Production quality performance of manufacturing systems with in-line product traceability and rework

Colledani M, Angius A

This is a post-peer-review, pre-copyedit version of an article published in CIRP Annals. The final authenticated version is available online at: http://dx.doi.org/10.1016/j.cirp.2020.04.018

Production quality performance of manufacturing systems with in-line product traceability and rework

Marcello Colledani[1](2), Alessio Angius[1]

1 Politecnico di Milano, Department of Mechanical Engineering, Via la Masa, 1, 20156, Milan, Italy.

In-line product inspection and identification are emerging technologies supporting the observability and traceability of the product state evolution in multi-stage manufacturing systems. These technologies enable the dynamic implementation of defect management actions, such as product rework. However, their impact on the system quality and production logistics performance needs to be assessed in order to justify the related investments. This paper presents a theory and a methodology to predict the effective throughput of manufacturing systems with product traceability and rework, jointly considering the product and system state dynamics. The industrial benefits are validated in a real system in the semiconductor industry.

Quality, Manufacturing system; Modelling

1. Introduction, motivation and objectives

Modern manufacturing systems are endowed with on-line product inspection and identification technologies [1], enabling to observe the product state evolution along the stages of a process-chain. These technologies include both optical measuring devices, to gather information on relevant dimensional and geometrical product features, and product tagging devices, e.g. RFID, to trace the location of the product in the system, in specific time instants. The gathered data, stored in data management systems, support the timely identification of emerging defects [2] and the subsequent introduction of defect management strategies, such as product repair or rework operations, affecting both the quality and the part flow logistics, thus targeting the achievement of desired *production quality* performance in the system [3].

In spite of the emerging interest for these technologies within the Industry 4.0 paradigm, traditional manufacturing system modeling environments fail in capturing the joint dynamics of the product and system state evolution. This prevents from predicting the impact of advanced quality and production flow control strategies on the production quality performance of the system. Therefore, performing cost-benefit analysis to justify the related investment remains a challenge. According to a recent survey, manufacturing firms indicate difficulties in assessing the performance and the potential business return of such data gathering technologies and this currently constitutes the most important barrier to their industrial uptake [4].

For example, in [5] a modeling approach to analyze the production logistics performance of a system operating under autonomous production flow control policies was proposed. The implication on product quality was not considered. In [6] an integrated quality, maintenance, and production logistics model to predict the production quality performance of manufacturing systems was proposed. In [7] a production line model was proposed to estimate the production quality performance of systems processing deteriorating products. In both works, the application of product rework strategies was neglected. While all these works are important to shed light on the problem, the underlying models usually ground on a detailed description of the system state dynamics while only aggregated statistics on the product state, such as defect rates, are considered. However, as products characteristics get more complex [8] and the correlation among process stages generates more significant inter-stage product variation patterns [9], such settings are not sufficient to capture the propagation of defects in the system. As a matter of fact, a methodology to support the analysis and design of manufacturing systems with on-line product identification and rework has never been proposed, directly mining the performance of industrial systems operating under these conditions. Important questions like ``What is the impact of different rework strategies on the production rate of conforming products?'' remain unsolved, resulting in sub-performing system configurations.

To overcome these limitations, in this paper a general model of manufacturing systems with rework, integrating the dynamics of the product and system state evolution, is proposed for the first time to predict the throughput and the lead time distribution of conforming products in these systems. The approach makes it possible to design improved rework strategies in order to achieve desired production quality performance levels in real systems.

2. System description

The considered system is formed by K manufacturing stages and $K-1$ buffers of finite capacity, N_k, configured in serial layout (Fig. 1). Stages are denoted as M_k, with $k = 1, \ldots, K$, and buffers are denoted as B_k, with $k = 1, \ldots, K - 1$. At each stage k, *Identification and Actuation Devices (IADs)*, A_k, i.e. handling robots or actuating valves, can be present to inspect and identify the product, detect its state and control the flow of parts in the system, according to specific criteria and policies. IADs can move parts upstream in the line, in case the product needs rework, or outside the system, in case of scrap. Additional buffers B_k, $k=K,\ldots,K+K_a-1$, with finite capacity N_k, are present outside the main line to store parts that are manipulated by the IADs.

