CHAPTER 20

THERMAL CONTROL

> We should be careful to get out of an experience only the wisdom that is in it—and stop there; lest we be like the cat that sits down on a hot stove−lid. She will never sit down on a hot stove lid again—and that is well; but also she will never sit down on a cold one anymore.
>
> — Mark Twain

20.1 INTRODUCTION

We often need to control the temperature of some subsystem in an instrument; typical examples are power semiconductors, diode lasers, infrared detectors, and mechanical stages requiring extreme accuracy. These different jobs need different mixes of capacity and precision [1].

Power transistors which dissipate large amounts of heat need to be kept at temperatures below 150°C, but operate well over a wide range; a large passive heat sink suffices for the job. Infrared detectors must often be kept very cold, but normally the control requirements are modest. Their leakage is usually acceptable provided the temperature is kept below a certain upper limit, but accurate calibration of their dark current and shunt resistance may require control at the 0.1°to 1°C level.

[1] Most practical thermal control tasks involve cooling more than heating; accordingly, we will usually refer to the temperature-controlled surface as the cold plate.

Building Electro-Optical Systems, Second Edition. By Philip C. D. Hobbs
ISBN 0-471-24681-6

Single mode diode lasers can be run at room temperature, but their high thermal tuning sensitivity, of the order of 0.1 $cm^{-}1$/K (between mode jumps), makes millikelvin-level stability necessary for the best tuning accuracy. To compound the problem, diode lasers dissipate power themselves, which can lead to considerable thermal forcing if their operating power is changed.

Translation stages also work near room temperature, and have no significant thermal forcing, but they must be able to move freely, so you can't just swaddle them with styrofoam.

There are some instruments, such as detector systems for infrared telescopes, which must be cryogenically cooled. Cryogenic cooling is bound up inextricably with vacuum technology, which is a lore-intensive subject but beyond our scope; accordingly, this chapter concentrates on heating and cooling a bit closer to room temperature, mainly using heaters and thermoelectric coolers.

20.2 WHY DO I CARE ABOUT THERMAL CONTROL?

20.2.1 Thermal Expansion

The best known consequence of temperature swing is dimensional changes: objects expand when heated. The expansion increases all dimensions of the object by the same fractional amount (assuming it is of uniform composition). For example, heating a washer causes the hole diameter to grow as well as the outer diameter; the material does not expand into the hole. The fractional length change per degree, the coefficient of thermal expansion (CTE), is a material characteristic. When parts of your of your optical system change in length, bad things may happen. Diode lasers go out of collimation; etalons become mistuned; interferometers shift by many fringes' worth; lenses become loose in their mounts; translation stages bind. In serious cases, lenses shatter or leftover globs of epoxy tear chunks of glass out of prisms and mirrors.

These effects are of two basic kinds: mismatch between the CTEs of elements in contact, leading to stress, and thermal changes of path length. The normal tendency of things to continue to work when their dimensions increase by a part in 10^4 make uniform temperature changes benign for the most part. Where it matters, we can combat thermal expansion in a couple of ways: force the temperature to be constant (temperature control) or balance the expansions of different parts so that they sum to 0 (temperature compensation).

■ **EXAMPLE 20.1**

Stress Due To CTE Mismatch

Consider a BK-7 glass window (CTE=$8 \cdot 10^{-6}$/°C), 5 mm thick, held in an aluminum tube by a threaded aluminum ring which just touches the glass at 60°C. If the assembly is cooled to -40°, the window will contract by $5\text{mm} \cdot 100°\text{C} \cdot 8 \cdot 10^{-6} \approx 4\mu\text{m}$, while the space it occupies in the aluminum will contract by $5\text{mm} \cdot 100°\text{C} \cdot 23 \cdot 10^{-6} \approx 12\mu\text{m}$. The 8μm difference is taken up principally by stretching the aluminum. The axial stress S_A in the aluminum is found from Hooke's Law, assuming a uniform cross-sectional area:

$$S_A = E\frac{(12\mu m - 4\mu m)}{5mm} \approx 1.1 \cdot 10^8 \mathrm{N/m}(16,000\psi) \qquad (20.1)$$

where E is Young's modulus. If the tube has thick walls, the total force may be very large, enough to damage fine threads or even shatter the window. Even with thinner stuff, the window is likely to show pronounced stress birefringence near its edges (Yoder recommends 500 psi as the "birefringence tolerance" of glass).

20.2.2 Athermalization

[Huyghens' bimetallic pendulum; difficulty of cancelling nonlinear thermal expansions accurately over a wide range of temperature; inaccuracies due to CTE tolerances; mechanical instability due to long serpentine mechanical path; a good technique for cancelling small residuals, e.g. a brass or aluminum spacer in a quartz tube–you need only 1/45 as much aluminum as you do quartz.}

■ **EXAMPLE 20.2**

Transient and gradient response of a bimetallic structure

20.2.3 Thermal gradients

Gradients and transients are less benign than uniform thermal expansion, and they can do some surprising things. Acton gives the example of a 1 mile long railway track (L= 5280 ft), fixed at both ends. During the night, some practical joker comes and welds in an extra 1 foot of track ($\epsilon{\cdot}L$ = 1 ft). The (extremely stiff) track bends into the arc of a circle; how high off the ground is the peak of the arc? (take a guess before looking at the footnote)[2]. This is analogous to what happens when one side of an object grows a bit due to a temperature gradient.

Thermal gradients cause objects to bend by making one side longer than the other, and the results can be similarly surprising. An initially straight rod, subjected to a uniform gradient across its thickness, will curl into a circle of radius

$R = \frac{1}{CTE\frac{dT}{dz}}$

so that a length L held at one end will warp away by $L^2/2R$, or

$\Delta Z = \frac{L^2 CTE}{2}\frac{dT}{dz}$

If the rod rests on its two ends, the height of the arch in the middle is $L^2/8R$.

■ **EXAMPLE 20.3**

The Hot Dog Effect [3]

Consider a pair of 304 stainless steel rails for a translation stage, of length 10 cm, width 4 mm, and roughly square cross-section, sitting on a cold optical table. A warm He-Ne laser is attached to the slider of the stage. The laser dissipates 15W, mostly via conduction through the stage. From Table &1MaterialPropertiesTable., we find that α=15W/K/m, and the CTE is $9.6 \cdot 10^{-6}$/K. The temperature gradient in the steel is

[2] It's *44.5 feet*—the height is $h{\approx}L(3\epsilon/8)^{1/2}$.

[3] For those unfamiliar with North American food, barbecued hot dogs bend slightly.

$$\frac{dT}{dz} = \frac{P}{\alpha A} \tag{20.2}$$

or 1250K/m, a temperature drop of 5°C over 4 mm thickness. The rails will want to bend into circles of radius 83m, a runout of about 60 μm in 10 cm (15μm arch height), which is enough to make the stage bind if it is not constrained. If the upper rails are designed so as to bend to the same radius, the stage will still run freely, although lateral gradients in the rest of the stage may result in some twisting.

If the rails were held rigidly at the ends (as in the railroad track example), e.g. in a massive aluminum block, the average change in length of 2.5°C would produce an arch height of 10cm$(2.5° \cdot 9.6 \cdot 10^{-6}/°C \cdot 3/8)^{1/2}$ or 30μm, which is about twice that due to the hot dog effect in this example.

The bending problem is so prevalent, that if you're trying to track down a thermally induced mechanical problem, always look first for bending[4]. The figure of merit for resisting distortion due to thermal gradients is α/CTE (higher is better). A high α reduces the gradient and a low CTE reduces its bending effect. Brass rails would have been 3 times better, and hard aluminum five times, although they are much softer.

Aside: Teflon Teflon (polytetrafluoroethylene) is amazing stuff for some uses, but dimensional stability is not one of them. Teflon is probably the hot dog champion–CTE as high as 2900 ppm/K and α=0.25 W/m/K, making it almost 20,000 times worse than stainless steel for a constant heat flux. Fortunately, it squashes easily and is seldom used as a framing material.

20.3 HEAT FLOW

20.3.1 Heat Conduction in Solids

In media that do not themselves flow, heat transfer follows Fourier's law of heat conduction, where the heat energy $\dot{Q}$ flowing out of a surface element *dA* in unit time is

$$\dot{Q} = -\alpha \nabla T \cdot \mathbf{n} dA \tag{20.3}$$

where **n** is the outward-directed unit normal vector and α is the thermal conductivity. Using the divergence theorem, this becomes the heat equation,

$$\frac{\partial T}{\partial t} = \kappa \nabla^2 T \tag{20.4}$$

The constant κ is the thermal diffusivity, which is related to the mass density ρ and heat capacity at constant pressure c_P by

$$\kappa = \frac{\alpha}{\rho c_P} \tag{20.5}$$

One special case is a uniform 1-D thermal gradient, which simplifies to

$$\dot{Q} = -\alpha \frac{\partial T}{\partial z} \tag{20.6}$$

[4]The author heard this piece of advice from Dr. Erwin Loewen, of Richardson Grating Laboratories, a master of spectrographs and ruling engines.

This is the most commonly used formula for doing thermal transfer calculations by hand, because we are usually dealing with complicated shapes and with poorly understood interfaces—the approximation is as good as our knowledge. In this section we'll normalize everything to unit cross-sectional area—don't forget to put the actual area back in when you're choosing the heater or cooler.

[WORKING ON FIXING THIS SECTION UP TO USE POSITIVE OMEGA– JUST SWITCHED FROM exp(-i omega) to exp(j omega) AND TOOK COMPLEX CONJUGATES]

For temperature control, we need the frequency response of heat transfer, too (since we're eventually doing electrical engineering, we'll use $e^{j\omega t}$ for the time dependence). A half-space (z>0) of material, whose surface temperature goes as $e^{j\omega t}$ has *T* given by

$$T(z,t) = e^{j\omega t} e^{\frac{-1-j}{\sqrt{2}}\sqrt{\frac{\omega}{\kappa}}z} \tag{20.7}$$

Thus as a sinusoidal frequency component propagates into the material, it falls off as exp $(-(\omega/2\kappa)^{1/2}z)$, and suffers a phase delay of $(\omega/2\kappa)^{1/2}z$ radians. For a sheet of thickness *d*, insulated on the other surface (so $\dot{Q}$ =0 there), we can patch the $\pm$(1+j) solutions to get

$$T(z,t) = e^{j\omega t}\frac{\cosh\left(\beta\sqrt{\omega}\,(z-d)\right)}{\cosh\left(\beta\sqrt{\omega}d\right)} \tag{20.8}$$

where β is given by

$$\beta = \frac{-1-i}{\sqrt{2\kappa}} \tag{20.9}$$

If we compute $\partial T/\partial z$ at the surface in the two cases, we get

$$\dot{Q}(t) = \alpha\beta A\sqrt{\omega}e^{j\omega t}\dot{Q}(t)|z = 0 = \alpha\beta A\sqrt{\omega}e^{j\omega t}\tanh\left(\beta\sqrt{\omega}d\right) \tag{20.10}$$

and so we can compute *T*(*z*,*t*) vs ω for constant heating power, and get the transfer function from heating power in to temperature change out, with *z* as a parameter:

$$H(\omega|z) = \frac{T(z,\omega)}{\dot{Q}e^{j\omega t}} = \frac{1}{\alpha\beta\sqrt{\omega}A}e^{\beta\sqrt{\omega}z} \tag{20.11}$$

$$H(\omega|z,d) = \frac{T(z,\omega|d)}{\dot{Q}e^{j\omega t}} = \frac{\cosh\left(\beta\sqrt{\omega}(z-d)\right)}{\alpha\beta\sqrt{\omega}A\sinh\left(\beta\sqrt{\omega}d\right)} \tag{20.12}$$

Thermal mass m_{th}=mc_{P} represents the energy required to increase the (well insulated) object's temperature by 1°. Assuming that thermal diffusion is fast, the rate of temperature increase is given by

$$\frac{dT}{dt} = \frac{\dot{Q}}{m_{th}} \tag{20.13}$$

which agrees with the low frequency limit of (20.12). Of course, the temperature will not continue to increase indefinitely, because the plate is not perfectly insulated. If the thermal resistance from cold plate to hot is θ, the low frequency response goes as $1/(1+i2\pi f\theta m_{\mathrm{th}})$, and is asymptotically constant at f=0. For a sufficiently well insulated cold plate, this pole appears well below the loop bandwidth, and so is of little importance.

