Gas Identification System using Graded Temperature Sensor and Neural Net Interpretation

Lanwai Wong, Toshikazu Takemori, and M. W. Siegel

CMU-RI-TR-20-89

Intelligent Sensors Laboratory

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

May 1997

Abstract

An innovative approach is described for enhancing the selectivity of an integrated multi-element thick film gas sensor. A temperature gradient maintained along the sensor surface induces spatial sensitivity and selectivity gradients. The response map of the sensor is examined in the vapor of several organic compounds and their binary mixtures. Subtle but recognizable and separable signatures are observed. Emphasis is on identification, since quantitation of identified mixtures is straightforward. An effective and robust classification technique using a neural network trained via the back propagation method is described.

Introduction

The demand for selectively sensing vapors and gases is keen. For example, it would be valuable to distinguish toxic exhaust gases from fuel vapors and warn of either of these in the passenger compartment of automobiles, boats, and airplanes. Similarly, it is important for a domestic gas leakage detection system to avoid false alarms by differentiating between the odors of alcoholic beverages and the utility gas. Additional examples that require selectively sensing gases and vapors are found in military, industrial, and domestic applications. Most chemical analysis systems for gaseous samples, such as gas chromatography, are expensive, complicated to use, and at best marginally portable. In contrast, relatively inexpensive, simple, highly portable gas sensing devices based on resistivity changes of semiconducting materials are in common use. One of the most common materials is tin oxide, SnO_2. It has high sensitivity to a large number of gases and is relatively easy to use. However its high sensitivity to many gases is a weak point when it comes to sample discrimination and mixture analysis. Substantial efforts have been directed at improving the selectivity of SnO_2 based sensors by various approaches using filters, special absorbing layers, and various catalysts or promoters.

Temperature programming is among the many techniques that have proven useful for enhancing selectivity. The sensing mechanism of SnO_2 to reducing agents is believed to be due to the chemical reactions between their molecules and surface adsorbed and ionized oxygen as O^- and perhaps other species. There is an optimum temperature at which sensitivity maximizes. If the temperature is lower, the reaction is too slow to give high sensitivity; if the temperature is higher, the overall oxidation reaction proceeds too rapidly. In the latter case, diffusion of the reducing agent is confined to a thin layer near the surface, and the effective concentration of reducing agent seen by the sensor bulk is decreased. Thus if an SnO_2 sensor is ramped through an appropriate temperature range, the sensitivity to a gaseous sample will show a peak at a particular temperature. Sensitivity peaks associated with the various gases present will generally appear at different temperatures in this temperature programming technique. However, hysteresis in the sensor resistance during temperature cycling generally broadens these peaks. Thus, temperature programming of the sensor is not a common technique used for selectivity enhancement; a notable exception is the Figaro Corporation approach to CO alarming, which employs a Taguchi sensor and a complex temperature programming method [1].

We are approaching discrimination by temperature from another direction. Instead of cycling the temperature of a single sensor, we apply a static temperature gradient along a linear sensor array. In fact, the "array" is a continuous film on a ceramic substrate, and the discrete array elements are defined only by the contact metalization. Since the sensing elements on the substrates are fabricated under exactly the same conditions, their sensing properties should be extremely uniform at constant temperature. But when these sensors are at different temperatures, their relative responses to the components of gas mixtures differ, providing recognizable signatures.[1] Significantly different response patterns of an SnO_2 array to ethanol, methanol, heptane and their binary mixtures were observed using the temperature gradient technique. The reproducibility of these response patterns is acceptable for laboratory studies, but stability improvement is a subject of ongoing research toward field applications.

The coarse spatial resolution of our sensor (seventeen elements in 23 mm), the difficulty in measuring precisely the temperature distribution, and the small but non-negligible intrinsic structural differences among the sensing elements, makes the development of an analytical model to describe the sensing properties of the integrated thick film sensor very difficult. An alternative approach is to interpret pragmatically the experimental data via machine learning, pattern recognition, and perhaps

---

1. An alternative approach, operating at constant temperature with a gradient in catalyst concentration or composition, and a combination approach with perpendicular temperature and catalyst gradients is also employed and will be described in future reports. In fact, some of our devices employ both approaches, with perpendicular gradients in temperature and catalyst composition.

knowledge based artificial intelligence methods. In this report we test this approach when implemented via the neural network [2] with training by back propagation method [3].

Sensors

The sensor arrays we use in these experiments are fabricated on commercial alumina ceramic substrates intended for hybrid circuits. The area of the substrate is compatible with a 20-pin dual-in-line integrated circuit. A gold electrode metallization pattern is printed, then fired at high temperature onto the substrate, as shown in Figure 1. Next a layer of a modified commercial SnO_2 semiconductor ink is screen printed on top of the electrodes. The gold pattern provides stable electrical contact between the sensing layer and the measurement equipment. The modified SnO_2 ink and the screen printing technique provide a cost-effective method to fabricate a high quality sensor array. Catalysts such as fine Pt metallic powder are blended into the commercial inks to enhance sensitivity. The organic binders in the ink are burned away by firing in an oven. A detailed description of the fabrication procedure, which is actually carried out by collaborators at Oak Ridge National Laboratories, is given in [4], [5], [6].

