Scientific
Research
Publishing

Table of Contents

Volume 7   Number 2   February 2015

The figure on the front cover is from the article published in Engineering, 2015, Vol. 7, No. 2, pp. 47-57 by Tetsuya Akitsu, Akira Kudo and Tomio Sato.

Engineering (ENG)

Journal Information

SUBSCRIPTIONS

The *Engineering* (Online at Scientific Research Publishing, www.SciRP.org) is published monthly by Scientific Research Publishing, Inc., USA.

SERVICES

Advertisements
Advertisement Sales Department, E-mail: service@scirp.org

Reprints (minimum quantity 100 copies)
Reprints Co-ordinator, Scientific Research Publishing, Inc., USA.
E-mail: sub@scirp.org

COPYRIGHT

PRODUCTION INFORMATION

For manuscripts that have been accepted for publication, please contact:
E-mail: eng@scirp.org

Scientific
Research
Publishing

Compatibility of the Active Inductance Double Resonance Quartz Oscillator with Q-MEMS Temperature Sensor

Tetsuya Akitsu[1*], Akira Kudo[2], Tomio Sato[1]

[1]Human, Environment and Medical Engineering, Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, Takeda 4-3-11, Kofu, Japan
[2]SEIKO EPSON Corporation, Head-Quarter Owa 3-3-5, Suwa, Japan
Email: [*]akitsu@yamanashi.ac.jp

Abstract

Low-frequency double-resonance quartz crystal oscillator circuit was developed with active inductance aiming the quick start-up in the intermittent operation on the sensor circuit and DC isolation using a Q-MEMS sensing crystal HTS-206. Allan standard deviation indicated 5×10^{-12}, showing short range stability of the sensor circuit sufficient for the ubiquitous environmental sensor network.

1. Introduction

Environmental sensing awaits solutions to reduce the electric-power in monitoring under the limitation of the power source. The ubiquitous sensor network is realized with varieties of sensor circuit and a wireless network. The temperature measurement in the environmental sensing is realized by several methods: thermistors, platinum wire or sheet resistor and semiconductor sensor devices. The temperature is measured as the amplitude of low level DC or modulated signals and faces the difficulty which arises from the drift. Q-MEMS crystal temperature sensor can realize high resolution in the sensing of environmental temperature. Direct digital temperature measurements have been developed by several research groups, usint crystal cut LT, SC-cut or equivalent cut [1]-[7]. In a recent work we reported quick start of quartz crystal oscillator. The quick start of the crystal sensor

---

[*]Corresponding author.

circuit allows intermittent excitation of the sensor system meeting the request for the power management in the environmental sensing. In this work, we aimed several engineering issues: 1) Quick start of low frequency quartz crystal oscillator circuit in the intermittent operation; 2) Reduction of drift by the direct digital measurement; 3) DC isolation between the sensor and the electronic circuit. Tuning fork type Q-MEMS quartz crystal sensor realizes the direct digital sensing of temperature, where Q-MEMS is a combination word of Quartz and MEMS (Micro Electro Mechanical System), a fabrication process offering high performance in a compact package, with *CI* values as low as those on ordinary-sized crystals. In recent works, double-resonance quartz crystal oscillator was reported for the enhancement of the frequency pulling [8], the mode separation of the multimode quartz crystal resonator [9] [10], and the start-up acceleration of low frequency quartz crystal oscillator circuit [11]. Start-up acceleration of several Mega Hertz is studied by the gain control in the quartz crystal oscillator using a cascade circuit [12]-[14]. Few works treats the acceleration of the start-up of the low-frequency quartz crystal oscillator. We aim to test the conformity of the acceleration scheme with the Q-MEMS crystal sensor. Stability of the oscillation frequency is discussed based on the moving average of the variance determined for the discrete samples following the proto call of the modified Allan standard deviation for moving average of finite length data is employed as the measuring rule of the short range stability [15]-[17].

2. Design and Analysis of Quartz Crystal Oscillator

2.1. Acceleration of the Start-Up of a Quartz Crystal Oscillator

Figure 1 shows a circuit diagram of the active inductance double resonance oscillator circuit. The initial oscillation is generated by a part this oscillator acting as a CR oscillator, and after the stable oscillation of the quartz crystal resonator current starts, the double resonance is established between the quartz crystal and an active inductance combined with the parallel capacitance resulting in the generation of negative resistance. IC_1 and IC_2 are CMOS inverter, where CMOS is Complementary Metal Oxide Semiconductor. HTS-206 Q-MEMS temperature sensing crystal is connected through coupling capacitors which are formed between metallic sheets, 20 mm square in dimension attached on both sides of a Pyrex glass plate. The value is approximately C_{16}, C_{17} = 18 pF.

Essential circuit constants R_2, C_{10}, and C_0 determines the resonance condition, where C_0 is the parallel capacitance of the quartz crystal resonator. R_2 settles the bias in the initial stage of the oscillation. C_{10} stores the ground potential at the activation of the V_{cc} voltage, inserted between the node connecting two inverters. The oscillation frequency is determined by a recharging-time constant R_2 multiplied by C_{10}. Capacitors C_2 and C_3 are load capacitors which is necessary for the generation of negative resistance. C_5 and C_6 are pass-capacitors between the bus-line and the circuit ground. C_0 and C_1 are reserved for the parallel capacitance of the resonator and the series capacitor of the motion arm. The conductance is controlled by negative feedback resistors $R_f = R_3$, R_4, R_5, and R_6.

$$g_{mf} = \frac{g_m}{1 + R_f g_m}.$$

(1)

The problem is if the active inductance can generate the negative resistance, and if the negative resistance is large enough to realize the short start-up time. Practical question is the shift of the resonance frequency of the crystal sensor by the series capacitors. **Figure 2** shows simplified equivalent circuit-1. CMOS inverter IC_1 and IC_2 is replaced by two current sources controlled by the gate voltage V_{in} and V_g.

Applying Kirchhoff's law, the relations for I_{out} and V_{in} are found. V_{in} is the input voltage of IC_1 and I_{out} is the output current of IC_2.

$$I_{out} = 2g_{mf}Z_4\left(I_5 + 2g_{mf}V_{in}\right) = 2g_{mf}V_g = 2g_{mf}Z_4I_4$$

(2)

$$V_{in} + R_2I_5 + Z_4I_4 = 0, \quad \text{where } I_4 = I_5 + 2g_{mf}V_{in}.$$

(3)

$$I_5 = -\frac{1 + 2g_{mf}Z_4}{R_2 + Z_4}V_{in}.$$

(4)

$$I_{out} = 2g_{mf}Z_4\left(-\frac{1 + 2g_{mf}Z_4}{R_2 + Z_4}\right)V_{in} + 4g_{mf}^2Z_4V_{in}.$$

(5)
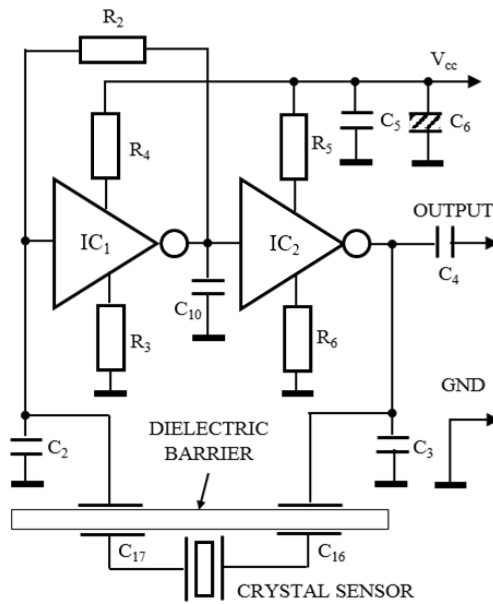
Figure 1. Circuit diagram of the quartz crystal oscillator. Circuit constants: Bias resistors R_3, R_4, R_5 and R_6 = 3.3 kΩ, feedback resistor R_2 = 3 MΩ, varied for the optimum setting; C_2, C_3 = 10 pF; C_{10} = 10 pF; C_5 = 0.1 μF; C_6 = 10 μF; C_7 = 100 pF; C_{16}, C_{17} = 18 pF. Inverter IC_1 and IC_2 TC7SHU04F; V_{cc} = 3 V. Equivalent circuit constant of the quartz crystal resonator: f_1 = 40 kHz; L_1 = 12,000 H; R_1 = 47.6 kΩ; C_1 = 1.326 fF, C_0 = 803 fF.
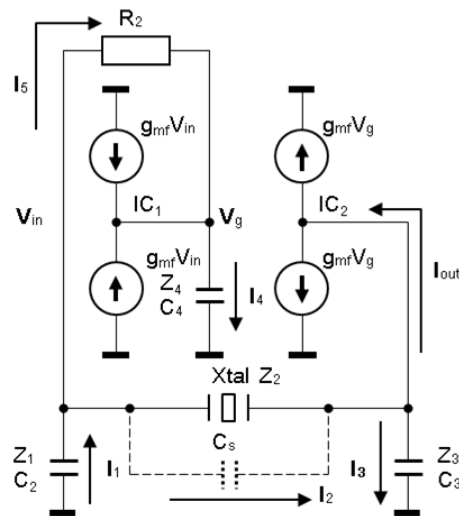
Figure 2. Simplified circuit: equivalent circuit-1.

Solving for the relation between I_{out} and V_{in}, total conductance G_M is found.

$$G_M = \frac{I_{out}}{V_{in}} = \frac{2g_{mf}\left(2g_{mf}R_2 - 1\right)}{1 + \dfrac{R_2}{Z_4}}. \tag{6}$$

Then the following relation is found. Current I_2, I_3 are expressed in the terms of I_1.

$$V_{in} = Z_1 I_1. \tag{7}$$

$$I_2 = I_1 - I_5 = \left(1 + \frac{1 + 2g_{mf}Z_4}{R_2 + Z_4} Z_1\right) I_1. \tag{8}$$

$$I_3 = I_2 - I_{out} = I_2 - G_M Z_1 I_1 = \left(1 + \frac{1 + 2g_{mf} Z_4}{R_2 + Z_4} Z_1 - G_M Z_1\right) I_1. \tag{9}$$

Rearranging the expression, relation (11) is found.

$$Z_1 + Z_3 - G_M Z_1 Z_3 = Z_c, \quad Z_c = R_c + \frac{1}{j\omega C_c}. \tag{10}$$

$$Z_2 + \frac{\left(\dfrac{R_2}{Z_4} + 1\right) Z_c + \left(\dfrac{1}{Z_4} + 2g_{mf}\right) Z_1 Z_3}{\dfrac{R_2}{Z_4} + 1 + \left(\dfrac{1}{Z_4} + 2g_{mf}\right) Z_1} = 0. \tag{11}$$

Z_2 is the impedance of a quartz crystal resonator (Z_{xt}), and impedance for other components is defined as in (12). The composed impedance Z_{cc} of the active circuit is found, substituting the impedance. From the condition for the non-zero solution of current, the oscillation condition results in (13). The impedance of the circuit is divided into resistive and reactance parts.

$$Z_2 = Z_{xt}, Z_1 = \frac{1}{j\omega C_2}, Z_3 = \frac{1}{j\omega C_3}, Z_4 = \frac{1}{j\omega C_4}, \omega_r = \frac{1}{C_4 R_2}. \tag{12}$$

$$Z_{xt} + Z_{cc} = 0 \tag{13}$$

$$Z_{cc} = R_{cc} + jX_{cc} = \frac{R_c + \dfrac{1}{\omega_r C_c} - \dfrac{2g_{mf}}{\omega^2 C_2 C_3} + j\left(\dfrac{\omega R_c}{\omega_r} - \dfrac{C_4}{\omega C_2 C_3} - \dfrac{1}{\omega C_c}\right)}{1 + \dfrac{C_4}{C_2} + j\left(\dfrac{\omega}{\omega_r} - \dfrac{2g_{mf}}{\omega C_2}\right)}. \tag{14}$$

The equivalent resistance and the reactance of the circuit are found. Equivalent inductance L_{cc} or capacitance C_{cc} is determined depending on sign of reactance X_{cc}.

$$R_{cc} = \frac{ac + bd}{a^2 + b^2}, X_{cc} = \frac{ad - bc}{a^2 + b^2}, \left(L_{cc} = \frac{X_{cc}}{\omega}, C_{cc} = -\frac{1}{\omega X_{cc}}\right). \tag{15}$$

Factors "a", "b", "c" and "d" are introduced for the simplicity of the expression, where factors "c" and "d" have the dimension of "Ω" and factors "a" and "b" are dimensionless numbers.

$$a = 1 + \frac{C_4}{C_2}, b = \frac{\omega}{\omega_r} - \frac{2g_{mf}}{\omega C_2}, c = \frac{1}{\omega_r C_c} + R_c - \frac{2g_{mf}}{\omega^2 C_2 C_3}, d = \frac{\omega R_c}{\omega_r} - \frac{C_4}{\omega C_2 C_3} - \frac{1}{\omega C_c}. \tag{16}$$

G_M is separated into real and imaginary parts.

$$G_M = g_s + jb_s. \tag{17}$$

$$g_s = \frac{2g_{mf}\left(2g_{mf} R_2 - 1\right)}{1 + \left(\dfrac{\omega}{\omega_r}\right)^2}, \quad b_s = \frac{\omega}{\omega_r} \frac{2g_{mf}\left(1 - 2g_{mf} R_2\right)}{1 + \left(\dfrac{\omega}{\omega_r}\right)^2}. \tag{18}$$

Introducing (13) and (19) into Z_{cc}, the impedance of the active circuit is found.

$$R_{cc} = \frac{g_s}{\omega^2 C_2 C_3}, \quad \frac{1}{C_{cc}} = \frac{1}{C_2} + \frac{1}{C_3} - \frac{b_s}{\omega C_2 C_3}. \tag{19}$$

Figure 3 shows simplified diagram of equivalent circuit-2 and equivalent circuit-3. The active circuit is indicated with R_{cc} and reactance C_{cc} or L_{cc} depending on the sign. The resonator consists of parallel capacitance C_0 and the motion arm, L_1, C_1, and R_1, the equivalent series inductor, capacitor, and resistor respectively. C_S is a stray capacitance. Calculating the parallel composition of C_0 and C_s with the active circuit, equivalent circuit-3
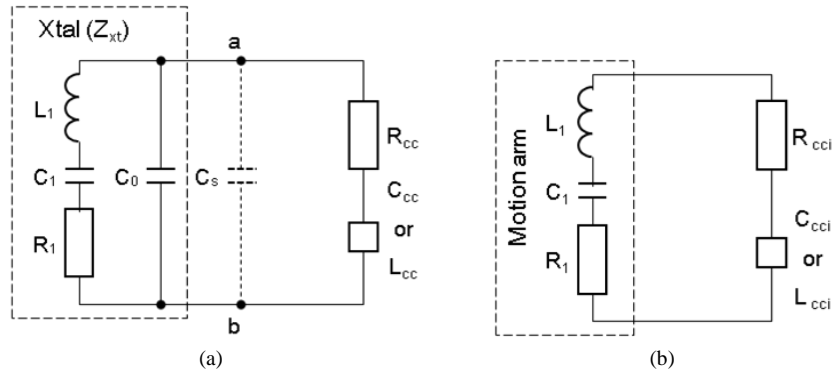
Figure 3. Simplified equivalent circuit. (a) Equivalent circuit-2; (b) Equivalent circuit-3.

is found. Composed equivalent resistance R_{cci} and capacitance C_{cci} are found.

$$R_{cci} = \frac{R_{cc}}{\left(1+\frac{C_{0s}}{C_{cc}}\right)^2 + \left(\omega C_{0s}R_{cc}\right)^2}, \quad \frac{1}{C_{cci}} = \frac{\frac{1}{C_{cc}}\left(1+\frac{C_{0s}}{C_{cc}}\right) + \frac{\left(\omega C_{0s}R_{cc}\right)^2}{C_{0s}}}{\left(1+\frac{C_{0s}}{C_{cc}}\right)^2 + \left(\omega C_{0s}R_{cc}\right)^2}. \tag{20}$$

Negatively signed capacitance is converted to an active inductance by relation (21).

$$L_{cci} = -\frac{1}{\omega^2 C_{cci}}. \tag{21}$$

The denominator of negative resistance R_{cci} has quadratic dependence on R_{cc}. The maximum value of the absolute value is reached at a specific value of R_{cc} determined by C_{0s} and C_{cc}. The following relation is fulfilled.

$$\left|R_{cci}\right|_{max} = \frac{1}{2\omega C_{0s}\left(1+\frac{C_{0s}}{C_{cc}}\right)} \quad at \ \left|R_{cc}\right| = \frac{1}{\omega C_{0s}}\left(1+\frac{C_{0s}}{C_{cc}}\right). \tag{22}$$

The active inductance appears in the vicinity of the resonance frequency, while capacitance C_{cc} is negative. The resonance frequency is determined by L_{cc}, C_{0S}, and the sum of C_0 and C_s. In this simplified form, the absolute value of negative resistance R_{cci} becomes infinitely large, if C_{cc} approaches $-C_{0S}$ and condition (23) is fulfilled.

$$C_{cc} = -\frac{1}{\omega^2 L_{cc}}. \tag{23}$$

At the resonance frequency determined by L_{cc} and C_{0S}, the absolute value of negative resistance determines the growth of signal. The suppression of negative resistance by inductance L_1 establishes the stability and inhibitory action against the signal growth. Temperature sensing crystal HTS-206 is a tuning-fork type resonator, 2 mm in diameter and 6 mm in length of the exterior size, produced for low power oscillation of 0.1 μW typically. **Table 1** shows the equivalent circuit constant of the quartz crystal resonator.

Temperature dependence of the crystal sensor is explained in the experimental part. **Figure 4** and **Figure 5** compare the absolute values of negative resistance R_{cc} and R_{cci} as functions of frequency. In **Figure 4**, the absolute value of negative resistance and the active inductance is compared as functions of frequency. The parameter g_{mf} is selected at 4.1 μA/V. The active inductance disappears at 55 kHz for $g_{mf} = 4$ μA/V, and 110 kHz for $g_{mf} = 8$ μA/V. The frequency limit is 40 kHz for $g_{mf} = 3$ μA/V and 110 kHz for $g_{mf} = 8$ μA/V. Larger gain is necessary for the negative resistance and the active inductance.

$$C_0 - C_{cc} = 0. \tag{24}$$

Table 1. Equivalent circuit constant of the temperature sensing crystal (HTS-206).

Resonance frequency	Equivalent circuit constant				
	L_1	C_1	R_1	C_0	Q_1
39.992508 kHz	12 kH	1.326 fF	12.6 kΩ	803.279 fF	239,149
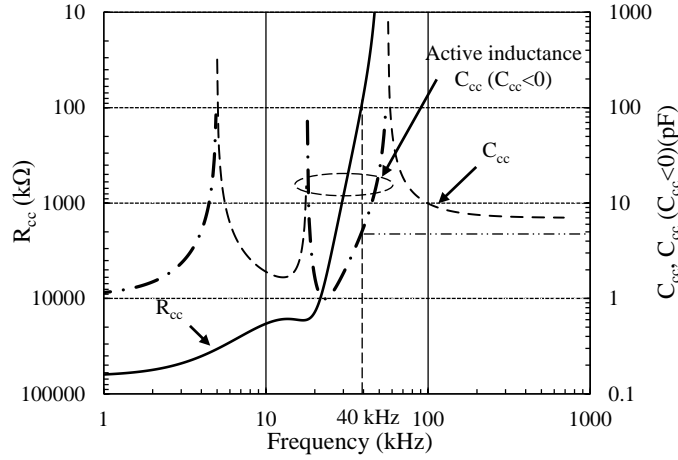
Figure 4. Absolute value of negative resistance R_{cc} and reactance C_{cc} as functions of frequency and gain. Circuit constant: C_2, C_3 = 14 pF; C_{10} = 18 pF; R_2 = 1.0 MΩ, g_{mf} = 4 μA/V.
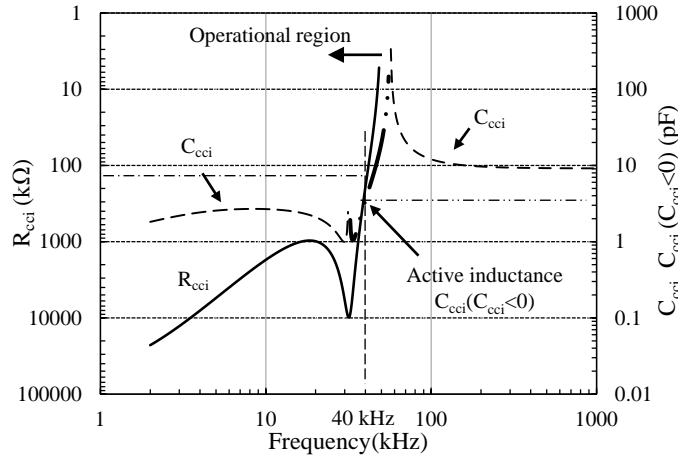
Figure 5. Absolute value of negative resistance R_{cci} and reactance as functions of frequency and g_{mf}. Circuit constant: R_2 = 1.9 MΩ; C_2, C_3 = 14 pF; C_0 = 1.14 pF; C_4 = 18 pF; C_s = 1 pF. g_{mf} = 4 μA/V.

In **Figure 5**, the absolute value of negative resistance R_{cci} and the active inductance is compared as functions of on frequency. The parameter g_{mf} = 4.1 μA/V was optimized for 32.768 kHz realizes the maximum value of negative resistance R_{cci} approximately of 150 kΩ and the reactance: C_{cci} = −3 to −4 pF.

Figure 6 shows that the active inductance is generated in narrow range of gain corresponding to the resonance frequency and higher gain is necessarily compared with the case of 32.768 kHz, for better performance. The absolute value of negative resistance ranges to R_{cc} = 2 × 10³ kΩ and the reactance is inductive C_{cc} = −0.6 pF at g_{mf} = 8 μA/V.

In this analysis, the terminal impedance at a - b is expressed with R_{cc} and R_{cci}. The parallel capacitance C_0 and stray capacitance C_s included in the impedance R_{cci}. From relation (22), R_{cci} becomes infinitely large at C_{cc} = −C_{0s}. This result must be interpreted carefully, because the optimum condition is not realized in the context of

(a)

(b)
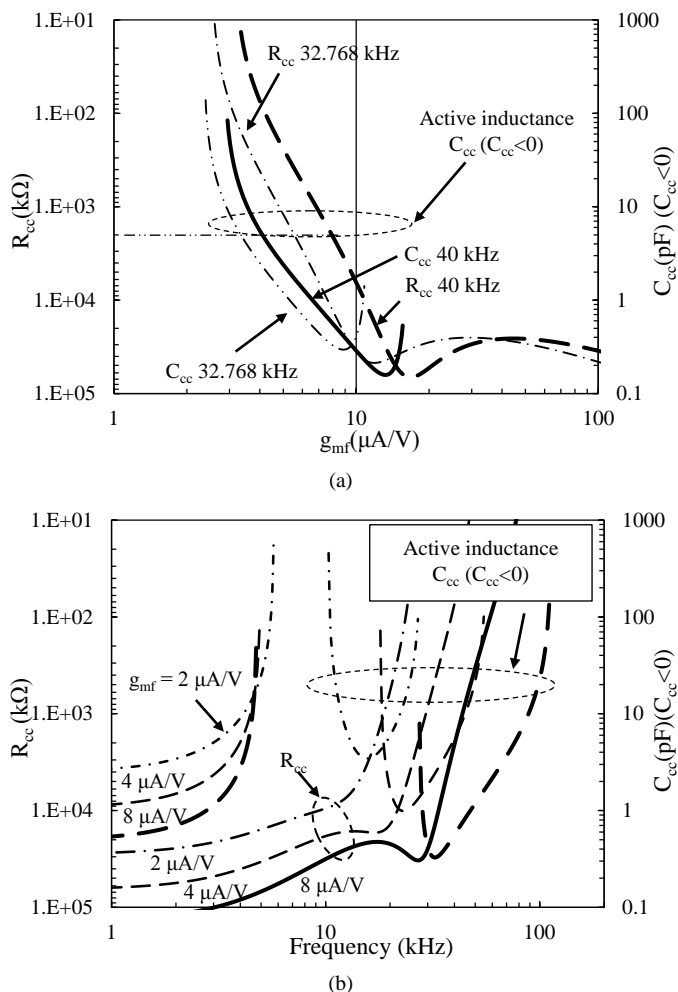
Figure 6. Comparison of negative resistance for different values of g_{mf} and f_r. (a) Dependence of negative resistance on g_{mf} for different values of f_r. Frequency f_r = 32.768 kHz and 40 kHz. Circuit constants: C_2, C_3 = 14 pF; C_4 = 18 pF; R_2 = 1.9 MΩ; (b) Frequency dependence of negative resistance for different values of g_{mf}. g_m = 2, 4, and 8 μA/V. Circuit constants: C_2, C_3 = 10 pF; C_4 = 18 pF. R_2 = 1.9 MΩ.

the actual circuit design. The idea given in this result is that the active inductance can generate large negative resistance compared to the capacitive region. Actually, R_{cc} is determined under the limitation of the circuit constants and the oscillation frequency. The strength of the oscillation is limited within the linear region of the active circuit.