20.3.2 3-D Heat Conduction

[Talk about 3-D thermal conduction, and how much faster it is–a 1 micron spot drops half its delta-T across the first 1 micron of material, which makes quite an impressive gradient, so that small thermal devices mounted on substrates are *fast*. (The Green's function for Laplace's equation in a spherical geometry is 1/r, so 1/(1 micron) is twice as large as 1/(2 microns), and that doesn't even count the difference between a sphere and a flat surface, which eventually gives you another factor of 2. The net effect is that for a uniform disk of radius a, in homogeneous material thick compared to a, the temperature rise is

$$\Delta T = P/(\pi K a) \tag{20.14}$$

20.3.3 Thermal Properties of Materials

Table 24: Thermal properties of common materials. These have been collected from various (and often contradictory) sources, so use them as a guide to material selection and early design, but check carefully with your supplier or do your own measurements in critical applications.
[Table &1MaterialPropertiesTable.]

Conductor Material: Copper Phosphor−Bronze Constantan Manganin Nichrome
Composition: "pure" 95% Cu, 5% Sn 60% Cu, 40% Ni 83% Cu 80%Ni, 20%Cr
Thermal Conductivity (W/m*K) 410.0 34.0 21.0 16.0 9.5
Electrical Resistivity (uOhm*m) 0.016 0.115 0.489 0.490 1.103
Inverse Product (K/(W*Ohm)) 150421 255754 97430 127634 95469

Material	α W/(K·m)	Linear CTE ppm/K	c_P J/(kg·K)	$10^{-3}\cdot\rho$ kg/m^3	$10^5\cdot\kappa$ (m^2/s)	ρ_E $\mu\Omega\cdot m$ ($\dagger\epsilon_r$)	1/(α K/(μ
Metals							
Gold	317	14.2	129	19.3	12.733	0.022	143.
Silver	417	18.9	236	10.5	16.828	0.014	171.
OFHC Copper	390	17	390	8.96	11.2	0.017	150.
Copper Wire	120-220	17	390	8.96	3.4-6.3	0.0175	
Phosphor Bronze (94.8 Cu, 5 Sn, 0.2 P)	48	18	376	8.9	1.43	0.115	181.
Free-Cutting Brass	110	20	390	8.5	3.3	0.06	151.
1100-T0 soft Aluminum	240	23	900	2.7	9.9	0.03	138.
6061-T6 Aluminum	180	23	900	2.7	7.4	0.05	111.
Nickel	90	13	440	8.9	2.298	0.07	158.
German silver (82Ni 18Ag, trace Zn)	23	16	400	8.7	0.661	2.5	17.
304 Stainless	15	9.6	470	8.0	0.40	0.57	116.
Carbon steel (0.4% C)	80	12	450	7.8	2.279	0.45	27.
Manganin	23	18	410	8.4	0.668	0.42	103.
Nichrome	13	17	430	9	0.336	1.08	71.
Constantan	22	17	400	8.4	0.655	0.71	64.
Material	α W/(K·m)	Linear CTE ppm/K	c_P J/(kg·K)	$10^{-3}\cdot\rho$ kg/m^3	$10^5\cdot\kappa$ (m^2/s)	ρ_E $\mu\Omega\cdot m$ ($\dagger\epsilon_r$)	1/(α K/(μ
Inorganic Dielectrics							
Aluminum Nitride (sintered)	180	4.6	750	3.3	7.273	9†	
Alumina (sintered)	35	7	800	3.6	1.215	10†	
Beryllia	300	8.5	1100	2.9	9.404	7†	
Fused Silica	1.38	0.52	750	2.2	0.084	†	
Corning ULE glass	1.3	±0.03	745	2.21	0.079		
BK-7 Glass	1.1	8	700	2.5	0.063	†	
Sapphire (crystalline Al_2O_3)	32-35 anisotro-pic	7	770	3.6	−116.703		
Pure Silicon	146	2.6	712	2.33	8.801		
Highly doped Silicon	100	2.6	712	2.33	6.028		
β-Silicon Carbide	42 (ceramic) 350 (single-crystal)			3.2	??	100-200	
Silicon Nitride			711		??		
	630 (strongly						

(For references quoting α and k in English units, 1 BTU/hr = 293.1 mW, 1 BTU·in/(hr·ft^2·°F) = 0.1443 W/m/K, so 1 W/m/K is just about 7 BTU·in/hr·ft^2·°F. For older metric units, 1 cal/cm2/sec/°C = 418.7 W/m/K.)

■ **EXAMPLE 20.4**

How fast is heat conduction? Let's plug a few numbers into (20.7). If the material is a thick piece of 1100-T0 aluminum, with κ =9.9 · 10 $^{-5}$m^2/s, a 1 Hz excitation will decrease by 1/*e* in 5.6 mm, and will be phase shifted by 1 radian in the process. In 304 stainless (κ =0.4 · 10 $^{-5}$), it's only 1.1 mm, and in plastic or generic thermal grease (e.g. that joining the sensor to the heater or cooler), only 100-400 μm. Thin layers behave a bit better than this would suggest, because of the effect of the opposite boundary, but don't try using a thick layer of glue to attach your sensor.

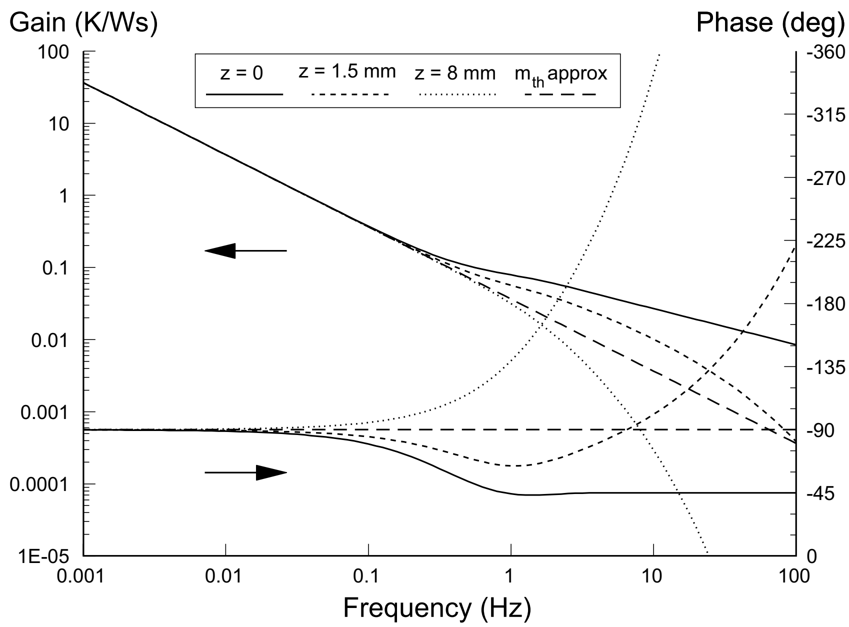

Figure 20.1. Gain and phase transfer function of thermal diffusion in an 8mm thick plate of 6061 aluminum, for sensors mounted at z=1.5 mm and z=8 mm.

Note that because of the power law, you lose bandwidth quadratically with thickness: doubling that 5.6mm of aluminum reduces the 1/*e* bandwidth to 0.25 Hz.

If the aluminum block is 1 cm square and 8mm thick, we can compute the gain and phase shift due to thermal diffusion for temperature sensors right at the heater (*z*=0), placed in a small drilled hole at *z*=1.5mm, and clamped on top at *z*=8 mm, as shown in Figure 20.1.. The curves go like 1/*f* near dc, because the response is dominated by the thermal mass. The fun starts where the effects of the other surface begin to die out. The curve for z=0 just goes smoothly from f^{-1} to $f^{-1/2}$, because the mass of material being heated at that rate starts declining as $f^{\frac{1}{2}}$, which is okay. As soon as there is any material at all in between, though, the exponential falloff kicks in and high frequencies become inaccessible. Note however that the phase shift 1.5mm into the 8 mm of aluminum is acceptable out to $f \approx 15$ Hz, so that a really fast sensor is needed to get the best performance—we'll almost certainly be limited by the diffusion of heat in the glue or grease if we're not careful.

20.3.4 Thermal Interfaces

Heat sink grease is actually a very poor thermal conductor, but it's better than air; use it for low-performance applications only, and keep it very thin. If you need better heat transfer, consider soldering or using silver-filled epoxy such as Circuit Works 2400. Solder is best if you can use it—high thermal conductivity, low cost, very convenient. It doesn't stick to everything, though, and does require high temperatures. Indium solder wets glass and many ceramics, and melts at much lower temperatures, so it may be a good alternative when tin-lead or tin-cadmium can't be used.

Silver epoxy can be very good, but is much more expensive and considerably touchier as regards curing—your results can easily vary by a factor of 10 depending on formulation and application. Cure it properly, by baking at 100°C or higher for several minutes. Improvements of 8°to 10°C in T_{cold} have been reported with a single stage TEC by replacing the greased joint between the TEC and the (water-cooled) heat sink with a thin layer of silver epoxy. **NB:** That "properly cured" bit isn't automatic, and is vitally important for its strength, dimensional stability, and electrical and thermal conductivity; be sure to bake the epoxy according to the manufacturer's instructions, and measure it to be sure you're getting what's claimed. Sometimes there are unpleasant surprises lurking.

For the best thermal resistance, it is important to lap parts together with fine polishing compound and clean carefully before final assembly. If you lap with abrasive in aqueous solvent, rinse with a nonpolar organic solvent such as naphtha, and vice versa. This is an old piece of black magic that seems to prevent the grit from becoming embedded in the surfaces and wedging them apart. Ideally, the surfaces should stick together when assembled, as Johansson gauge blocks do. Then you know they're really matched.

The order of preference for readily-available thermal interface materials (best to worst) goes: No interface, solder, indium solder, indium-gallium slurry, good silver epoxy, heat sink epoxy, heat sink grease, Krazy Glue, elastomer pads (e.g. Silpads), any other glue, air, vacuum.

20.3.5 Interfacial Thermal Resistance

If you calculate the expected thermal properties of your assembly by adding up the thermal resistances of all the thicknesses of all the materials in the thermal path, the answer you get will probably be slightly optimistic, especially with thermal grease, thermal epoxy, and fluxed solder paste joints. In fact, if you plot the thermal resistance of a joint made with these materials as a function of their thickness, you get the slope you'd expect from Table &1MaterialPropertiesTable., but with a significant positive offset, seldom less than 30 μ Km ^{2} /W and as much as 10 $^{-4}$ m^{2}K/W (0.05 to 0.2 in^2 °C/W, for readers of American heat sink catalogues).

Aside: Phonons There is a lot of disagreement as to the origin of this effect. Physicists talk about phonon impedance mismatches and diffuse phonon scattering from the interface. This appears right for liquid-solid interfaces, especially when the liquid is a simple one like helium. For solids-solid interfaces, the theory works much less well, e.g. predicting a $1/T^3$ temperature dependence, whereas the measured values tend to be constant above about 150 K, and to go roughly as T below that. Other contributions are poor wetting, voids, and (for greases and epoxies) changes in their average composition and the geometric relationship of the high thermal conductivity filler particles to the surface.

Silver is always used for electrically-conductive epoxy, because it oxidizes quite slowly, and its oxide also has high electrical conductivity, an unusual property. This is potentially less important for thermal applications. The reason that filled greases and epoxies are so delicate is that the filler's α is three orders of magnitude better than that of the matrix, so the thermal properties are almost entirely determined by the small gaps between them. (A similar effect stabilizes the inductance of coils on gapped cores, see Section 14.3.7.) The joints between the silver filler particles depend on their shape, size distribution, and previous history. The material actually used is ball-milled flakes of 2-30 μm diameter. Interlocking surfaces are required for best performance, and that isn't going to happen as easily at a flat boundary. Roughening the surfaces with an acid wash and applying pressure both help. Some epoxies are improved by curing them very hot, to the point where they start to lose just a little bit of mass ($\approx$160°C), enough to bring the flakes closer but not enough to make the epoxy porous.

{the particle stacking is a key parameter–you want a distribution of particle sizes, so that the small ones can fill the gaps caused by the stacking of the bigger ones. One problem is that regions where the particles form columns tend to have poorer thermal conductivity–which is a bit surprising. What seems to be happening is that as the grease squeezes out, it washes the small particles out of the stacks, which reduces the thermal conductivity.}

20.3.6 Dry Interfaces

In Section 21.7, we saw that the microscopic contact area between two solid surfaces is approximately equal to the normal force divided by the yield strength of the material. Thus one way to improve thermal contact is to clamp the surfaces together harder. A better method is to lap them together with finer and finer polishing compound until they stick together when clean, and use a very thin layer of good thermal compound. Note that even reasonably good thermal compounds have 100 times the thermal resistance of aluminum, so thin means really thin—a few microns, preferably.

