Chapter 9
Adaptive Plasma and Machine Learning

Taeyoung Lee and Michael Keidar

Contents

Abstract Cold Atmospheric Plasma (CAP) jets provide a unique combination of reactive oxygen species, reactive nitrogen species, photons, and electric fields that exhibit desirable properties of triggering cell death pathway, selectively for cancer cells. However, the effects of CAP on cancer cells vary substantially depending on the particular type of cancer cell under treatments as well as various properties of plasma jet, such as gas composition, discharge voltage, and treatment duration. On the other hand, artificial intelligence has demonstrated remarkable capabilities in decision-making under uncertainties. Adaptive plasma in conjunction with artificial intelligence could lead to breakthroughs in autonomous, personalized cancer treatments. This chapter presents mathematical modeling of plasma cancer treatment, and summarizes the recent results in adaptive learning plasma.

T. Lee (✉)
Mechanical and Aerospace Engineering, The George Washington University, Washington, DC, USA
e-mail: tylee@gwu.edu

M. Keidar
Mechanical and Aerospace Engineering, School of Engineering and Applied Science, The George Washington University, Washington, DC, USA
e-mail: keidar@gwu.edu

M. Keidar (ed.), *Plasma Cancer Therapy*, Springer Series on Atomic, Optical, and Plasma Physics 115, https://doi.org/10.1007/978-3-030-49966-2_9

9.1 Introduction

Cold atmospheric plasma (CAP) jet is formed by ionization of noble gases, such as helium and argon, initiated when the gas jet flows through a high electric field.

The recent progress has led to a generation of CAP with the corresponding ion temperature close to the room temperature in a laboratory setting [1], and it is also referred to as non-thermal plasma jet or non-equilibrium plasma jet [2].

CAP jet has attracted much attention in the past decade due to its potential application in cancer therapy: it triggers cell death pathway in cancer cells while leaving normal cells unharmed. Various researchers have pointed that its therapeutic effects are related with the reactive oxygen and nitrogen species (RONS), including atomic nitrogen and oxygen, hydroxyl (OH), singlet delta oxygen, superoxide, and nitric oxide (NO) [3–7].

In particular, it is shown that the CAP jet eliminates cancer cells in vitro selectively with minimal damage to healthy cells, and furthermore, it significantly reduces tumor size in vivo [8]. The selectivity is suggested to be from the synergistic effect of selective diffusion of RONS into tumor cells and the high basic RONS level in tumor cells [9, 10]. As the level of RONS is already increased in cancer cells, the additional RONS delivered by CAP causes leads to death pathways especially for cancer cells. More specifically, these cause damage in DNA, causing cell cycle arrest in G2/M or programmed cell death referred to as apoptosis.

Despite success with various in vitro and in vivo experiments, there are several challenges in reliable CAP cancer treatment [11]. First, the therapeutic effect of the CAP jet is susceptible to the variability of plasma parameters (such as discharge voltage, flow rate, and frequency) and exogenous disturbances (such as temperature, target properties, and gas composition of surrounding environment) [12–14]. Second, different types of cancers exhibit different responses when exposed to the same CAP treatment conditions. Change of cancer type, properties of the medium in the device, and the treating duration exposure by plasma jet can drastically influence the plasma characteristics and their effects on the viability of cancer cells [15]. Third, the underlying biological mechanisms of novel therapy approaches have not been fully understood and guidelines on how to schedule these therapies may need to be established. Due to the complexity in clinical trials, the scheduling of treatments is often guided by exhaustive and expensive trial-and-error approaches.

One of the unique features of CAP compared to other sources of reactivity is the ability to rapidly change the reactive species production. As such, it is possible to adjust the CAP parameters such that the reactive species delivered to cells are customized in real-time according to the cancer cell response. Based on these, the idea of adaptive plasma for medical application was proposed recently in [16, 17]. More specifically, the objective is to develop an autonomous, self-adaptive therapeutic system that determines the parameters in the device generating plasma, after diagnosing the response of the cancer cells to CAP in a particular patience under treatment.

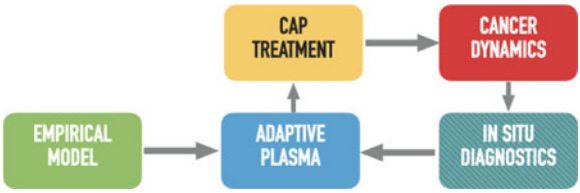

Fig. 9.1 Schematics of adaptive plasma

The component of the adaptive plasma framework is illustrated in Fig. 9.1. First the prior knowledge of the cancer cell response to CAP is represented in the modeling block. This corresponds to our best projection for the behavior of cancer in the future, which can be constructed by a mathematical model of cancer cell dynamics or prior experiences in CAP cancer treatments. Then, according to the model, the adaptive plasma system is designed to plan the parameters to generate CAP for a specific treatment goal. Next, the actual cancer cell response is diagnosed in real-time, and according to the discrepancy between the empirical model and the in situ diagnostics, the adaptive plasma system learns the characteristics of the particular cancer cell under treatments and adjusts the treatment conditions accordingly.

This framework essentially corresponds to a real-time feedback mechanism in control system engineering or a decision-making process in machine learning. However, there is a unique challenge in developing the adaptive plasma system particularly due to the complexities of cancer cell dynamics. The conventional control systems rely on a mathematical model of the dynamic system, which is often formulated as an ordinary differential equation that describes the time evolution of its state, derived according to the first principles. While there have been adaptive control techniques to handle uncertainties in the mathematical model, or robust control to mitigate the effects of unknown disturbances, they focus on a certain class of parametric uncertainties or additive disturbance perturbing the given dynamics. The cancer cell response to CAP involves various complicated processes in biochemistry under a spectrum of time scales, and as such, it is not feasible to construct a mathematical model with sufficient reliability and simplicity required for real-time implementation of such control systems.

On the other hand, reinforcement learning has been successfully applied to sequential decision-making for a Markov process [18]. Reinforcement learning aims to construct a so-called action-value function to evaluate the projected outcome of each possible action, in selecting the optimal action at the current time. While this approach is comparable to the conventional optimal control, the distinct feature is that the action-value function is constructed in situ based on the prior experiences, possibly avoiding the need to construct any dynamic model in prior. However, successful implementation of reinforcement learning often requires numerous trials before the value function converges, and the number of trials exponentially increases as the complexities of the problem is intensified. It is impossible to administer multiple trials for cancer patient in clinics. Even conducting several in vitro

experiments requires nontrivial efforts and costs. As such, it is not practical to implement reinforcement learning directly to the proposed adaptive plasma system.

Considering the above challenges of control system engineering and reinforcement learning in adaptive plasma, the most reasonable choice is integrating both approaches in a synergistic fashion. More specifically, we propose to construct an empirical model by utilizing a limited set of experimental data or by characterizing the overall behaviors from the current knowledge of treatment mechanism. Starting from this empirical model, learning control techniques can be designed to generate self-adaptive CAP treatments that adjust both of the treatment parameters and the empirical model in real-time. This can be further extended to reinforcement learning that does not require extensive pre-training.

Finally, our approach should be distinguished from model predictive control and machine learning presented in [19–21], which are designed to regulate treatment conditions of a device producing CAP jet, such as substrate temperature, plasma current, and power with no consideration on the actual cellular and tissue response. In contrast, we focus on adjusting plasma treatment conditions adaptive to the cellular response of cancer.

This chapter is organized as follows. In the first part, we present two mathematical models, namely an oxidative DNA stress model and an empirical model that represent the dynamic response of cancer cells to CAP. Next, three control approaches based on adaptive learning control and reinforcement learning are presented.

9.2 Mathematical Model of Plasma Cancer Treatment

As discussed above, it is implausible to derive a dynamic model describing the cancer cell responses to CAP from first principles. However, a dynamic representing our current knowledge of the corresponding mechanism will be critical for the successful implementation of adaptive plasma, in terms of scheduling an initial treatment plan to be updated or reducing the amount of data required for reinforcement learning. This section presents two particular attempts: an analytical model based on the oxidative DNA stress caused by CAP, and an empirical model constructed by a limited set of experiments.

9.2.1 Oxidative DNA Stress Model

The cell cycle is a series of phases that a cell goes through when it is divided into two daughter cells [22], which happen aggressively in cancer cells. It is composed of the first growing phase G_1, the synthesis S to replicate DNA, the second growing phase G_2, and the mitosis phase M where the cell is literally divided. To guarantee successful cell division over these delicate processes, there are several mechanisms

to ensure its proper progression, referred to as cell cycle check points. In particular, multiple check points are involved in the transition from G_2 to M to examine DNA for the possible damage or incomplete replication. The presence of DNA damage detected in these check points prevents the cells from transitioning into the next phase for division, causing the cell cycle arrest, which eventually leads to a programmed cell death called apoptosis.

In [23, 24], the following hypotheses are presented to explain the effects of CAP on the cell cycle:

- the plasma treatment causes oxidative DNA stress at the most vulnerable S phase;
- the increased damage in DNA results in G_2/M cell cycle arrest and apoptosis.

