



Original Article

Computational fluid dynamics modeling of a wafer etch temperature control system

Henrique Oyama ^a, Kip Nieman ^a, Anh Tran ^b, Bernard Keville ^b, Yewei Wu ^b, Helen Durand ^{a,*}^a Department of Chemical Engineering and Materials Science, Wayne State University, Detroit, MI 48202, United States of America^b Lam Research Corporation, United States of America

ARTICLE INFO

Keywords:

Computational Fluid Dynamics
Process control
Reduced-order modeling

ABSTRACT

Next-generation etching processes for semiconductor manufacturing exploit the potential of a variety of operating conditions, including cryogenic conditions at which high etch rates of silicon and very low etch rates of the photoresist are achieved. Thus, tight control of wafer temperature must be maintained. However, large and fast changes in the operating conditions make the wafer temperature control very challenging to be performed using typical etch cooling systems. The selection and evaluation of control tunings, material, and operating costs must be considered for next-generation etching processes under different operating strategies. These evaluations can be performed using digital twin environments (which we define in this paper to be a model that captures the major characteristics expected of a typical industrial process). Motivated by this, this project discusses the development of a computational fluid dynamics (CFD) model of a wafer temperature control (WTC) system that we will refer to as a “digital twin” due to its ability to capture major characteristics of typical wafer temperature control processes. The steps to develop the digital twin using the fluid simulation software ANSYS Fluent are described. Mesh and time independence tests are performed with a subsequent benchmark of the proposed ANSYS model with etch cooling system responses that meet expectations of a typical industrial cooling system. In addition, to quickly test different operating strategies, we propose a reduced-order model in Python based on ANSYS simulation data that is much faster to simulate than the ANSYS model itself. The reduced-order model captures the major features of the WTC system demonstrated in the CFD simulation results. Once the operating strategy is selected, this could be implemented in the digital twin using ANSYS to view flow and temperature profiles in depth.

1. Introduction

In the semiconductor manufacturing industry, plasma etching is a critical step that selectively removes materials to generate increasingly small and complex features such as desired high-aspect ratio on the wafer (Wu et al., 2010). This step is becoming increasingly sophisticated with advances in plasma sources and control designs (Kanarik et al., 2018). Specifically, the modules for typical etching processes are developed to increase wafer etching performance and to control material critical dimensions and microstructures (Tachi et al., 1988). In particular, as we can see in Fig. 1 (Tachi et al., 1988), there is an exponential relationship between etch rate and temperature for the silicon and photoresist materials. This demonstrates that precise control of the wafer temperature is needed to meet industrial specifications in terms of the shape, sharpness, and precision of features at modern technology nodes (Tachi et al., 1988). Moreover, we can observe that high etch rates of silicon and very low etch rates of the photoresist are achieved

by keeping the wafer temperature under cryogenic conditions (e.g., below -100°C) (Tachi et al., 1988; Tinck et al., 2017). In light of this, the next-generation etching process will require careful consideration of cryogenic process modules that may be incorporated into current and novel wafer temperature control (WTC) systems to operate in a wider range of operating conditions. To achieve this goal, the selection and evaluation of control designs, material, and operating costs under different operating strategies are needed. As indicated in Lee et al. (2014), as the technology for plasma etching advances, there will be an ongoing need to decrease overall cost structure and increase innovation in multiple areas of the wafer manufacturing process, which includes exploration of advanced wafer temperature control strategies.

Modeling of atomic layer deposition (e.g., Ding et al. (2019, 2021, 2020)) and advances in wafer temperature measurements for process control (e.g., Sun and Gabriel (2002), Wang et al. (2006)) have been reported in the literature. However, to the best of the authors’

* Corresponding author.

E-mail address: helen.durand@wayne.edu (H. Durand).

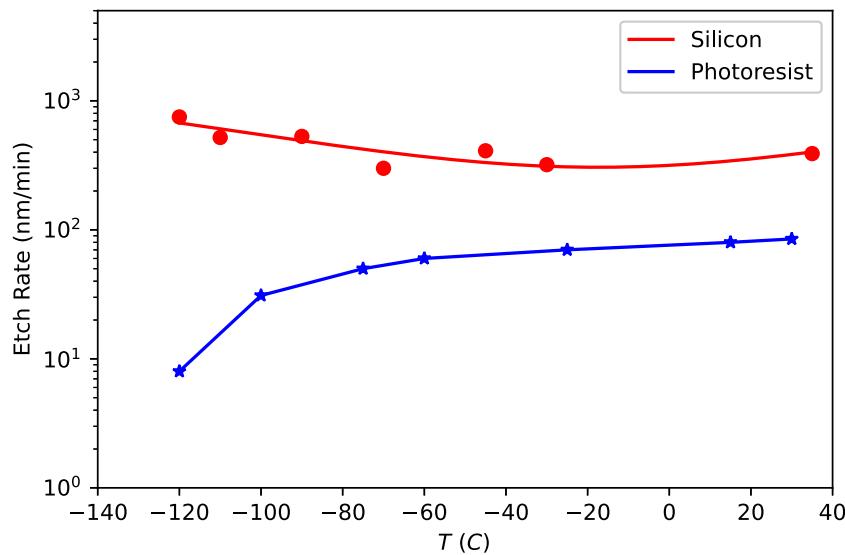


Fig. 1. Relationship between etch rate and temperature for silicon and photoresist (Tachi et al., 1988).

knowledge, there have not been studies on the development of a fully digital model of typical WTC systems with the aim to explore control strategies for wafer temperature control. Applications of etch process control include the works of Mozumder and Barna (1994) and Zhang et al. (2008). In Mozumder and Barna (1994), a statistical feedback control that targets quality characteristics (e.g., etch rate) instead of desired process states (e.g., temperature and flow rates) has been developed which utilizes fitting polynomial models for statistical quality control. In Zhang et al. (2008), a model-based controller for the etch process has been proposed to reduce the critical dimension variations on across-wafer levels. The process/equipment models developed for testing the controllers above were reduced-order models based on experimental data. To allow for more flexibility in exploring different control strategies in etching systems that do not need to rely on the actual physical system for control testing with a high level of confidence, a digital twin is needed for data generation and further analysis of operating strategies. As a step towards this direction, the computational fluid dynamics (CFD) method using ANSYS Fluent may be used to build a digital model of a typical etch cooling system.

Motivated by the above, we develop a modular ANSYS/CFD model that captures the major features of typical etch cooling systems. We benchmarked the results with expected typical etch cooling system responses after mesh and time independence tests were performed. Furthermore, as the ANSYS model takes significant computing time to be simulated, we developed a reduced-order model based on ANSYS simulation data that is much faster to be simulated using the Python programming language, which utilizes several assumptions but still captures the major features of the WTC system developed in ANSYS/CFD software. This can enable fast testing of a variety of potential operating conditions for the WTC with a high level of confidence. Future work will investigate other control strategies, in particular model predictive control (MPC) schemes that may be used to minimize energy costs of the WTC system, using the digital twin developed in ANSYS and depict details of flow and temperature patterns that would result under control strategies determined to be suitable when selected with the Python model. We expect that the advanced control formulations such as MPC would outperform classical PID controllers typically used in semiconductor manufacturing in cases where there are significant changes in the operating condition (e.g., rapidly entering regions of operation that the original tuning parameters of the PID controllers may not be designed for), and would provide flexibility to allow for changes in operating conditions without the need to set up several tuning strategies or other heuristic approaches. We expect that this

project can be a baseline for developing digital models in ANSYS, in particular for etch cooling systems, and may serve as a platform to explore and analyze the design of the control and cooling systems for ensuring that manufacturing specifications are met while reducing costs.

2. Wafer temperature control for plasma etching: Process overview

In this work, we consider part of a temperature control system for the wafer etching process (to be subsequently referred to as the wafer temperature control system or WTC) with the structure in Fig. 2. This WTC manipulates valves for a heating/cooling system. The goal of this system is to: 1) maintain the temperature set-point for the etching process and 2) achieve the fast temperature switching required throughout the etching process, for the range of temperatures currently utilized, by providing cooling fluid at the required temperature to the electrostatic chuck (ESC) in the etching process. Specifically, for this design, as shown in the figure, two fluids are mixed before being routed to the ESC, where one is colder than the other (the fluid exiting Temperature Control Unit (TCU) Channel (Ch.) 1 is colder than the fluid exiting Temperature Control Unit (TCU) Ch. 2). The temperatures leaving TCU Ch. 1 and TCU Ch. 2 are controlled using internal model controllers (IMC). Though the flow rates exiting TCU Ch. 1 and TCU Ch. 2 are also controlled, the dynamics of the flow controllers are considered to be fast enough that the flow rate exiting each TCU channel is considered to be a constant value equal to its set-point.

The fluids leaving TCU Ch. 1 and TCU Ch. 2 are diverted via a three-way diverter valve in one of two directions: one which travels toward the T-junction which precedes the electrostatic chuck (ESC), and one which is recycled back to the TCU Channel from which it came. The positions of these diverter valves are set by a multivariable PID control setup to cause the temperature and flow rate out of the T-junction to hit desired set-points. A level controller splits the fluid leaving the ESC in two directions (toward either TCU Ch. 1 or TCU Ch. 2) using a diverter valve to attempt to keep the levels in both TCU channels close to each other. Together, the four controllers keep the temperatures of the fluids exiting the two TCU's at the values of T_{1sp} and T_{2sp} , which change over time to facilitate the switches in temperature required at the ESC.

3. Modeling of the wafer temperature control for plasma etching

The model of the wafer temperature control system described in this work is developed in the CFD software ANSYS Fluent. However, aspects

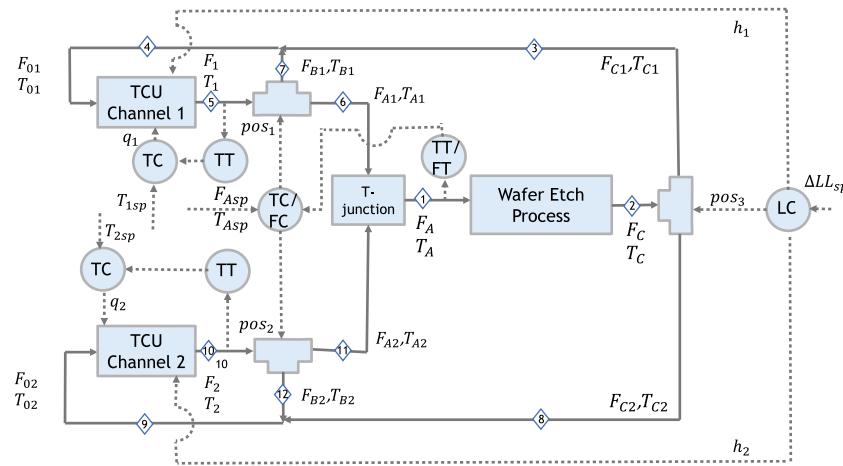


Fig. 2. Process diagram of the WTC.

of the wafer temperature control process are represented in ANSYS Fluent using user-defined functions (UDF's). Therefore, the description of the model of the WTC process has two components: one is a set of model equations used in the UDF's to represent dynamics of the TCU's, boundary conditions for the valves, boundary conditions for the ESC outlet, and controllers for the process; another is a CFD simulation including features such as the turbulence models selected and geometry and meshing used for ANSYS Fluent. This section will describe the equations which are used in the UDF's of the ANSYS Fluent model; the next section will describe the CFD simulation in greater depth.

3.1. Dynamic models for TCU Channels 1 and 2

The following equations represent how the fluid level, volume, and temperature change over time in the two TCU Channels:

$$\frac{dh_i}{dt} = \frac{f_{0i} - f_i}{A_i} \quad (1a)$$

$$V_i = h_i A_i \quad (1b)$$

$$\frac{dT_i}{dt} = \frac{f_{0i}T_{0i} - f_iT_i + q_i(t - \theta)}{V_i} \quad (1c)$$

where h_i is the fluid level in the TCU Ch. i ($i = 1, 2$), V_i is the volume of the fluid in the TCU Ch. i , T_i is the outlet temperature of TCU Ch. i (in $^{\circ}\text{C}$), f_{0i} is the volumetric flow rate entering TCU Ch. i , f_i is the volumetric flow rate leaving TCU Ch. i (assumed to be fixed), A_i is the cross-sectional area of the TCU Ch. i , T_{0i} is the inlet temperature of TCU Ch. i (in $^{\circ}\text{C}$), ρ is the coolant density, C_p is the coolant heat capacity, q_i is the heat rate provided to the TCU Ch. i , and θ is a time delay. The reason that the temperature in Eq. (1c) is assumed to be in degrees C is that Eq. (1c) is derived from an energy balance where a term containing the reference temperature for the enthalpy of the coolant is neglected. If this reference temperature is assumed to be 0°C , the neglect of this term can be warranted. The time delay is applied because many processes can be described as first-order-plus-dead-time processes, and industrial chillers are known to have a large number of potential sources of delay (e.g., [Controls \(2021\)](#)), so that we postulate that the chiller models can also be described this way (i.e., there is a delay between the change in the heat input and its impact on the chiller outlet temperature). To cause the response of the process to a set-point change to delay, we apply the lag in the input to achieve the same effect as having the process dynamics lag at an input change. We will consider a delay of slightly less than a minute (40 s). Thus, Eq. (1c) can be transformed to the Laplace domain to give the following first-order-plus-dead-time form:

$$\bar{T}_i(s) = \frac{1}{F_{is}C_p} \frac{e^{-\theta s}}{\tau s + 1} \bar{q}_i(s) \quad (2)$$

Table 1
Parameters for the TCU Ch. i ($i = 1, 2$) models.

Parameter	Value	Unit
ρ	1793	kg/m^3
C_p	1038	J/kg K
f_i	16.8	L/min
A_i	0.0984	m^2
τ	50	s

where $\bar{T}_i = T_i - T_{is}$ (for a steady-state temperature in TCU Ch. i of T_{is}), $\bar{q}_i = q_i - q_{is}$ (for a steady-state heat rate input in TCU Ch. i of q_{is}), $F_{is} = f_{is}\rho$ is the steady-state mass flow rate entering TCU Ch i and the time constant $\tau = V_{is}/f_{is}$ is obtained by dividing a steady-state volume of the fluid in the TCU Ch. i (assumed to be $V_{is} = 0.014 \text{ m}^3$ or 14 L) by the steady-state volumetric flow rate entering TCU Ch. i (taken to be $0.00028 \text{ m}^3/\text{s}$, or 16.8 L/min).

To obtain the physical properties of the coolant such as ρ and C_p that appear in the above equations, a coolant had to be selected for the simulations. The coolant Fluorinert has found important applications as an electronic liquid for etching processes ([Acota, 2021](#); [Watanabe et al., 1997](#)) and was therefore selected for this study. Specifically, we consider Fluorinert FC-770 due to its wide operating temperature range ([3M, 2021](#)), which provides flexibility for this study. Its relevant physical parameters are presented in Table 1, along with other process parameters assumed for the TCU's (the areas noted are approximate, as they are obtained by assuming that the maximum reservoir volume for the chiller is 20 L (0.02 m^3), and that the total height of the reservoir is 8 inches (0.2032 m)). The heat which can be supplied to the TCU's will be limited; it is therefore important to include upper and lower bounds on the heat rates for TCU Channels 1 and 2 (which are assumed to be -10 kW and 10 kW , to be on an order of magnitude with cooling capacities typically reported for chillers such as [BROAD Group \(2021\)](#)), to examine the effect on the system's ability to heat and cool the heat transfer fluid in the WTC system. For example, to investigate how large the assumed cooling capacity is expected to be compared to what would be needed to remove energy from the heat transfer fluid, consider a case with a heat exchanger instead of a reservoir used for removing energy. To estimate the heat duty required in such a case, we use the equation for heat duty of a heat exchanger:

$$Q = f \rho C_p \Delta T \quad (3)$$

where Q is the rate of heat input/removal in a heat exchanger, f is a volumetric flow rate through the heat exchanger, and ΔT is the expected difference in temperature between the inlet and outlet of the heating/cooling system. Given that a typical temperature range of plasma etching is $20\text{--}50^{\circ}\text{C}$ ([Kuo and Lee, 2001, 2004](#); [Roozeboom](#)

et al., 2015; Sun et al., 2015), the heat duty required to cool Fluorinert FC-770 from 50°C to 20°C at a volumetric flow rate of 17 L/min is approximately 16 kW. This heating duty is more than the maximum which could be provided by q_1 , indicating that bounds on q_1 and q_2 may play a role in the effectiveness of the cooling/heating system.

For the control designs, internal model control (IMC) schemes are used to control the temperatures T_1 and T_2 leaving the TCU Ch. 1 and 2, respectively, by adjusting the heat rates applied to TCU Ch. 1 and 2. This will be described in more detail in a subsequent section.

Remark 1. T_{0i} defined above is not fixed and it accounts for the mixing effects at the T-junctions between the fluid leaving the ESC unit via valve pos_3 in Fig. 2 and the fluid returning to the TCU Channels via the valves pos_1 and pos_2 . In particular, the temperature at the mixing points has been modeled based on a flow rate-weighted average temperature as will be clarified in the next section (e.g., Eq. (7b)).

3.2. Dynamic models for three-way valves, mixing points, velocity and temperature profiles, sensors, and ESC unit

We will now describe the modeling for the next stage of the WTC system, in which the fluids that are leaving the TCU Channels, at mass flow rates F_1 and F_2 and temperatures T_1 and T_2 , travel through piping to the diverter valves labeled pos_1 and pos_2 (since the valve positions would be adjusted by a controller). The mass flow rates leaving the three-way diverter valves named pos_1 and pos_2 in the process diagram of Fig. 2 are represented by the following equations:

$$F_{Ai} = F_i \gamma_i \quad (4a)$$

$$F_{Bi} = F_i (1 - \gamma_i) \quad (4b)$$

where γ_i ($i = 1, 2$) is the fraction of mass flow from TCU Ch. i that goes to the T-junction, F_i is the mass flow rate entering the valve pos_i (we note that this is distinct from f_i , which is the volumetric flow rate), F_{Ai} is the mass flow rate that goes to the T-junction, and F_{Bi} is the mass flow rate recycled back to the TCU Ch. i . Because the Fluorinert density is taken to be constant throughout the cooling/heating system, the mass flow rates in Eq. (4a) can be divided by the fluid density to give volumetric flow rates through the valves. Eq. (4a) neglects any effects of pressure drop on the fluid flow rate through the valve. γ_1 and γ_2 , as percentages of flow diverted by the valve, will be adjusted based on the set-points $\gamma_{1,sp}$ and $\gamma_{2,sp}$ computed by controllers. It is assumed that the valve setup is such that the position of the valve does not respond instantaneously to the controller signal. Rather, the controllers will compute set-point values for the valve positions and communicate these to the valves. The valves will then be modeled as first-order processes with a time constant $\tau_\gamma = 1$ s. The dynamic equations representing the values of $\bar{\gamma}_1$ and $\bar{\gamma}_2$ (where $\bar{\gamma}_1 = \gamma_1 - \gamma_{1s}$ and $\bar{\gamma}_2 = \gamma_2 - \gamma_{2s}$, with γ_{1s} and γ_{2s} as the steady-state values of γ_1 and γ_2) are given by the following differential equations:

$$\frac{d\bar{\gamma}_i}{dt} = \frac{1}{\tau_\gamma} (\bar{\gamma}_{i,sp} - \bar{\gamma}_i) \quad (5)$$

for $i = 1, 2$, where the set-point values of the valve position from the controllers are denoted by $\bar{\gamma}_{1,sp}$ and $\bar{\gamma}_{2,sp}$. To model how the temperature changes along each pipe, we observe that its time constant and time delay depend on the flow rate. In particular, a higher flow rate would cause the fluid to travel faster throughout a pipe compared to a case where a lower flow rate is applied to the same pipe. This means that a change in the inlet temperature of the pipe would propagate faster to the end of a pipe if a high velocity flow is implemented, indicating that space as well as time are key for modeling the temperature at each position in the pipe. As we consider average temperatures at a cross-sectional area of a pipe, only variations in the horizontal direction are assumed. The temperature changes along a pipe over time can be modeled as a plug flow equation for temperature as follows:

$$\frac{\partial T_j}{\partial t} = -v_j \frac{\partial T_j}{\partial z} \quad (6)$$

where T_j is the temperature at a certain position z in the pipe and v_j is the velocity flow in the pipe. We assume that the change in mass flow throughout a pipe is instantaneous. The mass flows (with rates F_{A1} and F_{A2}) leaving the three-way diverter valves pos_1 and pos_2 go through several feet of piping and then are mixed in a T-junction. A mass balance around the T-junction assuming that there is no accumulation of mass within the T-junction gives the following equation:

$$F_A = F_{A1} + F_{A2} \quad (7a)$$

$$T_A|_{z=initial} = \frac{F_{A1}T_{A1}|_{z=end} + F_{A2}T_{A2}|_{z=end}}{F_{A1} + F_{A2}} \quad (7b)$$

where F_A is the total mass flow rate leaving the T-junction and $T_A|_{z=initial}$ is the temperature at the mixing point of the T-junction (a flow rate-weighted average temperature), and $T_{A1}|_{z=end}$ and $T_{A2}|_{z=end}$ are the fluid temperatures from the cold and hot valves, respectively, right before the mixing point of the T-junction. After the fluid leaves the T-junction, it travels through several feet of piping to the ESC. It is assumed that the temperature and flow rate sensors are placed at the end of this piping, right before the ESC, so that the temperature and flow rate are measured after the fluid has reached fully developed flow behavior. Specifically, the flow rate sensor will be modeled as a first-order process with no time delay and a time constant of τ_{F_i} , and the temperature sensor will be modeled as a first-order process with no time delay and a time constant of τ_{T_i} . The equations describing the flow rate sensor reading $F_{A,sensor}$ and the temperature sensor reading $T_{A,sensor}$ are as follows:

$$\frac{dF_{A,sensor}}{dt} = \frac{1}{\tau_{F_i}} (F_A - F_{A,sensor}) \quad (8)$$

$$\frac{dT_{A,sensor}}{dt} = \frac{1}{\tau_{T_i}} (T_A|_{z=end} - T_{A,sensor}) \quad (9)$$

where $T_A|_{z=end}$ is a temperature of the fluid after the T-junction.

