

# COMPARATIVE STUDY OF VIBRATION CONDITION INDICATORS FOR DETECTING CRACKS IN SPUR GEARS

**Nenad Nanadic**

nxnasp@rit.edu  
Rochester Institute of Technology  
Rochester, NY

**Michael Thurston**

Rochester Institute of Technology  
Rochester, NY

**Paul Ardis**

Rochester Institute of Technology  
Rochester, NY

**Anindya Ghoshal**

U.S. Army Research Laboratory  
Aberdeen, MD

**Adrian Hood**

U.S. Army Research Laboratory  
Aberdeen, MD

**David Lewicki**

NASA Glenn Research Center  
Cleveland, OH

This paper reports the results of an empirical study on the tooth breakage failure mode in spur gears. Of four dominant gear failure modes (breakage, wear, pitting, and scoring), tooth breakage is the most precipitous and often leads to catastrophic failures. The cracks were initiated using a fatigue tester and a custom-designed single-tooth bending fixture to simulate over-load conditions, instead of traditional notching using wire electrical discharge machining (EDM). The cracks were then propagated on a dynamometer. The ground truth of damage level during crack propagation was monitored with crack-propagation sensors. Ten crack propagations have been performed to compare the existing condition indicators (CIs) with respect to their: ability to detect a crack, ability to assess the damage, and sensitivity to sensor placement. Of more than thirty computed CIs, this paper compares five commonly used: raw *RMS*, *FM0*, *NA4*, raw kurtosis, and *NP4*. The performance of combined CIs was also investigated, using linear, logistic, and boosted regression trees based feature fusion.

## Introduction

The U.S. Army has the goal of transitioning from time-based to condition-based maintenance for its fleet of vehicles in an effort to improve safety and reduce costs. The premise is that critical components are serviced when indicators reveal that they can no longer function as designed. A major focus has been on the helicopter transmission and researchers have proposed vibration-based CIs to detect faulty components, particularly bearing and gears. (Refs. 1–10) However the availability of damage cases to evaluate and validate them are scarce. This research addresses the need for seeded fault data, as it pertains to gear tooth cracks, by developing a well-documented, statistically significant database of monitored cracks from inception to failure.

Gear service life can be divided into two phases: crack initiation and crack propagation. (Refs. 11–13) Because crack initiation takes considerably longer than propagation, an accelerated method was employed by subjecting a single tooth to a cyclic force above its intended operating range. (Refs. 14, 15) Traditional crack seeding uses wire Electrical Discharge Machining (EDM), where a gear is notched. (Refs. 16, 17) However, in this study, a single tooth was subjected to fatigue crack initiation. A potential advantage of the fatigue-based crack seeding vs. notching is that the radius of curvature of the initiated cracks are considerably smaller. The smaller cracks propagate faster and better represent real field failures. Seeding cracks have an additional advantage: it isolates the loca-

tion of breakage and thus enables manageable monitoring of the ground truth.

The objective of the current study is to compare the performance of vibration based CIs to detect incipient cracks in gear teeth. It is an extension of a methodology described in (Ref. 15) in which crack initiation was conducted on a high cycle fatigue test rig and propagated on a 4-square spur gear fatigue test rig. This current study also employed crack-propagation sensors (CPS) to obtain the ground truth information on the damage as was done in (Ref. 18).

Although many CIs are computed and stored in a database, this report will focus on five. The rationale for selecting this subset is given as follows: Root Mean Square (*RMS*) and Kurtosis (*Kur*) are chosen for their simplicity and long history of use in vibration monitoring. *FM0* was selected for being one of the first CIs, first proposed by Stewart in 1977 (Ref. 5). *NA4* was chosen because it has had considerable success, after being introduced by Zakrajek et al. in 1998 (Ref. 3). Finally, *NP4*, a more recent features, introduced by Polyshchuk et al. in 2003 uses a time-frequency method. (Ref. 9). Descriptions of many CIs can be found in (Refs. 1, 2).

Table 1 summarizes formulae for the five CIs considered. The *RMS*, Kurtosis, and *NP4* are based on raw signals while *NA4*, and *FM0* are based on time synchronously averaged data.  $x(t)$  is the measured signal,  $N$  is the number of data points,  $P$  to  $P$  is the peak-to-peak amplitude of  $x(t)$ ,  $A_k$  is the amplitude of the gear harmonics  $k$ ,  $r(t)$  is the residual signal derived by removing known frequency components,  $P_n$  is the signal power and  $n$  represents the number of gear mesh harmonics removed in  $r(t)$  and  $M$  is the current time record in run ensemble.

Table 1: CI Table

Metric	Formula
RMS	$\sqrt{\frac{1}{N} \sum_{i=1}^N x^2(t_i)}$
$FM0$	$P \text{ to } P / \sum A_k$
$NA4(M)$	$\frac{N \sum_{i=1}^N (r_i - \bar{r})^4}{\frac{1}{M} \sum_{j=1}^M [\sum_{i=1}^N (r_{ij} - \bar{r}_j)^2]^2}$
Kurtosis	$\frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^2}$
$NP4(n)=$	$\frac{1}{N} \sum_{i=1}^N \left(\frac{P_n(t_i) - \bar{P}_n}{\sigma}\right)^4 - 3$

## Test Description

The gears were designed and manufactured according to NASA drawing and specifications with 28 teeth, a diametrical pitch of 8, and a pressure angle of 20 degrees. More details on the specifications can be found in Townsend and Shimski (Ref. 19). Figure 1 depicts a flowchart of the testing procedure. For each test set, the two test gears were spun

