

PARAMETRIC ESTIMATION OF SPECTRUM DRIVEN BY AN EXOGENOUS SIGNAL

Tom Dupré la Tour, Yves Grenier, Alexandre Gramfort

Télécom ParisTech, LTCI, CNRS, Université Paris-Saclay, Paris, France

ABSTRACT

In this paper, we introduce new parametric generative driven auto-regressive (DAR) models. DAR models provide a non-linear and non-stationary spectral estimation of a signal, conditionally to another exogenous signal. We detail how inference can be done efficiently while guaranteeing model stability. We show how model comparison and hyper-parameter selection can be done using likelihood estimates. We also point out the limits of DAR models when the exogenous signal contains too high frequencies. Finally, we illustrate how DAR models can be applied on neuro-physiologic signals to characterize phase-amplitude coupling.

Index Terms— non-linear auto-regressive models, non-stationary, spectrum estimation, electrophysiology, phase-amplitude coupling

1. INTRODUCTION

Auto-regressive (AR) models are stochastic signal models that have proved their usefulness in many applications: from speech and audio applications, econometrics or even physiology and biological signal processing. One of their advantages is the existence of fast inference algorithms [1] and to provide a compact representation of the spectral content of a signal (e.g. using 10 to 16 parameters for coding 32 ms of speech).

Standard AR models are so-called stationary, meaning that the statistics of the signal are assumed to be stable over time. When working with such models, the spectrum is therefore not a function of time. In many applications, this modeling assumption is not adapted to describe the interesting dynamics of the physical system observed. This is for example the case in the field of econometrics [2] where time-varying or non-linear AR models were first studied, but it is also the case for physiological signals as it will be illustrated below with a phenomena known as phase-amplitude coupling [3]. The type of signal we propose to study here is illustrated in Fig. 1.

Related non-linear auto-regressive models Various non-linear auto-regressive models have been proposed in audio signal processing or econometrics. Models with a conditional heteroskedasticity (ARCH [4], GARCH [5]) are extremely popular in econometrics, since they are able to model a varying amplitude for the entire spectrum. However, these models

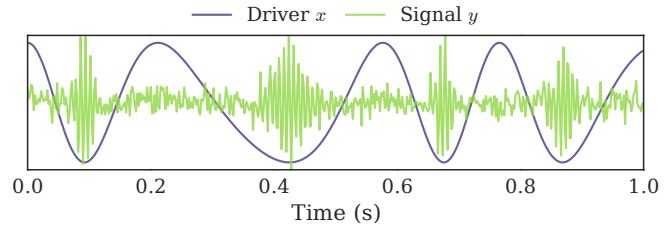


Fig. 1. Simulated signal with time-varying spectrum driven by exogenous signal.

do not model variations in the spectrum itself. To do so, a simple way is to define several regimes: at each time, the instantaneous model is a simple linear AR model, yet the AR coefficients change depending on a *non-linear* function of the past of the signal. The first models based on this idea are the SETAR models [2], which estimate several AR models depending on the amplitude of the signal with respect to some thresholds. To get a smoother transition between regimes, SETAR models have inspired other models like EXPAR [6] or STAR [7], in which the AR coefficients change continuously depending on a non-linear function of the past of the signal. FAR models [8] are even more general, since they estimate the different regimes independently from each other. Motivations behind these models are shared with the DAR models presented in this paper. However, DAR models do not require to infer the driving behavior from the signal itself but rely on the prior knowledge of an exogenous driving signal. As we demonstrate below, this increases both speed and robustness of the estimation, while we guarantee the local stability of the models thanks to a parametrization using *parcor* coefficients.

2. DRIVEN AUTO REGRESSIVE MODELS

2.1. Models definition

An auto-regressive (AR) model specifies that a signal y depends linearly on its own p past values, where p is the *order* of the model:

$$\forall t \in [p+1, T] \quad y(t) + \sum_{i=1}^p a_i y(t-i) = \varepsilon(t) \quad (1)$$

where T is the length of the signal and ε is the *innovation* (or *residual*), modeled with a Gaussian white noise: $\varepsilon(t) \sim$

$\mathcal{N}(0, \sigma(t)^2)$. To extend this AR model to a non-stationary model, one can assume [9, 10] that the AR coefficients a_i are driven by a polynomial function of a given exogenous signal x , here called the *driver*:

$$\forall i \in [1, p] \quad a_i(t) = \sum_{j=0}^m a_{ij} x(t)^j = A_i^\top X(t) \quad (2)$$

where $x(t)^j$ is $x(t)$ to the power j . The choice of the polynomial basis is discussed in section 3.2. Note that A_i is a vector, as the upper-case letter suggests, composed of the scalars a_{ij} . We note \hat{A}_i the estimated value, and A_i^\top the transpose.

