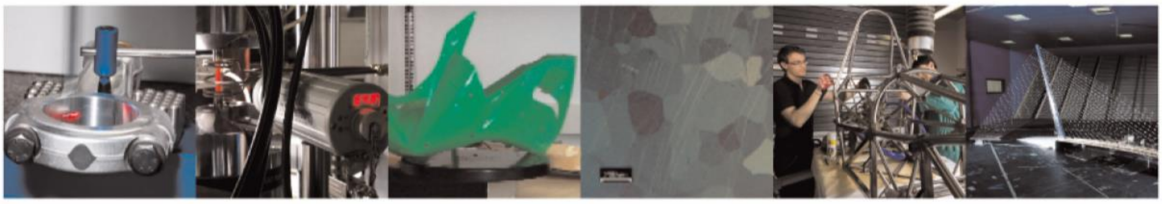
POLITECNICO
MILANO 1863

DIPARTIMENTO DI MECCANICA

mecc



Energy-Efficient Control Policy for Parallel and Identical Machines with Availability Constraint

Alberto Loffredo; Nicla Frigerio; Ettore Lanzarone; Andrea Matta

This is a post-peer-review, pre-copyedit version of an article published in IEEE Robotics and Automation Letters. The final authenticated version is available online at:
http://dx.doi.org/10.1109/LRA.2021.3085169

DIPARTIMENTO DI MECCANICA ■ POLITECNICO DI MILANO
via G. La Masa, 1 ■ 20156 Milano ■ EMAIL (PEC): pecmecc@cert.polimi.it
http://www.mecc.polimi.it
Rev. 0

Energy-Efficient Control Policy for Parallel and Identical Machines with Availability Constraint

Alberto Loffredo[1], Nicla Frigerio[1], Ettore Lanzarone[2,3], and Andrea Matta[1]

Abstract—**Energy efficiency is becoming a key subject in manufacturing, especially if related to the major environmental impact of machining activities. Nowadays, great research efforts are performed to find new methodologies to improve sustainability of manufacturing processes. Energy consumption can be lowered controlling machine state during idle periods. This can be achieved with Energy-Efficient Control (EEC) policies that switch off/on the machine. The same approach can be applied simultaneously to more identical devices in a parallel machines workstation. This paper presents a novel model to identify EEC policies for manufacturing systems composed of parallel and identical machines with finite buffer capacity. In this configuration, each machine can be switched on with a stochastic startup time and switched off instantaneously also considering an availability constraint. The proposed model reduces energy consumption while assuring a target availability level. The control is executed using buffer level information. Numerical results confirm model benefits when applied to a real industrial system from the automotive sector.**

Index Terms—**Sustainable Production and Service Automation; Energy and Environment-Aware Automation.**

I. INTRODUCTION

In recent years, the energy saving topic is becoming increasingly important in industry: environmental impact and sustainability of processes are, nowadays, considered critical factors for this field. Recent studies estimated that around the 40% of the overall electricity use in industry corresponds to manufacturing activities and this value appears to enhance in future years [1]. Therefore, companies are investing more and more resources to improve energy efficiency of their manufacturing systems. A major trend to achieve this goal is the Energy-Efficient Control (EEC) of machines executing machining operations ("machines" from now on in the paper). Commonly, in manufacturing systems, machines are kept in ready-for-process conditions during idle periods (i.e. when the part flow is interrupted and the machine is not operating on parts). This policy is known as *Always On* policy. However, the machine is not-productive during idle periods but it still has a high energy request to maintain the ready-for-process conditions. The idea behind machine EEC is to provide policies for controlling machine state towards the optimum trade-off between system production rate and energy demand. Specifically, the machine might be switched off to reduce the energy consumed during idle periods; then, it might be switched on when the production has to be resumed or the machine has to be ready before the arrival of a part. Nevertheless, the switching on transition requires a startup time to resume the service and, consequently, produces additional energy consumption. A proper EEC policy should improve machine sustainability while assuring a target production rate. The latter is strictly related to the system availability, since the machine can process parts only when available. Thus, in manufacturing systems the achievement of a target availability level is considered a main requirement.

The switch off/on control approach can be applied simultaneously to more identical devices in a parallel machines workstation. This configuration is widely used in industry to have balanced manufacturing systems in terms of workstations workload. For this reason, this work is focused on this type of configuration. The goal is to obtain an EEC policy using buffer level information to switch off/on identical parallel machines with a common upstream finite capacity buffer while satisfying a target level on the system availability.

