Draft
Minutes
ASC OP1 ASC OP/TF 2, Performance Based Optical Imperfections Task Force Draft Standard Meeting
Sunday August 26, 2007, 8:30 a.m. — 12 Noon

Present	Attendees (13 of 17 Entities)	
☒	**Committee Members**	**Representing**
☒	David Aikens	Zygo Corporation
☒	Gordon Boultbee	JDSU Corporation
☐	Andrei Brunfeld	Xyrtex
☐	Benjamin Catching (Alternate)	JDSU Corporation
☒	Walter Czajkowski	APOMA (Edmund Optics)
☒	Frank Dombrowski (by phone)	Gage-Line Technology, Inc.
☒	Marla Dowell	IEEE/LEOS (NIST)
☒	Lincoln Endelman	SPIE, (Endelman Enterprises)
☐	Charles Gaugh	Davidson Optronics, Inc.
☒	John Hamilton (by phone)	Northrop Grumman
☒	Hal Johnson (by phone)	Harold Johnson Optical Lab
☐	Rudolf Hartman	Retired
☒	Alan Krisiloff	Triptar Lens Co., Inc.
☒	Jonathan McGuire (Alternate)	Northrop Grumman Laser Systems
☐	Michael Morrill	Lockheed Martin Space Systems Company
☒	Bruce Netherton	Lockheed Martin Coherent Technologies
☐	Sam Richman (Alternate)	Research Electro-Optics, Inc.
☒	William Royall (by phone)	Eastman Kodak Company, Retired
☐	Trey Turner	Research Electro-Optics, Inc.
☐	Steve VanKerkhove	Corning Tropel
☒	Ray Williamson	Ray Williamson Consulting
	Observers (1)	
☒	Gene Kohlenberg	OEOSC

Auditor's Summary of Meeting

Two representatives from Northrop Grumman presented the results from their examination of the optical surface imperfection readings recorded by a group of optical inspectors. The conclusion is that the current qualitative test is less accurate than the optics industry assumes. There was considerable discussion concerning how the conclusion affects the surface imperfection evaluation that has been traditionally used to grade optical components.

New laser and micro-optics components require the elimination of surface scratches and digs that traditionally were acceptable for consumer and military products. The current test does not include reference samples small enough to accommodate these new products. There was concern that the revision of this standard would be delayed significantly if it were not released until a decision concerning micro-optics and laser components could be completed. The current draft will be finished at the next meeting so that it can be balloted and then the Task Force will turn to the micro-optics issue.

Eastman Kodak Company had marketed a molded imperfection reference for evaluation of optical surfaces. The business was sold and the new vendor uses a different method for producing the reference. There was voiced concern that the new reference did not yield as repeatable results as was formerly obtained using the Kodak product, and yet the product is sold under the same product number. The companies selling the reference were urged to change the part number so that optics manufactures would not assume that it is the same product.

The task force agree to meet again in San Jose, CA on January 20, 2008.

1 Welcome and Introductions

G. Boultbee opened the meeting at 9:00 a.m. Since there were several new persons attending the meeting, each one was introduced.

2 Adoption of Agenda

G. Boultbee noticed that C. Gaugh was not present to discuss his item. D. Aikens asked if the status of the Scratch and Dig paddle could be discussed. G. Boultbee asked for a motion to approve the published agenda with addition of the item concerning Scratch and Dig Paddle. A. Krisiloff made the motion and M. Dowell provided the second. The motion carried unanimously.

3 Approval of the Monday, May 14, 2007 ASC OP/SC 1, BSR/OEOSC-OP1.002, Optics and Electro-Optical unanimously. Instruments – Optical Elements and Assemblies — Appearance Imperfections Draft Review Minutes

The minutes had been posted on the web site. The Task Force Leader asked if there were any additions or corrections to the minutes. G. Boultbee offered two changes to the minutes. The first was in section 4, where the word “approve” was changed to “improve.” The second change was in section 5: the sentence was changed to say, “G. Boultbee called for a break at 10:30 a.m. and then he suggested that the task force address the run of the mill dimensional inspection, then if there is time, return to T. Turner's method.” J. Hamilton moved that the minutes be approved and M. Dowell seconded the motion. The motion carried unanimously.

4 Scratch and Dig Round Robin

C. Gaugh was not available for this discussion. He had offered to supply artifacts that NIST could use to develop a round robin test.

