DriPP: Driven Point Processes to Model Stimuli Induced Patterns in M/EEG Signals

How a visual flash affects your brain? A statistical model based on point processes.

Cédric Allain

PhD student, Inria Saclay - Mind Team

Advisors: Alexandre Gramfort & Thomas Moreau

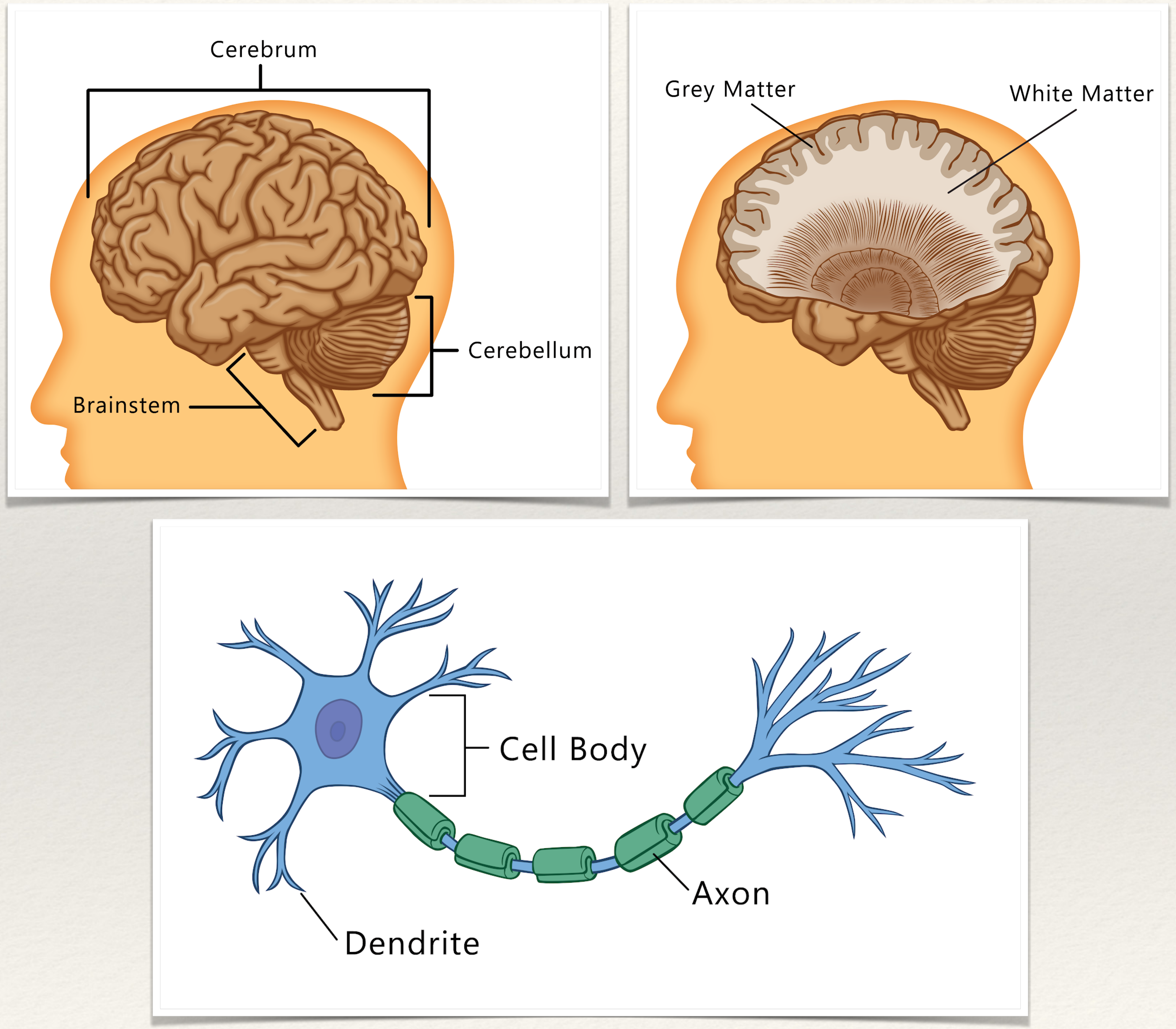
1st December 2022

From the M/EEG recording to DriPP

- Paradigm of a M/EEG experiment
- Convolutional Dictionary Learning for recurrent patterns identification in M/EEG signals
- Background on Point Processes
- Methodology of DriPP
- Results on synthetic and real datasets

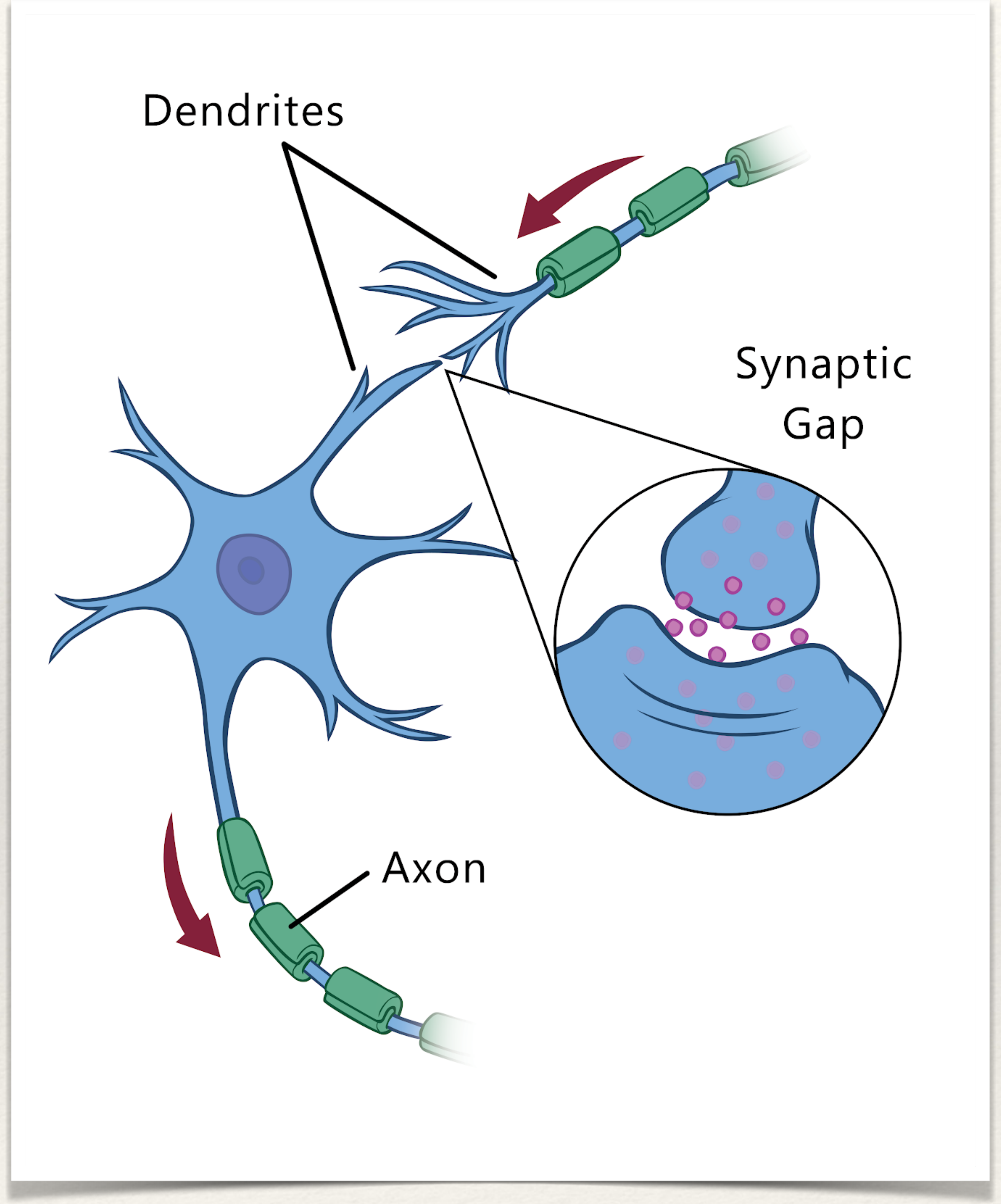
Paradigm of a M/EEG experiment

- Brain constitution:
  60% white matter, 40% grey matter
- **Grey matter**: in the brain, forms the cortex and is composed mainly of neuronal cell bodies and synapses
- **White matter**: found in deeper regions, composed of bundles of axons and connects one part of the brain to another



Composition of the brain and scheme of a neuron
Source: thepartnershipineducation.com

Paradigm of a M/EEG experiment



Synaptic interaction
Source: [thepartnershipineducation.com]

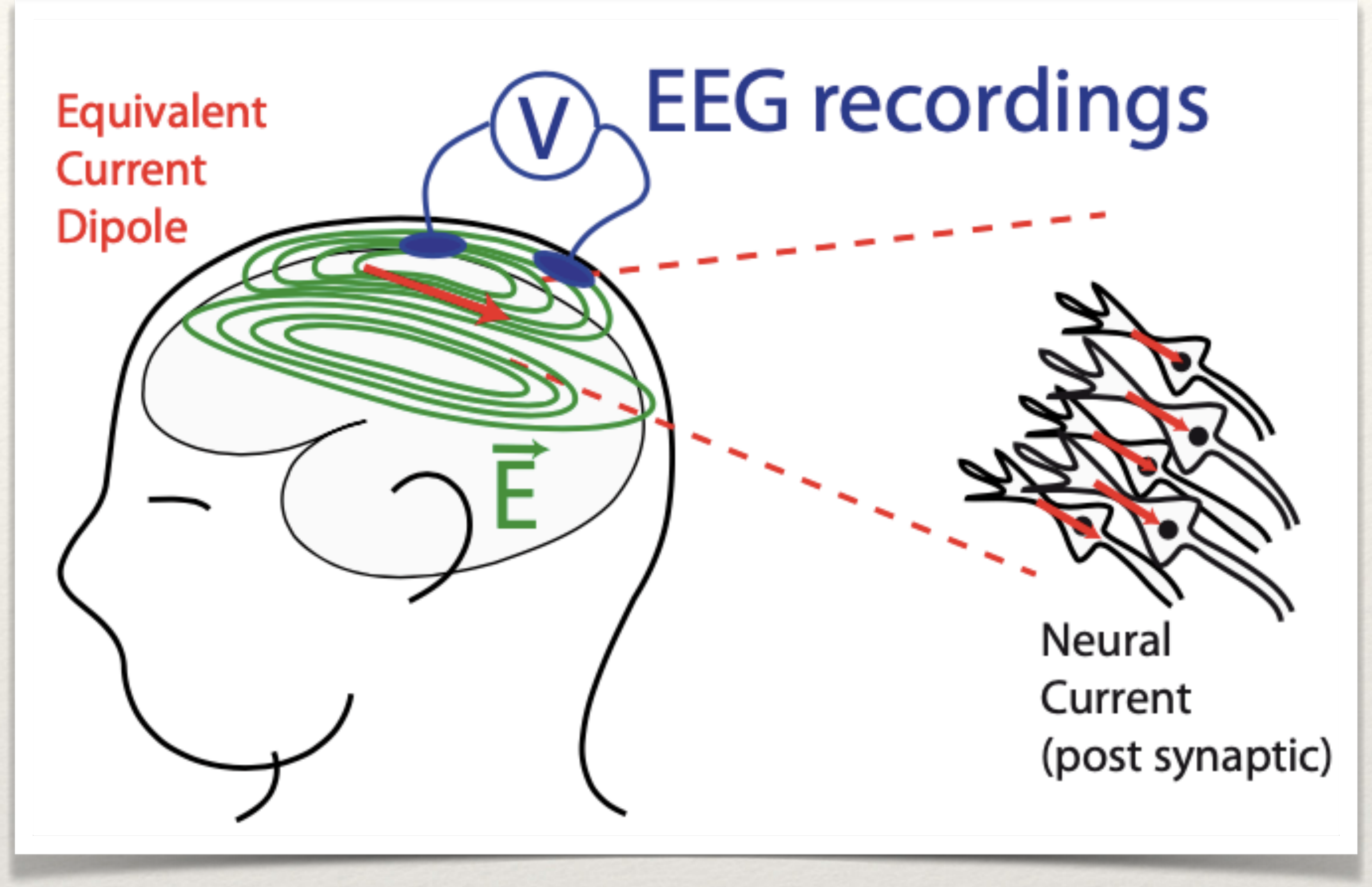
- The reception of neurotransmitters from another neuron at the synaptic gap triggers an electrical signal that travels along the axon.
- The magneto and electro signals derive from the net effect of ionic currents flowing in the dendrites of neurons during synaptic transmission.

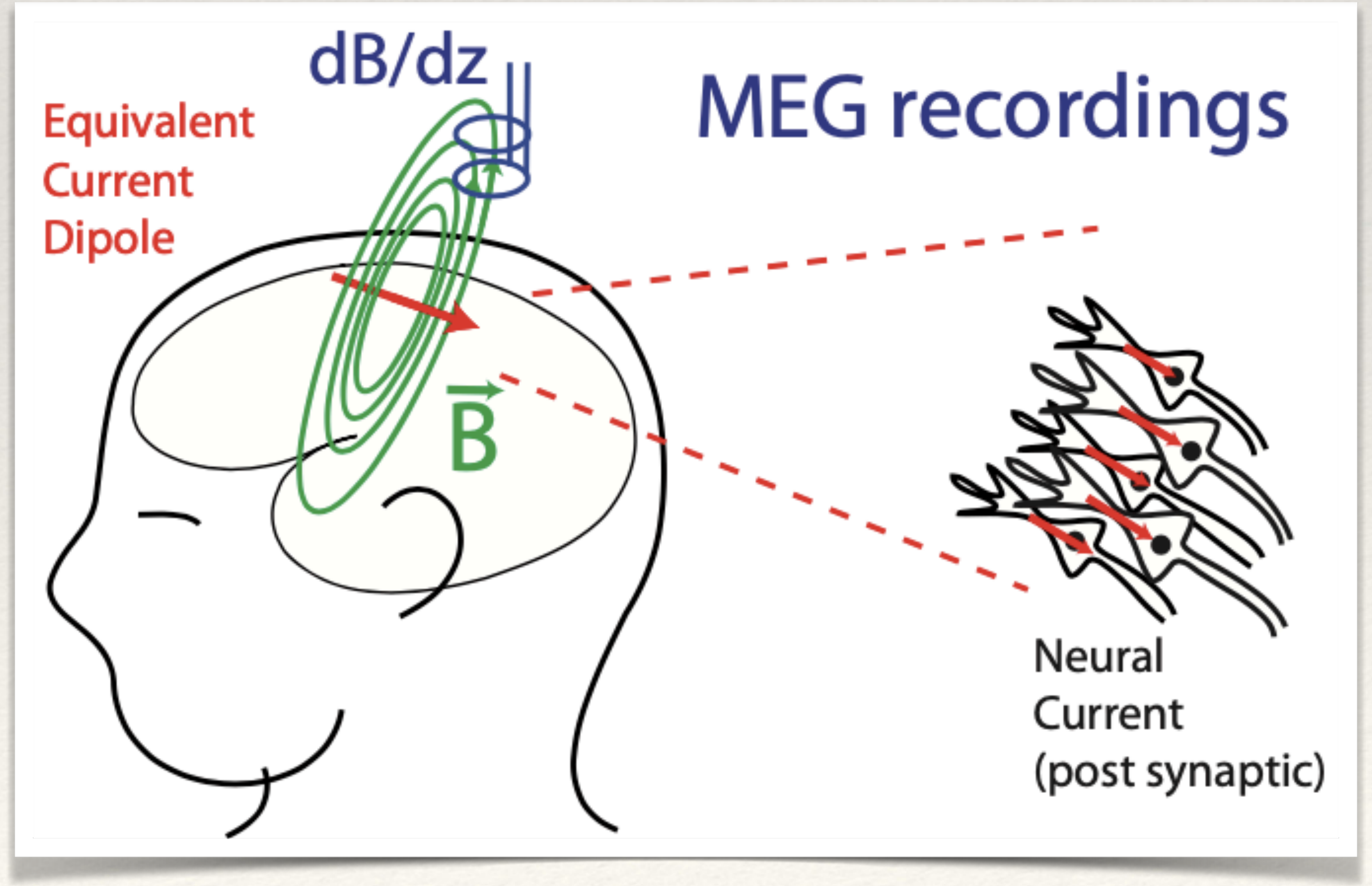
Paradigm of a M/EEG experiment

- A group of neurons in the gray matter form a current generator that produces an electrical and a magnetic field.
- A large number of simultaneously active neurons (~ 50 000) are needed to generate a measurable M/EEG signal [2].
- **Magneto- and electro-encephalography (M/EEG)**: non-invasive recording methods to capture magnetic & electric fields produced by brain neurons, at a high temporal resolution (often between 250 and 1000 Hz).
- **Non-invasive**: unlike ectrocorticography (ECoG) that uses electrodes placed directly on the exposed surface of the brain.

[1] Alexandre Gramfort. Lecture notes in M/EEG: Functional brain imaging with MEG, EEG and sEEG

[2] Saskia Helbling. Lecture notes in SPM course: What are we measuring with M/EEG?, May 2014.

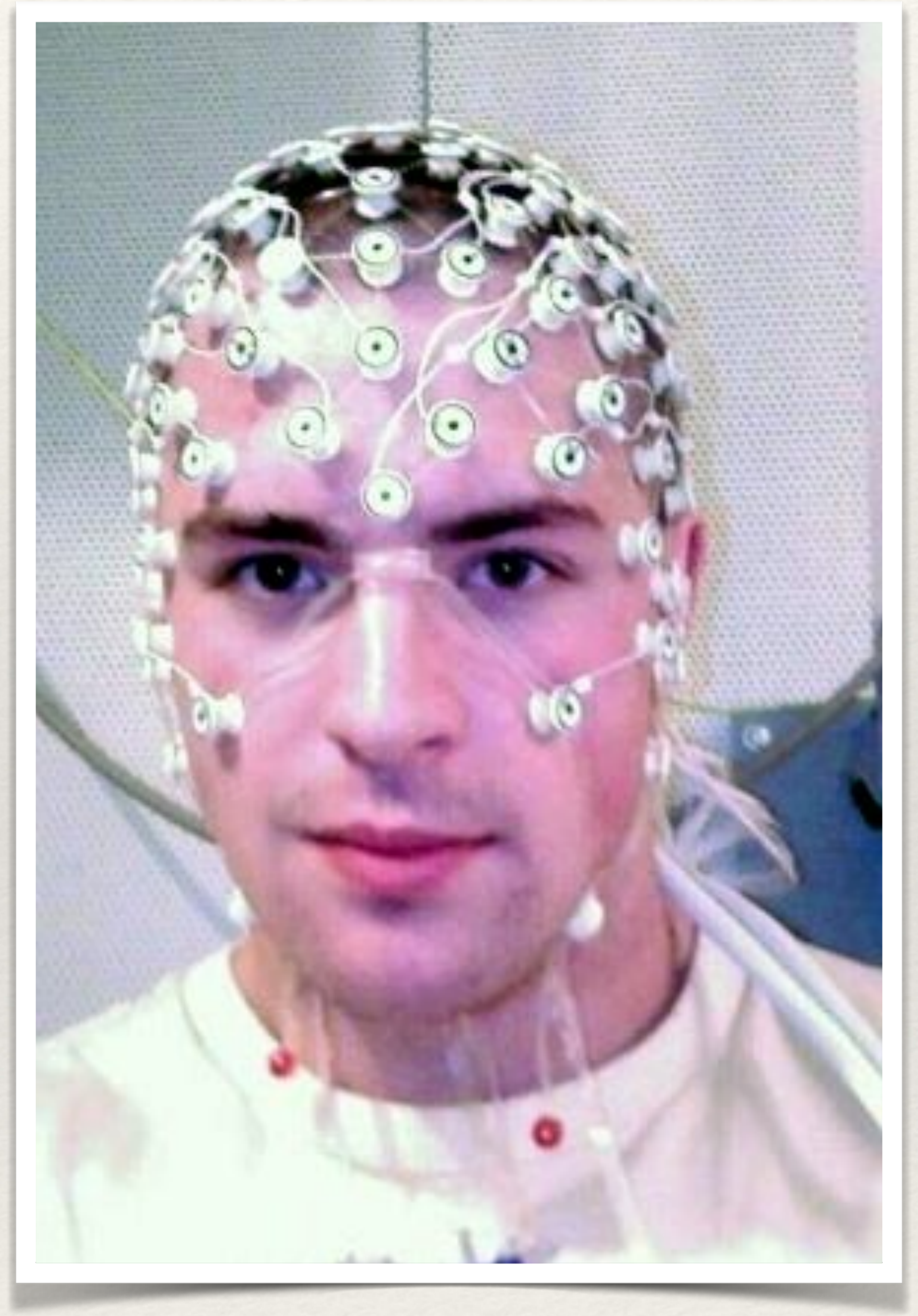




Source: [1]

