ABSTRACT
In this paper a control algorithm, that generates adequate machine settings for the Stencil Printing Process (SPP) in Surface Mount Technology manufacturing, is presented. The SPP is characterized by high process-noise levels, and by requiring constant solder-paste volume deposition at all times. Due to the prohibitive cost associated with extensive testing, only a limited amount of information can be used in the decision-making. The major constraint when controlling the SPP is that it is not possible to evaluate the output of the system an arbitrary number of times, since for each calculation the printing and inspection of at least one or even two new boards is required. These difficulties are addressed by the controller proposed in this paper, which is based on an algorithm for generating a sequence of iterative values that converge to an optimum set of machine parameters for a desired solder paste volume. The merit of the control is that it minimizes the variance and the steady state error of the weighted sample mean versus the desired height, which improves the quality of the process. In addition, it considers print direction and different component types independently. Finally, this controller also automatically corrects for the process errors associated with the loss of the solder-paste working-viscosity-point, which can occur due to unscheduled stops on the line, and for incorrect specification of machine parameters given by the operator.

Keywords: Stencil printing process control, least squares, affine estimator.

I. INTRODUCTION
Consider a process with control variable $c$ and with measurements $y$ of a quality characteristic of the process given by

$$y = F(c) + v$$

where $F$ is an unknown function of the control variable $c$ and $v$ is process noise. In general, the function $F$ may vary with time; that is, the process is time varying, due for example to changes in environmental conditions.

Given some desired value $H_d$ for the quality characteristic, the objective is to determine a value $C_d$ of the control variable $c$ so that $|F(C_d) - H_d|$ is suitably small. In many industrial processes, the value of the control variable $c$ (the machine setting) is determined by the equipment operator using past experience obtained from running the process. However, this procedure does not guarantee that the chosen value of $c$ will be a “good value,” or that the chosen value of $c$ will always work while the process is in operation. To achieve the best possible performance, there has been a considerable degree of interest in generating feedback control laws of the form $c = G(y)$ for some function $G$, so that $|F(G(y)) - H_d|$ is suitably small during process operation. If the function $F$ is known and has inverse $F^{-1}$, an obvious choice for the control variable is $c = F^{-1}(y)$. Of course, this solution blatantly neglects any kind of noise $v$ in the process described in (1).

Unfortunately, in many industrial processes, the function $F$ cannot be neither determined from a first-principles analysis nor identified without having to generate a large number of measurements for various values of the control variable $c$. A major difficulty in generating a large number of measurements is the cost of having to run the experiments to produce the measurements. Often, only a very limited number of measurements are available for determining the control parameter values.

In this paper, we generate some schemes for producing values of the control variable using search algorithms that require only a small number of measurements. This paper also explains how a weak-search algorithm can produce adequate control values for a specific kind of process. The merit of employing a weak-search algorithm is that it uses minimal information [4] regarding the nature of the process, allowing the algorithm to be applicable to a variety of problems [19].

The last part of the paper deals with a procedure for fine tuning the control variable when the quality measurements $y$ are in some desired performance band. The feasibility of the proposed control algorithm is supported by simulations and by experimental results.

The particular process under consideration in this paper is the SPP, which is characterized by having very significant amounts of noise and by requiring a bounded and smooth output at all times. Furthermore, due to the prohibitive cost associated with extensive testing, only the available (minimal) information [4] will be used for decision-making without trying to construct any concrete objective function. The solution for this kind of problem is called Optimization with Minimal Information (OMI) [4]. Another way that OMI can be interpreted in this case is, for example, when it is not possible to evaluate the output of the system an arbitrary number of times. In addition, the number of output evaluations for asymptotic convergence must be minimized and they should be constrained to a specific permissible operation range.
A defect that occurs in the earlier stages of the SMT manufacturing, like the SPP, will propagate causing overcosts of approximately 10 times for each additional step in the process that the PCB goes through without being detected as defective. Hence, this shows the importance of early detection of not only obvious printing errors (e.g. extreme lack or excess of solder paste in a solder brick) but also of possible causes of other defects due to degradation of the quality of the solder paste, loss of the working viscosity point or even machine related failures.

