

Statistically Enhanced Analogue and Mixed-Signal Design and Test

Pedro L. Ramos

José Machado da Silva

Diogo R. Ferreira

Marcelino B. Santos

FEUP

Universidade do Porto
Porto, Portugal

pedro.lima.amos@outlook.com

INESC TEC, FEUP

Universidade do Porto
Porto, Portugal

jms@fe.up.pt

INESC-ID, IST

Universidade de Lisboa
Lisboa, Portugal

diogomiguel0@hotmail.com

INESC-ID, IST

Universidade de Lisboa and SiliconGate
Lisboa, Portugal

marcelino.santos@ist.utl.pt

Abstract—The design, manufacture and operational characteristics (e.g., yield, performance, and reliability) of modern electronic integrated systems exhibit extreme levels of complexity that cannot be easily modelled or predicted. Different mathematical methodologies have been explored to address this issue. Monte Carlo simulation is the most widely employed and straightforward approach to evaluate the circuits' performance statistics. However, the high number of trial cases and the long simulations times required to obtain results for complex circuits with a ppm resolution, lead to very long analysis times. The present work addresses the evaluation of alternative statistical inference methodologies which allow obtaining similar results departing from a smaller dimension data set of Monte Carlo simulations from which the overall population is estimated. These methodologies include the use of Bayesian inference, Expectation-Minimization, and Kolmogorov-Smirnov tests. Results are presented which show the validity of these approaches.

I. INTRODUCTION

Within new and future nano-scale technologies, the ability to control critical device parameters is becoming increasingly difficult and the effects of significant process variations are inevitable [1]. The prohibitively high cost associated with pre-silicon validation and post-silicon tuning of current analogue and mixed-signal (AMS) circuits is a growing problem as geometries continue to shrink and the relative magnitude of critical process fluctuations continues to grow. These circuits tend to become large-scale complex systems that must be tuned and adaptively adjusted over time. Hence, there is an immediate need to develop new methodologies and tools to support the validation and tuning of nano-scale AMS circuits for future technology generations.

Most AMS circuits' performance metrics are extremely sensitive to inter-die and/or intra-die variations associated with today's manufacturing processes. For this reason, AMS circuits are not as scalable as digital circuits and they have been considered as major bottlenecks for future IC technology scaling [2]. On the other hand, with such large variations testing issues related to yield loss and test escapes tend to be aggravated [1]. Each AMS circuit becomes now a larger and complex system that can adaptively vary over time namely by means of digitally assisted schemes to eliminate defective

redundant parts or to re-adjust the operating conditions in order to increase production yield and/or operating performance. It, in turn, brings up enormous fundamental technical problems related to computer-aided design of tunable AMS circuits in terms of: (1) time needed to perform a large number of simulations to optimally design the circuit; (2) accuracy to find the best parameters and control nodes to be used for observation and control purposes in the testing and tuning schemes to be adopted; (3) select the best parameters to be adopted in the test production stages in order to maximize manufacturing yield and minimize test escapes.

Different mathematical methodologies have been explored to address these issues. In the most recent approaches, statistical modelling has been explored where an initial sample of the circuits behaviour under parametric variations is obtained by Monte Carlo (MC) simulation, after which a statistical model is formulated to estimate a single joint probability density function (pdf) for all performance parameters being specified. Methodologies based on the multivariate normal law, Gaussian copulas [3], kernel density estimation [4], and univariate extreme value statistics [5] were proposed. Nevertheless, these methods are sensitive to the dimension of the initial sample and still require performing a large number of simulations of a model obtained from the preliminary sample of the circuit performance.

The major drawback to MC analysis is simulation time. A large number of trials is needed, particularly if one needs to accurately evaluate the tails of the distribution. While this is not an issue for small circuits or individual circuit blocks found after partitioning, it can represent a significant burden in design and test development time for large circuits.

The objective of the preliminary work presented here is to evaluate alternative statistical methods to reduce the time and effort to estimate performance distribution and parametric yield for pre-silicon verification. Among these alternatives is the use of an approach based on Bayesian inference and Expectation-Maximization (EM) algorithms. The key idea is to start from a set of early-stage data to first learn the prior knowledge and statistically encode it as a prior distribution. The EM algorithm allows obtaining the response surfaces of both mean and variance of a process even when departing from

incomplete data. Next section presents the theoretical concepts behind this approach and section III presents preliminary results that confirm its validity. Section IV highlights the main conclusions.

