Chapter 5

Drug Dosage Design

ANTIRETROVIRAL therapy generally entails the application of drugs in a fixed dosage regimen. A typical initial regimen would consist of 2 reverse transcriptase inhibitors - nucleoside/nucleotide (NRTI/NtRTI) as the basis of the regimen, plus either a protease inhibitor (PI) or a non-nucleoside reverse transcriptase inhibitor (NNRTI) [1]. But generally, selection of an initial regimen is subject to considerations such as efficacy, toxicity, resistance profile and durability. Other issues to address, besides what regimen to start with, are when to change the initial regimen; and when deciding to change, what should the regimen be change to?

From a control theoretic point of view, HIV chemotherapy is control of a time varying nonlinear dynamical system with constrained controls. Once a suitable model has been developed or identified, control system theoretical concepts and design principles can be applied. The adopted control approach or strategy depends primarily on the control objectives, performance specifications and the control constraints. In principle, the designed control system can then be validated with clinical data.

This chapter focuses on the rational sequencing of antiretroviral drugs in order to maximize the benefits of therapy, when replication cycle based HAART (reverse transcriptase inhibitors - RTIs, protease inhibitors - PIs) and immune based therapies - IBT (Hydroxyurea - HU, Interleuken 2 - IL2) are available and therapy is initiated during the asymptomatic stage of the infection. This will be done taking into account the constraints on the admissible controls imposed by the narrow therapeutic range of the antiretroviral agents, as well as the possible emergence of drug resistant mutants with inadequate viral load suppression. The practicality of applying the derived dosage schemes to a patient will be discussed.

5.1 The Dynamical System to be Controlled

Equations (5.1) represent the HIV/AIDS system to be controlled. This system has been described in detail in the preceding chapters, with two notable exceptions - the parameter u_H, which takes the binary value 1 when HAART is ON, and 0 when HAART is OFF, and the parameter u_I, which takes the binary value 1 when IBT is ON, and 0 when IBT is OFF. The $CD4^+$ T cell and macrophage counts prior to infection (pre-infection) as well as the system states at the initiation of therapy (pre-treatment/pre-HAART measurements) are presented in Table 5.2. Model parameters are re-presented in Table 5.1 for ease of reference.

$$\Sigma_{Eco} = \begin{cases} \frac{dT}{dt} &= s_T + (1-u_I\eta_{ps})pT(1-\frac{T}{T_m}) - d_T T - \beta_T T V_i \\ \frac{dT_l}{dt} &= (1-u_H\eta_{rt})q_l\beta_T T V_i - (1+u_I\eta_{da})\delta_l T_l - kT_l \\ \frac{dT_a}{dt} &= (1-u_H\eta_{rt})q_a\beta_T T V_i - (1+u_I\eta_{da})\delta_a T_a + kT_l \\ \frac{dM}{dt} &= s_M - d_M M - \beta_M M V_i \\ \frac{dM^*}{dt} &= (1-u_H\alpha_{rt}\eta_{rt})\beta_M M V_i - \mu M^* \\ \frac{dV_i}{dt} &= (1-u_H\eta_{pi})r_T T_a + (1-u_H\alpha_{pi}\eta_{pi})r_M M^* - cV_i \\ \frac{dV_n}{dt} &= u_H\eta_{pi}r_T T_a + u_H\alpha_{pi}\eta_{pi}r_M M^* - cV_n \end{cases} \tag{5.1}$$

Table 5.1: Parameter estimates

Parameter	Value	Parameter	Value
s_T	10^4 mL^{-1}day^{-1}	s_M	150 mL^{-1}day^{-1}
d_T	0.01 day^{-1}	d_M	0.005 day^{-1}
β_T	4.5×10^{-8} mL day^{-1}	β_M	1.75×10^{-8} mL day^{-1}
p	0.02 day^{-1}	q_M	1
T_m	10^6 mL^{-1}	μ	0.05 day^{-1}
q_l	0.005	r_M	35 cell^{-1}day^{-1}
q_a	0.55	α_{rt}	0.85
δ_l	0.01 day^{-1}	α_{pi}	0.55
δ_a	0.5 day^{-1}		
k	0.025 day^{-1}		
r_T	240 cell^{-1}day^{-1}	c	5 day^{-1}

[46, 71, 50, 176]

u_H	0	Off HAART
	1	On HAART
u_I	0	Off IBT
	1	On IBT

Table 5.2: System states prior to infection and at the initiation of therapy

Variable	Value	Note
Pre-infection		
$T(0)$	10^6 mL^{-1}	CD4$^+$ T cell count
$M(0)$	3x10^4 mL^{-1}	Macrophage cell count
Pre-treatment/Pre-HAART		
$\bar{T}$	4.08x10^5 mL^{-1}	Uninfected CD4$^+$ T cell count
$\bar{T}_l$	1.54x10^3 mL^{-1}	Latently infected CD4$^+$ T cell count
$\bar{T}_a$	1.19x10^4 mL^{-1}	Actively infected CD4$^+$ T cell count
$\bar{M}$	9.85x10^3 mL^{-1}	Uninfected macrophage cell count
$\bar{M}^*$	2.02x10^3 mL^{-1}	Infected macrophage cell count
$\bar{V}_i$	5.94x10^5 mL^{-1}	Infectious viral load
$\bar{V}_n$	0	Non-infectious viral load

Figure 5.1 shows the control system block diagram where the patient's virus and host cell dynamics are modelled by (5.1). The model is known to be have some inaccuracies and there are some un-modelled dynamics. The model parameters are also inaccurate and are also thought to change as the infection progresses.

It is clear from the illustrations in figure 5.1 that there are many issues that need to be addressed or taken into consideration when designing a controller. There are disturbances, many sources of error and generally much uncertainty in the system. Furthermore, the objectives of therapy are multiple and conflicting.

In essence then, one has to design a drug regimen that optimizes a performance criterion, and also explores future dosing and sampling scenarios in advance. Besides optimizing performance, the intention should also be to optimize the process of learning about the patient. All this has to be done while having to treat the patient at the same time [180].

5.2 Modelling Antiretroviral Drugs As Control Inputs

Administering an antiretroviral drug is equivalent to introducing an input signal that perturbs the HIV dynamics. A point to note is that these drugs are administered periodically, and that the instantaneous inhibitory effect of the antiretroviral agent is a time varying function [43, 77, 97, 182]. Furthermore, HIV drug pharmacodynamics,

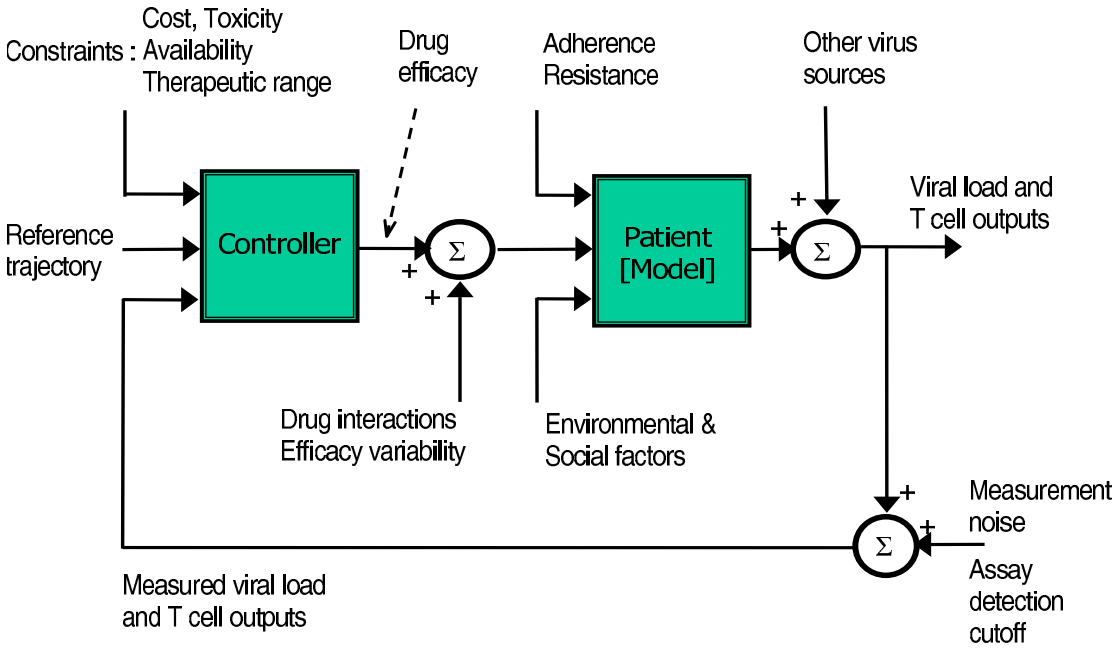


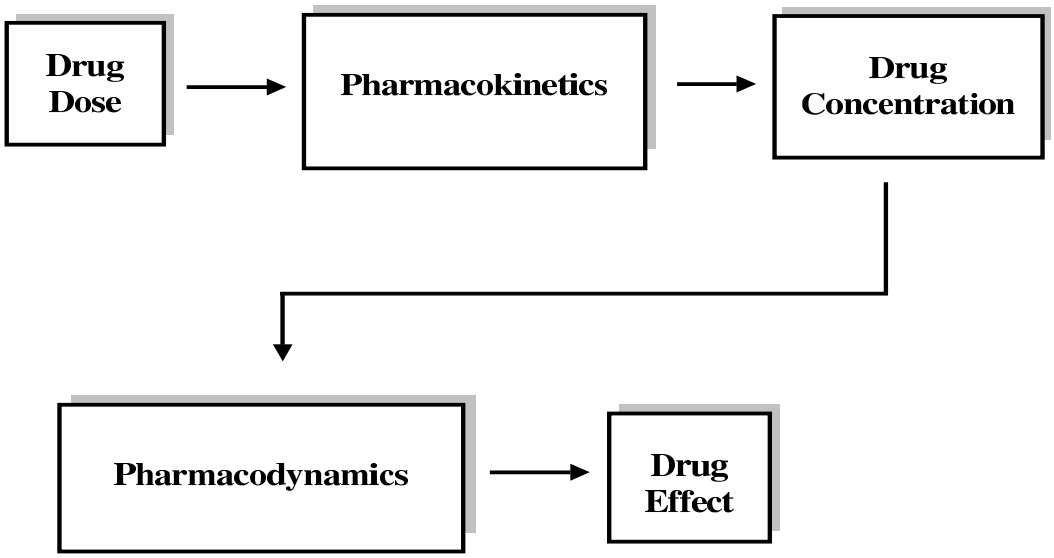
Figure 5.1: Control system block diagram.

pharmacokinetics and adverse reactions are genetically predisposed [41, 126].

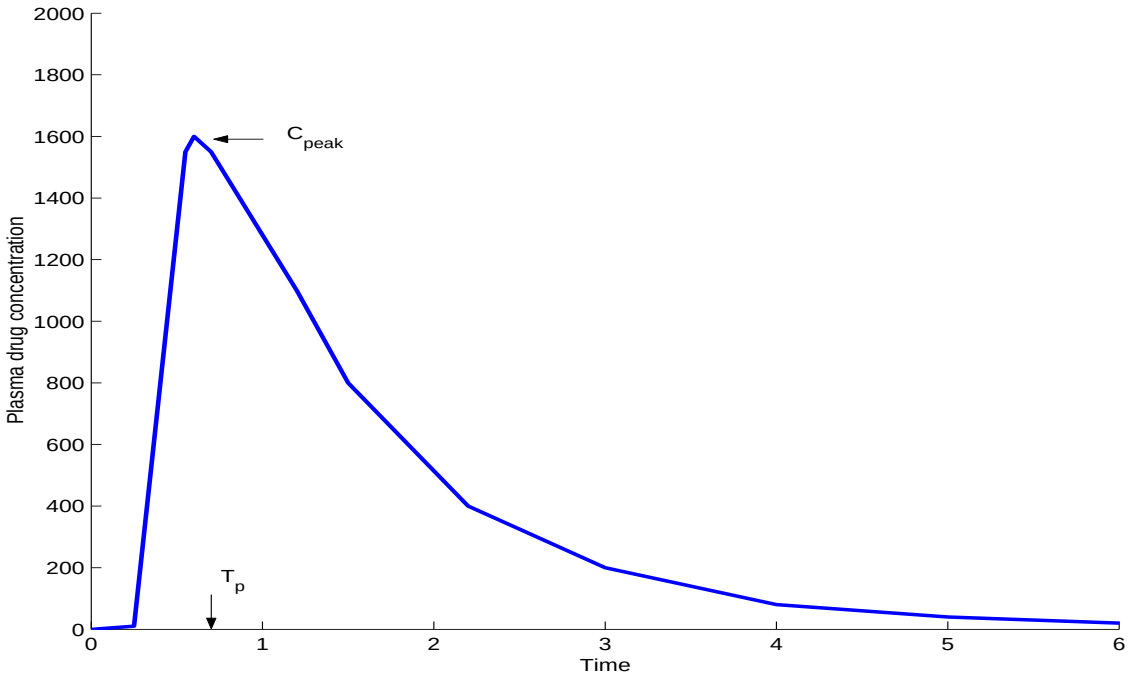
5.2.1 Drug Pharmacology

Drug pharmacokinetics deals with the relationship between dosage and plasma concentration of the drug, and involves the processes of absorption, distribution, metabolism and excretion of the drug. Pharmacodynamics on the other hand, deals with the relationship between the drug plasma concentration and the end point effect of the drug. For a particular drug dose, there will be inter individual variations in drug plasma concentrations. Similarly, for a particular plasma drug concentration, there will be inter individual variations in the end point effect of the drug.

The drug dosage to end effect process is illustrated in figure 5.2(a). From ingestion, the increase in plasma drug concentration is determined by the rate at which the drug enters the plasma by absorption and simultaneously removed from plasma by either distribution to other body compartments or by elimination. A peak concentration, as illustrated in figure 5.2(b), is reached when there is a balance between the entry and removal rates of the drug into and out of the plasma. Thereafter, the drug plasma concentration declines because the combined rate at which it is distributed and elimi-



(a)



(b)

Figure 5.2: (a) Drug dosage to end effect process [181]. (b) Single dose plasma drug concentration variation [177]. C_{peak} : peak drug concentration, T_p : time to reach peak concentration.

Table 5.3: Suggested minimum target trough concentrations for persons with wild-type HIV-1

Drug	**Concentration (ng/mL)**
Amprenavir (Agenerase)	400
Indinavir (Crixivan)	100
Lopinavir/ritonavir (Kaletra)	1000
Nelfinavir (Viracept)[a]	800
Ritonavir (Norvir)[b]	2100
Saquinavir (Fortovase, Invirase)	100−250
Efavirenz (Sustiva)	1000
Nevirapine (Viramune)	3400

Reproduced from [1].

nated exceeds the rate at which it is absorbed [177]. The expression for the plasma drug concentration for a single dose can be represented as

$$\text{conc}_0(t) = a_0(e^{-t/\tau_x} - e^{-t/\tau_a}) \tag{5.2}$$

where τ_x and τ_a, $\tau_x > \tau_a$ are the respective elimination and absorption time constants, and a_0 is drug and dosage dependent.

Figure 5.3(a) shows how the plasma drug concentration is expected to vary between multiple doses. This actually is the accumulation of all the single dosages taken at specified intervals and can be expressed as

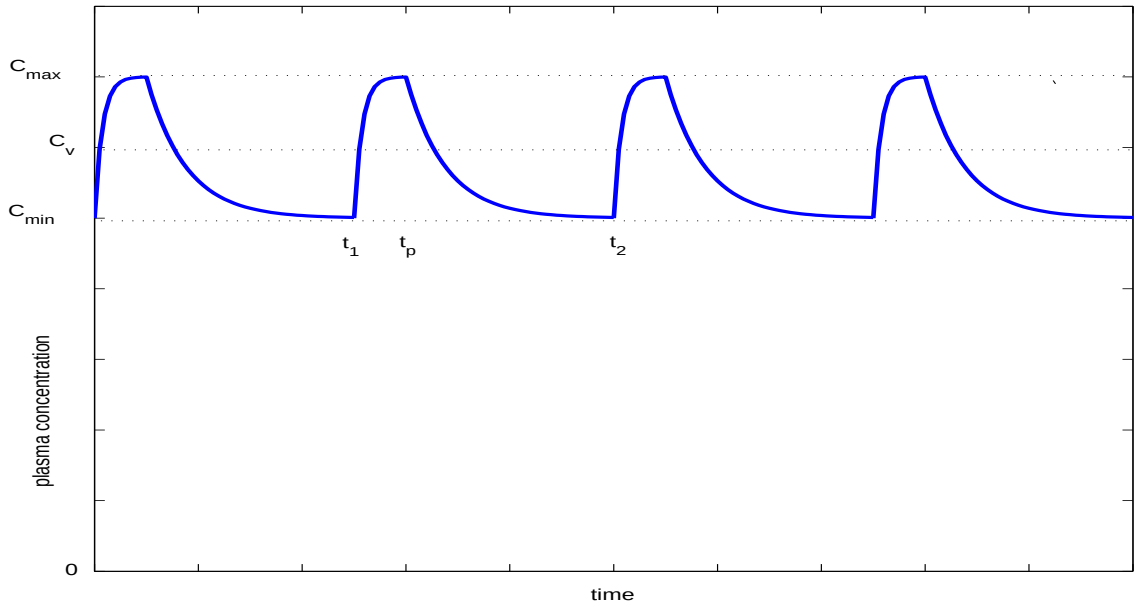
$$\text{conc}(t) = \sum_{i=0}^{n-1} a_i(e^{-(t-t_i)/\tau_x} - e^{-(t-t_i)/\tau_a}) \tag{5.3}$$

where t_i is the instance when the dosage is administered.
Table 5.3 presents a summary of the recommended minimum concentrations for some antiretroviral drugs [1].

