Boucheny, Brunel, Arleo 2004

Replication : A Continuous Attractor Network Model Without Recurrent Excitation: Maintenance and Integration in the Head Direction Cell System

A few weeks ago, Guillaume, the computational post-doc at the lab did an excellent review of over 30 years of head-direction (HD) system modeling. It might be surprising, but despite this extensive effort the entire pathway for HD coding has yet to be completely understood. It is even more surprising considering how appropriate this system is to ask theoretical questions. An attractor dynamic is notably present in the system [1], which is used code for a very easily understood 1D variable: the head-direction, enforcing the stability of the population activity to be confined onto a ring. The pathway extends from the thalamus to the cortex, and the study of the different cells types along it enables to shed light on how they each have a role for the computation. Nonetheless, the main modeling effort was focused on the circuit which first integrates the angular velocity signal coming from vestibular inputs and generate the attractor dynamic. It was striking to see in Guillaume's presentation how the model designs were guided by the experimental data and progressively shaped our current understanding of this circuit. The modeling effort diminished after 2005 as Pengcheng Song and Xiao-Jing Wang model accounted for the remaining empirical evidences [2], notably the absence of recurrent excitation which had until then been overlooked was replaced by lateral connections: $i \rightarrow i, e \rightarrow i$, $i \rightarrow e$. Shortly after, in the same year, Bouchemy, Brunel and Arleo [3] completed the analysis by considering how strong the lateral $i \rightarrow i$ connections could be. In this tiny report, I have tried to reimplement this last model, it was really interesting to go through the same early path as these computational neuroscientists which have now transformed their field. My goal is not to go through the article, or explain it. Instead I will stress details where I am in a disagreement with the authors, and how well I managed to replicate the study. Please feel free to make any comments you wish, either through mail or by github, these reimplementations are a way for me to get experiences and feedbacks are wellcome.
We will divide the analysis in two, first I stress a few (that I believe to be) mistakes in the analyses and then I will present problem I had to create the code.

Mathematical analysis

The paper is perfect for a beginner as it first goes through a simpler model to show the main idea, and progressively complexify it. Nonetheless a few inconsistency made it difficult for me to believe it until I managed to replicate the figures.
Notably, they model the system with the following set of equations:

$$\tau \dot{s}_e(\theta,t) = -s_e(\theta,t) + f_e(\theta,t) \quad (11)$$

$$\tau \dot{s}_l(\theta,t) = -s_l(\theta,t) + f_l(\theta,t) \quad (12)$$

$$\tau \dot{s}_r(\theta,t) = -s_r(\theta,t) + f_r(\theta,t) \quad (13)$$

$$f_e(\theta,t) = \left[I_E(t) - \int_{-\pi}^{\pi} \frac{K(\theta-\theta'-\pi+\alpha)s_l(\theta',t) + K(\theta-\theta'-\pi-\alpha)s_r(\theta',t)}{4\pi} d\theta' \right]_+ \quad (14)$$

$$f_l(\theta,t) = \left[I_I + I_l(t) + \int_{-\pi}^{\pi} \frac{H(\theta-\theta')s_e(\theta',t) + L_0 s_r(\theta',t)}{2\pi} d\theta' \right]_+ \quad (15)$$

$$f_r(\theta,t) = \left[I_I + I_r(t) + \int_{-\pi}^{\pi} \frac{H(\theta-\theta')s_e(\theta',t) - L_0 s_l(\theta',t)}{2\pi} d\theta' \right]_+ \quad (16)$$

Where $K(\theta) = K_0 + K_1 \cos(\theta)$ and $H(\theta) = H_0 + H_1 \cos(\theta)$. As we will see, there is a mistake in equation (15): the sign in front of L_0 should be a -. Let us see why.
For $A = e, l, r$ the author then define:

$S_{A0}(t) = \int_{-\pi}^{\pi} \frac{s_A(\theta,t)}{2\pi} d\theta$, (17)

