Alternative Microcleanliness Measurement Methods For Wheel Steel

Jay Galbraith, Ph. D.
Melting Metallurgist
Standard Steel, LLC
500 North Walnut Street
Burnham, PA, USA 17009

Cameron Lonsdale
Vice President-Technical
Standard Steel, LLC
500 North Walnut Street
Burnham, PA, USA 17009

Jean-Marc Bohlen and Brad Cooley
Thermo Electron Corporation
En Vallaire Ouest C, Ecublens, CH
and 15300 Rotunda Dr. Suite 301
Dearborn, MI, USA

Paul Meyer, Ph. D.
Director of Sensor Research
GE Inspection Technologies, Inc.
50 Industrial Park Road
Lewistown, PA 17044

Summary: Shattered rim wheel failures, although relatively rare in North American railway freight service, remain a major concern for railways and wheel manufacturers since such failures can occur under a train at high speed, and can result in dangerous, expensive derailments. Shattered rim fatigue cracks typically initiate and grow at a depth of 12 to 20 mm (1/2 to 3/4 inches) below the wheel tread surface. To help further reduce the incidence of these wheel failures, the Association of American Railroads (AAR) recently adopted a wheel steel microcleanliness specification based upon ASTM E1245. A major North American railroad has used this microcleanliness evaluation method, which involves using quantitative image analysis techniques to analyze six polished metallographic samples removed from the wheel rim, for several years. Alternative effective methods of microcleanliness evaluation, which could be performed faster in-house, would be desirable from a wheel manufacturing point of view. This paper discusses the use of two alternative microcleanliness measurement methods, phased array ultrasonic testing and the Spark-Dat system, to determine the cleanliness of wheel steel samples. An ultrasonic evaluation of several wheel rims is first conducted to provide counts and sizes of indications within the rim. From this information an ultrasonic cleanliness determination is made. Next, the Spark-Dat system is described and the basic principles of its operation are reviewed. Samples removed from the ultrasonically tested wheel rims are then examined using the ASTM 1245/BNSF test, and the Spark-Dat system subsequently evaluates the same samples. Data from the ultrasonic and Spark-Dat tests are then compared to data obtained using the ASTM E1245 microcleanliness method, and statistical comparisons are made.

1. INTRODUCTION

Railway wheel failures in North American freight service are not a common occurrence, but can cause derailments and serious damage when they do take place. Several types of broken wheels can occur in service including shattered rims, spread rims, cracked/broken rims, vertical split rims, cracked flanges, cracked plates, etc. The Association of American Railroads (AAR) has established Why Made Codes for categorizing wheel failures in the AAR Car Repair Billing (CRB) system. Shattered rims (Why Made Code 71) are fatigue cracks that initiate 12 to 20 mm (1/2 inches to 3/4 inches) below the tread surface and propagate generally parallel to the tread surface until a piece of the rim breaks off. The fatigue crack origin is typically porosity in cast wheels and inclusions in wrought wheels.

AAR CRB data for year 2002 showed only 174 shattered rims out of nearly 447,000 total wheel removals [1]. We recognize that other types of wheel failures such as cracked rims (Why Made Code 68 - 423

removals in 2002) and spread rims (Why Made Code 72 - 104 removals in 2002) could be improperly identified by field car repair personnel and could have crack origins similar to that of shattered rims. Also, we note that CRB data does not contain all wheel changeouts in North America – only those wheel changes made by railroads to "foreign line" cars and to privately owned cars are included in the data (with few exceptions). However, the total number of shattered rim failures is still relatively small.

Several excellent papers have been written on the subject of shattered rims in recent years and some are referenced here [2-8]. A significant amount of discussion in these papers has been centered on the initiation and propagation mechanisms for shattered rims [2-7]. Also, several authors have discussed the size of discontinuity necessary to initiate a shattered rim fatigue crack in service. Lunden [2] in Sweden and Marais [3] in South Africa have suggested that a 1 mm (0.040 inches) diameter discontinuity can initiate shattered rims. Baretta et al. [7] (Italy) stated that the typical dimensions of original discontinuities (elongated aluminum oxides) in wrought wheel shattered rims had a length of 1 to 5 mm (0.040 to 0.2 inches) and width of 0.3 to 1 mm (0.012 to 0.040 inches). Stone and Dahlman [8] have provided micrographic evidence that a shattered rim initiated at a cast wheel void of only 0.64-mm (0.025 inches) in diameter.

