Learning by trial-and-error in the cerebellum

V Hakim
(LPS ENS, Paris)

Boris Barbour, Guy Bouvier, J Aljadeff (U Chicago), C Bimbard, N Brunel (Duke U), C Clopath (Imperial Col), JP Nadal, J Ranft

Brain Workshop, KITP, september 2018

Learning and credit-assignment

Example: learning a complex movement (walking, riding a bike, skiing….)

A task generally composed of many coordinated sub-tasks but the error is often only global.

Many muscle involved, which muscle command should be modified to avoid falling?

A supplementary difficulty : the sign of the error is not known i.e. was one particular muscle too contracted or not enough?

A possible solution : trial and error.

-During the task, perturb a few elements (cells, synapses,...)
(...Doya and Sejnowski, 1995;...; Seung, 2004; Fiete et al, 2007;..)

-> Changes in the task performance can be attributed to the elements that were perturbed

An attractive framework that has been primarily considered in the context of song learning in birds but detailed implementation still to be worked out (how is the song evaluated, where are the sites of plasticity,...?)

Here : examine the question for learning complex coordinated movements at the level of the cerebellum.

Learning by trial-and-error/stochastic gradient descent in the cerebellum?

The questions we would like to anwer :

-During the task, a few cells should be perturbed ->
Question I : what is the source of this perturbation?

-Evaluation of the current performance needed to evaluate whether the perturbation has improved or deteriorated it ->
Question II : where is current performance stored/how is the comparison made?

Question III: characteristics of the resulting learning algorithm (convergence, speed, ...)?

Question IV: specific experimentally testable predictions?

Outline.

- Cerebellum and movement learning : cerebellar anatomy and the classical Marr-Albus theory

- A new proposal

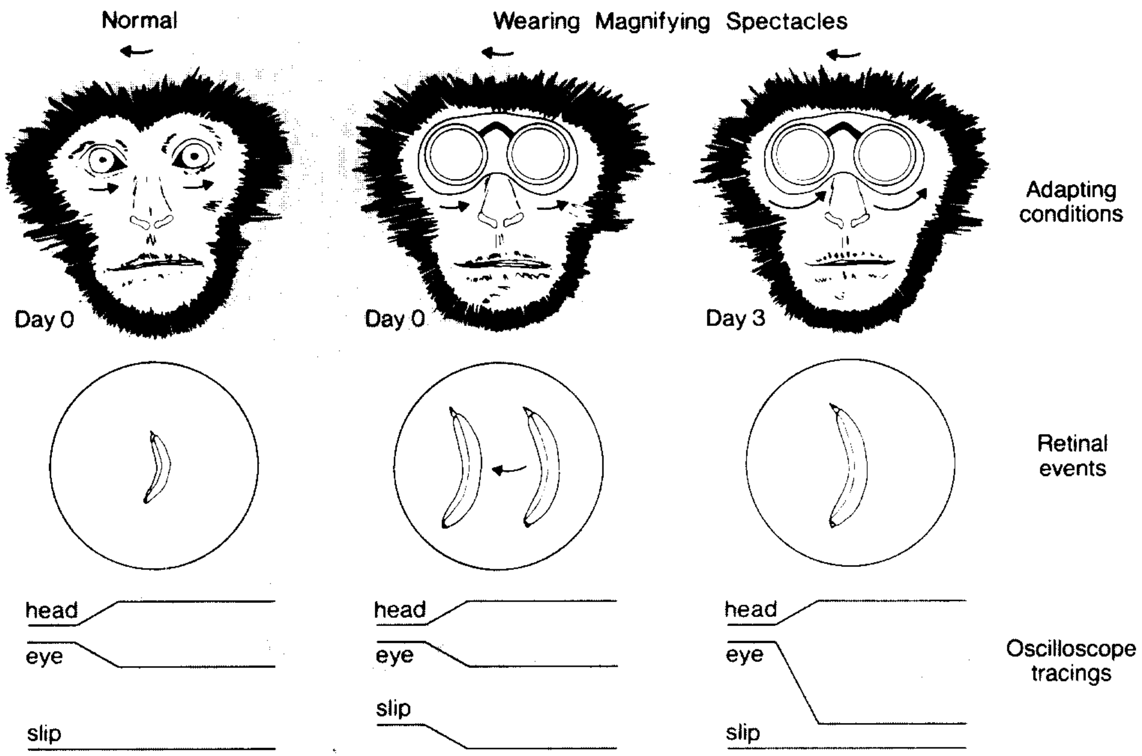
- Some experimental results

- Simple mathematical implementations and analysis

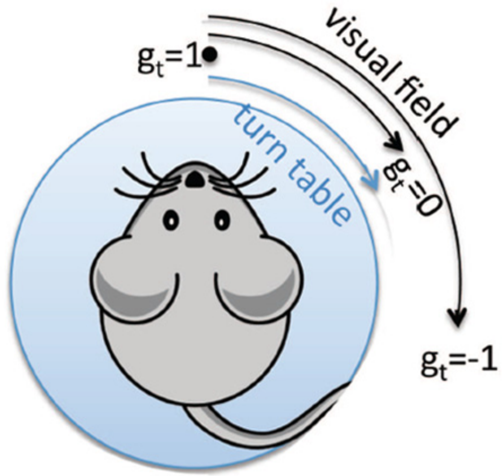
- Further theoretical questions and experiments

The cerebellum is a main site of motor learning

Vestibulo-ocular reflex (VOR)



