Automated analysis of large (>4Gb) spectral images with efficient out-of-core-RAM algorithms

P.G. Kotula and M.R. Keenan

Sandia National Laboratories, PO Box 5800, Albuquerque, NM 87185-0886

Spectral imaging combined with automated multivariate statistical analysis (MSA) is a powerful characterization tool [1]. The primary limitation of the technique at this point is the size of the spectral image data sets. Small spectral images typically comprise 64 by 64 pixels (4096 spectra) or 128 by 128 pixels (16384 spectra), each with 1000 or more energy channels. Such data sets easily fit into available RAM on standard personal computers. Under most circumstances 0.5-1µm/pixel or finer resolution is needed to effectively sample a microstructure in the SEM, implying that only small regions of a microstructure could be comprehensively analyzed (50-100µm). A prime goal of utilizing spectral images in routine and comprehensive microanalysis is therefore to surpass 1000 by 1000-pixel (10^6 spectra) spectral images. This would allow full analysis of 500-1000µm^2 regions simultaneously. In the present paper we describe the extension of previous work [1] to larger data sets, in excess of 1 million spectra or 1 billion individual data points, where the data can be noisy.

One of the limitations of current PCs imposed by the Windows ^{TM} operating system is the amount addressable (useful) core-RAM memory, which is 2Gbytes. This is the space in memory that would be available to an automated spectral image analysis calculation. One strategy for overcoming this core-memory limitation is to compress the spectral image data on the fly prior to subjecting it to MSA. In this way then, the size (in RAM) of the problem never exceeds available resources. Compression can be done both spatially and spectrally. Simple binning of groups of pixels or channels is one way to accomplish compression and is the method that will be discussed here. One benefit of compression is an improvement in signal to noise (S/N). If we start with a raw spectral image that is 1024x1024pixels x 1024 energy channels we can compress it by a factor of 64 (spatially) for analysis to 128x128 pixels x 1024 energy channels and realize a S/N improvement of ~√64=8. Since the overall goal of the analysis is to maintain the original high resolution of the raw data, the solution calculated automatically on the compressed data set is fit (via least squares) to the raw data at its original resolution. In this way then, we gain the benefit of improved S/N for analysis of compressed data without loosing any image resolution. An early form of this out-of-core-RAM analysis has successfully been applied to a 512x512pixels x 1024 energy-channel data set from a geologic material [2]. In the current paper we examine a 1024x1024pixels (0.5µm/pixel) by 1024 energy-channel data set from a cutting tool where each raw pixel (spectrum) has on average 5 counts. An eigenanalysis for various levels of compression in Figure 1 shows that compression does indeed improve detectability of minor chemical components (improvement in S/N). Additionally, the computation time of compressed data sets is considerably shorter-300 sec for 1024x1024 and 60 sec for 128x128 compression. Figure 2 shows the two carbide components (also phases in this case). Even with such noisy raw data spectral correlations are easily seen-WC has W-M and –L lines and Co, TaC has Ta-M and –L lines and Ti. Not shown are Fe (substrate), Cu and Ag (bonding layer).

[1]   P.G. Kotula et al., *Microsc. Microanal.* **9** [1](2003) 1-17.
[2]   P. G. Kotula, et al. *Microsc. Microanal.8 Suppl.2 Proceedings* (2002) 1554CD.
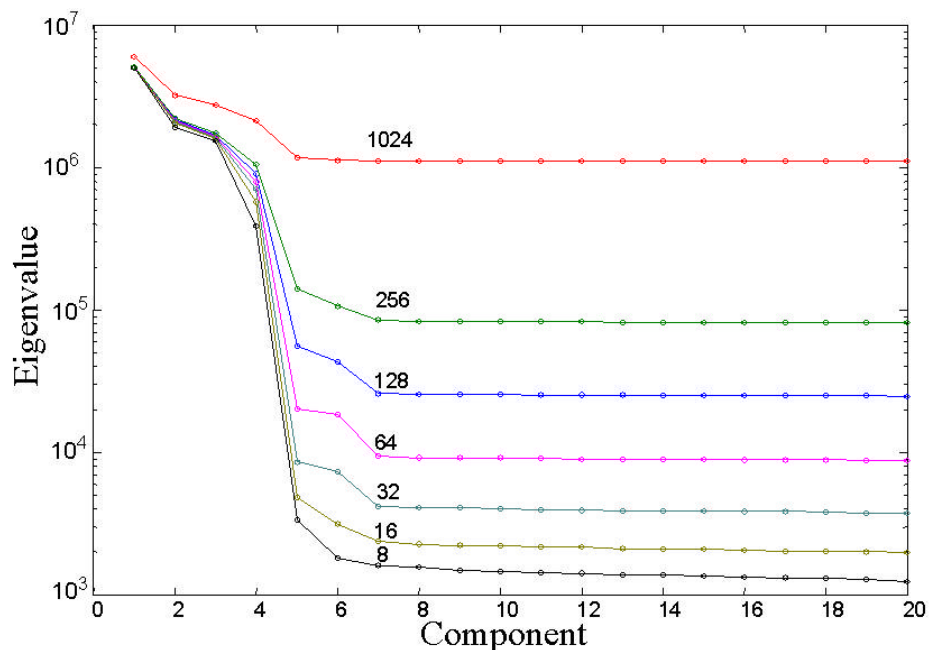[2]   Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin

