Temperature Measurement in Dimensional Metrology – Why the Steinhart-Hart Equation works so well

Michael Matus[1]

[1] *BEV, Bundesamt für Eich- und Vermessungswesen, Arltgasse 35, 1160 Wien, Austria*

Abstract
The functional form of the Steinhart-Hart equation which describes the electrical characteristics of NTC thermistors, is studied particularly with regard to its metrological properties. This equation is almost exclusively used in the community and it is even hard-wired in many commercial instruments. However this equation (in its original form) does not follow the rules of quantity calculus and it is not unit invariant. Therefore the naïve use is error prone. In this contribution proper metrological concepts to avoid pitfalls are presented.

1 Introduction

Dimensional metrology without precision thermometry is simply unthinkable. Already the development of the very first scientific temperature scale (by Chappuis at the BIPM in 1884) was driven by the need to use metre prototypes to disseminate the unit of length. And still nowadays the measurement uncertainty in dimensional metrology is very often dominated by temperature effects.

For high precision dimensional metrology a few temperature measuring techniques are well established: Industrial resistance thermometers (IPRTs), thermocouples together with a reference thermometer and thermistors (mostly of the negative temperature coefficient type, denoted by NTC).

Out of those, NTCs have some advantages for the use in dimensional metrology – they can be manufactured in small size, have tenfold sensitivity as compared to IPRTs, self heating effects can be controlled, lead compensation is usually unnecessary and the demands on the reading instruments are somewhat relaxed. In addition the long term stability can be quite good. A typical sensor of this kind is shown in Figure 1.

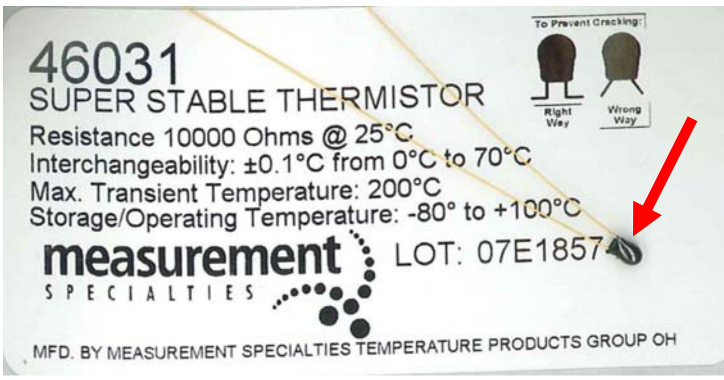

Figure 1: Typical NTC (arrow) as used in dimensional metrology placed on its package label. Note the strange style of the resistance specification (10000 Ω at 25 °C)

The electrical characteristic of a thermistor is very nonlinear, even for small temperature ranges. Different equations to model this characteristic were recommended over the years, but the so-called Steinhart-Hart equation evolved as a de facto standard in this field. This empirical three-parameter equation was published by Steinhart and Hart in a (for dimensional metrologist somewhat obscure) journal in 1968 [1]. This equation works fine in a limited temperature range and it is even implemented in a variety of commercial instruments. However this very equation violates quantity calculus and should therefore not be used to describe a physical phenomenon. Although this fact was pointed out already 3 years later (in the same obscure journal) [2] it is commonly ignored in literature, standards and by users. Beside of being a misuse in the eye of a metrologist, this habit is also error prone. Even publications of the highest metrological authority [3] do not state this fact explicitly. By simply expressing the resistance in kΩ instead of Ω, for example, the Steinhart-Hart equation fails to describe the characteristics of the thermistor in a proper way. Depending on the actual circumstances the numerical errors introduced are either negligible or can cause subtle systematic errors which might be difficult to detect.

It will be shown that the success and widespread use of this empirical equation bases in a fortunate coincident (interplay) between the numerical value of the unit of resistance (the ohm) and the intrinsic properties of the conduction processes in semiconductors. This fact seems to have been overlooked in the 40 years history on this subject, probably because of a sloppy use of numerical value equations versus quantity equations. Considering the confusing misuse of metrological concepts within one single equation it is not surprising that it has been even used as a model equation for evaluating measurement uncertainty [4].

The paper is organized as follows. In section 2 the Steinhart-Hart equation and its derivation will be revisited. Section 3 deals with the quantity calculus properties of this equation while in section 4 its scale transformation properties will be discussed in detail. Some numerical examples on the scale transformation are presented in section 5. Finally conclusions and recommendations for a proper formulation are given in section 6 which may be used in future standardization work, reviews and textbooks on this subject.