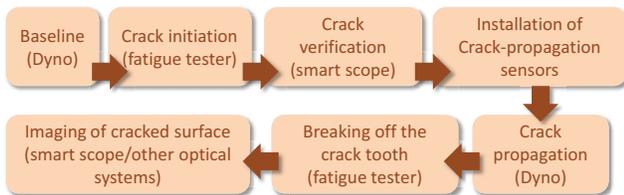


Fig. 1: Testing Flowchart

together for 54.5s in a gear box instrumented with four accelerometers whose positions are shown in Figure 2. Some of these are placed at locations very sensitive to gear cracks, and some at locations that are expected to have relatively poor sensitivity to gear crack. Sub-optimal placement was motivated by the fact that one must use pragmatic, sub-optimal sensor placement in practical applications due to space and other constraints. The torque and angular speed profile is given in Figure 3. The repeating cycle is indicated in the graph with  $T$ . This established the baseline for the test. The context (operating condition) profiles were designed to propagate the cracks effectively, but also to gain understanding of the effect of speed and torque on gear fault detection condition indicators. The dynamic portion of the profile is used to determine the dependence of CIs on changing operating conditions. Only CI values for fixed operating conditions of  $\tau = 170$  ft-lb and  $\omega = 1500$  rpm will be presented.

After the baseline tests were completed, the top gear was removed from the gearbox, mounted in the single-tooth, fatigue-tester fixture shown in Figure 4, and subjected to a cyclic load of 100 to 3100 lbs at 10Hz. The anvil applied a normal force to a single tooth at the highest point of single tooth contact (HPSTC) as shown in Figure 4c. Controlling the crack size was achieved using the measured compliance of the fatigue rig as feedback. (Ref. 14) It was desired to create a crack approximately equal in size.



Fig. 2: Dynamometer test stand (gearbox) with the location of the accelerometers.

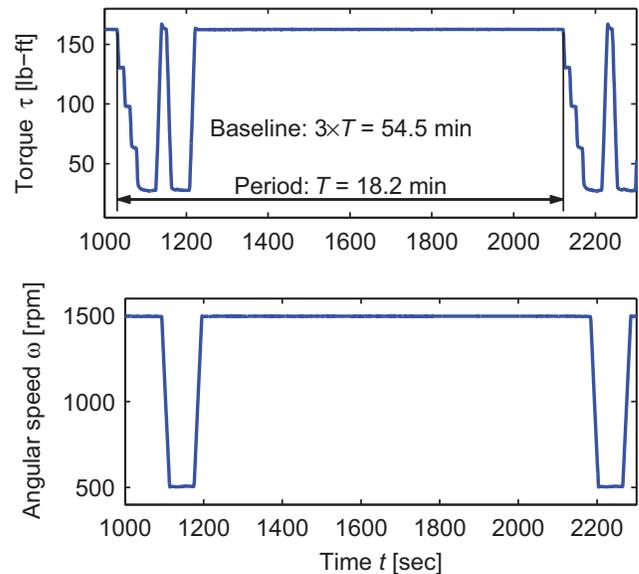


Fig. 3: The segment of the dynamometer operating conditions, viz. torque and angular speed. The repeating cycle is indicated in the graph with  $T$ .

The initiated cracks were verified in a non-destructive manner using florescent magnetic particle inspection. Before inspection, the gear is placed in a custom fixture designed to place the tooth in tension of approximately 500 lb. The fixture is shown in Figure 5. The tooth is covered with a layer of florescent iron powder, placed in strong magnetic field, and observed under ultraviolet (UV) light. The magnetic particles coalesce at the flux leakage caused by the crack and become detectable as shown in 6a.

Once the crack was initiated and verified, the gear was equipped with two crack-propagation sensors, one on each face as shown in Figure 6b. The gear was then re-assembled in the gearbox and operated according to the same profile of Figure 3. Testing was concluded when the crack propagated beyond the range of both crack-propagation sensors. The propagation times ranged from 39 minutes to over 3 days and a total of 10 set of gears were tested as shown in Table 2. Often, the

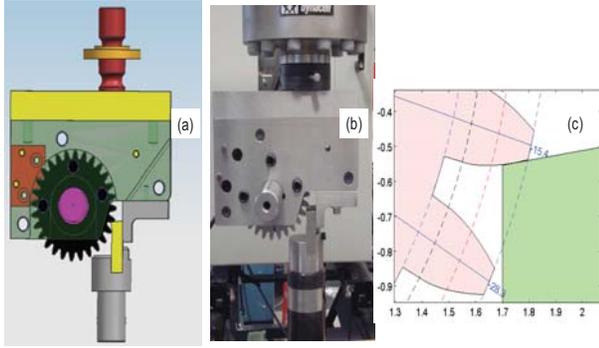


Fig. 4: Fatigue-tester-based, single-tooth fixture. (a) CAD model (b) Photo of the built fixture. (c) The force is applied approximately perpendicularly at HPSTC

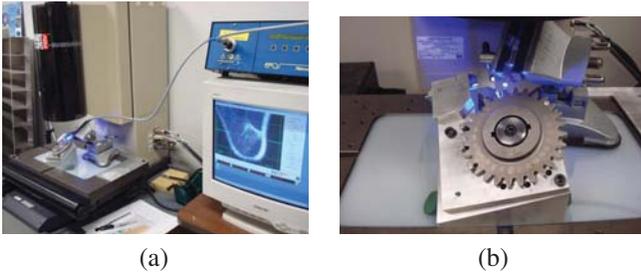


Fig. 5: Non-destructive method for crack verification (a) Imaging system (b) Custom gear fixture

system needed to shut down in the evening and be restarted in the morning. Propagation was run until crack propagation sensors indicated full propagation. At the completion of each test, the tooth was then completely severed using the fatigue tester in order to gain a metallurgical understanding of the fatigue cracks. The bottom image of the tooth was imaged as shown in Figure 7.