A. Related Literature

Research efforts focused on the EEC of a single-buffer-single-machine manufacturing system. In literature, this analysis is addressed at two distinct levels: considering the stand-alone machine without focusing on its interactions with the shop floor (recent examples in [2], [3]) and considering the overall production system, i.e. a series of single machines interspersed by single finite buffers (recent examples in [4], [5], [6]). In this paper, the focus is on a completely different manufacturing system architecture, where parts wait in the same upstream buffer and can be processed indifferently by one of the identical parallel machines. Hence, the system behavior is not immediate and straightforward as in the single machine case and the EEC assumes a totally different perspective. Especially in production periods with few parts in the buffer, more machines in the workstation might be switched off for energy saving purposes. Thus, in this case, the EEC focuses on how many machines in the station should be switched on (and off) in each instant to improve the energy efficiency while maintaining a certain availability level. In literature, a comprehensive study for this problem in manufacturing field is missing.

Using queueing theory, the identical parallel machines system under investigation can be modeled as a multi-server queue with a common upstream buffer of finite capacity

[1]A. Loffredo, N. Frigerio and A. Matta are with the Department of Mechanical Engineering, Politecnico di Milano, Milan, Italy. A. Loffredo is the corresponding author alberto.loffredo@polimi.it.

[2]E. Lanzarone is with the Department of Management, Information and Production Engineering, University of Bergamo, Dalmine, (BG), Italy.

[3]E. Lanzarone is with the Institute for Applied Mathematics and Information Technologies (IMATI), National Research Council of Italy (CNR), Milan, Italy.

and part arrivals and processing times determined by Poisson processes (M/M/c/K queue). The parallel machines are represented by the multiple servers. Queueing theory literature for controlling this type of system is wide, even though the majority is not related to manufacturing field. To the best of our knowledge, Szarkowicz et al. [7] were the first to propose a control policy for an M/M/c/K system with controllable servers, where each of them can be switched off and on instantaneously. This model was extended in more recent years by Printezis et al. [8] and Feinberg et al. [9] for an M/M/c/K queue modeling a customer-provider situation. In this case, the switch off/on policy is applied to understand when service providers should work or not. Optimal pricing policies to serve customers and maximize the profit are computed starting from that. The mentioned works on M/M/c/K queues did not consider startup times and a target availability level to be satisfied.

A different approach can be found considering control policies for M/M/c systems for server farms and data-centers. Gandhi et al. [10], [11] introduced a *staggered-setup* policy for an M/M/c queue with exponential startup time. This policy imposes a number of switched on servers equal to the number of items waiting to be processed and servers to be switched off as soon as they become idle. The same model was analyzed by Mitrani [12], where a certain number of servers are switched on when the number of elements in the system exceeds a certain threshold; then, they are switched off when the number of parts becomes lower than another threshold. Another approach was used by Xu and Tian [13] with (e,d)-control policies for M/M/c systems: when d servers are idle, e servers are switched off. Lastly, an extensive analysis for M/M/c queues in server farms and data-centers was performed by Maccio and Down [14], [15], [16], [17]. In [14] and [16] the authors extracted structural properties pertaining to the optimal policy to be applied for controlling M/M/c queues. Then, in [15] and [16], they analyzed system performances when applying two specific policies satisfying those structural properties. In their most recent work on this topic [17], policies for an M/M/c configuration are studied under the asymptotic regime where the number of servers approaches infinity while the load per server remains constant. However, all the mentioned works on server farms and data-centers are characterized by assumptions not properly aligned with manufacturing field, as part processing and startup procedure that can be interrupted. Moreover, in server farms and data-centers related works the authors did not consider finite buffer capacity and a target availability level to be satisfied. Thus, all the extracted properties and policies might be not suitable for manufacturing systems.

B. Contribution

Identical parallel machines configuration is widely used in manufacturing. Nevertheless, literature does not address deeply the EEC approach for this type of system and models to identify EEC policies reducing the overall workstation energy consumption are not present. Furthermore, providing EEC policies that also guarantee the achievement of a target availability level is considered a main requirement in manufacturing. This type of need is not discussed in literature for the identical parallel machines configuration. To fill this gap, this work proposes a novel model leading to an EEC policy using buffer level information to switch off/on identical parallel machines with finite buffer capacity. An availability constraint is included in the proposed model, to ensure the achievement of a target availability level with the obtained EEC policy. Dynamic Programming (DP) methodology is used to solve the problem. Control actions can be taken at specific instants in time and the problem is formalized as a Discrete Time (DT) Markov Decision Process (MDP). The proposed model represents the first-time application of a well-proven method, i.e. the MDP, to obtain an EEC policy for a workstation composed by an upstream buffer of finite capacity and multiple identical parallel machines operating in the manufacturing context. A real case of identical parallel machines configuration from the automotive sector is analyzed, to show how the proposed model provides benefits in an industrial system.

The remainder of the paper is organized as follows. Section II includes a detailed description of the system under investigation. In Section III a complete formulation of the proposed model is reported. Section IV provides a numerical validation of the proposed model. Further numerical analysis is presented in Section V, showing the resulting benefits of the model when applied to the industrial case along with sensitivity analysis of results. Conclusions and further developments are discussed in Section VI.