The final portion of the WTC system to be modeled is after the ESC, in which a controller (the control law parameters will be defined in a subsequent section) is used to adjust the return valve position (for the valve labeled pos_3 in Fig. 2) to maintain an expected difference between the volumes (levels) in the two TCU Channels. This part of the process introduces several new components, including the temperature rise across the ESC, the three-way diverter valve labeled pos_3 , and the controller for pos_3 . $T_A|_{z=end}$ is computed based on Eq. (6) with $T_A|_{z=initial}$ as a boundary condition. The mass flow rate of the coolant leaving the ESC is assumed to be the same as that entering the ESC (i.e., $F_C = F_A$ in Fig. 2), but there is a temperature rise across the ESC as the cooling fluid absorbs heat from the wafer etch process. The temperature of the fluid leaving the ESC ($T_C|_{z=initial}$) must be higher than the temperature $T_A|_{z=end}$ of fluid entering. If $T_A|_{z=end}$ and $T_C|_{z=initial}$ are at steady-state, we assume that the temperature increase across the ESC can be modeled using the simplified model of a heat duty applied at the ESC. Specifically, again using Eq. (3) to determine the temperature of the fluid leaving the wafer etch process (WEP), we compute the steady-state value of $T_C|_{z=initial}$ for a given value of $T_A|_{z=end}$ as:

$$T_C|_{z=initial} = T_A|_{z=end} + \frac{q_c}{F_A C_p} \quad (10)$$

where q_c is the WEP heat rate. This means at the initial time, $T_{C,ss} = T_{Ass} + \frac{q_c}{F_A C_p}$, where $T_{Ass} = 30^\circ\text{C}$ is the steady-state value of $T_A|_{z=end}$, and $T_{C,ss}$ is the steady-state value of $T_C|_{z=initial}$. However, if there is a step change in T_A , we assume that there is some lag before $T_C|_{z=initial}$ reaches its final steady-state value due to factors such as transport lag, conduction, convection, and heat transfer from a solid to a fluid. The dynamics of the temperature change are assumed to be captured by a first-order model with a time constant of τ_{T_c} as follows:

$$\frac{d\bar{T}_C|_{z=initial}}{dt} = \frac{1}{\tau_{T_c}} (\bar{T}_A|_{z=end} - \bar{T}_C|_{z=initial}) \quad (11)$$

where $\bar{T}_C|_{z=initial}$ and $\bar{T}_A|_{z=end}$ are the deviation variables representing $T_C|_{z=initial}$ and $T_A|_{z=end}$ in deviation form from their steady-states $T_{Ass} + \frac{q_e}{F_A C_p}$ and T_{Ass} , respectively, and $\bar{T}_C|_{z=initial}$ is delayed by a time θ_{T_c} before being used as the boundary condition for the ESC outlet.

A PID controller with a derivative filter computes set-points for the valve position of the three-way diverter valve pos_3 . The mass flow rates of the fluids leaving pos_3 are represented by the following equations:

$$F_{C1} = F_C \gamma_3 \quad (12a)$$

$$F_{C2} = F_C (1 - \gamma_3) \quad (12b)$$

where γ_3 is the fraction of mass flow recycled to TCU Ch. 1, F_C is the mass flow rate entering the valve pos_3 , F_{C1} is the mass flow rate that goes back to TCU Ch 1, and F_{C2} is the mass flow rate that goes back to TCU Ch 2. Because the Fluorinert density is taken to be constant throughout the cooling/heating system, the mass flow rates in Eq. (12b) can be divided by the fluid density to give volumetric flow rate through the valve. Eq. (12b) neglects any effects of pressure drop on the fluid flow rate through the valve. γ_3 can be considered to have a relationship to the valve, since it is the percent of flow diverted in the direction of TCU Channel 1. Its steady-state value γ_{3s} is 0.5. Though a set-point value $\gamma_{3,sp}$ is computed by a controller, the valve position does not respond instantaneously, but instead is given by the following first-order transfer function:

$$\frac{d\bar{\gamma}_3}{dt} = \frac{1}{\tau_\gamma} (\bar{\gamma}_{3,sp} - \bar{\gamma}_3) \quad (13)$$

where $\tau_\gamma = 1$ s, $\bar{\gamma}_3 = \gamma_3 - \gamma_{3s}$ and $\bar{\gamma}_{3,sp} = \gamma_{3,sp} - \gamma_{3s}$ (the set-point value for γ_3 computed by the controller).

Remark 2. The dynamic models for the mixing points and plug flow equation for the temperature profiles along the pipes were not included in the UDF's since they are captured directly in the ANSYS Fluent simulations. However, these dynamic models will be used to develop a reduced-order model of the WTC system, which will be described in Section 5.

3.3. Temperature, flow, and level control in the WTC system

For the control designs, IMC formulations are used to control the temperatures T_1 and T_2 leaving the TCU Ch. 1 and 2, respectively, and maintain their values at the desired set-points. The set-points in this study are selected to be $T_{1,sp}$ and $T_{2,sp}$, both of which must be within the typical operating range of Fluorinert FC-770, with the second temperature at the upper bound of that operating range (3M, 2021). The IMC's were designed based on the first-order-plus-dead-time process models of Eq. (2), following a procedure similar to that in Bequette (1999) (Example 7.3) for the IMC design. The equivalent transfer function for IMC is given by $G_{ci}(s) = \frac{G_{II}(s)}{1 - G_{pi}(s)G_{II}(s)}$, where $G_{pi}(s)$ is the assumed process transfer function for the i th TCU channel, and $G_{II}(s)$ is a filter for that TCU channel multiplied by the invertible part of G_{pi} .

To obtain G_{ci} , the dead time in the original model (which is the part of the right-hand side of Eq. (2) which multiplies \bar{q}_i) is approximated by a first-order Padé approximation ($e^{-\theta s} \approx \frac{1-0.5\theta s}{1+0.5\theta s}$) before splitting the model into the invertible and non-invertible parts. The model with the first-order Padé approximation applied is G_{pi} . Subsequently, the invertible part of the process model is taken to be all of G_{pi} except $1 - 0.5\theta s$ (which forms the non-invertible part). The filters are taken to be $\frac{1}{\lambda_i s + 1}$, and a PID equivalent of the resulting controller is obtained by expanding the terms in the numerator of the resulting G_c and dividing by the denominator to give three terms which, when transformed back to the time domain, give the following control law:

$$\bar{q}_i(t) = F_{is} C_p \left[\frac{\tau + 0.5\theta}{\lambda_i + 0.5\theta} (T_{isp} - T_i) + \right. \quad (14a)$$

$$\left. \frac{0.5\tau\theta}{\lambda_i + 0.5\theta} \frac{d(T_{isp} - T_i)}{dt} \right]$$

where, as shown in Eq. (1c), $q_i = \bar{q}_i + q_{is}$ is delayed by $\theta = 40$ s and applied to the TCU's of Eq. (1) to represent the process delay under the control action of Eq. (14). $q_i(t)$ is saturated at its minimum or maximum bounds (-10 kW to 10 kW).

To manipulate the three-way diverter valves before the ESC, PID controllers with derivative filters (and accounting for controller interactions) are used. Specifically, two PID controllers with derivative filters determine the signals u_1 and u_2 , which are then combined to set the set-point values of $\bar{\gamma}_1$ and $\bar{\gamma}_2$ to control the temperature $T_A|_{z=end}$ and flow rate F_A entering the ESC. The control laws u_1 and u_2 in the Laplace domain are:

$$u_i(s) = K_i \left(1 + \frac{1}{\tau_{Ii}s} + \frac{\tau_{Di}s}{\alpha_i \tau_{Di}s + 1} \right) e_i(s) \quad (15)$$

where $e_i(s)$ is the Laplace transform of $T_{Asp} - T_{A,sensor}$ (for $i = 1$) or of $F_{Asp} - F_{A,sensor}$ (for $i = 2$). K_i , τ_{Ii} , and τ_{Di} ($i = 1, 2$) are the proportional gain, integral time, and derivative time of the i th PID controller, and α_i is the derivative filter parameter that reduces the sensitivity of the control calculations to measurement noise (Seborg et al., 2017). The set-point T_{Asp} is taken to be 30 C (303.15 K), since that is toward the lower end of the typical 20–50 C range for plasma etching mentioned in Section 3.1. The set-point F_{Asp} is taken to be 16.8 L/min (0.50204 kg/s for Fluorinert FC-770). Upon multiplying the denominator of the right-hand sides on both sides and taking the inverse Laplace transform, the following control laws are developed:

$$\begin{aligned} \alpha_1 \tau_{I1} \tau_{D1} \frac{d^2 u_1(t)}{dt^2} + \tau_{I1} \frac{du_1(t)}{dt} = K_1 & \left[(T_{Asp} - T_{A,sensor}) + \right. \\ & \left. (\tau_{I1} + \alpha_1 \tau_{D1}) \frac{d(T_{Asp} - T_{A,sensor})}{dt} + \right. \\ & \left. (1 + \alpha_1) \tau_{I1} \tau_{D1} \frac{d^2(T_{Asp} - T_{A,sensor})}{dt^2} \right] \end{aligned} \quad (16)$$

$$\begin{aligned} \alpha_2 \tau_{I2} \tau_{D2} \frac{d^2 u_2(t)}{dt^2} + \tau_{I2} \frac{du_2(t)}{dt} = K_2 & \left[(F_{Asp} - F_{A,sensor}) + \right. \\ & \left. (\tau_{I2} + \alpha_2 \tau_{D2}) \frac{d(F_{Asp} - F_{A,sensor})}{dt} + \right. \\ & \left. (1 + \alpha_2) \tau_{I2} \tau_{D2} \frac{d^2(F_{Asp} - F_{A,sensor})}{dt^2} \right] \end{aligned} \quad (17)$$

where it is assumed that $u_1(0) = u_2(0) = \frac{du_1}{dt}(0) = \frac{du_2}{dt}(0) = 0$. Since changes in γ_1 and γ_2 impact both F_A and T_A , $\bar{\gamma}_{2,sp}$ is set to half the output of the PID flow controller of Eq. (17) plus half the output of the PID temperature controller of Eq. (16) and $\bar{\gamma}_{1,sp}$ is set to half the output of the PID flow controller of Eq. (17) minus half the output of the PID temperature controller of Eq. (16). Thus, under these control actions, an increased flow demand would result in both control valves opening further, and an increase in temperature demand would result in the three-way diverter valve pos_2 opening and the three-way diverter valve pos_1 closing. This gives the following equations for $\bar{\gamma}_{1,sp}$ and $\bar{\gamma}_{2,sp}$:

$$\begin{aligned} \bar{\gamma}_{1,sp} &= \frac{u_2 - u_1}{2} \\ \bar{\gamma}_{2,sp} &= \frac{u_2 + u_1}{2} \end{aligned} \quad (18a)$$

The valves were assumed to all be half open at the initial condition corresponding to steady-state operation (i.e., $\gamma_1 = 0.5$ and $\gamma_2 = 0.5$, where the steady-state values of γ_1 and γ_2 (γ_{1s} and γ_{2s}) are also 0.5). Bounds were placed on the process inputs such that $\gamma_{1,sp}$ and $\gamma_{2,sp}$ could only vary between 0 and 1. These bounds were used to saturate $\gamma_{1,sp}$ and $\gamma_{2,sp}$ (as well as γ_1 and γ_2) at their bounds if the computed inputs exceeded these bounds.

The control actions u_1 and u_2 were assumed to have zero initial values. u_1 was derived using the following equations reflecting backward finite difference approximations of all derivatives in Eq. (16):

$$u_1(t) = \frac{(-\alpha_1 \tau_{I1} \tau_{D1})(-2u_1(t - h_p) + u_1(t - 2h_p))}{DEN_1 h_p^2} + \frac{\tau_{I1} u_1(t - h_p)}{DEN_1 h_p} + \frac{K_1(T_{Asp} - T_{A,sensor}(t))}{DEN_1} + \frac{K_1(\tau_{I1} + \alpha_1 \tau_{D1})(T_{A,sensor}(t - h_p) - T_{A,sensor}(t))}{DEN_1 h_p} + (K_1(1 + \alpha_1) \tau_{I1} \tau_{D1}) * \frac{(2T_{A,sensor}(t - h_p) - T_{A,sensor}(t) - T_{A,sensor}(t - 2h_p))}{DEN_1 h_p^2} \quad (19)$$

where h_p is the sampling period size and

$$DEN_1 = \frac{\alpha_1 \tau_{I1} \tau_{D1}}{h_p^2} + \frac{\tau_{I1}}{h_p} \quad (20)$$

The same equation was used for u_2 , except with the subscript 1 replaced by 2 for the controller parameters, T_{Asp} replaced by F_{Asp} , and $T_{A,sensor}$ replaced by $F_{A,sensor}$. Though Eqs. (19)–(20) are a potential implementation of u_1 and u_2 , the implementation of u_1 and u_2 used multiplies the numerator and denominator of each term in Eqs. (19)–(20) by h_p^2 to remove the integration step size from the denominator.

Finally, the PID controller with a derivative filter that is used to adjust $\tilde{\gamma}_3$ to control the TCU's fluid levels (and therefore volume since the cross-sectional area is constant) is as follows:

$$\alpha_3 \tau_{I3} \tau_{D3} \frac{d^2 \tilde{\gamma}_{3,sp}(t)}{dt^2} + \tau_{I3} \frac{d \tilde{\gamma}_{3,sp}(t)}{dt} = K_3 \left[\epsilon + (\tau_{I3} + \alpha_3 \tau_{D3}) \frac{d \epsilon}{dt} + (1 + \alpha_3) \tau_{I3} \tau_{D3} \frac{d^2 \epsilon}{dt^2} \right] \quad (21a)$$

where $\epsilon = \Delta LL_{sp} - (h_1 - h_2)$, ΔLL_{sp} is the set-point for the difference in liquid level between TCU Ch. 1 and TCU Ch. 2 (which is set to be zero), and K_3 , τ_{I3} , and τ_{D3} are the proportional gain, integral time, and derivative time of the controllers, with α_3 as the derivative filter parameter.

Remark 3. The control tuning values for each IMC and PID controller are presented later in Section 5 after testing the performance and robustness of the controllers.

4. Wafer temperature control for plasma etching: ANSYS model

The WTC system described in the prior sections has been developed in ANSYS Workbench 2020 R1 and run using the Wayne State Grid (or Wayne State's High Performance Computing cluster). The Wayne State Grid is connected to the global research community via Wayne State's science DMZ, Internet2, and the Michigan LambdaRail, and utilizes the Slurm job scheduler for job submission. The Wayne State Grid contains several resources, including the possibility to run ANSYS Fluent simulations using either batch scripts or interactive applications with ANSYS Workbench. We utilize PuTTY (a Windows-based client) to connect to the Wayne State Grid and WinSCP for file transfers. In particular, as part of the CFD methodology, the following steps must be rigorously followed: (1) Problem identification step: define goals and identify domain; (2) Pre-processing step: create a solid model of the domain, generate mesh, set up the physics and solver settings; (3) Post-processing step: examine the results and consider revisions to the model. In terms of modeling goals, we are seeking to develop a modular digital twin, whose dynamics and control tunings can be adjusted easily, and that can serve as a testbed for exploring different operating strategies for typical WTC systems. This digital system has been modeled using ANSYS Fluent with a number of assumptions in the modeling which will be clarified in the next sections involving the pre-processing procedure.

Remark 4. A digital twin is typically defined as a virtual representation of a process system that spans its lifecycle, is built/updated from real-time process data, and utilizes simulation tools and reasoning to aid decision-making and guide novel solutions (Singh et al., 2021). Though the modeling of the WTC proposed in this work does not strictly follow this definition, the WTC model using ANSYS Fluent is named "digital twin" due to its ability to capture major process characteristics of typical wafer temperature control systems.

Remark 5. To benchmark the proposed WTC model, which will be presented in subsequent sections, the typical process responses (e.g., time constants and time delays described for the WTC system) are based on normalized industrial data from our industry collaborators.

4.1. ANSYS model: Digital twin geometry

The system developed in this work is based on a typical industrial-scale dynamic wafer temperature control system. In particular, the WTC system contains connected rows of pipes with an internal diameter of 2 inches (50.8 mm). The pipe lengths have been chosen such that the fluid flow is fully developed towards the end of each pipe in the WTC system. We define that a fluid achieved a fully developed flow when the velocity profiles appear similar along the axial direction at the end of the pipe and the temperature across cross-sectional areas towards the end of the pipe do not vary more than 2%. The final geometry was chosen after simulating alternative pipe lengths for the WTC system (the post-processing simulation results are shown in a subsequent section). Inside these pipes, the coolant Fluorinert FC-770 flows to regulate the temperature of the ESC. The geometry of the WTC system has been created via SpaceClaim in ANSYS. We determined that, to have a fully developed flow in the pipes leaving the T-junctions, a pipe length of at least 1 m is needed when the realizable $k-\epsilon$ turbulence model with enhanced wall treatment is used (the choice of this model is justified in Section 4.3). Fig. 3 shows a WTC geometry in which the pipes leaving the mixing points are 1 m long. To reduce computing time and still have the fully developed flow at each pipe in the WTC, the sizes of the return pipes after the elbows were decreased, as the flow was checked to be fully developed before reaching the mixing points. The selected WTC geometry is represented in Fig. 4, where the length of the return pipes after the elbows have been reduced. This geometry of the WTC system has been used as the baseline case for the simulations described in this paper.

In Fig. 4, T_{01} and T_{02} represent the inlet temperature whereas T_1 and T_2 correspond to the outlet temperature of the TCU Ch. 1 and TCU Ch. 2, respectively. T_A is the inlet temperature of the ESC and T_c is the outlet temperature of the ESC. T_{A1} , T_{B1} , T_{A2} , T_{B2} , T_{C1} , and T_{C2} are the temperatures leaving the cold, hot, and level three-way valves. The dynamics of the valves, TCU Channels, and ESC have been modeled via user-defined functions (UDFs), according to the dynamic equations described in Section 3, which use a C program that can be dynamically loaded with ANSYS Fluent. Specifically, in this system representation, the following "gaps" in Fig. 4 are replaced by the dynamic equations described in Section 3 via UDFs: the spaces between T_{A1} and T_{B1} , T_{A2} and T_{B2} , and T_{C1} and T_{C2} correspond to the dynamics of the three-way valves, the spaces between T_{01} and T_1 and T_{02} and T_2 correspond to the dynamics of the TCU Ch. 1 and TCU Ch. 2, respectively, and the space between T_A and T_c corresponds to the ESC unit. The digital twin model developed in this work is meant to be flexible in the sense that it is able to be adjusted for different geometries or operating conditions. As part of investigating its flexibility, we will analyze its performance with different process tuning parameters, which will be indicated in each section. Later, we will showcase the flexibility of the digital twin for testing various scenarios.

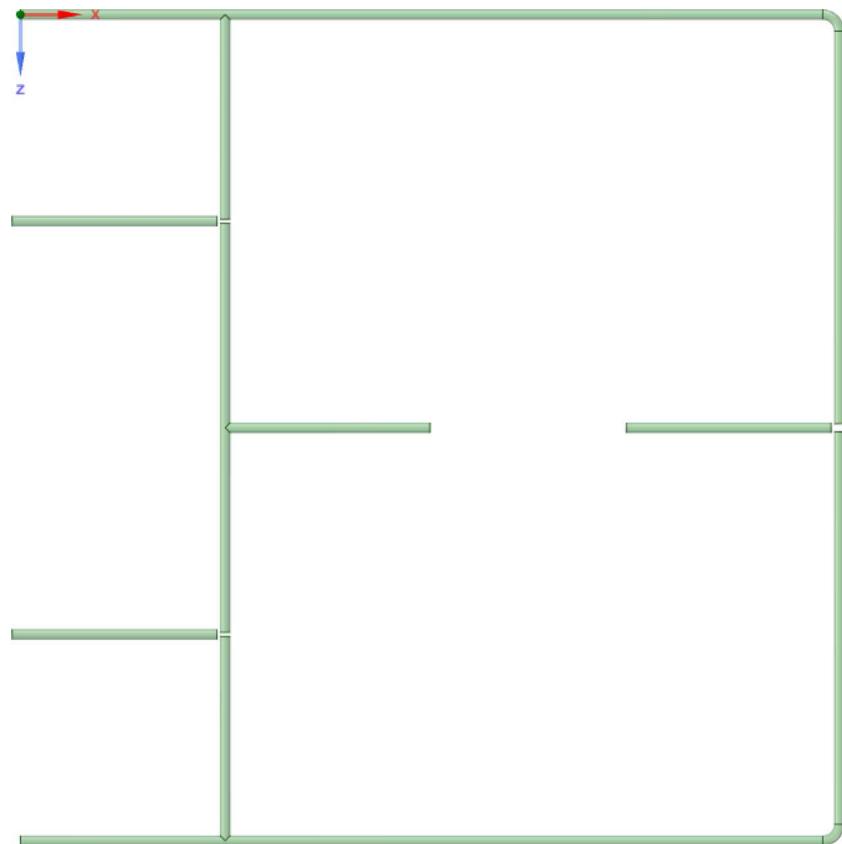


Fig. 3. A WTC geometry.

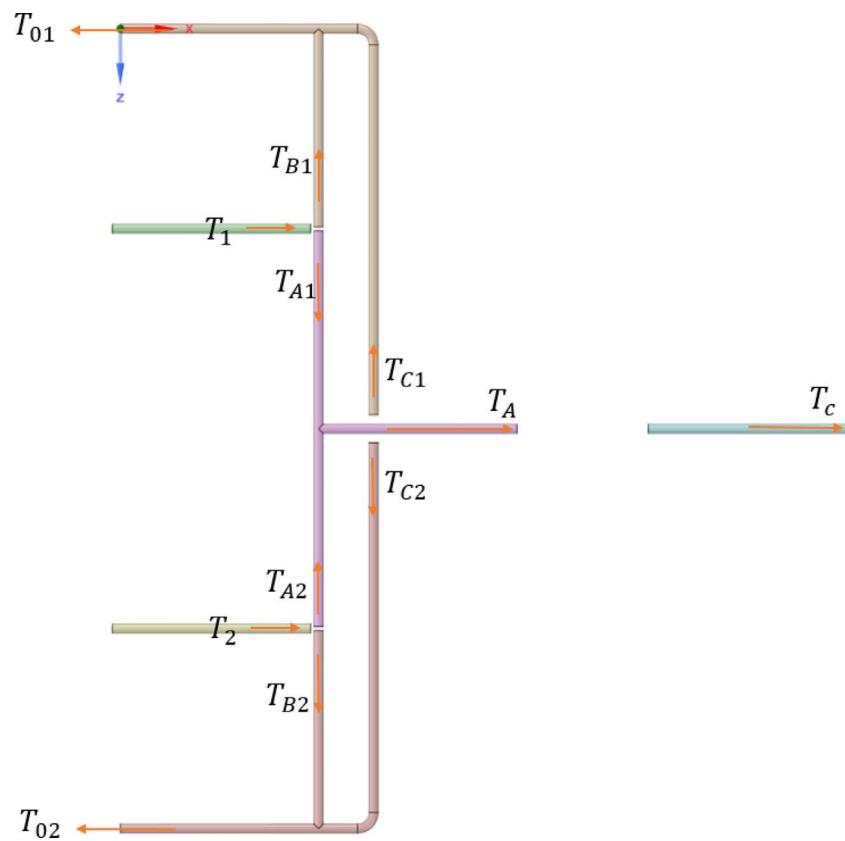


Fig. 4. Selected WTC geometry.