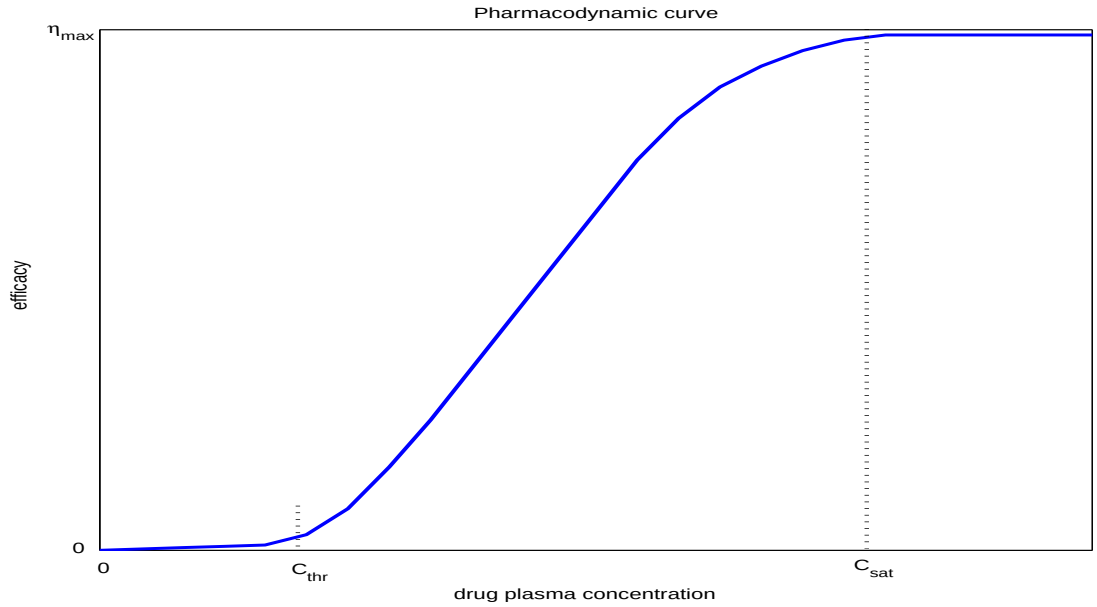
If one assumes that the plasma concentration of a drug at a particular point in time is an indication of the efficacy or instantaneous inhibition of virus replication by the drug, as illustrated in figure 5.3(b), then the variation of the drug's efficacy $\eta(t)$ over one dosing cycle can be approximated as

$$\eta(t) = \begin{cases} \eta_v + \eta_w(\frac{1}{2} - e^{-(t-t_1)/\tau_r}), & t_1 \leq t \leq t_p \\ \eta_v + \eta_w(e^{-(t-t_p)/\tau_d} - \frac{1}{2}), & t_p < t < t_2 \end{cases} \tag{5.4}$$

where η_w is the difference between the upper and lower efficacy bounds, η_v is the average efficacy value, while τ_r and τ_d are the rise and decay time constants after each dose.



(a)



(b)

Figure 5.3: (a) Plasma drug concentration variation with time for multiple doses. C_{max} : maximum concentration, C_v : average concentration, C_{min} : minimum concentration. (b) Drug pharmacodynamics: Variation of end point drug efficacy with drug plasma concentration. η_{max} : maximum attainable efficacy, C_{thr} : threshold drug concentration, C_{sat} : saturation drug concentration.

Refer to, for example, [95, 183] for the relationship between plasma drug concentration and efficacy, and to [173] for a summary of the clinical pharmacokinetic properties of some antiretroviral drugs. Techniques to estimate the instantaneous inhibitory effect from end point data are proposed in [182]. The drug efficacy is therefore an oscillatory function of time. If however, the interval between dosing, or the release mechanism of the drug is such that the variation between the peak drug concentration and the minimum concentration is minimal, then the average value a_v in equation (5.4) gives a good approximation of the drug's efficacy over time, and the efficacy can therefore be assumed to be constant.

The drug is usually effective until the time when resistance emerges. The effectiveness η can also be assumed to decline exponentially with time, and the expression for $\eta(t)$ over one cycle can be modified to

$$\eta(t) = \begin{cases} (\eta_v + \eta_w(\frac{1}{2} - e^{-(t-t_1)/\tau_r}))e^{-(t-t_i)/\tau_e}, & t_1 \leq t \leq t_p \\ (\eta_v + \eta_w(e^{-(t-t_p)/\tau_d} - \frac{1}{2}))e^{-(t-t_i)/\tau_e}, & t_p < t < t_2 \end{cases} \tag{5.5}$$

where t_i is the time when resistance emerges and the time constant τ_e of the drug's effectiveness due to resistance is drug dependent. Alternative descriptions for the drug dosage to end point effect are available in [77, 97, 182].

Besides the development of resistance to drugs, there are other factors that can affect the end point drug efficacy, such as non-adherence to a regimen and drug-drug interactions that arise when more than one drug is taken. The reader is referred to the guidelines [1] for allowable drug combinations and drug-drug interactions between some compounds.

5.2.2 Therapeutic Range

Antiretroviral agents are known to have, or there are indications that they have a narrow therapeutic range [184], where the therapeutic range for a drug is the dosage range within which most patients will experience significant therapeutic effect without an undesirable degree of adverse reactions. It is therefore desirable to administer drug doses that are within this therapeutic range. However, one needs to bear in mind that the therapeutic range for a particular drug is the average for many individuals on the therapeutic trial, and that it approximates single drug therapy when the dosage is as prescribed. The combined therapeutic effect when more than one drug is used to treat a single condition needs to be considered.

A point to note is that inadequate viral load suppression, inadequate concentrations/efficacy of each class of antiretroviral agent in the regimen and the pre-existence

of drug resistant mutants are the primary reasons why drug resistance emerges [118], and not necessarily the use of low drug doses. If for a particular individual, a drug dose that is below the specified therapeutic range can adequately suppress the viral load, then there is no need to increase the dosage.

5.3 Prioritization of Objectives of Therapy

Eradication of HIV infection cannot be achieved with available antiretroviral regimens and now the focus has shifted from virus eradication to managing a chronic infection. Therefore, once the decision is made to initiate therapy, the primary goals of antiretroviral therapy, according to [1] are to:

- reduce HIV-related morbidity and mortality,
- improve quality of life,
- restore and preserve immunologic function, and
- maximally and durably suppress viral load.

The maximum count to which the T cells can rebound depends on the state of the immune system or the extent to which the immune system is repairable, and not necessarily on the ability of the drugs to suppress the viral load. However, it is currently not clear how one can quantify the health of the immune system, especially so for the chronically infected patient. One could intuitively, expect the health of the immune system to correlate with the source rates s_T and s_M at which uninfected $CD4^+$ T cells and macrophages are produced. However, this expectation can not be upheld because the $CD4^+$ T cell steady state value does not depend on parameter s_T, as illustrated by equation (4.1) in section 4.1.1.

Given that there are problems associated with current antiretroviral agents, therapy should ideally entail the use of minimal drug dosage schemes that reduce toxic effects, and simultaneously maximally/effectively suppress the viral load for as long as possible. However, minimal dosing and maximal suppression of the viral load are conflicting objectives. There is therefore a need to prioritize ones objectives in order to strike a balance between aggressive therapy and toxicity reduction.

Numerous clinical trial have shown that if one can suppress the viral load, there will be some degree of $CD4^+$ T cell and macrophage cell rebounds. It is therefore, more practical to aim to suppress the viral load than to try and maintain a pre-determined $CD4^+$ T cell count. That is, more emphasis should be placed on viral load control. A reference trajectory that specifies the limits to the time available to suppress the viral load from when therapy is initiated should be determined, bearing in mind that

antiretroviral drugs are generally toxic.

5.4 Model Predictive Control

Model Predictive Control - MPC is a technique in which the control is determined by solving, at each sampling instance, a finite-horizon optimal control problem [53]. The optimal solution is attained while respecting the constraints on the system. MPC's "constraint-tolerance" is what differentiates it from other optimal control strategies [185].

The MPC controller is a discrete-time controller and only computes the control move at regularly spaced, discrete time instants. Furthermore, the controller does not accept updates on the measured plant outputs between sampling instances. Once the optimal control sequence is computed, it is sent to the plant, but only for one sampling period. The plant operates with this constant input until the next sampling instant [185]. At the next sampling instant, the controller accepts new plant measurements and a new optimal solution is computed. It is hence, assumed that the output measurement is available when computing the control sequence.

Predictive control is model based in the sense that the controller explicitly uses an internal plant model to make predictions of future plant behaviour [186]. Model based control could therefore, be a drawback if the model or parameters are not accurate, as is the case with HIV/AIDS models. However, MPC has a certain degree of robustness to model inaccuracies [187].

5.4.1 Overview

The following summary of how Model Predictive Control works is a direct extraction from [186, 187]:

A predicative controller assumes a discrete time setting and the current time is referred to as k.
For a discrete system of the form $X_{k+1} = f(X_k, u_k)$, and a current state X_k a length H_u sequence $U = \{u_k, u_{k+1}, ..., u_{k+H_u-1}\}$ is derived which minimizes a cost function of the form

$$V(k) = \sum_{j=H_w}^{H_p} \parallel \hat{z}(k+j|k) - r(k+j|k) \parallel^2_{Q(j)} + \sum_{j=0}^{H_u-1} \parallel \Delta\hat{u}(k+j|k) \parallel^2_{R(j)}$$

The cost function V penalizes deviations of the predicted controlled outputs $\hat{z}(k+j|k)$ from a (vector) reference trajectory $r(k+j|k)$. The prediction horizon has length H_p, but

it is not necessary to start penalizing deviations of z from r immediately (if $H_w > 1$). The prediction and control horizons H_p and H_u, the 'window' parameter H_w, the weights $Q(j)$ and $R(j)$, and the reference trajectory $r(k+j)$, all affect the behaviour of the closed-loop combination of plant and predictive controller. Some of these parameters, particularly the weights, may be dictated by the economic objectives of the control system, but usually they are in effect tuning parameters which are adjusted to give satisfactory dynamic performance.

The controller has an internal model which is used to predict the behaviour of the plant, starting at the current time, over a future prediction horizon. This predicted behaviour depends on the assumed input trajectory $\hat{u}(k+j|k)$, $(j = 0, 1, ..., H_p - 1)$ that is to be applied over the prediction horizon, and the idea is to select that input which promises the best predictive behaviour. The procedure therefore entails calculation of a control sequence minimizing an objective function, and relies on the explicit use of the model to predict the process output at future time instants. It is assumed that the output measurement $y(k)$ is available when deciding the value of the input.

Once a future input trajectory has been chosen, only the first element of that trajectory is applied as the input signal to the plant. Then the whole cycle of output measurement, prediction and input trajectory is repeated, one sampling interval later: a new output measurement $y(k+1)$ is obtained; a new reference trajectory $r(k+j|k+1)$, $(j = 2, 3, ...)$ is defined; predictions are made over the horizon $k+1+j$, with $(j = 1, 2, ..., H_p)$; a new input trajectory $\hat{u}(k+1+j|k+1)$, with $(j = 0, 1, ..., H_p - 1)$ is chosen; and finally the next input is applied to the plant: $u(k+1) = \hat{u}(k+1|k+1)$. Since the prediction horizon remains the same length as before, but slides along by one sampling interval at each step, this way of controlling a plant is often called a receding horizon strategy. This concept is illustrated in figure 5.4.

Constraints of the following form are assumed to hold over the control and prediction horizons:

$$E\ vec(\Delta\hat{u}(k|k), \ldots, \Delta\hat{u}(k+H_u-1|k), 1) \leq vec(0)$$

$$F\ vec(\hat{u}(k|k), \ldots, \hat{u}(k+H_u-1|k), 1) \leq vec(0)$$

$$G\ vec(\hat{z}(k+H_w|k), \ldots, \hat{z}(k+H_p|k), 1) \leq vec(0)$$

It is also possible to have the converse situation, of variables which are constrained but do not appear in the cost function (*zone* objectives). This is only likely to occur with the z variables, and is represented in the standard formulation by having appropriate zero entries in the weighting matrices *Q(j)*.

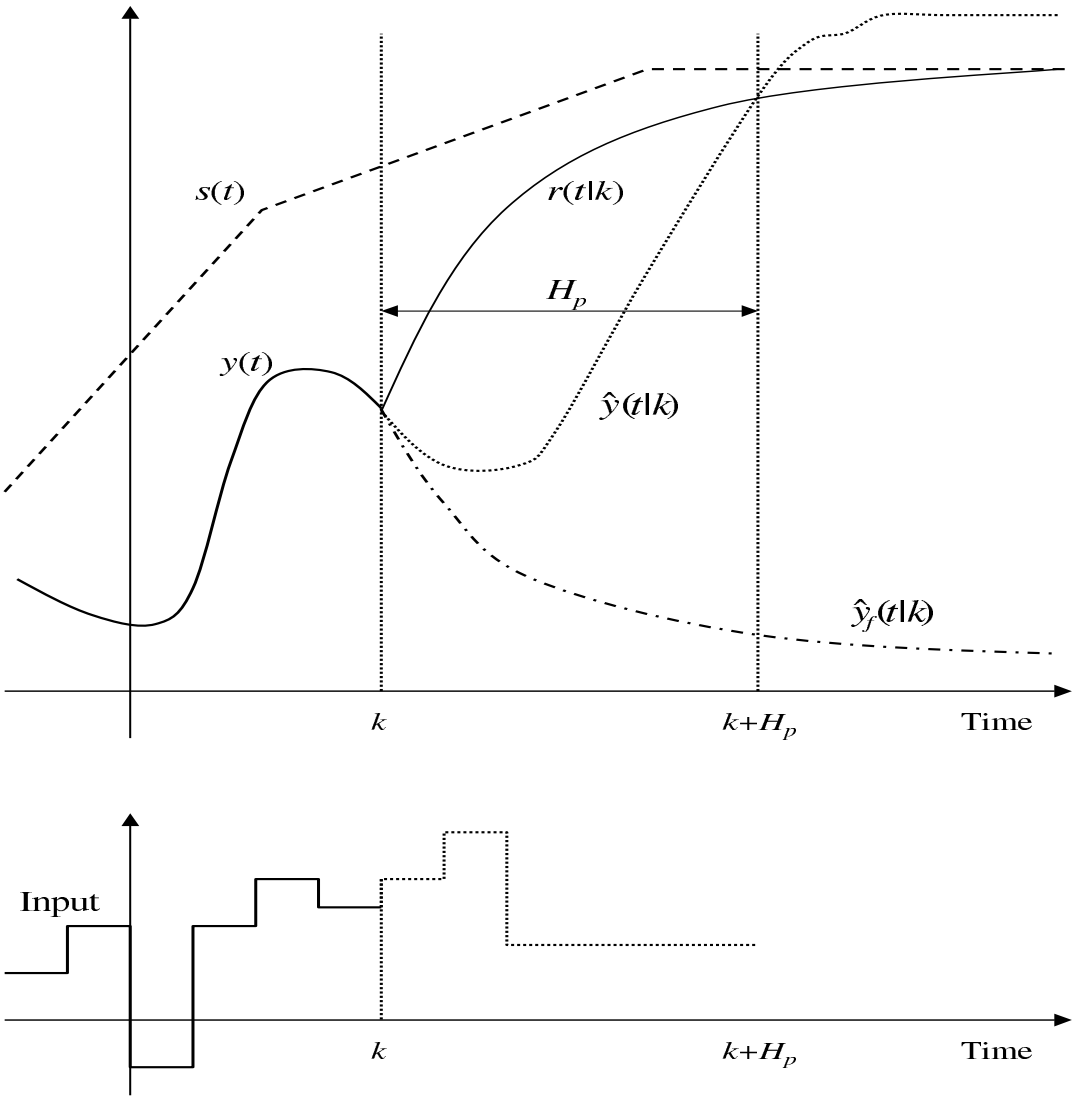


Figure 5.4: Predictive control - receding horizon concept. The current time is referred to as step time k. $y(t)$: plant output, $s(t)$: set-point trajectory, $r(t|k)$: reference trajectory, $\hat{y}(t|k)$: Assumed plant output, H_p: prediction horizon [186].

5.4.2 Suitability for HIV/AIDS Drug Dosage Design

MPC has many uses in drug delivery systems. [53] used MPC to derive an On/Off dosage sequence that enhances immune response to HIV infection, while [181] used MPC in a complex physiologically based drug model. MPC is attractive to use in drug delivery systems because it has a certain degree of robustness to model inaccuracies and can be easily tuned [187].

For HIV/AIDS control, MPC offers the flexibility to incorporate some clinically observed but not modelled phenomena, such as the fact that the $CD4^+$ T cell count does not usually rebound to pre-infection values, even with prolonged viral load suppression. Weights can also be adjusted to favour the use of one class of antiretroviral inhibitor over the other, and the limits for the operating therapeutic range can be directly set. The time between samples is long enough to allow for computation and an observer can be used for states that are not readily measured. MPC is easily adaptable and this allows for other improved HIV/AIDS models to be integrated as they are developed [53].

Recalculation of a new optimal control sequence at each sampling instance serves viral load control very well, since the predictions that were made during the previous optimization stage are not perfect. As illustrated in figure 5.1, the plant is subjected to various unexpected disturbances. Measurement feedback and reformulation therefore allows the controller to correct for these errors and unexpected disturbances.
MPC is therefore well suited for HIV control.

5.5 Sampling

After therapy is initiated, blood samples are taken after about 2-8 weeks in order to assess the initial effect of the drugs. Thereafter, guidelines call for sampling every 3-4 months to monitor for virologic failure and the emergence of resistance, as presented in Table 5.4 [1].

Clearly, the recommended measurement schedule is inadequate for feedback control. There are also indications from various studies that antiretroviral therapy should be individualized and depend on the individuals response to therapy [50, 51, 177, 184]. This in turn calls for more frequent or adequate sampling of the viral load and host cell counts. As a rule of thumb, one should sample at a rate that is 5-10 times the signal bandwidth, for feedback control [188]. However, due to the invasive nature of taking blood samples, it is desirable to keep sampling at a minimum with due consideration to patient discomfort, the need for one to avail themselves and cost.