5 Northrop Grumman Laser Systems Gage R&R

J. Hamilton stated that the accuracy of imperfection evaluation has become more difficult as optical requirements become more stringent. Optical components for precision systems are becoming much more expensive, and disagreements concerning surface imperfections are more critical because of potential expensive scrap losses. Even though the military imperfection specifications stipulate that they are for appearance purposes only, in practice the specifications are treated as if the imperfections are measured values. Northrop Grumman has had suppliers' engineers spend as much as a week with the company trying to correlate what a “60” scratch is on a manufactured component. J. Hamilton said that this is the continuation of a fifty-year debate which is quantified by the various revisions to the military artifact drawings over the years.

Northrop Grumman decided to conduct an in-house study of the problem by treating the inspection methodology as an attribute gage study rather than an optical problem. The earlier NIST study by Mat Young treated the evaluation technique as an optical problem. Lionel Baker's work also came from an optical perspective.

Because there was no obvious source of visual evaluation discrepancies by a group of students who were being trained to inspect optical components, Northrop Grumman looked at these inspectors' eye sight, skill level, and experience. They evaluated the inspection environment and examined the military's and Kodak's reference samples. They then ordered certified sets of reference samples from Brysen, cut them in half, and used one half as the test sample. This ensured that “apples” were being compared to “apples.” A similar test was designed for Kodak paddles to evaluate those reference samples. All three inspection environments described in MIL-O-13830 were set up. (The incandescent and fluorescent inspection environments produce almost the same results.) The Brysen references were evaluated outside of the cases that the military encloses them in to see if repeatability could be improved.

The intent of the evaluation was to determine how often a failure was missed and how often an acceptable part was failed. They wanted to determine the reason for reading errors. For the test five trained optical inspectors and five optical technicians not trained for inspection were selected so that the skill level question could be answered. The eye-sight of each inspector was tested. Testing done by J. Hamilton's group covered a larger inspector population than the testing done by J. McGuire's group.

At this point J. McGuire continued the presentation. His testing had to be piggy-backed on other projects. That minimized the amount of time that could be devoted to the evaluation. He said that the data he was showing at this meeting, and the data that J. Hamilton would be presenting later show the need for improvement in the inspection system.

J. McGuire used six inspectors ─ three from the receiving inspection department and three quality engineers from the production floor. All six have been making these types of inspections for a number of years. J. McGuire used the same Brysen samples that J. Hamilton had prepared. Samples were labeled A, B, C… so that the inspectors had no idea which scratch number each sample represented. There were ten samples, two each of 10, 20, 40 ,60, and 80. Two samples had a mirror coating. J. McGuire performed an SCM analysis to confirm that the samples were not compromised when they were cut in half. Where as J. Hamilton used both military references and Kodak paddles as the comparators, J. McGuire used a

new and old set of military references (encased in a wooden box). There were 240 test data points, half with the new reference and half with the old reference.

The small sample size leads to large confidence intervals. In order to minimize errors caused by the viewing system, the same black box with curtains, 40 watt light with opal glass and black matte bars was used. There were a few lint particles inside the reference cases, but nothing that would compromise the test. The test was explained to the inspectors so that the unusual event would not confuse them. The test samples were cleaned using standard techniques. J. Hamilton said that the same cleaning procedure was followed for his test. J. McGuire said that each inspector evaluated each test sample four times: twice with the old references and twice with the new references. R. Williamson suggested that for any future testing that the samples be relabeled between multiple readings. J. McGuire said that he did the test himself, and could not remember what numbers he had assigned to each sample when he did a repeat evaluation. The test procedures used by both Northrop Grumman divisions presented the samples to each inspector in a random fashion so that there would not be a tenancy for the inspector to remember the evaluations.

J. McGuire presented charts of the test results showing how each inspector agreed with him or herself as well as agreement with the reference standard. The inspectors had a higher level of agreement with their own results than they did with the reference samples. The data was presented as the percentage agreement, i.e., if an inspector chose the same scratch number three out of five times, then his score was 60%. Bars encompassing the plotted reading illustrated the 95% confidence interval. A. Krisiloff suggested that the four readings of a sample could be plotted on a graph with each reading occupying a spot on the x-axis while the four readings would be plotted on the y-axis. Then the standard deviation of those four readings could be computed to see if there is a tighter correlation.

J. Hamilton said that his data also indicate that inspectors are more consistent with their own readings than they are with the actual reading.

R. Williamson asked if there was a good correlation of a individual inspector across the two samples that had the same scratch number. J. Hamilton said that that relationship has not yet been evaluated.