2 The Steinhart-Hart Equation

This equation is used to convert resistance reading of a thermistor to thermodynamic temperature. It contains three parameters which have to be determined empirically, e.g. by a fit procedure. These parameters do not have a physical meaning.

$$\frac{1}{T} = b_0 + b_1 \ln R + b_3 \ln^3 R \tag{1}$$

The formula was derived using fundamental principles of solid state physics. NTCs are polycrystalline semiconductor devices; the resistance should follow the classical temperature activated form.

$$\frac{1}{R} = F \exp\left(-\frac{\Delta E}{2kT}\right) \quad \text{thus} \quad \frac{1}{T} = b_0 + b_1 \ln R \tag{2}$$

Even for intrinsic semiconductors F is itself temperature dependent. Moreover (2) is only an approximation since grain boundaries and doping can change the simple behavior. In such cases physicists are tempted to apply a series expansion. One obvious form of such a expansion is shown in equation (3).

$$\frac{1}{T} = \sum_{i=0}^{n} b_i \ln^i R \tag{3}$$

The authors of [1] found that, at least for their application, there is no loss in precision by terminating the expansion after $n = 3$. Moreover b_2 proved to be statistically insignificant. So the authors recommend (1) which is a special case of the more general equation (3) with $n = 3$ and $b_2 = 0$. This fact was later on confirmed by other authors [5, 6, 7] and (1) eventually found its way to literature, data sheets, and metrological practice. Nowadays it is often implemented in the software of industrial temperature measurement systems. In this paper the term “Steinhart-Hart equation” is synonymous with (1).

3 Quantity calculus

R is a physical quantity but the logarithmic terms on the right hand side of (1) and (3) are not. With other words it is not a quantity equation in the sense of the VIM [8]. If at all, this has been fixed in different ways in the past.

3.1 Pseudo units

One can formally introduce “pseudo units” for the b_i. For example:

$$[b_3] = \ln^{-3}\Omega \cdot \mathrm{K}^{-1} \tag{4}$$

This procedure is recommended in [9], thus making (1) a quantity equation of make-believe.

3.2 Numerical value equation

One can explicitly (or more often implicitly) assume that the symbol R represents the numerical value rather than a quantity when using specified units. In the terminology of the VIM:

$$R \equiv \{R\}_{\Omega} \tag{5}$$

This notation is error prone since the unit used does not display in the equation. It is the way the authors of [1] have interpreted their equation.

3.3 Normalizing to a reference value

By normalizing the resistance R to some reference resistance R_0 one obtains a proper quantity equation.

$$\frac{1}{T} = b_0 + b_1 \ln\left(\frac{R}{R_0}\right) + b_3 \ln^3\left(\frac{R}{R_0}\right) \tag{6}$$

The normalized quantities are of the unit one therefore there are no problems with the logarithms. Often the reference R_0 is set (implicitly) to 1 Ω, in which case the original Steinhart-Hart equation is reproduced. But as long as R_0 does not appear explicitly in the equation this solution is equally error prone as the former one.

No matter which way is used to make (1) a proper quantity equation, this does not guaranty it will also be a viable sensor model. This fact will be analyzed in the next section.

4 Scale transformation properties

More important to the user is however the following point since it can lead to considerable measurement errors. Even if (1) describes the behavior of a given thermistor using a given resistance unit (say ohm) perfectly well, it will no longer do so when choosing any other unit (e.g. kiloohm). This rather trivial but sometimes startling fact can be easily verified by applying a conversion factor k between units (scale transformation) and equating coefficients. Primed symbols correspond to the transformed system.

$$b_0' = b_0 + b_1 \ln k + b_2 \ln^2 k + b_3 \ln^3 k \tag{7}$$

$$b_1' = b_1 + 2b_2 \ln k + 3b_3 \ln^2 k \tag{8}$$

$$b_2' = b_2 + 3b_3 \ln k \tag{9}$$

$$b_3' = b_3 \tag{10}$$

From equation (9) it is obvious that for $b_2 = 0$ and for any choice of $k \neq 1$ the squared term no longer vanishes in the primed system, i.e. $b_2' \neq 0$. Also the reverse holds: for any equation (3) there is exactly one transformation where b_2' vanishes, in this system the Steinhart-Hart equation describes the thermistor as well as the non-primed equation. Therefore one can wonder why (1) proved to be so successful.

It should be noted that a scale transformation can be described via equation (6) by the parameter R_0 also. The algebraic derivation can by found in [2].

It's remarkable to find that actual thermistor data is tacitly given in Ω, even accepting insignificant digits for high numerical values. This is true for publications, user manuals, normative documents and data sheets (see for example Figure 1). Also in their original work Steinhart and Hart state observed resistance data in a form like 10 558 700 Ω [1].