## CI Database

A MySQL database was developed with the intention of serving the gear research community by providing the ground truth for further development of diagnostics and prognostics of gear crack failure mode based on vibration signals. The tables containing measured data cannot be altered by the user, but the database schema allows the user to recompute CIs as well as append and compute additional CIs. The

Table 2: Summary of test runs

GearID	Total propagation time [min]
13	1809.1
16	255.8
20	506.2
104	4693.3
106	614.5
108	52.6
112	99.8
114	79.9
116	47.5
118	38.7

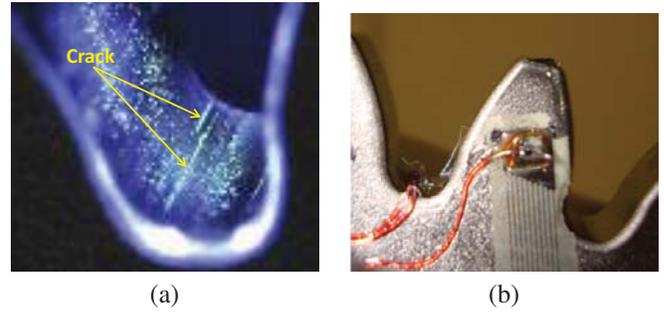


Fig. 6: (a) Initiated crack (b) Gear tooth with crack propagation sensor



Fig. 7: The bottom surface of a cracked tooth.

database tables are shown in Figure 8. The crack initiation data is stored into two main tables: *instron\_test\_details* and *instron\_test\_data*. The former table contains the information on the test parameters (date, force magnitude, force frequency, data acquisition parameters, etc.). The latter table shows the measurements (force, displacements, computed compliance, etc.). Similarly the dynamometer data has a separate table for test parameters and for data. However, the high-frequency data bursts are stored in linked binary files.

## CI Computation and Analysis

The CIs are compared with respect to their: ability to detect a crack early, ability to assess the damage, and sensitivity to sensor placement. While one can speculate that these CI attributes may be related, and expect that a CI that is more sensitive initially will remain more sensitive through the propagation, there is currently no evidence to support this. Moreover, a previous study (Ref. 10) suggested fusion of features that showed more sensitivity for crack initiation and features more sensitive to crack propagation.

### CI Comparison: Early Crack Detection

Crack detection is examined from two different points of view: 1) ability to detect cracks early, and 2) ability to rapidly achieve high confidence of the crack as the crack grows. To compare CIs' ability to detect the crack early, the features computed during the baseline test were labeled as no fault and

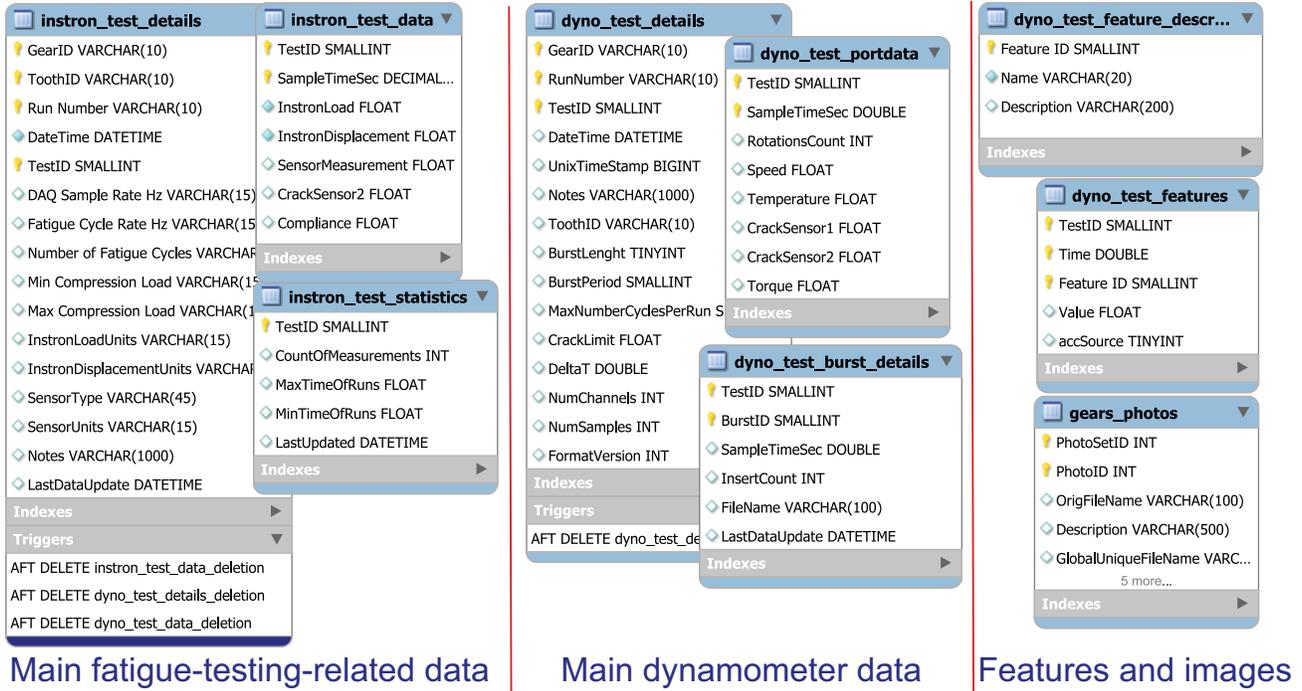


Fig. 8: The main tables in the database.

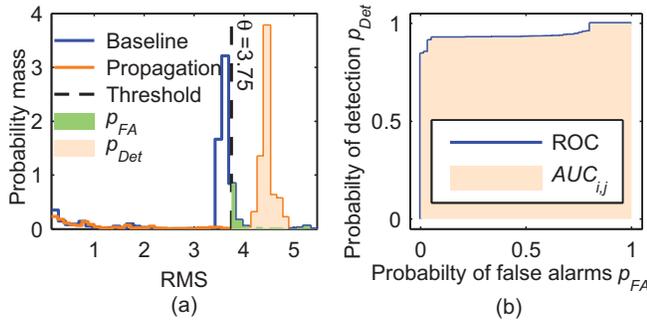


Fig. 9: RMS: (a) Computing an ROC point ( $p_{FA}$ ,  $p_{Det}$ ). (b) ROC and  $AUC_{i,j}$

the features computed during the propagation test prior to the first detected CP sensor strand breakage as fault.