This proposition describes the effects of CAP reasonably without excessive complication. Here, we present a mathematical model representing the above effects of CAP on the cell cycle.

Cell Cycle Dynamics

To formulate the above hypotheses mathematically, we first present a cell cycle model that specifies the population density of cells with respect to the stress level at each cell cycle. In other words, we specify the distribution of the oxidative stress for the cells going through a specific cell cycle.

More specifically, let $x \in [0, \infty)$ be the level of oxidative DNA stress, represented by a positive real number. We assume $x = 0$ represents no stress, and the stress is more intense as x becomes greater. The density of cells at time t for the specific stress level x is denoted by $g_1(t, x)$, $s(t, x)$, and $g_2(t, x) \in [0, \infty)$, respectively, for the cell cycle G_1, S, and G_2/M. We do not distinguish M from G_2, as the conventional flow cytometry is not able to separate those two cycles. However, the proposed model is readily extended to four cell cycles.

According to the presented density model, cell population at the stress interval $[x, x + dx]$ is given by $g_1(t, x)dx$ for the G_1 phase, and the total cell population at G_1 is given by

$$G_1(t) = \int_0^\infty g_1(t, x)dx.$$

Similarly, the population for S and G_2/M are given by

$$S(t) = \int_0^\infty s(t, x)dx, \quad G_2(t) = \int_0^\infty g_2(t, x)dx.$$

The proposed mathematical model of the cell cycle with oxidative stress is given by

$$\frac{\partial g_1(t,x)}{\partial t} = 2k_2(x)g_2(t,x) - k_1 g_1(t,x), \tag{9.1}$$

$$\frac{\partial s(t,x)}{\partial t} = -v(u)\frac{\partial s(t,x)}{\partial x} + d(u)\frac{\partial^2 s(t,x)}{\partial x^2} + k_1 g_1(t,x) - k_s s(t,x), \tag{9.2}$$

$$\frac{\partial g_2(t,x)}{\partial t} = k_s s(t,x) - k_2(x)g_2(t,x) - \mu_2(x)a(t,x;T_a), \tag{9.3}$$

where the cells subject to apoptosis is given by

$$a(t,x;T_a) = k_s \exp(-k_2(x)T_a)s(t-T_a,x), \tag{9.4}$$

for a time delay $T_a > 0$, and the boundary condition for (9.2) is

$$d(u)\frac{\partial s(t,0)}{\partial x} - v(u)s(t,0) = 0. \tag{9.5}$$

Here, the parameters k denote the rate of transition from one cycle to the next one. For example, k_1 is the rate of transition from G_1 to S, and k_s is the rate of transition from S to G_2/M. In (9.1), the population density at G_1 is reduced by the rate of $k_1 g_1$ leaving G_1, and it is increased by the rate $2k_2 g_2$, where the factor 2 implies that the cells leaving G_2/M at the rate of $k_2 g_2$ are divided. In other words, as there is cell division in the M phase, the influx rate to G_1, namely $2k_2$, is twice of the efflux rate k_2 from G_2/M. Next, in (9.2), the first two terms represent the effects of CAP denoted by u, describing the increase and the diffusion of the stress over the S phase. Finally, (9.3) is for the G_2/M phase, where the last term μ_2 corresponds to the effects of the cell cycle arrest and the apoptosis, which are discussed below. See Fig. 9.2 for an illustration of the above dynamics model.

Effects of CAP Treatments

Now, we discuss how the presented mathematical model reflects the hypotheses on the cell cycle dynamics. Equations (9.1)–(9.3) are defined such that the following three effects are formulated mathematically: the increase of the oxidative stress at S; G_2/M cell cycle arrest; the resulting apoptosis.

Increased Oxidative DNA Stress in S

As discussed above, it is considered that CAP treatment increases the oxidative DNA stress at S. This is modeled by the overall shift and the diffusion of the stress at (9.2) as follows. Assume the intensity or the gas composition of the plasma treatment, such as the voltage and the helium gas flow rate, is parameterized by $u \in \mathbb{R}^m$. The treatment is modeled by the terms $v(u) \in \mathbb{R}$ and $d(u) \in \mathbb{R}$ in (9.2)

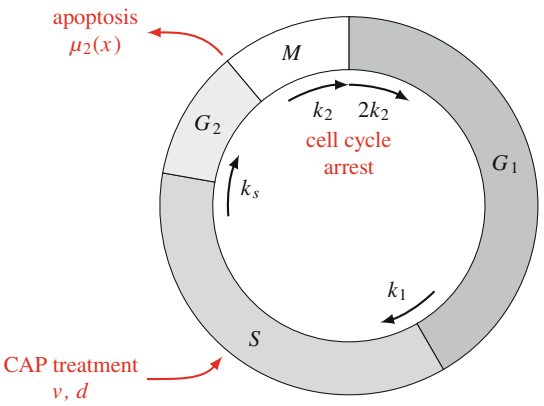

Fig. 9.2 The effects of CAP treatment to the cell cycle dynamics: under nominal cancer growths, the cells go through the phase G_1, S, and G_2/M at the exponential rates of k_1, k_s, and k_2, respectively. Due to the cell divisions, the influx of G_1 is twice of the efflux of G_2/M. As illustrated by red colors, it is assumed that CAP treatments cause oxidative DNA stress at S_2, which result in G_2/M cell cycle arrest represented by reduction of k_2, and apoptosis modeled by μ_2

that correspond to the advection or shift of the stress and the diffusion of the stress, respectively. In other words, the distribution of the stress will shift toward higher x with the rate $v(u)$, and it will be smoothed according to $d(u)$, and the rates of the advection and diffusion are considered to be dependent of the treatment conditions parameterized by u. The given boundary condition (9.5) ensures that no stress is created arbitrarily at $x = 0$.

Cell Cycle Arrest

In the presented cell cycle model, the cell division is represented by the fact that the rate of influx to G_1, namely $2k_2$, is twice to that of the efflux from G_2/M. The effects of the increased cell stress at S to the other cell cycle are accounted by formulating the cell division rate k_2 and the death rate μ_2 of G_2/M as a function of stress x so that only the cells with lower stress go through the mitosis to enter G_1, and the other cells with higher stress are destroyed.

More explicitly, consider the step function $\rho(x)$ defined in the appendix, that is, a smooth function satisfying (9.31). For constants $k_{2_0} \in \mathbb{R}$ and $0 < x_0 < x_1 \in \mathbb{R}$, the cell division rate is defined as

$$k_2(x) = k_{2_0}(1 - \rho(x; x_0, x_1)). \tag{9.6}$$

According to the property of the step function ρ summarized at (9.31),

$$k_2(x) = k_{2_0}, \quad x \le x_0,$$

$$k_2(x) < k_{2_0}, \quad x_0 < x < x_1,$$
$$k_2(x) = 0, \quad x \geq x_1.$$

As such, the cell with the stress lower than x_0 completes the cell division at the fixed rate k_{2_0}, and proceeds to G_1. For other cells with stress greater than x_0, either the cell division rate is discounted for $x > x_0$ or no cell division occurs for $x \geq x_1$. Consequently, the cells with higher stress cannot complete the cell division to G_1, and remain at G_2/M, thereby causing G_2/M cell cycle arrest.

Apoptosis

For mathematical modeling of apoptosis caused by the above cell cycle arrest, the death rate at G_2/M, namely $\mu_2(x)$, is defined as a function of the stress as follows:

$$\mu_2(x) = \mu_{2_0}\rho(x; x_2, x_3), \tag{9.7}$$

for constants $\mu_{2_0}, x_2, x_3 \in \mathbb{R}$ satisfying $x_2 \leq x_3$. Therefore,

$$\mu_2(x) = 0, \quad x \leq x_2,$$
$$\mu_2(x) < \mu_{2_0}, \quad x_2 < x < x_3,$$
$$\mu_2(x) = \mu_{2_0}, \quad x \geq x_3.$$

This implies that no apoptosis occurs when $x \leq x_2$, and the rate is increased to μ_{2_0} for $x \geq x_3$. As such, the cells with higher stress go through apoptosis.

It is further assumed that the apoptosis is completed with the time delay of $T_a > 0$. The main motivation of introducing the time delay is the experimental results showing that the apoptosis occurs a certain period after CAP treatment. As such, the cells subject to apoptosis at t had entered G_2/M at $t - T_a$ with the rate of $k_s s(t - T_a, x)$, and they have evolved according to

$$\frac{\partial a(\tau, x; T_a)}{\partial \tau} = -k_2(x)a(\tau, x; T_a),$$

for $\tau \in [t - T_a, t]$, with the boundary condition $a(t - T_a, x; T_a) = k_s s(t - T_a, x)$. This is identical to (9.3) with $k_s = \mu_2 = 0$.

The above equation is linear, and it yields an explicit solution given by

$$a(t, x; T_a) = \exp(-k_2(x)T_a)k_s s(t - T_a, x). \tag{9.8}$$

The above expression can be implemented even when $t \leq T_a$, assuming $s(t, x) = s(t, 0)$ for $t < 0$.