Notable problems with certain freight car wheels during the past decade, and several large wheel-related derailments on railroads in the western United States, have led to increased scrutiny for all wheels in North America. With increased freight car axle loads, train speeds and car utilization, wheels are under greater stress than ever before. There are many efforts now underway to prevent wheel failures in freight service, some of which are described in a recent International Heavy Haul Association paper by Dahlman and Lonsdale [9]. For example, in 1999 the AAR tightened the ultrasonic inspection requirement for newly manufactured wheel rims and also recently adopted ultrasonic testing of tread turned wheels at wheel mounting shops.

2. PHASED ARRAY ULTRASONIC INSPECTION

Ultrasonic inspection systems are well established for evaluation of manufactured components, are non-destructive and can efficiently and effectively test the entire volume of a part. Use of modern phased array ultrasonic inspection techniques to evaluate new wheel rims was outlined in a paper given at the 13th International Wheelset Congress [10]. Standard Steel was the first wheel manufacturer in the world to use phased array systems to test wheels. These phased array systems use small individual ultrasonic elements and sophisticated electronics to provide a quality inspection.

A total of 128 transducers are available to inspect the rim radially from the tread surface and 128 transducers are available to inspect the rim axially from the back rim face. The number and location of transducers used depends on the wheel design and dimensions being inspected. Advantages of the new system include improved volumetric coverage, improved sensitivity, a more versatile inspection system and a higher quality evaluation of the rim. New software available from the equipment vendor enables data collection and analysis, and for ultrasonic scans to be saved. Such parameters as total indication count, etc., can be determined for various indication amplitude levels, and suggest a possible method to determine rim cleanliness.

3. MICROCLEANLINESS TESTING

One of the strategies to reduce wheel failures used by Burlington Northern Santa Fe (BNSF) Railway has been to employ microcleanliness testing of new wheel rims. This customized quantitative image analysis method is performed using BNSF and ASTM 1245 procedures and counts the microscopic porosity and inclusion content of metallographic samples. Six samples are cut and removed from the central tread area 1/2 inches (12.7 mm) below the tread surface of a wheel and the samples are ground and subsequently polished using an automated method. The surface for evaluation measures 7/8 inches (22.2 mm) in the circumferential direction and 3/4 inches (19 mm) wide through the thickness of the rim and testing takes place over an area approximately 161 mm^2 in size. The image analysis work is typically performed using a Clemex system with gray scale levels established by BNSF. With these gray scale levels, the system distinguishes between voids, oxides and sulfides in the matrix. Voids and oxides are

darker constituents while sulfides are lighter gray. Wheel manufacturers are required to destructively test at least two wheels per quarter in this fashion, and the BNSF Research and Test Department at Topeka, Kansas, then re-evaluates the samples. AAR recently adopted such testing for new wheels.

Mean volume percent and maximum volume percent (worst field) are collected for voids, oxides and sulfides on each sample's polished surface. Sulfides are soft inclusions and have not been associated with wheel fatigue failures such as shattered rims. However, voids and oxides have been found to cause shattered rims in cast and forged wheels, respectively, and wheel manufacturers are therefore concerned about the possible presence of such discontinuities. Given the nature of the forged wheel manufacturing process where the wheel blank is exposed to two operations in a 10,000 ton press, a rolling operation in a vertical wheel rolling mill, and a final operation in a 4,000 ton press, Standard Steel is not as concerned about the possibility of internal voids. Therefore, most attention is paid to oxides as potential discontinuities in the wheel rim. For BNSF and AAR requirements, acceptable microcleanliness results must be less than 0.1 percent mean volume percent voids + oxides. The maximum volume percent voids + oxides (worst field) must be less than 0.2 percent for BNSF and less than 1.0 percent for AAR. Microcleanliness results for numerous wheels were recently published by Dahlman and Lonsdale [9].