Table 5.4: Indications for plasma HIV RNA testing

Clinical Indication	**Information**	**Use**
Syndrome consistent with acute HIV infection	Establishes diagnosis when HIV antibody test is negative or indeterminate	Diagnosis
Initial evaluation of newly diagnosed HIV infection	Baseline viral load set-point	Use in conjunction with $CD4^+$ T cell count for decision to start or defer therapy
Every 3−4 months in patients not on therapy	Changes in viral load	Use in conjunction with $CD4^+$ T cell count for decision to start therapy
2−8 weeks after initiation of or change in antiretroviral therapy	Initial assessment of drug efficacy	Decision to continue or change therapy
3−4 months after start of therapy	Assessment of virologic effect of therapy	Decision to continue or change therapy
Every 3−4 months in patients on therapy	Durability of antiretroviral effect	Decision to continue or change therapy
Clinical event or significant decline in $CD4^+$ T cells	Association with changing or stable viral load	Decision to continue, initiate, or change therapy

Reproduced from [1].

As stated before, there are inter-individual variations in parameter estimates. These variations in parameter values lead to variations in steady state cell counts and viral load set points from one individual to the other [171] and consequently, to inter individual variations in the response to therapy. Furthermore, these parameter ranges are very wide for some of the model parameters. This explains, in part, why some individuals have virologic failure on therapy that is highly effective on others. There is a need therefore, to determine the individual's sampling interval from one's parameters.

5.6 A Sequential Perturbation Approach to Dosage Design

5.6.1 Strategy

The strategy, in light of the objectives of therapy and the issues discussed in sections 5.1 - 5.3 is: Determine the operating therapeutic range and an appropriate sampling interval for the individual. The next step is then to design or select a desired viral load trajectory. Finally, a control input sequence to attain the desired viral load trajectory is obtained using MPC with the appropriate constraints in place. Only the use of replication cycle based HAART (RTI and PI containing regimens) will be considered in this section. That is, $u_H = 1$, while $u_I = 0$.

5.6.2 Objective Function and Constraints

Measured Outputs:
The measured outputs are taken as the uninfected $CD4^+$ T cells T and the infectious virus particles V_i.

In current practice however, discriminatory T cell count measurements are not readily attainable. The total cell count $T_{tot} = T + T_l + T_a$ is what is measured. However, it has been observed that, under therapy, the infected cells make a very small percentage of the total T cell count. Therefore, in this case, it is reasonable to assume that $T \simeq T_{tot}$.

The measured viral load is also the total of the infectious and noninfectious particles, that is, $V_{tot} = V_i + V_n$ is what is currently measured. If reverse transcriptase inhibitors are exclusively used in the regimen, then the measured viral load is the infectious virus particles, and $V_{tot} = V_i$. If the regimen includes protease inhibitors with a known efficacy, then one can calculate the infectious virus particles from the total measured viral load.

Objective Function:

The intention of therapy is to suppress the viral load as much as possible. However, given the cost and undesirable side effects of the antiretroviral drugs, it would be best if such viral load suppression is attained with the least amount of drugs. No attempt is made to control the uninfected $CD4^+$ T cell count.
The objective function is therefore selected as

$$min\ J(v,\eta) \quad = \quad \sum_{j=1}^{N_p} \delta(j)[\hat{v}(t+j|t) - r(t+j)]^2 + \sum_{j=0}^{N_\eta - 1} \lambda(j)[\eta(t+j-1)]^2 \qquad (5.6)$$

where N_p is the prediction horizon and N_η is the control horizon. The intention is that the future viral load output $(\hat{v})$ on the considered horizon should follow a determined reference trajectory (r), and at the same time, the drug dosage (η) necessary for doing so should be penalized [187]. $\delta(j)$ and $\lambda(j)$ are weights that consider the future behaviour and are used as tuning parameters.

Sampling Interval:
There are numerous reports of viral load rebounds for individuals on HAART, whose viral load had been previously suppressed to below detectable levels. This phenomena is referred to as viral 'blips' [22, 23, 25]. Some rebounds are transient, while others lead to virologic failure. Predicting the timing and frequency of these blips, as well establishing which blips are transient and which are indicative of virologic failure, is currently the subject of intense research.

A step response analysis of the system has shown that it is also possible to obtain short term viral load suppression with low drug doses due to the transient undershoot attained when therapy is initiated [104], as illustrated in figure 5.5. This figure 5.5 was generated with the nonlinear model when therapy is initiated 200 days from initial infection, using a constant low efficacy dosage. The point being made is that there is a transient undershoot after therapy is initiated and the dosage has to be adjusted before the viral load rebounds, with the arrow indicating the desired viral load direction.

The sampling interval was accordingly determined as the time to reach the undershoot, as the dosage needs to be adjusted before the viral load rebounds. The intention of the proposed sampling interval is to have as few samples as possible, with due consideration to the invasive nature of taking blood samples and patient discomfort. The sampling interval was estimated from the step response of the linearized system when therapy is initiated at the asymptomatic stage. The sampling interval thus chosen is parameter dependent and will vary from one individual to the other. For the parameter set presented in Table 5.1, the sampling interval was determined to be 15 days, ie. $t_s = 15$ days.

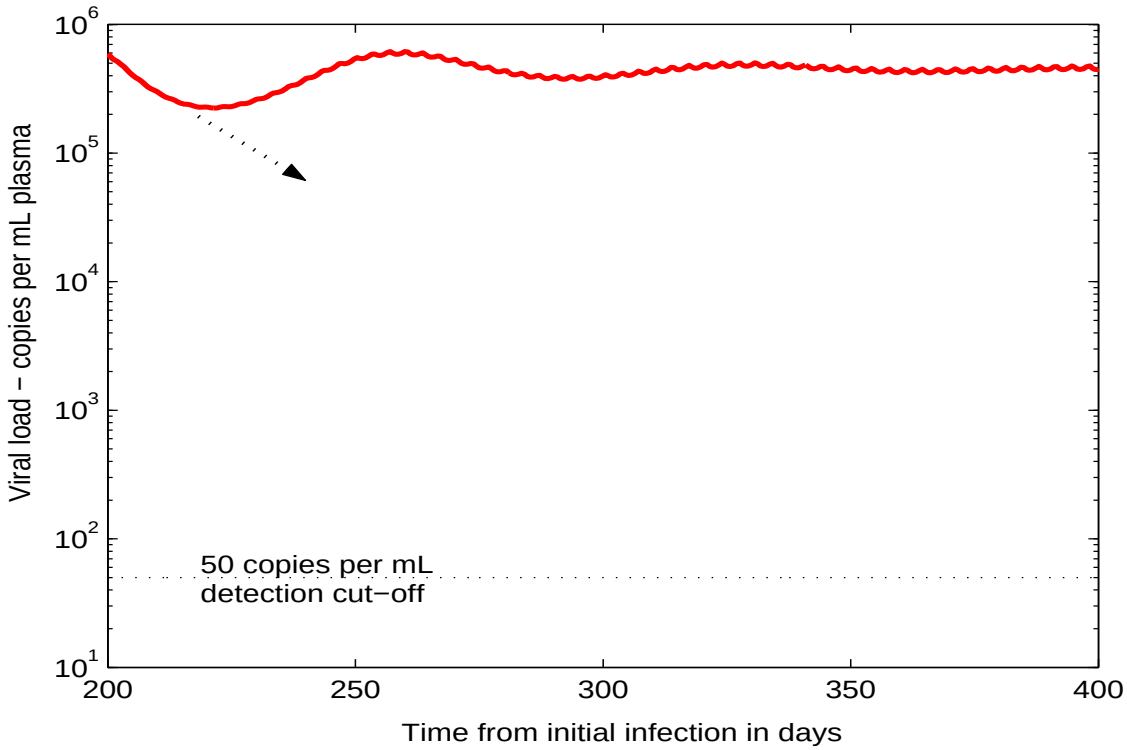


Figure 5.5: Transient effect of a fixed low efficacy dosage on viral load when therapy is initiated.

Prediction and Control Horizons:
Prediction was done over 120 days. With the sampling interval determined as $t_s = 15$ days, then $N_p = 8$ and the control horizon of $N_\eta = 3$ was selected. This configuration is motivated by the guidelines' recommendation on plasma HIV testing (presented as Table 5.4). The guidelines call for testing 2-8 weeks after initiation of therapy for the initial assessment of drug efficacy. Thereafter, test 3-4 months after start of therapy for assessment of virologic effect of therapy and the durability of antiretroviral effect.

Reference Trajectory:
The desired viral load trajectory was selected based on the guidelines notion of a controllable viral load. The guidelines consider antiretroviral therapy to be effective if it can reduce the viral load by 90% in less than 8 weeks and continue to suppress it to below 50 copies per mL of plasma in less than 6 months [1].

These conditions are loosely stated, hence for this thesis, an additional requirement that such viral load suppression to below 50 copies per mL of plasma, should not be

attained before 5 months of initiating therapy is added.

The weight $\delta(j)$, penalizing any deviations of the viral load ($\hat{v}$) from the reference trajectory (r), was selected as the increasing function

$$\delta(j) = \sigma^{N_p - j} \qquad 0 < \sigma < 1 \tag{5.7}$$

because more emphasis is placed on viral load suppression than on the initial viral load response.

Efficacy Constraints:
As stated before, perfect inhibition of virus replication occurs if $\eta = 1$, and there is no inhibition if $\eta = 0$. The assumption is usually that perfect inhibition is not possible with the currently available antiretroviral agents. That is $0 \leq \eta < 1$ for both types of inhibitors.

Therapeutic Range Constraints:
The upper limit for the operating therapeutic range was determined from the steady state analysis that was carried out in section 4.1.1, as the 50 copies per mL viral load steady drug efficacy η_{sup} as given by equation (4.15). The rationale for selecting η_{sup} as the upper therapeutic range limit was presented in that section. For the parameter set presented in Table 5.1, the therapeutic limit was determined to be $\eta_{th} = \eta_{sup} = 0.72$. This constraint is lifted however, when only one class of inhibitor is available for use in the regimen.

CD4$^+$ T Cell Count Rebound:
No attempt is made to control either the CD4$^+$ T cell or macrophage cell counts. The assumption is that cell counts will increase from their pre-treatment values. However, the CD4$^+$ T cell count under therapy can not rebound to values higher than the pre-infection steady state value. That is, $\bar{T} \leq T \leq T(0)$ during therapy. Values for the pre-infection CD4$^+$ T cell count $T(0)$, the pre-treatment CD4$^+$ T cell measurement, T_{ss} and the pre-treatment viral load measurement, $\bar{V}_i$ are presented in Table 5.2.

5.6.3 Dosage Sequence Design

MPC is applied to derive the required drug efficacy input sequences for the HIV/AIDS model Σ_{Eco} (5.1) with multiple target cells and differential drug penetration into these cells. Only the extended cell model is used in this section because this model can realis-

Table 5.5: Summary of constraints

Type	Lower and Upper Limits	Notes
Efficacy - RTI	$0 \leq \eta_{rt} < 1$	Perfect inhibition not possible.
Efficacy - PI	$0 \leq \eta_{pi} < 1$	Perfect inhibition not possible.
Therapeutic - RTI	$\eta_{th(rt)} = 0.72$	Lower limit not established.
Therapeutic - PI	$\eta_{th(pi)} = 0.72$	Lower limit not established.
$CD4^+$ T cell rebound	$\bar{T} \leq T \leq T(0)$	$T(0)$: Pre-infection $CD4^+$ T cell count.
Viral load	$0 < V_i \leq \bar{V}_i$	$\bar{V}_i$: Pre-treatment viral load measurement.

t_s	15 days
N_p	8 (120 days)
N_η	3
λ	**1**

tically simulate low viral loads under potent HAART. The assumption is that both the reverse transcriptase and protease inhibitors are available for therapy.

- Since MPC normally requires a linear discrete-time controller, the extended model Σ_{Eco} (5.1) was first linearized, then discretized to obtain the internal controller model.
- The nonlinear plant was maintained as is given by system Σ_{Eco} (5.1).
- The applicable constraints are summarized in Table 5.5.

5.6.4 Results

For dosage sequence design, the algorithm presented in the preceding section was implemented using Matlab. Unless otherwise stated, parameters and constraints are as presented in Table 5.1 and Table 5.5.

As a bench mark, dosage efficacy sequences were derived with no therapeutic range constraints. That is, when $0 \leq \eta_{rt} < 1$, $0 \leq \eta_{pi} < 1$ was the only constraint on the inputs. Simulations were performed for a prediction horizon of 120 days and 3 control moves ($N_p = 8$ and $N_\eta = 3$). The results are presented in figures 5.6 and 5.7.

Please note that mono class therapy does not necessarily imply the use of a single inhibitor as in mono drug therapy. An example of mono class therapy is the currently

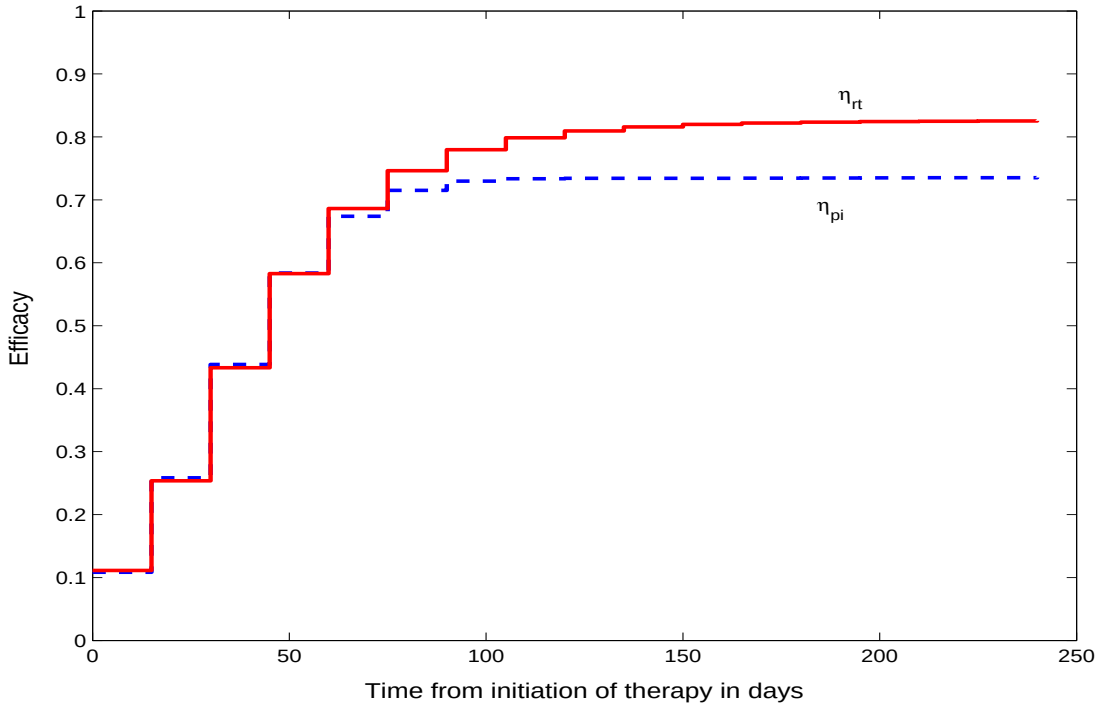


Figure 5.6: Control input sequence for mono (single) class therapy with no therapeutic range constraint. η_{rt} : RTI efficacy; η_{pi} : PI efficacy. Sampling interval t_s=15 days.

used 2NRTI plus 1NNRTI regimen consisting of didanosine, lamivudine and efavirenz [1], where all drugs are from the reverse transcriptase class of antiretroviral agents.

For mono class therapy, figure 5.6 suggests that from an end point efficacy perspective, protease inhibitors are slightly better than reverse transcriptase inhibitors at controlling the infectious viral load because the end efficacy for the PIs is less than than of the RTIs. This is understandable when one considers that under reverse transcriptase inhibitors, the viral load is exclusively infectious, while the viral load is split between infectious and noninfectious types when protease inhibitors are used. However, care should be taken not to confuse a lower end point efficacy of an inhibitor to automatically imply a lower pill intake.

When both classes of inhibitors are available or used in the regimen, then for this combined therapy, figure 5.7 shows that for an individual with a typical set of parameters as in Table 5.1, the resulting dosage efficacy sequence suggests starting therapy with a low efficacy reverse transcriptase inhibitor, and sequentially increasing the efficacy over time. The protease inhibitors are later added to the regimen. The protease inhibitor efficacy also starts low, then is sequentially increased while the reverse transcriptase inhibitors

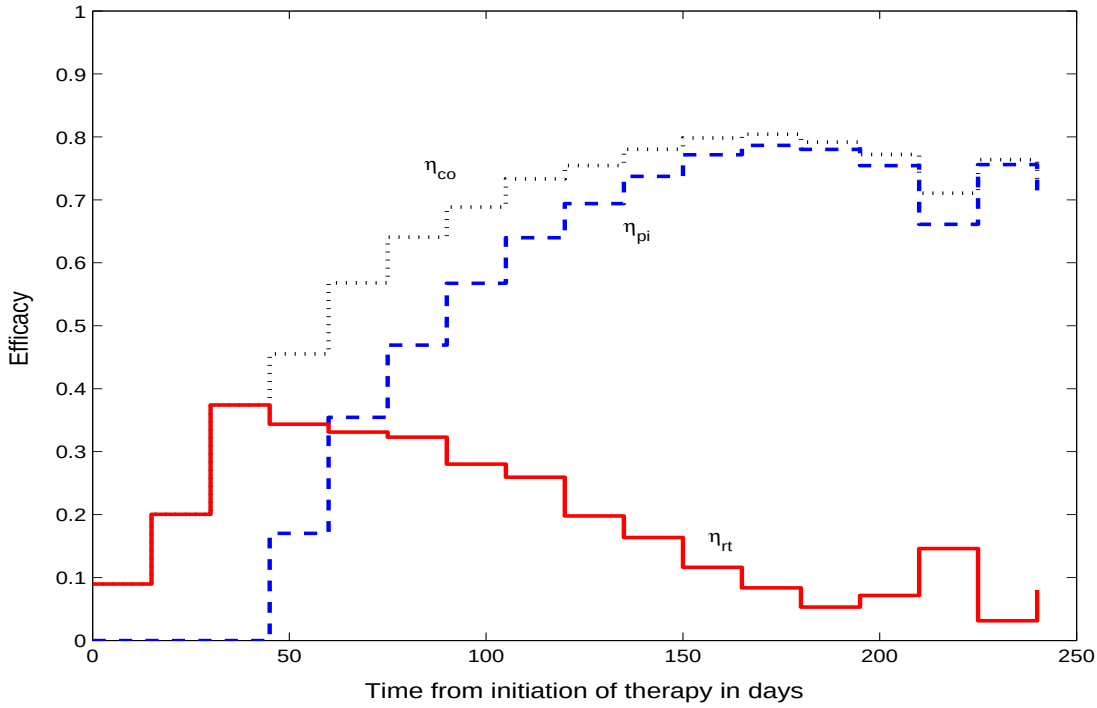


Figure 5.7: Control input sequence for combined therapy with no therapeutic range constraint. η_{rt} : RTI efficacy; η_{pi} : PI efficacy; η_{co} : Combined efficacy. Sampling interval t_s=15 days.

are eventually removed from the regimen. Of interest is the distinct switching from the reverse transcriptase class of inhibitors to the protease inhibitors during therapy. This switch could be explained by figure 5.6 because the eventual efficacy for the protease inhibitors is lower than that of the reverse transcriptase inhibitors.