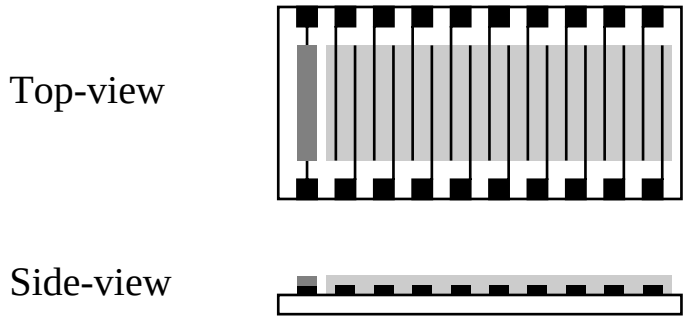

Figure 1. METALLIZATION PATTERN FOR CONTACTS BETWEEN INSTRUMENTATION AND SENSING LAYER

Measurements

The integrated sensor is mounted on a specially designed chip holder that incorporates pressure contact probes, electrical heaters, and temperature sensing devices. The whole assembly is installed in an air-tight metal test chamber supplied with a constant flow of either pure synthetic air or dilute sample vapor in synthetic air. Experiments have been performed on the integrated sensor with the vapor of ethanol, methanol, heptane and their binary mixtures. These vapors are generated by saturating a secondary synthetic air flow in a bubbler, then diluting it in a precision flow controlled mixing system. The binary mixtures are made by turbulent mixing of the outputs of two bubblers. The concentration of the sample is computed from the volume of sample mixed with synthetic air, temperature measurements, and vapor pressure tables. The result is confirmed by measurement with a total hydrocarbon analyzer located down-stream of the test chamber. A simplified block diagram of the gas delivery system is shown in Figure 2.

A temperature gradient is established along the length of the sensor by preferentially heating one end of the substrate with an asymmetrically located, small, high power ceramic heater. The temperature difference between the hot end and cold end of the sensor surface is typically about 100 C. Before

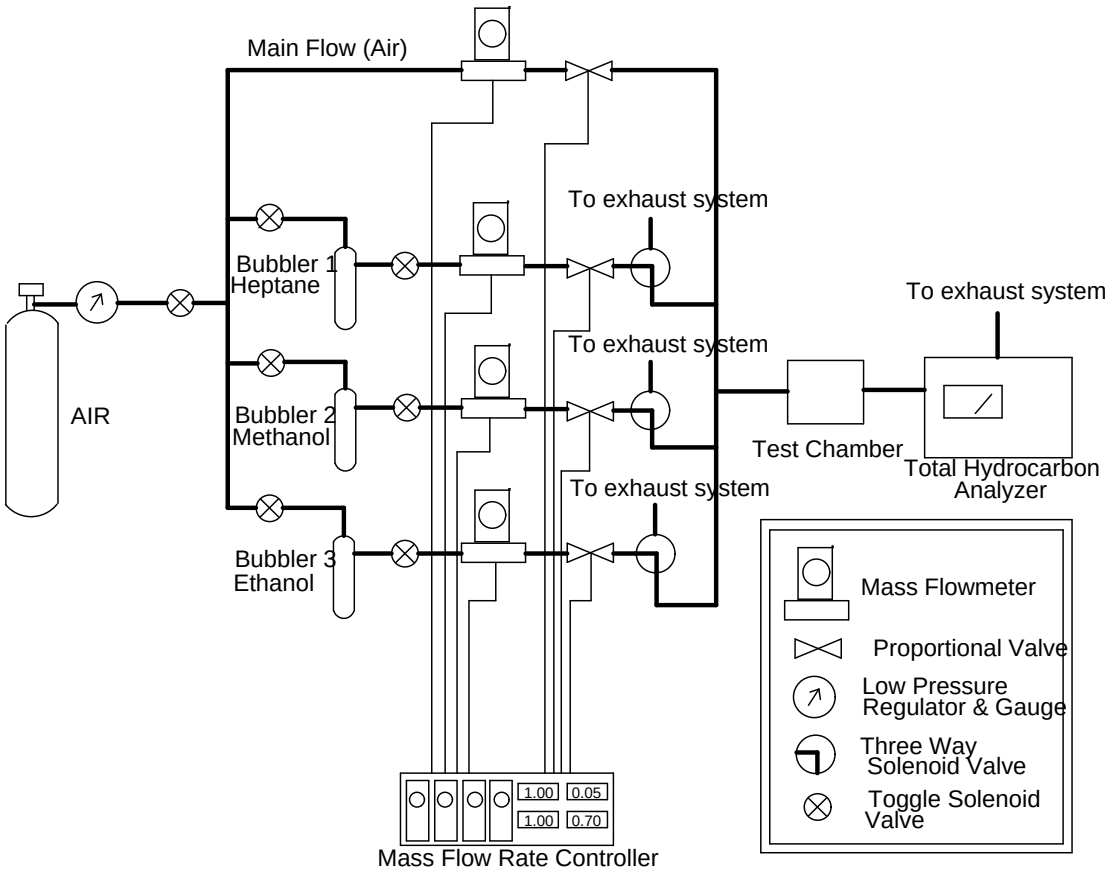

Figure 2. BLOCK DIAGRAM OF THE GAS DELIVERY SYSTEM

Solenoid valves and flow rate controller set points are
controlled by the IBM-XT via the Keithley data acquisition system.

taking any measurements, the test chamber is flushed with synthetic air at 1000 ml/min flow rate for at least five hours. Then in a typical experiment the sample gas is turbulently mixed with the main air flow and is introduced into the test chamber for 50 min. The response patterns of the sensor associated with various gases are recorded by measuring the resistances cyclically. Data are collected with an automated data acquisition system using an HP 3421A IEEE 488 unit and a Keithley 500 memory mapped data acquisition system controlled by an IBM-XT. Preliminary data reduction is done on-line by the XT. Final processing, as well as methods development, is done off line on a 750 VAX to which the IBM-XT is connected by ethernet. A simplified block diagram of the data acquisition and control system is shown in Figure 3.

