Advanced Cryocoolers For Next Generation VLA

Larry R. D'Addario
California Institute of Technology
2017 September 29

I. Introduction and Historical Review

This report gives the results of an unfunded Community Study on behalf of the Next Generation VLA (ngVLA) project of the National Radio Astronomy Observatory (NRAO). It is based on the Call for Proposals of August 2016 [1] and a proposal submitted on 2016 September 26. It was approved by the NRAO on 2016 November 2 [2] and was carried out between then and 2017 August 31 (10 months). A preliminary report [3] was presented in Socorro on 2017 June 26.

Cryocoolers have been used in radio astronomy since the 1960s [4], and perhaps earlier. Commercially-available two-stage Gifford-McMahon (GM) refrigerators were used to cool mixers and parametric amplifiers, and later GaAs FET amplifiers, to ~20K. Prior to that, paramps and masers were cooled with liquid cryogens [5-6], and for many years liquid helium continued to be used to reach ~4K for superconducting magnets on masers. By 1975, closed-cycle cryocoolers were developed at the DSN for cooling masers to 4K [7], and similar systems began to be adopted at radio observatories. These involved the construction of a custom-built Joule-Thompson (JT) refrigeration stage that was added to a commercial two-stage GM cryocooler. To support such non-commercially-available systems, cryocooler development and construction laboratories were maintained at NRAO (Green Bank, WV) and CSIRO (Epping, NSW, Australia) through the mid-1990s. These laboratories were able to design and build the JT refrigeration stages from basic materials and to construct large helium compressors from parts intended for other purposes.

Masers were soon supplanted in cm-wavelength radio astronomy by HEMT amplifiers [8] due to their larger instantaneous bandwidth; these work well when cooled to 20K, avoiding the need for a three-stage cryocooler and making the commercial two-stage GM devices adequate. However, at about the same time (mid-1980s), superconducting (SIS) mixers were being developed for use at shorter (mm) wavelengths [9], where they were able to provide lower noise temperature than HEMT amplifiers. These required cooling to ~4K in order to be well below the critical temperature of Pb (7.2K) or Nb (8.7K), so custom-built JT stages continued to be needed.

By the early 1990s, academic and industrial research in cryocooler technology had found that it is possible to achieve much lower temperatures with two-stage GM crycoolers by using rare-earth alloys in the regenerator (initially ErNi or ErNiCo, and later HoCu) [10]. However, this technology was not adopted in radio astronomy until after it was commercialized by Sumitomo is 1997. The ALMA telescope, which makes extensive use of SIS mixers, choose to use a 3-stage GM cryocooler [11] that was custom-manufactured by Sumitomo to meet their requirements for cooling at ~100K and ~15K as well as 4K.

Meanwhile, outside radio astronomy, extensive research and development in cryocooler technology was occurring for military and space applications. This led to devices having much better power efficiency than the GM coolers used at radio telescopes, but they were expensive. The commercial devices used in radio astronomy were developed for other applications with mass markets (>20,000 units per year), yielding lower cost due to economies of scale.

By the late 1990s, the availability of commercial products covering the full range of temperatures needed for radio telescope receivers led to the abandonment of the cryocooler labs at NRAO and CSIRO, so that today the worldwide radio astronomy community has very little in-house development capacity in this area. It is even difficult for today's observatories to test the commercial offerings adequately [12].

During most of this history, all of radio astronomy needed only a few cryocoolers, since many telescopes used single antennas or small arrays. The original VLA, with 28 antennas, had one cryocooler per antenna — still a small number. The power consumption of the devices being used (2-5 kW each) was easily tolerated. After several rounds of receiver upgrades, the VLA now has 8 cryocoolers per antenna or 224 in total, and the power consumption has become a significant burden. So has the maintenance cost, since each of the GM cryocoolers has seals and valves that require replacement every 12-24 months. Reliability is also an issue; problems with the cryogenic subsystem are one of the main reasons for technician visits to antennas.

We are now embarking on a big expansion of the number of antennas per telescope, with SKA1 and ngVLA each expected to have about 200 dishes with cooled receivers. These telescopes also include long baselines, so some antennas will be in remote places. Obtaining adequate long-term funding to operate such telescopes is expected to be challenging. For these reasons, it is worthwhile to consider whether cryocooler technologies not previously used in radio astronomy can significantly reduce the operating costs of power and maintenance. Reduced operating cost may justify higher capital cost. It may be cost-effective to develop a custom design, since the development cost can be amortized across a significant number of units. This study was intended to explore these issues, with emphasis on application to the ngVLA.

I begin by considering *de novo* the question of what temperatures are really required and how much heat must be extracted (Section II). Given ngVLA's frequency range (1.2 to 116 GHz), I assume that all front ends will use transistor amplifiers (not superconducting devices). Next, I review the most promising cryocooler types from first principles (III). Then I look at what specific cryocooler designs are available, considering not only commercial products but also published designs developed through academic research and for military or space applications (IV). Finally, I provide a set of recommendations for ngVLA (V).

II. Requirements

A. Noise Temperature vs. Physical Temperature

In general, the colder we make the temperature-sensitive components of a receiver, the lower its noise temperature. But getting colder can be costly, so there is a point of diminishing returns where the incremental benefit is not justified by the incremental cost. We want to understand that tradeoff.

Historically, receivers were operated at the lowest temperature that a convenient cryocooler (or liquid cryogen) could produce. The possible temperatures tended to be roughly quantized, e.g., LN_2 at 77K and LHe at 4.2K. Single-stage GM crycoolers (with steel mesh or lead ball regenerators) could reach only about 50K, and ran at about 70-80K with significant load — about the same as LN_2. Two-stage GM cryocoolers (prior to the advent of rare-earth regenerators) had a minimum temperature of about 15K and operated near 20K with significant load. Three-stage GM/JT cryocoolers and LHe could get below 4K, but were expensive, power hungry, and somewhat unreliable. So there were basically three temperatures below 100K for which practical receivers were designed: ~77K, ~20K, and ~4K. Cooling to 4K was hard to justify and was done only if the receiver would not otherwise function, such as when it depends on superconducting devices.

In reality, a continuous range of temperatures is available with mechanical cryocoolers. A given machine has a minimum (no-load) temperature, but its cold stage can be raised to several times that absolute temperature by applying sufficient heat load. There will be a load at which efficiency is maximized. Therefore, we consider first noise temperature vs. physical temperature, assuming that any physical temperature can be produced (at some cost that that decreases with increasing temperature).

A fair analysis considers the system noise temperature T_{sys}, rather than the receiver or cooled amplifier noise temperature (T_r or T_a), because of diminishing returns; e.g., if T_{sys} = 100K with T_a = 5K, little is gained by cooling the amplifier further to achieve T_a = 2K. Some components of T_{sys} cannot be reduced by cryocooling regardless of temperature, including noise from the sky and thermal noise from the ground near the antenna. For other components, cryocooling may be completely impractical (like the reflectors of a large dish) and for still others it may be possible but not cost-effective (like a large feed horn). Often a compromise design makes sense, where small but noisy components (like an amplifier) are cooled to a low temperature and large but not-very-noisy components are cooled to a higher temperature. These complexities make it difficult to consider all possibilities, so the analysis here is limited to a simplified model based on the following assumptions, which I believe are representative of what we expect for ngVLA.

1. Frequency range 1.2 to 50 GHz (neglecting the 70-116 GHz range of ngVLA for simplicity).
2. Sky noise from the Butler model [13] for the VLA site with 6 mm precipitable water vapor.
3. Antenna pointed at 45º elevation angle.
4. Spillover (ground radiation) noise is 3K (1%), independent of frequency. The total spillover loss of the feed-antenna combination is likely to be 5% to 10%, but it is assumed that most of this couples to the sky rather than to the ground.
5. The receiver is in a vacuum chamber with an RF window that is at 300K and has 1% absorptive loss, independent of frequency.
6. There is only one temperature, T_{phys}, to which both the feed antenna and low-noise amplifier (LNA) are cooled. (Depending on the receiver design, this may not be practical at the low end of the frequency range where the feed is relatively large.)
7. The feed has a resistive loss of 1%, independent of frequency, and is connected directly to the LNA with no intervening components.
8. The LNA is based on InP HEMT transistors from the "cryo3" wafer, since those transistors provide near-state-of-the-art performance in this frequency range and are well modeled [14].
9. The LNA noise is dominated by that of its input transistor, which is always operated at its optimum source impedance (for each frequency and temperature) so as to achieve its minimum noise temperature.
10. The LNA transistors are subject to self-heating of up to 18K [15]. This means that the active parts of the transistor remain at 18K when its substrate is cooled to 0 K.

Only the noise of the LNA and the thermal noise of the feed are affected by the cryocooling temperature T_{phys}. Everything else is independent of T_{phys} but sky noise depends on frequency.

The LNA noise temperature is taken to be the minimum noise temperature of its input transistor (assumption 9), which is modeled as explained in [14] to include thermal noise from its intrinsic gate resistance at its physical temperature; non-thermal noise from its intrinsic drain resistance at a fixed effective temperature of 400K; and thermal noise from extrinsic gate, source, and drain resistances at its physical temperature. The model shows that the optimum source impedance varies significantly with temperature, so we are assuming that at each temperature we are using a somewhat different amplifier, one designed to present the optimum source impedance for that temperature. Thus the model is not expected to reproduce experimental results found in the literature for amplifier noise vs. physical temperature, since those were measured with a fixed amplifier.

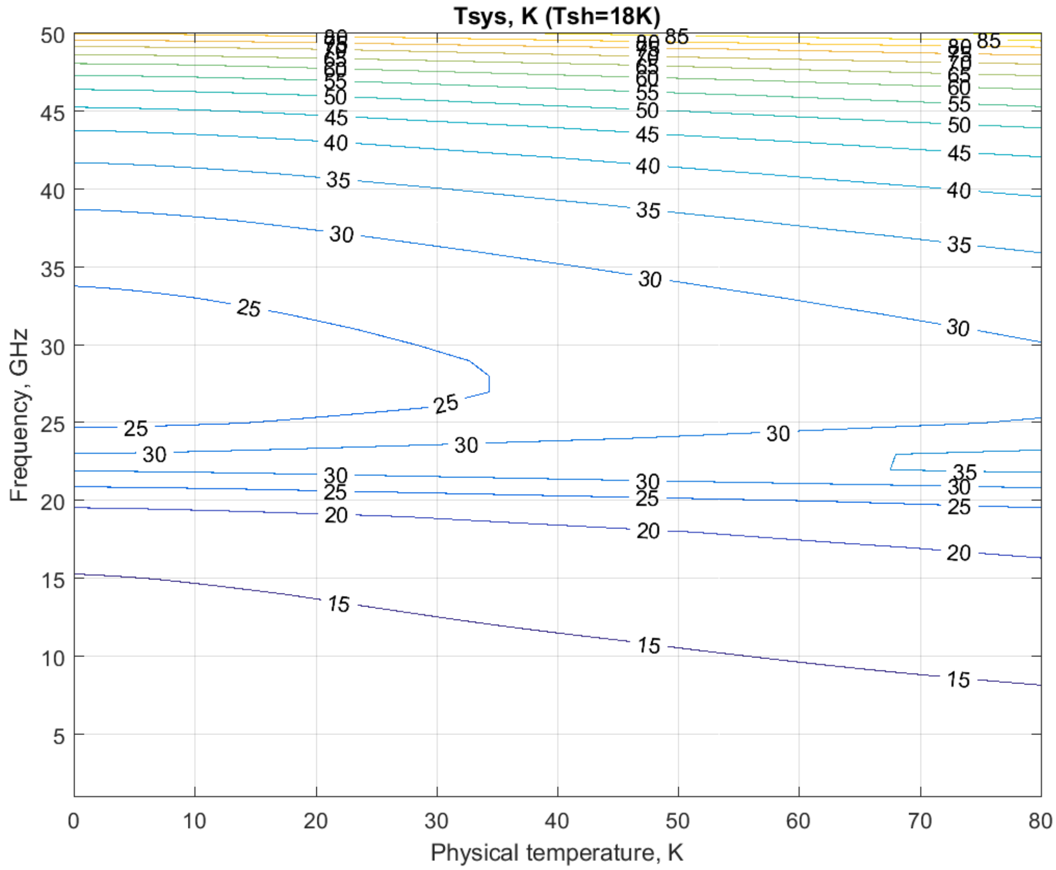

Figure 1. System noise temperature vs. physical temperature and frequency, calculated from a model based on the assumptions given in the text, including transistor self-heating of 18K.

