Anal. Chem. **1999,** *71,* 5157−5164

Identification of Solid Materials by Correlation Analysis Using a Microscopic Laser-Induced Plasma Spectrometer

I. B. Gornushkin, B. W. Smith, H. Nasajpour, and J. D. Winefordner*

Department of Chemistry, P.O. Box 117200, University of Florida, Gainesville, Florida 32611

The primary goal of this work was the instant identification of solid materials on the basis of spectral libraries stored in a computer using laser-induced plasma spectroscopy. The libraries were obtained prior to analysis and consisted of representative spectra from different groups of materials to be analyzed. Special attention was paid to identification of samples with very similar chemical composition, such as certain series of stainless steel and cast iron standards. Both linear and rank correlation methods were applied. Rank correlation proved to be more reliable, yielding probability of correct identification close to unity for almost all studied samples. This technique should have applications in the metallurgical, mineralogical, and semiconductor industries and in medical and forensic sciences.

Most applications of laser-induced plasma spectrometry (LIPS) have been aimed toward quantitative elemental characterization of materials.[1−2] Indeed, for the analytical chemist, LIPS possesses such attractive features as speed and the capability of analyzing any sort of material regardless of its aggregate state. Speed is a result of a significant reduction in the time usually needed for a sample preparation. LIPS allows on-line measurements of unprepared materials in real time. The capability of analyzing any material relates to the extremely high energy of a tightly focused laser beam; the high temperature of the laser plasma vaporizes and excites samples of any nature. However, for quantitative measurements, several limitations do exist. Most important, the reproducibility is relatively poor (10% RSD is typical) and severe matrix effects are common. Reliable quantitative results can only be achieved under well-controlled conditions with carefully selected standards. A vast literature on the subject already exists and continues to grow.[3]

In this work, we examine LIPS from a slightly different perspective. Realizing the great importance of this technique for quantitative spectrochemical analysis, we note that not much attention has been paid so far to the potential of LIPS as a powerful tool for rapid, reliable qualitative analysis. In this sense, we do not seek a detailed chemical composition but rather that the material be instantly identified using its unique laser breakdown spectral "fingerprint." This "fingerprint" is a LIP spectrum chosen from many others by a robust correlation procedure. This, in turn, implies that corresponding databases exist that contain LIP spectra of a variety of compounds, such as the libraries that exist for IR, X-ray, or mass spectrometry. The necessity of using such spectral libraries, in addition to well-known tables of spectral lines, stems, first, from the nature of LIPS emission spectra, which are usually dominated by ionic lines, and, second, from the complex matrix dependence of LIPS emission spectra, which often inhibits the direct relation between the elemental concentrations in a sample and the intensities of spectral lines.

Assuming the availability of LIPS libraries, then we need to know which correlation technique is most suitable for compound identification. Among different correlation methods, one can consider relatively simple linear and nonparametric correlation procedures, Fourier analysis,[4] and neural network[5] and wavelet approaches.[6] The choice of correlation method is determined by the particular experimental arrangement, the type of data obtained, and time requirements. In our case, where 2000 data points (pixels) represent a sample spectrum and where the libraries were small, not exceeding tens of spectra, we found the use of simple linear and nonparametric correlations to be entirely satisfactory for our objectives.

The goals of this work can therefore be formulated as follows. First, we constructed a very simple, inexpensive, and compact (ideally—portable) microscopic LIP spectrometer for rapid, real-time identification of a variety of solid and particulate materials. Second, we compiled LIPS spectral libraries which served for reliable material identification. Third, we developed a user-friendly software for simple spectrometer operation and rapid data processing. Fourth, we analyzed almost identical materials (cast iron and stainless steel standards) in order to demonstrate the robustness of the proposed methodology.

EXPERIMENTAL SECTION

Our version of a microscopic LIP spectrometer is shown in Figure 1. A compact 17-mJ, 4-ns pulse width, Nd:YAG laser (model

(1) Schechter, I. *Rev. Anal. Chem*. **1997**, *16*, 173−298.
(2) Rusak, D. A.; Castle, B. C.; Smith, B. W.; Winefordner, J. D. *Crit. Rev. Anal. Chem*. **1997**, *27*, 257−290.
(3) Rusak, D. A.; Castle, B. C.; Smith, B. W.; Winefordner, J. D. *Trends Anal. Chem*. **1998**, *17*, 453−461.
(4) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes: the Art of Scientific Computing*; Cambridge University Press: New York, 1986; pp 381−453, 454−497.
(5) Ruisanchez, I.; Potokar, P.; Zupan, J.; Smolej, V. *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 214−220.
(6) Walczak, B.; van den Bogaert, B.; Massart, D. L. *Anal. Chem*. **1996**, *68*, 1742−1747.

