An Introduction to Mathematical Physiology

Lecturer: Ian Griffiths

Authors: S. J. Chapman, A. C. Fowler, R. Hinch and S. L. Waters

Contents

Chapter 1

Enzyme kinetics

Many of the processes which occur in the human body are enabled by chemical reactions, and many of these involve large complex molecules called *proteins*. Proteins consist of chains of *polypeptides*, which are themselves formed of chains of *amino acids*. A particular type of protein is the *enzyme*, which takes part in reactions by acting as a *catalyst*. A catalyst is a molecule that helps to convert other molecules (called *substrates*) into products, but is not itself used up in the reaction. An example is the conversion of ATP to ADP during glycolysis, to which we return later in the chapter. Enzymes are extremely efficient catalysts, often giving increases in the rate of reaction by a factor of 10^7 or more. Just as importantly, they are highly specific, catalysing only one reaction of one specific substrate or family of substrates. Enzymes are also regulated, responding to a complicated network of positive and negative feedback mechanisms, thus allowing the rate of reaction to be precisely controlled.

An enzyme works by lowering the activation energy of the reaction. This it may do by a number of different mechanisms. For example, it may aid in overcoming the electrostatic repulsion of like-charged molecules, or it may help in breaking existing bonds within the substrate.

1.1 Law of mass action

The way in which chemical reactions and the consequent evolving concentrations of their reactants are quantified is by the *law of mass action*. Suppose that two chemicals, A and B say, react together on collision to produce the product C. The *law of mass action* states that the rate at which the reaction takes place is proportional to the number of sufficiently energetic collisions between the molecules A and B per unit time, which in turn is taken to be proportional to the concentrations of A and B. Thus we write

$$\mathrm{A} + \mathrm{B} \xrightarrow{k} \mathrm{C} \tag{1.1}$$

and, taking A, B, C to be the concentrations of A, B and C respectively,

$$\frac{dC}{dt} = kAB. \tag{1.2}$$

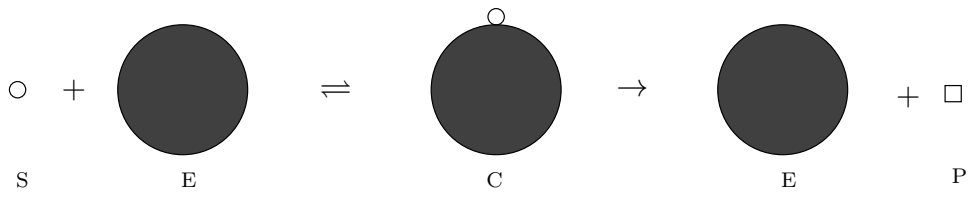

Figure 1.1: Schematic representation of the Michaelis-Menten reaction scheme.

The constant of proportionality k is known as the *rate constant* for the reaction, and depends on the geometrical shapes and sizes of the reactant molecules or ions, and on the temperature of the mixture.

While the law of mass action is extremely useful, there are many reactions to which it cannot be applied, usually because the reaction proceeds by a complex mechanism involving many elementary steps of the form (1.1). Often for biochemical reactions things are further complicated by the fact that many of the intermediate steps are unknown.

1.2 Michaelis-Menten kinetics

A simple model for the action of an enzyme on a single substrate was formulated by Michaelis and Menten in 1913. They proposed that the reaction proceeds in two steps, as shown in figure 1.1. Firstly, the enzyme E may bind with the substrate S to form a complex C. Secondly, the complex C may break down into the product P, releasing the enzyme at the same time. The basic Michaelis-Menten reaction scheme is

$$\mathrm{S} + \mathrm{E} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{C} \overset{k_2}{\rightarrow} \mathrm{E} + \mathrm{P}. \tag{1.3}$$

Applying the law of mass action to each of the steps separately, with S = concentration of S, etc., gives

$$\begin{aligned} \frac{dS}{dt} &= k_{-1}C - k_1SE, \\ \frac{dE}{dt} &= (k_{-1} + k_2)C - k_1SE, \\ \frac{dC}{dt} &= k_1SE - (k_2 + k_{-1})C, \\ \frac{dP}{dt} &= k_2C. \end{aligned} \tag{1.4}$$

In writing down equations such as (1.4) (and indeed (1.2)) we are assuming that the medium in which the reaction is taking place is well-stirred, so that the concentrations of reactants are uniform in space. If this were not the case then we would need to allow the concentrations to be functions of position as well as time (so that we

would have partial differential equations rather than ordinary differential equations) and we would need to take into account the diffusion and convection of reactants.

The overall reaction may be denoted as

$$\mathrm{S} \overset{r}{\to} \mathrm{P}, \tag{1.5}$$

where r denotes the overall reaction rate. Although (1.5) has the appearance of a first order reaction, the reaction rate $r = k_2 C$ is not constant (we shall see that it is effectively a function of S); this is a consequence of the fact that the overall reaction consists of a number of intermediate steps. Often biochemical reactions are modelled by the *Hill equation*

$$r = \frac{r_0 S^n}{K^n + S^n}, \tag{1.6}$$

an example being in the modelling of blood cell production. A particular example of this can be derived in models of enzyme kinetics, and in particular for those of cooperative enzymes, where the case $n > 1$ may be derived.

The set of four equations in (1.4) is nonlinear and apparently intractable, but it may be simplified by noting firstly that the equation for P uncouples from the others, i. e., it can be found by direct integration once the other three equations for E, C and S have been solved. Secondly, if we add equations $(1.4)_2$ and $(1.4)_3$, then we see that

$$E + C = E_0 \tag{1.7}$$

is constant. This expresses the conservation of enzyme, and is a consequence of the observation that the enzyme is neither produced nor consumed, the total of bound and unbound enzyme being a constant quantity. This, together with the uncoupling of P, allows the system to be reduced to just two equations for C and S:

$$\begin{aligned} \frac{dS}{dt} &= k_{-1} C - k_1 S(E_0 - C), \\ \frac{dC}{dt} &= k_1 S(E_0 - C) - (k_2 + k_{-1}) C. \end{aligned} \tag{1.8}$$

Typical initial conditions for the system (1.4) would consist of given concentrations of substrate and enzyme, and no product or complex, that is, $P = C = 0$, $S = S_0$, $E = E_0$. For (1.8), we therefore have

$$S = S_0, \quad C = 0 \quad \text{at} \quad t = 0. \tag{1.9}$$

The first step in a systematic mathematical analysis is to nondimensionalise the system. We set

$$S = S_0 s, \quad C = E_0 c, \quad t = \frac{t'}{k_1 E_0}, \tag{1.10}$$

to give

$$\frac{ds}{dt'} = -s + c(s + K' - \lambda), \tag{1.11}$$

$$\varepsilon \frac{dc}{dt'} = s - (s + K')c, \tag{1.12}$$

with initial conditions

$$s(0) = 1, \qquad c(0) = 0, \tag{1.13}$$

where

$$K' = \frac{k_{-1} + k_2}{k_1 S_0}, \qquad \lambda = \frac{k_2}{k_1 S_0}, \qquad \varepsilon = \frac{E_0}{S_0}.$$

The remarkable effectiveness of enzymes as catalysts is reflected in the extremely small concentrations needed in comparison to the substrate. Thus the parameter ε is small, typically in the range 10^{-2} to 10^{-7}. This means that the reaction (1.12) equilibriates very rapidly by comparison with (1.11), and remains near equilibrium even as s changes. Thus, to a first approximation we may take $\varepsilon \dfrac{dc}{dt'} = 0$, so that

$$c = \frac{s}{s + K'}, \tag{1.14}$$

$$\frac{ds}{dt'} = -\lambda c = -\frac{\lambda s}{s + K'}. \tag{1.15}$$

The approximation above is known as the *quasi-steady state* approximation. Note that we are not claiming that $dc/dt' = 0$; c will vary through equation (1.14) as s varies.

The quasi-steady state approximation was first used by Briggs and Haldane in 1925, and is the basis for most present-day descriptions of enzyme reactions[1]. Equation (1.15) describes the rate of transformation of the substrate, and is known as a *Michaelis-Menten law*. It has the general property of enzyme-catalysed reactions that for small concentrations of substrates the reaction rate is linear in the substrate concentration, as in the law of mass action, but for large substrate concentrations the reaction rate approaches a constant value (as the enzyme is working at full capacity). Dimensionally, under the quasi-steady state approximation, the rate of reaction $r = \dfrac{dP}{dt} = -\dfrac{dS}{dt} = -S_0 E_0 k_1 \dfrac{ds}{dt'}$, and thus

$$r = \frac{k_2 E_0 S}{K + S}, \tag{1.16}$$

where the Michaelis constant is

$$K = \frac{k_{-1} + k_2}{k_1}. \tag{1.17}$$

Although the individual reaction rate constants are difficult to measure, the ratio K can be measured relatively easily due to the observation that the initial reaction rate r_0 at $t = 0$ is given by

$$\frac{1}{r_0} = \frac{1}{k_2 E_0} + \frac{K}{k_2 E_0} \frac{1}{S_0},$$

[1]In fact Briggs and Haldane arrived at the approximation by the erroneous argument that the rates of formation and breakdown of complex were essentially equal at all times, so that dC/dt should be zero.

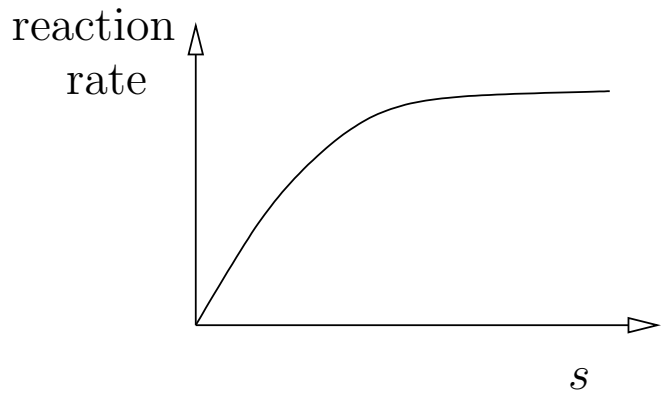

Figure 1.2: A typical Michaelis-Menten reaction rate.

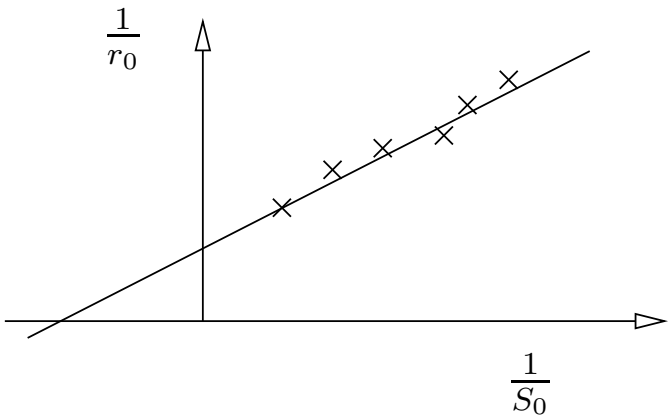

Figure 1.3: A Lineweaver-Burk plot.

so that $1/r_0$ is a linear function of $1/S_0$. Plots of $1/r_0$ against $1/S_0$ are known as *Lineweaver-Burk* plots; from them K and k_2E_0 can be found. Data such as that indicated in figure 1.3 is obtained by repeating the experiment for a number of different initial substrate concentrations, and then plotting the initial (apparent) reaction rate as a function of initial substrate concentration.

Note that with c given by (1.14),

$$c(0) = \frac{s(0)}{s(0) + K'} = \frac{1}{1 + K'} \neq 0, \tag{1.18}$$

so that the initial condition is not satisfied. There is an initial rapid transient (a boundary layer in time) when $t' = O(\varepsilon)$, during which the quasi-steady state approximation does not hold. To examine this transient we rescale the time variable by

writing $t' = \varepsilon\tau$ to give

$$\frac{ds}{d\tau} = \varepsilon\left(-s + c(s + K' - \lambda)\right), \tag{1.19}$$

$$\frac{dc}{d\tau} = s - (s + K')c. \tag{1.20}$$

Thus to leading order $ds/d\tau = 0$, so that s is constant. Since $s(0) = 1$ we therefore have $s \equiv 1$, giving

$$\frac{dc}{d\tau} = 1 - (1 + K')c, \tag{1.21}$$

so that

$$c = \frac{1}{1 + K'}\left(1 - e^{-(1+K')\tau}\right). \tag{1.22}$$

This short-time behaviour satisfies the initial condition $c(0) = 0$, and gives $c \sim 1/(1 + K')$ as we move out of the boundary layer (as $\tau \sim \infty$) in agreement with (1.18).

1.3 Inhibitors

An enzyme *inhibitor* is a substance which inhibits the catalytic action of the enzyme. Inhibition is a common feature of enzyme catalysed reactions, and is a means by which the activity of enzymes may be controlled.

There are many different types of enzyme inhibitors. Two common types of enzyme inhibitor which may be easily modelled are *competitive* inhibitors and *allosteric* inhibitors. To understand the way that an inhibitor works, and the distinction between competitive and allosteric inhibition, it is useful to recall that enzymes are usually large proteins (usually much larger than the substrate molecule), and that their catalytic properties are believed to arise from *active sites* embedded in the enzyme to which the substrate can bind. The active sites arise as a result of the three-dimensional structure of the enzyme molecule, and are highly specific, with the substrate matching the site in a "lock-and-key" fashion. However, if another molecule has a similar structure to the substrate molecule, it may also bind to the active site, preventing the binding of the substrate, and decreasing the effectivity of the enzyme. Because the inhibitor molecule binds to the active site in competition with the substrate, such inhibition is called competitive inhibition.

However, enzymes usually have many other binding sites, distinct from the active site. These other binding sites are known as *allosteric* or *regulatory* binding sites. When a molecule binds to one of these other sites it may alter the three-dimensional shape of the enzyme, thus affecting the binding of the substrate at the active site. The molecules that bind at the allosteric sites are called *effectors* or *modifiers*. They may increase the effectiveness of the active site, in which case they are called allosteric activators, or they may decrease the effectiveness, in which case they are called allosteric inhibitors.

1.3.1 Competitive inhibition

The simplest model example of a competitive inhibitor is one in which the substrate cannot bind when the inhibitor is bound to the enzyme, so that the reaction stops. Labelling the inhibitor as I, and denoting the enzyme complex with the substrate $\mathrm{C_S}$ and the enzyme complex with the inhibitor $\mathrm{C_I}$, the reaction scheme is

$$\mathrm{S + E} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{C_S} \overset{k_2}{\rightarrow} \mathrm{E + P}, \quad \mathrm{E + I} \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} \mathrm{C_I}. \tag{1.23}$$

Using the law of mass action gives

$$\begin{aligned}
\frac{dS}{dt} &= k_{-1}C_S - k_1SE, \\
\frac{dI}{dt} &= k_{-3}C_I - k_3IE, \\
\frac{dE}{dt} &= (k_{-1} + k_2)C_S - k_1SE + k_{-3}C_I - k_3IE, \\
\frac{dC_S}{dt} &= k_1SE - (k_2 + k_{-1})C_S, \\
\frac{dC_I}{dt} &= k_3IE - k_{-3}C_I, \\
\frac{dP}{dt} &= k_2C_S.
\end{aligned} \tag{1.24}$$

As before the equation for P decouples and $\dfrac{d(E + C_S + C_I)}{dt} = 0$, so that enzyme is conserved and

$$E + C_S + C_I = E_0. \tag{1.25}$$

Under the quasi-steady state approximation

$$\begin{aligned}
C_S &= \frac{K_iE_0S}{K_mI + K_iS + K_mK_i}, \\
C_I &= \frac{K_mE_0I}{K_mI + K_iS + K_mK_i},
\end{aligned} \tag{1.26}$$

where

$$K_m = \frac{k_2 + k_{-1}}{k_1}, \quad K_i = \frac{k_{-3}}{k_3}, \tag{1.27}$$

and the rate of reaction is

$$r = \frac{k_2E_0SK_i}{K_mI + K_iS + K_mK_i}. \tag{1.28}$$

The effect of the inhibitor is to increase the effective equilibrium constant of the enzyme by a factor of $1 + I/K_i$ from K_m to $K_m(1 + I/K_i)$.

1.3.2 Allosteric inhibition

If the inhibitor binds at a different site from the active site (i. e., at an allosteric site), then it is possible for the enzyme to be bound to both the inhibitor and the substrate at the same time, and there are four possible states for the enzyme, which we denote by E (unbound), C_S (bound to the substrate only), C_I (bound to the inhibitor only) and C_{IS} (bound to the substrate and the inhibitor). The reaction scheme is then

$$\begin{gathered} \mathrm{S} + \mathrm{E} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{C_S} \overset{k_2}{\rightarrow} \mathrm{E} + \mathrm{P}, \\ \\ \mathrm{E} + \mathrm{I} \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} \mathrm{C_I}, \\ \\ \mathrm{C_S} + \mathrm{I} \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} \mathrm{C_{IS}} \underset{k_1}{\overset{k_{-1}}{\rightleftharpoons}} \mathrm{C_I} + \mathrm{S}. \end{gathered} \tag{1.29}$$

The possible states of the enzyme and the rates of transition between these states are

$$\begin{array}{ccccc} \mathrm{E} & \underset{k_{-1}}{\overset{k_1 S}{\rightleftharpoons}} & \mathrm{C_S} & \overset{k_2}{\rightarrow} & \mathrm{E} + \mathrm{P} \\ {\scriptstyle k_3 I} \downarrow\uparrow {\scriptstyle k_{-3}} & & {\scriptstyle k_3 I} \downarrow\uparrow {\scriptstyle k_{-3}} & & \\ \mathrm{C_I} & \underset{k_{-1}}{\overset{k_1 S}{\rightleftharpoons}} & \mathrm{C_{IS}}\,. & & \end{array} \tag{1.30}$$

The analysis of the model now proceeds in much the same way as before. Under the quasi-steady state approximation the rate of reaction is

$$r = \left(\frac{r_{\max} K_3}{I + K_3}\right)\left(\frac{S(k_{-1} + k_3 I + k_1 S + k_{-3})}{k_1(S + K_1)^2 + (S + K_1)(k_3 I + k_{-3} + k_2) + k_2 k_{-3}/k_1}\right), \tag{1.31}$$

where $K_3 = k_{-3}/k_3$ and $K_1 = k_{-1}/k_1$.

1.4 Cooperative systems

Often the reaction rates of enzyme-catalysed reactions are more sigmoidal in nature than is predicted by the simple Michaelis-Menten law. This can result from cooperative effects.

Many enzymes have more than one active site, so that many substrate molecules can bind to the enzyme at the same time. Moreover, if a substrate molecule is bound at one active site, this can affect the binding of substrate molecules at other active sites (as in allosteric activation/inhibition).

Consider the simplest cooperative system of an enzyme with two active sites, as shown in figure 1.4. Assuming the two active sites to be identical, and denoting the

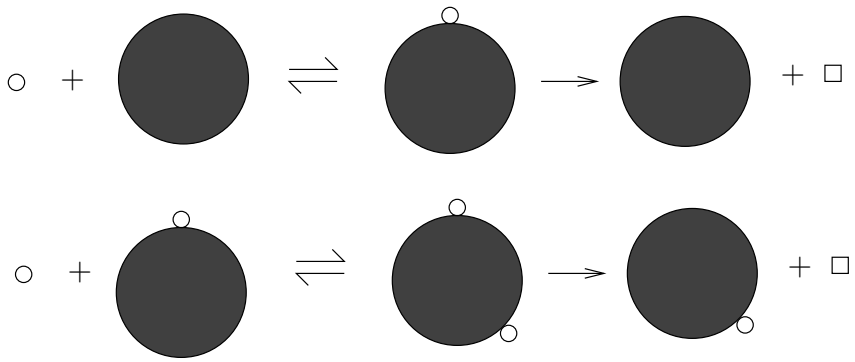

Figure 1.4: Schematic representation of an enzyme with two active sites.

enzyme complex with one substrate molecule (at either site) by C_1 and with two substrate molecules by C_2, the reaction scheme is

$$\begin{aligned} \mathrm{S} + \mathrm{E} &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{C}_1 \overset{k_2}{\to} \mathrm{E} + \mathrm{P}, \\ \mathrm{S} + \mathrm{C}_1 &\underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} \mathrm{C}_2 \overset{k_4}{\to} \mathrm{C}_1 + \mathrm{P}. \end{aligned} \tag{1.32}$$

As usual P will uncouple and enzyme will be conserved, so that $E + C_1 + C_2 = E_0$. Thus we only need to write down the remaining equations for S, C_1 and C_2. These are

$$\begin{aligned} \frac{dS}{dt} &= k_{-1}C_1 - k_1SE + k_{-3}C_2 - k_3SC_1, \\ \frac{dC_1}{dt} &= k_1SE - (k_2 + k_{-1})C_1 - k_3SC_1 + (k_4 + k_{-3})C_2, \\ \frac{dC_2}{dt} &= k_3SC_1 - (k_4 + k_{-3})C_2. \end{aligned} \tag{1.33}$$

Nondimensionalising by setting

$$S = sS_0, \quad C_1 = c_1E_0, \quad C_2 = c_2E_0, \quad t = t'/k_1E_0, \tag{1.34}$$

gives

$$\begin{aligned} \frac{ds}{dt'} &= \frac{k_{-1}}{k_1S_0}c_1 - s(1 - c_1 - c_2) + \frac{k_{-3}}{k_1S_0}c_2 - \frac{k_3}{k_1}sc_1, \\ \varepsilon\frac{dc_1}{dt'} &= s(1 - c_1 - c_2) - \frac{k_2 + k_{-1}}{k_1S_0}c_1 - \frac{k_3}{k_1}sc_1 + \frac{k_4 + k_{-3}}{k_1S_0}c_2, \\ \varepsilon\frac{dc_2}{dt'} &= \frac{k_3}{k_1}sc_1 - \frac{k_4 + k_{-3}}{k_1S_0}c_2. \end{aligned} \tag{1.35}$$

As before, we suppose $\varepsilon \ll 1$, whence we deduce the quasi-steady state approximation

$$\begin{aligned} 0 &\approx s(1-c_1-c_2) - \frac{k_2+k_{-1}}{k_1 S_0} c_1 - \frac{k_3}{k_1} s c_1 + \frac{k_4+k_{-3}}{k_1 S_0} c_2, \\ 0 &\approx \frac{k_3}{k_1} s c_1 - \frac{k_4+k_{-3}}{k_1 S_0} c_2, \end{aligned} \tag{1.36}$$

i. e.,

$$\begin{aligned} c_1 &= \frac{K_2' s}{s^2 + K_2' s + K_1' K_2'}, \\ c_2 &= \frac{s^2}{s^2 + K_2' s + K_1' K_2'}, \end{aligned} \tag{1.37}$$

where

$$K_1' = \frac{k_{-1}+k_2}{k_1 S_0}, \quad K_2' = \frac{k_4+k_{-3}}{k_3 S_0}. \tag{1.38}$$

Dimensionally, under the quasi-steady state assumption, the rate of reaction is found after some algebra to be

$$r = \frac{(k_2 K_2 + k_4 S) E_0 S}{K_1 K_2 + K_2 S + S^2}, \tag{1.39}$$

where

$$K_1 = \frac{k_{-1}+k_2}{k_1}, \quad K_2 = \frac{k_4+k_{-3}}{k_3}. \tag{1.40}$$

If the rates of binding (and reaction) at each site are identical and independent then

$$k_1 = 2k_3, \quad k_{-3} = 2k_{-1}, \quad k_4 = 2k_2. \tag{1.41}$$

The reaction rate is then given by

$$r = \frac{2k_2 E_0 S}{K + S}, \tag{1.42}$$

where

$$K = \frac{2(k_{-1}+k_2)}{k_1}. \tag{1.43}$$

Thus, as expected, the reaction rate is exactly twice that for an enzyme with a single active site.

If cooperativity is large then the rate of binding of the first substrate molecule is small, but the rate of binding of the second once the first is attached is large. This corresponds to the limit $k_1 \sim 0$, $k_3 \sim \infty$, with $k_1 k_3$ finite, so that $K_1 \sim \infty$, $K_2 \sim 0$, with $K_1 K_2$ finite. In this limit the reaction rate

$$r \approx \frac{k_4 E_0 S^2}{K_1 K_2 + S^2}, \tag{1.44}$$

which is a Hill equation with exponent 2.

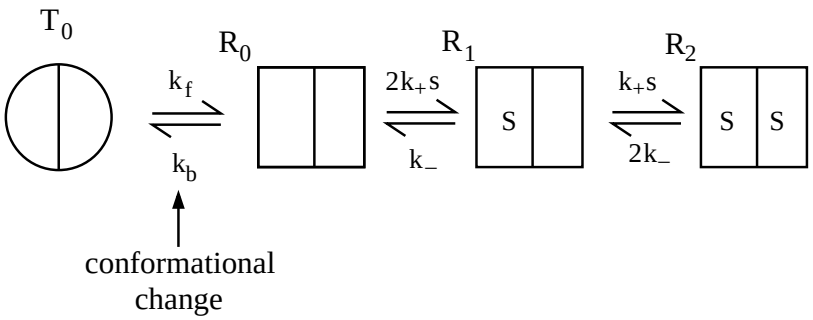

Figure 1.5: A schematic diagram of the reaction scheme for a simple Monod-Wyman-Changeux model.

If the enzyme has binding sites for n substrate molecules, then in the same limit of large cooperativity $k_i \sim 0$ for $i < n$, $k_n \sim \infty$, with $\prod_{i=1}^{n} k_i$ finite (so that $K_i \sim \infty$ for $i < n$, $K_n \sim 0$ with $\prod_{i=1}^{n} K_i$ finite) the reaction rate is approximately

$$r = \frac{r_{\max} S^n}{\prod_{i=1}^{n} K_i + S^n}, \tag{1.45}$$

which is a Hill equation with exponent n. Such a Hill equation is often used to model the reaction rate when details of the intermediate steps are not known, but where cooperative behaviour is suspected, with the parameters $r_{\max}$ and n fitted from experiment.

While the model (1.33) can predict the overall reaction rate given the individual rate constants, it gives no explanation of why cooperative behaviour should occur, i. e., why the rate constant k_3 should be larger than $\frac{1}{2}k_1$. One of the first models that was proposed to explain cooperativity was the allosteric theory of Monod-Wyman-Changeux, which is illustrated in figure 1.5. It assumes that the protein has two conformational states (denoted in the figure by a circle and a square), and that these two states differ in their ability to bind to the substrate molecules. In the simplest model (shown in the diagram), only when the protein is in one of the conformational states (the square) can the substrate bind to the active sites. However, the protein can only switch between conformations when no substrate molecules are bound, so that once one substrate molecule is bound the protein is "locked" in that confirmation until the substrate molecule unbinds.

It is straightforward to write down rate equations for the reaction scheme illustrated in figure 1.5. The rate of reaction is again a sigmoidal function of substrate concentration S. This is illustrated in detail in section 1.5 below.

1.5 Glycolysis

We have seen that the Michaelis-Menten law for the rate of enzyme-catalysed reactions differs from the simple law of mass action, and that cooperative and inhibitory

effects may lead to even more complicated reaction rates as functions of substrate concentration. However, with some reactions things may be more complicated still, with positive feedback loops leading to oscillations in the concentrations.

An example of a biochemical pathway in which oscillations can occur is glycolysis. This is part of the sequence of reactions converting foodstuffs to energy – e. g., the oxidation of glucose:

$$\mathrm{C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O + \text{energy}}. \tag{1.46}$$

In some circumstances, oscillations can occur in the concentrations. One part of the pathway consists of the transformation of the substrate ATP (adenosine triphosphate) to ADP (adenosine diphosphate) via the action of the allosteric enzyme PFK (phosphofructokinase). The rate of this reaction is, however, modified by the ADP itself in an autocatalytic fashion.

We initially consider as an example the precise situation shown in figure 1.5, in which we suppose that the allosteric enzyme PFK is a *dimer* consisting of two subunits (*protomers*) which can exist (together) in either of two conformations, T and R. We suppose that ATP can only bind to PFK in the R form, and that each protomer of R can bind one molecule of ATP. We denote the concentration of ATP by s and that of ADP by p, and initally ignore binding of ADP to PFK. There are four states of the enzyme, which we denote as $\mathrm{T_0}$, $\mathrm{R_0}$, $\mathrm{R_1}$ and $\mathrm{R_2}$, with the subscript denoting the number of molecules of ATP bound to the enzyme. The reaction scheme between these different states is thus

$$\mathrm{T_0} \underset{k_b}{\overset{k_f}{\rightleftharpoons}} \mathrm{R_0} \underset{k_-}{\overset{2k_+s}{\rightleftharpoons}} \mathrm{R_1} \underset{2k_-}{\overset{k_+s}{\rightleftharpoons}} \mathrm{R_2}, \tag{1.47}$$

where the factors of two arise through the number of bound or unbound sites available.

It is straightforward to write down the rate equations for these variables, and these are

$$\begin{aligned}
\dot{T}_0 &= -k_f T_0 + k_b R_0, \\
\dot{R}_0 &= k_f T_0 - k_b R_0 - 2k_+ s R_0 + k_- R_1, \\
\dot{R}_1 &= 2k_+ s R_0 - k_- R_1 - k_+ s R_1 + 2k_- R_2, \\
\dot{R}_2 &= k_+ s R_1 - 2k_- R_2, \\
\dot{s} &= -2k_+ s R_0 + k_- R_1 - k_+ s R_1 + 2k_- R_2.
\end{aligned} \tag{1.48}$$

In keeping with the quasi-steady state assumption, we assume the first four of these equations are in equilibrium, and then after some algebra, we find that

$$T_0 = \frac{LK^2 R_2}{s^2}, \quad R_0 = \frac{K^2 R_2}{s^2}, \quad R_1 = \frac{2KR_2}{s}, \tag{1.49}$$

where

$$L = \frac{k_b}{k_f}, \quad K = \frac{k_-}{k_+}. \tag{1.50}$$

The fraction of bound sites on all forms of the enzyme is $Y = \dfrac{R_1 + 2R_2}{2(T_0 + R_0 + R_1 + R_2)}$, and using (1.49) we find that this is

$$Y = \frac{S(1+S)}{L + (1+S)^2}, \tag{1.51}$$

where

$$S = \frac{s}{K}. \tag{1.52}$$

Now suppose that the product ADP (denoted P) with concentration p is produced at a rate k per bound site. Then we have the additional reactions

$$\begin{aligned} \mathrm{R}_1 &\overset{k}{\rightarrow} \mathrm{R}_0 + \mathrm{P}, \\ \mathrm{R}_2 &\overset{2k}{\rightarrow} \mathrm{R}_1 + \mathrm{P}, \end{aligned} \tag{1.53}$$

and these modify the model in (1.48) simply by replacing k_- by $k + k_-$. Thus in this case the fraction of bound sites is still given by (1.51), but where now

$$S = \frac{s}{K_m}, \tag{1.54}$$

and

$$K_m = \frac{k_- + k}{k_+}. \tag{1.55}$$

The rate of reaction r in the overall reaction $\mathrm{S} \rightarrow \mathrm{P}$ is

$$r = kR_1 + 2kR_2 = 2kR_2 \left(1 + \frac{1}{S}\right). \tag{1.56}$$

We can relate this to Y by noting that the enzyme is conserved; $T_0 + R_0 + R_1 + R_2 = e_0$ is constant, so that, using (1.49), we find

$$r = 2ke_0 Y. \tag{1.57}$$

Because Y is a sigmoidal function of S, so also is r, as we mentioned earlier.

Now let us add in the effect of feedback, that is we allow the product P to bind as a second substrate to the enzyme, and specifically we will suppose that binding of S prevents further binding of P. No production of P is yet invoked. The reaction scheme is shown in figure 1.6, and it is a simple if tedious matter to write down the corresponding rate equations. Assuming the corresponding ten enzyme equations are in equilibrium, we find, after some algebra[2], that the fraction of bound sites is now given by

$$Y = \frac{S(1+S)(1+P)^2}{L + (1+S)^2(1+P)^2}, \tag{1.58}$$

[2]It is simplest to first write the nine overall reaction rates $r_{00} = k_f T - k_b R_{00}$, etc., and then the rate equations are $\dot{T} = -r_{00}$, etc.; equilibrium shows that all $r_{ij} = 0$, which yields a sequence of expressions for $R_{ij} \propto R_{00}$. Conservation of enzyme yields R_{00}, and the number of S-bound sites follows easily.

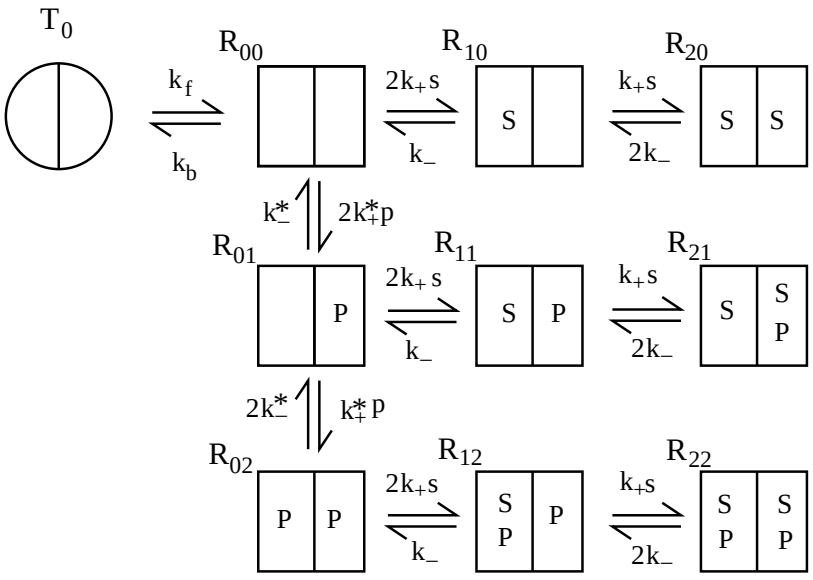

Figure 1.6: Reaction scheme for the Monod-Wyman-Changeux dimer with product feedback. Binding of S (ATP) is assumed to inhibit further binding of P (ADP).

where

$$S = \frac{s}{K_s}, \quad P = \frac{p}{K_p}, \tag{1.59}$$

and in terms of the reaction rates indicated in figure 1.6,

$$L = \frac{k_b}{k_f}, \quad K_p = \frac{k_-^*}{k_+^*}, \quad K_s = \frac{k_-}{k_+}. \tag{1.60}$$

If we now add production of P to the scheme via reactions involving enzymes with bound S such as

$$\mathrm{R}_{10} \overset{k}{\rightarrow} \mathrm{R}_{00} + \mathrm{P}, \tag{1.61}$$

this adds terms to the equations of the form

$$\dot{p} = kR_{10} + \ldots, \quad \dot{R}_{10} = -kR_{10} + \ldots, \quad \dot{R}_{00} = kR_{10} + \ldots . \tag{1.62}$$

These latter terms combine with the dissociation reactions such as

$$R_{10} \overset{k_-}{\rightarrow} R_{00}, \tag{1.63}$$

and it is clear that the effect of production of p is simply to change k_- to $k_- + k$ in the enzyme model of figure 1.6. Hence the overall reaction rate in the reaction $\mathrm{S} \overset{r}{\rightarrow} \mathrm{P}$ is just

$$r = 2ke_0Y \tag{1.64}$$

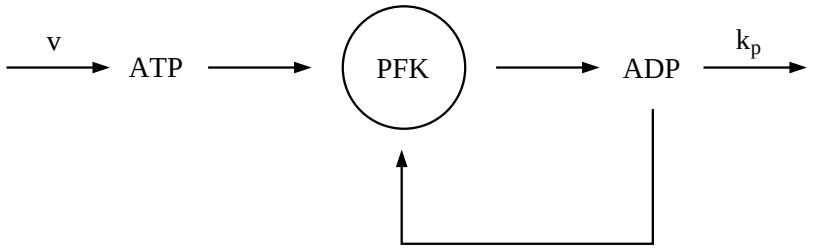

Figure 1.7: A part of the glycolytic path of reactions involving glucose. ATP is converted to ADP; however, this can bind to the allosteric enzyme PFK, which can then affect the rate of reaction.

as before, the bound site fraction is still given by (1.58), but S is now defined by

$$S = \frac{s}{K_m}, \tag{1.65}$$

where

$$K_m = \frac{k_- + k}{k_+}. \tag{1.66}$$

We can now write a model for the ATP–ADP reaction scheme indicated in figure 1.7. If, as indicated, S is produced at a rate v, while P is removed by first order decay with rate coefficient k_p, then a suitable model is

$$\begin{aligned} \dot{s} &= v - r, \\ \dot{p} &= -k_p p + r, \end{aligned} \tag{1.67}$$

where r is the reaction rate in (1.64). It is convenient to write these equations in dimensionless form by scaling

$$s \sim K_m, \quad p \sim K_p, \quad t \sim \frac{1}{k_p}. \tag{1.68}$$

From this we derive the dimensionless model

$$\begin{aligned} \dot{S} &= \mu - \phi(S, P), \\ \dot{P} &= -P + K\phi(S, P), \end{aligned} \tag{1.69}$$

where

$$\phi = \frac{\beta S(1+S)(1+P)^2}{L + (1+S)^2(1+P)^2}, \quad \beta = \frac{2ke_0}{k_p K_m}, \quad \mu = \frac{v}{k_p K_m}, \quad K = \frac{K_m}{K_p}, \quad L = \frac{k_b}{k_f}. \tag{1.70}$$

As we shall see, this model supports self-sustained oscillatory solutions.

1.5.1 Glycolytic oscillations

It is straightforward in principle to analyse the solution in the (S, P) phase plane, but this is complicated by the complexity of the rate function ϕ; and this function is even more complicated for more complex models of the enzyme. Simplification ensues by using realistic values for the dimensionless parameters μ, L, β and K. Typical values of these (for example, for ATP–ADP conversion in yeast) satisfy $L \gg \beta \gg 1$, with $\mu \sim K \sim 1$, and we shall assume these orders of magnitude. Typical quoted values are in the range $L \sim 10^6$, $\beta \sim 10^3$, for example.

It is simple to see from (1.69) that S and P remain positive since $\dot{S} > 0$ on $S = 0$ and $\dot{P} > 0$ on $P = 0$, $S > 0$. Further, there is a unique fixed point in the positive quadrant, since we must have $P = K\mu$, and ϕ is certainly an increasing function of S if $L \gg \beta$. Next, we study the shape of the nullclines (where the derivatives of S and P are zero), assuming $L \gg \beta \gg 1$ and $K \sim \mu \sim O(1)$. The S nullcline is given by

$$\frac{\beta S(1+S)(1+P)^2}{L + (1+S)^2(1+P)^2} = \mu. \tag{1.71}$$

Suppose that $S, P \gg 1$. Then this is approximated by

$$\frac{\beta S^2 P^2}{L + S^2 P^2} = \mu, \tag{1.72}$$

i. e., the S nullcline at large P is given by the hyperbola

$$SP \approx \left[\frac{\mu L}{\beta - \mu}\right]^{1/2}. \tag{1.73}$$

Note that this is consistent with the assumption that $S, P \gg 1$, since the right hand side is of $O(L/\beta)^{1/2} \gg 1$. The approximations break down, but not dramatically, if $S \to 0$ or $P \to 0$. As $P \to 0$, the asymptote at $P = 0$ is in fact at $P = -1$, and as $P \to 0$, S cuts the S axis at $S \approx \left(\dfrac{\mu L}{\beta - \mu}\right)^{1/2} \gg 1$. As S becomes small at large P ($\gg L^{1/2}$), we see that $S \to \mu/\beta \ll 1$. Thus the S nullcline is monotonically decreasing as P increases, and reasonably approximated everywhere by (1.73).

The P nullcline where $\dot{P} = 0$ is given by

$$P = \frac{\beta K S(1+S)(1+P)^2}{L + (1+S)^2(1+P)^2}. \tag{1.74}$$

Now we see in the S nullcline from (1.73) that $SP \ll \sqrt{L}$, and with this assumption, the P nullcline takes the approximate form

$$S + S^2 \approx \frac{L}{\beta K} \frac{P}{(1+P)^2}. \tag{1.75}$$

Since the left hand side is an increasing function of S, this gives S as a unimodal (one-humped) function of P, increasing from zero at $P = 0$ to a maximum, and then decreasing towards zero as $P \to \infty$. While $P \sim O(1)$, S is large, thus

$$S \approx \left[\frac{L}{\beta K}\,\frac{P}{(1+P)^2}\right]^{1/2}, \tag{1.76}$$

and $SP \ll \sqrt{L}$ as assumed. As P becomes large, the approximation becomes

$$S \approx \left[\frac{L}{\beta K P}\right]^{1/2}, \tag{1.77}$$

and the assumption that $SP \ll \sqrt{L}$ remains valid until $P \gtrsim \beta$.

For $P \gtrsim \beta \gg 1$ and while S is still large, we anticipate that then $SP \gtrsim \sqrt{L}$, and the nullcline is approximated by $P \approx \dfrac{\beta K S^2 P^2}{L + S^2 P^2}$, i. e.,

$$\frac{1}{\beta K S} = \frac{SP}{L + S^2 P^2}. \tag{1.78}$$

The right hand side is a unimodal function of SP, so that S decreases from ∞ to a minimum and then increases again as SP increases. This then implies the same behaviour for S as P increases (think graphically!). Further, as $P \to 0$, then $S \to \infty$ and $SP \to 0$ in this approximation, so that $\dfrac{1}{\beta K S} \approx \dfrac{SP}{L}$ which is identical to (1.75), which is itself the large P limit of the unimodal approximation when $P \ll O(\beta)$. Thus these two approximations match to each other and provide a uniform approximation for the P nullcline, which has a pseudo-cubic shape as shown in figure 1.8. A uniform approximation which is also approximated by both (1.76) and (1.78) is

$$S \approx \left[\frac{LP}{\beta K(1+P)^2 - P^3}\right]^{1/2}. \tag{1.79}$$

Note from (1.79) that $S \to \infty$ as $P \to \beta K$, approximately. In fact this can be seen from (1.74) since the maximum of the right hand side is βK when $S \to \infty$.

We denote the local maximum of the P nullcline as U and the local minimum as V (see figure 1.8). The stability of the fixed point then depends on whether the fixed point of the system lies to the left of U, to the right of V, or between U and V. It is easy to use the approximations described above to estimate the locations of these points. The results are that U is approximately at $P = 1$, $S = \left(\dfrac{L}{4\beta K}\right)^{1/2}$, while V is approximately at $S = \dfrac{2L^{1/2}}{\beta K}$, $P = \dfrac{\beta K}{2}$.