4. SPARK-DAT PRINCIPLES OF OPERATION AND BACKGROUND

Optical emission spectrometry (OES) is a well-established method for measuring elemental concentrations in metals. The technique is fast and reliable and allows measuring all the elements from trace to major element levels. It has therefore been adopted by most of the industries that perform production control. In traditional quantitative OES analysis, the sample is excited by several thousand sparks and the light emitted is integrated over the total excitation period. The elementary emission signals are converted to concentrations from calibration curves. In the last decade, a new route has been opened for optical emission spectrometers. This is the case on Thermo's ARL 4460 equipped with Spark-DAT (Spark Data Analysis and Treatment). Spark-DAT carries out digital acquisition and treatment of every spark signal separately. Considering that a single spark hits a very small area, the resultant light pulses contain information on the local material composition. Therefore, Spark-DAT has the potential of providing new analytical information such as cleanliness indices and inclusion determination. In homogeneous samples the intensities of the individual signals follow a gauss shaped distribution around an average value corresponding to the concentration. In samples containing inclusions, the chemical composition is altered when there is an inclusion. When a spark hits an area containing an inclusion, peaks of higher intensities may appear, if the inclusion is large enough (and if the concentration is larger in the inclusion than in the matrix), for the elements composing the inclusion. Spark-DAT not only allows counting these high intensities, it also makes it possible to correlate simultaneous occurrences of high signals on several channels in order to determine the inclusion composition. To date, Spark-DAT has been successful in different applications, for instance:

- Cleanness assessments in matrices like Fe, Al, Cu, Pb, Zn, Ni, etc.
- Analysis of low oxygen (<10 ppm) in some bearing steels
- Analysis of soluble / insoluble Al in low alloy steels
- Determination of inclusions' compositions
- Replacement of lengthy methods like fatigue tests by correlating inclusions counted by Spark-DAT with fatigue test results.

Currently, many efforts are being made to replace traditional methods of inclusion identification and subsequent determination of number density and size distribution. In Spark-DAT, each sample is typically analyzed 3-5 times to obtain data on the inclusion content over an area of 50-100 mm^2. A typical Spark-DAT analysis acquires 2000 individual signals from the different elemental channels in 10 seconds. On the ARL 4460, the results can be analyzed online through pre-selected algorithms. Inclusion signals are identified and counted according to the following principle:

a) Calculate the mean (m) and standard deviation (SD) on the elemental channels
b) Count intensities larger than $m + K \cdot SD$, where $K = 3$ (commonly used in statistical treatment)
c) Remove these intensities and repeat steps a through c iteratively

5. INITIAL SPARK-DAT TESTING AT STANDARD STEEL

An effort was first made to determine if results from the Spark-Dat system correlated to quantitative image analysis microcleanliness testing results obtained from an outside vendor, Cambridge Materials Testing, Ltd., in Canada. Six metallographic samples were removed from four different wheel rims and were tested in accordance with ASTM E1245 and BNSF requirements. Two of the selected samples (from the same wheel rim) had failing results for the mean volume percent voids + oxides parameter, while the other four samples passed the parameter. All four wheels passed current AAR ultrasonic testing requirements prior to removal of the metallographic samples. Following microcleanliness testing, the same six samples were sent to Thermo Electron Corporation for evaluation in their Spark-Dat system. It must be noted that Thermo Electron used a general set of algorithms for the Spark-Dat evaluations - an optimized set of algorithms established for the specific application was not used at this point.

Four Spark-Dat readings were taken on each of the six metallographic samples and averages for each sample were then calculated. Values were obtained for numerous elements and oxide compounds including aluminum, oxygen, Al_2O_3, titanium, TiO, etc. Four Spark-Dat readings were taken on ground sample surfaces, and four Spark-Dat readings were re-taken on the same metallographic sample after polishing of the surface. A correlation coefficient (the statistic r) was calculated between the average Spark-Dat values and the microcleanliness parameter determined for each of the six samples. Table 1 shows selected results for Spark-Dat value vs. mean volume percent voids + oxides and Table 2 shows results for Spark-Dat value vs. maximum volume percent voids + oxides (worst field).

	Correlation Coefficient, r		
Sample condition	**Oxygen**	**Aluminum**	**Al_2O_3**
Polished surface	0.717	0.801	0.607
Ground surface	0.333	0.313	0.329

Table 1. Correlation for mean volume percent voids + oxides vs. Spark-Dat value.

	Correlation Coefficient, r		
Sample Condition	**Oxygen**	**Aluminum**	**Al_2O_3**
Polished surface	0.522	0.466	0.394
Ground surface	0.542	0.435	0.541

Table 2. Correlation for maximum volume percent voids + oxides vs. Spark-Dat value.