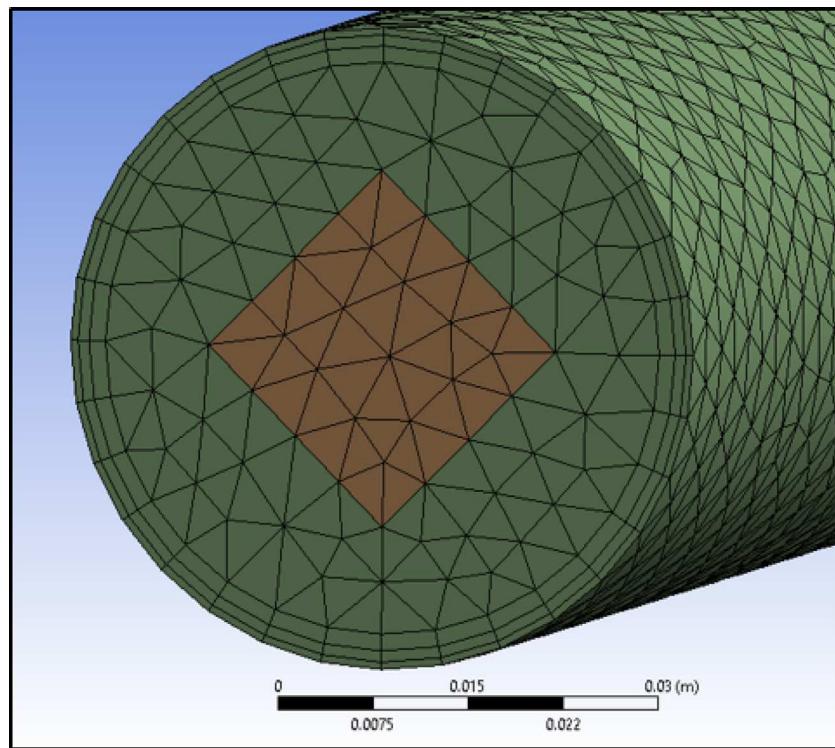


Fig. 5. WTC system with medium mesh size.

Remark 6. In contrast to works incorporating semiconductor-relevant process modeling (e.g., atomic layer deposition (Ding et al., 2019, 2021, 2020)) where the reactive portion of the process has been modeled, in this work, the internal dynamics of the ESC have not been modeled. Since the focus is on the cooling system for the wafer etch process, the impact of the wafer etch process on the cooling system has been modeled as a heat load. This effect has been included in a user-defined function in ANSYS Fluent.

4.2. ANSYS model: Digital twin mesh

In the CFD study of the WTC system, the WTC volume is discretized into a finite set of cells also known as grids or subdomains (the cells' distribution in the domain is referred to as a mesh). Then, the set of governing mathematical equations (e.g., conservation of mass, momentum, and energy) is numerically solved within the set of cells to describe the fluid flow and temperature fields. The numerical solution at each set of cells are then combined to reconstruct the solution of the entire domain. For this reason, a mesh with acceptable mesh quality is a critical aspect of the CFD modeling that determines its degree of success. In general, a CFD model built upon a poor quality mesh has not only a slower solution convergence, but also tends to converge to an inaccurate solution.

In ANSYS Fluent, there are two main classes of meshing implementation: the unstructured tetrahedral meshing (also denoted as “unstructured meshing”) and the multiblock structured hexahedral meshing (also denoted as “structured meshing”). The latter creates a set of hexahedral cells that are organized in a pattern specified by the user, while the first procedure generates a set of primarily tetrahedral cells which are distributed in an irregular pattern. We utilized the unstructured meshing strategy since it is considered generally more effective at approximating complex geometries. To enable more control of the cell distribution in the center of each pipe and prevent grid distortions with abnormal shapes, we created a subdomain along the center of each tube in the form of a “rectangle-shaped tube” as shown in Fig. 5. In

addition, boundary layers have been added to the mesh to capture the viscous forces in the regions near the wall.

In the system mesh, the grids are not uniformly distributed, but are denser in regions expected to have large momentum or temperature gradients, such as in the neighborhood of the mixing points (where the fluid velocity changes according to the inlet velocities before mixing and the heat transfer from the hot to the cold fluid is expected to create temperature gradients that must be captured through a finer mesh as shown in Fig. 6).

The quality of the resulting mesh is evaluated based on two mesh quality criteria: the orthogonal quality metric and skewness mesh metric. These two metrics are recommended by ANSYS Inc. of the commercial CFD software package utilized to develop the digital CFD model in this work. In general, low orthogonal quality or high skewness values are not recommended. In particular, if the values of the two metrics are within the ranges from “acceptable” to “excellent” shown in Table 2, the mesh can be considered to have reasonably good overall quality and thus can be used to start generating CFD results for subsequent refinement.

For the mesh represented by Fig. 5, the skewness metric values were between 0 and 0.78 and the orthogonal quality metric values were between 0.22 and 0.99. As the metric values are within the ranges from “good” to “excellent” among all subdomains, the mesh is considered to have reasonably good overall quality (this is further validated by the mesh and time independence tests performed in a subsequent section and the good agreement of the ANSYS/CFD data generated using this mesh and the expected typical industrial response). Finally, in the CFD method, the system mesh must be discretized into a sufficient number of cells such that the ANSYS/CFD simulation data becomes mesh independent. This study is provided in Section 4.4 after the physical models and solver settings are elucidated in the next section.

Remark 7. Another possibility of a meshing strategy that would allow superior manipulation of the grids at each pipe is the structured meshing implementation using an O-grid Block function. The O-grid Block

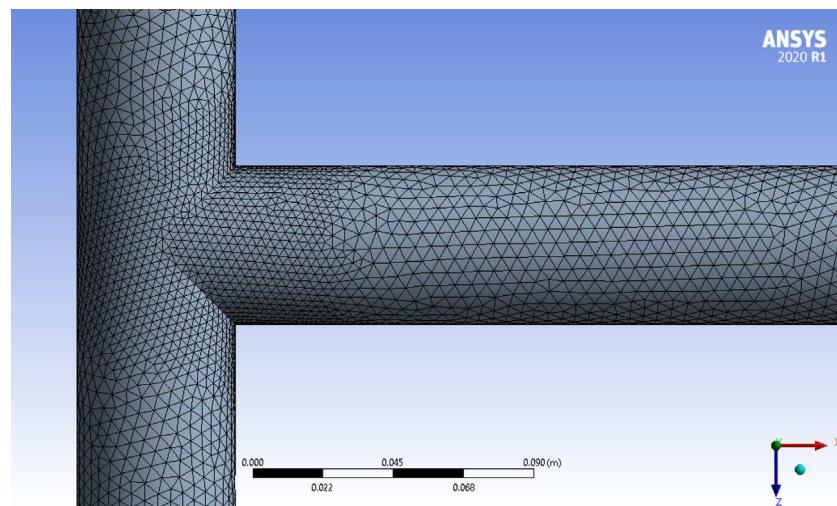


Fig. 6. A sample of the top view of a portion of the WTC system where the grid is denser compared to the rest of the mesh.

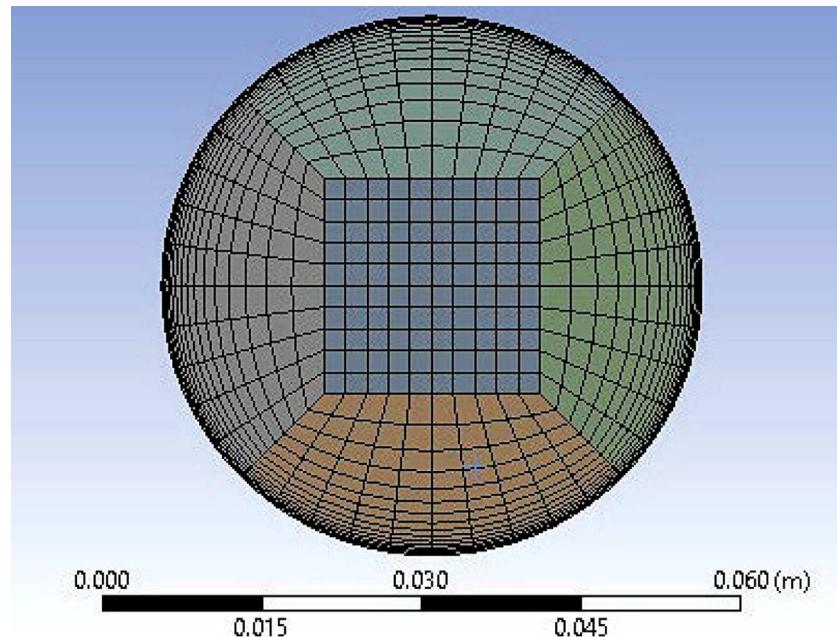


Fig. 7. Illustration of an O-grid Block function that could be used in meshing the WTC system.

Table 2
Skewness and orthogonal quality mesh metrics spectrum.

Skewness					
Excellent	Very good	Good	Acceptable	Bad	Unacceptable
0–0.25	0.25–0.50	0.50–0.80	0.80–0.94	0.95–0.97	0.98–1.00
Orthogonal quality					
Excellent	Very good	Good	Acceptable	Bad	Unacceptable
0.95–1.00	0.70–0.95	0.20–0.69	15–0.20	0.001–0.14	0–0.001

can improve the ability to approximate complex geometry features by reorganizing grid lines into an “O” shape structure to enhance overall mesh quality as shown in [Fig. 7](#). As the unstructured meshing strategy described in this section achieved a reasonably good overall mesh quality to capture the relevant system phenomena (which will be validated in a subsequent section), the O-grid Block was not pursued in this work.

4.3. ANSYS model: Set up the physics and solver settings

Steady-state and transient ANSYS simulations have been performed for mesh and time independence tests in the next section and for validation of the etch cooling system based on industrial data expected to be representative of etch cooling system responses. First, an ANSYS steady-state simulation was performed, for which the results were

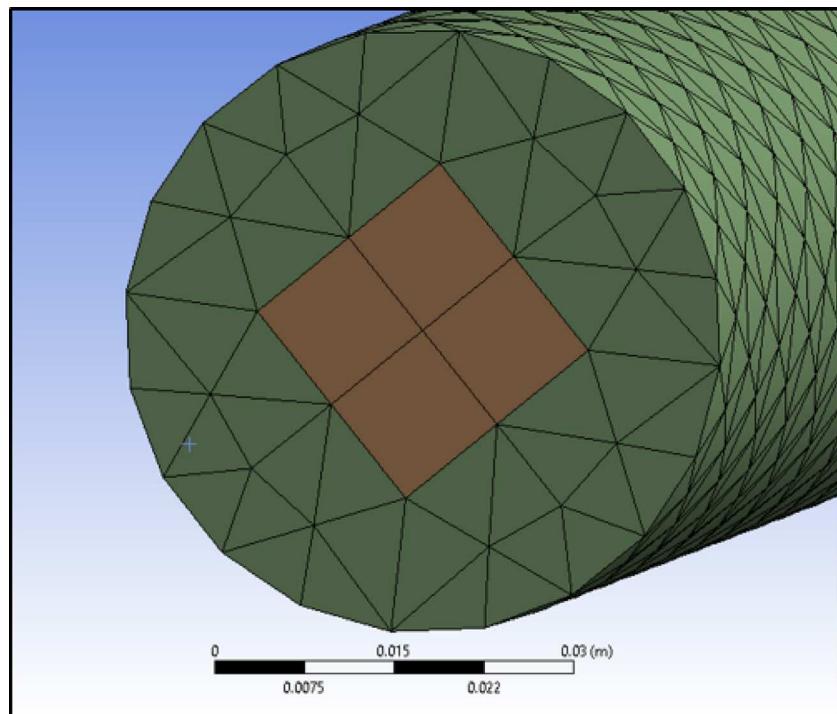


Fig. 8. WTC system with large mesh size.

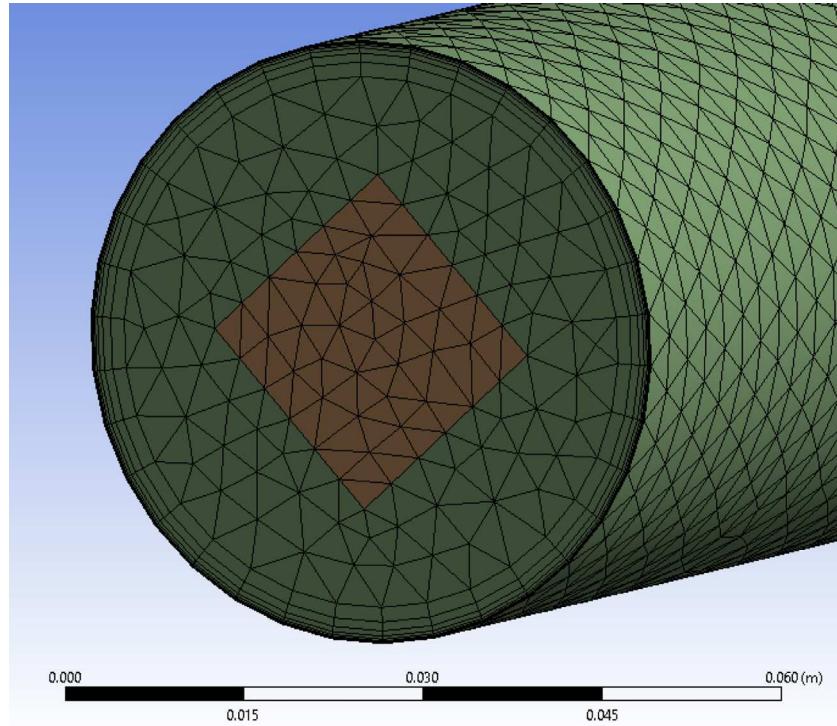


Fig. 9. WTC system with small mesh size.

then used as the initialization for the subsequent ANSYS transient simulations. In this section, we provide the details of the ANSYS Fluent setup for both the steady-state and transient studies. In particular, a journal file has been created to automate the sequence of ANSYS

Fluent commands for the ANSYS etch cooling system (similarly, the ANSYS user could enter the commands interactively through the GUI (graphical user interface) or TUI (text user interface) instead of using a journal file).

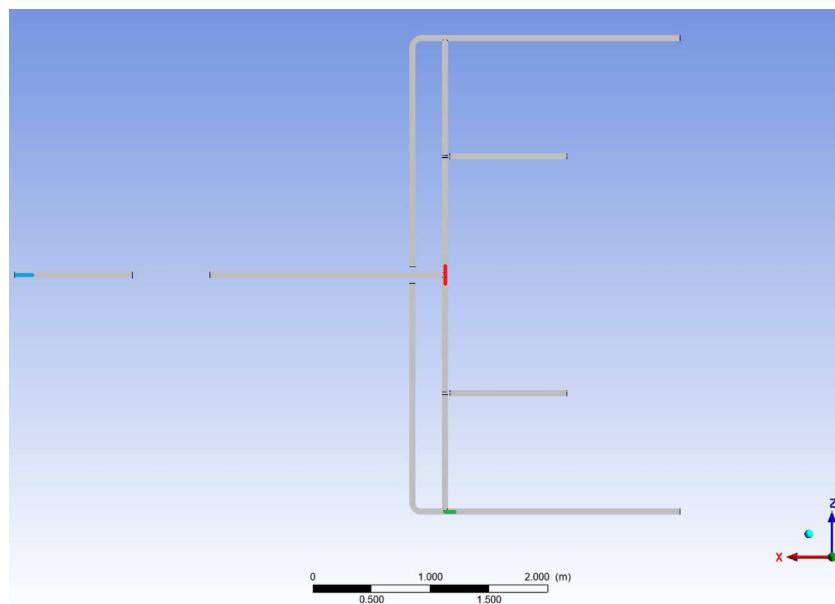


Fig. 10. Lines on which mesh independence tests were performed.

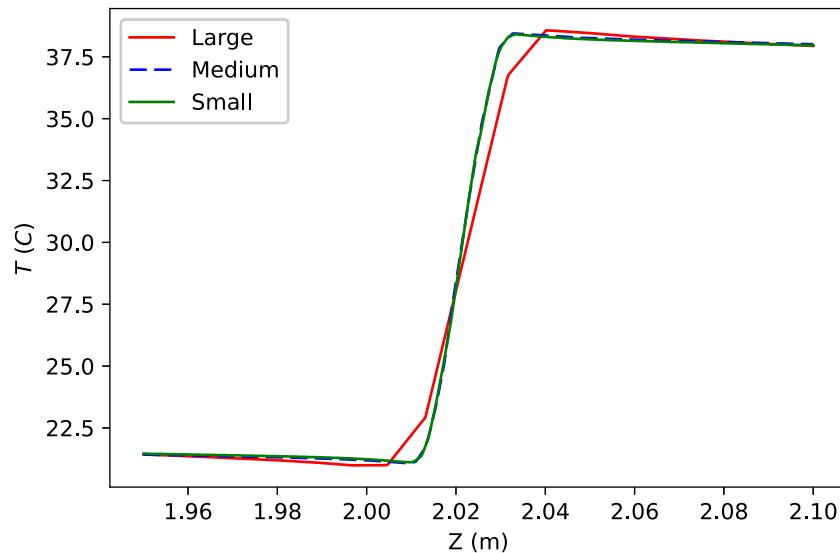


Fig. 11. Temperature profile on the red line indicated in Fig. 10 using different mesh sizes.

For the ANSYS steady-state setup via the journal file, we defined the temperature units to be in degree Celsius, the energy model was enabled, and the realizable $k - \epsilon$ turbulence model with enhanced wall treatment was used. In addition, the properties of the coolant Fluorinert-770 have been added to the ANSYS Fluent database and used as the fluid material. The user-defined function (UDF) for the steady-state simulation, written in C language, was then compiled and loaded via an ANSYS Fluent command in the journal file. All the fluid zones were defined using the coolant Fluorinert-770. The following zone types were defined: the pipe inlets were defined as mass-flow-inlet boundary conditions and the pipe outlets were defined as pressure-outlet boundary conditions. Then, UDF profiles were written in terms of “mass-flow-inlet” and “pressure-outlet” boundary conditions. Specifically, the faces corresponding to the outflow of the

three-way diverter valves (which are pipe inlets) have “mass-flow-inlet” boundary conditions that provide the mass flow rate values coded by the user. The specification of the mass flow permits the total pressure to vary in response to the interior solution. Faces corresponding to the outlet of the pipes have “pressure-outlet” boundary conditions using gauge pressure equal to zero and weak average pressure enforcement (which are the default settings in ANSYS Fluent). Finally, for the walls of the pipes, stationary and no-slip wall with roughness height zero, which corresponds to smooth walls, were utilized. Then, we defined and created report files that contain the current and past history of mass flows and area-weighted temperature values over time at each zone described above. These report files were then used as a means of monitoring surfaces and checking convergence criteria. In particular, the convergence conditions are met if all the stopping criteria for the mass-flow and area-weighted temperature values defined in the

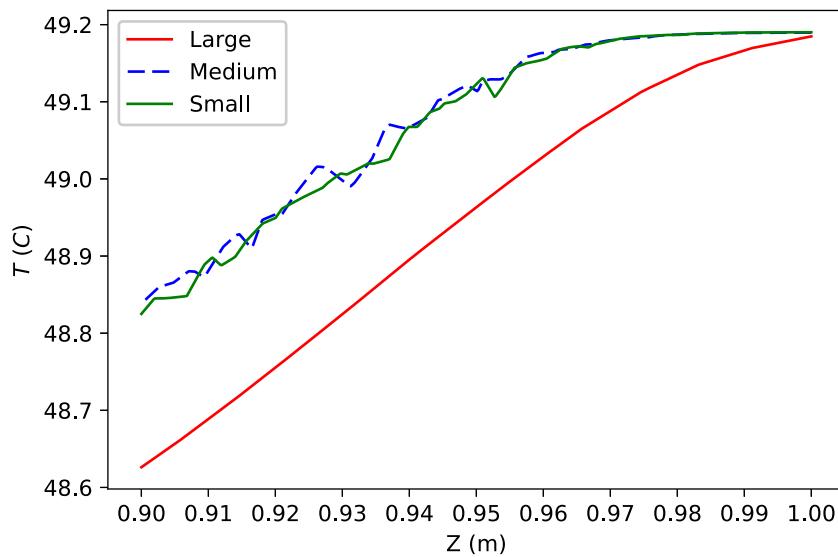


Fig. 12. Temperature profile on the green line indicated in Fig. 10 using different mesh sizes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

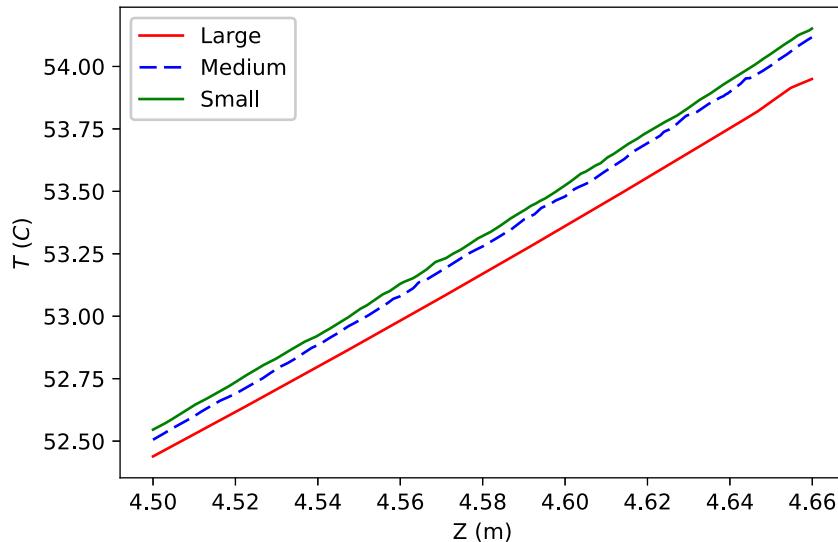


Fig. 13. Temperature profile on the blue line indicated in Fig. 10 using different mesh sizes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

report files are satisfied. Specifically, if the difference between an area-weighted temperature value at a given iteration and a prior iteration is below 10^{-6} (after ignoring the first 20 initial iterations), then the stopping criterion is satisfied. The same stopping criteria were defined for the mass-flow rates at each specified zone. The maximum number of iterations was set to be 10^6 and the converged solution was saved in a case file.