II. SYSTEM DESCRIPTION

A workstation composed by an upstream buffer of finite capacity and multiple identical parallel machines working on a single part type is considered as the system to be controlled (Figure 1). Parts arrive to the buffer following a stochastic process with rate λ. First come first served rule is applied. The machine is *busy* while working on parts with stochastic processing times with rate μ; on the other hand, the machine is *idle* when it is in ready-for-process conditions but it is not operating on parts. *Busy* and *idle* are two sub-states composing the machine *working* state. From the *working* state, the machine can be switched off instantaneously going into the *standby* state: a lower power request state where only emergency services are active, so that the machine cannot process parts and the service is interrupted. From the *standby* state, the machine must go through the *startup* state to come back in *working* state: a stochastic startup time with rate δ is required to resume inactive machine components.

III. ASSUMPTIONS AND MODEL FORMULATION

The proposed model is composed of two modules.

First Module - Sections from III-A to III-E describe the first module. The problem is formalized as a Continuous Time (CT) MDP and then converted into a Discrete Time (DT) MDP with the uniformization technique. DP methodology is used to solve the problem. The identified solution of the MDP problem corresponds to the optimal EEC policy minimizing the expected total discounted energy cost over an infinite

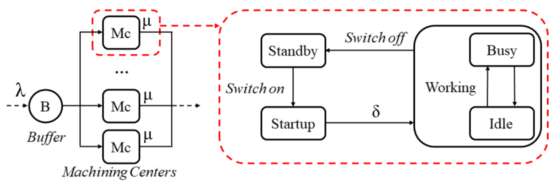


Fig. 1. The identical parallel machines system under analysis along with the machine model. Parts arrive to the buffer following a stochastic process with rate λ. Machines have stochastic processing and startup times with, respectively, rates μ and δ.

horizon. The goal of the first module is to identify the optimal EEC policy based on buffer level information.

Second Module - In Section III-F an availability constraint is introduced. When the optimal EEC policy found out with MDP does not guarantee a target level on system availability, it is not suitable to be applied to the system under investigation. Starting from the optimal EEC policy, the second module iteratively modifies the policy to be applied until the desired availability level is reached. In this way, a new and suitable EEC policy is obtained. This EEC policy does not guarantee the global optimum in terms of energy saving, but it improves system sustainability while assuring a target availability level.

A. Assumptions

The following assumptions are introduced to be aligned with the manufacturing system under investigation and are assumed to be valid in this paper:

(i) The system is composed by c identical parallel machines and one shared upstream buffer with finite capacity K.

(ii) There is an infinite capacity buffer downstream the system. Hence, machines cannot be blocked and processed parts leave the system immediately after the process completion.

(iii) The stochastic processes involved in the system are assumed to be Poisson processes, independent of each other and stationary. Thus, machines are characterized by exponentially distributed startup time with rate δ and processing times exponentially distributed with rate μ; parts arrive to the buffer following a Poisson process with rate λ.

(iv) Machines work on a single part type and each machine can be switched off instantaneously.

(v) Each machine is characterized by the following states: *working*, *standby* and *startup* (w, sb and su, respectively). The *working* state is composed by two sub-states: *idle* (i) and *busy* (b).

(vi) w_s represents the power requested in machine state $s = \{w, sb, su\}$. w_w is non-negative and depends on the power requested in the two sub-states i and b (w_i and w_b, respectively). w_{sb}, w_{su}, w_i and w_b are constant and they represent the average power requested by the machine in those states or sub-states. Moreover $w_b > w_i > w_{sb} \simeq 0$ and $w_{su} > w_i$.

(vii) System failures are modeled by embedding them into machine processing times.

(viii) Part processing and the startup procedure cannot be interrupted by the control.

(ix) Part arrival process stops when the buffer is full.

B. Decision Epochs

The time horizon is divided into periods $k = \{1, 2, \dots\}$ of variable length, according to the occurrence of event $y \in \mathbb{Y}$. Define the event set $\mathbb{Y} = \{A, D, E\}$: part arrival to the buffer ($y = A$), part departure due to process completion ($y = D$), and startup completion ($y = E$). Trivially, departures cannot happen when the system is empty, arrivals cannot happen when the system is full and startup completion occurs only when at least one machine is in *startup* state. The event y_k happens at the end of period k and the decision epochs k correspond to instances of the event y_k.

C. State Space and Action Space

The system state $s_k \in \mathbb{S}$, at the beginning of period k, is denoted by the ordered triple $s_k = \{n_k, x_k, z_k\}$. The number of parts in the upstream buffer is represented with the integer variable $n \in \{0, 1, 2, \dots, K\}$. The number of enabled machines is represented with the integer variable $x \in \{0, 1, 2, \dots, c\}$. The enabled machines are the switched on machines, i.e. that are in *working* or *startup* state. Lastly, $z \in \{0, 1, 2, \dots, c\}$ counts how many machines are in *startup* state. Consequently, the number of machines in *working* state is $(x - z) \in \{0, 1, 2, \dots, c\}$. Each machine has a service rate equal to $\mu > 0$ (when in *working* state) or 0 (when in *startup* or *standby* state). For this reason, the service rate of the overall workstation is $\mu_{tot} = (x - z)\mu \in \{0, \mu, 2\mu, ..., c\mu\}$.