The derived dosage sequence as depicted in figure 5.7 therefore inherently favours a protease inhibitor intensive regimen for the eventual suppression of the viral load, and to keep it below detectable levels. This may pose a problem when one considers that protease inhibitors are generally more toxic (long term) than the reverse transcriptase inhibitors [189], and most regimens are currently more inclined towards the use of reverse transcriptase inhibitors. Furthermore, the eventual elimination of the reverse transcriptase inhibitors from the regimen may be undesirable, as the protease inhibitor's efficacy ends up being high. The eventual combined drug efficacy

$$\eta_{co} = 1 - (1 - \eta_{rt})(1 - \eta_{pi}) \tag{5.8}$$

is the efficacy that attains a viral load steady state below 50 copies per mL of plasma, as is illustrated in figure 5.7. Figure 5.8(a) shows the resulting viral load outputs for

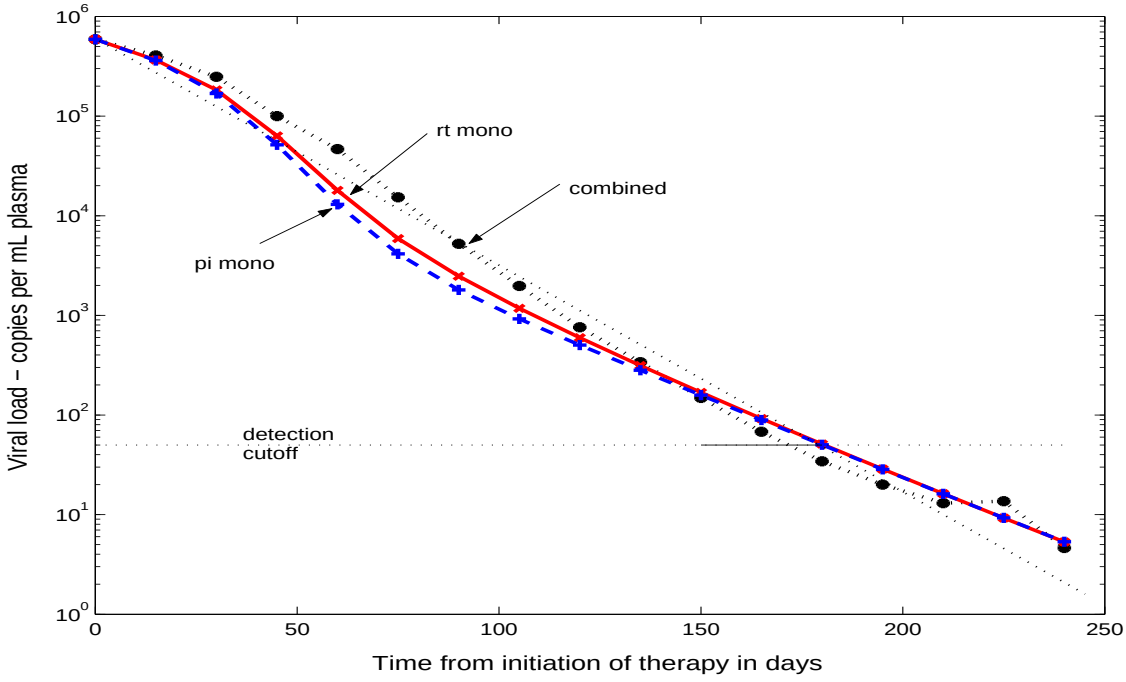
the dosage input sequences in figures 5.6 and 5.7, and shows that combined therapy has better viral load control when compared with mono class therapy. However, there is no significant difference in $CD4^+$ T cell count outputs between mono class and combined therapy, as illustrated in figure 5.8(b).

So, for this particular case, if drug resistance and long term protease inhibitor toxicity were not an issue, then the results suggest that reverse transcriptase inhibitors should be the choice basis for the starting regimen, while protease inhibitors should be the choice basis for the subsequent suppression and maintenance regimen. The general outlook is that there is no need to start treatment with high dosages. The dosage should start low and be sequentially increased over time. Also, there may be no need to start therapy with inhibitors from both classes. Therapy can start with one class, and the other class can be added later.

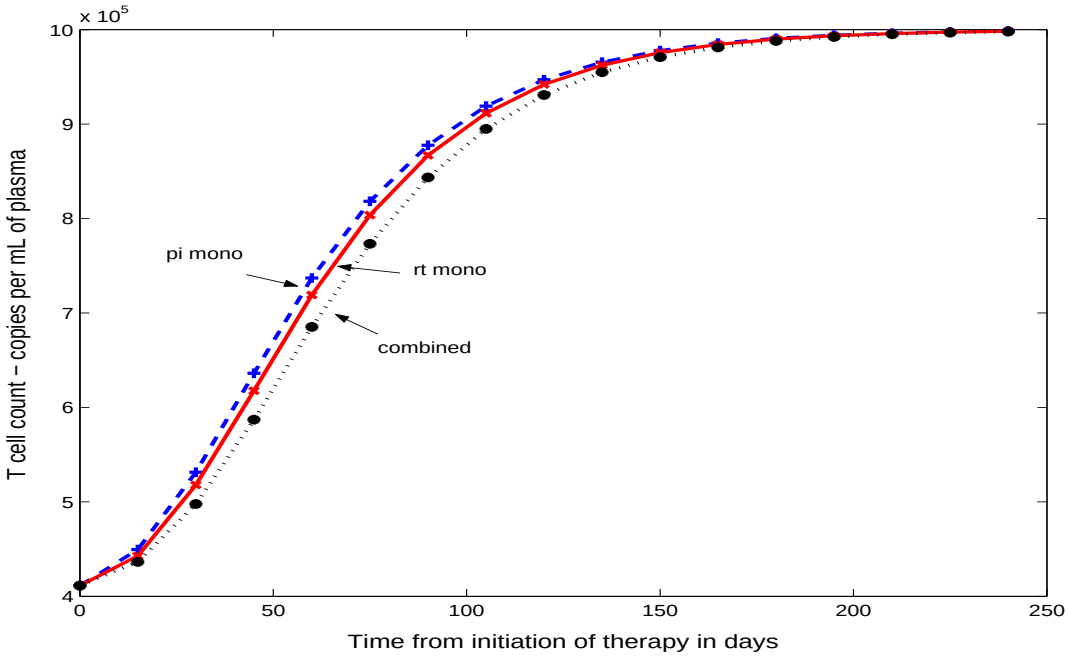
The problem of high efficacies for a particular class of antiretroviral agent can be reduced by imposing an individualized operating therapeutic efficacy range. A lower efficacy limit will prevent the elimination or exclusion of the class from the regimen. An upper efficacy limit should reduce cumulative and instantaneous toxicity, and could also prevent class exclusion or elimination. Figure 5.9 shows the resulting dosage efficacy sequence when an upper operating limit is imposed on the protease inhibitor efficacy. Still, the derived dosage sequence favours the use of protease inhibitors for the eventual suppression of the viral load. The upper limit on the protease inhibitor efficacy however, keeps the reverse transcriptase inhibitors as part of the regimen. This however, has no significant effect on the viral load and $CD4^+$ T cell count output. Figure 5.9 also shows the fixed dosage efficacy η_{fix} that is equally capable of suppressing the viral load.

A comparison between the MPC derived dosage input sequence and the fixed dosage approach to therapy is made in figures 5.10, 5.11 and 5.12. The initial viral load and $CD4^+$ T cell responses are much faster for the fixed dosage regimen. This is understandable considering that the fixed regimen starts therapy with a much higher efficacy. However, the viral loads for both therapy approaches reach the 50 copies per mL detection cutoff almost at the same time. For the infected $CD4^+$ T cells, the decline pattern for the actively infected $CD4^+$ T cells as illustrated by figure 5.11(b) is similar to that of the viral load (figure 5.10(a)). Macrophage rebound and decline is slower than that of the $CD4^+$ T cells.

What is clear from these simulations is that, if the intention is to reduce drug toxicities, the logical way to schedule drugs is to take advantage of the initial viral load undershoot when dynamics are perturbed and start therapy with a low drug efficacy.



(a)



(b)

Figure 5.8: System outputs with no therapeutic range constraint. rt mono : RTI mono class therapy; pi mono : PI mono class therapy; combined : Combined therapy; Sampling interval $t_s = 15$ days. (a) Viral load (b) $CD4^+$ T cells

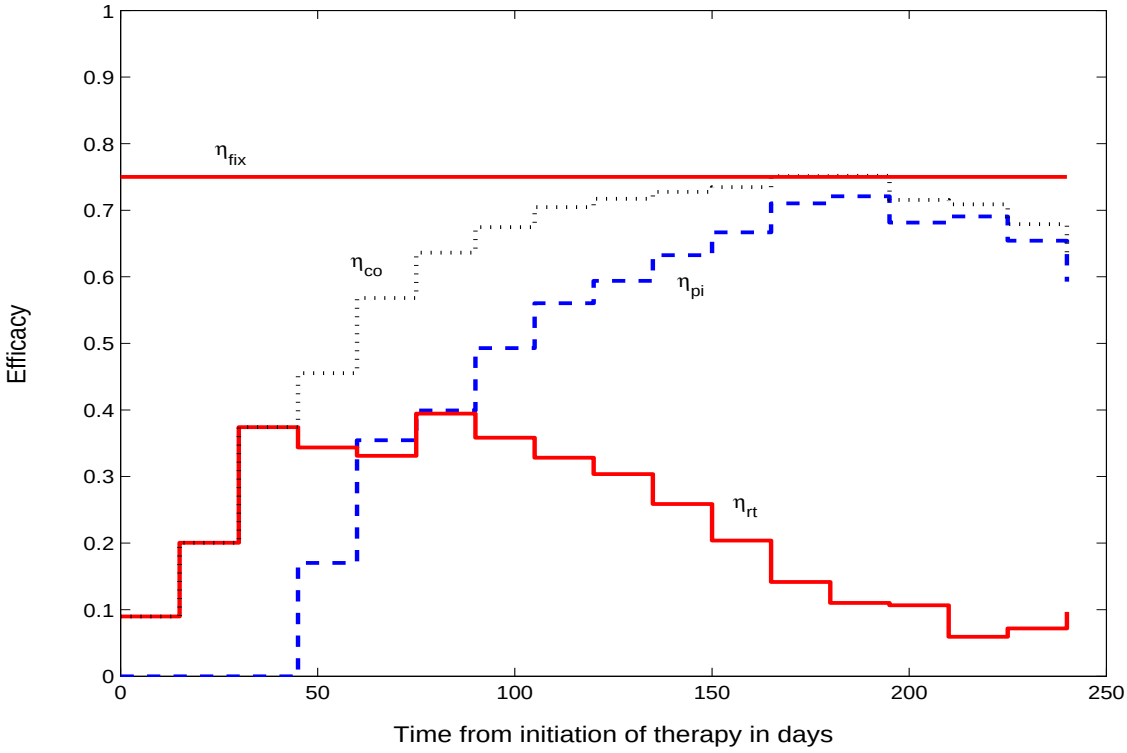
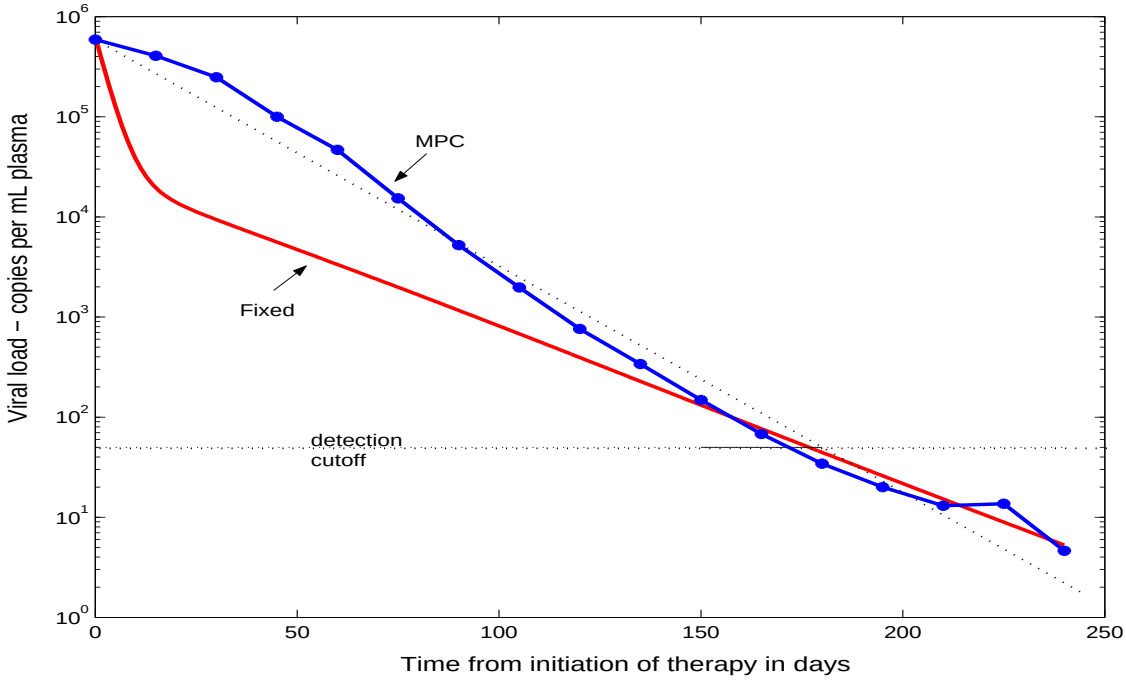


Figure 5.9: Control input sequence for therapeutic range constrained combined therapy ($t_s = 15$ days) and fixed dosage therapy. Fixed efficacy $\eta_{fix} = 0.75$.

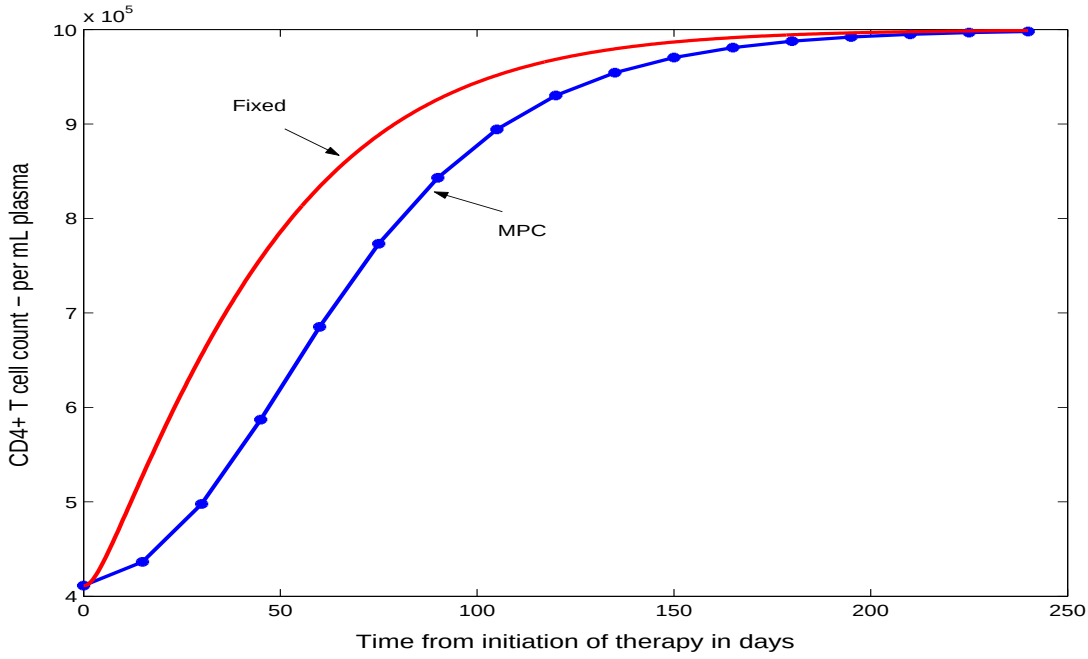
Then the dosage must be sequentially increased over time to prevent the viral load from rebounding, increased again to reduce the viral load further, and eventually increased to suppress and keep the viral load below detectable levels. This results in a sequential perturbation approach to antiretroviral therapy.

The derived dosage sequences as presented in figures 5.6, 5.7 and 5.9 suggest that the dosage should start low and be sequentially increased over time. This suggested dosage schedule could be seen as being contrary to how ARVs are currently scheduled. However, one needs to take note of the following points:

1. Starting therapy with a low efficacy dosage does not necessarily imply that it will take longer to suppress the viral load. This point is illustrated in figures 5.10 and from that perspective, the proposed MPC dosage schedule and the fixed, high all the time efficacy dosage are equally aggressive.
2. Given that the objective is to strike a balance between aggressive therapy and toxicity reduction (instantaneous and accumulative), then an equally aggressive therapy schedule that uses less drugs is better.