A correlation coefficient closest to one shows a stronger relationship between two parameters. The highest correlation coefficients are for the mean volume percent voids + oxides with a polished sample surface. The initial tests suggested that Spark-Dat should be further evaluated as a cleanliness measure.

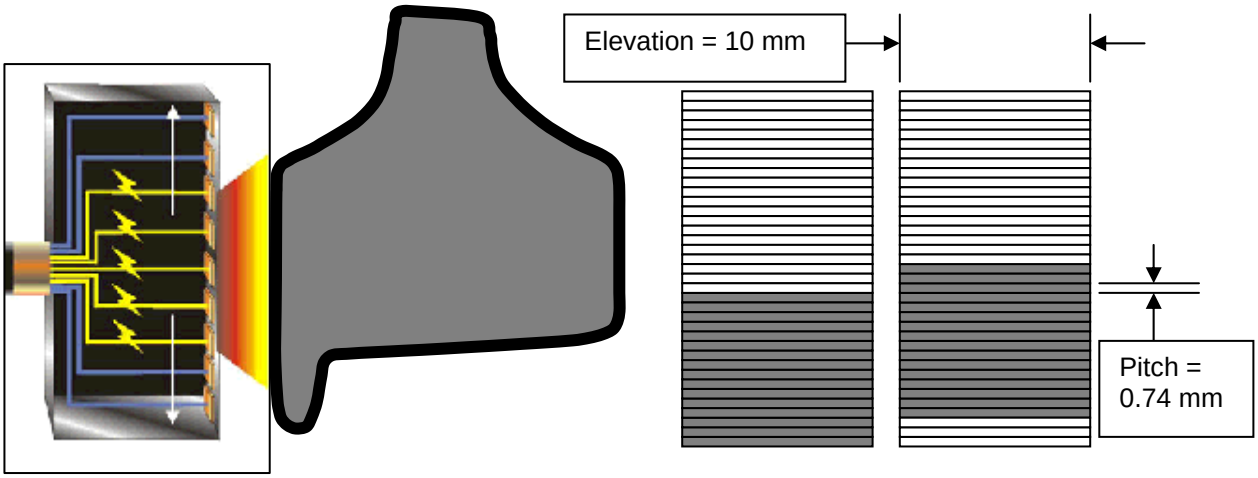
6. ULTRASONIC EVALUATION, RESULTS AND DISCUSSION

Six forged locomotive wheels were selected for comparative evaluations using phased array ultrasonic systems, microcleanliness testing, and Spark-Dat. Two of the wheels were D42 wheel design with 63.5 mm (2-1/2 inches) minimum rim thickness while four were E42 wheel designs with 89 mm (3-1/2 inches) minimum rim thickness. Table 3 shows a summary of wheel information.

Wheel Designation	Wheel Design Type	Serial Number
A	E42	9-3-1979
B	D42	7-3-2070
C	E42	6-3-11988
D	E42	9-3-1954
E	E42	9-3-15943
F	D42	9-3-15919

Table 3. Summary of information for six test wheels.

The six wheels were shipped to GE Inspection Technologies in Lewistown, PA for detailed scanning of the wheel rims and collection of indication information. Parameters including total number of indications, size of indications, etc., were collected and subsequently analyzed. The six test wheels were inspected using an ultrasonic phased array inspection system. This test permits the sound beam to be focused to a depth in the wheel, as well as scanned to different positions in the wheel as shown in Figure 1.



a - Position of the array relative to the wheel

b- Array details

Figure 1. Schematic arrangement of the ultrasound phased array and the tested wheel, and the details of the array – scan is axial, from the back rim face.

In addition to scanning and focusing, the sound field can also be steered with this system, but steering was not needed for these tests.

The transducer is a 5 MHz, 128 element array fixtured with a 30 mm (1.2 inches) water column between the transducer and the wheel. The array pitch is 0.74 mm (0.029 inches) pitch and operated with a 16-element "virtual probe." The "virtual probe" is the group of elements pulsed on each repetition cycle of the instrument, and is colored gray in Figure 1b. On the first repetition cycle, the virtual probe is positioned just inside the outside radius of the rim. On the next repetition cycle, the instrument will index the virtual probe by 3 elements, moving the sound beam approximately 2.2 mm (0.09 inches) (3 * element pitch) closer to the inside radius of the rim. The indexing repeats until the array is indexed to the inside radius of the wheel rim, a position that was pre-selected in accordance with the wheel geometry. Then, the sound beam is repositioned to its starting point at the outside radius of the wheel and the process repeats. During these successive repetition cycles, the wheel is rotating with a circumferential speed of 63 mm/second (2.48 inches/second). The resulting scan yields a 2.2 mm (0.09 inches) radial step size and a 1.5 mm (0.06 inches) circumferential step size.