Whether the transistor is subject to self-heating (assumption 10) may be controversial. The assumption is based on the thermal simulations in [15]. I discuss below how the results are affected if this assumption is relaxed. The model takes the effective physical temperature of the transistor to be

$$T_{tr} = T_{phys} + T_{sh} \exp(-T_{phys}/20\text{K})$$

where T_{sh} = 18K is the self-heating temperature.

The model's result for T_{sys} vs. physical temperature T_{phys} and frequency is plotted in Figure 1. At the high end of the frequency range, T_{sys} is dominated by absorption in the atmosphere. The peak near 22 GHz is due to H_2O absorption and the steady increase above 30 GHz is due to O_2 absorption, which peaks near 60 GHz.

The same data are re-plotted in Figure 2 to emphasize the variation with cooling temperature at a few selected frequencies. This shows that at low frequencies (below a few GHz) only a few percent improvement can be obtained by cooling below 80K, but that at higher frequencies cooling from 80K to the 10K-20K range reduces T_{sys} by up to 25%. There is very little further improvement for cooling below about 15K. This is because, at low frequencies, the noise from the feed loss and LNA (the only components affected by cooling) at 80K are already a small fraction of the fixed noise sources. At high frequencies, the LNA noise is much more significant, in spite of the fact that the fixed noise from the atmosphere is also higher.

Figure 3 shows the effect of eliminating self-heating of the LNA transistors (T_{sh} = 0, T_{tr} = T_{phys}). At frequencies below a few GHz, it remains true that there is little benefit to cooling below 80K, but above 30 GHz we can now obtain about 10% improvement in T_{sys} by cooling from 20K to 5K.

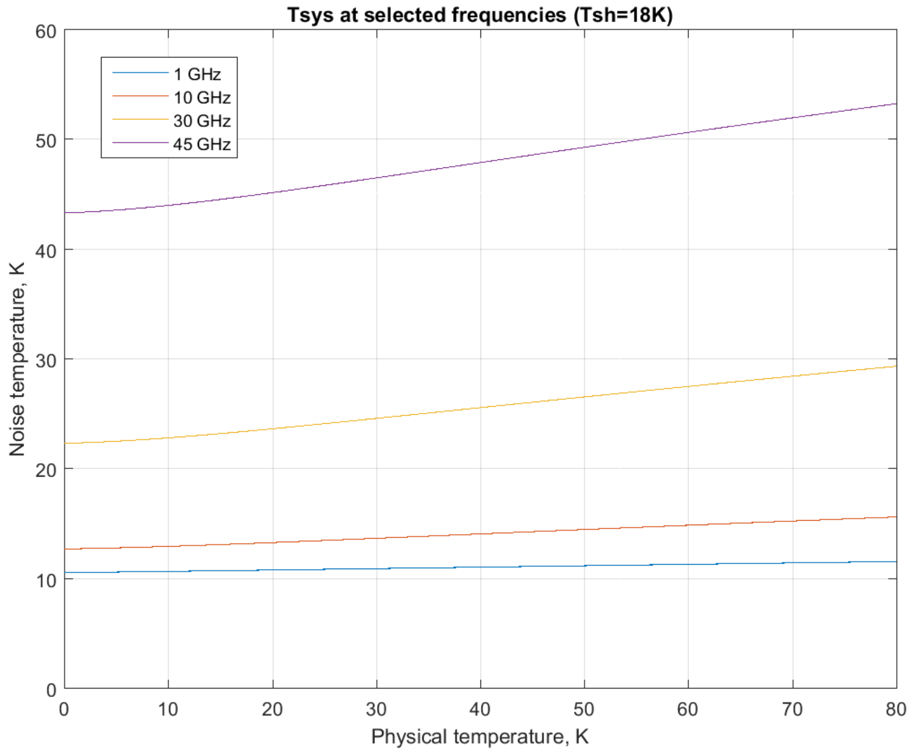

Figure 2 System noise temperature vs. physical temperature at selected frequencies (T_{sh} = 18K).

B. Cooling Capacity

In the steady state, a receiver has several sources of heat that must be extracted by the cryocooler in order to maintain the desired temperature: dissipation from internal electronics (primary the LNAs); solid conduction from outside through wires, transmission lines, and support structures; conduction through residual gas in the vacuum chamber; and radiation from the room-temperature walls of the vacuum chamber and from the window through which the feed illuminates the antenna. Solid conduction can be made negligible by good design, and gas conduction is negligible if a hard vacuum is maintained. Radiation from the walls can be made very small by multi-layer insulation, but MLI cannot be used over the feed window. Thus, the sources that are fundamental are electrical dissipation and radiation through the window.

A survey of commercial cryogenic LNAs from two companies [16][17] shows that devices covering most of the ngVLA range (1.2-50 GHz and 70-116 GHz) can provide >30 dB of gain while dissipating 10-30 mW. The current ngVLA reference design [18] requires 6 receivers, or 12 LNAs with two polarizations. Using 30 mW per amplifier as a worst-case estimate, we find a total dissipation of 360 mW. The lowest frequency receiver in that design covers 1.2 to 3.9 GHz, and from the previous section this need only be cooled to about 80K. The remaining amplifiers should be cooled to about 20K. So we have dissipation of 60 mW at 80K and 300 mW at 20K.

The window diameter needed for each feed is proportional to $(\theta f_{min})^{-1}$ where f_{min} is the lowest frequency of that receiver and θ is the antenna illumination angle as seen from the feed. Primarily for this reason, the ngVLA antenna is being designed so that θ is relatively large, ~100º. In that case, recent design work [19] implies that window diameters of

$$D = (48 \text{ cm-GHz}) / f_{min}$$

are sufficient. Using f_{min} = {1.2, 3.9, 12.6, 21, 30.5, 70} GHz for the various bands [18] then implies window diameters of D = {40, 10, 3.8, 2.3, 1.6, 0.7} cm. The black-body thermal

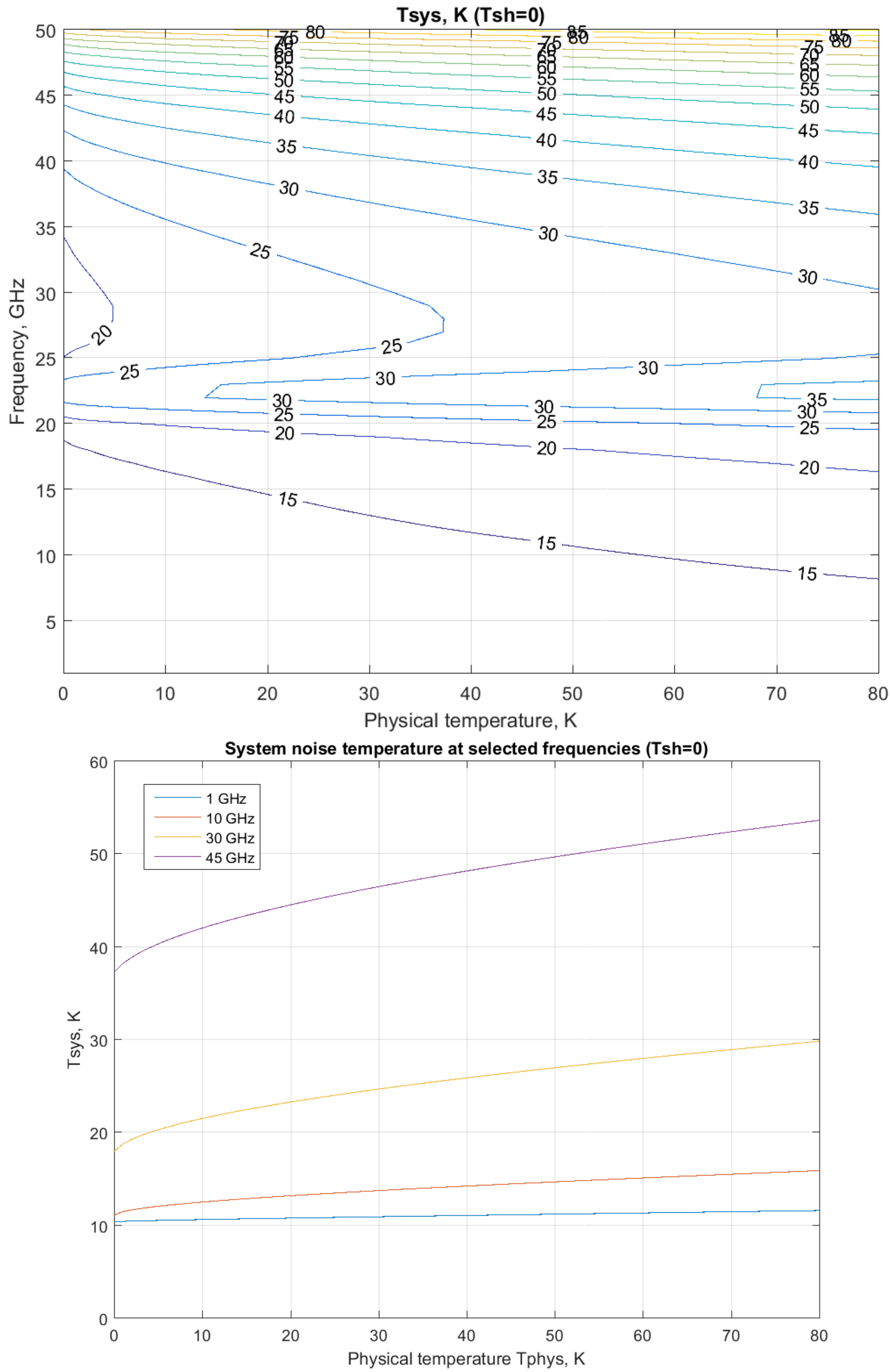

Figure 3. Same as Fig. 1 and Fig. 2 without self-heating of LNA transistors.

radiation impinging on such a window is given by

$$P_{IR} = \pi D^2 \sigma T^4/4$$

where $\sigma = 5.67\text{e}{-8}$ W m^{-2} K^{-4} is the Stefan-Boltzman constant and T is the temperature of the surrounding material. At $T = 300$K this is 57.7 W for the lowest frequency feed and a total of 4.4 W for all 5 of the other feeds. Not all of this power needs to be absorbed by the cryocooler. Some is reflected back out by the feed, which can be polished to high reflectivity, but due to its complex shape multiple reflections will occur and the net reflectivity is unlikely to exceed 0.5. In addition, it is possible to build filters that reflect (or absorb and re-radiate outward) most of the incoming infrared radiation while transmitting most of the desired microwave signals. The performance achievable with such filters is not yet well understood, but it is being actively investigated [19]. For now, we assume that 10% of the radiation impinging on the window must ultimately be absorbed by the cryocooler.