10.1021/ac9905524 CCC: $18.00
Published on Web 10/15/1999

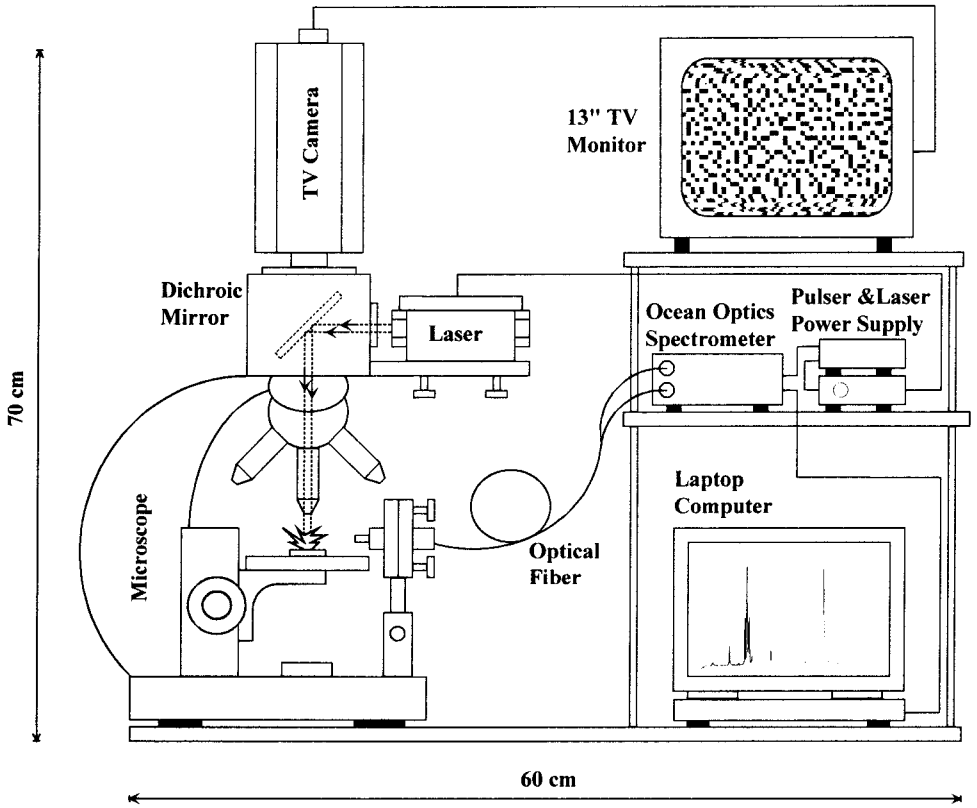

Figure 1. Schematic diagram of the microscopic LIP instrument.

MK-367, Kigre, Inc.) was attached to a modified microscope (Zeiss) and its output beam was aligned with the microscope optical system using a dichroic mirror placed inside the microscope cubic enclosure. The mirror reflected 99% of the laser radiation at 1064 nm and transmitted visible light from the sample surface which was illuminated by a fiber-optic illuminator (Schott KL 1500 not shown in Figure 1). A magnified image of the sample surface was monitored with a TV camera (VCC 3700, Sanyo Electric Inc., Compton, CA) mounted on the top of the microscope and connected to a TV monitor (CT-1020-M, Panasonic). Microscope magnification could be varied with a set of objectives placed in a rotatable mount. The maximum achievable magnification was 2000× with the 40× objective. However, for the present work, we used only the 10× objective (magnification 500×) with a working distance of 5 mm between the objective tip and the sample surface. At higher magnifications, this distance was dangerously short so that the laser spark induced on the surface could damage the objective. With the 10× working objective, the laser was focused to a ~20-μm-diameter spot on the sample surface which corresponded to an irradiance of $\sim 10^{12}$ W/cm^2. This irradiance was sufficient for a stable breakdown even in air. The conic crater left by the laser on a solid surface (e.g., brass or steel) was ~30 μm deep, and mass removal was on the order of 0.1 μg/laser shot. The microscope stage was motorized by means of a stepping motor belted to the *x*-axis stage adjustment knob. The stage (and, hence, the sample) could be translated in a horizontal direction with a minimum speed of 4 μm/s. The laser and the spectrometer were synchronized by a trigger pulse from a homemade compact pulse generator working either in a single-pulse mode or at 0.5-Hz repetition rate. This low repetition rate was necessary to allow operation of the laser without water cooling, which would have complicated the entire spectrometer design.

The radiation from the laser spark was collected with a bifurcated optical fiber (2 separate 600-μm fiber bundles in a "Y"configuration) placed at a distance of 2 cm from the spark region. The fiber head was terminated with a collection lens (5-mm diameter, *f*/2) and mounted in a precision adjustable mount attached to the microscope platform. One end of the bifurcated optical cable was connected to the high-resolution and the other to the low-resolution channel of an Ocean Optics minispectrometer. The spectrometer used in the experiment was a dual-channel high-sensitivity UV−visible fiber-optic spectrometer (SD2000, Ocean Optics, Inc., Dunedin, FL) having the following specifications: channel one, 230−310-nm spectral range, 3600-mm^{-1} holographic grating, 25-μm slit, 0.16-nm spectral resolution, 2048-pixel linear CCD array; channel two, 200−850-nm spectral range, 600-mm^{-1} grating blazed at 400 nm, 25-μm slit, 1.3-nm spectral resolution, 2048-pixel linear CCD array. The spectrometer was driven from a laptop computer (Travel Pro supplied with a Pentium II processor at 266 MHz, AMS Tech) via a DAQCard-700 interface (National Instruments).

Solid samples (steel, brass, glass, cast iron, etc.) did not require any preparation and were used in the form of small chunks or solid blocks. While the stage was moved, the sample surface was kept in focus by adjusting the microscope objective.