Data

Typical resistance responses of several sensing elements to ethanol, methanol and heptane are shown in Figure 4. The resistance of each integrated sensor element is reduced by each vapor as expected for an n-type semiconductor material. The response plotted in Figure 4 is the ratio of the change in resistance to the original resistance in air; this is the conventional definition of response used in the solid state gas sensing field. Each sensitivity to each test compound increases with operating temperature in the temperature range of these experiments. At the hot end, each absolute sensitivity is three to four times higher than at the cold end, and of a similar magnitude to the commercial Figaro 812 sensor that is installed in the test chamber for comparison. To minimize the irreversible changes due to abrupt environmental shock to the sensor, several measurement cycles

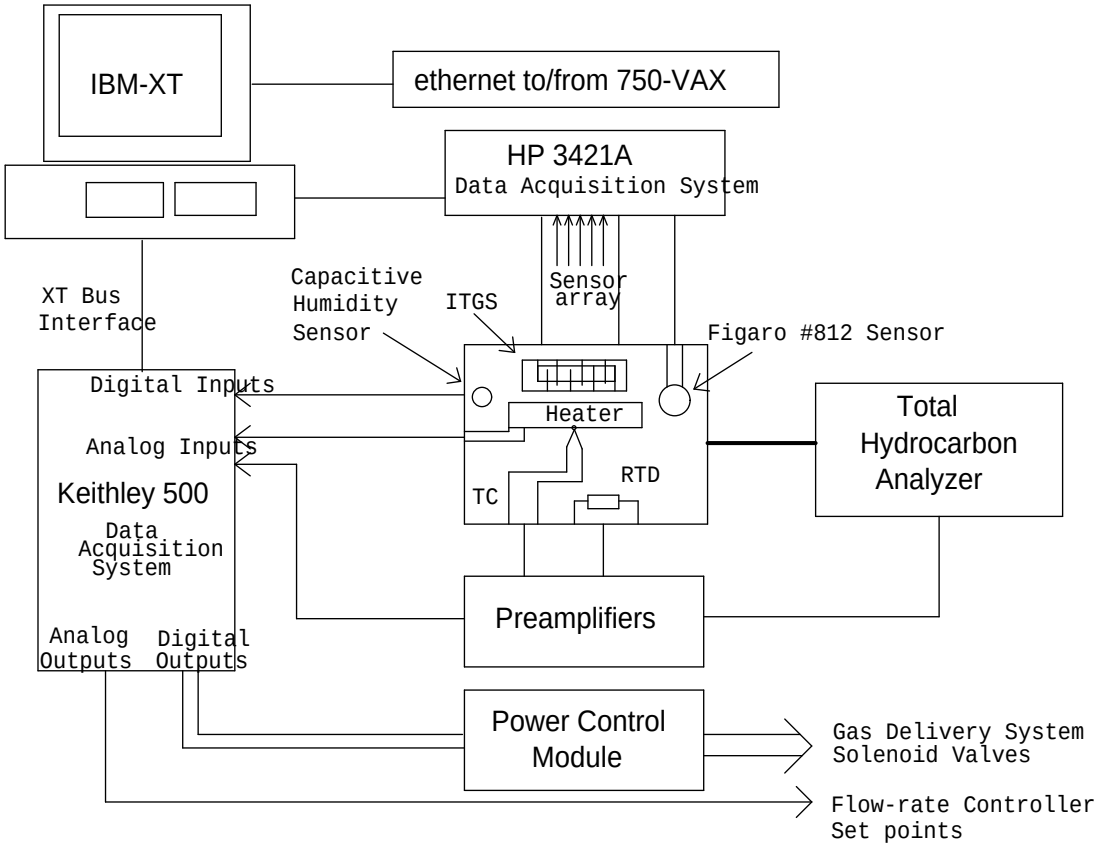

Figure 3. BLOCK DIAGRAM OF DATA ACQUISITION AND CONTROL SYSTEM COMPONENTS

High precision measurement of sensor element resistances is via the HP 3421A data acquisition system on the IEEE-4888 bus controlled by the IBM-XT. Moderate and low precision monitoring of housekeeping and support sensors, and control of valves and heaters, is via the Keithley 500 data acquisition system, whose functionality is memory mapped on the IBM-XT bus. Preamplifiers for thermocouples and RTDs are standard modules for these functions. The total hydrocarbon analyser (a flame ionization detector manufactured by Mine Safety Appliances) and the Figaro #812 Taguchi-type sensor are used for checking and comparison. The capacitive humidity sensor is monitored via a frequency measurement applied to the capacitance-to-frequency output of a capacitance meter.

are made with each sample vapor before changing to another vapor. This method, as opposed to more frequent cycling of sample type, increases vulnerability to slow drift. To alleviate the effect of long term drift, the individual sensitivities are normalized by the difference between the maximum and minimum sensor element sensitivities found in each experiment separately. Sensitivity distributions normalized this way are very consistent with respect to repeated runs with the same sample species, and distinctly different from sample species to sample species.