Paradigm of a M/EEG experiment

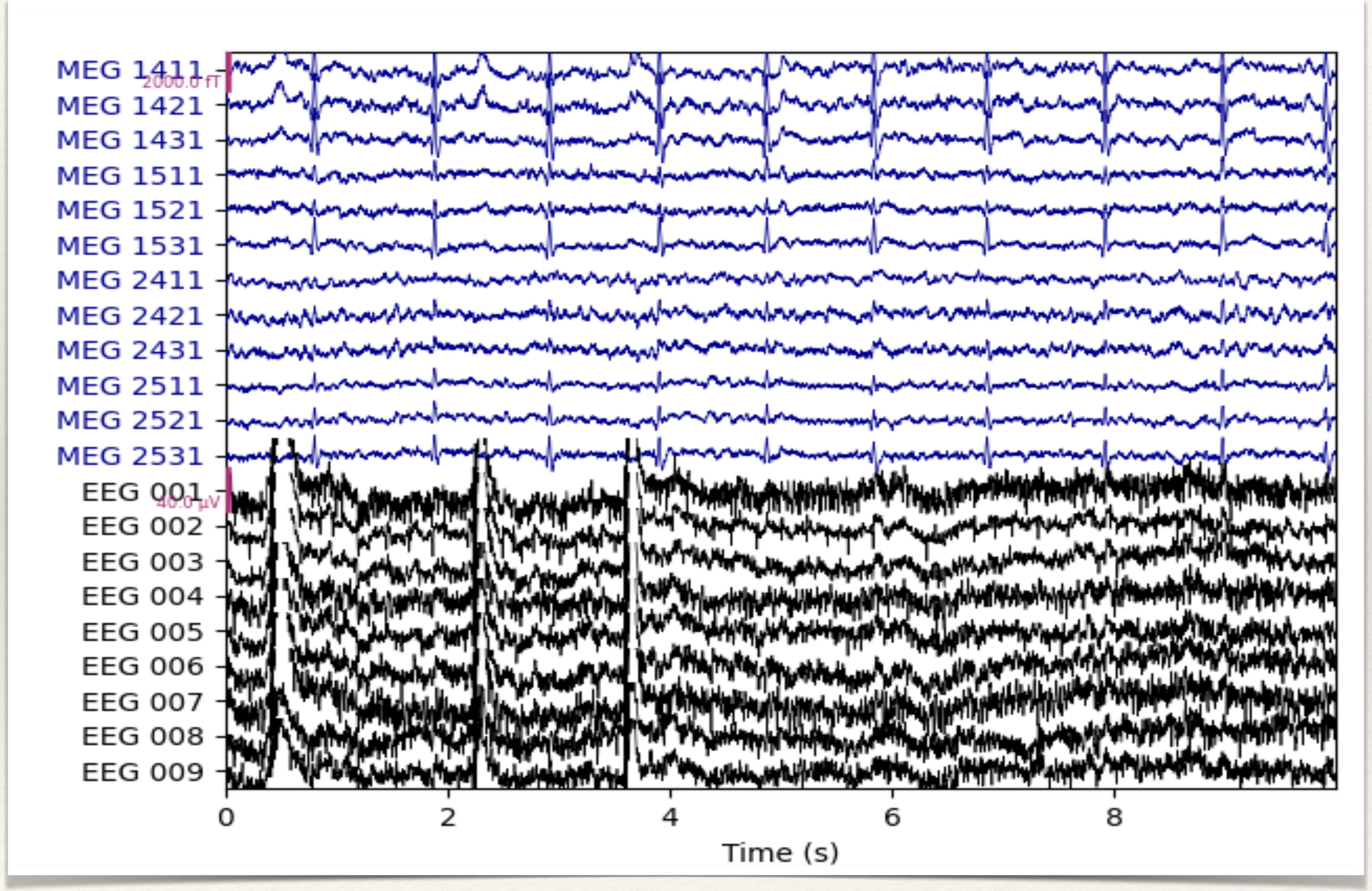
- M/EEG high temporal resolution: attractive for functional study of the brain
- Rather poor spatial resolution: only a few hundred data positions can be acquired simultaneously (MEG: ~ 300/400 sensors; EEG: up to 250 electrodes)
- Recordings last from few minutes (active experiment) to several hours (sleep stages analysis).
- During a task experiment, some (external) stimuli are presented to the subject (auditory and/or visual signals, somatosensory, etc.) and active tasks may be asked (*e.g.*, cues to press a button)



An EEG recording setup

Person undergoing an MEG

Data from a M/EEG experiment

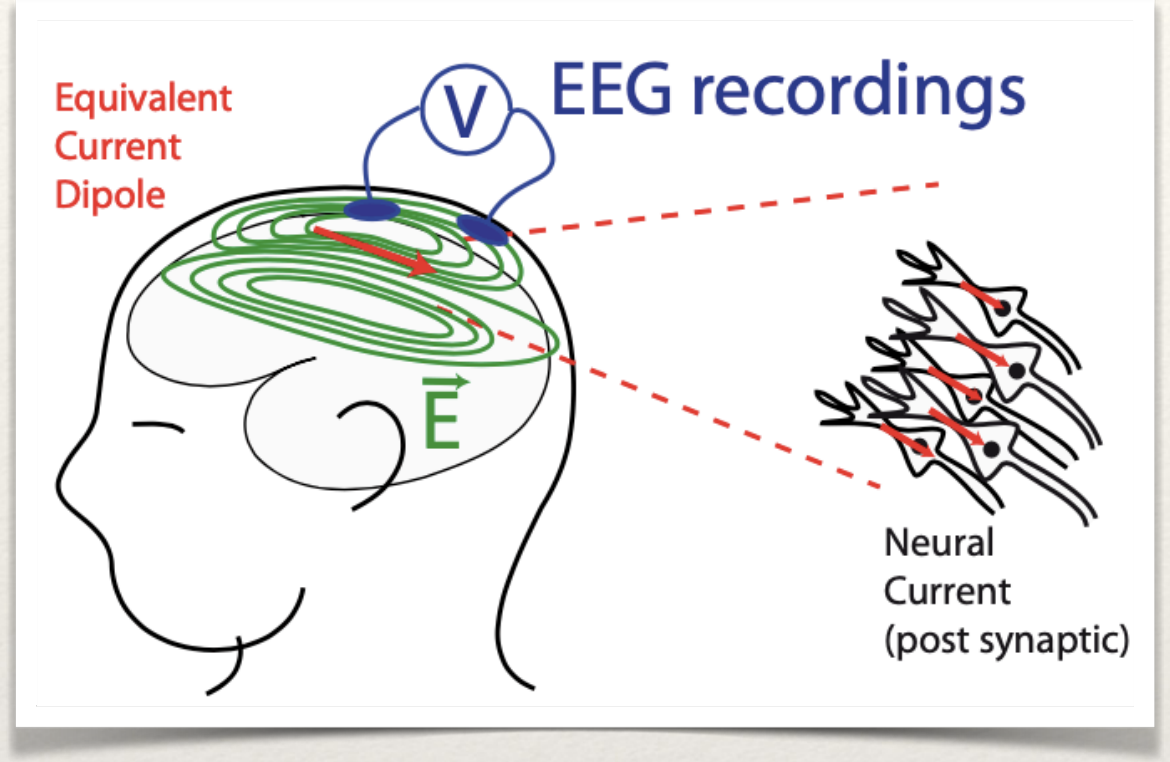


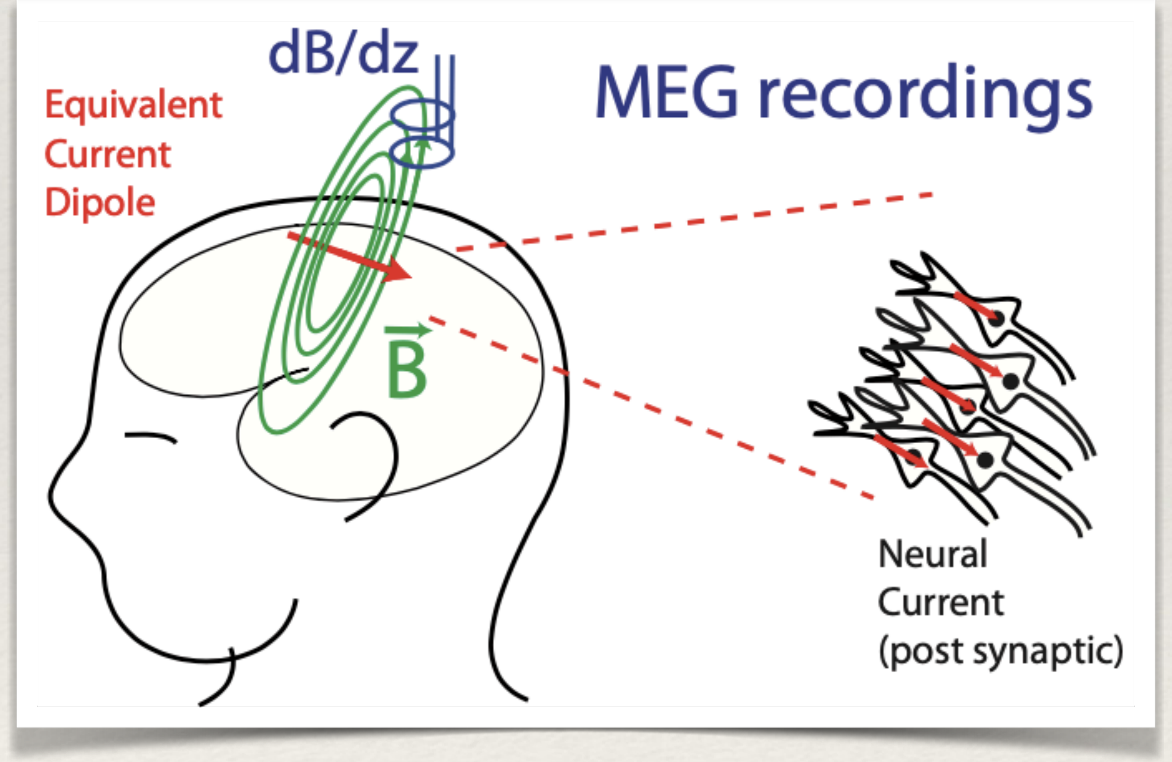
Raw signals over some M/EEG sensors

Source: MNE sample dataset

- Observed signal: $X \in \mathbb{R}^{P \times T}$ with P sensors over T timestamps
- Raw signals are too long and noisy (heavy noise bursts, low signal-to-noise ratio) to be directly analyzed
- Some artifact may be present in the signals, thus corrupting the data (here, clear presence of the heartbeat artifact)
- Pre-processing is necessary
- Artifact can be manually identified and removed using ICA

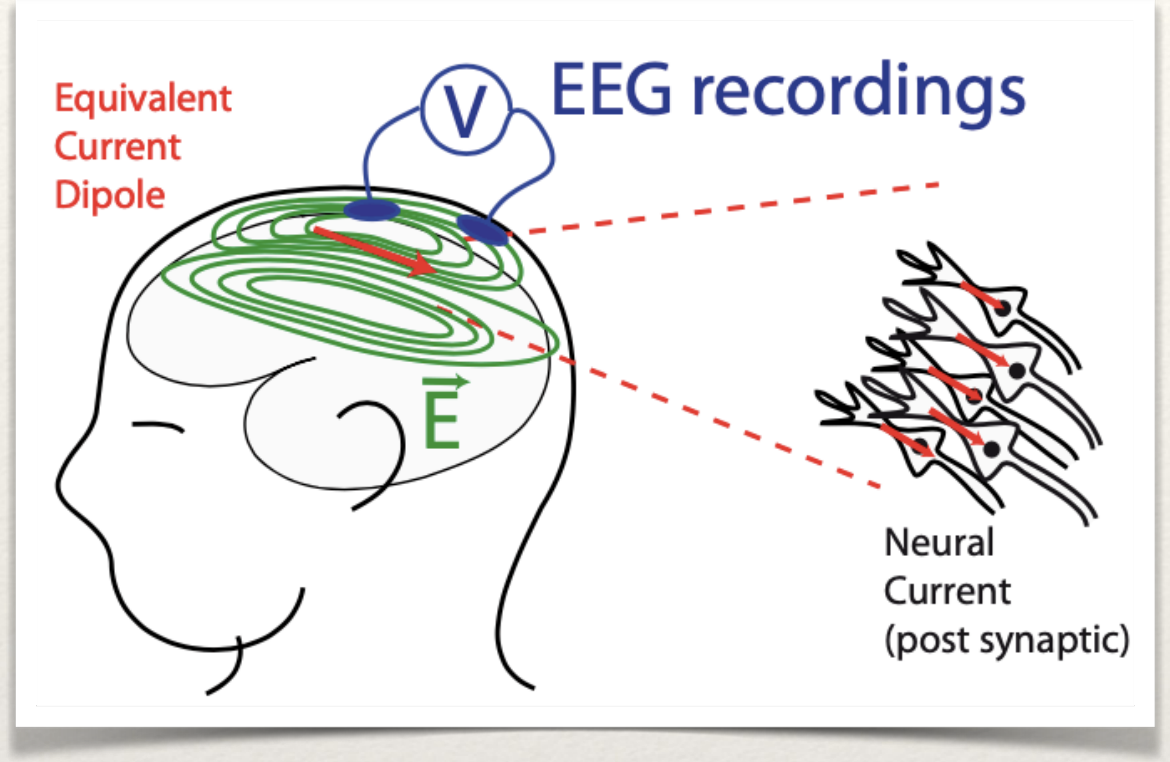
Paradigm of a M/EEG experiment

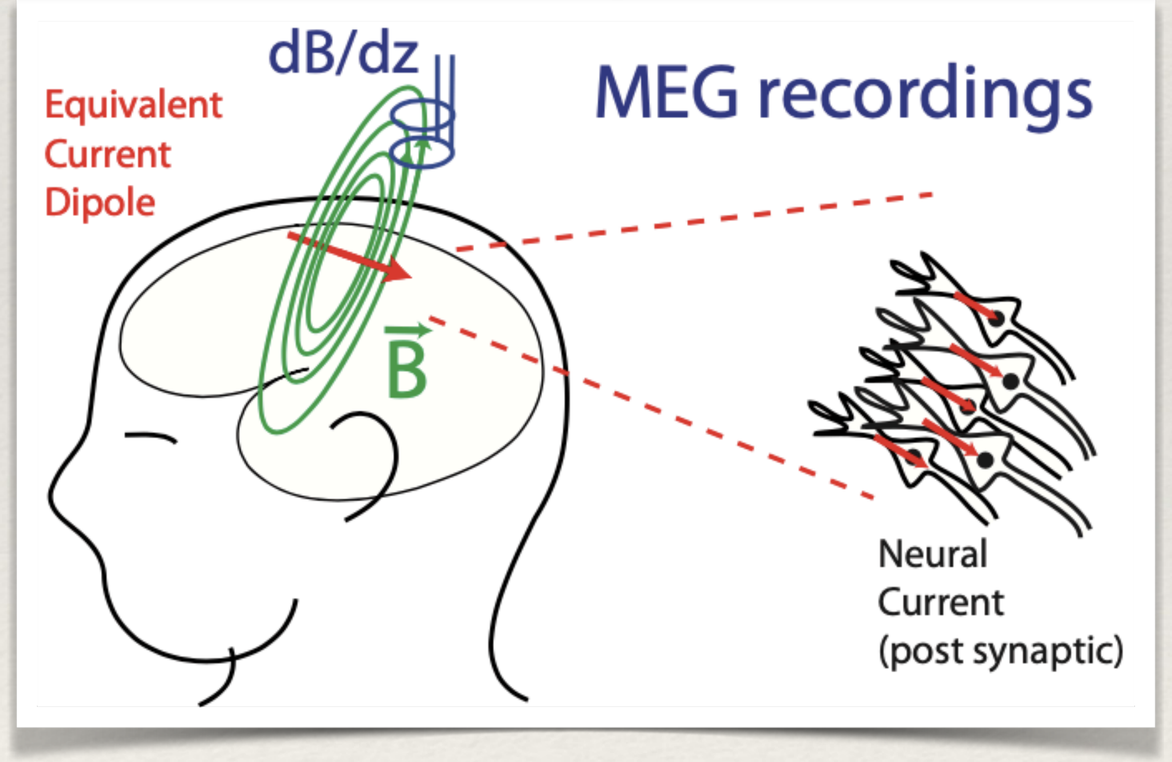




Stimuli: visual, auditory, somatosensory, etc.
Task: attend / ignore, detect + react, etc.

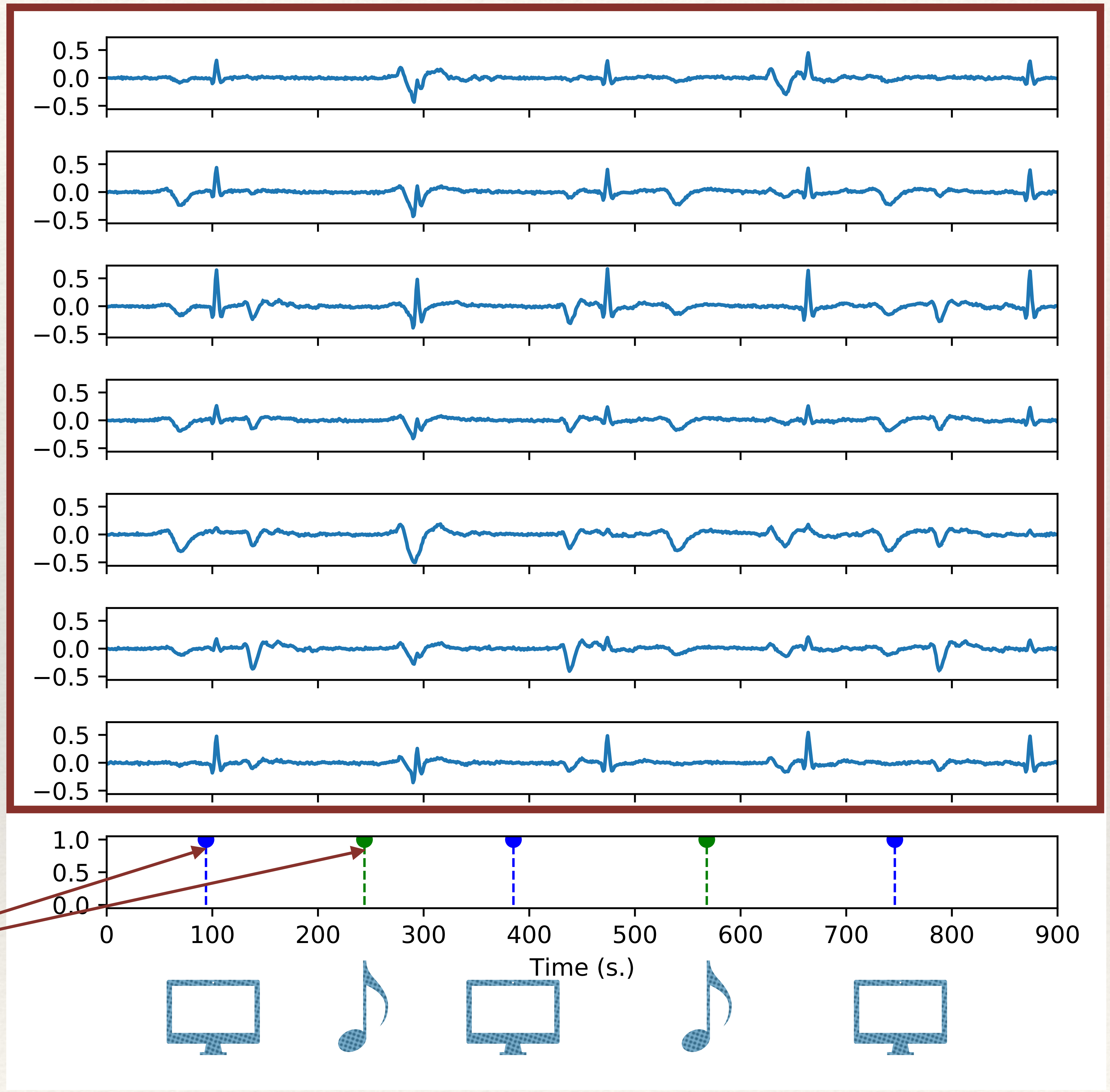
Paradigm of a M/EEG experiment





MEG recording

Sensors
(~300)



Stimuli events

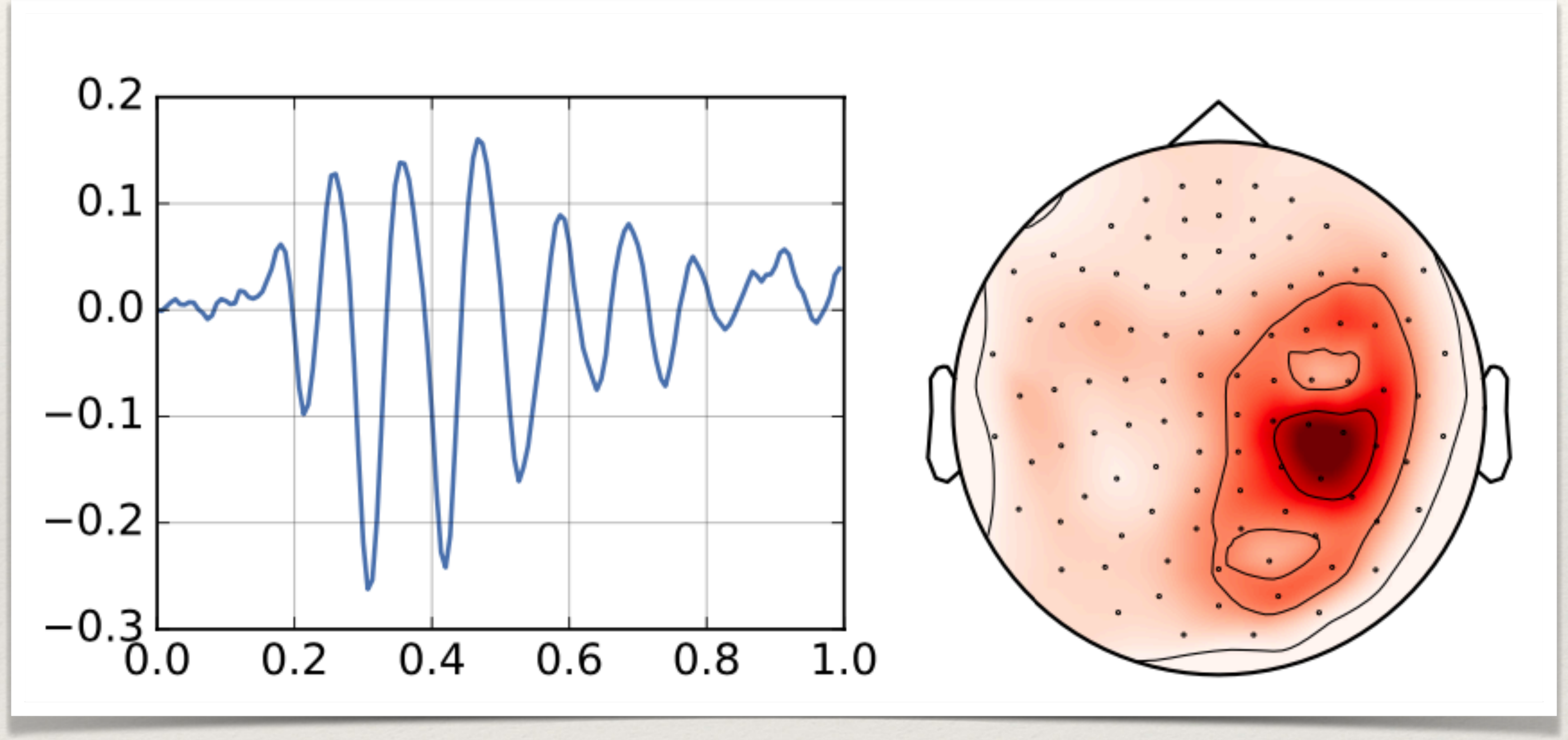
Stimuli: visual, auditory, somatosensory, etc.
Task: attend / ignore, detect + react, etc.