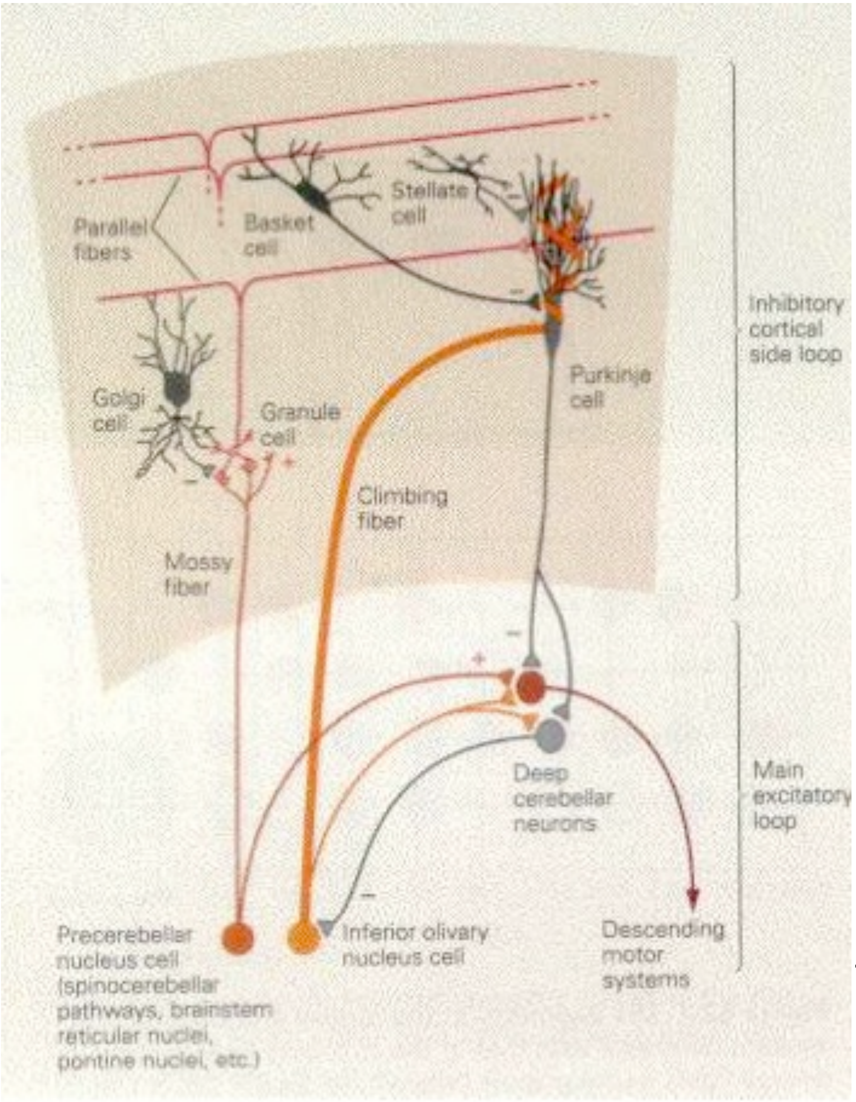
Lisberger, TINS 1988

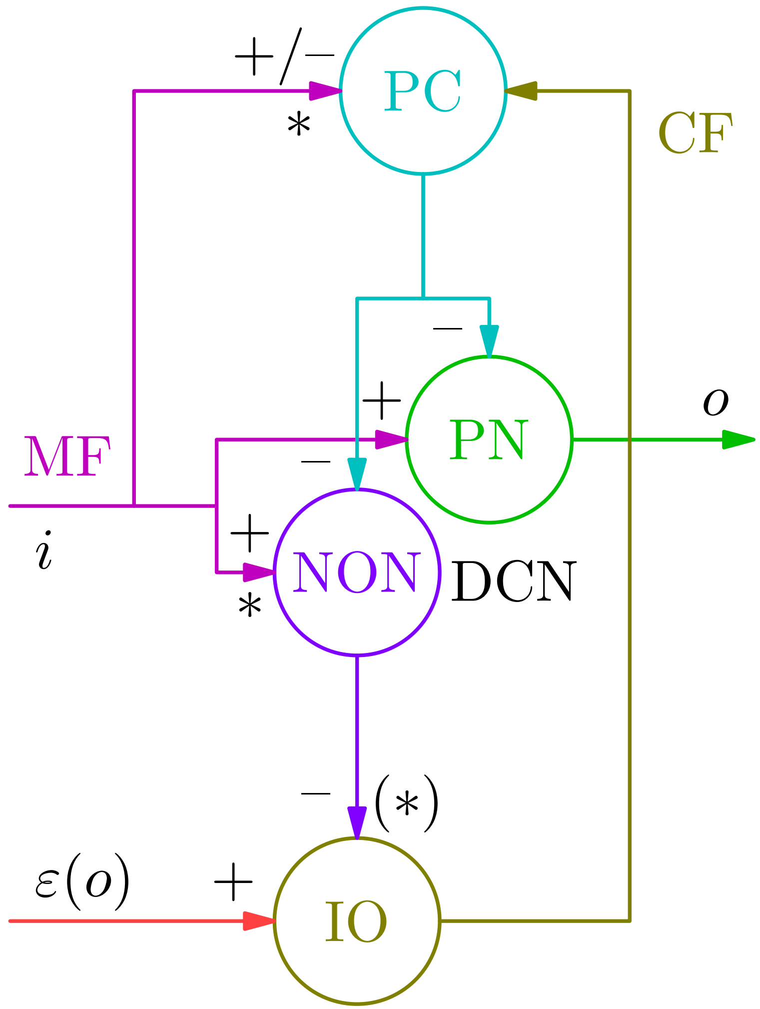


De Zeeuw et al, Neuron (2010), ...; Nat Rev Neur (2012)

Learning a gain change in the VOR is dependent on cerebellum

Cerebellum circuitry



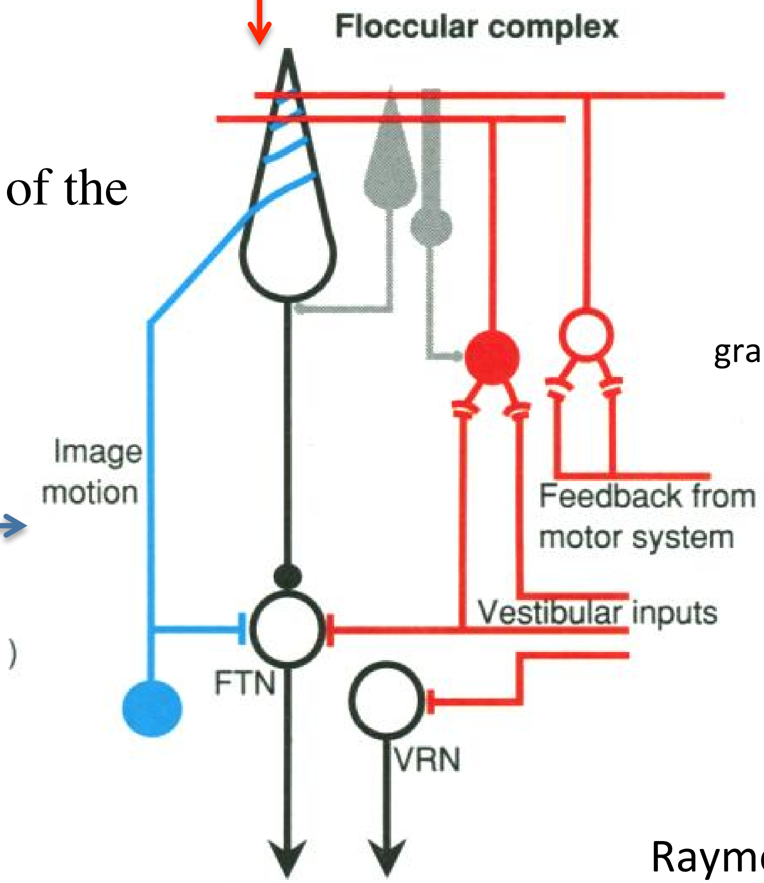


Two types of spikes in Purkinje cells (PC):
-simple spikes (SS) from granule cells/parallel fibers
-complex spikes (CS) from inferior olive/climbing fibers

Classic theory of cerebellar learning (Marr-Albus-Ito):

Plasticity of the parallel fibre (PF)-Purkinje cell synapse

under the control of the
inferior olive
(climbing fibers)



granule cell

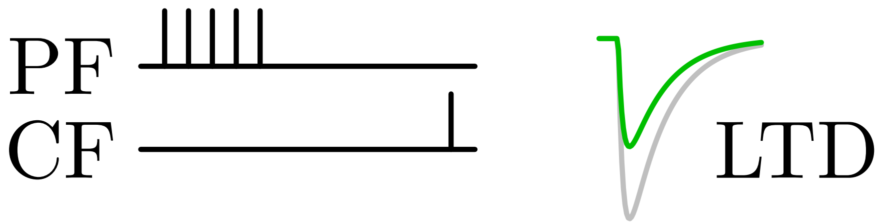
Supervised learning
(teacher: climbing fiber)

Raymond, Lisberger, Mauk, Science (1996)

Classic theory of cerebellar learning (Marr-Albus-Ito)

Climbing fibers : error signal -> depression of PF-PC synapse

In cerebellar slices:



Compensating potentiation : high frequency PF activity
with no error signal

Marr-Albus-Ito theory : some questions/difficulties

- In the theory of M-A-I, errors always lead to synaptic depression. Is it the case that errors always come from too strong synapses?

- the VOR is a simple task, sign of the error clear (given by retinal slip). For subtasks of complex movements less so, how can one determine if a synapse should be depressed or potentiated?

- For complex movements, how does the brain determine the precise source of error, the precise muscle that was not enough or too much contracted?

« Credit assignment problem » (Minsky, 1961)

The present proposal :

the cerebellum learns complex movements by performing stochastic gradient descent guided by a global estimate of the movement performance.

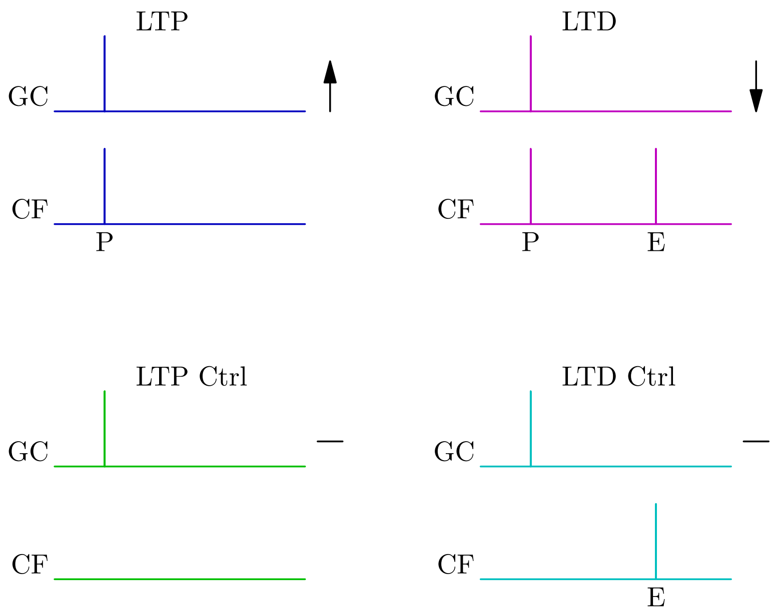
Proposal I : perturbation source

Climbing fiber- induced complex spikes serve both as perturbations of movements and as a signal errors (as classically proposed)

