Journal of Process Control 68 (2018) 218–225

Contents lists available at ScienceDirect

Journal of Process Control

journal homepage: www.elsevier.com/locate/jprocont

A Finite Horizon Markov Decision Process Based Reinforcement Learning Control of a Rapid Thermal Processing system

D. John Pradeep[a], Mathew Mithra Noel[b,*]

[a] School of Electronics Engineering, VIT University, India
[b] School of Electrical Engineering, VIT University, India

ARTICLE INFO

Article history:
Received 17 July 2017
Received in revised form 5 May 2018
Accepted 6 June 2018

Keywords:
Reinforcement Learning
Rapid Thermal Processing
Nonlinear control
Markov Decision Process
Process control
Multivariable control

ABSTRACT

Manufacture of ultra large-scale integrated circuits involves accurate control of a challenging nonlinear Rapid Thermal Processing (RTP) system. Precise control of temperature profile and rapid ramp-up and ramp-down rates demanded by a RTP system cannot be achieved with conventional control strategies due to nonlinear and multi time-scale effects. In this paper the control of a RTP system is reformulated as an optimal multi-step sequential decision problem using the framework of finite horizon Markov decision processes and solved using a Reinforcement Learning (RL) algorithm. Three increasingly complex RL based control strategies are explored and compared with the existing state-of-the-art approach for controlling RTPs. Simulation results indicate that the approach proposed in this paper achieves superior control of the temperature profile and ramp-up and ramp-down rates for the RTP system.

1. Introduction

Modern large-scale manufacture of integrated circuits involves the challenging problem of control of nonlinear thermal processes. To achieve smaller critical lengths and flexibility in semiconductor manufacturing, precise control of temperature profiles and temperature ramp-up and ramp-down rates is necessary. This is achieved by exposing the wafer to a flexible heat source called a Rapid Thermal Processor (RTP) (Roozeboom, 1990; Vandenabeele and Maex, 1991). Control of RTP systems is critical for semiconductor industry processes such as nitration, annealing, and Chemical Vapour Deposition (CVD) to impart necessary structural and electrical properties to silicon wafer (Cho et al., 2005). RTP demands stringent temperature profile control, uniformity of wafer temperature and high ramp up/down rates (Balakrishnan, 2000).

Two temperature profiles of interest in RTP are the spike and soak shaped profiles shown in Figs. 1 and 2 respectively. A Spike shaped temperature profile is often chosen as it targets ultra-shallow junctions with precision. The features of spike shaped profile involve rapidly increasing the wafer temperature to a predefined set point T_{ref} followed by rapid cooling as shown in Fig. 1.

Soak shaped profile involves increasing temperature to a reference value T_{ref} and maintaining the temperature constant for a predefined duration, followed by cooling of the wafer surface as shown in Fig. 2. Disparate time constants associated with various components of a RTP and nonlinear radiation effects make control of RTP systems a challenging problem. RTP controllers can be designed to either achieve a target temperature profile or achieve a precise thermal budget. Thermal budget refers to the total heat energy transferred from the lamp source to the wafer during RTP.Tracking a target temperature profile closely is in general more challenging than targeting thermal budget indices (Jeng and Chen, 2013), so most controllers are designed to track precise thermal budget indices. Also Soak temperature profile control is widely discussed in literature than Spike profile control as the latter is more complex. Model based control (Balakrishnan, 2000), iterative learning control (Cho et al., 2005; Choi and Do, 2001), internal model control (Schaper and Kailath, 1999), non-linear model predictive control (Dassau et al., 2006), fuzzy based controller design for spatiotemporal control of RTP is reported in (Zhang et al., 2017), model based controller, combining a linear quadratic Gaussian (LQG) controller, a constrained iterative learning controller (ILC), and a model parameter estimator for RTP (Won et al., 2017) and gain scheduled controller for an RTP system modelled as a linear parameter varying model (Trudgen et al., 2016) are few of the important control strategies used to design an RTP controller to ensure soak shaped temperature profile on the wafer.

* Corresponding author.
E-mail addresses: johnpradeepdarsy@gmail.com (D.J. Pradeep), mathew.mithra@gmail.com (M.M. Noel).

https://doi.org/10.1016/j.jprocont.2018.06.002
0959-1524/

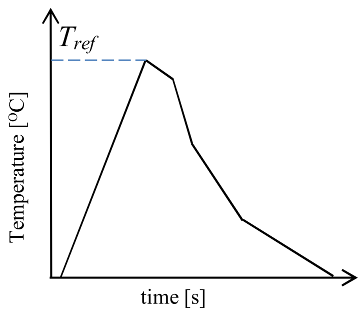

Fig. 1. Spike temperature profile.

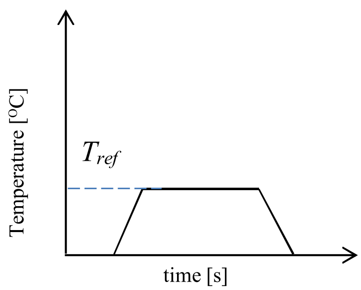

Fig. 2. Soak temperature profile.