**Figure 1. Stencil Printing Process**

The SPP is illustrated in Figure 1. In step (a), a metallic stencil is placed over the PCB and solder paste is kneaded on one side of the stencil. In step (b), the squeegee is pushed over the stencil with a specific pressure and then it is moved from one side to the other of the stencil with a specific speed. This procedure makes the solder paste roll to fill the apertures in the stencil and finally the squeegee blade removes the excess of material. The final step before the components are placed over the solder bricks in step (c), is to separate the stencil from the PCB at a very slow rate; such rate is known as the snap-off speed.

**B. Problem Definition**

The SPP objective is to apply adequate amounts of solder paste over the PCB pads where components are going to be connected. The technique is rather simple; it uses a metallic stencil with apertures of different shapes and sizes that allow the solder paste to be deposited in the areas where it is required. The stencil has a specific height, and because the solder brick areas are given by the aperture size, a specific volume of the solder paste is deposited on each pad. In practice, because there are so many variables affecting the SPP, it is impossible to achieve a “perfect printing.”

General recommendations [10][13][16] are that the printing speed should be around 1-2 in/sec (2.54-5.08 cm/sec) and that the squeegee pressure should be about 1 lb/in (0.113...
kg/m) of linear squeegee length. These values can usually be adjusted iteratively by process engineers, but this method relies heavily on heuristics or experience. In order to improve the quality of the printing, some machine parameters can be adjusted in a more formal way by the use of a mathematical criterion. However, direct performance evaluation of the optimization based in such criteria is not readily achievable as a direct measurement of a physical variable but often as the form of DPMO in the final product; and even then, a quantitative evaluation can be hard to produce.

A basic assumption that is made here is that 3-D inspection capabilities are available. This is required because accurate and repeatable estimation of the height of the solder bricks is mandatory. With just basic 2-D inspection, it is only possible to make strong claims about the area of the solder bricks but not about their actual volume. This does not imply that 100% 3-D inspection should be implemented; however, it is preferred as a part of an integral strategy for SPP defect reduction.

C. Data Analysis

Through the analysis of the data collected from several experiments, it has been shown that the main factors to be considered by a SPP controller include:

1. Stencil printer
   - Speed
   - Pressure
   - Snap-off speed and distance
   - Print direction of the squeegee blade relative to the angle of the solder brick
   - Print direction (backward, forward)
   - Environmental conditions (temperature-humidity)

2. Solder paste
   - Type (3,4, etc)
   - Working viscosity point

3. Measurement Tool
   - Precision
   - Accuracy
   - Speed
   - 2D vs. 3D inspection

A very complete description of most of the factors present in the SPP can be found in [7].

Volume of the deposited solder paste is generally accepted as a measure of the quality of the process. However, a more strict analysis shows that even when the parameters of the Probability Density Function (PDF) of the volume seem much easier to identify, they are also hiding information about the underlying nature of the process. Most types of inspection equipment rely on indirect measurement of volume. They measure an approximation of the solder brick area and then calculate the mean of the height of that surface giving an indirect volume measurement. The present analysis will consider area and height of the solder bricks independently without using a direct volumetric consideration of the deposited paste.

III. MATHEMATICAL APPROACH

The process under consideration could be characterized as follows. When the \( n^{th} \) board is printed, this is a realization of \( Q \) Random Variables (RV) of unknown PDF, where \( Q \) is the total number of component pads on the board. The main issue here is that additional samples of the same RV’s cannot be obtained in a strict sense. When a new board is printed and then measured, a new set of samples is created; these cannot be considered as subsequent sample realizations of \( Q \) RVs associated with the \( n^{th} \) board because of the following reasons:

- The \( n+1 \) board produced is printed in an opposite direction and it has been shown from experimental data that this fact can cause a discrepancy even greater than 10% in the volume of the solder paste that is deposited.
- The \( n+2 \) board is different from the \( n^{th} \) board by the effects of the printing of the \( n+1 \) board; therefore, it does not match either the printing conditions or the measured volume of the solder paste that was deposited in the \( n^{th} \) board.

