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

All instruments are trying to be thermometers. That is, besides what we want them to see, our measurement gizmos see temperature too, not usually in a useful way. The trick is to make temperature sensors sense only temperature, and all other sensors ignore temperature completely.

There are three basic ways of doing this. First, we can make the sensor as nearly temperature-independent as we can. The optically contacted,[†] air-spaced Zerodur etalon of Example 2.2 and Section 4.2.3 is a good example. Second, we can *athermalize* the measurement by arranging things so their temperature dependences cancel, as in Harrison's gridiron pendulum clock (Example 20.2 and Section 20.2.3). Third, we can take our temperature-sensitive gizmo and keep its temperature very, very constant (temperature control). For the highest-stability measurements, we may well need to do all three.

[†]Section 12.13.3.

20.1.1 Temperature Control Regimes

We're always needing to control the temperature of various bits of an instrument, such as power semiconductors, diode lasers, infrared detectors, and mechanical stages requiring extreme accuracy. These different jobs need different mixes of capacity and precision.†

Power transistors which dissipate large amounts of heat need to be kept at junction temperatures below 150° C. They operate well over a wide range, so a passive heat sink and an overtemperature cutout are usually all we need. 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.

Single mode diode lasers usually like running 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. They're also liable to bind up if you push heat through them.

Some instruments, such as detector systems for infrared telescopes, have to be cryogenically cooled. Cryogenic cooling is bound up inextricably with vacuum technology, which is a hugely lore-intensive subject that is beyond our scope; accordingly, this chapter concentrates on heating and cooling a bit closer to room temperature, using resistive heaters and thermoelectric coolers (and the occasional bit of dry ice).

20.2 Thermal Problems And Solutions

20.2.1 Thermal Expansion

Besides the ice melting in your drink, the best known temperature effect is probably thermal expansion. The first thing to remember is that uniform heating affects all dimensions of an object by the same fractional amount (assuming its composition is also uniform). For example, if you heat a flat washer until its outer diameter grows by 1%, its inner diameter grows 1% as well, regardless of their relative sizes; the material does not expand into the hole. The fractional length change per degree, the *coefficient of thermal expansion* (CTE), is a material characteristic.

The normal tendency of things to continue to work when their dimensions slowly increase by a part in 10^4 makes uniform temperature changes benign for the most part. However, when parts of your of your optical system change in length, bad things can 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, at low temperatures threads strip, lenses shatter, and leftover globs of epoxy tear chunks out of prisms and mirrors.

†Most practical thermal control tasks involve cooling more than heating; accordingly, we will usually refer to the temperature-controlled volume as the *cold plate*.

These effects are of three basic kinds: thermal gradients or CTE mismatch in elements in contact, both of which lead to stress and bending, and thermal changes of path length, which makes focal lengths and fringes shift and operating points drift around.

Where it matters, we can use athermalization or temperature control.

EXAMPLE 20.1 Stress Due To CTE Mismatch

Consider a BK-7 glass window (CTE= $8 \cdot 10^{-6}$/K), 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

$$5 \text{ mm} \cdot 100^\circ \text{ C} \cdot 8 \cdot 10^{-6} \approx 4\mu\text{m},$$

while the space it occupies in the aluminum will contract by

$$5\text{mm} \cdot 100^\circ \text{ C} \cdot 23 \cdot 10^{-6} \approx 12 \ \mu\text{m}.$$

The 8 μm 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:

$$S_A = E \frac{(12\mu\text{m} - 4\mu\text{m})}{5\text{mm}} \approx 1.1 \cdot 10^8 \text{N/m}^2 \ \ (16{,}000\,\text{lb/in}^2), \tag{20.1}$$

where E is Young's modulus (69 GPa for 6061 Al). 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). That means that (unless other measures are taken) the average wall thickness of the aluminum cell needs to be at least 32 times narrower than the width of the contact patch between the lens and retaining ring, which is fairly surprising.

20.2.2 Compliant Mounts

To avoid the destructive effects of temperature extremes, we have to make sure that our optical mounts keep at least a minimum preload and never become so tight as to do damage. We usually do this by making a selected few of the mechanical parts slightly soft or springy. Lens tubes with the walls thinned way down between the threaded sections work well. Retaining rings can be made of plastic, for instance, or have flexures built in using slots.

Since the thermal expansion properties are known, it's possible to athermalize mounts, as we'll see in the next section.

In noncritical cases such as filters and windows, you can use an O-ring between the retainer and the optic, but of course that throws away most of the self-centering action of hard mounts (see Section 12.4.16), so you usually don't want to do it with lenses. It's possible, though fiddly, to do both by using two hard mounts during setup, gluing the optic to one of them, and then replacing the other with a compliant mount once the glue has set up, just to maintain the preload.

Once we've made sure the optics are safe, there are other things to worry about, such as defocus and etalon fringes due to thermal expansion and the TCN of the glass and the air.

20.2.3 Athermalization

It's possible to athermalize a structure in the same sort of way we made achromatic lenses in Section 4.12.1: use different materials arranged so their CTEs cancel. Like refractive indices, CTEs are nonlinear, so it's hard to do this really accurately over a wide temperature range, but usually we can make it good enough to be very useful.

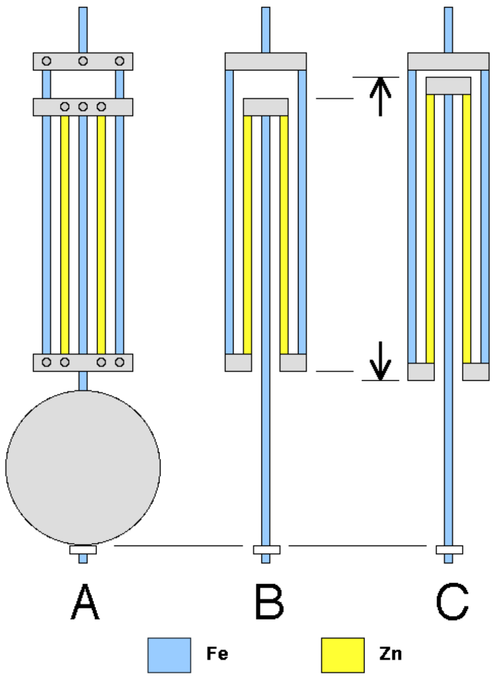

Figure 20.1 A 'banjo' pendulum using iron and zinc rods. The zinc rods have a much larger CTE than the iron but are much shorter overall, allowing the overall temperature coefficient of clock rate to be adjusted to nearly zero. (a) actual pendulum; (b) at low temperature the zinc contracts more than the iron, keeping the total length constant (exaggerated for clarity); (c) similarly at high temperature the zinc expands more, with the same result. The thermal transient response is far inferior to the Harrison gridiron pendulum owing to the bob being suspended from the bottom rather than its center of mass. (The image is from Wikimedia Commons.)

EXAMPLE 20.2 Thermal response of a bimetallic structure

A pendulum oscillating with a small amplitude has a period of $2\pi\sqrt{l/g}$, which is very nearly equal to two seconds for a 1-m clock pendulum. Because the period goes as $\sqrt{l}$, its tempco is half that of the pendulum material, which amounts to several ppm/K. Thus a 10 K shift results in an error of 30 minutes per year for an iron pendulum (CTE $\approx$ 12 ppm/K) or almost an hour for yellow brass ($\approx$ 20 ppm/K).

The celebrated English clockmaker John Harrison used this difference to stabilize the rate of his clocks with his 1726 invention of the *gridiron pendulum*. Harrison's brilliant insight is that a back-and-forth structure arranged so that iron wires expanding downwards compensated for brass wires expanding upwards could be designed so that their temperature coefficient canceled exactly. A gridiron using a back-and-forth pattern of five pairs of iron wires and four pairs of brass wires of equal length has a net CTE of 5· 12 - 4 · 20 = -20 ppm/K. The length of the gridiron section can thus be chosen to be about 3/8 of the length of the entire pendulum (the remainder being iron), so that the temperature coefficient of the clock rate can be tuned to near zero. Harrison suspended the pendulum bob from its center of mass, so that its expansion did not affect the period to leading order. (It isn't that easy for air circulation to produce large gradients in a solid metal disc weighing a pound or more, so this is pretty good.)

Figure 20.2.3 shows a so-called *banjo pendulum*. It has an iron/zinc gridiron section, but suspends the bob from the bottom, rather than from its center of mass. This is easier to build but has a non-obvious problem.

Sufficiently slow temperature changes can in principle be compensated as nearly as you like, but here's the non-obvious problem: structures like this have *horrible* thermal transient response. Thermal expansion of the bob and the extra shaft length (from the bob's center to its bottom) affect the period of the pendulum along with the wires. Because the wires have much more surface area per unit mass than the bob, they respond to ambient temperature much faster. For instance, a rapid temperature increase makes the pendulum run slow for awhile, because the bob's center of mass is too far from the pivot until its expansion catches up. Thus sudden changes in ambient temperature cause frequency transients that damp out very slowly. Athermalization techniques relying on modern low-CTE materials rather than cancellation have much better transient response as well as much reduced bending due to gradients.[†]

The gridiron principle can be applied to lens mounts too; by making the retaining ring deeper, we can put the mounting surface at a different axial position from the threads. (The mount might wind up looking like a shot glass with threads on the rim and a lens in the bottom.) Using different materials for the tube and the ring, we can in principle give the air gap between the bearing surfaces any CTE we like—positive, negative, or nearly zero. The ring can be thinner-walled than the tube could be, because it doesn't have to handle the main structural tasks.

20.2.4 Thermal Gradients and Bending

As we've seen, thermal transients are a good deal less benign than uniform thermal expansion. Similarly, spatial gradients in temperature can do some surprising things. Acton[‡] gives the example of a 1-mile-long railway track ($L = 5280$ ft), fixed very firmly at both ends. During the night, some practical joker comes and welds in an extra 1 foot of track ($\varepsilon \cdot L = 1$ ft). The (extremely stiff) track bends upwards 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.)[§] 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

[†] See Example 20.3.

[‡] Forman Acton, *Numerical Methods That Work*

[§] It's *44.5 feet*—the height is $h \approx L\sqrt{3\varepsilon/8}$.

$$R = \left[CTE \frac{dT}{dz} \right]^{-1} \tag{20.2}$$

independent of its length. A rod of length L held at one end will warp away by $L^2/2R$, or

$$\Delta Z = \frac{L^2 CTE}{2} \frac{dT}{dz}. \tag{20.3}$$

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

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 15 W, mostly via conduction through the stage. From Table 20.5, we find that $\alpha = 15$ W/m/K and the CTE is $9.6 \cdot 10^{-6}$/K. The temperature gradient in the steel is

$$\frac{dT}{dz} = \frac{P}{\alpha A} \tag{20.4}$$

or 1250 K/m, a temperature drop of 5° C over 4 mm thickness. The rails will want to bend into circles of radius 83 m, a runout of about 60 μm in 10 cm (15 μm arch height), which may well be 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 due to that 2.5° C would produce an arch height of 10 cm $(2.5^\circ \cdot 9.6 \cdot 10^{-6}/\text{K} \cdot 3/8)^{1/2}$ or 30 μm, 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.‡ 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 than steel. (Yes, the wisdom here is: Don't use your translation stage as a heat sink.)

Aside: Teflon

Teflon (polytetrafluoroethylene) is amazing stuff in some ways, but dimensional stability is not one of them. Teflon is probably the hot dog champion—highly variable CTE (reaching 2900 ppm/K) and $\alpha = 0.25$ W/m/K, making it almost 20,000 times worse than stainless steel for a constant heat flux. Fortunately, it's seldom used as a framing material, and since it's mechanically weak it's easy to squash any nonsense it might try to pull.

†This is in the spherical-cow simulation universe, of course—in real life, once thermal contact between the sides of the rails and Al block was lost the problem would change fairly radically.

‡The author heard this piece of advice from the late Erwin Loewen, a master of spectrographs and ruling engines, whom we met in Section 19.12.18.

20.2.5 Temperature and Young's Modulus

Fused quartz and Invar-type steels have very low CTEs, but that isn't the whole story. In optomechanical systems you sometimes care very much about the temperature dependence of the elastic properties, and the thermoelastic coefficients of low CTE materials aren't particularly small.