The control action a is applied to control x: the number of machines to be enabled. Consequently, also the number of not-enabled machines, equal to $(c - x)$, is controlled. The not-enabled machines are the switched off machines, i.e. that are in *standby* state. In addition, x and z are strictly correlated: when a machine is switched on from the *standby* state, x increases and also z increases because the machine goes in *startup* state after the switch on transition. On the other hand, when a machine is switched off, x decreases but z does not vary, since the machine immediately goes in *standby* state. Thus, $z = z(a)$. The set of feasible action is $\mathbb{A}(s, y)$, depending on system state s and event occurrence y. $\mathbb{A}(s, y)$ is determined by the assumption (viii) and trivial boundaries (e.g. when $x = 0$ it is not possible to switch off any machine and when $x = c$ it is not possible to switch on any machine). At the end of the period k, after the event y_k is observed, x_k can be controlled with the action a_k. The optimal policy π^* maps the optimal action $a_k^*(s_k, y_k)$ from system state s_k and event y_k.

D. System Dynamics and Uniformization

System dynamics is assumed to be stationary and it is represented by $\mathbb{Z} : \mathbb{S} \text{ x } \mathbb{Y} \text{ x } \mathbb{A}(s, y) \to \mathbb{S}$. Starting from system state s_k, the event y_k and the control action a_k, the next system state $s_{k+1} = \{n_{k+1}, x_{k+1}, z_{k+1}\}$ is defined as follows:

$$s_{k+1} = \begin{cases} \{\min[n_k+1,K], a_k, z(z_k,a_k)\} & if \;\; y_k = A \\ \{\max[n_k-1,0], a_k, z(z_k,a_k)\} & if \;\; y_k = D \\ \{n_k, a_k, \max[z(z_k-1,a_k),0]\} & if \;\; y_k = E \end{cases} \quad (1)$$

Thus, according to the event y_k, the transition probabilities $p(s_k, s_{k+1}, y_k, a_k)$ for the MDP problem are:

$$p(s_k, s_{k+1}, y_k = A, a_k) = \begin{cases} 0 & if \;\; n_k = K \\ \lambda & otherwise \end{cases} \quad (2)$$

$$p(s_k, s_{k+1}, y_k = D, a_k) = \begin{cases} 0 & if \;\; n_k = 0 \\ (x_k - z_k)\mu & otherwise \end{cases} \quad (3)$$

$$p(s_k, s_{k+1}, y_k = E, a_k) = \begin{cases} 0 & if \;\; z_k = 0 \\ \delta & otherwise \end{cases} \quad (4)$$

In this way, the system is described by a Continuous Time Markov Chain (CTMC). With the uniformization technique [18], it is possible to convert the CTMC into a Discrete Time Markov Chain (DTMC) using a uniform transition rate ν defined as follows:

$$\nu = \lambda + c(\delta + \mu) \quad (5)$$

ν represents the exponential parameter that yields the minimum expected transition duration. The transition probabilities become $p'(s_k, s_{k+1}, y_k, a_k) = \frac{1}{\nu} p(s_k, s_{k+1}, y_k, a_k)$.

E. Payoff Function and Optimality Equation

The payoff function consists of three elements: the machine power cost per unit time, the holding cost and the startup cost. The machine power request in working state is $w_w(s, w_i, w_b)$ (assumption vi). Defining $m_b = [\min[(x-z), n]$ as the number of *busy* machines and $m_i = (x-z) - m_b$ as the number of *idle* machines, $w_w(s, w_i, w_b)$ can be defined as $w_w(s, w_i, w_b) = m_b w_b + m_i w_i$. Thus, the machine power cost per unit time is $c_w(s, w_i, w_b) \propto w_w(s, w_i, w_b)$. A non-negative constant power consumption w_h is associated to the buffer level n: this value determines a linear holding cost $c_h(n) \propto w_h n$. $c_h(n)$ is used to represent a penalty imposed to the system for maintaining parts in the buffer and not processing them; in this way, as w_h increases, the control is prone to be more productive. Being the power consumption in *standby* state w_{sb} close to zero (assumption (vi)), the correspondent standby cost c_{sb} is considered null. The startup cost is non-negative, constant and represented by $c_{su} \propto w_{su}$.