Receiver operating characteristic (ROC) plots were used for performance comparisons. Two probability distribution functions (PDFs) of the CIs are produced from the data: one associated with the known undamaged case and one for the known damaged case as shown in Figure 9a. The ROC can be thought of as a measure of overlap of the two PDFs. The ratio of the detected faults to all faults is plotted against the ratio of false detections, as the threshold was varied (Refs. 20–22) Figure 9a illustrates what a point on an ROC curve represents. The CI used in this example is the *RMS*. An example ROC curve with the associated area under the curve (*AUC*) are shown in Figure 9b. Figure 10a illustrate dependence on resulting ROC on the sensor placement and Figure 10b shows how this ROC varies for different gears using the signal from  $a_4$ .

These plots provide an initial glimpse of the ability of dif-

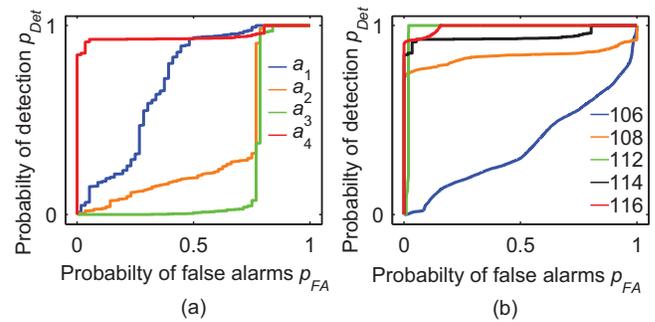


Fig. 10: RMS: (a) ROCs from four different accelerometers and a single gear (b) ROCs from several gears (using  $a_4$ ).

ferent CIs to detect small cracks. The ROC is a two dimensional depiction of classifier performance, which makes these plots somewhat busy for multiple CIs and multiple accelerometers. A more compact comparison is enabled using the area under the ROC curve, abbreviated as  $AUC_{i,j}$  for gear  $i$  and accelerometer  $j$ . (Ref. 23) *AUC* is a single scalar obtained by integrating ROC. Higher *AUC* indicates better performance, with  $AUC = 1$  being the maximum, associated with perfect detection. The summary comparison plot for early detection is shown in Figure 11a, which shows  $M_{AUC,i}$  vs CI, where

$$M_{AUC,i} = \max_j(AUC_{i,j}). \quad (1)$$

Each gear is shown with a unique marker. A histogram is also provided to indicate the number of overlapping markers.

Within our sample, raw RMS exhibits the best average performance, smallest gear-to-gear variation, and does not show

## Data Fusion

To demonstrate the benefit of CI fusion, the complementary nature of CI information, by way of a panel of feature fusion experiments, was evaluated. In these experiments, the provided CIs were fused via three different predictive models: LiNear Regression (LNR) (Ref. 24), LoGistic Regression (LGR) (Ref. 25), and Boosted Regression Trees (BRT) (Ref. 26). LNR and LGR fit singular lower-dimensional models to the data as follows: LNR finds a two-parameter linear (slope and intercept) model that minimizes the sum of squared error between modeled values and observed values, while LGR finds a similar parametrization for a logit function in place of the linear function. The logit function,  $\text{logit}(x)$ , is the inverse of the sigmoid function, defined as

$$\text{logit}(x) = \ln\left(\frac{x}{1-x}\right), \quad (2)$$

BRT, in contrast, learn a fixed (constant) number of lower-dimensional models over discrete regions of the feature space, forming a single piecewise function. In the experiments performed, the BRT region boundaries and individual region functions were linear functions, with parameter selection based upon Adaptive Boosting and the number of regions fixed at 10. For all algorithms, the regression targets were based upon the observed wire breakages, with final classification results performed by thresholding the predicted output value for a given set of input CI readings. In all cases, the three top-performing CIs were used as inputs, with the hypothesis that this arrangement should outperform single-CI crack detection. Features computed from one accelerometer were fused during these experiments, although additional benefit may be achieved by fusing features computed from vibration signals at different locations. As before, the largest of four AUCs was selected.

Figure 12a shows the results after fusing three of the best performing CIs: *RMS*, *NA4*, *NP4*. These results show that performance trends with the best constituent CIs, but that CIs with strongly varying performance can reduce the overall scores and introduce variance in the final result.

### CI Comparison: Sensor Placement

To compare CIs with respect to their sensitivity to accelerometer placement, plots of the standard deviation of AUC's for individual cracks defined as

$$\sigma_{AUC_i} = \sqrt{\frac{1}{N-1} \sum_{j=1}^4 (AUC_{i,j} - \mu_{AUC_i})^2} \quad (3)$$