Spike shaped temperature profile on wafer surface reduce transient offset triggered dopant diffusion and improves electrical activation (Fiory, 2002). Thus control algorithms that ensure spike profile on the wafer surface are of interest. Linear Quadratic Gaussian controller (Emami-Naeini et al., 2003), nonlinear Wiener filter model based multivariable control strategy, developed by targeting thermal budget indices (Jeng and Chen, 2013), learning control approach based on dominant modes of the system state (Xiao and Li, 2015) are few control methods reported for spike annealing. None of the models target servo control of spike shaped temperature profile. Thus new control strategies that attempt to achieve a target temperature profile directly (and hence not targetting thermal budget indices) and deal with the challenging problem of achieving the Spike temperature profile are of interest.

In the following, an improved control strategy for RTP systems based on reformulating the control problem as an optimal multistep sequential decision problem is proposed. A Reinforcement Learning (RL) algorithm is used to find an optimal nonlinear policy by executing various actions and observing the desirability of the resulting state transitions and rewards.

The paper is organized as follows: An introduction to RTP system and its non-linear model is presented in section 2. In section 3, concepts of Finite Horizon Markov Decision Process are presented. In section 4, formulation of RTP control problem as a sequential decision problem using the framework of Markov decision processes and a reinforcement learning algorithm, applied to compute an optimal control policy are presented. Finally in section 6, simulation results comparing different control strategies and discussion of result is presented.

2. Rapid thermal processing (RTP)

An RTP system consists of a chamber in which all the wafer processing operations are done. The chamber consists of heating elements (usually tungsten-halogen lamps), heat transferring quartz window and the semiconductor to be processed. Effective baking of the wafer is achieved by controlled heating from a lamp. An RTP controller must sense the temperature at different parts of the wafer and suitably adjust the power to the lamp source in order to achieve the desired temperature profile. The effect of the heating lamp source on the wafer surface temperature is complex and has been widely discussed (Huang et al 2000, Jeng chen 2013). In this paper, the lumped model given by (Huang et al 2000) is used to model the temperature dynamics on the wafer. In this model, the tungsten lamps are assumed to be grouped into 'm' equal radius annular lamp zones that can be powered independently. Wafer temperature ($\boldsymbol{x}$) uniformity can be achieved, by controlling the power supplied to each lamp zone ($\boldsymbol{p}$). The operation of a RTP system is depicted in Fig. 3. The wafer is assumed to be divided into 'n' equal radius annular rings and the relationship between the temperature on i-th wafer ring 'T_i' and the power supplied to j-th lamp zone 'P_j' is given by Eq. (1) (Jeng and Chen, 2013).

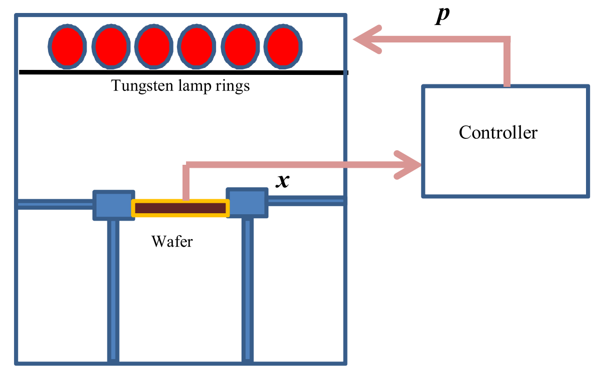

Fig. 3. A Rapid Thermal Processing (RTP) system with controller.

$$\frac{dT_i}{dt} = \frac{-\sigma}{\rho C_p d}\left(\sum_{j=1}^{m}\phi_{ij}\right)T_i^4 - \frac{-h_c}{\rho C_p d}(T_i - T_a) + \frac{1}{\rho C_p d}\left(\sum_{j=1}^{m}\phi_{ij}P_j\right) \tag{1}$$

where, σ is Stefan-Boltzmann constant, ρ is the wafer density, C_p is the heat capacity, d is the wafer thickness, T_a is the ambient temperature of cooling gas, P_j is the power given as input to j-th lamp zone and h_C is convective heat transfer coefficient.

The components φ_{ij} can be computed using Eq. (2).

$$\phi_{ij} = \frac{1}{\frac{1-\varepsilon}{\varepsilon} + \frac{1}{F_{ij}}} \tag{2}$$

Where, ε is the emissivity and F_{ij} is form factor computed using Eq. (3).

$$F_{ij} = \frac{1}{\pi}\int_{r_{L,j-1}}^{r_{L,j}}\int_{0}^{2\pi}\frac{H^2}{S_{ij}^4}r_L d\theta_{ij} dr_L \tag{3}$$

$r_{L,j}$ is the radius of the j-th lamp zone and $r_{L,j\text{-}1}$ is the radius of the j-1 th lamp zone. S_{ij} is given by Eq. (4).

$$S_{ij}^2 = H^2 + r_{w,i}^2 + r_L^2 - 2r_{w,i}r_L\cos\theta_{ij} \tag{4}$$

H is the distance between the lamp and wafer, $r_{w,i}$ is the radius of the i-th wafer zone and r_L is the lamp radius. The values of RTP system parameters used in this study and their units are provided in Table 1.

3. Finite horizon Markov decision process

Reinforcement Learning (RL) addresses the general problem of learning an optimal control policy by interacting with the environment by performing random actions and receiving rewards which indicate the desirability of the actions. RL (Watkins and Dayan, 1992; Bertsekas and Tsitsiklis, 1996; Mitchell, 1997) mimics the behaviour of an intelligent agent that learns to achieve a goal by