CUED - Engineering Tripos Part IIB 2019-2020

Module Coursework

Module	4G3	Title of report	4G3 - Coursework Two - Hopfield network and spiking neurons

Date submitted: 26/3/20	Assessment for this module is ☒ 100% / ☐ 25% coursework of which this assignment forms _____ %

UNDERGRADUATE STUDENTS ONLY		**POST GRADUATE STUDENTS ONLY**	
Candidate number: 5584F		Name:	College:

Feedback to the student

☐ **See also comments in the text**

		Very good	**Good**	Needs improvmt
C O N T E N T	**Completeness, quantity of content:** Has the report covered all aspects of the lab? Has the analysis been carried out thoroughly?			
	Correctness, quality of content Is the data correct? Is the analysis of the data correct? Are the conclusions correct?			
	Depth of understanding, quality of discussion Does the report show a good technical understanding? Have all the relevant conclusions been drawn?			
	Comments:			
P R E S E N T A T I O N	**Attention to detail, typesetting and typographical errors** Is the report free of typographical errors? Are the figures/tables/references presented professionally?			
	Comments:			

Overall assessment (circle grade)	A*	**A**	B	C	D
Guideline standard	>75%	**65-75%**	55-65%	40-55%	<40%
Penalty for lateness:		*20% of marks per week or part week that the work is late.*			

Marker: Date:

4G3 - Coursework Two - Hopfield network and spiking neurons

Candidate Number: 5584F

March 28, 2020

Abstract

Hopfield networks and firing using the Hodgkin-Huxley model were investigated. In a Hopfield network partial memories can be retrieved as stored memories act as attractors for convergence on average. Errors can occur in retrieval of noisy memories as other attractors e.g inverted memories are converged upon. The Hodgkin-Huxley model includes the effect of different ion channels on membrane currents and potentials. This extra complexity allows the modelling of behaviours that oscillate in different limit cycles, respond to resonant frequencies and have sharp input thresholds to induce firing.

I. Introduction

Associative memory is the process of retrieving a memory from partial data. This is to be investigated using a Hopfield network, to consider the effect of noise and instability within the network.

A physiologically realistic model of a spiking neuron was then investigated using the Hodgkin-Huxley model to discover the differences from more simplistic models such as leaky integrate and fire neurons.

II. Hopfield network

A binary Hopfield network was investigated. This network is fully connected with each neuron having an activity of 0 or 1. This network can be used to store binary strings (memories), which are retrievable with partial information about the memory. This retrieval is possible as the network weights are chosen as in equation 1, this choice causes memories to be stable fixed points on average and thus can be converged upon. The success of this convergence was analysed.
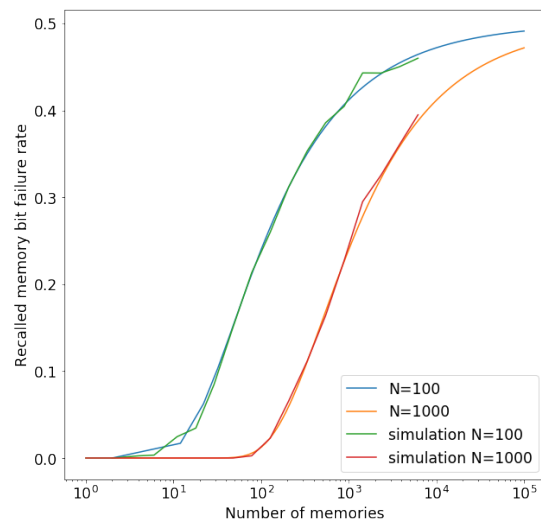
i. Theoretical error probability

Figure 1: *Simulations one step*

Errors during memory recall can occur from:

- Points that are not memories act as attractors and cause convergence, these spurious attractors take the form of combinations of stored memories or inverted memories
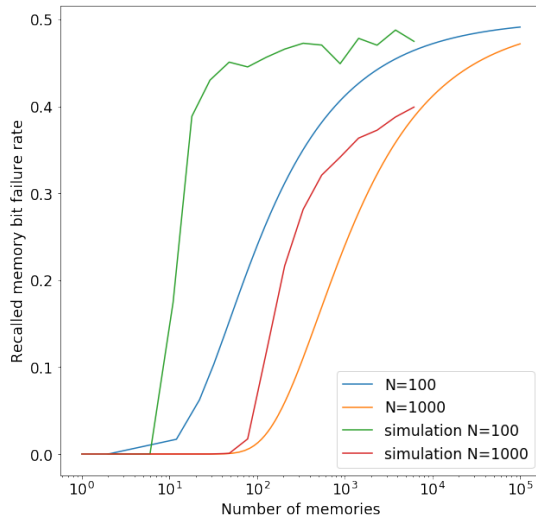
1

Figure 2: *Simulations 3000 step*

- Stored memories are unstable, and therefore with a disturbance the retrieval will drift or oscillate
- Stored memories are not fixed points therefore convergence to them doesn't occur

Within the space of dimension N exists the local field on each bit $H_k = \sum_{j \neq k} W_{kj} r_j(t)$ this is prevalent in the update term in equation 2. This field is constructed as in equation 1 such that the field points towards stored memories on average. This helps the retrieval of partial memories. The field can be broken down into two terms:

$$H_k = (r_k^{(\mu)} - \frac{1}{2}) \sum_{j \neq k} r_j^{(\mu)}(r_j^{(\mu)} - \frac{1}{2})$$
$$+ \sum_{j \neq k} r_j^{(\mu)} \sum_{m \neq \mu} ((r_j^{(m)} - \frac{1}{2})(r_k^{(m)} - \frac{1}{2})$$