Special Case: Low-Stress

Here we consider a special case when the oxidative stress is sufficiently low. This corresponds to the case where the cancer cells proliferate naturally without CAP treatments. More explicitly, we assume $x \leq \min\{x_0, x_2\}$ so that

$$k_2(x) = k_{2_0}, \quad \mu_2(x) = 0.$$

In other words, all of the cells complete the cell division and there is no apoptosis. Throughout this subsection, as it is independent of x, $k_2(x)$ is denoted by $k_2(x) = k_2$ for convenience.

From (9.1) and (9.3), it is straightforward to show

$$\dot{G}_1(t) = 2k_2 G_2(t) - k_1 G_1(t),$$
$$\dot{G}_2(t) = k_2 S(t) - k_2 G_2(t).$$

Also, integrating (9.2) with respect to x,

$$\begin{aligned}\dot{S}(t) &= \frac{d}{dt}\int_0^\infty s(t,x)dx = \int \frac{\partial s(t)}{\partial t}dx \\ &= \int \frac{\partial}{\partial x}\left(-vs + d\frac{\partial s}{\partial x}\right) + k_1 G_1(t) - k_s S(t) \\ &= -vs + d\frac{\partial s}{\partial x}\Big|_0^\infty + k_1 G_1(t) - k_s S(t).\end{aligned}$$

Therefore, for given boundary condition (9.5) of zero flux, this reduces to

$$\dot{S}(t) = k_1 G_1(t) - k_s S(t).$$

As such, the dynamics of the cell population is given by the following linear time-invariant system:

$$\begin{bmatrix}\dot{G}_1 \\ \dot{S} \\ \dot{G}_2\end{bmatrix} = \begin{bmatrix}-k_1 & 0 & 2k_2 \\ k_1 & -k_s & 0 \\ 0 & k_s & -k_2\end{bmatrix}\begin{bmatrix}G_1 \\ S \\ G_2\end{bmatrix}. \tag{9.9}$$

The characteristic equation of the above system matrix is given by

$$\lambda^3 + \sum_{i\in\mathcal{I}} k_i\lambda^2 + \frac{1}{2}\sum_{i,j\in\mathcal{I}, i\neq j} k_i k_j \lambda - k_1 k_s k_2 = 0,$$

with $\mathcal{I} = \{1, s, 2\}$.

One can show that there is one positive real eigenvalue, namely λ. Suppose the initial condition is chosen such that $[G_1(0), S(0), G_2(0)]$ is parallel to the corresponding eigenvector of the positive read eigenvalue, i.e., $(A - \lambda I)[G_1(0), S(0), G_2(0)]^T = 0$. Then, the corresponding solution of (9.9) is given by

$$G_1(t) = e^{\lambda_t} G_1(0), \quad S(t) = e^{\lambda t} S(0), \quad G_2(t) = e^{\lambda t} G_2(0). \tag{9.10}$$

Let $C(t) = G_1(t) + S(t) + G_2(t)$ be the total cell population. The above implies

$$C(t) = e^{\lambda t} C(0). \tag{9.11}$$

Therefore, in the proposed model when the oxidative stress is sufficiently low, all of the cell population at each cell cycle and the total cell population grow exponentially with the same rate λ. Also the ratio of each cell cycle to the total cell population remains unchanged.

Furthermore, the parameters k_1, k_s, k_2 can be determined by the cell cycle ratio and the exponential growth factor. More specifically, let $f_1, f_s, f_2 \in [0, 1]$ be the ratio of the cell population at each cycle, i.e.,

$$f_1 = \frac{G_1}{G_1 + S + G_2}, \quad f_s = \frac{S}{G_1 + S + G_2}, \quad f_2 = \frac{G_2}{G_1 + S + G_2}. \tag{9.12}$$

Then, it is straightforward to show k_1, k_s, k_2 are given explicitly as

$$k_1 = \frac{2 - f_1}{f_1}\lambda, \quad k_s = \frac{f_2 + 1}{f_s}\lambda, \quad k_2 = \frac{\lambda}{f_2}. \tag{9.13}$$

As such, the above parameters can be easily identified by the flow cytometry of untreated cancer cells.

Numerical Example

Several numerical examples are presented. Throughout this section, the unit of time is hours if unspecified. The cell cycle ratio is chosen as $f_1 = 0.5$, $f_s = 0.3$, and $f_2 = 0.2$, and the overall growth rate is chosen as $\lambda = \frac{\log 2}{24}$, which represents that the doubling time is 24 h. From (9.13), the corresponding cell cycle transition rates are given by

$$k_1 = 0.0866, \quad k_s = 0.1155, \quad k_{2_0} = 0.1444.$$

The apoptosis rate is $\mu_{2_0} = 1$, and the delay is $T_a = 12$. The parameters to define the step function in $k_2(x)$ and $\mu_2(x)$ are given by $x_0 = 0.6$, $x_1 = 1.4$, $x_2 = 0.7$,

and $x_3 = 1.3$. For CAP treatment, the rate of advection and diffusion is selected as $v = 30$ and $d = 3$, and when there is no treatment, they are changed to zero.

The initial conditions are chosen as

$$g_1(0,x) = f_1 \frac{2}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right), \quad s(0,x) = \frac{f_s}{f_1} g_1(0,x),$$
$$g_2(0,x) = \frac{f_2}{f_1} g_1(0,x),$$

with $\sigma = 0.2$. In other words, the stress is distributed according to the Gaussian distribution, and scaled according to the cell cycle ratio. The resulting initial total cell population is $C(0) = G_1(0) + S(0) + G_2(0) = 1$.

No CAP Treatment

We first consider the case of no CAP treatment. This represents the natural growth of the cancer cells. For the selected initial stress distribution and the parameters of the step function, the majority of cells, excluding less than 0.3% of the population, have low stress less than $\min\{x_0, x_1, x_2, x_3\} = 0.6$. As such, this case is well approximated by the results presented in section "Special Case: Low-Stress."

Figure 9.3 illustrates the simulation results. As presented in section "Special Case: Low-Stress," the total cell population and the population at each cell cycle grow exponentially with the rate λ, and consequently, the cell cycle ratio remains fixed.

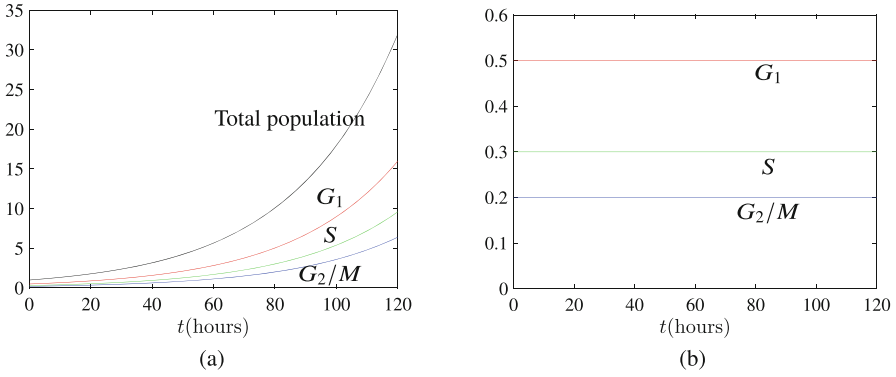

Fig. 9.3 Simulation results for no CAP treatment. (**a**) Cell population growth. (**b**) Cell cycle ratio

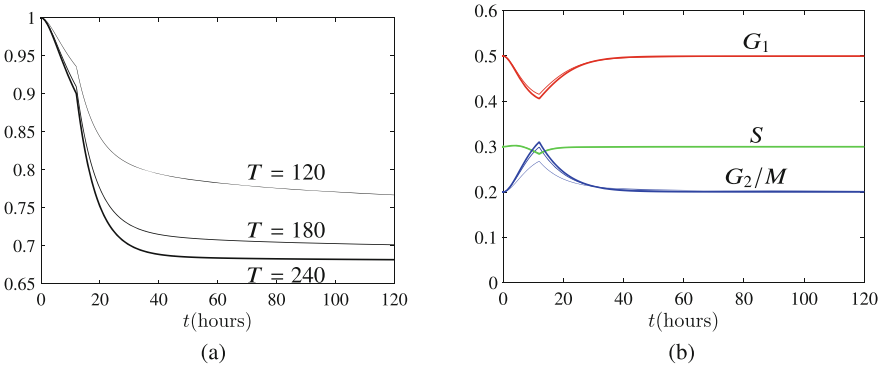

Fig. 9.4 Simulation results with CAP treatments. (**a**) The ratio of the total cell population to the untreated growth for varying treatment duration. (**b**) Cell cycle ratio ($T = 120$:thin, $T = 180$:medium, $T = 240$:thick): the G_2/M cell cycle arrest occurs when $t = 12$ h where the population of G_2/M exceeds those of S

CAP Treatments

Next, we consider the simulation results with CAP treatments. Three cases are presented for varying treatment duration of 120, 180, and 240 s.