The total steady-state cryocooler loads from dissipation and window radiation are then

$$0.06 \text{ W} + 5.77 \text{ W} = 5.83 \text{ W at } 80\text{K}$$
$$0.30 \text{ W} + 0.44 \text{ W} = 0.73 \text{ W at } 20\text{K}.$$

This can be achieved by a single two-stage cryocooler. There will be some residual load from wall radiation and from solid conduction, and some margin must be provided, so <u>a cryocooler that achieves about 8W at 80K and 1W at 20K should be adequate with careful design</u>.

In addition to handling the steady-state load, the cryocooler must be able to reach the desired equilibrium temperatures starting from room temperature. This is a highly non-linear process, so it is not obvious that a cryocooler sized for the steady-state conditions will achieve it. One issue is conduction through the residual gas. Vacuum chambers for radio telescope receivers are rarely built to ultra-high-vacuum standards, so they can be slightly dirty, contain materials subject to outgassing, and may even have very small leaks. A hard vacuum is nevertheless achieved in the steady state by cryopumping; if any part of the interior is below about 77K, then the main residual gasses (H_2O, O_2, N_2) freeze out on that part. Prior to reaching that temperature, the vacuum achieved by a mechanical roughing pump may allow many watts of conduction through the gas. Another issue is the thermal mass of passive parts. The ngVLA feed for 1.2 GHz, with about 40 cm aperture diameter and similar length, is likely to have a mass of several kg. (A design for 1.2-4.2 GHz has an estimated mass of 2.5 kg [20].) The average heat capacity of aluminum from 300K to 70K is about 750 J/kg-K, so 172,500 J must be removed to cool 1 kg to 70K; with 8W of cooling this takes 6 hours per kg. Fortunately, the cooling capacity of cryocoolers is generally greater when the load is warmer, which helps overcome the residual gas conduction and reduce the cooling time, but the actual high-temperature capacities are rarely included in specifications. Therefore cooling time for each practical design must be obtained experimentally.

In order to ensure that cooldown will occur in a reasonable time in the face of residual gas conduction and thermal mass, radio astronomy receivers are sometimes built with cryocoolers much larger than needed for the steady-state load. To avoid this, there have recently been proposals to use crycoolers of variable capacity [21]. The idea is to install a large unit but throttle it down once the target temperatures are reached, thereby reducing its power consumption. The difficulty with this approach is that the throttled-down efficiency is much lower than that of a smaller unit, so the power saving is less than it could be. Another approach is to install an auxiliary cryocooler that can be completely turned off when the target temperatures are reached.

C. Cost

Cooling to lower temperature leads to higher capital and operating costs. If it did not, we

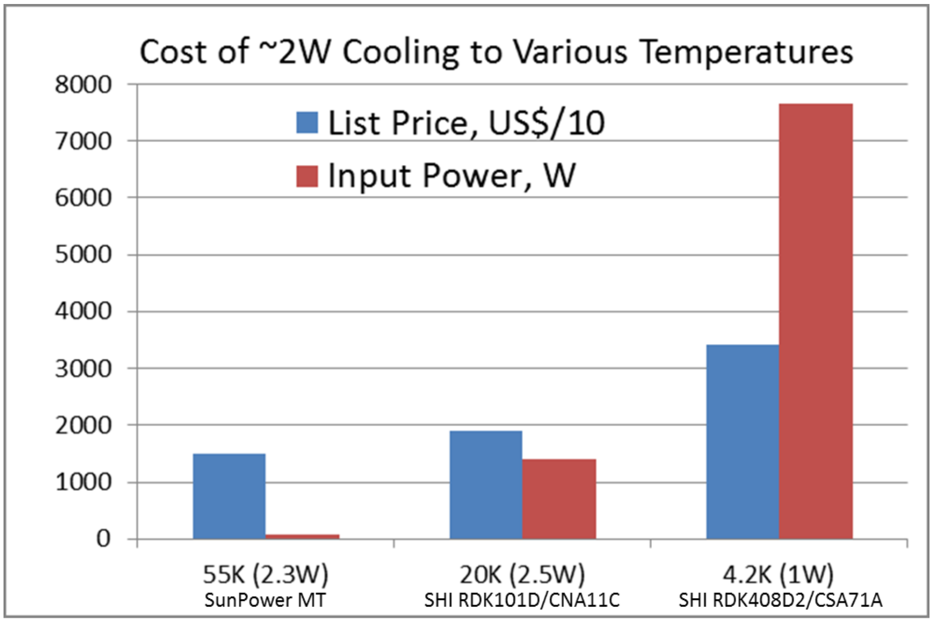

Figure 4. Cost and power consumption of several commercially-available cryocoolers of similar cooling capacity at different temperatures.

would prefer to cool receivers to the lowest possible temperatures. This is illustrated by Figure 4, which shows the list prices and power consumptions of three commercial cryocoolers that provide 1 W to 2.5 W of cooling at temperatures of 77K, 20K, and 4.2K, respectively.

Although the capital cost increases only modestly at lower temperatures, the power consumption increases by 16.5x from 55K to 20K and by an additional 5.9x from 20K to 4.2K. Since the 4.2K cooler has a capacity of only 1W, the latter ratio becomes 14.7 if scaled by capacity. On the other hand, the 55K cooler here is a one-stage device and the 20K and 4.2K coolers used here are two-stage devices and no credit has been given for the higher-temperature cooling of the first stage.

D. Discussion

Below 3 GHz, it appears that cooling to 80K (and perhaps higher) is sufficient; only a few percent reduction in system temperature is achieved at lower temperatures. A similar result is found in the 21-23 GHz range, where the emission from the 22 GHz H_2O line in the atmosphere dominates T_{sys}, even if the receiver is cooled only to 80K. At other frequencies in the 1-50 GHz range, substantial improvement in T_{sys} is obtained by cooling to 10K-20K. Whether cooling to even lower temperatures is cost-effective depends on whether self-heating in the LNA transistors is significant.

Modern radio astronomy receivers have bandwidths much larger than 10%, so behavior in the narrow range 21-23 GHz is not likely to drive the design. A receiver that includes that range likely includes other frequencies where tropospheric emission is lower and there is more benefit to receiver cooling.

Infrared radiation into the feed should be the dominant load on a cryocooler at sufficiently low frequencies, since that power is proportional to $(\theta f_{min})^{-2}$. Recall that we are assuming that the feed is cooled to the same temperature as the LNA. For designs where θ is small, the load is often so high that cooling of the feed is not practical. Whereas the ngVLA antenna is being designed with $\theta \approx 100^{\circ}$, cooling of the feed should be practical even at the lowest frequency (1.2 GHz).

Fortunately, receivers for lower frequencies where the radiation load is high can be operated at higher temperatures. Since ngVLA is unlikely to be able to cover the 1.2 GHz to 50 GHz range with a single receiver, cooling the 1.2 to ~4 GHz receiver(s) to ~80K and cooling the higher-frequency receivers to ~20K is a near-optimum choice. This can be accomplished with a single two-stage cryocooler if it is properly sized. Whether it is cost-effective to cool the high-frequency receivers to even lower temperatures requires more investigation. The main uncertainty is whether self-heating of LNA transistors is significant; there seems to be only one simulation of this [15], on which I relied for this study, and no experimental data. Published empirical results on amplifier noise vs. physical temperature do not provide the answer because they measure the amplifier at a fixed source impedance, not at the optimum source impedance for each temperature.

Another design approach, not modeled here, is to cool the LNA to a lower temperature than the feed. This includes the possibility of leaving the feed at room temperature. Doing this can substantially reduce the load on the cryocooler, but only if the transmission line from feed to LNA is of high thermal resistance or has a thermal break. High thermal resistance is accompanied by electrical loss, so that the transmission line then introduces additional noise. For bandwidths greater than about 1.6:1, thermal breaks with low microwave loss are not feasible. Therefore, with the exception of narrow-bandwidth receivers for which the feed is at room temperature, this seems to me to be a poor compromise. For wide bandwidths or for receivers where the feed is cooled at all, it is best to keep the feed and LNA at the same temperature and keep the transmission line between them short enough to have negligible loss.

Our results might be different if the atmosphere has much more or much less water vapor. The model here assumes a vertical column of 6 mm of condensed water, which is the median value for the VLA site. In contrast, the median value for the ALMA site is 1.1 mm.

Another issue with the model used here is that it assumes that LNAs achieve the minimum noise temperature of the cryo3 transistors at all frequencies. While this might be approached in practice for narrow-bandwidth LNAs, modern telescopes require wide-bandwidth receivers. It is not practical to achieve the minimum noise temperature over a 2:1 or larger bandwidth. It is more realistic to assume that the minimum noise temperature at the high end of the band is achieved over the whole band. This has the effect of increasing the LNA noise at most frequencies, which makes cooling to lower temperatures more beneficial. This seems unlikely to alter the conclusions of this study, but more detailed modeling is needed to be sure.

III. Cryocooler Technologies

A. General ideas and the ideal (Carnot) cycle

Closed-cycle refrigeration relies on changes to the thermodynamic state of a working fluid, which absorbs heat from the load and dissipates heat to the environment. This is shown conceptually in Figure 5. Work W must be done on the system in order to absorb heat Q_L from the low-temperature load and discharge heat Q_H to the high-temperature environment. From the first law of thermodynamics (conservation of energy), $Q_H = Q_L + W$. We define the power efficiency (also known as coefficient of performance, CoP) as

$$\eta = Q_L/W = (dQ_L/dt) / (dW/dt) = (dQ_L/dt) / P, \tag{1}$$

where $P = dW/dt$ is the input power. Efficiency is maximized when the state changes of the working fluid are *reversible* (entropy is constant except during heat transfer). In that case, the same machine can also be run as a heat engine, where the heat flow is in the opposite direction and work is extracted. The state of the working fluid can be described by the pressure-volume (P-V) diagram in Fig. 5, where for refrigeration the cycle proceeds counter-clockwise (4-3-2-1-4...). The compression (3 to 2) and expansion (1 to 4) phases must be performed adiabatically and the heat transfer phases (4 to 3 and 2 to 1) must be isothermal for the process to be

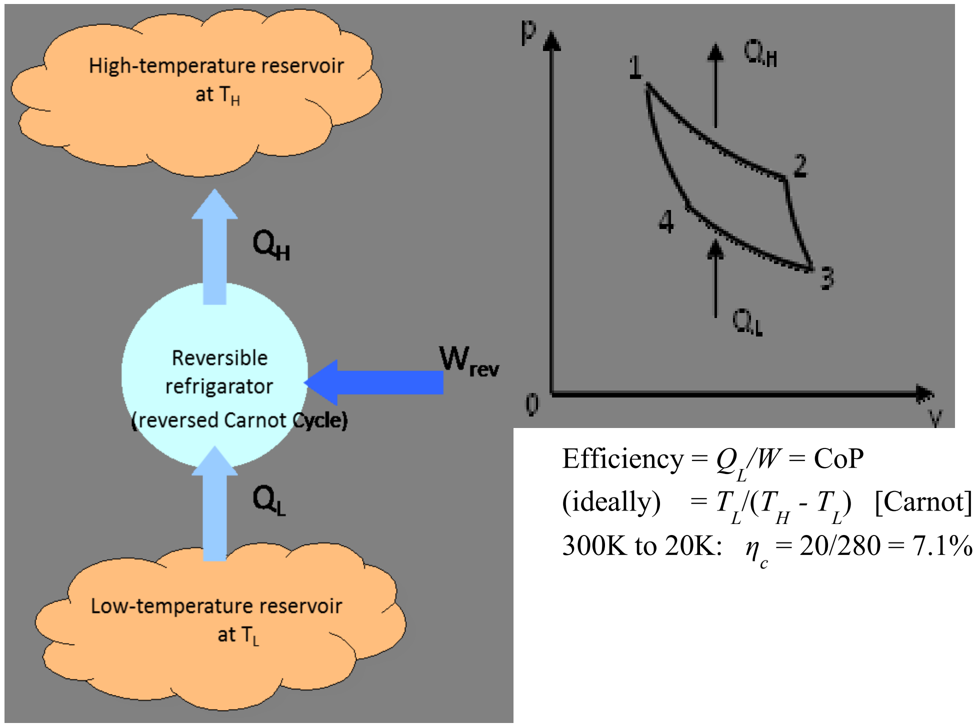

Figure 5. The Carnot refrigeration cycle.

reversible, in which case the whole process is called the *Carnot cycle* [22]. The efficiency is then called the Carnot efficiency and is given (for refrigeration) by

$$\eta_C = T_L/(T_H - T_L). \tag{2}$$

For practical refrigerators, we always have $\eta \leq \eta_C$ and often $\eta \ll \eta_C$.