Consider an object whose center of mass is suspended from a very stable support by a long, fine quartz fiber of length ℓ *in vacuo*. Used as a gravity pendulum with constant amplitude, its period goes as $\sqrt{\ell}$ to leading order, and so is stable with temperature to the 10^{-7}/K level. As a torsion balance, its period and sensitivity go as $1/\sqrt{E}$. The tempco of E near room temperature is $+125$ ppm/° C, about 1000 times worse than the gravity case.[†]

On the other hand, if the suspension is made of Alloy 42,[‡] the temperature coefficient of E can be made very very low, so that spring constants and resonant frequencies are much better controlled. This is a win in many situations—see Section 15.14.

20.3 Heat Flow

20.3.1 Heat Conduction in Solids

In the absence of mass motion, heat transfer follows Fourier's law of heat conduction: the heat power $d\dot{Q}$ flowing out of a surface element dA in unit time is

$$d\dot{Q} = -\alpha \hat{\mathbf{n}} \cdot \nabla T \, dA, \tag{20.5}$$

where $\hat{\mathbf{n}}$ is the outward-directed unit normal vector and α is the the thermal conductivity. Using the divergence theorem and assuming that the material is linear, homogeneous and isotropic, this becomes the heat equation,

$$\frac{\partial T}{\partial t} = \kappa \nabla^2 T. \tag{20.6}$$

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

$$\kappa = \frac{\alpha}{\rho c_P}. \tag{20.7}$$

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

$$d\dot{Q}/dA = -\alpha \frac{\partial T}{\partial z}. \tag{20.8}$$

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.

[†] H. D. H. Drane, "Elastic constants of fused quartz. Change of young's modulus with temperature", *Proc. R. Soc. Lond.* **A122** pp. 274–282 (1929). (The slope of $\sqrt{x}$ at $x = 1$ is 0.5, so the ratio of the sensitivities is the same irrespective of the square root dependence.)

[‡] M. E. Fine and W. C. Ellis, "Young's modulus and its temperature dependence in 36 to 52 pct nickel-iron alloys", *JOM* **2**:9, pp. 1120–1125 (1950)

[§] George Herold points out that near room temperature, all fully dense solids have roughly the same heat capacity per unit volume–about 3 J/K/cm^3, within about $\pm$ 50%. Doesn't apply to styrofoam!

For temperature control, we also need the frequency response of heat transfer (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

$$T(z,t) = e^{j\omega t} \exp\left(\frac{-1-j}{\sqrt{2}}\sqrt{\frac{\omega}{\kappa}}z\right). \tag{20.9}$$

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

$$T(z,t) = e^{j\omega t}\frac{\cosh\left(\beta\sqrt{\omega}\,(z-d)\right)}{\cosh\left(\beta\sqrt{\omega}d\right)}, \tag{20.10}$$

where β is given by

$$\beta = \frac{-1-i}{\sqrt{2\kappa}}. \tag{20.11}$$

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

$$\dot{Q}(t) = \alpha\beta A\sqrt{\omega}e^{j\omega t}\dot{Q}(t)|z = 0 = \alpha\beta A\sqrt{\omega}e^{j\omega t}\tanh\left(\beta\sqrt{\omega}d\right), \tag{20.12}$$

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:

$$H(\omega|z) = \frac{T(z,\omega)}{\dot{Q}e^{j\omega t}} = \frac{1}{\alpha\beta\sqrt{\omega}A}e^{\beta\sqrt{\omega}z} \tag{20.13}$$

$$H(\omega|z,d) = \frac{T(z,\omega|d)}{\dot{Q}e^{j\omega t}} = \frac{\cosh\left(\beta\sqrt{\omega}(z-d)\right)}{\alpha\beta\sqrt{\omega}A\sinh\left(\beta\sqrt{\omega}d\right)}. \tag{20.14}$$

20.3.2 'Thermal Mass'

It's not necessarily that easy to visualize the solutions of (20.14), though we'll do some plots and discuss its general properties in a minute, and of course your favorite math program can do as many more plots as you want. For sufficiently slowly varying problems involving partially insulated systems, we can often use the *thermal mass approximation*, in which we treat the temperatures and heat flows like lumped circuit elements in electronics.† Thermal mass $m_{\text{th}} = mc_P$ represents the energy required to increase the (well insulated) object's temperature by 1°. Assuming that thermal diffusion is fast enough to ignore, the rate of temperature increase is given by

$$\frac{dT}{dt} = \frac{\dot{Q}}{m_{\text{th}}}, \tag{20.15}$$

which agrees with the low frequency limit of (20.14). 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 ambient is θ, the low frequency response goes as $1/(1+i2\pi f\theta m_{\text{th}})$, 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.

†The analogy can be pressed somewhat further, as we'll see in Section 20.9.6.

20.3.3 3D Heat Conduction

In 3D situations and small geometries, such as near a focused laser spot on a surface, heat conduction is dramatically faster and stronger.

Heat dissipated in a 1-μm diameter hemisphere drops half its ΔT across the next 1 μm of material, which makes quite an impressive gradient. Thus 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 μm) is twice as large as 1/(2 μm), 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 if you dump a constant power P uniformly into a disc of radius a on a half-space of material with thermal conductivity α, the temperature rise is about

$$\Delta T = P/(\pi \alpha a). \tag{20.16}$$

20.3.4 Thermal Properties of Materials

At the end of this chapter there are tables of thermal properties of a selection of useful materials for building instruments.

EXAMPLE 20.4 How fast is heat conduction?

Let's plug a few numbers into (20.9). If the material is a thick piece of 1100-T0 aluminum, with $\kappa = 9.9 \cdot 10^{-5}\mathrm{m}^2/\mathrm{s}$, a 1-Hz excitation will decrease by $1/e$ in the first 5.6 mm, and will be phase shifted by 1 rad in the process. In 304 stainless ($\kappa = 0.4 \cdot 10^{-5}$), 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.

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 8 mm 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 $z = 1.5$mm, and clamped on top at $z = 8$ mm, as shown in Figure 20.2. 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 $1/f$ to $1/\sqrt{f}$, because the mass of material being heated at that rate starts declining as $1/\sqrt{f}$, which is fine. 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 1.5 mm 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.

Aside: Thermal Conductivity Integral

All of these thermal characteristics are nonlinear, *i.e.* they depend on the temperature. Using room-temperature numbers to calculate thermal conductance in a cryogenic system will give a huge overestimate for brass and a huge underestimate for copper. It's especially

[†] Area is πr^2 on a plane and $2\pi r^2$ on a hemisphere.

[‡] The math is very similar to the derivation of (20.10) or frustrated TIR.

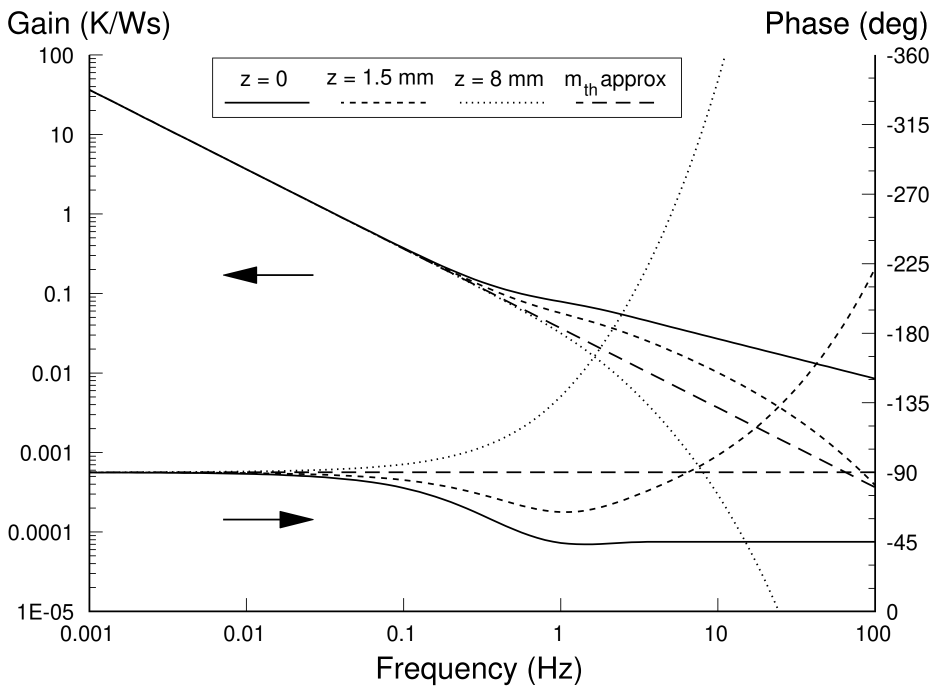

Figure 20.2 Gain and phase transfer function of thermal diffusion in an 8mm thick plate of 6061 aluminum, for sensors mounted at z=1.5 mm and z=8 mm, compared to the thermal mass approximation.

difficult with mounts and other parts that span a large temperature differential, because the thermal conductivity changes along the length. Lakeshore Cryogenics tabulates the *thermal conductivity integral* θ for various materials, which allows you to calculate heat flow in a uniform cross section as

$$\dot{Q} = \frac{A}{L}(\theta_2 - \theta_1). \tag{20.17}$$

20.3.5 Thermal Interfaces

Heat sink grease is actually a pretty 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 thermal 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. The resulting thermal expansion will lead to residual stress on cooling if the two sides of the interface have mismatched CTEs. 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. It's also soft, which helps relieve stress. Indium corrodes readily, though, so don't use it where it might get damp. (The author likes InSn eutectic solder for attaching small TECs.)

If you use paste, large thermal joints will have a lot of small flux inclusions, which are pretty good insulators. Thus soldered thermal joints are best made by tinning both sides and then using at most a tiny amount of flux when joining the two with a quick sliding motion. Because the flux is localized at the interface and the liquid metal's surface tension is very large, the flux residue will get expelled pretty thoroughly if you do it that way.

Silver thermal 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 20 depending on formulation, application, and curing. 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 good silver epoxy. *Note:* 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.

You also have to buy the right material. Electrically conductive silver epoxy is *not* the same stuff at all—it's filled with silver-plated glass spheres instead of bulk quantities of silver metal flakes. (The really cheap stuff uses carbon-coated spheres.) The very best silver epoxy, such as Diemat DM6030HK or Epo-Tek H20E, has enough metal in it that it conducts electricity as well as heat, but most silver thermal epoxies don't.

Silver is good 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 particles. (A similar effect stabilizes the inductance of coils wound on gapped cores; see Section 14.3.8.) 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 order of preference for readily available thermal interface materials (best to worst) goes: No interface, tin-based solder, indium solder, indium-gallium slurry, good silver epoxy, silver- or AlN-loaded grease, good elastomer pads, heat sink epoxy, heat sink grease, mineral oil, cyanoacrylate glue, poor elastomer pads (*e.g.* Silpads), any other glue, air, vacuum.

Aside: Bond Line Control

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 with oil-based abrasive, use a polar solvent such as methanol. 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.

Of course, if you're using epoxy rather than grease, ultra-thin bond lines make the joint much more likely to delaminate due to CTE mismatch or thermal gradients. A very common way to fix that problem is to load up the glue with about 1% by weight of solid glass spheres of the appropriate diameter. When the surfaces are clamped together for gluing, the spheres wedge them apart by just the calculated amount, without affecting the mechanical or thermal properties of the bond otherwise.

20.3.6 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 joints made with these materials as a function of the bond line thickness, you get the slope you'd expect from Table 20.5, but with a significant positive offset, seldom less than 30 $\mu\text{Km}^2/\text{W}$ and as much as 100 $\mu\text{Km}^2/\text{W}$ (0.05 to 0.2 in^2 $^\circ\text{C}/\text{W}$ for readers of American heat sink datasheets).

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 solid-solid interfaces, the theory works much less well, *e.g.* predicting a $1/T^3$ temperature dependence when the measured values tend to be constant above about 150 K and to go roughly as T below that. Other contributions are poor wetting of the surface by the TIM, 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.

20.3.7 Dry *vs.* Greased Interfaces

The microscopic contact area between two flat solid surfaces is approximately equal to the normal force divided by the yield strength of the weaker 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 the best thermal compounds have 20 times the thermal resistance of aluminum, so thin means really thin—preferably under 10 μm for joints that won't see much cycling.

