SSC19-WKIV-07

Design and Implementation of a Thermoelectric Cooling Solution for a CCD-based NUV Spectrograph

Nicholas DeCicco, Nicholas Nell, Kevin France, Stefan Ulrich, Arika Egan, Brian Fleming, Rick Kohnert
Laboratory for Atmospheric & Space Physics
1234 Innovation Drive, Boulder, CO 80303; 1-303-735-8214
Nicholas.DeCicco@lasp.colorado.edu

ABSTRACT

The Colorado Ultraviolet Transit Experiment (CUTE) is a 6U CubeSat designed to obtain transit spectra of more than ten close-orbiting exoplanets. To this end, CUTE houses a near-ultraviolet (~250 – 330 nm) spectrograph based around a novel rectangular Cassegrain telescope; the spectrograph sensor is an off-the-shelf Teledyne e2v CCD. To achieve desired spectral signal-to-noise ratio (SNR), dark current is reduced by cooling the CCD to a temperature of −50 °C with a thermoelectric cooler (TEC). The TEC is driven by a constant current buck converter with an H-bridge topology for bidirectional current control. The packaging of the CCD imposes a maximum time rate of change of temperature of 5 K/min. A cascaded software control loop (discussed here) was developed that constrains this time rate of change within allowable bounds while simultaneously driving the CCD temperature to a desired setpoint. Criteria for sizing a TEC to the application and initial laboratory results are discussed, as well as digital filtering methods employed and possible solutions to integral wind-up.

OVERVIEW

Charged coupled devices (CCDs) have been widely used in astronomical imaging and spectroscopy applications. However, all CCDs are subject to noise in the form of an internally generated dark current. This current can be reduced by lowering the detector temperature: previous work[1] established that, for the Teledyne e2v CCD42-10 to be used in CUTE, a device temperature of approximately −50 °C would yield an SNR sufficient for CUTE's science goals. Preliminary system-level thermal simulation showed an on-orbit spacecraft temperature of around 0 °C, thus a system to produce a roughly 50 K delta between the spacecraft temperature and CCD was required. A thermoelectric cooler (TEC) was chosen as the cooling solution due to their small size, relative ease of implementation, and ability to produce the required temperature delta.

While producing a drive current to operate a TEC is relatively straight-forward, the control system is complicated by the requirement that the CCD42-10 maximum rate of heating or cooling of 5 K/min (0.083 K/sec) never be exceeded. Simple PID control system designs do not provide mechanisms that would adequately guarantee that this requirement always be met while simultaneously providing (relatively) fast settling times, so a nested control system that would more assuredly avoid large $\mathrm{d}T/\mathrm{d}t$ (along with reasonable settling times) was developed.

A BRIEF OVERVIEW OF THERMOELECTRIC COOLERS (TECs)

The typical single-stage TEC is a two-terminal device consisting of an array of rectangular thermoelectric elements wired electrically in series and arranged thermally in parallel. The application of electric current to the TEC generates a heat flux through the TEC (in the axis normal to the plane of the TEC faces) by way of the Peltier—Seebeck effect. This heat flux creates a temperature gradient through the TEC. The temperature gradient across each element, coupled with the Seebeck (and therefore Peltier) coefficients' temperature dependency, results in an additional heat flux term known as the Thompson effect, which is small in value and therefore *generally* ignored, though there is argument that it should not be.[2]

The energy balance equations governing CCD operation have been well-established[2,3,4,5,6] and will not be discussed in detail here, but arguably most important is the relationship between the hot and cold side heat fluxes, q_h and q_c:

$$q_h = q_c + P_{\text{in}} \tag{1}$$

where P_{in} is the power input to the device. While much of the input power is dispelled as waste heat by Joule heating, the remainder of this power is what generates the temperature delta across the device by the Peltier—Seebeck effect:

$$P_{\text{in}} = IS_m\Delta T + I^2R_m,$$

where S_m is the *device* Seebeck coefficient, I is the current through the device, ΔT is the temperature delta across the TEC, and R_m is the Ohmic resistance of the device. The voltage across the device is

$$V = \frac{P}{I} = S_m \Delta T + I R_m,$$

which is to say that the voltage across the device is not due entirely to the Ohmic behavior of the device.

Note that the overall energy balance equation (1) does not provide a full description of the fluxes q_c and q_h nor fluxes internal to the TEC (i.e., due to thermal conductivity), but does provide a sufficient description of the TEC when "performance curves" (graphs of ΔT vs. I for a family of q_c) and I-V curve is provided by the manufacturer. The methods of modeling single- and multi-stage TECs described in the literature[2,3,4,5,6] require knowledge of TEC parameters (S_m and thermal conductivity k_m) not generally divulged in manufacturer-provided datasheets. Readily available techniques exist[5,6] for deriving these parameters from datasheet parameters for single-stage TECs, but not for multi-stage TECs such as the three-stage II-VI Marlow SP2402 that was selected for use in CUTE.