The evaluation of the ultrasound signals is identical to many flaw detectors. An interface gate monitors the position of the flanged side of the wheel rim. A data acquisition gate is timed to start just beyond the flange side of the wheel rim and to end just before the opposite side of the wheel rim. The instrument records any reflections in this time window. As mentioned earlier, the virtual probe is selected so the sound field is scanned from the outside radius of the wheel rim to the inside of the rim.

The instrument is calibrated using a 3.2 mm (0.125 inches) diameter flat-bottom hole in a calibration wheel. The reflection from this target was set to 80 percent screen height, and the six test wheels were then scanned at these instrument settings. The six wheels were then rescanned with the instrument amplification increased + 6 dB. The wheels were scanned two additional times. In these scans, the sound field was focused to a depth of 63 mm (2.48 inches), the wheels were scanned, and then the amplifier gain was increased +6 dB and the wheels were scanned again. There was no insight by the authors leading them to these additional scan settings, but while acquiring the data it was deemed

prudent to compare data acquired at multiple instrument settings. The peak amplitude for each a-scan was digitally plotted as a color-coded map as shown in Figure 2. In this image the reflector amplitudes are color coded – the color code is shown in the lower portion of the image.

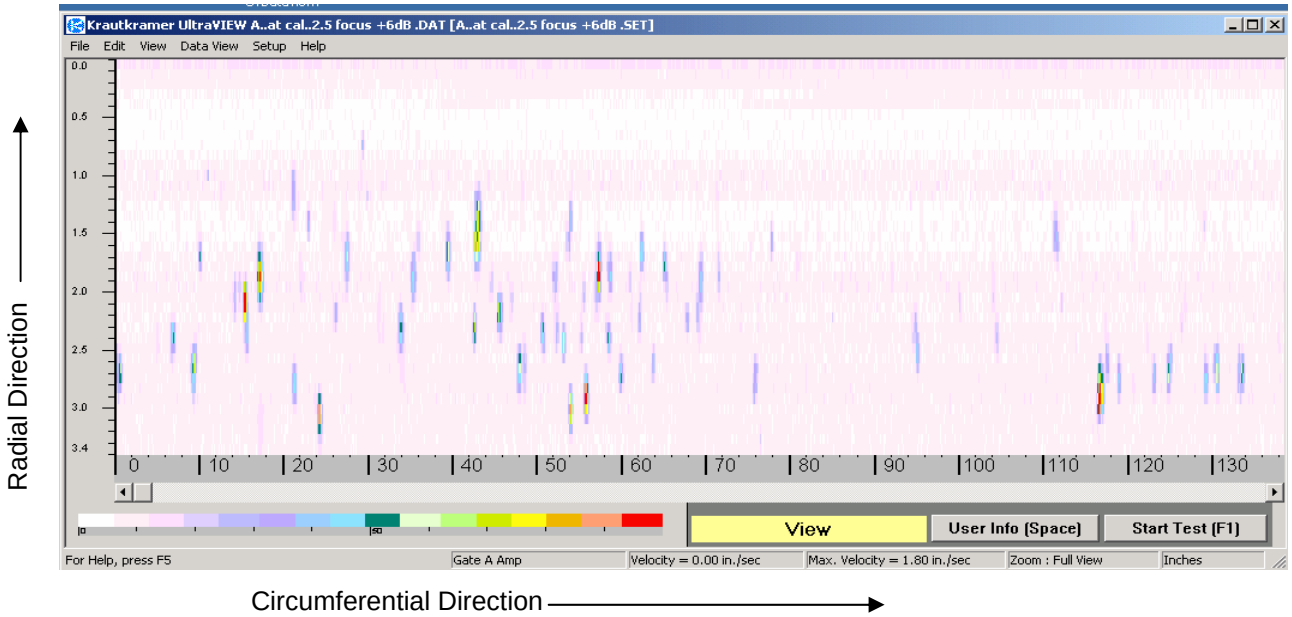


Figure 2. Example color-coded amplitude scan of a test wheel, focused scan with 6 dB added.