Defining a discount factor $0 < \rho < 1$ and $\eta = \rho + \nu$ it is possible to define the transition probabilities for the infinite horizon discounted cost scenario as $\widetilde{p}(s_k, s_{k+1}, y_k, a_k) = \frac{\nu}{\eta} p'(s_k, s_{k+1}, y_k, a_k)$. Thereby, the Bellman's optimality equation for the infinite horizon discounted cost can be expressed as:

$$V^*(s_k) = \min_{a_k} \Bigg[g_k(s_k) + \sum_{s_{k+1} \in \mathbb{S}} \widetilde{p}(s_k, s_{k+1}, y_k, a_k)(V^*(s_{k+1}) + \max[0, (a_k - x_k)] c_{su})) \Bigg] \quad (6)$$

Where $g_k(s_k)$ is the stage cost:

$$g_k(s_k) = \frac{c_w(s, w_i, w_b) + c_h(n)}{\eta} \quad (7)$$

The value iteration method [19] can be used to numerically approximate Equation (6). This numerical approximation represents the minimum expected cost that the system, starting from state s_k, will incur when the optimal control action a_k^* is applied. Thus, the optimal policy π^* indicates the optimal action $a_k^*(s_k, y_k)$ to be implemented, i.e. the number of machines to be enabled from system state s_k and event y_k to minimize the energy consumption. It must be noticed that, being the system dynamics stationary and the cost functions independent from the period k considered, also the obtained policy is independent from the period k considered. As said, π^* is based on s and y; however, system state is represented by $s = \{n, x, z\}$ and, as stated in Section III-C, a determines both x and z. Thus, the only non-controlled parameter in s is the buffer level n. Hence, π^* can be based on buffer level n and event y. In this way, the optimal EEC policy using buffer level information to switch off/on identical parallel machines is obtained.

F. Availability Constraint

When a policy π is applied to the system, a Markov Chain (MC) can be generated: it describes the system behavior when π is imposed. From the MC, system performance indicators can be computed. In particular, the system availability when π is applied can be extracted. It is represented with the continuous variable $u_\pi \in [0, 100\%]$. $u_\pi = 0\%$ if all the machines are always not-enabled and $u_\pi = 100\%$ if all the machines are always enabled. If u_{π^*} is higher or equal than the target availability level to be guaranteed u_{target}, π^* is suitable for the system under investigation. On the contrary, if the requested availability level is not satisfied, π^* must be modified. Hence, an availability constraint is introduced within the second module composing the proposed model of this work (along with the MDP previously presented). Starting from π^*, this module iteratively modifies the EEC policy to be applied, until the target availability level is satisfied. New variables are defined:

- a_n: the number of machines to be enabled, according to a policy π, when the buffer level is n;
- n_s: the highest value of n for which a_n is lower than c;
- a_{n_s}: the number of machines to be enabled, when the buffer level is n_s.

In each iteration, a_{n_s} is increased. In this way, a modified policy π' is obtained along with a new system availability $u_{\pi'}$. This procedure is reiterated until a suitable EEC policy π_{suit} is obtained. π_{suit} guarantees $u_{\pi_{suit}} > \; u_{target}$; thus, it reduces the energy consumption of the identical parallel machines, with a switch off/on approach, while satisfying the target level on the system availability. A complete overview on the model proposed in this work is shown in Algorithm 1 and it includes the two modules presented in Section III.

Algorithm 1 The proposed model to obtain an EEC policy for identical parallel machines with availability constraint. The two modules composing the model are presented.

```
First Module: Solve Eq. (6) with the value iteration method and find π*
Impose π* in the system
Compute u_{π*}
if u_{π*} > u_target then
    π_suit = π*.
else
    Second Module:
    π = π* & u_π = u_{π*}
    while u_π < u_target do
        From π, find n_s = max[ n | a_n ≠ c ]
        Impose a_{n_s} = a_{n_s} + 1
        New policy π' obtained
        Compute availability u_{π'}
        π = π' & u_π = u_{π'}
    end while
    π_suit = π
end if
The policy to be applied is π_suit
```

IV. VALIDATION OF THE PROPOSED MODEL

A numerical analysis is carried out to show the validity of the proposed model for general cases of identical parallel machines manufacturing systems. A full factorial design with 8 factors at two levels [20] is used to generate 256 different cases where the proposed model has been applied. In all the systems, the upstream buffer has a finite capacity equal to 10, a target availability level of 80% is required and the discount factor ρ is equal to 0.80. For the varying factors, the mean processing, part arrival and startup times have the same low level equal to 10 s and the same high level equal to 60 s. The corresponding rates μ, λ and δ, along with the other factors and levels are reported in Table I. In all the experiments, the value iteration method is used to numerically approximate optimality Equation (6) with *Matlab (Mathworks, Natick, MA, US)* software (10^3 iterations). The mean computational time for one experiment is lower than 2 s.

TABLE I
FACTORS AND LEVELS FOR THE FULL FACTORIAL DESIGN.