Selection of an appropriately sized TEC

The heat flux generated (and resulting temperature gradient) allows the TEC to be used in applications that require either heat pumping with no temperature differential (in which case the TEC is capable of pumping the maximum amount of heat), or applications which require a component to be held at a particular temperature (in addition to removal of heat). Cooling a CCD to a desired temperature is the latter case, which means that the TEC is incapable of pumping its maximum rated heat capacity, and so the first tradeoff makes itself apparent: obtaining large temperature deltas requires selecting a TEC with a $Q_{\max}$ appreciably higher than the anticipated cold side heat load.

Second, the maximum no-load temperature delta achievable with a TEC is primarily limited by the number of TEC stages. Though this is dependent also on the choice of semiconductor used in construction, as most TECs are made with Bi_2Te_3 elements, this maximum is relatively constant across 1-, 2-, and 3-stage TECs, with single-stage TECs generally capable of achieving deltas of around 60-70 °C, and multi-stage coolers capable of 130-150 °C.[7]

However, as TEC elements are (typically) wired in series, if TEC element density per unit area is regarded as constant, TEC electrical resistance R_m increases linearly with increasing TEC area and, likewise, for increasing number of stages with hot side area held constant. Power dissipated by the TEC for constant input current I increases by the square of the resistance, so a decision to use a multi-stage TEC for more $\Delta T_{\max}$ margin is generally made at the expense of higher hot side heat loads q_h to be removed by the system for the same q_c capacity.

Adequate margin is required to avoid control loop induced thermal runaway, which occurs if I is increased past the point where $\mathrm{d}\Delta T/\mathrm{d}I = 0$ ($\Delta T_{\max}$); there is a point past which more heat is produced by Joule heating than is pumped by the module, decreasing ΔT. This is especially an issue for the control loops discussed herein, as the integral terms in these control loops will increase TEC current ever higher if the desired ΔT cannot be obtained. A hardware limit on TEC current is therefore advisable.

Finally, some amount of ΔT margin should be set aside for degradation of the TEC performance over time, which can be as much as about 10% per year in continuous use. Ripple voltage also negatively affects performance, so it is imperative that the TEC supply have as little ripple as possible.[8]

These considerations in mind, any design starts with the two primary requirements: ΔT required and the magnitude of q_c that is needed to be removed from the device. The TEC must be able to remove not only the heat that is generated by the object that it is primarily intended to cool (which, for CUTE, is the CCD) but also any heat which is parasitically conducted (q_{cond}) into or radiated (q_{rad}) onto the cold side of the TEC—including anything in contact with the cold side of the TEC, for that matter. Thus, the total heat to be removed is

$$q_c = q_{\mathrm{device}} + q_{\mathrm{rad}} + q_{\mathrm{cond}}.$$

Radiative heat loads can be reduced by careful selection of surface finishes and composition to increase their IR radiation shielding properties: i.e., wherever possible, the cooled object should be either constructed from or shielded by materials with very low IR emissivities. Copper, for example, when given a smooth finish and removed of surface oxides, can have an emissivity as low as ~0.03.[9] Reduction of cross-sectional area of the cooled object and preferential orientation of the cooled object with respect to the environment it is contained in (i.e., to minimize radiation view factor) can also reduce radiative heat loads.

It is advisable to compute an upper bound on possible radiative heat loads by assuming worst case view factors (i.e., $F = 1$), worst case emissivities ($\varepsilon = 1$ for all surfaces), and over-estimate surface area of affected

objects. The radiative heat load upper bound is simply then

$$q_{\text{rad}} = A\sigma(T_{\text{ambient}}^4 - T_c^4),$$

where σ is the Stefan-Boltzmann constant. Thus, for CUTE's hot side temperature of approximately 0 °C and cold side temperature of approximately −50 °C,

$$\frac{q_{\text{rad}}}{A} \approx 17.5 \text{ mW/cm}^2.$$

As the cooled object has an area on the order of square centimeters, this gives a hard upper bound of likely no greater than ~100 mW.

Conductive heat loads can be diminished by removing any parasitically conductive paths between the TEC hot and cold sides, and, if these paths cannot be eliminated (e.g., due to mechanical considerations, vibration chief among those in aerospace applications[3]), materials with low thermal conductivities should be employed.

DESIGN OF THE CONTROL LOOP

The control loop must achieve two goals:

1. It must drive the CCD (TEC cold side) temperature to a target temperature T.
2. It must strive to limit changes in temperature per time T' to less than some maximum (here, 5 K/min).