Algorithms

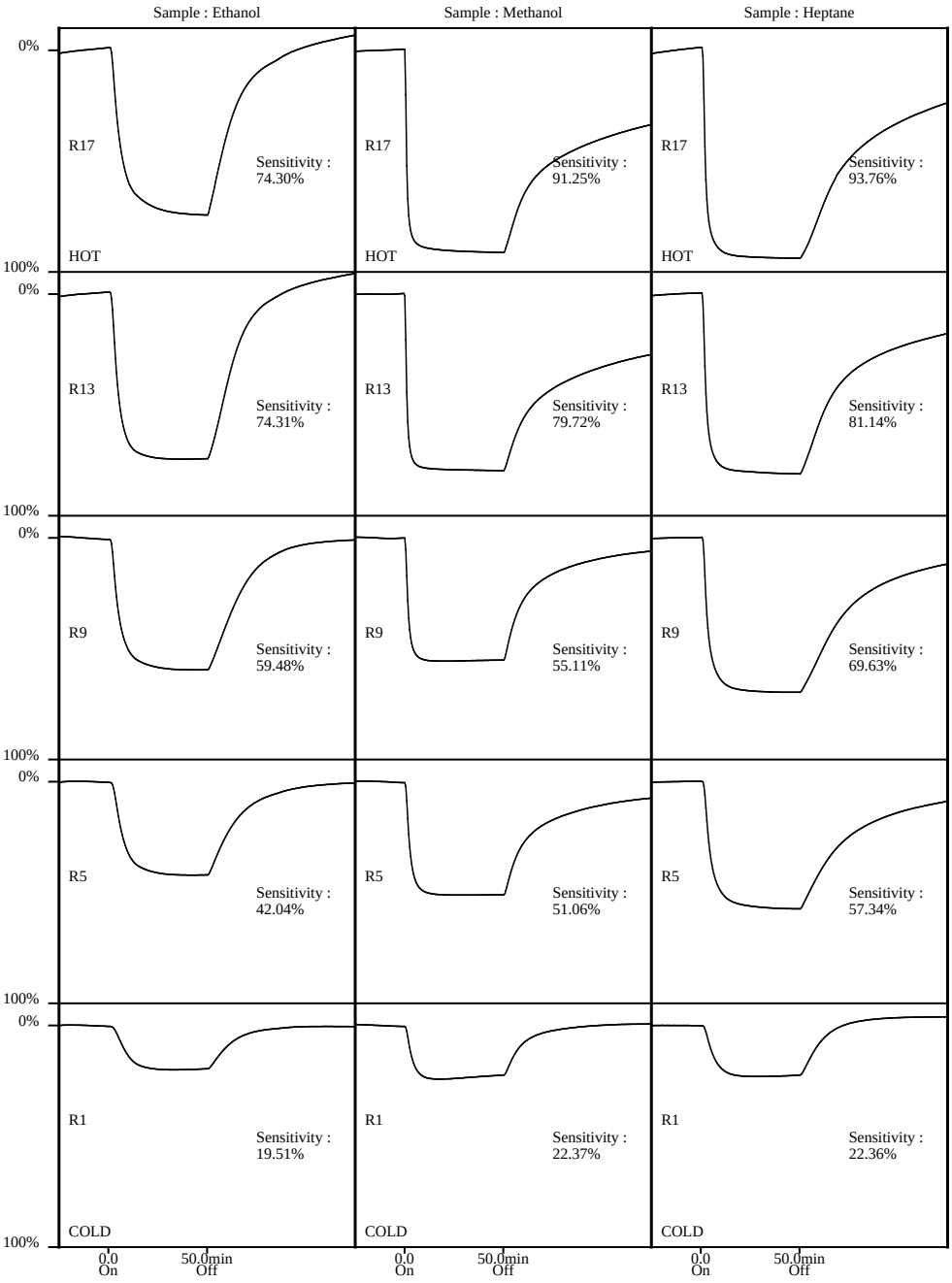

Figure 4. TYPICAL RESPONSES OF THE SENSOR TO ETHANOL, METHANOL AND HEPTANE VAPORS

Sample vapor is mixed into the artificial air flow through the test chamber for 50 minutes. The "sensitivity" is by convention the resistance difference between the sensor in air and the sensor in air with sample vapor divided by resistance in air, *i.e.,* $\Delta R/R_0$. The time constant of the rapid resistance drop at the hot end upon methanol and heptane introduction is close to the limit set by the ratio of test chamber volume to volumetric flow rate. The longer fall times for ethanol, and for methanol and heptane on the cooler parts of the sensor, and the longer rise times for all samples at all temperatures upon termination of sample introduction are characteristic of the sensor.

Data Interpretation

A three-layered neural network, the simplest non-trivial implementation [7, 8], is investigated for data interpretation with our integrated thick film sensor. It has one input layer, one hidden layer, and one output layer. The topology of this neural network is illustrated in Figure 5. Each unit above the input layer is completely connected to all units in the layer one level below.

The learning capacity of a neural network is bounded by its complexity in number of levels and number of units in each level. The information stored in a neural network is distributed in the weights of the connections and the thresholds of the units. Initially, all weights and thresholds in the network are randomly set to small values. Then a set of training data (in a demonstration of this sort a subset of the experimental data) is presented to the input units of the network. The weights and thresholds are adjusted by the back propagation [3] learning algorithm until the outputs converge on the known input classifications. The result of a classification is interpreted by a decision rule that relates the numerical outputs of the top layer to a set of symbolic descriptions of the systems activating the inputs, *e.g.,* "methanol/ethanol mixture". The generality of the result is tested by measuring the trained network's classification performance on the remainder of the experimental data, *i.e.,* the part not in the training set.

Quantitation is straightforward once identification is achieved, since any one sensor is sufficient to quantitate an identified unary sample, any two linearly independent sensors are sufficient to quantitate an identified binary mixture, etc.