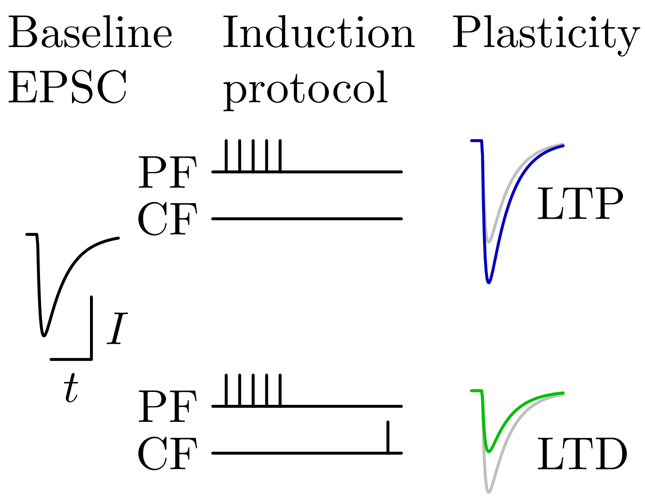
Olivary cells are spontaneously active at 1Hz

Consequences for synaptic plasticity rules

Predicted rules



Classical theory/results



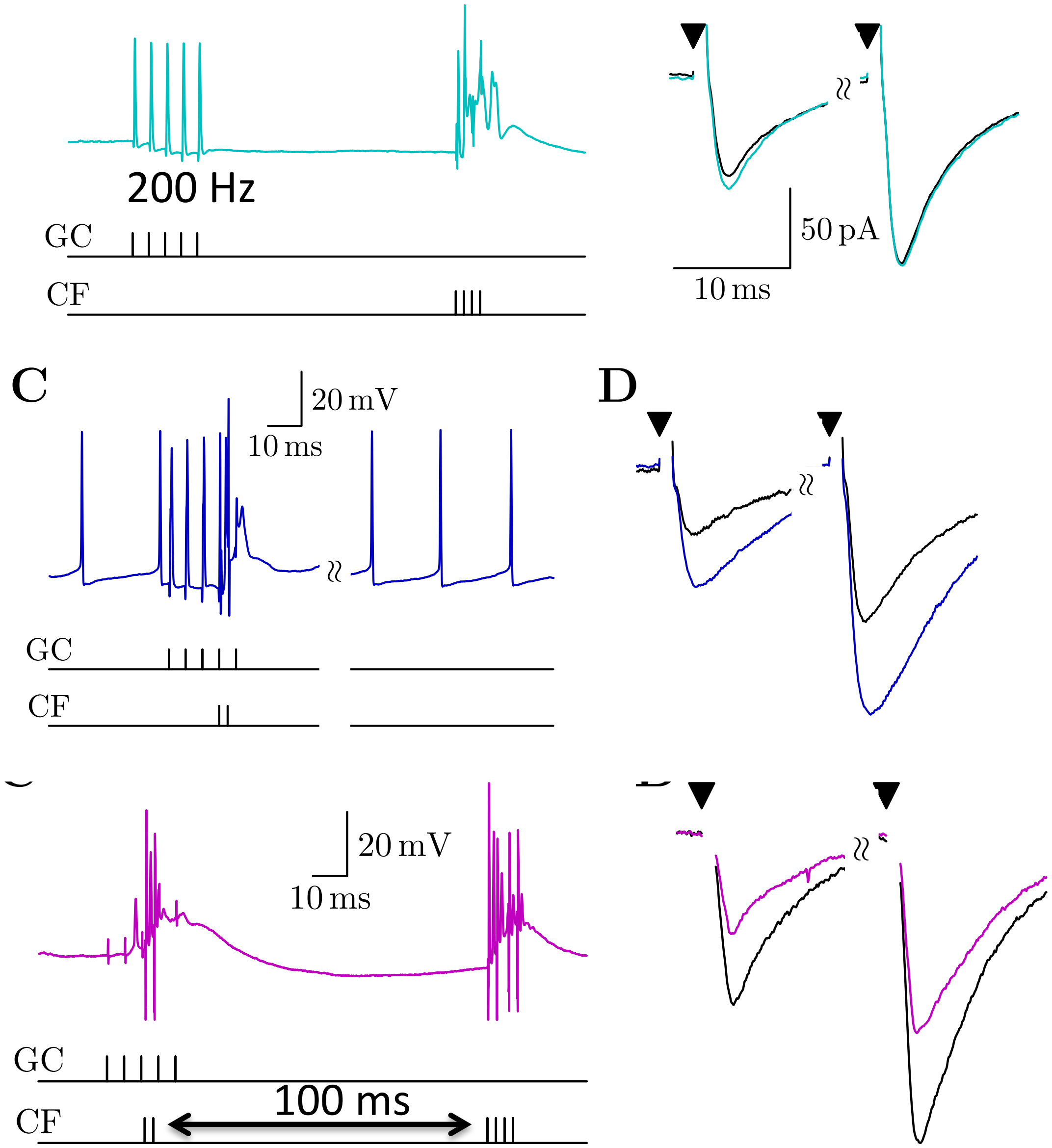
Change to standard protocols (to approach more physiological conditions) :
- -lowered extracellular calcium (2mM->1.5mM)
- -Weak stimulation of granule cell layer (instead of parallel fibers)
- -inhibition not blocked

Plasticity protocol

Test (pair of PF pulses before and after stimulation)

repeated every 2s between 10mn i.e. 300 times

No change



paired-pulse stimuli
(50ms interval)