As we parametrize the AR coefficients, the estimated models are not guaranteed to be stable. To examine stability, the standard technique is to consider the *parcor* coefficients k_i , introduced in the fast Levinson-Durbin algorithm [1] that estimates the a_i coefficients recursively from previous orders: the a_i at order p (noted $a_i^{(p)}$ in Eq. 3) can be computed from k_p and the AR coefficients at previous order $a_i^{(p-1)}$:

$$a_p^{(p)} = k_p; \quad \forall i \in [1, p-1], \quad a_i^{(p)} = a_i^{(p-1)} + k_p a_{p-i}^{(p-1)} \quad (3)$$

From this so-called *lattice* representation, we can define a second non-stationary parametrization, where the $k_i(t)$ are directly driven by the exogenous signal x , instead of the $a_i(t)$:

$$\forall i \in [1, p] \quad k_i(t) = \sum_{j=0}^m k_{ij} x(t)^j = K_i^\top X(t) \quad (4)$$

This parametrization is used as an intermediate step during model estimation. The lattice representation brings a simple condition (necessary and sufficient) for stability: $-1 < k_i < 1$ [11]. To enforce this condition, we use the log-area ratios (LAR) γ_i [12], as suggested in [13]:

$$\gamma_i = \log \left(\frac{1 + k_i}{1 - k_i} \right) \iff k_i = \frac{e^{\gamma_i} - 1}{e^{\gamma_i} + 1} \quad (5)$$

We use the LAR coefficients $\gamma_i(t)$ in order to force the $k_i(t)$ to be within $]-1, 1[$, which enforces the stability of the instantaneous model for all t . Our third and final non-stationary parametrization is:

$$\forall i \in [1, p] \quad \gamma_i(t) = \sum_{j=0}^m \gamma_{ij} x(t)^j = \Gamma_i^\top X(t) \quad (6)$$

We call this model a driven auto-regressive (DAR) model.

2.2. Model estimation

For DAR models, as the innovation is assumed to be Gaussian white noise, the likelihood L is obtained via:

$$L = \prod_{t=p+1}^T \frac{1}{\sqrt{2\pi}\sigma(t)} \exp \left(-\frac{\varepsilon(t)^2}{2\sigma(t)^2} \right) \quad (7)$$

$$-2 \log(L) = T \log(2\pi) + \sum_{t=p+1}^T \frac{\varepsilon(t)^2}{\sigma(t)^2} + 2 \sum_{t=p+1}^T \log(\sigma(t))$$

To estimate a DAR model, we maximize the likelihood iteratively from order 1 to order p , by adding each time a lattice cell. First, we assume that the innovation variance is constant $\sigma(t)^2 = \sigma^2$ and equal to the signal's empirical variance. So for each lattice cell i from $i = 1$ to p , we estimate the DAR coefficients $\Gamma_i \in \mathbb{R}^{m+1}$ by minimizing a least square criterion over the forward residual:

$$J_1(\Gamma_i) = \sum_{t=i+1}^T \frac{\varepsilon_i^+(t)^2}{\sigma(t)^2} \quad (8)$$

where the forward and backward residuals at the output of the i -th lattice cell are given by:

$$\forall i \in [1, p] \quad \begin{cases} \varepsilon_i^+(t) = \varepsilon_{i-1}^+(t) + k_i(t) \varepsilon_{i-1}^-(t-1) \\ \varepsilon_i^-(t) = \varepsilon_{i-1}^-(t-1) + k_i(t) \varepsilon_{i-1}^+(t) \end{cases} \quad (9)$$

and where we initialize with $\varepsilon_0^+(t) = \varepsilon_0^-(t) = y(t)$. We minimize $J_1(\Gamma_i)$ with a Newton-Raphson procedure since the gradient and Hessian can be computed easily. To start with a good initialization, we approximate the problem with a two-step algorithm. In the first step, we solve the linear problem $\arg\min_{K_i} J_1(K_i)$ using (4), which leads to a *normal equation*:

$$\left(\sum_{t=i+1}^T \frac{\varepsilon_{i-1}^-(t-1)^2}{\sigma(t)^2} X(t) X(t)^\top \right) \hat{K}_i = - \sum_{t=i+1}^T \frac{1}{\sigma(t)^2} \varepsilon_{i-1}^+(t) \varepsilon_{i-1}^-(t-1) X(t) \quad (10)$$