If we linearise the equations (1.69) about the fixed point (S_0, P_0) by writing $S = S_0 + \sigma$, $P = P_0 + \pi$, then we find

$$\begin{pmatrix} \dot\sigma \\ \dot\pi \end{pmatrix} = \begin{pmatrix} -\phi_S & -\phi_P \\ K\phi_S & K\phi_P - 1 \end{pmatrix} \begin{pmatrix} \sigma \\ \pi \end{pmatrix}. \tag{1.80}$$

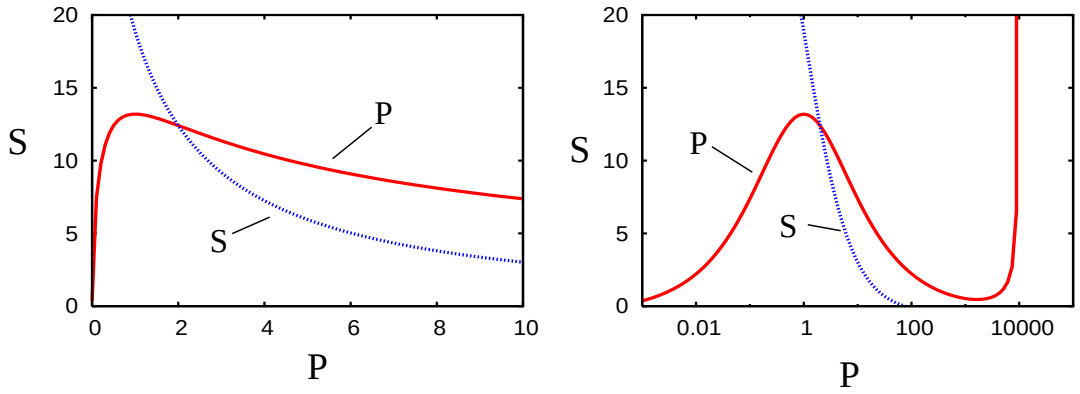

Figure 1.8: S and P nullclines (linear and log plots) when $\mu = 2$, $K = 1$, $\beta = 10^4$ and $L = 0.75 \times 10^7$.

The determinant of the matrix in this equation is ϕ_S which is positive, and therefore instability occurs (as a Hopf bifurcation) if and only if the trace of the matrix is positive, i. e., $K\phi_P > \phi_S + 1$. Since the slope of the P nullcline S'_P is easily computed to be

$$S'_P = \frac{1 - K\phi_P}{K\phi_S}, \tag{1.81}$$

we can deduce that instability occurs if and only if $-S'_P > \dfrac{1}{K}$. In particular the P nullcline must have negative slope for instability, i. e., the fixed point must lie between U and V.

We use (1.73) and (1.76) to calculate the values of S, P and S'_P at the fixed point. P is given implicitly by

$$\frac{P^3}{(1+P)^2} = \mu K, \tag{1.82}$$

and

$$S \approx \left(\frac{\mu L}{\beta}\right)^{1/2} \frac{1}{P}, \tag{1.83}$$

whence

$$-2SS'_P = \frac{L}{\beta K}\left[\frac{P-1}{(1+P)^3}\right]. \tag{1.84}$$

Instability thus occurs if

$$\frac{P(P-1)}{(1+P)^3} > 2\sqrt{\frac{\mu\beta}{L}}, \tag{1.85}$$

and thus approximately if $P \gtrsim 1$. Bearing in mind (1.82), this implies that instability occurs for

$$\mu K \gtrsim \tfrac{1}{4}. \tag{1.86}$$

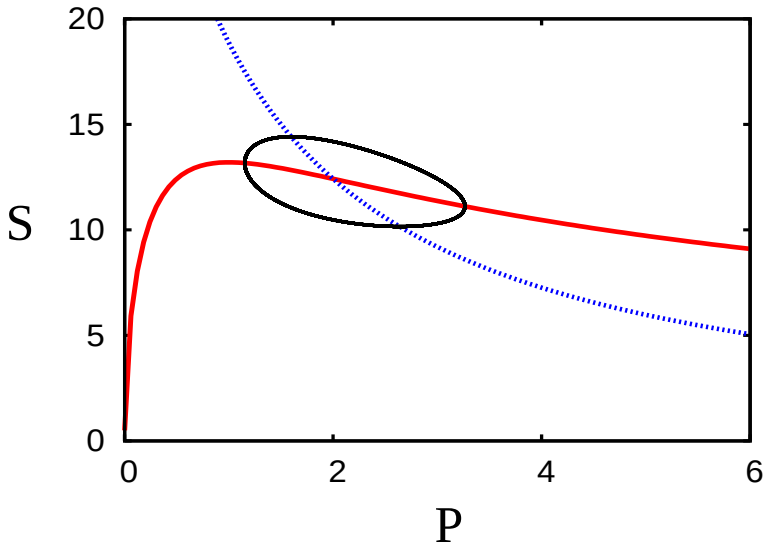

Figure 1.9: Limit cycle solution of (1.69) with $\mu = 2$, $K = 1$, $L = 0.75 \times 10^7$, and $\beta = 10^4$.

1.5.2 Limit cycles

Figure 1.9 shows the limit cycle oscillation which ensues when we use the values $\mu = 2$, $K = 1$, $L = 0.75 \times 10^7$, and $\beta = 10^4$. The presence of the large parameters L and β suggest the possibility of asymptotic methods for the solution of (1.69), although figure 1.9 itself shows no sign of any limiting behaviour (such as a relaxation oscillation).

The periodic orbit shown lies close to both nullclines, and this suggests a rescaling in the form

$$S = \frac{\psi}{\delta}, \tag{1.87}$$

where

$$\delta = \sqrt{\frac{\beta}{L}} \ll 1. \tag{1.88}$$

With this definition,

$$\phi = \frac{\psi(\psi + \delta)(1 + P)^2}{1 + \varepsilon(\psi + \delta)^2(1 + P)^2}, \tag{1.89}$$

where

$$\varepsilon = \frac{1}{\beta} \ll 1. \tag{1.90}$$

In addition, we rescale $t \sim \dfrac{1}{\delta}$, to find

$$\begin{aligned} \dot{\psi} &= \mu - \phi, \\ \delta \dot{P} &= -P + K\phi, \end{aligned} \tag{1.91}$$

and the overdot now denotes differentiation with respect to this rescaled time.

This is now in the classic form of a relaxation oscillation. Since $\delta \ll 1$, we expect P to relax rapidly to the P-nullcline, and since this curve has the classic pseudo-cubic shape (similar to the 'slow manifold' of the Van der Pol oscillator), we might expect a similar kind of switching behaviour. In reality, we can see from figure 1.9, for which the value of $\delta \approx 0.04$, that this behaviour is not attained. The orbit stays relatively close to the P-nullcline, but does not drift past the turning points.

However, if we reduce the value of β to 10^2, for which $\delta \approx 0.004$, then the motion becomes more clearly relaxational, as shown in figures 1.10 and 1.11. The trajectory hugs the left part of the P-nullcline, but is unable to reach the right part, which lies in $P \sim \beta \gg 1$. (Further reduction of β does not help, since then the P-nullcline loses its non-monotonicity.) As seen in figure 1.11, the solution for S becomes relaxational, while that for P develops a series of isolated pulses.

It is easy to see that the slow branch of the oscillation where P is small is described by the quasi-steady approximation for (1.89) and (1.91), where we put $\delta = 0$. It is less obvious how to describe the rapid pulses of figure 1.11. Apparently, ψ remains $O(1)$, but P is large. If $P \sim \beta = \dfrac{1}{\varepsilon}$ as we might expect, then $\phi \approx \dfrac{1}{\varepsilon}$ is constant, and P would approach βK. Evidently, this is prevented in figure 1.10 by the fact that the trajectory crosses the P-nullcline.

We need to keep P and ϕ the same size in the pulse, and this requires that formally ψ is small. It is not then difficult to show that a distinguished rescaling which keeps the P–ϕ balance in the P equation, while also allowing ψ to decrease on the same

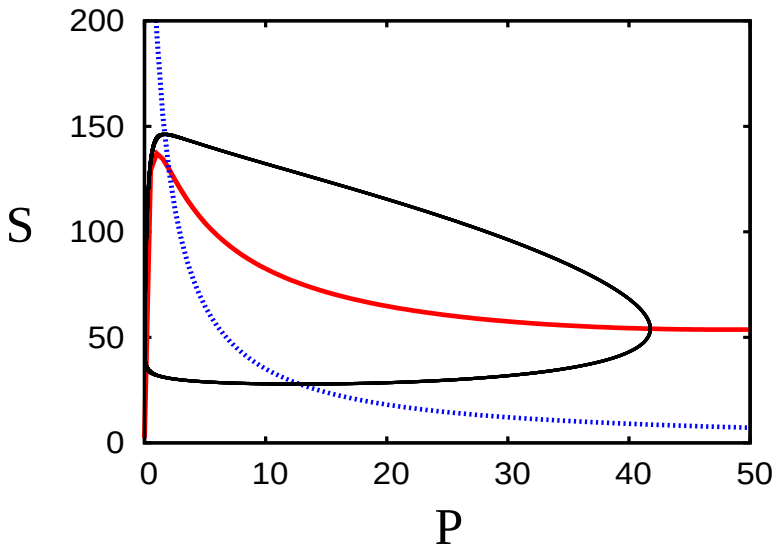

Figure 1.10: Limit cycle solution of (1.69) with $\mu = 2$, $K = 1$, $L = 0.75 \times 10^7$, and $\beta = 10^2$.

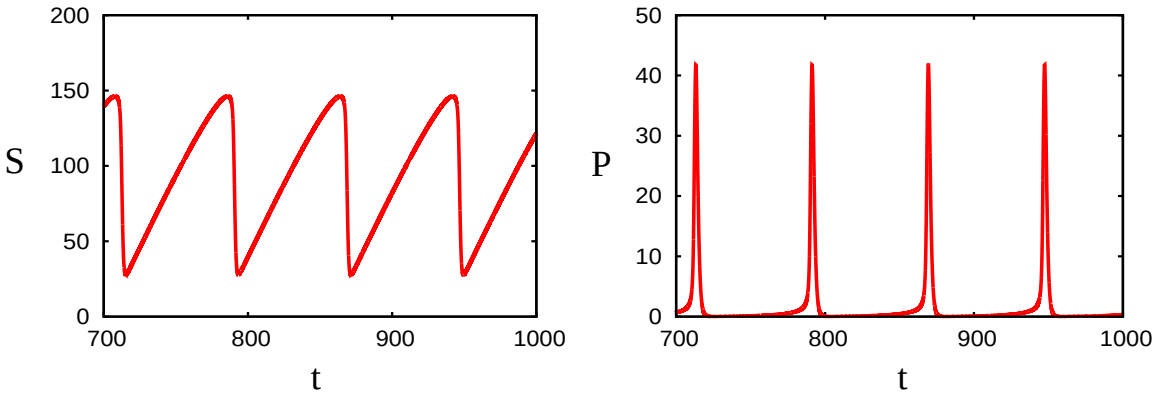

Figure 1.11: Time series for S and P in the solution of (1.69) with $\mu = 2$, $K = 1$, $L = 0.75 \times 10^7$, and $\beta = 10^2$, as in figure 1.10.

time scale, can be found by defining

$$\psi = \delta^{1/3}\Psi, \quad P = \frac{\Pi}{\delta^{2/3}}, \quad t = \delta T, \quad \phi = \frac{\Phi}{\delta^{2/3}}, \tag{1.92}$$

whence we obtain

$$\begin{aligned} \Psi' &= \delta^{2/3}\mu - \Phi, \\ \Pi' &= -\Pi + K\Phi, \end{aligned} \tag{1.93}$$

and Φ is given by

$$\Phi = \frac{\Psi(\Psi + \delta^{2/3})(\Pi + \delta^{2/3})^2}{1 + \gamma(\Psi + \delta^{2/3})^2(\Pi + \delta^{2/3})^2}, \tag{1.94}$$

where

$$\gamma = \frac{\varepsilon}{\delta^{2/3}} = \frac{L^{1/3}}{\beta^{4/3}}; \tag{1.95}$$

for $L = 0.75 \times 10^7$ and $\beta = 10^2$, $\gamma = 0.42$. This rescaling is appropriate if $\varepsilon \lesssim \delta^{2/3}$, i. e., $L \lesssim \beta^4$. Neglecting terms of $O(\delta^{2/3})$, we find that at leading order (1.93) reduces to

$$\begin{aligned} \Psi' &\approx -\frac{\Psi^2\Pi^2}{1 + \gamma\Psi^2\Pi^2}, \\ \Pi' &\approx -\Pi + \frac{K\Psi^2\Pi^2}{1 + \gamma\Psi^2\Pi^2}. \end{aligned} \tag{1.96}$$

Numerical simulations of (1.96) show that it gives a very good approximation to the pulse, providing the initial condition is chosen appropriately (for example, with $S = 154$, $P = 4$). Formally, the initial conditions must come from matching the pulse to the slow phase of the oscillation.

Chapter 2

Trans-membrane ion transport

Living organisms consist of cells, and complex organisms such as the human body consist of organs formed by cells (the brain, the heart, the kidneys, the skin, and so on) which reside in a medium of extracellular fluid. Different cells have different functions (nerve cells, muscle cells, stem cells, etc.) which are enabled by a constant supply of a nutrient (oxygen) which provides energy through reaction with organic carbon in the form of sugars, for example, with the waste product being carbon dioxide; this process is called *respiration*.

Cells are essentially bags of water which contain various constituent parts. There is a nucleus, surrounded by *cytoplasm*, which contains various structures called organelles, such as the mitochondria, the endoplasmic reticulum, the Golgi apparatus, and so on. The water itself is a dilute aqueous solution of dissolved salts, mainly sodium chloride (common salt – NaCl) and potassium chloride (KCl), which dissociate into sodium (Na^+), potassium (K^+) and chloride (Cl^-) ions. (The same is true of the extracellular fluid.) As a consequence both extra- and intra-cellular fluids carry an electrical charge. These differ, so that there is an electrical potential difference between them.

In consequence, much of the function of a cell is enabled by electrochemical processes within the cell, which involve transport of ions through the *cell membrane* which separates the intracellular environment from the extracellular one.

2.1 Membrane transport

The cell membrane is a *phospholipid double layer* about 7.5 nm thick separating the cell interior (the cytoplasm) from the extracellular environment (see figure 2.1). The term *lipid* refers to a class of water-insoluble, energy-rich macromolecules, typically fats, waxes and oils. The most important property of the cell membrane is its selective permeability; it allows the passage of some molecules but restricts the passage of others, thereby regulating the passage of materials into and out of the cell.

The membrane contains water-filled pores with diameters of about 0.8 nm, and protein-lined pores, called *channels* or *gates*, which allow the passage of specific

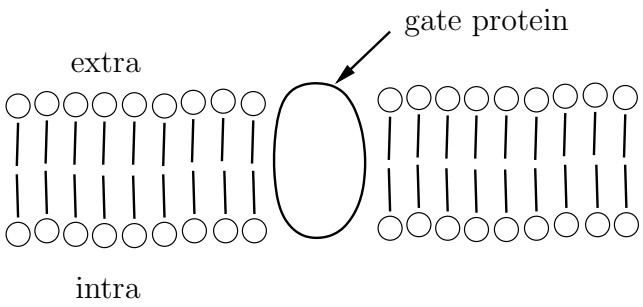

Figure 2.1: Schematic diagram of the cell membrane.

molecules. The cell membrane acts as a barrier to the free flow of dissolved ionic species, and to the flow of water.

The many mechanisms that exist for transporting molecules from one side of the membrane to the other can be divided up into active and passive processes. An active process is one which requires the expenditure of energy (for example, convection), while a passive process is one which results solely from the random motion of molecules (for example, diffusion).

Passive mechanisms by which molecules are transported across the cell membrane include osmosis, diffusion, and carrier-mediated mechanisms. *Osmosis* is the mechanism by which water is transported across the cell membrane. Small molecules such as chloride ions may diffuse through pores in the membrane, while lipid-soluble molecules such as oxygen and carbon dioxide may diffuse directly through the membrane itself. *Carrier-mediated diffusion* refers to a process by which a molecule 'hitches a lift' by binding to a carrier molecule which is lipid soluble and can move readily through the membrane. *Carrier-mediated transport* (figure 2.2) occurs when a protein which sits in the membrane has an active site which may be exposed either on the exterior or interior side of the membrane depending on the conformational state of the protein. A substrate may bind to the protein in one conformation, the protein undergoes a conformational change, and the substrate unbinds on the other side of the membrane.

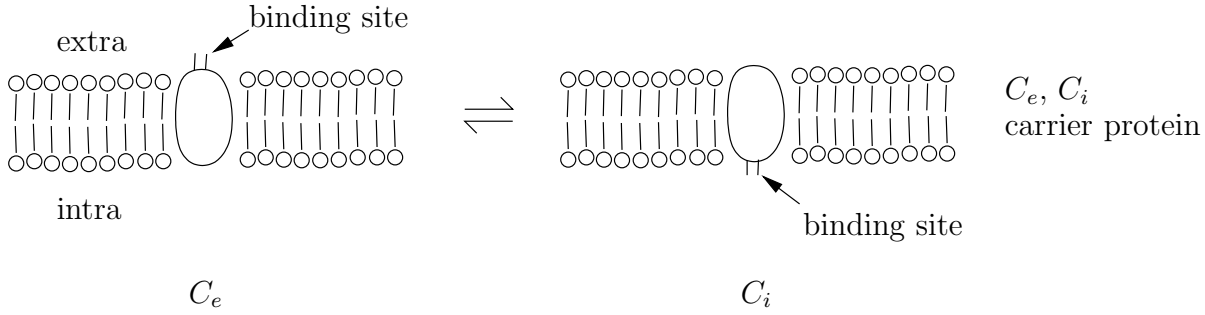

Figure 2.2: Carrier protein switches from binding-site-outwards to binding-site-inwards.

Ionic species	extra	intra
Na^+ (sodium)	437	50
K^+ (potassium)	20	397
Cl^- (chloride)	556	40
Mg^{2+} (magnesium)	53	80

Table 2.1: Typical intracellular and extracellular ionic concentrations for the squid giant axon. Units are mM = millimolar = 10^{-3} M. 1 M = 1 molar = 1 mole litre^{-1}.

The transport of glucose and amino acids across the cell membrane is thought to be by a carried-mediated process.

The concentration differences that exist between the intracellular and extracellular environments are set up and maintained by active processes. One of the most important of these is the Na^+-K^+ pump, which uses the energy stored in ATP molecules to pump Na^+ out of the cell and K^+ in. There are also a variety of exchange pumps, which use the concentration gradient of one ion to pump another ion against its concentration gradient, such as the Na^+-Ca^{2+} exchanger (figure 2.3), which removes Ca^{2+} from the cell at the expense of allowing Na^+ in. Differences in interior and exterior ionic concentrations create a potential difference across the cell which also drives an ionic current down ion-specific membrane channels. Some typical intracellular and extracellular ionic concentrations are shown in Table 2.1.

2.1.1 Carrier mediated transport

We describe here a simple model for carrier mediated transport. We suppose that the carrier protein has two conformational states, and that in the first state, labelled C_i, the substrate binding site is exposed on the cell interior, while in the second state, labelled C_e, the substrate binding site is exposed on the cell exterior (see figure 2.2). We suppose that substrate molecules outside the cell (concentration S_e) can bind to C_e to produce a complex P_e, and that substrate molecules inside the cell

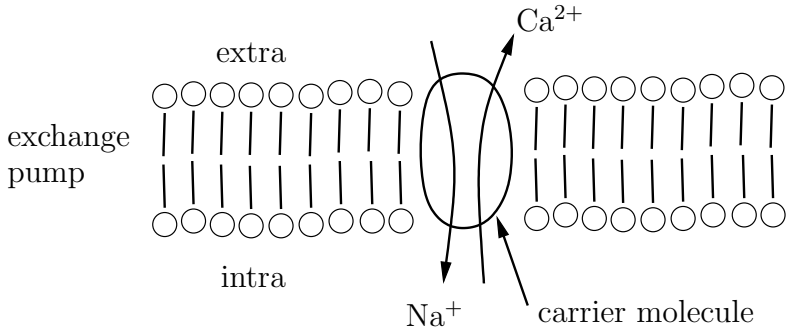

Figure 2.3: The sodium–calcium exchanger.

(concentration S_i) can bind to C_i to produce a complex P_i . Furthermore, we assume that P_i can conformally change into P_e and vice versa, and at the same rate as the conformal changes of C_i and C_e (so that the binding of the substrate does not affect the conformal changes of the protein). Thus the reaction scheme is

$$\begin{aligned} S_i + C_i \underset{k_-}{\overset{k_+}{\rightleftharpoons}} P_i \quad &\underset{k}{\overset{k}{\rightleftharpoons}} \quad P_e \underset{k_+}{\overset{k_-}{\rightleftharpoons}} S_e + C_e \\ C_i \quad &\underset{k}{\overset{k}{\rightleftharpoons}} \quad C_e, \end{aligned} \tag{2.1}$$

where we have assumed that the binding affinity of S_i to C_i is the same as that of S_e to C_e, and that the two conformational states C_i and C_e are equally likely. To avoid the system simply settling down to a steady state with zero flux, we assume that the substrate is supplied at a constant rate J on the exterior and taken away at the same rate from the interior, and we wish to determine this flux of substrate through the membrane as a function of the interior and exterior concentrations.

Using the law of mass action the reaction kinetics are given by

$$\begin{aligned} \frac{dS_i}{dt} &= k_- P_i - k_+ S_i C_i - J, \\ \frac{dS_e}{dt} &= k_- P_e - k_+ S_e C_e + J, \\ \frac{dP_i}{dt} &= kP_e - kP_i + k_+ S_i C_i - k_- P_i, \\ \frac{dP_e}{dt} &= kP_i - kP_e + k_+ S_e C_e - k_- P_e, \\ \frac{dC_i}{dt} &= kC_e - kC_i - k_+ S_i C_i + k_- P_i, \\ \frac{dC_e}{dt} &= kC_i - kC_e - k_+ S_e C_e + k_- P_e. \end{aligned} \tag{2.2}$$

Adding $(2.2)_3$-$(2.2)_6$ gives

$$\frac{d}{dt}\left(P_i + P_e + C_i + C_e\right) = 0, \tag{2.3}$$

so that

$$P_i + P_e + C_i + C_e = C_0 = \text{ constant}, \tag{2.4}$$

i. e., the total amount of carrier is conserved. Adding $(2.2)_1$-$(2.2)_4$ gives

$$\frac{d}{dt}\left(S_i + S_e + P_i + P_e\right) = 0, \tag{2.5}$$

whence

$$S_i + S_e + P_i + P_e = S_0 = \text{ constant}, \tag{2.6}$$

and the total amount of substrate is conserved. In the steady state, assuming J is unknown, we have six equations for seven unknowns, so that there is a one parameter

family of solutions, even when C_0 and S_0 are given. However, solving the linear system $(2.2)_2$-$(2.2)_5$ and (2.4) for J, P_i, P_e, C_i, C_e in terms of S_i and S_e gives

$$J = \frac{k_- k C_0}{2k_+} \frac{S_e - S_i}{(K_m + S_i)(K_m + S_e) - K_d^2}, \tag{2.7}$$

where

$$K_m = \frac{k_- + k}{k_+}, \quad K_d = \frac{k}{k_+}. \tag{2.8}$$

There are also other, more complicated, mechanisms for carrier-mediated transport. The carrier molecule may have binding sites for more than one substrate molecule, leading to cooperative behaviour and more complicated kinetics. It may also have binding sites for different types of substrate molecules. These different substrates may both be transported across the membrane in the same direction (*symport transporters*), or may be transported in opposite directions (*antiport transporters*). A key assumption in modelling such transporters is that the carrier may only undergo the conformational change when either all or none of the substrate molecules are bound. While the different states of the carrier and the transitions between them are easy to write down, the resulting system of equations is sufficiently large that the expression for the flux as a function of concentration is very complicated.

2.1.2 Active transport: the sodium-potassium pump

The carrier-mediated transport described above moves molecules down chemical gradients. Any process which move molecules against a chemical or electrical gradient requires the expenditure of energy, and is known as an active transport mechanism. One of the most important of these is the Na^+–K^+ pump, which pumps sodium ions out of the cell against a steep electrochemical gradient, while pumping potassium ions in. In fact, this pump alone consumes almost a third of the energy requirement of a typical animal cell.

The pump uses the energy stored in ATP which is released when it is dephosphorylated into ADP, through the overall reaction scheme

$$\mathrm{ATP} + 3\mathrm{Na}_i^+ + 2\mathrm{K}_e^+ \rightarrow \mathrm{ADP} + \mathrm{P}_i + 3\mathrm{Na}_e^+ + 2\mathrm{K}_i^+, \tag{2.9}$$

where the subscripts i and e denote intracellular and extracellular ions respectively. The individual components of the reaction are thought to be as follows. When the carrier protein (Na^+-K^+ATPase) is in its dephosphorylated state, three sodium binding sites are exposed to the cell's interior. When all three binding sites are filled, the carrier protein is phosphorylated by the hydrolysis of ATP into ADP. This phosphorylation induces a change in conformation, so that the sodium binding sites are exposed to the cell exterior, and their binding affinity is reduced, causing the release of the sodium ions. At the same time, two potassium sites are exposed to the cell's exterior. When potassium ions have bound to these two sites, the carrier protein is dephosphorylated, inducing the reverse conformational change, exposing the potassium

binding sites to the cell interior and reducing the binding affinity so that potassium is released. If we simplify the process slightly, assuming that there is a single binding site for sodium and potassium, leading to a one-for-one exchange rather than the three-for-two which actually happens, then the detailed reaction scheme is

$$\mathrm{Na}_i^+ + \mathrm{C} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{NaC} \overset{\mathrm{ATP} \overset{k_p}{\to} \mathrm{ADP}}{\longrightarrow} \mathrm{NaCP} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{Na}_e^+ + \mathrm{CP},$$

$$\mathrm{CP} + \mathrm{K}_e^+ \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} \mathrm{KCP} \underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} \mathrm{P} + \mathrm{KC},$$

$$\mathrm{KC} \underset{k_{-5}}{\overset{k_5}{\rightleftharpoons}} \mathrm{K}_i^+ + \mathrm{C}, \tag{2.10}$$

where the carrier protein is represented by C in its unphosphorylated, unbound state, CP in its phosphorylated unbound state, NaC when bound to sodium and unphosphorylated, NaCP when bound to sodium and phosphorylated, KC when bound to potassium and unphosphorylated, and KCP when bound to potassium and phosphorylated. Using mass action kinetics, assuming that intracellular potassium and extracellular sodium are removed at a constant rate J, leads to a steady-state flow of ions through the pump

$$J = \frac{J_0\,[\mathrm{Na}_i^+][\mathrm{K}_e^+]}{[\mathrm{Na}_i^+] + \alpha[\mathrm{K}_i^+] + \beta[\mathrm{K}_e^+]} \tag{2.11}$$

where

$$J_0 = \frac{C_0 k_3 k_4 k_5}{k_{-3}k_{-3}[\mathrm{P}] + k_{-3}k_5 + k_4 k_5}, \quad \alpha = \frac{(k_{-1}k_p + k_2 k_{-1} + k_2 k_p)k_{-3}k_{-4}k_{-5}}{k_1 k_2 k_p (k_{-3}k_{-3}[\mathrm{P}] + k_{-3}k_5 + k_4 k_5)},$$

$$\beta = \frac{(k_{-1}k_p + k_2 k_{-1} + k_2 k_p)k_3 k_4 k_5}{k_1 k_2 k_p (k_{-3}k_{-3}[\mathrm{P}] + k_{-3}k_5 + k_4 k_5)}, \tag{2.12}$$

C_0 is the total concentration of carrier molecule, and $[\mathrm{Na}_e^+]$ denotes the concentration of sodium ions in the extracellular medium, etc. The important thing to note is that for both carrier-mediated transport (2.7) and active transport (2.11), pump fluxes have the same features as enzyme catalysed reactions, that is, they are linear in the concentrations at small concentrations, but saturate to a maximum value at large concentrations.

2.1.3 The membrane potential

The differences in the concentration of various ions between the extracellular and intracellular medium which are set up and maintained by active transport mechanisms can cause a potential difference to be generated between the inside and the outside of the cell.

Suppose we have two reservoirs containing different concentrations of a positively charged ion X^+. We suppose both reservoirs are electrically neutral to begin with, so

that there is an equal concentration of a negatively charged ion Y^-. Now suppose the reservoirs are separated by a semi-permeable membrane which is permeable to X^+ but not to Y^-. Then the difference in concentration of X^+ on each side will lead to the flow of X^+ across the membrane. However, because Y^- cannot diffuse through the membrane, this will lead to a build up of charge on one side. This charge imbalance sets up an electric field, which produces a force on the ions opposing further diffusion of X^+. It is important to realise that the actual amount of X^+ which diffuses through the membrane is small, and the excess charge all accumulates near the interface, so that to a good approximation the solutions on either side remain electrically neutral. The potential difference at which equilibrium is established and diffusion and electric-field-generated fluxes balance is known as the *Nernst potential*. We may derive an expression for it as follows.

If c denotes the concentration of an ion S then the flux of ions due to diffusion is $\mathbf{J} = -D\boldsymbol{\nabla} c$, where D is the diffusion coefficient. To this we must add the flux due to the fact that the ion carries a charge and is in the presence of an electric field, which is given by

$$\mathbf{J} = -\frac{uzc}{|z|}\boldsymbol{\nabla}\phi, \tag{2.13}$$

where u is the mobility of the ion (defined as the velocity under a constant unit electric field), z is the valence of the ion (so that $z/|z|$ is either $+1$ or -1 and gives the sign of the force on the ion; positive ions move down potential gradients, negative ions move up potential gradients), and ϕ is the electric potential, so that $-\boldsymbol{\nabla}\phi$ is the electric field. Thus the total flux is given by

$$\mathbf{J} = -D\boldsymbol{\nabla} c - \frac{uzc}{|z|}\boldsymbol{\nabla}\phi. \tag{2.14}$$

Now, there is a relationship (determined by Einstein) between the ionic mobility and the diffusion coefficient, which is

$$D = \frac{uRT}{|z|F}, \tag{2.15}$$

where R is the universal gas constant, T is the absolute temperature and F is Faraday's constant. Furthermore, since the membrane is thin we can replace the Nernst-Planck equation (2.14) with the one-dimensional version

$$J = -D\left(\frac{\partial c}{\partial x} + \frac{zFc}{RT}\frac{\partial \phi}{\partial x}\right), \tag{2.16}$$

where x is a coordinate normal to the membrane. Now, at equilibrium the flux J is zero, giving

$$\frac{\partial c}{\partial x} + \frac{zFc}{RT}\frac{\partial \phi}{\partial x} = 0. \tag{2.17}$$

Assuming that the interior of the membrane is at $x = 0$ while the exterior is at $x = L$, we may integrate from 0 to L to give

$$\phi_i - \phi_e = \frac{RT}{zF}\ln\left(\frac{c_e}{c_i}\right), \tag{2.18}$$

where the subscripts e and i denote extracellular and intracellular quantities respectively. If we follow the standard convention of defining the potential difference across the cell membrane as $V = \phi_i - \phi_e$, then the Nernst potential for S is

$$V_S = \frac{RT}{zF} \ln \frac{[\mathrm{S}_e]}{[\mathrm{S}_i]}. \tag{2.19}$$

Using the values of intra- and extracellular concentrations given in table 2.1, typical Nernst potentials for the squid giant axon for potassium and sodium are $V_{\mathrm{K}} = -77$ mV (millivolt) and $V_{\mathrm{Na}} = 56$ mV. Note that when more than one ion is present, and they have different Nernst potentials, the flux of each individual ion will not be zero even when there is no net current across the membrane. For example, when -77 mV $< V <$ 56 mV there will be a flux of K^+ out of the cell and Na^+ into the cell through ion-specific channels. This flux is balanced by the action of the Na^+-K^+ pump. This is illustrated in the following section.

The resting potential

Let us now consider a cell in which there are sodium channels with outward ionic flux (equals current since the valency of sodium is one) J_{Na}, potassium channels with outward flux J_{K}, and sodium-potassium pumps in which the flux of sodium outwards is J_P (and thus the outwards potassium flux is $-J_P$). J_{Na} and J_{K} are still given by (2.16), but are no longer zero in equilibrium. Rather, we have

$$J_{\mathrm{Na}} + J_P = 0, \quad J_{\mathrm{K}} - J_P = 0. \tag{2.20}$$

As in (2.11), we presume J_P is given in terms of extracellular and intracellular concentrations, and is thus constant when integrating (2.20). Additionally, we need an equation in the channels for ϕ. In a steady state, the constant current is proportional to the potential gradient, which is therefore also constant. This means that the potential is

$$\phi = V\left(1 - \frac{x}{L}\right), \tag{2.21}$$

where V is the internal potential and L is the channel length, so that in (2.16), $\dfrac{\partial \phi}{\partial x} = -\dfrac{V}{L}$. We thus have to solve the two equations

$$\frac{\partial s}{\partial x} = \frac{Vs}{V_0 L} + \frac{J_P}{D}, \quad \frac{\partial p}{\partial x} = \frac{Vp}{V_0 L} - \frac{J_P}{D}, \tag{2.22}$$

where s and p are the sodium and potassium concentrations, and we have written

$$V_0 = \frac{RT}{F}. \tag{2.23}$$

To be specific, let us suppose that

$$J_P = J_0 s_i p_e, \tag{2.24}$$

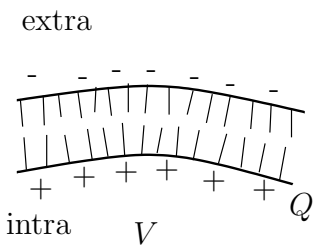

Figure 2.4: Schematic diagram of the build up of charge on the cell membrane.

where the subscripts indicate the intracellular and extracellular values. We then find, on integrating (2.22) and using the definitions of the Nernst potential in (2.19), that

$$\begin{aligned} V - V_{\mathrm{Na}} &= -V_0 \ln\left[1 - \frac{Bp_e}{V}\left(e^{-V/V_0} - 1\right)\right], \\ V - V_{\mathrm{K}} &= V_0 \ln\left[1 - \frac{Bs_i}{V}\left(1 - e^{V/V_0}\right)\right], \end{aligned} \tag{2.25}$$

where

$$B = \frac{J_0 V_0 L}{D}, \tag{2.26}$$

and these define the resting potential V_{eq}. We see from this that if $V < 0$, then $V_{\mathrm{K}} < V < V_{\mathrm{Na}}$, as is the case.

If we suppose the extracellular ionic concentrations are prescribed, then (2.25) provides two equations for s_i and p_i, and further such equations would arise if other ions are present. What then determines V? The answer is that the electric potential inside the cell is just due to the total ionic charge, so V is also a function of the ionic concentrations.

2.1.4 Ionic currents

The flow of ions through the cell membrane due to concentration differences leads to a build up of charge near the membrane and a potential difference (see figure 2.4). Thus the cell membrane is effectively acting as a capacitor. The voltage across any capacitor is related to the charge stored Q by

$$V = \frac{Q}{C}, \tag{2.27}$$

where C is the capacitance. If I is the ionic current out of the cell (the rate of flow of positive charges outwards) then the stored charge changes according to

$$I = -\frac{dQ}{dt}. \tag{2.28}$$

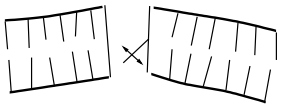

Figure 2.5: Schematic diagram of channel gating.

Thus, assuming the capacitance is constant,

$$C\frac{dV}{dt} + I = 0. \tag{2.29}$$

This equation is the basis for much theoretical electrophysiology. The difference between the various models arises in the expression used for the ionic current I.

The simplest model to use is to assume a linear dependence of I on V (as in Ohm's law). For a single ion S, with Nernst potential V_S, this gives an ionic current

$$I_S = g_S(V - V_S), \tag{2.30}$$

where the constant g_S is the ion-specific membrane conductance, since the current must be zero when $V = V_S$. If more than one ion is present the currents from different ions are simply added together to produce the total ionic current I. Note that the action of pumps such as the sodium-potassium pump do not contribute, as no charge is transferred.

2.1.5 Gating

It is found experimentally that g_S is not constant, but depends on both V and time t. One proposed explanation for this is that the channels are not always open, but may be open or closed, and that the transition rates between open and closed states depend on the potential difference V (see figure 2.5). The membrane conductance may then be written as ng_S, where g_S is the constant conductance which would result if all channels were open, and n is the fraction of open channels.

Denoting the open channels by O and the closed channels by C, the reaction scheme is simply

$$\mathrm{C} \underset{\beta(V)}{\overset{\alpha(V)}{\rightleftharpoons}} \mathrm{O}, \tag{2.31}$$

giving

$$\frac{dn}{dt} = \alpha(V)(1 - n) - \beta(V)n, \tag{2.32}$$

or equivalently

$$\tau_n(V)\frac{dn}{dt} = n_\infty(V) - n. \tag{2.33}$$

where $n_\infty(V) = \alpha/(\alpha + \beta)$ is the equilibrium value of n and $\tau_n(V) = 1/(\alpha + \beta)$ is the time scale for approach to this equilibrium. Both n_∞ and τ_n can be determined experimentally.

2.1.6 Multiple gates

The simple model presented in Section 2.1.5 can be generalised to channels which contain multiple identical subunits, each of which can be in either the open or closed state (see figure 2.6). For example, suppose as shown that there are two such gates. If we let S_i denote the density of states with i open gates then the transition between channel states is governed by the reaction scheme

$$S_0 \underset{\beta}{\overset{2\alpha}{\rightleftharpoons}} S_1 \underset{2\beta}{\overset{\alpha}{\rightleftharpoons}} S_2. \tag{2.34}$$

The coefficients of two arise because there are two possible states with one gate open and one gate closed. Since each gate is identical we have lumped these two states into one variable S_1. Using mass action kinetics gives

$$\begin{aligned} \frac{dS_0}{dt} &= \beta S_1 - 2\alpha S_0, \\ \frac{dS_2}{dt} &= \alpha S_1 - 2\beta S_2. \end{aligned} \tag{2.35}$$

We could write down an equation for S_1 also, but this equation is superfluous since S_1 can be determined from the conservation of channels

$$S_0 + S_1 + S_2 = 1. \tag{2.36}$$

Simple substitution shows that (2.35)-(2.36) are satisfied by

$$S_0 = (1-n)^2, \quad S_1 = 2n(1-n), \quad S_2 = n^2, \tag{2.37}$$

providing

$$\frac{dn}{dt} = \alpha(1-n) - \beta n. \tag{2.38}$$

Thus n satisfies the differential equation (2.32) for a single gate, and the proportion of open channels is n^2.

Equation (2.38) is a first order equation, whereas (2.35)-(2.36) is a second order system. Thus not all solutions to (2.35)-(2.36) are of the form (2.37); only if the initial conditions are compatible with (2.37) will the solution be of this form. However, if we write

$$\begin{aligned} S_0 &= (1-n)^2 + y_0, \\ S_1 &= 2n(1-n) - y_0 - y_2, \\ S_2 &= n^2 + y_2, \end{aligned} \tag{2.39}$$

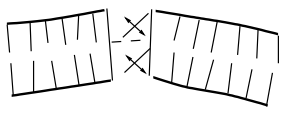

Figure 2.6: A channel with two identical gate units.

then y_0, y_2 satisfy

$$\begin{aligned} \frac{dy_0}{dt} &= -2\alpha y_0 - \beta(y_0 + y_2), \\ \frac{dy_2}{dt} &= -\alpha(y_0 + y_2) - 2\beta y_2. \end{aligned} \tag{2.40}$$

This linear system of equations has eigenvalues $-(\alpha+\beta)$, $-2(\alpha+\beta)$ so that y_0, y_2 decay exponentially to zero. Thus even if an initial condition not compatible with (2.37) is given, the solution will still approach exponentially that given by (2.37)-(2.38) (that is, (2.37)-(2.38) defines a stable invariant manifold of the full system (2.35)-(2.36)).

The analysis of a two-gated channel above generalises easily to channels containing more gates. In the case of k identical gates the fraction of open channels is n^k, where n again satisfies (2.38).

2.1.7 Non-identical gates

Often channels are controlled by more than one protein, with each protein controlling a set of identical gates, but with the gates of each protein different and independent. Consider, for example, the case of a channel with two types of gate, m and h say, each of which may be open or closed. Suppose also that there each channel has two m-gates and one h-gate. If S_{ij} denotes a channel with i open gates of type m and j open gates of type h, then the reaction scheme is

$$\begin{array}{ccccc} S_{00} & \underset{\beta}{\overset{2\alpha}{\rightleftharpoons}} & S_{10} & \underset{2\beta}{\overset{\alpha}{\rightleftharpoons}} & S_{20} \\ \gamma \uparrow\downarrow \delta & & \gamma \uparrow\downarrow \delta & & \gamma \uparrow\downarrow \delta \\ S_{01} & \underset{\beta}{\overset{2\alpha}{\rightleftharpoons}} & S_{11} & \underset{2\beta}{\overset{\alpha}{\rightleftharpoons}} & S_{21} \end{array} \quad . \tag{2.41}$$

Simple substitution shows that the rate equations are satisfied by

$$S_{00} = (1-m)^2(1-h), \quad S_{10} = 2m(1-m)(1-h), \quad S_{20} = m^2(1-h),$$

$$S_{01} = (1-m)^2 h, \quad S_{11} = 2m(1-m)h, \quad S_{21} = m^2 h, \tag{2.42}$$

so that the proportion of open channels is m^2h, providing

$$\begin{aligned} \frac{dm}{dt} &= \alpha(1-m) - \beta m, \\ \frac{dh}{dt} &= \gamma(1-h) - \delta h. \end{aligned} \tag{2.43}$$

As before, not all solutions of the rate equations (which form a fifth-order system) are solutions of the second order system (2.43). However, such solutions again form a stable invariant manifold.

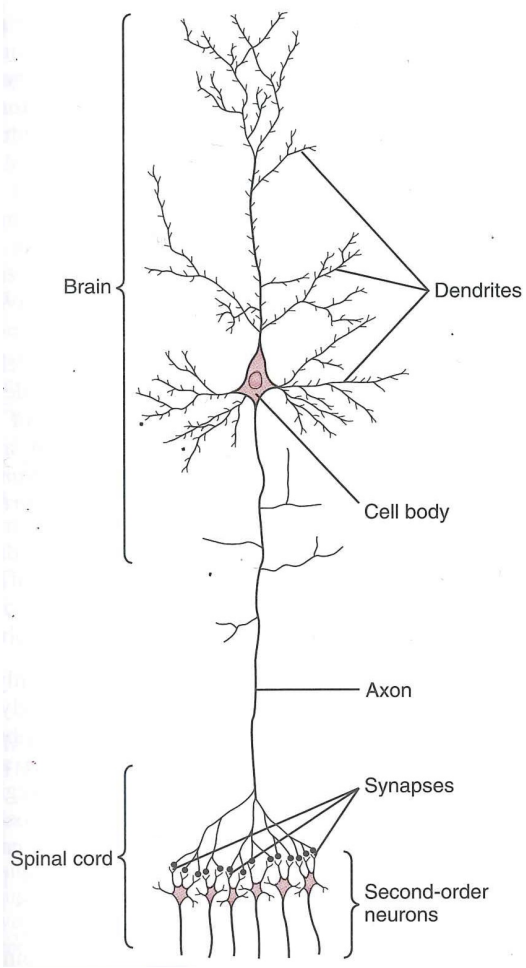

Figure 2.7: Schematic illustration of a large neuron in the brain (Guyton and Hall 2000, figure 45-1). Signals propagate from the cell body downwards along the axon, and are transmitted at the synapses to the dendrites of further neurons.