The total cell population relative to the untreated growth is shown at Fig. 9.4a. It is also observed that the longer the treatment duration is, the ratio reduces further. There are two phases in the decrease of the ratio. During the first $T_a = 12$ h, the ratio decreases slightly due to the reduced cell division rate, namely $k_2(x)$ for cells with higher stress. Afterwards, the apoptosis contributes to further decrease.

Figure 9.4b illustrates the cell cycle ratio. The ratio for G_2/M increases until $t = 12$ h, and it exceeds the S phase temporarily, representing the G_2/M cell cycle arrest. After the cells with higher stress are destroyed due to apoptosis, the cell cycle ratio asymptotically converges to the initial value, indicating that the remaining cells with lower stress proliferate in the same fashion presented in the preceding section for the natural growth without treatment.

Figures 9.5, 9.6 and 9.7 show the evolution of the stress distribution for three time segments, when the treatment period is 180 s. In those figures, the gradual increase of the intensity of color represents the time evolution. In Fig. 9.5, it is shown that the stress distribution of S increases due to CAP treatment. The next figure, Fig. 9.6, the cells with the increased stress are transferred from S to G_2/M, and they remain at G_2/M as the transition rate from G_2/M to G_1, namely $k_2(x)$, is discounted for higher stress level. This corresponds to the G_2/M cell cycle arrest, and it explains the peak of the G_2/M cell cycle ratio in Fig. 9.4b. Finally, in Fig. 9.7, the cells with higher stress that have been accumulated at G_2/M go through apoptosis and are destroyed.

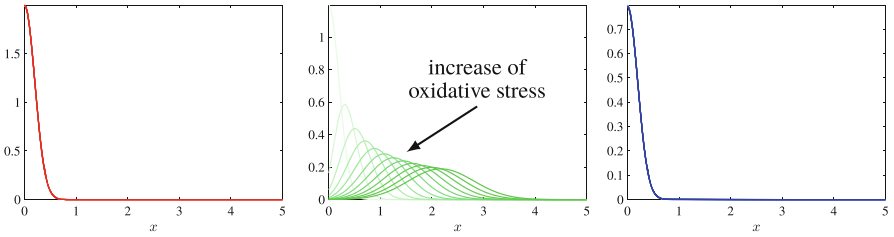

Fig. 9.5 Evolution of oxidative stress distribution for $0 \leq t \leq 180\,\mathrm{s}$: the stress at S is increased due to CAP treatment. (**a**) Population density g_1. (**b**) Population density s. (**c**) Population density g_2

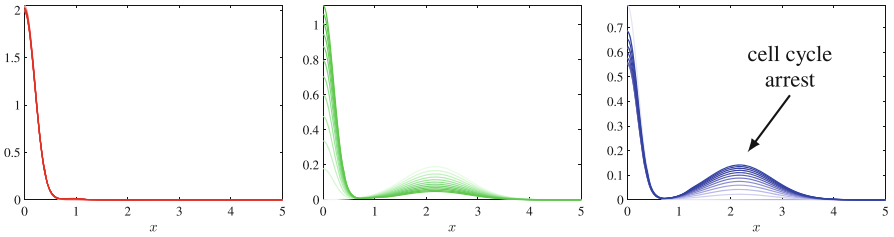

Fig. 9.6 Evolution of oxidative stress distribution for $180\,\mathrm{s} \leq t \leq 12\,\mathrm{h}$: the cells with higher stress level are transferred from S to G_2/M, and they become accumulated at G_2/M, while representing G_2/M cell cycle arrest. (**a**) Population density g_1. (**b**) Population density s. (**c**) Population density g_2

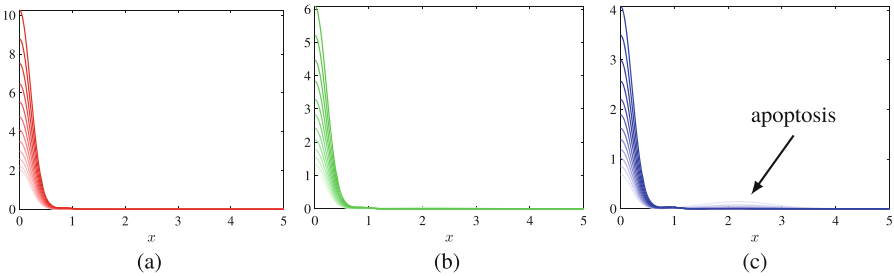

Fig. 9.7 Evolution of oxidative stress distribution for $12\,\mathrm{h}\,\mathrm{s} \leq t \leq 72\,\mathrm{h}$: the cells with higher stress level that have been accumulated at G_2/M go through apoptosis. (**a**) Population density g_1. (**b**) Population density s. (**c**) Population density g_2

9.2.2 Empirical Model

Next, we present another mathematical model of CAP treatments [25]. This is developed to represent the dynamic response of cancer cells under CAP treatment, based on the raw data from in vitro experiments presented in [26]. Cancer cell

response to CAP is monitored over the course of 48 h for two types of cancer cell, namely U87 and MDA-MB-231, where the treatment duration is varied from 0 to 180 s, and the plasma discharge voltage is selected from 3.16 and 3.71 kV. The resulting CAP-induced cell death was investigated by RealTime-Glo MT Cell Viability Assay several times.

To generalize the experimental data over arbitrary treatment conditions and time, the following form of growth model is considered:

$$\dot{p} = pF(t, p), \tag{9.14}$$

where $p \in \mathbb{R}$ denotes the population of cancer cell measured in terms of the metabolic activities of cells. To have the consistent value of p for several experiments presented in [26], we normalize the cancer cell viability under CAP treatments with the initial cancer cell viability just before the CAP exposure. Therefore the initial value is $p(0) = 1$ always, and the variable p is unit-less. Next, $F : \mathbb{R} \to \mathbb{R}$ models its net exponential proliferation rate depending on the current viability and the time.

The objective is to find an analytical expression of F that characterizes the viability of cancer cells under CAP treatment as reported in [26], which exhibit the following properties.

- immediately after CAP treatment, cell viability is reduced instantaneously;
- shortly afterwards, from 0 min to 6 h, the cell viability increases rapidly;
- from 6 to 24 h, the cell viability decreases when the treatment duration is sufficiently large;
- from 24 to 48 h, the cell viability approaches its steady state value;
- for the effect of treating duration and voltage, the cell numbers decrease with the increase of the treating duration and voltage.

Based on these common features, we formulate an expression for the net proliferation rate as follows. To represent the instantaneous reduction of the cell viability, the cell viability immediately after the treatment is given by $p(0^+) = p_0$ for $p_0 \in \mathbb{R}$. Afterwards, the cell viability evolves according to (9.14), where the net proliferation rate is chosen as

$$F(t, p) = (c_1 - c_2 t) \exp\left(- c_3^{-t} p^{c_4} \right) - c_5, \tag{9.15}$$

where $c_1, c_2, c_3, c_4, c_5 \in \mathbb{R}$ are parameters to be determined. The above expression is applied to both types of cancer cells, namely U87 and MDA-MB-231, but c_5 is set to zero for U87.

Next, the values of the free parameters in (9.15) are determined according to optimal system identification [27]. This is to minimize the discrepancy between the experimental data and (9.14) measured by

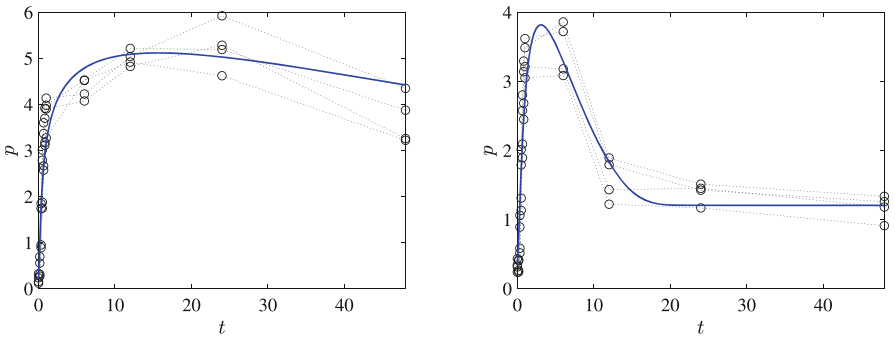

Fig. 9.8 Normalized cell viability for U87 with the discharge voltage of 3.16 kV (blue: analytical model of (9.14); black circle: experimental data). (**a**) 60 s treatment. (**b**) 180 s treatment

$$J(c) = \sum_{i=1}^{n} \int_0^{48} \| p_{\exp_i}(t) - p(t; c) \|^2 dt, \tag{9.16}$$

where $p_{\exp_i}(t)$ denotes the cell viability at t for the i-th experimental data, and $p(t; c)$ corresponds to the value obtained by the mathematical model (9.14) with a given parameter $c = (c_1, c_2, c_3, c_4, c_5) \in \mathbb{R}^5$. The optimal values of the free parameters are selected by

$$c_{\text{opt}} = \arg\min\{J(c)\}. \tag{9.17}$$

This is solved by the nonlinear programming solver, namely `fmincon` in MATLAB for each discharge voltage of $U = 3.16$ and 3.71 kV.

