in mind that the step size, i.e., $\Delta\eta$, sets the Nyquist frequency $\sigma_{\text{Nyq.}} = 1/(2\Delta\eta)$, which is the highest frequency to which we are sensitive. The frequency resolution is set by the greatest extent of the mirror from the ZPD position, i.e., $\Delta\sigma = 1/(2\eta_{\text{max.}})$.

The Fourier transforms of some raw interferograms, which have not been flattened, phase corrected, linearized nor apodized, are shown in Fig. 3.28(a). If the spectrum extended further to the left, we would see large low-frequency responses below 10 GHz, which are due to the tilt in the interferograms that looks like a long-wavelength signal. The spectra. The 90- and especially the 150-GHz bands show 2F bumps, which are due to non-linear bolometer responses. All three spectra show fast ringing with a period of a few GHz, which is due to the lack of a smooth window function. The effect of an absent phase correction is not as obvious as the other effects. It tends to be small, since we set up the FTS so as to drive the required phase correction to zero to begin with. Essentially, though, a translated interferogram will steal power from the real part of the Fourier transform and move it to the unphysical imaginary part.

The Fourier transforms of the same interferograms but now with all of the processing described above are shown in Fig. 3.28(b). It cannot be seen in this image, but the spurious low-frequency responses have been suppressed. The 2F bumps have disappeared. The fast ringing has disappeared as well.

(a) Liquid-N_2 chopped source.



(b) XY-stage.

Figure 3.29: Beam mapper.

3.4.4 Beam map

The beam of an antenna is its sensitivity to optical signals as a function of angle. This is sometimes called a radiation pattern when the system has a time-reversal symmetry, i.e., if we were to excite the antenna, it would radiate in each direction with an intensity proportional to the beam. We map the beam by moving a source in front of the detector. It is common to use a chopped thermal source similar to what is used for FTS measurements. Some ceramic heaters, however, have an internal structure which can affect the beam measurement. If this is the case, it is better to use, e.g., a liquid-nitrogen source. This can be constructed as a basin lined with absorbing material, e.g, ECCOSORB, that is filled with liquid nitrogen. Above this basin, we place a room-temperature chopper. Above the chopper, we place an absorbing (from the perspective of the detector) plate with a small aperture in it. This type of setup is shown in Fig. 3.29(a). The detector sees a room-temperature absorber with a small aperture in it, which is chopping between room temperature and 77 K. Just as for the FTS, the chopping frequency is controlled by the user, and the bolometer time stream is demodulated at this frequency using a lock-in amplifier. This source is bolted to the top of two orthogonal stepper-motor tracks, so the source can be moved in an xy-plane in front of the dewar window. This is shown in Fig. 3.29(b). The source is commanded to raster scan in front of the detector, and the bolometer response is recorded at each position.

In post-processing, we often remove two geometric effects. One is a simple projection effect, i.e., away from the zenith the effective area of the source as seen by the detector is smaller by a factor of $\cos\theta$, where θ is the polar angle. The other effect is due to the fact that the source is not at the same distance from the detector at all points in the scan. Assuming the source intensity falls off as an inverse-square law, the signal received by the detector will be reduced by a factor of $\cos^2\theta$. Overall, then, the data is related to the true beam by

$$I_{\text{true}} = \frac{I_{\text{meas.}}}{\cos^3\theta}. \tag{3.30}$$

To find the center of the beam, it would be possible to achieve a very careful mechanical alignment given the measured positions of all of the relevant components. This is, however, an impractical method. It is more convenient to simply record the data and only in post-processing estimate the beam center. It is typical to fit the beam to an elliptical Gaussian. We introduce a number of free parameters for this fit. First of all, we allow the orientation of the ellipticity to be rotated with respect to the xy-coordinates defined by the stepper motors. So we have the new coordinates

$$x' = \cos\phi_0(x - x_0) + \sin\phi_0(y - y_0), \qquad y' = -\sin\phi_0(x - x_0) + \cos\phi_0(y - y_0), \tag{3.31}$$

where ϕ_0 is the azimuthal rotation angle of the ellipticity relative to the beam-map raster scan and (x_0, y_0) are the true coordinates of the beam center. So the new coordinates x' and y' are defined relative to the beam center. Let $s = \sqrt{x'^2 + y'^2}$ be the in-plane distance from the beam center, and let z be the distance from the detector to the beam-mapper source. Then we can define the polar angle θ and azimuthal angle ϕ with

$$\cos\theta = \frac{z}{\sqrt{s^2 + z^2}}, \qquad \sin\theta = \frac{s}{\sqrt{s^2 + z^2}}, \qquad \cos\phi = \frac{x'}{s}, \qquad \sin\phi = \frac{y'}{s}. \tag{3.32}$$

If we assume an elliptical Gaussian near-field illumination of the form

$$E \propto \exp\left[-\left(\frac{x'}{w_{0,x}}\right)^2 - \left(\frac{y'}{w_{0,y}}\right)^2\right], \tag{3.33}$$

where $w_{0,x}$ and $w_{0,y}$ are often called *Gaussian beam waists*, then the far field has the form

$$E \propto \exp\left[-\left(\frac{w_{0,x}k\sin\theta\cos\phi}{2}\right)^2 - \left(\frac{w_{0,y}k\sin\theta\sin\phi}{2}\right)^2\right], \tag{3.34}$$

which is just the Fourier transform of the near field as prescribed by, e.g., the Huygens-Fresnel principle. To be clear, $k\sin\theta\cos\phi$ is just the x-component of the wavevector $\mathbf{k}$, and $k\sin\theta\sin\phi$ is just the y-component. The intensity is proportional to the squared magnitude of the electric field. Adding the suppression due to the geometric effects described above and an overall normalization I_0, the model for the measured intensity becomes

$$I_{\text{meas.}} = I_0\cos^3\theta\exp\left[-\left(\frac{w_{0,x}k\sin\theta\cos\phi}{\sqrt{2}}\right)^2 - \left(\frac{w_{0,y}k\sin\theta\sin\phi}{\sqrt{2}}\right)^2\right]. \tag{3.35}$$

This is really a six-parameter model; we can vary ϕ_0, x_0, y_0, I_0, $w_{0,x}$ and $w_{0,y}$. We take many more than six data points, though.

It should be emphasized, though, that the main utility of this Gaussian beam model is not to test whether our detectors perform as designed, since the beams do not need to be and, in many cases, are not even expected to be Gaussian. The main utility here is simply in locating the beam center. Even when the Gaussian fit is not very good, e.g., the beam is peakier than a Gaussian, the center is still found quite robustly. Once we have the beam center, we have a zenith from which to measure angles. The angular sensitivity of our detectors is directly tied to the overall sensitivity of an experiment.

3.4.5 Polarization

We are also interested in the polarization response of our detectors, since we are, after all, trying to measure the polarized fluctuations in the CMB. It is not actually necessary for the detectors to be perfectly polarized. Any polarization leakage will just translate to an efficiency hit in making polarized maps of the CMB. We typically characterize the polarization response by rotating a polarizer in front of the detectors. The detector response is approximately sinusoidal in the rotation angle of the polarizer. When the polarizer is aligned with the detector polarization, we see a large signal. When the polarizer is perpendicular to the detector polarization, we see a small signal. The peak-to-trough ratio indicates the polarization efficiency of the system. Polarization efficiency is not consistently defined in our field, so it is a good idea to be clear about what one means precisely when reporting a value. Sometimes the ratio of the minimum to the maximum is reported, i.e., min./max., which gives some idea of the polarization leakage; the smaller this ratio, the more polarized the detector. Sometimes the ratio of the difference to the sum is reported, i.e., $(\text{max.} - \text{min.})/(\text{max.} + \text{min.})$, which gives some idea of the polarization efficiency; the closer this value is to unity, the better polarized the detector is. For a small minimum-to-maximum ratio, this second convention produces $1 - 2$ min./max., so the inferred "leakage" is actually twice that of the first convention.

There are several non-idealities that complicate these types of measurements. First of all, the polarizer is usually not perfect, and it is difficult to estimate exactly how polarized it really is. Most non-idealities in the system reduce the measured polarization efficiency, so every measurement is a kind of lower bound on the polarization efficiency of the polarizer. We tend to use a wire-grid polarizer, which consists of closely-spaced (compared with a wavelength) parallel wires. For electric fields parallel to the wires, the wire grid acts like a conducting boundary and reflects most of the incident power. For electric fields perpendicular to the wires, the electrons in the wires cannot rearrange themselves to respond, so the power is mostly transmitted. The polarization leakage increases with frequency, since the smaller wavelengths can resolve the gaps in the wire grid. It is typical for such wire grids to have an intrinsic polarization efficiency of 97% or higher, i.e., $1 - 2$ min./max. $= 97\%$.

Since the polarizer reflects a substantial amount of power, it is advisable to tilt it about an axis perpendicular to the optical axis in order to suppress standing waves. A common symptom of standing waves is a non-sinusoidal response as a function of the polarizer rotation angle. Often, the troughs become wider, and the peaks narrower. Usually, it is sufficient to tilt the polarizer by 10-15°. One must, however, account for this tilt in analyzing the measurements, since the projected polarizer rotation angle differs from the commanded angle. The relationship is

$$\phi_{\text{eff.}} = \arctan\left(\frac{\tan\phi}{\cos\theta}\right), \tag{3.36}$$

where ϕ is the commanded polarization angle of the polarizer, θ is the tilt angle and $\phi_{\text{eff.}}$ is the effective polarization angle of the polarizer.

The polarizer test setup can be combined with the beam mapper as shown in Fig. 3.30. A wire grid controlled by a stepper motor is placed directly on top of the beam mapper. Typical choices for the thermal source are heaters or liquid-nitrogen baths. An xy location is chosen, usually the beam center, although we could, in principle, perform this measurement

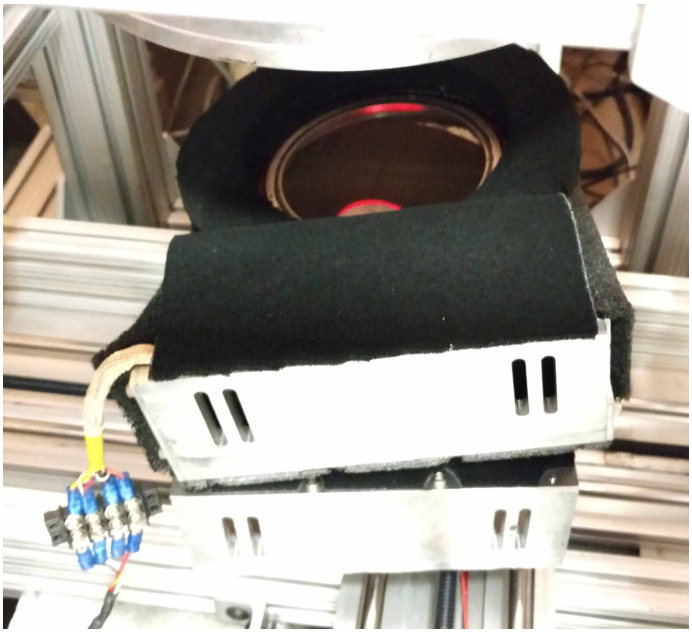


Figure 3.30: Rotating wire-grid polarizer.

at every xy location to produce a *polarized* beam map. The wire grid is tilted 10-15° out of the xy-plane to suppress standing waves. Then we simply command the stepper motor to rotate the wire grid and take data at many rotation angles.

We provide here a model for analyzing polarization measurements. First note that a perfect polarizer would produce a response in a perfectly polarized detector that goes as $\cos^2(\phi - \phi_{\text{det.}})$, where $\phi_{\text{det.}}$ is the polarization angle of the detector. But now we allow for all manner of imperfections. Suppose the source is somewhat but not necessarily completely polarized with ϕ_{source} as the angle of maximum polarization. Let a be the polarized fraction of the power emitted by the source. Then $1 - a$ is the unpolarized fraction. No matter what angle the polarizer has been rotated to, half of the source power is polarized parallel to the polarizer. Then the power emerging from the polarizer that is polarized parallel to the polarizer is

$$P_{\parallel}^{\text{pol.}} = \epsilon_{\parallel}^{\text{pol.}} \left[\frac{1}{2}(1-a) + a\cos^2(\phi_{\text{eff.}} - \phi_{\text{source}}) \right], \tag{3.37}$$

where $\epsilon_{\parallel}^{\text{pol.}}$ is the fraction of co-polarized power transmitted through the polarizer; we expect $\epsilon_{\parallel}^{\text{pol.}} \approx 100\%$. The power emerging from the polarizer that is polarized perpendicular to the polarizer is

$$P_{\perp}^{\text{pol.}} = \epsilon_{\perp}^{\text{pol.}} \left[\frac{1}{2}(1-a) + a\sin^2(\phi_{\text{eff.}} - \phi_{\text{source}}) \right], \tag{3.38}$$

where $\epsilon_{\perp}^{\text{pol.}}$ is the fraction of cross-polarized power transmitted through the polarizer; we expect $\epsilon_{\perp}^{\text{pol.}} \approx 2\%$. Now we move on to the detector. The power received by the detector that is polarized parallel to the detector's polarization axis is

$$P_{\parallel}^{\text{det.}} = \epsilon_{\parallel}^{\text{det.}} \left[P_{\parallel}^{\text{pol.}} \cos^2(\phi_{\text{eff.}} - \phi_{\text{det.}}) + P_{\perp}^{\text{pol.}} \sin^2(\phi_{\text{eff.}} - \phi_{\text{det.}}) \right], \tag{3.39}$$

where $\epsilon_{\parallel}^{\text{det.}}$ is the fraction of incident co-polarized power received by the detector. Similarly, the power received by the detector that is polarized perpendicular to the detector's polarization axis is

$$P_{\perp}^{\text{det.}} = \epsilon_{\perp}^{\text{det.}} \left[P_{\parallel}^{\text{pol.}} \sin^2(\phi_{\text{eff.}} - \phi_{\text{det.}}) + P_{\perp}^{\text{pol.}} \cos^2(\phi_{\text{eff.}} - \phi_{\text{det.}}) \right], \tag{3.40}$$

where $\epsilon_{\perp}^{\text{det.}}$ is the fraction of incident cross-polarized power received by the detector. The total power received by the detector is

$$P^{\text{det.}} = P_{\parallel}^{\text{det.}} + P_{\perp}^{\text{det.}} \tag{3.41}$$

There are typically large degeneracies among all of the parameters in this model, so it is helpful to have a prior on some of them. Many of the parameters are independent of the detector, so they can be estimated from multiple measurements with different detectors.

3.4.6 Optical efficiency

We can estimate the optical power received by a TES by appealing to the P-R curves described above. In the constant-power portion of the TES transition, the sum of the electrical and optical power is constant, i.e.,

$$P_{\text{elec.}} + P_{\text{opt.}} = \text{const.} \tag{3.42}$$

We know $P_{\text{elec.}}$, because we can measure the resistance and we know how much voltage bias we are applying. It is trickier to measure $P_{\text{opt.}}$ directly. However, we can measure *differences* in $P_{\text{opt.}}$ by measuring differences in $P_{\text{elec.}}$. We have the simple relation

$$\Delta P_{\text{opt.}} = -\Delta P_{\text{elec.}}. \tag{3.43}$$

For well-understood optical loads like a beam-filling blackbody of known temperature, we can predict $\Delta P_{\text{opt.}}$ for a given bandwidth up to an overall constant of proportionality, which is the efficiency. An example of such a measurement is shown in Fig. 3.31, which shows the P-R curves for the same bolometer subjected to blackbody optical loads of different temperatures. We see that the constant-power portion of the transition is noticeably smaller for the higher-temperature load.

There is no standard definition in our field for optical efficiency, which causes quite a lot of confusion and misunderstanding. Here we give some recommendations.

If a single number is required, then it should be the ratio of the received optical power to the design value. In this case, the shape of the spectra should be ignored so long as there is no significant response outside of the design frequency range.

If it is desired to maintain information about the shape of the spectrum, then the efficiency as a *function* of frequency should be computed. With this function, the efficiency in any bandwidth can be computed simply by integrating. Typically, the spectrum measured with an FTS has an uncalibrated normalization, which is why FTS spectra are often peak-normalized. Below, we outline a procedure for calibrating these spectra.

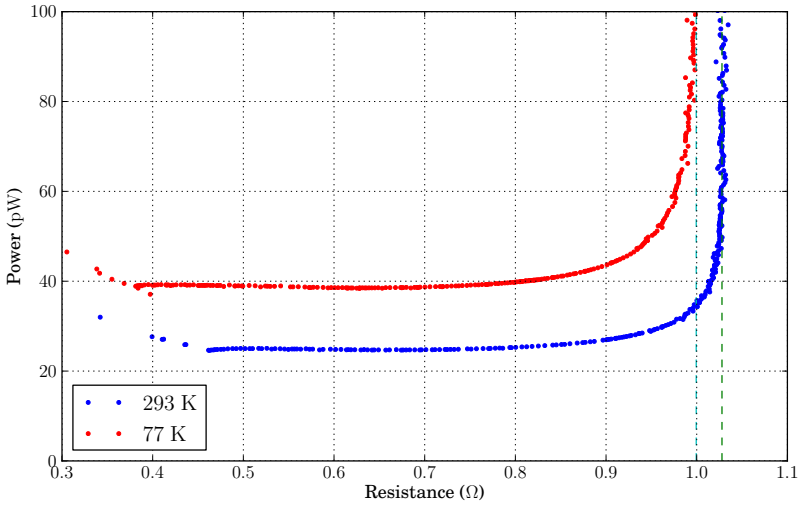


Figure 3.31: Example P-R curves with 77- and 293-K optical loads.

Definitions and assumptions

The measured spectrum with arbitrary normalization is denoted $\eta_{\text{meas.}}(\nu)$, where ν is frequency. The calibrated spectrum is denoted $\eta(\nu)$. This is a dimensionless quantity for the efficiency as a function of frequency. We take the calibrated spectrum to be proportional to the measured spectrum. The constant of proportionality is denoted

$$\eta_0 = \eta(\nu)/\eta_{\text{meas.}}(\nu). \tag{3.44}$$

The optical efficiency is measured by chopping between two sources of different temperatures and measuring the change in electrical power on the bolometer. The temperatures of the sources are assumed to be high enough that the measured spectrum is safely in the Rayleigh-Jeans limit. The difference of these two temperatures is denoted ΔT. The bolometer is assumed to be operating at the same point in the transition for both sources, so the magnitude of the change in electrical power, which we denote ΔP, is equal to the magnitude of the change in optical power.

Calibrating the spectrum

The power received by the detector from a single mode of a perfect blackbody source of temperature T is

$$P = \int_0^\infty d\nu\, \eta(\nu) \frac{h\nu}{e^{h\nu/(k_B T)} - 1,} \tag{3.45}$$

where h is Planck's constant and k_B is Boltzmann's constant. In the Rayleigh-Jeans limit, this simplifies to

$$P = \int_0^\infty d\nu\, \eta(\nu) k_B T. \tag{3.46}$$

Then the change in optical power from chopping between two sources of different temperatures is

$$\Delta P = k_B \Delta T \int_0^\infty d\nu\, \eta(\nu). \tag{3.47}$$

The problem is that we measure an uncalibrated spectrum $\eta_{\text{meas.}}(\nu)$, which is related to the real efficiency by

$$\eta(\nu) = \eta_0 \eta_{\text{meas.}}(\nu). \tag{3.48}$$

We can plug this into Eq. 3.47 to get

$$\eta_0 = \frac{\Delta P}{k_B \Delta T \int_0^\infty d\nu\, \eta_{\text{meas.}}(\nu)}. \tag{3.49}$$

Notice that, if $\eta_{\text{meas.}}(\nu)$ is peak-normalized, then η_0 is the *actual* peak efficiency. We can now rewrite Eq. 3.48 as

$$\boxed{\eta(\nu) = \frac{\Delta P}{k_B \Delta T \int_0^\infty d\nu\, \eta_{\text{meas.}}(\nu)} \eta_{\text{meas.}}(\nu).} \tag{3.50}$$

Here we have an expression for the efficiency as a function of frequency that is written entirely in terms of measured quantities.

Top-hat spectrum In the special case of a peak-normalized top-hat spectrum for $\eta_{\text{meas.}}(\nu)$, the calibrated efficiency becomes

$$\eta(\nu) = \frac{\Delta P}{k_B \Delta T \Delta \nu}, \tag{3.51}$$

where $\Delta\nu$ is the width of the top hat. This same sort of relation could be used to define an *effective* bandwidth, i.e., the bandwidth over which perfect efficiency would yield the same measured power difference.

Procedure

Measure a spectrum $\eta_{\text{meas.}}(\nu)$ with the FTS with arbitrary normalization. Chop between two temperature sources with temperature difference ΔT, and measure the magnitude of the change in electrical power ΔP at the same point in the transition. Then the efficiency as a function of frequency is

$$\eta(\nu) = \frac{\Delta P}{k_B \Delta T \int_0^\infty d\nu\, \eta_{\text{meas.}}(\nu)} \eta_{\text{meas.}}(\nu). \tag{3.52}$$

Chapter 4

Multichroic pixels

In this chapter, we discuss the development of *multichroic pixels*, i.e., pixels that are simultaneously sensitive to multiple frequency bands. This is an efficient design for increasing the sensitivity of an experiment, since the focal-plane area is shared by all frequency bands, which reduces the number of wafers and, ultimately, the number of receivers that must be fabricated.

4.1 Extended hemispherical lenslet

There are several advantages to coupling a planar antenna to a dielectric *lenslet*, which is nothing more than a small lens placed on top of the antenna.

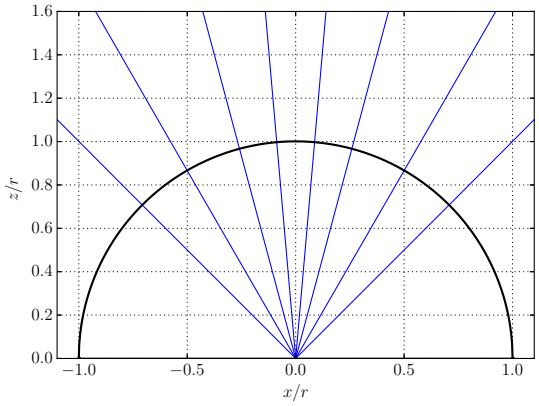
The dielectric will pull the electric fields to one side of the antenna plane, so the radiation pattern will be enhanced on one side and suppressed on the other. The idea is to create what looks locally like an infinite dielectric half space, which will pull the fields preferentially to one side. The higher the dielectric constant, the more pronounced the effect. We typically use materials with high dielectric constants like alumina ($\epsilon \approx 10$) and silicon ($\epsilon \approx 12$).

This lens will also serve to focus the beam further than the planar antenna would have on its own. But it does so above the plane of the antenna, so it does not take up valuable focal-plane real estate for microwave circuitry, bolometers, wiring, etc.

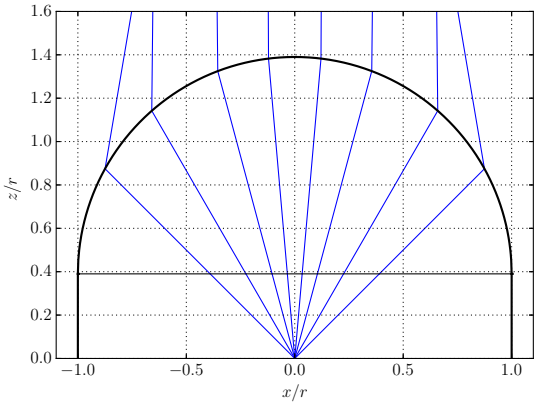
We use an extended hemisphere, since the geometries involved are relatively simple to fabricate. A more ideal lens might have been designed to be ellipsoidal, but a profile like that would be difficult to machine. The extended hemisphere can be thought of as a kind of approximation to an ellipsoid. The extension length can be tuned to maximize the forward gain of the antenna. The hemisphere can be taken from a precision ball bearing that is cut or ground in half.

Some ray diagrams of for extended hemispherical silicon ($\epsilon = 11.7$) lenslets are shown in Fig. 4.1.[1] An interesting parameter to optimize is the ratio of the extension length L to the radius of the hemisphere R. When there is no extension length, i.e., $L/R = 0$, all rays from the central source, which we can imagine to be our planar antenna, strike the lens-air interface at normal incidence and, therefore, experience no refraction; there is no focusing in this case. As we increase L/R, we reach a regime in which the central rays are refracted

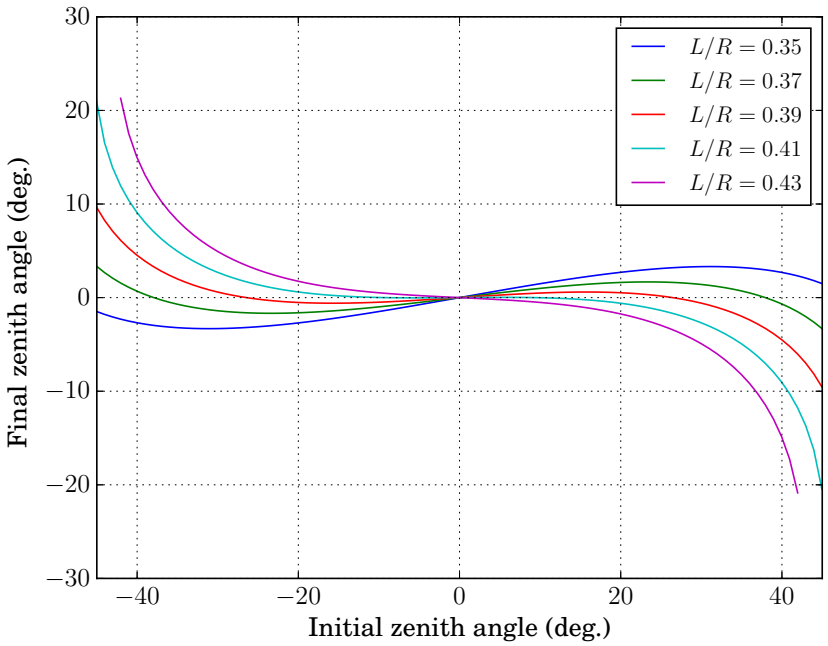
[1] For a copy of the script that generates such ray diagrams, email `ajcukierman@berkeley.edu`.



(a) $L/R = 0$. The lens has no refractive power.

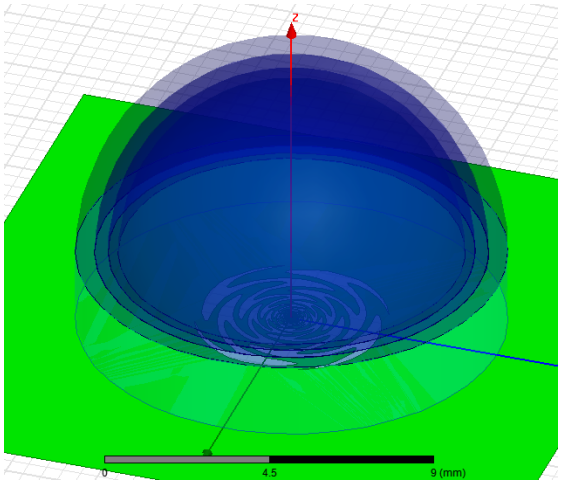


(b) $L/R = 0.39$. The lens focuses incoming parallel rays.

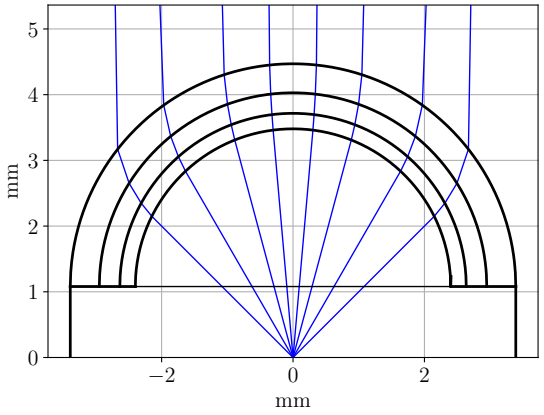


(c) Final zenith angle, i.e., the angle between the zenith direction and the ray outside the lens, as a function of the initial zenith angle, i.e., the angle near the source. Sweep in the ratio L/R to find an optimum, for which the central rays (low initial zenith angle) are mapped to small final zenith angles. The optimum in this case appears to be near $L/R \approx 0.37$.

Figure 4.1: Performance of extended hemispherical silicon ($\epsilon = 11.7$) lenses.



(a) HFSS model of a sinuous antenna coupled to a lenslet with three layers of anti-reflection coating.



(b) Ray diagram for an extended hemispherical lenslet with three layers of anti-reflection coating. The lenslet diameter is *not* the same as that in the HFSS model at left.

Figure 4.2: Planar antenna coupled to an extended hemispherical lenslet with anti-reflection coating.

toward the zenith. If we continue increasing L/R, we will begin to suffer from total internal reflection at the lens-air interface. At extreme values of L/R, the rays will simply miss the hemisphere altogether (at least for the first reflection or refraction point).

We can create a mapping from the ray angle exiting the source and the ray angle outside the lens. We measure these angles relative to the zenith. A lens that focuses well will map every initial ray to an angle close to 0°, i.e., a ray propagating nearly straight upwards. In Fig. 4.1(c), we show this mapping for various values of L/R for extended hemispherical silicon lenslets. For $L/R \approx 0.37$, the initial angles from approximately $-40°$ to $40°$ are mapped to within a few degrees of the zenith. Since our planar antennas our typically designed to radiate most strongly near the zenith anyway, i.e., the central rays account for the bulk of the radiation, the degradation in the performance of the lens for large initial zenith angles is not worrisome.

To increase the refractive power of the lens, we use materials with high dielectric constants, e.g., alumina ($\epsilon \approx 9.7$) and silicon ($\epsilon \approx 11.7$). At normal incidence, approximately 30% of the radiation will be reflected at the lens-air interface. To ease the transition from the lens dielectric to free space, we apply an anti-reflection coating to the surface of the hemisphere. Some diagrams are shown in Fig. 4.2. As will be discussed later, the anti-reflection coating for the lenslets tends to much more forgiving than that for a plane wave incident on a flat multilayer stack of dielectrics. Nevertheless, care must be taken to apply the anti-reflection coating uniformly to avoid distorting the beam or creating scattering sites.

4.1.1 Simulation

Some results from a finite-element-method (FEM) simulation using HFSS are shown in Fig. 4.3. Near the planar antenna, the wave fronts are curved. The lenslet flattens the wave fronts at the tip of the lenslet. We sometimes refer to this flattened electric-field structure as the focal-plane *illumination*. The illumination acts as an effective diffraction aperture, the Fourier transform of which is the far-field radiation pattern. In general, we want the effective diffraction aperture to be as close as possible in size to the lenslet footprint, since we are otherwise wasting focal-plane real estate. Typically, the effective diameter of the illumination is 60-70% of that of the lenslet. We usually set up our HFSS simulations, so that the walls of the box are so-called *radiation boundaries*, which do not reflect the radiating component of the near-field pattern. The far-field radiation pattern is calculated based on the field structure at these near-field radiation boundaries. It can be seen in Fig. 4.3 that some power is traveling horizontally. For electrically large lenslets, e.g., with diameters greater than three free-space wavelengths, this effect is negligible. For electrically small lenslets, however, the lenslet-coupled antenna will begin to couple to its neighbors. We call this *mutual coupling*, i.e., when the performance of the antenna depends on the presence of neighboring structures. This will be discussed in more detail in Sec. 4.4.

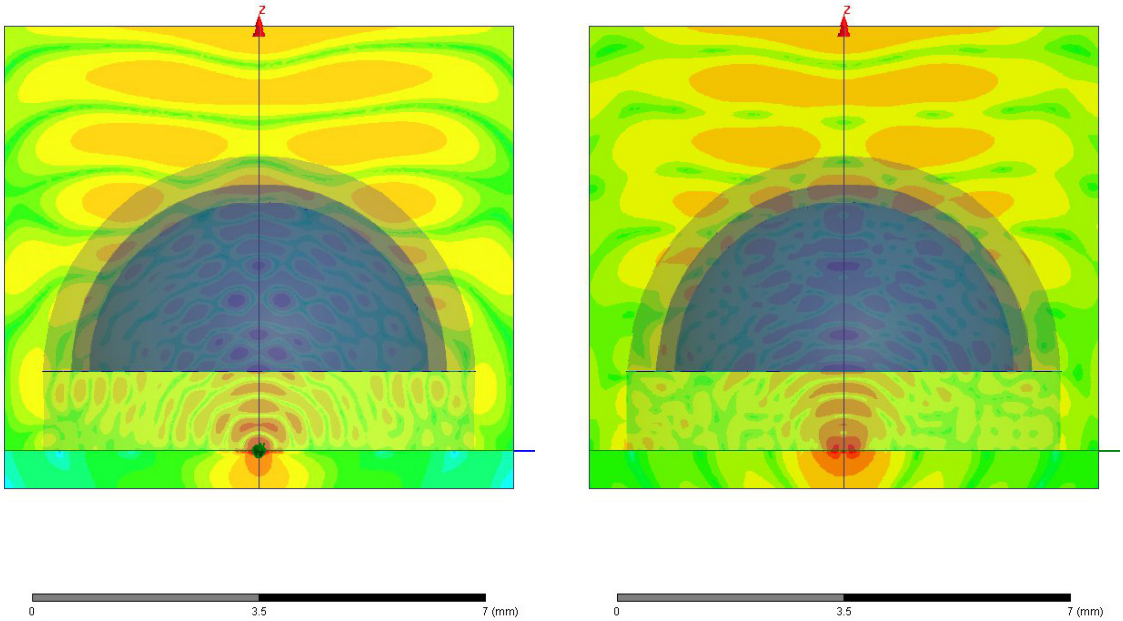
4.2 Double-slot-dipole antenna

The double-slot dipole is a relatively narrowband planar antenna that was chosen for use in PB-1 [9, 10, 66, 67]. Due to its small size and simple design, it is useful for prototype pixels that do not require a large bandwidth. The antenna relies on half-wavelength slot dipole antennas.[2] The beam of a single dipole antenna is very broad, so this design places two dipole approximately one quarter-wavelength away from each other to form a miniature antenna array that narrows the beam. For sensitivity to both linear polarization, two such pairs of orthogonal slot dipoles are superimposed to form the tic-tac-toe patterns shown in Fig. 4.4. In this design, each pair of parallel slots is fed by two out-of-phase microstrip lines that enter from opposite directions; since they excite the dipoles in opposite directions but are also 180° out of phase, the two dipoles are actually excited completely in phase. The radiation patterns of the two dipoles experience maximal constructive interference at the zenith and nadir.

We use an extended hemispherical lenslet to enhance the radiation pattern on one side and suppress it on the other. With a silicon lenslet, the slot dipoles are resonant with lengths around $\lambda_0/3$, where λ_0 is the free-space wavelength. In the limit that the lenslet is large compared with a wavelength, each pair of dipoles forms a beam within the silicon which is then focused further by the lenslet to produce a free-space radiation pattern [32, 106].

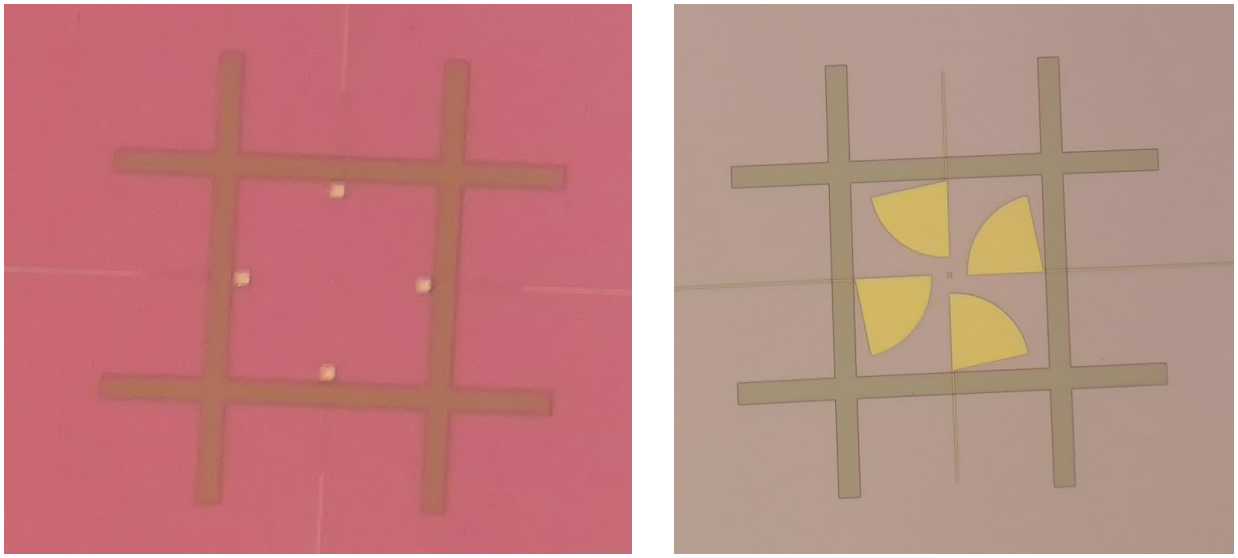
We have used two methods to couple microstrip lines to the slot dipoles. The goal is to set up a voltage across the midpoint of the slot. In Fig. 4.4(a), the microstrip lines are shorted to the ground plane on one side through a via in the microstrip dielectric, which is in this case

[2] If the length of the slot is set to exactly half of a free-space wavelength when there are no dielectrics around, the resonance will be about 25% too high in frequency. The length should be tuned to get the slot to look like a half-wavelength dipole at the desired frequency. As soon as dielectrics are included in the design, the length must be retuned to compensate for the effective dielectric constant.



(a) E-plane (b) H-plane

Figure 4.3: Simulated electric-field cross sections for a PB2 lenslet-coupled sinuous antenna at 160 GHz. The colors indicate the magnitude of the electric field and span five orders of magnitude from blue to red. The color scale has been chosen to saturate near the excitation, so the free-space waves can be seen more clearly. This configuration is similar to what is shown in Fig. 4.2(a) but with only two layers of anti-reflection (AR) coating. The lenslet and extension length are both composed of silicon ($\epsilon = 11.8$). The dielectric constants of the AR coating layers are 5 and 2, where the dielectric constants decrease monotonically from the lenslet to free space, and the thicknesses are chosen to be a quarter wavelength at 120 GHz. The E-plane is conventionally defined so that the excitation lies within it, although the sinuous antenna actually produces a far-field polarization angle that is perpendicular to the excitation. So the far-field polarization vector actually lies in the H-plane in this case. We see that the lenslet produces an approximately flat wavefront at the tip of the lenslet. It is in this sense that the lenslet "focuses" incident radiation onto the planar antenna. As the phase of the excitation is varied, the wavefronts evolve upwards.



(a) Via (b) Radial stub

Figure 4.4: Fabricated double slot dipoles.

silicon dioxide. The voltage between the strip layer and the ground plane is then converted to a voltage across the slot. This method is useful when the fabrication process allows for vias. When a via is undesirable, it is possible to use a quarter-wavelength open-circuited stub in its place. The idea is that the signal will travel to the end of the stub and back for a total path length of a half wavelength; this reflected signal, therefore, interferes destructively with the incoming signal and creates a node in the voltage between the strip layer and the ground plane layer on the stub side of the slot. A node in the voltage is electrically equivalent to a short circuit, which is exactly what the via accomplished. So a quarter-wavelength open-circuited stub acts just like a via but with a narrower bandwidth. There isn't much space in the interior of the double-slot dipole, so we use a *radial* stub as shown in Fig. 4.4(b) to fit within the available area. The radial stub is a typical bandwidth-increasing replacement for a quarter-wavelength stub.

4.2.1 Simulations

We can simulate a lenslet-coupled double slot dipole in HFSS. The simulation model is shown in Fig. 4.5. Due to the four-fold symmetry of the double slot dipole, it is possible in HFSS to reduce the simulation volume by a factor of four by imposing symmetry boundaries. This allows for a tremendous reduction in computation time. For this reason, the double slot dipole is attractive for investigating general effects related to lenslet-coupled planar antennas. For instance, the performance of the anti-reflection coatings and mutual coupling to neighboring lenslets can be characterized relatively quickly with the double slot dipole before simulating with, e.g., a sinuous antenna.

Some simulation results are shown in Fig. 4.6. The dipole lengths have been tuned to create a resonance, i.e., $\mathrm{Im}(Z) = 0$, at 150 GHz. The real part of the impedance is not

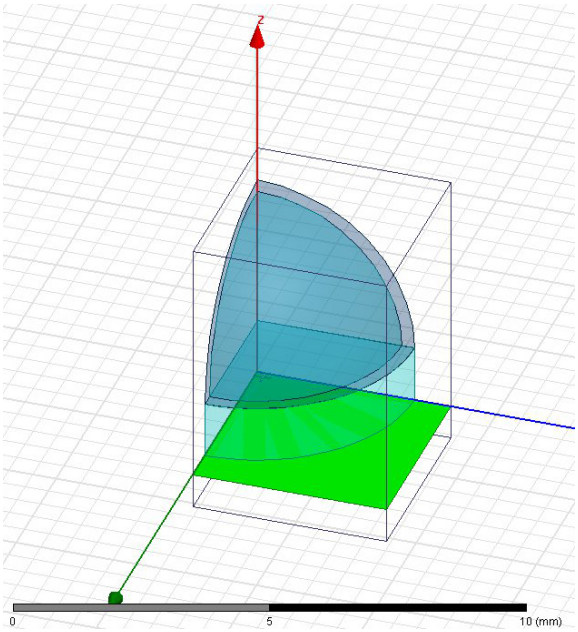


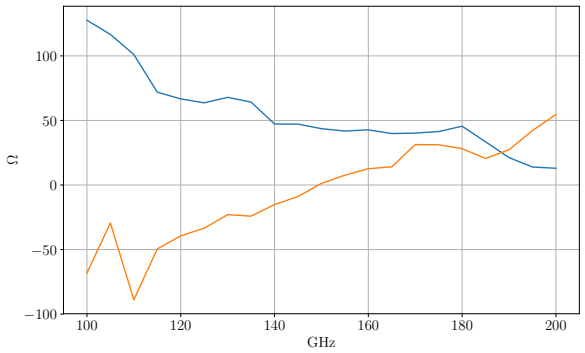
Figure 4.5: Simulation model of a lenslet-coupled double slot dipole in HFSS. We exploit the four-fold symmetry of the double slot dipole to reduce the computation time. We impose symmetry boundaries on two inner faces, and only one quarter of the total structure is explicitly simulated.

too steep around this resonance, so the coupling is better than 90% over approximately a 30% bandwidth. The dipoles also have higher harmonics at thrice the fundamental frequency. The radiation pattern at 150 GHz shows a round main beam with some minor rectangular distortions when the intensity drops to about 20% of its peak value. The polarization angle at zenith in the far field is shown in Fig. 4.7. The polarization angle is extremely stable over a broad range of frequencies. This is essentially enforced by symmetry.

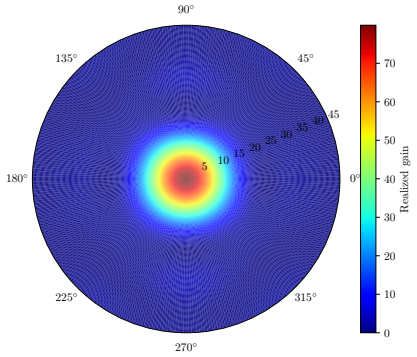
4.3 Sinuous antenna

The sinuous antenna is a type of *log-periodic antenna*, which we use to achieve sensitivity to a broad range of frequencies. The sinuous antenna can be implemented both in a two- and three-dimensional form, but we use the planar version in order to use the scheme described above which couples lenslets to planar slot antennas. A log-periodic antenna is fractal in the sense that it appears the same when scaled by a constant factor. Having no characteristic length scale, the performance is periodic in the logarithm of frequency. If we can achieve acceptable performance within one logarithmic period in frequency, then we have the same acceptable performance at all frequencies so long as we continue the geometry infinitely toward both smaller and larger length scales. In practice, of course, the geometry is truncated at both the large and small ends, and these truncations define the total bandwidth of the antenna.

The equation describing the sinuous antenna is giving in the polar coordinates r and ϕ



(a) Antenna impedance. The *orange* curve indicates the imaginary part of the impedance, and the *blue* curve indicates the real part. The null in $\mathrm{Im}(Z)$ is referred to as a resonance, since it is possible to perfectly couple a real-impedance transmission line at this frequency. The slope in $\mathrm{Re}(Z)$ is not too steep around this resonance, so it is possible to pull out about a 30% bandwidth of better-than-90% coupling using a transmission-line impedance of approximately 40-50Ω.



(b) Radiation pattern at the resonant frequency of 150 GHz. The radial coordinates in this plot indicate polar angle in degrees. The circumferential coordinates indicate azimuthal angle in degrees. The inner part of the beam is round. Some rectangular features are noticeable when the intensity has fallen off to approximately 20% of its peak value. There are also faint four-fold-symmetric features farther out.

Figure 4.6: Simulation results for a lenslet-coupled double slot dipole with a single layer of anti-reflection coating.