Factor	c	λ	μ	δ
Low Level	3	0.02	0.02	0.02
High Level	6	0.1	0.1	0.1
Factor	w_b	w_i	w_h	w_{su}
Low Level	20 (kW)	0.5 (kW)	0.5 (kW)	5.5 (kW)
High Level	60 (kW)	5 (kW)	25 (kW)	19.5 (kW)

Figure 2 shows the resulting energy saving values for the analyzed configurations. In 32 cases, due to specific conditions (e.g. fast arrival of parts and/or high machine processing time), the identified policy π_{suit} corresponds to the *Always On* policy. In these cases, the machines are almost never idle and, for this reason, they cannot be switched off for energy saving purposes. In all the other 224 cases, π_{suit} is different from the *Always On* policy and this leads to significant energy saving values, up to 15.60%. The highest savings are achieved with parameters leading to frequent production periods with few parts in the buffer (e.g. slow arrival of parts and/or low machine processing time). In these cases, more machines in the workstation are frequently idle and might be switched off to save energy. However, in all the analyzed situations, the application of the proposed model is able to offer an appropriate and effective EEC policy and, for this reason, the model effectiveness is verified.

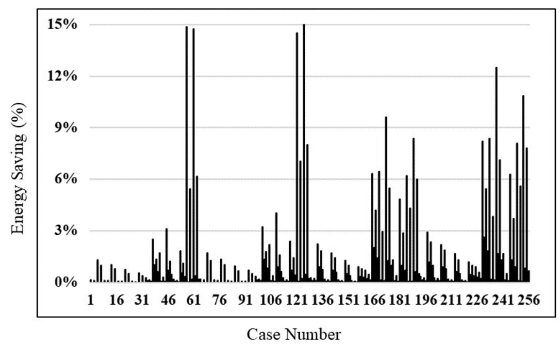


Fig. 2. Full factorial analysis: achieved energy saving in the different cases.

V. NUMERICAL EXPERIMENTS

A. Industrial Case Analysis

In this Section, the proposed model is applied to an industrial case to assess its benefits when implemented on a real system. The latter is an identical parallel machines workstation used in a manufacturing system producing cylinder heads in the automotive sector. The production line is composed by 21 total workstations and the one analyzed is used for machining operations. The workstation is composed by 6 identical parallel machines. The upstream buffer has a finite capacity equal to 10. For this system, a target availability level of 85% is required. The mean processing, part arrival and startup times are, respectively, equal to 83.70, 15 and 30 s. The corresponding rates μ, λ and δ, along with the other system parameters are reported in Table II. These parameters are provided by the company owning the industrial system under study. The value iteration method [19] is used to numerically approximate optimality Equation (6) with *Matlab* software (10^3 iterations). The computational time is lower than 2 s.

TABLE II
SYSTEM PARAMETERS FOR THE INDUSTRIAL CASE UNDER INVESTIGATION.

Parameter	**Value**	**Parameter**	**Value**
c	6	ν	0.34
K	10	w_b	15 (kW)
λ	0.07	w_i	9.30 (kW)
μ	0.01	w_h	0.75 (kW)
δ	0.03	w_{su}	10 (kW)
ρ	0.80	-	-

At first, the model is applied to the system without the availability constraint. The optimal EEC policy π^* is obtained

TABLE III
EEC POLICY π^* **(A)** AND π_{suit} **(B)** FOR THE INDUSTRIAL CASE UNDER INVESTIGATION.

	(A)	(B)
n	a_n	a_n
0	0	0
1	1	1
2	2	2
3	3	3
4	4	5
5	5	6
$6;7;8;9;10$	6	6
Energy Saving	9.38%	8.76%
System Availability	81.94%	86.49%
Target Availability Level	Not Present	85.00%

(reported in Table III-(A)). π^* indicates, for each possible n, the optimal action a_n to perform, i.e. the number of machines to be enabled in the workstation. The resulting average energy saving is 9.38% and it is computed comparing π^* to the *Always On* policy. However, the system availability (81.94%) is lower than the target (85%) and, for this reason, π^* is not suitable for the system under investigation. Afterwards, the complete version of the model is applied, inserting the availability constraint. Starting from π^*, 2 iterations are required to reach the target availability level required:

(1) Iteration 1 $\rightarrow n_s = 5 \rightarrow a_5 = a_5 + 1 = 6 \rightarrow u_{\pi'} = 83.95\% < u_{target}$.
(2) Iteration 2, starting from the policy π obtained in iteration 1 $\rightarrow n_s = 3 \rightarrow a_4 = a_4 + 1 = 5 \rightarrow u_{\pi'} = 86.49\% > u_{target}$. The policy to be applied is $\pi_{suit} = \pi'$.

In this way, a suitable policy π_{suit} is obtained (presented in Table III-(B)). The new availability level (86.49%) is higher than the target (85%): π_{suit} is suitable for the industrial case. Finally, the resulting average energy saving (8.76% in respect to the *Always On* policy) is lower than the previous case (9.38%), but still represents a significant value.

B. Sensitivity Analysis

A sensitivity analysis of results is performed to understand into details the effect of u_{target} and w_h (i.e. the target availability and the holding power consumption) on system performances.