20.3.7 Greased Joint Problems

Greased joints that undergo significant repetitive thermal cycling will often develop voids in the grease, because the thermal cycling pumps the grease out of the joint in the same way that freezing water cracks stone—expansion pushes very hard under heating, but contraction hardly pulls at all under cooling. Air bubbles thus form, which then expand on the next heating cycle, and so on until you have a big void and some very hot devices. Narrow grooves in the heat sink provide a reservoir of grease to allow expansion and contraction without void formation.

20.3.8 Radiative Transfer

In Section 2.4.1, we saw that surfaces at finite temperature give off electromagnetic radiation. A point on a surface will receive radiation from π steradians. This is because the thermal spectral radiance

$$L_\nu(\nu,T) = \frac{c}{n}\cos\theta e_0(\nu,T) = \frac{2hn^2\nu^3\cos\theta}{c^2 \quad e^{\frac{h\nu}{k_B T}}-1} \tag{20.15}$$

contains a cosine obliquity factor—otherwise it'd be 2π. A surface of area A at temperature T_1, completely enclosed by walls at temperature T_2 will receive a net heat influx of

$$P_{net} \approx \pi A\sigma\epsilon^2\left(T_1^4 - T_2^4\right) \tag{20.16}$$

where ϵ is the average emissivity of the surface. This approximation breaks down if the area A is not convex (think of a teakettle, where the inside has area but communicates only through the small spout), if the absorbed power is a large fraction of the total emission of the walls, or if significant amounts of A's radiation is reflected back to A, e.g. if A were a filament at the centre of a polished metal sphere. You'll have to put a fudge factor in to take care of cases like that, for example by using the outside area of the teakettle plus the area of the end of the spout. (We're usually doing one-significant-figure calculations anyway.)

Table 25: Thermal properties of gases

[Table &1HeatConductionInGases.]

Gas	α(W/K/m) 25°C †0°C	BP	ΔH_{vap} J/g
Hydrogen	0.171†	20.3K	452
Helium	0.143†	4.2K	20
Dry air	0.026		
Nitrogen	0.025	77.3K	200
Argon	0.016†	87.5K	163
Propane	0.016	-43°C	455
CO_2	0.015†	-78.5°C	618
$CHClF_2$ (dust-off)	0.011	-40.8°C	233
Freon-21 HCFC $CHCl_2F$	0.0097	+9°C	417
Krypton	0.0090	121K	108
Xenon	0.0055	-107°C	96

[DO AN EXAMPLE]

20.3.9 Heat Conduction in Gases

For gas layers thicker than 1 mean free path, the thermal conductivity of gases is independent of pressure; Table &1HeatConductionInGases. shows the thermal conductivity of common gases. At low pressures ($<\approx 1$ torr) the thermal conductivity starts to drop off towards 0 at 0 torr. Thermal conductivity for monatomic and diatomic gases goes roughly as $1/\sqrt{m}$, so that, for example, xenon has a much lower thermal conductivity than helium.

Xenon is pretty expensive stuff, so molecular gases such as butane or HCFCs are good alternatives at ordinary temperatures. Even CO_2 has 40% lower thermal conductivity than air, and is nearly as cheap; the HCFCs listed are competitive with krypton.

20.3.10 Convection

Natural convection is intrinsically complicated on short length scales, particularly in realistic surroundings. Heat loss from the heat sink depends on the temperature gradient at the surface and the thermal conductivity of the gas. Circulation brings cool air near the sink, thus sharpening the gradient and improving the cooling. Where air can flow easily, e.g. heat sink fins oriented vertically, convection is roughly linear for objects near room temperature, with the thermal resistance decreasing gradually as the object gets hotter, due to the gradual thinning of the boundary layer as the air flow increases, and the increased thermal conductivity of hot air vs. cold.

20.3.11 Getting Uniform Air Temperature

As you can see from the table, air is not a very good thermal conductor. This is great when we want to insulate our houses, but not so good when we need to eliminate gradients. What do we do?

The two basic strategies are isolation and stirring. If you have a room with air temperature gradients, putting a closed box around your setup, with a little fan inside stirring the air, will do a surprisingly good job of homogenizing the temperature. Two nested ones (each with its own fan) are even better; conduction through the box then becomes the limiting factor. Flowing air can be homogenized very effectively with a well-stirred plenum: a cardboard box plenum with a fan has been measured to exhibit temperature nonuniformities less than 3 millikelvins near room temperature[5].

20.4 INSULATION

There are two types of insulation: vacuum and still air (vacuum is better). Insulating materials such as spun glass batts and styrofoam are basically immobilized air: they work by preventing convective heat transport, and accordingly, the good ones all have about the same thermal conductivity as the gas filling their spaces. Near room temperature, they're especially useful for reducing the thermal forcing due to rapid air temperature changes, e.g. when somebody opens the lab door. Fibrous or loose particle insulation is usually unsuitable in instruments, as it's hard to control all the fluff and dust.

[5]Robert A. Pease, "What's all this box stuff, anyhow?", *Electronic Design*, August 22, 1991, p. 115

20.5 INSULATION AND THERMAL RADIATION

The other benefit of insulation is in suppressing thermal radiation. Thick layers consisting of strong optical scatterers (such as fibreglass or styrofoam) impede thermal radiation by diffusion, just as they do with visible light. Infrared-absorbing surfaces re-emit in a Lambertian pattern, which works just the same way as diffusion. This is a significant effect for layers thicker than a few millimetres. Vacuum systems and spacecraft sometimes use *superinsulation*, consisting of a series of layers of low-emissivity material (often gold-coated polyimide sheets). These are not diffusers, but rather more like separate insulators in series. All vacuum layers have the same thermal resistance, independent of thickness, so dividing the vacuum into N layers ought to improve the thermal resistance by N times. In practice it usually isn't this good, because all that extra surface outgasses a lot (especially since it's plastic underneath), so gas conduction increases. Superinsulation is also fairly hard to bake out, for the above reasons plus the lower temperature limits of even the best plastics.

20.5.1 Styrofoam

Styrofoam has good mechanical properties for an insulating material, being resilient and easily worked. Best of all, it is available everywhere at a very low cost. Its only major disadvantages are inflammability and low strength (NB: the fumes are toxic). Solid styrofoam is completely safe, chemically inert and moisture resistant. Use the soft stuff with the very fine holes, that squeaks when you rub it hard with your finger; the harder stuff with the bigger holes (that crunches instead of squeaking) is inferior unless you need its structural strength.

20.5.2 Dewars

The insulating quality of a decent vacuum is the basis for Dewar flasks, best known as Thermos bottles. Small Dewars are made of glass with a shiny metallic coating to reduce the thermal emissivity of the surface in the infrared. The low thermal conductivity of glass (especially at low T), and long and contorted thermal path minimize heat conduction.

Aside: Condensation Condensation is a common evil in cooled instruments. It corrodes mechanical parts, stains lenses, leaves deposits on optical surfaces, and promotes the growth of fungus and mildew. On circuit cards, it causes severe drift, leakage, and $1/f$ noise, and may destroy components and connectors eventually. All instruments must be condensation free, and that doesn't happen by accident. An instrument that stays dry in Scotland may drip in Louisiana.

Relative humidity is not an absolute measure of the water content of the air, but rather the ratio of the partial pressure of water vapour in the air to the equilibrium vapour pressure (i.e. the point at which fog just begins to form). Humid air in equilibrium with liquid is said to be saturated. The vapour pressure of water is a steeply increasing function of temperature. The equilibrium partial pressure of water is given by

$$p(\mathrm{H_2O})(T) = (0.061\mathrm{kPa})\exp\left(5209.5\left[\frac{1}{273.15\mathrm{K}} - \frac{1}{T}\right]\right) \tag{20.17}$$

At 30°C, air can hold three times more water than at 10°C. The dew point is the temperature at which a given body of humid air reaches saturation. The atmosphere forms a very large source reservoir, so if any exposed surface of your instrument is below the dew point, dew will form on it. Since $p(H_2O)$ increases so rapidly with temperature, mild heating is usually enough. If you rely on external insulation to prevent condensation, make sure it forms a hermetic seal, or it will eventually become wet, and the water will be trapped for long periods, which is especially bad. The best approach is to put the cooled system in a well sealed room-temperature can with a room temperature window, any necessary insulation being inside the can. Additional measures such as heating the window or even the can itself may be needed if the instrument is to be moved from cold environments to warm ones, e.g. airborne sensors or portable instruments brought inside during the winter. Dew point sensors can help determine when heating is necessary.

Condensation can occur even inside such a can, unless the air inside is very dry. Use HCFC (e.g. Dust-Off), dry gas, or really dry air. Dry air can be obtained by running room air through a cooled molecular sieve trap, or for limited temperature range prototypes, by putting the disassembled can in a chest-style domestic freezer until it equilibrates. Open the freezer a crack and put the cover on the can before taking it out. (Using a chest-style unit keeps the dry air from escaping when you open the freezer for assembly. A glove panel would be helpful, if you have one.)

Desiccants such as Drierite can be used to help control condensation, but watch out for their dust and make sure that the desiccant is kept cold; a -30°C cold finger can be a much stronger water vapour sink than room temperature desiccant, so all the water will eventually wind up there (warm silica gel is especially poor). In cryogenically cooled systems, molecular sieve materials are better, but with all desiccants, beware of dust; it's a poor trade to replace condensation with powdered desiccator dust. Permanently sealed vacuum systems usually use a getter such as metallic sodium to control residual water. If lasers are involved, it's usually best to use dry air, as this helps to control contamination (see Section 2.13.12).

In infrared systems, where a room temperature window may be a major problem, you can cool the window and use flowing dry gas (e.g. blowoff nitrogen from the cryostat) to keep the moist air away from the window. This will require some experimentation.

20.6 TEMPERATURE SENSORS

Ideally a temperature sensor has a response that depends only on temperature, has a fast response, negligible thermal mass, and imposes no thermal load, either by its own dissipation or by conduction through its leads. Real temperature sensors come in many sorts and kinds, but none is ideal. Of their nonideal characteristics, a few usually dominate, as listed below.

- Thermal contact Sensors read the sensor temperature, not the cold plate temperature. Accuracy depends on the two being identical, which requires excellent thermal contact and low dissipation.
- Response speed Most loops are limited in speed (and hence in control accuracy) by the thermal response time of their sensors and actuators. Small sensors are better, and IC sensors are very slow.

- Dissipation A sensor that dissipates significant power is an inaccurate sensor.
- Thermal Shorts Controlling the thermal conduction along lead wires is tough, because metal conducts heat well and often many wires are needed.
- Stress Effects All temperature sensors can be made to drift by mechanical stress. Since they are normally made of different materials than the cold plate, this is a serious source of error and drift with time. Solder the sensors to the cold plate by one lead only, use very small wire wire or flex circuit board for the other lead. Make sure the wire is thermally grounded to avoid thermocouple offsets.
- Thermocouple Offsets {A sensor connected by wires contains at least two thermocouple junctions in series with its sensing element, which will cause errors if they are not held at the same temperature. Glass- and metal-packaged devices are the worst for this, because their Kovar leads have an enormous thermocouple coefficient against copper. For thermocouple sensors, offsets and drift due to circuitry and unwanted junctions limits them to low-accuracy jobs.

20.6.1 IC sensors

There are a number of good IC temperature sensors, which produce a voltage or a current proportional to the temperature, such as the popular LM34/LM35 and LM335 (National) and AD590 (Analog Devices). They are stable and linear, so that $\partial V_{\text{out}}/\partial T$ is constant. Thus they do not contribute to the loop nonlinearity, which is an important benefit. Their output is very suitable for on-line checking, too, since their output slope is usually 10mV or 1μA per degree (F or C), so that the temperature can be read off directly from a DVM with fair accuracy ($\pm 2°$C typically).

On the other hand, they are not as stable as platinum RTDs or good quality glass bead thermistors, are somewhat sensitive to stress on their leads, have a narrower operating range, typically -40°to 100°C, and dissipate some heat (10μW to a few milliwatts) themselves. Their most serious disadvantages are that their thermal time constant is many seconds, which may easily dominate the loop response, and that their dominant thermal connection is via their leads, not by the air.