With the first term being the contribution from the desired memory μ and the second the contribution from all other stored memories and is classed as noise. The two terms have the following mean and variance, given the desired memory and with each state equally probable:

$$< signal >= \pm \frac{N-1}{8}$$
$$< noise >= 0$$
$$<< signal >>= 0$$
$$<< noise >>= \frac{(N-1)(M-1)}{32}$$

On average the noise terms are zero which means that memory states are on average a stable fixed point. However, if the noise term is larger then the signal mean, and resides in the correct tail the local field can point the wrong direction. This means that memory is not a fixed point. This error can be theoretically computed as $P(H_k < 0 | r_k^{(\mu)} = 1)$, where $H_k \sim \mathcal{G}(\frac{N-1}{8}, \frac{(N-1)(M-1)}{32})$. H_k is Gaussian as its the sum of many random variables so is Gaussian by the CLT. The error as a function of stored memories is shown in figure 1.

As the number of memories stored increases, memories and attractors become closer within the space ($\mathcal{R}^N$). This increases the weight of each attractor on a memory being retrieved and therefore increases the noise prevalent in convergence. This additional noise increases the error probability.

The calculated probability is conditional on the memory being the starting point, and then checking the field. The probability of a state being fixed but not stable is not considered. This probability will effect error rates when noise or full convergence is simulated.

ii. Simulations

To create a network that converges upon memories in the desired way connectivity weights are chosen as in equation 1. When recalling a memory at each iteration a single neuron is asynchronously updated using equation 2. These updates can be repeated until convergence.

$$W_{ij} = \begin{cases} \sum\limits_{m=1}^{M} (r_i^{(m)} - \frac{1}{2})(r_j^{(m)} - \frac{1}{2}) & i \neq j \\ 0 & i = j \end{cases} \tag{1}$$

$$r_i(t + \Delta t) = \frac{sign[\sum\limits_{j=1}^{N} W_{ij}r_j(t)] + 1}{2} \qquad (2)$$

When simulating network convergence you can't isolate one type of error as additional update steps have a chance of causing error from one of the three methods. This means that the bit error calculated will include differences caused by unstable/oscillatory memories in addition to non-fixed point memories as theoretically calculated. Errors caused in one bit can effect later update steps and cause more errors. However, if you simulate just one step then check if the memory has been recalled correctly you are testing on average if a bit converges correctly therefore checking fixed points. Figure 1 shows the results of this simulation for N=100 and 1000, averaged over 100 networks. This result is shown to agree very closely to the theoretical error term calculated above.

When recalling memories convergence is required to properly test recall ability. If the network is ran for 3000 steps to allow for convergence to occur the results in figure 2 are produced. This produces a steeper curve then the predicted error probability as spurious attractors and unstable/oscillatory memory states have a greater effect. This bit error rate is more useful to consider when calculating network capacities.

iii. Noise

Figure 3 shows the results from simulating 10 Hopfield networks for 3000 steps to allow for convergence, and calculating the average probability of a bit being incorrectly recalled from a noisy memory. The noisy memory is made by flipping each bit in a memory by a certain probability.

The results show that noise levels below 0.2 result in similar curves to a noiseless case, with levels above 0.8 being the same but mirrored in y=0.5. At a noise level of 0.5 the recall rate is quite noisy but can be seen to be a constant 0.5, independent of M. This shows that convergence is happening to random memories and
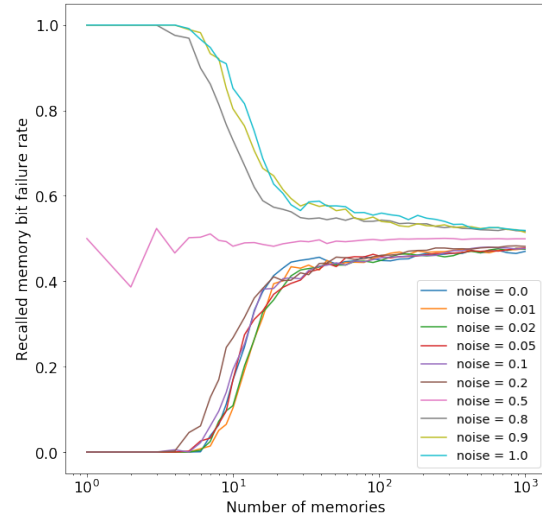

Figure 3: *Simulations with varying noise levels*

not the initial memory.

All curves converge upon a failure rate of 0.5 as the number of stored memories increases. This is due to the space ($\mathcal{R}^N$) becoming very crowded, meaning small changes due to noise cause corrupted memories to converge upon different memories or spurious attractors, and therefore have 50% of bits different on average. At a noise level above 0.5 the memory gets changed to be closer to its inverted memory. In a Hopfield network inverted stored memory are also attractors, this means that when retrieving the memory the attraction from the inverted memory is stronger then the true memory. Therefore convergence occurs to that and the bit failure rate is close to or equal 1.

Surprisingly when noise was 0 the network didn't perform the best. This would be expected as unstable fixed point memories would be greatly effected by the introduction of noise. Therefore causing greater bit errors. This difference could be explained due to insufficient averaging across networks, and the inherent randomness in testing.

