

Supplementary Information to “State-dependent representation of stimulus-
evoked activity in high-density recordings of neural cultures”

Thierry Nieuws^{1,4*¶}, Valeria D’Andrea^{2¶}, Hayder Amin¹, Stefano di Marco^{1#a}, Houman Safaai^{2,3},
Alessandro Maccione¹, Luca Berdondini^{1&}, Stefano Panzeri^{2&*}

¹ NetS³ Laboratory, Neuroscience and Brain Technologies Department., Istituto Italiano di
Tecnologia, Genova, Italy.

² Neural Computation Laboratory, Center for Neuroscience and Cognitive Systems
@UniTn, Istituto Italiano di Tecnologia, Rovereto, Italy.

³ Department of Neurobiology, Harvard Medical School, 02115 Boston, MA

⁴ Department of Biomedical and Clinical Sciences "Luigi Sacco", Università di Milano, Milano, Italy

^{#a} Current Address: Department Scienze Cliniche Applicate e Biotecnologiche, Università
dell’Aquila, Italy

*Corresponding authors

E-mails: thierry.nieuws@unimi.it (TN) (lead contact), stefano.panzeri@iit.it (SP)

¶ TN and VD are Joint First Authors.

& SP and LB are Joint Senior Authors

25 **Supplementary Methods**

26 27 **Preparation of neuronal cultures and CMOS-MEA recordings**

28 We performed electrophysiological measures of ongoing and electrically evoked responses from
29 cultured neuronal networks with primary neurons grown on high-density CMOS-MEAs (BioChip
30 4096E, from 3Brain AG, Switzerland), that provide up to 16 on-chip stimulation sites and 4096
31 recording microelectrodes, both of 21 μm x 21 μm in size. We prepared primary hippocampal
32 neurons from E18 rat embryos (Sprague Dawley from Charles River Laboratories International) ¹.
33 Before cellular seeding, we prepared chips as follow. First, we sterilized the outside of the chips
34 with a tissue moistened with EtOH 96%. Then, we sterilized the well of each device for 20 min. with
35 70% EtOH and we rinsed abundantly with double-distilled water (DDW) before drying. We pre-
36 conditioned the devices overnight in an incubator with the well filled with Complete Neurobasal
37 Medium, containing 2% B-27 1% penicillin/streptomycin and 1% GlutaMax supplements (all
38 reagents from Life Technologies). Successively, we prepared the active area of each device for
39 cellular seeding by coating the electrode array area with 50 $\mu\text{g}/\text{ml}$ poly-dl-ornithine (PDLO) (Sigma-
40 Aldrich) as adhesion promoting layer and incubated overnight at 37°C and 5% CO₂. The day after,
41 we rinsed the CMOS-MEAs 3 to 4 times with sterile DDW and we left them drying under a sterile
42 hood. We seeded cells on the pre-coated substrates and we maintained them in an incubator at
43 37°C with 5% CO₂ and 95% humidity. To do so, we diluted the cellular suspension in medium
44 (Neurobasal, from Invitrogen) at a nominal concentration of 1000 cell/ μL and we seeded drops of
45 90 μL on the active area of the chips. After 2 hours from cellular seeding, we added Neurobasal
46 supplemented with 2% of B27, 1% of Glutamax. This resulted in a neuronal seeding density of
47 ~ 2500 cell/ mm^2 . The medium was exchanged every four days by removing 1/3 of the total volume
48 and by replacing it with the fresh complete Neurobasal medium. Experiments started after 24 days-
49 in-vitro (DIVs) cell culture. To record and evoke spiking activity from CMOS-MEAs we used a
50 custom setup (sampling frequency of 7.7 KHz/electrode for full-array recordings) connected to a
51 Plexon Stimulator 2.0 (Plexon Inc., USA). The experimental protocol included three phases that
52 consisted in the recording of the spontaneous activity before electrical stimulation (10 min), the