20.3.8 Greased Joint Problems

Greased joints that undergo repetitive thermal cycling will often develop voids in the grease, because the thermal cycling pumps the grease out of the joint. It's a bit analogous to the way freezing water makes frost heaves in the soil—mechanical confinement makes the icy mud push up very hard when it freezes, but there's only gravity to push it down again. Heat sink grease behaves similarly; it's incompressible, so heating pushes it out of the confined joint volume, but once it's in the open and cools, there's only surface tension trying to pull it back in, which doesn't accomplish much since it's so thick and the joint is generally under a preload force. Air bubbles thus form and grow gradually 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. It's all a bit of a black art.†

20.3.9 Thermal Gap Pads

These days you aren't stuck with just greased *vs.* dry joints. Elastomeric thermal gap pads are available in a wide range of consistencies from hard rubber to nearly liquid, with thermal conductivities up to 8 W/m/K. They're a great way to cool a board with a lot of power dissipation in surface mount components, especially if it has parts on both sides. You can get pads up to several millimeters thick, so even fairly tall components can embed themselves in the pad. Some have low enough outgassing that you can use them in high vacuum (not UHV of course).

20.3.10 Heat Conduction in Gases

For gas layers thicker than a few times the mean free path, the thermal conductivity of gases is independent of pressure; Table 20.9.5 shows the thermal conductivity of common gases. At low pressures ($\lesssim$ 1 torr) the thermal conductivity starts to drop off towards zero. 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. (Heavier molecules transfer the same energy per collision, but their slower motion means they have fewer collisions per second.)

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.

Molecules move at the same average speed as atoms with the same mass. The reason they have higher thermal conductivity is that they have more degrees of freedom. At temperatures where xenon is in its ground state, it just has kinetic energy, whereas molecules have all sorts of rotational and vibrational resonances way below 25 meV (kT at room temperature), each of which contributes its $kT/2$ on average. All of that energy has the chance to be transferred in each collision.

20.3.11 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 shears the boundary layer, bringing cool air near the sink, and thus sharpening the gradient and improving the cooling.

Where air can flow easily, *e.g.* between vertically oriented heat sink fins, natural convection is roughly linear for objects near room temperature. The thermal resistance decreases 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.

†The author was fortunate to spend some years palling around with a bunch of very smart packaging and silicon folks at IBM Research, which was most educational.

20.3.12 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

$$L_\nu(\nu,T) = \frac{c}{n}\cos\theta e_0(\nu,T)$$

$$= \frac{2hn^2\nu^3\cos\theta}{c^2\left[\exp\left(\frac{h\nu}{k_BT}\right)-1\right]} \tag{20.18}$$

contains a cosine obliquity factor—otherwise it'd be 2π. A surface of area A and emissivity ε at temperature T_1, in an environment in equilibrium at temperature T_2 will emit a net heat flux density of

$$\dot{Q}_{\text{net}} \approx A\sigma\varepsilon\left(T_1^4 - T_2^4\right), \tag{20.19}$$

This approximation breaks down if the area A is not convex (think of a teakettle, where the inside has lots of area but communicates only through the small spout), or if the temperature of the radiation is not T_2, for instance if the absorbed power is a large fraction of the total emission of the walls, or if a large fraction of A's radiation is reflected back to A, *e.g.* if A were a filament at the center 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.)

EXAMPLE 20.5 Air Insulation

The Footprints sensor of Section 3.10.17 achieves its high sensitivity by insulating the pyroelectric pixels. These are made of carbon ink screen-printed onto a 9-μm thick PVDF film which is freely suspended in air about 5 mm away from the circuit board. Because the radiative coupling to the scene temperature via the optical system is weak, the pixels are always near ambient temperature. At 300 K, still air has a thermal conductivity of about 0.026 W/m/K, so a 5-mm air layer has a thermal conductance per unit area of 5.2 W/m^2/K. Is that more or less than the radiative coupling?

The pixels have an IR emissivity of about 0.5, mostly due to their thinness. The circuit board's will be close to 1.0 because of the solder mask and low percentage of top-level copper. Differentiating (20.19) with respect to T_1, we get

$$G_{\text{rad}} = d\dot{Q}_{\text{net}}/dT_1 \approx 4A\sigma\varepsilon T_1^3, \tag{20.20}$$

which at 300 K and $\varepsilon = 0.5$ is 3.1 W/m^2/K, which is of the same order as the air conductance. A metal foil radiation shield would improve the pixel insulation some, and so increase the low-frequency response. It would also give the transmitted thermal light another chance to be absorbed, further improving sensitivity by probably 10% or so at all frequencies.

20.3.13 Getting Uniform Air Temperature

As you can see from Table 20.9.5, 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 to 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 by stirring. This works even with very simple apparatus: a cardboard box plenum with a fan has been measured to exhibit temperature nonuniformities less than 3 mK near room temperature[†].

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 air, immobilized in an IR-opaque matrix: they work by preventing both convective and radiative heat transport, so 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. Foam can also outgas plasticizer and blowing agent, which might possibly condense as well.[‡]

20.4.1 Insulation and Thermal Radiation

The other benefit of insulation is in suppressing thermal radiation. Thick layers consisting of strong optical scatterers impede thermal radiation by diffusion, just as they do with visible light. In addition, common insulation materials are highly absorbing in the thermal IR.

Infrared-absorbing surfaces re-emit in a Lambertian pattern, which works just the same way as diffusion. As we saw in Example 20.5, this is important for layers thicker than a few millimeters, where radiation may well dominate otherwise, so that 1/2" insulation in your walls would be about as good as 6".

Aside: Superinsulation
Vacuum systems and spacecraft sometimes use *superinsulation*, consisting of a series of *N* layers of low-emissivity material (often gold-coated polyimide sheets) with some low-density web or mesh spacers in between.[§]

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

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

[‡]NASA maintains a public database of the outgassing properties of materials and components that they've tested. It moves around some, so a bit of searching is probably required.

[§]Sometimes it's just called *multilayer insulation*, but where's the fun in that?

into N layers ought to improve the thermal resistance by N times because the ΔT should be constant layer to layer, so that the net conductance should go down by a factor N.†

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.

20.4.2 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 low strength, inflammability, and solubility in polar solvents such as gasoline. Burning styrofoam emits toxic fumes, including styrene monomer, but if you don't set it on fire or dump fuel on it, styrofoam is 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.

20.4.3 Dewars

The insulating quality of a decent vacuum is the basis for Dewar flasks, best known as Thermos bottles. Small dewars are made of low-expansion 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 the long, contorted thermal path minimize heat conduction.

20.4.4 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 eventually destroy components, connectors, and even board traces galvanically. All instruments must be condensation-free, and that doesn't happen by accident. A box that stays dry in New York may drip in Scotland and grow fungus in Brazil or 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 vapor to the equilibrium vapor pressure (*i.e.* the point at which fog just begins to form). Humid air in equilibrium with liquid or ice is said to be *saturated*. The vapor pressure of water increases steeply with temperature, so the relative humidity does too.‡

At 30° C, air can hold three times as much water as 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 to keep things dry. If you rely on external insulation to

†The gold coatings on the first and last layers help improve the steady-state thermal isolation, but do the ones in the middle? Why or why not? (The radiation temperature in the spaces is the key.)

‡There's a bit of a cottage industry in producing empirical fits to the measured pressures, depending on whether it's over water or over ice, and so forth. There are lots of tables available online.

prevent condensation, make sure it forms a hermetic seal, or it will eventually become wet. (The water will be trapped for long periods, which is especially bad.) Usually 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 in from outdoors during the winter. You can keep track of the dew point using a little SMT temperature/humidity sensor such as the Sensirion SHTC3 plus a lookup table, which will let you know when heating is necessary. It's a pretty nice part: ±0.2 K, ±2% RH, 2x2 mm DFN package, $1 in quantity.

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 (in a pinch) by putting the disassembled can in a chest-style domestic freezer until it equilibrates. Open the freezer a crack, reach your arms in, and put the cover on the can before taking it out.

The author and his colleagues have had good success with home-made desiccators made from Tyvek/clear plastic bags sold for use in steam autoclaves, loaded with a mixture of 95% Linde 4A molecular sieve and 5% blue indicator silica gel. The desiccants are pretty well dust-free anyway, but the bags keep in any residual dust and prevent the grains from rolling around inside the instrument. Linde 4A will absorb its own mass in water before saturating, and its pores are too small for it to pump oil or other large molecules. You just toss a bag into the can, seal it up, and you're golden.[†]

The T/H sensor will also give you advance warning that your desiccant needs changing, which is a big help—regular maintenance is a lot more restful than random emergencies.

Do beware of dust; it's a poor trade to replace condensation with powdered desiccant. The autoclave bags are pretty good, or you can use a trick of one of the author's friends, who makes dust-free desiccant cakes by mixing zeolite with two-part silicone rubber; the water diffuses readily through the silicone and gets stuck in the zeolite.

Note that a -30° C cold finger can be a much stronger water vapor sink than room temperature desiccant, so some water will eventually wind up there. (Warm silica gel is especially poor.) Permanently sealed vacuum systems usually control residual water using a *getter*, *i.e.* a flashed film of highly reactive barium metal.[‡] If lasers are involved, it's usually best to use dry air, as this helps to control organic contamination (see Section 2.12.14).

In infrared systems, where radiation from 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–it's amazingly hard to prevent turbulence from entraining *some* room air.

Aside: Ingress Protection and You

If you're building instruments for a living, sooner or later somebody will want one that will work underwater, or survive years out in the elements, or accept washing with industrial disinfectant once a shift. That's where you have to start really worrying about *ingress*

[†]You can get all that stuff from Amazon and elsewhere, so there's no problem finding it—the point of going easy on the blue stuff is that you don't need much to see the color, and the indicator is cobalt chloride, not a super pleasant compound.

[‡]Getters are produced by putting a pellet of barium aluminum alloy in the chamber, on a tungsten wire heater. It's stable in room air and survives bakeout fine, but when you turn on the heater the barium sublimes and coats nearby surfaces.

protection. There's a fair amount of lore out there about it, including tables of what IP-soandso protects against and whatnot, so here we'll just say two main things: that plastic boxes and elastomer seals are not actually waterproof, because there is gas exchange via diffusion, and that those IP-soandsos are based on very brief tests with subjective pass/fail criteria. (You can read all about them in IEC Standard 60529.)

Pressure differences also drive gas exchange; in any given location the air pressure changes by about $\pm 5\%$ on time scales of a few days, and solar heating and cooling can do the same in a few hours. Thus the seals are constantly being worked, and leak a bit in the process. If they are wet, then instead of moist air coming in it'll be liquid water, which is $2000\times$ worse. In cases like that it's usually necessary to provide a strategically placed air vent, protected from liquid and fitted with a hydrophobic gas exchange membrane, and rely on the desiccant in the box to absorb the vapor that gets exchanged. (Submersible systems are of course in a different category, but folks in that business are used to doing lots of maintenance.)

20.5 Temperature Sensors

To control temperature, we need a good way of measuring it. An ideal temperature sensor senses only temperature, does it fast, has no thermal mass, and contributes no heat 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 *their own* 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. Calibration can help a lot, provided the dissipation is kept really constant. (Lots of times there's other dissipation on the cold plate anyway.)

Thermal shorts Controlling the thermal conduction along lead wires is tough, because metal conducts heat well and often many wires are needed. Air temperature sensors almost always measure the temperature of their lead wires, not that of the air.[†] Small SMT devices measure the temperature of the board instead, which can be good or bad. (See Section 15.5.)

Stress effects All temperature sensors can be made to drift by mechanical stress. Since they aren't normally made from the same materials as the cold plate, this can be a serious source of error and drift. Solder the sensors to the cold plate by one lead, and use very small wire wire or flex circuit board for the other lead. Make sure the wire is thermally grounded to avoid thermocouple offsets.

[†]See *e.g.* Section 4 of the National Semiconductor Temperature Measurement Handbook, https://web.archive.org/web/20070324135933/http://www.national.com/appinfo/tempsensors/files/temphb.pdf.

Thermocouple offsets Every junction between dissimilar conductors will cause a thermocouple offset, but as long as they're balanced it will cancel out. A sensor connected by copper wires contains at least two thermocouple junctions in series with its sensing element. Glass- and metal-packaged devices are the worst for this, because their Kovar leads have an enormous thermocouple coefficient against copper. For thermocouple sensors, you have to use the thermocouple alloy wires all the way to the meter, where offsets and drift due to circuitry and unwanted junctions still limit them to low-accuracy jobs.

20.5.1 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 LM335, which reads directly in kelvins with a ± 2 K initial accuracy and costs about $0.50 in volume. They are stable and linear, so that $\partial V_{\text{out}}/\partial T$ is constant. Thus they don't contribute to loop nonlinearity, which makes frequency compensation easier if your setpoint can vary a lot. (It makes very little difference in fixed-temperature applications, of course, since it's only the gain near the setpoint that matters very much.) 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 ($\pm 2^\circ$ C typically). For not much more money, you can get digital sensors, which have built-in ADCs and calibration tables. The $1 Sensirion SHTC3 that we met in Section 20.4.4 has 0.2 C initial accuracy, 16-bit readout with CRC, and comes with a quite reasonable humidity sensor built in.