5 Numeric results

Taking everything into account it is tempting to re-evaluate published data (given in ohm) using different measurement units for R. Three different data sets were chosen for this test. Calibration data designated "S4" from [1] with 17 T-R pairs in the range 0 °C to 35 °C, "BGS" also from [1] with 21 T-R pairs in the range –68 °C to 135 °C. The third set consists of 5 points in the range –40 °C to 80 °C these are synthetic data provided to validate fit algorithms for the b_i [9]. This set will be designated "DKD" in the following.

In figure 2 the effect of a unit change from Ω to kΩ is shown for the S4 data. Equation (1) is fitted to both, the original and the transformed data and the residuals are plotted as a function of temperature. For the original data given in ohm equation (1) is certainly a proper sensor model (this is exactly the message of [1]). But by simply changing the resistance unit to kΩ the residuals grow and show moreover a systematic deviation.

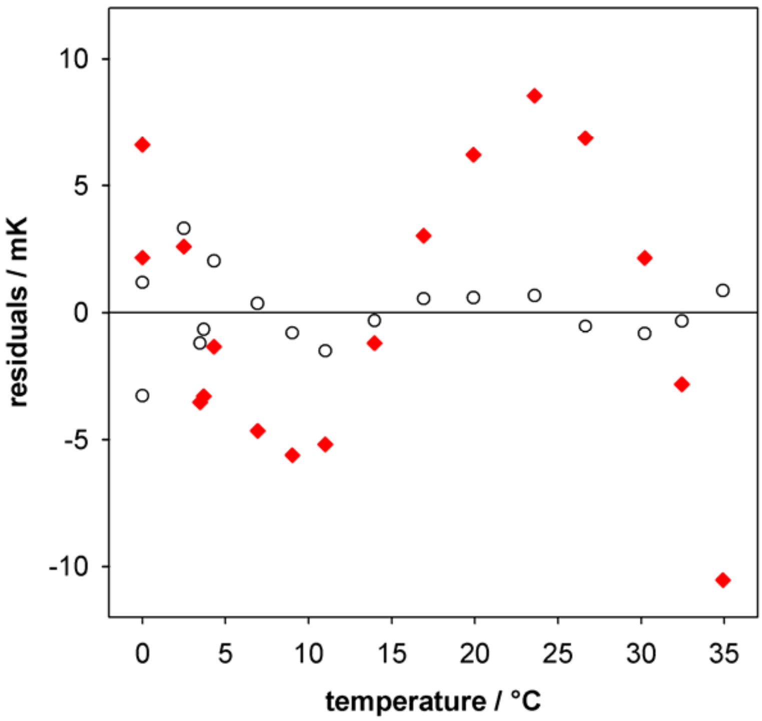

Figure 2: Residuals of fitting (1) to the S4 data. Open circles are for the original numerical data (R expressed in Ω), filled diamonds for residuals when R expressed in kΩ*.*

To study this behavior in more detail, equation (1) was fitted for the three data sets after changing the resistance unit between µΩ and MΩ. Since significant systematic deviations are expected, the span of the residuals (as opposed to the empirical standard deviation) was taken as a measure for the goodness of fit. The results are presented in Figure 3.

For a given set of data points the range of residuals can be quite large, up to two orders of magnitude in the presented data. The specific shape of the shown curves depend on many peculiarities, like range, number of points, scatter etc. At least for the data examined the overall shape shows a shallow minimum somewhere in the range corresponding to resistance units between dΩ and hΩ. This seems to be true also for otherthermistors [2]. A pronounced maximum ("bad fit") will be obtained by using a resistance unit which corresponds to a value of kΩ to 0.1 MΩ (lower numerical quantity value of R). The effect is less prominent when a smaller temperature range is considered or when the data to be fitted shows a high scatter.

The tremendous impact of temperature errors like in Figure 3 on dimensional measurements is obvious. The situation is even worse since fit residuals are of systematic nature.

