LETTER Communicated by Angus Silver

How Synaptic Release Probability Shapes Neuronal Transmission: Information-Theoretic Analysis in a Cerebellar Granule Cell

Angelo Arleo
angelo.arleo@upmc.fr
CNRS, UPMC, UMR 7102 Neurobiology of Adaptive Processes, F-75005, Paris, France, and Neuroscience Group, SONY Computer Science Laboratory, 75005, Paris, France

Thierry Nieus
thierry.nieus@iit.it
Neuroscience and Brain Technologies, Italian Institute of Technology, 16163 Genova, Italy, and Department of Physiology, University of Pavia, and IRCCS C. Mondino, 27100, Pavia, Italy

Michele Bezzi
michele.bezzi@gmail.com
Neuroscience Group, SONY Computer Science Laboratory, 75005, Paris, France

Anna D'Errico
derrico@sissa.it
Egidio D'Angelo
dangelo@unipv.it
Department of Physiology, University of Pavia, and IRCCS C. Mondino, 27100, Pavia, Italy

Olivier J.-M. D. Coenen
olivier@oliviercoenen.com
Neuroscience Group, SONY Computer Science Laboratory, 75005, Paris, France

A nerve cell receives multiple inputs from upstream neurons by way of its synapses. Neuron processing functions are thus influenced by changes in the biophysical properties of the synapse, such as long-term potentiation

---

Note: A. A., T. N., M. B., and O. J. M. D. C. contributed equally to this work. M. Bezzi is now at SAP Research Center, Sophia-Antipolis, France. A. D'Errico is now at Sector of Neurobiology, International School for Advanced Studies, Trieste, Italy. O. Coenen is now at Computational Neuroscience Research Team, Intelligent Systems Research Center, School of Computing and Intelligent Systems, University of Ulster, Londonderry, BT48 7JL, Northern Ireland.

(LTP) or depression (LTD). This observation has opened new perspectives on the biophysical basis of learning and memory, but its quantitative impact on the information transmission of a neuron remains partially elucidated. One major obstacle is the high dimensionality of the neuronal input-output space, which makes it unfeasible to perform a thorough computational analysis of a neuron with multiple synaptic inputs. In this work, information theory was employed to characterize the information transmission of a cerebellar granule cell over a region of its excitatory input space following synaptic changes. Granule cells have a small dendritic tree (on average, they receive only four mossy fiber afferents), which greatly bounds the input combinatorial space, reducing the complexity of information-theoretic calculations. Numerical simulations and LTP experiments quantified how changes in neurotransmitter release probability (p) modulated information transmission of a cerebellar granule cell. Numerical simulations showed that p shaped the neurotransmission landscape in unexpected ways. As p increased, the optimality of the information transmission of most stimuli did not increase strictly monotonically; instead it reached a plateau at intermediate p levels. Furthermore, our results showed that the spatiotemporal characteristics of the inputs determine the effect of p on neurotransmission, thus permitting the selection of distinctive preferred stimuli for different p values. These selective mechanisms may have important consequences on the encoding of cerebellar mossy fiber inputs and the plasticity and computation at the next circuit stage, including the parallel fiber–Purkinje cell synapses.

1 Introduction

Theoretically, neurons can be considered as transmitting devices encoding information in terms of digital spike trains. Spikes are transmitted between neurons at the synapses, where they are converted into analog signals by elaborate nonlinear transformations based on the time-dependent properties of neurotransmitter release and diffusion, postsynaptic receptor activation, and intrinsic electroresponsiveness. Synapses undergo plasticity via activity-dependent modifications, such as long-term potentiation (LTP) and depression (LTD) (Bliss & Collingridge, 1993; Bliss & Lomo, 1973; Malenka & Bear, 2004). These modifications affect the synaptic dynamics (e.g., by regulating the number of neurotransmitter quanta released or the phosphorylation of postsynaptic receptors), and ultimately they alter the computational and transmitting properties of the whole postsynaptic neuron (Abbott & Regehr, 2004; Tsodyks & Markram, 1997).

Neuronal computation can be analyzed in terms of information content by quantifying how much information the neural responses convey about

the input stimuli (Bialek, Rieke, de Ruyter van Steveninck, & Warland, 1991; Borst & Theunissen, 1999; Fuhrmann, Segev, Markram, & Tsodyks, 2002; Quian Quiroga & Panzeri, 2009). In this framework, neurons are treated as stochastic communication channels, and information theory (Cover & Thomas, 1991; Shannon, 1948) provides the mathematical tools to measure their transmitting properties. Information theory has been used to analyze neuronal computation and quantify the information transmitted by a neuron following sensory stimulation (e.g., in the cat visual cortex: Sharpee et al., 2006; in the fly visual system: Brenner, Bialek, & de Ruyter van Steveninck, 2000; Brenner, Strong, Koberle, Bialek, & de Ruyter van Steveninck, 2000; de Ruyter van Steveninck, Lewen, Strong, Koberle, & Bialek, 1997; in the mammalian auditory system: Lu & Wang, 2004; Smith & Lewicki, 2006; in the cricket cercal sensory system: Dimitrov, Miller, Gedeon, Aldworth, & Parker, 2003; Roddey & Jacobs, 1996; Theunissen, Roddey, Stufflebeam, Clague, & Miller, 1996; Theunissen & Miller, 1991; in the rat somatosensory system: Wan et al., 2004). It has also been used to characterize the relevant regions of a neuron tuning curve in relation to the variability in its sensory encoding properties (Butts & Goldman, 2006) or to characterize the adaptation in receptive fields to visual stimuli (Sharpee et al., 2006). Information theory has been found useful to measure the efficiency of information transmission at a single synapse (de la Rocha, Nevado, & Parga, 2002; Fuhrmann et al., 2002; Goldman, 2004; London, Schreibman, Hausser, Larkum, & Segev, 2002; Manwani, Steinmetz, & Koch, 2002; Tiesinga, 2001), as well as to characterize the information transmission in simplified integrate-and-fire neurons (Manwani et al., 2002; Zador, 1998), or in more complex models, but under specific constraints, for example, in the absence of interaction among presynaptic inputs (Manwani & Koch, 2001) or between inhibitory inputs and phase lags (Tiesinga, Fellous, Jose, & Sejnowski, 2002). In the current work, an information-theoretic approach was used to characterize the processing of a neuron over a region of its excitatory input space and investigate how neuronal processing changes following synaptic plasticity.

The mean information transmitted between the neuron responses r and its inputs s can be quantified by using Shannon's mutual information (MI) (Cover & Thomas, 1991; Rolls & Deco, 2002; Shannon, 1948):

$$MI(R, S) = \sum_{s \in S} \sum_{r \in R} p(s) p(r \mid s) \log_2 \frac{p(r \mid s)}{p(r)}, \tag{1.1}$$

where S and R are the input and output spaces, respectively; $p(s)$ and $p(r)$ are the a priori probability distributions; and $p(r \mid s)$ is the conditional probability distribution.

The MI measure can be used to estimate how much the neuron response reflects the input stimuli versus the intrinsic variability of the neuron (e.g.,

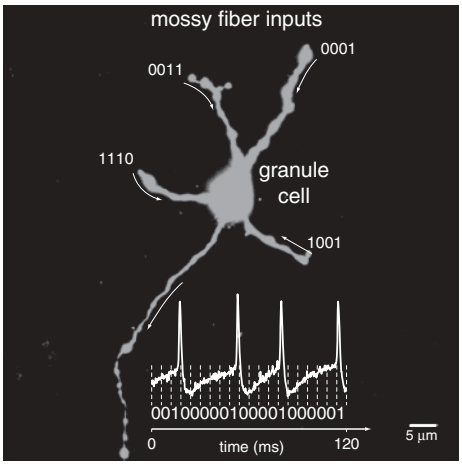


Figure 1: Cerebellar granule cell morphology and spike signal digitalization. Confocal image of a biocytin-stained granule cell (GC) in the rat cerebellum (slice preparation). Granule cells are characterized by a compact electrotonic structure (D'Angelo et al., 1995; Saviane & Silver, 2006) and have an exceptionally low number of synapses (four on average) (Eccles et al., 1967; Jakab & Hamori, 1988) receiving spike trains from the mossy fibers (MFs) (Chadderton et al., 2004). GCs are numerous ($\sim 10^{11}$ in humans) and constitute more than half of the neurons in the brain. White trace: The membrane potential of a GC recorded over 120 ms. Spike trains were digitized as strings of 0s and 1s, within discrete time windows of 6 ms (time bins).

if the response is independent from the inputs, MI will be zero). Thus, MI essentially measures the difference between the signal and noise entropy (Borst & Theunissen, 1999) and provides a sound statistical tool to dissect the relative contributions of different factors to neural information transmission (e.g., spike count versus spike timing) under different conditions (e.g., before and after LTP induction). One major factor that makes it difficult to estimate the MI (see equation 1.1), and assess how synaptic plasticity affects neuronal processing is the high dimensionality of the input-output space (Borst & Theunissen, 1999). Indeed, assessing MI requires determining the probability distribution of the output spike trains given any input spike train. In general, this is impracticable even for a single neuron due to the multiple mechanisms of nonlinear integration at individual synapses; the large number of synapses, typically 10^3 to 10^4; and their location on wide dendritic trees with complex electrotonic and active properties (Borst & Theunissen, 1999; Koch & Segev, 2000).

To overcome these issues, the cerebellar granule cell, GC (see Figure 1), was considered. GCs are tiny neurons (6 μm diameter) (Eccles, Ito, &

Szentagothai, 1967; Jakab & Hamori, 1988) located at the major input stage of the cerebellum, the granular layer. GCs play a major role in the early stages of cerebellar computation, and their synapses have been proposed to regulate the input-output relationship through gain modulation (Albus, 1971; Mitchell & Silver, 2003; Rothman, Cathala, Steuber, & Silver, 2009) and long-term adaptation (Hansel, Linden, & D'Angelo, 2001; Philipona & Coenen, 2004; Schweighofer, Doya, & Lay, 2001).

GCs have several remarkable properties. First, they have a compact electrotonic structure (D'Angelo, De Filippi, Rossi, & Taglietti, 1995; Saviane & Silver, 2006; Silver, Traynelis, & Cull-Candy, 1992), which maintains their whole cell membrane equipotential, eliminating spatial effects on computation (Koch & Segev, 2000). Second, they have a low number of mossy fiber (MF) afferents (4.17 on average) (Eccles et al., 1967; Jakab & Hamori, 1988), which generates a tractable number of presynaptic input combinations and greatly reduces the complexity of information-theoretic calculations. Third, they have a stereotyped synaptic and excitable behavior, which simplifies the implementation of models (D'Angelo et al., 1995). Fourth, MFs have been shown to respond with high-frequency bursts to punctuate stimulation (Arenz, Silver, Schaefer, & Margrie, 2008; Chadderton, Margrie, & Hausser, 2004; Jorntell & Ekerot, 2006; Rancz et al., 2007), which have been carefully characterized. Fifth, both GC output bursts and plasticity at MF-GC synapses are controlled by input patterns and Golgi cell inhibition (Mapelli & D'Angelo, 2007). The dynamics of repetitive stimulation have been clarified to a considerable extent (Nieus et al., 2006; Saviane & Silver, 2006; Sola, Prestori, Rossi, Taglietti, & D'Angelo, 2004). Finally, MF-GC synaptic transmission is based on nonlinear transformations determined by presynaptic short-term facilitation and depression, glutamate spillover, postsynaptic AMPA and NMDA receptor gating, and multiple voltage-dependent channel interactions regulating intrinsic electroresponsiveness (D'Angelo et al., 1995; Nielsen, DiGregorio, & Silver, 2004; Nieus et al., 2006; Sargent, Saviane, Nielsen, DiGregorio, & Silver, 2005; Sola et al., 2004). The mechanisms of synaptic transmission and plasticity at the MF-GC synapses have been intensely investigated, revealing that LTP is largely determined by a raise in the presynaptic neurotransmitter release probability (p) (D'Angelo et al., 2001; Nieus et al., 2006; Saviane & Silver, 2006; Sola et al., 2004).

The study presented here focused on the effects of release probability (p) changes at MF-GC synapses on the overall GC information transmission properties. The same information-theoretic quantification was applied to analyze data from numerical simulations of biophysical synaptic and GC models, as well as data issued from in vitro intracellular GC recordings. At a first level, the mutual information MI between MF inputs and GC responses was measured as a function of release probability. Expectedly, MI increased significantly with p, as MI is a measure of the information transmission averaged over the entire input set considered for examination. A second level of

analysis assessed the contribution of specific stimuli to information transmission. For this purpose, the stimulus-specific surprise measure (Butts & Goldman, 2006; DeWeese & Meister, 1999; Theunissen & Miller, 1991) was used to quantify the optimality of single stimulus transmission (see section 4). This analysis investigated which stimulus patterns were "preferred" by the neuron under different release probability conditions, and it demonstrated that maximum synaptic release probability did not constitute a necessary condition in order to achieve optimal transmission. Rather, for a significant set of stimuli, the surprise values saturated at intermediate p values (consistent with those found in brain recordings at MF-GC synapses: Sola et al., 2004; at CA3-CA1 hippocampal synapses: Dobrunz & Stevens, 1997; and at neocortical pyramid-to-pyramid connections: Koester & Johnston, 2005) with optimal information transmission occurring over a large range of release probabilities (from about 0.4 to the maximum value tested, i.e., 0.8). Finally, a third level of analysis concentrated on the spatiotemporal characteristics of the stimuli and quantified, for different p values, the spike timing contribution to information transmission by means of the surprise-per-spike measure (see section 4). The surprise per spike was typically higher for either long-correlated stimuli at low p or short-correlated stimuli at high p.

The work presented in this letter provides a set of tools to investigate neuronal coding and information transfer in the cerebellar granular layer network, which enables a quantitative exploration of the relative importance of the coding strategies for different input patterns and synaptic parameters.

2 Results

A first fundamental question is how the transmitting properties of a GC vary under different release probability conditions (e.g., LTP). To this aim, the average amount of information transmitted by the cell (i.e., the MI) was measured (see section 2.1) computationally with a GC model and experimentally. Experimental results were obtained with a limited set of stimuli; the numerical simulations were used for corroborating the experimental findings and extending them to a larger region of the input space. The question of how changes in release probability shape the neurotransmission of specific MF inputs is addressed in section 2.2 using the stimulus-specific surprise. Finally, section 2.3 investigates how correlations across input spike trains affect single-stimulus transmission and assesses the informative contribution of single pulses in the presence of distinct spatiotemporal stimulus structures and multiple presynaptic release probabilities.

2.1 Impact of Neurotransmitter Release Probability Changes on Information Transmission. The average information transmitted by a single GC was quantified before and after induction of long-term synaptic

plasticity at MF-GC synapses, a condition shown to modify release probability at the MF synaptic terminals (Sola et al., 2004). MI analysis was performed using experimental data obtained by whole-cell patch recordings of GCs during in vitro electrophysiological recordings (see section 4). To measure MI, one to four MFs were stimulated by a set of spike trains, according to a protocol inspired by the bursting discharge of GCs following punctuate tactile stimulation in vivo (Chadderton et al., 2004). Because our experimental techniques did not allow us to stimulate the four MFs independently, this analysis could be done only over a limited input set (16 distinct stimuli, each made of identical spike trains on the four MF inputs). The GCs responded with noisy spike trains that had higher average frequency and occurred earlier after LTP (see Figure 2A). The neurotransmitter release probability p was estimated before and after LTP induction (see Supplementary Material).[1] Because p was the average value over the different synapses, it will henceforth be indicated with $\bar{p}$. LTP caused an average MI increase of $32 \pm 4\%$ for $\bar{p}$ changing by $48 \pm 5\%$ ($n = 9$; paired student's t-test, $p < 0.05$). The average amount of information carried by the firing frequency was 51%, meaning that half of the information transfer was due to interspike temporal relationships.

The same stimulation protocol was employed to run the numerical simulations with the detailed GC model, and MI was measured for different p values at the model MF-GC synapses. As shown in Figure 2B, MI increased as a function of $\bar{p}$ for both experimental and simulated data. The numerical results predicted an increase in MI from 0.2 to 3.5 bits for $\bar{p}$ varying within the range [0.1, 1]. Due to the restricted input space explored (16 stimuli), MI tended to saturate when the number of simultaneously active MFs was greater than 2 and $\bar{p} >= 0.5$ (see below). The vectors representing the experimental MI changes during LTP fell within the limits of the model predictions for all the GCs examined. The similarity between experimental and simulation results indicated that the model could predict the information transfer dynamics following long-term synaptic plasticity. It also supported the hypothesis that the major noise source of GCs is stochastic neurotransmitter release, as anticipated by quantal analysis (Sola et al., 2004). A control analysis focusing on the role of presynaptic depression and postsynaptic receptor desensitization revealed that setting either one or the other to zero affected spike timing only to a small extent (see Supplementary Material, section S2.2), implying minor effects on information transmission.

The model permitted extending numerical simulations to the experimentally impracticable case of independent activation of the four MF afferents. Independent input spike trains (with a maximal frequency of 100 Hz) were generated at each of the four MFs, and all the possible

[1]Supplementary material referred to throughout the letter is available online at http://www.mitpressjournals.org/doi/suppl/10.1162/NECO_a_00006-Arleo.

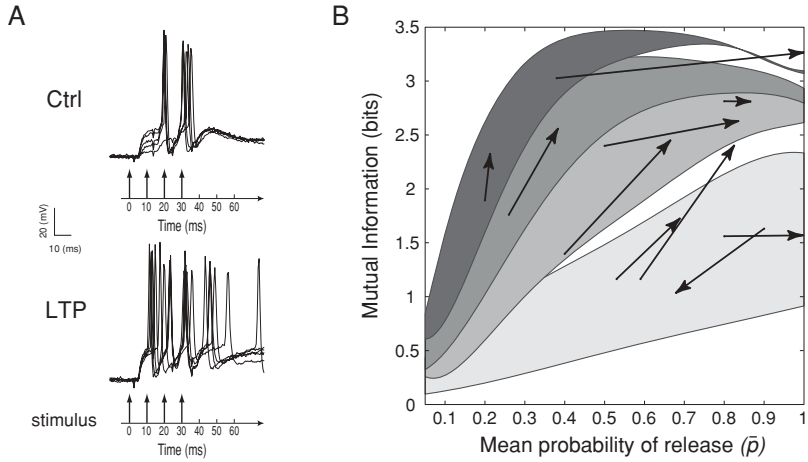


Figure 2: Increases in neurotransmitter release probability enhance mutual information in cerebellar granule cells: experiments and simulations with a restricted mossy fiber input set. (A) A GC was maintained at a membrane potential between −60 mV and −70 mV, and it was activated through MF stimulations at the times indicated by the arrows. This protocol was repeated 25 times. The responses were recorded before and after delivering a theta-burst stimulation (TBS) (Nieus et al., 2006) (five selected voltage traces are shown superposed). In this cell, release probability was $\bar{p} = 0.4$ before and $\bar{p} = 0.65$ after LTP induction (see Supplementary Material for the $\bar{p}$ estimation process). Effective long-term potentiation (LTP) induction is demonstrated by earlier and more intense spike activation (Nieus et al., 2006). (B) Mutual information (MI) changed significantly as a function of neurotransmitter release probability, $\bar{p}$. The shaded regions show the MI values obtained with the GC model, whereas the vectors indicate the experimental changes in MI before and after induction of long-term synaptic plasticity in 10 GCs. The four regions (from bottom to top) correspond to different numbers of active MFs (one to four). The lower and upper borders of each computed region were obtained by setting the resting potential of the GC model at −70 and −60 mV, the two experimental extremes, respectively. LTP, associated with a $\bar{p}$ increase, was observed in nine cases, while in one case with very high initial $\bar{p}$, LTD (long-term depression) was observed associated with a $\bar{p}$ decrease. Note that a 60 ms time window was used to sample both experimental and simulated GC responses.

input combinations were explored (yielding an extensive set of 65536 stimuli; see section 4). Neurotransmitter release probability was also regulated independently at each MF-GC synapse (from $p = 0.1$ to 0.8, in steps of 0.1). Consequently the information transmission analysis was performed for many different p combinations across the four MFs (e.g., $p_{MF_1} = 0.2$, $p_{MF_2} = 0.8$, $p_{MF_3} = 0.3$, $p_{MF_4} = 0.4$). Figure 3A displays a sample response of the

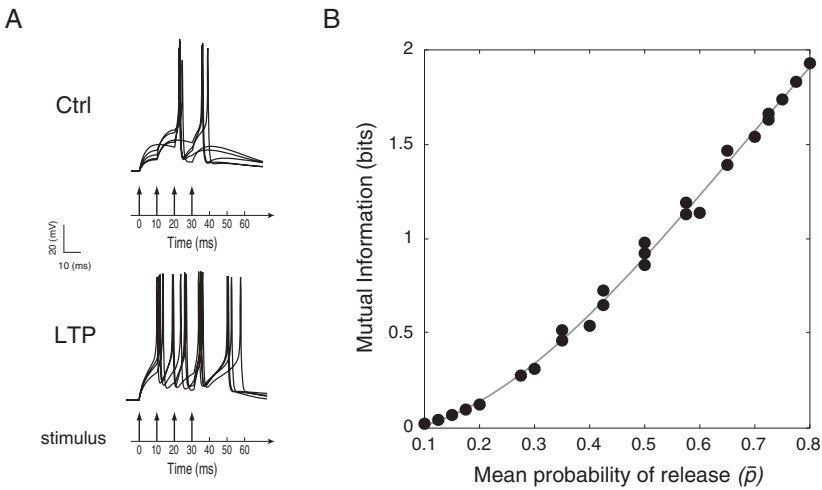


Figure 3: Increases in neurotransmitter release probability enhance mutual information in cerebellar granule cells: simulation results with an extended mossy fiber input set. (A) Responses of the model GC increased their time locking to the start of the input stimulation after a large increase in release probability $\bar{p}$ (five sample traces are shown superposed). (B) MI calculated by stimulating the MF-GC synapses with the extended input set ($k = 2^{16} = 65536$ stimuli) and by varying the release probability p on the four MFs independently, within the range [0.1, 0.8] (Sola et al., 2004). Each data point indicates the MI value corresponding to a different combination of p across the four MFs, and the x-axis provides the $\bar{p}$ averaged over the four MFs (therefore different MI values can coexist for any value of $\bar{p}$). The larger $\bar{p}$ is, the larger is MI. Note that MI shows a smooth increase rather independent from the specific p combination used over the different synapses.

model GC to the same stimulus before and after simulated LTP (i.e., for two different p combinations, with $\bar{p} = \langle p_{MF_i} \rangle_{i=1,4}$). Figure 3B shows that the MI computed over the extended stimulus set increased as a function of $\bar{p}$ without saturating (in contrast to the limited set case of Figure 2B), suggesting that the information transmission averaged over a large set of MF stimuli may benefit linearly from increasing $\bar{p}$ values.

2.2 Impact of Release Probability Changes on Single Stimulus Transmission. The surprise measure (see section 4) was used to study the influence of release probability changes on the transmission of specific MF inputs. For each $\bar{p}$ value, all stimuli were ranked according to their surprise value (e.g., in Figure 4A for $\bar{p} = 0.42$), and then different subsets of stimuli were considered. In the subset of stimuli with surprise larger than 90% of the maximum (see Figure 4B), after an initial rapid growth for $0.1 \leq \bar{p} \leq 0.5$,

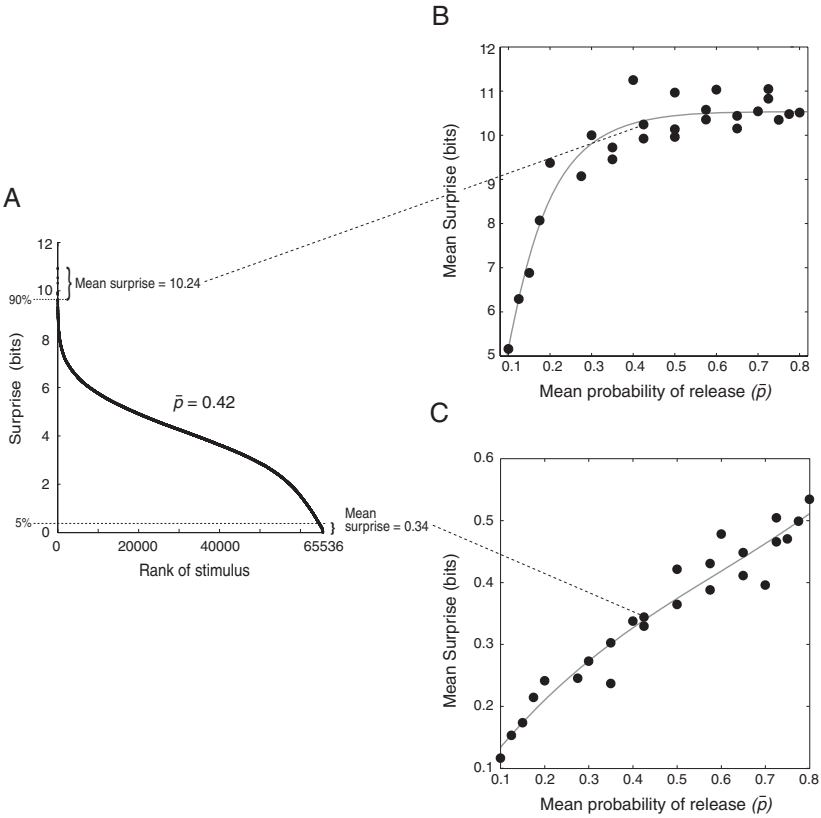


Figure 4: Surprise as a function of neurotransmitter release probability: simulation results. (A) The surprise measure was used to rank the extensive stimulus set ($k = 2^{16} = 65{,}536$ stimuli), which had p values set independently on the four MFs within the range [0.1,0.8]. The example given here had an average (over the four MF-GC synapses) release probability of $\bar{p} = 0.42$. (B) The mean surprise for the (top 10%) most informative stimuli tended to plateau as $\bar{p}$ increased, even though MI increased with larger $\bar{p}$ values (see Figure 3B). For a particular set of p values, the mean surprise was obtained by averaging over the subset of stimuli with a surprise larger than 90% of the maximum. Note that the subset of stimuli contributing to the mean surprise may change with different set of p values. (C) The mean surprise for the least informative stimuli—stimuli with a surprise less than 5% of the maximum—was the only subset for which the surprise was monotonously increasing with $\bar{p}$. See also Supplementary Material, figure S15 and section S2.4.

the average surprise tended to plateau around 10.5 bits. This saturation effect was observed for most of the stimuli (see Supplementary Material, Figure S15 and section S2.4). Only the surprise averaged over the subset of stimuli having a surprise smaller than 5% of the maximum showed a quasi-linear increase with $\bar{p}$ (see Figure 4C), suggesting that only these stimuli continued to benefit from increasing $\bar{p}$ values. Although on average the information transmitted by the GC increased monotonically with $\bar{p}$ (see Figure 3B), optimal single-stimulus transmission already occurred at intermediate release probability values for most MF inputs, and further increases in $\bar{p}$ did not result in larger surprise values.