B. Sterling and Gifford-McMahon Cycles

Many practical refrigeration cycles have been developed, but only the Sterling and Gifford-McMahon cycles are likely to be of value to radio telescope receivers, so the discussion here is limited to those[1]. In all cases of interest, the working fluid is helium 4 (He^4).

A Sterling-cycle refrigerator [23] is shown conceptually in Figure 6. The working fluid remains within one contiguous space whose volume is varied by a pair of pistons. There are four separate phases of operation. We begin with both pistons moved to the left so that none of the fluid is in the expansion space. Then the warm piston is moved part way to the right, compressing the fluid. Next, both pistons are moved together to the right, forcing the fluid completely out of the compression space, through the regenerator (described in a moment), and into the expansion space but keeping its volume unchanged. Then the cold piston is moved right, expanding the fluid. Finally, both pistons are moved fully left, forcing all of the fluid back

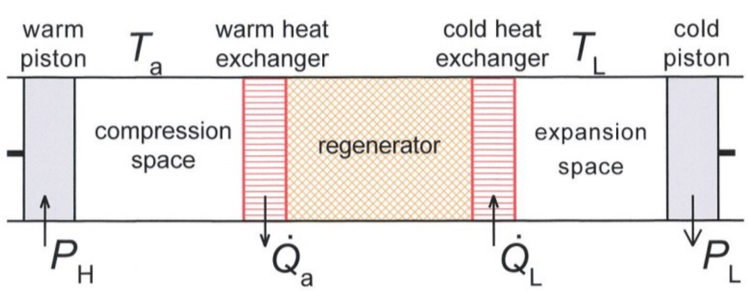

Figure 6. Sterling cycle refrigerator, conceptual diagram.

[1] As mentioned in Section I, the Joule-Thompson cycle has been used in the past.

through the regenerator and into the compression space, completing the cycle. During compression, heat is transferred to the hot reservoir; during expansion, heat is absorbed from the cold reservoir. Work at rate P_H must be done on the warm piston to cause compression, but some of that work is recovered at rate P_L from the cold piston during expansion, so the net power input is $P = P_H - P_L$.

The regenerator is a porous solid chosen to have high heat capacity, so as to act as a thermal buffer. It absorbs heat from the warm fluid as it passes from the compression space to the expansion space, and it returns heat to the cold fluid as it passes in the other direction.

A schematic of one practical implementation of a Sterling refrigerator is shown in Figure 7. Instead of having separate pistons for compression and expansion, a single pair of

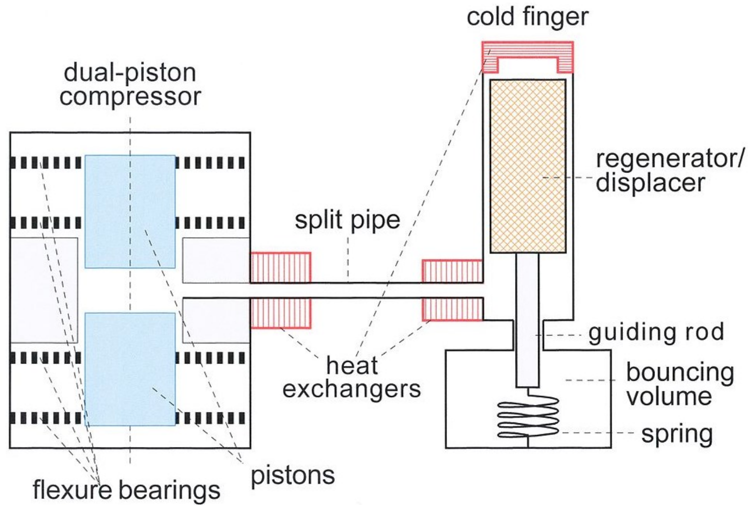

Figure 7. Schematic of one implementation of a Sterling refrigerator.

synchronized pistons does both. A separate "displacer" piston moves the fluid between the expansion space (top, cold) and the compression space (bottom, hot) without changing its volume. The displacer doubles as the regenerator. To implement the four phases described earlier, the two pistons should be driven about 90º out of phase. One way to do that, shown in Fig. 7, is to attach the displacer to a spring and let it be driven by the gas flow, where the mass and spring constant are chosen to achieve the desired phase relationship. Another approach is to drive the displacer with a separate motor; this allows more precise control of the phase and allows variable cycle frequency. The displacer motor should consume little power because the displacer neither compresses nor expands the fluid.

This arrangement allows the compressor and refrigerator to be slightly separated, which is often convenient, but the pipe between them cannot be very long because the working fluid within its volume does not contribute to the heat transport but must still be compressed and expanded; and friction within the pipe along with the inertia of its fluid make the pressure cycle amplitude smaller in the refrigerator than in the compressor. These things reduce efficiency.

Now consider the schematic of a Gifford-McMahon refrigerator [24] shown in Figure 8. It has a displacer/regenerator, cold expansion space, and hot compression space that are the same as in the Sterling refrigerator of Fig. 7. However, instead of a single contiguous volume for the working fluid, it is broken into separate spaces by a pair of valves. This allows a compressor to operate independently of the refrigeration unit and to be separated from it by two pipes of almost any length. The compressor must simply maintain a pressure difference between the two pipes. During the compression phase of the cycle, the high-pressure side valve is opened, admitting additional fluid and raising the pressure in the working space. During the expansion phase, the low-pressure-side valve is opened. During the other two phases, both valves are closed and the

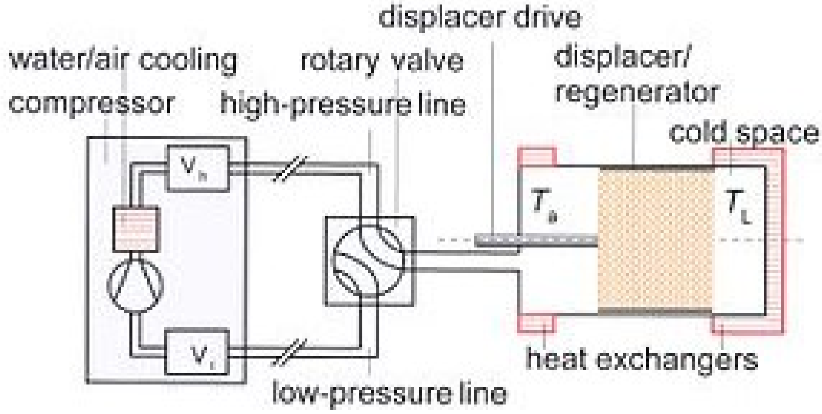

Figure 8. Schematic of a Gifford-McMahon refrigerator.

displacer/regenerator is moved. The refrigeration cycle within the working space is thus the same as for a Sterling refrigerator; only the compressor is different. The compressor's operation need not be synchronized with the refrigeration cycle. It can use a continuously-running pump (the usual design for modern units) or it can use a set of pistons and additional valves running at a rate independent of that of the refrigeration cycle.

One consequence of the GM arrangement is inherently lower efficiency compared with Sterling[2]. That is because in the Sterling cycle work is done on the compressor piston during the expansion phase, and some of that work can be recovered (e.g., stored in a spring and used during the next compression phase). Only the net work over a cycle (difference between work input during compression and work output during expansion) counts toward the input power P. In the GM refrigerator, there is no mechanism for work recovery during expansion. On the other hand, GM has the practical advantage of allowing most of the heat to be dissipated far from the cooling unit. Recall that we must have $\eta \leq \eta_C = T_L/(T_H - T_L)$. To reach cryogenic temperatures (<80K) from room temperature (300K) requires low efficiency ($\eta < 0.36$ and in practice far less), so much more heat must be dissipated than is removed from the load. In a Sterling refrigerator, all of it must be dissipated near the refrigerator, but in a GM refrigerator most of it can be dissipated at the (possibly well separated) compressor.

C. Pulse tubes

A relatively new approach to cryocooler design involves the use of pulse tubes. A pulse tube is simply a pipe filled with a gas (the working fluid) in which a cyclic pressure wave is generated in such a way that the gas is predominantly undergoing expansion (cooling) near one end of the pipe and compression (heating) near the other end. The result is that heat can be transported from the expansion end to the compression end. Early pulse tubes produced very little refrigeration, but sophisticated designs were eventually developed making them sufficiently efficient to make practical cryocoolers [25].

Modern pulse tube refrigerators are actually very similar to Sterling or GM refrigerators, as shown in Figure 9. To facilitate the comparison, the displacers and regenerators are separated for Sterling and GM [Figs. 9(a) and 9(c)]; the pulse tube [Fig. 9(b)] simply omits the displacer. Instead it relies on the inertia of the working gas to move itself from the cold end during expansion to the hot end during compression, in response to a pressure oscillation produced by the compressor. To achieve efficient refrigeration, it is necessary for the mass flow at any point

[2] I will show in Section IV that in practice $\eta/\eta_C \sim 5\%$ for GM refrigerators and 10% to 20% for Sterling refrigerators.

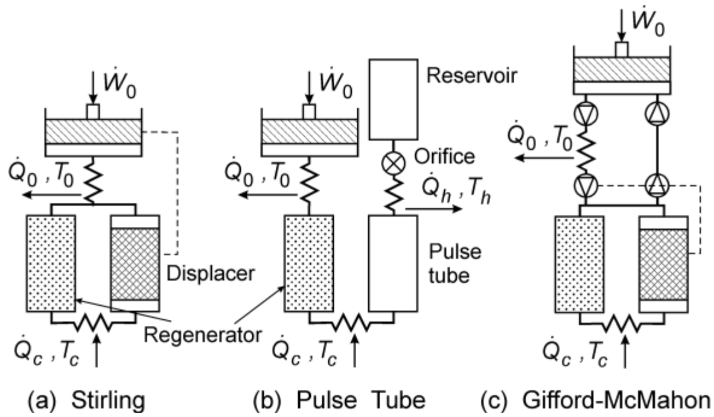

Figure 9. Schematic of a pulse tube refrigerator (b), compared with those of Sterling (a) and GM refrigerators (c) having separate regenerators and displacers. From [26].

in the pulse tube to be roughly in phase quadrature with the pressure oscillation at the same point. This is achieved by careful selection of the sizes of the orifice and the reservoir. An alternative design replaces the orifice with a thin tube of carefully selected length. The main advantage of the pulse tube is that it avoids having any moving parts inside the cold part of the refrigerator. A challenge in the design of Sterling and GM refrigerators is to maintain a good seal between the displacer/regenerator and the wall of the chamber both in equilibrium and during cool down from room temperature. Use of clearance seals requires maintaining precise tolerances and using materials with carefully controlled temperature coefficient of expansion. Use of rubbing seals creates a mechanism for wear and a need for periodic replacement. These issues are avoided with the pulse tube refrigerator, which should therefore be more reliable and essentially maintenance-free. In addition, the motion of the displacer causes some vibration, so pulse tubes are strongly favored in some vibration-sensitive applications; this is not large enough to be significant in most radio telescope receivers.

