FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

FREQUENCY-TEMPERATURE CONSIDERATIONS FOR DIGITAL TEMPERATURE COMPENSATION

Raymond L. Filler
US Army Electronics Technology & Devices Laboratory (LABCOM)
SLCET-EQ
Fort Monmouth, NJ 07703-5601

Abstract

A fundamental requirement for compensation of the frequency-temperature (f-T) characteristic of a resonator is that the scheme used to estimate the frequency of the resonator from the temperature measurement be able to provide the necessary precision. The difference between the measured f-T data and the estimate of the frequency is a major limiting factor in any compensation scheme. If the measured f-T characteristic were sufficiently modeled by a simple function, hysteresis would be the only residual. Unfortunately, experimental results show that, in general, a simple function, for example, a polynomial of reasonable degree such as 10th, will accomplish compensation to only about $5 \times 10^{-8}$.

In this paper a description will be given of the anomalous behavior of the f-T characteristics of quartz resonators at the pp10$^6$ level which necessitates a more sophisticated estimation scheme. A review will be presented of estimation schemes, including least squares regression, minmax, segmented polynomials, and interpolation algorithms such as spline and Akima.

Introduction

The mathematical model for the frequency-temperature characteristic (f-T) of quartz crystal resonators has historically been a cubic polynomial.[1] The cubic model has been satisfactory for temperature compensating oscillators to about 1 ppm. The Microcomputer Compensated Crystal Oscillator (MCCO) using the dual-harmonic mode self temperature sensing method [2] requires a model capable of accuracy in the ppb range. Conventional temperature compensated crystal oscillators use a temperature sensing element which is remote from the actual vibrating quartz resonator. If the temperature is changing, the different thermal time constants produce a temperature gradient between the temperature sensing element and the vibrating quartz which results in "apparent" thermal hysteresis.[3] This "apparent" hysteresis has been one of the limiting factors in temperature compensation schemes. The dual-mode temperature sensing method eliminates "apparent" hysteresis and opens the way for much better temperature compensation. The cubic model is not adequate for realizing the higher accuracies. The techniques described below are useful for any digitally temperature compensated oscillator although one employing the dual-mode temperature sensing technique will make the best use of the results.

Algorithms

In order to perform digital compensation, an algorithm is required which will approximate the actual oscillator frequency from the measured temperature. For the dual-mode temperature sensing technique the temperature is represented by $f_p$, where

$$f_p = 3f_1 - f_3$$  \hspace{1cm} (1)

and $f_1$ and $f_3$ are the fundamental and third overtone frequencies, respectively. The frequency estimated by the algorithm will be referred to as $\tilde{f}(f_p)$. In a digitally compensated oscillator, the difference between $\tilde{f}(f_p)$ and the desired frequency is calculated and used by the compensation electronics to operate on the oscillator output to remove the effect of temperature.

The residuals, $R_i$, of an algorithm are the differences between the measured frequency and the estimated frequency at each measured temperature. That is

$$R_i = f(f_p) - \tilde{f}(f_p); \ i = 1, ..., N$$  \hspace{1cm} (2)

where N is the number of calibration data points. The measure of quality of a temperature compensation algorithm is the maximum residual because the
definition of 1-T accuracy is the maximum deviation from the specified nominal frequency.[4] It is assumed that the compensation electronics can eliminate any offset so that a zero residual corresponds to an output frequency exactly at nominal.

Two general classes of algorithms are examined in this paper. The first is multiple regression with integer powers of \( f_0 \) as the basis functions, i.e., a polynomial fit. The familiar Method of Least Squares is an example of a multiple regression algorithm. The second class of algorithms is interpolation. For interpolation a function is found which passes through each data point while conforming to some continuity rules. Lagrange's polynomial formula and spline functions are examples of interpolation algorithms.[5]

Regression

To perform regression, a distance function with adjustable parameters is defined and then minimized with respect to those parameters.[6] The distance function used in this paper is the sum of the N residuals, \( R_i \), raised to the exponent \( p \). A weight \( w_i \) is included to increase or decrease influence of selected data points. The general distance function, \( L_p \), is

\[
L_p = \sum_{i=1}^{N} w_i |f_i - \hat{f}(A, f_0)|^p
\]  

(3)

If \( p \) is chosen to be 2, the algorithm is the commonly used Method of Least Squares. If \( p \) is chosen to be \( \infty \), the distance function is dominated by the maximum residual and the algorithm is a minimum maximum or a "minmax" algorithm.

To see that when \( p=\infty \) the only term which survives is the maximum residual, consider the normalized distance function where each term is divided by the maximum residual. All of the terms, except the one involving the maximum residual, are a fraction raised to the power of infinity, i.e., each term is vanishingly small except the one involving the maximum residual, which is equal to 1.