before

Potentiation

after

after

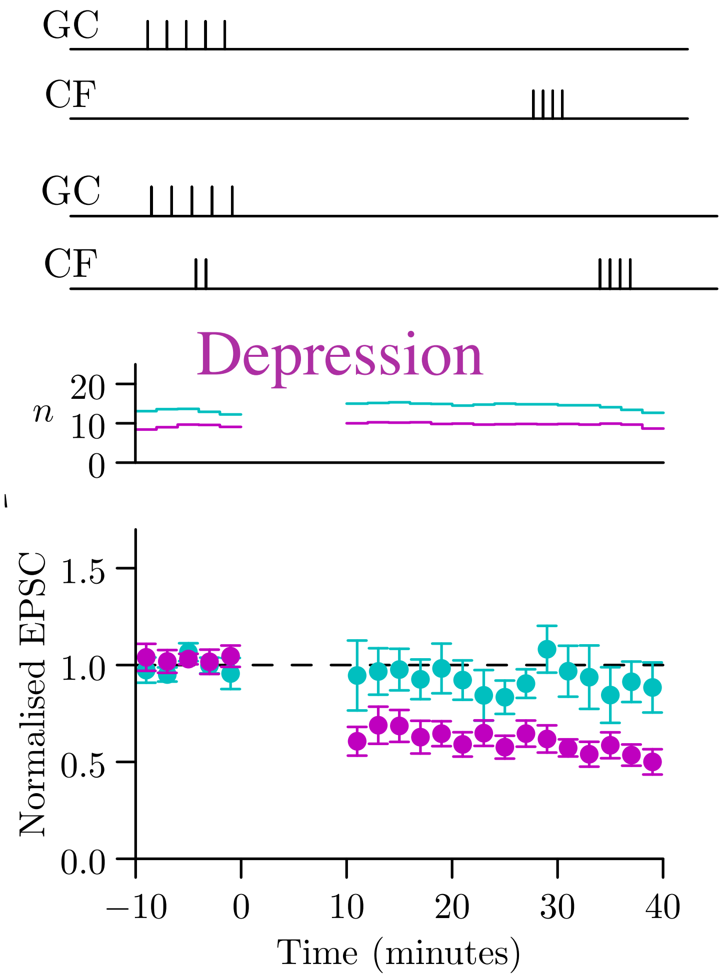
Depression

before

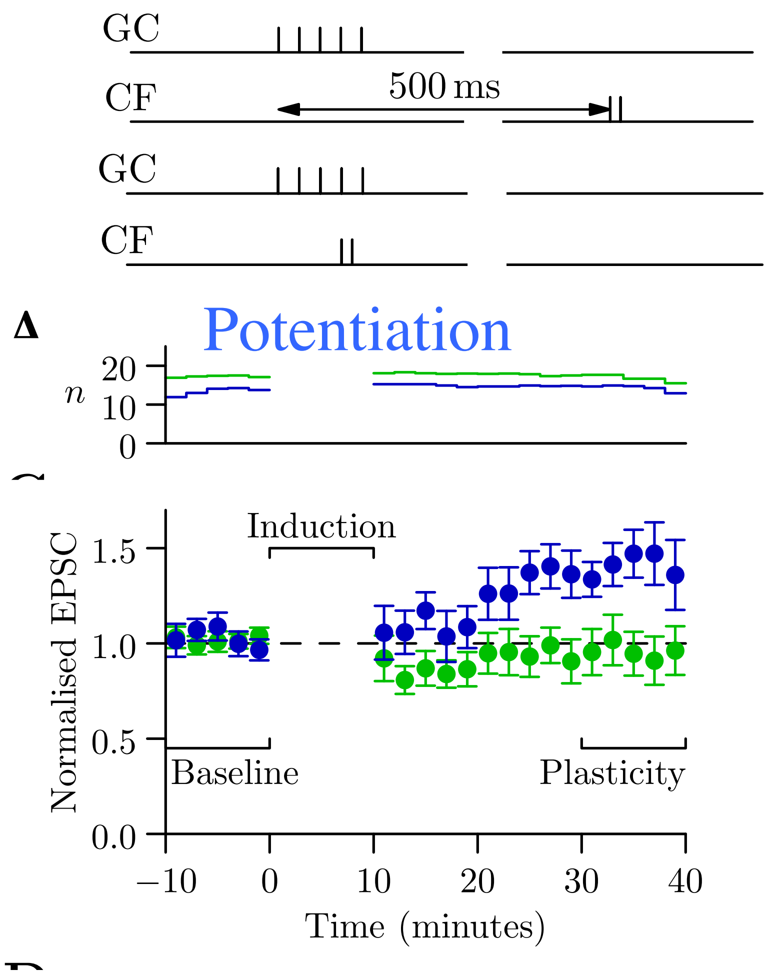
Guy Bouvier

Plasticity results summary

Guy Bouvier

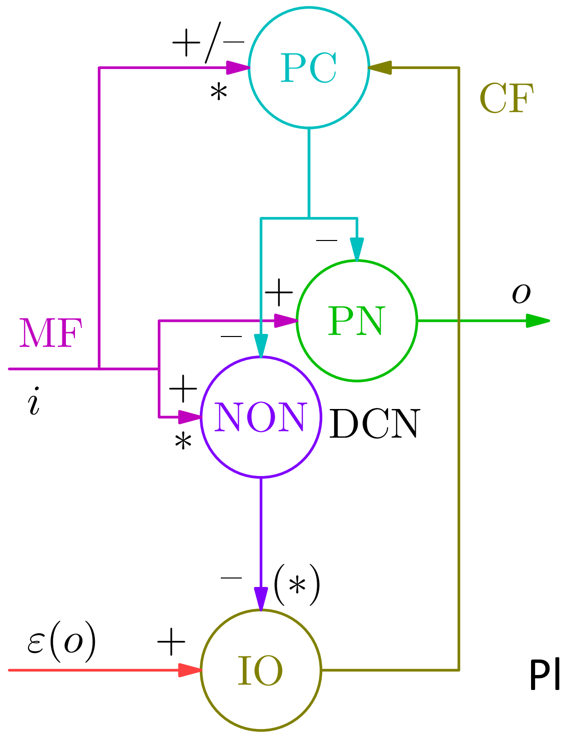


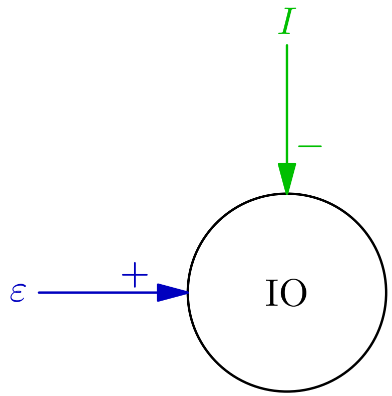
Controls

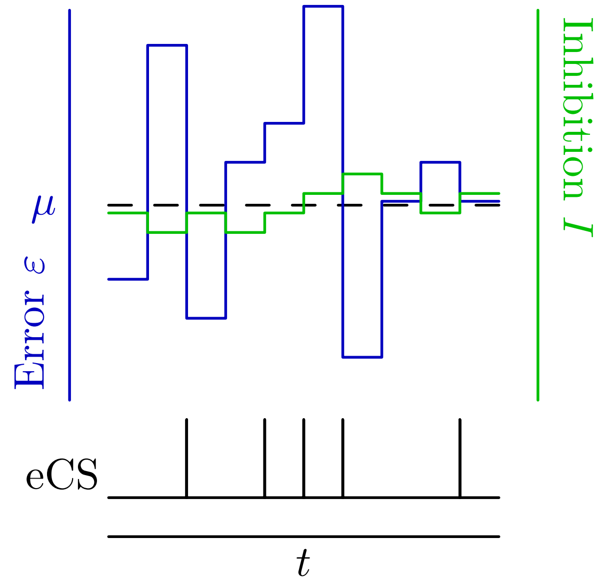


Proposal II : storage of current performance/ comparison with perturbed task

Proposal: the level of current performance is measured by the level of inhibition of olive neurones provided by NoN neurones







Plasticity site : - NoN –>IO synapses ?
- MF -> NoN synapses
(related plasticity observed at MF->PN synapses)

A toy model

N cells with firing rates $P_1, \ldots\ldots P_N$

Successful task firing rates $T_1, \ldots\ldots, T_N$

Mean error

$$E(t) = \frac{1}{N} \sum_i |P_i(t) - T_i|$$

Internal estimated value of the error : I

Learning task : bring the P_i's to their target values T_i's

A toy model

Learning algorithm (I):

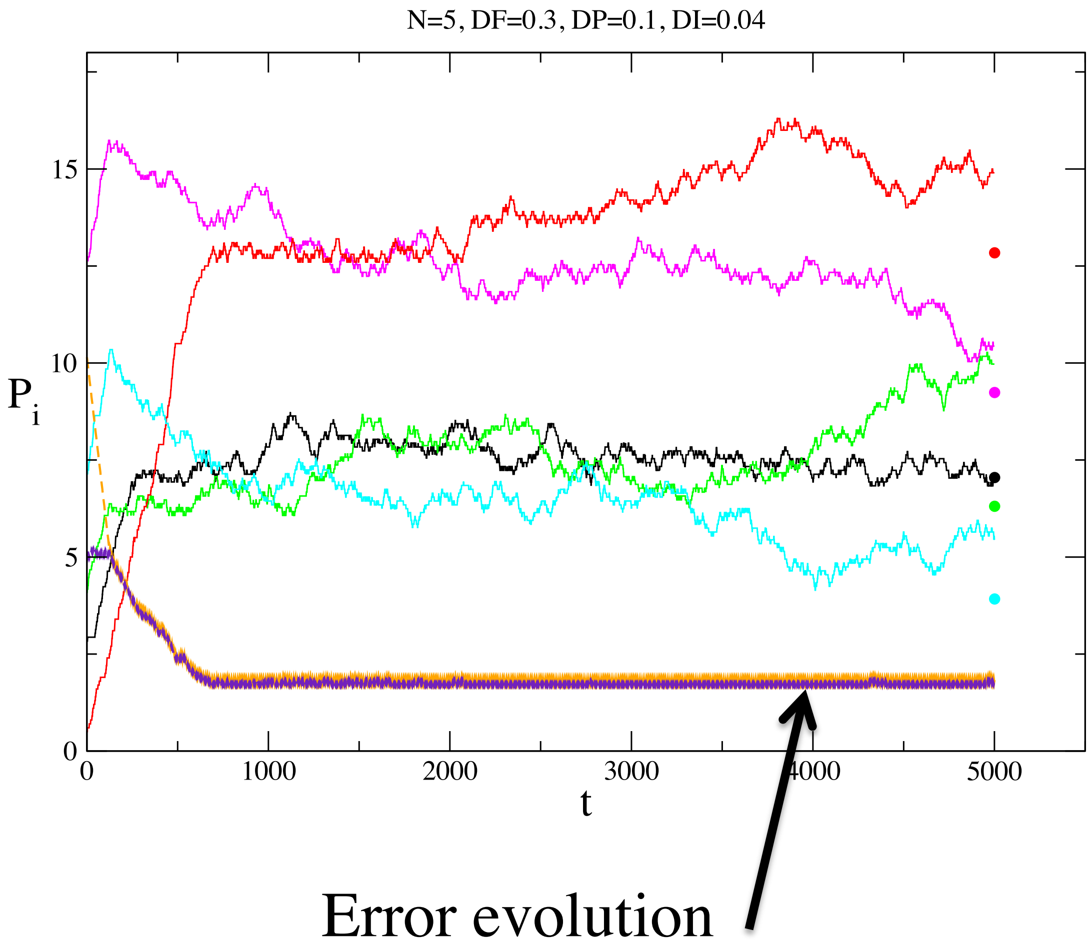
-choose randomly one of the rate i_c and perturb it by A>0: Pi_c=Pi_c+A
-Error with perturbation

$$E_p = \frac{1}{N} \sum_i |P_i(t) + \delta_{i,i_c} A - T_i|$$

-compare E_p estimated value of the error I

If E_p < I good to increase the rate, modify Pi_c -> Pi_c+ ΔP ,
update error estimation I -> I – ΔI

If E_p > I perturbation is bad, decrease Pi_c -> Pi_c- ΔP ,
I -> I + ΔI



With these moves only, conservation of:

$$C = I(t)\,\Delta P + \Delta I \sum_i P_i(t)$$ cannot converge!