In the second step, we estimate Γ_i with a regression over the trajectories $\gamma_i(t)$ and $\hat{k}_i(t)$: After clipping $\hat{k}_i(t) = \hat{K}_i^\top X(t)$ inside $[-1 + \eta, 1 - \eta]$ (e.g. $\eta = 10^{-6}$), we measure the quality of the approximation with a second least-squares criterion:

$$J_2(\Gamma_i) = \sum_{t=i+1}^T \left(\Gamma_i^\top X(t) - \log \left(\frac{1 + \hat{k}_i(t)}{1 - \hat{k}_i(t)} \right) \right)^2 \quad (11)$$

Again this leads to a normal equation:

$$\left(\sum_{t=i+1}^T X(t) X(t)^\top \right) \hat{\Gamma}_i = \sum_{t=i+1}^T X(t) \log \left(\frac{1 + \hat{k}_i(t)}{1 - \hat{k}_i(t)} \right)$$

These two steps are used to initialize the Newton-Raphson procedure, and largely speed-up the optimization. We iterate this over each lattice cell i : For $i > 1$, the residuals $\varepsilon_i^+(t)$ and $\varepsilon_i^-(t)$ are obtained with (9) using the $\hat{k}_i(t)$ derived from $\hat{\gamma}_i(t) = \hat{\Gamma}_i^\top X(t)$ using (5). We then obtain the DAR coefficients Γ_i for every order i from 1 to p .

2.3. Innovation variance estimation

During inference of the DAR coefficients Γ_i , we first assumed a fixed innovation variance σ^2 . We propose to model it as

driven by x , as in [9]:

$$\forall t \in [p+1, T] \quad \log(\sigma(t)) = \sum_{j=0}^m b_j x(t)^j = B^\top X(t) \quad (12)$$

The vector $B \in \mathbb{R}^{m+1}$ is estimated from the residual $\varepsilon_p^+(t)$ by maximum likelihood also using a Newton-Raphson procedure. We then iterate between the estimation of the Γ_i and of B . We observed in our experiments that two iterations are generally sufficient.

2.4. Power spectral density

After model estimation, we can compute the *conditional* power spectral density (PSD) of the model S_y : For a given driver's value x_0 , we compute the AR coefficients $a_i(x_0)$ from the DAR coefficients Γ_i using (6), (5) and (3), along with the innovation's standard deviation $\sigma(x_0)$ using (12). Since AR models with time-varying coefficients are locally stationary [14], we can compute the PSD for this driver's value :

$$S_y(x_0)(f) = \left| \sum_{i=0}^p \frac{a_i(x_0)}{\sigma(x_0)} e^{-j2\pi f i} \right|^{-2} \quad (13)$$

where $j^2 = -1$ and $a_0(x_0) = 1$. Fig. 2 shows an example of $S_y(x)$ for the observed range of value of the driver x (for robustness, we only use the value of x between the 5 and 95 percentiles of all x values) (the signal used in this estimation will be detailed in section 4). Using the fast Fourier transform (FFT) in (13), we estimate the PSD for a range of frequency $f \in [0, f_s/2]$. As model estimation is also very fast, we obtain the PSD very quickly: With $(p, m) = (10, 2)$, on a signal with $T = 10^6$ time points, the entire DAR estimation lasts 0.22 s.

3. MODEL AND HYPER-PARAMETER SELECTION

3.1. BIC selection

DAR models are probabilistic, thus we can define a likelihood function L (7), and perform model and hyper-parameter selection by maximizing it. To compare models with different number of hyper-parameters (i.e. degrees of freedom), we can also use the well known bayesian information criterion (BIC) [15], which reads $BIC = -2 \log(L) + d \log(T)$, where d is the number of degrees of freedom. In DAR models, we have $d = (p+1)(m+1)$. As model estimation is very fast, we can easily estimate on a full grid of hyper-parameters (p and m) and keep the model that leads to the lowest BIC.