LIBS Libraries. Several LIPS libraries were recorded and grouped depending on types of compounds to be analyzed. The library spectra for solids were averaged by taking spectra from 50 different spots on solid surfaces. Spectra were recorded using

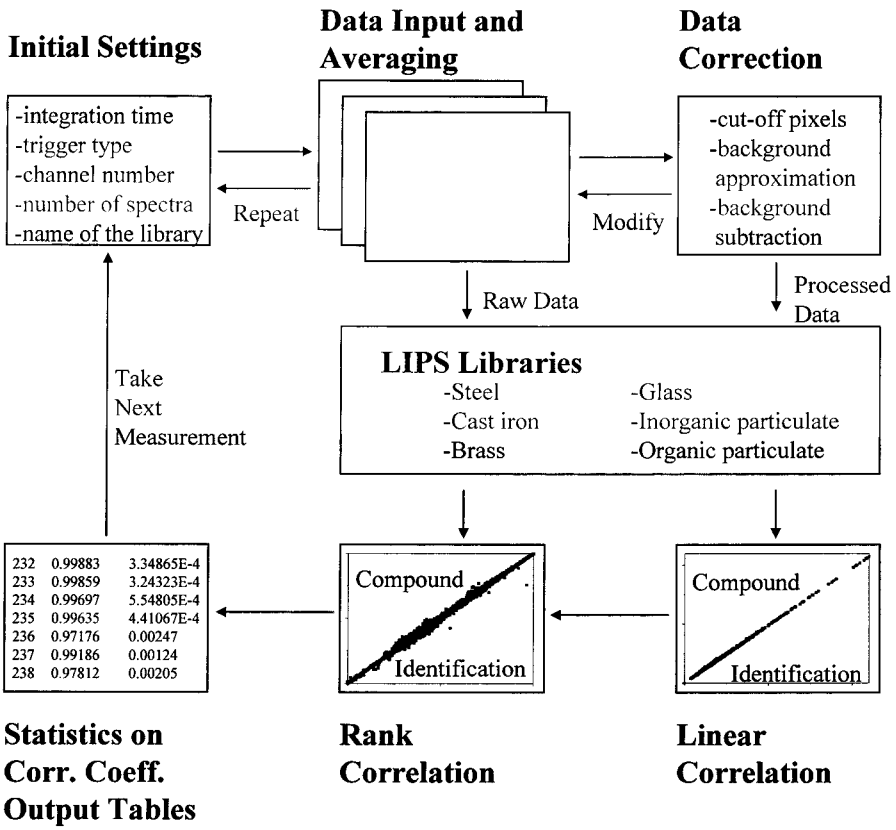

Figure 2. Flow chart for the software.

the two spectrometer channels alternatively. We also collected spectra of a wide variety of powdered inorganic materials. The library spectrum of a particulate material was obtained by inducing the laser spark on ~10 random particles and averaging the emission spectra obtained. All libraries were stored in a computer and could be chosen by an operator prior to analysis.

Software. A program was developed using Visual Basic 6.0 and the LabView drivers supplied with the Ocean Optics spectrometer. The software offers the following options to an operator (Figure 2). First, the operator chooses the desired experimental parameters: integration time, trigger type (internal or external), spectrometer channel (high or low resolution), number of spectra to average, and a library file. After the spectrum is displayed on the computer monitor, the operator has the option either to accept it or to repeat the measurement. For example, a repeated measurement is necessary when the laser misses the particle or the spectrum looks unusual. Once the spectrum is accepted, the operator can use it as a raw spectrum or pretreat it for subsequent correlation calculations. The pretreatment options include a reduction in the spectral range and a correction for continuous plasma background. Reduction of the spectral range may be necessary when a large portion of the recorded spectrum does not contain any useful information (no spectral lines), which often occurs when the low-resolution channel (200−850 nm) is used. Elimination of the uninformative portion of the spectrum simplifies the subsequent calculations. The background correction is carried out by an automatic search and selection of background points, approximating them with a polynomial, and subtracting the background from the initial spectrum. The power of the polynomial (up to 10) can either be set by the operator or be selected automatically. Of course, the library spectra have to be treated in the same way as the sample spectra. In the present work, we did not use the pretreatment option. We found that raw data are much more suitable for correlation than treated ones; any artificial manipulating with the data leads to loss of some crucial information and, as a result, a poorer correlation.

The operator selects a correlation method (linear or rank) and the computer calculates all mutual correlation coefficients between the current spectrum and all library spectra. The correlation plot, corresponding to the maximum correlation coefficient, is displayed along with the statistical parameters (correlation coefficients, errors, probabilities) and the name of the library file identified with the highest correlation probability. If desired, the correlation can be repeated with the use of another correlation method. These calculations take from a few to several tens of seconds for a library consisting of 50 substances. Finally, the output data are saved and stored in a hard drive and can be further processed, if desired.