## III. Physiological model of a spiking neuron

To build a physiological model of spiking neurons different channels into a neuron must be considered. In these experiments the Hodgkin Huxley model of action potentials will be investigated. This model includes channels for a leakage current (L), potassium ions (K) and sodium ions (Na). These channels have maximum conductance densities and reversal potentials associated with them defined by g_x, e_x respectively. The sodium and potassium channel conductances are functions of both time and voltage, and therefore require the additional m, h, and n terms. These terms are sodium channel activation, sodium channel deactivation and potassium channel activation respectively and are controlled by α and β functions. This model could be extended to include additional ion channels. Equations 3 define the forward euler update steps used during simulations.
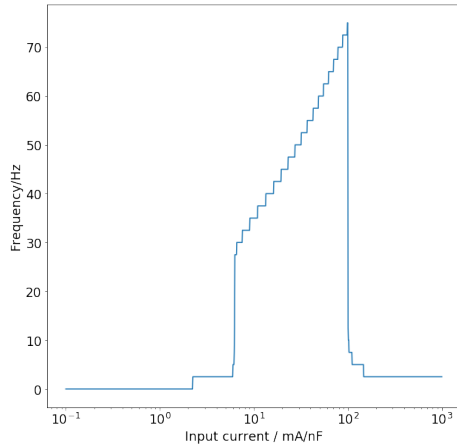
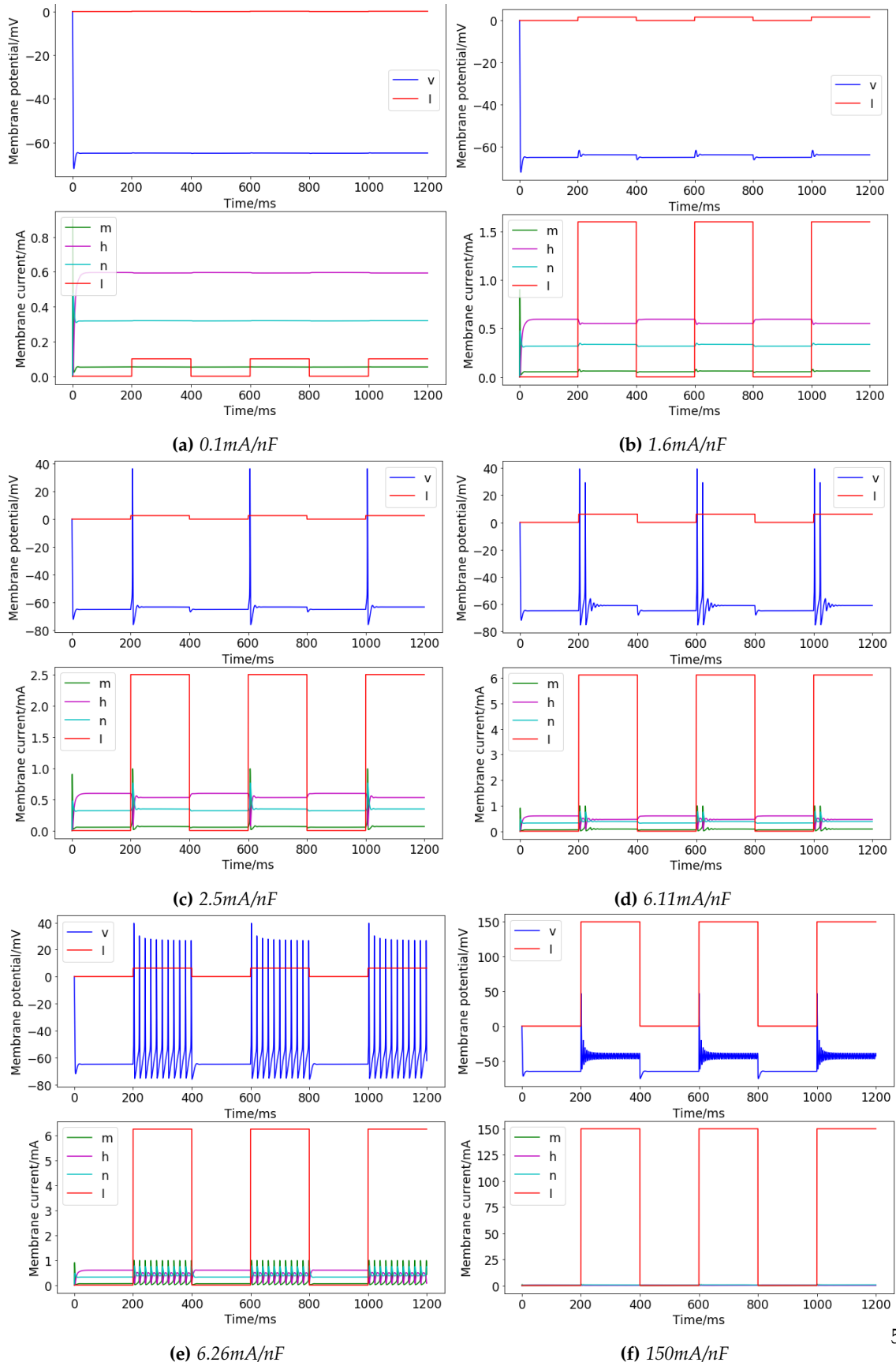**Figure 4:** *Response to 200ms pulse, 400ms period input*

$$v_{t+\Delta t} = (-g_{Na}.m_t^3 h_t(v_t - e_{Na}) - g_K.n_t^4(v_t - e_K)$$
$$-g_L(v_t - e_L) + I)\Delta t + v_t$$
$$m_{t+\Delta t} = (\alpha_m(v_t)(1 - m_t) - \beta_m(v_t)m_t)\Delta t + m_t$$
$$h_{t+\Delta t} = (\alpha_h(v_t)(1 - h_t) - \beta_h(v_t)h_t)\Delta t + h_t$$
$$n_{t+\Delta t} = (\alpha_n(v_t)(1 - n_t) - \beta_n(v_t)n_t)\Delta t + n_t$$
$$(3)$$