Method of Least Squares

The most familiar least squares algorithm is polynomial regression. It uses the integer powers of the independent variable as the basis functions. The trial frequency function is

\[
\hat{f}(f_0) = \sum_{k=0}^{D} A_k (f_0)^k
\]  

(4)

where the \( D+1 \) coefficients are contained in the vector \( A \), and \( D \), the highest exponent in the polynomial is the degree of the polynomial. The vector \( A \) containing the \( D+1 \) coefficients is determined by minimizing the distance function

\[
L_2 = \sum_{i=1}^{N} w_i |f_i - \hat{f}(A, f_0)|^2
\]  

(5)

with respect to those \( D+1 \) coefficients. The RMS (or average) error of all of the residuals is minimized in Least Squares Regression.

This minimization problem is linear in the coefficients and reduces to a set of linear simultaneous equations which has a closed form solution.[5] An interesting property of a least squares fit, (only strictly true if a constant term is included) is that the algebraic sum of the residuals is identically equal to zero.

Fig. 1 shows the residuals of the least squares fit using a 10th degree polynomial on a 10MHz, 3rd O/T, SC-cut resonator.

\[
\hat{f}(f_0) = \sum_{k=0}^{D} A_k (f_0)^k
\]  

(4)

where the \( D+1 \) coefficients are contained in the vector \( A \), and \( D \), the highest exponent in the polynomial is the degree of the polynomial. The vector \( A \) containing the \( D+1 \) coefficients is determined by minimizing the distance function

\[
L_2 = \sum_{i=1}^{N} w_i |f_i - \hat{f}(A, f_0)|^2
\]  

(5)

with respect to those \( D+1 \) coefficients. The RMS (or average) error of all of the residuals is minimized in Least Squares Regression.

This minimization problem is linear in the coefficients and reduces to a set of linear simultaneous equations which has a closed form solution.[5] An interesting property of a least squares fit, (only strictly true if a constant term is included) is that the algebraic sum of the residuals is identically equal to zero.

Fig. 1 shows the residuals of the least squares fit using a 10th degree polynomial on a 10MHz, 3rd O/T, SC-cut resonator.
task of designing an algorithm for temperature compensation would be simplified if these anomalies were smaller.

**Minmax**

A more appropriate distance function for the temperature compensation problem is one which conforms to the definition of stability, that is one which minimizes the maximum deviation or "minmax" criterion. The distance function to be minimized reduces to the minimization of

\[
L = \text{Max} \left\{ \left| R(f_h) - \hat{f}(A, f_h) \right| \right\}
\]  

(6)

with respect to \( A \). There is no closed form solution to this problem.[6]

**Lawson's Algorithm:** A technique known as Lawson's Algorithm uses the linear least squares technique described above with appropriate weights, \( w \), in a recursive manner to converge to the minmax fit.[7] Lawson's approach is to use as the weights of the \((j+1)\)th least square recursion the normalized residuals from the \(j\)th recursion. That is

\[
w^{j}_{i+1} = \frac{w^j_i |R^j_i|}{\sum_{i} w^j_i |R^j_i|}
\]  

(7)

When this method was applied to actual f-T data, the maximum residual did not always decrease monotonically, but after about 3 to 5 iterations a satisfactory minimum maximum was obtained.

Fig. 2 is a comparison of the least square fit from Fig. 1 to the fit using Lawson's Algorithm. The reduction of the maximum residual is about 30%. The penalty incurred is that the Lawson residuals are larger almost everywhere but at the maximum. This is a consequence of the fact that the Least Squares Method produces the minimum RMS residual and any other algorithm must have the same or larger RMS error.

**Direct Minmax:** To test the Lawson Algorithm, and to see if there was a more convenient algorithm to find the minmax coefficients, a direct search was made of \( A \) space using a simplex [5] routine with the magnitude of the maximum residual as the value of the function to be minimized. This approach only reduced the maximum residual a few percent and was exceedingly time consuming.

**Interpolation**

In interpolation the task is to find a function which passes through every data point. There is no smoothing effect since every point has equal weight. Outliers or sharp features can produce unexpected results. The interpolation function can be a single polynomial of order \( N \) (degree \( N-1 \)), where \( N \) is the number of data points, or the function can be piecewise continuous polynomials of lower degree. The continuity conditions and the degree of the polynomials define the different interpolation algorithms. The minimum continuity condition is that the function is continuous at each junction. Higher order continuity conditions consist of forcing the slope, or the slope and higher order derivatives to be continuous.

Interpolation algorithms fall into two broad classes, "local" and "global", depending on the relative influence of data far from the region where the interpolation is being applied. The more restricted the influence of remote data, the more "local" is the algorithm.