Simulations Using simulated data, we tested the hyper-parameters selection based on the BIC. We generated a driving signal x for $T = 10^4$ time points, by re-sampling a Gaussian white noise of length $T f_x / f_s$ (where $f_s = 10^3$ Hz is the

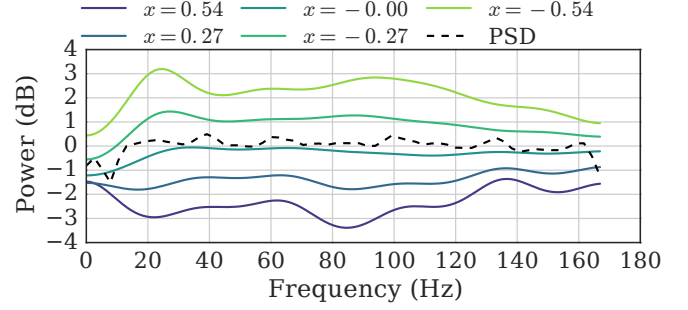


Fig. 2. Power spectral density (PSD) evaluated through a DAR model, for different driver's values x . In dashed black, we plot the PSD estimated with a Welch method over the entire signal. Note that the PSD is mostly flat since the signal was globally whitened as preprocessing, cf. 4 .

sampling frequency). We then drew a set of LAR coefficients $\gamma_i(t)$ using random walks: starting from a real number $\gamma_i(0) \in]-1, 1[$, we computed $\gamma_i(t+1) = \gamma_i(t) + \varepsilon_i(t)$ with $\varepsilon_i(t) \sim \mathcal{N}(0, 0.3)$. Note that the LAR coefficients do not need to be in $] -1, 1[$ to make the system stable (cf. section 2.1). However, to avoid having very large coefficients during a long period of time, we rescaled the trajectory to be inside $[-4, 4]$. To make the LAR coefficients dependent of the driver x , we projected them on the basis $X(t) = [1 \ x(t) \ \dots \ x(t)^m]^\top$:

$$\hat{\Gamma}_i = \underset{\Gamma_i \in \mathbb{R}^{m+1}}{\operatorname{argmin}} \sum_{t=1}^T (\Gamma_i^\top X(t) - \gamma_i(t))^2 \quad (14)$$

With the obtained DAR coefficients $\hat{\Gamma}_i$, we computed the instantaneous LAR coefficients $\hat{\gamma}_i(t)$, and generated a signal $y(t)$ by feeding the corresponding lattice filters defined by $k_i(t)$ with a Gaussian white noise. To focus on the spectral fluctuations, we used a constant innovation gain: $\sigma(t) = \sigma$.

We then fitted several DAR models on y , with $p \in [1, 20]$ and $m \in [0, 3]$, and compared the BIC of each fit. We tested if the BIC selects the correct hyper-parameters, and we tested it for different driver's frequency f_x . The results are presented in Fig. 3 : Interestingly, the BIC selection of m is correct in most cases, provided that the driver's frequency is not too high ($f_x < 50$ Hz). When the driver's frequency is too low, the BIC sometimes overestimate m since the time length is too short to see many driver's oscillations. The BIC selects p correctly in most cases at ± 2 (not shown), for all driver's frequencies.

3.2. Comparison with STAR models

The choice of a polynomial basis in (6) and (12) is motivated by simplicity, and practice also shows that a low order (≤ 3) is enough as stated by a BIC selection. For the sake of comparison, we also estimated a multi-regime LSTAR [16]:

$$\forall i \in [1, p] \quad a_i(t) = \sum_{j=0}^m a_{ij} F_j(x(t)) \quad (15)$$

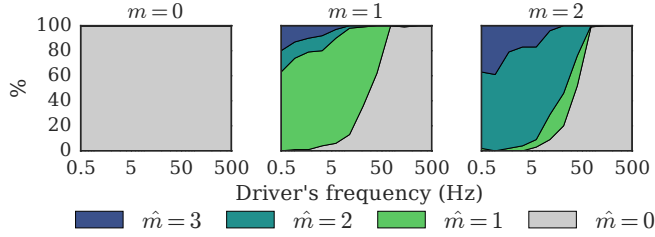


Fig. 3. Selection of hyper-parameter m in simulated data with respect to the driver's frequency. We simulated 100 signals from 100 DAR models with $p = 10$ and $m = 0$ (left), $m = 1$ (middle), $m = 2$ (right). We then estimated some DAR models on these signals, and we selected \hat{p} and \hat{m} that minimized the BIC. The graphs show the proportion of each \hat{m} selected. The hyper-parameter m is correctly estimated in most cases if the driver's frequency is not too high ($f_x < 50$ Hz).