Similarly, for the ANSYS transient simulations, the realizable $k - \epsilon$ turbulence model with enhanced wall treatment was used. The coolant Fluorinert-770 was used as the fluid material and the same set up for the boundary conditions described above was implemented. In addition, the report files defined above were created for the transient simulation cases. To check for convergence criteria, the residual equations were enabled, which allow for monitoring the continuity, x-velocity, y-velocity, z-velocity, energy, turbulent kinetic energy, and

rate of dissipation of the turbulent kinetic energy residuals. The convergence criteria are met when the residuals above reach 10^{-5} , 10^{-5} , 10^{-5} , 10^{-5} , 10^{-5} , 10^{-5} , and 10^{-5} , respectively. The maximum number of iterations per time step was selected to be 300 and the time step size was chosen to be 0.01 s (mesh and time independence tests, which will be described in next section, verified this choice). The converged solution for each time step was then recorded in case and data files.

Remark 8. Several tests have been conducted to evaluate the maximum temperature variation in the radial direction at different positions in the pipe. Using the realizable $k - \epsilon$ turbulence model with enhanced wall treatment, there is less than 1% of maximum temperature variation in the radial direction for a 1 m pipe. This indicates that a 1 m pipe is sufficient to capture the fully developed flow behavior in the

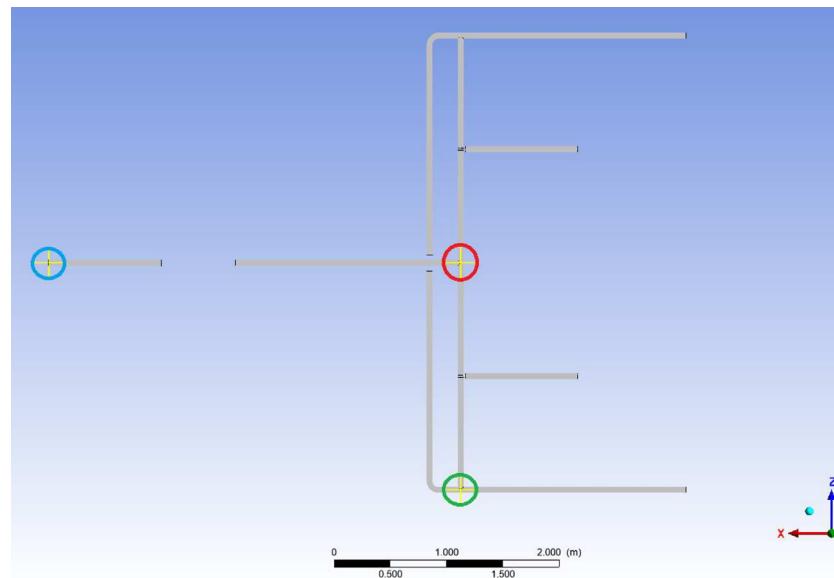


Fig. 14. Points at which time independence tests were performed.

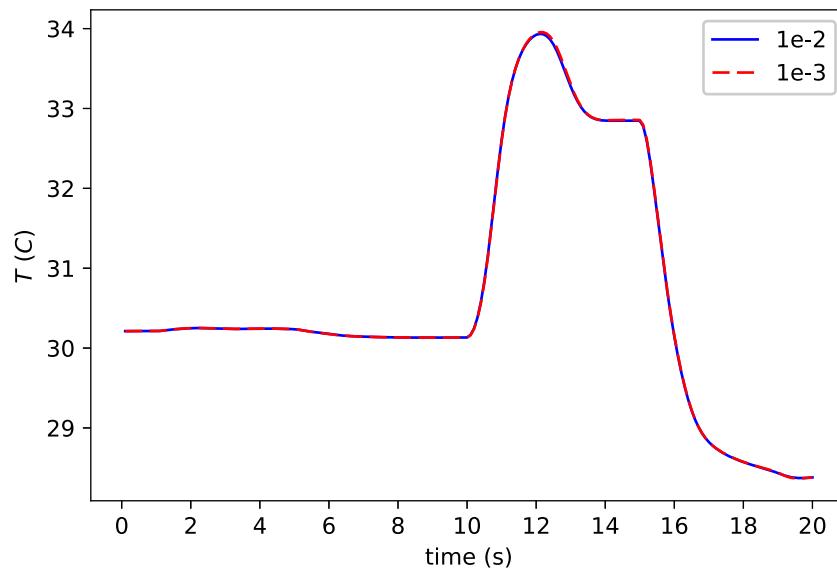


Fig. 15. Temperature over time at the red point indicated in Fig. 14 using different time steps. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

WTC system with the realizable $k - \epsilon$ turbulence model with enhanced wall treatment.

Remark 9. Turbulent flow behavior in the WTC system can be predicted by computing the Reynolds number (Re). In particular, this can be estimated for flow in pipes of the WTC system as follows: $Re = (FD_H)/(\mu A)$, where $F = 0.50204 \text{ kg/s}$ is the mass flow rate of the fluid, $D_H = 50.8 \times 10^{-3} \text{ m}$ is the diameter of the pipe, $\mu = 0.001359 \text{ kg/(m-s)}$ is the dynamic viscosity of the fluid, and A is the pipe's cross-sectional area. Thus, we can estimate that $Re \approx 9259$, which indicates that the flow is turbulent in the WTC system, and thus supports the use of a turbulence model in this simulation.

Remark 10. The selection of the turbulence model was based on two criteria: the accuracy of the results and the computation time to converge to the solution. The realizable $k - \epsilon$ turbulence model with enhanced wall treatment was selected to meet these criteria. According to ANSYS (ANSYS, 2023a), the realizable $k - \epsilon$ turbulence model is likely to provide superior performance for flows involving recirculation, which is the case for the WTC system. In addition, since the restriction that the near-wall mesh must be sufficiently fine everywhere imposes a considerably large computational requirement, we utilized enhanced wall treatment to achieve the goal of having a near-wall modeling approach for the selected mesh without a fine near-wall mesh. An alternative and popular turbulence model that could be used is the SST $k - \omega$ model (ANSYS, 2023b). However, since the results with the

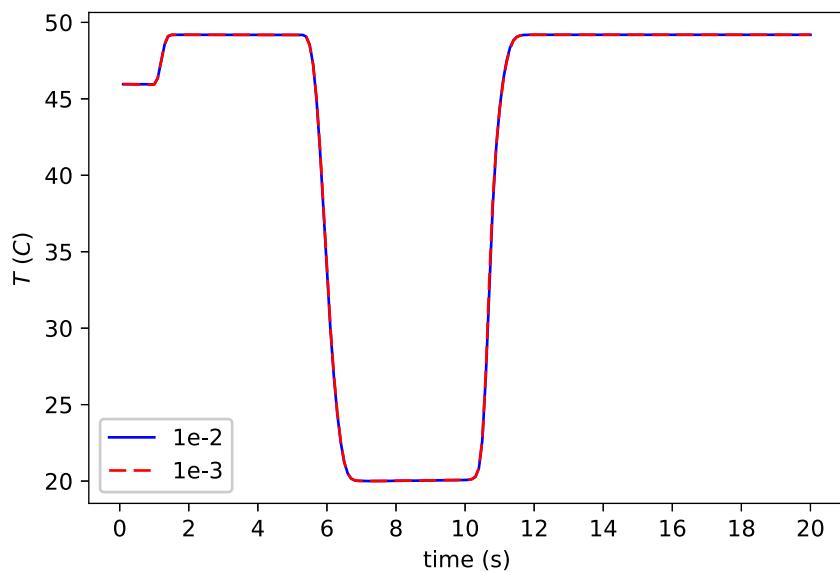


Fig. 16. Temperature over time at the green point indicated in Fig. 14 using different time steps. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

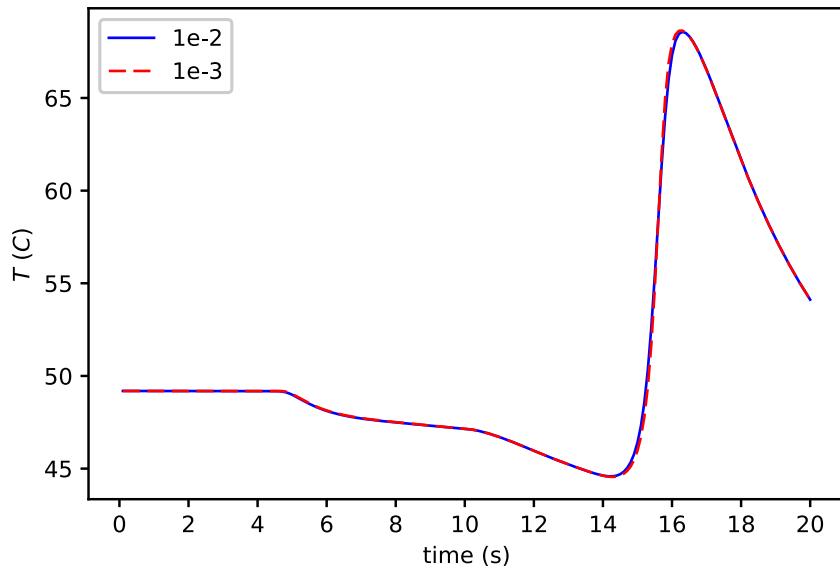


Fig. 17. Temperature over time at the blue point indicated in Fig. 14 using different time steps. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

realizable $k - \epsilon$ turbulence model provided adequate agreement with the typical industrial data, the realizable $k - \epsilon$ model was considered sufficient for the flow characteristics which it was desired to capture in this work.

4.4. ANSYS model: Mesh and time independence tests

As described in the prior sections, to check whether the simulation results are independent of the underlying mesh and will provide a physically meaningful description of the system, mesh and time independence tests have been performed. For these ANSYS simulation tests, we considered the following process parameters: $\tau_{F_i} = 5$ s, $\tau_{T_i} = 2$ s, $\tau_{T_c} = 5$ s, $\theta = 0$, and $\theta_{T_c} = 0$ s. The time delays θ and θ_{T_c} were set to zero so that a more quick response between the TCU Channels and ESC could be obtained for the mesh and time independence tests without waiting for the process response after the time delays θ and θ_{T_c} . In particular,

Table 3
Mesh independence cases.

Mesh case	Number of elements	Boundary layers
Large	374642	0
Medium	2908955	3
Small	5354001	10

three mesh cases with different sizes and number of boundary layers have been generated in ANSYS Fluent mesh, which can be visualized in Figs. 8–9. Table 3 describes the characteristics of the tested mesh cases.

Figs. 11–13 show the results of the mesh independence test for a fixed time step of 0.01 s computed at the lines indicated in Fig. 10. These results inform that a medium mesh size may be sufficient as the error between the small and medium mesh size was less than

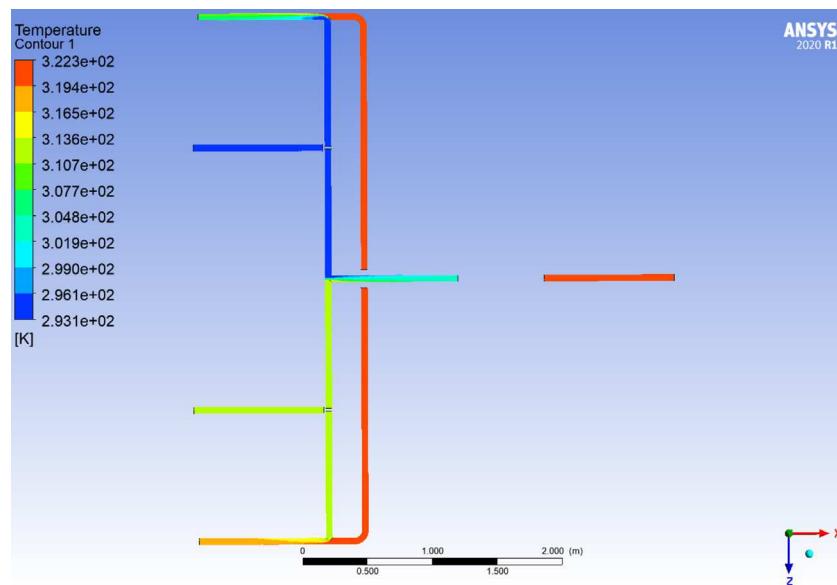


Fig. 18. Contours of temperature in the WTC system.

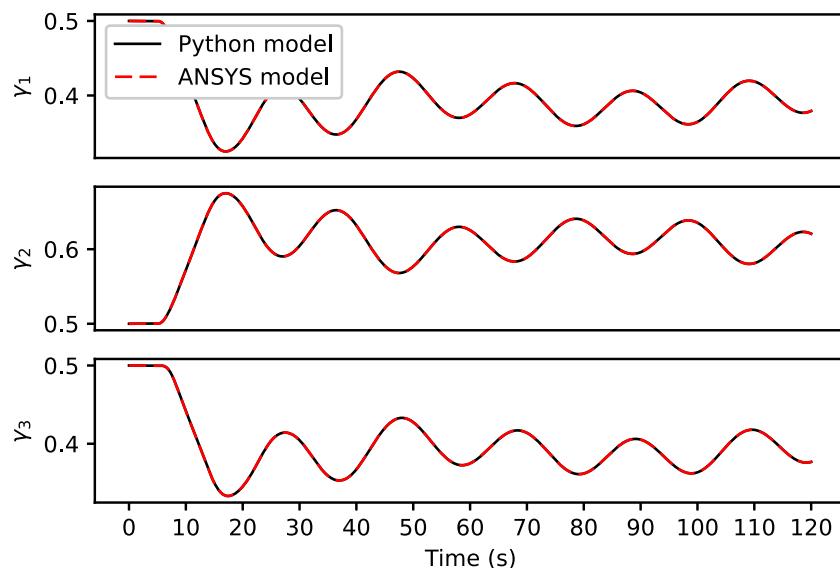


Fig. 19. Sequence of cold, hot, and level valve positions applied over time to Python and ANSYS models.

1%. Finally, to check whether the medium mesh case with a time step of 0.01 s is adequate, we performed a time independence test using the medium mesh size, which is displayed in Figs. 15–17, at the points indicated in Fig. 14. From Figs. 15–17, 0.01 s as the time step is sufficiently small for the medium mesh case because using a smaller step size does not appreciably change the result in the bulk fluid. We note also that no divergence was reported during the ANSYS Fluent simulations, and as will be described further below, the results of the time independence tests (after the selection of the mesh size) performed for both the Python and ANSYS models overlap, providing further evidence that the time step is sufficient for this process.

The results above demonstrate that a medium mesh size and a time step of 0.01 s are sufficient to capture the physics of the selected system in the bulk (the mesh independence tests performed were in

the bulk fluid). Thus, the choice of the system geometry described in Section 4.1 and “medium” mesh size allowed a much faster ANSYS Fluent simulation compared to the same system simulated with the geometry of Fig. 3. Finally, to check for fully developed flow behavior in the system, Fig. 18 displays the development of the flow for a steady-state simulation in the selected ANSYS model, in which the temperature profile presents almost no variation towards the end of each pipe.

Remark 11. For the mesh independence tests, the steady-state simulation result of the WTC system has been used. On the other hand, for the time independence tests, the dataset from the transient simulations was utilized. Specifically, the following step-changes were applied for the transient simulation: 1) between 1 s and 5 s of operating time, the valve position set-points of all three-way diverter valves changed from 50% to

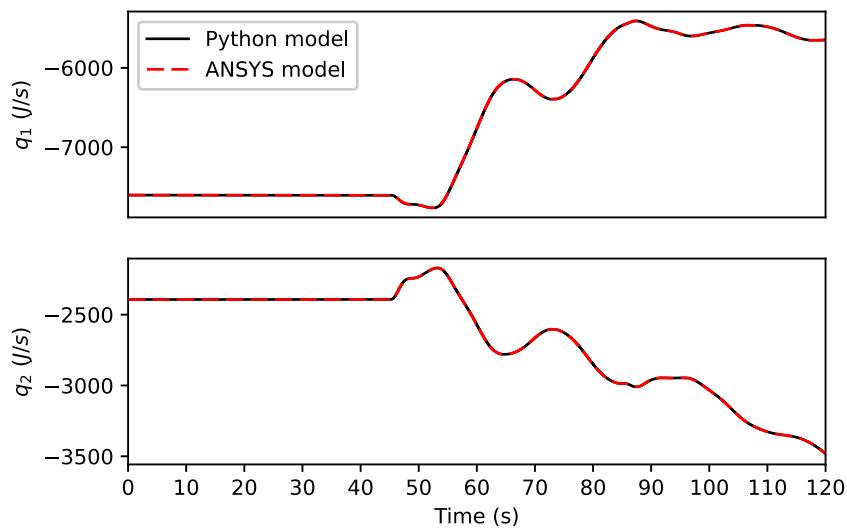


Fig. 20. Sequence of heat rates in the TCU Ch. 1 and 2 applied over time in Python and ANSYS models.

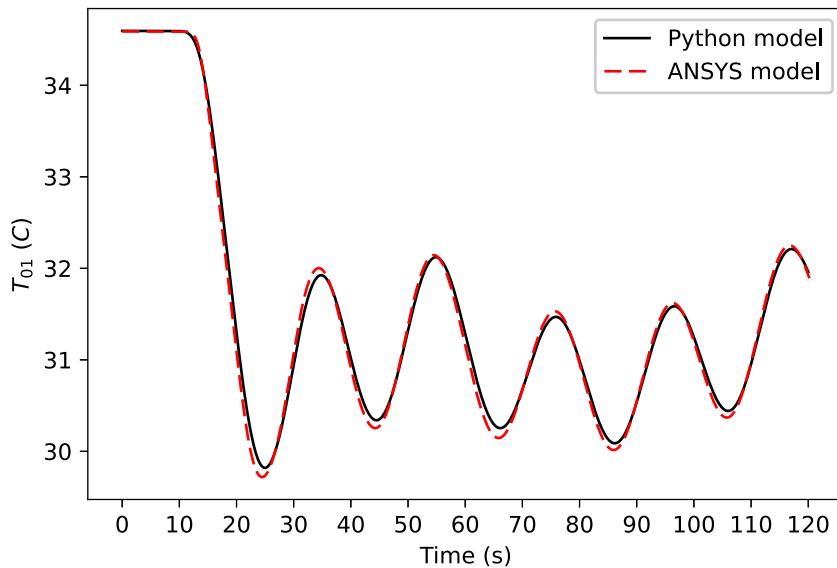


Fig. 21. Inlet temperature of the TCU Ch. 1 over time in Python and ANSYS simulations under the same inputs.

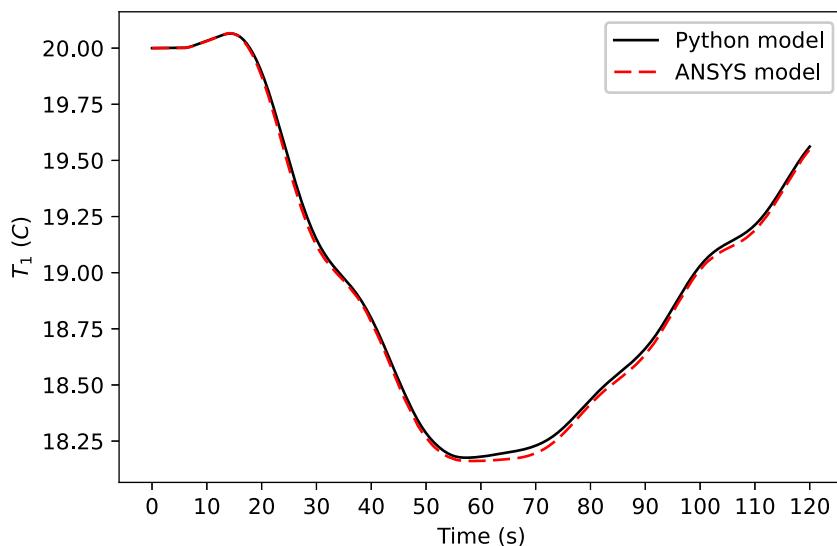


Fig. 22. Outlet temperature of the TCU Ch. 1 over time in Python and ANSYS simulations under the same inputs.

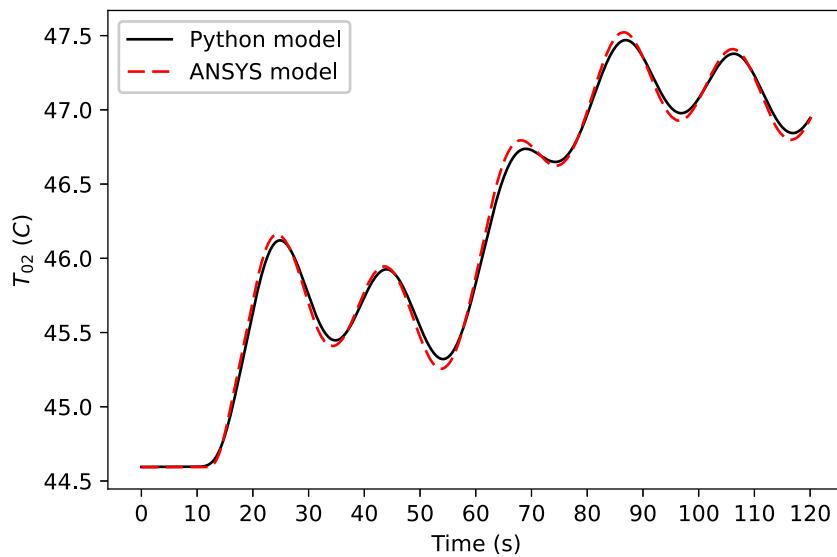


Fig. 23. Inlet temperature of the TCU Ch. 2 over time in Python and ANSYS simulations under the same inputs.

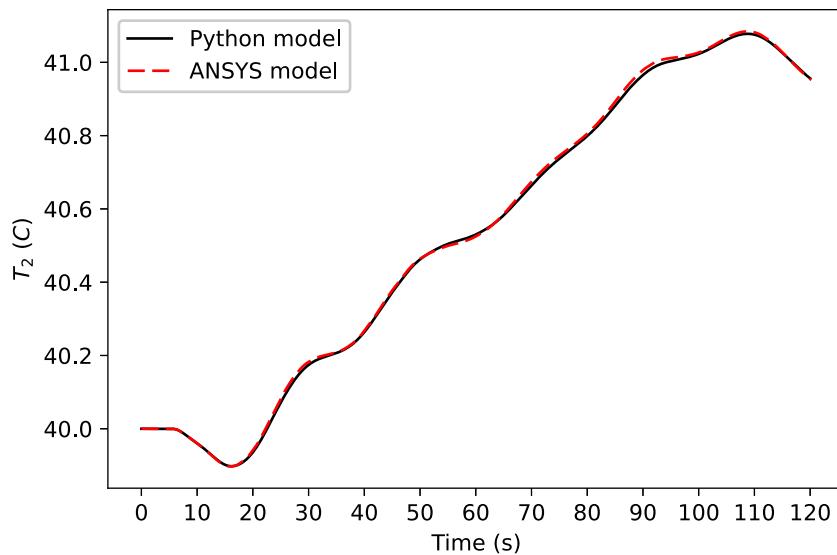


Fig. 24. Outlet temperature of the TCU Ch. 2 over time in Python and ANSYS simulations under the same inputs.