53 recording of the ongoing activity and electrically evoked responses (40 min., 60 trials per electrode)
54 and the recording of the post-stimulation activity (10 min.). The same protocol was repeated after
55 addition of norepinephrine (5 μ M , 10 min. incubation). We used a PDMS cap to ensure sterile and
56 stable conditions in the cell culture well during long experimental recordings. For the electrical
57 stimulation we applied rectangular biphasic current pulses (amplitude adjusted between 200-400
58 μ A; duration of 600 μ sec). The network was stimulated at 0.2 Hz, i.e. inter-stimulus was 5 seconds,
59 and each stimulus was delivered from 8 spatially distinct sites using randomized electrode
60 sequences. To program the stimulation sequences and to operate the electrical stimulator we
61 developed a software tool in C# language (Visual Studio 2010; .NET framework 3.5).

62

63 **Spike detection and first order statistics of network activity**

64 We performed spike detection by using the BrainWave software (3Brain AG, Switzerland) and the
65 Precise Timing Spike Detection (PTSD) algorithm ². Successively, we computed mean activity
66 parameters to characterize the network activity. We performed the first order statistics and network
67 burst detection as previously reported ³. This includes the computation of the mean firing rate (i.e.
68 total number of spikes over the recording duration), the mean bursting rate (i.e. total number of
69 burst over the recording duration), the network burst rate and the mean intra-burst firing rate (i.e.
70 the number of spikes fired in the network burst).

71

72 **Quantification of the response features**

73 We quantified the stimulus-evoked response of the network in terms of different features of the
74 spiking activity, computed as follows. First, for each electrode, the spiking activity was smoothed
75 by counting the spikes in temporal windows of fixed duration (i.e. 20 ms) and slid by 5 ms over the
76 time window of interest. This procedure yielded the spike count: $ST(t) =$
77 $[st_1(t), st_2(t), \dots, st_k(t), \dots st_N(t)]$, where $st_k(t)$ is the number of spikes elicited by the electrode k
78 ($k = 1, \dots, N$; with $N=4096$) in the time window $[t \ t+20]$ ms. Note that the first 10 ms time window
79 after the stimulus, the so called “early phase” of the response, was removed since in this time
80 interval the response is very reproducible across trials because it does not involve synaptic

transmission ^{4,5} and state dependent processing likely does not take place in this early response. The network spike count (MUA) was computed as the sum of $ST(t)$ across all electrodes. We further analysed the spike count $ST(t)$ in terms of the centre activity trajectory (CAT, see ⁶) and principal component analysis (PCA). With the CAT, the N-dimensional network activity is reduced to the centre of mass of the spiking activity (i.e. two coordinates) over time bins of 20 ms and it provides a representation of the propagating network events. With the PCA, the variance of the N-dimensional network activity is represented in a lower dimensional space and it has proven to provide an effective mean to cluster propagating activities in cell cultures ⁷. In the latter lower dimensional space the spike count $ST(t)$ can be approximated by $\sum_{j=1}^M \alpha_j(t) \cdot PC_j$ where $\alpha_j(t) = \langle ST(t), PC_j \rangle$ is the projection of $ST(t)$ onto the eigenvector (basis) PC_j and M is the dimension of the lower dimensional space ($M \ll N$). Examples of spatial representations of the PC_j are reported in Supplementary Fig. S4 ($j=1,2,3$) while the time course of the different PCs used to compute information is represented by the $\alpha_j(t)$ functions.

94

95 **Numerical procedures to compute direct estimates of mutual information and their** 96 **statistics**

97 Here we detail the numerical procedures we used to compute direct mutual information estimates
98 from Eqs (3-4) of the main text.

99 We first facilitated the sampling of probability by considering a low-dimensional representation of r
100 and θ , by discretizing r and θ in six equi-populated bins. The response could be either the MUA,
101 the CAT or a PC score at any given time point in a [0 500] ms post-stimulus time window. The
102 state variable θ was the time TB between last burst and stimulus application.

103 The estimation of the information from probabilities computed from a limited number of trials
104 suffers from a systematic error⁸. To correct for this problem, we used the Panzeri-Treves ⁹
105 correction that subtracts the bias of each information quantity.