i. Constant pulse
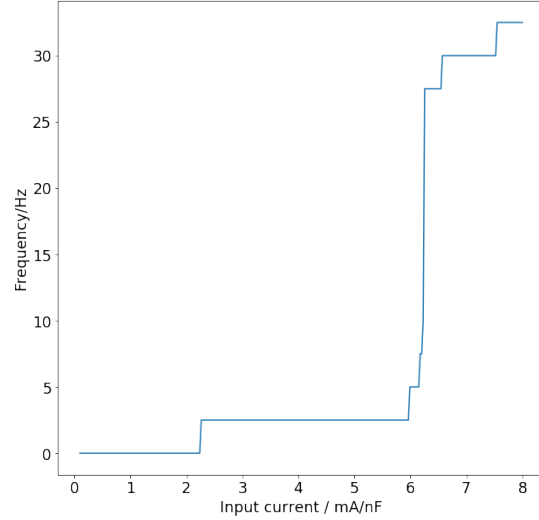
Figure 5: *Response to 200ms pulse, 400ms period input at low currents*

The external input I was varied from 0.1mA/nF - 1A/nF for pulse of length 200ms, and period 400ms. Figure 4 shows the firing frequency as a function of this input. A sharp threshold is observed for continuous firing to occur. This happens between 6.11-6.26mA/nF. This is an example of an all or none response to the input. The firing rate increases approximately logarithmically in steps until an input around 100mA/nF upon which the response follows the pulse shape with very small oscillations and one spike as in figure 6f. Example responses can be seen before the firing threshold in figures 6a and 6b, with 0.1mA/nF causing almost no change, and 1.6mA/nF causing small disturbances from the rest potential/current. With an input of 2.5mA/nF as in figure 6c single spikes are observed in membrane potentials and currents. A double spike is briefly possible at 6.11mA/nF (6d) before continuous firing is reached at 6.26mA/nF (6e). During continuous firing, spikes occur at regular intervals, this is due to the limit cycle prevalent in the model, which causes regular oscillations.

4

(a) *0.1mA/nF*

(b) *1.6mA/nF*

(c) *2.5mA/nF*

(d) *6.11mA/nF*

(e) *6.26mA/nF*

(f) *150mA/nF*

Figure 6: *Example responses to 200ms pulse, 400ms period for different applied currents*
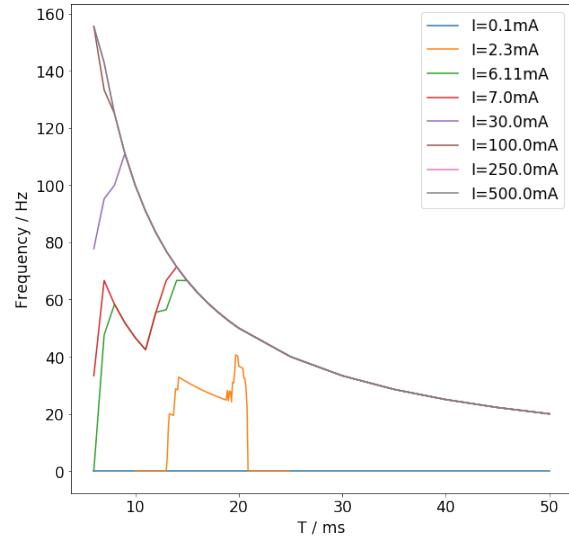
ii. Variable period

The input was next changed to allow for variation in period. A pulse length of 5ms was used with periods from 6ms to 50ms tested with different amplitudes, 15 periods were simulated. Figure 7a shows the spiking frequency of the investigations. As the frequency was dependant on a variable(T) it was clearer to instead compare spikes per pulse as in figure 7b. The test inputs were chosen to be at points of interest from the previous experiment.

At an input of 2.3mA/nF sharp threshold was once again observed at periods of 13.5ms and 21ms. Figures 8a-8c show example plots of this behaviour. The spike count never reached an average of one spike per peak for this input, this shows that there is some instability in the dynamics which cause some pulses to not initiate a spike as in figure 8c. At periods of 18.7ms this instability was very variable and caused the noisy point point on figure 7b.

For a small input of 0.1mA/nF no spiking response was seen as with a constant pulse. Inputs 6.11mA/nF and 7mA/nF were chosen as this was the threshold for continuous firing for a constant pulse. Figure 7b shows that the spiking behaviour of both was very similar, and initially started at no spiking or firing at 1 in 5 pulses and increased to one spike per pulse with a 14ms period. Alternate pulse spiking was observed before reaching consistent spiking. Figures 8d-8e show this behaviour. Additional combinations of spikes were observed at 6.11mA/nF where every fourth pulse resulted in no response at a 13ms period, this is shown in figure 8f. These extra periodic behaviours indicate more complicated limit cycles present in the system.