Input Feature Selection

We have investigated two ways of presenting the data to the neural network. We can categorize these as the "direct input method" (DIM) and the "extracted features input method" (EFIM). The difference between them is the degree of preprocessing of the experimental data before they are input to the neural network. DIM presents essentially raw data to the input layer of the neural network, while EFIM employs some form of preprocessing to compute (by analog or digital methods) selected features derived from the data. DIM is appropriate when there is a high level of ignorance about the physical system generating the data. EFIM reduces the burden by employing models of the physical system. Thus EFIM is comparable to the way humans operate in familiar contexts, while DIM may be comparable to the way humans operate when faced with data in new and confusing contexts.

DIM

Essentially raw data from a sensor array can be directly applied to the inputs of a neural network. However with the direct input method the trained neural network is observed to focus on the absolute differences among the input values in the data. The network is then easily misled: it tends to make its identification based on small and perhaps irrelevant differences over many units rather than to base it on truly discriminating features that are manifest over a small number of input units. To minimize these effects, which are in part concentration dependent, preprocessing that normalizes each pattern is advantageous.

The number of DIM input units is equal to the number of sensing elements. The number of hidden units is chosen empirically, roughly optimized by observing the activities of the hidden units in the trained network. The number of output units is just the number of classes into which the samples will be classified. The activity value of each output unit is interpreted as the probability of the sample being in the corresponding class. In our case, a network of seventeen input units, eight hidden units, and six output units was used to classify ethanol, methanol, heptane and their binary mixtures over a range of concentrations. The inputs are the normalized responses of individual sensing elements on the integrated thick film sensor. One output unit was assigned to each of the six possible "answers," and during training the corresponding output unit was assigned unit output, while the rest were assigned zero output.

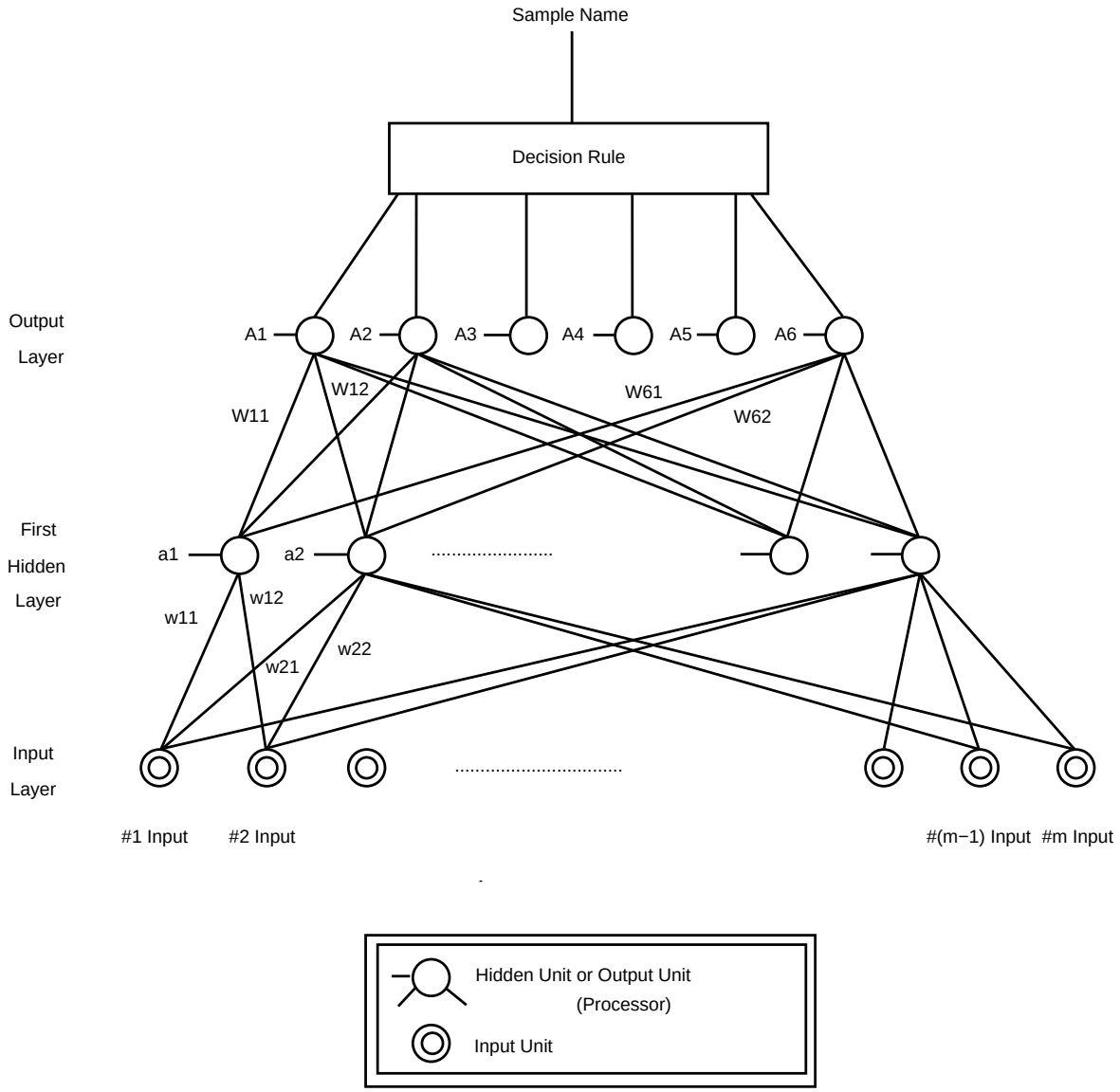

Figure 5. TOPOLOGY OF THREE LAYERED NEURAL NETWORK

For graphic simplicity only a few connections have been depicted;
in fact in the network we use every hidden unit is completely connected to
every input and output unit. a_i and A_i depict threshold
values for hidden and output units respectively. w_{ij} and
W_{kl} depict weights connecting input unit i to hidden unit j,
and hidden unit k to output unit l respectively.