FIG. 1. Eigenvalues versus component for different levels of compression. Number over each curve is the square root of the number of pixels in the compressed analysis. The contribution of the components (chemical signals) to the overall solution is seen relative to the flat baseline for each curve (due to noise). All data was weighted for Poisson statistics.
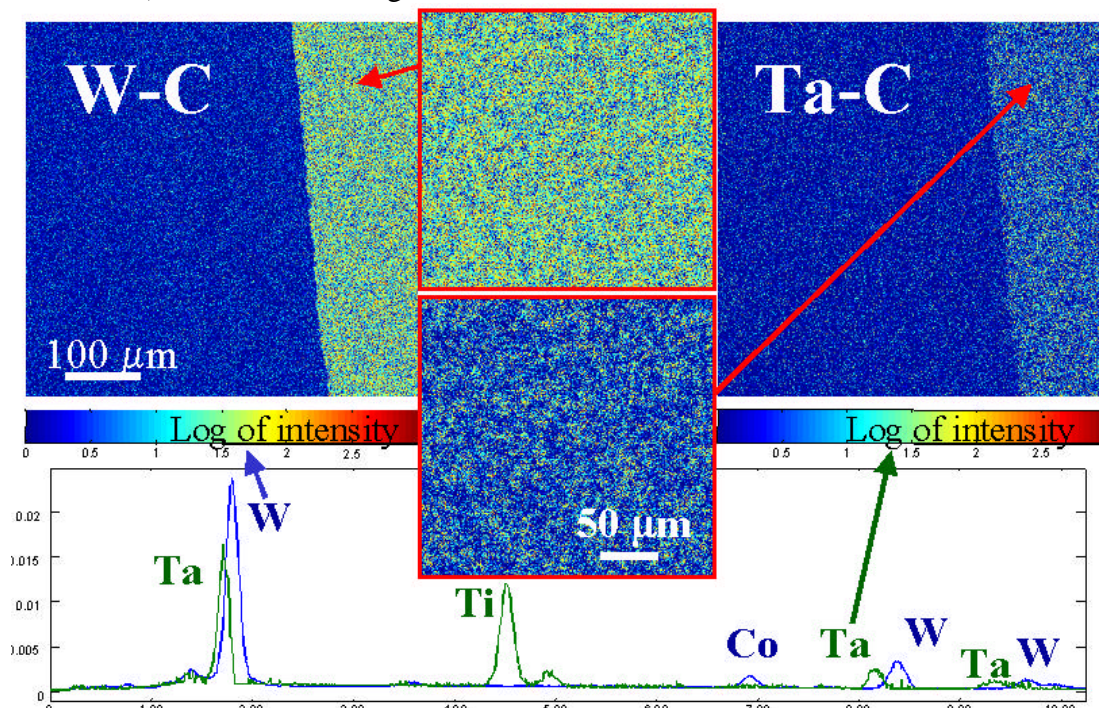

FIG. 2. Pure component images and spectra from the two carbide components (phases) with inset enlargements. W-C (blue/darker spectrum) contains Co and Ta-C (green/lighter spectrum) contains Ti.