J. McGuire said that the agreement among the inspectors was only 10%. Inspector agreement with the "correct" reading was 32%. 55% of the time the inspectors report the scratch to be larger than the accepted value. 13% of the time a smaller reading is reported. A number 10 scratch was always marked as larger than a 10. This was important to J. McGuire because a large portion of his laser optics have a specification of 10. Practically speaking, either an optical element was either perfect or it had a scratch that was larger than 10; there were no 10s. On a few occasions inspectors actually read 20s as if they were 10s.

The inspectors were told that the samples came from reference sets, so the inspectors knew that there were no samples greater than 80. Therefore, there were no rejections for 80s. A. Krisiloff observed that the data seemed to show that small scratches are over estimated and large scratches are underestimated. D. Aikens countered that the 40 sample is fairly accurately identified. There seems to be a psychological bias towards rejection. Since the test samples and the references came from the same pieces, there should be no bias. M. Dowell asked if a 40 scratch is the most frequently encountered imperfection so that the inspectors would be more familiar with that size scratch. J. McGuire said that in his organization the optical components are either very high quality or fairly loose, so examples on both ends of the spectrum should be more frequently encountered.

D. Aikens observed that the 10 sample was evaluated as high as 80; the 20 was consistently identified as a 40; the 40 was consistently identified as a 40; the 60 almost always was identified as an 80. The samples were no longer in a case, while the references were still in their case. Were the references harder to see?

J. Hamilton said that he saw the same type of distribution for his test using a larger number of inspectors. In order to answer the question about visibility of the references in the cases, he bought new references without the cases, and got essentially the same results.

A. Krisiloff surmised that the results show that the test is working as designed, because if there is doubt the inspector is instructed to choose the next higher number. In order to confirm this 5, 15, e.t.c, scratches should be created to test if the binning is working properly. The boundary conditions are being tested rather than the center of the bin. R. Williamson said that if the protocol said, "which scratch does the sample most closely match," then one would get a very different result.

J. Hamilton said that they modeled the illumination conditions and scratch geometry in Z-Max. When the scratch is evaluated using Z-Max the same problem occurs. The problem comes from the mono-width variable depth geometry for all of the scratches in the current military specification for the reference samples. When the range of scratches are evaluated by Z-Max the different scratch numbers can not be recognized. D. Aikens reported that the reference samples that Brysen sells are not unit-width, variable depth. The unit-width, variable-depth artifacts are limit samples that are used by Brysen in the manufacturing process. J. Hamilton countered that when he does electron microscope evaluation of the references, he finds that they are 7 μ – 10 μ in width.

D. Aikens said that when he does optical evaluations of reference samples, he sees considerable variation in the widths, although a 10 and a 20 are close to the same width. The 40, 60 and 80 are dramatically different in width.

J. McGuire said that he did similar testing on the samples in his possession, and found that the widths were in the 5 µ – 10 µ range with depths approximately 1 µ. When measured he determined that a 10 was less than a 60, but a 60 was wider than an 80. He could not say that there was a correlation between the scratch width and its apparent visibility. D. Aikens said that theoretically there did not have to be a correlation because the samples are binned by their visibility, not their physical characteristics.

A. Krisiloff asked if the inspectors could have observed a glint at the edge where the Brysen sample was cut to come to the conclusion that a 10 appears to be an 80. J. Hamilton said that the inspectors were instructed to ignore the test samples near the cut edge. J. McGuire did not remember if he gave the same instruction. Since results from the two Northrop Grumman facilities were similar, J. McGuire assumed that this was not an issue.

A. Krisiloff then asked J. Hamilton and J. McGuire how they explain the situation where 10s were observed as 80s. J. Hamilton attributed the observations to the fact that the widths of the 10s and 80s were similar, and the human eye cannot distinguish them as the SIRA instrument can.

D. Aikens made an observation about the data recorded for the older military references. The results for a 10 sample were noticeably different when the older military references were used. A 10 was never identified as a 10, five times it was reported as a 20, six times it was identified as a 40, six to eight times it was thought to be a 60, and only once called an 80. M. Dowell surmised that the reference that is used frequently may have some damage that the older military reference that sits on the shelf does not exhibit. The old standard would possibly provide a more accurate result. A. Krisiloff suggested that the profiles of the two sets of references might actually be different. J. Hamilton said that the military had difficulty making the original reference samples, and the profiles of the scratches is not known. D. Aikens reminded the group that Brysen reported at a previous meeting that the process for scribing the scratches has not been changed in 40 years. The scratches are made with differing pressure and then the resulting samples are visually sorted and binned.

G. Boultbee suggested that the samples should be given to Ari Siletz of CCDMETRIX so that they could be evaluated on the new CCDMETRIX instrument.