RESULTS AND DISCUSSION

As we have already mentioned, our primary goal was to identify a compound belonging to a known class of compounds stored in a certain spectral library. It would also not be difficult to make a search among all available libraries in order to find a subgroup into which the analyzed material falls, but for the sake of time we omitted this step. Such an identification may seem trivial at first glance: the probe spectrum can be sequentially superimposed with the library spectra and the difference or similarity will immediately show up even without using a computer. However, if one looks at Figure 3 (groups a and b), where LIPS spectra from eight cast iron standards (Spectrometric Reference Materials,

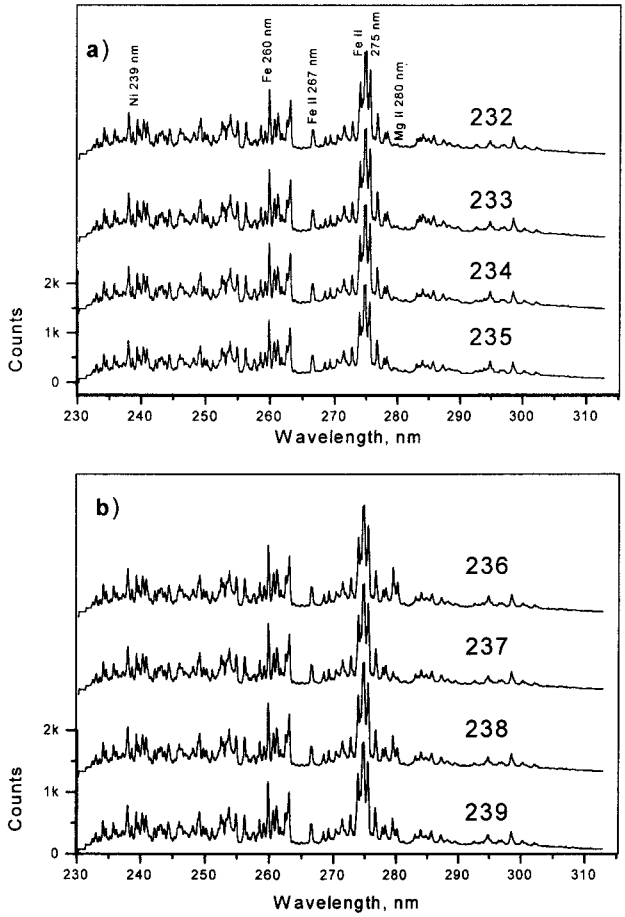

Figure 3. High-resolution channel emission spectra of the eight cast iron standards.

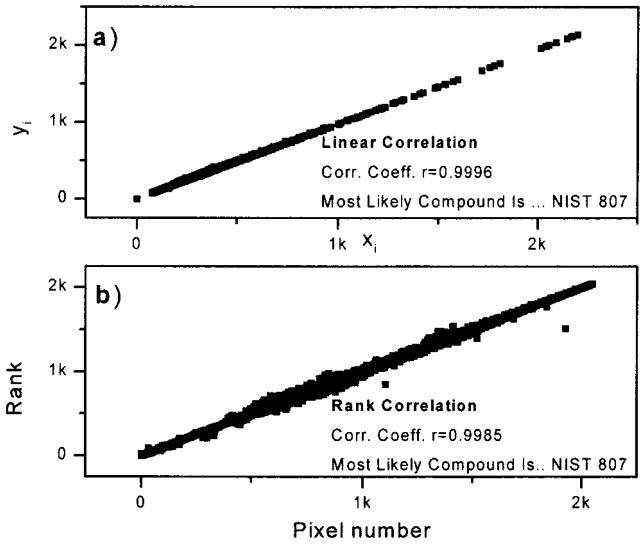

Figure 4. Linear and rank correlation plots for the NIST 807 stainless steel.

Cast Irons C18.8, Czechoslovakian Research Institute, series 232−239) are shown, visual identification is certainly not obvious. The spectra look totally identical. The compositions of these standards are very similar, as can be seen in Table 1. Therefore, powerful statistical methods are required in order to reliably identify such materials.

Table 1. Composition (%) of the Czechoslovakian Cast Iron Standards

std	C	Cr	Cu	Fe	Mg	Ni	P	Si
232	1.93	1.190	0.038	91.859	0.007	0.026	0.009	3.500
233	2.12	1.920	0.110	92.184	0.005	0.062	0.033	2.590
234	2.46	0.460	0.275	91.969	0.009	0.305	0.380	2.020
235	2.73	0.410	0.157	91.772	0.005	0.195	0.780	0.920
236	2.85	0.050	0.215	91.814	0.075	1.770	0.084	1.650
237	3.03	0.150	0.545	92.135	0.017	0.700	0.175	1.200
238	3.36	0.018	0.920	91.897	0.046	1.110	0.052	1.550
239	4.15	0.052	0.085	91.877	0.038	2.420	0.024	0.270

Because the spectra consist of 2048 points (pixels), enough statistical material is provided to permit the use of simple correlation methods such as linear correlation and nonparametric rank correlation. Linear correlation measures the association between variables, and the linear correlation coefficient r is expressed in the following form:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}} \quad (1)$$

where $\bar{x}$ is the mean of x_i's, and $\bar{y}$ is the mean of y_i's. The value of r lies between -1 and 1; $r = 1$ corresponds to complete positive correlation when the data points lie on a perfect straight line with positive slope, with x and y increasing together. The "goodness" of the correlation can be estimated from[4]

$$\alpha = \mathrm{erfc}\left(\frac{|r|\sqrt{N}}{\sqrt{2}}\right) \quad (2)$$

where α denotes the correlation significance, erfc(x) is the complementary error function, and N is the number of data points. A small value of α indicates that two variables are significantly correlated. Unfortunately, as pointed out by Press et al.,[4] the linear correlation coefficient r is a poor statistic for deciding whether an observed correlation is statistically significant or not because r is ignorant of the individual distributions of x and y. This may be important in our case where we compare fluctuating single-shot spectra (x's with different distributions) with stable library spectra (y's). In this situation, nonparametric rank correlation, where the x_i's and y_i's distributions are known a priori, is likely to be more robust. The equation for nonparametric correlation is the same as eq 1 but now the values of the x's and y's are replaced by their corresponding ranks R's and S's:

$$r = \frac{\sum_i (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_i (R_i - \bar{R})^2}\sqrt{\sum_i (S_i - \bar{S})^2}} \quad (3)$$