Multiple stimuli presented to the subject

From the M/EEG recording to DriPP

- Paradigm of a M/EEG experiment
- Convolutional Dictionary Learning for recurrent patterns identification in M/EEG signals
- Background on Point Processes
- Methodology of DriPP and results on synthetic and real datasets

Identify recurring patterns

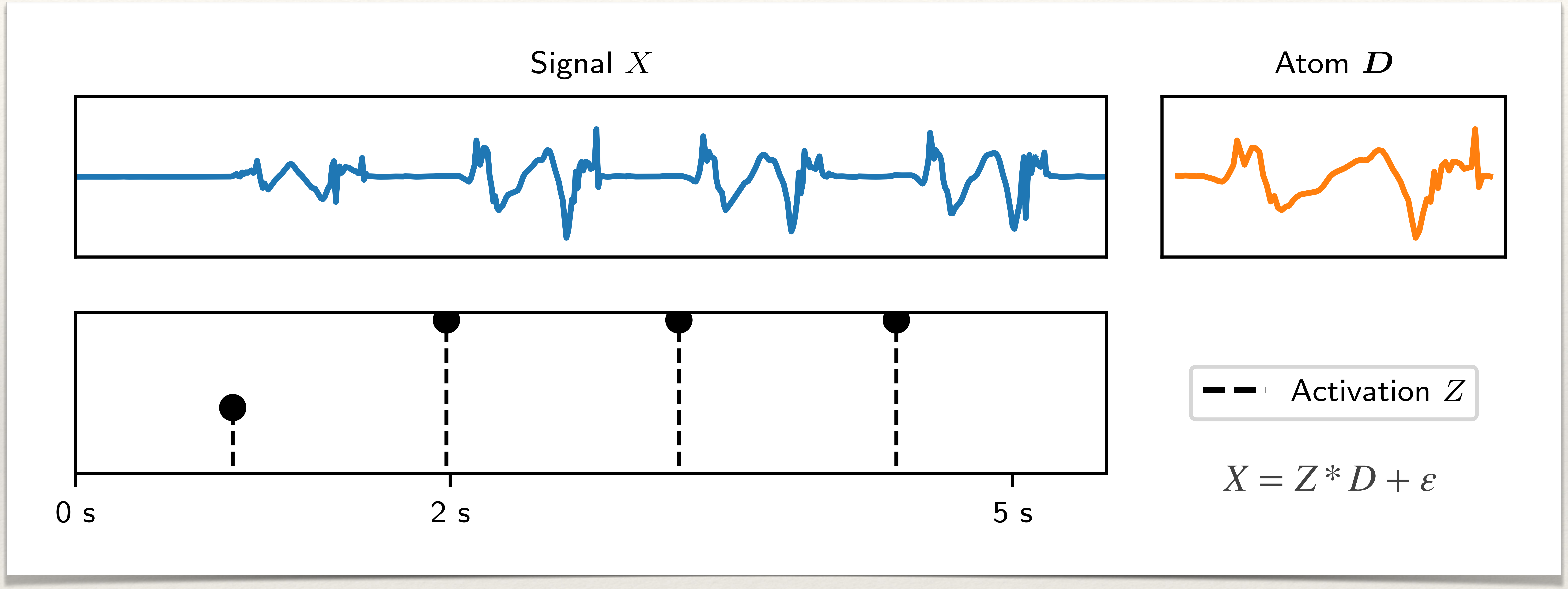


Spatial and temporal pattern learned using CDL with python `alphacsc` *package [1] on MNE somato dataset*

- Neural time-series data contain a wide variety of prototypical signal waveforms
- Using *dictionary learning*, neural signals are decomposed as combinations of time-invariant patterns, called *atoms*.
- Applied to neuroscience, atoms have one spatial representation (which sensors are the most activated) and one temporal representation (the temporal form of the signal)

[1] Dupré La Tour, T., Moreau, T., Jas, M., & Gramfort, A. (2018). Multivariate Convolutional Sparse Coding for Electromagnetic Brain Signals. Advances in Neural Information Processing Systems (NIPS).

Principle of Convolutional Dictionary Learning



Decomposition of a noiseless univariate signal X (blue) as the convolution $Z * D$ between a temporal pattern D (orange) and a sparse activation signal Z

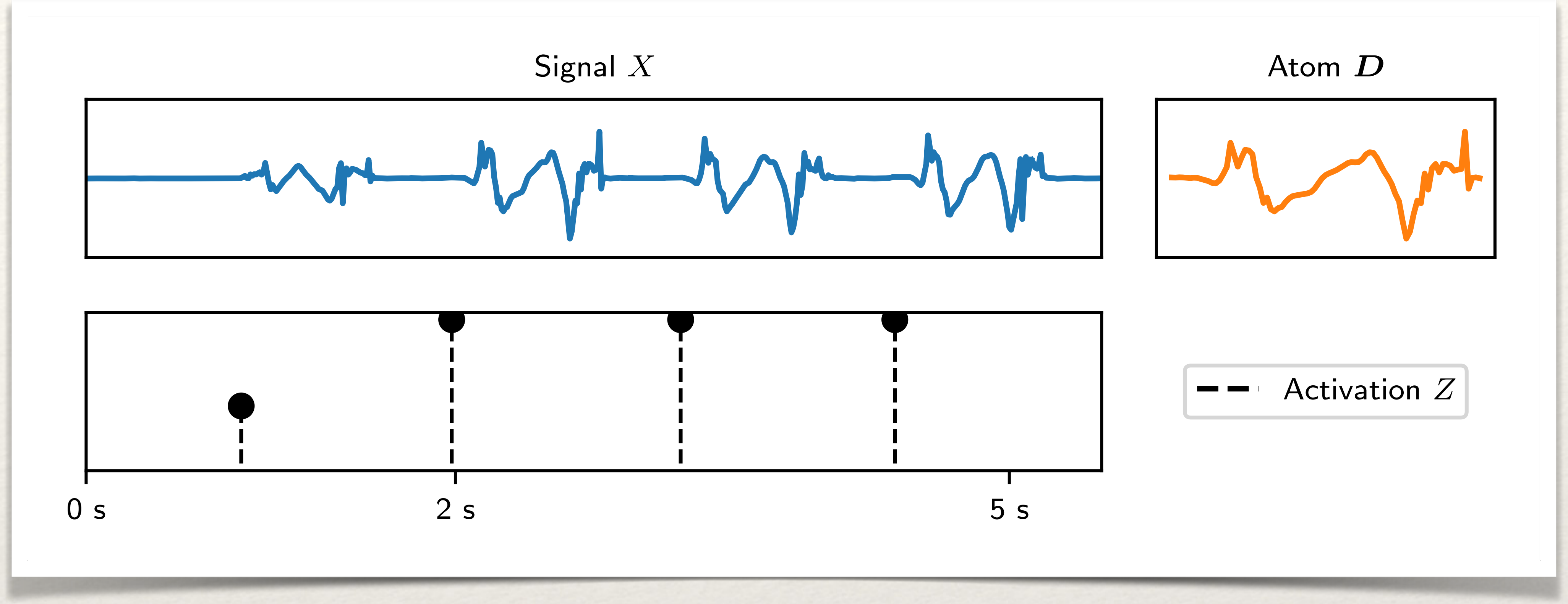
Principle of Convolutional Dictionary Learning

Multivariate CSC [1]:

$$\min_{D_k, z_k^n} \sum_{n=1}^{N} \frac{1}{2} \left\| X^n - \sum_{k=1}^{K} z_k^n * D_k \right\|_2^2 + \lambda \sum_{k=1}^{K} \left\| z_k^n \right\|_1,$$

$$\text{s.t.} \quad \left\| D_k \right\|_2^2 \leq 1 \text{ and } z_k^n \geq 0$$

- $\{X^n\}_{n=1}^N \subset \mathbb{R}^{P \times T}$: observed signals
- $\{D_k\}_{k=1}^K \subset \mathbb{R}^{P \times L}$: the spatio-temporal atoms
- $\{z_k^n\}_{k=1}^K \subset \mathbb{R}^{\widetilde{T}}$: the sparse activations associated with X^n, $\widetilde{T} = T - L + 1$
- $\lambda > 0$: regularization parameter



Decomposition of a noiseless univariate signal X (blue) as the convolution Z ∗ D between a temporal pattern D (orange) and a sparse activation signal Z

[1] R. Grosse, R. Raina, H. Kwong, and A. Y. Ng. Shift-invariant sparse coding for audio classification. In 23rd Conference on Uncertainty in Artificial Intelligence, 2007.

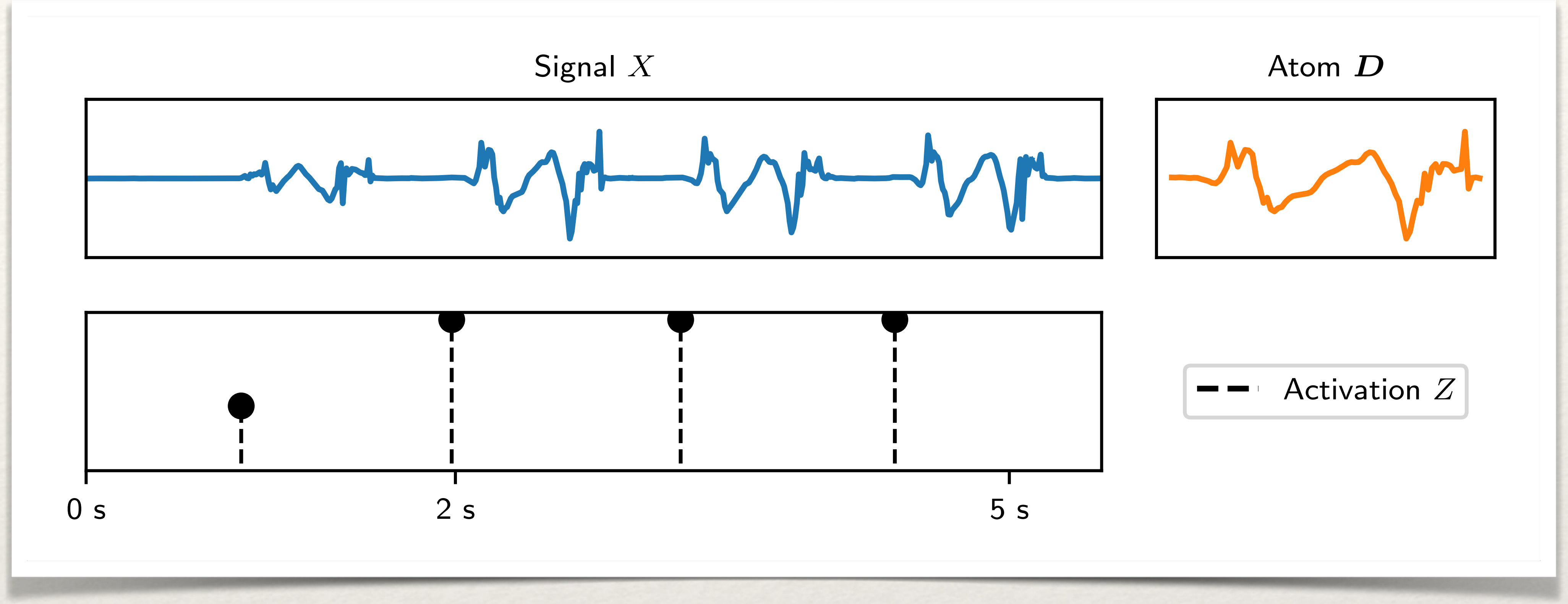
Principle of Convolutional Dictionary Learning

Multivariate CSC [1]:

$$\min_{D_k, z_k^n} \sum_{n=1}^{N} \frac{1}{2} \left\| X^n - \sum_{k=1}^{K} z_k^n * D_k \right\|_2^2 + \lambda \sum_{k=1}^{K} \left\| z_k^n \right\|_1,$$

$$\text{s.t.} \quad \left\| D_k \right\|_2^2 \leq 1 \text{ and } z_k^n \geq 0$$

- $\{X^n\}_{n=1}^N \subset \mathbb{R}^{P \times T}$: observed signals
- $\{D_k\}_{k=1}^K \subset \mathbb{R}^{P \times L}$: the spatio-temporal atoms
- $\{z_k^n\}_{k=1}^K \subset \mathbb{R}^{\widetilde{T}}$: the sparse activations associated with X^n, $\widetilde{T} = T - L + 1$
- $\lambda > 0$: regularization parameter



Decomposition of a noiseless univariate signal X (blue) as the convolution $Z * D$ between a temporal pattern D (orange) and a sparse activation signal Z

Bi-convex optimization problem solved with alternate minimization:

1. given K fixed atoms D_k and a regularization parameter $\lambda > 0$, retrieve the activation signals z_k^n by locally greedy coordinate descent (LGCD);
2. then, given NK fixed activation signals z_k^n, update the K atoms D_k, and so forth until convergence.

[1] R. Grosse, R. Raina, H. Kwong, and A. Y. Ng. Shift-invariant sparse coding for audio classification. In 23rd Conference on Uncertainty in Artificial Intelligence, 2007.

CDL with rank-1 constraint

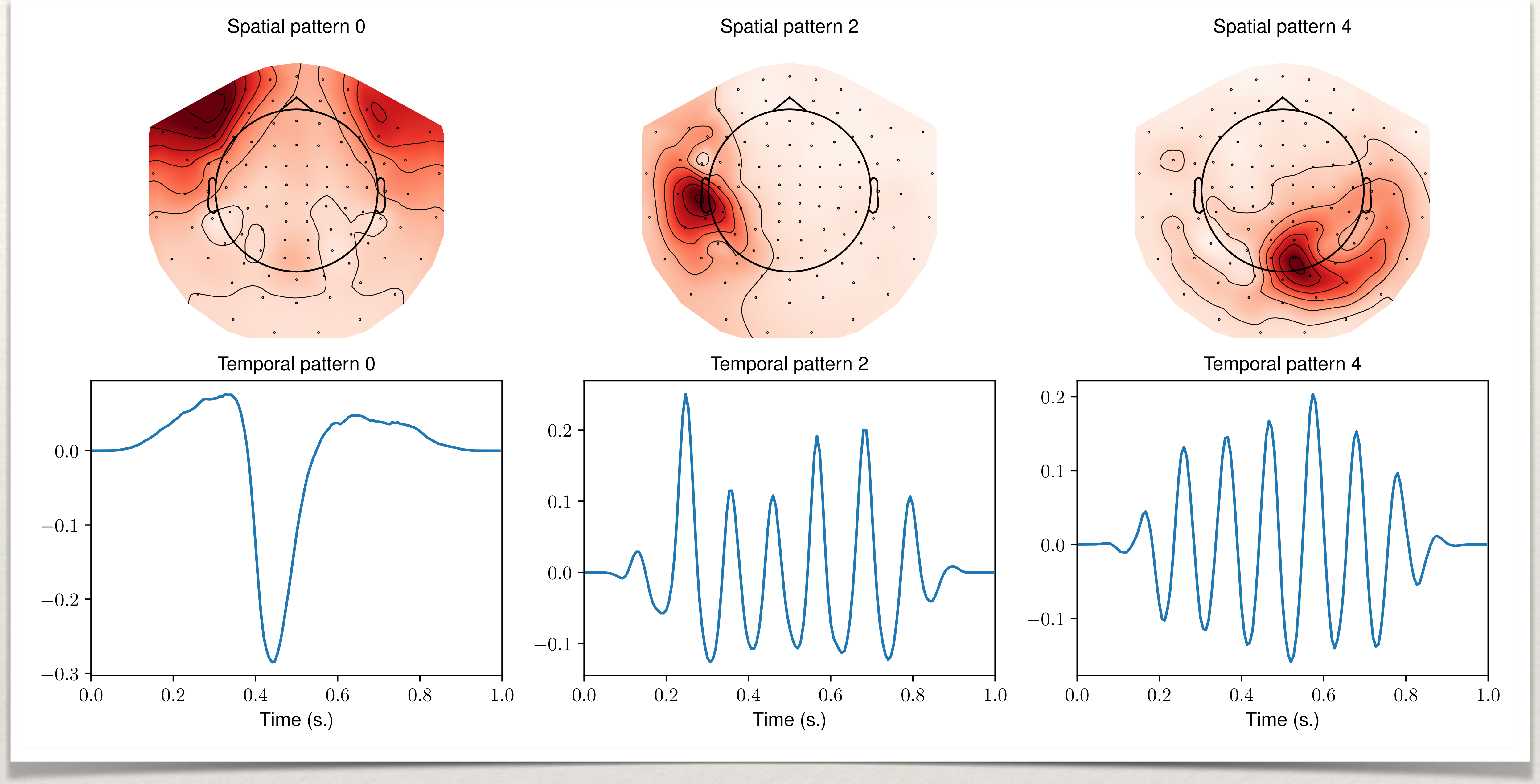
Multivariate CSC with rank-1 constraint [1]:

$$\boxed{D_k = u_k {v_k}^\top \in \mathbb{R}^{P \times L}}$$

- $u_k \in \mathbb{R}^P$: pattern over the channels (sensors);
- $v_k \in \mathbb{R}^L$: pattern over time.

$$\min_{u_k, v_k, z_k^n} \sum_{n=1}^{N} \frac{1}{2} \left\| X^n - \sum_{k=1}^{K} z_k^n * (u_k v_k^\top) \right\|_2^2 + \lambda \sum_{k=1}^{K} \| z_k^n \|_1 ,$$

$$\text{s.t.} \quad \| u_k \|_2^2 \leq 1, \ \| v_k \|_2^2 \leq 1 \text{ and } z_k^n \geq 0$$



Spatial and temporal patterns of 3 handpicked atoms extracted using CDL with rank-1 constraint on MNE somato dataset.

Remark: each atom comes along with its sparse activation vector

Python package: alphacsc

[1] *Dupré La Tour, T., Moreau, T., Jas, M., & Gramfort, A. Multivariate Convolutional Sparse Coding for Electromagnetic Brain Signals. NIPS, 2018.*

Convolutional Dictionary Learning

How the atoms' onsets are link to the stimuli events?

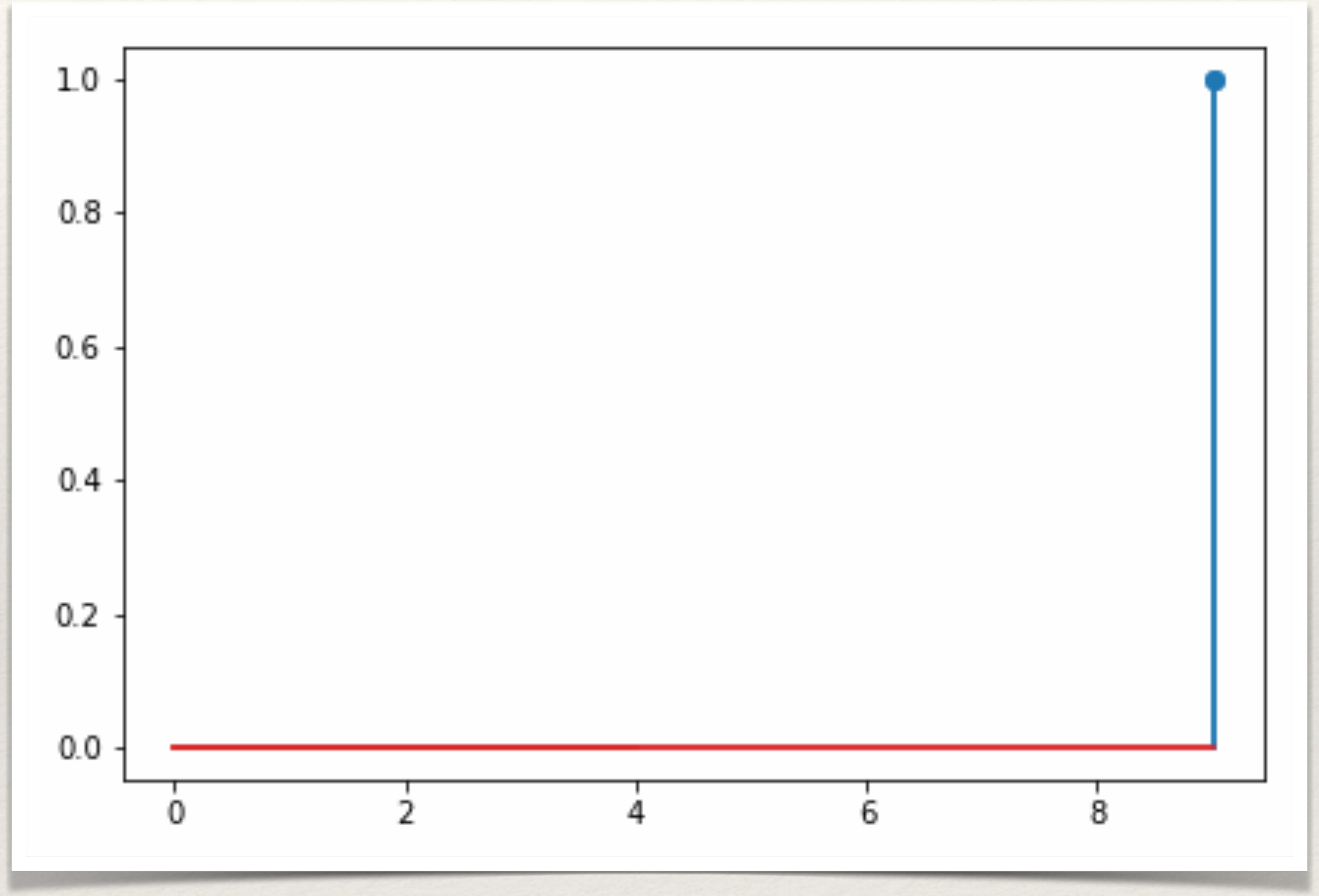
From the M/EEG recording to DriPP

- Paradigm of a M/EEG experiment
- Convolutional Dictionary Learning for recurrent patterns identification in M/EEG signals
- Background on Point Processes
- Methodology of DriPP
- Results on synthetic and real datasets

Background on Point Processes

Temporal point process (TPP): a stochastic, or random, process composed of a time series of binary events that occur in continuous time; $S = [0,T)$ the time interval [1].