2.2 Hodgkin-Huxley model

The nervous system is a communication system formed by nerve cells called *neurons* (figure 2.7). Information is propagated along long cylindrical segments called *axons* by electrochemical signals.

An important property of neurons is *excitability*. If a small current is applied to the cell for a short time, then the membrane potential changes slightly, but returns directly to its equilibrium potential (the *resting potential*) once the applied current is removed. However, if a sufficiently large current is applied for a short time the membrane potential undergoes a large excursion (called an *action potential*) before returning to its resting value. It is by the propagation of such action potentials along the axons of neurons that signals are transmitted.

In chapter 3 we will consider the propagation of an action potentials along the

axon. In building up to that model we first consider here the Hodgkin-Huxley model of the excitability of an axon without the extra complication of spatial variation in the membrane potential. Such a situation can be realised experimentally by inserting a thin electrode along the centre of the axon, which acts to short-circuit the cell and keep the membrane potential V constant along its length (this is known as the space clamp technique).

If we suppose that we are applying an external inward current I to the cell, then the total outward current is $I_i - I$, where I_i is the outward ionic current as before. Then equation (2.29) gives

$$C_m \frac{dV}{dt} = I - I_i, \tag{2.44}$$

where C_m is the membrane capacitance. In the squid giant axon, as in many neurons, the most important ionic currents are due to the motion of sodium and potassium ions. Experimentally it is found that the potassium conductance may be modelled by a gated channel of the form described in Section 2.1.5 with an exponent of 4, that is, the conductance is $g_{\mathrm{K}} n^4$ where

$$\tau_n(V) \frac{dn}{dt} = n_\infty(V) - n. \tag{2.45}$$

While this seems to imply that the potassium channel is controlled by a protein with four identical gates, it should be noted that the exponent is determined as a reasonable fit to measured ionic currents, and not from any detailed physiological knowledge of the potassium channel itself. The equilibrium value n_∞ is found to be an increasing function of V. Thus at elevated potentials n is increased, thereby turning on, or *activating*, the potassium current. The Nernst potential for potassium is below the resting potential, so that the potassium current is an outward current at potentials greater than the resting potential. The function $n(t)$ is called the potassium activation.

For the sodium conductance, experimental data suggests that there are two processes at work, one which turns on the sodium current and one which turns it off. The Hodgkin-Huxley model assumes a channel controlled by two proteins, with a conductance of the form $g_{\mathrm{Na}} m^3 h$ where

$$\begin{aligned} \tau_m(V) \frac{dm}{dt} &= m_\infty(V) - m, \\ \tau_h(V) \frac{dh}{dt} &= h_\infty(V) - h. \end{aligned} \tag{2.46}$$

Again, while this seems to imply a channel with three m gates and one h gate, the exponents are determined as a reasonable fit to measured ionic currents, and not from a detailed knowledge of the sodium channel itself. At the resting potential m_∞ is small and h_∞ is close to one. For elevated potentials h_∞ decreases and m_∞ increases. Thus m is called the *sodium activation*, and h is called the *sodium inactivation*. The Nernst potential for sodium is above the resting potential, so that the sodium current is an inward current at potentials greater than rest.

Symbol	Meaning	Typical value
g_{Na}	Sodium conductance	120 mS cm^{-2}
g_{K}	Potassium conductance	36 mS cm^{-2}
g_{L}	Leakage conductance	0.3 mS cm^{-2}
V_{Na}	Sodium Nernst potential	56 mV
V_{K}	Potassium Nernst potential	-77 mV
V_{L}	Leakage Nernst potential	-60 mV
V_{eq}	Equilibrium potential	-70 mV
C_m	Membrane capacitance	1 μF cm^{-2}
τ_n	Potassium gate time scale	5 ms

Table 2.2: Values of the parameters of the Hodgkin-Huxley model. Unfamiliar units are S, siemens, V, volts, F, farads, and these are defined in the text.

The Hodgkin-Huxley model for the outward ionic current is thus

$$I_i = \underbrace{g_{\mathrm{Na}} m^3 h (V - V_{\mathrm{Na}})}_{\mathrm{Na}^+ \text{ current}} + \underbrace{g_{\mathrm{K}} n^4 (V - V_{\mathrm{K}})}_{\mathrm{K}^+ \text{ current}} + \underbrace{g_L (V - V_{\mathrm{L}})}_{\text{leakage}}, \tag{2.47}$$

where, apart from the sodium and potassium currents detailed above, the final term represents a 'leakage' current due to other ions, particularly chloride, Cl^-. The *resting potential* is the value of V when the outward ionic current is zero, and this is denoted as V_{eq}, and is approximately -70 mV (though this and other such values depend on the particular animal species and nerve cell under consideration). The experimentally measured values of the conductances and the Nernst potentials for the squid giant axon, as given by Hodgkin and Huxley, are given in table 2.2, and the electrical units are as follows.

The basic electrical unit is that of charge, for which the unit is a *coulomb*, denoted as C. The unit of electrical current is the *amp* A, and 1 A = 1 C s^{-1}. The unit of electric potential is the *volt*, but it easier first to define the unit of *power*, which is the *watt* W, and 1 W = 1 V A = 1 J s^{-1}, and thus the volt has units of energy per charge, thus 1 V = 1 J C^{-1}.

Subsidiary units are used for the coefficients of electrical properties, such as capacitance, resistance and conductance; the corresponding units are the *farad* F, the *ohm* Ω, and the *siemens* S, and these are defined by 1 F = 1 C V^{-1}, 1 Ω = 1 V A^{-1}, and 1 S = 1 Ω^{-1} = 1 F s^{-1}. Note that conductance is the inverse of resistance.

The experimentally measured forms of the functions k_∞ and τ_∞, $k = m, h, n$, are shown in figure 2.8, and these can be used to explain the form of the action potential. If an inward current I is applied to the membrane, then we can see from (2.44) that the membrane potential begins to rise. Now we can see that τ_m is much smaller than the other response times, and thus the sodium m-gates respond rapidly. m_∞ increases and, because $V_{\mathrm{Na}} = 56$ mV $> V_{\mathrm{eq}} = -70$ mV, the open sodium gates allow sodium ions to flood into the axon, allowing the potential to rise further via this

positive feedback effect. There is as yet nothing to stop this rise in potential, and so V increases above the sodium Nernst potential; this causes a reverse in the sodium current, but more importantly, after the longer time scale $\tau_h \sim \tau_n \sim 2$ ms (note that this response time decreases as the potential increases), the n-gates start to open and the h-gates start to close. The opening of the n-gates means that the potassium channels open, while the closing of the h-gates shuts down the sodium channels; this causes a flood outwards of potassium current (since $V > V_{\mathrm{K}}$), which decreases the potential, and it continues to decrease and overshoot the resting potential (since $V_{\mathrm{K}} < V_{\mathrm{eq}}$); the fact that then $V < V_{\mathrm{Na}}$ (which would allow inward sodium current) is irrelevant, since although the decrease of V starts to open the h-gates, the effect is offset by the rapid (small τ_m) decrease of the open m-gates, which keeps the sodium channels shut. Finally, with $V < V_{\mathrm{eq}}$, the sodium channels remain fairly well shut, but the potassium channels allow a weak inflow which carries the potential back to its resting value. An illustration of an action potential is shown in figure 2.9.

This description of the time course of an action potential shows that the membrane is *excitable*. Insofar as the space-clamped system is described by the four ordinary differential equations (2.44)-(2.46), it is clear that the resting state $V = V_{\mathrm{eq}}$, $n = n_\infty(V_{\mathrm{eq}})$, etc., is a steady state of the system; we suppose (and will show) that it is stable; but a sufficiently large perturbation (and in practice not that large) causes the potential to undergo an excursion as described above: the action potential. It is of course difficult to analyse such excursions in a four-dimensional system, but luckily we can analytically reduce the system to a two-dimensional model, which is amenable to phase plane analysis. This model is called the *FitzHugh-Nagumo model*.

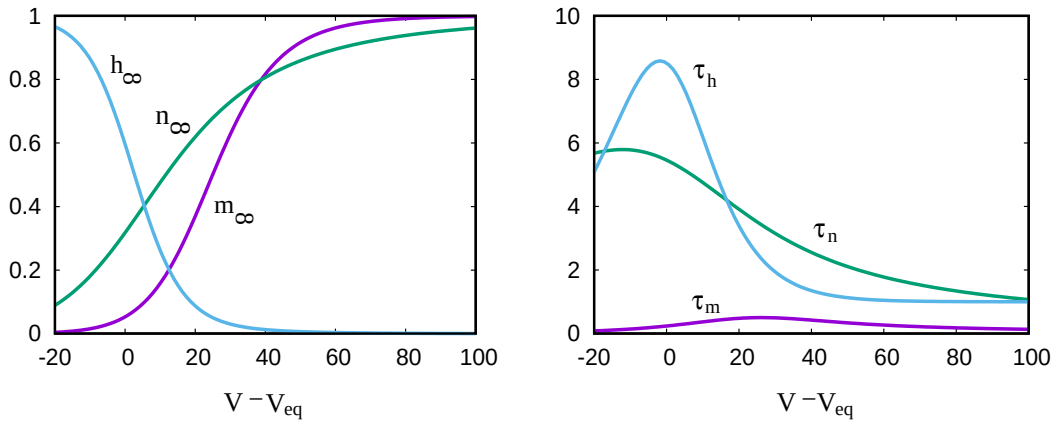

Figure 2.8: Experimentally measured values of the equilibrium gate variables (left) and their relaxation times (right) as functions of potential relative to the resting potential.

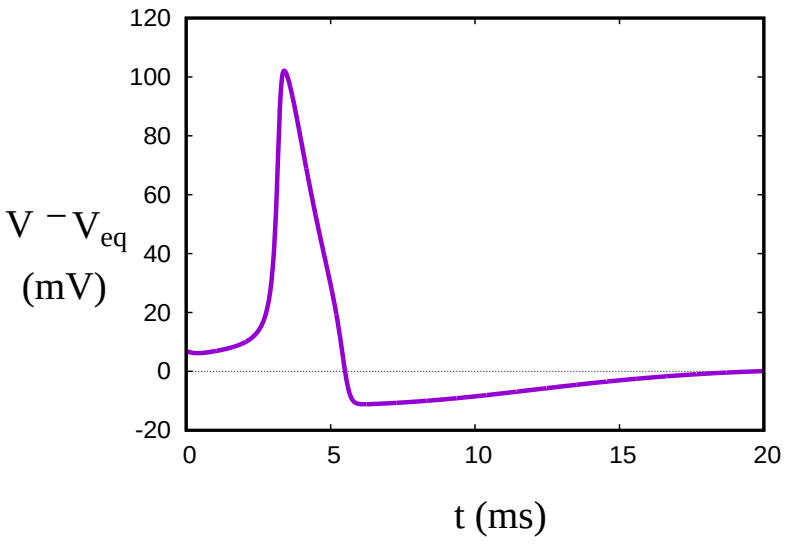

Figure 2.9: An action potential, obtained by solving the Hodgkin-Huxley model (2.44)-(2.46).

2.3 FitzHugh-Nagumo model

The FitzHugh-Nagumo model is an approximate asymptotic reduction of the Hodgkin-Huxley model, based on two key assumptions. The first is the observation, mentioned above, that

$$\tau_m \ll \tau_n, \tau_h; \tag{2.48}$$

thus $m \to m_\infty(V)$ rapidly, on a time scale of a fraction of a millisecond. As is suggested by figure 2.9, the action potential evolves over a much longer time scale of $\gtrsim 1$ ms; this can be confirmed by noting that (2.44) can be written in the form (if we ignore variation of the gate variables from their equilibrium values)

$$C_m \frac{dV}{dt} = I - \left(g_{\rm Na} m^3 h + g_{\rm K} n^4 + g_L\right)(V - V_{\rm eq}), \tag{2.49}$$

where

$$V_{\rm eq} = \frac{(g_{\rm Na} m^3 h V_{\rm Na} + g_{\rm K} n^4 V_{\rm k} + g_L V_L)}{(g_{\rm Na} m^3 h + g_{\rm K} n^4 + g_L)} \tag{2.50}$$

(which provides an implicit definition of $V_{\rm eq}$ since $m = m_\infty(V_{\rm eq})$, etc.); thus that the response time constant for V is

$$\tau_V = \frac{C_m}{g_{\rm Na} m^3 h + g_{\rm K} n^4 + g_L}. \tag{2.51}$$

To estimate this, note that at equilibrium, figure 2.8 suggests that the sodium channels are effectively shut, while $n \approx 0.3$ and thus $g_{\rm K} n^4 \approx 0.36$ mS cm^{-2}. Using the value

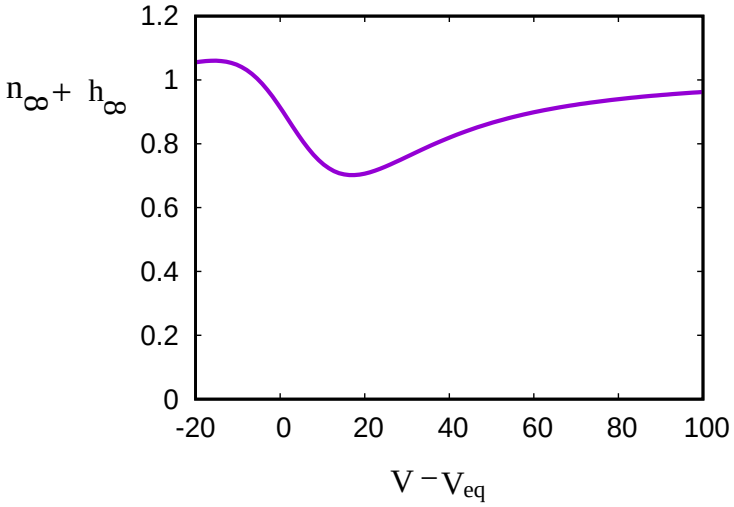

Figure 2.10: The variation of $n_\infty + h_\infty$ as a function of perturbation potential $V - V_{\rm eq}$.

for C_m in table 2.2, this suggests $\tau_V \sim 3$ ms; it follows therefore that also $\tau_m \ll \tau_V$, so that it is safe to suppose that $m = m_\infty(V)$ always.

Next we make the crude but feasible assumption that $\tau_n \approx \tau_h$. As can be seen in figure 2.8, this is not quantitatively accurate, but it is not hugely inaccurate; further, we suppose that

$$n_\infty + h_\infty = \bar{h}, \tag{2.52}$$

a constant. This is suggested by figure 2.8, and in fact we plot the corresponding graph of $n + h$ as a function of $V - V_{\rm eq}$ in figure 2.10. To be specific we will take $\bar{h} = 0.8$. Now since

$$\begin{aligned} \tau_n \frac{dn}{dt} &= n_\infty - n, \\ \tau_h \frac{dh}{dt} &= h_\infty - h, \end{aligned} \tag{2.53}$$

it follows with the two assumptions above that after an initial transient, we may take

$$n + h \equiv \bar{h}, \tag{2.54}$$

so that h may be eliminated in favour of n. The effect of all this is that the four-dimensional Hodgkin-Huxley model can be reduced to an approximately equivalent two-dimensional system,

$$\begin{aligned} C_m \frac{dV}{dt} &= I - \left\{ g_{\rm K}(V - V_{\rm K})n^4 + g_{\rm Na}(V - V_{\rm Na})m^3(V)(\bar{h} - n) + g_L(V - V_L) \right\}, \\ \tau_n(V) \frac{dn}{dt} &= n_\infty(V) - n. \end{aligned} \tag{2.55}$$

Parameter	Typical value
$v_{\rm K}^*$	0.1
$v_{\rm L}^*$	0.1
$\gamma_{\rm K}$	0.3
$\gamma_{\rm L}$	0.3×10^{-2}
ε	0.2×10^{-2}

Table 2.3: Typical values of the dimensionless parameters defined in (2.61).

Non-dimensionalisation

It is now helpful to non-dimensionalise the system. First we define the potential relative to the resting potential to be

$$v = V - V_{\rm eq}, \tag{2.56}$$

and there are corresponding values

$$v_{\rm Na} = V_{\rm Na} - V_{\rm eq} \sim 126 \text{ mV}, \quad v_{\rm K} \sim -7 \text{ mV}, \quad v_{\rm L} \sim 10 \text{ mV}. \tag{2.57}$$

We non-dimensionalise the model by scaling the variables as

$$v \sim v_{\rm Na}, \quad t \sim \tau_n, \tag{2.58}$$

and this leads to the dimensionless system

$$\begin{aligned} \frac{dn}{dt} &= n_\infty(v) - n, \\ \varepsilon\frac{dv}{dt} &= I^* - g(v,n), \end{aligned} \tag{2.59}$$

where

$$g(v,n) = \gamma_{\rm K}(v + v_{\rm K}^*)n^4 + \gamma_L(v - v_L^*) - (1-v)(\bar{h} - n)m^3(v), \tag{2.60}$$

and the dimensionless parameters are defined by

$$\begin{aligned} I^* &= \frac{I}{g_{\rm Na}v_{\rm Na}}, \quad \gamma_{\rm K} = \frac{g_{\rm K}}{g_{\rm Na}}, \quad \gamma_{\rm L} = \frac{g_{\rm L}}{g_{\rm Na}}, \\ v_{\rm K}^* &= -\frac{v_{\rm K}}{v_{\rm Na}}, \quad v_{\rm L}^* = \frac{v_L}{v_{\rm Na}}, \quad \varepsilon = \frac{C_m}{g_{\rm Na}\tau_n}. \end{aligned} \tag{2.61}$$

Typical values of these parameters based on the values in table 2.2 are given in table 2.3.

Because $\varepsilon \ll 1$, we see that v will rapidly relax to a quasi-equilibrium in which, if we take the applied current I^* to be zero, v is simply given by $g = 0$. This is in fact the v-nullcline for the pair of equations (2.59). In order to see what the nullcline

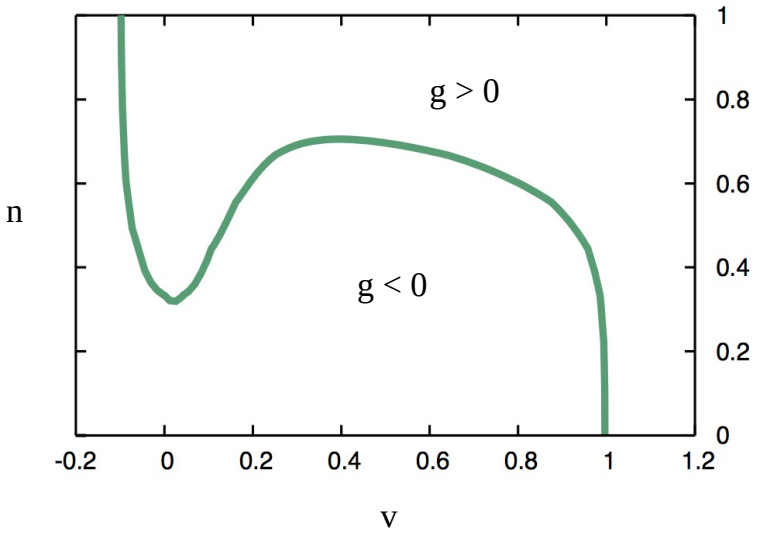

Figure 2.11: The v-nullcline $g(v, n) = 0$. This is the exact curve including the small $\gamma_{\rm L}$ term.

looks like, we first note that $\gamma_{\rm L} \ll 1$, and if we neglect the corresponding term, then (2.60) implies that the g-nullcline is approximately given by

$$\frac{n^4}{\bar{h} - n} \approx \frac{(1 - v)m^3(v)}{\gamma_{\rm K}(v + v_{\rm K}^*)}. \tag{2.62}$$

To see what this looks like, we note that as v increases towards one, the right hand side decreases towards zero. However as v decreases towards zero, m is very small so the right hand side also becomes very small; thus for $0 < v < 1$ the right hand side is a positive function with a single maximum. For $v < 0$, the right hand side increases again, tending to infinity as $v \to v_{\rm K}^*$.

This same shape is reflected in the nullcline, because the left hand side is an increasing function of n. The only difference is that as $v \to v_{\rm K}^*$, $n \to \bar{h}$ and n will reach its maximum possible value of $n = 1$ at some finite value of $v < v_{\rm K}^*$. The small term in $\gamma_{\rm L}$ does not affect this discussion, as it provides a regular perturbation near $v = 1$ and $v = -v_{\rm K}^*$, as is easy to show. Figure 2.11 shows the form of the g-nullcline. Note that at large n (for $-v_{\rm K}^* \lesssim v \lesssim 1$) g is positive, so that v is decreasing above the nullcline, and increasing below it. Thus the left and right hand branches are stable, and the intermediate branch where n increases with v is unstable.

Phase plane analysis

In order to elucidate the behaviour of the system, we need to add the other nullcline, that for n, on which $\dot{n} = 0$; this is just given by $n = n_\infty(v)$, and is shown in figure 2.12, using the original functional fits for the functions m_∞, etc., determined by

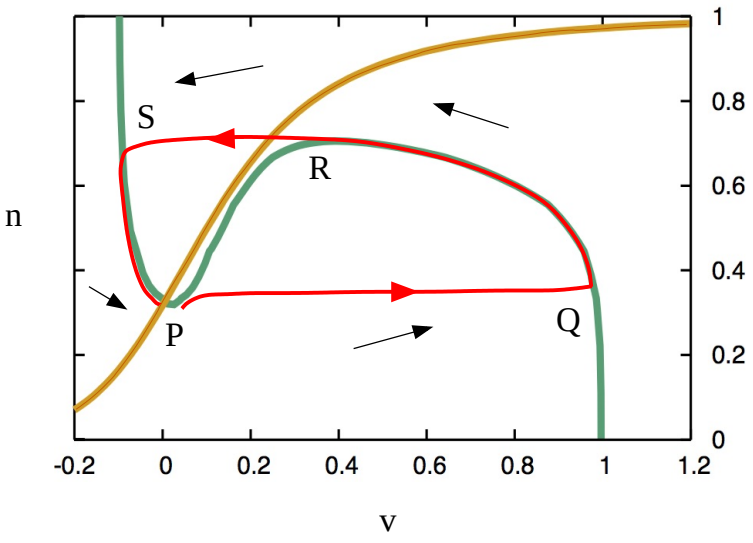

Figure 2.12: Phase plane for (2.59), together with the trajectory of an action potential.

Hodgkin and Huxley. There is a unique fixed point P, and trajectories cycle around it as indicated by the arrows. This tells us that P is not a saddle, but either a node or a focus. As is well known, the behaviour in the vicinity of the origin for the two-dimensional system $\dot{\mathbf{x}} = M\mathbf{x}$ (M is sometimes called the community matrix) is determined by $T = \operatorname{tr} M$ and $D = \det M$; specifically, $\mathbf{0}$ is a saddle if $D < 0$ and is stable (either node or spiral) if $T < 0$ and $D > 0$. Since we know already that P in figure 2.12 is not a saddle, its stability is determined by the trace of the matrix obtained by linearising (2.59) about P.

By definition, P is at $v = 0$, $n = n_\infty(0)$. We put $n = n_\infty(0) + N$ and linearise (2.59); this gives

$$\frac{d}{dt}\begin{pmatrix} N \\ v \end{pmatrix} = \begin{pmatrix} -1 & n'_\infty \\ -g_n/\varepsilon & -g_v/\varepsilon \end{pmatrix}\begin{pmatrix} N \\ v \end{pmatrix}, \tag{2.63}$$

and thus $T = -1 - g_v/\varepsilon$, so that P is stable if $g_v > -\varepsilon$.[1] Figure 2.12 suggests that the v-nullcline ($n = n_v(v)$) slope $n'_v < 0$, and thus, since $g_n > 0$ (figure 2.11), that $g_v = -n'_v g_n > 0$, and P is stable. Analytically, since $n_\infty(0)^4 \approx 10^{-2}$ while $m_\infty(0)^3 \approx 10^{-4}$, it follows that near $v = 0$,

$$g \approx \gamma_{\mathrm{K}}(v + v^*_{\mathrm{K}})n^4 + \gamma_L(v - v^*_L), \tag{2.64}$$

and thus

$$g_v \approx \gamma_{\mathrm{K}} n_\infty(0)^4 + \gamma_{\mathrm{L}} > 0 \quad \text{at} \quad v = 0. \tag{2.65}$$

[1]Figure 2.12 already tells us, purely geometrically, that P is not a saddle and thus that $D > 0$. But $D = g_v + g_n n'_\infty$ is a tremendously complicated function; how can we tell that it is positive? The answer to this that if we define the v-nullcline to be $n = n_v(v)$, then the above definition can be written in the form $D = g_n[n'_\infty - n'_v]$, and a glance at figure 2.11 shows that $g_n > 0$, while figure 2.11 shows that $n'_\infty > n'_v$ at P.

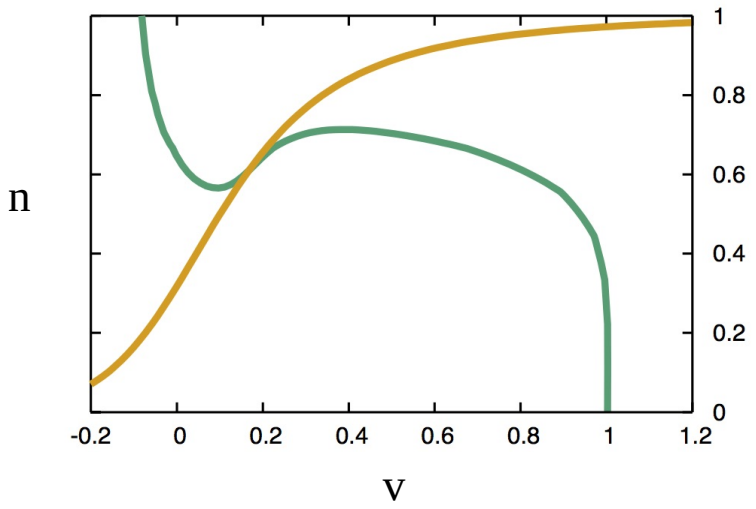

Figure 2.13: The v- and n- nullclines for the system (2.59), with $I^* = 0.005$.

Although P is stable, it is evident from figure 2.12 that a fairly small increase in v will lead to a large excursion: this is the *action potential*. v jumps rapidly to Q where $v \approx 1$, and then subsides along the green curve to R, where a second rapid transition to S occurs, overshooting the rest state; the trajectory then recovers along the green curve to P. The first part of the action potential describes depolarisation, because the resting potential is -70 mV, so that the increase causes the potential difference to drop.

Limit cycles

If we apply a current I^*, the v-nullcline is raised, as shown in figure 2.13. We then see that $n'_v > 0$, so that $g_v = -n'_v g_n < 0$, and it is possible that then there may be a range $I_- < I^* < I_+$ where the fixed point is oscillatorily unstable (if $g_v < -\varepsilon$); this would lead to a limit cycle, and thus periodic neuron firing. If we compare figures 2.12 and 2.13, we can crudely estimate that $g_n \sim 0.17$ and $n'_v \sim 0.4$, and thus (at least for $I^* = 0.005$) $g_v \sim -0.07$ and oscillations are predicted; they take the form of *relaxation oscillations* which mimic the behaviour of the red trajectory in figure 2.12, that is, they rapidly shuttle between the left and right branches of the 'slow manifold' $g = 0$. Relaxation oscillations typically have a distinctive appearance, as shown in figure 2.14. They indicate repetitive firing of the neurons.

FitzHugh-Nagumo equations

The FitzHugh-Nagumo model refers here to the reduction of the Hodgkin-Huxley model to a simplified two-dimensional system. The FitzHugh-Nagumo *equations* consist of an analytically simpler pair of equations which have the same behaviour as (2.59) due to a choice of nullclines which is topologically similar to those in figure

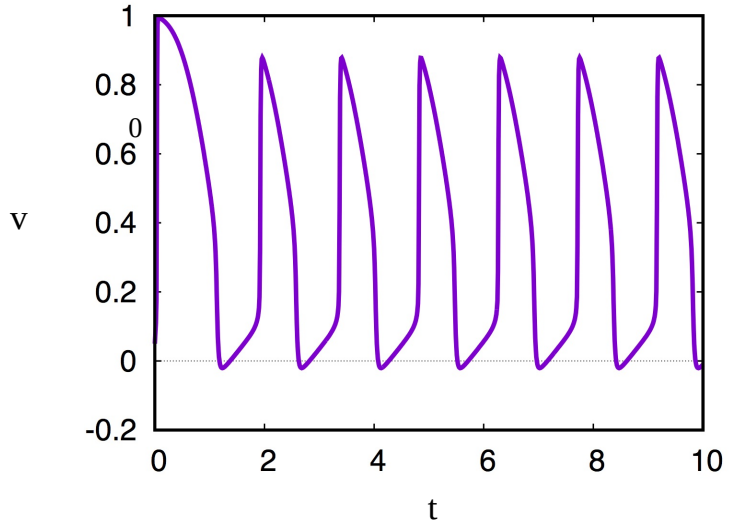

Figure 2.14: Oscillations of the reduced Hodgkin-Huxley model for $I^* = 0.005$.

2.12. We define $w = n - n_\infty(0)$, so that the rest state in (v, w) space is the origin; then the FitzHugh-Nagumo equations are

$$\begin{aligned} \varepsilon \dot{v} &= I^* + f(v) - w, \\ \dot{w} &= \gamma v - w, \end{aligned} \tag{2.66}$$

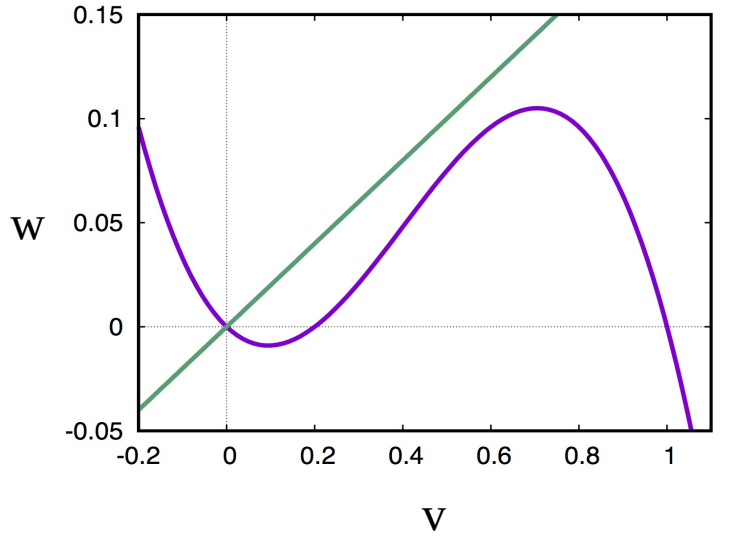

Figure 2.15: Nullclines of (2.66) and (2.67), for values $\gamma = a = 0.2$.

and $f(v)$ is commonly taken to be the cubic

$$f = v(v-a)(1-v), \quad a \in (0,1), \tag{2.67}$$

as shown in figure 2.15; the parameter γ is chosen sufficiently large that the origin is the only fixed point of the system.

Chapter 3

Wave propagation in neurons

3.1 Excitable media

We now turn to study the behaviour of the Hodgkin-Huxley model when spatial effects are allowed. What we will find is that a local action potential propagates along the axon, and this enables the transmission of signals along the nerve. Essentially, the axon acts like an electric cable, and the spatially dependent model is somewhat analogous to that for an electrical cable, which is described by the *telegraph equation*. The situation is shown in figure 3.1. The axon's internal electrical potential V is now a function of distance x along the cable and time t. The space dependence is one-dimensional because the axon is long and thin, and this causes the lateral variation of the potential in the interior to be negligible (but still different to the exterior: all the lateral resistance is in the cell membrane).

Before beginning, we need to recall some facts about electromagnetism. In an electrically conducting medium, current density $\mathbf{J}$ is related to the electric field $\mathbf{E}$ by the relation (Ohm's law)

$$\mathbf{J} = \sigma \mathbf{E}, \tag{3.1}$$

where σ is the electric conductivity. The units of $\mathbf{J}$ are A m^{-2} and those of $\mathbf{E} = -\boldsymbol{\nabla} V$ are V m^{-1}, so the units of σ are S m^{-1}. In our case, the axial current $I_\parallel$ is thus the

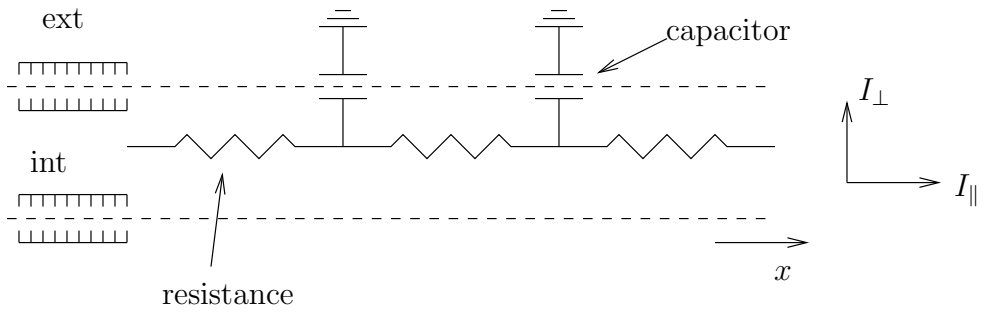

Figure 3.1: Schematic representation of the axon as a resistive cable.

current density thus defined multiplied by the cross-sectional area of the neuron,

$$I_{\parallel} = -\sigma A_i \frac{\partial V}{\partial x}, \tag{3.2}$$

and this can be written in the form

$$I_{\parallel} = -\frac{1}{R}\frac{\partial V}{\partial x}, \tag{3.3}$$

where

$$R = \frac{1}{\sigma A_i} \tag{3.4}$$

is the axial resistance, with units of Ω m^{-1}. Obviously this is not a property of the medium, but by defining the resistivity of the medium as

$$R_c = \frac{1}{\sigma}, \tag{3.5}$$

with units of Ω m, the two quantities are related by

$$R = \frac{R_c}{A_i}. \tag{3.6}$$

Equally, the membrane capacitance given in table 2.2 is a capacitance per unit area, but we need the capacitance per unit length, C; this is

$$C = p\, C_m, \tag{3.7}$$

where p is the length of the perimeter of the axon. The same applies to the ionic current; I_i defined in (2.47) is a current per unit area, but we require the transverse current per unit length, $I_{\perp}$, and this is given by

$$I_{\perp} = pI_i. \tag{3.8}$$

We now derive the cable equation for the Hodgkin-Huxley model. From first principles we consider the balance of charge in a short section of axon of length δx. The quantity of charge is $CV\,\delta x$, and its change in a short time interval δt is $C\delta V\,\delta x$, and this is due to the outward ionic current, a quantity $I_{\perp}\,\delta x\,\delta t$, and also to the axial current space increment, a quantity $\delta I_{\parallel}\,\delta t$; hence the cable equation takes the form

$$C\frac{\partial V}{\partial t} = -I_{\perp} - \frac{\partial I_{\parallel}}{\partial x}. \tag{3.9}$$

The axial current has already been defined in (3.3). Putting this and (3.9) together, we obtain the Hodgkin-Huxley cable equation in the form

$$C_m\frac{\partial V}{\partial t} = -I_i + \frac{d}{4R_c}\frac{\partial^2 V}{\partial x^2}, \tag{3.10}$$

where we have used (3.6), (3.7) and (3.8), together with the fact that for a circular cross section, the *hydraulic radius* is

$$\frac{A_i}{p} = \frac{d}{4}, \tag{3.11}$$

where d is the axon diameter. The effect of the spatial variation is simply to add a diffusion term to the equation for V. The gate equations for m, n and h remain the same, however.

Nondimensionalisation

We now non-dimensionalise the model, in the same way as in chapter 2. Specifically, we define $V - V_{\rm eq} = v$, and then scale the variables as

$$v \sim v_{\rm Na}, \quad t \sim \tau_n, \quad I_i = g_{\rm Na} v_{\rm Na}\, g(n, v), \quad x \sim l, \tag{3.12}$$

where l will be chosen below, and this leads to

$$\begin{aligned} \varepsilon \frac{\partial v}{\partial t} &= I^* - g(n, v) + \varepsilon^2 \frac{\partial^2 v}{\partial x^2}, \\ \frac{\partial n}{\partial t} &= n_\infty(v) - n. \end{aligned} \tag{3.13}$$

The length scale is then determined by the requirement that the diffusion term in (3.13) should be ε^2 (for reasons which will emerge below), and this leads to the choice

$$l = \frac{\tau_n}{2C_m} \left(\frac{d g_{\rm Na}}{R_c} \right)^{1/2}. \tag{3.14}$$

Values appropriate for the squid giant axon (most of these were given in table 2.2) are

$$\begin{gathered} d = 5 \times 10^{-2} \text{ cm}, \quad g_{\rm Na} = 120 \text{ mS cm}^{-2}, \quad R_c = 30\ \Omega \text{ cm}, \\ \tau_n = 5 \text{ ms}, \quad C_m = 1\ \mu\text{F cm}^{-2}, \end{gathered} \tag{3.15}$$

which imply that $l \sim 35$ cm! In practice, the effective length scale is less than this, because (as may be inferred from figure 2.13), although we assume $g = O(1)$, in practice it is somewhat smaller.

In more detail, figure 3.2 shows contours of g in the vicinity of the nullclines, which show that in fact $g \sim 0.1$. If we rescale $g \sim \bar{g}$, where $\bar{g} \approx 0.1$, then we find that (3.13) is regained, but with $\varepsilon \to \varepsilon/\bar{g}$, and the length l in (3.14) is rescaled with $\sqrt{\bar{g}}$.

3.2 Wave propagation in the FitzHugh-Nagumo model

We now study the propagation of an action potential for the FitzHugh-Nagumo model. Ideally we would use (3.13), but it is simpler for the sake of exposition to use instead

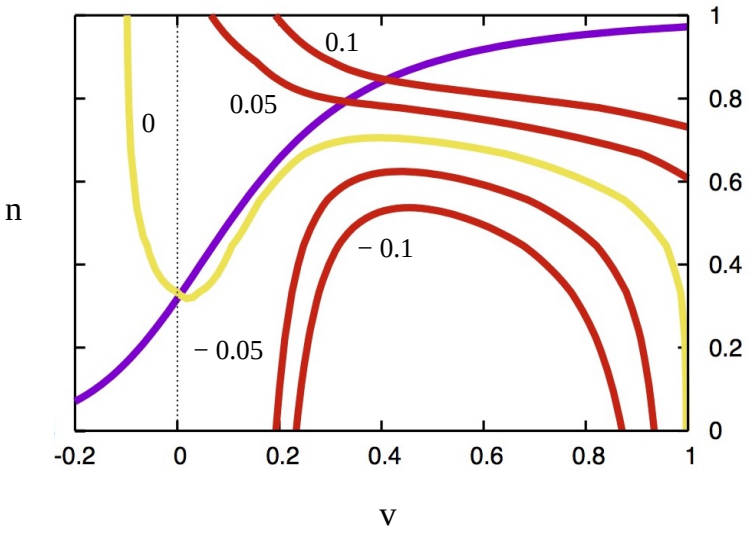

Figure 3.2: The n and v nullclines (the latter being the contour $g = 0$), together with contours of $g = -0.1, -0.05, 0.05, 0.1$ as indicated.

the spatially dependent version of (2.66), which can be written in the form

$$\begin{aligned} \varepsilon v_t &= f(v) - w + \varepsilon^2 v_{xx}, \\ w_t &= \gamma v - w, \end{aligned} \tag{3.16}$$

and $f(v)$ is a cubic, for example

$$f(v) = v(v-a)(1-v). \tag{3.17}$$

Here w masquerades as $n - n_\infty$.

The idea is that if a stimulus is applied at a point on the axon, then the potential there will undergo the action potential excursion of figure 2.9 or figure 2.12. The principal apparent effect of diffusion is then to raise the potential *ahead* of this point, which then causes a new action potential to begin there; in this way the action potential propagates its way down the axon. This scheme is illustrated in figure 3.3, where the wave representing the action potential propagates to the right. If a stimulus were applied in the middle of an axon, this theory would predict that a pair of waves would be set up (just as occurs in the Fisher equation), since the model has reflectional symmetry, but in reality signals propagate along the axon from the cell body, sometimes called the soma; this can be associated with the appropriate boundary conditions at either end of the axon: the synapses allow incoming signals to the dendrites but not *vice versa*.

As well as initiating an action potential, diffusion has a controlling effect on its spatial form, as we shall find below: in particular, the excursion from the rest state back to itself is different from that shown in figure 2.12. Just as for the Fisher

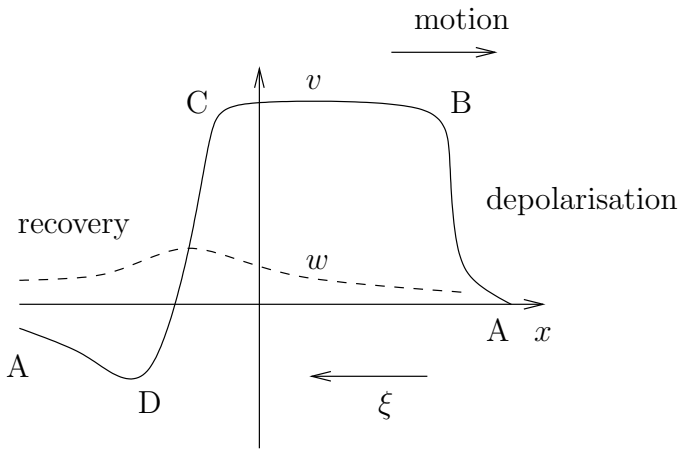

Figure 3.3: The form of the travelling wave of the action potential.

equation, an arbitrary disturbance at the cell body end of the axon will lead to a travelling wave being formed, although this needs to be confirmed numerically. We therefore look for a solution in which

$$v = v(\xi), \quad w = w(\xi), \quad \xi = ct - x, \tag{3.18}$$

as shown in figure 3.3; we presume the wave speed $c > 0$. With this assumption, the model reduces to the pair of ordinary differential equations

$$\begin{aligned} \varepsilon c v' &= f(v) - w + \varepsilon^2 v'', \\ c w' &= \gamma v - w, \end{aligned} \tag{3.19}$$

where the primes denote derivatives with respect to ξ. Since we seek a solitary wave in which the solution returns to the rest state at each end, the appropriate boundary conditions for (3.19) are

$$v = w = 0 \quad \text{at} \quad \pm\infty. \tag{3.20}$$

This problem is harder to solve than the equivalent second order differential equation which is appropriate for the Fisher equation, and the reason for this is that (3.19) is a *third* order system. Although it is not inconceivable to do phase portrait analysis in three (or more!) dimensions, it is very difficult. At least one reason for this is that in two dimensions, solution trajectories for non-autonomous systems cannot intersect themselves, so that there is a strong geometric constraint on where the trajectories can go. This is not true in three dimensions: it is still true that trajectories cannot intersect, but they can wind around indefinitely without self-intersecting, as is familiarly the case for chaotic trajectories.