On the other hand, IC sensors are not as stable as platinum RTDs or good-quality glass bead thermistors, are sensitive to mechanical stress, have a narrower operating range, typically -40° to 100° C, and dissipate some heat (10 μW 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 or through the case. With care, you can make small SMT ones faster by connecting their substrate leads to a copper pour, *e.g.* using an exposed-pad or SOT-23 package (see Section 15.5).

20.5.2 Thermistors

A thermistor is a type of mixed metal oxide resistor whose resistance value is a strong function of temperature, typically changing -3% to -4%/K. They come in a wide variety of styles and prices, from a few cents for an 0603 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 (1 s). These can be embedded in your cold plate with a thin layer of heat sink epoxy, and survive well, although their values may change slightly with mechanical stress. They come in resistor-linearized versions†, and in interchangeable types whose $R(T)$ curves are identical within $\pm 0.2^\circ$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 worthwhile for wider temperature

†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. Metal oxide thermistors are like this, but platinum RTDs are concave downwards, so they need negative shunt resistances. This requires active devices.

range applications since thermistors are so nonlinear that the temperature resolution at the upper limit will be poor otherwise. (You just have to pick one resistor, after all.)

You can get inexpensive thermistors in DO35 packages (like 1N4148 diodes), which have slugs of copper connecting to the faces of a thin oxide element, and are therefore fast if you use them right. These are pretty useful in protos: soldering one (short) lead to the thing you want to measure, and using really skinny conductors to connect to the other end, is the ticket. (Later on we'll use 0603 thermistors in the same sort of way.)

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 cold plate temperature since the thermistor is nonlinear anyway). Run it from a stable reference voltage.

One possible gotcha is that thermistors are slightly magnetic.

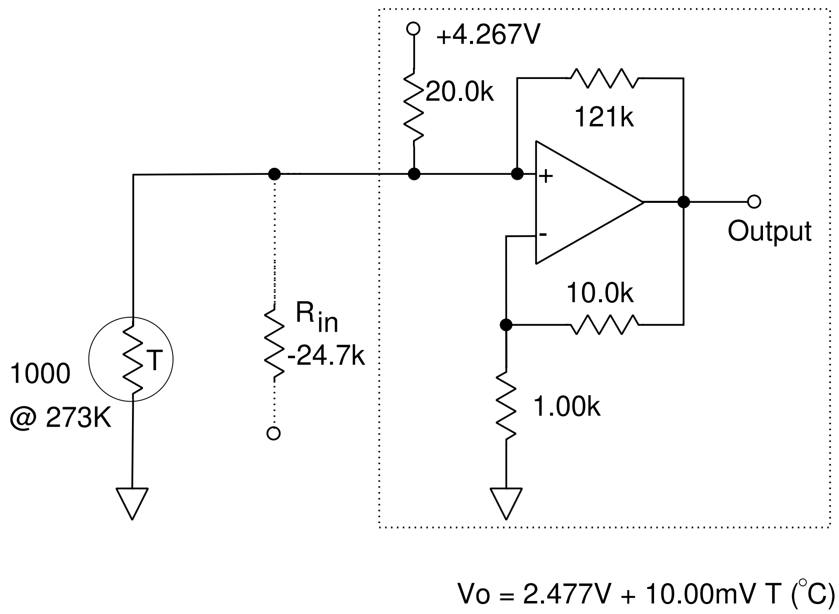

Figure 20.3 Negative resistance linearizer for platinum RTD; V *vs.* T is now linear to ± 0.2° from -150° C to +500° C.

20.5.3 Platinum RTDs

Platinum RTDs (resistance temperature detectors) are also thermistors, at least from the behaviour point of view. They come in wire-wound and thin film types, from 100Ω to 1kΩ. Wire-wound RTDs are still expensive, but the thin film ones are now much cheaper—some 0.1% units are as low as $1 in quantity. 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 an order of magnitude smaller than thermistors', about +0.35%/K. They work over a wide range of temperatures, but cannot be linearized as simply as ther-

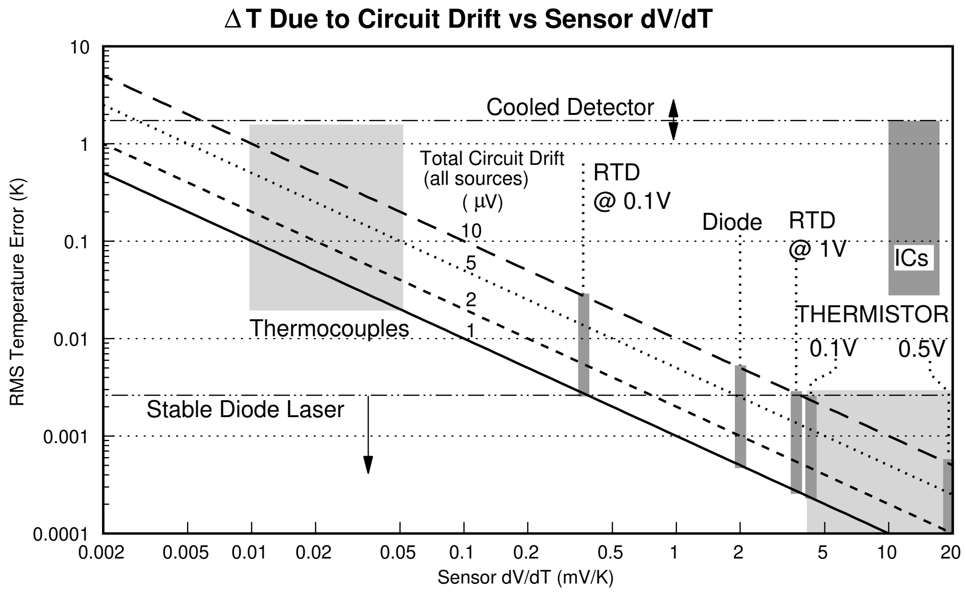

Figure 20.4 Attainable temperature stability *vs.* temperature sensor slope, for total circuit drifts from 1 μV (extremely hard to attain) to 10 μV (merely very difficult). Tight temperature control requires fast sensors with very high sensitivity—thermocouples don't make the grade except for the lowest-performance applications. Note that the absolute accuracy is nowhere near this good, but we usually don't care very much as long as it's repeatable.

mistors since $R(T)$ is concave downward. On the other hand, if you use a simple positive feedback circuit (*e.g.* the one in Section 20.5.3) 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 $\pm 0.2°$ C from $-150°$ C to $+500°$ C, and better over narrower ranges, as shown in Figure 20.5.2. The 10 mV/K output slope of the circuit makes it easy to make human-readable thermometers.

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 1 kΩ RTD's voltage sensitivity is 30 dB worse than a 10 kΩ oxide 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. (The copper pour approach of Section 20.9.4 is even better.)

The main remaining issue with RTDs is that they have a small amount of hysteresis due to thermal cycling. If you run near room temperature, this really isn't a worry.[†]

[†]K. S. Gam *et al.*, "Thermal Hysteresis in Thin-Film Platinum Resistance Thermometers", *Int J Thermophys* **32**, pp. 2388–2396 (2011)

20.5.4 Thermocouples

Thermocouples generate a voltage related to the temperature difference between two junctions of dissimilar metal wires, and are a pain in the neck.

For one thing, their sensitivity is very low, and for another, what they actually measure is some function of two junction temperatures, not one. To make a single-point measurement, you need 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 15.11.6. 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 catalog.

20.5.5 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/K, although here as so often, diode-connected 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 the photocurrent is less than 20 μA or so, you can forward bias it by 1 or 2 milliamps, so that the 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 response times way under 1 s for 9-mm diode packages, indistinguishable from the response of a 75 μm wire thermocouple silver-epoxied to the header. You can even use a simple chopping circuit to sense both the monitor photocurrent and the temperature, either by forward biasing or (for higher photocurrents) by alternately measuring the open circuit voltage and the short circuit current; modeling the photodiode as a current source in parallel with a forward-biased diode allows calculating both laser power and temperature.[‡]

[†] http://ww1.microchip.com/downloads/en/AppNotes/00001838A.pdf

[‡] Some adventurous folk use the laser itself as the temperature-sensing diode, but you have to be quick, because 3D heat conduction makes its junction cool extremely rapidly when power is removed. (See Section 20.3.3).

20.5.6 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, but you do have to worry about their transient response—a window could easily ice up in the time it takes a largish ice cell to start freezing. Now that good T/H sensors are cheap and easy to use, it makes much more sense to use a lookup table instead.

20.5.7 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; a switching regulator fails and blows up the temperature controller; 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 lots of devices available for this. The two oldest are thermal switches, where the motion of a bimetallic strip or snap disc opens the circuit when the temperature gets too high, and thermal fuses, where a fusible-alloy wire melts to open the circuit. Thermal switches are resettable, either manually or automatically upon cooling, but blown fuses have to be replaced. (Something mechanical is very comforting to have as a last resort.)

Other candidates are based on positive temperature coefficient (PTC) thermistors, whose resistance increases enormously (25-50%/K) above some transition temperature. PTC devices in series with a load have a snap action similar to thermal runaway in transistors: once the resistance starts increasing, the PTC's dissipation goes up, increasing the resistance further until the PTC comes to rest in a high-temperature, high resistance, low current state. If the requirements are sufficiently loose, this can be a good technique, but for most things the switching is too sloppy.

There are also a whole lot of IC thermostat chips, usually with open-drain outputs to allow them to be wire-ORed to form a single over-temperature alarm signal to an MCU, a solid-state relay, or the enable pin of a switching regulator. As long as the rest of the circuitry is working, these work really well and are cheap insurance.

20.6 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

†Safety standards and electrical codes look odd and arbitrary, but they contain a lot of non-obvious and very expensive wisdom.

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.† Pulse-width modulation will make the dissipation linear in the duty cycle, for much the same reason, but will produce more EMI.

20.6.1 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 and giant arc furnaces for melting steel scrap. The most useful type in instruments is just a thin Nichrome film embedded 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 maintain 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, KOA Speer's RK73B series surface mount resistors make nice flat-topped ceramic heaters of 2010 size, 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. Regular resistors work too, if you mount them upside-down. (In production that would probably require re-reeling.) **Manufacturers:** Minco, Omega, Chromalox, Kyocera, Caddock.

Aside: Thermal Forcing

In equilibrium, the temperature of an object is uniquely determined by its total heat energy, so if there is no net energy input, the temperature will stay perfectly constant.‡ What we're trying to do with a temperature control loop is to eliminate the temperature changes by getting rid of the thermal forcing.

20.6.2 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 25–50%/K 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

†A drawback of this approach is that the dissipation moves around as the heat goes up and down, so the gradients will go all over the place. One approach is to leave out the heater and just use the transistor.

‡Neglecting chemical and nuclear reactions and so forth, of course.

powers available range from a watt to several kilowatts per device. You can't beat it for simplicity and reliability, and since it's a linear 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 vapor lamps warm, that sort of thing.

PTC thermistors have poor settability, because the device switches where it feels like it. Big ones are hard to get leads on, because the thermal shock at turn-on kills soldered leads very quickly by metal fatigue, so use clamps instead. There will be a certain amount of sliding around between the clamped thermistor and its mount, so pay attention to the manufacturer's recommendations for materials and clamp arrangements. (Fretting corrosion can spoil your day.)

20.6.3 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 40 mm 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.

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. Slightly more expensive ones constructed with hard solder will work reliably near 200° C, and stand lots of temperature cycling. Historically, Melcor/Laird uses soft solder, while Marlow, Ferrotec, and Tellurex use hard solder. You won't find this crucial info in the datasheet, so be sure to ask your vendor before specifying.

20.6.4 TECs, Thermal Loads, and Heat Leaks

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 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 improving the heat exchanger and losing the TEC.†

Another big advantage of TECs is that (near ambient) you can get lots of $\dot{Q}$ without much of a heat leak. In a system using heaters alone, the heater output governs the maximum positive slew rate, but the negative slew rate (in the cooling direction) depends only on the heat leak, so you can't insulate it too well or it'll stop working properly. That makes the system inherently far more vulnerable to thermal forcing than a TE-cooled system. Combining the two (as in Section 20.7.7) can get us both fast slewing and good forcing resistance.