**Linear Interpolation**

The simplest interpolation scheme is to use straight lines between successive data points. This technique is not satisfactory because, as can be seen from Fig. 3, all of the residuals have the same sign. There will be no reduction of the accumulated error due to self-cancellation.
Spline Interpolation

A very popular interpolation algorithm is the spline. A spline is a polynomial between neighboring pairs of points, with continuous higher order derivatives. A cubic spline is continuous through the second derivative. A cubic polynomial has four coefficients. The condition of continuous second derivative allows computation of the first derivative at each data point. The two first derivatives and the values of the two data points are enough information to determine the four coefficients of each cubic interpolating function. A spline uses all of the data to determine the coefficients, so it is, therefore, a "global" algorithm.

Akima Interpolation

The problem with global interpolation algorithms is unwarranted undulations in the resulting curve. Fig. 5 shows a spline function applied to a constructed f-T anomaly.

It can be seen that there are unwarranted undulations in the resulting curve. Akima [9] devised an algorithm which uses 5 nearest-neighbor points to eliminate the unwarranted undulations. Akima defined the slope at each data point as

$$ t_i = \frac{m_{i-1} - m_i}{|m_{i-1} - m_i| + |m_i - m_{i+1}|} $$

with the special case of

$$ t_i = \frac{m_{i-1} + m_i}{2} $$

when $m_{i-1} = m_i = m_{i+1}$, where $m_i$ is the slope of the straight line from $x_{i-1}$ to $x_{i+1}$. The cubic polynomial defined for the interval $x_i \leq x \leq x_{i+1}$ is

$$ f(x) = \frac{x - x_i}{x_{i+1} - x_i} (a_{i+1} (x - x_i)^3 + a_i (x - x_{i+1})(x_{i+1} - x_i)) + \frac{x_{i+1} - x}{x_{i+1} - x_i} (a_{i+1} (x_{i+1} - x)^3 + a_{i+1} (x_{i+1} - x_i)(x - x_i)) $$

where $a_i = m_i$ and $a_{i+1} = m_{i+1}$.
Fig. 6 shows that the Akima algorithm applied to the same set of data as in Fig. 5 eliminated the undulations.

\[
y = \rho_0 + \rho_1 (x - x_j) + \rho_2 (x - x_j)^2 + \rho_3 (x - x_j)^3
\]

where

\[
\rho_0 = y_j, \\
\rho_1 = \frac{t_j}{x_j - x_{j-1}}, \\
\rho_2 = \frac{3(y_{j-1} - y_j)(x_{j-1} - x_j) - 2t_j - t_{j-1}}{x_{j-1} - x_j}, \\
\rho_3 = \frac{t_j + t_{j-1} - 2(y_{j-1} - y_j)(x_{j-1} - x_j)}{(x_{j-1} - x_j)^2}
\]

Fig. 6 shows that the Akima algorithm applied to the same set of data as in Fig. 5 eliminated the undulations.

Local anomalies in the f-T characteristic of some devices require a local interpolation scheme. The Akima algorithm is the best of the ones investigated.

Other Interpolations Algorithms

Other interpolations algorithms investigated were 1) the osculatory algorithm[9] which uses three nearest neighbors and 2) a simple sliding parabola with no continuity conditions imposed. These perform almost as well as the Akima algorithm.

Interpolation Results

Fig. 7 shows the results of the Akima interpolation on the same data as was used in Figs. 1 and 2. It can be seen that the maximum residual is significantly lower than in Fig. 2.

Regression with Multiple Segments

An oscillator which uses interpolation must store all of the calibration data in an ordered table, therefore, the memory requirements are larger than in a device which just stores regression coefficients. A technique which gives the superior results of interpolation with the ease of use of regression would be advantageous.

A compromise between interpolation and regression is to break up the temperature interval into segments and do a regression in each segment. This is similar to using a higher order polynomial but is more local and does not suffer from roundoff errors in the calculation. The calculation used in does not impose any continuity conditions at the breakpoints. This is not a problem as long as the residuals near the breakpoints are well within the required stability. Fig. 8 shows a three segment, 10th degree polynomial regression to the data used above. It can be seen that the residuals are similar to those of Fig. 7.

Fig. 9 shows the Akima interpolation and the three segment regression superimposed to show the similarity. An additional degree of freedom, which was not exercised in this example, is to vary the location of the breakpoints.

Segment No. vs. Degree vs. Algorithm

If the Akima algorithm is taken as the best maximum residual which can be obtained for a given f-T characteristic we can tabulate the degradation in the
Conclusions

There are many choices for f-T compensation algorithm which are superior to the usual least squares fit. These algorithms can be used with any digital compensation technique. The multisegment Lawson algorithm is the most efficient choice but any of the interpolation schemes will give optimum results at a cost of requiring more memory.

Smooth f-T characteristics and a good algorithm make it possible to achieve the limit set by the hysteresis of a device. Lower hysteresis is required to get the ultimate performance from digital temperature compensation.

References


Figure 10 - Number of segments vs. Degree of Polynomial vs. Algorithm Relative to Akima Interpolation.