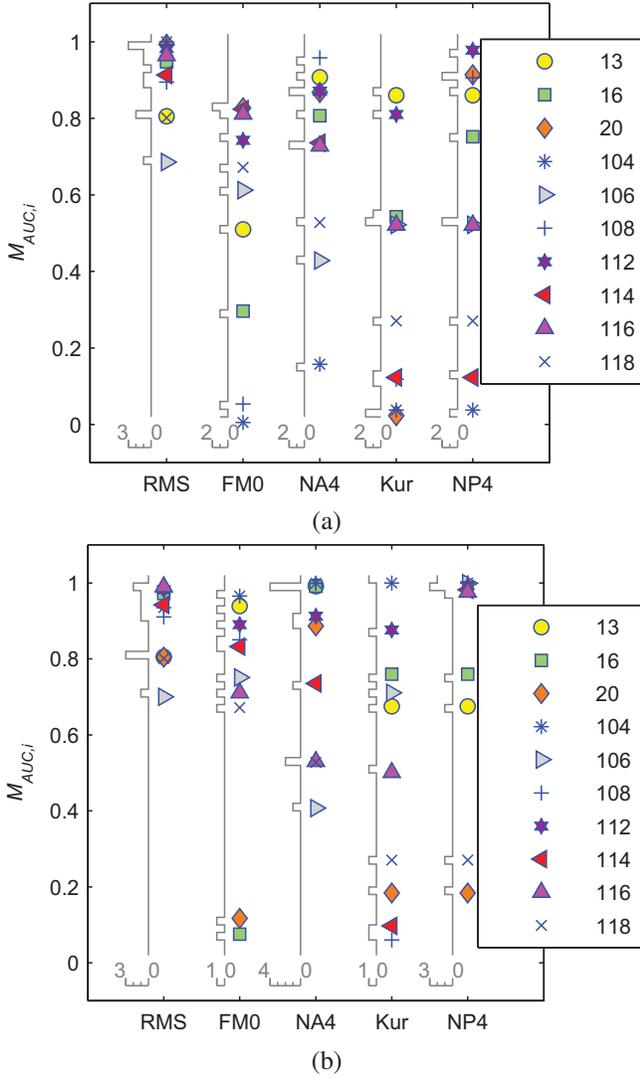


Fig. 11: Summary performance of different features. (a) Early detection. (b) Entire propagation

large outliers as other features. *FM0*, *NA4*, Kurtosis, and *NP4* all have cases of false positives, which is manifested by  $M_{AUC,i}$  outliers. It is also interesting to note that different CIs perform differently on different crack propagations. For example, the crack on GearID = 114 was missed by *NP4*, and Kurtosis, but not by *FM0*, *NA4*, and *RMS*.

It is of interest to see how the performance improves over time. Figure 11b shows the performance of the CIs when all crack propagation data is taken into account. While the performance of all indicators improved, *RMS* was still the only one without significant outliers. However, *NA4* and *NP4* show peaks in  $M_{AUC,i}$  distributions at high probability of detection (POD), indicating better performance for a subset of gears. Also, as mentioned above, CIs seem complementary. Thus, the results confirm earlier suggestions that detectors employing more than one CI are promising.

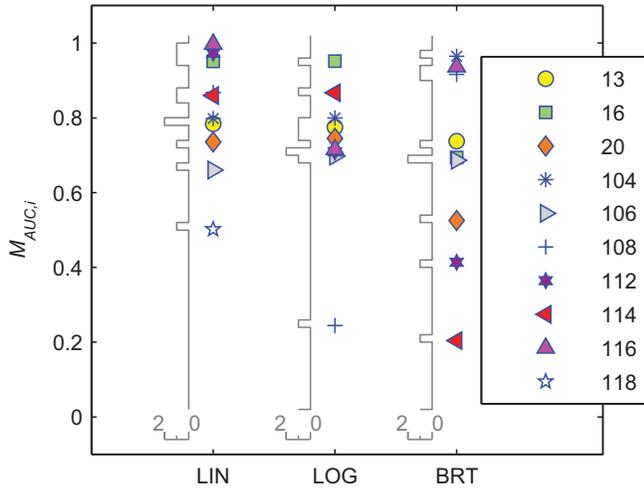


Fig. 12:  $MAUC_i$  after fusing top performers (raw  $RMS+NA4+NP4$ )

where  $AUC_i$  is their mean defined by

$$\mu_{AUC_i} = \frac{1}{N} \sum_{j=1}^4 AUC_{i,j} \quad (4)$$

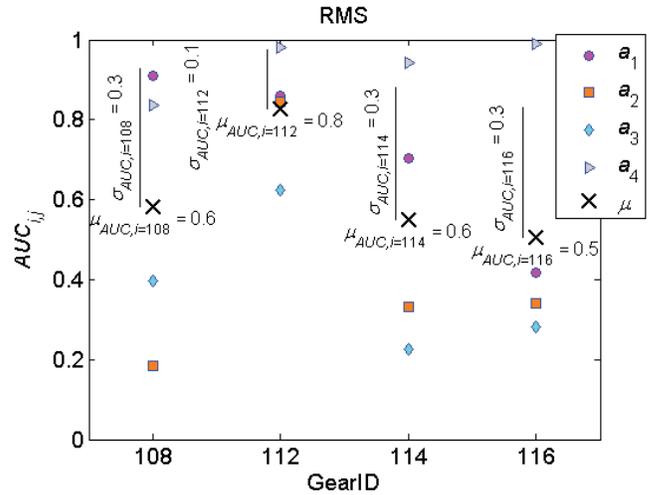
are produced.

Figure 13a illustrates the process for four select gears (GearID = 108, 112, 114, and 116). The gears are indicated on the x-axis and  $AUC_{i,j}$  values on the y-axis. The source accelerometers, denoted in the plot as  $a_1$  through  $a_4$ , are consistently labeled with unique markers. In addition, their mean,  $\mu_{AUC_i}$ , is also plotted and labeled with an 'x' marker'. The values for the mean and standard deviation  $\sigma_{AUC_i}$  are also indicated in the plot.