20.6.5 Heat Sinking TECs

The ΔT is measured from the hot plate to the cold plate, so a poor heat sink or poor thermal contact on the hot side 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, a heat pipe to a large metal object, or even water cooling.

20.6.6 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 120 N/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, especially if this can be made of foam glass.

Solder has better thermal conductivity than any kind of paste or glue, including silver epoxy. The hot side needs to transfer a lot more heat than the cold side, so if you have to choose, solder the TECs to the heat sink or hot-side spreader plate.

One 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‡ or silver thermal grease (or ordinary ZnO thermal grease if you must) and clamping screws instead. If you have more than one TEC, and are working at any significant ΔT, clamp those ones too, at least on one side. Otherwise the shear stress caused by the cold plate shrinking and the hot plate expanding will crack rigidly mounted TECs.

If you're determined to use thermal grease with TECs at large ΔT, make it extremely thin, 10 μm or less. To achieve really thin grease layers, you have to make sure that both

†Sometimes there's no choice, *e.g.* if your ambient can go above the maximum operating temperature of a diode laser, but fight it if you can.

‡Indalloy paste, from the Indium Corporation of America.

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.

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're liable to go out of focus on you as the heat sink temperature changes.

The screws can be a serious thermal short, especially if they're too large. Stainless steel is very strong, but you can use only so small a screw. A #2-56 stainless steel machine screw (1.9 mm mean diameter) spanning a 1.5 cm gap at $\Delta T = 65^\circ$ 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 Belleville spring washers and don't expect too much of them. More insidiously, if nylon dries out too much, it becomes very brittle. Screws made from other plastics are weirdly expensive—dollars per piece, versus a nickel or so for nylon. A largish plastic clamp held down by steel screws is a possible solution; at large ΔT it usually isn't sufficient just to use plastic washers for insulation.

Aside: Preload Force
When you're frying something on the stove, and you want it to cook faster, what do you do? You push on it with a spatula. The same principle works in instruments: you'll get a lower and more repeatable thermal resistance if you apply a well-controlled preload. With epoxy TIMs, you can probably get away with removing the preload after curing, but leave it on if you can—it helps prevent delamination.

20.6.7 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 use Kapton flex circuits with traces sized so as to equalize the I^2R dissipation and thermal conduction. Take great care in assembling the thermal interfaces (see Section 20.3.5).

The first stage runs nearly unloaded, but you want to get the same ΔT from each of the others 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 3 per stage, lateral drops in the alumina rapidly approach and then exceed $\Delta T_{\max}$, which means that the effective area becomes asymptotically constant. The heat load continues to increase as you go down the stack, so the attainable ΔT 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. (Flat heat pipes can help here.)

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 flex 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.

20.6.8 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 (127 μm diameter) has an electrical resistance of about 30mΩ, can handle about 250 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× coarser wire) and which is much stiffer. Diameters of 75 μm 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.

The best way to get wires in and out is fine-line Kapton flex circuit, preferably with FFC connectors. You can get 100 μm thick Kapton film (α_{sub}=0.12 W/m/K) with 0.25-oz copper on one side, and inexpensive connectors exist for it. Using 6-mil lines and spaces (150 μm), a 100-mm length of 4-mil flex with 20 conductors has a total thermal conductance of

$$G_T = \frac{20 \cdot \alpha_{\text{Cu}}(150\ \mu\text{m})(8.5\ \mu\text{m}) + 41 \cdot \alpha_{\text{sub}}(150\ \mu\text{m})(100\ \mu\text{m})}{0.1\ \text{m}} \approx 103\ \mu\text{W/K}, \quad (20.21)$$

of which the polyimide contributes only 0.7 μW. This is very hard to approach with individual wires. (The author and his colleagues are fond of the Molex 502598xx93 series, which have traces on 0.3 mm pitch, fanned out into two rows of 0.5 mm contacts.) Avoid the ones with individual rolled-wire conductors–they leak far more heat.

20.6.9 Modeling TECs

Over a good part of its range, a TEC is a nice linear actuator; you put in a bit of extra current, and pump a predictable amount of extra heat. The main issue there is the heat leak back through the TEC itself, but that's a forcing term and doesn't affect the loop dynamics much.

When you get near the limits, though, 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; half a volt does something very different when ΔT is 5° *vs.* 50°. Voltage is therefore not a good control parameter—use a variable current source instead. What's more, the amount of cooling you get per amp drops as you approach $\Delta T_{\max}$, which makes the control loop slow down and may even lead to oscillation.

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 usually prevent disaster. (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 datasheet 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 modeling 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 $\Delta T = 0$. From these relations, we can derive the very useful equation

$$\dot{Q}_{\text{cold}} = \dot{Q}_{\max}\left[\frac{-2I}{I_{\max}} + \frac{I^2}{I_{\max}^2} + \frac{\Delta T}{\Delta T_{\max}}\right], \tag{20.22}$$

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,

$$\dot{Q}_{\text{hot}} = VI - \dot{Q}_{\text{cold}} \tag{20.23}$$

and is equal to

$$\dot{Q}_{\text{hot}} = VI + \dot{Q}_{\text{load}} \tag{20.24}$$

in the steady state. The thermocouple coefficient of a single TEC junction is about 430μV/K, 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},

$$T_{\text{cold}} \approx T_A + \theta_{SA}\left[I^2\frac{V_{\max}}{I_{\max}} + I\cdot(430\,\mu V\cdot N\Delta T)\right] - \Delta T_{\max}\cdot\left[\frac{2I}{I_{\max}} - \frac{\dot{Q}_{\text{load}}}{\dot{Q}_{\max}} - \frac{I^2}{I_{\max}^2}\right], \quad (20.25)$$

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.†

Aside: Using TECs as Thermocouples
Because a TEC is basically a bunch of big thermocouples run backwards, you can use its open-circuit voltage to make a low-quality measurement of the ΔT between cold plate and heat sink. This requires disconnecting the drive briefly, of course, which may be inconvenient depending on your drive circuitry, and since there's some temperature drop in the rest of the thermal path, it doesn't quite measure what you care about. However, it does give an accurate way to tell how hard the TEC is actually working, and perhaps to double-check the control parameters so as to avoid I^2R thermal runaway.

20.7 Heat Sinks

20.7.1 Natural convection

What you're buying in a heat sink is a way to remove lots of power without a huge temperature rise, *i.e.* a low sink-to-ambient thermal resistance θ_{SA}, measured in kelvins per watt. 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 $T_{\text{sink}} = 100°$ C in a 20° C ambient. One gotcha is that the 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 at all 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.

20.7.2 Forced air

We won't talk much about forced air cooling except to note that it can improve θ_{SA} by a factor of 5 to 10 if you move air fast enough (3–6 m/s). Fan cooling brings with it all sorts of noise, vibration, dust, and reliability problems that you'll be much happier avoiding it if you can.

If not, 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, as used to cool high-performance processor chips, 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.

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

Fan-cooled equipment tends to die when the heat sinks clog up with dust, so plan for how to deal with that.

There was a bit of a vogue in the 1990s for writing theory papers on fan cooling, but the brutal truth is that it's an empirical business in most cases. Fan cooling causes so many problems that it's always a last resort. That means that few systems are designed for it initially, and so the result is a bodge; maybe we have a vertically mounted board where the transformer and the electrolytic caps are running too hot because we put the MOSFETs' heat sink right underneath them. (Oops.) We can't turn the board upside down because it would mess up the cabling and the front panel, so we cut a couple of holes and put in a fan.†

20.7.3 Water Cooling

Circulating water is about the best way to remove a lot of heat without a big temperature rise and without a lot of acoustic noise. 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. There are lots of things to worry about with water cooling, of course: biofilms; corrosion; frost; pump lifetimes; leaks; reservoirs; temperature, pressure and level sensors; and so on. Don't do it in production unless you really need to—according to the received wisdom, getting it good enough to put in the hands of untrained users is far more complicated than it may appear. On the other hand, for lab use, a $12 submersible aquarium pump sitting at the bottom of a bucket of water and connected with Tygon tubing can be a big help, and you can dump in a bag of party ice if you need to. (Commercial chillers are good, but they're noisy and expensive, they heat up the lab, and you don't get to put them back in the broom closet when you're done with them.)

20.7.4 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. Melting also provides a handy built-in thermometer; the volume changes at constant temperature, so you can use a pressure sensor to control the heating or cooling of the cell. You can stuff 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, $\Delta H_{fusion} =$ 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.* very much 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

†The author has never needed to do this, but it's done all the time.

its inflammability, and it boils at about +68° C) Acetone ($T_{MP} = -95°$ C, $\Delta H_{fusion} = 98$ J/g, T_{BP}=+56° C) is another good possibility. One fairly serious drawback is that since TECs are so weak mechanically, hanging a big mass on the cold plate with no other support will make the whole thing fragile, so a bit of thermomechanical ingenuity is required. (Foam glass insulation and quilted indium foil TIM on the cold plate may be your friends.)

EXAMPLE 20.6 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 more than 100 times worse than that.

Consider an alumina TEC plate, 30 mm square and 0.6 mm thick, with a localized heat load of 20 W 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

$$\delta T \approx \frac{\dot{Q}}{\alpha} \frac{10\text{mm}}{70\text{mm} \times 0.6\text{mm}} \approx 60° \; C, \tag{20.26}$$

which is far too high. The moral of the story is that you have to use big chunks of copper or 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 20.7 Flat Heat Pipe Spreaders

The best way to transport heat any distance is to physically move hot matter around. Forced air and water cooling are the most familiar methods, but for restricted temperature ranges there is another option: *heat pipes*. These use vapor transport in a sealed, partially evacuated cavity: fluid evaporates in the hot regions, recondenses in the cooler ones, depositing its heat of vaporization, and is transported back to the hot area by gravity or capillary action. These can be startlingly better than copper.

Most heat pipes are long and skinny, but you can also get them in flat plates, usually called 'vapor chambers'. Whatever you call them they make great spreaders, *e.g.* for attaching a Peltier cooler to a large heat sink. Cao *et al.*[†] claim thermal conductivities 40× that of copper for a Cu/H_2O flat heat pipe used as a spreader. You might worry that they might have rather strange and non-repeatable transient responses, but then you don't care too much in a closed-loop system, especially on the heat sink side.

You can get heat pipes from Digikey for cheap, so check them out.

[†]Cao, Y., *et al.* "Experiments and Analyses of Flat Miniature Heat Pipes", *J. Thermophys. Heat Transfer*, **11**, 2, pp. 158–164 (1996)

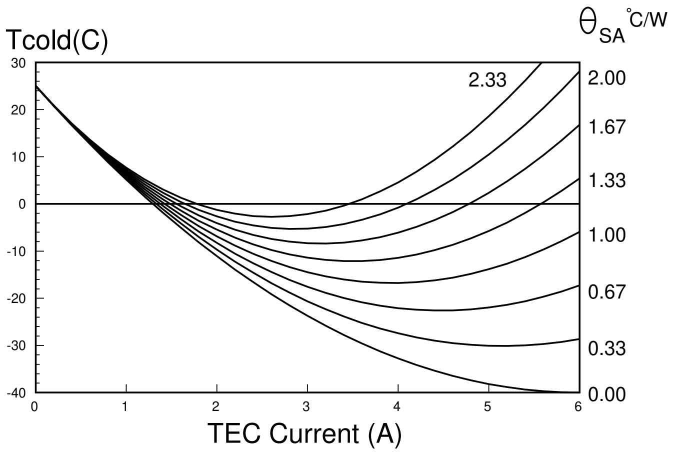

Figure 20.5 Designing a TE cooler system using a passive heat sink: cold plate temperature *vs.* drive current and heat sink thermal resistance θ_{SA}. (T_A=25° C)

EXAMPLE 20.8 TEC on a passive heat sink

Passive heat sinks are poor matches to TECs used at large ΔT, but are useful for applications near room temperature. Consider a 15×30 mm rectangular TEC with I_{max} = 6 A, V_{imax} = 4.5±0.5 V, ΔT_{max} = 65° C, and $\dot{Q}_{max}$ = 14 W.† 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 (20.25) by setting $\dot{Q}_{cold}$ to 0. As we see from Figure 20.5, 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 common 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.