2.1. Single stage dynamics

The dynamics of each manufacturing stage is modeled by a discrete-time and discrete-state Markov chain of general complexity. Specifically, each stage M_k is represented by S_k states, and thus the state indicator α_k assumes values in $\{1,\dots,S_k\}$. The set containing all the states of M_k is called S_k. The dynamics of each stage in visiting its states is captured by the transition probability matrix $\boldsymbol{\lambda_k}$, that is a square matrix of size S_k. Moreover, a quantity reward vector $\boldsymbol{\mu_k}$ is considered, with S_k entries. While in the generic state i, M_k produces $\mu_{k,i}$ parts per time unit. Therefore, state i with $\mu_{k,i}=1$ can be considered as an operational state for stage M_k, while state i with $\mu_{k,i}=0$ is a down state for stage M_k. The IADs are assumed reliable and they operate in negligible time, i.e. without interfering with machine process times. As described in [10], this representation provides the sufficient conditions for embedding of realistic stage dynamics with general distributions.

2.2. Product dynamics

Products flowing in the system are characterized by a set of states that are defined depending on specific attributes, including critical quality levels, part types, and timestamps. Such states evolve as long as the parts cross the different manufacturing stages of the system. In particular, products are characterized by state $P_{k,j}$, that describes the state of the part in the *j-th* slot of the *k-th* buffer, with $j=1,..,N_j$. The change of states occurs according to matrix Δ containing the transition probabilities of a change of state. A relevant change of product states may happen both when the part is waiting in a buffer, in the case of time-based product deterioration [6], or after the processing at a machine, in case of dimensional or geometrical product deviations [7]. Therefore, the generality of the model enables to deal with different root causes for product deterioration and defects.

2.3. Material flow dynamics

A discrete flow of parts is considered in the system. Stage M_k is starved if the buffer B_{k-1} is empty. Stage M_k is blocked if the buffer B_k is full or a IAD, immediately after it, is blocked. IADs are blocked whenever they have to move a part in an off-line buffer B_k, $k=K,\dots,K+K_a-1$, that is full. Operational Dependent Transitions are considered, i.e., a machine cannot make transitions to other states if it is starved or blocked.

2.4. Performance Measures

The main performance measures of interest are:

- Average total production rate of the system denoted by E^{Tot}.
- Probability that the lead time, LT, is equal to a given number of time units, h, i.e $P(LT=h)$.
- Average effective production rate of conforming parts, E^{Eff}.
- System yield, Y^{system}, i.e. fraction of conforming parts, (E^{Eff}/E^{Tot}).
- Total average inventory of the system, WIP.
- $SL(l,t)$: service level, i.e. the probability with which a lot of l parts is completed after t time units.

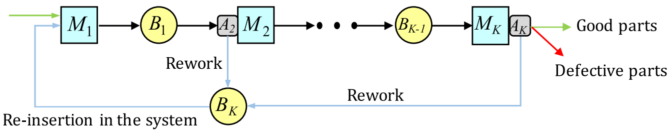

Figure 1. Representation of the analyzed manufacturing system.

3. Analysis of the system

In this paragraph, an efficient and exact analytical method to compute the target performance measures of the system is described. The rationale of the approach is explained in the following. Firstly, the integrated product - system state indicator is constructed. Secondly, the dynamics of the product - system while visiting different states is described. Thirdly, a discrete time - discrete state Markov chain is analyzed to compute the target performance measures of this system.

The generic state indicator for this product - system is described by the tuple $s=\left(\boldsymbol{b}_1,\boldsymbol{b}_2,\dots,\boldsymbol{b}_{K-1},\boldsymbol{b}_K,\dots,\boldsymbol{b}_{K+K_a-1},\alpha_1,\dots,\alpha_K\right)$ where the terms α_k represent the states of the manufacturing stages and $\boldsymbol{b_k}$ are vectors of the form $(P_{k,1},P_{k,2},\dots,P_{k,N_k})$, i.e. they represent the states of the parts stored in the buffers, introduced for simplicity of notation.