106 In the following we explain how we evaluated with non-parametric permutation statistics, if a state
107 variable significantly changes the stimulus dependence of neural responses. This happens if the

state dependent information $I(S; R, \Theta)$ is larger than $I(S; R)$. However, when comparing these measures we had to consider that $I(S; R)$ has a smaller sampling bias than $I(S; R, \Theta)$ because the space of possible neural responses to a stimulus is one-dimensional rather than two-dimensional^{8,9}. To avoid this problem that may bias the comparisons, we first estimated $I(S; R)$ as $I(S; R, \Theta_{SH})$, that is a measure of information about the stimuli obtained jointly from responses and states, but after randomly permuting the values of the state variables Θ across trials at fixed value of response variables R . This shuffling of Θ ensures that the quantity $I(S; R, \Theta_{SH})$ has the same asymptotic value as $I(S; R)$, for infinite number of trials, but has the same bias as $I(S; R, \Theta_{SH})$ for a finite number of trials. We repeated the procedure 200 times and we used the distribution of random shuffles as a null hypothesis distribution and we computed $I(S; R)$ as the 50th percentile of shuffled distribution $I(S; R, \Theta_{SH})$. The ranking of $I(S; R, \Theta)$ among the shuffled test statistic $I(S; R, \Theta_{SH})$ gives the p value of the null test hypothesis of no difference between $I(S; R, \Theta)$ and $I(S; R)$. When conducting multiple comparisons (that is, for example, when considering multiple time points) we corrected for multiple comparisons using a False Discovery Rate (FDR)¹⁰.

122

123

124 **Computation of information from the confusion matrix of a decoder**

125

When including several state variables, the direct calculation of information from the stimulus-state-response probabilities (as in Eqs 3,4 of the main text) became problematic because of limited sampling issues. In this case, we thus computed information through a decoding approach^{9,11}. For each trial, we decoded the predicted stimulus sp from the neural activity recorded in that trial (see below for details). From these results, we then computed the confusion matrix $Q(S, SP)$ of the joint probability of presenting stimulus s and decoding stimulus sp . Finally, we computed the information obtained through the decoder as the information contained in the confusion matrix¹¹, as follows:

133

$$134 \quad \sum_{s, sp} q(s, sp) \log_2 \frac{q(s, sp)}{p(s)q(sp)} \quad (S1)$$

135 The decoding was performed using a linear discriminant analysis (LDA) classifier, with a 5 fold
136 cross-validation to separate training data (those used to set the boundaries of the linear
137 discriminant) and test data (those used to decode stimuli and compute the confusion matrix). For
138 computation of information decoded from the neural responses alone, we used a one-dimensional
139 neural activity feature consisting of the binned state post-stimulus responses (6 bins). For
140 computation of information decoded from the neural responses and state variables, we used a
141 multidimensional neural activity feature consisting of the binned post-stimulus responses and the
142 considered state variables (6 bins for each feature). For consistency, the binning of the neural
143 activity features into 6 bins was the same applied for the direct information calculation (Eqs 3-4 of
144 the main text).

145 We corrected the bias of the decoded information with the same Panzeri-Treves ⁹ procedure used
146 for the direct information estimates of Eqs. (3,4) of the main text. To evaluate the significance of
147 the increase of information due to state dependence, we used for the decoded information the
148 same non-parametric permutation used for the direct information and detailed above.

149 The advantage of computing information through the decoding matrix is that this approach is data
150 robust ¹¹. However, the information in the decoder's confusion matrix, Supplementary Information
151 Eq. S1, is a lower bound to the total information in neural activity, which is quantified by the direct
152 information, Eqs 3-4 of main text ¹¹. In other words, the information computed through a decoder
153 cannot be higher, and is in general much less, than the total information in neural activity. As
154 detailed in ¹¹, the information loss when decoding can be due to either or both of two factors. The
155 first is the potential inaccuracy of the decoding model in describing the relationship between neural
156 activity and stimuli. For example, a non-linearity in this relationship cannot be fully captured by a
157 linear decoder. The second is that reporting only the identity of the most likely stimulus given
158 neural activity, as the decoder does, leads inevitably to an intrinsic loss of some of the available
159 information. For example, reporting only the most likely stimulus loses information about the
160 relative likelihood of each stimulus that is present in neural activity. Comparisons of information
161 quantities computed with the direct computation of information and through a decoder can be used

162 to evaluate the efficiency of a neural decoding algorithm in extracting information from neural
163 activity.