20.6.2 Thermistors

A thermistor is a type of carbon resistor whose value is a strong function of temperature, typically changing -3% to -4%/°C. These come in a wide variety of styles and prices, from less than $1 to $20 for the fanciest ones. For temperature control purposes, the best ones are small glass bead devices (1 mm diameter) with well specified temperature characteristics and reasonably fast time constant (1s). These can be imbedded in your cold plate with heat sink epoxy, and survive well, although their values may change slightly with mechanical stress. They come in resistor-linearized versions[6], and in interchangeable types whose $R(T)$ curves are identical within $\pm 0.2°$C, although strongly nonlinear. Their high stability, high sensitivity, and low noise makes thermistors the best choice for narrow-range temperature control systems, or those which don't need a human-readable indicator. Linearization is

[6] A resistance sensor of quantity x whose R vs. x curve is concave upwards can be linearized by putting a resistor in parallel with it. Carbon thermistors are like this, but platinum RTDs are concave downwards, so they need negative shunt resistances. This requires active devices.

worthwhile for wider temperature range applications since thermistors are so nonlinear that the temperature resolution at the upper limit will be poor otherwise.

Watch out how much heat you dissipate in the thermistor, and make sure that it doesn't change with time (it's OK if it changes a little with temperature since the thermistor is nonlinear anyway). Run it from a stable reference voltage.

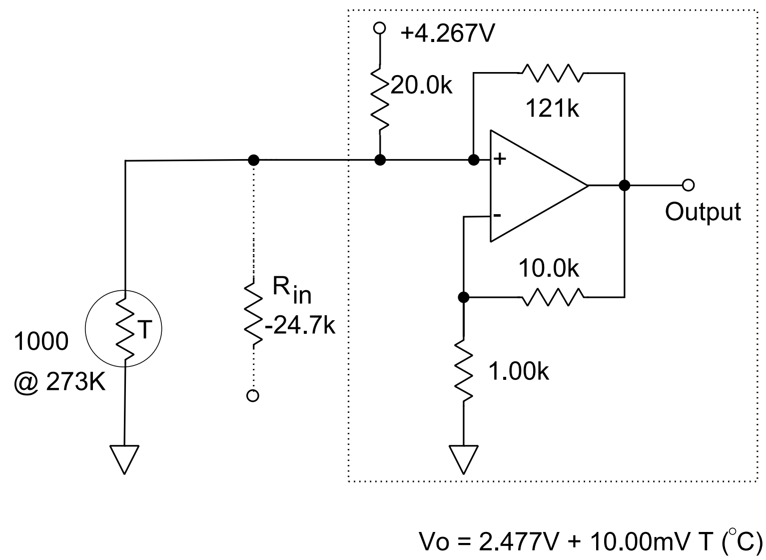

Figure 20.2. Negative resistance linearizer for platinum RTD; V vs. T is now linear to ± 0.2° from -150°C to +500°C.

20.6.3 Platinum RTDs

Platinum RTDs (resistance temperature detectors) are also thermistors, really. They come in wire wound and thin film types, from 100Ω to 1kΩ. Wire wound RTDs are very expensive, but the thin film ones are now below $25, and some as low as $6. Their characteristics are very stable and repeatable, not only over time and history, but unit-to-unit as well—after adjusting for the initial resistance tolerance, RTDs of the same type are interchangeable at the 0.02-0.1° level, limited by material purity and mechanical stress on the resistive element.

Their sensitivities are smaller than carbon thermistors', more like +0.35%/°C. They work over a wide range of temperatures, but cannot be linearized as simply as thermistors since $R(T)$ is concave downward. On the other hand, if you use a simple positive feedback circuit (e.g. the one in Section 20.6.3) to apply a negative load resistance of about −24.7 times the 0°C resistance of the RTD, you get a system linear to within ±0.2°C from −150°C to +500°C, and significantly better over narrower ranges. The 10 mV/K output slope of the circuit makes it easy to make human-readable thermometers.

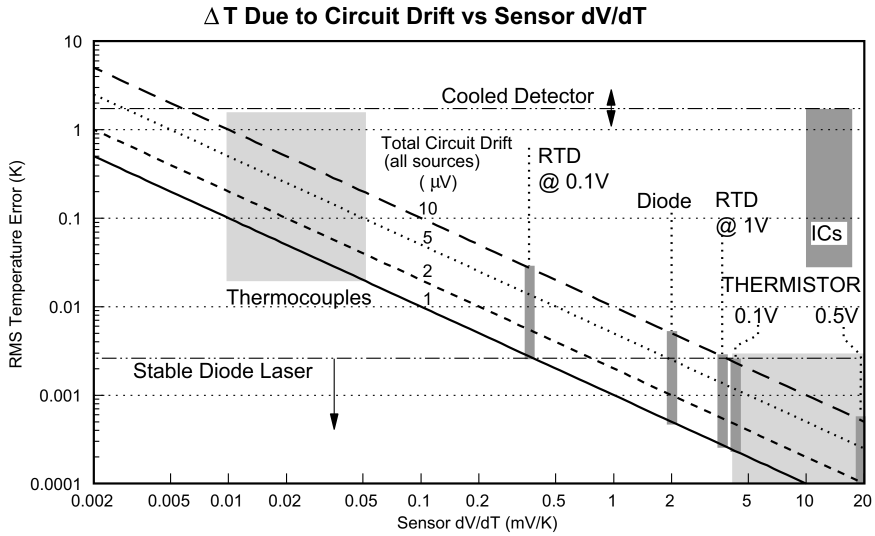

Figure 20.3. Attainable temperature accuracy vs. temperature sensor slope, for total circuit drifts from 1 μV (extremely hard to attain) to 10 μV (merely very difficult). Tight temperature control requires fast sensors with very high sensitivity–thermocouples don't make the grade except for the lowest-performance applications.

The very low 1/*f* noise of platinum RTDs (similar to metal film fixed resistors) makes them suitable for high stability control even though a 1kΩ RTD's voltage sensitivity is 30 dB worse than a 10kΩ carbon thermistor's for the same power dissipation (why?). Thin film RTDs are also fast; a 1mm alumina substrate has a thermal time constant of around 100ms, and thinner ones are faster. Their flat bottoms also make it easy to get a good bond with a thin layer of solder, glue, or grease. A fast RTD attached with solder or silver epoxy would be a good match for the thermal transfer function example above.

Table 26: Thermocouple Properties
[Table &1Thermocouples.]

Type	Composition	$\partial V/\partial T$ (0°C) μV/°C	$T_{\max}$
K	Chromel-Alumel	40	1250
T	Copper-Constantan	39	750
S	90%Pt/10%Rh-Pt	5.5	1800
J	Iron-Constantan	51	1400
E	Chromel-Constantan	59	900
N	Nicrosil-Nisil	26	1300
	Copper-Kovar		
	Silicon-Copper		
	Carbon-Copper		

20.6.4 Thermocouples

Thermocouples generate a voltage related to the temperature difference between two junctions of dissimilar metal wires, and are a pain in the neck. Their output is very small, and they require an outside temperature reference for the other thermocouple junction (the 'cold junction'), or a cold junction compensator based on an IC temperature sensor plus a judiciously chosen nonlinearity (you can't use resistor linearization because they aren't resistance sensors). Their low sensitivity makes them very susceptible to noise and pickup. They are the natural choice for sensing gradients, but apart from that their only advantages for instrument use are that they can be made very small, so that their thermal response can be fast, and that they dissipate no power. This matters in some applications, e.g. laser heating of delicate samples, where the heater has very fast response and the loop is detector-limited. They may also be useful in a composite control loop: use a dc-coupled integrating servo with a slower but more accurate sensor, but use a separate thermocouple and amplifier, ac coupled, to provide the high frequency response, something like the high voltage power supply stabilizer of Section 15.11.5. This avoids all the cold-junction problems, while keeping the speed. Copper-Constantan is a good choice for this sort of use, because both materials are solderable. For less specialized applications, avoid thermocouples like fleas. You can learn all you ever wanted to know about them from the Omega Engineering catalogue.

20.6.5 Diodes

A really cheap but not too accurate temperature sensor is an ordinary silicon diode, driven with a constant current. As we saw in Section 14.6.1, a diode has a temperature coefficient of about -2.1 mV/°C, although as usual transistors make better diodes than diodes do. The one great virtue of diodes as temperature sensors is this: a temperature sensing diode comes free inside every diode laser. You may know it better as the monitor photodiode, but it makes a really great temperature sensor. If you forward bias it by 1 or 2 milliamps, any plausible photocurrent won't perturb the measured temperature much; a 10μA photocurrent will cause a forward voltage shift of 100-200 μV, which is 0.05-0.1°C, and that can be corrected for anyway. This diode can be a remarkably fast sensor, because it is brazed right to the diode laser header—measurements indicate sub-second response times for 9 mm diode packages, indistinguishable from the response of a 75μm wire thermocouple silver-epoxied to the header. You can even use a simple chopping circuit to sense both the monitor photocurrent and the temperature.

20.6.6 Phase Change Sensors

Phase changes in materials are occasionally useful as sensors. Examples are dew point sensors, useful in preventing condensation, and ice point calibrators, which use the freezing point of some pure material (usually water) to provide a well buffered temperature reference. The best thing about these is that they're sure to correlate with what you're trying to measure.

20.6.7 Preventing Disasters: Thermal Cutouts

Instruments break. A circuit fails; someone blows up a power amplifier by dropping a screw down inside the case and causing a short; someone puts a magazine down over the cooling louvres and overheats the whole box; the list is endless. In power circuits, this can cause a fire, or collateral damage such as singeing the circuit card or delaminating lenses. It is therefore important to have some sort of thermal cutout to chop the power before this happens.

There are two basic kinds: thermal switches based on bimetallic strips (like thermostats) or metal-loaded polymers, and thermal fuses. Thermal fuses have to be replaced after blowing, but thermal switches either reset themselves or can be reset with a pushbutton. If your instrument can dissipate more than 50W under fault conditions without blowing a fuse, it should have one of these.

20.7 TEMPERATURE ACTUATORS: HEATERS AND COOLERS

Heaters are 100% efficient, and everybody knows how they work. There are three problems with using heaters for temperature control. The first is asymmetric slewing—if the temperature overshoots, it has to recover via thermal losses, not by refrigeration. Unless the heat loss is comparable to the heater capacity, the cooling rate will be smaller than the heating rate. Since the steady-state temperature change is proportional to the rate of heat input $\dot{Q}$, the slew rate in each direction is limited by how far from that temperature limit we are, regardless of the actuator we're using; with heaters, one temperature limit is at ambient, which is very inconvenient. The slew asymmetry of a heater becomes extreme within a few degrees of ambient, and that's a major impediment to good thermal control if heaters alone are used. Next most important is uniformity: how to make sure that the parts that need more heat get it, so that the whole assembly remains at constant temperature. The third one is nonlinearity; a resistance heater is approximately a square law device, since heat dissipated is I^2R, and R is nearly constant. You can get round this one if you use linear control, and put both the heater and the transistor on the same plate—since the supply voltage is constant, the total dissipation is VI.

20.7.1 Electric Heaters

Electric heaters come in a very wide variety, from a power resistor or a small coil of resistance wire used to prevent water from condensing on a window, up to big Chromalox elements for industrial ovens. The most useful type in instruments is just a thin Nichrome film imbedded in a silicone rubber or polyimide sheet. These can be tailored to give a desired profile of dissipation vs. position, which is very useful for maintaining temperature uniformity across the whole plate.

Heating is in principle nearly instantaneous, but in practice it isn't trivial to get good transient response with most heaters. They are usually made of low thermal diffusivity materials, and it's hard to make a good thermal interface between the heater and the hot plate, especially if the gradients under transient conditions are large.

For small-scale applications, e.g. temperature control of a single diode laser package, or vernier control of a TEC, you can use ceramic surface mount resistors, which have decent thermal performance because they're made of alumina. If you need a low power heater with flat surfaces to mount something on, Caddock's precision surface

mount resistors make nice little ceramic heaters of 2015 or 2520 size[7], and you can pick whatever resistance value suits your heat requirement and supply voltage. The resistance element is down inside the ceramic where it can't get damaged.
Manufacturers: Minco, Omega, Chromalox, Kyocera, Caddock.

20.7.2 PTC Thermistors

A PTC (positive temperature coefficient) thermistor's resistance increases by 3 orders of magnitude in a narrow temperature range (choices range 40°C-300°C), and typically has a slope of 25%/°C in that region. It can thus be used with a fixed bias voltage to make a nifty combination heater and temperature sensor. At minimum resistance, the heating powers available range from 1W to several kW per device. You can't beat it for simplicity and reliability, and since it's a proportional control, you get no switching transients.

The regulation is fairly poor. PTC thermistors are a good match to coarse temperature control applications such as heating hygroscopic IR windows to keep them dry, periodically baking out small vacuum systems, keeping metal vapour lamps warm, that sort of thing.

PTC thermistors have poor settability (the device switches where it feels like it) and are hard to get leads on. Because of the thermal shock at turn-on, soldered leads die very quickly from metal fatigue, so use clamps instead. There will be a certain amount of sliding around between the clamped thermistor and its mount, so make sure the mount is made of a soft material such as brass or aluminum.