A simpler and better approximation is to consider the measurements of the solder bricks of the same component type as multiple sample realizations of the same RV. This approximation is valid because the conditions that were present for all solder bricks printed on the same board are practically the same, and even when their locations are different, as long as they are partitioned in an appropriate way, a conclusive analysis can be performed over that kind of data. Therefore, some other factors such as pad orientation, density, location and type need to be considered. In addition, by the Central Limit Theorem, it can be assumed that if the number of measurements (from RVs with unknown PDF) is large enough for a specific component pad type, then the sample PDF obtained will be approximately Gaussian.

Now a mathematically rigorous approach can be established to analyze the problem. Let the experiment be measuring the height of a solder paste brick on the \( n^{th} \) board. Let \( Q \) be the number of pads on the \( n^{th} \) board and \( z(n,i) \) be the measured value of the height of the solder paste brick on pad \( i \) of the \( n^{th} \) board. The goal is to have the heights of the solder paste bricks be as close as possible to a desired height denoted by \( H_d \). The Mean Square Error (MSE) [8] between the height \( H(n) \) and the desired height \( H_d \) for the \( n^{th} \) board is given by

\[
MSE(n) = \text{Var}(z(n,i)) + (E[H(n)] - H_d)^2.
\]

From (2) it can be seen that in order to minimize the MSE, it is necessary to minimize the variance and also the error term between the expected value of the heights of the \( n^{th} \) board \( E[H(n)] \) and the desired height.
The sample mean square error of the \( n \)th board, denoted by \( \text{SMSE}(n) \), is given by

\[
\text{SMSE}(n) = \frac{1}{Q} \sum_{i=1}^{Q} (z(n,i) - H_d)^2 ;
\]

note that if \( z(n,i) = H_d \) for all \( i \), then \( \text{SMSE}(n) = 0 \). Unfortunately, due to process noise, this idealized case is not applicable. The problem is to select the machine control variables to minimize \( \text{SMSE}(n) \). To pursue this, we first express \( \text{SMSE}(n) \) on the form

\[
\text{SMSE}(n) = \frac{1}{Q} \sum_{i=1}^{Q} (z(n,i) - \text{SM}(n))^2 + \left[ \frac{1}{Q} \sum_{i=1}^{Q} z(n,i) - H_d \right]^2 \quad (3)
\]

where \( \text{SM}(n) \) is the sample mean of the solder paste brick heights for the \( n \)th board and it is given by

\[
\text{SM}(n) = \frac{1}{Q} \sum_{i=1}^{Q} z(n,i) , \quad (4)
\]

and the sample variance of the solder paste brick heights for the \( n \)th board is given by

\[
\text{SV}(n) = \frac{1}{Q-1} \sum_{i=1}^{Q} (z(n,i) - \text{SM}(n))^2 . \quad (5)
\]

Note that the first term in (3) is approximately the sample variance of \( H(n) \) given in (5). Rearranging coefficients in (4) and (5) the \( \text{SMSE}(n) \) can be expressed as

\[
\text{SMSE}(n) = \frac{Q-1}{Q} \text{SV}(n) + (\text{SM}(n) - H_d)^2 . \quad (6)
\]

Actually, for large values of \( Q \) the term \( (Q-1)/Q \approx 1 \). That is in fact an analogous result to the case when the true mean and variance are used rather than the sampled ones. The \( \text{SMSE}(n) \) error term should be combined with the error associated with the sample variance and mean for a specific number of measurements.