Target Availability - In this analysis u_{target} varies while the other system parameters remain unchanged (reported in Table II). Increasing u_{target} leads to increasing system throughput until the maximum possible throughput for the system under study is reached. This value is equal to the arrival rate of parts λ because of flow conservation (Figure 3-(a)). Moreover, when system availability increases, machine power cost enhances and energy saving decreases until a null value is reached (Figure 3-(b)). System availability equal to 100%, indeed, corresponds to the *Always On* policy condition; thus, the energy saving achieved in this case is null. Finally, the achievement of u_{target} is always guaranteed (Figure 3-(c)).

Holding Cost - In this case w_h varies and the other system parameters remain the same as in Table II. Figure 4 shows system performance for this analysis. For low values of w_h, the optimal EEC policy might indicate to keep the system availability low, since maintaining parts in the buffer (and not processing them) has a low total cost in this case. However, the availability constraint imposes the system availability to be above the target level. At the same time, the energy saving slightly decreases because of the increasing holding cost. For high values of w_h, the situation is different. The resulting high holding cost leads to EEC policies imposing high system availability (and high throughput): in this way, parts are processed and not maintained in the buffer. However, high system availability means enhancing machine power cost and decreasing energy saving for high $w_h(n)$ values. Finally, the achievement of u_{target} is always guaranteed.

C. Assumptions Modification

The policy π_{suit} identified in Section V-A is then applied to the industrial case but with different assumptions, in order to verify the effectiveness of the proposed model also for more general cases. Assumptions (iii) and (vii) are modified in this analysis: machines processing and startup times are imposed following two lognormal distributions equal to, respectively, LOGN(4.32, 0.47) and LOGN(3.28, 0.47). The two distributions lead to the same coefficient of variation (0.50) and mean values equal to, respectively, 83.70 and 30 s. The mean values are the same as in Section V-A, to compare two cases with aligned parameters, evaluating only the difference in terms of stochastic distribution involved. The lognormal distribution is selected because it is always statistically plausible to model a stochastic time distribution with a lognormal if its coefficient of variation is lower than 1 [21]. Thus, this distribution leads to a system modeling undoubtedly aligned to real manufacturing systems. All the other system parameters remain the same as in Table II. Discrete event simulation is used for performance evaluation: the number of experiments is equal to 25 with a simulation length of 100 days. The system simulation model is developed in *Arena* environment. Under these conditions, the resulting average energy saving is equal to 7.22$\pm$0.06% in respect to the *Always On* policy; this value is extracted with a confidence level of 95% on its confidence interval. Due to the different stochastic distribution involved, the resulting energy saving value is lower than the original case analyzed in Section V-A (8.76%) but still represents a significant value. This confirms the effectiveness of the proposed model also with a modification of the assumptions.

VI. CONCLUSIONS AND FURTHER DEVELOPMENTS

In this work, a novel model is presented: it is used to identify EEC policy based on buffer level information for identical parallel machines with finite buffer capacity used in manufacturing systems. The proposed model leads to the reduction of the energy consumption; at the same time, it ensures a target level on system availability without large computational efforts. Numerical results are presented, showing model benefits when applied to a real industrial case from the automotive sector.

To extend this work, an interesting study could be performed with more control actions such as the control of admitted parts.

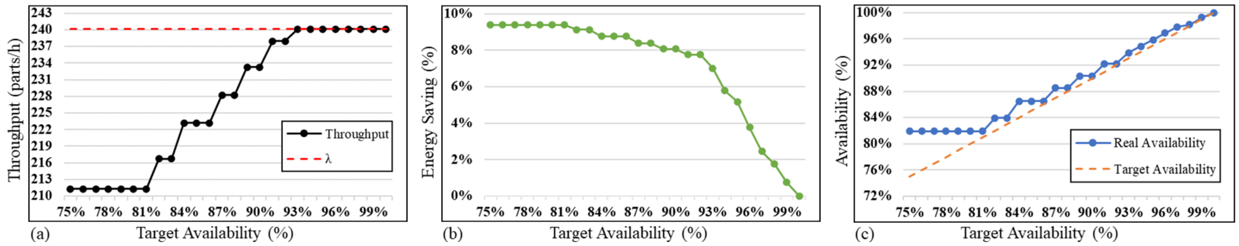


Fig. 3. Sensitivity analysis for the industrial case: throughput (a), energy saving (b) and system availability (c) achieved when u_{target} varies.

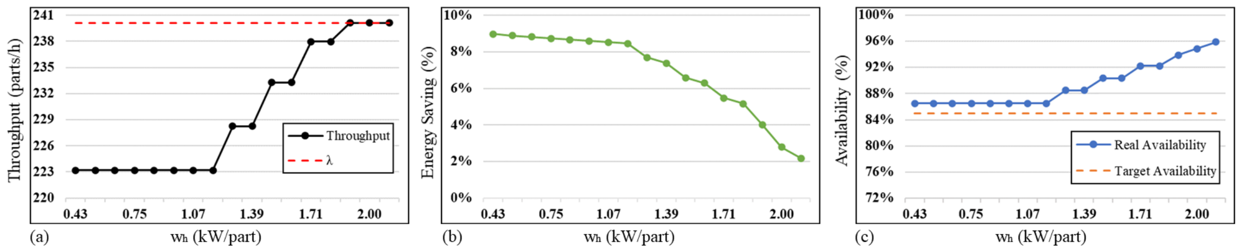


Fig. 4. Sensitivity analysis for the industrial case: throughput (a), energy saving (b) and system availability (c) achieved when w_h varies.