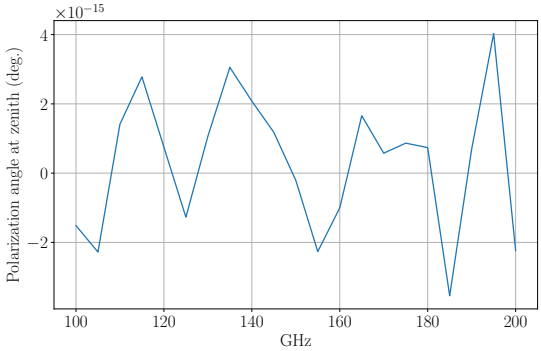


Figure 4.7: Polarization angle (relative to the x-axis) at zenith in the far field for the lenslet-coupled double slot dipole from Fig. 4.5. The polarization angle is remarkably stable. Note that the vertical axis in this plot is in units of $^\circ \times 10^{-15}$.

as

$$\phi = \alpha \sin\left[\pi \frac{\log(r/R_1)}{\log \tau}\right] \pm \delta, \tag{4.1}$$

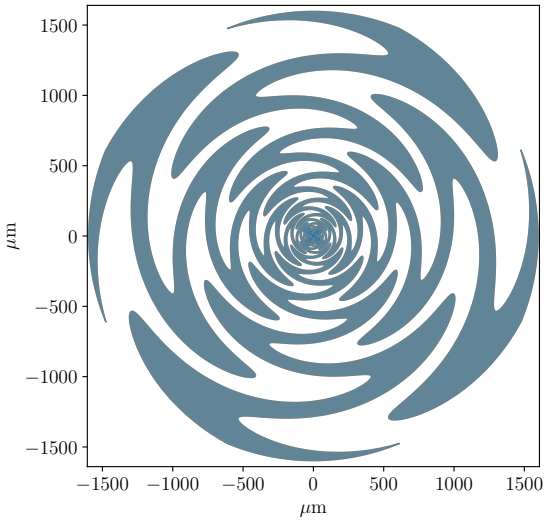
where α, R_1, τ and δ are parameters that can be tuned to alter the antenna performance. Notice that $r \to \tau^2 r$ leaves ϕ unchanged, since this will simply add a phase of 2π to the argument of the sine. It is in this sense that the sinuous antenna is logarithmically periodic, i.e., the antenna is symmetric under scalings by the factor τ^2. A scaling by a factor τ produces a mirror image, so τ is sometimes referred to as the logarithmic period. Equation 4.1 describes the sinuous profile. We implement the sinuous antenna as a slot in a ground plane, so we really need two profiles to describe each side of the slot. This is why we include a $\pm$ sign in Eq. 4.1. Each sign corresponds to a different side of the slot. When we connect these two profiles, we form a closed area which defines the sinuous slot. We implement the sinuous antenna with four such slots, so the antenna has a four-fold rotational symmetry. Some sketches are shown in Fig. 4.8 for finite sinuous antennas, i.e., for which the radius r varies over a finite range.

The sinuous antenna was invented by Raymond Duhamel in 1985 [29] and has been studied and adapted for use by our research group over the course of several years [31, 68, 99, 103]. It is being used by the PB-2 [100] and SPT-3G [15] collaborations and has been chosen for use in the LiteBIRD [64] and Simons Observatory projects.

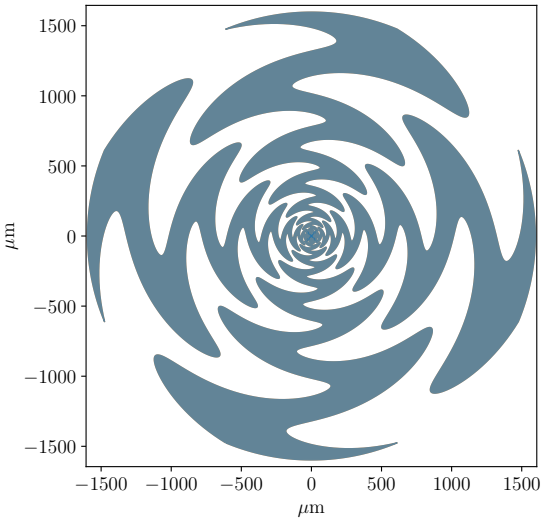
In our implementation, the sinuous antenna is fed with two out-of-phase microstrip lines that run along the ground plane into the center. Antennas that are fed by two out-of-phase transmission lines are sometimes said to *differentially fed*. This is shown in Fig. 4.9 for some prototype devices we fabricated. We set a minimum radius for the sinuous features, which sets the highest frequency to which the antenna is sensitive. For a pixel designed to be sensitive up to 200 GHz, we typically use $R_1 = 24$ μm. At smaller radii, we open the slot to allow the microstrip feeds to transfer power to the antenna. This shown in Fig. 4.9(b). Since the out-of-phase microstrip lines have a virtual ground at the center of the antenna, the microstrip lines for the two orthogonal polarizations are able to directly contact each other at this point.

The microstrip lines create an excitation in the slot, which propagates through the antenna. At a characteristic radius which depends on the wavelength, the sinuous antenna becomes resonant and radiates. Just as we did for the double-slot dipole, we couple the planar sinuous antenna to an extended hemispherical lenslet, usually composed of alumina or silicon. A sketch is shown in Fig. 4.2(a). The radiation is enhanced on the lenslet side and suppressed on the other.

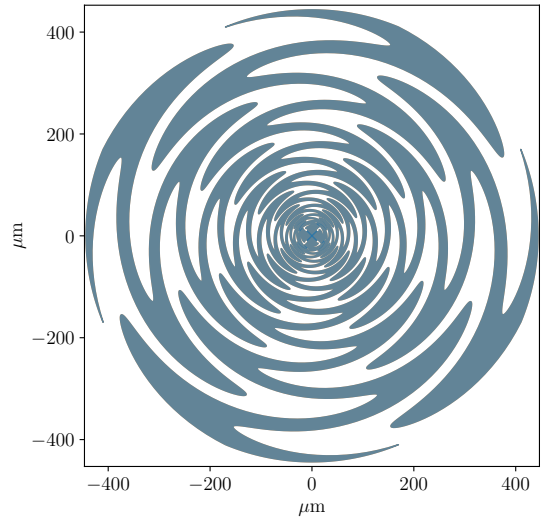
When the sinuous is embedded completely in a dielectric or simply in air, the self-complementarity creates a stable antenna impedance over wavelengths covered by the log-periodic sinuous profile [69, 98]. In air, the real part of the impedance is close to 200 Ω, and the imaginary part stays close to 0 Ω. Placing a dielectric lenslet on one side breaks the symmetry and creates ripples in the antenna impedance. We have tuned the sinuous geometric parameters to keep the amplitude of these oscillations at an acceptable level. For a silicon lenslet, the real part of the impedance oscillates around 100 Ω, so we feed the antenna differentially with two 50-Ω microstrip lines.



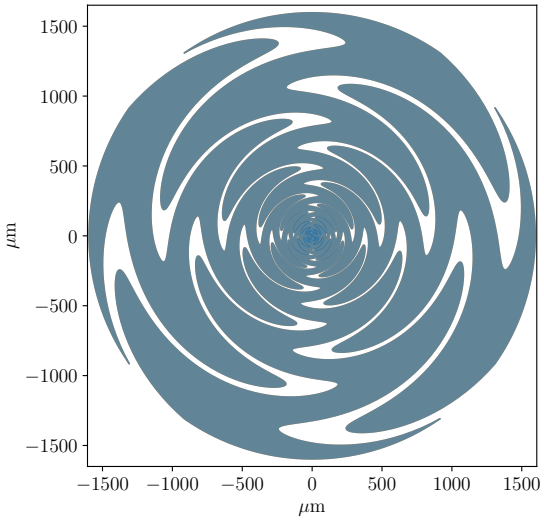
(a) Fiducial sinuous antenna, which is the design used in all of the devices reported on in this thesis.



(b) $\alpha = 30°$. This parameter controls the amplitude of the sinuous angular oscillations.



(c) $\tau = 1.2$. This parameter controls the logarithmic period, i.e., the radial frequency of the sinuous angular oscillations.



(d) $\delta = 35°$. This parameter controls how wide the slot is. For $\delta \neq 45°$, the antenna no longer achieves self-complementarity with the ground plane.

Figure 4.8: Sinuous-antenna variations. We perturb the antenna from a fiducial set of parameters: $\alpha = 45°$, $\tau = 1.3$, $\delta = 22.5°$. In each of these sketches, only one parameter has been altered from this fiducial set. The blue area is meant to indicate a slot in a ground plane, and the white area is the ground plane.

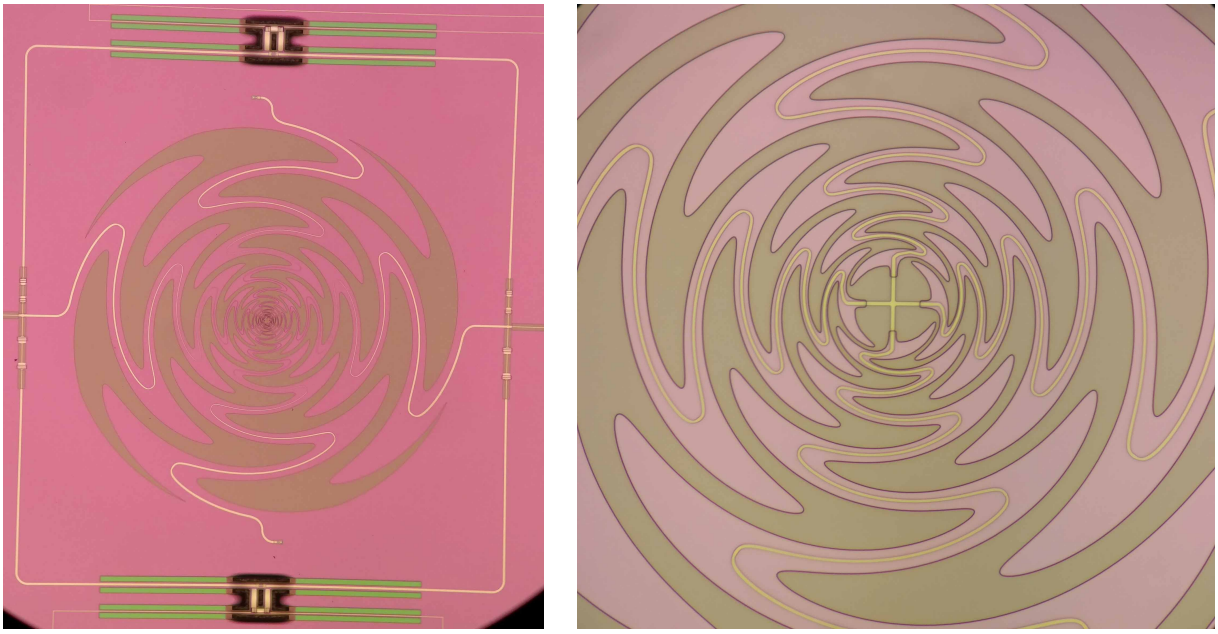


Figure 4.9: Fabricated sinuous-antenna microstrip feeds.

4.3.1 Simulations

We now discuss simulation results from the model presented in Fig. 4.3.

In free space, i.e., without a lenslet, the sinuous antenna has a near-constant pure real impedance over a bandwidth corresponding to the minimum and maximum radii of the sinuous profile. The addition of the dielectric lenslet creates ripples in the impedance. This is shown in Fig. 4.10. The breaks and multiple values come from the manner in which this antenna was simulated. Since the bandwidth is so large, the mesh of finite elements that is necessary at higher frequencies is overkill for lower frequencies and increases simulation time unnecessarily. To speed up the simulation, we generate a new mesh in each frequency range. We select a frequency at which to generate the mesh; then we use that mesh only down to 80% of the mesh frequency. The convergence criterion is that the S-parameters are stable to within 2%. The erratic fringing in these simulation results should, therefore, be viewed with a certain amount of skepticism. It is also not too worrisome that the results for different mesh frequencies do not agree perfectly. The real part of the impedance oscillates about 100 Ω, and the imaginary part oscillates about 0 Ω. We can, therefore, tune the transmission-line feed to 100 Ω to achieve better than 90% coupling over the bandwidth corresponding to these stable oscillations. At very low frequencies, the sinuous and lenslet are not large enough to support radiation, so the impedance becomes erratic. It is difficult to simulate to very high frequencies, because the required mesh of finite elements uses tremendously more memory. At a high enough frequency, however, which is probably near 300 GHz, the inner non-sinuous structure of the antenna will become resolved, which will destroy the stable impedance oscillations.

Some radiations patterns are shown in Fig. 4.11. We have chosen to present three rela-

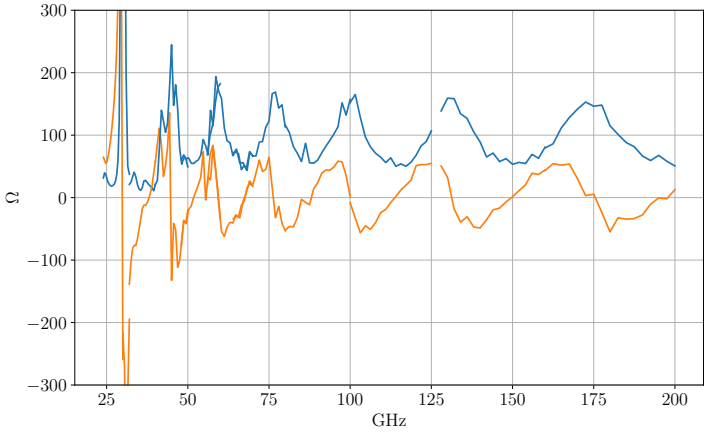
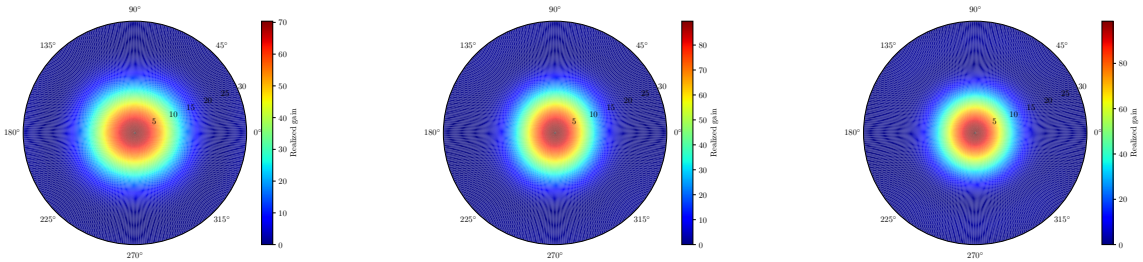


Figure 4.10: Antenna impedance for a PB2 lenslet-coupled sinuous antenna (the same described in Fig. 4.3). The *orange* curve indicates the imaginary part of the impedance, and the *blue* curve indicates the real part. See the text for an explanation of the breaks and multiple values. Notice the stable logarithmically periodic oscillations in both the real and imaginary components. At very low frequencies, the oscillations become unstable. It is computationally impractical to simulate to higher frequencies, but scaling from non-lenslet simulations and non-sinuous simulations leads us to believe that these oscillations will continue stably up to at least 300 GHz.



(a) 140 GHz (b) 150 GHz (c) 160 GHz

Figure 4.11: Some radiation patterns from a lenslet-coupled sinuous antenna showing the change in ellipticity as a function of frequency. The 140-GHz beam is slightly elliptical along $\phi = 0°$. The 150-GHz beam is elliptical along $\phi \approx 80°$. The 160-GHz is not noticeably elliptical at all. We call this property an ellipticity "wobble". At the same time, the beams can be seen to decrease in overall size as frequency increases.

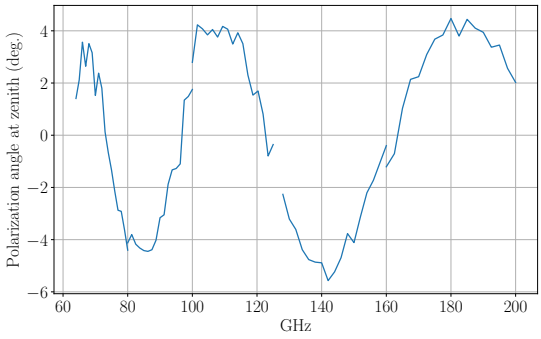


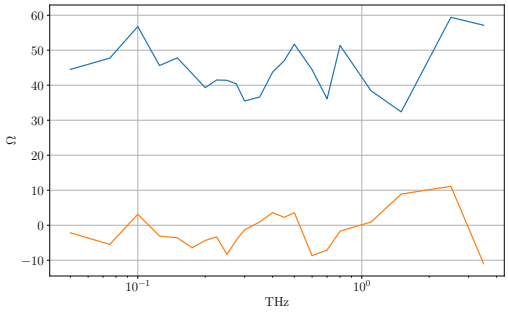
Figure 4.12: Polarization angle at zenith for a PB-2 lenslet-coupled sinuous antenna. There is some noise, which is at least partially due to the numerical precision of the simulation, but there is a clear log-periodic oscillation, which we call a *polarization wobble*. The amplitude is approximately 5°, and the oscillation repeats after a frequency ratio of $\tau^2 \approx 1.7$.

tively closely spaced frequencies in order to highlight the "wobble" of the ellipticity of the sinuous-antenna radiation pattern. The beam from the sinuous antenna has an ellipticity that varies with frequency both in amplitude and orientation. Over a 30% bandwidth, which is typical for our devices, the effect tends to average away and produce a round beam. For high-precision applications, however, this effect should be considered.

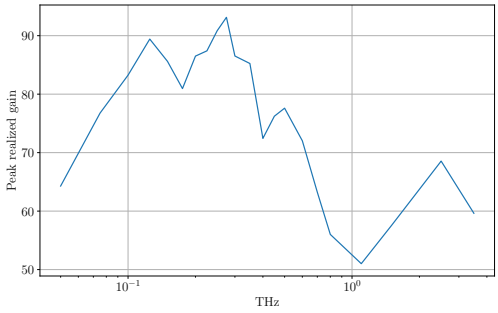
The polarization angle at zenith in the far field is shown in Fig. 4.12. Ignoring the noise in the simulation results, we see a clear log-periodic oscillation with an amplitude of approximately 5°. We call this oscillation a *polarization wobble*. We expect the properties of the sinuous antenna to persist when the frequency is scaled by $\tau^2 \approx 1.7$, and this appears to be approximately true for the polarization wobble. The lenslet diameter and AR-coating thickness set length scales in the antenna, so we expect the log-periodic behavior to become distorted at very small and very large frequencies. Since the sinuous antenna has a handedness, which comes from the choice of direction for the azimuthal swing at a given radius, the mirror-imaged version will have a polarization wobble that is equal and opposite. By designing a focal plane consisting of both variants, the effects of the polarization wobble can be mitigated statistically [98].

4.4 Electrically small lenslet-coupled antennas

In Sec. 4.1, we treated the extended hemispherical lenslet using geometric optics. In the limit that the lenslet diameter is comparable to a wavelength, however, this approximation breaks down. In this section, we discuss some effects that arise to the electrically small diameter of the lenslets used in our focal planes.



(a) Antenna impedance as a function of the AR-coating $\lambda/4$ frequency (*not* the simulation frequency). (*Blue*) Real part. (*Orange*) Imaginary part.



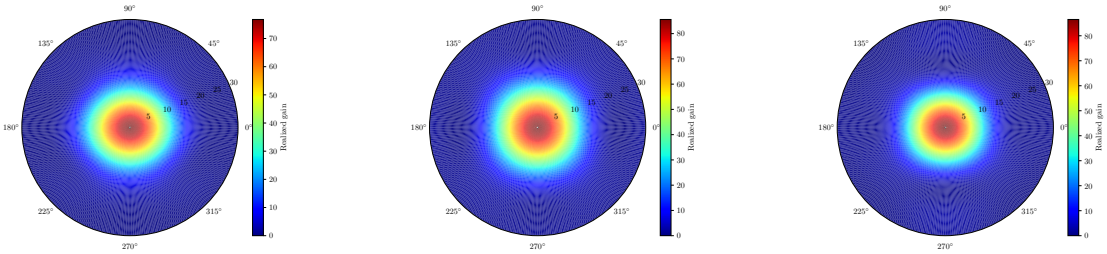
(b) Peak realized gain, i.e., realized gain at the zenith, as a function of the AR-coating $\lambda/4$ frequency.

Figure 4.13: Simulation results for a lenslet-coupled double slot dipole tuned to a resonant frequency of 150 GHz, where the $\lambda/4$ frequency of the AR coating is varied but the pixel diameter is held constant. The *simulation* frequency is 150 GHz. The lenslet is composed of silicon with $\epsilon = 11.8$, and the AR coating has $\epsilon = 3.4$. The pixel diameter is maintained at 6.789 mm.

4.4.1 Anti-reflection coating

When the radius of curvature of the lenslet is larger than but within an $\mathcal{O}(1)$ factor of a wavelength, the anti-reflection (AR) coating at the lenslet surface does not perform as would be expected based on an analysis of a flat interface with an incident plane wave. It is illegitimate in this case to approximate the lenslet-air interface as locally flat. Fortunately, the approximation breaks down in a way that increases the tolerance on the AR coating.

As an example of this phenomenon, we simulated a lenslet-coupled double slot dipole tuned to a resonant frequency 150 GHz. We keep the diameter of the pixel, i.e., the radius of the lenslet together with the AR coating, constant and vary the thickness of the AR coating. In this example, we take the AR coating to have a dielectric constant that is the geometric mean of the dielectric constant of the lenslet and the dielectric constant of free space, which is what would be optimal in the case of a flat interface with an incident plane wave. Normally, the thickness is chosen to be a quarter wavelength at the central frequency, which is in this case 150 GHz. We parameterize the thickness of the AR coating by the frequency for which the thickness is a quarter wavelength. The antenna impedance at 150 GHz as a function of this $\lambda/4$ frequency is shown in Fig. 4.13(a), where we see that it is remarkably stable over nearly two orders of magnitude. Coupling to a microstrip line of characteristic impedance 45 Ω, the radiation efficiency is greater than 90% over this entire range. The peak realized gain, i.e., the realized gain at the zenith, is shown in Fig. 4.13(b), where we see a fairly broad optimum and certainly no obvious maximum at 150 GHz. Some beam patterns are shown in Fig. 4.14, where we that a small ellipticity is introduced as the $\lambda/4$ frequency deviates substantially from 150 GHz. Otherwise, however, the beams are well-behaved. The beam widths become slightly smaller as the $\lambda/4$ frequency is increased. The AR-coating



(a) 75 GHz (b) 150 GHz (c) 300 GHz

Figure 4.14: Radiation patterns for the pixel described in Fig. 4.13. The beam patterns are shown for various $\lambda/4$ frequencies. The *simulation* frequency is in all cases 150 GHz.

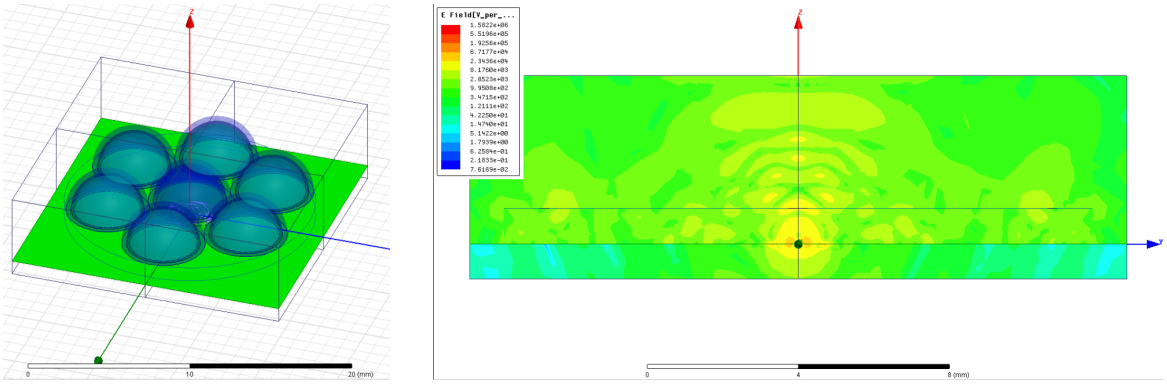
thickness becomes smaller, so the silicon lenslet becomes larger. The focusing of the beam is dominated by the materials with high dielectric constants like the lenslet. For a constant pixel diameter, then, a larger silicon lenslet produces a larger effective illumination. It is advantageous, therefore, to use thinner and fewer layers of AR coating.

4.4.2 Mutual coupling

We use a hexagonal close packing to fit as many pixels on the focal plane as possible. This is shown, for example, in Fig. 4.15(a). Each lenslet has six neighbors. For electrically small lenslets, the presence of the neighboring lenslets affects the performance of each pixel. We sometimes call this mutual coupling. Some antenna arrays depend on mutual coupling, e.g., a Vivaldi array or a current-sheet array, but in our focal planes it is a nuisance. There are really two effects that seem to be occurring simultaneously. The small lenslet allows the near-field excitation to couple to the extension-length substrate. This can propagate horizontally as a surface mode in the silicon. At the same time, some power couples into the neighboring lenslets. This is demonstrated in simulation in Fig. 4.15(b). These unintended couplings alter the effective focal-plane illumination and create pathological beam patterns such as those shown in Fig. 4.16. Notice that the main lobe disappears in some cases and that the beam pattern changes quite rapidly as a function of frequency. Averaging the beams over a 20-25% bandwidth, however, smooths much of this erratic behavior. If pixel diameters smaller than two free-space wavelengths are desired, then it will be necessary to measure the beams as a function of frequency to see if these simulated pathologies are real.

4.5 Bandpass filters

Once the incident radiation is transferred to microstrip lines, we can filter it before it reaches the bolometer. For this purpose, we design on-wafer bandpass filters. The passbands are chosen to avoid atmospheric absorption lines, to aid in constraining galactic foregrounds and to increase experimental sensitivity to the CMB.



(a) A seven-lenslet array with a pixel spacing of 5.3 mm. Only the central lenslet has a planar antenna underneath it.

(b) Cross section of the electric field along the H-plane. The lenslets have been hidden in this image. We see the beam formation in the center, but we also see some activity to the right and left.

Figure 4.15: Simulation of an array of lenslet-coupled antennas to characterize the effects of mutual coupling. To speed up the simulation, we use a double slot dipole instead of a sinuous antenna. The neighboring lenslets tend to distort the near-field pattern, which can manifest itself as a beam distortion in the far field.

4.5.1 Insertion-loss method

Our filter designs rely on the *insertion-loss method*. Essentially, this means that we specify a transfer function and then solve for the circuit-component values that achieve that transfer function. We are transferring power from port 1 to port 2, where both ports can be taken to be terminated in microstrip lines that travel either to the antenna or to the bolometer. The *power-loss ratio* is defined as

$$P_{\mathrm{LR}} \equiv \frac{1}{|S_{12}|^2}, \tag{4.2}$$

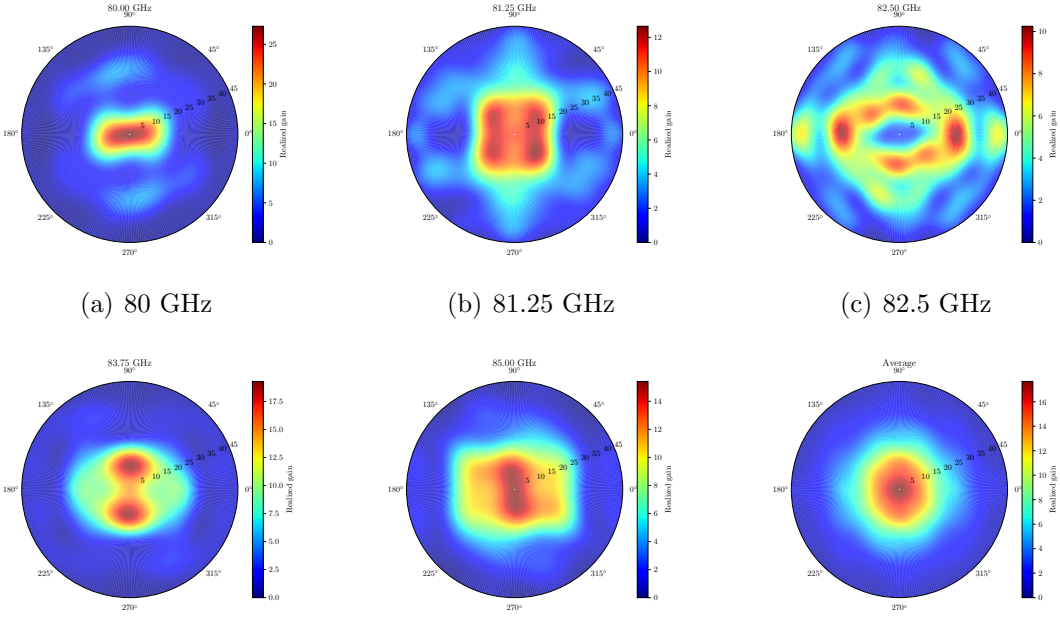
and the *insertion loss* is simply the decibel version of this quantity, i.e., $\mathrm{IL} \equiv 10 \log_{10} P_{\mathrm{LR}}$. The power-loss ratio ranges from 1 to ∞, where $P_{\mathrm{LR}} = 1$ means we have perfect transmission and $P_{\mathrm{LR}} \gg 1$ means we have negligible transmission. We assume in this discussion that the circuits are low loss, so the main possibilities are reflection and transmission only.

We begin by studying so-called low-pass prototypes. This means that we will initially design a low-pass filter, but there are simple transformations that can turn this low-pass filter into a high-pass, bandpass or bandstop filter.

We specify at the outset what we want the insertion loss or, equivalently, the power-loss ratio to be. Set $P_{\mathrm{LR}} = f(\omega/\omega_c)$ for some function f and some characteristic angular frequency ω_c. There are several common choices for the function f:

- A Butterworth filter, which is also often called a maximally flat filter, is defined by

$$P_{\mathrm{LR}} = 1 + k^2 \left(\frac{\omega}{\omega_c}\right)^{2N}, \tag{4.3}$$



(a) 80 GHz (b) 81.25 GHz (c) 82.5 GHz

(d) 83.75 GHz (e) 85 GHz (f) Average from 80 to 100 GHz

Figure 4.16: Beams from a lenslet-coupled sinuous antenna with one layer of anti-reflection coating spaced at 5.3 mm from its neighbors. The simulation includes the neighboring lenslets but not the neighboring sinuous antennas. In other simulations, we have shown that the antennas do not couple to each other at an appreciable level, i.e., $|S(1,2)| \ll 1$. We simulate in relatively small frequency increments to show the erratic behavior of the beams. The average is noticeably elliptical but not pathological like the single-frequency beams.

where the parameter k controls attenuation when $\omega = \omega_c$ and the parameter N controls the steepness of the roll-off. A larger N gives a steeper roll-off, which is usually desirable, but requires, as we will see, a more complex circuit. This filter is called maximally flat, because there are no ripples: the power-loss ratio increases monotonically as ω increases.

- A Chebyshev filter, which is also often called an equal-ripple filter, is defined by

$$P_{\mathrm{LR}} = 1 + \left[kT_N(\omega/\omega_c)\right]^2, \tag{4.4}$$

where $T_N(x)$ is the Chebyshev polynomial of the first kind of order N.[3] The parameter k and N play the same qualitative roles for the Chebyshev filter as for the Butterworth. This filter is called equal ripple, because the Chebyshev polynomials create ripples of equal amplitude in the passband. In exchange for these ripples, the Chebyshev filter has a steeper roll-off than the Butterworth. The parameter k controls both the ripple amplitude and the attenuation at $\omega = \omega_c$. A higher-order filter has more ripples and a stepper roll-off. Some Chebyshev transfer functions are plotted in Fig. 4.17 for a 0.5-dB ripple, i.e., the transmission is approximately 90% at minimum in the passband.

- An elliptic filter is defined by

$$P_{\mathrm{LR}} = 1 + [kR_N(\xi, \omega/\omega_c)]^2, \tag{4.6}$$

where $R_N(\xi, x)$ is the elliptic rational function of order N and k and ξ are free parameters. This type of filter has ripples in both the passband and the stopband but allows for a steeper roll-off than the Chebyshev filter. The parameters k and ξ control the trade-off between ripple amplitudes and roll-off steepness. This type of filter requires a qualitatively different circuit model than the Chebyshev and Butterworth filters and is, in general, more difficult to implement in microstrip format.

The order N of these filters is also referred to as the number of poles and is proportional to the number of reactive circuit components in the low-pass prototype. A good choice for our purposes is a 3-pole Chebyshev filter. This gives a relatively steep roll-off to help us avoid atmospheric lines but is also not too complicated. The ripples in the passband translate to an overall degradation of the optical efficiency.

Now we attempt to realize these transfer functions. With the low-pass prototype shown in Fig. 4.18(a), we will be able to match the transfer functions for 3-pole Butterworth and Chebyshev filters. Leaving the circuit-component values as free parameters, we can calculate the input impedance of the low-pass prototype. Given an incoming characteristic impedance Z_0, we can then extract a reflection coefficient S_{11}. Since this is a 2-port network in which loss has been assumed to be negligible, we immediately have $|S_{12}|^2 = 1 - |S_{11}|^2$, and

[3] For low N, the Chebyshev polynomials of the first kind can be generated quickly from

$$T_0(x) = 1, \qquad T_1(x) = x, \qquad T_{N+1}(x) = 2xT_N(x) - T_{N-1}(x). \tag{4.5}$$

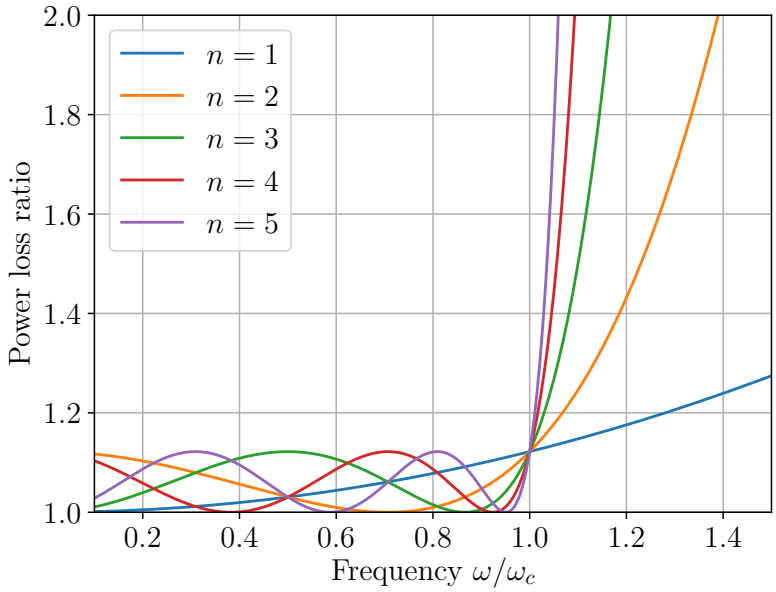
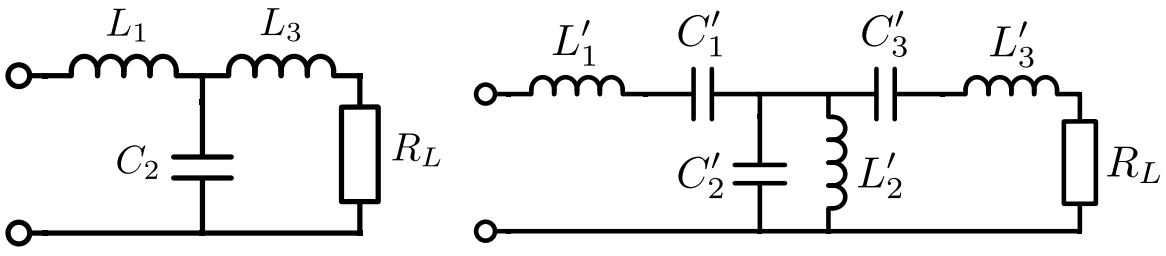


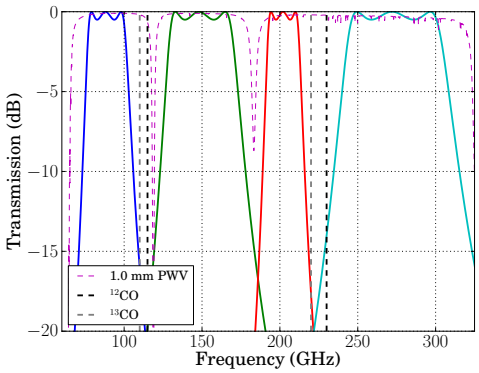
Figure 4.17: Power loss ratio for Chebyshev (equal-ripple) low-pass filters, where n is the order of the Chebyshev polynomial. Higher orders have more ripples and steeper roll-offs.



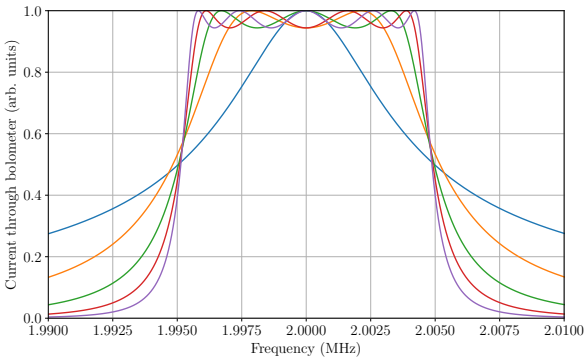
(a) Low-pass prototype

(b) Bandpass prototype

Figure 4.18: Filter circuit schematics.



(a) 3-pole passbands chosen to avoid atmospheric and galactic CO lines.



(b) Passbands with a varying number of poles but approximately the same 3-dB points. These filters were designed for a frequency-multiplexing readout system operating between 1 and 5 MHz.

Figure 4.19: Passbands for Chebyshev bandpass filters.

from $|S_{12}|^2$ we have $P_{\rm LR}$. At this point we have $P_{\rm LR}$ in terms of L_i and C_i. It is convenient to work in units for which $Z_0 = \omega_c = 1$. Then L_i and C_i are actually dimensionless. We have the power-loss ratio in terms of L_i and C_i, but we also have a desired, e.g., Butterworth or Chebyshev, power-loss ratio. So we can set these two expressions equal to each other and solve for L_i, C_i and R_L. Define $g_1 = L_1$, $g_2 = C_2$, $g_3 = L_3$, $g_4 = R_L$. The load impedance R_L is not necessarily equal to Z_0, although it happens to be true for the 3-pole Chebyshev filters that we typically design. The solutions for g_i are the same for the Π-network low-pass filter, which is the dual of the T-network we are using here. For the Π filter we have $C_1 = g_1$, $L_2 = g_2$, etc. These coefficients have been tabulated and are commonly available for standard ripple levels, e.g., 0.5 dB and 3 dB.

At this point, we will transform the low-pass prototype into a bandpass filter. At the level of the idealized transfer function, the transformation

$$\omega \rightarrow \frac{1}{\Delta}\left(\omega - \frac{1}{\omega}\right) \tag{4.7}$$

changes the insertion loss to that of a bandpass filter of central frequency 1 and fractional bandwidth Δ. Some examples of Chebyshev passbands are shown in Fig. 4.19. At the circuit level, this changes, e.g., the impedance of an inductor to

$$Z_L = j\omega L \rightarrow j\omega\frac{L}{\Delta} + \frac{L}{j\Delta\omega}, \tag{4.8}$$

which just looks like an inductor in series with a capacitor. We can make the same kind of transformation for the capacitors, and we will find that the new impedance looks like that of a capacitor in parallel with an inductor. So we can draw a bandpass prototype, which looks like the circuit shown in Fig. 4.18(b). The component values are

$$L_1' = g_1/\Delta, \qquad C_1' = \Delta/g_1, \qquad L_2' = \Delta/g_2, \qquad C_2' = g_2/\Delta, \qquad \text{etc.} \tag{4.9}$$

At this point, we can restore units to get

$$L_1' = \frac{g_1 Z_0}{\Delta\omega_0}, \qquad C_1' = \frac{\Delta}{g_1 Z_0 \omega_0}, \qquad L_2' = \frac{\Delta Z_0}{g_2\omega_0}, \qquad C_2' = \frac{g_2}{\Delta Z_0 \omega_0}, \qquad \text{etc.,} \tag{4.10}$$

where

$$\Delta = \frac{\omega_2 - \omega_1}{\omega_0}, \qquad\qquad \omega_0 = \sqrt{\omega_1\omega_2}. \tag{4.11}$$

4.5.2 Circuit transformations

To realize a lumped-element circuit, i.e., one in which there are discrete components corresponding to circuit elements, we will need to design and fabricate the inductors and capacitors shown in, e.g., Fig. 4.18(b). The component values for this circuit are, however, somewhat awkward to fabricate. As described below, we use coplanar-waveguide sections as inductors, for which a comfortable range for our fabrication process is 10-100 pH. Meanwhile, we use wide microstrip sections as shunt capacitors, for which a comfortable range is 10-100 fF. For $\Delta = 25\%$ and $\nu_0 = 90$ GHz, the design values for the shunt components are $L_2' = 0.40$ pH and $C_2' = 780$ fF. This requires a tiny inductor and an enormous capacitor.

For more comfortable component values, it is better to invert the shunt LC to series. This can be accomplished with the impedance inverters shown in Fig. 4.20(a). There are several methods for constructing an impedance inverter. We choose to use a capacitor T-network, where the series capacitors have effectively negative capacitance. This type of impedance inverter should, therefore, only be chosen when it is possible to combine these negative capacitors with others to produce physically realizable components. An impedance inverter is characterized by a parameter K such that the input impedance of the inverter followed by a load impedance Z_L is $Z_{\text{in}} = K^2/Z_L$. If we place inverters on either side of a shunt impedance, we can flip the shunt components to series subject to the condition $K = \sqrt{Z_{\text{shunt}} Z_{\text{series}}}$. We use this technique to flip the shunt LC in Fig. 4.18(b) to a series LC as in Fig. 4.20(a). If we choose

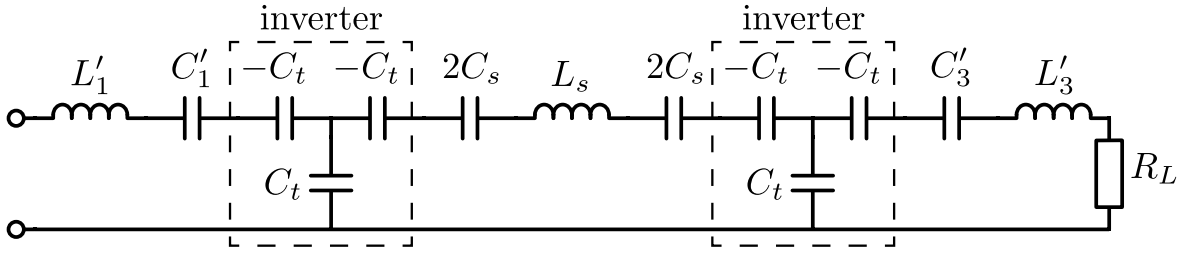
$$L_s = L_1', \qquad\qquad C_s = \frac{1}{L_s\omega_0^2}, \tag{4.12}$$

then we must have $K = \sqrt{L_s/C_2'}$. The condition on the capacitances in the inverter T-network is $K = 1/(\omega_0 C_t)$. To be clear, this inverter is narrowband.

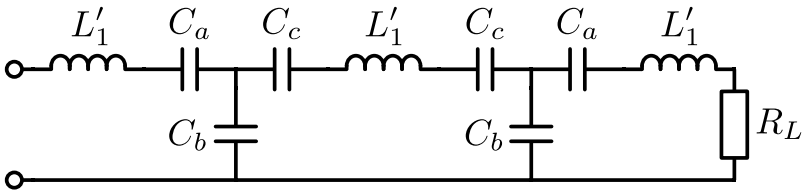
We can now combine capacitances to arrive at the circuit shown in Fig. 4.20(b). The new capacitances are

$$C_a = \left(\frac{1}{C_1'} - \frac{1}{C_t}\right)^{-1}, \qquad C_b = C_t, \qquad C_c = \left(\frac{1}{2C_s} - \frac{1}{C_t}\right)^{-1}. \tag{4.13}$$

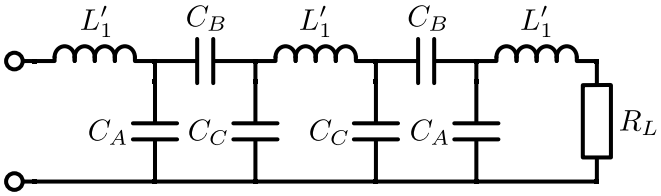
From these expressions, it is clear that we require $C_t > C_1'$ and $C_t > 2C_s$. At this stage, we find that the series capacitances are bit too large to be easily fabricated. For example, a 3-pole Chebyshev filter with $\nu_0 = 150$ GHz, $\Delta = 25\%$ and $Z_0 = 10\Omega$ requires $C_c = 50$ fF, whereas it would be more comfortable to have series capacitances in the range 1-10 fF.



(a) Invert the middle components from shunt to series and split the series capacitor C_s to keep the circuit symmetric.

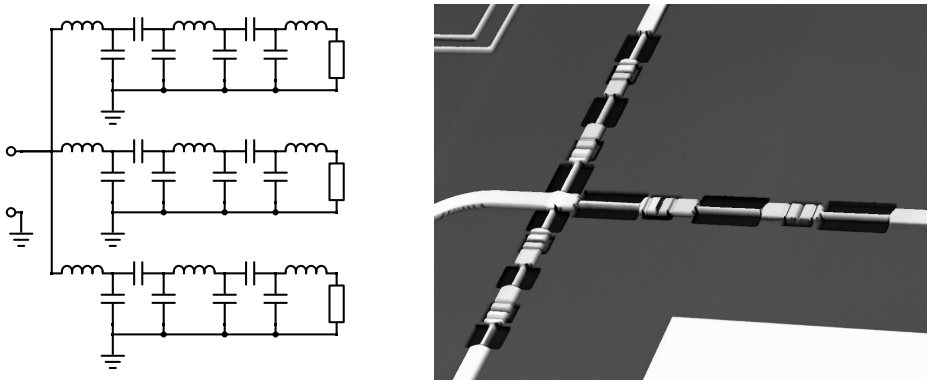


(b) Combine capacitors to get rid of the negative capacitances.