$S_{A1}(t) = \int_{-\pi}^{\pi} s_A(\theta,t)\cos(\theta - \psi_A(t)) \frac{1}{2\pi} d\theta$ (18)

$0 = \int_{-\pi}^{\pi} s_A(\theta,t)\sin(\theta - \psi_A(t)) \frac{1}{2\pi} d\theta$ (19)

Why is the last integral 0? We can take $\psi_A(t)$ to be the argument of $\frac{\int_{-\pi}^{\pi} s_A(\theta,t)e^{i\theta} d\theta}{2\pi}$, ie the phase, or center of mass of the activation variable.
The authors then presented that the integrals (14-16) become

$$f_e(\theta,t) = \Big[I_E(t) - \frac{K_0}{2}[s_{l0}(t) + s_{r0}(t)] + \frac{K_1}{2}[\cos(\theta + \alpha - \psi_l(t))s_{l1}(t) + \cos(\theta - \alpha - \psi_r(t))s_{r1}(t)] \Big]_+ \quad (20)$$

$$f_l(\theta,t) = [I_I + I_l(t) + H_0 s_{e0}(t) - L_0 s_{r0}(t) + H_1 \cos(\theta - \psi_e(t))s_{e1}(t)]_+ \quad (21)$$

$$f_r(\theta,t) = [I_I + I_r(t) + H_0 s_{e0}(t) - L_0 s_{l0}(t) + H_1 \cos(\theta - \psi_e(t))s_{e1}(t)]_+ \quad (22)$$

But when made carefully, we observe that in equation (21) it should be $+L_0 s_{r0}$ according to (15). My belief is that the mistake was not made here but instead in equation (15) and that equation (21) is correct.

Then more errors are found. Plugging back the above equations for the firing rates into (17-18-19) and integrating (11) to (13), the author obtained :

$$\tau \dot{s}_{A0}(t) = -s_{A0}(t) + \int_{-\pi}^{\pi} f_A(\theta, t) \frac{d\theta}{2\pi} \qquad (23)$$

And: .

$$\tau \dot{s}_{A1}(t) = -s_{A1}(t) + \int_{-\pi}^{\pi} f_A(\theta, t) \frac{\cos\theta d\theta}{2\pi} \quad (24)$$

$$\tau \dot{\psi}_A(t) s_{A1}(t) = \int_{-\pi}^{\pi} f_A(\theta, t) \frac{\sin\theta d\theta}{2\pi} \qquad (25)$$

But I do not agree with (24), instead it should be:

$$\tau s_{A1}^{'}(t) = -s_{A1}(t) + \int_{-\pi}^{\pi} \frac{f_A(\theta, t)\cos\left(\theta - \psi_A(t)\right) d\theta}{2\pi}$$

And similarly for (25) the ψ_A terms was missing.

These mistakes were quite minors and I think unpresent in the authors simulations. Nut now comes a more important one, which if not taken care of can dramatically alter the computations.

The author pushed further the analyses with the following:

We rewrite the equations for the firing rate profiles as

$$f_A(\theta, t) = [I_{A0}(t) + I_{A1}(t)\cos(\theta - \phi_A(t))]_+ \qquad A = e, l, r \qquad (26)$$

$$I_{e0}(t) = I_E(t) - \frac{K_0}{2}(s_{l0}(t) + s_{r0}(t)) \qquad (27)$$

$$I_{l0}(t) = I_I + I_l(t) + H_0 s_{e0}(t) - L_0 s_{r0}(t) \qquad (28)$$

$$I_{r0}(t) = I_I + I_r(t) + H_0 s_{e0}(t) - L_0 s_{l0}(t) \qquad (29)$$

$$I_{e1}(t) = \frac{K_1}{2}(\cos(\psi_l(t) - \alpha - \phi_e(t)) s_{l1} + \cos(\psi_r(t) + \alpha - \phi_e(t)) s_{r1}(t)) \qquad (30)$$