Consider a transition between two generic states s and s'. The probability to move the system from s to s' can be expressed as the product between the matrices describing the stage dynamics, λ_K, and the matrix describing the product dynamics, Δ. In mathematical form:

$$\gamma(s,s')=\prod_{k=1}^{K}\lambda_k\left[\alpha_k(s),\alpha_k(s')\right]\times\prod_{k=1}^{K+K_a-1}\prod_{j=1}^{N_k}\Delta\left[P_{\boldsymbol{k,j}}(s),P_{\boldsymbol{k,j}}(s')\right] \quad (1)$$

The dynamics of the integrated product – system state is then captured in the matrix Γ, with elements $\gamma(s,s')$.

Once the matrix Γ is built, all the relevant performance measures can be calculated by applying suitable numerical methods, such as the method reported in [6] for the lead time analysis, or the method reported in [11] for the computation of the effective and total throughputs, the system yield, the WIP, and the service level.

4. Numerical results and system behavior

In the first experiment, a scenario where the product state is characterized by the amount of time the product spends in a given portion of system, included between stages M_1 and M_K, is considered. It is assumed that the product is deteriorating, i.e. if the sojourn time of the parts within the system exceeds a certain threshold, τ, the part is defective. This situation is very common in the food industry, in the automotive industry, as well as in the semiconductor industry.

Three balanced systems having $K=2$, $K=3$, and $K=4$ stages respectively are considered. Parts are subject to a time threshold equal to $\tau=15$. If parts spend more than τ time units into the system, they are defective. If a IAD is positioned at stage M_K, it can detect a defective part and move it to a rework buffer B_K. Once in the rework buffer, parts wait for being reprocessed by machine M_1, and the time stamp is restarted. For each system, we analyzed two different scenarios. In the first scenario, the system is not equipped with IADs before stage M_K. Therefore, M_K may process parts that are already defective. In the second scenario, IADs prevent the processing of already deteriorated parts by removing them in the rework buffer before the service at M_K. The aim of the first experiment is then to evaluate the impact of the presence of IADs on the system effective throughput.

Case	N_1	N_2	N_3	E^{Eff} *gain*
1	10	-	-	3.61%
2	5	5	-	3.69%
3	3	3	4	2.46%

Table 1. Comparison of the system effective throughput between system with or without IADs.

As shown in Table 1, significant benefits can be brought by the use of IADs in the system. Indeed, by increasing the observability of the product state through the use of the identification and actuation devices, downstream production capacity can be saved thus avoiding processing parts that are already defective.

The second experiment aims at analyzing the impact of the IADs on the lead time distribution. The four-stage system in Case 3 is considered and the lead time distribution for both the parts crossing the system for the first time and the reworked parts is investigated. Figure 2 reports the lead time distribution for the system with the IADs (top figure) and without the IAD (bottom figure). As it can be noticed, the use of IADs cuts the tail of the distribution thus reducing the probability of exceeding the limiting residence time in the system.

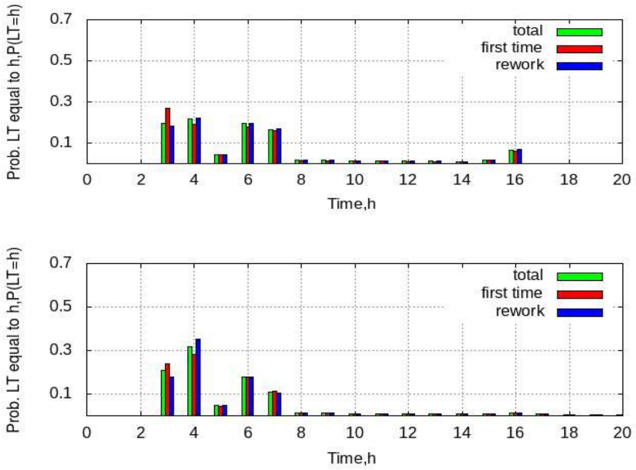

Figure 2. Lead time distributions of a four-stage systems with (top) and without IADs (bottom)