The ranks are numbers 1, 2, 3, ..., N, where N is a total number of data points (or pixels in our detector, 2048), which replace the

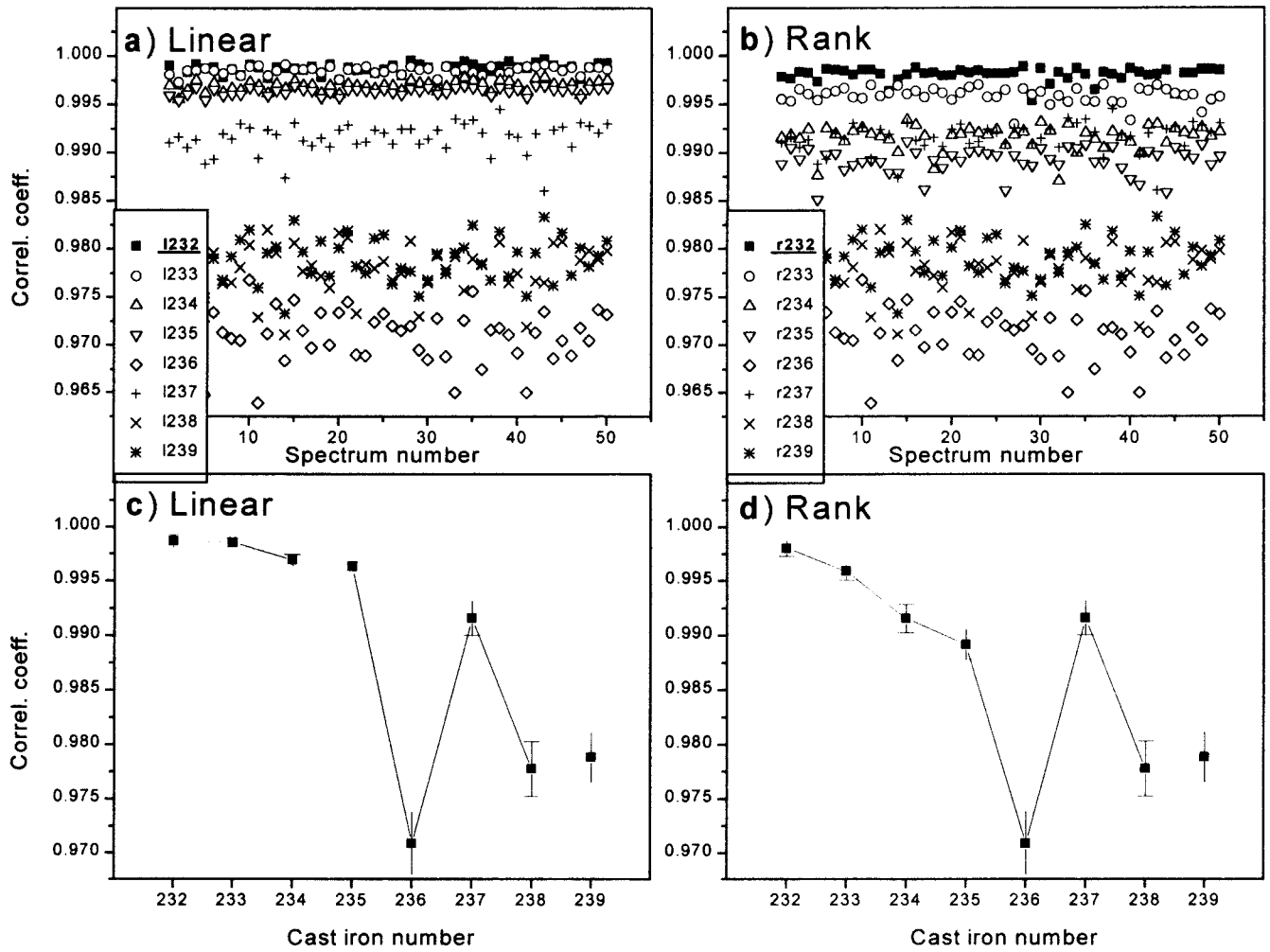

Figure 5. Linear and rank correlation coefficients for the eight cast iron standards. (a) and (b) are single-shot data and (c) and (d) are averages of 50 shots.

true values of *x*'s and *y*'s in accordance with their magnitudes. For example, the most intense pixel in a spectrum will be assigned the number 2048. Hence, the resulting list of numbers is drawn from a perfectly known distribution function, namely, uniformly from the integers between 1 and *N*. It is important to emphasize that if a correlation is proven nonparametrically then it really exists.[4]