This finding suggests that transmission of specific inputs was differentially affected by changes in presynaptic release probability. To begin characterizing the properties of stimuli with optimal transmission at different $\bar{p}$, simple features such as low versus high firing rate were first considered. Low-rate stimuli tended to benefit the most from $\bar{p}$ increases (see Figure S16). The hypothesis that high-frequency stimuli (as opposed to low-rate stimuli) had less to gain from an increasing release probability $\bar{p}$ was confirmed by quantitative analysis (see Figure S17). But how does the spatiotemporal structure of MF inputs affect neurotransmission in GCs?

2.3 Influence of MF Input Correlations and Spike Timing on Information Transmission. In order to reveal the effect of input correlations on information transmission, the relationship between the surprise of a stimulus and the correlation C across the four MF spike trains forming the stimulus was analyzed for each release probability, $\bar{p}$ (similar to Butts & Goldman, 2006). The coefficient C measured the average number of coincident input spikes across the four MFs (see section 4). The surprise tended to increase as a function of C for all $\bar{p}$ values, showing that correlated activity across MFs improves GC neurotransmission (see Figure 5). Moreover, these findings supplement the observations that GCs operate as coincidence detectors (Moreno-Bote & Parga, 2004) requiring the coactivation of two or more MFs (D'Angelo et al., 1995; Jorntell & Eckerot, 2006; Eccles et al., 1967).

In order to discard the influence of firing rate on information transmission and focus on the contribution of single spikes, the surprise per spike was employed (see section 4). The entire MF stimulus set was ranked according to the surprise-per-spike value for all release probabilities, $\bar{p}$ (see Figure 6A). As shown in Figure 6B.B1, the stimuli with the largest surprise per spike at low $\bar{p}$ were characterized by long-correlated MF spike trains (the blue-labeled patterns), whereas as $\bar{p}$ increased, the stimuli with greatest surprise per spike were short correlated trains (red-labeled stimuli). As shown in Figure 6B.B2, the stimuli with lowest surprise per spike at small $\bar{p}$ did not have a stereotyped structure, whereas as $\bar{p}$ increased, they rapidly became stimuli with at most one spike per bin and therefore no coactivated spikes.

These findings were corroborated by analyzing how the surprise per spike changed as a function of the correlation C for different release

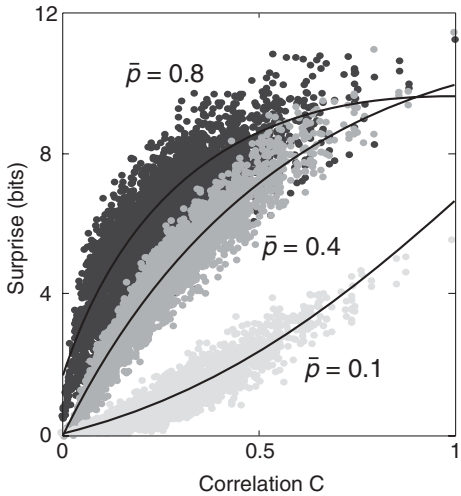


Figure 5: Surprise as a function of the correlation across the four mossy fiber inputs: simulation results. The larger the correlation coefficient C of the stimulus became, the larger was the information transmitted by the stimulus. C measures the average number of coincident spikes across the four MF afferents and over the four time bins (see section 4). This result holds for all release probability values $\bar{p} \in [0.1,0.8]$. Nevertheless, the larger the $\bar{p}$ value was, the more the surprise saturated with higher C values.

probability values (see Figure 6C). At low $\bar{p}$, the surprise per spike increased quasi-linearly with the MF spike correlation C, with the peak of the distribution occurring for long-correlated inputs (blue-labeled stimuli in Figure 6C, left). By contrast, as $\bar{p}$ increased, the distribution of the surprise per spike as a function of C became non linear, and the peak of the distribution increased and occurred for shorter correlated stimuli (red dots in Figure 6C, center and right). Thus, an increase in $\bar{p}$ enhanced the surprise per spike of short correlated stimuli, whereas it favored only partially that of long highly correlated stimuli, whose surprise per spike saturated and even decreased at high $\bar{p}$.

2.4 Influence of Specific Patterns on Information Transmission. Why did long correlated stimuli benefit less, in terms of information per spike, from $\bar{p}$ increases than short correlated ones? A possible explanation emerged by examining at different $\bar{p}$ values the GC responses to distinct MF stimuli. Figure 7A shows both the spikegram and the poststimulus time histogram (PSTH) of the GC response to a long correlated stimulus (with four bins with coactivated spikes). For $\bar{p} > 0.5$, spike doublets were elicited by the second set of input spikes (second arrow, $\bar{p} = 0.6$ and $\bar{p} = 0.8$). The

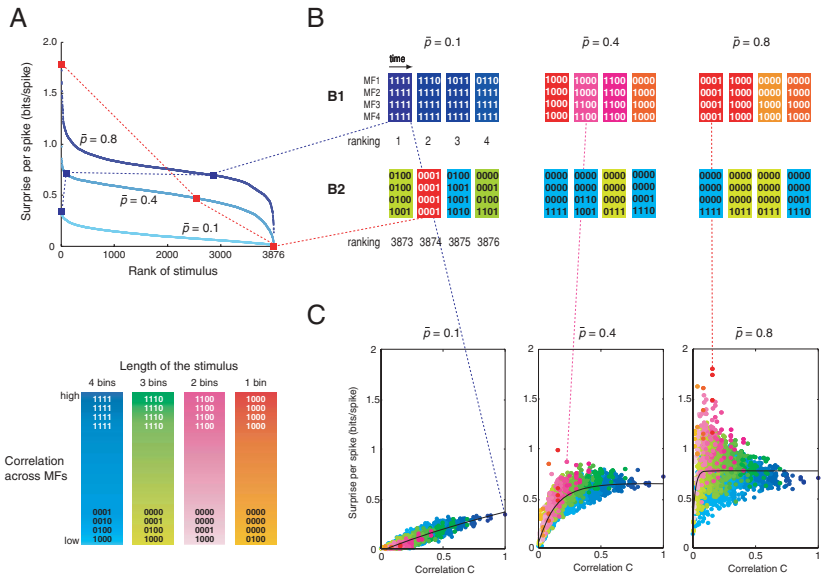


Figure 6: Surprise per spike as a function of release probability: simulation results. (A) The stimuli were ranked as a function of their surprise per spike for every $\bar{p}$ value [0.1, 0.8, step of 0.1]. Here, the release probability was equal for all four MF synapses at the GC; that is, any permutation of the four input spike trains was equivalent. This reduced the number of different stimuli to 3876 from the initial $2^{16} = 65{,}536$. (B) The sets providing the largest (B1) and the smallest (B2) contribution to the surprise per spike were selected for three different $\bar{p}$ values. A different color map was chosen to identify the number of bins occupied (blue, green, pink, red for 4, 3, 2, 1 bins, respectively). The number of spikes per bin modulated the color map (e.g., red to orange for four to one spikes per bin, within the red color map). At low $\bar{p}$, the stimuli with the largest surprise per spike were long-correlated spike trains: four bins with coactivated spikes at $\bar{p} = 0.1$ (e.g., the blue-labeled stimuli in B1). As $\bar{p}$ increased, the stimuli with largest surprise per spike were short correlated trains with two or three bins of coactivated spikes at $\bar{p} = 0.2$–0.3, one to two bins at $\bar{p} = 0.4$–0.5 (red and pink-labeled stimuli), and one bin for $\bar{p} > 0.5$ (red and orange-labeled stimuli). The stimuli with the smallest surprise per spike had no stereotyped structure at low $\bar{p}$, whereas they were long noncorrelated patterns at high $\bar{p}$ (e.g., the cyan-labeled stimuli in B2). (C) Whereas the surprise per spike increased almost linearly with correlation at the lowest $\bar{p}$ value ($\bar{p} = 0.1$), the peak of the distribution increased and moved back as $\bar{p}$ increased: starting with the highest correlated stimuli at $\bar{p} = 0.1$ with four coactivated spikes in all four bins (blue dots), the peak at $\bar{p} = 0.8$ was found for a stimulus with a smaller correlation value, with four coactivated spikes in only one bin (red dot and red-labeled stimuli). The distribution of points at $\bar{p} = 0.4$ already showed saturation for the long correlated stimuli while for the shorter correlated stimuli, the distribution started to peak.

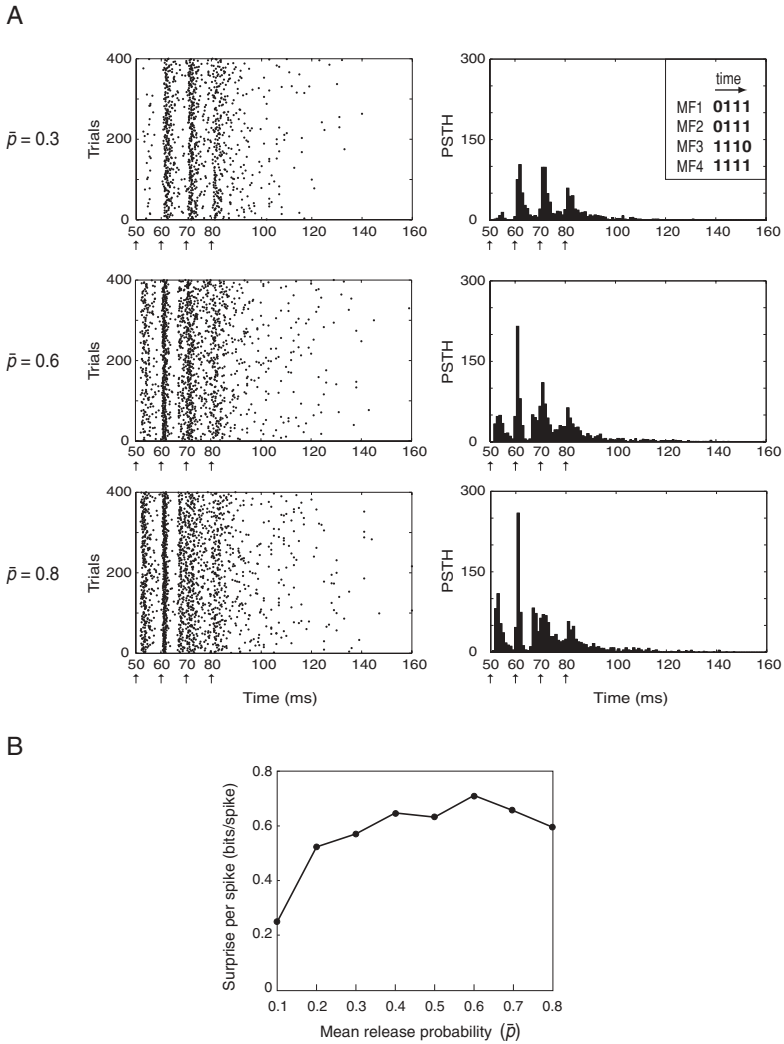


Figure 7: Spike doublet appearance in GC responses in the presence of high release probability: simulation results. (A) GC response (left: spikegram; right: poststimulus time histogram, PSTH) to a long correlated stimulus at three different $\bar{p}$ values. The stimulus structure is shown in the top-right inset, and arrows in the time line of the diagrams indicate the input spike timing. For $\bar{p} = 0.6$ and 0.8, the second set of spikes (second arrow) elicited a spike doublet, made of one spike occurring at ∼61 ms, followed by another spike at ∼67 ms (before the occurrence of the next set of input spikes at 70 ms). Note the double-peaked PSTH around 70 ms. (B) Surprise per spike of the long correlated stimulus used in *A* as a function of release probability.

doublets disrupted the precise spike timing of the response (evident in the PSTH after the second response spike), limiting the surprise per spike of the stimulus despite the improved timing of the first response spike with respect to the $\bar{p} < 0.5$ case (see Figure 7B). The longer the stimulus (i.e., the larger the number of bins with spikes), the higher was the probability of eliciting spike doublet responses at high $\bar{p}$ values (see Figure 8C for other spike doublet examples). This behavior was probably determined by the engagement of smoothly varying currents, such as the NMDA current and the persistent and resurgent Na^+ currents (D'Angelo et al., 2001; Magistretti, Castelli, Forti, & D'Angelo, 2006; Nieus et al., 2006), with increasing $\bar{p}$. These currents drive repetitive firing and decrease the probability of having precise stimulus-locked response spikes.

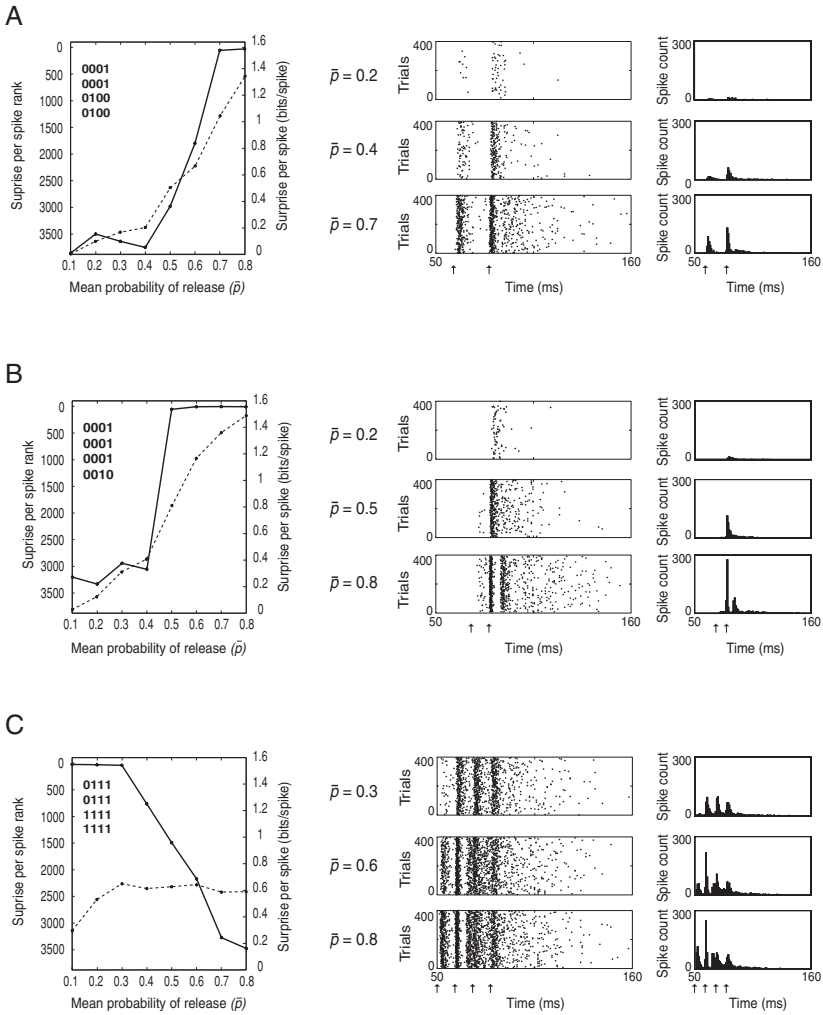
Which stimuli were the most sensitive to release probability changes? An answer was obtained by considering the stimuli showing the largest and the steepest ranking changes in their surprise-per-spike values following release probability modifications. The largest surprise-per-spike increases were observed for short stimuli with at least two coactivated spikes in one bin (see Figure 8A and Supplementary Material, Figure S6). A similar spatiotemporal structure was typically observed for the steepest, and therefore most rapid, ranking transitions in surprise per spike (see Figure 8B and Supplementary Material, Figure S7). For these short stimuli, their surprise per spike increased significantly following small plasticity changes ($\Delta\bar{p} = 0.1$). As shown by the spikegrams and the PSTHs of Figures 8A and 8B, the stimuli with the largest and steepest ranking increases of surprise per spike produced little activity at low $\bar{p}$, whereas they elicited better time-locked responses at high $\bar{p}$. Finally, the largest and the steepest surprise per spike ranking decreases were observed for long and highly correlated stimuli (for which most of the bins were filled with three or four coactivated spikes, Figure 8C and Supplementary Material, Figure S8).

3 Discussion

This letter provides an analysis of information transmission over a region of the excitatory input space of a neuron. The direct method to compute mutual information, MI (Zador, 1998), adopted in this work is usually computationally impracticable, and it was made possible by the peculiar structure of the studied system, the cerebellar granule cell (GC) (see Figure 1). Indeed, the fact that GCs are electronically compact and receive on average only four mossy fiber (MF) excitatory inputs reduces the state-space dimensionality drastically. Still, the MF-GC system is representative because MF-GC synapses call on the same complex mechanisms mediating information processing at most brain synapses (D'Angelo et al., 1995; Nielsen et al., 2004; Nieus et al., 2006; Sola et al., 2004).

The information-theoretic study presented here explored the relative contribution of various factors (e.g., spike timing, specific input patterns)

to neurotransmission and investigated how their interrelations changed following long-term plasticity. The analysis showed that the MF-GC relay transmitted information via firing rate, spike timing, and spike correlation of MF discharge. Along with its ability to regulate spike frequency and timing (Nieus et al., 2006), the increase of release probability (p) accompanying LTP regulated the amount of transmitted information (Zador, 1998). The main observation is that, on average, information was carried almost equally by spike frequency and spike timing, and it increased quasi-linearly with p. On the other hand, it was shown that maximizing release



probability is not a necessary condition to optimize stimulus-specific information for most stimuli. Indeed, single-stimulus transmission reached a plateau at intermediate p values for most MF inputs, with optimal transmission taking place over a large p range (from intermediate values to the maximum). Interestingly, in this range, the preferred stimulus patterns were very sensitive to p variations; in other words, small variations of p changed the subset of stimuli that are best transmitted by the cell. Such intermediate release probability values reflect the values observed experimentally in the cerebellar granular layer (Saviane & Silver, 2006; Sola et al., 2004), as well as in other brain areas (e.g., neocortical pyramid-to-pyramid connections: Koester & Johnston, 2005; CA3-CA1 hippocampal synapses: Dobrunz & Stevens, 1997).

Figure 8: Surprise-per-spike changes for exemplar stimuli with highest information transmission modifications following release probability increases: simulation results. (A–C) Information transmission for stimuli with the fourth largest increase (A), the fourth steepest increase (B), and the fourth largest decrease (C) in the rank of the surprise per spike as a function of $\bar{p}$ (left panels: the continuous lines indicate the changes in the rank, whereas the dashed curves denote the absolute values of the surprise per spike). The middle and right panels show the GC responses to the stimulus as spikegrams and PSTHs, respectively, for three different $\bar{p}$. The steepest increases and decreases typically occurred between 0.4–0.6 and 0.3–0.4, respectively. In *C*, at $\bar{p} = 0.6$ and 0.8, note the spike doublets in response to the second set of input spikes at ∼62 and ∼68 ms before the third set of input spikes at 70 ms. This doublet perturbed the timing of the following spikes as seen by the wider spread in the PSTH. This spread is largely responsible for the decline in rank for this stimulus at high $\bar{p}$. In *B*, note the persistent high rank for this stimulus after the rank increase at $\bar{p} = 0.5$ (left panel). A doublet to the three-spike bin (second arrow) starts appearing at $\bar{p} = 0.6$ (not shown) and is clear at $\bar{p} = 0.8$ with spikes at ∼82 and ∼88 ms in response to the second set of input spikes (middle and right panels). Hence, the doublet appearance at $\bar{p} = 0.6$ is not the cause of the rank increase at $\bar{p} = 0.5$ or of a transmission decrease as for the stimuli in *C* since its rank remains high beyond $\bar{p} = 0.6$. Thus, when *B* is compared with *A* and *C*, the effect of the appearance of doublets on surprise per spike is highly stimulus dependent. For the left panels, the 3876 different stimuli obtained with the same p over the four synapses ($\bar{p} = p$) were ranked according to the surprise per spike obtained at a specific p between 0.1 and 0.8. The 10 largest and steepest increases and decreases in rank were analyzed. In general, it was found that in short notation (number of spikes per bin with implicit permutation), the largest increases were for stimuli of the form x2xx or x202 with x = 0, 1; whereas the steepest increases were for x2xx or x3xx, x = 0, 1, 2; and the largest decreases were for xyy4, x = 1,2,3,4; y = 3,4.

In this study, the influence of the Golgi cells (GoCs) was omitted to focus on the GC encoding of its excitatory inputs only. In fact, the series of analyses presented here did not aim at providing a comprehensive characterization of the information transmission process in the granular layer; rather, it was meant to set forth an information-theoretic framework suitable for studying the efficacy of GC information processing quantitatively. At this stage, GoC inhibition was not included because a thorough understanding of the synaptic transmission at the MF-GoC relay is still missing (Forti, Cesana, Isope, Dieudonné, & D'Angelo, 2008). In addition, the study of quantal release properties of GoC-GC synapses has been addressed only very recently (Mapelli, Rossi, Nieus, & D'Angelo, 2009). Further analyses will be carried out to quantify the effects of Golgi inhibition onto GC neurotransmission by means of a cerebellar GABA synaptic model accounting for recent experimental evidence (Mapelli et al., 2009) at this synapse (see also Bezzi, Nieus, Arleo, D'Angelo, & Coenen, 2004, for preliminary work in this direction). GoCs fire autorhythmically at 2 to 4 Hz in vitro (Forti, Cesana, Mapelli, & D'Angelo, 2006) and at around 10 Hz in vivo (Vos, Wijnants, Taeymans, & de Schutter, 1999), and their discharge frequency can be raised up to 300 Hz on stimulation. It is worth mentioning that even under such strong inhibitory conditions, GCs are still able to display short-burst activity (D'Angelo & De Zeeuw, 2009). Following stimulation of the whisker pad, GoCs in anesthetized rats follow a stereotyped firing pattern characterized by bursts of two or three spikes, followed by pauses of around 100 ms. Moreover, the cerebellum displays oscillatory activity at theta frequency (Hartmann & Bower, 1998), and a cerebellar network model suggested that these oscillations are characterized by a highly synchronous activity of GoCs and GCs (Maex & Schutter, 1998). In all these studies, the GoCs appear to regulate the narrow time window during which GCs may fire (D'Angelo & De Zeeuw, 2009; Kistler & De Zeeuw, 2003; Solinas et al., 2007a, 2007b).

The synaptic model used here was stochastic (see section 4; see also Sun, Lyons, & Dobrunz, 2005). It presented both facilitation and depression depending on the release probability p and spike activity (Nieus et al., 2006; Tsodyks & Markram, 1997). Hence, the calculations of information presented here extend those that have thus far been investigated using depressing synapse (de la Rocha & Parga, 2005; Goldman, 2004), and constant-transmission-probability models (Manwani & Koch, 2001; Zador, 1998). Recently an information-theoretic measure, namely, the synaptic information efficacy (SIE) (London et al., 2002), has been introduced to quantify the transmitting properties of a single synapse. SIE is actually the mutual information (MI) between the presynaptic input spike trains and the postsynaptic responses. In the study by London et al. (2002), MI was measured across only a single synapse, while considering the rest of the dendritic input as background noise. Similarly, other studies using different analytical approaches have focused on the transmitting properties of single synapses

(de la Rocha et al., 2002; Goldman, 2004; Manwani et al., 2002; Tiesinga, 2001). In addition, although input correlations have been shown to be relevant to neuronal computation (de la Rocha & Parga, 2005; Schreiber, Fellous, Tiesinga, & Sejnowski, 2004; Tiesinga & Toups, 2005), many studies have investigated conditions in which they are absent (London et al., 2002; Manwani & Koch, 2001; Zador, 1998). This study complements these previous works by presenting a quantitative characterization of how the correlation across multiple synaptic inputs of a neuron may contribute to information transmission.

Our results corroborate the hypothesis that the correlation among the inputs plays a crucial role in affecting neurotransmission at all levels of release probability p. Our analysis strengthened this observation by showing how the spatiotemporal structure of MF inputs affected GC neurotransmission as a function of p. The most sensitive stimuli to release probability changes were identified in terms of both surprise and surprise per spike. Short correlated MF bursts resulted to benefit the most from p increases. Long correlated stimuli caused changes in neurotransmission and excitation dynamics at high p, which bounded their transmission reliability. These results are consistent with previous findings suggesting that when the synapses are tuned toward specific input stimuli, one of the roles of LTP and LTD may be that of generating spike train–specific nonlinear detectors (Natschlager & Maass, 2001; Sharpee et al., 2006), which would regulate the transmission of specific spatiotemporal input patterns at the level of the neuron. The preferential recoding of the information contained in certain patterns may be of biological relevance for computation at subsequent stages in the cerebellar cortex. For instance, two spikes in close succession lead to the opening of presynaptic NMDA channels and start a cascade of events leading to long-term plasticity at the synapses between the Purkinje cells (PCs) (Casado, Isope, & Ascher, 2002) and the parallel fibers (PFs), which are the axons of the GCs. Recoding following granular layer plasticity could also have the effect of reducing the length of PF spike trains, thereby increasing the timing accuracy of the PC responses.

Naturally, to provide definite answers beyond the neuronal level, coding strategies need to be evaluated within a cerebellar network (Coenen, Arnold, Sejnowski, & Jabri, 2001; Philipona & Coenen, 2004) by means of large-scale simulations, where plasticity may also regulate the average GC population firing and the duration of multiple PF activation necessary to elicit a PC response. Therefore, the information-theoretic approach presented here constitutes a step forward in the investigation of neural information transfer in the granular layer network of the cerebellum. Scaling from elementary cellular mechanisms such as synaptic release probability to network computation is fundamental to understanding how MF-GC long-term plasticity, by being instrumental in the control of information transmission, may regulate the operations to which the cerebellum participates (Dum, Li, & Strick, 2002).