II. THE EXPECTATION-MAXIMIZATION ALGORITHM

Knowing the expected probability distribution of a circuits performance is equivalent to finding the Maximum-Likelihood (ML) estimate of the parameters of an underlying distribution from a given data set, when the available data is incomplete or has missing values. The same approach can be applied to estimating the performance distribution of the fabricated circuits departing from a limited in size sample of fabricated circuits.

The EM algorithm [6], [7] allows finding the ML estimate of the parameters of a distribution x after a given (observed) data sample of that population. It can be said that data x comprises both incomplete observed data y and unobserved data z , i. e., $x = (y^T; z^T)^T$.

In a ML estimation approach one has a likelihood function and wants to find the parameter Θ that maximizes the probability of having obtained the observed data, i. e., that governs the probability density function (pdf). Θ is found by maximizing $L(\Theta) = p(\mathbf{y}|\Theta) = \prod_{i=1}^N p(y_i|\Theta)$, given the observed sample $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$ of that distribution. Within the EM approach one starts by finding an estimate for the likelihood function followed by maximizing the whole term, where each entry of z is as a random variable once it was not actually observed.

The expectation of the whole likelihood function, given the observed data \mathbf{y} is $\mathbb{E}[\log L(\Theta)|\mathbf{y}, \Theta^{(i)}]$. This expectation is maximized calculating $\Theta^{(i+1)} = \underset{\Theta}{\operatorname{argmax}} \mathbb{E}[\log L(\Theta)|\mathbf{y}, \Theta^{(i)}]$.

Let's suppose one intends to estimate parameter θ of population $X \sim \text{Normal}(\theta, \sigma^2)$. It is stipulated or believed true that $\sigma^2 = a$. Likewise, there is evidence that the most expected value of parameter θ is b , with variance d^2 . It is then credible to assume that *a priori* $\theta \sim \text{Normal}(b, d^2)$. Taking $\mathbf{y} = [y_1, \dots, y_n]$ as a particular case of a random i.i.d. sample $[X_1, \dots, X_n]$, then,

$$\begin{aligned} L(\theta, a, x) &= f_X(x | \theta, \sigma^2 = a) \\ &= P(X = x_1, \dots, X = x_n | \theta, \sigma^2 = a) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sqrt{a}} \exp\left[-\frac{1}{2} \frac{(x_i - \theta)^2}{a}\right] \\ &= (2\pi a)^{-\frac{n}{2}} \exp\left[-\frac{1}{2a} \sum_{i=1}^n (x_i - \theta)^2\right] \end{aligned}$$

On the other hand, using the Bayes rule, with Θ equal to the set of all values that θ can assume,

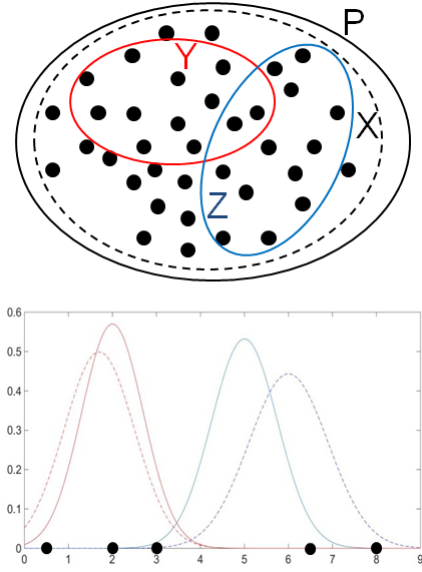


Fig. 1: Distribution of the x values over the two normal pdf.

$$\begin{aligned} P(\theta | x) &= \frac{P(X = x_1, \dots, X = x_n | \theta, \sigma^2 = a) p(\theta)}{\int_{\Theta} P(X = x_1, \dots, X = x_n | \theta, \sigma^2 = a) p(\theta) d\theta} \\ &= \frac{(2\pi a)^{-\frac{n}{2}} (2\pi d)^{-\frac{1}{2}} \exp\left[-\frac{1}{2a} \sum_{i=1}^n (x_i - \theta)^2\right]}{(2\pi a)^{-\frac{n}{2}} (2\pi d)^{-\frac{1}{2}} \int_{\Theta} \exp\left[-\frac{1}{2a} \sum_{i=1}^n (x_i - \theta)^2\right]} \\ &\quad \times \frac{\exp\left[-\frac{1}{2d^2} (\theta - c)^2\right]}{\exp\left[-\frac{1}{2d^2} (\theta - c)^2\right]} d\theta = f_Z(\theta) \end{aligned}$$