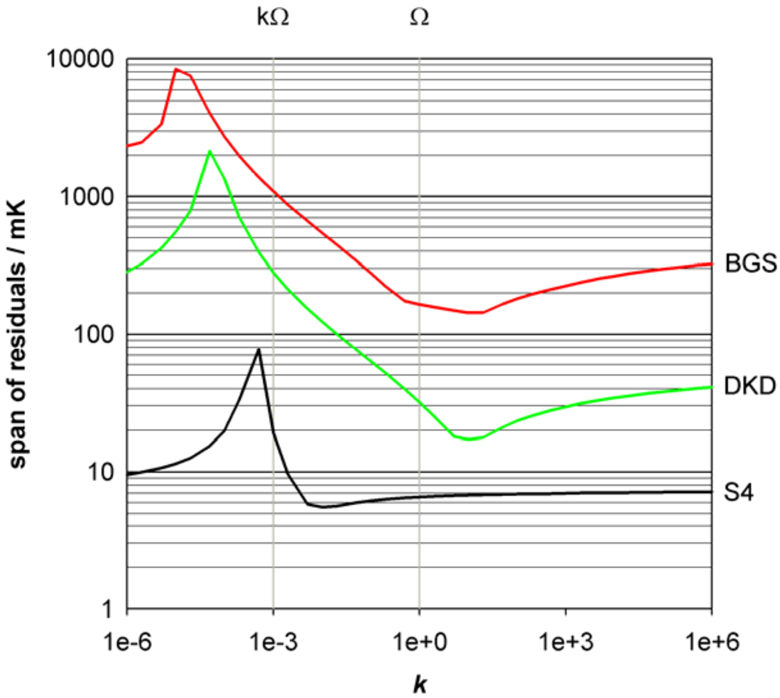

Figure 3: Span of residuals as a function of the conversion factor k for the three data sets analyzed. The two k-values corresponding to Ω *and* kΩ *are indicated by vertical lines.*

6 Conclusion

Steinhart and Hart probably would never have recommended the equation of the same name if for some historical coincidence the value of the resistance unit would have been 10^3 to 10^6 of the today Ohm. But since we have the Ohm as the resistance unit, we have to live with (1) and should at least use proper metrological concepts to avoid pitfalls when using the Steinhart-Hart equation.

The symbol R in the original work is actually a numerical quantity value for a given resistance unit. Moreover the design of (1) only works reasonably well for a limited numerical range of this resistance unit (around 1 Ω). Both facts should be addressed explicitly in future applications.

Formally this may be done by substituting the symbol R by $R/(1\ \Omega)$ (preferred) or $\{R\}_\Omega$ (unusual). But since also with this conventional choice of units (1) is only nearly optimal, one may choose (6). The actual value of this R_0 is now a parameter to be fitted and hence problem and sensor specific. It should not be confused with the use of a fixed reference value to obtain more convenient numbers, as recommended in some manufacturers' data sheets.

It follows that the only prudent way to describe the temperature-resistance characteristic is in general to use the complete form (3) with R used as described in section 3. It should be noted here that some of the literature already recommends the inclusion of the squared term but mainly for improving the fit in some circumstances. In this paper it is recommended for principal reasons even when the squared term proves to be insignificant in the actual application.

Acknowledgement

The author is gratefully indebted to Mrs. Eva-Maria Bachmann for her help in carrying out some of the calculations. The author would also like to thank Dr Peter Rosenkranz for discussions and his assistance in providing references.

References

[1] Steinhart J S and Hart S R 1968 Calibration curves for thermistors *Deep-Sea Research* **15** 497-503

[2] Bennett A S 1972 The calibration of thermistors over the temperature range 0 – 30 °C *Deep-Sea Research* **19** 157-163

[3] Bedford R E, Bonnier G, Maas H and Pavese F 1997 Techniques for Approximating the International Temperature scale of 1990 *BIPM monograph*

[4] Ospina J and Canuto E 2008 Uncertainty on differential measurements and its reduction using the calibration by comparison method *Metrologia* **45** 389-394

[5] Stephenson R J et al. 1999 Temperature Measurement, *CRC 1999*

[6] Sapoff M, Siwek W R, Johnson H C, Slepian J and Weber S 1982 The Exactness of Fit of Resistance-temperature Data of Thermistors with Third-degree Polynomials *Temperature, Its Measurement and Control in Science and Industry (American Institute of Physics, New York)* **5** 875-887

[7] Siwek W R, Sapoff M, Goldberg A, Johnson H C, Botting M, Lonsdorf R and Weber S 1992 A Precision Temperature Standard Based on the Exactness of Fit of Thermistor Resistance-Temperature Data Using Third Degree Polynomials *Temperature, Its Measurement and Control in Science and Industry (American Institute of Physics, New York)* **6** 491-496

[8] BIPM Joint Committee for Guides in Metrology (JCGM) 2008 International Vocabulary of Metrology – Basic and General Concepts and Associated Terms, VIM, 3rd edition, *JCGM* **200:2008**

[9] Deutscher Kalibrierdienst 2008 Bestimmung von Thermometerkennlinien *DKD-R* **5-6**