4 Materials and Methods

4.1 Biophysical Model of the MF-GC Synaptic Relay. A stochastic version of our previous models (D'Angelo et al., 2001; Nieus et al., 2006) was developed for this study (see also Sun et al., 2005). The GC model was provided with four independent MF-GC synaptic contacts endowed with stochastic neurotransmitter release mechanisms. Each synaptic contact consisted of three independent releasing sites (RS) (Saviane & Silver, 2006; Sola et al., 2004), each governed by a three-state presynaptic model (Tsodyks & Markram, 1997). Neurotransmitter release was modeled as a system of four first-order differential equations relating the probability of release (p) to the available (X), released (Y), and recovered (Z) neurotransmitter resources (Nieus et al., 2006; Tsodyks & Markram, 1997) (see Supplementary Material). The release was made probabilistic and modeled as an all-or-none process by comparing a random number (ε) drawn from a uniform [0, 1] probability distribution with the released resources Y. Hence, a wave of glutamate (Nieus et al., 2006) was released from the RS whenever ε was less than Y, so that averaging over numerous independent Ys led to recovery of the behavior of the deterministic release model (Nieus et al., 2006; Saviane & Silver, 2006; Sola et al., 2004). Each RS activated an independent postsynaptic site endowed with AMPA and NMDA receptors. Because three RS/contact (Saviane & Silver, 2006) were introduced, each postsynaptic site was calibrated to contribute one-third of the total conductance (Nieus et al., 2006). Since no evidence for p heterogeneity was reported (Sola et al., 2004), p was set at the same value for all releasing sites at the same synapse. (Further details on the biophysical model can be found in Supplementary Material, section S1.1.)

First, a series of simulations was run by adopting the same stimulation protocol used for the patch clamp experiments (see section 2 and Figure 2). Then the input space was extended by considering the four MF afferents as four independent spike trains and regulating the neurotransmitter release probability p at each MF-GC synapse independently (e.g., see Figure 3). All possible combinations of input spike trains lasting up to 40 ms, as binary words of 10 ms bins (i.e. with a maximum rate of 100 Hz), were then created. Thus, each MF spike train was encoded as a four-bit binary word, and the cell input was a 4 (number of MFs) $\times$ 4 (number of bits for each MF) $=$ 16 bit binary word. Accordingly, the entire input space consisted of $2^{16} = 65536$ stimuli. GC responses were digitalized using temporal bins of 6 ms over a period of 120 ms (see Supplementary Material, section S2.1, for a discussion on the dependence of the results upon the bin size used to sample GC responses). Each stimulus was presented 400 times, and the effect of limited sampling on MI computation was taken into account (see Supplementary Material, section S1.3).

To simulate natural firing activity in MFs, all 65,536 stimuli occurring with different probabilities, $p(s)$, were considered. The actual firing

distribution of MFs is not known, but following Treves, Panzeri, Rolls, Booth, and Wakeman (1999), a continuous unimodal distribution (with a single peak close to the spontaneous activity, i.e., close to zero in our case) with an exponential tail was assumed. This distribution has been observed in other brain areas, among them, the frontal cortex (Abeles, Vaadia, & Bergman, 1990), hippocampus and close structures (Barnes, McNaughton, Mizumori, Leonard, & Lin, 1990), visual cortex (Baddeley et al., 1997; Franco, Rolls, Aggelopoulos, & Jerez, 2007), as well as used in previous neural network models (Treves & Rolls, 1991). For this study, a decreasing exponential distribution of firing rate with a mean of 10 Hz was chosen. Another distribution with similar features (i.e., Poisson) was also tested, and compatible results were obtained (see Supplementary Material, section S2.3).

The model was implemented with NEURON 5.4. The most demanding numerical simulations were run on a cluster of 20 CPUs (1.7 GHz each).

4.2 Electrophysiological Recordings. Whole-cell patch clamp recordings of GCs were performed from acute cerebellar slices of P18-P23 Wistar rats according to published procedures (D'Angelo et al., 1995; Sola et al., 2004) (see Supplementary Material, section S1.2). Varying the intensity of MF stimulation permitted generating simultaneous activity in one to four MFs. The core experiment was carried out in current-clamp mode. MFs were stimulated by a set of spike trains lasting 40 ms and with a frequency up to 100 Hz. To investigate the effect of long-term synaptic plasticity on MI, the same set of stimuli was presented before and after LTP induction by theta-burst stimulation (TBS: 8×100 Hz bursts lasting 100 ms every 250 ms) (Nieus et al., 2006; Sola et al., 2004) (see Supplementary Material, Figure S2). The mean MF-GC release probability $\bar{p}$ was estimated in voltage clamp, at the beginning and at the end of each experiment, by analyzing the excitatory postsynaptic currents (EPSCs) elicited by four stimuli at 100 Hz (Nieus et al., 2006; see Supplementary Material, Figure S3).

GC responses were spike-sorted via a threshold crossing procedure (Igor; WaveMetrics Inc.). To compute MI, both input (MF) and output (GC) spike trains were transformed into binary words (see Figure 1). MF spike trains were digitized by using a temporal bin of 10 ms. Because each stimulus lasted 40 ms and all MFs received the same spike train, the input set consisted of $2^4 = 16$ stimuli. Each stimulus was presented 25 times. The GC responses were digitized by using a 6 ms bin width. A 60 ms time window was used for sampling the GC output; the spikes occurring beyond this time window (less than 1% of all spikes) were not taken into account to reduce the dimensionality of the sampling space. The significance of MI measurements was assessed by a bootstrapping procedure (Lu & Wang, 2004) of 1000 repetitions. MI proved to be statistically significant ($p < 0.05$) for MI > 0.4 bits.

All experiments were carried out according to the guidelines and regulations laid down by the institution's animal welfare committee.

4.3 Theoretical Analysis

4.3.1 Mutual Information and "Surprise." The mean information transmitted between the GC responses r and the MF inputs s was calculated by using Shannon's mutual information (MI), equation 1.1, where both the stimuli s and the responses r were represented as either binary words (see Figure 1) or spike counts. The binary word coding preserves the information about spike timing, up to a certain precision, whereas the spike count coding simply assesses the information transmitted by the neuron average firing. The data processing inequality (Cover & Thomas, 1991) ensures that MI obtained using spike counts is always not greater than the MI obtained using binary words.

To isolate the specific contribution of a single stimulus s to MI, the stimulus specific surprise (called simply *surprise* here) was used (Butts & Goldman, 2006; DeWeese & Meister, 1999):

$$I_1(s) = \sum_{r \in R} p(r \mid s) \log_2 \frac{p(r \mid s)}{p(r)}. \tag{4.1}$$

The surprise measures how much the conditional distribution $p(r \mid s)$ differs from the prior probability distribution $p(r)$, corresponding to the so-called Kullback-Leibler distance. The surprise per spike was then computed by dividing the surprise by the spike count of the input stimulus. Normalizing by the number of input spikes essentially removes the linear part of the contribution of MF firing rates to information transmission.

Both MI and the surprise measure have a systematic bias due to the limited data samples available (Nemenman, Bialek, & de Ruyter van Steveninck, 2004; Paninski, 2003; Strong, Koberle, de Ruyter van Steveninck, & Bialek, 1998; Treves & Panzeri, 1995). This issue is treated in the Supplementary Material, section S1.3.

4.3.2 Correlation Measure. To estimate the number of coincident spikes across the four MF inputs, the normalized average of the pair-wise correlation over all the different pairs of MFs was computed. Let $s \in S$ be a stimulus constituted by four MF patterns, and let x_i and x_j denote the input spike trains representing the activity on the ith and the jth MF, respectively. The correlation coefficient $C(s)$ was taken as

$$C(s) = \frac{1}{N} \cdot \langle (K(x_i) - \bar{x}) \cdot (K(x_j) - \bar{x})' \rangle_{i>j},$$

where $\bar{x}$ represents the average firing rate, N is the normalization factor equal to the number of input bins, and $K(x)$ denotes a smoothing function (i.e., an exponential kernel with τ equal to the input bin width) that takes into account the effects of the short-time integration over the membrane

time constant. The $C(s)$ vector, for all $s \in S$, was then rescaled into the range [0, 1]. For example, $C(s) = 0.4039$ for $s = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$, whereas $C(s) = 0.1561$ for $s = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$.

Acknowledgments

We kindly acknowledge the contribution of David Gall (Université Libre de Bruxelles) for the confocal image in Figure 1. This work was supported by European projects SENSOPAC IST-2005–028056 (www.sensopac.org) and SpikeFORCE IST-2001-35271 (www.spikeforce.org), both created by O.J.-M.D.C., by Sony CSL-Paris, and by the Italian MIUR project PRIN-2004053317 and CNR-INFM project FIRB-RBNE01AAS7 to E.D'A.

References

Abbott, L. F., & Regehr, W. G. (2004). Synaptic computation. *Nature*, *431*, 796–803.

Abeles, M., Vaadia, E., & Bergman, H. (1990). Firing patterns of single units in the prefrontal cortex and neural network models. *Network*, *1*, 13–25.

Albus, J. S. (1971). A theory of cerebellar function. *Math. Biosc.*, *10*, 25–61.

Arenz, A., Silver, R. A., Schaefer, A. T., & Margrie, T. W. (2008). The contribution of single synapses to sensory representation in vivo. *Science*, *321*, 977–980.

Baddeley, R., Abbott, L. F., Booth, M. C., Sengpiel, F., Freeman, T., Wakeman, E. A., et al. (1997). Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. Biol. Sci.*, *264*, 1775–1783.

Barnes, C. A., McNaughton, B. L., Mizumori, S. J., Leonard, B. W., & Lin, L. H. (1990). Comparison of spatial and temporal characteristics of neuronal activity in sequential stages of hippocampal processing. *Prog. Brain Res.*, *83*, 287–300.

Bezzi, M., Nieus, T., Arleo, A., D'Angelo, E., & Coenen, O. J. M. (2004). *Information transfer at the mossy fiber-granule cell synapse of the cerebellum.* Society for Neuroscience Abstracts No. 827.5, San Diego.

Bialek, W., Rieke, F., de Ruyter van Steveninck, R. R., & Warland, D. (1991). Reading a neural code. *Science*, *252*, 1854–1857.

Bliss, T. V., & Collingridge, G. L. (1993). A synaptic model of memory: Long-term potentiation in the hippocampus. *Nature*, *361*, 31–39.

Bliss, T. V., & Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J. Physiol.*, *232*, 331–356.

Borst, A., & Theunissen, F. E. (1999). Information theory and neural coding. *Nat. Neurosci.*, *2*, 947–957.

Brenner, N., Bialek, W., & de Ruyter van Steveninck, R. (2000). Adaptive rescaling maximizes information transmission. *Neuron*, *26*, 695–702.

Brenner, N., Strong, S. P., Koberle, R., Bialek, W., & de Ruyter van Steveninck, R. R. (2000). Synergy in a neural code. *Neural. Comput.*, *12*, 1531–1552.

Butts, D. A., & Goldman, M. S. (2006). Tuning curves, neuronal variability, and sensory coding. *PLoS Biol.*, *4*, e92.

Casado, M., Isope, P., & Ascher, P. (2002). Involvement of presynaptic N-methyl-D-aspartate receptors in cerebellar long-term depression. *Neuron*, *33*, 123–130.

Chadderton, P., Margrie, T. W., & Hausser, M. (2004). Integration of quanta in cerebellar granule cells during sensory processing. *Nature*, *428*, 856–860.

Coenen, O. J.-M. D., Arnold, M. P., Sejnowski, T. J., & Jabri, M. A. (2001). Parallel fiber coding in the cerebellum for life-long learning. *Autonomous Robots*, *11*, 291–297.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory.* New York: Wiley.

D'Angelo, E., De Filippi, G., Rossi, P., & Taglietti, V. (1995). Synaptic excitation of individual rat cerebellar granule cells in situ: Evidence for the role of NMDA receptors. *J. Physiol.*, *484*, 397–413.

D'Angelo, E., & De Zeeuw, C. I. (2009). Timing and plasticity in the cerebellum: Focus on the granular layer. *Trends Neurosci.*, *32*, 30–40.

D'Angelo, E., Nieus, T., Maffei, A., Armano, S., Rossi, P., Taglietti, V., et al. (2001). Theta-frequency bursting and resonance in cerebellar granule cells: Experimental evidence and modeling of a slow k+-dependent mechanism. *J. Neurosci.*, *21*, 759–770.

de la Rocha, J., Nevado, A., & Parga, N. (2002). Information transmission by stochastic synapses with short-term depression: Neural coding and optimization. *Neurocomputing*, *44–46*, 85–90.

de la Rocha, J., & Parga, N. (2005). Short-term synaptic depression causes a non-monotonic response to correlated stimuli. *J. Neurosci.*, *25*, 8416–8431.

de Ruyter van Steveninck, R. R., Lewen, G. D., Strong, S. P., Koberle, R., & Bialek, W. (1997). Reproducibility and variability in neural spike trains. *Science*, *275*, 1805–1808.

DeWeese, M. R., & Meister, M. (1999). How to measure the information gained from one symbol. *Network*, *10*, 325–340.

Dimitrov, A. G., Miller, J. P., Gedeon, T., Aldworth, Z., & Parker, A. E. (2003). Analysis of neural coding through quantization with an information-based distortion measure. *Network*, *14*, 151–176.

Dobrunz, L. E., & Stevens, C. F. (1997). Heterogeneity of release probability, facilitation, and depletion at central synapses. *Neuron*, *18*, 995–1008.

Dum, R. P., Li, C., & Strick, P. L. (2002). Motor and nonmotor domains in the monkey dentate. *Ann. N. Y. Acad. Sci.*, *978*, 289–301.

Eccles, J. C., Ito, M., & Szentagothai, J. (1967). *The cerebellum as a neuronal machine.* Berlin: Springer Verlag.

Forti, L., Cesana, E., Isope, P., Dieudonné, S., & D'Angelo, E. (2008). The mossy fiber input to Golgi cells in the cerebellum. FENS Abst., vol. 4, 221.5.

Forti, L., Cesana, E., Mapelli, J., & D'Angelo, E. (2006). Ionic mechanisms of autorhythmic firing in rat cerebellar Golgi cells. *J. Physiol.*, *574*, 711–729.

Franco, L., Rolls, E. T., Aggelopoulos, N. C., & Jerez, J. M. (2007). Neuronal selectivity, population sparseness, and ergodicity in the inferior temporal visual cortex. *Biol. Cybern.*, *96*, 547–560.

Fuhrmann, G., Segev, I., Markram, H., & Tsodyks, M. (2002). Coding of temporal information by activity-dependent synapses. *J. Neurophysiol.*, *87*, 140–148.

Goldman, M. S. (2004). Enhancement of information transmission efficiency by synaptic failures. *Neural Comput.*, *16*, 1137–1162.

Hansel, C., Linden, D. J., & D'Angelo, E. (2001). Beyond parallel fiber LTD: The diversity of synaptic and non-synaptic plasticity in the cerebellum. *Nat. Neurosci.*, *4*, 467–475.

Hartmann, M. J., & Bower, J. M. (1998). Oscillatory activity in the cerebellar hemispheres of unrestrained rats. *J. Neurophysiol.*, *80*, 1598–1604.

Jakab, R. L., & Hamori, J. (1988). Quantitative morphology and synaptology of cerebellar glomeruli in the rat. *Anat. Embryol. (Berl.)*, *179*, 81–88.

Jorntell, H., & Ekerot, C. F. (2006). Properties of somatosensory synaptic integration in cerebellar granule cells in vivo. *J. Neurosci.*, *26*, 11786–11797.

Kistler, W. M., & De Zeeuw, C. I. (2003). Time windows and reverberating loops: A reverse-engineering approach to cerebellar function. *Cerebellum*, *2*, 44–54.

Koch, C., & Segev, I. (2000). The role of single neurons in information processing. *Nat. Neurosci.*, *3* (Suppl.), 1171–1177.

Koester, H. J., & Johnston, D. (2005). Target cell–dependent normalization of transmitter release at neocortical synapses. *Science*, *308*, 863–866.

London, M., Schreibman, A., Hausser, M., Larkum, M. E., & Segev, I. (2002). The information efficacy of a synapse. *Nat. Neurosci.*, *5*, 332–340.

Lu, T., & Wang, X. (2004). Information content of auditory cortical responses to time-varying acoustic stimuli. *J. Neurophysiol.*, *91*, 301–313.

Maex, R., & Schutter, E. D. (1998). Synchronization of Golgi and granule cell firing in a detailed network model of the cerebellar granule cell layer. *J. Neurophysiol.*, *80*, 2521–2537.

Magistretti, J., Castelli, L., Forti, L., & D'Angelo, E. (2006). Kinetic and functional analysis of transient, persistent, and resurgent sodium currents in rat cerebellar granule cells in situ: An electrophysiological and modelling study. *J. Physiol.*, *573*, 83–106.

Malenka, R. C., & Bear, M. F. (2004). LTP and LTD: An embarrassment of riches. *Neuron*, *44*, 5–21.

Manwani, A., & Koch, C. (2001). Detecting and estimating signals over noisy and unreliable synapses: Information-theoretic analysis. *Neural Comput.*, *13*, 1–33.

Manwani, A., Steinmetz, P. N., & Koch, C. (2002). The impact of spike timing variability on the signal-encoding performance of neural spiking models. *Neural Comput.*, *14*, 347–367.

Mapelli, J., & D'Angelo, E. (2007). The spatial organization of long-term synaptic plasticity at the input stage of cerebellum. *J. Neurosci.*, *27*, 1285–1296.

Mapelli, L., Rossi, P., Nieus, T., & D'Angelo, E. (2009). Tonic activation of GABA-B receptors reduces release probability at inhibitory connections in the cerebellar glomerulus. *J. Neurophysiol.*, *101*(6), 3089–3099.

Mitchell, S. J., & Silver, R. A. (2003). Shunting inhibition modulates neuronal gain during synaptic excitation. *Neuron*, *38*(3), 433–445.

Moreno-Bote, R., & Parga, N. (2004). Role of synaptic filtering on the firing response of simple model neurons. *Phys. Rev. Lett.*, *92*, 028102.

Natschlager, T., & Maass, W. (2001). Computing the optimally fitted spike train for a synapse. *Neural Comput.*, *13*, 2477–2494.

Nemenman, I., Bialek, W., & de Ruyter van Steveninck, R. R. (2004). Entropy and information in neural spike trains: Progress on the sampling problem. *Phys. Rev. E*, *69*, 056111.

Nielsen, T. A., DiGregorio, D. A., & Silver, R. A. (2004). Modulation of glutamate mobility reveals the mechanism underlying slow-rising AMPAR EPSCs and the diffusion coefficient in the synaptic cleft. *Neuron*, *42*, 757–771.

Nieus, T., Sola, E., Mapelli, J., Saftenku, E., Rossi, P., & D'Angelo, E. (2006). LTP regulates burst initiation and frequency at mossy fiber–granule cell synapses of rat cerebellum: Experimental observations and theoretical predictions. *J. Neurophysiol.*, *95*, 686–699.

Paninski, L. (2003). Estimation of entropy and mutual information. *Neural Comput.*, *15*, 1191–1254.

Philipona, D., & Coenen, O. J.-M. D. (2004). Model of granular layer encoding of the cerebellum. *Neurocomputing*, *58–60*, 575–580.

Quian Quiroga, R., & Panzeri, S. (2009). Extracting information from neuronal populations: Information theory and decoding approaches. *Nat. Rev Neurosci.*, *10*, 173–185.

Rancz, E. A., Ishikawa, T., Duguid, I., Chadderton, P., Mahon, S., & Hausser, M. (2007). High-fidelity transmission of sensory information by single cerebellar mossy fibre boutons. *Nature*, *450*, 1245–1248.

Roddey, J. C., & Jacobs, G. A. (1996). Information theoretic analysis of dynamical encoding by filiform mechanoreceptors in the cricket cercal system. *J. Neurophysiol.*, *75*, 1365–1376.

Rolls, E. T., & Deco, G. (2002). *Computational neuroscience of vision.* Oxford: Oxford University Press.

Rothman, J. S., Cathala, L., Steuber, V., & Silver, R. A. (2009). Synaptic depression enables neuronal gain control. *Nature*, *457*, 1015–1018.

Sargent, P. B., Saviane, C., Nielsen, T. A., DiGregorio, D. A., & Silver, R. A. (2005). Rapid vesicular release, quantal variability, and spillover contribute to the precision and reliability of transmission at a glomerular synapse. *J. Neurosci.*, *25*, 8173–8187.

Saviane, C., & Silver, R. A. (2006). Fast vesicle reloading and a large pool sustain high bandwidth transmission at a central synapse. *Nature*, *439*, 983–987.

Schreiber, S., Fellous, J. M., Tiesinga, P., & Sejnowski, T. J. (2004). Influence of ionic conductances on spike timing reliability of cortical neurons for suprathreshold rhythmic inputs. *J. Neurophysiol.*, *91*, 194–205.

Schweighofer, N., Doya, K., & Lay, F. (2001). Unsupervised learning of granule cell sparse codes enhances cerebellar adaptive control. *Neuroscience*, *103*, 35–50.

Shannon, C. E. (1948). The mathematical theory of communication. *Bell Syst. Tech. J.*, *27*, 379–423.

Sharpee, T. O., Sugihara, H., Kurgansky, A. V., Rebrik, S. P., Stryker, M. P., & Miller, K. D. (2006). Adaptive filtering enhances information transmission in visual cortex. *Nature*, *439*, 936–942.

Silver, R. A., Traynelis, S. F., & Cull-Candy, S. G. (1992). Rapid-time-course miniature and evoked excitatory currents at cerebellar synapses in situ. *Nature*, *355*, 163–166.

Smith, E. C., & Lewicki, M. S. (2006). Efficient auditory coding. *Nature*, *439*, 978–982.

Sola, E., Prestori, F., Rossi, P., Taglietti, V., & D'Angelo, E. (2004). Increased neurotransmitter release during long-term potentiation at mossy fibre-granule cell synapses in rat cerebellum. *J. Physiol.*, *557*, 843–861.

Solinas, S., Forti, L., Cesana, E., Mapelli, J., De Schutter, E., & D'Angelo, E. (2007a). Computational reconstruction of pacemaking and intrinsic electroresponsiveness in cerebellar Golgi cells. *Front Cell Neurosci.*, *1*, 2.

Solinas, S., Forti, L., Cesana, E., Mapelli, J., De Schutter, E., & D'Angelo, E. (2007b). Fast-reset of pacemaking and theta-frequency resonance patterns in cerebellar Golgi cells: Simulations of their impact in vivo. *Front Cell Neurosci.*, *1*, 4.

Strong, S. P., Koberle, R., de Ruyter van Steveninck, R. R., & Bialek, W. (1998). Entropy and information in neural spike trains. *Phys. Rev. Lett.*, *80*(1), 197–200.

Sun, H. Y., Lyons, S. A., & Dobrunz, L. E. (2005). Mechanisms of target-cell specific short-term plasticity at Schaffer collateral synapses onto interneurones versus pyramidal cells in juvenile rats. *J. Physiol.*, *568*, 815–840.

Theunissen, F. E., & Miller, J. P. (1991). Representation of sensory information in the cricket cercal sensory system. II. Information theoretic calculation of system accuracy and optimal tuning-curve widths of four primary interneurons. *J. Neurophysiol.*, *66*, 1690–1703.

Theunissen, F., Roddey, J. C., Stufflebeam, S., Clague, H., & Miller, J. P. (1996). Information theoretic analysis of dynamical encoding by four identified primary sensory interneurons in the cricket cercal system. *J. Neurophysiol.*, *75*, 1345–1364.

Tiesinga, P. H. (2001). Information transmission and recovery in neural communication channels revisited. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, *64*, 012901.

Tiesinga, P. H., Fellous, J. M., Jose, J. V., & Sejnowski, T. J. (2002). Information transfer in entrained cortical neurons. *Network*, *13*, 41–66.

Tiesinga, P. H., & Toups, J. V. (2005). The possible role of spike patterns in cortical information processing. *J. Comput. Neurosci.*, *18*, 275–286.

Treves, A., & Panzeri, S. (1995). The upward bias in measures of information derived from limited data samples. *Neural Comput.*, *7*, 399–407.

Treves, A., Panzeri, S., Rolls, E. T., Booth, M., & Wakeman, E. A. (1999). Firing rate distributions and efficiency of information transmission of inferior temporal cortex neurons to natural visual stimuli. *Neural Comput.*, *11*, 601–632.

Treves, A., & Rolls, E. T. (1991). What determines the capacity of autoassociative memories in the brain? *Network*, *2*, 371–398.

Tsodyks, M. V., & Markram, H. (1997). The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc. Natl. Acad. Sci. USA*, *94*, 719–723.

Vos, B. P., Wijnants, M., Taeymans, S., & de Schutter, E. (1999). Miniature carrier with six independently moveable electrodes for recording of multiple single-units in the cerebellar cortex of awake rats. *J. Neurosci. Methods*, *94*, 19–26.

Wan, Y. H., Jian, Z., Wen, Z. H., Wang, Y. Y., Han, S., Duan, Y. B., et al. (2004). Synaptic transmission of chaotic spike trains between primary afferent fiber and spinal dorsal horn neuron in the rat. *Neuroscience*, *125*, 1051–1060.

Zador, A. (1998). Impact of synaptic unreliability on the information transmitted by spiking neurons. *J. Neurophysiol.*, *79*, 1219–1229.

Received September 18, 2007; accepted February 16, 2010.

Supplement to "How Synaptic Release Probability Shapes Neuronal Transmission: Information-Theoretic Analysis in a Cerebellar Granule Cell" by Angelo Arleo, Thierry Nieus, Michele Bezzi, Anna D'Errico, Egidio D'Angelo, and Olivier J.-M. D. Coenen, *Neural Computation*, Vol. 22, No. 8 (August 2010), pp. 2031–2058.