with $Z \sim \text{Normal}(\frac{cb+n\psi\bar{x}}{c+n\psi}, \frac{1}{c+n\psi})$, $c = \frac{1}{d^2}$ e $\psi = \frac{1}{a}$. Then, $\hat{\theta} = E[\theta | x] = \frac{cb+n\psi\bar{x}}{c+n\psi}$, with $c = \frac{1}{d^2}$ e $\psi = \frac{1}{a}$.

Next, $P(\theta | x)$ is used as the new *a priori* distribution for θ , being the process repeated until the values for $\hat{\theta}$ converge.

The key idea is to start from a set of early-stage data to first learn the prior knowledge and statistically encode it as a prior distribution. Even if MC simulations are used to obtain a relatively small set of samples that will act as iteration 0.

A. Modelling of a population after fusion of Gaussian sub-sets

The EM approach can be used to model an unknown population after fusing Gaussian sub-sets of that population [8]. An unknown population P can be represented from sampled sub-sets Y and Z in the form $P = w_1 Y + w_2 Z$, with $Y \sim \text{Normal}(\theta_1, \sigma_1^2)$, $Z \sim \text{Normal}(\theta_2, \sigma_2^2)$, where coefficients w_1 and w_2 are real numbers. The EM algorithm allows estimating these coefficients after using iteratively Bayesian statistics. The population sub-sets Y and Z can actually be obtained from a larger sample X of P .

To illustrate this, consider the universe P from which a sample comprised by the set of values X was obtained, as shown in figure 1. From the X sample two sub-sets Y and Z , both with normal pdf, were extracted. Given a value x_i , one

can calculate the probabilities of this value belonging to Y , p_{Y_i} , and Z , $p_{Z_i} = 1 - p_{Y_i}$:

$$p_{Y_i} = P(p_Y|x_i) = \frac{P(x_i|Y)P(Y)}{P(x_i|Y)P(Y) + P(x_i|Z)P(Z)}$$

$$\text{with } P(x_i|Y) = \frac{1}{\sigma_Y\sqrt{2\pi}} \exp\left(-\frac{(x_i-\theta_Y)^2}{2\sigma_Y^2}\right) \quad (1)$$

If $p_{Y_i} > p_{Z_i}$, then x_i is initially allocated to the Y pdf, and to Z otherwise. This procedure is repeated for all the other points, being in the end two new Y and Z clusters obtained (continuous lines in figure 1). Next, new mean and variance values are calculated for the two normal distributions (equations 2).

$$\theta_y = \frac{y_1x_1 + y_2x_2 + \dots + y_nx_n}{y_1 + y_2 + \dots + y_n};$$

$$\sigma_y^2 = \frac{y_1(x_1 - \theta_y)^2 + \dots + y_n(x_n - \theta_y)^2}{y_1 + y_2 + \dots + y_n};$$

$$\theta_z = \frac{z_1x_1 + z_2x_2 + \dots + z_nx_n}{z_1 + z_2 + \dots + z_n};$$

$$\sigma_z^2 = \frac{z_1(x_1 - \theta_z)^2 + \dots + z_n(x_n - \theta_z)^2}{z_1 + z_2 + \dots + z_n} \quad (2)$$

The process is repeated until convergence is found. In the end, the P pdf is expressed as $P = p_Y N(\mu_Y; \sigma_Y) + p_Z N(\mu_Z; \sigma_Z)$. It is also possible to implement the process using more than two sub-sets.

B. Kolmogorov-Smirnov test

Once the parameters of the unknown population have been estimated, the Kolmogorov-Smirnov (KS) test can be applied to evaluate whether a sample of dimension n is representative of a population X with a (*fdp*) $Normal(\mu, \sigma^2)$. For that, the following hypothesis can be formulated:

$$H_0: F_X(x) = F_Y;$$

$$H_1: F_X \neq F_Y,$$

where $F_X(x)$ is the accumulated distribution function (*fda*) of X and F_Y that of a random variable $Y \sim Normal(\mu, \sigma^2)$. If the obtained KS test value is lower than p , then the probability of rejecting H_0 , being that the true hypothesis, is lower than p , i. e., $P(Rejecting H_0 | H_0 \text{ is true}) < p$; as such, the null hypothesis is rejected; otherwise H_0 is not rejected. Iteratively, one can find the (*fdp*) of which x comes from, or of which it represents an occurrence.