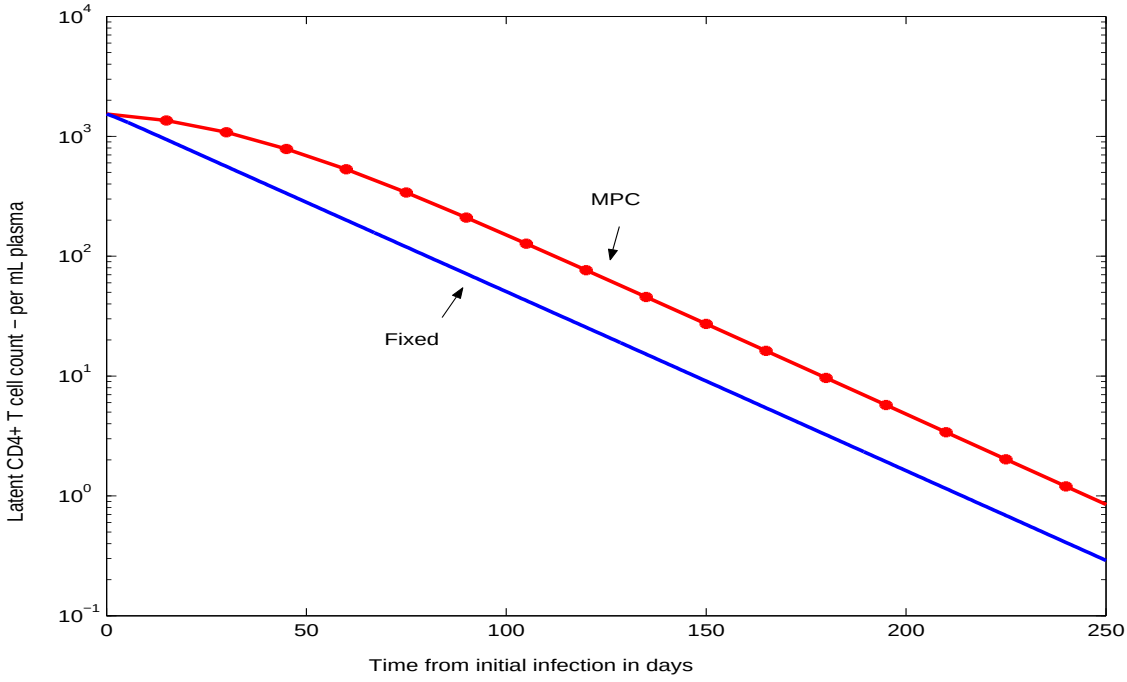


(a)

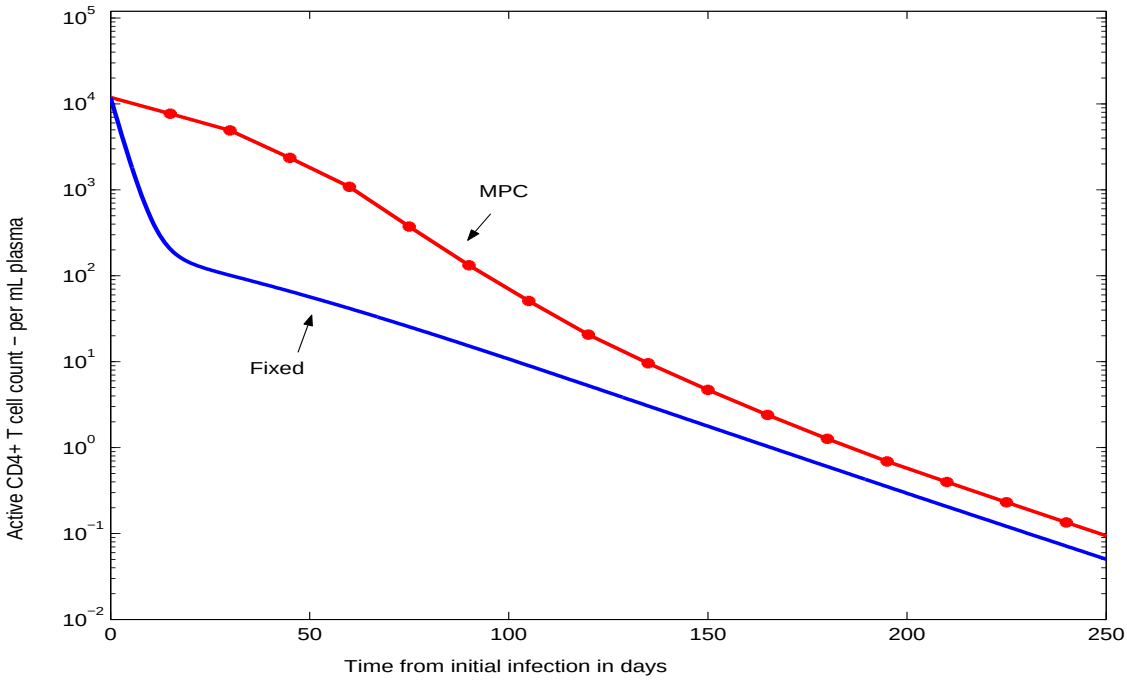


(b)

Figure 5.10: System outputs for therapeutic range constrained combined therapy ($t_s = 15$ days) and fixed dosage therapy. Control input sequences are in figure 5.9. (a) Viral load, (b) $CD4^+$ T cells.

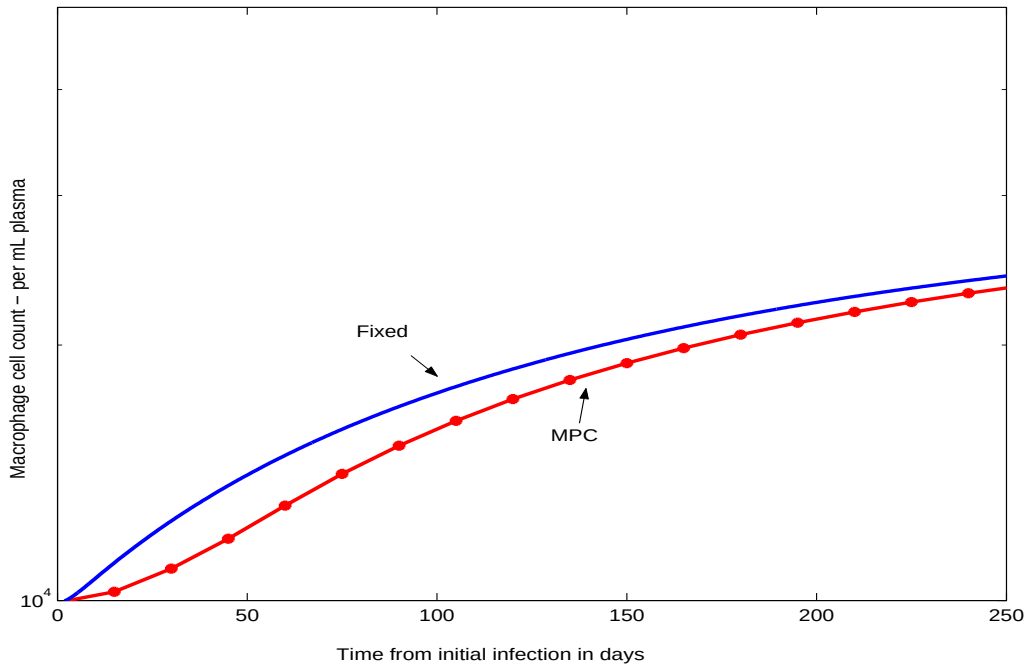


(a)

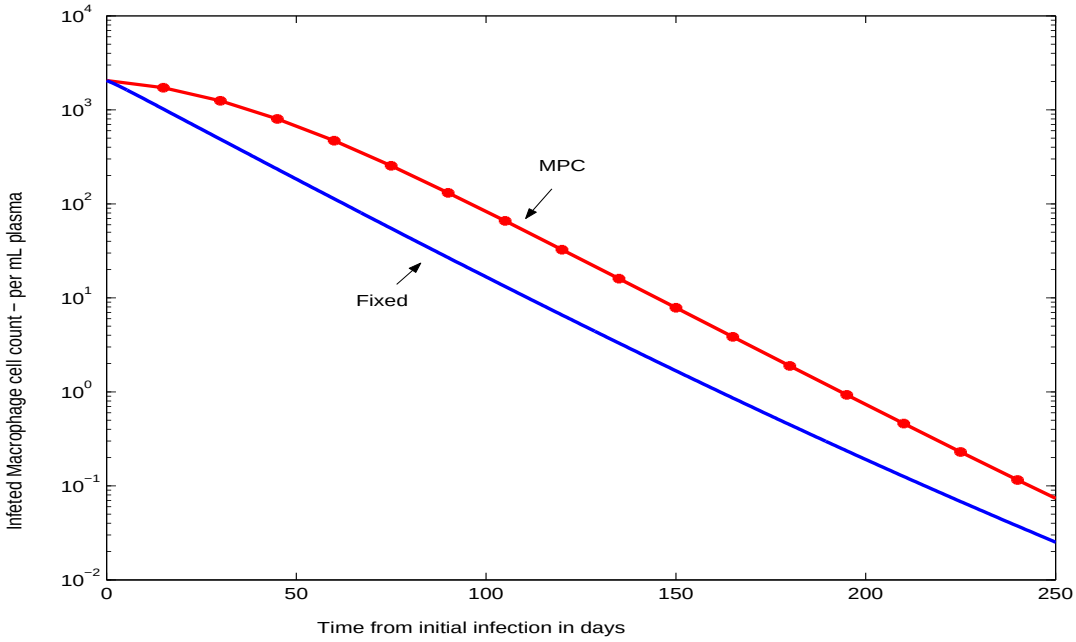


(b)

Figure 5.11: System outputs for therapeutic range constrained combined therapy ($t_s = 15$ days) and fixed dosage therapy. Control input sequences are in figure 5.9. (a) Latently infected CD4$^+$ T cells, (b) Actively infected CD4$^+$ T cells.



(a)



(b)

Figure 5.12: System outputs for therapeutic range constrained combined therapy ($t_s = 15$ days) and fixed dosage therapy. Control input sequences are in figure 5.9. (a) Uninfected macrophage cells, (b) Infected macrophage cells.

Table 5.6: Initial efficacy and drug exposure variation with sampling interval

Sampling interval	Initial efficacy	Drug exposure
15 days	0.10	0.73
30 days	0.43	0.88
60 days	0.70	0.97
Fixed dosage	0.75	1

Calculated from initiation of therapy up to day 180

5.6.5 Effect of Inadequate Sampling

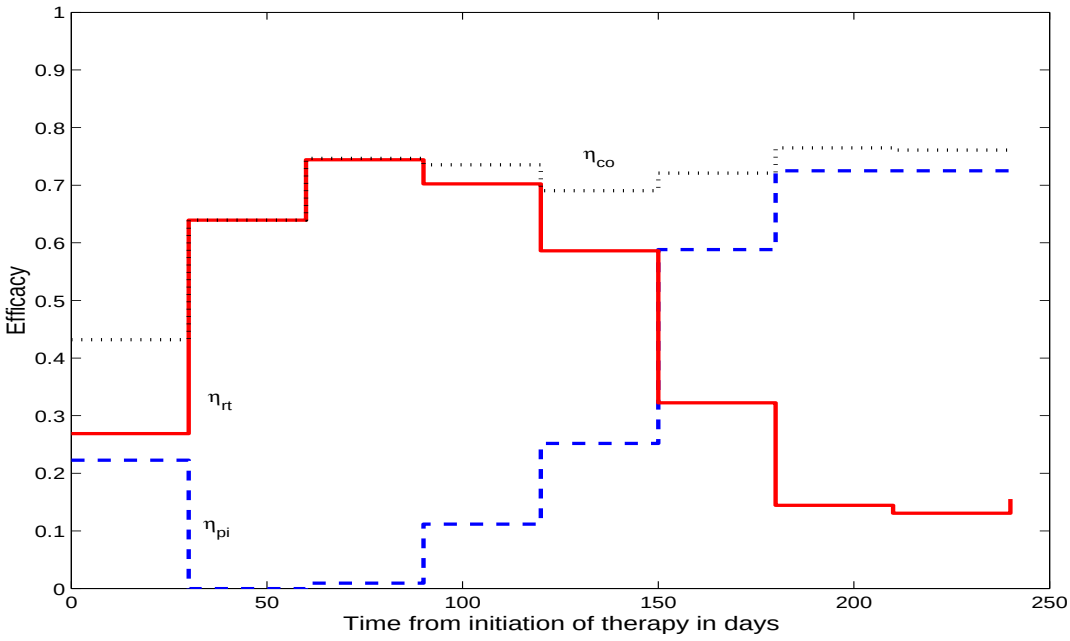
In this section, the sampling interval was varied in order to assess its effect on the resulting dosage efficacy input sequence. Figures 5.13(a) and 5.13(b) show the resulting input sequences when samples were taken every month (30 days) and every 2 months (60 days), respectively. Increasing the interval between taking samples increases the required initial drug efficacy when starting therapy from as low as 11% for a 2 week sampling interval, to 35% for a monthly interval, and as high as 68% when samples are taken once every 2 months.

The once every 2 months dosage efficacy sequence is more like how antiretroviral therapy is currently administered, as it is more inclined towards a fixed dosage regimen. This puts an emphasis on adequate sampling in order to reduce total drug intake. This could explain why current regimens dictate the use of potent HAART from the start. The relative increase in total drug exposure due to increasing the sampling interval is summarized in Table 5.6. Adequate viral load sampling therefore can reduce total drug exposure by up to 27%, as is the case for a 15 day sampling interval.

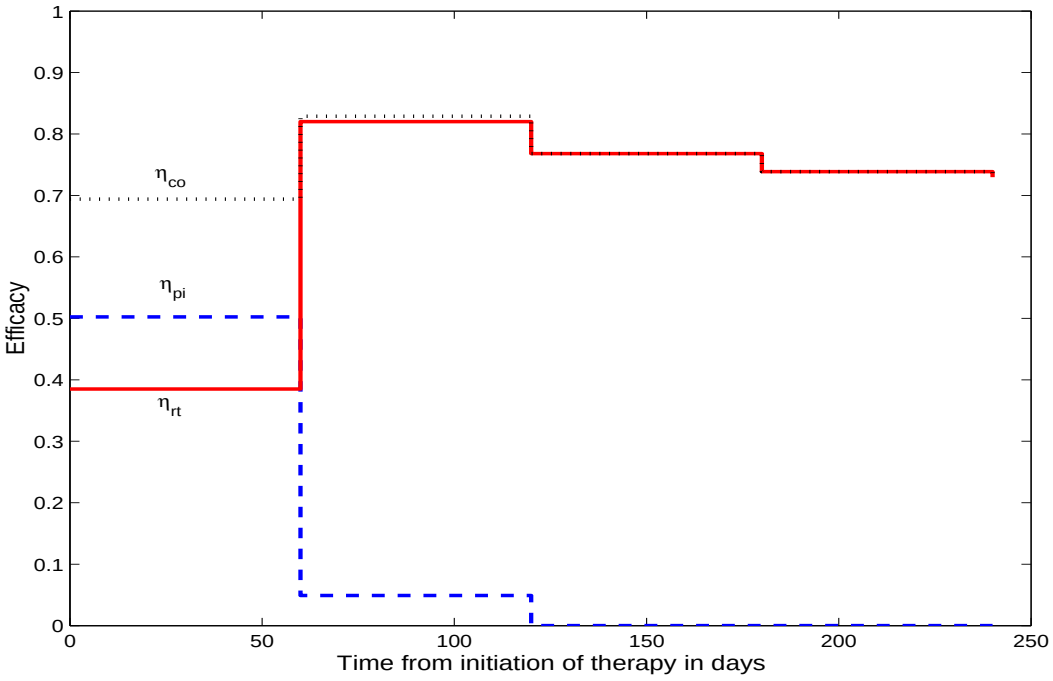
Figure 5.14(a) shows the resulting viral load and figure 5.14(b) the $CD4^+$ T cell outputs for the dosage schedules in figures 5.13(a) and 5.13(b), respectively.