The curve indicated as 32.768 kHz shows the result calculated using the equivalent circuit constant of a timebase quartz resonator analyzed in Ref. 6. Comparing the dependence on frequency and gain, larger gain is needed for the appropriate design of the active inductance and negative resistance, when the resonance frequency is higher.

2.2. Modelling of the Start-Up of Oscillation

Computer simulation was carried out using LTspice IV for Windows (Linear Technology Corporation, 1630 McCarthy Blvd., Milpitas, CA, USA) [18]. **Figure 7** shows the circuit layout for the simulation. CMOS inverter IC_1 and IC_2 are replaced with pairs of complementary MOSFETs (Metal Oxide Semiconductor Field Effect Transistor). Because original Q is too high for the stable simulation, the motional capacitance and inductance are scaled with the resonance frequency fixed. The motional components do not correspond to the values assigned
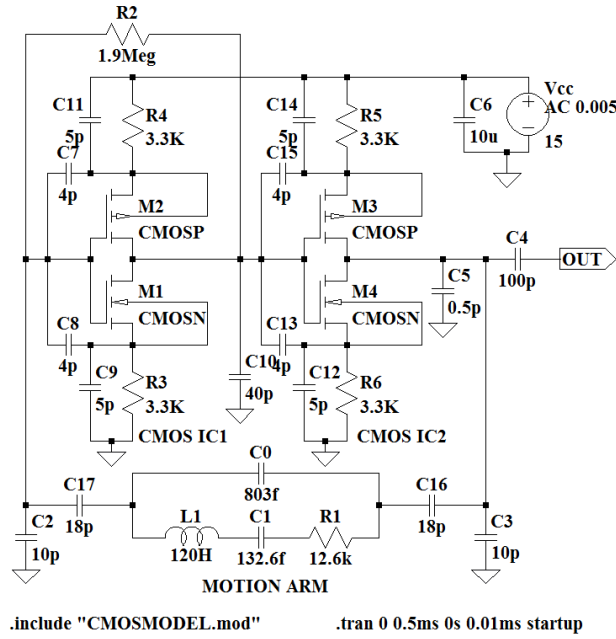
Figure 7. Equivalent circuit for the LTspice simulation. Courtesy of Linear Technology Co.

in the analysis and experiment, neither the delayed connection of the motion arm is considered. When the motion arm is removed, this circuit forms a CR oscillator. The oscillation frequency is determined by the reactance of the parallel capacitance of the quartz resonator and feedback resistor R_2. **Figure 8** shows a typical wave form of the CR oscillator and FFT spectral peak of the initial CR oscillation.

Figure 9 shows the transient excitation of the crystal current with miss-matched frequency setting of the CR oscillation. As the peak of the CR oscillation at 70 kHz is higher than the resonance frequency, phase mismatch can be observed in a few cycles of the initial oscillation. In the optimum setting, the crystal current in the motion arm grows faster, and the oscillation frequency of the oscillator circuit is locked to the resonance frequency within several oscillation cycles.

3. Experimental Result and Discussions

The stability of the stable oscillation of the double resonance oscillator is evaluated experimentally. The stability of the oscillation frequency is analyzed with 53230A universal frequency counter (Agilent Technologies, Santa Clara, Ca, USA) synchronized with external rubidium oscillator with long period stability $< 2 \times 10^{-11}$/month and short period stability $< 1 \times 10^{-11}$/s.

3.1. Stability of the Double Resonance Quartz Crystal Oscillator

Figure 10 shows the modified two-sample Allan standard deviation as a measure for the short-time frequency stability showing sufficiently high short range stability. The protocol is defined in (25), following IEEE Standard 1139. The frequency of oscillator circuit f_k is the discrete sample of oscillation frequency. τ is the gate time and k is the sequential number of samples. Dimensionless parameter is defined from frequency deviation normalized by the moving average of finite length data over 10 sequential samples [17]. As a sensor system, the frequency is measured at 100 to 1000 ms range. Allan standard deviation of 5×10^{-12} indicates sufficiently high stability when the sensor is isolated in a constant temperature vessel. The Allan standard deviation shows increase in the range from 1000 to 10,000 ms indicating the increase in the environmental drift.

$$\sigma_y(\tau) = \left[\frac{2}{M-1} \sum_{k=1}^{M-1} \left(\frac{f_{k+1} - f_k}{f_{k+1} + f_k} \right)^2 \right]^{\frac{1}{2}}. \tag{25}$$
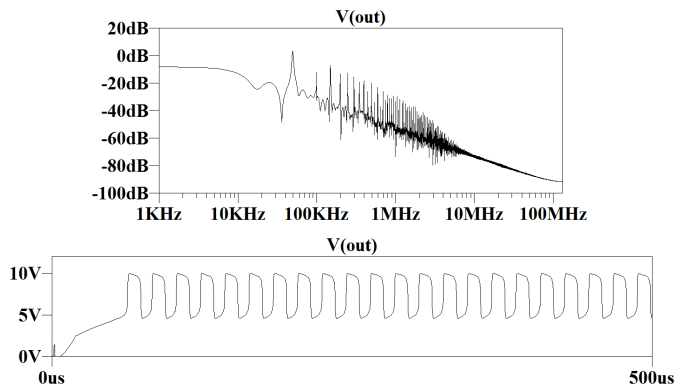
Figure 8. Waveform of the CR oscillator and the FFT spectral analysis $R_2 = 3.0$ MΩ.
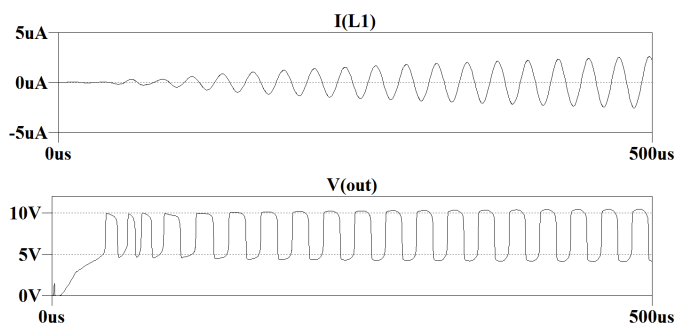
Figure 9. Growth of the crystal current and the frequency locking. Upper track: Current I(L_1) fowing through motional inductance L_1. Lower track: Output voltage. $R_2 = 1.9$ MΩ.
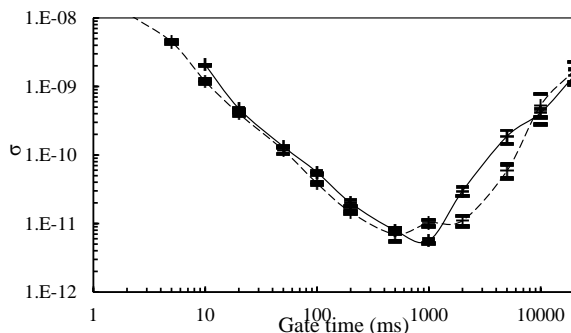
Figure 10. Allan standard deviation showing high short range stability of the crystal oscillator.

3.2. Temperature Sensing through Coupling Capacitors

Figure 11 shows the normalized frequency shift as functions of temperature.

In **Figure 11**, solid line shows the regression curve of the frequency shift in the direct connection normalized with respect to the oscillation frequency at 0 degree C. Cross symbols indicate the frequency shift when the crystal sensor is connected through coupling capacitors: C_{16}, $C_{17} = 18$ pF. Once the regression curve is determined, temperature can be calculated from the oscillation frequency. It is necessary to calibrate the temperature dependence curve must with different standard, because the temperature dependence varies depending on the value of the coupling capacitor. **Figure 12** shows the dependence of the resonance frequency on temperature measured by the impedance curve. The active circuit satisfies the oscillation condition for the variation of motional impedance CI over operating temperature.
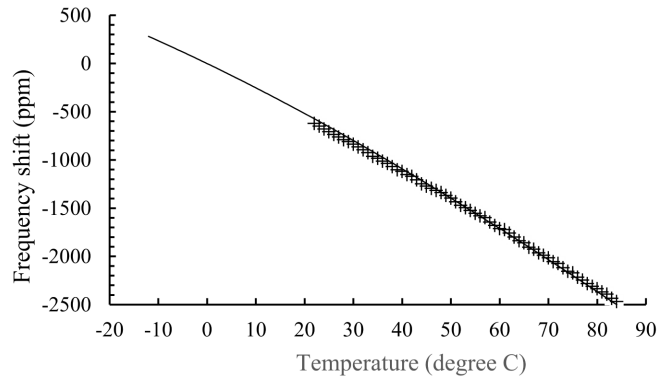
Figure 11. Temperature sensing. Solid line: Direct sensing; y = 2E−4x^3 − 0.0808x^2 − 24.37x − 0.4608. Cross symbols: Capacitively coupled sensing, C_{16}, C_{17} = 18 pF; y = −4E−05x^3 − 0.0495x^2 − 24.128x + 72.718.
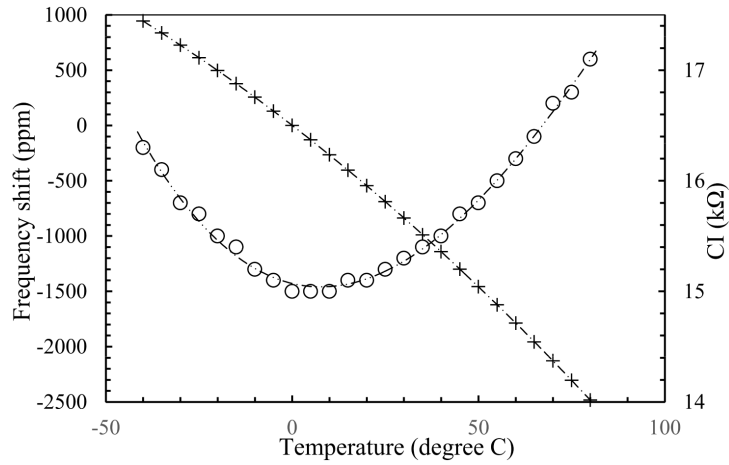
Figure 12. Temperature dependence. Cross symbols: Resonance frequency; Regression curve: y = 7E−7x^3 − 0.0623x^2 − 26.069x + 0.7698. Open circles: CI. HTS-206 sample-F10. Courtesy of SEIKO EPSON Corp.

4. Conclusion

Environmental sensing awaits solutions to reduce the electric-power in monitoring under the limitation of the power source. The quick start of the crystal sensor circuit allows intermittent excitation of the sensor system meeting the request for the power management in the environmental sensing. In this work, active inductance double resonance circuit resolved the engineering issues for the quick start-up: 1) Large negative resistance; 2) Low distortion and linearity; 3) Triggering circuit. The quartz crystal oscillator is triggered with a CR oscillator, and transferred to a stable excitation within several period. The maximum negative resistance ranges to 2 MΩ at specified gain of the active CMOS inverter circuit. The composed reactance of the active circuit negative capacitance C_{cc} = −0.6 pF. Simulation showed the rapid start-up of the oscillation by the energy transfer by the initial CR oscillation. The oscillation condition was examined by the analysis, the start-up in the computer simulation and examined by the experiment. The stability of the double-resonance oscillator showed short range stability of 5×10^{-12} which satisfied the industrial requirement for the resolution of the standard quartz crystal sensor.

Acknowledgements

References

[1] Nakazawa, M., Takemae, T., Miyahara, A. and Matsuyama, K. (1985) A Study of Quartz Temperature Sensors Characterized by Ultralinear Frequency-Temperature Responses. *IEEE Transactions on Sonics and Ultrasonics*, **32**, 828-834. http://dx.doi.org/10.1109/T-SU.1985.31672

[2] Ueda, T., Kohsaka, F., Iino, T. and Yamazaki, D. (1986) Temperature Sensor Using Quartz Tuning Fork Resonator. *The* 40*th Annual Symposium on Frequency Control*, 28-30 May 1986, Philadelphia, 224-229.

[3] Matko, V. and Safsric R. (2009) Major Improvements of Quartz Crystal Pulling Sensitivity and Linearity Using Series Reactance. *Sensors*, **9**, 8263-8270. http://dx.doi.org/10.3390/s91008263

[4] Filler, R.L. and Vig, J.R. (1993) Long-Term Aging of Oscillators. *IEEE Transactions on Ultrasonics*, *Ferroelectrics*, *and Frequency Control*, **40**, 387-394.

[5] Matko, V. (2011) Next Generation AT-Cut Quartz Crystal Sensing Devices. *Sensors*, **11**, 4474-4482. http://dx.doi.org/10.3390/s110504474

[6] Walls, F.L. and Vig J.R. (1995) Fundamental Limits on the Frequency Stabilities of Crystal Oscillators. *IEEE Transactions on Ultrasonics*, *Ferroelectrics*, *and Frequency Control*, **42**, 576-589. http://dx.doi.org/10.1109/58.393101

[7] Matko, V. and Milanovic, M. (2014) Temperature-Compensated Capacitance-Frequency Converter with High Resolution. *Sensors and Actuators A*: *Physical*, **220**, 262-269.

[8] Ehara, K., Tanaka, H., Sato, T. and Kanno, Y. (2006) Development of Inductive Quartz Crystal Oscillator Circuit with CMOS Inverter. *IEEE International Conference on Systems*, *Man*, *and Cybernetics*, Vol. 2, Taipei, 8-11 October, 1425-1430.

[9] Izyan-Ruslan, R., Sato, T. and Akitsu, T. (2012) Voltage Controlled Narrow Band Wide-Variable Range Four Segment Quartz Crystal Oscillator. *IEEE Transactions on Ultrasonics*, *Ferroelectrics, and Frequency Control* (*UFFC*), **59**, 564-572. http://dx.doi.org/10.1109/TUFFC.2012.2230

[10] Sato, T. and Akitsu, T. (2014) Simultaneous Multimode Oscillation of Stress-Compensated Cut Quartz Resonator with Narrow-Band Wide Variable-Range Quartz Crystal Oscillator. *Engineering*, **6**, 973-988. http://dx.doi.org/10.4236/eng.2014.613088

[11] Sato, T., Kudo, A. and Akitsu, T. (2015) Start-Up Acceleration of Quartz Crystal Oscillator Using Active Inductance Double Resonance and Embedded Triggering Circuit. *Journal of Sensor Technology*, **5**.

[12] Lu, W., Tse, T. and Chan, W.-C. (2006) Crystal Oscillator Circuit Having a Start-up Time Reduction Circuit. US Patent 7034628 B1.

[13] Izumiya, S., Asaki, J. and Adachi, T. (2003) Algebraic Analysis Method of Start-Up Characteristics of Cascade Crystal Oscillator. *Electronics and Communications in Japan* (*Part II*: *Electronic*), **86**, 32-43. http://dx.doi.org/10.1002/ecjb.10166

[14] Ulmer, R.W. (1990) Stat-Up Circuit for Low Power MOS Crystal Oscillator. US Patent 4956618.

[15] Barnes, J.A., Chi, A.R., Cutler, L.S., Healey, D.J., Leeson, D.B., McGunigal, T.E., Mullen Jr., J.A., Smith, W.L., Sydnor, R.L., Vessot, R.F.C. and Winkler, G.M.R. (1971) Characterization of Frequency Stability. *IEEE Transactions on Instrumentation and Measurement*, **IM-20**, 105-120.

[16] Allan, D.W. (1966) Statistics of Atomic Frequency Standards. *Proceedings of the IEEE*, **54**, 221-230. http://dx.doi.org/10.1109/PROC.1966.4634

[17] IEEE Standard 1139. Definitions of Physical Quantities for Fundamental Frequency and Time Metrology—Random Instabilities, E.E. ISBN: 0-7381-1754-4.

[18] Linear Technology Corporation (2014) LTspice IV for Windows. http://www.linear-tech.co.jp/designtools/software/

Mathematical Models to Simultaneously Determine Overtime Requirements and Schedule Cells

Gürsel A. Süer, Kush Mathur

ISE Department, Ohio University, Athens, OH, USA
Email: suer@ohio.edu, km166610@ohio.edu

Abstract

The problem studied in this paper was inspired from an actual textile company. The problem is more complex than usual scheduling problems in that we compute overtime requirements and make scheduling decisions simultaneously. Since having tardy jobs is not desirable, we allow overtime to minimize the number of tardy jobs. The overall objective is to maximize profits. We present various mathematical models to solve this problem. Each mathematical model reflects different overtime workforce hiring practices. An experimentation has been carried out using eight different data sets from the samples of real data collected in the above mentioned textile company. Mathematical Model 2 was the best mathematical model with respect to both profit and execution time. This model considered partial overtime periods and also allowed different overtime periods on cells. We could solve problems up to 90 jobs per period. This was much more than what the mentioned textile company had to handle on a weekly basis. As a result, these models can be used to make these decisions in many industrial settings.

Keywords

Overtime Decisions, Cell Loading, Cell Scheduling, Mathematical Modeling

1. Introduction

In manufacturing systems, capacity allocation and machine scheduling play an essential role. Typically, capacity requirements are determined first and then detailed scheduling is performed to fit into the determined capacity. The capacity planning establishes the number of machines/workers needed in a process layout environment, number of production lines or cells needed in a cellular environment, number of shifts to be utilized, whether

overtime will be done or not, etc. Various approaches have been developed for capacity planning purposes such as capacity bills, resources profiles, rate-based capacity planning, and mathematical modeling, etc. On the other hand, scheduling deals with the detailed allocation of resources established in the capacity planning step to various jobs over a period of time. Scheduling establishes start and completion times of jobs on resources, *i.e.* machines, cells, production lines and workers.

In a company, each on-time delivered product brings revenue and customer satisfaction; both are essential to maintain the profitability of any company over a long term period. Making a job tardy is highly undesirable for any company. Establishing capacity requirements is a difficult task and it gets even harder when capacity requirements vary from one period to the next. In this situation, using overtime is a feasible method and many companies use that. Therefore, overtime can be used to adjust the capacity to handle the jobs which otherwise become tardy and may lead to poor reputation or in worst case loosing up of valuable customers. Overtimes are of two types: weekend overtime and weekday overtime. This distinction is important in that their costs usually vary. Weekend overtimes are usually more expensive than the weekday overtime because of the change in hourly workforce rate. Therefore, each overtime decision comes with a cost associated with it. This all makes the problem difficult to solve and lead to a NP-hard optimization problem where the weekend and weekday overtime decisions as well as cell loading and job sequencing are performed to maximize the overall net profit. We need to mention that if overtime is needed week after week, this may be an indicator that organization needs to re-assess capacity requirements and adjust regular capacity levels, *i.e.*, increase number of cells, number of machines, etc.

2. Problem Definition

The problem studied in this paper was inspired from an actual textile company. In every planning period (week) there are *n* jobs to complete using *m* number of identical cells. Manufacturing cells are popular in the manufacturing world. Cellular Manufacturing (CM) aims to obtain the flexibility to produce a high or moderate variety of low or moderate demand products with high productivity. CM is a type of manufacturing system that consists of manufacturing cell(s) with dissimilar machines needed to produce part family/families. Generally, the products grouped together form a product family. The benefits of CM are lower setup times, smaller lot sizes, lower work-in-process inventory and less space, reduced material handling, and shorter flow time, simpler work flow (Suresh & Kay, 1998 [1]). Cellular Manufacturing also simplifies planning and control of various tasks floating on the shop floor. Cells are ideally equipped with all of the machines needed to produce a productor product family (independent) and usually follow uni-directional flow (flowshop).

Many textile companies also use cells to enjoy the benefits the cells have to offer. Machines used in cells in textile industry are usually limited to different types of sewing machines such as overlock machines, stitch machines, automatic flat lock sewing machines. In this study, each cell can perform eight operations. The operations performed are briefly discussed in this paragraph. The first operation is cutting and the next three operations require sewing machines. The operations that follow are finishing, ironing, cleaning and packing. Operations 2 and 4 each were performed on two identical machines as they were the slowest operations. It is easy to re-configure cells in the textile industry once products with different requirements are produced as machines and work benches are easier to move around.

In this paper, it is assumed that each job has its own individual due date. The modeling gets complicated in that we need to decide additional capacity requirements at the same time we make scheduling decisions. Any job that cannot be completed by its due date becomes tardy and is assumed lost sales. Having tardy jobs is not desirable as this adversely affects relationships with customers and thus reputation of the firm in the long term. As a result, we allow overtime to minimize the number of tardy jobs. Overtime decisions are made for shifts on a daily basis. Overtime can be done during the weekends (Saturday and Sunday) as well as weekdays. Weekend overtime is done prior to upcoming week. It is also assumed that weekend overtime cost is higher than weekday overtime cost. The overall objective is to maximize profits. Tardy jobs are avoided as long as profit to be made is higher than the corresponding production/overtime costs. There are limitations on weekday and weekend overtime capacities. In this paper, the number of shifts considered is 2. Models can easily be modified to accommodate 1 shift. However, if the company wants to adopt a 3-shift policy, then weekday overtime option has to be eliminated and overtime can only be done during weekends. Lot splitting is not allowed in this study.

We propose mathematical models based on the current overtime workforce handling practices followed by the industries to solve this problem and then compare their performance in this study.

3. Literature Review

Several researchers focused on cell loading problem. Among them, Süer, Saiz, Dagli & Gonzalez [2] andSuer, Saiz & Gonzalez [3] developed several simple cell loading rules for connected and independent cells, respectively. They have tested the performance of these rules against four different performance measures, namely number of tardy jobs, maximum tardiness, total tardiness and utilization. Süer, Vazquez & Cortes [4] developed a hybrid approach of Genetic Algorithms (GA) and local optimizer to minimize number of tardy jobs in a multi-cell environment. In the same study, they have used three different GA approaches; basic GA approach, hybrid GA with local optimizers and finally GA with learning feature. Hybrid GA outperformed basic GA, however, GA with learning improved solutions only marginally. Süer, Arikan & Babayiğit [5] [6] studied cell loading subject to manpower restrictions and developed fuzzy mathematical models to minimize number of tardy jobs and total manpower levels. They have used various fuzzy operators and identified fuzzy operators that can find non-dominated solutions. In their work, number of cells and crew size for each cell were identified in addition to product sequencing decisions.

There are a number of works reported where both cell loading and product sequencing tasks are carried out. Süer & Dagli [7] and Süer, Cosner & Patten [8] discussed models to minimize makespan, machine requirements and manpower transfers. These works first assigned products to cells to minimize makespan and then re-sequenced the products to minimize manpower requirements. Yarimoglu [9] developed a mathematical model and a genetic algorithm to minimize manpower shortages in cells with synchronized material flow. Synchronized material flow allowed material transfers between stages at regular intervals (e.g., 4 hour in his case).