5. Real case study in the semiconductor industry

The proposed approach has been applied to a scenario of lead-time constrained system in the semiconductor industry at Robert Bosch, within the fab in Reutlingen, Germany. The portion of system under investigation is commonly known as *diffusion area.* It is usually the most critical area in semiconductor fabs. In this area, a number of wafers is placed in a cylindrical reactor, cleaning stage M_1, which is then sealed, heated, and filled with a carrier gas to allow dopant atoms present in the gas to diffuse into the exposed layer of the wafers, altering their electrical and chemical characteristics. Wafers are processed in standard lots whose size is dictated by material handling considerations. The time constraint starts immediately after the cleaning process. If the batches do not enter the downstream ovens, stage M_2, within a specific time window they must be reprocessed at the cleaning stage. The considered system in the diffusion area can be modelled as depicted in Figure 3.

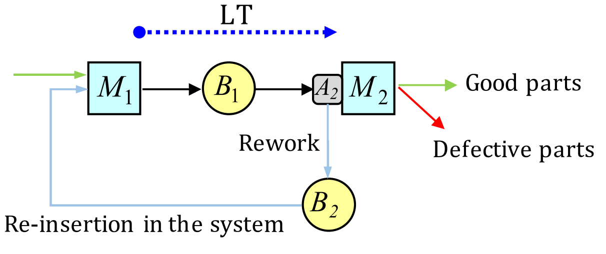

Figure 3. Modeled system in the diffusion area.

For the purposes of this experiment, data logs have been analysed and fitted by using the method described in [12] to obtain failures and repairs parameters, included in the matrices $\boldsymbol{\lambda}$, as well as the processing times of the machines, included in the vectors $\boldsymbol{\mu}$.
Three different types of products are considered. The first deteriorates after spending 8 time units in the buffer, whereas the second and the third deteriorate after 12 and 4 time units, respectively. The time length is omitted due to confidentiality reasons. The aim of the experiment was to find the value N_1^{opt} i.e. the capacity of buffer B_1 that maximizes the system throughput. The rework buffer has been considered fixed because evidence from the plant showed that it does not play a crucial role in this scenario. Therefore, N_r was set in such a way that the blocking of parts waiting for rework was negligible. Results are reported in Figure 4.

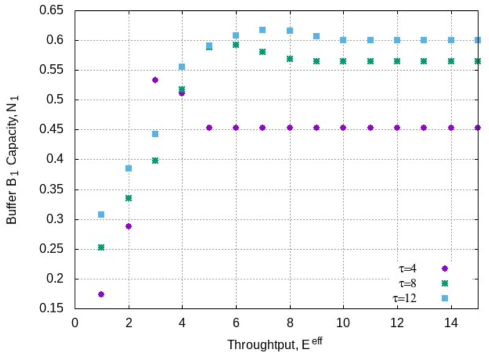

Figure 4. Effective throughput as function of the buffer capacity.

The following considerations hold:

- All the cases present a point of maximum followed by a plateau. Indeed, while small buffers reduce the probability of exceeding the residence time threshold, they negatively affect the production rate of the system. On the contrary, large buffers are beneficial for the production rate of the system, but the fraction of defective parts, exceeding the time threshold that need to be reworked, increases. Therefore, a trade-off is generated that is suitably solved by the proposed approach.
- The case with $\tau = 4$ is the one that benefits the most from the buffer optimization, since it has a difference between the optimal and the plateau area of about 20%.
- Although less marked, the difference is not negligible also for the case $\tau = 8$ and is rather small for the last case of $\tau = 12$. This is consistent with the fact that the tighter is the time constraint, the more important is the identification of the optimal buffer capacity.
- Despite this, it is possible to notice that opting for a conservative strategy by using a small buffer capacity is extremely penalizing for all the three cases.