100%, and the steady-state heat rates, q_{1s} and q_{2s} , were applied to TCU Ch. 1 and 2, respectively; 2) between 5 s and 10 s of operating time, the valve position set-points of all three-way diverter valves changed from 100% to 10%, while keeping the steady-state heat rates, q_{1s} and q_{2s} , applied to TCU Ch. 1 and 2, respectively; 3) between 10 s and 15 s of operating time, the hot valve position set-point changed from 10% to 60%, the cold valve position set-point changed from 10% to 70%, the level valve position set-point changed from 10% to 80%, and the upper bounds of the heat rates, $q_1 = 10 \text{ kW}$ and $q_2 = 10 \text{ kW}$, were applied to TCU Ch. 1 and 2, respectively; 4) between 15 s and 20 s of operating time, the hot valve position set-point changed from 60% to 100%, the cold valve position set-point changed from 70% to 90%, the level valve position set-point changed from 80% to 60%, and the lower bounds of the heat rates, $q_1 = -10 \text{ kW}$ and $q_2 = -10 \text{ kW}$, were applied to TCU Ch. 1 and 2, respectively.

5. Wafer temperature control for plasma etching: Python model

The ANSYS/CFD model proposed above takes several days to execute around 100 s of physical operation time on the computing

resources that were utilized (e.g., a high-performance computing cluster at Wayne State University with one node and 64 cores takes about a week to run 100 s of physical operation). However, the ANSYS simulations have enabled us to generate data that could be used to develop a reduced-order model for the etch cooling system with a computation time that is orders of magnitude faster than the ANSYS system. Thus, based on the ANSYS results, a reduced-order model was built that is able to capture key features of the digital system. Specifically, this model is a set of partial differential equations of the form of Eq. (6) that consider flow in the axial direction of the pipe, along with updates to the velocities in the pipes based on valve positions. In particular, Eqs. (1)–(13) were explicitly coded in Python. The ordinary differential equations (ODEs) were solved using Euler's method with an integration step size of 0.01 s (which is validated in the next section), whereas for the partial differential equations (PDEs), the method of lines was utilized. In particular, the backward finite difference was used to replace the spatial derivative of the temperature with an algebraic approximation, as follows:

$$\frac{\partial T_i}{\partial z} \approx \frac{T_i - T_{i-1}}{\Delta z} \quad (22)$$

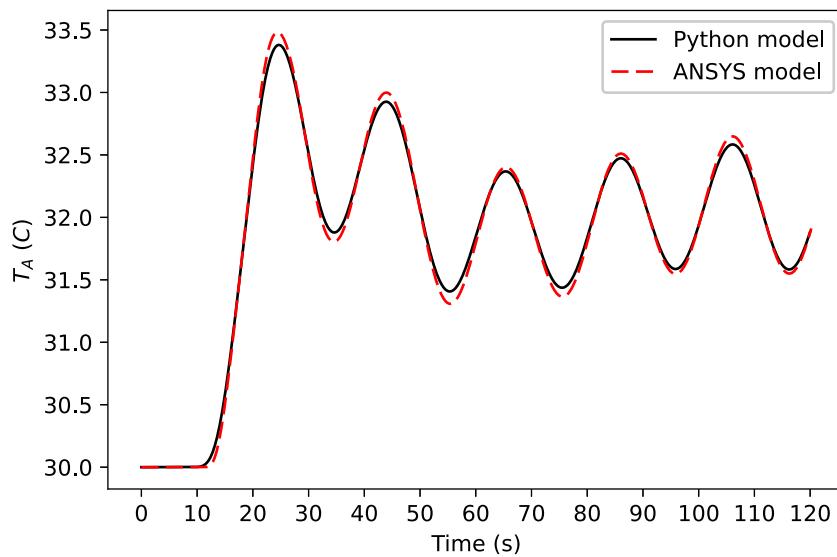


Fig. 25. Inlet temperature of the ESC over time in Python and ANSYS simulations under the same inputs.

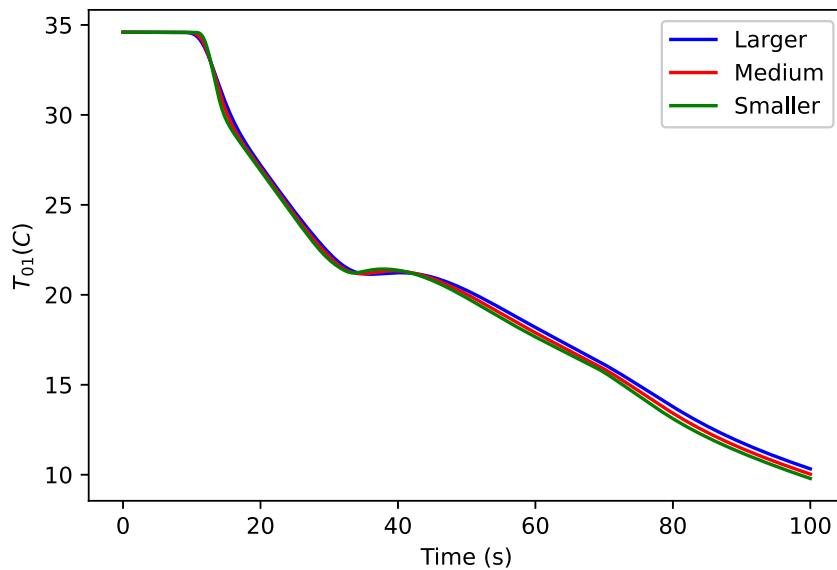


Fig. 26. Inlet temperature profile of the TCU Ch. 1 under different numbers of nodes.

where the index i designates a position along a grid in z and Δz is the spacing in z along the grid, such that Eq. (6) can be approximated as a set of ordinary differential equations. Then, Euler's method was applied to numerically integrate the resulting ODE models.

The reduced-order model was run in Python and used to evaluate various operating options for the etch cooling system. The faster model in Python takes the order of seconds (e.g., around 15 s) to execute around 100 s of physical operation time on a desktop computer. As this model is based on the physics observed in the ANSYS/CFD simulations and incorporates all the assumptions described in Section 3, we checked whether these assumptions are suitable for capturing the major behavior of interest of the etch cooling system (such as an average temperature at the inlet of the ESC and the average flow rate at the inlet of the ESC). In particular, we compared data from the model implemented in Python with that from ANSYS for the same inputs as shown in Figs. 19–20 using the following process parameters: $\tau_{F_i} = 5$ s, $\tau_{T_i} = 12$ s, $\tau_{T_c} = 12$ s, $\theta = 40$ s, and $\theta_{T_c} = 0.5$ s. The results are depicted in Figs. 21–25, which demonstrate good agreement. We therefore consider that the fast Python simulation captures the major features of the digital twin and may be used in benchmarking how the digital twin

would behave under different control strategies. This enables us to test a variety of potential control strategies for the etch cooling system much more rapidly before the implementation in ANSYS Fluent while having a high level of confidence that they are providing results and insights that may be used for retuning the controllers or changing operating strategies. Then, a final testing of the control policies in the ANSYS simulation environment may be performed to show important details of flow and temperature patterns that would result under the control strategies determined to be suitable when selected with the faster Python model.

Remark 12. The selection of the numerical method and programming language used to simulate the reduced-order model of the WTC system may play a role in the speed of the simulations. Though the method of lines was used to solve the set of partial differential equations and Euler's method was applied to solve the resulting set of ordinary differential equations (ODEs), alternative ODE solving methods (and languages besides Python such as compiled languages) could be compared in terms of performance and computing time for the reduced-order model of the WTC system.

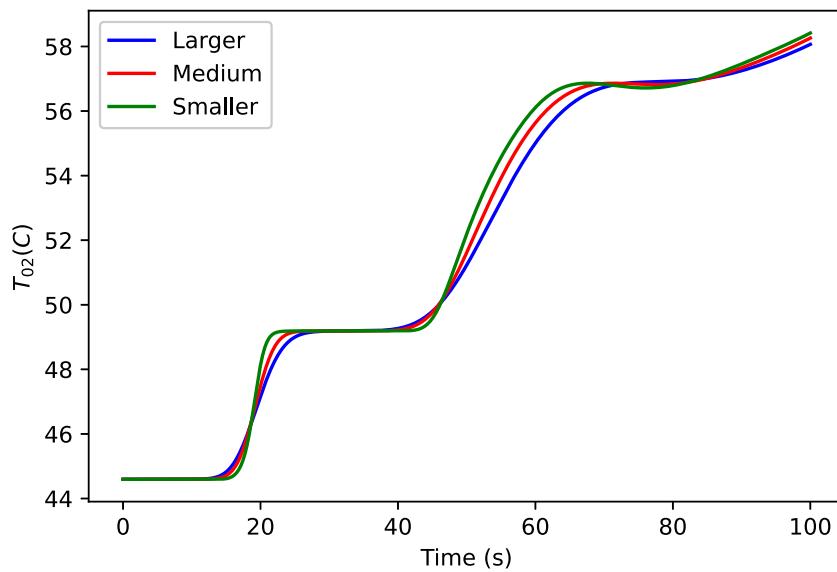


Fig. 27. Inlet temperature of the TCU Ch. 2 over time under different numbers of nodes.

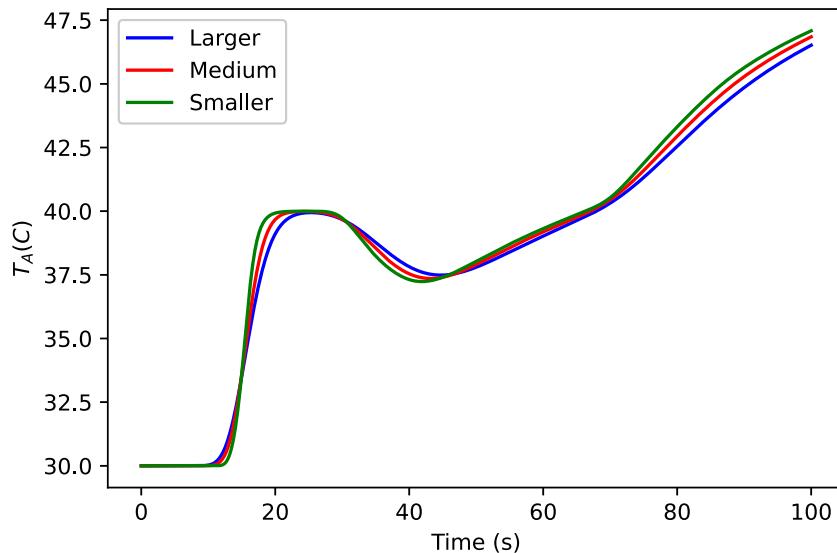


Fig. 28. Inlet temperature of the ESC over time under different numbers of nodes.

5.1. Python model: Selection of number of nodes and integration step size

The WTC model defined in Section 3 has been implemented in Python 3.6 and benchmarked with typical WTC systems in terms of open-loop and closed-loop responses. Prior to this evaluation, an appropriate number of nodes and integration step for the dynamic system model described in Section 3 needed to be selected. First, several tests have been performed using different numbers of nodes, according to Table 4, and fixed integration step size of 0.01 s, using the process parameters $\tau_{F_i} = 5$ s, $\tau_{T_i} = 12$ s, $\tau_{T_c} = 12$ s, $\theta = 40$ s, and $\theta_{T_c} = 0.5$ s. Figs. 26–28 show the results of temperature profiles at different points in the WTC model under different numbers of nodes. We can see that the “Larger” grid case is not fine enough to capture well all the dynamics of the system. The “Medium” grid case has been selected as the baseline case since the dynamic response of the WTC system is considered to be sufficiently close to the dynamic response of the WTC system under the finer grid case (“Smaller” grid case).

Table 4
Tests to check number of nodes.

Tests	Larger	Medium	Smaller
1 m pipe	10 nodes	20 nodes	100 nodes
≈2 m pipe (2.144 m)	20 nodes	40 nodes	200 nodes

Finally, to check whether the number of nodes and integration step size of 0.01 s are adequate, time independence tests have been performed using the closed-loop system under the control parameters from Table 5 with a set-point change such that T_{Asp} changes from 30°C to 32°C at 100 s, and then T_{1sp} and T_{2sp} change from 20°C and 40°C to 22°C and 42°C, respectively, at 300 s, while F_{Asp} changes from 16.8 L/min to 19 L/min at 300 s, using different integration steps for the “Medium” grid case. Figs. 29–33 depict the temperature profiles at different points in the WTC model using different integration step

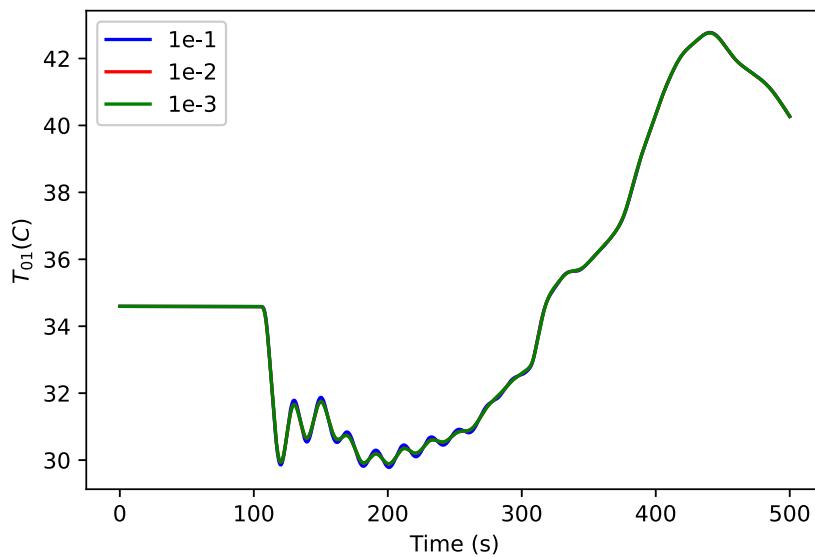


Fig. 29. Inlet temperature of the TCU Ch. 1 over time under different integration steps.

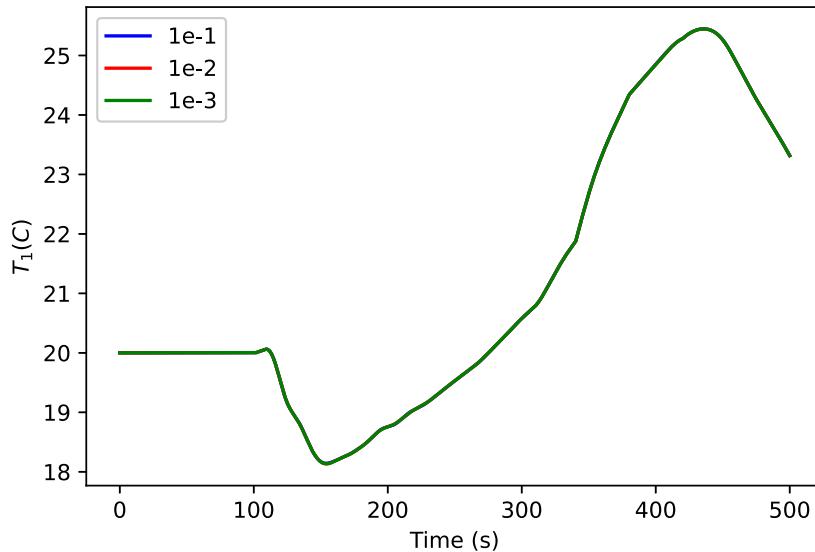


Fig. 30. Outlet temperature of the TCU Ch. 1 over time under different integration steps.

sizes. These results indicate that 0.01 s as the integration step for the “Medium” grid case is sufficient to represent the underlying dynamics of the WTC model.

Remark 13. We check the node number with a pre-specified input policy. In closed-loop, numerical errors can cause different state trajectories, creating different inputs because the inputs rely on state measurements. The goal of the Python model is to form a suitably accurate model for use in, for example, determining controller parameters for the ANSYS simulation. We considered that reasonable numerical accuracy of the Python model in the open-loop simulations would be sufficient for aiding with the tuning task.

5.2. Python model: Benchmarking Python WTC system model with ANSYS WTC system model

Although the Python and ANSYS simulations provided good agreement under the same input trajectories over time as depicted in Section 5, the Python closed-loop system under the PID controllers still

needs to be benchmarked with closed-loop ANSYS simulation data. In particular, we are interested to have a reasonable control performance in which the inlet temperature to the ESC must reach the specified set-point within 20 s of operation. To attempt to achieve this performance goal, the control parameters were first tuned according to Table 5, using the process parameters $\tau_{F_i} = 5$ s, $\tau_{T_i} = 12$ s, $\tau_{T_c} = 12$ s, $\theta = 40$ s, and $\theta_{T_c} = 0.5$ s. A set-point change from $T_{Asp} = 30^\circ\text{C}$ to $T_{Asp} = 32^\circ\text{C}$ was applied after 5 s of operation and the closed-loop simulation results in Python and ANSYS are shown in Figs. 34–40. We can observe from Fig. 40 that the inlet temperature to the ESC, simulated using the Python model, first reached the set-point in less than 20 s of operation after the set-point change with an overshoot of about 4%. However, the closed-loop ANSYS simulation results revealed considerable discrepancies between both closed-loop ANSYS and Python simulations under the same PID controllers, in particular with larger oscillatory behavior under the ANSYS model. This suggests that the control tunings in Table 5 are closer to causing instability in the presence of small differences in time constants and/or time delays (or numerical inaccuracies) between the Python and ANSYS models. To

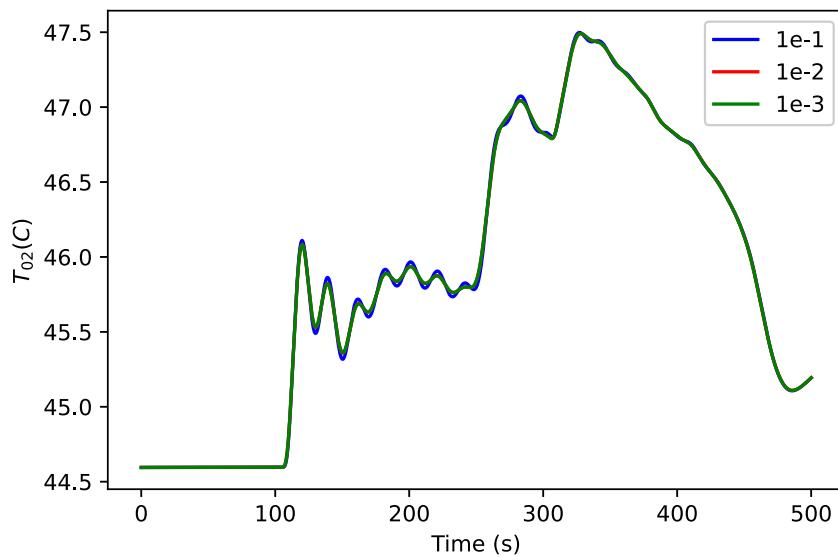


Fig. 31. Inlet temperature of the TCU Ch. 2 over time under different integration steps.

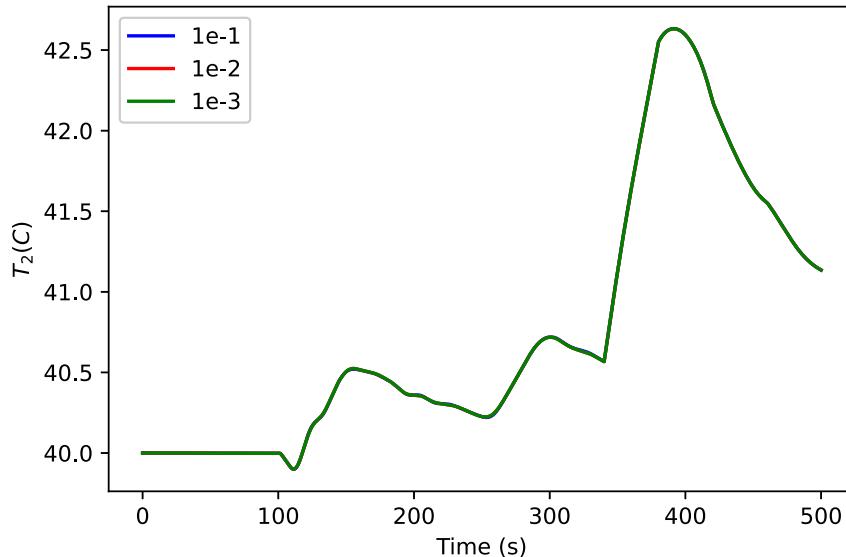


Fig. 32. Outlet temperature of the TCU Ch. 2 over time under different integration steps.

investigate this, two closed-loop Python simulations were performed with a set-point change from $T_{Asp} = 30^\circ\text{C}$ to $T_{Asp} = 32^\circ\text{C}$ applied after 5 s of operation: the first was conducted based on the Python model described in Section 5 (where $\tau_{F_i} = 5$ s, $\tau_{T_i} = 12$ s, $\tau_{T_c} = 12$ s, $\theta = 40$ s, and $\theta_{T_c} = 0.5$ s) with the control tunings in Table 5, whereas the second closed-loop Python simulation was performed with the Python model in Section 5 but 1 s subtracted from the temperature sensor time constant (τ_{T_i}) and time constant of the ESC/WEP (τ_{T_c}) with the same control tunings in Table 5. The results are shown in Figs. 41–47. It can be seen that the relatively small differences in time constants cause the closed-loop response to have larger oscillation amplitudes around the set-point value of 32°C , which is similar to the closed-loop response displayed in Figs. 34–40. This indicates that when using the Python model to suggest controller tunings, the controllers should not be overly aggressive so that the controllers are robust to these small variations in the process dynamics.

As noted earlier in this work, a benefit of the digital twin is its ability to flexibly test closed-loop responses in various scenarios, including different time constant and delay parameters. At this point, we will begin

Table 5
Control parameters attempt for the WTC system.

Control parameter	Value
λ_i	25
K_{c3}	0.2
τ_{I3}	9.8
τ_{D3}	5.6
α_3	0.1
K_{c4}	35
τ_{I4}	0.1
τ_{D4}	0.001
α_4	0.1
K_{c5}	35
τ_{I5}	100
τ_{D5}	60
α_5	0.1

to evaluate the performance of the digital twin with a new set of process parameters ($\tau_{F_i} = 5$ s, $\tau_{T_i} = 9$ s, $\tau_{T_c} = 9$ s, and $\theta_{T_c} = 1.7$ s), keeping in

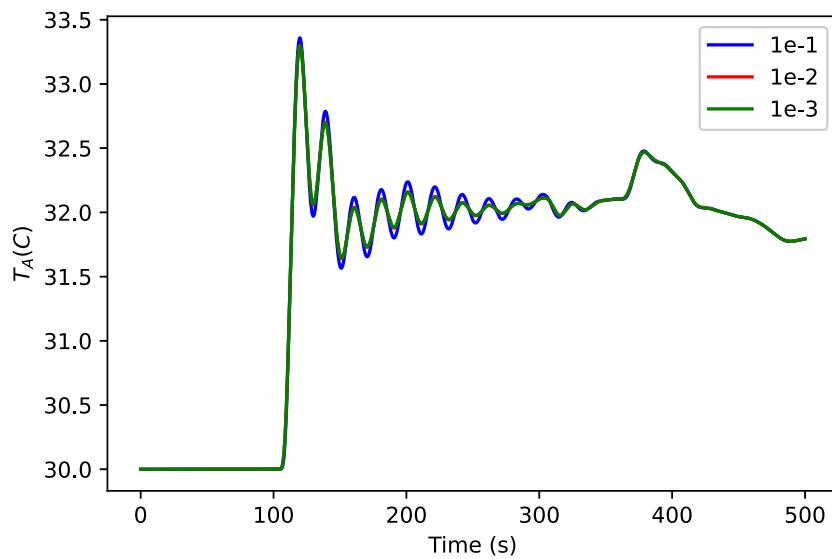


Fig. 33. Inlet temperature of the ESC over time under different integration steps.