The effect of control variable changes on the first term in \( \text{SMSE}(n) \) can be computed and can be taken into account in determining appropriate settings for the machine control variables. Additional constraints must be added to the previous algorithm to be able to meet the requirements of the process, specifically, the maximum and minimum values of all samples \( z(n,i) \).

In practical terms, a larger weight must be given to small, fine-pitch or problematic pads so a deviation from its nominal value is heavily penalized. Now it is necessary to include the definitions of weighted (W) sample mean, variance and MSE

\[
\text{SM}_w(n) = \frac{1}{Q} \sum_{i=1}^{Q} w(z(n,i))
\]

\[
\text{SV}_w(n) = \frac{1}{Q-1} \sum_{i=1}^{Q} w(z(n,i) - \text{SM}(n))^2
\]

\[
\text{SMSE}_w(n) = \frac{Q-1}{Q} \text{SV}_w(n) + (\text{SM}_w(n) - H_d)^2 .
\]

Furthermore, the sample variance of the sample mean \( \sigma_{\text{SM}(n)}^2 \) will be determined by the number of samples used to calculate \( \text{SM}(n) \) [15]. This is proved as follows:

\[
\sigma_{\text{SM}(n)}^2 = \frac{1}{Q} \sum_{i=1}^{Q} \sigma^2(n)
\]

\[
E[\text{SV}(n)] = \sigma^2(n)
\]

If

\[
\sigma^2(n) \equiv \sigma^2 \quad \text{for all } n
\]

then

\[
\sigma_{\text{SM}}^2 = \frac{\sigma^2}{Q} . \quad (8)
\]

Equation (8) implies that if the variance of the height measurements for each board is approximately 0.4 mils (10.16µm), then the expected variation of the sample mean over several boards will be that value divided by the number of measurements taken from that board. This demonstrates that not even in the ideal case, with uncorrelated samples, will the variation of the process be zero. However, this was assuming that all of the heights of the solder bricks were independent from each other; in reality, this is not true because they were printed on the same machine and with the same squeegees; thus some characteristics of the solder paste are quite similar from one board to the other.

Based on experimental results it was observed that the sample mean variation is bounded by its mean plus and minus half of the sample variance of an individual board. This means that for the 0.4 mils (10.16µm) board variance case, the mean will be confined to the interval \( \pm 0.2 \) mils \((\pm 5.08\mu m)\) from the mean of the mean over all boards.
The following results are based on Monte Carlo simulations but they also can be obtained analytically.

**Algorithm 1**
1. Define \( \Delta \) to be the minimum change in the control parameter that can produce a detectable change in the output variable.
2. Inspect initial board.
3. Check if the sample of the height is above or below the desired target. This is the search direction.
4. Change the control value by \( \Delta \) in the search direction.
5. Print a new board.
6. Find the slope of the output function.
7. Determine if the target has been reached; if the target height can be attained then exit and switch to the affine estimator.
8. Determine if the target cannot be reached; if the desired height is impossible to reach then exit and use the best approximation given by the quadratic estimator.
9. Go to 4.

It should be noted that there is an independent controller, which performs the optimization procedures, for each different direction. This is valid not only for the global but also for the local optimization algorithm.

**V. STEADY STATE OPTIMIZATION**

Once a good approximation of the operational band has been reached, in steady state optimization it is necessary to fine-tune the control values generated from Algorithm 1. At this point, a Local Affine Least Squares (LALS) algorithm takes control of the process and performs fine-tuning of the control variable. The LALS is inspired from the LLLS [18][12][20] algorithms but the implementation is completely different in this case; the global search is performed using Algorithm 1 and the Kalman Filter recursion from [18] is replaced for a modified block form of the LS algorithm. A LALS estimator is only applied locally because of the high levels of noise that make the use of LS prohibitive on a global scale without having very biased and/or unbounded estimations. However, it is necessary to keep the values of the two final iterations of Algorithm 1 to maintain the slope direction estimation; this is required because extreme-random noisy measurements can invert the direction of the slope.