Moreover, structural properties pertaining to the optimal EEC policy could be identified. Finally, another challenging topic might be the extension of the proposed model to control a larger manufacturing system. In this case, the final aim might be the EEC of the overall production system, i.e. a series of workstations composed of identical parallel machines with a common upstream finite capacity buffer.

ACKNOWLEDGMENT

We would like to thank Comau SpA for providing the industrial case studied in Section V.

REFERENCES

[1] B. P. Center, "Annual energy outlook 2020," *Energy Information Administration, Washington, DC*, 2020.

[2] N. Frigerio and A. Matta, "Analysis on energy efficient switching of machine tool with stochastic arrivals and buffer information," *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 1, pp. 238–246, 2015.

[3] Z. Jia, L. Zhang, J. Arinez, and G. Xiao, "Performance analysis for serial production lines with bernoulli machines and real-time wip-based machine switch-on/off control," *International Journal of Production Research*, vol. 54, no. 21, pp. 6285–6301, 2016.

[4] L. Li and Z. Sun, "Dynamic energy control for energy efficiency improvement of sustainable manufacturing systems using markov decision process," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 43, no. 5, pp. 1195–1205, 2013.

[5] M. P. Brundage, Q. Chang, Y. Li, G. Xiao, and J. Arinez, "Energy efficiency management of an integrated serial production line and hvac system," *IEEE transactions on automation science and engineering*, vol. 11, no. 3, pp. 789–797, 2013.

[6] Y. Zhang, Z. Sun, R. Qin, and H. Xiong, "Idle duration prediction for manufacturing system using a gaussian mixture model integrated neural network for energy efficiency improvement," *IEEE Transactions on Automation Science and Engineering*, 2019.

[7] D. S. Szarkowicz and T. W. Knowles, "Optimal control of an M/M/s queueing system," *Operations research*, vol. 33, no. 3, pp. 644–660, 1985.

[8] A. Printezis and A. Burnetas, "Priority option pricing in an M/M/m queue," *Operations research letters*, vol. 36, no. 6, pp. 700–704, 2008.

[9] E. A. Feinber and F. Yang, "Dynamic price optimization for an M/M/k/n queue with several customer types," *ACM SIGMETRICS Performance Evaluation Review*, vol. 41, no. 3, pp. 25–27, 2014.

[10] A. Gandhi, M. Harchol-Balter, and I. Adan, "Server farms with setup costs," *Performance Evaluation*, vol. 67, no. 11, pp. 1123–1138, 2010.

[11] A. Gandhi, M. Harchol-Balter, and I. Adan, "Analysis of an M/M/k system with exponential setup times under staggered boot up," in *Madrid Conference on Queueing Theory*, vol. 2, 2010, p. 6.

[12] I. Mitrani, "Managing performance and power consumption in a server farm," *Annals of Operations Research*, vol. 202, no. 1, pp. 121–134, 2013.

[13] X. Xu and N. Tian, "The M/M/c queue with (e, d) setup time," *Journal of Systems Science and Complexity*, vol. 21, no. 3, pp. 446–455, 2008.

[14] V. J. Maccio and D. G. Down, "On optimal control for energy-aware queueing systems," in *2015 27th International Teletraffic Congress*. IEEE, 2015, pp. 98–106.

[15] V. J. Maccio and D. G. Down, "Exact analysis of energy-aware multiserver queueing systems with setup times," in *2016 IEEE 24th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS)*. IEEE, 2016, pp. 11–20.

[16] V. J. Maccio and D. G. Down, "Structural properties and exact analysis of energy-aware multiserver queueing systems with setup times," *Performance Evaluation*, vol. 121, pp. 48–66, 2018.

[17] V. J. Maccio and D. G. Down, "Asymptotic performance of energy-aware multiserver queueing systems with setup times," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 6266–6272.

[18] S. A. Lippman, "Applying a new device in the optimization of exponential queuing systems," *Operations research*, vol. 23, no. 4, pp. 687–710, 1975.

[19] R. Bellman, "A markovian decision process," *Journal of mathematics and mechanics*, vol. 6, no. 5, pp. 679–684, 1957.

[20] D. C. Montgomery, *Design and analysis of experiments*. John wiley & sons, 2017.

[21] D. Trietsch and K. R. Baker, "Pert 21: Fitting pert/cpm for use in the 21st century," *International journal of project management*, vol. 30, no. 4, pp. 490–502, 2012.