In the present case, we are able to circumvent this difficulty by dividing the trajectory into four segments, each of which lives in a different two-dimensional section of the three-dimensional phase space. The reason we are able to do this is because of

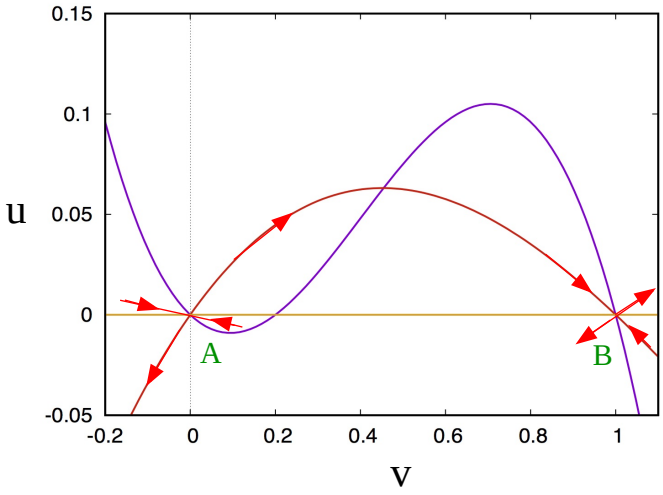

Figure 3.4: The passage from A to B.

the parameter $\varepsilon \ll 1$. The wave indicated in figure 3.3 consists of four segments: AB, BC, CD, DA. As suggested in the figure, AB and CD are fast segments in which it is appropriate to rescale the independent variable as $\xi = \varepsilon X$, and the rescaled form of (3.19) is then

$$\begin{aligned} cv' &= f(v) - w + v'', \\ cw' &= \varepsilon(\gamma v - w); \end{aligned} \tag{3.21}$$

here the primes refer to X derivatives, and we will anticipate that the unknown wave speed $c = O(1)$. We will consider the four segments of the wave indicated in figure 3.3 in turn.

(i) The first part of the wave is the passage from A to B. This is a fast segment, and thus described by (3.21). We see that to leading order, w is constant and thus zero, and the passage from A to B is described by the second order equation

$$cv' = f(v) + v'', \tag{3.22}$$

and this is amenable to phase plane analysis. We define $u = v'$, so that (3.22) can be written as the pair

$$\begin{aligned} v' &= u, \\ u' &= cu - f(v). \end{aligned} \tag{3.23}$$

As we assume $c > 0$, the phase plane is as shown in figure 3.4. There are three equilibria, which are the zeros of $f(v)$, with $u = 0$. Simply by ascertaining the directions of the trajectories away from the nullclines, it is straightforward to

see that A ($v = 0$) and B ($v = 1$) are saddles. The stability of the intermediate fixed point, which is evidently a node or a spiral, is determined by linearising about it, thus $v = v^* + V$, ($v^* = 0.2$ in figure 3.4), so that the linearised system is

$$\begin{pmatrix} \dot{V} \\ u \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -f' & c \end{pmatrix} \begin{pmatrix} V \\ u \end{pmatrix}, \tag{3.24}$$

whose stability is determined by the trace of the matrix, which is $c > 0$. Thus the intermediate fixed point is an unstable node or spiral, so that the separatrix from A extends past this fixed point, and the transition is effected by connecting the unstable separatrix from A to the stable separatrix to B.

But generally such a connection will not occur. This is where the unknown wave speed c comes in. The idea is that we can choose c such that the unstable separatrix from A actually lands on B, as shown in figure 3.4. It is quite straightforward to prove this using a comparison argument; the same method can be used to prove that the travelling wave of the Fisher equation is monotonic.

We write (3.23) in the form

$$\frac{du}{dv} = c - \frac{f(v)}{u}, \tag{3.25}$$

with initial condition

$$u \sim \lambda_+ v \quad \text{as} \quad v \to 0, \tag{3.26}$$

where

$$\lambda_+ = \tfrac{1}{2}[c + \{c^2 + 4|f'(0)|\}^{1/2}], \tag{3.27}$$

which is the initial condition for the unstable separatrix from A. Now we consider the family of such solutions $u(v, c)$ as c varies. Since λ_+ is an increasing function of c, it is clear that if $c_1 < c_2$, then at least for small v, the corresponding solutions satisfy $u_1 < u_2$. Next, suppose that $u_1 = u_2$ for some $v > 0$. Then at this value, we must have $u_1' \geq u_2'$ and thus from (3.25), $c_1 \geq c_2$, in contradiction to our assumption. Thus in fact no such intersection can occur, and this proves that u increases monotonically with c.

Next we consider the limits as $c \to \infty$ and $c \to 0$. As $c \to \infty$, we have firstly $u \sim cv$, and the consequent correction

$$u \approx cv - \frac{1}{c} \int_0^v \frac{f(v')\, dv'}{v'}, \tag{3.28}$$

and evidently $u \to \infty$ as $v \to \infty$. On the other hand, as $c \to 0$, the limiting trajectory satisfies (3.25) with $c = 0$, and thus

$$\tfrac{1}{2}u^2 = -\int_0^v f(v')\, dv'. \tag{3.29}$$

We see in figure 3.4 that f is initially negative, so that the right hand side of (3.29) is positive, and (*providing* $\int_0^1 f(v)\, dv > 0$) u defined by (3.29) reaches zero

before $v = 1$; this, together with the large c behaviour and the monotonicity result, shows that there is a unique value of c for which the separatrix from A connects to that into B providing $\int_0^1 f(v)\,dv > 0$, which for the FitzHugh-Nagumo cubic in (3.17) requires $a < \frac{1}{2}$. It is evident from (2.14) that this applies for the Hodgkin-Huxley model as well.

(ii) When the trajectory reaches B the system switches to a slow phase, in which the appropriate scaling is as in (3.19). Approximately, we have

$$w = f(v), \quad cw' = \gamma v - w, \tag{3.30}$$

and as indicated in figure 3.5, this causes a slow migration up the 'slow manifold' $w = f(v)$ on which w increases since $w < \gamma v$. Since $u = \varepsilon v_\xi$, u remains close to zero.

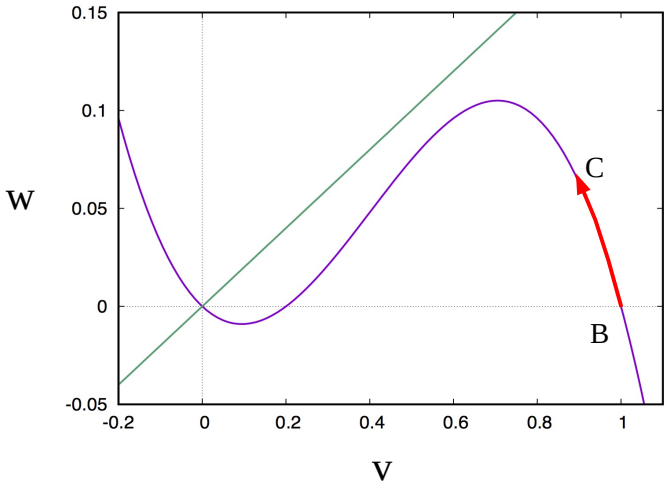

Figure 3.5: The passage from B to C.

In the purely time-dependent system, the action potential climbs up the slow manifold all the way to the top, but for the travelling wave, this is not the case, as we shall see. We suppose that w reaches a value w_C, which is yet to be determined. Evidently this is positive, and less than (or equal to) the maximum of $f(v)$.

(iii) Next we propose another fast phase in which (3.21) is appropriate, but now $w \approx w_C$ is constant, and thus (3.22) is modified to

$$cv' = f(v) - w_C + v'', \tag{3.31}$$

and the phase plane system (3.23) is modified to

$$\begin{aligned} v' &= u, \\ u' &= cu - [f(v) - w_C]. \end{aligned} \tag{3.32}$$

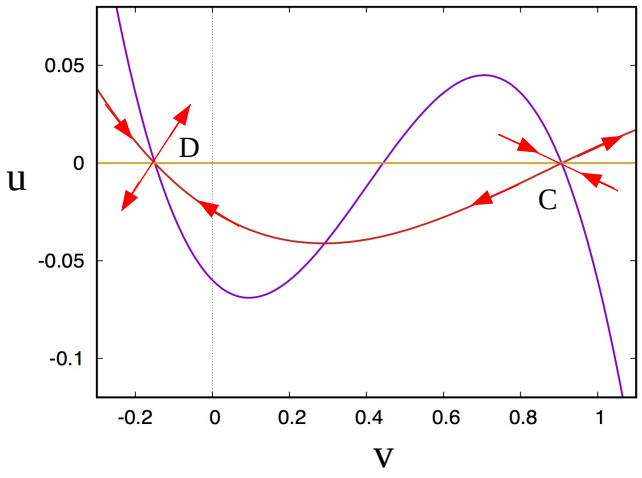

Figure 3.6: The passage from C to D.

The phase plane for this system is shown in figure 3.6. This is similar to figure 3.4, with the exception that the v-nullcline is lower. The object is to find a trajectory from the slow equilibrium at C (which is at $u = 0$, $w = f(v) = w_C$) to that at D. As for AB, both are saddles, and in general they will not connect. This is where the choice of w_C comes in. The idea, analogous to that for the wave speed selection in the transition from A to B, is to choose w_C in order to effect the connection. The proof that there is a unique w_C which enables this connection is similar to that used to show that there is a unique value for which A connects to B, and it is consigned to an exercise.

(iiii) Given that there is indeed a value of w_C such that C connects to D in figure

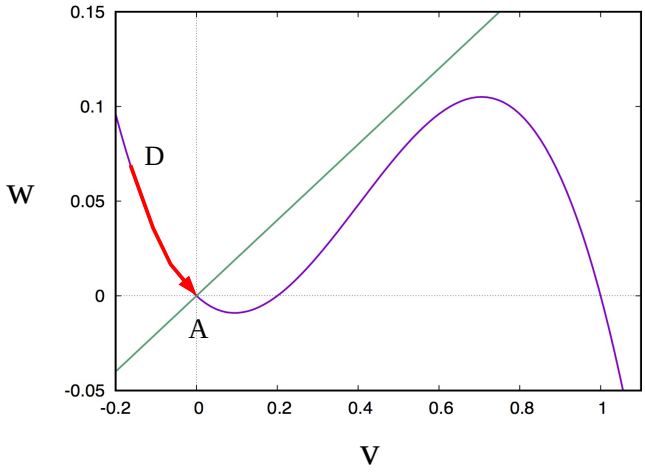

Figure 3.7: The passage from D to A.

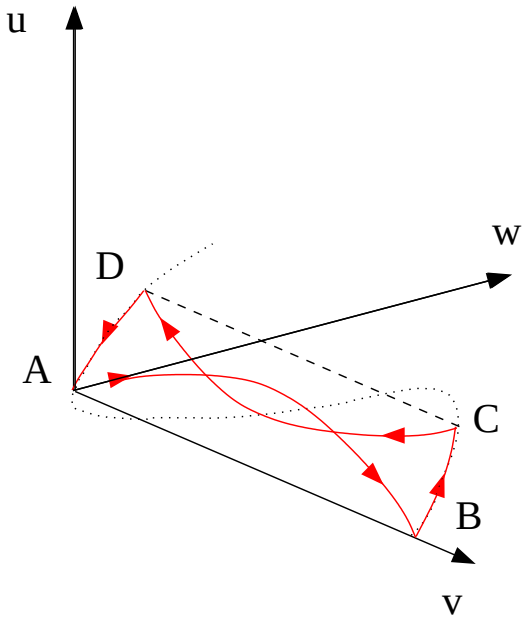

Figure 3.8: A representation of the trajectory of the wave in the (u, v, w) phase space.

3.6, it remains to connect D back to A in figure 3.7. This is again effected by using the slow manifold equations (3.30), but now since the initial point D lies above the w-nullcline, the motion is that of a slow relaxation back towards the rest state. This completes the description of the form of the travelling wave.

We see that in the (u, v, w) phase space, the fast trajectory AB is on the plane $w = 0$, while the slow trajectory BC is on the plane $u = 0$. Similarly the fast trajectory CD is on the plane $w = w_C$ (and $u < 0$) and the concluding slow trajectory is on the plane $u = 0$. A schematic representation in the three-dimensional phase space of the trajectory is shown in figure 3.8.

Chapter 4

Calcium dynamics

So far the story has been largely about sodium and potassium, but another important ion is that of calcium, Ca^{2+}. Calcium in sea water has a concentration similar to that of potassium, around 10 mM.[1] The typical extracellular concentration is lower than this, around 1 mM, but this is much higher than the normal intracellular concentration, around 10^{-4} mM; the difference is maintained in similar fashion to that of other ions.

Intracellular calcium is important in a number of physiological systems, for example in muscle contraction and cardiac signalling. Most of the calcium in the body is stored in bones, whence it is released by hormonal stimulation to maintain the extracellular Ca^{2+} concentration. However, because intracellular Ca^{2+} concentrations are so low, control mechanisms exist both to maintain them at such levels, but also to enable rapid release when required. Both functions are enabled by storing intracellular calcium in various compartments. The one we will focus on is the the *sarcoplasmic reticulum* (SR) (found in skeletal and cardiac muscle cells). Other *calcium-sequestering compartments* store calcium in non-muscle cells.

Skeletal muscle consists of bundles of muscle fibres, which are cells containing numerous filamentary structures called *myofibrils*, as shown in figure 4.1. The myofibrils themselves contain myosin and actin filaments, and it is the action of calcium on these which causes muscles to contract. So in normal circumstances, the intracellular fluid matrix (called the *sarcoplasm* in muscle cells) must maintain the low level of calcium (cramp is an example of when this control fails), but also must be able to release calcium rapidly from the internal stores when stimulated to do so (and pump it back rapidly when the stimulus ceases).

The way in which the muscle fibres are activated by nervous stimulation is by the release of *acetylcholine*, a neurotransmitter, which causes membrane channels to open and sodium to flood in to the cells; this causes an action potential to travel along the muscle fibre much as they do in nerve cells, and it is this which causes the release of calcium from the internal stores. The precise mechanism by which this release occurs is a subject of ongoing research (there are numerous different mechanisms at play),

[1] Recall that 1 M (molar) is 1 mole l^{-1}.

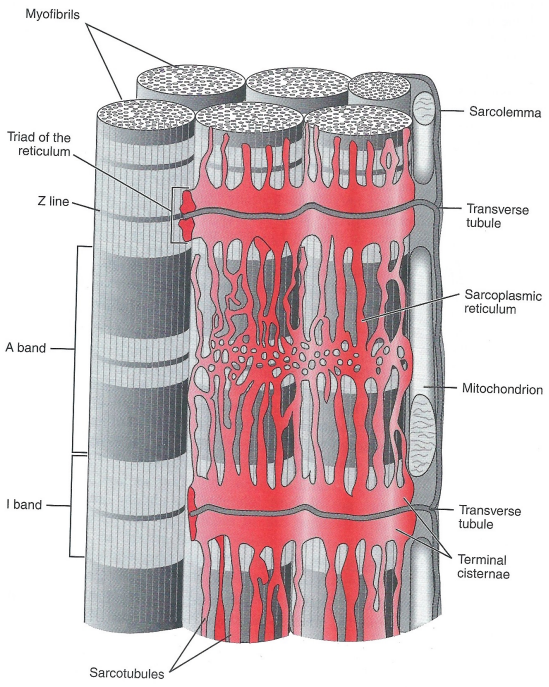

Figure 4.1: Diagram representing the sarcoplasmic reticulum in a muscle cell (Guyton and Hall 2000). Despite appearances, the SR lies within the cell membrane (the sarcolemma) and sheathes the constituent myofibrils.

and we shall limit ourselves to the description of one particular mechanism, that of *calcium-induced calcium release*.

4.1 Calcium-induced calcium release

One way in which the release of intracellular calcium can occur is called *calcium-induced calcium release* (CICR), and it forms the principal focus of this chapter. When the sarcolemma (the muscle cell membrane) is stimulated by the agonist acetylcholine, channels open and as well as sodium, this allows calcium in. The release may then occur due to the calcium sensitivity of the internal store (the SR). Such releases and uptake are enabled by differing kinds of receptors, called IP_3 (inositol (1,4,5)-triphosphate) receptors and ryanodine receptors (RyR), located on the SR. So what we have in mind is that there should be a rest state in which sarcoplasmic Ca^{2+} is low and SR Ca^{2+} is high, but that under a stimulus, this rest state is excitable.

4.1.1 Intracellular oscillations

In certain circumstances, experimental work has shown that periodic oscillations in the sarcoplasmicic Ca^{2+} can occur. Some examples are shown in figure 4.2, and have

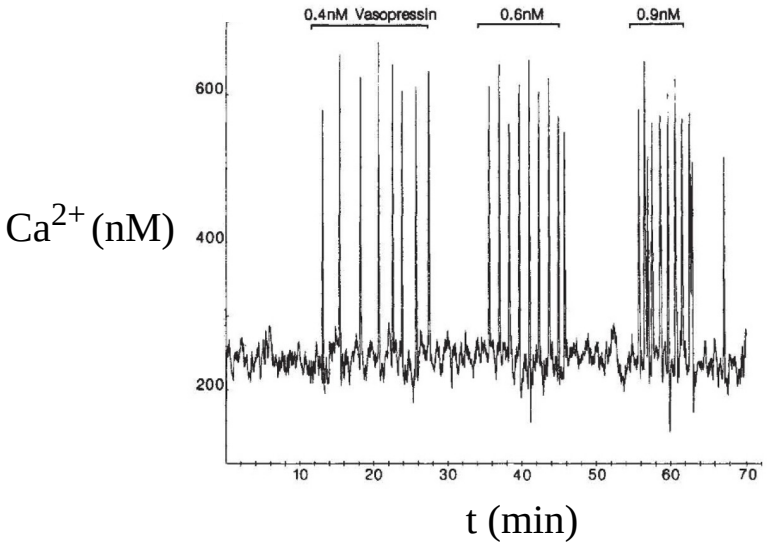

Figure 4.2: Oscillations of Ca^{2+} in a rat liver cell (hepatocyte) under stimulation by vasopressin, from Woods *et al.* (1986). Increasing stimulation (left to right) leads to a decrease in the period.

a typical spiky form, with periods which are quite variable, but generally of the order of tens to hundreds of seconds. A number of such examples have been documented from various sources by Berridge and Galione (1988). The one illustrated in the figure shows the interesting feature that the frequency increases noticeably as the concentration of the effecting agonist (here the hormone vasopressin) is increased, whereas the amplitude remains essentially the same. A further feature of such oscillations is that they do not occur if the stimulus is too high or too low. Both of these features are ones which we would like a model of the process to be able to explain.

4.1.2 The two pool model

The model we will build is based on the above description, and is illustrated in figure 4.3; it is called the two-pool model, as it describes the interaction between two internal calcium pools, one of which is the sarcoplasm itself, and the other is the Ca^{2+} sensitive SR.

A hormone, called an agonist because it *acts* on the cell membrane, causes the opening of channels at the cell membrane, and these admit calcium, which diffuses through the sarcoplasm to IP_3 and RyR receptors on the SR, which has the effect of opening gates which allow the release of Ca^{2+}. The key to the model we now describe lies in the prescription of two transport terms: J_+, the rate at which sarcoplasmic Ca^{2+} is taken up by the SR, and J_-, the rate at which the SR releases its internal store.

To build a model to describe the interactions indicated in figure 4.3, we define c

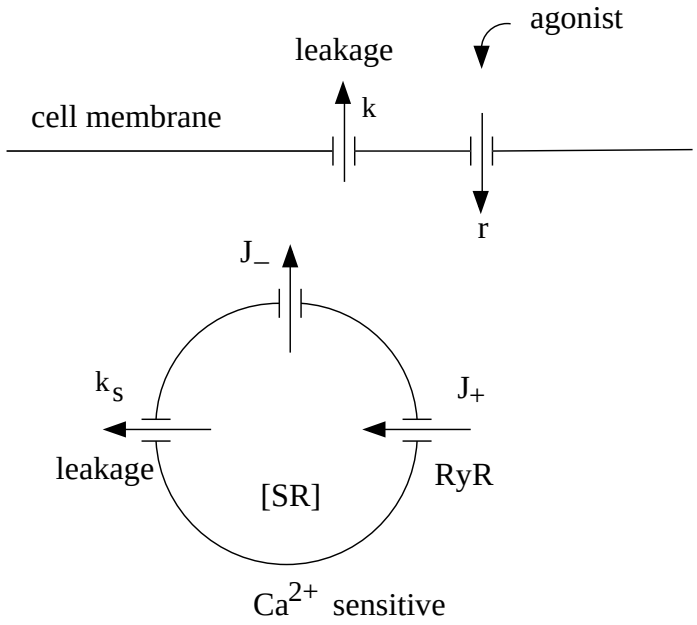

Figure 4.3: The two-pool model.

to be the concentration of Ca^{2+} in the sarcoplasm, and c_s to be the concentration of Ca^{2+} in the Ca^{2+} sensitive SR. The effect of the agonist is simply modelled as an influx rate of Ca^{2+} from the extracellular fluid, and there is an equivalent pumping out term (effectively a leakage). The influx term is denoted by r, and represents the experimental effect of the applied stimulus. As mentioned, we denote the uptake rate of Ca^{2+} by the ryanodine receptors as J_+, and the calcium-induced calcium release rate from the Ca^{2+} sensitive store as J_-. Finally, and importantly, there is a leakage term from the SR. As we shall see, this term and the sarcoplasmic leakage term are of crucial importance.

The model thus enunciated is described by the system

$$\begin{aligned} \frac{dc}{dt} &= r - kc - F, \\ \frac{dc_s}{dt} &= F, \end{aligned} \tag{4.1}$$

where F is the calcium flux into the calcium sensitive store, and is given by

$$F = J_+ - J_- - k_s c_s. \tag{4.2}$$

It remains to choose specific forms for the uptake and output from the Ca^{2+} sensitive store. We choose these to be given by Hill functions

$$\begin{aligned} J_+ &= \frac{V_1 c^n}{K_1^n + c^n}, \\ J_- &= \left(\frac{V_2 c_s^m}{K_2^m + c_s^m}\right)\left(\frac{c^p}{K_3^p + c^p}\right). \end{aligned} \tag{4.3}$$

The choice of J_+ is unremarkable, as is the first parenthesis of the J_- term. It is the second part which induces the CICR. It states that the release from the SR *increases* as the external cytosolic concentration increases, i. e., Ca^{2+} stimulates its own release. Clearly this acts in the opposite direction to what one would expect from diffusion (higher cytosolic Ca^{2+} would diminish the gradient flow from the SR), and thus is a positive feedback term, and it is because of this that we shall find that oscillations can occur.

Nondimensionalisation

The first task is to non-dimensionalise the equations. To do this, we define

$$c = K_1 u, \quad t \sim \frac{1}{k}, \quad c_s = K_2 v, \quad F = V_2 f, \tag{4.4}$$

which choice is made to sensibly balance various terms in the equations, and this leads to the dimensionless system

$$\begin{aligned} \dot{u} &= \mu - u - \frac{\gamma}{\varepsilon} f(u,v), \\ \dot{v} &= \frac{1}{\varepsilon} f(u,v), \end{aligned} \tag{4.5}$$

where

$$f = \beta \left(\frac{u^n}{1+u^n} \right) - \left(\frac{v^m}{1+v^m} \right) \left(\frac{u^p}{\alpha^p + u^p} \right) - \delta v. \tag{4.6}$$

The dimensionless parameters in this model are defined by

$$\begin{gathered} \mu = \frac{r}{kK_1}, \quad \gamma = \frac{K_2}{K_1}, \quad \varepsilon = \frac{kK_2}{V_2}, \\ \alpha = \frac{K_3}{K_1}, \quad \beta = \frac{V_1}{V_2}, \quad \delta = \frac{k_s K_2}{V_2}. \end{gathered} \tag{4.7}$$

Representative experimentally derived values of the parameter values are taken to be

$$\begin{gathered} k = 10 \text{ s}^{-1}, \quad K_1 = 1 \ \mu\text{M}, \quad K_2 = 2 \ \mu\text{M}, \quad K_3 = 0.9 \ \mu\text{M}, \\ V_1 = 65 \ \mu\text{M s}^{-1}, \quad V_2 = 500 \ \mu\text{M s}^{-1}, \quad k_s = 1 \text{ s}^{-1}, \\ m = 2, \quad n = 2, \quad p = 4, \end{gathered} \tag{4.8}$$

and these suggest the following estimates for the dimensionless parameters:

$$\alpha \sim 0.9, \quad \beta \sim 0.13, \quad \gamma \sim 2, \quad \delta \sim 0.004, \quad \varepsilon \sim 0.04. \tag{4.9}$$

We do not suggest an estimate for μ because this represents the size of the stimulus and is under experimental control, but it is natural to suppose that interesting behaviour will occur when $\mu = O(1)$, and we will assume this. Note that $\mu = 1$ if $r = 10 \ \mu\text{M s}^{-1}$.

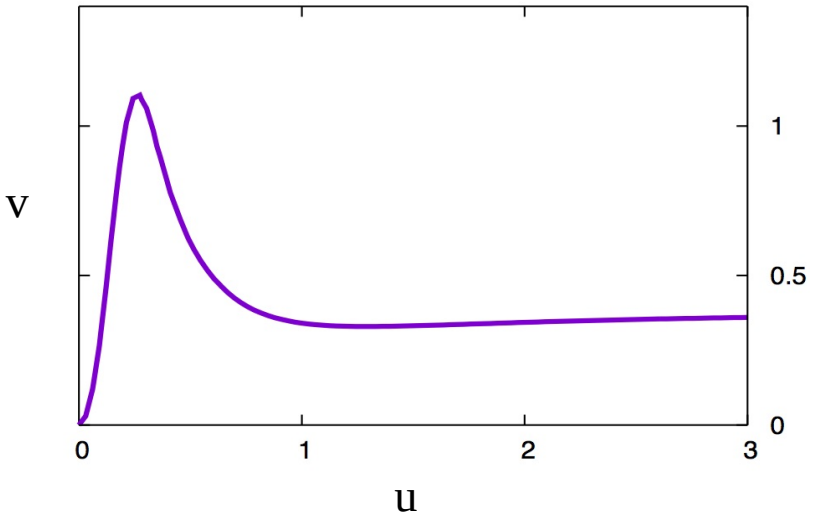

Figure 4.4: The v-nullcline of (4.5), given by (4.10).

Phase plane analysis

Being a two-dimensional system, the dynamics can be investigated by using phase plane analysis, and to this end we need to analyse the shape of the nullclines. Additionally the fact that $\varepsilon \ll 1$ means that the principal dynamics are controlled by the slow v-nullcline, because we expect v to rapidly approach it. This is treated in the following section.

The v-nullcline is given by $f = 0$, and thus

$$\beta\left(\frac{u^n}{1+u^n}\right) - \left(\frac{v^m}{1+v^m}\right)\left(\frac{u^p}{\alpha^p+u^p}\right) - \delta v = 0, \tag{4.10}$$

and is shown in figure 4.4. It is instructive to explain the form of this curve. We can do this using the fact that $\delta \ll 1$. Ignoring the term in δ, (4.10) can be written in the form

$$L(v) \equiv \frac{v^m}{1+v^m} \approx \frac{J(u)}{K(u)}, \tag{4.11}$$

whence

$$v = \left[\frac{J(u)}{K(u)-J(u)}\right]^{1/m}, \tag{4.12}$$

in which

$$J(u) = \frac{\beta u^n}{1+u^n}, \quad K(u) = \frac{u^p}{\alpha^p+u^p}. \tag{4.13}$$

Now, as shown in figure 4.5, J and K are monotonically increasing sigmoidal functions, with $J/K \sim u^{n-p}$ as $u \to 0$. With the important assumption that $n < p$, it follows that $J/K \to \infty$ as $u \to 0$. As $u \to \infty$, $J/K \to \beta$, but because $J = \beta + O(u^{-n})$,

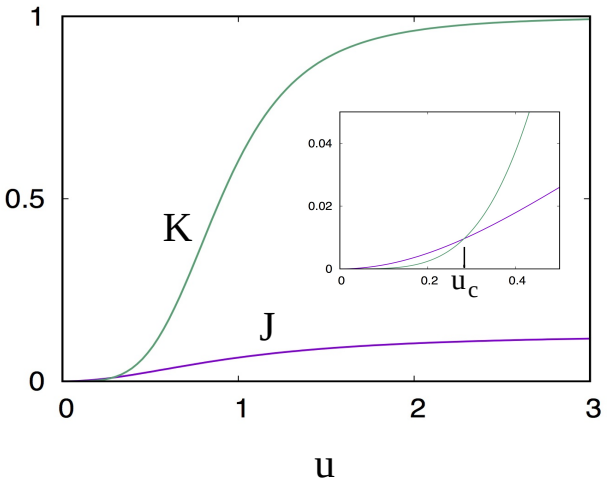

Figure 4.5: Behaviour of the functions J and K defined by (4.11). The inset shows a close-up near the origin.

while $K = 1 + O(u^{-p})$, we see that J/K increases towards its limit. Therefore, J/K decreases from infinity to a (positive) minimum, and then increases to its far field asymptote.

Because $L(v)$ is monotonic in v, $v(u)$ defined by (4.12) has a similar shape, and this is illustrated in figure 4.6. There is one significant difference between the graphs of $v = L^{-1}\{J(u)/K(u)\}$ and $L(v) = J(u)/K(u)$ as functions of u, which is due to the fact that as $v \to \infty$, $L \to 1$. At the point u_c (figure 4.5) where $J = K$, $v(u)$ given by (4.12) tends to infinity. The approximation (4.12) thus only applies for $u > u_c$.

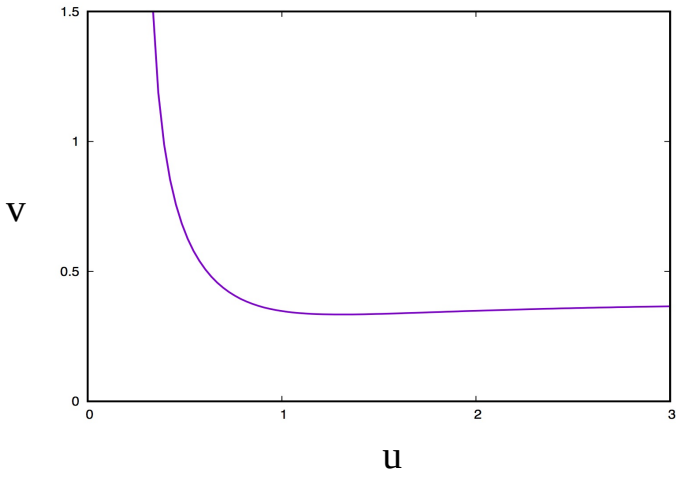

Figure 4.6: The function $v(u)$ defined by (4.12), using $\alpha = 0.9$, $\beta = 0.13$, $m = n = 2$, $p = 4$.

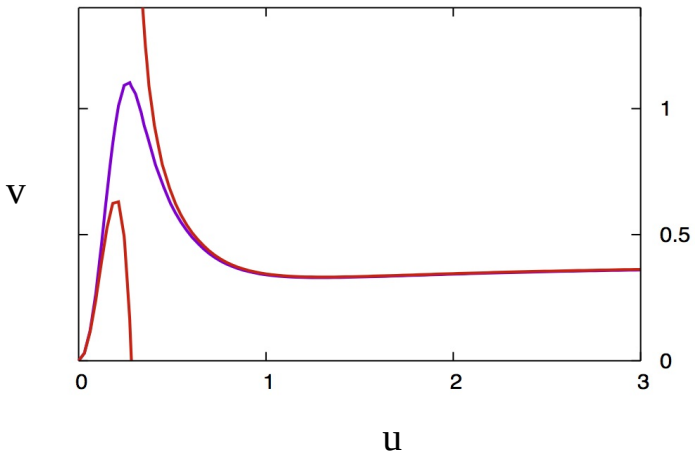

Figure 4.7: Small-v and large-v approximations to the v-nullcline in $u < u_c$ and $u > u_c$, together with the exact curve.

What then happens for $u < u_c$? The key thing to notice is that the assumption that the term in δ in (4.10) is negligible is only valid if v is bounded, and this assumption becomes invalid on the asymptote in figure 4.6. So we need to consider a different approximation when v is large, and specifically of $O(1/\delta)$. We thus define

$$v = \frac{V}{\delta}, \tag{4.14}$$

so that $L(v) \approx 1$, and (4.10) reduces to the new approximation

$$V \approx J(u) - K(u). \tag{4.15}$$

We can see from figure 4.5 that $V > 0$ only for $u < u_c$, so that this approximation works precisely when the earlier one does not.

Figure 4.7 shows both the original approximation and the large-v approximation, together with the exact curve. Two things stand out. The first is that although the two approximating curves are disjoint, the actual nullcline is continuous. The second is that the large-v approximation is rather poor, and indeed for the parameters used in (4.9), v is not even larger than one!

The reason for this second observation is easily understood from figure 4.5. The value of $u_c \approx 0.28$, but in fact this value (which we might think to be $O(1)$) is actually small, in the sense that $J = K \approx 0.0096$ there, and in fact V given by (4.15) has a maximum at $u \approx 0.199$ which is just 0.00257. We would actually need δ to be much smaller than this in order for the large-v approximation to be a good one.

The first observation is accommodated by the fact that when $u \approx u_c$, both approximations break down, and there is a *distinguished limit* in which elements of both approximations are appropriate. In this overlap region, we suppose that both $v \gg 1$

and $u \approx u_c$. Expanding (4.10) in this region, and noting that $J(u_c) = K(u_c)$, we find

$$J'(u_c)(u-u_c) + \ldots + \frac{K(u_c)}{v^m} - K'(u_c)(u-u_c) + \ldots - \delta v = 0, \tag{4.16}$$

and thus at leading order in the overlap region,

$$u - u_c \approx \frac{1}{\{K'(u_c) - J'(u_c)\}} \left\{ \frac{K(u_c)}{v^m} - \delta v \right\}. \tag{4.17}$$

It is clear that the overlap region is where we have the distinguished limit

$$v \sim \frac{1}{\delta^{\frac{1}{m+1}}}, \quad u - u_c \sim \delta^{\frac{m}{m+1}}. \tag{4.18}$$

We leave it as an exercise to show that this approximation matches to the two other approximations, and thus provides a suitable approximation which can describe how they join up.

4.1.3 Relaxation oscillations

We now return to the dynamics of the model (4.5), which we can write in the form

$$\begin{aligned} \dot{u} + \gamma \dot{v} &= \mu - u, \\ \varepsilon \dot{v} &= f(u, v). \end{aligned} \tag{4.19}$$

Since $\varepsilon \ll 1$, it follows that v rapidly approaches the v-nullcline, but the novelty here is that this approach is along lines $u + \gamma v =$ constant. The resultant motion on the slow manifold $f = 0$ then depends on the value of the stimulus parameter μ. There is a finite interval of values $\mu_- < \mu < \mu_+$ for which self-sustained oscillations occur.

Before commenting on these oscillations, we note that in the normal unstressed state where $\mu < \mu_-$, the steady state is stable but *excitable*. Sufficient elevation of μ for a sufficiently long time will cause an excursion in the sarcoplasmic calcium concentration, which corresponds to muscle contraction. Note also that if $\mu > \mu_+$ (and particularly $\mu \gg \mu_+$), then the intracellular Ca^{2+} remains high, and the muscle will permanently contract. Presumably this corresponds to cramp (and *rigor mortis*).

For the case in which $\mu_- < \mu < \mu_+$, figure 4.8 shows how the dynamics works in the (u, v) phase plane. The two arrowed lines are the tangents to the nullcline of slope $\dfrac{dv}{du} = -\dfrac{1}{\gamma}$ (thus $u+\gamma v$ is constant on these lines). The values of μ where these tangents touch the nullcline define $\mu_\pm$, as shown; their values are $\mu_+ = 0.69$, $\mu_- = 0.28$, approximately (always using the values of the parameters in (4.8) and (4.9)). For an arbitrary initial condition, for example at P in the figure, v changes rapidly, and since $f < 0$ above the nullcline, the trajectory rapidly (on a time scale $t \sim \varepsilon$) approaches the nullcline at A, and remains on it. Now, if we suppose $\mu_- < \mu < \mu_+$, we see that $\dot{u} + \gamma \dot{v} < 0$ at A since $u > \mu$ there, and therefore the solution migrates along

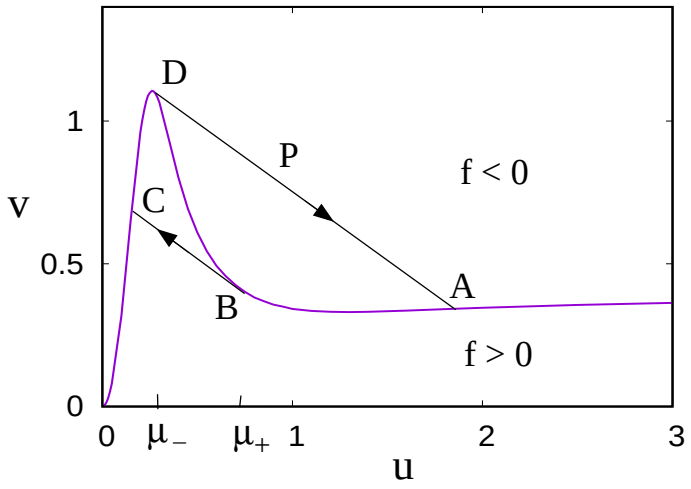

Figure 4.8: Phase portrait for (4.19).

the nullcline to the left on the slow time scale $t \sim O(1)$. If μ were greater than μ_+, this migration would end when $u = \mu$, which then defines a stable steady state of the system. But if $\mu < \mu_+$, the trajectory reaches B, and can not continue along the nullcline, since on the nullcline, $u + \gamma v$ starts to increase again. The only possibility is that the slow phase AB finishes and there is another phase of rapid change. This is BC; again v changes rapidly while $u + \gamma v$ remains constant, until the trajectory lands on the nullcline again (note that here v increases, because BC lies below the nullcline, where $f > 0$).

We revert to another slow phase, but now, if $\mu > \mu_-$, $\dot{u} + \gamma\dot{v} > 0$ because $u < \mu_- < \mu$ at C, and so the trajectory moves up along the nullcline and slightly past its maximum to D, where again it must enter a rapid transition phase. If μ were $< \mu_-$, then the trajectory would reach a stable steady state where $u = \mu$, but if $\mu > \mu_-$, this does not occur. In this way, the trajectory is doomed to cycle repeatedly around $ABCD$, forming a periodic limit cycle which, because of the alternation of fast and slow components, is called a *relaxation* oscillation (the solution alternately relaxes to the slow phases AB and CD).

How do we describe this in the more conventional fast and slow variable formulation? If we define

$$w = u + \gamma v, \tag{4.20}$$

then the system takes the form

$$\begin{aligned} \dot{w} &= \mu - w + \gamma v, \\ \varepsilon\dot{v} &= F(w, v) \equiv f(w - \gamma v, v), \end{aligned} \tag{4.21}$$

and it is not difficult to see that the nullcline $F = 0$ will be multi-valued as in figure 4.9. The other nullcline $u = \mu$ is equivalently $v = (w - \mu)/\gamma$, and an example for

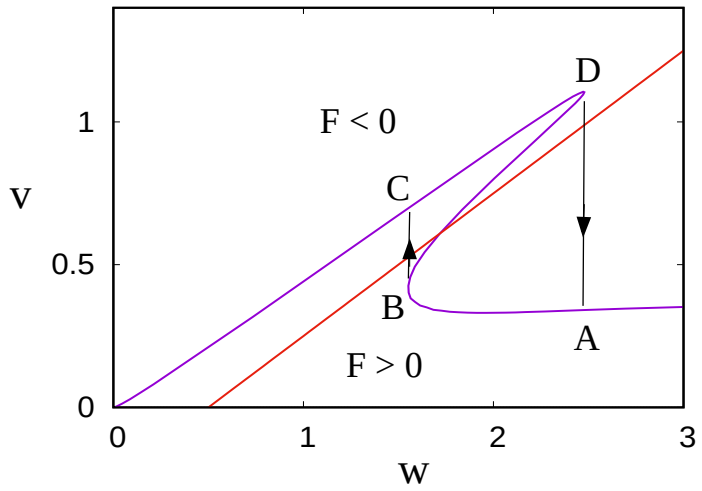

Figure 4.9: The nullclines for (4.21). The red w nullcline is drawn for the value $\mu = 0.5 \in (\mu_-, \mu_+)$.

which $\mu_- < \mu < \mu_+$ is shown in the figure as the red line. This is the classical sort of situation where a relaxation oscillation occurs. The fixed point on the intermediate branch of the v-nullcline is unstable, and the limit cycle $ABCD$ cycles back and forth between the upper and lower branches.

While ordinary phase plane analysis can be done in figure 4.9, it is less easy to do in figure 4.8. First, let us try to do this in figure 4.9. The trajectories go through the v-nullcline horizontally and the red w-nullcline vertically. Then the trajectories cycle clockwise around the unstable fixed point. If we try and then draw in the relaxation limit cycle for $\varepsilon \ll 1$, we see that the trajectory on AB must lie marginally above the v-nullcline, and on CD, marginally below; the relaxation oscillation is thus consistent with the demands of phase plane analysis.

But it is less easy to do this for figure 4.8. This is not to say that the description above is inaccurate, but think about this: trajectories must cross the v-nullcline in figure 4.8 horizontally. Thus as in figure 4.9, the section AB must lie marginally above the v-nullcline. Equally, since $u + \gamma v$ decreases along DA, the trajectory there dips slightly more rapidly than the straight line DA. So the trajectory must switch rapidly through the vertical close to A, and this is where the u-nullcline must be. So where is the u-nullcline?

From (4.19), we see that it is given by

$$f(u, v) = \frac{\varepsilon}{\gamma}(\mu - u), \tag{4.22}$$

and thus it lies very close to the v-nullcline. In fact, if we denote the v-nullcline as $v = g(u)$, then expansion of (4.22) leads to the approximate formula for the u-

nullcline,

$$v = g(u) + \frac{\varepsilon(u-\mu)}{\gamma\left[\delta + L'\{g(u)\}K(u)\right]}, \tag{4.23}$$

and this allows us to complete a phase plane analysis satisfactorily.

What do the solutions look like? According to the theory, the cytosolic Ca^{2+} oscillations (thus u) should be spiky, like those shown in figure 4.2. With the parameter values we have been using, figure 4.10 shows that this is mildly true, but not dramatically so. The reason for this is that ε is not small enough; indeed the phase plane on the left shows that the trajectory dresses the v-nullcline, but is not glued to it.