As the input increased a smaller period was required to reach consistent firing on each pulse. This indicated that the input has an effect of the recovery time of the system.
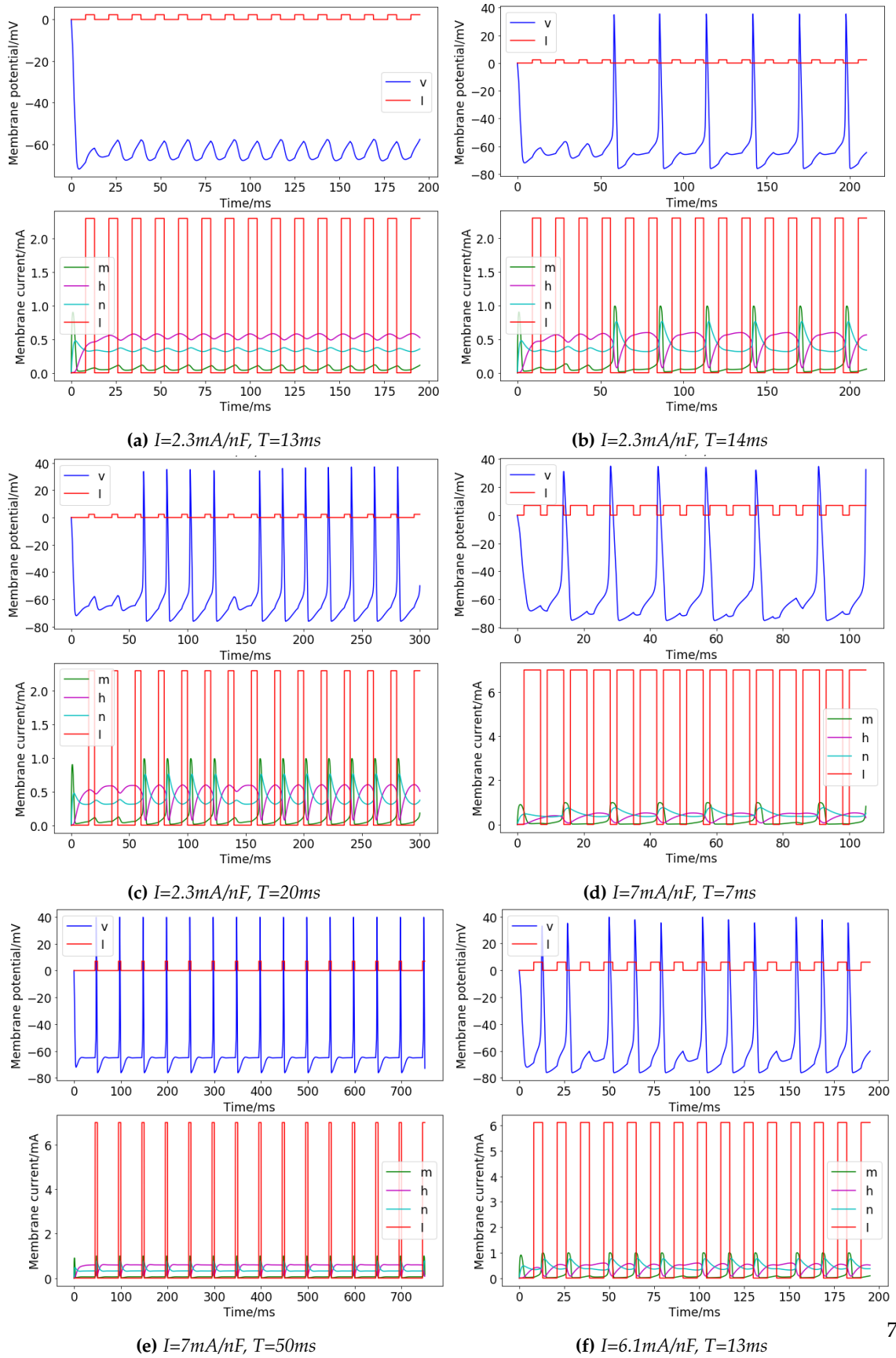

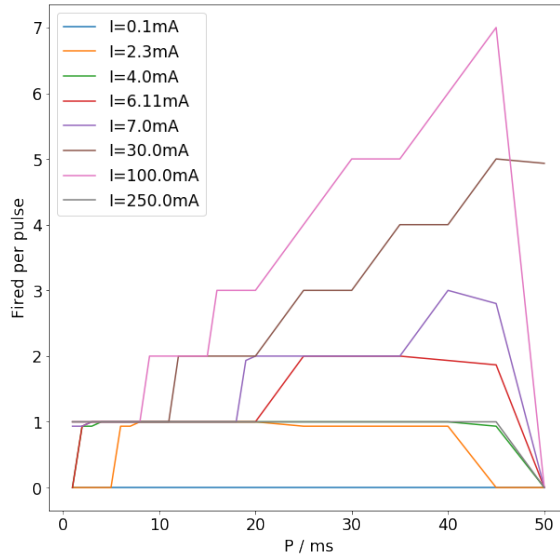**(a)** *Spike frequency*
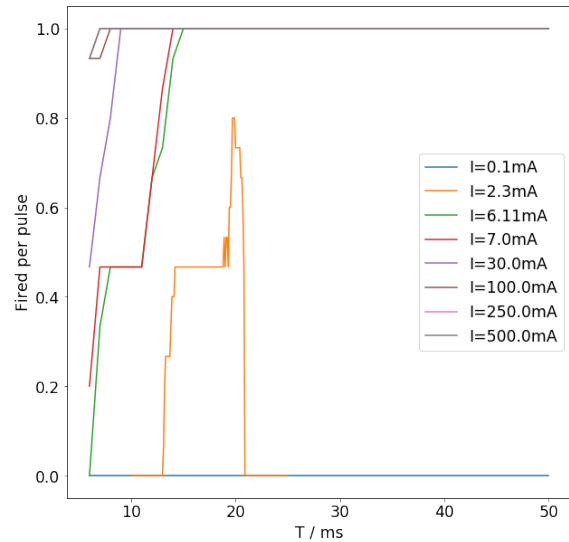
(b) *Spikes per pulse*