III. RESULTS

The EM based performance variation estimation approach presented above was evaluated considering the simulation of two particular circuits, a cascode current mirror and a low dropout voltage (LDO) regulator.

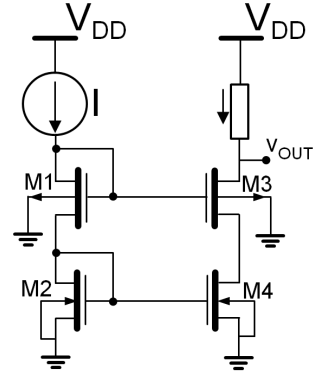


Fig. 2: Cascode current mirror circuit used as an example.

A. Cascode current mirror

Figure 2 shows a cascode current mirror considered in the first evaluation case. The circuit was simulated applying a DC current I superimposed with an AC component. MC simulations with different number of cases were carried out for 10% Gaussian random variations at 3σ of the transistors' widths, and the rms value of the output voltage V_{out} was taken for each case. An MC population with 10^5 cases taken as reference, presents $V_{out} \sim Normal(\theta = 1.0924V, \sigma^2 = 23.14 mV)$.

Figures 3a and 3b show the evolution of $E(V_{out})$ and $Std(V_{out})$, for the different number of simulation cases. It can be seen that with the 4050 cases MC simulation, the average value obtained is very close to the expected value of 1.0924 V. Also, although the average value obtained with EM is smaller, nevertheless, the probability distribution curve (figure 3c) is closer to that given by the 10^5 MC simulation than the one provided by the 4050 cases.

Table I shows the evolution of $E[V_{out}]$ and $Std[V_{out}]$ when using the EM algorithm with 100 iterations for different number N of samples (sub-sets, each with 300 elements), for the 4050 cases data set. These values show that the EM algorithm provides estimates that closely approach the true values $\theta = 1.0924 V$ and $\sigma = 23.14 V$, as N increases.

It was then estimated that, using the EM algorithm applied to the sample obtained from the 4050 trials MC simulation, the overall population (here characterized by the 10^5 MC simulation) presents $E[X] = 1.09265 V$ and $Std[X] = 0.02226 V$. To evaluate whether it is likely that X follows a normal distribution with the parameters provided by the EM algorithm applied to the 4050 MC simulation data set, the KS test was applied, i. e.,

$$H_0: F_X(x) = F_Y;$$

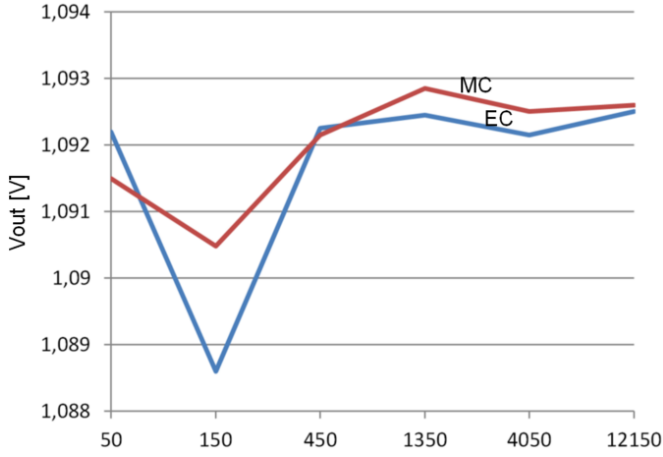
$$H_1: F_X \neq F_Y,$$

where $F_X(x)$ is the accumulated distribution function (*fda*) of X and F_Y the (*fda*) of a random variable $Y \sim Normal(1.0925, 0.02226^2)$. As $p\text{-valor} = 0.3174$ the null hypothesis is credible, what confirms the robustness of the EM algorithm.