which is parametrized with sigmoid functions:

$$\forall j \in [1, m] \quad F_j(x(t)) = (1 + e^{-\gamma_j(x(t) - c_j)})^{-1} \quad (16)$$

and with $F_0(x(t)) = 1$. Note that to fairly compare the models, we use the same exogenous signal x to drive the AR coefficients, when standard LSTAR models use the delayed signal $y(t - d)$, which results in a worse fit (cf. Fig 4). We also extended them to model the innovation variance $\sigma(t)^2$ with sigmoids:

$$\forall t \in [p + 1, T] \quad \log(\sigma(t)) = \sum_{j=0}^m b_j F_j(x(t)) \quad (17)$$

The results are presented in Fig. 4, and show that the polynomial and the sigmoid parametrizations achieve comparable performances. However, the optimization of the LSTAR is much slower, since there is no good initialization procedure for the transitions parameters γ_j and c_j : the non-linear optimization is slow and may contain several local minima, and the cost of an initial grid-search is large, or even prohibitive as soon as $m \geq 2$. On the contrary, the DAR models estimation is very fast and does not suffer from multiple minima.

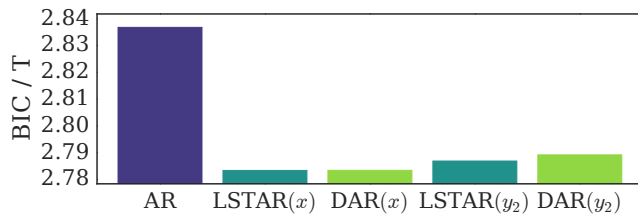


Fig. 4. BIC by time point for different models and drivers (lower is better), using signals of section 4. The signal x is a better driver than the best delayed raw signal $z_2(t) = z(t - 2)$. The hyper-parameters (p, m) , chosen through a BIC selection, are respectively: $(18, \emptyset)$, $(10, 1)$, $(10, 2)$, $(14, 2)$, $(10, 2)$.

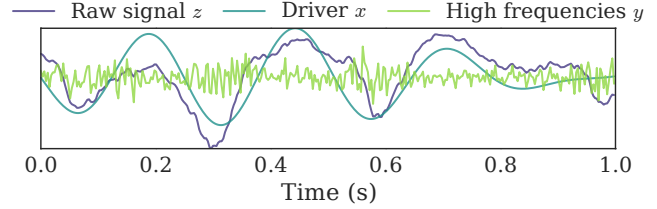


Fig. 5. The driver x is extracted from the raw signal z with a band-pass filter. The remaining high frequencies are whitened to form the modeled signal y .

4. APPLICATION TO NEUROSCIENCE

In neuroscience, phase-amplitude coupling (PAC) refers to the interaction between the phase of a slow neural oscillation and the amplitude of high frequencies within the same signal or at a distinct brain location. High frequencies driven by slow signal fluctuations have been observed in animal [17] and human [3] studies and have been reported as a key element for neural communication during complex cognitive processes [18, 19]. Yet, PAC is usually measured through empirical and somehow *ad hoc* metrics [3, 20]. To give a proper model to PAC, we applied DAR models on a human electro-corticogram (ECoG) channel from [3] (730 seconds at 333 Hz).

After decimation and removal of the electrical network noise, we extracted the driver x with a zero-phase bandpass filter ($f_x = 3.8$ Hz, $\Delta f_x = 1$ Hz). We removed a wider bandpass signal around it ($\Delta f = 4$ Hz) to prevent spurious correlations between x and y , and we filled the gap with white noise filtered with the same filter. We then whitened y with a simple AR model, by applying the inverse AR filter to the signal. This whitening step is not necessary, yet it reduces the need of a high order p in DAR, which reduces both the computational cost and the variance of the model. After an exhaustive grid search with $p \in [1, 50]$ and $m \in [0, 4]$, the hyper-parameters $(p, m) = (10, 2)$ are selected by BIC, hence revealing the phase-amplitude coupling present in the data.