There is no "time limit" which constrains how quickly the first goal must be achieved, so long as the amount of time required to reach the target temperature is not unreasonable: if we cooled the CCD to −50 °C from a hot side temperature of +20 °C, a rate of 2.5 K/min would require 28 minutes to reach the target temperature, but (for sake of comparison) a rate of 0.1 K/min would require nearly half a day to cool the CCD.

A traditional P-, PI-, PD-, or PID controller applied in the typical fashion (whereby the error term $E(s)$ is proportional the difference between the target temperature T_0 and the system temperature $T(s)$) can relatively easily be made to achieve one of these goals—i.e., minimizing the error term—but it is difficult to guarantee that the maximum time *rate* of change of temperature will never exceed some maximum T'_{max}. Further, that these types of controllers will *not* exceed said maximum can only be proven for some limited operating regime: system variables must be assumed to not stray outside of certain bounds, and must also be assumed not to change at rates above some limits. For example, one could design a P controller that slews output temperature slowly so long as the delta between setpoint temperature T_0 and system temperature $T(s)$ never exceeds some maximum, but it should be easy to conceive of situations in which insufficient margin is designed into the system and actual on-orbit conditions result in excess slew rate.

Thus, to help ensure that the temperature slew rate is limited, a nested control loop was developed which can achieve both goals without discontinuities in control output. This nested control loop was evolved from an earlier iteration which achieved the same behavior, but had undesirable control discontinuities.

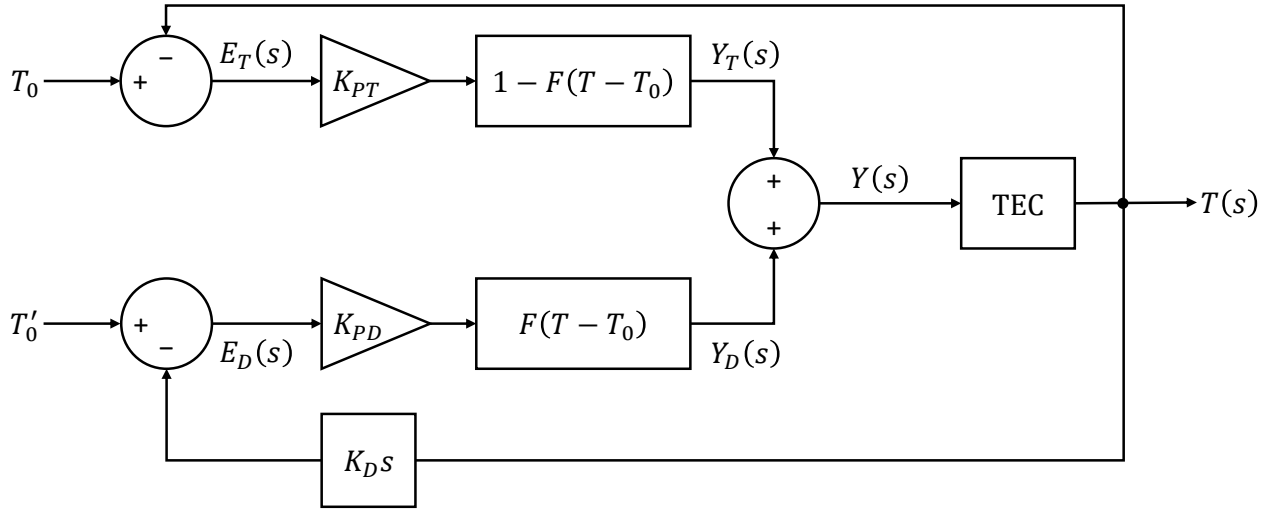

Figure 1: An initial attempt at devising a control loop capable of achieving the goals stated in the text.

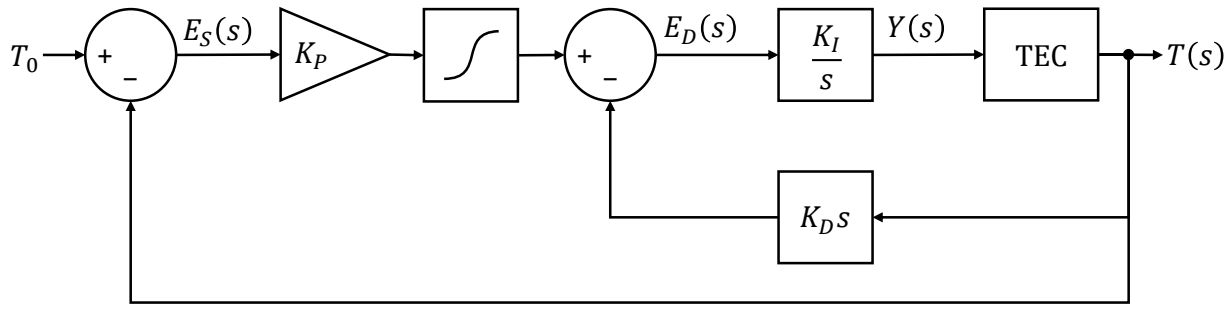

Figure 2: The improved controller.