The time evolution of the cell viability predicted by (9.14) is illustrated in Fig. 9.8 against the experimental data of [26] for two selected treatment durations. While the experimental data are noisy, the presented analytical model reflects the overall trend of the data successfully.

The above parameters are optimized for the particular set of the treatment duration and discharge voltages considered in [26]. However, it can be generalized for arbitrary treatment conditions, by assuming that such parameters vary linearly. For example, the cell viability for the treatment duration $\Delta t = 100$ s can be constructed by interpolating the parameters of $\Delta = 90$ and $\Delta t = 180$. Similarly, the effects of the discharge voltage can be generalized as well. They are illustrated in Fig. 9.9.

In Fig. 9.9a, the treatment duration is varied linearly from $\Delta t = 60$ (cyan) to $\Delta t = 180$ (purple) for a fixed discharge voltage $U = 3.16$ kV. Similarly, in Fig. 9.9b, the discharge voltage is varied from $U = 3.16$ kV (cyan) to $U = 3.71$ kV (purple) when $\Delta t = 90$ s. As such, the presented empirical model (9.14) can be utilized to predict the dynamics of cancer cell viability for arbitrary treatment conditions.

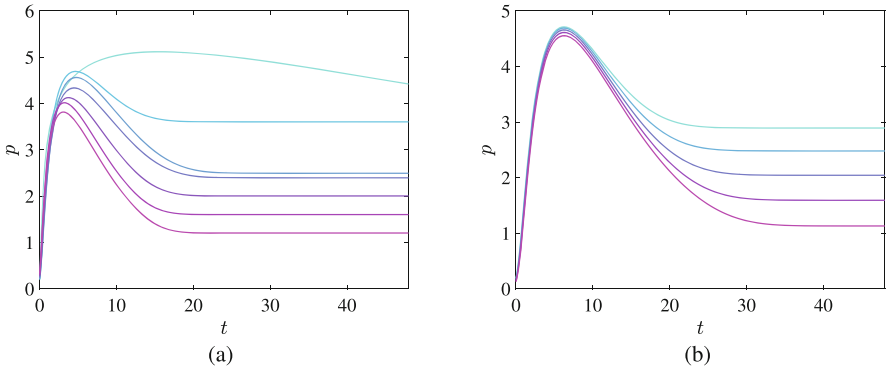

Fig. 9.9 Normalized cell viability for U87 generalized for arbitrary treatment conditions (blue: smaller values; purple: larger values). (**a**) Treatment duration varying from 60 to 180 s. (**b**) Discharge voltage varying from 3.16 to 3.71 kV

9.3 Adaptive Plasma

The mathematical model of the prior section can be utilized to plan a baseline treatment schedule for a given objective. For example, the treatment duration and the plasma discharge voltage can be chosen such that the cancer cell viability is reduced to a desired level after a certain period. However, the effects of CAP on cancer cells vary substantially depending on the various factors, such as the size and the type of the cancer cell under treatment, ambient temperature, and humidity. It is also susceptible to exogenous disturbances not accounted in the mathematical model. Consequently, it cannot be expected that the actual cancer response would follow the one predicted by the model. The key component of adaptive plasma is to address this issue by adjusting the treatment conditions based on in situ diagnostics of the actual response to compensate the discrepancy between the model and the actual response. This section presents three such approaches, namely model predictive control, adaptive learning control, and reinforcement learning, based on the empirical model in Sect. 9.2.2.

9.3.1 Model Predictive Control

A model predictive control (MPC) is a control strategy to convert the solution of open-loop optimal control into a feedback control [28]. The idea is applying the optimal control repeatedly over a certain time period. As each optimization is initiated by the current state vector, the corresponding control input constructed from MPC is a feedback.

We first formulate an optimal control problem as follows. Let the control parameters be the CAP treatment duration Δt. The objective is to minimize the treatment time, while ensuring that the cancer cell viability is reduced to the desired level. This is to maximize the therapeutic effect of CAP treatments for a prescribed level of cancer growth inhibition. More explicitly, the objective function that is to be minimized is

$$J(\Delta t) = \Delta t^2. \tag{9.18}$$

It is subject to a terminal inequality constraint to reduce the ratio of the terminal cancer viability to the untreated case up to a desired ratio, i.e.,

$$\frac{p(t = 48\,\mathrm{h};\, \Delta t)}{p(t = 48\,\mathrm{h};\, \Delta t = 0)} \leq r_d, \tag{9.19}$$

where $r_d \in \mathbb{R}$ is the desired ratio of the cancer cell viability.

Once the value of Δt is given, the above objective function and the inequality constraint can be evaluated by integrating the dynamic model (9.14). As such, the presented optimization can be addressed by any numerical parameter optimization tool. Figure 9.10a illustrates the corresponding results, showing the treatment duration Δt required to reduce the relative cancer cell viability to the desired value r_d, where Δt increases as r_d decreases.

These provide a CAP treatment schedule for a specific level of cancer cell growth inhibition. However, cancer cell response to CAP treatments depends on various intrinsic and extrinsic factors, and the presented mathematical model may not accurately characterize the actual response of the cancer cells under treatments. This may cause that the terminal value of the relative cancer cell viability becomes greater than the desired level, or it may yield unnecessarily intensive CAP treatments.

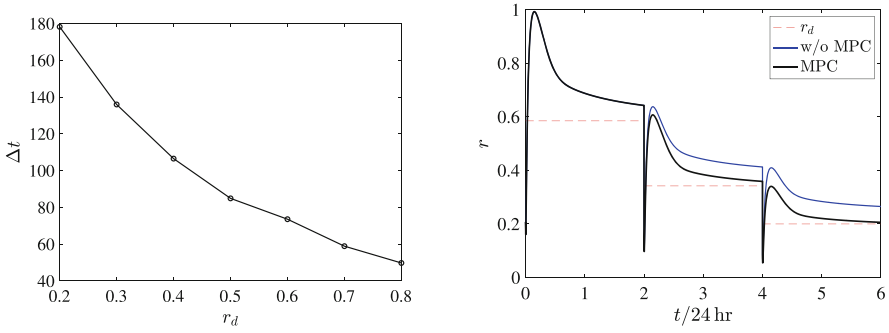

Fig. 9.10 Simulation results of model predictive control. (**a**) Optimal control to determine the treatment duration in seconds for a given desired ratio. (**b**) Model predictive control

We address this by optimal feedback framework based on model predictive controls (MPC). The objective is to adjust the treatment parameters adaptively based on the actual cell response.

A specific case is considered as below. We assume that a series of three CAP treatments are conducted at the interval of 48 h such that the terminal ratio at the end of 144 h reduces to 0.2. If there is no modeling error, the ratio of the cell viability can be reduced by $(0.2)^{1/3}$ as each treatment. Instead, in the presented MPC, the desired relative cell viability is chosen as

$$r_d = \begin{cases} 0.2^{1/3} & \text{first treatment,} \\ 0.2^{1/3} \times \dfrac{\text{(predicted ratio of cell viability)}}{\text{(actual ratio of cell viability)}} & \text{remaining treatments.} \end{cases} \tag{9.20}$$

The first treatment is scheduled based on the mathematical model. At the end of the first treatment period, the predicted cell viability ratio is compared against the actual value. For the next treatment, the treatment objective is adjusted as in (9.20) to compensate the corresponding discrepancy. For example, this reduces the desired viability of the next treatment further if the actual cell viability at the end of the preceding treatment is greater than its predicted value.

The proposed approach is verified by a numerical simulation, where the preceding mathematical model is considered as the actual cancer response, and the parameters of the mathematical model are altered to represent a mathematical model available to MPC. Therefore, the mathematical model available to the controller is different from the dynamic model representing the actual cancer cell response.

In Table 9.1, the three columns from the second to the fourth correspond to the ideal case when the exact model is available to MPC. In this case, the relative cancer viability reduces exactly by the desired factor r_d at each treatment. The treatment duration for all four treatments is identical to $\Delta t = 75.64$, and the terminal relative viability after four treatments is 0.2 as desired. The next three columns are the results of MPC when the exact model is not available to MPC. The first treatment duration is $\Delta t = 67.5$ for the first treatment, and it is less than the ideal case due to the modeling error. Consequently, the terminal viability ratio 0.64 becomes greater than the desirable value of 0.58. To compensate this, the goal of the second treatment is reduced to 0.47, which results in the viability ratio 0.36 that is slightly greater than the ideal value of 0.34. The final treatment is adjusted similarly so that the actual viability ratio at the end of three treatments achieves the treatment goal of 0.2, just as the ideal case.

Table 9.1 Modeling predictive control of U87

	Ideal case			MPC		
Treatment	r_d	Δt	Viability ratio	r_d	Δt	Viability ratio
1	0.58	75.64	0.58	0.58	67.5	0.64
2	0.58	75.64	0.34	0.47	78.8	0.36
3	0.58	75.64	0.20	0.50	77.0	0.20

Figure 9.10b illustrates the temporal response of the relative cell viability. The red, dotted lines show the desired relative cell viability at the end of each period. The blue lines are the results of optimization without MPC. Due to the modeling error, the actual cancer cell viability is greater than the desired value. Finally, the black lines correspond to MPC, where the treatment goal is achieved at the end of the three periods. These simulation results suggest that by adjusting CAP treatment conditions adaptive to the actual cancer response, the adverse effects of modeling errors can be mitigated.