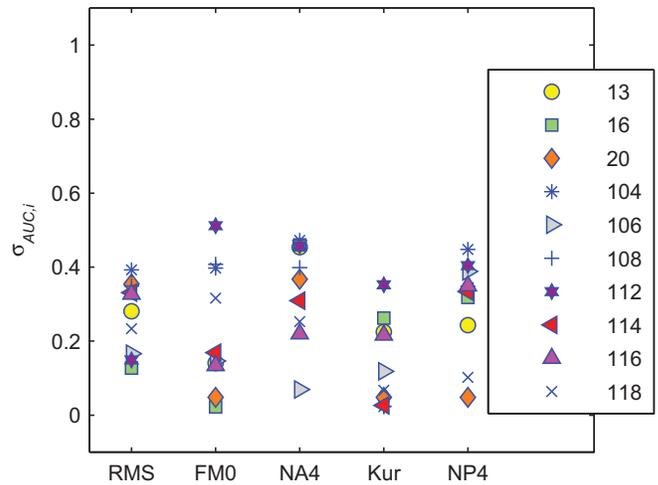
If a CI was perfectly insensitive to sensor placement, the standard deviation would be, ignoring minute differences due to sensor calibration and the processing channels, zero, because CIs originating from differently placed accelerometers would be indistinguishable. Figure 13b shows the summary plot of standard deviations of  $AUC$ s of features computed from different data captured by differently placed accelerometers for the same propagations. The CIs exhibit similar dependence on sensor placement, as estimated by  $\sigma_{AUC_i}$ , but raw kurtosis seems slightly better than the others. The dependence varies sample-to-sample, and the raw  $RMS$  displays the least gear-to-gear variation. Note that very low sensitivity to sensor placement often corresponds to a missed alarm. For example, in Figure 13b, GearID = 16 has the smallest  $\sigma_{AUC}$  for  $FM0$ , but Figure 11b shows that this crack detection was missed by the CI.

### CI Comparison: Damage Assessment

CIs are correlated with estimated crack size in order to compare them with respect to their ability to assess the dam-



(a)



(b)

Fig. 13: (a)  $AUC_{i,j}$  vs. GearID for four accelerometers (b) Sensitivity of CIs to sensor placement

age. The first-order ground truth information on damage was contained in the crack propagation sensor signals,  $CP_1$  and  $CP_2$ . Figure 14 is a schematic for the CPSs. There are two CP sensors, one on each gear face. Figure 15 is an example

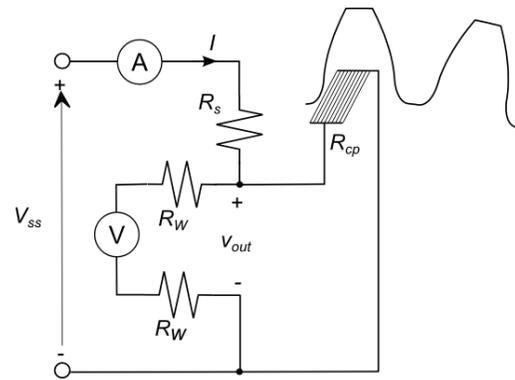


Fig. 14: Crack Propagation Sensor circuit schematic

output. The resulting voltage is noisy and requires averaging

and some signal processing, including filtering and peak detection. Breaking of CP wires give rise to the recorded voltage level as shown in Figure 16a. The spacing between centerlines is .25mm. The CPS estimates are saved in the same database table as the CIs, *dyno-test-features*.

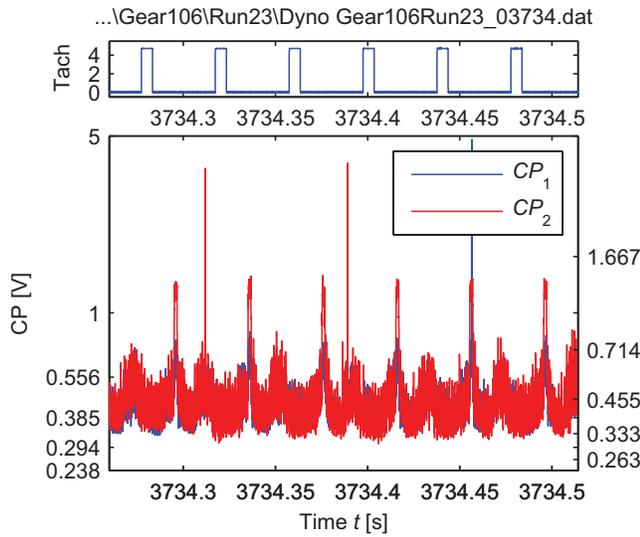


Fig. 15: CPS signal

The CP1-CP2 plane is a parameter space spanned by the two voltages CP sensors. It maps directly into the crack lengths as measured on the two gear faces. Figure 16b shows three different crack propagations (gearID = 108, 114, and 116) and illustrates that a crack can propagate fairly symmetrically (Gear 108), but also asymmetrically (114 and 116). Note that the asymmetry for 114 and 116 is in different directions.

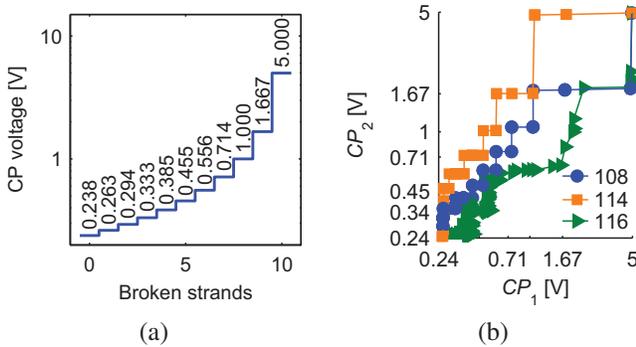


Fig. 16: (a) theoretical CP levels (b) Three crack surface paths in  $CP_1$ - $CP_2$  plane.

The following plots show each CI as a function of crack length. All CIs are computed from the data from the most sensitive accelerometer. Figures 17a-b plots raw *RMS* vs.  $CP_1/CP_2$  and Figures 17c-d plots *FM0* vs.  $CP_1/CP_2$ . In a like manner, Figures 18a-b plots the results for *NA4* and Figures 18c-d for the Kurtosis. The *NP4* results are given in Figure 19. All CIs showed relatively weak dependence on the level of damage, as estimated by surface sensors  $CP_1$  and  $CP_2$ .