Fig. 5 shows a portion of the signals, while Fig. 2 shows the PSD (in dB) depending on the driver's value, estimated through a DAR model. PAC can be identified in the difference of $S_y(x)$ as the driver x varies: the PSD has more power for negative driver's values than for positive driver's values, and PSD shapes are also different.

5. CONCLUSION

In this paper, we introduced driven auto-regressive (DAR) models, to provide a parametric spectrum estimation of a signal, conditionally to another exogenous driving signal. The chosen parametrization guarantees local stability, and allows fast inference and easy hyper-parameter selection. We also showed how DAR models can be used to characterize phase-amplitude coupling on neuro-physiologic signals.

6. REFERENCES

- [1] James Durbin, "The fitting of time-series models," *Review of the Int. statistical institute*, pp. 233–244, 1960.
- [2] Howell Tong and Keng S Lim, "Threshold autoregression, limit cycles and cyclical data," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 245–292, 1980.
- [3] Ryan T Canolty, Erik Edwards, Sarang S Dalal, Maryam Soltani, Srikantan S Nagarajan, Heidi E Kirsch, Mitchel S Berger, Nicholas M Barbaro, and Robert T Knight, "High gamma power is phase-locked to theta oscillations in human neocortex," *Science*, vol. 313, no. 5793, pp. 1626–1628, 2006.
- [4] Robert F Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation," *Econometrica: Journal of the Econometric Society*, pp. 987–1007, 1982.
- [5] Tim Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of econometrics*, vol. 31, no. 3, pp. 307–327, 1986.
- [6] Valérie Haggan and Tohru Ozaki, "Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model," *Biometrika*, vol. 68, no. 1, pp. 189–196, 1981.
- [7] Kung Sik Chan and Howell Tong, "On estimating thresholds in autoregressive models," *Journal of Time Series Analysis*, vol. 7, no. 3, pp. 179–190, 1986.
- [8] Rong Chen and Ruey S Tsay, "Functional-coefficient autoregressive models," *Journal of the American Statistical Association*, vol. 88, no. 421, pp. 298–308, 1993.
- [9] Yves Grenier, "Estimating an AR Model with Exogenous Driver," Tech. Rep. 2013D007, Telecom ParisTech, 2013.
- [10] Yves Grenier, "Time-dependent ARMA modeling of nonstationary signals," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 31, no. 4, pp. 899–911, 1983.
- [11] Jacob Benesty, M Mohan Sondhi, and Yiteng Huang, *Springer handbook of speech processing*, Springer, 2007.
- [12] H. Wakita, "Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms," *IEEE Transactions on Audio and Electroacoustics*, vol. 21, no. 5, pp. 417–427, Oct 1973.
- [13] Yves Grenier and Marie-Christine Omnes-Chevalier, "Autoregressive models with time-dependent log area ratios," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 36, no. 10, pp. 1602–1612, 1988.
- [14] Rainer Dahlhaus, "On the kullback-leibler information divergence of locally stationary processes," *Stochastic Processes and their Applications*, vol. 62, no. 1, pp. 139–168, 1996.
- [15] Gideon Schwarz et al., "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, 1978.
- [16] Dick van Dijk, Timo Teräsvirta, and Philip Hans Franses, "Smooth transition autoregressive models—a survey of recent developments," *Econometric reviews*, vol. 21, no. 1, pp. 1–47, 2002.
- [17] Anatol Bragin, Gábor Jandó, Zoltán Nádasdy, Jámille Hetke, K Wise, and Gy Buzsáki, "Gamma (40–100 Hz) oscillation in the hippocampus of the behaving rat," *The Journal of Neuroscience*, vol. 15, no. 1, pp. 47–60, 1995.
- [18] Alexandre Hyafil, Anne-Lise Giraud, Lorenzo Fontolan, and Boris Gutkin, "Neural Cross-Frequency Coupling: Connecting Architectures, Mechanisms, and Functions," *Trends in Neurosciences*, vol. 38, no. 11, pp. 725–740, 2015.
- [19] John E Lisman and Ole Jensen, "The theta-gamma neural code," *Neuron*, vol. 77, no. 6, pp. 1002–1016, 2013.
- [20] Adriano BL Tort, Robert Komorowski, Howard Eichenbaum, and Nancy Kopell, "Measuring phase-amplitude coupling between neuronal oscillations of different frequencies," *J. Neurophysiol.*, vol. 104, no. 2, pp. 1195–1210, 2010.