(c) Convert the capacitor T-networks to Π-networks (Y-Δ transformation). This is the final form.

Figure 4.20: Circuit transformations for a 3-pole bandpass filter (Fig. 4.18(b)).



(a) Circuit schematic for triplexing bandpass filters.

(b) Fabricated 220/280/350 triplexer. (Photo credit: Ben Westbrook.)

Figure 4.21: Triplexer layout.

It is straightforward to implement a wye-delta transform to replace the capacitor T-network with a Π-network as shown in Fig. 4.20(c). The new capacitances are

$$C_A = \frac{1}{aC_c}, \qquad C_B = \frac{1}{aC_b}, \qquad C_C = \frac{1}{aC_a}, \tag{4.14}$$

where

$$a = (C_aC_b)^{-1} + (C_aC_c)^{-1} + (C_bC_c)^{-1}. \tag{4.15}$$

The circuit diagram shown in Fig. 4.20(c) is the form we use for fabrication.

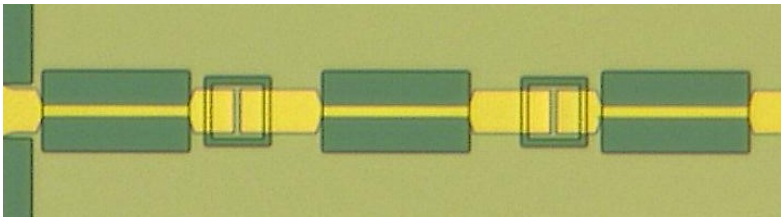
4.5.3 Multiplexing filter

For a multichroic pixel, we require multiple bandpass filters stemming from the same microstrip line. The idea is that the antenna receives radiation and transfers the power to a microstrip line. The power is then split into multiple frequency bands with bandpass filters. Each filter sends power to a separate bolometer.

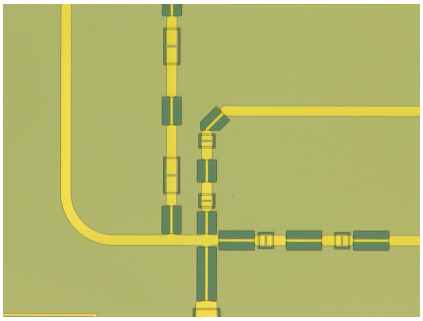
For three or fewer bands, we can connect the bandpass filters to a common node as shown in Fig. 4.21. So long as the passbands do not overlap too much, any given frequency will see at least two of these bandpass filters as open circuits.[4] We call this device a *multiplexing bandpass filter*. When we split the power into two bands, we call the device a *diplexer*. For three bands, we call the device a *triplexer*. For four bands, the device is called either a *quadplexer* (or sometimes a *quadruplexer*) or a *tetraplexer*.[5]

[4]There is a somewhat subtle point here, which often goes unmentioned. What is really happening is that each frequency attempts to enter all of the bandpass filters but is reflected from all but possibly one of them. The reflected wave, however, could be phase shifted and interfere with the component traveling through the resonant filter. For the bandpass filters that we design, this is fortunately not the case, i.e., the reflected wave is not phase shifted. It is in this sense that the non-resonant bandpass filters act like open circuits. For low-pass filters, however, this is actually *not* the case, and one must be careful with the reflections at the out-of-band frequencies.

[5]I personally prefer the term *tetraplexer*, since this keeps all of the numerical prefixes Greek, i.e., di-, tri-, tetra-, penta-, hexa-, etc. as opposed to the Latin du-, tri-, quadru-, quinque-, sexa-, etc.



(a) 220-GHz lumped-element bandpass filter. The corresponding circuit diagram is shown in Fig. 4.20(c).



(b) 90/150/220/280-GHz tetraplexer. The microstrip line at left comes from a broadband sinuous antenna.

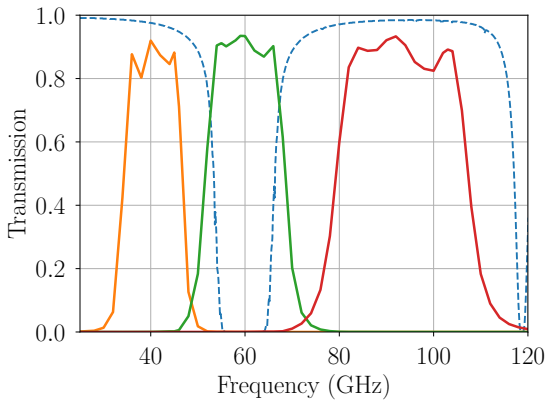
Figure 4.22: Fabricated tetraplexing bandpass filters. (Photo credits: Ben Westbrook.)

For more than four bands, it is impractical to connect all of the bandpass filters to a common node. In this case, we offset some of the filters and tune the phase delay to achieve acceptable performance. An example of a tetraplexer sensitive to 90, 150, 220 and 280 GHz is shown in Fig. 4.22, where the 90-GHz filter is offset from what would have been the common node. We chose to offset the lowest-frequency filter, since the extra distance is a smaller fraction of a wavelength than for the other bands.

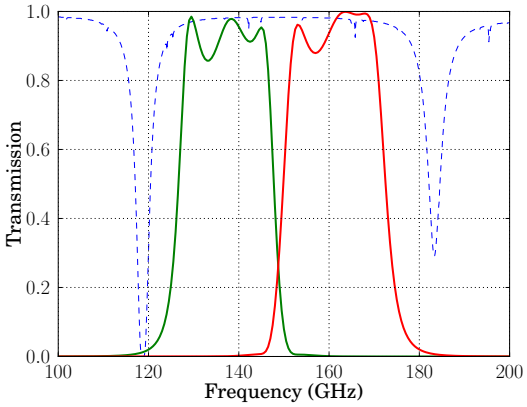
4.5.4 Microstrip realization

There are two main approaches to realizing a bandpass filter in a manner that is compatible with microstrip fabrication. The approach we take here is sometimes called the *lumped-element* method. In this paradigm, each circuit element shown in, e.g., Fig. 4.20(c), appears as a discrete component in the fabricated filter. For example, we use short coplanar-waveguide (CPW) sections to act as little inductors and short wide microstrip sections to act as shunt capacitors. These lumped-element filters are to be contrasted with so-called *distributed* or *stub* filters, which rely on quarter-wavelength short-circuited stubs separated by quarter-wavelength impedance inverters. The stub design relies on the fact that a quarter-wavelength short-circuited segment has an input impedance similar to that of a shunted parallel LC resonator. The best low-pass prototype for these types of filters is the Π-network version rather than the T-network that we opted for above. In the bandpass version, the shunted LC resonators can be redrawn as short-circuited quarter-wavelength stubs. The series LC resonator can be flipped down to shunt with quarter-wavelength impedance inverters that act in a similar way to the capacitor T-networks we used above to flip from shunt to series. Distributed filters are simpler to design, since one can derive equations for the required impedances of the microstrip stubs. The stub filters tend to occupy more space than their lumped-element counterparts, and they are more susceptible to variations in the kinetic inductance of the superconducting microstrip metal. For these reasons, we prefer lumped-element filters.

We use CPW segments as inductors. Since a CPW, like any transmission line, has some



(a) Low-frequency triplexer. The 60-GHz band is sitting right on top of a very strong atmospheric line. This triplexer would only be useful in a space-based application, for which the atmosphere is no longer a foreground.



(b) Diplexer splitting the 150-GHz atmospheric window.

Figure 4.23: Simulated performance of lumped-element bandpass filters. The simulations were run with Sonnet Software.

shunt capacitance which must be tuned out, the filter design process is an iterative one. With Sonnet Software, we draw a geometry, simulate, adjust the geometry, resimulate, etc. It is helpful to use a program like Keysight's Advanced Design System (ADS) alongside Sonnet. In ADS we can add parasitic impedances to see what problems need to be addressed in the microstrip version drawn in Sonnet. Since ADS simulates virtually instantaneously, this is a very cheap way to diagnose design problems.

The shunt capacitors are simply short microstrip segments. Often we widen these far beyond the width of the incoming microstrip in order to increase the capacitance and decrease the inductance. Just as we cannot completely rid the CPW of some shunt capacitance, we cannot rid a microstrip segment of some series inductance. This is another source of parasitic impedances that must be tuned out iteratively in the design and simulation cycle.

The series capacitors are formed by creating an island on the ground-plane layer and running microstrip lines across the gap to form parallel-plate capacitors with the island. In, e.g., Fig. 4.22, the breaks in the microstrip line indicate locations where we have these series capacitors. By creating an island on the ground-plane layer, we have a metal area that is not electrically connected to ground and can, therefore, become part of the series components of the bandpass filter. We put two such capacitors in series with each other, so the series capacitor is symmetric.

Some results from Sonnet simulations of lumped-element bandpass filters are shown in Fig. 4.23. The Chebyshev ripple structure is clearly visible but with some distortions. It is not necessary to replicate the Chebyshev transfer function exactly. All we really need are bands that are strongly attenuated (< -10 dB) at the atmospheric lines and receiving a majority of the photons within the atmospheric windows.

Narrow passbands

The primary CMB observations bands tend to be centered on 90 and 150 GHz. The corresponding atmospheric windows have fractional bandwidths of approximately 30%. Often, bands placed at other frequencies have similar fractional bandwidths, even if the atmospheric windows don't force that to be true. Now, if we keep the fractional bandwidth constant, we can scale the circuit-component values inversely with frequency to shift the passband, i.e., $L\omega_0$ and $C\omega_0$ are constant for constant Z_0 and constant Δ. This is useful to remember when designing filters for multiple frequency bands. As soon as one filter is designed, it is relatively trivial to produce the others, so long as the fractional bandwidths and the incoming transmission-line impedances are similar.

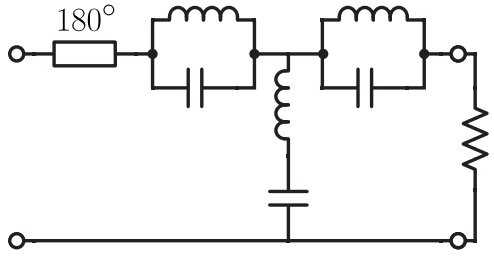
But what happens when we reduce the fractional bandwidth? We see from Eq. 4.10 that the required inductance scales *inversely* with the fractional bandwidth, i.e., $L \propto 1/\Delta$. This is not so surprising when we remember that a series LCR circuit (which is a 1-pole bandpass filter) satisfies the well-known relation $L\omega_0/R = Q = 1/\Delta$, where Q is the *quality factor* of the resonance. So a reduced fractional bandwidth translates to an increased inductance requirement. But our lumped-element inductors are just CPW segments, where the inductance is approximately proportional to the length. If the required inductance is too large, though, then these segments will no longer appear electrically small, which is another way of saying that the lumped-element approximation is no longer valid. In this limit, it becomes impractical to realize anything resembling a Chebyshev transfer function.

The best way to deal with this problem is to note that the required inductance scales with the incoming transmission-line impedance, i.e., $L \propto Z_0$ (Eq. 4.10). To put it another way, the characteristic impedance of the incoming transmission line is the dimensionful quantity that sets the scale for all of the circuit components (once a central frequency ω_0 has been chosen). So we can reduce the required inductance simply by reducing the incoming characteristic impedance. For implementation, this translates to a widening of the incoming and outgoing microstrip lines. Since other parts of the optical chain may require other impedances for various reasons, it is useful to be able to design transformers to vary the impedance over an order of magnitude or more if need be.

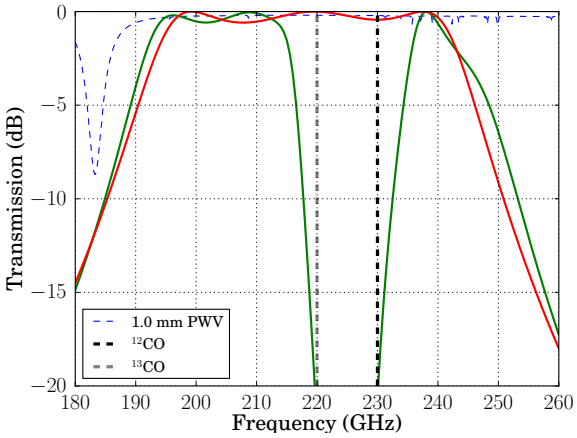
For example, the diplexer whose S parameters are shown in Fig. 4.23(b) uses incoming microstrip lines that are 26-μm wide, which produces a characteristic impedance of 4 Ω. The antenna, however, which these bandpass filters are connected to is fed with 2-μm microstrip lines ($Z_0 = 40$ Ω), and the bolometer terminations are 10 Ω, requiring a microstrip width of 10 μm. The relatively low characteristic impedance entering and exiting the bandpass filters enables fractional bandwidths of approximately 12%, which is necessary for splitting the 150-GHz atmospheric window into two frequency bands.

4.6 Bandstop filters

It may be desirable in some circumstances to explicitly reject a narrow band of frequencies. These filters can be designed according to the insertion-loss method described in Sec. 4.5.1. Just as we did for bandpass filters, we begin with a low-pass prototype. But now we make



(a) A 3-pole bandstop filter that is meant to connect to a bandpass filter at left (not shown). The half-wavelength transmission-line segment on the left is crucial, although there is a relatively large tolerance on the exact length.



(b) Simulated performance (ADS) with and without a bandstop filter in series.

Figure 4.24: A design for a bandstop filter to reject CO lines from a 220-GHz band. The bandstop filter is placed in series with a bandpass filter.

the transformation

$$\omega \to -\Delta \left(\omega - \frac{1}{\omega} \right)^{-1}, \tag{4.16}$$

where we are working in units for which the central frequency $\omega_0 = 1$. This expression is the bandstop analog of Eq. 4.7 and creates circuits like the one shown in Fig. 4.24(a). Notice that parallel LC pairs are now in series and vice versa relative to the bandpass filters of Fig. 4.18(b).

4.6.1 Galactic CO

An example that has been discussed in the POLARBEAR collaboration is that of galactic carbon-monoxide (CO) lines. There are multiple isotopologues of CO, but the most common are ^{12}CO and ^{13}CO, which have rotational modes at integer multiples of 115 GHz and 110 GHz, respectively. The *Planck* collaboration saw significant CO contributions to almost all of their HFI channels [76, 79]. There are large variations in CO intensity across the sky. When observing in molecular clouds and the galactic plane, as much as 50% of the contribution to the 100-GHz channel comes from CO. At the same time, the CO contribution at high galactic latitudes is negligible. The *Planck* collaboration uses a fiducial ^{13}CO-to-^{12}CO ratio of 0.2.

Diatomic molecules of relatively low atomic numbers tend to have moments of inertia that put their lowest rotational modes in the mm-wave regime. For example, carbon monosulfide (CS) has resonances at integer multiples of 49 GHz. The *Planck* collaboration also

mentions HCN, CN and HCO+ as molecules that contribute toward the galactic nucleus. A compilation of such resonances can be found on `splatalogue.net`. It is actually not hard to convince oneself that such molecular lines will show up at mm-wavelengths. An order-of-magnitude estimate of the lowest rotational transition in CO can be made by estimating the moment of inertia to be

$$I \approx m_{\mathrm{C}}(d/2)^2 + m_{\mathrm{O}}(d/2)^2, \tag{4.17}$$

where $m_{\mathrm{C}} = 12$ u is the mass of a carbon atom, $m_{\mathrm{O}} = 16$ u is the mass of an oxygen atom and d is the distance between the two atoms. Without knowing much about CO, we can guess that d is on the order of 1 Å. Again, we are just trying to estimate the order of magnitude of the rotational modes. If we take the energies to be

$$E_J = \frac{J(J+1)\hbar^2}{2I} \tag{4.18}$$

for non-negative integer J, then the $J = 1 \to 0$ transition has a frequency of 140 GHz, which differs from the actual value of 115 GHz by only about 20%.[6] Since the photon is a spin-1 particle, the dominant transitions are $J \to J - 1$, which emit photons of energy

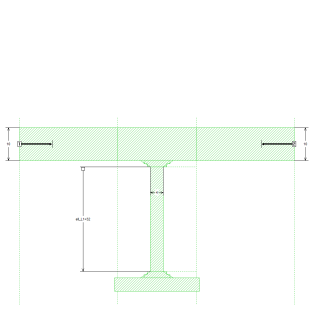
$$\Delta E_{J\to J-1} = J\frac{\hbar^2}{I} = J\Delta E_{1\to 0}, \tag{4.19}$$

i.e., the transitions are all harmonics of a fundamental. Galactic CO is not hot enough to create redshift broadening anywhere close to the width of the frequency bands used in CMB experiments, so these resonances can be taken to be essentially δ-functions centered on the rest-frame frequencies.

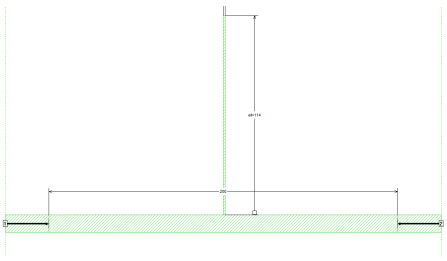
In intensity, CO is a substantial foreground in the *Planck* maps, but *Planck* does not detect CO polarization. In fact, only a few tentative detections of CO polarization have been made at all [35, 36, 38, 57]. The molecular emission is hypothesized to be polarized by magnetic fields in what is called the Goldreich-Kylafis effect. Polarized CO could then be used to map galactic magnetic fields. In the tentative detections mentioned above, the polarization varies significantly over even a few arcseconds, so this is likely to average down in the relatively large beams of CMB experiments.

To be conservative, the POLARBEAR collaboration has chosen to avoid the strongest CO line, which is the $J = 1 \to 0$ resonance, and to accept contamination at the second resonance, which is the $J = 2 \to 1$ resonance. The $J = 1 \to 0$ lines, which encompass the range 110-115 GHz, are easily avoided, since they are near an atmospheric line that we avoid anyway. We simply shift the band a little farther from this atmospheric line to be safe. It would be a major disruption, however, to a 220-GHz band to attempt to avoid the $J = 2 \to 1$ lines, which encompass the range 220-230 GHz. One idea is to narrow this band and center it on something closer to 200 GHz. Another option is to use a bandstop filter in series with the nominal 220-GHz bandpass filter. Ideally, this circuit will remove power only in the range 220-230 GHz but keep the band otherwise intact. A possible design is shown in Fig. 4.24.

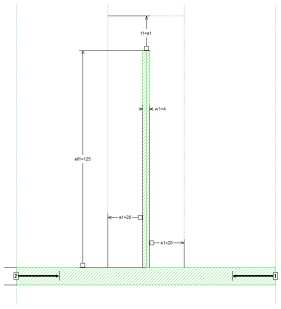
[6] The actual bond length for CO is 1.13 Å. Using this value and configuring the rotation about the center of mass, we estimate 113 GHz, which is within 2% of the true value.



(a) Lumped-element LC. The upper element is a CPW-style short inductor, and the lower element is a wide microstrip-style shunt capacitor.



(b) Microstrip stub. The thin vertical line is nothing more than a quarter-wavelength microstrip segment.



(c) Coplanar-waveguide (CPW) stub. The dashed lines show that the ground plane has been cut away underneath the stub to create the CPW.

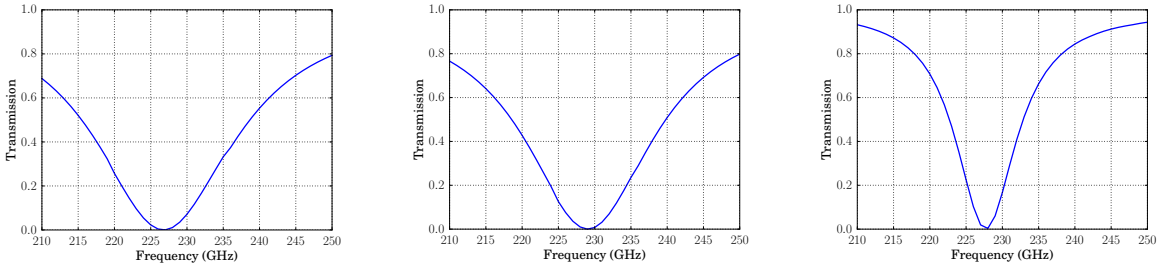
Figure 4.25: Three types of notch filters. Each is a 1-pole bandstop filter and functions as a shunted LC resonator. A microstrip lines runs from right to left; some power is shunted through the resonator to create the stopband.

4.6.2 Notch filter

When a bandstop filter is especially narrowband, it is sometimes referred to as a *notch filter*. When we with to remove a single frequency, an attractive design is a 1-pole bandstop filter, since it is simple and removes power at a particular frequency. A 1-pole bandstop filter can be constructed from a shunt LC pair, where the inductor and capacitor are in series with each other. Imagine removing everything in Fig. 4.24(a) except for the shunt components. This is shown as a Sonnet-Software layout in Fig. 4.25(a). Since a quarter-wavelength open-circuited transmission-line stub has approximately the same input impedance as a series LC resonator, we can also construct a 1-pole bandstop filter without any lumped components. The higher the impedance of the stub, the narrower the stopband, i.e., $Z_{0,\text{stub}} \propto 1/\Delta$. For a microstrip realization, this pushes the design toward narrow lines as shown, for example, in Fig. 4.25(b). To boost the impedance, it is possible to use a coplanar waveguide (CPW) for the stub instead of a microstrip line. A layout for a CPW notch filter is shown in Fig. 4.25(c). So there are three types of notch filters easily available: a lumped LC resonator, an open-circuited microstrip stub and an open-circuited coplanar-waveguide stub. The simulated transmission curves ($S(1,2)$) for these notch filters are shown in Fig. 4.26.

4.6.3 Material characterization

In this section, we describe a proposed method for characterizing on-wafer materials using notch filters. This method has not been properly utilized, since the performance of devices fabricated in the Marvell Nanolab became satisfactory before the relevant material-characterization measurements were made. For future devices that use new materials, this method may prove useful, and for this reason we provide a description.



(a) Lumped element (b) Microstrip (c) CPW

Figure 4.26: Transmission curves ($S(1,2)$) for the filters shown in Fig. 4.25. Notice that the CPW notch filter is noticeably narrower than the other two.

There are multiple material properties that affect the performance of our microwave circuits, e.g., the dielectric constant of the silicon wafer, the dielectric constant of the lenslet, the dielectric constant of the lenslet anti-reflection coating, the sheet resistance of the titanium film, the dielectric constant of the microstrip dielectric, the dielectric constant of the low-stress silicon nitride, the sheet inductance of the niobium, etc. Many of these we think we know pretty well or are able to test in other ways. The three that are most uncertain are the dielectric constant of the microstrip dielectric, which we will denote by $\epsilon_{\rm MS}$, the dielectric constant of the low-stress silicon nitride, which we will denote by $\epsilon_{\rm LSN}$, and the sheet inductance of the niobium film, which we will denote by L_s. We have some idea of what these quantities are. The method outline below looks for deviations from these expectations.

As these material properties are changed, the notch-filter resonances will shift in frequency space. If the different types of notch filters depend differently on the material properties, there is some hope that we could invert the measurements to recover the material properties. The 1-pole notch filters are relatively simple structures that pick out a well-defined frequency, which makes it relatively easy to identify a frequency shift.

Let the resonant frequency be denoted ν_0. We want to know how the resonant frequency changes as a function of the material properties. In a linear approximation, then, it is sufficient to know $\partial\nu_0/\partial L_s$, $\partial\nu_0/\partial\epsilon_{\rm MS}$ and $\partial\nu_0/\partial\epsilon_{\rm LSN}$. We can estimate these quantities in simulation. We take small deviations from fiducial values, i.e., our best current estimates, and treat them as linear perturbations. For example, we denote the departure of the sheet inductance from its fiducial value as ΔL_s. So we are estimating the partial derivatives near the fiducial values. Then the shift in the resonant frequency for the ith structure is

$$\Delta\nu_{0,i} = \frac{\partial\nu_{0,i}}{\partial L_s}\Delta L_s + \frac{\partial\nu_{0,i}}{\partial\epsilon_{\rm MS}}\Delta\epsilon_{\rm MS} + \frac{\partial\nu_{0,i}}{\partial\epsilon_{\rm LSN}}\Delta\epsilon_{\rm LSN}. \tag{4.20}$$

This looks just like a linear transformation. We need three independent structures to solve, so that we have the invertible mapping

$$(\Delta L_s, \Delta\epsilon_{\rm MS}, \Delta\epsilon_{\rm LSN}) \leftrightarrow (\Delta\nu_{0,1}, \Delta\nu_{0,2}, \Delta\nu_{0,3}). \tag{4.21}$$

So we define the linear transformation

$$A \equiv \begin{pmatrix} \frac{\partial \nu_{0,1}}{\partial L_s} & \frac{\partial \nu_{0,1}}{\partial \epsilon_{\mathrm{SNMS}}} & \frac{\partial \nu_{0,1}}{\partial \epsilon_{\mathrm{LSN}}} \\ \frac{\partial \nu_{0,2}}{\partial L_s} & \frac{\partial \nu_{0,2}}{\partial \epsilon_{\mathrm{SNMS}}} & \frac{\partial \nu_{0,2}}{\partial \epsilon_{\mathrm{LSN}}} \\ \frac{\partial \nu_{0,3}}{\partial L_s} & \frac{\partial \nu_{0,3}}{\partial \epsilon_{\mathrm{SNMS}}} & \frac{\partial \nu_{0,3}}{\partial \epsilon_{\mathrm{LSN}}} \end{pmatrix}. \tag{4.22}$$

Define a column vector for the material-property deviations from fiducial to be

$$q \equiv \begin{pmatrix} \Delta L_s \\ \Delta \epsilon_{\mathrm{SNMS}} \\ \Delta \epsilon_{\mathrm{LSN}} \end{pmatrix}, \tag{4.23}$$

and define a column vector for the resonant-frequency shifts to be

$$p \equiv \begin{pmatrix} \Delta \nu_{0,1} \\ \Delta \nu_{0,2} \\ \Delta \nu_{0,3} \end{pmatrix}. \tag{4.24}$$

Then $p = Aq$, where we measure p and simulate A. Then we can compute

$$q = A^{-1}p, \tag{4.25}$$

but only if A is invertible. Since the frequency shifts are proportional to the resonant frequency, it is not possible to simply use three notch filters of the same type but tuned to different frequencies. This type of system would produce redundant information, and the matrix A would be singular. It would, however, be useful to measure at multiple frequencies as a way of characterizing statistical errors. With a broadband antenna like a sinuous antenna, it would be possible to put notch filters of different frequencies in series, since they will not interact with each other if placed far apart in frequency space. This takes up little extra space, and the measurements are performed simultaneously. To ensure that the matrix A is invertible, we need three resonant structures that shift frequencies differently as a function of the material properties. For this purpose, we can use the microstrip, CPW and lumped-element 1-pole notch filters described above.

Simulated example To see how well this method can be expected to work, we ran some simulations. The layer stack-up include a silicon wafer of dielectric constant 11.8. We use silicon nitride as a microstrip dielectric. The fiducial material properties are

$$\epsilon_{\mathrm{LSN}} = 7.3, \qquad \epsilon_{\mathrm{MS}} = 7.3, \qquad L_s = 0.13 \text{ pH}/\square. \tag{4.26}$$

By perturbing these material properties in Sonnet simulations, we estimate

$$A = \begin{pmatrix} -320 \frac{1}{\mathrm{pH}/\square} & -15 & -0.38 \\ -33 \frac{1}{\mathrm{pH}/\square} & -2.5 & -2.3 \\ -50 \frac{1}{\mathrm{pH}/\square} & -15 & 0 \end{pmatrix} \text{ GHz}, \tag{4.27}$$

where the top row provides the shifts for the microstrip filter, the middle row for the CPW filter and the bottom row for the lumped-element filter. It can be seen at a glance that these

rows are linearly independent, so we can invert this matrix. We now input some deviations in the material properties from their fiducial values and see how well this method recovers those inputs. We input

$$\Delta L_s = 0.04 \text{ pH}/\square, \qquad \Delta\epsilon_{\text{MS}} = -0.8, \qquad \Delta\epsilon_{\text{LSN}} = 0.6. \tag{4.28}$$

We simulate with these material-property shifts and treat the results as the measurements for this proof of principle. The electromagnetic simulations yield the following resonance shifts:

$$\Delta\nu_{0,MS} = 1 \text{ GHz}, \qquad \Delta\nu_{0,\text{CPW}} = -1 \text{ GHz}, \qquad \Delta\nu_{0,\text{lumped}} = 10 \text{ GHz}. \tag{4.29}$$

Applying the matrix A^{-1} gives

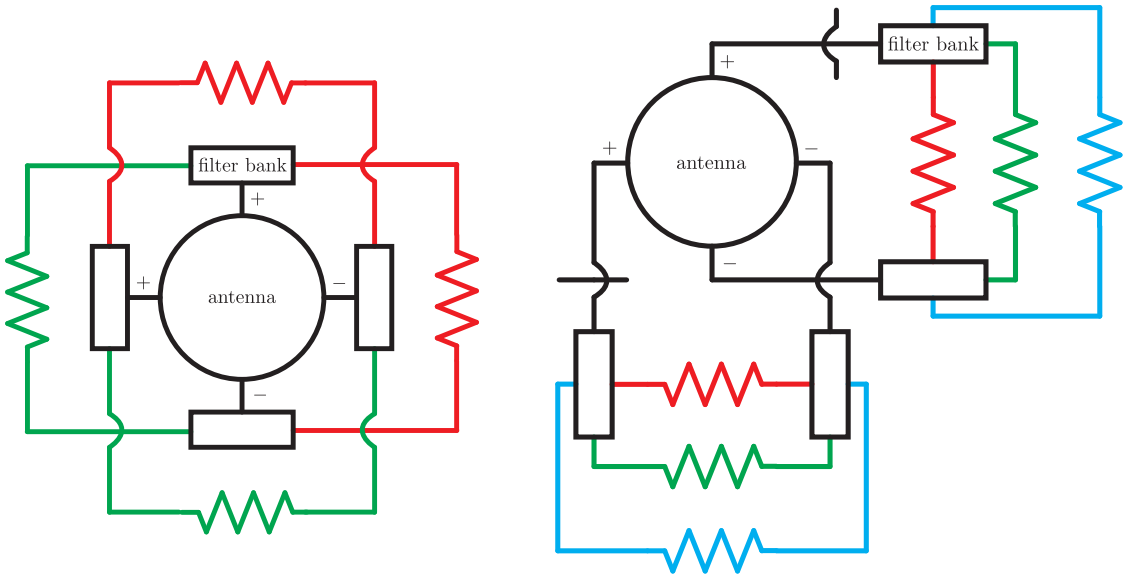
$$\Delta L_s = 0.032 \text{ pH}/\square, \qquad \Delta\epsilon_{\text{MS}} = -0.77, \qquad \Delta\epsilon_{\text{LSN}} = 0.81. \tag{4.30}$$

So we see that we are recovering the input parameters to 25-30%. Keep in mind that we have here estimated the *deviations* in these material properties from their fiducial values, so our estimates of the actual values of these parameters are significantly better. In this scenario, there was a negligible change in the microstrip and CPW resonators but a large change in the lumped-element resonator; if we saw this in a real measurement, it would mean that we underestimated the niobium sheet inductance and overestimated the microstrip dielectric constant.

4.7 Microstrip crossover

Since our antennas are dual-polarized and also differentially fed, it is impossible to lay out a strictly planar microwave circuit. This is simply a graph-theoretic statement and is illustrated in Fig. 4.27, where we have been forced to include microstrip *crossovers* to achieve all of the necessary electrical connections. If we could feed our antennas with a single transmission line for each polarization, then we would not need any crossovers. Unfortunately, we've been unable to show good performance from a sinuous antenna with a single-ended feed, so we have reverted to the more natural and symmetric differential feed. The double slot dipole is also naturally fed in this way. The issues related to multichroic pixels are common to the double slot dipole, where the only caveat is that the total bandwidth is severely restricted relative to the sinuous antenna. A dichroic design allows for an approximate four-fold symmetry in the placement of the bolometers (Fig. 4.27(a)). When three or more frequency bands are desired, it is more natural to switch to the kind of design shown in Fig. 4.27(b). We have only drawn this design for three frequency bands, but the generalization to more is straightforward. Since the bolometers for all but one frequency band are trapped inside rings of transmission lines, it is necessary to include additional crossovers in order to connect electrical leads to the bolometers. Another method for accessing these inner bolometers is to use wire bonds, which are essentially just gigantic post-fabrication crossovers.

A crossover requires two extra layers in the fabrication process: a metal layer to make the electrical connection over the original strip layer and a dielectric layer to insulate the two metal strip layers. Our devices have no other need for these two extra layers, and we



(a) Dichroic topology. The bolometers can be accessed with electrical leads *without* additional crossovers.

(b) Trichroic topology. The green and red bolometers *cannot* be accessed with electrical leads without additional crossovers. To keep the transmission-line signals as symmetric as possible, crossovers are added even where they are not strictly required.

Figure 4.27: Topology of multichroic pixels with a differentially-fed antenna. The colors indicate different frequency bands. For each linear polarization, the antenna is fed with two out-of-phase signals, which are denoted by the ± signs. The filter bank breaks the signal into frequency bands. The resistors indicate the transmission-line termination on the bolometer island. The bolometers are not explicitly drawn. Crossovers are unavoidable in these designs.

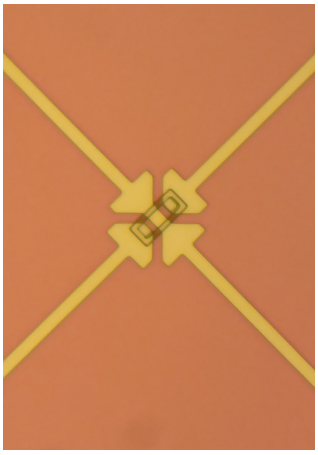


Figure 4.28: Fabricated crossunder. At the crossing point, the ground plane has been etched away, so the southwest-northeast line can via to the lower metal layer and avoid electrically contacting the southeast-northwest microstrip line.

have only seven other layers. A substantial fraction of the fabrication process, then, must be devoted purely to the creation of the crossover. For this reason, many of our prototype devices were made *without* crossovers in order to accelerate the fabrication process. Without crossovers, however, the microstrip lines for one linear polarization must be left open. This is why many of our prototypes are effectively single-polarization devices.

4.7.1 Crossunder

To avoid the additional two layers required for crossovers and still achieve sensitivity to two polarizations, we developed microstrip cross*under*s that use only the three layers required for the original microstrip lines, i.e., the ground plane, the microstrip dielectric and the strip layer. This is achieved by opening up the ground plane at the crossing point and sending one microstrip line underneath the other as shown in Fig. 4.28. There is a via which connects one of the microstrip lines to the ground-plane *layer* but not to the *electrical* ground. The microstrip impedance should be held approximately constant through the crossunder to prevent reflections. At the point where the two microstrip lines cross, it is desirable to use thin lines to prevent crosstalk. This creates a high inductance, however. To compensate for this increased inductance, we widen the microstrip lines away from the crossing point. If this can be done over a length much smaller than wavelength, the effective input impedance at the crossunder can be tuned to the characteristic impedance of the incoming microstrip lines. Since there is no longer a proper ground plane at the crossing point, there tends to be a very high inductance even from a short crossing segment. This pushes the natural characteristic impedance of the incoming microstrip lines to larger values, e.g., 20-30 Ω, which corresponds to 3-4 μm in width for typical dielectric thicknesses. Otherwise, the compensating capacitive segments must be too large to keep the crossunder electrically small. This design has now been baselined for the PB-2 device wafers.

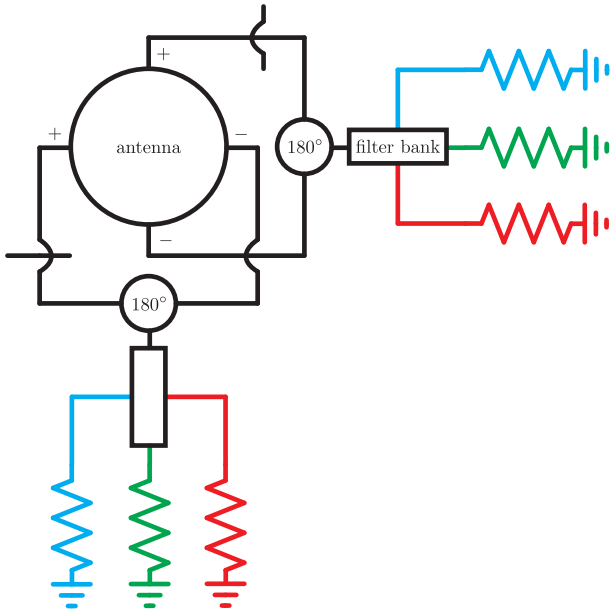


Figure 4.29: Topology of a trichroic pixel using a broadband 180° hybrid (cf. Fig. 4.27). All of the bolometers can be accessed with electrical leads *without* additional crossovers. There is only one filter bank per polarization, and there are no symmetry constraints after the signal exits the hybrid (on the bolometer side). The transmission-line terminations are now single-ended, which can be accomplished either with a lossy meander or a lumped resistor connected to the ground plane with a via through the microstrip dielectric.

4.7.2 Broadband 180° hybrid

For three or more frequency bands, the pixel topology requires large numbers of crossovers (or crossunders) as demonstrated in Fig. 4.27(b). Recall that there are additional crossovers required beyond what is shown in the diagram, since the bolometer leads must also make their way out of the pixel to the edge of the wafer for wire bonding. Additionally, the microwave circuitry becomes more congested with more frequency bands, since there are two microstrip lines per polarization per frequency band.

We can simplify the microwave circuit with the use of a broadband 180° hybrid. A 180° hybrid combines two out-of-phase transmission lines into one. In the reverse-time sense, the power is split evenly into two out-of-phase signals. The topology of a multichroic pixel that makes use of a broadband 180° hybrid is shown in Fig. 4.29. In this design, there is only one filter bank per polarization. The microstrip line breaks out to frequency bands just before the bolometers instead of near the antenna. The bolometer leads require no additional crossovers. In fact, the number of crossovers is independent of the number of frequency bands. The microstrip termination on the bolometer islands are no longer differential; instead, they are single-ended, which can be accomplished either with a lumped resistor that is connected to the ground plane with a via through the microstrip dielectric or with a lossy meander.

It is common in microwave engineering to make use of relatively narrowband 180° hybrids like the ring hybrids described in Sec. 5.3.2. It is difficult and less common to design a hybrid

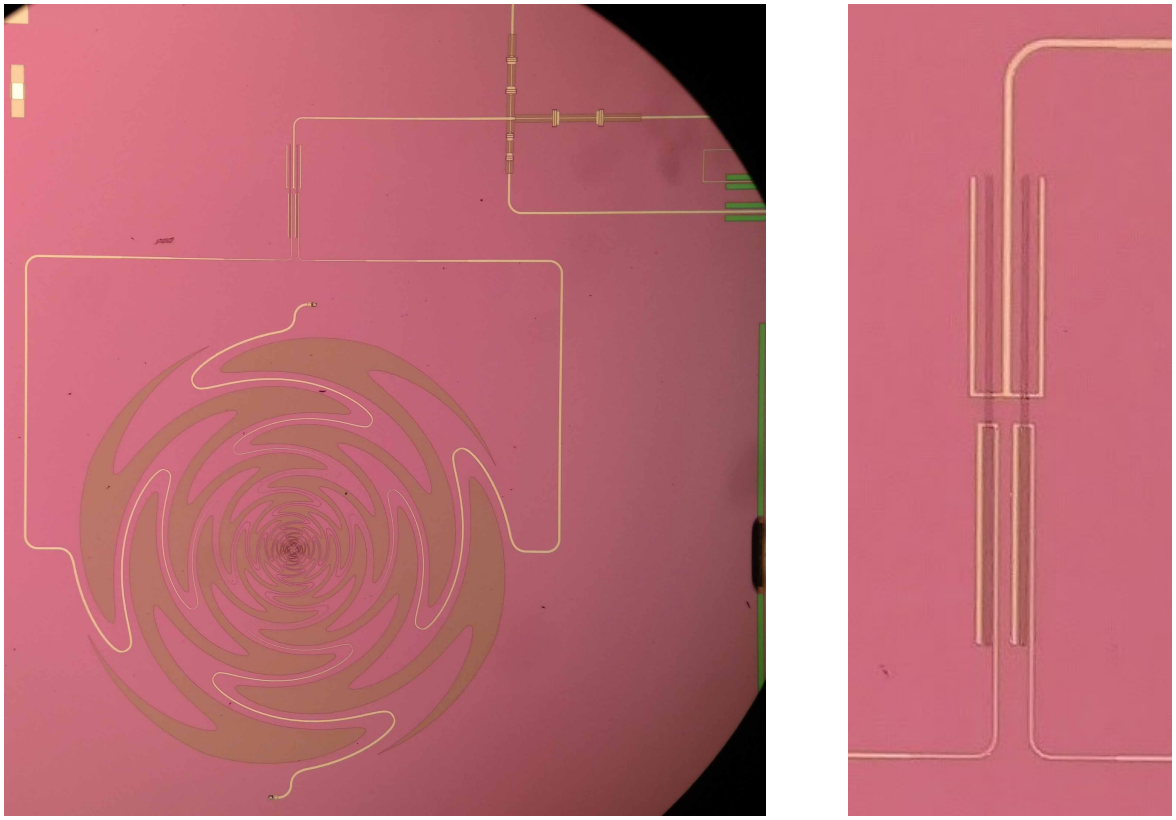


Figure 4.30: A broadband 180° hybrid relying on a microstrip-slotline-microstrip coupling. The two lines at bottom couple asymmetrically to the two slotlines; the line at top couples symmetrically. A signal injected at top is split evenly into the two bottom lines but with a 180° phase shift between them. The two lines at bottom feed the sinuous antenna differentially. The line at top enters bandpass filters that split the signal into frequency bands. The power in each frequency band is then terminated on a separate TES bolometer.

that operates over a large bandwidth.

We designed a 180° hybrid that operates across a 4:1 bandwidth. This component saves space, simplifies the design and reduces the number of microstrip cross-overs. Our design, which is shown in Fig. 4.30, relies on a microstrip-slotline-microstrip coupling. In the reverse-time sense, i.e., traveling from the bolometer toward the antenna, the power is split evenly and symmetrically into two slotlines. The power is then picked up from these slotlines by two separate microstrip lines, but the coupling is asymmetric, i.e., one microstrip line comes from the region outside the two slotlines and one from the region in between. This asymmetry creates a 180° phase difference between these two microstrip lines, which are then used to feed the antenna.

We simulated this design using Sonnet Software, and the results are shown in Fig. 4.31. The power is split evenly, i.e., $S(1,2) = S(1,3) \approx -3$ dB from 70 to 280 GHz with a phase difference within 1° of 180°. The coupling between the out-of-phase lines is at an average level of approximately −6 dB, which is substantially larger than for the ring hybrids

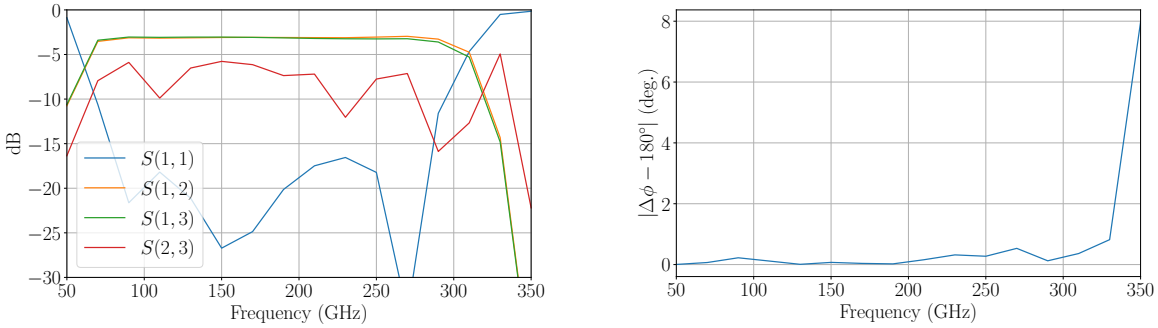


Figure 4.31: Simulated performance of the broadband 180° hybrid. (*Left*) The S-parameters, where port 1 is the microstrip on the right side of Fig. 4.30 and ports 2 and 3 are the microstrips on the left side. The power is split evenly ($S(1,2) = S(1,3) \approx -3$ dB) from 70 to 280 GHz. Compared with four-port ring hybrids described in Sec. 5.3.2, this hybrid has greater coupling between the output ports ($S(2,3)$). (*Right*) The phase difference between the output ports relative to 180°. A perfect 180° hybrid would show a flat line at 0°, which has been achieved here in simulation within 1° across 50-300 GHz.

of Sec. 5.3.2. This is mainly due to the lack of a fourth port to suppress crosstalk. It is important, therefore, to achieve a good impedance matching on the antenna side, so there are no reflections bouncing between these out-of-phase lines.

4.8 Prototype devices

In this section, we report on measurements of several prototype devices based on the designs described in the preceding sections.

4.8.1 Extending sinuous-antenna coverage

We designed a test chip, which is shown in Fig. 4.32, to extend the coverage of the sinuous antenna over approximately a decade in frequency.[7] The fabrication process used for this test chip is described in Westbrook et al. [103] and differs in some important ways from the process described in Sec. 3.3.5, which can be uncontroversially said to be superior. The fabrication yield for these test chips was relatively low. Many bolometers broke or were electrically open. The TESs that worked were often unstable. Since a 6″ wafer contains a large number of 20-mm dies, there were enough working devices to make the measurements presented in this section.

The test chip contained sinuous antennas tuned to three different frequency regimes, which can be roughly said to be low, middle and high frequencies. The target applications for these devices were in future balloon- and space-based telescopes, so we more or less ignore the atmospheric lines that constrain ground-based devices. The low-frequency device was

[7]This particular test chip was designed by the author but fabricated and tested by Benjamin Westbrook. We include some of his unpublished plots with his permission.