20.7.3 Thermoelectric Coolers

A thermoelectric (TE) cooler is a thermocouple run backwards. Although this wouldn't work too well with chromel/alumel thermocouples, when you use P- and N-doped bismuth telluride ceramic, you wind up with a cheap, effective cooler that can support a maximum temperature drop of about 60−70°C (newer Bi-Te-Cs materials appear to be better). TECs can heat as well as cool, by reversing the current. Once cooling is admitted, the slew asymmetry problem gets better, but is not eliminated since the available heating and cooling capacity are strong functions of the temperatures of the controlled box and the environment.

Since each thermocouple drops only a tiny voltage, TECs are made by wiring dozens of them in series. Each junction forms a finger joining the hot and cold plates, which are typically made of alumina, lapped very flat to facilitate mounting. The entire sandwich is usually 3−8 mm thick, and sizes from 10 mm to 40mm square are readily available.

TE coolers are not as good as real refrigerators, due to I^2R losses and heat conduction through the fingers. Convection inside the active device contributes at large ΔT. This of course assumes that we don't do anything too dumb, for example using short pieces of large-diameter copper wire to connect to the device on the cold plate, or mounting the TEC upside down, so that the fat power supply wires wind up on the cold side of the device.

[7]These are EIA surface mount outline designations; 2015 is 0.2×0.15 inch (3.8×5.1 mm) and 2520 is 0.25×0.2 inch (6.4×5.1 mm).

TE coolers are especially suitable for use right around room temperature, with near zero temperature drop across them. There, the insulation requirements are modest, the power dissipation is relatively small, and the cooling capacity is large, so that the cold plate can be slewed up or down nearly symmetrically.

Cheap TECs don't stand repeated wide range thermal cycling very well. A hundred cycles from -40 to +85°C (with a fixed heat sink temperature of 10°C) will kill one. There exist slightly more expensive ones constructed with hard solder that will work reliably near 200°C, and stand lots of temperature cycling.

One very important point: there's no such thing as a heat sink. Heat is energy, and energy is neither created nor destroyed. Heat near ambient temperature is especially hard to get rid of, so don't make it worse with a poorly thought-out cooling strategy. TECs are great for *control*, and good though less than great for cooling low-dissipation devices somewhat below ambient, but they are terrible for cooling high-dissipation devices. If you have something that's running too far above ambient, putting a TEC on it will make it much worse rather than better–the TEC's dissipation will get added to the device dissipation, and all of it will have to go through some sort of heat exchanger, e.g. a fan-cooled aluminum extrusion. In that case, you're much better off with a better heat exchanger and no TEC.

Manufacturers: Melcor, Marlow, Ferrotec, Tellurex.

20.7.4 Mounting TECs

TECs are mechanically somewhat fragile. Bismuth telluride isn't the world's strongest ceramic, and due to the cooler design, the weak ceramic is the only source of mechanical strength. Like most ceramics, it is strong in compression, weaker in tension, and hopeless in shear or peel. Cracking occurs where the ceramic is under tension, so applying a compressive preload of about 120N/cm^2 (200 psi) to the cooler helps its shear strength, as well as improving thermal transfer by thinning out the grease and helping to eliminate voids. Unfortunately, it is sometimes difficult to do this without dumping large amounts of heat into the cold plate by conduction through the clamp. One approach is to apply the preload via the cold-plate insulation.

TE coolers have fairly large CTEs in the thickness direction. Although the cold plate may be at a well-regulated temperature, the rest of the TEC isn't, so there is a net expansion as the heat sink warms up. Beware especially of diode lasers sold in TO-3 cans with integral TE coolers—they'll go out of focus on you as the heat sink temperature changes. Another gotcha is the small CTE of the alumina end plates, which causes a big stress concentration at the heat sink and cold plate interfaces. A rule of thumb is that for large ΔT, don't use silver epoxy or solder to mount coolers larger than 15mm square. Use indium paste[8] (or thermal grease if you must) and clamping screws instead. If you have more than one TEC, and are working at significant ΔT, clamp those ones too. Otherwise the shear stress caused by the cold plate shrinking and the hot plate expanding will crack rigidly mounted TECs. If you need better location of the cold plate than clamping can provide, use silver epoxy on one TEC and clamps on the others (remember the big CTE of the TEC in the thickness direction).

The screws can be a serious thermal short, especially if they are too large. Stainless steel is very strong, but you can use only so small a screw. A #2-56 stainless steel

[8] Indalloy paste, from the Indium Corporation of America.

machine screw (1.9 mm mean diameter) spanning a 1.5 cm gap at ΔT=65°C will conduct 180 mW, and a brass one nearly 1W. If your application allows it, use nylon screws; a #6 nylon screw (2.6 mm) in the same place would conduct only 6 mW. Nylon screws tend to loosen with time and temperature cycling, so use spring washers and don't expect too much of them. Stretched steel wires with spring-loaded mounts would be a possible solution.

20.7.5 Heat Sinking TECs

The ΔT is measured from the hot plate to the cold plate, so a poor heat sink or poor thermal contact will limit the minimum temperature we can reach. TECs working at large ΔT produce a huge amount of waste heat, so if you need a low temperature, you'll need a really good heat sink. A little aluminum plate with fins like the ones used on CPUs won't cut the mustard; you'll probably need forced air cooling, or water cooling in extreme cases.

If you're determined to use thermal grease with TECs at large ΔT, make it extremely thin, 10μm or less. To achieve really thin grease layers, you have to make sure that both sides are very smooth and flat. Buy TECs with specified flatness, and use a fly cutter on a milling machine to remove any anodizing and surface irregularities from the heat sink. The fly cutter should do a good enough job on its own, but, if not, finish up by polishing the sink with crocus cloth, or with abrasive slurry and a spare TEC as a lap. Lapping the sink and TEC together with slurry (e.g. fine-grade automotive valve grinding compound) works well if you haven't got a machine shop.

20.7.6 Stacking TECs

TECs can be used in multi-stage arrays, where a large cooler cools the hot plate of a smaller cooler, which can then reach a lower temperature. You can buy them like that, or build your own. Some manufacturers are good at stacking, but some aren't—get specs from a number of them before buying a multistage cooler. The problem in building your own stack is getting the sizes of the units right and overcoming thermal drops in the interfaces and end plates.

If you're building your own, all stages should be of the same family, with the number of junctions going up by a factor of three per stage. Wire all the stages in series, thermally ground the wires at each stage, mount the TECs so that the leads come off the bottom of each stage, and size the wires so as to equalize the I^2R dissipation and thermal conduction. Take great care in assembling the thermal interfaces (see Section 20.3.4).

Asymptotically, you want to get the same delta-T per stage as you go down the stack, at least if we ignore the variation in the thermoelectric figure of merit. Since the area of the TEC goes up by a factor of 2 or 3 per stage, lateral drops in the alumina rapidly approach and then exceed ΔT_{max}, which means that the effective area becomes asymptotically constant. The heat load continues to increase as you go down the stack, so the attainable delta-T per stage eventually goes negative. The only way to combat this is to use spreaders. Eventually, if one were going to go to very many stages, one would need to space the upper-stage junctions further and further to avoid the lateral drops in the spreaders themselves. There really is no alternative to conquering the interface thermal resistance and using spreaders.

You don't get as much improvement as you'd expect with multiple stages, because the lower ones all operate with big $\dot{Q}_{\text{load}}$. A 4-stage cooler with a water-cooled heat sink can reach -100°C from room temperature, but not much farther than this, due to the poor efficiency of the coolers—the early stages just dissipate too much heat for the later ones to cope with easily. Use multi-stage coolers *in vacuo*, and be sure to thermally ground the power and signal wires to each stage, with enough slack to prevent really bad thermal shorts. Detailed design of these multistage coolers is nontrivial once you get beyond these rules of thumb.

20.7.7 Connecting to TEC Stages

The detector or source on the cold plate will need power, ground, and signal leads, which have to go from the last stage of the TEC to ambient temperature.

The figure of merit for interconnecting leads is the product of electrical resistivity ρ and thermal conductivity α. Although copper has an α 20 times that of some alloys, its electrical resistivity is so low that it wins the $\rho\alpha$ race. Copper is also highly solderable; the only problem is that you may find yourself using extremely fine wire, whose mechanical fragility makes it difficult to handle.

A 2.5 cm piece of #36 AWG copper magnet wire (127μm diameter) has an electrical resistance of about 30mΩ, can handle about 50 mA, and will conduct about 10 mW of heat across a ΔT of 65°C. This sounds small but can add up if you need lots of wires (small CCDs need around 20). Very thin copper wire is too weak and floppy to handle easily, so try fine solderable thermocouple wire such as constantan, whose α is 10 times lower (so you can use 3$\times$ coarser wire) and which is much stiffer. Diameters of 75μm or even less are easily available, e.g. from Omega Engineering. Make sure you use the same kind in both sides of the circuit, to eliminate thermocouple offsets. You don't have to worry about the junctions at the ends of the wire; there are an even number of them, wired so that their voltages cancel. The cancellation is perfect provided that the two hot junctions are at exactly the same temperature, and the cold ones are too. Make sure you *thermally ground* the leads by clamping or gluing them to their respective plate. If you're doing low-level dc voltage measurements, use manganin wire instead. It has very low thermocouple sensitivity with copper, and low α, but can't be soldered; use silver epoxy, spot welds, or ultrasonic bonding to attach it. Of course, if you have access to a wire bonder, very thin gold wires work well too.

A recent development is fine-line flexible circuit board. You can get 100μm thick Kapton film with 0.5 oz copper on one side, and connectors exist for it. Thin conductors etched on one side of that are probably the best of all worlds for wiring to cold fingers, especially if you have something huge like a CCD mosaic, where you might need dozens of wires.

20.7.8 Modelling TECs

The control characteristics of TECs are peculiar. Because they are thermocouples, they exhibit a voltage offset that varies with ΔT (in fact they are sometimes used as generators, as in spacecraft nuclear power supplies). Their I-V curve is also complicated and nonlinear; an extra half volt does something very different when ΔT is 5°vs. 50°. Voltage is therefore not a good control parameter—use a settable current source instead.

The heating and cooling characteristics of a TEC are a complicated mixture of Peltier cooling, I^2R heating, and thermal conduction, so that a simple control law is difficult to generate. They are strong functions of hot and cold plate temperature and current; if you drive a TEC too hard, I^2R will dominate and the cold plate will start warming up again—whoops, your negative feedback system just turned into a positive feedback one, and something will either rail or melt, in a hurry. This sounds unlikely, but it can easily happen if your heat sink cooling water is interrupted, for example. A thermal cutout on the heat sink temperature will prevent this. (We'll talk more about control circuits in a bit.)

This strong nonlinearity makes it useful to have a model for how TECs work, and one that can be extracted easily from data sheet parameters: maximum cooling power $\dot{Q}_{\max}$, $\Delta T_{\max}$, and $I_{\max}$. TECs get a bit more effective as the temperature goes up, but not enough to prevent T_{cold} from going up as T_{hot} does; ΔT goes up about a quarter as fast as T_{hot}. These fine points make detailed modelling complicated, but we can get an approximate expression for the cooling power. We know that the cooling capacity goes quadratically to 0 at $\dot{Q} = 0$, I=$I_{\max}$, and $\Delta T_{\max}$, and also at $\dot{Q}_{\text{cold}}$= $\dot{Q}_{\max}$, $I = I_{\max}$, and ΔT=0. From these relations, we can derive the very useful equation

$$\dot{Q}_{cold} = \dot{Q}_{\max}\left(\frac{-2I}{I_{\max}} + I^{\frac{2}{I^2_{\max}}} + \frac{\Delta T}{\Delta T_{\max}}\right) \tag{20.18}$$

where $\dot{Q}_{\text{cold}}$ is the heat flowing into the cold plate (negative for cooling). The heat flowing into the hot plate $\dot{Q}_{\text{hot}}$ is equal to the VI product plus the net heat extracted from the cold plate,

$$\dot{Q}_{hot} = VI - \dot{Q}_{cold} \tag{20.19}$$

and is equal to

$$\dot{Q}_{hot} = VI + \dot{Q}_{load} \tag{20.20}$$

in the steady state. The thermocouple coefficient of a single TEC junction is about 430μV/°C, and this dominates V for low currents. At higher currents, this voltage offset is reduced, and the voltage is dominated by the resistance, $R_{\text{TEC}} \approx V_{\max}/I_{\max}$. Conservatively, we'll sum the two to get a rough upper bound for VI, which gives us an upper bound for T_{cold},

$$T_{cold} \approx T_A + \theta_{SA}\left(\frac{I^2 V_{max}}{I_{max}} + I\cdot(430\mu V\cdot N\Delta T)\right) - \Delta T_{\max}\cdot\left(\frac{2I}{I_{\max}} - \frac{\dot{Q}_{\text{load}}}{\dot{Q}_{\max}} - \frac{I^2}{I^2_{\max}}\right) \tag{20.21}$$

where N is the total number of thermocouple junctions in the TEC and θ_{SA} is the thermal resistance of the heat sink. For CCD camera cooling specifically, see Petrick.[9]

Aside: Transient Response of Peltiers

[9] S. Walter Petrick, "Generalized approach to cooling charge-coupled devices using thermoelectric coolers", *Optical Engineering*, Oct 1987, V26 N10, p. 965.