9.3.2 Adaptive Learning Control

In the above model predictive control, comparing the actual viability at the end treatment period with the predicted value constitutes the feedback mechanism. As such, the evolution of the actual viability after a CAP treatment is not accounted. Furthermore, the mathematical model itself remains unchanged even after the discrepancy is observed, as the only part that is adjusted is the treatment global of optimization.

It would be more desirable if the adaptive plasma system *learns* the dynamic characteristics of the particular cancer cell under treatment, so that the mathematical model can be refined as the treatment is repeated. This provides more accurate mathematical model that can be used to adjust the prospective treatment plan accordingly.

The conventional adaptive controls deal with unknown parameters in the equations of motion, and as such it is not suitable for the cancer cell dynamics whose uncertainties cannot be represented in a structured form with parametric uncertainties. Next, it is also critical to evaluate the degree of confidence in the learned model, as we cannot rely on untrustworthy information in planning cancer treatment. In other words, the learned model is useful only if we are confident about its validity.

To address these, we propose to utilize Bayesian machine learning, which is a broad field in artificial intelligence to account uncertainties in data-based learning [29]. This is useful as an arbitrary, non-structured model and can be represented and learned in a probabilistic formulation that considers uncertainty distribution explicitly, thereby resolving the aforementioned issues of the conventional adaptive control or deterministic learning.

Gaussian Process

In particular, here we adopt the Gaussian process [30] to represent uncertainties in the mathematical model. A Gaussian process is a stochastic process, defined such that any finite number of collection is jointly Gaussian. It is completely described by second-order statistics as follows. Define a mean function $\mathbf{m}(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ and

a positive-definite covariance function $\mathbf{K}(x, x') : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, which is referred to as a kernel function. The corresponding Gaussian process is denoted by

$$g(x) \sim \mathcal{G}(\mathbf{m}(x), \mathbf{K}(x, x')). \tag{9.21}$$

Let $\mathcal{D} = \{(x_i, g_i, \sigma_{g_i})\}_{i \in 1,\dots N}$ be a set of data, where $g_i \in \mathbb{R}$ is a sample value of $g(x)$ when $x = x_i$, after corrupted by an additive, independent noise. More explicitly,

$$g_i \sim g(x_i) + \epsilon_{g_i}, \tag{9.22}$$

with $\epsilon_{g_i} \sim \mathcal{N}(0, \sigma_{g_i}^2)$.

Define $\mathbf{g}$, $\mathbf{x}$, and $\mathbf{m}(\mathbf{x}) \in \mathbb{R}^N$ be the concatenation of g_i, x_i, and $\mathbf{m}(x_i)$ for $i \in \{1, \dots, N\}$, respectively. Also, let the matrix $\mathbf{K}(\mathbf{x}, \mathbf{x}) \in \mathbb{R}^{N \times N}$ be defied such that its i, j-th element is $\mathbf{K}(x_i, x_j)$, and let $\Sigma_\mathbf{g} = \text{diag}[\sigma_{g_1}^2, \dots, \sigma_{g_N}^2] \in \mathbb{R}^{N \times N}$. From the definition of the Gaussian process, we have

$$\mathbf{g} \sim \mathcal{N}(\mathbf{m}(\mathbf{g}), \mathbf{K}(\mathbf{x}, \mathbf{x}) + \Sigma_\mathbf{g}). \tag{9.23}$$

Let $g_* \in \mathbb{R}$ be a sample value when $x = x_*$. It is jointly Gaussian with $\mathbf{g}$ as

$$\begin{bmatrix} \mathbf{g} \\ g_* \end{bmatrix} = \mathcal{N}\left(\begin{bmatrix} \mathbf{m}(\mathbf{x}) \\ \mathbf{m}(x_*) \end{bmatrix}, \begin{bmatrix} \mathbf{K}(\mathbf{x}, \mathbf{x}) + \Sigma_\mathbf{g} & \mathbf{K}(\mathbf{x}, x_*) \\ \mathbf{K}(x_*, \mathbf{x}) & \mathbf{K}(x_*, x_*) \end{bmatrix} \right). \tag{9.24}$$

Therefore, from the conditional distribution of joint Gaussian distributions, the regression equation for g_* is

$$\begin{aligned} g_* | \mathcal{D}, x_* \sim \mathcal{N}(&\mathbf{m}_* + \mathbf{K}_{*\mathbf{x}}(\mathbf{K}_{\mathbf{xx}} + \Sigma_\mathbf{g})^{-1}(\mathbf{g} - \mathbf{m}_\mathbf{x}), \\ &\mathbf{K}_{**} - \mathbf{K}_{*\mathbf{x}}(\mathbf{K}_{\mathbf{xx}} + \Sigma_\mathbf{g})^{-1}\mathbf{K}_{\mathbf{x}*}), \end{aligned} \tag{9.25}$$

where the subscripts for $\mathbf{m}$ and $\mathbf{K}$ denote the input arguments, e.g., $\mathbf{K}_{*\mathbf{x}} = \mathbf{K}(x_*, \mathbf{x}) \in \mathbb{R}^{1 \times N}$.

The desirable feature is that a Gaussian process may represent an arbitrary function explicitly as in (9.25), without need for training or numerical optimization required for common multi-layer neural networks. The uncertainties are represented by Gaussian distributions that are provided by various properties, which can be utilized to simplify the required mathematical analysis.

Adaptive Learning Control with Gaussian Process

Consider the empirical model (9.14) perturbed by the unknown disturbance or modeling error represented by $\Delta(t, p) \in \mathbb{R}^2 \times \mathbb{R}$:

Table 9.2 Adaptive learning control of U87

	Ideal case			Adaptive learning control		
Treatment	r_d	Δt	Viability ratio	r_d	Δt	Viability ratio
1	0.44	90	0.44	0.44	75.9	0.58
2	0.44	90	0.20	0.37	127.6	0.20

$$\dot{p} = pF(t, p) + \Delta(t, p). \tag{9.26}$$

To simplify the following discussion, suppose the above continuous-time differential equation is discretize over a time sequence $\{t_0, t_1, \ldots, t_N\}$ into

$$p_{k+1} = p_k F_d(t_k, p_k) + \Delta_d(t_k, p_k), \tag{9.27}$$

for $F_d, \Delta_d : \mathbb{R}^2 \times \mathbb{R}$ and the subscript k denotes the value of a variable at t_k. Here the first term on the right-hand side $p_k F_d(t_k, p_k)$ corresponds to the mathematical model, and the second term $\Delta_d(t_k, p_k)$ denotes the unknown modeling error or disturbance that may be dependent of the current viability and time. This can be generalized to incorporate other intrinsic and extrinsic factors.

Whenever the viability is measured, the above equation yields a sample data $\{(t_k, p_k), \Delta_d(t_k, p_k)\}$ to be used to represent the unknown disturbance with a Gaussian process described above. The desirable property is that as the treatment is repeated more data become available so that the Gaussian process models the unknown part more accurately, thereby executing the learning process. Once the model is updated, any control strategy can be applied.

The proposed adaptive learning control is applied to the CAP treatment problem formulated in Sect. 9.3.1. The objective is to reduce the relative cancer viability into 0.2. But instead of three consecutive treatment, here we consider two treatments over the interval of 96 h. The first treatment is scheduled based on the optimization with the empirical model. After the first treatment, the cancer viability is measured five times at $t \in \{0.2, 0.4, \ldots, 1\}$ h to produce a set of sample data for a Gaussian process. The second treatment at $t = 48$ is planned based on the learned dynamic model.

The corresponding simulation results are summarized in Table 9.2. The two columns from the second to the fourth are for the ideal case when $\Delta = 0$ without adaptive learning. The desired ratio of each treatment is $r_d = 0.2^{1/2} = 0.44$, which is achieved exactly with $\Delta t = 90$ s. The resulting terminal viability ratio reduces to the desired value exactly. The next three columns are for the proposed adaptive learning control in the presence of non-zero modeling error Δ. Due to the modeling error, the first treatment time $\Delta t = 75.9$ is less than the ideal value of 90. However, the cancer cell response to the first treatment is monitored, and the discrepancy is accounted by the Gaussian process. Based on the learned dynamic model, the second treatment is planned. As the error is properly compensated, the second treatment from the learned model achieves the desired viability ratio exactly.