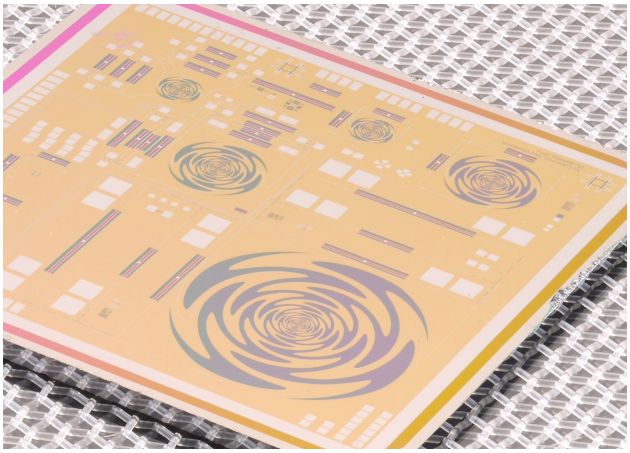


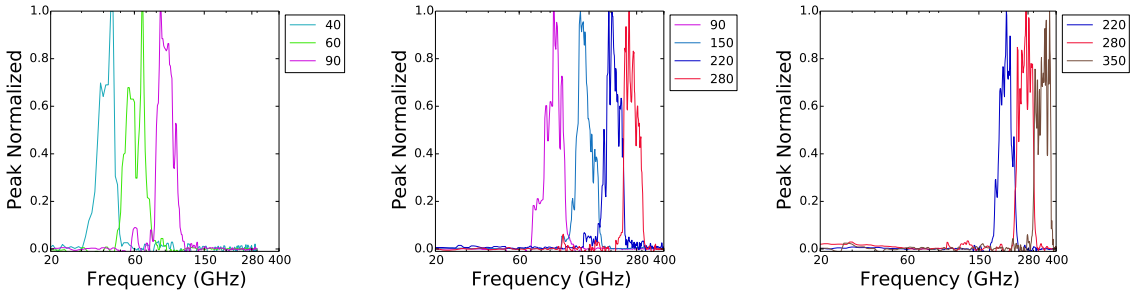
Figure 4.32: Test chip for extending the frequency coverage of the sinuous-antenna technology. The differently sized sinuous antennas target different frequency ranges. The chip is 20 mm on a side. (This photograph was taken with a wide-field microscope camera by our colleagues at NIST, Boulder.)

a triplexer with bands centered on 40, 60 and 90 GHz. The mid-frequency device was a tetraplexer with bands centered on 90, 150, 220 and 280 GHz. The high-frequency device was a triplexer with bands centered on 220, 280 and 350 GHz.

The frequency responses for these devices are shown in Fig. 4.33. Together the devices cover approximately a decade in frequency. The spectra have been peak-normalized, since the optical-efficiency measurements were confusing and erratic. We later discovered that the niobium film stress is crucial to maintaining high optical efficiency. When the niobium film stress is tensile, the efficiency can be very low due to an effective series resistance in the microstrip transmission lines, i.e., the niobium is not acting as a lossless superconductor. When the niobium film stress is compressive, the optical efficiency is more stable and at the levels we expect, but we made this discovery only after this chip had been tested. The polarization response was characterized using a rotating wire-grid polarizer. Some representative results are shown in Fig. 4.34. A typical minimum-to-maximum ratio for these polarization measurements is 2%, where at least 1% of the cross-polar response is due to leakage in the polarizing wire grid. We expect at least another 1% from the wobble of the polarization axis of the sinuous antenna. Some representative beam maps are shown in Fig. 4.35. The beams show a small amount of ellipticity. Some of this may be due to structure in the beam-mapper source.

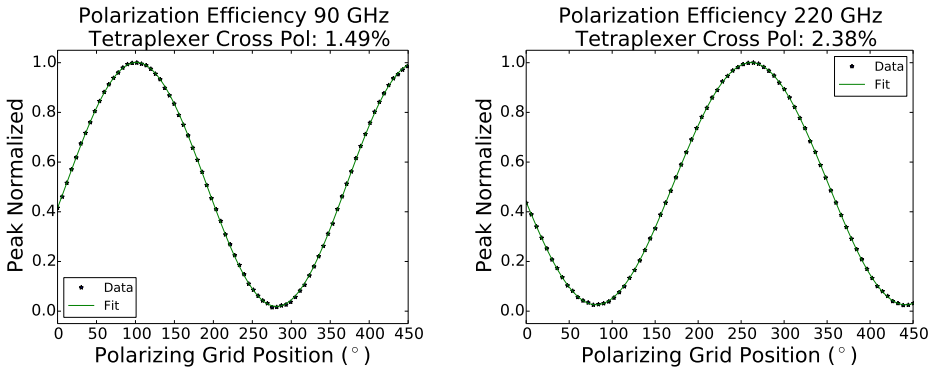
4.8.2 Split atmospheric window

The cosmological reach of CMB experiments is beginning to become limited by characterization of the galactic foregrounds rather than by raw sensitivity. For foreground removal, it is necessary to observe in multiple frequency bands so as to be able to fit for the foreground components. Usually, we place only one band in the 90- and 150-GHz atmospheric windows. With two bands, we could provide fine frequency coverage for precision foreground model-



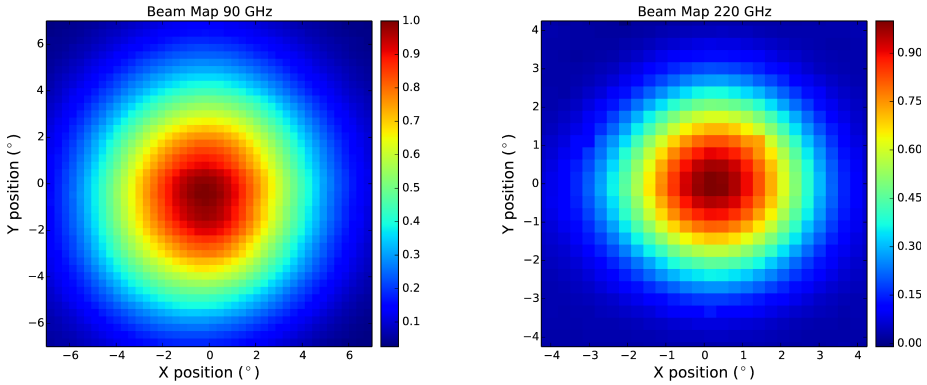
(a) Low-frequency triplexer (b) Mid-frequency tetraplexer (c) High-frequency triplexer

Figure 4.33: Spectra from the test chip of Fig. 4.32 covering a decade in frequency. The x-axes are the same in all plots. The spectra have been peak-normalized. (Plot credits: Ben Westbrook.)



(a) 90 GHz (b) 220 GHz

Figure 4.34: Polarization measurements from the mid-frequency tetraplexer. (Plot credits: Ben Westbrook.)



(a) 90 GHz (b) 220 GHz

Figure 4.35: Beam maps from the mid-frequency tetraplexer. (Plot credits: Ben Westbrook.)

ing. This is unlikely to be useful in the near future, since the foregrounds are so poorly constrained that it is more useful to cover a larger total bandwidth than to observe with fine frequency coverage. Nevertheless, we designed and fabricated a prototype pixel that splits the 150-GHz atmospheric window. The pixel uses a lenslet-coupled double slot dipole tuned to a resonant frequency of 150 GHz. The antenna feed lines are split with diplexing bandpass filters into an upper and a lower band. Since the fractional bandwidths are half as large as we typically design, the microstrip impedance was lowered to relax the inductance requirements as described in Sec. 4.5.4. The microstrip dielectric had either the wrong thickness or an unexpected dielectric constant, so the frequency bands shifted down by about 10 GHz from the designed band centers. The measured passbands are shown in Fig. 4.36. The bands have shifted down, but this shift could relatively easily be tuned away in one or two fabrication cycles. Increasing the dielectric thickness will shift the bands upwards without any changes to the photolithographic masks. Although the measured bands do not fit neatly in the atmospheric window, we achieved narrow enough passbands to fit two bands into a comparable fractional bandwidth. A simple shift in frequency would make this design viable.

4.8.3 Broadband 180° hybrid

Here we report on measurements of the broadband 180° hybrid described in Sec. 4.7.2. Measurements of the bands are shown in Fig. 4.37(a). The spectra show some degradation but are correctly placed. Measurements of the beam profiles are shown in Fig. 4.37(b). The beams are round, and they are all co-centered to better than 1°. The lack of frequency-dependent beam steering indicates that the hybrid splits power evenly with a 180° phase difference over the entire frequency range. The full beam maps are shown in Fig. 4.38. The optical efficiency of the devices can be seen in Fig. 4.37(a). This plot shows the end-to-end efficiency and includes losses in the dewar optics. Even with that caveat, however, the efficiency is too low for a deployable device, and the cause is unclear. This problem will need to be solved before this device can be considered viable.

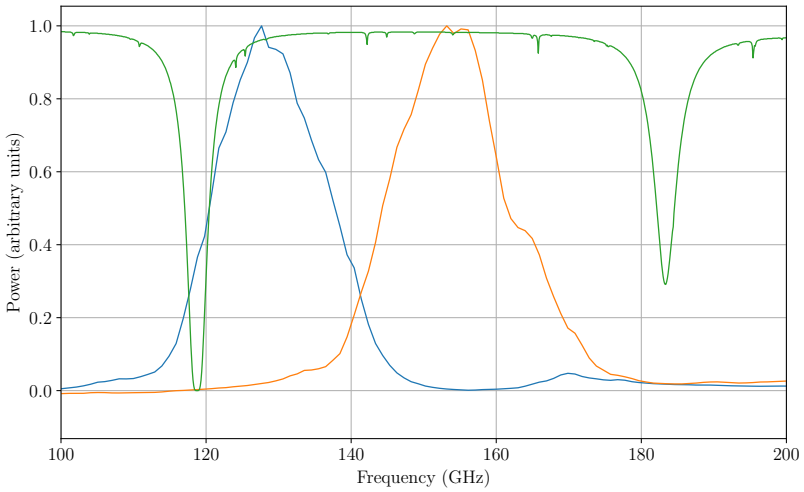
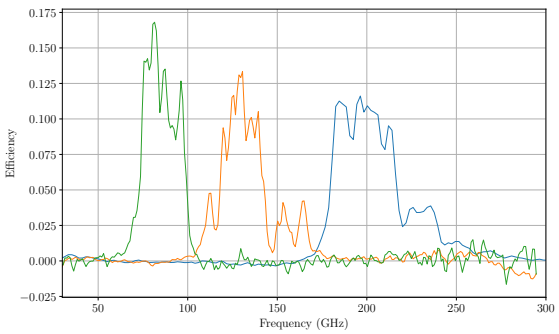
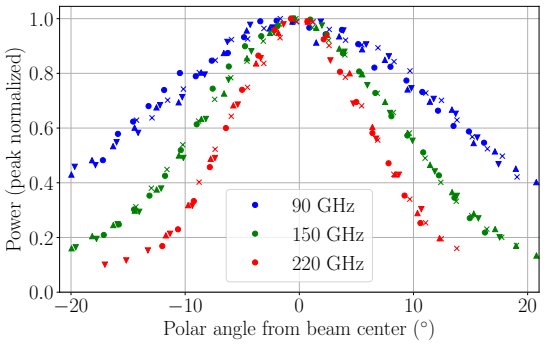


Figure 4.36: Frequency response from a prototype 136/162-GHz diplexer.

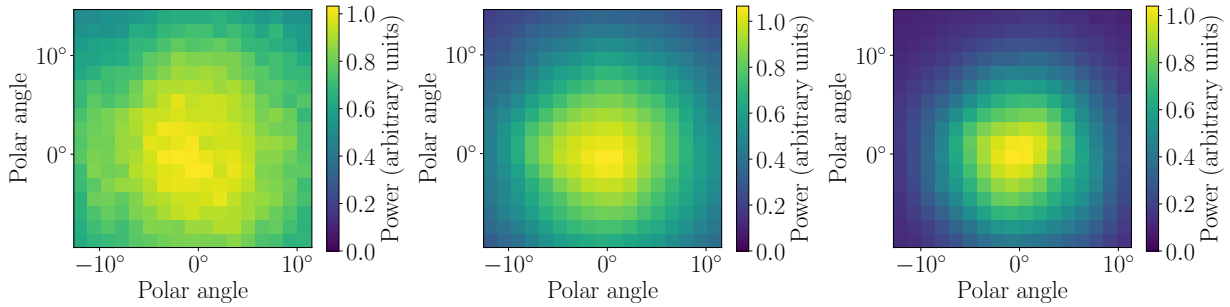


(a) Spectra showing 90-, 150- and 220-GHz bands. These curves have been normalized according to the prescription presented in Sec. 3.4.6. The y-axis is now the end-to-end efficiency including losses in the dewar optics.



(b) Beam profiles. The different markers correspond to different azimuthal cuts, i.e., 0°, 45°, 90° and 135°.

Figure 4.37: Measurements of a prototype broadband 180° hybrid.



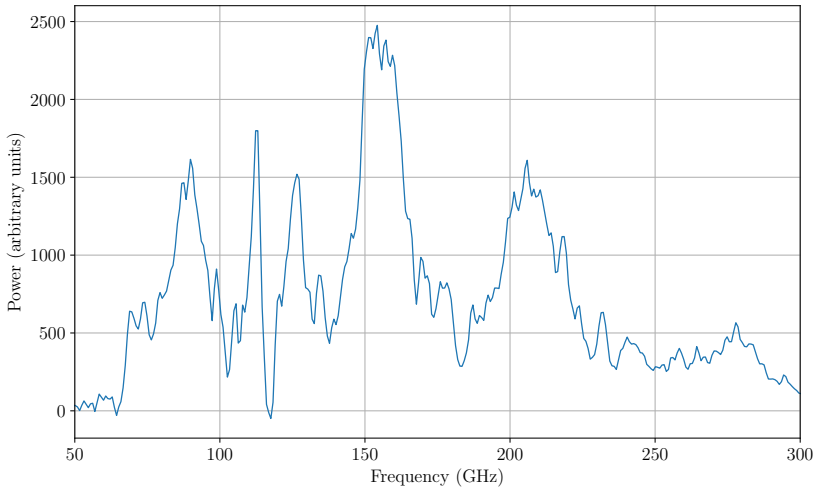
(a) 90 GHz (b) 150 GHz (c) 220 GHz

Figure 4.38: Beam maps from a device with a prototype broadband 180° hybrid.

4.8.4 Notch filter

We tested a 1-pole CPW bandstop filter connected to a filterless sinuous antenna. A photograph of the fabricated notch filter is shown in Fig. 4.39(a). The filter was designed to reject a narrow band near the lowest rotational mode of ^{12}CO, i.e., 115 GHz. The results of an FTS scan are shown in Fig. 4.39(b), where we see a clear notch at approximately the design frequency. The measured notch is actually centered around 117 GHz. Since only this one measurement was made, there is not enough information to meaningfully constrain the material properties used in the microwave circuitry; see Sec. 4.6.3 for a description of how to use notch filters for material characterization.

(a) A fabricated CPW 1-pole bandstop filter.



(b) Measured spectrum of a filterless sinuous antenna with a CPW notch filter designed for 115 GHz. The measured notch is centered on 117 GHz.

Figure 4.39: Prototype 115-GHz notch filter, which could be used to reject the $J = 1 \rightarrow 0$ line of ^{12}CO. Since there is also a ^{13}CO line at 110 GHz, a future design should perhaps use a multipole bandstop filter with a larger bandwidth.

Chapter 5

Multiscale focal planes

We describe a new architecture for multichroic focal planes. Some of the text in this chapter was taken from Cukierman et al. [23].

5.1 Mapping speed

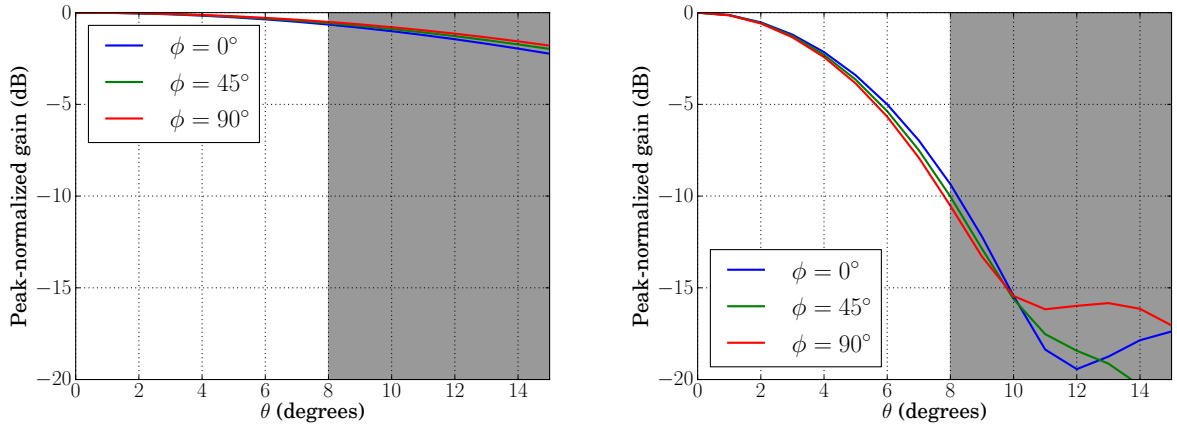
In this section, we describe a limitation to the sensitivity of a conventional multichroic focal plane. In the following section, we outline a scheme for solving this problem.

5.1.1 Beam variation

A lenslet-coupled sinuous antenna has a beam waist that is approximately constant with frequency. We can gain an intuition for this by appealing to the case in which the lenslet-air interface is in the far field of the sinuous antenna. The radiation radius $r_{\text{rad.}}$ on the sinuous antenna is proportional to wavelength λ. Then $\lambda/r_{\text{rad.}}$ is approximately constant, which means that the beam radiated into a silicon half-space will have approximately the same width at all frequencies. The beam, therefore, illuminates the lenslet-air interface in the same way at all frequencies. This illumination acts as an effective aperture which radiates to free space. Since this illumination is approximately constant as a function of frequency, the effective beam waist is constant with frequency.

The pixel sizes we typically use in CMB experiments are electrically small, so the lenslet-air interface cannot be taken to be in the far field. Nevertheless, it is true both in electromagnetic simulations and in measurements that the effective beam waist is approximately constant with frequency. A typical value for the beam waist, defined as the effective radius at the focal plane at which the illumination has reach $1/e^2$ of its peak value, is about 60% of the pixel radius.

Since the beam waist is approximately constant, the beam width varies roughly inversely with frequency much in the same way as the radiation from a circular diffraction aperture ($\theta \sim \lambda/D$). Shown in Fig. 5.1 are simulated beam profiles from an ultrawideband lenslet-coupled sinuous antenna. At 195 GHz, the beam hits the Lyot aperture at its -10-dB point. At 40 GHz, the beam hits the Lyot aperture closer to 0.5 dB. This indicates that the coupling to the receiver optics will vary wildly across the frequency range. In addition to possible



(a) 40 GHz (b) 195 GHz

Figure 5.1: Beam profiles for a LiteBIRD-style pixel. The shaded region represents the Lyot stop.

systematic effects in the on-sky beam and spurious side lobes, there is a direct degradation of the experimental sensitivity due to this beam-width variation, which we discuss in the following section.

5.1.2 Sensitivity optimization

The noise-equivalent temperature (NET), often given in units of $\mu\mathrm{K}\sqrt{\mathrm{s}}$, is a measure of the instantaneous sensitivity of an experiment to a sky signal such as the CMB.[1] This is the quantity that one divides by the square root of the integration time to find the RMS temperature fluctuations in a measurement of a sky pixel. A typical NET for a single-moded detector is on the order of 500 $\mu\mathrm{K}\sqrt{s}$. The noise in an experiment's maps of the sky can be reduced by integrating for longer lengths of time. For a white noise spectrum, the noise-equivalent power (NEP) is constant with frequency, and we have

$$\mathrm{NET} = \left.\frac{\partial T}{\partial P}\right|_{P=\langle P\rangle} \frac{\mathrm{NEP}}{\sqrt{2}}, \tag{5.1}$$

where the factor of 2 ultimately stems from the fact that there are two independent modes at each frequency, e.g., sine and cosine, and the derivative is evaluated at the mean power giving a linear approximation to the power-to-temperature conversion. The NEP is a spectral quantity often given in units of W $\mathrm{Hz}^{-1/2}$. This is the quantity that one squares and integrates over frequency to find the variance in power in a given frequency band. A typical NEP for a single-moded detector is on the order of 100 aW $\mathrm{Hz}^{-1/2}$. The mapping speed

$$\mathrm{MS} = \frac{1}{\mathrm{NET}^2}, \tag{5.2}$$

[1] In other contexts, the NET is defined as a spectral quantity, e.g., with units of K $\mathrm{Hz}^{-1/2}$, analogous to the conventional definition of noise-equivalent power (NEP) but with power converted to temperature.

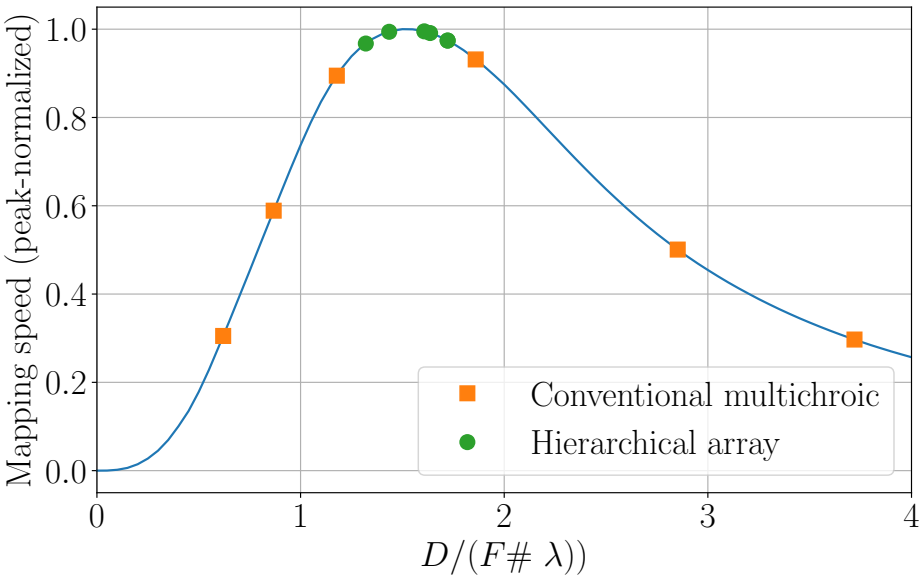


Figure 5.2: Mapping speed vs. pixel diameter (*solid*) in units of f-number ($F\#$) of the detector-aperture system times wavelength (λ) for an example multichroic focal plane for a space-based telescope [64]: 2.7-K sky with a 4-K aperture stop and bands centered on 50, 70, 95, 150, 230 and 300 GHz. A more precise treatment would yield slightly different curves for different frequencies, but the optimal diameter typically varies by less than 10% in units of $F\#\ \lambda$. *Square* markers indicate where the frequency bands of a conventional multichroic pixel would fall on this curve when the pixel diameter is optimized for 120 GHz; *circular* markers indicate the same for a hierarchical focal plane based on triangular arrays for which the array-element diameter is optimized for 270 GHz. The hierarchical array allows near-optimal mapping speed in all frequency bands.

often given in units of $\mu\mathrm{K}^{-2}\ \mathrm{s}^{-1}$, is a measure of the rate at which the map noise is reduced. An experiment with twice the mapping speed can achieve the same map noise in half the time.

The dependence of the mapping speed on pixel size is well known in the CMB community [39]. Here we consider the common case of single-moded detectors and a cryogenic aperture stop with the angular size of the telescope's field of view held constant. A smaller pixel size allows for more pixels, which tends to increase the mapping speed. However, a smaller pixel has a wider beam and, therefore, receives a smaller fraction of power from the sky and a larger fraction from the aperture stop in the telescope optics, which contributes only noise at a level which depends on the effective temperature. There are, therefore, two competing concerns in choosing a pixel size. A plot of mapping speed vs. pixel size for an example experiment is shown in Fig. 5.2, which shows a clear peak indicating an optimal pixel diameter. The pixel can be treated roughly as a diffraction aperture, so the diameter-to-wavelength ratio sets the beam width. As a result, the optimal pixel diameter is approximately proportional to wavelength. The target beam width depends on the distance to and the size of the aperture the detector is meant to couple to. We can define an effec-

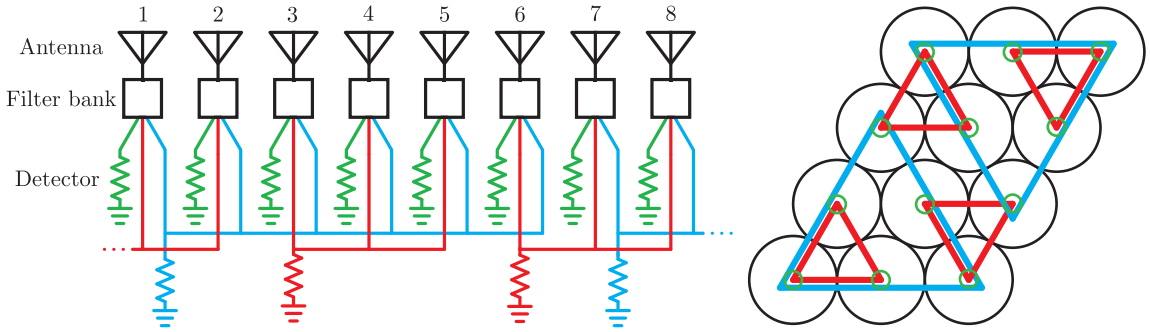


Figure 5.3: (*Left*) Example topology of a three-level hierarchical phased array. The highest frequencies (*green*) dissipate power to bolometers in a conventional way. The middle frequencies (*red*) sum three antennas coherently before dissipating power. The lowest frequencies (*blue*) sum six antennas. (*Right*) Effective pixels for a three-level hierarchy based on triangular arrays. *Black circles* represent individual antennas. The *small green circles* represent the effective pixels for the highest frequencies, *small red triangles* the middle frequencies and *large blue triangles* the lowest frequencies. The triangular hierarchy naturally creates a second hierarchy of rhombi, which can be used to tile the wafer.

tive f-number, often denoted $F\#$, of the detector-aperture optical system.[2] A higher $F\#$ requires a narrower beam and, therefore, a larger pixel diameter D. This relationship is approximately linear, so we often plot mapping speed vs. pixel size in units of of $F\#\ \lambda$.

A common problem with a conventional multichroic pixel, which is sensitive to multiple frequency bands but has a single fixed pixel diameter, is that the beam width and, therefore, the optimal pixel diameter are frequency-dependent. It is usually impossible, then, to choose a pixel diameter that is optimal in every frequency band. A *multiscale* focal plane solves this problem by creating pixels whose effective diameters scale with wavelength.

5.2 Hierarchical phased array

A multiscale focal plane can be realized through the use of a *hierarchical phased array*, which consists of superimposed antenna arrays operating in different frequency bands and increasing in size approximately inversely with frequency. An example topology is shown in Fig. 5.3 for a three-level hierarchy based on triangular arrays. The antennas couple free-space radiation to transmission lines, at which point the power is split into frequency bands via the filter banks. Transmission-line signals from neighboring antennas are then summed coherently before dissipating the power on a bolometer. The summing network includes a different number of antenna elements for each frequency band. The estimated improvement to mapping speed is shown in Fig. 5.2.

[2]The f-number of a lens is the ratio of the focal length to the lens diameter.

5.2.1 Antenna array

The radiating electric field of an antenna located at a position $\mathbf{x}$ can be written as

$$E = V f(\hat{\mathbf{n}}) \frac{e^{jk|\mathbf{r}-\mathbf{x}|}}{|\mathbf{r}-\mathbf{x}|}, \tag{5.3}$$

where $f(\hat{\mathbf{n}})$ is the angular radiation pattern and V is a complex amplitude with units of voltage. We assume that $\mathbf{x}$ is sufficiently close to the coordinate origin that the angular vector $\hat{\mathbf{n}}$ is approximately the same regardless of whether it is referenced to $\mathbf{x}$ or to $(0,0,0)$. Then the electric field radiated by an array of antennas is given by

$$E = \sum_i V_i f_i(\hat{\mathbf{n}}) \frac{e^{jk|\mathbf{r}-\mathbf{x}_i|}}{|\mathbf{r}-\mathbf{x}_i|}. \tag{5.4}$$

This expression is not particularly useful. We have already assumed that the antennas are all close to the origin. By that, of course, we mean that they are close to the origin compared with the distance to the far field. Our approximation is, then, $x_i \ll r$. Then we have

$$E = \frac{e^{jkr}}{r} \sum_i V_i f_i(\hat{\mathbf{n}}) e^{-j\mathbf{k}\cdot\mathbf{x}_i}. \tag{5.5}$$

Remember that $\mathbf{k} = k\hat{\mathbf{n}}$. If all of the antennas are identical, then we can factor out the angular pattern to get

$$\boxed{E = \frac{e^{jkr}}{r} f(\hat{\mathbf{n}}) \mathrm{AF}(\hat{\mathbf{n}}),} \tag{5.6}$$

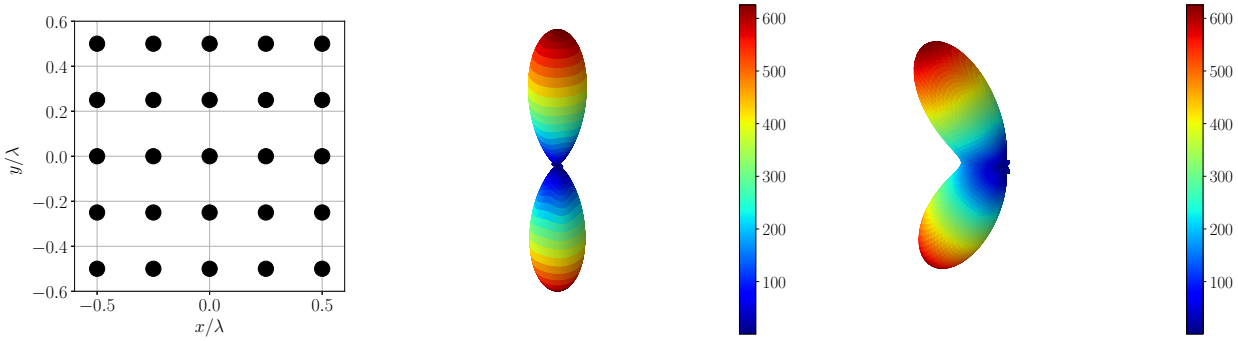
where

$$\boxed{\mathrm{AF}(\hat{\mathbf{n}}) = \sum_i V_i e^{-j\mathbf{k}\cdot\mathbf{x}_i}} \tag{5.7}$$

is called the *array factor*. The array factor is the angular radiation pattern of a set of isotropic radiators, i.e., $f(\hat{\mathbf{n}}) = \text{const}$. An example is given in Fig. 5.4 for a square array. When there are no phase delays, the array factor produces symmetric beams pointing toward the zenith and nadir. When we introduce phase delays δ_x, the phase delay between neighboring columns, and δ_y, the phase delay between neighboring rows, the beams steer off-zenith. These phase delays are useful for interferometric experiments, but we will always choose to have no delays, so that our beams point directly toward the zenith. For our devices, the phases of the antenna elements are set by the microstrip lengths between the bolometers and the antennas. To avoid any phase delays, we simply ensure that these microstrip lengths are all equal.

From Eq. 5.6, the beam produced by an antenna array is proportional to the individual element's beam multiplied by the array factor. In this way, we can remove, for instance, the back-lobe response of the beam shown in Fig. 5.4(b): we simply choose an antenna element that has a negligible back lobe to begin with.

When the array spacing is larger than $\lambda/2$, the array factor contains additional local maxima, which are sometimes called *grating lobes*. An example is shown in Fig. 5.5, which



(a) Configuration (b) No phase delays (c) $\delta_x = 45°$, $\delta_y = 15°$

Figure 5.4: A 5×5 rectangular array with equal amplitudes.

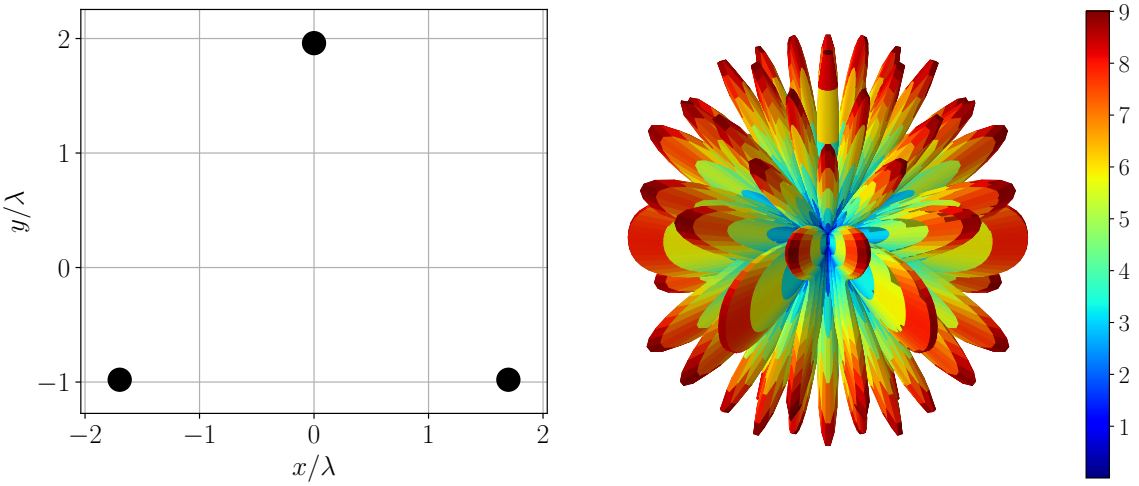
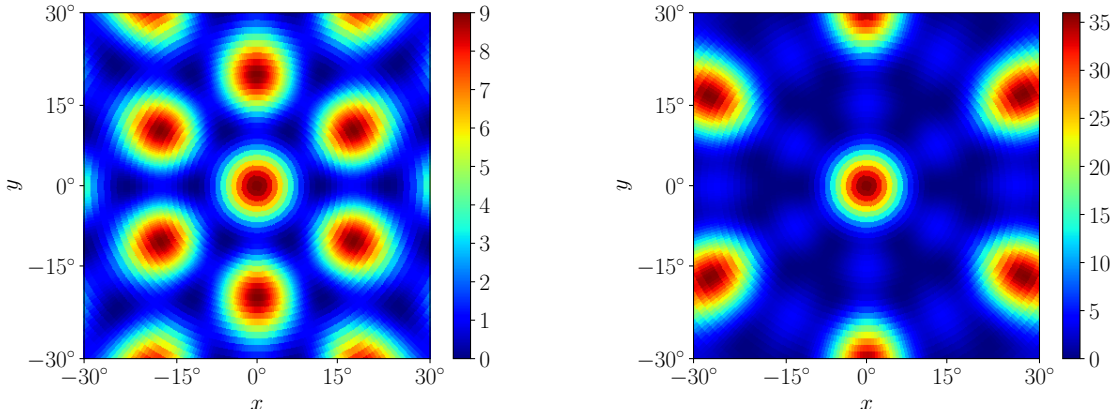


Figure 5.5: A 3-cell 150-GHz array with 6.8-mm spacing.



(a) 3-cell 150 GHz

(b) 6-cell 90 GHz

Figure 5.6: Array factors for triangular arrays with 6.8-mm spacing.

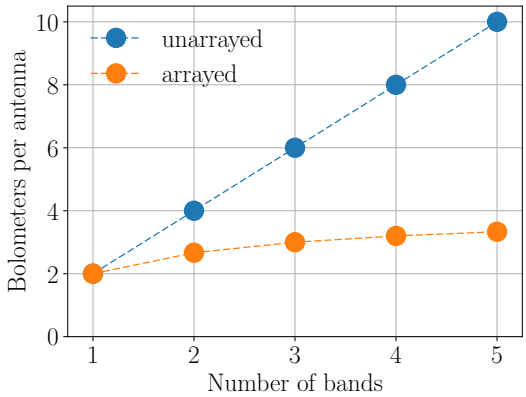
is the array factor for the 150-GHz array in the prototype to be discussed in Sec. 5.3. We see that the array factor contains a large number of grating lobes. This is, however, not a problem, so long as the beam width of the antenna element is smaller than lobe spacing. The grating lobes will be suppressed by the element beam, and the element beam will be narrowed by the array factor. Simulated beam maps of the array factors for the prototype of Sec. 5.3 are shown in Fig. 5.6.[3] Notice that the lobe spacing is significantly larger for the 90-GHz array compared with the 150-GHz array.

5.2.2 Readout savings

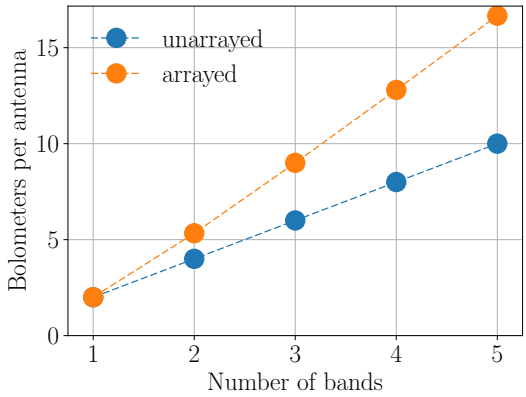
An additional benefit of a hierarchical focal plane is that the detector count per unit area may be reduced relative to a conventional design, even though the sensitivity may be improved. This is possible when the conventional design is optimized for the highest frequencies. In this case, the hierarchical design simply combines antennas at lower frequencies, which reduces the number of detectors per antenna. Whereas the conventional design would have a detector count that increases approximately linearly with the number of frequency bands, the hierarchical design has a detector count that increases approximately logarithmically as shown in Fig. 5.7(a). Under these conditions, a three-band experiment could reduce its detector count by half.

This paradigm is particularly attractive for satellite- and balloon-based experiments, for which increased weight and power are expensive. Additionally, the mapping speed tends to have a narrower optimum due to the larger optical loading from the aperture stop relative to the sky in the absence of atmospheric loading. For ground-based experiments, the reduction in system complexity, coming, e.g., from decreased power consumption and wire count for readout electronics, is also attractive.

[3]For a copy of the software that produces graphical representations of array factors such as those shown in Figs. 5.4, 5.5 and 5.6, email `ajcukierman@berkeley.edu`.



(a) The unarrayed pixel spacing is optimized for the highest frequency. In this case, there are significant readout savings to be gained from the hierarchical array.



(b) The unarrayed pixel spacing is optimized for the geometric mean of the frequency range. In this case, the number of readout channels is similar for the two cases and, in fact, a little larger for the hierarchical array.

Figure 5.7: Comparison of the number of readout channels per antenna for arrayed and unarrayed focal planes. We assume that the arrays are triangular, that each band corresponds to a different array size, such that the nth band has n antennas at the base of its triangle, and that the band centers scale inversely with the effective size of the arrays. There is freedom in the *unarrayed* design to optimize the pixel spacing for a certain frequency band. Depending on this choice, the number of bolometers per antenna may increase or decrease when the focal plane is redesigned with hierarchical phased arrays.

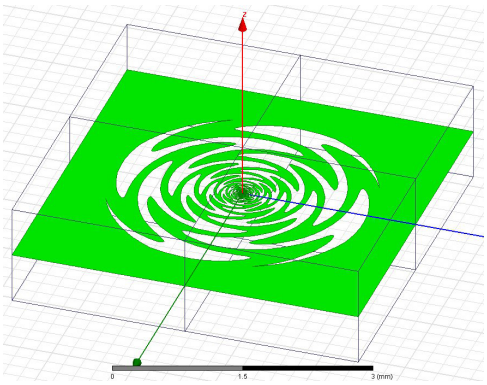
The subtlety in the discussion of readout savings is that the pixel spacing of a conventional multichroic focal plane is not uniquely defined, even when the band centers are set. Because the conventional design can only be optimal in at most one frequency band, there is freedom to choose which band is to be the optimal one if any. In discussing the potential readout savings from a hierarchical array, we need to be clear about what we are comparing to. In the above discussion, we assumed that the conventional design was optimized for the highest frequency, in which case the comparison is shown in Fig. 5.7(a), where we see that the hierarchical array provides a significant reduction in the number of readout channels. If, however, the conventional design is optimized for, e.g., the geometric mean of the frequency range, then the comparison is shown in Fig. 5.7(b), where we see that the number of readout channels is similar for the two cases and even a little larger for the hierarchical array. Of course, these are rough numbers that rely on several assumptions, so the conclusion here should be that the number of readout channels is approximately the same when the conventional array is optimized for the geometric mean of the frequency range. If the conventional array is optimized for the *lowest* frequency, then the number of readout channels is dramatically larger for the hierarchical array. This is because the conventional array had oversized pixels at high frequencies, and the hierarchical array now populates that area with a large number of tiny pixels.

A plausible frequency schedule that would yield readout savings if implemented as a hierarchical array is one that covers, e.g., 20-120 GHz. Suppose we have band centers at 30, 40, 60 and 90 GHz, where the 60-GHz band is, of course, only useful above the Earth's atmosphere. Then a conventional focal plane might optimize the pixel spacing for the higher frequencies, since they are less contaminated by galactic foregrounds. When we redesign with a hierarchical array, we will simply reduce the number of low-frequency bolometers and, therefore, enjoy a reduction in the total number of readout channels.

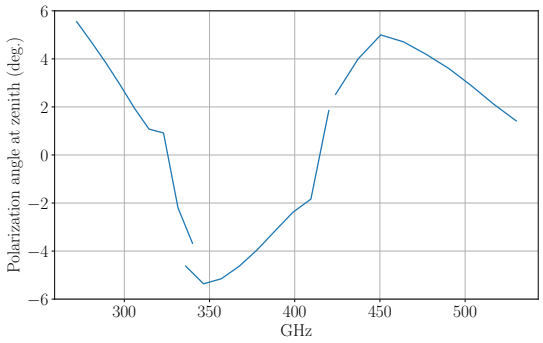
5.2.3 Polarization- and ellipticity-wobble cancellation

A hierarchical focal plane can also mitigate systematic errors related to beam shape and polarization purity. The far-field beam is the product of the array factor and the beam of the individual antenna element. Beam non-idealities can be compensated for by choosing an appropriate array configuration. Due to this compensating effect, we are able to undersize the sinuous antenna by nearly a factor of two. In this case, the individual element's beam may show some distortion [98], but the array factor will round out these imperfections.

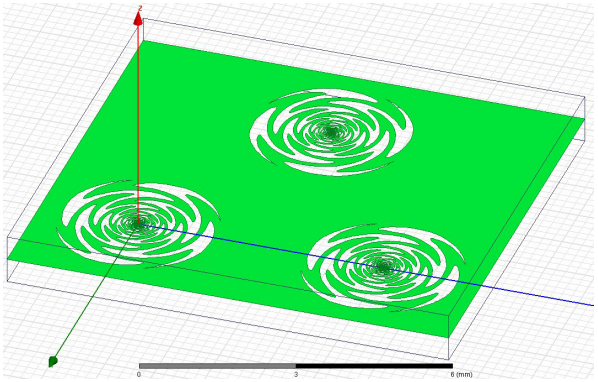
Some antennas, including the sinuous antenna discussed below, have a polarization response that varies or "wobbles" with frequency (see Sec. 4.3.1). By constructing an array consisting partially of mirror-imaged versions of the antenna element, the polarization wobble can be reduced or even cancelled completely. In Fig. 5.8, we show a simulated example of this mitigation strategy. In this example, we use sinuous antennas *without* lenslets in order to decrease the simulation time. We see that the three-cell array reduces the amplitude of the polarization wobble at zenith by nearly a factor of three. For an even-numbered array, the wobble can be cancelled completely. A rough estimate of the wobble amplitude of the polarization at zenith is shown in Fig. 5.9 as a function of the number of elements in the array. For even-numbered arrays, the wobble can be cancelled completely (at the zenith). For odd-numbered arrays, the wobble cannot be completely cancelled, but it can be diluted.



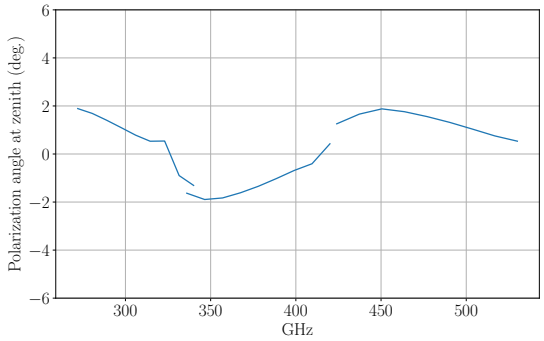
(a) HFSS model of a single sinuous antenna.



(b) The polarization angle at zenith for a single sinuous antenna.



(c) HFSS model of a three-cell array of sinuous antennas, where the antenna in the upper right has the opposite handedness relative to the other two. This antenna is also fed 180° out of phase relative to the others.



(d) The polarization angle at zenith for a three-cell array, where one antenna has the opposite handedness relative to the other two.

Figure 5.8: Simulated demonstration of polarization-wobble cancellation for sinuous-antenna phased arrays. To accelerate the simulations, we do *not* include lenslets. The absence of dielectrics increases the resonant frequencies associated with each cell of the sinuous antenna, so we simulate in the range 270-530 GHz. The three-cell array reduces the amplitude of the polarization wobble from approximately 5° to approximately 2°. An even-numbered array would cancel the wobble completely, which can be argued from symmetry.