Learning algorithm (II)

Remedy :

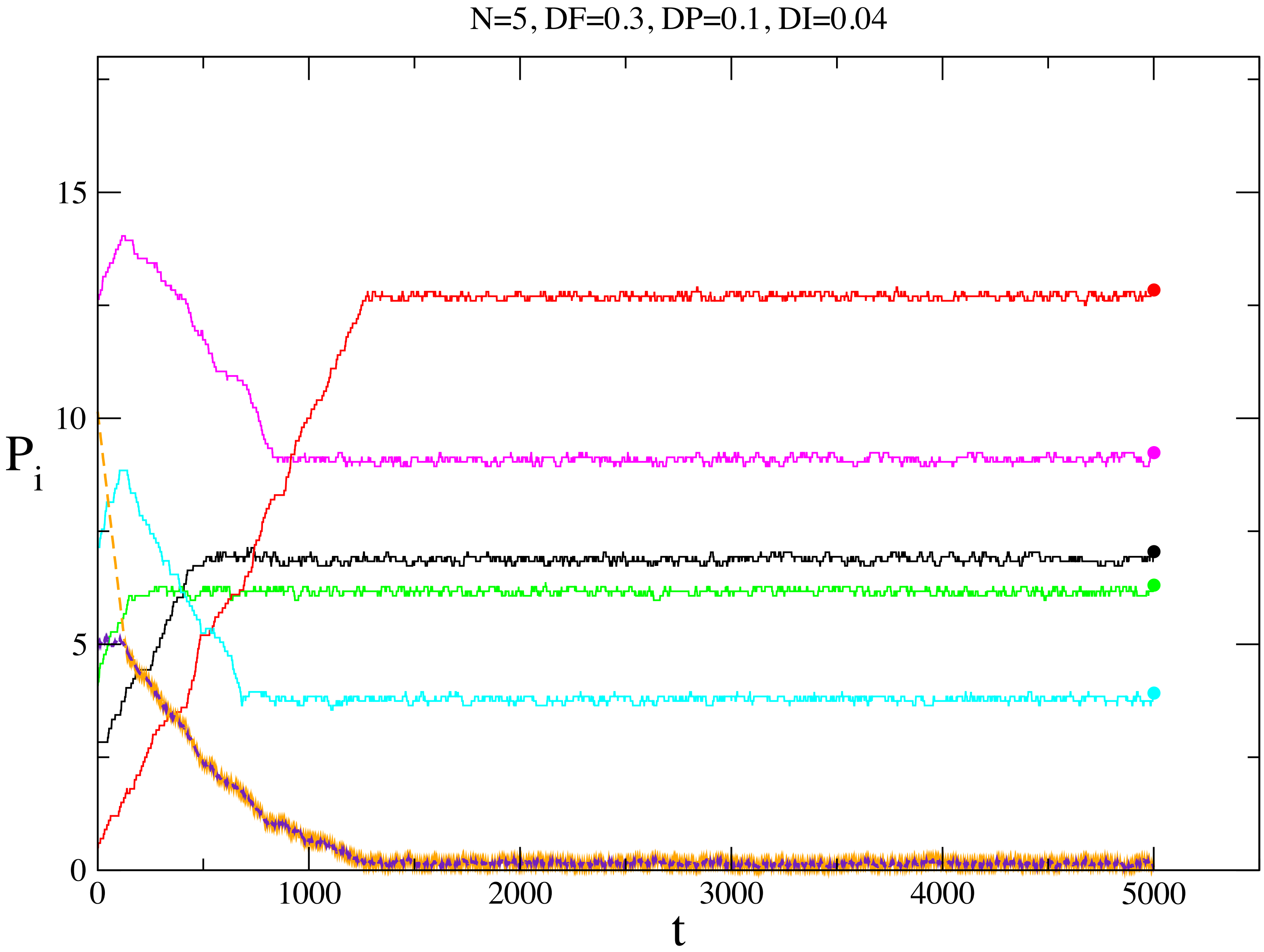
- make the previous moves with perturbation (« type A ») with probability ρ

- Make moves with no perturbation and just error estimation adjustment (« Type B ») with probability $1-\rho$

« Type B » moves:

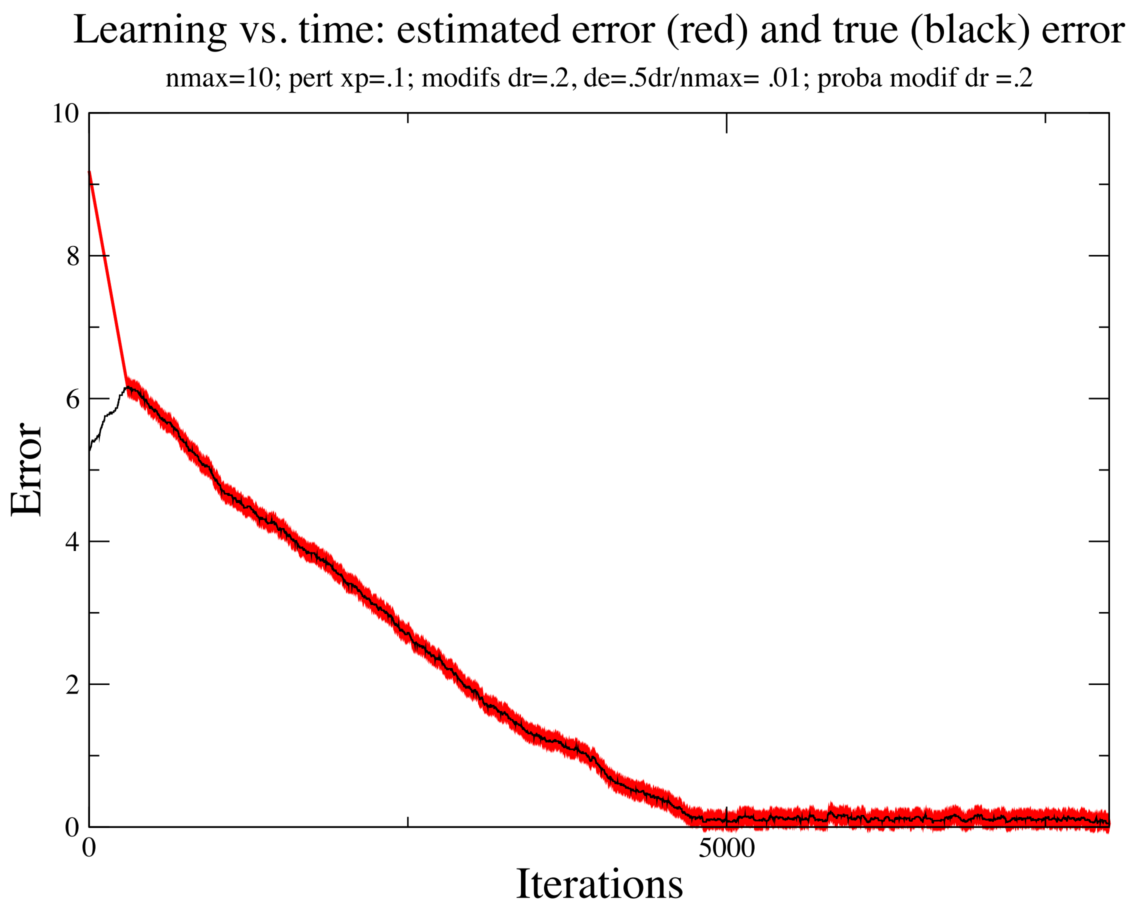
If $E < I$ then $I \rightarrow I - \Delta I$

If $E > I$ $I \rightarrow I + \Delta I$



N=5, DF=0.3, DP=0.1, DI=0.04
P_i
t
0
5
10
15
1000
2000
3000
4000
5000

Toy model : convergence



Three successive phases in the learning dynamics

Role of the different parameters? convergence speed?