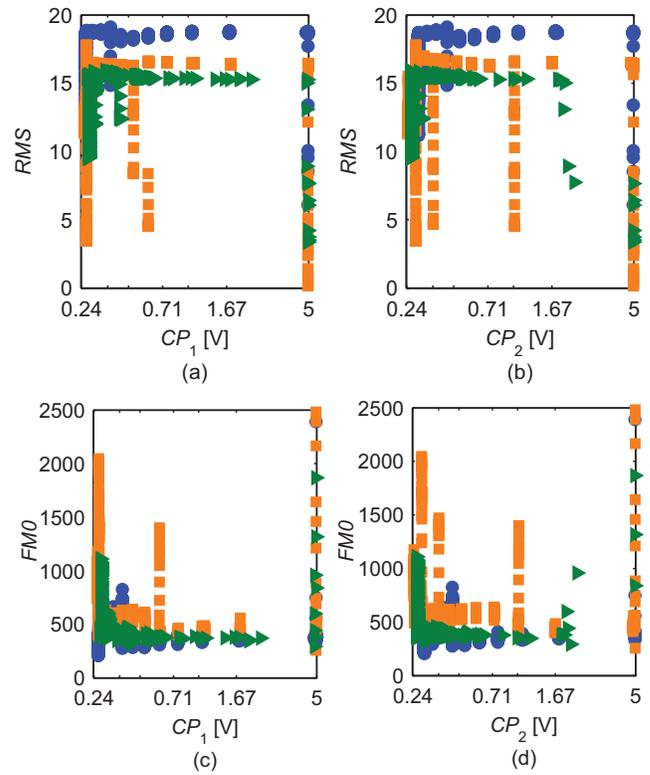


Fig. 17: Dependence of CIs on the damage level, measured by  $CP_1$  and  $CP_2$  for four representative propagations. *RMS* and *FM0* as a function of  $CP_1 / CP_2$   $\circ$ : 108,  $\square$ : 114,  $\triangle$ : 116

## Conclusions

The research project consisted of accelerated crack growth in spur gears and measuring vibrations using accelerometers mounted on the gearbox. The cracks were initiated using a fatigue tester. Crack propagation sensors, mounted on each face of the cracked tooth, captured the ground truth information on crack propagation. Signals from four accelerometers and the tachometer were used to compute condition indicators. Five CIs were selected from a larger set and compared with respect to their ability to detect small cracks, their sensitivity to sensor placement, and their ability to assess the damage. While simple raw *RMS* was found to be the most robust for early crack detection, more advanced CIs exhibited higher *AUCs* for a subset of gears. In addition, evidence of the complementary detection power of different features encouraged feature fusion for improved performance. It was found that combining all features does not always improve the overall performance. The compared features showed similar sensitivity to sensor placement. The initial analysis did not detect a consistent significant sensitivity to damage of any of the selected CIs.

## References

<sup>1</sup>Mosher, M., Pryor, A., and Huff, E. M., "Evaluation of standard gear metrics in helicopter flight operation," 56th

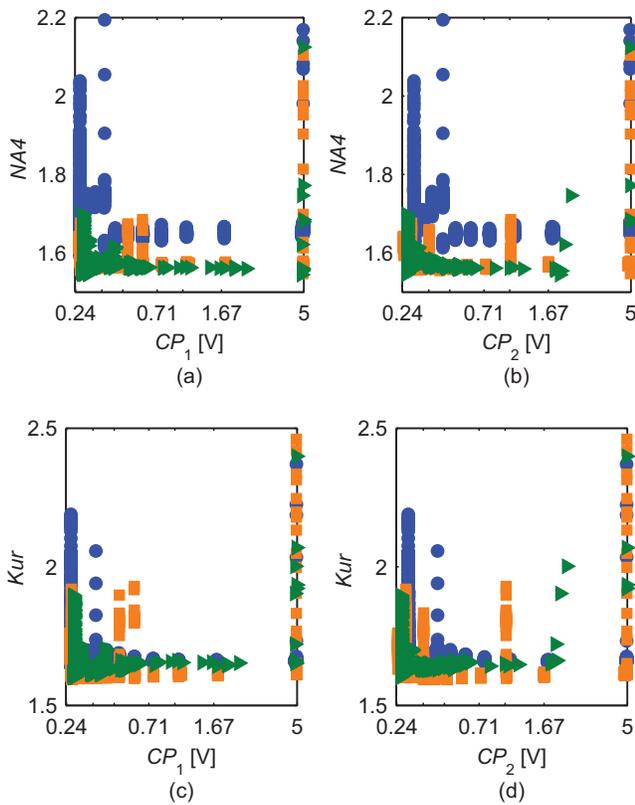


Fig. 18: Dependence of CIs on the damage level, measured by  $CP_1$  and  $CP_2$  for four representative propagations. NA4 and Kur as a function of  $CP_1 / CP_2$  ○:108, □: 114, ▷: 116

Meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, VA, 2002.

<sup>2</sup>Samuel, P. and Pines, D., “A Review of Vibration-Based Techniques for Helicopter Transmission Diagnostics,” *Journal of Sound and Vibration*, Vol. 282, 2005, pp. 475–408.

<sup>3</sup>Zakrajsek, J. J., Townsend, D. P., and Decker, H. J., “An Analysis of Gear Fault Detection Methods as Applied to Pitting Fatigue Failure Data,” Technical Memorandum 105950, 1993.

<sup>4</sup>Decker, H., Handchuh, R., and Zakrajsek, J., “An Enhancement to the NA4 Gear Vibration Diagnostic Parameter,” Technical Memorandum/Technical Report NASA TM 106553/ARL-TR-389, 1994.