One intent of this work is comparison of the ultrasound data to microcleanliness (quantitative image analysis) test data determine whether there might be a correlation. As a first attempt, the authors chose to develop a histogram of the amplitude data for each wheel and then compare it to microcleanliness data acquired from metal samples subsequently cut from each of these wheels. Because the wheels are the same size, and because they were scanned with the same settings, there are approximately the same number of pixels in each resulting image. Those that were slightly larger were "clipped" at the edges so that all six wheels had 90090 pixels. The original a-scan data for each pixel was digitized using a 8-bit A/D converter and a amplitude histogram was developed for each wheel. The amplitude of various pixels were collected and recorded for each wheel. Table 4 shows the number of pixel indications for various ultrasonic amplitude (percent screen height) ranges. Testing was non-focused with no additional gain added.

Ultrasonic Amplitude (% screen height)	Wheel A	Wheel B	Wheel C	Wheel D	Wheel E	Wheel F
0-9	88406	87744	87391	89766	88032	86887
10-19	1303	2027	2224	265	2028	3172
20-29	277	269	332	51	28	23
30-39	70	43	91	8	2	8
40-49	22	1	32	0	0	0
50-59	8	5	11	0	0	0
60-69	4	1	9	0	0	0

Table 4. Number of indications by amplitude for six test wheels.

Based upon the results shown in Table 4, three of six wheels failed the AAR's ultrasonic rejection requirement. This requirement is 50 percent of the amplitude associated with a 1/8-inch (3.2 mm) flat bottom hole, and due to 80 percent maximum screen height used, rejection would take place at 40

percent screen height amplitude. Note that the first three wheels were found to have a greater number of larger indications (40 percent amplitude and greater).

7. MICROCLEANLINESS RESULTS AND DISCUSSION

Three samples were removed from each of the six test wheels and were sent to Cambridge Materials Testing, Ltd., in Cambridge, Ontario, Canada, for quantitative image analysis testing using BNSF and ASTM E1245 procedures. Results for the tests are shown in Table 5. Results are averages of the three samples for all parameters except the "worst field" parameters - "Maximum volume percent voids plus oxides" and "Maximum volume percent sulfides." For these parameters, the single worst value encountered in the three metallographic samples is entered.

Inclusion Type	Microcleanliness Parameter	Wheel A	Wheel B	Wheel C	Wheel D	Wheel E	Wheel F
Voids + Oxides	Mean Volume %	0.072	0.078	0.05	0.052	0.086	0.067
	Count per square mm	49.7	50.7	32	31.7	57.3	45
	Maximum Volume % (worst field)	0.16	0.26	0.14	0.13	0.18	0.18
Sulfides	Mean Volume %	0.117	0.12	0.103	0.088	0.123	0.113
	Count per square mm	37.3	35.7	30	33.7	40.3	39.3
	Maximum Volume % (worst field)	0.24	0.26	0.21	0.22	0.39	0.23

Table 5. Microcleanliness results for six test wheels, three samples taken per wheel.

An attempt was then made to correlate the various microcleanliness parameters with ultrasonic indication counts for each wheel. Indications less than 10 percent amplitude were disregarded due to their very small size and the difficulty in resolving them from noise, etc.

The number of indication counts from 10 percent amplitude to 69 percent amplitude were summed for each wheel and the statistic "r" was calculated in a spreadsheet vs. the various microcleanliness parameters. Results are contained in Table 6.

Correlation Coefficient	Voids + Oxides			Sulfides		
	Mean Volume %	Count per square mm	Maximum Volume %	Mean Volume %	Count per square mm	Maximum Volume %
r value	0.231	0.280	0.396	0.574	0.177	0.031

Table 6. Correlation coefficient for correlation between number of ultrasonic counts (10% to 69% amplitude) and the various microcleanliness parameters.

Results show that there is poor correlation between all comparisons, except perhaps between the number of ultrasonic counts and the mean volume percent of sulfides. We offer some possible reasons for the poor correlation between ultrasonic testing and microcleanliness testing, as follows:

1. The microcleanliness samples are analyzed on a polished plane that is parallel to the wheel tread surface. During the wheel forging process, any non-metallic discontinuities in the rim are "flattened out" approximately parallel to the tread surface. This means that such discontinuities are best detected by the axial ultrasonic scan (from the back rim face). Due to the nature of the forging process, rejectable indications are seldom found using the radial ultrasonic scan (from the tread surface). Therefore, since the ultrasonic test used in this work was evaluating the plane perpendicular to the tread surface, orientation of discontinuities is likely an important factor.