The one cell case

One pattern, cell fires with rate P, desired firing rate T.
Current inhibition on olive I

Two types of trials :

-Perturbation with probability ρ, $P \to P+A$, $E = |P+A-T|$

If $E > I$, error CS, $P \to P - \Delta P$, $I \to I + \Delta I$

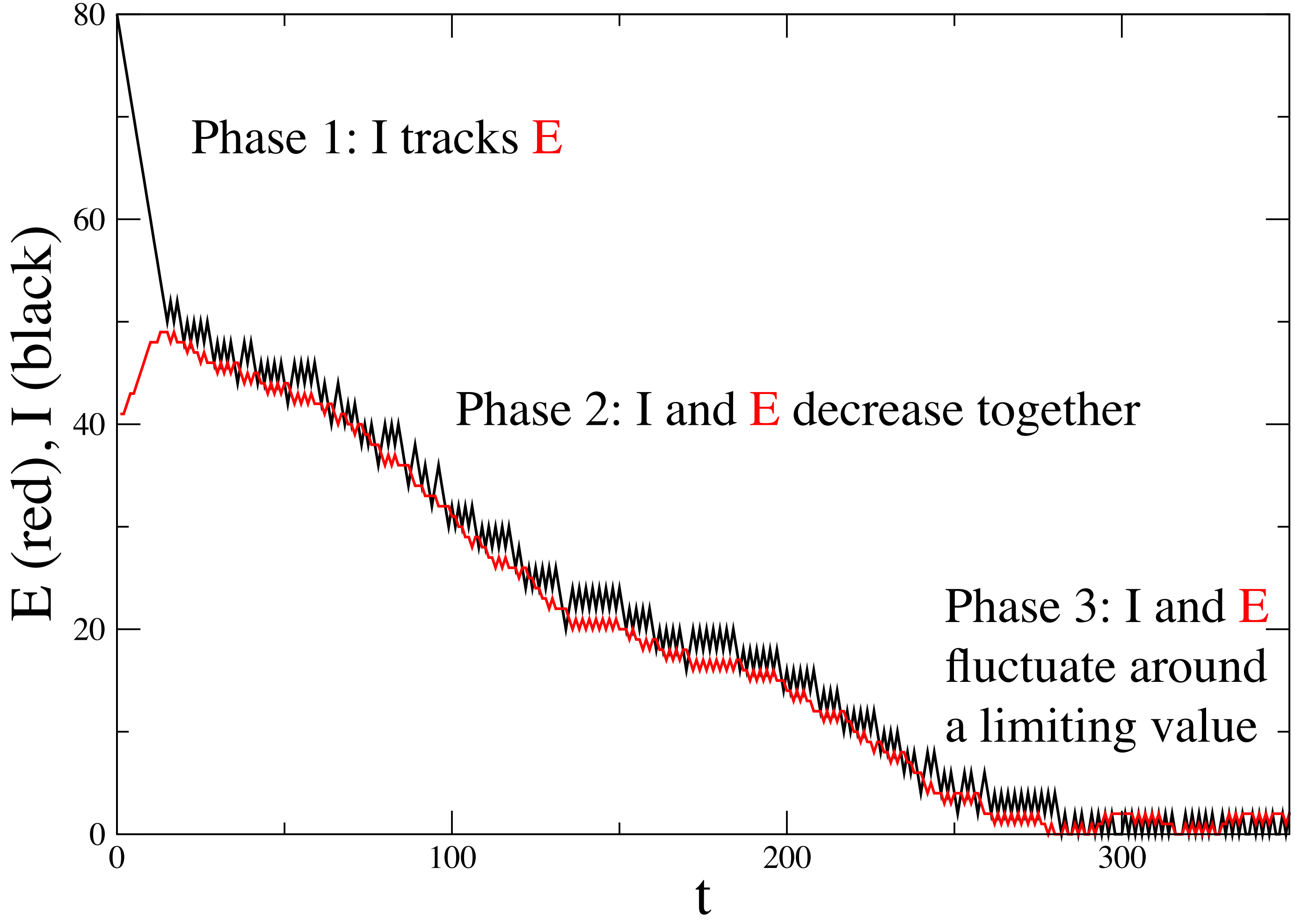
$E < I$, no error CS, $P \to P + \Delta P$, $I \to I - \Delta I$

-No perturbation with probability $1-\rho$, $E = |P-T|$

If $E > I$, error CS, $I \to I + \Delta I$ (P unchanged)

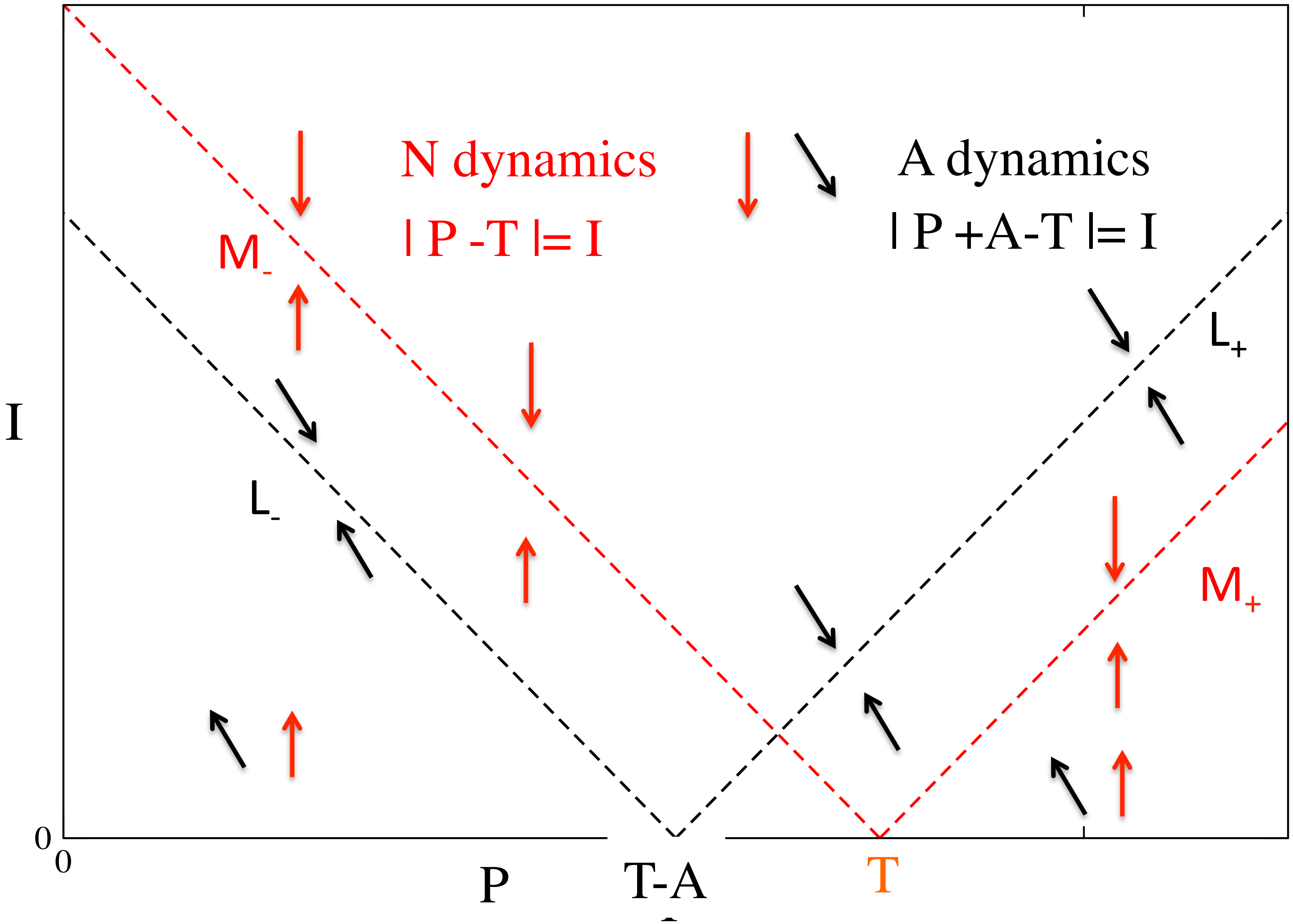
$E < I$, no error CS, $I \to I + \Delta I$ (P unchanged)

Convergence in 3 phases



- Origin of the 3 successive phases?
- Role of the different parameters?
- convergence speed?