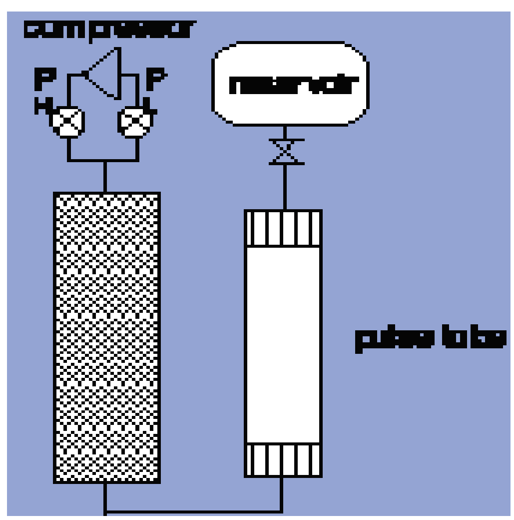

Figure 10. Schematic of a GM-style pulse tube refrigerator.

The pulse tube refrigerator in Fig. 9(b) is similar to the Sterling refrigerator because both are driven by a compression/expansion piston near the refrigerator. Pulse tube refrigerators can also be constructed in the style of a GM refrigerator, using valves to generate the required pressure oscillation, as shown in Figure 10. GM-style pulse tubes are inherently less efficient than Sterling-style for the same reason that GM refrigerators are inherently less efficient than Sterling: the lack of a mechanism for work recovery during expansion.

D. Multi-stage refrigerators

The discussion so far has considered only single-stage refrigerators, with one cold load temperature and one hot reservoir temperature. For several reasons involving the properties of materials, it is difficult to reach very low temperatures in this way. When $T_L \ll T_H$, conduction through the walls of the refrigerator and through the regenerator/displacer causes increasing loss of efficiency; the regenerator material cannot maintain high heat capacity over the whole temperature range; and it is difficult for seals to work well over a large temperature range due to thermal expansion. These and other effects make it advantageous to construct the refrigerator as a cascade of two stages, with $T_L < T_M < T_H$ where the first stage operates between T_H and T_M and the second stage operates between T_M and T_L. The same concept can be extended to three or more stages. In many designs the stages are coupled, for example using a common compressor and allowing movement of working fluid between stages, but in principle they can be independent and can use different refrigeration cycles.

Having more than one stage can have practical advantages, since it allows some of the heat load to be absorbed at higher temperatures and thus higher efficiency, reducing the cooling requirement at the lowest temperature. This is a natural arrangement for radio telescope receivers, since only the components that benefit the most from cooling (LNAs) need to be at the lowest temperature.

It is easy to compare the efficiencies of two single-stage cryocoolers, even if they cool to slightly different temperatures, by considering $E = \eta/\eta_C$, the ratio of actual efficiency to ideal (Carnot) efficiency. This is more difficult for multi-stage cryocoolers, since it is often not obvious how to allocate the input power among the stages. To provide a single overall "normalized" efficiency measure for multi-stage cryocoolers, the following definition [26] has been adopted by the cryocooler community:

$$E = \frac{1}{P}\sum_{i=1}^{n} \frac{dQ_i/dt}{\eta_{Ci}} \tag{3}$$

where there are n stages and dQ_i/dt is the heat flow absorbed (load) in stage i at temperature T_i and $\eta_{Ci} = (T_H - T_i)/T_H$ is the Carnot efficiency for that stage. This gives "credit" for the cooling provided at each temperature in proportion to the inherent difficulty of achieving it. For a two-stage cooler, this reduces to

$$E = \frac{1}{P}\left(\frac{dQ_1}{dt}\frac{T_H - T_1}{T_1} + \frac{dQ_2}{dt}\frac{T_H - T_2}{T_2}\right). \tag{4}$$

E. Comparison

Table 1 gives a comparison of the types of cryocoolers considered here. The efficiency difference has already been discussed and will be seen for specific units in Section IV. The

Table 1: Sterling vs. Gifford-McMahon

Sterling and Sterling-style pulse tube	Gifford-McMahon and GM-style pulse tube
Efficiency relative to Carnot ~10%	Efficiency relative to Carnot ~3%
Compressor adjacent to cold head	Compressor can be remote
Little to no maintenance for >10 years	Replacement of seals every 12-24 months
Vary power input / capacity by amplitude or frequency	Vary power input / capacity by frequency
PT version may have slight orientation dependence	PT version may have large orientation dependence

compressor location and maintenance differences have also been discussed. I consider the other points now.

For any of these refrigerators, the cooling capacity at a given temperature varies with the cycle rate. As the frequency increases from low speed, the rate of gas flow between the expansion and compression chambers increases, producing larger heat flow. However, as the speed is further increased the inertia of the gas and friction in the regenerator result in incomplete gas transfer. For GM coolers, there may not be enough time during the compression phase for enough gas to pass through the input valve to achieve the full inlet pressure, and similarly during the expansion phase. For either type at high frequencies, the gas may not spend enough time in the regenerator to release/absorb enough heat. For all these reasons, further increases in speed result in smaller increases in heat flow, and eventually in a reduction in heat flow. Meanwhile, the power input needed to drive the refrigerator increases monotonically with speed. The net result is that there is a speed at which efficiency is maximized. For reasons that are not quite clear to me, this speed tends to be much higher for Sterling and Sterling-style PT refrigerators (10s of Hz) than for GM and GM-style PT refrigerators (1-2 Hz).

Sometimes the capacity at optimum speed is more than required, in which case the speed can be reduced to lower the input power. Efficiency will be lower, but that may not be important. In this situation, the Sterling refrigerator has another method of reducing capacity to save power. If the compressor/expander piston is driven by a linear motor, the input power can be reduced without changing the speed. The piston stroke and the pressure amplitude are then reduced, reducing the cooling capacity but maintaining nearly the same efficiency. This method is not available for a GM cooler (line 4 in Table 1).

For pulse tube coolers, the lack of a displacer to force the gas to move between the expansion space and the compression space can lead to a performance dependence on the direction of gravity. The motion is nominally in response to the pressure oscillation, but whereas the cold gas is more dense than the warm gas, it is also affected by buoyancy forces. Whether the effect is significant depends on the pulse tube's design, including its length to diameter ratio, but also on the cycle rate. At high speed, the pressure-driven motion is much faster than any buoyancy-driven motion and dominates. At low speed, this may not be the case.

For this reason, GM-style pulse tube refrigerators (which operate at low speed) tend to have more orientation sensitivity than Sterling-style pulse tube refrigerators (line 5 in Table 1).

E. Cryocoolers in existing radio telescopes

Today, almost all cryocoolers in centimeter-wavelength telescopes, both single-dish and synthesis arrays, are two-stage GM types, reaching 15-25K depending on the heat load. Many use commercial models whose designs are 30-40 years old. The GM cryocooler was invented in 1960 [31], and two-stage versions were commercialized around 1970 when CTI introduced the model 1020; that design is still in production. Two systems using such refrigerators are shown in Figure 11.

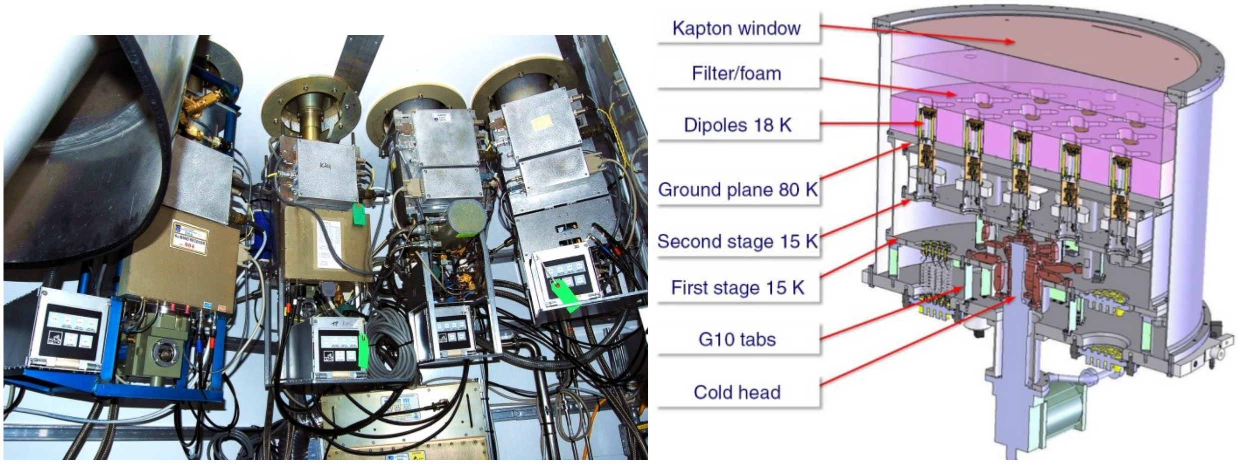

Figure 11. Current receiver designs using commercial 2-stage GM cryrcoolers. Left: JVLA, with 8 receivers (4 shown), each in a separate vacuum chamber with its own cryocooler. Right: 19-element phased array receiver for GBT [32].

At millimeter and sub-millimeter wavelengths, superconducting receivers are often used (SIS mixers and some bolometers), requiring cooling to ~4K. Until about 15 years ago, these relied on open-cycle cooling with liquid He or adding a Joule-Thompson third stage to a 2-stage GM. Now 2-stage GM refrigerators are able to reach temperatures <4K by employing exotic regenerator materials (ErNi, HoCu). The ALMA mm and submm array chose to use a specially-designed 3-stage GM cryocooler for its complex, multi-band receiver (Figure 12).

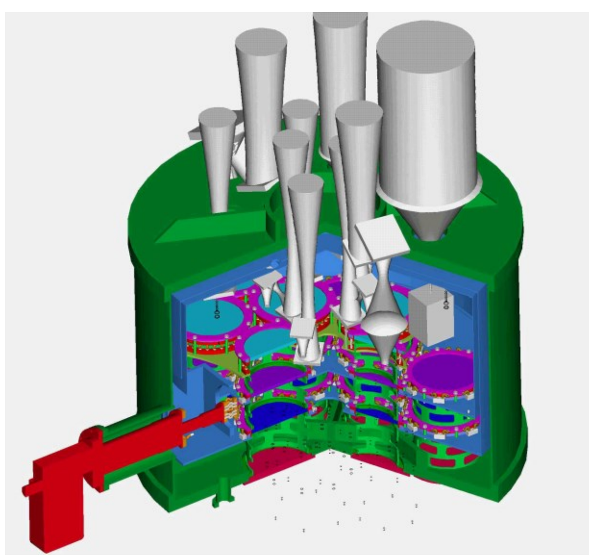

Figure 12. The ALMA front end assembly. Up to 10 receiveers are accomodated in a single vacuum chamber with a single 3-stage GM cryocooler (red) achieving cooling of 33W at 68K, 8W at 13.7K, and 1W at 4.2K.