2. The microcleanliness test evaluates only one polished plane of wheel rim material per sample, while the ultrasonic test evaluates nearly the entire volume of rim material. Thus more total discontinuities will be detected using ultrasonic methods while the polished planes may not be representative of the wheel rim as a whole.

3. By disregarding indications less than 10 percent of the amplitude, we may be ignoring a significant number of small inclusions that would otherwise be detected and measured using quantitative image analysis techniques. For example, typical alumina inclusions in wheel steel are less than 0.125-mm in length. Thus, significant numbers of inclusions counted and included in the microcleanliness data may be ignored in the ultrasonic data.

8. SPARK-DAT RESULTS AND DISCUSSION

The same eighteen polished metallographic samples (three samples each, for six wheels) used for microcleanliness testing were shipped to Thermo Electron Corporation for evaluation using their Spark-Dat analysis method. Samples were shipped and tested using Spark-Dat within approximately one to two weeks of the earlier microcleanliness tests at Cambridge. Spark-Dat total counts were collected for Ca, Al_2O_3, Al, O, CaO, MnS and S, along with other elements and compounds. Four Spark-Dat readings (10 seconds each) were taken on each sample – therefore twelve readings were collected for each wheel. An average total count for each of the six wheels (A through F) was then calculated.

Spark-Dat values were first collected on each sample's polished surface, and subsequently were also collected on the same specimen after it had been ground with 120 SiC grit. Given the better correlation shown for earlier test readings taken with a polished surface (see Table 1), polished surface results were used for subsequent correlations in this paper.

Results of statistical correlation calculations (made in a spreadsheet) between total ultrasonic indication counts (10 percent amplitude to 69 percent amplitude) vs. average Spark-Dat counts for various elements are shown in Table 7. The highest correlation values are seen for calcium, titanium oxide, aluminum, calcium oxide and oxygen.

Correlation Coefficient	Ca	Al_2O_3	Al	O	CaO	TiO	MnS	S
r value	0.91	0.57	0.65	0.60	0.63	0.87	0.27	0.37

Table 7. Correlation coefficient for ultrasonic counts vs. Spark-Dat counts.

Results of statistical correlation calculations between average microcleanliness parameters and average Spark-Dat counts are shown in Table 8. Voids and oxides microcleanliness parameters were correlated to those elements and compounds associated with non-metallic oxides, while sulfide microcleanliness parameters were correlated to manganese sulfides and sulfur. The best correlation is seen for sulfides count per square mm vs. Spark-Dat manganese sulfides (r = 0.84) and sulfides count per square mm vs. Spark-Dat sulfides (r = -0.71). Other relationships showed a weaker correlation.

Inclusion Type	Microcleanliness Parameter	Correlation Coefficient, r value							
		Ca	Al_2O_3	Al	O	CaO	TiO	MnS	S
Voids + Oxides	Mean Volume %	0.33	-0.08	0.013	-0.13	-0.02	0.47		
	Count per square mm	0.38	-0.14	0.002	-0.16	-0.07	0.53		
	Max. Vol. % (worst field)	0.17	0.30	0.007	0.31	0.31	0.26		
Sulfides	Mean Volume %							0.38	0.018
	Count per square mm							0.84	-0.71
	Max. Vol. % (worst field)							0.56	-0.20

Table 8. Correlation coefficient for microcleanliness parameters vs. Spark-Dat counts.

The results in Tables 7 and 8 show a generally better correlation between ultrasonic counts and Spark-Dat values than between microcleanliness data and Spark-Dat values. It must be noted that algorithms for Spark-Dat have not been optimized for wheel steel. Further development work in this area could lead to better correlation with other methods. Also, as mentioned earlier in the previous section, a significant number of small inclusions that would be detected and measured using Spark-DAT may be ignored in the ultrasonic data by disregarding indications less than 10 percent of the amplitude. Typically, Spark-DAT is best suited for detecting and identifying inclusions in the range of 0.002 to 0.02-mm.