One cell (P, I) phase-plane dynamics



- Two convergence « corridors »

Different parameter regimes
A, ρ, ΔP, ΔI
(large perturbation A>> ΔP, ΔI)

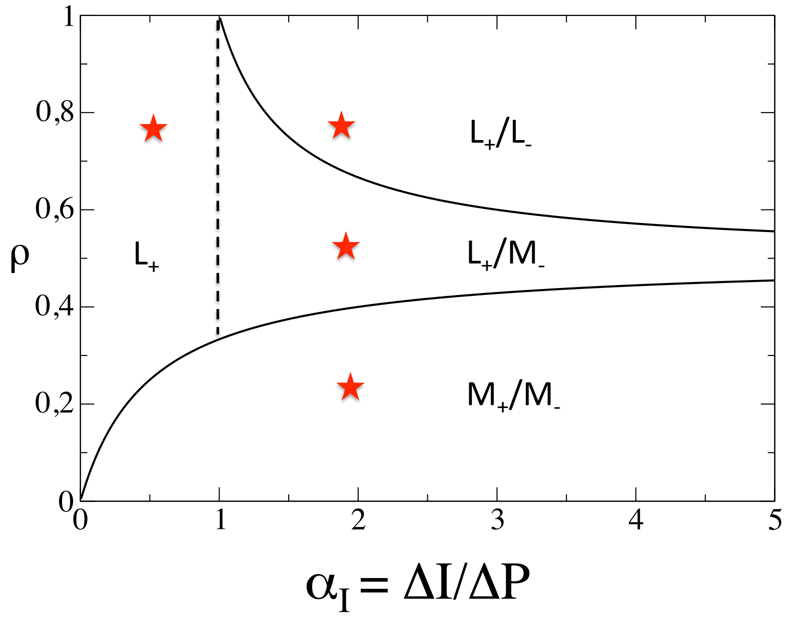
Convergence speed

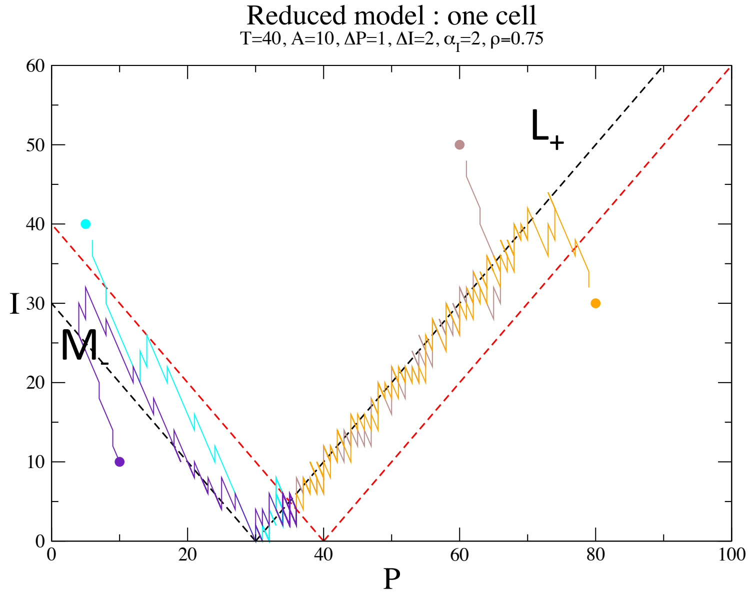
$$L_+ : \quad v_c = -\frac{1-\rho}{1+\alpha_I}\Delta I$$

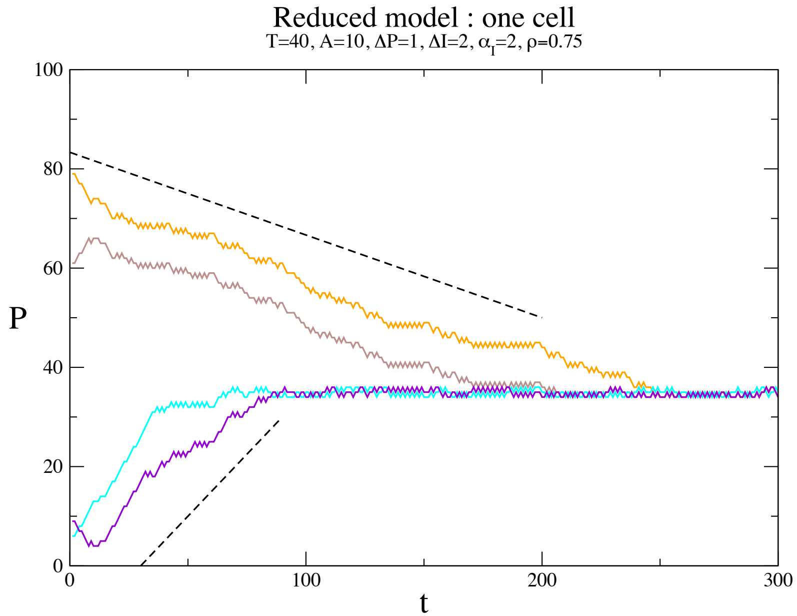
Dominated by error estimation

$$M_- : \quad v_c = \rho \Delta P$$

Dominated by rate update







Simple mathematical description

- Purkinje cell as an analog perceptron

g_i^μ, i=1,…,N_g inputs 0,1 with prob. f

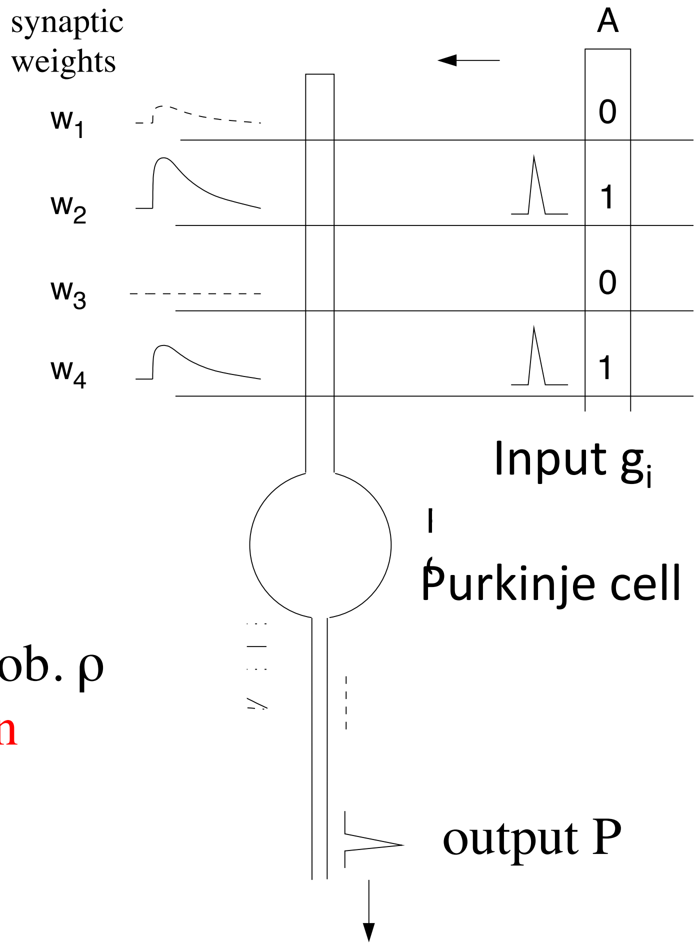
μ=1,…,N_p patterns, target rates T^μ

$$P = \left[\frac{1}{\sqrt{N_g}} \left(\sum_i w_i g_i^\mu - \theta N_g \right) \right]_+ + \eta A$$

η=0,1 with prob. ρ

A perturbation

$$E = |P - T^\mu|$$ Error for pattern μ



- Input to DCN

$$D = E - I, \quad \text{with} \quad I = \left[\frac{1}{\sqrt{N_g}} \left(\sum_i q_i g_i^\mu - \theta N_g \right) \right]_+$$

Perceptron plasticity rules

$$P = \left[\frac{1}{\sqrt{N_g}} \left(\sum_i w_i g_i^\mu - \theta N_g \right) \right]_+ + \eta A$$

η=0,1 with prob. ρ perturbation A

$$D = E - I, \quad \text{with} \quad I = \left[\frac{1}{\sqrt{N_g}} \left(\sum_i q_i g_i^\mu - \theta N_g \right) \right]_+$$

$$E = |P - T^\mu|$$

Error for pattern μ

Synaptic weight changes :

