Measurement of Decarburization of Heat Treated Steel Surfaces

George F. Vander Voort
Consultant – Struers Inc., Wadsworth, Illinois USA

Decarburization is a significant problem in the heat treatment of steels as decarburization is detrimental to wear life and fatigue life of components. This presentation concentrates on its measurement by four procedures. Test results from an interlaboratory test of steels with spheroidize annealed, lamellar pearlite (normalized) and quenched and tempered martensitic microstructures were evaluated in an interlaboratory program. Opposite end portions of each specimen were turned incrementally and the turnings were analyzed for percent carbon. Then, the central part was sectioned, mounted and polished producing six mounts of each type. These specimens were rated by about a dozen different people in two ways. First, a “qualitative” evaluation was made where the metallographer, using a light optical microscope (LOM), scanned around the bar periphery looking for the greatest maximum affected depth (MAD), a qualitative assessment. Then, each operator made 25 randomly selected measurements of the MAD going around the bar periphery, a quantitative assessment. Microindentation hardness traverses were also made on each specimen to define the maximum affected hardness depth. The talk will compare the LOM qualitative with the LOM quantitative assessment of the MAD and also compare these results to the MAD determined by chemical analysis of the incremental turnings and the MAD determined by microindentation hardness testing as a function of matrix microstructure and alloy carbon content.

Qualitative ratings are often performed by mill metallographers. The rater scans the edge of the specimen using a LOM looking for areas that appear to be typical of either an apparent average or the deepest loss of carbon. Qualitative measurements tend to be subjective and may even be deliberately biased. For quantitative measurements, the rater makes a reasonably large number of measurements at randomly chosen locations spaced around the periphery of the cross section. The larger the cross section, the greater the number of measurements. Here, the rater does not try to select regions that appear to be “typical” or the “worst.” Quantitative measurements are more reproducible than qualitative measurements and should be unbiased if the measurement location is selected randomly. In most cases, a reticule is used to make the measurement. Alternatively, many software programs for capturing images allow the operator to make point-to-point distance measurements. These systems must be properly calibrated. Automated image analysis measurements may also be performed but recognition of the location where the carbon content becomes constant may be a challenge depending upon the microstructure, etchant used and the carbon content of the unaffected matrix.

Figure 1 shows an example of the results for a spheroidize annealed specimen of W1 tool steel. The MAD based on chemical analysis of incremental turnings was 0.64 mm while a microhardness traverse revealed an MAD of 0.51 mm (79.7% Lower). The average of the qualitative estimates of MAD was 0.406 mm (63.4% of the MAD according to the chemical analysis) while the average for the mean of the 25 randomly chosen fields by 12 raters was 0.433 mm (67.7% of the MAD based on chemical analysis of the incremental turnings).
(a) Spheroidized Carbides in ferrite (4% Picral)
(b) Carbon analysis of incremental turnings
(c) Unusual microhardness traverse
(d) Mean Qualitative MAD = 0.406 mm
Mean Quantitative MAD = 0.433 mm

Figure 1. Example of the results of the study of decarburization using (b) incremental turnings from the bar periphery followed by traditional chemical analysis for carbon; (c) a 200 gf Knoop microindentation hardness traverse which revealed that the non-spheroidized decarburized zone was actually harder than the matrix; and (d) the mean results of qualitative LOM measurements by 11 of the 12 raters picking out what they believed was the greatest depth of decarburization which was lower than the mean of 25 randomly chosen locations measured by the same 12 raters.