URL: http://www.mitpressjournals.org/doi/abs/10.1162/NECO_a_00006-Arleo

Supplementary Material

S1. Supplementary Methods

S1.1 Biophysical model of the mossy fiber – granule cell (MF – GC) synaptic relay: equations and parameter settings

Cell model

The GC model was based on the Hodgkin-Huxley formalism according to which the membrane potential V (for a single compartment) was computed by solving the ordinary differential Eqs.1,3,5:

$$\frac{dV}{dt} = -\frac{1}{C_m}\left\{\sum_i g_i(V - V_i) + I_{INJ}\right\} \tag{1}$$

where C_m is the membrane capacitance, g_i and V_i are the ionic conductance and reversal potential of channel i, respectively, and I_{INJ} denotes an external driving current. The dynamics of each ionic conductance g_i was taken equal to:

$$g_i = g_{\max} \cdot x_i^m \cdot y_i^n \tag{2}$$

with $g_{\max}$ being the maximum ionic conductance, x_i and y_i gating variables (probabilities) for activation and inactivation, respectively, and the power values m, n corresponding to the number of gating particles in the ionic channel i. The kinetics of the gating variables x and y (with the index i omitted) were defined by:

$$\frac{dx}{dt} = \frac{x_\infty - x}{\tau_x}$$
$$\frac{dy}{dt} = \frac{y_\infty - y}{\tau_y} \tag{3}$$

where the steady state values x_∞, y_∞ and the time constants τ_x, τ_y for activation and inactivation, respectively, were given by:

$$x_\infty = \frac{\alpha_x}{(\alpha_x + \beta_x)} \qquad \tau_x = \frac{1}{(\alpha_x + \beta_x)}$$
$$y_\infty = \frac{\alpha_y}{(\alpha_y + \beta_y)} \qquad \tau_y = \frac{1}{(\alpha_y + \beta_y)} \tag{4}$$

with α_x, β_x and α_y, β_y being functions of the membrane potential V. All the equations and parameter settings characterizing the ionic channels included in the GC model are reported on Table S1 (a more detailed description of their biological significance can be found in D'Angelo et al., 2001).

The ionic channel K-Ca is a potassium channel whose activation depends on both voltage and inner calcium concentration. Therefore, an auxiliary ODE to describe calcium transients was included in the model:

$$\frac{d[Ca^{2+}]}{dt} = -\frac{I_{Ca}}{2FAD} - \beta_{Ca}([Ca^{2+}] - [Ca^{2+}]_0) \tag{5}$$

where F is the Faraday constant, D denotes the depth of a shell adjacent to the cell surface of area A, β_{Ca} determines the loss of Ca^{2+} from the shell, and $[Ca^{2+}]_0$ is the resting calcium concentration.

Synaptic model

Synaptic currents were described according to the following conductance-based approach:

$$g_j = g_{\max} 0_j \tag{6}$$

where j denotes either AMPA or NMDA conductances, and 0_j was obtained from the kinetic scheme shown on Figure S1 (A) and the kinetic rates reported in Table S2 (see Nieus et al., 2006 for a more detailed description).

[Figure S1 about here.]

A simplified model of pre-synaptic vesicle recycling dynamics was adopted in order to reproduce short-time plasticity processes (both depression and facilitation) known to occur at MF – GC synapses. The model was governed by the following equations:

$$\begin{aligned}
\frac{dX}{dt} &= \frac{Z}{\tau_R} - PX\delta(t - t_{sp}) \\
\frac{dY}{dt} &= -\frac{Y}{\tau_I} + PX\delta(t - t_{sp}) \\
\frac{dZ}{dt} &= \frac{Y}{\tau_I} - \frac{Z}{\tau_R} \\
\frac{dP}{dt} &= -\frac{P}{\tau_F} + p(1 - P)\delta(t - t_{sp})
\end{aligned} \tag{7}$$

where X, Y, Z represent the fraction of neurotransmitter vesicles available, released, and recycled, respectively, and τ_R and τ_F denote the recovery and the facilitation time constants, respectively. Release probability is indicated by P, whose initial value is p.

The pre- and post-synaptic dynamics (Eqs. 6,7) are coupled through the following equations describing the neurotransmitter concentration T:

$$\begin{aligned}
T &= P([GLUT]_P + [GLUT]_D) = P(\delta(t - t_{sp})) + \frac{M}{h4\pi D_{eff}(t - t_{sp})} e^{-\frac{r^2}{4D_{eff}(t - t_{sp})}} \\
S(T) &= \frac{T^2}{(T + K_B)^2}
\end{aligned} \tag{8}$$

Neurotransmitter concentration T is obtained by summing a brief pulse ($[GLUT]_P$ of duration 0.3 ms *and amplitude* 1 mM), corresponding to direct release, to a diffusive term ($[GLUT]_D$) which accounts for glutamate spilling from distal releasing sites. In Eq. 8 M corresponds to the number of neurotransmitter molecules released, r is the distance from which it diffuses, D_{eff} is the diffusion coefficient for glutamate (it also accounts for glomerulus tortuosity), h is the size of the synaptic cleft and t_{sp} corresponds to the time at

which the diffusion of neurotransmitter occurs. In the $S(T)$ equation, the constant K_B corresponds to the affinity to glutamate. Glutamate concentration, calculated using Eq. 8, permits to calculate the AMPA and NMDA open state through the kinetic schemes shown in Figure S1A.

Model plausibility

The control analysis presented below aimed at assessing to what extent the model response variability reflected the variability observed experimentally. We compared the PSTHs of real and model GC responses when stimulating the system with the limited MF input set (i.e. 16 distinct stimuli, each constituted of four identical spike trains). To do so, we selected the longest recording we had (cell A5321), which provided us with 50 trials in both CTRL and LTP conditions. A statistical *t*-test was performed to compare model and experimental PSTHs (relative to the same stimulus, a binsize of 6 ms was adopted for the MI measurements) under both CTRL and LTP conditions. The test confirmed that almost all experimental and model PSTHs did not differ significantly (P_{VAL}>0.1, 14 out of 15 in CTRL, 15 out of 15 in LTP). We also computed the cross-correlograms of experimental PSTHs against model PSTHs for each stimulus. Then, we measured the absolute correlation peak (C_{MAX}) and its time lag $C_{MAX\text{-}LAG}$. For a given stimulus, a good match between experimental and model PSTHs would result in high cross-correlation values as well as in small values of the time lag. Figure S1B presents the result of this cross-correlogram analysis. It is shown that the matching between experimental and model PSTHs tended to increase after LTP induction, as indicated by the smaller $C_{MAX\text{-}LAG}$ values. This is in agreement with known effects of LTP induction, which not only increases the probability of eliciting GC responses to MF stimulation but also improves the time-locking of GC spikes with respect to the stimulus (Nieus et al., 2006).

S1.2 Patch-clamp experiments

In this section we report the fundamental properties of granule cell (GC) patch-clamp recordings. Patch-clamp recordings in acute cerebellar slices were performed as previously reported (Armano et al., 2000; D'Angelo et al., 1995; D'Angelo et al., 1999). Slices were cut in the sagittal plane from the cerebellar vermis of 18-to-23 day-old Wistar rats. Recordings were performed at room temperature (19-21 °C). Krebs solution for slice cutting and recovery contained (mM): NaCl 120, KCl 2, $MgSO_4$ 1.2, $NaHCO_3$ 26, KH_2PO_4 1.2, $CaCl_2$ 2, glucose 11; it was equilibrated with 95% O_2 and 5% CO_2 (pH 7.4). During recordings, the GABA-A receptor blocker, 10 μM bicuculline (Sigma), was added to the solution. The patch-clamp pipette solution contained (mM): K-gluconate 126, NaCl 4, $MgSO_4$ 1, $CaCl_2$ 0.05, BAPTA 0.1, glucose 15, ATP-Mg 3, GTP 0.1, HEPES 5. This solution maintained resting free [Ca^{2+}] at 100 nM and pH was adjusted to 7.2 with KOH. Patch-clamp pipettes filled with this solution had a resistance of 5-8 MΩ before seal formation.

All electrophysiological recordings were performed with an Axopatch 200-B amplifier and signals were sampled with a Labmaster 1200-B interface (sampling rate = 10 kHz). Current and voltage traces were digitally filtered at 5 kHz and analyzed off-line with P-Clamp (Axon Instruments), Igor (Wavemetrics), and NEURON software. Mossy fibers were stimulated with a bipolar tungsten electrode via an isolation unit. Most of the recordings were carried out in current-clamp, except for 2 sets of tracings taken in voltage-clamp at the beginning and the end of the experiment (see below). The number of active mossy fibers was determined as explained previously (D'Angelo et al., 1995; Sola et al., 2004).

In current-clamp, the GCs were maintained between –70 mV and –60 mV, and were then stimulated to generate synaptic responses. The 16 different stimulus patterns,

constructed by setting to either 1 or 0 the values in a 4-digit string, were applied randomly every 3 seconds. Since the bins lasted 10 ms, the maximum internal frequency of each pattern was 100 Hz. Each series of patterns was repeated 25 times. Each pattern was preceded by a current step to monitor intrinsic excitability (see Figure S2A). Monoexponential fitting of the voltage responses allowed us to monitor the apparent input resistance and the membrane time constant, and hence the input capacitance, $C_{in}=\tau_{in}/R_{in}$ (Figure S2B). A recording was considered stable if these parameters did not change by more than 10% during the experiment (Armano et al., 2000). LTP was induced by 8 bursts of 10 impulses at 100 Hz, which were repeated every 250 milliseconds (theta-burst stimulation, TBS). The efficiency of induction was monitored by verifying that TBS was able to elicit action potential bursts (Figure S2C, inset) (Armano et al., 2000). LTP appeared as an increase of the probability of spike generation evaluated from the response to the first stimulus in each pattern used to calculate MI (Figure S2C).

[Figure S2 about here.]

In voltage-clamp recordings at –70 mV, 30 regular spike trains consisting of 4 pulses were delivered at 100 Hz every 3 sec. The trains were averaged off-line and fitted following the procedure reported in Nieus et al. (2006) (Figure S3), which permitted to measure the release probability used to construct the diagram of Figure 2B. The previous deterministic version of the model (Nieus et al., 2006) was used for this procedure. In voltage-clamp recordings a second assessment of recording stability was performed. The cerebellar GC has a compact structure and behaves like a lumped electrotonic compartment (D'Angelo et al., 1995; Sola et al., 2004). The analysis of the passive current transients induced by 10 mV hyperpolarizing voltage steps from the holding potential of -70 mV (low-pass filter = 5 kHz, sampling rate = 100 kHz) yielded an input capacitance C_{in}= 4.64±0.12 pF, a membrane input resistance R_{in}=1.35±0.16 GΩ and a

series resistance R_s=13.51±1.42 MΩ. These parameters did not vary significantly during the experiments, demonstrating the recording stability.

[Figure S3 about here.]

S1.3 Information theoretical analysis

In the following we describe the methods we used to estimate the *mutual information* (MI) and the *surprise* between the input and output spike trains. As explained in the main text, the most relevant issue consisted of finding a reliable estimator for $p(r \mid s)$, i.e. the conditional probability distribution relating the responses of the cell r to the incoming stimuli s. Section S1.3.1 introduces the main features of information transmission measures. Section S1.3.2 describes the analysis we performed for the estimation of the sampling bias (direct method with a second order extrapolation). This procedure, along with the high reliability of the granule cell (GC) system, allowed us to evaluate mutual information and surprise with a limited number of trials.

S1.3.1 Mutual information and surprise

Shannon mutual information (MI) (Cover and Thomas, 1991; Shannon, 1948; Shannon and Weaver, 1949) provides a measure of *how much* information is contained in the neural spike patterns. In practice, in either an experiment or a simulation, we chose a set of stimuli S, and we recorded the elicited neural responses $r \in R$ when one stimulus $s \in S$ was repeatedly presented with a known prior probability $p(s)$. For example r can be defined as the number of spikes recorded within a fixed time window after the stimulus presentation, or the interspike intervals, or a binary word representation of the spike train. Once we had collected all the data, we could estimate the corresponding joint probability,

$p(r,s)$, and the probability distribution of the responses averaged over the stimuli, $p(r)$. Then, we could compute the mutual information according to:

$$MI(R;S) = \sum_{s \in S} \sum_{r \in R} p(r,s) \log_2 \frac{p(r,s)}{p(r)p(s)} \tag{9}$$

Mutual information summarizes how much we can tell about the stimuli by looking at the neural responses (or vice-versa). For example, in the trivial case in which stimulus and response are completely uncorrelated, we have $p(r,s) = p(r)\ p(s)$ and $MI(R;S) = 0$.

Eq. 9 may be rewritten in terms of conditional probability using Bayes rule ($p(r \mid s) = p(r,s)/p(s)$):

$$MI(R;S) = \sum_{s \in S} p(s) \sum_{r \in R} p(r \mid s) \log_2 \frac{p(r \mid s)}{p(r)} \tag{10}$$

or, introducing the entropy $H(R) = -\sum_{r \in R} p(r) \log_2 p(r)$ of the response probability distribution, we may rewrite Eq. 10 as:

$$MI(R;S) = H(R) - H(R \mid S) \tag{11}$$

where $H(R \mid S) = \sum_{s \in S} p(s) H(R \mid s)$ is the conditional entropy. This measure quantifies the average variability of the responses given a stimulus. If there is a deterministic mapping from stimuli to responses, then the conditional entropy is zero, and mutual information is maximal and coincides with the entropy of the responses.

In many cases it may be interesting to assess stimulus specific contribution to information transfer (Machens et al., 2005). To this extent we used the stimulus specific surprise (I_1), which is defined as:

$$I_1(s) = \sum_r p(r \mid s) \log_2 \frac{p(r \mid s)}{p(r)} \tag{12}$$

This measure tells us how much the conditional probability distribution $p(r \mid s)$ differs from an *a priori* distribution $p(r)$, hence the term "surprise". Other definitions of event specific neural coding have been proposed (Bezzi et al., 2002; Butts, 2003; DeWeese and Meister, 1999). There are various advantages in choosing the stimulus specific surprise. First, it is a simple definition that can be immediately inferred from the mutual information definition (Eq. 10) by simply taking the stimulus-dependent term of the sum. Second, it has been shown (DeWeese and Meister, 1999) that it is the only possible definition that always assumes non-negative values. Third, in the limit of very unlikely stimuli (i.e. $p(s) << 1$ for each stimulus s, as in the present study), it may be interpreted as the mutual information between the response set and a stimulus set composed by two clusters: one constituted by a single stimulus s and the other by all the rest of the stimuli (Bezzi et al., 2002).

The main issue when calculating information quantities is to estimate the conditional probabilities $p(r \mid s)$. Because the stimulus and especially the response space are in general of high dimensions, this procedure needs a very large amount of samples, rarely available in neurophysiological experiments or even in simulations. Different techniques have been proposed to overcome this problem and they are discussed in the next section.

S1.3.2 Estimation of the mutual information between spike trains

Estimating information from empirical distributions is a hard task. The main difficulty is to derive a reliable estimator for $p(r \mid s)$, given that $p(s)$ is assumed *a priori*, and $p(r)$ can be calculated from the conditional probability according to $p(r) = \sum_{s \in S} p(s) p(r \mid s)$.

The naïve procedure consists of replacing $p(r \mid s)$ by the frequency of occurrence of the corresponding response for each stimulus (direct method). This corresponds to the maximum likelihood estimator, but it faces the problem of the huge number of samples needed due the high dimensionality of the input and output spike train representations. To preserve the temporal information encoded in the spike train, we must keep track of the times of occurrence of the spiking events with a certain degree of accuracy. For example, by choosing a temporal resolution of 6 ms (as in the main text) we could represent spike trains without any loss of information up to 160 Hz; by considering a 100 ms time window, this would result in 2^{16} possible different output sequences. Thus, a robust estimation of mutual information using the direct method would require a number of sample N >> 2^{16} for each stimulus. Therefore, in the presence of a large stimulus set, as in our simulation (65536 stimuli, e.g. see Figure 3 in the main text) the number of trials needed would be larger than a few billions.

Although the application of the direct method appears to be unrealistic due to these computational issues, there are two important features of the GC system that make this approach suitable for our specific case.

First of all, the degree of stochasticity of our system is low, that is the behavior of the GC is *almost* deterministic. This could be assessed quantitatively calculating the conditional (noise) entropy for any stimulus. In a typical simulation (65536 stimuli, 400 trials), the average conditional entropy contribution $H(R \mid S)$ was 1.56 bits. By increasing the number of trials up to 50 000, we obtained a similar value: 1.59 bits. The maximal

conditional entropy over the whole stimulus set was 7.1 bits. These values were much lower than the maximal entropy corresponding to a uniform distribution (16 bits). In other words, the conditional probability distribution $p(r \mid s)$ was concentrated around few significant, and well represented, responses. Accordingly, sampling this typical set gave a good approximation of the entropy of the distribution. Brautbar and Samorodnitsky (2005) introduced a parameter of a distribution called *effective alphabet size* q_{eff}, which is a function of the entropy of the distribution. They showed that a consistent maximum likelihood estimator converges to the entropy of the distribution asymptotically for N>> q_{eff},. In our case, q_{eff} = 54 for $p(r)$ and the average q_{eff} for the conditional probability distributions $p(r \mid s)$ was 104 symbols with a maximum at 216 symbols.

Secondly, by examining the behavior of mutual information as a function of the sampling size we found a smooth (decreasing) curve that could be extrapolated to the infinite data limit (see details below).

When applying the direct method, we introduce a systematic negative bias in the entropy estimation and a positive bias in the mutual information (Nemenman et al., 2004; Paninski, 2003; Strong et al., 1998; Treves and Panzeri, 1995). This bias can be evaluated analytically in some special cases (Treves and Panzeri, 1995), otherwise it has to be computed numerically. Following Strong et al. (1998) to evaluate this correction, we performed the analysis using only a fraction of the trials, then we plotted the information values over the inverse data fraction (*1/N*, where *N* is the number of trials), and finally we extrapolated the values to the infinite data set limit, as *1/N* goes to zero, with a second order extrapolation method. All the simulations described in the main text were done with *400* trials, and the infinite size limit values for the mutual information were obtained by extrapolation using the *N=100-, 200-, 300-* and *400*-trial values.

To estimate the error, we performed the same analysis using a larger dataset (*N=50000* trials), and compared with the extrapolated values of the mutual information and the surprise obtained with the different sampling sizes. Figure S4 compares the MI estimate (for a release probability *p=0.5)* computed by extrapolating from a subset of samples (*N=100, 200, 300* and *400* trials) to the infinite limit number of trials, with the MI numerical results obtained with *N=50000* trials (black arrow). The left (A) and right (B) panels refer to MI calculation using spike counts and binary words, respectively. The estimated errors (the difference between the second order extrapolation and the numerical value obtained with 50000 trial simulations) were only 0.3% for spike counts and 1% for binary words. The extra data points computed for this particular case (p=0.5) for *N* between *400* and *50000* shown in the figure are all near the extrapolated line. Together, these results show that the bias estimation by extrapolating from the subset of samples (*N=100, 200, 300* and *400 trials*) was effective.

Applying the same method to estimate the error in the surprise computation, we found an average error of 10% for spike count and 10% for binary word.

[Figure S4 about here.]

S1.3.3 Conclusions

The main problem of measuring information transmission with spiking neurons is to estimate the conditional probability density of binary words from a limited set of trials. We replaced these distributions with the corresponding frequency histograms obtained by numerical simulations, and we estimated the limited sampling bias by means of a second order extrapolation. This procedure, together with the low stochasticity that characterizes GC processing, allowed us to successfully evaluate the mutual information and the surprise by using a limited number of trials.

S2 Supplementary Results

S2.1 Dependence of the results upon the temporal bin-size used to sample granule cell responses

The 6 ms bin-size used to sample granule cell (GC) responses was empirically set to characterize the output state space without discarding any relevant spike-timing information. The control analysis reported here was performed to assess to what extent our findings might be influenced by the binning procedure.

First, we quantified the fraction of omitted spikes (*FOS*), i.e. the fraction of action potentials disregarded during the information theoretic measurements due the 6 ms bin-width (multiple action potentials might eventually occur within the same temporal bin). Figure S5 displays the outcome of this analysis. We computed $\langle FOS \rangle$ by averaging over all the different release probabilities $\overline{p}$ (range [0.1-0.8], with p varying independently across the four mossy fibers, MFs), and over all 65536 x 400 stimulus presentations (65536 stimuli each presented 400 times). We found that $\langle FOS \rangle$=2.34 spikes per thousand were omitted (std=2 spikes per thousand). The maximum fraction of lost spikes $FOS_{\max}$ was equal, on average, to 6.47 spikes per thousand and it occurred (as expected) in the presence of the largest release probability (i.e., $\overline{p}$=0.8) on the four MFs. This control analysis suggested that a bin-width of 6 ms permitted to recover the majority of the information transmitted by the GC.

[Figure S5 about here.]

To further investigate this issue we focused on the results reported on Figure 8, which aimed at identifying the stimulus features that produced the largest and the sharpest changes of surprise-per-spike following release probability increases. A new series of

analyses was carried out to assess whether or not the findings of Figure 8 were likely to depend on the width of temporal bin. Figures S6-S8 compare the results obtained with a bin-width of 3 ms against those obtained with a bin-width of 6 ms. This qualitative analysis suggested that reducing (by a factor of 2) the bin-size does not effect our results about the type of stimuli that benefited the most (i.e., short correlated stimuli) or the least (i.e., long correlated stimuli) from release probability increases. In order to provide a more quantitative assessment, we compared the correlation coefficient C (see Materials and Methods) of the stimuli providing the largest increases and decreases of surprise-per-spike with increasing releasing probability in the case of 6 and 3 ms bin-sizes. The diagram of Figure S9 shows a good consistency between the findings under the two conditions.

[Figure S6 about here.]

[Figure S7 about here.]

[Figure S8 about here.]

[Figure S9 about here.]

S2.2 Relative role of AMPA receptor desensitization *vs* pre-synaptic vesicle turnover in determining the transmission properties of granule cells before and after LTP induction

Short-term plasticity (STP) at mossy fiber – granule cell (MF – GC) synapses is likely to be mediated by both post-synaptic AMPA receptor desensitization (Saviane and Silver, 2006) and pre-synaptic vesicle dynamics (Nieus et al., 2006). In order to ascertain the relative role of AMPA desensitization *vs* vesicle turnover, a series of numerical simulations was run in which the recovery time constant (τ_R) and the AMPA desensitization (DES) process were modified in turn. Interestingly, setting either τ_R or DES to zero did not cause any significant differences in GC spike timing, because action

potentials tended to fell within the same time bins with respect to the CTRL situation (Figure S10; notice that in order to elucidate the effect of setting either τ_R or DES to zero, the strongest stimulus, i.e. four pulses at 100Hz on the four MFs, was taken). As a consequence, no significant MI differences were revealed in either testing conditions.

[Figure S10 about here.]

S2.3 Dependence of the results upon the probability distribution underlying MF activity patterns

As mentioned in the Material and Methods section, the actual distribution underlying MF firing activity under natural conditions remains unknown. It has been shown that GCs do not fire spontaneously but they rather require the co-activation of 3 to 4 MFs to elicit a spike (D'Angelo et al., 1995, Nieus et al., 2006). It has also known, since the 80's (Kase et al., 1980), that MFs discharge high-frequency bursts (>100Hz) during eyelid conditioning experiments. Recent *in vivo* studies have confirmed that such high-frequency bursts (average ISI 12.1ms, Rancz et al., 2007) are also generated in response to mechanical stimulation of rat's whiskers (Rancz et al., 2007). In both the aforementioned experimental investigations the high frequency bursts are a footprint of sensorial input reaching the cerebellum. At rest, when no specific sensorial input reaches the cerebellum, the MF background activity is almost quiescent (3.9 Hz, Rancz et al., 2007). These findings point towards the hypothesis of a non-uniform distribution *P(s)* describing MF activity patterns and suggest that a low probability is likely to be associated with high frequency inputs. Similar to the theoretical approach put forth by Treves et al. (1999), we assumed that the probability distribution *P(s)* associated to MF inputs would follow a general trend: to be continuous, unimodal (with a single peak close to the spontaneous activity, i.e. close to zero in our case), and with an exponential tail.

The findings presented in the main part of the paper were obtained by assigning to each MF input train *s* a *P(s)* drawn from a decreasing exponential distribution with a mean firing rate of 10Hz. Here we present the results of a complementary series of simulations aimed at testing the effect of using a different probability distribution, namely a Poisson distribution *P(s)* with a mean firing rate of 10Hz. Figures S11-S14 show that the findings obtained with the Poissonian *P(s)* are consistent with those obtained with the exponential *P(s)*.

[Figure S11 about here.]

[Figure S12 about here.]

[Figure S13 about here.]

[Figure S14 about here.]

S2.4 Contribution of subsets of stimuli to information transmission

When focusing on MI, one may predict a monotonic-like relationship between the synaptic efficacy of MF – GC projections and the average amount of information transmitted by the GC. Indeed, LTP is known to increase not only the post-synaptic firing but also the time locking of the output spikes relative to the afferent signals. On the other hand, when looking at more stimulus-specific measures (such as the surprise) the way LTP shapes neuronal information transmission becomes less predictable. Thus, questions like “how does the LTP-dependent shaping process account for the spatio-temporal structure of specific stimuli?” can be asked. Figure 4 helped us to begin answering this question and suggested that only the surprise of the least informative stimuli (i.e., those having a surprise value smaller than 5% of the maximum surprise) showed a linear relationship with the release probability $\overline{p}$. By contrast, optimal information transmission for most inputs occurred already at intermediate values of release probability. Figure S15A is an extended version of Figure 4: it shows the relationship between the surprise values and

$\overline{p}$ for all the subsets of stimuli considered for this analysis. Data points were obtained by averaging the surprise over distinct subsets of stimuli, from the most informative ones (i.e., those with surprise larger than 90% of the maximum) to the least informative ones (i.e., those having a surprise less than 5% of the maximum). Note that the size of the input subsets used to calculate the mean surprise values varied across data points.

We quantified the contribution of each mean surprise data point to the overall MI. That is, let S_i be the subset of stimuli considered to generate a given data point i of Figure S15A. We computed $MI(R;S_i) = \sum_{s \in S_i} p(s) \cdot I_1(R;s)$, where $p(s)$ denotes the prior probability of a stimulus s (drawn from an exponential distribution of firing rates, see Material and Methods as well as Supporting Information), and $I_1(R;s)$ is the surprise associated to stimulus s. Figure S15B displays how the $MI(R;S_i)$ contributions are distributed over the sampled data points. It is shown that the contribution of least informative stimuli becomes significantly dominant as the release probability increases, which is consistent with the findings of Figure 3B suggesting a monotonic behavior of MI following LTP (for a given release probability, mutual information is given by $MI(R;S) = \sum_i MI(R;S_i)$).

[Figure S15 about here.]

[Figure S16 about here.]

[Figure S17 about here.]