Figure 7: *Neuron response to 5ms pulse with variable period*

(a) *I=2.3mA/nF, T=13ms*

(b) *I=2.3mA/nF, T=14ms*

(c) *I=2.3mA/nF, T=20ms*

(d) *I=7mA/nF, T=7ms*

(e) *I=7mA/nF, T=50ms*
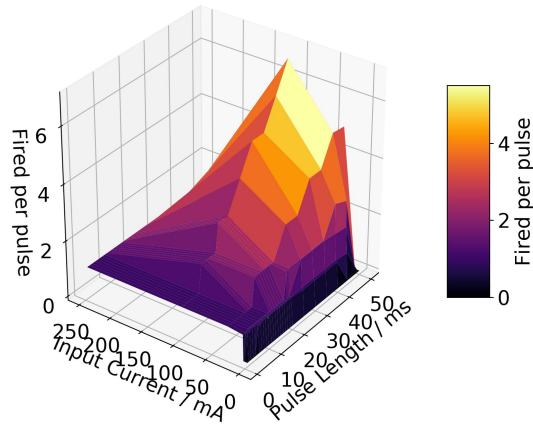
(f) *I=6.1mA/nF, T=13ms*

7

Figure 8: *Example responses for various periods(T) and applied currents(I) for a 5ms pulse*

iii. Variable pulse

(a)

(b) *Surface plot*

Figure 9: *Response when period was fixed to T=50ms*

A variable pulse length was investigated next, the period was fixed at 50ms and different currents were once again tested. Figure 9 shows the counts per pulse achieved in this investigation. A complex surface is observed in figure 9b showing peak response doesn't occur at maximum pulse length or current. This surface would include additional thresholds between currents if more simulations had been run. All responses decreased when a contin-

uous input was applied, except at 30mA/nF. With this input and 50ms pulse length continuous firing was still occurring. This is seen in figure 11e, but looking at other inputs such as 100mA/nF as in figure 11f its clear that the input causes an oscillatory response of different amplitudes at other currents which weren't large enough to be counted as a spike in my tests.

The spike position relative to pulse start time appears constant for a given input. Figures 11a and 11b show that a spike occurs within 2ms of a a pulse for an input of 7mA/nF. Figures 11c and 11d show that a spike occurs within 6ms of a a pulse for an input of 2.3mA/nF. This is expected as the membrane currents depend on the potential, which in turn is affected by the input current. As membrane currents change the time constant of the system will also be effected.
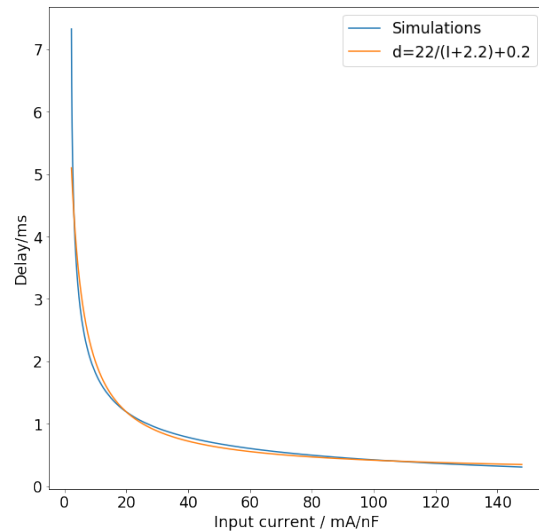
Figure 10: *Spike delay with 200ms pulse, 400ms period input*