If we reduce ε, however, the spikiness increases, as shown in figure 4.11, in which $\varepsilon = 0.005$, and the solution follows the v-nullcline much more closely. Normally we think of $\varepsilon = 0.04$ as being in practice small, but this example shows that one must always take ideas of numerical largeness or smallness with a degree of circumspection. In fact, there is a good reason for the inaccuracy, which is to do with the confounding effect of δ also being small. We already saw on page 64 how the smallness of δ was compromised in approximating the v-nullcline by the 'smallness' of u_c. Here a different confusion occurs, because the smallness of δ interferes with the smallness of ε.

We can use our various approximations to estimate such quantities as the amplitude and the period of the oscillation when ε and δ are small, though we should bear in mind the issues above. First, it is clear that as ε is reduced, the shape of the limit cycle in the phase plane becomes fixed, and thus the ampltude is fixed as ε is reduced and is independent of μ. The period also becomes independent of ε, but depends, as we shall see, on μ; it is dominated by the passage of the trajectory through the slow phases AB and CD. In fact, figure 4.11 suggests that most of the time is spent on the low branch CD.

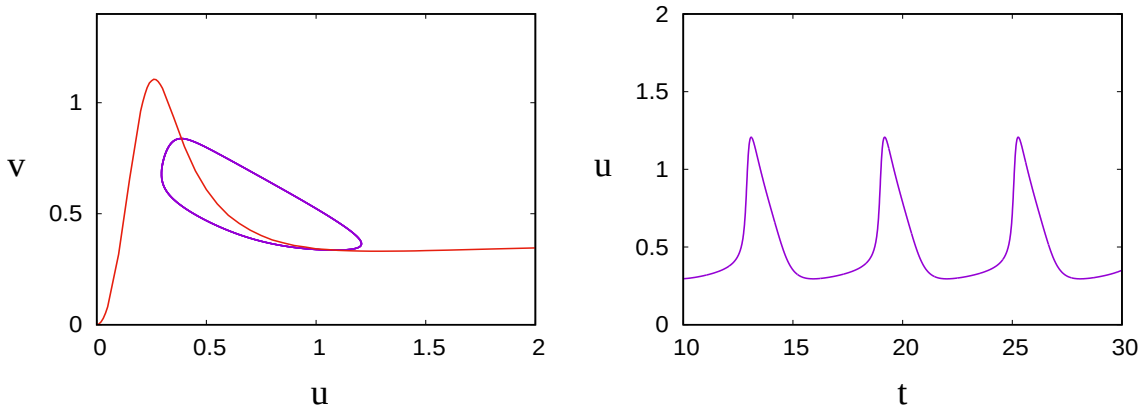

Figure 4.10: Phase plane and time series for (4.19), standard parameter values, including $\varepsilon = 0.04$.

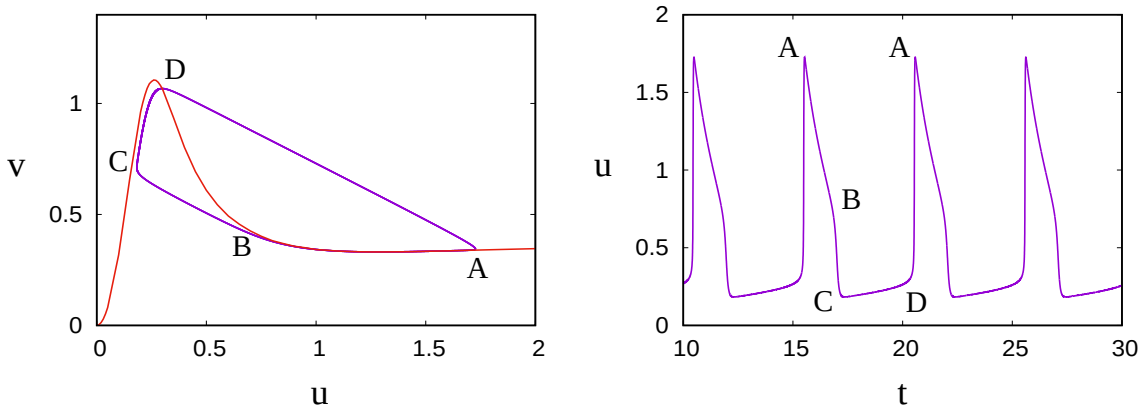

Figure 4.11: Phase plane and time series for (4.19), standard parameter values, except that $\varepsilon = 0.005$.

To quantify this, we note that on the slow branch CD, u is $O(1)$ but v is $O(1/\delta)$. It is appropriate to rescale the time also, thus we put

$$v = \frac{V}{\delta}, \quad t = \frac{T}{\delta}, \quad f \approx J(u) - K(u) - V, \tag{4.24}$$

and (4.19) is approximately (overdots now being differentiation with respect to T)

$$\begin{aligned} \delta \dot{u} &= \mu - u - \gamma \dot{V}, \\ \varepsilon \dot{V} &= J(u) - K(u) - V. \end{aligned} \tag{4.25}$$

This suggests that, on this longer time scale, V relaxes to its equilibrium

$$V \approx G(u) = J(u) - K(u), \tag{4.26}$$

and then u satisfies

$$\dot{u} \approx \frac{\mu - u}{\gamma G'(u)}. \tag{4.27}$$

The duration of this phase is thus (in terms of t)

$$t_{CD} \approx \frac{1}{\delta} \int_{u_C}^{u_D} \frac{\gamma G'(u)\, du}{\mu - u}, \tag{4.28}$$

where $u = u_C$ at C and $u = u_D$ at D. Note that $G'(u) \gtrsim 0$ for $u < u_D$, and that also $u < \mu$ in this range, so the integrand is positive and bounded.

Next, since $v \sim \dfrac{1}{\delta}$ at D, it follows that $u \sim \dfrac{1}{\delta}$ at A, and we can put

$$u = \frac{U}{\delta}, \quad f \approx \beta - L(v), \tag{4.29}$$

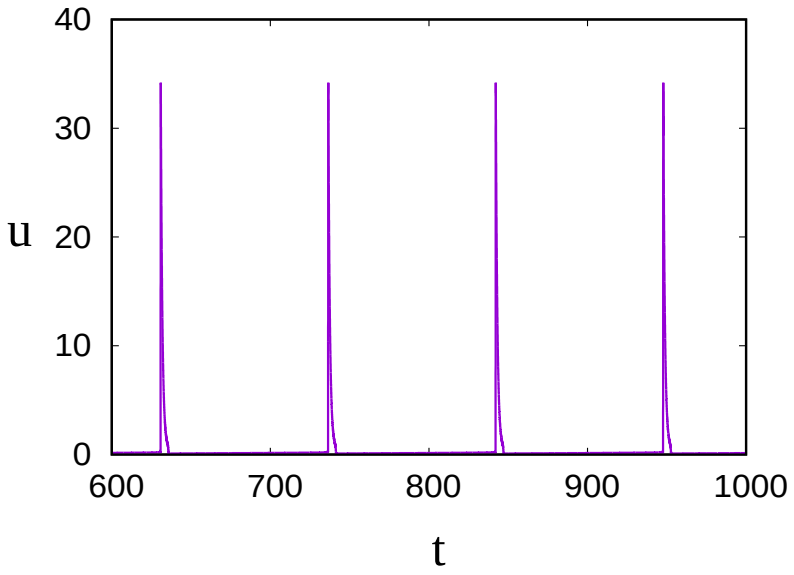

Figure 4.12: Time series for (4.19), standard parameter values, except that $\varepsilon = 0.005$, $\delta = 0.0001$.

and therefore the system (4.19) takes the form

$$\begin{aligned} \dot{U} &= -U + \delta(\mu - \gamma\dot{v}), \\ \varepsilon\dot{v} &= \beta - L(v), \end{aligned} \tag{4.30}$$

whence $v \approx L^{-1}(\beta)$ is constant and $\dot{U} \approx \delta\mu - U$. We could go on with this, but the main point is that the duration of CD is $\sim \frac{1}{\delta}$, while that of AB is $\sim \ln\left(\frac{1}{\delta}\right)$, which is the time for U in (4.30) to reach $O(\delta)$. The oscillation becomes increasingly spiky as δ us reduced, and this is shown in figure 4.12, where we take $\delta = 0.0001$ and $\varepsilon = 0.005$.

At leading order, the period is thus approximately given by t_{CD}, with a correction of $O\left(\ln\frac{1}{\delta}\right)$, and it is a decreasing function of μ. This is consistent with the data shown in figure 4.2.

4.2 Wave propagation

Ca^{2+} oscillations can cause waves to propagate inside cells at speeds 10-100 μm s^{-1}. These are particularly evident in large cells (e. g., oocytes (developing eggs)) where periodic wave trains can be seen. Spiral waves have been seen in Xenopus oocytes.

The simplest model to describe such waves is the two pool model where we also allow diffusion of cytosolic Ca^{2+}. Thus (non-dimensionally), and in one spatial di-

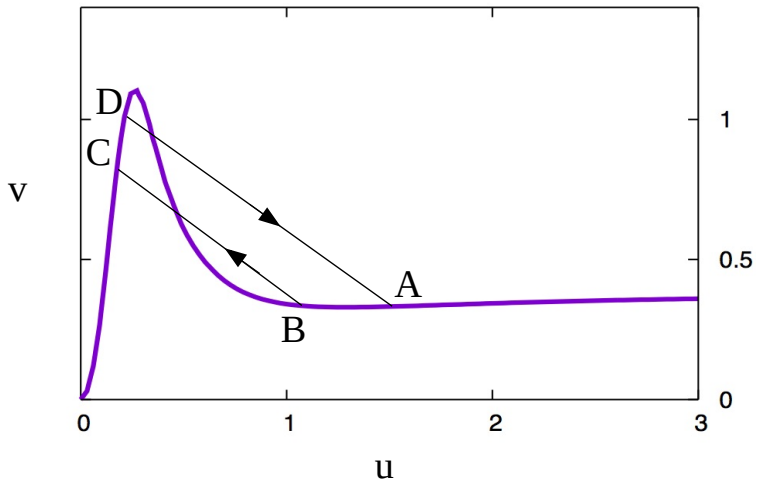

Figure 4.13: The sought-after trajectory of a periodic travelling wave in the (u, v) phase plane.

mension,

$$\begin{aligned} u_t + \gamma v_t &= \mu - u + \nu u_{xx}, \\ \varepsilon v_t &= f(u, v), \end{aligned} \tag{4.31}$$

where

$$\nu = \frac{D}{l^2 k}, \tag{4.32}$$

l is the length scale used to non-dimensionalise x, and D is the diffusion coefficient. In what follows, we seek travelling wave solutions of the form

$$u = u(\xi), \quad v = v(\xi), \quad \xi = x + st; \tag{4.33}$$

our discussion will be very similar to that used in discussing travelling waves in the FitzHugh–Nagumo model (section 3.2, page 49); the difference is that there the diffusion term was included in the fast equation: here it is the other way round. In addition, the excitable steady state led to a solitary travelling wave; here the limit cycle of the phase plane is expected to yield periodic travelling waves. In somewhat similar fashion to section 3.2, we will aim to construct a relaxational wave having the cyclic phase plane trajectory indicated in figure 4.13, where the values of u_D and u_B are as yet unconstrained. If we anticipate that $s > 0$, then we expect the wave profile to be as indicated in figure 4.14.

In keeping with the FitzHugh–Nagumo equation analysis, we need to choose the size of the diffusion term so that it becomes important in the fast transition (between the two quasi-steady branches AB and CD; evidently this will involve rescaling $\xi \sim \varepsilon$, and therefore we choose $\nu = \varepsilon$, which thus implies

$$l = \left(\frac{D}{k\varepsilon} \right)^{1/2}. \tag{4.34}$$

Using values $D \sim 20\ \mu\text{m}^2\ \text{s}^{-1}$, $\varepsilon = 0.04$, $\text{k} = 10\ \text{s}^{-1}$, we find that $l \sim 7\ \mu\text{m}$, and this suggests a wave speed $\sim lk \sim 70\ \mu\text{m s}^{-1}$ if $s \approx 1$.

We now sketch the structure of the solution. Substituting (4.33) into (4.19), we obtain (with $\nu = \varepsilon$)

$$\begin{aligned} s(u' + \gamma v') &= \mu - u + \varepsilon u'', \\ \varepsilon s v' &= f(u, v). \end{aligned} \tag{4.35}$$

In the fast wavefront DA, we put $\xi = \varepsilon X$; $(4.35)_1$ becomes approximately (primes now denoting differentiation with respect to X)

$$s(u' + \gamma v') = u'', \tag{4.36}$$

and a first integral of this together with $(4.35)_2$ gives the system

$$\begin{aligned} u' &= s[u - u_D + \gamma(v - v_D)], \\ sv' &= f(u, v), \end{aligned} \tag{4.37}$$

where u_D and v_D are the values of u and v at D. The object now is to find a solution of this which connects the fixed point (u_D, v_D) to the corresponding fixed point (u_A, v_A). As for the FitzHugh–Nagumo case, we expect this to require a particular choice of s.

It is convenient to shift D to the origin by writing

$$U = u - u_D, \quad V = v - v_D, \quad f(u, v) = F(U, V), \tag{4.38}$$

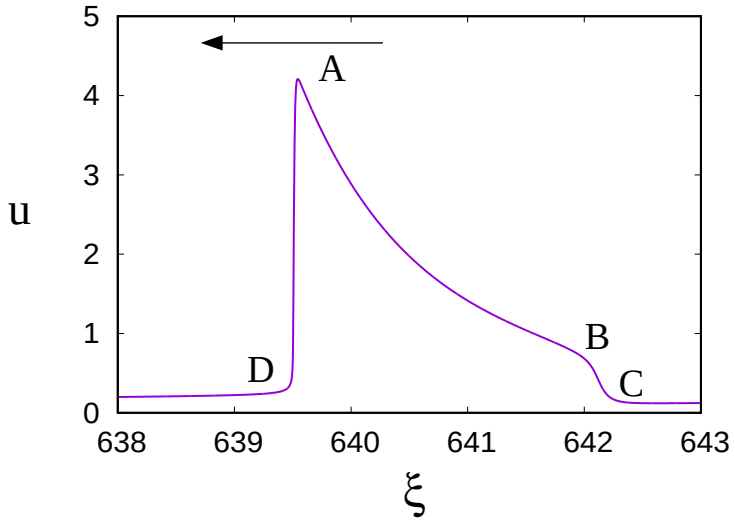

Figure 4.14: Presumed shape of the wave. (This figure is actually the time series of the limit cycle when $\delta = 0.001$.) As indicated, the wave moves backwards, thus $s > 0$ in (4.33). Note that it repeats periodically.

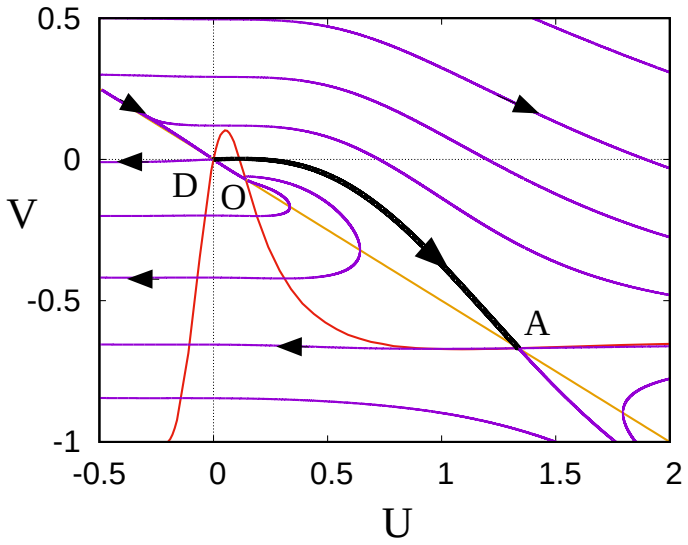

Figure 4.15: Phase portrait for (4.39). The default parameters are used, and the choice of s which connects D to A for this is $s = 0.816$, approximately. The values of u_D and v_D used are $u_D = 0.21$ and $v_D = 1.003$.

so that

$$\begin{aligned} U' &= s(U + \gamma V), \\ sV' &= F(U, V). \end{aligned} \tag{4.39}$$

The phase portrait for this system is shown in figure 4.15. D is a saddle, as is A, and the fixed point between them is an unstable node or focus. (In this case it is a node.) In order to connect D to A as required, the unstable separatrix in $U > 0$ from D must connect to the stable separatrix to A. In general this will not occur, but as we decrease s, the separatrix trajectory becomes lower, and at one particular value it connects to A. For the default parameters of the model, this is at $s \approx 0.816$. (In order to construct figure 4.15 numerically, it is necessary to determine s to seven decimal places, but while these are necessary to construct the figure, the later places have little meaning, since they depend on the details of the numerical method used.)

To prove such a value of s exists, one can use comparison arguments. To do this, we would write (4.39) in the form

$$\frac{dV}{dU} = \frac{F(U, V)}{s^2(U + \gamma V)}, \quad V \sim kU \quad \text{as} \quad U \to 0+, \tag{4.40}$$

where k is the positive root of

$$k = \frac{F_U - |F_V|k}{s^2(1 + \gamma k)}, \tag{4.41}$$

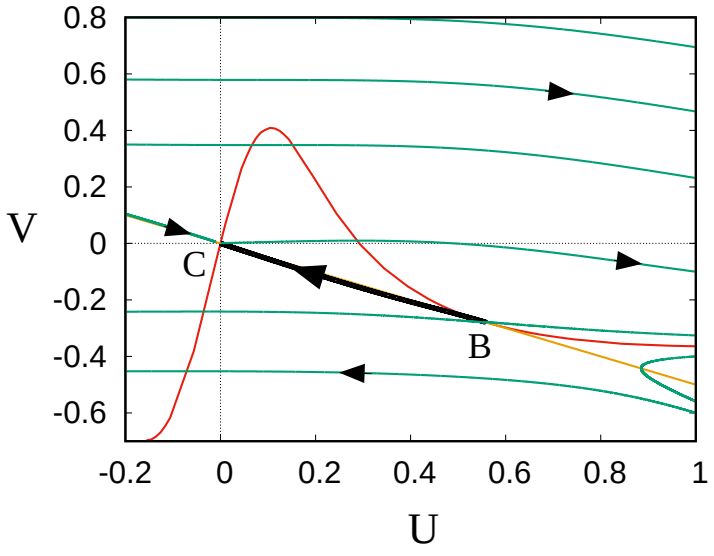

Figure 4.16: Phase portrait for (4.39). The choice of u_B which connects B to C for this is $u_B = 0.714$, approximately. At C, $u_C \approx 0.156$, $v_C \approx 0.697$.

the derivatives being evaluated at $U = V = 0$; this selects the required separatrix solution. It is easy to see by graphical means that k decreases as s increases, so that, for small U where $F > 0$, $V_-(U) > V_+(U)$, where $V_\pm$ are solutions of (4.40) corresponding to two values $s_- < s_+$. Equally, for solutions of (4.40) which connect to the stable separatrix to A, $V_- > V_+$. Now as $s \to \infty$, the separatrix solution from D becomes $V \approx 0$, while as $s \to 0$, it follows the v-nullcline to the (now stable) node; so by continuity arguments there is at least one value of s for which the separatrices are connected. Often one can demonstrate uniqueness using monotonicity arguments, but that is not so easy here because of the change of sign of F. Numerically (at least for these parameter values), it seems there is only one value of s, however.

In fact, it is noticeable in figure 4.15 that the trajectories to the left of the figure are very flat. The reason for this is that for u less than about 0.5, f and thus F is very small (which, as we noted earlier, is why the small δ approximation in figure 4.7 is not very accurate). In particular, the unstable separatrix from D in $U > 0$ in figure 4.15 is very close to $V = 0$; in fact, where it crosses the nullcline again into the region where $F < 0$, $U = 0.115$ and $V = 0.002$. If we assume that as s varies, this value of V stays constant, then the part of the trajectory in $F < 0$ varies monotonically with s, which then proves the uniqueness of the value of s which connects D to A. One supposes this will in general be true for sufficiently small but non-constant values of V where the separatrix crosses the nullcline. In fact, it is possible to prove uniqueness using a slightly different formulation of the problem.

Once the trajectory reaches A, it enters a slow phase, in which we return to the variable ξ, and at leading order (4.35) implies $f(u, v) \approx 0$, whence $v \approx g(u)$ (the

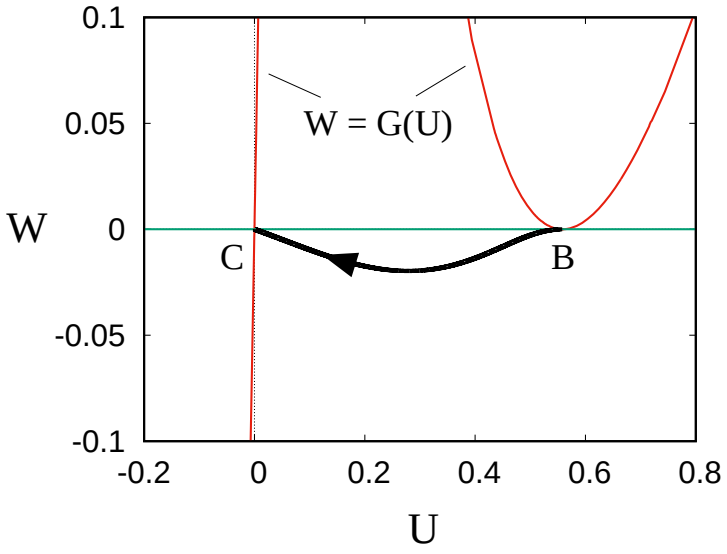

Figure 4.17: The connecting trajectory in the (U, W) phase plane, where $W = U + \gamma V$. Note the vertical scale. The indicated curve $W = G(U)$ is where $F(U, V) = F\left(U, \dfrac{W - U}{\gamma}\right) = 0$.

v-nullcline), and

$$u' \approx \frac{\mu - u}{s(1 + \gamma g'(u)}; \tag{4.42}$$

this is negative, and u decreases along the v-nullcline until the point B, where another fast phase is initiated. The value of u_B is unknown, and must be chosen in a similar manner to s, in order to connect the separatrices of the two outer fixed points of (4.37), which now takes the form

$$\begin{aligned} u' &= s[u - u_B + \gamma(v - v_B)], \\ sv' &= f(u, v). \end{aligned} \tag{4.43}$$

The phase plane is shown in figure 4.16, in terms of the variables

$$U = u - u_C, \quad V = v - v_C, \tag{4.44}$$

so that the point C is the origin, and U and V satisfy (4.39); the construction of the phase plane is left as an exercise.

Unlike figure 4.13, figure 4.16 appears to show that the point B is actually where the U-nullcline $U + \gamma V = 0$ is tangent to the V-nullcline, and numerically, this appears to be the case. Figure 4.17 shows the connecting trajectory and figure 4.18 shows a close-up. For visualisation purposes, the variables have been changed again to U and $W = U + \gamma V$, so that the curve $F(U, V) = 0$ becomes $W = G(U)$. In interpreting

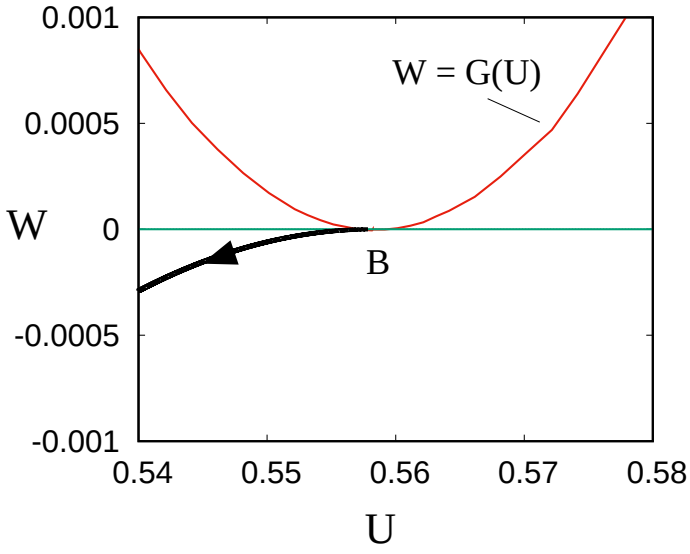

Figure 4.18: A close-up near B of the connecting trajectory.

these figures, it should be noted that the system is

$$\begin{aligned} U' &= sW, \\ W' &= sW + \frac{\gamma F}{s}, \end{aligned} \tag{4.45}$$

so that while $W = 0$ is the U-nullcline, $F = 0$ (or $W = G(U)$) is *not* the W-nullcline, which is given by $W = -\dfrac{\gamma F}{s^2}$; for phase plane purposes we can think of the W-nullcline in figure as being like $W = -G(U)$.

The further close-up in figure 4.18 provides more evidence that the correct choice of B is at the tangent, as also indicated in figure 4.16. It is worth commenting on how such figures are produced. As mentioned, to the left of this figure, most of the trajectories are horizontal, because F is small there. The exception is near the U-nullcline (of (4.39)), which is why the stable separatrices to C almost lie on it (but not exactly, as shown in figure 4.17). But this fact explains why the choice of B must be close to the tangent, for if it were not (as in figure 4.13), then the stable separatrix to C would follow the U-nullcline to the unstable node and terminate there.

So B needs to be close to the tangent; but why it should appear to lie exactly at the tangent is not entirely obvious. A clue lies in the fact that if F is small enough (as here) then B will be at the tangent, but that for larger F it will move away from the tangent, as suggested in figure 4.13. In calculating the trajectory BC in figure 4.16, we adjusted B nearer and nearer to the tangent, until finally we put it precisely at the tangent, and integrated the equations (4.39) backwards (in X) from near C. When B is at the tangent, it becomes a degenerate saddle.

Figure 4.19: Spiral waves in the Belousov-Zhabotinskii reaction. Image by Art Winfree.

4.2.1 Waves in higher dimensions

We finish this chapter with some comments on wave propagation in higher dimensions. In reality, waves such as described above propagate in a three-dimensional medium. To be identifiable as a wave, solutions need to be quasi-one-dimensional, but the wave fronts can now be curved. Various kinds of modifications can occur, but the principal ones are target patterns and spiral waves, and both are essentially two-dimensional in nature. As their name suggests, target patterns are circularly symmetric concentric rings like those on an archer's target. Spiral waves are similar, but are not rotationally symmetric in the angular variable θ.

Both kinds of wave can be described in cylindrical polar coordinates by a representation of the form

$$u = u[\Omega t + m\theta - \psi(r)]. \tag{4.46}$$

If $m = 0$ these are target patterns, and typically they propagate outwards, thus $\psi'(r) > 0$ if $\Omega > 0$. The reason for this is that typically the waves are organised by inhomogeneities or other entities at their centres. If $m \neq 0$, (4.46) represents a spiral wave. The spiral shape is given by $m\theta = \psi(r)$, and the spiral rotates; sometimes such waves are called rotating waves, and rotating Ca^{2+} waves have been observed in Xenopus oocytes.

The best-known examples of target patterns and spiral waves occur in the Belousov-Zhabotinskii reaction, which is a chemical reaction which spontaneously oscillates (the oscillation is visible because of colour change in the reacting medium, see figure 4.19).

When the experiment is done in a thin layer of reagent in a petri dish, the oscillations cause waves to propagate, generally of target or spiral form. It is commonly thought that the centres of the waves are associated with impurities in the medium, for example.

Another, and more pertinent, example is the electrochemical action of the heart. We shall have more to say on this in chapter 5. The heartbeat is generated by a pacemaker (oscillator) in the sino-atrial node which acts as an organising centre, and causes a wave to propagate through the excitable *atria* and *ventricles*. Because the medium is excitable, a solitary wave propagates, and in normal function this would be something like a target pattern. However, blockage of conduction paths by diseased heart tissue can lead to 're-entrant' spiral waves (which cycle round the diseased tissue), causing *ventricular tachycardia*. In the increasingly diseased heart, such spiral waves can become chaotic, leading to ventricular fibrillation and death.

Chapter 5

The electrochemical action of the heart

The purpose of the heart is to pump blood through the body. The blood carries nutrients to the tissues and carries away waste products; principally, the nutrient is oxygen (O_2) and the waste is carbon dioxide (CO_2). Exchange of these gases between the body and the atmosphere occurs via respiratory exchange at the lungs, and is effected by a perfusion of the blood through a *capillary bed* in the pulmonary circulation.

There are two parts of the heart function which we will focus on in these notes. The first is the electrochemical action of the heart, that is to say, the way in which electrochemical signals cause muscle contraction in the *myocardium*, which enables the heart to pump blood round the body. The second is the mechanical action of the heart, i. e., the way in which this contraction enables a uni-directional circulation of the blood via a system of valves in the heart. We deal with the electrochemical action in this chapter, and the pump action in the next.

5.1 Action potentials and the heart beat

The heart contains four chambers, comprising two atria and two ventricles. Blood collects in the atria and is pumped to the ventricles, which in turn pump the blood to the lungs and around the body. For the heart to act effectively it is necessary for the sequence of contraction of the chambers to be synchronised. This is achieved by electrical signals (cardiac action potentials) which are transmitted through the myocardium (see figure 5.1). The heart is made of billions of individual cells. Each cell is surrounded by a membrane which is electrically polarised (the membrane potential). The electrical signals that stimulate contraction cause the membrane potential to depolarise and a second electrical signal repolarises the membrane. The *electrocardiogram* (ECG) is a measurement of these electrical signals on the surface of the body and consists of a series of waves (see figure 5.2).

Each wave in the ECG is linked with a depolarisation wave or a repolarisation wave

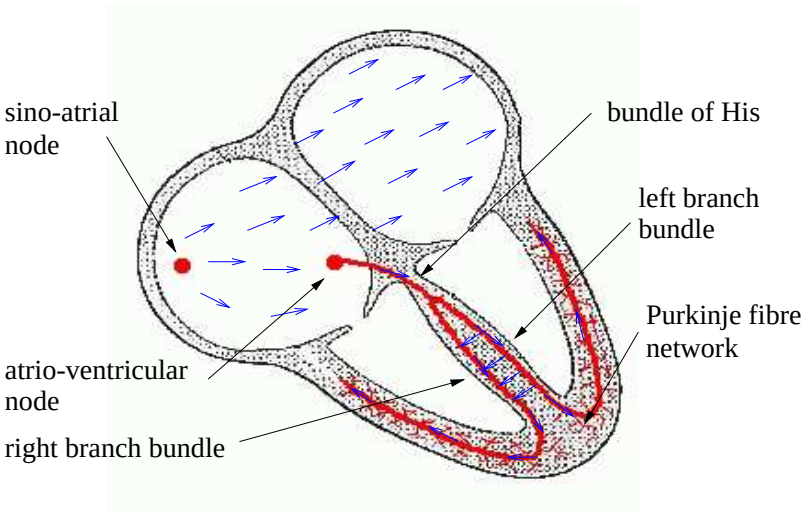

Figure 5.1: The electric pathways in the heart. The red lines are the Purkinje fibre network, and the blue arrows show the direction of the electric signal (adapted from Houghton and Gray (1997)).

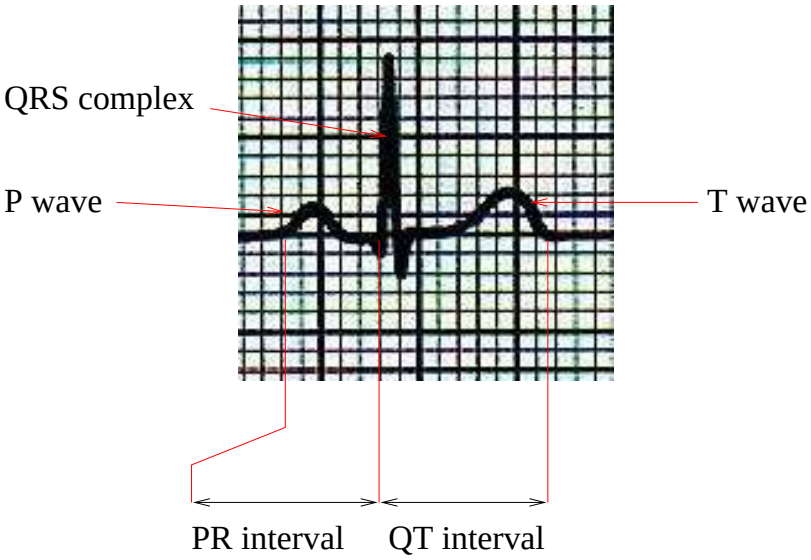

Figure 5.2: The electrocardiogram viewed from lead-I. The separate P wave, QRS complex and T wave, signifying atrial depolarisation, ventricular depolarisation and ventricular repolarisation respectively, are clearly visible (adapted from Houghton and Gray (1997)).

in the atria or ventricles. The first wave is the *P-wave* and is caused by a depolarisation wave in the atria which originates from the sino-atrial (SA) node (figure 5.1). The sino-atrial node is an electrical oscillator which is the heart's pacemaker. The second wave is the *QRS-complex* and is caused by the ventricles depolarising. There is a gap between the P-wave and QRS-complex (the PR interval) which is due to a pause in the transmission of the depolarisation wave in the atrio-ventricular (AV) node. The depolarisation wave is transmitted rapidly though the ventricles by the Purkinje fibre network which leads to the whole ventricle contracting simultaneously. The final wave is the *T-wave* which is caused by the ventricles repolarising.

5.2 Cardiac cells

Similarly to neurons (see chapter 2), cardiac cells are electrically active. There are five main types of heart cells, each type fulfilling a specific function. The cells are listed below in the order in which they depolarise during a cardiac activation sequence:

1. *Sino-atrial node cells* are the pacemakers of the heart, and have oscillatory action potentials. SA node cells do not contract. The SA node is situated on the right atrium.

2. *Atrial myocytes* are excitable cells which conduct the action potential (the P-wave) with a velocity of about 0.5 m s^{-1}. Atrial myocytes are muscle cells and contract when their membrane potential is depolarised.

3. *Atrio-ventricular node cells* are excitable cells which conduct action potentials with a velocity of about 0.05 m s^{-1}. The main role of the AV node is to create a pause between the contraction of the atria and ventricles (the PR-interval). However, if the SA node fails, then the AV node can take over as the pacemaker.

4. *Purkinje fibres* are excitable cells and conduct the action potential rapidly (at about 5 m s^{-1}). After the action potential has passed through the AV-node, the rapid conduction in the Purkinje fibres makes the depolarisation of the ventricles synchronous (this is why the QRS-complex is narrow).

5. *Ventricular myocytes* are excitable cells which conduct the action potential slowly (at about 0.5 m s^{-1}). Ventricular cells are muscle cells and contract when their membrane potential is depolarised.

In the folowing sections, we shall see how the underlying ionic currents give rise to the distinctive action potentials of the SA-node cells and of the ventricular myocytes. The Nernst potentials for sodium, potassium and calcium in cardiac cells are approximately $V_{\mathrm{Na}} = 60$ mV, $V_{\mathrm{K}} = -95$ mV and $V_{\mathrm{Ca}} = 130$ mV.

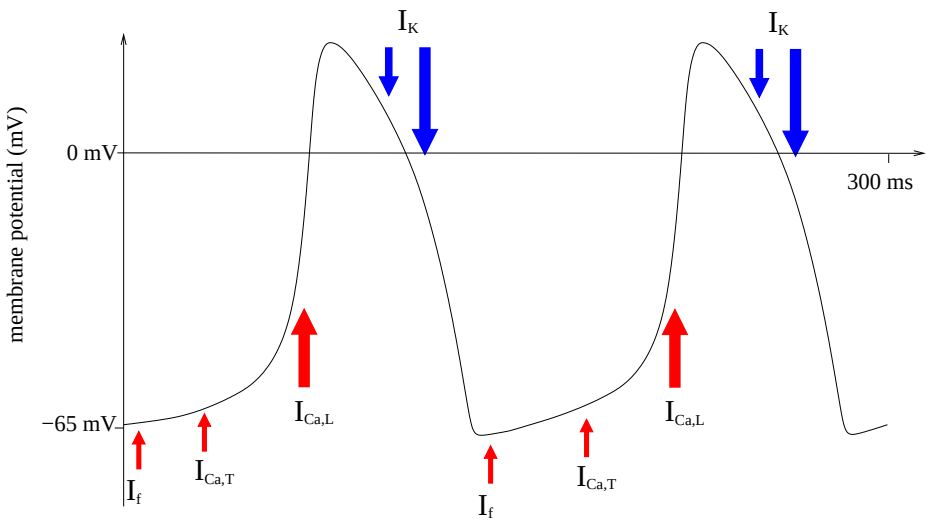

Figure 5.3: The action potential and principal ionic currents of the SA node cells. The upward arrows represent depolarising currents, the downward arrows represent repolarising currents, and the size of an arrow represents the relative size of the current.

5.2.1 Sino-atrial node cells

The SA-node is the pacemaker of the heart, due to the fact that the SA-node cells have oscillatory action potentials. The action potential contains three principal depolarising ionic currents and one principal repolarising current. Note that depolarising currents are also called inward currents, and cause the membrane potential to increase, while repolarising currents are also called outward or rectifying currents and cause the membrane potential to decrease. Each of the ionic currents is voltage-gated (see chapter 2) and they act at different times during the action potential (see figure 5.3).

The main currents are the following:

1. I_{f} is the 'funny' depolarising current. This depolarising current is activated (i. e., it contains voltage-gates which open) at low potentials (≈ -65 mV) and it occurs at the beginning of the depolarising phase of the action potential.

2. $I_{\mathrm{Ca,T}}$ is the transient Ca^{2+} current, which contributes to the depolarisation at potentials between -60 mV and -45 mV. At higher potentials it is inactivated.

3. $I_{\mathrm{Ca,L}}$ is the long-lasting Ca^{2+} current and is the most important pacemaking current. The current is activated when the membrane potential rises above -55 mV.

4. I_{K} is the time-dependent potassium current. This current is responsible for the repolarisation of the action potential and is activated when the potential rises

above −40 mV. However, the voltage-gates take approximately 100 ms to fully open.

Individual isolated SA node cells oscillate with different time periods (between 170 ms and 350 ms). However, in the SA node neighbouring cells are electrically coupled, which synchronises their action potentials. The SA node is electrically coupled to the right atrium which allows the action potential to spread from the SA node. This electrical coupling has a second effect: during the initial depolarisation phase it acts as a current sink, which slows the rate of depolarisation and thus reduces the frequency of the pacemaker. The frequency of the pacemaker can be increased by increasing the size of the depolarising current $I_{\mathrm{Ca,L}}$, $I_{\mathrm{Ca,T}}$ and I_{f}. For example catecholamines (such as adrenaline and noradrenaline), which are neurotransmitters released when the sympathetic nerves are stimulated, dramatically increase $I_{\mathrm{Ca,L}}$ and can increase the frequency of the pacemaker by a factor of three. Conversely, acetylcholine (Ach), which is a neurotransmitter released when the parasympathetic nerves are stimulated, reduces both $I_{\mathrm{Ca,L}}$ and I_{f}, and decreases the frequency of the pacemaker.

5.2.2 Ventricular myocytes

The ventricular muscle is made of billions of individual ventricular myocytes. The ventricles contract when all the myocytes contract simultaneously. The electrophysiology of ventricular myocytes therefore must play two rôles. First, it must allow electrical signals to pass between the cells to synchronise the contractions, and second, it must stimulate the contraction. The resting potential for ventricular myocytes is typically −90 mV, which is close to the Nernst potential for the potassium ions. The principal currents involved in the ventricular action potential are shown in figure 5.4. The action potential starts with a rapid depolarisation phase which is followed by the long plateau phase where the potential changes slowly. After approximately 300 ms the plateau phase ends and the cell membrane rapidly repolarises. The exact make-up and balance of the ionic currents during the plateau phase is still an area of active research and is known to differ between species.

1. I_{Na} is the fast inward sodium current which is the main current during the depolarisation phase. As the membrane potential rises above −65 mV, this current is rapidly activated and then deactivated by voltage-gates. The current is very large leading the membrane to depolarise fully in less than 1 ms.

2. I_{to} is the transient outward current which causes a small but rapid repolarisation immediately after the initial depolarisation.

3. $I_{\mathrm{Ca,L}}$ is the long-lasting Ca^{2+} current which is a depolarising current. It is activated when the potential rises above −40 mV, and is deactivated by the increased intracellular Ca^{2+} during an action potential and the membrane potential. The current triggers the release of Ca^{2+} from the *sarcoplasmic reticulum*, which is the internal store of Ca^{2+} (see chapter 4). This process is called

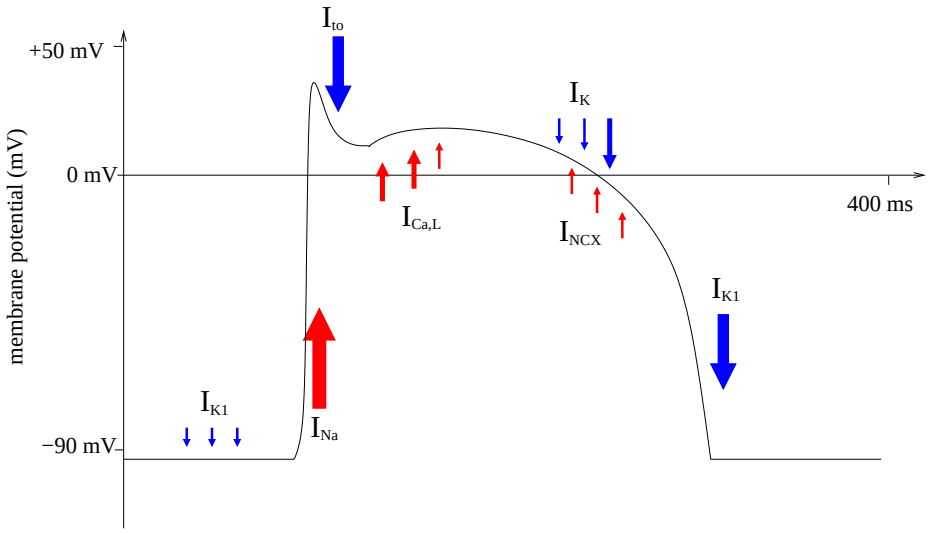

Figure 5.4: The action potential and principal ionic currents of the ventricular myocyte. The upward arrows represent depolarising currents, the downward arrows represent repolarising currents, and the size of an arrow represents the relative size of the current.

calcium-induced calcium release (CICR) and is essential in electro-contraction coupling (E-C coupling) because the high levels of Ca^{2+} it produces cause the cell to contract.

4. I_{NCX} is the sodium-calcium exchanger and is responsible for removing Ca^{2+} ions from the cell (note Ca^{2+} enters the cell via $I_{\mathrm{Ca,L}}$, so it must eventually be removed in order to prevent a build up of Ca^{2+}). The I_{NCX} pumps Ca^{2+} against the concentration gradient (the extracellular $[Ca^{2+}]$ is over 10,000 times greater than the intracellular $[Ca^{2+}]$) by allowing three sodium ions to enter the cell for each calcium ion which is removed. When Ca^{2+} is being removed from the cell there is a net inward current which helps to maintain the action potential.