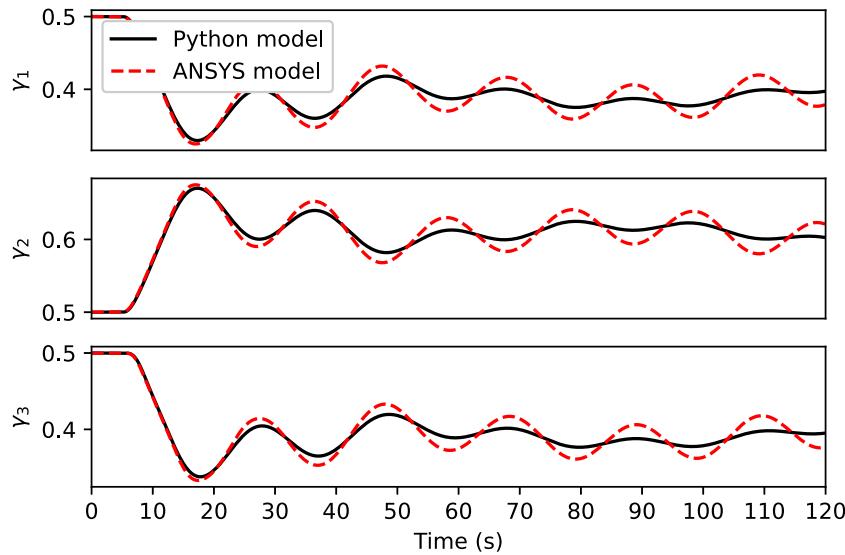


Fig. 34. Cold, hot, and level valve position set-points computed using the Python model and ANSYS model under the same PID controllers with the control tunings indicated in Table 5.

mind the insights gained with the other set above and demonstrating the flexibility of the digital twin strategy. To account for the insights described above using the perturbed parameters in the Python model, we expect that tuning a controller for the updated process parameters will require that the controller not be overly aggressive to prevent significant disagreement between the Python and ANSYS simulations. We therefore propose the PID control settings in Table 6, in which the integral (τ_{I3} and τ_{I4}) and derivative (τ_{D3} and τ_{D4}) terms are based on heuristics in Bequette (1999) but adjusted for fixed values of K_{c3} and K_{c4} to provide a reasonable control performance and robustness. To show that these updated control tunings would provide an adequate performance and robustness, we conducted a similar test as described

for Figs. 41–47. Specifically, two closed-loop Python simulations were performed with a set-point change from $T_{Asp} = 30^\circ\text{C}$ to $T_{Asp} = 32^\circ\text{C}$ implemented after 0.01 s of operation: the first was conducted based on the Python model described in Section 5 (where $\tau_{F_i} = 5$ s, $\tau_{T_i} = 9$ s, $\tau_{T_c} = 9$ s, $\theta = 40$ s, and $\theta_{T_c} = 1.7$ s for these tuning benchmarking simulations) with the control tunings in Table 6, whereas the second closed-loop Python simulation was performed with the Python model in Section 5 but 1 s subtracted from the temperature sensor time constant (τ_{T_i}) and time constant of the ESC/WEP (τ_{T_c}) with the same control tunings in Table 6. The simulation results are depicted in Figs. 48–54. The good closed-loop response agreement in Figs. 48–54 between the two closed-loop Python simulations is an indicator that the

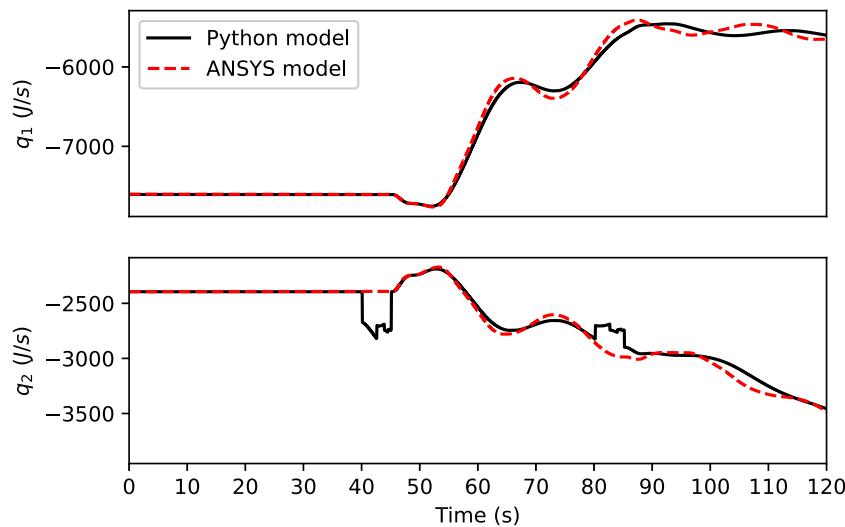


Fig. 35. Heat rate inputs computed using the Python model and a perturbed version of the Python model under the same PID controllers with the control tunings indicated in Table 5.

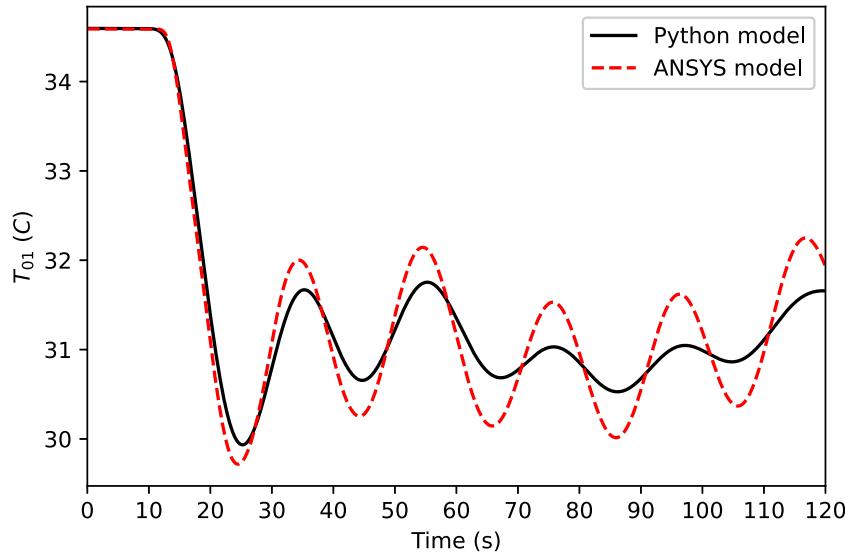


Fig. 36. Inlet temperature of the TCU Ch. 1 over time using the Python model and a perturbed version of the Python model under the same PID controllers with the control tunings indicated in Table 5.

proposed PID control settings may be applied in the closed-loop ANSYS simulation as the controllers not only provided good performance (the inlet temperature of the ESC reaches the set-point in about 17 s with about 3% of overshoot), but also demonstrated good robustness to the variation in the process dynamics.

Finally, to benchmark the closed-loop Python simulation with the corresponding closed-loop ANSYS simulation, the same test described above using the process parameters $\tau_{F_i} = 5$ s, $\tau_{T_i} = 9$ s, $\tau_{T_c} = 9$ s, $\theta = 40$ s, and $\theta_{T_c} = 1.7$ s was performed using the ANSYS model in place of the perturbed Python model under the control tunings indicated in Table 6. These simulation results are shown in Figs. 55–61. Both the Python and ANSYS closed-loop responses are close and thus provide the level of confidence necessary to explore other simulation studies using either the Python or ANSYS model.

Remark 14. Though we did not make a rigorous analysis of each PID controller in terms of performance and robustness under different control tunings, an acceptable closed-loop response is obtained with the control parameters indicated in Table 6 that can be used for both the Python and ANSYS closed-loop simulations. In addition, since both controller tunings as well as process parameters are adjusted throughout this manuscript, the results of how well the node number and integration step checks with more and less nodes match one another can change. However, we deemed that despite different discrepancies between results with more or less nodes and different integration steps as the controller and process parameters changed, the chosen step size and node number provided an adequate trade-off between prediction and computation time. We considered that the relatively good agreement between the Python and ANSYS closed-loop simulation

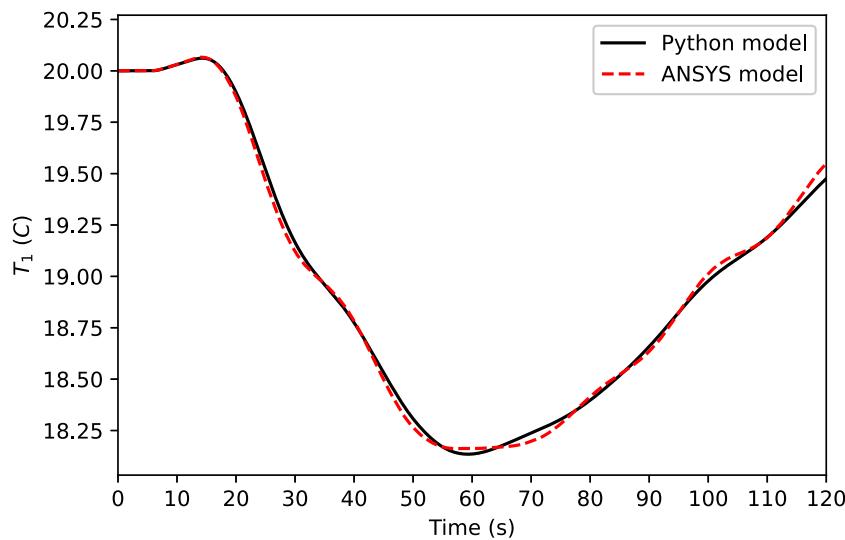


Fig. 37. Outlet temperature of the TCU Ch. 1 over time using the Python model and a perturbed version of the Python model under the same PID controllers with the control tunings indicated in Table 5.

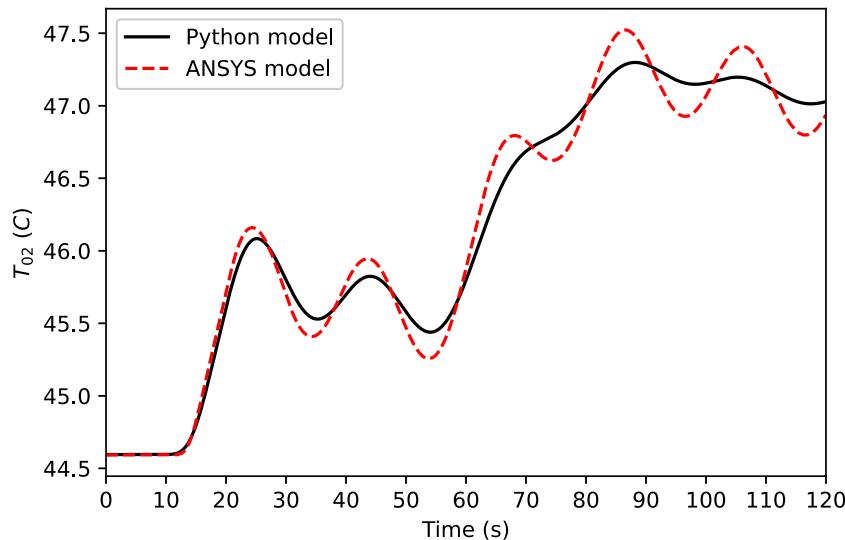


Fig. 38. Inlet temperature of the TCU Ch. 2 over time using the Python model and a perturbed version of the Python model under the same PID controllers with the control tunings indicated in Table 5.

Table 6
Control parameters for the WTC system.

Control parameter	Value
λ_i	25
K_{c3}	0.1
τ_{I3}	7.5
τ_{D3}	2.6
α_3	0.1
K_{c4}	35.0
τ_{I4}	0.1
τ_{D4}	0
α_4	0.1
K_{c5}	20
τ_{I5}	17
τ_{D5}	3
α_5	0.1

data validates that despite any numerical inaccuracies in the Python model, it maintains relatively good predictions.

Remark 15. The values of τ_{T_i} , τ_{T_c} , and θ_{T_c} were updated after the benchmarking performed in Figs. 41–47 to match the closed-loop system response of a typical industrial etch cooling system more closely, which will be discussed in the subsequent sections.

5.3. Python model: Benchmarking open-loop dynamic model with typical industrial etch cooling system

The prior sections benchmarked the open and closed-loop Python simulations with open and closed-loop ANSYS simulations. It is important to note that the reduced-order model of the etch cooling system proposed in this work based on ANSYS simulation data of the digital twin can be easily adjusted by changing the time constant τ_y of the valve dynamics, time constants τ_{T_i} of the temperature sensor dynamics,

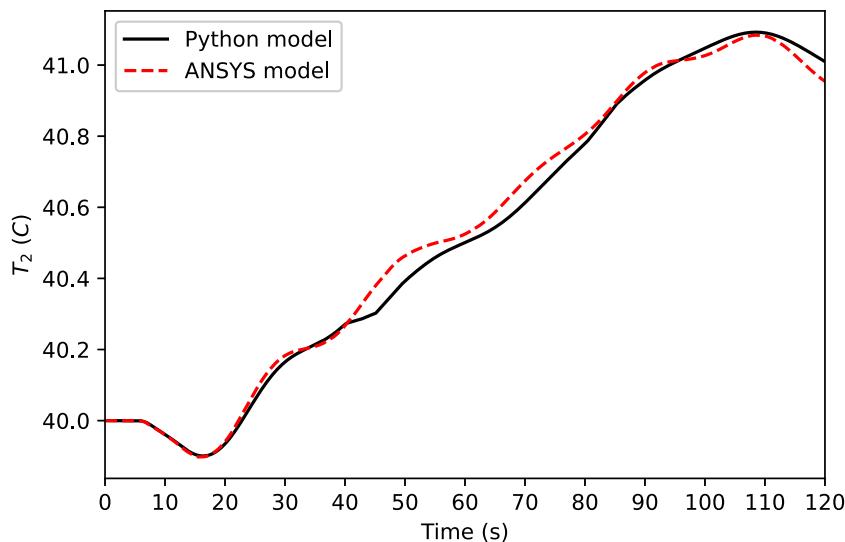


Fig. 39. Outlet temperature of the TCU Ch. 2 over time using the Python model and a perturbed version of the Python model under the same PID controllers with the control tunings indicated in Table 5.

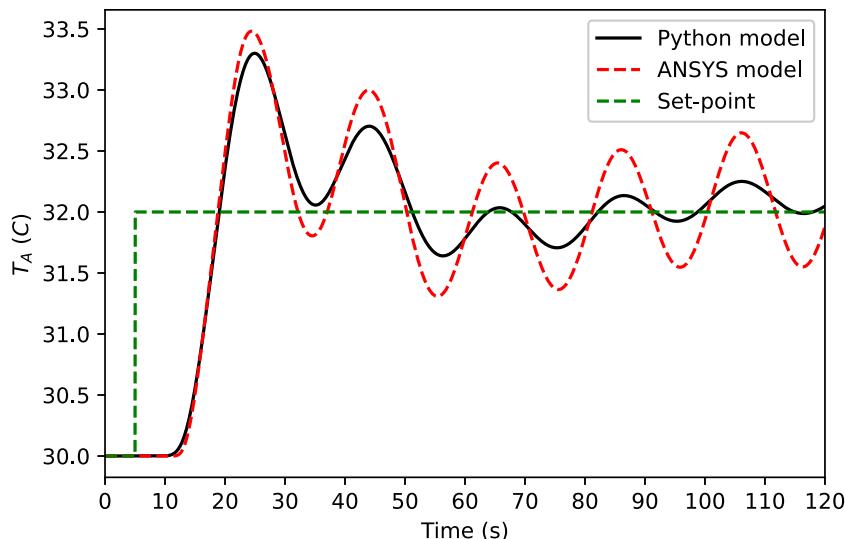


Fig. 40. Inlet temperature of the ESC over time using the Python model and a perturbed version of the Python model under the same PID controllers with the control tunings indicated in Table 5.

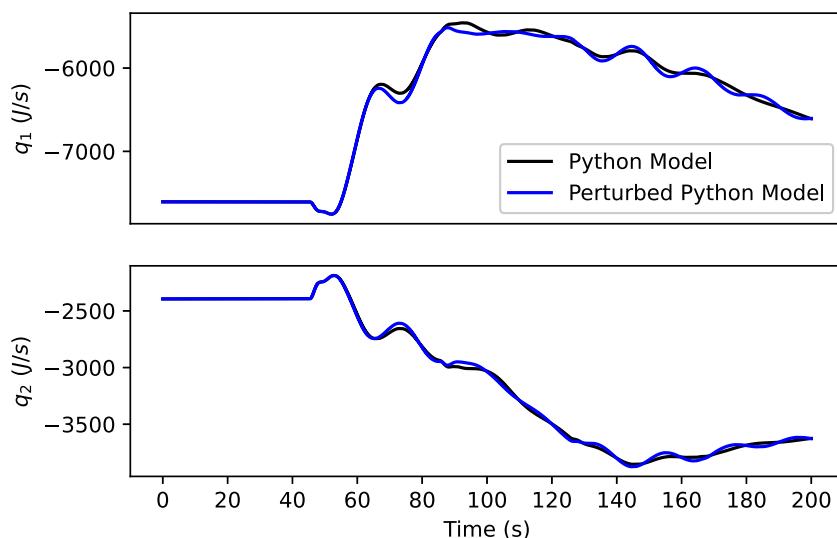


Fig. 41. Heat rate inputs computed using the Python model and a perturbed version of the Python model under the same PID controllers with the control tunings indicated in Table 5.

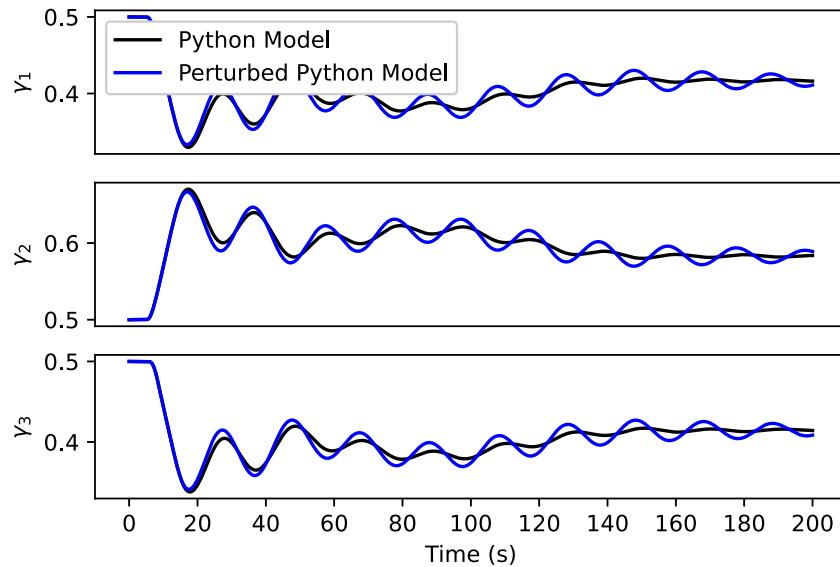


Fig. 42. Cold, hot, and level valve position set-points computed using the Python model and a perturbed version of the Python model under the same PID controllers with the control tunings indicated in Table 5.

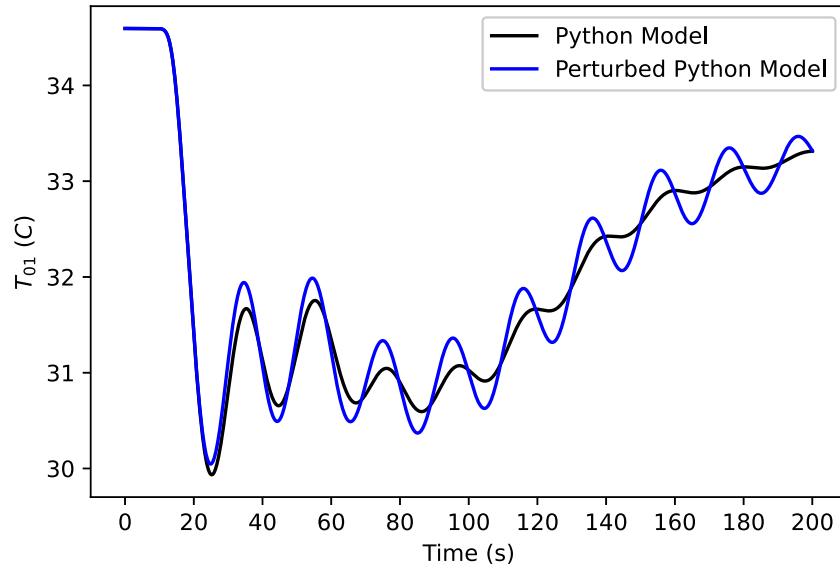


Fig. 43. Inlet temperature of the TCU Ch. 1 over time using the Python model and a perturbed version of the Python model under the same PID controllers with the control tunings indicated in Table 5.

time constants τ_{F_i} of the flow sensor dynamics, time constant τ_{T_c} and time delay θ_{T_c} of the WEP/ESC, and time delay θ of the TCU Channels. The next step is to validate the selection of all process parameters for the proposed etch cooling system. To do this, the portion corresponding to the T-junction and ESC (i.e., the pipe segments with T_{A1} , T_{A2} , T_A and T_C from Fig. 4) of the open-loop Python model described in Section 3 (with the heat rate to the ESC unit set to zero) using the process parameters from Table 7 were benchmarked against industrial responses that might be expected for an etch cooling system (i.e., involving reasonably small time constants and time delays, generally on the order

of magnitude of several seconds), in particular a normalized response of our industrial partner. In particular, the process parameters in Table 7 were chosen.