EFIM

The extracted feature input method can be a very powerful technique if one can identify and precompute, before presentation to the neural net inputs, features similar to those which humans seem to use to make classification decisions. EFIM is less general than DIM in the sense that it requires feature identification in advance by a human. As with conventional pattern recognition, classification with neural nets generally works better if a human selects features to precompute than if essentially raw data are presented.

Our experiments have focused on the spatial first derivatives of normalized sensitivity distributions as

EFIM input data. The number of EFIM input units is the number of chosen features, in this case sixteen spatial first derivatives from seventeen sensors. The number of hidden units is chosen by the same empirical method used in DIM. The rest of the network, consisting of eight hidden units and six output units, was the same as in DIM. In a sense, our EFIM approach is to insert a fourth layer between the original input and hidden layers, this extra hidden layer being prewired to output differences between adjacent inputs.

Output Classification

The output value or activity of each output unit in our trained network is by definition proportional to the probability of the sample being in the corresponding class. Since the activity of each output unit ranges between zero and one, a plausible top level discriminant function P_i can be defined as

$$P_i = o_i \prod_{j=1,\, j \neq i}^{n} (1 - o_j)$$

where o_i is the activity of the i^{th} output unit corresponding to the i^{th} sample class, i indicates a unary gas species or a specific binary mixture, j is a dummy index over the same space, and n is the number of sample classes, *i.e.,* the total number of unary plus binary samples in the output space. P_i is essentially the joint probability that the sample is in class i and not in any other class.

Training

Learning by back propagation employs two traversals of the net, one forward and one backward, each time each input vector in the training set is presented. In the forward pass activity propagates (as usual) from the input units through the hidden layers to generate the output vector. In the backward pass the error (the difference between the output vector and the known truth) is back propagated through the same connections from output layer toward input layer. This facilitates computing, for each connection, the gradient of the output error with respect to the weight of the connection. The weights are then changed so as to reduce the error. In effect learning works by gradient descent on an error surface in weight space.

The total input to unit j, denoted x_j, is by definition a weighted sum of the inputs y_i of the units in the previous layer[2]. With the weights denoted w_{ji}:

$$x_j = \sum_i w_{ji} y_i$$

Each unit is taken as having an output, y_j, defined to be a sigmoidal function (analytically approximating the instantaneous step function) of its total input. A suitable form (with slope and x-offset suppressed for simplicity) is:

$$y_j = \frac{1}{1 + e^{-x_j}}$$

which varies between 0 at large negative values of x_j and 1 at large positive values of x_j, and has value 0.5 and slope 1 at $x_j = 0$.

The goal of the training procedure is to find a set of weights that ensure that for each input vector in the training set the output vector produced by the network is (in some sense depending on the application and context) sufficiently close to the output vector corresponding to the known truth. The neural net literature analyzes in detail the network configurations and training schemes under which

2. Connection schemes that interconnect units of the same layer, skip layers, loop back, etc., are sometimes considered.

training data sets can be learned perfectly, and, more importantly, the circumstances under which data not in the original training set will be correctly classified. Our experiments are oriented toward demonstrating and categorizing the robustness and generality of the method with real data from real sensors.

A heuristic measure of performance is provided by extending the hill climbing analogy to a potential energy function. If there is a fixed, finite set of input-output cases, the total error in the performance of the network with a particular set of weights can be computed by comparing the actual and desired output vectors for every case. The total error E is defined as

$$E = \frac{1}{2}\sum_{c}\sum_{j}(y_{j,c}-d_{j,c})^2$$

where c is an index over cases (input-output pairs), j is an index over output units, $y_{j,c}$ is the actual state of an output unit, and $d_{j,c}$ is the state that output unit $y_{j,c}$ would be in were it reporting the truth.

The straightforward approach to gradient descent is to change each weight by an amount proportional to its influence on the error:

$$\Delta w_{ij} = -\varepsilon\frac{\partial E}{\partial w_{ij}} \qquad (1)$$

where ε is an empirically selected constant.

Convergence is more rapid if higher order derivatives are taken into account. The second derivative, for example, is accounted for by revising equation 1 to

$$\Delta w_{ij}(n) = -\varepsilon\frac{\partial E}{\partial w_{ij}}(n) + \alpha\Delta w(n-1)$$

where n is incremented by 1 for each sweep through the whole set of input-output cases, and α is an exponential "forgetting" factor between 0 and 1.

gas/date/id	E	M	H	E+M	M+H	E+H	Air	%correct
E 6.16a	97	0	0	0	0	1	2	99
E 6.16b	97	0	0	0	0	0	3	100
E 6.16c	98	0	0	0	0	0	2	100
M 6.10e	0	99	0	0	0	0	1	100
M 6.10f	0	98	0	0	0	0	2	100
M 6.10g	0	99	0	0	0	0	1	100
M 6.17a	0	63	0	36	0	0	1	64
H 6.14a	11	3	84	0	0	0	2	86
H 6.14b	14	2	82	0	0	0	2	84
H 6.14d	12	4	82	0	0	0	2	84
H 6.14e	15	3	80	0	0	0	2	82
E+M 6.21a	0	7	0	92	0	0	1	93
E+M 6.21b	0	6	0	93	0	0	1	94
E+M 6.21c	0	6	0	93	0	0	1	94
M+H 6.23a	8	0	0	0	91	0	1	92
M+H 6.23b	14	0	0	0	84	0	2	86
E+H 6.22a	2	0	0	0	0	97	1	98
E+H 6.22b	3	0	0	0	62	34	1	35
E+H 6.22c	3	0	0	0	80	16	1	17