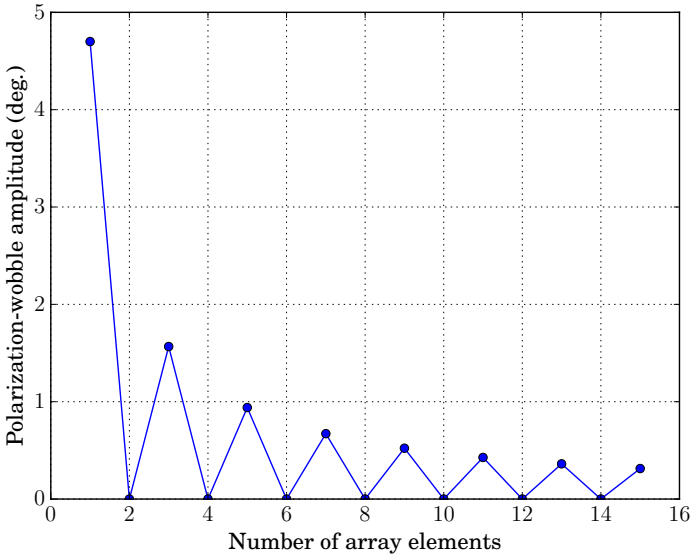


Figure 5.9: Polarization-wobble amplitude as a function of the number of elements in the array.

5.2.4 Limitations

In this implementation of a hierarchical phased array, each antenna element is, in principle, uncoupled to its neighboring antennas in the sense that its performance is approximately independent of the presence of these neighboring antennas. This is to be contrasted with antenna arrays in which the mutual coupling of neighboring antennas is tolerable or even necessary, e.g., Vivaldi arrays and current-sheet arrays [11, 50]. We require the individual element to operate across the entire bandwidth. Since most antennas experience significant degradation when the antenna size is reduced to become comparable to a wavelength, there is a minimum size of the individual element given a design bandwidth. This minimum size is driven by the lowest frequency (largest wavelength). At the same time, however, the sensitivity optimization may pick a pixel spacing for the highest frequencies that is smaller than this minimum size. This is especially likely to be true when the temperature of the aperture stop is much lower than that of the sky; in this case, the sensitivity is optimized by having many small pixels with very broad beams, since there is only an efficiency hit but no noise penalty for spilling over onto the aperture stop. In such a scenario, it will be difficult to implement a hierarchical phased array in the manner described here, since it would be necessary to build antenna elements with diameters comparable to or perhaps even smaller than a wavelength at the lowest frequencies. At these small pixel sizes, the performance of each individual antenna is severely degraded as shown in Sec. 4.4.2.

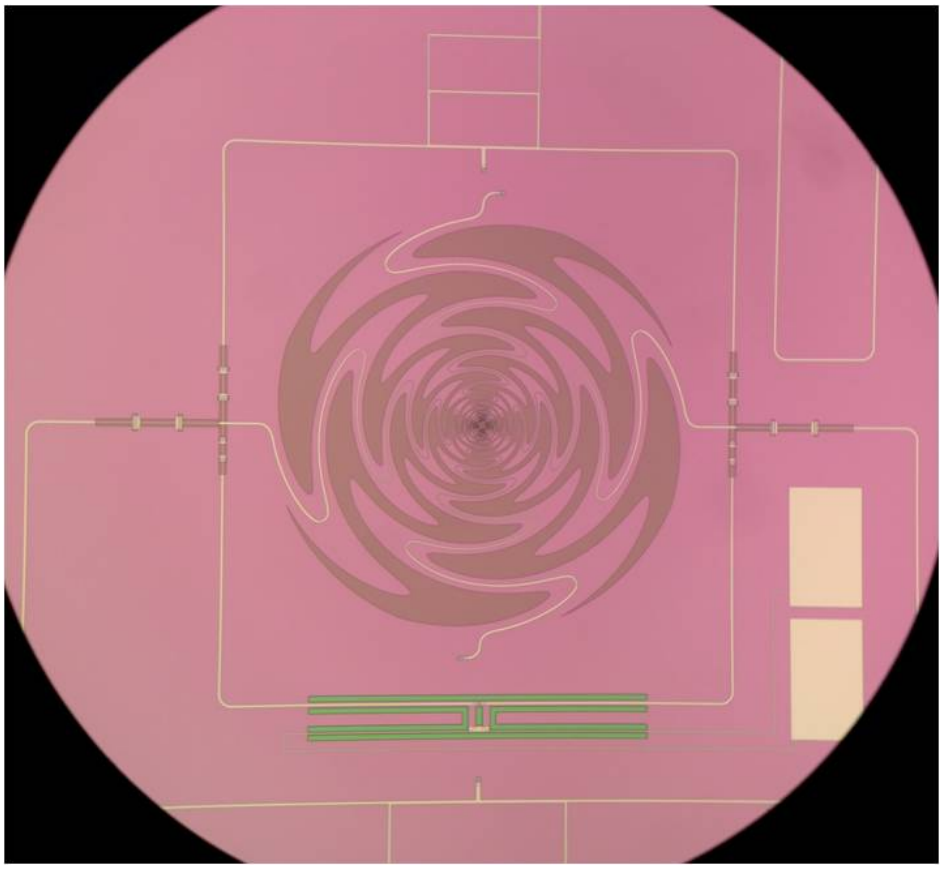


Figure 5.10: Unit cell of the prototype hierarchical array: a sinuous antenna coupled to out-of-phase microstrip lines, triplexing bandpass filters (*middle, left and right*) and double-ring hybrids (*top and bottom*). The 220-GHz signal is sent directly to a bolometer (*bottom center*). For simplicity of fabrication, only one polarization is active in this prototype.

5.3 Prototype design

We have developed a multiscale architecture based on hierarchical phased arrays that allows for dual-polarized wideband sensitivity at millimeter wavelengths.

5.3.1 Unit cell

As the unit cell of these arrays, we choose the lenslet-coupled sinuous antenna on account of its broad bandwidth, frequency-independent beam waist and dual-polarized capability.[31, 68, 103] Since the beam waist is determined mainly by the lens diameter, the on-chip planar antenna need only occupy a fraction of the total pixel area, which makes space for microwave circuitry and detectors. The on-chip components of the unit cell are shown in Fig. 5.10. We find that the beam forming of the antenna arrays compensates for beam non-idealities in the unit cell, so the sinuous antenna can be substantially undersized relative to the case of a conventional multichroic pixel; here we use a 1.7-mm diameter (cf. 3-mm for a conventional

pixel operating at the same frequencies[15]).

To simplify and accelerate the fabrication process for this prototype, we chose to avoid microstrip crossovers, since this would have added extra layers. This prototype is, consequently, a single-polarization device, although the architecture can be extended, by adding well-proven microstrip crossovers, to two polarizations. Separately, we have been developing microstrip cross*under*s, which perform the same function as crossovers but without requiring additional layers. We plan to make use of these crossunders in future hierarchical devices.

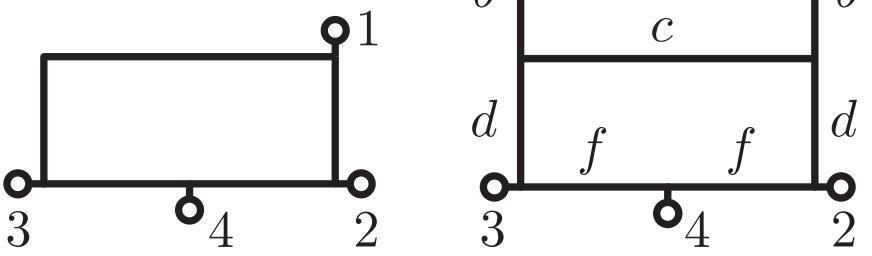
The sinuous antenna is coupled to two out-of-phase microstrip lines that route the signal to triplexing lumped-element bandpass filters, which split the signal into bands centered on 90, 150 and 220 GHz with bandwidths of approximately 25%. The 220-GHz signal is brought directly to a bolometer, where the microstrip lines are terminated differentially on a lumped resistor (second-from-right image in Fig. 5.17). The out-of-phase lines for 90 and 150 GHz are combined in double-ring hybrids to form single microstrip lines, which enter the summing networks.

5.3.2 180° hybrid

A 180° hybrid splits a transmission-line signal into two out-of-phase components, usually with equal amplitudes. Alternatively, the hybrid can combine two out-of-phase signals into one transmission line. This is useful for a differentially-fed antenna such as the sinuous antenna, which requires two out-of-phase transmission lines for the excitation. To simplify the microwave circuitry away from the antenna, it is sometimes useful to use a 180° hybrid to cut down the number of transmission lines traveling between the antennas and the detectors.

A common implementation of a 180° hybrid is in the form of the so-called *ring hybrid*, which is often also called a *rat-race coupler*. The conventional ring hybrid is shown in Fig. 5.11(a). Imagine injecting two signals at ports 2 and 3. The two signals traverse an equal electrical length to get to port 4, so port 4 is often called the *sum port*. On the other hand, the signal from port 3 will arrive at port 1 with a 180° phase shift relative to the signal from port 2, so port 1 is often called the *difference port*. A signal injected at port 1 will have a node at port 4 due to the 180° phase difference between the two paths available from port 1 to 4. But this signal will experience constructive experience at both ports 2 and 3 with the phase at port 3 exactly 180° off of the phase at port 2. In this way, the hybrid has split the incoming signal into two equal-amplitude but out-of-phase components. Reversing time, we can inject out-of-phase signals at ports 2 and 3 to get a combined signal at port 1. The impedance of port 4 is important in determining the coupling between ports 2 and 3. We want S_{23} to be small in case there are spurious reflections on the transmission lines that cause signals to bounce around the circuit. The performance of a conventional ring hybrid is shown as dashed lines in Fig. 5.12.

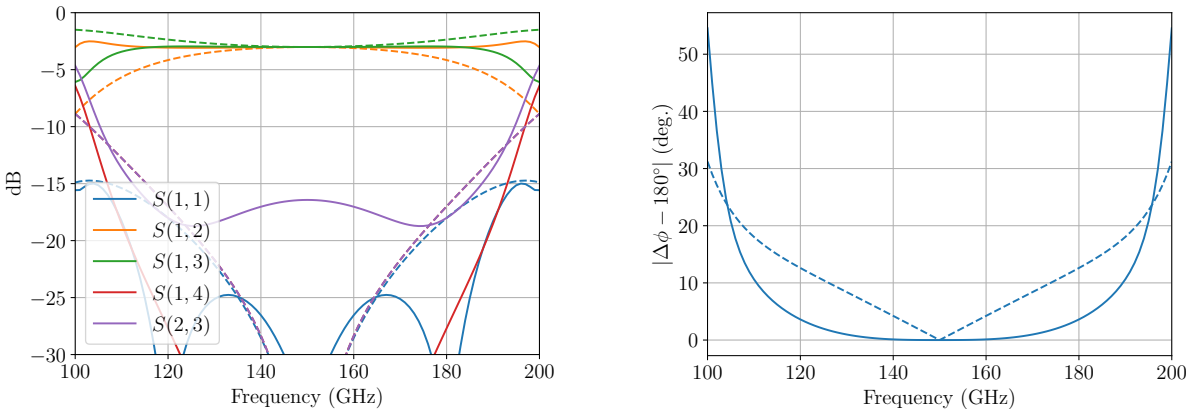
We can extend the bandwidth of the ring hybrid by adding an additional ring. This is shown in Fig. 5.11(b). The impedances of the transmission-line segments can be quickly tuned in a program like Keysight's Advanced Design System (ADS). The parameters shown in the diagram are dimensionless impedances in units of Z_0, the impedance of ports 1, 2 and 3. For example, the impedance of the segment labeled a is aZ_0. We leave the port impedance of port 4 as a free parameter ($Z_{0,4} = gZ_0$), which controls the crosstalk between ports 2 and 3. The performance of this *double-ring hybrid* is shown as solid lines in Fig. 5.12.



(a) Conventional ring hybrid. All port impedances are Z_0. All interior segments have impedance $\sqrt{2}Z_0$.

(b) Double-ring hybrid. The impedances of the interior segments are parameterized by the Latin characters, which give the impedance of each segment in units of Z_0. It is also useful to leave the impedance of port 4 as a free parameter $Z_{0,4} = gZ_0$. All other ports have impedance Z_0.

Figure 5.11: Ring-hybrid circuit diagrams. The shorter segments are $\lambda/4$ in length. The longer segments are $\lambda/2$. Port 1 is the difference port, and port 4 is the sum port. A signal injected at port 1 is split into two equal-amplitude but out-of-phase signals exiting the hybrid at ports 2 and 3. Ports 2 and 3 are isolated by tuning the impedance of port 4.



(a) S parameters. We design for $S_{12} = S_{13} = -3$ dB and S_{11} and S_{23} as low as possible.

(b) Phase difference of the output ports (2 and 3) relative to 180°. We design for this quantity to be as close to 0° as possible, i.e., $\Delta\phi = 180°$.

Figure 5.12: Performance of ring hybrids as simulated in ADS. The dashed lines correspond to a conventional ring hybrid, and the solid lines correspond to a double-ring hybrid. We have tuned the double-ring parameters to $a = 1.26$, $b = 1.70$, $c = 1.14$, $d = 1.17$, $f = 0.85$, $g = 0.71$ ($Z_{0,4} = gZ_0$). The double-ring version acts as a 180° hybrid over a broader range of frequencies.

We see that the device operates as a 180° hybrid over a broader range of frequencies than the conventional ring hybrid.

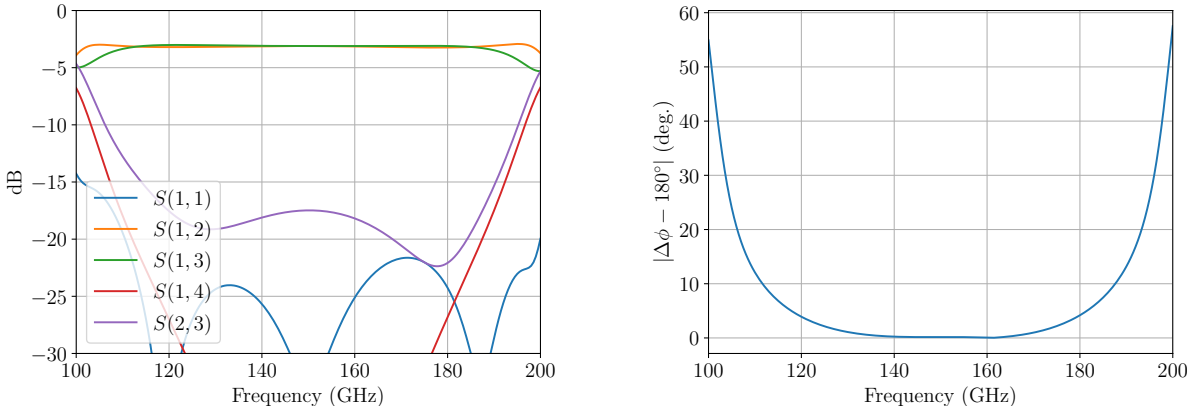
What if we add more rings? Unfortunately, it turns out that cascading rings quickly reaches diminishing returns in extending the bandwidth of these hybrids. It is basically not worth it to even go to a triple-ring hybrid.

A fabricated microstrip realization of a double-ring hybrid is shown in Fig. 5.13. The impedances of the transmission-line segments are set by the microstrip widths. The sum port (port 4) is terminated with a lumped titanium resistor shorted to the ground plane with a via through the silicon-oxide microstrip dielectric. We can simulate the performance of this microstrip realization with Sonnet Software. The results are shown in Fig. 5.14 and are to be compared with the solid lines of Fig. 5.12. The agreement is actually better than it really needs to be to accept this design.

5.3.3 Wilkinson splitter

In the summing network, we must combine microstrip lines from neighboring antennas. For this purpose, we must design 2- and 3-way power splitters, so called because, in the reverse-time sense, they split power. Unlike the ring hybrids discussed above, these splitters must produce identical phases at the outputs. A common power splitter is the so-called *Wilkinson divider*, whose circuit diagram is shown in Fig. 5.15(a). An advantage of these dividers is that the output ports are well-isolated from each other, which suppresses unwanted reflections between antennas.

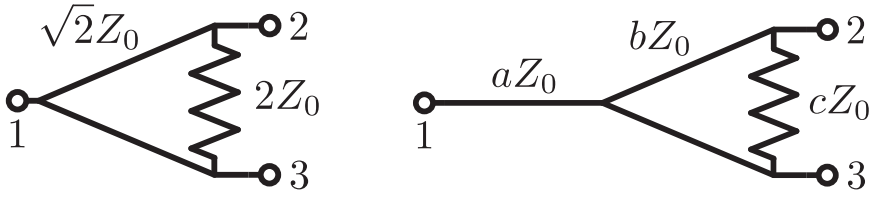
Figure 5.13: Fabricated double-ring hybrid tuned for a 150-GHz band. The orientation of this image is the same as that of the circuit diagram in Fig. 5.11(b). The sum port (port 4) is terminated with a lumped titanium resistor. Ports 2 and 3 feed the sinuous antenna differentially, and port 1 enters the summing network of the hierarchical phased arrays.



(a) S parameters.

(b) Phase difference of the output ports (2 and 3) relative to 180°.

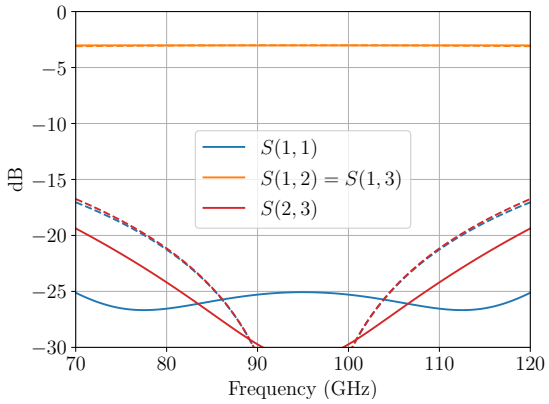
Figure 5.14: Simulated performance of a double-ring hybrid with Sonnet.



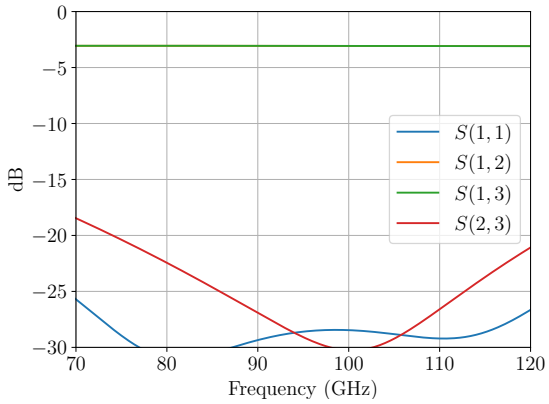
(a) Conventional Wilkinson divider

(b) 2-stage Wilkinson divider

Figure 5.15: Wilkinson-divider circuit schematics. All of the port impedances are Z_0.



(a) S parameters as simulated in ADS. We design for S_{11} and S_{23} to be minimized and $S_{12} = S_{13} = -3$ dB. Solid lines correspond to the two-stage and dashed to the conventional Wilkinson divider.



(b) Sonnet simulation of a microstrip realization of a two-stage Wilkinson power divider.

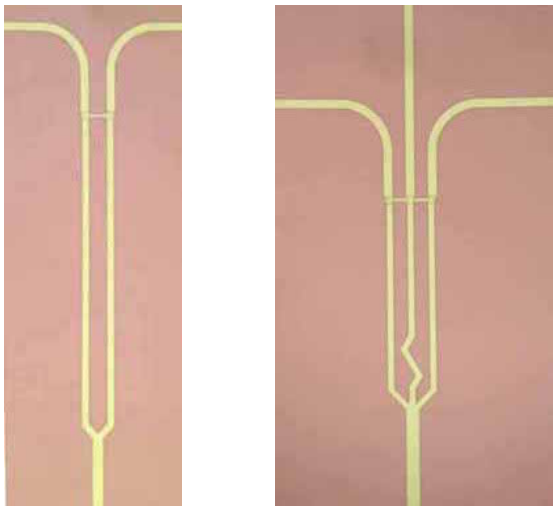
Figure 5.16: Performance of Wilkinson dividers. For the two-stage dividers, we use $a = 0.88$, $b = 1.3$, $c = 2.0$ as defined in Fig. 5.15(b).

We can extend the bandwidth of a Wilkinson divider by adding a quarter-wavelength section as shown in Fig. 5.15(b). We tune the parameters a, b and c in Keysight Advanced Design System (ADS). A comparison in performance to the conventional Wilkinson divider is shown in Fig. 5.16(a), where dashed lines correspond to the conventional and solid lines to the two-stage Wilkinson divider. The improvement in frequency coverage is somewhat modest over the $\sim$ 25% bandwidth we typically design for, but the quarter-wavelength section is a cheap and simple addition. A microstrip realization is shown in Fig. 5.17(a). The performance can be simulated in Sonnet Software and is shown in Fig. 5.16(b). The microstrip version differs only at the percent level from the ideal circuit.

We can generalize the Wilkinson divider to an n-way splitter. A Wilkinson *trivider* is shown in Fig. 5.18(a). This circuit is, however, incompatible with a planar layout due to the resistor connecting ports 2 and 4. We can planarize this circuit by removing this resistor and retuning the impedances of the transmission-line segments and the resistors. We can also broaden the bandwidth by adding a first-stage quarter-wavelength section as we did for the Wilkinson divider. This circuit is shown in Fig. 5.18(b). The performance is shown in Fig. 5.19(a). A microstrip realization of a two-stage Wilkinson trivider is shown in Fig. 5.17(b). To ensure equal electrical lengths along all three paths, we designed a zig-zag section in the middle segment. This microstrip realization can be simulated with Sonnet Software; the results are shown in Fig. 5.19(b).

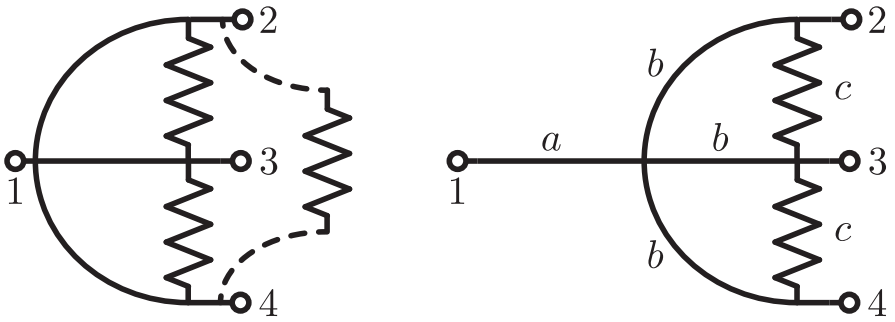
5.3.4 Summing network

After leaving the 180° hybrids, then microstrip lines enter a summing network, in which they combine with the lines from neighboring lenslet-coupled antennas. This summation



(a) Divider. (b) Planarized trivider.

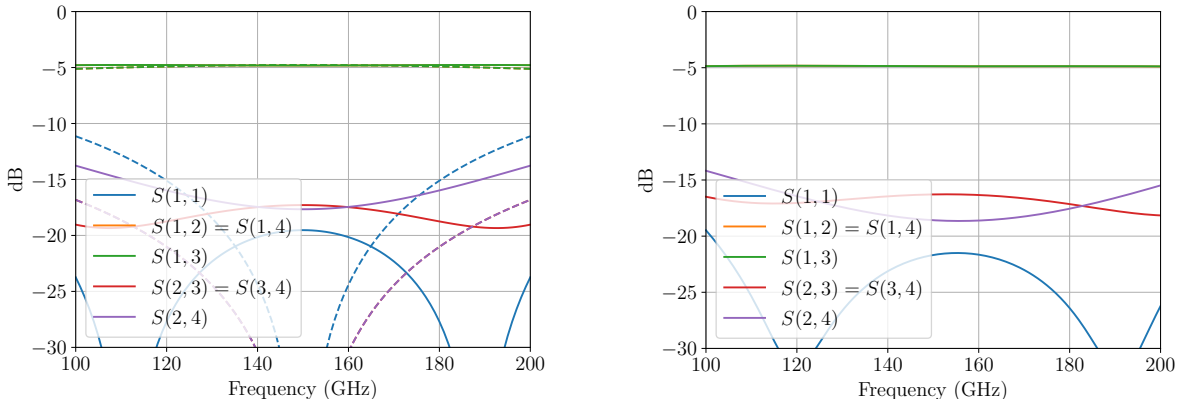
Figure 5.17: Fabricated two-stage Wilkinson splitters consisting of superconducting Nb microstrip lines and Ti resistors (connecting the output lines near the top). The first-stage quarter-wavelength sections are cut off in these images. The zig-zag in the trivider ensures equal electrical lengths along all three paths.



(a) Conventional Wilkinson trivider. A dashed line indicates a short circuit. The long solid segments have impedance $\sqrt{3}Z_0$, and the lumped resistances are $3Z_0$.

(b) Planarized 2-stage Wilkinson trivider. The resistor connecting ports 2 and 4 has been removed, but the impedances can be tuned to compensate.

Figure 5.18: Wilkinson-trivider circuit schematics. The long solid segments are all $\lambda/4$ in length. All port impedances are Z_0. The impedance of the segment labeled a is aZ_0, etc.



(a) S parameters as simulated in ADS. We design for $S_{12} = S_{13} = S_{14} = -4.8$ dB and all other S parameters as low as possible.

(b) Sonnet simulation of a microstrip realization of a two-stage Wilkinson trivider.

Figure 5.19: Simulated performance of Wilkinson trividers: $a = 0.73$, $b = 1.4$, $c = 1.6$.

is performed independently for the 90- and 150-GHz bands. An overview of the summing networks is shown in Fig. 5.20.

The 3-antenna 150-GHz array is formed by combining the microstrip lines from three neighboring antennas with a Wilkinson trivider; this combined signal is then sent to a bolometer. The 6-antenna 90-GHz array is formed by combining pairs of antennas with Wilkinson dividers and then bringing these three combined signals together in a Wilkinson trivider; this signal is then sent to a bolometer. The microstrip lengths between the antennas and the splitters are identical within each frequency band to avoid phase delays that steer the beam off-zenith. The trombone-like microstrip sections are included to ensure equal path lengths.

The summed signal is terminated on a lumped resistor on a bolometer island, where the signal is detected by an AlMn TES with a critical temperature of approximately 450 mK. The fabrication process and bolometer design are similar to those developed for the POLARBEAR-2 (PB-2) detector wafers [100, 103].

For the microstrip ground plane and strip layer, we use superconducting Nb. As a microstrip dielectric, we used SiO for this prototype, but we intend to use SiN for future devices, since it shows ~ 10 times lower loss and has been successfully implemented in the wafers recently fabricated for PB-2.

5.4 Prototype measurements

In this section, we report on the measured performance of the prototype sinuous-antenna phased array. We used a conventional trichroic sinuous-antenna pixel as a control.

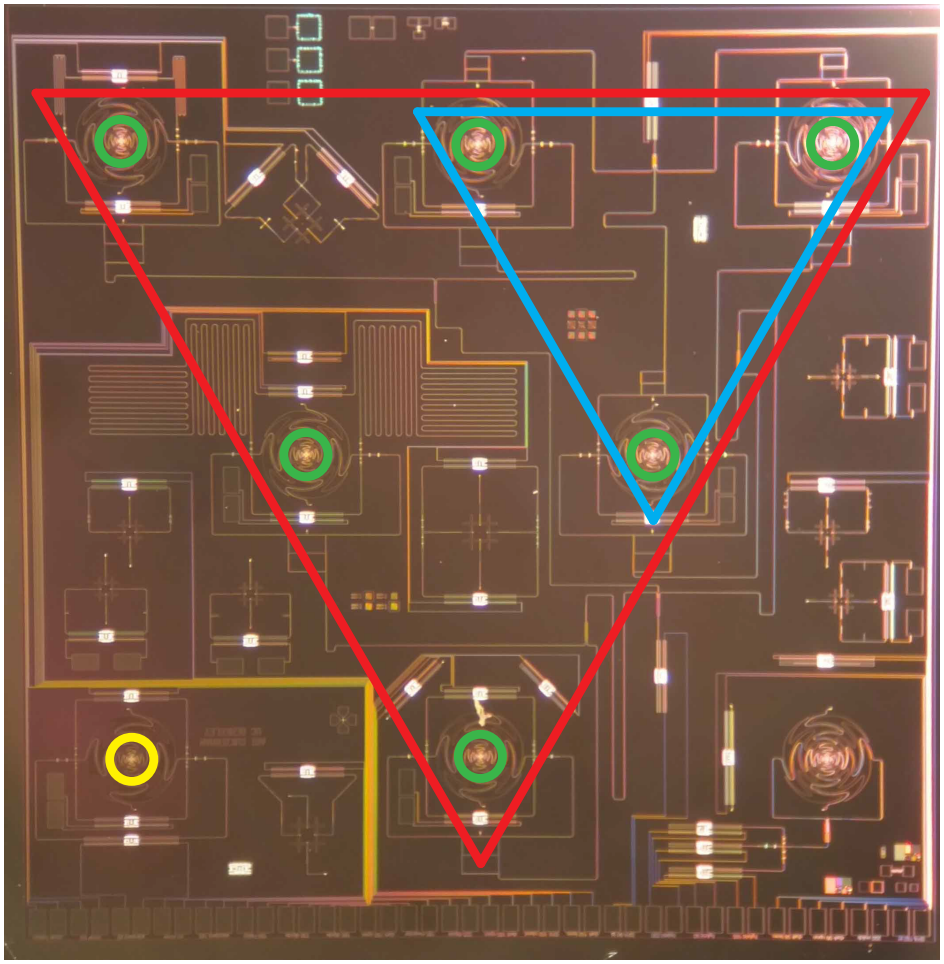
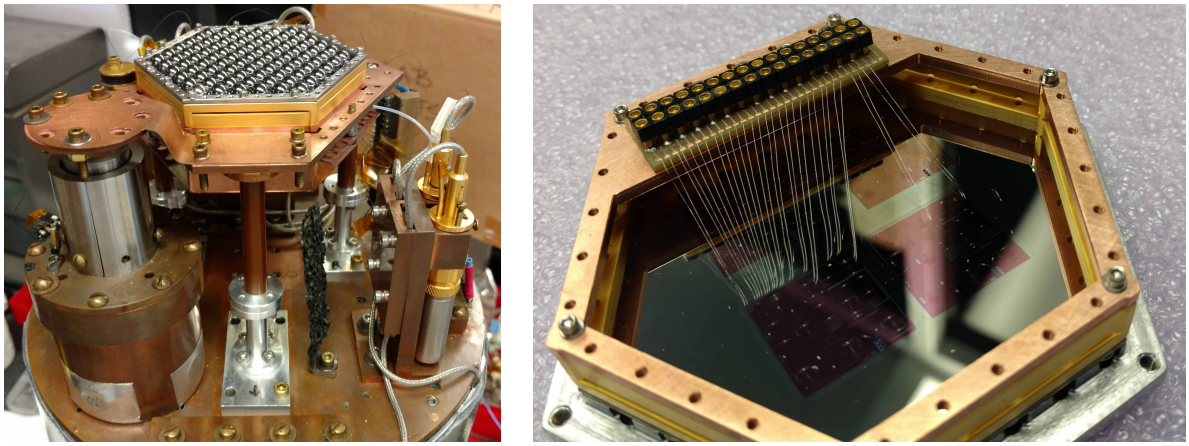


Figure 5.20: Fabricated prototype chip with overlaid hierarchical structure: 220-GHz pixels (*green*), 150-GHz array (*blue*) and 90-GHz array (*red*). A conventional trichroic pixel (*yellow*) is included as a control. The other features on the chip are for test pixels that were included in the same mask design but are electrically independent of the hierarchical array. For example, the sinuous-antenna pixel in the bottom right uses a prototype broadband 180° hybrid. The double-slot-dipole pixel near the center uses a prototype diplexer that splits the 150-GHz atmospheric window.



(a) The wafer was installed in an optical wet dewar and cooled to 270 mK with a ^{3}He sorption refrigerator (*left*).

(b) "Heroic" wire bonds are made directly from the devices to a printed-circuit board.

Figure 5.21: Test setup for the prototype sinuous-antenna phased array. The device wafer was mated to a PB-1 lenslet array using PB-1 invar-holder parts.

5.4.1 Test setup

The wafer was aligned to a PB-1 lenslet array [84] using PB-1 invar-holder parts and installed in an optical test cryostat, which cools the detector wafer to 270 mK using a ^{3}He sorption refrigerator (see Sec. 3.4.2). The test setup is shown in Fig. 5.21. The bolometers are biased with a DC voltage supply and read out with DC superconducting quantum interference devices (SQUIDs).

5.4.2 Spectra

The frequency response was measured with a Fourier-transform spectrometer (see Sec. 3.4.3). The results (Fig. 5.22) are similar for the conventional and hierarchical pixels, showing that the hierarchical design is not introducing a spurious frequency response. The fringes in the 220-GHz channel are not understood but could be due to an impedance mismatch somewhere in the optical chain, i.e., between the lenslet-air interface and the bolometer. Relative to the design, the measured bands shifted down in frequency by 5-10%. This is common for a first prototype and is likely due to an incorrect dielectric constant or dielectric thickness in simulation. The shift can be tuned away by varying the dielectric thickness on subsequent iterations. The simulated bands shown in Fig. 5.22 use a 15% thinner microstrip dielectric than originally intended. A similar shift can be achieved by increasing the dielectric constnat by 15%.

5.4.3 Beams

The beams were measured using a chopped liquid-nitrogen source on an XY stage (see Sec. 3.4.4). The results are shown in Fig. 5.23 for all three frequency bands. The array has

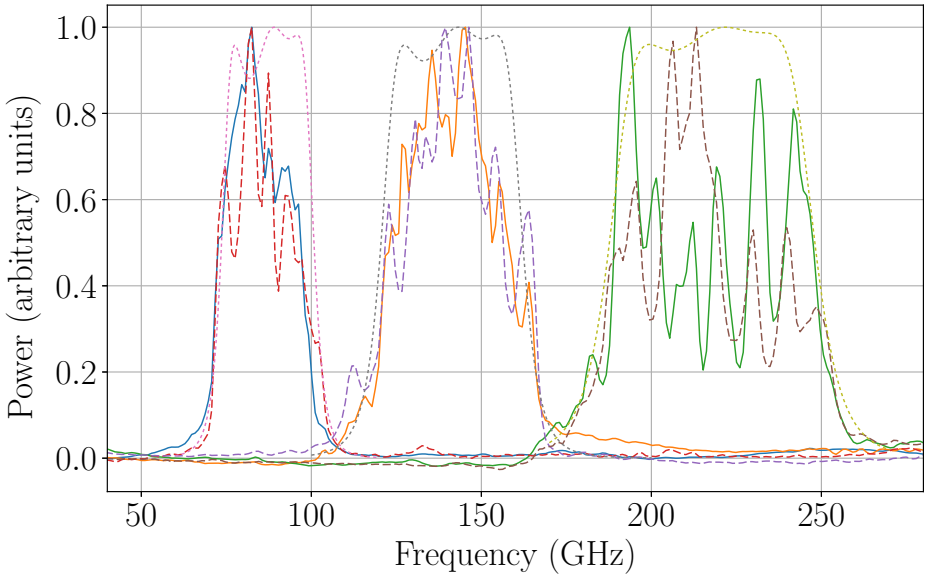


Figure 5.22: The frequency response shows bands centered at 90, 150 and 220 GHz for both the conventional trichroic pixel (*dashed*) and the hierarchical array (*solid*), which show similar band edges to the simulated bandpass filters (*dotted*). The spectra have been peak-normalized, since we were unable to make reliable measurements of the optical efficiency (see text for details).

created round beams of approximately the same size at all frequencies with widths that are consistent with expectations from EM simulations. The faint hexagonal pattern is expected from the array factor of triangular antenna arrays (see Sec. 5.2.1). The beam profiles are shown in Fig. 5.24, where we see that the hierarchical design has created approximately constant beam widths across the 90, 150 and 220-GHz bands. The 220-GHz beam is a bit broader than desired, and this is not understood. A possible explanation is that the anti-reflection coating on the PB-1 lenslets was delaminating and creating scattering sites, which would affect the higher-frequency channels more than the lower-frequency channels. This explanation is made more plausible by the observation that several lenslets were delaminated when the dewar was warmed up and reopened. This is not necessarily a relevant discovery for PB-1, since the lenslet arrays I used had been sitting in unpurged air for several years unlike the barely aged lenslets that have been sitting at sub-Kelvin temperatures for most of the past six years.

The centers of the beams are found by fitting to a two-dimensional Gaussian, although the fits were not good at some frequencies. Several beams have a peak that rises above the Gaussian. As the beams are reasonably round, the Gaussian still does a good job of finding the beam centers. It is not a requirement for CMB experiments that the pixel beams are Gaussian, although it would be convenient; it is much more important that the beams are round. By knowing the distance from the focal plane to the beam-mapper plane, we can convert the x- and y-axes of the beam maps to angular coordinates as shown in Fig. 5.23.

The beam centers in Fig. 5.23 are not co-located. There are three reasons for this. First,

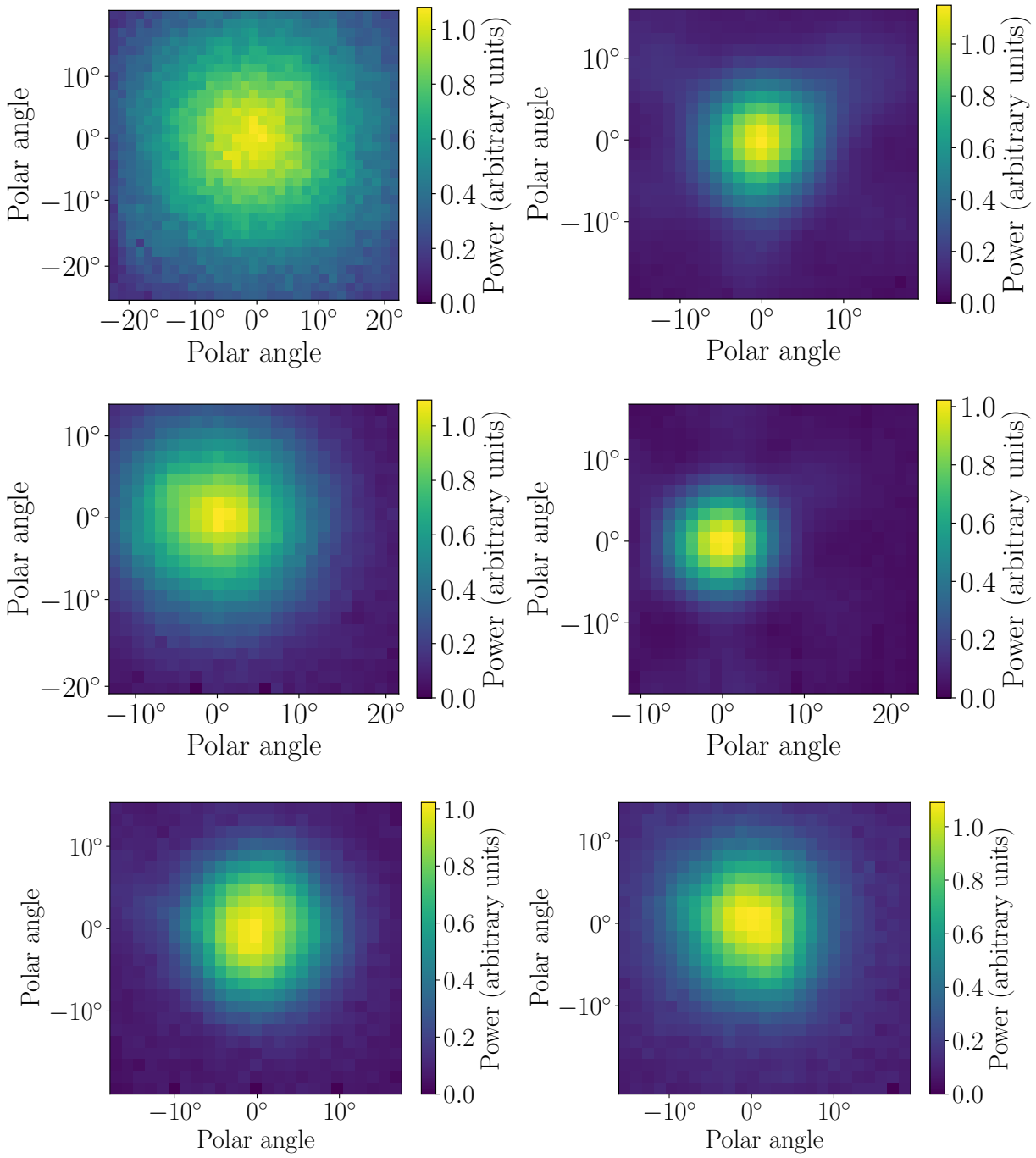


Figure 5.23: Beam maps for 90 GHz (*top row*), 150 GHz (*middle row*) and 220 GHz (*bottom row*). The conventional trichroic pixel (*left column*) shows frequency-dependent beam widths. The hierarchical array (*right column*) shows roughly constant beam widths.

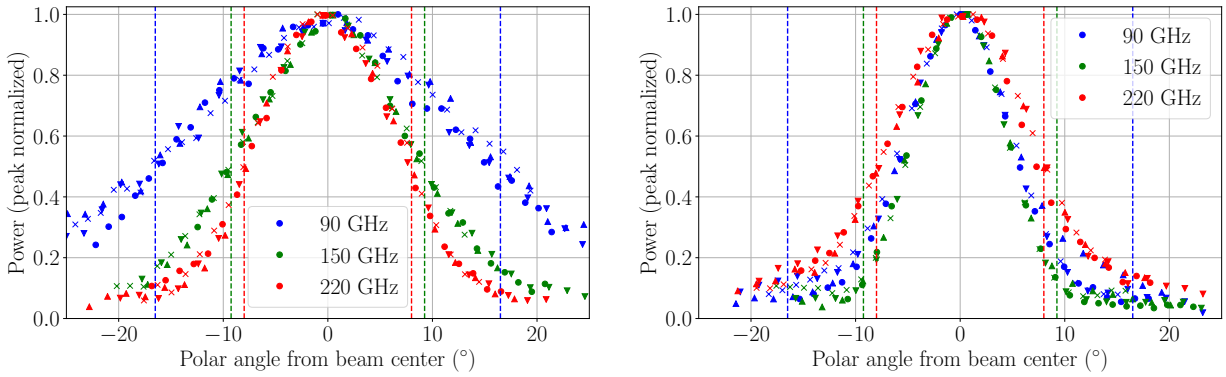


Figure 5.24: Beam profiles for the conventional trichroic pixel (*left*) and the hierarchical array (*right*). The different markers correspond to different azimuthal angles, i.e., 0°, 45°, 90° and 135°. To guide the eye, the vertical *dashed* lines indicate the half maxima for the conventional pixel. We see that the beam profiles vary much less for the hierarchical phased array over the approximately 3 : 1 bandwidth.

the actual coordinates of the raster scans are not the same for all of these beam maps. Second, the hierarchical phased array has different phase centers for different frequencies, so the beams are not even expected to be co-centered. Third, the bolometers used for these measurements did not all come from the same part of the wafer due to the fact that several more conveniently located bolometers were open-circuited by electrostatic discharge during the assembly process.[4] In the coordinates of the beam mapper, all of the measured devices have beam centers relative to each that are consistent with expectations to better than 1°.

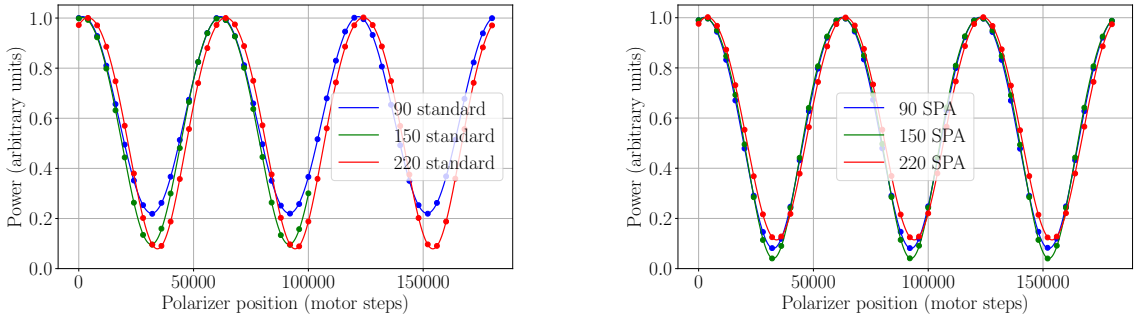
We plot the beam profiles in Fig. 5.24. The different marker styles correspond to different azimuthal cuts. That these markers all appear to be drawn from the same distribution is an indication of the roundness of the beam.

5.4.4 Optical efficiency

We were unable to make a clean measurement of optical efficiency, since the optical loading from the room-temperature equipment exceeded the saturation power of the TESs. All of the measurements presented were performed, therefore, above the superconducting transition, where there is still a non-zero dR/dT that creates a reasonable optical responsivity. Non-linearities in the detector response were removed in post-processing.

For a second device that was fabricated alongside the first, we heated the wafer to 225° C before releasing it from the clean room. This raises the critical temperature of the AlMn and, therefore, the saturation power of the TESs. This allowed for the IV-curve method of estimating optical efficiency (see Sec. 3.4.6). The spectra, however, were somewhat distorted relative to the first wafer, and several lenslets were observed to be damaged when the dewar was warmed up and reopened. In fact, every device tested in this cooldown was affected by damaged lenslets. The damage was mainly in the form of delaminated anti-reflection coating

[4]Dear future graduate students, do not assemble wafers on a sheet of foam.



(a) Conventional trichroic pixel.

(b) Prototype sinuous phased array.