Figure 10 shows the delay in ms between pulse start and firing, for a 200ms pulse with 400ms period. The relationship observed can be approximately modelled by an equation of the form: delay = $a/(I+b)$ +c. As this delay is predictable it can be useful in networks to add a desired delay especial at low currents where delay is highly sensitive.
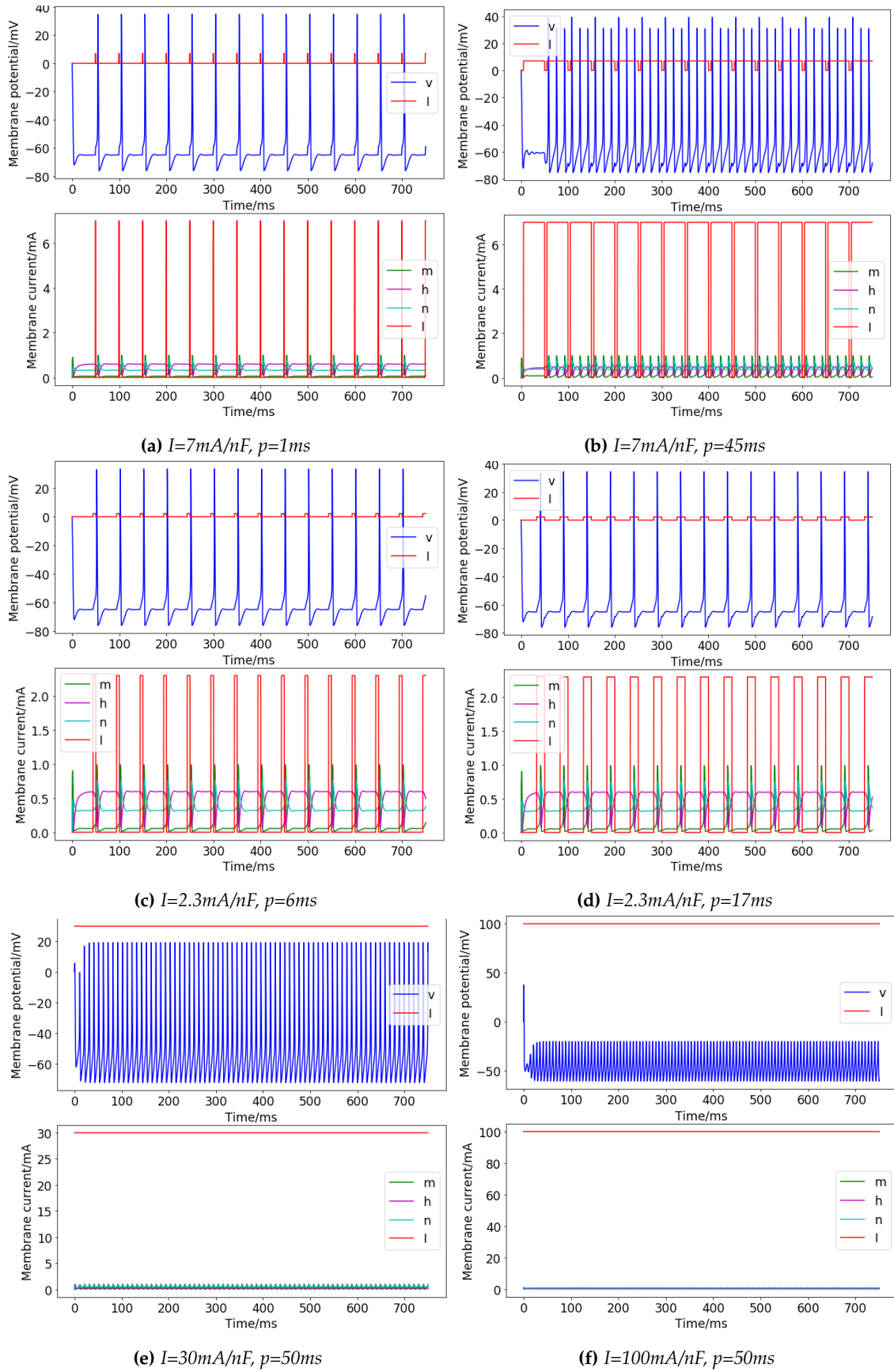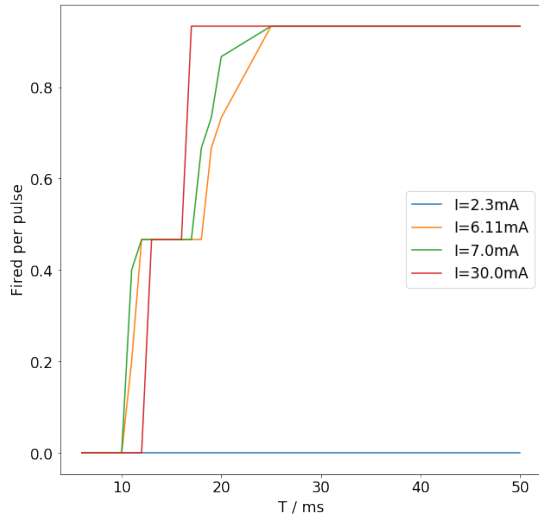
(a) *I=7mA/nF, p=1ms*

(b) *I=7mA/nF, p=45ms*

(c) *I=2.3mA/nF, p=6ms*

(d) *I=2.3mA/nF, p=17ms*

(e) *I=30mA/nF, p=50ms*

(f) *I=100mA/nF, p=50ms*

Figure 11: *Example responses to different pulse lengths with a 50ms period*

iv. Negative pulse

Figure 12: *Neuron response to negative current input*

Negative current inputs were investigated, it was found that interesting spiking behaviour and oscillatory responses were still possible at low negative currents. When the current got too large the membrane potential increased exponentially and broke simulations. Figure 13 shows the effect of period length on spiking when a 5ms pulse is applied. A rate of 1 spike per peak is incorrectly not observed as the positive spike would have occurred after simulation time for the final pulse as an extra delay was present with negative inputs.