IV. PERFORMANCE OF COMMERCIAL AND PUBLISHED DESIGNS

A. Markets and manufacturers

Some perspective on the place of radio astronomy in the cryocooler marketplace is given by Table 2, which shows the main applications for which commercial cryocoolers are produced. It should be clear that no part of the market is driven by radio astronomy. Even an exceptionally large telescope like ngVLA or SKA creates a negligible demand compared with other applications. For this reason, we cannot expect commercial manufacturers to produce cryocoolers tailored to our requirements, and it would require some luck to find a model already in production that is a good match to a particular receiver. Especially with two-stage cryocoolers, use of an off-the-shelf model usually results in one or both stages being larger than necessary. On the other hand, we can benefit from the cost savings of volume production.

Development of 4K cryocoolers was driven by the MRI market and is now achieved with two-stage GM machines. Prior to the 1990s, GM coolers could not produce significant cooling below about 15K because the heat capacity of materials being used in regenerators (lead and stainless steel) fell to near zero at lower temperatures. This was changed by the introduction of exotic rare-earth alloys (e.g., ErNi and HoCu [10]). However, as shown in Section I, the non-superconducting receivers with transistor-based LNAs now planned for ngVLA do not require cooling to such low temperatures.

Table 2: Cryocooler Markets

Market	*Typical Requirements*	*Volume*
Magnetic resonance imaging (MRI): superconducting magnets	1 W at 4K	>40,000 4K GM cryocoolers since 1995.
Infrared sensor cooling: primarily military applications night vision, weapon guidance	0.3W–1.75W at 65K–120K	~180,000 cryocoolers since 1970s.
High vacuum (cryopumping): mainly semiconductor processing	~3W at 15K	~20,000 units per year at peak of semiconductor business
Space: IR and X-ray detectors other specialized applications	Mission-specific. Low power, low mass.	small quantities, very expensive per unit

To date, radio telescope cryocoolers have mostly used cryocoolers intended for the high-vacuum (cryopumping) market. These are also two-stage GM machines.

Commercial Sterling cycle refrigerators are almost always single-stage designs[3] intended to cool loads at 70-80K. Some can reach no-load temperatures near 40K. Many have been built to cool infrared sensors in military weapons, where only a very short working lifetime is needed. In recent years, long-life units have been produced for other applications, such as cooling high-temperature superconductor filters for cell phone towers.

Single-stage and two-stage pulse tube refrigerators are also available, but they are almost all GM-style so they have no efficiency advantage over similar-size GM machines. They tend to be more expensive. Therefore they have a limited market, primarily in vibration-sensitive applications.

In recent years, there has been little innovation in cryocoolers for these high-volume markets. Unlike a large radio telescope array, none of these applications is especially sensitive to power consumption nor does it place units in remote locations where long-term operation

[3] I found one company, Sterling Cryogenics, that makes two models of two-stage Sterling machines, but they are very large, achieving 75W and 300W at 20K.

without maintenance is important. However, there has been considerable innovation in university-based research and in the aerospace industry. In these contexts, many designs for Sterling and Sterling-style pulse tube refrigerators have been published, including two-stage versions. These achieve much higher efficiency than COTS units of similar size.

There are not many manufacturers of cryocoolers. Most of them are listed in Table 3. CTI has made GM cryocoolers of the same designs and model numbers since the 1970s, and these are still available. They have been used in most cryocooled receivers built at the NRAO. The company was eventually purchased by Helix Technologies which was then purchased by Brooks Automation, so the products are sometimes found under those names. SHI Cryogenics now has the largest catalog of cryocooler products, including GM and PT machines of various sizes. This company (along with academic laboratories) pioneered the use of rare-earth regenerators and has most of the market for 4K cryocoolers. The exact regenerator materials are considered proprietary, but they are mostly disclosed in published papers.

The other manufacturers in Table 3 serve smaller markets.

Table 3: Cryocooler Manufacturers

Company	*Products*
CTI => Helix Technologies => Brooks Automation	1-stage and 2-stage GM
SHI Cryogenics (Sumitomo)	1-stage and 2-state GM and GM-type PT
Cryomech	1-stage GM; 1- and 2-stage GM-type PT
Oxford Cryosystems	1-stage and 2-stage GM
Qdrive	large single-stage Sterling
Sterling Cryogenics	large 1-stage and 2-stage Sterling
Sunpower	small single-stage Sterling
Ball Aerospace	military, space
Raytheon	military, space
Northrop Grumman	military, space
Honeywell Aerospace	military, space

B. Survey of available designs

For this study, data on a wide range of existing cryocoolers have been collected from the data sheets of manufacturers and from published papers. Results for the subset of these most likely to be useful in radio telescopes, namely two-stage devices that cool to temperatures near 70K and 20K, are plotted in Figure 13. For each cryocooler, the normalized efficiency relative to Carnot is plotted against the second stage load. The loads used correspond to temperatures as close to 70K and 20K as the available data allow. The power consumption is as reported, but in some cases the manufacturer provides a choice of several compressors that can be used with the same refrigerator; we then use the total power of the smallest applicable air-cooled compressor. All pulse tube refrigerators are Sterling-style; GM-style pulse tubes were not included. The raw data on which this plot are based, with additional details, are available [27]. This can be compared with an older published survey of single-stage cryocoolers, reproduced in Figure 14.

Both Fig. 13 and Fig. 14 show that larger machines have higher efficiency. This is because losses grow less rapidly than capacity upon scaling up. They also show that Sterling and Sterling-type PT are much more efficient than GM and GM-type PT at similar size.

The largest normalized two-stage efficiency found in this survey is 19.7%. This is the simulated performance of a published Sterling design [28] for a large machine (100W at 77K and 50W at 30K). Two others, near 13%, are from data sheets of Sterling Cryogenics; they are also

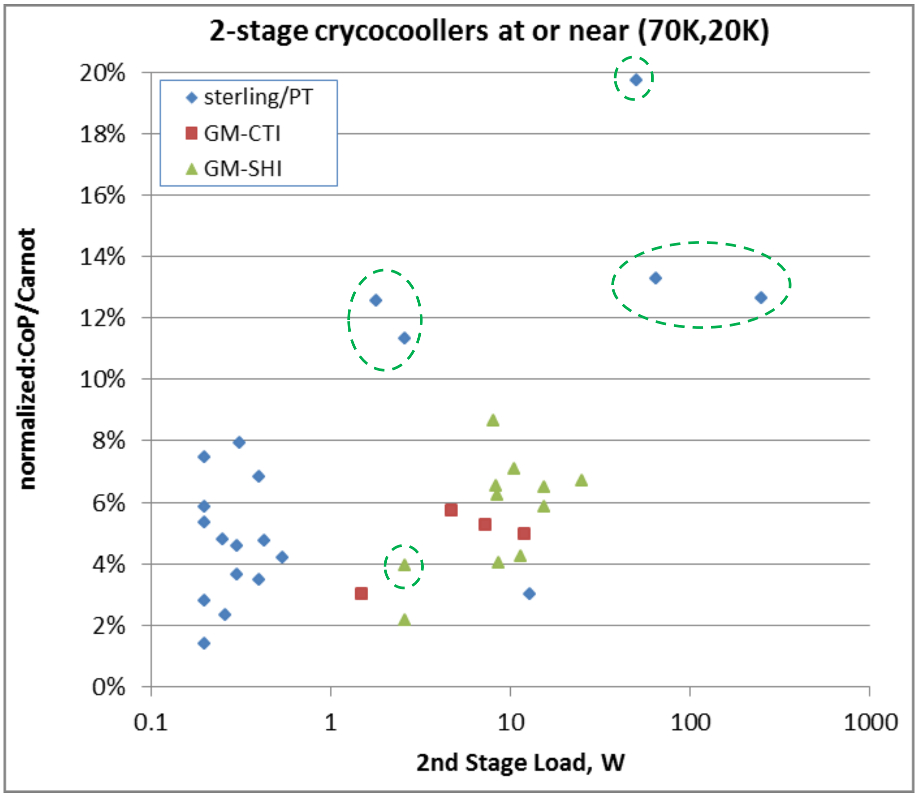

Figure 13. Survey of published performance of two-stage cryocoolers for temperatures near 70K and 20K, as a function of 2nd stage load. The ordinate is the normalized efficiency relative to Carnot, from equation (4). Certain points enclosed in dashed lines are discussed in the test.

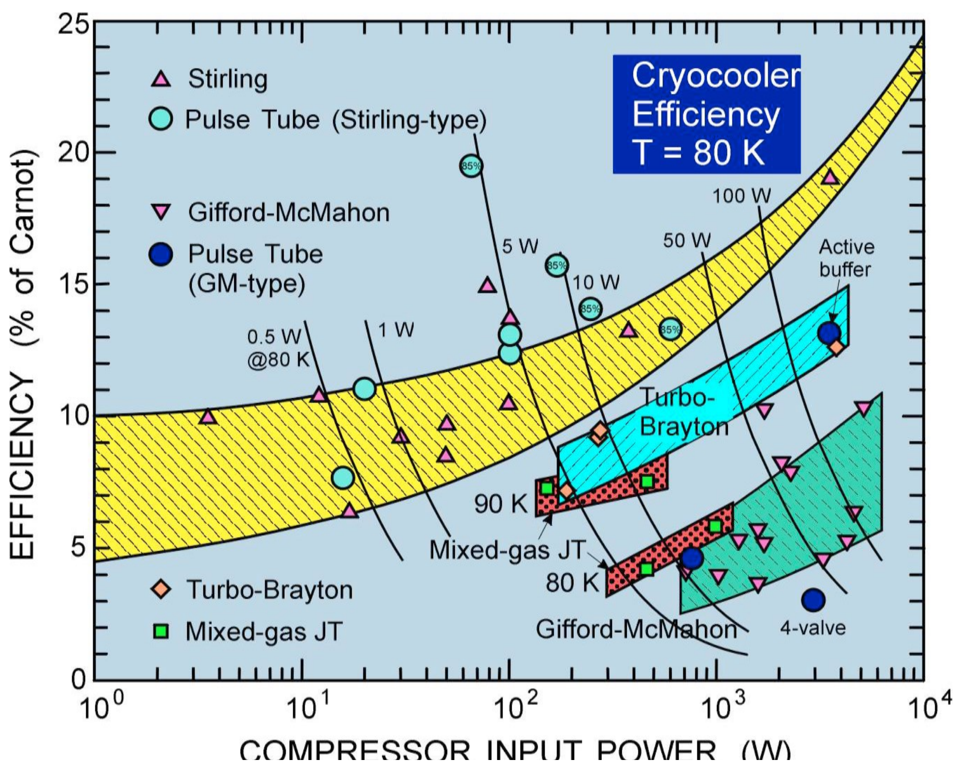

Figure 14. Efficiency of various single-stage cryocoolers, from [25].

very large. Among smaller machines, the best efficiency found is for the Raytheon LT-RSP2 [29], originally developed for space applications. The two points near 2W and 12% are both for this machine under different operating conditions. It is a hybrid cooler, with a Sterling refrigerator for the first stage and a Sterling-type pulse tube for the second stage. It is further described below. Its cooling capacity is nearly the same as the SHI GM machine plotted at 2.6W and 3.9% in Fig. 13, but the power consumptions of the two machines are 550W and 1600W, respectively.