An affine estimator is used because a linear one has very poor performance due to the nature of the process. A quadratic estimator will be used in the instances when a minimum or maximum of the function is reached in the search without attaining the desired operational band. An important remark is that during the transition between the two modes of operation, to have continuity in the control values, the local algorithm uses previous estimations generated with Algorithm 1 during the interface stage. This stage is generally about the same length of the window for the final block form of the LS estimator. To guarantee a smooth output, an intermediate step is necessary for enabling to switch between Algorithm 1 and the LALS algorithm.

**IV. GLOBAL OPTIMIZATION**

In this paper, a global search algorithm is proposed to achieve the operational band in the SPP. It uses a fixed step length and it can have two different kinds of outcomes depending on whether the desired height can be reached or not. The details of this algorithm are presented in [3]. The general outline of this algorithm is stated as follows:

![Figure 2. Convergence of the sample mean](image)

**Figure 2. Convergence of the sample mean**

![Figure 3. Convergence of the sample variance](image)

**Figure 3. Convergence of the sample variance**

The height measurement can be approximated as a Gaussian distribution; the plots for sample mean and variances of the convergence for a given number of samples are shown in Figure 2 and Figure 3. Figure 2 implies that, for example, on average, for a 5% error in the sample mean, over 250 measurements are required, but for only 10% accuracy, less than 60 will suffice. This is the mean for the worse case scenario when there is no correlation among samples; therefore, for the SPP the results will be more optimistic. In the case of the sample variance in Figure 3, the results are slightly better; for a 5% error in the sample variance, less than 120 measurements are required, and for only 10% accuracy, less than 42 will be enough.
A. Affine Regression
Once the operational band had been obtained by Algorithm 1, a LS algorithm is used to estimate the parameters needed in order to apply the affine estimator. The classic least squares solution \([8][5][1]\) for the problem \(A\theta = z\) is given by

\[
\hat{\theta} = \left[ \hat{\theta}_0 \right. \left. \hat{\theta}_1 \right] = \left( A^T A \right)^{-1} A^T z, \tag{9}
\]

where \(\hat{\theta}\) are the estimated parameters.

For an affine estimator, a simple closed-form solution can be found for the parameters as function of the previous measurements and control values. The estimator is given by

\[
H_z = \hat{\theta}_0 + \hat{\theta}_1 c(n+1), \tag{10}
\]

where \(c(n+1)\) is the control value for the next board in order to achieve an optimum value for the height in the least squares sense. Note, however, that the affine estimator is only applicable when the desired height is achievable given the range of variation of the control value and its effect on the height. However, if an unrealistic value for the desired height is specified, then a quadratic estimator should be used to provide the solution that minimizes the error even when the target operational point is not attainable.

B. Quadratic Regression
In this case, the parameter estimation can be done either by direct matrix inversion as in (9), or by QR decomposition as in (11) for computational efficiency purposes \([5]\). If \(Q\) and \(R\) correspond to the QR decomposition of \(A\), then

\[
A = QR
\]

\[
\hat{\theta} = \left[ \hat{\theta}_0 \right. \left. \hat{\theta}_1 \right] = R^{-1}Q^T z. \tag{11}
\]

Notice that the inverse of \(R\) is easily calculated by using, for example, row back-substitution, because \(R\) is an upper-triangular matrix. Once \(\hat{\theta}\) is known, then the quadratic estimator of the signal can yield the control value required to produce the closest possible value to the desired height using

\[
H_z = \theta_0 + \theta_1 c(n) + \theta_2 c(n)^2. \tag{12}
\]

Note that from the difference between the estimated and the actual sample mean height, an estimate of the variance of the process noise can be obtained if it is required. When the local optimization algorithm is running and there are not large changes in the system response, it is not necessary to go back to the global search algorithm and its thresholding. However, if the process is restarted, to avoid out-of-bounds control parameters, it is necessary to go back to Algorithm 1 to find the correct operational band.