Table 1

Performance under training by DIM (17 input units)

Results

Table 1 shows the result of tests on ethanol (E), methanol (M), heptane (H), and their binary mixtures generating data of the type illustrated in Figure 4 using the DIM method, *i.e.,* seventeen input units corresponding to seventeen sensor array elements. The seventeen inputs are re-normalized to a minimum value of 0 and a maximum value of 1 in each cycle to minimize the effect of slow drifts in baseline and sensitivity. Identification is by the probability method outlined in the *Output Classification* section, with the modification that if the largest response ($\Delta R/R_o$) seen among the seventeen sensors is less that 0.01 then the sample is classified as air, *i.e.,* in effect classification is not attempted. Training data are two randomly selected members of the run for each gas, *i.e.,* for ethanol two members of the set designated gas *E*, date *6.16*, id *a*. Table 1 summarizes 1900 data vector measurements collected in 19 separate cycles of 100 measurements each. The data were collected on seven separate days during a two week period. Out of 1900 possible classifications, 1608 or 85% were correct.

gas/date/id	E	M	H	E+M	M+H	E+H	Air	%correct
E 6.16a	98	0	0	0	0	0	2	100
E 6.16b	97	0	0	0	0	0	3	100
E 6.16c	98	0	0	0	0	0	2	100
M 6.10e	0	99	0	0	0	0	1	100
M 6.10f	0	98	0	0	0	0	2	100
M 6.10g	0	99	0	0	0	0	1	100
M 6.17a	0	99	0	0	0	0	1	100
H 6.14a	9	1	88	0	0	0	2	90
H 6.14b	11	1	86	0	0	0	2	88
H 6.14d	9	2	87	0	0	0	2	89
H 6.14e	12	1	85	0	0	0	2	87
E+M 6.21a	0	3	0	96	0	0	1	97
E+M 6.21b	0	2	0	97	0	0	1	98
E+M 6.21c	1	0	0	98	0	0	1	99
M+H 6.23a	6	0	0	0	93	0	1	94
M+H 6.23b	11	0	0	0	87	0	2	89
E+H 6.22a	2	0	0	0	4	93	1	94
E+H 6.22b	3	0	0	0	3	93	1	94
E+H 6.22c	3	0	0	0	2	94	1	95

Table 2

Performance under training by EFIM (16 input units)

Table 2 shows the result of tests on ethanol (E), methanol (M), heptane (H), and their binary mixtures generating data of the type illustrated in Figure 4 using the EFIM method, *i.e.,* sixteen input units corresponding to sixteen spatial first derivatives across the seventeen sensor array elements. The sixteen inputs are re-normalized to a minimum value of 0 and a maximum value of 1 in each cycle to minimize the effect of slow drifts in baseline and sensitivity. Identification is by the probability method outlined in the *Output Classification* section, with the modification that if the largest response ($\Delta R/R_o$) seen among the seventeen sensors is less that 0.01 then the sample is classified as air, i.e., in effect classification is not attempted. Training data are two randomly selected members of the run for each gas, *i.e.,* for ethanol two members of the set designated gas *E*, date *6.16*, id *a*. Table 2 summarizes 1900 data vector measurements collected in 19 separate cycles of 100 measurements each. The data were collected on seven separate days during a two week period. Out of 1900 possible classifications, 1814 or 96% were correct.

gas/date/id		E	M	H	E+M	M+H	E+H	Air	%correct
E	6.16a	97	0	0	0	0	1	2	99
E	6.16b	97	0	0	0	0	0	3	100
E	6.16c	98	0	0	0	0	0	2	100
M	6.10e	0	99	0	0	0	0	1	100
M	6.10f	0	98	0	0	0	0	2	100
M	6.10g	0	99	0	0	0	0	1	100
M	6.17a	0	99	0	0	0	0	1	100
H	6.14a	12	1	85	0	0	0	2	87
H	6.14b	16	0	82	0	0	0	2	84
H	6.14d	14	1	83	0	0	0	2	85
H	6.14e	17	0	81	0	0	0	2	83
E+M	6.21a	0	6	0	93	0	0	1	94
E+M	6.21b	0	5	0	94	0	0	1	95
E+M	6.21c	0	5	0	94	0	0	1	95
M+H	6.23a	12	0	0	0	87	0	1	88
M+H	6.23b	24	0	0	0	74	0	2	76
E+H	6.22a	3	0	0	0	0	96	1	97
E+H	6.22b	4	0	0	0	9	86	1	87
E+H	6.22c	5	0	0	0	18	76	1	77

Table 3

Performance under training by DIM + EFIM (33 input units)

Table 3 shows the result of tests on ethanol (E), methanol (M), heptane (H), and their binary mixtures generating data of the type illustrated in Figure 4 using a combination of the DIM and EFIM methods, *i.e.,* thirty three input units corresponding to the seventeen sensor elements plus their sixteen spatial first derivatives. The DIM and EFIM input sets are individually re-normalized to a minimum value of 0 and a maximum value of 1 in each cycle to minimize the effect of slow drifts in baseline and sensitivity. Identification is by the probability method outlined in the *Output Classification* section, with the modification that if the largest response ($\Delta R/R_o$) seen among the seventeen sensors is less that 0.01 then the sample is classified as air, i.e., in effect classification is not attempted. Training data are two randomly selected members of the run for each gas, *i.e.,* for ethanol two members of the set designated gas *E*, date *6.16*, id *a*. Table 3 summarizes 1900 data vector measurements collected in 19 separate cycles of 100 measurements each. The data were collected on seven separate days during a two week period. Out of 1900 possible classifications, 1747 or 92% were correct.