164

165 **Information gain**

166 We defined the information gain due to the knowledge of state as the difference $I(S; R, \theta) - I(S; R)$.
167 This measure has the advantage of concentrating the effect of θ on the stimulus dependence of r
168 and this can be appreciated by considering that the following information-theoretic equality holds ¹²
169 :

170

$$171 \quad I(S; R, \theta) - I(S; R) - I(S; \theta) = \langle I(R; \theta|s) \rangle_s - I(R; \theta) \quad (S2)$$

172

173 where $\langle I(R; \theta|s) \rangle_s$ is the mutual information between R and θ given the stimulus s . Making the
174 reasonable hypothesis that $I(S; \theta) \cong 0$ (that is verified in our data, see Supplementary Fig. S1) it
175 follows that, if the relationship between response r and state θ depends on the stimulus, the
176 difference $I(S; R, \theta) - I(S; R)$ will be positive, and it will be zero otherwise. Hence difference
177 between $I(S; R, \theta)$ and $I(S; R)$ captures the effect of state on stimulus-dependent response
178 relationships. We also defined the percentage information gain as $100 * [I(S; R, \theta) - I(S; R)] /$
179 $I(S; R)$ and we evaluated the amount of information explained by the linear model by computing the
180 discount information ratio as: $I(S; R_d) / I(S; R, \theta)$, where R_d are the responses after discounting the
181 predicted single-trial variability due to state dependency .

182

183 **Spike synchronization**

184 We quantified the synchrony of the spontaneous spiking activity and how it changes when
185 norepinephrine is added to cell cultures. To this aim we analyzed the spontaneous activity when no
186 stimulus was applied to the cultures and we used the software SPIKY ¹³ to compute the
187 synchronization coefficient. Given the large number of electrodes of CMOS-MEAs, we sub-
188 sampled the electrode array and computed the synchronization coefficient on $N=100$ electrodes
189 randomly selected from the whole array. The random selection was repeated 100 times and the

190 synchronization coefficient was estimated as the average across all repetitions. The
191 synchronization coefficient computed by the SPIKY software was obtained by averaging over the
192 whole recording duration the time-resolved measure called SPIKE-synchronization. The latter is a
193 measure of similarity between simultaneously recorded spike trains that quantifies the fraction of
194 coincident spikes among all spike trains. SPIKE-synchronization values range between zero (spike
195 trains do not contain any coincidences in a small time window) and one (each spike in every spike
196 train has one matching spike with the other spike trains).

197

198 **Functional connectivity analysis and properties of the graphs**

199 We performed a cross-correlation analysis ¹⁴ to quantify the strength (i.e. cross-correlation peak) of
200 the functional connections (FC) between all pairs of electrodes. In order to retain significant
201 functional links, we introduced a null model obtained by shuffling the original spike trains while
202 maintaining the original ISI distribution. This procedure was applied to each pair of spike trains
203 (n=200 repetitions) and the 95th percentile of the shuffled distribution was used to define a
204 threshold (CC_{95-sh}) of acceptance for the functional links (i.e. only the links with $CC > CC_{95-sh}$ were
205 retained for further analysis). From the significant links, we built a functional graph and we
206 quantified the mean path length (i.e. mean of the shortest paths between pairs of electrodes, MPL).
207 Importantly, in the calculation of the MPL among electrodes belonging to a certain group we
208 allowed the paths to be constituted of electrodes falling outside the specified group. For instance,
209 paths among electrodes of a given cluster could extend beyond the cluster itself.