EXAMPLE 20.9 Sizing a TEC

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

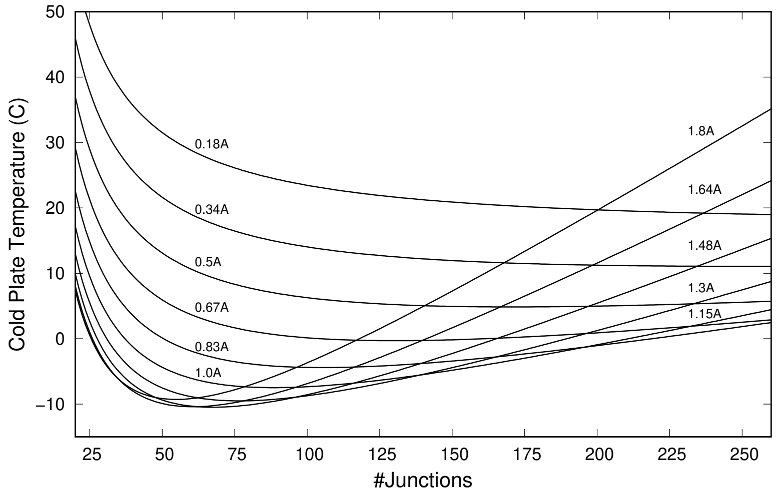

Figure 20.6 Choosing the right TEC for a given $\dot{Q}_{\text{load}}$ and Θ_{SA}. With $\Theta_{SA} = 3.3$ K/W and $\dot{Q}_{\text{load}} = 0.8$ W, the cold plate temperature bottoms out at about -10° C over a fairly broad TEC size range of 50–75 junctions. (T_A=25° C)

Let's come at this the other way round. TECs in a given series have different areas, but the same pillar design. The pillars are wired in series, so all the parameters stay the same except that $\dot{Q}_{\text{max}}$, R, N, and V_{max} scale as the area. We can thus optimize the TEC size for a given cold-plate dissipation $\dot{Q}_{\text{load}}$ and heat sink Θ_{SA}. Say we have a quite decent heat sink with $\Theta_{SA} = 3.3$ K/W and a maximum $\dot{Q}_{\text{load}} = 0.8$ W due to a large MPPC detector array running near maximum output current.

We specialize to the Marlow series represented by the NL1025T-01AC (9 x 11 mm), which has $I_{\text{max}} = 1.8$ A, $\Delta T_{\text{max}} = 67$ K, $\dot{Q}_{\text{max}} = 5.2$ W, $R_{\text{TEC}} = 2.1\,\Omega$, $N = 80$. From (20.25) we can plot the cold plate temperature as a function of N for various currents, as shown in Figure 20.6. The optimum N is 50–75, so the 80-junction NL1025 is the closest.

If no single TEC fits the bill, of course you can use more than one, ideally wired in series and positioned so as to minimize time-dependent thermal gradients.

Aside: Thermoacoustic Refrigerators

Thermoacoustics are an intriguing way to do high power refrigeration with no moving parts. They're less efficient than Stirling engines, but far more durable and cheaper to make.†

20.7.5 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

†Interested readers may wish to look at the publications of Greg Swift, especially his monograph *Thermoacoustics*, available for download from the Acoustical Society of America, and the useful list of references given by Steven Garrett, *Am. J. Phys.* **72** (1), January 2004.

ceramic packages with metal lead frames, and they are very expensive and 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 thin circuit board with lots of via stitching to a ground plane on the back that is bonded directly to the cold plate with silver grease. Relieve the solder mask on the contact area. Vias covering a few percent of the board area is enough copper for most cold-plate purposes, but you can calculate it pretty easily. Ask your board house for the drill and finished hole diameters, and you can calculate how much copper is in there. It's not rocket surgery, just bookkeeping.

If you're adventurous, you may be able to find an indium solder that bonds well to the ceramic package. That can help a lot, and is much easier to remove than silver epoxy.

20.7.6 The TEC Control Problem

Example 20.8 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} and the ambient temperature.

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. Since the math is pretty simple, it's worth having your MCU run a continuous TEC model to back out what $\dot{Q}_{\text{load}}$ is, and maybe even adjusting the loop filter to match. See Section 20.6.9 for more on that.

How Not To Drive A TEC 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.[†]

Open-loop 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.

20.7.7 Controlling TECs

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 actual temperature control job (the heater can be bang-bang

[†]PWM control of a TEC is an insidious source of FM noise and RIN in diode lasers. Even if you filter it, enough will probably get through to cause serious FM problems in high stability applications. Commercial controllers are generally far from quiet enough—remember: 'there's switcher quiet, and then there's instrument quiet.'

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 slow, complicated, and hard to compensate.

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 depend on fancy 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 millihertz bandwidths, we won't even be able to get decent rejection of diurnal temperature changes, which are important even in an air-conditioned building.

By putting a small surface mount resistor on a small silver or copper cold plate, you can get control bandwidths of a hertz, and by using a MELF-package 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.

You don't have to dump a lot of power into the heater, provided that the temperature setpoint doesn't change rapidly and the thermal forcing is smallish. If you need to drop the temperature quickly, 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). Remember that the heater dissipation goes as I^2, so the loop gain will go all over the place unless you put a square-rooter someplace in the loop. (PWMing the heater gets you a linear control law, but is often a no-no due to pickup problems, especially since you can't filter it very much without losing the linearity.)

If the goal is to get T_{cold} as low as possible, just use a big heat sink and run the TEC just to the low-current side of the minimum of the $T(I)$ curve (Figure 20.5). 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.† 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 definitely want to use hard-soldered TECs if you do this.

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

20.7.8 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, lasers, and optical detectors.

20.7.9 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 20.9.5. 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 LN_2 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 awhile 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—see Section 6.10.9.

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 that are easily excited by fluid convection or vibration of the pump or the table. Cryopumps are especially bad for this.

[†]Cables and resistors don't mind being dunked straight into the liquid, but you can destroy detectors with a thermal shock like that. InSb diodes don't grow on trees, so be careful.

20.8 Temperature Controller Design

One important distinction that needs to be made at the outset is between temperature *control* and temperature *stabilization*. Say you're in a standards lab measuring absolute temperature dependence of the bandgap of silicon. In that case, you really care that the sample temperature is exactly what the thermometer says, to the best available accuracy.

Inside an instrument, however, we normally care far more about stabilization. For instance, a nominally single-mode diode laser exhibits islands of stability,† where its AM and FM noise are minimized. Since we have no *a priori* idea where those are, we have to go looking for them. Once we've found a good one, we want to sit there very accurately, come what may. Stabilizing etalons is another example of the same thing—if it stays still, we stay happy.

20.8.1 High-Precision Control

If you look up the older literature on ultra-precise temperature control, you'll find a lot of care and complexity being devoted to minimizing dissipation on the cold plate and especially to minimizing temperature errors caused by thermocouple offsets—AC excitation, transformer coupling, AC amps and phase-sensitive detectors, all that sort of stuff.

Cold-plate dissipation can still be an issue today, though usually not much of one, as we'll see shortly. But all that AC stuff is now blessedly unnecessary. The point of it was that the control electronics and amplifiers took up a lot of space and dissipated a lot of power, so they had to be on the ambient side. This presented a design problem that the AC method solved rather elegantly for the time. (It was first done back in the vacuum-tube days.)

However, it's now much more reasonable to do a bit of creative goal-post-moving, and put the whole sensing circuit and chop amp on the cold plate. A modern chop amp such as an OPA378‡ dissipates way under 1 mW. The sensor's $1/f$ noise comes from conductivity fluctuations, and so is not improved by AC excitation. A Pt1000 RTD has very low $1/f$ noise, with the wirewound ones being better.

Dropping a volt across the PT1000 will dissipate a milliwatt, so the total thermal forcing from the chop amp and sensing circuit is easily maintained at no more than 3 mW or so. A fairly low-grade voltage reference such as an AZ431A (6 cents in reels) has a tempco of ± 20 ppm/K, which will keep the dissipation constant to ± 60 nW/K. (Remember that the leading-order effects will get tracked out by the loop, leaving only the residual gradients from that 60 nW.)

The temperature-sensing circuit has a bridge consisting of the sensor and three resistors, driven by a voltage reference. With the whole bridge on the cold plate, thermocouple offsets go away to any accuracy remotely required, and the temperature measurement is purely ratiometric, so reference drift contributes only a tiny gain error. "Fine,", you say, "but what about the room-temperature circuitry that closes the loop? There are still TC offsets there."

Here's where the goal posts get moved: we use the chop amp as the integrating servo as well. That way, thermocouple offsets, resistance tempcos, and other perturbations in the room-temperature circuitry don't matter anymore because they all get tracked out by the

†Section 2.12.5.

‡It's not the absolute newest or quietest in the flatband, but really exhibits no $1/f$ noise down to the tens of microhertz at least.

loop in the same way as the bridge dissipation. We can make it settable and possibly add some other digital smarts by using a digital pot or multiplying DAC for half of the bridge (still on the cold plate).

This can be taken even further by adding a second (lead-lag) integrator to the loop, on the room-temperature side, so that the chop amp's output stays the same as well. That way there is no opportunity for its dissipation to change due to changes in its internal biasing.

Some systems, such as stabilized diode lasers, will have widely varying power dissipation on the cold plate. In that case some more attention will need to be paid to reducing gradients across the sensing circuit, but this isn't very hard—a bit of creative slot-cutting will usually do it.

20.8.2 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.

One caution is needed: by adding gain, we can increase the loop bandwidth as we like, limited by the aforementioned excess phase shift from diffusion. However, this does zilch to increase the slew rate–that requires using a driver with more stooch. Wideband loops with slow slewing are vulnerable to windup, as we'll see in Example 20.11.

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.

20.8.3 Local Feedback Loops

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 a temperature sensor right at the heater or cooler, even if the heater temperature isn't what we care about.

This is a wonderful opportunity for local feedback (see Section 15.5). Wrapping a proportional-only loop around a TEC and thermistor gives a very well-behaved virtual actuator with probably 10× the speed of the real one, and a nice first-order-style response besides. It also has good resistance to thermal forcing, on account of its high speed. You can combine these and control them far more easily than if the outer loop has to manage all their idiosyncrasies, and the forcing resistance can't be beat.

20.8.4 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.

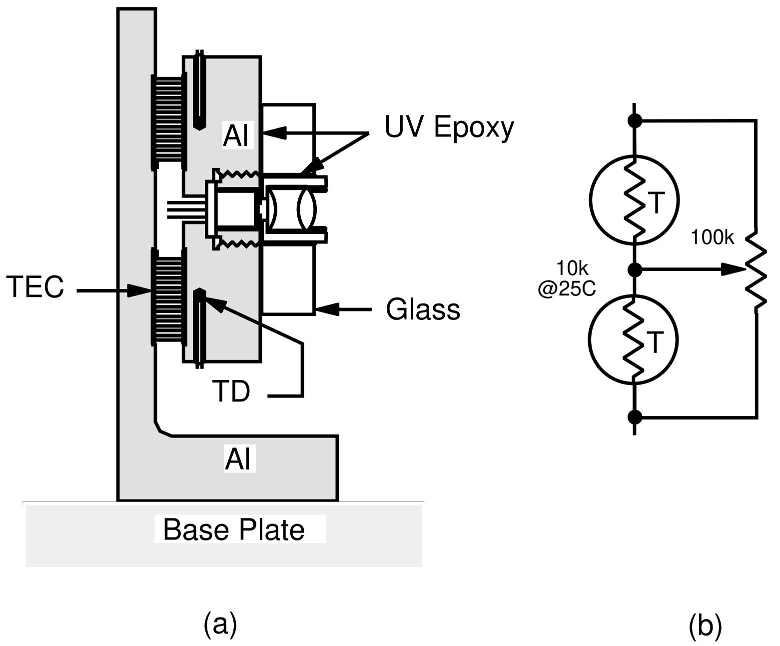

Figure 20.7 Cancellation of temperature gradient by common centroid design. (a) Thermistors placed symmetrically about diode laser. (b) Series connection gives $T_{avg} \approx T_{laser}$, allows vernier adjustment.

EXAMPLE 20.10 Common Centroid Design Of A Diode Laser Mount

The prototype ISICL sensor of Example 1.12 used a diode laser mount as shown in Figure 20.7. For packaging reasons, the mount had to stand upright. The laser's temperature was near 25° C, but had to be stable to 10 mK 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 center of the TECs, solved the problem.† The thermistors were

†If you do this, use *soft* epoxy—see Section 12.4.20.

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 10 kΩ 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 100 kΩ potentiometer, wired as shown, to null out any residual gradient sensitivity. NB: The sensor mounting is critical—differences in the bond line (glue thickness) can change the speed of response by factors of two or more.