The Raytheon LT-RSP2 is an example of the innovations being implemented in the aerospace industry. It has been built in several versions, optimized for different second-stage temperatures. A photograph of a version for 20K is shown in Figure 15. Simulated performance of this version, calculated at my request for this study, is plotted in Figure 16 [30].

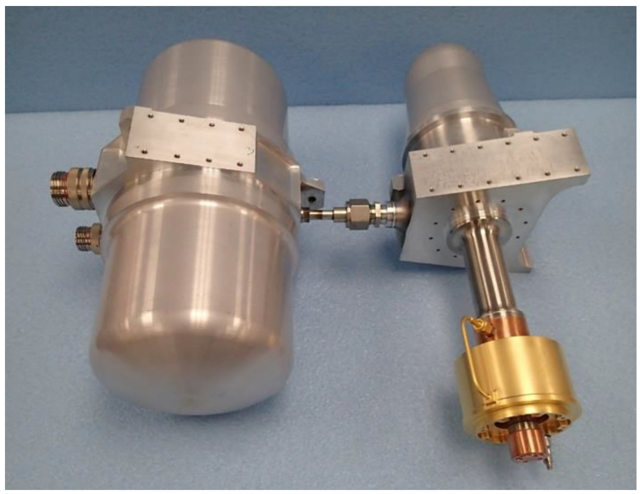

Figure 15. Photo of one version of the Raytheon LT-RSP2 hybrid cryocooler with Sterling first stage and Sterling-type pulse tube second stage. The compressor (left) is common to both stages.

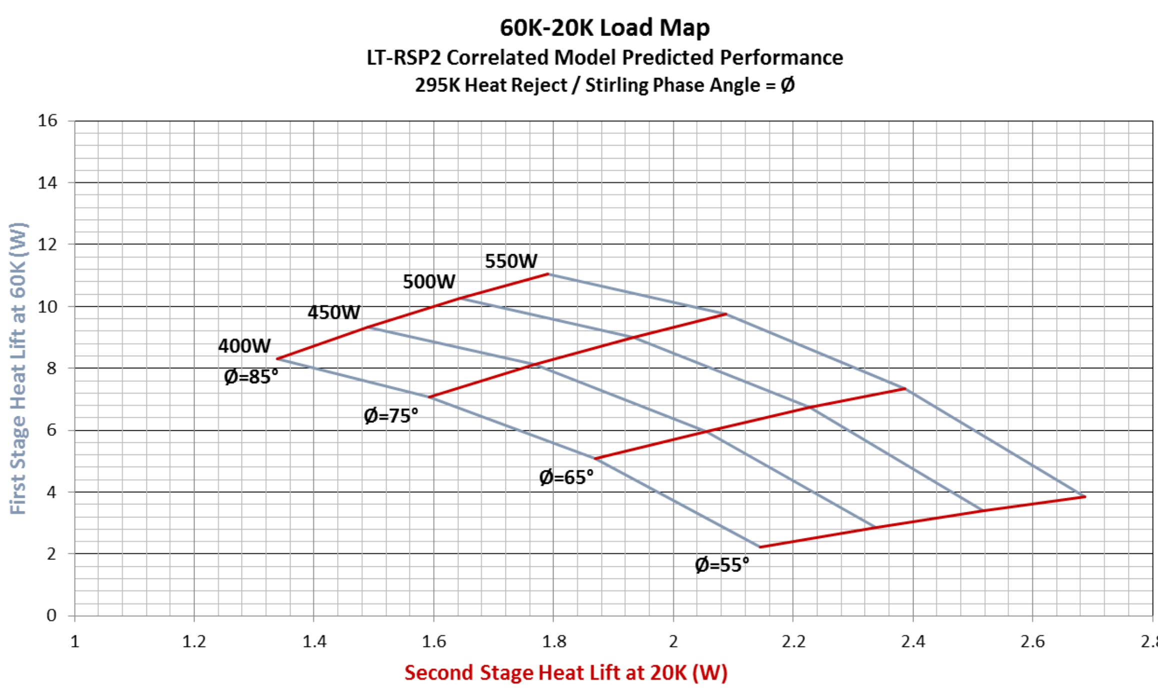

Figure 16. Cooling power on each stage of the LT-RSP2 cryocooler at fixed temperatures vs. compressor input power and phase angle of the first stage displacer [30].

A useful feature is that the phase difference between displacer motion of the first (Sterling) stage and the compressor is adjustable (since each is driven by a separate motor). This allows cooling capacity to be shifted between the two stages; as the Sterling stage is made less efficient, more of the gas is driven to the second (pulse tube) stage, increasing its cooling power at the same total input power. Input power can also be reduced while maintaining good efficiency, resulting in reduced cooling on both stages.

In principle this approach to load shifting between stages could be applied to any cooler where the displacement phase can be controlled separately for each stage. When both stages have mechanical displacers (e.g., where each is either Sterling or GM), this increases complexity quite a bit. To my knowledge, for all existing designs with two displacers both displacers are connected to a common drive shaft and driven together by the same motor. In many GM or GM-type PT designs, all displacers and both valves are driven by the same motor with the timing controlled by a cam.

V. Conclusions and Recommendations

The main conclusion of this study is that cryocooling for centimeter wavelength telescopes can use 4 to 40 times less power than has been achieved in recent years. This can be done with a combination of careful thermal design of the receiver to minimize the cooling load and use of more efficient refrigerators. For example, the current JVLA cryocooolers consume about 18 kW per antenna (partly because cooling is needed in 8 separate receiver packages), but it looks like an ngVLA antenna need only consume about 0.5 kW while covering a wider frequency range (with all receivers in one package and cooled by a single two-stage refrigerator).

These specific recommendations apply to the ngVLA project:

- Do not try to achieve very low temperatures, especially for receivers below 3 GHz. 80-100K for feeds and LNAs is sufficient below 3 GHz, and 20K is enough for receivers at higher frequencies.
- Careful thermal design of the receiver package is essential, especially with respect to radiation loading. This should allow a 6-receiver package covering the entire ngVLA range to produce an estimated steady-state load of about 1 W at 20K and 8W at 80K. However, more capacity may be needed to achieve acceptable cooldown time from room temperature.
- Consider use of Sterling and Sterling-type pulse tube refrigerators, even though such machines are not available off-the-shelf in the desired sizes. Some custom development may be needed, but it is worthwhile in a project as large as ngVLA.
  - Saving 10 kW per antenna for 200 antennas reduces electricity cost by 1.75 M$ per year at 0.10 $/kWh, so an investment of 3 M$ will be recovered in the first 21 months of operation.
  - Compressors and displacers with clearance seals and flexure bearings have no known wear mechanism, reducing maintenance cost.
- Allocate funds for non-recurring engineering, and begin development early.
  - Develop hardware matched to our specific needs;
  - Do not be limited by what can be purchased off-the-shelf.
  - Consider spending more on construction and NRE to save on operations
- Engage with the aerospace industry. That is where all the innovative work in cryocoolers has occurred in recent years. Designs originally intended for space applications can be adapted for mass production.
- Hire an in-house cryocooler expert.
  - That person could supervise an in-house development team

- If development is done by contracting, an in-house expert is still needed for proper oversight.

REFERENCES

[1] E. Murphy, "ngVLA Community Studies Program: Call for Proposals," issued August 2016. Version dated 2016 Sep 22 available as of 2017 Aug 24 at https://science.nrao.edu/futures/ngvla/ngvla-community-studies-program-call-for-proposals .

[2] E. Murphy, letter to Larry D'Addario, dated 2016 Nov 2.[3] L. D'Addario, "Advanced Cryocooling Technologies." Presentation at *Developing the ngVLA Science Program Workshop*, Socorro, NM, 2017 June 26. Slides available at https://science.nrao.edu/science/meetings/2017/ngvla-science-program/presentations/monday/advancedCryocoolers-Socorro-June2017.pdf .

[3] L. D'Addario, "Advanced Crycooler Technology." Next Generation VLA Science Workshop, Socorro, June 2017.

[4] S. Weinreb, "Front End Options." VLA Electronics Memo 101, 1972 Apr 10.

[5] J. V. Jelley, "The potentialities and present status of masers and parametric amplifiers in radio astronomy." *Proceedings of the IEEE*, **51**:30-45, Jan. 1963.

[6] J. A. Giordmaine, L. E. . A. Giordmaine, L. E. Alsop, C. H. Mayer and C. H. Townes, "A Maser Amplifier for Radio Astronomy at X-Band," *Proceedings of the IRE*, **47**:1062-1069, June 1959.

[7] E. R. Wiebe, "Low-Noise Receivers: 3-Kelvin Refrigerator Development for Improved Microwave Maser Performance." *Deep Space Network Progress Report* vol. 42-28, May-June 1975. Available: https://ipnpr.jpl.nasa.gov/progress_report2/42-28/28H.PDF.

[8] T. Mimura, "The Early History of the High Electron Mobility Transistor (HEMT)." *IEEE Trans. on Microwave Theory and Techniques*, **50**:780-782, March 2002.

[9] L. R. D'Addario, "An SIS Mixer for 90-120 GHz with Gain and Wide Bandwidth." *Int. J. of Infrared and Millimeter Waves*, **5**:1419-1422, 1984.

[10] T. Inaguchi, M. Nagao, and H. Yoshimura, "Two-Stage Gifford-McMahon cycle Cryocooler Operating at about 2 K." *Proc. Sixth International Cryocoolers Conf.*, Vol. II, p 25, 1991.

[11] A. Orlowska, M. Harman, B. Ellison, "Receiver Cryogenic System." Chap. 6 of *ALMA Project Book*, 2001 Jan 29. Available: https://www.cv.nrao.edu/~demerson/almapbk/construc/chap6/chap6.pdf.

[12] L. D'Addario, "Recommended Tests of Cryocoolers and Compressors." Informal memo dated 20 July 2017. Reproduced as Attachment 1 of this report.

[13] B. Butler, "Some Issues for Water Vapor Radiometry at the VLA." VLA Scientific Memo. No. 177, 1999.

[14] M. Pospieszalski, "On the Dependence of FET Noise Model Parameters on Ambient Temperature." IEEE Radio and Wireless Symposium, Phoenix, AZ, Jan. 2013.

[15] J. Schleeh et al., "Phonon black-body radiation limit for heat dissipation in electronics." *Nature Materials*, **14**:187-192, 2015. doi:10.1038/nmat4126.

[16] Low Noise Factory, http://www.lownoisefactory.com .

[17] Cosmic Microwave Technology, http://cosmicmicrowavetechnology.com .

[18] R. Selina and E. Murphy, "ngVLA Reference Design Development and Performance Estimates." ngVLA Memo #17, 2017 July 18

[19] S. Weinreb and H. Mani, "Low Cost 1.2 to 116 GHz Receiver System – a Benchmark for ngVLA." Next Generation VLA Science Workshop, Socorro, June 2017.

[20] H. Mani, private communication (email of 9/10/2017).