Some other researchers studied the sequencing of families in a single cell or machine. Nakamura, Yoshida & Hitomi [10] considered sequence-independent group setup in his work and his objective was to minimize total tardiness. Hitomi & Ham [11] also considered sequence-independent setup times for a single machine scheduling problem. Ham, Hitomi, Nakamura & Yoshida [12] developed a branch-and-bound algorithm for the optimal group and job sequence to minimize total flow time keeping the number of tardy jobs minimum. Pan and Wu [13] focused on single machine scheduling problem to minimize mean flow time of all jobs subject to due date satisfaction. They categorized the jobs into groups without family splitting. Gupta and Chantaravarapan [14] studied the single machine scheduling problem to minimize total tardiness considering group technology. Individual due dates and independent family setup times have been used in their problem with no family splitting. Süer and Mese [15] studied the same problem but allowed family splitting as it would be beneficial in many cases.

A number of researchers focused on shoe manufacturing industry and reported various results. These works heavily focused on scheduling jobs on the Rotary Injection Molding Machines. Süer. Santos and Vazquez [16] have developed a three-phase Heuristic Procedure to minimize makespan. Süer, Subramanian and Huang [17] included some heuristic procedures and mathematical models for cell loading and scheduling problem. The objective was to minimize makespan and unlimited availability of the molds was assumed. Later, Huang, Suer and Urs [18] introduced the limited mold availability into the problem for the same objective.

The most important feature of the scheduling problem studied in this paper is that both overtime decisions and detailed scheduling tasks including cell loading and product sequencing are performed simultaneously. Even though there were previous works that considered capacity and scheduling decisions simultaneously before (see Süer, Arikan & Babayiğit [5] [6]), there is only one work that considers both overtime and scheduling decisions simultaneously (see Mathur and Süer [19]) in the literature to the best knowledge of the authors. We propose new mathematical models that can be used under varying conditions. The objective is to maximize the profit and it is closely related to finishing jobs on time.

4. Integer Programming Formulation

Maxwell [20] developed a mathematical model to minimize number of tardy jobs on a single machine schedule. Later, Süer, Pico and Santiago [21] extended this model for parallel machine scheduling. In a later study, Süer, Arikan and Babayiğit [5] [6] adapted this model to cell loading problem where jobs are assigned to cells and manpower levels for each cell are identified simultaneously. Süer, Mathur and Ji [19] extended these models where due date equations are modified to allow for overtime decisions. In the models proposed in this paper, the due dates of jobs are modified in a similar way to meet different working conditions. This modification allows us to consider overtime decisions along with cell loading and scheduling. The main logic behind this formula-

tion is that the due time for a job is extended (increased) by the total overtime hours planned before this job is due.

Assume that a job is due Wednesday noon. Assuming only one shift work, then the due time for this job is calculated as 20 hours (8 hours available on Monday, 8 hours available on Tuesday and 4 hours available on Wednesday). Assume that mathematical model proposes to do overtime both on Monday and Tuesday, 3 hours each day. Then the due time of this job becomes 26 hours (=20 + 3 + 3). The proposed mathematical models accommodate this possibility in the formulation. The objective is to maximize profit. All of the jobs should be ordered by the Earliest Due Date (EDD) before model is run ($d_{[1]} \leq d_{[2]} \leq d_{[3]} \leq \ldots \leq d_{[n]}$). Similarly $p_{[i]}$ is the processing time of the job in the ith order. It is also assumed that weekend overtime work is done prior to planning week.

In this section, four different integer programming formulations are introduced. Each one of them is based on the observation on overtime workforce handling practices followed by the industries. Indices and parameters used throughout the paper are listed below.

Indices:
i: job index
j: cell index
t: day index
q: shift index
Parameters:
n: number of jobs
m: number of identical cells
p_i: processing time of job i (*in hours*)
d_i: due time of job i (*in hours*)
$shift_{[i]v}$: shift in which i^{th} job is due
s_i: sales price of job i
cwd: weekday overtime cost per shift
cwe: weekend overtime cost per shift
day_i: due day of job i depending on due time
otwd: length of weekday overtime (3 hours)
otwe: length of weekend overtime (8 hours)

4.1. Mathematical Model 1

The unique feature about this model is that it allows different weekday and weekend overtime decisions with respect to different cells and also workers are paid for the complete period of overtime slot even if they finish the work early in the overtime period. The model is capable of determining the schedule of jobs on each cell as well as handling the capacity of each cell by making overtime decisions.

The objective is to maximize the net profit as shown in Equation (1). This is computed by subtracting all weekend and weekday overtime costs from the sales revenue of the products finished by their due time. Weekday overtime can be done on any of the five working days and weekend overtime can be done on Saturday and/or Sunday. The first set of constraints (Equation (2)) enforce that early jobs are completed before their due times (original or extended). The second set of constraints (Equation (3)) guarantees that a job can assigned to at most one cell. The values of decision variables x_{ij} and ot_{tjq} will be known once the mathematical model is solved.

Decision Variables:
x_{ij}: 1 *if job i is assigned to cell j*, 0 *otherwise*
ot_{tjq}: 1 *if there is overtime on day t in shift q of cell j*, 0 *otherwise*
Objective Function:

$$\text{Max } z = \sum_{i=1}^{n}\sum_{j=1}^{m} s_i x_{ij} - cwd\sum_{t=1}^{5}\sum_{j=1}^{m}\sum_{q=1}^{2} ot_{tjq} - cwe\sum_{t=6}^{7}\sum_{j=1}^{m}\sum_{q=1}^{2} ot_{tjq} \tag{1}$$

Subject to:

$$\sum_{i=1}^{k} p_{[i]} x_{[i]j} \leq d_{[k]} + otwd\sum_{t=1}^{\text{day}_{[k]}}\sum_{q=1}^{\text{shift}_{[k]}} ot_{tjq} + otwe\sum_{t=6}^{7}\sum_{q=1}^{2} ot_{tjq} \qquad \forall k = 1, 2, \cdots, n, \forall j = 1, 2, \cdots, m \tag{2}$$

$$\sum_{j=1}^{m}x_{ij} \leq 1 \qquad \forall i = 1, 2, \cdots n \tag{3}$$

4.2. Mathematical Model 2

Model 2 is similar to Model 1 in that it allows different weekday and weekend overtime work with respect to different cells. However, the difference between two models comes from the practice that overtime workforces can be hired and paid for the fraction of the overtime slot in this Model. In this model, overtime decision variable is modified as given below.

ot_{tjq}: *fraction of overtime to do on day t in shift q of cell j between* 0 *and* 1

4.3. Mathematical Model 3

The model has been first proposed by Süer, Mathur and Ji [19]. The unique feature about this model is that a uniform overtime strategy is applied, *i.e.*, all of the cells get to do the same overtime whenever an overtime decision is made. This may be important in some companies where all the workers are given the same opportunity to do overtime and gain additional income. Overtime workforces are hired and paid for the complete period of overtime slot on all cells. In this case, overtime decision variable is simplified as shown below.

ot_{tq}: 1 *if there is overtime on day t in shift q*, 0 *otherwise*

The objective is to maximize the net profit as shown in Equation (4). The first set of constraints (Equation (5)) enforce that early jobs are completed before their due times (original or extended). The second set of constraints (Equation (6)) guarantees that a job can assigned to at most one cell.

Objective Function:

$$\text{Max } z = \sum_{i=1}^{n}\sum_{j=1}^{m}s_i x_{ij} - cwd\ m\sum_{t=1}^{5}\sum_{q=1}^{2}ot_{tq} - cwe\ m\sum_{t=6}^{7}\sum_{q=1}^{2}ot_{tq} \tag{4}$$

Subject to:

$$\sum_{i=1}^{k}p_{[i]}x_{[i]j} \leq d_{[k]} + otwd\sum_{t=1}^{\text{day}_{[k]}}\sum_{q=1}^{\text{shift}_{[k]}}ot_{tq} + otwe\sum_{t=6}^{7}\sum_{q=1}^{2}ot_{tq} \qquad \forall k = 1, 2, \cdots, n, \forall j = 1, 2, \cdots, m \tag{5}$$

$$\sum_{j=1}^{m}x_{ij} \leq 1 \qquad \forall i = 1, 2, \cdots, n \tag{6}$$

4.4. Mathematical Model 4

Model 4 is similar to Model 3 in that all of the cells get to do the same overtime whenever an overtime decision is made. However, the difference between two models is that overtime workforces can be hired and paid for the fraction of the overtime slot in Model 4. In this case, overtime decision variable is modified as shown below.

ot_{tq}: *fraction of overtime to do on day t in shift q between* 0 *and* 1

The main features of four mathematical models are summarized in **Table 1**.

5. A Numerical Illustration

An example case is illustrated in this section. A 20-job and 2-cell problem is solved by all four mathematical models using ILOG OPL 6.3 optimization tool. In each mathematical model illustration, three tables and one figure are given. The first table shows the sequence of jobs assigned to each cell; the second table shows the corresponding overtime decisions for the problem; the third table shows the original due day, due shift and due time; revenue generated from each job, processing times, detailed computations of completion times and revised due dates (D_i*) based on overtime decisions; and the Gantt chart is given in the corresponding figure. It is important to note that the weekend overtime starts before the week starts. For example, to complete a job before the set due date (say on Monday evening), overtime can be performed on Saturday and/or Sunday shifts.

5.1. Solution by Mathematical Model 1

This model assigned 18 out of 20 jobs to cells. Jobs 8 and 14 are not assigned to any cell (hence tardy and in this

paper lost customer orders). The jobs assigned to each cell are summarized in **Table 2**. **Table 3** shows when and on what cell we need to do overtime during the week. **Table 4** shows the details of all computations. When a job is tardy, its completion time is not listed since it is assumed as lost sales (for example for job 8 in **Table 4**). Job 1 in **Table 4** is due Monday, shift 1, and original due time is 5.79 hours and it has been assigned to cell 1. Since we allow overtime on Sunday shift 1 for eight hours on cell 1, the due date is taken as (D_1^*=) 13.79 hours (=5.79 + 8). Similar adjustment are made for all jobs, *i.e.* jobs original due dates are modified (increased) by considering the overtimes scheduled before their due date. On cell 1, Monday shift1 is of regular eight hours whereas on cell 2, Monday shift 1 is of 11 hours (including overtime). On both the cells, Monday shift 2 are 11 hours (including overtimes); Tuesday shift 1 and shift 2 are 11 (including overtimes) and eight hours respectively; and the remaining weekdays are all regular eight hours. Whereas on the weekends, only cell 1 needs eight hours of overtime only on Sunday shift1 as shown in **Table 3** and **Figure 1**.

Table 1. Characteristics of Mathematical models over Overtime decision variables.

Mathematical Models	Overtime			
	Different on different cells	Same on all cells	Complete duration	Partial duration
Mathematical Model 1	✓		✓	
Mathematical Model 2	✓			✓
Mathematical Model 3		✓	✓	
Mathematical Model 4		✓		✓

Table 2. Sequence of jobs assigned to each cell for model 1.

Cells	Jobs
C1	1, 3, 4, 6, 9, 10, 11, 12, 16, 19
C2	2, 5, 7, 13, 15, 17, 18, 20

Table 3. Overtime decisions for model 1.

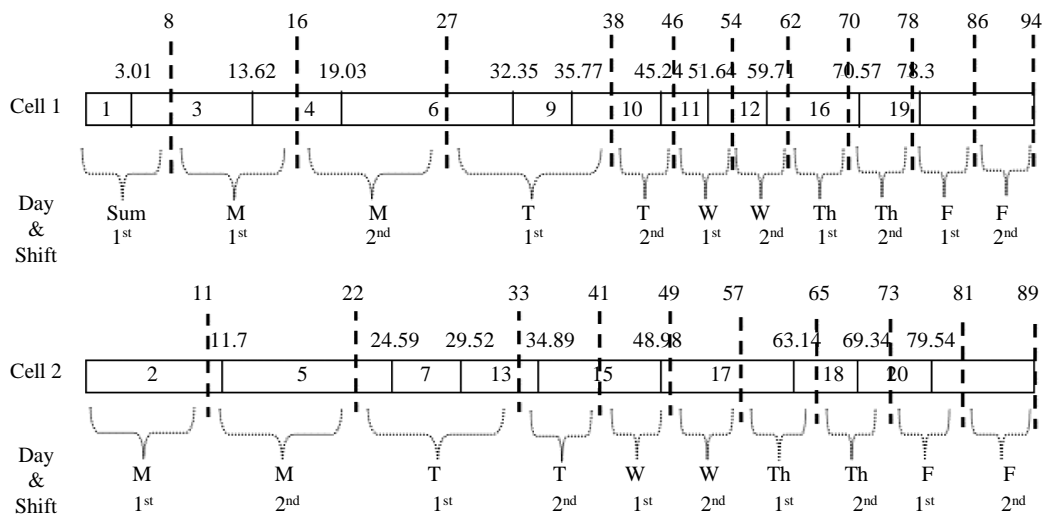|		Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Cells	C1	Shift 1	0	1	0	0	0	0	1
		Shift 2	1	0	0	0	0	0	0
	C2	Shift 1	1	1	0	0	0	0	0
		Shift 2	1	0	0	0	0	0	0

Figure 1. Gantt chart for model 1 solution.

Table 4. Detailed computations for model 1 solution.

Jobs	Due Day	Due Shift	Processing time	Due time	D_i^*	Completion Time (C_i)	Sales Price (s_i) in $	Early/Tardy
1	Monday	1	3.01	5.79	13.79	3.01	2424.38	Early
2	Monday	2	11.70	13.75	19.75	11.70	3539.11	Early
3	Tuesday	1	10.61	17.11	31.11	13.62	2451.80	Early
4	Tuesday	1	5.41	18.44	32.44	19.03	1792.35	Early
5	Tuesday	1	12.89	19.08	28.08	24.59	3733.76	Early
6	Tuesday	1	13.32	19.49	33.49	32.35	3748.54	Early
7	Tuesday	1	4.93	20.96	29.96	29.52	1696.70	Early
8	Tuesday	1	15.35	21.04	-	-	2147.59	*Tardy*
9	Tuesday	1	3.42	22.21	36.21	35.77	1430.30	Early
10	Wednesday	2	9.47	45.36	59.36	45.24	1299.49	Early
11	Wednesday	2	6.40	45.90	59.90	51.64	1913.32	Early
12	Wednesday	2	8.07	46.06	60.06	59.71	3092.46	Early
13	Thursday	1	5.37	52.81	61.81	34.89	2799.01	Early
14	Thursday	1	11.74	54.82	-	-	1629.80	*Tardy*
15	Thursday	2	14.09	58.18	67.18	48.98	3914.95	Early
16	Thursday	2	10.86	59.30	73.30	70.57	2483.75	Early
17	Thursday	2	14.16	60.35	69.35	63.14	2026.68	Early
18	Friday	1	6.20	64.15	73.15	69.34	1799.68	Early
19	Friday	1	7.73	64.62	78.62	78.30	1110.93	Early
20	Friday	1	10.20	70.69	79.69	79.54	3483.41	Early

5.2. Solution by Mathematical Model 2

This model also assigned 18 out of 20 jobs to cells as shown in **Table 5**. On both cells, Monday shifts are 11 hours (including overtimes); Tuesday shift1 and shift 2 are 11 (including overtimes) and 8 hours, respectively; and the remaining weekdays except for Friday shift 1 are all regular eight hours. On cell 1, Friday shift 1 is of 8.39 hours (=8 + 0.13*3) whereas on cell 2 it is of 8.02 hours (=8 + 0.0067*3) including overtimes. On the weekends, overtimes are required on Sunday shift 1 of 3 hours (=0.375*8) on cell 1 and on Saturday shift 1 of 1.12 hours (=0.14*8) on cell 2 as shown in **Table 6**. In this case, the due date for job 1 (D_1^*=) is computed as 11.79 hours since we allow partial overtime on Sunday shift 1 for three hours (=0.375*8) and also allow three hour overtime on shift1 on Monday on cell 1 (=5.79 + 3 + 3). **Table 7** shows the details of all computations and the corresponding Gantt chart is given in **Figure 2**.

5.3. Solution by Mathematical Model 3

This model resulted in three tardy jobs. The sequence of jobs assigned to cells is given in **Table 8**. Overtime decisions are summarized in **Table 9**. On both the cells, Monday as well as Tuesday shift 1 and shift 2 are of 11 hours (including overtimes) and eight hours respectively; and the remaining days are all regular eight hours. On the other hand, no overtime work is scheduled for weekends. In this model, the due date for job 1 (D_1^*=) is taken as 8.79 hours (=5.79 + 3) as we allow three-hour overtime on Monday shift 1 on both cells. The detailed computations are summarized in **Table 10** and the corresponding Gantt chart is given in **Figure 3**.

Table 5. Sequence of jobs assigned to each cell for model 2.

Cells	Jobs
C1	1, 2, 6, 7, 12, 15, 17, 19
C2	3, 4, 5, 9, 10, 11, 13, 16, 18, 20

Table 6. Overtime decisions for model 2.

		Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Cells	C1	*Shift* 1	1	1	0	0	0.13	0	0.375
		Shift 2	1	0	0	0	0	0	0
	C2	*Shift* 1	1	1	0	0	0.0067	0.14	0
		Shift 2	1	0	0	0	0	0	0

Table 7. Detailed computations for model 2 solution.

Jobs	Due Day	Due Shift	Processing time	Due time	D_i*	Completion Time (C_i)	Sales Price (s_i) in $	Early/Tardy
1	Monday	1	3.01	5.79	11.79	3.01	2424.38	Early
2	Monday	2	11.70	13.75	22.75	14.71	3539.11	Early
3	Tuesday	1	10.61	17.11	27.23	10.61	2451.80	Early
4	Tuesday	1	5.41	18.44	28.56	16.02	1792.35	Early
5	Tuesday	1	12.89	19.08	29.20	28.91	3733.76	Early
6	Tuesday	1	13.32	19.49	31.49	28.03	3748.54	Early
7	Tuesday	1	4.93	20.96	32.96	32.96	1696.70	Early
8	Tuesday	1	15.35	21.04	-	-	2147.59	*Tardy*
9	Tuesday	1	3.42	22.21	32.33	32.33	1430.30	Early
10	Wednesday	2	9.47	45.36	55.48	41.80	1299.49	Early
11	Wednesday	2	6.40	45.90	56.02	48.20	1913.32	Early
12	Wednesday	2	8.07	46.06	58.06	41.03	3092.46	Early
13	Thursday	1	5.37	52.81	62.93	53.57	2799.01	Early
14	Thursday	1	11.74	54.82	-	-	1629.80	*Tardy*
15	Thursday	2	14.09	58.18	70.18	55.12	3914.95	Early
16	Thursday	2	10.86	59.30	69.42	64.43	2483.75	Early
17	Thursday	2	14.16	60.35	72.35	69.28	2026.68	Early
18	Friday	1	6.20	64.15	74.29	70.63	1799.68	Early
19	Friday	1	7.73	64.62	77.01	77.01	1110.93	Early
20	Friday	1	10.20	70.69	80.83	80.83	3483.41	Early

Table 8. Sequence of jobs assigned to each cell for model 3.

Cells	Jobs
C1	2, 5, 9, 10, 11, 12, 16, 19
C2	1, 4, 6, 7, 13, 15, 17, 18, 20

Table 9. Overtime decisions for model 3.

Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Shift 1	1	1	0	0	0	0	0
Shift 2	0	0	0	0	0	0	0

Table 10. Detailed computations for model 3 solution.

| Jobs	Due Day	Due Shift	Processing time	Due time	D_i^*	Completion Time (C_i)	Sales Price (s_i) in $	Early/Tardy
1	Monday	1	3.01	5.79	8.79	3.01	2424.38	Early
2	Monday	2	11.70	13.75	16.75	11.70	3539.11	Early
3	Tuesday	1	10.61	17.11	-	-	2451.80	*Tardy*
4	Tuesday	1	5.41	18.44	24.44	8.42	1792.35	Early
5	Tuesday	1	12.89	19.08	25.08	24.59	3733.76	Early
6	Tuesday	1	13.32	19.49	25.49	21.74	3748.54	Early
7	Tuesday	1	4.93	20.96	26.96	26.67	1696.70	Early
8	Tuesday	1	15.35	21.04	-	-	2147.59	*Tardy*
9	Tuesday	1	3.42	22.21	28.21	28.01	1430.30	Early
10	Wednesday	2	9.47	45.36	51.36	37.48	1299.49	Early
11	Wednesday	2	6.40	45.90	51.90	43.88	1913.32	Early
12	Wednesday	2	8.07	46.06	52.06	51.95	3092.46	Early
13	Thursday	1	5.37	52.81	58.81	32.04	2799.01	Early
14	Thursday	1	11.74	54.82	-	-	1629.80	*Tardy*
15	Thursday	2	14.09	58.18	64.18	46.13	3914.95	Early
16	Thursday	2	10.86	59.30	65.30	62.81	2483.75	Early
17	Thursday	2	14.16	60.35	66.35	60.29	2026.68	Early
18	Friday	1	6.20	64.15	70.15	66.49	1799.68	Early
19	Friday	1	7.73	64.62	70.62	70.54	1110.93	Early
20	Friday	1	10.20	70.69	76.69	76.69	3483.41	Early
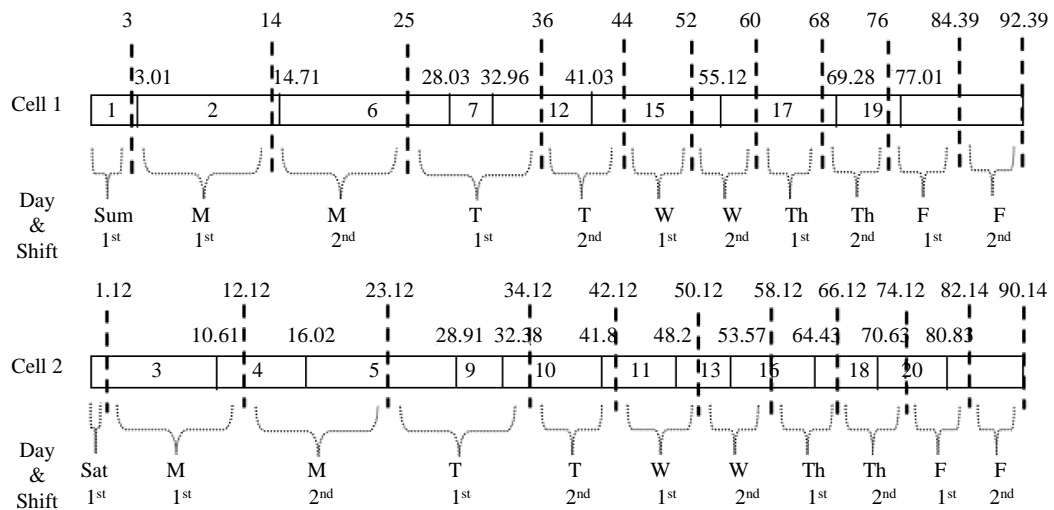
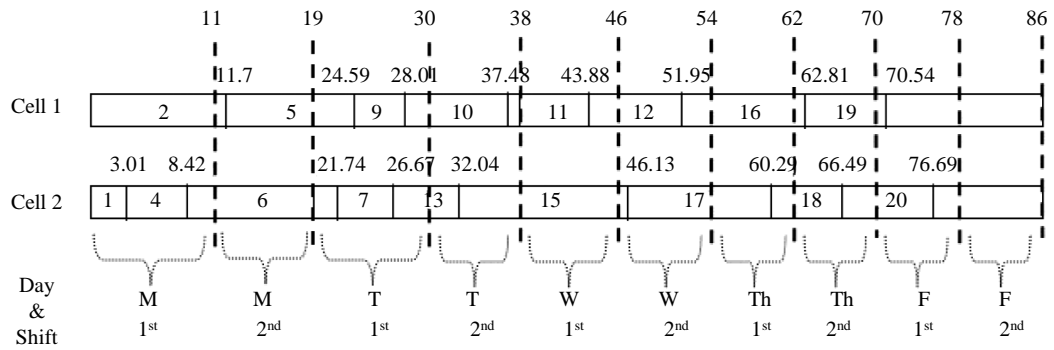**Figure 2.** Gantt chart for model 2 solution.