S2.5 Considerations concerning the transmission of specific input examples

S2.5.1 Output firing rate and information transmission

The output firing frequency constitutes a relevant facet of the manifold transmission process occurring at the MF-GC synaptic relay. Indeed, it is shown (main body, Sec. 2.1)

that the GC firing rate conveyed about 50% of the overall amount of information transmitted. Yet, the rest of the paper focuses on the contribution to information transfer of the spatiotemporal structure of the input-output patterns. Indeed, the spike count alone cannot be employed to understand the results provided by the MI, Surprise, and Surprise-per-spike measures, which are based on the binary-string decoding scheme. For example, at release probability $\overline{p} = 0.8$, the GC responses to the stimuli $s_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$ and $s_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ are characterized by equivalent mean firing rates and standard deviations (Figure S18A). By contrast, the surprise and surprise-per-spike measures (Figures S18B,C) capture the significant information-wise difference between the two stimuli, which reflects the different spatiotemporal characteristics of the input and output patterns displayed on Figures S18D,E.

[Figure S18 about here.]

S2.5.2 Transmission of stimuli having spikes in adjacent *vs.* non-adjacent time bins

Let us consider the two following stimuli:

$s_1 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$, with $C(s_1) = 0.4039$, and $s_2 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$, with $C(s_2) = 0.3048$.

Figure S19 displays the spike count of the GC response to s_1 and s_2 as a function of the mean release probability $\overline{p}$. It is shown that, on average, the two stimuli elicit a different number of spikes at low (e.g. $\overline{p} = 0.3$) and high (e.g. $\overline{p} = 0.8$) release probabilities. This

GC behaviour is likely to depend, among other factors, on short-term synaptic facilitation and depression mechanisms. Figure S20 presents an information-wise comparison of the encoding of the two stimuli at different $\overline{p}$. A coherent picture is provided by the surprise and surprise-per-spike measures –see A-B for absolute values, see C-D for rankings. The transmission of s_1 benefits more rapidly from $\overline{p}$ increases, whereas s_2 is better transmitted at high release probabilities ($\overline{p} > 0.6$). Finally, Figure S20E displays the shape of the distribution of the GC response patterns at $\overline{p} = 0.3$ and $\overline{p} = 0.8$. The PSTHs at low $\overline{p}$ show that the response distribution to s_1 is sharper –and better time locked– than the response to s_2. At high $\overline{p}$ the occurrence of doublets creates interferences in the GC response to the second spike of s_1 –which spreads the response probability distribution and results in the decrease or saturation of information transmission. By contrast, the occurrence of doublets does not create any interference with the subsequent spike production in the case of s_2 and the surprise and surprise-per-spike values continue to grow at high $\overline{p}$.

[Figure S19 about here.]

[Figure S20 about here.]

Other two examples of transmission of specific stimuli having spikes in adjacent input bins are considered in Figure S21. First, the surprise-per-spike and its ranking (Figure S21 A and B, respectively) are measured over $\overline{p}$ for the three following stimuli:

$s_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$, $s_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, and $s_3 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. As expected according to our previous findings, the short-and-correlated stimulus s_1 is the best transmitted, compared to s_2 and s_3, as the release probability increases. By contrast, the stimulus s_3, which is likely to induce monosynaptic temporal summation and short-term

facilitation/depression mechanisms, benefits the least from $\bar{p}$ increases. Figures S21B-C compare the surprise-per-spike and its ranking, respectively, of the following stimuli:

$s_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$, $s_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$, and $s_3 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$. Consistent with the results of the previous example, the stimulus s_3 is the one which gains the least, in terms of information transmitted, as the release probability augments.

In the last two examples, the stimulus s_1 was compared first to inputs with the same number of spikes but lower correlation, and then to stimuli with larger spike counts and higher correlation. The results of these comparisons (Figure S21) suggest that the stimuli that benefit the most, in terms of surprise, from release probability increases are the ones that set the best compromise between correlation and spike count, which is consistent with the general finding in the main body concerning the relevance of short correlated stimuli.

[Figure S21 about here.]

S2.5.3 Discussion

Correlation and surprise/spike measure

Figures 5 & 6 present important information regarding the relationships between input correlation and cell response. The surprise-per-spike and its ranking provides a color-coded classification scheme, which illustrates the archetype of the “structure” of the stimuli falling in a category. It indicates also how it relates to the correlation of the stimuli. It is clear from Figure 6C that correlation is only part of the story. Indeed, for the same correlation C= 0.2, one can find almost all categories of stimuli since almost all colors are

represented along this value. The surprise-per-spike differentiates between them along another dimension.

Intuitive presentation of surprise measures

The surprise describes whether the response associated with a particular stimulus was expected or not when compared to our prior knowledge, hence the surprise terminology. Since probabilities are used, the results for one stimulus are always in relation to all other stimuli. Once normalized, the PSTHs presented in some of the figures are a good approximation to the probability distribution of the response given the stimulus, $p(r \mid s)$. A high reliable response is one that has high and sharp peak(s) in its probability distribution or PSTH: the response spikes occur at about the same time for every stimulus presentation. A low reliable response is one that has low and wide probability distribution peak(s): the timing of the response spikes varies significantly across presentations of the same stimulus.

As other neurons, the granule cell is more likely to fire after spike summation in time. Nevertheless, the analysis indicates that a response spike can also occur whenever there are 3 or 4 input spikes in the same time bin (correlated spikes) with no summation required. Intuitive analysis of spike responses can be understood as the approximation of two principal cases. As release probability $\overline{p}$ increases, when there is only one spike in the input or when the spikes are far apart, the output spike(s) occur(s) earlier and the response probability distribution increases and remains sharp, since there is no or minimal interference between output spikes. There is a monotonic increase in information transmission as $\overline{p}$ increases. With multiple input spikes close to one another, there may be interference in spike production. The interference can be the result of the dynamics of the cell, such as when doublets occur, the refractory period, etc. As $\overline{p}$ increases, the response probability distribution spreads and its peak decreases in amplitude. This results

in a decrease or saturation of information transmission at large $\bar{p}$. Most spike responses can be understood in terms of particular combinations of the two basic cases above, which we will refer to as case #1 and #2.

In the following, we examine closely three examples of spike priming. The first is given in Figure 8A. Neither of the two sets of correlated spikes alone produces a spike response at low probability of release $\bar{p}$ (see spikegram for $\bar{p} = 0.2$). As $\bar{p}$ increases beyond 0.4, the ranking increases as the responses become more and more "unexpected" (or surprising). Two coincident spikes are rarely enough to fire the cell. Here, the first set of two spikes prime the cell so that the next set can make it fire reliably at $\bar{p} = 0.4$. Notice that there are some responses to the first set of two spikes at $\bar{p} = 0.4$, but these are not reliable as their PSTH is very low and wide. The reliability of the response is not high relative to other stimuli, that is why the surprise-per-spike ranking is still very low. As $\bar{p}$ increase above 0.4, the overall reliability of the cell response increases (i.e. the reliability of the responses to both sets of spikes) and the response becomes more and more unexpected. The timing is about just right. The first set of two spikes at high $\bar{p}$ leads to case #1 above: the response does not suffer from interference. Moreover, the distance between the next set of spikes (2 time bins) is about just right to minimize interference from the 1st spike production. This suggests that this 2nd set of spikes also acts like a case #1 above. Nevertheless, one can see in the PSTH at $\bar{p} = 0.7$ that the response to the 2nd set of spikes starts to spread after the high peak. Since this tail was not present at lower $\bar{p}$, this suggests a starting interference from the previous response spike, which becomes stronger as $\bar{p}$ increases (see the difference in peak between $\bar{p} = 0.4$ and $\bar{p} = 0.7$). This interference is nonetheless not sufficient to modify the high ranking in the surprise-per-spike for this stimulus (left diagram).

In conclusion, this is an example where priming of spikes occurs to permit appearance of later spikes at low $\bar{p}$. At high $\bar{p}$, the two sets of spikes are well separated in time to permit both to act like a first set of spikes and do not produce much interference. This means that the overall response of the cell can remain reliable at high $\bar{p}$ with two sharp peaks. This is the reason of the highest change in rank of surprise-per-spike: practically no response at low $\bar{p}$, and very high and reliable response at high $\bar{p}$.

The second example of priming is provided in Figure 8C. The spikegram and PSTH at $\bar{p} = 0.8$ indicate that the sharpest peak occurs for the 2nd set of spikes, not the first. This is because there are only 2 input spikes in the first set. They do not elicit a strong response and do not produce interference with the 2nd set of spikes. Consequently, the 2nd set of spikes, with 4 correlated spikes, retains a highly reliable response at high $\bar{p}$ even if they produce doublets. This is not the case for the following spikes: they suffer from the previous presence of doublets and their response is not reliable, similarly for the 4th set of spikes. The overall ranking of this stimulus drops dramatically as a result (left diagram).

The 3rd and last example is given in Figure 8B. It shows another example of a 2nd set of spikes that behaves like case #1 above (first spikes). Here, even at high $\bar{p}$, the lonely spike does not produce much of a response, yet it primes the response to the 2nd set of spikes. The change in rank can occur very quickly: the response to the 2nd set of spikes both appears suddenly and the peak is high and sharp (very reliable response). Since there are no following spikes, the doublet does not produce any interference and the surprise-per-spike ranking remains high.

Figure 8 presents the key characteristics of the stimuli that responded to the criterion selected: highest jump, sharpest jump, and lowest jump. The analysis by surprise and surprise-per-spike permitted to uncover the descriptions above, it also provided a categorization of the stimuli in ways that were unexpected at first. A posteriori, it is

relatively easy to make sense of it all. Nevertheless, the direction in reverse is less than obvious (predicting which stimuli would fulfill the criterion). Let's recall that our interest was to uncover how information transmission changed as release probability changed, which goes beyond just doing a detailed analysis of when the cell responds.

Figure Legends

Figure S1. *Synaptic model of the mossy fiber – granule cell projection.*

(A) AMPA & NMDA kinetic schemes. S(T) is a Michaelsen Menten function (see text), D and C are deactivated and closed states, respectively. (B) Consistency between model and experimental PSTHs before and after LTP induction. The cross-correlation (measured here by the $C_{MAX\text{-}LAG}$ values, see text) between experimental and model PSTHs tended to be larger after LTP induction (red data) than in CTRL conditions (blue data).

Figure S2. *LTP Induction at the Mossy Fiber – Granule Cell Synapses. Experimental results.*

(A) Each EPSP was preceded by a -2 pA current step lasting about 150 ms. Note that, after theta-burst stimulation (TBS), the EPSP increased while the control step did not change. (B) Plots of Rin and Cin values obtained from monoexponential fitting of the passive voltage response. Notice that Rin and Cin were stable and did not change after TBS (arrow). (C) The time-course of LTP for the same cell shown in A-B. The probability of spike firing during low-frequency synaptic transmission was enhanced following TBS – in this example, a decaying PTP/STP (post-tetanic potentiation / short-term potentiation) phase is observed before full LTP develops. The inset shows the GC response to TBS, revealing strong action potential discharge (Armano et al., 2000). Two of these protocols (x 2) were delivered in close sequence to induce LTP.

Figure S3. *Fitting of EPSC trains*.

EPSC trains of 4 pulses at 100 Hz were elicited in voltage-clamp at the beginning and at the end of recordings. The traces are averaging of 30 sweeps and were fitted with the model reported previously (Nieus et al., 2006). Fittings permitted to estimate the

neurotransmission parameters and, in particular, release probability. In the example reported in this figure, release probability changed from 0.58 to 0.81, while the other parameters (including the effective glutamate diffusion constant, the time constants for synaptic facilitation and vesicle recovery, and glomerular geometry) were held constant to their control values.

Figure S4. *Finite sampling bias.*

Mutual information computed using spike counts (A) and binary words (B) is displayed as a function of the sampling size. The sampling size is shown as the inverse fraction of data used, *1/N*, where *N* is the number of samples. The mutual information (diamonds) was computed by using 65536 different stimuli, each of which presented up to 50000 times (trials), with a probability of release *p=0.5*. The dotted line is a second order fit obtained from the 4 rightmost points, *N = 100, 200, 300, 400* trials (as described in the main text). The infinite data sampling extrapolation limit, as *1/N* goes to zero, has values 0.416 (spike count) and 0.84 bits (binary word). The measured values using 50000 trials (black arrows) were close to the infinite data values: 0.415 (spike count) and 0.832 bits (binary word).

Figure S5. *Mean number of GC spikes per thousand omitted from the analysis due to the 6 ms bin-width. Simulation results.*

The number of omitted spikes per thousand was measured as a function of the mean release probability $\bar{p}$ (range [0.1-0.8], with p varying independently across the four mossy fiber synaptic terminals). Each data point was computed by averaging over the entire set of 65536 x 400 stimulus presentations (i.e., 65536 stimuli with 400 trials per stimulus).

Figure S6. ***Largest*** *surprise-per-spike* ***increases*** *following release probability increases. Comparison between results obtained with 6 and 3 ms bin-width. Simulation results.*

The 10 largest increases (in rank) of surprise-per-spike following $\bar{p}$ increases when employing a bin-size equal to 6 ms (A) and 3 ms (B).

Figure S7. ***Sharpest*** *surprise-per-spike* ***increases*** *following release probability increases. Comparison between results obtained with 6 and 3 ms bin-width. Simulation results.*

The 10 sharpest increases (in rank) of surprise-per-spike following $\bar{p}$ increases when taking a bin-size equal to 6 ms (A) and 3 ms (B).

Figure S8. ***Largest*** *surprise-per-spike* ***decreases*** *following release probability increases. Comparison between results obtained with 6 and 3 ms bin-width. Simulation results.*

The 10 largest decreases (in rank) of surprise-per-spike following $\bar{p}$ increases when taking a bin-size equal to 6 ms (A) and 3 ms (B).

Figure S9. ***Largest*** *surprise-per-spike* ***changes*** *following release probability increases. Comparison between results obtained with 6 and 3 ms bin-width. Simulation results.*

Correlation coefficient of the stimuli that provided the largest increases (red data) and the largest decreases (blue data) of surprise-per-spike following LTP.

Figure S10. *Effect of manipulating the AMPA receptor desentization process and the pre-synaptic vesicle dynamics on GC spike timing. Simulation results.*

A) Spike time delays were measured by computing t_{CTRL}-t_{TEST}, being negative when the test condition caused an advance of the output action potential. B) Effect of setting either the recovery time constant (τ_R) or the AMPA desentization (DES) to zero under two probability release conditions (*p=0.4* and *p=0.6*). In terms of spike timing, the effect of both manipulations was negligible (given the 6 ms bin-size used to sample GC responses).

Figure S11. *Release probability increases enhance mutual information in GCs. Simulation results.*

No significant difference was revealed when comparing the case of exponential (blue curve) *vs* Poisson (red curve) probability distribution underlying MF firing activity.

Figure S12. *Surprise changes following release probability increases. Simulation results.*

Comparing the case of exponential (A) *vs* Poissonian (B) probability distribution of the MF activity suggested a coherent effect of release probability increases upon the surprise values.

Figure S13. *Surprise changes as a function of both release probability and correlation across the four MF inputs. Simulation results.*

The comparison between the results obtained with an exponential (A) *vs* a Poisson (B) probability distribution of the MF firing rate suggested a consistent behavior under both assumptions. Similar to Figure 6, a color code was employed to characterize the number of bins occupied (blue, green, pink, red for 4, 3, 2, 1 bins, respectively). Within the same color (e.g., red), the number of spikes/bin modulated the color nuances (e.g. red to orange for 4 to 1 spikes/bin).

Figure S14. *Changes in surprise-per-spike as a function of both release probability and correlation across the four MF inputs. Simulation results.*

The results obtained by assuming an exponential (A) and a Poisson (B) probability distribution of the MF firing rate are consistent with each other. The color map is the same than the one used for Figure 6 and S13.

Figure S15. *Contribution of subsets of stimuli to information transmission. Simulation results.*

Average surprise of stimulus subsets as a function of mean neurotransmitter release probability $\bar{p}$. (A) For each mean release probability, data points were obtained by averaging the surprise over subsets of stimuli, from those with surprise larger than 90% of the maximum to those having a surprise less than 5% of the maximum. Optimal information transmission of most inputs occurred already at intermediate values of release probability. Only the mean surprise of stimuli with a surprise less than 5% of the maximum was monotonously increasing with $\bar{p}$ (see bottom-right diagram). (B) Distribution of the contributions to the overall MI over the sampled data points. The stimuli with lowest surprise become dominant following release probability increases.

Figure S16. *Largest increases (A) and decreases (B) of* ***surprise*** *ranking following release probability increases. Simulation results.*

Figure S17. *Characterization in terms of spike count of the stimuli that provided the largest* ***increases*** *(red data) and the largest* ***decreases*** *(blue data) of surprise following LTP. Simulation results.*

Figure S18. *Example showing that the output spike count alone cannot explain the results about most informative stimuli. Simulation results.*

At release probability $\bar{p} = 0.8$, the two MF stimuli s_1 and s_2 elicit, on average, the same number of GC spikes although their surprise and surprise-per-spike computed based on the binary string decoding scheme are significantly different. The two stimuli have correlation coefficients equal to $C(s_1) = 0.2548$ and $C(s_1) = 0.0647$, respectively.

Figure S19. *Example showing the output spike count as a function of release probability for two specific stimuli. Simulation results.*

Figure S20. *Surprise and surprise-per-spike as a function of release probability for the same stimuli used in the example of Figure S19. Simulation results.*

Figure S21. *Two examples of information-wise comparison for specific stimuli having spikes in adjacent input bins. Simulation results.*

Table S1. *Equations and parameter settings determining the dynamics of the ionic channels of the granule cell (GC) model.*

Table S2. *Kinetic rates of AMPA and NMDA postsynaptic receptors of the model.* Dimension units are ms^{-1}, except for the star-labeled ones that are ms^{-1} mM^{-1}.

References

Armano, S., Rossi, P., Taglietti, V., and D'Angelo, E. (2000). Long-term potentiation of intrinsic excitability at the mossy fiber-granule cell synapse of rat cerebellum. J Neurosci 20, 5208-5216.

Bezzi, M., Samengo, I., Leutgeb, S., and Mizumori, S.J. (2002). Measuring information spatial densities. Neural Comput 14, 405-420.

Brautbar, M., and Samorodnitsky, A. (2005). Approximating the Entropy of Large Alphabets. In Electronic Colloquium on Computational Complexity.

Butts, D.A. (2003). How much information is associated with a particular stimulus? Network 14, 177-187.

Cover, T.M., and Thomas, J.A. (1991). Elements of information theory (Wiley).

D'Angelo, E., De Filippi, G., Rossi, P., and Taglietti, V. (1995). Synaptic excitation of individual rat cerebellar granule cells in situ: evidence for the role of NMDA receptors. J Physiol 484 (Pt 2), 397-413.

D'Angelo, E., Nieus, T., Maffei, A., Armano, S., Rossi, P., Taglietti, V., Fontana, A., and Naldi, G. (2001). Theta-frequency bursting and resonance in cerebellar granule cells: experimental evidence and modeling of a slow k+-dependent mechanism. J Neurosci 21, 759-770.

D'Angelo, E., Rossi, P., Armano, S., and Taglietti, V. (1999). Evidence for NMDA and mGlu receptor-dependent long-term potentiation of mossy fiber-granule cell transmission in rat cerebellum. J Neurophysiol 81, 277-287.

DeWeese, M.R., and Meister, M. (1999). How to measure the information gained from one symbol. Network 10, 325-340.

Kase, M., Miller, D.C., and Noda, H. (1980). Discharges of Purkinje cells and mossy fibres in the cerebellar vermis of the monkey during saccadic eye movements and fixation. J Physiol 300, 539-555.

Machens, C.K., Gollisch, T., Kolesnikova, O., and Herz, A.V. (2005). Testing the efficiency of sensory coding with optimal stimulus ensembles. Neuron 47, 447-456.

Nemenman, I., Bialek, W., and van Steveninck, R.R.R. (2004). Entropy and information in neural spike trains: Progress on the sampling problem. Phys Rev E 69, 056111.

Nieus, T., Sola, E., Mapelli, J., Saftenku, E., Rossi, P., and D'Angelo, E. (2006). LTP regulates burst initiation and frequency at mossy fiber - granule cell synapses of rat cerebellum: experimental observations and theoretical predictions. J Neurophysiol 95, 686-699.

Paninski, L. (2003). Estimation of entropy and mutual information. Neural Comput 15, 1191-1254.

Rancz, E.A., Ishikawa, T., Duguid, I., Chadderton, P., Mahon, S., and Hausser, M. (2007). High-fidelity transmission of sensory information by single cerebellar mossy fibre boutons. Nature 450, 1245-1248.

Saviane, C., and Silver, R.A. (2006). Fast vesicle reloading and a large pool sustain high bandwidth transmission at a central synapse. Nature 439, 983-987.

Shannon, C.E. (1948). The mathematical theory of communication. Bell Syst Tech J 27, 379-423.

Shannon, C.E., and Weaver, W. (1949). The mathematical theory of communication (University of Illinois Press).

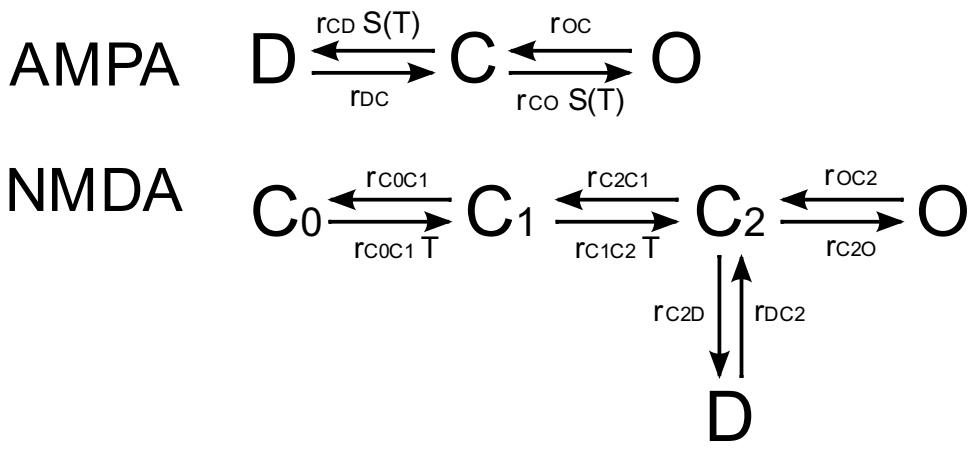
Sola, E., Prestori, F., Rossi, P., Taglietti, V., and D'Angelo, E. (2004). Increased neurotransmitter release during long-term potentiation at mossy fibre-granule cell synapses in rat cerebellum. J Physiol 557, 843-861.

Strong, S.P., Koberle, R., de Ruyter van Steveninck, R.R., and Bialek, W. (1998). Entropy and information in neural spike trains. Phys Rev Lett 80(1), 197-200.

Treves, A., and Panzeri, S. (1995). The upward bias in measures of information derived from limited data samples. Neural Comput 7, 399-407.

Treves, A., Panzeri, S., Rolls, E.T., Booth, M., and Wakeman, E.A. (1999). Firing rate distributions and efficiency of information transmission of inferior temporal cortex neurons to natural visual stimuli. Neural Comput 11, 601-632.