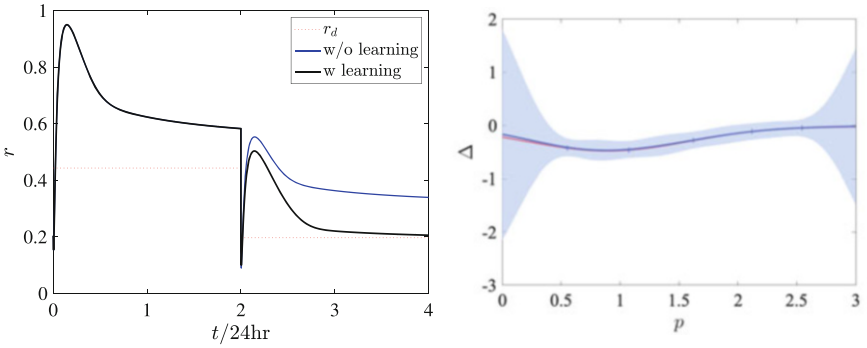

Fig. 9.11 Simulation results of adaptive learning control. (**a**) Cell viability ratio. (**b**) Gaussian process learning for modeling error

Figure 9.11a illustrates the evolution of the viability ratio, where the red dotted lines are for the desired treatment goal. The blue lines are for treatments without the Gaussian process learning, and it exhibits a substantial error: the terminal viability ratio is 0.34, which exceeds the desired value by 70%. Finally, the black lines are for the proposed adaptive learning control, which achieves the treatment goal at the end of the second treatment. The next Fig. 9.11b illustrates the modeling error learned by the Gaussian process, where the true values are represented by the red curve, and the learned model is given by the blue curves with 3σ bounds represented by a shaded area, and the training data denoted by the marks $+$. The learned model accurately represents the true unknown modeling error. More importantly, the 3σ bounds show that there is less uncertainty over the interval covered by the training data, and outside of the interval is greater uncertainty. As such, it provides the level of confidence in the learned model depending on the given domain.

This information can be utilized in the trade between performance and safety: over the region of smaller uncertainties, we can plan an aggressive plan with confidence; for the region of greater uncertainties, the treatment can be scheduled more conservatively. Accounting uncertainties and gauging confidence are one of the important attributes of the proposed adaptive learning control for adaptive plasma. While this section relies on the Gaussian process, any multi-layer neural network can be utilized to represent the modeling uncertainties. With Bayesian machine learning, it is more suitable to represent more complicated modeling errors and uncertainties depending on various factors.

9.3.3 Reinforcement Learning

Reinforcement learning (RL) deals with optimal strategies for an agent in an environment to take the action to maximize a notion of cumulative reward [18].

The interaction between the agent and the environment is typically modeled by a Markov decision process. As such it is closely related to the dynamic programming in optimal control. The unique feature of RL is that it may not assume the complete knowledge of the dynamics, and the optimal policy does not have to be completely determined in prior. Instead, the optimal policy that may be initialed randomly is revised through each trial after evaluating the possible actions from experience. Therefore, the theory and the practice of RL can be utilized in adaptive plasma, to adjust the treatment plan in an optimal fashion to suppress the growth of the particular cancer cell under treatments, without need for the complete knowledge of the mechanism behind it. However, the successful implementation of RL often requires a large number of trials, which leads to a challenge in cancer treatments.

This section provides a formulation of RL for adaptive plasma. It is based on a partially observable Markov decision process (POMDP) defined by the following components:

- State s_k: It represents the current status of the cancer cell. For example, for the empirical model in Sect. 9.2.2, the cancer viability corresponds to the state.
- Action a_k: The parameters of CAP that can be adjusted arbitrarily are called action. This may include the treatment duration, plasma discharge voltage, or gas composition.
- Transition probability $P(s_{k+1}|s_k, a_k)$: According to the assumption of the Markov process, the future state is completely determined by the current state and the action taken, without need for the prior history. The transition probability describes the distribution of the state in the next time step, for the given current state and action.
- Reward $R(s_k, a_k)$: This represents the reward by choosing the action a_k at the given state s_k. This can be designed to reflect the objectives of the treatment. For example, a reward can be the amount of the reduction of cancer cells.
- Observation o_k: The observation corresponds to the characteristics of the cancer that can be measured by a sensor. For example, the level of metabolism is measured in [26]. Recently, the impedance is utilized for real-time diagnostics.
- Observation probability $O(o_k|s_k, a_k)$: this characterizes the type of sensor used. It represents the distribution of the measurement for a given state and action.

In short, an agent (adaptive plasma) in an environment with the state s_k (cancer) chooses an action a_k (CAP), which causes the environment to transfer the state according to the transition probability (cancer dynamics). At the same time, the agent receives a reward (cancer treatment) and an observation (measurement) from the environment. The objective of the agent is to choose an action at each time step that maximizes its expected cumulative reward, or *return* defined as

$$G_k = \mathrm{E}\left[\sum_{l=0}^{\infty} \gamma^l R(s_{k+l+1}, a_{k+l+1})\right], \tag{9.28}$$

where $0 < \gamma < 1$ is a discount factor. The strategy of the agent is called *policy* $\pi(a|s)$, which describes the distribution of the action for a given state at each step. As the state is not directly measured, the belief of the state is determined by its history of actions and observations. For the belief of the current state $b(s_k)$, the next one is updated by

$$b(s_{k+1}) \propto O(o_k|s_k, a_k) \sum_{s_k} P(s_{k+1}|s_k, a_k) b(s_k).$$

The problem of constructing the optimal policy maximizing (9.28) can be addressed by the dynamic programming in optimal control. However, it is subjected to complexities, referred to as *curse of dimensionality*, especially when the system is of higher dimensions, and it requires the complete knowledge of the dynamics.

The heart of RL is avoiding these issues by iteratively updating the policy as described below. For a given policy π, let $Q^\pi(s_k, a_k)$ be the expected return when an arbitrary action a_k chosen as the k-th step, and the prospective actions at the $k+1$-th step and afterwards are chosen from the policy π, i.e.,

$$Q^\pi(s, a) = \mathrm{E}_\pi[G_k|s_k = s, a_k = a]. \tag{9.29}$$

Once the Q function is computed, the policy can be improved by

$$\pi' = \arg\max\, Q^\pi(s, a), \tag{9.30}$$

which is a *greedy* policy improvement that seeks for the best possible alternative choice from π always. Instead of seeking the optimal action at every step greedily, the above policy improvement can be relaxed to explore other possibilities.

However, constructing the Q function from (9.29) is not practical due to the same issues of the dynamic programming: this may take nontrivial computational efforts and it requires the complete knowledge of the dynamics. RL circumvents these by revising the Q function continuously. First, the Q function is initialized by some random values. Let the action a be chosen from the state s so that the state is transferred to s', while resulting in the reward r. From this experience, the best guess of the correct value of $Q(s, a)$ is $r + \gamma \max_{a'} Q(s', a')$, which is referred to as the *TD target*. Then the Q function is updated from the current value toward the TD target according to

$$Q(s, a) = Q(s, a) + \alpha\{\underbrace{r + \gamma \max_{a'} Q(s', a')}_{\text{TD target}} - Q(s, a)\},$$

where $\alpha > 0$ corresponds to a learning rate. Once the Q function is updated, the policy can be improved according to (9.30). These process for improving the Q function and the policy is repeated at each time step. It is a learning process as both are improved by an experience represented by the set (s, a, s', r).

To implement this, the Q function can be defined as a lookup-table when there a limited number of discrete states and actions. For continuous state and action, it can be represented by any function approximator. In particular, when the dimension of state and action is large, a deep neural network can be adopted for the Q function, resulting in the so-called deep Q-learning [31].

As discussed above, the desirable feature is that it does not require the complete prior knowledge of the cancer dynamics, and the optimal policy representing adaptive plasma is improved as the set of experiences accumulates. The challenge is that it may take a lot of trials until the Q function converges. Also to ensure convergence to the global optimum and to facilitate the process, the agent may need to take an unreasonable action through the course. This might be problematic for cancer treatment, where it is infeasible to repeat numerous CAP treatments, and the treatment should remain within a reasonable bound. This might be mitigated by pre-training the Q function based on the mathematical models formulated in Sect. 9.2. Even with the potential challenges, it is expected that an innovative advance in adaptive plasma can be achieved by utilizing the theory and the practice of RL.

9.4 Conclusions

This chapter has presented mathematical models and control strategies for adaptive plasma.

Mathematical Modeling
The mathematical models can be used to predict the cancer cell response to CAP for a given treatment conditions, namely the treatment duration and the plasma discharge voltage. The first oxidative DNA stress model is constructed to mathematically formulate the effects of the CAP on the underlying cell cycle. This represents one of the current understandings for the mechanism how CAP eliminates cancer cells, and as such, there is a great potential that this model is further generalized and revised to account various treatment conditions and exogenous factors to reflect cancer cell dynamics accurately.

However, identifying the parameters would require a set of experiments to monitor the evolution of cell cycle population.

On the other hand, the next empirical model is solely based on the observation of cancer cell response, after treating the cancer cell dynamics as a completely unknown system. As such, it does not reflect our comprehension behind the CAP cancer treatment. However, it is relatively simple, and it can be adjusted to make a reasonable prediction consistent with the experimental data.

Adaptive Plasma
Next, three control strategies, namely model predictive control, adaptive learning control, and reinforcement learning, are presented. Model predictive control is aligned with the traditional control engineering, and it can ensure optimality under the feedback controls. However, the feedback mechanism is focused on adjusting

the desired goal of optimization, and there are no adjustments in the mathematical model and the control algorithm.