210

211 **Conventions for plotting the results of Tukey's HSD multiple comparison's test.**

212 In Fig. 2b, 3a-b, 5b-c, 6c-d-e-f and 7a-b the one-way ANOVA test was used to assess significant
213 differences ($p < 0.05$) when two or more groups were compared. The ANOVA was then followed by
214 Tukey's HSD multiple comparison test to determine which group of data did not show significantly
215 different means among groups. The symbols $\{ > < ^ \vee = - \# \}$ were used to mark the data groups
216 (among those listed on the x axis) that have similar means between groups, according to Tukey's

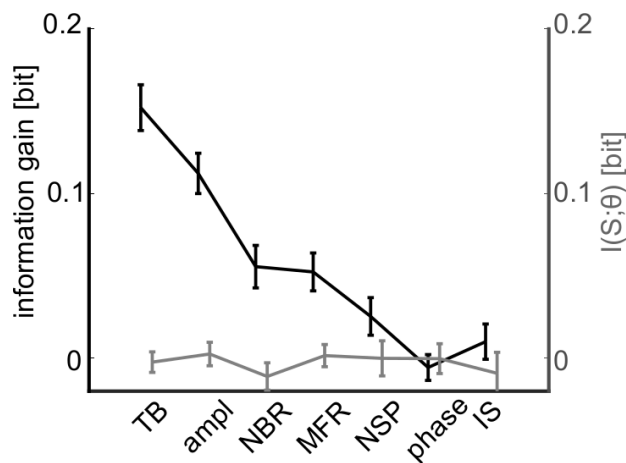
217 HSD multiple comparison test, $p < 0.05$. For example, in Fig.6C the symbol $>$ indicates that the
218 mean information across PC2 and PC3 is not significantly different, and the symbol $<$ indicates that
219 the mean information across PC3, PC4, PC5, PC6, PC7 and CAT is not significantly different, etc.