9. CONCLUDING REMARKS

The authors realize that this study uses a relatively small sample size of six wheels. Further, only three samples per wheel were used for microcleanliness and Spark-Dat testing - normally six samples per wheel are used for microcleanliness testing. Therefore, it is difficult to reach definitive conclusions regarding the suitability or superiority of a method. However, some general statements and recommendations for future work are made as follows:

1. Although AAR has required wheel manufacturers to ultrasonically test new wheels for many years and more recently has required them to perform microcleanliness testing on new wheels, we found poor correlation between ultrasonic testing and the various microcleanliness parameters. Given that ultrasonic testing systems, particularly those automated systems that use new modern phased array transducers, have the ability to test the volume of a wheel rim, and are accepted for industrial product evaluation, ultrasonic testing provides a useful benchmark for evaluation of alternative methodologies.

2. New computer software for ultrasonic phased array systems provides the ability to save scans and to obtain additional information regarding product quality. Thus, there is the potential to collect and evaluate data as part of a process-improvement effort, as suggested by Giammarise and Gilmore [11] at the 13th International Wheelset Congress. Further development work with ultrasonic testing techniques could lead to manufacturing process improvements.

3. Correlation between ultrasonic testing and Spark-Dat values was somewhat better. This suggests that further development of Spark-Dat algorithms specific to wheel steel may be useful as an alternative cleanliness measure. Correlation between microcleanliness and Spark-Dat was poor except for with respect to sulfide inclusions. Future work in this area should focus on the effect of specimen preparation, development of calibration standards, and studies to determine the required sample size to obtain acceptable confidence levels.

10. ACKNOWLEDGEMENTS

Thanks to Mr. Tom Carodiskey of GE Inspection Technologies, Lewistown, PA for his ultrasonic testing work on this project.

11. REFERENCES

1. Richard Sullivan, Cameron Lonsdale and John Oliver, “2002 AAR Car Repair Billing Wheel Removal Analysis,” Proceedings of 2003 Railway Wheel Manufacturers’ Engineering Committee Technical Conference, C.P. Lonsdale, editor, Chicago, IL, September 23, 2003, pp. 19-28.

2. R. Lunden, “Cracks in Railway Wheels Under Rolling Contact Load,” *Proceedings of the 10th International Wheelset Congress*, Australia, 1992, pp. 163-168.

3. J. J. Marais, “Wheel Failures on Heavy Haul Freight Wheels Due to Subsurface Defects,” *Proceedings of the 12th International Wheelset Congress*, Quingdao, China, 1998, pp. 306-314.

4. Xing Lixian, Zhang Yingzhi and Lin Jizhong, “Study on Rim Fatigue Crack and Prevention,” *Proceedings of the 12th International Wheelset Congress*, Quingdao, China, 1998, pp. 53-58.

5. A. Ekberg and J. Marais, “Effects of Imperfections on Fatigue Initiation in Railway Wheels,” *Proceedings of the Institution of Mechanical Engineers*, 2000, Vol. 214, Part F, pp. 45-54.

6. D. H. Stone, “Wheel Shattered Rims – An Interpretive Review,” *Wheels and Axles, Cost Effective Engineering*, ImechE Seminar Publication 2000-20, London, 2000, pp. 75-84.

7. S. Baretta et al., “Deep Shelling in Railway Wheels,” *Proceedings of the 13th International Wheelset Congress*, Rome, Italy, September 2001.

8. Daniel H. Stone and Geoffrey E. Dahlman, “The Effect of Discontinuity Size on the Initiation of Shattered Rim Defects,” *ASME Rail Transportation Division Volume 19*, 2000, pp. 7-14.

9. Geoffrey E. Dahlman and Cameron Lonsdale, “Strategies to Prevent Heavy Haul Wheel Failures,” *Proceedings of the International Heavy Haul Association Specialist Technical Session*, Dallas, Texas, May 4, 2003, pp. 2.9 - 2.15.

10. Cameron Lonsdale, John Traxler, Richard Wagner and Paul Meyer, “Use of Phased Arrays for Ultrasonic Inspection of New Wrought Railroad Wheel Rims,” *Proceedings of the 13th International Wheelset Congress*, Rome, Italy, September 2001.

11. A. W. Giammarise and Robert Gilmore, “Wheel Quality, A North American Locomotive Builder’s Perspective,” *Proceedings of the 13th International Wheelset Congress*, Rome, Italy, September 2001.