J. Hamilton reminded the group that the military only allows vendors to use the current edition of reference samples. He was able to keep the old set by agreeing to not use them in production.

D. Aikens asked J. Hamilton if the data generated using the Kodak paddle showed the same bias for rejection. J. Hamilton did not have the data in front of him, but he thought that the Kodak paddle results were marginally better. D. Aikens said that this observation would support M. Dowell's hypothesis that inspectors are more careful when using a less familiar reference to grade samples. They have to look at the reference more frequently.

J. Hamilton interjected that this system does not work as a gage device. J. McGuire reinforced that statement by stating that the old reference results were correct only one-third of the time.

A. Krisiloff said that the sample data looked as if there was the expected bias toward rejection since, in general, samples were binned one number larger than the actual reading. He asked J. Hamilton how that observation is wrong. J. Hamilton replied that for $10,000 optics, a company cannot afford an evaluation system that has a 55% chance of rejecting good product. A. Krisiloff said that the qualitative evaluation, which has large error bars for observations, should be revised to compensate for the error on the high side. Instead of presenting a 20 reference to the inspector, an 18 should be the reference. J. McGuire said that there is variability in the Brysen reference samples, so that a 20 reference may be an 18. A. Krisiloff said that that fact indicts the qualitative test even more. J. Hamilton replied that the problem with the current test is that it does not take into consideration the tolerance of the reference sample, and it does not account for the variability introduced by the inspector.

D. Aikens observed that if one has an expensive optical component, then the current qualitative standard should not be used. The specification should be for scratch widths, and the part should be evaluated under a microscope. J. Hamilton said that in practice his group uses the Kodak paddle as the first sorting test. He can then distinguish the questionable samples from the good ones. The scratch width would be measured for the questionable samples.

D. Aikens observed that there is a need for a quick visual check; however, this system is even questionable for fast, inexpensive testing. A. Krisiloff observed that the current system is valid except for the 10 case. J. McGuire countered that if all of the observations that were one bin high were removed from the sample population, there were still 31 data samples that were incorrect. 25% of the time the observation is more than one bin above the correct value. D. Aikens added that this is a large error bar for commercial transactions. A. Krisiloff said that skipping to a second higher bin is unacceptable. M. Dowell countered that this result is only true for the new artifact standard; it is not true for the older artifact standard. D. Aikens said that the 10 specification is meaningless because it is often observed as higher values. G. Boultbee said that this discussion was mirroring the discussions at earlier meetings where the laser and micro-optic applications were determined to not be adequately represented by this standard.

G. Boultbee closed the discussion by saying that he was hearing two groups of comments. There were questions for J. McGuire concerning data that he had not presented, and there were questions concerning what should be done next with

regard to additional studies of the test methodology, and consideration of the acceptability of this standard. He proposed that the Task Force document the questions that it has

- for J. McGuire about other ways to look at his data, and
- concerning future studies to resolve issues with the standard.

J. McGuire said that J. Hamilton was planning a report that documented the tests and test results. D. Aikens said that he would like to see the data that J. McGuire presented released to the optics community. J. Hamilton said that his organization planned an in-house symposium for November that would include this topic. He said that he always presumed that the results would be reported in a peer-reviewed publication such as one of the SPIE journals.

J. McGuire said that under the current system vendors double the charge of their components because they expect the customer to reject half of the components supplied to them.

L. Endelman said that if the methodology and test equipment are not changed, then no amount of additional evaluation of the test system will change the end result. Everyone agreed.

G. Boultbee said that the Task Force needs to anticipate the questions that may be asked the next time the subject is presented. J. McGuire said that he would like to extended the testing to include 0s and imperfections in excess of 80, and do the evaluation in a more blind fashion. J. Hamilton's testing did address issues such as cased reference artifacts vs. uncased artifacts and the Kodak paddle vs. Brysen reference artifacts. Those who are experienced with surface imperfection testing new that the current method is qualitative and flawed, but these tests indicate that it is more flawed than previously assumed.

D. Aikens asked if the Task Force was prepared for him to include the results of this report in the upcoming Scratch and Dig class in Boston? J. McGuire said that Northrop Grumman considers what he presented at this meeting is now part of the public domain. M. Dowell said that the students in the class are a different audience than those who are on this Task Force. The students are looking for answers and this report creates more questions. D. Aikens agreed to not use the data in the class. G. Boultbee said that D. Aikens could caution the students to be aware of number 10 scratch evaluations. J. Hamilton urged D. Aikens to wait until after the Northrop Grumman report is published (in the November time frame) because he could better describe the problem.