220

221

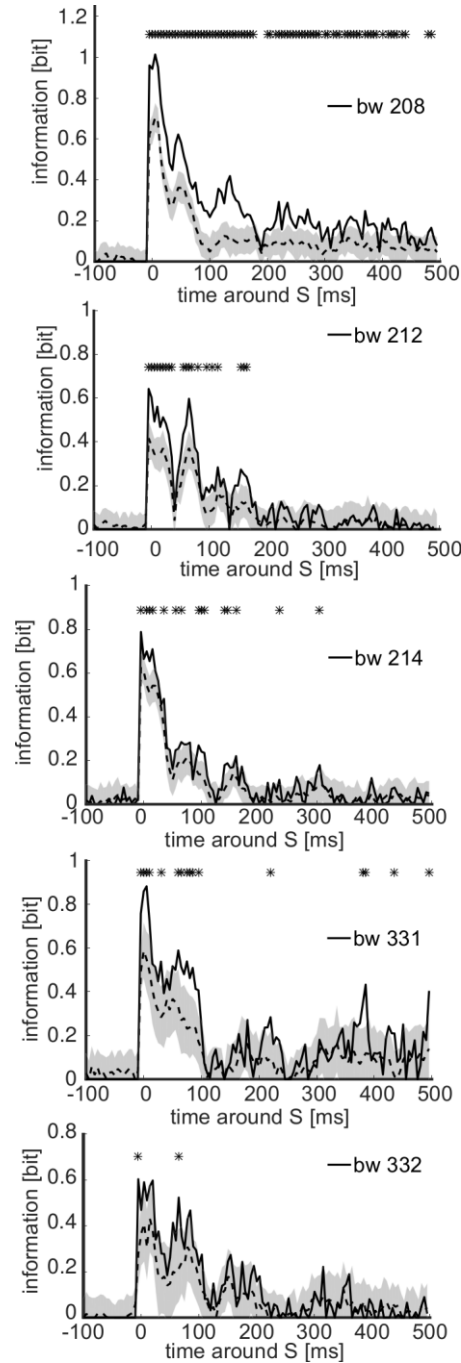
222

223 **Supplementary Figures**



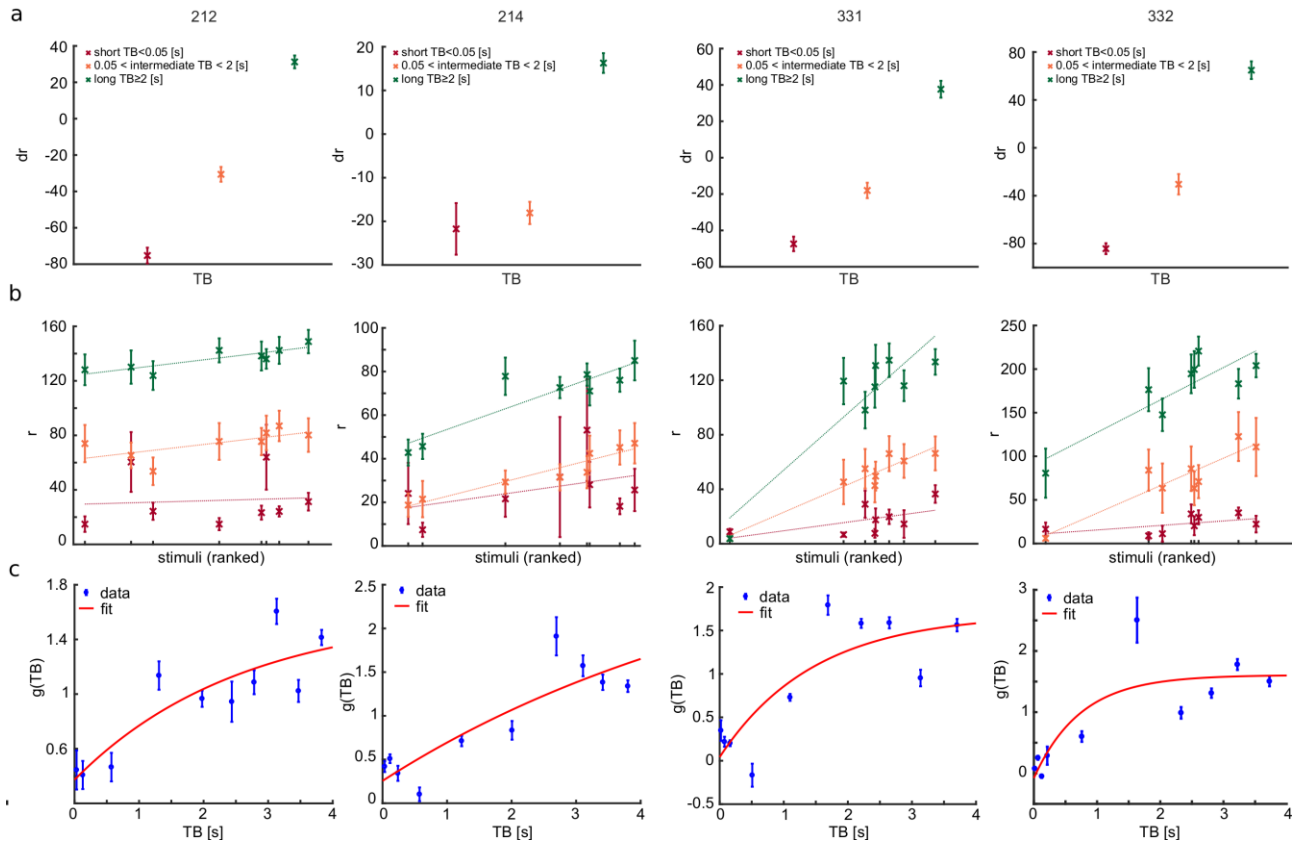
Supplementary Figure S1. The pre-stimulus state variables θ carry information about stimuli only through the synergistic interaction with the post-stimulus responses R . None of the tested state variables carry *per se* stimulus information ($I(S;\theta)$, grey line). When considering θ together with the post-stimulus responses R in the same trial, we measured an increase in the information about stimuli (information gain, black line). The mean $\pm$ SEM across all experiments in basal condition are reported. The reported response representation is MUA.