A similar behaviour was observed in inputs of -6.11mA/nF and -7mA/nF as in figure 7b, however spiking occurred at a period of around 14ms compared to 5ms. No responses was observed to an input of -2.3mA/nF indicating the threshold was different for negative currents. Figure 13c shows this lack of spikes. For larger currents that worked in simulations such as -30mA/nF the spiking frequency was reduced to be similar to that of inputs around -6 to -7mA/nF. This input current also produced interesting oscillatory behaviour seen in figures 13d-13f. Negative saw-wave oscillations, and negative spiking in addition to positive spikes are present in various frequencies. These responses could prove useful in situations where negative pulses are desired as well, creating a neuron that can act as inhibitory and excitatory. For example dopamine can be both a inhibitory and excitatory neurotransmitter.
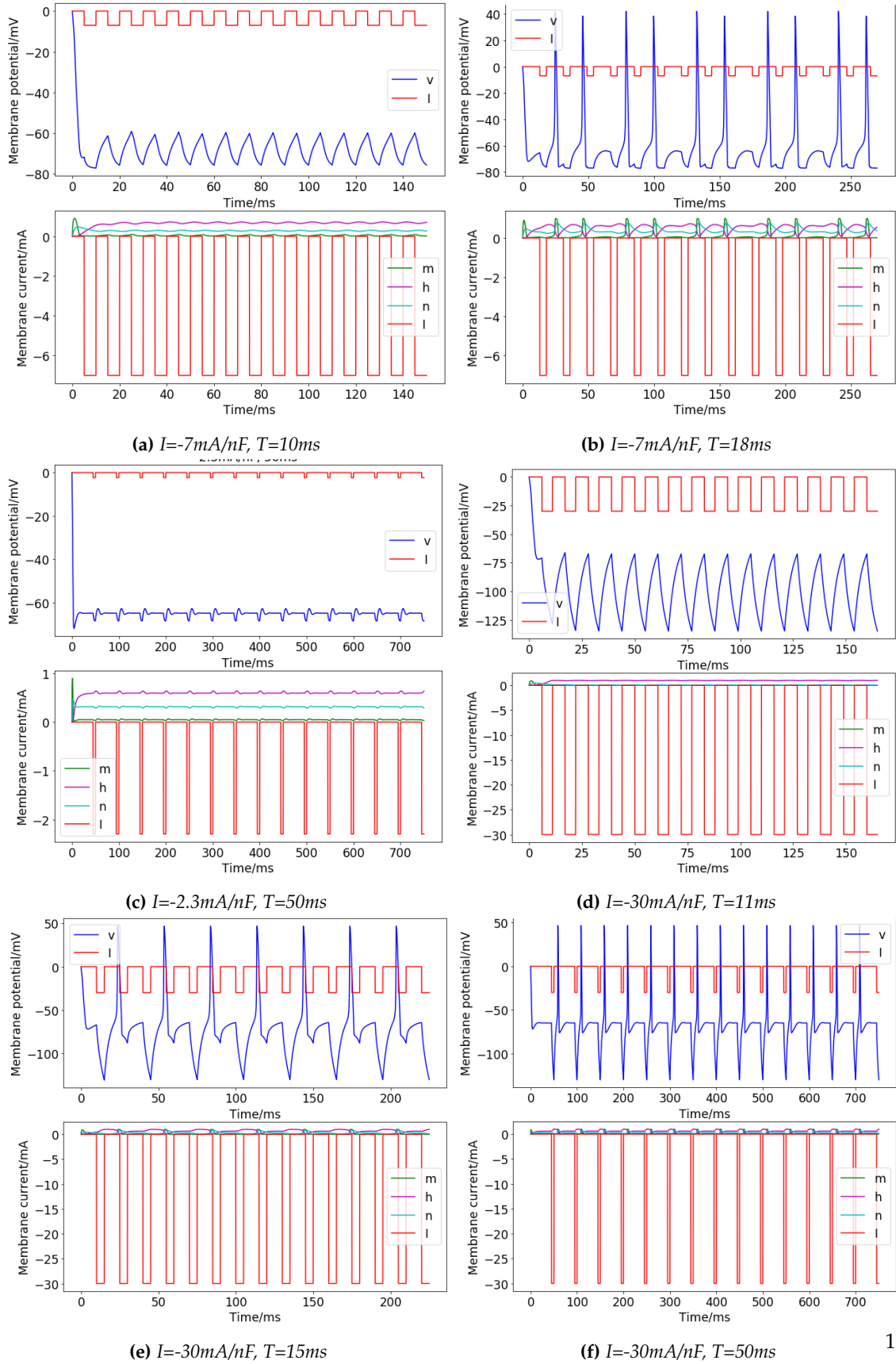
A threshold of I= -2.8mA/nF was observed to initiate firing when using a 200ms pulse with 400ms period. Figure 14 shows this threshold and that a regular 1 spike per period is achieved for all currents that exceed this threshold.

v. Conclusions

- A sharp current threshold is observed to initiate continuous firing
- Firing rate doesn't increase forever, another threshold is reached that reduced spikes down to 1 per pulse
- The system resonates and fires
- Cyclic patterns of different orders are observed when using small pulse lengths, and negative currents
- These patterns are the result of limit cycles present in the dynamics
- Firing delay is a decreasing function of input current
- Negative currents can cause both negative and positive spikes
- Low potential oscillations can be produced with negative currents or when at currents that don't cause spikes

The wide variety of behaviours is useful in biology as it means neurons can be controlled with different inputs or modulators to produce the behaviour required in the current situation. This makes them very adaptable.

Resonators such as this model differ from more simplistic integrate and fire models as the firing rate is more likely when the input is at its resonant frequency, opposed to a higher frequency causing additional firing in integrators. This means that using the correct model for the biological system being studied can have great effect on the simulated responses. The simpler model while computationally faster will not include the complex behaviour present in resonators, so would be unsuitable in certain situations.

(a) *I=-7mA/nF, T=10ms*

(b) *I=-7mA/nF, T=18ms*

(c) *I=-2.3mA/nF, T=50ms*

(d) *I=-30mA/nF, T=11ms*

(e) *I=-30mA/nF, T=15ms*

(f) *I=-30mA/nF, T=50ms*

11

Figure 13: *Example responses to negative currents and different periods when p=5ms*
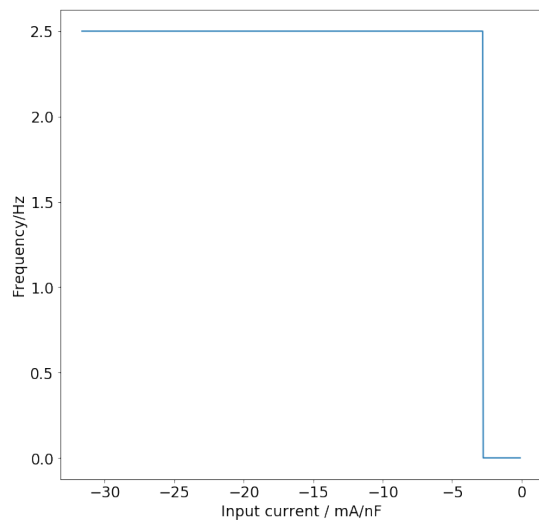
Figure 14: *Negative current input to regular pulse*