- A realization of a TPP is a set of distinct time points $\xi = \{t_1, \ldots t_n\}$ occurring before T.
- To ξ, we associate the counting process $N_t = \sum_{t \in \xi} 1_{t_i \leq t}$, *i.e.*, the number of points in the time interval $[0,t]$.
- N_t : a random process which evolves over time by jumps of size 1.
- Studying TPP $\Leftrightarrow$ analyzing when this jumps occur.

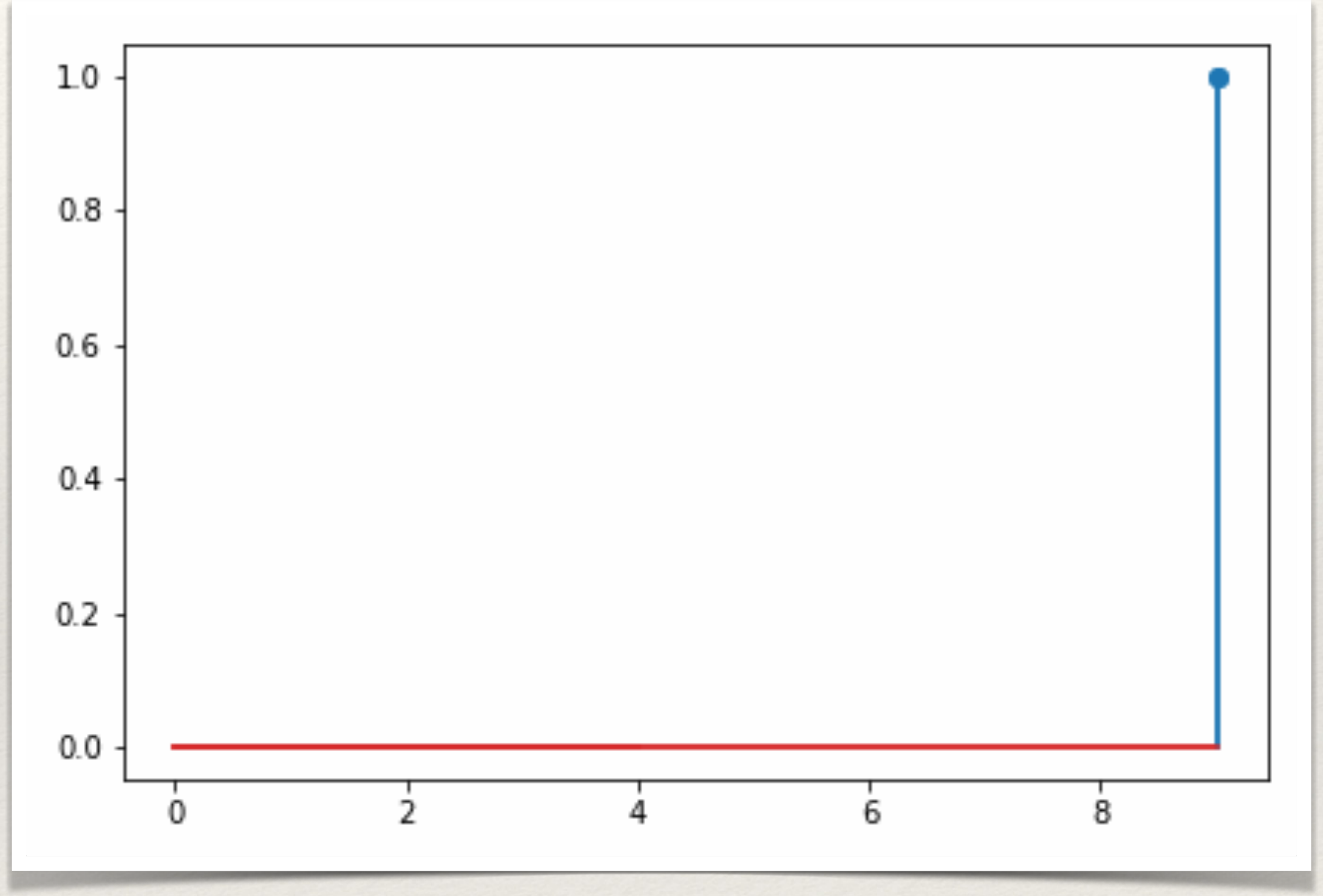


[1] *D. J. Daley and D. Vere-Jones. An introduction to the theory of point processes. Volume I: Elementary theory and methods. Probability and Its Applications. Springer-Verlag New York, 2003.*

Background on Point Processes

Temporal point process (TPP): a stochastic, or random, process composed of a time series of binary events that occur in continuous time; $S = [0,T)$ the time interval [1].

- A realization of a TPP is a set of distinct time points $\xi = \{t_1, \ldots t_n\}$ occurring before T.
- To ξ, we associate the counting process $N_t = \sum_{t \in \xi} 1_{t_i \leq t}$, *i.e.*, the number of points in the time interval $[0,t]$.
- N_t : a random process which evolves over time by jumps of size 1.
- Studying TPP $\Leftrightarrow$ analyzing when this jumps occur.

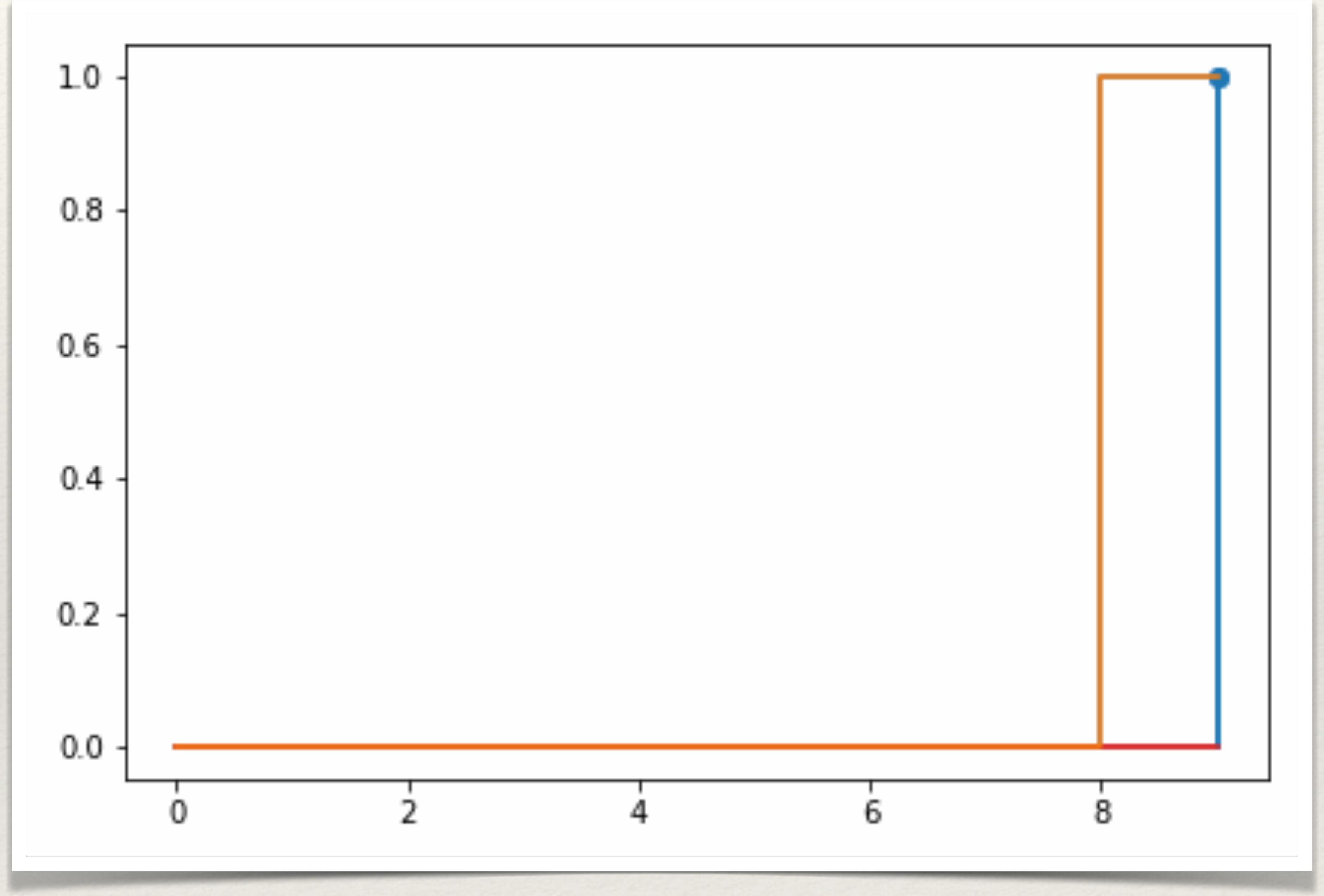


[1] D. J. Daley and D. Vere-Jones. An introduction to the theory of point processes. Volume I: Elementary theory and methods. Probability and Its Applications. Springer-Verlag New York, 2003.

Background on Point Processes

Temporal point process (TPP): a stochastic, or random, process composed of a time series of binary events that occur in continuous time; $S = [0,T)$ the time interval [1].

- A realization of a TPP is a set of distinct time points $\xi = \{t_1, \ldots t_n\}$ occurring before T.
- To ξ, we associate the counting process $N_t = \sum_{t \in \xi} 1_{t_i \leq t}$, *i.e.*, the number of points in the time interval $[0,t]$.
- N_t : a random process which evolves over time by jumps of size 1.
- Studying TPP ⇔ analyzing when this jumps occur.



[1] D. J. Daley and D. Vere-Jones. An introduction to the theory of point processes. Volume I: Elementary theory and methods. Probability and Its Applications. Springer-Verlag New York, 2003.

Background on Point Processes

Intensity function

$$\boxed{\lambda\left(t \mid \mathcal{F}_t\right) = \lim_{\mathrm{d}t \to 0} \frac{\mathbb{P}\left(N_{t+\mathrm{d}t} - N_t = 1 \mid \mathcal{F}_t\right)}{\mathrm{d}t}}$$

$N_t = \sum_{t \in \xi} 1_{t_i \leq t}$ the counting process, $\mathcal{F}_t = \{t_i, t_i < t\}$

Homogeneous Poisson process: $\lambda\left(t \mid \mathcal{F}_t\right) \equiv \lambda^*$, $\left(N_{t+\Delta t} - N_t\right) \sim \mathit{Poisson}\,(\Delta t)$

Inhomogeneous Poisson process: $\lambda\left(t \mid \mathcal{F}_t\right) \equiv \lambda^*(t) \geq 0$

Goal: find a statistical model to uncover the underlying generating model

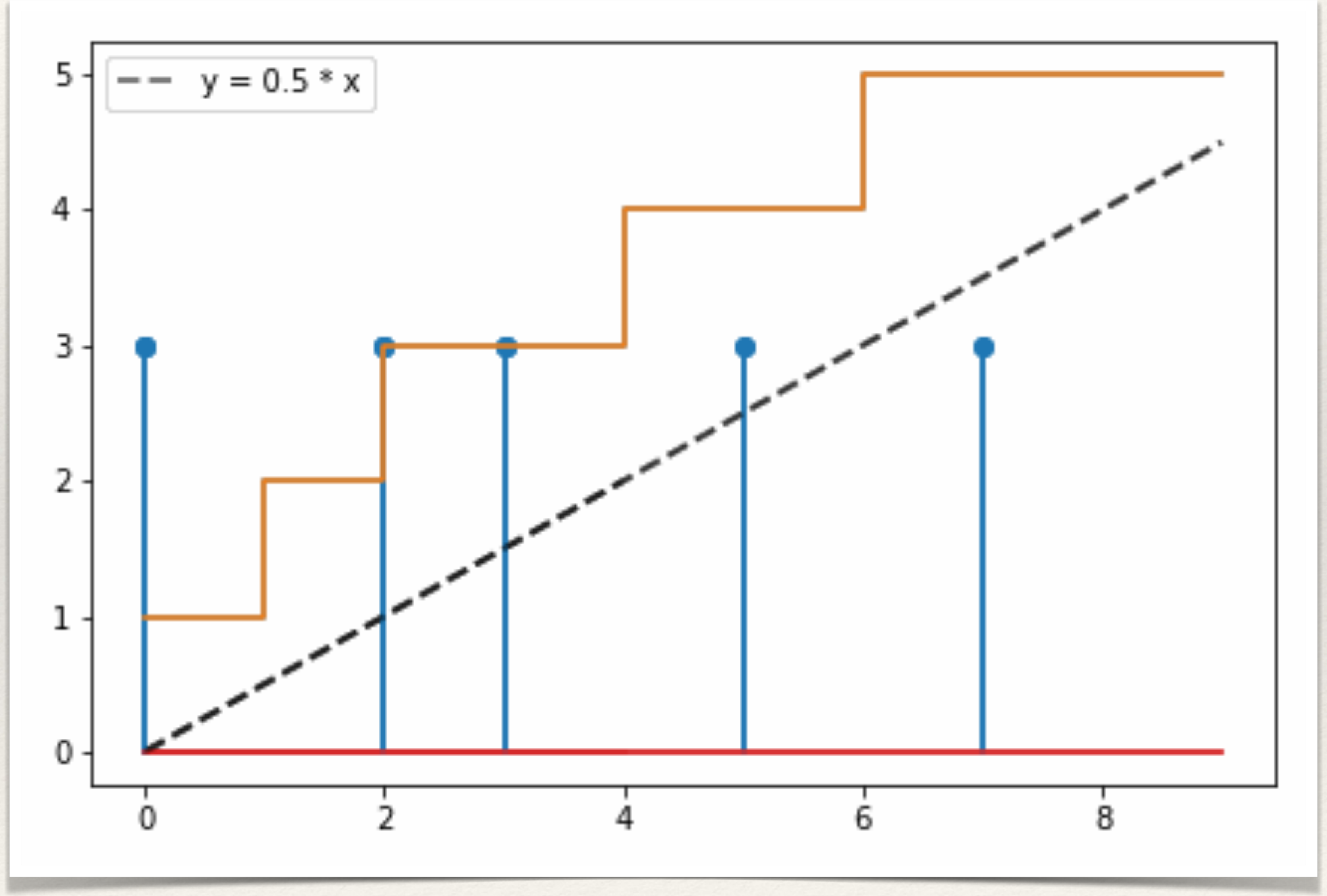
Background on Point Processes

- **Homogeneous Poisson process**:

  $\lambda\left(t \mid \mathcal{F}_t\right) \equiv \lambda,\ \left(N_{t+\Delta t} - N_t\right) \sim Poisson\left(\Delta t\right)$

  $\forall B \in \mathcal{B}(\mathbb{R}), M(B) = \lambda \lvert B \rvert$, where $\lvert \cdot \rvert$ is the Lebesgue measure on $\left(S, \mathcal{B}(\mathbb{R})\right)$

  $\forall B \in \mathcal{B}(\mathbb{R})$, the mean measure $M(B) = \mathbb{E}\left[N\left(B\right)\right]$ is the expected number of events in B.



Simulation of a homogeneous Poisson process of parameter $\lambda = 0.5$

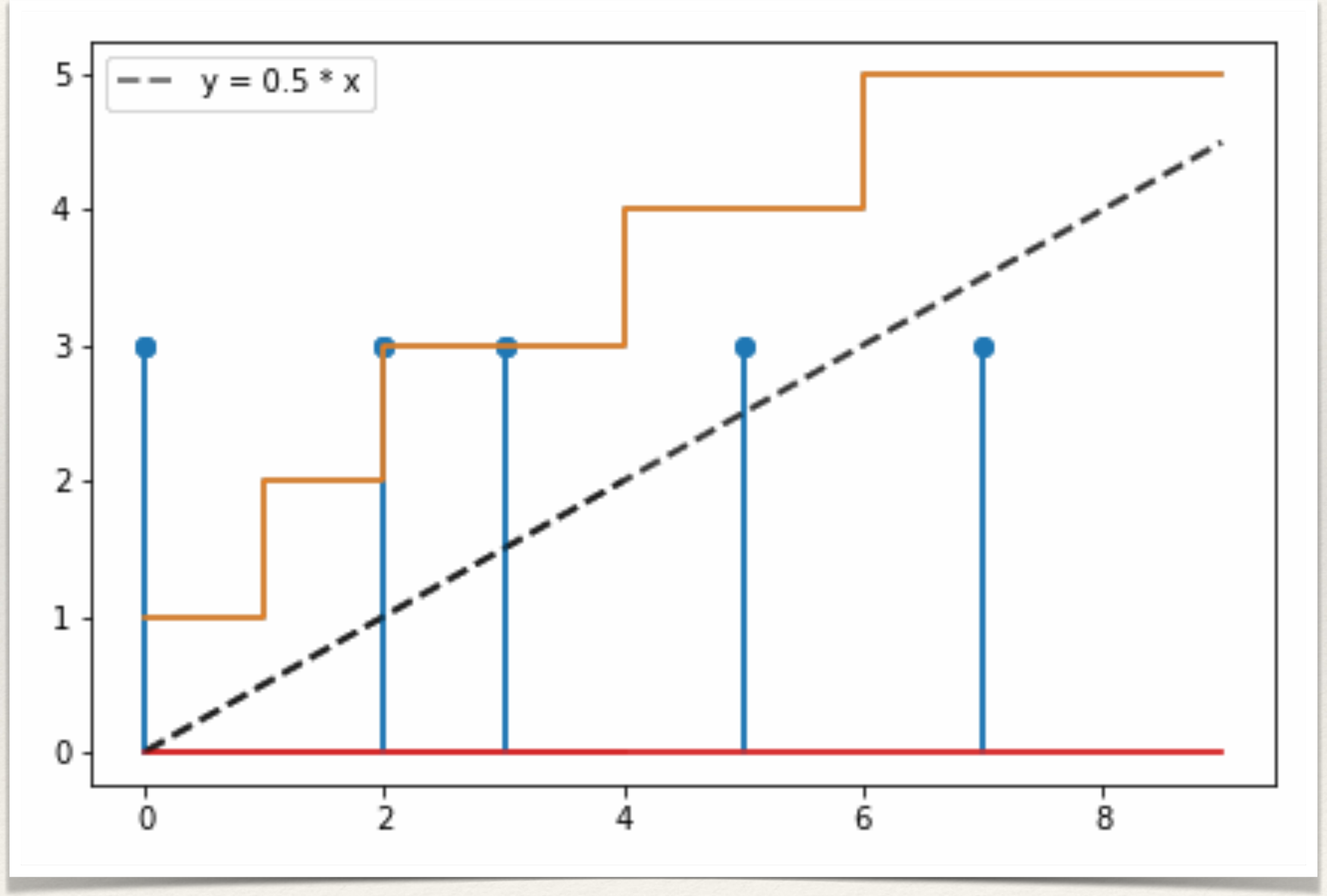
Background on Point Processes

- **Homogeneous Poisson process**:

  $\lambda\left(t \mid \mathscr{F}_t\right) \equiv \lambda, \left(N_{t+\Delta t} - N_t\right) \sim Poisson\left(\Delta t\right)$

  $\forall B \in \mathscr{B}(\mathbb{R}), M(B) = \lambda |B|$, where $|\cdot|$ is the Lebesgue measure on $\left(S, \mathscr{B}(\mathbb{R})\right)$

  $\forall B \in \mathscr{B}(\mathbb{R})$, the mean measure $M(B) = \mathbb{E}\left[N(B)\right]$ is the expected number of events in B.