In contrast, the adaptive learning control attempts to utilize all of the information available from the experience to refine the model and the control. Furthermore, the proposed stochastic framework evaluates the level of uncertainties throughout the information fusion. As such, the objective of control can be strategically adjusted between higher performance in the region of lower uncertainties and safety in the area of higher uncertainties.

Finally, reinforcement learning can be utilized for the complete model free control. However, a successful implementation of reinforcement learning may require numerous experiments, which can be mitigated by integrating with a mathematical model. As the field of artificial intelligence is rapidly advancing, there is a great potential for reinforcement learning utilized in innovative adaptive plasma.

Remarks

All of these mathematical models and control strategy would greatly benefit from real-time in situ diagnostics. As it is nearly impossible to characterize the cancer cell response to CAP from fundamental principles, any of the mathematical model should rely on the experimental data, and the accuracy of such models is limited by the richness and the quality of data. However, it is often that a cell viability and proliferation assay or flow cytometry need to destroy the cell to detect and measure its chemical and physical properties. As such, to measure the time evolution of cancer response to CAP over multiple instances, a set of experiments should be performed in parallel under the identical condition. Any real-time diagnostics can be utilized to generate a variety of valuable data to be adopted for more reliable mathematical models.

Furthermore, the critical component of adaptive plasma is monitoring the cellular response in real-time so that the treatment conditions are adjusted accordingly. The information that can be acquired from real-time diagnostics will play a critical role in the success of adaptive plasma.

Appendix

Step Function

Here, we construct a smooth step function $\rho(x; x_0, x_1) : \mathbb{R} \to [0, 1]$ such that

$$\rho(x; x_0, x_1) = \begin{cases} 0 & x \leq x_0, \\ 1 & x \geq x_1, \end{cases} \tag{9.31}$$

for $x_0 < x_1$.

We introduce a C^∞ function,

$$f(x) = \begin{cases} e^{-1/t} & t > 0, \\ 0 & t \leq 0. \end{cases}$$

Consider

$$g(x) = \frac{f(x)}{f(x) + f(1-x)},$$

which is a smooth step function from 0 for $x \leq 0$ to 1 for $x \geq 1$. Utilizing this, we can define a function satisfying (9.31) as

$$\rho(x; x_0, x_1) = g\left(\frac{x - x_0}{x_1 - x_0}\right). \tag{9.32}$$

References

1. G. Fridman, D. Dobrynin, G. Friedman, A. Fridman, Physical and biological mechanisms of plasma interaction with living tissue, in *2008 IEEE 35th International Conference on Plasma Science* (IEEE, Piscataway, 2008), pp. 1–1
2. M. Keidar, I. Beilis, *Plasma Engineering: Applications from Aerospace to Bio and Nanotechnology* (Academic, Cambridge, 2013)
3. N. Knake, K. Niemi, S. Reuter, V. Schulz-von der Gathen, J. Winter, Absolute atomic oxygen density profiles in the discharge core of a microscale atmospheric pressure plasma jet. Appl. Phys. Lett. **93**(13), 131503 (2008)
4. J. Sousa, K. Niemi, L. Cox, Q.T. Algwari, T. Gans, D. O'connell, Cold atmospheric pressure plasma jets as sources of singlet delta oxygen for biomedical applications. J. Appl. Phys. **109**(12), 123302 (2011)
5. G.E. Conway, A. Casey, V. Milosavljevic, Y. Liu, O. Howe, P.J. Cullen, J.F. Curtin, Non-thermal atmospheric plasma induces ROS-independent cell death in U373MG glioma cells and augments the cytotoxicity of temozolomide. Br. J. Cancer **114**(4), 435 (2016)
6. M. Vandamme, E. Robert, S. Lerondel, V. Sarron, D. Ries, S. Dozias, J. Sobilo, D. Gosset, C. Kieda, B. Legrain, et al., ROS implication in a new antitumor strategy based on non-thermal plasma. Int. J. Cancer **130**(9), 2185–2194 (2012)
7. J. Schlegel, J. Köritzer, V. Boxhammer, Plasma in cancer treatment. Clin. Plasma Med. **1**(2), 2–7 (2013)
8. M. Keidar, R. Walk, A. Shashurin, P. Srinivasan, A. Sandler, S. Dasgupta, R. Ravi, R. Guerrero-Preston, B. Trink, Cold plasma selectivity and the possibility of a paradigm shift in cancer therapy. Br. J. Cancer **105**(9), 1295–1301 (2011)
9. J.A. Cook, D. Gius, D.A. Wink, M.C. Krishna, A. Russo, J.B. Mitchell, Oxidative stress, redox, and the tumor microenvironment, in *Seminars in Radiation Oncology*, vol. 14, no. 3 (Elsevier, Amsterdam, 2004), pp. 259–266
10. P.T. Schumacker, Reactive oxygen species in cancer cells: live by the sword, die by the sword. Cancer Cell **10**(3), 175–176 (2006)
11. M. Keidar, A prospectus on innovations in the plasma treatment of cancer. Phys. Plasmas **25**(8), 083504 (2018)

12. J.W. Bradley, J.-S. Oh, O.T. Olabanji, C. Hale, R. Mariani, K. Kontis, Schlieren photography of the outflow from a plasma jet. IEEE Trans. Plasma Sci. **39**(11), 2312–2313 (2011)
13. T. Darny, J.-M. Pouvesle, J. Fontane, L. Joly, S. Dozias, E. Robert, Plasma action on helium flow in cold atmospheric pressure plasma jet experiments. Plasma Sources Sci. Technol. **26**(10), 105001 (2017)
14. B. Klarenaar, O. Guaitella, R. Engeln, A. Sobota, How dielectric, metallic and liquid targets influence the evolution of electron properties in a pulsed he jet measured by Thomson and Raman scattering. Plasma Sources Sci. Technol. **27**(8), 085004 (2018)
15. N. Georgescu, A.R. Lupu, Tumoral and normal cells treatment with high-voltage pulsed cold atmospheric plasma jets. IEEE Trans. Plasma Sci. **38**(8), 1949–1955 (2010)
16. M. Keidar, Therapeutic approaches based on plasmas and nanoparticles. J. Nanomed. Res **3**, 3–5 (2016)
17. M. Keidar, D. Yan, I.I. Beilis, B. Trink, J.H. Sherman, Plasmas for treating cancer: opportunities for adaptive and self-adaptive approaches. Trends Biotechnol. **36**(6), 586–593 (2018)
18. R.S. Sutton, A.G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, 2018)
19. D. Gidon, B. Curtis, J.A. Paulson, D.B. Graves, A. Mesbah, Model-based feedback control of a kHz-excited atmospheric pressure plasma jet. IEEE Trans. Radiat. Plasma Med. Sci. **2**(2), 129–137 (2018)
20. D. Gidon, D.B. Graves, A. Mesbah, Effective dose delivery in atmospheric pressure plasma jets for plasma medicine: a model predictive control approach. Plasma Sources Sci. Technol. **26**(8), 085005 (2017)
21. A. Mesbah, D.B. Graves, Machine learning for modeling, diagnostics, and control of non-equilibrium plasmas. J. Phys. D Appl. Phys. **52**(30), 30LT02 (2019)
22. D.O. Morgan, *The Cell Cycle: Principles of Control* (New Science Press, London, 2007)
23. O. Volotskova, T.S. Hawley, M.A. Stepp, M. Keidar, Targeting the cancer cell cycle by cold atmospheric plasma. Sci. Rep. **2**, 636 (2012)
24. X. Yan, F. Zou, S. Zhao, X. Lu, G. He, Z. Xiong, Q. Xiong, Q. Zhao, P. Deng, J. Huang, et al., On the mechanism of plasma inducing cell apoptosis. IEEE Trans. Plasma Sci. **38**(9), 2451–2457 (2010)
25. Y. Lyu, L. Lin, E. Gjika, T. Lee, M. Keidar, Mathematical modeling and control for cancer treatment with cold atmospheric plasma jet. J. Phys. D Appl. Phys. **52**(18), 185202 (2019)
26. E. Gjika, S. Pal-Ghosh, A. Tang, M. Kirschner, G. Tadvalkar, J. Canady, M.A. Stepp, M. Keidar, Adaptation of operational parameters of cold atmospheric plasma for in vitro treatment of cancer cells. ACS Appl. Mater. Interfaces **10**(11), 9269–9279 (2018)
27. H. Bryson, *Applied Optimal Control: Optimization, Estimation, and Control* (Taylor & Francis, Abingdon-on-Thames, 1975)
28. W. Kwon, S. Han, *Receding Horizon Control: Model Predictive Control for State Models* (Springer, Berlin, 2005)
29. D. Barber, *Bayesian Reasoning and Machine Learning* (Cambridge University Press, Cambridge, 2012)
30. C. Rasmussen, C. Williams, *Gaussian Process for Machine Learning* (MIT Press, Cambridge, 2006)
31. M. Hausknecht, P. Stone, Deep recurrent Q-learning for partially observable MDPS, in *2015 AAAI Fall Symposium Series* (2015)