VI. RESULTS
To verify the validity of the control scheme two kinds of results are provided: simulation and real experimental data. Because of the elevated cost of extensive online testing, the controller was initially tested and fine-tuned in a simulator. For such tests, values of variances and means obtained from real board runs were used. All height values referred in this section are sample means \(SM(n)\) of all solder bricks inspected on each board and they are calculated as in (4). As performance comparison, the \(SMSE(n)\) defined in (6) was used.

A. Simulation
The following graphics show the evolution of the process in terms of the iterations; random samples were generated based on the PDF of the height and then used to calculate the control values using the global and local optimization algorithms.

Figure 4. Simulated process output

Figure 5. Simulated controller Output
In Figure 5 the complete output of the controller is shown in a simulation where for a constant pressure of 12 lb/in² (0.07 kg/cm²) applied with a 12 inch squeegee blade, the initial value of the speed was set at 0.5 in/sec (12.7 mm/sec). These are furthermore the conditions used during the experiments.

The controller initially uses the global search algorithm and after it reaches the operational band, the LS estimation algorithm provides the affine estimator with appropriate values for the printing speed.

In addition, in Figure 6, the case when a large disturbance is introduced was considered. As a result, the controller adapted its output value to a near optimal value in approximately 3 iterations. Notice that the previous simulations only show the output of the controller and the process in one direction. Hence, the real process will take twice as many boards to converge.

**B. Experimental Results**

To show the effectiveness of the controller and to demonstrate how it can help to recover even from incorrect machine settings programmed by an operator the following experiment was designed. Using a 5 mil (0.127mm) laser-cut stencil, a 12 in (30.48cm) metallic squeegee and solder paste Type IV, the experiment was run until steady state values in the control parameter were reached.

With a desired height $H_d = 5.5$ mil (0.14mm), the controller was set to run at a constant pressure of 12 lb/in² (0.07 kg/cm²) and using speed as control variable. The initial value for the speed was set to 3 in/sec (7.62cm/sec); obviously this value of speed is too high for the selected value of pressure if a consistent and accurate height needs to be achieved.

The objective of the controller is to adjust the printing speed independently for each direction until it reaches the optimum value in the least squares sense for an affine estimator. In order for the measurements to be statistically significant for each board, about 700 solder bricks of 8 different component types were inspected. In case that different weights need to be assigned to each specific pad type to give priority to problematic solder bricks, (7) can be used instead of (4), (5) and (6).

The sample measurements were taken using a 3D laser inspection system with a measured total error of less than 0.1% with respect to the mean height, which indicates that the measurement error was neglectable.

Figure 7 shows the evolution of the sample mean of the height for each direction given the speed values generated by the controller as is shown in Figure 8. In contrast to standard practice, the suggested controller produces steady state values for the speed that are different for each direction; this is consistent with the height distributions shown in Figure 10.
The possible limits for the desired height can be obtained from the sampled input-output relationship in Figure 9. Thus, it can be inferred that the attainable values for $H_d$ are between about 5.3 mil and 6 mil (134.62µm and 152.4µm).

![Process Output Diagram]

**Figure 9. Relationship of Height vs. Speed**

In order to verify that the value of speed produced by the controller is optimal, it is necessary to compare the sample probability distribution of the height with the desired height using for this all boards produced under the local optimization algorithm as shown in Figure 10.

![Output Histogram Diagram]

**Figure 10. Output histogram**

VII. CONCLUSIONS

In this paper, a general control algorithm is proposed. With the use of this algorithm operating not only in simulations but also in real experiments, it was shown that it is possible to control the SPP, to obtain a desired sample mean of the height, and therefore to control the solder paste deposition volume. It has also been found that it is possible to compensate for differences on height on backward and forward direction by using independent values of print speed for each direction. The ability of the controller to recover a process from abnormal equipment set points is potentially useful in the manufacturing industry.

REFERENCES