We first applied both correlation methods for identification of the Czechoslovakian cast iron standards (Figure 3 and Table 1). Because of time considerations and also bearing in mind the potential of the technique for single-particle identification (where a single shot may be the only sampling event), we did not average spectra and used single-shot spectra to correlate with the cast iron spectral library. Typical correlation plots for the two correlation methods are shown in Figure 4. Typical values for all "single-shot" correlation coefficients as a function of a laser shot number are depicted in Figure 5a for linear correlation and in Figure 5b for rank correlation. Here the spectra from cast iron 232 have been correlated with all library spectra; the solid squares correspond to correlation of sample 232 with itself (232 vs 232), the open symbols to correlation with all other standards (232 vs 233−239). One can see from Figure 5a that linear correlation yields almost indistinguishable values for correlation coefficients taken for pairs 232 vs 232 (solid squares) and 232 vs 233 (open circles), whereas rank correlation separates these values completely (compare to Figure 5b). Plots in Figure 5c and d show the same trends for the averaged (over 50 laser shots) linear and rank correlation coefficients. It is also clearly seen from Figure 5c and d that all standards fall into two groups: one contains samples 232−235 and the other 236−239. This could be predicted on the basis of visual inspection of the spectra: the Mg II 280-nm line is absent in the first group (Figure 3a) and appears in the second group (Figure 3b).

Besides the apparent differences in correlation coefficients, strict statistical criteria must be applied in order to quantify the level of significance. A simple Student's *t*-test is a suitable solution when the distribution of a variable (the correlation coefficient) is close to normal. The normality of the distribution was checked by fitting a Gaussian function to our *r*-values. Two examples of such a fit are shown in Figure 6a) for a linear correlation and (b) for a rank correlation. As can be seen, both distributions are quite close to normal and therefore the use of a simple Student's *t*-test is appropriate.

The values for a Student's *t* should be calculated differently depending on whether the two distributions have the same or different variances. To check this, an *F*-test was applied (*F* denoting the ratio of the variances). If the calculated significance of *F* did not exceed 0.1, the difference in variances was considered

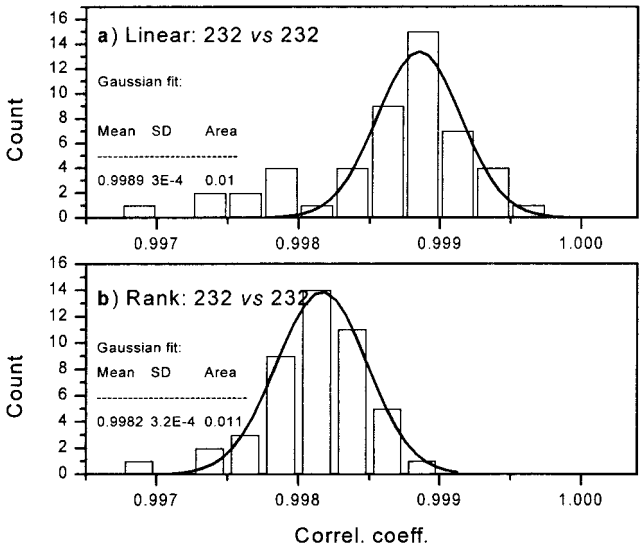

Figure 6. Correlations coefficient distributions for linear and rank correlations, for the No. 232 cast iron sample.

Table 2. Calculated Probabilities That Differences between Cast Iron Samples Were Detected Using Single-Shot Probe Spectra Correlated to 50-Shot Averaged Library Spectra[a]

Samples	232	233	234	235	236	237	238	239
232	0	0.8402	1	1	1	1	1	1
	0	1	1	1	1	1	1	1
233	1	0	1	1	1	1	1	1
	1	0	1	1	1	1	1	1
234	1	1	0	0.9275	1	1	1	1
	1	1	0	0.9699	1	1	1	1
235	1	1	0.9901	0	1	1	1	1
	1	1	0.882	0	1	1	1	1
236	1	1	1	1	0	1	1	1
	1	1	1	1	0	1	1	1
237	1	1	1	1	1	0	1	1
	1	1	1	1	1	0	1	1
238	1	1	1	1	1	1	0	1
	1	1	1	1	1	1	0	1
239	1	1	1	1	1	1	1	0
	1	1	1	1	1	1	1	0

[a] First number in each cell corresponds to the linear correlation routine, second number- to the rank correlation routine.

as significant and *t* was calculated in a slightly different way than in the case where the distributions had the same variances.[4] On the basis of these *t*-values, the probabilities that two distributions of correlation coefficients had different means were calculated. Table 2 shows the results of these calculations. It is clear that the same means have zero probability of being different (diagonal elements in Table 2). On the other hand, all the probabilities given in the table as unity differ from unity by negligibly small numbers, less than 10^{-8}. As can be seen from Table 2, for a rank correlation, only two samples are slightly confused with each other, 234 and 235, and even in this case, the probability of difference between them is ~90%. This is a remarkable result if we take into account that only single-shot spectra have been used.