224



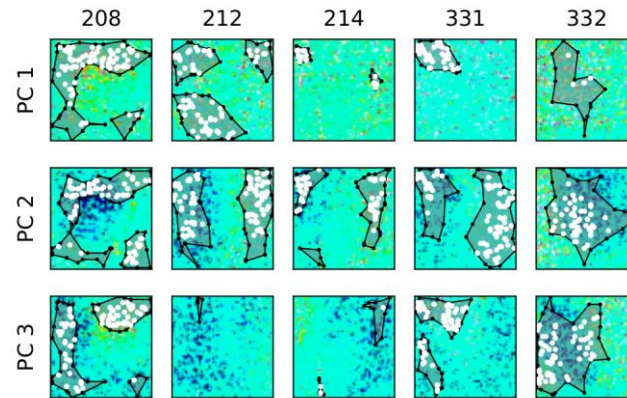
Supplementary Figure S2. Time course of information for different experiments in basal conditions. For each experiment, black solid line is the information $I(S; R, \theta)$ about stimuli carried by the joint response features $R=MUA$ and state variables $\theta=TB$, black dashed line is the information $I(S; R)$ between stimuli and responses and grey area delimits the 5th and 95th percentiles of the distribution $I(S; R, \theta_{SH})$ obtained with $N=200$ random permutations of state variables θ across trials. Stars indicate time points in which $I(S; R, \theta)$ is significantly higher than

$I(S; R)$ (one-tailed permutation test, $p < 0.05$, FDR corrected).

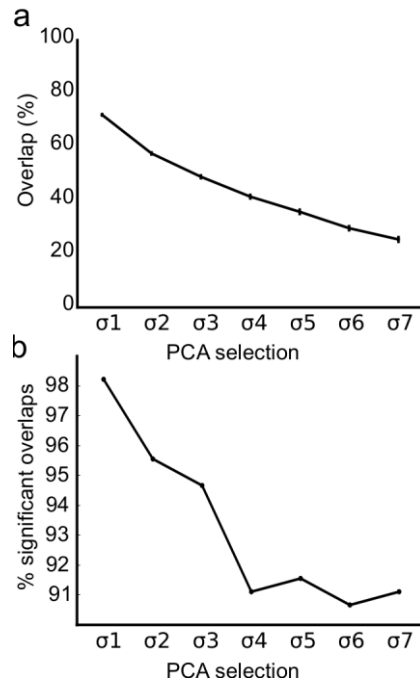


Supplementary Figure S3. Modeling single trial response as a function of state variable TB.

Panels in different columns refer to results in different recording sessions (results for the example session 208 are reported in Fig. 4). **(a)** Mean $\pm$ SEM across trials of trial-to-trial variability dr of MUA as a function of TB, computed in [0 100] ms time interval after stimulus. State variable values are divided in three intervals: low TB if $TB < 0.05$ s (red point), high TB (green point) if $TB \geq 2$ s, intermediate TB otherwise (yellow point). **(b)** Single trial response r as a function of the mean response at fixed stimulus $\langle r \rangle_s$. Mean $\pm$ SEM across trials are reported for long, intermediate and short TB (lines represent best linear fits). **(c)** Mean $\pm$ SEM of $g(TB)$ computed in a 100 ms time window after the stimulus as a function state variable TB binned in 10 equi-populated intervals. Red line shows bi-exponential fit to the data points.

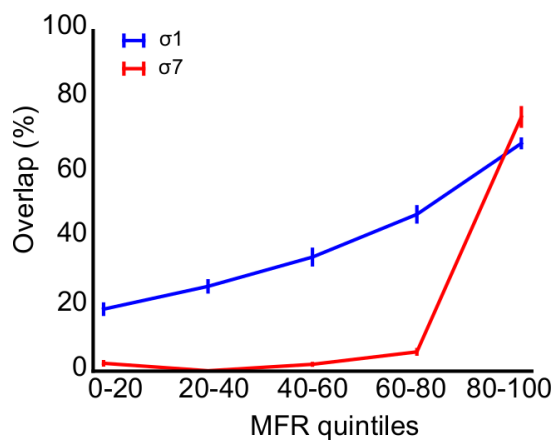


Supplementary Figure S4. The PCA reveals that cell cultures are characterized by specific regions in the network that respond more actively to electrical stimulation. The first three eigenvectors of the PCA decomposition are reported for five experiments (208,212,214,331,332) together with the electrodes selected with the σ_3 criteria (black and white circles). The shaded areas delimit the clusters of selected electrodes (black, on the border; white, within the area). The weights of the electrodes in the PCA maps are encoded by the colors (blue, negative; red positive). For a given experiment the shaded areas are quite comparable across the different PCs.