A


AMPA
D
C
O
rCD S(T)
rDC
rOC
rCO S(T)
NMDA
C0
C1
C2
O
D
rC0C1
rC0C1 T
rC2C1
rC1C2 T
rOC2
rC2O
rC2D
rDC2

B

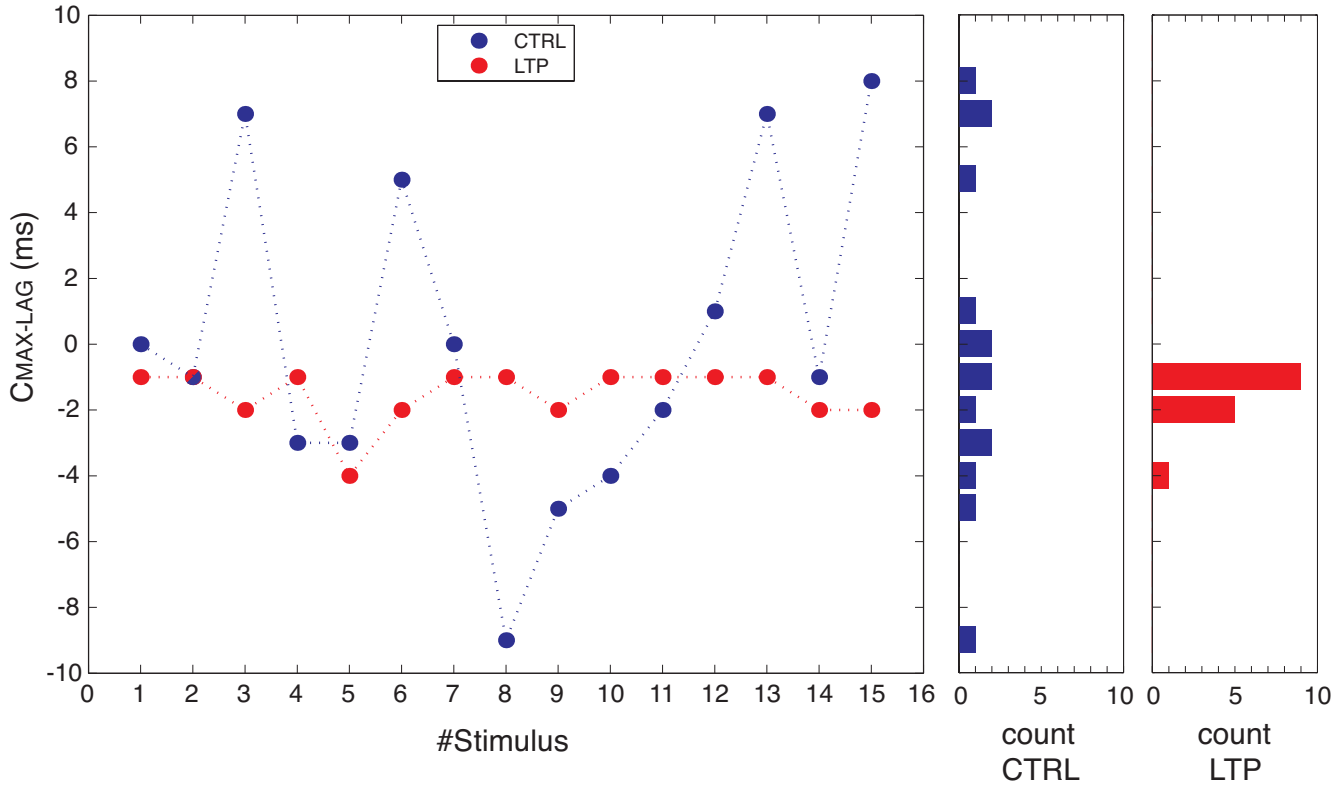

CTRL
LTP
CMAX-LAG (ms)
#Stimulus
count
CTRL
count
LTP

Figure S1

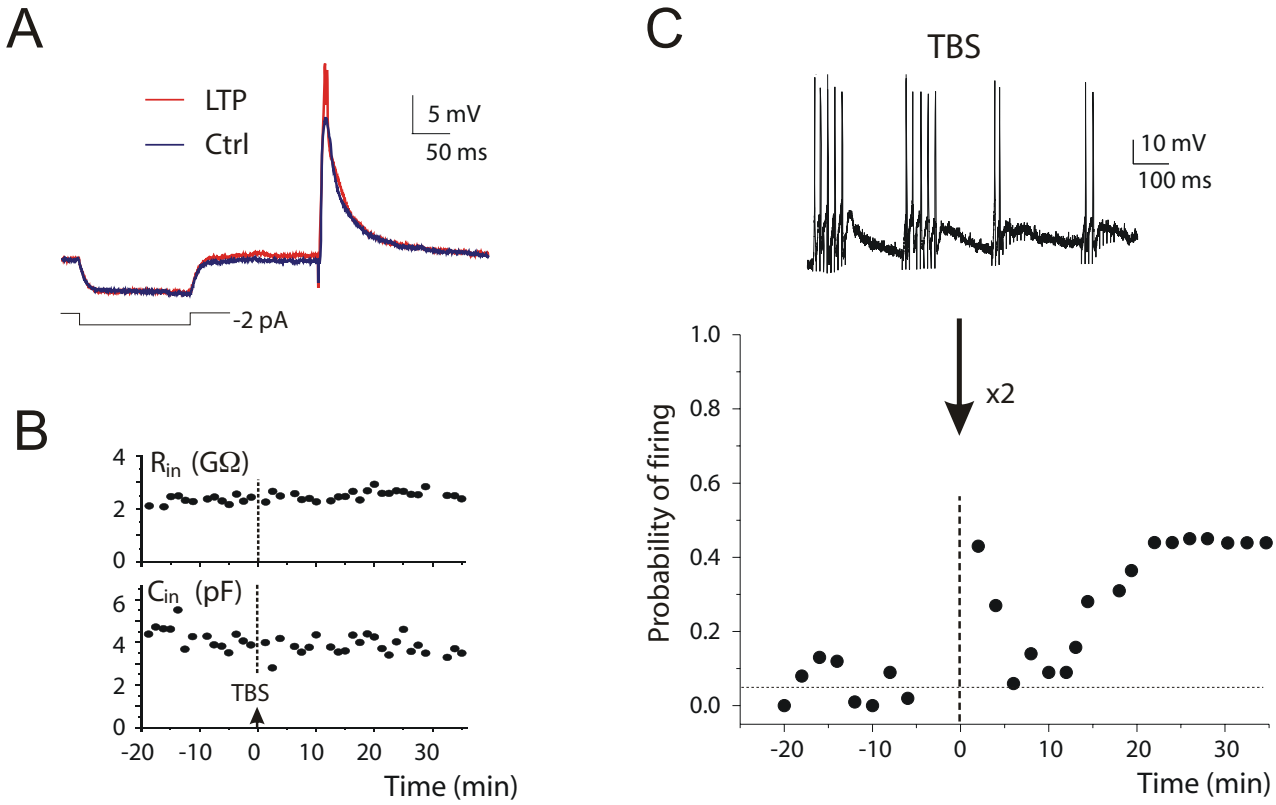

A
LTP
Ctrl
5 mV
50 ms
-2 pA
B
R_{in} (GΩ)
C_{in} (pF)
TBS
Time (min)
C
TBS
10 mV
100 ms
x2
Probability of firing
Time (min)

Figure S2

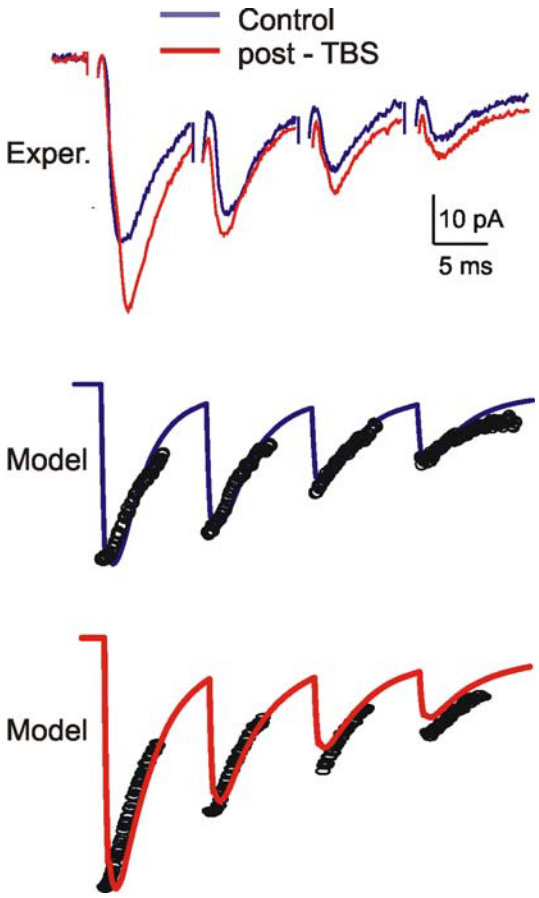


Figure S3

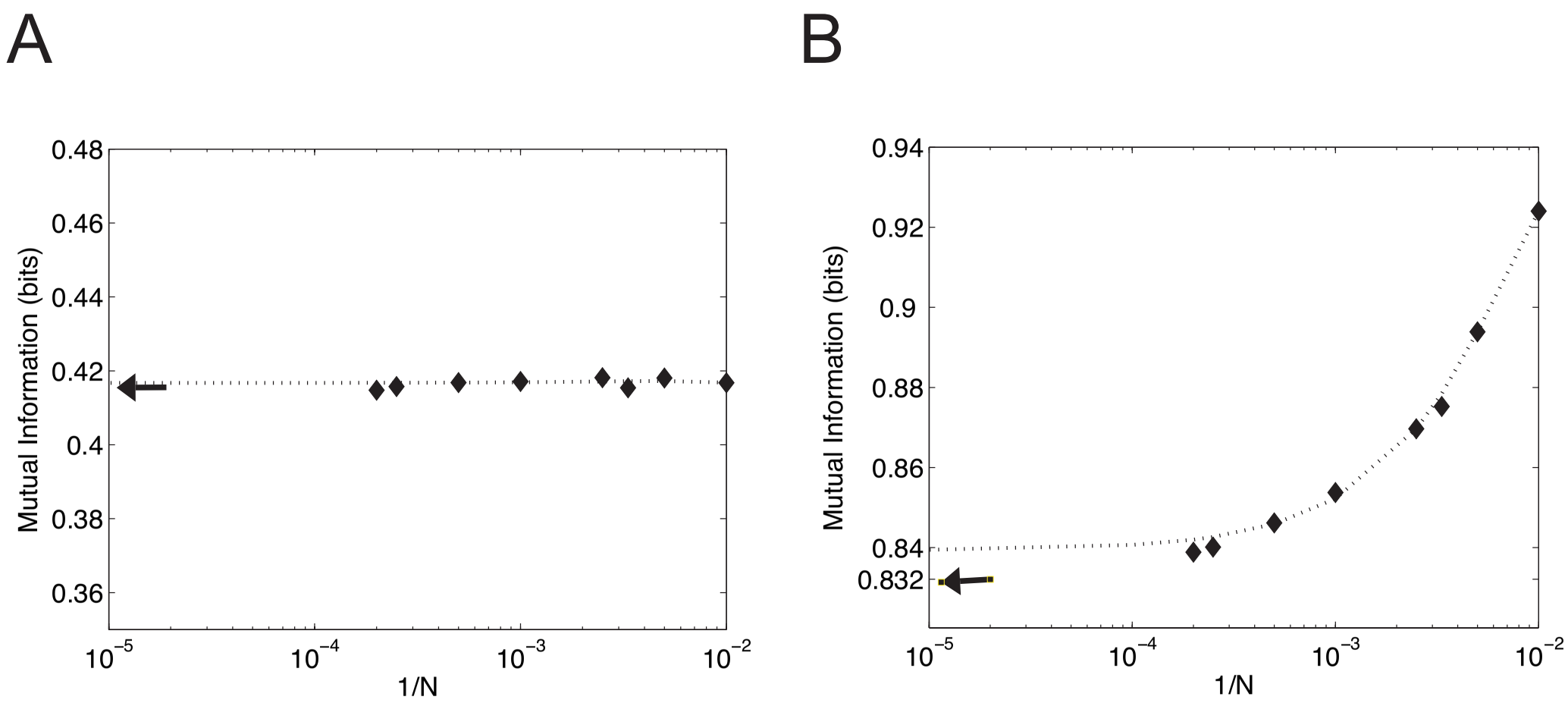

A
B
Mutual Information (bits)
0.48
0.46
0.44
0.42
0.4
0.38
0.36
10^{-5}
10^{-4}
10^{-3}
10^{-2}
1/N
Mutual Information (bits)
0.94
0.92
0.9
0.88
0.86
0.84
0.832
10^{-5}
10^{-4}
10^{-3}
10^{-2}
1/N

Figure S4

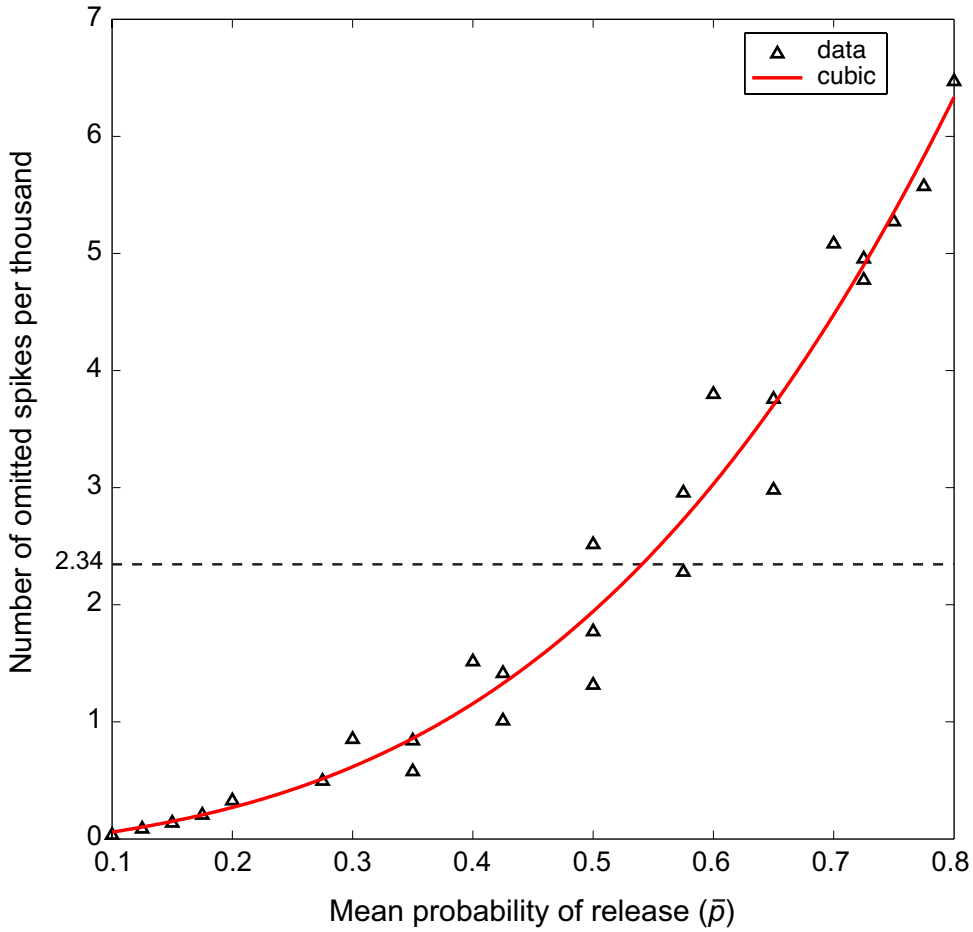

data
cubic
Number of omitted spikes per thousand
Mean probability of release ($\bar{p}$)
0
1
2
2.34
3
4
5
6
7
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

Figure S5

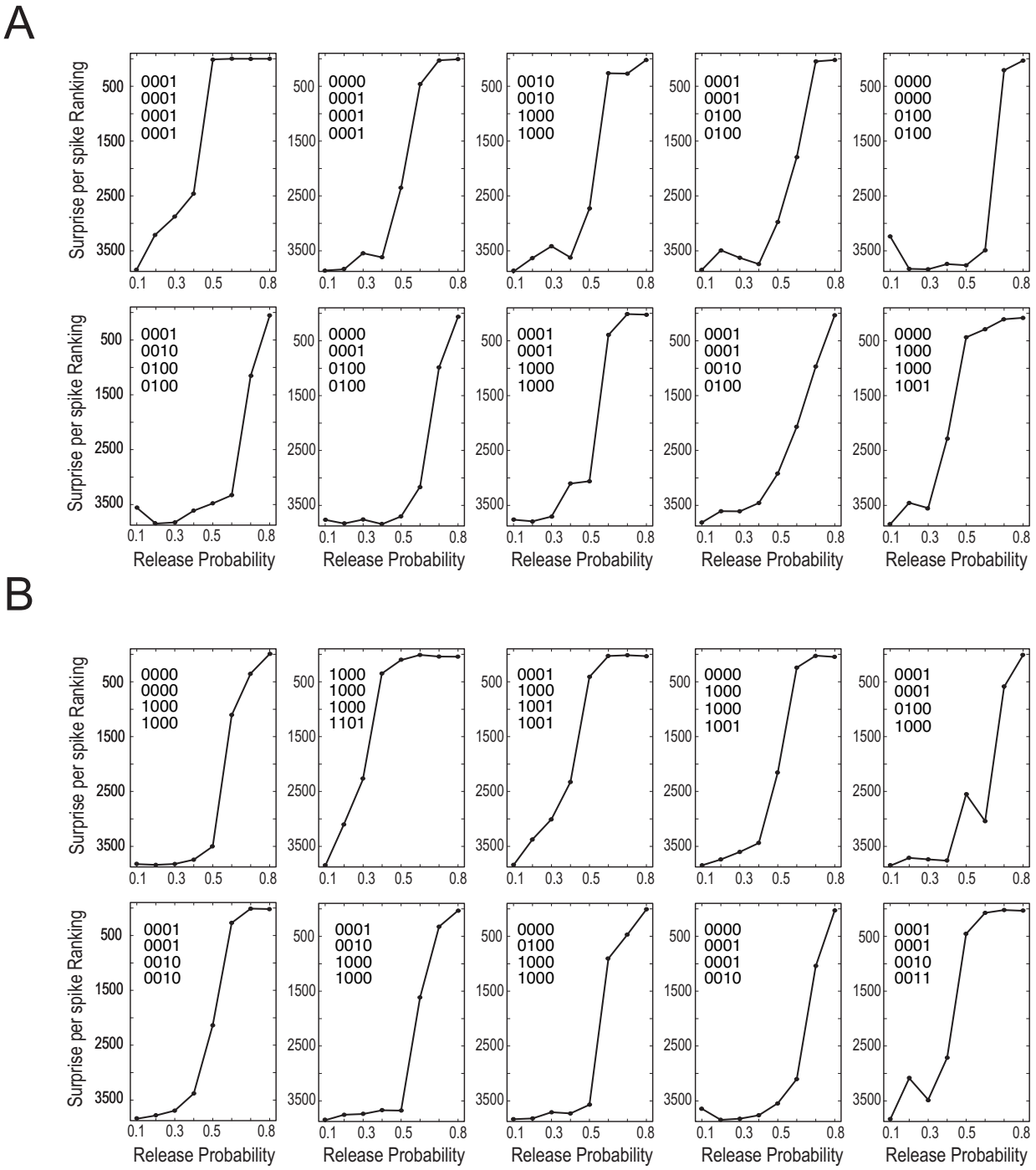

A
B
Surprise per spike Ranking
Release Probability
0001 0001 0001 0001
0000 0001 0001 0001
0010 0010 1000 1000
0001 0001 0100 0100
0000 0000 0100 0100
0001 0010 0100 0100
0000 0001 0100 0100
0001 0001 1000 1000
0001 0001 0010 0100
0000 1000 1000 1001
0000 0000 1000 1000
1000 1000 1000 1101
0001 1000 1001 1001
0000 1000 1000 1001
0001 0001 0100 1000
0001 0001 0010 0010
0001 0010 1000 1000
0000 0100 1000 1000
0000 0001 0001 0010
0001 0001 0010 0011

Figure S6

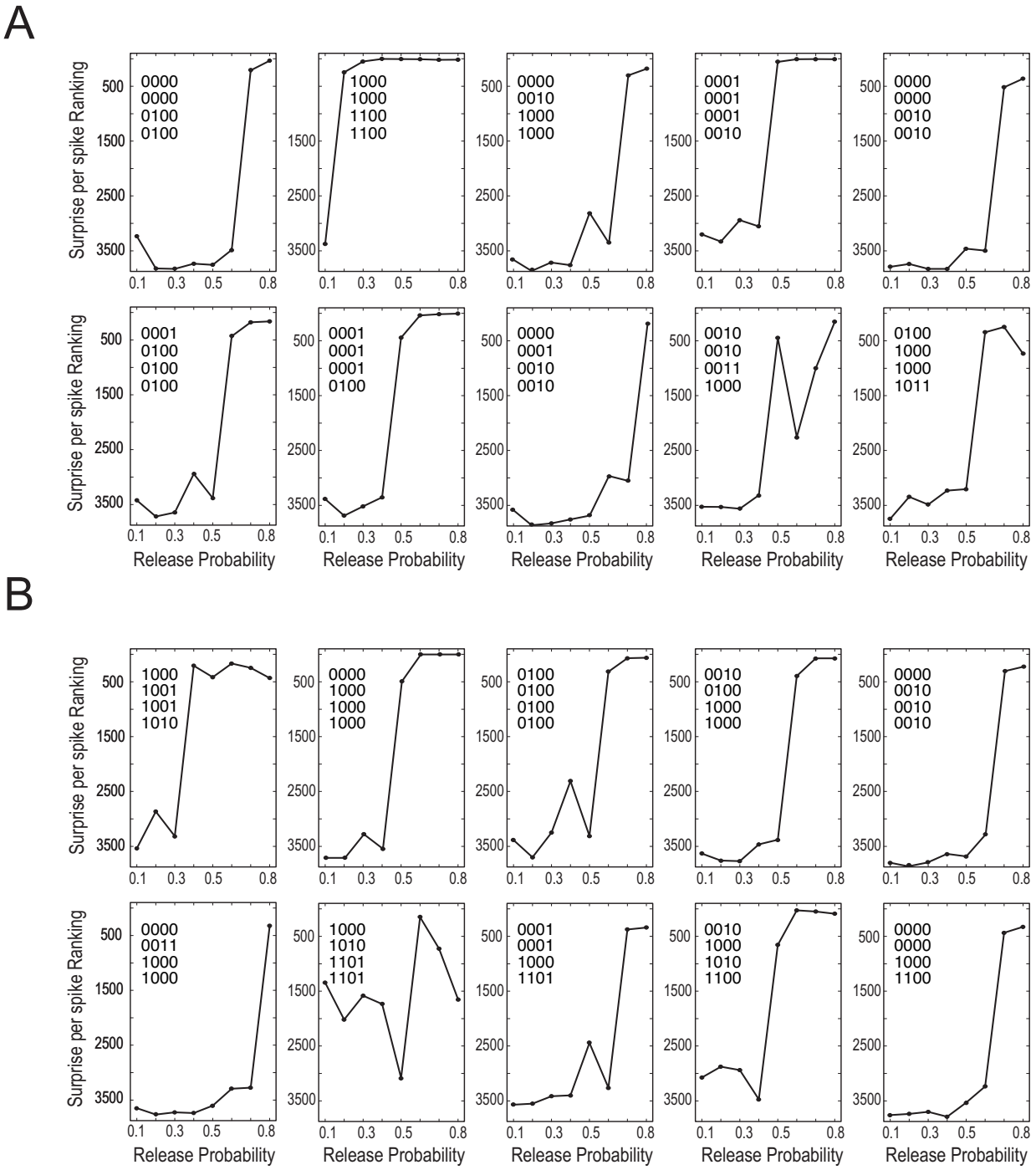

A
B
Surprise per spike Ranking
Release Probability
0000 0000 0100 0100
1000 1000 1100 1100
0000 0010 1000 1000
0001 0001 0001 0010
0000 0000 0010 0010
0001 0100 0100 0100
0001 0001 0001 0100
0000 0001 0010 0010
0010 0010 0011 1000
0100 1000 1000 1011
1000 1001 1001 1010
0000 1000 1000 1000
0100 0100 0100 0100
0010 0100 1000 1000
0000 0010 0010 0010
0000 0011 1000 1000
1000 1010 1101 1101
0001 0001 1000 1101
0010 1000 1010 1100
0000 0000 1000 1100

Figure S7

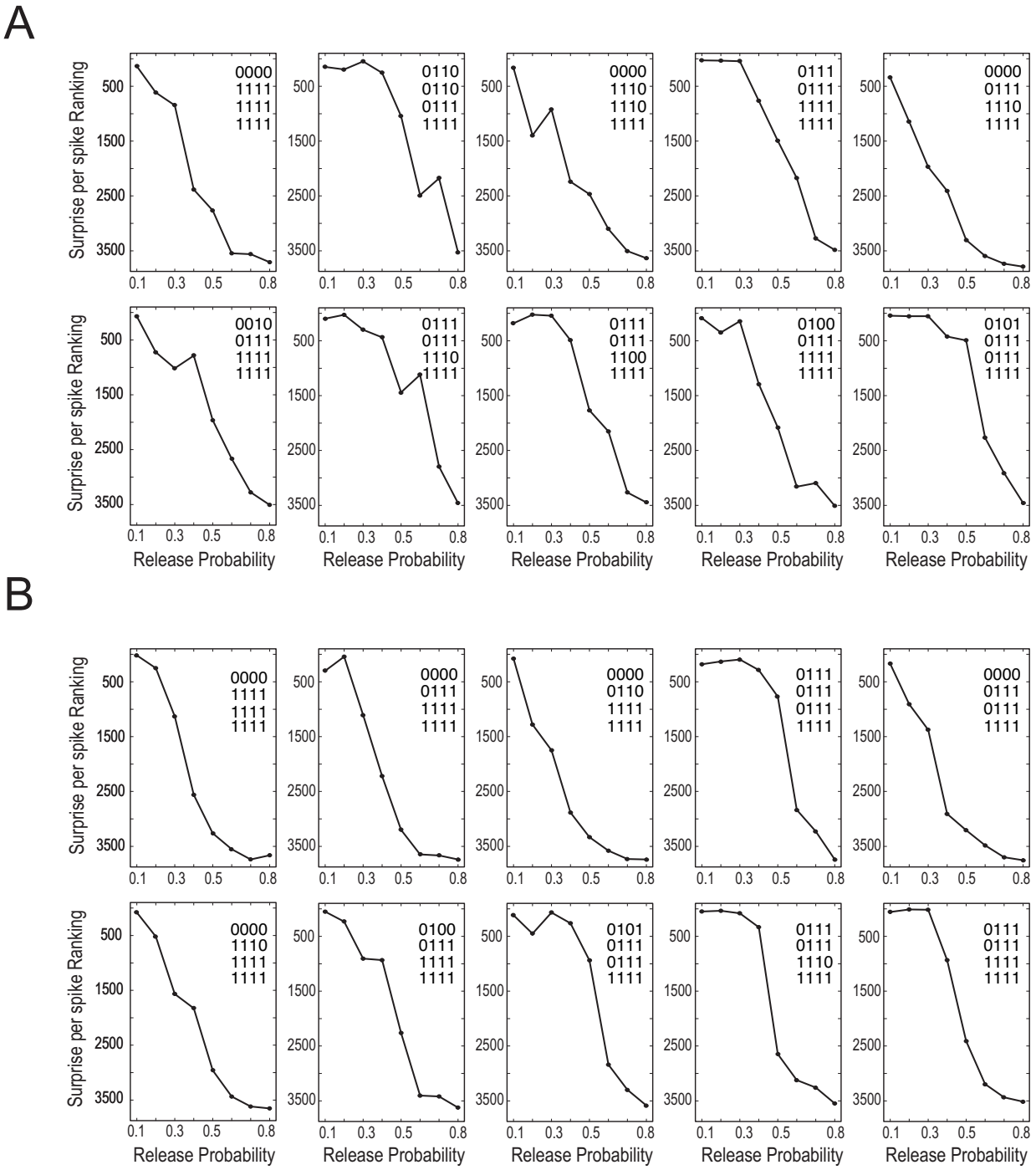

A
B
Surprise per spike Ranking
Release Probability
0000 1111 1111 1111
0110 0110 0111 1111
0000 1110 1110 1111
0111 0111 1111 1111
0000 0111 1110 1111
0010 0111 1111 1111
0111 0111 1110 1111
0111 0111 1100 1111
0100 0111 1111 1111
0101 0111 0111 1111
0000 1111 1111 1111
0000 0111 1111 1111
0000 0110 1111 1111
0111 0111 0111 1111
0000 0111 0111 1111
0000 1110 1111 1111
0100 0111 1111 1111
0101 0111 0111 1111
0111 0111 1110 1111
0111 0111 1111 1111