Specifically, the inlet flow rate and temperature to the ESC must be close to typical industrial responses for an etch cooling system, which is the main feature of the process dynamics that we are interested to capture in the reduced-order model for control design considerations. The benchmark results of the model with a typical industrial response from our industrial partner are shown in Fig. 62 (the industrial data is normalized). Specifically, a step-change in the set-point of the hot

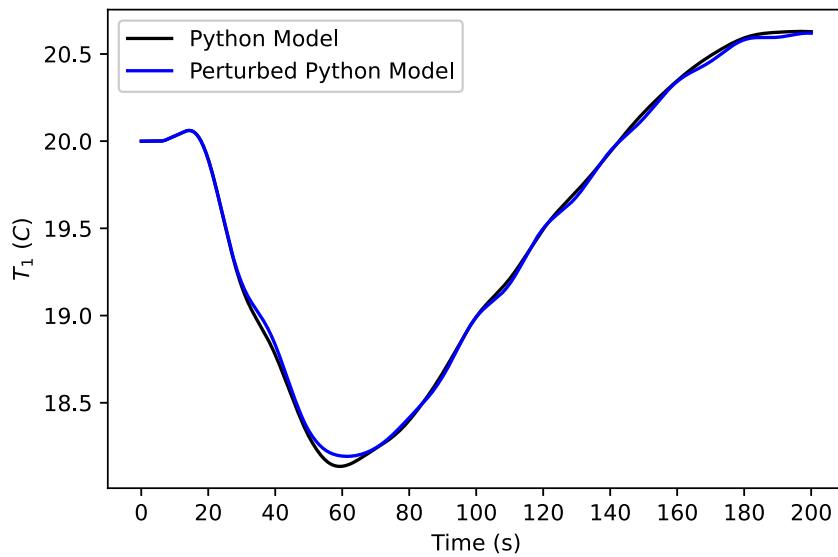


Fig. 44. Outlet temperature of the TCU Ch. 1 over time using the Python model and a perturbed version of the Python model under the same PID controllers with the control tunings indicated in Table 5.

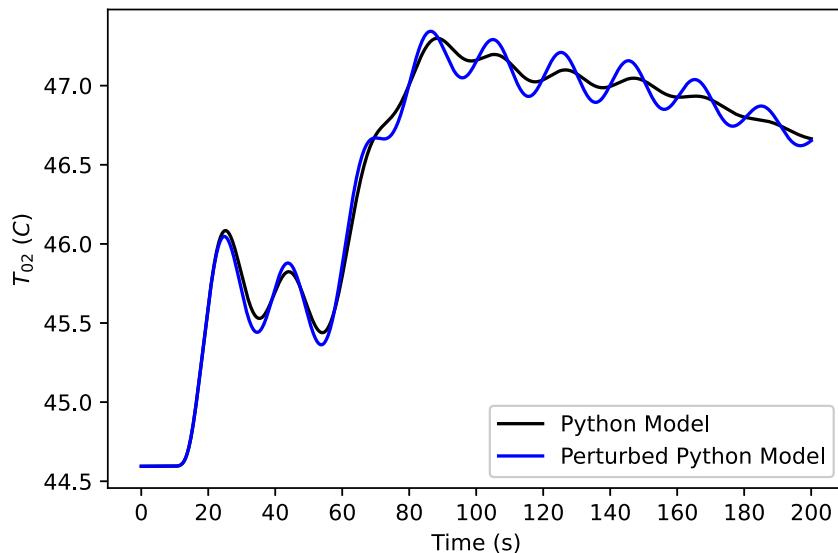


Fig. 45. Inlet temperature of the TCU Ch. 2 over time using the Python model and a perturbed version of the Python model under the same PID controllers with the control tunings indicated in Table 5.

Table 7
Parameters for the WTC system.

Parameter	Value	Unit
τ_y	1	s
τ_{T_i}	9	s
τ_{F_i}	5	s
τ_{T_c}	9	s
θ_{T_c}	1.7	s
θ	40	s

valve position from 50% open to 30% open (in which the 70% of the fluid flow goes to the T-junction before the ESC) was applied after 114 s of operation in the Python model. We observed that the measured temperature leaving the T-junction has a time constant of about 13 s (which matches the typical industrial time constant presented in Fig. 62) and a time delay of 6.5 s, which is 2.5 s higher than the typical industrial response presented in Fig. 62. The time delay of 6.5 s was not adjusted to match exactly the typical industrial time delay as this would require the modification of the pipe lengths that form the T-junction section before the ESC, which would cause the fully developed flow behavior in the pipe entering the ESC not to be satisfied. In addition,

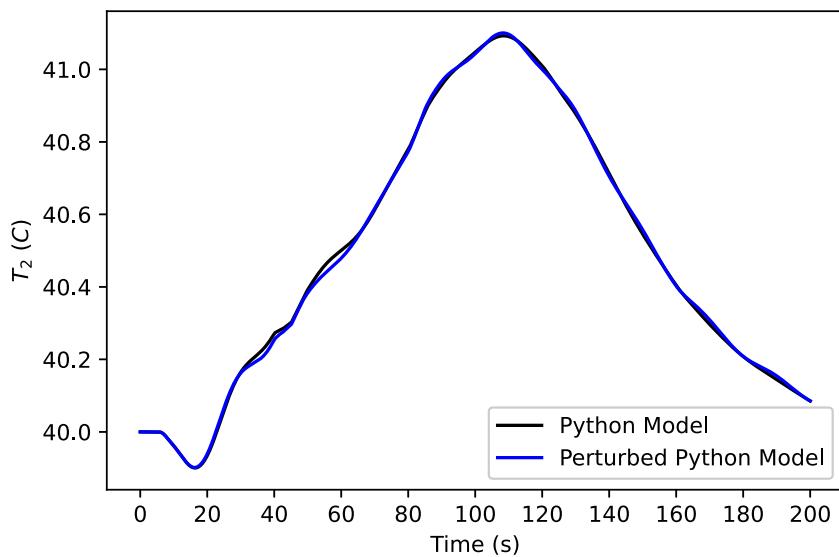


Fig. 46. Outlet temperature of the TCU Ch. 2 over time using the Python model and a perturbed version of the Python model under the same PID controllers with the control tunings indicated in Table 5.

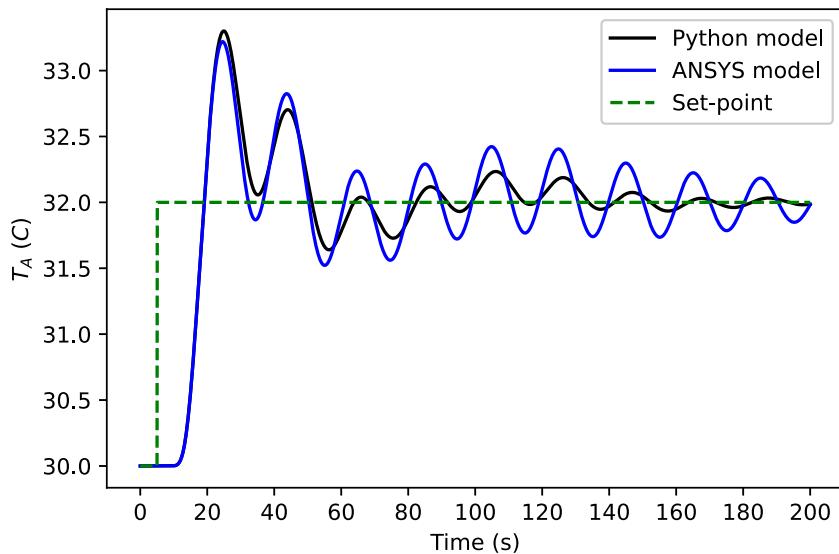


Fig. 47. Inlet temperature of the ESC over time using the Python model and a perturbed version of the Python model under the same PID controllers with the control tunings indicated in Table 5.

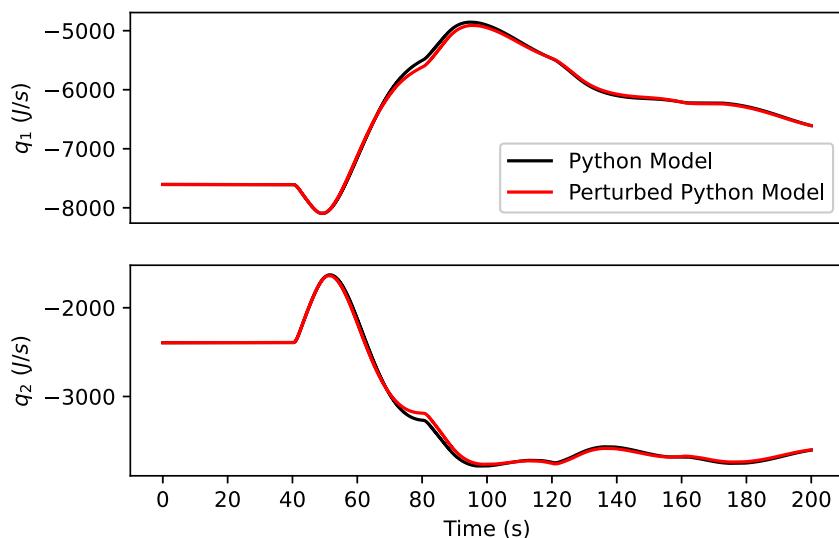


Fig. 48. Heat rate inputs computed using the Python model and a perturbed version of the Python model under the same PID controllers with the control tunings indicated in Table 6.

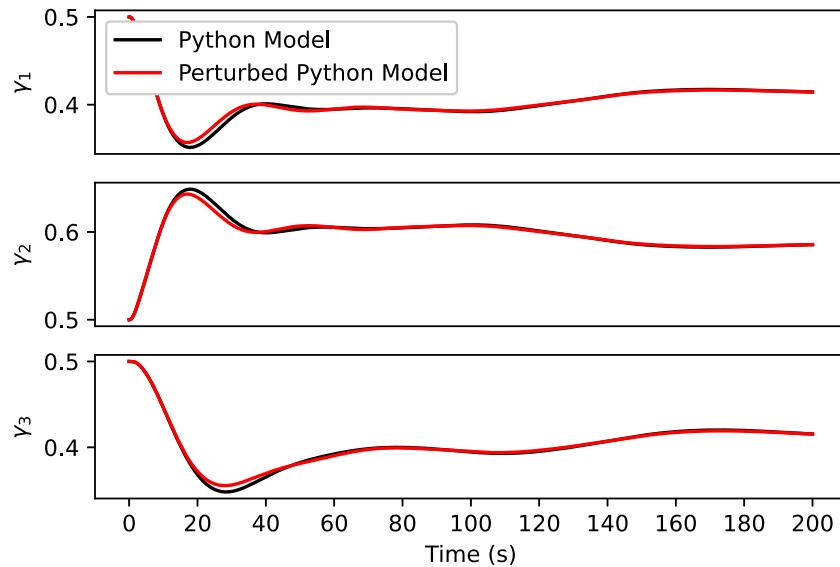


Fig. 49. Cold, hot, and level valve position set-points computed using the Python model and a perturbed version of the Python model under the same PID controllers with the control tunings indicated in Table 6.

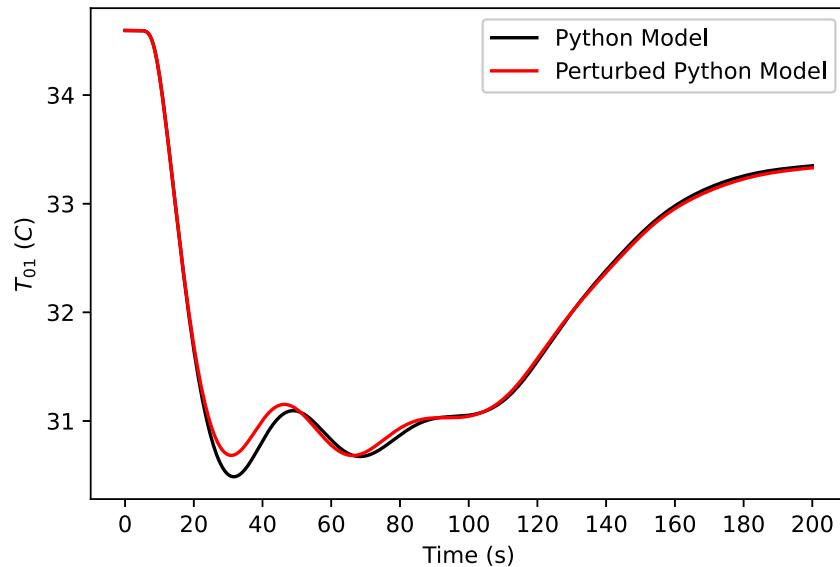


Fig. 50. Inlet temperature of the TCU Ch. 1 over time using the Python model and a perturbed version of the Python model under the same PID controllers with the control tunings indicated in Table 6.

the measured flow rate has no time delay and a time constant of about 6 s, which is in accordance with typical industrial information.

5.4. Python model: Testing closed-loop wafer temperature control system

In this section, a sequence of set-point changes in the flow rate and temperature entering the ESC, and temperatures leaving the TCU's Ch. 1 and 2 is applied to check whether the control tunings provided in Table 6 are adequate to control the WTC system using the process

parameters from Table 7. In particular, we are interested in rapidly adjusting the inlet flow rate and temperature of the ESC upon set-point changes, and keeping these variables close to the set-points. Specifically, the following set-point changes have been implemented: (1) T_{Asp} changes from 30°C to 32°C at 500 s; (2) T_{1sp} , T_{2sp} change from 20°C and 40°C to 22°C and 42°C, respectively, at 1500 s, while F_{Asp} changes from 16.8 L/min to 19 L/min between 1500 s and 2000 s; (3) F_{Asp} changes from 19 L/min back to 16.8 L/min at 2000 s. Figs. 63–64 depict the closed-loop response of the WTC system under

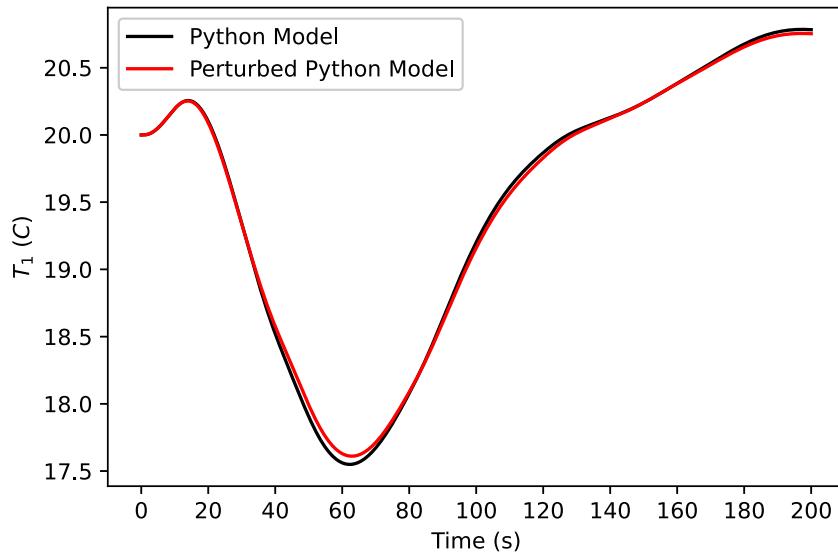


Fig. 51. Outlet temperature of the TCU Ch. 1 over time using the Python model and a perturbed version of the Python model under the same PID controllers with the control tunings indicated in Table 6.

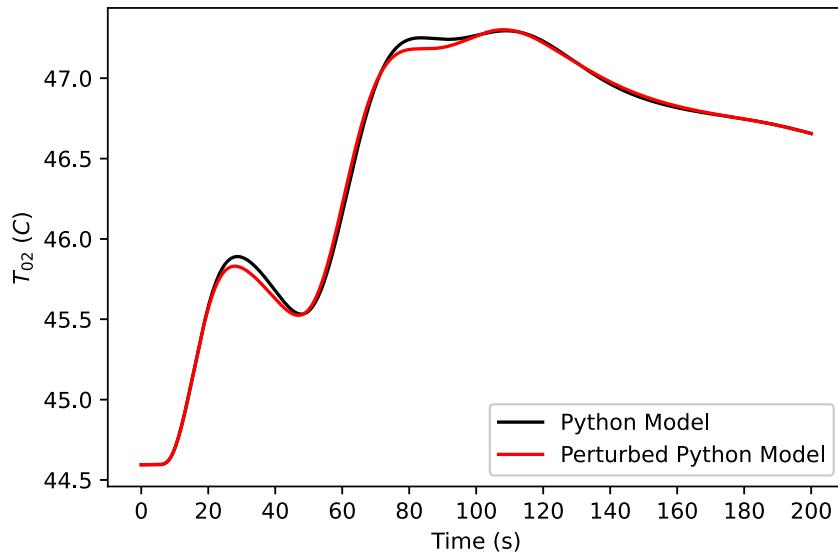


Fig. 52. Inlet temperature of the TCU Ch. 2 over time using the Python model and a perturbed version of the Python model under the same PID controllers with the control tunings indicated in Table 6.

the step-changes in different set-points described above at the times highlighted by the vertical dashed lines. We can observe a fast control tracking response of both coolant supply temperature and flow rate to the ESC, which corresponds to an important requirement for typical WTC systems. In particular, the closed-loop inlet flow rate to the ESC reaches the set-point within a few seconds while the closed-loop inlet temperature to the ESC first reaches the set-point in about 19 s (subtracting the time delay). As expected, the outlet temperatures of TCU Ch. 1 and TCU Ch. 2, and volume difference in the TCU Channels, reach their set-points under the IMC/PID control system. In addition, all the valve positions are adjusted appropriately (no valve saturation) under the IMC/PID control tunings. Thus, a satisfactory

set-point tracking response is achieved for the proposed WTC system using the control parameters in Table 6.

6. Conclusion

The present work detailed the development of a digital model of a wafer temperature control system in ANSYS Fluent and described the procedure for modeling this CFD model. This allows the prediction and evaluation of the expected flow and temperature patterns to select operating strategies based on process data generated from CFD simulations. The digital model developed in this work was benchmarked with typical industrial responses and can be considered an adequate

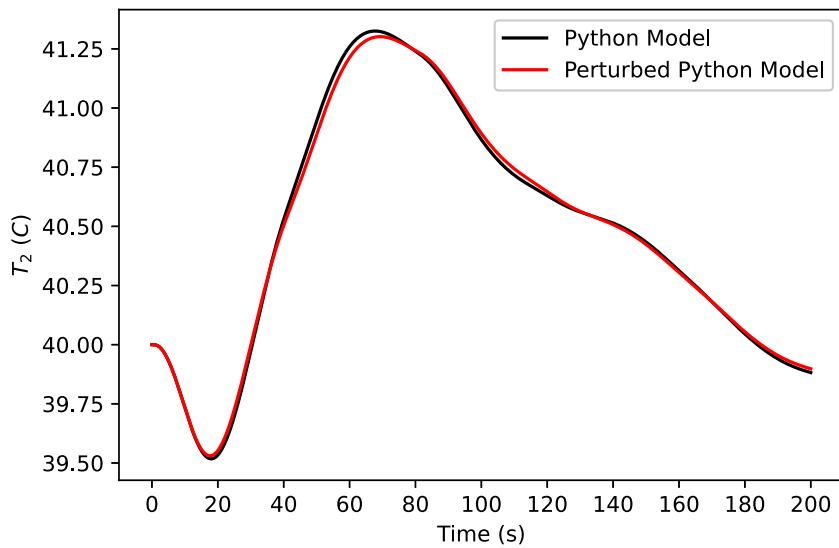


Fig. 53. Outlet temperature of the TCU Ch. 2 over time using the Python model and a perturbed version of the Python model under the same PID controllers with the control tunings indicated in Table 6.

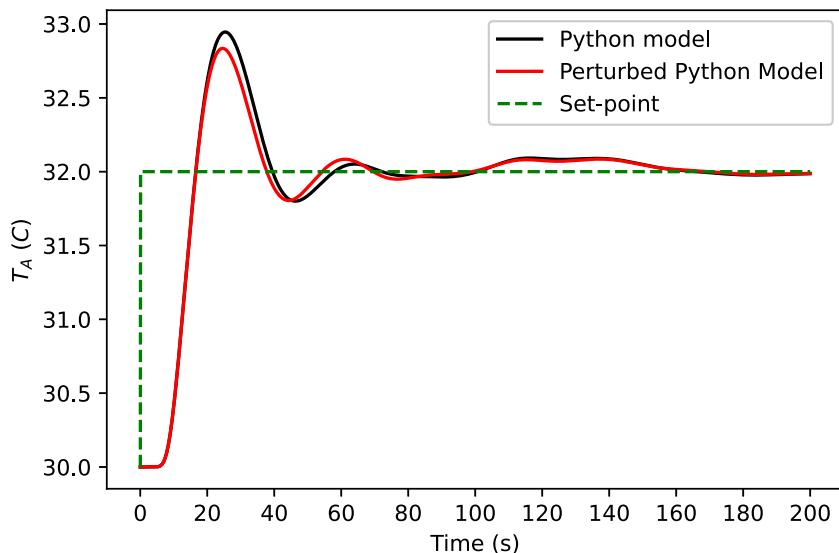


Fig. 54. Inlet temperature of the ESC over time using the Python model and a perturbed version of the Python model under the same PID controllers with the control tunings indicated in Table 6.

representation of an online wafer etching control system. Alternatively, a reduced-order model of the etch cooling system was then developed built upon the data generated from ANSYS simulations. This faster model was implemented in Python for rapid testing of control tunings and designs before implementation in the ANSYS Fluent environment. This series of steps from developing the reduced-order model and the digital twin system can be effective for exploring other operating conditions, potentially those which are more profitable and sustainable.

Future work will explore advanced control designs, in particular using model predictive control (MPC) for the etching control system. MPC is an optimization-based control design, which computes optimal

control actions to apply to a system while accounting for process constraints and predictions of how the system will behave under the control actions in the future. The MPC design could utilize a metric that causes the etch cooling system model to achieve a goal similar to that of classical PID control for this process (i.e., to drive the temperature and flow rate at the inlet of the ESC to their set-point values). However, MPC would implement this with knowledge of the system behavior and constraints when deciding on the control actions (which is not a capability of PID controllers). Furthermore, the flexibility of MPC enables alternative metrics to be optimized in the optimization problem that can represent different operating objectives. For example, a metric

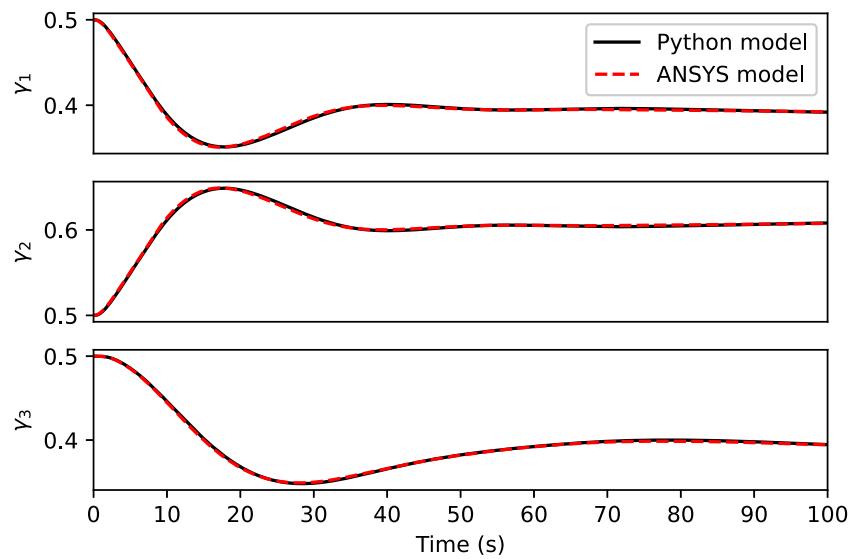


Fig. 55. Cold, hot, and level valve position set-points computed using the Python model and ANSYS model under the same PID controllers with the control tunings indicated in Table 6.