Figure 5.25: Polarization measurements from the test chip containing the prototype sinuous phased array. The results show a large cross-polarized response, which is discussed in the text. The solid lines are fits that show the bulk of the polarization leakage is most likely on the focal plane.

and cracked silicon lenslets. The beams were similar to those of the first run, though. The distortion in the spectra could also be due to the heating of the wafer, which was done in open air on a hot plate in the Marvell Nanolab and may have encouraged chemical reactions or structural changes in the on-chip materials. For these reasons, we take the optical efficiencies measured in this cooldown to be lower bounds on the optical efficiencies of the devices from the first cooldown.

We place conservative lower bounds on the optical efficiency relative to top-hat bands of fractional width 25% centered on 90, 150 and 220 GHz. For these estimates, we have divided out an estimated 70% efficiency hit due to the dewar optics; there is a large uncertainty in this number. We estimate $>$ 50% optical efficiency at 90 GHz, $>$ 30% at 150 GHz and $>$ 25% at 220 GHz. Future efforts should include test features that can be used to identify the source of the efficiency hits. Using SiN instead of SiO should make an improvement: the PB-2 optical efficiencies increased by nearly a factor of two when this change was made.

5.4.5 Polarization

We measured the polarization response of the prototype detectors with a wire-grid polarizer, and the results including fits (see Sec. 3.4.5) are shown in Fig. 5.25. All of the devices, including the conventional trichroic pixel, showed a spurious cross-polarized response at the $\sim$ 10% level. It is not clear what this is due to. It could be from direct stimulation of the bolometer island. This is, however, unlikely, since the spectra showed no spurious out-of-band features. The PB-2 and SPT-3G pixels, which are extremely similar to this prototype, show no such spurious cross-polarized response. The problem is, therefore, probably not fundamental to the design of a lenslet-coupled sinuous antenna. For future devices, it will be important to understand this better and possibly to demonstrate the cancellation of polarization wobble (see Sec. 5.2.3).

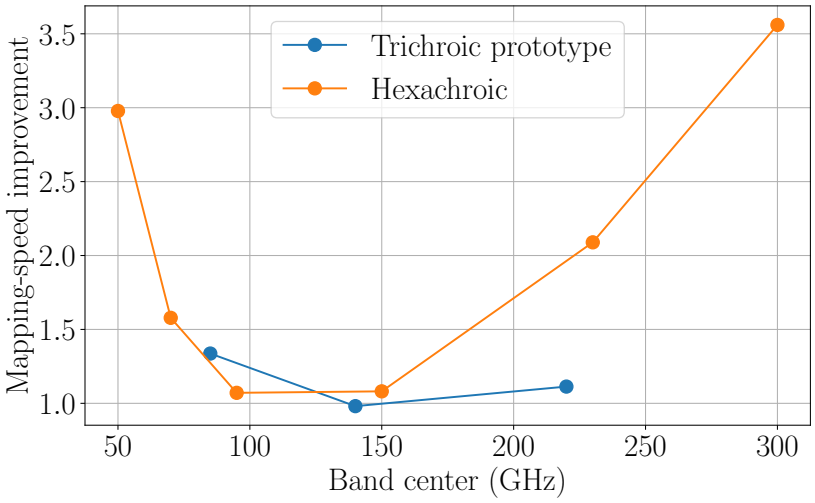


Figure 5.26: Mapping-speed improvement factor from hierarchical phased arrays for the prototype described above, where the effective pixel size has been chosen to be optimal at 120 GHz, and for the ultrawideband design described in Fig. 5.2. The improvement becomes more significant for larger total bandwidths. The prototype produces the relatively modest improvements of 10% at 220 and 30% at 90 GHz.

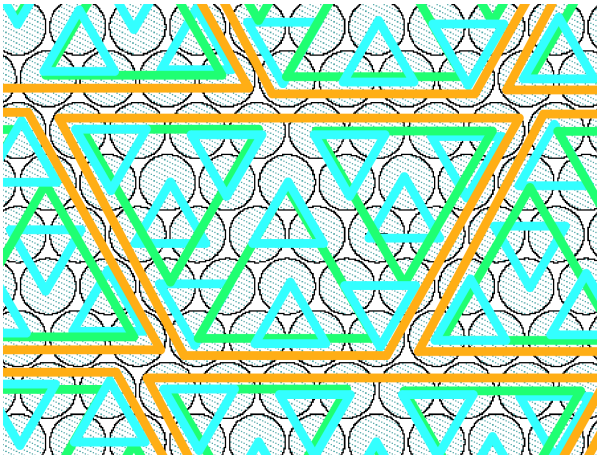
5.4.6 Mapping-speed improvement

In Fig. 5.26, we show the expected mapping-speed improvement from the trichroic prototype described above and from the notional hexachroic design described in the caption of Fig. 5.2. The prototype yields a modest improvement of approximately 10% in the 220-GHz band and 30% in the 90-GHz band. For such a small total bandwidth, there is not much more room for improvement. The mapping-speed improvement is much more dramatic when the total bandwidth is larger than approximately 3 : 1, which is shown by the hexachroic design in Fig. 5.26. In this figure, the pixel spacing has been optimized for 120 GHz. For frequencies more than an octave away, the improvement is greater than 50% and increasing steeply. The hierarchical array is most beneficial for large total bandwidths.

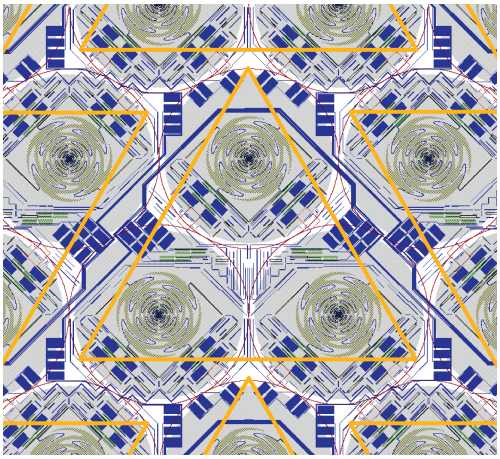
5.5 Scalability

The prototype described above demonstrated the essential features of a hierarchical sinuous-antenna phased array. The next step is to adapt this technology for a deployable wafer.

First of all, we will need to include both polarizations. With a good design for a microstrip crossunder, which we developed for PB-2, there is no fundamental problem here. The only challenge is in the complexity of the summing networks, since the microstrip lines will have to avoid each other. Additionally, for the sake of symmetry on all microstrip segments, we will want to keep track of how many crossunders occur along each path and at which points. We can then add dummy crossunders to the the other paths, so each path is identical. To keep



(a) Triangular hierarchy for tiling the plane. The triangles increase in size approximately in octaves. The rhombi tile the plane.



(b) A design for a two-level hierarchy sensitive to 90, 150, 220 and 270 GHz.

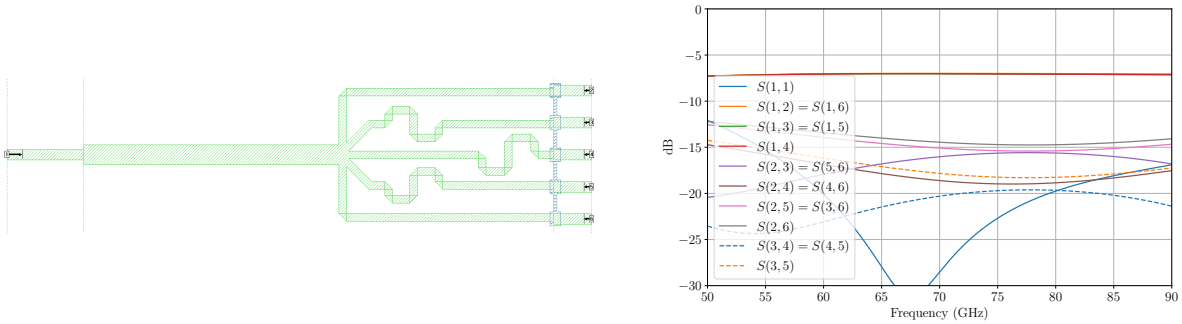
Figure 5.27: Extending the concept of hierarchical phased arrays to deployable wafers.

track of all of this, we have written a script in `C++` for use with the layout program L-Edit. It allows for semi-automatic implementation of the summing network. Email `ajcukierman@berkeley.edu` for a copy.

We need a hierarchical structure that can be repeated across the wafer. An example is shown in Fig. 5.27(a), where we have three levels of triangular arrays that combine to form tileable rhombi. In this case, the effective diameters of the triangles increase in factor of two. In such a design, every octave that is added to the bandwidth of the focal plane would require an extra level of the hierarchy. For the 10-cell triangle, it will be necessary to include a *pentavider*. This can be designed by analogy to the trividers presented in the previous sections. Simulations in both ADS and Sonnet show good performance. See Fig. 5.28 for a Sonnet simulation.

With more and more frequency bands, the space on the wafer will start to become limited. The two-level hierarchy shown in Fig. 5.27(b) is sensitive to bands centered on 90, 150, 220 and 280 GHz, where the 90- and 150-GHz bands are arrayed. Notice how congested the wafer is. To make space available, the double-ring hybrids can be compatified by folding the microstrip segments. A close-up of the pixels from Fig. 5.27(b) is shown in Fig. 5.29, where the hybrids have been folded to take up less total area. Electrically, these double-ring hybrids are identical to those discussed in Sec. 5.3.2. The microstrip segments have simply been folded to reduce the total required area. The corners can be mitered to reduce reflections.

To reduce loss in the long microstrip paths necessary for the summing networks, we will need a dielectric with lower loss than the silicon oxide used in the prototype described above. For PB-2, the Berkeley group has developed devices based on silicon nitride, which shows a dielectric loss tangent that is approximately ten times lower than that of silicon oxide. At this point, there is little risk in switching to nitride for future hierarchical phased arrays.



(a) Layout for simulation in Sonnet Software.

(b) Simulated S-parameters.

Figure 5.28: A design for a two-stage Wilkinson pentavider. The meanders are necessary to ensure the electrical length of each path is identical. The S-parameters show that the power is being split evenly five ways ($S(1,i) \approx -7$ dB for $i \neq 1$).

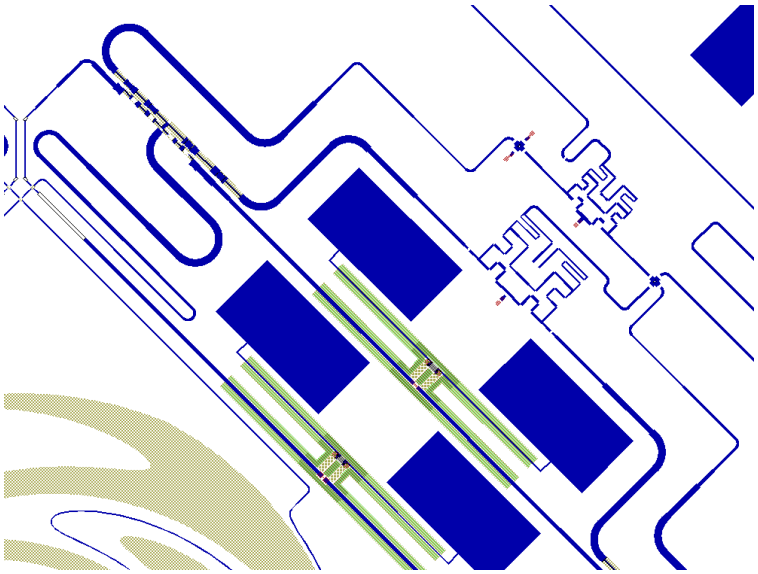


Figure 5.29: A zoomed-in view of the pixels from Fig. 5.27(b). The squiggly features in the upper right are double-ring hybrids. The larger is for the 90-GHz band, and the smaller is for the 150-GHz band. In the upper left, oriented along the northwest-southeast diagonal, is a tetraplexing bandpass filter that splits the signal from the antenna into frequency bands centered on 90, 150, 220 and 270 GHz. The two (green) bolometers are for the two highest frequencies, which are not arrayed in this design. The 90- and 150-GHz signals enter summing networks, which are shown in overview in Fig. 5.27(b).

Chapter 6

Focal-plane structure

This chapter is affectionately dedicated to McMaster-Carr's 0-80 3/16" stainless-steel socket-head cap screw.

The device wafers must be integrated with the readout electronics and installed in the focal plane, which is the coldest temperature stage in the receiver cryostat. The design of the focal plane is a compromise among competing cryogenic, electrical, mechanical and logistical concerns.

6.1 POLARBEAR-2 wafer module

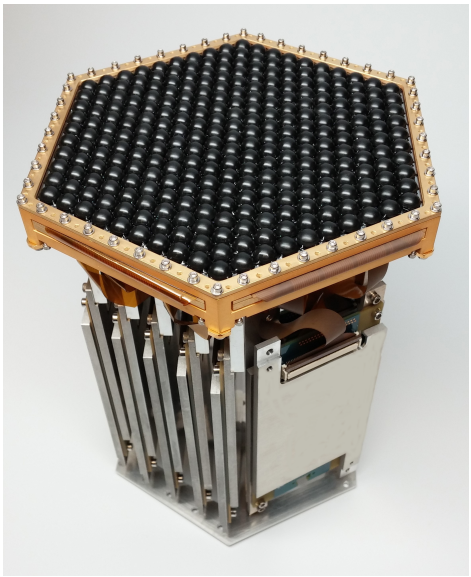
The *wafer module* integrates the device (bolometer) wafer, the lenslet array and the *LC* chips. Several wafer modules are installed onto the focal plane, but each can be removed, transported and tested separately. We designed the PB-2 wafer module, an overview of which is shown in Fig. 6.1. Some photographs of these wafer modules installed in the PB-2 main receiver for testing at KEK are shown in Fig. 6.2. The PB-2 focal plane consists of seven wafer modules in a hexagonal tiling. The modules, which are bolted to the 250-mK focal-plane plate, are connected by niobium-titanium (NbTi) *striplines*[1] to the 4-K SQUID amplifiers.

6.1.1 Invar holder

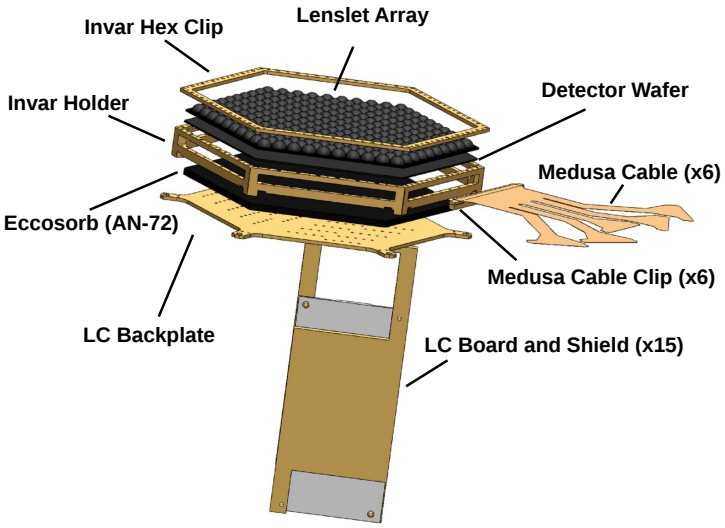
The device wafer is clamped to the lenslet array in a holder made of *invar*, an alloy of nickel ($\sim 36\%$) and iron, which experiences negligible thermal contraction.[2] This property is im-

[1] This terminology is conventional in the PB-2 and SPT-3G collaborations but somewhat misleading, since these cables are *not* what microwave engineers would consider to be "striplines". For a microwave engineer, a "stripline" is essentially a microstrip line that has *two* ground planes: one above the strip and one below. The NbTi striplines are more similar, at least in geometry, to what microwave engineers would call *parallel-plate transmission lines*. The wavelengths are so large in this circuit, however, that transmission-line properties of the NbTi striplines are basically irrelevant.

[2] The name *invar* is meant to be reminiscent of words like *invariant*.

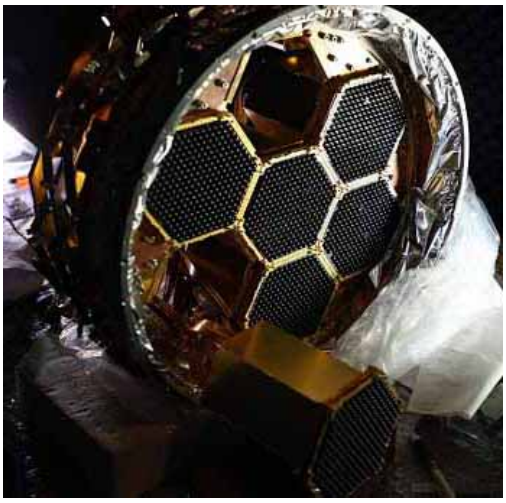


(a) The first assembled PB2 wafer module (June 2015). Since this photograph was taken, the hex clip was replaced with six smaller corner clips (see, e.g., Fig. 6.2(b)).

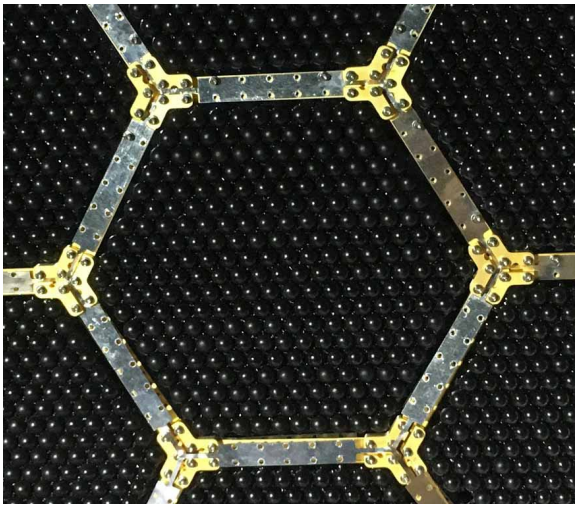


(b) Exploded CAD image of the wafer module. Wire bonds are made from the wafer to the medusa cable. The medusa cable connects the bolometers to the *LC* boards. The ECCOSORB terminates the back lobe of the antennas. The real module has medusa cables on all six sides and 15 *LC* boards. (Image credit: John Groh.)

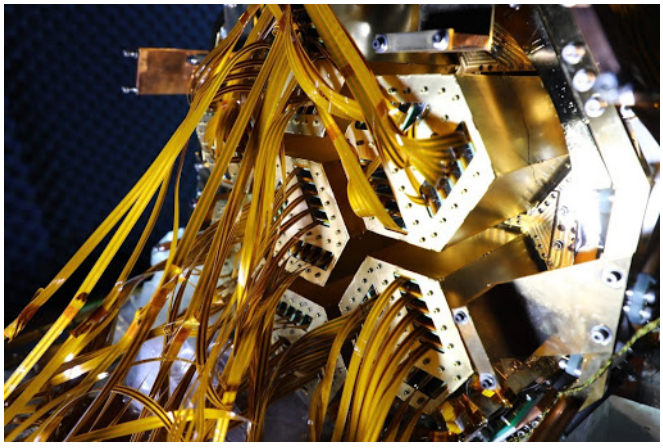
Figure 6.1: PB2 detector module integrating the device wafer, lenslet array and *LC* boards.



(a) Populating the PB2 focal plane. At bottom, a sixth module is being guided into its slot in the focal plane. The focal-plane tower is assembled and operated in a horizontal orientation.



(b) The gaps between the invar holders are covered with aluminum foil to create a Faraday cage around the cold electronics, especially the SQUID amplifiers.



(c) The back (non-sky) side of the focal plane showing the NbTi striplines that connect the 250-mK wafer to the 4-K SQUID amplifiers. There are fifteen striplines per module, and there are seven modules in the focal plane.

Figure 6.2: Some images of the PB-2 focal-plane tower in KEK from summer 2017. The particular wafers that were used in this test may not be deployed in PB-2, but the focal-plane design is likely to remain essentially identical. (Image credits: Masaya Hasegawa.)

portant, since the wafers also experience negligible thermal contraction but, being composed of single-crystal silicon, are very brittle. Invar normally has an iron-like color, but we typically gold plate the holders to prevent oxidation and increase thermal conductance. The invar holder really consists of two parts. The larger part, which is the one usually referred to as the "invar holder", contains a shelf in which the wafers are placed. Then we bolt down an invar clip to clamp the two wafers together. At this point, we have a sub-module consisting of the wafers and invar parts.

6.1.2 Lenslet-array alignment

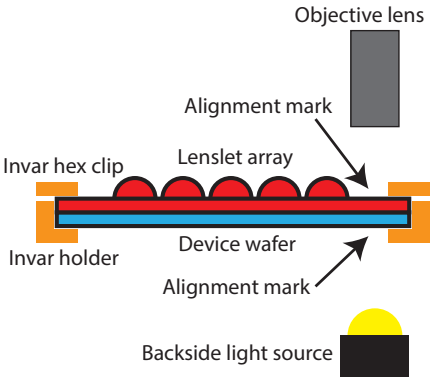
The device wafer and lenslet array must be aligned so that the lenslets and antennas are centered on each other. A typical tolerance for this alignment is 25 μm, but this is a soft requirement. The main effect of a small misalignment is a steering of the beam off-zenith.

In the corners of the device wafer, there are tic-tac-toe alignment marks that mate with marks in the corners of the lenslet array. For PB-1, the lenslet arrays were made at Berkeley by Erin Quealy [84]. For PB-2, the lenslet arrays were made at UC San Diego by Praween Siritinasak [93]. For SPT-3G, the lenslet arrays were made at the University of Illinois, Urbana-Champaign by Joaquin Vieira's group. In the device wafer, the alignment marks are defined as slots in the niobium ground plane. In the lenslet array, the alignment marks are defined as trench etches in the silicon that are created at the same time as the seating pockets for the silicon hemispheres.

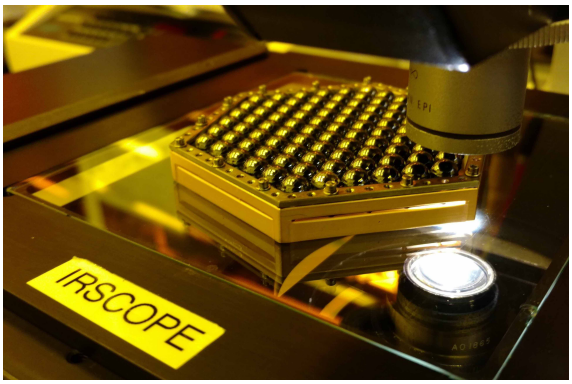
In order to be able to see both sets of alignment marks simultaneously, we use a back-illuminated infrared microscope. In the Marvell Nanolab, we use a tool called irscope. This is really nothing more than a regular back-illuminated microscope with a CCD camera whose infrared filter has been removed. The geometry of the alignment procedure is depicted in Fig. 6.3(a), and a corresponding photograph is shown in Fig. 6.3(b). The underside illumination shines light through the slots in the ground plane. Although this light must travel through two wafers' worth of silicon, the infrared-sensitive CCD picks it up quite easily. At the same time, we use a top-side illumination from near the objective lens to pick up the lenslet-array alignment marks. Some images from the CCD camera are shown in Fig. 6.3(c) and Fig. 6.3(d). There are alignment marks in each of the six corners of the hexagon. The alignment procedure is iterative: we align one corner, move to the opposite corner, align the opposite corner, return to the first corner, etc. Then we check the other pairs of opposite corners. As we tighten the bolts that clamp the invar hex clip down, we also check all six corners to make sure the wafers have not shifted. The clamping is remarkably robust. Even after a few screws have been tightened, it is surprisingly difficult to move the lenslet array relative to the device wafer. We have occasionally checked the alignment after shipping and thermal cycling, and we have not seen any noticeable shifts.

6.1.3 Flex cable

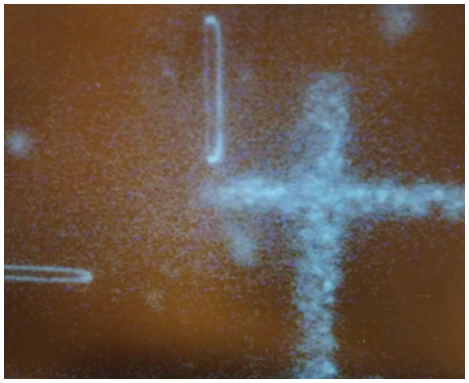
The cable the connects the device wafer to the *LC* boards is referred to as the "flex cable" in the POLARBEAR and SPT collaborations. For PB-2 and SPT-3G, the cable breaks out into multiple connectors and is, therefore, sometimes referred to as a "medusa cable". Some photographs of a PB-2 flex cable are shown in Fig. 6.4. We designed the cables to



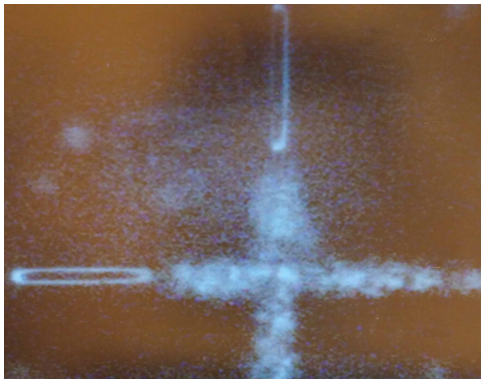
(a) Cartoon of the wafer-alignment geometry. The invar hex clip is loosely bolted until the wafers are well aligned. Then we tighten the screws to clamp the device wafer and lenslet array.



(b) Photograph of a PB-1 lenslet array being aligned to a prototype wafer that was designed to have mating alignment marks. While the hex-clip screws are loose, we prod the lenslet array carefully with plastic or carbon-fiber tweezers to move it into good alignment with the device wafer. The image from the CCD camera updates in real time.

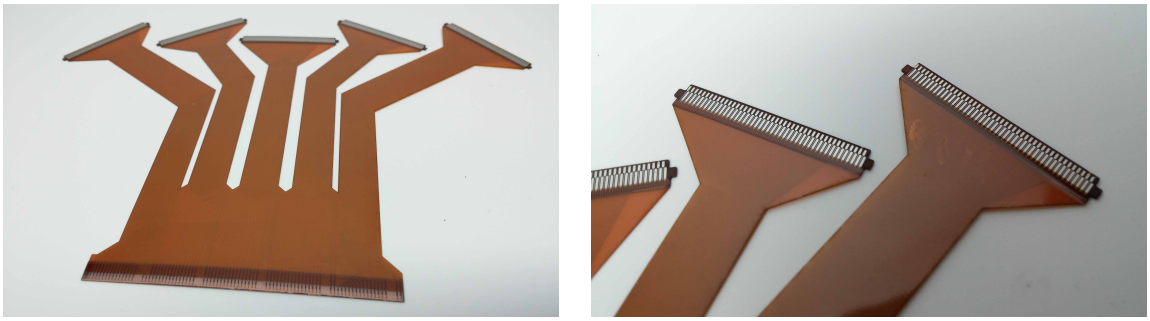


(c) An image from the CCD camera showing the two sets of alignment marks when the wafers are misaligned. The white chalk-like features are the back-illuminated slots in the ground plane of the device wafer. The thinner outlined streaks are the trench etches in the lenslet array.



(d) An image from the CCD camera showing the two sets of alignment marks when the wafer are well aligned. The ground-plane slots are approximately 30 μm in width; from this image, then, we can say that the alignment is good to approximately 10 μm.

Figure 6.3: Alignment procedure for the device wafer and lenslet array.



(a) The flex cable is wire bonded at bottom to the device wafer and breaks out into five heads for connecting to the *LC* boards.

(b) Detail of the connector ends.

Figure 6.4: A PB-2 flex cable. Since these photographs were taken, a gold plating was introduced on the device-wafer end to make wire bonding more reliable.

match our wire-bonding capabilities and *LC*-board connectors.[3] The cables are fabricated by Tech-Etch, Inc.

The cables are constructed from polyimide on which copper, tin and a small amount of gold plating have been deposited. Tin is a superconductor at the operating temperature of these cables, but it is difficult to bond to. We, therefore, etch away the tin at the bond pads to expose copper. The small amount of series resistance that this introduces to the circuit is negligible. Since copper forms an oxide, it is necessary to clean the bond pads 24-48 hours prior to bonding or to store the cables in a dry-nitrogen environment. In more recent iterations, however, the copper bond pads have been gold plated to remove these needs. At the bonding end, extra layers of polyimide are deposited to make the bonding surface stiffer.

The *LC*-board connectors are zero-insertion-force (ZIF) connectors that have become standard in the POLARBEAR and SPT collaborations. The connectors use fine pins to press against the cable traces.

Each head of the PB-2 flex cable connects to 38 bolometers. On each frequency-multiplexing comb, there are two calibration-resistor channels, so the total multiplexing factor is 40. The ZIF connectors actually contain 45 pairs of pins, so seven are simply unused. The standardization of connectors is more important than the wasted connections.

6.1.4 Wire bonding

The flex cable is electrically connected to the device wafer with aluminum wire bonds. For PB-1, these bonds were made manually with a West-Bond ultrasonic wire bonder. For PB-2, the bonds are made automatically with a Hesse-Knipps ultrasonic wire bonder at LBNL.

A *wire bond* is a type of electrical connection that is especially useful for delicate small circuitry. Wire bonding is especially useful when many finely-spaced connections must be made. The wires are typically made of aluminum or gold with diameters on the order

[3] Grantland Hall was directly responsible for laying out the flex-cable design.

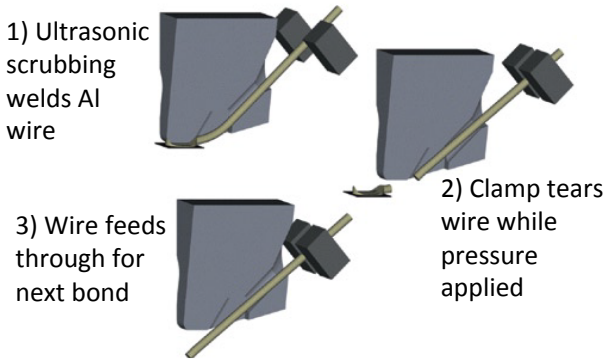


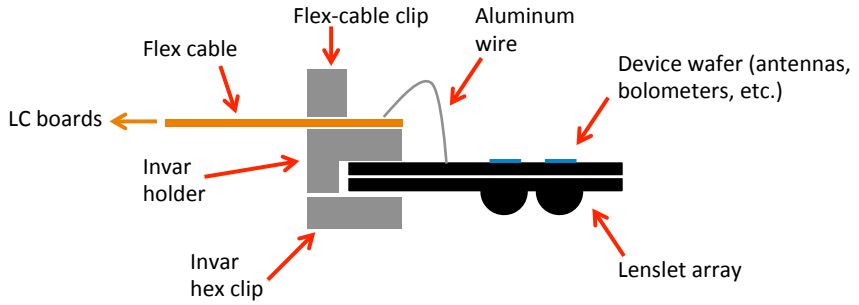
Figure 6.5: A cartoon of the process for ultrasonic wire bonding. Each wire is bonded on two ends. The first bond is made with steps 1 and 3 only, i.e., we move to the second bonding site without breaking the wire. For the second bond, all three steps are used. (Image credit: Small Precision Tools (SPT) for the cartoons; the text is *not* from SPT.)

of 25 μm. Other materials have been used and are chosen based on the particular application [43]. In our experiments, we use exclusively aluminum wires with a diameter of 25 μm. There are several methods for actually making the bond. We use *ultrasonic* bonding. A cartoon of the process is shown in Fig. 6.5. The idea is to weld the wire to the bond pad with ultrasonic scrubbing that heats up the wire only locally. This is to be contrasted with conventional soldering which heats up the surroundings in a relatively uncontrolled manner. Ultrasonic wire bonding can be considered a kind of room-temperature welding, since effectively nothing other than a small section of the wire is heated. The first side of the connection is made by welding the wire to the surface. The wire clamp is then loosened, so the wire can be fed through while the bonding head moves to the second bonding site. At the second bonding site, another weld is made. This time, however, the clamp is engaged to break the wire. More wire is fed through to prepare for the next connection.

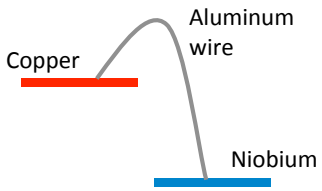
The wire-bonding geometry for PB-2 is depicted in Fig. 6.6. The flex cable is held down to the invar holder with double-sided tape for wire bonding. After bonding, a much stronger stress relief is provided by the flex-cable clip. This additional pressure is only applied *after* bonding, since it tends to bend the cable upwards and create a floppy surface, which can be fatal for wire bonding. After the bonds have been made, however, it is fine to allow the cable to bend upwards, since the bonds have some slack. The flex cable must be aligned to the wafer, so the bond pads line up within approximately 50 μm. To facilitate this process, we designed an alignment jig relying on a four-axis micrometer stage (x, y, z and ϕ) that lowers the flex cable into place before it is pressed down onto the adhesive.[4] We developed a procedure for this flex-cable installation, which we transferred to the LBNL technicians.

We worked with several technicians at LBNL to develop the automatic wire-bonding process for PB-2. Some photographs are shown in Fig. 6.7. The bonder can be used in a fully-automatic mode, in which an entire edge of the wafer (380 bonds) is programmed without any human intervention. Bonding parameters such as loop height, time, pressure

[4]The CAD design and the assembly of this jig are the work of John Groh.

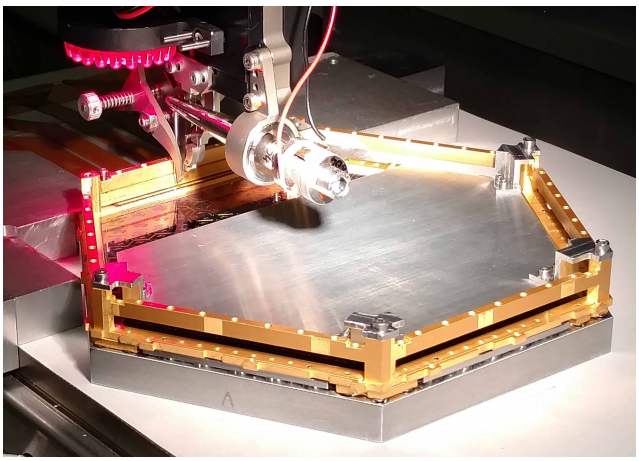


(a) Cross section of the wafer module with the wire bonds highly exaggerated. The flex-cable clip provides stress relief on the flex cable, so the wire bonds are not broken in handling. The flex cable is additionally held down to the invar holder with double-sided tape.



(b) Zoom in on the wire bond. In more recent modules, the flex-cable bond pads have been gold plated, which generally improves the reliability of the bonds.

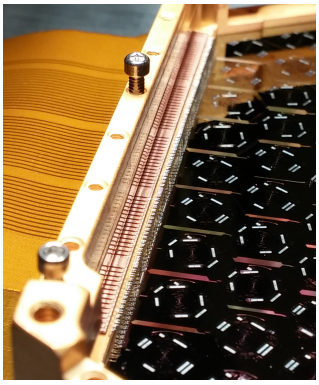
Figure 6.6: PB2 wire-bonding basics. Wire bonds are made from the device wafer to the flex cable, which ultimately plugs into the *LC* boards.



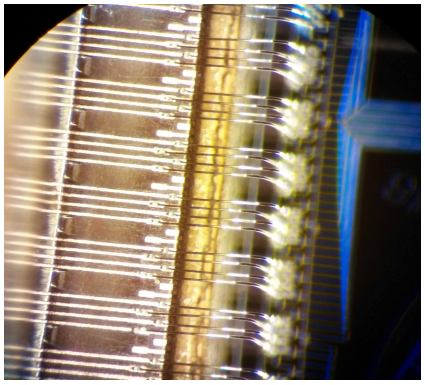
(a) The bonder in action. The aluminum that is covering approximately 2/3 of the wafer is simply for protection and has no influence on wire bonding. In this photograph, a bond is being made in the middle of the hexagon side in the upper left.

(b) A close-up of the bonding head showing the wire feed-through and the wedge.

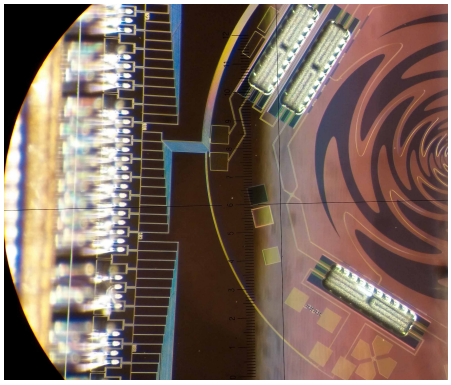
Figure 6.7: Automatic wire bonding at LBNL.



(a) A fully bonded wafer side. There are 380 wire bonds for 190 bolometers.



(b) The flex-cable side of the bonds. One out of every three traces on the wafer and flex cable are open-circuited. This made the PB-2 design, which is dichroic, easily adapted to SPT-3G, which is trichroic.



(c) The wafer side of the bonds. It is difficult to see the bond pads due to the steep vertical drop in the bonds. This aspect ratio cause many difficulties early on in the development of automatic wire bonding.

Figure 6.8: Completed wire bonds on a PB-2 wafer.

and ultrasonic power are programmed for each side of the bonds. The bond locations are also pre-programmed. The bonder has pattern-recognition software, which can compensate for small perturbations to the geometry, e.g., a rotation or a shift of the wafer relative to the flex cable. Generally, the bonding is performed under the supervision of a technician, who makes sure that every bond is made properly. It is common for $\mathcal{O}(1\%)$ of the bonds to fail. When this happens, the technician stops the program, takes any necessary action and then restarts it at the failure point.

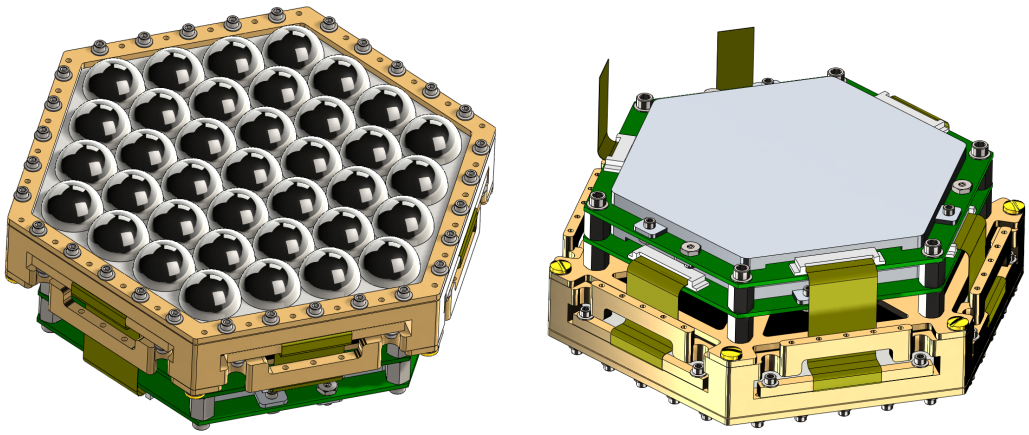
Before bonding an entire side of a wafer, some test bonds are made and their strengths are measured with a calibrated pull tester. A good rule of thumb is to ensure the bond strengths are above 6 g.[5] Sometimes the bond pads have a poor surface condition, which brings the strengths down to as low as 2 or even 1 g. This can sometimes be overcome by adjusting the bonding parameters.

Bonding aluminum wire to niobium is non-standard in the semiconductor industry. To make reliable bonds to the niobium, we found that it was necessary to increase the ultrasonic power while decreasing the bond force (pressure).

Some photographs of a bonded wafer are shown in Fig. 6.8. The bonds are spaced 100 μm apart and are slightly staggered, so the bonding wedge does not bump the neighboring bonds. The aspect ratio of the bonds (vertical distance relative to horizontal distance) is relatively large and is a primary source of frustration in wire bonding PB-2 wafers. When we bond to the same surfaces but with a smaller aspect ratio, we see an increase in bond strength and success rate. In a future design, if possible, it would be better to decrease this aspect ratio.

When the bonds stick, the bonding takes about 10 minutes per side. Much more time

[5]Obviously, what is meant is a gram-*weight*, which happens to be approximately equal to a centinewton (cN), incidentally. The Hesse-Knipps bonders actually set bond force in cN.



(a) Sky side. (b) Readout side.

Figure 6.9: Wafer-module design for the LiteBIRD mid-frequency detectors. The low-frequency wafer module is similar but with larger pixel diameters. The invar-holder design is similar to that of PB-2. The wafers are 4″ in diameter instead of 6″, though. The *LC* boards are designed to be stacked horizontally instead of vertically. The pixel diameter is for the mid-frequency detectors is 12 mm. The truncated cables are striplines that will actually be connected to the 4-K SQUID amplifiers.

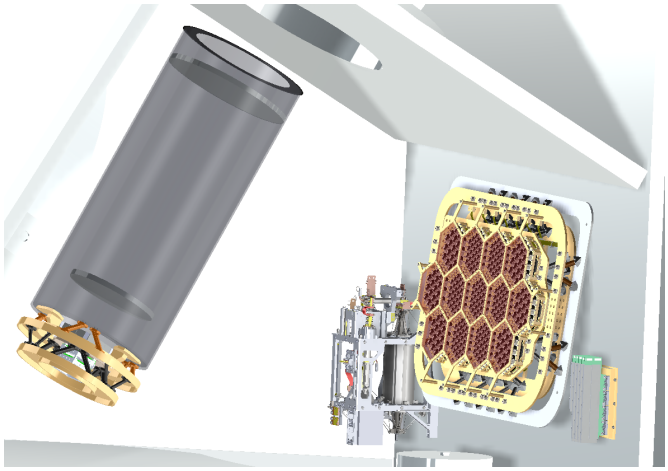
is spent simply *preparing* the module for bonding and checking the electrical connections post-bonding.

6.1.5 SPT-3G

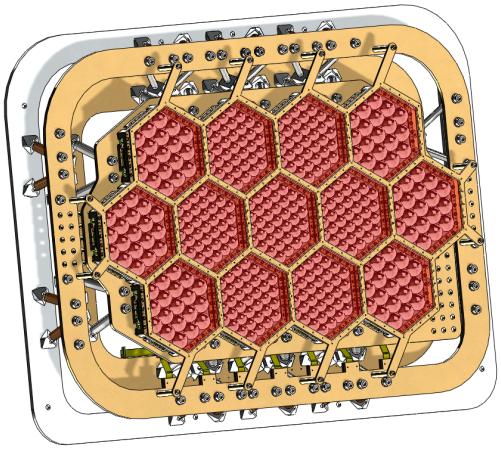
This wafer-module design was shared with the SPT collaboration to aid in the design of the SPT-3G focal plane. Some modifications were made to match their needs. The SPT-3G pixels are trichroic but otherwise similar to the PB-2 pixels. There are, therefore, 50% more bolometers per wafer. The flex cable was modified accordingly and has eight connector heads. For easier wire bonding, the invar holder was split into two pieces, so there need be no metal above the flex-cable surface during bonding. This allows flexibility in the particular bonder that is used.

6.1.6 LiteBIRD

The LiteBIRD wafer modules for the low- and mid-frequency channels will be similar to those described above for PB-2 and SPT-3G. Some CAD figures are shown in Fig. 6.9, which were included in the Concept Study Report (CSR) submitted to NASA in 2016. The LiteBIRD wafers are 4″ in diameter, and the pixel spacing is 12 mm. Contrast this with 6″ and 6.789 mm for the diameter and pixel spacing, respectively, of the PB-2 wafers. The number of bolometers per wafer is, therefore, much lower. To make the modules more compact, we design the *LC* boards to lie horizontally. Only two striplines are required for each mid-frequency wafer module.



(a) Telescope configuration. The low- and mid-frequency telescope is a crossed Dragone, whose mirrors are only faintly visible in the foreground, which couples to the rectangular focal-plane tower at right. The cylindrical refracting telescope is in the background and completely out of the optical path of the crossed Dragone; this is the high-frequency telescope, which couples to a single wafer of feedhorns coupled to orthomode transducers (OMTs). To the left of the low- and mid-frequency focal plane is the ADR; to the right are the SQUID cards.



(b) Low- and mid-frequency focal-plane tower. The focal-plane stage is held at 100 mK. The tower consists of thermally isolated temperature stages. The white stage is bolted to the $\sim$ 5-K wall of the cryostat. The other stages are held at $\sim$ 300 mK and 2 K. The stages are connected thermally insulating Vespel struts. Just above the lenslets are (red) metal-mesh low-pass filters sunk to the 300-mK stage.

Figure 6.10: LiteBIRD receiver design as of the submission of the Concept Study Report in 2016. The design is likely to change over time, especially the distribution of frequency bands in the two telescopes, the number of wafer in each telescope and even the number of independent telescopes.

6.2 LiteBIRD focal-plane tower

We designed the focal-plane towers for the Concept Study Report (CSR) for "U.S. Participation in the LiteBIRD Cosmic Microwave Background Polarization Survey" that was submitted to NASA in 2016. This design should be thought of as a feasibility argument and not as the design to which the LiteBIRD collaboration is necessarily committed.

The focal-plane tower is the thermal isolation structure that separates the typically 100-300 mK focal plane from the typically 4- or 5-K cryostat shell. The baseline focal-plane temperature for LiteBIRD is 100 mK, and the cryostat shell is likely to be closer to 5 K. The main competing concerns in designing a focal-plane tower are mechanical and cryogenic. There must be enough thermal isolation to reduce the cooling-power requirements on the coldest stages, but there must also be enough rigidity to avoid low-frequency ($\lesssim$ 50 Hz) mechanical resonances. For a space mission especially, there are the additional concerns related to weight and volume.

The design developed for the CSR is shown in Fig. 6.10. We use Vespel struts to connect the temperature stages on account of its favorable ratio of elastic modulus to thermal conductivity [52, 88]. The temperature stages are constructed from gold-plated aluminum.