$$0 = \sin(\psi_l(t) - \alpha - \phi_e(t)) s_{l1}(t) + \sin(\psi_r(t) + \alpha - \phi_e(t)) s_{r1}(t) \qquad (31)$$

$$I_{l1}(t) = H_1 s_{e1}(t) \qquad (32)$$

$$I_{r1}(t) = H_1 s_{e1}(t) \qquad (33)$$

$$\phi_l(t) = \psi_e(t) \qquad (34)$$

$$\phi_r(t) = \psi_e(t) \qquad (35)$$

$$\int_{-\theta_A}^{\theta_A} \frac{d\theta}{2\pi}(I_{A0} + I_{A1}\cos(\theta)) = I_{A1} f_0(\theta_A) \qquad (36)$$

$$\int_{-\theta_A}^{\theta_A} \frac{d\theta}{2\pi}(I_{A0} + I_{A1}\cos(\theta))\cos\theta = I_{A1} f_1(\theta_A) \qquad (37)$$

where f_0 and f_1 are given by

$$f_0(x) = \frac{1}{\pi}(\sin x - x\cos x) \qquad (38)$$

$$f_1(x) = \frac{1}{2\pi}\left(x - \frac{\sin 2x}{2}\right) \qquad (39)$$

Then define $\theta_A(t) = arcos\left(-\frac{I_{A0}(t)}{I_{A1}(t)}\right)$

The computation of (36) seems false.

Let us do it step by step:

$$\int_{-\theta_A}^{\theta_A} \frac{d\theta}{2\pi}\left(I_{A0} + I_{A1}\cos(\theta)\right) = \frac{I_{A0}\theta_A}{\pi} + \frac{I_{A1}}{2\pi}[\sin(\theta)]_{-\theta A}^{\theta_A} = \frac{I_{A0}\theta_A}{\pi} + \frac{I_{A1}\sin\left(\theta_A\right)}{\pi}$$

Now the author probably wanted to use that:

$\theta_A = arcos\left(-\frac{I_{A0}}{I_{A1}}\right)$, and simplify $I_{A0} = \cos\left(\theta_A\right) I_{A1}$, the issue is that this is not exactly true.

To be more precise, we will integrate $f_A(\theta, t)$ over $[-\pi, \pi]$ and define its support as $[\phi_A - \theta_a, \phi_A + \theta_A]$, the issue is that for some values of I_{A0} and I_{A1} the support might be $\emptyset$. So we will need to integrate over $[-\pi, \pi]$, in which case it is false that $\pi = arcos\left(-\frac{I_{A0}}{I_{A1}}\right)$ and so $I_{A0} \neq \cos(\pi) I_{A1}$

This <u>will</u> happen during the resolution of the problem.

As such I use the unsimplified version of (36):

$\int_{-\theta_A}^{\theta_A} \frac{d\theta}{2\pi}\left(I_{A0} + I_{A1}\cos(\theta)\right) = \frac{I_{A0}\theta_A}{\pi} + \frac{I_{A1}\sin(\theta_A)}{\pi}$ *(36)* where $\theta_A = \pi \; if \; support(f_A) = \emptyset$

Let us check (37):

$$\int_{-\theta_A}^{\theta_A} \frac{d\theta}{2\pi}\left(I_{A0} + I_{A1}\cos(\theta)\right)\cos(\theta) = \frac{I_{A0}\sin(\theta_A)}{\pi} + \frac{I_{A1}}{2\pi}\int_{-\theta_A}^{\theta_A}\cos^2(x)dx$$

With an integration by part:

$$\int_{-\theta_A}^{\theta_A} \cos^2(x)dx = [\cos(x)\sin(x)]_{-\theta_A}^{\theta_A} - \int_{-\vartheta_A}^{\theta_A} -\sin(x)\sin(x)\,dx = \sin(2\theta_A) + \int_{-\vartheta_A}^{\theta_A} 1 - \cos^2(x)\,dx$$