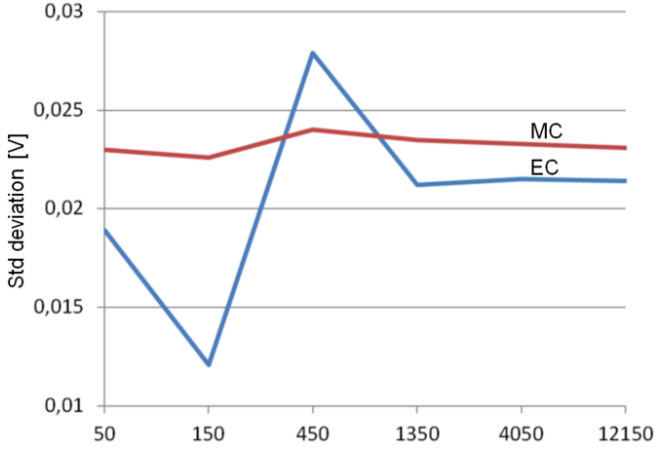
Regarding simulation time, while the 4050 MC simulation takes 190 seconds cpu time, the 10^5 MC simulation takes 50

TABLE I: Estimates of the output voltage average and standard deviation provided by EM for the cascoded current mirror.

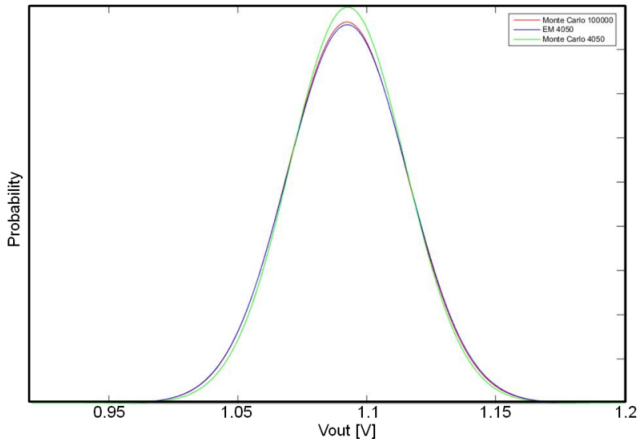
N	10	20	50	100	200	300	500	600
$E[X]$	1.08363	1.09423	1.09319	1.09317	1.09275	1.09270	1.09269	1.0926
$Std[X]$	0,01261	0,01547	0,01599	0,01567	0,01544	0,02092	0,02073	0,02226



(a) Average values.



(b) Standard deviation values.



(c) Probability distribution curves.

Fig. 3: Comparison between values given by MC simulations of the cascoded current mirror with 10^5 and 4050 trials, and those estimated with EM.

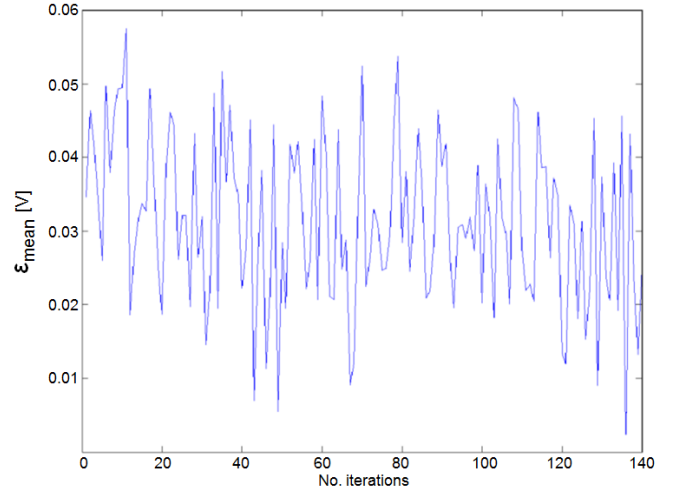


Fig. 4: Evolution of the average values provided by EM with the number of iterations.

times more. The execution time to run the EM algorithm is 3.8 seconds.

B. Low dropout voltage regulator

The second evaluation case concerns an ultra low-power LDO voltage regulator. A 10^4 MC simulation of the circuit provides an output voltage with average value and variance $v_{OUT_X} \sim Normal(1.152, 0.02357^2)$ V. The simulation time is 11784 seconds. Two MC simulations with 200 cases each were carried out that present, respectively, $v_{OUT_Y} \sim Normal(1.156, 0.02295^2)$ V and $v_{OUT_Z} \sim Normal(1.150, 0.02144^2)$ V, and take 179,6 seconds.