Simulation of a homogeneous Poisson process of parameter $\lambda = 0.5$

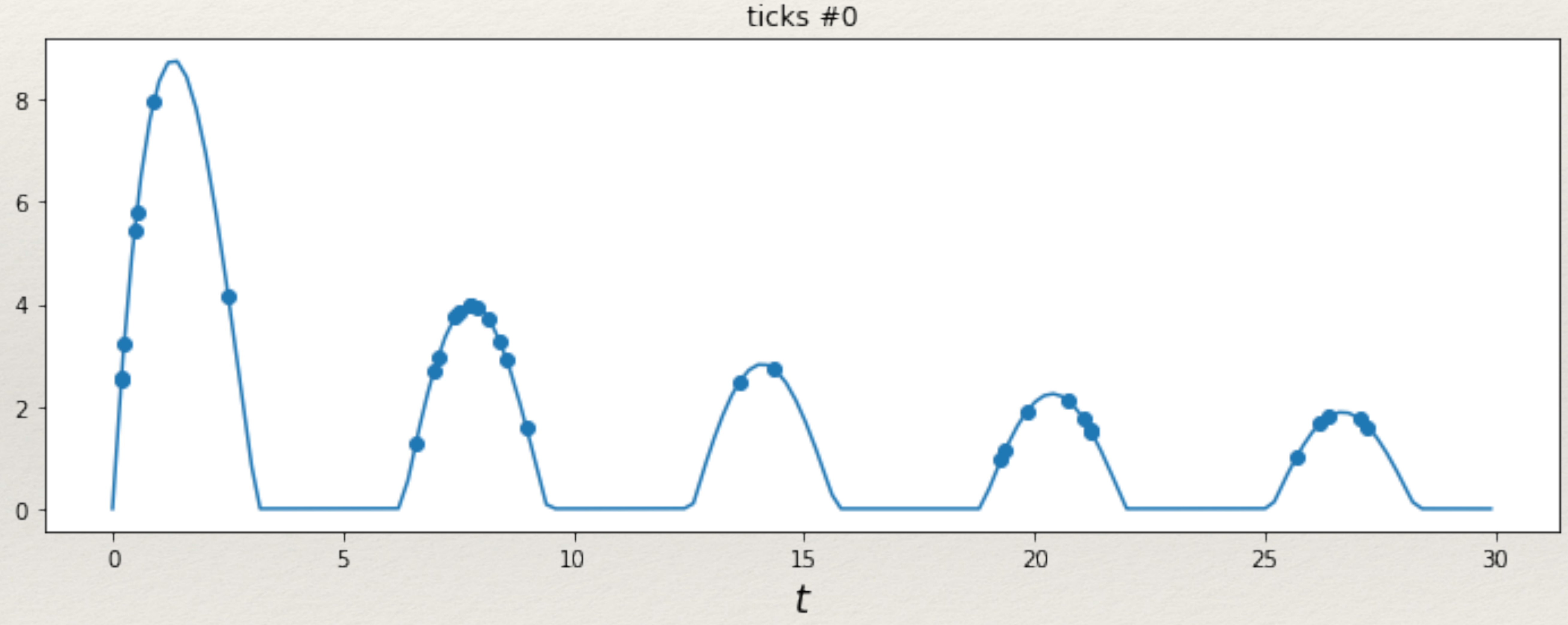
Background on Point Processes

- **Inhomogeneous Poisson process:**

$$\lambda\left(t \mid \mathscr{F}_t\right) \equiv \lambda^*(t)$$

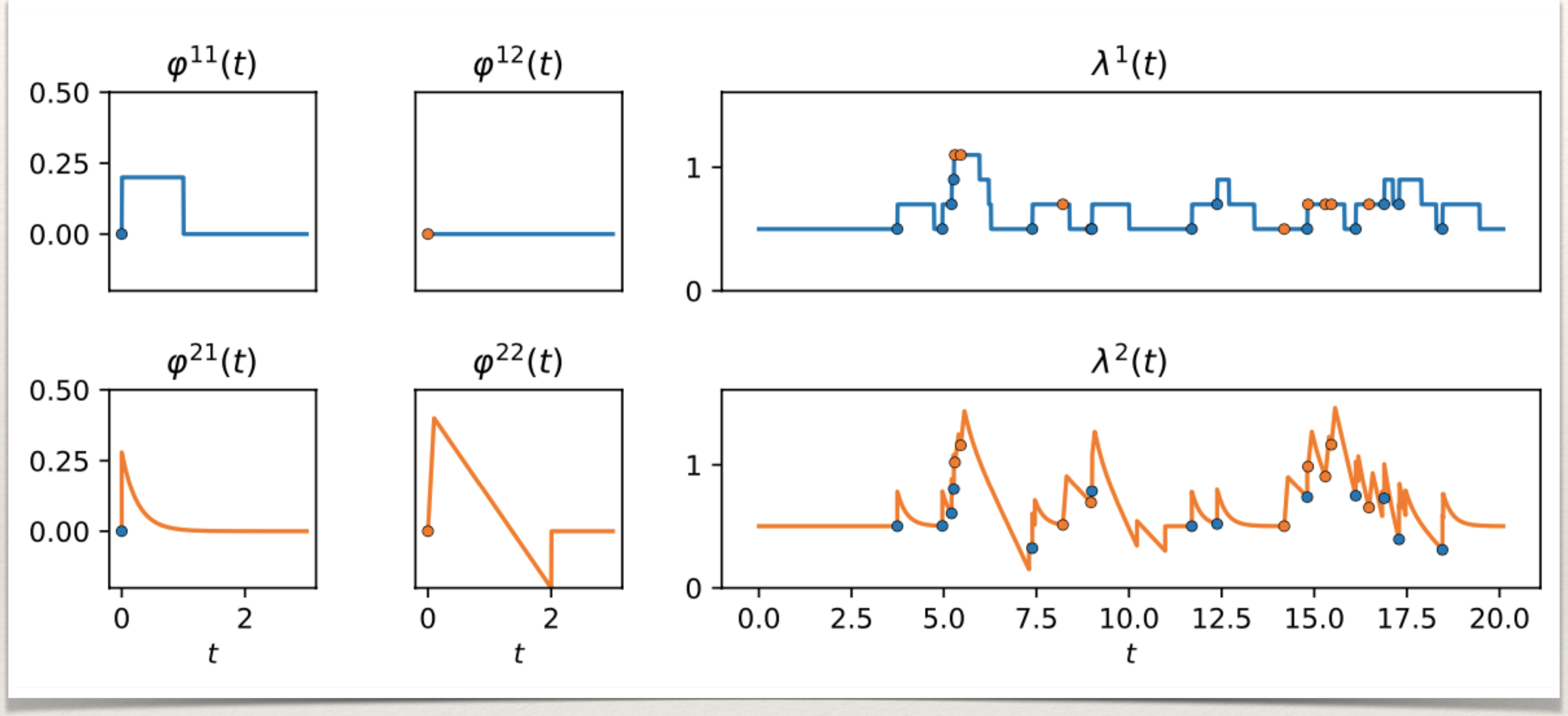
$$\forall B \in \mathscr{B}(\mathbb{R}), M(B) = \int_B \lambda^*(x)\mathrm{d}x$$

$\forall B \in \mathscr{B}(\mathbb{R})$, the mean measure $M(B) = \mathbb{E}\left[N(B)\right]$ is the expected number of events in B.



Simulation of a inhomogeneous Poisson process

Multivariate Hawkes Processes



Python package: tick

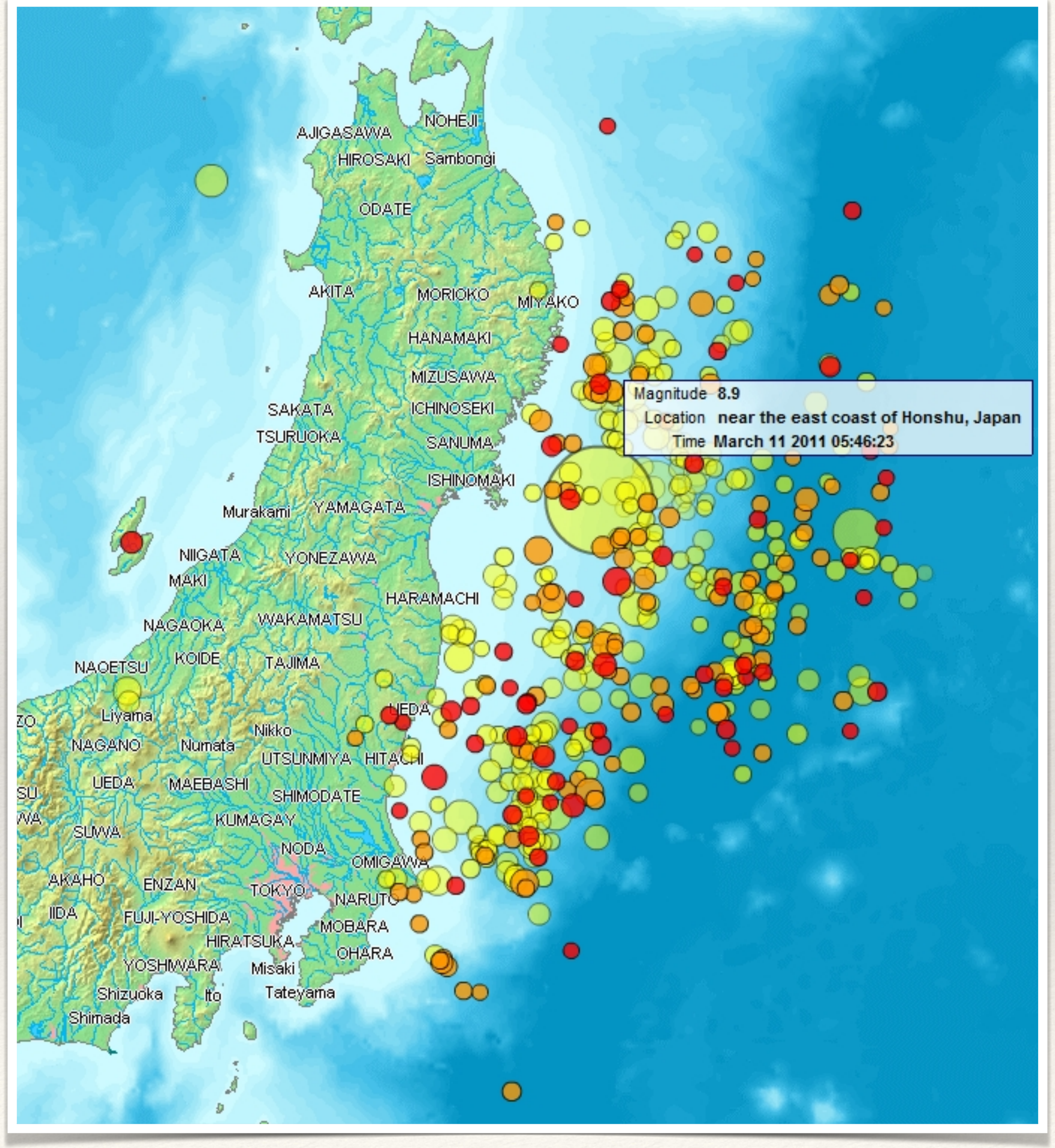
Source: Bompaire M., Machine learning based on Hawkes processes and stochastic optimization, thesis, 2019

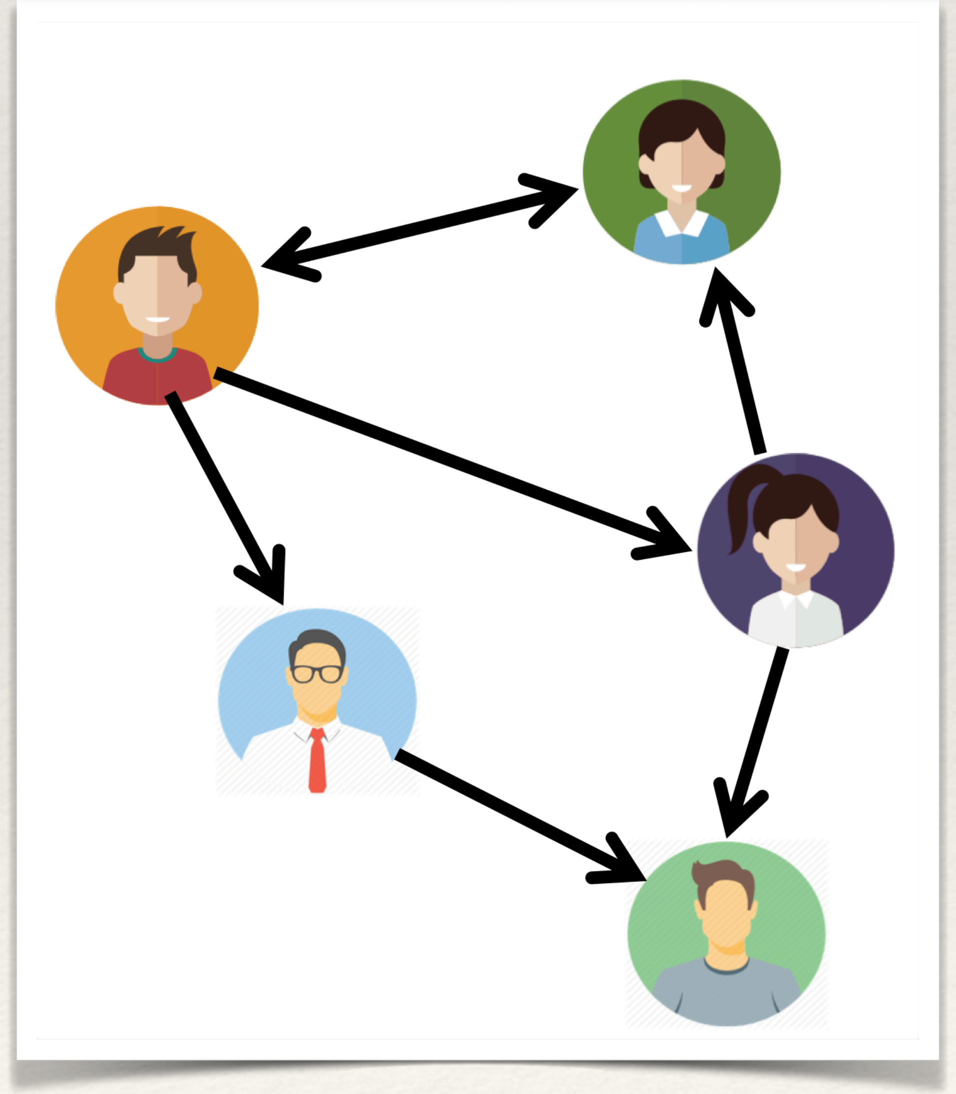
- Model interaction between multiple PP:

$\phi_{i,j}$: influence of j on i

$$\lambda_i\left(t \,|\, \mathcal{F}_t\right) = \mu_i + \sum_{j=1}^{K} \int_{-\infty}^{t} \phi_{i,j}(t-s)\mathrm{d}N_s^{(j)}, \quad \mu_i \geq 0$$

$$e.g., \ \phi_{i,j}(t) = \alpha_{i,j}\beta \exp\left(-\beta t\right), \quad \alpha_{i,j} > 0, \beta > 0$$





Multiple usages for multi-dimensional Hawkes processes:

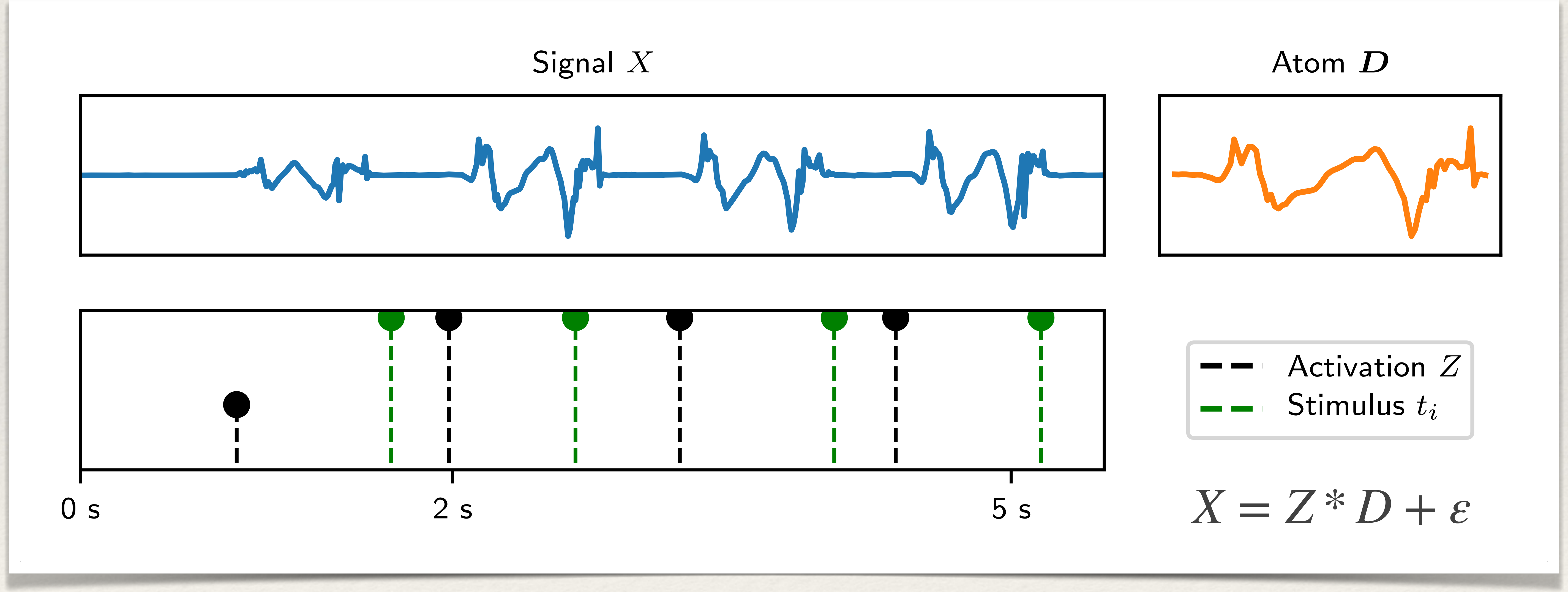
- Earthquake propagation and replicas [Vere-Jones, 1970; Ogata 1999];
- Market stocks [Bacry *et al.*, 2013 & 2015];
- Social network interactions [Crane and Sornette, 2008; Mitchell and Cates, 2009];
- etc.

From the M/EEG recording to DriPP

- Paradigm of a M/EEG experiment
- Convolutional Dictionary Learning for recurrent patterns identification in M/EEG signals
- Background on Point Processes
- Methodology of DriPP
- Results on synthetic and real datasets

DriPP, or how to link two Point Processes

1. During an M/EEG experiment, several stimuli (auditory, visual, etc.) are presented to the subject;
2. Using Convolutional Dictionary Learning (CDL), recurrent patterns called *atoms* are extracted from the raw signal;
3. Each atom comes along with its sparse activation vector and is composed of a spatial and a temporal representation.



How to uncover the link between stimuli (green) and atom's activations (black)?

DriPP, or how to link two Point Processes

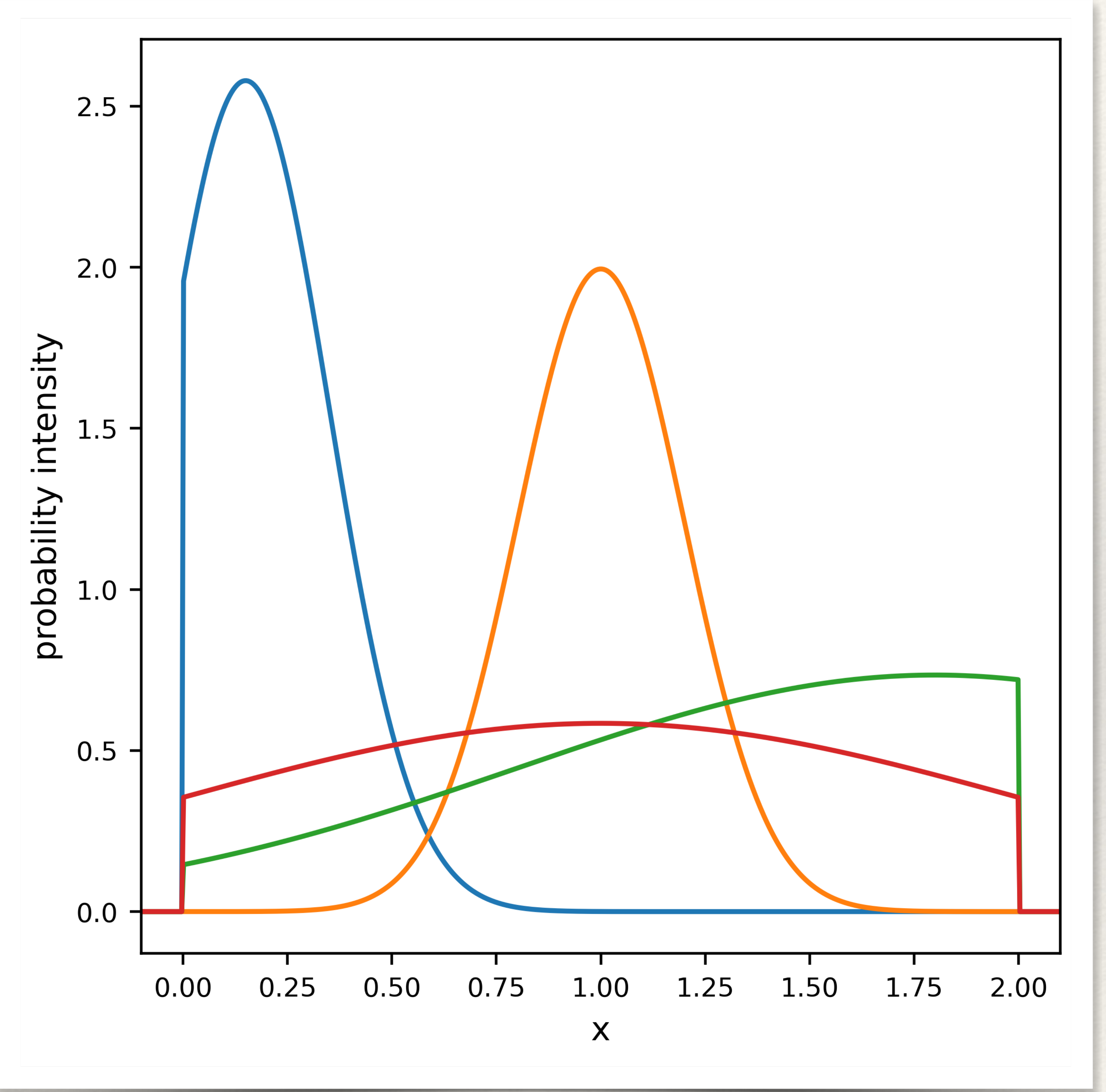
- Intensity function that links a PP k to a potential set of triggering PP $\mathscr{P}$, called **drivers**:

$$\lambda_{k,\mathscr{P}} = \mu_k + \sum_{p\in\mathscr{P}} \sum_{i,t_i^{(p)}\leq t} \alpha_{k,p}\kappa_{k,p}\left(t - t_i^{(p)}\right), \quad \mu_k \geq 0, \alpha_{k,p} \geq 0$$

- Use of a **truncated Gaussian kernel**, to capture some **latency**:

$$\boxed{\kappa_{k,p} = \mathscr{N}_{[a,b]}\left(m_{k,p}, \sigma_{k,p}^2\right)}$$

- Derived an **EM algorithm** to learn the *few* parameters $\Theta_{k,\mathscr{P}} = \left(\mu_k, \boldsymbol{\alpha}_{k,\mathscr{P}}, \boldsymbol{m}_{k,\mathscr{P}}, \boldsymbol{\sigma}_{k,\mathscr{P}}\right)$



Probability density functions for the truncated normal distribution for different sets of parameters. In all cases, $a = 0$ and $b = 2$.

DriPP - EM algorithm

Define the negative log-likelihood as the loss function:

$$\mathscr{L}_{k,\mathscr{P}}\left(\Theta_{k,\mathscr{P}}\right) = \int_0^T \lambda_{k,\mathscr{P}}(t)\mathrm{d}t - \sum_{t\in\mathscr{A}_k} \log \lambda_{k,\mathscr{P}}(t)$$

where $\mathscr{A}_k$ denotes the set of events of stochastic process k.

Recall that we defined the intensity by $\lambda_{k,\mathscr{P}} = \mu_k + \sum_{p\in\mathscr{P}} \sum_{i,t_i^{(p)}\leq t} \alpha_{k,p}\kappa_{k,p}\left(t - t_i^{(p)}\right)$, hence,

$$\mathscr{L}_{k,\mathscr{P}}\left(\Theta_{k,\mathscr{P}}\right) = \mu_k T + \sum_{p\in\mathscr{P}} \alpha_{k,p} n_p - \sum_{t\in\mathscr{A}_k} \log\left(\mu_k + \sum_{p\in\mathscr{P}} \sum_{i,t_i^{(p)}\leq t} \alpha_{k,p}\kappa_{k,p}\left(t - t_i^{(p)}\right)\right)$$

By cancelling the loss derivatives with respect to each parameters, we obtained the EM update equations.