An initial attempt

The first attempt (illustrated in figure 1) consisted of an isolated (separate) pair of control loops: the first (the "temperature setpoint controller" or just "setpoint controller"), with proportional term K_{PT} in the forward path, seeks to minimize the error term $E_T(s) = T(s) - T_0$; the second (the "derivative setpoint controller" or just "derivative controller"), with proportional term K_{PD} in its forward path, attempts to minimize the error term $E_D(s) = sT - T_0'$. These two separate control loops are then summed to form the control signal $Y(s)$, which is fed to the plant (labeled "TEC"), the output of which is a cold-side (CCD) temperature $T(s)$. (For the purpose of this paper, the input to the plant is a unitless control signal of arbitrary scale; in practice, this signal is proportional to TEC current.)

The function $F(T - T_0)$ "blends" the output of the controller $Y(s)$ between the control signals $Y_T(s)$ and $Y_D(s)$, such that at some moment t, the controller output $y(t)$ is equal to either $y_T(t)$, $y_D(t)$, or some linear combination thereof, dependent on the difference between the temperature T and its setpoint T_0.

This function might ideally be a sigmoid, but in its simplest form can be a discontinuous piecewise function:

$$F(T - T_0) = \begin{cases} 1, & |T - T_0| > T_p \\ 0, & |T - T_0| < T_p \end{cases}.$$

With this definition, if the current temperature T is further than some proximity T_p from the temperature setpoint T_0, then the setpoint controller will be effectively "switched off," and the derivative setpoint controller will be "switched on," causing the overall behavior of the entire control loop to attempt to drive the temperature time derivative $T'(t)$ to the setpoint T_0'. When the derivative controller has driven the cold-side temperature T sufficiently close (less than T_p away from) T_0, the output of F "inverts," granting the setpoint controller full control of the plant.

Note that for proper operation both above and below the temperature setpoint T_0, T_0' must switch in sign; i.e.,

$$T_0' = \begin{cases} -|T_0'|, & T > T_0 \\ |T_0'|, & T < T_0 \end{cases}.$$

There are a number of problems with the practical implementation of this design, however. The first is a lack of hysteresis in F. Without hysteresis, when control switches from the derivative setpoint controller to the temperature setpoint controller, if the setpoint controller forward path gain is not sufficiently large to generate a control signal equal to or greater than that produced by the derivative controller, then the temperature will climb such that $T - T_0 > T_p$, and the derivative controller will be handed control once more. This will result in rapid oscillations of control output as control switches from $y_T(t)$ to $y_D(t)$.

A solution to this problem is to alter the setpoint controller to be a PI-controller instead of just a P-controller.

However, while the application of these "fixes" minimize discontinuous jumps in control output, sudden changes in heat load or environmental changes could result in undesirable behavior from the temperature setpoint controller.

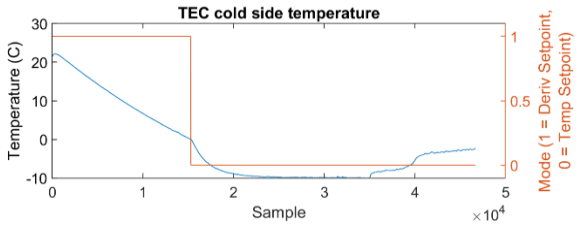

Figure 3: A test of the original controller using the Marlow RC3-2.5 single-stage TEC. The right *y* axis shows the value of $F(T - T_0)$. The perturbation at sample 35,000 corresponds to a 100W incandescent light bulb directed towards the TEC cold side (as a radiative heat load) being switched on.

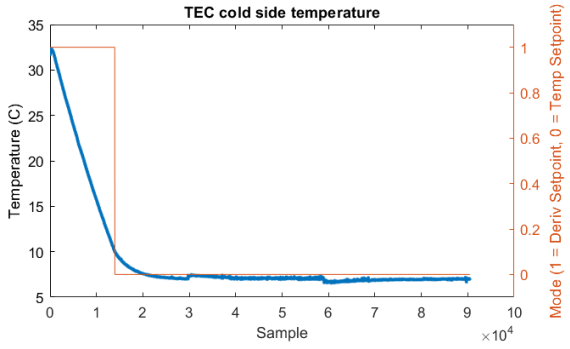

Figure 4: A test of the original controller, plus the addition of an integral term, hysteresis, and adjustment of the integral term at the cross-over point to eliminate discontinuities in control, using the Marlow RC3-2.5. The right *y* axis shows the value of $F(T - T_0)$. The perturbation at samples 30,000-60,000 corresponds to a 100W incandescent light bulb directed towards the TEC cold side being first switched on, then off.