Figure 3. Gantt chart for model 3 solution.

5.4. Solution by Mathematical Model 4

This model also led to 18 jobs being completed before their due dates. The sequence of jobs assigned to cells is given in **Table 11**. Overtime decisions are summarized in **Table 12**. In **Table 13**, for example Job 1, is due Monday, shift 1, and original due time is 5.79 hours. Since we allow partial overtime of 2.21 hours (=0.27625*8) on Saturday shift1 and also allow three hour overtime on Monday shift 1 on both the cells, the due date is taken as (D_1*=) 11 hours (=5.79 + 2.21 + 3). Similar adjustment are made for all jobs, *i.e.* jobs original due dates are modified (increased) by considering the overtimes scheduled before their due date. On both the cells, Monday shift 1 and shift 2 are 11 hours (including overtimes) each; Tuesday shift 1 and shift 2 are 11 hours (including overtimes) and regular of eight hours, respectively; Thursday shift 1 and shift 2 are regular eight hours and 8.09 hours (=8 + 0.03*3) (including overtime) respectively and the remaining days are all regular eight hours. Whereas on the weekend on both the cells, overtime is required only on Saturday shift 1 of 2.21 hours (=0.27625*8) as shown in **Figure 4**.

6. Experimentation Results

In this section, experimentation results are discussed. Eight problems were formed with 20, 30, 40, 50, 60, 70, 80 and 90 jobs; and 2, 2, 2, 2, 2, 4, 4 and 5 cells, respectively (problems 1, 2, 3, 4, 5, 6, 7 and 8 respectively). The data generated is inspired from the Textile Company where the problem was observed. Each product undergoes eight different operations. The upper and lower range of processing times for each operation were identified based on the sample data provided by the company. The processing times were generated randomly from this interval for the corresponding operation in each data set. Processing times for operations 1, 2, 3, 4, 5, 6, 7 and 8 follow uniform distributions UD(2, 4), US(1, 2.5), UD(1, 3.5), UD(1, 2), UD(1, 2), UD(1, 2), UD(1, 2), and UD(0.2, 2), respectively. The batch size of each product follows uniform distribution UD(50, 250). The production rate for each product is identified based on the bottleneck machine. The processing time (hr) for a job is computed dividing its batch size with its production rate. The due time of each job is generated by using Equation (7). Products' due days (considering one week as a period) as well as a due shifts (1 or 2) are calculated on the basis of due time. The sale price of each job is assigned by using Equation (8). The hourly labor rates for regular time, weekday overtime and weekend overtime are taken as $10, $15, and $20, respectively. The weekend overtime is limited to 8 hours per shift whereas weekday overtime is restricted to 3 hours per shift.

$$d_i = p_i + \left(\frac{\sum_{i=1}^{n} p_i}{n} \right) \times UD(1,10) \tag{7}$$

$$s_i = p_i \times 200 \times \left[1 + UD(0.25, 0.5) \right] \tag{8}$$

6.1. Results of Mathematical Model 1

The results of Mathematical Model1 are shown in **Table 14**. The results show that mathematical model produced tardy jobs in spite of the fact that overtime was done both during the weekdays and weekend for some problems (problems 1, 4, and 5). In problems 2, 7, 8, weekend overtime was not done even though there was a

Table 11. Sequence of jobs assigned to each cell for model 4.

Cells	Jobs
C1	2, 4, 5, 9, 10, 11, 13, 16, 18, 20
C2	1, 3, 6, 7, 12, 15, 17, 19

Table 12. Overtime decisions for model 4.

Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Shift 1	1	1	0	0	0	0.27625	0
Shift 2	1	0	0	0.03	0	0	0

Table 13. Detailed computations for model 4 solution.

| Jobs	Due Day	Due Shift	Processing time	Due time	D_i^*	Completion Time (C_i)	Sales Price (s_i) in $	Early/Tardy
1	Monday	1	3.01	5.79	11.00	3.01	2424.38	Early
2	Monday	2	11.70	13.75	21.96	11.70	3539.11	Early
3	Tuesday	1	10.61	17.11	28.32	13.62	2451.80	Early
4	Tuesday	1	5.41	18.44	29.65	17.11	1792.35	Early
5	Tuesday	1	12.89	19.08	30.29	30.00	3733.76	Early
6	Tuesday	1	13.32	19.49	30.70	26.94	3748.54	Early
7	Tuesday	1	4.93	20.96	32.17	31.87	1696.70	Early
8	Tuesday	1	15.35	21.04	-	-	2147.59	*Tardy*
9	Tuesday	1	3.42	22.21	33.42	33.42	1430.30	Early
10	Wednesday	2	9.47	45.36	56.56	42.89	1299.49	Early
11	Wednesday	2	6.40	45.90	57.11	49.29	1913.32	Early
12	Wednesday	2	8.07	46.06	57.27	39.94	3092.46	Early
13	Thursday	1	5.37	52.81	64.02	54.66	2799.01	Early
14	Thursday	1	11.74	54.82	-	-	1629.80	*Tardy*
15	Thursday	2	14.09	58.18	69.48	54.03	3914.95	Early
16	Thursday	2	10.86	59.30	70.60	65.52	2483.75	Early
17	Thursday	2	14.16	60.35	71.65	68.19	2026.68	Early
18	Friday	1	6.20	64.15	75.45	71.72	1799.68	Early
19	Friday	1	7.73	64.62	75.92	75.92	1110.93	Early
20	Friday	1	10.20	70.69	81.99	81.92	3483.41	Early
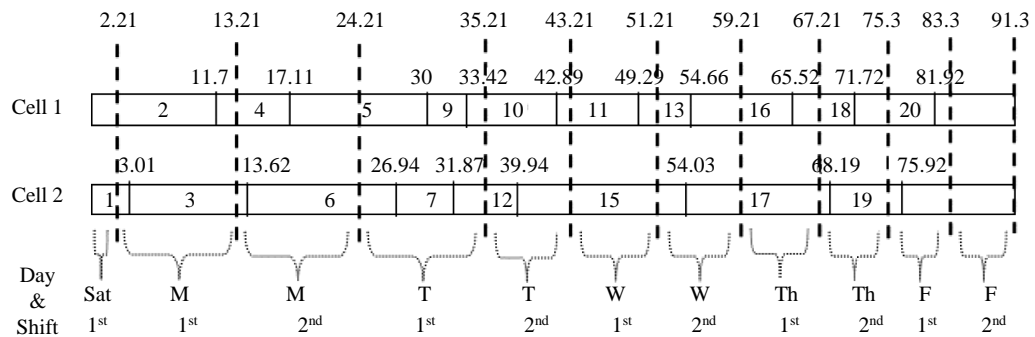
Figure 4. Gantt chart for model 4 solution.

Table 14. Mathematical model 1 results.

Problem #	# of jobs	# of cells	nT	Doing weekday overtime (Y/N)	Doing weekend overtime (Y/N)	Objective function value	Computation time
1	20	2	2	Y	Y	40890.62	3 s
2	30	2	1	Y	N	30085.93	2 m 8 s
3	40	2	0	Y	Y	36465.85	<1 s
4	50	2	2	Y	Y	48188.39	27 s
5	60	2	6	Y	Y	50869.05	56 m 58 s
6	70	4	0	Y	N	88216.51	<1 s
7	80	4	1	Y	N	93810.16	4 m 54 s
8	90	5	1	Y	N	109658.39	4 m 19 s

tardy job in the system. Doing overtime would cost more than profit to be obtained from this tardy in these problems. In problem 3, tardy job(s) were avoided by doing overtime both weekdays and weekends. Finally, in problem 6, weekday overtime was sufficient to avoid tardy jobs. The execution times of models varied from 1 second to almost 57 minutes.

6.2. Results of Mathematical Model 2

The results of Mathematical Model 2 are summarized in **Table 15**. There were tardy jobs in problems 1, 2, and 5 even though overtime was done both during the weekdays and weekend. In problems 3 and 4, tardy job(s) were avoided by doing overtime both weekdays and weekend. Finally, in problems 6, 7 and 8, weekday overtime was sufficient to avoid tardy jobs. The execution times of models were less than 2 seconds.

6.3. Results of Mathematical Model 3

The results of Mathematical Model 3 are given in **Table 16**. In problems 4 and 5, there were tardy jobs even though overtime was done both during the weekdays and weekend. In problems 1 and 2, weekend overtime was not done even though there were tardy jobs in the system. Finally, in problems 3, 6 and 8, tardy jobs were avoided by doing overtime either on weekdays or both weekdays and weekend. Problem 7 could not be solved due to memory restrictions. The execution times of models varied from 1 second to almost 2 hours 4 minutes.

6.4. Results of Mathematical Model 4

The results of Mathematical Model 4 are shown in **Table 17**. The results show that mathematical model produced tardy jobs in spite of the fact that overtime was done both during the weekdays and weekend for some problems (problems 1, 4, and 5). In problems 2, 7, 8, weekend overtime was not done even though there was a tardy job in the system. Doing overtime would cost more than profit to be obtained from this tardy in these problems. In problem 3, tardy job(s) were avoided by doing overtime both weekdays and weekends. Finally, in problem 6, weekday overtime was sufficient to avoid tardy jobs. The execution times of models varied from 1 second to almost 57 minutes.

6.5. Comparison of Mathematical Models

In this section, the results of four mathematical models are compared. The highest profit for each problem was obtained with mathematical model 2 (MM2). MM4 found the second highest profit in six of the problems tested and tied with MM2 in problems 2 and 4. The main reason why MM2 and MM4 obtained better results is due to the fact that overtime decision variables are allowed to take fractional values thus lowering the overtime cost. On the other hand, MM1 got the third highest profit values in five problems and tied with MM3 in the remaining three problems. Based on the results, the preference of models can be listed as 1) MM2; 2) MM4; 3) MM1; and 4) MM3. It is also important to note that the difference between the best and worst profit values in the problems tested here is less than 0.80% in seven out of eight problems and it is 1.62% in the remaining problem. The results of four mathematical models are summarized in **Table 18**.

Table 15. Mathematical model 2 results.

Problem #	# of jobs	# of cells	nT	Doing weekday overtime (Y/N)	Doing weekend overtime (Y/N)	Objective function value	Computation time
1	20	2	2	Y	Y	41155.12	<1 s
2	30	2	1	Y	Y	30243.78	<1 s
3	40	2	0	Y	Y	36622.85	<1 s
4	50	2	0	Y	Y	48557.07	<1 s
5	60	2	6	Y	Y	50876.50	9 s
6	70	4	0	Y	N	88311.01	<1 s
7	80	4	0	Y	N	94068.04	2 s
8	90	5	0	Y	N	109881.77	2 s

Table 16. Mathematical model 3 results.

Problem #	# of jobs	# of cells	nT	Doing weekday overtime (Y/N)	Doing weekend overtime (Y/N)	Objective function value	Computation time
1	20	2	3	Y	N	40488.82	<1 s
2	30	2	1	Y	N	30081.08	4 m 24 s
3	40	2	0	Y	Y	36465.85	<1 s
4	50	2	1	Y	Y	48168.69	<1 s
5	60	2	6	Y	Y	50869.05	2 h 3 m 31 s
6	70	4	0	Y	N	88216.51	<1 s
7	80	4		out of memory			
8	90	5	0	Y	N	109632.80	<1 s

Table 17. Mathematical model 4 results.

Problem #	# of jobs	# of cells	nT	Doing weekday overtime (Y/N)	Doing weekend overtime (Y/N)	Objective function value	Computation time
1	20	2	2	Y	Y	41155.12	<1 s
2	30	2	1	Y	Y	30243.78	<1 s
3	40	2	0	Y	Y	36622.85	<1 s
4	50	2	0	Y	Y	48557.07	<1 s
5	60	2	6	Y	Y	50876.50	6 m 41 s
6	70	4	0	Y	N	88311.01	3 m 27 s
7	80	4	0	Y	Y	94068.04	17 s
8	90	5	0	Y	N	109881.77	15 m 36 s

Table 18. Performance comparison between mathematical models.

Problem #	# of jobs	# of cells	Mathematical Model 1		Mathematical Model 2		Mathematical Model 3		Mathematical Model 4	
			nT	Objective function value	nT	Objective function value	nT	Objective function value	nT	Objective function value
1	20	2	2	40890.62	2	41155.12	3	40488.82	2	41155.12
2	30	2	1	30085.93	1	30243.78	1	30081.08	1	30243.78
3	40	2	0	36465.85	0	36622.85	0	36465.85	0	36622.85
4	50	2	2	48188.39	0	48557.07	1	48168.69	0	48557.07
5	60	2	6	50869.05	6	50876.50	6	50869.05	6	50876.50
6	70	4	0	88216.51	0	88311.01	0	88216.51	0	88311.01
7	80	4	1	93810.16	0	94068.04		Out of memory	0	94068.04
8	90	5	1	109658.39	0	109881.77	0	109632.80	0	109881.77

7. Conclusions and Future Work

In this study, various mathematical models are proposed to determine overtime requirements, assign jobs to cells and also determine the sequence of jobs simultaneously in each cell. The objective is to maximize the profit to be obtained from the completion of products on time. These mathematical models represent different overtime workforce handling practices followed by industries. The problem has been inspired from a textile company. Data used in the experimentation has been generated based on the collected sample real data from the same company.

The result of a mathematical model gives us a weekly complete schedule with the overtime decisions on each weekend and weekday on each shift and on each cell. Moreover, it tells us the load on cells and also the schedule of jobs on each cell. The selection of an appropriate mathematical model obviously depends on practices of different companies. However, if there is a chance to select among these models, then Mathematical Model 2 would be the best choice as it outperforms all other mathematical models in all of the problems tested. Mathematical Model 2 allows partial overtime periods. It also allows different overtime periods on different cells. This flexibility in the problem formulation led to the best results. Mathematical model 4 tied the results in two of the problems and got the second best results in the rest of the problems. Indeed, the gap between the best and worst results when all models were considered was smaller than 1.62%.

The experimentation also showed that mathematical models could solve large problems relatively fast. Especially Mathematical Model 2 took less than nine seconds to find the optimal solution in all of the problems tested. This may be another factor that would make Mathematical Model 2 a desirable one in addition to the fact that it found the highest profits among all models. We could solve problems up to 90 jobs per period. The number of jobs mentioned in the textile company had to handle on a weekly basis roughly varied 20 - 40 jobs. As a result, these models can be used to make these decisions in similar industrial settings.

Some of the future work planned includes addition of setup times to the models. Allowing lot splitting is another option. We may also consider a minimum limit on overtime in the future models. Another possible extension is to allow tardy jobs but apply penalty (tardiness cost) when that occurs. In this study we assumed that cells were identical. It is also possible to consider cell size alternatives. One last note about the topic is that if a company uses overtime regularly, it may be worthwhile to adjust capacity levels (increase) to avoid additional overtime costs.

References

[1]   Suresh, N.C. and Kay, J.M. (1998) Group Technology and Cellular Manufacturing: A State-of-the-Art Synthesis of Research and Practice. Kluwer Academic, Boston. http://dx.doi.org/10.1007/978-1-4615-5467-7

[2]   Süer, G.A., Saiz, M., Dagli, C. and Gonzalez, W. (1995) Manufacturing Cell Loading Rules and Algorithms for Connected Cells. *Manufacturing Research and Technology Journal*, **24**, 97-127. http://dx.doi.org/10.1016/S1572-4417(06)80038-0

[3]   Süer, G.A., Saiz, M. and Gonzalez, W. (1999) Evaluation of Manufacturing Cell Loading Rules for Independent Cells. *International Journal of Production Research*, **37**, 3445-3468. http://dx.doi.org/10.1080/002075499190130

[4]   Süer, G.A., Vazquez, R. and Cortes, M. (2005) A Hybrid Approach of Genetic Algorithms and Local Optimizers in Cell Loading. *Computers and Industrial Engineering*, **48**, 625-641. http://dx.doi.org/10.1016/j.cie.2003.03.005

[5]   Süer, G.A., Arikan, F. and Babayiğit, C. (2008) Bi-Objective Cell Loading Problem with Non-Zero Setup Times with Fuzzy Aspiration Levels in Labour Intensive Manufacturing Cells. *International Journal of Production Research*, **46**, 371-404. http://dx.doi.org/10.1080/00207540601138460

[6]   Süer, G.A., Arikan, F. and Babayiğit, C. (2009) Effects of Different Fuzzy Operators on Fuzzy Bi-Objective Cell Loading Problem in Labor-Intensive Manufacturing Cells. *Computers & Industrial Engineering*, **56**, 476-488. http://dx.doi.org/10.1016/j.cie.2008.02.001

[7]   Süer, G.A. and Dagli, C. (2005) Intra-Cell Manpower Transfers and Cell Loading in Labor-Intensive Manufacturing Cells. *Computers & Industrial Engineering*, **48**, 643-655. http://dx.doi.org/10.1016/j.cie.2003.03.006

[8]   Süer, G.A., Cosner, J. and Patten, A. (2009) Models for Cell Loading and Product Sequencing in Labor-Intensive Cells. *Computers & Industrial Engineering*, **56**, 97-105. http://dx.doi.org/10.1016/j.cie.2008.04.002

[9]   Yarimoglu, F. (2009) Cell Loading and Product Sequencing Subject to Manpower Restrictions in Synchronized Manufacturing Cells, Ohio University. http://etd.ohiolink.edu/send-pdf.cgi/Yarimoglu%20Fatih.pdf?ohiou1234899632

[10] Nakamura, N., Yoshida, T. and Hitomi, K. (1978) Group Production Scheduling for Minimum Total Tardiness Part(I),

AIIE Transactions, **10**, 157-162.

[11] Hitomi, K. and Ham, I. (1976) Operations Scheduling for Group Technology Applications. *CIRP Annals*, **25**, 419-422.

[12] Ham, I., Hitomi, K., Nakamura, N. and Yoshida, T. (1979) Optimal Group Scheduling and Machining-Speed Decision under Due-Date Constraints. *Transactions of the ASME Journal of Engineering for Industry*, **101**, 128-134.

[13] Pan, J.C.H. and Wu, C.C. (1998) Single Machine Group Scheduling to Minimize Mean Flow Time Subject to Due Date Constraints. *Production Planning and Control*, **9**, 366-370. http://dx.doi.org/10.1080/095372898234091

[14] Gupta, J.N.D. and Chantaravarapan, S. (2008) Single Machine Group Scheduling with Family Setups to Minimize Total Tardiness. *International Journal of Production Research*, **46**, 1707-1722. http://dx.doi.org/10.1080/00207540601009976

[15] Süer, G.A. and Mese, E.M. (2011) Cell Loading and Family Scheduling for Jobs with Individual Due Dates. In: Modrák, V. and Pandian, R.S., Eds., *Operations Management Research and Cellular Manufacturing Systems*: *Innovative Methods and Approaches*, IGI Global, Hershey, 208-226.

[16] Süer, G.A., Santos, J. and Vazquez, R. (1999) Scheduling Rotary Injection Molding Machines. *Proceedings of the Second Asia-Pacific Conference on Industrial Engineering and Management Systems*, Kanazawa, 30-31 October 1999, 319-322.

[17] Süer, G.A., Subramanian, A. and Huang, J. (2009) Heuristic Procedures and Mathematical Models for Cell Loading and Scheduling in a Shoe Manufacturing Company. *Computers & Industrial Engineering*, **56**, 462-475. http://dx.doi.org/10.1016/j.cie.2008.10.008

[18] Huang, J., Süer, G.A. and Urs, S.B.R. (2011) Genetic Algorithm for Rotary Machine Scheduling with Dependent Processing Times. *Journal of Intelligent Manufacturing*, **23**, 1931-1948. http://dx.doi.org/10.1007/s10845-011-0521-9

[19] Mathur, K. and Süer, G.A. (2013) Math Modeling and GA Approach to Simultaneously Make Overtime Decisions, Load Cells and Sequence Products. *Computers & Industrial Engineering*, **66**, 614-624. http://dx.doi.org/10.1016/j.cie.2013.08.012

[20] Maxwell, W.L. (1970) On Sequencing *n* Jobs on One Machine to Minimize the Number of Late Jobs. *Management Science*, **16**, 295-297. http://dx.doi.org/10.1287/mnsc.16.5.295

[21] Süer, G.A., Pico, F. and Santiago, A. (1997) Identical Machine Scheduling to Minimize the Number of Tardy Jobs When Lot-Splitting Is Allowed. *Computers & Industrial Engineering*, **33**, 277-280. http://dx.doi.org/10.1016/S0360-8352(97)00092-2

Scientific
Research
Publishing

Thermal Behavior of Externally Driven Spindle: Experimental Study and Modelling

Christian Brecher[1], Yair Shneor[2], Stephan Neus[1], Kolja Bakarinow[1], Marcel Fey[1]

[1]RWTH Aachen University, Laboratory for Machine Tools and Production Engineering, Chair of Machine Tools, Aachen, Germany
[2]CAMT, Center for Advanced Manufacturing Technologies, Rotem Industries, Rotem Industrial Park, Mishor Yamin, D.N. Arava, Israel
Email: camt@rotemi.co.il, s.neus@wzl.rwth-aachen.de

Abstract

This paper focuses on model development for computer analysis of the thermal behavior of an externally driven spindle. The aim of the developed model is to enable efficient quantitative estimation of the thermal characteristics of the main spindle unit in an early stage of the development process. The presented work includes an experimental validation of the simulation model using a custom-built test rig. Specifically, the effects of the heat generated in the bearings and the heat flux from the bearing to the adjacent spindle system elements are investigated. Simulation and experimental results are compared and demonstrate good accordance. The proposed model is a useful, efficient and validated tool for quantitative simulation of thermal behavior of a main spindle system.

Keywords

Machine Tool, Thermal Behavior, Heat Transfer, Spindle Modelling

1. Introduction

The accuracy of a machine tool is influenced by its static, dynamic and thermo-elastic behavior [1]. The influence of a machine tool's thermal behavior on the machining process accuracy and repeatability has been the focus of research since the 1960s [2]. Main spindle systems contribute significantly to the final workpiece accuracy. The thermal properties of the main spindle unit as well as its rotational accuracy and static and dynamic stiffness affect the accuracy of the manufactured part. Advances in mechatronics have led to a higher precision of machine tools. However, the predictability of the thermal stability of machine tools becomes more critical,

especially to avoid costly design modifications in latter stages of machine development based on experimental studies [3]. Due to the significance of the main spindle's thermal behavior on machine tool manufacturing accuracy, a good understanding of the thermal interaction among different spindle components in practical spindle systems is required [4].

The spindle system's overall behavior strongly depends on the design and structure of the employed components. Particularly, the behavior of high-speed spindles can vary significantly according to the direction or arrangement of its components. For example, the arrangement of bearing sets, motor placement and motor coolant jackets in the housing, fits between bearings, housing design and bearing preload mainly influence the spindle behavior during high speed operation.

Spindle bearing friction is the main reason for bearing heat generation, which limits the maximum achievable spindle speed. As speed increases, there is an increase in generated heat in the bearings, the motor and the cutting surface. This additional heat causes thermal expansion. The amount of heat generated in the bearing should be estimated within the design process in order to choose the proper types of bearings and drives.

Angular contact ball bearings are most widely used for high speed spindles due to their properties under high speed conditions [5]. The bearings outer ring temperature is commonly used to detect early bearing failure, since expansion causes increased bearing loads and can lead to seizure [6].

Therefore, it is important to include the spindle's and bearing's thermal effects on the prediction of the overall response at elevated rotational speeds. The thermal models of bearing and spindle must be combined to provide a comprehensive representation of the heat transfer mechanisms.

Many attempts have been made to model the thermal behavior of the spindle and bearings for more than 60 years [7]. The Finite Element Method (FEM) is widely used to formulate the thermo-mechanical model of the bearing and spindle system [8]-[10]. The FEM models include the dynamic stiffness of the bearings, contact forces, temperature distributions and thermal expansions. The main drawback of this method is that it cannot be applied and adapted to different types of systems and conditions of operation in an efficient and fast manner. Besides, the standard software does not incorporate the models for numerical estimation of heat generation in the bearings and requires a lot of preparatory work [10].