DriPP - EM algorithm

- **Expectation step**: compute the events' assignation, *i.e.*, the probability that an event comes from either the kernel or the baseline intensity.

$$P_k^{(n)} = \frac{\mu_k^{(n)}}{\lambda_{k,\mathscr{P}}^{(n)}(t)} \quad \text{and } \forall p \in \mathscr{P}, P_p^{(n)} = \frac{\alpha_{k,\mathscr{P}}^{(n)} \sum_{i, t_i^{(p)} < t} \kappa_{k,\mathscr{P}}^{(n)} \left(t - t_i^{(p)}\right)}{\lambda_{k,\mathscr{P}}^{(n)}(t)}$$

- **Maximisation step**: we fix the probabilities $P_k^{(n)}$ and $P_p^{(n)}$, and cancel the loss with respect to each parameter.

$$\mu^{(n+1)} = \frac{1}{T} \sum_{t \in \mathscr{A}_k} P_k^{(n)}(t) \quad \text{and } \alpha^{(n+1)} = \pi_{\mathbb{R}^+} \left(\frac{1}{n_p} \sum_{t \in \mathscr{A}_k} P_p^{(n)}(t) \right)$$

Algorithm 1: EM-based algorithm

input : $\mathcal{A}_k, \mathcal{T}_p, a, b, T, N$
output : The estimated values for parameters $\mu, \boldsymbol{\alpha}, \boldsymbol{m}$ and $\boldsymbol{\sigma}$

Initialize $\mu^{(0)}, \boldsymbol{\alpha}^{(0)}, \boldsymbol{m}^{(0)}, \boldsymbol{\sigma}^{(0)}$;
for $i = 0, \ldots, N-1$ **do**
  if $\boldsymbol{\alpha}^{(i)} = \mathbf{0}_{\mathbb{R}^{\#\mathcal{P}}}$ **then**
    $\mu^{(i+1)} = \mu^{(\mathrm{MLE})}$; $\mu_k^{(MLE)} = \#\mathscr{A}_k / T$
    break;
  end
  Define $\lambda^{(i)}$; Compute $\mu^{(i+1)}, \boldsymbol{\alpha}^{(i+1)}, \boldsymbol{m}^{(i+1)}, \boldsymbol{\sigma}^{(i+1)}$;
end
return $\mu^{(i+1)}, \boldsymbol{\alpha}^{(i+1)}, \boldsymbol{m}^{(i+1)}, \boldsymbol{\sigma}^{(i+1)}$

Pseudo-code of the EM algorithm

DriPP - EM algorithm

Moment-matching initialization: parameters are initialized based on their « role » in the model

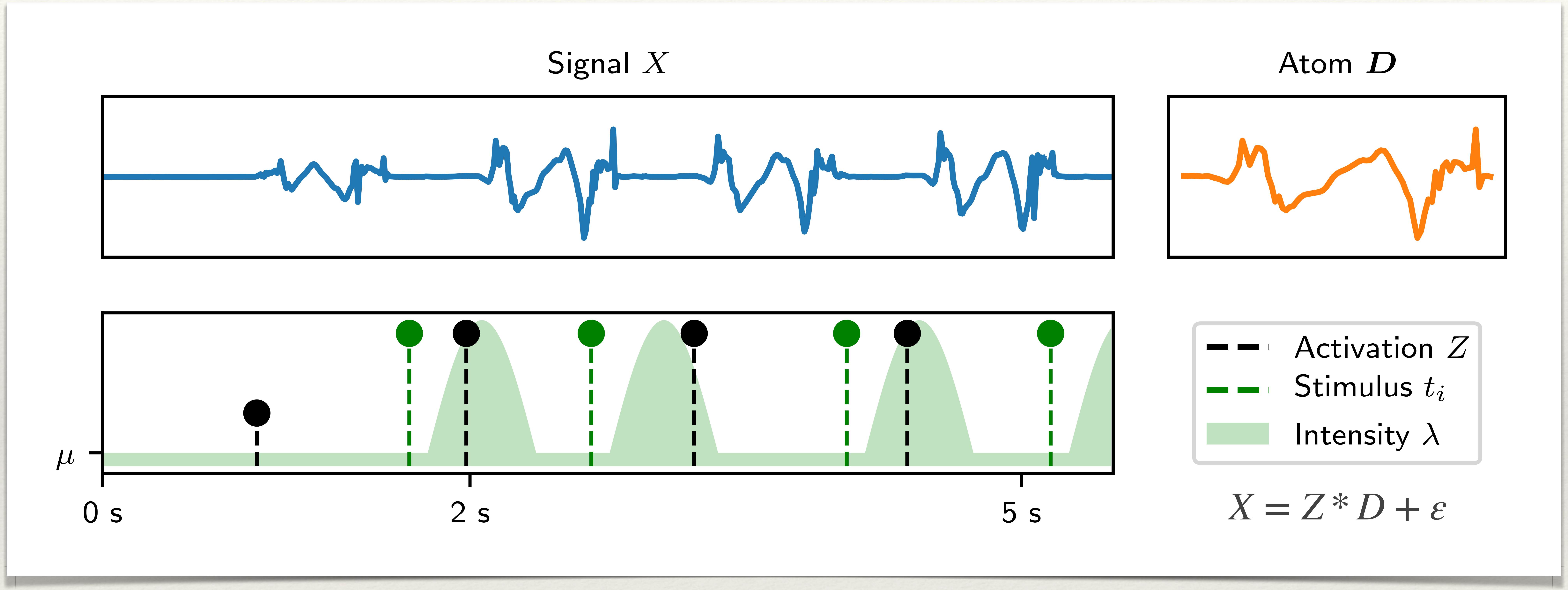
$$m_{k,p}^{(0)} = \frac{1}{\#\mathcal{D}_{k,p}} \sum_{d \in \mathcal{D}_{k,p}} d \quad \text{and} \quad \sigma_{k,p}^{(0)} = \sqrt{\frac{1}{\#\mathcal{D}_{k,p}} \sum_{d \in \mathcal{D}_{k,p}} |d - m_{k,p}^{(0)}|^2}$$

$$\mu_k^{(0)} = \frac{\#\mathcal{A}_k - \#\left(\bigcup_{p \in \mathcal{P}} \mathcal{D}_{k,p}\right)}{T - \lambda\left(\bigcup_{p \in \mathcal{P}} \bigcup_{t' \in \mathcal{T}_p} [t' + a, t' + b]\right)}$$

$$\alpha_{k,p}^{(0)} = \frac{\#\mathcal{D}_{k,p}}{\lambda\left(\bigcup_{t' \in \mathcal{T}_p} [t' + a, t' + b]\right)} - \mu_k^{(0)}, \quad \forall p \in \mathcal{P}$$

where $\mathcal{D}_{k,p} = \left\{ t - t_*^{(p)}(t), t \in \mathcal{A}_k \right\} \cap [a, b]$ the set of all empirical delays possibly linked to the driver p, and $t_*^{(p)}(t) = \max \left\{ t', t' \in \mathcal{T}_p, t' \leq t \right\}$ the timestamp of the last event on the driver p that occurred before time t

How to link two Point Processes



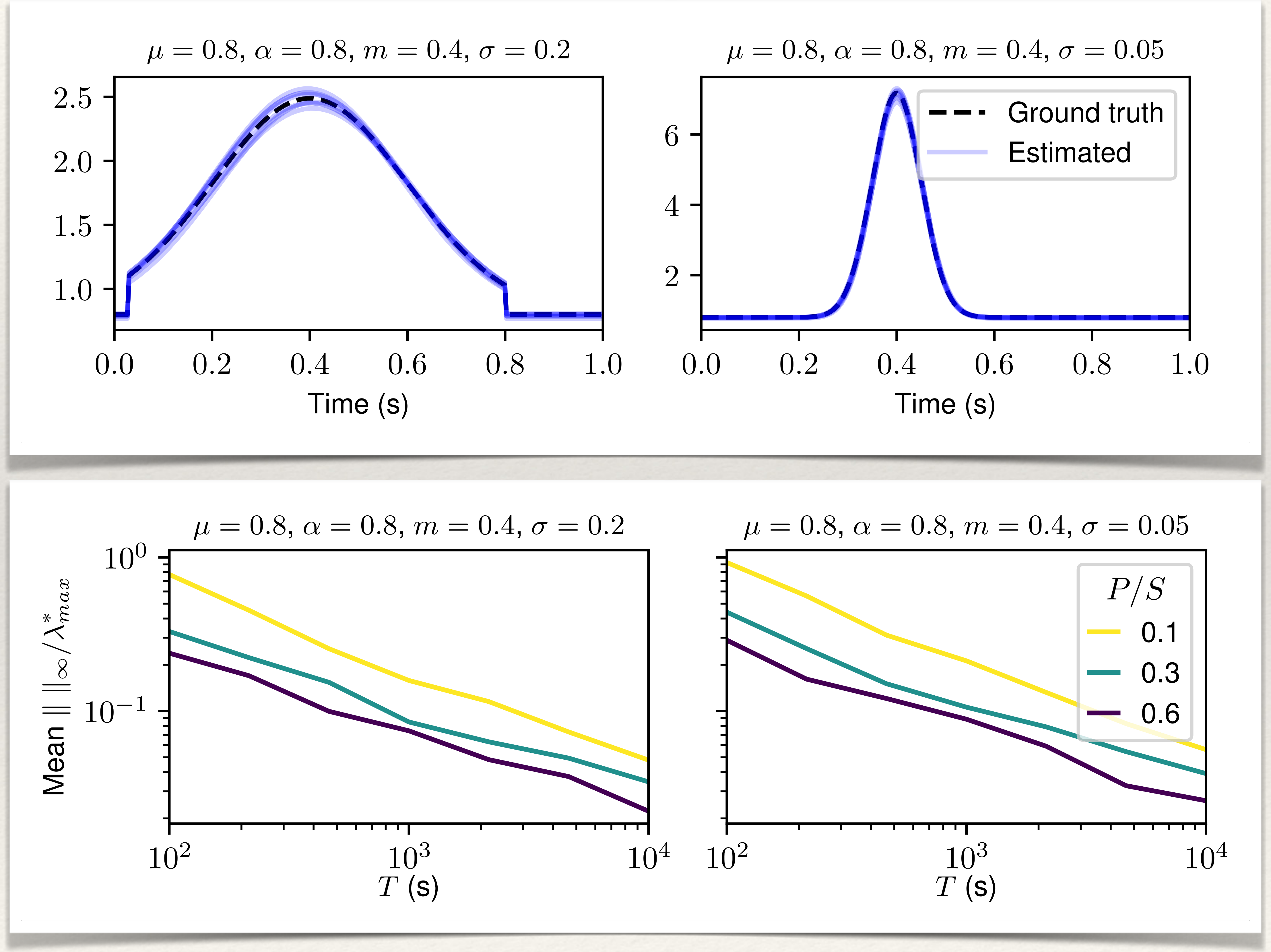
An intensity function is fitted between the stimuli PP (green) and the activations PP (black).

From the M/EEG recording to DriPP

- Paradigm of a M/EEG experiment
- Convolutional Dictionary Learning for recurrent patterns identification in M/EEG signals
- Background on Point Processes
- Methodology of DriPP
- Results on synthetic and real datasets

DriPP - Results on synthetic data

- Intensity with 2 drivers: wide and sharp kernels
- Data simulation following Lewis' thinning algorithm [1], for **multiple process duration** and **data intensity**.
- Relative norm: $||\lambda^* - \widehat{\lambda}||_\infty / \lambda^*_{\max}$
- The accuracy of the EM estimates increases with longer and denser processes.



[1] *PA W Lewis and Gerald S Shedler. Simulation of nonhomogeneous Poisson processes by thinning. Naval research logistics quarterly, 26(3):403–413, 1979.*

DriPP - Results on real data

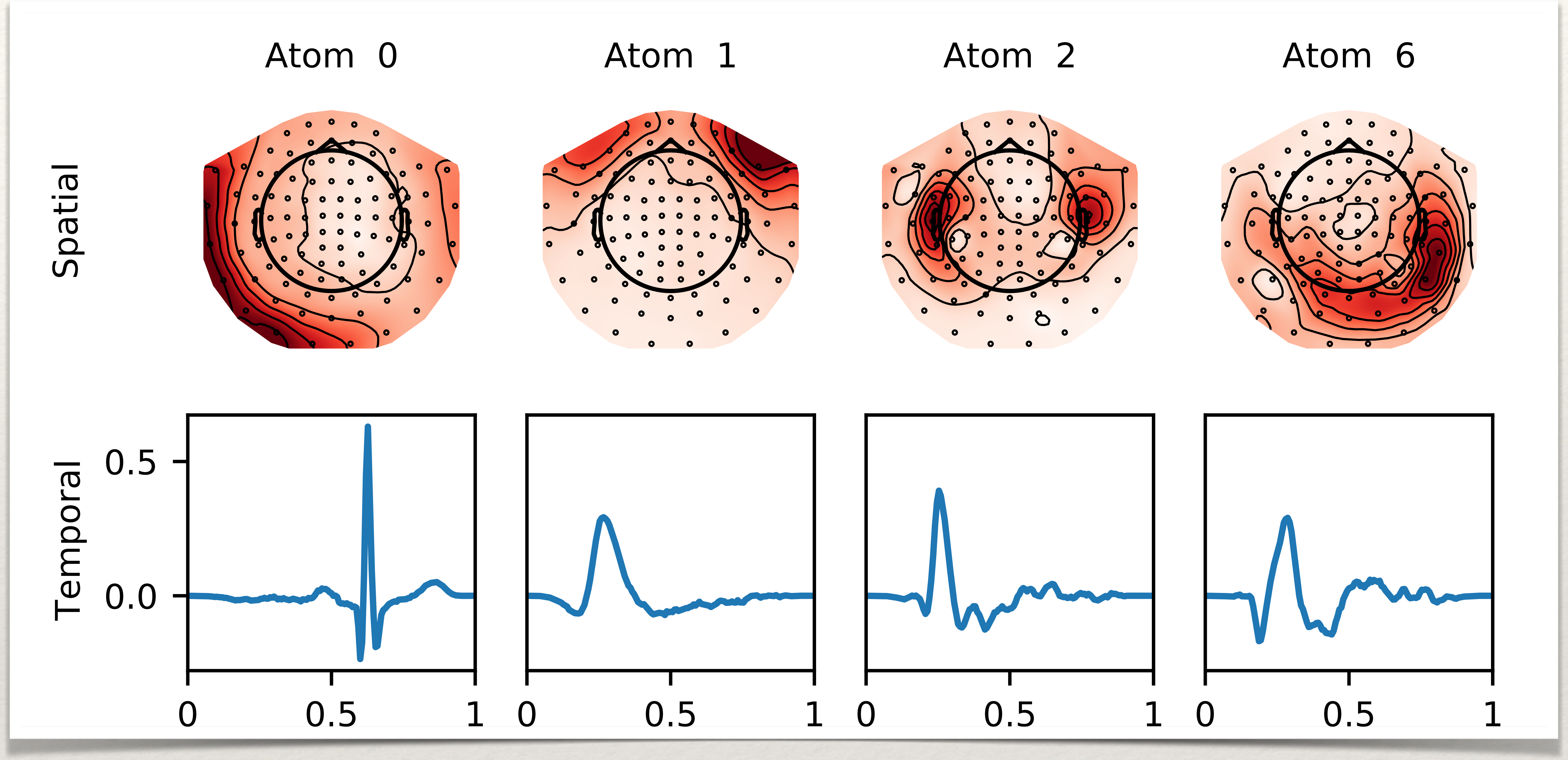
Results on 3 real MEG datasets:

- MNE sample: checkerboard patterns are presented to the subject in the left and right visual field, interspersed by tones to the left or right ear; 4.6 min long; 70 stimuli per type.
- MNE somato(sensory): 111 stimulations of a human subject left median nerve; 15 min
- Cam-CAN [1]: 643 human subjects submitted to audio and visual stimuli (120 bimodal audio/visual trials and eight unimodal trials); 4 min

[1] Shafto et al., The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing, 2014

DriPP - Results on audio-visual paradigm

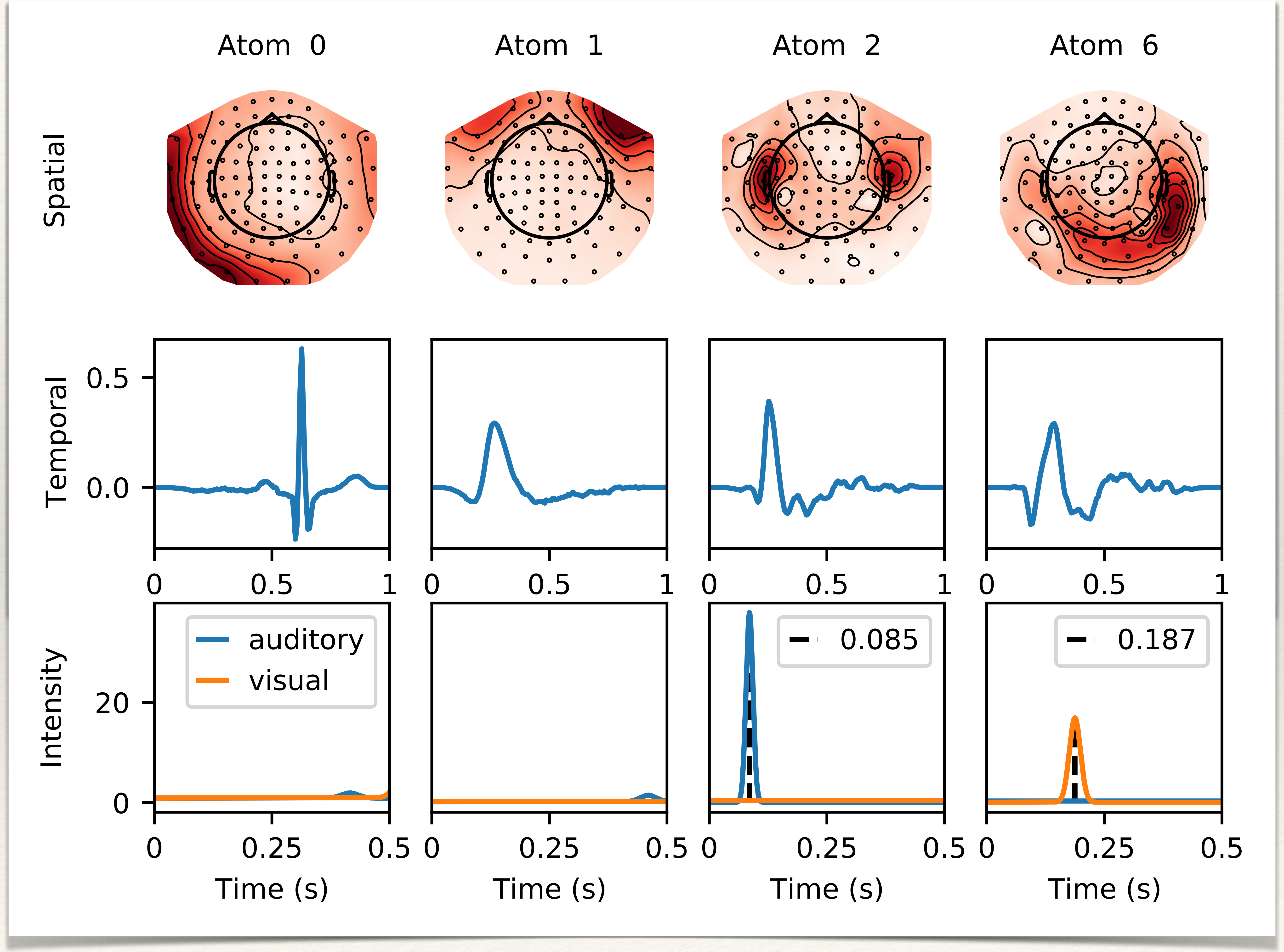
- Subject is randomly exposed to either visual or auditory stimuli for 6 mins.
- MEG Data collected on 306 sensors, 40 atoms of 0.5 s extracted



Spatial and temporal patterns of 4 atoms from MNE sample dataset, and their respective estimated intensity functions following an auditory or visual stimulus. Stimuli happen at time = 0 s.

DriPP - Results on audio-visual paradigm

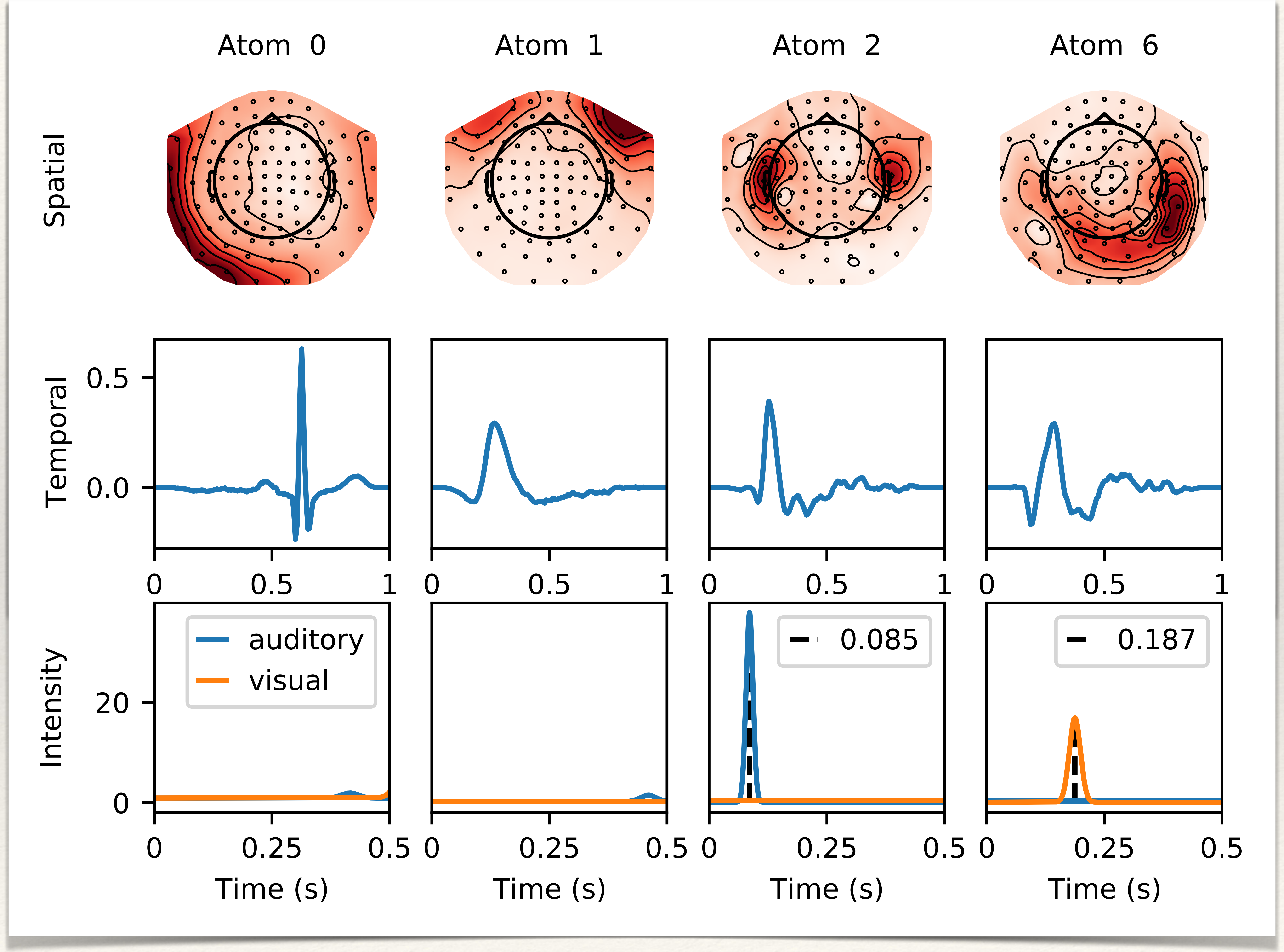
- Subject is randomly exposed to either visual or auditory stimuli for 6 mins.
- MEG Data collected on 306 sensors, 40 atoms of 0.5 s extracted



Spatial and temporal patterns of 4 atoms from MNE sample dataset, and their respective estimated intensity functions following an auditory or visual stimulus. Stimuli happen at time = 0 s.