A different approach

As the limitations of the previous approach became more apparent, a different approach was explored. This is shown in figure 2. Note that we assume K_P, K_I and K_D to be positive in the discussion that follows. In this control loop, a saturation block is visible. The time-domain behavior of this block is that of a sigmoid function,

$$s_0(x) = \frac{2}{1 + e^{-x}} - 1$$

that has been scaled and shifted like so:

$$s(x) = -T_0' s_0\left(\frac{-2x}{T_0'}\right).$$

The leading T_0' scale factor sets the upper and lower bounds of the sigmoid (more formally, the limits of $s(x)$ as $x \to \pm 0$) to the derivative set point on the upper bound and the negative of the derivative set point on the lower bound. The -2 scale factor adjusts the steepness of the sigmoid and can be altered to adjust system response.

The overall behavior of this control loop can be described quantitatively for two primary cases.

First, consider when $K_P|T - T_0| > T_0' + \varepsilon_1$, where $s(T_0' + \varepsilon_1) \approx T_0'$. In this case, the sigmoid saturates at (approximately) $\pm T_0'$ (depending on the sign of $T - T_0$). The outer control loop is effectively "disabled" so long as this condition persists, and the inner control loop minimizes the $e_D(t)$ (derivative setpoint error) term to make T' approach the derivative setpoint T_0'.

An expression for temperature in the s-domain in this limit ($K_P|T - T_0| > T_0' - \varepsilon_1$) dependent on the combined TEC controller, TEC, and cooled object transfer function $P(s)$ is

$$T(s) \approx \frac{K_I T_0 P(s)}{s(K_D K_I P(s) + 1)}.$$

Next, consider when $K_P|T - T_0| < T_0' - \varepsilon_2$, where $\varepsilon_2 \ll T_0'$ such that the sigmoid is roughly linear over the interval $s(x) \in [-T_0' + \varepsilon, +T_0' - \varepsilon]$. Here, the setpoint control signal $K_P e_S(t)$ "sneaks through" the sigmoid function such that $e_D(t) \approx K_P e_S(t)$ (assuming the limit of $\mathrm{d}s(x)/\mathrm{d}x$ is 1 as $x \to 0$; otherwise, $e_D(t) \approx \beta K_P e_S(t)$, where β is the value of said limit). The behavior of the control loop in this regime be described qualitatively by two sub-cases that each can be further split into three additional cases:

1. $T > T_0$
   a. $T' = 0$
   b. $T' > 0$
   c. $T' < 0$
2. $T < T_0$
   a. $T' = 0$
   b. $T' > 0$
   c. $T' < 0$

We will examine cases 1(a) through 1(c): cases 2(a) through 2(c) are identical in behavior but for sign reversals.

In case 1(a), $e_D(t) \approx K_P e_S(t)$, so the overall control response is that of the simple temperature setpoint P-controller.

In case 1(b), a negative control signal $e_D(t)$ is required to minimize the setpoint error, and the temperature is trending upwards. $K_P e_S(t)$ is negative in sign, as is $-k_D T'$, so both the inner and outer control loop behaviors sum additively to result in a larger control signal than would have been obtained with just the simple temperature setpoint P-controller; the inner control loop "helps along" the outer control loop, increasing gain, to combat the upward trend of T, but as soon as T levels out (before trending down), case 1(a) will be in effect, and the inner control loop will not contribute additional gain.

In case 1(c), a likewise negative control signal $e_D(t)$ is required to minimize $e_S(t)$, and temperature is heading in the correct direction (down, towards T_0). The negative sign of $K_D T'$ is canceled by the sum block, resulting in a value of $e_D(t)$ that is less than the control signal leaving the sigmoid (approximately equal to $e_S(t)$). Thus, the inner control loop "slows down" the outer control loop.

An expression for temperature in the s-domain in this limit ($K_P|T - T_0| < T_0' - \varepsilon_1$) is

$$T(s) \approx \frac{K_I K_P T_0 P(s)}{s + P(s)K_I(K_P + K_D s)}.$$

For a typical system, the behavior of $P(s)$ is dominated by the term which models the heat capacity of the object being cooled: the settling time of the object temperature for a step input is typically appreciably slower than the settling time of the TEC controller and TEC step response with no load.

In all cases, the effect of the integral term K_I/s is to "translate" the derivative control signal $E_D(s)$ back into a "non-derivative" control signal.

In practice, a small amount of oscillation is visible in steady-state. For lack of more detailed analysis, the exact cause of this is uncertain. A proper analysis would require derivation of the plant transfer function $P(s)$, which is non-trivial.