20.8 HEAT SINKS

20.8.1 Natural convection

What you're buying in a heat sink is a way to dissipate lots of power without a huge temperature rise, i.e. a low sink-to-ambient thermal resistance θ_{SA}, measured in °C/W. Decent air cooled heat sinks operating with natural convection have a θ_{SA} roughly inversely proportional to the volume they take up. Their θ_{SA} decreases with temperature, by about 40% from its zero dissipation value at T_{sink}=100°C. One gotcha is that quoted θ_{SA} is the average from 0 to 100°C, so that close to room temperature, you have to add around 30% to the claimed value.

Natural convection is very easy to screw up by putting things in its way: it doesn't turn corners well. A louvred box, for example, can easily triple θ_{SA} if you're not very careful. Especially watch out that the heat sink fins are oriented vertically, so the air can easily flow upwards between them, and don't do anything to plug the ends even a little.

20.8.2 Forced air

Forced air can improve θ_{SA} by a factor of 5 to 10 if you move air fast enough (3-6 m/s). Resist the temptation to save money on the fan; noisy or failed fans are common causes of field failures. Forced air heat sinks operate very far from the thermodynamic limit, because they are really designed for natural convection and retrofitted for forced air; a properly designed heat sink (not available commercially unfortunately) can be a factor of 10 better for the same volume, albeit at a higher back pressure. Watch out for fan vibrations and noise in your measurement apparatus, and be sure to investigate how much back pressure your fan will encounter—quiet fans don't handle back pressure well. {FAN OPERATING POINT, FAN LAWS (ROTRON/kordyban)}

sic (corrected) fan laws
When fan speed changes When air density changes
Air flow CFM2 = CFM1 (RPM2/RPM1) MASSFLOW2 = MASSFLOW1 (density2/density1) Pressure P2 =P1 (RPM2/RPM1)2 P2 = P1 (density2/density1)
Power HP2 = HP1 (RPM2/RPM1)3 HP2 = HP1 (density2/density1)
Noise N2 = N1 + 50 log10 (RPM2/RPM1) N2 = N1 + 20 log10 (density2/density1)

20.8.3 Water Cooling

Circulating water is about the best way to remove a lot of heat without a big temperature rise. The high thermal conductivity and specific heat of water make it an excellent coolant. Its low toxicity is helpful too. Particularly when using thermoelectric coolers, which put out a great deal of waste heat and have a limited ΔT, a water-cooled heat sink will make a big difference to your cold finger temperature.

20.8.4 Phase change cells

If your thermal load has a small duty cycle, consider using a phase change cell to spread it out in time. Many substances melt at a very well defined temperature, and absorb a lot of heat doing it. The result is just like a big thermal mass in a small package. The phase change is a built-in thermometer; phase changes involve volume changes, so you can use a pressure sensor to control the heating or cooling of the cell. Fill the cell with copper wool or the equivalent to improve the thermal diffusivity.

For example, consider a focal plane array that needs to dissipate a few watts at $-90°$C. This is an enormous load for a stacked TEC, so it looks like a job for a mechanical 'fridge, a LN_2 cold finger, or (if we can stand going to -78°C) dry ice and acetone. However, if the duty cycle is 1%, the average $\dot{Q}$ of the 4-stage TEC would probably be large enough, and a small reservoir of frozen heptane (C_7H_{16}) on the cold plate (T_{MP}=-90.6°C, ΔH_{fusion}=141 J/g) would absorb the peak load and keep the sensor from warming up. The tradeoff is cost, maintenance, and complexity for the conventional approaches vs. longer cooldown times for the TEC and phase change cell. Be sure to choose a non-toxic material with a liquid range that extends well above room temperature, and put in a thermal cutout and a rupture disc; you don't want a nasty explosion if your controller breaks and the cold plate gets hot (heptane is a major constituent of gasoline, so it's pretty safe stuff apart from its inflammability, and it boils at about +68°C) Acetone (T_{MP}=-95°C, ΔH_{fusion}=98 J/g, T_{BP}=+56°C) is another good possibility. Watch out for vibration sensitivity and fragility when you put a big mass on the cold plate.

■ **EXAMPLE 20.5**

Thermal Spreader Plate.

Heaters and coolers tend to be somewhat nonuniform in their heat output. Even if they're perfect, any external heat load will cause gradients, which can be a real problem; make sure you use an aluminum or copper spreader plate to even it out. The base material of TECs is usually alumina, which is polycrystalline sapphire (Al_2O_3), *G*=35 W/m/K. Thin-film heaters are usually polyimide or silicone rubber, which are nearly 100 times worse than that.

Consider an alumina TEC plate, 30 mm square and 0.6 mm thick, with a localized heat load of 20W uniformly distributed across a 17 mm square patch in the middle (e.g. from another TEC stage). Solving the exact temperature profile is complicated, but a very simple approximation will show the problem: assuming that the heat has to travel laterally by 1 cm on average, and that the perimeter of the heat flow region can be taken to be 70 mm, the temperature drop will be of the order of

$$\delta T \approx \frac{\dot{Q}}{\alpha} \frac{10mm}{70mm \times 0.6mm} \approx 60°C \tag{20.22}$$

which is far too high. The moral of the story is that you have to use big chunks of aluminum to even out the temperature drop, if you're dissipating any amount of heat on the cold plate. Lateral gradients this large are not fantastic: people have destroyed multistage TECs built without spreader plates by simply turning them on too suddenly, causing these huge lateral gradients to develop and the hot areas to melt.

■ **EXAMPLE 20.6**

TEC on a passive heat sink

Passive heat sinks are poor matches to TECs used at large ΔTs, but are useful for applications near room temperature. Consider a 15×30mm rectangular TEC with

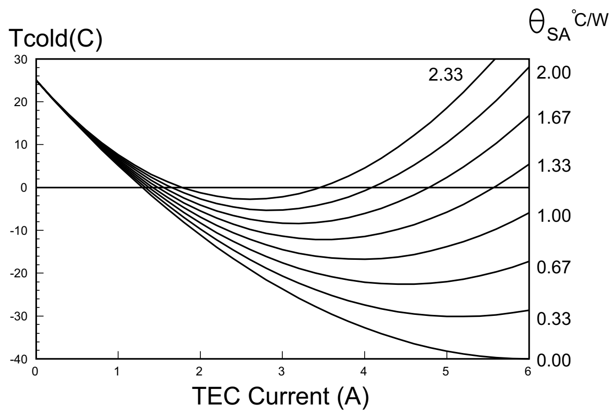

Figure 20.4. Designing a TE cooler system using a passive heat sink: cold plate temperature vs. I and heat sink thermal resistance θ_{SA}. (T_A=25°C)

I_{max}=6A, V_{imax}=4.5±0.5V, ΔT_{max}=65°C, and $\dot{Q}_{max}$=14W, [10] used with a finned aluminum sink at 25°C ambient. What should the thermal resistance θ_{SA} of the heat sink be? We can find this from (20.21) by setting $\dot{Q}_{cold}$ to 0. As we see from 20.6, T_{cold} is seriously degraded by a relatively small heat sink thermal resistance. If we need a cold plate temperature of -15°C, then we'd just make it with a 1°C/W sink, e.g a Thermalloy 6159B, 104×152×45 mm (which is fairly big). If the interior of our instrument can get to 40°C, which is not uncommon in racks or during the summer, the best T_{cold} we can get will be around -3°C with this heat sink. On the other hand, a fan can make the 2.33°C/W curve improve to the level of the 0.33°C/W one. The calculated heat sink temperature at the upper right is above 100°C, which will cook some TECs.

For smaller TECs than this, the little CPU coolers with integrated fans can be a good match, if you can stand the vibration.

20.8.5 Attaching Devices

Most of the time, the reason we're working this hard on cooling is so that we can get some device down to a temperature where it works better. Diode lasers are no problem, because they come in solid metal cases that are easily mounted. Detectors are trickier, especially imaging devices such as CCDs and infrared arrays. These usually come in ceramic packages with metal lead frames, and they are very expensive and

[10]Similar to a Ferrotec 6300/035/060A (around $12 in quantity 100).

very delicate, so we're not going to try silver-epoxying the package to the cold plate. One good feature of these packages is that the thermal conductance of the leads is surprisingly high. Thermally ground the leads by using a very thin circuit board with a ground plane on the back that is bonded directly to the cold plate with silver grease. Alumina works well for this. If you're adventurous, you may be able to find an indium solder that bonds well to the ceramic package, which is an additional help, and which is much easier to remove than silver epoxy.

20.8.6 Controlling TECs

The previous example shows that controlling TECs in high ΔT applications is a bit fraught. The sign of the loop gain changes at the minima of the T_{cold} vs I_{TEC} curves, causing immediate destruction of the TEC due to runaway heating unless we have a thermal cutout or well-chosen current limit. The problem is that the minimum moves, depending on θ_{SA}. The right approach here is to sense T_{sink} as well as T_{cold}, and watch for ΔT getting too large, either moving the set point, shutting down, or whatever is sensible in the application. Lower ΔT applications can just put in a current limit at some value below the lowest plausible value for the minimum T_{cold}, and watch for the current hitting the limit, because of a high ambient or reduced cooling.

How the drive power is applied matters too. Whatever you do, don't use bang-bang drive, e.g. a thermostat or unfiltered pulse width modulator controller, with a TEC.

The ripple current from the PWM (or other high-ripple source) transfers no heat but does cause I^2R dissipation, seriously degrading the ΔT. It may also hasten the death of the cooler by electromigration. Slower cycling, e.g. a dead-band controller like a domestic thermostat, is even worse; the massive thermal cycling encountered by the junctions and the solder in the TEC will send it to an early grave.

Voltage control is also a disaster. TECs are thermocouples after all, so their voltage drop depends on ΔT; they also have a low impedance, so that the current will go all over the map if you use voltage control.

Two control strategies work well: linear current control, where the operating current is varied in real time to keep the T_{cold} steady, or constant current drive with a variable heater on the cold plate to do the real controlling (the heater can be bang-bang controlled if you like). If you need a wide range or very symmetrical slewing, then linear control is the way to go. On the other hand, if you're operating in a narrow range near ΔT_{max}, the heater approach works very well, and can have faster response since the heater can be thin and laid on top of the cold plate. The heater approach is (somewhat paradoxically) also best for multistage TECs, since their transient response tends to be extremely complicated and hard to compensate.

The disadvantage is that in order for the temperature to be able to drop rapidly under transient conditions, the heater and TEC must be fighting each other quite hard under quiescent conditions, which degrades ΔT and increases power consumption (a partial solution would be to use linear control only on the last TEC stage).

Rejection of environmental perturbations requires high loop gain at the frequency of the perturbation. For a 40 dB improvement, you'll need a bandwidth 100 times larger than the forcing frequency. (You don't dare try anything fancy by way of frequency compensation–the loop slows down badly at the extremes of its range.) Since the ugly transient response of a multistage TEC limits us to sub-mHz bandwidths, we won't even be able to reject diurnal temperature changes, which are significant even in an air conditioned building.

By putting a small surface mount resistor on a small silver cold plate, you can get control bandwidths of a hertz, and by using a diode instead, you can dissipate power and measure temperature in exactly the same device, which is even faster. Another benefit is that the transient response is constant and simple, and the power dissipated in the diode is nearly linear in the applied current.

If the goal is to get T_{cold} as low as possible, just run the thing flat out, but when stability is a concern, you'll want a temperature as low as can be accurately maintained, not a few degrees lower at the price of drift. Since the heater's dissipation will go to zero at the limit of its control range, you don't lose anything except some electrical power by heating the last stage.