5.6.6 Conclusions

Now that virus eradication does not seem attainable, the focus has shifted to striking a balance between adequately suppressive therapy and toxicity reduction. Therefore, a logical way to minimize cumulative toxicities is to start therapy with a low dosage (take advantage of the transient undershoot when dynamics are perturbed) and sequentially increase the dosage to further suppress and maintain the viral load below detectable levels. This results in a sequential perturbation approach to therapy. The eventual dosage required to keep the viral load below detectable levels will be high. However, once viral load suppression is attained, other strategies like Structured Treatment Interruptions

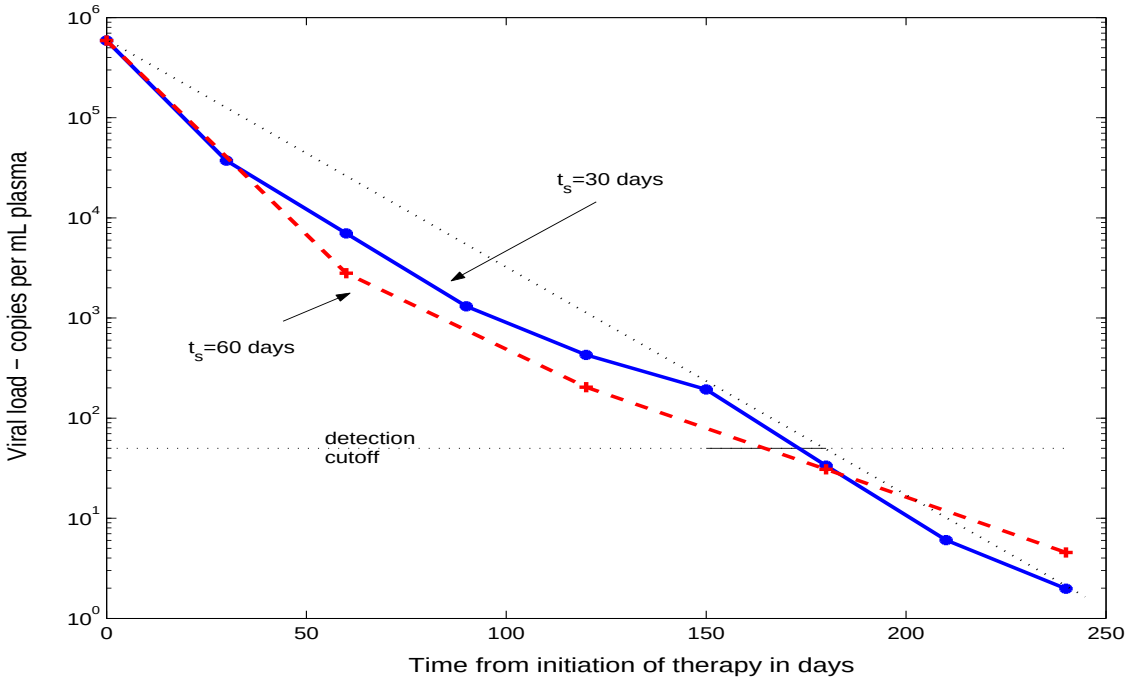


(a)

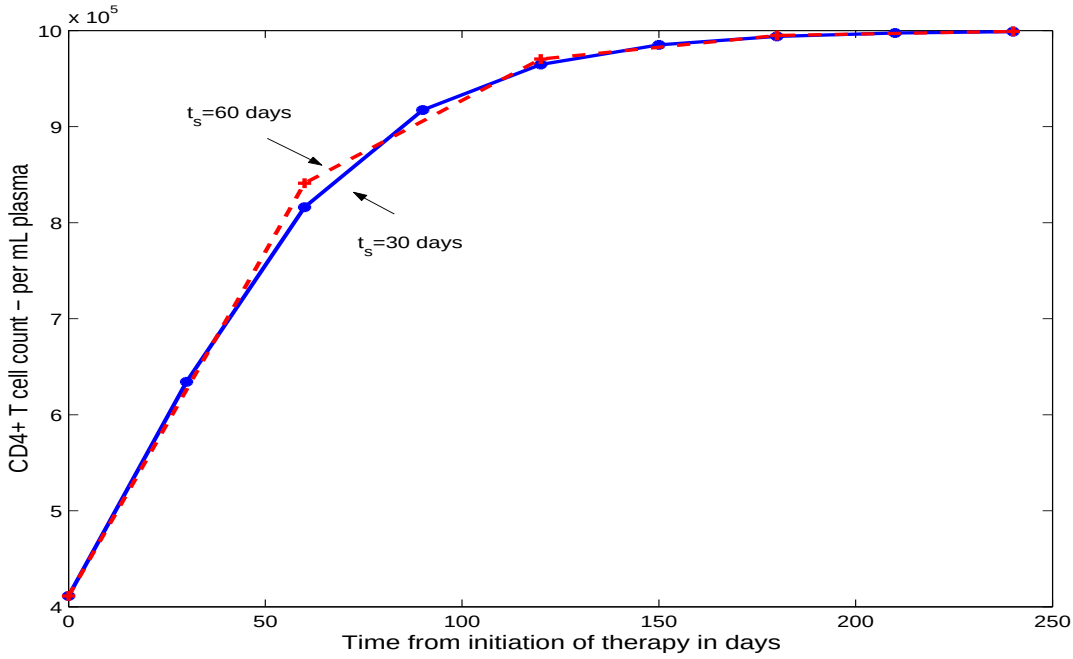


(b)

Figure 5.13: Control input sequences for combined therapy with no therapeutic range constraint when sampling interval is increased. η_{rt} : RTI efficacy; η_{pi} : PI efficacy; η_{co} : combined efficacy. (a) t_s=30 days (b) t_s=60 days.



(a)



(b)

Figure 5.14: System's outputs for combined therapy with no therapeutic range constraint when sampling interval is increased. ts : Sampling interval. Control input sequences are in figure 5.13. (a) Viral load. (b) $CD4^+$ T cell count.

can be employed to further reduce total drug use.

Firstly, the study suggests that the selection of both the starting and the subsequent regimens will depend on the frequency at which the viral load is sampled. The initial dosage efficacy also depends on the sampling interval, and the prescribed regimen is more inclined towards a fixed dosage regimen as the sampling interval increases. And as expected, the sampling interval is parameter dependent and will vary from one individual to the other. This puts emphasis on the need for more frequent sampling.

Secondly, for mono class therapy, the eventual protease inhibitor efficacy is lower than that of the reverse transcriptase inhibitors. This could explain why for combined therapy, protease inhibitors are preferred for the eventual suppression of the viral load, even though the initial regimen was more inclined towards the use of reverse transcriptase inhibitors.

When both drug options are available for therapy, there is a distinct switch from one class of antiretroviral agent to the other during treatment. The underlying dynamics, or criteria for switching the regimen basis from one class to the other, and the eventual elimination of one class from the regimen in some cases, requires further investigation. Further investigation could add some insight into the 'what to start with, when and what to change to' questions that arise when selecting regimens.

The issue is whether such a dosage scheme can be practically implemented. Given that an individual's model parameter estimates can be obtained, the results from this study indicate that if the sampling interval is reasonable, which it is in this case, then such a dosage scheme can be practical. However, even though the models predict that any drug efficacy can result in some degree of viral load reduction, one needs to consider or investigate the clinical implications of under dosing, or more generally, the implications of not taking drugs as prescribed.

5.7 Interruptible Drug Dosage Design

The preceding section 5.6 has shown how viral load suppression can be attained with a reduction in total drug exposure. This consequently minimizes cumulative toxicities and the cost of therapy. The approach that was presented also eliminated the problem of viral load rebounds after suppression has been attained. However, as pointed out, the eventual dosage efficacy required to keep the viral load suppressed at below detectable levels will generally be high. This eventually high drug efficacy reintroduces the toxicity problem. Other strategies that can further reduce drug exposure, once suppression is attained and maintained, need to be employed. One such strategy that has been tried

out in numerous clinical trials is structured treatment interruptions - STI.

As explained before in section 2.4, STI for reducing total drug exposure and invoking immunologic control of the virus, has been shown to have more success if HAART was initiated while the patient was in the acute infection stage, than the chronic infection stage. However, the acute infection stage does not last long enough, and most HIV infected persons are in the chronic infection stage. Furthermore, virus suppression with HAART, especially at the chronic infection stage, does not necessarily imply a reconstitution of the immune system. There is therefore, a need to employ other therapeutic options that will slow down viral load rebound and/or $CD4^+$ T cell decline during HAART interruption for the chronically infected individual.

This section presents a model based approach to STI protocol design for patients who initiate therapy during the chronic infection stage. For these patients, the purpose of treatment interruptions is primarily to reduce the time on HAART and the related toxicity. This reduction in total drug exposure should be attained while keeping the viral load either below the level of detection, or below some preset viral load cut off value.

The resulting STI schedules will be compared with STI protocols that have been tried out in clinical trials. The results from this section will therefore, be interpreted in the context of these trials, the objective being to highlight the strengths and weaknesses of those trials. This section also shows how, if need be, the duration of the STI cycles can be adjusted by $CD4^+$ T cell specific therapies, as well as how these $CD4^+$ T cell specific therapies can be used to make a patient conform to pre-arranged STI schedules.

5.7.1 Bottlenecks and Advances in STI Protocol Design

Clinical trials have been conducted in an attempt to determine the immunological and virological benefits of STI for patients with chronic HIV infection, as well as determine the best protocol. However, coming up with an STI protocol that will benefit most patients is still ellusive. It is clear from almost all the STI trial that have been carried out to date, that STI protocols with fixed or predetermined on/off periods are not for every body.

Bottlenecks in STI protocol design, generally can be summarized as follows:

- The viral load rebound and T decline rates during treatment interruption will vary from one individual to the other [15, 137, 144]. In essence, there is variability in response between individuals within a trial, as well as between individuals in trials with similar protocols. In clinical trial settings where poor adherence has been

ruled out, this variation in response is primarily due to inter-individual variations in viral and cell parameters, as well as the individual's stage of infection, and not necessarily on the stimulation of HIV specific immunity [143].

- Most assays used in clinical practice have a detection cutoff limit of 50 copies per mL of plasma. This makes timing viral load rebounds and deriving a suitable STI schedule problematic because the virus dynamics under consideration, occur at viral loads that are below the level of detection. The current situation is such that, attempts are being made to maintain a variable at a level that can not be readily measured. In other words, the maximality of virus suppression below the detection cutoff is unknown because there usually is no model that one can use as a guide. This is problematic because these unknown initial system conditions when therapy is interrupted, do influence its transient response.
- In the absence of a model and parameters, how then can a practitioner pre-empt the viral load response when HAART in interrupted? This means that, in order to understand the underlying viral dynamics below the level of detection, one will have to use viral load measurements that are obtainable only when the viral load has rebounded to above detectable levels. In this case, how then is the practitioner supposed to keep the viral load below detection and simultaneously collect data? This problem has in many STI trial instances, made it necessary for viral load cut off points to be increased from 50 copies per mL of plasma, to 200, 500 and even 5000 copies per mL.

In spite of all the problems with STI protocols, significant advances have been made:

- It has long been observed that re-initiation of HAART after an interruption results in re-suppression of the viral load to below detectable levels (for patients who had prior viral load suppression before treatment interruption) [154]. This observation has since been supported by almost all the clinical trials referenced above.
- As supported by the outcome of many clinical trials, STI does not necessarily facilitate or accelerate the selection of resistant mutants [156], though there are some instances of the emergence of drug resistance during repeated STI [128, 129, 130, 131].
- Even though coming up with an STI protocol that will benefit most patients is still illusive, there are compelling reasons to believe that short cycle STI schedules have more success than schedules with long cycles.

5.7.2 Strategy

Having highlighted the bottlenecks in STI protocol design, the strategy then, is as follows: Determine the conditions for interrupting and resuming therapy. The next step is then to identify STI schedule options that would work for the chronically infected individual, under the stipulated conditions.
For each schedule option, the following should be evaluated or considered:

- Percentage reduction in total drug exposure when compared with continuous HAART.
- The ease of implementation, or lack thereof.
- Ways of improving the appeal of the schedule.

Finally, ways of making an individual conform to pre-defined STI schedules, as is the case in clinical trials, needs to be investigated.

Assumptions:

- The assumption in most clinical trials is usually that the individual has been on HAART long enough and has a record of sustained viral load suppression below 50 copies per mL of plasma. This assumption is also made in this thesis.
- When HAART is re-initiated after an interruption, it is also assumed that the patient will resume the regimen they were previously on. In this case, this will be the RTI and PI containing HAART regimen that was derived in section 5.6.
- By interrupting HAART, it is understood to imply the discontinuation of all replication cycle based drugs (RTI and PI) in the regimen.
- Also, by interrupting IBT, it is understood to imply the discontinuation of all immune based drugs in the regimen.

This section is therefore, a logical continuation of the preceding section. So HAART is considered to have been initiated and day 300 from initial infection for a period of 300 days. The end point data, that is, the combined RTI and PI drug efficacy and the system states at day 600 are presented at the bottom of Table 5.7.

5.7.3 Off/On HAART: Getting the Timing Right

When HAART is discontinued, the viral load will in most cases rebound and target cell counts decline to pre-HAART levels [152], as illustrated in figure 5.15. The objective then, is to maintain the viral load between an upper V_{max} and lower V_{min} cut off limit by Off/On control.

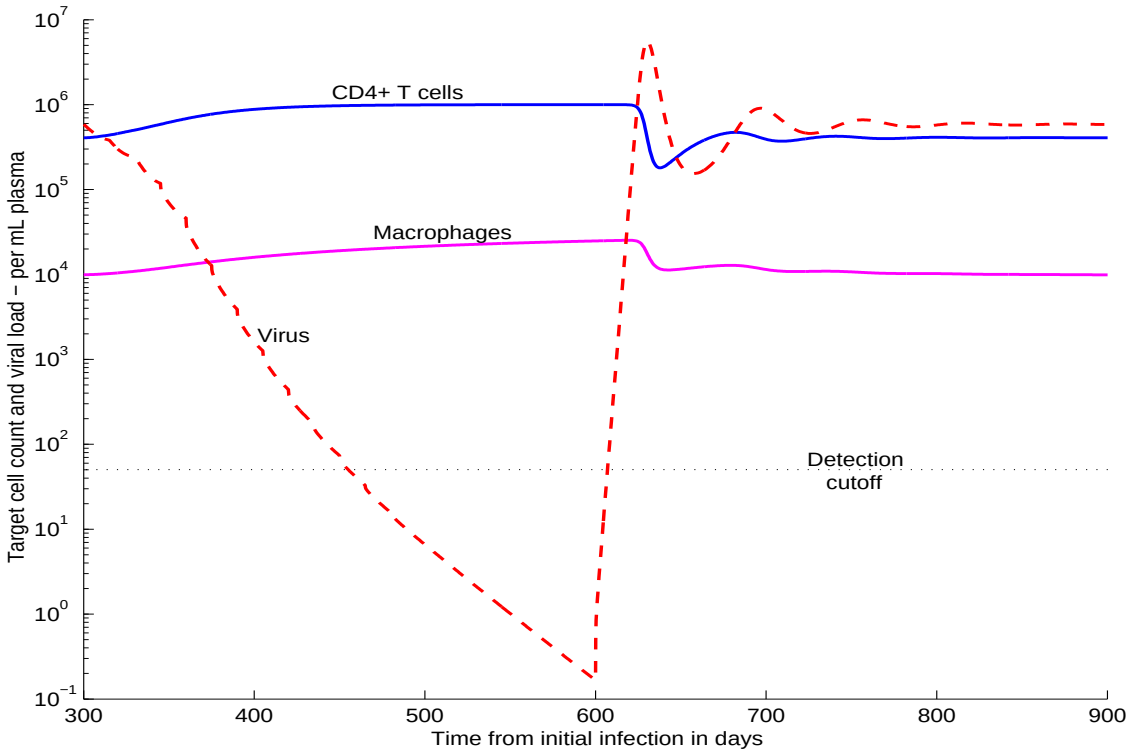


Figure 5.15: Viral load rebound and cell loss to pre-HAART values when therapy is discontinued. HAART was On from day 300 for 300 days, then discontinued at day 600.

Upper Viral Load Limit:

Choices for the upper viral load limit V_{max}, was based on similar choices that were made in various clinical trials, where V_{max} was selected as, for example, 200 copies per mL plasma [139] and even as high as 5000 copies per mL plasma [135].

In this thesis, the upper viral load limit varied from V_{max} =50 copies per mL plasma, to V_{max} =200 copies per mL plasma, and to V_{max} =500 copies per mL plasma.
The appropriateness of imposing each upper viral limit will be assessed.

Lower Viral Load Limit:

The lower limit, V_{min} will simply be taken as the viral load measurement when STI is initiated (pre-STI). This is the viral load measurement at day 600.

CD4$^+$ T cells decline and rebound:

It is undesirable for the CD4$^+$ T cells count to drop during HAART interruption, by more than half of the cell gain that has been attained prior to STI. That is CD4$^+$ T cells count should not drop below $\hat{T} = \bar{T} + \frac{T_H - \bar{T}}{2}$, where T_H is the pre-STI CD4$^+$ T cells count.

Table 5.7: Summary of constraints for interruptible dosage scheduling

Type	Lower and Upper Limits	Notes
HAART Efficacy	$\eta_{co} = 0.82$	Fixed. Pre-determined from section 5.6.
IBT Efficacy	$0 \leq \eta_{ps} < 1$ $0 \leq \eta_{da} < 1$	Varies. Varies.
u_H	0 1	HAART interrupted. HAART resumed.
u_I	0 1	IBT interrupted. IBT resumed.
$CD4^+$ T cell decline and rebound	$\hat{T} \leq T \leq T(0)$	Maintain asymptomatic status. $\hat{T} = \bar{T} + \frac{\bar{T}+T_H}{2}$
Viral load	$V_{iH} < V_i \leq V_{max}$	V_{max} : [50 200 500].

Pre-HAART conditions	
$\bar{T}$	408 cells per μL plasma
Pre-STI conditions	
T_H	1000 cells per μL plasma
V_{iH}	0.17 copies per mL plasma
η_{co}	0.82

The assumption is that $\hat{T} > 350$ cells per μL of plasma, or equivalently $\hat{T} > 3.5\text{x}10^5$ cells per mL of plasma. The intention is to keep the patient clearly in the asymptomatic stage.
The $CD4^+$ T cells are not expected to rebound to higher than pre-infection values.