In order to achieve a better understanding of the thermal behavior of the main spindle and its influence on the surroundings, it is not sufficient to use specially designed test rigs for the testing of individual spindle components. Beyond that, an examination of the entire spindle system is required. For this reason a custom-designed modular test rig was built. Different machine tool spindles can be operated on this test rig under a variety of testing conditions. During the early stages of design and as a part of the comprehensive effort to define the error caused by thermal expansion due to spindle operation, a thermal model of the spindle system was developed, following the purpose of increasing the efficiency of evaluating the thermal behavior of such spindle systems. Using the model and software developed, a fast numerical evaluation of the effect of machining operation parameters on the heat transfer mechanism within the main spindle system was obtained.

For the purpose of model validation, an experimental study was conducted using a custom-built test rig. In this paper the test rig configuration and experimental work for the externally driven spindle's thermal behavior are investigated. In this spindle as a special case study, the only heat generation sources are the bearings. The experimental work is followed by modelling and simulation of the thermal behavior of the spindle system using Wolfram Mathematica® software.

2. Simulation Model Development

The purpose of the spindle bearing model is the calculation of heat generation, heat transfer and temperature distribution over the spindle and housing elements. The developed model is coupled with the thermal models of the bearing to obtain a thermal response of the whole spindle system. The following assumptions were made [6]:

- Shaft and housing are assumed to be radially axisymmetric about the centerline of the spindle.
- The primary analysis will be one-dimensional in the axial direction. Radial heat loss in the housing will also be considered.
- Any heat generation (or cooling) is assumed to occur at the bearing contact zone, the center of the spindle (as in a centrally located motor), or the tips of the spindle (cutting heat and motor heat).

Although there are many simplifications, this heat transfer model can be used to examine the temperature fields in the bearings, housing and spindle shaft by examining the actual spindle bearing assembly. Without loss

of generality, the proposed heat transfer model is developed based on an externally driven grinding spindle with maximum spindle speed of 7000 rpm and a bearing bore diameter of 85 mm.

2.1. Bearing Heat Generation

The externally driven grinding spindle investigated in this research is assembled with sealed universal bearings with small steel balls. **Figure 1** illustrates the basic method used for defining heat transfer through the spindle and housing.

The heat is mainly generated in the contact between bearing raceways and balls due to frictional losses and rolling friction, influenced by speed, preload and lubricant. The cutting process is also a heat source. All three types of heat are the result of rotating motion and they are calculated from torque and speeds. The equation combining all three sources of heat is [8]:

$$q_{brg} = q_{brg\_l} + q_{brg\_v} + q_{brg\_s} \tag{1}$$

where q_{brg} is the heat flux from the bearing, l, v and s stand for load, viscosity and rotational speed, respectively. All three types of heat are generated at the contact between the ball surfaces and bearing raceways. Prior to the solution of thermal steady-state behavior, bearing heat generation must be calculated. According to Palmgren [7] and Harris [11], the total friction torque in the bearings is related to load, viscosity and rotational speed. By replacing the bearing loading forces with bearing contact loads expressions, the equations for inner and outer rolling resistance torque, $M_{i(\text{inner ring})}$, $M_{e(\text{outer ring})}$, related to load and viscosity at each ball can be expressed as [6]:

$$M_{fi} = M_{i(\text{inner ring})} = [0.675 f_0](\eta \omega_{roll})^{2/3} D^3 + f_1 \left(\frac{Q_i}{Q_{i\_max}} \right)^{1/3} Q_i D;$$

$$M_{fo} = M_{e(\text{outer ring})} = [0.675 f_0](\eta \omega_{roll})^{2/3} D^3 + f_1 \left(\frac{Q_e}{Q_{e\_max}} \right)^{1/3} Q_e D. \tag{2}$$

The bearing parameters f_0 and f_1 are dependent on the bearing and lubrication type. For angular contact ball bearings, f_0 is equal to 2.0 for grease and f_1 is equal to 0.001 for oil-air. The kinematic viscosity η (m^2/s) is found using the operating lubricant temperature. The ball's rolling velocity is

$$\omega_{roll} = \omega d_m / D \tag{3}$$

where D is the ball diameter (m), ω is the inner ring velocity (rad/s). Q_i, Q_e and Q_{i_max}, Q_{e_max} (N), are the ball contact load and ball maximum contact load at the contact point of the inner ring and the ball and the outer ring and the ball, respectively. The rotational speed related rolling resistance torque, $M_{si(\text{inner ring})}$, $M_{se(\text{outer ring})}$ can be expressed as:

$$M_{si(\text{inner})} = \frac{3k_s Q_i a_i \lambda}{8};$$

$$M_{se(\text{outer})} = \frac{3k_s Q_e a_e \lambda}{8} \tag{4}$$
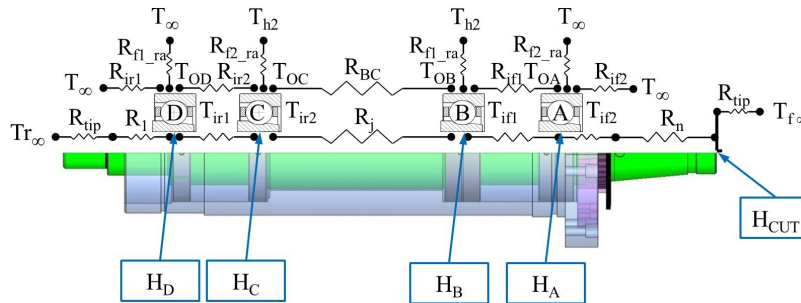

Figure 1. The investigated spindle and housing geometry definition including heat sources.

where k_s is the friction coefficient, a_e, a_i are the major axes of the inner and outer contact ellipse area and λ is the contact area geometry factor [6]. The heat generation in the inner ring H_{ik} and outer ring H_{ek} contact zone for ball k, is found from the power loss equation, where power is rotational speed times moment.

$$H_{ik} = \omega_{\text{roll}} M_{fi} + \omega_{si} M_{si};$$
$$H_{ek} = \omega_{\text{roll}} M_{fo} + \omega_{se} M_{se}. \tag{5}$$

It is assumed that for grease lubrication all heat generation enters the combined ball/ring network. The overall bearing heat generation is the summation of the contact heat generation at the inner and outer rings [8]. The housing will experience the heat generated from the outer ring H_e, while the spindle shaft receives heat generation from the inner ring H_i.

$$H_i = \sum_{k=1}^{N_b} H_{ik}, \quad H_e = \sum_{k=1}^{N_b} H_{ek} \tag{6}$$

where N_b is the total number of balls in the bearing. The Hertzian contact theory is used in this study to compute the contact loads Q_i, Q_e between bearing balls and bearing rings. The Hertzian inner and outer contact loads for each ball are shown in **Figure 2** and expressed by [12]-[14]:

$$Q_{ik} = K_i \delta_{ik}^{3/2};$$
$$Q_{ek} = K_o \delta_{ok}^{3/2} \tag{7}$$

where k is the ball index, δ_{ik}, δ_{ok} are the relative normal contact displacement between ball, inner and outer ring. K_i and K_o, the load deflection parameters can be obtained using the equations shown in Appendix A [11] [14].

The bearing displacement vector $\{\delta\}$ describes the three linear and two rotational motions, $\left[\delta_x \delta_y \delta_z \gamma_y \gamma_z\right]^{\text{T}}$. Due to the external forces applied to the bearings, displacements appear within the inner and outer rings and the distance between the curvature centers of the bearing rings change, as shown in **Figure 3** [14]. The relative displacement between the inner ring and the outer ring can be expressed as:

$$\Delta\delta_x = \delta_x^i - \delta_x^0; \quad \Delta\delta_y = \delta_y^i - \delta_y^o; \quad \Delta\delta_z = \delta_z^i - \delta_z^o; \quad \Delta\gamma_y = \gamma_y^i - \gamma_y^o; \quad \Delta\gamma_z = \gamma_z^i - \gamma_z^o \tag{8}$$

where B_d is the distance between the curvature centers of the inner and outer rings before deformation of the bearing and can be represented as:

$$B_d = BD = \left(f_o + f_i - 1\right)D. \tag{9}$$

$$f_i = \frac{r_{in}}{D}; \quad f_o = \frac{r_{ou}}{D} \tag{10}$$

As shown in **Figure 3**, where k is the ball index. Δ_{ik} and Δ_{ok} can be expressed as:

$$\Delta_{ik} = r_{in} - D/2 + \delta_{ik} = \left(f_{in} - 0.5\right)D + \delta_{ik};$$
$$\Delta_{ok} = r_{ou} - D/2 + \delta_{ok} = \left(f_o - 0.5\right)D + \delta_{ok}. \tag{11}$$

The following relations are derived from **Figure 3**:

$$\sin\theta_{ok} = \frac{U_k}{\left(r_o/D - 0.5D\right) + \delta_{ok}};$$

$$\cos\theta_{ok} = \frac{V_k}{\left(r_o/D - 0.5D\right) + \delta_{ok}};$$

$$\sin\theta_{ik} = \frac{U_{ik} - U_k}{\left(r_i/D - 0.5D\right) + \delta_{ik}};$$

$$\cos\theta_{ik} = \frac{V_{ik} - V_k}{\left(r_i/D - 0.5D\right) + \delta_{ik}}. \tag{12}$$

The equilibrium equations for the bearing ball as shown in **Figure 2(c)** can be expressed as [14]:
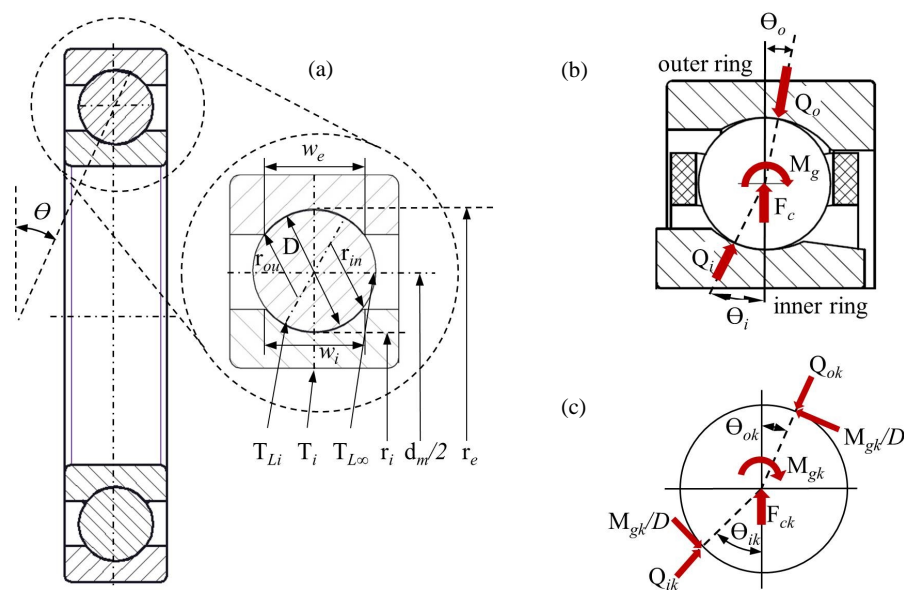
Figure 2. Angular contact ball bearings. (a) Geometry and temperature terms; (b) Loads diagram [15]; (c) Forces acting on the bearing ball [14].
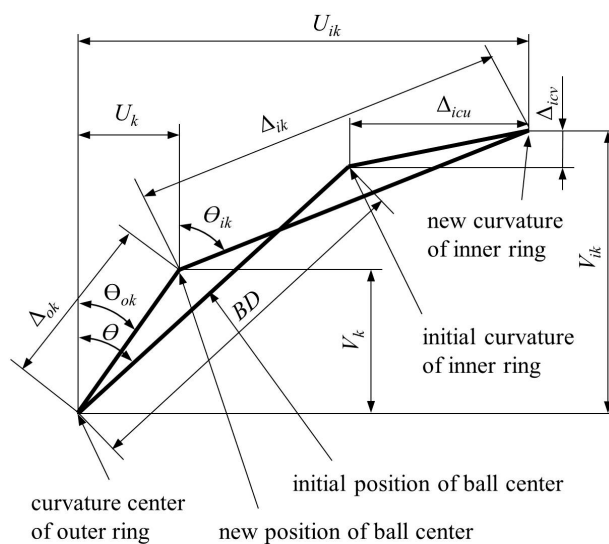

Figure 3. Displacement relationships between curvature centers of bearing rings under load [14].

$$Q_{ok}\cos\theta_{ok} - \frac{M_{gk}}{D}\sin\theta_{ok} - Q_{ik}\cos\theta_{ik} + \frac{M_{gk}}{D}\sin\theta_{ik} - F_{ck} = 0;$$

$$Q_{ok}\sin\theta_{ok} + \frac{M_{gk}}{D}\cos\theta_{ok} - Q_{ik}\sin\theta_{ik} - \frac{M_{gk}}{D}\cos\theta_{ik} = 0$$

(13)

where F_{ck} and M_{gk} are the centrifugal force and gyroscopic moment, respectively [14]. The displacement equations for the bearing ball as shown in **Figure 3** can be expressed as [14]:

$$\left(U_{ik} - U_k\right)^2 + \left(V_{ik} - V_k\right)^2 - \Delta_{ik}^2 = 0;$$

$$\left(U_k\right)^2 + \left(V_k\right)^2 - \Delta_{ok}^2 = 0.$$

(14)

For $\varphi_k = 2\pi k/N$, N is the total number of balls in the bearing, θ is the initial contact angle, U_{ik}, V_{ik} in Equation (14) will be:

$$U_{ik} = BD\sin\theta + \Delta\delta_x - \Delta\gamma_z r_{ic}\cos\varphi_k + \Delta\gamma_y r_{ic}\sin\varphi_k;$$
$$V_{ik} = BD\cos\theta + \Delta\delta_y\cos\varphi_k + \Delta\delta_z\sin\varphi_k. \tag{15}$$

$$r_{ic} = \frac{D_m}{2} + (f_i - 0.5)D\cos\theta \tag{16}$$

r_{ic} in Equation (16) is the radius from the bearing center to the inner ring curvature center shown in **Figure 2(a)**. The four unknown parameters, $\{\delta_k\} = \{U_k, V_k, \delta_{ok}, \delta_{ik}\}^{\mathrm{T}}$ must be obtained by solving Equations (13), (14) iteratively using the approach [14]:

$$\{\delta_k^{(n+1)}\} = \{\delta_k^{(n)}\} - [a_{ij}^{(n)}]^{-1}\{\varepsilon_k^{(n)}\}; \quad (n = 0,1,2,\cdots; \ i,j = 1,2,3,4) \tag{17}$$

$\{\varepsilon_k\} = \{\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4\}^{\mathrm{T}}$ is the error vector and should converge to a specifically defined small value. The coefficients matrix $[a_{ij}]$ can be expressed as:

$$[a_{ij}] = \left[\frac{\partial\{\varepsilon_k\}_i}{\partial\{\delta_k\}_{ij}}\right]; \quad (i,j = 1,2,3,4). \tag{18}$$

The elements of the error vector $\{\varepsilon_k\}$ are:

$$(U_{ik} - U_k)^2 + (V_{ik} - V_k)^2 - \Delta_{ik}^2 = \varepsilon_1;$$
$$(U_k)^2 + (V_k)^2 - \Delta_{ok}^2 = \varepsilon_2;$$
$$Q_{ok}\cos\theta_{ok} - \frac{M_{gk}}{D}\sin\theta_{ok} - Q_{ik}\cos\theta_{ik} + \frac{M_{gk}}{D}\sin\theta_{ik} - F_{ck} = \varepsilon_3; \tag{19}$$
$$Q_{ok}\sin\theta_{ok} + \frac{M_{gk}}{D}\cos\theta_{ok} - Q_{ik}\sin\theta_{ik} - \frac{M_{gk}}{D}\cos\theta_{ik} = \varepsilon_4.$$

The bearing's contact loads Q_{ik}, Q_{ok} in Equation (7) are obtained using the calculated values of $\{\delta_k\}$. The bearing heat generation can be computed by Equations (2)-(6) using the calculated variables values. Knowing the bearing heat generation quantity, we can move forward to the thermal model of spindle and housing.

2.2. Spindle and Housing Thermal Model

The developed model is based on representing the spindle and the housing using axisymmetric thermal resistance elements. For the solution, the input parameters are: geometry, material parameters, air temperature, heat generation due to the bearings and initial temperature of the system. The thermal resistance is calculated for each housing and spindle element. Linear thermal resistance is defined according to the following equation [6]:

$$R = \frac{L}{KA} \tag{20}$$

where L is the element length, K is the material thermal conductivity and A is the cross-sectional area.

2.2.1. Bearing Heat Transfer Model
Figure 4 shows the bearing geometry and temperature nodes definition as well as the bearing heat transfer model. The externally driven spindle is grease lubricated. With grease-packed lubrication, heat convection is less significant and conduction is more prominent. The radial heat transfer through the spindle, rolling element and housing is described similar to the thermal resistance network described by Jorgensen [6] and is shown in **Figure 4**.
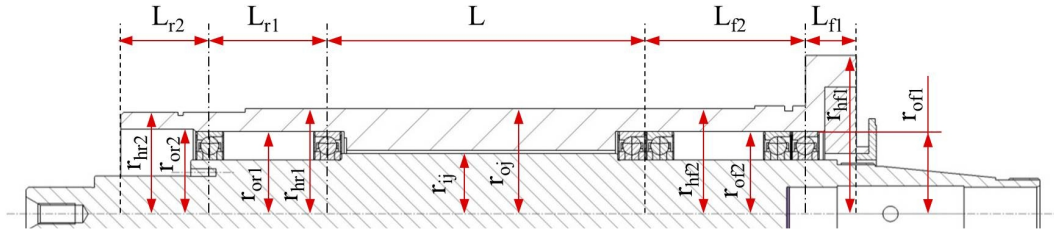
Figure 4. Bearing heat resistances model.

$$R_{Li} = \frac{r_b}{k_l\left(\frac{2\pi}{n}r_iW_i - \pi r_b^2\right)}; \quad R_{Le} = \frac{r_b}{k_l\left(\frac{2\pi}{n}r_eW_e - \pi r_b^2\right)}; \quad R_b = \frac{1}{k_b\pi r_b}; \quad R_1 = \frac{1}{\frac{1}{R_{Le}} + \frac{1}{R_b/2}} = \frac{R_{Le}\left(R_b/2\right)}{R_{Le} + R_b/2};$$

$$R_2 = \frac{1}{\frac{1}{R_{Li}} + \frac{1}{R_b/2}} = \frac{R_{Li}\left(R_b/2\right)}{R_{Li} + R_b/2}; \quad R_e = \frac{1}{k_e}\frac{n\ln\left(r_o/r_e\right)}{2\pi W}; \quad R_i = \frac{1}{k_i}\frac{n\ln\left(r_i/r_s\right)}{2\pi W}. \tag{21}$$

The heat transfer equations use temperatures at the edge of the inner and outer rings as boundary conditions. As an alternative solution method, the outer ring temperature can be easily measured and applied to the solution, allowing the housing approximations to be ignored. Steady state outer ring temperatures can be used to provide a reference point for the thermal equations [6]. The matrix form of the heat transfer equations can be expressed as:

$$\left[R_b\right]\{T\} = \{H\} \tag{22}$$

where $\left[R_b\right]$ is the system heat resistance matrix; $\{T\}$ is the vector of unknown temperatures, in this case three unknowns, and $\{H\}$ are the heat generation inputs. As shown in **Figure 4**, Equation (22) can be written as:

$$\begin{bmatrix} R_{b11} & R_{b12} & 0 \\ R_{b21} & R_{b22} & R_{b23} \\ 0 & R_{b32} & R_{b33} \end{bmatrix} \begin{Bmatrix} T_{Le} \\ T_b \\ T_{Li} \end{Bmatrix} = \begin{Bmatrix} H_{Le} \\ H_b \\ H_{Li} \end{Bmatrix} = \begin{Bmatrix} H_e + T_O/R_e \\ 0.0 \\ H_i + T_i/R_i \end{Bmatrix} \tag{23}$$

The resistance matrix elements, R_{bij} as well as the expressions for the generated heat can be found in Appendix B. Solving Equation (22) allows to predict the contact points' temperatures T_{Le}, T_{Li} as well as the bearing ball temperature T_b. Although the heat transfer equations are approximations, sufficient accuracy is attainable for thermal expansion predictions [6].

2.2.2. Housing Heat Transfer Model

The housing heat transfer model based on quasi-two-dimensional analysis of heat flow is shown in **Figure 1**. The geometry parameters are shown in **Figure 5**. The four unknown nodal points' temperatures, T_{oA}, T_{oB}, T_{oC}, T_{oD}, shown in **Figure 1**, were calculated by applying axial and radial conduction and convection resistance elements. The radial and axial resistances for an element can be expressed as:

$$R_{rad} = \frac{\ln\left(\frac{r_h}{r_o}\right)}{k_h 2\pi L} + \frac{1}{h_{air}L2\pi r_h}; \tag{24}$$

$$R_{ax} = \frac{L}{k_h\pi\left(r_h^2 - r_o^2\right)} + \frac{1}{h_{air}\pi\left(r_h^2 - r_o^2\right)} \tag{25}$$

where r_h and r_o are the outer (housing) and inner (housing) element radius. L is the element length, k_h is the housing element thermal conductivity and k_{air} is the heat convection coefficient in air. The first part in Equations (24) and (25) is the radial and axial conduction resistance, respectively. The second part in Equations (24) and (25) is the radial and axial convection resistance. The second parts will be equal to zero if there are no free
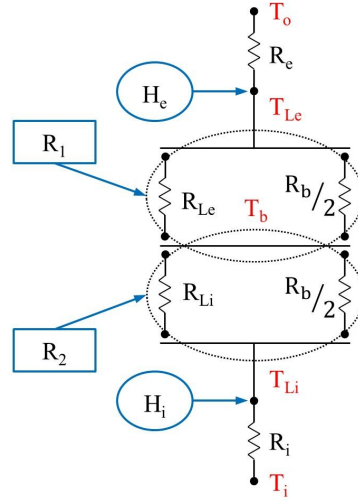
Figure 5. Spindle and housing 2D configuration.

ends in the element. In order to simplify the heat transfer equilibrium equation the axial and radial element resistance are combined as shown in the following equation:

$$R_{\text{total}} = \frac{R_{ax} \times R_{\text{rad}}}{R_{\text{rad}} + R_{ax}} \tag{26}$$

R_{total} is the element total resistance. The housing heat transfer equations in matrix can be expressed as:

$$[R_h]\{T\} = \{H\} \tag{27}$$

where R_A and R_D are the total resistance of the front and rear housing elements. R_{BC_rad} is the resistance of the connection element between points B and C. The resistance matrix elements R_{hij} as well as the expressions for generated heat and the resistances R_A, R_D and R_{BC_rad} can be found in Appendix C. By solving the set of Equation (27), the temperature values at the four most heated points along the housing outside surface are obtained.

$$\begin{bmatrix} R_{h11} & R_{h12} & 0 & 0 \\ R_{h21} & R_{h22} & R_{h23} & 0 \\ 0 & R_{h32} & R_{h33} & R_{h34} \\ 0 & 0 & R_{h43} & R_{h44} \end{bmatrix} \begin{Bmatrix} T_{OA} \\ T_{OB} \\ T_{OC} \\ T_{OD} \end{Bmatrix} = \begin{Bmatrix} H_{OA} + \dfrac{T_{\infty}}{R_A} \\ H_{OB} + \dfrac{T_{h2}}{R_{BC\_\text{rad}}} \\ H_{OC} + \dfrac{T_{h2}}{R_{BC\_\text{rad}}} \\ H_{OD} + \dfrac{T_{\infty}}{R_D} \end{Bmatrix} \tag{28}$$

2.2.3. Spindle Shaft Heat Transfer Model

As shown in **Figure 6**, the spindle shaft is represented by a nine resistance element model. By applying the same matrix notation as above, the matrix notation of the heat generation balance is obtained:

$$[R_i]\{T\} = \{H\} \tag{29}$$

where $[R_i]$ is the spindle shaft heat resistance matrix, $\{T\}$ is the vector of unknown temperatures in points of interest along the shaft axis, in this case six unknowns, and $\{H\}$ is the heat generation inputs. The spindle shaft heat transfer set of equations can be expressed as:
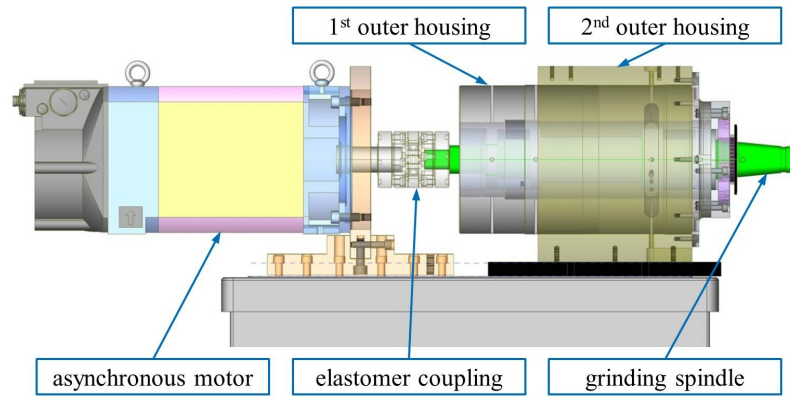
Figure 6. Spindle shaft heat transfer net.

$$
\begin{bmatrix}
R_{i11} & R_{i12} & & & & \\
R_{i21} & R_{i22} & R_{i23} & & 0 & \\
0 & R_{i32} & R_{i33} & R_{i34} & & \\
& & R_{i43} & R_{i44} & R_{i45} & \\
& 0 & & R_{i54} & R_{i55} & R_{i56} \\
& & & & R_{i65} & R_{i66}
\end{bmatrix}
\begin{Bmatrix}
T_{tail} \\
T_{ir1} \\
T_{ir2} \\
T_{if1} \\
T_{if2} \\
T_{tip}
\end{Bmatrix}
=
\begin{Bmatrix}
T_{r\infty}/R_{tail} \\
H_{iD} \\
H_{iC} \\
H_{iB} \\
H_{iA} \\
T_{f\infty}/R_{tip}
\end{Bmatrix}
\tag{30}
$$

All matrix and vector elements presented in Equation (30) are expressed in Appendix D. By solving the set of Equation (30), the temperature values in six points of interest along the spindle shaft centre are obtained.