6. Real case study in the railway industry

By exploiting the generality of the proposed approach, it has been applied to a system producing axles for the railway industry at CAF, in Spain. The materials used to produce the axles is carbon steel and low alloyed steel. The production system is composed by three main production area: *forging, machining* and *painting*.
In this second industrial case, the product state is not linked to the residence time of the parts in the system, but to dimensional and geometrical product variations, leading to out-of-tolerance product features, needing rework operations or scrap, depending on the related quality levels. The IAD, positioned at the end of the line, inspects the product, determines its quality level and can change the flow of the parts within the system according to the identification outcomes, sending them to rework or to scrap.
In this case, the focus of the analysis is on the machining area which has the most complex material flow due to the presence of Molybdenum axles. In the current configuration, after the marking operation, these axles are extracted from the main line and processed in a dedicated station for specific surface treatments. Then, these parts go back to the main line at the machining stage, sharing the remaining part of the system with the non-Molybdenum parts. All parts then pass through different

inspection stages including ultrasonic, magnetic particle inspection, and dimensional measurements. If the axles do not pass the dimensional inspection, they are reworked offline. If the variation entity is too extensive, parts are scrapped. According to historical data, scrapped parts are around the 5% and rework involves the 20% of the total part flow.

In order to simplify the machining area configuration and improve its performance, the company required to investigate the possibility to reconfigure the line by adding a second dedicated portion of system only for the Molybdenum axles, in order to avoid resource sharing with the non-Molybdenum axles. Such reconfiguration will require investment in additional equipment, but it is expected to be beneficial from the system performance point of view. Therefore, the method presented in this paper was adopted to quantify the benefits of such reconfiguration.

Figure 5 shows a simplified version of the complex flow of the reconfigured machining area. At the first centering stage a tag is attached to the axles enabling to discriminate between normal and Molybdenum axles. A first IAD splits the two flows according to this label. Then, the axles follow their paths until the inspection where a second tag is provided to determine if they are conforming, rework or scrap according to their quality level. Rework parts are processed in an off-line additional stage and then re-inserted upstream in the system.

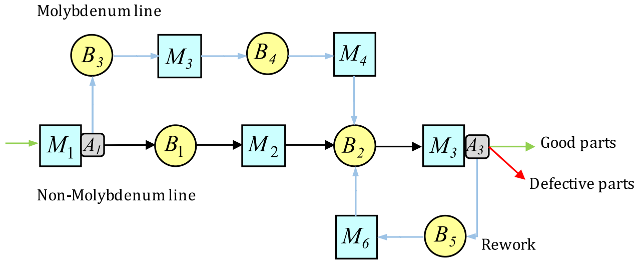

Figure 5. Machining area for the production of axles.

The method presented in this paper was used to analyse the service level of lots having different production mixes, both for the current configuration and the reconfigured system with the dedicated line. In particular, we considered three lots with the 5%, 15% and 30% of Molybdenum axles. The results are shown in Figure 6. As it can be noticed, a considerable difference in performance is shown. For example, it is possible to observe that the reconfigured system with the dedicated line has improved service level and is less sensible to the variation of the mixing. By extracting the 90% percentile from Figure 6, it can be seen that the new configuration saves the 45% of time units to complete the lot having 5% Molybdenum axles, whereas it saves the 42% and 29% of time units in case of lots with 15% and 30% Molybdenum axles respectively. This suggests that the gain obtained by adding a dedicated line drops more than linearly as function of the percentage of Molybdenum axles in the mix.

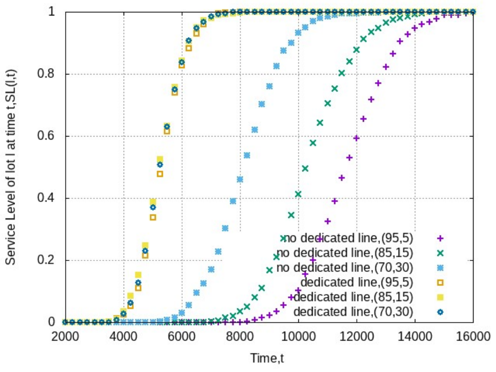

Figure 6. Comparison of service levels for the scenarios with and without the dedicated line.

7. Conclusions

This paper proposes a modeling framework and a methodology to evaluate the production quality performance of manufacturing systems operating under product observability, enabled by in-line product inspection and identification technologies, and rework. The proposed model captures the integrated dynamics of the product and system state evolution for the first time, thus enabling to estimate the system performance under several different settings. The numerical results show that the method provides interesting insights of the behavior of these complex systems. Moreover, the generality of the approach makes it possible to estimate the performance of the system within different manufacturing contexts and product deterioration scenarios, including residence time based as well as feature deviation related defect causes.