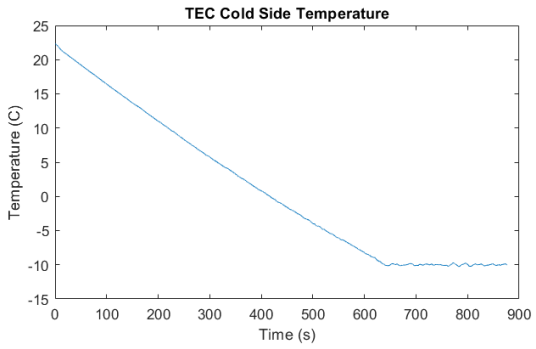

Figure 5: Behavior of the final control loop design for a Marlow NL2012T with no thermal mass attached to the cold side in ambient air conditions and $T_h = 22$ °C.

INTEGRAL ANTI-WINDUP

As the control system here described contains an integral term in its forward path, the problem of integral wind-up exists and therefore consequences of its ill effects must be considered. For example, if the software control loop were to stop running for a period of time for some arbitrary reason (e.g., due to some unforeseen complexities of the operating system scheduler) while environmental conditions were to change, the value accumulated by the integral term could result in discontinuous operation when the software control loop resumes running.

A possible solution may be to pragmatically store the TEC performance curves in memory, then periodically check for either an excess of elapsed time since the last iteration through the control loop, or a large discrepancy between the commanded TEC current and observed ΔT (for some assumed q_c). If either is observed, the value accumulated by the integrator could then be altered to yield a current (I) control signal value predicted by the performance curves for the observed ΔT.

SYSTEM IMPLEMENTATION

The TEC driver system being developed for CUTE consists of a software control loop which runs in Micrium µC/OS-III on the ARM microprocessor core of a Xilinx Zynq system-on-a-chip (SoC) at a fixed frequency of 20 Hz. Experiments were performed using the MAX1968 TEC driver IC to produce the TEC drive current, but it is the intention of the authors to use the LTC1923 in the final design: a design using this IC is currently in progress. Both ICs drive the TEC using four MOSFETs in an h-bridge configuration; PWM signals are supplied to the MOSFET gates, and a pair of LC low-pass filters smooth the h-bridge output to produce a DC voltage across the TEC.

A thermistor is used for hot-side temperature measurement, and a platinum resistance temperature detector (RTD) for cold-side temperature measurement. A combination of digital low-pass and moving average filters are used to filter the RTD signal measured with an ADC: this is a requirement if the ADC lacks sufficient precision to permit computing the temperature derivative without excessive noise.

Results of testing displayed in this paper were performed with no thermal mass attached to the TEC cold side, except the aforementioned RTD. The TECs were affixed to an aluminum heat sink $7.25 \times 2.25 \times 4.125''$ in size with a thin layer of Wakefield Vette no. 120 thermal paste, and secured with a small piece of Kapton tape. The thermistor was placed adjacent to the TEC on the surface of the aluminum heat sink.

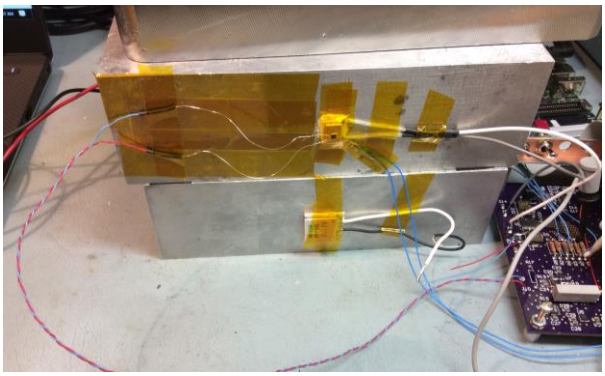

Figure 6: Experimental test setup.