And so we see that $\int_{-\theta_A}^{\theta_A} \cos^2(x)dx = \frac{\sin(2\theta_A)}{2} + \vartheta_A$.
Bringing:
$\int_{-\theta_A}^{\theta_A} \frac{d\theta}{2\pi}\left(I_{A0} + I_{A1}\cos(\theta)\right)\cos(\theta) = \frac{I_{A0}\sin(\theta_A)}{\pi} + \frac{I_{A1}\theta_A}{2\pi} + I_{A1}\frac{\sin(2\theta_A)}{4\pi}$ (37), again, further simplification here would be misleading for the numerical integrations!

We will also need:
$\int_{-\theta_A}^{\theta_A} \frac{d\theta}{2\pi}\left(I_{A0} + I_{A1}\cos(\theta)\right)\sin(\theta) = 0 + \frac{I_{A1}}{2\pi}\int_{-\theta_A}^{\theta_A} \cos(x)\sin(x)dx = \frac{I_{A1}}{2\pi}\int_{-\theta_A}^{\theta_A} \frac{\sin(2x)}{2}dx = \frac{I_{A1}}{2\pi}\left[-\frac{\cos(2x)}{4}\right]_{-\theta_A}^{\theta_A} = 0$ *(our 38)* (can also be seen from a parity argument)

Next, to obtain the result one need to take into account my first correction, and we will see that they are also modified by our last three integrals.
The authors claim:

$$\begin{aligned} \tau\dot{s}_{A0}(t) &= -s_{A0}(t) + I_{A1}(t)f_0(\theta_A(t)) && (40) \\ \tau\dot{s}_{A1}(t) &= -s_{A1}(t) \\ &\quad + I_{A1}(t)f_1(\theta_A(t))\cos(\phi_A(t) - \psi_A(t)) \\ & && (41) \\ \tau\dot{\psi}_A(t)s_{A1}(t) &= I_{A1}(t)f_1(\theta_A(t))\sin(\phi_A(t) - \psi_A(t)) \\ & && (42) \end{aligned}$$

We have :

$$\tau\dot{s}_{A0}(t) = -s_{A0}(t) + \int_{-\pi}^{\pi} f_A(\theta, t)\frac{d\theta}{2\pi} \qquad (23)$$

So $\tau s_{A0}'(t) = -s_{A0}(t) + \int_{-\pi}^{\pi} \frac{f_A(\theta,t)d\theta}{2\pi} = -s_{A0}(t) + \int_{\phi_A-\theta_A}^{\phi_A+\theta_A} I_{A0}(t) + I_{A1}(t)\cos(\theta - \phi_A(t))d\theta$
We do a change of variable, and can use (36):
$\tau s_{A0}'(t) = -s_{A0}(t) + \int_{-\theta_A}^{\theta_A} I_{A0}(t) + I_{A1}(t)\cos(\theta)d\theta = -s_{A0} + \frac{I_{A0}\theta_A}{\pi} + \frac{I_{A1}\sin(\theta_A)}{\pi}$ *(our 40)*
We can proceed similarly and we thus find:

$$\tau s_{A1}'(t) = -s_{A1}(t) + \int_{-\pi}^{\pi} \frac{f_A(\theta,t)\cos\left(\theta - \psi_A(t)\right)d\theta}{2\pi} = -s_{A1}(t) + \int_{-\theta_A+\phi_A}^{\theta_A+\phi_A} \frac{(I_{A0}(t) + I_{A1}(t)\cos\left(\theta - \phi_A(t)\right))\cos(\theta - \psi_A(t))\,d\theta}{2\pi}$$

Let us compute the second integral:

$$\int_{-\theta_A+\phi_A}^{\theta_A+\phi_A} \frac{I_{A0}(t) + I_{A1}(t)\cos\left(\theta - \phi_A(t)\right))\cos(\theta - \psi_A(t))d\theta}{2\pi} = \int_{-\theta_A}^{\theta_A} \frac{(I_{A0}(t) + I_{A1}(t)\cos(\theta))\cos\left(\theta + \phi_A(t) - \psi_A(t)\right)d\theta}{2\pi}$$