J. McGuire said that the testing scope could be expanded to include more operators, operators from different companies, 0 samples, samples in excess of 80 to expand the imperfection range.

B. Netherton said that he has seen similar testing results for the past 27 years, and would prefer that Northrop Grumman and Lockheed Martin would work together to conceptualize alternatives to the current test procedure. D. Aikens added that the ISO 10110 method should be checked to see if it has similar attributes as the current military method or if it is more robust. He also noted that this Task Force has spent considerable time expanding the current standard to include physical measurement methods; do they work? Perhaps effort should be directed toward the measurement method to see if it is more reliable.

L. Endelman summarized the discussion by saying that the current qualitative testing system is better than nothing, but the optics community does not know how much better than nothing.

At this point G. Boultbee proposed a 10 minute break.

6 Review of Revised OP1.002

G. Boultbee reviewed the status of the draft document by stating that the agreed upon changes to clause 3.7 and Annex B were incorporated in the 8/26/07 version. Annex C still had to be addressed. The Task Force also must address T. Turner's RES proposal. G. Boultbee proposed that the draft of the standard with the incorporation of Annex C can be used for traditional, conventionally-sized optics. What T. Turner proposes for laser/micro-optics applications would be better served by a new standard dealing directly with imperfections that are on the order of 10-5 and tighter surface quality. The current draft would be delayed excessively if the effort to incorporate the 10-5 venue were undertaken.

M. Dowell suggested that a notification of the limitation concerning laser and micro-optics be added to the foreword of the current document. From a laser standpoint, she would prefer to see the 10-5 regime added to a future version of the current standard because it is all about the qualification of optics. The T. Turner proposal would need to have a similar evaluation as Northrop Grumman has done for the current visual system before it were to be incorporated into the standard.

G. Boultbee suggested that the Task Force deal with the wording in Annex C, and then go back to the scope to address any constraints.

D. Aikens said that the Task Force has not addressed functional performance for laser optics. B. Netherton said that in practice he specifies the use of a laser to inspect. He also uses a scatter test, which has no standard. A. Krisiloff said that it sounds as if there should be a scatter standard. D. Aikens said that the Task Force is not in the position to address scatter. M. Dowell asked about an integrating sphere method of measuring scatter for laser rods.

D. Aikens asked if there is a need for a width specification. J. McGuire said that a measured width standard would be of use.

D. Aikens said that if the Task Force is going to open up the lower end of the artifact range, then there could be letters "AA" (2 µ), "AAA" (1 µ). M. Dowell countered that a statement specifying the actual number to be called out. To address this concept, D. Aikens suggested that an "A" followed by a size could be incorporated; an "A2" is a 2 µ scratch. J. McGuire countered that the letter should be "W", for width, to remove confusion. G. Boultbee said that at some point, say less than 5µ, microscope viewing is required so that imperfections can be seen.

Because the meeting time limit was approaching, G. Boultbee agreed to draft a proposal for specifying imperfections that are smaller than the existing range. The Task Force will look at the Foreword, which needs to be updated. D. Aikens suggested that we give a notation and caution the user about viable measurement in this regime. R. Williamson said he preferred retaining the Foreword of the first addition and adding a second edition Foreword for the new version. D. Aikens volunteered to write the text of the Foreword for the second edition down to the paragraph that starts, "Suggestions for improvement of this standard are welcome." He will give the wording to the Secretary who will incorporate it into the document.

R. Williamson identified a typographical error in clause 3.1.6, which should read 0.05 – 0.25.

7 Scratch and Dig Paddle Status

W. Royall said that the original Kodak paddle is out of production. The new Edmund paddles have curved scratches. He did not have time to evaluate them. D. Aikens has looked at new the paddles, and he did not see as good a correlation between the old and new. They are manufactured using an entirely different technology than Kodak used. Edmund and Thor labs sell the paddles under the same part number that was formerly used for the Kodak paddle. D. Aikens urged that the part numbers be changed. W. Czajkowski said that he would look into this issue at Edmund. D. Aikens will also investigate.

W. Czajkowski suggested that ultimately reference samples would best be created by replication rather than scribing.

8 Time and Place of next TF 2 Meeting

The Task Force agreed to meet next in San Jose,CA, January 20, 2008, 8:30 a.m. – noon.

9 Adjourn

W. Czajkowski moved that the meeting be adjourned; M. Dowell seconded the motion, which carried unanimously. The meeting adjourned at 12:20 p.m.