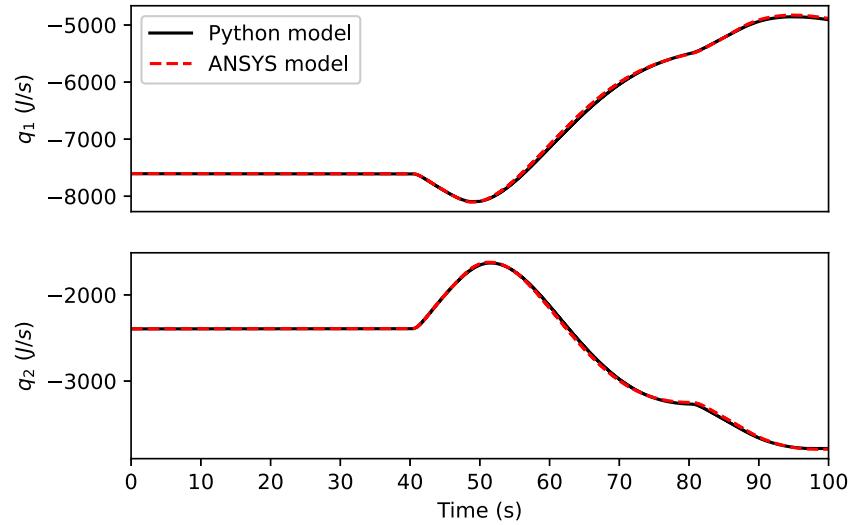


Fig. 56. Heat rate inputs computed using the Python model and ANSYS model under the same PID controllers with the control tunings indicated in Table 6.

representing energy usage could be penalized in the objective function to encourage the controller to choose control actions that require as little energy use as possible. A two-layer MPC structure may be also considered where an upper layer searches for operating conditions that minimize energy use and then a lower layer seeks to track the energy reduction policy as closely as possible while driving the temperature and flow rate of the cooling fluid entering the ESC to its set-point value.

We expect two major benefits from the use of MPC: (1) MPC can handle constraints (such as limits on the heat removal rates in the Temperature Control Units (TCU's) in the WTC) and make predictions of how the process will behave under different control actions; we expect that this would be useful for large changes in the desired temperature of the cooling fluid entering the ESC; (2) MPC can optimize a desired process metric such as energy usage. MPC is expected to outperform PID controllers in cases where there are significant changes in the operating condition (rapidly entering regions that the original tuning parameters of the PID controllers may not be designed for), and would provide flexibility in enabling changing operating conditions without a need to set up many tuning strategies or other heuristic

approaches (i.e., an MPC structure would be able to account for all of the conditions at once in one standardized control framework). We also expect that this control strategy would be able to reduce energy usage via the two-tier framework described above through its foresight of how the process will act under different inputs, and its ability to therefore schedule a variety of changes to the heat removal rates in the TCU's while still considering whether these different heat removal rates would allow the temperature and flow rate specifications of the fluid entering the ESC to be met. Thus, MPC provides flexibility for optimization and may enable new types of operating conditions without a need to re-evaluate the control framework. The proposed reduced-order model may be used in the design and evaluation of the MPC strategy. Though this model requires much less computation time than the ANSYS model, it still requires knowledge of the conditions of the transport fields at points in the system that would not typically be measured (e.g., flow and temperature measurements along each pipe and at mixing points). In this case, a state estimator (e.g., a moving horizon estimator) may be designed for estimating the unmeasured variables from measured variables in the wafer temperature control system.

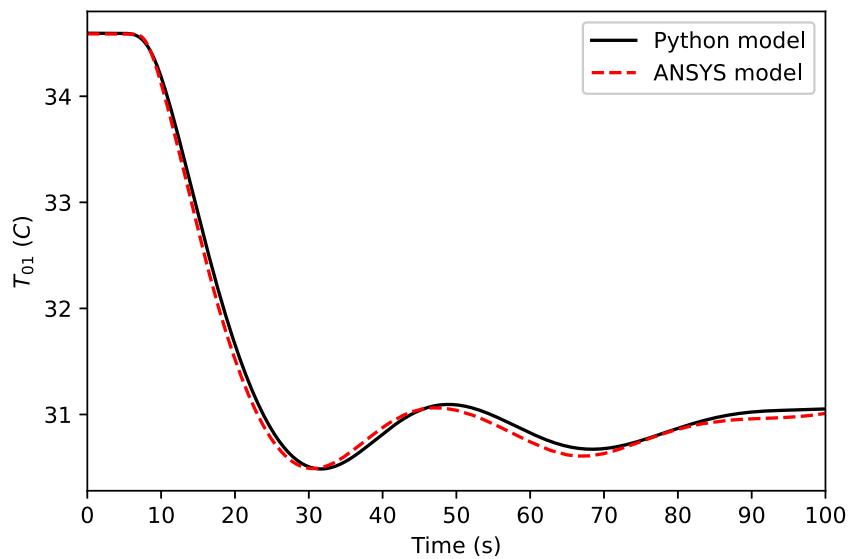


Fig. 57. Inlet temperature of the TCU Ch. 1 over time using the Python model and ANSYS model under the same PID controllers with the control tunings indicated in Table 6.

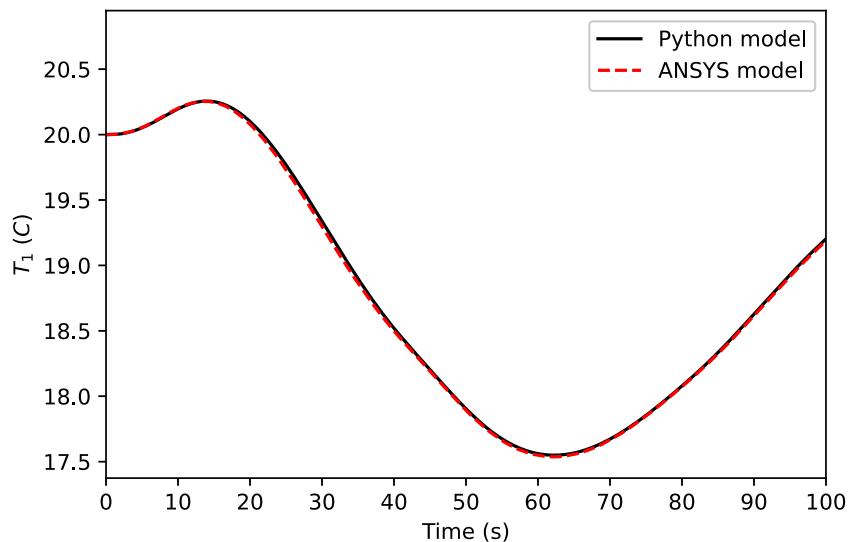


Fig. 58. Outlet temperature of the TCU Ch. 1 over time using the Python model and ANSYS model under the same PID controllers with the control tunings indicated in Table 6.

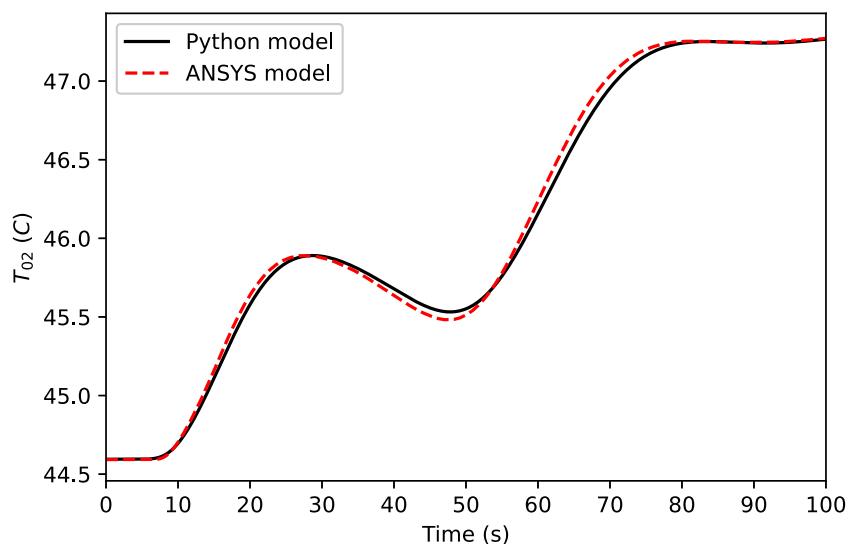


Fig. 59. Inlet temperature of the TCU Ch. 2 over time using the Python model and ANSYS model under the same PID controllers with the control tunings indicated in Table 6.

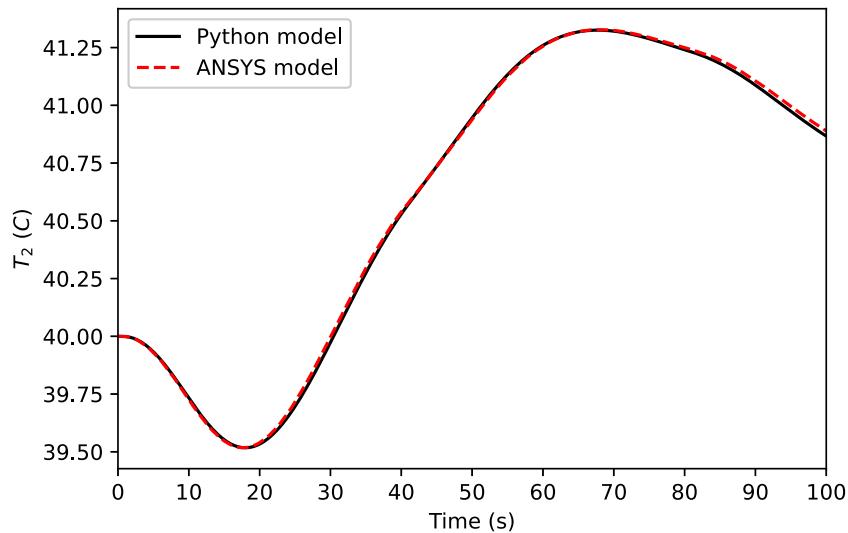


Fig. 60. Outlet temperature of the TCU Ch. 2 over time using the Python model and ANSYS model under the same PID controllers with the control tunings indicated in Table 6.

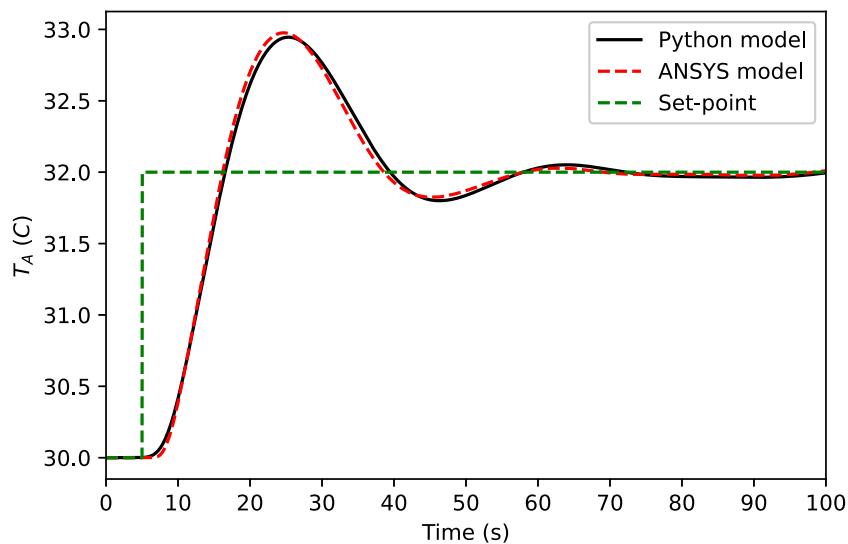


Fig. 61. Inlet temperature of the ESC over time using the Python model and ANSYS model under the same PID controllers with the control tunings indicated in Table 6.

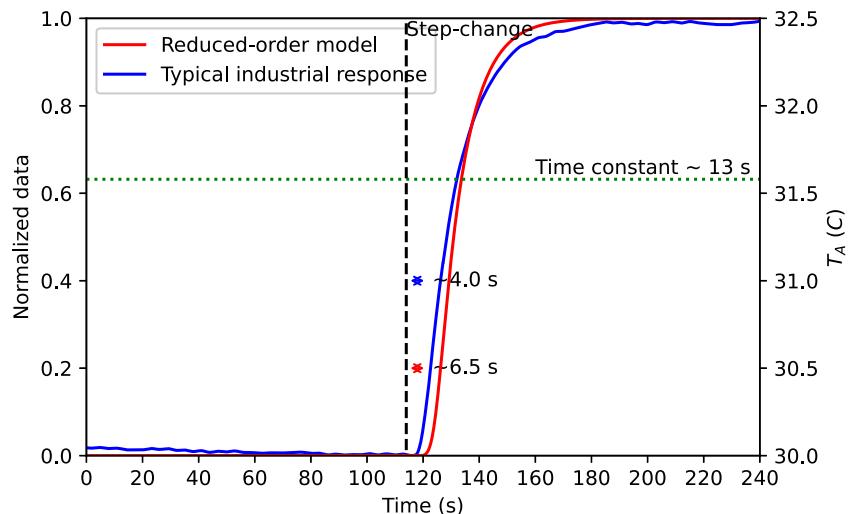


Fig. 62. Open-loop model validation for inlet temperature to the ESC.

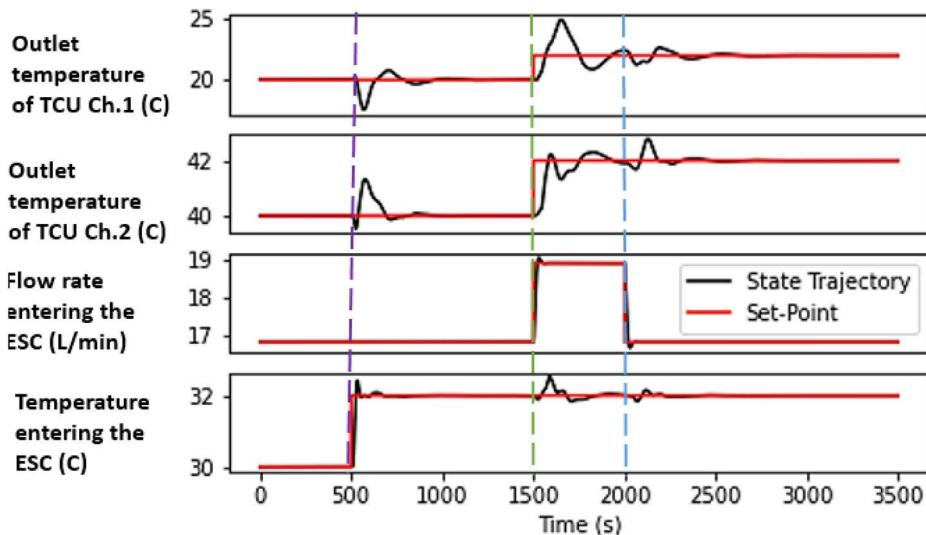


Fig. 63. Closed-loop response of the WTC system under the control tunings indicated in Table 6.

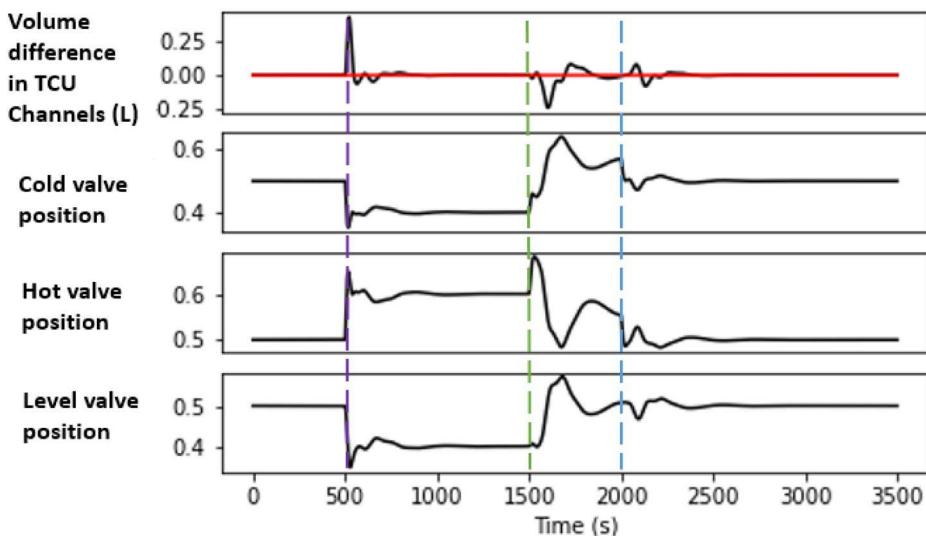


Fig. 64. Valve positions over time and closed-loop response of the WTC system under the control tunings indicated in Table 6.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Helen Durand reports financial support was provided by Lam Research Corp. Dr. Durand is on the editorial board of Digital Chemical Engineering and is one of the co-editors for the special issue on Autonomy, Safety, and Security for Cyber-Physical Systems in the Process Industries that this is being submitted to.

Acknowledgments

Financial support from the Lam Research Corporation, United States, the Wayne State Graduate School Summer Dissertation Fellowship, the National Science Foundation, United States CBET-1839675 and CNS-1932026, and Wayne State University, United States is gratefully acknowledged.

References

Ding, Yangyao, Zhang, Yichi, Chung, Ho Yeon, Christofides, Panagiotis D., 2021. Machine learning-based modeling and operation of plasma-enhanced atomic layer deposition of hafnium oxide thin films. *Comput. Chem. Eng.* 144, 107148.

Ding, Yangyao, Zhang, Yichi, Kim, Keegan, Tran, Anh, Wu, Zhe, Christofides, Panagiotis D., 2019. Microscopic modeling and optimal operation of thermal atomic layer deposition. *Chem. Eng. Res. Des.* 145, 159–172.

Ding, Yangyao, Zhang, Yichi, Orkoulas, Gerasimos, Christofides, Panagiotis D., 2020. Microscopic modeling and optimal operation of plasma enhanced atomic layer deposition. *Chem. Eng. Res. Des.* 159, 439–454.

Kanarik, Keren J., Tan, Samantha, Gottscho, Richard A., 2018. Atomic layer etching: rethinking the art of etch. *J. Phys. Chem. Lett.* 9 (16), 4814–4821.

Kuo, Yue, Lee, Sangheon, 2001. Room-temperature copper etching based on a plasma-copper reaction. *Appl. Phys. Lett.* 78 (7), 1002–1004.

Kuo, Yue, Lee, Sangheon, 2004. A new, room-temperature, high-rate plasma-based copper etch process. *Vacuum* 74 (3–4), 473–477.

Lee, Chris G.N., Kanarik, Keren J., Gottscho, Richard A., 2014. The grand challenges of plasma etching: a manufacturing perspective. *J. Phys. D: Appl. Phys.* 47 (27), 273001.

Mozumder, Purnendu K., Barna, Gabriel G., 1994. Statistical feedback control of a plasma etch process. *IEEE Trans. Semicond. Manuf.* 7 (1), 1–11.

Roozeboom, F., van den Bruele, F., Creyghton, Y., Poodt, P., Kessels, W.M.M., 2015. Cyclic etch/passivation-deposition as an all-spatial concept toward high-rate room temperature atomic layer etching. *ECS J. Solid State Sci. Technol.* 4 (6), N5067.

Seborg, Dale E., Edgar, Thomas F., Mellichamp, Duncan A., Doyle, Francis J., 2017. *Process Dynamics and Control*. John Wiley & Sons.

Singh, Maulshree, Fuenmayor, Evert, Hinchy, Eoin P., Qiao, Yuansong, Murray, Niall, Devine, Declan, 2021. Digital twin: Origin to future. *Appl. Syst. Innov.* 4 (2), 36.

Sun, Mei, Gabriel, Calvin, 2002. Direct wafer temperature measurements for etch chamber diagnostics and process control. In: 13th Annual IEEE/SEMI Advanced Semiconductor Manufacturing Conference. *Advancing the Science and Technology of Semiconductor Manufacturing*. ASMC 2002 (Cat. No. 02CH37259). IEEE, pp. 134–139.

Sun, Jian, Iwasaki, Takuya, Muruganathan, Manoharan, Mizuta, Hiroshi, 2015. Lateral plasma etching enhanced on/off ratio in graphene nanoribbon field-effect transistor. *Appl. Phys. Lett.* 106 (3), 033509.

Tachi, Shinichi, Tsujimoto, Kazunori, Okudaira, Sadayuki, 1988. Low-temperature reactive ion etching and microwave plasma etching of silicon. *Appl. Phys. Lett.* 52 (8), 616–618.

Tinck, Stefan, Tilocher, Thomas, Georgieva, Violeta, Dussart, Rémi, Neyts, Erik, Bogaerts, Annemie, 2017. Concurrent effects of wafer temperature and oxygen fraction on cryogenic silicon etching with SF₆/O₂ plasmas. *Plasma Process. Polym.* 14 (9), 1700018.

Wang, Shi-Qing, Macdonald, Paul, Kruger, Michiel, 2006. Plasma etch process diagnosis and control by wireless sensor wafer in semiconductor chip manufacturing. In: 2006 8th International Conference on Solid-State and Integrated Circuit Technology Proceedings. IEEE, pp. 2175–2180.

Watanabe, Satori, Esashi, Masayoshi, Yamashita, Yoshio, 1997. Fabrication methods for high aspect ratio microstructures. *J. Intell. Mater. Syst. Struct.* 8 (2), 173–176.

Wu, Banqiu, Kumar, Ajay, Pamarthi, Sharma, 2010. High aspect ratio silicon etch: A review. *J. Appl. Phys.* 108 (5), 9.

Zhang, Qiaolin, Poolla, Kameshwar, Spanos, Costas J., 2008. One step forward from run-to-run critical dimension control: Across-wafer level critical dimension control through lithography and etch process. *J. Process Control* 18 (10), 937–945.