The aluminum is chosen for its relatively low mass density, and the gold plating is applied to increase the thermal conductance. In addition to the 5-K cryostat shell, a 1.8-K stage will be provided by a Joule-Thomson (JT) cooler. The sub-Kelvin cooler will provide a 300-mK stage with a ^{3}He sorption cooler and a 100-mK with an adiabatic demagnetization refrigerator (ADR). The sub-Kelvin cooler will be provided by the French Alternative Energies and Atomic Energy Commission (CEA) group, which developed the cooler for the SPICA-SAFARI instrument [30].

6.2.1 Mechanical design

Vibrations can couple into the focal-plane tower and cause spurious heating of the detectors, which we sometimes call *microphonic pick-up*. Since bolometers are essentially thermometers, this is simply an additional noise source in the detector timestreams. The exact requirements depend on the vibration spectrum of the cryogenic receiver. Typically, the power has a roll-off frequency at $\mathcal{O}(10)$ Hz, so any mechanical resonances in the focal-plane tower need to be at much higher frequencies.

For a satellite-based telescope, an additional mechanical requirement comes from the need to survive the large accelerations present during launch. *Launch locks* can be used to strengthen the structures during launch but allow for thermal isolation once in orbit. One option is to use Frangibolts, which use shape-memory alloys that expand on heating and break bolts to create gaps. These types of bolts are common in space applications. To survive launch, the focal-plane structure including launch locks should be designed to withstand accelerations of $50g$.

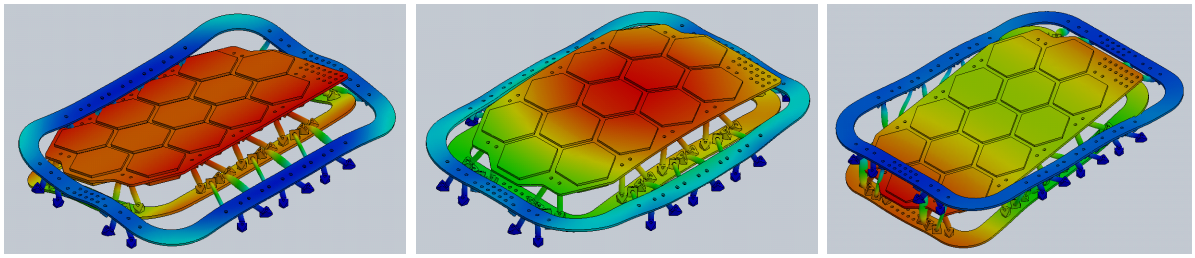
Vibration analysis

At the design level, the normal modes can be characterized through simulations using software such as SolidWorks. SolidWorks simulations of the LiteBIRD low-frequency focal-plane structure are shown in Fig. 6.11. To accelerate the simulations, we often replace the wafer modules with their equivalent weight in aluminum. Of course, this choice alters the resonances, but the qualitative modes and rough resonant frequencies should persist. The simulations are only meant to provide a baseline design before machining and construction. In Fig. 6.11, we show the normal modes with the lowest three resonant frequencies, which are all above 70 Hz. To qualify this design, it will be necessary to to build a demonstration model and subject it to a shake-table test.

Vibration testing

We performed some preliminary vibration testing using a *shake table* at the Space Sciences Laboratory (SSL) at UC Berkeley. The most fragile components in the focal-plane structure are the wafers, which are composed of single-crystal silicon and supported only at their edges, leaving them susceptible to drum-head modes.

We assembled a mock wafer module using PB-1 hardware for the lenslets and invar holder. In place a device wafer, we used a blank silicon wafer, to which we glued an accelerometer. The setup is shown in Fig. 6.12. We bolt the module to the table which can be controlled

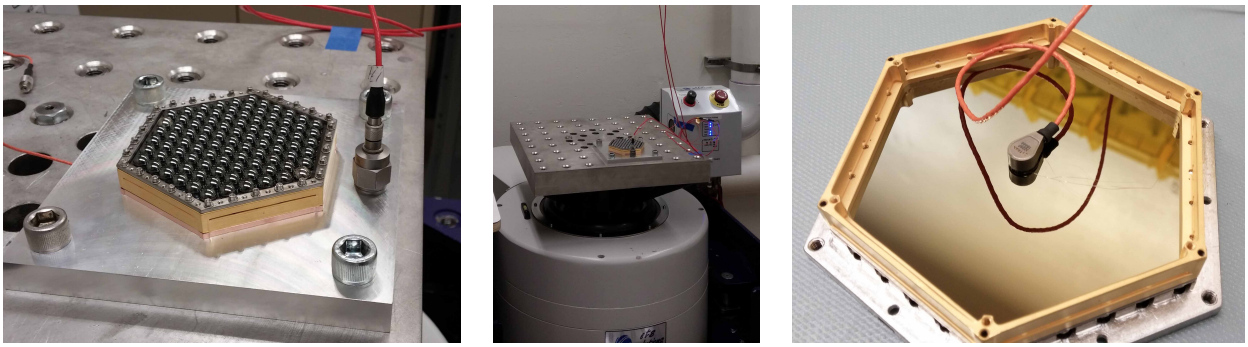


(a) 71 Hz. The focal plane rotates along its long axis.

(b) 87 Hz. This is a kind of drum-head or trampoline mode: the center bounces up and down.

(c) 89 Hz. This is a seesaw mode: the focal plane rotates around its short axis.

Figure 6.11: SolidWorks mechanical simulations of the LiteBIRD low- and mid-frequency focal-plane structure. SolidWorks solves for the normal modes of the structure subject to certain boundary conditions. These images show exaggerated images of the deformations that define each mode. To shorten the computation time, the wafer modules have been replaced with their equivalent weight in aluminum.



(a) The module bolted to the shake table. The orange wire at right is for a control accelerometer mounted directly to the shake table. The orange wire at left is for an accelerometer glued to the mock device wafer.

(b) The shake table at SSL. For a sense of scale, recall that the PB-1 wafers are approximately 4″ in diameter.

(c) The device-wafer side of the wafer module. We used a blank silicon wafer and glued an accelerometer to the center.

Figure 6.12: Shake-table testing at SSL with a PB-1 wafer module.

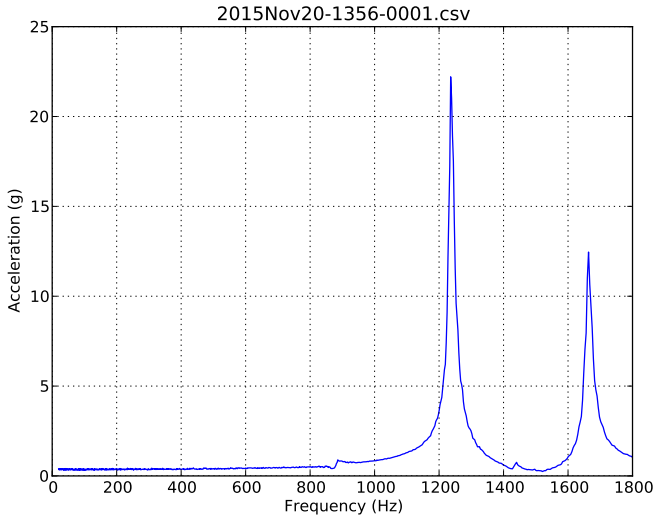


Figure 6.13: Sine sweep of a PB-1 module. We see clear resonances above 1 kHz.

to vibrate with a given amplitude and frequency. We performed a test called a *sine sweep*, in which a low-amplitude vibration is swept over frequency to find mechanical resonances. This test does not directly probe the module's ability to survive launch. Knowing the launch vibration profile, however, this information can be used to make plausibility arguments. The accelerometer response is plotted in Fig. 6.13. We see two clear resonances below 1800 Hz, one at 1250 Hz and another at 1650 Hz. These are in the tail of a typical launch profile, which is shown in Fig. 6.14. When coupled to the focal-plane structure, however, these resonances can change, since the boundary conditions are different.

A standard test for launch survival is a *random vibration test*, in which the payload is subject to the acceleration spectral density profile shown in Fig. 6.14. This is a standard NASA test. As the project progresses and more realistic vibration profiles are estimated, the test may be modified. In this test, we can expect the payload to experience up to $50g$ of acceleration. Because we didn't want to break the PB-1 lenslet array shown in Fig. 6.12, we made a mock lenslet array using hemispheres made of aluminum, which has a similar density to silicon. The mock module was subjected to a random vibration test, and no damage at all was noticed. Perhaps the next test should check for misalignment between the device wafer and the lenslet array. Obviously, the tests will eventually need to use silicon lenslets in a similar configuration to what is intended for the flight model.

6.2.2 Cryogenic design

In the CSR, we selected a sub-Kelvin cooler made by the French Alternative Energies and Atomic Energy Commission (CEA) that was developed for the SPICA-SAFARI instrument [30]. This is a hybrid cooler that uses a single ADR stage, which will be operated at 100 mK for LiteBIRD, and a 300-mK ^{3}He sorption stage. The cooler was designed to use thermal interfaces at 1.8 K and 4.9 K. We worked with the CEA group to design a focal-plane

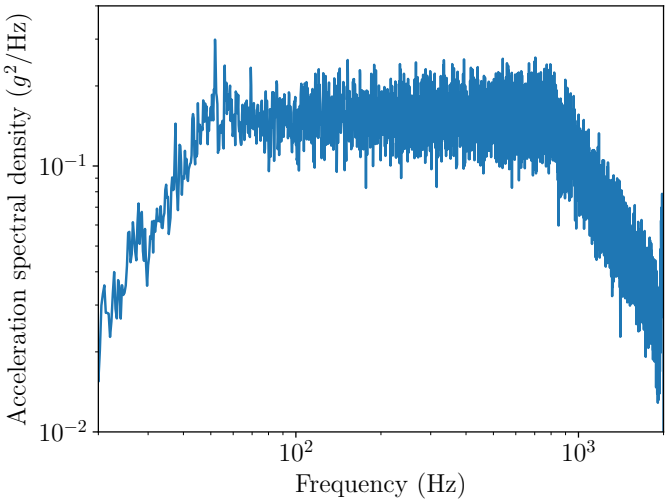


Figure 6.14: Acceleration spectral density for a *random vibration test*. The RMS acceleration is $15g$. This is a standard profile for NASA vibration tests. In a few seconds of testing, then, the payload may experience up to $\sim 50g$.

structure with thermal loads that allow for maximal fridge hold times. We developed a thermal model of the LiteBIRD focal-plane structures, which was used to estimate thermal loads. A cartoon of this model is shown in Fig. 6.15.

Much of what follows comes from a memo that was written for the LiteBIRD collaboration. A major decision leading up to the submission of the CSR was the selection of the sub-Kelvin cooler, so it was important to have a well-understood and carefully constructed thermal model.

Sources of thermal loading

We describe the major sources of expected thermal loading on the LiteBIRD focal-plane tower.

Radiative load The primary source of radiative loading on all of the stages of the focal-plane tower comes from the cryostat shell, which will most likely have a temperature of 5 K. To reduce stray reflections, the shell will be absorptive, which increases the radiative thermal loading substantially. Radiative loading is most severe for the focal-plane stage, i.e., the coldest stage. For other stages, it is a minor contribution. From the perspective of the focal plane, the telescope optics occupy a relatively small solid angle. As a worst-case assumption, then, we take the absorptive cryostat shell to occupy the entire 2π steradians visible to any area element on the focal-plane tower. In estimating the radiative load, we include all of the surface area at each temperature stage with appropriate emissivities. This includes the lenslet arrays, the temperature-stage structural rings and the insulating struts.

The struts have a non-negligible surface area. We assume that any radiative loading on

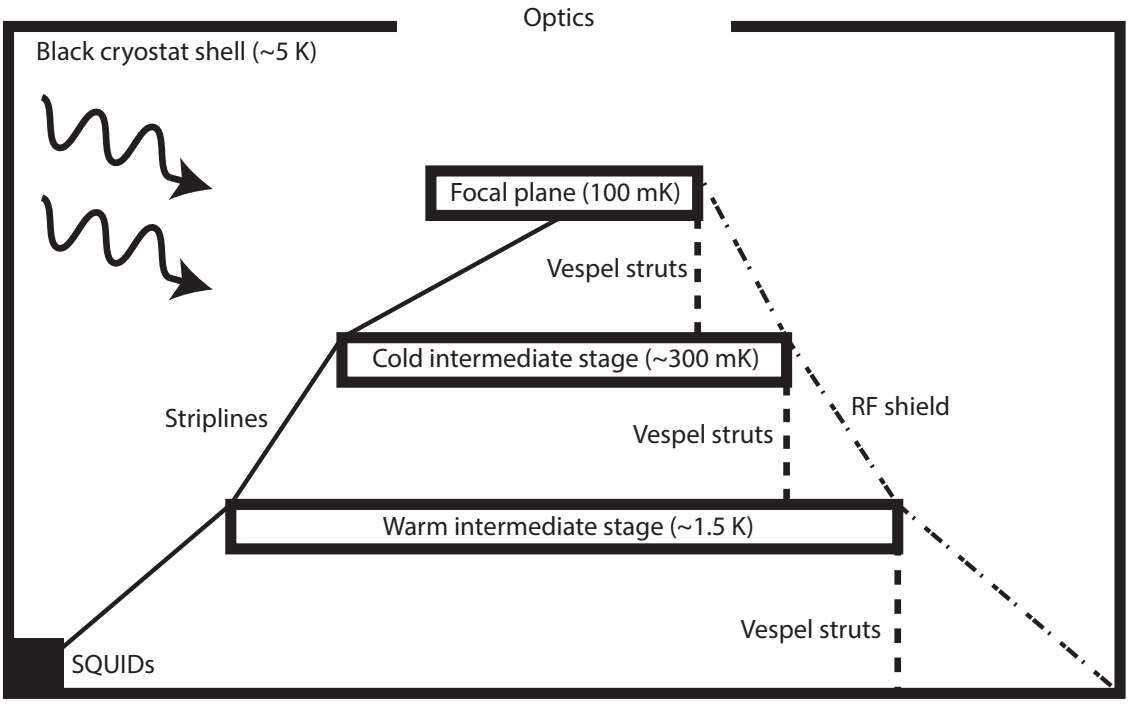


Figure 6.15: LiteBIRD thermal model showing the main sources of loading.

the struts ultimately flows to the lower-temperature stage. For example, the radiative load on the struts connecting the 100- and 300-mK stages flows to the 100-mK stage. We will find that this is a significant contribution to the thermal load at low temperatures. To suppress it, we may need to engineer the struts so that they present a reflective surface to the cryostat shell. This can be done by a choice of material, e.g., NbTi, or by miniature radiation shields. Thin aluminized tape has been used by the Astro-H engineers at Goddard Space Flight Center (GSFC). The Goddard engineers have also recommended the alloy Ti 15333, which has a critical temperature of approximately 3.4 K.

We take the gold-plated surfaces such as the stage rings to have an emissivity of 3%. This is a number that the Astro-H engineers use. It may be a conservative overestimate.

The lenslet-array emissivity is difficult to estimate. Although the silicon is low loss, off-zenith rays can experience total internal reflection in the lenslets. This increases the path length that the rays take through the silicon and increases the absorptivity. This "greenhouse effect" creates an *effective* emissivity significantly larger than one would naively calculate based on the loss tangent of silicon. Simulations suggest that our lenslet arrays have an effective emissivity of approximately 30%.[6] This assumes that the loss tangent of the silicon lenslets is similar to what our group has measured at cryogenic temperatures. The anti-reflection coating tends to have a higher loss tangent by about an order of magnitude, which

[6]Mike Myers ran these simulations to understand the radiative loading on the POLARBEAR focal plane. It was found empirically that a 350-mK metal mesh filter just above the focal plane was necessary in order to allow the wafer to cool to the base temperature of the ^{3}He sorption refrigerator.

will increase the effective emissivity.

Metal-mesh filters have been assumed above each wafer. The cut-off is 120 GHz for the low-frequency wafers, 300 GHz for the mid-frequency wafers and 480 GHz for the high-frequency horns. For the following estimates, we have assumed tophat passbands.

Structural load Structural loading is defined here as the power that flows through the insulating struts of the focal-plane tower. The final design of the tower will be a compromise between cryogenic and mechanical concerns, since weaker struts are more insulating but also tend to lower the resonant frequencies. The focal-plane design used for the following estimates is the result of an iterative series of mechanical simulations that give a lowest resonant frequency of approximately 70 Hz as shown in Fig. 6.11.

The materials chosen for the struts should be both thermally insulating and mechanically stiff. The best material for the struts is probably Vespel SP-22 below about 2 K [88]. Above 2 K, the best material is probably Vespel SP-1. The *Planck* focal plane used NbTi struts, but NbTi is about an order of magnitude more conductive than Vespel [71]. Another promising material is carbon fiber, which is a better insulator than NbTi and very strong [88]. In the following estimates we assume Vespel SP-1 for the struts stretching from the cryostat shell ($\sim$ 5 K) to the warm intermediate stage ($\sim$ 1.8 K). The wall thickness is set to 350 μm, which is used in PB-2. The colder struts are assumed to be Vespel SP-22 with the same wall thickness.

Wiring load The striplines connecting the SQUIDs to the *LC* boards are responsible for the majority of the wiring thermal load. LiteBIRD will contain approximately 2000 bolometers. If a multiplexing factor of 40 is used, then 55 striplines are required. The multiplexing factor is likely to be larger than 40, but this gives an idea of the number of striplines required. We use the PB-2 stripline geometry with literature thermal conductivities for Kapton and NbTi [12, 71]. We assume the temperature intercepts are separated by 20 cm of stripline length.

We assume the wiring load from thermometry is negligible. This can be thought of as a requirement rather than an estimate.

RF-shield load The RF shield is composed of aluminized mylar and blocks the cold electronics, especially the SQUIDs, from the RF environment of the cryostat. It is not clear what the environment will be for LiteBIRD. It might not be necessary to have an RF shield.

Estimating the thermal conductivity through the RF shield is difficult. Little data exists for the thermal conductivity of mylar below 1 K. The thermal conductivities of aluminum vary by many orders of magnitude in the literature depending on purity and heat treating. Finally, the geometry of the RF shield has not been determined for LiteBIRD.

For now, we estimate the order of magnitude of the RF-shield thermal load by using the estimates made for the PB-1 cryostat [54]. We substitute the stage temperatures for those appropriate for LiteBIRD. The geometry, however, could be totally different for LiteBIRD, and the thermal conductivities could be wrong. This contribution to the heat load, therefore, has the largest uncertainty.

Other contributions Cosmic rays will deposit energy in the focal plane. The total power has been estimated and determined to be in the range of 1-10 nW on the focal plane, which is a minor contribution and essentially negligible.[7]

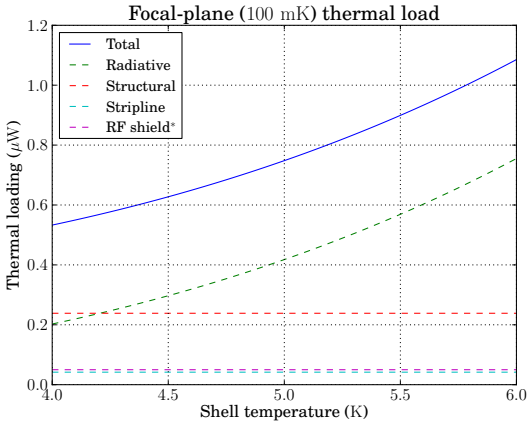
Vibrational energy can couple to the focal-plane tower, especially if there are mechanical resonances at low frequencies. The vibrational spectra of, e.g., the 5-K cooler and telescope controller need to be known in order to estimate or measure this contribution more precisely. In principle, the focal-plane tower is designed to have resonances at high-enough frequencies to avoid any significant vibrational coupling.

Thermal-loading estimates

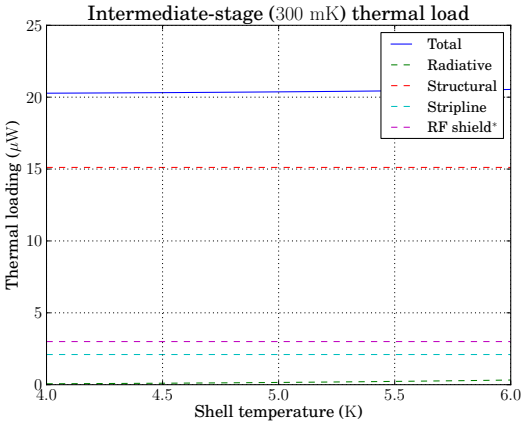
Since the temperature of the cryostat shell controls the radiative loading, which is the dominant source of loading on the focal-plane stage. We produce our estimates as a function of the shell temperature, which we allow to vary around 5 K. The estimates for the focal-plane stage, the intermediate ^{3}He-sorption stage and the 1.8-K JT-cooler stage are shown in Fig. 6.16. The load on the 100-mK focal-plane stage tends to be in the range of 1 μW. The load on the ^{3}He-sorption stage tends to be much higher, i.e., $\sim$ 20 μW. The load on the JT cooler is in the range of 0.3-1 mW.

For the CSR, we doubled these estimates to provide 100% margin. The doubled estimates evaluated at a shell temperature of 5 K were used as the requirements for the sub-Kelvin cooler selection.

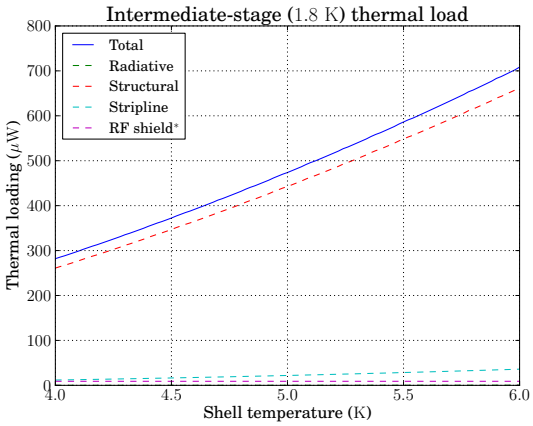
[7]This estimate comes from Tucker Elleflot.



(a) Thermal load on the 100-mK focal-plane stage.



(b) Thermal load on the 300-mK ^{3}He-sorption stage.



(c) Thermal load on the 1.8-K JT-cooler stage.

Figure 6.16: Estimated heat loads on the temperature stages of the LiteBIRD focal-plane towers. These estimates include contributions from both the low- and high-frequency focal planes. The four main contributions are shown as dashed curves in distinct colors. The total heat load is shown as a solid curve. We provide these estimates as a function of the shell temperature, since the loading on the focal-plane stage is strongly dependent on it.

Bibliography

[1] Kevork N. Abazajian et al. CMB-S4 Science Book, First Edition. 2016.

[2] Maximilian H. Abitbol et al. CMB-S4 Technology Book, First Edition. 2017.

[3] W. S. Adams. Some Results with the COUDÉ Spectrograph of the Mount Wilson Observatory. *Astrophys.J.*, 93:11, January 1941. doi: 10.1086/144237.

[4] P. A. R. Ade et al. BICEP2 II: Experiment and Three-Year Data Set. *Astrophys. J.*, 792(1):62, 2014. doi: 10.1088/0004-637X/792/1/62.

[5] P. A. R. Ade et al. Evidence for Gravitational Lensing of the Cosmic Microwave Background Polarization from Cross-correlation with the Cosmic Infrared Background. *Phys. Rev. Lett.*, 112:131302, 2014. doi: 10.1103/PhysRevLett.112.131302.

[6] P. A. R. Ade et al. Measurement of the Cosmic Microwave Background Polarization Lensing Power Spectrum with the POLARBEAR experiment. *Phys. Rev. Lett.*, 113: 021301, 2014. doi: 10.1103/PhysRevLett.113.021301.

[7] P. A. R. Ade et al. Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.*, 594:A13, 2016. doi: 10.1051/0004-6361/201525830.

[8] P. A. R. Ade et al. A Measurement of the Cosmic Microwave Background B-Mode Polarization Power Spectrum at Sub-Degree Scales from 2 years of POLARBEAR Data. *Astrophys. J.*, 848(2):121, 2017. doi: 10.3847/1538-4357/aa8e9f.

[9] K. Arnold, P. A. R. Ade, A. E. Anthony, D. Barron, D. Boettger, J. Borrill, S. Chapman, Y. Chinone, M. A. Dobbs, J. Errard, G. Fabbian, D. Flanigan, G. Fuller, A. Ghribi, W. Grainger, N. Halverson, M. Hasegawa, K. Hattori, M. Hazumi, W. L. Holzapfel, J. Howard, P. Hyland, A. Jaffe, B. Keating, Z. Kermish, T. Kisner, M. Le Jeune, A. T. Lee, E. Linder, M. Lungu, F. Matsuda, T. Matsumura, N. J. Miller, X. Meng, H. Morii, S. Moyerman, M. J. Myers, H. Nishino, H. Paar, E. Quealy, C. Reichardt, P. L. Richards, C. Ross, A. Shimizu, C. Shimmin, M. Shimon, M. Sholl, P. Siritanasak, H. Speiler, N. Stebor, B. Steinbach, R. Stompor, A. Suzuki, T. Tomaru, C. Tucker, and O. Zahn. The bolometric focal plane array of the polarbear cmb experiment, 2012. URL `http://dx.doi.org/10.1117/12.927057`.

[10] K. S. Arnold. *Design and Deployment of the Polarbear Cosmic Microwave Background Polarization Experiment.* PhD thesis, University of California, Berkeley, 2010.

[11] P. Baldonero, A. Manna, F. Trotta, and R. Flamini. *UWB Multifunction Antennas*. INTECH Open Access Publisher, 2011. ISBN 9789533074528. URL `https://books.google.com/books?id=nrW5oAEACAAJ`.

[12] M Barucci, E Gottardi, I Peroni, and G Ventura. Low temperature thermal conductivity of kapton and upilex. *Cryogenics*, 40(2):145 – 147, 2000. ISSN 0011-2275. doi: https://doi.org/10.1016/S0011-2275(00)00013-8. URL `http://www.sciencedirect.com/science/article/pii/S0011227500000138`.

[13] A. N. Bender, J.-F. Cliche, T. de Haan, M. A. Dobbs, A. J. Gilbert, J. Montgomery, N. Rowlands, G. M. Smecher, K. Smith, and A. Wilson. Digital frequency domain multiplexing readout electronics for the next generation of millimeter telescopes. In *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII*, volume 9153 of *"Proc. SPIE"*, page 91531A, July 2014. doi: 10.1117/12.2054949.

[14] C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, D. N. Spergel, G. S. Tucker, E. Wollack, E. L. Wright, C. Barnes, M. R. Greason, R. S. Hill, E. Komatsu, M. R. Nolta, N. Odegard, H. V. Peiris, L. Verde, and J. L. Weiland. First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results. *"Astrophys. J. Supp."*, 148:1–27, September 2003. doi: 10.1086/377253.

[15] B. A. Benson et al. SPT-3G: a next-generation cosmic microwave background polarization experiment on the South Pole telescope. In *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII*, volume 9153 of *Proc. SPIE Int. Soc. Opt. Eng.*, page 91531P, July 2014. doi: 10.1117/12.2057305.

[16] S. Betts, W. R. Blanchard, R. H. Carnevale, C. Chang, C. Chen, S. Chidzik, L. Ciebiera, P. Cloessner, A. Cocco, A. Cohen, J. Dong, R. Klemmer, M. Komor, C. Gentile, B. Harrop, A. Hopkins, N. Jarosik, G. Mangano, M. Messina, B. Osherson, Y. Raitses, W. Sands, M. Schaefer, J. Taylor, C. G. Tully, R. Woolley, and A. Zwicker. Development of a Relic Neutrino Detection Experiment at PTOLEMY: Princeton Tritium Observatory for Light, Early-Universe, Massive-Neutrino Yield. *ArXiv e-prints*, July 2013.

[17] R. S. Bhatia, S. T. Chase, S. F. Edgington, J. Glenn, W. C. Jones, A. E. Lange, B. Maffei, A. K. Mainzer, P. D. Mauskopf, B. J. Philhour, and B. K. Rownd. A three-stage helium sorption refrigerator for cooling of infrared detectors to 280 mK. *Cryogenics*, 40:685–691, 2000. doi: 10.1016/S0011-2275(00)00072-2.

[18] BICEP2 Collaboration, Keck Array Collaboration, P. A. R. Ade, Z. Ahmed, R. W. Aikin, K. D. Alexander, D. Barkats, S. J. Benton, C. A. Bischoff, J. J. Bock, R. Bowens-Rubin, J. A. Brevik, I. Buder, E. Bullock, V. Buza, J. Connors, B. P. Crill, L. Duband, C. Dvorkin, J. P. Filippini, S. Fliescher, J. Grayson, M. Halpern, S. Harrison, G. C. Hilton, H. Hui, K. D. Irwin, K. S. Karkare, E. Karpel, J. P. Kaufman, B. G. Keating, S. Kefeli, S. A. Kernasovskiy, J. M. Kovac, C. L. Kuo, E. M. Leitch, M. Lueker, K. G. Megerian, C. B. Netterfield, H. T. Nguyen, R. O'Brient, R. W. Ogburn, A. Orlando,

C. Pryke, S. Richter, R. Schwarz, C. D. Sheehy, Z. K. Staniszewski, B. Steinbach, R. V. Sudiwala, G. P. Teply, K. L. Thompson, J. E. Tolan, C. Tucker, A. D. Turner, A. G. Vieregg, A. C. Weber, D. V. Wiebe, J. Willmert, C. L. Wong, W. L. K. Wu, and K. W. Yoon. Improved Constraints on Cosmology and Foregrounds from BICEP2 and Keck Array Cosmic Microwave Background Data with Inclusion of 95 GHz Band. *Physical Review Letters*, 116(3):031302, January 2016. doi: 10.1103/PhysRevLett.116.031302.

[19] BICEP2 Collaboration, Keck Array Collaboration, P. A. R. Ade, Z. Ahmed, R. W. Aikin, K. D. Alexander, D. Barkats, S. J. Benton, C. A. Bischoff, J. J. Bock, R. Bowens-Rubin, J. A. Brevik, I. Buder, E. Bullock, V. Buza, J. Connors, B. P. Crill, L. Duband, C. Dvorkin, J. P. Filippini, S. Fliescher, J. Grayson, M. Halpern, S. Harrison, S. R. Hildebrandt, G. C. Hilton, H. Hui, K. D. Irwin, J. Kang, K. S. Karkare, E. Karpel, J. P. Kaufman, B. G. Keating, S. Kefeli, S. A. Kernasovskiy, J. M. Kovac, C. L. Kuo, E. M. Leitch, M. Lueker, K. G. Megerian, T. Namikawa, C. B. Netterfield, H. T. Nguyen, R. O'Brient, R. W. Ogburn, IV, A. Orlando, C. Pryke, S. Richter, R. Schwarz, C. D. Sheehy, Z. K. Staniszewski, B. Steinbach, R. V. Sudiwala, G. P. Teply, K. L. Thompson, J. E. Tolan, C. Tucker, A. D. Turner, A. G. Vieregg, A. C. Weber, D. V. Wiebe, J. Willmert, C. L. Wong, W. L. K. Wu, and K. W. Yoon. BICEP2/Keck Array VIII: Measurement of Gravitational Lensing from Large-scale B-mode Polarization. *Astrophys.J.*, 833:228, December 2016. doi: 10.3847/1538-4357/833/2/228.

[20] BICEP2/Keck Collaboration, Planck Collaboration, P. A. R. Ade, N. Aghanim, Z. Ahmed, R. W. Aikin, K. D. Alexander, M. Arnaud, J. Aumont, C. Baccigalupi, and et al. Joint Analysis of BICEP2/Keck Array and Planck Data. *Physical Review Letters*, 114(10):101301, March 2015. doi: 10.1103/PhysRevLett.114.101301.

[21] Judd D. Bowman, Alan E. E. Rogers, Raul A. Monsalve, Thomas J. Mozdzen, and Nivedita Mahesh. An absorption profile centred at 78 megahertz in the sky-averaged spectrum. *Nature*, 555(7694):67–70, 2018. doi: 10.1038/nature25792.

[22] D. T. Chuss, A. Ali, M. Amiri, J. Appel, C. L. Bennett, F. Colazo, K. L. Denis, R. Dünner, T. Essinger-Hileman, J. Eimer, P. Fluxa, D. Gothe, M. Halpern, K. Harrington, G. Hilton, G. Hinshaw, J. Hubmayr, J. Iuliano, T. A. Marriage, N. Miller, S. H. Moseley, G. Mumby, M. Petroff, C. Reintsema, K. Rostem, K. U-Yen, D. Watts, E. Wagner, E. J. Wollack, Z. Xu, and L. Zeng. Cosmology large angular scale surveyor (class) focal plane development. *Journal of Low Temperature Physics*, 184(3):759–764, 2016. ISSN 1573-7357. doi: 10.1007/s10909-015-1368-9. URL `http://dx.doi.org/10.1007/s10909-015-1368-9`.

[23] Ari Cukierman, Adrian T. Lee, Christopher Raum, Aritoki Suzuki, and Benjamin Westbrook. Hierarchical sinuous-antenna phased array for millimeter wavelengths. *Applied Physics Letters*, 112(13):132601, 2018. doi: 10.1063/1.5021962. URL `https://doi.org/10.1063/1.5021962`.

[24] Richard H. Cyburt, Brian D. Fields, Keith A. Olive, and Tsung-Han Yeh. Big bang nucleosynthesis: Present status. *Rev. Mod. Phys.*, 88:015004, Feb 2016.

doi: 10.1103/RevModPhys.88.015004. URL https://link.aps.org/doi/10.1103/RevModPhys.88.015004.

[25] S. Das, T. Louis, M. R. Nolta, G. E. Addison, E. S. Battistelli, J. R. Bond, E. Calabrese, D. Crichton, M. J. Devlin, S. Dicker, J. Dunkley, R. Dünner, J. W. Fowler, M. Gralla, A. Hajian, M. Halpern, M. Hasselfield, M. Hilton, A. D. Hincks, R. Hlozek, K. M. Huffenberger, J. P. Hughes, K. D. Irwin, A. Kosowsky, R. H. Lupton, T. A. Marriage, D. Marsden, F. Menanteau, K. Moodley, M. D. Niemack, L. A. Page, B. Partridge, E. D. Reese, B. L. Schmitt, N. Sehgal, B. D. Sherwin, J. L. Sievers, D. N. Spergel, S. T. Staggs, D. S. Swetz, E. R. Switzer, R. Thornton, H. Trac, and E. Wollack. The Atacama Cosmology Telescope: temperature and gravitational lensing power spectrum measurements from three seasons of data. *"J. Cosmol. Astropart. Phys."*, 4:014, April 2014. doi: 10.1088/1475-7516/2014/04/014.

[26] P. de Bernardis, P. A. R. Ade, J. J. Bock, J. R. Bond, J. Borrill, A. Boscaleri, K. Coble, B. P. Crill, G. De Gasperis, P. C. Farese, P. G. Ferreira, K. Ganga, M. Giacometti, E. Hivon, V. V. Hristov, A. Iacoangeli, A. H. Jaffe, A. E. Lange, L. Martinis, S. Masi, P. V. Mason, P. D. Mauskopf, A. Melchiorri, L. Miglio, T. Montroy, C. B. Netterfield, E. Pascale, F. Piacentini, D. Pogosyan, S. Prunet, S. Rao, G. Romeo, J. E. Ruhl, F. Scaramuzzi, D. Sforna, and N. Vittorio. A flat Universe from high-resolution maps of the cosmic microwave background radiation. *Nature*, 404:955–959, April 2000. doi: 10.1038/35010035.

[27] R.H. Dicke, P.J.E. Peebles, P.G. Roll, and D.T. Wilkinson. Cosmic Black-Body Radiation. *Astrophys.J.*, 142:414–419, 1965. doi: 10.1086/148306.

[28] M. A. Dobbs, M. Lueker, K. A. Aird, A. N. Bender, B. A. Benson, L. E. Bleem, J. E. Carlstrom, C. L. Chang, H.-M. Cho, J. Clarke, T. M. Crawford, A. T. Crites, D. I. Flanigan, T. de Haan, E. M. George, N. W. Halverson, W. L. Holzapfel, J. D. Hrubes, B. R. Johnson, J. Joseph, R. Keisler, J. Kennedy, Z. Kermish, T. M. Lanting, A. T. Lee, E. M. Leitch, D. Luong-Van, J. J. McMahon, J. Mehl, S. S. Meyer, T. E. Montroy, S. Padin, T. Plagge, C. Pryke, P. L. Richards, J. E. Ruhl, K. K. Schaffer, D. Schwan, E. Shirokoff, H. G. Spieler, Z. Staniszewski, A. A. Stark, K. Vanderlinde, J. D. Vieira, C. Vu, B. Westbrook, and R. Williamson. Frequency multiplexed superconducting quantum interference device readout of large bolometer arrays for cosmic microwave background measurements. *Review of Scientific Instruments*, 83(7):073113-073113-24, July 2012. doi: 10.1063/1.4737629.

[29] Raymond H DuHamel. Dual polarized sinuous antennas, April 14 1987. US Patent 4,658,262.

[30] J-M Duval, L Duband, and A Attard. Qualification campaign of the 50 mk hybrid sorption-adr cooler for spica/safari. *IOP Conference Series: Materials Science and Engineering*, 101(1):012010, 2015. URL http://stacks.iop.org/1757-899X/101/i=1/a=012010.

[31] J. M. Edwards, R. O'Brient, A. T. Lee, and G. M. Rebeiz. Dual-Polarized Sinuous Antennas on Extended Hemispherical Silicon Lenses. *IEEE Transactions on Antennas and Propagation*, 60:4082–4091, September 2012. doi: 10.1109/TAP.2012.2207048.

[32] D. F. Filipovic, S. S. Gearhart, and G. M. Rebeiz. Double-slot antennas on extended hemispherical and elliptical silicon dielectric lenses. *IEEE Transactions on Microwave Theory and Techniques*, 41(10):1738–1749, Oct 1993. ISSN 0018-9480. doi: 10.1109/22.247919.

[33] D.J. Fixsen. The Temperature of the Cosmic Microwave Background. *Astrophys.J.*, 707:916–920, 2009. doi: 10.1088/0004-637X/707/2/916.

[34] D.J. Fixsen, E.S. Cheng, J.M. Gales, John C. Mather, R.A. Shafer, et al. The Cosmic Microwave Background spectrum from the full COBE FIRAS data set. *Astrophys.J.*, 473:576, 1996. doi: 10.1086/178173.

[35] Jan Forbrich, Helmut Wiesemeyer, Clemens Thum, Arnaud Belloche, and Karl M Menten. Observations of the goldreich-kylafis effect in star-forming regions with xpol at the iram 30 m telescope. *Astronomy & Astrophysics*, 492(3):757–766, 2008.

[36] José M Girart, Richard M Crutcher, and Ramprasad Rao. Detection of polarized co emission from the molecular outflow in ngc 1333 iras 4a. *The Astrophysical Journal Letters*, 525(2):L109, 1999.

[37] E. Grace, J. Beall, J. R. Bond, H. M. Cho, R. Datta, M. J. Devlin, R. Dünner, A. E. Fox, P. Gallardo, M. Hasselfield, S. Henderson, G. C. Hilton, A. D. Hincks, R. Hlozek, J. Hubmayr, K. Irwin, J. Klein, B. Koopman, D. Li, M. Lungu, L. Newburgh, J. P. Nibarger, M. D. Niemack, L. Maurin, J. McMahon, S. Naess, L. A. Page, C. Pappas, B. L. Schmitt, J. Sievers, S. T. Staggs, R. Thornton, J. Van Lanen, and E. J. Wollack. ACTPol: on-sky performance and characterization. In *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII*, volume 9153 of *Proc. SPIE*, page 915310, August 2014. doi: 10.1117/12.2057243.

[38] J. S. Greaves, W. S. Holland, P. Friberg, and W. R. F. Dent. Polarized CO Emission from Molecular Clouds. *"Astrophys. J. Lett."*, 512:L139–L142, February 1999. doi: 10.1086/311888.

[39] M. J. Griffin, J. J. Bock, and W. K. Gear. Relative performance of filled and feedhorn-coupled focal-plane architectures. *Appl. Opt.*, 41:6543–6554, November 2002. doi: 10.1364/AO.41.006543.

[40] Alan H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23:347–356, Jan 1981. doi: 10.1103/PhysRevD.23.347. URL `https://link.aps.org/doi/10.1103/PhysRevD.23.347`.

[41] S. Hanany, P. Ade, A. Balbi, J. Bock, J. Borrill, A. Boscaleri, P. de Bernardis, P. G. Ferreira, V. V. Hristov, A. H. Jaffe, A. E. Lange, A. T. Lee, P. D. Mauskopf, C. B. Netterfield, S. Oh, E. Pascale, B. Rabii, P. L. Richards, G. F. Smoot, R. Stompor,

C. D. Winant, and J. H. P. Wu. MAXIMA-1: A Measurement of the Cosmic Microwave Background Anisotropy on Angular Scales of 10 arcminutes to 5 degrees. *Astrophys. J. Lett.*, 545:L5–L9, December 2000. doi: 10.1086/317322.

[42] D. Hanson, S. Hoover, A. Crites, P. A. R. Ade, K. A. Aird, J. E. Austermann, J. A. Beall, A. N. Bender, B. A. Benson, L. E. Bleem, J. J. Bock, J. E. Carlstrom, C. L. Chang, H. C. Chiang, H.-M. Cho, A. Conley, T. M. Crawford, T. de Haan, M. A. Dobbs, W. Everett, J. Gallicchio, J. Gao, E. M. George, N. W. Halverson, N. Harrington, J. W. Henning, G. C. Hilton, G. P. Holder, W. L. Holzapfel, J. D. Hrubes, N. Huang, J. Hubmayr, K. D. Irwin, R. Keisler, L. Knox, A. T. Lee, E. Leitch, D. Li, C. Liang, D. Luong-Van, G. Marsden, J. J. McMahon, J. Mehl, S. S. Meyer, L. Mocanu, T. E. Montroy, T. Natoli, J. P. Nibarger, V. Novosad, S. Padin, C. Pryke, C. L. Reichardt, J. E. Ruhl, B. R. Saliwanchik, J. T. Sayre, K. K. Schaffer, B. Schulz, G. Smecher, A. A. Stark, K. T. Story, C. Tucker, K. Vanderlinde, J. D. Vieira, M. P. Viero, G. Wang, V. Yefremenko, O. Zahn, and M. Zemcov. Detection of B-Mode Polarization in the Cosmic Microwave Background with Data from the South Pole Telescope. *Physical Review Letters*, 111(14):141301, October 2013. doi: 10.1103/PhysRevLett.111.141301.

[43] G. Harman. *Wire Bonding in Microelectronics*. McGraw-Hill Education, 2009. ISBN 9780071642651. URL `https://books.google.com/books?id=rAmJbSQ7BmEC`.

[44] J. W. Henning, P. Ade, K. A. Aird, J. E. Austermann, J. A. Beall, D. Becker, B. A. Benson, L. E. Bleem, J. Britton, J. E. Carlstrom, C. L. Chang, H.-M. Cho, T. M. Crawford, A. T. Crites, A. Datesman, T. de Haan, M. A. Dobbs, W. Everett, A. Ewall-Wice, E. M. George, N. W. Halverson, N. Harrington, G. C. Hilton, W. L. Holzapfel, J. Hubmayr, K. D. Irwin, M. Karfunkle, R. Keisler, J. Kennedy, A. T. Lee, E. Leitch, D. Li, M. Lueker, D. P. Marrone, J. J. McMahon, J. Mehl, S. S. Meyer, J. Montgomery, T. E. Montroy, J. Nagy, T. Natoli, J. P. Nibarger, M. D. Niemack, V. Novosad, S. Padin, C. Pryke, C. L. Reichardt, J. E. Ruhl, B. R. Saliwanchik, J. T. Sayre, K. K. Schaffer, E. Shirokoff, K. Story, C. Tucker, K. Vanderlinde, J. D. Vieira, G. Wang, R. Williamson, V. Yefremenko, K. W. Yoon, and E. Young. Feedhorn-coupled TES polarimeter camera modules at 150 GHz for CMB polarization measurements with SPTpol. In *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VI*, volume 8452 of *Proc. SPIE*, page 84523A, September 2012. doi: 10.1117/12.927172.

[45] J. W. Henning, J. T. Sayre, C. L. Reichardt, P. A. R. Ade, A. J. Anderson, J. E. Austermann, J. A. Beall, A. N. Bender, B. A. Benson, L. E. Bleem, J. E. Carlstrom, C. L. Chang, H. C. Chiang, H.-M. Cho, R. Citron, C. Corbett Moran, T. M. Crawford, A. T. Crites, T. de Haan, M. A. Dobbs, W. Everett, J. Gallicchio, E. M. George, A. Gilbert, N. W. Halverson, N. Harrington, G. C. Hilton, G. P. Holder, W. L. Holzapfel, S. Hoover, Z. Hou, J. D. Hrubes, N. Huang, J. Hubmayr, K. D. Irwin, R. Keisler, L. Knox, A. T. Lee, E. M. Leitch, D. Li, A. Lowitz, A. Manzotti, J. J. McMahon, S. S. Meyer, L. Mocanu, J. Montgomery, A. Nadolski, T. Natoli, J. P. Nibarger, V. Novosad, S. Padin, C. Pryke, J. E. Ruhl, B. R. Saliwanchik, K. K. Schaffer,

C. Sievers, G. Smecher, A. A. Stark, K. T. Story, C. Tucker, K. Vanderlinde, T. Veach, J. D. Vieira, G. Wang, N. Whitehorn, W. L. K. Wu, and V. Yefremenko. Measurements of the Temperature and E-mode Polarization of the CMB from 500 Square Degrees of SPTpol Data. *"Astrophys. J"*, 852:97, January 2018. doi: 10.3847/1538-4357/aa9ff4.