Figure S8

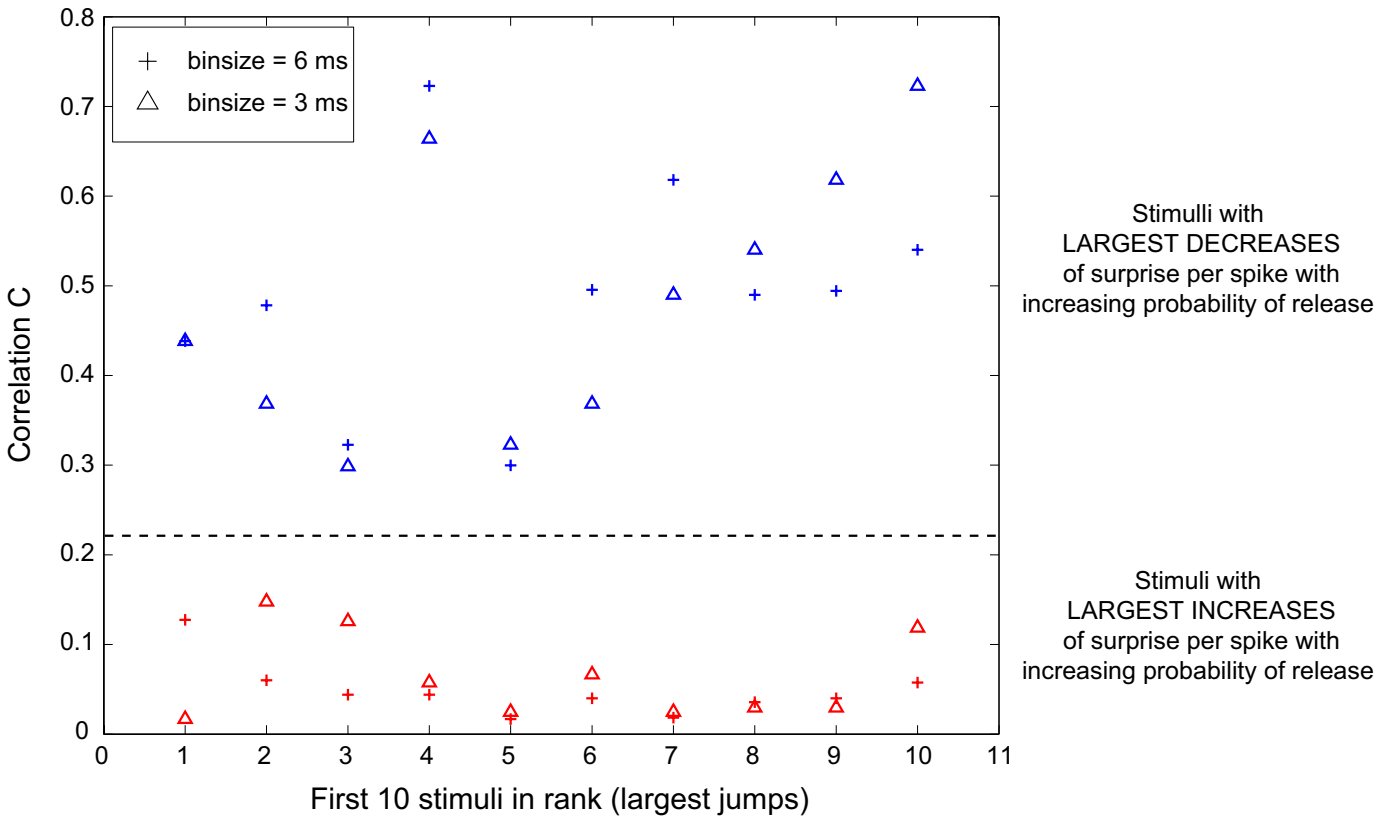

binsize = 6 ms
binsize = 3 ms
Correlation C
First 10 stimuli in rank (largest jumps)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
6
7
8
9
10
11
Stimulli with
LARGEST DECREASES
of surprise per spike with
increasing probability of release
Stimuli with
LARGEST INCREASES
of surprise per spike with
increasing probability of release

Figure S9

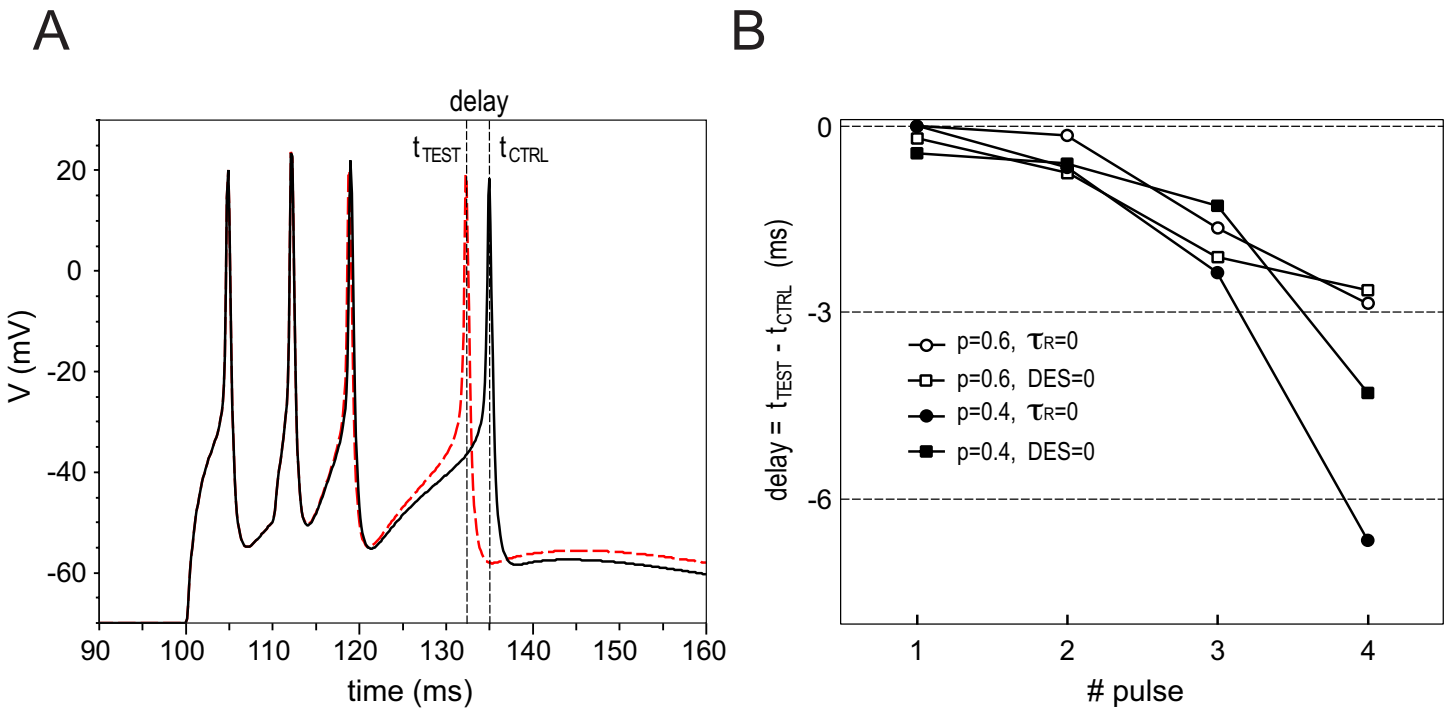

A
B
delay
t_{TEST}
t_{CTRL}
V (mV)
time (ms)
delay = t_{TEST} - t_{CTRL} (ms)
# pulse
p=0.6, τ_R=0
p=0.6, DES=0
p=0.4, τ_R=0
p=0.4, DES=0

Figure S10

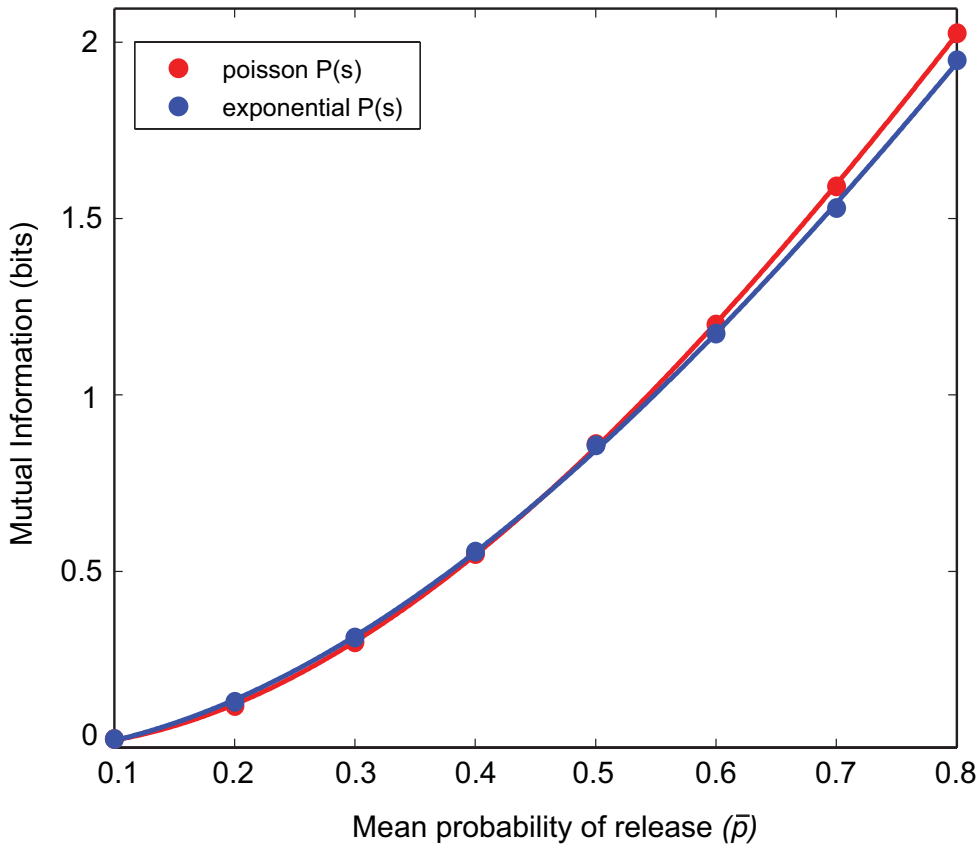

poisson P(s)
exponential P(s)
Mutual Information (bits)
Mean probability of release ($\bar{p}$)
0
0.5
1
1.5
2
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

Figure S11

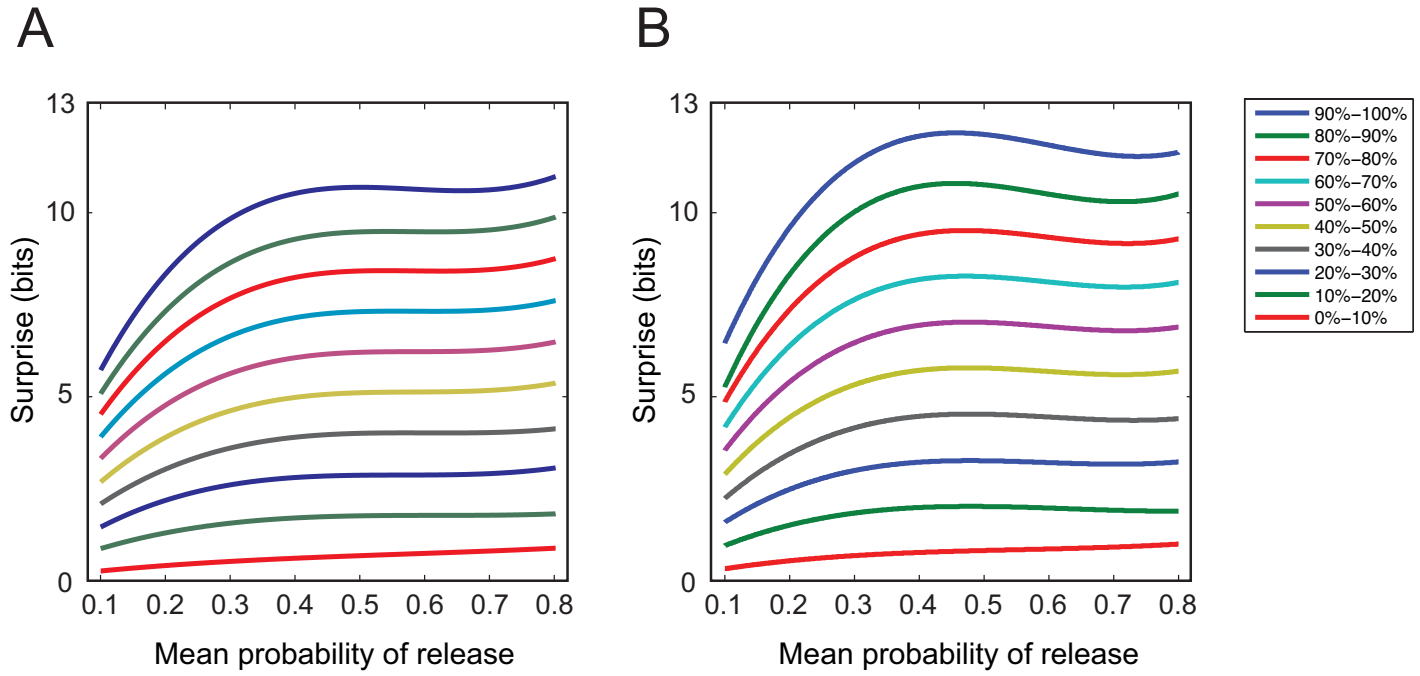
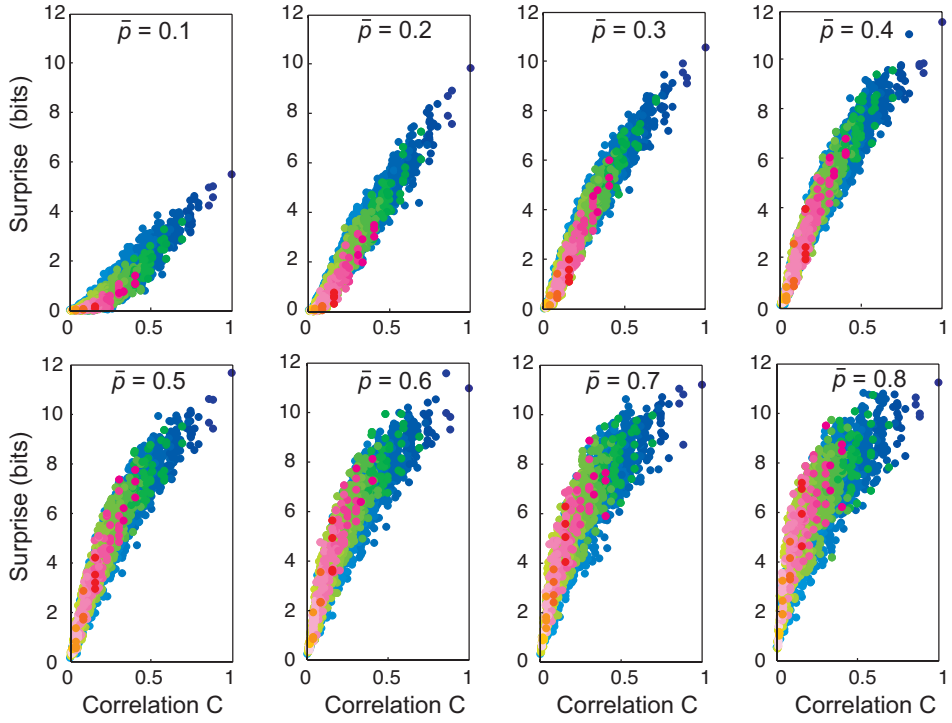

A
B
13
10
5
0
Surprise (bits)
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Mean probability of release
90%–100%
80%–90%
70%–80%
60%–70%
50%–60%
40%–50%
30%–40%
20%–30%
10%–20%
0%–10%

Figure S12

A



B

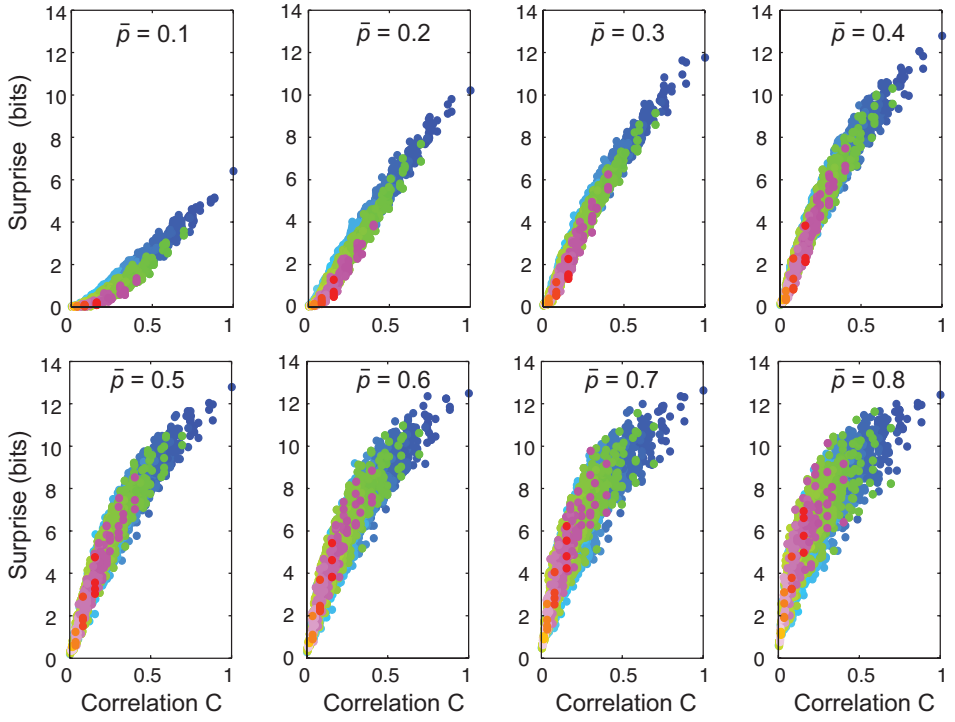
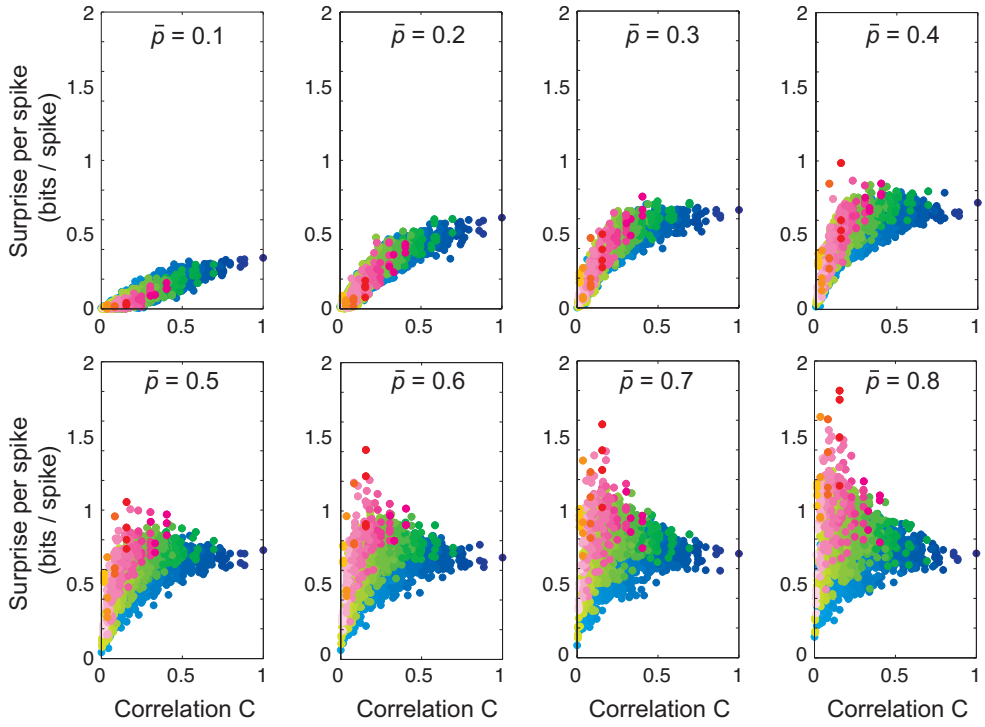


Figure S13

A



B

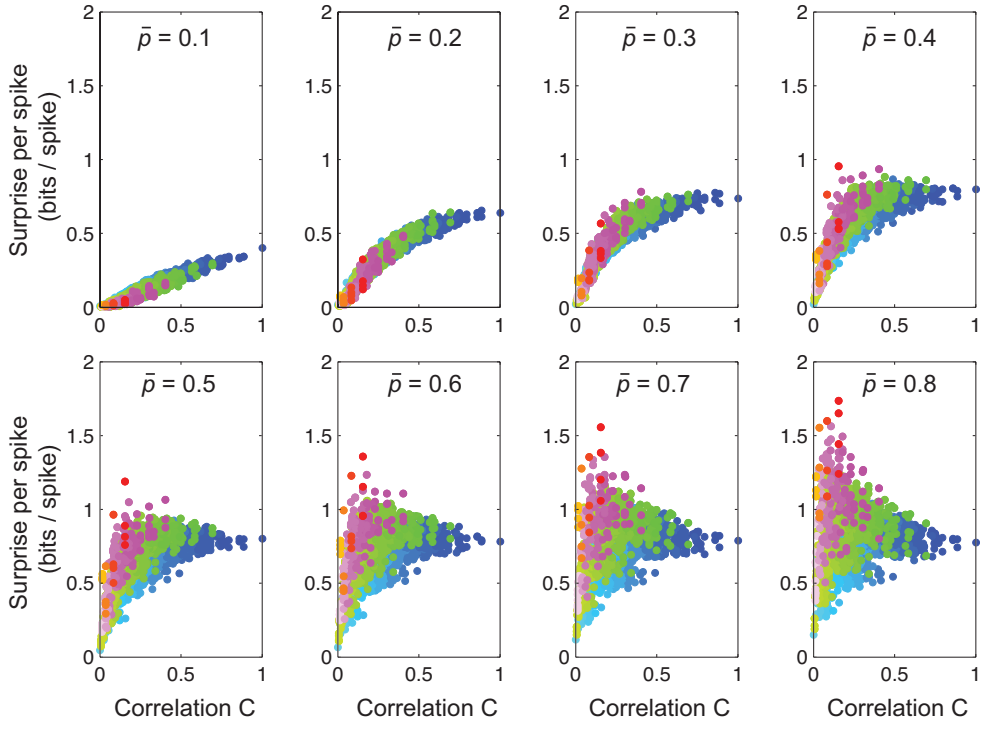
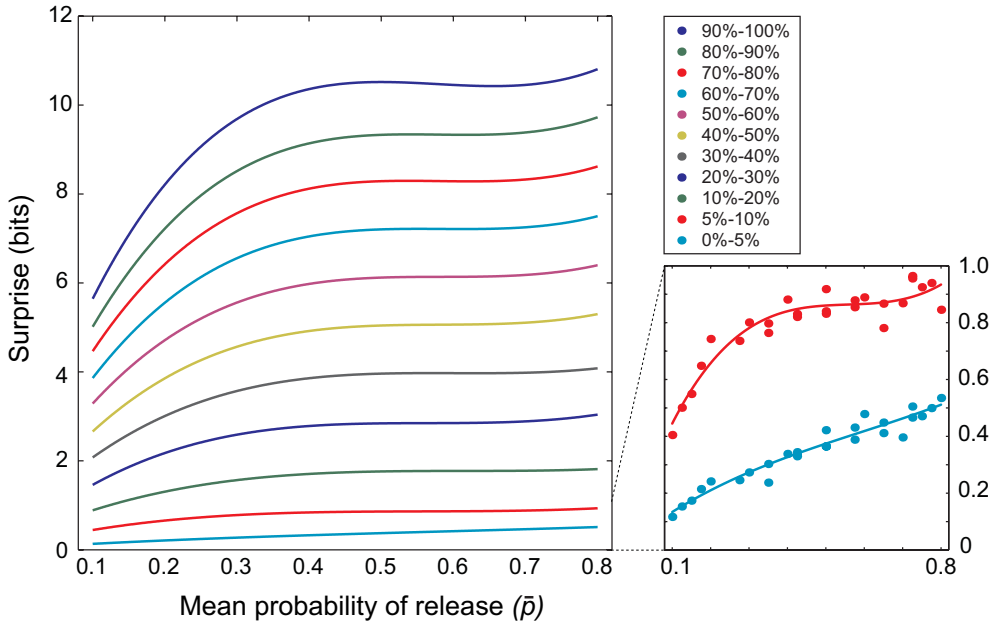