5. I_{K} is the outward potassium current which is a repolarisation current. During the plateau phase it balances the inward $I_{\mathrm{Ca,L}}$ and I_{NCX}.

6. $I_{\mathrm{K_1}}$ is the background potassium current and is a repolarisation current which is inactivated at high membrane potentials. It is responsible for the rapid repolarisation at the end of the plateau phase and for maintaining the resting potential.

The plateau phase ends when the outward currents I_{K} and $I_{\mathrm{K_1}}$ become larger then the balancing inward currents $I_{\mathrm{Ca,L}}$ and I_{NCX}.

5.2.3 The Noble model

Mathematical models (e. g., those of Noble (1962), Beeler and Reuter (1977), or Luo and Rudy (1994)) of cardiac cells can be constructed describing these ionic currents and their associated voltage-gates. These models are similar to the Hodgkin-Huxley model for nerve cells, but involve more variables due to the greater number of currents involved. However, the basic feature of cardiac cells is that they are excitable and that the upstroke of the action potential is much more rapid than repolarisation during the plateau phase. The FitzHugh-Nagumo model (chapter 2) also has these properties and can be used as a crude approximation to model cardiac action potential.

As one example of such models, we here describe the Noble model, not for its veracity, but because it is relatively simple, and also it sets the historical precedent for later models such as those cited above. In particular, note that the action potential illustrated in figure 5.4 has quite a different shape to that of the SA node cells as illustrated in figure 5.3.

The Noble model was designed for Purkinje fibres, but was used by him to describe pacemaker activity, that is, self-sustained oscillations; it is quite similar to the Hodgkin-Huxley model, and represents the ionic current as the sum of potassium, sodium and leakage (what Noble calls anionic, i. e., of negatively charged ions) currents, thus

$$C_m \dot{V} = -I_i, \quad I_i = I_{\mathrm{Na}} + I_{\mathrm{K}} + I_{\mathrm{L}}. \tag{5.1}$$

(Note the difference between this and figure 5.4.) The distinction lies in the forms used for the gating variables, and also specifically in considering the potassium currents to be of two types (through two different channels), to accommodate the fact that 'the fibres differ from squid nerve in that depolarisation *decreases* the potassium permeability of the membrane'. The form of the model Noble suggested was based on experimental results, but the precise forms of the functions used were somewhat empirical. This difference from the Hodgkin-Huxley model appears to be significant. Specifically, Noble prescribes (units of V are mV)

$$\begin{aligned} I_{\mathrm{Na}} &= \left[g_0 + g_{\mathrm{Na}} m^3 h\right](V - V_{\mathrm{Na}}), \\ I_{\mathrm{K}} &= (f_K + g_K n^4)(V - V_{\mathrm{K}}), \\ I_{\mathrm{L}} &= g_L (V - V_{\mathrm{L}}), \end{aligned} \tag{5.2}$$

where the fast potassium conductance is given by (units are mS cm^{-2} and V in mV)

$$f_K = 1.2 \exp\left[-\frac{(V+90)}{50}\right] + 0.015 \exp\left[\frac{V+90}{60}\right], \tag{5.3}$$

and the gate variables m, h, n satisfy

$$\begin{aligned} \dot{m} &= \alpha_m (1-m) - \beta_m m, \\ \dot{h} &= \alpha_h (1-h) - \beta_h h, \\ \dot{n} &= \alpha_n (1-n) - \beta_n n, \end{aligned} \tag{5.4}$$

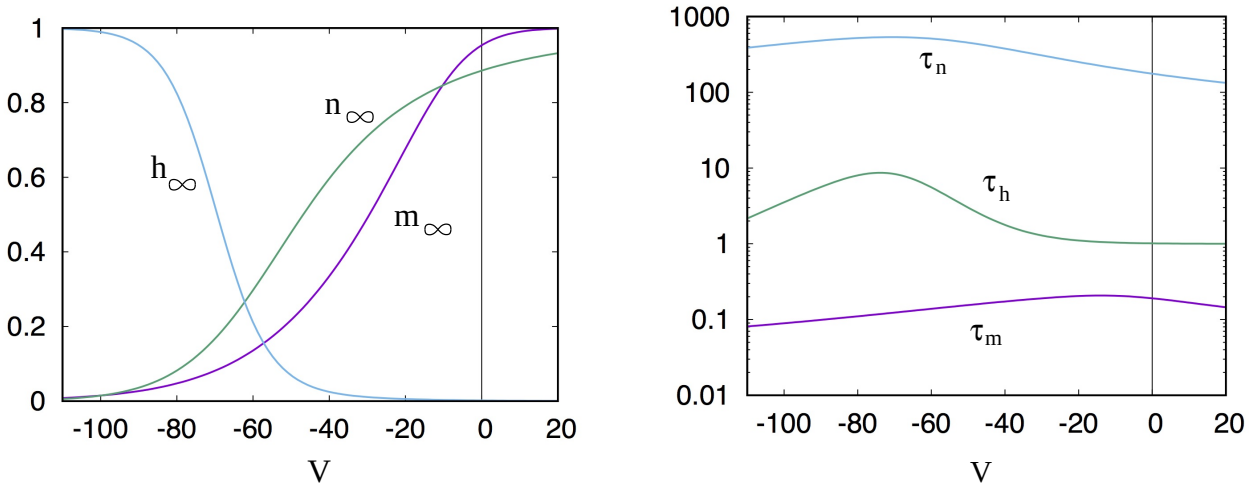

Figure 5.5: Experimentally measured values of the equilibrium gate variables (left) and their relaxation times (right) as functions of potential, as given by Noble (1962). V in mV, τ_k in ms; note the logarithmic scale for τ_k. The resting potential in practice is typically $V_{\text{eq}} \approx -90$ mV.

where the coefficient functions are

$$\alpha_m = \frac{0.1(V+48)}{1-\exp\left[-\dfrac{(V+48)}{15}\right]}, \qquad \beta_m = \frac{0.12(V+8)}{\exp\left[\dfrac{V+8}{5}\right]-1},$$

$$\alpha_h = 0.17\exp\left[-\frac{V+90)}{20}\right], \qquad \beta_h = \frac{1}{1+\exp\left[-\dfrac{(V+42)}{10}\right]},$$

$$\alpha_n = \frac{10^{-4}(V+50)}{1-\exp\left[-\dfrac{(V+50)}{10}\right]}, \qquad \beta_n = 0.002\exp\left[-\frac{(V+90)}{80}\right]. \tag{5.5}$$

For each $k = m, h, n$ we define the time constant τ_k and limit k_∞ by

$$\tau_k = \frac{1}{\alpha_k + \beta_k}, \quad k_\infty = \alpha_k \tau_k, \tag{5.6}$$

so that these equations can be written in the form

$$\tau_k \dot{k} = k_\infty - k. \tag{5.7}$$

The graphs of these time constants and limit functions are shown in figure 5.5. Figure 5.6 shows the fast potassium conductance f_K.

As for the Hodgkin-Huxley equations, there are four equations to solve, but unlike them, there is no clear FitzHugh-Nagumo type reduction to a more amenable two-dimensional system. Instead, a different approach is possible. This is based on the

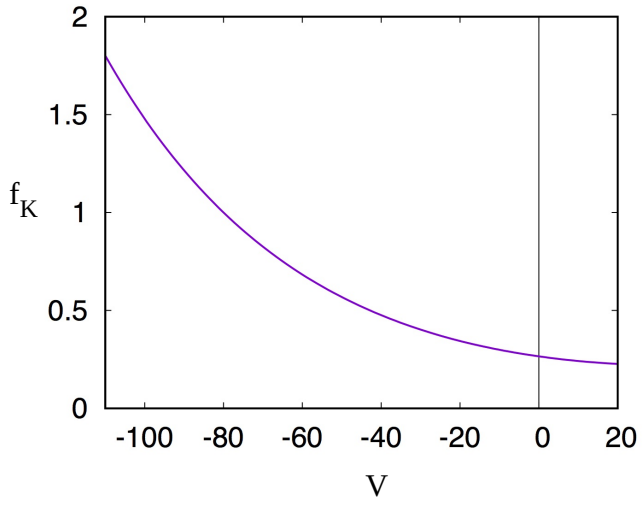

Figure 5.6: The fast potassium conductance f_K (units mS cm^{-2}): V in mV.

observation (see figure 5.5) that the time scales for the gating variables are very different; roughly $\tau_m \sim 0.25$ ms, $\tau_h \sim 8$ ms, and $\tau_n \sim 500$ ms. This allows for successive relaxation of the gate variables, with the n-gate being the slowest-acting.

What is the time scale t_V for the potential V? We define the sodium and potassium conductances to be g and f,

$$g = g_0 + g_{\rm Na} m^3 h, \quad f = f_K + g_K n^4, \tag{5.8}$$

and then it seems, on inspecting table 5.1, that we might expect $g \sim 400$ mS cm^{-2}, and, noting figure 5.6, $f \sim 1$ mS cm^{-2}, hence $t_V \sim \dfrac{C_m}{g} \sim 0.03$ ms, and thus even faster than the m-gate variable. This observation must be tempered with the fact that (see figure 5.5) the product $g_{\rm Na} m_\infty^3 h_\infty$ is actually ~ 1 mS cm^{-2}; in fact it increases monotonically from a value ~ 0.01 mS cm^{-2} at -90 mV to a value of 0.8 mS cm^{-2} at ≈ -10 mV, and decreases above this. So when depolarised, the apparent time scale is about 15 ms, and 1,200 ms at the resting potential! So a degree of wariness is appropriate. However, we shall find that the assumption that $t_V \lesssim \tau_h$ is generally reasonable.

Figure 5.7 reproduces figure 6 in Noble's 1962 paper,[1] and shows self-sustained (pacemaker) periodic behaviour with $g_{\rm L} = 0$. The basis for ignoring the leakage current is that it is small, though in fact it has quite an effect on the value of the resting potential. Indeed, computation of $I_i(V)$ when the gate variables are in equilibrium shows that I_i is a non-monotonic function of V which, if $g_{\rm L} = 0$, has a single steady state at $V_{\rm eq} \approx -35$ mV. Oops! What happened to -90 mV?. It seems this is due to

[1]Sixty years ago, Noble used a Runge-Kutta method, and with a time step of 0.1 ms, his calculations on a Mercury computer took an hour; the computation in figure 5.7 with a time step of 0.001 ms took about two seconds on a 2016 Macbook Pro.

Parameter	Typical value
C_m	12 μF cm^{-2}
g_{K}	1.2 mS cm^{-2}
g_{L}	$\lesssim$ 0.405 mS cm^{-2}
g_{Na}	400 mS cm^{-2}
g_0	0.14 mS cm^{-2}
V_{Na}	40 mV
V_{K}	-100 mV
V_{L}	-60 mV

Table 5.1: Parameters in the Noble model.

the specific choice of the constituent functions. The function $I_i(V)$ (using m_∞, n_∞ and h_∞) resembles a cubic. When $g_{\mathrm{L}} = 0$, its maximum is at about -78 mV, and only slightly negative. If g_{L} is *reduced*, two further roots appear, and if $g_{\mathrm{L}} = -0.234$ mS cm^{-2}, the lower of these is at -90 mV. There are of course then three steady states of the system. It seems this is due to the rather sensitive prescription of the model in terms of the various coefficient functions chosen.

Figure 5.8 shows a close up of the peak, and helps us to understand the dynamics of

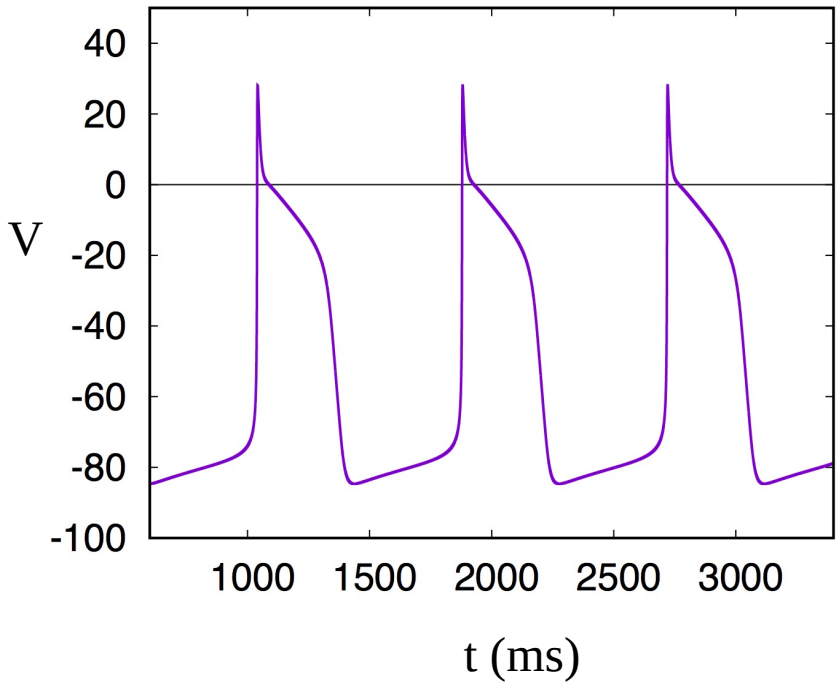

Figure 5.7: Periodic solutions of the Noble model; parameters as in table 5.1, with $g_{\mathrm{L}} = 0$.

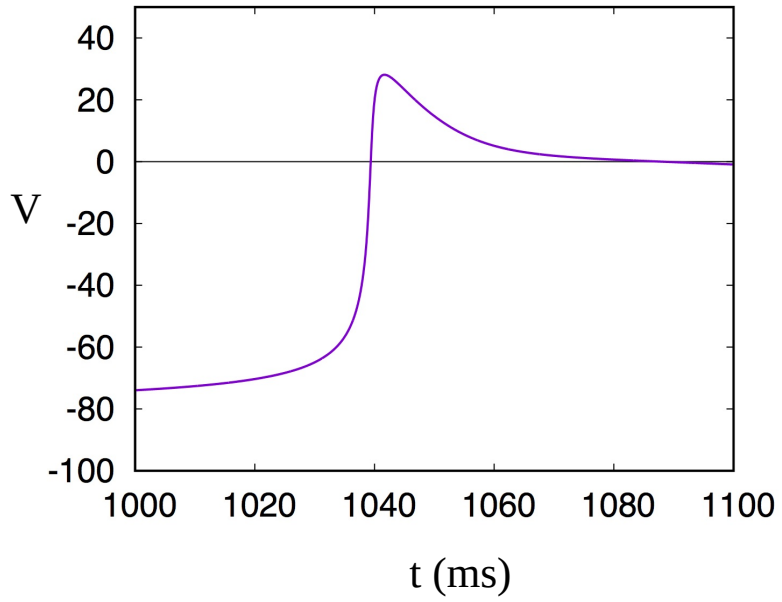

Figure 5.8: Close-up of the fast depolarisation of the solution in figure 5.7. The units of V are mV.

the model. Starting at the minimum of the oscillation, at around -80 mV, we can take $m \approx m_\infty(V)$, and indeed this assumption remains approximately valid throughout the cycle.[2] On the slow part of the oscillation, h and V are in approximate equilibrium also, and n slowly relaxes towards n_∞: this is similar to the FitzHugh-Nagumo model.

The fast phase occurs on the time scale τ_h. On this time scale $n \approx$ constant, and the fast phase is determined by the second order system (putting $g_{\rm L} = 0$)

$$\begin{aligned} \tau_h \dot{h} &= h_\infty(V) - h, \\ C_m \dot{V} &= -\left[\{g_0 + g_{\rm Na} m_\infty^3(V) h\}(V - V_{\rm Na}) + f(V)(V - V_{\rm K})\right], \end{aligned} \tag{5.9}$$

and we might aim to seek a trajectory in the (h, V) phase plane connecting two fixed points of these equations: this is essentially what happens, with a slight twist. Thus it seems that an analysis of the Noble model is possibly by its reduction to a third order system, and an analysis similar to that used in analysing periodic travelling waves.

To proceed formally, we take the gate times τ_k to be constant, $m = m_\infty(V)$, and non-dimensionalise the system by scaling the variables as

$$V \sim |V_{\rm K}|, \quad t \sim \tau_h, \tag{5.10}$$

[2]Examination of the numerical solution shows that it is very accurate everywhere, except in the fast depolarisation phase, where $m_\infty - m$ rises to a maximum of about 0.12 before decreasing again. Nevertheless, we retain the assumption that $m \approx m_\infty$ throughout.

Parameter	Typical value
τ_m	0.25 ms
τ_h	8 ms
τ_n	500 ms
$v_{\rm Na}$	0.4
$v_{\rm L}$	0.6
γ_0	0.09
$\gamma_{\rm K}$	0.8
$\gamma_{\rm L}$	< 0.27
$\gamma_{\rm Na}$	267
ε	0.016
ϕ_K	< 1

Table 5.2: Time scales and dimensionless parameters. The estimate of the decreasing function ϕ_K is based on $f_K < 1.5$ mS cm^{-2} for $V > -100$ mV.

so that the equations become

$$\begin{aligned} \dot{n} &= \varepsilon(n_\infty - n), \\ \dot{h} &= h_\infty - h, \\ \dot{V} &= -G(V,h,n), \end{aligned} \tag{5.11}$$

where

$$G(V,h,n) = \{\gamma_0+\gamma_{\rm Na}m_\infty^3(V)h\}(V-v_{\rm Na})+\{\phi_K(V)+\gamma_{\rm K}n^4\}(V+1)+\gamma_{\rm L}(V+v_{\rm L}), \tag{5.12}$$

and

$$\varepsilon = \frac{\tau_h}{\tau_n}, \quad v_{\rm Na} = \frac{V_{\rm Na}}{|V_{\rm K}|}, \quad v_{\rm L} = \frac{|V_{\rm L}|}{|V_{\rm K}|}, \quad \gamma_0 = \frac{g_0\tau_h}{C_m},$$
$$\gamma_{\rm Na} = \frac{g_{\rm Na}\tau_h}{C_m}, \quad \phi_K(V) = \frac{\tau_h f_K(V)}{C_m}, \quad \gamma_{\rm K} = \frac{g_{\rm K}\tau_h}{C_m}, \quad \gamma_{\rm L} = \frac{g_{\rm L}\tau_h}{C_m}. \tag{5.13}$$

We formally assume that $G \gtrsim 1$ (and specifically $\gg \varepsilon$). Values of the parameters are given in table 5.2, which shows that indeed $G \gtrsim 1$. The fact that $\gamma_{\rm Na} \gg 1$ does not matter, since if G is large it simply emphasises the validity of the fast phase approximation. (In principle, one should also pay attention to the validity of the m-equilibrium, as mentioned in the footnote on page 89.)

We limit ourselves to a discussion of the fast upstroke of the action potential. We have $n \approx$ constant, and with $\gamma_{\rm L} = 0$, the equations are

$$\begin{aligned} \dot{h} &= h_\infty - h, \\ \dot{V} &= -G(V,h,n), \\ G(V,h,n) &= -\{\gamma_0 + \gamma_{\rm Na}m_\infty^3(V)h\}(v_{\rm Na} - V) + \phi(V+1), \end{aligned} \tag{5.14}$$

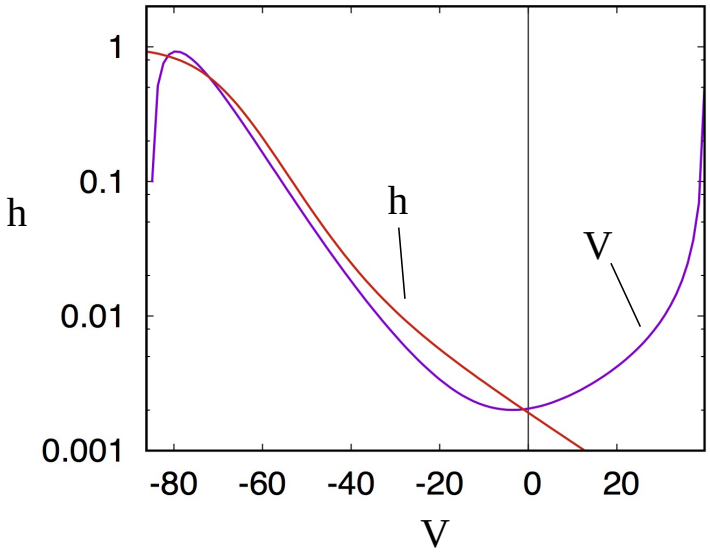

Figure 5.9: The h and V nullclines of (5.14) for values of $g_{\rm L} = 0$ and $n = 0.5$. V is plotted in dimensional units. Note the logarithmic scale for h.

where we define

$$\phi = \phi_K(V) + \gamma_{\rm K} n^4. \tag{5.15}$$

In the phase plane, the h-nullcline is just $h = h_\infty(V)$, as shown in figure 5.5. The V-nullcline is given by

$$h = h_0(V) = \frac{S}{\gamma_{\rm Na} m_\infty^3(V)}, \quad S = \frac{\phi(V+1)}{v_{\rm Na} - V} - \gamma_0. \tag{5.16}$$

For $V < v_{\rm Na}$, S is a monotonically increasing function of V, $S = 0$ for $V \approx -1$ and $S \to \infty$ as $V \to v_{\rm Na}$. The V-nullcline would then have the same shape, except that as V decreases, m_∞ decreases dramatically, and this causes h_0 to increase again before it reaches zero. The possibility thus arises of the V-nullcline intersecting the h-nullcline not only at large V, as it must, but also at small V. It turns out, for the particular choice of functions used by Noble, that this indeed occurs. Figure 5.9 shows that this is indeed the case (the h values are plotted logarithmically, as the two nullclines are so close). It is clear that the precise way in which the nullclines intersect is very sensitive to the precise forms of the constituent functions, and in particular m_∞.

In this figure, it is not difficult to show that the extreme fixed points are stable, and the intermediate one is a saddle: we are *not* trying to construct connecting orbits here! Now when $V = -80$ mV, figure 5.5 shows that $n_\infty \approx 0.1$, whereas when $V \approx 0$ mV, $n_\infty \approx 0.9$. Thus in figure 5.9 (where $n = 0.5$), if we are at the left hand stable fixed point, $\dot{n} < 0$ and thus ϕ decreases, and this decreases $h_0(V)$: the V-nullcline gradually lowers. At a critical value of n (about 0.4), the two left hand fixed points coalesce, and there is a rapid transient to the other fixed point this is the fast

depolarisation of the action potential. Moreover this transient starts off slowly, but when $V > -60$ mV, the sodium conductance term becomes large, and the trajectory in the phase plane becomes horizontal and rapidly approaches the V-nullcline on a dimensionless time scale $\sim \dfrac{1}{\gamma_{\mathrm{Na}}}$; it then slowly migrates down the V-nullcline towards the remaining steady state on the longer dimensionless time scale ~ 1. This provides the explanation for the very fast upstroke and the slightly slower relaxation that can be seen in figure 5.8.

Having reached a quasi-steady state near $V = 0$ mV, we still have $n \approx 0.4$ but now $n_\infty \approx 0.9$, so n starts to increase, and as it does so, ϕ and thus h_0 start to increase. Consequently the two left hand fixed points reappear, and as the V-nullcline continues to rise, eventually the two *right-hand* fixed points coalesce and disappear (at about $n = 0.7$), and there follows a rapid decrease of V back to the next slow, polarised phase. In this way the oscillation continues.

However, it is obviously sensitive to the precise choice of functions used. It is fairly clear how minor adjustments to the constituent functions can be made in order that the resting potential is about -90 mV, and that it is stable but excitable. This is also why inclusion of the leakage term causes the oscillations to disappear. As g_{L} increases, the slope of the V-nullcline, given by (5.16) in the form

$$h = h_0(V) = \frac{S}{\gamma_{\mathrm{Na}} m_\infty^3(V)}, \quad S = \frac{\phi(V+1) + \gamma_{\mathrm{L}}(V + v_{\mathrm{L}})}{v_{\mathrm{Na}} - V} - \gamma_0, \tag{5.17}$$

increases, so that when $g_{\mathrm{L}} = 0.4$ mS cm^{-2}, there is only one intersection of the nullclines, and consequently only one fixed point, and it is stable.

5.3 Wave propagation in two dimensions

Ventricular myocytes are very small, typically only 100 μm long and 10 μm wide. The heart muscle is made of billions of interconnected ventricular myocytes which divide the tissue into two regions: the intracellular space and the extracellular space. The myocytes are aligned approximately parallel to each other defining the *fibre direction*. The fibre structure of cardiac tissue makes it anisotropic. Adjacent cells form connections in regions called *intercalated discs*. Within these discs are *gap junctions*, which form electrical connections between the cells. The majority of the gap junctions are concentrated at the ends of the cells, leading to a larger electrical conductivity in the direction of the cardiac fibres compared with the direction perpendicular to the fibres. In theory it is possible to write down equations that model the fine structure of the intercellular connections; however, to solve these equations is computationally impossible in practice. The macroscopic length scale over which the potential varies is much greater than the microscopic length scale of the cellular structure, and therefore homogenisation techniques can be used to yield a continuum description of the tissue. Without going into detail, we will model this spatial transmission of electric potential by a diffusion term, just as we did in the Hodgkin-Huxley model. Our main

focus in this section is to provide a geometrical theory of wave propagation which allows the study of wave propagation in more than one dimension. To be specific, we will begin by considering periodic travelling wave trains, which, as shown in chapter 4, are the common consequence of adding diffusion to oscillatory reaction kinetics, such as describe the pacemaker potential of the cells in the sino-atrial node. The propagation of solitary waves in this context can be understood as a limit in which the period of the underlying limit cycle tends to infinity. In a second section, we provide a description for the propagation of waves which are quasi-one-dimensional, in the sense that the radius of curvature of the wavefront is much greater than the thickness of the wave profile. Unfortunately, the presentation of the material is rather abstract.

5.3.1 Periodic wave propagation

Suppose that a vector of concentrations $\mathbf{w}$ satisfies the general (non-dimensional) reaction-diffusion equation

$$\mathbf{w}_t = \mathbf{f}(\mathbf{w}) + \varepsilon \nabla^2 \mathbf{w}, \tag{5.18}$$

where $\mathbf{w} \in \mathbf{R}^n$ and ε is a small parameter. The assumption here is that if D is the diffusion coefficient and T is the reaction time scale, then the macroscopic length scale $l \gg \sqrt{DT}$. While this may be true in chemical reactions, for example, it is not actually appropriate in the heart. In the simplest case $n = 2$, and while it is not necessary that the diffusivities of the different species are all the same, it is simpler to suppose this for the purposes of exposition.

Suppose that the reaction kinetics admit an attractive limit cycle for the underlying system $\mathbf{w}_t = \mathbf{f}(\mathbf{w})$ of period T, and denote this as $\mathbf{W}_0(t)$, thus

$$\mathbf{W}_0' = \mathbf{f}(\mathbf{W}_0); \tag{5.19}$$

here the prime denotes differentiation with respect to the argument of $\mathbf{W}_0$. Now because ε is small, different parts of the medium (different cells in our case) will oscillate with the same period but with different phase. Such spatial phase gradients can be expected to lead to solutions in which the phase varies in space, but also (because diffusion is small), the spatial variation of the phase will evolve on a slow time scale. This motivates the introduction of a slow time variabled e τ given by

$$\tau = \varepsilon t. \tag{5.20}$$

The procedure we are about to adopt is called the *method of multiple time scales*. Formally, we seek solutions of (5.18) in the form $\mathbf{w}(\mathbf{x}, t, \tau)$, where t and τ are considered to be independent variables. What we are doing is embedding the problem in a higher dimensional space, but we will eventually project the resulting solution on to the line $\tau = \varepsilon t$. The reason for doing this is that we want to find a uniform asymptotic expansion for $\mathbf{w}$, and a straightforward approach will fail (as we will see).

Introducing this second slow time variable means that the equation takes the form

$$\mathbf{w}_t + \varepsilon \mathbf{w}_\tau = \mathbf{f}(\mathbf{w}) + \varepsilon \nabla^2 \mathbf{w}. \tag{5.21}$$

Expanding $\mathbf{w}$ as

$$\mathbf{w} \sim \mathbf{w}_0 + \varepsilon \mathbf{w}_1 + \dots \tag{5.22}$$

leads to

$$\begin{aligned} \mathbf{w}_{0t} &= \mathbf{f}(\mathbf{w}_0), \\ \mathbf{w}_{1t} - J\mathbf{w}_1 &= -\mathbf{w}_{0\tau} + \nabla^2 \mathbf{w}_0, \end{aligned} \tag{5.23}$$

and so on; here $J = D\mathbf{f}(\mathbf{w}_0)$ is the Jacobian of $\mathbf{f}$ at $\mathbf{w}_0$. After an initial transient, we may take

$$\mathbf{w}_0 = \mathbf{W}_0(t + \psi), \tag{5.24}$$

where $\psi(\tau, \mathbf{x})$ is the slowly-varying phase, and the Jacobian $J = D\mathbf{f}(\mathbf{W}_0)$ is a time-periodic matrix. Thus we find that $\mathbf{w}_1$ satisfies

$$\mathbf{w}_{1t} - J\mathbf{w}_1 = -(\psi_\tau - \nabla^2 \psi)\mathbf{W}_0' + |\boldsymbol{\nabla}\psi|^2 \mathbf{W}_0''. \tag{5.25}$$

Note that $\mathbf{s} = \mathbf{W}_0'$ satisfies the homogeneous equation $\mathbf{s}_t - J\mathbf{s} = \mathbf{0}$. It follows that the solution of (5.25) is

$$\mathbf{w}_1 = -t(\psi_\tau - \nabla^2 \psi)\mathbf{s} + |\boldsymbol{\nabla}\psi|^2 \mathbf{u}, \tag{5.26}$$

where

$$\mathbf{u} = M(t) \int_0^t M^{-1}(\theta) J(\theta) \mathbf{s}(\theta)\, d\theta + M(t), \tag{5.27}$$

and M is a fundamental matrix for the homogeneous equation, i. e., $M' = JM$, $M(0) = I$. Floquet's theorem implies that

$$M = P e^{t\Lambda}, \tag{5.28}$$

where P is a periodic matrix of period T (the same as that of the limit cycle $\mathbf{W}_0$).[3] We can take the matrix Λ to be diagonal if the characteristic multipliers are distinct, and since we assume $\mathbf{W}_0$ is attracting, the eigenvalues of Λ will all have negative real part, except one of zero corresponding to $\mathbf{s}$. With a suitable choice of basis, we then have

$$(e^{t\Lambda})_{ij} \to \delta_{i1}\delta_{j1} \quad \text{as} \quad t \to \infty, \tag{5.29}$$

i. e., a matrix with the single non-zero element being unity in the first element. In this case the first column of P is $\mathbf{s}$, i. e., $P_{i1} = s_i$.

From (5.27), we have

$$\mathbf{u} = P(t) \int_0^t e^{\eta\Lambda} P^{-1}(t - \eta) J(t - \eta) \mathbf{s}(t - \eta)\, d\eta + M\mathbf{c}. \tag{5.30}$$

[3]This is not so complicated. Suppose $M(T)$ has eigenvalues $e^{\lambda_i T}$. By a change of coordinates if necessary, we take $M(T) = e^{T\Lambda}$ to be diagonal, $\Lambda = \operatorname{diag}(\lambda_i)$. Then if $F = e^{t\Lambda}$, we have $F(0) = M(0) = I$, $F(T) = M(T)$, and so the matrix $P = MF^{-1}$ is periodic: hence (5.28).

The effect of the transient dies away as $t \to \infty$, and if we ignore it, then we can take $M_{ij} = s_i \delta_{j1}$, whence $M\mathbf{c} = c_1 \mathbf{s}$, and thus

$$\mathbf{u} = \mathbf{s} \left[\int_0^t \alpha(\eta)\, d\eta + c_1 \right], \tag{5.31}$$

where the periodic function α is given by[4]

$$\alpha = (P^{-1})_{1m} J_{mj} s_j. \tag{5.32}$$

We define the mean of α to be

$$\bar{\alpha} = \frac{1}{T} \int_0^T \alpha(\eta)\, d\eta, \tag{5.33}$$

so that

$$\beta = \int_0^t (\alpha - \bar{\alpha})\, d\eta \tag{5.34}$$

is periodic with period T. Then (5.26) is

$$\mathbf{w}_1 = \big[t\{ -\psi_\tau + \nabla^2 \psi + \bar{\alpha} |\boldsymbol{\nabla} \psi|^2 \} + c_1 + \beta(t) \big] \mathbf{s}. \tag{5.35}$$

Aside from all the technical jargon, what have we actually done here? (5.35) provides the second term in the expansion in powers of ε. Suppose we had done this straightforwardly, without the slow time variable; then we would have obtained the same result (via all the technical wizardry) but without the term ψ_τ. And then what? The term $\propto t$ would grow in time until, when $t \sim \dfrac{1}{\varepsilon}$, the asymptotic nature of the expansion breaks down. Well, unless $\nabla^2 \psi + \bar{\alpha} |\boldsymbol{\nabla} \psi|^2 = 0$. But that would imply $\nabla^2 e^{\bar{\alpha} \psi} = 0$, and thus (with no boundary forcing) $\psi =$ constant: but there is no reason this should be the case. The whole point of the method of multiple scales is to provide the flexibility via the extra slow time variable to suppress these growth terms, called *secular terms*, and we can do this by requiring the phase ψ to satisfy the evolution equation

$$\psi_\tau = \nabla^2 \psi + \bar{\alpha} |\boldsymbol{\nabla} \psi|^2. \tag{5.36}$$

This is an integrated form of Burgers' equation; in one dimension, $u = -\dfrac{\psi_X}{2\bar{\alpha}}$ satisfies $u_\tau + u u_X = u_{XX}$. Disturbances will form shocks, which are jumps of phase gradient. More generally, if $\mathbf{u} = -\dfrac{\boldsymbol{\nabla} \psi}{2\bar{\alpha}}$, then (bearing in mind that $\boldsymbol{\nabla} \times \mathbf{u} = \mathbf{0}$) we find

$$\mathbf{u}_\tau + (\mathbf{u}.\boldsymbol{\nabla})\mathbf{u} = \nabla^2 \mathbf{u}, \tag{5.37}$$

which is the Navier-Stokes equation with no pressure term. Phase gradients move down phase gradients, and form defects where the (sub-)characteristics intersect.[5] (5.36) has both target and spiral wave patterns, as we now show. (Target and spiral waves were discussed in section 4.2.1 on page 77, and were visualised in figure 4.19.)

[4] We use the summation convention, which implies summation over repeated suffixes.

[5] Physicists call (5.36) the KPZ equation (after Kardar *et al.* (1986)). The substitution $\phi = \exp(\bar{\alpha}\psi)$ reduces it to the diffusion equation $\phi_\tau = \nabla^2 \phi$ for ϕ; this is the Hopf–Cole transformation (see Whitham 1974).

5.3.2 Target patterns and spiral waves

Solutions of (5.36) which vary with $\mathbf{x}$ correspond to travelling wave trains. For example, in one dimension, waves travel locally at speed $\dfrac{dx}{dt} \approx -\left(\dfrac{\partial\psi}{\partial x}\right)^{-1}$. In the heart, a more relevant situation is to suppose that, by analogy with the SA node pacemaker, ψ is prescribed at a central nucleus, which we take to be the origin. In this case, *target patterns* are formed; these are circular wavefronts which originate from a point. To be specific, suppose that

$$\psi = \tau \quad \text{at} \quad r = b, \tag{5.38}$$

where r is the polar radius in two dimensions; we then seek a solution $\psi = \tau - f(r)$, where f satisfies

$$f'' + \frac{1}{r}f' - \bar{\alpha}f'^2 + 1 = 0, \tag{5.39}$$

together with $f(b) = 0$ and an appropriate condition at large r; this is prescribed by restricting attention to a target pattern originating near the origin and thus suppressing incoming waves (this is known as a radiation condition). The relevant solution if $\bar{\alpha} > 0$ is

$$f(r) = -\frac{1}{\bar{\alpha}} \ln\left[\frac{K_0(\sqrt{\bar{\alpha}}r)}{K_0(\sqrt{\bar{\alpha}}b)}\right], \tag{5.40}$$

where K_0 is the modified Bessel function of the second kind of order zero. The other Bessel function I_0 is suppressed because of the radiation condition (it produces incoming waves). At large r, $\psi \sim \tau - r/\sqrt{\bar{\alpha}}$, which represents an outward travelling wave of speed $dr/dt \approx \sqrt{\bar{\alpha}}$. If, on the other hand, $\bar{\alpha} < 0$, then K_0 is replaced by a combination of the Bessel functions J_0 and Y_0, and the solution is singular at finite r, and target pattern solutions of this type do not exist. More generally, if $\psi = \Omega\tau$ on $r = b$, then target patterns exist if $\bar{\alpha}\Omega > 0$.

Spiral waves can also be described as solutions of (5.36). We put

$$\psi = \Omega\tau + m\theta - g(r), \tag{5.41}$$

where m is an integer, associated with an appropriate boundary condition at a central 'dead' core at $r = b$. Again we require $\bar{\alpha}\Omega > 0$, and the solution, if we suppose $\Omega > 0$, is

$$g(r) = -\frac{1}{\bar{\alpha}} \ln\left[\frac{K_\nu(\sqrt{\bar{\alpha}\Omega}r)}{K_\nu(\sqrt{\bar{\alpha}\Omega}b)}\right], \quad \nu = i\alpha m, \tag{5.42}$$

if we suppose $g = 0$ at $r = b$. If $\Omega < 0$, we select the modified Bessel function I_0 instead of K_0 (to maintain the radiation condition). At large r,

$$\psi \sim \Omega\tau + m\theta - \sqrt{\frac{\Omega}{\bar{\alpha}}}r. \tag{5.43}$$

This is an Archimedean spiral, which winds outward clockwise if $m > 0$, and then rotates anti-clockwise. In practice, and as shown in figure 5.11, the existence of spiral waves is associated with a core which allows the maintenance of a travelling wave circulating round its boundary.

5.3.3 Curved front propagation

We saw in chapter 3 that action potentials (solitary waves) propagating in neurons consist of two wave fronts of width ~ 5 mm separated by a depolarised region of length ~ 30 cm. In practice, therefore, the solitary wave effectively consists of two separate travelling waves which are respectively activating and de-activating. The same is true in the heart. The upstroke of the action potential of the sino-atrial node causes propagation of an activation wave across the heart, causing contraction, and the downstroke sends another wave, causing relaxation. Thus in practice it is sufficient in studying the propagation of curvilinear waves to use just the activation part of a FitzHugh–Nagumo type model.

To be specific, we consider the equation

$$v_t = f(v) + \nabla^2 v, \tag{5.44}$$

which we suppose admits a travelling wave solution in one dimension of the form

$$v = V(\xi), \quad \xi = ct - x, \quad c > 0, \quad V(\infty) = 1, \quad V(-\infty) = 0, \tag{5.45}$$

thus

$$V'' - cV' + f(V) = 0, \tag{5.46}$$

and c is uniquely determined if f has a cubic-type shape.

Now we want to find a solution in two (or three) dimensions which is slowly varying in directions transverse to the direction of propagation (think for example of a circularly expanding wavefront of large radius). On the large scale, the travelling wave front at each instant is approximately a surface. This family of surfaces, parameterised by time, defines a function $\psi(\mathbf{x}, t)$, such that the front position can be taken to be

$$\psi(\mathbf{x}, t) = 0, \tag{5.47}$$

and we will suppose $\psi < 0$ ahead of the front. To think of a specific example, we might think of a target pattern with $\psi = ct - r$. We want to define a curvilinear coordinate system in which ξ measures distance along the normals to the family of curves $\psi =$ constant, where we define the unit normal to be in the direction of propagation, thus

$$\mathbf{n} = -\frac{\boldsymbol{\nabla}\psi}{|\boldsymbol{\nabla}\psi|}, \tag{5.48}$$

but ξ points in the opposite direction (thus in the same direction as increasing ψ). We then have (at fixed t)

$$\delta\psi = \boldsymbol{\nabla}\psi.\,\delta\mathbf{x} = -|\boldsymbol{\nabla}\psi|\,\mathbf{n}.\delta\mathbf{x} = |\boldsymbol{\nabla}\psi|\,\delta\xi. \tag{5.49}$$

Next, we seek a solution to (5.44) in the form $v = V[\psi(\mathbf{x}, t)]$; it follows that

$$V'(\psi_t - \nabla^2\psi) = f(V) + V''|\boldsymbol{\nabla}\psi|^2, \tag{5.50}$$

where $V'(\psi) = \dfrac{dV}{d\psi}$. We want to transform this to an ordinary differential equation in physical space; to do this we use (5.49), and specifically we take

$$V_\xi \approx |\boldsymbol{\nabla}\psi| V'(\psi). \tag{5.51}$$

There is a subtle point here. It is tempting to think that the wavefront locations $\psi = 0$ themselves define the orthogonal curvilinear system used at any fixed t, but this is not necessarily the case (though it can be, as in a target pattern). Think, for example, of a wave front which is planar (say, $x = x_0$) at time t_0 but which is curved at time $t > t_0$. The natural coordinate system at t_0 is Cartesian, but not at later time. So the transformation from a coordinate system (ψ, θ, χ), say, to the system (ξ, η, ζ) has $\psi = \psi(\xi, \eta, \zeta)$, and the quasi-one-dimensional assumption in (5.51) is based on the assumption that the wavefront is slowly varying in its transverse directions η and ζ compared to the variation of ψ within it: the thickness of the wave is much less than its curvature. A slightly different version of this argument is given in the following section.

Adopting (5.51), it follows that

$$V_{\xi\xi} + \frac{V_\xi}{|\boldsymbol{\nabla}\psi|}\left\{ -\frac{\partial |\boldsymbol{\nabla}\psi|}{\partial \xi} + \nabla^2\psi - \psi_t \right\} + f(V) = 0, \tag{5.52}$$

and the unique solution requires

$$\psi_t = \nabla^2\psi - \frac{\partial |\boldsymbol{\nabla}\psi|}{\partial \xi} + c|\boldsymbol{\nabla}\psi|. \tag{5.53}$$

This equation can be compared with (5.36), since ψ here can be identified as the phase of the wave. It can be written as

$$\frac{\psi_t}{|\boldsymbol{\nabla}\psi|} = c + \frac{1}{|\boldsymbol{\nabla}\psi|}\left[\nabla^2\psi - \frac{\partial |\boldsymbol{\nabla}\psi|}{\partial n} \right]. \tag{5.54}$$

Now the second term on the right hand side of this expression is just the negative of the divergence of the normal

$$\mathbf{n} = -\frac{\boldsymbol{\nabla}\psi}{|\boldsymbol{\nabla}\psi|}, \tag{5.55}$$

and thus

$$v_n = c - \boldsymbol{\nabla}.\,\mathbf{n}, \quad v_n = \frac{\psi_t}{|\boldsymbol{\nabla}\psi|}; \tag{5.56}$$

v_n is the normal velocity of the front, and $\boldsymbol{\nabla}.\,\mathbf{n}$ is its curvature. The equation (5.56) is sometimes called the *eikonal equation.*[6] It relates the normal velocity v_n of the surface

[6]More commonly, the eikonal equation is $|\boldsymbol{\nabla}\phi| = 1$. This is regained if $\psi = \phi(\mathbf{x}) - t$ and the curvature term is ignored.

to the curvature $\boldsymbol{\nabla}.\,\mathbf{n}$. At large distance and time, we can rescale $\psi \sim t \sim \mathbf{x} \sim \dfrac{1}{\varepsilon}$, $\varepsilon \ll 1$, and then (5.56) takes the form

$$\psi_t = c|\boldsymbol{\nabla}\psi| + \varepsilon|\boldsymbol{\nabla}\psi|\boldsymbol{\nabla}.\left(\frac{\boldsymbol{\nabla}\psi}{|\boldsymbol{\nabla}\psi|}\right), \tag{5.57}$$

which again can be compared to (5.36).