Adding additional training data, particularly data selected from cycles on which performance is poor, improves performance statistics substantially, since the overall "grade point average" in each case suffers badly from a few extremely poor performances. This is of course in keeping with the well known result that perceptron-like neural nets are guaranteed to learn their training data perfectly. However it is observed that it is not only the learning of the new training data that contributes to improved overall performance: there is general improvement in performance on all data in the cycles from which the new training data are taken. This illustrates the ability of machines of this type in some situations to generalize, and to abstract classes from disjoint sets.

A somewhat surprising result is that performance on the set of inputs consisting of the DIM plus the EFIM data is better than the performance on the DIM data alone, but worse than on the EFIM data alone: in fact it is just a little better than their average. Understanding whether this effect is of a general nature or an anomaly of these particular experiments awaits further work.

As a final comment, it is clear were we to define any reasonable method for assigning (in teaching parlance) "partial credit," effectively to recognize that the classes are not inherently orthogonal, this statistic would obviously be improved.

Conclusion

An array of identical metal oxide gas sensors, each operated at a different temperature, showed sufficient differential sensitivity to discriminate among several simple organic vapors and their binary mixtures. The array was implemented by a spatial temperature gradient along the surface of a single large area thick film device. A three layer neural network, trained by the back propagation algorithm, was able automatically to generate the classification scheme. This was true for input of either raw sensor responses (the direct input method) or of preprocessed sensor responses (the extracted feature input method) in the form of the spatial derivative of response.

The difference between a multilayer neural network and its antecedents in conventional pattern recognition with linear discriminants is that the neural network is able to perform recognition tasks on nonlinearly separable data. The difference between neural network classifiers and classification by symbol manipulation based artificial intelligence lies in its ability to distinguish subtle quantitative differences among the input features, and the existence of quantitative cues embedded in the data which cannot easily be represented symbolically. In situations requiring classification based on multiple sensor responses, the neural network may be a particularly good choice even in a noisy environment.

Acknowledgements

The sensors were fabricated by our collaborators, Robert J. Lauf and Barbara S. Hoffheins, with the assistance of their colleagues at Oak Ridge National Laboratories. Support for the work at Oak Ridge, and at CMU for developing the concept, building the apparatus, and initiating the laboratory work, was provided by Cabot Corporation (Boston) and enthusiastically monitored by C. Law McCabe. Invaluble laboratory assistance and searching of the commericial products literature was provided by Alan D. Guisewite. Toshikazu Takemori's participation in our laboratory was supported by Osaka Gas Company (Japan). An unrestricted grant provided by Rudolf E. Kubli in behalf of Mettler Instrumente AG (Switzerland) partially supported the implementation of the neural net classification approach. Charles E. Mennell made the manuscript readable.

References

1. Murakami, Takabata and Seiyama, ‘‘Selective Detection of CO by SnO_2 Gas Sensor Using Periodic Temperature Change’’, *Proceedings of the 1987 Meeting*, IEEE - Transducers’87, 4th International Conference on Solid-State Sensors and Actuators, June 1987, pp. 618-21, Figaro #213 type CO Gas Sensor

2. H. D. Block, ‘‘The Perceptron: A Model for Brain Functioning’’, *Reviews of Modern Physics*, Vol. 34, No. 1, January 1962, pp. 123-35.

3. David E. Rumelhard, James L. McClelland, et al, *Parallel Distributed Processing,* The MIT Press, Cambridge MA, Vol. 1, 1988.

4. B. S. Hoffheins, R. J. Lauf, M. W. Siegel, ‘‘Intelligent Thick Film Gas Sensor’’, *Hybrid Circuits*, Vol. 14, September 1987, pp. 8-12, Invited by journal editor.

5. M. W. Siegel, R. J. Lauf, and B. S. Hoffheins, ‘‘Dual Gradient Thick-Film Metal Oxide Gas Sensors’’, *Proceedings of the Tokyo meeting*, Transducers ’87, June 1987, pp. 599-604.

6. B. S. Hoffheins, R. J. Lauf, M. W. Siegel, ‘‘Intelligent Thick Film Gas Sensor’’, *Proceedings of the Atlanta Meeting*, International Society for Hybrid Microelectronics, October 1986.

7. Frank Rosenblatt, *Principles of Neurodynamics,* Spartan Books, Washington DC, 1962.

8. Marvin Minsky and Seymour Pappert, *Perceptrons,* The MIT Press, Cambridge, MA, 1969.

Contents

Abstract 1

Introduction 2

Sensors 3

Measurements 3

Data 4

Algorithms 5

- Data Interpretation 7
  - Input Feature Selection 7
- Output Classification 9
- Training 9

Results 11

Conclusion 14

Acknowledgements 14

References 15

Illustrations

Figures

1 Metallization pattern for contacts between instrumentation and sensing layer 3
2 Block diagram of the gas delivery system 4
3 Block diagram of data acquisition and control system components 5
4 Typical responses of the sensor to ethanol, methanol and heptane vapors 6
5 Topology of three layered neural network 8

Tables

1 Performance under training by DIM (17 input units) 11

2 Performance under training by EFIM (16 input units) 12

3 Performance under training by DIM + EFIM (33 input units) 13