Aside: Transient Response of Peltiers There is one exception to the run-it-flat-out rule: you can get a few extra degrees for a few seconds by pulsing the cooler.[11] Peltiers absorb heat right at the junctions, which are soldered to the cold plate. Their I^2R heating is rather uniformly generated throughout the length of the bismuth telluride bars. One of the most important properties of a Peltier material is poor heat conduction, so by goosing the current, you can get several seconds' worth of lower temperature on the cold plate, before the extra heat generated further down the bars can warm it up. While this is a small advantage, it may might be quite helpful in marginal situations, e.g. ones involving a phase change or at the ends of the temperature-tuning range of a diode laser. Watch out for reduced lifetime due to the cycling, as with bang-bang controllers—you'll want to use hard-soldered TECs if you do this.

20.8.7 Mechanical Refrigerators

Mechanical 'fridges come in several kinds. The most popular is the Stirling cycle, a closed cycle with a floating piston. 'Fridges are a lot more efficient than TECs, but a lot more expensive and less convenient to use. If you don't need cryogenic cooling, you're better off saving power by using a really small TEC with good insulation than going to a mechanical 'fridge. You shouldn't consider using a 'fridge for lab use unless you have to go below 77K, which is pretty infrequent in the optical instrument business.

One problem with fridges is that they often have a lot of acoustic and electrical noise from their motors and motor controllers. The acoustics are mostly down in the 100 Hz-10 kHz region, but the electrical junk can get up to 1 MHz or so due to switching spikes. Unfortunately, of course, that junk gets coupled straight into your most sensitive point, the optical detector. If you use one, be aware that getting the grounding, shielding, filtering, and acoustic damping right will be nontrivial. Watch especially for the leads to the temperature sensors and optical detectors.

Table 27: Cryogens and low temperature mixtures

[Table &1Cryogens.]

Recipe Temperature K (°C)

Ice water 273.15 (0)

Ice and NaCl, 2:1 255 (-18)

Ice and $CaCl_2$, 1:1 225 (-48)

Dry Ice (CO_2) and Acetone 195 (-78)

[11] G. Jeffrey Snyder *et al.*, "Supercooling of Peltier cooler using a current pulse", *J. Appl. Phys.* **92**,3, 1564-1569 (2002)

Boiling Liquid N_2 (1 atm) 77.3 (-195.8)
Boiling Liquid He (1 atm) 4.2 (-269)

20.8.8 Expendable Coolant Systems

Open-cycle cooling systems are a better way to get low temperatures in the lab. The melting or boiling points of common safe substances are very convenient, and holes in the warmer end of the list of temperatures available can often be filled in with mixtures, as shown in Table &1Cryogens.. Note that the thermal diffusivities of these mixtures are very low, so that either stirring or a heat spreader made of copper or aluminum mesh will provide much better temperature stability (the boiling of LN2 and LHe provide stirring automatically).

Liquid nitrogen is the default cryogen. It costs about a buck a quart, boils at 77.3K, and has a reasonable heat of vaporization, 200 kJ/kg (water's is 2.26 MJ/kg). Nitrogen temperature is cold enough for almost anything optical, except far-IR photoconductors, which makes it a good choice. LN_2 is easy to handle, needing only neoprene gloves and a face shield to protect against splashes; it lasts for a while in an open-topped Dewar, so you can easily do simple experiments with it, such as freezing a metal resistor to do noise measurements, or measuring the shunt resistance of your InSb photodiode.

Liquid helium is not for the faint of heart. It is very expensive, has a small heat of vaporization (20 kJ/kg), and requires special Dewars that are also not cheap. Changing something in a helium Dewar takes a whole day if you have to bring it up to room temperature. Purpose-built "optical Dewars" are available with windows in them, but getting your beam in and out of one is nontrivial. The clear diameter is small and the path long, so you have to work near normal incidence; the multiple layers of windows wind up looking like a hall of mirrors, all aligned perpendicular to your beam. The etalon fringes are thus very large if you're using a laser. Circular polarizers can help somewhat.

The other problem with optical cryostats is vibration. The sample tends to be mounted on a long narrow cantilever, leading to resonances in the tens of hertz, which are easily excited by fluid convection or table vibration.

20.9 THERMAL DESIGN

20.9.1 How fast can we go?

For accurate control in the face of strong perturbations, we want lots of bandwidth. There are two reasons for this: firstly, we want fast response to rapidly changing thermal loads, such as a TE cooled diode laser being switched on; secondly, we need high loop gain at lower frequencies. As we saw in Section 15.4.1, phase shift is what limits how fast we can roll off the gain with frequency. That means that low frequency loop gain is purchased by stability at high frequencies—i.e. low and stable phase shifts.

Accordingly, thermal design is an exercise in maximizing bandwidth and reducing thermal forcing. There are two main bandwidth limiters: slow thermal diffusion, which we've seen already, and large thermal masses being controlled by small-capacity heaters and coolers.

Assuming for the moment that $\dot{Q}$ is locally linear in some control voltage, a temperature control loop has an integrator already in it, since we control $\dot{Q}$ but measure something roughly proportional to its time integral. The unity gain crossover frequency f_0 can be moved by introducing additional gain, and the phase shift can be changed locally by using an RC circuit in the feedback loop, just as we did in frequency compensating amplifiers and PLLs. The limit to these sorts of games is set by loop nonlinearity and excess phase shifts due to thermal diffusion.

Loop nonlinearity causes trouble in two ways: by making the turn-on transient response unpredictable, and by changing the loop gain as a function of hot and cold plate temperature. Any other source of parametric variation, e.g. cooling water temperature changes or changes in thermal mass, will do the same.

20.9.2 Material Choice

The thermal properties of different materials vary widely. Plastics are slow, because {UNDER CONSTRUCTION}

20.9.3 Placement of temperature sensors

As we saw in Example 20.4, the slow diffusion of heat puts a limit on temperature control bandwidth. Accordingly, we have no choice but to put the temperature sensors right at the heater or cooler, even if the heater temperature isn't what we care about.

20.9.4 Handling Gradients

We can't usually put the temperature sensor right on the laser or detector, but we can use common centroid design to cancel gradients. A symmetrical layout of sensors and actuators, wired up so that the effects of gradients cancel at the position of the active device can do an excellent job of correcting for external inputs.

Example 20.7: **Common Centroid Design Of A Diode Laser Mount**
The ISICL sensor of Example 1.13 used a diode laser mount as shown in Figure 20.5.. For packaging reasons, the mount had to stand upright. The laser's temperature was near 25°C, but had to be stable to 10mK or better, with heat sink temperatures from 15°C-45°C. Because of the longer heat conduction path of the upper TEC, a temperature gradient of a few degrees could exist, especially near the upper limit of T_{hot}, where the TEC is putting out a lot of waste heat. A symmetrical arrangement of the two TECs, with a matched pair of glass bead thermistors potted into drilled holes in the cold plate, just above the centre of the TECs, solved the problem. The thermistors were wired in series, so that when a gradient made TD_1's resistance increase slightly and TD_2's decrease by the same amount, the series combination continued to reflect the temperature at the midpoint, which was where the laser was. The thermistors were nominally 10kΩ at 25°C, and interchangeable at 0.2°accuracy; the position of the neutral point could be adjusted up and down the cold plate with a 100K potentiometer, wired as shown, to null out any residual gradient sensitivity.

The collimating lens and the spacer were also temperature controlled, which kept the focal length and the state of focus highly stable. The lens mount is an annular disc of fused quartz held on with a very thin (10 μm) glue layer.

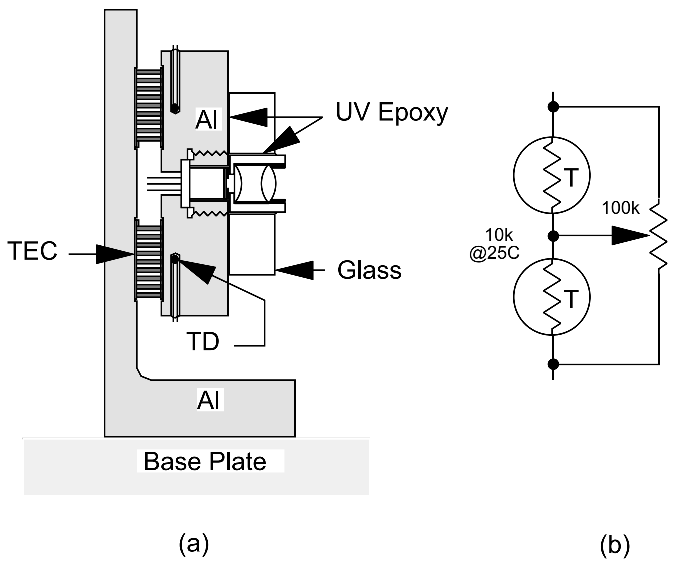

Figure 20.5. Cancellation of temperature gradient by common centroid design. (a) Thermistors placed symmetrically about diode laser. (b) Series connection gives $T_{\mathrm{avg}} \approx T_{\mathrm{laser}}$, allows vernier adjustment.

20.9.5 Is the Sensor Temperature What You Care About?

Controlling the temperature of an object with a single control loop is in some sense an ill-posed problem, since temperature is a function of position as well as time. What we care about is the temperature of the active region of the device (laser or detector) mounted on the cold plate. Because of thermal gradients and the speed of thermal diffusion, this may bear only an oblique connection to the temperature of the sensor.

You can make them more similar by making the cold plate out of copper, aluminum, or beryllia, making the thermal path from actuator to active device short and fat, and keeping the cold plate dissipation constant. More complicated solutions include using the actuator-coupled sensor for ac control, and a second one near or in the active device package to sense the device temperature (e.g. occasional measurement of the forward voltage drop of the monitor photodiode in the laser package, with the laser off).

Gradients can be reduced further by insulating and shielding the cold plate very carefully.

20.9.6 Dissipation On The Cold Plate

Since temperature control loops never have enough bandwidth to suit us, we usually have to work hard to reduce the high-frequency thermal forcing. High frequency forcing rarely comes from the environment, unless your system has to work even after being chucked into the ocean. The perturbations usually come from turning on the temperature controller or the active devices on the cold plate.

The turn-on transient can be dealt with in the controller, e.g. by using a baby-scale two speed loop (see Section 15.4.7), by feedforward, or just by enduring it, since it is usually infrequent. Active device turn-on or modulation should be nulled out with a heater whose dissipation keeps the total $\dot{Q}_{\text{load}}$ constant. Ideally, the spatial distribution of $\dot{Q}_{\text{load}}$ should be constant in time as well. For example, a diode laser mount should have a small heater right at the laser, driven so as to hold the total dissipation constant. This technique is a nice match to the slow cooler/fast heater approach to TEC control, since the load dissipation can be measured, and the heater power adjusted, much faster than the thermal system will respond. Slight errors in the computation (which is usually done with resistors) will be tracked out by the feedback loop eventually.

It is also possible to use a subsidiary control loop, based on a local heater and sensor, to control the active device temperature, but this approach needs very careful testing. Beware using two loops with comparable bandwidths—this leads to oscillations and flakiness.

20.10 TEMPERATURE CONTROLLERS

20.10.1 Bang-Bang Controllers: Thermostats

All this stuff about frequency compensating loops and linearizing $\dot{Q}(I)$ may strike you as overkill for your application, and maybe it is. After all, domestic heating and cooling are controlled with thermostats, and they seem able to handle a window being open or the oven being on. Why not just use a thermostat?

Thermostats are bang-bang (i.e. on-off) controllers with some temperature hysteresis built in; the heat goes on at 66°F and off at 69°. The temperature oscillates irregularly with time, but is usually between these limits, provided the heater has the right capacity and the thermostat is properly placed. Too large a heater, or slow heater-sensor coupling, will lead to pronounced overshoot, especially on heating, so that the peak to peak temperature ripple may be much larger than the deadband. If your application can live with these limitations, a thermostat can be just the right medicine (but see Section 20.8.6 for a caution).

20.10.2 Linear Control

A linear controller[12] continuously adjusts the heating or cooling to maintain a constant temperature. Because the controller has no dead zone, the temperature is much better defined, but the loop is more difficult to design, because oscillation must be avoided, and good transient response maintained. There are three broad types of strategies, referred to as proportional, proportional-integral (P−I), and proportional-integral-derivative (PID), depending on the time dependence of the loop filter. The names are a historical accident, dating from the time when the three components needed three separate modules, whose outputs were summed. Having separate adjustments of P, I, and D is convenient for setup, but a mistake in a commercial instrument: whatever can be adjusted can be misadjusted.