5.3.4 A more formal derivation

The above discussion is accurate but rather slapdash. Dare one say, it takes an engineer's, or even a physicist's, approach. It gets the right answer but without due regard to the error corrections associated with geometric nonlinearity. To be satisfied mathematically, we need to come at this another way.

We return to the equation (5.44), but in its 'outer' scaled form (cf. (3.16))

$$\varepsilon v_t = f(v) + \varepsilon^2\nabla^2 v, \tag{5.58}$$

and we suppose that $f(0) = f(1) = 0$ as before. An 'outer' approximation to (5.58) is just $f(v) = 0$, and thus $v = 0$ or $v = 1$. As a particular case, we consider a piecewise continous solution in which v jumps from zero to one across a surface. We suppose this surface can move in time, and thus it defines a family of non-intersecting surfaces $\psi(\mathbf{x}, t) = 0$, and we suppose that $v = 0$ for $\psi < 0$, $v = 1$ for $\psi > 0$. Because of the non-intersection of this family, we can suppose that there is a unique value of t for each $\mathbf{x}$ for which $\psi = 0$. We denote this function as $T(\mathbf{x})$, thus $\psi[\mathbf{x}, T(\mathbf{x})] \equiv 0$.

The family of surfaces can be used to construct a set of orthogonal curvilinear coordinates $\mathbf{X} = (X_1, X_2, X_3)$. First we construct the X_1-axis, with unit vector $\mathbf{n}$. The X_1-axis is normal to each member of the family $\psi(\mathbf{x}, t) = 0$. How do we know we can construct this? We simply solve (in principle) the equation

$$\frac{\partial \mathbf{X}}{\partial s} = \lambda\mathbf{n}[\mathbf{X}, T(\mathbf{X})], \tag{5.59}$$

where the unit normal $\mathbf{n}$ is defined by (5.55); this is subject to the initial condition $\mathbf{X} = \mathbf{X}_0(\eta, \zeta)$ at $s = 0$, where η and ζ are orthogonal curvilinear coordinates on the initial (at $t = 0$) surface, i. e., $\psi[\mathbf{X}_0(\eta, \zeta), 0] \equiv 0$. The coefficient λ is arbitrary, and can generally be a function of $\mathbf{X}$. It is obviously convenient to choose it so that $s = t$, and this requires $\lambda = v_n$, where v_n is the normal velocity of the interface; we assume this. Distance along these curves is denoted $X_1 = s$, and the corresponding unit vector is denoted $\mathbf{E}_1 = \mathbf{n}[\mathbf{X}, T(\mathbf{X})]$. By choice of $\lambda = v_n$, the front position is simply $X_1 = t$.

The other two coordinates can be taken to be η and ζ at $t = 0$, but these will not in general extend orthogonally to $t > 0$. Instead, one can for example select the unit vectors

$$\mathbf{E}_2 = \left|\frac{\partial \mathbf{X}}{\partial \eta}\right|^{-1}\frac{\partial \mathbf{X}}{\partial \eta}, \quad \mathbf{E}_3 = \mathbf{E}_1 \times \mathbf{E}_2, \tag{5.60}$$

with corresponding coordinates X_2 and X_3. Here, $\mathbf{X} = \mathbf{X}(X_1, \eta, \zeta)$ is the solution of (5.59). $\mathbf{E}_2$ is a tangent vector to the front.

We now transform (5.58) to the orthogonal curvilinear coordinates (X_1, X_2, X_3). Note that the change of coordinates is independent of t. The result of this is

$$\varepsilon v_t = f(v) + \frac{\varepsilon^2}{h_1 h_2 h_3}\left[\frac{\partial}{\partial X_1}\left(\frac{h_2 h_3}{h_1}\frac{\partial v}{\partial X_1}\right) + \frac{\partial}{\partial X_2}\left(\frac{h_3 h_1}{h_2}\frac{\partial v}{\partial X_2}\right) + \frac{\partial}{\partial X_3}\left(\frac{h_1 h_2}{h_3}\frac{\partial v}{\partial X_3}\right)\right], \tag{5.61}$$

where the metric coefficients are given by

$$h_i = \left|\frac{\partial \mathbf{X}}{\partial X_i}\right|. \tag{5.62}$$

We can assume these coefficients and their derivatives are $O(1)$. Also, because X_2 and X_3 are functions of η and ζ, our earlier choice of the metric λ on the X_1-axis implies

$$h_1 = v_n. \tag{5.63}$$

The front is defined by $X_1 = t$, and so we seek a solution for the front structure by putting

$$X_1 = t - \frac{\varepsilon \xi}{h_1}, \tag{5.64}$$

under which the derivatives transform as

$$\begin{aligned} \frac{\partial}{\partial t} &\to \frac{\partial}{\partial t} + \frac{h_1}{\varepsilon}\frac{\partial}{\partial \xi}, \qquad \frac{\partial}{\partial X_1} \to -\frac{h_1}{\varepsilon}\frac{\partial}{\partial \xi} + \frac{\xi}{h_1}\frac{\partial h_1}{\partial X_1}\frac{\partial}{\partial \xi}, \\ \frac{\partial}{\partial X_2} &\to \frac{\partial}{\partial X_2} + \frac{t}{\varepsilon}\frac{\partial h_1}{\partial X_2}\frac{\partial}{\partial \xi}, \qquad \frac{\partial}{\partial X_3} \to \frac{\partial}{\partial X_3} + \frac{t}{\varepsilon}\frac{\partial h_1}{\partial X_3}\frac{\partial}{\partial \xi}. \end{aligned} \tag{5.65}$$

We now anticipate that in fact $h_1 = v_n$ is approximately constant, so that its derivatives are small, and explicitly $O(\varepsilon)$. In that case, (5.61) takes the form

$$\varepsilon v_t + h_1 v_\xi = f(v) + v_{\xi\xi} - \frac{\varepsilon v_\xi}{h_1 h_2 h_3}\frac{\partial (h_2 h_3)}{\partial X_1} + O(\varepsilon^2). \tag{5.66}$$

Now in fact

$$\boldsymbol{\nabla}.\,\mathbf{n} = \frac{1}{h_1 h_2 h_3}\frac{\partial (h_2 h_3)}{\partial X_1}, \tag{5.67}$$

thus correct to $O(\varepsilon^2)$, (5.66) is

$$\varepsilon v_t = v_{\xi\xi} - \left[h_1 + \varepsilon \boldsymbol{\nabla}.\,\mathbf{n}\right] v_\xi + f(v), \tag{5.68}$$

and consistency with (5.46) now shows that

$$h_1 = v_n = c - \varepsilon \boldsymbol{\nabla}.\,\mathbf{n} + O(\varepsilon^2), \tag{5.69}$$

which reproduces (5.57). Our *ansatz* that $v_n \approx$ constant is thus vindicated.

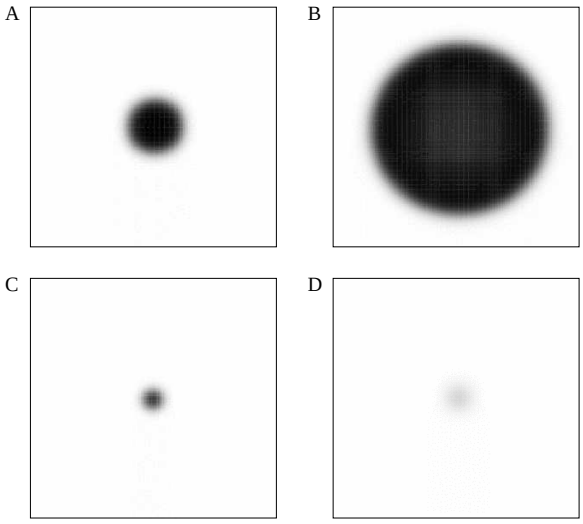

Figure 5.10: Solutions of the FitzHugh-Nagumo equation in the form $\varepsilon v_t = f(v) + \varepsilon^2 \nabla^2 v$, with $f(v) = v(v-a)(1-v)$, showing target patterns (white is $v = 0$; black is $v = 1$). When a large circular wavefront is excited (A), the wave propagates outwards (B). However, when a small circular wavefront is excited (C) the wave front decreases in size (D) due to *curvature blocking*.

5.3.5 Target patterns

An example of a target pattern in the heart is the depolarisation wave which originates from the sino-atrial node (the pacemaker region). Further examples of target patterns occur in monomorphic ventricular tachycardia and Wolff-Parkinson-White disease. A numerical solution of the FitzHugh-Nagumo equations (using a cubic function for $f(v)$ with stable roots at $v = 0$ and $v = 1$) showing a target pattern is shown in figure 5.10. The initial conditions used were $v = 0$ when $r > r_0$ and $v = 1$ when $r < r_0$ (i. e., a circular region of tissue was excited). When the external medium is excitable, then the initial perturbation needs to be large enough to cause propagation, otherwise it dies away: this is shown in the figure, and is associated with the existence of a critical initial patch size.

To understand this, we consider a two-dimensional wave front at $r = R(t)$; then (5.56) implies simply that

$$\dot{R} = c - \frac{1}{R}, \tag{5.70}$$

and thus if $R(0) < \dfrac{1}{c}$, the patch will shrink. If it is bigger than this, the wave will propagate outwards indefinitely, and the wave speed approaches the Cartesian limit as $R \to \infty$. Consequently, an important question to ask is how large must the sino-atrial node be for the depolarisation wave to successfully transmit from the sino-atrial node into the atria. Normally the sino-atrial node is larger than the critical value of the patch radius. However in conditions such as hyperkalaemia, when the excitability

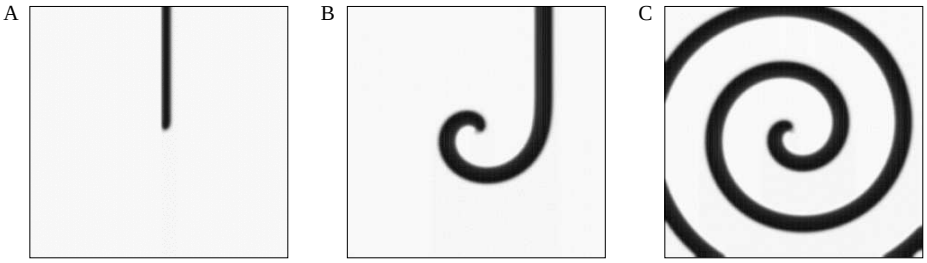

Figure 5.11: Solutions of the FitzHugh-Nagumo equations as in figure 5.10 showing spiral waves (white is $v = 0$; black is $v = 1$). The initial condition (A) is a plane wave in half the domain which wraps around a central core forming a re-entrant spiral wave (C) which continues to re-excite the tissue. In the centre of the domain is a core consisting of a small disk of 'dead tissue' (where $f(v) = 0$) which pins the spiral wave to the centre.

of the tissue is greatly reduced, it is possible for the wave to fail to transmit from the sino-atrial node: this is called sino-atrial block.

5.3.6 Spiral waves

Spiral waves are self-replicating patterns which consist of a rotating spiral and occur in the heart during certain types of polymorphic tachycardia (for example, *Torsades de pointes*). Figure 5.11 shows a solution of the FitzHugh-Nagumo equations exhibiting a spiral wave. The fact that spiral waves rotate means that spiral wave solutions are periodic in time.

Spiral waves are important because they lead to re-entrant behaviour without pacemaking cells. Re-entrant behaviour is when one part of tissue is continually re-excited. Spiral waves are thought to be the cause of polymorphic ventricular tachycardia. Instabilities in spiral waves can cause them to break up and form multiple wavelets. This is a possible explanation for the breakdown of ventricular tachycardia into ventricular fibrillation, which is fatal if not treated immediately.

Can the curved front equation describe spiral waves? This is not as simple to assert as it was for target waves, or in the pacemaker phase model (5.36). If we suppose a two-dimensional wavefront $r = R(\theta, t)$, then (5.56) implies

$$R_t = \frac{c\,(R^2 + R_\theta^2)^{1/2}}{R} - \frac{(R^2 + 2R_\theta^2 - RR_{\theta\theta})}{R\,(R^2 + R_\theta^2)}. \tag{5.71}$$

The target pattern solution $R = \bar{R}(t)$ of (5.70) is clear, but whether spiral wave solutions exist is less obvious. One idea is to seek solutions of the form

$$R = R(\eta), \quad \eta = \omega t - \theta \tag{5.72}$$

(by analogy with (5.41)), and then R satisfies the second order system

$$\omega R' = \frac{c\left(R^2 + R'^2\right)^{1/2}}{R} - \frac{\left(R^2 + 2R'^2 - RR''\right)}{R\left(R^2 + R'^2\right)}. \tag{5.73}$$

The detail of what happens for small R is clearly complicated, but a solution when R is large (which is in fact consistent with the assumption that the wave thickness is small compared to the curvature) is available by expansion of (5.73); the result of this is

$$R \approx \frac{c\eta}{\omega} - \frac{1}{c}\ln\eta + \ldots, \tag{5.74}$$

which represents an Archimedean spiral wave, consistent with what can be seen in figure 5.10.

Chapter 6

The heart as a pump

The rhythmic contraction of the heart described in the preceding chapter causes blood to be expelled into the arterial system. The heart, together with the *arteries* (which carry oxygenated blood to the tissues) and the *veins* (which carry the de-oxygenated blood back to the heart), form a closed system of some five litres in volume, and in order for contraction of the heart to effect a one way pulsatile flow, a system of valves is necessary in order to prevent back flow. In this chapter we describe this system, and also how it is controlled.

6.1 The circulation

A semi-schematic illustration of the human circulation is shown in figure 6.1. Blood is pumped from the heart through the pulmonary capillary bed, where gas exchange in the *alveoli* of the lung (described in chapter **??**) removes metabolically produced carbon dioxide and charges the blood with oxygen. The red, oxygenated blood returns to the heart where it is then pumped to the tissues of the body via the arteries; the blood dumps its oxygen in the tissues, and acquires a load of CO_2, which it takes back to the heart in the veins.

In order to function in this way, the heart really consists of two parts, the right heart (which sends the blood to the lungs) and the left heart (which sends the blood to the tissues). Each side consists of an atrium, where the incoming blood is stored, and a ventricle, where the blood is pumped. Thus the heart contains four chambers: the right and left atria, and the right and left ventricles. The wall between the left and right ventricles is called the septum.

The detailed pathway taken by the blood is shown in figure 6.2; de-oxygenated blood returns to the heart through the vena cava into the right atrium. From there it is sucked into the right ventricle, before being ejected into the pulmonary artery. On returning to the heart via the pulmonary vein, the oxygenated blood flows into the left atrium, from where it is sucked into the left ventricle before being ejected into the aorta under high pressure. The ventricular walls are much thicker than the atrial walls, since they are responsible for creating the majority of the blood pressure.

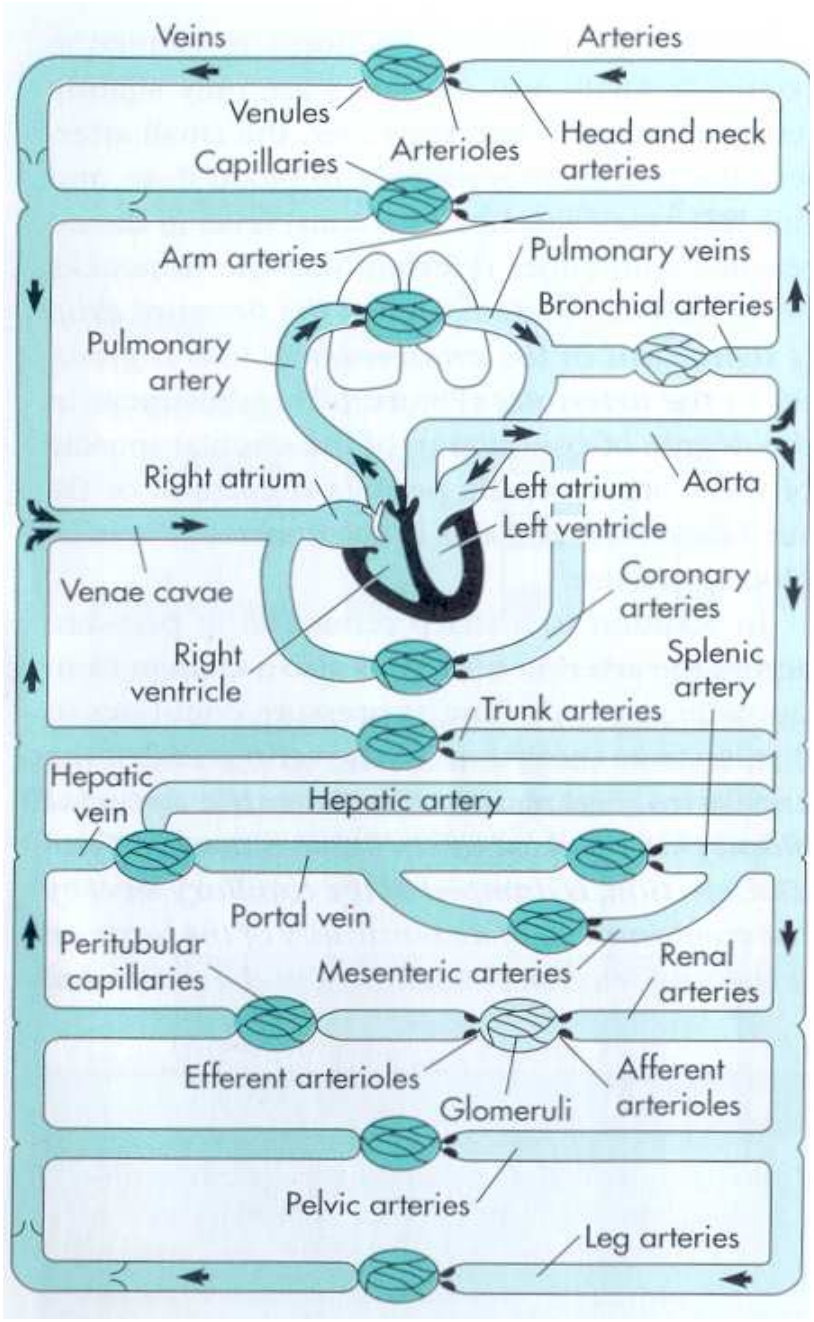

Figure 6.1: The human circulation. Blood flows from the heart to the lungs and back (the pulmonary circulation), before carrying the consequently oxygenated blood through the arteries to the tissues, and then back to the heart via the veins.

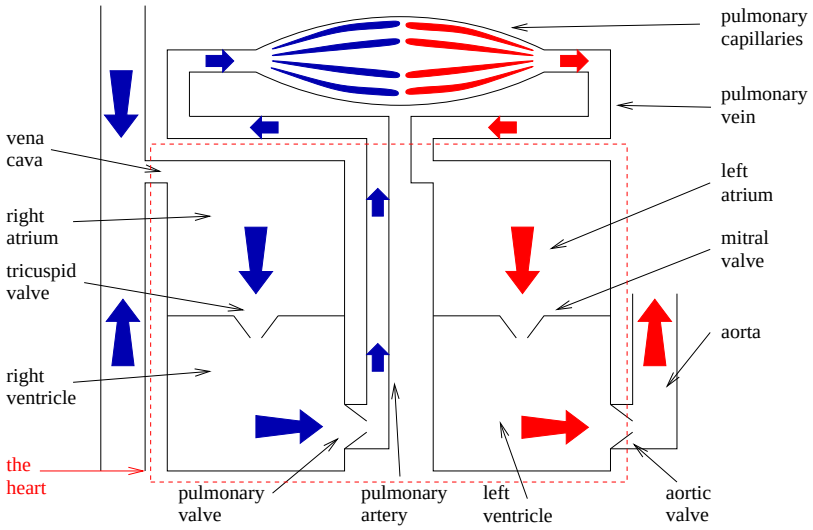

Figure 6.2: A schematic diagram of the chambers and valves of the heart, and the connecting arteries and veins. The direction of the blood flow is shown, the blue arrows represent de-oxygenated blood and the red arrows represent oxygenated blood.

Additionally the left ventricular wall is much thicker than the right ventricular wall because it is responsible for pumping blood around the whole body as opposed to just the lungs.

Fluid flow in the circulation is driven by a pressure gradient which descends from arteries to microcirculation in the tissues to veins, and because the circulation is closed, there must always be regions of contrary pressure gradient. It is precisely in order that a back flow is prevented that the heart contains valves. There are, in fact, four valves which enable the flow to proceed in the manner described above, and these are indicated in figure 6.2. The tricuspid valve prevents back-flow into the right atrium, the pulmonary valve prevents back-flow into the right ventricle, the mitral valve prevents back-flow into the left atrium, and the aortic valve prevents back-flow into the left ventricle.

Figure 6.3 indicates how pressure varies in the circulation. The left ventricular pressure varies from a maximum of about 120 mmHg to a minimum close to zero. The pressure in the arteries also oscillates, but less dramatically, a typical range being from 80 to 120 mmHg (hence the typical blood pressure reference as “120 over 80”). The varying pressure causes waves to propagate down the deformable arteries, but their amplitude is quickly attenuated, and the pressure drop through the capillary microcirculation is essentially constant. Moreover, it is in the capillaries that the bulk of the pressure drop occurs, because of their small diameters. Because of their small volume, one can think of the capillaries as providing (by analogy with an electrical circuit) a resistance to the flow.

The actual sequence of events in the heart itself during a heartbeat is best described with reference to the left ventricular pressure–volume diagram, which repre-

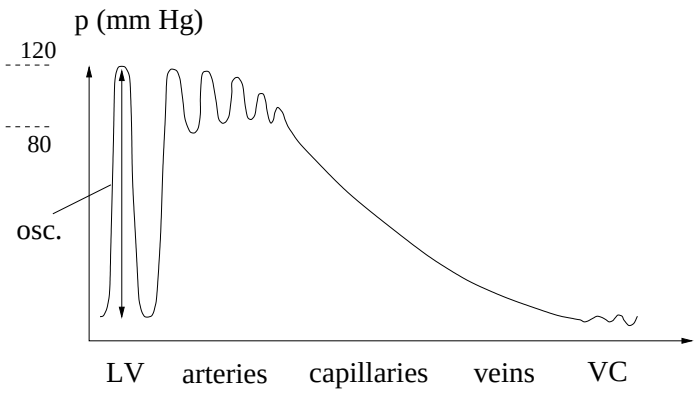

Figure 6.3: Illustration of the variation of pressure with arterial distance from the heart. LV: left ventricle; VC: Vena Cava.

sents how the pressure and volume of the left ventricle vary through the cycle. This is shown in figure 6.4. The left and right hearts act more or less synchronously, so that it suffices to describe the cycle in the left heart.

The heartbeat has two phases, called *systole* and *diastole* (the last 'e' is pronounced in each word). Systole refers to the contraction phase, when the ventricular pressure rises, and diastole refers to the relaxation phase, when ventricular pressure is low. Each of these phases is further subdivided, depending on whether the two left

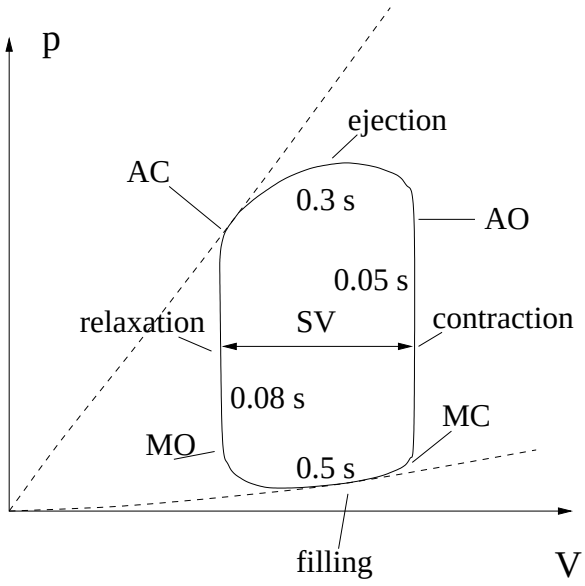

Figure 6.4: Left ventricular pressure volume diagram. AC: aortic valve closes; MO: mitral valve opens; MC: mitral valve closes; AO: aortic valve opens. SV denotes stroke volume.

ventricular valves are open or closed.

The beginning part of systole is the isovolumetric contraction phase (between MC and AO); both valves are shut, so that the volume V_{LV} of the left ventricle remains constant (because the contained blood is incompressible). In this phase the muscles of the ventricular myocardium contract because of the cardiac action potential, and as a consequence the left ventricular pressure p_{LV} increases. This contraction phase is rapid, taking about 0.05 seconds. When the pressure increases beyond the aortic arterial pressure, the aortic valve opens and ejection of the blood into the aorta begins. Contraction continues for a further short period before the ventricular pressure starts to decline as a consequence of the ejection. During this ejection phase, of some 0.3 s duration, the ventricular volume decreases from about 120 ml to about 50 ml; the volume of ejected blood is the *stroke volume*, about 70 ml.

The end of the ejection phase usually occurs at the same time that the action potential drops, and is due to the relaxation of heart muscle as the transient intracellular calcium concentration decreases, and the crossbridges break (see chapter 5). The declining ventricular pressure decreases below the arterial pressure once again, thus closing the aortic valve. There now follows diastole: first an isovolumetric relaxation phase, during which the pressure drops sharply in a time of about 0.08 s, until the ventricular pressure decreases below the left atrial pressure. At this point (MO) the mitral valve opens (the aortic valve is still closed), and the fourth phase of the cycle, the filling phase, begins; it lasts about 0.5 s. In this the ventricle is filled from the atrium, and the ventricular volume increases once more towards its pre-systolic value. The cardiac pacemaker fires, and the cycle begins again.

6.2 A simple one-chamber compartment model

The simplest model of the circulation is a compartment model, in which the veins, arteries, capillary beds, and the chambers of the heart form separate compartments. The model then simply traces the volume changes of the various compartments due to the flow between them. The simplest of such models are those with the fewest compartments, and the most basic one which retains the concept of the heart as a pump is illustrated in figure 6.5. The pulmonary circulation is ignored, and the heart is taken to have a single compartment, the left ventricle. The other compartments are the veins, arteries and a capillary bed, whose primary function is in providing resistance to the flow. The capillary volume is small, and is ignored in the following discussion. In keeping with our assumption of a one chamber heart, there are only two valves, the mitral and aortic valves.

We let V and p denote chamber volumes, and suffixes a, v, and LV refer to arteries, veins and left ventricle, respectively. Blood flow rates to and from the left ventricle are denoted by Q_- and Q_+, respectively, and the blood flow through the capillaries is denoted by Q_c (for both inflow and outflow, since we assume incompressible blood and zero capillary volume). Conservation of blood volume is then expressed by the

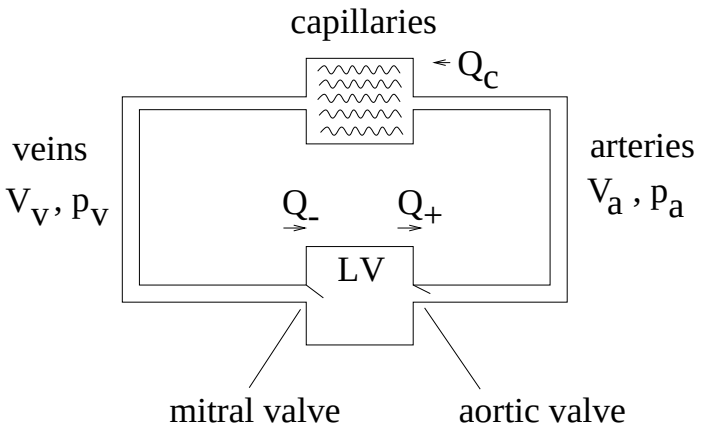

Figure 6.5: Simple compartment model of the circulation.

equations

$$\begin{aligned} \dot{V}_a &= Q_+ - Q_c, \\ \dot{V}_v &= Q_c - Q_-, \\ \dot{V}_{\mathrm{LV}} &= Q_- - Q_+, \end{aligned} \tag{6.1}$$

whence evidently total blood volume is conserved.

The capillary resistance is denoted by R_c, so that the capillary blood flow is

$$Q_c = \frac{p_a - p_v}{R_c}. \tag{6.2}$$

There are also, similarly, resistances associated with the flow to and from the left ventricle. We denote these by R_v and R_a, so that

$$Q_+ = \frac{[p_{\mathrm{LV}} - p_a]_+}{R_a}, \quad Q_- = \frac{[p_v - p_{\mathrm{LV}}]_+}{R_v}, \tag{6.3}$$

and $[x]_+ = \max\{x, 0\}$; this represents the effects of the two valves, which do not allow backflow.

The anatomically 'correct' figure 6.5 is not really consistent with this description, since it portrays arteries and veins as passageways. In order for (6.3) to make sense, we need to interpret p_a and p_v as pressures either side of the capillary bed, but in consideration of the arteries and veins as compartments, they need to be thought of as averages. This blurring is a necessary consequence of the neglect of spatial variation of pressure with distance along the blood vessels.

In order for incompressible blood to circulate, it is necessary that compartment volumes can change, and this is enabled by compliance of the blood vessels. This is the balloon-like property of blood vessel walls, which allows their distension under increased internal pressure. The simplest assumption is one of linear dependence, and

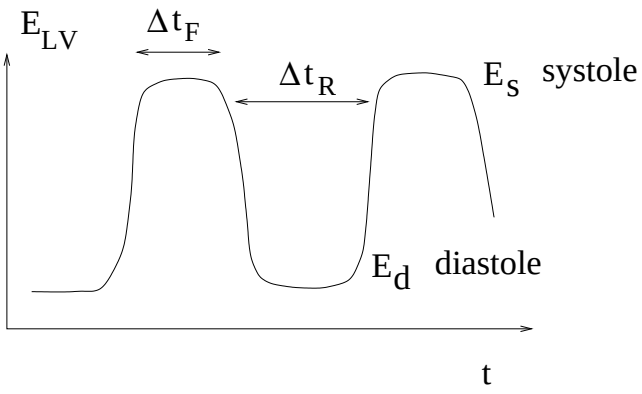

Figure 6.6: Effect of periodic ventricular contraction on elastance.

thus we write

$$\begin{aligned} V_a &= V_{a0} + C_a p_a, \\ V_v &= V_{v0} + C_v p_v, \\ V_{\rm LV} &= V_0 + C_{\rm LV} p_{\rm LV}, \end{aligned} \tag{6.4}$$

and the quantities C_k are called *compliances*. Their inverses ($E_k = 1/C_k$) are called *elastances*, and we will use elastance in discussing left ventricular volume.

This completes the simple mechanical description of the circulation, except that the driving force for the heart beat is not present. As described in chapter 5, the heart beat consists of a ventricular contraction, driven by the passage of a wave of contraction through the atria and ventricles which originates in the sino-atrial node. This periodic firing, and the resulting contraction of heart muscle, causes a stiffening of the ventricles, and this reduces ventricular compliance. Thus the regular firing and consequent contraction of the ventricles causes a periodic change in ventricular elastance, which we suppose to take the form shown in figure 6.6. Essentially the elastance jumps from a very low value E_d to a very high value E_s, and back, with a period of about 0.9 s.

6.2.1 An approximate solution

The model above is simple but nevertheless nonlinear. We are hoping that its solution will mimic the observed pressure-volume cycle shown in figure 6.4. This is shown again in figure 6.7, where also we indicate the way the arterial and venous pressures cycle. The heart valves open and close where these latter two curves touch the ventricular pressure curves. Note that the (aortic) arterial pressure cycles between values of 120 and 80 mmHg, whereas the venous pressure is much lower, around 10 mmHg. Figure 6.8 shows how these pressures vary with time, together with left ventricular volume and the ECG.

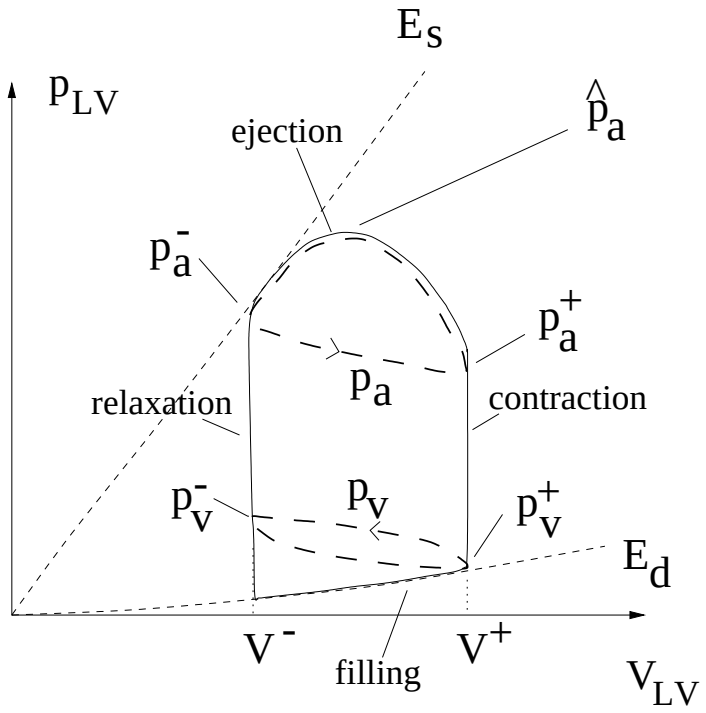

Figure 6.7: Left ventricular pressure-volume curve, showing also arterial and venous pressures. p_a^+ is the end diastolic arterial pressure, V^+ is the corresponding left ventricular volume. The diagram is idealised on the assumption that ventricular elastance shuttles rapidly between two constant values. More realistically, left ventricular pressure follows a curved path during ejection and filling, similar to that shown in figure 6.4.

The equations (6.1)–(6.4) combine to give the model

$$\begin{aligned} R_c C_a \dot{p}_a &= -(p_a - p_v) + \frac{R_c}{R_a}\left[p_{\rm LV} - p_a\right]_+ , \\ R_c C_v \dot{p}_v &= (p_a - p_v) - \frac{R_c}{R_v}\left[p_v - p_{\rm LV}\right]_+ , \\ \left(\frac{\dot{p_{\rm LV}}}{E_{\rm LV}}\right) &= \frac{\left[p_v - p_{\rm LV}\right]_+}{R_v} - \frac{\left[p_{\rm LV} - p_a\right]_+}{R_a}, \end{aligned} \tag{6.5}$$

in which $E_{\rm LV}$ varies between the diastolic value E_d and the systolic value E_s, as shown in figure 6.6. Typical values of the parameters are given in table 6.1.

There are some notable features of the values in table 6.1. Most of the blood volume resides in the venous system, which, with its large compliance, is like a large soggy bag. The venous system is like an air mattress, while the arterial system is like the bicycle pump with which you blow it up (and then the capillary system is the nozzle of the pump). It is because of this disparity that the arterial pressure is so high, while the venous pressure sits round a pressure of about 8 mmHg, and is

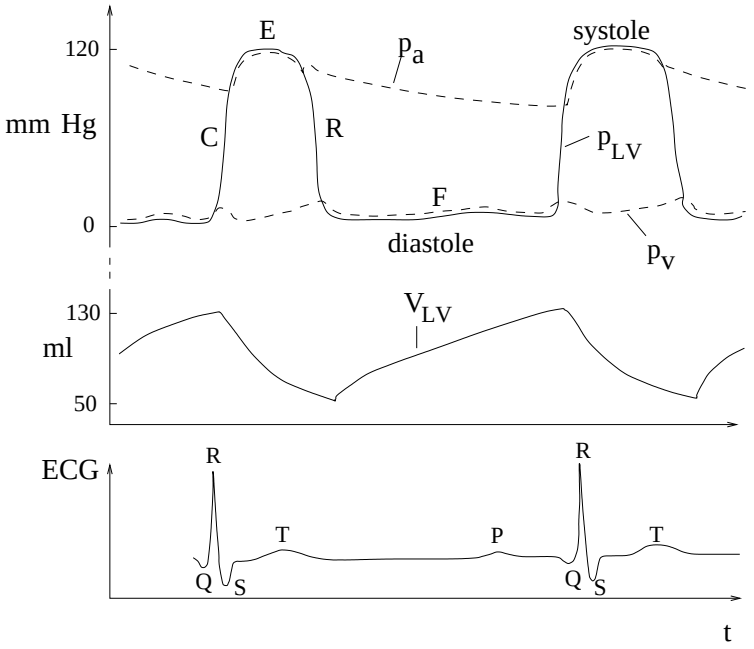

Figure 6.8: Arterial, venous, and left ventricular pressures as a function of time through the heart beat. Also shown is the left ventricular volume, and the ECG. The letters C, R, E and F on the pressure curves stand for contraction, relaxation, ejection and filling, respectively. Wiggles in the pressure curves are associated with valve closure, in particular the *dicrotic notch* in the arterial curve at the end of ejection, when the aortic valve closes.

more or less constant at this level. Another feature is the high value of the capillary resistance compared with those of arteries and veins. It is a consequence of this that the left ventricular pressure is close to the arterial pressure when the aortic valve is open, and to the venous pressure when the mitral valve is open, as shown in figure 6.8. Finally, we notice the extreme change in the left ventricular elastance between diastole and systole. These dramatic variations in the parameters will allow us to derive approximate solutions of the model. We now seek to do this by solving for each of the four phases of the heart beat in turn.

Isovolumetric contraction

We suppose, to begin with (see figure 6.7), that $p_{\rm LV} < p_v < p_a$ at the end of diastole. We suppose that the end diastolic arterial and venous pressures are p_a^+ and p_v^+ respectively, and that contraction has just commenced, with the left ventricular volume being $V^+ = V_0 + C_d {p_{\rm LV}}^+$. Because $p_v \ll p_a$, arterial pressure is approximately determined by

$$\dot{p}_a \approx -\frac{p_a}{R_c C_a}. \tag{6.6}$$

Parameter	Value	Units
V_{a0}	0	ml
V_{v0}	4500	ml
V_0	17	ml
R_a	0.06	mmHg s ml^{-1}
R_v	0.016	mmHg s ml^{-1}
R_c	1.2	mmHg s ml^{-1}
C_a	1.5	ml mmHg^{-1}
C_v	50	ml mmHg^{-1}
E_d	0.06	mmHg ml^{-1}
C_d	16	ml mmHg^{-1}
E_s	3	mmHg ml^{-1}
C_s	0.3	ml mmHg^{-1}
Δt_F	0.3	s
Δt_R	0.6	s

Table 6.1: Values of the constants in (6.5). These are based on values used by Ursino (1998), but modified to account for the simpler description used here (i. e., with fewer compartments).

From table 6.1, $R_cC_a \approx 1.8$ s, while the contraction phase is of duration 0.05 s. Therefore $p_a \approx p_a^+$ during contraction. Similarly, p_v hardly changes, and $p_{\rm LV}/E_{\rm LV}$ remains constant, thus

$$p_{\rm LV} \approx E_{\rm LV}(V^+ - V_0). \tag{6.7}$$

As $E_{\rm LV}$ rises rapidly during contraction, so also does $p_{\rm LV}$, and the isovolumetric contraction phase continues until $p_{\rm LV}$ reaches p_a^+, the aortic valve opens, and the ejection phase begins. This will occur provided $E_s(V^+ - V_0) > p_a^+$ (which is indeed the case).

Ejection (systole)

In the ejection phase, the aortic valve is open and $p_{\rm LV} > p_a > p_v$; outflow Q_+ is positive, but inflow $Q_- = 0$. The governing equations are thus

$$\begin{aligned} R_cC_a\dot{p}_a &= -(p_a - p_v) + \frac{R_c}{R_a}\left[p_{\rm LV} - p_a\right], \\ R_cC_v\dot{p}_v &= (p_a - p_v), \\ R_c\left(\frac{\dot{p}_{\rm LV}}{E_{\rm LV}}\right) &= -\frac{R_c}{R_a}\left[p_{\rm LV} - p_a\right]. \end{aligned} \tag{6.8}$$

Estimated values of the parameters are, from table 6.1, $R_cC_a \approx 1.8$ s, $R_cC_v \approx 60$ s, $R_c/R_a \approx 20$, $R_c/E_s \approx 0.4$ s. Bearing in mind that the ejection phase lasts 0.3

s, $(6.8)_3$ implies that $p_{\rm LV} \approx p_a$ since $R_c/R_a \gg 1$; addition of $(6.8)_1$ and $(6.8)_3$ then implies that

$$R_c C_a \dot{p}_a + R_c \left(\frac{\dot{p}_a}{E_{\rm LV}} \right) \approx -p_a, \tag{6.9}$$

and this continues in the ejection phase until $p_{\rm LV} = p_a$ (exactly) again, which is when

$$\left(\frac{\dot{p}_{\rm LV}}{E_{\rm LV}} \right) \approx \left(\frac{\dot{p}_a}{E_{\rm LV}} \right) = 0. \tag{6.10}$$

Because $R_c C_v$ is large, p_v changes little during the ejection phase, and remains low.

(6.9) can be integrated explicitly to find p_a, given $E_{\rm LV}(t)$. For simplicity, we focus on the case where the transition of $E_{\rm LV}$ from diastolic to systolic values is rapid, as is the transition back. From (6.7), the aortic valve opens when

$$E_{\rm LV} \approx \frac{p_a}{(V^+ - V_0)}, \tag{6.11}$$

which is about 1.25 mmHg ml^{-1}. Supposing that the continuing rise of $E_{\rm LV}$ to its peak value of about $E_s = 3$ mmHg ml^{-1} is rapid, it follows that during this rapid ascent phase,

$$R_c C_a p_a + \frac{R_c p_a}{E_{\rm LV}} = R_c C_a p_a^+ + R_c (V^+ - V_0), \tag{6.12}$$

and thus p_a jumps to the value

$$\hat{p}_a = \frac{C_a p_a^+ + V^+ - V_0}{C_a + C_s}, \tag{6.13}$$

bearing in mind that $C_s = 1/E_s$. This is the peak arterial pressure in systole.