Table 3. Composition (%) of the NIST 800-Series Stainless Steel Standards

std	Mn	Si	Cu	Ni	Cr	V	Mo	Al
805	1.38	0.19	0.12	0.15	0.18	0.01	0.01	
807	0.79	0.29	0.09	0.15	1.15	0.19	0.035	0.055
808	0.62	0.22	0.12	1.21	0.64			
818	0.52	0.28		0.11	0.96		0.22	

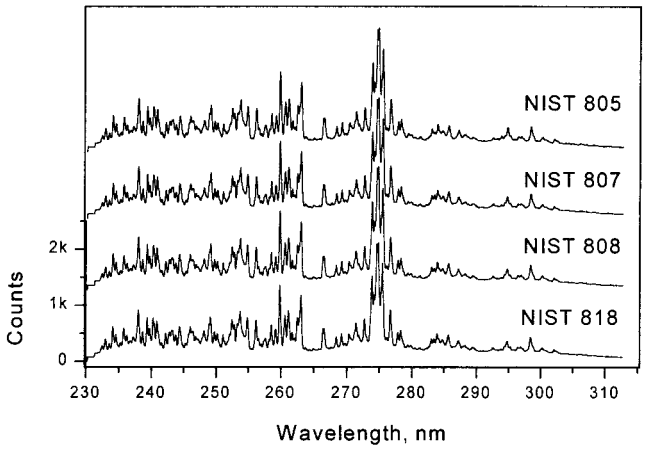

Figure 7. High-resolution channel emission spectra of the four NIST stainless steel standard samples.

The same methodology was applied to a set of 800-series stainless steel NIST samples (Table 3). LIPS spectra from these samples are shown in Figure 7. Even a very thorough visual inspection does not reveal any discernible differences between them. However, correlation analysis is able to reliably identify them. Figure 8 (like Figure 5) shows single and averaged correlation coefficients for the correlation of NIST 805 vs NIST 805−818. Again, rank correlation is better (compare panels c and d of Figure 8), giving a larger separation between correlation coefficients for different standard pairs. Table 4 shows data for the statistical significance of the separation in the form of probabilities. Compared to cast irons (Table 2), these data are slightly poorer because of the greater similarity in the composition of these standards (compare Table 1 and Table 3).

Here, it is important to mention that both the cast iron and steel spectra were recorded using a 1.0 neutral density filter (10-fold attenuation) placed in front of the spectrometer. The filter was used in order to avoid saturation of the detector at some strong iron lines (for example, Fe II 275 nm), otherwise some important information for correlation would be lost. This, in turn, decreased the instrument's sensitivity. Therefore, weak spectra from elements with concentrations well below 1% (almost all additives in the steels, Table 4) did not play the same crucial role for correlation as stronger spectral lines from elements with concentrations above 1% (several additives in the cast irons, Table 1). However, saturation of the detector yielded even worse correlation.

Despite these minor difficulties, the results in Table 4 look very encouraging, especially for rank correlation. All standards except one (NIST 807 was somewhat confused with NIST 818) were identified with probabilities better than 90%. It is interesting to note that averaging over 10 laser shots (performed for "difficult" standards 807 and 808) did not significantly improve the prob-