$$c = \text{sign}(E - I)$$

climbing fiber discharge/presence or absence of perturbation CS

$$w_i \to [w_i - \alpha_w \, c \, \eta \, g_i^\mu]_+$$

$$q_i \to [q_i + \alpha_q \, c \, g_i^\mu]_+$$

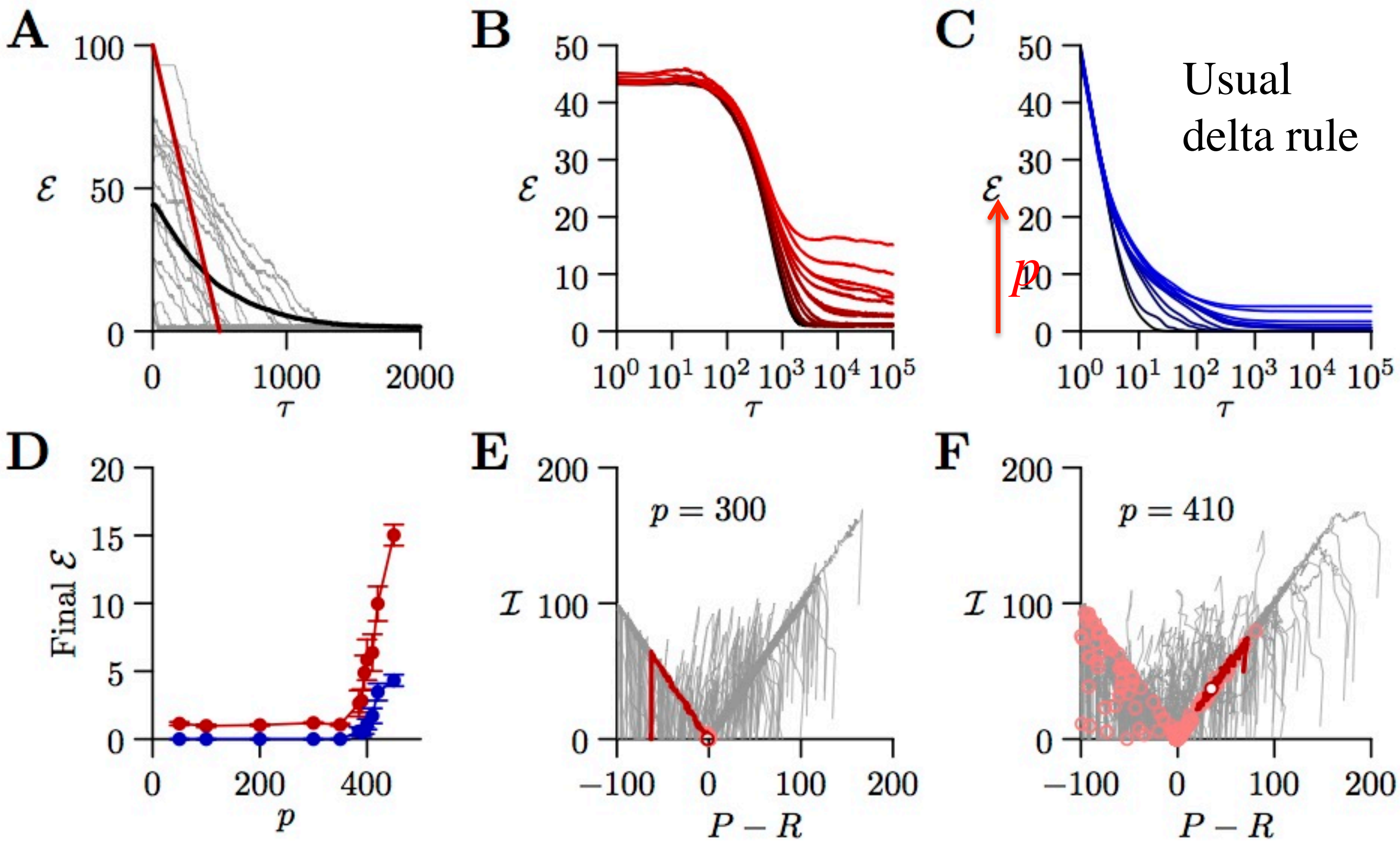
Relation to toy model

$$\Delta P = \alpha_w f \sqrt{N_g}$$

$$\Delta I = \alpha_q f \sqrt{N_g}$$

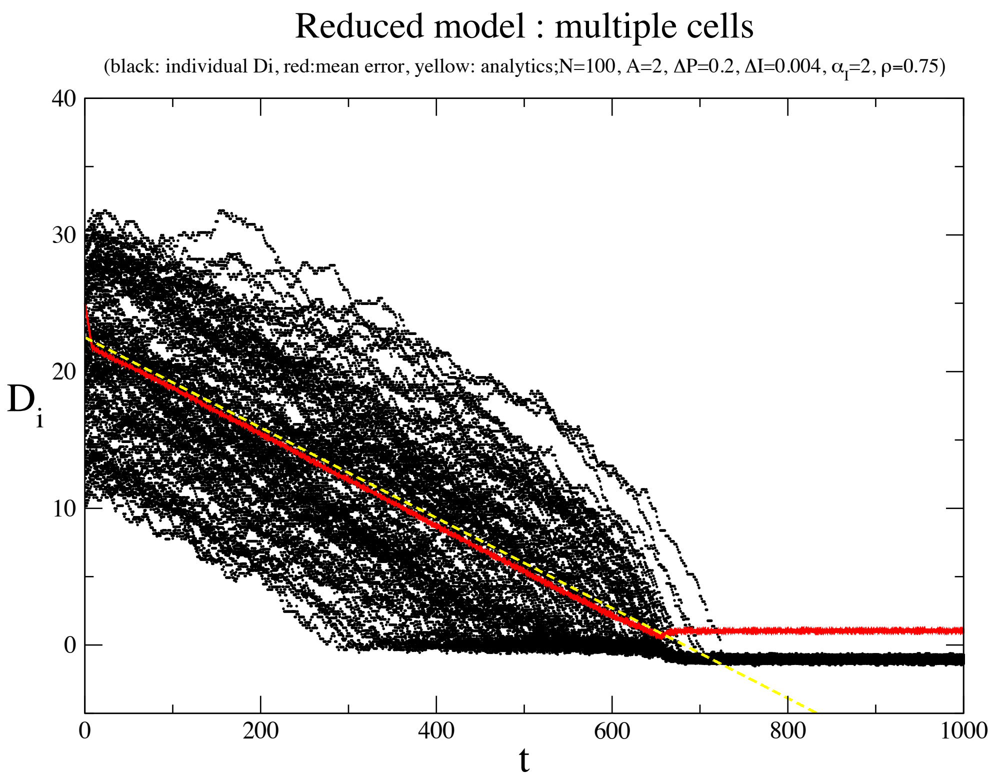
Perceptron : learning p patterns with estimated error

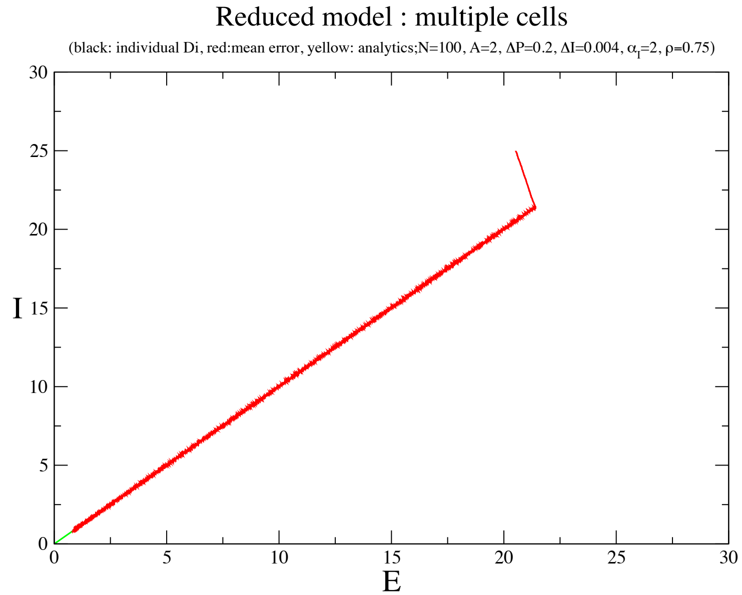
N=1000, coding fraction f=0.2

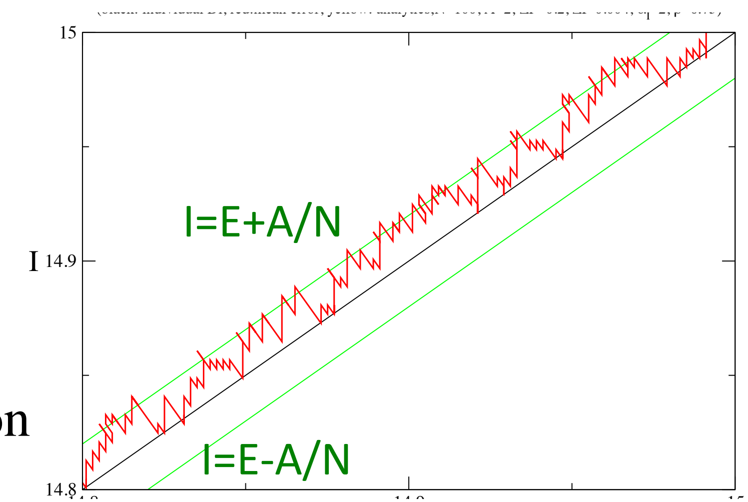


N cells, estimated global error

Many cells







Generalized analysis possible, cells successively reach their limiting rates convergence curve depend on initial condition

What’s next ?

Theory :

- convergence dynamics with ‘interfering’ patterns
- different forms of global error

Experiment : in vivo data

- evidence from perturbation during movement
- extraction of plasticity rules from data

Both : - Other structures where a similar type of learning may apply? (e.g. basal ganglia; dopamine release signals reward but also promotes movement initiation, plasticity rules involving this two successive releases?).

Thank you!

Nicolas Brunel

Claudia Clopath

Jean-Pierre Nadal

Yonatan Aljadeff

Jonas Ranft

Boris Barbour

Guy Bouvier

Celian Bimbard