After $E_{\rm LV}$ reaches its peak, (6.9) still applies, but now with $E_{\rm LV} \approx E_s$. Thus

$$p_a \approx \hat{p}_a \exp\left[-\frac{t}{R_c(C_a + C_s)} \right] \tag{6.14}$$

(starting from $t = 0$). The ejection phase finishes when (6.10) occurs, and this is essentially where $E_{\rm LV}$ starts to drop, at the end of the firing sequence, at $t = \Delta t_F$. The arterial pressure at the end of the ejection phase is thus

$$p_a \approx p_a^- = \hat{p}_a \exp\left[-\frac{\Delta t_F}{R_c(C_a + C_s)} \right], \tag{6.15}$$

and this is the end systolic arterial pressure, and Δt_F is the firing time. At this point the left ventricular volume is

$$V_{\rm LV} = V^- \approx V_0 + C_s p_a^-, \tag{6.16}$$

and thus the stroke volume $\Delta V = V^+ - V^-$ is, using (6.13) and (6.15),

$$\Delta V = V^+ - V_0 - \left(\frac{C_s}{C_a + C_s} \right) \left[C_a p_a^+ + V^+ - V_0 \right] \exp\left[-\frac{\Delta t_F}{R_c(C_a + C_s)} \right]. \tag{6.17}$$

Isovolumetric relaxation

At the end of the ejection phase, left ventricular elastance drops very rapidly, and the aortic valve closes. With both valves now closed, $p_v < p_{\rm LV} < p_a$, inflow Q_- and outflow Q_+ to and from the left ventricle are both zero, and left ventricular volume V_- remians constant. As for the contraction phase, both p_a and p_v are virtually constant, and left ventricular pressure is given by

$$p_{\rm LV} \approx E_{\rm LV}(V^- - V_0). \tag{6.18}$$

The mitral valve opens and filling commences when $p_{\rm LV} = p_v$, and this is when

$$E_{\rm LV} \approx \frac{p_v}{(V^- - V_0)}, \tag{6.19}$$

and this is about 0.25 mmHg ml^{-1}.

Filling (diastole)

During the filling phase, of about 0.5 s duration, $p_{\rm LV} < p_v < p_a$, $Q_+ = 0$, and the governing equations reduce to

$$\begin{aligned} R_c C_a \dot{p}_a &= -(p_a - p_v), \\ R_v C_v \dot{p}_v &= \frac{R_v}{R_c}(p_a - p_v) - (p_v - p_{\rm LV}), \\ R_v \left(\frac{\dot{p}_{\rm LV}}{E_{\rm LV}} \right) &= p_v - p_{\rm LV}. \end{aligned} \tag{6.20}$$

Relevant constants are $R_c C_a \approx 1.8$ s, $R_v C_v \approx 0.8$ s, $R_v/R_c \approx 0.013$, and $R_v/E_d \approx 0.27$ s. As for the ejection phase, $E_{\rm LV}$ continues to fall rapidly to E_d, and during this short period

$$p_{\rm LV} \approx E_{\rm LV}(V^- - V_0), \tag{6.21}$$

while p_v and p_a hardly change. Thus the left ventricular pressure reaches the value

$$p_{\rm LV} \approx E_d(V^- - V_0). \tag{6.22}$$

Following this, we have $p_{\rm LV} < p_v \ll p_a$, so that p_a is approximately

$$p_a \approx p_a^- \exp\left[-\frac{t}{R_c C_a} \right] \tag{6.23}$$

(measuring t from the beginning of filling). The second and third equations in (6.20) then give a pair of linear equations for p_v and $p_{\rm LV}$.

From (6.23), the subsequent value of end diastolic arterial pressure is

$$p_a^{++} = p_a^- \exp\left[-\frac{\Delta t_R}{R_c C_a} \right], \tag{6.24}$$

where Δt_R is the refractory period. This provides a map from the previous value p_a^+ to p_a^{++}. Conjoining (6.13), (6.15) and (6.24), we have

$$p_a^{++} = \left(\frac{C_a p_a^+ + C_d p_v^+}{C_a + C_s} \right) \exp \left[- \left\{ \frac{\Delta t_F}{R_c(C_a + C_s)} + \frac{\Delta t_R}{R_c C_a} \right\} \right], \tag{6.25}$$

where p_v^+ is the end diastolic venous pressure (initially, where $p_a = p_a^+$), and we have supposed that end diastole occurs when $p_{\rm LV} = p_v$, and that $V^+ = V_0 + C_d p_v^+$. More generally, the ventricular pressure will be less than p_v. If venous pressure is known, this provides a linear first order difference equation for p_a^+, which has a stable fixed point. From (6.17), stroke volume is given in terms of venous pressure by

$$\Delta V = C_d p_v^+ - \left(\frac{C_s}{C_a + C_s} \right) \left[C_a p_a^+ + C_d p_v^+ \right] \exp \left[- \frac{\Delta t_F}{R_c(C_a + C_s)} \right]. \tag{6.26}$$

Venous pressure

Although venous pressure is small and does not vary much, it is important in the determination of arterial pressure and stroke volume through the formulae (6.25) and (6.26), because the diastolic compliance is so high. In fact, $C_d p_v$ is of the same size as $C_a p_a$. Therefore we need to retrace our steps to calculate the corresponding map, $p_v^+ \to p_v^{++}$.

To begin with, the change of p_v in the contraction phase is very small, and is ignored. In the ejection phase, after $E_{\rm LV}$ reaches E_s, p_v is given by $(6.8)_2$, thus, approximately,

$$R_c C_v \dot{p}_v = p_a, \tag{6.27}$$

where p_a is given by (6.14), and thus

$$C_v(p_v - p_v^+) = (C_a + C_s)(\hat{p}_a - p_a). \tag{6.28}$$

At the end of systole, $p_v = p_v^-$, where from (6.15),

$$C_v(p_v^- - p_v^+) = (C_a + C_s)(\hat{p}_a - p_a^-). \tag{6.29}$$

In the relaxation phase, p_v is unchanged; then in the filling phase, $(6.20)_2$ implies (since $p_a \gg p_v$) that p_v and $p_{\rm LV}$ satisfy

$$\begin{aligned} R_v C_v \dot{p}_v &\approx \frac{R_v p_a}{R_c} - (p_v - p_{\rm LV}), \\ R_v C_d \dot{p}_{\rm LV} &= p_v - p_{\rm LV}. \end{aligned} \tag{6.30}$$

The two time constants are $R_v C_v \approx 0.8$ s, and $R_v C_d \approx 0.27$ s, both comparable to the filling time of about 0.5 s. Adding the two, and using (6.23), we get

$$C_d \dot{p}_{\rm LV} + C_v \dot{p}_v + C_a \dot{p}_a = 0 \tag{6.31}$$

(as we should!), and thus the values at beginning and end diastole are related by

$$(C_d + C_v)p_v^{++} + C_a p_a^{++} = C_v p_v^{-} + (C_a + C_s)p_a^{-}, \tag{6.32}$$

where we use (6.16) and (6.22) to determine the initial condition for p_{LV}, and use the fact that $p_{\mathrm{LV}} = p_v$ at end diastole. (6.32) provides the other part of the map we seek. More generally, one also needs to consider the variation of p_{LV}.

6.3 Nervous control of the heart

Cardiac output (blood flow as volume per unit time) is equal to the stroke volume times the heart rate; it is therefore the control of these quantities which controls the blood flow to the body tissues. Since blood is the agent of nutrient (oxygen) supply, it is evident that this needs to be tightly controlled. In the previous section we saw that the pumping mechanism of the heart is itself an agent for control of stroke volume and arterial pressure. Heart rate is controlled by the period of the oscillatory sino-atrial node firing, and this (as well as other parts of the circulatory system) is mainly controlled by the actions of the autonomic nervous system.

There are two parts to the nervous system which control cardiac output. These are the *sympathetic* and *parasympathetic* systems. Each in turn consists of *afferent* nerves (*ad + fero*, I carry to) taking messages to the brain, and *efferent* nerves (*ex + fero*, I carry from) taking messages from the brainstem or spinal cord. The afferent nerves transmit information from various kinds of receptors (*baroreceptors*, *chemoreceptors*); the efferent nerves innervate different parts of the system. One speaks of the tonic activity of these systems, meaning the rate of firing of the nerve fibres. Muscle tone refers to the same notion.

The sympathetic nervous system has two sub-systems, the α-sympathetic system, which innervates the peripheral vasculature, and the β-sympathetic system, which innervates ventricular muscle. The sympathetic nerves act by releasing *noradrenaline*[1] and other neurotransmitters. In the peripheral circulation, these cause *vasoconstriction*, i. e., they constrict blood vessels, and thus increase peripheral resistance. This has the effect of decreasing cardiac output.

The release of noradrenaline and also adrenaline (together, these are called the *catecholamines*) by the β-sympathetic system in ventricular muscle has two effects. The *chronotropic* effect is that on the firing rate of the sino-atrial node. The effect of the release of catecholamines is to increase the inward calcium current to the *myocytes*, which in turn increases the firing frequency of the SA node. Thus the sympathetic system increases heart rate.

The other effect is the *inotropic* effect. Increasing sympathetic tone increases the contractility, or elastance, of ventricular muscle (by increasing the intracellular calcium transients). This then has the effect of increasing stroke volume.

[1]Noradrenaline is called *norepinephrine* in the American literature; similarly, *adrenaline* is called *epinephrine.*

In summary, the sympathetic system acts, via the release of catecholamines, to increase cardiac output. The system acts *slowly* over a time scale of the order of ten seconds, because this is the time scale for uptake of the catecholamines, and their effect on intercellular calcium.

In contrast, the parasympathetic system acts to decrease cardiac flow. It innervates the heart, particularly the sino-atrial node and the atrio-ventricular node, through the left and right *vagus* nerves. The vagus nerves act by releasing another transmitter, *acetylcholine*, which has an immediate effect on heart rate, causing it to decrease (*bradycardia*) by altering the characteristics of the pacemaker firing oscillation; acetylcholine is an inhibitor for the pacemaking currents $i_{Ca,L}$ and i_f. The parasympathetic system also innervates peripheral blood vessels, having a *vasodilative* effect, in contrast to the vasoconstrictive effect of the sympathetic system.

Baroreceptor reflex

Sympathetic and parasympathetic tone is determined by signals transmitted along afferent nerves from various kinds of receptors. For the control of blood flow, the most important of these are the *baroreceptors*, located in the aortic arch in the chest, and in the *carotid sinus* in the neck. As their name indicates, baroreceptors respond to arterial pressure (via its effect of stretching the arterial wall), and the control effected through the feedback via the baroreceptors is called the *baroreflex*. Figure 6.9 shows the multiple feedback control loops which the above description embodies.

6.4 Oscillatory patterns

The heart rate oscillates on a variety of time scales, conventionally separated into three different magnitudes, each ascribed to a different control system.

The first of these is *respiratory sinus arrhythmia* (RSA), which is an oscillation of frequency 0.2–0.4 Hz, and is due to the coupling between respiration and heart rate. Specifically, heart rate increases on inspiration, and decreases during expiration. The simplest cause for this is that during inspiration, the intrathoracic pressure is low, and this correspondingly causes the filling (venous) pressure to be low. From (6.25), this causes arterial pressure to drop, and hence the vagal feedback proposed in figure 6.9 leads to an immediate rise in heart rate.

The second oscillation, with a period of about ten seconds (frequency 0.1 Hz) goes by the name of *Mayer waves*; because the time scale is comparable to the response time of the sympathetic system, it is generally thought that Mayer waves are due to the sympathetic system, although (as we shall see) there are different ways in which their occurrence can be modelled.

The third frequency, $\lesssim 0.1$ Hz, is associated with long term thermo-regulatory control, and is not discussed further.

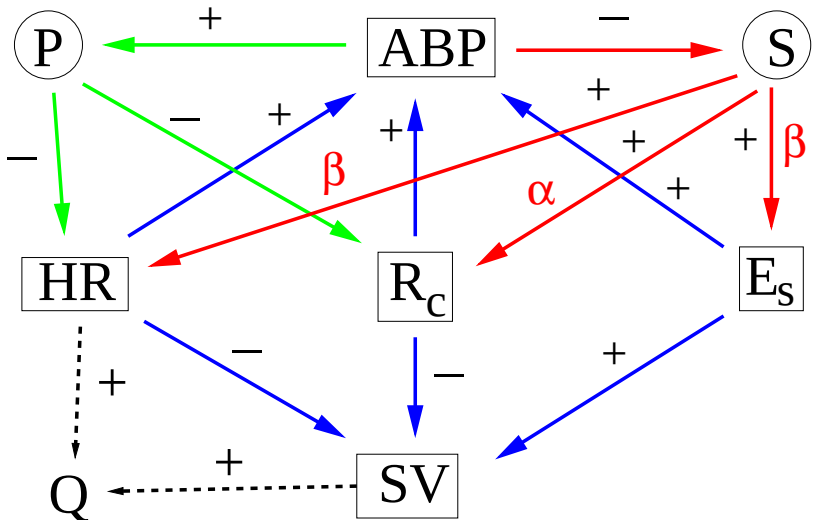

Figure 6.9: A schematic representation of the autonomic and nervous control systems of the heart. P and S represent parasympathetic and sympathetic systems, ABP is arterial blood pressure, HR is heart rate, R_c is peripheral resistance, E_s is systolic ventricular elastance, SV is stroke volume; heart rate and stroke volume combine to give cardiac output Q. Plus and minus signs indicate activating or inhibiting effect; blue arrows represent autonomic (pump action) control, red is the sympathetic system, green is the parasympathetic system.

6.5 Mathematical models of the baroreflex

The fundamental conceptual difficulty which arises in modelling the control of heart rate is that the basic pumping mechanism itself is not smooth, because of valve opening and closure. As we have seen above, this naturally leads to a pump action model which gives a map: in our case, of the values of end diastolic arterial and venous pressures from one cycle to the next. Vagal control fits naturally into this scheme, because the response is virtually instantaneous, and this leads to a beat-to-beat model, exemplified by the De Boer model which is discussed below in section 6.5.2. Such a model would hope to explain RSA, for example.

It is less easy to put sympathetic control into a beat-to-beat model, because it is effected continuously by the uptake of catecholamines released at nerve endings. The best one can do is to represent its effect by a distributed delayed effect over a number of heart beats, and this is what the De Boer model does.

The natural alternative for longer term sympathetic control is to suppose that the arterial and venous blood pressures and other quantities of the system vary slowly under the influence of the slowly acting sympathetic system, so that their evolution can be modelled by differential equations, and this is what the Ottesen model, described next, does. Ideally, the resulting continuous model is derived from a more realistic beat-to-beat model by a formal process of averaging, but in practice this

procedure is short circuited, and one writes down the continuous model directly from first principles, on the basis that pump action is continuous. modelling respiratory control (see chapter **??**).

6.5.1 Ottesen model

This model is described by Ottesen (1997). There are three variables, the (averaged) arterial and venous pressures p_a and p_v, and the heart rate H. Control is effected by sympathetic and parasympathetic tones (i. e., firing frequencies), and these are assumed to be given by

$$T_s = g(p_a^\tau), \quad T_p = 1 - g(p_a), \tag{6.33}$$

where $p_a^\tau = p_a(t - \tau)$ is the arterial pressure a time τ in the past. The choice of the controlling function g is the sigmoidal Hill function given by

$$g(p) = \frac{1}{1 + (p/p_0)^n}, \tag{6.34}$$

and the choice of the Hill exponent n is quite high (for example $n = 7$) in order that control be effected sharply.

The representation of sympathetic control in terms of delayed arterial pressure is a simple surrogate to represent the slow response to release of catecholamines by the sympathetic nerves. A more realistic model would have a distributed delay, and ultimately one would want to incorporate details of the sino-atrial node cell firing oscillation in terms of intracellular calcium, potassium and sodium concentrations.

The essence of the model lies in the way in which H responds to T_s and T_p. Ottesen chooses $\dot{H} = F(T_s, T_p)$, and specifically (by way of example)

$$\dot{H} = \frac{\lambda_H T_s}{1 + \gamma T_p} - \mu_H T_p, \tag{6.35}$$

which represents the effect of sympathetic and parasympathetic tone on rate of change of heart rate, and also includes the inhibiting effect of the vagal response on the sympathetic response, through the coefficient γ.

There is an odd feature of this choice. In reality, if one cuts (or deactivates) both the sympathetic and parasympathetic systems in humans, then the heart rate settles at a steady hundred beats per minute. In the model, this correponds to putting both T_s and T_p to zero, but we see that there is then no preferred heart rate, and it is neutrally stable. This is because there is no damping term in the equation for H, which is something in a real physical system that one might expect. It is suggestive of structural instability in the model and, as we shall see, it can lead to unphysical behaviour.

We therefore propose a modified version of the Ottesen model, which is

$$\dot{H} = \delta_H(H_0 - H) + \frac{\lambda_H T_s}{1 + \gamma T_p} - \mu_H T_p, \tag{6.36}$$

where H_0 denotes the natural resting heart rate in the absence of nervous tone. A motivation for this choice can be found by consideration of what actually determines heart rate. This was discussed in chapter 5. The heart rate is the inverse of the period P of a limit cycle oscillation involving intracellular concentrations of calcium, potassium and sodium in the sino-atrial node cells. The essence of any such model is captured by the Landau-Ginzburg equation

$$\frac{dz}{dt} = az - b|z|^2 z, \tag{6.37}$$

where a and b are complex, and a_R $(= \mathrm{Re}\, a) > 0$, $b_R > 0$.[2] If we define the period of the evolving oscillation as the interval P between values of zero phase ($\arg z = 0 \mod 2\pi$), then it is straightforward to show that $P = P(A)$, where A is the amplitude of z, and thus that $H = 1/P$ satisfies an evolution equation of the form

$$\dot{H} = r(H)\,[H_0 - H]\,. \tag{6.38}$$

This provides the motivation for the form of (6.36) in the absence of nervous tone.

The model for heart rate is supplemented by the two blood pressure equations

$$\begin{aligned} C_a \dot{p}_a &= -\frac{(p_a - p_v)}{R_c} + H\Delta V, \\ C_v \dot{p}_v &= \frac{p_a - p_v}{R_c} - \frac{p_v}{R_v}, \end{aligned} \tag{6.39}$$

which can be compared directly to the first two equations in (6.5); the compliances C_k and resistances R_k carry the same meaning as before, as does the stroke volume ΔV.

The first equation is the continuous version of conservation of arterial blood volume, since $Q = H\Delta V$ is the cardiac output. The second equation represents conservation of venous blood volume, if ventricular (or, more properly, atrial) pressure is ignored in (6.20) during filling.

The model can be simplified in much the same way as the pump action model, by observing that a balance of terms on the right hand side of (6.39) suggests $p_a - p_v \sim R_c Q$ and $p_v \sim R_v Q$, and thus that $p_v/p_a \sim R_v/R_c \sim 10^{-2}$, if we use the values for resistance in table 6.1. (Since $1/R_v$ is an average conductance over the heart beat, the value we use for it in the continuous model should be reduced, and thus the value for R_v should be increased in (6.39); a mild increase is still consistent with the observation that $p_v \ll p_a$. In fact, Ottesen uses a value four times higher for R_v, but a similar value for R_c.)

Allowing that $p_v \ll p_a$ enables $(6.39)_1$ to be approximated as

$$R_c C_a \dot{p}_a = -p_a + R_c \Delta V\, H. \tag{6.40}$$

[2]This equation universally describes the amplitude of periodic solutions in the vicinity of a Hopf bifurcation.

We note from table 6.1 that we have typical values $R_cC_a \sim 1.8$ s, and for cardiac output $Q = 5$ l min$^{-1} \sim 80$ ml s^{-1}, $R_cQ \sim 100$ mmHg.

There are three time constants in (6.36): δ_H, λ_H/H_0 and μ_H/H_0. We choose $H_0 = 100$ min$^{-1} = 1.7$ s^{-1}, as the resting heart rate in the absence of nervous control. The choice of δ_H relies on a detailed model of the sino-atrial cell firing oscillation. In the absence of any other information, the natural choice for δ_H is the time constant of the oscillator, which is thus simply $\delta_H = H_0$. Ottesen's preferred values for λ_H and μ_H are $\lambda_H = 0.84$ s^{-2} and $\mu_H = 1.17$ s^{-2}, and we will use these for illustration. We then have

$$\lambda = \frac{\lambda_H}{H_0^2} \sim 0.3, \quad \mu = \frac{\mu_H}{H_0^2} \sim 0.4, \quad \delta = \frac{\delta_H}{H_0} \sim 1. \tag{6.41}$$

He also chooses to put $\gamma = 0$, and we will do this also. In this simple model, the stroke volume is taken as constant. More generally, we would take ΔV as a function of arterial pressure.

We now non-dimensionalise the approximate (i. e., neglecting p_v) model by writing

$$H = H_0 h, \quad p_a = p_0 p, \quad t \sim \tau \tag{6.42}$$

(p_0 is defined in (6.34)); the corresponding dimensionless model is then

$$\begin{aligned} \dot{p} &= \kappa[-p + \nu h], \\ \varepsilon\dot{h} &= \delta(1-h) + \frac{\lambda g(p_1)}{1+\gamma\{1-g(p)\}} - \mu\{1-g(p)\}, \end{aligned} \tag{6.43}$$

where now

$$g(p) = \frac{1}{1+p^n}, \tag{6.44}$$

and the notation p_1 denotes the delayed arterial pressure $p(t-1)$. The additional parameters in (6.43) are given by

$$\kappa = \frac{\tau}{R_cC_a}, \quad \varepsilon = \frac{1}{H_0\tau}, \quad \nu = \frac{R_c\Delta V\, H_0}{p_0}, \tag{6.45}$$

and we can take typical values $\kappa \sim 5.6$, $\varepsilon \sim 0.06$, $\nu \sim 1$ (the latter because we can undoubtedly assume that the nervous controls are effective at a typical value of arterial pressure).

We now take advantage of the fact that ε is relatively small. This suggests that h rapidly approaches a quasi-equilibrium state. If we take $\gamma = 0$ and $\delta = 1$, then this quasi-steady state is

$$h \approx 1 + \lambda g(p_1) - \mu\{1-g(p)\}, \tag{6.46}$$

and substituting this into $(6.43)_1$ yields the delay recruitment type equation

$$\dot{p} = \kappa[\nu(1-\mu) - p + \nu\{\lambda g(p_1) + \mu g(p)\}]. \tag{6.47}$$

The steady state solution $p = p^*$ satisfies

$$g(p) = \frac{p - \nu(1-\mu)}{\nu(\lambda+\mu)}, \tag{6.48}$$

and is uniquely defined.

Stability of the steady state is ascertained by linearising the model about it. We write

$$p = p^* + P, \tag{6.49}$$

so that the linear equation for P is

$$\dot{P} = \kappa[-P - \nu s(\lambda P_1 + \mu P)], \quad s = -g'(p^*) > 0. \tag{6.50}$$

Solutions of this equation proportional to $\exp(\sigma t)$ exist provided

$$\sigma = -B - Ge^{-\sigma}, \tag{6.51}$$

where

$$B = \kappa(1 + \nu s\mu), \quad G = \kappa\nu s\lambda. \tag{6.52}$$

Instability then occurs if $\operatorname{Re}\sigma > 0$, and the instability is oscillatory if $\operatorname{Im}\sigma \neq 0$.

This is an equation which we will meet again and again in the following chapters. (6.51) is a transcendental equation, which has an infinite number of complex roots, no more than two of which are real. The roots accumulate at the essential singularity at $\sigma = \infty$, and $\operatorname{Re}\sigma \to -\infty$ as $\sigma \to \infty$, thus the set of $\operatorname{Re}\sigma$ is bounded above. There is an instability criterion which determines when *all* the roots σ have negative real part, and this is indicated in figure 6.10.

Evidently both G and B here are positive, and so the steady state is unstable if $G > \gamma_1(B)$ in the figure. To estimate G and B in the present case, we suppose that $p^* = 1.3$, $\nu = 1.85$, $\mu = 0.4$, $\lambda = 0.29$. Then (6.48) implies $g^* = g(p^*) = 0.149$, and to be consistent with the definition of g, this requires $n = 6.65$. Using this value together with the fact that $s = \dfrac{ng^*(1-g^*)}{p^*}$ suggests $s \approx 0.65$, and thus $B \approx 8.3$, $G \approx 1.95$. For large B, one can show that $\gamma_1(B) \approx B + \dfrac{\pi^2}{2B}$, so the Ottesen model is stable for these parameter values. Since for large B, instability requires $G \gtrsim B$, we see that instability would require $\lambda > \mu$ (and then a sufficiently large value of s). Note that the dimensionless period of the resulting bifurcating periodic solution is approximately 2 for large B; the dimensional period is simply $\approx 2\tau$. This suggests that if this model is to be used to predict 10 second Mayer waves, then the delay in the sympathetic response needs to be chosen to be $\tau \approx 5$ s.

The limit $\delta \to 0$

It is evident from (6.43) that if $\delta \neq 1$, one simply replaces ε, λ and μ by ε/δ, λ/δ and μ/δ. As long as $\varepsilon \ll \delta$, h still rapidly approaches equilibrium, and the subsequent

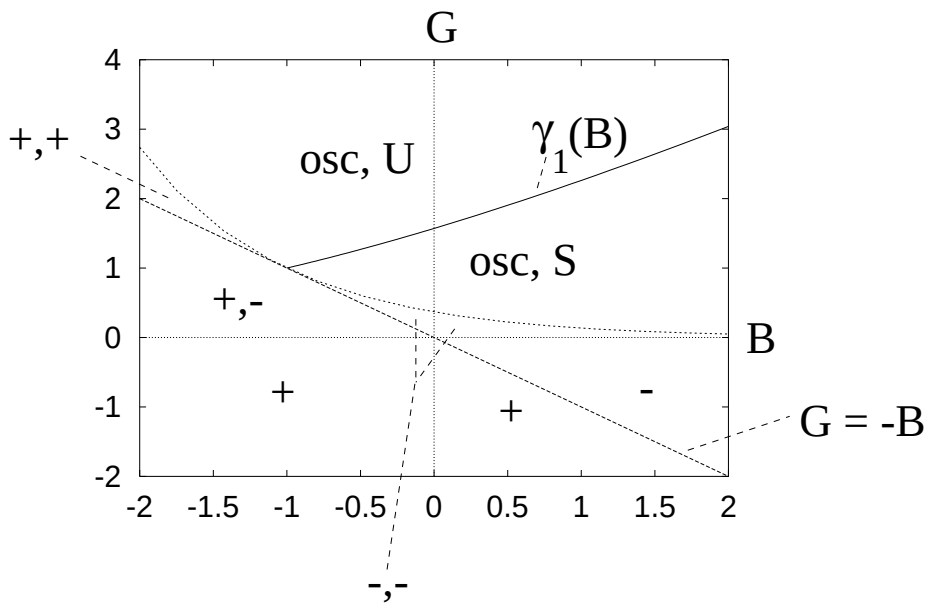

Figure 6.10: Stability map for the solutions of (6.51). The curves demarcate the behaviours of the two roots involved in transition to instability. 'osc' means complex conjugate, and 'U' is unstable, 'S' is stable; '+' and '-' refer to the signs of real roots; where only one sign is indicated, the other root has disappeared by tending to infinity. The diagram indicates that stability occurs only if G lies within the sector bounded by the two curves $G = -B$ and $G = \gamma_1(B)$.

discussion above of steady states and stability is unaffected, except that one uses λ/δ and μ/δ in (6.48) and (6.52).

On the other hand, suppose that $\delta \lesssim \varepsilon$, and let us in fact take $\delta = 0$. The equations (6.43) are thus (also with $\gamma = 0$)

$$\begin{aligned} \dot{p} &= \kappa[-p + \nu h], \\ \varepsilon\dot{h} &= \lambda g(p_1) - \mu\{1 - g(p)\}. \end{aligned} \tag{6.53}$$

This is a decidedly odd-looking pair, since h no longer appears on the right hand side of its own equation. Apparently, for small ε, the h equation defines an approximate map $p_1 \to p$, which has a unique fixed point, but experience with, for example, the delay-recruitment equation $\varepsilon\dot{x} = -x + f(x_1)$ when ε is small suggests the possibility of boundary layer behaviour, where p jumps rapidly from one iterate to the next (see Fowler (1997), p. 360); however, that is not possible here, since on the face of it, p cannot change rapidly.

The apparent resolution is to rescale the variables as

$$h = H\sqrt{\frac{\mu}{\nu\kappa\varepsilon}}, \quad t \sim \sqrt{\frac{\varepsilon}{\nu\kappa\mu}}, \tag{6.54}$$

which leads to the pair

$$\begin{aligned} \dot{p} &= H - \beta p, \\ \dot{H} &= \zeta g(p_T) - \{1 - g(p)\}, \end{aligned} \tag{6.55}$$

where

$$T = \sqrt{\frac{\nu\kappa\mu}{\varepsilon}}, \quad \beta = \sqrt{\frac{\varepsilon\kappa}{\nu\mu}}, \quad \zeta = \frac{\lambda}{\mu}. \tag{6.56}$$

For small ε, β is small and T is large; for our values of $\varepsilon = 0.06$, $\kappa = 5.6$, $\nu = 1.85$, $\mu = 0.4$, we have $T \approx 8.31$, $\zeta \approx 0.73$, $\beta \approx 0.67$, so one might keep $\beta \sim O(1)$. The fact that the delay T is large suggests that the steady state will be unstable.

To assess stability, it is simplest to write (6.56) as the single second order equation

$$\ddot{p} + \beta\dot{p} + \{1 - g(p)\} - \zeta g(p_T). \tag{6.57}$$

Linearising about the steady state and seeking solutions $e^{\sigma t}$, we obtain

$$\sigma^2 + \beta\sigma + s\left(1 + \zeta e^{-\sigma T}\right) = 0. \tag{6.58}$$

For small s, all the roots have $\mathrm{Re}\,\sigma < 0$, but instability will occur for sufficiently large s. Formally, as $\varepsilon \to 0$ and thus also $\beta \to 0$, the steady state will always be unstable.

6.5.2 De Boer model

The model of heart rate due to De Boer, Karemaker and Strackee (1987) is a beat-to-beat model, which relates successive values of peak systolic pressure P and end diastolic pressure D. In terms of the notation we used earlier, $p_a^+ = D$, $\hat{p}_a = S$. While the Ottesen model focuses on the control of heart rate by the nervous systems, the De Boer model also allows for its effect on peripheral resistance R_c. This is done through an equation for the time constant T, which in our notation is given by $T = C_a R_c$. Finally heart rate is included by having an equation for the beat-to-beat interval I, conventionally measured as the RR interval (see figure 6.8)

In the model, the effect of peak systolic arterial pressure on baroreceptor response is represented by a function $F(S)$ with units of pressure, which is taken as a sigmoidal function increasing from about 90 mmHg to about 150 mmHg as S increases through normal values of 120 mmHg. Thus F is essentially a shifted version of the Hill function, $F = A - Bg$, and in some way it represents the firing rate of the efferent nerves. In terms of this firing rate, the RR interval is taken to be

$$I_n = I^* + \sum_{k \geq 0} a_k F_{n-k}, \tag{6.59}$$

where n indexes the sequence of heart beats. This is an analogue of (6.36). Both sympathetic and parasympathetic systems decrease heart rate (thus increase RR interval) on increasing arterial blood pressure, and both effects are indicated by (6.59) if

the coefficients a_k are positive. The index $k = 0$ then represents the vagal effect, and $k > 0$ represents the delayed sympathetic response. The continuum limit of (6.59) would be

$$I(t) = I^* + A_V p_a(t) + \int_0^\infty A(s) f\left[p_a(t-s)\right]\, ds, \tag{6.60}$$

with some suitable redefinition of a_k and F, and this would be consistent with $(6.43)_2$ if there were a distributed delay in the Ottesen model. De Boer *et al.* choose values of a_k for $k > 0$ distributed round a maximum at $k = 4$. The corresponding discrete delay in the Ottesen model would then be $4/H$, about 3.5 seconds. The value a_0 is manifested as the vagal coefficient A_V.

Consulting figure 6.9, we see that the blood pressure has an inhibitory effect on the peripheral resistance due to the sympathetic system, and thus on the time constant $T = C_a R_c$. In the De Boer model, this is effected through the equation

$$T_n = T^* - \sum_{k>0} b_k F_{n-k}, \tag{6.61}$$

and the b_k's are taken to be a multiple (twice) of the a_k's. There is no correponding effect in the Ottesen model, though it would simply be included by suitable functional dependence of R_c in (6.39). The continuum limit of (6.61) would be

$$T(t) = T^* - \int_0^\infty B(s) f\left[p_a(t-s)\right]\, ds. \tag{6.62}$$

The model is completed by two equations for S and D which describe the pump action of the heart. The equations (6.39) play the same rôle in the Ottesen model, although the De Boer model omits reference to the venous or filling pressure. The first equation is the *Windkessel* model,

$$D_n = cS_{n-1} \exp\left[-\frac{I_{n-1}}{T_{n-1}}\right]. \tag{6.63}$$

This equation comes directly from consideration of pump action, and is written, in the notation of section 6.2, as

$$p_a^{++} = c\hat{p}_a \exp\left[-\frac{\Delta t_F + \Delta t_R}{R_c C_a}\right], \tag{6.64}$$

whereas from (6.15) and (6.24) we have, in fact,

$$p_a^{++} = \hat{p}_a \exp\left[-\frac{\Delta t_F}{R_c(C_a + C_s)} - \frac{\Delta t_R}{R_c C_a}\right]. \tag{6.65}$$

The Windkessel model thus follows from our simple pump model providing $C_s \ll C_a$, which is not unreasonable according to table 6.1.

The second equation relates pulse pressure $S - D$ to the length of the preceding RR interval, thus

$$S_n = D_n + \gamma I_{n-1} + P^*. \tag{6.66}$$

In our notation this is

$$\hat{p}_a = p_a^+ + \gamma(\Delta t_F + \Delta t_R) + P^*, \tag{6.67}$$

whereas (6.13) implies

$$\hat{p}_a = \frac{C_a p_a^+ + C_d p_v^+}{C_a + C_s}. \tag{6.68}$$

De Boer *et al.*'s motivation for (6.66) is that a longer filling interval leads to a more powerful contraction (thus an increased stroke volume) via Starling's law, and also via the contractility of the myocardium, which also increases following a longer filling interval. This discussion muddies the waters, since figure 6.9 suggests no direct connection between heart rate and ventricular elastance. Starling's law is indeed manifested by (6.68), insofar as the increase of $\hat{p}_a$ with p_v^+ leads to an increase in stroke volume (cf. (6.17)) and thus cardiac output, but the effect on ventricular elastance is relatively small, and more importantly is effected through the slow sympathetic system. In fact, (6.68) suggests that the De Boer model should properly include a beat-to-beat model of the venous pressure, which would follow from (6.32).

For suitable choices of the model parameters, De Boer *et al.* found that model simulations produced Mayer waves, and also respiratory sinus arrhythmia, when respiratory forcing was included by allowing P^* in (6.66) to vary in time with the respiratory frequency (the assumption being that respiration affects the filling pressure via its effect on the intra-thoracic pressure).

Glossary

Acetylcholine

ADP

Adrenaline (epinephrine)

Afferent nerves

Alveoli

Aorta

Apnea

Absence of breathing.

Apoptosis

Pre-programmed cell death.

Apneustic centre

Arteries

These are the blood vessels which carry freshly oxygenated blood from the (left ventricle of the) heart to the tissues, where the oxygen load is consumed, and metabolically produced carbon dioxide is taken up.

ATP

Autonomic control

Baroreceptors

Baroreflex

Basophil

Blast cells

Bradycardia

A slow heart rate.

Bronchi

Capillary bed

As the arteries transport blood away from the heart, they divide into finer and finer passageways, first arterioles, and finally capillaries, extremely thin tubes which perfuse tissue and which can efficiently exchange blood gases with the tissue cells.

Carbonic anhydrase

An enzyme contained in red blood cells which facilitates the reaction of water and carbon dioxide to form bicarbonate ions, in which form most of the CO_2 in the blood is transported.

Carotid arteries

Carotid sinus

Catecholamines

Cerebro-spinal fluid (CSF)

Chemoreceptors

Cheyne-Stokes respiration

This is an oscillatory kind of breathing in which a waxing and waning pattern of breathing alternates with periods of complete *apnea*, with a typical period of a minute or so. Common causes of Cheyne-Stokes breathing are heart failure, stroke, and ascent to high altitude, when the period is less.

Chronotropic effect

Compliance

Dead space

Diaphragm

Diastole

Dicrotic notch

Dorsal respiratory neurons

Efferent nerves

Elastance

Eosinophil

Erythrocytes

Erythropoietin

A hormone which controls red blood cell production. Low blood oxygen levels stimulate production of erythropoietin in the kidneys, and this in turn stimulates erythrocyte production in the marrow, by encouraging the production of proerythroblasts, and by quickening their rate of maturation to form erythrocytes.

Glossopharyngeal nerves

Granulocytes

A collective name for the three types of white blood cell having a granular appearance: neutrophils, basophils, and eosinophils.

Granulopoietin

Haematocrit

The percentage of blood consisting of cells.

Haematopoietic stem cells

These are the most primitive cells resident in the bone marrow, which collectively give rise to the various kinds of blood cells: *erythrocytes*, *platelets* and white blood cells. The process of maturation which they undergo is called differentiation. Isolation of stem cells is difficult, insofar as no real most primitive cell has ever been really identified. It is likely that stem cells are very scarce, and that they can survive in the resting phase for a long time.

Haemoglobin

Hydrolysis

Hypercapnea

Hyperpnea

Inotropic effect

Intercostal muscles

Leukocytes

General term for white blood cells.

Lymph nodes

Lymphocytes

Mayer waves

Megakaryocytes

Minute ventilation

Monocyte

Myelocytes

Myocytes

Myocardium

Neutrophil

Noradrenaline (norepinephrine)

Parasympathetic nervous system

Phagocytosis

The process of cell ingestion by means of which granulocytes and monocytes destroy antigens.

Plasma cells

Platelets

Platelets are cell fragments which are formed by the disintegration of *megakaryocytes*. They circulate in the blood with a life span of about ten days, and are instrumental in clotting during wound healing.

Pneumotaxic centre

Residual volume

Respiratory sinus arrhythmia

Reticulocytes

Stroke volume

Sympathetic nervous system

Systole

Thoracic cavity

Thrombocytes

Another name for *platelets*.

Thrombocytopenia

A disease signalled by low circulating numbers of thrombocytes, or platelets.

Thrombocytopenic purpura

Another name for thrombocytopenia, arising from the purplish blotches on the skin of those afflicted.

Thrombopoietin

Tidal volume

References

Batchelor, G. K. 1967 An introduction to fluid dynamics. C. U. P., Cambridge.

Batzel, J. J. and H. T. Tran 2000a Modeling variable delay and instability in the control system for human respiration: application to infant non-rem sleep. J. Appl. Math. Comp. **110**, 1-51.

Beeler, G. W. and H. Reuter 1977 Reconstruction of the action potential of ventricular myocardial fibres. J. Physiol. **268**, 177-210.

Berne, R. M. and M. N. Levy 1996 Principles of physiology, 2nd ed. Mosby, St. Louis.

Berridge, M. J. and A. Galione 1988 Cytosolic calcium oscillators. The FASEB J. **2**, 3,074-3,082.

Brown, H. and R. Kozlowski 1997 Physiology and pharmacology of the heart. Blackwell Science, Oxford.

De Boer, R. W., J. M. Karemaker and J. Strackee 1987 Hemodynamic fluctuations and baroreflex sensitivity in humans: a beat-to-beat model. Amer. J. Physiol. **253**, 680-689.

Fowler, A. C. 2013 Note on a paper by Omta *et al.* on sawtooth oscillations. SeMA Journal **62**, 1-13.

Fowler, A. C. 2014 Starvation kinetics of oscillating microbial populations. Math. Proc. R. Ir. Acad. **114** (2), 173-189.

Glass, L. and M. C. Mackey 1988 From clocks to chaos. Princeton University Press, Princeton, New Jersey.

Grindrod, P. 1991 Patterns and waves. O. U. P., Oxford.

Grodins, F. S., J. Buell and A. J. Bart 1967 Mathematical analysis and digital simulation of the respiratory control system. J. Appl. Physiol. **22**, 260-276.

Guyton, A. C. and J. E. Hall 2000 Textbook of Medical Physiology, 10th ed. W. B. Saunders, Philadelphia.

Houghton, A. R. and D. Gray 1997 Making sense of the ECG, a hands-on guide. Arnold, London.

Huppert, A., B. Blasius, R. Olinky and L. Stone 2005 A model for seasonal phytoplankton blooms. J. Theor. Biol. **236**, 276-290.

Julian, D. G., J. C. Cowan and J. M. McLenachan 1998 Cardiology. Saunders, Edinburgh.

Katz, B. 1966 Nerve, muscle and synapse. McGraw-Hill, New York.

Keener, J. P. 1986 A geometrical theory for spiral waves in excitable media. SIAM J. Appl. Math, **46** (6), 1,039-1,056.

Keener, J. and J. Sneyd 1998 Mathematical physiology. Springer-Verlag, Berlin.

Khoo, M. C. K., R. E. Kronauer, K. P. Strohl, and A. S. Slutsky 1982 Factors inducing periodic breathing in humans: a general model. J. Appl. Physiol. **53**, 644-659.

Levick, J. R. 2000 An introduction to cardiovascular physiology, 3rd ed. Butterworth-Heinemann, Oxford.

Lewis, T. 1925 Mechanism and graphical registration of the heart beat. Shaw and Sons, London.

Luo, C.-H. and Y. Rudy 1994 A dynamic model of the cardiac ventricular action potential. I. Simulations of ionic currents and concentration changes. Circ. Res. **74**, 1,071-1,096.

Mackey, M. C. and L. Glass 1977 Oscillations and chaos in physiological control systems. Science **197**, 287-289.

Murray, J. D. 1993 Mathematical biology, 2nd ed. Springer-Verlag, Berlin.

Noble, D. 1962 A modification of the Hodgkin-Huxley equations applicable to Purkinje fibre action and pace-maker potentials. J. Physiol. **160**, 317-352.

Omta, A. W., G. A. K. van Voorn, R. E. M. Rickaby and M. J. Follows 2013 On the potential role of marine calcifiers in glacial-interglacial dynamics. Global Biogeochem. Cycles **27**, 692-704.

Ottesen, J. T. 1997 Modelling of the baroreflex-feedback mechanism with time-delay. J. Math. Biol. **36**, 41-63.

Segel, L. A. 1984 Modeling dynamic phenomena in molecular and cellular biology. C. U. P., Cambridge.

Ursino, M. 1998 Interaction between carotid baroregulation and the pulsating heart: a mathematical model. Amer. J. Physiol. **275**, H1733-H1747.

Whitham, G. B. 1974 Linear and nonlinear waves. John Wiley, New York.

Woods, N. M., K. S. R. Cuthbertson and P. H. Cobbold 1986 Repetitive transient rises in cytoplasmic free calcium in hormone-stimulated hepatocytes. Nature **319** (6,054), 600-602.