[21] D. Urbain *et al*., "Improved Power Efficiency for Cryogenics at the Very Large Array." *Cryocoolers 19*, Proc. of 19th International Cryocooler Conf., 2016.
[22] Sadi Carnot, *Réflexions sur la puissance motrice du feu et sur les machines propres à développer cette puissance (Reflections on the Motive Power of Fire and on Machines Fitted to Develop that Power)*. Paris, 1824.
[23] Sterling's original paper.
[24] Gifford and McMahon, original paper.
[25] R. Radebaugh, "Development of the Pulse Tube Refrigerator as an Efficient and Reliable Cryocooler." *Institute of Refrigeration Proceedings*, v 96, London 2000.
[26] C. S. Kirkconnel and K. D. Price, "Thermodynamic Optimization of Multi-Stage Cryocoolers." *Cryocoolers 11*, R. G. Ross Jr, ed., Kluwer/Plenum Pub., 2001.
[27] L. D'Addario, file 'cyrocoolers.xls'. Available from the author upon request.
[28] Sterling Cryogenics, unit of DH Industries BV, models SPC-1T and SPC-4T: https://www.stirlingcryogenics.com/en/products/cryocoolers/spc-1t-cryogenerator, https://www.stirlingcryogenics.com/en/products/cryocoolers/spc-4t-cryogenerator.
[29] T. Conrad, B. Schaefer, R. Yates, D. Bruckman, M. Barr, M. Kieffer, "Low Temperature RSP2 Production Cryocooler and Electronics Performance." *Cryocoolers 19*, Proc. of 19th International Cryocooler Conf., 2016.
[30] Brian Schaefer and Ted Conrad, Raytheon Corp., El Segundo, CA. Private communication (email of 6/23/2017).
[31] William E. Gifford, "The Gifford-McMahon Cycle." Cryogenic Engineering Conference, Rice University, Houston, Aug 23-25, 1965.
[32] G. Cortes-Medellin *et al.*, "A Fully Cryogenic Phased Array Camera for Radio Astronomy," *IEEE Trans on Ant. and Prop*, **63**:2471-2481, June 2015.

ATTACHMENT 1: RECOMMENDED TESTS

A reproduction of reference [12] is presented on the following pages.

Recommended Tests of Cryocoolers and Compressors

Larry R. D'Addario
California Institute of Technology
20 July 2017

I. Introduction

This memo describes tests needed to characterize the performance of closed-cycle helium cryocoolers. It focusses on crycoolers with continuous-flow compressors, such as those based on scroll pumps or reciprocating pistons with valves. Such compressors are appropriate for Gifford-McMahon (GM) cryocoolers or GM-style pulse tube cryocoolers as well as a few others.

For cyclic-flow compressors and the corresponding Sterling cycle or Sterling-style pulse tube cryocoolers, similar tests can be done by measuring a subset of the parameters described here. For example, only one pressure needs to be monitored, and measurement of helium flow is not necessary.

Manufacturers generally specify the performance of particular compressor-cold head combinations, providing little or no data on the behavior of either part separately. Because it is possible to use different combinations than the manufacturer intended (including using a cold head and compressor from different manufacturers), and because the operating speeds of compressors and cold heads can be independently varied, a full understanding requires characterizing each part separately. This can be done without actually testing them separately provided that the gas conditions at the interface are fully determined. To do that, it is necessary to know the gas temperature, pressure, and mass flow rate. With continuous flow, it is sufficient to determine the average values over the refrigeration cycle; the pressures and temperatures are needed for both the supply and return lines, but mass flow need only be measured in one place since (by conservation of mass) its average value is the same everywhere.

II. Initial Conditions

It will be helpful to know the total quantity of helium in the system. If the total system volume is known, it is sufficient to measure the static charge pressure (in hydrostatic equilibrium at room temperature with everything off). Often the volume is not known, in which case a good way to know the helium quantity is to evacuate the system to a rough vacuum and then refill from a helium tank of known volume whose pressure and temperature are measured before and after.

III. Instrumentation

Instruments should be in place to measure the following quantities.

- P_s, T_s: Helium supply pressure and temperature[4].
- P_r, T_r: Helium return pressure and temperature[4].
- dm/dt: Helium mass flow rate.
- E: Compressor electrical input power. (This may require measuring load factor as well as rms voltage and current.)
- T_i: Temperature at refrigeration stage i for all stages.
- $H_i = dQ_i/dt$: Heat load at refrigeration stage i for all stages. This should be the heat dissipated in resistors whose electrical input powers are accurately controlled.

The cold stages should be carefully insulated to avoid any uncontrolled additional loads.

[4] Sometimes it is known that the gas temperatures at the interface are very nearly equal to the ambient (room) temperature, in which case it may not be necessary to measure them directly.

They should be inside a good vacuum to avoid gas conduction; they should be enclosed in multi-layer insulation to avoid radiation; and the sizes, lengths, and materials of wires for temperature sensors and heaters should be carefully chosen to make solid conduction negligible. Thermal mass of materials attached to the cold stages should be minimized to allow the system to reach equilibrium rapidly when conditions are changed during testing.

A system is needed for sampling all sensor outputs and recording their readings under software control. Several different physical and logical interfaces or protocols may need to be supported, since sensors are likely to be from different manufacturers. Some of the measurements may be available from sensors built into the compressor or cold head.

Whereas we are primarily interested in average values of gas pressures, temperatures, and flow rates over the refrigeration cycle, it is not necessary for those sensors to have fast response times. However, the sensor outputs should be sampled at least 3 times per time constant and those readings should be averaged over several refrigeration cycles to produce one measurement. Usually the same is true of the sensors measuring the cold stage temperatures T_i , but sometimes there is significant variation over a cycle and a desire to observe that variation.

IV. Tests

A. Load Map. A load map can be created by systematically varying the heat loads H_i and recording the resulting stage temperatures T_i after thermal equilibrium is reached. This can easily be automated if the heating resistors are connected to computer-controllable power supplies. If there is only one stage, the map is a 1:1 function $T_1(H_1)$, which can be represented as a Cartesian plot. For two stages, we have two functions $T_1(H_1,H_2)$ and $T_2(H_1,H_2)$, which can be graphically represented in several ways, e.g. as two contour maps plotted on the same axes. For three or more stages, graphical representation is more complicated but data can be collected in a similar way.

However, unlike the load maps published by cryocooler manufacturers, ours should include measurements of P_s, T_s, P_r, T_r, dm/dt, and E at each load point $\{H_i\}$. This allows some disentangling of the performance of the compressor and cold head. If possible, load maps should be measured with one or more of the interface variables held constant. For example, supply pressure P_s could be held constant by adjusting an orifice-type bypass valve from supply to return, and return pressure P_r could be held constant by adjusting the compressor speed.

B. Cool Down. It is useful to monitor all variables during cool down from room temperature. In this case, it is also necessary to monitor the residual pressure (vacuum) in the chamber, since it is in practice difficult to obtain a high enough vacuum to avoid significant conduction through the residual gas until the temperature at some internal point is low enough to freeze out the dominant molecules in air (N_2, O_2, H_2O). Cool downs should be observed with no loads ($H_i = 0$ for all i) and with all loads near their maximum useful values. If the latter prevents reaching the temperatures achieved in a load map at the same loads, then it may be because of residual gas conduction; in that case, the loads should be progressively reduced and the cool down repeated until that limitation is avoided.

C. Compressor only. The compressor can be accurately characterized if an adjustable orifice-type valve is connected from supply to return in place of the cold head. By measuring P_s, T_s, P_r, T_r, dm/dt, and E over a range of valve settings, the compressor efficiency can be determined for each setting.

D. Variation of charge. The manufacture of a specific compressor-cold head pair usually specifies a nominal helium static pressure (supply and return) when the system is not operating and everything is at room temperature. This is a way of setting the total helium charge. However, that might not be optimum if the cold head is used with a different compressor or in a multi-compressor, multi-cold head system. To understand this, load maps (IV.A) and compressor performance (IV.C) should be measured over a range of charge levels, with the total amount of helium in the system determined (Section II).

E. Variation of compressor pumping rate. If the compressor speed is variable, then load maps (IV.A) and compressor performance (IV.C) should be measured over a range of compressor speed.

F. Variation of refrigeration cycle rate. If the cycle rate of the cold head is variable, then load maps (IV.A) should be measured over a range of cycle rates.

V. ANALYSES

A. Overall efficiency (CoP). The coefficient of performance H_i/E can be computed for any stage at each set of loads in the load map and at each measured condition (cycle rate, compressor speed, helium charge). For multi-stage refrigerators, it is usually most interesting to do this for the coldest stage, but it still depends on the loads on the other stages. A single CoP that is useful in that situation is given by [1]

$$CoP = \frac{1}{E}\sum_{i=1}^{K}\frac{T_K}{T_H - T_K}\frac{T_H - T_i}{T_i}H_i \tag{1}$$

where there are K stages (stage K being coldest) and T_H is the heat rejection temperature. This reduces to H_K/E if the only load is on the last (coldest) stage, but it gives normalized "credit" (based on Carnot efficiency) for cooling any higher-temperature loads. This enables us to find the conditions for optimum efficiency and to compare different cryocoolers.

B. Compressor efficiency. If compressor-only characterization has been carried out (IV.C), then it is possible to compute the thermodynamic efficiency of just the compressor. This is the ratio of rate of adding energy to the gas to the electrical input power E, and is given (see Appendix A) by

$$\eta_c = \frac{1}{E}\frac{R}{2m_H}(P_1 + P_2)\left(\frac{T_1}{P_1} - \frac{T_2}{P_2}\right)\frac{dm}{dt} \tag{2}$$

where subscripts 1 and 2 refer to the compressor inlet and outlet (or return and supply) conditions, respectively; R = 8.313 J/K/mol is the ideal gas constant; and m_H = 4.003 g is the mass of one mole of He^4 atoms. If the compressed helium at the outlet has been cooled to ambient temperature T_a by heat exchangers so that $T_2 = T_1 = T_a$, then this reduces to

$$\eta_c = \frac{1}{E}\frac{RT_a}{2m_H}\left(\frac{P_1^2 + P_2^2}{P_1 P_2}\right)\frac{dm}{dt} \quad . \tag{3}$$

REFERENCE

[1] C. S. Kirkconnel and K. D. Price, "Thermodynamic Optimization of Multi-Stage Cryocoolers." *Cryocoolers 11*, ed. R. G. Ross, Kluwer/Plenum Publishers, 2002.

Appendix A: Compressor Efficiency Derivation

The ideal gas law describes the relationship among pressure, volume, and temperature of and isolated set of gas molecules as

$$PV = NRT \tag{A1}$$

where N is the number of molecules. Thus the change in volume when the gas is compressed is

$$\Delta V = NR\left(\frac{T_1}{P_1} - \frac{T_2}{P_2}\right). \tag{A2}$$

The work done during compression is

$$\begin{aligned} W &= \int_{V_1}^{V_2} PdV \approx \frac{1}{2}(P_1 + P_2)\,\Delta V \\ &= \frac{NR}{2}(P1 + P2)\left(\frac{T_1}{P_1} - \frac{T_2}{P_2}\right). \end{aligned} \tag{A3}$$

When the gas is continuously flowing, the rate of work is

$$\begin{aligned} \frac{dW}{dt} &= \frac{R}{2}(P1 + P2)\left(\frac{T_1}{P_1} - \frac{T_2}{P_2}\right)\frac{dN}{dt} \\ &= \frac{R}{2m_H}(P1 + P2)\left(\frac{T_1}{P_1} - \frac{T_2}{P_2}\right)\frac{dm}{dt}. \end{aligned} \tag{A4}$$

The ratio of rate of work on the gas dW/dt to the electrical power into the compressor E is the compressor efficiency defined in Section V, giving equation (2).

The National Radio Astronomy Observatory and Green Bank Observatory are facilities of the U.S. National Science Foundation operated under cooperative agreement by Associated Universities, Inc. This work was supported by awards AST-2034328 (MSIP Prototype Antenna) and AST-2334267 (ngVLA Design Activities); NRAO related activities are funded under award AST-1647378 (NRAO Operations/Development).