[46] C. A. Hill, S. Beckman, Y. Chinone, N. Goeckner-Wald, M. Hazumi, B. Keating, A. Kusaka, A. T. Lee, F. Matsuda, R. Plambeck, A. Suzuki, and S. Takakura. Design and development of an ambient-temperature continuously-rotating achromatic half-wave plate for CMB polarization modulation on the POLARBEAR-2 experiment. In *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VIII*, volume 9914 of *"Proc. SPIE"*, page 99142U, July 2016. doi: 10.1117/12.2232280.

[47] G. Hinshaw, J. L. Weiland, R. S. Hill, N. Odegard, D. Larson, C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, N. Jarosik, E. Komatsu, M. R. Nolta, L. Page, D. N. Spergel, E. Wollack, M. Halpern, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright. Five-Year Wilkinson Microwave Anisotropy Probe Observations: Data Processing, Sky Maps, and Basic Results. *Astrophys.J.Supp.*, 180:225–245, February 2009. doi: 10.1088/0067-0049/180/2/225.

[48] W. Hu and M. White. A CMB polarization primer. *New Astronomy*, 2:323–344, October 1997. doi: 10.1016/S1384-1076(97)00022-5.

[49] E. Hubble. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15:168–173, March 1929. doi: 10.1073/pnas.15.3.168.

[50] M. Jones and J. Rawnick. A new approach to broadband array design using tightly coupled elements. In *MILCOM 2007 - IEEE Military Communications Conference*, pages 1–7, Oct 2007. doi: 10.1109/MILCOM.2007.4454764.

[51] R. Keisler, S. Hoover, N. Harrington, J. W. Henning, P. A. R. Ade, K. A. Aird, J. E. Austermann, J. A. Beall, A. N. Bender, B. A. Benson, L. E. Bleem, J. E. Carlstrom, C. L. Chang, H. C. Chiang, H.-M. Cho, R. Citron, T. M. Crawford, A. T. Crites, T. de Haan, M. A. Dobbs, W. Everett, J. Gallicchio, J. Gao, E. M. George, A. Gilbert, N. W. Halverson, D. Hanson, G. C. Hilton, G. P. Holder, W. L. Holzapfel, Z. Hou, J. D. Hrubes, N. Huang, J. Hubmayr, K. D. Irwin, L. Knox, A. T. Lee, E. M. Leitch, D. Li, D. Luong-Van, D. P. Marrone, J. J. McMahon, J. Mehl, S. S. Meyer, L. Mocanu, T. Natoli, J. P. Nibarger, V. Novosad, S. Padin, C. Pryke, C. L. Reichardt, J. E. Ruhl, B. R. Saliwanchik, J. T. Sayre, K. K. Schaffer, E. Shirokoff, G. Smecher, A. A. Stark, K. T. Story, C. Tucker, K. Vanderlinde, J. D. Vieira, G. Wang, N. Whitehorn, V. Yefremenko, and O. Zahn. Measurements of Sub-degree B-mode Polarization in the Cosmic Microwave Background from 100 Square Degrees of SPTpol Data. *Astrophys.J.*, 807:151, July 2015. doi: 10.1088/0004-637X/807/2/151.

[52] N. Kellaris, M. Daal, M. Epland, M. Pepin, O. Kamaev, P. Cushman, E. Kramer, B. Sadoulet, N. Mirabolfathi, S. Golwala, and M. Runyan. Sub-Kelvin Thermal Conductivity and Radioactivity of Some Useful Materials in Low Background Cryogenic

Experiments. *Journal of Low Temperature Physics*, 176:201–208, August 2014. doi: 10.1007/s10909-013-1048-6.

[53] Z. D. Kermish et al. The POLARBEAR experiment. In *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VI*, volume 8452 of *Proc. SPIE Int. Soc. Opt. Eng.*, page 84521C, September 2012. doi: 10.1117/12.926354.

[54] Z.D. Kermish. *The POLARBEAR Experiment: Design and Characterization*. 2012. URL `https://books.google.com/books?id=stZLAQAACAAJ`.

[55] L. Knox and M. S. Turner. Detectability of tensor perturbations through anisotropy of the cosmic background radiation. *Physical Review Letters*, 73:3347–3350, December 1994. doi: 10.1103/PhysRevLett.73.3347.

[56] P L. Richards. Bolometers for infrared and millimeter waves. 76:1 – 24, 08 1994.

[57] S.-P. Lai, J. M. Girart, and R. M. Crutcher. Interferometric Mapping of Magnetic Fields in Star-forming Regions. III. Dust and CO Polarization in DR 21(OH). *"Astrophys. J."*, 598:392–399, November 2003. doi: 10.1086/378769.

[58] Antony Lewis and Sarah Bridle. Cosmological parameters from CMB and other data: A Monte Carlo approach. *Phys. Rev.*, D66:103511, 2002. doi: 10.1103/PhysRevD.66.103511.

[59] T. Louis, E. Grace, M. Hasselfield, M. Lungu, L. Maurin, G. E. Addison, P. A. R. Ade, S. Aiola, R. Allison, M. Amiri, E. Angile, N. Battaglia, J. A. Beall, F. de Bernardis, J. R. Bond, J. Britton, E. Calabrese, H.-m. Cho, S. K. Choi, K. Coughlin, D. Crichton, K. Crowley, R. Datta, M. J. Devlin, S. R. Dicker, J. Dunkley, R. Dünner, S. Ferraro, A. E. Fox, P. Gallardo, M. Gralla, M. Halpern, S. Henderson, J. C. Hill, G. C. Hilton, M. Hilton, A. D. Hincks, R. Hlozek, S. P. P. Ho, Z. Huang, J. Hubmayr, K. M. Huffenberger, J. P. Hughes, L. Infante, K. Irwin, S. Muya Kasanda, J. Klein, B. Koopman, A. Kosowsky, D. Li, M. Madhavacheril, T. A. Marriage, J. McMahon, F. Menanteau, K. Moodley, C. Munson, S. Naess, F. Nati, L. Newburgh, J. Nibarger, M. D. Niemack, M. R. Nolta, C. Nuñez, L. A. Page, C. Pappas, B. Partridge, F. Rojas, E. Schaan, B. L. Schmitt, N. Sehgal, B. D. Sherwin, J. Sievers, S. Simon, D. N. Spergel, S. T. Staggs, E. R. Switzer, R. Thornton, H. Trac, J. Treu, C. Tucker, A. Van Engelen, J. T. Ward, and E. J. Wollack. The Atacama Cosmology Telescope: two-season ACTPol spectra and parameters. *J. Cosmol. Astropart. Phys.*, 6:031, June 2017. doi: 10.1088/1475-7516/2017/06/031.

[60] D. H. Lyth. What Would We Learn by Detecting a Gravitational Wave Signal in the Cosmic Microwave Background Anisotropy? *Physical Review Letters*, 78:1861–1863, March 1997. doi: 10.1103/PhysRevLett.78.1861.

[61] A. Manzotti, K. T. Story, W. L. K. Wu, J. E. Austermann, J. A. Beall, A. N. Bender, B. A. Benson, L. E. Bleem, J. J. Bock, J. E. Carlstrom, C. L. Chang, H. C. Chiang, H.-M. Cho, R. Citron, A. Conley, T. M. Crawford, A. T. Crites, T. de Haan, M. A. Dobbs, S. Dodelson, W. Everett, J. Gallicchio, E. M. George, A. Gilbert, N. W. Halverson,

N. Harrington, J. W. Henning, G. C. Hilton, G. P. Holder, W. L. Holzapfel, S. Hoover, Z. Hou, J. D. Hrubes, N. Huang, J. Hubmayr, K. D. Irwin, R. Keisler, L. Knox, A. T. Lee, E. M. Leitch, D. Li, J. J. McMahon, S. S. Meyer, L. M. Mocanu, T. Natoli, J. P. Nibarger, V. Novosad, S. Padin, C. Pryke, C. L. Reichardt, J. E. Ruhl, B. R. Saliwanchik, J. T. Sayre, K. K. Schaffer, G. Smecher, A. A. Stark, K. Vanderlinde, J. D. Vieira, M. P. Viero, G. Wang, N. Whitehorn, V. Yefremenko, and M. Zemcov. CMB Polarization B-mode Delensing with SPTpol and Herschel. *Astrophys.J.*, 846: 45, September 2017. doi: 10.3847/1538-4357/aa82bb.

[62] John C. Mather, E.S. Cheng, R.A. Shafer, C.L. Bennett, N.W. Boggess, et al. A Preliminary measurement of the Cosmic Microwave Background spectrum by the Cosmic Background Explorer (COBE) satellite. *Astrophys.J.*, 354:L37–L40, 1990. doi: 10.1086/185717.

[63] John C. Mather, E.S. Cheng, D.A. Cottingham, R.E. Eplee, D.J. Fixsen, et al. Measurement of the Cosmic Microwave Background spectrum by the COBE FIRAS instrument. *Astrophys.J.*, 420:439–444, 1994. doi: 10.1086/173574.

[64] T. Matsumura et al. LiteBIRD: Mission Overview and Focal Plane Layout. *Journal of Low Temperature Physics*, 184:824–831, August 2016. doi: 10.1007/s10909-016-1542-8.

[65] A. McKellar. Molecular Lines from the Lowest States of Diatomic Molecules Composed of Atoms Probably Present in Interstellar Space. *Publications of the Dominion Astrophysical Observatory Victoria*, 7, 1941.

[66] M. J. Myers, W. Holzapfel, A. T. Lee, R. O'Brient, P. L. Richards, H. T. Tran, P. Ade, G. Engargiola, A. Smith, and H. Spieler. An antenna-coupled bolometer with an integrated microstrip bandpass filter. *Applied Physics Letters*, 86(11):114103, March 2005. doi: 10.1063/1.1879115.

[67] M.J. Myers. *Antenna-coupled Superconducting Bolometers for Observations of the Cosmic Microwave Background Polarization.* PhD thesis, University of California, Berkeley, 2010.

[68] Roger O'Brient, Peter Ade, Kam Arnold, Jennifer Edwards, Greg Engargiola, William L. Holzapfel, Adrian T. Lee, Michael J. Myers, Erin Quealy, Gabriel Rebeiz, Paul Richards, and Aritoki Suzuki. A dual-polarized broadband planar antenna and channelizing filter bank for millimeter wavelengths. *Applied Physics Letters*, 102(6): 063506, 2013. doi: 10.1063/1.4791692. URL `https://doi.org/10.1063/1.4791692`.

[69] Roger C. OBrient. *A log-periodic focal-plane architecture for cosmic microwave background polarimetry.* PhD thesis, University of California, Berkeley, 2010.

[70] K.A. Olive et al. Review of Particle Physics. *Chin.Phys.*, C38:090001, 2014. doi: 10.1088/1674-1137/38/9/090001.

[71] J.R. Olson. Thermal conductivity of some common cryostat materials between 0.05 and 2 k. *Cryogenics*, 33(7):729 – 731, 1993. ISSN 0011-2275. doi: https://doi.org/10.1016/

0011-2275(93)90027-L. URL `http://www.sciencedirect.com/science/article/pii/001122759390027L`.

[72] C. Patrignani et al. Review of Particle Physics. *Chin. Phys.*, C40(10):100001, 2016. doi: 10.1088/1674-1137/40/10/100001.

[73] Arno A. Penzias and Robert Woodrow Wilson. A Measurement of excess antenna temperature at 4080-Mc/s. *Astrophys.J.*, 142:419–421, 1965. doi: 10.1086/148307.

[74] S. Perlmutter et al. Measurements of Omega and Lambda from 42 high redshift supernovae. *Astrophys.J.*, 517:565–586, 1999. doi: 10.1086/307221.

[75] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. I. R. Alves, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, H. Aussel, and et al. Planck 2013 results. I. Overview of products and scientific results. *"Astron. Astrophys."*, 571: A1, November 2014. doi: 10.1051/0004-6361/201321529.

[76] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. I. R. Alves, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, and et al. Planck 2013 results. XIII. Galactic CO emission. *"Astron. Astrophys."*, 571:A13, November 2014. doi: 10.1051/0004-6361/201321553.

[77] Planck Collaboration, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, K. Benabed, A. Benoit-Lévy, J.-P. Bernard, M. Bersanelli, P. Bielewicz, J. Bobin, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, M. Bridges, C. Burigana, R. C. Butler, J.-F. Cardoso, A. Catalano, A. Challinor, A. Chamballu, H. C. Chiang, L.-Y. Chiang, P. R. Christensen, D. L. Clements, L. P. L. Colombo, F. Couchot, B. P. Crill, A. Curto, F. Cuttaia, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, J. M. Diego, S. Donzelli, O. Doré, X. Dupac, G. Efstathiou, T. A. Enßlin, H. K. Eriksen, F. Finelli, O. Forni, M. Frailis, E. Franceschi, S. Galeotta, K. Ganga, M. Giard, G. Giardino, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gregorio, A. Gruppuso, F. K. Hansen, D. Hanson, D. L. Harrison, G. Helou, S. R. Hildebrandt, E. Hivon, M. Hobson, W. A. Holmes, W. Hovest, K. M. Huffenberger, W. C. Jones, M. Juvela, E. Keihänen, R. Keskitalo, T. S. Kisner, J. Knoche, L. Knox, M. Kunz, H. Kurki-Suonio, A. Lähteenmäki, J.-M. Lamarre, A. Lasenby, R. J. Laureijs, C. R. Lawrence, R. Leonardi, A. Lewis, M. Liguori, P. B. Lilje, M. Linden-Vørnle, M. López-Caniego, P. M. Lubin, J. F. Macías-Pérez, N. Mandolesi, M. Maris, D. J. Marshall, P. G. Martin, E. Martínez-González, S. Masi, M. Massardi, S. Matarrese, P. Mazzotta, P. R. Meinhold, A. Melchiorri, L. Mendes, M. Migliaccio, S. Mitra, A. Moneti, L. Montier, G. Morgante, D. Mortlock, A. Moss, D. Munshi, P. Naselsky, F. Nati, P. Natoli, H. U. Nørgaard-Nielsen, F. Noviello, D. Novikov, I. Novikov, S. Osborne, C. A. Oxborrow, L. Pagano, F. Pajot, D. Paoletti, F. Pasian, G. Patanchon, O. Perdereau, F. Perrotta, F. Piacentini, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, E. Pointecouteau, G. Polenta, N. Ponthieu, L. Popa, G. W. Pratt, G. Prézeau, J.-L. Puget, J. P. Rachen, W. T. Reach, M. Reinecke, S. Ricciardi, T. Riller, I. Ristorcelli, G. Rocha, C. Rosset, J. A.

Rubiño-Martín, B. Rusholme, D. Santos, G. Savini, D. Scott, M. D. Seiffert, E. P. S. Shellard, L. D. Spencer, R. Sunyaev, F. Sureau, A.-S. Suur-Uski, J.-F. Sygnet, J. A. Tauber, D. Tavagnacco, L. Terenzi, L. Toffolatti, M. Tomasi, M. Tristram, M. Tucci, M. Türler, L. Valenziano, J. Valiviita, B. Van Tent, P. Vielva, F. Villa, N. Vittorio, L. A. Wade, B. D. Wandelt, M. White, D. Yvon, A. Zacchei, J. P. Zibin, and A. Zonca. Planck 2013 results. XXVII. Doppler boosting of the CMB: Eppur si muove. *Astron. Astrophys.*, 571:A27, November 2014. doi: 10.1051/0004-6361/201321556.

[78] Planck Collaboration, R. Adam, P. A. R. Ade, N. Aghanim, Y. Akrami, M. I. R. Alves, F. Argüeso, M. Arnaud, F. Arroja, M. Ashdown, and et al. Planck 2015 results. I. Overview of products and scientific results. *Astron. Astrophys.*, 594:A1, September 2016. doi: 10.1051/0004-6361/201527101.

[79] Planck Collaboration, R. Adam, P. A. R. Ade, N. Aghanim, M. I. R. Alves, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, and et al. Planck 2015 results. X. Diffuse component separation: Foreground maps. *"Astron. Astrophys."*, 594:A10, September 2016. doi: 10.1051/0004-6361/201525967.

[80] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, and et al. Planck 2015 results. XX. Constraints on inflation. *Astron. Astrophys.*, 594:A20, September 2016. doi: 10.1051/0004-6361/201525898.

[81] Polarbear Collaboration, P. A. R. Ade, Y. Akiba, A. E. Anthony, K. Arnold, M. Atlas, D. Barron, D. Boettger, J. Borrill, S. Chapman, Y. Chinone, M. Dobbs, T. Elleflot, J. Errard, G. Fabbian, C. Feng, D. Flanigan, A. Gilbert, W. Grainger, N. W. Halverson, M. Hasegawa, K. Hattori, M. Hazumi, W. L. Holzapfel, Y. Hori, J. Howard, P. Hyland, Y. Inoue, G. C. Jaehnig, A. H. Jaffe, B. Keating, Z. Kermish, R. Keskitalo, T. Kisner, M. Le Jeune, A. T. Lee, E. M. Leitch, E. Linder, M. Lungu, F. Matsuda, T. Matsumura, X. Meng, N. J. Miller, H. Morii, S. Moyerman, M. J. Myers, M. Navaroli, H. Nishino, A. Orlando, H. Paar, J. Peloton, D. Poletti, E. Quealy, G. Rebeiz, C. L. Reichardt, P. L. Richards, C. Ross, I. Schanning, D. E. Schenck, B. D. Sherwin, A. Shimizu, C. Shimmin, M. Shimon, P. Siritanasak, G. Smecher, H. Spieler, N. Stebor, B. Steinbach, R. Stompor, A. Suzuki, S. Takakura, T. Tomaru, B. Wilson, A. Yadav, and O. Zahn. A Measurement of the Cosmic Microwave Background B-mode Polarization Power Spectrum at Sub-degree Scales with POLARBEAR. *Astrophys.J.*, 794:171, October 2014. doi: 10.1088/0004-637X/794/2/171.

[82] C M Posada, P A R Ade, Z Ahmed, K Arnold, J E Austermann, A N Bender, L E Bleem, B A Benson, K Byrum, J E Carlstrom, C L Chang, H M Cho, S T Ciocys, J F Cliche, T M Crawford, A Cukierman, D Czaplewski, J Ding, R Divan, T de Haan, M A Dobbs, D Dutcher, W Everett, A Gilbert, N W Halverson, N L Harrington, K Hattori, J W Henning, G C Hilton, W L Holzapfel, J Hubmayr, K D Irwin, O Jeong, R Keisler, D Kubik, C L Kuo, A T Lee, E M Leitch, S Lendinez, S S Meyer, C S Miller, J Montgomery, M Myers, A Nadolski, T Natoli, H Nguyen, V Novosad, S Padin, Z Pan, J Pearson, J E Ruhl, B R Saliwanchik, G Smecher, J T Sayre, E Shirokoff, L Stan, A A

Stark, J Sobrin, K Story, A Suzuki, K L Thompson, C Tucker, K Vanderlinde, J D Vieira, G Wang, N Whitehorn, V Yefremenko, K W Yoon, and K E Ziegler. Fabrication of large dual-polarized multichroic tes bolometer arrays for cmb measurements with the spt-3g camera. *Superconductor Science and Technology*, 28(9):094002, 2015. URL `http://stacks.iop.org/0953-2048/28/i=9/a=094002`.

[83] J. R. Pritchard and A. Loeb. 21 cm cosmology in the 21st century. *Reports on Progress in Physics*, 75(8):086901, August 2012. doi: 10.1088/0034-4885/75/8/086901.

[84] E. E. Quealy. *The POLARBEAR Cosmic Microwave Background Polarization Experiment and Anti-Reflection Coatings for Millimeter Wave Observations*. PhD thesis, University of California, Berkeley, 2012.

[85] A. G. Riess, L. M. Macri, S. L. Hoffmann, D. Scolnic, S. Casertano, A. V. Filippenko, B. E. Tucker, M. J. Reid, D. O. Jones, J. M. Silverman, R. Chornock, P. Challis, W. Yuan, P. J. Brown, and R. J. Foley. A 2.4% Determination of the Local Value of the Hubble Constant. *Astrophys.J.*, 826:56, July 2016. doi: 10.3847/0004-637X/826/1/56.

[86] Adam G. Riess et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astron.J.*, 116:1009–1038, 1998. doi: 10.1086/300499.

[87] K. Rotermund, B. Barch, S. Chapman, K. Hattori, A. Lee, N. Palaio, I. Shirley, A. Suzuki, and C. Tran. Planar Lithographed Superconducting LC Resonators for Frequency-Domain Multiplexed Readout Systems. *Journal of Low Temperature Physics*, 184:486–491, July 2016. doi: 10.1007/s10909-016-1554-4.

[88] M. C. Runyan and W. C. Jones. Thermal conductivity of thermally-isolating polymeric and composite structural support materials between 0.3 and 4 K. *Cryogenics*, 48:448–454, September 2008. doi: 10.1016/j.cryogenics.2008.06.002.

[89] U. Seljak and C. M. Hirata. Gravitational lensing as a contaminant of the gravity wave signal in the CMB. *Phys. Rev. D*, 69(4):043005, February 2004. doi: 10.1103/PhysRevD.69.043005.

[90] Uros Seljak and Matias Zaldarriaga. Signature of gravity waves in polarization of the microwave background. *Phys. Rev. Lett.*, 78:2054–2057, 1997. doi: 10.1103/PhysRevLett.78.2054.

[91] B. D. Sherwin and M. Schmittfull. Delensing the CMB with the cosmic infrared background. *Phys. Rev. D*, 92(4):043005, August 2015. doi: 10.1103/PhysRevD.92.043005.

[92] S. M. Simon, S. Raghunathan, J. W. Appel, D. T. Becker, L. E. Campusano, H. M. Cho, T. Essinger-Hileman, S. P. Ho, K. D. Irwin, N. Jarosik, A. Kusaka, M. D. Niemack, G. W. Nixon, M. R. Nolta, L. A. Page, G. A. Palma, L. P. Parker, J. L. Sievers, S. T. Staggs, and K. Visnjic. Characterization of the Atacama B-mode Search. In *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII*, volume 9153 of *Proc. SPIE*, July 2014.

[93] P. Siritanasak, C. Aleman, K. Arnold, A. Cukierman, M. Hazumi, K. Kazemzadeh, B. Keating, T. Matsumura, A. T. Lee, C. Lee, E. Quealy, D. Rosen, N. Stebor, and A. Suzuki. The Broadband Anti-reflection Coated Extended Hemispherical Silicon Lenses for Polarbear-2 Experiment. *Journal of Low Temperature Physics*, 184:553–558, August 2016. doi: 10.1007/s10909-015-1386-7.

[94] K. M. Smith, D. Hanson, M. LoVerde, C. M. Hirata, and O. Zahn. Delensing CMB polarization with external datasets. *"J. Cosmol. Astropart. Phys."*, 6:014, June 2012. doi: 10.1088/1475-7516/2012/06/014.

[95] G. F. Smoot, M. V. Gorenstein, and R. A. Muller. Detection of anisotropy in the cosmic blackbody radiation. *Physical Review Letters*, 39:898–901, October 1977. doi: 10.1103/PhysRevLett.39.898.

[96] G. F. Smoot, C. L. Bennett, A. Kogut, E. L. Wright, J. Aymon, N. W. Boggess, E. S. Cheng, G. de Amici, S. Gulkis, M. G. Hauser, G. Hinshaw, P. D. Jackson, M. Janssen, E. Kaita, T. Kelsall, P. Keegstra, C. Lineweaver, K. Loewenstein, P. Lubin, J. Mather, S. S. Meyer, S. H. Moseley, T. Murdock, L. Rokke, R. F. Silverberg, L. Tenorio, R. Weiss, and D. T. Wilkinson. Structure in the COBE differential microwave radiometer first-year maps. *Astrophys. J. Lett.*, 396:L1–L5, September 1992. doi: 10.1086/186504.

[97] K. T. Story, C. L. Reichardt, Z. Hou, R. Keisler, K. A. Aird, B. A. Benson, L. E. Bleem, J. E. Carlstrom, C. L. Chang, H.-M. Cho, T. M. Crawford, A. T. Crites, T. de Haan, M. A. Dobbs, J. Dudley, B. Follin, E. M. George, N. W. Halverson, G. P. Holder, W. L. Holzapfel, S. Hoover, J. D. Hrubes, M. Joy, L. Knox, A. T. Lee, E. M. Leitch, M. Lueker, D. Luong-Van, J. J. McMahon, J. Mehl, S. S. Meyer, M. Millea, J. J. Mohr, T. E. Montroy, S. Padin, T. Plagge, C. Pryke, J. E. Ruhl, J. T. Sayre, K. K. Schaffer, L. Shaw, E. Shirokoff, H. G. Spieler, Z. Staniszewski, A. A. Stark, A. van Engelen, K. Vanderlinde, J. D. Vieira, R. Williamson, and O. Zahn. A Measurement of the Cosmic Microwave Background Damping Tail from the 2500-Square-Degree SPT-SZ Survey. *"Astrophys. J."*, 779:86, December 2013. doi: 10.1088/0004-637X/779/1/86.

[98] A. Suzuki. *Multichroic Bolometric Detector Architecture for Cosmic Microwave Background Polarimetry Experiments*. PhD thesis, University of California, Berkeley, 2013.

[99] A. Suzuki, K. Arnold, J. Edwards, G. Engargiola, A. Ghribi, W. Holzapfel, A. Lee, X. Meng, M. Myers, R. O'Brient, E. Quealy, G. Rebeiz, and P. Richards. Multichroic Dual-Polarization Bolometric Focal Plane for Studies of the Cosmic Microwave Background. *Journal of Low Temperature Physics*, 167:852–858, June 2012. doi: 10.1007/s10909-012-0602-y.

[100] A. Suzuki et al. The Polarbear-2 and the Simons Array Experiments. *Journal of Low Temperature Physics*, 184:805–810, August 2016. doi: 10.1007/s10909-015-1425-4.

[101] S. Takakura, M. Aguilar, Y. Akiba, K. Arnold, C. Baccigalupi, D. Barron, S. Beckman, D. Boettger, J. Borrill, S. Chapman, Y. Chinone, A. Cukierman, A. Ducout, T. Elleflot,

J. Errard, G. Fabbian, T. Fujino, N. Galitzki, N. Goeckner-Wald, N. W. Halverson, M. Hasegawa, K. Hattori, M. Hazumi, C. Hill, L. Howe, Y. Inoue, A. H. Jaffe, O. Jeong, D. Kaneko, N. Katayama, B. Keating, R. Keskitalo, T. Kisner, N. Krachmalnicoff, A. Kusaka, A. T. Lee, D. Leon, L. Lowry, F. Matsuda, T. Matsumura, M. Navaroli, H. Nishino, H. Paar, J. Peloton, D. Poletti, G. Puglisi, C. L. Reichardt, C. Ross, P. Siritanasak, A. Suzuki, O. Tajima, S. Takatori, and G. Teply. Performance of a continuously rotating half-wave plate on the POLARBEAR telescope. *"J. Cosmol. Astropart. Phys."*, 5:008, May 2017. doi: 10.1088/1475-7516/2017/05/008.

[102] Steven Weinberg. Universal neutrino degeneracy. *Phys. Rev.*, 128:1457–1473, Nov 1962. doi: 10.1103/PhysRev.128.1457. URL `https://link.aps.org/doi/10.1103/PhysRev.128.1457`.

[103] B. Westbrook, A. Cukierman, A. Lee, A. Suzuki, C. Raum, and W. Holzapfel. Development of the next generation of multi-chroic antenna-coupled transition edge sensor detectors for cmb polarimetry. *Journal of Low Temperature Physics*, 184(1): 74–81, Jul 2016. ISSN 1573-7357. doi: 10.1007/s10909-016-1508-x. URL `https://doi.org/10.1007/s10909-016-1508-x`.

[104] D. P. Woody and P. L. Richards. Spectrum of the cosmic background radiation. *Phys. Rev. Lett.*, 42:925–929, Apr 1979. doi: 10.1103/PhysRevLett.42.925. URL `https://link.aps.org/doi/10.1103/PhysRevLett.42.925`.

[105] M. Zaldarriaga and U. Seljak. Gravitational lensing effect on cosmic microwave background polarization. *Phys. Rev. D*, 58(2):023003, July 1998. doi: 10.1103/PhysRevD.58.023003.

[106] J. Zmuidzinas and H. G. LeDuc. Quasi-optical slot antenna sis mixers. *IEEE Transactions on Microwave Theory and Techniques*, 40(9):1797–1804, Sep 1992. ISSN 0018-9480. doi: 10.1109/22.156607.

Appendix A

LC layout

We advertise here a script for generating *LC*-chip masks in L-Edit, which is a layout software produced by Mentor Graphics. The script is written in `C++` as a T-cell in L-Edit, which means that there are special packages available that are particular to L-Edit and that the script is compiled and executed within L-Edit. This script is approximately 1000 lines long, so we do not include it here. To request a copy, send an email to `ajcukierman@berkeley.edu`.

The script produces a complete layout for an *LC* chip including all of the resonators, the wiring and the bond pads. The resonator type, the line width, the space between resonators and the frequency schedule are taken as inputs. We developed this script to automate what was previously the tedious process of designing *LC* chips manually.

In L-Edit 2016.1, the following line is included automatically when a new T-cell is created:

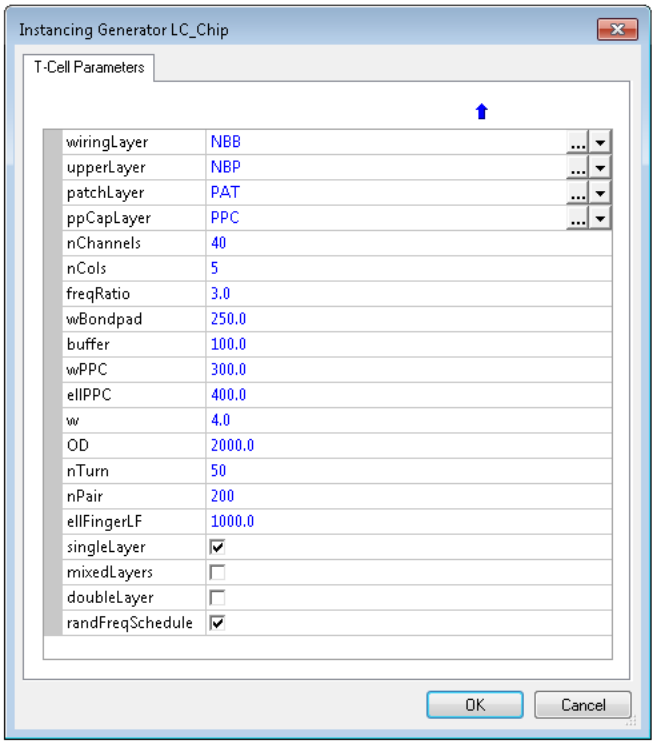
```
LC_InitializeStateFromCell(cellCurrent);
```

In L-Edit 16.0, this line must be replaced with the following:

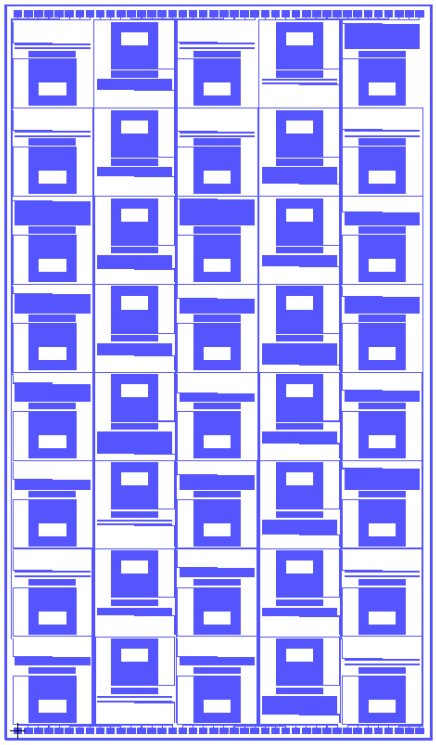
```
LC_InitializeState();
LC_CurrentCell = cellCurrent;
```

The replacement actually works in L-Edit 2016.1, so the program is backwards but not forwards compatible. The moral is to be aware that the proper syntax changes as L-Edit is updated.

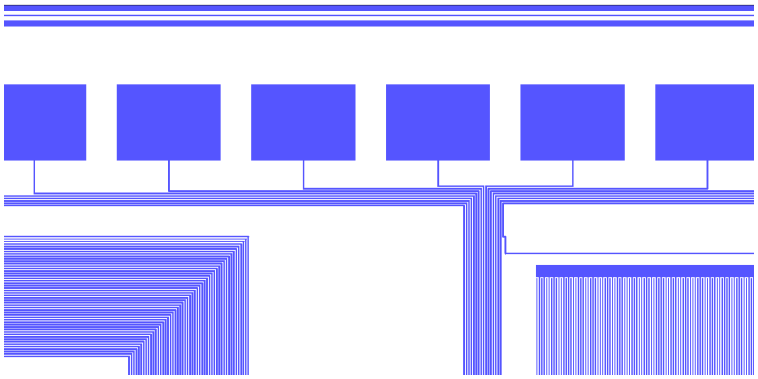
When the T-cell is instanced, several parameters are presented in a window such as the one shown in Fig. A.1(a). The layout that this particular set of parameters would produce is shown in Figs. A.1(b) and A.1(c). Some of the parameters are irrelevant, unless certain booleans are set to true. For now, suppose we have set `singleLayer` to true as in Fig. A.1(a). In this case, the T-cell will create single-layer spiral inductors and interdigital capacitors similar to those in use in PB-2 and SPT-3G. The features are created on the layer set by the parameter `wiringLayer`. The other layer parameters are irrelevant when `singleLayer` is true. The integer `nChannels` sets the number of resonators. The resonators are arranged in rectilinear rows and columns. The integer `nCols` sets the number of columns. The number of rows is chosen automatically to accommodate all of the resonators. If `nCols` is not an integer divisor of `nChannels`, then the last column is left only partially filled. An example in which `nCols` has been set to 7 is shown in Fig. A.2(a). The inductor is the same for all resonators, and its geometry is set by parameters, which we will discuss below. The



(a) Parameters defining the *LC* chip. All lengths are in microns.

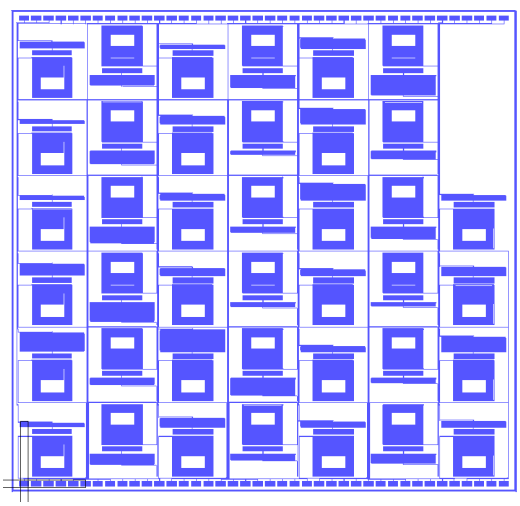


(b) Layout produced by the parameters of Fig. A.1(a). The rectangles with large holes in them are spiral inductors, whose traces are not resolved in this image. The wide solid-looking rectangles are interdigital capacitors, whose traces are also not resolved in this image.

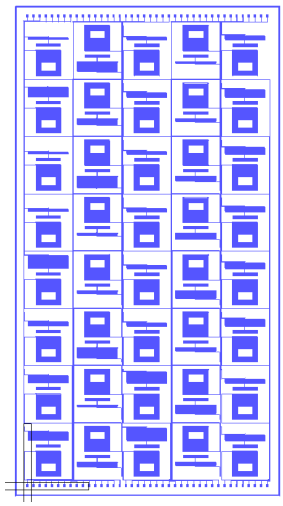


(c) Detail of the layout showing a corner of the spiral inductor (*bottom left*), a corner of the interdigital capacitor (*bottom right*), bond pads (*middle, running horizontally*) and the dicing lane (*top*). Notice the wiring that connects the resonators to the bond pads, which is all drawn automatically.

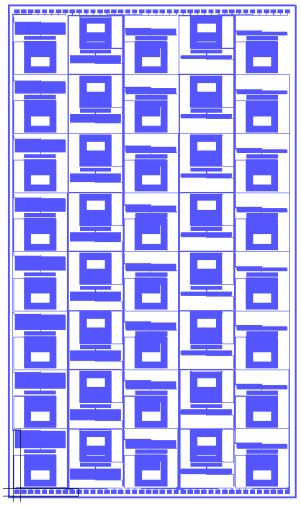
Figure A.1: Example of how the *LC* layout code is used. The default parameters prescribe a random frequency schedule. The capacitor length controls the frequency of each resonator, so the capacitors are seen to be of random lengths. The inductance is held constant.



(a) Prescribing seven columns instead of five. The code is able to handle a number of columns that does not evenly divide the total number of resonators.



(b) Increasing the buffer distance from 100 to 300 μm creates more space between components.



(c) Monotonic frequency schedule instead of random. The capacitor lengths now increase monotonically from the bottom left up each column.

Figure A.2: Variations on the *LC* chip instanced in Fig. A.1.

geometry of the largest capacitor, which determines the lowest resonant frequency, is also set by parameters discussed below. The capacitor length is then varied to create resonances at higher frequencies. The ratio of the highest to lowest frequency is set by `freqRatio`. Notice, therefore, that the explicit frequencies are not set by this code. It is the user's responsibility to set the geometry of the inductor and lowest-frequency capacitor to achieve the correct *absolute* frequency range. All lengths are set in microns. The width of the bond pads in the direction toward and away from the resonators is set by `wBondpad`. The extra space between resonators is set by `buffer`. This also controls the space between bond pads. An example, in which the buffer has been increased to 300 μm, is shown in Fig. A.2(b). The variables ending in `...PPC` are irrelevant when `singleLayer` is true. The line width and line spacing used in the spiral inductors, interdigital capacitors and wiring is set by `w`. For PB-2 and SPT-3G, we use 4 μm, although an aspiration is to reduce the size of the resonators by decreasing the line width and spacing to, e.g., 2 μm. The outer diameter of the spiral inductor is set by `OD`. The spiral inductor is actually rectangular, so this "outer diameter" really refers to the side length. The number of turns in the spiral inductor is set by the integer `nTurn`. Given an outer diameter and line width, there is a maximum number of turns possible, since the traces are spiraling toward the center. The program does not throw an error if `nTurns` is too large for the given `OD`. It is the user's responsibility to choose these variables appropriately. The single-layer inductors require a wire bond to the center of the spiral in order to complete the circuit. Increasing the number of turns will reduce the size of the central bond pad. The interdigital capacitors are defined by the

number of pairs of fingers, which is set by the integer `nPair`, and the length of the fingers, which is set by `ellFingerLF`. The frequency schedule, i.e., the ordering of the resonant frequencies in physical space, can be toggled between a monotonic setting, in which case the resonant frequencies increase from the bottom left moving upwards in each column as shown in Fig. A.2(c), and a random setting, which is what is shown in Fig. A.1(b). This toggle is set by the boolean `randFreqSchedule`.

For a custom frequency schedule, it will be necessary to modify the script directly. The following lines are currently used to define the frequency schedule:

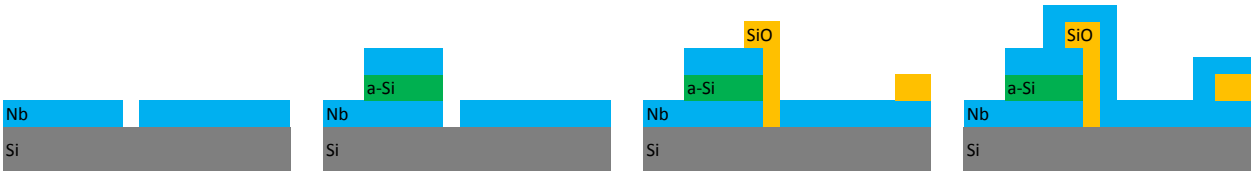
```
int *freqSchedule;
freqSchedule = new int[nChannels];
for (int i = 0; i < nChannels; i++) {
  freqSchedule[i] = i;
  }

if (randFreqSchedule) {
  KnuthShuffle(freqSchedule,nChannels);
  }
```

The array `freqSchedule` simply contains integers between `0` and `nChannels - 1`. The integers can be permuted randomly by setting `randFreqSchedule` to true. The frequencies are taken to be logarithmically spaced between the lowest frequency, which is defined by the geometric parameters described above, and the highest frequency, which is related to the lowest frequency by the parameter `freqRatio`. Again, the program doesn't know anything about the absolute resonant frequencies. Everything is relative to the lowest-frequency resonator, and the other resonances are accessed by varying the capacitor length only. For a custom frequency schedule, it is only necessary to alter the `freqSchedule` array after the lines shown above. If a logarithmic frequency schedule is unacceptable, it will be necessary to alter the code in a more serious way. In that case, send an email to `ajcukierman@berkeley.edu`.

It is extremely desirable to reduce the size of the *LC* resonators, so the wafer modules can be more compactly integrated. One method, which was mentioned above, is to reduce the line width and spacing. This reduces quadratically the footprint of the spiral inductors and interdigital capacitors. This move is non-trivial, however, since the fractional tolerance on the line widths remains the same, which means that the *absolute* tolerance is smaller. Another angle is to include additional layers on the chip, which would allow for parallel-plate capacitors and double-layer spiral inductors. The parallel-plate capacitors have a much larger capacitance per unit area as compared with the interdigital capacitors. Stacking two spiral inductors on top of each other and connecting them at the center boosts the inductance by a factor of four. Making all of these changes simultaneously would allow us to reduce the required area of a PB-2 *LC* chip by more than order of magnitude.

We have been developing a process with a company called STARCryo to test all of these avenues for resonator compactification. The process steps are illustrated in Fig. A.3. In order to avoid oxide growth between the electrodes of the parallel-plate capacitor, which would introduce loss and degrade the quality factor of the resonators, the capacitor dielectric and top-layer niobium are deposited in the same chamber without breaking vacuum. Most likely, we will use amorphous silicon (a – Si) as the capacitor dielectric. A sputter etch can be



(a) Deposit and pattern bottom-layer niobium.

(b) Deposit amorphous silicon and top-layer niobium together and pattern.

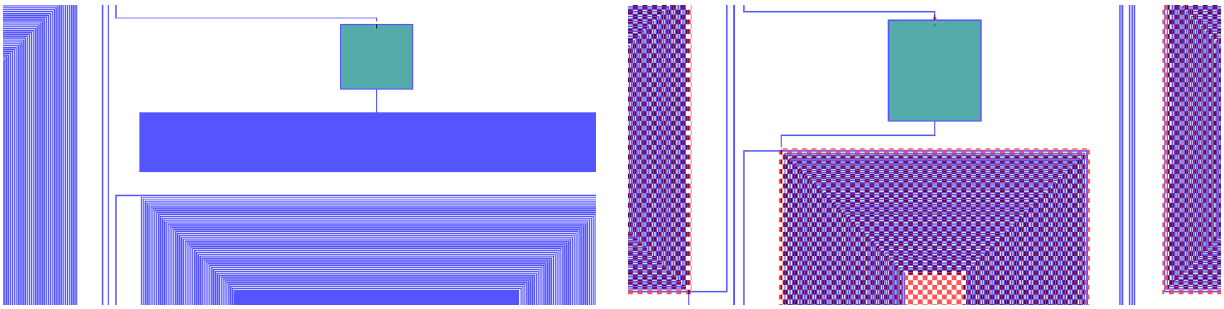
(c) Deposit and pattern patch dielectric (SiO).

(d) Deposit and pattern patch niobium.

Figure A.3: Process for fabricating multilayer inductors and capacitors. The parallel-plate capacitors are formed with amorphous silicon (a − Si) as a dielectric. To avoid oxide growth, the a − Si and top-layer niobium are deposited without breaking vacuum and must, therefore, be patterned together. Dielectric patches are needed to connect the top-layer niobium to the wire traces without shorting out the capacitors. These extra layers allow for double-layer spiral inductors, where the two layers are connected by a via in the patch dielectric.

performed before depositing the dielectric, so any oxide growth on the bottom-layer niobium is removed. To make a connection from the wire traces to the top-layer niobium, it is necessary to fabricate small patches of dielectric to allow a third niobium layer to connect without shorting out the capacitor. The loss requirements on this patch dielectric are not as stringent, so we will begin by choosing a dielectric which is relatively convenient to deposit. This might be a simple silicon oxide (SiO).

The T-cell allows for the use of these multilayer components. By toggling `mixedLayers` to true, the interdigital capacitors are replaced by parallel-plate capacitors. If `singleLayer` is still true, then it will trump `mixedLayers`. Similarly, `mixedLayers` will trump `doubleLayer`. When only `doubleLayer` is true, the resonators are composed of parallel-plate capacitors and *double-layer* spiral inductors. The mask layer that defines the parallel-plate capacitors is set by `ppCapLayer`. Recall that this mask defines both the capacitor dielectric and top-layer electrode. The patch dielectric is defined by the layer set by `patchLayer`, and the final metal layer, which is mainly responsible for connecting the top-layer electrodes to the wiring, is set by `upperLayer`. Some images of the layout for these multilayer variations are shown in Fig. A.4. The double-layer spiral inductor includes a via near the center, which connects the bottom and top layers. This removes the need for the on-chip wire bond used with the single-layer spiral inductors.



(a) The "mixedLayers" setting uses parallel-plate capacitors in place of interdigital capacitors.

(b) The "doubleLayer" setting uses parallel-plate capacitors in place of interdigital capacitors as well as double-layer spiral inductors in place of single-layer.

Figure A.4: The code allows for parallel-plate capacitors and double-layer spiral inductors.