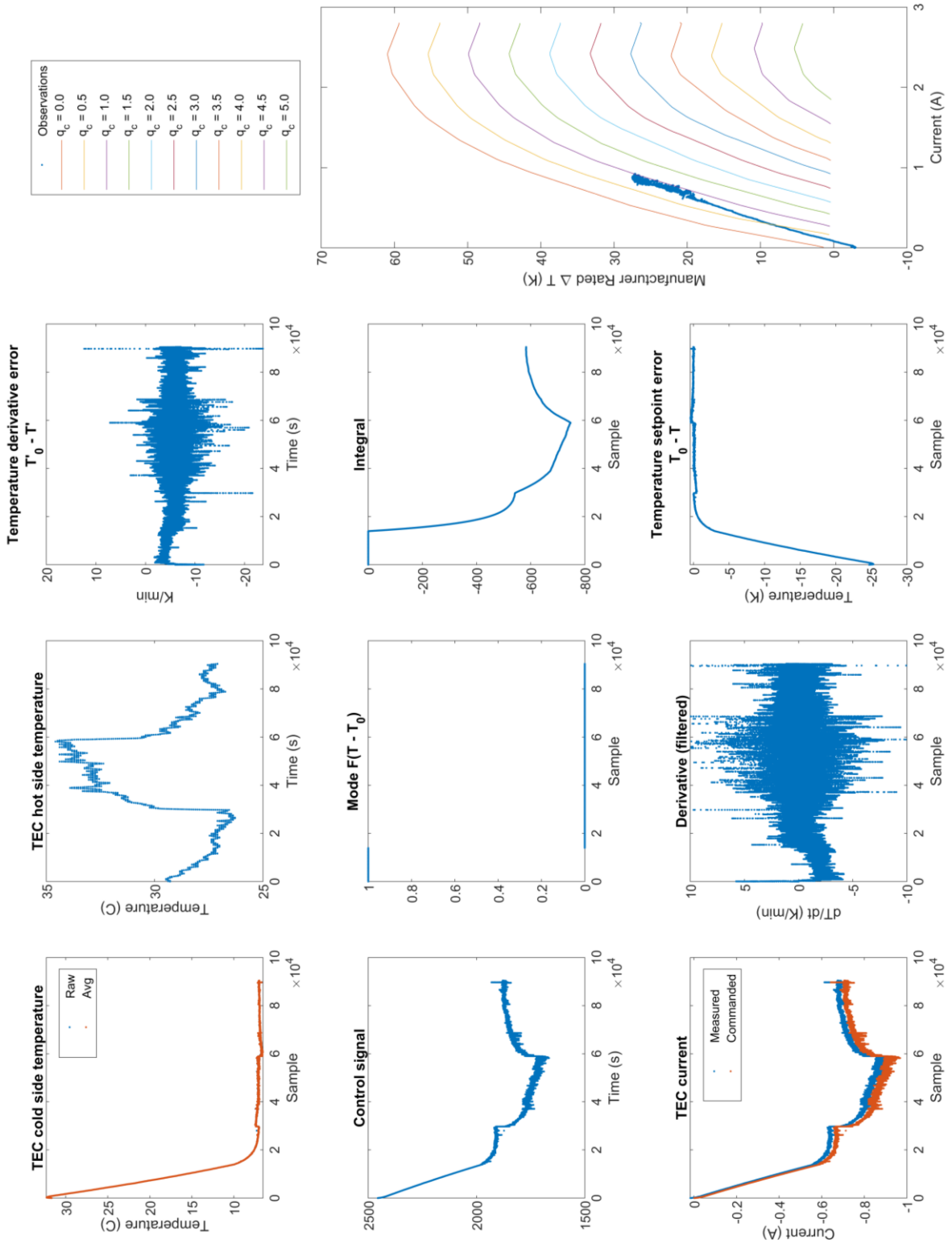

Figure 7: Detail view of the first iteration control loop behavior for the same conditions as in figure 4.