Supplementary Figure S5. The selected electrodes across different PCs are the same. (A)

The overlaps (i.e. intersections) of the electrodes selected for different PCs decrease with increasing selection thresholds (from σ_1 to σ_7). (B) The overlap decreases with the selection criteria (from ~70%, for σ_1 , to ~20%, for σ_7) but it is highly significant for most the computed overlaps (i.e. for more than the 90% of the computed overlaps, $p < 0.001$, hyper-geometric test).



Supplementary Figure S6. The selected electrodes correspond to the most spontaneously

firing ones. Plot of the mean $\pm$ sem over experiments of the overlap between the selected electrodes (i.e. thresholding the PC weights) in different quintiles of the distribution of mean firing

234 rates (MFR) within each experiment. The selected electrodes with σ_7 (~300 electrodes) have
235 approximately 80% overlap with the electrodes in the highest quintile of firing.

236 References

- 237 1. Banker, G. & Goslin, K. in *Cellular and molecular neuroscience series* (MIT Press, 1998).
238 doi:36877411
- 239 2. Berdondini, L. *et al.* Active pixel sensor array for high spatio-temporal resolution
240 electrophysiological recordings from single cell to large scale neuronal networks. *Lab Chip*
241 **9**, 2644 (2009).
- 242 3. Bologna, L. L. *et al.* Low-frequency stimulation enhances burst activity in cortical cultures
243 during development. *Neuroscience* **165**, 692–704 (2010).
- 244 4. Kermany, E. *et al.* Tradeoffs and constraints on neural representation in networks of cortical
245 neurons. *J. Neurosci.* **30**, 9588–96 (2010).
- 246 5. Jimbo, Y., Kawana, A., Parodi, P. & Torre, V. The dynamics of a neuronal culture of
247 dissociated cortical neurons of neonatal rats. *Biol. Cybern.* **83**, 1–20 (2000).
- 248 6. Gandolfo, M., Maccione, A., Tedesco, M., Martinoia, S. & Berdondini, L. Tracking burst
249 patterns in hippocampal cultures with high-density CMOS-MEAs. *J. Neural Eng.* **7**, 56001
250 (2010).
- 251 7. Nieuws, T., Di Marco, S., Maccione, A., Amin, H. & Berdondini, L. Investigating cell culture
252 dynamics combining high density recordings with dimensional reduction techniques. *Proc.*
253 *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS* 3759–3762 (2015).
254 doi:10.1109/EMBC.2015.7319211
- 255 8. Panzeri, S., Senatore, R., Montemurro, M. A. & Petersen, R. S. Correcting for the sampling
256 bias problem in spike train information measures. *J. Neurophysiol.* **98**, 1064–72 (2007).
- 257 9. Panzeri, S. & Treves, A. Analytical estimates of limited sampling biases in different

258 information measures. *Netw. Comput. Neural Syst.* **7**, 87–107 (1996).

259 10. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful
260 approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).

261 11. Quiñero, R. & Panzeri, S. Extracting information from neuronal populations:
262 information theory and decoding approaches. *Nat. Rev. Neurosci.* **10**, 173–85 (2009).

263 12. Schneidman, E., Bialek, W. & Berry, M. J. Synergy, Redundancy, and Independence in
264 Population Codes. *J. Neurosci.* **23**, 11539–11553 (2003).

265 13. Kreuz, T., Mulansky, M. & Bozanic, N. SPIKY: a graphical user interface for monitoring spike
266 train synchrony. *J. Neurophysiol.* **113**, 3432–3445 (2015).

267 14. Garofalo, M., Nieuwenhuis, T., Massobrio, P. & Martinoia, S. Evaluation of the Performance of
268 Information Theory-Based Methods and Cross-Correlation to Estimate the Functional
269 Connectivity in Cortical Networks. *PLoS One* **4**, e6482 (2009).