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) bond line.

20.8.5 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 AlN, 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). (It's usually best to do this via a local proportional-only loop wrapped round the actuator and its local sensor—that way you get faster and more predictable response, so that it's easier to close the outer loop—see Sections 15.5 and 20.8.3.)

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

20.8.6 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 15.4.6), by feedforward, or just by enduring it, since it is usually infrequent. Active device turn-on or modulation can 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 could 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

[†]R. K. Karlquist *et al.*, "The Theory Of Zero-Gradient Crystal Ovens" *Proc. Int'l Frequency Control Symp.* pp. 898–908 (1997)

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 local control loop, based on a local heater and sensor, to control the active device temperature, but this approach needs very careful testing. If you're considering two loops with comparable bandwidths, beware—this leads to oscillations and flakiness.

20.9 Temperature Controllers

20.9.1 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 69° F and off at 71°. The temperature oscillates irregularly with time, but is usually between these limits, provided the heater has roughly the right capacity and the thermostat is properly placed. Too large a heater, or slow heater-sensor coupling, will lead to pronounced overshoot, so that the peak-to-peak temperature ripple may be much larger than the deadband. A less obvious benefit of bang-bang control is that it is well behaved even in the face of serious variations in the *plant*, *i.e.* the heater and the controlled volume. Your furnace doesn't go nuts if you have a big Christmas party, for instance, although all those people have a large thermal mass and produce a lot of heat themselves, not to mention the kitchen ovens and opening the front door a lot. If your application can live with its limitations, a thermostat-controlled heating loop can be just the right medicine. Just don't do it with TECs—see Section 20.7.6.

20.9.2 Linear Control

A linear controller[†] 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 (PI), and proportional-integral-derivative (PID), depending on the time dependence of the loop filter. Back in the day the three functions were often generated separately (Figure 20.9.2) and then summed. Nowadays it's generally better to use frequency-compensation ideas (Section 20.9.3) to design loops. Most temperature control loops in instruments are digital, so this is easy to do even though we need poles and zeros at very low frequencies. (Having to use 10 MΩ resistors with 10 μF capacitors gets old pretty fast.)

Proportional Loops A proportional loop uses 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 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

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

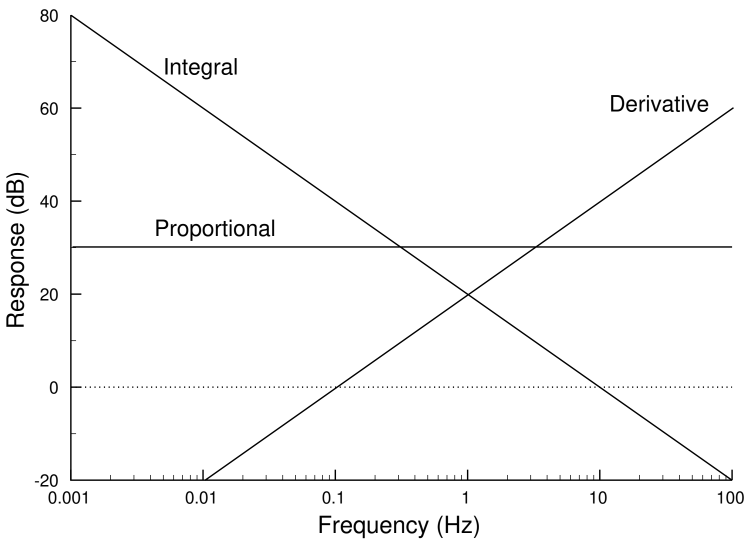

Figure 20.8 Proportional, integral, and derivative

static temperature error (or *gain error*), and dialing the proportional gain up to reduce the error will lead to instability.

Integrating Loops This error can be eliminated by using an integrating (P−I) loop, which in analog 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 with a slow plant, a straightforward integrating loop usually has a much uglier turn-on transient than a proportional-only loop. Unless the integrator gain is quite small, the integrated error signal builds up to a huge value during the slew period, and has to be discharged by an equal and opposite error after the setpoint has been crossed, a phenomenon called *windup* that we dealt with in Section 15.4.5. Note that windup is nonlinear, and is quite different from the ringing of an ordinary underdamped loop. A loop with 60° phase margin exhibits very little overshoot or ringing in normal operation; however, if it winds up, even such a well-behaved loop can go on banging the rails till you put it out of its misery.

Derivative Terms There are some plants, *e.g.* a current-controlled motor, where the actuator itself has a two-pole response.† A two-pole rolloff has a phase shift of 180°, so as we saw in Example 15.1, we need a zero in the loop filter to be able to close the loop stably. In control-system lingo, this is called a *proportional-integral-derivative* (PID) loop.

Now for a couple of points deserving emphasis. First, PID can be helpful in temperature control if the system is physically small enough that thermal diffusion doesn't become

†The current sets the torque, which couples to the second derivative of shaft angle.

important at frequencies inside the required loop bandwidth, but it's useless otherwise. (See Example 20.4).

Second, eliminating static error and increasing the loop bandwidth are very nice, but not a complete solution. Remember that we're still only controlling the temperature of the sensors; connecting that with the temperature we actually care about is entirely up to us. Nonetheless, a single layer of temperature control, with a well-insulated cold plate, can easily achieve 10 mK stability in the lab.

20.9.3 Frequency Compensation

Temperature control loops are intrinsically more subtle than ordinary op amp circuits because we are measuring T but controlling $\dot{Q}$, which makes the temperature controller act like an integrator at lowish 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 sensor temperature, perhaps with a fixed offset due to device dissipation.

You can estimate what the open loop transfer function is from (20.14); don't be daunted by the unintuitive form—your favorite math program (or 20 lines of Python or 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 and dump enough heat that you probably can't ignore the heat sink response. Typical control bandwidths for those are in the low millihertz.)

EXAMPLE 20.11 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 $K_h = \partial \dot{Q}/\partial V_h = 200$ mW/V, and we'll use a silicon diode sensor with K_s = -2.1 mV/K.

The response calculated from (20.14) and the desired overall response are shown in Figure 20.9. The loop filter is a simple lead-lag network† with a zero at 0.13 rad/s to give us high accuracy at low frequency, while not destabilizing the loop. Now we need to make sure that the overall gain, which includes the TEC's $\dot{Q}$, temperature sensor gain, and loop filter gain, make the open loop unity gain crossover occur near 13 rad/s (2 Hz). The required high frequency value is 4000 W/K.

The loop gain is the product of all the individual gains, $A_{\mathrm{VL}} = K_h \cdot K_s \cdot H_{\mathrm{plate}} \cdot H_{\mathrm{amp}}$. Since H_{amp} has flattened out well before the unity gain cross, but before the thermal

†See Example 15.1.

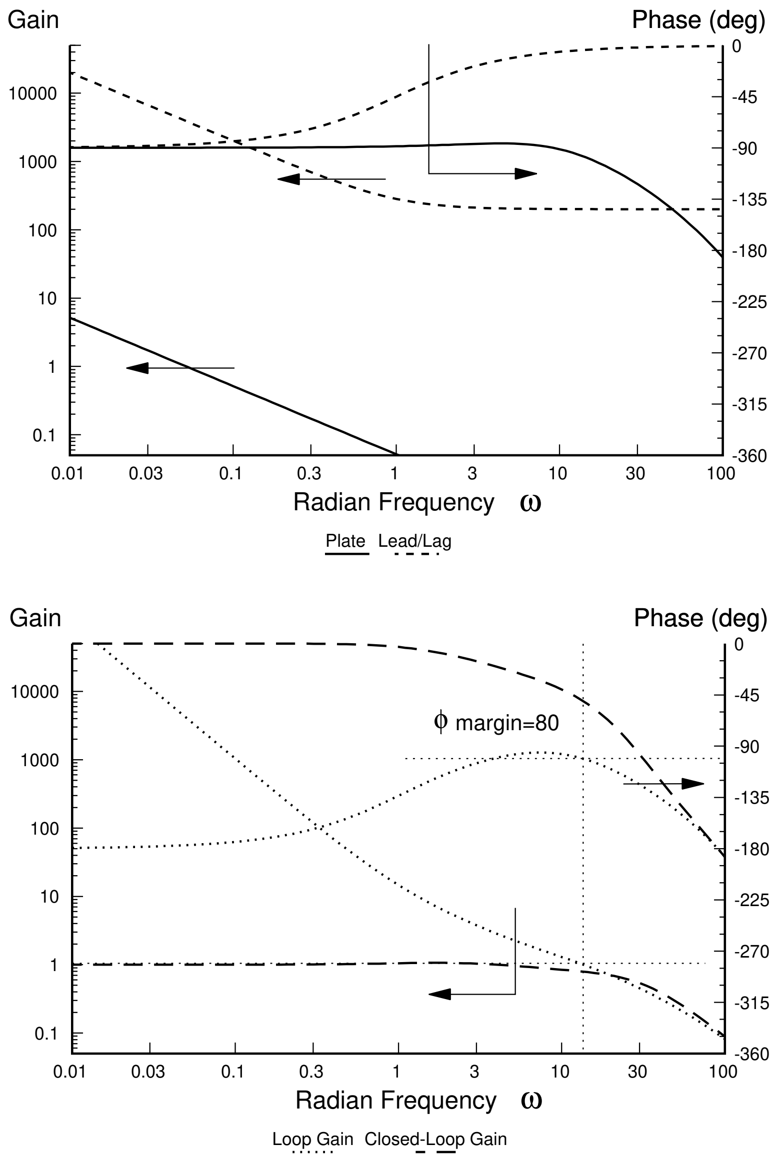

Figure 20.9 Bode plot of the 8-mm thick aluminum plate of Figure 20.2, with the temperature sensor at z=3mm, using lead-lag compensation. The unity gain bandwidth is over 2 Hz.

mass approximation fails, this is easy to solve approximately. Neglecting heat loss, the low-frequency limit of the plate's response is

$$H(\omega'|z) \sim \frac{1}{j\omega m_{\text{th}}} \omega \to 0 \tag{20.27}$$

with $m_{\text{th}} = \rho c_P V$, which is 8 $\cdot 10^{-}6$ m^3 $\cdot$ 2.7$\cdot 10^3$ kg/m^3 $\cdot$ 900 J/(kg $\cdot$K) = 1.94 J/K. Thus the high frequency limit of H_{amp} is

$$H_{amp} \approx \frac{13\,\mathrm{rad/s} \cdot 1.94\,\mathrm{J/K}}{0.2\,\mathrm{W/V} \cdot 2.1 \cdot 10^{-3}\,\mathrm{V/K}} \approx 6 \cdot 10^4. \quad (20.28)$$

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 transient is going to exhibit horrific windup unless we do something about it. We dealt with this problem in Section 15.4.5. Another approach is to use a two-speed loop, which we discussed in Example 15.4.6.

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, switch to pulse-width modulation (PWM), which is inherently linear, or use a more complicated loop filter that has lower gain at low frequencies. The best approach is probably to stop updating the integrator when the loop is slewing rapidly. In an analog loop, you can use a separate RRIO op amp for the integrator contribution. By making it very fast and attenuating its output, you can get the required *I* contribution in the steady state, but the *I* amp will rail before it can wind up much at all. (You do have to be careful to consider what happens to the summing junction when its output rails–some sort of clamp may be needed, *e.g.* antiparallel diodes between the op amp inputs.) In a digital loop, you can bound *I* arithmetically to a couple of times the maximum gain error of a P-only loop with the same constants.

20.9.4 Frequency Compensating Slow Loops: Integrator With Time Delay

In control theory, the actual physical system that accepts control inputs and produces response outputs is called the *plant*. In our diode laser system, the plant is the TEC, thermistor, and thermistor amplifier. You put a current in and get a change in voltage out.

The temporal response characteristics of a TEC + thermistor plant can vary very widely depending on the thermal mass of the cold plate and the relative placement of the TEC and thermistor. Thermal diffusion is slow, and gets slower quadratically with distance. If the thermal response is too slow, the temperature control loop may become unstable. Adjusting the servo amp to make the loop stable and well behaved is called *frequency compensation* (see Section 15.4.1). The simple way to do this is by way of a *bump test*: hit the TEC with a current step and watch how the temperature sensor responds. As we'll see, this works well even in systems where thermal diffusion dominates the high frequency response.

For frequency compensation purposes, we really only care about the early parts of the curve, before it starts to roll over and reach a steady temperature. (We assume that the sign convention is that voltage goes the same way as temperature, *i.e.* higher temperatures produce higher voltages.)