Then we can use $\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$ with a $=\theta$, b $= \phi_A - \psi_A$. The second term will disappear thanks to (our 38), letting:
$\tau s_{A1}'(t) = -s_{A1}(t) + \left(\frac{I_{A0}}{2\pi}\sin(\vartheta_A) + \frac{I_{A1}\theta_A}{2\pi} + I_{A1}\frac{\sin(2\theta_A)}{4\pi}\right)\cos(\varphi_A - \psi_A)$ (our 41)
And similarly:

$$\tau\psi_A'(t)s_{A1}(t) = \int_{-\theta_A}^{\theta_A} \frac{(I_{A0}(t) + I_{A1}(t)\cos(\theta))\sin\left(\theta + \phi_A(t) - \psi_A(t)\right)d\theta}{2\pi}$$

$\tau\psi_A'(t)s_{A1}(t) = \left(\frac{I_{A0}}{2\pi}\sin(\vartheta_A) + \frac{I_{A1}\theta_A}{2\pi} + I_{A1}\frac{\sin(2\theta_A)}{4\pi}\right)\sin(\phi_a - \psi_A)$. (our 42)

It was nice to find this subtilty that was a bit overlooked in the paper, I was not able to replicate the limited devlopment that came next for the perturbation analysis, it seemed to be a bit more challenging than what the author reported.

Simulations

I could replicate most of the simulations and skipped the traveling Bump solutions (3.2) part.
One part that I found challenging was the plot of figure 4. I first tried to solve it using the above derived equations, but failed. I had to come back to the integrals of equations (14) to (16).
The code can be found on github:
https://github.com/PierreOrhan/ReplicationsProjects/tree/master/src/replicate/replicateBrunel2004
and was made in Julia.

I was not really sure how they computed the velocity variance. It is in fact quite difficult to estimate the velocity of the bump from the raster plot, and even more when their is multiple bump evolving at the same time! What I did is that I unrolled it to be in R instead of $[-\pi, \pi]$, and then fitted a line through the unrolled scatter plot.

Additional explorations

In this kind of modeling, the author does not study plasticity, they define a connectivity profile and explore the network dynamics. But we can easily explore a bit more by changing the weight matrix. Notably I increased the frequency of the cos function used to define the weights. As expected I could now see n bump emerging, where n is the frequency. I think that the neurons would then present tuning curves with n folds, unfortunately I did not take the time to obtain it, I could not find a good way to separate the bump and estimate their position…

Conclusions

Despite the few mathematical errors, the analysis was quite deep, and I could successfully replicate most of the paper results. It was interesting to note that all of this is based on the connectivity matrix first chosen. I wonder what would be the plasticity mechanisms that really supports the circuits, but more on that latter in a review which will focus less on one paper details.

References:

[1] Rishidev Chaudhuri et al., « The Intrinsic Attractor Manifold and Population Dynamics of a Canonical Cognitive Circuit across Waking and Sleep », *Nature Neuroscience* 22, nº 9 (septembre 2019): 1512-20, https://doi.org/10.1038/s41593-019-0460-x.

[2] Pengcheng Song et Xiao-Jing Wang, « Angular Path Integration by Moving "Hill of Activity": A Spiking Neuron Model without Recurrent Excitation of the Head-Direction System », *The Journal of Neuroscience* 25, nº 4 (26 janvier 2005): 1002-14, https://doi.org/10.1523/JNEUROSCI.4172-04.2005.

[3] Boucheny, C., Brunel, N. & Arleo, A. A Continuous Attractor Network Model Without Recurrent Excitation: Maintenance and Integration in the Head Direction Cell System. J Comput Neurosci 18, 205–227 (2005). https://doi.org/10.1007/s10827-005-6559-y