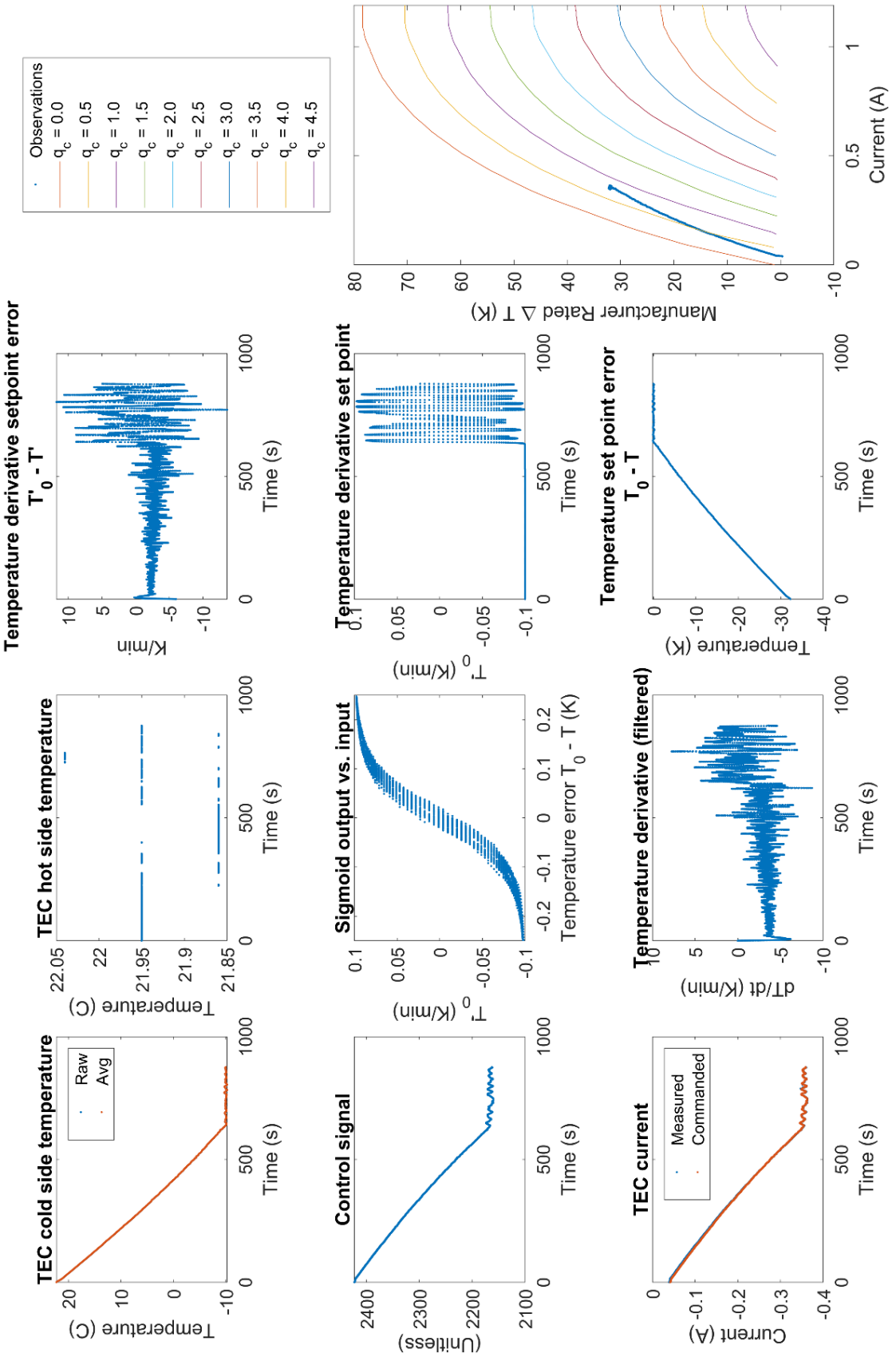

Figure 8: Detail view of the final control loop behavior for the same conditions as in figure 5.

CONCLUSION

The problem of developing a TEC-based cooling system poses a number of challenges as explored in this paper. The complexities of TEC control are complicated by a need to regulate temperature time derivative in addition to temperature setpoint. Two control loops proven by experiment to be capable of this were herein described. The application of finite element analysis to the study of the TEC mechanical mounting arrangement and numerical modeling of the TEC could permit software simulation and tuning of the control loop.

REFERENCES

1. Fleming, B. T. "Colorado Ultraviolet Transit Experiment: a dedicated CubeSat mission to study exoplanetary mass loss and magnetic fields," Journal of Astronomical Telescopes, Instruments, and Systems, vol. 4, no. 1, January-March 2018.
2. McCarty, R., "A Comparison Between Numerical and Simplified Thermoelectric Cooler Models," Journal of Electronic Materials, vol. 39, no. 9 February 2010.
3. Ferrotec, "Thermoelectric Reference Guide," [Online]. Available: https://thermal.ferrotec.com/technology/thermoelectric-reference-guide/ [Accessed: Jun. 6, 2019]
4. Chavez, J. A. et al., "SPICE Model of Thermoelectric Elements Including Thermal Effects" in *Proc. 17th IEEE Instr. and Meas. Tech. Conf.* Baltimore: IEEE, 2000, vol. 2., pp. 1019-1023.
5. Lineykin, S. and S. Ben-Yaakov, "Analysis of Thermoelectric Coolers by a Spice-Compatible Equivalent-Circuit Model," IEEE Power Electronics Letters, vol. 3, no. 2. June 2005.
6. Lineykin, S. and S. Ben-Yaakov, "Modeling and Analysis of Thermoelectric Modules," IEEE Transactions on Industry Applications, vol. 43, no. 2. March-April 2007.
7. Zhao, D. and Tan, G, "A review of thermoelectric cooling: Materials, modeling and applications," J. App. Thermal Eng., vol. 66, iss. 1-2, May 2014.
8. Texas Instruments, "Closed-Loop Temperature Regulation Using the UC3638 H-Bridge Motor Controller and a Thermoelectric Cooler," September 2001. [Online]. Available: http://www.ti.com/lit/an/slua202a/slua202a.pdf [Accessed June 11, 2019].
9. Reid, R. L. and C. W. Coon, "Hemispherical Thermal Emittance of Copper as a Function of Oxidation Conditions," in *Heat Transfer and Spacecraft Thermal Control*, J. W. Lucas and M. Summerfield, Eds. Cambridge: MIT Press, 1971, pp. 184-204.