If the curve looks just like a plain decaying exponential,

$$\Delta V(t) = K(1 - \exp(-\gamma t)), \quad (20.29)$$

then the plant's transfer function is

$$H_{\text{plant}} = \frac{K}{1 + j2\pi f/\gamma}, \quad (20.30)$$

If there is any significant delay time before the sensor starts responding, you've got a delay τ followed by an integrator. The transfer function for that is

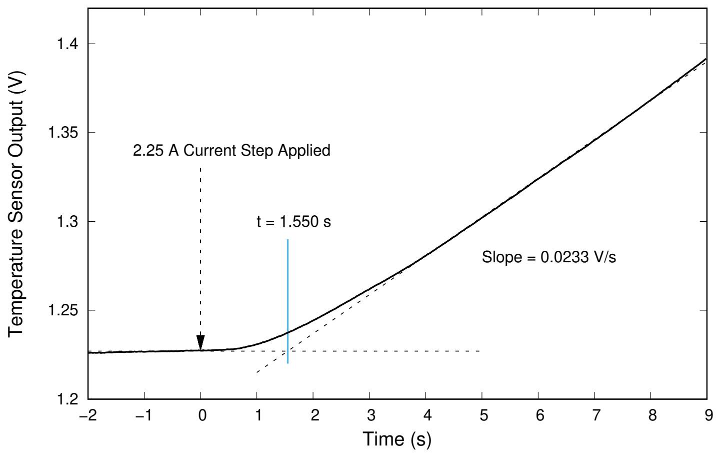

Figure 20.10 Bump test for estimating the plant parameters of a TE-cooled diode laser with a SMT thermistor soldered to a bare copper pour in intimate contact with the cold plate. A 2.25 A step change was applied to the TEC current at $T = 0$. The measured temperature stayed nearly constant for a second or so due to slow thermal diffusion, then started ramping up nearly linearly. (At some point it rolls over and becomes asymptotically constant. That's what sets the temperature limits, but it isn't very relevant for loop compensation.)

$$H_{\text{plant}} = \exp(-j2\pi f\tau) \cdot \frac{K}{j2\pi f}. \tag{20.31}$$

Of course you don't get a sharp corner; after the delay, the temperature trace curves up fairly quickly, straightens out, and then begins to curve back towards horizontal. What you need is the part near the inflection point. With a ruler, you can extend a tangent line down to the original baseline temperature value and note where the lines intersect. The time between this point and the step input is the delay τ, and the slope of the rising curve is $I \cdot K$, where I is the height of the current step.

Figure 20.11 shows one measurement of this sort for a built-up diode laser head with TE cooling.

As we know from op amp frequency compensation (Section 15.4.1) the main factor affecting loop stability is the detailed behavior of the loop gain A_{VL} near the point where $|A_{\text{VL}}| = 1$, *i.e.* the unity-gain crossing frequency f_0. In that section we talked about phase margin, lead-lag compensation, and other things of that sort, and it's just the same here, only slower and less linear. You want about 60° phase margin to avoid overshoot, and that has to happen over your full design range of heat sink temperature and ΔT.

20.9.5 Testing and Optimization of Temperature Controllers

For a lab system, it's enough to double-check the closed-loop transient response of your controller over the anticipated range of cold plate and heat sink temperatures, heat load, and transient conditions.†

If you're building temperature-controlled instruments for other people to use, final tweaking is more of a job because failures are very expensive. The best way to proceed is 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 repeat the bump test to get the step response. Taking the Fourier transform of the result, divided by the transform of the 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). If you use the slow loop method, the measured low frequency response will be different from A_{VOL} since the loop is not really open down there. However, since in the real circuit the loop gain will be high at those frequencies, that isn't much of a problem.

There are dedicated instruments called *dynamic signal analyzers* that are perfect for this job; you want at least two input channels, one for the step function and one for the temperature sensor. More channels will let you follow the behavior of the local feedback loops as well.‡ Nice boat anchors for this job include the HP 35665A and HP 3562A, which have built-in software for the feedback loop application. (Look for that if you're buying a new model.)

Make sure that you repeat the bump testing 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 lower temperature limits. Near the maximum ΔT your TEC/heat sink combination can reach, every additional milliamp pumps less and less heat because more of it goes into I^2R heating, so K_{TEC} goes to zero. Your loop is liable to overshoot massively or even oscillate near there if you've done anything at all fancy with the frequency compensation, and of course that's bound to push the loop into the meltdown region of Figure 20.5. (You did remember the current limiter and thermal cutout, right?)

Once you have a measurement of the open loop transfer function, proceed just as we did in Section 15.4.1, and don't get too ambitious; 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.

20.9.6 Thermal Simulations: A Hack

Circuit simulators such as the excellent (and free) LTspice from Linear Technology are quite useful for doing thermal models. The heat equation is a first-order PDE, but many systems can be modeled as heat sources, thermal conductors, and thermal masses. The mathematics of that is just like *RC* circuits, except that all the capacitors have to have one end grounded.

The author's colleague John Larkin of Highland Technology likes to use SPICE for simple thermal simulations. He points out the interesting equivalences in Table 20.9.5.

†A 20% overshoot isn't serious for small signals, but will do bad things if you try swinging from a hot temperature to the low-temperature limit.

‡People have homebrewed these by hacking a PC sound card to remove the AC-coupling capacitors.

Table 20.1 Thermal properties of gases

Gas	α (W/m/K) @ 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
$CHCl_2F$ (Freon-21)	0.0097	+9° C	417
Krypton	0.0090	121K	108
Xenon	0.0055	-107° C	96

Table 20.2 Thermocouple Properties

Type	Composition	$\partial V/\partial T$ @0° C (μV/K)	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-(any metal)	700–1200†	250
	Graphite-Copper	± 10–20	500
	Graphite-Boron Carbide	40	2500
	Copper-Copper wire	±0.2	
	Copper-Silver	0.3	
	Copper-Gold	0.3	
	Copper-Kovar	40	

With a bit more work, *RC* transmission lines can be used to model 1D thermal diffusion at some level, though the number of sections becomes large if you start wanting spatial resolution or an accurate transient response. A larger problem is that SPICE is a pretty capable solver for systems of nonlinear ordinary differential equations (ODEs) but can't do transport problems in any generality; there's too much internal state that it doesn't know about. (Transmission lines were hacked in as a special case in SPICE—if you try modeling them as *LC* ladder networks, the number of sections required for a given accuracy goes up as the square of the length in wavelengths.)

Table 20.3 Cryogens and low temperature mixtures

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

.

Table 20.4 Approximate electrical equivalences for thermal quantities which enable the use of SPICE for thermal simulations as long as diffusion isn't important. Note that since the heat equation is first-order in time, there is nothing corresponding to inductance, and all capacitors must connect to ground. (Current flowing into the capacitors represents heat flow, but the current flowing to ground is unphysical.

ELECTRICAL	THERMAL
1 A	1 W
1 F	1 g Al†
1 V	1° C
1 s	1 s
1 Ω	1 K/W

Table 20.5 Thermal properties of some commonly used metals. These have been collected from various (and sometimes inconsistent) 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. The tradeoff between electrical and thermal conductivity is expressed by the parameter $1/(\alpha\rho_E)$ (higher is better for getting signals on and off the cold plate).

Material	α W/m/K	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 $\mu\Omega$ m	1/($\alpha\rho_E$) K/(mΩW)	TC ρ_E ppm/K
Gold	317	14.2	129	19.3	12.7	0.022	144	+3900
Silver	417	18.9	236	10.5	16.8	0.014	170	
OFHC Copper	390	17	390	8.96	11.2	0.017	150	+4300
Copper Wire	120–220	17	390	8.96	3.4–6.3	0.018	250-460	+4300
Phosphor Bronze (94.8Cu 5Sn P)	48	18	376	8.9	1.43	0.12	180	
Free-Machining Brass (3% Pb)	110	20	390	8.5	3.3	0.06	150	+1500
1100-T0 soft Aluminum	240	23	900	2.7	9.9	0.03	140	+4000
6061-T6 hard Aluminum	180	23	900	2.7	7.4	0.05	110	+4000
Nickel	90	13	440	8.9	2.298	0.07	160	
German Silver (82Ni 18Cu Zn)	23	16	400	8.7	0.661	2.5	17	
304 Stainless	15	9.6	470	8.0	0.40	0.57	120	
Mild Steel (0.4% Carbon)	80	12	450	7.8	2.279	0.45	28	
Manganin	23	18	410	8.4	0.668	0.42	100	±10
Nichrome	13	17	430	9	0.336	1.08	71	+110
Constantan	22	17	400	8.4	0.655	0.71	64	-30

Table 20.6 Thermal properties of common inorganic nonmetals. 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.

Material	α W/(K·m)	CTE (ppm/K)	c_P J/kg/K	$10^{-3} \cdot \rho$ kg/m^3	$10^5 \cdot \kappa$ m^2/s	ρ_E $\mu\Omega \cdot m$ (†ε_r)
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	3.8†
Corning ULE glass	1.3	±0.03	745	2.21	0.079	4.0†
BK-7 Glass	1.1	8	700	2.5	0.063	†
Sapphire (crystalline Al_2O_3)	32–35 anisotropic	7	770	3.6	1.2	$9.4 \perp, 11.6 \parallel$ †
Pure Silicon	146	2.6	712	2.33	8.801	11.7†
Highly doped Silicon	100	2.6	712	2.33	6.028	
Silicon Nitride	27	3	711	3.3	1.2	7†
Natural Diamond (300K)	630 (strongly dependent on purity)	0.8	506	3.5	36	5.7†
Graphite	110–190 anisotropic	8	715	1.7	9-16	1000–2000
Macor	1.46	8.5	790	2.5	1.0	6†
Boron Nitride	20	1.2	790	2.3	1.1	4†

Table 20.7 Thermal properties of common plastics and foams. 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.

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)	ε_r
Nylon-66	0.25	80	1700	1.1	0.013	4
LD Polyethylene	0.33	100-200	2100	0.92	0.017	2.3
HD Polyethylene	0.50	100-200	2300	0.95	0.023	2.3
Polytetrafluoro-ethylene (PTFE)	0.25	300	1000	2.2	0.011	2.0
Worst-case PTFE	0.25	2900	1000	2.2	0.011	2.0
Delrin	0.38	68	1500	1.42	0.018	3.7
Polyimide	0.12	20	1100	1.42	0.008	3–4
Polypropylene	0.2	90	900	0.90	0.025	2.2
Silicone rubber	0.2	600	770	1	0.03	
Kel-F	0.84	70	1400	1.4	0.043	2.6
Vespel	0.29	30	1130	1.43	0.018	3.5
PET (Mylar)	0.21	60	1300	1.4	0.012	3.2
Polycarbonate	0.2	66	1200	1.3	0.013	3
Acrylic	0.2	72	1500	1.4	0.010	2.6
Polystyrene	0.15	34–210	1400	1.06	0.010	2.5
HD Styrofoam	0.04		1400	0.03	0.095	1.03
LD Styrofoam	0.03		1400	0.005	0.429	1.0
Glass foam	0.04-0.055	10	820	0.09	0.05–0.07	1.2

Table 20.8 Thermal properties of common thermal interface materials (TIMs) and a few others of interest. 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.

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	$1/(\alpha\rho_E)$ K/($\mu\Omega$W)
Oriented Carbon Fiber Tape (Vel-Bond)	100						
Carbon Fiber Composite	30–200 anisotropic	6	1.2	1.85			
Sn/Pb Solder	50	24.1	170	9.3	3.2		
Silver Epoxy	1–5	50					
Premium Silver Epoxy (Diemat DM6030HK 83% Ag (claimed))	60	26		3.8-4.5			
Indium-tin eutectic solder, 52 In/48 Sn (Indalloy 1E)	34	20	230	7.3	2.025	0.15	0.20
Indium	85	29	233	7.3	4.997	0.07	0.17
Indium Paste (Indalloy)	(≈30) (varies)						
Omega OB-200 Thermal Epoxy	1.3	38					
Epo-Tek H20E Thermal Epoxy	29	$31 < T_g$ $160 > T_g$				4	
Circuit Works 7100 Silver Grease	7.2					10	
Omega OB-201 Thermal Grease	2.3		2000	2.5	0.046		
Ordinary Thermal Grease	0.8		2100	2.4	0.016		
Bismuth Telluride	1.5	13	550	7.5	0.036		
Water	0.60	(21)	4200	1.0	0.014		