Off/On Control Sequence:
The criteria for switching OFF and ON HAART, or the control sequence rather, can be summarized as follows:

- Interrupt HAART (u_H=0) until $V_i \geq V_{max}$ [50, 200 or 500 mL^{-1}]
  or $T \leq \hat{T}$
- Resume and maintain HAART (u_H=1) until $V_i \leq V_{min}$
- Repeat Off/On cycle for 300 days

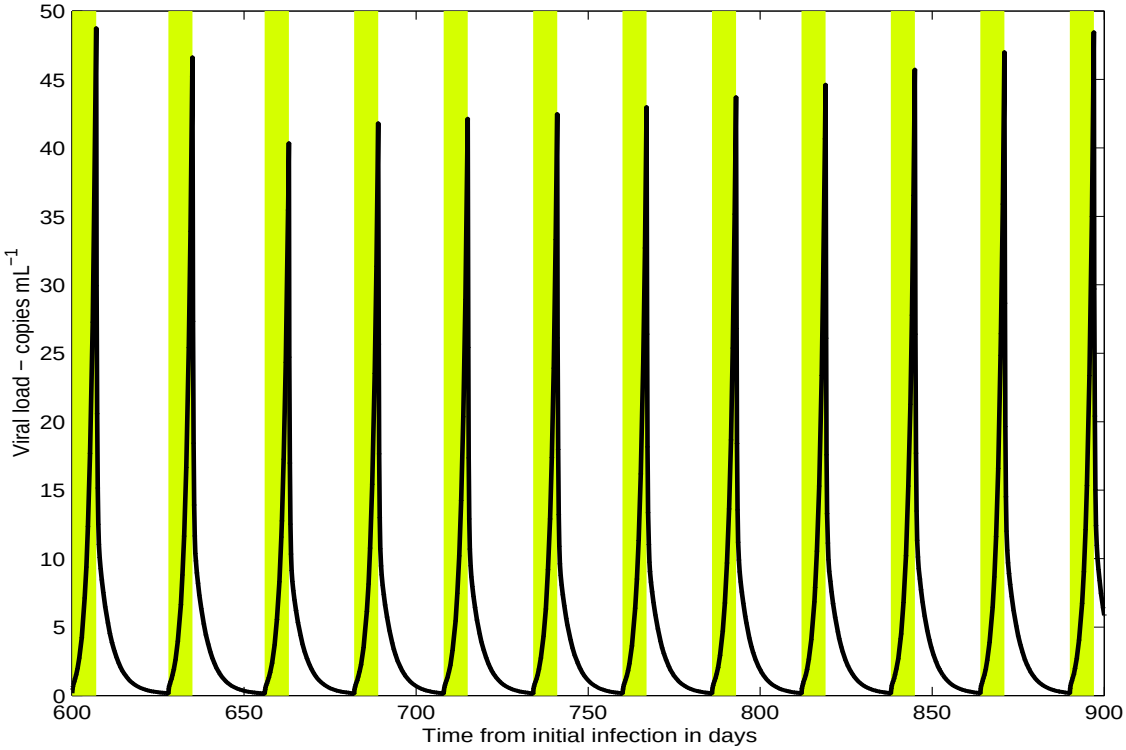


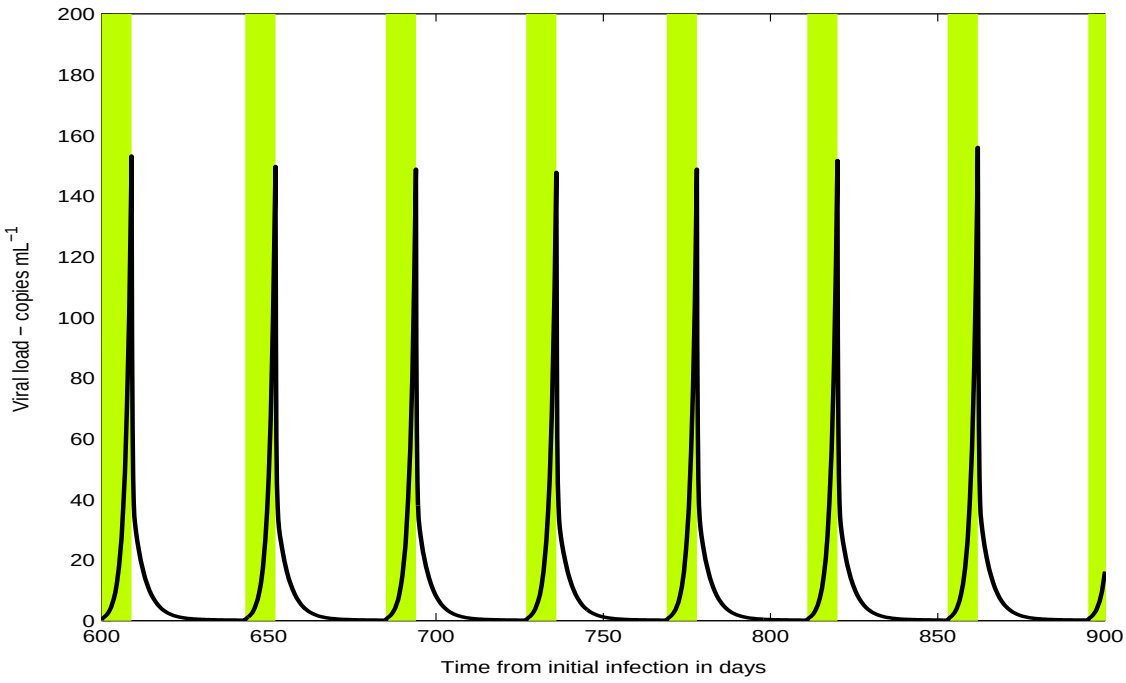
Figure 5.16: Getting the timing right. $V_{max} = 50$. Pre-STI conditions are presented in Table 5.7. Shaded areas indicate Off HAART periods. $\eta_{co} = 0.82$.

u_H is as presented by Σ_{Eco} (5.1) in section 5.1. The applicable constraints, together with the pre-STI viral load and CD4$^+$ T cell count are summarized in Table 5.7.

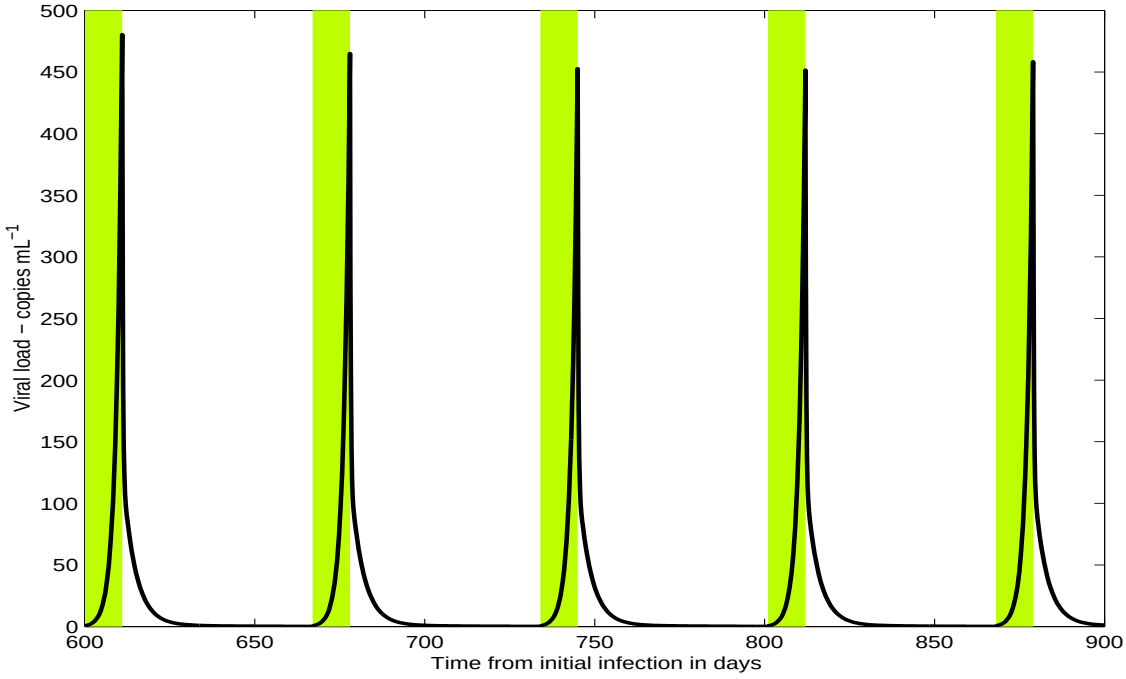
5.7.4 Off/On HAART: Results

Off/On HAART sequence was first derived for the 50 copies per mL of plasma upper viral load cutoff limit $V_{max} = 50$, then for $V_{max} = 200$ and $V_{max} = 500$. The assumption that only replication cycle based HAART is used still applies. Therapy is considered to have failed if the viral load can not be kept below the respective V_{max}.

Figure 5.16 shows the viral load response for $V_{max} = 50$ copies per mL plasma. The shaded area indicates Off HAART periods. The resulting STI input sequence starts with an Off/On cycle with 7 days Off HAART/21 days On HAART (7/21), for the first two cycles. Thereafter, the On HAART period decreases slightly to 19 days, while the time Off HAART remains as 7 days. But generally, the resulting input sequence approximates the fixed 7/21 Off/On schedule that was used in the clinical trial by Lori *et al* [14]. For practical purposes, it is better to have STI schedules with fixed Off/On cycles, as they are easier to administer.



(a)



(b)

Figure 5.17: Getting the timing right. Viral load response with varying V_{max}. Pre-STI conditions are presented in Table 5.7. $\eta_{co} = 0.82$. (a) V_{max}=200; (b) V_{max}=500

Table 5.8: Getting the timing right with varying viral load upper cutoff limits

Viral load cutoff	**Off/On HAART**	**Drug Reduction***
$V_{max} = 50$ mL^{-1}	7/21 days (28 day cycle)	25%
$V_{max} = 200$ mL^{-1}	9/34 days (43 day cycle)	20.9%
$V_{max} = 500$ mL^{-1}	11/56 days (67 day cycle)	16.4%

*Percentage reduction in total drug exposure when compared with continuous HAART.

Similarly, figure 5.17 shows the viral load response for $V_{max} = 200$ and $V_{max} = 500$ copies per mL plasma. For $V_{max} = 200$ (figure 5.17(a)), the derived STI input sequence starts with an Off/On cycle with 9 days Off HAART/34 days On HAART (9/34), for the first cycle only, which is followed by a slightly shorter 9/33 day Off/On cycle for the subsequent cycles. The resulting input sequence can be approximated very well by the fixed 9/34 day Off/On schedule.

For $V_{max} = 500$ (figure 5.17(b)), the derived STI input sequence has fixed 11 days Off HAART/56 days On HAART (11/56) cycles. All the resulting fixed duration Off/On schedules that can be used to approximate the derived input sequences and the percentage reduction in total drug intake are summarized in Table 5.8. Figure 5.18 shows the resulting viral load response for the various upper cut off limits when the approximate fixed duration Off/On schedules are used.

It seems, from figure 5.18, figure 5.17 and Table 5.8, that there is no obvious correlation between the Off HAART period and the On HAART period. However, it is clear that it takes longer to re-suppress the viral load after an interruption, than it does for the viral load to rebound during an interruption. This is reflected by the systematically longer On HAART periods with relatively shorter Off HAART periods.

This simple fact, seems to suggest that in the absence of immune control of the virus, as is usually the case in chronic infection, STI clinical trials that have been carried out with equal Off/On periods, were designed to fail. Examples of such designed to fail trials are the 30 days On HAART/30 days Off HAART (30/30) trial by Ortiz *et al* [142] and the 7 days On HAART/7 days Off HAART (7/7) trial by Ananworanich *et al* [20]. Seemingly unexplainable, is the positive outcome of the 7/7 trial by Dybul *et al* in which the participants attained viral load control. However, analysis of the trial's protocol shows that the HAART regimen that was used during STI was more potent that the one originally used to suppress the viral load. Refer to Tables 2.5, 2.6 and 2.4, respectively in section 2.4 for a summary of these trials.

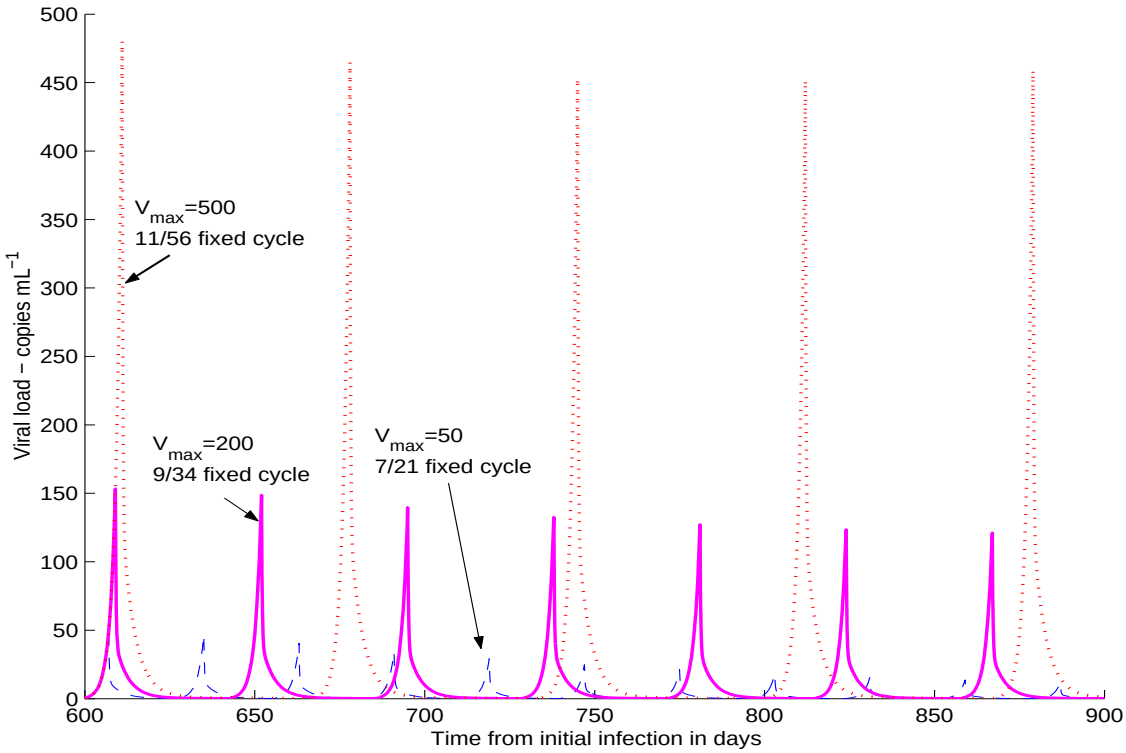


Figure 5.18: Getting the timing right. Viral load response with varying V_{max}. Pre-STI conditions are presented in Table 5.7. $\eta_{co} = 0.82$.

The reason for failure is that On HAART period, relative to the Off HAART period, was not long enough to adequately re-suppress the virus when therapy is re-initiated after an interruption. This effect is illustrated in figure 5.19.

Results also show that the resulting input sequences for lower V_{max} have shorter Off/On cycles than the derived input sequences with higher V_{max}. For example, setting $V_{max} = 50$ copies per mL plasma results in a 28 day (7/21) Off/On cycle, while setting $V_{max} = 500$ copies per mL plasma results in a 67 day (11/56) Off/On cycle. This can be intuitively deduced because, the higher you let the virus rebound, the longer it will take to re-suppress it, and this will result in a longer Off/On cycle.

The real advantage of setting a lower value for V_{max}, is that the resulting Off/On schedule has a relatively higher percentage reduction in total drug intake. For example, setting $V_{max} = 50$ copies per mL plasma results in at least a 25% reduction in total drug use when compared with continuous HAART, while setting $V_{max} = 500$ copies per mL plasma reduces drug exposure by only 16%. This seems to suggest that short cycle STI schedules are better than schedules with long Off/On cycles.

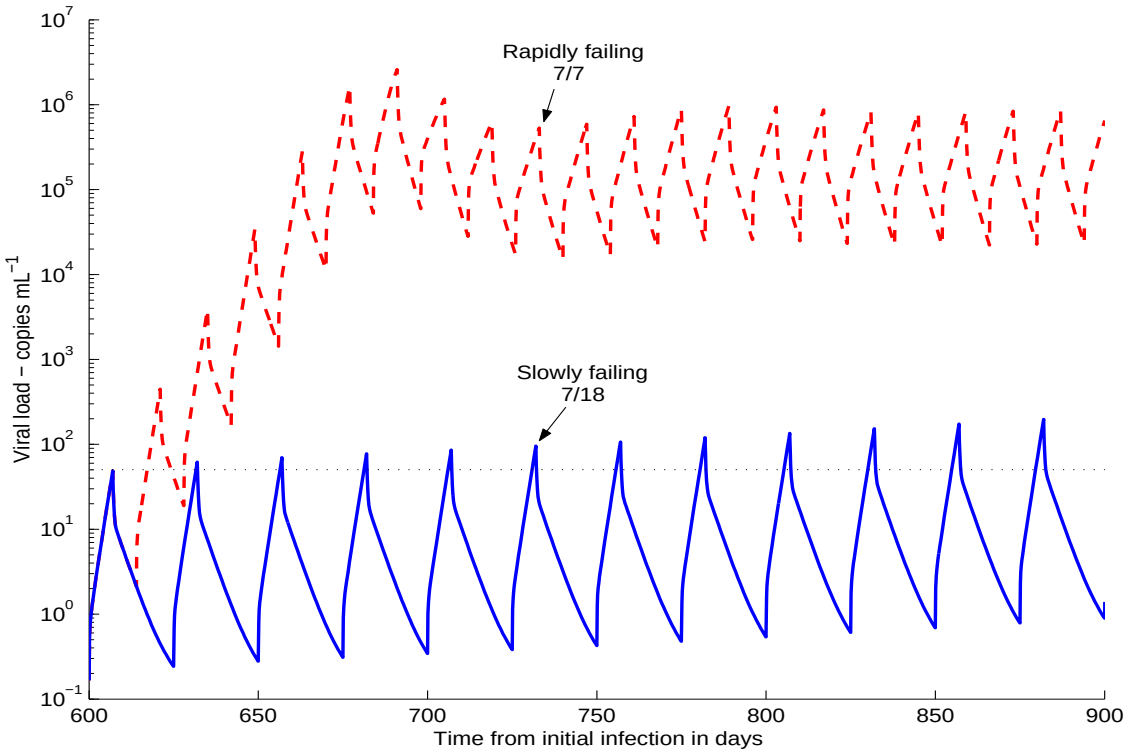


Figure 5.19: Designed to fail: Effect of failure to adequately re-suppress the viral load after an interruption. 7/7: 7 days Off HAART/7 days On HAART; 7/18: 7 days Off HAART/18 days On HAART.

As it appears, increasing the upper viral load cutoff limit, is a case of increased risk of drug resistance with no reward. However, this increase in V_{max} does not seem to have any adverse effect of the CD4$^+$ T cell count, as long as there is viral load control, as illustrated in figure 5.20.

The results suggests that an individual with a parameter set similar to the one presented in this chapter (Table 5.1) then, would have had a negative outcome in the following clinical trials:

- the 7days/7days trial by Ananworanich, *et al* [20] because the 7 days given to re-suppress the viral load would have been inadequate. This could explain why the trial's outcome was negative.
- the 2weeks/8weeks trials by Fischer, *et al* [137] and Oxenius, *et al* [143] because the 2 weeks off therapy is too long and would allow the viral load to rebound to high levels.