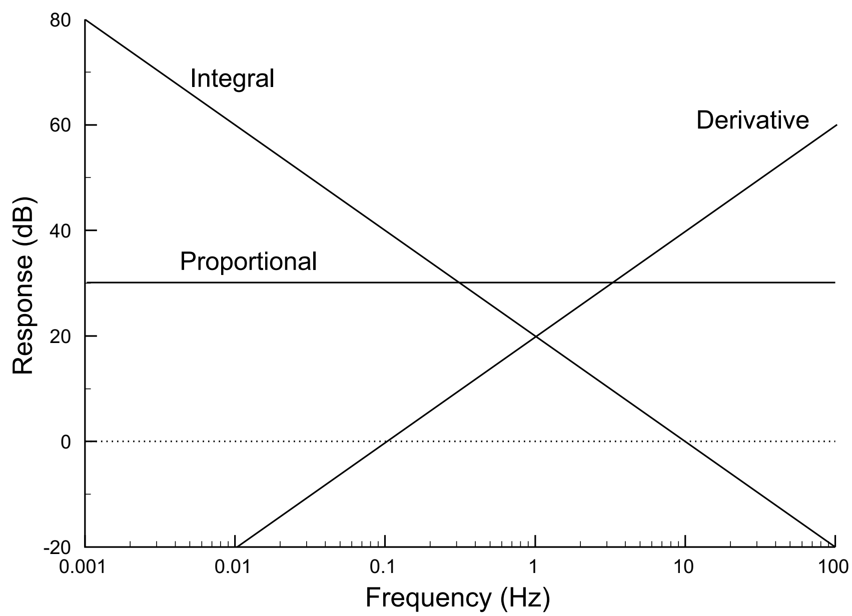

Figure 20.6. Proportional, integral, and derivative

A proportional loop is just a dc amplifier with constant gain A_{Vcl}, so that the transfer function $H(\omega|z,d)$ provides all the filtering; you just set the gain to a value

[12] They're commonly called *proportional controllers*, since the error signal is linear in the error, but this leads to confusion with proportional as opposed to integral and derivative.

where the loop remains acceptably stable over the range of loads and temperatures to be encountered. Its finite dc gain means that any thermal load will lead to a static temperature error (or *gain error*), and dialling the proportional gain up to reduce the error will lead to instability.

This error can be eliminated by using an integrating (P−I) loop, which is nothing more exciting than putting a largish capacitor in series with the feedback resistor of the loop amplifier. This makes the amplifier gain A_{Vol} extremely large at dc without messing up the high frequency performance. The extra dc gain kills the static error, which lets us reduce the proportional gain to improve stability at high frequency.

It is worth mentioning that an integrating loop usually has a much uglier turn-on transient than a proportional-only loop, because the integrated error signal builds up to a huge value during the slew period, which has to be discharged by an equal and opposite error after the setpoint has been crossed. In motor control loops this is called *windup*, and can cause the destruction of the controlled system. Windup is best prevented by using proportional term to do most of the work, adding in a small integral term to zero out the dc error, and limiting the total value of the integral term to a few times the maximum gain error of the proportional-only loop. A common method is to put back-to-back diodes across the integrating capacitor, and choose the other circuit constants to make the error of the proportional loop about ± 0.2 V. Alternatively, you can use a separate integrator that is allowed to saturate, and attenuate its output to the right range. You'll need to reduce its time constant by the same factor, so that its net gain stays the same in the linear region.

For devices like current-controlled motors, where the actuator has a two-pole response itself (the current sets the torque, which is proportional to the second derivative of shaft angle), we need a derivative term as well, so that the loop filter must be rising near the unity gain crossing. Loops that have this bathtub-shaped loop filter gain are called proportional-integral-derivative (PID) loops. They're worse than useless for high performance temperature control, because of the extremely large phase shifts at high frequency—the last thing we want to do is jack up the gain out there. Nowadays we have lots more control over our loop filter than just P, I, and D, but the name stuck.

Eliminating static error and increasing the loop bandwidth are very nice, but not a complete solution, remember, because we're still only controlling the temperature of the heaters or coolers. Nonetheless, a single layer of temperature control, with an insulated cold plate, can easily achieve 10 mK stability in the lab.

20.10.3 Frequency Compensation

Temperature control loops are intrinsically more subtle than ordinary op amps because we are measuring T but controlling $\dot{Q}$, which makes the temperature controller inherently integrating like a PLL at low frequency. Loss of communication with the edges of the plate adds a phase lead at mid-frequencies, and thermal diffusion between the actuator and the cooler adds an extra phase lag that can be very large at high frequencies. The main job is to make the loop as stable and wideband as possible, then work hard to make the controlled temperature equal the active device temperature, perhaps with a fixed offset due to device dissipation.

You can estimate what the open loop transfer function is from (20.12); don't be daunted by the unintuitive form—your favourite math program (or 20 lines of C++) will give you nice plots of it. Remember that $\dot{Q}$ is roughly proportional to I for a TE cooler, but to I^2 for a heater (both have a parabolic nonlinearity at large signals); you may need to linearize that parabola in order to avoid huge bandwidth variations and possible loop instability.

Ignore the heat sink response at ac, because the heat sink is normally much larger than the cold plate, and high frequencies ideally contribute no net heat over a cycle. This is a useful guide to the early stages of design, because it will help you estimate how much bandwidth you can achieve with a given mechanical design. (Multistage TECs are slow enough that you probably can't ignore the heat sink response. Typical control bandwidths for those are in the millihertz.)

Example 20.8: *Temperature Controller*
Let's try temperature-controlling the 1 cm^2 by 8 mm thick aluminum plate, with a fast sensor 3 mm above a fast ceramic thin-film heater of 25Ω resistance. We'll work well above room temperature, where the quiescent heater power is 0.25 W, so that V_h = 2.5 V and its gain K_h= $\frac{\partial \dot{Q}}{\partial V_h}$ = 200 mW/V, and we'll use a silicon diode sensor with K_s = -2.1 mV/K.

The response calculated from (20.12) and the desired overall response are shown in 20.10.3. The loop filter is a simple lead-lag network[13]

with a zero at 0.13 rads^{-1} to give us high accuracy at low frequency, while not destabilizing the loop. Now we need to make sure that the overall gain, which is including the TEC capacity, temperature sensor gain, and loop filter gain, make the open loop unity gain crossover occur near 13 rads/s (2 Hz). The required high frequency value is 4000 W/K.

The loop gain is the product of all the individual gains, A_VL = K_h · K_s · H_plate · H_amp. Since H_amp has flattened out well before the unity gain cross, but before the thermal mass approximation fails, this is easy to solve approximately. The low-frequency limit of the plate's response is

$$H(\omega' \,|z\,) \sim \frac{1}{j\omega m_{th}} \omega \to 0 \tag{20.23}$$

with m_th= $\rho c_\mathrm{P} V$, which is 8 × 10 $^{-6}$ m^3 · 2.7×10^3 kg/m^3 · 900 J/(kg ·K) = 1.94 J/K. Thus the high frequency limit of H_amp is

$$H_{amp} \approx \frac{13\mathrm{rads}^{-1} \cdot 1.94\mathrm{JK}^{-1}}{0.2\mathrm{WV}^{-1} \cdot 2.1 \times 10^{-3}\mathrm{VK}^{-1}} \approx 6 \times 10^4 \tag{20.24}$$

This is a pretty big number. If we say that the linear range of our heater is a volt, the loop will remain linear for temperature excursions of only 8 mK, so the settling behaviour is liable to be rather peculiar. This is a good application for a two-speed loop, which we discussed in Example 15.4.7. The difference here is that in this case, we want the fast loop time constant after settling instead of the slow one. To get a wider linear range, we can relax the speed requirement, use a bigger heater, linearize the parabolic heater characteristic, or use a more complicated loop filter that has lower gain at low frequencies.

Final optimization of the loop filter will be necessary, because (20.12) will be wrong in detail, especially at high frequency (say above 1 Hz). You will need to measure the amplitude and phase of the open loop response. Use hand-tweaking or a very, very slow loop filter to get into the right operating regime, and measure the step response of the combination when you put a small current step directly into the heater or cooler. Taking the Fourier transform of the result, divided by the transform of the

[13] See Example 15.1.

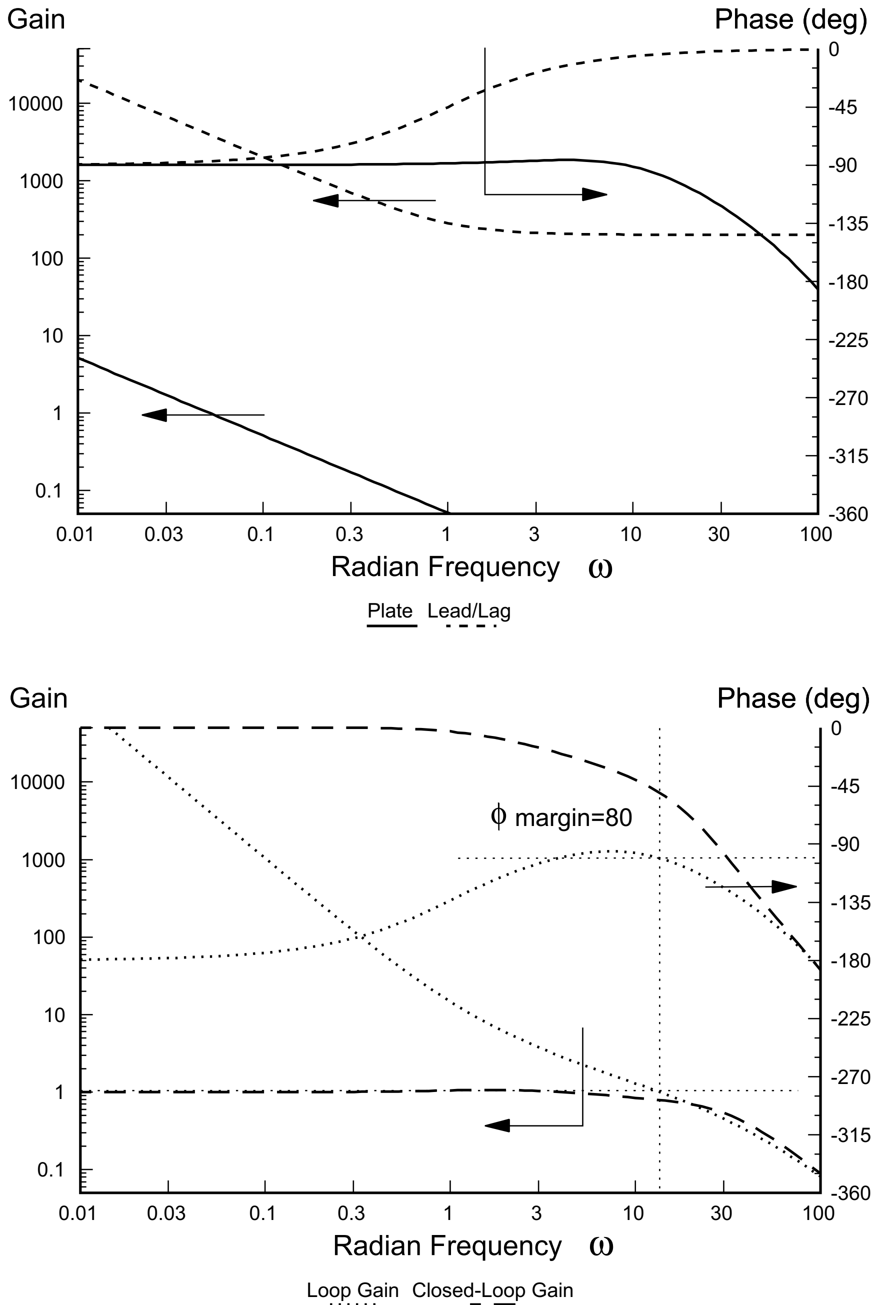

Figure 20.7. Bode plot of the 8 mm thick aluminum plate of Figure igure 20.1., with the temperature sensor at z=3mm, using lead-lag compensation. The unity gain bandwidth is over 2 Hz.

step function input, will yield a good estimate of the open loop response of the cooler-plate-sensor combination (remember to unwrap the phase before using it). The low frequency response will of course be wrong, since the loop is not really open down there, but since in the real circuit, the loop gain will be high at those frequencies, that isn't much of a problem. Make sure that you repeat this measurement over a sufficiently wide range of conditions (e.g. heat sink temperature and cold plate dissipation) that you have a good set of worst-case limits—you want the worst case to happen in the lab and not in the field. In particular, watch out for the way that loops slow down near the upper and lower temperature limits—your loop is liable to oscillate there if you've done anything very fancy with the frequency compensation.

Once you have a measurement of the open loop transfer function, proceed just as we did in Section 15.4.1, but use care; temperature control loops whose transfer functions are too tightly tweaked may fail due to unit-to-unit variations in TEC efficiency, thermal interface resistance, and so on.