The EM algorithm was applied to the two sets of data obtained with the 200 cases MC simulations, being then the average of each set expressed as $\bar{Y} = w_{y1}\bar{Y}_1 + w_{y2}\bar{Y}_2$ and $\bar{Z} = w_{z1}\bar{Z}_1 + w_{z2}\bar{Z}_2$. In this case, the EM execution time is 1.43 seconds for each set Y and Z .

The value estimated for the output voltage $E[X]$, after combining the four averages and coefficients, $E[X] = (w_{y1}\bar{Y}_1 + w_{y2}\bar{Y}_2 + w_{z1}\bar{Z}_1 + w_{z2}\bar{Z}_2)/2$, obtained using EM after 140 iterations is $E[X] = 1.1532$ V, i. e., an approach with a relative error of $\epsilon_{rel} = 0.1\%$. It can be seen that the EM estimated value provides a very small error, which is also smaller than those given by the 200 cases MC simulations, which present errors of $\epsilon_{rel} = 0.35\%$ and $\epsilon_{rel} = -0.17\%$, respectively. After 1300 iterations $E[X]$ is almost the same. In fact, as shown in figure 4 the average values provided by EM after 140 iterations are very close to the expected reference value.

Regarding simulation time, instead of 3 hours and 16 minutes required for the 10^4 MC simulation, a total of about

6 minutes is needed to run the two 200 cases MC simulation and execute EM.

IV. CONCLUSIONS

The use of the Expectation-Maximization algorithm has been explored as a means to, in a more expedite form, find the distribution of a circuits performance variation, either at the design stage or in production after fabrication. In the case of pre-silicon verification, this allows us to reduce the number of Monte Carlo simulations to a smaller dimension data set from which the overall population is estimated. In the post-silicon case, yield can be estimated after fabrication of a smaller number of circuits.

Two circuit examples are presented for which it is shown that significant simulation time can be saved when a large number of Monte Carlo simulations is replaced for a seed, lower dimension, Monte Carlo simulation followed by the application of the Expectation-Maximization algorithm. In one of the evaluation cases a gain of execution time of about 50 with an estimation error of 0.02% for the average value and 3.8% for the standard deviation is obtained, and for the other an execution time gain of 35 is obtained with an estimation error of 0.1%.

This approach can also be used to estimate test tolerance bands and to identify adjustment ranges for controlling variables of tuning schemes.

ACKNOWLEDGEMENTS

This work is financed by the by FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology), through the ERDF - European Regional Development Fund, COMPETE Programme (operational programme for competitiveness), and national funds within project CMUP-EPB/TIC/0038/2013, SEAD (Statistically Enhanced Mixed-Signal and Analog Design), in the framework of the CMU-Portugal Early Bird Projects program.

REFERENCES

- [1] "Yield enhancement," International Technology Roadmap for Semiconductors, ITRS, Tech. Rep., 2013.
- [2] X. Jiang, *Digitally-Assisted Analog and Analog-Assisted Digital IC Design*. Cambridge University Press, 2015. [Online]. Available: <https://books.google.pt/books?id=wlcZCgAAQBAJ>
- [3] A. Bounceur, S. Mir, and H. G. Stratigopoulos, "Estimation of analog parametric test metrics using copulas," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 30, no. 9, pp. 1400–1410, Sept 2011.
- [4] H. G. Stratigopoulos, J. Tongbong, and S. Mir, "A general method to evaluate rf bist techniques based on non-parametric density estimation," in *2008 Design, Automation and Test in Europe*, March 2008, pp. 68–73.
- [5] H. G. Stratigopoulos, "Test metrics model for analog test development," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 31, no. 7, pp. 116–1128, July 2012.
- [6] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.
- [7] C. F. J. Wu, "On the convergence properties of the EM algorithm," *The Annals of Statistics*, vol. 11, no. 1, pp. 95–103, March 1983.
- [8] T. Kimura, T. Tokuda, Y. Nakada, T. Nokajima, T. Matsumoto, and A. Doucet, "Expectation-maximization algorithms for inference in dirichlet processes mixture," *Pattern Analysis and Applications*, vol. 16, no. 1, pp. 55–67, 2013. [Online]. Available: <http://dx.doi.org/10.1007/s10044-011-0256-4>