On a similar note, the said individual would have had a positive outcome in the following clinical trials:

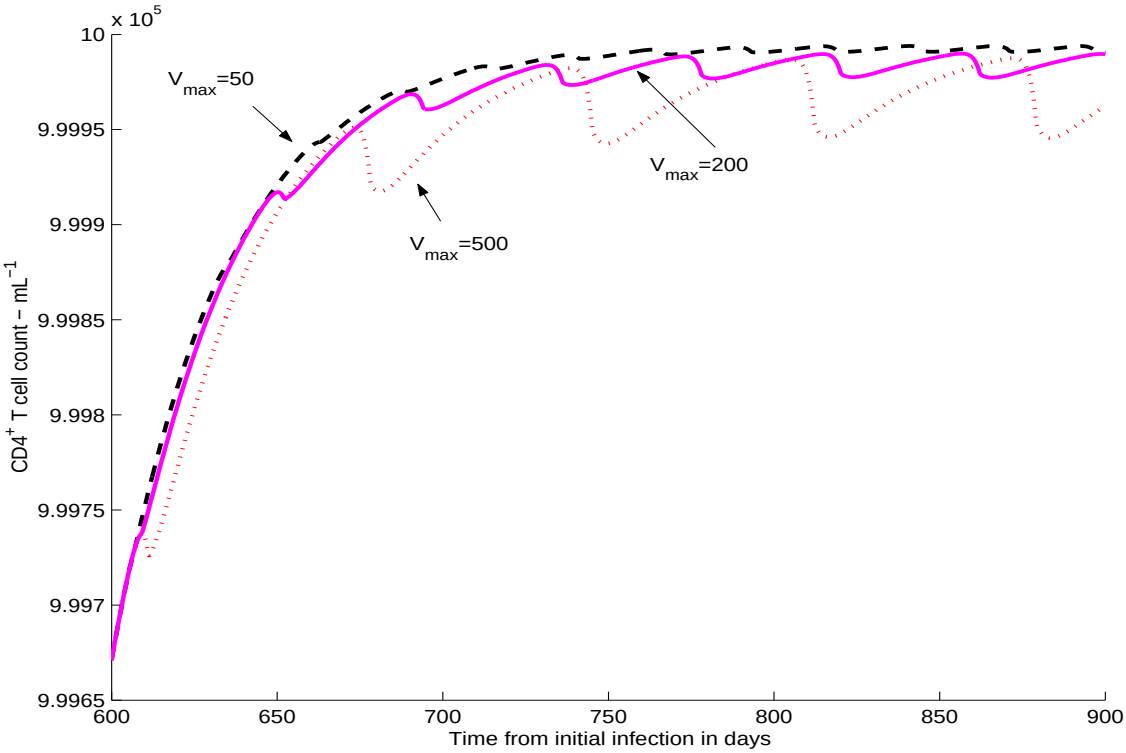


Figure 5.20: Getting the timing right. CD4$^+$ T cell response with varying V_{max}. Pre-STI conditions are presented in Table 5.7. $\eta_{co} = 0.82$.

- the 1week/3weeks trial by Lori, *et al* [14] though no conclusions about resistance can be drawn in this case.

It is acknowledged that 50, 200 and 500 are not the only viable options for V_{max}. Table 5.9 presents a summary of the derived Off/On schedules for various viral load cut off limits, and mostly for those that are below the level of detection, that is for $V_{max} < 50$ copies mL^{-1}. As expected, setting lower values for V_{max} further shortens the duration of the Off/On cycles. However, this does not imply a further reduction in total drug exposure. Care should therefore be taken when selecting the value for V_{max}.

5.7.5 Including Protease Inhibitors in the STI Regimen

As pointed out, in current practice, there is no differential measuring of neither the viral load nor the CD4$^+$ T cell count. The total viral load, which is the sum of the infectious and noninfectious virus particles is what is measured. This therefore, nullifies the effect of protease inhibitors in the regimen, and consequently, could result in over prescribed regimens. Similarly, measuring the total CD4$^+$ T cell count over estimates the availability of target cells.

Table 5.9: STI schedule options for $V_{max} < 50$ copies mL^{-1}

V_{max}	Off/On Schedule	Drug Reduction*
500 mL^{-1}	11/56 (67 day cycle)	16.4%
300 mL^{-1}	10/45 (65 day cycle)	18.2%
200 mL^{-1}	9/34 (43 day cycle)	20.9%
100 mL^{-1}	8/27 (35 day cycle)	22.9%
50 mL^{-1}	7/21 (28 day cycle)	25%
30 mL^{-1}	6/18 (24 day cycle)	25%
20 mL^{-1}	5/16 (21 day cycle)	23.8%
10 mL^{-1}	4/13 (17 day cycle)	23.5%
5 mL^{-1}	3/11 (14 day cycle)	21.4%
3 mL^{-1}	2/9 (11 day cycle)	18.2%
2 mL^{-1}	1/6 (7 day cycle)	14.3%

*Percentage reduction in total drug exposure when compared with continuous HAART.

Figure 5.21 shows the infectious and total viral load responses to the previously derived 7/21 STI schedule with PI containing regimens. The peak for the total viral load is slightly higher than that of the infectious particles. This is so because no noninfectious virus particles are produced when therapy is interrupted. So whether or not PIs are used in the regimen should not determine the failure or success of the schedule, especially so when operating at viral loads below levels of detection.

5.7.6 Immune Based Therapy to Augment HAART Interruptions

The preliminary analysis that was carried out in section 4.3 has suggested the potential use of immune based therapies - IBT, to augment HAART interruptions. This section presents simulation results when infected $CD4^+$ T cell death rates δ_l and δ_a are accelerated, and when $CD4^+$ T cell proliferation is suppressed. Such immune based therapies include the use of hydroxyurea (HU) and interluekin-2 (IL-2).

The rationale for accelerating the death rates of infected cells is clear. $CD4^+$ T cell proliferation suppressing therapies in any case, can also be considered because even though they are not capable of long term enhancement of antiviral efficacy, they do have an effect on the viral load transient response, with and without HAART.
The objective of using IBT to augment HAART interruptions could be to:

- Slow down viral load rebound when HAART is interrupted, and consequently, extend the Off HAART period. This implies alternating HAART with IBT such

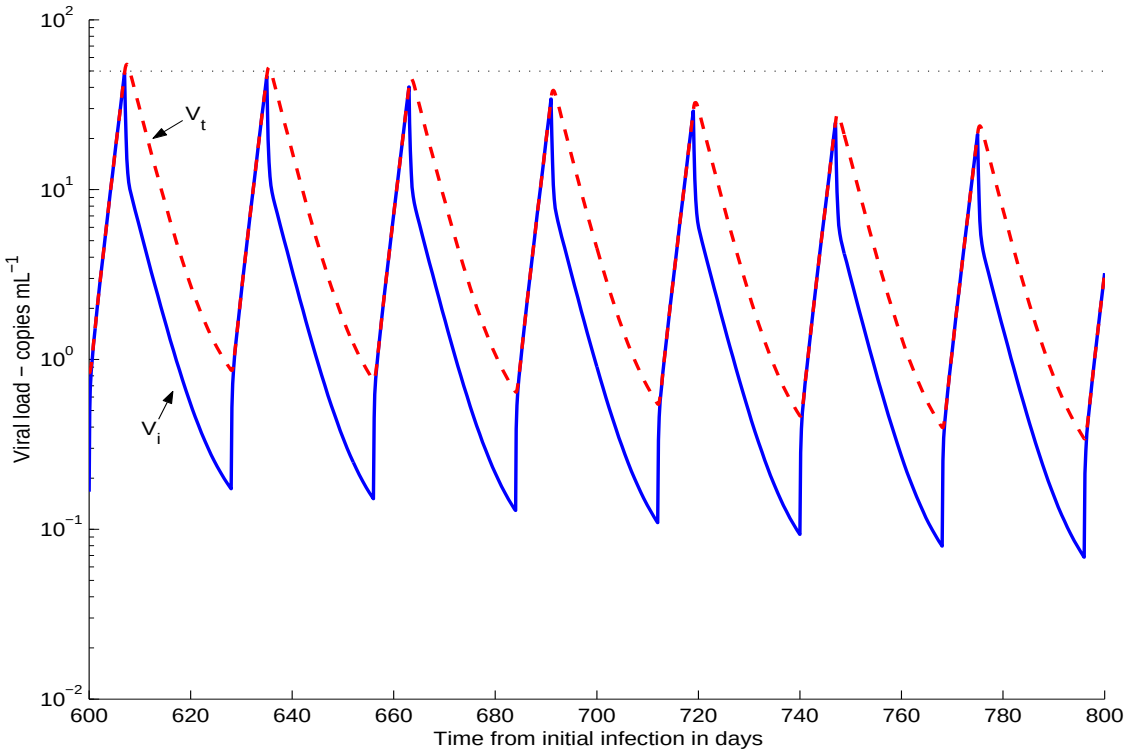


Figure 5.21: Viral load response to Off/On cycles with PI containing regimens. $\eta_{pi} = 0.8$, $\eta_{rt} = 0.1$. Pre-STI conditions are presented in Table 5.7. V_t: Total viral load. V_i: Infectious virus particles.

that, when HAART is interrupted, IBT is initiated and when HAART is resumed, IBT is interrupted. The switching is as follows :

$$(u_H = 1, u_I = 0)/(u_H = 0, u_I = 1)$$

As this entails the use of IBT during HAART interruptions, this approach will sacrifice one's drug free days, or drug holidays.

- Accelerate viral load re-suppression after HAART interruption, and consequently, reduce the On HAART period. This implies the concomitant use of HAART and IBT such that, when HAART is interrupted, IBT is also interrupted and when HAART is resumed, IBT is also resumed. The switching is as follows :

  $$(u_H = 1, u_I = 1)/(u_H = 0, u_I = 0)$$

  As this entails the use of IBT during HAART periods, this approach will preserve one's drug free days.

Only the option that preserves drug holidays will be presented here.

Even though maximal reduction of the time on therapy is desirable, the use of IBT must be limited, so as not to make an already potent regimen more toxic. The advantages

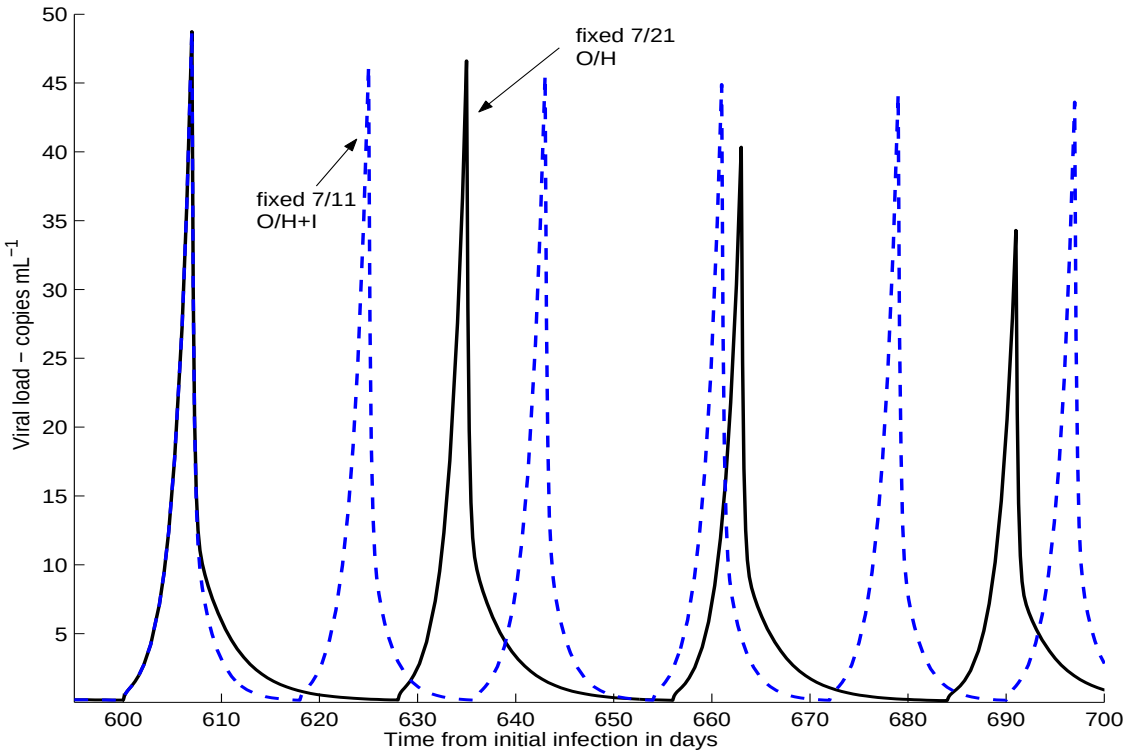


Figure 5.22: Viral load response when IBT is added to HAART when therapy is On. $\eta_{da} = 0.5$. Pre-STI conditions are as presented in Table 5.7. 7/21 O/H: 7 days Off all therapies /7 days On HAART only; 7/11 O/H+I: 7 days Off all therapies/11 days On HAART plus IBT.

and disadvantages of adding IBT to a HAART regimen that is already capable of viral load suppression should be evaluated. That is, one should asses if further reducing HAART exposure is worth the possible compounded regimen toxicity.

Figure 5.22 shows how concurrently administering HAART and IBT (cell death accelerators) can accelerate viral load suppression and consequently, reduce the time On therapy. However, one needs to weigh the advantages of reducing On therapy periods against the sacrifice of increased toxicity. The results are summarized in Table 5.10.

This approach can also be used to force conformation to STI schedules with predetermined Off/On periods.

5.7.7 Conclusions

In this section, structured treatment interruptions schedules of HAART were derived for when varying upper viral load limits for the virus rebound were set. The intention was to reduce drug exposure (when compared with continuous HAART), the associated toxicity and cost of therapy. The appropriateness of imposing each viral load upper limit

Table 5.10: STI schedules with HAART plus IBT. $V_{max} = 50$ copies mL^{-1}.

Schedule	**Cycle**	**Drug Reduction***
Off/HAART	7/21 days (28 day cycle)	25%
Off/(HAART+IBT)	7/11 days (18 day cycle)	39%†

*Percentage reduction in total drug exposure when compared with continuous HAART.
†Increased IBT exposure.

was assessed, and the derived Off/On HAART schedules were compared with schedules that have been used in some STI clinical trials.

The preliminary analysis that was carried out in section 4.3 has suggested the potential use of immune based therapies to augment HAART interruptions. STI schedules with shorter cycles were derived when IBT was added to HAART when therapy was resumed after an interruption.

The following conclusions can be drawn for this section:

1. It is clear that it takes longer to re-suppress the viral load with HAART after an interruption, than it does for the viral load to rebound during an interruption. This means that for STI schedule design with HAART, the Off HAART period will always be shorter than the On HAART period, for all viral load upper cutoff limits.
2. STI schedules with equal Off/On periods that have been used in some clinical trials were designed to fail.
3. Results show that setting lower viral load limits result in Off/On schedules with shorter cycles. This can be intuitively deduced because, the higher you let the virus rebound, the longer it will take to re-suppress it, and this will result in long Off/On cycles.
4. It can also be seen that these short Off/On cycle STI schedules have a relatively higher percentage reduction in total drug intake. As it appears, increasing the viral load cutoff limit is a case of increased risk of drug resistance with no reward.
5. From a total drug intake perspective, selecting a viral load upper limit $V_{max} = 50$ copies per mL of plasma produced the best STI schedule, as it resulted in the highest reduction in total drug intake of at least 25%. Another advantage with this schedule is that it can be approximate very well by a 28 day (1 month) cycle with 7 days (1 week) Off HAART and 21 days (3 weeks) On HAART. This Off/On cycle can be easily implemented as it is convenient to use.

6. The use of IBT to augment HAART interruptions that was presented in this section was to accelerate viral load re-suppression after an interruption, and consequently reduce the time on therapy. This resulted in shorter Off/On cycles and further reduction in HAART exposure. However the resulting regimen was more potent.
7. The approach presented above can be used to force conformation to STI schedules with pre-determined Off/On periods. Similar results can be achieved by selecting a more potent HAART regimen for use during STI.

5.8 Chapter Summary

It is clear that the rational sequencing of antiretroviral drugs is needed. There is also a pressing need to individualize therapy and schedule drugs depending on the individual's response. Furthermore, drug dosages should be within the individuals operating or therapeutic range, and this operating efficacy range for an individual can be determined if the model parameters are known.

Antiretroviral drugs have a narrow therapeutic range. So while it is desirable to administer drugs within the prescribed therapeutic range, the fact that this range is an average taken over many clinical trials should be taken into consideration. Furthermore, and as pointed out by the guidelines, a regimen that is not effective or is considered alternative for a certain individual can turn out to be another individuals preferred and effective regimen. Better response to therapy can therefore be attained by true individualization of therapy, where the underlying causes for variability in response are factored in when selecting a regimen.

It would not be practical to determine the dosage to end point efficacy for each drug for each individual because the co-administration of multiple drugs leads to drug-drug interactions. A more viable option would be to determine the dosage to end point efficacy for a particular regimen, when administered to an individual. That is, defining the dosage to end point efficacy relationship for the administered regimen would be a step in that direction.

This study has provided some insights on the way antiretroviral agents could be administered. Firstly, the study suggests that the selection of both the starting and the subsequent regimens will depend on the frequency at which the viral load is sampled. The initial dosage efficacy also depends on the sampling interval, and the prescribed regimen is more inclined towards a fixed dosage regimen as the sampling interval increases. And as expected, the sampling interval is parameter dependent and will vary from one individual

to the other. Secondly, for mono class therapy, the eventual protease inhibitor efficacy is lower than that of the reverse transcriptase inhibitors. When both drug options are available for therapy, there is a distinct switch from one class of antiretroviral agent to the other during treatment.

The study did not address the issue of stability for the model predictive controller. Furthermore, for the linearized internal model, the exact linearization about a non-equilibrium system trajectory is a time varying linear model, since the state does not remain constant [186]. Even though one does not have to re-linearize at each sampling instance, re-linearizing often enough could improve the performance of the controller.

The issue was whether such a dosage scheme can be practically implemented. Given that an individual's model parameter estimates can be obtained, the results from this study indicate that if the sampling interval is reasonable, which it is in this case, then such a dosage scheme can be practical.

The eventual dosage required to keep the viral load below detectable levels will be high. However, once viral load suppression is attained, other strategies like Structured Treatment Interruptions were employed to further reduce total drug use.

Setting lower viral load cutoff limits resulted with STI input sequences with shorter Off/On cycles and relatively higher percentage reduction in total drug intake. As it appears, increasing the viral load cutoff limit is a case of increased risk of drug resistance with no reward.

The viral load rebound when HAART is interrupted may be too rapid for some individuals. This may render STI of HAART an inviable option. A possible way to prolong Off HAART periods for these individuals is to alternate HAART with immune based therapy. However, this could sacrifice the drug free days or holidays that one had. A more viable option is the co-administration of IBT with HAART when therapy is resumed after an interruption. This further reduced the duration of the Off/On cycle, further reduced HAART exposure and preserved one's drug free days.