The theoretical work developed in this paper opens the way for applications in several different fields. For example, systems where the product state evolution is affected by specific in-line compensation and product repair practices can be easily investigated. Moreover, new production planning approaches that take into account the exact distribution of the residence time of parts in the system can be developed. In addition, specific approaches integrating product design, tolerancing and production planning [13] can be proposed by exploiting the generality of the model. Finally, new production flow control policies to provide a direct control on the production quality performance of the system [14] can be elaborated starting from the capability provided by the developed approach.

References

[1] Scholz-Reiter, B., Gorges, M. Philipp, T. 2009, Autonomously Controlled Production Systems – Influence of Autonomous Control Level on Logistic Performance, CIRP Annals – Manufacturing Technology: 58(1): 395–398.

[2] Putz, M., Wiener, T., Pierer, A., Hoffmann, M., 2018, A multi-sensor approach for failure identification during production enabled by parallel data monitoring, CIRP Annals - Manufacturing Technology: 67 (1): 491–494.

[3] Colledani M., Fischer A., Iung B., Lanza G., Schmitt R., Tolio. T., Vancza J., 2014, Design and Management of Manufacturing Systems for Production Quality, *CIRP Annals - Manufacturing Technology*, 63(2): 773-796.

[4] Kroll, H., Copani, G., Van de Velde, E., Simons, M., Horvat, D., Jäger, A., Wastyn, A., Pourabdollahian, G., Naumanen, M., 2016, An analysis of drivers, barriers and readiness factors of EU companies for adopting advanced manufacturing products and technologies, EU Commission, Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs.

[5] Scholz-Reiter, B., Rekersbrink, H., Gorges, M., 2010, Dynamic Flexible Flow Shop Problems – Scheduling Heuristics vs. Autonomous Control, CIRP Annals – Manufacturing Technology 59(1): 465–468.

[6] Angius, A., Colledani, M., Horvath, A., 2015, Production quality performance in manufacturing systems processing deteriorating products. CIRP Annals – Manufacturing Technology, 64 (1): 431–434

[7] Colledani, M., Tolio, T., 2012, Integrated quality, production logistics and maintenance analysis of multi-stage asynchronous manufacturing systems with degrading machines, CIRP Annals - Manufacturing Technology, 59(1): 455-458.

[8] Thiede, S., Turetskyy, A., Kwade, A., Kara, S., Herrmann, C., 2019, Data mining in battery production chains towards multi-criterial quality prediction, CIRP Annals - Manufacturing Technology, 68(1): 463–466.

[9] Djurdjanovic, D., Haq, A.-U., Magnanini, M., Majstorovic , V., 2019, Robust model-based control of multistage manufacturing processes, CIRP Annals - Manufacturing Technology, 68 (1): 479–482.

[10] Angius, A., Colledani, M., 2019, Integrated production and reconfiguration planning in modular plug-and-produce production systems. CIRP Annals - Manufacturing Technology, 68 (1): 435–438.

[11] Angius, A., Colledani, M., Horvath, A. 2014, Moments of accumulated reward and completion time in Markovian models with application to unreliable manufacturing systems, Performance Evaluation, 75: 69-88.

[12] G. Horvath, P. Buchholz and M. Telek, 2005, A MAP fitting approach with independent approximation of the inter-arrival time distribution and the lag correlation, *Second International Conference on the Quantitative Evaluation of Systems (QEST'05)*, Torino: 124-133.

[13] Tsutsumi, D., Gyulai, D., Kovács, A., Tipary, B., Ueno, Y., Nonaka, Y., Monostori, L., 2018, Towards joint optimization of product design, process planning and production planning in multi-product assembly, CIRP Annals - Manufacturing Technology, 67 (1): 441-446.

[14] Lanza, G., Haefner, B., Kraemer, A., 2015, Optimization of selective assembly and adaptive manufacturing by means of cyber-physical system based matching, CIRP Annals - Manufacturing Technology, 64 (1): 399-402.