DriPP - Results on audio-visual paradigm

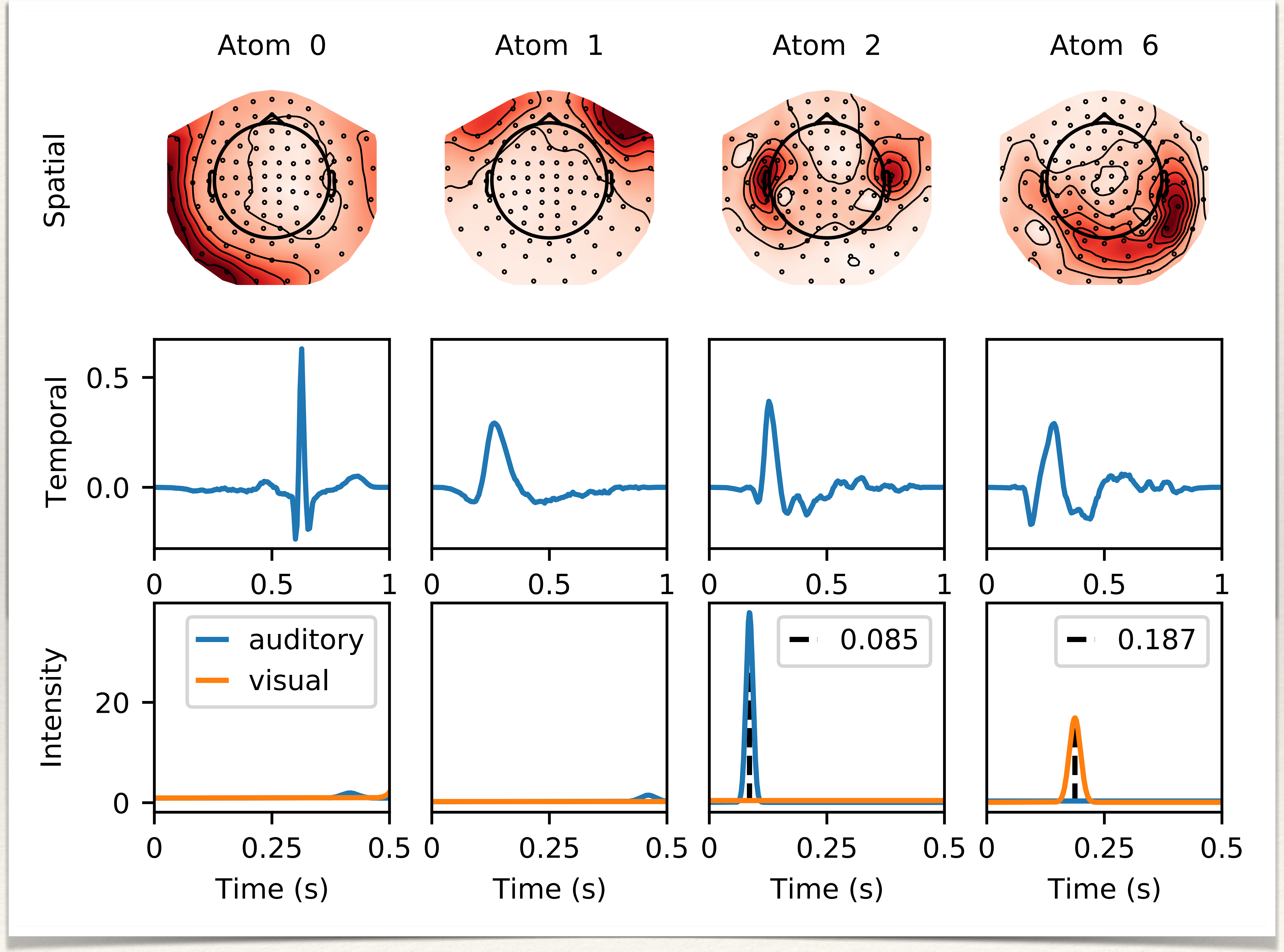
- Subject is randomly exposed to either visual or auditory stimuli for 6 mins.
- MEG Data collected on 306 sensors, 40 atoms of 0.5 s extracted
- Heartbeat (atom 0) and eye-blink (atom 1) artifacts are not link to any stimuli.



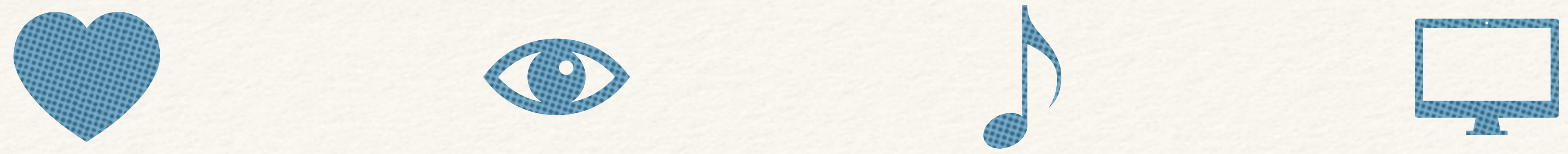
Spatial and temporal patterns of 4 atoms from MNE sample dataset, and their respective estimated intensity functions following an auditory or visual stimulus. Stimuli happen at time = 0 s.

DriPP - Results on audio-visual paradigm

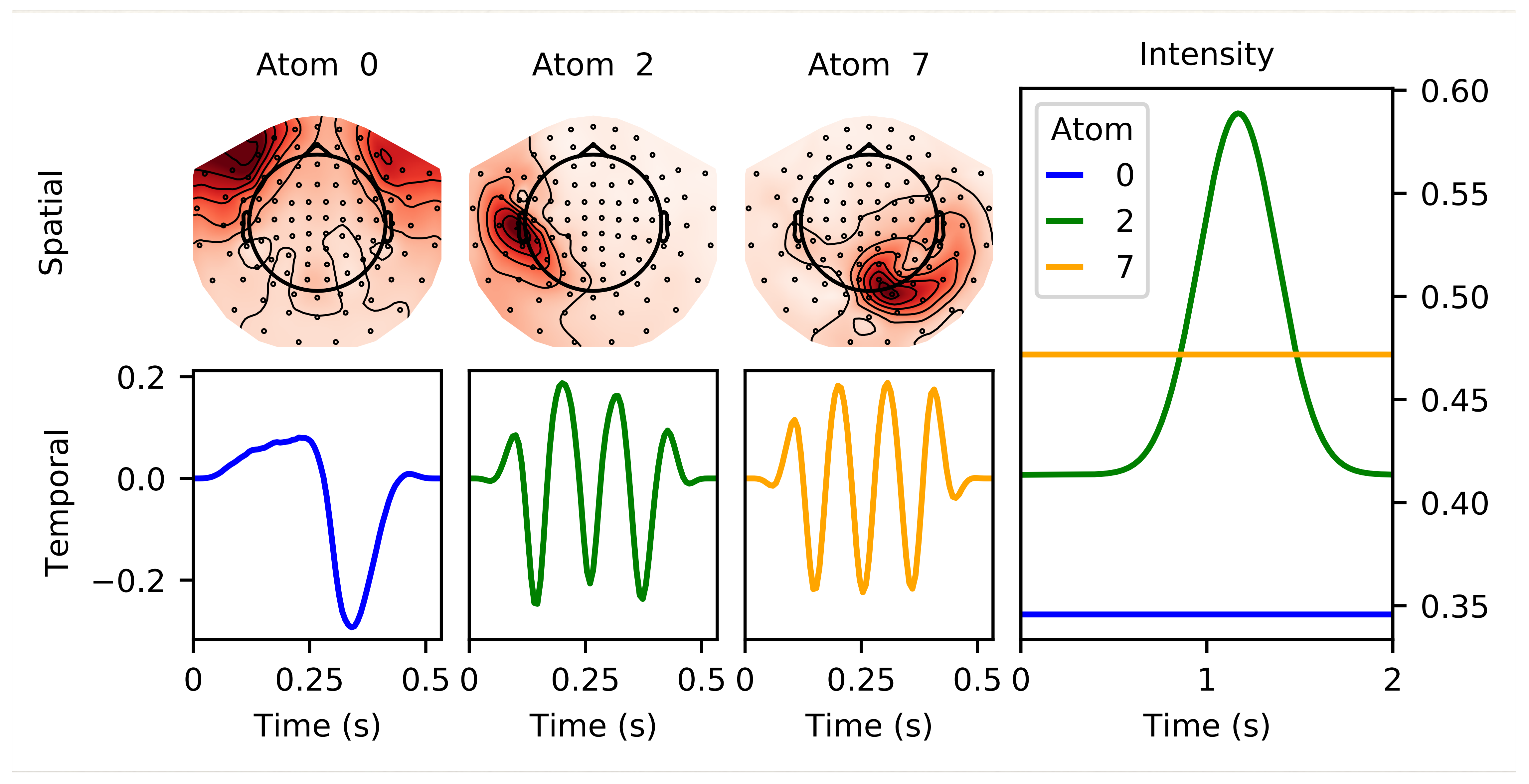
- Subject is randomly exposed to either visual or auditory stimuli for 6 mins.
- MEG Data collected on 306 sensors, 40 atoms of 0.5 s extracted
- Heartbeat (atom 0) and eye-blink (atom 1) artifacts are not link to any stimuli.
- A visual stimulus will induce a neural response of pattern atom 6, with a 187 ms latency in average

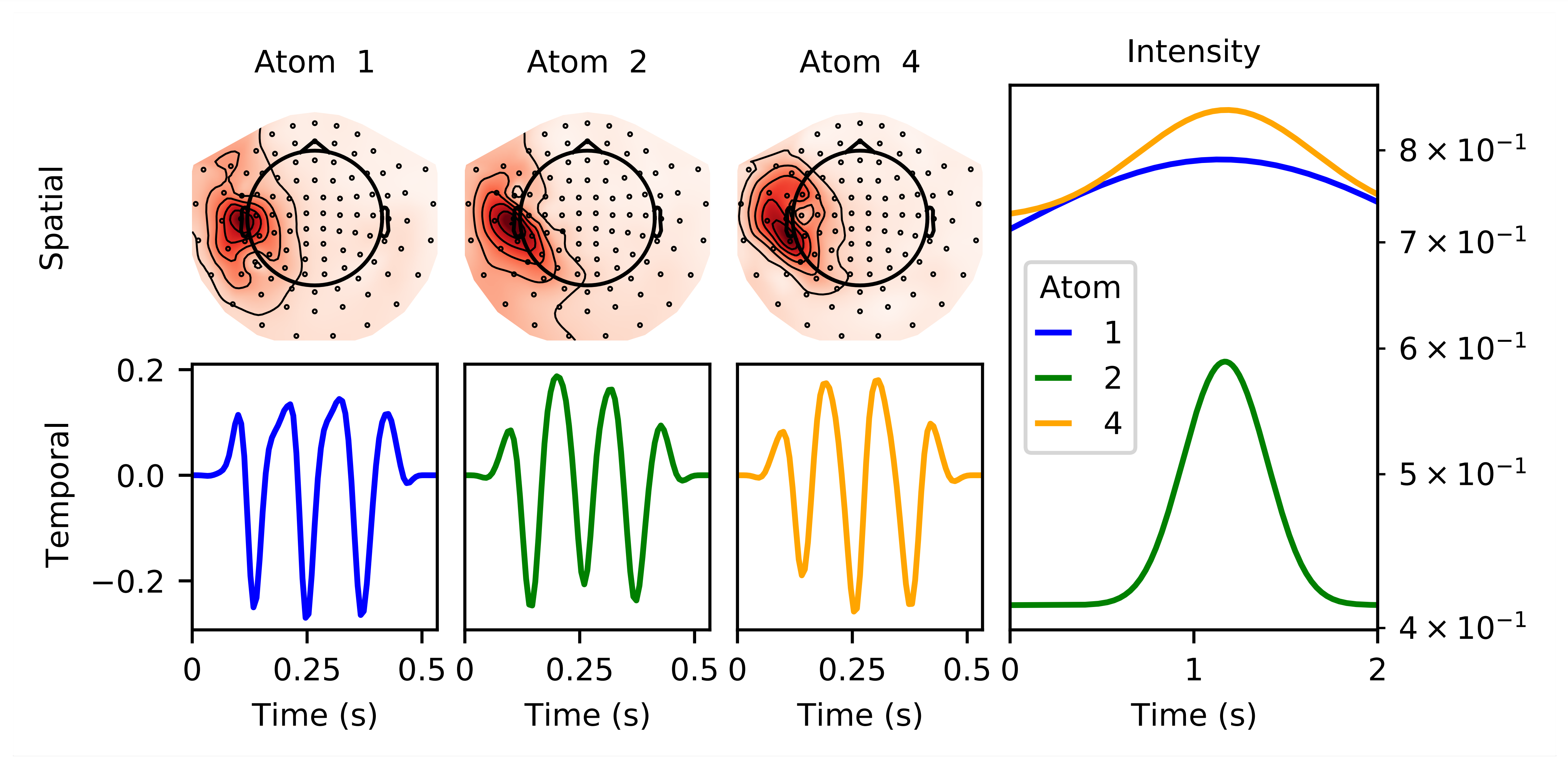


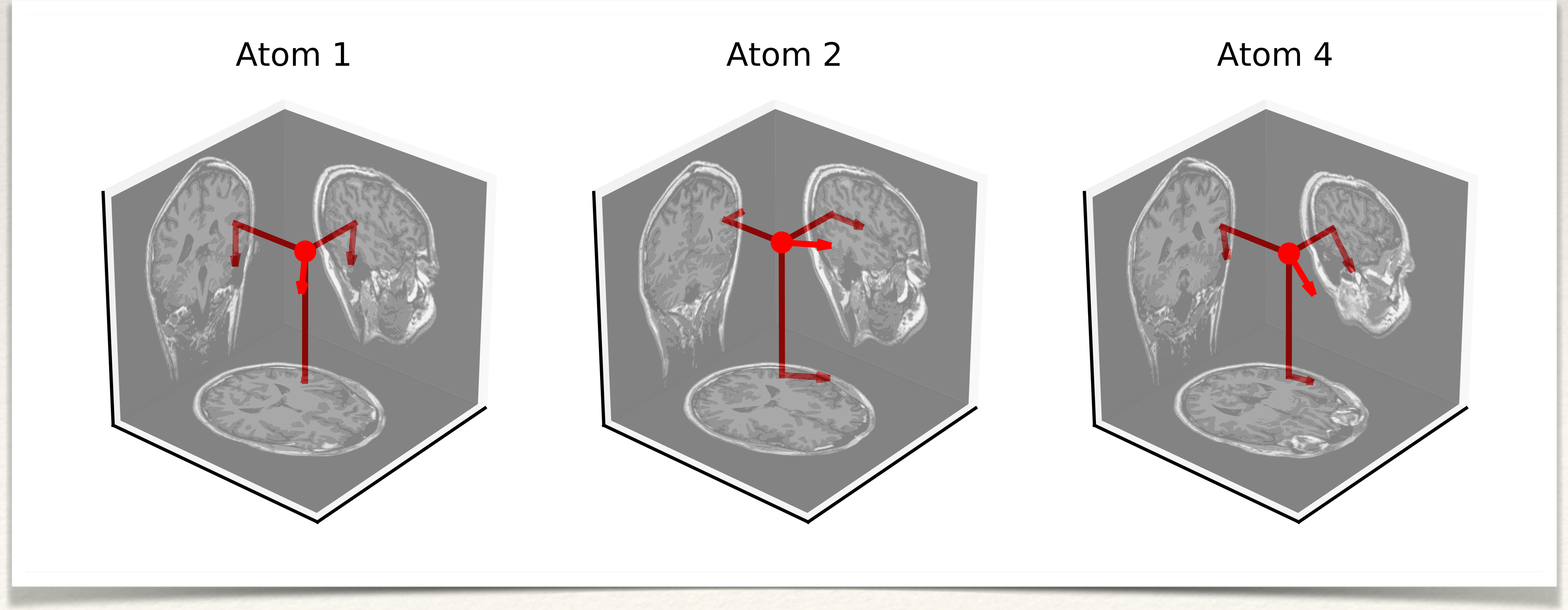
Spatial and temporal patterns of 4 atoms from MNE sample dataset, and their respective estimated intensity functions following an auditory or visual stimulus. Stimuli happen at time = 0 s.



DriPP - Results on MNE somato(sensory)



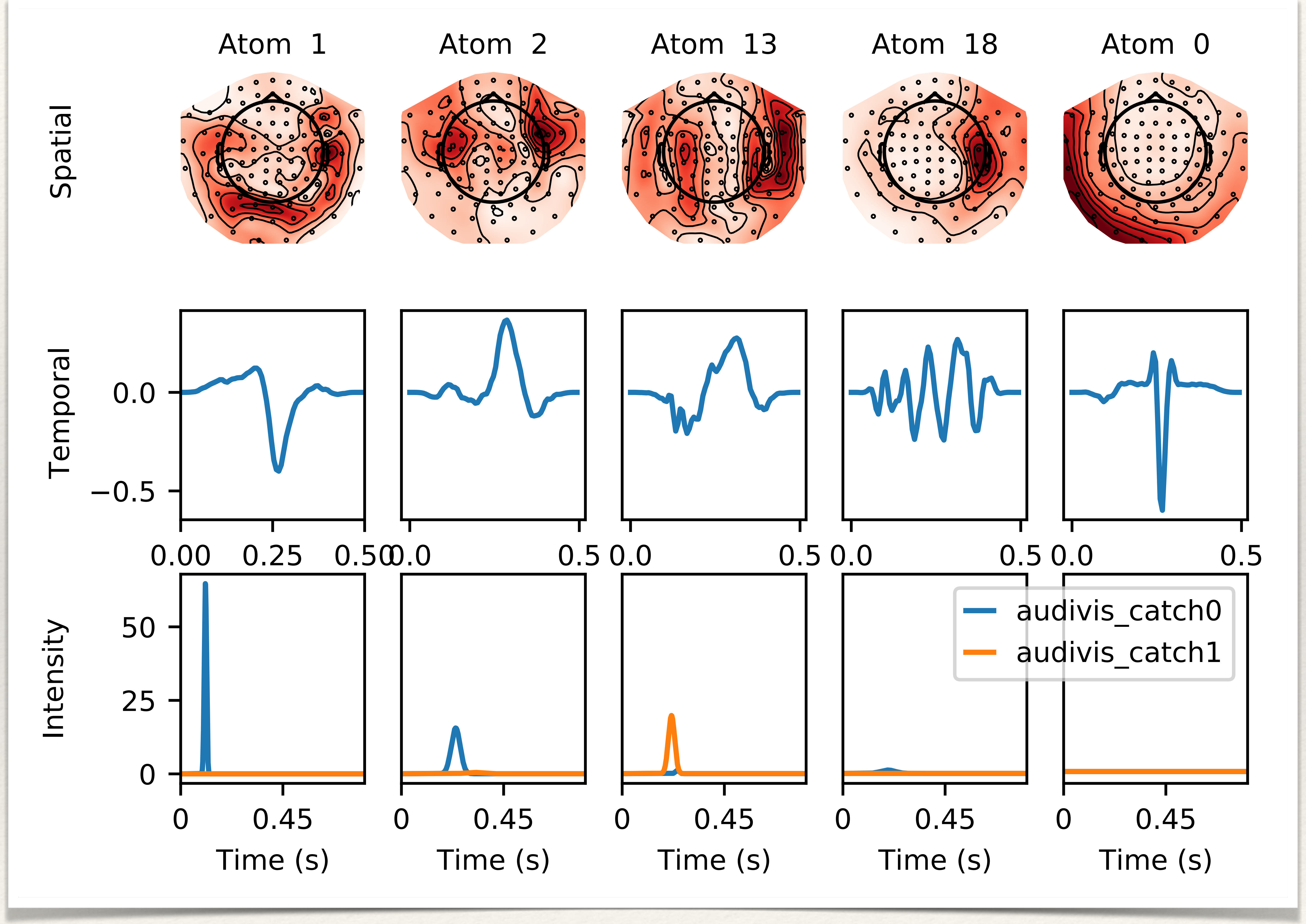




- The eye-blink artifact (atom 0) and the α-wave (atom 7) are not linked to the stimulus.
- A somatosensory stimulus will induce a neural response similar to atom 2, with a mean latency of 1 s.
- Atoms 1, 2 and 4 correspond to a μ-waves and their respective intensity functions characterize an induced response.