Figure S14

A



B

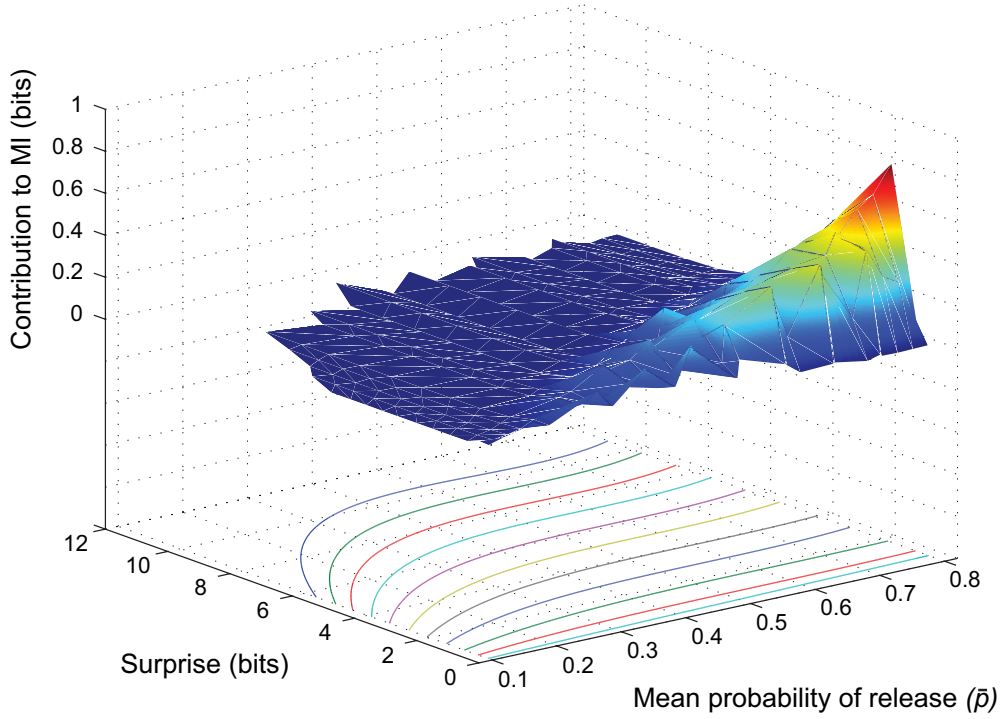


Figure S15

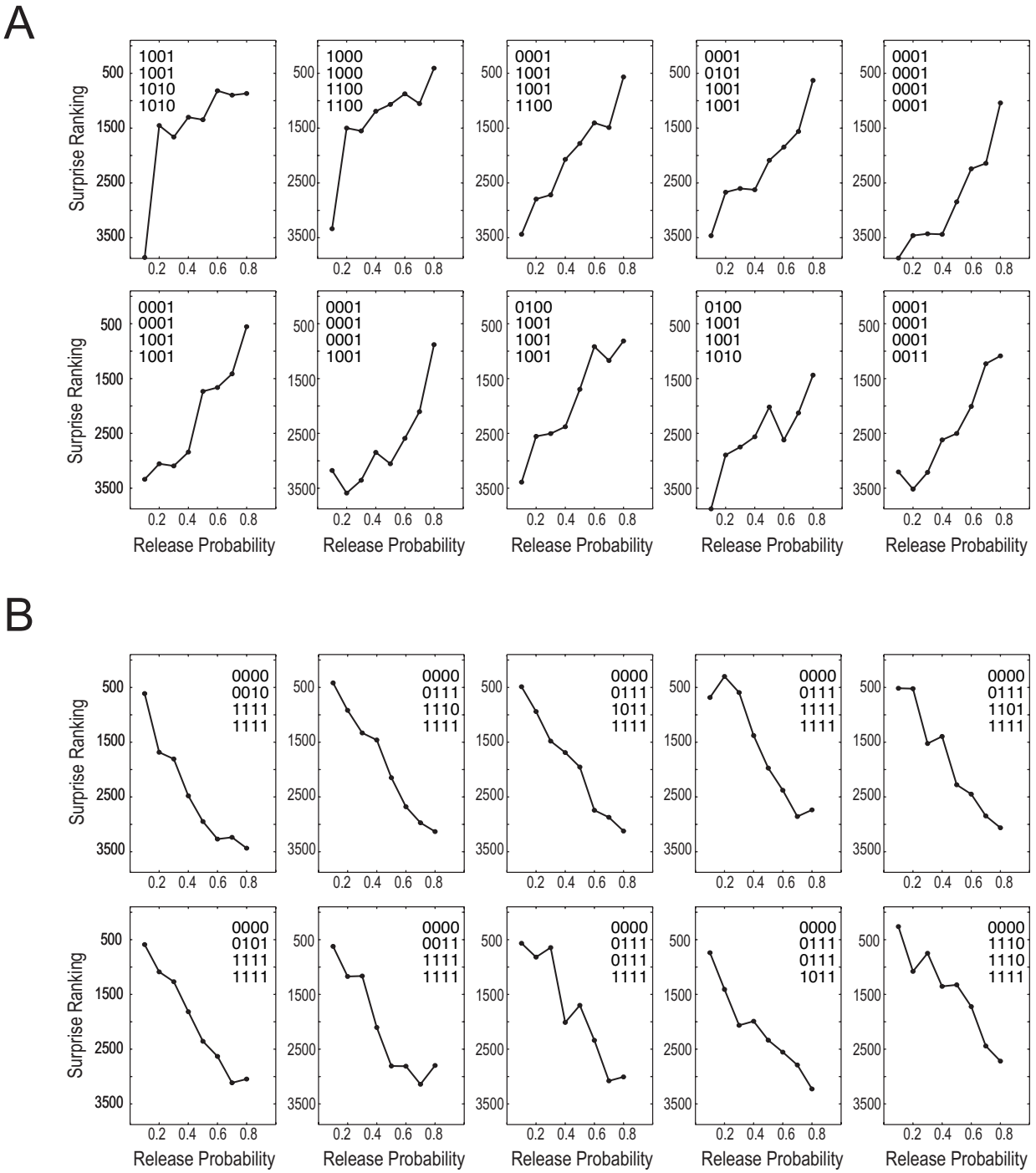

A
1001
1001
1010
1010
1000
1000
1100
1100
0001
1001
1001
1100
0001
0101
1001
1001
0001
0001
0001
0001
0001
0001
1001
1001
0001
0001
0001
1001
0100
1001
1001
1001
0100
1001
1001
1010
0001
0001
0001
0011
B
0000
0010
1111
1111
0000
0111
1110
1111
0000
0111
1011
1111
0000
0111
1111
1111
0000
0111
1101
1111
0000
0101
1111
1111
0000
0011
1111
1111
0000
0111
0111
1111
0000
0111
0111
1011
0000
1110
1110
1111
Surprise Ranking
Release Probability
500
1500
2500
3500
0.2
0.4
0.6
0.8

Figure S16

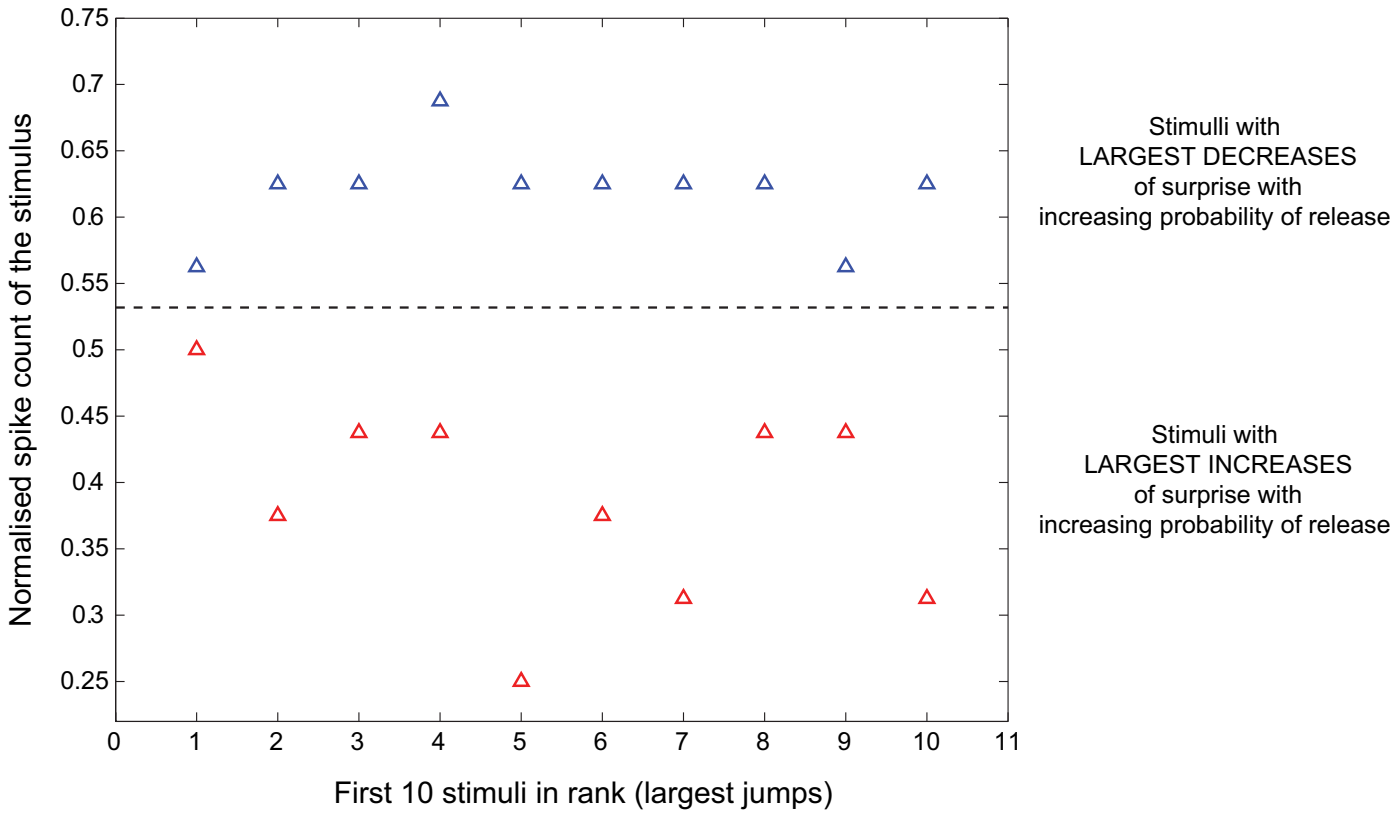

0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
Normalised spike count of the stimulus
0
1
2
3
4
5
6
7
8
9
10
11
First 10 stimuli in rank (largest jumps)
Stimulli with
LARGEST DECREASES
of surprise with
increasing probability of release
Stimuli with
LARGEST INCREASES
of surprise with
increasing probability of release

Figure S17

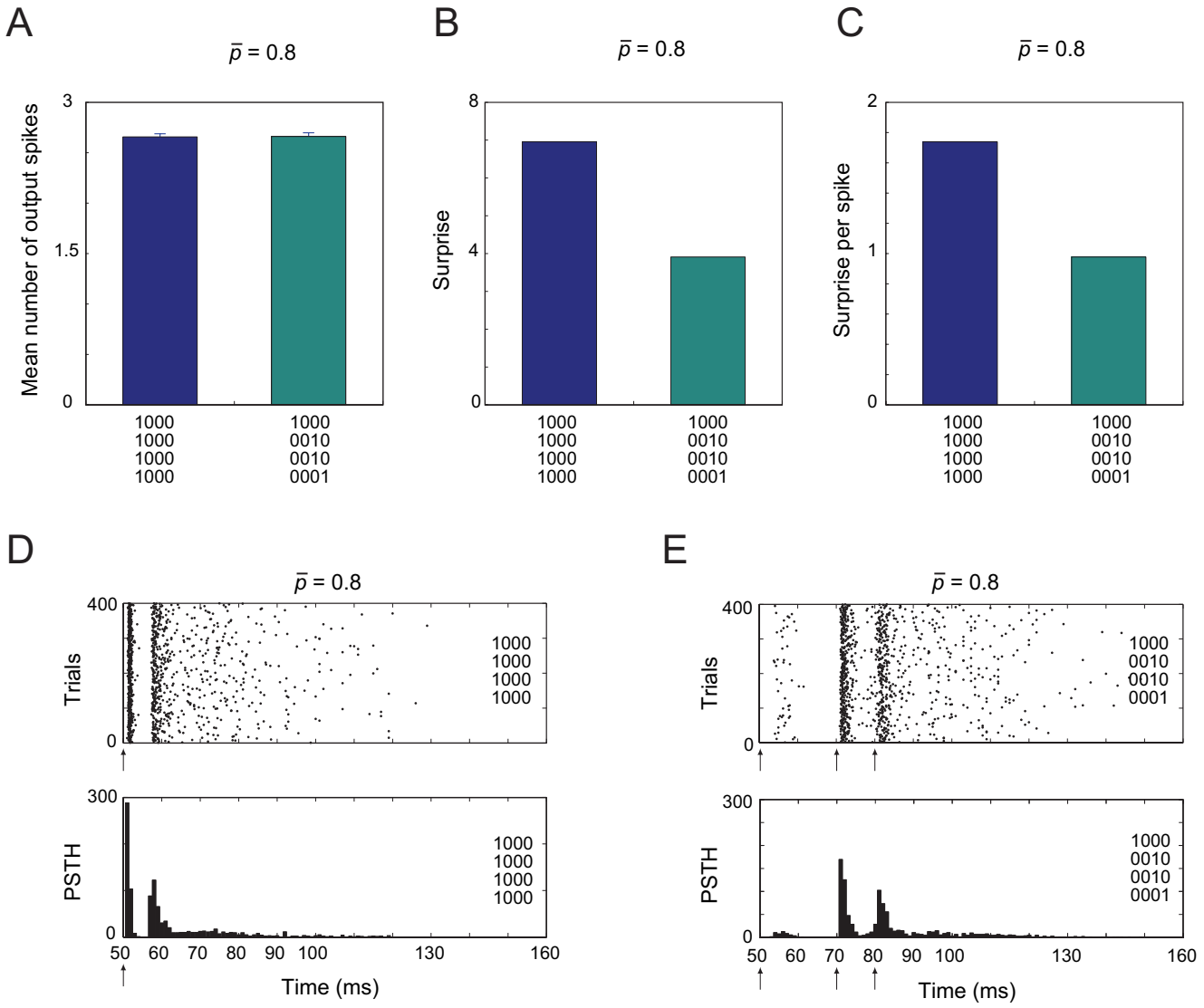

A
$\bar{p}$ = 0.8
Mean number of output spikes
3
1.5
0
1000 1000 1000 1000
1000 0010 0010 0001
B
$\bar{p}$ = 0.8
Surprise
8
4
0
1000 1000 1000 1000
1000 0010 0010 0001
C
$\bar{p}$ = 0.8
Surprise per spike
2
1
0
1000 1000 1000 1000
1000 0010 0010 0001
D
$\bar{p}$ = 0.8
Trials
400
0
1000 1000 1000 1000
PSTH
300
0
1000 1000 1000 1000
50 60 70 80 90 100 130 160
Time (ms)
E
$\bar{p}$ = 0.8
Trials
400
0
1000 0010 0010 0001
PSTH
300
0
1000 0010 0010 0001
50 60 70 80 90 100 130 160
Time (ms)

Figure S18

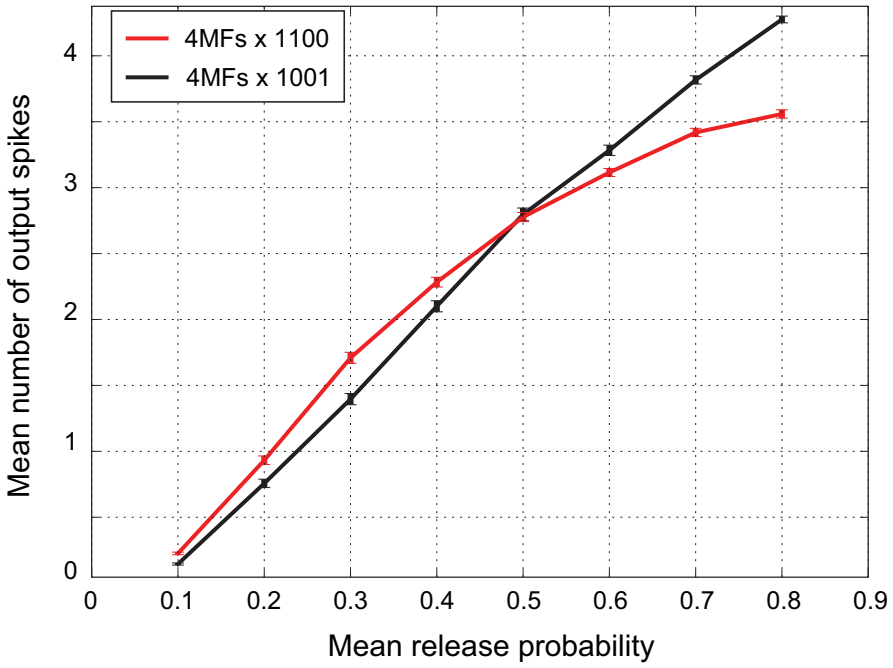

4MFs x 1100
4MFs x 1001
Mean number of output spikes
Mean release probability
0
1
2
3
4
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

Figure S19

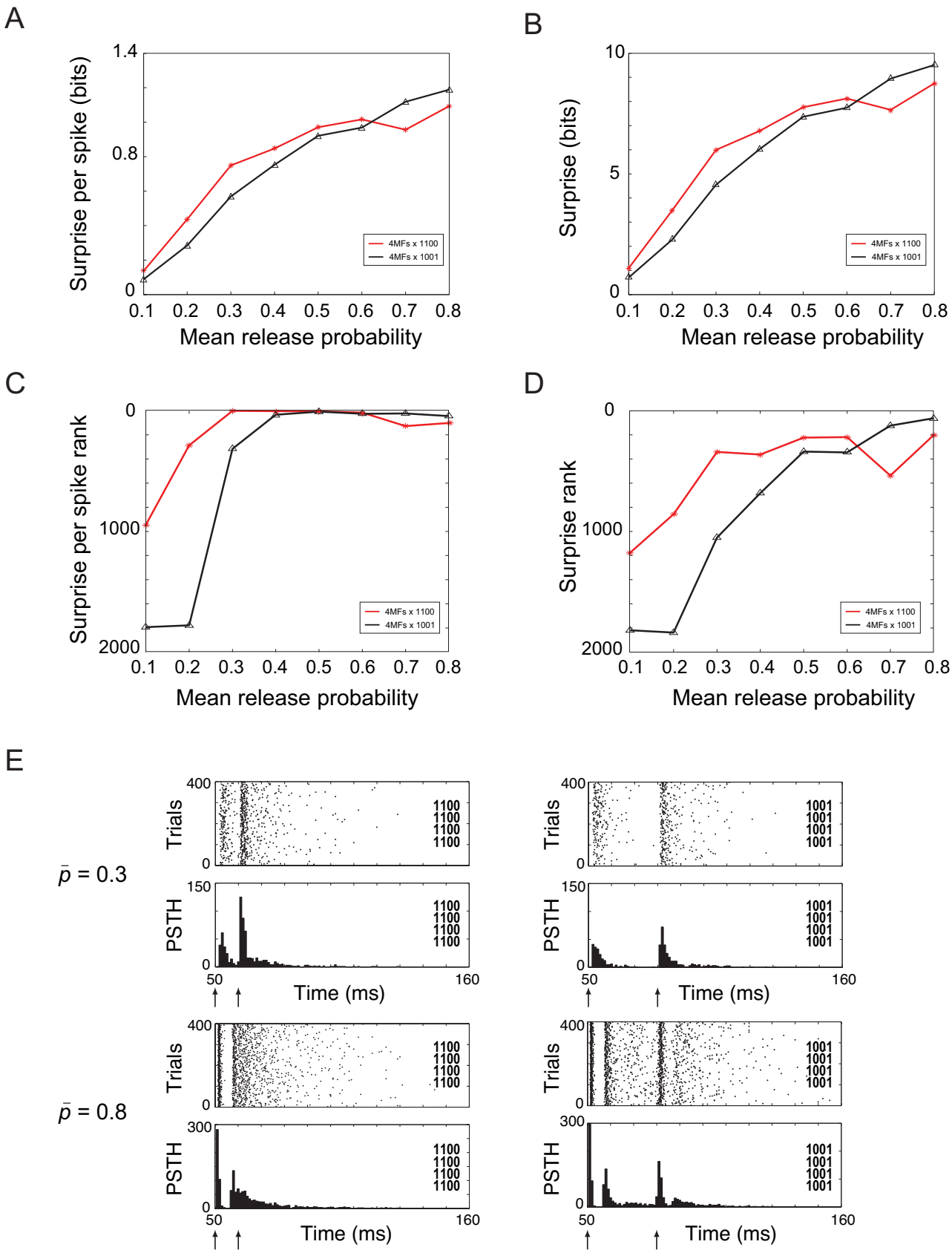

A
Surprise per spike (bits)
1.4
0.8
0
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8
Mean release probability
4MFs x 1100
4MFs x 1001
B
Surprise (bits)
10
5
0
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8
Mean release probability
4MFs x 1100
4MFs x 1001
C
Surprise per spike rank
0
1000
2000
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8
Mean release probability
4MFs x 1100
4MFs x 1001
D
Surprise rank
0
1000
2000
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8
Mean release probability
4MFs x 1100
4MFs x 1001
E
$\bar{p}$ = 0.3
$\bar{p}$ = 0.8
Trials
PSTH
400
150
300
0
1100
1001
50
160
Time (ms)

Figure S20

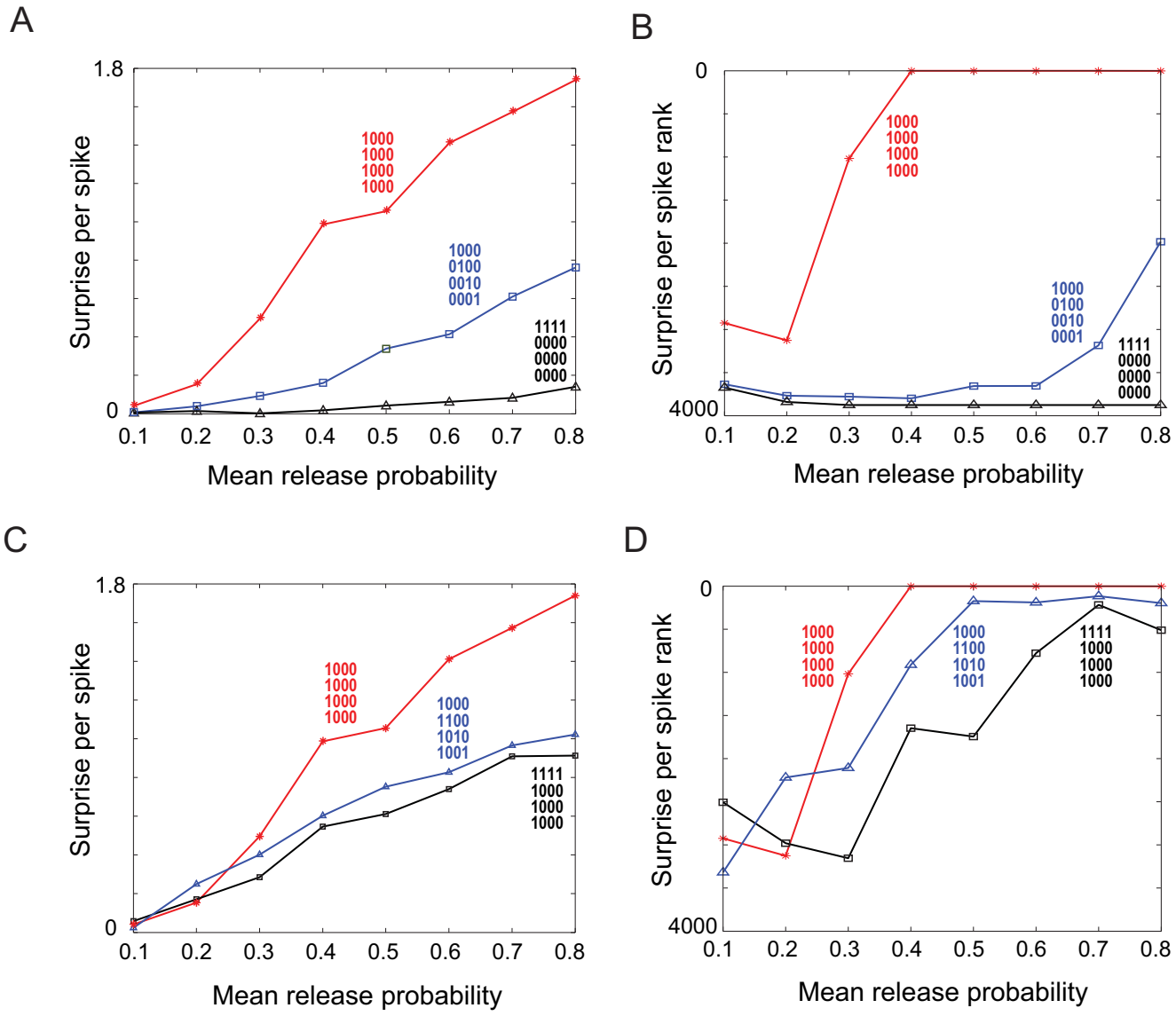

A
1.8
0
Surprise per spike
1000
1000
1000
1000
1000
0100
0010
0001
1111
0000
0000
0000
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8
Mean release probability
B
0
4000
Surprise per spike rank
1000
1000
1000
1000
1000
0100
0010
0001
1111
0000
0000
0000
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8
Mean release probability
C
1.8
0
Surprise per spike
1000
1000
1000
1000
1000
1100
1010
1001
1111
1000
1000
1000
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8
Mean release probability
D
0
4000
Surprise per spike rank
1000
1000
1000
1000
1000
1100
1010
1001
1111
1000
1000
1000
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8
Mean release probability

Figure S21

State variables	m n	g_{max} (S/cm²)	V_{rev} (mV)	α (s⁻¹)	β (s⁻¹)
$g_{Na\text{-}f}$		.013	87.39		
activation	3			0.9(V+19)/(1-exp(-(V+19)/10))	36 exp(-0.055(V+44))
inactivation	1			0.315 exp(-0.3(V+44))	4.5/(1+exp(-(V+11)/5))
$g_{Na\text{-}p}$		2e-5	87.39		
activation	1			0.91(V+40)/(1-exp(-(V+40)/5)) x_∞=1/(1+exp(-(V+40)/5)) τ_∞=5/(α+β)	-0.62(V+40)/(1-exp((V+40)/5))
$g_{Na\text{-}r}$		2e-4	87.39		
activation	1			2.4e-4-0.015(V+4.5) / ((exp(-(V+4.5)/6.8)-1))	0.14+0.047(V+44)/(exp((V+44)/0.11)-1))
inactivation	1			0.96 exp(-(V+80)/62.5)	0.03 exp((V+83.3)/16.1)
$g_{K\text{-}V}$		.003	-84.69		
activation	4			0.135 (V+25)/(1-exp(-(V+25)/10))	1.69 exp(-0.0125(V+35))
$g_{K\text{-}A}$		.0032	-84.69		
activation	3			2.44/(1+exp(-(V+9.17)/23.32))	0.5 exp(-(V+18.28)/19.47)
inactivation	1			0.11/(1+exp((V+111.33)/12.84) x_∞=1/(1+exp(-(V+38)/17)) y_∞=1/(1+exp((V+78.8)/8.4))	0.1/(1+exp(-(V+49.95)/8.9))
$g_{K\text{-}IR}$		.0009	-84.69		
activation	1			0.4 exp(-0.041(V+83.94))	0.51 exp(0.028(V+83.94))
$g_{K\text{-}Ca}$		.003	-84.69		
activation	1			2.5/(1+1.5e-3/[Ca] exp(-0.085V))	1.5/(1+[Ca]/(0.15e-3 exp(-0.085V)))
g_{Ca}		4.6e-4	129.3		
activation	2			0.148 exp((v+29.06)/ 15.9)	0.249*exp(-(v+18.66)/25.6)
inactivation	1			3.9e-3exp(-(v+48)/18.2)	3.9e-3 exp((v+48)/83.3)
$g_{K\text{-}slow}$		2.5e-4	-84.69		
activation	1			0.01 exp(0.025(V+30)) x_∞=1/(1+exp(-(V+35)/6))	0.01 exp(-0.05(V+30))
$g_{Leakage}$		5.68e-5	-16.5		
g_{GABA}		3e-5	-65		
Calcium concentration		d=200 nm β_{Ca}=1.5 $[Ca^{2+}]_0$ = 100 nM			

Table S1

AMPA	
rDC, rCD	0.013, 1.12
rCO, rOC	5.4, 0.82
NMDA	
rC0C1*, rC1C0	5, 0.1
rC1C2*, rC2C1	5, 0.1
rC2D, rDC2	0.00012, 0.009
rOC2, rC2D	0.966, 0.03

Table S2