<sup>5</sup>Stewart, R. M., “Some Useful Data Analysis Techniques for Gearbox Diagnostics,” Technical Report MHM/R/10/77, 1977.

<sup>6</sup>Martin, H., “Statistical Moment Analysis As a Means of Surface Damage Detection,” Proceedings of the 7th International Modal Analysis Conference, 1989.

<sup>7</sup>Swanson, N., “Application of Vibration Signal Analysis Techniques to Signal Monitoring,” Conference on Friction and Wear Engineering, 1980.

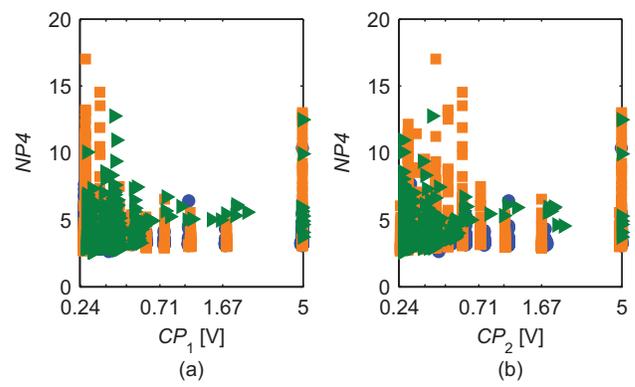


Fig. 19: Dependence of CIs on the damage level, measured by  $CP_1$  and  $CP_2$  for four representative propagations. NP4 as a function of  $CP_1 / CP_2$  ○:108, □: 114, ▷: 116

<sup>8</sup>Haile, M., A., G., and Dykas, B., “Blind Sources Separation and Blind Equation,” *Proceedings of the American Helicopter Society 68th Annual Forum*, Ft. Worth, TX, 2012.

<sup>9</sup>Choy, F. K., Polyshchuk, V., Zakrajsek, J. J., Handschuh, R. F., and Townsend, D. P., “Analysis of the effects of surface pitting and wear on the vibration of a gear transmission system,” *Tribology International*, Vol. 29, (1), 1996, pp. 77–83.

<sup>10</sup>Lebold, M., McClintic, K., Cambell, R., Byington, C., and Maynard, K., “Review of Vibration Analysis Methods for Gearbox Diagnostics and Prognostics,” Proceedings of the 54th Meeting of the Society for Machinery Failure Prevention Technology, 2000.

<sup>11</sup>Kramberger, J., Sraml, M., Glodez, S., Flasker, J., and Potrc, I., “Computational model for the analysis of bending fatigue in gears,” *Computers and Structures*, Vol. 82, (23-26), 2003, pp. 2261–2269.

<sup>12</sup>Kramberger, J., Sraml, M., Potrc, I., and Flasker, J., “Numerical calculation of bending fatigue life of thin-rim spur gears,” *Engineering Fracture Mechanics*, Vol. 71, (4-6), 2004, pp. 647–656.

<sup>13</sup>Podrug, S., Glodez, S., and Jelaska, D., “Numerical Modelling of Crack Growth in a Gear Tooth Root,” *Journal of Mechanical Engineering*, 2011.

<sup>14</sup>Nenadic, N., Wodenscheck, J., Thurston, M., and Lewicki, D., “Seeding Cracks Using a Fatigue Tester for Accelerated Gear Tooth Breaking Rotating Machinery,” *Structural Health Monitoring, Shock and Vibration*, Vol. 5, 2011.

<sup>15</sup>Stringer, D. B., LaBerge, K. E., Burdick, C. J., and Fields, B. A., “Natural Fatigue Crack Initiation and Detection in High Quality Spur Gears,” *Proceedings of the American Helicopter Society 68th Annual Forum*, Ft. Worth, TX, 2012.

<sup>16</sup>Decker, H. and Lewicki, D., “Spiral Bevel Pinion Crack Detection in a Helicopter Gearbox,” Proceedings of the 59th American Helicopter Society Annual Forum, 2003.

<sup>17</sup>Lee, C. and Lee, H., “Gear Fatigue Crack Prognosis using Embedded Model, Gear Dynamic Model and Fracture Mechanics,” *Mechanical Systems and Signal Processing*, Vol. 19, 2005, pp. 836–846.

<sup>18</sup>Lewicki, D. and Ballarini, R., “Rim Thickness Effects on Gear Crack Propagation Life,” *International Journal of Fracture*, Vol. 87(1), Dec. 1997, pp. 59–86.

<sup>19</sup>Townsend, D. and Shimski, J., “Evaluation of Advanced Lubricants for Aircraft Applications Using Gear Surface Fatigue Tests,” , 1991.

<sup>20</sup>Duda, R. O., Hart, P. E., and Stork, D. G., *Pattern classification*, Wiley, New York, second edition, 2001.

<sup>21</sup>Kay, S. M., *Fundamentals of statistical signal processing*, Prentice Hall signal processing series, Prentice-Hall PTR, Englewood Cliffs, N.J., 1993.

<sup>22</sup>Fawcett, T., “An introduction to ROC analysis,” *Pattern recognition letters*, Vol. 27, (8), 2006, pp. 861–874.

<sup>23</sup>Bradley, A. P., “The use of the area under the ROC curve in the evaluation of machine learning algorithms,” *Pattern Recognition*, Vol. 30, (7), 1997, pp. 1145–1159.

<sup>24</sup>Neter, J., Kutner, M., Wasserman, W., and Nachtsheim, C., *Applied Linear Statistical Models*, 1996.

<sup>25</sup>Hosmer, D. W. and Lemeshow, S., *Applied Logistic Regression*, John Wiley and Sons, Inc., 2000.

<sup>26</sup>Elith, J., Leathwick, J. R., and Hastie, T., “A working guide to boosted regression trees,” *Journal of Animal Ecology*, Vol. 77, (4), 2008, pp. 802–813.