3. Experimental Setup for Spindle System Analysis

The configuration of the test rig used for experimental investigations of an externally driven spindle is shown in **Figure 7** and **Figure 8**. An air-cooled asynchronous motor is connected via an elastomer coupling to the spindle. In addition to the original spindle housing two adjacent outer housings were used to mount the spindle on to the test rig.

The motor, friction in the bearings and the cutting process are the main causes of heat generated by the spindle system during machining operation. The motor is cooled using air flow in the direction from the drive end of the motor to the non-drive end. By assuming that most of the cutting heat is taken out by the coolant and that the heat generated by the motor in the configuration shown in **Figure 7** and **Figure 8** is kept away from the machine tool spindle, it follows that the heat generated by the bearings is the dominant heat source causing the thermal distortion of the spindle shaft. The motor's maximum continuous duty speed is 7000 rpm, with a rated power output of 12 kW. During machining operations such as external cylindrical rough grinding high cutting capacity is required, while during fine grinding a high standard of form and surface quality should be maintained. In order to achieve both mentioned requirements a high degree of rigidity and running accuracy as well as good damping and speed suitability are the main criteria for the spindle bearing arrangement. These requirements are met by the precision bearings assembled in the test spindle. Sealed universal spindle bearings with small steel balls (HSS) are used:

- at the tool end: 1 spindle bearing set, in double O arrangement (Tandem), as fixed bearings
- at the drive end: 1 spindle bearing set, in O arrangement, as floating bearings

The employed contact angle of 15° is suitable for high radial rigidity. The universal design bearings are lightly preloaded. The sealed spindle bearings require no maintenance and are lubricated for life with roller bearing grease [5].

Studying the heat flux mechanisms during spindle operation, the most interesting spindle elements are the spindle bearings, housing and shaft. In order to measure the temperature distribution in axial and radial direction and investigate the heat flux into the machine tool components adjacent to the spindle elements, the spindle system was divided into four cross-sections for temperature measurements. **Figure 9** illustrates the measurement sections that were positioned at the bearings axial positions along the spindle system's rotational axis. In radial direction the sensors were positioned along the normal direction of the spindle's outer cylindrical surface, on circles of constant diameter, beginning from the outer ring of the bearing through the original spindle housing and the 1st outer housing up to the 2nd outer housing supporting the spindle as shown in **Figure 10**. The data acquisition system enables simultaneous measurement, recording and display of the output temperatures from a total of 31 Pt-100 resistance temperature sensors attached in 4-Wire configuration. All 31 sensor positions are shown in

Figure 7. Test rig configuration for the externally driven grinding spindle.
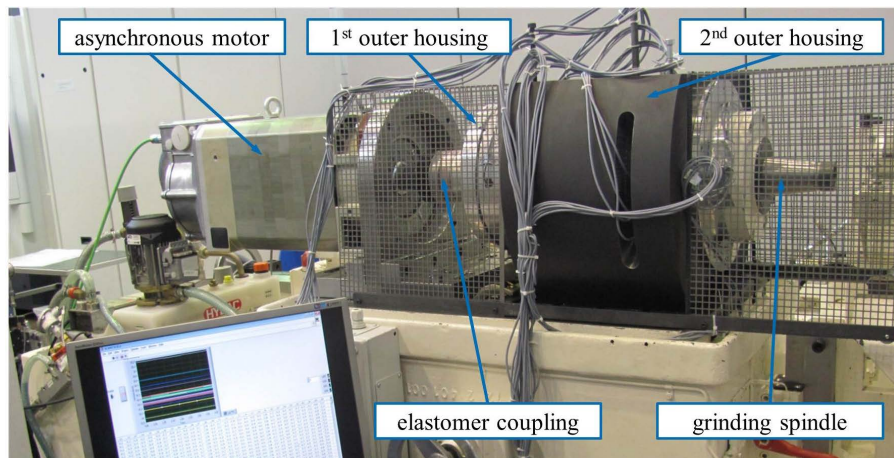
Figure 8. Experimental setup of the sensors and the data acquisition arrangement.
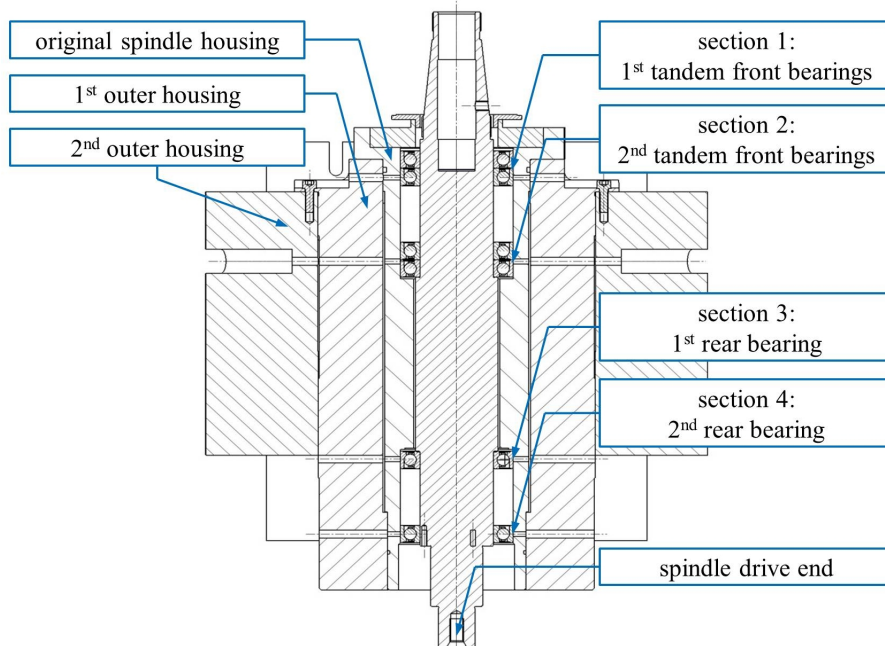
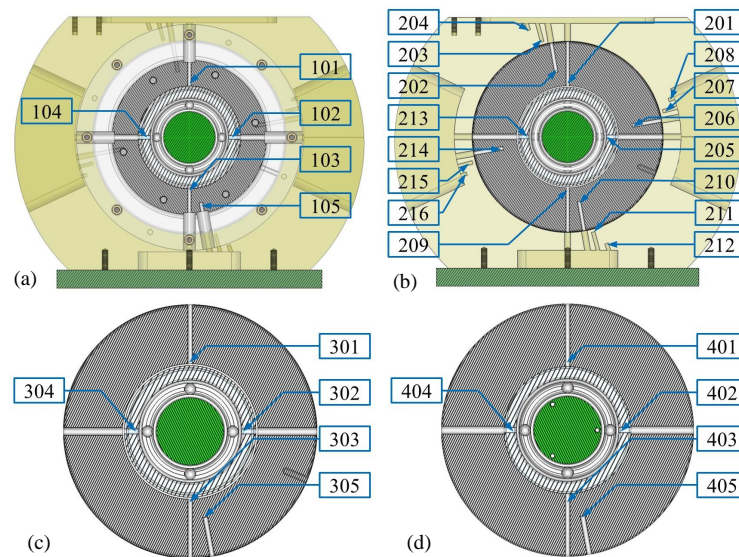**Figure 9.** Measuring sections along the spindle axis.

Figure 10. Sensor positions: (a) Section 1; (b) Section 2; (c) Section 3; (d) Section 4.

Figure 10 and the wiring is shown in **Figure 8**. As mentioned before, the only heat source within the spindle system, during the externally driven spindle operation in the configuration described above is the bearing friction. The heat generation of internal components as a function of spindle speed is only investigated using the test rig described. The research phase is applied in order to gain fundamental knowledge about the heat generation in the spindle system with less complexity. The next step of the investigations will include load-dependent heat generation and heat flux as well as the motorized spindle system's thermal behavior.

4. Spindle System Thermal Behavior—Experimental Study

The thermo-elastic behavior of a machine tool is strongly nonlinear and has to be modelled for several operating points. Hence, the speed spectrum has to be divided into several sections, each having its own operating point [2]. Sensitivity analysis under variety of operating modes has been designed and carried out. The spindle speed spectrum up to 7000 rpm is divided into seven operating points starting at 1000 rpm with steps of 1000 rpm. The duration of a single step was chosen to be 4 h. The sensor area with the highest temperature in each measurement section was the one located on the bearing outer ring. **Figure 11** illustrates the highest measured temperature obtained at spindle speeds from 1000 to 7000 rpm. The sensors mounted in Section 2 present the highest temperature over all spindle speeds. **Figure 12** illustrates the steady state temperatures measured in the sections mentioned above according to 3 sensor measurements in each section at 7000 rpm spindle speed. Sensor no. 205 mounted in section 2 at the position directly adjacent to the bearing outer ring. In this position the highest temperature (38.2˚C) was achieved. The second highest temperature, 35.4˚C, was measured by sensor no. 302 and 31.8˚C was measured by sensor no. 102. All 3 sensors mentioned above were mounted in touch with the bearing's outer ring in each section. As shown in **Figure 10**, 16 sensors are mounted in Section 2. The graph in **Figure 13** represents the temperature measured in 5 of the 16 different sensor positions in Section 2. All points across the section from the deepest measuring point, the outer bearing ring, through all 3 housing elements, to the outermost measuring point (sensor no. 204), were heated. Even the outmost point on the outer housing in the 2nd section was affected by the heat generated due to the bearing friction.

By changing the spindle speed over the speed spectrum, from 1000 rpm to 7000 rpm, it is clear from **Figure 14** that the highest rate of change in the maximum temperature occurs when the rotational speed is increased from 5000 to 7000 rpm. **Figure 15** illustrates the influence of spindle speed on the measured temperature in Section 2, sensors no. 201 to 205. As already mentioned, the thermo-elastic behavior of the main spindle system in a machine tool is strongly nonlinear and has to be investigated and modelled in several simplifying steps. First, the thermal behavior of an externally driven spindle was investigated in order to verify and apply a simple simulation model, based only on the heat generated by the friction in the bearings. The heat flux from the bearings into
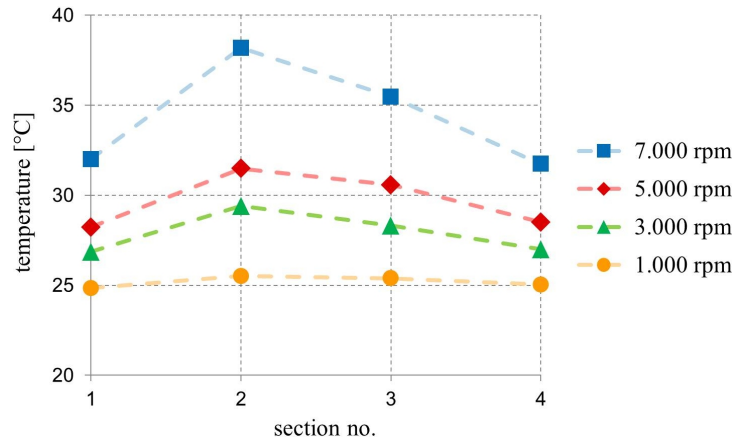
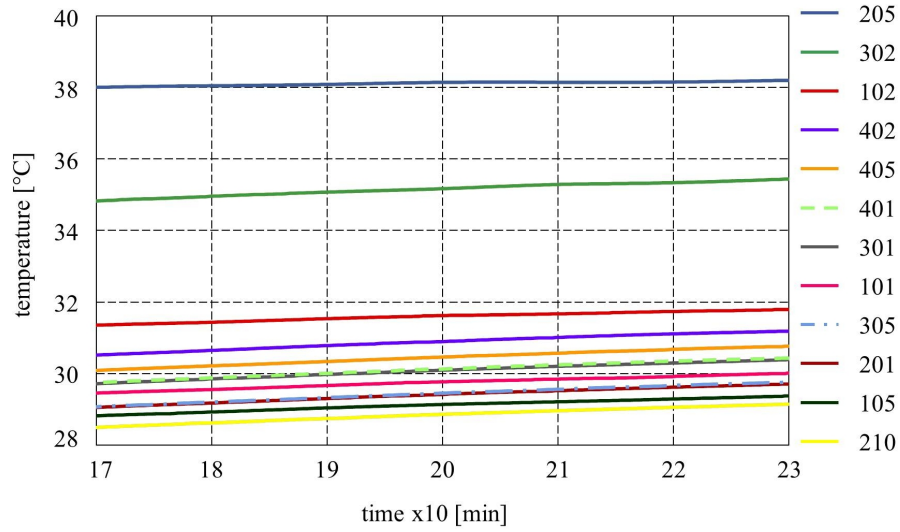**Figure 11.** Measured temperature in the most heated sensors among all sections over changing spindle speeds.

Figure 12. Measured temperature at steady state in 4 Sections; 3 sensors each section.
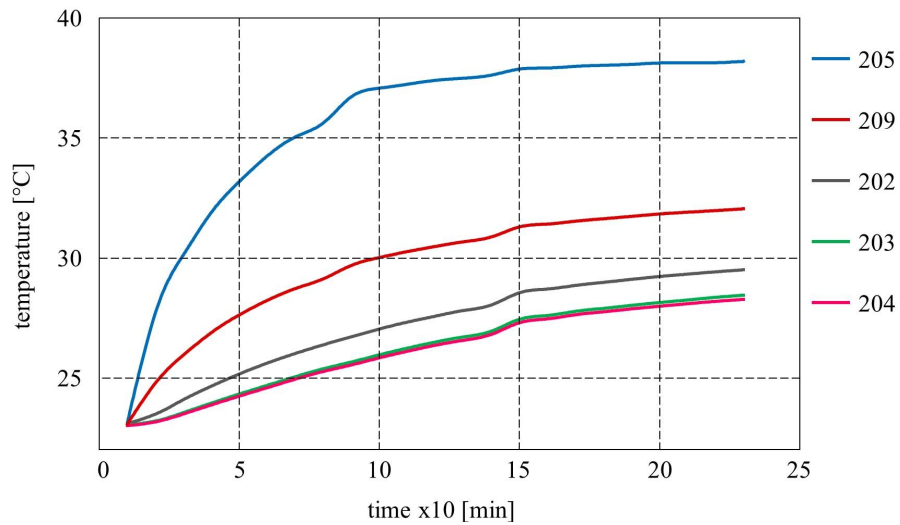

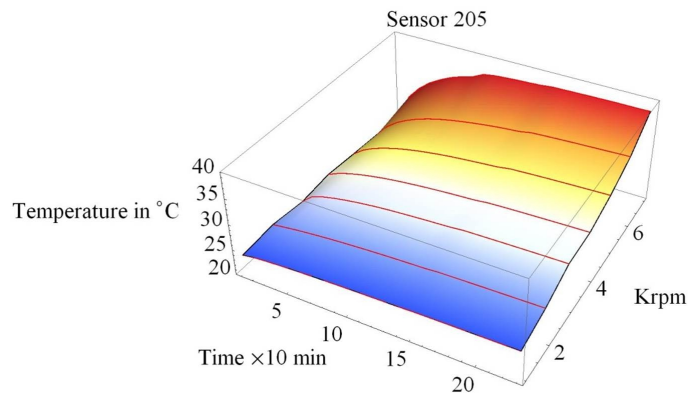**Figure 13.** Temperature measured by 5 out of 16 sensors in Section 2.

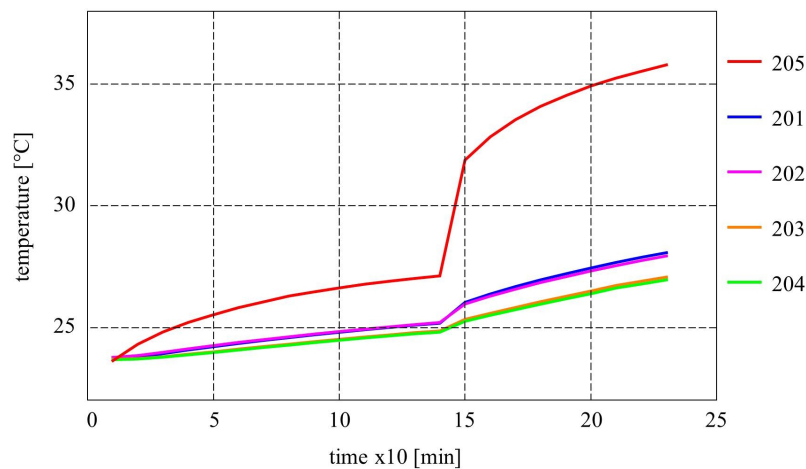**Figure 14.** Temperature measured by sensor no. 205, as a function of time and spindle speed.

Figure 15. Measured temperature in Section 2 while changing the spindle speed from 3.000 to 7.000 rpm.

several outer housings is investigated as well. By verifying the heat transfer from the inner bearings element to the outer housings, one is able to understand the heat flux mechanism and the influence of the heat generated in the bearings on the spindle-adjacent machine elements.

5. Numerical Simulations

Based on the model described in Section 2 and Appendices A to D, a simulation program was developed. The numerical computations as well as the graphical presentations and investigations were carried out using the Mathematica® software [16]. Using the algorithm and software developed, numerical studies of the heat transfer mechanism and temperature variations along the external driven spindle system were made. Some simplifications were made in the simulation programs above. No motor and no external cutting load exists in this stage of the study, in order to enable the use of this model as a base line for further simulations including the cutting process effects and motorized spindle simulations.

5.1. Validation of the Simulation Model

Due to the fact that a commercial grinding machine spindle was used for the investigations, as described already in Section 3, in this stage of the experimental studies, we focused our experiments on the heat transfer mechanism from the bearings towards the outer housing elements as a function of spindle rotational speed. As stated, the spindle's original geometry, preload, lubrication viscosity, structure or functional elements can not be varied during the investigations. The simulation results are compared with the results obtained in the experimental stu-

dies described in Section 4. The numerical simulations investigate the effect of operating conditions such as rotational speed (rpm), axial preload, lubricant viscosity, as well as geometrical parameters of the spindle system structural components. Comparison between measured temperatures at the housing outer surface, sensors 101, 201, 301 and 401, and simulated temperature along the axis of the spindle system in the bearings' positions is shown in **Figure 16** and **Figure 17** for different ranges of spindle rotational speeds.

From **Figure 16** and **Figure 17** it follows that there is a good fit between the simulation results and the experimental results obtained. It can be seen that the differences are in the range of 1˚C to 2˚C over the full range of spindle speeds. It should be mentioned that the main goal of the simulation model is to be fast and efficient and should give first approximation only during an early stage of spindle design. It should not be accurate as a FEM or other simulation software programs, which are much more complex and time consuming.

5.2. Simulations under Speed Variations

Internal heat generation and heat flow is strongly speed-dependent. **Figure 18** shows the temperature developed in the spindle housing outer surface, T_{oA}, T_{oB}, T_{oC}, T_{oD}, at the sections shown in **Figure 10**. The results can be achieved only by simulation because the spindle and bearings speed limit is 7000 rpm.

The temperature profiles in **Figure 19** identify the temperatures T_{le} in the ball bearings' point of contact with the outer and inner rings. These contact points show the highest temperature within the spindle system.
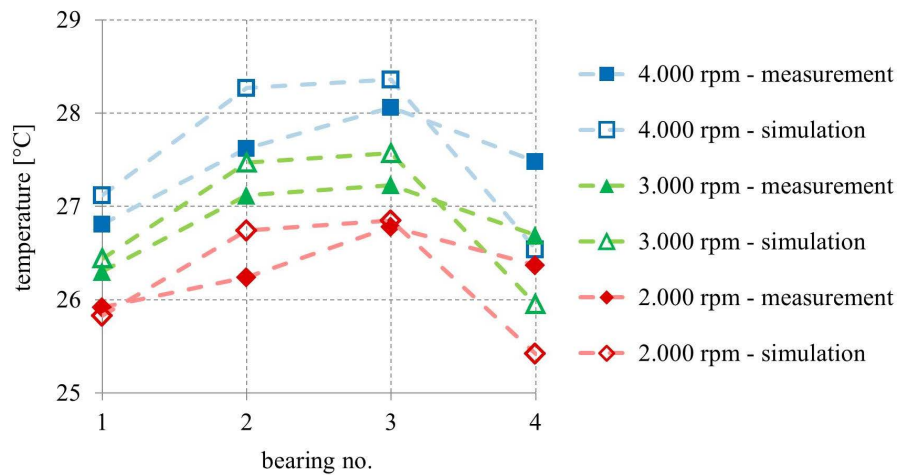
Figure 16. Comparison between measured temperatures and simulated temperature as a function of spindle speeds between 2.000 and 4.000 rpm.
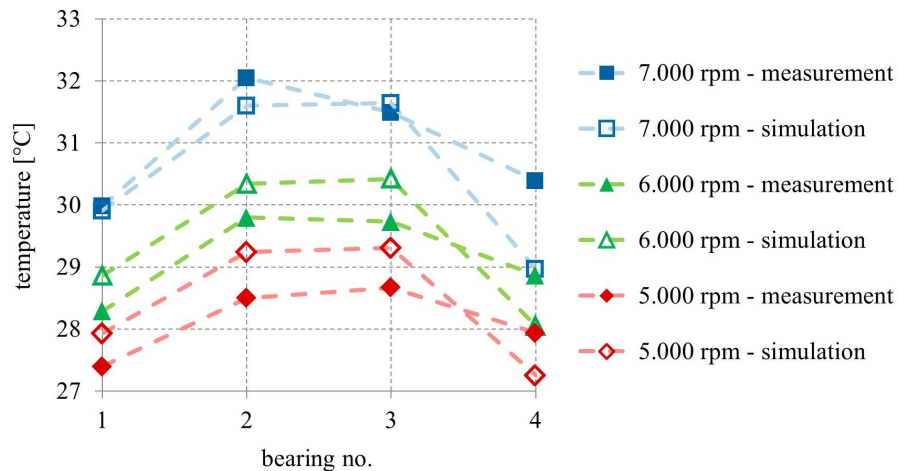

Figure 17. Comparison between measured temperatures and simulated temperature as a function of spindle speeds between 5.000 and 7.000 rpm.