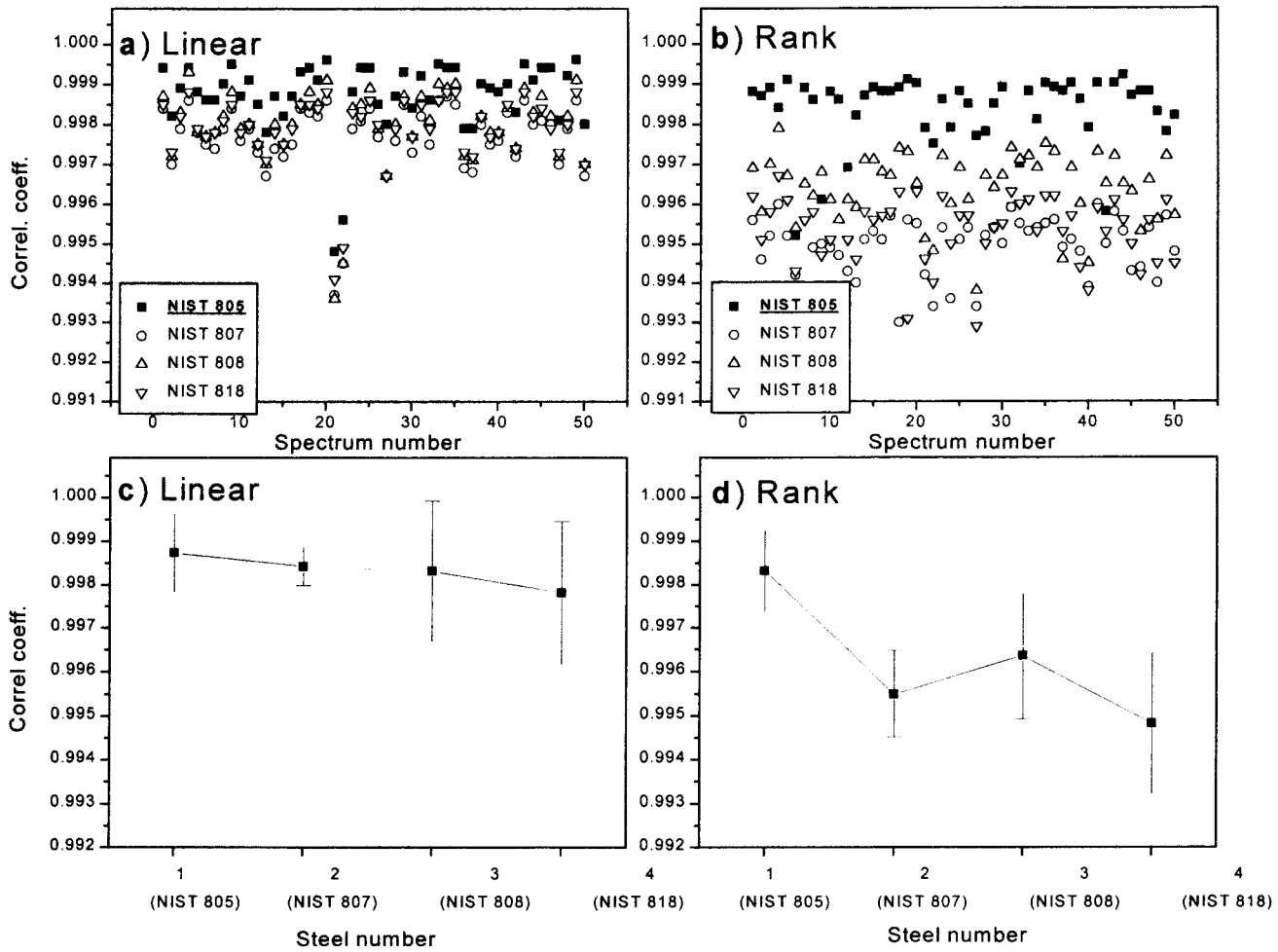

Figure 8. Linear and rank correlation coefficients for the four NIST stainless steel standards. (a) and (b) are single-shot data and (c) and (d) are averages of 50 shots.

Table 4. Calculated Probabilities That Differences between Stainless Steel Samples Were Detected Using Single-Shot Probe Spectra and 50-Shot-Averaged Library Spectra[a]

Corr. Type	Linear correlation with:				Rank correlation with:			
Samples	805	807	808	818	805	807	808	818
805	0	1	0.9998	1	0	1	1	1
807	1	0	0.9993	0.1414	1	0	1	.7571
	1*	0*	1*	0.2034*	1*	0*	0.9995*	0.7613*
808	0.8693	0.7199	0	0.1689	1	1	0	.9168
	1*	1*	0*	0.9834*	1*	1*	0*	.9754*
818	0.9974	0.625	0.6625	0	1	0.9984	0.9394	0

[a] Asterisk indicates 10-shot-averaged spectrum.

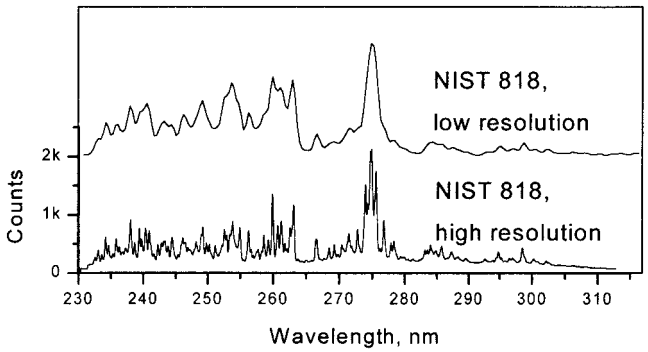

Figure 9. Emission spectra for the NIST 818 stainless steel, showing the data from both the low- and high-resolution spectrometer channels.

abilities of identification (the numbers marked with an asterisk in Table 4), except for the linear correlation for the pair 808 vs 818. This is likely the result of the high degree of homogeneity of these standards.

There is significant potential for improvement in the robustness of this technique. First, the higher spectrometer resolution results in an increased probability of correct identification. This can be illustrated by the example in Figure 9 where two spectra are shown, obtained with the use of the two spectrometer channels. It can be seen from Figure 9 that the high-resolution spectrum is much more informative in terms of individual spectral features, which are so important for correct identification. For example, when we attempted to identify the same steel and cast iron standards using the low-resolution channel, the ratio of correct to incorrect identifications was ~1. Although the probability of correct identification was still better than random (0.25 for the four steel standards and 0.125 for the eight cast irons), this result cannot even compare to the almost 100% correct identification with the use of a high-resolution channel. The statement above should,

however, be taken with precaution. If a spectrum of a certain sample is not concentrated in a narrow UV range as was case for the steels and cast irons, but widely distributed over the UV−visible region or contains broad molecular bands, the use of a low resolution but wide spectral window channel may be more advantageous. Ideally, high resolution and wide range, combined together (as in an echelle spectrometer) would provide the best possible performance for the proposed correlation routine.

Finally, we have observed, in working with this instrument and procedure over a period of several months, that the correlation is quite robust with respect to variations in instrumental parameters such as laser pulse energy and focus. The instrument was completely dismantled and rebuilt with a new laser head, and the correlation results with respect to the previously saved spectral libraries were virtually unchanged.

CONCLUSIONS

A compact laser-induced breakdown spectrometer with microscopic sample imaging has been developed for instant reliable classification of different groups of solid materials. The robustness of this technique was demonstrated by the nearly 100% reliable identification of two groups of materials of close composition: NIST stainless steel standards and Czechoslovakian cast iron standards. A software package was developed combining both data acquisition and data treatment functions. Linear and nonparametric (rank) correlation techniques were applied for classification of spectral data. The rank correlation was demonstrated to be more reliable compared to the linear correlation. The technique has also excellent potential for similar rapid identification of particulate materials. This work is currently under way in our laboratory.

ACKNOWLEDGMENT

This work was supported by the Engineering Research Center (ERC) for Particle Science and Technology at the University of Florida, the National Science Foundation (NSF Grant EEC 94-02989), the Industrial Partners of the ERC, and the Florida Institute of Phosphate Research (Grant FIPR 96-04-057R).

Received for review May 25, 1999. Accepted September 13, 1999.

AC9905524