DriPP - Results on Cam-CAN

- catch_0: with 4 **auditory** unimodal stimuli
- catch_1: with 4 **visual** unimodal stimuli
- atom 0: heartbeat artefact



subject CC520597, a 64.25-year-old male

Atoms are ordered by their bigger ratio α/μ

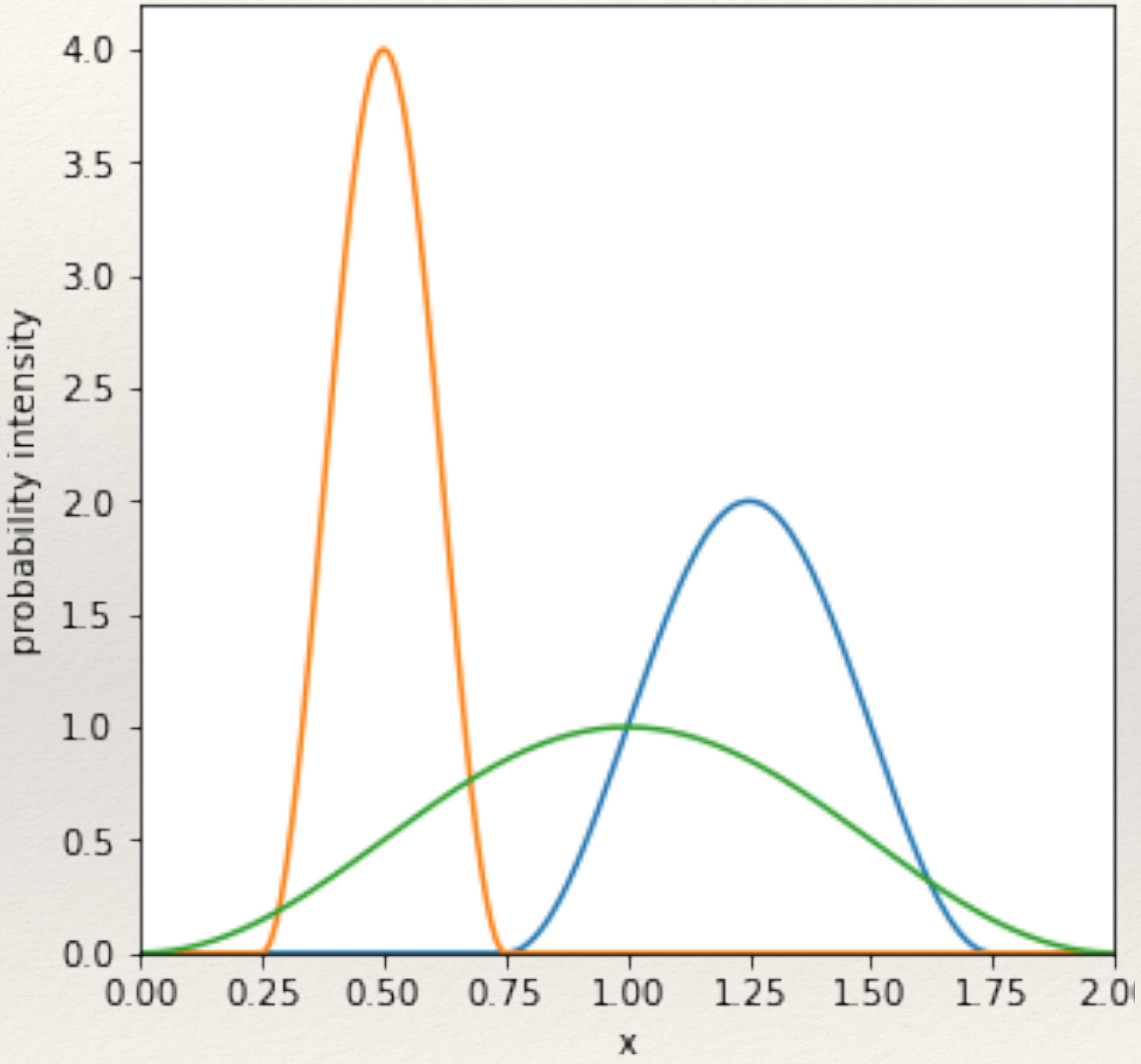
From TG to RC

- A drawback of Truncated Gaussian kernel: the support $[a, b]$ need to be pre-determined

- Raised Cosine kernel as a self-determined support:

$$\kappa(x) = \frac{1}{2\sigma}\left[1 + cos\left(\frac{x-m}{\sigma}\pi\right)\right] 1_{x \in [m-\sigma, m+\sigma]}$$

New Fast Discretized Inference (FaDIn) method for Hawkes parametric kernels



Probability density functions for the raised cosine distribution for different sets of parameters.

From TG to RC

- L2-based loss

$$\mathscr{L}\left(\theta, \mathscr{F}_T\right) = \frac{1}{N_T} \sum_{i=1}^{p} \left(\int_0^T \lambda_i(s)^2 \mathrm{d}s - 2 \sum_{t_n^i \in \mathscr{F}_T^i} \lambda_i\left(t_n^i\right) \right)$$

$N_T = \sum_{i=1}^{p} N_T^i$ the total number of timestamps across all processes

- Such loss allows pre-computations to enhance computation time
- Study of the discretization impact on estimates quality:

$$\left\| \widehat{\theta_\Delta} - \theta^* \right\|_2 \leq \left\| \widehat{\theta_c} - \theta^* \right\|_2 + \left\| \widehat{\theta_\Delta} - \widehat{\theta_c} \right\|_2$$

Stearman G., Allain C., Gramfort A., Moreau T., *FaDIn: Fast Discretized Inference For Hawkes Processes With General Parametric Kernels*, submitted to ICLR 2023

From TG to RC

- L2-based loss

$$\mathscr{L}\left(\theta, \mathscr{F}_T\right) = \frac{1}{N_T} \sum_{i=1}^{p} \left(\int_0^T \lambda_i(s)^2 \mathrm{d}s - 2 \sum_{t_n^i \in \mathscr{F}_T^i} \lambda_i\left(t_n^i\right) \right)$$

$N_T = \sum_{i=1}^{p} N_T^i$ the total number of timestamps across all processes

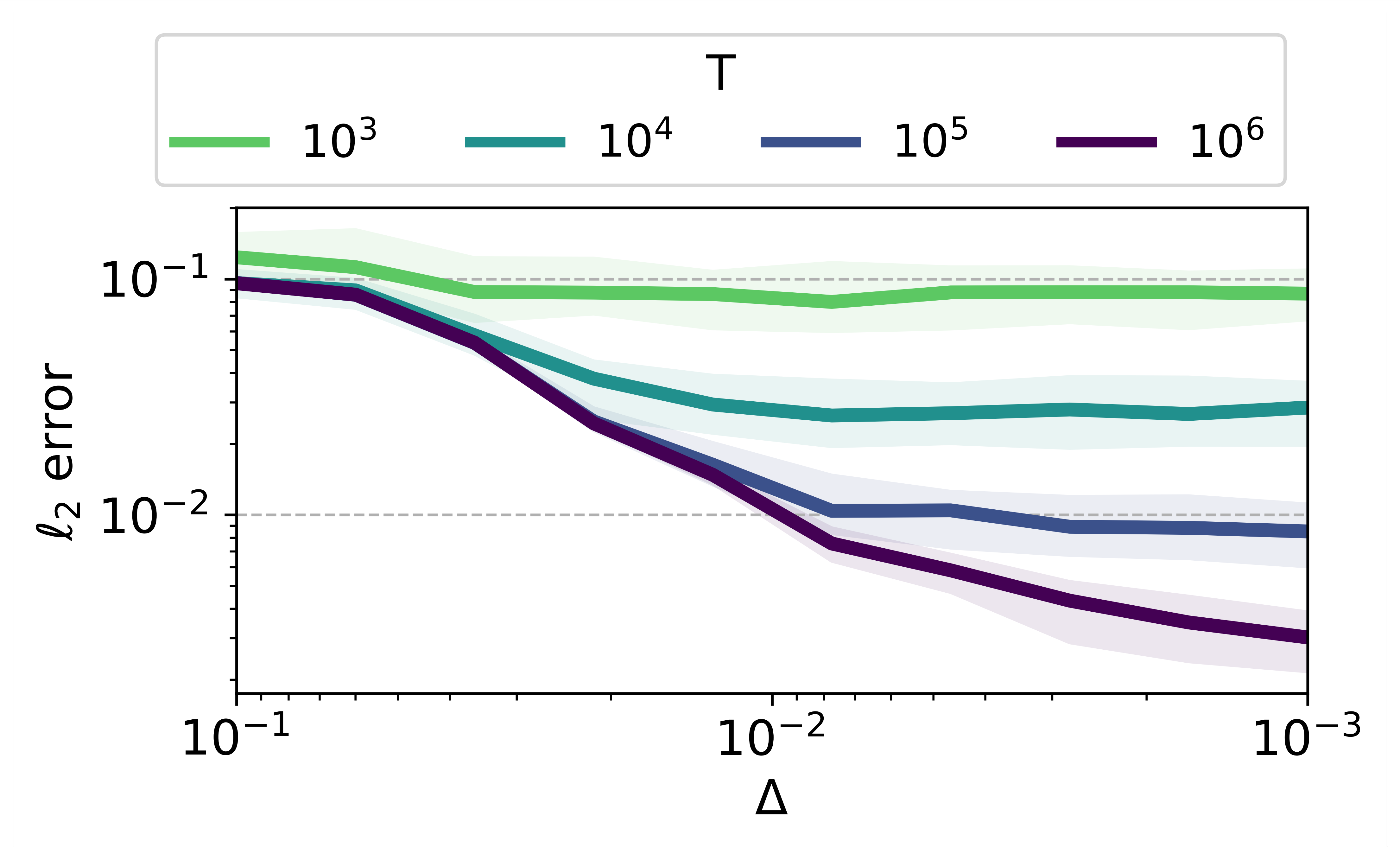
- Such loss allows pre-computations to enhance computation time
- Study of the discretization impact on estimates quality:

$$\left\| \widehat{\theta_\Delta} - \theta^* \right\|_2 \leq \left\| \widehat{\theta_c} - \theta^* \right\|_2 + \boxed{\left\| \widehat{\theta_\Delta} - \widehat{\theta_c} \right\|_2}$$

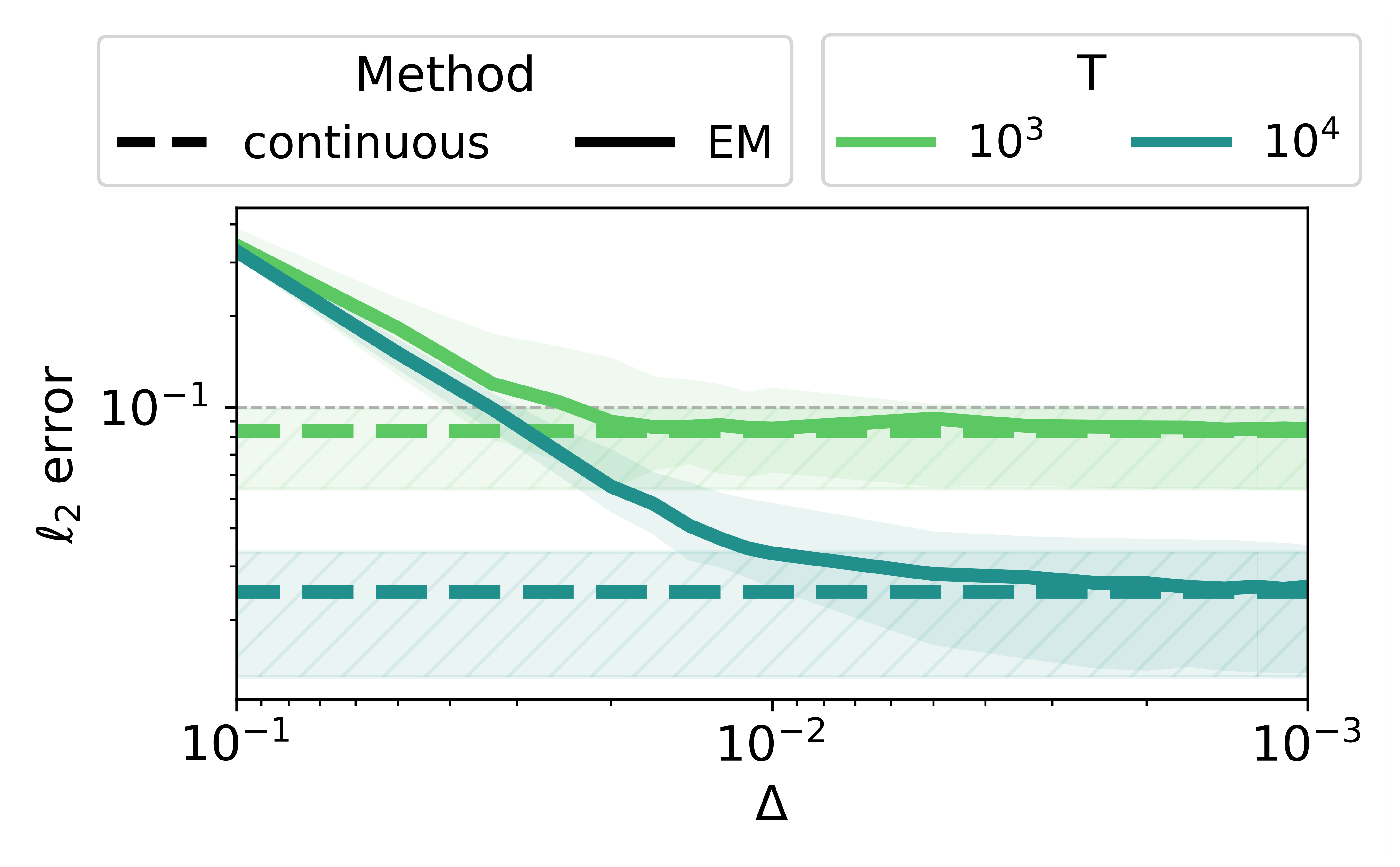
Stearman G., Allain C., Gramfort A., Moreau T., *FaDIn: Fast Discretized Inference For Hawkes Processes With General Parametric Kernels*, submitted to ICLR 2023

From TG to RC

Δ: discretization parameter

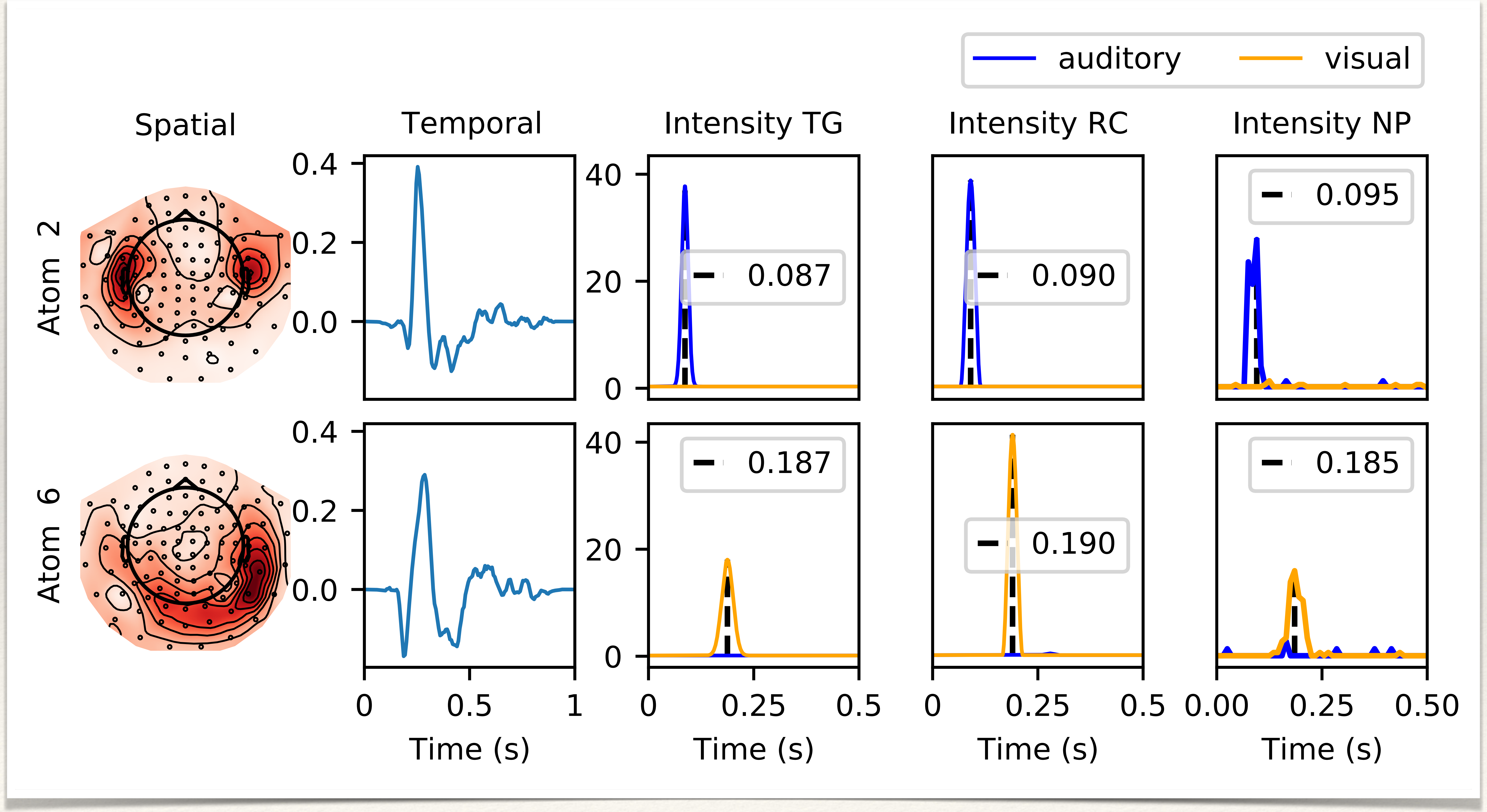


FaDIn



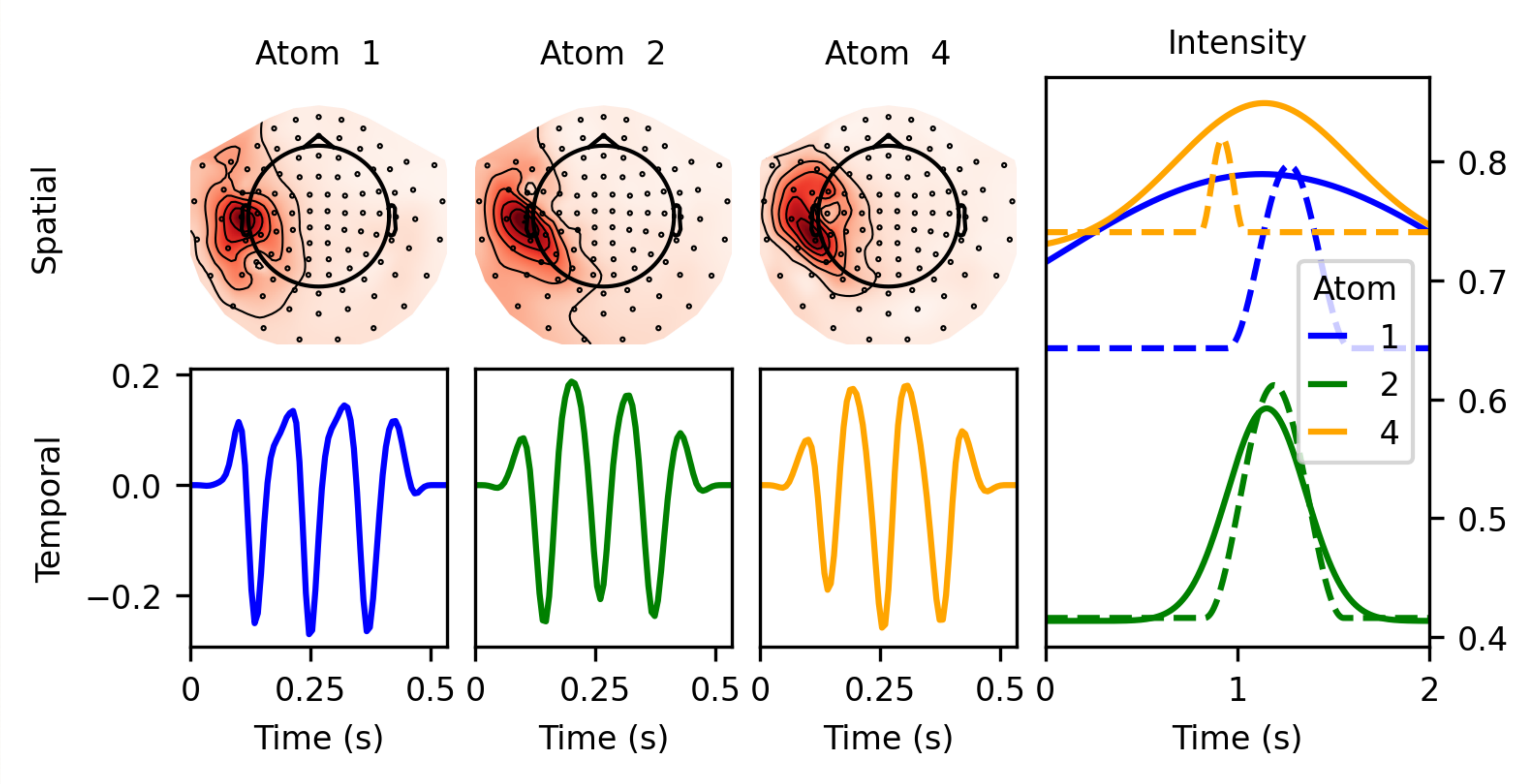
EM

From TG to RC



- Similar latencies
- Non-parametric is less interpretable

From TG to RC



Atom 1
Atom 2
Atom 4
Intensity
Spatial
Temporal
0.2
0.0
−0.2
0
0.25
0.5
Time (s)
0.8
0.7
0.6
0.5
0.4
1
2
Atom
1
2
4

Conclusion

- Direct statistical characterisation of when and how each stimulus is responsible for the occurrences of neural responses.
- Unified approach to extract waveforms and automatically select the ones that are likely to be triggered by the considered stimuli.
- Well adapted to M/EEG experiments that have tens or hundreds of events at most.
- **Futur work**: CDL on population + use TPP for more accurate atom extraction