Figure 20 depicts the heat generation H_e at the bearing outer ring due to the frictional moments between the bearing balls and the bearing outer ring. According to the developed model and as mentioned in the literature [6], it is assumed that part of the heat generated in the contact flows into the bearing ring and another part flows into the bearing ball. **Figure 21** shows the temperature profile along points of interest along the spindle shaft axis, as shown in **Figure 6**, for various spindle rotational speeds.
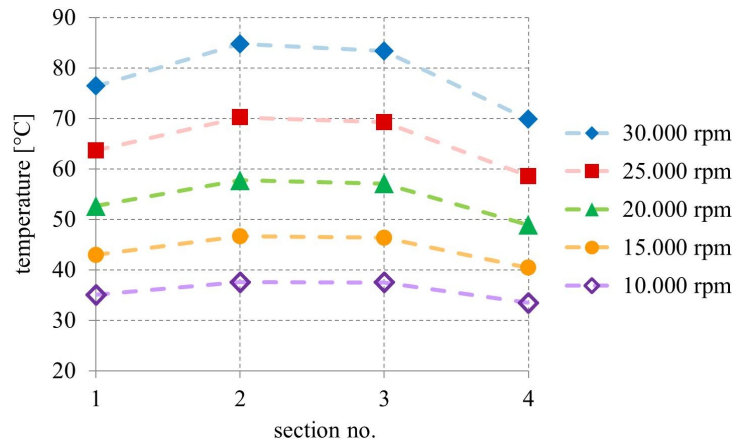
Figure 18. Simulated temperature at changed spindle speed up to 30.000 rpm.
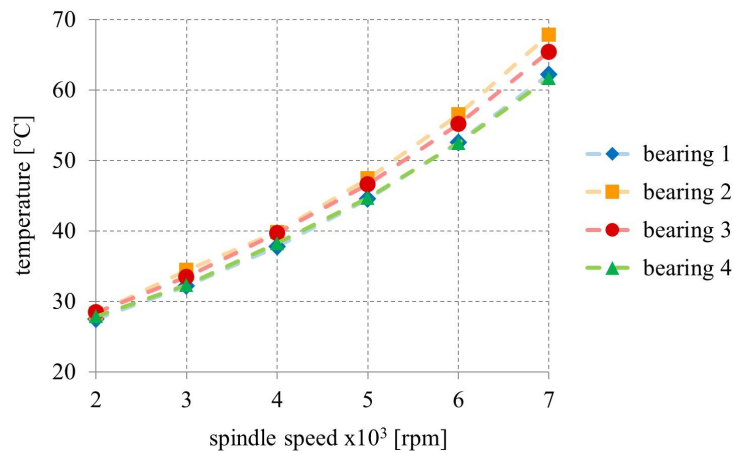
Figure 19. Simulated temperature T_{le} at the contact point between the bearing ball and the bearing outer ring.
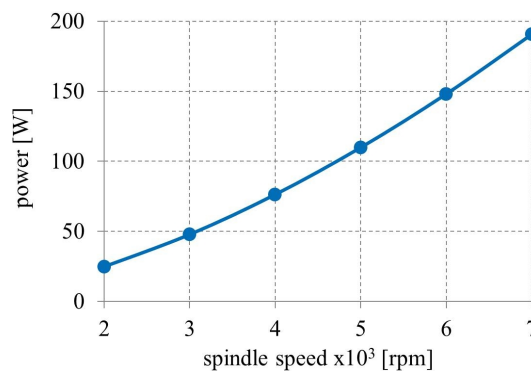
Figure 20. Calculated heat generation at the bearing outer ring as a function of spindle rotational speed.
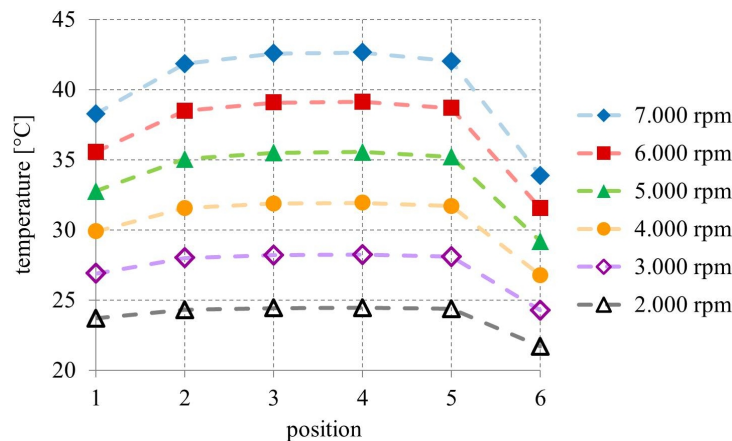
Figure 21. Simulated temperature at 6 points along the spindle shaft axis. T_{tail}, T_{ir1}, T_{ir2}, T_{if1}, T_{if2}, T_{tip} over spindle speeds spectrum from 2.000 to 7.000 rpm.

6. Conclusions

The theoretical and experimental investigations of the heat transfer mechanism in the externally driven grinding spindle have enabled the following conclusions.

1) The use of the bearing heat generation terms in each model, bearing model, spindle model and housing model enabled investigation of the thermal behavior of the entire spindle system under the effect of the main source of heat generation: the bearings' friction.

2) Experimental investigation of the heat transfer in the spindle system and the machine elements adjacent to the spindle allows obtaining the data needed for better understanding the heat generation and flux due to the bearing friction.

3) The simplifying assumptions of the method presented in this study are mainly the one-dimensional heat flow approximations, simplified housing geometry, external drive motor, excluding the cutting process heat and excluding cutting process loads. These spindle-housing approximations are acceptable since the most important temperature predictions are those within the bearings.

4) Using the models presented, the temperature field can be numerically found for most spindle-housing configurations in a fast and efficient manner. The simulated and validated temperature fields can be used for numerical simulations of spindles thermal expansion and machine elements adjacent to the spindle, as well as for machine tool precision prediction during the development process.

Acknowledgements

References

[1]   Brecher, C. and Wissmann, A. (2011) Compensation of Thermo-Dependent Machine Tool Deformations Due to Spindle Load: Investigation of the Optimal Transfer Function in Consideration of Rough Machining. *Production Engineering*, **5**, 565-574. http://dx.doi.org/10.1007/s11740-011-0311-4

[2]   Bryan, J. (1990) International Status of Thermal Error Research. *CIRP Annals—Manufacturing Technology*, **39**, 645-656. http://dx.doi.org/10.1016/S0007-8506(07)63001-7

[3]   Mayr, J., Jedrzejewski, J., Uhlmann, E., Donmez, M.A., Knapp, W., Härtig, F., *et al.* (2012) Thermal Issues in Machine Tools. *CIRP Annals—Manufacturing Technology*, **61**, 771-791. http://dx.doi.org/10.1016/j.cirp.2012.05.008

[4]   Brecher, C., Fey, M., Neus, S., Shneor, Y. and Bakarinow, K. (2014) Influences on the Thermal Behavior of Linear Guides and Externally Driven Spindle Systems. *Production Engineering*, **9**, 133-141.

[5] Li, H. and Shin, Y.C. (2004) Analysis of Bearing Configuration Effects on High Speed Spindles Using an Integrated Dynamic Thermo-Mechanical Spindle Model. *International Journal of Machine Tools and Manufacture*, **44**, 347-364. http://dx.doi.org/10.1016/j.ijmachtools.2003.10.011

[6] Jorgensen, B. (1996) Robust Modeling of High-Speed Spindle Bearing Dynamics under Operating Conditions. Ph.D. Thesis, Purdue University, West Lafayette.

[7] Palmgren, A. (1959) Ball and Roller Bearing Engineering. 4th Edition, S.H. Burbank, Philadelphia.

[8] Bossmans, B. and Tu, J. (1999) A Thermal Model for High Speed Motorized Spindles. *International Journal of Machine Tools and Manufacture*, **39**, 1345-1366. http://dx.doi.org/10.1016/S0890-6955(99)00005-X

[9] Li, H. and Shin, Y.C. (2004) Integrated Dynamic Thermo-Mechanical Modeling of High Speed Spindles. Part 1: Model Development. *Journal of Manufacturing Science and Engineering*, **126**, 148-158. http://dx.doi.org/10.1115/1.1644545

[10] Kim, J.D., Zverv, I. and Lee, K.B. (2003) Thermal Model of High-Speed Spindle Units. *KSME International Journal*, **17**, 668-678.

[11] Harris, T.A. (1991) Rolling Bearing Analysis. 3rd Edition, John Wiley & Sons, Hoboken.

[12] Zhao, H.T., Yang, J.G. and Shen, J.H. (2007) Simulation of Thermal Behavior of a CNC Machine Tool Spindle. *International Journal of Machine Tools and Manufacture*, **47**, 1003-1010.

[13] Harris, T.A. and Kotzalas, M.N. (2005) Advanced Concepts of Bearing Technology, Rolling Bearing Analysis. 5th Edition, Taylor & Francis Corporations, London.

[14] Cao, Y. (2006) Modeling of High-Speed Machine-Tool Spindle System. Ph.D. Thesis, British Columbia University, British Columbia.

[15] Altintas, Y. and Cao, Y. (2005) Virtual Design and Optimization of Machine Tool Spindles. *CIRP Annals—Manufacturing Technology*, **54**, 379-382.

[16] Wolfram Research (2012) Wolfram Mathematica® 9. Documentation Center.

Appendix A

$$K = \frac{\pi \kappa E'}{3F} \sqrt{\frac{2ER_{xy}}{F}};$$

$$\kappa = 1.0339 \left(\frac{R_y}{R_x} \right)^{0.636};$$

$$F = 1.5277 + 0.6023 \ln \left(\frac{R_x}{R_y} \right);$$

$$E = 1.0003 + 0.5968 \left(\frac{R_x}{R_y} \right);$$

$$E' = \frac{2}{\left(\frac{\left(1 - \mu_a^2\right)}{Ea} + \frac{\left(1 - \mu_b^2\right)}{Eb} \right)};$$

$$R_{xy} = \frac{R_x R_y}{\left(R_x + R_y \right)}.$$

(A.1)

where μ_a, μ_b are Poisson's ratio of the bearing ball and the rings, respectively. For the inner and outer ring contacts, R_x and R_y are:

$$R_{xik} = \frac{D\left(1 - \gamma_{ik}\right)}{2}; \quad R_{xok} = \frac{D \cdot f_{ou}}{2f_{ou} - 1};$$

$$R_{yik} = \frac{D \cdot f_{in}}{2f_{in} - 1}; \quad R_{yok} = \frac{D\left(1 + \gamma_{ok}\right)}{2};$$

$$\gamma_{ik} = \frac{D\cos\theta_{ik}}{D_m}; \quad \gamma_{ok} = \frac{D\cos\theta_{ok}}{D_m}.$$

(A.2)

Appendix B

$$H_e = \frac{T_{Le} - T_o}{R_e} + \frac{T_{Le} - T_b}{R_1}; \quad H_b = \frac{T_b - T_{Le}}{R_1} + \frac{T_b - T_{Li}}{R_2}; \quad H_i = \frac{T_{Li} - T_b}{R_2} + \frac{T_{Li} - T_i}{R_i}$$

(B.1)

$$R_{b11} = \frac{1}{R_e} + \frac{1}{R_1}; \quad R_{b12} = -\frac{1}{R_1}; R_{b21} = -\frac{1}{R_1}; \quad R_{b22} = \frac{1}{R_1} + \frac{1}{R_2}; \quad R_{b23} = -\frac{1}{R_2};$$

$$R_{b32} = -\frac{1}{R_2}; \quad R_{b33} = \frac{1}{R_2} + \frac{1}{R_i}.$$

(B.2)

Appendix C

$$H_{oD} = \frac{T_{oD} - T_\infty}{R_{D_{total}}} + \frac{T_{oD} - T_{oC}}{R_{C_{total}}}; \quad H_{oC} = \frac{T_{oC} - T_{oD}}{R_{C_{total}}} + \frac{T_{oC} - T_{oB}}{R_{BC_{ax}}} + \frac{T_{oC} - T_{h2}}{R_{BC_{rad}}};$$

(C.1)

$$H_{oB} = \frac{T_{oB} - T_{oC}}{R_{BC\_ax}} + \frac{T_{oB} - T_{h2}}{R_{BC\_rad}} + \frac{T_{oB} - T_{oA}}{R_{B\_total}}; \quad H_{oA} = \frac{T_{oA} - T_\infty}{R_{A\_total}} + \frac{T_{oA} - T_{oB}}{R_{B\_total}};$$

(C.2)

$$R_{h11} = \frac{1}{R_A} + \frac{1}{R_B}; \quad R_{h12} = -\frac{1}{R_B};$$

$$R_{h21} = -\frac{1}{R_B}; \qquad R_{h22} = \frac{1}{R_B} + \frac{1}{R_{BC\_ax}} + \frac{1}{R_{BC\_rad}}; \quad R_{h23} = -\frac{1}{R_{BC\_ax}};$$

$$R_{h32} = -\frac{1}{R_{BC\_ax}}; \quad R_{h33} = \frac{1}{R_C} + \frac{1}{R_{BC\_ax}} + \frac{1}{R_{BC\_rad}}; \quad R_{h34} = -\frac{1}{R_C}; \tag{C.3}$$

$$R_{h43} = -\frac{1}{R_C}; \qquad R_{h44} = \frac{1}{R_C} + \frac{1}{R_D}.$$

$$R_{A\_ax} = \frac{L_{f1}}{k_h \pi \left(r_{Ah}^2 - r_{Ao}^2 \right)} + \frac{1}{h_{air} \pi \left(r_{Ah}^2 - r_{Ao}^2 \right)};$$

$$R_{A\_rad} = \frac{\ln\left(\dfrac{r_{Ah}}{r_{Ao}} \right)}{k_h 2\pi L_{f1}} + \frac{1}{h_{air} L_{f1} 2\pi r_{Ah}};$$

$$R_{D\_ax} = \frac{L_{r2}}{k_h \pi \left(r_{Dh}^2 - r_{Do}^2 \right)} + \frac{1}{h_{air} \pi \left(r_{Dh}^2 - r_{Do}^2 \right)}; \tag{C.4}$$

$$R_{D\_rad} = \frac{\ln\left(\dfrac{r_{Dh}}{r_{Do}} \right)}{k_h 2\pi L_{r2}} + \frac{1}{h_{air} L_{r2} 2\pi r_{Dh}}.$$

Housing free convection [6]:

$$h_{air} = 17.63 \left(T - T_\infty \right)^{\frac{1}{3}};$$

$$R_{B\_ax} = \frac{L_{f2}}{k_h \pi \left(r_{Bh}^2 - r_{Bo}^2 \right)} + 0;$$

$$R_{B\_rad} = \frac{\ln\left(\dfrac{r_{Bh}}{r_{Bo}} \right)}{k_h 2\pi L_{f2}} + \frac{1}{h_{air} L_{f2} 2\pi r_{Bh}};$$

$$R_{C\_ax} = \frac{L_{r1}}{k_h \pi \left(r_{Ch}^2 - r_{Co}^2 \right)} + 0;$$

$$R_{C\_rad} = \frac{\ln\left(\dfrac{r_{Ch}}{r_{Co}} \right)}{k_h 2\pi L_{r1}} + \frac{1}{h_{air} L_{r1} 2\pi r_{Ch}};$$

$$R_{BC\_rad} = \frac{\ln\left(\dfrac{r_{BCh}}{r_{BCo}} \right)}{k_h 2\pi L} + \frac{1}{h_{air} L 2\pi r_{BCh}};$$

$$R_{BC\_ax} = \frac{L}{k_h \pi \left(r_{BCh}^2 - r_{BCo}^2 \right)} + 0. \tag{C.5}$$

$$R_{A\_total} = \frac{R_{A\_ax} \times R_{A\_rad}}{R_{A\_ax} + R_{A\_rad}}; \quad R_{B\_total} = \frac{R_{B\_ax} \times R_{B\_rad}}{R_{B\_ax} + R_{B\_rad}};$$

$$R_{C\_total} = \frac{R_{C\_ax} \times R_{C\_rad}}{R_{C\_ax} + R_{C\_rad}}; \quad R_{D\_total} = \frac{R_{D\_ax} \times R_{D\_rad}}{R_{D\_ax} + R_{D\_rad}}. \tag{C.6}$$

Appendix D

$$0 = \frac{T_{\text{tail}} - T_{r\infty}}{R_{\text{tail}}} + \frac{T_{\text{tail}} - T_{ir1}}{R_M}; \quad H_{iD} = \frac{T_{ir1} - T_{\text{tail}}}{R_M} + \frac{T_{ir1} - T_{ir2}}{R_N}; \quad H_{iC} = \frac{T_{ir2} - T_{ir1}}{R_N} + \frac{T_{ir2} - T_{if1}}{R_O};$$

$$H_{iB} = \frac{T_{if1} - T_{ir2}}{R_O} + \frac{T_{if1} - T_{if2}}{R_P}; \quad H_{iA} = \frac{T_{if2} - T_{if1}}{R_P} + \frac{T_{if2} - T_{tip}}{R_Q}; \quad 0 = \frac{T_{tip} - T_{if2}}{R_Q} + \frac{T_{tip} - T_{f\infty}}{R_{tip}}. \tag{D.1}$$

$$R_M = \sum_{j=1}^{ir1-1} R_j + \frac{1}{2} R_{ir1} = R_1 + R_2 + \frac{1}{2} R_{ir1};$$

$$R_N = \sum_{j=ir1+1}^{ir2-1} R_j + \frac{1}{2} R_{ir1} + \frac{1}{2} R_{ir2} = R_3 + \frac{1}{2} R_{ir1} + \frac{1}{2} R_{ir2};$$

$$R_O = \sum_{j=ir2+1}^{if1-1} R_j + \frac{1}{2} R_{ir2} + \frac{1}{2} R_{if1} = R_4 + \frac{1}{2} R_{ir2} + \frac{1}{2} R_{if1}; \tag{D.2}$$

$$R_P = \sum_{j=if1+1}^{if2-1} R_j + \frac{1}{2} R_{if1} + \frac{1}{2} R_{if2} = R_5 + \frac{1}{2} R_{if1} + \frac{1}{2} R_{if2};$$

$$R_Q = \sum_{j=if2+1}^{n} R_j + \frac{1}{2} R_{if2} = R_6 + R_7 + \frac{1}{2} R_{if2}.$$

$$R_{i11} = \frac{1}{R_{tail}} + \frac{1}{R_M}; \quad R_{i12} = -\frac{1}{R_M};$$

$$R_{i21} = -\frac{1}{R_M}; \quad R_{i22} = \frac{1}{R_M} + \frac{1}{R_N}; \quad R_{i23} = -\frac{1}{R_N};$$

$$R_{i32} = -\frac{1}{R_N}; \quad R_{i33} = \frac{1}{R_N} + \frac{1}{R_O}; \quad R_{i34} = -\frac{1}{R_O};$$

$$R_{i43} = -\frac{1}{R_O}; \quad R_{i44} = \frac{1}{R_O} + \frac{1}{R_P}; \quad R_{i45} = -\frac{1}{R_P}; \tag{D.3}$$

$$R_{i54} = -\frac{1}{R_P}; \quad R_{i55} = \frac{1}{R_P} + \frac{1}{R_Q}; \quad R_{i56} = -\frac{1}{R_Q};$$

$$R_{i65} = -\frac{1}{R_Q}; \quad R_{i66} = \frac{1}{R_Q} + \frac{1}{R_{tip}}.$$

A Multiobjective Optimization Method for Designing *M*-Channel NPR Cosine Modulated Filter Bank for Image Compression

Anamika Jain, Aditya Goel

Department of Electronics and Communication Engineering, Maulana Azad National Institute of Technology, Bhopal, India
Email: ajain_bpl@yahoo.co.in, adityagoel2@rediffmail.com

Abstract

This paper proposes a method to design multichannel cosine modulated filter bank for image compression using multiobjective optimization technique. The design problem is a combination of stopband residual energy, least square error of the overall transfer function of the filter bank, coding gain with dc leakage free condition as constraint. The proposed algorithm uses Non-dominated Sorting Genetic Algorithm (NSGA) to minimize the mutually contradictory objective function by minimizing filter tap weights of prototype filter. The algorithm solves this problem by searching solutions that achieve the best compromise between the different objectives criteria. The performance of this algorithm is evaluated in terms of coding gain and peak signal to noise ratio (PSNR). Simulation results on different images are included to illustrate the effectiveness of the proposed algorithm for image compression application.

1. Introduction

Filter banks find application in various fields of signal, image and video processing for subband/transform coding [1]-[3]. In subband coding of images, filter banks are used to decompose the input image into several subimages in terms of different frequency bands. These subimages are then quantized and decoded to have compres-

sion with least distortion in reconstructed image quality. To get better resolution, decomposition in M bands can be achieved in one/two stage via *M*-channel filter bank or via repetitive use of wavelets (2 channel filter bank) in five/six stages [4]-[6]. Thus wavelet results in increased delay with nonuniform band decomposition. Filter banks are the most important component in image processing as the quality of the reconstructed image depends mainly on analysis/synthesis filter bank structure. Filter banks are designed as a transform to provide maximum energy compaction and decorrelation. To fulfil both the requirements, filter bank should have least stopband energy with least square error in overall transfer function. However, stopband attenuation increases if the perfect reconstruction condition is relaxed [7]. Amongst various types of transforms/filter banks cosine modulated, filter banks are the most popular choice as the designing and implementation both are easy as compared to others [8] [9]. Many approaches have been proposed for the design of 2 channel filter bank and *M*-channel CMFB filter banks with perfect and near perfect reconstruction condition since long time [10]-[11]. These approaches are based on the standard constrained, unconstrained, iterative least square optimization techniques which are in general complex, depend on the starting point and lead to local solution and may not fulfill all the objectives criterion. Further, quantization and coding of the transformed coefficients are achieved via progressive block based embedded image coder.

In this paper, an approach based on multiobjective optimization *i.e.* non-dominated sorting genetic algorithm [12] [13] is used to design *M*-channel uniform near perfect reconstruction Cosine modulated filter banks with coding gain given the highest priority amongst other objectives and Shapiro's EZW coder with modifications suitable for block based transform coding is used for coding/decoding.

The organization of the paper is as follows: in Section 2, a relevant brief analysis of the pseudo cosine modulated filter bank is given. In Section 3 optimization problem is formulated. A brief overview of multiobjective optimization algorithm is explained in Section 4. In Section 5, design examples (cases) and results are presented with their application to image coding and conclusions are drawn in Section 6.

2. Problem Description

2.1. Design of Cosine Modulated Filter Bank

A typical *M*-band filter bank is shown in **Figure 1**. Input signal $(x[n])$ is decomposed into *M* subband by analysis filters $(H_0(z) \cdots H_{M-1}(z))$ in transmitter side, these subbands are then decimated by a factor of *M*. At the receiver side, these subband are up sampled and recombined to get the reconstruction output $y[n]$ with the aid of synthesis filters $(F_0(z) \cdots F_{M-1}(z))$. When these analysis /synthesis filters are cosine modulated version of a low pass prototype filter, then these filter banks are termed as cosine modulated filter banks. Based on input/output relationship of filter bank the reconstruction output is expressible as [14]

$$\hat{X}(z) = \frac{1}{M} T_0(z) X(z) + \frac{1}{M} \sum_{k=0}^{M-1} T_r(z) X(W_M^r) \tag{1}$$

where $W_{2M} = e^{(-j\pi/M)}$, *M* is no. of channels, $T_0(z)$ is a distortion transfer function (determines overall amplitude distortion to input signal) and $T_r(z)$ is an aliasing distortion transfer function. In perfect reconstruction output signal is same as input except some delay *i.e.* $y[n] = x[n - n_0]$ for some integer n_0.
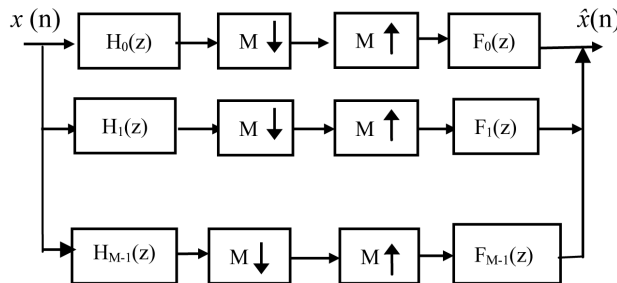

Figure 1. *M*-channel maximally-decimated Filter bank.

In cosine modulated filter bank (CMFB) impulse response of all the analysis and synthesis filters generated from the prototype filter $h(n)$ via cosine modulation, can be expressed as follows

$$h_k(n) = 2h(n)\cos\left((2k+1)\frac{\pi}{2M}\left(n - \frac{N-1}{2}\right) + (-1)^k \frac{\pi}{4}\right),$$

$$f_k(n) = 2h(n)\cos\left((2k+1)\frac{\pi}{2M}\left(n - \frac{N-1}{2}\right) - (-1)^k \frac{\pi}{4}\right).$$

(2)

For $k = 0, 1, \cdots, (M-1)$ and $n = 0, 1, \cdots, N-1$. The length N of $H_k(z)$ and $F_k(z)$ are the same and multiple of $2M$ to satisfy linear phase property in prototype filters.

Design criteria for effective filter bank are:

2.1.1. NPR Condition

In M-band cosine modulated filter bank perfect reconstruction conditions [10] for the impulse response of the prototype filter $h[n]$ is

$$2 \sum_{r=0}^{2m-2k-1} h(n+rM) h(n+rM+2kM) = \delta(k)$$

(3)

where $0 \le k \le E\left\lfloor \dfrac{N}{2M} \right\rfloor$, $\delta(k) = \begin{cases} 1, & k = 0; \\ 0, & k \ne 0. \end{cases}$, $n = 0, \cdots, N$ and $\lfloor E[n] \rfloor$ is the integer part of n.

The near perfect reconstruction condition measure is expressed as

$$\text{NPRV} = \sum_{k=0}^{k=E\left\lfloor \frac{N}{2M} \right\rfloor} \left(\delta(k) - 2 \sum_{r=0}^{N-2Mk} h(r) h(2Mk + r)\right)$$

(4)

2.1.2. Coding Gain

A widely accepted measure of coding performance is the Coding Gain (CG) which measures the energy concentration capability of filter banks. By modeling a natural image as a one-dimensional Markovian source with a correlation factor $\rho = 0.95$ and by assuming uncorrelated quantization errors, Katto and Yasuda [15] derived a filter dependent coding gain given by:

$$\text{CG}(\rho) = 10\log_{10}\left(\prod_{k=0}^{M-1} (A_k B_k)^{\frac{1}{\alpha_k}}\right)$$

(5)

where: $A_k = \sum_i \sum_j h_k(i) h_k(j) \rho^{|j-i|}$, $B_k = \sum_i g_k(i)^2$ and h_k and f_k are the kth analysis and synthesis filter of the M-channel uniform filter bank α_k is the corresponding subsampling ratio, and ρ is the correlation factor.

2.1.3. Frequency Selectivity

As suggested in [7], for high quality cosine modulated filter banks stop band mean energy of the prototype filter should be minimum and the magnitude distortion should be completely cancelled. Aliasing level in Pseudo CMFB can be kept small by keeping the attenuation large for $\omega \ge \dfrac{\pi}{M}$. The stopband energy can be evaluated by the following expression

$$E_s = \frac{1}{\pi - \omega_s} \int_{\omega_s}^{\pi} \left| H\left(e^{j\omega}\right)\right|^2 d\omega$$

(6)

where ω_s is the stopband edge of the prototype filter.

2.1.4. DC Leakage Free

In image coding, dc leakage free condition does not yield significant improvements in image quality [16]; how-

ever, to minimize the checkerboard and ringing artifacts dc leakage in subbands other than the lowest one is desired. DC-leakage free property in the cosine modulated filter bank can be expressed as

$$h^t c(\omega) = 0, \quad \forall \omega = \frac{(2k+1)\pi}{2M}, \ 1 \leq k \leq M-1 \tag{7}$$

3. A Multiobjective Genetic Algorithm for the Design of NPR Cosine Modulated Filter Bank

For optimizing a number of conflicting objective functions simultaneously, multiobjective optimization techniques are used. These techniques provide Pareto-optimal solutions instead of unique optimal solution. These solutions are optimal in the sense that no other solutions in the parameter space are superior to them when all the objective functions are considered [17]-[19].

A multiobjective optimization can be stated mathematically as the minimization problem to minimize

$$y = f(x) = \left[f_1(x), f_2(x), \cdots, f_M(x) \right]^{\mathrm{T}} \tag{8}$$

where $x = \left[x_1, x_2, \cdots, x_N \right]^{\mathrm{T}} \in X$ and $y = \left[y_1, y_2, \cdots, y_M \right]^{\mathrm{T}} \in Y$.

Vectors x and y represent the independent variables and the corresponding values of the individual objective functions respectively; on the other hand, X and Y are called the solution space and objective space, respectively.

The notion of Pareto-optimality is an important concept for multiobjective and explained in terms of a "dominance relation" as a solution x_i is said to dominate another solution x_j if the following conditions hold:

$$f_k(x_i) \leq f_k(x_j) \quad \text{For all } k = 1, 2, \cdots, M;$$
$$f_k(x_i) < f_k(x_j) \quad \text{For all } k = 1, 2, \cdots, N. \tag{9}$$

If a solution is not dominated by any other solutions among the entire objective functions than that solution is said to be a nondominated solution. The solutions that are nondominated regarding the entire parameter space are called Pareto-optimal solutions.

In our design, we search prototype filter h that minimize the three individual objective functions, namely, the energy of the filter stopband, near perfect reconstruction measure and the coding gain. To confine the search process in a feasible solution space, we impose dc leakage free condition as constraint. We have to determine only half of the coefficients since the remaining coefficients are obtained, by applying symmetry.

Our multi-objective optimization problem is formulated as follows:

$$\min/\max_{h_0} \left(\mathrm{objf}_1, \mathrm{objf}_2, \mathrm{objf}_3 \right) \text{ and } \begin{cases} \mathrm{objf}_1 = 1/\mathrm{CG}^2; \\ \mathrm{objf}_2 = E_s; \\ \mathrm{objf}_3 = E_t. \end{cases} \tag{10}$$

Subject to the constraint: $h^t c(\omega) = 0$

In our case, a set of coefficients of the analysis filter bank (h) are treated as chromosomes which are optimized by the constrained multi-objective genetic algorithm (C-NSGA) to obtain a set of filter banks that minimize all the prescribed objective functions with a satisfactory level. This algorithm is based on the NSGA approach [12].

4. Case Study

4.1. Filter Design Examples

Filter banks designed using non dominated sorting genetic algorithm here use some important parameters *i.e.* real valued filter coefficients for chromosome construction, crossover operator with distribution index of 20 and probability of 0.9, a polynomial mutation of distribution index 20 and probability of 0.01 are applied [12] [13]. The population size is set to 100 to find Pareto optimal solutions. Chromosomes of the initial population of filter coefficients are set as $h_0 = \left[h(0), \cdots, h(N) \right]^t$ obtained by setting the starting vector as $h(N-1)/2 = \left(1/\sqrt{2} \right)$

else are equal to zero, satisfying the unit energy constraint between [−1, 1]. NPR condition is set by keeping the distortion to 10^{-5}. The maximum generation (G_{max}) is set to 2000 generations. The programme was run using MATLAB 7.8 and the obtained coefficients were used for designing NPR cosine modulated filter bank. The amplitude response of the designed filter banks for 8 channel with 16 and 24 taps, with and without dc leakage condition are plotted in **Figure 2** and **Figure 3**. The result of response indicates the superiority of the longer tap filters as the stopband attenuation increase with the increase in no. of taps. Secondly, incorporating dc leakage deteriorate the response of the filter but is good from the image coding point as dc leakage free condition reduce the aliasing errors and all the subband filters except the lowpass one, has at least one zero at $z = 1$ and thus all other bandpass and highpass filters are dc transmission free. Both the filter banks are designed giving highest priority to coding gain amongst the multiple objectives. The 8×16 CMFB in design Example I, with dc leakage (optimized for pure coding gain); attain 9.3749 dB, which almost equals the coding gain of optimal biorthogonal systems. However, applying dc leakage free constraints on the filter coefficients ensure a high level of perceptual performance in image coding. In design example II, the 8×24 case, our CMFB achieves an improvement of 0.001 dB in coding gain. CMFB's with filter length 16 and 24 are shown in **Figure 2** and **Figure 3**. Increasing the length only improves the coding gain marginally, it helps in the case of stopband attenuation (where longer filters are always beneficial), which result in more complexity in design and implementation.
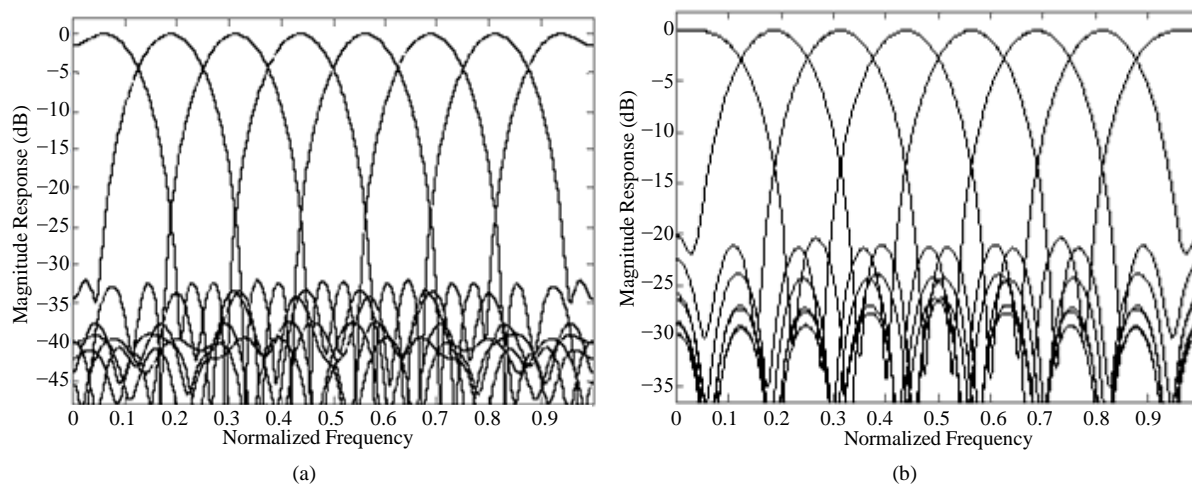

Figure 2. Magnitude Response of the NPR Cosine modulated filter bank with dc leakage free constraint for (a) 8×16; (b) 8×24.
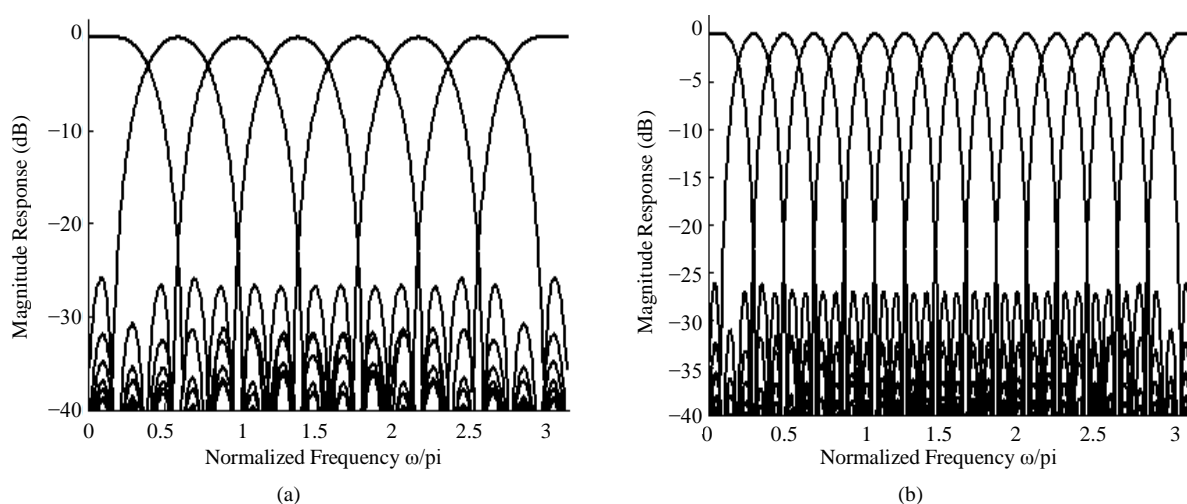

Figure 3. Magnitude response of the NPR pseudo QMF filter bank with DC leakage constraint designed using multiobjective optimization method for (a) 8×32; (b) 16×64.

Magnitude response of 8 Channel and 16 Channel NPR Pseudo QMF filter banks with 32 and 64 taps respectively, with dc leakage free constraint, designed using multiobjective optimization method, have been shown in **Figure 3(a)** and **Figure 3(b)** respectively. Although stopband attenuation is almost same for both the filter banks, clarity of magnitude response shows that the designed filter bank have least dc leakage with coding gain 8.8872 dB (8 × 32) and 9.03632 dB (16 × 64).

4.2. Application in Image Coding

The performance of our designed Cosine Modulated Filter Bank's is evaluated through a progressive block based EZW image coder [20] merely replacing DCT by designed filter bank. In all simulations, we employed a 2 level decomposition of the 8-channel filter bank (1-D transform) separable implementation and symmetric extension at the edges. The coder is chosen to encode the transform coefficients because it provides some of the best performances of any image codec by using the zero tree approach for block based transform and generates an embedded coding which can be helpful in the reconstruction of images at different bitrates.

The images chosen for the coding experiments are Barbara and Boat. Both of them are standard, well known 512×512 8-bit gray-scale test images. The objective distortion measure is the popular peak signal-to-noise ratio (PSNR)

$$PSNR = 10\log_{10}\left(\frac{255}{MSE}\right) dB \tag{11}$$

where, MSE denotes the mean squared error. Filter banks 8 × 16 and 8 × 24 were included in the tests, and both were optimized for maximum coding gain. Reconstructed images "Boat" and "Barbara" using designed filter bank (8 × 16) and (8 × 24) are shown in **Figure 4(b)**, **Figure 4(c)** and **Figure 5(b)**, **Figure 5(c)** respectively.

It is clear that longer filter banks with more no. of taps have the best characteristics with increase in PSNR. In this context, it is important to note that the filter length cannot be arbitrarily increased in order to increase PSNRs and achieve better frequency selectivity because long filters cause a ringing around edges artefacts when high frequency subband are coarsely quantized. Reconstructed images are better in visual quality as the frequency selectivity improves and overall distortion reduced, however, as the no. of taps increased artifacts become pronounced and PSNR decreases as the bit rates further reduced.

As shown in **Figure 4** and **Figure 5** for both the images performance of the designed CMFB is good as in low bit-rate coding image is reconstructed using only nonzero coefficients which are few in number and are very concentrated in lowest frequency component (lowpass filter output). Even though coding gain become lower, inclusion of dc leakage free condition leads to smooth reconstructed image as details are preserved.

Objective results for improvement in performance are shown in **Table 1** and **Table 2**. In **Table 1** designed filter bank with 24 taps is compared with baseline JPEG using (8 × 8 DCT) as transform and LOT and GENLOT [21]. It is clear that using multiobjective optimization the designed filter bank outperforms as both the coding gain and stopband attenuation are higher as compared other existing filter bank with same no. of parameters.

Results shown in **Table 2** indicate the improvement on filter bank performance as the additional criterion of dc leakage free is incorporated in the design. Using multiple objectives our designed filter bank shows a PSNR gain of around 2.5 dB over a wide range of bit rates at the expense of PR violation.

| (a) | (b) | (c) |

Figure 4. (a) Original "Boat" Image; (b) Reconstructed image with (8 × 16) NPR CMFB filter bank; (c) Reconstructed image with (8 × 24) NPR CMFB filter bank.
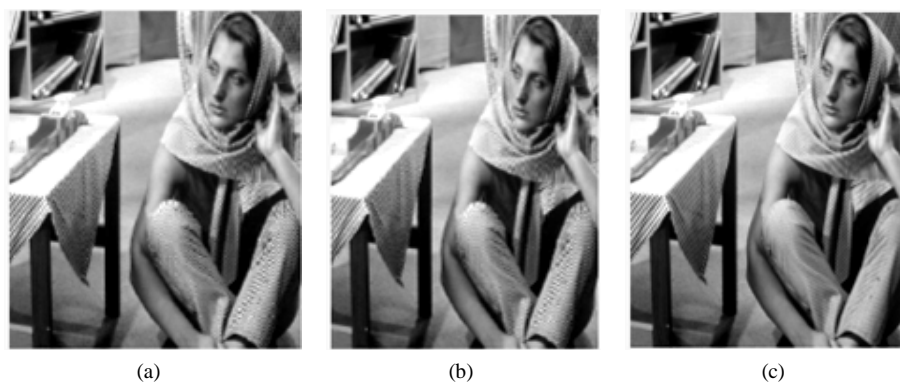
Figure 5. (a) Original "Barbara" image; (b) Reconstructed image with (8×16) NPR CMFB filter bank; (c) Reconstructed image with (8×24) NPR CMFB filter bank.

Table 1. Comparison of different transform's performance.

Transform	Coding Gain	Stopband Attenuation (dB)
8×8 DCT	8.83	9.96
8×16 LOT	9.22	19.38
8×24 GenLOT	9.35	23.20
Proposed (8×24)	9.3749	26.026

Table 2. Image coding results for "Barbara" image at 0.25 bpp.

M No. of channels	Length of prototype filter	With dc leakage		Without dc leakage	
		PSNR	CG	PSNR	CG
8	16	28.02	8.89	29.99	8.85
8	24	30.11	9.37	31.18	9.21
8	32	32.02	9.374	33.63	9.257

5. Conclusion

The design of M-channel uniform cosine modulated filter banks using genetic algorithm is discussed in this paper. The design problem having multiple objectives is solved using non-dominated sorting algorithm with regularity constraint. Although dc leakage free condition is not an essential requirement for image processing in M-channel filter bank as PSNRs don't improve considerably, its imposition identifies infeasible solutions thereby confines the search for the NPR filter banks simultaneously, and improve visual quality of images at low bit rates. From simulation result, it is shown that the algorithm results in lower tap filters with high stopband attenuation, the least overall distortion and higher coding gain for image compression. Use of progressive image coder based on block-based EZW improves PSNR via reducing blocking and ringing artefacts. The work can be further extended to design of biorthogonal cosine modulated filter banks for image coding application.

References

[1] Esteban, D. and Galand, C. (1977) Application of Quadrature Mirror Filters to Split Band Voice Coding Schemes. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, ASSP*, May 1977, 191-195.

[2] Smith, M.J.T. and Eddins, S.L. (1990) Analysis/Synthesis Techniques for Sub-Band Image Coding. *IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP*-38, August 1990, 1446-1456. http://dx.doi.org/10.1109/29.57579

[3] Woods, J.W. and O'Neil, S.D. (1995) Sub-Band Coding of Images. *IEEE Trans. Acoustics Speech*, PTR, Englewood Cliffs.

[4] Vatterli, M. and Kovacevic, J. (1995) Wavelet and Subband Coding. Prentice Hall PTR, Englewood Cliffs.

[5]   Steffen, P., Heller, P.N., Gopinath, R.A. and Burrus, C.S. (1993) Theory of Regular M-Band Wavelet Bases. *IEEE Transactions on Signal Processing*, **41**, 3497-3510. http://dx.doi.org/10.1109/78.258088

[6]   Gopinath, R.A. and Burrus, C.S. (1994) On Cosine-Modulated Wavelet Orthonormal Bases. *IEEE Transactions on Image Processing*, **4**, 162-175. http://dx.doi.org/10.1109/83.342190

[7]   Nguyen, T.Q. (1994) Near Perfect Reconstruction Pseudo-QMF Banks. *IEEE Transactions on Image Processing*, **42**, 65-75. http://dx.doi.org/10.1109/78.258122

[8]   Lin, Y.P. and Vaidyanathan, P.P. (1995) Linear Phase Cosine Modulated Maximally Decimated Filter Banks with Perfect Reconstruction. *IEEE Transactions on Image Processing*, **43**, 2525-2539. http://dx.doi.org/10.1109/78.482104

[9]   Vaidyanathan, P.P. (1993) Multirate Systems and Filter Banks. Prentice Hall, Englewood Cliffs.

[10] Rossi M., Zhan J.Y. and Steenaart, W. (1996) Iterative Constrained Least Square Design of Near Perfect Reconstruction Pseudo qmf Banks. *IEEE CCECE*, 1996, 766-769.

[11] Jain, A. and Goel, A. (2012) Unconstrained Optimization Method for Design of M Channel Perfect Reconstruction CMFB Filter Bank. *Proceedings of the International Conference on Signal*, *Image and Video Processing*, Patna, 13-15 January 2012, 302-308.

[12] Boukhobza, A., Ahmed, T., Bounoua, A. and Taleb, N. (2009) A Filter Banks Design Using a Multi-Objective Genetic Algorithm for Embedded Image Coding Scheme. *Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis*, Salzburg, 16-18 September 2009, 776-781.

[13] Boukhobza, A., Ahmed, T., Bounoua, A. and Taleb, N. (2009) A Filter Banks Design Using a Multi-Objective Genetic Algorithm for an Image Coding Scheme. *Proceedings of the 16th IEEE International Conference on Image Processing*, Cairo, 7-10 November 2009, 1933-1936.

[14] Heller, P.N., Karp, T. and Nguyen, T.Q. (1999) A General Formulation for Modulated Filter Banks. *IEEE Transactions on Signal Processing*, **47**, 986-1002. http://dx.doi.org/10.1109/78.752597

[15] Katto, J. and Yasuda, Y. (1991) Performance Evaluation of Subband Coding and Optimization of Its Filter Coefficients. *Proceedings of the SPIE Symposium on Visual Communications and Image Processing*, **1605**, 95-106.

[16] Karp, T. and Mertins, A. (1998) Biorthogonal Cosine-Modulated Filter Banks without DC Leakage. *Proceedings of the 1998 IEEE International Conference on Acoustics*, *Speech and Signal Processing*, Seattle, 12-15 May 1998, 1457-1460.

[17] Deb, K. (1999) Evolutionary Algorithms for Multi-Criterion Optimization in Engineering Design. *Proceedings of Evolutionary Algorithms in Engineering and Computer Science* (*EUROGEN*-99), Jyväskylä, 29 May-3 June 1999, 135-161.

[18] Deb, K., Pratab, A., Agarwal, S. and Meyarivan, T. (2002) A Fast and Elitist Multi-Objective Genetic Algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, **6**, 182-197. http://dx.doi.org/10.1109/4235.996017

[19] Goldberg, D.E. (1989) Genetic Algorithms in Search Optimization and Machine Learning. Addison-Wesley, Boston.

[20] Xiong, Z., Galerius, O.G. and Orchard, M.T. (1996) A DCT Based Embedded Image Coder. *IEEE Signal Processing Letters*, **3**, 289-290. http://dx.doi.org/10.1109/97.542157

[21] de Queiroz, R.L., Nguyen, T.Q. and Rao, K.R. (1996) The GenLOT: Generalized Linear-Phase Lapped Orthogonal Transform. *IEEE Transactions on Signal Processing*, **44**, 497-507. http://dx.doi.org/10.1109/78.489023